Bana Yalan Söylediler



SETH STEPHENS-DAVIDOWITZ

Bana Yalan Söylediler

İnternet ve Gerçek Yüzümüz

SETH STEPHENS-DAVIDOWITZ

BANA YALAN SÖYLEDİLER

İNTERNET VE GERÇEK YÜZÜMÜZ

Koç Üniversitesi Yayınları: 165 UÇBEYLERİ

SOSYOLOJİ | VERİ BİLİM

Bana Yalan Söylediler: İnternet ve Gerçek Yüzümüz

Seth Stephens-Davidowitz

İngilizceden çeviren: Ferit Burak Aydar

Yayına hazırlayan: Berna Akkıyal

Düzelti: M. Cihangir Balkır

Kitap tasarımı: Gökçen Ergüven

Kapak tasarımı: Evan Gaffney

Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are

© 2017 by Seth Stephens-Davidowitz. This edition arranged through Kayı Agency.

© Türkçe yayın hakları: Koç Üniversitesi Yayınları, 2017

1. baskı: İstanbul, Nisan 2018

Bu kitabın yazarı, eserin kendi orijinal yaratımı olduğunu ve eserde dile getirilen tüm görüşlerin kendine ait olduğunu, bunlardan dolayı kendinden başka kimsenin sorumlu tutulamayacağını; eserde üçüncü şahısların haklarını ihlal edebilecek kısımlar olmadığını kabul eder.

Baskı: A4 Ofset Sertifika No: 12168

Otosanayi Sitesi, Donanma Sk. No: 16 Seyrantepe/İstanbul +90 212 281 64 48

Koç Üniversitesi Yayınları Sertifika no: 18318

Rumelifeneri Yolu 34450 Sarıyer/İstanbul +90 212 338 1000

kup@ku.edu.tr • www.kocuniversitypress.com • www.kocuniversitesiyayinlari.com

Koç University Suna Kıraç Library Cataloging-in-Publication Data

Stephens-Davidowitz, Seth

Bana yalan söylediler : internet ve gerçek yüzümüz = Everybody lies : big data, new data, and what the Internet can tell us about who we really are / Seth Stephens-Davidowitz ; çeviren Ferit Burak Aydar ; yayına hazırlayan Berna Akkıyal.-- İstanbul : Koç Üniversitesi, 2018.

256 pages ; 16,5x24 cm.-- Koç Üniversitesi Yayınları ; 165. Teknoloji, Sosyoloji.

ISBN 978-605-2116-45-6

1. Data mining--Social aspects. 2. Big data--Social aspects. 3. Internet--Social aspects. I. Aydar, Ferit Burak. II. Akkıyal, Berna. III. Title.

QA76.9.D343 S6820 2018

Bana Yalan Söylediler

İnternet ve Gerçek Yüzümüz

SETH STEPHENS-DAVIDOWITZ

İngilizceden çeviren: Ferit Burak Aydar

KÜY

Anneme ve Babama

İçindekiler

Önsöz 9

GİRİŞ 11
Bir Devrimin Anahatları

BİRİNCİ KISIM 27
Büyük Veri, Küçük Veri

BİRİNCİ BÖLÜM 29
Kabahat Altıncı Hissinizde

İKİNCİ KISIM 41
BÜYÜK VERİ'NİN GÜÇLERİ

İKİNCİ BÖLÜM 43
Freud Haklı mıydı?

ÜÇÜNCÜ BÖLÜM 51
Verileri Yeniden Düşünmek

DÖRDÜNCÜ BÖLÜM 85
Dijital Hakikat Serumu

BEŞİNCİ BÖLÜM 127
Yakından Bakmak

ALTINCI BÖLÜM 157
Bütün Dünya Bir Laboratuvardır

ÜÇÜNCÜ KISIM 181
Büyük Veri: Özenle Ele Alınız

YEDİNCİ BÖLÜM 183
Büyük Veri, Büyük Şema mı? Elinden Gelmeyenler

SEKİZİNCİ BÖLÜM 193
Daha Fazla Veri, Daha Fazla Sorun mu? Ne Yapmamalı?

SONUÇ 203
Kaç İnsan Başladığı Kitabı Bitirebiliyor?

Teşekkür 213
Notlar 217
Dizin 243

Önsöz

Filozoflar insanın düşüncelerini ekrana yansıtacağı söylenen "serebroskop" adlı muhayyel alet hakkında spekülasyonlara daldığı günden beri, sosyal bilimciler de insan doğasının işleyişini ortaya serecek aletler arıyorlar. Bir deneysel psikolog olarak, meslek hayatım boyunca çeşit çeşit aletin moda olduğunu gördüm; sonradan gözden düşen bu aletlerin hepsini ben de denedim: Değerleme ölçeği, tepki ~~zamanları~~, gözbebeğinin büyümesi, işlevsel nörogörüntüleme, hatta vücuduna elektrotlar yerleştirilmiş halde epilepsi krizi gelsin diye beklerken bir dil deneyinde memnuniyetle saatler harcayan epilepsi hastaları.

Ama bu yöntemlerden hiçbiri, zihne dair engelleri ortadan kaldıran bir yaklaşım sunmaz. Burada sorun kaçınılmaz bir seçim yapma zorunluluğudur. İnsanın düşünceleri karmaşık önermelerdir; *Savaş ve Barış*'ı hızlı okuma yöntemiyle hatmeden Woody Allen'ın aksine, salt "Ruslarla ilgili bir şey işte" diye düşünmeyiz. Ama tüm o girift, çokboyutlu muazzam önermeleri bilimcilerin tahlil etmesi zordur. Kuşkusuz, insanlar kalplerinden geçeni söyleyip içlerini döktüklerinde, bilinç akışlarının zenginliğini idrak ederiz; ancak monologlar hipotezleri test etme konusunda ideal veri kümeleri değildir. Diğer yandan, kolaylıkla niceliksel olarak ifade edilebilecek ölçülere (insanların kelimelere tepki süresi ya da derilerinin resimlere verdiği yanıt gibi) yoğunlaşırsak istatistik çıkarabiliriz; fakat bu yöntemle biliş yetisinin karmaşık dokusunu tek bir rakama indirgemiş oluruz. En karmaşık nörogörüntüleme yöntemleri bize bir düşüncenin üç boyutlu uzamda nasıl serildiğini anlatabilir, ama düşüncenin nelerden oluştuğunu söyleyemez.

Sanki kolay kontrol edilebilirlik ile zenginlik arasında seçim yapma zorunluluğu yeterince kötü değilmiş gibi, insan doğası bilimcilerini canından bezdiren bir şey de Küçük Sayılar Kanunu'dur: Amos Tversky ve Daniel Kahneman bir nüfusun, ne kadar küçük olursa olsun, özelliklerinin her örneklemde yansımasını bulacağına dair düşünce yanıltısına bu adı veriyorlar. Sayılarla arası en iyi olan bilimciler bile, rastgele tuhaflıkları ve çıkıntıları soyutlayarak tüm Amerikalılar ve hatta tüm *Homo sapiens* için genelleme yapmak isteyen bir çalışmanın kaç deneğe ihtiyaç duyduğuna dair, esef verici derecede kusurlu sezgilere sahiptir. Bu deneyi yapan bilimci,

örneklemi uyarına geldiği şekilde (örneğin derslerimizde ikinci sınıf öğrencilerine üç beş kuruş vererek) topladığında, mesele daha da sıkıntılı bir hal almaktadır.

Bu kitap zihin incelemelerine yönelik bütünüyle yeni bir yöntemi ele almaktadır. İnternet araştırmalarından ve çeşitli *online* yanıtlardan elde edilen Büyük Veri bir serebroskop değildir; ama Seth Stephens-Davidowitz bu verilerin insan ruhuna dair eşi menendi bulunmaz bir içgörü sunduğunu gösteriyor. İnsanlar klavyelerinin mahremiyetinde en tuhaf şeyleri itiraf ederler, bazen (flört sitelerinde ya da mesleki tavsiye arayışlarında olduğu gibi) gerçek yaşama yönelik sonuçları olduğundan, ama bazen de tastamam hiçbir sonuç doğurmayacağından ötürü: İnsanlar karşılarında dehşete kapılarak tepki verecek gerçek bir insan olmadığında, bazı istek ya da korkularının yükünden kurtulabiliyorlar. Her iki durumda da insanlar sadece bir tuşa basmıyor ya da kolu çevirmiyor, bilakis patlayıcı, birleşimsel bir genişliğe sahip düşüncelerini dile getirmek için trilyonlarca karakter silsilesinden bazılarını tercih ediyorlar. Daha da önemlisi, bu dijital izleri kolayca toplanacak ve tahlil edilecek bir biçimde ortaya koyuyorlar. Hayatın bütün alanlarından geliyorlar. Gerçek hayatta uyaranları çeşitlendiren ve yanıtları tablolara döken mütevazı deneylerde yer aldıkları oluyor. Ve devasa sayılardaki bu verileri zevkle sunuyorlar.

Bana Yalan Söylediler bir kavramın kanıtlanmasından fazlasını ifade ediyor. Stephens-Davidowitz'in keşifleri ülkem ve ait olduğum tür hakkındaki peşin hükümleri kim bilir kaç kere altüst etti bilmiyorum. Donald Trump'a beklenmedik desteğin kaynağı nedir? 1976'da okurlarına çocuk sahibi oldukları için pişman olup olmadıklarını soran ve çoğunun pişman olduğunu öğrendiğinde sarsılan Ann Landers, temsili niteliği olmayan, kendi seçtiği bir örneklem tarafından yanlış mı yönlendirilmişti? 2010'ların ağdalı bir dille adlandırılmış krizinin ("filtre balonu") sorumlusu internet midir? Nefret suçlarını tetikleyen nedir? İnsanların neşeleri yerine gelsin diye internette şaka aradıkları olur mu? Hiçbir şeyin beni şoke edemeyeceğini düşünmekten haz duysam da, internetin insan cinselliği hakkında ortaya çıkardıkları (her ay çok sayıda kadının "seks yapan dolgu hayvan" aradığını keşfetmek dahil) beni epey bir şoke etmiştir. Tepki süresi, gözbebeğinin büyümesi ya da işlevsel nörogörüntülemeyle yapılan hiçbir deney bu gerçeği ortaya çıkaramazdı.

Bana Yalan Söylediler'i herkesin seveceğine eminim. Stephens-Davidowitz dinmek bilmez merakı ve insanı mest eden kıvrak zekâsıyla 21. yüzyılda sosyal bilimler için yeni bir yola işaret ediyor. İnsanın takıntılarına açılan bu uçsuz bucaksız büyüleyicilikteki pencere dururken, serebroskoba ne gerek var?

—Steven Pinker, 2017

GİRİŞ

Bir Devrimin Anahatları

Kesin kaybedecek, dediler.

Cumhuriyetçilerin 2016'daki başkan adayı ön seçimlerinde, anket şirketleri Donald Trump'a hiç şans vermiyorlardı. Sonuçta, Trump bir sürü azınlığa hakaretler yağdırmıştı. Anketler ve yorumcular, bu tür çıkışları çok az Amerikalının tasvip ettiğini söylüyordu.

O dönemde birçok anket şirketi Trump'ın genel seçimlerde kaybedeceğini düşünüyordu. Birçok muhtemel seçmen onun hem tavırlarını hem de görüşlerini itici buluyordu.

Yine de, Trump'ın hem ön seçimleri hem de genel seçimi kazanabileceğine dair bazı ipuçları vardı – ama bu ipuçları internetteydi.

* * *

Bendeniz internet veri uzmanıyım. Her gün, insanların internette dolaşırken bıraktıkları dijital izlerin peşine düşüyorum. Tıkladığımız tuşlardan ya da klavyedeki harflerden hareketle, gerçekte ne istediğimizi, gerçekten ne yapacağımızı ve gerçekte kim olduğumuzu anlamaya çalışıyorum. Bu sıradışı yola nasıl çıktığımı anlatarak başlayayım.

Hikâye (sanki üstünden asırlar geçmiş gibi) 2008 ABD Başkanlık Seçimleri'yle ve sosyal bilimlerde yıllardır cevabı bulunamamış bir soruyla başlıyor: ABD'de ırkçı önyargılar ne kadar dikkate değerdir?

Barack Obama büyük bir partinin ilk Afrikalı-Amerikalı başkan adayı olarak yarışıyordu. Seçimi kazandı, üstelik kolay bir şekilde. Anketler Amerikalıların oy verme örüntülerinde ırkın bir etken olmadığını söylüyordu. Örneğin Gallup, Obama seçilmeden önce ve seçildikten sonra sayısız anket yapmıştı. Sonuç? Amerikalı seçmenlerin çoğu Barack Obama'nın siyah olmasını umursamıyordu. Seçimlerden kısa süre sonra, Berkeley'deki California Üniversitesi'nden iki tanınmış profesör, daha incelikli veri-sondaj teknikleri kullanarak, kamuoyu yoklamalarından gelen diğer verileri didiklemeye başladılar ve onlar da benzer bir sonuca vardılar.

Dolayısıyla Obama'nın başkanlığı sırasında, medyanın genelinde ve geniş akademik çevrelerde geleneksel görüş bu oldu. Medyanın ve sosyal bilimcilerin seksen küsur yıldır dünyayı anlamak için yararlandığı kaynaklar, Amerikalıların ezici çoğunluğunun Obama'nın başkan olup olmaması gerektiğini değerlendirirken siyah olmasını umursamadığını söylüyordu.

Uzun yıllar boyunca kölelikle ve ırk ayrımcısı yasalarla yaşamış bu ülke, nihayet, insanları derilerinin rengine göre yargılamayı bırakmış görünüyordu. ABD'de ırkçılık son demlerini yaşıyor gibiydi. Öyle ki bazı uzmanlar ırkçılığın ortadan kalktığı bir toplumda yaşadığımızı ilan etmişlerdi.

2012'de iktisat bölümünde yüksek lisans öğrencisiydim; hayatın içinde kaybolup gitmiş, çalışmaktan sürmenaj olmuş ve dünyanın nasıl döndüğünü, insanların 21. yüzyılda ne düşündüklerini ve neleri önemsediklerini gayet iyi bildiğinden emin, hatta küstah biriydim. Sıra bu önyargı meselesine gelince, psikoloji ve siyaset biliminde okuduğum şeyleri temel alarak, açıktan ırkçılığın Amerikalıların küçük bir kesimiyle (çoğunluğu muhafazakâr Cumhuriyetçi olan ve çoğu Güney'in iç bölgelerinde yaşayan bir azınlıkla) sınırlı olduğuna kanaat getirmiştim.

Sonra, Google Trends'le (Google Trendler) tanıştım.

2009'da epey bir debdebeyle tanıtılan bu araç, kullanıcılara farklı zamanlarda ve farklı yerlerde bir kelimenin ya da kelime öbeğinin ne kadar sık arandığını söylüyor. Eğlencelik bir araç olarak sunulmuştu; mesela arkadaş gruplarında hangi ünlünün en popüler olduğunun ya da o ara hangi modanın birdenbire sükse yaptığının tartışılmasını sağlayacak bir araçtı. Google Trends'in ilk versiyonlarında "İnsanlar doktora tezinizi bu verilerle yazmanızı istemezler," şeklinde oyunbaz bir ikaz da vardı ki, ben hemen bu tembihi teşvik olarak alıp tezimi Google Trends'le yazmaya karar verdim.*

Öte yandan, Google arama verileri "ciddi" akademik araştırmalar için düzgün bir bilgi kaynağı gibi görünmüyordu. Anketlerin aksine, Google arama verileri insan psikesini anlamamıza yardımcı olsun diye oluşturulmuş değildi. Google insanlar

---

* Verilerimin çoğunun kaynağı Google Trends. Fakat bu araç sadece farklı araştırmaların göreli sıklığını karşılaştırmaya izni verip, belli bir araştırmanın kesin sayısını vermediğinden, ben genellikle tamamlayıcı olarak Google Adwords'ten (Reklam kelimeleri) de yararlanıyorum ve bu sayede her araştırmanın ne kadar sıklıkla yapıldığını kesinkes öğreniyorum. Birçok örnekte, Google Trends'i temel alan kendi algoritmamın da yardımıyla tabloyu daha da netleştirebiliyorum. Bu algoritmayı "Essays Using Google Data" başlıklı tezimde ve *Journal of Public Economics*'te çıkan "The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data" başlıklı yazımda anlatmıştım. Tezim, yazının linki ve bu kitaptaki tüm özgün araştırmalarda kullanılan verilerin ve kodların eksiksiz bir izahı internet sitem sethsd.com'da yer alıyor.

dünya hakkında yeni bilgiler edinsinler diye icat edilmişti, araştırmacılar insanlar hakkında bir şeyler öğrensinler diye değil. Ama internette bilgi ararken ardımızda bıraktığımız izler son derece bilgilendirici olabilir.

Başka bir deyişle, insanların bilgi arayışı kendi içinde bir bilgidir. Görünen o ki, nerede ve ne zaman olgu, alıntı, şaka, yer, insan, şey ya da yardım aradıkları bize gerçekte ne düşündükleri, gerçekte ne istedikleri, gerçekte nelerden korktukları ve gerçekte ne yaptıkları hakkında kimsenin tahmin edemeyeceği kadar çok şey anlatabilir. İnsanların zaman zaman Google'da bir şey aramaktan ziyade Google'la sırlarını paylaştıklarını düşününce, bu tespitin doğruluğu iyice açığa çıkıyor: "Patronumdan nefret ediyorum." "Sarhoşum." "Babam bana vurdu."

Her gün dikdörtgen şeklindeki beyaz bir kutucuğun içine bir kelime ya da cümleciğin yazılması, bunun milyonlarca kişi tarafından yapıldığı düşünülürse, hakikatin küçük de olsa bir parçasının izlerini belirgin kılıyor ve nihayetinde derin gerçeklikler ortaya serilmiş oluyor. Google Trends'e yazdığım ilk kelime "Tanrı"ydı. Google'da Tanrı kelimesini en fazla arayan eyaletlerin Kitabı Mukaddes Kuşağı diye bilinen Alabama, Mississippi ve Arkansas olduğunu öğrendim. Bu aramaların en sık yapıldığı gün Pazar'dı. Hiçbiri şaşırtıcı değildi, ama arama verilerinin ortaya böyle net bir örüntü serebiliyor olması ilgi çekiciydi. New Yorklu basketbol takımı "Knicks"i yazdığımda en çok New York City'de google'landığını gördüm. Bu da şaşırtıcı değildi. Sonra kendi adımı yazdım. "Üzgünüz" diyordu Google Trends, "aramanız burada görüntülenecek kadar veriye sahip değil." Öğrendim ki, Google Trends ancak çok sayıda insan aynı aramayı yaptığında veri sağlıyormuş.

Ama Google aramalarının gücü, bize Tanrı'nın ABD'nin güneyinde popüler olduğunu, Knicks'in New York City'de popüler olduğunu ya da benim herhangi bir yerde popüler olmadığımı söyleyebilmesinden ileri gelmiyor. Her araştırma, her kamuoyu yoklaması size bunu söyleyebilir. Google verilerinin gücü insanların başka birilerine söylemek istemedikleri şeyleri dev arama motoruna yazmalarından kaynaklanıyor.

Örneğin seks (kitabın ileriki sayfalarında çok daha ayrıntılı irdeleyeceğim bir konu). Seks yaşamlarımıza ilişkin hakikati anketlerden öğrenemeyiz. Amerikalıların davranışlarına dair en etkili ve yetkili bilgi kaynaklarından biri olarak görülen General Social Survey'in verilerini tahlil ettim. Bu araştırmaya göre, heteroseksüel seks söz konusu olduğunda, kadınlar yılda ortalama elli beş kez seks yaptıklarını ve bunun yüzde on altısında kondom kullandıklarını söylüyorlar. Buna göre yılda 1,1 milyar kondom kullanılması gerekiyor. Ama heteroseksüel erkeklerin beyanına

bakılırsa, her yıl 1,6 milyar kondom kullanılıyor. Oysa bu rakamların ister istemez aynı olması gerekirdi. O halde, kim doğru söylüyor, kadınlar mı erkekler mi?

Anlaşılan o ki, hiçbiri. Tüketici davranışlarını araştıran küresel bilgi ve ölçüm şirketi Nielsen'e göre, her yıl satılan kondom sayısı 600 milyondan azdır. Dolayısıyla herkes yalan söylüyor; tek fark, kimin daha çok yalan söylediğidir.

Yalan söylemek aslında yaygın bir alışkanlıktır. Hiç evlenmemiş erkekler yılda ortalama yirmi dokuz kondom kullandıklarını iddia ediyorlar: Toplama vurulduğunda, bu, ABD'de evli ve bekâr insanların toplamda satın aldıkları kondom sayısından daha fazladır. Evli insanlar da muhtemelen kaç kez seks yaptıkları sorulduğunda abartılı bir cevap veriyorlardır. Ortalamaya vurulduğunda, altmış beş yaşın altındaki evli erkekler haftada bir kez seks yaptıklarını söylüyorlar. Sadece yüzde biri son bir yılda hiç seks yapmadığını söylüyor. Evli kadınlar biraz daha az seks yaptıklarını söylüyorlar, ama aradaki fark çok değil.

Google aramaları, evlilikte seks konusunda çok daha cansız ve bence çok daha doğru bir tablo sunuyor. Google'da evlilik konusundaki şikâyetlerin başında seks yapmamak geliyor. "Sekssiz evlilik" aramaları "mutsuz evlilikler"den üç buçuk kat, "aşksız evlilikler"den sekiz kat daha fazladır. Evli olmayan çiftlerin bile seks yapmadıkları için epey sık yakındıklarını görüyoruz. Google'da "sekssiz ilişkiler"den daha çok aranan tek bir başlık bulunuyor: "İlişkide taciz". (Bu verilerin hepsinin anonim olduğunu belirtmeliyim. Google elbette hiçbir bireyin aramaları hakkında veri sunmuyor.)

Google aramaları, anketlerin sunduğu ırkçılığın ortadan kalktığı ütopyavari ortamdan çarpıcı derecede farklı bir ABD tablosu çiziyordu. "Nigger" ["Zenci"] kelimesini ilk kez Google Trends'e yazdığım günü hatırlıyorum da, ne kadar safmışım. Bunun ne kadar tepki yaratan, rahatsız edici bir kelime olduğu dikkate alındığında, aramada karşıma pek bir şey çıkmayacağından çok emindim. Heyhat, yanılmışım! ABD'de "nigger" ya da çoğulu "niggers" kelimesi ile "migren", "ekonomist" ve "Lakers" kelimeleri hemen hemen aynı sayıda aranıyormuş. Acaba rap şarkılarının sözleri arama sonuçlarını çarpıtıyor olabilir mi diye düşündüm. Hayır! Rap şarkılarında neredeyse her zaman nigger'ın bozulmuş hali olan "nigga(s)" kelimesi kullanılıyordu. O halde, Amerikalıların "nigger" kelimesini aramalarının arkasındaki saik neydi? Sık sık, Afrikalı-Amerikalılarla dalga geçen şakalar arıyorlardı. Aslında, "nigger" kelimesiyle yapılan aramaların yüzde yirmisinde "şaka" kelimesi de vardı. Diğer bir yaygın arama da "aptal zenciler" ve "zencilerden nefret ediyorum" idi.

Her yıl bu şekilde milyonlarca arama yapılıyor. Çok sayıda Amerikalı evinin mahrem alanında insanı hayrete düşürecek derecede ırkçı aramalar yapıyor. Şahsen araştırdıkça daha da rahatsız edici bilgilere eriştim.

Obama'nın ilk kez seçildiği gece, yorumların çoğunda Obama övülürken ve seçilmesinin tarihi niteliği vurgulanırken, Google'da yapılan her yüz Obama aramasından kabaca birinde "kkk" [Ku Kulux Klan] ya da "nigger" kelimesi de aranmıştı. Belki bu oran çok yüksek gelmeyebilir, ama seçimde kazanma şansı düşük olan, güzel bir aileye sahip bu adamın dünyanın en güçlü işini devralmakta olduğu esnada bir Google aramasına konu edilmesinin ırkçılıkla alakası olmayan binlerce nedenini bir düşünün. Seçim gecesindeki aramaların ve ABD'de şaşırtıcı derecede popüler olan beyaz milliyetçisi Stormfront web sitesine yeni üyeliklerin sayısı normalin on kat üstündeydi. Bazı eyaletlerde "zenci başkan" araması "ilk siyah başkan"dan daha fazlaydı.

Geleneksel kaynaklarda görülmeyen, ama insanların yaptıkları aramaların bariz şekilde ortaya koyduğu bir karanlık ve nefret söz konusuydu.

Bu araştırmaların yapıldığı bir toplumda ırkçılığın önemsiz bir etken olduğunu söylemek zordur. 2012'de Donald J. Trump'ı esasen bir işadamı ve *reality show* programcısı olarak tanıyordum. Dört yıl sonra ciddi bir başkan adayı olacağına herkes gibi ben de ihtimal vermiyordum. Ama Google'daki bu çirkin aramaları göçmenlere saldırarak, hınç ve öfke nöbetleri geçirerek çoğu zaman insanların en berbat temayüllerine oynayan bir adayın başarısıyla bağdaştırmak zor değil.

* * *

Google aramaları bize ırkçılığın görüldüğü yerler konusunda düşündüklerimizin de çoğunlukla yanlış olduğunu söylüyordu. Anketler ve geleneksel yaklaşım, günümüzde ırkçılığın ağırlıklı olarak Güney'de ve çoğunlukla da Cumhuriyetçiler arasında görüldüğünü telkin ediyordu. Ama ırkçı aramaların en yüksek olduğu yerler arasında New York'un kuzeyi, Ohio'nun doğusu, sanayi merkezi Michigan ve Illinois'in kırsal kesimi ile Batı Virginia, Louisiana'nın güneyi ve Mississippi vardı. Google arama verilerinin söylediğine göre, gerçek bölünme Güney-Kuzey şeklinde değil, Doğu-Batı şeklindeydi. Mississippi'den batıya doğru ilerledikçe bu tür şeylere rastlanmaz. Ayrıca ırkçılık Cumhuriyetçilerle sınırlı değildir. Aslına bakılırsa, Cumhuriyetçilerin sayıca güçlü olduğu yerlerde yapılan ırkçı aramalar Demokratların sayıca güçlü olduğu yerlerdekinden daha fazla değildi. Başka bir deyişle, Google aramalarına bakılarak ABD'de ırkçılığın yeni bir haritasının çizilmesi mümkündü ve bu harita tahmin edebileceğinizden çok farklı görünüyordu.

Güney'deki Cumhuriyetçiler ırkçılıklarını kabul etmeye daha meyilli olabilirler; fakat aslında Kuzey'deki Demokratların pek çoğu da benzer tutumlar sergilemektedir.

Dört yıl sonra, bu harita Trump'ın siyasi başarısını açıklamakta çok önemli bir yer tutacaktı.

2012'de Google aramalarından yararlanarak geliştirdiğim bu ırkçılık haritasını Obama'nın ırkının oynadığı rolü kesin olarak yeniden değerlendirmek için kullandım. Obama, ırkçı aramaların sayısının yüksek olduğu bölgelerde Demokratların dört yıl önceki beyaz başkan adayı John Kerry kadar başarılı olamamıştı. İlişki bu bölgelere dair başka herhangi bir etkenle (eğitim düzeyleri, yaş, kiliseye gitme sıklığı ya da silah sahibi olma) açıklanmıyordu. Irkçı aramalar başka bir Demokrat adayın değil, sadece Obama'nın kötü performans sergileyeceği tahmininde bulunuyordu.

Sonuçlar ise daha büyük bir etkiyi ortaya koymuştu. Obama sadece aleni ırkçılıktan ötürü ülke çapında yüzde dört civarında oy kaybetmişti. Anketleri temel alan herhangi bir beklentiden çok daha yüksek bir sayıydı bu. Elbette Barack Obama, Demokratlar için bazı elverişli koşulların da yardımıyla bir değil iki kez başkan seçildi, ama bunu yapabilmek için geleneksel veri kaynaklarından yararlanan birinin (aslında herkesin de diyebiliriz) fark ettiğinden biraz daha fazlasını aşması gerekti. Demokratlar için o kadar da uygun koşullarda gerçekleşmeyen bir genel seçimin sonuçlarına ilişkin ipucu verecek ya da adaylığın kazanılmasına yardımcı olacak kadar ırkçı vardı.

Çalışmam ilk başta beş akademik dergi tarafından reddedildi. Dergilerin hakemlerinden birçoğu, yüksek müsaadelerinizle biraz dert yanacağım, bu kadar çok Amerikalının böyle iğrenç bir ırkçılıkla malul olduğuna inanmanın imkânsız olduğunu söylemişti. Benim araştırmam insanların söylediklerileyle uyuşmuyordu. Ayrıca Google aramaları çok tuhaf bir veri kümesi gibi görünüyordu.

Ama Donald J. Trump'ın başkanlık koltuğuna oturduğu günümüzde verilerim daha akla yatkın görünüyor.

* * *

Araştırmamı derinleştirdikçe, Google'ın anketlerde çıkmayan ama bir seçimi (ya da birçok, diğer pek çok konuyu) anlamamıza yardımcı olabilecek çeşitli bilgiler sunduğunu öğrendim.

Kimin gerçekten sandığa gidip oy kullanacağının bilgisi mevcut. Oy vermeyen yurttaşların yarısından çoğu seçimden hemen önceki anketlerde, sandığa gitme niyetinde olduğunu söylüyor ve dolayısıyla seçim tahminimizi yanıltıyorlar; oysa Google'da seçimlerden bir hafta önceki "nasıl oy verilir" ya da "nerede oy verilir"

aramaları ülkenin hangi kısımlarının seçimlere daha büyük katılım sağlayacağını kesinkes söyleyebilir.

Öyle ki kime oy vereceklerini bile öğrenmemiz mümkün. İnsanların salt ne aradıklarından hareketle hangi adaya oy vereceklerini gerçekten tahmin edebilir miyiz? Çok açık ki, en sık hangi adayların arandığını araştırmakla yetinemeyiz. Bir adayı sırf sevdiği için Google'da arayan birçok insan vardır. Benzer sayıda insan da bir adayı sırf nefret ettiği için Google'da arıyor. Bunu aklımızın bir köşesine yazıp, Los Angeles'taki California Üniversitesi'nden finans profesörü Stuart Gabriel ile birlikte insanların hangi adaya oy vermeyi planladıklarına ilişkin şaşırtıcı bir ipucu bulduk. Seçimle alakalı aramaların büyük bir kısmında her iki adayın da adı geçiyordu. 2016'da Trump ile Hillary Clinton'ın yarıştığı seçimler sırasında bazı insanlar "Trump Clinton seçimler" diye arama yapmıştı. Bazılarıysa "Clinton Trump tartışması"nın özetlerini aramıştı. Nitekim "Trump" aramalarının yüzde on ikisinde "Clinton" kelimesi de geçiyordu. "Clinton"lı aramaların dörtte birinden fazlasında da "Trump" kelimesi vardı.

Görünüşte bir anlam ifade etmeyen aramaların bize bir insanın hangi adayı desteklediğine dair bazı ipuçları sunduğunu gördük.

Bu nasıl oluyordu? Adayların yazılma sırasına göre. Araştırmamız bir insanın arama yaparken desteklediği insanın ismini diğer adayınkinden önce yazma ihtimalinin çok daha yüksek olduğunu gösteriyor.

Önceki üç seçimde, ismi ilk sırada en çok yazılan aday en fazla oyu almıştı. İşin daha da ilginç yanı, adayların Google'da hangi sırayla arandığı o eyaletin hangi yöne gideceğinin de habercisiydi.

Adayların aranma sırası, anketlerin yakalayamadığı bilgiler de içeriyor. 2012'de Obama ile Cumhuriyetçi aday Mitt Romney arasındaki seçim yarışında, duayen bir istatistikçi ve gazeteci olan Nate Silver elli eyaletin tamamındaki sonuçları doğru tahmin etmişti. Gelgelelim Google aramalarında Romney'nin genellikle Obama'dan önce yazıldığı eyaletlerde, Romney seçimlerde Silver'ın tahmin ettiğinden daha başarılı oldu. Obama'nın Romney'den önce yazıldığı aramaların daha fazla olduğu eyaletlerdeyse, Obama Silver'ın tahmin ettiğinden daha fazla oy almıştı.

Bu gösterge tam da seçmenler kendilerine yalan söyledikleri ya da anketörlere gerçek tercihlerini söylerken rahat hissetmedikleri için anketlerde atlanan bilgileri içerebiliyor. Belki de 2012'de kararsız olduğunu söyleyip, bir yandan sürekli "Romney Obama seçimleri", "Romney Obama tartışması" ve "Romney Obama seçimi" araması yapan bir kişi, en başından beri Romney'ye oy vermeyi planlıyordu.

O halde, Google ABD seçimlerini Trump'ın kazanacağını öngörmüş müydü? Doğrusu, seçim sonuçlarını tahmin edebilmek için Google'dan en iyi nasıl yararlanacağımızı öğrenmek için daha kırk fırın ekmek yememiz gerekiyor (ve daha pek çok araştırmacının bana destek vermesi gerekiyor). Bu yeni bir bilim ve sadece birkaç seçimde bu verilerden yararlandık. Ayrıca seçim sonuçlarını tahmin etme konusunda kamuoyu yoklamalarını yardımcı araç kategorisinden tamamen çıkarabilecek bir noktada olduğumuzu veya bir gün gelip o noktaya varabileceğimizi kesinlikle iddia etmiyorum.

Ama internette birçok noktada, Trump'ın kamuoyu yoklamalarında tahmin edilenden daha başarılı olacağına dair birçok emare kesinlikle vardı.

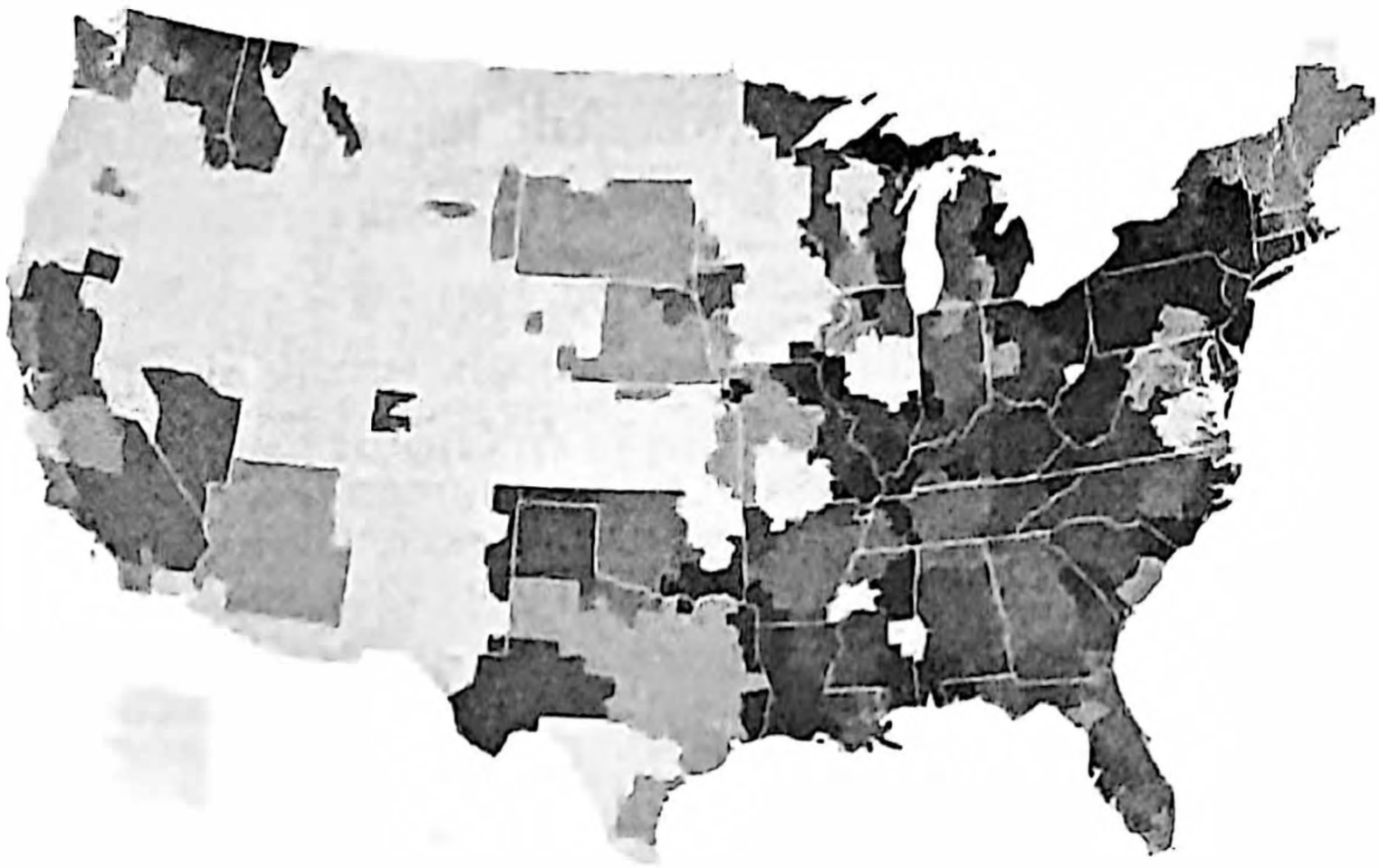
Genel seçimler sırasında seçmenlerin Trump için elverişli bir kitle olabileceğine dair ipuçları bulunuyordu. Siyah Amerikalılar Trump'ı durdurmak için akın akın sandığa gideceklerini söylüyorlardı. Ama siyahların yoğun olarak yaşadığı bölgelerde oy verme konusunda bilgi almak için yapılan Google aramaları çok düşüktü. Seçim günü Clinton siyahların düşük katılımından ötürü ağır yara alacaktı.

Kararsız olduğunu söyleyen seçmenlerin Trump'tan yana oy kullanacaklarına dair işaretler bile vardı. Gabriel ile birlikte, Clinton'ın kazanması beklenen Ortabatı'daki kilit eyaletlerde "Trump Clinton" aramalarının "Clinton Trump" aramalarından daha fazla olduğunu tespit ettik. Sonuçta, Trump zaferini bu eyaletlerde anketlerde görülene kıyasla çok büyük başarı elde etmiş olmasına borçluydu.

Ama iddiam o ki, Trump'ın –öncelikle başkan adaylığı seçimlerinde olmak üzere– başarılı bir aday olacağına dair esas ipucu, benim Obama çalışmamın ortaya çıkardığı gizli ırkçılıkla alakalıdır. Google aramaları, kamuoyu araştırmacılarının yıllardır gözden kaçırdıkları, azımsanmayacak sayıda Amerikalı arasında bir husumet ve nefret olduğunu ortaya koyuyordu. Arama verilerinin ortaya çıkardığı dünya ile kamuoyu yoklamalarından hareket eden akademisyenlerin ve gazetecilerin yaşadığımızı düşündükleri dünya çok farklıydı. Arama verileri, ortada tehlikeli, korkutucu ve yaygın bir öfkenin olduğuna işaret ediyordu; gereken tek şey, adaylardan birinin bu öfkenin sözcülüğüne soyunmasıydı.

İnsanlar sık sık yalan söylerler, hem kendilerine hem de başkalarına. 2008'de Amerikalılar kamuoyu yoklamalarına artık ırkı umursamadıklarını söylemişlerdi. Sekiz yıl sonra, Donald J. Trump adında, beyaz Amerikalıların katillerinin çoğunun siyahlar olduğuna dair bir yalanı 'retweet'leyen, bir mitinginde Siyahların Yaşamları Değerlidir protestocularından birine müdahale etmiş destekçilerine arka çıkan ve Ku Klux Klan'ın sabık liderlerinden birinin desteğini istemeye iste-

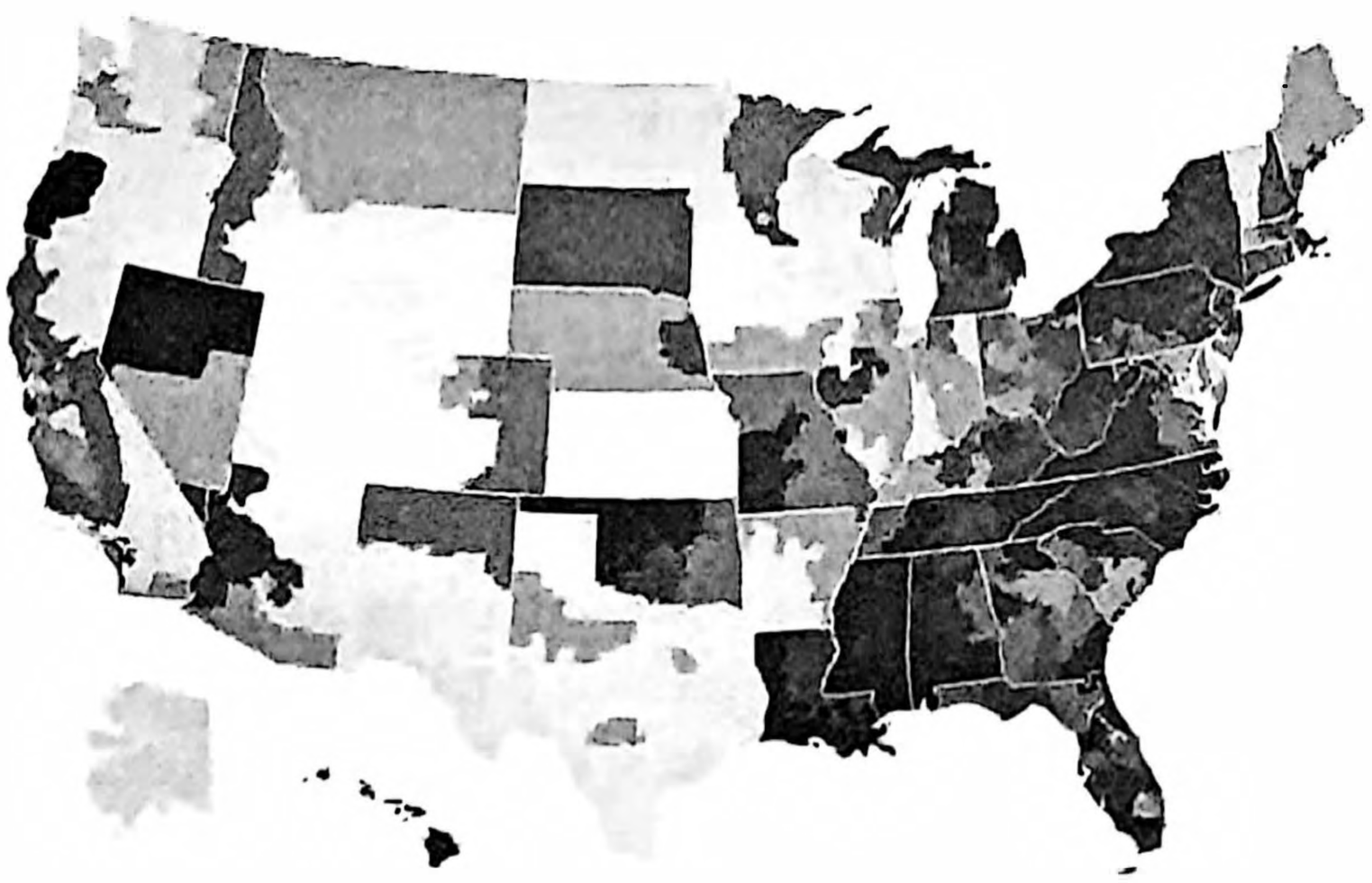
Irkçı Arama Oranları



Arama Oranı

En alt %20 En üst %20

Cumhuriyetçi başkan adayı seçimlerinde Donald Trump'a destek



Trump'ı desteklediği tahmin edilen Cumhuriyetçi seçmenler

%28 %31 %34 %37

meye reddeden bir adamı başkan olarak seçtiler. Barack Obama'ya zarar veren gizli ırkçılık Donald Trump'ın işine yaramıştı.

Nate Silver başkanlık adayı seçimlerinin başlarında, Trump'ın kazanma şansının sıfır olduğunu iddia ederek gündeme oturmuştu. Başkanlık adayı seçimleri ilerleyip de Trump'ın geniş bir destek topladığı giderek belli olunca, Silver olayın

gelişimini yorumlamak üzere verilere bakmaya karar vermişti. Trump nasıl olur da bu kadar başarılı olurdu?

Silver, Trump'ın en başarılı olduğu bölgelerin haritasını oluşturduğunda tuhaflığı fark etti. Trump hem Güney'de hem de Kuzeydoğu'nun ve sanayi bölgesi olan Ortabatı'nın belli kısımlarında başarılıydı. Batı'daysa çok başarısızdı. Silver bu haritayı açıklayacak değişkenler aramaya başladı. İşsizlik olabilir miydi? Din? Silah sahibi olma? Yoksa sebep göç oranları ya da Obama karşıtlığı mıydı?

Silver, Donald Trump'ın Cumhuriyetçilerin başkan adayı seçimlerindeki desteğiyle en çok örtüşen etkenin, benim dört yıl önce keşfettiğim ölçüt olduğunu buldu. Trump'ın en kitlesel destek gördüğü bölgeler Google'da en fazla "nigger" araması yapılan yerlerdi.

* * *

Dört yıldır hemen her günümü Google verilerini tahlil ederek geçirdim. Buna Google'da veri bilimcisi olarak çalıştığım dönem de dahil: Google bu ırkçılık araştırmamı öğrenince beni işe almıştı. Ayrıca bu verileri *New York Times*'ta veri gazetecisi ve köşe yazarı olarak araştırmaya devam ediyorum. Bilgi akışı hiç durmuyor. Akıl hastalığı; insan cinselliği; çocuk istismarı; kürtaj; reklamcılık; din; sağlık. Gördüğünüz gibi, bunlar hiç de önemsiz konular değil; yirmi-otuz yıl önce var olmayan bu veri kümesi her birine ilişkin şaşırtıcı yeni veriler sunuyor. İktisatçılar ve diğer sosyal bilimciler daima yeni veri kaynakları peşinde koşuyorlar, bu yüzden eğip bükmeden söyleyeyim: Google aramalarının insan psikesiyle ilgili gelmiş geçmiş en önemli veri kümesi olduğuna artık kaniyim.

Ne var ki bu veri kümesi internetin dünyamızı anlamak için sunduğu yegâne araç değildir. Çok geçmeden, dijital alanda başka altın madenleri olduğunu da fark ettim. Tüm Wikipedia'yı indirdim, Facebook profillerini gezdim ve Stormfront sitesini didik didik ettim. Ayrıca internetteki en büyük porno sitelerinden biri olan PornHub, dünyanın dört bir tarafından ismi saklı tutulan kişilerin aramalarına ve video görüntülemelerine ait verilerin tamamını benimle paylaştı. Başka bir deyişle, artık Büyük Veri adı verilen şeyin çok derinlerine dalmış durumdayım. Ayrıca benim gibi bu yeni alanları araştıran onlarca başka insanla (akademisyen, veri gazetecisi ve müteşebbisle) mülakatlar yaptım. Bu araştırmaların birçoğunu burada ele alacağım.

Ama öncelikle bir itirafta bulunayım: Büyük Veri'nin ne olduğuna dair kesin bir tanım sunmayacağım. Neden? Kendi içinde muğlak bir kavram da ondan. Büyük de ne kadar büyük? 18.462 gözlem Küçük Veri iken, 18.463 gözlem Büyük Veri mi? Neyin bu kapsama gireceğine dair kapsayıcı bir görüşü benimsemiş durumdayım:

Kurcaladığım verilerin çoğu internetten olmakla birlikte, başka kaynakları da ele alacağım. Çağımızda her türlü mevcut bilginin hem niceliğinde hem de niteliğinde bir patlama yaşanıyor. Yeni bilgilerin çoğu Google'dan ve sosyal medyadan geliyor. Bunlardan bazıları bir süre dosya ve klasörlerde saklanmış olan bilgilerin dijitalleştirilmesinin bir ürünüdür. Bazılarıysa pazar araştırmasına ayrılmış kaynaklardaki artışın sonucudur. Bu kitapta ele alınan araştırmalardan bazıları hiç de devasa veri kümelerine dayanmıyor, bunun yerine verilere yönelik yeni ve yaratıcı bir yaklaşım sergiliyor: Bilgiyle dolup taşan bir çağda can alıcı yaklaşımlar.

O halde, Büyük Veri tam olarak ne kadar güçlüdür? Belirli bir günde internete dağılan tüm bilgileri bir düşünün: İnternette ne kadar bilgi olduğuna dair elimizde bir sayı var aslında. 21. yüzyılın başlarında ortalama bir günde insanlar 2,5 trilyon bayt veri yaratıyorlar.

Ve bu baytlar ipucu mahiyetindedir.

Kadının biri bir perşembe günü sıkıntıdan patlarken, Google'a "bel altı olmayan komik şakalar" yazıyor. E-postasını kontrol ediyor. Twitter'a giriyor. Google'a "Zenci şakaları" yazıyor.

Adamın birinin canı sıkkın. "Depresyon belirtileri" ve "depresyon hikâyeleri" diye google'lıyor. Solitaire oynuyor.

Bir kadın Facebook'ta bir arkadaşının nişanlandığını öğreniyor. Bekâr olan kadın, arkadaşını blokluyor.

Adamın biri arama motoruna NFL ve rap müzik konularında yazarken, birdenbire "rüyada öpüşen erkekler görmek normal midir?" sorusunu giriyor.

Bir kadın "en minnoş 15 kedi"nin gösterildiği bir Buzzfeed hikâyesine tıklıyor.

Bir adam kediler hakkındaki aynı hikâyeyi görüyor. Ama ekranında "en müthiş 15 kedi" yazıyor. Tıklamıyor.

Bir kadın "Oğlum dâhi mi?" yazıp google'lıyor.

Bir adam, "Kızım nasıl kilo verir?" yazıp google'lıyor.

Kadının biri en iyi altı kadın arkadaşıyla tatile çıkıyor. Arkadaşlarının hepsi ne kadar çok eğlendiklerini söyleyip duruyor. Ama o çaktırmadan Google'a, "kocandan uzaktayken yalnızlık" yazıyor.

Bir adam, demin bahsi geçen kadının kocası, en iyi altı erkek arkadaşıyla tatilde. "Karınızın sizi aldattığına dair belirtiler" yazıp google'lıyor.

Bu verilerden bazıları normalde kimseye itiraf edilemeyecek bilgiler içeriyor. Hepsini toplar, somut bir bireyin arzularını, korkularını ya da davranışlarını ele vermeyecek şekilde saklı tutup birtakım veri bilimlerini işe koşarsak, insanlara

(davranışlarına, arzularına ve doğalarına) dair yeni bir görüş elde ederiz. Aslında, kulağa abartılı gelmesi riskine rağmen, yaşadığımız dijital çağda giderek daha fazla ulaşılabilir olan yeni verilerin insanlık anlayışımızı köklü şekilde genişleteceğine inandığımı söylemem gerek. Mikroskop bize bir damla havuz suyunda düşündüğümüzden daha çok şey olduğunu göstermişti. Teleskop gece gökyüzünde gördüğümüzü düşündüğümüzden daha fazla şey olduğunu gösterdi. Yeni, dijital verilerse insan toplumunda gördüğümüzü düşündüğümüzden daha fazla şey olduğunu gösteriyor. İnternet, çağımızın mikroskobu ya da teleskobu olabilir: Önemli, hatta devrimci içgörüleri mümkün kılıyor.

Bu tür beyanlarda bulunmanın bir riski daha var: Sadece abartılı görünmekle kalmayıp aynı zamanda revaçta olan bir risk. Birçok insan Büyük Veri'nin gücü hakkında büyük iddialarda bulunuyor. Ama kanıttan yana sıkıntı var.

Bu durum Büyük Veri'ye şüpheyle yaklaşanların, ki böyle epey insan var, daha büyük veri kümeleri arayışını reddetmelerine vesile olmuştur. Deneme yazarı ve istatistikçi Nassim Taleb şöyle diyor: "Ben Büyük Veri'de tek kelime bilgi olmadığını söylemiyorum. Bir yığın bilgi var. Sorun ve esas mesele, giderek büyüyen bir samanlıkta iğne aramaya benzemesi."

Dolayısıyla bu kitabın başlıca hedeflerinden biri Büyük Veri'yle neler yapılabileceğinin eksik de olsa bir kanıtını sunmaktır: Her geçen gün daha da büyüyen bu samanlıkta iğneleri nasıl bulabiliriz, tabii eğer bulacaksak ve bulabilirsek? Büyük Veri hakkında insan psikolojisi ve davranışına dair yeni içgörüler sunan yeterli sayıda örnek aktarmayı ve bu sayede gerçekten devrimci bir olguyu ana hatlarıyla gösterebilmeyi umuyorum.

Tam şu anda, "Dur bir dakika, Seth," diyor olabilirsiniz. "Bize devrimle geliyorsun, bu büyük ve yeni veri kümeleri hakkında şiirsel şiirsel konuşuyorsun. Ama şu ana kadar, bu olağanüstü, inanılmaz, nefes kesici, çığır açıcı verilerden yararlanarak temelde iki şey söyledin: Amerika'da bir sürü ırkçı var ve insanlar, özellikle de erkekler ne kadar seks yaptıkları sorusuna abartarak cevap veriyorlar."

Yeni verilerin kimi zaman malumu ilam etmekten öteye geçmediğini kabul ediyorum. Eğer bu bulguların zaten aşikâr olduğunu düşünüyorsanız, Dördüncü Bölüm'e kadar bekleyin; zira orada erkeklerin –sıkı durun– penislerinin büyüklüğü konusunda muazzam bir kaygı ve güvensizlik duyduğuna dair Google aramalarından açık, reddedilemez kanıtlar sunacağım.

Zaten şüphelendiğiniz ama elinizde çok az kanıt olan şeylerin kanıtlanmasının değersiz olmadığını düşünüyorum. Bir şeyden şüphe duymak ayrı, onu ispatlamak ayrıdır. Ama eğer Büyük Veri'nin yapabileceği yegâne şey şüphelerinizi doğrulamak-

sa, devrimci olduğu söylenemez. Neyse ki, Büyük Veri bundan çok daha fazlasını yapabilir. Veriler dünyanın benim tahmin ettiğimin tam tersi yönde işlediğini gösteriyor. İşte size daha şaşırtıcı bulabileceğiniz bazı örnekler.

Irkçılığın temel nedenlerinden birinin ekonomik güvencesizlik ve savunmasızlık olduğunu düşünebilirsiniz. Bu durumda, insanların işlerini kaybettiklerinde ırkçılığın arttığından doğal olarak şüphelenmeniz gerekir. Oysa ırkçı aramalarda ve Stormfront üyeliklerinde görülen artışla işsizliğin artışı arasında bir ilişki yoktur.

Eğitim düzeyinin en yüksek olduğu büyük şehirlerde anksiyetenin yüksek oranlara ulaştığını düşünüyor olabilirsiniz. Kentlilerin nevrotikliği meşhur bir klişedir. Ama endişeyle alakalı Google aramalarının ("anksiyete belirtileri" ya da "anksiyete yardımı" gibi) eğitim seviyesinin düşük olduğu, alt orta gelirli ve nüfusun çoğunluğunun kırsal bölgelerde yaşadığı yerlerde genellikle daha yüksek olduğu görülüyor. New York'un kırsala doğru, daha dış bölgelerinde anksiyete aramaları New York'un merkezine kıyasla daha yüksektir.

Onlarca ya da yüzlerce insanın öldüğü bir terör saldırısının ardından kitlesel düzeyde, yaygın bir anksiyete görüleceğini düşünüyor olabilirsiniz. Terörizm, tanımı gereği, bir terör (dehşet) hissi uyandırır. Google'daki anksiyeteyle alakalı aramalara baktım. Bu aramaların Avrupa'da ya da Amerika'da 2004'ten bu yana terör saldırılarından sonraki günlerde, haftalarda ve aylarda bir ülkede ne kadar yükseldiğini test ettim. Ortalamaya vurulduğunda, anksiyeteyle alakalı aramalar ne kadar yükselmişti? Hiç. Sıfır.

İnsanların üzgünken daha sık "komik şaka" araması yaptığını düşünüyor olabilirsiniz. Tarihin en büyük düşünürlerinden birçoğu, acıdan kurtulmak için mizaha döndüğümüzü iddia etmiştir. Mizah uzun zamandır başarısızlıklarla, acılarla, hayatın kaçınılmaz hüsranlarıyla başa çıkmanın bir yolu olarak görülmüştür. Charlie Chaplin'in ifade ettiği şekliyle, "Gülme acıya ara vermedir, rahatlamadır, kuvvet verici bir ilaçtır."

Gelgelelim internette komik şaka aramalarının en az olduğu gün, insanların en mutsuz olduklarını söyledikleri Pazartesi günleridir. En az arama, bulutlu ve yağmurlu günlerde yapılmaktadır. Dahası 2013 Boston Maratonu'nda iki bombanın üç kişinin ölümüne, yüzlerce insanın da yaralanmasına yol açtığı saldırı gibi büyük bir trajedinin ardından aramalar dibe vurmaktadır. İnsanlar işler yolunda gittiğinde Google'da şaka aramaya daha meyilli oluyorlar.

Kimi zaman yeni bir veri kümesi benim aklımın ucundan bile geçmeyecek bir davranış, arzu ya da kaygıyı ortaya seriyor. Bu kategoriye giren sayısız cinsel eğilim var. Örneğin Hindistan'da "kocamın en çok istediği şey" ile başlayan aramalar

içinde en çok tekrar edenin "kocamın en çok istediği şey onu emzirmem" şeklinde olduğunu biliyor muydunuz? Bu yorum, Hindistan'da başka hiçbir ülkede olmadığı kadar yaygındır. Ayrıca kadınların erkekleri emzirdiği porno aramaları da Hindistan ve Bangladeş'te diğer bütün ülkelerden dört kat daha fazladır. Verileri görmesem böyle bir şeye hayatta inanmazdım.

Keza erkeklerin penis büyüklükleri konusunda takıntılı olduğunu duyduğunuzda pek şaşırmayabilirsiniz, ama kadınların bedenleri konusunda, Google'da ifade bulduğu haliyle, en güvensiz oldukları şey gerçekten şaşırtıcıdır. Bu yeni veriler temelinde, erkeğin penisi konusunda kaygı duymasının kadınlardaki muadili –sıkı durun!– vajinasının kokup kokmadığıdır. Kadınların cinsel organlarıyla ilgili kaygı belirten arama yapma sayısı, erkeklerin cinsel organlarıyla ilgili arama yapma sayısına neredeyse denktir. Ayrıca kadınların başlıca kaygı konusu vajinalarının kokusu ve bunun nasıl iyileştirileceğidir. Doğrusu verileri görmeden önce böyle bir bilgim yoktu.

Yeni veriler kimi zaman, aklımın ucundan bile geçmeyecek kültürel farklılıkları ortaya koyuyor. Bir örnek: Dünyanın dört bir tarafındaki erkekler eşlerinin hamile olduğunu öğrendiklerinde çok farklı şekillerde yanıt veriyorlar. Meksika'da "karım hamile" ifadesiyle birlikte en fazla yapılan arama "frases de amor para mi esposa embarazada" (hamile eşime sevgi sözleri) ve "poemas para mi esposa embarazada" (hamile eşime şiirler) oluyor. ABD'deyse arama motoruna "karım hamile, şimdi n'olcak" ve "karım hamile ne yapmalıyım" yazılıyor.

Yine de, bu kitap tuhaf olgulara ya da bir kez uygulanmış araştırmalara epey yer verse de, bunların bir derlemesinden daha fazlasıdır. Bu metodolojiler çok yeni olduğundan ve ancak ileride güçleneceğinden, bunların nasıl işlediğine ve neden çığır açıcı olduklarına dair bazı fikirler öne süreceğim. Ayrıca yeri geldiğinde Büyük Veri'nin sınırlılıklarını da kabul edeceğim.

Veri devriminin potansiyelinden duyulan coşku kısmen yersizdir. Büyük Veri'ye âşık olanların çoğu bu veri kümelerinin ne denli muazzam olduğu konusundaki övgülerinde ipin ucunu kaçırıyorlar. Veri kümeleri konusundaki bu takıntı yeni değildir. Google, Amazon ve Facebook'tan önce, yani "Büyük Veri" tabiri ortaya çıkmadan önce, Teksas'ın Dallas kentinde "Büyük ve Karmaşık Veri Kümeleri" üzerine bir konferans düzenlenmişti. Stanford'da istatistik profesörü ve Google'dan mesai arkadaşım olan Jerry Friedman 1977 yılındaki bu konferansı hiç unutmuyor. Toplantıda söz sırası seçkin bir profesöre gelir. İnanılmaz, dudak uçuklatıcı miktarda, tam beş gigabaytlık veri biriktirdiğini anlatır. Sonraki seçkin istatistikçi söz aldığındaysa, "Son konuşmacının gigabaytları var. O ne ki? Bende terabaytlar var"

der. Başka bir deyişle, iki konuşmacı da bilgiyle ne yapmak istediğini ya da hangi soruları cevaplamayı planlandığını değil, ne kadar bilgi biriktirebildiğini vurgular. "O dönem bunu komik bulmuştum," diyor Friedman, "insanların etkilenmesini bekledikleri şey veri kümelerinin ne kadar büyük olduğuydu. Bunu hâlâ görüyoruz."

Bugün birçok veri bilimcisi muazzam veri kümeleri biriktiriyor, ama bize önem taşıyan çok az şey anlatıyorlar (Knicks'in New York'ta çok popüler olduğu gibi). Birçok işletme, bilgi yığını altında boğuluyor. Terabaytları pek çok, ama büyük içgörüleri çok az. Kanımca veri kümelerinin büyüklüğüne genellikle haddinden fazla değer veriliyor. Bunun basit ama önemli bir açıklaması var. Bir sonuç ya da etki ne kadar büyükse, onu tespit etmek için gerekli gözlem sayısı o kadar azdır. Sıcak bir fırına dokunmanın tehlikeli olduğunu anlamak için bir kez dokunmanız yeterlidir. Kahvenin sizde baş ağrısına yol açıp açmadığını belirlemek içinse yüzlerce kez kahve içmeniz gerekebilir. Hangi ders daha önemlidir? Elbette, doğurduğu etkinin yoğunluğundan ötürü fırının sıcaklığı çok çabuk ve çok az veriyle kendini gösterir.

Aslında, en akıllı Büyük Veri şirketleri çoğu zaman verilerini azaltıyorlar. Google'da en önemli kararlar tüm verilerin sadece küçük bir örneklemine dayalıdır. Önemli içgörülere ulaşmak için her zaman bir dolu veriye gerek yoktur. Doğru verilere ihtiyaç vardır. Google aramalarının bu denli değerli olmasının temel nedenlerinden biri sayıca bol olmaları değildir; insanların aramalarında dürüst olmalarıdır. İnsanlar arkadaşlarına, sevgililerine, doktorlarına, anketörlere ve kendilerine yalan söylerler. Ama Google'da kendilerini utandıran bilgileri paylaşabilirler. Mesela evliliklerinde cinselliğin olmamasını, sağlık problemlerini, güvensizliklerini ve siyahlara karşı düşmanlıklarını Google ile paylaşabilirler.

En önemlisi, Büyük Veri'den içgörü çekip çıkarabilmek için doğru soruları sormanız gerekir. Nasıl ki teleskobu gece gökyüzüne rastgele doğrultup Plüton'u keşfetmesini bekleyemezseniz, aynı şekilde yığınla bilgiyi indirip insan doğasının sırlarına vakıf olmayı da bekleyemezsiniz. Ümit vaat eden yerlere bakmanız gerekir: Örneğin Hindistan'da "kocam benden ... istiyor" şeklindeki aramalara.

* * *

Elinizdeki kitap Büyük Veri'nin nasıl en iyi şekilde kullanabileceğini gösterip, neden bu denli güçlü olabileceğini ayrıntılarıyla açıklayacak. Bunu yaparken de, benim ve başkalarının hâlihazırda şunları keşfettiğimizi de öğreneceksiniz:

- Kaç gey erkek var?
- Reklam vermek işe yarıyor mu?
- Amerikan Firavunu neden büyük bir yarış atıydı?
- Medya yanlı mı?

- Freudcu dil sürçmeleri gerçek mi?
- Kimler vergi kaçırıyor?
- Hangi üniversiteye gittiğinin bir önemi var mı?
- Borsayı kandırabilir miyiz?
- Çocuk büyütmek için en iyi yerler nereleri?
- Bir hikâyeyi viral yapan nedir?
- İlk randevuda nelerden bahsetmeliyim ki ikincisi olsun?

Ve çok çok daha fazlası...

Fakat tüm bunlara girmeden önce, daha temel bir soruyu tartışmamız gerekiyor: Neden verilere ihtiyaç duyuyoruz? Bunu tartışmak için de, sizi büyükannemle tanıştıracağım.

Birinci Kısım

Büyük Veri, Küçük Veri

BİRİNCİ BÖLÜM

Kabahat Altıncı Hissinizde

Eğer otuz üç yaşındaysanız ve Şükran Günü'ne birkaç sene üst üste sevgiliniz olmadan katıldıysanız, hiç merak etmeyin "Ee, evlilik ne zaman?" sorusu er geç gündeme gelecektir. Ve hemen herkes nasıl bir eş bulmanız gerektiği konusunda fikir beyan edecektir.

Kız kardeşim: "Seth'e kendisi gibi deli bir kız lazım."

Erkek kardeşim: "Deli misin? Seth'e onu dengeleyecek normal bir kız lazım."

Annem: "Seth deli falan değil."

Babam: "Delisin sen! Elbette Seth deli."

Ve sonra, o utangaç, yumuşak yumuşak konuşan büyükannem birdenbire sofradaki sessizliğini bozup konuşmaya başlıyor. Gürültücü, saldırgan New Yorkluların sesleri kesiliyor ve tüm gözler kısa saçlı, sarışın Doğu Avrupalı aksanını tam olarak atamamış bu yaşlı, minyon hanımefendiye çevriliyor. "Seth, sana hoş bir kız lazım. Çok güzel olması şart değil. Ama çok akıllı olmalı. İnsanlarla iyi geçinmeli. Sosyal olmalı ki, sen de işlerini yapabilesin. Mizahtan anlamalı, çünkü senin mizah anlayışın iyidir."

Bu yaşlı hanımefendinin nasihatleri neden aile fertlerim tarafından bu kadar dikkate alınıp saygı görüyor? Açıkçası, seksen yaşındaki büyükannem masadaki herkesten daha çok görmüş geçirmiş biri. Daha fazla evlilik yapmış, bazıları yürümüş, bazıları yürümemiş. Ve uzun yıllardır, doğru ilişkiler için gerekli nitelikleri sıralayıp katalog haline getirmiş. O Şükran Günü sofrasında, bu sorun özelinde, en fazla veriye erişim puanına sahip olan büyükannemdi. Büyükannem Büyük Veri'ydi.

Bu kitapta veri biliminin gizemini çözmek istiyorum. İster beğenin ister beğenmeyin, veriler tüm hayatımızda giderek daha önemli bir rol oynuyor ve bu rol daha da artacak. Gazetelerde sadece verilere geniş sütunlar ayrılıyor. Şirketler, işi yalnızca veri analizi olan ekipler istihdam ediyorlar. Yatırımcılar daha fazla veri depolayabilecek genç şirketlere on milyonlarca dolar aktarıyorlar. Regresyonun nasıl yapıldığını asla öğrenemeseniz ya da güven aralığını hesaplayamasanız bile,

okuduğunuz sayfalarda, katıldığınız iş toplantılarında, soğutucunun yanında su içerken duyduğunuz dedikodularda olsun bir sürü veriyle karşılaşacaksınız.

Bu gelişme birçok insanı endişeye sevk ediyor. Veriler insanların gözünü korkutuyor, sayılar dünyasında kolaylıkla kaybolup kafaları karışıyor. Dünyanın nicel olarak anlaşılmasının kendilerinin değil, küçük bir seçkin azınlığı oluşturan sol beyinlilerin işi olduğunu düşünüyorlar. Sayılarla karşılaşır karşılaşmaz sayfayı çevirmeye, toplantıyı bitirmeye ya da sohbetin konusunu değiştirmeye hazırlar.

Ama ben veri analizi işinde on yıldır çalışıyorum ve bu alanın en iyi isimleriyle birlikte mesai yapma fırsatı elde ettim. Öğrendiğim en önemli derslerden biri şudur: İyi veri bilimi, insanların zannettiği kadar karmaşık değildir. Aslında en iyi veri bilimi şaşırtıcı derecede sezgiseldir.

Veri bilimini sezgisel yapan nedir? Veri bilimi, özünde, örüntüleri tespit edip bir değişkenin diğerini nasıl etkileyeceğini kestirmekle alakalıdır. İnsanlar bunu her zaman yaparlar!

Büyükannemin bana verdiği ilişki nasihatini bir düşünün. Yaklaşık bir asırlık yaşamında ilişkilere dair ailesinden, arkadaşlarından ve tanıdıklarından duyduğu hikâyelerden beynine yüklediği geniş veritabanından yararlanmıştı. Analizini erkeklerin benimle birçok ortak özelliğe sahip olduğu ilişkilerin bir örneklemiyle sınırladı: Duyarlı bir mizaç, kendini başkalarından yalıtma eğilimi, mizah duyusu. Kadınlarınsa kilit özelliklerini nişan aldı: Ne kadar nazik oldukları, ne kadar akıllı oldukları, ne kadar güzel oldukları. Kadınların bu kilit özellikleriyle ilişkinin kilit bir özelliğinin bağıntısını kurdu: İyi bir ilişki mi, değil mi? Son olarak, sonuçlarını raporladı. Başka bir deyişle, örüntüler saptadı ve bir değişkenin diğerini nasıl etkileyeceğini tahmin etti. Büyükannem bir veri bilimcisiydi.

Siz de bir veri bilimcisisiniz. Çocukken ağladığınızda annenizin sizinle ilgilendiğini fark etmiştiniz. Bu bir veri bilimidir. Büyüdüğünüzde, çok fazla şikâyet ederseniz insanların sizinle daha az takılmak istediğini fark ettiniz. Bu da veri bilimidir. İnsanlar sizinle daha az takıldığında, daha mutsuz olduğunuzu fark ettiniz. Daha mutsuz olduğunuzda, eskisi kadar cana yakın olmadığınızı fark ettiniz. Eskisi kadar cana yakın olmadığınızdaysa, insanlar sizinle daha da az takılmak istiyorlardı. Veri bilimi. Veri bilimi. Veri bilimi.

Veri bilimi bu denli doğal bir şey olduğundan, en iyi Büyük Veri çalışmalarının hemen her akıllı insan tarafından anlaşılabileceğini keşfettim. Eğer bir çalışmayı anlayamıyorsanız, sorun muhtemelen çalışmadadır, sizde değil.

Azametli veri biliminin sezgisel olma eğilimine dair kanıt mı istiyorsunuz? Geçenlerde son birkaç yılda yürütülmüş en önemli çalışmalardan birine denk

geldim. Ayrıca gördüğüm en sezgisel çalışmalardan da biriydi. Sizden sadece çalışmanın önemini değil, aynı zamanda ne kadar doğal ve büyükannemsi olduğunu da düşünmenizi isteyeceğim.

Çalışmayı Columbia Üniversitesi ve Microsoft'tan bir araştırma ekibi yürüttü. Bu araştırmacılar hangi belirtilerin pankreas kanserinin habercisi olduğunu bulmak istiyorlardı. Bu hastalığa yakalananların hayatta kalma süresi beş yıldır (üstelik sadece yüzde üçü), ama erken tanı hastanın şansını iki katına çıkarabilir.

Araştırmacıların yöntemi neydi? Microsoft'un arama motoru olan Bing'in on binlerce anonim kullanıcısının verilerinden yararlandılar. "Biraz önce bana pankreas tanısı kondu" ya da "pankreas kanseri olduğumu öğrendim, ne yapmalıyım?" gibi yanlış anlaşılmaya mahal vermeyecek aramalar temelinde yakın zaman önce pankreas kanseri tanısı konmuş bir kullanıcı kodladılar.

Ardından, araştırmacılar sağlık belirtileri aramalarına baktılar. Sonradan pankreas kanseri tanısı konduğunu söyleyen az sayıdaki kullanıcıyı bu tanının konmadığı kişilerle karşılaştırdılar. Başka bir deyişle, hangi belirtiler birkaç hafta ya da ay içinde bir kullanıcının kanser tanısından bahsedeceğini haber veriyordu?

Sonuçlar çarpıcıydı. Arama kutucuğuna sırt ağrısı ve ardından derinin sararması yazmak pankreas kanserinin işaretiydi; sadece sırt ağrısı aramak ise pankreas kanserine yakalandığınız anlamına gelmiyordu. Benzer şekilde, hazımsızlık ve ardından karın ağrısı aramak pankreas kanserinin kanıtıydı, ama karın ağrısı olmadan sadece hazımsızlık aramak değildi. Araştırmacılar neredeyse hiçbiri yanlış pozitif (testin hatalı olarak hasta dediği kişiler) olmayan vakaların yüzde beş ila on beşini tespit edebilmişlerdi. Şimdi, bunun büyük bir oran olmadığını düşünebilirsiniz; ama eğer pankreas kanseriyseniz, hayatta kalma şansınızı muhtemelen ikiye katlayacak yüzde onluk bir ihtimal bile çölde vaha gibi gelecektir.

Bu incelemenin ayrıntılarını ortaya koyan yazıyı uzman olmayanların anlamlandırması zordur. Bir sürü teknik jargon içeriyor, mesela Kolmogorov-Smirnov testi – itiraf edeyim ben bile unutmuştum neyin nesi olduğunu! (Şaka yahu! Bir modelin verilerle uyumunun doğru olup olmadığını belirlemenin bir yoludur.)

Gelgelelim bu kayda değer çalışmanın en temel düzeyde ne kadar doğal ve sezgisel olduğunu belirtmek gerekiyor. Araştırmacılar çok çeşitli tıbbi vakalara bakmış ve belirtileri belli bir hastalığa bağlamaya çalışmıştı. Başka kimler bir insanın hasta olup olmadığını anlamaya çalışırken bu yöntemlerden yararlanır? Eşler, anne-babalar, hemşireler ve doktorlar. Ateş, baş ağrısı, burun akıntısı veya boğaz ağrısını deneyim ve bilgi temelinde çeşitli hastalıklara bağlamaya çalışırlar. Başka bir deyişle, Columbia Üniversitesi ve Microsoft'tan araştırmacılar herkesin hastalık

tanısında bulunmak için yararlandığı doğal, bariz yöntemlerden yararlanarak çığır açıcı bir çalışmaya imza atmışlardı.

Ama bir durun! Ağır ağır gidelim. Eğer en iyi veri biliminin yöntemleri, iddia ettiğim gibi, sık sık doğal ve sezgiselse, Büyük Veri'nin değeri konusunda temel bir sorun var demektir. Eğer insanlar doğal olarak veri bilimcisiyse, eğer veri bilimi sezgiselse, neden bilgisayarlara ve istatistik yazılımlarına ihtiyacımız var? Neden Kolmogorov-Smirnov testine ihtiyacımız var? Sadece içgüdülerimizle hareket etsek olmuyor mu? Biz de büyükannem gibi, hemşire ve doktorlar gibi yapamaz mıyız?

Bu soru bizi Malcolm Gladwell'in altıncı hislerin sihrini göklere çığıran çoksatar kitabı *Blink*'in* yayımlanmasından sonra kızışan bir tartışmaya götürüyor. Gladwell sadece altıncı hislerine güvenerek bir heykelin sahte olup olmadığını; bir tenisçinin topa vurmadan önce hata yapıp yapmayacağını; bir müşterinin ne kadar ödemeye hazır olduğunu söyleyebilen insanların hikâyesini anlatır. *Blink*'teki kahramanların regresyonla işi olmaz; güven aralıklarını hesaplamazlar; Kolmogorov-Smirnov testleri yapmazlar. Ama genellikle kayda değer tahminlerde bulunurlar. Birçok insan Gladwell'in sezgi savunmasını sezgisel olarak desteklemiştir: İçgüdülerine ve hislerine güvenirler. *Blink* hayranları büyükannemin bilgisayarların yardımı olmadan ilişki tavsiyesi vermesindeki bilgeliği yüceltebilirler. *Blink* hayranları benim çalışmalarıma ya da bu kitapta sergilenen bilgisayar merkezli diğer çalışmalaraysa sanırım daha soğuk yaklaşacaklardır. Eğer Büyük Veri (büyükanne türü değil de bilgisayar türü olanı) bir devrimse, Gladwell'in dikkat çektiği gibi çoğu zaman kayda değer olabilen yardımsız sezgimizden daha güçlü olduğunu kanıtlaması gerekir.

Columbia ile Microsoft'un çalışması bilgisayarların bize sadece altıncı hislerimizle asla bulamayacağımız şeyleri öğretmesine ve sıkı veri bilimine dair net bir örnek sunar. Bu aynı zamanda veri kümesinin büyüklüğünün önem arz ettiği bir örnektir. Kimi zaman elimizde yardımsız altıncı hissimizin yararlanacağı kadar deneyim yoktur. Sizin (ya da yakın arkadaşlarınızın veya aile fertlerinizin) karın ağrısını takiben hazımsızlık ile sadece hazımsızlık arasındaki farkı anlayacak kadar pankreas kanseri vakası görmüş olduğunuzu sanmıyorum. Gerçekten de Bing veri kümesi büyüdükçe araştırmacıların –şu veya bu hastalıkta– belirtilerin zamanlamasında doktorların bile gözden kaçırabileceği bir sürü incelikli örüntüyü yakalamaları kaçınılmazdır.

Diğer yandan, altıncı hissimiz bize dünyanın nasıl işlediğine dair genellikle iyi bir genel his verse de, bu çoğu zaman kesin değildir. Tabloyu netleştirmek için ve-

---

* Türkçesi: *Blink – Düşünmeden Düşünebilmenin Gücü*, çev. Dilan Sarıoğlu, İstanbul: Mediacat Yayıncılık, 2014.

rilere ihtiyacımız vardır. Örneğin hava durumunun ruh haline etkilerini düşünün. İnsanların havanın -10 derece olduğu günlerde 20 derece olduğu günlerden daha mahzun hissetmeleri muhtemeldir. Bu gerçekten de doğrudur. Ama bu sıcaklık farkının ne kadar büyük bir etkide bulunabileceğini tahmin edemezsiniz. Bir bölgenin Google'daki depresyon aramaları ile ekonomik koşulları, eğitim seviyeleri ve kiliseye ne kadar sık gidildiği dahil olmak üzere çok çeşitli etkenler arasındaki bağıntıları araştırdım. Kış mevsimi hepsini silip süpürüyor. Kış aylarında, Hawaii'deki Honolulu gibi sıcak iklimlerde depresyon aramaları Illinois'daki Chicago gibi soğuk iklimlere kıyasla yüzde kırk daha az oluyor. Bu sonuç ne kadar önemlidir peki? Antidepresanların faydasına dair iyimser bir okuma, en etkili ilaçların depresyon sıklığını sadece yüzde yirmi kadar düşürdüğünü gösterecektir. Oysa Google rakamlarına bakarak söylersek, kış depresyonunu savuşturmak için ilaç almaktansa Chicago'dan Honolulu'ya taşınmak en az iki kat daha yararlı olacaktır.*

Kimi zaman altıncı hislerimiz, dikkatli bilgisayar analizinin kılavuzluğu olmadığı takdirde ölümcül yanlışlar yapabilir. Kendi deneyimlerimiz ve önyargılarımız gözümüzü kör edebilir. Nitekim büyükannem onlarca yıllık deneyimi sayesinde ailemin geri kalanından daha iyi tavsiyelerde bulunabiliyor olmasına karşın, yine de bir ilişkinin nasıl kalıcı kılınacağına ilişkin birtakım şüpheli görüşlere sahiptir. Örneğin ortak arkadaşlara sahip olmanın önemini çok sık vurgular. Evliliğin başarısında bunun kilit bir etken olduğuna inanır: En güzel akşamlarını kocasıyla (dedemle) Queens'teki küçük arka bahçelerinde katlanır sandalyelerinde oturup, sıkı bir arkadaş grubu oluşturdukları komşularıyla dedikodu yaparak geçirmişler.

Ne var ki kendi büyükannemin okkanın altına gitmesi pahasına belirtmeliyim ki, veri bilimi büyükannemin teorisinin yanıldığını söylüyor. Yakın zaman önce bir grup bilgisayar mühendisi, insan ilişkileri hakkında toplanmış en büyük veri kümesini analiz etti: Facebook. Bir dönem "ilişkisi var" olan çok sayıda çifti incelediler. Bu çiftlerden bazıları "ilişkisi var" olarak kalmayı sürdürürken, diğerleri durumlarını "ilişkisi yok"a çevirmişti. Araştırmacılar çekirdek kadro bir arkadaş grubuna sahip olmanın bir ilişkinin uzun sürmeyeceğine dair güçlü bir emare olduğunu tespit ettiler. Belki de her gece sevgilinizle ve aynı insan grubuyla takılmak o kadar da iyi bir şey değildir; ayrı toplumsal çevreler ilişkilerin güçlenmesine yardımcı olabilir.

Göreceğiniz üzere, sezgilerimiz tek başına, bilgisayarlardan uzak durup altıncı hislerimizle hareket ettiğimizde kimi zaman bizi hayretler içinde bırakabilir. Ama

---

* Gerçeği saklayacak değilim: Bu çalışmayı tamamladıktan kısa süre sonra, California'dan New York'a taşındım. Ne yapmak gerektiğini öğrenmek için verilerden yararlanmak çoğu zaman kolaydır; uygulamaya geçirmekse zordur.

büyük yanlışlar da yapabilir. Büyükannem bir bilişsel tuzağa düşmüş olabilir: Kendi deneyimlerimizin konuyla alakasını genellikle abartırız. Veri bilimcilerin diliyle söylersek, verilerimizi tartarız ve belli bir veri göstergesine çok fazla ağırlık veririz: Kendimize.

Büyükannem dedemle ve arkadaşlarıyla geçirdiği akşam sefalarına o kadar odaklanmıştı ki, diğer çiftleri yeterince hesaba katmamıştı. Çoğu akşamını az sayıda kişinin mensup olduğu aynı arkadaş grubuyla laklak yaparak geçiren ama sık sık kavga edip en sonunda boşanan kayınbiraderini ve eşini dikkate almamıştı mesela. Anne-babamın, yani kızının ve damadının durumunu görmezden gelmişti. Annemle babam çoğu akşam ayrı takılırlar: Babam caz kulübüne ya da arkadaşlarıyla maça gider, annem ise kendi arkadaşlarıyla sinemaya ya da bir restorana gider; ama mutlu bir evlilikleri var.

Altıncı hissimize güvenirken, insanın temel bir özelliği olan dramatik, çarpıcı olana hayranlık da başımızı ağrıtıyor olabilir. Unutulmaz bir hikâye kapsamına giren şeylerin yaygınlığını genellikle çok abartırız. Örneğin kamuoyu yoklamalarında insanlar astıma kıyasla kasırgaların daha yaygın bir ölüm nedeni olduğunu söylüyorlar. Oysa astım yaklaşık yetmiş kat daha fazla ölüme sebep olmaktadır. Astımdan kaynaklanan ölümler göze çarpmıyor ve haber konusu olmuyor. Kasırga ölümleriyse öyle değil.

Başka bir deyişle, dünyanın nasıl işlediğine dair sadece kişisel olarak duyduklarımıza ya da kişisel olarak deneyimlediklerimize güvendiğimizde çoğu zaman yanılıyoruz. İyi veri biliminin yöntemleri çoğu zaman sezgisel olmakla birlikte, sonuçlar sık sık sezgilere aykırıdır. Veri bilimi doğal ve sezgisel bir insani süreci (örüntüleri saptamak ve anlamlandırmak) alır ve bunları steroidlerle doldurarak, dünyanın bizim düşündüğümüzden tamamen farklı bir tarzda işlediğini potansiyel olarak gösterir. Basketbolda başarıyı kestiren öğeleri araştırdığımda olan da buydu.

* * *

Küçükken tek bir hayalim vardı: Büyüyünce iktisatçı ve veri bilimcisi olacaktım. Yok be! Şaka yapıyorum. Profesyonel basketbolcu olmak, New York Knicks'in yıldızlar karmasına seçilmiş pivotu olan kahramanım Patrick Ewing'in izinden gitmek için yanıp tutuşuyordum.

Her veri bilimcisinin içinde çocukluk düşlerinin neden gerçekleşmediğini anlamaya çalışan bir çocuk yattığından şüphelendiğim oluyor. Bu nedenle yakın zaman önce, NBA'de oynamak için neler gerektiğini araştırmış olmam şaşırtıcı değil. Araştırmamın sonuçlarıysa şaşırtıcıydı. Aslında, iyi veri biliminin dünyaya

dair algılayışınızı nasıl değiştirebileceğini ve sayıların ne kadar sezgilere aykırı olabileceğini bir kez daha kanıtlıyor.

Özellikle odaklandığım soru şuydu: Yoksul ailelerde yetişen çocukların mı yoksa orta halli ailelerde yetişen çocukların mı NBA'e gitme ihtimali daha yüksektir?

Çoğu insan ilki diyecektir. Geleneksel akıl bize zor koşullarda yetişen, belki de daha kendisi çocukluğunu tamamlamamış bir annenin tek başına büyüttüğü çocuklarda, yoğun rekabetin yaşandığı bu sporun en üst basamaklarına tırmanmak için gerekli dürtünün daha fazla beslendiğini söylemektedir.

Philadelphia'da bir lisede basketbol antrenörü olan William Ellerbee, *Sports Illustrated*'a verdiği mülakatta tam da bu görüşü dillendirir. "Banliyölerde büyüyen çocuklar genellikle zevkine oynuyorlar. Şehrin içindeki çocuklar ise basketbola hayat memat gözüyle bakıyorlar." Heyhat, ben New Jersey banliyölerinden birinde yaşayan bir çift tarafından yetiştirildim. Benim kuşağımın en iyi oyuncusu olan LeBron James ise Ohio'nun Akron şehrinde tek başına yaşayan on altı yaşındaki bir annenin çocuğu olarak dünyaya geldi.

Gerçekten de internette yaptığım araştırma Amerikalıların çoğunluğunun koç Allerbee'yle ve benimle aynı şeyi düşündüğünü gösteriyor: Çoğu NBA oyuncusu yoksul ailelerden gelmektedir.

Bu geleneksel akıl doğru mudur?

Verilere bakalım. NBA oyuncularının toplumsal-ekonomik durumuna ilişkin kapsamlı bir veri kaynağı yoktur. Ama veri detektifliği yapıp bir sürü kaynaktan (basketball-reference.com, ancestry.com, ABD Nüfus Dairesi ve diğerleri) yararlanarak hangi ailelerden gelenlerin NBA'e gitmek konusunda daha avantajlı olduğunu saptayabiliriz. Fark edeceğiniz üzere, bu çalışmada bazıları büyük bazıları küçük, bazıları internetten bazıları internet dışından olmak üzere çeşitli veri kaynaklarından yararlanılıyor. Yeni dijital kaynaklardan bazıları heyecan verici olsa da, iyi bir veri bilimcisi işe yarayacak eski moda kaynaklara başvurmaktan çekinmez. Bir soruya doğru cevap vermenin en iyi yolu mevcut tüm verileri birleştirmektir.

Konuyla ilgili ilk veri oyuncuların doğum yerleridir. ABD'deki tek tek her yönetim bölgesinde 1980'lerde kaç tane siyah ve beyaz erkeğin doğduğunu kayda geçirdim. Ardından, kaç tanesinin NBA'e gittiğini kayda geçirdim. Bunu bir bölgenin ortalama hane halkı geliriyle karşılaştırdım. Ayrıca her bölgenin ırk demografisini kontrol ettim; zira (bu da ayrı bir kitabın konusudur) siyah erkeklerin NBA'e gitme ihtimali beyaz erkeklerden yaklaşık kırk kat daha fazladır.

Veriler zengin bir bölgede doğan birinin NBA'e gitme şansının çok daha yüksek olduğunu söylüyor. Örneğin ABD'deki en zengin bölgelerden birinde doğmuş

bir siyah çocuğun NBA'e gitme ihtimali, en yoksul bölgelerden birinde doğmuş bir siyah çocuğunkinden en az iki kat daha fazladır. Beyaz bir çocuk söz konusu olduğundaysa, aynı karşılaştırmalı oran yüzde altmıştır.

Bu demektir ki, geleneksel aklın aksine, yoksul erkekler gerçekte NBA'de daha az temsil edilmektedir. Fakat bu veri kusursuz değildir; zira ABD'deki birçok zengin bölgede de (örneğin Manhattan'daki New York County'de) Harlem gibi yoksul mahalleler vardır. Dolayısıyla zorlu bir çocukluğun NBA'de oynamaya yardımcı olma ihtimali hâlâ ortadadır. Daha fazla ipucuna, daha fazla veriye ihtiyacımız var.

Bu nedenle NBA oyuncularının aile geçmişlerini araştırdım. Bu bilgiye gazete yazılarından ve sosyal ağlardan ulaştım. Bu metodoloji epey bir zamanımı alacağı için, analizimi 1980'lerde doğup en fazla sayı atmış olan yüz Afrikalı-Amerikalı NBA oyuncusuyla sınırlandırdım. ABD'deki ortalama siyah erkeklerle kıyaslandığında NBA süperstarlarının küçük yaşta ya da evli olmayan bir annenin çocuğu olarak doğma ihtimali yüzde otuz daha azdı. Başka bir deyişle, NBA'deki en iyi siyah oyuncuların aile geçmişleri de rahat bir ailede yetişmiş olmanın başarıya giden yolda büyük bir üstünlük sağladığını gösteriyor.

Bunu söyledikten sonra, sınırlı sayıda oyuncu üzerinden yapılan bir araştırmada ne yönetim bölgesi bazlı doğum verilerinin ne de aile geçmişinin tüm NBA oyuncularının çocuklukları hakkında kusursuz bir bilgi sunacağını belirtmek gerekiyor. Bu nedenle, ebeveyni boşanmamış, orta sınıf ailelerin tek ebeveynli, yoksul ailelerden daha fazla NBA yıldızı çıkaracağına hâlâ ikna olmamıştım. Bu soruna ne kadar çok veri yüklersek, o kadar iyiydi.

Neden sonra, bir erkeğin mazisine dair daha anlamlı ipuçları sunabilecek bir veri göstergesini daha hatırladım. Roland Fryer ve Steven Levitt adında iki iktisatçının kaleme aldığı bir yazıda bir siyahın ilk adının sosyo-ekonomik kökenine dair bir işaret olduğu yazıyordu. Fryer ve Levitt 1980'lerde California'daki doğum belgelerini incelemiş ve Afrikalı-Amerikalılar arasında yoksul, eğitimsiz ve bekâr annelerin çocuklarına orta sınıftan eğitimli ve evli ebeveynlerden daha farklı adlar koyduğunu tespit etmişti.

Hali vakti yerinde ailelerden gelen çocukların Kevin, Chris ve John gibi bilindik adlar almaları daha muhtemeldi. Düşük gelirli ailelerde yetişen çocuklaraysa Knowshon, Uneek ve Breionshay gibi az görülen adların koyulması daha yüksek ihtimaldi. Yoksul ailelerde yetişen Afrikalı-Amerikalı çocukların aynı yıl doğmuş başka hiçbir çocuğa koyulmayan bir ada sahip olma ihtimali neredeyse iki kat daha yüksektir.

Peki, ya NBA'deki siyah oyuncuların ilk adları? Kulağa orta sınıf mı yoksa yoksul siyah adı gibi mi geliyor? Aynı zaman dilimine baktığımızda, California'da doğmuş NBA oyuncularının özgün adlara sahip olma oranı ortalama siyah erkeklere kıyasla yüzde elli daha fazlaydı ki bu, istatistiksel açıdan önemli bir farktır.

NBA'in gettodan gelen çocukların ligi olduğunu düşünen bir tanıdığınız var mı? Ona radyoda bir sonraki maçı daha dikkatli dinlemesini söyleyin. Russell'ın Dwight'tan sıyrıldığını, ters taraftan yardıma gelen Josh'u geçerek Kevin'a pas verdiğini ne kadar sık duyduğuna dikkat etmesini söyleyin. Eğer NBA gerçekten de yoksul siyah erkeklerle dolu bir lig olsaydı, çok daha değişik adlar duyardık. LeBron gibi adları olan oyuncular daha fazla olurdu.

Şimdi, üç farklı türde kanıt topladık: Doğduğu bölge, en skorer isimlerin annelerinin medeni durumu ve oyuncuların ilk adları. Kaynakların hiçbiri kusursuz değildir. Ama üçü de aynı hikâyeyi destekliyor. Daha iyi sosyo-ekonomik durum NBA'e gitme şansını artırıyor. Başka bir deyişle, geleneksel akıl yanılıyor.

1980'lerde doğan Afrikalı-Amerikalıların yaklaşık yüzde 60'ının anne-babası evli değildi. Ama o on yıllık dilimde doğup NBA'e giden Afrikalı-Amerikalıların birçoğunun anne-babasının evli olduğunu tahmin ediyorum. Başka bir deyişle, NBA esasen LeBron James gibi ailelerden gelen erkeklerden oluşmuyor. Teksas'ta anne-babası tarafından elektronik cihazlara olan merakı teşvik edilerek yetiştirilen Chris Bosh gibi ya da Kuzey Carolina'daki Lewisville'de orta sınıf bir ailenin ikinci oğlu olarak doğan ve 2011'de ailesiyle birlikte *Family Feud*'un ["Aileler Yarışıyor"] bir bölümünde oynayan Chris Paul gibi isimler daha fazladır.

Bir veri bilimcisinin hedefi dünyayı anlamaktır. Sezgilere aykırı bir sonuçla karşılaştığımızda, dünyanın neden göründüğü gibi olmadığını açıklamamıza yardımcı olacak daha fazla veriden yararlanabiliriz. Örneğin neden orta sınıf erkekler basketbolda yoksul erkeklerden daha iyidir? Bunu en az iki nedenle açıklayabiliriz.

Birincisi, yoksul erkeklerin genellikle daha kısa olmalarıdır. Uzmanlar çocukluk dönemindeki bakımın ve beslenmenin yetişkin sağlığında büyük bir rol oynadığını uzun zamandır biliyorlar. Tam da bu nedenle, gelişmiş dünyada ortalama erkek boyu bugün yüz elli yıl öncesine göre on santimetre daha uzundur. Veriler yoksul ailelerden gelen Amerikalıların çocukken maruz kaldıkları kötü beslenmeden ve yetersiz sağlık hizmetinden ötürü kısa kaldığını söylüyor.

Veriler NBA'e gidebilmek için boyun etkisinin ne olduğunu da söylüyor. Eminim, sezgileriniz uzun boyluların basketbolcu olma ihtimalinin daha yüksek olduğunu söylüyor. Denemesi bedava! Korttaki tipik bir basketbolcunun boyuyla tribünlerdeki

tipik bir taraftarın boyunu karşılaştırın. (NBA'deki ortalama bir oyuncunun boyu 2,01'ken, Amerikalı ortalama bir erkeğin boyu 1,76'dır.)

Boy ne kadar önemlidir? NBA oyuncularının boyları konusunda küçük yalanlar söyledikleri oluyor, ayrıca Amerikalı erkeklerin boy dağılımının tam bir listesine sahip değiliz. Ama bu dağılımın kaba bir matematiksel tahmininin nasıl bir sonuç vereceğini ve NBA'in kendi rakamlarıyla çalışınca, boyun muazzam, hatta sandığımızdan da çok etkisi olduğunu doğrulamak kolaydır. Kabaca her üç santimde bir NBA'e gitme şansınızın iki kat arttığını tahmin ediyorum ve bu tüm boy dağılımları için geçerlidir. 1,80 boyundaki bir erkeğin NBA'e gitme şansı, 1,78'lik bir erkeğinkinden iki kat daha fazladır. 2,11'lik bir adamın NBA'e gitme şansı 2,08'lik bir adamınkinden iki kat daha fazladır. Görünen o ki, 1,82'nin altındaki erkekler arasında her iki milyon erkekten sadece biri NBA'e gidebiliyor. 2,13'ün üstündeki erkekler arasındaysa, benim ve diğerlerinin tahminlerimize göre, beşte bir gibi bir orandan bahsediyoruz.

Fark etmişsinizdir, veriler benim basketbol yıldızı olma hayallerimin neden suya düştüğünü netleştiriyor. Mesele banliyöde yetişmiş olmam değil; 1,78 boyunda bir beyaz (ayrıca yavaş) olmamdı. Ayrıca tembelimdir. Kondisyonum iyi değildir, kötü şut atıyorum ve top elime geldiğinde zaman zaman panik atak geçiriyorum.

Zorlu bir çocukluk geçirmiş olan çocukların NBA'e gitmekte zorlanmalarının ikinci bir nedeniyse bazılarının belli sosyal yeteneklerden yoksun olmalarıdır. Binlerce öğrenciye dair verilerden yararlanan iktisatçılar orta sınıfa mensup iki ebeveynli ailelerin güven sorunu yaşamayan, disiplinli, sebatkâr, odaklanmış ve düzenli çocuklar yetiştirme konusunda ortalamanın çok üstünde olduklarını tespit etmişlerdir.

O halde, kötü sosyal beceriler diğer açılardan ümit vaat eden bir basketbolcu için ne tür bir sorun teşkil eder?

1990'ların en yetenekli basketbolcu adaylarından biri olan Doug Wrenn'in hikâyesine bakalım. Wrenn'in Connecticut Üniversitesi'ndeki koçu olan ve geleceğin NBA süperstarlarını yetiştiren Jim Calhoun, hayatında Wrenn kadar yükseğe zıplayabilen bir oyuncuyla çalışmadığını söylüyordu. Ama Wrenn zorlu bir çocukluk geçirmişti. Seattle'ın en olaylı mahallelerinden biri olan Blood Alley'de annesi tarafından tek başına büyütülmüştü. Connecticut'ta çevresindekilerle sürekli takışırdı. Oyunculara sataşır, koçlarını sorgular ve takım kurallarını ihlal ederek bol kıyafetler giyerdi. Ayrıca başı yasalarla dertteydi: Bir dükkândan ayakkabı çalmış ve polisleri darp etmişti. Nihayet, Calhoun'un sabrı taşmış ve Wrenn'i takımdan atmıştı.

:nn, Washington Üniversitesi'nde ikinci bir şans daha buldu. Ama orada da zliği yüzünden başarısız oldu. Oyunda kalma süresi ve şut tercihleri yüzünden kavga etti ve bu takımdan da atıldı. Wrenn, NBA Seçmeleri'nde hiçbir takım lan istenmedi, alt liglerde takıldı, annesinin yanına taşındı ve sonunda darp an ötürü hapse atıldı. 2009'da *Seattle Times*'a, "Kariyerim bitti," diyordu. lerim, heveslerim bitti. Dog Wrenn öldü. O basketbol oyuncusu, o adam dü. Bitti." Wrenn sıradan bir NBA oyuncusu değil, büyük, hatta efsanevi bir olcu olacak kadar yetenekliydi. Ama hiçbir zaman bir üniversite takımında kadar bile güçlü bir mizaç sergilemedi ya da uyumlu davranamadı. Belki sorunsuz bir çocukluk geçirmiş olsaydı, Michael Jordan'ın halefi olabilirdi.

hael Jordan'ın da elbette muazzam bir dikey zıplama yeteneği vardı. Ayrıca bir egosu vardı ve çok rekabetçiydi: Zaman zaman Wrenn'den çok da farklı n bir kişilikti. Jordan da muhtemelen zor bir çocuktu. On iki yaşında kavga :in okuldan atılmıştı. Ama Jordan'ın Wrenn'de eksikliği göze çarpan en az liği vardı: Sırtını yaslayabileceği sağlam bir orta sınıf ailede yetişmişti. Babası l Electric'te ekipman danışmanı, annesiyse bankacıydı; ikisi de kariyerini dirmesine yardımcı oldular.

nda Jordan yaşamı boyunca ailesi tarafından büyük, rekabetçi bir yeteneğin eceği tuzaklardan kurtarılmıştır. Jordan okuldan atıldıktan sonra, annesi onu a işe götürmeye başlar. Arabadan çıkması yasak olduğundan park yerinde kitap okumak zorunda kalır. Chicago Bulls tarafından NBA'e seçildikten şhret ve paranın cazibesine kapılıp kendini kaybetmemesi için anne-babası eşleri onu sık sık ziyaret ederler.

an'ın kariyeri Wrenn'inki gibi *Seattle Times*'ın pek okunmayan bir haberiyle nedi. Basketbol Şöhretler Kadrosu'na (Hall of Fame) dahil edilmesi şerefine ve milyonlarca insanın izlediği bir konuşmayla son buldu. Jordan o konuş- "yaşama dair iyi şeylere odaklanma"ya çalıştığını söyledi: "Bilirsiniz işte, rın sizi nasıl algıladıkları, size nasıl saygı duydukları … toplum tarafından gılandığınız gibi. Bir dakika durup yaptığınız şeyleri düşünün. Ve ben tüm anne-babama borçluyum."

ler Jordan'ın orta sınıf mensubu ebeveynlerine şükran duymakta sonuna aklı olduğunu söylüyor. Veriler hali vakti yerinde olmayan ailelerde, kötü elerde NBA'e gidemeyen ama NBA seviyesinde olan yetenekler bulunduğunu r. Bu adamlarda gerekli genler var, heves var, ama basketbolda süperstar ne gelmelerini sağlayacak mizacı geliştiremiyorlar.

Hayır, hayır! Sezgilerimiz ne söylüyor olursa olsun, basketbolun "hayat memat meselesi" olduğu umutsuz bir yaşam hiç de ön açıcı değildir. Doug Wrenn'in hikâyesi gibi pek çok ibretlik örnek mevcuttur ve veriler de bunun kanıtıdır.

Haziran 2013'te LeBron James NBA'de ikinci şampiyonluğunu kazandıktan sonra bir televizyon kanalı kendisiyle röportaj yaptı. (Sonrasında üçüncü şampiyonluk yüzüğünü de takacaktı.) "Ben LeBron James," diye başladı, "Ohio, Akronluyum. Akron'un içindenim. Bir gün buralara geleceğimi söyleseler kimse inanmazdı." Twitter ve diğer sosyal ağlar eleştirilerle yıkıldı. Böyle olağanüstü yetenekleri olan, daha çok küçük yaştan itibaren geleceğin basketbol yıldızı olarak gösterilen biri nasıl olurdu da geriden gelip zoru başardığını iddia edebilirdi? Aslında zorlu bir çevrede yetişmiş herkes, sportif yetenekleri ne olursa olsun, mücadeleye bir-sıfır yenik başlar. Başka bir deyişle, James'in başarıları ilk başta göründüğünden bile daha istisnaidir. Veriler de bunu kanıtlıyor.

İKİNCİ KISIM

BÜYÜK VERİ'NİN GÜÇLERİ

İKİNCİ BÖLÜM

Freud Haklı mıydı?

Geçenlerde, sokakta yürüyen birinin "penistrian" diye tarif edildiğini gördüm. Yanlış duymadınız, "pedestrian" [yaya] değil "penistrian". İnsanların yaptıkları yazım yanlışlarıyla ilgili büyük bir veri kümesinin içinde gördüm. Bir insan yürüyen birini görüyor ve "penis" kelimesini yazıyor. Bir şey kastediyor olmalı, değil mi?

Geçenlerde karısıyla evlenmeye sunağa yürürken muz yemeyi hayal eden bir adam olduğunu öğrendim. Bunu da insanların bir uygulamaya kaydettikleri hayallerle ilgili büyük bir veri kümesinde gördüm. Bir adam fallus şekilli bir yiyecek yerken bir kadınla evlendiğini hayal ediyor. Bunun da bir anlamı olmalı, değil mi?

Sigmund Freud haklı mıydı? Teorileri kamuoyunun dikkatini çektiğinden beri, bu soruya verilmiş en dürüst cevap, "Valla' bilmiyorum ki!" olacaktır. Bu hususu en açık şekilde ortaya koyan, Avusturya kökenli Britanyalı felsefeci Karl Popper'dı. Popper'ın meşhur iddiası, Freud'un teorilerinin çürütülemeyeceği yönündeydi. Bunların doğru mu yanlış mı olduğunu bilmenin bir yolu yoktu.

Freud "penistrian" yazan kişinin muhtemelen bastırılmış bir cinsel arzuyu dışavurduğunu söylerdi. O kişi de bir şeyi açığa vurmadığı, sadece masum bir yazım hatası yaptığı yanıtını verebilirdi: "'Penistrian' yerine 'pedaltrian' da yazabilirdim." Tam bir, "Sen onu dedin, ben bunu dedim" durumu. Freud düğününde muz yemeyi hayal eden adamın gizliden gizliye penis yemeyi düşündüğünü, aslında bir kadından ziyade bir erkekle evlenmeyi arzuladığını söyleyebilirdi. O adam da sadece canının muz çektiğini söyleyebilirdi. Sunağa doğru yürürken muz yerine pekâlâ bir elma yemeyi de hayal ediyor olabilirdi. Tam bir, "Sen onu dedin, ben bunu dedim" durumu. Freud'un teorisini gerçekten sınamanın bir yolu yoktu.

Ta ki şimdiye kadar.

Veri bilimi Freud'u birçok açıdan çürütülebilir hale getiriyor; ünlü teorilerinden birçoğunu test edebilmemizi sağlıyor. Rüyalardaki fallik sembollerle başlayalım. Kaydedilmiş rüyalardan oluşan devasa bir veri kümesinden yararlanarak, fallus şekilli nesnelerin ne kadar sık görüldüğünü kolaylıkla saptayabiliriz. Yiyecekler, bu çalışmaya odaklanmak için iyi bir alandır. Birçok insan rüyasında yiyecek görür ve

birçok yiyecek de fallusa benzer: Muz, salatalık, sosis, vb. O halde belli yiyecekleri rüyamızda daha çok görüyor olmamızı sağlayan etkenleri ölçebiliriz: Bu yiyecekler ne kadar sık yeniyor, çoğu insan bunların tadını ne kadar seviyor ve evet, bunlar doğası bakımından fallik mi değil mi?

Aynı derecede popüler olan ama sadece biri fallus şeklinde olan iki yiyeceğin rüyalarda farklı sıklıklarda görülüp görülmediğini test edebiliriz. Eğer fallus şekilli yiyeceklerin rüyalarda görülme ihtimali diğer yiyeceklerin görülme ihtimalinden daha düşük değilse, o zaman fallik semboller rüyalarımızda kayda değer bir etken değildir. Büyük Veri sayesinde Freud'un teorisinin bu kısmı gerçekten de çürütülebilir.

Verilerimi Shadow adında, kullanıcılarından rüyalarını kaydetmelerini isteyen bir uygulamadan aldım. On binlerce rüyada yer alan yiyecekleri kodladım.

Nihayetinde, bir yiyeceği hayal etmemizi sağlayan nedir? Esas tahmin unsuru bu yiyecekleri ne kadar sık tükettiğimizdir. Rüyamızda en çok gördüğümüz madde sudur. En fazla görülen yirmi yiyecek ise tavuk, ekmek, sandviç ve pirinçtir: Hepsi de belirgin şekilde Freudcu olmayan yiyecekler.

Bir yiyeceğin rüyalarda ne kadar sık görüleceğinin ikinci belirtisiyse insanların o yiyeceği ne kadar lezzetli bulduklarıdır. En fazla rüyamıza giren iki yiyecek, hiç de Freudcu değildir ama lezzetiyle nam salmıştır: Çikolata ve pizza.

Peki, o halde fallus şekilli yiyecekler? Rüyalarımıza beklenmedik sıklıkta mı sızarlar? Hayır.

Muz, rüyalarımızda en sık gördüğümüz ikinci meyvedir. Ama aynı zamanda en çok tüketilen ikinci meyvedir. Dolayısıyla rüyalarımızda çok sık muz görüyor olmamızı açıklamak için Freud'a ihtiyacımız yok. Salatalık da en sık rüyası görülen sebzeler arasında yedinci sıradadır. Keza en sık tüketilen yedinci sebzedir. Dolayısıyla bir kez daha, uyurken zihinlerimizdeki varlıklarını açıklamak için şekillerine bakmak zorunlu değildir. Sosisliler rüyalarımıza hamburgerler kadar sık girmiyor. İnsanların sosisliden çok burger yediğini düşünmek bile bunun doğruluğunu destekler.

Toplama bakıldığında, bütün meyve ve sebzeleri kapsayan bir regresyon analizine (sosyal bilimcilere çoklu etkenleri ayırma olanağı veren bir yöntemle) göre, bir yiyeceğin fallus şeklinde olması, popülerlik kriteriyle kıyaslandığında, rüyalarda gözükme ihtimalini beklendiğinden daha fazla artırmaz. Freud'un bu teorisi çürütülebilir ve en azından benim verilere bakışıma göre ele alınırsa, yanlıştır.

Bir de, Freudcu lapsusları (dil ve kalem sürçmelerini) düşünün. Psikoloğumuzun hipotezi, hatalarımızı (kastettiğimizi yanlış söylememizi ya da yanlış yazmamı-

zı) çoğu zaman cinsel içerikli olan bilinçdışı arzularımızı açığa çıkarmak için kullandığımız yönündeydi. Bunu test etmek için Büyük Veri'den yararlanabilir miyiz? İşte bunun bir yolu: Hatalarımızın (lapsuslarımızın) müstehcene meyledip meyletmediğine bakalım. Eğer gömülü cinsel arzularımız lapsuslarımızda ortaya çıkıyorsa, "penis", "çük", "seks" gibi kelimelerin yer aldığı hataların orantısız düzeyde fazla olması gerekir.

Tam da bu nedenle Microsoft araştırmacıları tarafından derlenen, 40 binden fazla yazım hatasının yer aldığı bir veri kümesini inceledim. Bu veri kümesinde insanların yaptıkları ama hemen düzelttikleri hatalar vardı. Bu binlerce hatada cinsel içerikli hatalar yapan bir sürü insan vardı. Yukarıda bahsi geçen "penistrian" de bunlardan biriydi. Keza "security" yerine "sexurity" ve "rocks" yerine "cocks" yazan biri de vardı. Ama aynı zamanda bir sürü masum dil sürçmesi de vardı. İnsanlar "pindows", "fegetables," "aftermoons", "refrideratorsˆ" yazmıştı.

O halde, cinsel içerikli dil sürçmelerinin sayısı olağandışı mıydı?

Bunu test etmek için öncelikle insanların belli harfleri sehven değiştirme sıklıklarını bir örüntüye oturtacak olan Microsoft veri kümesinden yararlandım. Yani *t* harfinin *s* harfiyle, *g* harfinin *h* harfiyle ne kadar sık karıştırıldığını hesapladım. Ardından, insanlarla aynı tarzda yanlışlar yapan bir bilgisayar programı yarattım. Buna Hata Botu adını verelim. Hata Botu *t* harfi yerine *s* harfini Microsoft'taki insanlarla aynı sıklıkta kullanıyordu; *g* harfini *h* harfiyle aynı sıklıkta karıştırıyordu ve sair. Programı Microsoft çalışmasındaki insanların yanlış kullandıkları kelime kadarıyla sınırlı tuttum. Başka bir deyişle, Hata Botu, "pedestrian", "rocks," "windows" ve "refrigerator" kelimelerini telaffuz etmeye çalışmıştı. Ama *r* yerine *t* harfini insanlar kadar sık değiştirmişti ve örneğin "tocks" yazmıştı. Hata Botu da *r* yerine *c* harfini insanlar kadar sık kullanmış ve "cocks" yazmıştı.

O halde Hata Botu ile normalde dikkatsiz insanları karşılaştırınca ne öğreniyoruz? Hata Botu sadece insanlar gibi birkaç harfi yanlış kullanarak birkaç milyon hata yaptıktan sonra, Freudcu nitelikte sayısız yanlış ortaya çıkmıştı. "Seashell"i "sexshell" olarak, "lipstick"i "lipsdick" ve "luckiest"i "fuckiest" olarak yazmıştı ve buna benzer başka birçok yanlış mevcuttu.† İşin can alıcı kısmı, Hata Botu'nun, elbette bilinçdışı bulunmayan bir program olmasına karşın, cinsel olarak algılana-

---

* *Security* (güvenlik) kelimesindeki "sek" sesbirimi seks olmuştur. *Rocks* (kaya), *cocks* (çük) haline gelmiştir. *Windows* (pencere), *vegetables* (sebze), *afternoons* (öğleden sonra), *refrigerator* (buzdolabı) da yanlış yazılmıştır. (ç.n.)

† *Seashell*'deki (deniz kabuğu) *sea* (deniz) *sex* (seks) olmuş; *lipstick*'teki (ruj, dudak boyası) *stick* (boya) kelimesi *dick* (çük) diye yazılmış; *luckiest* (en şanslı) ise *fuckiest* (en sikik) haline gelmiştir.

bilecek hatalar yapma ihtimalinin en az gerçek insanlarınki kadar olmasıydı. Biz sosyal bilimcilerin söylemeyi sevdiği gibi, daha fazla araştırma yapmak gerektiği uyarısında bulunarak şu kadarını söyleyebiliriz: Cinsel içerikli hatalar tesadüften daha fazlasına işaret etmiyor.

Başka bir deyişle, insanların "penistrian", "sexurity" ve "cocks" gibi hatalar yapması için yanlışlar ile yasaklı şeyler arasında bir bağlantı olması, insanların gizli arzularını hataları yoluyla açığa çıkardıkları bir zihin teorisi olması şart değildir. Bu parmak kaymaları tamamen harf hatalarının tipik sıklığıyla açıklanabilir. İnsanlar bir dolu yanlış yaparlar. Eğer sık yanlış yapan biriyseniz, sonunda "lipsdick", "fuckiest" ve "penistrian" gibi şeyleri söylemeye de başlarsınız. Bir maymun yeterince uzun süre yazacak olsa, sonunda "olmak ya da olmamak"a varacaktır. Bir insan uzun uzun yazdığında, sonunda "penistrian" sözcüğü ortaya çıkacaktır.

Freud'un hataların bilinçdışı isteklerimizi ifşa ettiği teorisi gerçekten de çürütülebilir ve benim veri analizime göre çürütülmüştür.

Büyük Veri bize muzun her zaman sadece muz olduğunu ve "penistrian"ın da sadece "pedestrian"ın yanlış telaffuz edilmiş hali olduğunu söylüyor.

* * *

O halde, Freud bütün teorilerinde hedefi ıskalamış mıydı? O kadar da uzun boylu değil. PornHub verilerine ilk kez erişme imkânı bulduğumda, en azından birazcık Freudcu diyebileceğim bir vahiyle dumura uğradım. Aslında veri araştırmalarım sırasında tespit ettiğim en şaşırtıcı şeylerden birisi budur: Anaakım porno sitelerini ziyaret eden akıl almaz sayıda insan ensest temasını arıyordu.

En popüler porno sitelerinden biri olan PornHub'da erkeklerin yaptığı yüz aramadan on altısı ensest temalı videolar içindir. Hassas içerik uyarısı: Biraz daha somutlaştırmak gerekirse, bunlar "erkek kardeş ile kız kardeş", "üvey anne ile oğlu sikişiyor", "anne ile oğul", "anne ile oğlu sikişiyor" ve "gerçek erkek kardeş ile kız kardeş" şeklinde aramalardır. Erkeklerin ensest konulu aramalarında ekseriyetle anne ile oğlun yer aldığı sahnelerden söz ediliyor. Peki, kadınlar? Kadınların PornHub'da yaptığı yüz aramadan dokuzu ensest temalı videodur ve benzer fanteziler içerir; tek fark, bahsi geçen çocuk ve ebeveynin cinsiyetinin genellikle değişmesidir. Dolayısıyla kadınların yaptığı ensest aramalarının ekseriyeti babalar ile kızların yer aldığı sahnelere yöneliktir.

Bu verilerde Freud'un Oidipal kompleksinin belli belirsiz bir yankısını bulmak zor olmasa gerek. Freud çocuklukta karşıt cinsteki ebeveynle cinsel münasebet arzusunun neredeyse evrensel olduğu, ama sonradan bastırıldığı tezini savunmuştu.

Keşke bu Viyanalı psikolog analiz yeteneklerini, karşıt cinsten ebeveyne ilginin yetişkinler tarafından –büyük bir açıklıkla– kanıtlandığının görüldüğü ve çok az şeyin bastırıldığı PornHub verilerine uygulayacak kadar uzun yaşasaydı.

Elbette PornHub verileri bize bu tür videoları izlerken insanların kimlerin fantezisini kurduğu hakkında kesin bir şey söyleyemez. Gerçekten kendi ebeveynleriyle sevişmek mi istiyorlar? Google aramaları bu tür arzuları olan bir yığın insan olduğuna dair biraz daha fazla ipucu verebilir.

"Falancamla seks yapmak istiyorum" biçimindeki tüm aramaları düşünün. Bu aramalarda en çok tekrar edilen ifade "annemle"dir. Toplamda, bu şekilde yapılan aramaların dörtte üçünden çoğu ensest kapsamındadır ve mesele şu tabirin değil de bu tabirin kullanılması değildir. Örneğin "falancadan hoşlanıyorum" biçimindeki aramalarda ensest arzuların ikrarının daha da baskın olduğunu görüyoruz. Şimdi, Freud beyefendiyi hayal kırıklığına uğratma pahasına, bunların özellikle yaygın aramalar olmadığını kabul ediyorum: ABD'de her yıl birkaç bin insan annesinden hoşlandığını kabul ediyor. Kitabın ileriki kısımlarında ele alınacağı üzere, Google aramalarının kimi zaman yasaklı olana meylettiği haberini de Freud'a uçurmak gerekiyor.

Ama dahası da var. İnsanlarda görülen daha bir sürü uygunsuz hoşlanma var ki, aramalarda daha sık bahsinin geçmesini beklerdim. Patrondan? Çalışanından? Öğrencisinden? Terapistinden? Ebeveyninden? Karısının en iyi arkadaşından? Kızının en iyi arkadaşından? Baldızından? En iyi arkadaşının karısından? İtiraf edilmiş bu arzuların hiçbiri anneyle yarışamaz. Belki de PornHub verileriyle birleştirildiğinde bunun gerçekten de bir anlamı vardır.

Diğer yandan, Freud'un cinselliğin çocukluk deneyimleriyle şekillendirilebileceğine dair genel iddiasını, Google ve PornHub verilerindeki başka kısımlar da desteklemektedir. Buna göre, en azından erkekler, çocuklukla ilgili orantısız sayıda fantezi kurmayı sürdürüyorlar. Kadınların kocalarıyla ilgili yaptıkları aramalara göre, yetişkin erkeklerin en büyük fantezilerinden bazıları bebek bezi giymek ve emzirilmektir, özellikle de yukarıda belirtildiği üzere, Hindistan'da. Dahası porno animasyonlar (ergen erkekler arasında popüler olan karakterlerin yer aldığı aleni seks sahneleri içeren animasyonlar) son derece popülerdir. Veya erkeklerin porno aramalarında en sık aradıkları kadın mesleklerini düşünün. 18-24 yaş arasındaki erkeklerin en sık aradığı kadınlar bebek bakıcılarıdır. Aynısı 25-64 yaş arası erkekler için de geçerlidir. Keza 65 yaşından büyük erkekler için de öyle. Ayrıca her yaş grubundan erkek için, öğretmen ve ponpon kız fantezisi ilk dörttedir. Hayatın

ilk yıllarının erkeklerin yetişkin fantezilerinde fazlasıyla büyük bir rol oynadığı açıkça görülüyor.

Yetişkin cinselliğine dair bu benzersiz verilerden, cinsel tercihlerin nasıl biçimlendiğini kesin olarak kavrayabilecek şekilde istifade edebilmiş değilim. Önümüzdeki yirmi-otuz yılda başka sosyal bilimcilerle birlikte yetişkin cinselliğine dair yeni, çürütülebilir teoriler üretip bunları gerçek verilerle sınayacağız.

Ama şimdiden, kesinlikle yetişkin cinselliğine dair veri temelli bir teorinin parçası olabilecek bazı temel izlekleri kestirebiliyorum. Bunun, Freud'un anlattığı, tikel, ayrıntılarıyla tanımlanmış, çocukluk ve bastırma gibi evrensel aşamalar varsayan hikâyeyle aynı olmayacağı açıktır. Ama PornHub verilerine şöyle bir bakınca, yetişkin cinselliğine dair son yargının Freud tarafından vurgulanan bazı kilit izlekleri içereceğine adım gibi eminim. Çocukluk asli bir rol oynayacak. Anneler de öyle.

* * *

On yıl önce Freud'u bu şekilde analiz etmek herhalde mümkün olmazdı. Seksen yıl önce Freud hayattayken yapmak ise kesinlikle imkânsızdı. O halde, bu veri kaynaklarının ne şekilde işe yaradığına bakalım. Bunu yaparak Büyük Veri'nin neden bu kadar güçlü olduğunu da anlayabiliriz.

Hatırlayacak olursak, sırf veri dağları yaratmanın otomatik olarak içgörü doğurmadığını söylemiştik. Veri büyüklüğü kendi içinde haddinden fazla önemsenir. Peki, o halde Büyük Veri neden bu kadar güçlüdür? Neden kendimize bakışımızda bir devrime yol açacaktır? Büyük Veri'nin dört benzersiz gücü olduğunu savunuyorum. Bu Freud analizi, söz konusu dört güce dair iyi bir örnek teşkil ediyor.

Başlangıç olarak, bu Freud değerlendirmesinde pornografiyi ciddiye aldığımızı fark etmiş olmalısınız. Bu kitapta pornografi verilerinden sık sık yararlanacağız. Biraz şaşırtıcı olsa da, sosyologlar porno verilerinden nadiren yararlanırlar; çoğu rahatını bozmadan, kariyerini üzerine kurduğu geleneksel veri kümeleriyle idare eder. Ama bir an durup düşününce, pornodan yaygın şekilde yararlanmanın (ve onunla birlikte gelen arama ve izleme verilerinin) insan cinselliğini anlama yeteneğimizdeki en önemli gelişme, hatta belki de gelmiş geçmiş en önemli gelişme olduğunu görürüz. Schopenhauer, Nietzsche, Freud veya Foucault yaşıyor olsaydı hepsinin ağzının suyu akardı. Onların yaşadığı dönemde bu veriler yoktu. On-yirmi yıl önce de yoktu. Ama şimdi var. Öncesinde sadece tahmin yürütebileceğimiz alanlara pencere açan çok çeşitli konularda birçok benzersiz veri kaynağı mevcut. *Yeni türde veriler sunması Büyük Veri'nin birinci güçlü yanıdır.*

Porno verileri ve Google arama verileri yeni olmakla kalmaz, aynı zamanda dürüsttür de. Dijital çağdan önce, insanlar utanç verici düşüncelerini diğerlerinden saklarlardı. Dijital çağda da saklamaya devam ediyorlar, ama internetten ve özellikle de anonim kalmalarını sağlayan Google ve PornHub gibi sitelerden saklamıyorlar. Bu siteler bir tür dijital hakikat serumu işlevi görmektedir; enseste olan merakın yaygın olduğunu açığa çıkarabilmemiz bu sayede mümkündür. Büyük Veri bize insanların gerçekten ne istediklerini ve ne yaptıklarını, ne istediklerini söylediklerini ve ne yaptıklarını söylediklerini nihayet görme imkânı tanıyor. *Dürüst veriler sağlaması Büyük Veri'nin ikinci güçlü yanıdır.*

Artık elimizin altında çok fazla veri olduğundan, bir topluluğun çok küçük kesimleri hakkında bile anlamlı bilgiye sahibiz. Mesela rüyasında salatalık gören insan sayısı ile domates gören insan sayısını karşılaştırabiliriz. *İnsanlarla ilgili küçük altkümelere odaklanma fırsatı sunması Büyük Veri'nin üçüncü güçlü yanıdır.*

Büyük Veri'nin etkileyici olan bir yanı daha var; Freud'a dair ilk hızlı incelememde yararlanılmayan ama ileride yararlanılabilecek bir özellik: Bize hızlı, kontrollü deneyler yürütme imkânı sunuyor. Bu sayede sadece bağıntıları değil, sebep-sonuç ilişkilerini de test edebiliriz. Bu türde testler bugün en çok şirketler tarafından kullanılıyor, ama ileride sosyal bilimciler için de güçlü bir araç olduğu anlaşılacaktır. *Bize birçok sebep-sonuç deneyi yapma imkânı sunması da Büyük Veri'nin dördüncü güçlü yanıdır.*

Şimdi bu güçlü yanlardan her birini ayrıntılarıyla ele alıp Büyük Veri'nin neden önemli olduğunu ortaya koyma vakti.

ÜÇÜNCÜ BÖLÜM

Verileri Yeniden Düşünmek

Her ayın belli bir cuma günü sabah 6'da Manhattan'ın çoğu sokağı ıssızdır. Bu sokaklarda yan yana dizili olan dükkânlar kapalıdır, kepenkleri indirilmiş, sıkı güvenlik önlemleri alınmıştır ve üstündeki daireler karanlık ve sessizdir.

Manhattan'ın aşağısındaki küresel yatırım bankası Goldman Sachs'ın ışıkları ıssa açıktır ve asansörleri binlerce çalışanı masasına taşır. Saat 7 olduğunda bu masaların çoğu doludur.

Başka günlerde şehrin bu kısmının bu saatte uyku mahmuru olduğunu söylemek haksızlık olmayacaktır. Bu cuma sabahındaysa havada bir enerji ve heyecan dalgasının gezindiği hissedilir. Çünkü o gün, borsayı muazzam etkileyecek bir bilginin açıklandığı gündür.

Bilgi paylaşıldıktan dakikalar sonra, haber sitelerinde yerini alacak. Paylaşıldıktan saniyeler sonra, Goldman'da ve diğer yüzlerce finans şirketinde değerlendirilecek, tartışılacak ve teşrih masasına yatırılacak. Ama finans camiasında gerçek hareketlilik bugünlerde saliselerle ölçülüyor. Goldman ve diğer finans şirketleri bilginin Chicago'dan New Jersey'ye aktarımını sadece dört milisaniye azaltan (17'den 13'e düşüren) fiberoptik kablolara on milyonlarca dolar ödediler. Finans şirketleri bilgiyi ve buna dayalı ticareti okuyacak algoritmalara sahip ve hepsi de milisaniyelerle ölçülüyor. Bu hayati bilgi paylaşıldıktan sonra, piyasa siz gözünüzü açıp kapayıncaya kadar hareketlenmiş olacak.

Peki, Goldman ve sayısız finans şirketi için bu kadar değerli olan bu hayati bilgi nedir?

Aylık işsizlik oranı.

Fakat bu oran (borsayı öyle derinden etkiliyor ki, finans kurumları bu bilgiyi alma, analiz etme ve işleme hızlarını azamiye çıkarmak için ellerinden geleni yapıyorlar) İşgücü İstatistikleri Bürosu'nun (Bureau of Labor Statistics - BLS) yürüttüğü bir telefon anketiyle elde ediliyor ve bilgi kamuoyuyla paylaşıldığı sırada zaten üç hafta (ya da 2 milyar milisaniye) kadar eskimiş oluyor.

Şirketler bilgi akışını bir milisaniye hızlandırmak için milyon dolarlar harcarken, hükümetin işsizlik oranını hesaplamasının bu kadar uzun sürmesi size biraz tuhaf gelebilir.

Aslında, 2011'de Obama'nın Ekonomi Danışmanları Konseyi'nin başkanı olan Alan Krueger'ın gündemindeki temel başlıklardan biri bu hayati sayıları daha çabuk elde etmekti. Fakat başarısız oldu. Krueger bu durumu "Ya BLS'nin kaynak sorunu var ya da 20. yüzyıl düşünme tarzından kurtulamıyor," diyerek yorumladı.

Hükümetin öyle kolayca hız kazanamayacağı açık olduğuna göre, işsizlik istatistiklerinin en azından kaba bir ölçümünü daha hızlı yapmanın bir yolu var mı? Herhangi bir insanın internetteki her tıklamasının bir yerlerde kaydedildiği bu yüksek teknoloji çağında, kaç insanın işsiz olduğunu öğrenmek için gerçekten de haftalarca beklemek zorunda mıyız?

Potansiyel çözümlerden birine eski Google mühendisi Jeremy Ginsberg'ün bir çalışması ilham vermişti. Ginsberg sağlık verilerinin, tıpkı işsizlik verileri gibi, hükümet tarafından gecikmeli olarak paylaşıldığını fark etmişti. Hastalık Kontrol ve Engelleme Merkezleri'nin grip verilerini paylaşması bir hafta alıyor; oysa verilerin çok daha erken paylaşılması doktorların ve hastanelerin işini kolaylaştırır.

Ginsberg grip olan kişilerin griple ilgili arama yapma ihtimalinin yüksek olduğunu düşünmüştü. Kısacası, belirtilerini Google'a yazacaklardı. Bu aramaların mevcut grip oranı hakkında makul düzeyde isabetli bir ölçüm yapabileceğini düşünüyordu. Gerçekten de "grip belirtileri" ve "kas ağrıları" gibi aramaların gribin yayılma hızına ilişkin önemli göstergeler olduğu kanıtlanmıştır.*

Öte yandan, Google mühendisleri Google Correlate adında, dışarıdaki araştırmacılara sadece sağlık alanında değil, çok çeşitli alanlarda aynı türde analizlerle deneyler yapma imkânı sunan bir hizmet geliştirdiler. Araştırmacılar zaman içinde peşine düştükleri herhangi bir veri dizisinin en çok hangi Google aramasıyla bağıntılı olduğunu görebiliyorlar.

Sözgelimi Google'ın baş iktisatçısı Hal Varian ile birlikte Google Correlate'ten yararlanarak, hangi aramaların konut fiyatlarını en yakından takip ettiğini gösterebildik. Konut fiyatları yükseldiğinde, Amerikalılar "80/20 mortgage", "yeni ev yapan müteahhitler" ve "memnuniyet oranı" gibi tabirleri arıyorlar. Konut fiyatları düştüğündeyse, genellikle "açıktan satış süreci", "değerinin altında mortgage" ve "mortgage borcunun düşürülmesi" gibi aramalar yapıyorlar.

---

* Google Flu'nun (Grip) ilk versiyonunun ciddi kusurları olmakla beraber, araştırmacılar son dönemde modeli geliştirerek daha başarılı hale getirdiler.

O halde, Google aramaları konut fiyatları ya da gripte olduğu gibi işsizlikte de kesin veriler sunabilir mi? Sadece insanların neyi google'ladıklarına bakarak ne kadar insanın işsiz olduğunu söyleyebilir miyiz ve bunu hükümet kendi araştırma sonuçlarını paylaşmadan çok önce yapabilir miyiz?

Bir gün, 2004-2011 arasında ABD'deki işsizlik oranlarını Google Correlate'e yazdım.

Bu dönemdeki trilyonlarca Google aramasından işsizlikle en sıkı bağlantısı olan şey ne çıksa beğenirsiniz? "İşsizlik bürosu" ya da benzer bir şey mi? Evet, üst sıralardaydı, ama en tepede değil. "Yeni işler" olabilir mi? O da yüksekti, ama o da en tepede değildi.

Arama yaptığım dönemde en yükseği (ki bu terimler değişiklik gösteriyor) "Slutload"du. Doğru duydunuz, en sık yapılan arama bir porno sitesiydi. İlk bakışta bu durum tuhaf görünebilir, ama işsiz insanların muhtemelen çok fazla boş zamanı oluyor! Çoğu eve hapsoluyor, yalnız kalıyor ve sıkılıyor. Yüksek çıkan aramalardan bir diğeri de (bu müstehcen olmayan alandaydı) "Spider Solitaire"di. Çok fazla boş zamanı olduğunu tahmin edebileceğimiz bir insan grubu için bu da şaşırtıcı değil.

Şimdi, bu tekil analiz temelinde "Slutload" veya "Spider Solitaire" kelimelerinin izini sürmenin işsizlik oranını tahmin etmenin en iyi yolu olduğunu iddia etmiyorum. İşsiz insanların boş zaman meşgaleleri değişebilir (bir dönem "Rawtube" adında başka bir porno sitesi en güçlü bağıntılardan biriydi) ve bu kelimelerden hiçbiri kendi başına işsizlik rakamlarına ulaşmamızı sağlayamaz. Ama boş zaman meşgaleleriyle alakalı birtakım aramaların bir karmasının işsizlik oranının izini sürebileceğini ve bu oranı tahmin eden en iyi modelin bir parçası olacağını gördüm.

Bu örnek Büyük Veri'nin ilk gücünü açıklıyor: Veri olarak nitelendirilen şeylerin yeniden tahayyül edilmesi. Çoğu zaman, Büyük Veri'nin değeri büyüklüğünden değil, size incelenecek yeni türde bilgiler (öncesinde hiç derlenmemiş bilgiler) sunmasından ileri geliyor.

Google'dan önce belli boş zaman faaliyetleri hakkında, insanların ne kadar boş zamana sahip olduğuna dair bazı ipuçları sunacak bilgiler mevcuttu (örneğin sinema bileti satışları). Ama ne kadar çok Solitaire oynandığını ya da porno izlendiğini bilme fırsatını yeni yeni elde ediyoruz ve bu güçlü bir veridir. Bu örnekte, elimizdeki veri ekonominin gidişatını daha çabuk (en azından hükümetin bir araştırma yapıp verileri toplamasından daha çabuk) bir şekilde ölçmemize yardımcı olabilir.

* * *

California Mountain View'daki Google kampüsünde hayat Goldman Sachs'ın Manhattan'daki karargâhındakinden çok farklıdır. Sabah 9'da Google ofisleri neredeyse boştur. Etrafta çalışanlardan birileri olsa bile, onlar da muhtemelen beleş kahvaltı için gelmiştir: Muzlu-yaban mersinli krep, çırpılmış yumurta akı, salatalık suyu. Bazı çalışanlar şehir dışında olabilir: Boulder ya da Las Vegas'ta bir toplantıda ya da Tahoe Gölü'ne ücretsiz gezide. Öğle yemeği vaktinde, kum voleybolu sahaları ve çim futbol sahası dolu olur. Hayatımda en iyi 'burrito'yu Google'ın Meksika restoranında yemiştim.

Dünyadaki en büyük ve en rekabetçi teknoloji şirketlerinden birinin görünüşte bu kadar gevşek ve cömert olabilmesi nasıl mümkün oluyor? Büyük Veri'den başka hiçbir şirkette görülmeyen şekilde yararlanmış olan Google, otomatiğe bağlanmış bir para akışı sağlamış durumda. Bu şirket elinizdeki kitapta hayati bir rol oynuyor; zira Google aramaları Büyük Veri'nin bu zamana kadarki en üstün kaynağıdır. Ama Google'ın başarısının bizatihi yeni verilerin toplanmasına dayandığını unutmamak da önem arz ediyor.

Eğer interneti 20. yüzyılda kullanmış kişilerdenseniz, o dönemde var olan çeşitli arama motorlarını hatırlarsınız: MetaCrawler, Lycos, AltaVista bunlardan birkaçıdır. Keza bu arama motorlarının, en yumuşak ifadeyle söylersek, ancak genel hatlarıyla güvenilir olduğunu da hatırlarsınız. Bunlar ancak zaman zaman, eğer şanslıysanız, aradığınızı bulmanızı sağlardı. Çoğu zamansa, bu fırsatı sunmazlardı. 1990'ların sonunda en popüler arama motorlarına "Bill Clinton" yazdığınızda, en üstteki sonuçlar arasında "Bill Clinton sıçtı" demekle yetinen gelişigüzel bir site ya da kötü bir Clinton şakasının yer aldığı bir site de yer alırdı. O zamanki ABD başkanıyla pek de alakalı olduğu söylenemeyecek bilgiler.

1998'de Google faaliyete geçti. Google'ın arama sonuçları rakiplerinin her birinden reddedilemeyecek kadar daha iyiydi. 1998'de Google'a Bill Clinton yazdığınızda, Clinton'ın internet sitesi, Beyaz Saray e-posta adresi ve internette yer alan en kapsamlı biyografileri çıkıyordu. Google sihirli değnek gibi bir şeydi.

Google'ın kurucuları Sergey Brin ve Larry Page neyi farklı yapmıştı?

Diğer arama motorları, kullanıcılarına aradıkları kelime grubunun en çok yer aldığı internet sitelerini öneriyordu. Eğer "Bill Clinton" hakkında bilgi arıyorsanız, o arama motorları Bill Clinton'a tüm internette en fazla göndermede bulunan internet sitelerini buluyordu. Bu sıralama sisteminin kusurlu olmasının birçok nedeni vardı ve bunlardan biri de sistemi kandırmanın kolay olmasıydı. Sayfanın

bir yerinde "Bill Clinton Bill Clinton Bill Clinton Bill Clinton Bill Clinton" yazan bir şaka sitesi Beyaz Saray'ın resmî internet sitesinden daha yukarıda yer alıyordu.*

Brin ve Page'in yaptığı, basit bir kelime sayımından çok daha değerli olan yeni tür bir bilgiyi kaydetmenin yolunu bulmak oldu. İnternet siteleri bir konuyu tartışırken çoğu zaman konunun anlaşılmasına en fazla yardımcı olduğunu düşündükleri sitelere bağlantı veriyorlardı. Örneğin *New York Times*, Bill Clinton'dan bahsederken, onun adına tıklayan okurlarını Beyaz Saray'ın resmî internet sitesine yönlendirir.

Bu bağlantılardan birini yaratan her internet sitesi, bir anlamda, Bill Clinton hakkındaki en iyi bilgiye dair kendi fikrini sunuyordu. Brin ve Page her konuda tüm bu fikirleri toplamayı başardı. *New York Times*'ın, milyonlarca Listservs'ün, yüzlerce blogcunun ve internetteki herkesin görüşlerini içerik oluşturmak için kullanabilirdi. Eğer bir sürü insan "Bill Clinton" için en önemli bağlantının Clinton'ın resmî internet sitesi olduğunu düşünüyorsa, o zaman "Bill Clinton"ı arayan çoğu insanın görmek isteyeceği internet sitesi muhtemelen buydu.

Diğer arama motorlarının dikkate bile almadığı verileri sunan bu tür bağlantılardan yararlanan Google, herhangi bir konuda en yararlı bilgiyi kestirme konusunda inanılmaz başarılıydı. Burada mesele, Google'ın herkesten daha çok veri toplayarak arama dünyasına egemen olması değildi. Google bunu daha iyi türde bir veri bularak yapmıştı. Google kuruluşunun üstünden iki yıl geçmeden, bağlantı analizinin verdiği güçle internetin en popüler arama motoru haline geldi. Bugün Brin ve Page'in toplam serveti 60 milyar doların üstündedir.

Tıpkı Google'da olduğu gibi, dünyayı anlamak için verilerden yararlanmaya alışan herkes için durum aynıdır: Büyük Veri devrimi sürekli daha fazla veri toplamaktan ziyade, doğru verileri toplamakla alakalıdır.

Ama internet yeni veriler toplayabileceğiniz ve doğru verilere erişmenin doğuracağı sonuçların kargaşaya yol açacağı yegâne yer değildir. Bu kitap esasen internetteki verilerin insanları daha iyi anlamamıza nasıl yardımcı olabileceğini konu alıyor. Fakat bir sonraki kısmın internet verileriyle bir alakası yok. Aslında insanlarla da pek bir alakası yok. Ama yine de bize bu bölümün meramını anlatma konusunda yardımcı oluyor: Yeni, geleneksel olmayan verilerin boyundan büyük değeri. Ayrıca öğrettiği ilkeler dijital temelli veri devrimini anlamamıza yardımcı olacaktır.

---

* 1998'de Google'dan önceki popüler arama motorlarında "araba "diye arama yaptığınızda, karşınıza bir yığın porno sitesi çıkıyordu. Bu porno siteleri arama motorunu kandırmak için beyaz arkaplana beyaz harflerle defalarca "araba" yazıyorlardı. Böylece araba almak isteyen ama porno sitesine denk gelince dikkati dağılan kullanıcılardan fazladan birkaç tık alıyorlardı.

Veri Olarak Beden

2013 Yaz'ında büyüklüğü ortalamanın üzerinde olan, siyah yeleli kızıl-kahverengi bir at New York'un dışındaki küçük bir ahırda öyle duruyordu. Ağustos ayında Saratoga Springs'te düzenlenen Fasig-Tipton Seçkin Tay Satışı'nda görücüye çıkan 1 yaşındaki 152 attan ve o yıl mezatta satılacak olan bir yaşındaki 10 bin attan biriydi.

Varlıklı erkek ve kadınlar at yarışına büyük paralar akıtırken, atın adını seçme şerefine nail olmayı da isterler. Dolayısıyla kızıl-kahverengi atın henüz bir adı yoktu ve mezattaki çoğu at gibi ahır numarasıyla çağrılıyordu: 85.

Bu mezatta No. 85'i diğerlerinden öne çıkaracak pek bir şey yoktu. Şeceresi iyiydi ama müthiş değildi. Babası Pioneerof [metinde böyle] the Nile (Nil'in Öncüsü) en önemli yarış atlarından biriydi, ama Pioneerof the Nile'ın diğer evlatları yarışta çok fazla başarı elde edememişti. Ayrıca No. 85'in görünüşüne dair de şüpheler vardı. Örneğin talipler ayak bileğinde bir sıyrığın bir sakatlık belirtisi olduğundan endişe edebilirdi.

No. 85'in sahibi Mısırlı bir bira üreticisi olan Ahmed Zayat'tı. Kendisi New York civarına atını satıp yeni birkaç at almak için gelmişti.

Hemen her at sahibi gibi, Zayat da at seçiminde kendisine yardımcı olması için bir uzman ekiple anlaşmıştı. Ama onun uzmanları diğerlerininkinden biraz farklıydı. Bunun gibi bir etkinlikte karşılaşacağınız tipik at uzmanları orta yaşlı, genellikle Kentucky'den ya da Florida kırsalından gelen, fazla eğitim almamış ama ailesi atçılıkla uğraşan erkekler olurdu. Oysa Zayat'ın uzmanları EQB adında küçük bir şirkettendi. EQB'nin başında eski ekolden bir at terbiyecisi yoktu. EQB'nin sahibi, Jeff Seder adında Harvard mezunu eksantrik bir Philadelphia'lıydı.

Zayat önceden de EQB'yle çalışmıştı, dolayısıyla sürece aşinaydı. Seder'ın ekibi atları birkaç gün değerlendirdikten sonra, Zayat'a gelip No. 85'in yerine alabileceği beş-altı at önerecekti.

Fakat bu kez durum farklıydı. Seder'ın ekibi Zayat'a, talebini karşılayamayacağını söylemişti. O gün satışa çıkarılan 151 attan hiçbirini öneremiyorlardı. Bunun yerine, beklenmedik bir biçimde ve aşırı heyecanla bir ricada bulunmuşlardı. Zayat kesinlikle, asla ve asla No. 85'i satmamalıydı. EQB'nin dediğine göre, bu at mezattaki en iyi at olması dışında, o yılın, hatta belki de on yılın en iyi atıydı. EAQ Zayat'a, "Evinizi satın ama bu atı satmayın," diyerek adeta yalvarmıştı.

Ertesi gün, No. 85 adlı at, kendine Incardo Bloodstock diyen bir adam tarafından, tantana yaratılmadan 300 bin dolara satın alındı. Sonradan anlaşıldı ki, Bloodstock, Ahmed Zayat'ın kullandığı bir müstear admış. Seder'ın ricalarını

kıramayan Zayat, neredeyse eşi görülmedik bir hamleyle kendi atını satın almıştı. (Mezat kuralları gereği Zayat atını geri çekemiyordu ve dolayısıyla takma adla satın almak zorunda kalmıştı.) Mezatta No. 85'ten daha pahalıya giden altmış iki at vardı ki, bunlardan ikisi 1'er milyon dolara satılmıştı.

Zayat üç ay sonra nihayet No. 85'e bir ad koydu: Amerikan Firavunu. 18 ay sonra, New York City banliyölerinde sıcaklığın 24 derece olduğu bir cumartesi akşamı, Amerikan Firavunu son otuz küsur yılda Triple Crown'u kazanan ilk at oldu.

Jeff Seder No. 85 hakkında kimsenin bilmediği ne biliyordu? Bu Harvardlı adam atlardan nasıl bu kadar iyi anlar hale gelmişti?

Amerikan Firavunu'nun Triple Crown'u kazanmasından bir yıl sonra, Florida'ya bağlı Ocala'da Haziran ayının kavurucu öğle sıcağında Seder'la tanıştım. Seder o zaman altmış dört yaşındaydı. İki yaşındaki atların bir hafta boyunca görücüye çıktığı ve Zayat'ın kendi atını satın aldığı 2013'teki etkinlik gibi mezatla tamamlanacak bir organizasyondaydık.

Seder'ın Mel Brooks'u andıran borazan gibi bir sesi vardı, çok gür saçlıydı ve adım atarken göze batar şekilde sekiyordu. Askıyla tutturulmuş hâki bir pantolon ve üzerinde şirket logosunun bulunduğu siyah bir gömlek giymişti, ayrıca işitme cihazı kullanıyordu.

Üç gün boyunca, bana hayat hikâyesini ve atları tahmin etme konusunda nasıl bu kadar iyi hale geldiğini anlattı. Epey engeli aşarak buralara gelmişti. Harvard'dan ikincilikle mezun olduktan sonra yine Harvard'da hukuk ve işletme okumuştu. Yirmi altı yaşında New York'ta Citigroup'ta analist olarak çalışıyordu ama kendini mutsuz ve tükenmiş hissediyordu. Bir gün, şirketin Lexington Caddesi'ndeki yeni ofisinde avluda otururken kendisini geniş bir arazinin resmedildiği büyük bir duvar resmini incelerken buldu. Resim ona kırsalı ve atları ne kadar çok sevdiğini hatırlattı. Eve gidince aynada üç parçalı takım elbisesine baktı. Bankacı olmak istemediğini ve New York'ta yaşayamayacağını o an anlamıştı. Ertesi sabah, işi bıraktı.

Seder, Pennsylvania kırsalına taşındı ve tekstilden spor ilaçlarına kadar çeşitli sektörlerde çalıştıktan sonra tüm zamanını büyük tutkusuna vakfetmeye karar verdi: Yarış atlarının başarılarını tahmin etmek. At yarışındaki rakamlar kabataslaktır. Ülkenin en prestijli mezatlarından olan Ocala'da görücüye çıkan iki yaşındaki bin attan muhtemelen beşi ileride büyük ödüllü bir yarışı kazanacaktır. Diğer 995 at ne olacak? Kabaca üçte birinin çok yavaş olduğu anlaşılacak. Diğer üçte biri sakatlanacak; zira çoğu at dörtnala koşmanın muazzam baskısına dayanamadığından bileklerinden sakatlanmaktadır. (Her yıl yüzlerce at, çoğunlukla da bacakları kırıldığı için Amerikan yarışlarında hayatını kaybetmektedir.) Geri kalan üçte bir

ise Bartleby sendromu diyebileceğimiz bir duruma yakalanır. Hermann Melville'in olağanüstü öyküsündeki katip Bartleby çalışmayı bırakır ve patronundan gelen her talebe, "Yapmamayı tercih ederim" diye yanıt verir. Görünen o ki, birçok at, yarış kariyerinin başlarında eğer istemiyorsa koşmak zorunda olmadığını fark etmeye başlar. Bir yarışa hızla başlayabilirler, ama bir noktada, yavaşlayacak ya da koşmayı tamamen bırakacaklardır. Özellikle de toynakları ve dizleri ağrırken neden var gücüyle koşsun ki? "Yapmamayı tercih ederim" derler. (At olsun insan olsun, Bartleby'lere bir zaafım var.)

Koşullar at sahiplerinin bu denli aleyhineyken kâr getirecek bir atı nasıl seçebilirler? Geçmişten bu yana, bir atın başarılı olup olmayacağını tahmin etmenin en iyi yolunun şeceresini analiz etmekten geçtiğine inanılır. At uzmanı olmak demek bir atın anası, babası, büyükanneleri, büyükbabaları, kardeşleri hakkında bilinebilecek her şeyi bilmek demektir. Sözgelimi menajerler bir atın anne tarafında bir sürü büyük at varsa, o atın büyük olmasını gayet normal kabul ederler.

Fakat bir sorun var. Şecere önem arz etmekle birlikte, yarış atının başarısının ancak küçük bir kısmını açıklayabilir. At yarışlarının en prestijli ödülü olan Yılın Atı Ödülü'nü almış tüm atların kardeşlerinin listesine bakalım. Bu atlar olabilecek en iyi şecerelere sahiptir: Dünya çapında tarihe geçecek atlardan oluşan benzer aile geçmişleri vardır. Yine de, dörtte üçünden daha fazlası hiçbir zaman büyük bir yarış kazanamaz. Veriler atların başarılarını tahmin etmenin geleneksel yolunun epey geliştirilmesi gerektiğini gösteriyor.

Şecerenin o kadar da iyi bir gösterge olmaması aslında şaşırtıcı değildir. İnsanları düşünün. Bir NBA takımı sahibinin, gelecekteki takımını şecerelerine baktığı on yaşındaki çocuklardan kurduğunu düşünün. Earvin "Magic" Johnson'ın oğlu Earvin Johnson III.'ü izlemesi için bir menajerle anlaşması gerekirdi. O menajer "Boyu, kalıbı şu ana kadar iyi" diyebilirdi. "Gerektiği kadar iri, Johnson'ın soyundan geliyor. Saha görüşü müthiş olacaktır, bencil oynamayacaktır, boyu ve hızı iyi olacaktır. Sosyal biri, müthiş bir kişilik haline gelecektir. Yürüyüşü güvenli. Cana yakın. Müthiş bir yatırım." Ne yazık ki on dört yıl sonra, bu takım sahibinin elinde 1.88'lik (profesyonel bir basketbolcu için kısa boylu), *E!*'de çalışan bir moda blogcusu olacaktır. Sonuçta, Earvin Johnson III. forma tasarımında müthiş yardımcı olabilir, ama muhtemelen sahada çok pek işe yaramazdı.

At sahiplerinin yöntemiyle kadro kuran bir NBA takımı sahibi, moda blogcusu çocuğun yanı sıra, Michael Jordan'ın üniversitedeyken vasat basketbolcular olan iki oğlu Jeffrey ve Marcus Jordan'ı da havada kapardı. Talih Cleveland Cavaliers'ın yüzüne gülmüş: Takımın as oyuncusu LeBron James'in annesinin boyu sadece

1.64'tür. Veya liderlerini şecerelerine bakarak seçen bir ülke tahayyül edin. Başımızda George W. Bush gibileri olurdu. (Kusura bakmayın, dayanamadım!)

At menajerleri şecere dışında başka bilgilerden de yararlanırlar. Örneğin iki yaşındaki atların yürüyüşüne bakar ve atları görsel olarak incelerler. Ocala'da birçok menajerle yaptığım uzun sohbetlerde, aslında ne aradıklarına dair pek bir fikir birliği olmadığını anladım.

Bu bariz çelişkilere ve belirsizliklere bir de bazı at alıcılarının sınırsız görünen bir para kaynağına sahip olmalarını eklediğinizde, ortaya epey verimsiz bir piyasa çıkar. On yıl önce No. 153 diğer tüm atlardan daha hızlı koşan, birçok menajere güzel görünen ve mükemmel bir şeceresi olan (gelmiş geçmiş en büyük yarış atlarından ikisi olan Kuzey Dansçısı ile Secretariat'ın soyundan geliyordu) iki yaşında bir attı. Hem İrlandalı bir milyarder hem de Dubaili bir şeyh bu atı satın almak istedi. Fiyat teklifi kapışması bir süre sonra bir gurur savaşına dönüştü. Yüzlerce at terbiyecisinin şaşkın bakışları karşısında fiyat teklifi giderek arttı ve nihayet iki yaşındaki at gelmiş geçmiş en yüksek fiyata, 16 milyon dolara satıldı. Gelgelelim Yeşil Maymun adı verilen No. 153 üç yarış koştu, sadece 10 bin dolar kazandırdı ve emekliye ayrıldı.

Seder ise at değerlendirmelerinde geleneksel yöntemlere hiçbir zaman itibar etmemişti. O sadece verilerle ilgileniyordu. Yarış atlarının çeşitli özelliklerini ölçüp hangilerinin performansla bağıntılı olduğunu görmeyi planlıyordu. Seder'ın bu planını internet icat edilmeden beş yıl önce geliştirdiğini belirtmek önemlidir. Ama stratejisi büyük oranda veri bilimine dayalıydı. Hikâyesinden çıkarılacak dersler de Büyük Veri'den yararlanan herkes için geçerlidir.

Seder'ın çabaları uzun yıllar hep hüsranla sonuçlandı. Atların burun deliklerini ölçen Seder, burun deliklerinin büyüklüğü ve toplamda kazandıkları ödül hakkında dünyanın ilk ve en büyük veritabanını oluşturdu. Burun büyüklüğünün atların başarısına dair bir gösterge olmadığını tespit etmişti. Atların kalplerini incelemek için EKG çektirmiş ve hızlı kasılan kaslarının büyüklüğünü ölçmek için ölü atların kol ve bacaklarını kesti. Bir keresinde bir ahırın dışında eline küreği alıp atların dışkı büyüklüklerini belirlemeye çalışmıştı, zira bir etkinlikten önce çok fazla dışkılamanın bir atı yavaşlatabileceğine dair bir teorisi vardı. Bunların hiçbirinin yarışlarda başarıyla bağıntısı olmadığını anladı.

Sonra, bundan on iki yıl önce, ilk büyük atılımını gerçekleştirdi. Seder atların iç organlarının büyüklüğünü ölçmeye karar vermişti. Mevcut teknolojiyle bunu yapmak imkânsız olduğundan, kendisi bir portatif ultrason geliştirdi. Sonuçlar kayda değerdi. Kalp büyüklüğünün, özellikle de sol karıncık büyüklüğünün atın

başarısının (en önemli değişkenin) en belirgin habercisi olduğunu bulmuştu. Keza dalak da önemli bir organdı: Dalağı küçük olan atlar neredeyse hiçbir şey kazandırmıyordu.

Seder birkaç isabetli atış daha gerçekleştirdi. Dörtnala koşan atların binlerce videosunu dijital ortama aktardı ve belli yürüyüş tarzlarının yarıştaki başarıyla bağıntılı olduğunu tespit etti. Ayrıca iki yaşındaki atlardan bazılarının iki yüz metre koştuktan sonra hırıltılı sesler çıkardığını keşfetti. Bu tür atlar bazen milyon dolarlara satılıyordu, ama Seder'ın verileri hırıldayan atların asla büyük paralar kazandırmayacağı yönündeydi. Bu nedenle bir yardımcı tutup ona bitiş çizgisinde hırıldayan atları tespit etme görevi vermişti.

Ocala müzayedesindeki bin kadar attan kabaca on tanesi Seder'ın testlerini geçebildi. Seder şecereleriyle bir atın kaça satılacağını etkilemesi dışında hiç ilgilenmiyor. Şöyle diyor: "Şecere bize bir atın müthiş bir kariyere sahip olma konusunda çok küçük bir şansı olabileceğini söyler. Ama eğer müthiş olacağını görebiliyorsam, hangi aileden geldiğinin ne önemi var ki?"

Seder bir gece beni Ocala'da Hilton otelindeki odasına davet etti. Odada bana çocukluğundan, ailesinden ve kariyerinden bahsetti. Karısının, kızının ve oğlunun fotoğraflarını gösterdi. Philadelphia'da gittiği lisesinin üç Yahudi öğrencisinden biri olduğunu ve okula başladığında boyunun sadece 1.24 olduğunu anlattı. (Üniversitede 1.77 olmuştu.) En gözde atından bahsetti: Pinky Pizwaanski. Seder bu atı neşeli bir biniciden almış ve adını da bu şekilde koymuştu. Pinky'nin en başarılı at olmasa bile her zaman varını yoğunu ortaya koyduğunu düşünüyordu.

Son olarak, No. 85'le ilgili tuttuğu tüm verilerin yer aldığı dosyayı, yani kariyerinin en büyük tahminini getiren dosyayı gösterdi. İşinin sırrını mı açıyordu bana? Kim bilir, belki de öyleydi, ama bunu umursamadığını söyledi. Onun gözünde sırlarını saklamaktan daha önemlisi, haklı çıktığının kanıtlanmasıydı ve

No. 85 (adı Amerikan Firavunu olarak değiştirildi) Bir Yaşındayken, Yüzdelik Dilim

	YÜZDELİK DİLİM
Boy	56
Kilo	61
Şecere	70
Sol karıncık	99,61

tüm dünyaya yirmi yıl boyunca kırık uzuvlarla, bok püsürle ve eğreti ultrasonlarla uğraşmanın bu çabaya değdiğini göstermekti.

Seder ve ekibinin No. 85 konusundaki takıntısının sebepleri ayan beyan ortadaydı. Sol karıncığının yüzdelik dilimi 99,61'di!

Hepsi bu kadar da değil; diğer önemli organlarının hepsi (kalbinin ve dalağının geri kalanı dahil) bir o kadar büyüktü. Genel olarak konuşmak gerekirse, yarış söz konusu olduğunda sol karıncık ne kadar büyükse atın o kadar iyi olduğunu tespit etmişti. Ama bunun kadar büyük bir sol karıncık, diğer organlar küçükse, hastalık belirtisi de olabilir. Amerikan Firavunu'nda tüm kilit organlar ortalamanın üstündeydi ve sol karıncık muazzam büyüktü. Veriler No. 85'in yüz binde bir, hatta milyonda bir bulunan bir at olduğunu bağırıyordu.

* * *

Veri bilimcileri Seder'ın projesinden ne gibi dersler çıkarabilirler?

Birincisi ve belki de en önemlisi, eğer bir alanda devrim yapmak için yeni verilerden yararlanmaya çalışacaksanız, eski yöntemlerin berbat olduğu bir alana girmek en iyisidir. Seder'ın alt ettiği şecere takıntılı at menajerleri, gelişme kaydedilecek geniş bir alan bırakmışlardı. Aynısı Google'ın alt ettiği kelime sayma takıntılı arama motorları için de geçerliydi.

Google'ın arama verilerinden yararlanarak grip hastalığını tahmin etme girişiminin zayıf yönlerinden biri, sadece geçen haftanın verilerini kullanıp basit bir mevsimsel düzeltme yaparak da gribi önceden tahmin edebiliyor olmanızdır. Bu basit ama güçlü modele arama verilerinin ne kadar veri eklediği hâlâ tartışma konusudur. Kanaatimce, Google aramaları mevcut verilerin zayıf olduğu ve dolayısıyla Google STD [*sexually transmitted diseases* – cinsel yolla bulaşan hastalıklar] gibi bir şeyin uzun vadede Google Flu'dan daha değerli olduğunu kanıtlayabileceği sağlık durumunu ölçme konusunda daha fazla ümit vaat ediyor.

Çıkarılacak ikinci ders ise, kestirimlerde bulunmaya çalışırken, modellerinizin neden işe yaradığı konusunu çok fazla kafaya takmamanız gerektiğidir. Seder sol karıncığın bir atın başarısını tahmin etme konusunda neden bu denli önemli olduğunu tam olarak açıklayamamıştı. Keza dalağın değerini de öyle. Belki bir gün gelecek, at kardiyologları ve hematologları bu gizemleri çözecekler. Ama şu anda bunun pek bir önemi yok. Seder tahmin işinde çalışıyor, açıklama işinde değil. Ve tahmin işinde bilmeniz gereken, bir şeyin işe yarayıp yaramadığıdır, niye işe yaradığı değil.

Örneğin Walmart hangi ürünleri rafa koyacağını belirlemek için tüm dükkânlarındaki satış verilerinden yararlanır. Walmart 2004'te ABD'nin güney-

doğusunu vuran dehşetengiz Frances Kasırgası'ndan önce doğru bir akıl yürütmeyle, fırtınanın yaklaştığı bir şehirde yaşayan insanların alışveriş alışkanlıklarının değişebileceği tahmininde bulunmuştu. Önceki kasırgalarla ilgili verileri didik didik ederek insanların neler almak isteyebileceği tespit edilmişti. Kallavi bir cevap mı geldi dersiniz? Hayır! Çilekli Pop-Tart. Bu ürün kasırga öncesi günlerde normalden yedi kat daha fazla satılıyordu.

Bu analizi temel alan Walmart kamyonlarını çilekli Pop-Tart'la doldurup kasırganın güzergâhında yer alan dükkânlara yönlendirdi. Ve gerçekten de Pop-Tart çok iyi sattı.

Neden Pop-Tart? Belki de buzdolabı ya da pişirme gerektirmediği için. Peki, neden çilekli? Hiçbir ipucu yok. Ama kasırga bir bölgeyi vurduğunda, insanlar belli ki çilekli Pop-Tart'a yükleniyorlar. Dolayısıyla kasırgadan önceki günlerde Walmart raflarını kutu kutu Pop-Tart'la dolduracak şekilde stok yapıyor. Bu ilişkinin nedeninin bir önemi yok; ama ilişkinin önemi var. Belki bir gün gıda bilimcileri kasırgalar ile çilekli jöle kaplı hamur işi arasındaki çağrışımı belirleyecektir. Ama bu tür bir açıklama beklerken, Walmart yine de kasırga yaklaştığında raflarını çilekli Pop-Tart'la doldurup, Rice Krispies'i güneşli günlere saklasa iyi olur.

Bu dersi Orley Ashenfelter'ın hikâyesinden de rahatlıkla çıkarabiliriz. Seder ile atlar arasındaki ilişki neyse, Princeton'da iktisatçı olan Ashenfelter ile şaraplar arasındaki ilişki de o olabilir.

On küsur yıl önce Ashenfelter asfalyaları atacak raddeye gelmişti. Fransa'nın Bordeaux bölgesinden bir sürü kırmızı şarap alıyordu. Bazen şarap harika oluyor, verdiği paraya değiyordu; ama çoğu zaman tam bir hayal kırıklığı yaşıyordu.

Ashenfelter neden bu kadar farklılık gösteren şaraplara aynı parayı ödediğini merak etmeye başlamıştı.

Bir gün, aynı zamanda şarap uzmanı olan bir gazeteci arkadaşından bir ipucu aldı. Bir şarabın iyi çıkıp çıkmayacağını anlamanın bir yolu gerçekten de vardı. Ashenfelter'ın arkadaşının dediğine göre, kilit unsur üzümlerin yetişme mevsimindeki hava durumuydu.

Bu tespit Ashenfelter'ın çok ilgisini çekti. Arkadaşının söylediklerinin doğru olup olmadığını ve düzenli şekilde iyi şarap alıp alamayacağını anlamak için bir arayışa girdi. Bordeaux bölgesinin son otuz yıllık hava durumuna dair verileri indirdi. Ayrıca şarapların müzayede fiyatlarını topladı. Şarap ilk başta satıldıktan yıllar sonra gerçekleşen müzayedeler size şarabın akıbetinin ne olduğunu söylemesi bakımından önemliydi.

Sonuçlar inanılmazdı. Bir şarabın kalitesini büyük oranda sadece üzümlerin yetişme mevsimindeki hava durumuyla anlamak mümkündü.

Aslında bir şarabın kalitesi basit bir formüle indirgenebilir. Buna Bağcılığın İlk Yasası adını verebiliriz:

Fiyat = 12,145 + 0,00117 kışın yağış miktarı + 0,0614 yetişme mevsiminde ortalama sıcaklık – 0,00386 hasat dönemi yağış miktarı.

Peki, neden Bordeaux bölgesinde şarap kalitesi bu formüle göre işliyor? Bağcılığın İlk Yasası'nı nasıl açıklamak gerekir? Ashenfelter'ın şarap formülünün bir açıklaması var: Sıcaklık ve erken sulama üzümlerin düzgün olgunlaşması için zorunludur.

Ama Ashenfelter'ın kestirimsel formülünün kesin ayrıntıları her türlü teorinin çok ötesindedir ve alanın uzmanlarınca bile muhtemelen hiçbir zaman tam olarak anlaşılmayacaktır.

Neden kışın bir santimetre yağmur, ortalamaya vurulduğunda, tamamen olgunlaşmış bir şişe kırmızı şarabın fiyatına tastamam 0,1 sentlik artış olarak yansır? Neden 0,2 sent değil? Neden 0,05 değil? Bu soruları kimse cevaplayamaz. Ama bir kış mevsiminde fazladan 1.000 santimetre yağmur yağarsa, şaraba şişe başına fazladan 1 dolar ödemeniz gerekecektir.

Gerçekten de Ashenfelter kendi regresyon modelinin neden işe yaradığını tam olarak bilmemesine karşın, şarap almak için bu modelden yararlanmıştı. Ashenfelter'a göre, "acayip işe yaramıştı." İçtiği şarapların kalitesi belirgin şekilde artmıştı.

Eğer amacınız geleceği kestirmekse (hangi şarap iyi çıkacak, hangi ürünler satacak, hangi atlar hızlı koşacak?), modelinizin neden bu şekilde işlediğini çok umursamamanız gerekir. Sayıları yerli yerine oturtun yeter. Jeff Seder'ın at hikâyesinden çıkan ikinci ders budur.

Seder'ın Triple Crown'u kazanacak potansiyel atı tahmin etme girişimindeki başarıdan çıkarılacak son ders ise, nelerin veri sayılacağını belirlerken açık ve esnek olmaktır. Mesele eskinin at menajerlerinin Seder çıkagelmeden önce verilere kapalı olmaları değildir. Onlar da yarış zamanlarını ve şecereleri incelemişlerdi. Ama Seder'ın dehası başkalarının öncesinde bakmadığı yerde veri araması, geleneksel olmayan veri kaynaklarını düşünmesiydi. Bir veri bilimcisi için taze ve özgün bir perspektif, semeresini verebilir.

Veri Olarak Kelimeler

2004'te bir gün, o dönem Harvard'da doktora yapan ve medya alanında uzman olan iki genç iktisatçı Massachusetts'te gey evliliklerini yasal güvence altına alan bir mahkeme kararını okuyorlardı.

Matt Gentzkow ve Jesse Shapiro adlı bu iki iktisatçı ilginç bir şey fark ettiler: İki gazete aynı haberi anlatmak için çarpıcı derecede farklı bir dil kullanmıştı. Muhafazakârlığıyla nam salmış olan *Washington Times*'ın haber başlığı, "Eşcinseller Massachusetts'te 'Evleniyor'"du. Liberal olarak bilinen *Washington Post* ise "hemcins çiftler" için bir zaferden bahsediyordu.

Farklı haber kaynaklarının farklı yönlere meyletmeleri, gazetelerin aynı haberi farklı noktalara odaklanarak sunmaları hiç de şaşırtıcı değildir. Aslında Gentzkow ve Shapiro yıllardır, aldıkları iktisat eğitiminin medyanın yanlılığını anlamaya yardımcı olup olamayacağını düşünüyorlardı. Neden bazı haber kanalları daha liberal bir görüşe sahipken, diğerleri daha muhafazakârdır?

Gentzkow ve Shapiro bu sorunla nasıl başa çıkabileceklerini gerçekten bilmiyorlardı; medyadaki öznelliği sistemli ve nesnel bir şekilde nasıl ölçebileceklerini kestiremiyorlardı.

Dolayısıyla Gentzkow ve Shapiro'nun gey evlilikleri hakkında ilginç buldukları şey haber kanallarının aynı haberi farklı şekilde sunmaları değil, gazetelerin haberi nasıl farklı sunduklarıydı: Farklılık dönüp dolaşıp kelime tercihindeki bariz bir değişikliğe dayanıyordu. 2004'te, *Washington Times*'ın kullandığı "eşcinseller" ifadesi, geçmişte gey bireyleri tarif etmek için kullanılan aşağılayıcı bir tabirdi; oysa *Washington Post*'un kullandığı "hemcins çiftler" ifadesi ise gey ilişkilerin aşk ilişkisinin bir biçimi olduğunu vurguluyordu.

Bu iki akademisyen, dilin bu önyargıyı anlamanın kilidi olup olamayacağını merak etmişti. Liberaller ve muhafazakârlar tutarlı şekilde farklı ifadeler mi kullanıyorlardı? Gazetelerin haberlerde kullandıkları kelimeler veriye dönüştürülebilir miydi? Bu durum Amerikan basını hakkında bize ne anlatabilirdi? Basının liberal mi yoksa muhafazakâr mı olduğu konusunda bir yargıya varılabilir miydi? Dahası bunun nedeni açıklanabilir miydi? 2004'te bunlar önemsiz sorular değildi. Amerikan gazetelerindeki milyarlarca kelime kâğıda ya da mikrofilme hapsolmaktan çıkmıştı. Belli internet siteleri artık ABD'deki hemen her gazetedeki her haberde yer alan kelimeleri tek tek kaydediyordu. Gentzkow ve Shapiro bu siteleri tarayıp, dilin gazetelerin taraflılığını ya da önyargılarını hangi dereceye kadar ölçebileceğini hızla test edebilirlerdi. Ayrıca bunu yaparak haber medyasının nasıl işlediğine dair anlayışımızı da genişletebilirlerdi.

Fakat ne bulduklarını tarif etmeden önce, Gentzkow ve Shapiro'nun hikâyesini ve gazetelerdeki dili rakamlara dökme girişimlerini şimdilik bir kenara bırakıp çok farklı alanlardan araştırmacıların insan doğasını daha iyi anlamak amacıyla bu yeni tür veriden (kelimelerden) nasıl yararlandıklarını ele alalım.

* * *

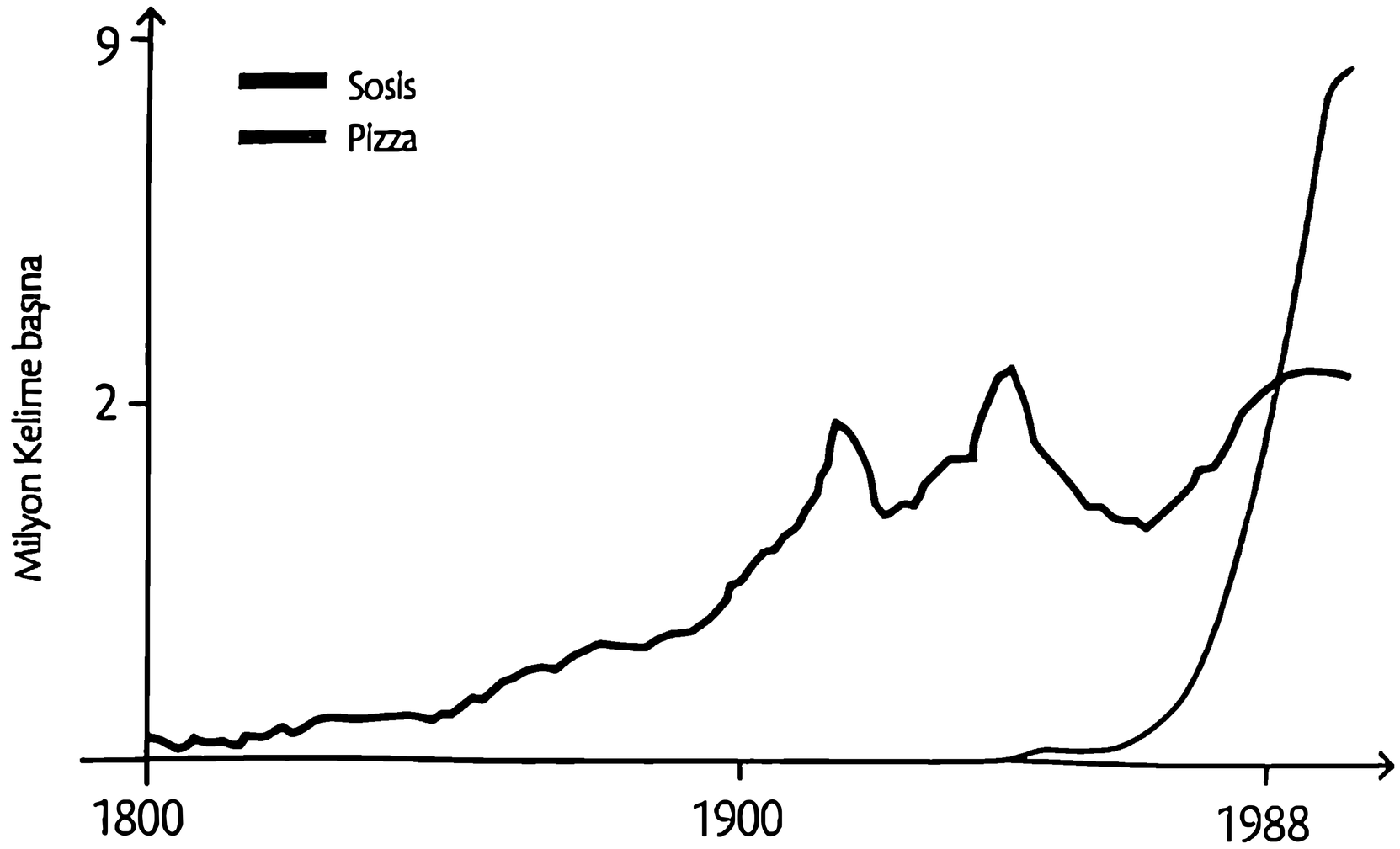
Dil, elbette, sosyal bilimler için her zaman ilgi çekici bir konu olmuştur. Ne var ki genel olarak dil incelemeleri yakın metin okumaları gerektiriyordu ve devasa metin yığınlarını verilere dönüştürmek pek mümkün olmuyordu. Ama şimdi bilgisayarlar ve dijitalleşme sayesinde devasa belge yığınlarındaki kelimeleri tablolara dökmek kolaydır. Böylece dil, Büyük Veri analizine tabi hale gelmiştir. Google'ın yararlandığı bağlantılar kelimelerden oluşuyordu. Keza benim incelediğim Google aramaları da. Bu kitapta kelimeler sık sık karşımıza çıkıyor. Ama dil Büyük Veri devriminde öyle önemli bir yere sahiptir ki, başlı başına bir kısım ayrılmasını hak ediyor. Aslında şu an o kadar çok kullanılmaktadır ki, sadece ona ayrılmış bir alan bile vardır: "Veri olarak metin".

Bu alandaki temel gelişmelerden biri Google Ngrams'tır. Birkaç yıl önce Erez Aiden ve Jean-Baptiste Michel adında iki genç biyolog, kelimelerin belli kullanımlarının nasıl yaygınlaştığına dair yeni içgörülere ulaşmak için asistanlarına eski, tozlu metinlerdeki kelimeleri tek tek saydırdılar. Bir gün, Google'ın dünyadaki kitapların büyük bir kısmını dijital ortama aktarmak üzere yeni bir projeye giriştiğini duyan Aiden ve Michel, dilin tarihini anlamak için bunun çok daha kolay bir yol olacağını hemen anladılar.

Aiden *Discover* dergisine şöyle diyordu: "Yöntemlerimizin iflah olmaz derecede köhnemiş olduğunu fark etmiştik. Bu dijitalleştirme deviyle rekabet edemeyeceğiniz açıktı." Dolayısıyla arama şirketiyle işbirliği yapmaya karar verdiler. Google mühendislerinin yardımıyla, dijital ortama aktarılmış milyonlarca kitapta belli kelimeleri ya da deyimleri arayan bir sistem geliştirdiler. Araştırmacılar bu sayede, bir kelimenin ya da deyimin 1800'den 2000'e kadar her yıl ne sıklıkla kullanıldığını öğrenebilecekti.

Peki, kelimelerin ya da deyimlerin farklı yıllarda çıkmış kitaplardaki kullanılma sıklığından neler öğrenebiliriz? Bir defa, sosisin yavaş yavaş popülerleştiğini, pizzanın ise görece yakın dönemde hızla popüler hale geldiğini öğrenebiliriz.

Ama bundan daha büyük dersler de çıkarılabilir. Sözgelimi Google Ngrams bize ulusal kimliğin nasıl oluştuğunu öğretebilir. Aiden ve Michel'in kitabı *Uncharted*'da büyüleyici bir örneğe yer verilmiştir.



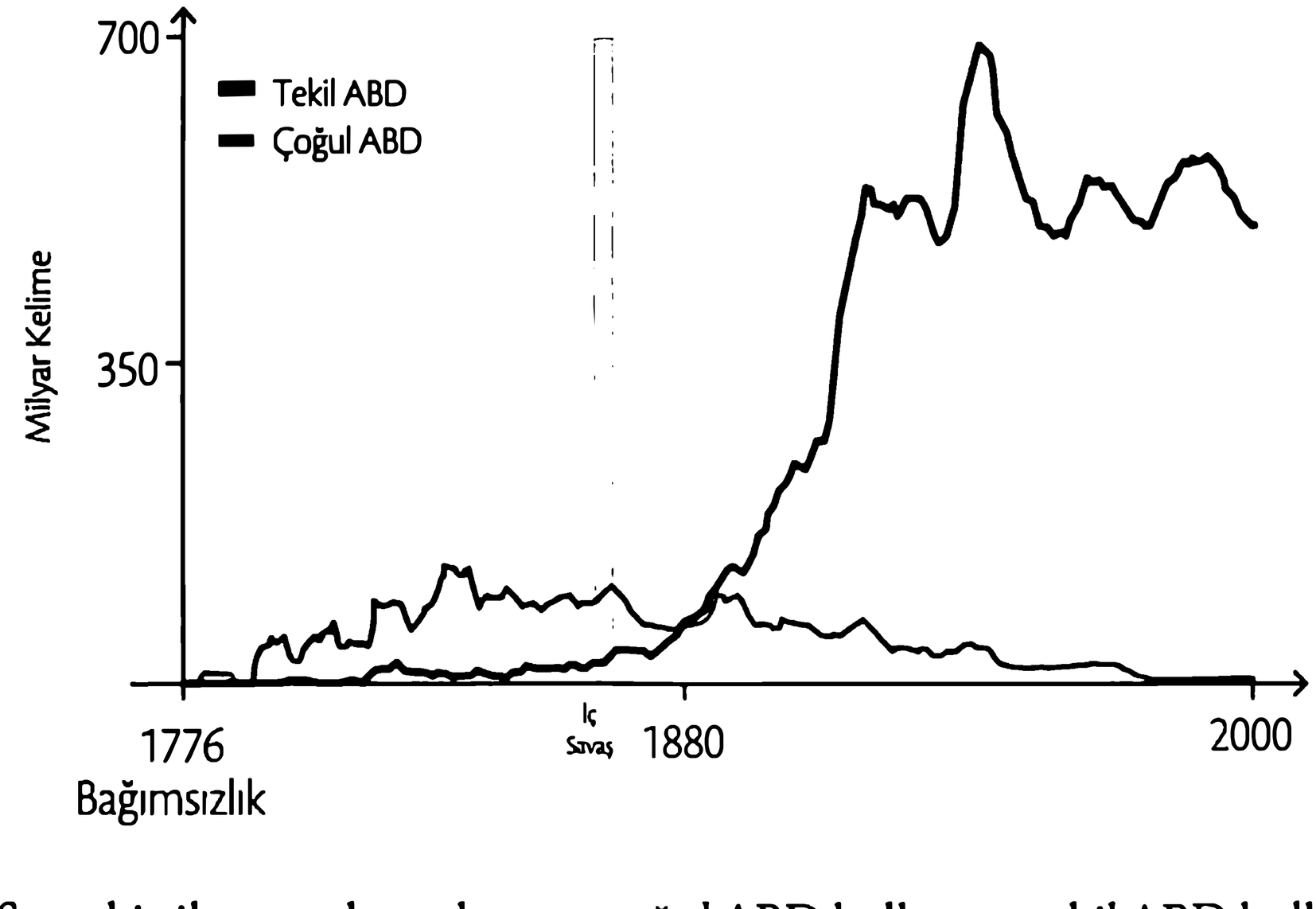
Birincisi, hızlı bir soru-cevap. ABD'nin şu anda birleşik bir ülke mi yoksa bölünmüş bir ülke mi olduğunu düşünüyorsunuz? Eğer siz de çoğu insan gibiyseniz, bugünlerde siyasal kutuplaşmanın iyice artmış olmasından ötürü ABD'nin bölünmüş olduğunu söylersiniz. Öyle ki ülkenin hiç olmadığı kadar bölünmüş olduğunu dahi söyleyebilirsiniz. ABD nihayetinde artık renk kodludur: Kırmızı eyaletler Cumhuriyetçi; mavi eyaletler Demokrattır. Ama Aiden ve Michel *Uncharted*'da ABD'nin bir zamanlar çok daha fazla bölünmüş olduğunu gösteren büyüleyici bir veri göstergesine dikkat çekerler. Bu veri göstergesi insanların ülke meselelerinden bahsederken kullandıkları dildir.

Önceki paragrafta ülkenin ne kadar bölünmüş olduğunu ele alırken kullandığım kelimeleri düşünün. ABD'den tekil bir ad gibi bahsettim. Gayet doğal; doğru gramer ve standart kullanım budur. Ama siz bunu fark etmediniz bile.

Oysa Amerikalılar her zaman bu şekilde konuşmuyorlardı. Ülkenin ilk günlerinde Amerikalılar Amerika Birleşik Devletleri'nden çoğul olarak bahsediyorlardı. Örneğin John Adams 1799'daki Birliğin Durumu konuşmasında, "Amerika Birleşik Devletleri ve onların Britanyalı Majesteleriyle yaptıkları anlaşmalar"dan bahseder. Eğer benim kitabım 1800'de yazılmış olsaydı, ben de ABD'yi çoğul olarak kullanacaktım. Bu küçük kullanım farkı uzun zamandır tarihçileri büyüleyen bir husustur, zira Amerika'nın belli bir noktada kendisini bir devletler toplamı olarak görmeyi bırakıp tek bir ulus olarak tahayyül etmeye başladığını gösterir.

Peki, bu ne zaman gerçekleşti? *Uncharted*'da anlatıldığına göre, tarihçiler kesin bir tarih veremiyorlardı; zira bunu sınamanın sistemli bir yolu yoktu. Ama birçoğu İç Savaş'ın dönüm noktası olduğunu düşünüyordu. Aslında Amerikan Tarih Derneği'nin Pulitzer ödüllü eski başkanı James McPherson net konuşmuştu: "İç Savaş, ABD'nin tekil isme geçişinin başlangıcıdır."

Ama McPherson'ın yanıldığı anlaşılıyor. Google Ngrams sayesinde, Aiden ve Michel bunu sistemli bir yolla kontrol edebildi. Amerika'da yayımlanmış kitaplarda, ülke tarihi boyunca her yıl ABD isminin kaç kez tekil, kaç kez çoğul kullanıldığını görebildiler. Dönüşüm daha tedriciydi ve İç Savaş sona erdikten epey sonra hızlanmıştı.



İç Savaş bittikten on beş yıl sonra, çoğul ABD kullanımı tekil ABD kullanımından hâlâ daha fazlaydı ve ülkenin dilsel açıdan hâlâ bölünmüş olduğunu gösteriyordu. Zihniyet dünyasındaki değişimler askeri zaferler kadar çabuk gerçekleşmez.

* * *

Bir ülkenin nasıl birleşik hale geldiği konusunda bu kadar söz yeter. Peki, bir erkek ile bir kadın nasıl birleşir? Kelimeler burada da yardımcı olabilir.

Örneğin bir erkek ile bir kadının ilk buluşmada nasıl konuştuklarına bakarak ikinci kez buluşup buluşmayacaklarını tahmin edebiliriz.

Bu işin altında Stanford ve Northwestern'den disiplinlerarası bir grup bilimcinin imzası vardı: Daniel McFarland, Dan Jurafsky ve Craig Rawlings. Yüzlerce heteroseksüel Speeddate kullanıcısını incelemiş ve karşılarındaki adaya yakınlık hissedip hissedemediklerini ve ikinci bir buluşma isteyip istemediklerini tahmin etmenin yolunu saptamaya çalışmışlardı.

İlkin geleneksel verilerden yararlandılar. Kişilere boy, kilo ve hobilerini sorup, bu etkenlerin aşk kıvılcımı hissettiğini söyleyen kişilerle bağıntısını sınadılar. Ortalama bir kadın kendinden daha uzun olan ve hobilerini paylaşan erkekleri tercih ediyordu; ortalama bir erkek ise daha zayıf olan ve hobilerini anlatan kadınları tercih ediyordu. Yeni bir şey yoktu yani.

Ama bilimcilerimiz yeni tür bir veri de toplamışlardı. Her iki cinsten de yanlarında kayıt cihazı götürmelerini istemişlerdi. Buluşmanın ses kayıtları sonradan dijital ortama aktarılacak, böylece bilimciler kullanılan kelimelerin, gülmelerin ve ses tonlarının şifresini çözebileceklerdi. Hem kadınların hem erkeklerin ilgilendiklerini nasıl belli ettiklerini ve partnerlerin bu ilgiyi nasıl doğurduklarını test edebileceklerdi.

O halde, dilsel veriler bize ne anlatıyor? Birincisi, bir erkek ya da kadının ilgisini nasıl aktardığını. Bir erkeğin karşı tarafı çekici bulduğunu belli etme yollarından biri açıktır: Kadının esprilerine güler. Bir diğeri o kadar bariz değildir: Erkek konuşurken sesinin iniş çıkışlarını sınırlandırır. Bir araştırmaya göre, kadınlar tekdüze bir sesi erkeksi buluyorlar; yani erkekler belki de bilinçdışı düzlemde, bir kadından hoşlandıklarında erkekliklerini abartıyorlar.

Bilimciler, bir kadının ilgisini, ses tonunu renklendirerek, daha yumuşak bir ses tonuyla konuşup, daha kısa sürede söz alarak belli ettiklerini tespit etmişlerdi. Ayrıca özel olarak kullandıkları kelimeler temelinde de büyük ipuçları elde edilebiliyordu. Kadınların "sanırım" ya da "muhtemelen" gibi kaçamaklı kelimeler kullandıklarında genellikle ilgi duymadıkları tespit edilmişti.

Ey erkekler! Eğer bir kadın herhangi bir konuda kaçamaklı konuşuyorsa, başka bir deyişle içeceği konusundaki yorumu "şöyle böyle"yse, "ne bili'im" işte biraz üşüdüyse ya da "sanırım" başka bir atıştırmalık daha almayacaksa, "ne bili'im", "sanırım" bu işin oluru yoktur.

Bir kadın kendisinden bahsetmeye başladıysa muhtemelen size ilgi duyuyordur. İlişki kurmaya çalışan bir erkeğin bir kadının ağzından duyabileceği en güzel kelimenin "ben" olduğu görülmüştür: Kendini rahat hissettiğinin bir işaretidir. Ayrıca bir kadının "Bilirsin ya", "Anlatabildim mi?" gibi kendine göndermeli tabirler kullanması da muhtemelen hoşlandığını gösterir. Neden? Bilimciler bu deyimlerin karşı tarafın dikkatini çekmeyi amaçladığını vurguluyorlar. Bunlar sıcak, cana yakın deyimlerdir ve bir insanın bağ kurmaya çalıştığına işaret eder, bilmem anlatabildim mi?

Şimdi, erkekler ve kadınlar ilgi duydukları birinden randevu koparmak için nasıl iletişim kurabilirler? Veriler bir erkeğin karşısındaki kadının ilgisini çekme şansını artırmak için epey farklı şekillerde konuşabileceğini göstermektedir. Kadınlar kendi açtıkları yoldan yürüyen erkeklerden hoşlanırlar. Bir kadının esprilerine gülen ve konuşmayı ikide bir kendi istediği konuya getirmektense kadının açtığı konudan devam ettiren erkeklerden hoşlanma ihtimallerinin daha yüksek olması sanırım kimseyi şaşırtmayacaktır.* Kadınlar destek sunan ve duygularına ortak olan er-

---

* Üzerinde çalıştığım teorilerden biri: Büyük Veri, rahmetli Leonard Cohen'in söylediklerinin tamamını onaylamaktan başka bir şey yapmıyor. Örneğin Leonard Cohen bir

keklerden de hoşlanırlar. Eğer bir erkek, "Bu müthiş bir şey!" ya da "Bu gerçekten iyiymiş," diyorsa, kadının yakınlık hissettiğini söyleme ihtimali çok daha yüksektir. Keza "Bu da zormuş," ya da "Üzülmüş olmalısın," gibi sözler de öyle.

Bu noktada kadınlara kötü bir haberim var; zira veriler erkekler hakkında nahoş bir hakikati doğruluyor görünmektedir. Sohbetin, erkeklerin kadınlardan etkilenmesi hususundaki payı oldukça zayıftır. Bir erkeğin bağ kurabildiğini belirttiği durumlarda fiziksel görüntü her şeyden daha baskın rol oynamaktadır. Bununla beraber, kadınların bir erkeğin ilgisini çekmek için şansını en azından birazcık artıracak bir kelime var ve onu da söylemiştik: "Ben". Erkekler, kendisinden bahseden kadınlardan daha çok hoşlanıyorlar. Daha önce belirtildiği üzere, bir kadının da kendisinden bahsettiği bir buluşmadan sonra olumlu bir yorumda bulunma ihtimali daha yüksektir. Dolayısıyla ilk buluşmada konuşmanın önemli bir kısmının kadına ayrılması iyiye işarettir. Kadın rahat olduğunun işaretini vermiş olur ve muhtemelen erkeğin konuşmayı kendi gündemiyle boğmamasını takdir eder. Erkek ise kadının açılmasından hoşlanır. İkinci buluşma o zaman muhtemeldir.

Son olarak, randevu dökümlerindeki bir şeyin bariz bir sıkıntıya işaret ettiği anlaşılmıştır: Soru işareti. Eğer bir buluşmada bir sürü soru soruluyorsa, erkek ile kadının bir bağ yakalama ihtimali düşüktür. Bu durum sezgilere aykırı görünüyor olabilir; soru sormanın ilgi işareti olduğunu düşünebilirsiniz. Ama ilk buluşmada değil. İlk buluşmada çoğu soru sıkıcı geçtiğine işaret eder. "Hobilerin neler?" "Kaç kardeşsiniz?" Bunlar insanların sohbet tıkandığında söyledikleri şeylerdir. Harika bir ilk buluşmada sadece en sonda bir soru sorulur: "Benimle yeniden çıkar mısın?" Eğer buluşmada sorulan tek soru buysa, cevap muhtemelen "evet" olacaktır.

Ayrıca erkekler ve kadınlar sadece birbirlerini tavlamaya çalıştıklarında farklı şekillerde konuşmazlar. Erkekler ve kadınlar zaten genel olarak farklı şekillerde konuşurlar.

Bir grup psikolog yüz binlerce Facebook gönderisinde kullanılan kelimeleri analiz etmişti. Erkeklerin ve kadınların bir kelimeyi ne kadar sık kullandığını ölçmüş ve ardından İngilizcedeki en eril ve en dişil kelimeleri tespit etmişlerdi.

Ne var ki bu kelime tercihlerinin çoğu apaçık farkları ortaya koyuyordu. Örneğin kadınlar "alışveriş"ten ve "saçlarım"dan erkeklere kıyasla daha çok bahsediyor, erkekler ise "futbol"dan ve "Xbox"tan daha çok bahsediyorlardı. Sanırım bunu bilmek için Büyük Veri analizi yapan bir grup psikoloğa danışmak gerekmez.

---

keresinde yeğenine kadınları tavlamak için şu nasihatte bulunmuştu: "İyi bir dinleyici ol. Sonra, biraz daha dinlemeyi bil. Ve yeterince dinlediğine kanaat getirdiğin an, biraz daha dinle." Cohen bu bilimcilerin bulduklarıyla kabaca aynı şeyi söylüyordu.

Ne var ki bulgulardan bazıları daha ilginçti. Kadınlar "yarın" kelimesini erkeklerden daha sık kullanıyorlardı, muhtemelen erkekler ileriyi düşünmek konusunda çok iyi olmadıkları için. "So" [çok] kelimesinin sonuna "o" harfinin eklenmesi en dişil dil özelliklerinden biridir. Kadınların en orantısız derecede kullandıkları kelimeler arasında "soo", "sooo", "soooo", "sooooo" ve "soooooo" vardır.

Belki de küçüklüğümde ara ara sinkaf etmekten korkmayan kadınlara maruz kalmamdandır, ama her zaman sövmenin kadınla erkeği eşit kılan bir şey olduğunu düşünmüştüm. Öyle değilmiş. Erkeklerin kadınlardan daha sık kullandıkları kelimeler arasında, "siktir", "sıçayım", "sikeyim", "hay bok", "siktiğimin" ve "amk" vardır.

İşte erkeklerin en sık kullandıkları kelimeler ile kadınların en sık kullandıkları kelimeler: Bir kelimenin büyüklüğü, ne kadar çok kullanıldığını göstermektedir.

Bu çalışmanın sevdiğim kısmı, yeni verinin bizi uzun zamandır var olan ama her zaman farkında olmadığımız örüntüler hakkında bilgilendirmesidir. Erkekler ile kadınlar her zaman farklı şekillerde konuşmuşlardır. Ama on binlerce yıldır, ses dalgaları uzayda kaybolduğu anda bu veriler de kaybolmuştur. Oysa şimdi bu veriler bilgisayarlarda muhafaza ediliyor ve bilgisayarlar tarafından analiz edilebiliyor.

Erkekler



Kadınlar



İÇ. ÇALIŞ. DUA ET.

19 - 22 yaş arası



23 - 29 yaş arası



30 - 65 yaş arası



Veya belki de bir erkek olarak şöyle demem gerekirdi: "Bu kelimeler eskiden kaybolurdu amk. Şimdi futbol izleyip Xbox oynamaya ara verip bu boku öğrenebiliriz. Tabii sikine takan olursa."

Farklı şekillerde konuşan sadece erkekler ile kadınlar değildir. İnsanlar yaşlandıkça da farklı şekilde konuşmaya başlarlar. Bu bize yaşlanma sürecinin nasıl gerçekleştiği konusunda bile bilgi verebilir. İşte aynı çalışmadan, değişik yaş gruplarındaki insanların Facebook'ta orantısız şekilde kullandıkları kelimeler. Ben bu grafiğe "İç. Çalış. İbadet et" diyorum. İnsanlar gençken içiyorlar. Yirmilerinde çalışıyorlar. Otuzlarında ve sonrasında ibadet ediyorlar.

* * *

Metin analizindeki güçlü ve yeni araçlardan birine duygu analizi adı veriliyor. Bilimciler artık bir metnin belli bir kısmının ne kadar mutlu ya da üzücü olduğunu değerlendirebiliyorlar.

Bunu nasıl yapıyorlar? Bilimciler çok sayıda insandan İngilizcedeki binlerce kelimeyi olumlu ya da olumsuz diye kodlamasını istediler. Bu metodolojiye göre, en olumlu kelimeler arasında "mutlu", "sevgi", "harika" vardı. En olumsuz kelimelerse "üzücü", "ölüm" ve "depresyon"du. Böylece çok sayıda kelimenin ruh haline göre bir indeksini çıkardılar.

Bu indeksten yararlanarak, metnin bir kısmındaki kelimelerin ortalama ruh halini ölçebiliyorlar. Eğer birisi "Mutluyum, âşığım ve harika hissediyorum," diyorsa, duygu analizi bunu aşırı mutlu bir metin olarak kodlayacaktır. Birisi, "Dünyadaki ölümleri ve depresyonları düşündükçe üzülüyorum," dediğindeyse duygu analizi bunu son derece üzücü bir metin olarak kodlayacaktır. Metnin diğer parçaları ise bu ikisinin ortasında bir yerde olacaktır.

Peki, metnin ruh halini kodlayarak ne öğreneceksiniz? Facebook veri bilimcileri heyecan verici olanaklardan birine işaret ediyorlar. Bir ülkenin gün başına düşen Gayri Safi Milli Mutluluğu'nu tahmin edebiliyorlar. Eğer insanların statü mesajları genellikle olumluysa, ülkenin o gün mutlu olduğu varsayılıyor; eğer olumsuzsa, mutsuz olduğu varsayılıyor.

Facebook veri bilimcilerinin bulgularından bazıları şunlardır: Noel, yılın en mutlu günlerinden biridir. Şimdi, ben bu analizin doğruluğundan kuşku duyuyorum ve aslında bu projenin tamamına biraz şüpheyle bakıyorum. Genel itibariyle, Noel'de birçok insan yalnız ya da ailesiyle kavgalı olduğu için gizliden gizliye üzgündür. Daha genele bakıldığındaysa, bir sonraki bölümde ele alacağım nedenlerden

ötürü, Facebook'taki durum güncellemelerine itibar etmiyorum: Sosyal medyada hayatlarımız hakkında genellikle yalan söylüyoruz.

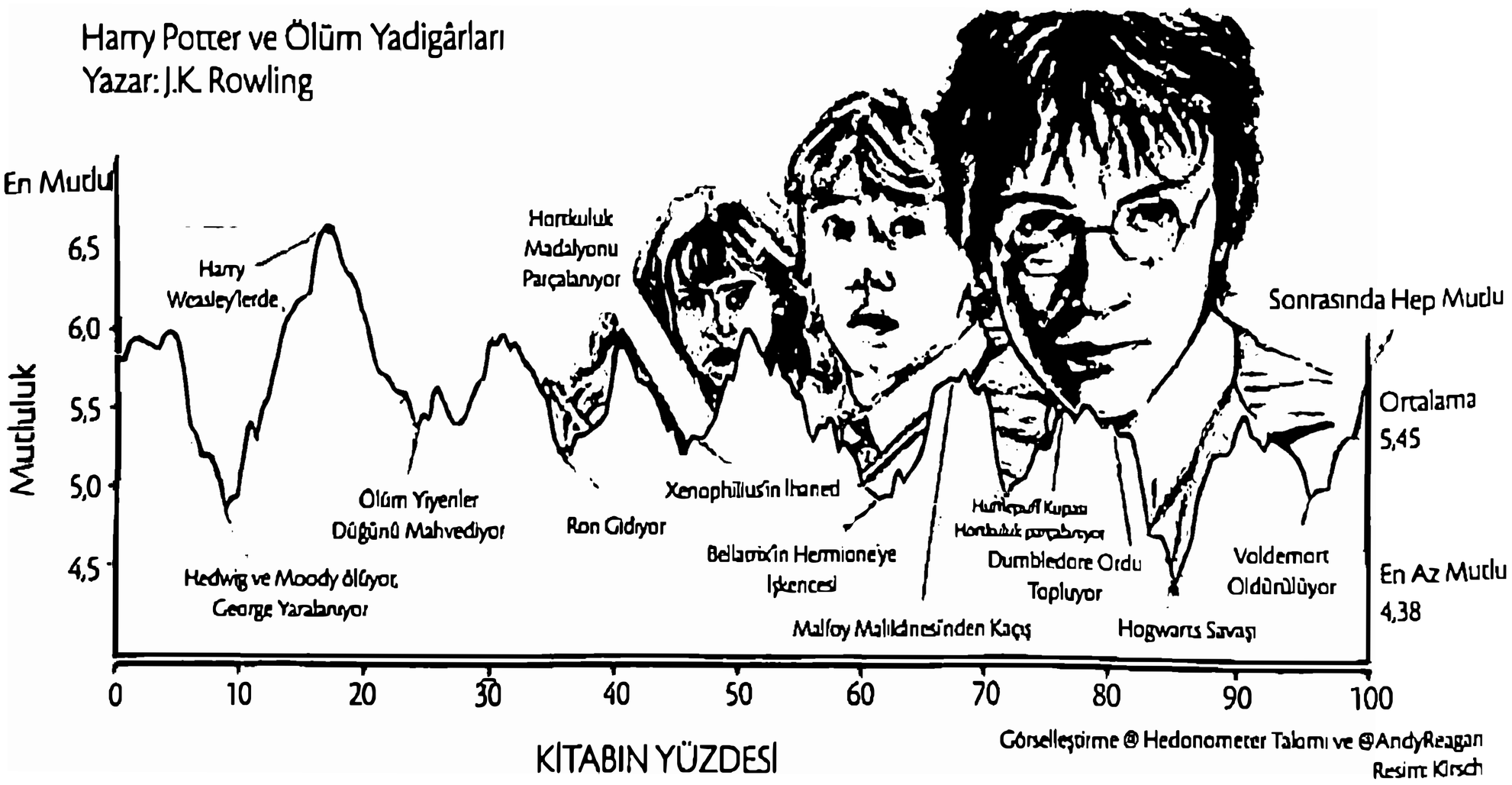
Eğer Noel'de yalnız ve acınası halde olsanız, ne kadar mutsuz olduğunuzu bütün arkadaşlarınıza duyurarak onları rahatsız etmek ister miydiniz? Tatsız tuzsuz bir Noel geçirirken yine de Facebook'una "harika, olağanüstü, mutlu bir yaşam"ı olduğunu yazan birçok insan var gibime geliyor. Bu kişiler ABD'nin Gayri Safi Milli Mutluluğu'nu artırmaktadır. Eğer Gayri Safi Milli Mutluluk'u gerçekten kodlayacaksak, sadece Facebook durum güncellemelerinden daha fazla kaynak kullanmamız gerekir.

Bununla beraber, Noel'in üç aşağı beş yukarı neşeli bir güne vesile olduğu doğru görünüyor. Google'daki depresyon aramaları ve Gallup anketleri de Noel'in yılın en mutlu günlerinden biri olduğunu söylüyor. Ayrıca intiharların tatillerde arttığı sadece bir şehir efsanesidir. Noel'de bazı insanlar üzgün ve yalnız olsa bile, neşeli insanlar çok daha fazladır.

Bu günlerde insanlar oturup okuduklarında, çoğu zaman Facebook'taki durum güncellemelerine bakıyorlar. Ama bir zamanlar, çok değil yakın zaman öncesine kadar insanlar, bazısı kitaplardan olmak üzere, hikâye okurlardı. Duygu analizi bize burada da çok şey öğretebilir.

Şu an California Üniversitesi'ne bağlı Berkeley School of Information'da çalışan Andy Reagan'ın başını çektiği bir grup bilimci binlerce kitabın ve film senaryosunun metnini internetten indirerek hikâyenin her bir noktasının ne kadar mutlu ya da üzgün olduğunu kodlamaya çalıştı.

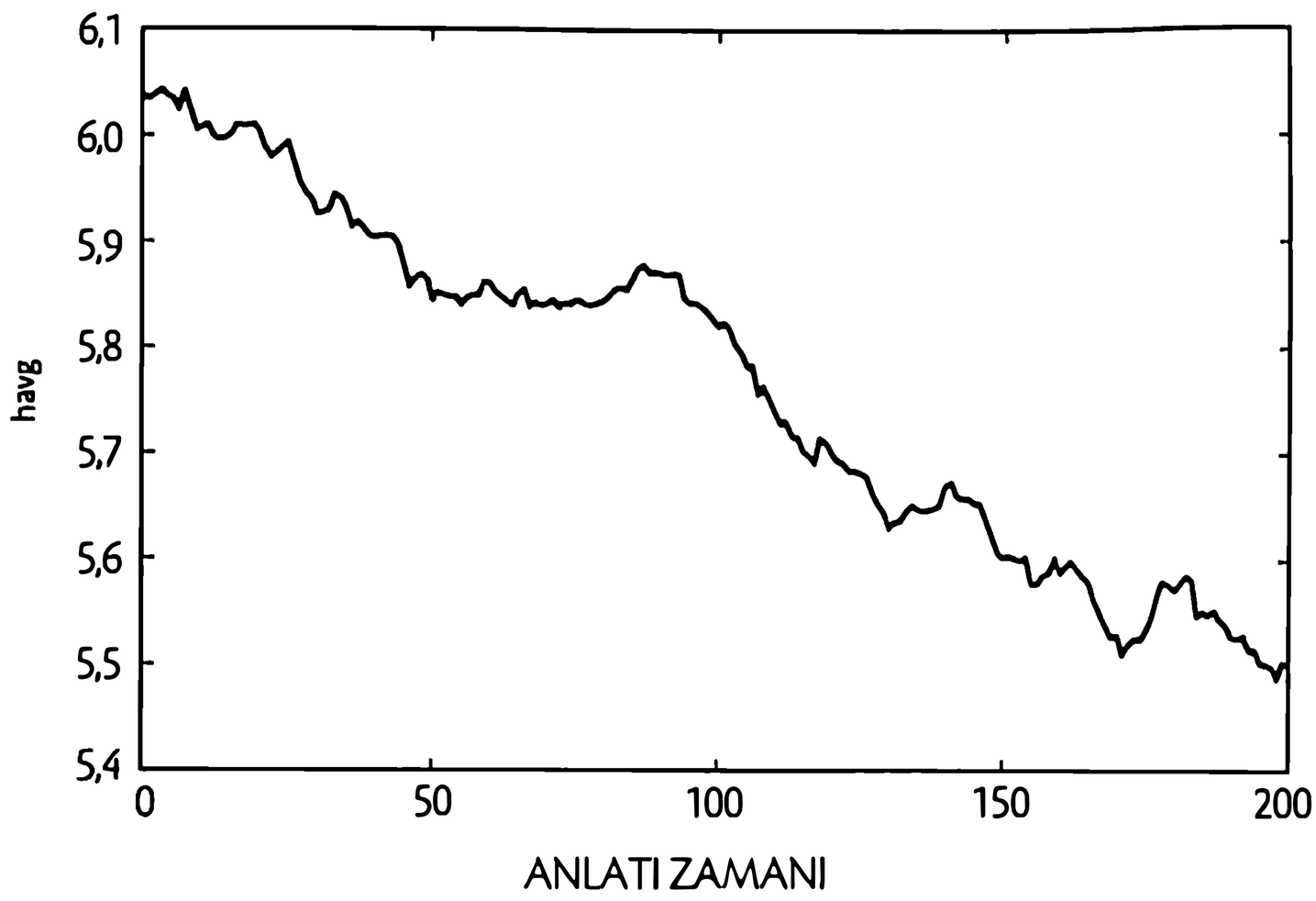
Örneğin *Harry Potter ve Ölüm Yadigârları* kitabını düşünün. İşte bu bilimcilerden oluşan ekibin olay örgüsünün kilit noktalarını tasvir ederek hikâyenin ruh halinin nasıl değiştiğine dair söyledikleri:



Duygu analizinin ruh halinde tespit ettiği birçok yükseliş ve alçalışın kilit olaylara tekabül ettiğine dikkat çekmek gerekiyor.

Çoğu hikâyenin yapısı daha basittir. Örneğin Shakespeare'in *Kral John* tragedyasını ele alalım. Bu oyunda her şey ters gider. İngiltere Kralı John'a tahttan feragat etmesi söylenir. Papaya karşı geldiği için aforoz edilir. Savaş çıkar. Yeğeni ölür, belki de öldürülür. Başka ölenler de olur. Son olarak, John huysuz keşiş tarafından zehirlenir.

İşte oyunun duygu analizi:



Başka bir deyişle, bilgisayar sadece kelimelere bakarak kötüden daha kötüye ve en kötüye giden şeyleri tespit edebilmişti.

Veya *127 Saat* filmini düşünün. Bu filmin olay örgüsünün kısa bir özeti şöyledir:

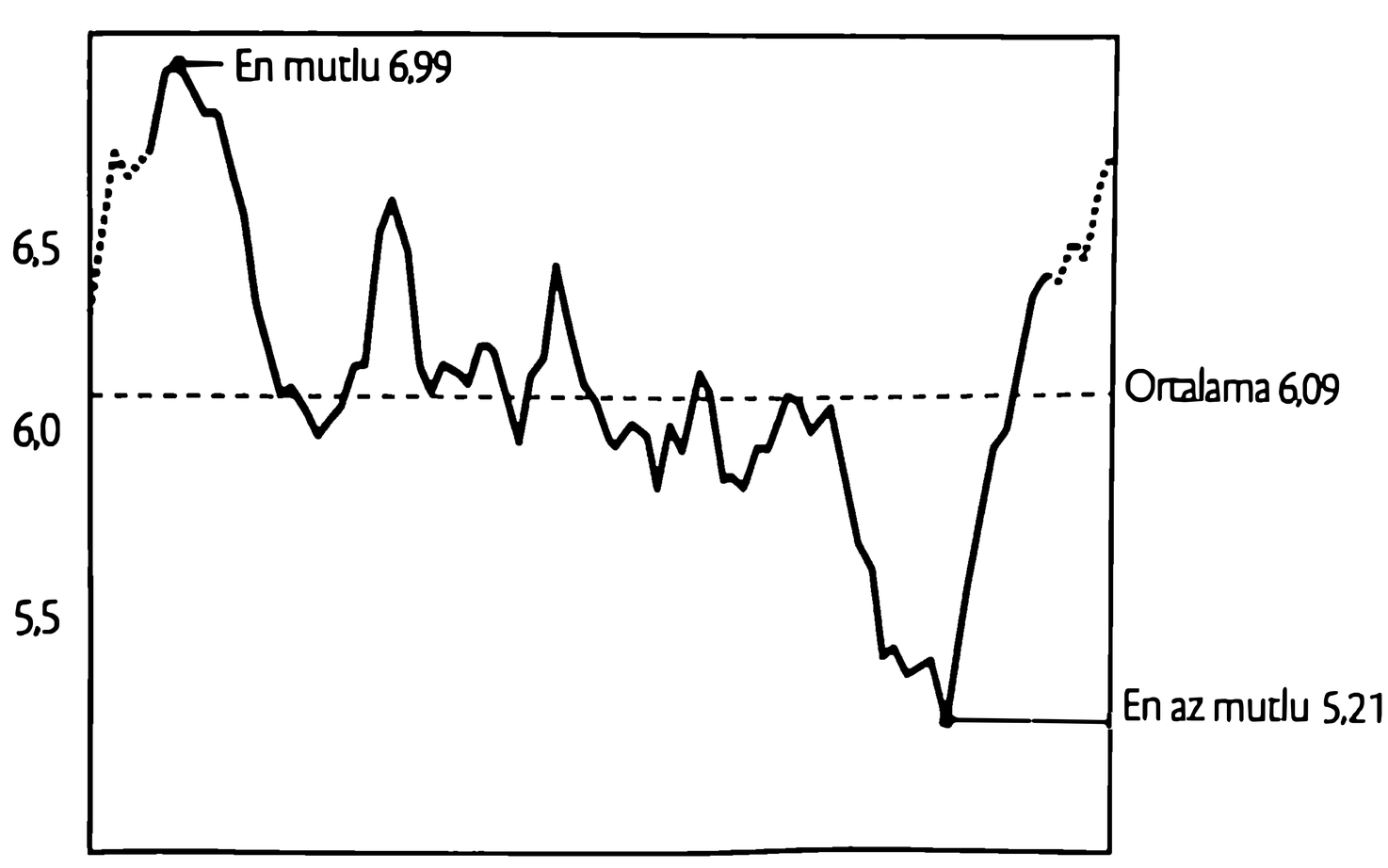
> Bir dağcı Utah'taki Canyonlands Ulusal Parkı'na yürüyüşe gider. Başka yürüyüşçülerle arkadaş olur ama sonra onlardan ayrılır. Birdenbire, kayıp düşer ve eliyle bileğini sıkışan kaya parçasına kaptırır. Birçok kez kurtulmaya çalışır ama her seferinde başarısız olur. Bunalıma girer. Nihayetinde, kolunu kesip kurtulur. Evlenir, aile kurar ve dağcılığa devam eder, ama bir daha asla giderken not bırakmayı ihmal etmez.

Bu basit şemanın yakalayamadığı küçük ayrıntılar olabilir. Örneğin *127 Saat* filmi kapana kısılma hikâyesine örnek teşkil eder, ama hikâyenin seyri içinde duyguların bir süreliğine iyiye gittiği anlar da vardır. Birçok hikâyenin genel yapısı bu altı kategoriden birine uyar. *Harry Potter ve Ölüm Yadigârları* ise istisnadır.

Cevaplayabileceğimiz daha birçok soru var. Örneğin hikâyelerin yapısı zaman içinde nasıl değişmiştir? Hikâyeler yıllar geçtikte daha karmaşık bir hal almış mıdır?

çok hangi tür hikâyeleri severler? Erkekler ile kadınlara farklı hikâye yapıları mı çekici gelir? Peki ya farklı ülkelerdeki insanlar?

İşte size, bir kez daha Reagan'ın takımının hazırladığı filmin duygu analizi:



Nihayetinde, veri olarak metin bize okur kitlesinin ve izleyicilerin gerçekte ne istediğine dair eşi görülmedik içgörüler kazandırabilir ve bu isteklerin yazarların ya da yönetmenlerin düşündüklerinden farklı olabileceğini düşündüğümüzde, önemi anlaşılır. Şimdiden bu yöne işaret eden ipuçları vardır.

Wharton School'da profesör olan Jonah Berger ve Katherine L. Milkman'ın hangi tür haberlerin paylaşıldığı konulu çalışmalarını düşünün. Bu iki isim, olumlu haberlerin mi yoksa olumsuz haberlerin mi *New York Times*'ın en çok e-postalanan listesine girme ihtimalinin daha yüksek olduğunu test etmişlerdi. Üç aylık dönemde *Times*'ta çıkan her yazıyı indirmiş, duygu analizinden yararlanarak yazıların ruh halini kodlamışlardı. Olumlu haberler arasında, "Şehre Yeni Gelmiş Safdiller Şehre Âşık Olursa" ve "Tony İnsancıllık Ödülü" vardı. "Koreli Aktrisin Ölümünde İnternet Dedikodularının Parmağı Var" ve "Almanya: Bebek Kutupayısının Besleyicisi Öldü" ise, beklendiği üzere, olumsuzdu.

Fakirlikten Zenginliğe (yükseliş)
Zenginlikten Fakirliğe (düşüş)
Kapana Kısılma (düşüş, sonra yükseliş)
İkarus (yükseliş, sonra düşüş)
Sinderella (yükseliş, sonra düşüş, sonra tekrar yükseliş)
Oidipus (düşüş, sonra yükseliş, sonra tekrar düşüş)

Profesörler haberin nereye yerleştirildiği konusunda da bilgi sahibiydi. Ana sayfada mıydı? Sağ üst köşede miydi? Sol üst köşede miydi? Ayrıca haberin ne zaman yayınlandığı hakkında da bilgileri vardı. Salı gece yarısı mı? Pazartesi sabahı mı?

Times'ın sitesinde benzer bir yerde yayınlanmış ve benzer bir zamanda çıkmış olan biri olumlu biri olumsuz iki yazıyı karşılaştırıp hangisinin e-postalanma ihtimalinin daha yüksek olduğunu bulabilirlerdi.

Peki, hangileri paylaşılıyor, olumlu yazılar mı olumsuz yazılar mı?

Olumlu yazılar. Yazarların vardığı sonuç: "Bir içerik ne kadar olumlu olursa viral olma ihtimali o kadar yüksektir."

Oysa geleneksel gazetecilik aklı, insanların şiddet ve felaket içeren haberlere daha fazla yakınlık duyduğunu iddia eder. Basın kuruluşlarının insanlara bir dolu iç karartıcı haber aktardığı doğru olabilir. Haber merkezlerindeki "daha fazla kan, daha fazla kazan" sözü boşa söylenmemiş olabilir. Fakat Whartonlı profesörlerin çalışması insanların daha neşeli haberler görmek isteyebileceğini söylüyor. Belki de yeni bir darbımesele ihtiyacımız var: "Gülümset ki e-postaya düşesin" – kafiye yapamadım ama olsun, idare edin!

* * *

Hüzünlü ve mutlu metinler hakkında bu kadarı yeter. Peki, hangi kelimelerin liberal, hangilerinin muhafazakâr olduğunu nasıl anlarız? Ve bu bize modern basın dünyası hakkında neler söyler? Biraz daha karmaşık olan bu konu bizi Gentzkow ve Shapiro'ya geri götürüyor. Hatırlarsanız, Gentzkow ve Shapiro gey evliliklerinin iki farklı gazetede farklı şekillerde anlatıldığını görüp, siyasal tarafgirliği açığa çıkarmak üzere dilden yararlanmanın mümkün olup olmayacağını merak eden iki iktisatçıydı.

Bu iki azimli genç araştırmacının yaptıkları ilk şey, Kongre Kayıtları'nın tutanaklarını incelemek oldu. Bu kayıt zaten dijital ortama aktarılmış olduğundan, Kongre'nin Demokrat üyelerinin 2005'te kullandıkları her bir kelimeyi ve Kongre'nin Cumhuriyetçi üyelerinin 2005'te kullandıkları her bir kelimeyi indirebildiler. Bu sayede, belli söz kalıplarının Demokratlar ya da Cumhuriyetçiler tarafından kullanılma ihtimalinin daha yüksek olup olmadığını görebileceklerdi.

Bazıları gerçekten de öyleydi. İşte her kategoriden birkaç örnek:

Demokratların Çok Daha Fazla Kullandıkları Söz Kalıpları	Cumhuriyetçilerin Çok Daha Fazla Kullandıkları Söz Kalıpları
Mülk vergisi	Ölüm vergisi
Sosyal sigortanın özelleştirilmesi	Sosyal sigorta reformu
Rosa Parks	Saddam Hüseyin
İşçi hakları	Özel mülkiyet hakları
Yoksul halk	Devlet harcamaları

Dildeki bu farklılıkların açıklaması ne? Demokratlar ve Cumhuriyetçiler zaman zaman aynı kavramı tarif etmek için farklı kelimeler kullanırlar. 2005'te Cumhuriyetçiler federal intikal vergisinde kesintiye gitmek istiyorlardı. Bu yüzden veraset vergisini Ölüm Vergisi diye adlandırıyorlardı (yeni ölmüş kişilere vergi bindirildiği ima ediliyordu). Demokratlar ise Mülk Vergisi diyerek zenginlere vergi kesildiğini ima etmiş oluyorlardı. Benzer şekilde, Cumhuriyetçiler Sosyal Sigorta'yı bireysel emeklilik hesaplarına kaydırmaya çalışıyorlardı. Cumhuriyetçilerin gözünde bu bir "reform"du. Demokratlar ise bunu kulağa daha tehlikeli gelen bir kelimeyle tanımlamayı tercih ediyorlardı: "Özelleştirme".

Dil farklılıkları kimi zaman da bir vurgu meselesidir. Cumhuriyetçiler de Demokratlar da bir yurttaşlık hakları kahramanı olan Rosa Parks'a muhtemelen büyük saygı duyuyorlardır. Ama Demokratlar onun adını daha çok anıyorlardı. Aynı şekilde, Demokratlar da Cumhuriyetçiler de muhtemelen, Irak'ın sabık lideri Saddam Hüseyin'in pis bir diktatör olduğunu düşünüyordur. Ama Cumhuriyetçiler Irak Savaşı'nı meşrulaştırmak amacıyla Saddam ismini sürekli gündemde tutuyorlardı. Keza işçi hakları ve "yoksul halk"a olan ilgi Demokrat Parti'nin temel ilkeleridir. "Özel mülkiyet hakları" ve "devlet harcamaları"nda kesintiyse Cumhuriyetçilerin temel ilkeleridir.

Dildeki bu farklılıklar azımsanmayacak öneme sahiptir. Örneğin 2005'te Cumhuriyetçi kongre üyeleri "ölüm vergisi" lafını 365 kez, "mülk vergisi" lafınıysa sadece 46 kez kullanmışlardı. Demokrat kongre üyelerindeyse örüntü tam tersiydi. "Ölüm vergisi" lafını sadece 35 kez kullanırken, "mülk vergisi"ni 195 kez kullanmışlardı.

Bu kelimeler bize bir kongre üyesinin Demokrat mı yoksa Cumhuriyetçi mi olduğunu anlatabiliyorsa, bir gazetenin solcu mu yoksa sağcı mı olduğunu da anlatabilirdi; araştırmacıların yaptıkları çıkarım buydu. Nasıl ki Kongre'nin Cumhuriyetçi

üyeleri "ölüm vergisi" lafını daha çok kullanarak halkın buna karşı çıkmasını sağlamaya çalışıyorsa, muhafazakâr gazeteler de aynısını yapabilirdi. Görece liberal olan *Washington Post* "mülk vergisi"ni "ölüm vergisi" lafından 13,7 kat daha sık kullanmıştı. Muhafazakâr *Washington Times* ise "ölüm vergisi" ile "veraset vergisi"ni hemen hemen aynı sıklıkta kullanmıştı.

Gentzkow ve Shapiro internetin mucizeleri sayesinde, ülkedeki çok sayıda gazetede kullanılan dili analiz edebildiler. Araştırmacılar toplamda 433 gazeteyi dijital ortama aktaran newslibrary.com ve proquest.com sitelerinden yararlanmışlardı. Ardından, gazetelerin siyasal eğilimini ölçmek amacıyla, bu türden siyasal anlam yüklü kelime ya da kelime gruplarından bin tanesinin gazetelerde ne kadar sık kullanıldığını hesapladılar. Bu ölçüme göre en liberal gazete *Philadelphia Daily News* iken, en muhafazakârı *Billings* (Montana) *Gazette*'ti.

Böyle çok geniş bir mecrada medyadaki tarafgirliğin ilk kapsamlı ölçümünü yaptığınızda, belki de basın hakkındaki en önemli soruyu cevaplayabilirsiniz: Neden bazı gazeteler sola meylederken, bazıları sağa meyleder?

İktisatçılar hemen kilit bir etken tespit etmişlerdi: Söz konusu bölgenin siyasi iklimi. Eğer bir bölge Philadelphia ve Detroit gibi genel olarak liberalse, gazetelerindeki baskın eğilim de liberal olma yönündedir. Eğer bir bölge Billings ve Amerillo (Teksas) gibi muhafazakârsa, gazetelerdeki baskın eğilim de muhafazakârlık olmaktadır. Başka bir deyişle, kanıtlar, gazetelerin okurlarına ne istiyorlarsa onu verdiğini gösteriyor.

Bir gazetenin sahibinin yayın politikasına etki edeceğini düşünebilirsiniz; fakat genel itibariyle söylersek, bir gazetenin sahibinin gazetenin siyasal çizgisine etkisi düşündüğümüzden daha azdır. Aynı kişi ya da şirket farklı pazarlarda gazete sahibi olduğunda ne olduğuna bakalım. New York Times Şirketi'ni düşünün. Gentzkow ve Shapiro'nun bulgularına göre, bu şirket, nüfusun kabaca yüzde 70'inin Demokrat partili olduğu New York'ta çıkan liberal eğilimli *New York Times*'ın sahibidir. Ama aynı zamanda, çalışmanın yapıldığı dönemde, bu ölçüme göre, muhafazakâr eğilimli olan *Spartanburg Herald-Journal*'ın da sahibiydi ve bu gazete nüfusun kabaca yüzde 70'inin Cumhuriyetçi partiye oy verdiği Spartanburg'da (Güney Carolina) yayımlanıyordu. Elbette istisnalar var: Rupert Murdoch'ın News Corporation'ı hemen herkesin muhafazakâr bulduğu *New York Post*'un sahibidir. Ama genele bakıldığında, bulgular gazetelerin eğilimini sahiplerinden çok piyasanın belirlediğini gösteriyor.

Söz konusu çalışmanın basın dünyası hakkındaki düşüncelerimize derin bir etkisi vardır. Birçok insan, özellikle de Marksistler, Amerikan gazeteciliğini zenginlerin

ya da şirketlerin kitleleri etkilemek, belki de halkı kendi siyasi görüşlerine çekmek amacıyla elinde bulundurduğu bir alan olarak görüyor. Gentzkow ve Shapiro ise gazete sahiplerinin esas saikinin bu olmadığını savunuyorlar. Amerikan basınının sahipleri esasen daha da zenginleşebilmek için kitlelerin istediği türde bir tutumu benimsiyor.

Aa, bir soru daha var; büyük, tartışmalı, hatta belki de tahrik edici bir soru. Amerikan medya kuruluşları, ortalamaya vurulduğunda, solcu mudur, sağcı mıdır? Medya ortalamaya vurulduğunda liberal midir, muhafazakâr mıdır?

Gentzkow ve Shapiro gazetelerin sol eğilimli olduğunu tespit ettiler. Ortalama bir gazete, kullandığı kelimeler bakımından, bir Cumhuriyetçi kongre üyesinden çok Demokrat kongre üyesine benziyordu.

Muhafazakâr okurlar "Ya! Gördünüz mü işte, ben demiştim!" diye haykırmaya hazırlanıyor olabilir. Birçok muhafazakâr, geçmişten beri, gazetelerin kitleleri sol görüşleri destekleyecek şekilde yönlendirdiğinden işkillenmektedir.

Fakat Gentzkow ve Shapiro "Hiç de öyle değil," diyor. Aslında liberal eğilim tam da gazete okurlarının isteklerini yansıtmaktadır. Gazete okurları, ortalamaya vurulduğunda, biraz sola meylediyorlar. (Buna dair veriler mevcuttur.) Gazeteler de, ortalamaya vurulduğunda, okurlarına talep ettikleri bakış açılarını verecek şekilde biraz sol eğilimlidir.

Ortada büyük bir kumpas falan yok. Bu sadece kapitalizm.

Gentzkow ve Shapiro'nun sonuçları, basın kuruluşlarının işleyişinin çoğu zaman gezegenimizdeki diğer tüm işletmelerinkine benzediğini ima ediyor. Nasıl ki süpermarketler insanların hangi dondurmayı istediğini bulup raflarını onunla dolduruyorsa, gazeteler de insanların hangi görüşleri istediğini tespit edip sayfalarını onunla dolduruyorlar. Shapiro bana, "Bunun herhangi bir ticari işten farkı yok," demişti. Haber, analiz ve düşünce gibi anlaşılması zor meseleleri parçalarına ayırıp sayılara döktüğünüzde, yani kelimeler olarak kavramlaştırdığınızda öğrenebileceğiniz budur.

Veri Olarak Resimler

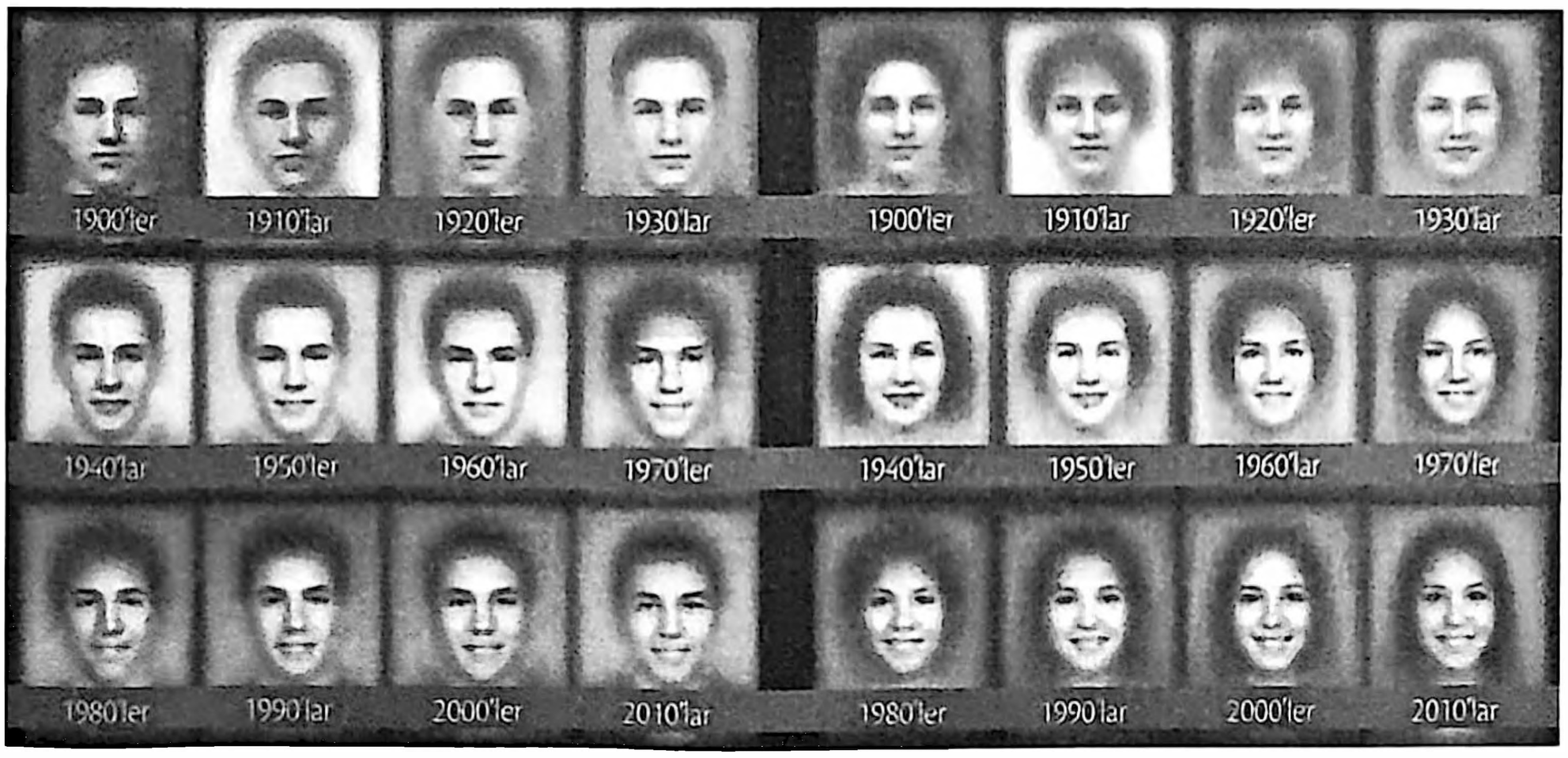
Akademisyenlerin ya da işadamlarının veriye ihtiyaç duyduklarında anket yapmaları âdettendir. Eskiden veriler, anketlerdeki işaret kutucuklarından ya da sayılardan çıkarılan, temiz ve derli toplu şeylerdi. Oysa artık durum farklıdır. Yapılandırılmış, apaçık, basit, anket temelli verilere paydos! Bu yeni çağda hayatın yolu üzerinde bıraktığımız dağınık izler verilerin başlıca kaynağı haline geliyor.

Yukarıda gördüğümüz üzere, kelimeler birer veridir. Tıklamalar birer veridir. Bağlantılar birer veridir. Yazım hataları birer veridir. Rüyada görülen muzlar birer veridir. Ses tonu veridir. Hırıltı veridir. Kalp çarpması veridir. Dalak büyüklüğü veridir. Aramalar ise, bence, en önemli veridir.

Görünen o ki, resimler de veridir.

Bir zamanlar tozlu raflardaki kitaplara ve dergilere hapsolmuş olan kelimelerin dijital ortama aktarılması gibi, resimler de albümlere ve kartpostallara hapsolmaktan kurtulmuştur. Resimler baytlara dönüştürülmüş ve cloud'a atılmıştır. Ayrıca metinler bize tarih dersi verebileceği gibi (örneğin insanların konuşma tarzlarının değiştiğini gösterirler) resimler de bize tarih dersi verebilirler (örneğin insanların poz verme şekillerinin değiştiğini gösterirler).

Brown ve Berkeley'deki dört bilgisayar bilimcisinden oluşan takımın yürüttüğü dâhiyane çalışmayı ele alalım. Dijital çağdaki bir gelişmeden istifade etmişlerdi: Birçok lise, öğrenci yıllıklarını tarayıp internete aktarmıştı. Araştırmacılar internette 1905-2013 yıllarını kapsayan 949 okul yıllığı buldular. Buna on binlerce son sınıf vesikalığı da dahildi. Bilgisayar yazılımından yararlanarak her on yıla ait resimlerden bir "ortalama" yüz yaratmayı başardılar. Başka bir deyişle, insanların burunlarının, gözlerinin, dudaklarının ve saçlarının ortalama konumunu ve konfigürasyonunu çıkardılar. İşte size son yüz küsur yılın ortalama kadın-erkek yüzleri:



Dikkatinizi çeken bir şey oldu mu? Amerikalılar –özellikle de kadınlar– gülümsemeye başlamışlar. 20. yüzyıl başındaki neredeyse sirke satan suratlar gitmiş, yerini çiçek açan suratlar almış.

Peki, değişikliğin sebebi ne? Amerikalıların daha mutlu insanlar haline gelmiş olması mı?

Alakası yok! Başka araştırmacılar bu sorunun cevaplanmasına yardımcı olabilir. Değişimin nedeni, en azından bana göre, hayranlık uyandırıcıdır. Fotoğraf ilk icat edildiğinde, insanlar fotoğraf ile tabloları aynı şey olarak görüyorlardı. Karşılaştırma yapacak başka bir şey yoktu ki! Bu nedenle fotoğraflardaki özneler tablolardaki öznelerin bir suretiydi. Portre için oturan insanlar resimleri çizilene kadar gülümseyerek poz veremeyeceklerinden, ciddi bir bakış takınıyorlardı. Fotoğraflardaki özneler de aynı bakışı benimsemişti.

Nihayetinde değişikliğe yol açan neydi? Elbette ticaret, kâr, pazarlama. 20. yüzyılın ortalarında insanların sınırlı sayıda fotoğraf çektirmeleri, film ve fotoğraf makinesi şirketi Kodak'ı hüsrana uğratıyordu. Nihayet, Kodak insanların daha fazla fotoğraf çekmeleri için bir strateji geliştirdi. Kodak, reklam kampanyasına fotoğrafla mutluluğu özdeşleştirerek başladı. Amaç insanlara çok güzel vakit geçirdiklerini başkalarına göstermek istedikleri her durumda resim çekme alışkanlığı kazandırmaktı. Tüm o mütebessim okul yıllıklarındaki fotolar bu başarılı kampanyanın bir sonucudur (keza bugün Facebook ve Instagram'da gördüğünüz çoğu fotoğraf da öyle).

Ama veri olarak fotoğraflar bize lise son sınıfların ne zaman "peynir" demeye başladığından çok daha fazlasını anlatabilir. Şaşırtıcı gelse de, resimler bize ekonominin ne âlemde olduğunu bile anlatabilir.

Şu kışkırtıcı başlıklı akademik yazıyı ele alalım: "Ekonomik Büyümeyi Uzaydan Ölçmek". Bir yazıya bunun gibi bir başlık koyulursa, ben kesin okurum. Bu makalenin yazarları J. Vernon Henderson, Adam Soreygard ve David N. Weil, gelişmekte olan ülkelerin çoğunda mevcut GSYH (Gayri Safi Yurtiçi Hasıla) ölçümlerinin yetersiz olduğu fikrinden yola çıkmışlardı. Bunun sebebi ekonomik faaliyetin büyük oranda kayıt dışı yürütülmesi ve ekonomik faaliyeti ölçmesi gereken devlet dairelerinin kaynaklarının yetersiz olmasıydı.

Yazarların kulağa epey sıradışı gelen fikri neydi? Bu ülkelerde geceleri ne kadar ışık yakıldığına bakarak GSYH'nin ölçülmesine yardımcı olabilirlerdi. Bu bilgiye dünyanın etrafını her gün on dört kez dönen ABD Hava Kuvvetleri uydusunun çektiği fotoğraflardan ulaştılar.

Geceleri yanan ışıklar neden GSYH'nin sağlam bir ölçüsü olabilir? Doğrusu, dünyanın çok yoksul bölgelerinde insanlar elektrik faturalarını ödemekte zorlanıyorlar. Bu nedenle ekonomik koşullar kötüyken, evlerde ve köylerde geceleri kullanılan ışık miktarı ciddi ölçüde azalıyor.

1998'de Asya finans krizi sırasında Endonezya'da gece ışık kullanımında keskin bir düşüş kaydedilmişti. Güney Kore'de, ekonominin fevkalade yükseliş gösterdiği

1992-2008 arasında gece ışık kullanımı yüzde 72 artmıştı. Kuzey Kore'deyse ekonominin çok kötü olduğu aynı dönemde gece ışık kullanımı azalmıştı.

1998'de Madagaskar'ın güneyinde büyük bir yakut ve gökyakut kaynağı keşfedildi. Ilakaka küçük bir kamyoncu durağı iken, büyük bir ticaret merkezi haline geldi. 1998'den önce Ilakaka'da geceleri neredeyse hiç ışık yanmazdı. Sonraki beş yıldaysa patlama yaşandı.

Makalenin yazarları geceleri ışık kullanımı verilerinin ekonomik verimliliğin kusursuz bir ölçüsü olmadığını kabul ediyorlar. Salt uyduların gece yakaladıkları ışık miktarından hareketle bir ekonominin ne durumda olduğunu kesinkes söyleyemezsiniz. Yazarlar ABD gibi gelişmiş ülkeler için bu ölçüm yöntemini kesinlikle tavsiye etmiyorlar; zira mevcut ekonomik veriler bu ülkelerde daha kesindir. Haksızlık olmasın, gelişmekte olan ülkelerde bile gece ışığının sadece resmî ölçümler kadar yararlı olduğunu da söylüyorlar. Ama hatalı resmî verilerle kusurlu gece ışığı verilerini birleştirmek bu kaynaklardan sadece birinin sunabileceğinden daha iyi bir tahmin imkânı yaratır. Başka bir deyişle, uzaydan çekilen resimlerden yararlanarak gelişmekte olan ekonomiler hakkındaki kavrayışınızı geliştirebilirsiniz.

Bilgisayar mühendisliği doktorası yapmış, sakin sesiyle dikkat çeken bir akademisyen olan Joseph Reisinger, tıpkı gece ışığını veri alan yazarlar gibi, gelişmekte olan ülkelerdeki ekonomilere dair mevcut veri kümelerinin yetersizliğinden rahatsızlık duyuyor. Reisinger, Nijerya'nın önceden gözden kaçırılmış olabilecek yeni sektörleri dikkate alarak, 2014'ün Nisan ayında GSYH tahminini güncellediğini belirtiyor.

"Nijerya Afrika'daki en büyük ekonomidir," diyor Reisinger, sesini hafifçe yükselterek, "Ama henüz bu ülke hakkında bilmek istediğimiz en basit şeyi bile bilmiyoruz."

Reisinger ekonomik performans hakkında daha net bir bilgi edinmenin yolunu arıyordu. Bulduğu çözüm, veri oluşturan şeylerin nasıl yeniden tahayyül edileceğine ve bunu yapmanın değerine ilişkin çok iyi bir örnektir.

Reisinger, Premise adında bir şirket kurdu. Bu şirkette gelişmekte olan ülkelerden bir grup işçiyi akıllı telefonlar vererek çalıştırmaya başladı. Çalışanların görevi neydi? Çevrelerinde gördükleri ekonomik değeri olabilecek ilginç şeylerin resimlerini çekmek.

Çalışanlar benzincinin ya da süpermarketlerdeki meyve reyonlarının fotoğraflarını çekiyorlar. Aynı yerlerin resimlerini tekrar tekrar çektikleri de oluyor. Resimler Premise'e yollanıyor, burada başka bir grup çalışan (bilgisayar uzmanları) fotoğrafları veriye dönüştürüyor. Şirket analistleri benzincilerdeki kuyruk uzunluklarından süpermarketlerde kaç tane elma olduğuna, bu elmaların tazeliğinden tezgâhta belirtilen fiyatlara kadar her şeyi kodluyorlar. Premise her türlü faaliyetin fotoğraflarını temel alarak ekonomik verim ve enflasyon tahminleri yapabiliyor.

Gelişmekte olan ülkelerde benzincilerdeki uzun kuyruklar ekonomik sıkıntının önemli işaretlerinden biridir. Keza elma olmaması ya da olgunlaşmamış elmalar da öyle. Premise'in Çin'de çektiği resimler 2011'de gıda enflasyonunu, 2012'de de gıda deflasyonunu resmî veriler gelmeden çok önce doğru tahmin etmesini sağlamıştı.

Premise bu bilgiyi bankalara ya da serbest yatırım fonlarına satıyor ve ayrıca Dünya Bankası'yla ortak çalışıyor.

Birçok güzel fikir gibi Premise'inki de kazandırmaya devam eden bir nimettir. Dünya Bankası yakın zaman önce Filipinler'deki kaçak sigara ekonomisinin büyüklüğünü inceleme altına aldı. Özellikle de Filipinler hükümetinin son dönemde vergi ödemeden sigara üreten üreticilerin tepesine binme çabalarının (rastgele baskınlar dahil) sonuçlarını öğrenmek istiyordu. Premise'in akıllı fikri neydi? Sokakta görülen tüm sigara paketlerinin fotoğraflarını çekmek. Çekilen fotoğrafların kaçında yasal sigaralarda olduğu gibi bandrol olduğuna bakmak. Yeraltı ekonomisinin bu yönünün 2015'te büyük olmakla birlikte, 2016'da epeyce azaldığını tespit ettiler. Her ne kadar genellikle gizlenen bir şeyi (kaçak sigaraları) görmek yeni veriler gerektiriyor olsa da, hükümetin çabaları işe yaramıştı.

* * *

Gördüğümüz gibi, veri olarak alınabilecek şeyler dijital çağda hiç beklenmedik şekillerde yeniden tasavvur edilmiş ve bu yeni bilgiden epey içgörü elde edilmiştir. Medya taraflılığının arkasındaki saiki, ilk buluşmanın nasıl iyi geçebileceğini ve gelişmekte olan ekonomilerin gerçekte ne durumda olduğunu öğrenmek henüz sadece başlangıçtır.

Bu tür verilerden Brin ve Page beyefendilerin on milyarlarca doları başta olmak üzere epey bir para kazanılmış olması da tesadüf değildir. Joseph Reisinger de fena iş yapmadı. Gözlemciler Premise'in şu an yılda on milyonlarca dolar gelir elde ettiğini söylüyorlar. Yatırımcılar yakın zaman önce şirkete 50 milyon dolar aktardılar. Demek ki bazı yatırımcılar Premise'i esasen fotoğraf çekip satma işinde, yani *Playboy*'la aynı ligde dünyanın en değerli işletmeleri arasında görüyorlar.

Başka bir deyişle, gerek araştırmacılar gerek girişimciler için şu an mevcut olan yeni türde verilerden yararlanmanın, nelerin veri olarak görüleceğine geniş bir perspektiften bakmanın değeri paha biçilmezdir. Bu günlerde bir veri bilimci veriler hakkında dar ya da geleneksel görüşle sınırlı kalmamalıdır. Süpermarket sepetlerinin doluluğu bir veridir. Elmaların olgunluğu veridir. Uzaydan çekilen fotoğraflar veridir. Dudakların bükülüğü veridir. Her şey veridir!

Ve tüm bu yeni verilerle, nihayet insanların yalanlarını yakalayabiliriz.

DÖRDÜNCÜ BÖLÜM

Dijital Hakikat Serumu

Herkes yalan söylüyor.

İnsanlar eve dönüş yolunda kaç tane içtikleri hakkında yalan söylüyorlar. Spor salonuna ne kadar sık gittikleri, yeni ayakkabılarının kaç para olduğu, o kitabı okuyup okumadıkları hakkında yalan söylüyorlar. Hasta değilken hastayım diyorlar. İrtibatı koparmayalım diyorlar, ama koparacaklarını biliyorlar. Sorun sende değil bende diyorlar. Sevmiyorlar ama seviyorum diyorlar. Üzüntüden geberiyorlar ama mutluyum diyorlar. Aslında erkeklerden hoşlanıyorlar ama kadınlardan hoşlandıklarını söylüyorlar.

İnsanlar arkadaşlarına yalan söylüyorlar. İnsanlar patronlarına yalan söylüyorlar. İnsanlar çocuklarına yalan söylüyorlar. İnsanlar anne-babalarına yalan söylüyorlar. İnsanlar doktorlarına yalan söylüyorlar. İnsanlar kocalarına yalan söylüyorlar. İnsanlar karılarına yalan söylüyorlar. İnsanlar kendilerine yalan söylüyorlar.

Ve kalıbımı basarım, insanlar anketlerde yalan söylüyorlar.

İşte sizin için hazırladığım kısa bir anket:

Öğrenciyken hiç kopya çektiniz mi? _____

Hiç aklınızdan birini öldürmek geçti mi? _____

Yalan söyleyesiniz geldi mi? Birçok insan utandığı davranışları ve düşünceleri hakkında anketlerde eksik beyanda bulunuyor. Çoğu anket isimleri gizli tutsa da, iyi görünmek istiyorlar. Buna toplumsal arzulanırlık taraflılığı deniyor.

1950 yılında yayımlanan önemli bir yazıda anketlerin bu türde taraflılıklara nasıl kurban gidebileceğine dair güçlü kanıtlar sunulmuştu. Araştırmacılar resmî kaynaklara başvurarak Denver sakinlerine dair veriler toplamıştı: Yüzde kaçı oy vermişti, bağış yapmıştı ve kütüphane kartına sahipti? Ardından yüzdelerin tutup tutmadığını anlamak için Denver sakinleri arasında anket yapmışlardı. Sonuçlar o dönemde şoke ediciydi. Sakinlerin anketlerde söyledikleriyle araştırmacıların elde ettikleri veriler apayrıydı. Kimse adını söylememiş olsa da, insanlar (az buz

değil çok sayıda insan) seçmen kaydı durumu, oy verip vermediği ve bağış yapıp yapmadığı hakkında abartılı beyanda bulunmayı seçmişti.

	ANKET	RESMİ SAYIM
Oy vermek için kaydolmuş seçmen	%83	%69
Son başkanlık seçimlerinde oy verenler	%73	%61
Son yerel seçimde oy verenler	%63	%36
Kütüphane kartı olanlar	%20	%13
Son bağış kampanyasına katılanlar	%67	%33

Geçen altmış beş yılda değişen bir şey var mı? İnternet çağında bir kütüphane kartı sahibi olmamak utanç verici olmaktan çıkmıştır. Ama utanç verici ya da istenir olan şeyler değişse de, insanların anketleri yanıltma eğilimi gücünden bir şey kaybetmemiştir.

Yakın dönemdeki bir ankette Maryland Üniversitesi mezunlarına üniversite deneyimleri hakkında çeşitli sorular soruldu. Cevaplar resmî kayıtlarla karşılaştırıldı. İnsanlar kendilerini olduklarından daha iyi gösterecek şekilde yanlış bilgi vermişlerdi. Yüzde ikiden azı 2,5'tan düşük not ortalamasıyla mezun olduğunu söylemişti (gerçek oran yüzde 11 civarındaydı). Yüzde 44'üyse geçen yıl üniversiteye bağış yaptığını söylemişti (gerçek oran yüzde 28 civarıydı).

Nitekim Donald Trump'ın 2016 yılındaki zaferini anketlerin bilememesinde de yalan söyleme alışkanlığının rolü olduğunu söylemek kesinlikle mümkündür. Anketler, Trump'a olan desteği ortalama yüzde iki civarında eksik söylemişti. Bazı insanlar Trump'ı desteklemeyi düşündüğünü söylemeye utanmış olabilir. Bazıları da aslında Trump'la aynı kafada olmasına rağmen henüz karar vermediğini söylemiş olabilir.

İnsanlar isimlerinin yer almayacağı anketlere neden yanlış bilgi verirler? Toplumsal arzulanırlık taraflılığı konusunda belki de dünyadaki en önemli uzman olan Michigan Üniversitesi'nden emekli profesör Roger Tourangeau'ya sordum. Sorunun önemli bir kısmının "pembe yalanlar"a olan zaafımızdan kaynaklandığını söyledi. "İnsanların gerçek hayatta söylediklerinin üçte bir kadarı yalandır. İnsanların huyları anketlere de yansıyor."

Bir de şöyle bir garip huyumuz var: Zaman zaman kendimize yalan söylüyoruz. "Sözgelimi çoğu insan okulda başarısız bir öğrenci olduğunu kendine itiraf etmeye yanaşmaz," diyor Tourangeau.

İnsanın kendine yalan söylemesi, neden bu kadar çok insanın ortalamanın üstünde olduğunu belirttiğine bir açıklama getirebilir. Bu problemin ölçeği nedir? Bir şirketteki mühendislerin yüzde kırkından fazlası ilk beşte olduğunu söylemişti. Üniversitedeki profesörlerin yüzde doksanından fazlası ortalamanın üstünde iş yaptığını söylüyor. Lise son sınıfların dörtte biri başkalarıyla geçinebilme becerileri konusunda yüzde birlik azınlığa dâhil olduğunu düşünüyor. Eğer kendinizi kandırıyorsanız, anketlerde de dürüst olamazsınız.

Anketlerde yalan söylememizde rolü olan etkenlerden bir diğeri de anketi yapan yabancı üzerinde iyi bir izlenim bırakma arzumuzun güçlü olmasıdır, tabii anketi yapan biri varsa. Tourangeau'nun ifade ettiği şekliyle, "en sevdiğiniz dayınıza benzeyen biri yaklaşıyor... En sevdiğiniz dayınıza geçen ay ot içtiğinizi söylemek ister misiniz?"* O eski güzel okulunuza para vermediğinizi kabul etmek ister misiniz?

Bu nedenle koşullar ne kadar kişisel meselelerden uzaklaşırsa, insanlar o kadar dürüst olacaktır. Doğru cevaplar almak için internet anketleri telefon anketlerinden, telefon anketleri de yüz yüze anketlerden daha iyidir. İnsanlar tek başlarına olduklarında, odada başkaları olduğunda kabul ettiklerinden daha fazlasını kabul edeceklerdir.

Gelgelelim, bütün anket yöntemleri hassas konularda azımsanmayacak derecede yanlış bilgiyle karşılaşacaktır. Tourangeau burada iktisatçıların sık kullandığı bir kelimeye başvurdu: "Teşvik". İnsanlara anketlerde doğruyu söyletecek bir teşvik yoktur.

O halde insanların gerçekte ne düşünüp ne yaptığını öğrenmenin yolu nedir?

Bazı örneklerde hakikate ulaşmak için atıfta bulunabileceğimiz resmî veri kaynakları vardır. İnsanlar örneğin bağış konusunda yalan söyleseler bile, belli bir bölgedeki gerçek sayıları bizatihi bağış sayılarından elde edebiliriz. Ama resmî kayıtlarda yer almayan davranışlar hakkında bir şey öğrenmeye ya da insanların ne düşündüğünü (gerçek inançlarını, hislerini ve arzularını) öğrenmeye çalışırken,

---

* Yalan söylemenin bir diğer nedeniyse anketlere gıcık olmaktır. Bu, ergenlerle ilgili her türlü araştırmada karşılaşılan çok ciddi bir sorundur ve söz konusu yaş grubunu anlayabilmemizi fevkalade zorlaştırmaktadır. Araştırmacılar ilk başlarda, bir ergenin evlat edinilmesiyle çeşitli olumsuz davranışlar (hap kullanma, alkol alma ve okuldan kaçma) arasında bir bağıntı olduğunu tespit etmişlerdi. Ama sonraki araştırmalar bu bağıntının aslında evlatlık olmadığı halde evlatlık olduğunu söyleyen ergenler (yüzde 19 oranında bir kitle) tarafından yapılan yanlış yönlendirmeyle açıklanabileceğini ortaya koydu. Sonraki araştırmalarda anlaşıldı ki, ergenlerin önemli bir kısmı anketlerde 2,10'dan uzun, 180 kilodan ağır olduğunu ya da üç çocuk sahibi olduğunu (ellerinizden öperler!) söylemişti. Başka bir anketteyse protez kolu olduğunu söyleyerek akademik araştırmacılarla dalga geçen öğrenci oranı yüzde 99 çıkmıştı.

insanların anketlere söylemeye tenezzül ettiklerinden başka bir bilgi kaynağı yoktur. Daha doğrusu, bu zamana kadar yoktu.

Büyük Veri'nin ikinci güçlü yanı budur: Belli internet kaynakları insanların başka bir yerde kabul etmeyecekleri şeyleri kabul ettirir. Bunlara dijital hakikat serumu diyebiliriz. Google aramalarını düşünün. İnsanları daha dürüst kılan koşulları düşünün. İnternet? Evet. Yalnız başına olma? Evet. Anket yapan kimsenin olmaması? Evet.

Ayrıca Google aramalarının insanlara doğruyu söyletme konusunda devasa bir üstünlüğü daha bulunuyor: Teşvikler. Eğer ırkçı şakalardan hoşlanıyorsanız, bilgisayarla ilgisi olmayan bu olguyu bir ankette paylaşmanızı teşvik edecek hiçbir şey yoktur, ama yeni ırkçı şakaların en iyilerini internette aramanız için vardır. Eğer depresyonda olduğunuzdan şüpheleniyorsanız, sizi bunu bir ankette kabul etmeye teşvik edecek hiçbir şey yoktur, ama Google'a depresyonun belirtilerini ve potansiyel tedavi yollarını yazıp arama yapmak için vardır.

Kendinize yalan söylüyor olsanız bile, Google yine de hakikati bilebilir. Seçimlerden birkaç gün önce, siz de komşularınızdan bazıları da, sandığa gidip oy kullanacağınızı düşünmekte haklı olabilirler. Ama eğer hiçbiriniz nasıl veya nerede oy verileceğine dair bilgi edinmek için arama yapmadıysanız, benim gibi veri bilimciler yaşadığınız bölgede seçime katılım oranının gerçekte düşük çıkacağını kestirebilirler. Benzer şekilde, depresyonda olabileceğinizi kendinize bile itiraf etmemişken, ağlama nöbetleri ve yataktan çıkmakta zorlanma gibi şeyleri google'layabilirsiniz. Böylece kitabın başında analiz ettiğim bir bölgedeki depresyonla alakalı aramalarda yerinizi alırsınız.

Kendi Google kullanma deneyimlerinizi düşünün. O arama kutusuna ara ara kibar çevrelerde kabullenmekte tereddüt edeceğiniz bir davranışınızı ya da düşüncenizi açığa vuran şeyler yazdığınızı tahmin ediyorum. Amerikalıların büyük bir çoğunluğunun Google'a çok kişisel şeyler söylediğine dair kanıtlar reddedilemeyecek mahiyettedir. Sözgelimi Amerikalılar "porno" aramasını "hava durumu" aramasından daha çok yapıyorlar. Bu gerçeği erkeklerin sadece yüzde 25'inin, kadınların ise sadece yüzde 8'inin pornografi izlediğini söyleyen anket verileriyle bağdaştırmak zordur.

Keza bu arama motorunun aramaları otomatik tamamlama yöntemine baktığınızda da belli bir dürüstlük dikkatinizi çekmiş olabilir. Öneriler diğer insanların yaptıkları en yaygın aramalar temelinde sunuluyor. Dolayısıyla otomatik doldurma bize insanların neleri google'ladığına dair ipuçları sunuyor. Aslında otomatik doldurma biraz yanıltıcı olabilir. Google "çük", "sikmek" ve "porno" gibi uygunsuz

addettiği belli kelimeleri önermeyecektir. Bu demektir ki, otomatik arama, insanların Google'a açtıkları düşüncelerini gerçekte olduğu kadar müstehcen göstermez. Böyle bile olsa, yine de bazı hassas içerikler kendini belli eder.

Eğer Google'a "Why is..." ["Neden..."] yazarsanız, otomatik tamamlamada en üst sırada "Neden gökyüzü beyazdır?" ve "Neden bir artık gün var?" çıkacaktır; başka bir deyişle, en çok aranan ilgili başlıklar bunlardır. Üçüncüsü ise "Neden kakam yeşil geliyor?"dur. Diğer yandan, Google otomatik tamamlama rahatsız edici olabilir. Bugün, "Is it normal to want to..." ["...istemek normal midir?"] yazarsanız, ilk öneri "öldürmek"tir. "Is it normal to want to kill..." ["...öldürmeyi istemek normal midir?"] yazdığınızdaysa, ilk öneri "ailemi"dir.

Google aramalarının dünyaya dair genellikle gördüğümüzden daha farklı bir tablo sunduğuna dair daha fazla kanıta ihtiyacımız var mı? Çocuk sahibi olma ya da olmama konusundaki pişmanlıklar konusundaki aramaları ele alalım. Bazı insanlar karar vermeden önce yanlış tercih yapmaktan korkuyorlar. Ve hemen her zaman, çocuk sahibi olmadıklarına pişman olup olmayacaklarını soruyorlar. İnsanların Google'a çocuk sahibi "olmadıklarına" pişman olup olmayacaklarını sorma ihtimalleri, çocuk sahibi "olduklarına" pişman olup olmayacaklarını sorma ihtimallerinden yedi kat fazladır.

Bazı insanlar kararlarını (çocuk yapma ya da evlat edinme) verdikten sonra Google'a pişman olduklarını itiraf ediyorlar. Bu sizi biraz şaşkına çevirebilir, ama karar sonrası rakamlar tersine dönmektedir. Çocuklu yetişkinlerin kararlarından pişmanlık duyduklarını Google'a söyleme ihtimalleri çocuksuz yetişkinlerinkinden 3,6 kat fazladır.

Bu bölüm boyunca hatırda tutmamız gereken bir uyarı: Google yakışıksız düşünceler, insanların başkalarına açamayacaklarını hissettikleri düşünceler karşısında taraflı olabilir. Ne var ki saklı düşünceleri açığa çıkarmaya çalışıyorsak, Google'ın bunları açığa çıkarma becerisi yararlı olabilir. Çocuk sahibi olmaktan pişmanlık duymak ile olmamaktan pişmanlık duymak arasındaki büyük uçurum bu örnekteki yakışık almayan düşüncenin kayda değer olduğunu söylemektedir.

Bir an durup, "Çocuk sahibi olduğuma pişmanım" gibi bir arama yapmanın ne anlama geldiğini düşünelim. Google kendini, hava durumu gibi, dün akşamki maçı kimin kazandığı gibi ya da Özgürlük Heykeli'nin ne zaman dikildiği gibi konularda doğrudan bilgi alabileceğimiz bir kaynak olarak sunar. Ama kimi zaman sansürsüz düşüncelerimizi de, bize yardımcı olabileceğine dair çok da umuda kapılmadan Google'a yazarız. Bu noktada, arama penceresi bir tür günah çıkarma hücresine dönüşür.

Her yıl, örneğin "soğuk havadan nefret ediyorum", "insanlar canımı sıkıyor" ve "üzgünüm" gibi binlerce arama yapılıyor. Elbette Google'daki bu binlerce "üzgünüm" araması belli bir günde üzgün hisseden yüz milyonlarca insanın sadece küçük bir kesimidir. Yaptığım araştırma sonucunda, bilgi aramaktan ziyade düşünceleri ifade eden aramaların, bu düşünceyi aklından geçiren insanların sadece küçük bir kesimi tarafından yapıldığını tespit ettim. Keza Amerikalıların her yıl yaptığı yedi bin "çocuk yaptığıma pişmanım" aramasının bu düşünceyi aklından geçirmiş olanların küçük bir örneklemini temsil ettiği anlaşılıyor.

Çocuklar birçok insan, hatta çoğu insan için büyük bir mutluluktur diye düşünüyorum. Annem, "Sen ve senin salak veri analizin" dediği şeyin sahip olacağı torun sayısını sınırlamasından korkmasına karşın, bu araştırma benim çocuk sahibi olma isteğimi değiştirmedi. Ama bu nahoş pişmanlık ilgi çekicidir ve insanlığın geleneksel veri kümelerinde görmemeye meylettiğimiz bir diğer veçhesidir. Kültürümüz mutlu ve mükemmel aile tablolarını mütemadiyen gözümüze gözümüze sokuyor. Birçok insan çocuk sahibi olmayı asla pişmanlık duymayacağı bir şey olarak görecektir. Ama bazıları böyle düşünmüyor. Bunu kimseye itiraf edemeyebilirler, Google hariç.

Seks Hakkındaki Hakikatler

Amerikalı erkeklerin kaçı geydir? Cinsellik araştırmalarında efsanevi bir sorudur bu. Ama sosyal bilimcilerin cevaplamakta en çok zorlandığı sorulardan da biridir. Psikologlar örneklemlerinde mahpusları ve fahişeleri sayıca daha çok dikkate alan Alfred Kinsey'in Amerikalı erkeklerin yüzde 10'unun gey olduğu tahminine artık inanmıyorlar. Temsili anketler bize bu oranın yüzde 2 ila 3 oranında olduğunu söylüyor. Ama cinsel tercih uzun zamandan beri, insanların yalan söyleme eğiliminde olduğu konulardan biri olmuştur. Kanaatimce Büyük Veri'den yararlanarak bu soruya hiç olmadığı kadar isabetli bir yanıt verebiliriz.

Birincisi, bu anket verileri hakkında bir şeyler daha söylemeliyim. Anketler hoşgörüsüz eyaletlere kıyasla hoşgörülü eyaletlerde daha fazla gey erkek olduğunu söylüyor. Örneğin Gallup'un yaptığı bir kamuoyu yoklamasına göre, gey evliliğine desteğin en yüksek olduğu eyalet olan Rhode Island'da nüfusun gey oranı, gey evliliklerine desteğin en düşük olduğu Mississippi eyaletinden neredeyse iki kat daha fazladır.

Bunu iki şekilde açıklayabiliriz. Birincisi, hoşgörüsüz eyaletlerde doğan gey erkekler hoşgörülü eyaletlere taşınıyor olabilirler. İkincisi, hoşgörüsüz eyaletlerdeki gey erkekler gey olduklarını açıkça beyan etmiyor olabilirler; hatta yalan söyleme ihtimalleri daha yüksektir.

Bir numaralı açıklamaya (geylerin hoşgörülü eyaletlere taşınması) dair bir içgörüyü yine bir Büyük Veri kaynağı olan Facebook'tan elde edebiliriz. Facebook, kullanıcılarına hangi cinsiyete ilgi duyduklarını belirtme imkânı tanıyor. İlgi duydukları cinsiyeti belirten erkek Facebook kullanıcılarının yaklaşık yüzde 2,5'i erkeklerle ilgilendiğini söylüyor; bu sayı kabaca anketlerde çıkan sayıya denk düşüyor. Ayrıca hoşgörülü eyaletler ile hoşgörüsüz eyaletlerdeki gey nüfus Facebook'ta da büyük farklılıklar gösteriyor: Facebook'ta Rhode Island'daki gey nüfusu Mississippi'dekinin iki katından daha fazladır.

Facebook insanların nasıl ve neden taşındıkları konusunda da bilgi sağlayabilir. Gey olduğunu açıkça söyleyen bir grup Facebook kullanıcısının doğduğu yerleri kodladım. Bu sayede hoşgörüsüz eyaletlerden hoşgörülü eyaletlere kaç tane gey erkeğin taşındığını doğrudan tahmin etme imkânı elde ettim. Cevap? Belli bir hareketlilik olduğu açıktır – örneğin Oklahoma City'den San Francisco'ya. Ama tahminimce, Judy Garland CD'lerini toplayıp daha açıkgörüşlü bir yere taşınan erkeklerin sayısı, hoşgörüsüz eyaletlerdeki açıktan gey nüfus ile hoşgörülü eyaletlerdekinin farkının yarısını bile açıklayamaz.*

Ayrıca Facebook bize lise öğrencilerine odaklanma imkânı veriyor. Bu özel bir grup, çünkü liseli erkekler çoğunlukla yaşayacakları yeri seçme imkânına sahip değillerdir. Eğer taşınma gey olduğunu açıkça söyleyen nüfusun eyaletten eyalete farklılık göstermesinin nedenini açıklıyorsa, bu farklılıkların liseliler arasında olmaması gerekir. Peki, lise verileri ne söylüyor? Hoşgörüsüz eyaletlerde liseli erkekler arasında gey olduğunu açıkça söyleyenlerin sayısı çok daha azdır. Mississippi'de erkek liselilerin sadece binde ikisi gey olduğunu açıkça söylüyor. Dolayısıyla mesele sadece taşınma değildir.

Eğer her eyalette benzer sayıda gey erkek doğuyorsa ve bazı eyaletlerde gey olduğunu açıklayan erkek sayısının neden çok daha fazla olduğunu taşınma tamamen açıklayamıyorsa, cinsel kimliğini saklamak büyük bir rol oynuyor olmalıdır. Bu bizi Google'a, yani birçok insanın bir sürü şahsi bilgisini paylaşmak istediğinin kanıtlandığı adrese geri götürüyor.

Porno araştırmalarını farklı eyaletlerde gerçekte kaç tane gey erkek olduğunu test etmek için kullanmanın bir yolu olabilir mi? Gerçekten böyle bir şeye imkân

---

* Bir erkeğin Judy Garland'ı tercih etmesini erkeklerle sevişme tercihleriyle özdeşleştirmemi, şaka yollu bile olsa rahatsız edici bulanlar olabilir. Gerçi kesinlikle gey erkeklerin tamamının –hatta çoğunun– divalara hasta olduğunu ima etmiyorum. Ama arama verilerinde klişeleri destekleyici bazı bulgular var. Judy Garland hakkında bilgi arayan bir erkeğin gey pornosu arama ihtimalinin düz porno arama ihtimalinden üç kat daha fazla olduğunu tahmin ediyorum. Büyük Veri bazı klişelerin doğru olduğunu söylüyor.

var. Ülke genelinde Google aramalarından ve Google AdWords'ten aldığım verileri kullanarak ulaştığım tahminlerime göre, erkek porno aramalarının yaklaşık yüzde beşinde gey erkek pornosu aranmaktadır. (Bunların arasında "gey pornosu" ifadesi dışında, popüler bir gey porno sitesi olan "Rocket Tube" gibi aramalar da vardır.)

Peki, bu durum ülkenin farklı bölgelerinde nasıl bir farklılık göstermektedir? Toplama vurulduğunda, açıktan gey porno aramaları hoşgörülü eyaletlerde hoşgörüsüz eyaletlere kıyasla daha fazladır. Bazı gey erkeklerin hoşgörüsüz eyaletlerden hoşgörülü eyaletlere taşındığı düşünüldüğünde, bu durum akla yatkındır. Ama farklar anketlerin ya da Facebook'un söylediği kadar büyük değildir. Tahminlerime göre, Mississippi'de erkeklerin porno aramalarının yüzde 4,8'i gey pornosudur: Kamuoyu yoklamalarının ya da Facebook'un ortaya koyduğu rakamlardan çok daha yüksek, Rhode Island'daki gey pornosu aramalarına dair rakamlara (porno aramalarının yüzde 5,2'si) ise makul derecede yakındır.

O halde, ABD'de kaç gey erkek vardır? Erkeklerin porno aramalarına dair bu ölçüm (kabaca yüzde 5, hemcins seks arıyor), ABD'deki gey nüfusunun gerçek büyüklüğüne dair makul bir tahmin olarak alınabilir. Bu rakama ulaşmanın daha dolaylı bir yolu da var. Biraz veri bilimi gerektiriyor. Hoşgörü ile gey olduğunu açıkça söyleyen nüfus arasındaki ilişkiden yararlanabiliriz. Bana biraz daha müsaade edin, sabrın sonu selamet!

Ön araştırmalarıma göre, belli bir eyalette gey evliliğine desteğin her yüzde yirmilik dilimi, o eyalette yaklaşık bir buçuk kat erkeğin gey olduğunu Facebook'ta açıkça bildirmesi anlamına geliyor. Bunu temel alarak, tamamen hoşgörülü olan farazi bir eyalette doğmuş kaç erkeğin (sözgelimi insanların yüzde yüzünün gey evliliğini desteklediği yerde) açıktan gey olacağını tahmin edebiliriz. Tahminlerim yüzde beş kadar olacağını gösteriyor ki, bu rakam porno aramaları verileriyle örtüşüyor. California'ya bağlı Bay Area'daki liseli erkekler, bütünüyle hoşgörülü bir çevrede büyümeye en fazla yaklaşabileceğimiz örneği teşkil ediyor. Bu öğrencilerin yaklaşık yüzde 4'ü gey olduğunu Facebook'ta açıkça beyan etmiştir. Bu da benim hesaplamalarımla uyuşuyor diyebiliriz.

Kadınlar için hemcinsini çekici bulmaya dair benzer bir tahmine henüz ulaşamadığımı belirtmeliyim. Burada porno rakamları daha yararsızdır; zira porno izleyen kadın sayısı çok daha az olduğundan örneklemin temsiliyeti azalmaktadır. Dahası porno izleyen kadınlar arasında, gerçek hayatta esasen erkekleri çekici bulan kadınların bile lezbiyen pornosu izlemekten hoşlandıkları görülmektedir. PornHub'da kadınların izlediği porno videolarının yüzde yirmisi lezbiyen pornosudur.

Amerikalı gey erkeklerin yüzde beşi elbette bir tahmindir. Bazı erkekler biseksüeldir; bazıları –özellikle de gençken– ne olduklarından emin değildir. Çok açık ki, bu sayıyı hesaplamak oy veren ya da bir filme giden insanların sayısını hesaplamak kadar kolay değildir.

Ama hesaplamamdan çıkan sonuçlardan biri açıktır: ABD'de, özellikle de hoşgörüsüz eyaletlerde çok fazla sayıda erkek hâlâ cinsel kimliğini saklamaktadır. Cinsel tercihlerini Facebook'ta açıklamıyorlar. Anketlerde kabullenmiyorlar. Dahası birçoğunun bir kadınla evli olması bile muhtemeldir.

Kadınların kocalarının gey olduğundan çok sık şüphelendikleri görülüyor. Bu işkillenmeyi şaşırtıcı derecede yaygın olan bir arama kanıtlıyor: "Kocam gey mi?" Google aramalarında "gey" kelimesinin "acaba kocam...?" ile birlikte yazılma ihtimali, ikinci sırada gelen "beni aldatıyor mu?" ifadesiyle tamamlanma ihtimalinden yüzde on daha fazladır. "Alkolik mi?" diye tamamlanmasından sekiz kat, "bunalımda"dan ise on kat daha yaygındır.

Belki de en manidar olanı, kocasının cinselliğini sorgulayan aramaların en hoşgörüsüz eyaletlerde çok daha yaygın olmasıdır. Bu soruyu soran kadınların oranının en yüksek olduğu eyaletler Güney Carolina ve Louisiana'dır. Aslında, bu sorunun en sık sorulduğu yirmi beş eyaletten yirmi birinde, gey evliliğine destek ülke ortalamasından daha düşüktür.

Google ve porno aramaları erkek cinselliği söz konusu olduğunda yegâne yararlı veri kaynağı değildir. Büyük Veri'de insanın gizli kapaklı yaşamasının ne anlama geldiğine dair çok daha fazla kanıt vardır. "Rastlaşmak" isteyen erkekler için Craigslist'e konan ilanları analiz ettim. Erkeklerle tanışma konulu bu ilanların oranı, hoşgörüsüz eyaletlerde genellikle daha fazladır. Yüzdenin en yüksek olduğu eyaletler arasında Kentucky, Louisiana ve Alabama vardır.

Gizli kapaklı yaşam hakkında daha fazla bilgi edinmek amacıyla, Google arama verilerine dönüp biraz daha ayrıntıya girelim. "Gey pornosu"ndan hemen önce ya da sonra yapılan en yaygın aramalardan biri "gey testi"dir. (Bu testlerin erkeklere eşcinsel olup olmadıklarını söylediği zannediliyor.) "Gey testi" aramaları hoşgörüsüz eyaletlerde neredeyse iki kat daha yaygındır.

"Gey pornosu" aramak ile "gey testi" aramak arasında gidip gelmek ne anlama geliyor? Muhtemelen, zihinsel ıstıraba değilse bile epey bir kafa karışıklığına delalettir. Herhalde şunu söylemek yanlış olmaz: Bu erkeklerden bazıları gey pornosuna duydukları ilginin gerçekten gey oldukları anlamına gelmediğini teyit etmeyi umuyorlar.

Google arama verileri bize belli bir kullanıcının arama tarihçesinin zaman içerisindeki seyrini görme imkânı sunmuyor. Ne var ki AOL 2006'da bir grup kullanıcısının aramalarını akademik araştırmacılara açmıştı. İşte ismi saklı bir grup kullanıcının altı günlük dönemdeki aramalarından bazıları.

Cuma 03:49:55 ücretsiz gey foto

Cuma 03:59:37 soyunma odası gey foto

Cuma 04:00:14 gey foto

Cuma 04:00:35 gey seks foto

Cuma 05:08:23 uzun bir gey testi

Cuma 05:10:00 iyi bir gey testi

Cuma 05:25:07 kafası karışık bir adam için gey testleri

Cuma 05:26:38 gey testleri

Cuma 05:27:22 ben gey miyim testleri

Cuma 05:29:18 gey resimleri

Cuma 05:30:01 çıplak erkek resimleri

Cuma 05:32:27 ücretsiz çıplak erkek resimleri

Cuma 05:38:19 ateşli gey seks resimleri

Cuma 05:41:34 ateşli erkek anal seks

Çarşamba 13:37:37 gey miyim testleri

Çarşamba 13:41:20 gey testleri

Çarşamba 13:47:49 ateşli erkek arkadan seks

Çarşamba 13:50:31 ücretsiz gey seks videoları

Bu aramaları yapan erkeğin cinselliğinden rahatsız olduğu kesindir. Ayrıca Google verileri onun gibi daha birçok erkek olduğunu söylüyor. Bu erkeklerin birçoğu aslında hemcins ilişkilere en az hoşgörü gösteren eyaletlerde yaşamaktadır.

Bu sayıların gerisindeki insanlara daha da yakından bakmak amacıyla, Mississippi'de kimliğini saklayan gey erkeklere yardım konusunda uzmanlaşmış bir psikiyatra, hastalarından herhangi birinin benimle konuşmak isteyip istemeyeceğini sordum. İçlerinden biri bana ulaştı. Altmışlarında bir emekli profesör olduğunu ve kırk küsur yıldır aynı kadınla evli olduğunu anlattı.

On yıl kadar önce, yaşadığı stresi artık kaldıramaz hale gelince, psikiyatra gitmiş ve nihayet cinselliğini kabul etmişti. Erkekleri çekici bulduğunu hep biliyormuş, ama bunun genel bir şey olduğunu ve her erkeğin sakladığını düşünüyormuş. Terapiye başladıktan kısa süre sonra, otuz yaşına merdiven dayamış bir eski öğrencisiyle ilk ve tek gey cinsel ilişkisini yaşamış; dediğine göre, "harikulade"ymiş.

Karısıyla seks yapmıyorlar. Evliliğini bitirirse ya da bir erkekle açıktan ilişki kurarsa ömür boyu kendini suçlu hissedeceğini söylüyor. Hayatı boyunca aldığı büyük kararların her birinden pişmanlık duyuyor.

Emekli profesör ve karısı aşksız, sekssiz bir gece daha geçirecekler. Muazzam ilerlemeye karşın hoşgörüsüzlüğün sürmesi milyonlarca Amerikalının aynısını yapmasına yol açacak.

* * *

Erkeklerin yüzde beşinin gey olduğunu ve pek çoğunun da kimliğini gizlediğini öğrenmek sizi afallatmıyor olabilir. Ama insanların büyük kısmının bunu duyduğunda afallayacağı zamanlardan bugünlere geldik ve bugün bile bazı yerlerde insanların çoğu bu hisse kapılacaktır.

2007'de, o zamanki İran Cumhurbaşkanı Mahmud Ahmedinejad, "İran'da sizin ülkenizde olduğu gibi eşcinsel yok. Bizde öyle şeyler olmaz," demişti. Aynı şekilde, 2014'te Rusya'nın Kış Olimpiyatları'na evsahipliği yapmaya hazırlandığı dönemde, Soçi Belediye Başkanı Anatoliy Pahomov gey bireyler hakkında, "Bizde onlardan bulunmaz," demişti. Ama internet kullanıcılarının etkinliği, Soçi'de ve İran'da gey pornosuna epey bir ilgi olduğunu ortaya koyuyor.

Bu durum bariz bir soruyu gündeme getiriyor: Bugün ABD'de hâlâ şaşırtıcı görülen yaygın bir cinsel eğilim var mıdır? Neyi yaygın gördüğünüze ve ne kadar kolay şaşırdığınıza göre değişir.

PornHub'da en fazla aranan şeylerin çoğu şaşırtıcı değildir: Erkeklerde "genç", "üçlü" ve "*blowjob*" gibi kelimeler, kadınlardaysa "ateşli sevişme", "meme ucu emme" ve "*pussy* yiyen erkek"tir.

Ana akımı bırakırsak, PornHub verileri aklınızın ucundan bile geçmeyebilecek bazı fetişlerin var olduğuna işaret eder. "Anal elma" ve "seks yapan dolgu hayvan" arayan kadınlar vardır. "Sümük fetişi" ve "çıplak çarmıha germe" araması yapan erkekler vardır. Ama bu aramalar nadirdir; bahsi geçen devasa porno sitesinde bile ayda on kadardır.

Başka bir alakalı konu da PornHub verilerini gözden geçirdiğimizde ortaya çıkıyor: Orada herkes için biri vardır. Kadınların çoğu zaman "uzun" erkek, "koyu tenli" erkek ve "yakışıklı" erkek aramaları şaşırtıcı değildir. Ama kimi zaman da "kısa" erkek, "açık tenli" erkek ve "çirkin" erkek arıyorlar. "Sakat" erkek, "siki küçük tombul erkek" ve "çirkin şişman yaşlı erkek" arayan kadınlar vardır. Erkekler sık sık "ince" kadın, "büyük memeli" kadın ve "sarışın" kadın arıyorlar. Ama bazen de "şişman" kadın, "küçük memeli" kadın ve "yeşil saçlı" kadın arıyorlar. "Kel"

kadın, "cüce" kadın ve "memesiz" kadın arayan erkekler de vardır. Bu veriler uzun, koyu tenli ve yakışıklı ya da ince, göğüsleri büyük ve sarışın olmayanlar için sevindirici olabilir.*

Peki, hem yaygın hem de şaşırtıcı olan diğer aramalar nelerdir? Erkekler arasında en yaygın olan 150 aramadan bana göre en şaşırtıcı olanı, Freud'dan bahsettiğim kısımda ele aldığım ensest aramalarıdır. Erkek arzularının az tartışılan diğer başlıklarından ikisi de "*shemale*" [trans kadın] (en fazla arananlarda 77. sırada) ve "nine"dir (110. sırada). Toplama bakıldığında, erkeklerin PornHub aramalarının yaklaşık yüzde 1,4'ü penisli kadın aramasıdır; yaklaşık yüzde 0,6'sı (otuz dört yaşın altındaki erkeklerde yüzde 0,4'ü) ise yaşlı kadın aramasıdır. Erkeklerin yaptığı 24 bin PornHub aramasından sadece 1'i açıkça 12 yaştan küçük çocuk aramasıdır; bunun nedeni, malum, PornHub'da her türlü çocuk pornosunun yasak olması ve çocuk pornosu bulundurmanın yasadışı olması olabilir.

Kadınların PornHub'da yaptıkları aramaların başında, en baştan uyarayım, birçok okuru rahatsız edecek olan bir pornografi türü geliyor: Kadına yönelik şiddet içeren seks. Kadınların düz porno aramalarının tam yüzde 25'i kadınlarda acı ve/veya aşağılanmayı öne çıkarıyor: "Acıtan anal bağırma", "herkesin içinde utandırma" ve "aşırı vahşi *gangbang*" bu örneklerden bazılarıdır. Yüzde beşi karşı tarafın rızası dışında seks ("tecavüz" ya da "zorla" seks) arıyor, ama bu videolar PornHub'da yasaklıdır. Tüm bu terimler için arama oranları erkeklere kıyasla kadınlar arasında en az iki kat daha fazladır. Veri analizlerime göre, eğer kadına şiddet uygulanan bir porno türü varsa, kadınlara neredeyse her zaman orantısız şekilde çekici geldiği tartışma götürmez.

Elbette bunu kabullenmeye çalışırken, fantezi ile gerçek yaşam arasında bir fark olduğunu unutmamak gerçekten önemlidir. Evet, PornHub'u ziyaret eden az sayıdaki kadın arasında tecavüz görüntülerini arayan –ama bulamayan– bir altküme vardır. Malumu ilam etmek gibi olacak, ama bu durum kadınların gerçek hayatta tecavüze uğramak istedikleri anlamına gelmez ve kesinlikle tecavüzü olduğundan daha önemsiz bir suç olarak göstermez. Porno verilerinin bize anlattığı şey şudur: İnsanların kimi zaman, keşke kurmasam dedikleri ve asla başkalarına bahsedemeyecekleri fantezileri olabilir.

* * *

---

* Bu verilerin optimum randevu stratejileri için de anlamı olduğunu düşünüyorum. Birilerine teklif edip epey bir reddedilseniz bile bunu üstünüze alınmamanız gerektiği açıktır. Bu süreç size nihayetinde sizin gibi birini gerçekten çekici bulacak eşi bulma imkânı sağlayacaktır. Bir kez daha, nasıl görünürseniz görünün, size göre bir insan var. Bana güvenin.

Sırlar sadece fantezi yuvaları değildir. Seks söz konusu olunca, insanlar sır küpü gibidir: Örneğin ne kadar sık seks yaptıklarını saklarlar.

Giriş kısmında, Amerikalıların her yıl satılan kondom sayısından çok daha fazla kondom kullandıklarını söylediklerine dikkat çekmiştim. Bu yüzden bunun sadece seks sırasında gerçekte olduğundan daha çok kondom kullandıklarını söylemeleri anlamına geldiğini düşünebilirsiniz. Kanıtlar ne kadar sık seks yaptıklarını da abarttıklarını gösteriyor. On beş ila kırk dört yaş arasındaki kadınların yaklaşık yüzde 11'i cinsel açıdan aktif olduğunu, şu an hamile olmadığını ve doğum kontrol hapı kullanmadığını söylüyor. Bilimciler, çiftlerin kaç defa seks yaptıklarına dair görece tutucu varsayımlarda bulunulsa bile, her ay yüzde onunun hamile kalması gerektiğini düşünüyor. Bu sayı ABD'deki toplam hamilelik sayısını aşacaktır (çocuk doğurma yaşına gelmiş kadınlarda 113'te 1'i). Seks takıntılı kültürümüzde o kadar da sık seks yapmadığınızı kabullenmek zor olabilir.

Ama anlayış ya da tavsiye arıyorsanız, bu konuda da Google'a açılmak için bir teşvik bulabilirsiniz. Google'da seks yapmak isteyen eş konusundaki aramalar, konuşmak istemeyen evli partnerle ilgili aramalardan on altı kat daha fazladır. Nikahsız yaşayan çiftler söz konusu olduğunda, seks istemeyen partnerden yana şikâyet, mesajına yanıt vermeyen partnere ilişkin şikâyetten beş buçuk kat daha fazladır.

Google aramaları bu sekssiz ilişkilerin birçoğunda sanık sandalyesine beklenmedik birini oturtuyor. Erkek arkadaşının seks yapmak istemediğinden şikâyet edenlerin sayısı, kız arkadaşının seks yapmak istemediğinden şikâyet edenlerin iki katıdır. Şu zamana kadar, erkek arkadaşlar konusundaki bir numaralı şikâyet, "erkek arkadaşım benimle sevişmiyor" aramasıdır. (Google aramaları cinsiyete göre dağılmıyor, ama bir önceki analiz erkeklerin yüzde 95'inin heteroseksüel olduğunu söylediğine göre, "erkek arkadaş" konulu aramaların pek azının erkeklerden geldiğini tahmin edebiliriz.)

Bunu nasıl yorumlamak gerekir? Gerçekten de kadınlara kıyasla erkeklerin seksten daha uzak durdukları anlamına gelir mi? Her zaman değil. Daha önce belirtildiği üzere, Google aramaları insanların konuşmayı pek sevmedikleri meselelerde taraflı olabilir. Erkekler kız arkadaşlarının cinsel ilgisizliğinden arkadaşlarına bahsetme konusunda, kadınların arkadaşlarına erkek arkadaşlarının cinsel isteksizliğinden bahsetmelerine kıyasla daha rahat olabilirler. Yine de, Google verileri erkek partnerlerin seksten uzak durma ihtimallerinin kadın partnerlerden iki kat daha fazla olduğunu söylemiyor olsa bile, insanların düşündüğünden daha çok sayıda erkeğin seksten uzak durduğunu pekâlâ söylüyor.

Google verileri insanların seksten sık sık uzak durmalarını açıklayabilecek bir neden de sunuyor: Büyük ve çoğu yersiz endişeler. Erkeklerin endişeleriyle başlayalım. Erkeklerin penislerinin büyüklüğü konusunda tedirginlik duyduklarını eskiden beri biliyoruz, ama bu endişenin ne kadar derin olduğunu bilmiyor olabiliriz.

Erkekler Google'a vücutlarındaki hiçbir organı penislerini sordukları kadar çok sormuyorlar: Akciğerleri, karaciğerleri, ayakları, kulakları, burunları, boğazları ve beyinleri hakkında sordukları soruların toplamından daha fazla. Erkeklerin penislerini nasıl büyütecekleri konusunda yaptıkları aramalar, bir gitarın akordunun nasıl yapılacağı, nasıl omlet pişirildiği ya da tekerlek değiştirildiği konusundaki aramalarından daha çoktur. Erkeklerin steroidler hakkında en çok google'ladıkları kaygıları sağlıklarına zarar verip vermeyeceği değil, daha ziyade steroid almanın penislerinin boyunu küçültüp küçültmeyeceğidir. Erkeklerin yaşlandıkça bedenleri ya da zihinlerinin nasıl değişeceğiyle ilgili en fazla google'ladıkları soru, penislerinin küçülüp küçülmeyeceğidir.

Ufak bir not: Erkeklerin cinsel organıyla ilgili Google'a en çok sorulan sorulardan biri, "penisim ne kadar büyük?"tür. Erkeklerin cetvele değil de Google'a başvurmaları, kanaatimce, yaşadığımız dijital çağın özsel bir ifadesidir.*

Kadınlar erkeklerin penis büyüklüğünü umursuyorlar mı? Google aramalarına göre, nadiren. Kadınların partnerlerinin fallusuyla ilgili yaptıkları her bir aramaya karşılık, erkekler penisleri hakkında kabaca 170 arama yapıyor. Kadınlar nadiren de olsa partnerlerinin penisi hakkında kaygılarını dile getirdiklerinde, konu büyüklüğü oluyor, ama illa küçüklüğünden yakınmıyorlar. Google'da partnerinin penis büyüklüğüne ilişkin şikâyette bulunanların yüzde 40'ından fazlası çok büyük olduğunu söylüyor. "Seks sırasında" ile en fazla google'lanan kelimelerin başında "acı" geliyor. ("Kanama", "işeme", "ağlama" ve "osurma" da ilk beştedir.) Oysa erkeklerin penis büyüklüğünü değiştirmekle alakalı yaptıkları aramaların sadece yüzde birinde "nasıl küçültebilirim?" sorusu sorulmaktadır.

Erkeklerin seksle ilgili en çok sordukları ikinci soru, cinsel münasebetlerini nasıl daha uzun tutabilecekleridir. Bir kez daha, erkeklerin güvensizlikleriyle kadınların kaygılarının örtüşmediği görünüyor. Erkek arkadaşını nasıl en çabuk şekilde orgazma ulaştırabileceğini soran arama sayısıyla nasıl daha uzun sürede orgazm olabileceğini soran arama sayısı kabaca aynıdır. Aslında kadınların erkek arkadaş-

---

* Kitabıma *Penisim Ne Kadar Büyük? Google Aramalarının İnsan Doğası Hakkında Öğrettikleri* adını koymak istemiştim, ama editörüm bu başlığa sahip bir kitabı satmanın zor olduğunu, örneğin havaalanındaki bir kitapçıda insanların bu başlıkta bir kitabı almaya utanacaklarını söyledi. Katılıyor musunuz?

larının orgazmları konusundaki en yaygın kaygısı ne kadar zamanda olduğu değil, neden olmadığıyla alakalıdır.

Erkekler söz konusu olduğunda bedensel görünüm meselelerinden sıkça bahsetmeyiz. İnsanın görünüşü konusunda kadınlarda daha yoğun bir ilgi olduğu doğrudur, ama klişelerin söylediği kadar tek yanlı bir durum arz etmez. İnsanların ziyaret ettikleri internet sitelerini ölçen Google AdWords analizime göre, güzel ve fit olmaya duyulan ilginin yüzde 42'si, kilo vermeye duyulan ilginin yüzde 33'ü ve estetik ameliyata duyulan ilginin de yüzde 39'u erkeklere aittir. Göğüsle ilgili "nasıl yapılır?" aramalarının yaklaşık yüzde 20'si erkekte göğüsten kurtulmanın nasıl mümkün olduğuyla alakalıdır.

Ama bedenlerine güven duymayan erkeklerin sayısı çoğu insanın düşündüğünden fazla olsa bile, yine de kadınlar görünüşlerinden yana duydukları güvensizlik söz konusu olduğunda erkekleri solluyorlar. Peki, bu dijital hakikat serumu kadınların kendileri hakkındaki şüphelerine dair ne anlatıyor? ABD'de her yıl yedi milyondan fazla meme implantı araması yapılıyor. Resmî istatistikler yaklaşık 300 bin kadının her yıl bu ameliyatı olduğunu söylüyor.

Son dönemlerde birçok kadının neyi sevip neyi sevmediği konusundaki görüşleri tersine dönmüş olsa da, kadınlar arka tarafları konusunda da epey bir güvensizlik içindeler.

2004'te ABD'nin bazı bölgelerinde kalça ameliyatıyla alakalı en yaygın aramada "nasıl küçültürüm?" sorusu soruluyordu. Kalçasını büyütme arzusu siyah nüfusun ağırlıklı olduğu bölgelerde yoğunlaşmıştı. Ama 2010'dan itibaren daha büyük bir kalçaya sahip olma arzusu ABD'nin geri kalanında da güçlendi. Sonraki dört yılda bu ilgi üç katına çıktı, ama aynısı kalça büyüklüğü için oldu mu bilmiyorum. 2014'te kalçasını nasıl daha büyük hale getirebileceğini soran aramalar her eyalette nasıl küçültebileceğini soran aramalardan daha fazlaydı. Bugünlerde, ABD'de meme implantı için yapılan her beş Google aramasına karşılık, bir tane kalça implantı araması yapılmaktadır. (Teşekkürler Kim Kardashian!)

Kadınların büyük bir popoya sahip olma tercihlerindeki artış erkeklerin tercihleriyle örtüşüyor mu? İlginçtir, evet. "Büyük göt pornosu" aramaları, yine eskiden siyahların yaşadıkları bölgelerde yoğunlaşmışken, son dönemde tüm ABD genelinde popüler hale gelmiştir.

Erkekler kadınların bedenlerinde başka ne görmek isterler? Daha önce belirtildiği ve çoğunuzun gayet iyi bildiği üzere, erkekler büyük memelere bayılırlar. Jenerik olmayan pornografik aramaların yaklaşık yüzde 12'si "büyük meme" diyor. Bu rakam, "küçük meme"li porno arama rakamlarının neredeyse yirmi katıdır.

Bununla birlikte, erkekler kadınların meme implantı taktırmalarını istiyor diyemeyiz. Büyük memeli porno aramalarının yüzde üç kadarı doğal meme görmek istediğini açıkça söylüyor.

Google'da erkeklerin karıları ve memeleri hakkında yaptığı aramalarda, "onu implant yaptırmaya nasıl ikna ederim?" sorusu ile karılarının neden implant istediğine dair şaşkınlık ifadesi eşit şekilde dağılmıştır.

Veya erkeklerin kız arkadaşlarının memeleri hakkındaki en yaygın aramalarına bakalım: "Kız arkadaşımın memelerini seviyorum." Erkeklerin bu aramayı yaparken Google'da ne bulmayı umdukları ise muammadır.

Erkekler gibi kadınlar da cinsel organları hakkında sorular soruyorlar. Aslında kadınlar vajinaları hakkında neredeyse erkeklerin penisleri hakkında sorduğu kadar çok soru soruyorlar. Kadınların vajinaları hakkındaki endişeleri çoğu zaman sağlıkla alakalıdır. Ama sorularının en az yüzde 30'u başka kaygıları ele veriyor. Kadınlar vajinalarını nasıl tıraş edecekleri, nasıl sıkılaştıracakları ve tadını nasıl daha güzel yapabilecekleriyle ilgileniyorlar. Daha önce bahsedildiği üzere, çarpıcı derecede yaygın kaygılardan biri kokusunu nasıl giderebilecekleridir.

Kadınların bu konuda en sık kaygı duydukları konuların başında vajinalarının balık gibi kokması geliyor; ardından sirke, soğan, amonyak, sarımsak, peynir, ter, çiş, ekmek, çamaşır suyu, bok, ter, metal, ayak, çöp ve çürümüş et kokusu geliyor.

Genel itibariyle, erkekler partnerlerinin cinsel organlarına ilişkin pek fazla Google araması yapmıyorlar. Erkeklerin kız arkadaşlarının vajinası hakkında yaptıkları arama sayısı kadınların erkek arkadaşlarının penisi hakkında yaptığı arama sayısıyla hemen hemen aynıdır.

Erkekler partnerlerinin vajinası hakkında arama yaptıklarında, bu genellikle kadınların en fazla kaygı duydukları konuda yakınmak için oluyor: Kokusu. Erkekler, çoğunlukla, bir kadına vajinasından kötü koku geldiğini kalbini kırmadan nasıl söyleyebileceklerini bulmaya çalışıyorlar. Ama kimi zaman da, erkeklerin koku hakkındaki soruları kendi güvensizliklerini ele veriyor. Erkekler ara ara kokudan yararlanarak aldatılıp aldatılmadıklarını öğrenmeye çalışıyorlar: Mesela kondom ya da başka bir erkeğin sperminin kokusunu alıyorlarsa?

Tüm bu gizli güvensizlikleri nereye oturtmalıyız? Bu noktada size güzel bir haberim var. Google bize bu kadar endişelenmememiz için haklı nedenler olduğunu söylüyor. Cinsel partnerlerimizin bizi nasıl algıladığına dair en derin korkularımızın çoğu yersizdir. Partnerler tek başınayken, yani bilgisayarın başına oturduklarında, yalan söylemeleri için hiçbir neden olmadığında, yüzeysellikten epey uzak ve anlayışlı olduklarını gösteriyorlar. Aslında kendi bedenlerimizi yar-

gılamakla o kadar meşgulüz ki, başkalarınınkini yargılamaya pek dermanımız kalmıyor.

Ayrıca Google'da cinsellikle alakalı aramalarda ortaya çıkan büyük kaygılardan ikisi arasında muhtemelen bir bağlantı vardır: seks yapılmaması ile kişinin cinsel çekiciliği ve performansı hakkında duyduğu kaygı. Bunlar birbiriyle ilişkili olabilir. Belki de cinsellik konusunda biraz daha az kaygı duysak, daha fazla seks yapabiliriz.

Google aramaları seks konusunda başka ne anlatıyor olabilir? Kimin daha verici olduğunu görmek için bir cinsler arası muharebe yapabiliriz. Karşı cinse daha iyi oral seks yapma yollarını bulmak için yapılan aramaları ele alalım. Erkekler mi daha çok ipucu arıyor, yoksa kadınlar mı? Cinsel açıdan kim daha vericidir, erkekler mi kadınlar mı? Kadınlar tabii, ne sandınız! Tüm artısını eksisini katıp hesapladığımda, partnerlerine nasıl daha iyi oral seks yapabileceği konusunda tavsiye arayan kadınların oranı erkeklerin iki katıdır.

Erkekler oral seks yapma konusunda ipuçları aradıklarındaysa, genellikle başka bir insanı nasıl memnun edebilecekleriyle ilgilenmiyorlar. Erkeklerin kendilerine oral seks yapılması konusunda yaptıkları arama sayısı, bir kadına nasıl orgazm yaşatabileceklerine dair aramalarına eşittir. (Google arama verilerinde ulaştığım en gözde verilerimden biri buydu.)

Nefret ve Önyargılar Hakkındaki Hakikatler

Utancın gölgesinden çıkamayan ve dolayısıyla insanların sırlarının olduğu tek konu seks ve aşk değildir. Birçok insan önyargılarını kendine saklama eğilimindedir; bunun sebepsiz yere olduğu da söylenemez. Bugün birçok insanın başka bir insanı etnik kimliği, cinsel yönelimi ya da dininden ötürü yargıladığını kabul ettiği takdirde hukuk önüne çıkacağını düşünmesini bir ilerleme olarak gördüğünüzü varsayıyorum. Ama birçok Amerikalı hâlâ bu şekilde yargılarda bulunmaktadır. (Okuru uyarayım, bu kısımda da rahatsız edici malzemelerle karşılaşacaksınız.)

Bunu kullanıcıların kimi zaman "Siyahlar neden kabalar?" ya da "Yahudiler neden kötüler?" gibi sorular sordukları Google'da görebilirsiniz. Aşağıda, çeşitli gruplarla ilgili aramalarda en çok kullanılan beş olumsuz kelime sırayla verilmiştir:

	1.	2.	3.	4.	5.
Afrikalı-Amerikalılar	kaba	ırkçı	aptal	çirkin	tembel
Yahudiler	günahkâr	ırkçı	çirkin	düzenbaz	açgözlü
Müslümanlar	günahkâr	terörist	kötü	şiddet eğilimli	tehlikeli
Meksikalılar	ırkçı	aptal	çirkin	tembel	mal
Asyalılar	çirkin	ırkçı	gıcık	aptal	düzenbaz
Geyler	günahkâr	sapkın	aptal	gıcık	bencil
Hıristiyanlar	aptal	deli	mal	uçmuş	yanlış yolda

Bu klişeler arasında birkaç örüntü öne çıkıyor. Örneğin Afrikalı Amerikalılar "kaba" klişesiyle muhatap olan tek gruptur. Neredeyse her grup "aptal" klişesine maruz kalıyor; Yahudiler ve Müslümanlar hariç. "Günahkâr" klişesiyse Yahudiler, Müslümanlar ve geyler için kullanılıyor, ama Siyahlar, Meksikalılar, Asyalılar ve Hıristiyanlar için kullanılmıyor.

Müslümanlar terörist diye damgalanan tek gruptur. Bir Müslüman Amerikalı bu klişeyi besleyecek şeyler yaptığında, hemen korkunç tepkiler verilmektedir. Google arama verileri sayesinde, bu tür nefret temelli öfke patlamalarını dakika dakika gözlemleyebiliyoruz.

2 Aralık 2015'te California'ya bağlı San Bernardino'da insanların üzerine rastgele ateş açılan saldırıdan kısa süre sonra olanları düşünelim. Rizwan Farook ve Tashfeen Malik, o sabah, Farook'un mesai arkadaşlarının yaptığı toplantıyı yarı otomatik silah ve yarı otomatik tüfeklerle basıp on dört kişiyi öldürdüler. Akşamleyin, saldırganların Müslüman olduklarını düşündüren adları medyada duyurulduktan dakikalar sonra, rahatsız edici sayıda Californialı, Müslümanları ne yapmak istediğine karar vermişti: Öldürmek.

O sırada California'da yapılan "Müslümanlar"la alakalı aramaların başında "Müslümanları öldürmek" geliyordu. Toplama bakıldığındaysa, Amerikalılar "Müslümanları öldürmek" kalıbını "martini tarifi", "migren belirtileri" ve "Cowboys kadrosu" ile hemen hemen aynı sıklıkta arıyorlardı. San Bernardino saldırısından sonraki günlerde, "İslamofobi" kaygısı duyan her Amerikalıya karşı, "Müslümanları öldürme" araması yapan bir Amerikalı vardı. Nefret aramaları saldırıdan önce Müslümanlarla ilgili tüm aramaların takriben yüzde 20'sini oluştururken, saldırıdan sonraki saatlerde bu oran yüzde 50'yi aşmıştı.

Bu dakika dakika dökülmüş arama verileri bu öfkeyi sakinleştirmenin ne kadar zor olabileceğine dair de ipuçları sunabilir. Saldırıdan dört gün sonra, o zamanki başkan Obama, televizyonun en çok izlendiği saatlerde halka hitaben bir konuşma yaptı. Amerikalıları hükümetin hem terörizmi durdurabileceğine, hem de daha önemlisi bu tehlikeli İslamofobiyi dindirebileceğine ikna etmek istemişti.

Obama sağduyumuza seslenerek, kapsayıcılık ve hoşgörünün öneminden dem vurmuştu. Retoriği hem güçlü hem de etkileyiciydi. *Los Angeles Times* Obama'yı, "Korkunun yargılarımızı gölgelemesine izin vermenin tehlikelerine karşı" uyardığı için övmüştü. *New York Times* konuşmayı hem "ödün vermez" hem "sakinleştirici" bulmuştu. Think Progress adlı internet sitesi de konuşmayı, "Amerikalı Müslümanların yaşamlarını korumaya odaklanmış iyi yönetişimin zorunlu bir aracı" olarak niteleyip övmüştü. Başka bir deyişle, Obama'nın konuşması büyük bir başarı olarak görülmüştü. Peki, öyle miydi?

Google arama verileri böyle söylemiyor. O dönemde Princeton'da çalışan Evan Soltas'la birlikte verileri inceledik. Başkanın konuşmasında, "Hangi inançtan olursa olsun tüm Amerikalılar ayrımcılığı reddetmekle mükelleftir" denmişti. Ama Müslümanları "terörist", "kötü", "şiddet eğilimli", "zararlı" diye damgalayan aramalar konuşma sırasında ve hemen sonrasında iki katına çıkmıştı. Başkan Obama şunları da söylemişti: "Bu ülkeye kimi kabul edeceğimizi belirlerken din testi yapılmasını reddetmek de sorumluluğumuzdur." Ama o dönem yana döne başlarını sokacak bir yer arayan, çoğunluğunu Müslümanların oluşturduğu Suriyeli mültecilere ilişkin olumsuz aramalar yüzde 60 artarken, Suriyeli mültecilere nasıl yardım edileceğini soran aramalar yüzde 35 azalmıştı. Obama Amerikalılara "Unutmayın!" diyordu, "Özgürlük korkudan daha güçlüdür." Ama "Müslümanları öldürmek" konulu aramalar konuşma sırasında üç katına fırlamıştı. Aslında Müslümanlara ilişkin test mahiyetinde görebileceğimiz hemen her negatif arama Obama'nın konuşması sırasında ve sonrasında ciddi bir artış gösterirken, olumlu aramalar ise azalmıştı.

Başka bir deyişle, Obama bütün doğru şeyleri söylüyor gibiydi; geleneksel medya Obama'yı birleştirici sözlerinden ötürü kutluyordu. Ama internetten gelen ve dijital hakikat serumu sunan yeni veriler aslında konuşmasının ters teptiğini gösteriyordu. Herkesin düşündüğünün aksine, Obama'nın konuşması kızgın kalabalığı sakinleştireceğine, internet verilerinin söylediğine göre, daha da öfkelendirmişti. İşe yaradığını düşündüğümüz şeyler beklediğimizin tam tersi etkiyi doğurabilir. Kimi zaman içgüdülerimizin hatasını düzeltmesi için internet verilerine ihtiyacımız vardır.

O halde, ABD'de şu an virüs gibi yayılmış olan bu nefret biçimini durdurmak için Obama ne dese iyi olurdu? Bu konuya bilahare döneceğiz. Öncelikle, ABD'deki

çok köklü bir önyargıya, aslında diğer hepsinden daha fazla öne çıkan, en yıkıcı sonuçları doğurmuş bir nefret biçimine, kitabın başında ele aldığım araştırma konusuna daha yakından bakacağız. Google arama verileriyle yaptığım çalışmada, internette nefretle ilişkili olarak bulduğum en önemli veri "*nigger*" [zenci] kelimesinin popülerliğidir.

İster tekil ister çoğul biçimiyle kullanılsın "*nigger*" kelimesi her yıl yedi milyon Amerikalının Google aramalarında yer almaktadır. (Bir kez daha belirteyim, rap şarkılarında kullanılan kelime "*nigger*" değil, "*nigga*"dır ve dolayısıyla yaygınlık hip-hop şarkı sözlerinin etkisiyle açıklanamaz.) "*Nigger* şakaları" için yapılan aramalar "*kike* [Yahudi] şakaları", "*gook* [Uzakdoğulu] şakaları", "*spic* [Latin Amerikalı] şakaları", "*chink* [Çinli] şakaları" ve "*fag* [eşcinsel] şakaları"yla ilgili aramaların toplamından on yedi kat daha fazladır.

"*Nigger*" ya da "*nigger* şakaları"yla ilgili aramalar en çok ne zaman artıyor? Afrikalı-Amerikalıların haberlere konu olduğu her zaman. Bu tür aramaların en fazla olduğu dönemlerden biri, televizyon ve gazetelerde New Orleans'ta hayatta kalmak için çırpınan siyahların görüntülerinin sıkça yer aldığı Katrina Kasırgası'nın hemen sonrasıydı. Ayrıca Obama'nın ilk seçim zaferi sırasında da tepe noktasına ulaşmıştı. Ayrıca "*nigger* şakaları" aramaları Martin Luther King Jr. Günü'nde ortalama yüzde otuz kadar artmaktadır.

Bu ırkçı aramaların korkutucu yaygınlığı, ırkçılık konusundaki mevcut anlayışları şüpheli hale getirmektedir.

Herhangi bir ırkçılık teorisinin ABD'deki bu büyük bulmacayı açıklaması lazım. Bir yandan, siyah Amerikalıların ezici çoğunluğu önyargıların mağduru olduğunu düşünüyor ve polis kontrollerinde, iş mülakatlarında ve mahkeme jürilerinin kararlarında ayrımcılığa maruz kaldığına dair mebzul miktarda kanıt sunuyor. Diğer yandan, ırkçı olduğunu kabul etmeye yanaşacak beyaz Amerikalı sayısı çok azdır.

Son dönemde siyaset bilimciler arasındaki egemen açıklama bunun büyük oranda yaygın *örtük* önyargılardan kaynaklandığıdır. Amerikalıların içinde fesatlık olmasa da, bilinçaltlarında bir taraflılık ya da önyargı varmış ve bu da siyah Amerikalılara davranışlarını etkiliyormuş. Akademisyenler bu önyargıyı test etmenin dâhiyane bir yolunu buldular. Buna örtük çağrışım testi adını veriyorlar.

Testler, çoğu insanın siyahların yüzlerini "iyi" gibi olumlu kelimelerle özdeşleştirmelerinin "berbat" gibi olumsuz kelimelerle özdeşleştirmelerinden milisaniyelerce daha uzun sürdüğünü göstermektedir. Beyazların yüzleri içinse tam tersi geçerlidir. Aradaki bu fazladan zaman, insanların örtük önyargılarının (kişinin farkında bile olamayabileceği önyargıların) kanıtıdır.

Ama Afrikalı-Amerikalıların hissedip beyazların reddettiği ayrımcılık için alternatif bir açıklama da mevcut: Saklı *aleni* ırkçılık. İnsanların epey bir farkında oldukları ama itiraf etmeyecekleri (hele hele anketlerde asla itiraf etmeyecekleri) epey yaygın bir bilinçli ırkçılık olduğunu varsayalım. Arama verilerinin tam da bunu söylediğini görüyoruz. "*Nigger* şakaları" aramanın örtük bir yanı yoktur. Ayrıca Amerikalıların "*nigger*" kelimesini "migren" ve "ekonomist "kelimesiyle aynı sıklıkta google'ladıkları bir ortamda, *aleni* ırkçılığın Afrikalı-Amerikalılar üzerinde büyük bir etkisi olmadığını düşünmek zordur. Google verilerinden önce, bu yaygın düşmanlık hakkında ikna edici bir ölçüte sahip değildik. Şimdi sahibiz. Bu nedenle açıklayıcılığını görebilecek konumdayız.

Daha önce belirtildiği üzere, Obama'nın 2008 ve 2012'deki oy oranlarının birçok bölgede düşük çıkmasını bununla açıklayabiliriz. Bu durum ayrıca, bir grup iktisatçının kısa süre önce aktardığı, siyahlar ile beyazlar arasındaki ücret farkıyla da bağıntılıdır. Başka bir deyişle, en fazla ırkçı arama yapıldığını tespit ettiğim bölgelerde siyahlara ödenen ücretler daha düşüktür. Tabii bir de Donald Trump'ın başkan adaylığı olgusu var. Giriş'te belirttiğim üzere, anketörlerin duayeni Nate Silver, Cumhuriyetçilerin 2016'daki başkan adayı seçimlerinde Trump'a verilen destekle en güçlü bağıntıya sahip coğrafi değişkeni, benim geliştirdiğim ırkçılık haritasında bulmuştu. Bu değişken "*nigger*" aramalarıydı.

Araştırmacılar son dönemde Siyahilere yönelik örtük önyargılara dair eyalet bazında bir ölçüm yaptılar; ben de bu sayede Google aramalarının ölçtüğü şekliyle aleni ırkçılığın ve örtük önyargının etkilerini karşılaştırma imkânı elde ettim. Örneğin Obama'nın iki başkanlık seçiminde de hangisinden en fazla zarar gördüğünü test ettim. Regresyon analizinden yararlanarak, Obama'nın düşük performans sergilediği yerleri tahmin etmek için bir bölgedeki ırkçı Google aramalarının çok şey açıkladığını tespit ettim. Bir bölgenin örtük çağrışım testlerindeki performansı bu tabloya çok az etkide bulunuyordu.

Kışkırtıcı olmak ve bu alanda daha fazla araştırmayı teşvik etmek amacıyla, çok çeşitli alanlardan araştırmacılar tarafından test edilmeyi bekleyen şu tahmini sizinle paylaşacağım: Bugün Afrikalı-Amerikalılara yönelik ayrımcılığın esas açıklaması, laboratuvar deneylerine katılmayı kabul eden insanların olumsuz kelimeler ile Siyahiler arasında bilinçdışı çağrışımlar kurmaları değil, milyonlarca beyaz Amerikalının "*nigger* şakaları" aramak gibi şeyler yapmaya devam etmesidir.

* * *

Siyah insanların ABD'de düzenli olarak karşılaştıkları ayrımcılığın aleni (ve belki de saklı) düşmanlıktan daha fazla beslendiği görülmektedir. Ama diğer gruplar söz

konusu olduğunda, bilinçdışı önyargıların daha temel bir etkisi olabilir. Örneğin Google aramalarından yararlanarak nüfusun başka bir kesimini hedef alan örtük önyargının kanıtlarını bulabildim: Genç kızlar.

Peki, diye sorabilirsiniz, kim kızlara önyargıyla yaklaşıyor?

Ebeveynleri.

Küçük çocukların ebeveynlerinin çocuklarının yetenekli olduğunu düşünüp heyecan duymaları pek şaşırtıcı değildir. "2 yaşındaki kızım"la başlayan Google aramalarının hepsinde, en yaygın kullanılan tamamlayıcı kelime "yetenekli mi?" oluyor. Ama bu soru oğlan çocuklarıyla kız çocukları arasında eşit şekilde dağılmıyor. Ebeveynlerin, "oğlum yetenekli mi?" diye sorma ihtimalleri "kızım yetenekli mi?" diye sorma ihtimallerinden iki buçuk kat daha fazladır. Ebeveynler zekâyla alakalı olup sesli söylemekten çekinecekleri başka tabirler kullanırken de ("oğlum dâhi mi?" gibi) benzer bir önyargı sergiliyorlar.

Ebeveynler küçük kızlar ile oğlanlar arasındaki reddedilemeyecek farklardan mı hareket ediyorlar? Belki de oğlan çocuklarının büyük laflar etme ya da yeteneğe ilişkin nesnel işaretler göstermeleri kız çocuklarına kıyasla daha muhtemeldir. Hayır! İlla böyle bir şey söylenecekse, tam tersi geçerlidir. Küçük yaştaki kız çocuklarının kelime haznesinin daha geniş olduğu ve daha karmaşık cümleler kurdukları görülmektedir. Amerikan okullarında kızların yetenek gerektiren programlarda yer alma ihtimalleri oğlan çocuklarına göre yüzde 9 daha fazladır. Tüm bunlara karşın, ebeveynler kız çocuklarındansa oğlan çocuklarını daha yetenekli görüyorlar.* Aslında zekâyla ilgili test ettiğim her arama teriminde (zekâ yokluğu dahil) ebeveynler genellikle kızlarındansa oğullarıyla ilgili arama yapıyorlardı. Ayrıca "oğlum ... geride mi?" ya da "aptal mı?" gibi aramalar, kız çocukları için yapılan benzer aramalardan daha fazladır. Ama oğlan çocukları hakkında "geride" ve "aptal" gibi olumsuz kelimelerle yapılan aramalar, "yetenekli" ya da "dâhi" gibi olumlu kelimelerle yapılanlardan çok daha az sayıdadır.

O halde, ebeveynlerin kız çocuklarına ilişkin temel kaygıları nelerdir? En başta, görünüşle ilgili her şey. Bir çocuğun kilosuyla ilgili sorunları düşünün. Ebeveynler, "kızım kilolu mu?" ifadesini, "oğlum kilolu mu?" sorusundan kabaca iki kat daha fazla google'lıyorlar. Ebeveynlerin kızlarına nasıl kilo verdirebileceklerini sorma oranları, oğulları için sorma oranlarının neredeyse iki katıdır. Tıpkı yeteneklilik meselesinde olduğu gibi, bu toplumsal cinsiyet yanlılığı da gerçeklikten beslenmi-

---

* Ebeveynlerin farklı cinsten çocuklara farklı davrandıkları hipotezini daha da sınamak adına, ebeveyn sitelerinden veri edinmeye uğraşıyorum. Bu tikel, özgül aramaları yapan ebeveynlerden çok daha fazla sayıda ebeveyni ele almak zorundayım.

yor. Kız çocuklarının yaklaşık yüzde 28'i fazla kiloluyken, oğlan çocuklarında bu oran yüzde 35'tir. Veriler oğlan çocuklarının kız çocuklarına kıyasla daha kilolu olduğunu söylese de, ebeveynler oğullarından çok, kızlarının fazla kilolarını görüyorlar ya da bu konuda endişeleniyorlar.

Keza ebeveynlerin, "kızım güzel mi?" sorusunu sorma oranları, "oğlum yakışıklı mı?" sorusunu sorma oranlarının bir buçuk katıdır. Ayrıca kızlarının çirkin olup olmadığını sorma ihtimalleri de oğullarının çirkin olup olmadığını sorma ihtimallerinin neredeyse üç katıdır. (Nasıl olup da, Google'ın bir çocuğun güzel mi çirkin mi olduğunu bilmesini bekliyorlar, orası meçhul.)

Genel olarak, ebeveynlerin oğullarıyla ilgili sorularda daha olumlu kelimeler kullandıkları görülüyor. Oğullarının "mutlu" olup olmadığını sorma eğilimleri, oğullarının "bunalımda" olup olmadığını sorma eğilimlerinden daha güçlüdür.

Liberal okurlar bu yanlı tutumun ülkenin daha muhafazakâr kesimlerine özgü olduğunu düşünebilirler, ama ben buna dair bir kanıta rastlamadım. Aslında, bu yanlı tutumlardan herhangi biriyle bir eyaletin siyasal ya da kültürel meşrebi arasında kaydadeğer bir ilişkiye rastlamadım. Nitekim bu önyargıların 2004'ten bu yana, yani Google'ın arama verilerinin ilk kez kullanıma açık hale geldiği yıldan bu yana azaldığına dair kanıt da yok. Kızlara karşı bu önyargının inanmak istediğimizden daha yaygın ve derine işlemiş olduğu görünüyor.

* * *

Cinsiyetçilik, önyargılar hakkındaki klişelerimizin çuvallayabileceği yegâne yer değildir.

Vikingmaiden88 yirmi altı yaşında. Tarihle ilgili şeyler okuyup şiir yazmaktan hoşlanıyor. İmza alıntısı Shakespeare'den. Tüm bunları profilinden ve ABD'nin internetteki en popüler nefret sitesi olan Stormfront.org'da yazdıklarından öğrendim. Ayrıca Vikingmaiden88'in çalıştığım gazetenin (*New York Times*) içeriğinden hoşlandığını da öğrendim. *Times*'taki bir haber hakkında heyecanlı bir yorum yazmıştı.

Son dönemde bu türde on binlerce Stormfront profilini analiz ettim. Bu profillerde kayıtlı üyelerin yaşadıkları yeri, doğum tarihlerini, ilgi alanlarını ve diğer bilgilerini görebiliyorsunuz.

Stormfront eski bir Ku Klux Klan lideri olan Don Black tarafından 1995'te kuruldu. En popüler "sosyal gruplar"ı "Nasyonal Sosyalistler Birliği" ve "Adolf Hitler Hayranları ve Destekçileri"dir. Quantcast'a göre, geçen yıl siteyi ayda 200 bin ila 400 bin arasında Amerikalı ziyaret etti. Southern Poverty Law Center'ın

yakın tarihli bir raporuna göre, son beş yıldaki yaklaşık yüz cinayet Stormfront'un kayıtlı üyelerince işlenmişti.

Stormfront üyeleri kimdir diye sorsalar, hayatta bilemezdim.

En azından profillerinde verdikleri doğum tarihlerine bakılırsa, genellikle genç bir kitlesi var. On dokuz yaşındakiler en fazla katılım sergileyen grubu oluşturuyor. On dokuz yaşındakiler kırk yaşındakilerden dört kat daha fazla. İnternet ve sosyal medya kullanıcıları zaten genellikle gençlerden oluşuyor, ama bu kadar genç bir kitle denemez.

Profillerde cinsiyet seçeneği yok. Ama Amerikalı kullanıcılardan rastgele bir örneklemle tüm gönderilere ve profillere eksiksizce baktım ve üyelerin çoğunun cinsiyetinin belirlenebileceğini gördüm: Stormfront üyelerinin yüzde 30 kadarının kadın olduğunu tahmin ediyorum.

En fazla üyesi olan eyaletler Montana, Alaska ve Idaho. Bu eyaletlerde ağırlıklı olarak beyazlar yaşıyor. Buradan çeşitliliğin az olduğu yerlerde büyümenin nefreti beslediği sonucu çıkar mı?

Sanırım çıkmaz. Daha ziyade, bu eyaletlerde Yahudi olmayan beyazların oranı yüksek olduğundan, Yahudilere ve beyaz olmayanlara saldıran bir grubun üye devşirebilme potansiyeli daha çoktur. Stormfront'un hedef kitlesine giren ve sonunda siteye katılanların oranı aslında daha fazla azınlığın olduğu bölgelerde daha yüksektir. Bu durum bilhassa, Stormfront'un on sekiz yaşında ve daha küçük olan ve dolayısıyla yaşadıkları yeri kendileri seçmeyen üyelerine bakıldığında doğrudur.

Bu yaş grubu arasında, azınlık nüfusunun en yüksek olduğu eyaletlerden California'nın üye oranı yüzde 25'tir ki bu, ulusal ortalamanın üstündedir.

Sitedeki en popüler sosyal gruplardan biri, "Yahudi Karşıtlığı Destekçileri"dir. Bu gruba üye olanların oranı bir eyaletin Yahudi nüfusuyla pozitif bağıntılıdır. Yahudi nüfusunun en yüksek olduğu eyalet olan New York'tan bu gruba katılanların sayısı ortalamanın üstündedir.

Dna88, 2001'de Stormfront'a katıldı. Kendisini "yakışıklı, ırksal farkındalığı olan", Jew* York City'de yaşayan otuz yaşında bir internet geliştiricisi olarak tarif ediyor. Sonraki dört ayda, "Yahudilerin İnsanlığa Karşı İşledikleri Suçlar" ve "Yahudi Kan Parası" gibi iki yüzden fazla gönderisi oldu ve insanları, jewwatch.com adında "Siyonist suçları" üzerine "akademik bir kütüphane" olduğunu iddia eden bir internet sitesine yönlendirdi.

---

* New York'taki *New* (yeni) yerine *Jew* (Yahudi) kullanarak yapılan bir kelime oyunu. (ç.n.)

Stormfront üyeleri azınlıkların farklı diller konuşup suça bulaşmalarından yakınıyorlar. Ama benim en ilginç bulduğum şey, flört piyasasındaki rekabetten yakınmalarıydı.

Kendisini "Kanada beyaz adamların ülkesi olarak kalmalıdır" sözüyle tanınan sabık Kanada başbakanına atıfla William Lyon Mackenzie King diye adlandıran bir adam, 2003'te beyaz bir kadının "yarı siyah kırma bir ucubeyle dolaştığı"nı görünce "öfke"sini "dizginlemek"te zorlanmıştı. Los Angeles'tan 41 yaşında bir kadın öğrenci olan Whitepride26, "Siyahları, Latinleri ve bazen de Asyalıları hiç sevmiyorum, özellikle de erkekler onları beyaz bir kadından daha çekici bulduklarında" diyor.

Belli siyasal gelişmelerin de rol oynadığını görmek lazım. Stormfront'un tarihinde bu zamana kadar en fazla üye artışının yaşandığı gün 5 Kasım 2008'di, yani Barack Obama'nın başkan seçildiğinin ertesi günüydü. Gelgelelim Donald Trump'ın adaylığı sırasında Stormfront'a ilgi artmamış, seçildikten hemen sonra da üyelik sayısında sadece küçük bir artış yaşanmıştı. Trump beyaz milliyetçiliği dalgasından yararlanarak iktidara geldi. Ama kendisinin bir beyaz milliyetçiliği dalgası yarattığına dair kanıt yoktur.

Obama'nın seçilmesi beyaz milliyetçisi harekette bir yükselişe yol açmıştı. Trump'ın seçilmesi buna bir yanıt gibi görünüyor.

Bu siteyle ilgili önem arz etmeyen bir şey varsa, herhalde o da ekonomidir. Aylık üyelik kayıtları ile bir eyaletin işsizlik oranı arasında hiçbir ilişki yoktu. Krizden orantısız şekilde etkilenen eyaletlerde Google'daki Stormfront aramaları göreli bir artış kaydetmemişti.

Ama belki de en ilginç ve şaşırtıcı olanı Stormfront üyelerinin konuştukları konulardı. Konuştukları kimi konular benim arkadaşlarımla konuştuğum konulara benziyordu. Belki de benim naifliğimdir ama beyaz milliyetçilerinin bizimkinden farklı bir evrende yaşadıklarını zannederdim. Uzun uzun *Game of Thrones* öven ve PlentyOfFish ve OkCupid gibi internetteki flört sitelerinin olumlu yanlarının da olduğunu söyleyen yazışmalara rastladım.

Stormfront üyelerinin benimle ve arkadaşlarımla benzer evrenlerde yaşıyor olmalarının gösterdiği kilit olgu şudur: *New York Times*'ın Stormfront üyeleri arasındaki popülerliği. *New York Times* sitesinde takılan tek Stormfront üyesi VikingMaiden88 değildir. Gazetenin internet sitesi Stormfront üyeleri arasında popülerdir. Aslında Stormfront kullanıcılarını Yahoo haber sitesini ziyaret edenlerle karşılaştırdığımda, Stormfrontçuların nytimes.com'u ziyaret etme ihtimalinin iki kat daha fazla olduğunu gördüm.

Bir nefret sitesinin üyeleri liberalliğiyle meşhur nytimes.com'a mı giriyormuş? Böyle bir şey olabilir mi? Eğer Stormfront üyelerinin önemli bir kısmı haberleri nytimes.com'dan alıyorsa, beyaz milliyetçileri hakkındaki geleneksel düşüncelerimiz yanlış demektir. Bu ayrıca, internetin nasıl işlediğine dair geleneksel görüşlerimizin de yanlış olduğu anlamına gelir.

İnternetle İlgili Hakikatler

İnternetin Amerikalıları kutuplaştırdığı, çoğu insanı sadece kendisine benzeyen kişilerin yer aldığı deliklere hapsettiği konusunda neredeyse herkes hemfikirdir. Harvard Hukuk Fakültesi'nden Cass Sunstein bu durumu şöyle tarif ediyor: "İletişim pazarımız insanların kendilerini kendi bakış açılarıyla sınırladıkları bir duruma doğru hızla ilerliyor: Liberaller çoğunlukla ya da sadece liberalleri; ılımlılar ılımlıları; muhafazakârlar muhafazakârları; Neo-Naziler de Neo-Nazileri izleyip okuyorlar."

Bu görüş oldukça anlamlı. Nihayetinde, internet bize neredeyse sınırsız sayıda imkân sunuyor ve bunların içinden istediğimiz haberi seçme şansımız var. Ne istersem onu okuyabilirim. Siz de ne isterseniz onu okuyabilirsiniz. VikingMaiden88 ne isterse onu okuyabilir. İnsanlar kendi araçlarıyla baş başa kaldıklarında, inandıkları şeyleri doğrulayacak görüşleri aramaya meylederler. Dolayısıyla internet siyasal kutuplaşmayı uç noktasına taşıyor olmalıdır, kesin!

Fakat bu görüşün sorunlu bir yanı var. Veriler bunun o kadar da doğru olmadığını söylüyor.

Bu geleneksel görüşe karşı çıkan kanıtlar, yukarıda bahsi geçen iktisatçılar Jesse Shapiro ve Matt Genzkow'un yaptıkları bir çalışmayla ortaya çıktı.

Genzkow ve Shapiro Amerika'da geniş bir kitlenin internette gezinme davranışlarına dair veriler topladılar. Elde ettikleri veri kümesinde, deneklerin –kendilerinin ifade ettikleri– ideolojileri, kendilerini daha liberal mi yoksa daha muhafazakâr mı gördükleri konusundaki fikirleri de yer alıyordu. Bu verileri internetteki siyasal kutuplaşmayı ölçmek için kullandılar.

Bunu nasıl yaptılar? İlgi çekici bir düşünce deneyiyle.

Aynı haber sitesini ziyaret ettiğini öğrendiğiniz iki Amerikalıyı rastgele seçtiğinizi varsayalım. Bunlardan birinin liberal diğerinin muhafazakâr olma olasılığı nedir? Başka bir deyişle, liberaller ile muhafazakârlar ne kadar sıklıkla aynı haber sitelerinde "karşılaşırlar"?

Bunu daha da irdelemek amacıyla, internetteki liberaller ile muhafazakârların çevrimiçi haberlerini asla aynı yerden almadıklarını varsayalım. Başka bir deyişle, liberaller sadece liberal internet sitelerini, muhafazakârlar da sadece muhafazakâr

internet sitelerini ziyaret ediyor olsun. Eğer durum buysa, belli bir haber sitesindeki iki Amerikalının karşıt siyasal görüşlere sahip olma ihtimali yüzde sıfır olacaktır. İnternette kusursuz bir *ayrışma, kutuplaşma* olacaktır. Liberaller ile muhafazakârlar asla karışıp barışmayacaktır.

Ama bir de liberaller ile muhafazakârların haberlere ulaşma biçimlerinin ve araçlarının aynı olduğunu varsayalım. Başka bir deyişle, bir liberal ile bir muhafazakârın belli bir haber sitesini ziyaret etme ihtimalleri eşit olsun. Eğer böyleyse, belli bir haber sitesindeki iki Amerikalının karşıt siyasal görüşlere sahip olma ihtimalleri kabaca yüzde 50 olacaktır. İnternette *ayrışma* ve *kutuplaşma tamamen ortadan kalkacaktır.* Liberaller ile muhafazakârlar karışıp barışacaktır.

Peki, veriler bize ne anlatıyor? Genzkow ve Shapiro'ya göre, ABD'de aynı haber sitesini ziyaret eden iki insanın farklı siyasal görüşlere sahip olma olasılığı yüzde 45 civarındadır. Başka bir deyişle, internet ayrışmanın ortadan kalkmasına, ayrışmadan çok daha yakındır. Liberaller ile muhafazakârlar internette birbirleriyle sürekli "karşılaşıyorlar".

İnternetteki ayrışmama durumunu yerli yerine oturtmanın en iyi yolu, bunu yaşamlarımızın diğer alanlarındaki ayrışmayla karşılaştırmaktır. Gentzkow ve Shapiro analizlerini çeşitli çevrimdışı etkileşimlerle tekrarlama imkânı buldular. Aynı ailenin iki ferdinin farklı siyasal görüşlere sahip olma ihtimali nedir? Veya iki komşunun? İki meslektaşın? İki arkadaşın?

TANIŞTIĞIMIZ KİŞİNİN KARŞIT SİYASAL GÖRÜŞLERE SAHİP OLMA OLASILIĞI

Bir internet haber sitesinde	%45,2
Aynı işyerinde	%41,6
Fiziksel ortamdaki komşu	%40,3
Aile ferdi	%37
Arkadaş	%34,7

Gentzkow ve Shapiro General Social Survey'in verilerinden yararlanarak, tüm bu rakamların bir internet haber sitesinde karşılaşan iki insanın farklı siyasi görüşlere sahip olma olasılığından düşük olduğunu tespit ettiler.

Başka bir deyişle, internette karşıt görüşlü birine rastlama ihtimaliniz internet dışında rastlama ihtimalinizden daha yüksektir.

İnternette ayrışma ve kutuplaşma neden daha fazla değildir? İnternette siyasal ayrışma ve kutuplaşmayı sınırlayan iki etken bulunuyor.

Birincisi, biraz şaşırtıcı gelebilir, ama internet haber sektörüne birkaç devasa site hükmediyor. İnternetin genellikle marjinallere hitap eden bir yer olduğunu düşünürüz. Gerçekten de herkese hitap eden siteler var, görüşünüz ne olursa olsun. Silah karşıtları için, silah yanlıları için, puro tüttürme hakkı savunucuları, madeni dolar eylemcileri için, anarşistler için ve beyaz milliyetçileri için, herkes için bir yer vardır. Ama bu siteler toplama vurulduğunda internetin haber trafiğinin ancak küçük bir kısmını oluşturuyor. Gerçekteyse bir tekel mevcut: 2009'da dört site (Yahoo News, AOL News, msnbc.com ve cnn.com) haberlerin yarısından çoğunu bünyesinde toplamıştı. Yahoo News, Amerika'daki en popüler haber sitesidir ve her ay 90 milyona yakın farklı kullanıcı tarafından ziyaret edilmektedir ki, bu sayı Stormfront kitlesinin 600 katı kadardır. Yahoo News gibi kitle iletişim siteleri geniş, siyasal açıdan farklı kesimlere hitap ediyor.

İnternetin o kadar da ayrışıp kutuplaşmış bir yer olmamasının ikinci nedeni güçlü siyasal görüşlere sahip insanların sırf sinirlenip tartışmak için bile olsa karşıt görüşlü siteleri ziyaret etmeleridir. Siyaset bağımlıları sadece kendilerine hitap eden sitelerle sınırlı kalmıyorlar. Sözgelimi thinkprogress.com ve moveon.org gibi aşırı liberal iki siteyi ziyaret eden birisinin sağ eğilimli foxnews.com'u ziyaret etme ihtimali ortalamanın üstündedir. Keza rushlimbaugh.com veya glennbeck.com gibi iki aşırı muhafazakâr siteyi ziyaret eden birinin daha liberal bir site olan nytimes.com'u ziyaret etme ihtimali ortalama internet kullanıcısının üstündedir.

Gentzkow ve Shapiro'nun çalışması internet tarihinde görece erken bir dönem olan 2004-2009 arasındaki verilere dayalıydı. İnternet o dönemden beri daha fazla ayrışmış olabilir mi? Sosyal medya ve özellikle de Facebook, ulaştıkları bu sonucu değiştirmiş olabilir mi? Çok açık ki, eğer arkadaşlarımız genellikle bizimle aynı siyasal görüşteyse, sosyal medyanın yükselişi aynı görüşten kişilerin sürekli bir arada bulunduğu fanusların oluşması anlamına gelecektir, öyle değil mi?

Bir kez daha, hikâyenin o kadar basit olmadığını hatırlatmalıyız. Doğru, insanların Facebook arkadaşlarının aynı siyasal görüşlere sahip olma ihtimali, olmama ihtimalinden daha yüksektir; fakat bir grup veri bilimcisi (Eytan Bakshy, Solomon Messing ve Lada Adamic) insanların Facebook'ta edindikleri bilgilerin şaşırtıcı bir kısmının karşıt görüşlü insanlardan geldiğini buldular.

Bu nasıl olabiliyor? Arkadaşlarımızla aynı siyasal görüşlere sahip değil miyiz? Orası öyle. Ama Facebook'un fiziksel dünyadaki sosyalleşmeden daha çeşitli siyasal tartışmalara yol açabilmesinin can alıcı bir nedeni bulunuyor. İnsanlar, ortalamaya

vurulduğunda, Facebook'ta gündelik hayatlarında olduğundan daha çok arkadaşa sahipler ve Facebook'un sağladığı bu zayıf bağların karşıt siyasal görüşlerden kişiler olması daha muhtemeldir.

Başka bir deyişle, Facebook bizi zayıf sosyal bağlantılara açık hale getiriyor: Liseden tanıdıklarımız, uzaktan kuzeniniz olan bir manyak, arkadaşınızın arkadaşının arkadaşının dıdısının dıdısı falan. Bunlar asla bowling oynamaya ya da mangal yapmaya yanaşmayacağınız kimselerdir. Onlarla bir akşam yemeğe gitmezsiniz. Ama Facebook'ta arkadaşsınızdır. Ve sizin aklınızın ucundan bile geçmeyecek görüşlerin yer aldığı yazıların bağlantılarını verdiklerini görürsünüz.

Özetle, internet aslında farklı siyasal görüşlerden insanları bir araya getiriyor. Ortalama bir liberal, sabahlarını liberal kocası ve liberal çocuklarıyla, öğleden sonrasını liberal mesai arkadaşlarıyla geçiriyor olabilir; işe gidip geldiği güzergâhta liberal afişlere rastlıyor olabilir; akşamını yoga sınıfından liberal yoga arkadaşlarıyla geçiriyor olabilir. Ama eve gelip cnn.com'daki birkaç muhafazakâr yoruma göz gezdirdiğinde ya da lisedeki Cumhuriyetçi tanıdığının paylaştığı bir Facebook bağlantısını gördüğünde, muhafazakârlığa maruz kalmış olur.

Brooklyn'deki gözde kahvecimde muhtemelen beyaz milliyetçilerine hiçbir zaman rastlamıyorum. Ama VikingMaiden88 ve ben *New York Times* sitesini sık sık ziyaret ediyoruz.

Çocuk İstismarı ve Kürtaj Hakkındaki Hakikatler

İnternet bize sadece rahatsız edici tutumlar hakkında değil, rahatsız edici davranışlar hakkında da içgörüler sunabilir. Gerçekten de Google verilerine başvurarak tüm alışıldık kaynakların gözden kaçırdığı krizlerden haberdar olabiliriz. Nihayetinde insanların başı belaya girdiğinde yüzlerini döndükleri yer Google oluyor.

2007 sonlarında başlayan ekonomik kriz sırasında yaşanan çocuk istismarı vakalarını ele alalım.

2007'nin sonlarında büyük ekonomik gerileme başladığında, birçok uzman doğal olarak bunun çocuklar üzerinde doğurabileceği etkileri düşünüp endişelenmişti. Sonuçta, birçok anne-baba strese girip bunalım yaşayacaktır ve bunlar da çocuklara kötü muamele için ciddi risk faktörleridir. Çocuk istismarı tavan yapabilirdi.

Sonra resmî veriler açıklandı; endişenin yersiz olduğu görülüyordu. Çocuk koruma hizmeti veren kuruluşlar istismar vakalarının azaldığını söylüyordu. Ayrıca bu düşüşler ekonomik gerilemenin en fazla etkilendiği eyaletlerde daha fazlaydı. Pennsylvania Üniversitesi'nden çocuk sağlığı uzmanı olan Richard Gelles 2011'de Associated Press'e, "Bittik, mahvolduk tahminleri doğru çıkmadı," diyordu. Evet,

sezgilere aykırı olsa da, çocuk istismarı ekonomik kriz döneminde gerilemiş görünüyordu.

Ama bu kadar fazla yetişkinin işsiz kalıp aşırı baskı altında olduğu bir dönemde çocuk istismarı gerçekten gerilemiş olabilir miydi? Buna inanmakta zorlanıyordum. Bu yüzden Google verilerine döndüm.

Bazı çocuklar Google'da trajik, insanın içini burkan aramalar yapıyorlardı: "Annem beni dövdü" ya da "babam bana vurdu" gibi. Ve bu aramalar o dönemde yaşananlara dair farklı (ve üzücü) bir tablo sunuyor. Bunun gibi aramaların sayısı ekonomik kriz sırasında tepe noktasına ulaşmıştı ve işsizlik oranına paralel olarak yükseliyordu.

Bence olup bitenleri şöyle özetleyebiliriz: Çocuk istismarı değil, çocuk istismarı bildirimleri azalmıştı. Nihayetinde, çocuk istismarı vakalarının zaten küçük bir kısmı yetkililere bildiriliyor. Bir ekonomik kriz sırasındaysa, genelde çocuk istismarı vakalarını bildiren kişilerin çoğu (öğretmenler ve polisler gibi) ya da bu vakalarla ilgilenenler (çocuk koruma hizmetlerinde çalışanlar) ya fazla mesai yapıyorlar ya da işsiz oluyorlar.

Ekonomik kriz sırasında potansiyel vakaları bildirmeye çalışırken uzun süre beklemek zorunda kalıp vazgeçen birçok insan vardı.

Gerçekten de çocuk istismarının ekonomik kriz sırasında arttığına dair pek çok kanıt vardır (bu kez Google'dan değil). Bir çocuk istismar ya da ihmalden ötürü öldüğünde, bunun bildirilip rapor edilmesi gerekir. Bu tür ölümler nadir olsa da, ekonomik krizin en çok etkilediği eyaletlerde gerçekten de artmıştı.

Bir de Google'dan gelen ve krizden en fazla etkilenen bölgelerde istismarın arttığını ele veren kanıtlar vardı. Ekonomik krizden önceki oranları ve ulusal eğilimleri dikkate aldığımızda, görece en fazla etkilenen eyaletlerde çocuk istismarı ve ihmaliyle ilgili arama oranları artmıştı. İşsizlik oranındaki her binde birlik artış, "çocuk istismarı" ya da "çocuğun ihmal edilmesi arama oranlarında yüzde 3'lük bir artış getiriyordu. Muhtemelen, bu eyaletlerde düşüş en fazla olduğuna göre, bu insanların çoğu, vakaları hiçbir zaman bildirmemiş ya da bildirememişlerdi.

Mağdur çocuklar tarafından yapılan aramalar artmıştı. Çocuk ölüm oranları tavan yapmıştı. En fazla etkilenen eyaletlerde istismardan şüphelenen insanların yaptıkları aramalar da artmıştı. Ama vaka bildirimleri azalmıştı. Görünen o ki, ekonomik krizden ötürü, daha fazla çocuk Google'a ebeveynlerinin onları dövdüğünü ya da onlara tokat attığını söylüyor ve birçok insan da yaşadığı çevrede çocukların istimara maruz kaldığından şüpheleniyordu. Ama fazla iş yükü altında ezilen kuruluşlar giderek daha az vakayla ilgilenebiliyorlardı.

Kanaatimce, geleneksel ölçümler farklı sonuçlara varsa da, ekonomik krizin çocuk istismarını daha da kötüye götürdüğünü söylemek mümkündür.

* * *

Bir grubun uğradığı kötü muamelenin ya da çektiği acının kayıtlarda yer almadığından şüphelendiğimde, Google verilerine başvuruyorum. Bu yeni verilerin ve bunları nasıl yorumlayacağınızı bilmenin potansiyel yararlarından biri, normalde yetkililer tarafından görmezden gelinebilecek olan kişilere yardım eli uzatma olanağı sunmasıdır.

Sözgelimi kısa süre önce Anayasa Mahkemesi yasaların kürtaj yaptırmayı daha da zorlaştırması meselesini ele alırken, anket verilerine döndüm. Bu yasadan etkilenen kadınların gebeliklerini sonlandırmak için kayıt dışı yolları araştırabileceklerinden şüphelenmiştim. Haklıydım. Kürtajı sınırlandırıcı yasalar geçiren eyaletlerde bu tür aramaların sayısı çok daha yüksekti.

Arama verileri bu noktada hem yararlıdır hem de başa beladır.

2015'te ABD'de "kendi kendine kürtaj nasıl yapılır?" diye 700 binden fazla Google araması yapılmıştı. Karşılaştırma yaparsak, kürtaj yapan klinikler için yapılan arama 3,4 milyon kadardı. Bu demektir ki, kürtaj yaptırmayı düşünen kadınların kayda değer bir kısmı bunu kendi kendine yapmayı aklından geçirmişti.

Kadınlar kürtaj haplarını gayriresmî kanallardan nasıl elde edebileceklerini yaklaşık 160 bin kez aramışlardı: "İnternetten kürtaj hapı alma" ve "ücretsiz kürtaj hapları". Google'da maydanoz gibi şifalı otları ya da C vitamini yoluyla kürtajın nasıl olduğunu aramışlardı. Elbise askısıyla yapılan düşükler için yönlendirici bilgi arayan 4 bin kişi vardı ki, bunların 1.300 kadarı "elbise askısında düşük nasıl yapılır?" diye tam olarak ne aradığını bilerek arama yapmıştı. Ayrıca rahmini çamaşır suyuyla yıkayarak ya da midesini yumruklayarak nasıl çocuk düşürüleceğini arayan birkaç yüz kişi vardı.

Kendi kendine kürtaja olan ilgiyi tetikleyen nedir? Google aramalarının coğrafyası ve zamanlaması muhtemel suçluya işaret ediyor: Resmî yollarla kürtaj yaptırmak zorlaştığında, kadınlar kayıt dışı yaklaşımlara yöneliyorlar.

Kendi kendine kürtaj için arama oranları 2004-2007 arasında hep aynı çizgide seyretmişti. Finans krizi ve akabinde gelen resesyonla örtüşecek şekilde 2008 sonlarındaysa artmaya başlamıştı. 2011'de yüzde 40'lık bir artışla büyük bir yükseliş kaydetmişti. Bir doğum hakları kuruluşu olan Guttmacher Enstitüsü ülkede kürtaj hakkına son dönemdeki saldırıların başlangıcı olarak 2011'i işaret ediyor; o yıl 92 eyalet, kürtaja erişimi kısıtlayan yasaları yürürlüğe sokmuştu. Kürtaj hakkına böyle

bir saldırının olmadığı Kanada'yla kıyaslarsak, o dönemde kendi kendine kürtaj aramalarında benzer bir artış yaşanmamıştı.

Google'da kendi kendine kürtaj aramalarının en fazla artığı eyalet Mississippi'ydi: Şu an kabaca üç milyon insanın yaşadığı ama sadece tek bir kürtaj kliniğinin bulunduğu eyalet. Kendi kendine kürtaj aramalarının oranının en yüksek olduğu on eyaletten sekizi Guttmacher Enstitüsü tarafından kürtaja düşman ya da aşırı düşman olarak görülüyor. Oysa kendi kendine kürtajla ilgili arama oranlarının en düşük olduğu on eyaletten hiçbiri bu kategoriye girmiyor.

Elbette Google aramalarından kaç kadının kendi kendine kürtaj olmayı başardığını öğrenemeyiz, ama kanıtlar sayının kayda değer olabileceğini gösteriyor. Bunu aydınlatmanın yollarından biri kürtaj ve ölüm verilerini karşılaştırmaktır

2011'de, yani eyalet düzeyinde kürtaj verilerinin eksiksiz olduğu son yılda, kürtaj kliniklerinin çok az olduğu eyaletlerde yaşayan kadınlar çok daha az yasal kürtaj yaptırmıştı.

Kişi başına en fazla kürtaj kliniğinin düştüğü on eyaleti (bu listede New York ve California da vardır), kişi başına düşen kürtaj kliniğinin en az olduğu on eyaletle (bu listede Mississippi ve Oklahoma da vardır) karşılaştıralım. Kürtaj kliniği sayısının en az olduğu eyaletlerde yaşayan kadınların yasal kürtaj yaptırma oranı yüzde 54 daha azdı; on beş ile kırk dört yaş arasındaki her bin kadında on bir kürtajlık bir fark demektir bu. Kürtaj kliniği sayısının en az olduğu eyaletlerde yaşayan kadınlar daha fazla canlı doğum gerçekleştirmişti. Ne var ki aradaki fark düşük kürtaj sayısını açıklamak için yeterli değildi. Çocuk doğurma yaşına gelmiş her bin kadın başına fazladan altı canlı doğum söz konusuydu.

Başka bir deyişle, kürtaj yaptırmanın zor olduğu bazı yerlerde eksik gebelikler var gibi görünüyordu. Resmî kaynaklar kürtaj yaptırmanın zor olduğu eyaletlerde her bin kadın başına düşen beş eksik doğumla ilgili bilgi vermiyor.

Google ise bazı önemli ipuçları sunuyor.

Devlet verilerine öyle körü körüne güvenemeyiz. Hükümet çocuk istismarının ya da düşüklerin düştüğünü söyleyebilir ve siyasetçiler bu başarıyla övünebilirler. Ama gördüğümüz sonuçlar veri toplama yöntemlerindeki kusurların bir ürünü olabilir. Hakikat farklı ve bazen de çok daha kötü olabilir.

Facebook Arkadaşlarınız Hakkındaki Hakikatler

Bu kitap genel olarak Büyük Veri'yi konu alıyor. Ama bu bölümde çoğunlukla Google aramalarını vurguladım, Google aramalarının gördüğümüzü düşündüğümüzden çok farklı olan bir saklı dünyayı ele verdiğini savundum. O halde, başka

Büyük Veri kaynakları da dijital hakikat serumları mıdır? Gerçek şu ki, Facebook gibi birçok Büyük Veri kaynağı çoğu zaman dijital hakikat serumunun zıddıdır.

Anketlerde olduğu gibi sosyal medyada da sizi hakikati söylemeye teşvik eden hiçbir şey yoktur. Sosyal medyada, anketlerin de çok ötesinde, kendinizi daha iyi göstermeniz için pek çok neden vardır. Sonuçta gündelik hayattaki mevcudiyetiniz anonim değildir. Bir kitleye oynuyorsunuz ve arkadaşlarınıza, aile fertlerinize, meslektaşlarınıza ve yabancılara kim olduğunuzu anlatıyorsunuz.

Sosyal medyadan elde edilen verilerin ne kadar taraflı olabileceğini görmek için, saygın ve elit bir aylık dergi olan *Atlantic* ile çoğu zaman sansasyon peşinde koşan *National Enquirer*'ın göreli popülerliğini ele alalım. İki yayının tirajı da birkaç yüz bin satış rakamıyla benzer ortalamalara sahiptir. (*National Enquirer* haftalık olduğundan toplamda daha fazla satış yapmaktadır.) Ayrıca iki derginin de Google aramaları benzer sayılardadır.

Gelgelelim Facebook'ta *Atlantic*'i beğenen ya da profilinde *Atlantic* makalelerini paylaşan insan sayısı kabaca 1,5 milyonken, *Enquirer* için aynı sayı 50 bin kadardır.

FARKLI KAYNAKLARA KIYASLA
ATLANTIC İLE *NATIONAL ENQUIRER*'IN POPÜLERLİĞİ

Tiraj	Kabaca her 1 *National Enquirer*'a karşı 1 *Atlantic*
Google Aramaları	her 1 *National Enquirer*'a karşı 1 *Atlantic*
Facebook beğenileri	her 1 *National Enquirer*'a karşı 27 *Atlantic*

Dergilerin popülerliğini ölçmek için tiraj verileri temel hakikattir. Google verileri bununla hemen hemen örtüşüyor. Facebook verileriyse boyalı basının ağır bir yenilgi aldığına işaret ediyor ve insanların gerçekte neleri beğendiklerinin belirlenmesinde en isabetsiz veriyi ortaya koyuyor.

Okuma tercihlerinde neyse, hayatta da öyle. Facebook'ta gerçekte kim olduğumuzu değil, kültürlü kimliğimizi gösteririz. Bu kitapta, hatta bu bölümde Facebook verilerinden yararlanıyorum, ama bu gerçeği asla unutmuyorum.

* * *

Sosyal medyanın neyi gözden kaçırdığını daha iyi anlamak için, kısa süreliğine pornografiye dönelim. Bir defa, internetin pislik kaynadığı şeklindeki yaygın kanıyı ele almamız gerekiyor. Bu doğru değildir. İnternetteki içeriğin çoğunluğu

pornografik değildir. Sözgelimi en fazla ziyaret edilen on internet sitesinden biri bile porno sitesi değildir. Dolayısıyla pornonun popülerliği muazzam olmakla birlikte abartılmamalıdır.

Ama bunu söyledikten sonra, pornografiyi nasıl beğenip paylaştığımıza yakından bakmak Facebook, Instagram ve Twitter'ın internette gerçekten popüler olana dair ancak sınırlı bir pencere açtığını ortaya koyacaktır. İnternette müthiş popüler olan, ama sosyal mevcudiyeti çok az olan geniş bir altküme vardır.

Gelmiş geçmiş en popüler video, dünyadaki trendleri yakından takip eden Korelilerin hicvedildiği PSY'ın "Gangnam Style" adlı gülünç pop müzik klibidir. 2012'de ilk yayınlandığı günden bu yana sadece YouTube'da 2,3 milyar kez izlenmiştir. Dahası hangi sitede olursanız olun popülerliğinin değişmediğini görebilirsiniz. Farklı sosyal medya platformlarında on milyonlarca kez paylaşılmıştır.

Gelmiş geçmiş en popüler porno videosuysa herhalde "Great Body, Great Sex, Great Blowjob"dır. 80 milyondan fazla izleyiciye ulaşmıştır. Başka bir deyişle, her otuz "Gangnam Style" görüntülenmesine, en azından bir "Great Body, Great Sex, Great Blowjob" görüntülenmesi denk düşüyor. Eğer sosyal medya insanların izlediği videolara dair doğru bir tablo sunuyor olsaydı, "Great Body, Great Sex, Great Blowjob"ın milyonlarca kez paylaşılmış olması gerekirdi. Ama bu video sosyal medyada yüz kez bile paylaşılmadı; paylaşanlar ise ortalama kullanıcılar değil yalnızca porno yıldızlarıydı. Belli ki insanlar bu videoya olan ilgilerini arkadaşlarına duyurma ihtiyacı hissetmiyorlar.

Facebook dijital alanda "arkadaşlarıma hayatımın ne kadar güzel olduğunu gösterip böbürleneyim serumu"dur. Facebook dünyasında ortalama bir yetişkinin mutlu bir evliliği vardır, tatilini Karayipler'de yapar ve *Atlantic* okur. Gerçek dünyadaysa pek çok insan kızgındır, süpermarketlerde kasa kuyruklarında bekler, *National Enquirer*'a göz gezdirir, eşinin aramalarını duymaz ve zaten aylardır da sevişmemişlerdir. Facebook dünyasında aile yaşamı kusursuz görünür. Gerçek dünyadaysa aile hayatı tam bir keşmekeştir. Bazen o kadar karman çorman bir hal alır ki, bazı insanlar –bunların sayıları az olsa da– çocuk yaptıklarına pişman olurlar. Facebook dünyasına inanacak olursak, cumartesi geceleri her genç yetişkin müthiş bir partide eğleniyordur. Gerçek dünyadaysa evde tek başına, Netflix'ten

DİJİTAL HAKİKAT	DİJİTAL YALANLAR
• Aramalar • Görüntülemeler • Tıklamalar • Kaydırmalar	• Sosyal medya paylaşımları • Sosyal medya beğenileri • Randevulaşma profilleri

dizi kovalıyordur. Facebook dünyasında bir kadın, erkek arkadaşıyla yaptıkları geziden 26 tane mutlu resim paylaşır. Gerçek hayattaysa, bunu paylaştıktan hemen sonra, Google'a "erkek arkadaşım benimle seks yapmıyor" yazar. Ve belki de hemen hemen aynı sırada, erkek arkadaşı da "Great Body, Great Sex, Great Blowjob"ı izliyordur.

Müşterileriniz Hakkındaki Hakikatler

Facebook 5 Eylül 2006'da sabah erken saatlerde, ana sayfasında çok büyük bir güncelleme yaptı. Facebook'un önceki versiyonları kullanıcılarına, arkadaşlarının ne yaptıklarını sadece profillerine tıklayarak öğrenme izni veriyordu. Büyük bir başarı olarak görülen Facebook'un o dönemde 9,4 milyon kullanıcısı vardı.

Ama aylar süren sıkı bir mesainin ardından, mühendisler "İçerik Akışı" adını verdikleri ve kullanıcılara tüm arkadaşlarının faaliyetleri hakkında güncellemeler sunan bir şey yarattılar.

Kullanıcıların ilk tepkisi İçerik Akışı'ndan nefret ettiklerini bildirmek oldu. Northwestern Üniversitesi'nde okuyan Ben Parr, "Öğrenciler Facebook İçerik Akışı'na Karşı" grubunu kurdu. "İçerik Akışı'nın çok ürpertici, adeta dikizleyici ve kaldırılması gereken bir özellik olduğunu" söylüyordu. Birkaç gün içinde grubun üye sayısı 700 bine çıkmıştı ve hepsi de Parr'la aynı duygulara sahipti. Michigan Üniversitesi'nden bir üçüncü sınıf öğrencisi, *Michigan Daily*'ye yaptığı açıklamada "Yeni Facebook gerçekten oha dedirtti. İnsana kendisini dikizci gibi hissettiriyor," demişti.

David Kirkpatrick, Facebook'ın resmî tarihi olan *The Facebook Effect: The Inside Story of the Company That Is Connecting the World* adlı kitabında bu hikayeyi anlatır. İçerik Akışı'nı "Facebook'un karşılaştığı en büyük kriz" diye nitelendirir. Ama Kirkpatrick hızla büyüyen bu şirketin kurucularından biri ve aynı zamanda başkanı olan Mark Zuckerberg'le görüştüğünde, Facebook CEO'su hiç istifini bozmaz.

Bunun sebebi nedir? Zuckerberg'in dijital hakikat serumuna erişimi vardı: İnsanların Facebook ziyaretlerine ve tıklamalarına dair sayılar elindeydi. Kirkpatrick'in sözleriyle,

> Zuckerberg insanların gruplarda ne söylediklerinden bağımsız olarak, aslında İçerik Akışı'nı sevdiklerini biliyordu. Bunu kanıtlayacak verileri vardı. İnsanlar İçerik Akışı geldikten sonra Facebook'ta daha fazla zaman geçirmeye başlamışlardı. Ve orada daha fazla, çok daha fazla şey yapıyorlardı. Ağustos'ta kullanıcılar Facebook'ta 12 milyar sayfayı ziyaret etmişlerdi. Oysa İçerik Akışı'yla birlikte Ekim ayına gelindiğinde, bu sayı 22 milyara çıkmıştı.

Zuckerberg'in elinin altındaki yegâne kanıt bu değildi. İçerik Akışı karşıtları grubunun hızla büyümesi bile İçerik Akışı'nın gücünün kanıtıydı. Grup, tam da birçok insan arkadaşlarının katıldığını duyduğu için bu kadar hızlı büyümüştü ve bu bilgiyi İçerik Akışı sayesinde öğrenmişlerdi.

Başka bir deyişle, insanlar Facebook'ta arkadaşlarının hayatlarının her ayrıntısını görmekten memnun olmadıklarına dair kopartılan yaygaraya katılmalarına karşın, arkadaşlarının hayatlarının tüm ayrıntılarını görmek üzere Facebook'a giriyorlardı. İçerik Akışı kaldırılmadı. Facebook'un şu anda günde bir milyardan fazla aktif kullanıcısı var.

Facebook'un ilk yatırımcılarından Peter Thiel *Zero to One* adlı kitabında, büyük işletmelerin sırlar üzerine kurulu olduğunu söyler: Bu sırlar ya kendilerinin içyüzü hakkında ya da insanlar hakkındadır. Üçüncü Bölüm'de ele alındığı üzere, Jeff Seder sol karıncık büyüklüğünün atların performansının bir göstergesi olduğuna dair doğal sırrı keşfetmişti. Google ise bağlantılardaki bilginin ne kadar güçlü olabileceğine dair doğal sırra erişmişti.

Thiel "insanlar hakkındaki sırlar"ı "insanların kendileri hakkında bilmedikleri şeyler ya da başkalarının bilmesini istemedikleri için sakladıkları şeyler" diye tanımlar. Başka bir deyişle, bu tür işletmeler insanların yalanları üzerine kuruludur.

Facebook'un, Zuckerberg'in Harvard'dayken insanlar hakkında öğrendiği nahoş bir sır üzerine kurulu olduğunu savunabilirsiniz. Zuckerberg ikinci sınıfın başlarında öğrencilere yönelik Facemash adında bir internet sitesi kurmuştu. "Am I Hot or Not?" [Çekici miyim, değil miyim?] adında bir siteyi model alan Facemash, iki Harvard öğrencisinin resimlerini gösterip, kimin daha güzel göründüğüne diğer öğrencilerin karar vermesini istiyordu.

Bu ikinci sınıf öğrencisinin sitesi infial yarattı. Genç Zuckerberg, *Harvard Crimson*'ın başyazısında, insanların "en kötü yanına oynamak"la suçlandı. Latin Amerika kökenli ve Afrikalı-Amerikalı gruplar tarafından cinsiyetçilikle ve ırkçılıkla itham edildi. Ama Harvard yöneticileri Zuckerberg'in internet erişimini kesene kadar (site kurulduktan birkaç saat sonra), siteyi 450 kişi ziyaret etmiş ve farklı görüntülere toplamda 22 bin oy vermişti. Zuckerberg önemli bir sırrı öğrenmişti: İnsanlar çok sinirlendiklerini iddia edebilirler, bir şeyin hiç hoş olmadığını haykırabilirler, ama yine de tıklarlar.

Gerçi bir şey daha öğrenmişti: Tüm o ciddiyet, sorumluluk ve başkalarının mahremiyetine saygı nutuklarına karşın, insanlar, hatta Harvard öğrencileri bile insanların görünüşlerini yargılamaya büyük ilgi duyuyorlardı. Görüntülemeler ve oylar bunu söylüyordu. Bilahare –Facemash çok ihtilaf yarattığından– insanların

İNSANLARIN SİZE SÖYLEDİKLERİNE KULAK ASMAMANIN BÜYÜK DEĞERİ

İNSANLARIN SÖYLEDİKLERİ	GERÇEKLİK	FİİLEN GERÇEKLEŞEN...
Arkadaşlarını stalk'lamak istemezler.	Arkadaşlarıyla temasta olmak ve onları yargılamak en çok istedikleri şeylerdendir.	Facebook'un kurucularından Mark Zuckerberg'ün serveti 55,2 milyar dolardır.
İnsanlar işçilerin köle gibi çalıştırıldığı fabrikalarda üretilen ürünleri almak istemezler.	Güzel, "fiyatı makul" olan ürünleri alırlar.	Nike'ın kurucularından Phil Knight'ın serveti 25,4 milyar dolardır.
Sabahları haberleri dinlemek isterler.	Sabahları, cücelerin porno yıldızlarıyla seks yaptığını duymak isterler.	Howard Stern'in serveti 500 milyon dolardır.
Esaret, bağlama, tahakküm ve sadomazoşizm hakkında bir şeyler okumak istemezler.	Genç bir üniversite öğrencisi ile büyük bir işadamı arasındaki BDSM'yi okumak isterler.	*Grinin Elli Tonu* ABD'de 125 milyon satmıştır.
Siyasetçilerin kendi siyasi tutumlarını açık ve net ortaya koymasını isterler.	Siyasetçilerin konuşmalarını ayrıntıya boğmayıp, sert ve kendinden emin görünmesini isterler.	Donald Trump

başkaları hakkında bir şekilde bildikleri yüzeysel gerçeklere ne kadar ilgi duyabileceklerine dair bu bilgiyi aldı ve kuşağının en başarılı şirketinin devindirici gücü yaptı.

Netflix de internet hayatına gözlerini ilk açtığı dönemde benzer bir ders almıştı: İnsanların size söylediklerine değil, yaptıklarına güvenin.

Şirket ilk başta kullanıcılarının gelecekte izlemek istedikleri ama şu anda zaman bulamadıkları filmlerin bir listesini yapmalarına izin vermişti. Bu yolla, daha uygun bir zaman bulduklarında, Netflix onlara bu filmleri hatırlatabiliyordu.

Gelgelelim Netflix verilerde garip bir şey fark etti. Kullanıcılar listelerine bir sürü film alıyorlardı. Ama günler sonra, listedeki filmler hatırlatıldığında, nadiren tıklıyorlardı.

Sorun neydi? Kullanıcılara birkaç gün içinde hangi filmi seyretmeyi planladıklarını sorun, listeyi sanat filmleriyle, entelektüel filmlerle, mesela siyah-beyaz çekilmiş İkinci Dünya Savaşı filmleriyle doldururlar. Ama birkaç gün geçtikten sonra, genellikle izlemek istedikleri filmler hangisiyse yine onları izlemek isterler: Hafif komediler ya da aşk filmleri. İnsanlar kendilerine istikrarlı şekilde yalan söylerler.

Netflix bu uçurumu fark edince, insanlara ileride ne izlemek istediklerini sormayı bıraktı ve benzer müşterilere ait milyonlarca tıklamaya ve görüntülemeye dayalı bir model inşa etmeye başladı. Şirket, kullanıcılarını neden hoşlandıklarına dair söyledikleri temelinde değil, neyi izleyebileceklerine dair verilerin söylediklerini temel alan film listeleri önererek karşılamaya başladı. Sonuç: Müşteriler Netflix'i daha sık ziyaret ediyor ve daha fazla film seyrediyorlardı.

"Algoritmalar sizi sizden daha iyi tanır," diyor bir dönem Netflix'in veri bilimcisi olan Xavier Amatriain.

Hakikatle Başa Çıkabiliyor muyuz?

Bu bölümün belli kısımlarını iç karartıcı bulabilirsiniz. Dijital hakikat serumu insanları görünüşlerine göre yargılamaya olan ilginin değişmez olduğunu; milyonlarca gey erkeğin cinsel kimliğini açıklamadan yaşamayı sürdürdüğünü; kadınların kayda değer bir kısmının tecavüz fantezisi kurduğunu; Afrikalı-Amerikalılara karşı yaygın bir nefret olduğunu; gizli bir çocuk istismarı ve kendi kendine kürtaj krizi olduğunu ve ABD başkanı hoşgörü çağrısında bulunduğunda daha da beter bir hal alan şiddetli bir İslamofobik öfkenin patlak verdiğini göstermiştir. Pek de yüzümüzü güldürecek şeyler değil. Araştırmam hakkında ne zaman bir konuşma yapsam, insanlar gelip, "Seth, hepsi de çok ilginç şeyler. Ama acayip moral bozucu," diyorlar.

Bu verilerin bazılarında bir karamsarlık olduğunu reddetmiyorum. İnsanlar bize işitmek istediğimizi düşündükleri şeyleri tutarlı şekilde anlatsalar, genellikle hakikatten daha rahatlatıcı olan şeyler duyarız. Dijital hakikat serumu, ortalamaya vurulduğunda, bize dünyanın düşündüğümüzden daha kötü olduğunu gösterecektir.

Bunu bilmemize gerek var mı? Google aramaları, porno verileri ve kimin neye tıkladığı hakkında bilgi sahibi olmak, "Bu harika bir şey. Gerçekte kim olduğumuzu böylece görüyoruz," diye düşünmenize yol açmayabilir. Bunun yerine, "Bu korkunç bir şey. Gerçekte kim olduğumuzu böylece görüyoruz," diye düşünebiliriz.

Ama hakikat işimize yarar, üstelik sadece Mark Zuckerberg'ün ya da tık almaya veya müşteri çekmeye çalışan diğerlerinin değil. Bu bilgi hayatlarımızı en az üç açıdan geliştirebilir.

Birincisi, güvensizlikleriniz ve utandırıcı davranışlarınızda yalnız olmadığınızı bilmek rahatlatıcı olabilir. Başkalarının kendi bedenleri hakkında duydukları güvensizliği bilmek hoş olabilir. Tüm dünyanın deli danalar gibi sevişmediğini bilmek herhalde birçok insan için –özellikle de fazla seks yapmayanlar için– hoş olacaktır. Ayrıca takımın oyun kurucusuna âşık olan Mississippi'deki bir erkek lise öğrencisi için, etrafındaki açıktan gey erkeklerin sayısı az olmasına karşın, başka birçok insanın da aynı tür hisler beslediğini bilmek değerli olabilir.

Google aramalarının yalnız olmadığınızı göstermeye yardımcı olabileceği bir alan daha var ki, henüz bunu anlatmadım. Gençken bir öğretmeniniz size, "Eğer sorun varsa parmak kaldırıp sor çünkü sen anlamadıysan başkaları da anlamamıştır," demiş olabilir. Eğer siz de benim gibiyseniz, öğretmeninizin tavsiyesine uymayıp, ağzınızı açmaya korkarak çıt çıkarmadan oturmuşsunuzdur. Sorularınızın çok aptalca olduğunu, diğerlerinin sorularınınsa çok derin olduğunu düşünmüşsünüzdür. Anonim kalan Google verileri öğretmenlerimizin ne kadar haklı olduğunu bir çırpıda söyleyebilir. Başkalarının aklında da derin denemeyecek bir sürü temel soru işareti vardır.

Amerikalıların Obama'nın 2014 tarihli Birliğin Durumu konuşmasıyla ilgili en fazla sordukları soruları ele alalım. (Kitabın sonundaki renkli fotoğrafa bakınız.)

TEK MERAK EDEN SİZ DEĞİLSİNİZ:
BİRLİĞİN DURUMU KONUŞMASI SIRASINDA EN FAZLA GOOGLE'LANAN SORULAR

Obama kaç yaşında?

Biden'in yanında oturan kim?

Boehner niye yeşil kravat takmış?

Boehner niye turuncu?

Şimdi, bu soruları okuyunca, demokrasimizin ne kadar kötü durumda olduğunu düşünebilirsiniz. Başkan Obama'nın konuşmasının içeriği yerine birinin kravatıyla ya da derisinin tonuyla ilgilenmek pek hayra alamet değildir. O dönem Temsilciler Meclisi Başkanı olan John Boehner'in kim olduğunu bilmemek de siyasal bağlılığımız konusunda çok şey anlatmaz.

Ben bunun yerine öğretmenlerimizin bilgeliğini gösteren bu tür sorulara kafa yormak istiyorum. Bunlar kulağa çok aptalca geldiği için insanların genellikle sormadığı türde sorulardır. Ama birçok insan bu tür sorular soruyor ve google'lıyor.

Aslında Büyük Veri'nin kişisel gelişim kitaplarında sık sık yer verilen şu meşhur söze, 21. yüzyıl güncellemesi yapabileceğini düşünüyorum: "Asla kendi içini başkalarının dışıyla karşılaştırma."

Büyük Veri güncellemesi şöyle olabilir: "Asla Google aramalarını başka birinin sosyal medya paylaşımlarıyla kıyaslama."

Örneğin insanların herkese açık sosyal medyada ve anonim kalan aramalarda kocalarını nasıl tasvir ettiklerini karşılaştıralım.

İNSANLAR KOCALARINI EN ÇOK NASIL TASVİR EDİYORLAR?

SOSYAL MEDYA PAYLAŞIMLARI	ARAMALAR
en iyi	gey
en iyi dostum	hödük
akıl almaz	akıl almaz
en müthiş	sinir bozucu
çok ince	cimri

Başkalarının sosyal medya paylaşımlarını görüp aramalarını görmediğimizden, kaç kadının kocalarını düzenli olarak "en iyi", "en müthiş" ve "çok ince" diye gördüğünü genellikle abartırız.* Kaç kadının kocasını "hödük", "sinir bozucu" ve "cimri" diye gördüğünü genellikle hafifseriz. Anonim verileri analiz ederek, evliliği ve hayatı zor bulma konusunda yalnız olmadığımızı anlayabiliriz. Aramalarımızı başkalarının sosyal medya paylaşımlarıyla karşılaştırmayı bırakabiliriz.

Dijital hakikat serumunun ikinci yararlı yanı ıstırap çeken insanları görmemizi sağlamasıdır. İnsan Hakları Kampanyası belli eyaletlerdeki gey erkekleri kimliklerini gizlemekten vazgeçirme konusunda verilecek eğitimlerde benden de yardım istedi. Anonim kalan Google arama verilerinden yararlanarak kaynaklarını nereye yönlendirmenin en doğru adım olacağını öğrenmek istiyorlardı. Benzer şekilde, çocuk esirgeme kurumları da ülkenin hangi bölgelerinde kayıtlarda yer aldığından çok daha fazla çocuk istismarı olduğunu öğrenmek için benimle temasa geçmişti.

Şöyle şaşırtıcı bir konuda da kapım çalındı: Vajina kokuları. *New York Times*'ta bu konuda ironik bir yazı yazmıştım. Beni de başkalarını da kıkırdatan bir yazı olmuştu.

---

* Twitter verilerini analiz ettim. Emma Pierson'a bunu indirmeme yardımcı olduğu için müteşekkirim. Sosyal medyada yaygın olsa da aramalarda pek anlam ifade etmeyecek olan, "kocam şu anda ne yapıyor?" temalı tasvirleri dahil etmedim. Gerçi bu tasvirler bile olumluya çalıyor. Sosyal medyada birinin kocasının o anda ne yaptığını tasvir etmek için en çok kullandığı sözler şunlar: "Çalışıyor" ve "yemek yapıyor".

Fakat sonradan, bu konuda arama yapıldığında ulaşılan forum sayfalarından bazılarını incelediğimde, vajina kokuları konusundaki endişelerinden ötürü hayatlarının mahvolduğunu düşünen genç kızların bu yazıyı çok kez paylaştığını gördüm. Şaka değil. Seks eğitimi uzmanları benimle temasa geçip, genç kızlar arasındaki bu paranoyayı azaltmak için bazı internet verilerinden yararlanmak istediklerini söylediler.

Ben tüm bu meselelerin biraz boyumu aştığını düşünsem de, ciddi meseleler olduğuna eminim ve veri biliminin bu konularda yardımcı olabileceğine inanıyorum.

Bu dijital hakikat serumunun son ve bence en değerli yanı, bizi problemlerden çözümlere götürebilmesidir. Anlayış gücümüz arttığında dünyanın çirkin tavırlar stoğunu tüketmenin yollarını bulabiliriz.

Obama'nın İslamofobi hakkındaki konuşmasına dönelim. Obama ne zaman Müslümanlara daha fazla saygı gösterilmesi gerektiğini savunsa, ulaşmaya çalıştığı insanlar daha da öfkeleniyordu.

Gelgelelim Google aramaları o zamanki başkan Obama'nın isteyebileceği türde bir yanıtı tetikleyen bir cümle olduğunu gösteriyordu. "Müslüman Amerikalılar bizim dostumuzdur, komşumuzdur, mesai arkadaşımızdır, spor kahramanlarımızdır ve evet, ülkemizi savunmak için canını vermeye hazır olan üniformalı erkek ve kadınlarımızdır," demişti.

Bu sözlerden sonra, bir yılı aşkın süredir ilk kez, Müslüman kelimesinin peşinden en fazla google'lanan kelime "terörist", "aşırılıkçı" ya da "mülteci" değil, "sporcu", ondan sonra da "asker" oldu. Aslında "sporcu" kelimesi sonraki gün boyunca zirvede kaldı.

Arama verileri, kızgın insanlara bir şeyler anlatmaya çalışırken öfkelerinin artabileceğini ima ediyor. Ama insanların merakını inceden inceye gıdıklamak, yeni bilgiler vermek ve öfke ateşlerine odun atan gruba dair yeni imgeler sunmak düşüncelerini farklı ve daha olumlu yönlere döndürebilir.

Obama ilk konuşmadan iki ay sonra, İslamofobi hakkında bir televizyon konuşması daha yaptı ve bu kez bir camiden seslendi. Başkanın ofisinden birisi, Solta'nın ve benim *Times*'ta çıkan ve neyin işe yarayıp neyin yaramadığını ele alan yazımızı okumuş olabilir; zira bu konuşmanın içeriği bariz bir şekilde farklıydı.

Obama hoşgörünün değerinden bahsettiği kısmı çok kısa tutmuş, bunun yerine ağırlıklı olarak insanların merakını tetiklemeye ve Müslüman Amerikalılarla ilgili algılarını değiştirmeye odaklanmıştı. Afrika'dan gelen kölelerin çoğu Müslümandı, demişti Obama; Thomas Jefferson ve John Adams'ın kütüphanelerinde Kuran'ları vardı; ABD topraklarındaki ilk cami Kuzey Dakota'daydı; Chicago'da gökdelenleri

bir Müslüman Amerikalı tasarlamıştı. Obama yine Müslüman atletlerden ve ordu mensuplarından bahsetmiş, ama aynı zamanda Müslüman polisleri ve itfaiyecileri, öğretmenleri ve doktorları da anmıştı.

Google aramalarına ilişkin analizim, bu konuşmanın öncekinden daha başarılı olduğunu gösteriyor. Müslüman aleyhtarı nefret dolu, öfkeli aramalar başkanın konuşmasından saatler sonra azalmıştı.

Nefreti nelerin doğurduğunu ya da azalttığını öğrenmek için arama verilerinden yararlanmanın başka potansiyel yolları da vardır. Örneğin bir şehirde siyah bir savunma oyuncusu takıma seçildiğinde ırkçı aramaların nasıl değiştiğine ya da bir kadın vekil seçildiğinde cinsiyetçi aramaların nasıl değiştiğine bakabiliriz. Irkçıların toplum destekli polisliğe nasıl yanıt verdiğini ya da cinsiyetçilerin yeni cinsel taciz yasalarına nasıl yanıt verdiğini görebiliriz.

Bilinçdışı önyargılarımız hakkında bir şeyler öğrenmek de yararlı olabilir. Örneğin hepimiz kız çocukların zihinlerini önemseyip, görünüşlerine daha az önem vermekten haz duymak için fazladan çaba harcayabiliriz. Google arama verileri ve internetteki diğer hakikat pınarları bizi insan ruhunun en karanlık köşelerine doğru bir yolculuğa çıkarır. Kabul, kimi zaman bununla yüzleşmek zordur. Ama yer yer insana güç de verebilir. Bu verilerden yararlanarak karanlıkla mücadele edebiliriz. Dünya sorunları hakkında zengin veriler toplamak bunları düzeltmeye giden yolda ilk adımdır.

BEŞİNCİ BÖLÜM

Yakından Bakmak

Kardeşim Noah benden dört yaş küçük.

Birçok insan ilk bakışta birbirimize acayip benzediğimizi düşünüyor. İkimiz de yüksek sesle konuşuyoruz, aynı şekilde kelleşiyoruz ve evlerimizi derli toplu tutmakta çok zorlanıyoruz.

Ama farklar da var: Ben hesabımı bilirim. Noah ise hep en iyisini alır. Ben Leonard Cohen ile Bob Dylan dinlerim. Noah için ise varsa yoksa Cake ve Beck.

Ama belki de aramızdaki en büyük fark beyzbolla ilişkimiz. Ben beyzbol manyağıyım, özellikle de New York Mets'e olan aşkım her zaman kimliğimin bir parçası olmuştur. Noah ise beyzbolu çok sıkıcı buluyor, tahammül edemediği bu spora duyduğu nefret uzun zamandır kimliğinin bir parçası olmuştur.*



Seth Stephens-Davidowitz
Beyzbol düşkünü



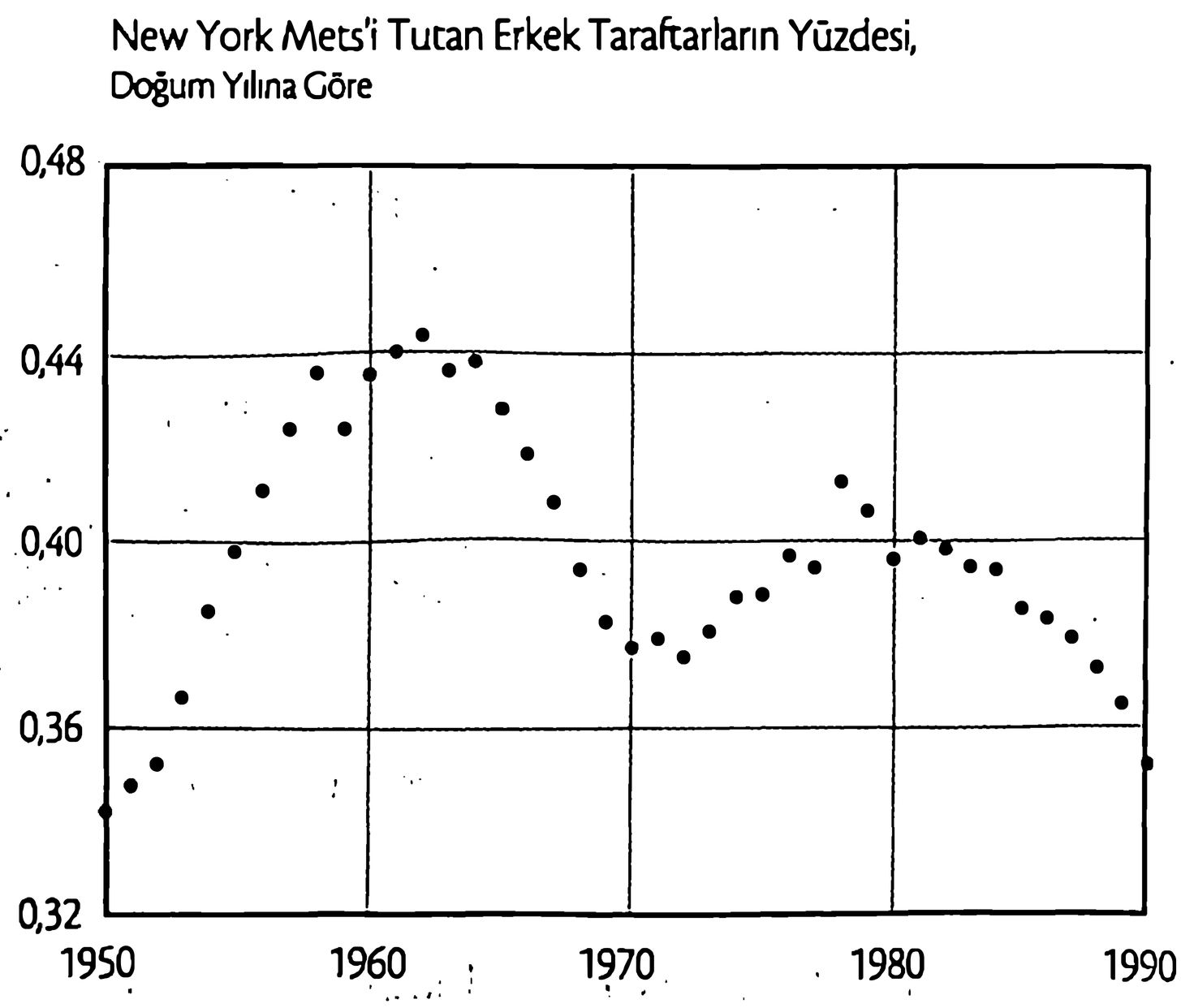
Noah Stephens-Davidowitz
Beyzbol düşmanı

* Gerçeği saklayacak değilim: Bu kitabın veri doğrulamasını yaptığım sırada Noah, ABD'nin gözde eğlencesine duyduğu nefretin kişiliğinin kilit bir parçası olduğunu reddetti. Beyzboldan nefret ettiğini kabul ediyor, ama kişiliğinin temel öğelerinin nezaket, çocuk sevgisi ve zekâsı olduğunu ve beyzbolla ilişkisinin ilk 10'a bile giremeyeceğini düşünüyor. Ne var ki ben insanın bazen kendi kimliğini nesnel bir şekilde göremeyeceği kanaatindeyim ve dışarıdan bir göz olarak, beyzboldan nefret etmesinin –o kabul etsin etmesin– Noah'nın kimliği açısından temel olduğunu görebiliyorum. Bu yüzden değişiklik yapmadım.

Böyle benzer genlere sahip, aynı anne-baba tarafından aynı kentte yetiştirilmiş iki adam nasıl olur da beyzbol konusunda böyle zıt duygular besler? Yetişkinlikteki halimizi belirleyen nedir? Daha temelde, Noah'ın nesi var? Gelişim psikolojisinde yetişkinlerle ilgili muazzam büyüklükte veritabanlarını derinlemesine inceleyerek, bulguları çocukluktaki kilit olaylarla bağıntılandıran, büyümekte olan bir alan var ki, bu meseleyi ve ilintili sorunları ele almamıza yardımcı olabilir. Psikolojik soruları cevaplamak için Büyük Veri'den bu şekilde giderek daha fazla yararlanılmasına Büyük Psike adını verebiliriz.

Bunun nasıl işlediğine bakmak için, çocukluk deneyimlerinin tuttuğunuz beyzbol takımı (ya da takım tutup tutmamanız) konusundaki etkisiyle ilgili yürüttüğüm bir çalışmaya bakalım. Bu çalışma için Facebook'taki beyzbol takımlarıyla ilgili "beğeni" verilerinden yararlandım. (Önceki bölümde Facebook verilerinin hassas konularda son derece yanıltıcı olabileceğini belirtmiştim. Bu çalışmamda hiç kimsenin, hatta Phillies taraftarlarının bile Facebook'ta hangi takımı tuttuğunu söylemekten utanmayacağını varsayıyorum.)

Öncelikle, New York'un iki beyzbol takımını da "beğenen" her yaştan kaç erkek olduğunu gösteren verileri indirdim. İşte doğum yıllarına göre Mets taraftarlarının yüzdesi:

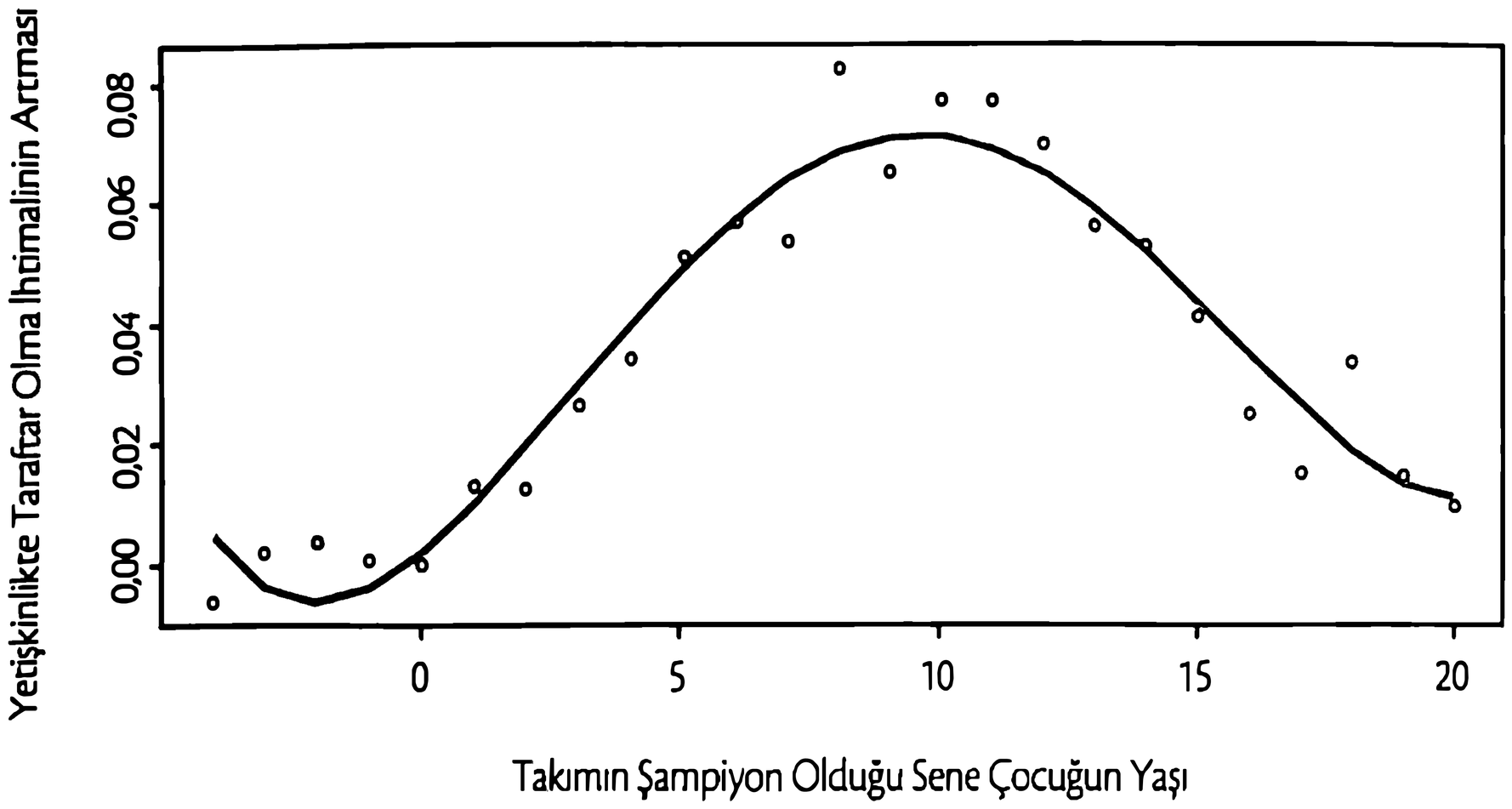


Nokta ne kadar yüksekse, Mets taraftarları da o kadar fazladır. Takımın popülerliği yükselip alçalıyor, sonra yeniden yükselip alçalıyor; nihayetinde, Mets'in 1962 ile 1978 doğumlular arasında çok popüler olduğunu görüyoruz. Beyzbol taraftarları bunun neden kaynaklandığını anlıyordur herhalde. Mets'in sadece iki

şampiyonluğu bulunuyor: Biri 1969'da, diğeri 1986'da. Bu erkekler Mets şampiyon olduğunda kabaca yedi-sekiz yaşındaydı. Dolayısıyla en azından erkek çocukların Mets taraftarı olmaları ihtimalini artıran en önemli etkenlerden biri, yedi-sekiz yaşlarındayken Mets'in şampiyon olduğunu görmeleridir.

Aslında, bu analizi genişletebiliriz. Facebook'ta her yaştan kaç tane taraftarın MLB (Amerikan Ulusal Beyzbol Ligi) takımlarından birini "beğen"diğini gösteren bilgileri indirdim.

Erkekler arasında 1962 doğumlu Baltimore Orioles taraftarları ile 1963 doğumlu Pittsburgh Pirates taraftarlarının sayısının da çok yüksek olduğunu tespit ettim. Bu adamlar, takımları şampiyon olduğunda sekiz yaşında birer çocuktu. Gerçekten de taraftarlarını incelediğim bütün takımların zirve yaptığı yaşı hesaplayıp, ardından bu taraftarların kaç yaşında olduklarını bulunca, şu grafiği çıkarmayı başardım:



Bir kez daha, bir erkeğin hayatı boyunca hangi beyzbol takımını tutacağının kesinleşmesi bakımından en önemli yaşın kabaca sekiz olduğunu görüyoruz. Genele bakıldığında, bir erkek çocuğunu kazanmak için beş ile on beş yaş arası kilit dönemdir. On dokuz-yirmi yaşına gelmiş bir erkek için şampiyonluk, sekiz yaşındaki bir çocuk üzerinde yarattığı etkiyle kıyaslandığında ancak sekizde bir oranında belirleyici olmaktadır. O yaşa gelmiş bir erkek, bir takımı ölene kadar tutup tutmamaya çoktan karar vermiştir.

"Peki, ya kadın beyzbol taraftarları?" diyebilirsiniz. Örüntüler az çok kesindir, ama zirve yaşının yirmi iki olduğu görülüyor.

Bu benim en gözde çalışmam. En sevdiğim iki konuyu birbirine bağlıyor: Beyzbol ile yetişkinlik çağındaki huzursuzluklarımın kaynakları. 1986'dan beri

ölümüne Mets taraftarıyım ve o günden beri de yüzüm gülmedi. Noah Allah'ın sevdiği kuluymuş ki, benden dört sene sonra doğdu da bu acıya ortak olmadı.

Beyzbol dünyadaki en önemli şey değildir ya da en azından tez danışmanlarım ikide bir böyle diyorlar. Ama bu metodoloji, benzer sorunları ele almamıza yardımcı olabilir: Mesela insanların siyasal tercihleri, cinsel yönelimleri, müzik zevkleri ve parasal konulardaki alışkanlıkları nasıl gelişir? (Kardeşimin şu son iki konu hakkındaki manyak fikirlerinin kaynaklarına bilhassa inmek isterdim.) Tahminim o ki, yetişkin olarak davranışlarımızın ve ilgilerimizin büyük bölümünün, hatta bizi biz yapan şeyler dediğimiz temel unsurların bile doğduğumuz tarihle ve gençliğimize denk gelen kilit tarihlerde olup bitenler gibi tesadüfi olgularla açıklanabileceğini göreceğiz.

Nitekim siyasal tercihlerin kökeni konusunda halihazırda yapılmış çalışmalar bulunuyor. Bir veri analiz şirketi olan Catalist'in yöneticisi Yair Ghitza ve Columbia Üniversitesi'nde siyaset bilimi ve istatistik bölümlerinde ders veren Andrew Gelman çoğu insanın ilk başlarda liberal olduğu ve yaşlandıkça muhafazakârlaştığı şeklindeki geleneksel fikri test etmeye çalışmıştı. Çoğu zaman Winston Churchill'e atfedilen meşhur alıntıdaki görüşten bahsediyorum: "30'undan önce liberal olmayan adamın kalbi yoktur; 30'undan sonra muhafazakâr olmayan adamınsa beyni yoktur."

Ghitza ve Gelman altmış yılık anket ve araştırma verilerini incelediler ve oy tercihleriyle ilgili 300 binden fazla gözlemden yararlandılar. Churchill'in iddiasının aksine, ergenlerin bazen liberalliğe bazen de muhafazakârlığa meylettiğini buldular. Aynısı orta yaşlı insanlar ve yaşlılar için de geçerliydi.

Bu araştırmacılar siyasal görüşlerin aslında hangi takımı tuttuğumuzla benzer bir tarzda şekillendiğini tespit etmişlerdi. Hayatta kritik bir dönem vardır ki, o dönemde olup bitenler ömür boyu peşimizden gelir. Kritik yaşlar olan on dört ile yirmi yaş arasındaki sayısız Amerikalı siyasi görüşlerini mevcut başkanın popülerliği temelinde şekillendirecek. Popüler bir Cumhuriyetçi ya da popüler olmayan bir Demokrat partili, birçok genç yetişkinin Cumhuriyetçi olmasını etkileyecektir. Popüler olmayan bir Cumhuriyetçi ya da popüler bir Demokrat partiliyse bu etkilenebilir grubu Demokratların safına yönlendirecektir.

Ayrıca bu kilit yıllardaki görüşler, ortalamaya vurulduğunda, ömür boyu sürecektir.

Bunun nasıl işlediğine bakmak için, 1941'de doğmuş Amerikalılar ile on yıl sonra doğmuş olanları karşılaştıralım.

İlk gruptakiler popüler bir Cumhuriyetçi olan Dwight D. Eisenhower'ın başkanlığı sırasında büyüdüler. 1960'ların başında, otuz yaşından küçük olmalarına karşın, bu

kuşağa mensup kişilerin çoğu sıkı birer Cumhuriyetçi oldu. Ayrıca bu yaş grubundakiler yaşları ilerledikçe de istikrarlı şekilde Cumhuriyetçi Parti'ye meylettiler.

On yıl sonra doğan Amerikalılar ise son derece popüler bir Demokrat partili olan John F. Kennedy'nin; ilk başlarda popüler bir Demokrat partili olan Lyndon B. Johnson'ın ve sonunda rezil olup istifa eden Cumhuriyetçi Richard M. Nixon'ın başkanlık dönemlerinde büyüdüler. Bu kuşağın üyeleri hayatları boyunca liberal oldular.

Araştırmacılar ellerindeki verilere dayanarak siyasal görüşlerin gelişimi açısından en önemli yaşın on sekiz olduğunu belirlediler.

Ayrıca kişinin yaşamına damga vuran bu etkilerin kayda değer sonuçlar yarattığını gördüler. Oluşturdukları modele göre, Eisenhower deneyimi 1941'de doğmuş Amerikalılar arasında ömür boyu Cumhuriyetçileri destekleyen kitleyi yüzde on oranında artırmıştı. Kennedy, Johnson ve Nixon deneyimiyse Demokratlara 1952 doğumlular arasında yüzde 7'lik bir üstünlük sağlamıştı.

Araştırma verilerine şüpheyle yaklaştığımı söylemiştim, ama burada incelenen çok sayıda yanıttan etkilenmedim değil. Aslında bu çalışma tek bir küçük kamuoyu yoklamasıyla gerçekleştirildi. Araştırmacılar insanlar yaşlandıkça tercihlerin nasıl değiştiğini görmek için çok sayıda kamuoyu yoklamasından toplanan yüz binlerce gözleme başvurmuşlardı.

Veri büyüklüğü de beyzbol çalışmam açısından hayati önemdeydi. Sadece tek tek bütün takımların taraftarlarına değil, her yaştan insana odaklanmam gerekti. Bunu yapmak için milyonlarca gözleme ihtiyaç var, ama neyse ki Facebook ve diğer dijital kaynaklar bu tedariki düzene oturtmuş durumdadır.

İşte Büyük Veri'nin büyüklüğü tam da burada devreye giriyor. Bir fotoğrafın küçük bir parçasına net bir şekilde odaklanmak için çok sayıda piksele ihtiyaç vardır. Benzer şekilde, bir veri kümesindeki küçük bir altkümeye (örneğin Mets'in 1978 doğumlular arasında ne kadar popüler olduğuna) net bir şekilde odaklanabilmek için de o veri kümesinde çok sayıda gözlemin bulunmasına ihtiyaç vardır. Birkaç bin insanla yapılan bir araştırma bu tür kişilerle ilgili yeterli büyüklükte bir örnekseme sunmayacaktır.

Büyük Veri'nin üçüncü güçlü yanı budur: Büyük Veri bir veri kümesinin küçük segmentlerine anlamlı şekilde odaklanma imkânı sunarak kim olduğumuza dair yeni içgörüler elde etmemizi sağlar. Ayrıca yaş haricinde diğer boyutlara da odaklanabiliriz. Eğer yeterli veriye sahipsek, belli kasaba ve şehirlerdeki insanların nasıl davrandıklarını görebiliriz. Ayrıca insanların saat saat, hatta dakika dakika neler yaptıklarını görebiliriz.

Bu bölümde insan davranışı mercek altına alınıyor.

Bölgelerimizde, Şehirlerimizde, Kasabalarımızda Gerçekte Neler Olup Bitiyor?

Bugün geriye dönüp bakıldığında şaşırtıcı gelse de, o dönem Harvard'da profesör olan Raj Chetty ve küçük araştırma ekibi epey büyük bir veri kümesini (1996'dan itibaren bütün Amerikalıların vergi kayıtlarını) ilk kez incelediklerinde, bundan bir şey çıkacağından emin değillerdi. IRS bu verileri araştırmacılara vermişti, çünkü ekiptekiler bunlardan veri politikasının sonuçlarının netleştirilmesine yardımcı olacak şekilde yararlanabileceklerini düşünüyorlardı.

Chetty ve ekibinin bu Büyük Veri'den yararlanma konusundaki ilk denemeleri aslında sayısız çıkmazla sonuçlandı. Hem eyalet düzeyinde hem de federal veri politikalarının sonuçlarına dair araştırmaları bu araştırmaları yapan hemen herkesle aynı sonuçlara varıyordu. Belki Chetty'nin cevaplarının, yüz milyonlarca IRS verisinden yararlanması sebebiyle biraz daha kesin olduğu söylenebilirdi, o kadar. Ama diğer herkesle aynı cevaplara ulaşmak, biraz daha kesinlik içerse de, sosyal bilimlerde büyük bir başarı sayılmaz; en önemli dergilerin yayımlamaya heves ettikleri türde çalışmalar kapsamına girmez.

Dahası tüm IRS verilerini düzenleyip analiz etmek zaman alan bir işti. Chetty ve ekibi verilerin altında boğularak herkesten çok zaman harcamalarına rağmen herkesle aynı cevaplara ulaşıyorlardı.

Büyük Veri'ye şüpheyle yaklaşanlar haklı çıkıyor gibiydi: Veri politikasını anlamak için yüz milyonlarca Amerikalıya ait verilere ihtiyaç yoktur; on bin kişiyle yapılacak bir araştırma yeter de artar bile. Bu nedenle, Chetty ve ekibinin hevesi kırılmıştı.

Ama nihayetinde, nerede yanlış yaptıklarını fark ettiler: "Büyük Veri'yle çalışmak, araştırmalarda ya da kamuoyu yoklamalarında yapılanların aynısını daha fazla veriyle yapmaktan ibaret değildir," diyor Chetty. Ellerine tutuşturulan devasa büyüklükteki verilere çok az veri sorusu soruyorlardı. "Büyük Veri bir kamuoyu yoklamasında olduğundan tamamen farklı tasarımlardan yararlanmanıza izin vermelidir. Örneğin belli coğrafi bölgelere daha fazla odaklanabilirsiniz."

Başka bir deyişle, Chetty ve ekibi yüz milyonlarca insana ait verilerle, şehirlerdeki, kasabalardaki ve mahallelerdeki irili ufaklı örüntüleri tespit edebilirlerdi.

Chetty her bir Amerikalının vergi kayıtlarından yararlandığı çalışmasının ilk sonuçlarını bir derslikte sunduğunda ben Harvard'da lisans öğrencisiydim. Sosyal bilimciler çalışmalarında gözlemlere (ne kadar veri göstergesine sahip oldukları-na) atıf yaparlar. Eğer bir sosyal bilimci araştırmasını sekiz yüz insanla yapıyorsa,

"Sekiz yüz gözleme sahibiz," diyecektir. Eğer yetmiş kişiyle bir laboratuvar deneyi yapıyorsa, "Yetmiş gözleme sahibiz," diyecektir.

Chetty ise söze "2,1 milyar gözleme sahibiz," diyerek girmişti. Dinleyiciler heyecanla kıkırdamışlardı.

Chetty ve mesai arkadaşları bu derslikte ve ardından birkaç yazıda ABD'nin nasıl işlediğine dair yeni ve önemli içgörüler sunmaya başladılar.

Şu soruyu düşünün: ABD bir fırsatlar ülkesi midir? Anne-babanız zengin değilse, sizin zengin olma şansınız var mıdır?

Bu soruyu cevaplamanın geleneksel yolu temsil niteliği taşıyan bir sayıda Amerikalıya bakıp bunu başka ülkelerden alınmış benzer verilerle karşılaştırmaktır.

İşte size fırsat eşitliği üzerine çeşitli ülkelerden veriler. Sorulan soru: Anne-babası gelir dağılımının en alt yüzde yirmilik diliminde bulunan bir insanın gelir dağılımının en üst yüzde yirmilik dilimine girme şansı nedir?

Yoksul Bir Ailede Yetişen Birinin Zengin Olma Şansı
(Seçilmiş Ülkeler)

ABD	7,5
Birleşik Krallık	9,0
Danimarka	11,7
Kanada	13,5

Göreceğiniz üzere, ABD'nin sicili pek de parlak değildir.

Ama bu basit analizle yetinirsek gerçek hikâyeyi gözden kaçırırız. Chetty'nin ekibi coğrafyaya da odaklanmıştı. ABD'nin neresinde doğduğunuza bağlı olarak oranlar epey değişiyordu.

Yoksul Bir Ailede Yetişen Birinin Zengin Olma Şansı
(ABD'den Seçilmiş Bölgeler)

San Jose, California	12,9
Washington, DC	10,5
ABD Ortalaması	7,5
Chicago, Illinois	6,5
Charlotte, Kuzey Carolina	4,4

ABD'nin bazı bölgelerinde yoksul bir çocuğun zengin olma şansı, dünyanın herhangi bir gelişmiş ülkesindeki kadar yüksektir. ABD'nin diğer bölgelerindeyse yoksul bir çocuğun zengin olma şansı dünyanın herhangi bir gelişmiş ülkesindekinden daha düşüktür.

Bu örüntülerin küçük bir araştırmayla görülmesi mümkün değildir; zira böyle bir araştırma Charlotte ve San Jose'deki az sayıda insanı dikkate alacak ve dolayısıyla bunun gibi bir odaklanmayı engelleyecekti.

Aslında Chetty'nin ekibi bölgelere daha da yakından bakabilirdi. Ellerinde çok fazla veri (tek tek her Amerikalıyla ilgili veri) olduğundan, bir şehirden başka bir şehre taşınan küçük insan gruplarına odaklanarak bu hamlenin zenginleşme şanslarını ne derece artırdığını görebilirlerdi: örneğin New York'tan Los Angeles'a, Milwaukee'den Atlanta'ya, San Jose'den Charlotte'a. Bu sayede sadece bağıntı değil, sebep-sonuç testi de yapabildiler (bu ayrımı bir sonraki bölümde ele alacağım). Ve evet, insanın gelişim yıllarında doğru şehre taşınması çok önemli bir fark yaratıyordu.

O halde, ABD bir "fırsatlar ülkesi" midir?

Cevap ne evettir ne hayır; bazı bölgelerde evet, bazı bölgelerde hayırdır.

Yazarların ifade ettiği şekliyle, "ABD'yi bir toplumlar toplamı olarak tarif etmek gerekir: Bazıları 'fırsatlar diyarı'dır ve kuşaktan kuşağa sınıf atlama oranları yüksektir; bazılarıysa çok az çocuğun yoksulluktan kurtulabildiği toplumlardır."

O halde, ABD'de gelir dağılımında değişikliğin yüksek olduğu bölgelerde farklı olan nedir? Oyun alanını eşit hale getirme, yoksul bir çocuğun gayet güzel bir yaşam sürme şansını artıran nedir? Eğitime daha fazla bütçe ayıran bölgeler yoksul çocuklara daha fazla şans sunuyor. Daha dindar insanların yaşadığı ve suç oranının düşük olduğu yerler daha başarılıdır. Siyahların daha fazla olduğu yerlerse daha başarısızdır. İşin ilginç yanı, bunun sadece siyah çocuklara değil o bölgede yaşayan beyaz çocuklara da etkisi vardır. Çocuğunu tek başına büyüten annelerin çok olduğu yerlerde durum daha kötüdür. Bu sadece bekâr annelerin çocukları için değil, o bölgede yaşayan evli ebeveynlerin çocukları için de geçerlidir. Bu sonuçlardan bazıları yoksul bir çocuğun akranlarının gayet önemli olduğunu söylüyor. Eğer arkadaşları zor koşullarda yetişmişse ve çok az fırsat sahibiyse, yoksullukran kaçmak için daha fazla çabalaması gerekebilir.

Veriler bize ABD'nin bazı bölgelerinin çocuklara yoksulluktan kaçmak için şans verme konusunda daha iyi olduğunu söylüyor. O halde kaderin sillesinden kurtulma şansı sunma konusunda en iyi yerler hangileridir?

* * *

Ölümü yüce eşitleyici olarak görmeyi severiz. Nihayetinde kimse ölümden kaçamaz. Ne yoksul ne kral, ne evsiz adam ne Mark Zuckerberg. Herkes ölür.

Ama zenginler ölümden kaçamıyor olsalar da, veriler artık en azından ölümü erteleyebildiklerini söylüyor. ABD'de gelir dağılımında en üst yüzde 1'lik dilimde yer alan kadınlar, ortalamaya vurulduğunda, en alt yüzde 1'lik dilimde bulunan Amerikalı kadınlardan on yıl daha fazla yaşıyorlar. Aynı oran erkeklerde on beş yıldır.

Bu örüntüler ABD'nin farklı bölgelerinde nasıl farklılık göstermektedir? Ortalama yaşam süreniz nerede yaşadığınıza göre farklılık gösterir mi? Bu farklılık zenginler için ayrı, yoksullar için ayrı mıdır? Raj Chetty'nin ekibi bir kez daha coğrafyaya odaklanarak cevaplara ulaştı.

İşin ilginç yanı, en zengin Amerikalıların nerede yaşadıkları ortalama yaşam sürelerini pek etkilenmiyor. Eğer fazla fazla paranız varsa, kadınlarda takriben seksen dokuz yaşına kadar, erkeklerde de takriben seksen yedi yaşına kadar yaşamak mümkündür. Zengin insanlar her yerde genellikle daha sağlıklı alışkanlıklara sahipler; genel itibariyle, daha fazla spor yapıyorlar, daha iyi yiyorlar, daha az sigara içiyorlar ve obez olma riskleri daha düşük oluyor. Zengin insanların parası yürüme bandı, organik avokado ve yoga dersi almaya yetiyor ve bunlar ABD'nin her yerinde satın alınabilir.

Yoksullardaysa durum farklıdır. En yoksul Amerikalılarda ortalama yaşam süresi nerede yaşadıklarına bağlı olarak muazzam farklılık göstermektedir. Aslında doğru yerde yaşamak yoksul bir insanın ortalama yaşam süresine beş yıl katabilir.

O halde, neden bazı yerlerin yoksullara daha fazla yaşama imkânı sunduğu görülmektedir? Yoksul insanların daha uzun yaşadıkları şehirlerin ortak özellikleri nelerdir?

İşte size bir şehrin dört özelliği; bunlardan üçü yoksul insanların ortalama yaşam süresiyle bağıntılı değildir, ama bir tanesi bağıntılıdır. Bilin bakalım hangisi?

Bir Şehirde Yoksul İnsanların Daha Uzun Yaşamasını Sağlayan Nedir?

Şehrin dindarlık düzeyinin yüksek olması.

Şehrin hava kirliliğinin düşük olması.

Şehrin sağlık sigortasına sahip sakinlerinin yüzdesinin yüksek olması.

Şehirde çok sayıda zenginin yaşaması.

İlk üçü (din, çevre ve sağlık sigortası) yoksullar için uzun yaşam süresiyle bağıntılı değildir. Chetty'ye ve bu çalışmada yer alan diğer araştırmacılara göre, önem arz eden değişken hangisidir? Bir şehirde kaç tane zenginin yaşadığıdır. Bir şehirde daha fazla zenginin yaşaması orada yaşayan yoksulların daha uzun yaşamaları anlamına geliyor. Örneğin New York'taki yoksullar Detroit'tekilerden daha uzun yaşıyorlar.

Neden zengin insanların varlığı yoksulların ortalama yaşam süreleri için böyle güçlü bir etkendir? Hipotezlerden biri bu çalışmanın yazarlarından ve aynı zamanda benim danışmanlarımdan biri olan David Cutler'a aittir, ama biraz spekülatiftir: İtici güçlerden birinin bulaşıcı davranış olabileceğini söylüyor.

Alışkanlıkların bulaşıcı olduğunu gösteren çok sayıda araştırma vardır. Dolayısıyla zenginlerin yanında yaşayan yoksullar, havasından mıdır suyundan mıdır bilinmez, onların alışkanlıklarını kapabilirler. Bu alışkanlıklardan bazıları (sözgelimi özenti dil) insanın sağlığını etkilemez. Diğerleri (çözüm odaklı çalışma gibi) kesinlikle olumlu etkide bulunacaktır. Gerçekten de zengin insanların yakınında yaşayan yoksullar daha fazla spor yapıyorlar, daha az sigara içiyorlar ve obez olma ihtimalleri daha düşüktür.

* * *

Raj Chetty'nin ekibinin vergi verilerinin incelenmesine dayalı çalışmasında şahsen benim en sevdiğim kısım, neden bazı insanların vergi konusunda dolandırıcılık yaparken bazılarının yapmadığına dair soruşturmalarıydı. Bu çalışma biraz daha karmaşık bir açıklama gerektirmektedir.

Kendi işinde çalışan tek çocuklu insanların devletten aldıkları parayı azamiye çekmelerinin kolay bir yolu olduğunu bilmek kilidi açacaktır. Eğer belli bir yıl içinde tastamam 9 bin dolar vergilendirilebilir gelire sahip olduğunuzu beyan ederseniz, hükümet size 1.377 dolarlık bir çek yazacaktır: Bu miktar, Kazanılmış Gelir Vergisi Kredisi adı verilen ve yoksul çalışanların kazancına ek geliri ifade eden paydan bordro vergisinin çıkarılmasıyla hesaplanır. Bundan daha fazla beyanda bulunduğunuz takdirde bordro verginiz artacaktır. Bundan daha az beyanda bulunursanız, Kazanılmış Gelir Vergisi Krediniz düşecektir. 9 bin dolarlık vergilendirilebilir gelir kritik eşiktir.

Ve ne tesadüftür ki, kendi işinde çalışan tek çocuklu kişilerin çoğunun beyan ettiği vergilendirilebilir gelir miktarı 9 bin dolardır!

Bu Amerikalılar çalışma takvimlerini, ne eksik ne fazla tam bu geliri elde edecek şekilde mi düzenliyorlardı? Hiç de değil! Bu çalışanlar rastgele denetlendiğinde

(çok nadir bir durumdur), neredeyse hiçbirinin 9 bin civarında kazanmadığı tespit edilmişti: Ya çok daha fazla ya da çok daha az kazanıyorlardı.

Başka bir deyişle, kendilerine hükümetten en dolgun çeki almalarını sağlayacak miktarı uydurarak dolandırıcılık yapıyorlardı.

O halde, bu vergi kaçakçılığı türü ne kadar yaygındır ve kendi işinin patronu olan tek çocuklu kişiler arasında bu suçu işlemesi en muhtemel olan kişiler kimlerdir? Chetty ve meslektaşlarının aktardığına göre, ABD'de bu dolandırıcılık türünün yaygınlığı bölgeden bölgeye muazzam farklılıklar gösteriyordu. Miami'de bu kategoriye dahil olan kişiler arasında yüzde 30 gibi şaşırtıcı oranda insan 9 bin dolar kazandığını beyan etmişti. Aynı oran Philadelphia'da yüzde 2'ydi.

Kimin dolandırıcılık yapacağını nasıl kestirebiliriz? En fazla dolandırıcının olduğu yerler ile en az dolandırıcının olduğu yerler arasındaki ayırıcı özellik nedir? Dolandırıcılık oranları ile şehir düzeyindeki diğer demografik özellikler arasında bağıntı kurabiliriz; bu durumda iki güçlü etken göze çarpacaktır: Kazanılmış Gelir Vergisi Kredisi'ne hak kazanmış insanların bir bölgede yoğun derecede yaşıyor olmaları ve bir mahallede mali müşavirlerin yoğun derecede yaşıyor olmaları.

Bu etkenler neyi gösterir? Chetty ve ekibinin bir açıklaması vardı. Vergi konusunda bu tür bir dolandırıcılık yapmanın gerisindeki kilit saik bilgi sahibi olmaktı.

Kendi işinin patronu olan tek çocuklu vergi mükelleflerinin çoğu hükümetten dolgun bir çek almak için sihirli meblağın 9 bin dolar olduğunu bilmiyordu. Ama bunu bilebilecek insanların (komşu ya da vergi danışmanları) yakınlarında yaşamak bu konuda bilgi sahibi olma ihtimallerini ciddi düzeyde artırmıştı.

Aslında Chetty'nin ekibi bu dolandırıcılık türünün bilgi sahibi olmakla bağıntılı olduğuna dair başka kanıtlar da bulmuşlardı. Amerikalılar bu vergi dolandırıcılığı türünün düşük olduğu bir bölgeden yüksek olduğu bir bölgeye taşındıklarında, bu dalavereyi öğrenip hayata geçiriyorlardı. Zaman içinde dolandırıcılık ABD'ye bölge bölge yayılmıştı. Vergi dolandırıcılığı tıpkı bir virüs gibi bulaşıcıdır.

Şimdi bir an durup bu çalışmanın ne kadar bilgilendirici olduğunu düşünelim. Kimlerin vergi dolandırıcılığı yapacağını öngörmeye çalışırken kilit husus, kimin dürüst olup kimin olmadığını belirlemek değildir. Mesele kimin dolandırıcılık yapmayı bilip kimin bilmediğini belirlemektir.

Dolayısıyla birisi size asla vergi dolandırıcılığı yapmayacağını söylediğinde, çok büyük ihtimalle (doğru tahmin!) yalan söylüyordur. Chetty'nin araştırması, dolandırıcılık yapmayı bilseydi birçok insan dolandırıcılık yapardı, diyor.

Eğer vergi dolandırıcılığı yapmak istiyorsanız (yanlış anlaşılma olmasın, zorla aklınıza sokmaya çalışmıyorum!), mali müşavirlerin ya da size nasıl dolandırıcılık yapılacağını gösterebilecek kişilerin yakınlarında yaşamalısınız. Eğer dünya çapında şöhrete sahip çocuklarınız olsun istiyorsanız, nerede yaşamanız gerekir? Verilere yakından bakabilmek ve gerçekten en ince ayrıntılara ulaşabilmek bu soruyu da cevaplandırmamızı sağlayacaktır.

* * *

En başarılı Amerikalıların nereli olduklarını çok merak ediyordum, bu yüzden bir gün oturup Wikipedia bilgilerini indirmeye karar verdim. (Bugünlerde bu tür şeyler yapabiliyorsunuz.)

Çok az kodlama yapmama karşın, Wikipedia editörlerinin başlık açmaya değer gördüğü 150 binden fazla Amerikalıyı kapsayan bir veri kümesi elde ettim. Veri kümesinde doğum yeri, doğum tarihi, meslek ve cinsiyet bilgileri vardı. Bunu Sağlık İstatistikleri Ulusal Merkezi'nin topladığı bölge çapında doğum verileriyle birleştirdim. ABD'deki her bölge için orada doğan birinin Wikipedia'da yer alma şansının ne olduğunu hesapladım.

Wikipedia'da profil sahibi olmak çok başarılı olduğunuzun anlamlı bir işareti midir? Kuşkusuz sınırlılıklar vardır. Wikipedia editörleri gençlere ve erkeklere ayrıcalık tanıyorlar ve bu da örneklemeler açısından taraflılık ifadesidir. Bazı kayda değer türlerin ise özel olarak değerli bir yanı yoktur. Örneğin Ted Bundy onlarca genç kadını öldürdüğü için Wikipedia maddesi olmaya hak kazanmıştır. Bu çerçevede, suçluları sonuçları fazla değiştirmeyecek şekilde ayıklamayı başardım.

Çalışmamı doğum oranının yüksek olduğu dönemde doğmuş olanlarla (1946-1964 arasında doğanlarla) sınırladım; çünkü bu gruptakiler kayda değer bir başarı elde etmeye yetecek kadar yaşadılar. Bu yıllar arasında doğmuş olan kabaca her 2.058 Amerikalıdan biri Wikipedia'da ayrı bir başlık olmayı hak edecek kadar başarılı görülmüştü. Yaklaşık yüzde 30'u sanat ya da eğlence alanında, yüzde 29'u sporda, yüzde 9'u siyasette, yüzde 3'ü de akademide ya da bilimde başarı elde etmişti.

Verilerde dikkatimi çeken ilk çarpıcı olgu en azından Wikipedia düzeyinde büyük bir başarı elde etme ihtimalinin coğrafi açıdan muazzam bir çeşitlilik arz etmesiydi. Kayda değer başarı elde etme şansınız nerede doğduğunuzla son derece bağlantılıydı.

1946-1964 yılları arasında California'da doğan kabaca her 1.209 kişiden biri Wikipedia'ya adını yazdırmıştı. Bu oran Batı Virginia'da 4.496'da birdi. Bölge bazında sonuçlara odaklanmak daha da bilgilendiricidir. Massachusetts'te Boston'ın

bulunduğu Suffolk County'de doğan her 748 kişiden biri Wikipedia'ya girmeyi başarmıştı. Başka bazı ilçelerdeyse başarı oranı yirmi kat daha düşüktü.

Ülkenin bazı bölgeleri neden ABD'nin gündemine yön verenler grubuna daha fazla üye yolluyor gibiydi? En başta gelen bölgeleri yakından inceledim. Neredeyse tamamının iki kategoriden birine dahil edilebileceğini gördüm.

Birinci (ve beni çok şaşırtan) kategori, bu bölgelerin birçoğunda büyük bir üniversitenin yer almasıydı. Listenin tepelerinde ismini duymadığım bir bölgenin adını her gördüğümde (Michigan'daki Washtenaw gibi), buranın klasik bir üniversite şehrine (bu örnekte Ann Arbor) ev sahipliği yaptığını fark ettim. Madison, Wisconsin; Athens, Georgia; Columbia, Missouri; Berkeley, California; Chapel Hill, North Carolina; Gainesville, Florida; Lexington, Kentucky; Ithaca, New York, hepsi de en üst yüzde 3'tedir.

Neden böyle? Bir açıdan gen havuzuyla açıklanabilir: Profesörlerin ve üniversite öğrencilerinin oğulları ve kızları genellikle akıllı oluyorlar (büyük başarı oyununda epey faydalı olabilecek bir özellik). Gerçekten de bir bölgede üniversite mezunlarının sayıca fazla olması orada doğmuş insanların başarısı açısından önemli bir öngöstergedir.

Ama büyük ihtimalle bunun başka sebepleri de vardır: Erken yaşta yenilikleri deneyimleme. Üniversite kentlerinin şöhret sahibi isimler çıkarma konusunda en başarılı olduğu alanlardan biri müziktir. Bir üniversite kentinde yetişen bir çocuk benzersiz konserlere gidecek, sıradışı radyo istasyonlarını dinleyecek, hatta plak dükkânlarından alışveriş edecektir. Bu karşılaşmalar sanatla da sınırlı değildir. Önde gelen işadamları ve işkadınları havuzunda da üniversite kentlerinin payına çok büyük bir dilim düşer. Belki de çağdaş sanata ve fikirlere küçük yaştan maruz kalmak onlar için de yararlıdır.

Üniversite kentlerinin başarısı sadece bölgeden bölgeye yayılan bir durum değildir. Etnik kökeni de kapsar. Afrikalı-Amerikalılar spor harici alanlarda, özellikle de iş hayatında ve bilimde hak ettikleri kadar yer almıyorlar. Kuşkusuz ayrımcılık bunun temel sebeplerindendir. Buna karşılık, 1950'de nüfusunun yüzde 84'ü siyah olan küçük bir bölgede 1946-1964 arasında doğanların içinden en tepedekilerle neredeyse aynı oranda ünlü çıkmıştır.

Bu dönemde Alabama'daki Macon County'de doğmuş olan 13 bine yakın kişiden on beşi, yani her 852 kişiden biri Wikipedia'da yerini almıştır. Bu kişilerin hepsi de siyahtır. On dördü Booker T. Washington'ın kurduğu ve geçmişten bu yana siyahların gittiği bir okul olan Tuskegee Üniversitesi'nin bulunduğu Tuskegee'den çıkmıştır. Listede hukukçular, yazarlar ve bilimciler vardır. Aslında Tuskegee'de

doğmuş bir siyah çocuğun spor dışındaki bir alanda başarılı olma ihtimali ile yüksek puanlı bir üniversitenin bulunduğu ve çoğunlukla beyazların yaşadığı kentlerden birinde doğmuş beyaz bir çocuğunki aynıydı.

Bir bölgenin sakinlerinin başarılı olma ihtimalini arttıran ikinci özellik o bölgede büyük bir şehrin olmasıydı. San Francisco, Los Angeles ya da New York'ta doğmuş olmak Wikipedia'ya girme olasılığını en çok artıran etkenler arasındaydı. (New York'un beş bölgesini birlikte değerlendirdim, çünkü birçok Wikipedia girdisi doğum yerinin hangi ilçe olduğunu belirtmiyordu.)

Kentsel alanlar başarılı örnekler bakımından çok güçlüydü. Gençken bir mesleği başarıyla icra edenlerin yakınlarında olmanın değerini anlamak için New York, Boston ve Los Angeles'ı karşılaştırın. Bu üçünden oran olarak en fazla ünlü gazeteciyi New York; en fazla ünlü bilimciyi Boston; en fazla ünlü aktörü de Los Angeles çıkarmıştır. Unutmayalım, orada doğmuş olanlardan bahsediyoruz, oraya taşınmış olanları dahil etmiyoruz. Ebeveynleri o alanda başarılı olan insanları çıkardıktan sonra bile bu söylenenler geçerliliğini korur.

Önemli üniversite kentlerinin yer almadığı banliyö bölgeleri kentsel muadillerine kıyasla çok daha kötü bir performans gösteriyordu. Kendi anne-babam, bu dönemde doğmuş birçok kişi gibi, kalabalık kaldırımlardan ağaçların gölgelediği sokaklara (bizim hikâyemizde Manhattan'dan New Jersey'deki Bergen County'ye) taşınarak üç çocuklarını yetiştirmişlerdi. Bu potansiyel açıdan yanlış bir adımdı, en azından ünlü çocuklara sahip olma açısından değerlendirildiğinde. New York'ta doğmuş bir çocuğun Wikipedia'ya girme şansı Bergen County'deki bir çocuğa kıyasla yüzde 80 daha fazladır. Elbette bunlar sadece birer bağıntıdır, ama büyük fikirlerin yakınında yöresinde büyümenin arka bahçesi büyük bir evde büyümekten daha iyi olduğunu gösteriyor.

Çocukluğun geçirildiği yerler konusunda daha sağlam verilere sahip olsaydım, bu konuda tespit ettiğim önemli etkiler daha da güçlü olabilirdi; zira birçok insan doğduğu yerde büyümüyor.

Üniversite kentlerinin ve büyük şehirlerin başarısı sadece verilere baktığınızda bile çarpıcıdır. Ama ben daha incelikli bir ampirik analiz yapmak amacıyla daha derin bir sondaj çalışmasına giriştim.

Bu sayede, Wikipedia'da yer alabilmenin güçlü bir öngöstergesi olan başka bir değişkene ulaştım: Doğduğunuz bölgedeki göçmenlerin oranı. Bir bölgede yabancı uyruklu sakinlerin oranı ne kadar yüksekse, orada doğan çocukların ileride kayda değer bir başarı elde etme oranı o kadar yüksektir. (Donald Trump, bu sana gelsin!)

Kent ve üniversite nüfusları benzer olan iki yerden daha fazla göçmenin bulunduğu yer daha önemli şahsiyetler çıkaracaktır. Bunu neyle açıklayabiliriz?

Başarının büyük bir kısmı doğrudan göçmenlerin çocuklarına atfedilebilir. Wikipedia verilerinden yararlanarak yürütülen MIT Pantheon projesine göre bu dönemde doğmuş en ünlü isimlerden yüzlercesinin biyografisini didik didik ederek inceledim. Bu kişilerin çoğu eğlence sektöründe çalışıyordu. En az on üçünün annesi yabancı uyrukluydu (Oliver Stone, Sandra Bullock ve Julianne Moore dahil). Bu oran söz konusu dönemde ülke ortalamasının üç katından fazladır. (Steve Jobs ve John Bellushi gibi birçoğunun babası göçmendi, ama bu veriyi ülke ortalamasıyla kıyaslamak daha zordu, zira doğum belgelerinde her zaman babalarla ilgili bilgi yer almıyor.)

Peki, ya başarıya etki etmeyen değişkenler? Şaşırtıcı bulduğum değişkenlerden biri bir eyaletin eğitime ne kadar bütçe ayırdığıydı. Kentsel alanlarda yaşayan nüfusu benzer oranlarda olan eyaletlerde eğitime ayrılan bütçe ile önemli yazar, sanatçı ya da iş dünyasının duayenlerini çıkarma oranları arasında bağıntı yoktu.

Wikipedia incelememi Chetty'nin ekibinin yukarıda bahsi geçen çalışmalarıyla karşılaştırmak ilgi çekici olacaktır. Hatırlayacak olursak, Chetty'nin ekibi üst orta sınıfa dahil olabilmek için hangi alanların ön açıcı olduğunu bulmaya çalışıyordu. Bense, şöhrete ulaşabilmek için hangi bölgelerin ön açıcı olduğuna odaklanmıştım. Sonuçlardaki farklılık çarpıcıdır.

Eğitime çok para harcamak çocukların ileride üst orta sınıfa dahil olmasına yardımcı oluyor. Ama önemli bir yazar, sanatçı ya da iş dünyasında lider hale gelmelerine pek etki etmiyor. Bu muazzam başarı hikâyelerinin çoğunda kişilerin okuldan nefret ettiğini görüyoruz. Öyle ki aralarında okulu bitirmemiş olanlar bile var.

Chetty'nin ekibinin bulgularına göre, çocuğunu üst orta sınıfa dahil etmek isteyen bir ebeveyn için New York hiç de iyi bir yer değildir. Benim çalışmama göreyse, şöhrete erişme şansı elde etmek için müthiş bir yerdir.

Başarının arkasındaki etkenlere baktığınızda, bölgeler arasındaki büyük farklılık anlam kazanıyor. Birçok bölge, başarı için gerekli temel bileşenlerin hepsine sahiptir. Bir kez daha Boston'a dönelim. Sayısız üniversitenin olduğu Boston yenilikçi fikirlerle dolup taşmaktadır. Gençlere başarıya giden yolu örneklerle gösteren son derece başarılı birçok insanın yaşadığı bir kentsel alandır. Ayrıca pek çok göçmen (ve başarılı örneklerden kendilerine ders çıkarıp uygulayacak olan çocukları) için çekim merkezidir.

Peki, ya bir bölgede bu niteliklerden hiçbiri yoksa? Daha az süperstar çıkarmaya mahkûm mudur? İlla böyle olması gerekmiyor. Başka bir yol daha var: Aşırı uzmanlaşma. Çok az yabancının yaşadığı, tek bir üniversitenin bile yer almadığı

kırsal bir bölge olan Minnesota'daki Roseau County iyi bir örnektir. Burada doğan kabaca her 740 kişiden biri Wikipedia'da yerini almıştır. Sırrı? Dokuzu da profesyonel hokey oyuncusuydu: Şüphesiz, bu bölgenin dünya çapındaki genç ve lise hokey programı sayesinde.

Dolayısıyla burada mesele (derdinizin bir hokey yıldızı yetiştirmek olmadığını varsayarak ilerliyorum) müstakbel çocuklarınıza azami üstünlük sağlamaksa, Boston ya da Tuskegee'ye taşındığınız takdirde pişman olmazsınız. Ama buradan çıkarılacak daha büyük dersler de var. İktisatçılar ve sosyologlar genellikle kötü sonuçlardan (yoksulluk ve suç gibi) nasıl kaçınılabileceğine odaklanırlar. Ama büyük bir toplumun hedefi geride az sayıda insanın kalmasını sağlamaktan ibaret değildir; hedef, olabildiğince çok insanın gerçekten öne çıkmasına yardım etmektir. Belki de en ünlü Amerikalıların yüz binlercesinin doğduğu bu yerlere yakından bakmaya çalışmak, başlangıç düzeyinde bazı stratejiler oluşturmamıza yardımcı olabilir: Göçün teşvik edilmesi, üniversitelerin sübvanse edilmesi ve sanatın desteklenmesi bunlardan birkaçıdır.

* * *

Ben genellikle ABD üzerine çalışıyorum. Dolayısıyla coğrafya temelinde odaklanmayı düşündüğümde, ABD'deki şehirlere ve bölgelere odaklanmayı, Alabama'daki Macon County ya da Minnesota'daki Roseau County gibi yerlere bakmayı düşünüyorum. Ama internetten elde edilen verilerin bir diğer devasa –ve giderek artan– faydası dünyanın dört bir tarafından veriler toplamanın kolaylığıdır. Bu sayede, ülkeler arasındaki farkları görebiliriz. Veri bilimcileri yavaş yavaş antropoloji alanına kayma fırsatını elde ediyorlar.

Son dönemde biraz rastgele irdelediğim konulardan biri şuydu: Dünyanın farklı ülkelerinde hamileliğin etkisi nedir? Hamile kadınların Google aramalarını inceledim. Elde ettiğim ilk bulgu kadınların şikâyet ettiği fiziksel emarelerdeki çarpıcı benzerlikti.

"Hamile" kelimesiyle birlikte çeşitli emarelerin ne kadar sık arandığını test ettim. Örneğin "hamile" kelimesinin "mide bulantısı", "sırt ağrısı" ya da kabızlık" ile birlikte aranma sıklığı neydi? Kanada'daki emareler ABD'dekilere çok yakındı. Britanya, Avustralya ve Hindistan gibi ülkelerdeki emareler de kabaca benzerdi.

Ayrıca tüm dünyada hamile kadınlar benzer şeyleri aşeriyorlardı. ABD'de bu kategoride en fazla google'lanan arama "hamilelik sırasında dondurma aşerme"dir. Ardından tuz, tatlı, meyve ve baharatlı yiyecekler geliyor. Avustralya'da bu aşermeler hiç değişiklik göstermiyor: Listede tuz, tatlı, çikolata, dondurma ve meyve vardır. Peki, ya Hindistan? Benzer bir hikâye: Baharatlı yiyecekler, tatlılar, çikolata, tuz ve dondurma. Aslında baktığım ülkelerin hepsinde ilk beş benzerdi.

Muvakkat kanıtlar dünyanın hiçbir bölgesinde hamileliğin fiziksel deneyimini köklü şekilde değiştiren bir beslenme tarzının ya da çevrenin söz konusu olmadığını gösteriyor.

Ama hamilelik etrafındaki düşünceler kesinlikle farklılık gösteriyor.

Hamile kadınların gönül rahatlığıyla neler yapabilecekleri sorusuyla başlayalım. ABD'de en çok sorulan sorular: Hamile kadınlar "karides yiyebilir mi", "şarap içebilir mi", "kahve içebilir mi" ya da "soğuk algınlığı ilacı kullanabilir mi"dir.

"Hamilelik sırasında nasıl...?" aramalarında ilk beş (sırayla)

ABD	HİNDİSTAN	AVUSTRALYA	BRİTANYA	NİJERYA	GÜNEY AFRİKA
deride çatlakları engellemek	uyku	deride çatlakları engellemek	kilo vermek	seks yapmak	seks yapmak
kilo vermek	sevişmek	kilo vermek	deride çatlakları engellemek	kilo vermek	kilo vermek
seks yapmak	seks yapmak	deride çatlakları engellemek	deride çatlakları engellemek	sevişmek	deride çatlakları engellemek
deride çatlakları engellemek	seks	uyumak	uyumak	sağlıklı kalmak	uyumak
formda kalmak	kendine bakmak	sevişmek	sevişmek	kusmayı durdurmak	kusmayı durdurmak

"Hamile kadınlar ... yapabilir mi?" aramalarında ilk beş

ABD	karides yemek	şarap içmek	kahve içmek	ağrı kesici almak	suşi yemek
BRİTANYA	büyük karides yemek	somo füme yemek	cheesecake yemek	mozzarella yemek	mayonez yemek
AVUSTRALYA	krem peyniri yemek	büyük karides yemek	domuz pastırması yemek	ekşi yemek	feta peyniri yemek

NİJERYA	soğuk su içmek	şarap içmek	kahve içmek	sevişmek	moringa bitkisi yemek
SİNGAPUR	yeşil çay içmek	dondurma yemek	durian yemek	kahve içmek	ananas yemek
İSPANYA	pâté yemek	domuz budu pastırması yemek	parasetamol almak	ton balığı yemek	güneş banyosu yapmak
ALMANYA	uçmak	salam yemek	saunaya girmek	bal yemek	mozzarella yemek
BREZİLYA	saçını boyamak	ağrı kesici almak	parasetamol almak	bisiklete binmek	uçmak

Bu tür kaygılar söz konusu olduğunda, diğer ülkelerin ABD'yle ya da kendi aralarında pek fazla ortak özelliği yoktur. Hamile kadınların şarap içip içemeyecekleri Kanada, Avustralya ya da Britanya'da ilk on soru arasında değildir. Avustralya'daki kadınların kaygıları çoğunlukla süt ürünleri, özellikle de krem peynir yemekle alakalıdır. Nüfusun yüzde 30'unun internet kullandığı Nijerya'da ilk soru, hamile kadınların soğuk su içip içemeyecekleridir.

Bunlar haklı kaygılar mıdır? Yerine göre. Hamile kadınların pastörize olmayan peynirlerden listeriya kapma riski çok yüksektir. Hamileyken çok fazla alkol almanın çocuk için olumsuz sonuçlar doğurduğu kanıtlanmıştır. Dünyanın bazı bölgelerinde soğuk su içmenin bebeğinizin zatürree olmasına yol açabileceğine inanılmaktadır; bu savın tıbben desteklenip desteklenmediğini bilmiyorum.

Dünyanın çeşitli bölgelerinde sorulan sorular arasındaki muazzam farkların sebebi büyük ihtimalle her ülkede farklı kaynaklardan gelen müthiş bilgi akışıdır: Meşru bilimsel çalışmalar, kısmen bilimsel denebilecek çalışmalar, kocakarı hikâyeleri ve mahalle dedikoduları. Kadınların neye odaklanacaklarını ya da neyi google'yacaklarını bilmeleri zordur.

"Hamilelik sırasında nasıl...?" için en çok yapılan aramalara baktığımızda, başka bir bariz farkı görüyoruz. ABD'de, Avustralya'da ve Kanada'da en çok yapılan arama, "hamilelik sırasında deride çatlaklar oluşması nasıl engellenir"dir. Ama Gana, Hindistan ve Nijerya'da deride çatlaklar oluşmasını engelleme kay-

gısı ilk beşte bile değildir. Bu ülkelerde genellikle nasıl seks yapılacağı ya da nasıl uyunacağı ilgi konusudur.

Kuşkusuz, dünyanın farklı köşelerinde sağlık ve kültürün veçhelerine odaklanarak öğrenilecek çok şey vardır. Ama benim ilk analizime göre, biyolojimizi aşmak söz konusu olduğunda, Büyük Veri bize insanların farkına vardığımız kadar bile güçlü olmadığını göstereceğini söylüyor. Fakat bunun ne anlama geldiğine ilişkin yorumlar kayda değer farklılıklar sergiliyor.

Dakikalarımızı ve Saatlerimizi Nasıl Doldururuz?

"Esas ilgi alanları tecavüz, aşırı şiddet ve Beethoven olan genç bir adamın maceraları."

Stanley Kubrick'in ihtilaflı filmi *Otomatik Portakal*'ın (1971) tanıtımı bu şekilde yapılmıştı. Filmin başkarakteri olan genç Alex DeLarge gözünü bile kırpmadan korkunç şiddet eylemleri gerçekleştirir. Filmin en meşum sahnelerinden birinde, bağıra çağıra "Singin' in the Rain" şarkısını söylerken bir kadına tecavüz eder.

Filmden neredeyse hemen sonra, gerçek hayatta birebir aynı hadiselerin yaşandığı haberleri gelmeye başlamıştı. Gerçekten de bir grup erkek aynı şarkıyı söyleyerek on yedi yaşında bir kıza tecavüz etmişti. Film birçok Avrupa ülkesinde yasaklandı, ABD'de en sarsıcı kısımları sansürlenerek yayınlandı.

Aslında gerçek hayatın sanatı taklit ettiği, insanların ekranda gördükleri şeyle hipnotize oldukları birçok örnek vardır. Gangster filmi *Renkler*'in bir gösteriminden sonra silahlar konuşmuştu. Keza bir diğer gangster filmi *New Jack City*'nin bir gösteriminden sonra da küçük çaplı bir ayaklanma çıkmıştı.

Belki de en rahatsız edici olanı, *Para Treni* gösterime girdikten dört gün sonra, insanlar neredeyse filmdeki bir sahneyi birebir taklit ederek metronun bilet gişelerini ateşe vermek için çakmak gazı kullanmışlardı. Kurmaca hayat ile gerçek dünya arasındaki tek fark kundaklama kısmıdır: Filmde gişe memuru kaçmayı başarır. Gerçek hayatta yanarak ölür.

Keza psikolojik deneyler de bir şiddet filmine maruz kalan deneklerin sahnelerden birini birebir taklit etmeseler bile, öfke ve nefret duyguları sergilemeye başladıklarına dair kanıtlar sunuyor.

Başka bir deyişle, anekdotlar ve deneyler şiddet filmlerinin şiddet davranışlarını kışkırtabileceğini söylüyor. Peki, bunların etkisi ne kadar büyüktür? On yılda bir-iki cinayetten mi, yoksa her yıl yüzlerce cinayetten mi bahsediyoruz? Anekdotlar ve deneyler bunun cevabını veremiyor.

İktisatçılar Gordon Dahl ve Stefano DellaVigna cevabı Büyük Veri'nin verip veremeyeceğini anlamak amacıyla, 1995-2004 arasındaki üç büyük Veri Kümesi'ni birleştirmişti: FBI'ın saat saat suç verileri, film gişe rakamları ve kids-in-mind.com'daki her filmden şiddet ölçümü.

Yararlandıkları bilgi tamdı: ABD'deki şehirlerde saat saat işlenen her suç ve her film. Bu çok önemliydi.

Çalışmalarının anahtarı bazı hafta sonlarında en popüler filmin bir şiddet filmi (mesela *Hannibal* ya da *Ölülerin Şafağı*) iken, diğer hafta sonlarında en popüler filmin *Kaçak Gelin* ya da *Oyuncak Hikâyesi* gibi şiddet içermeyen filmler olmasıydı.

Bu iki iktisatçı bilindik bir şiddet filminin gösterime girdiği hafta sonlarında kaç cinayet, tecavüz ve saldırı vakası görüldüğüne bakıp, bu verileri şiddet içermeyen bir filmin gösterime girdiği hafta sonlarında kaç tane cinayet, tecavüz ve saldırı vakası görüldüğüyle karşılaştırdılar.

Peki, ne buldular? Bir şiddet filmi gösterime girdiğinde, bazı deneylerin gösterdiği gibi suç sayısı artış gösteriyor muydu, yoksa aynı mı kalıyordu?

Popüler bir şiddet filminin gösterime girdiği hafta sonlarında suç oranları düşüyordu.

Yanlış okumadınız. Popüler bir şiddet filminin gösterime girdiği hafta sonlarında milyonlarca Amerikalı, bir insanın başka bir insanı öldürdüğü görüntülere maruz kalmasına karşın, suç oranları düşüyordu, hem de ciddi şekilde.

Böyle tuhaf ve beklenmedik bir sonuç elde ettiğinizde, aklınıza gelen ilk düşünce, "Acaba bir şeyi yanlış mı yaptım?" oluyor. İki iktisatçı da kodlamalarının üzerinden bir daha geçtiler. Yanlış yoktu. Hemen ardından akla gelen düşünceyse, bu sonuçları açıklayacak başka bir değişkenin olabileceğidir. Yılın belli dönemlerinde farklılık olup olmadığını kontrol ettiler. Sonuç yine değişmemişti. Hava durumuyla ilgili verilere baktılar, belki de ilişkinin devindirici gücü buydu. Hayır, bu da değildi.

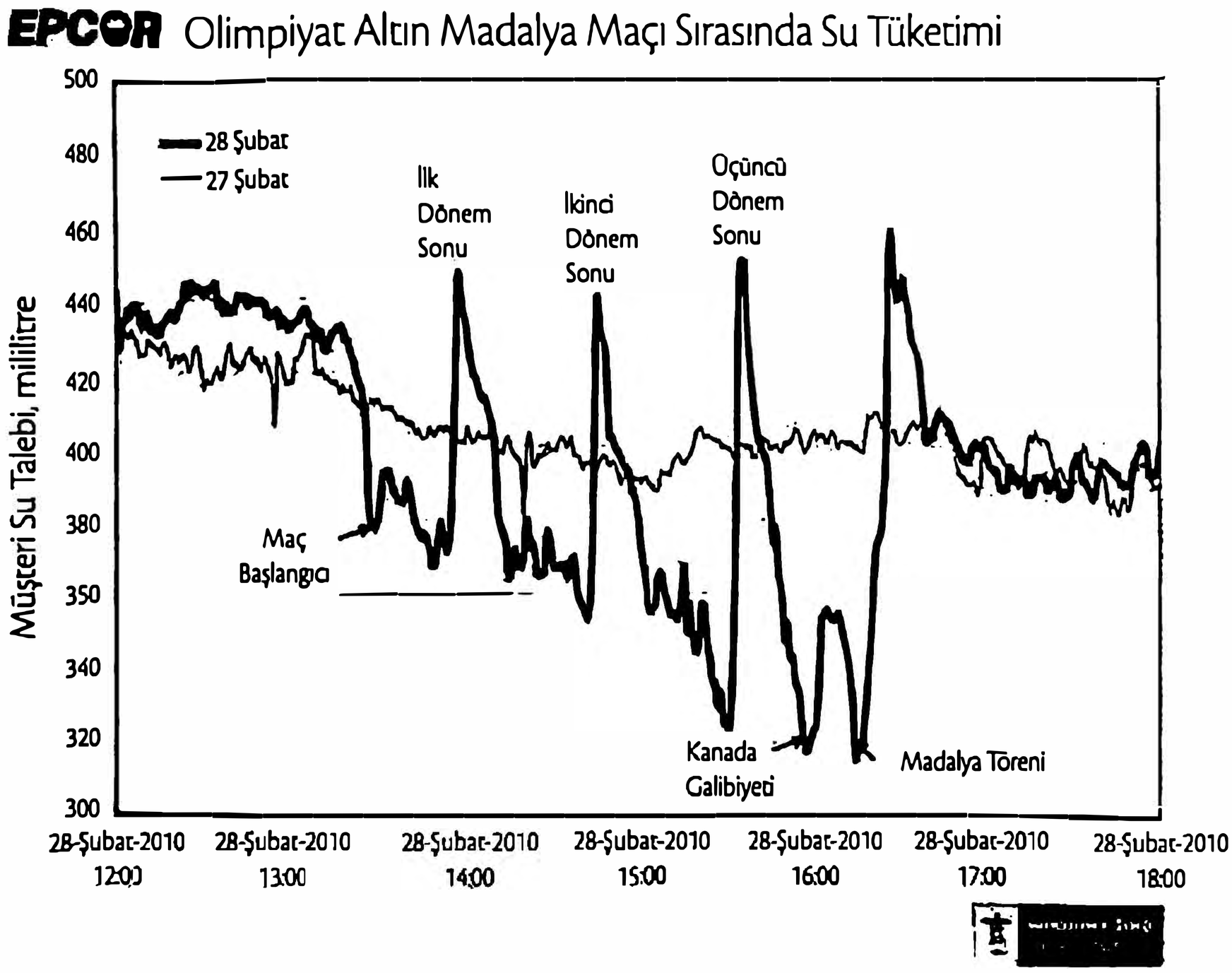
Dahl bana, "Bütün varsayımlarımızı, yaptığımız her şeyi kontrol ettik. Yanlış bir şey bulamadık," demişti.

Anekdotlara rağmen, laboratuvar kanıtlarına rağmen ve ne kadar tuhaf görünürse görünsün, bir şiddet filminin gösterilmesi suç oranlarında bir şekilde büyük bir düşüşe neden oluyordu. Böyle bir şey nasıl olabilirdi?

Dahl ve DellaVigna için kilidi açan etken Büyük Veri'den yararlanarak konuya biraz daha yakından bakmak oldu. Araştırma verileri geleneksel olarak yıllık ya da en iyi ihtimalle aylık bilgiler sunar. Eğer gerçekten şanslıysak, bir hafta sonuna ait veriler elde edebiliriz. Karşılaştırma yaptığımızda, küçük örneklemelerle yapılan

araştırmalardansa giderek daha kapsamlı veri kümelerinden yararlandığımız için, saatlik, hatta dakikalık verileri bile çalışmamıza dahil edebiliyoruz. Bu bize insan davranışları hakkında çok daha fazla şey öğrenme imkânı sunuyor.

Bazen zaman içindeki dalgalanmalar dünyayı yerinden oynatacak kadar büyük olmasa bile eğlenceli olabiliyor. Kanada'nın Edmonton şehrinde kamu hizmeti veren şirketlerden EPCOR, 2010 Olimpiyatları'nda ABD ile Kanada'nın hokey takımları arasında oynanan ve Kanadalıların yaklaşık yüzde sekseninin izlediği altın madalya maçında dakika dakika su tüketimi verilerini aktarmıştı. Veriler her periyodun bitiminden kısa süre sonra su tüketiminin arttığını gösteriyordu. Edmonton'daki tuvaletler resmen coşmuştu.



Google aramaları dakika dakika da analiz edilebilir ve bu süreçte ilginç birtakım örüntüler saptanabilir. Örneğin hafta sonları sabah 8'de "engellenmiş oyunlar" aramaları tepe noktasına varıyor ve öğleden sonra 3'e kadar bu seviyelerde kalıyor; kuşkusuz bu, okulların öğrencilerin cep telefonlarını yasaklamadan okul sınırları dahilinde internetten oyun oynamalarını engelleme girişimlerine bir yanıttır.

"Hava durumu", "dua" ve "haberler" ile ilgili arama oranları sabah 5.30'dan önce zirve noktasına ulaşıyor: Demek ki insanlar benden çok daha önce uyanıyorlar! "İntihar" arama oranları öğlen 12.36'da zirveye ulaşıyor ve sabah 9 civarında en düşük seviyelerde bulunuyor: Demek ki çoğu insan sabahları benim kadar sefil bir durumda değil.

Veriler büyük soruların gece 2 ila 4 arasında sorulduğunu gösteriyor: Bilincin anlamı nedir? Özgür irade var mıdır? Diğer gezegenlerde yaşam var mı? Gece geç saatlerde bu soruların popüler olması kısmen esrar kullanımının bir sonucu olabilir. "Ot nasıl sarılır?" aramaları gece 1 ile 2 arasında tepe noktasına varıyor.

Dahl ve Dellavigna ellerindeki büyük veri kümesi sayesinde, film gösteriminin olduğu hafta sonlarında suç oranlarının saat saat nasıl değiştiğini görebilmişlerdi. Popüler şiddet filmlerinin gösterildiği hafta sonlarında suç oranlarındaki düşüş (diğer hafta sonlarına kıyasla) sabah erken saatlerde başlıyordu. Başka bir deyişle, suç oranları daha şiddet sahneleri gösterilmeden, belki de sinemaya gidenler yoldayken düşüyordu.

Nedenini tahmin edebilir misiniz? İlkin, kimlerin şiddet filmlerine gitmeyi tercih edebileceğini düşünün. Genç erkekler, özellikle de genç, saldırgan erkekler.

Ardından, suçların genellikle nerede işlendiğini düşünün. Nadiren bir sinema salonunda. İstisnalar var tabii; en önemlisi de 2012'de Colorado'da bir sinema salonundaki planlı silahlı saldırıdır. Ama genel itibariyle, insanlar sinemaya silahsız gider ve sessizce otururlar.

Genç, saldırgan erkeklere *Hannibal*'ı izleme fırsatı sunarsanız, sinemaya gidecekler. Genç, saldırgan erkeklere bunun alternatifi olarak *Kaçak Gelin*'i önerirseniz, reddedecek ve bunun yerine dışarı çıkacak, belki bir bara, kulübe ya da bilardo salonuna gideceklerdir ki, bu yerlerde şiddet ve suç oranı daha yüksektir.

Şiddet filmleri, şiddete bulaşma potansiyeli yüksek insanları sokaklardan uzak tutuyor.

Bulmaca çözüldü, değil mi? Henüz değil. Verilerde tuhaf bir şey daha vardı. Bahsi geçen etkiler tam da filmler gösterilmeye başladığında kendini gösteriyordu; ama film bittikten ve sinema kapandıktan sonra durmuyordu. Şiddet filmlerinin gösterildiği hafta sonlarında suç oranları gecenin geç saatlerine kadar (gece yarısından sabah altıya kadar) daha düşüktü.

Suç oranları genç erkeklerin sinemada olduğu saatlerde daha düşük olsa bile, sinemadan çıktıklarında ve artık meşgaleleri olmadığında oranların yükselmesi gerekmez miydi? Sonuçta, daha biraz önce bir şiddet filmi izlemişlerdir ve deneyler bunun insanları daha kızgın ve saldırgan hale getirdiğini söylüyor.

Şiddet oranlarının film bittikten sonra da düşük kalmasını nasıl açıklayabiliriz? Aynı zamanda suç uzmanı olan yazarlar, epey bir düşündükten sonra, bir tane daha "Aha!" ânı yaşadılar. Alkolün suça ciddi bir katkısı olduğunu biliyorlardı. Yazarlar ABD'de neredeyse hiçbir sinema salonunda içki satılmadığını bilecek kadar sinemaya

gitmişlerdi. Nitekim yazarlar alkolle alakalı suçların şiddet filmlerinden sonra gece geç saatlerde düştüğünü tespit etmişlerdi.

Elbette Dahl ve DellaVigna'nın sonuçları sınırlıydı. Sözgelimi aylık, kalıcı sonuçları test edemiyor, suçtaki düşüşlerin ne kadar uzun sürdüğünü göremiyorlardı. Ayrıca şiddet filmlerine düzenli bir şekilde maruz kalmanın nihayetinde daha fazla şiddete yol açması hâlâ mümkündür. Gelgelelim çalışmaları bu deneylerin esas izleği olan şiddet filmlerinin doğrudan etkisini belli bir perspektife oturtuyor. Belki de şiddet filmleri bazı insanları gerçekten etkiliyor ve alışılmadık derecede sinirli ve saldırgan hale getiriyor. Peki, insanları kesin olarak şiddet doğrultusunda etkileyen şeyler nelerdir? Şiddet potansiyeli olanlarla takılmak ve içki içmek.*

İşte şimdi bu bir anlam ifade ediyor. Ama Dahl ve DellaVigna veri yığınlarını incelemeden önce anlam ifade etmiyordu.

Daha yakından baktığınızda açık hale gelen başka bir husus da dünyanın karmaşık olduğudur. Bugün girişebileceğimiz eylemlerin çoğu uzun vadede hiç de niyet etmediğimiz sonuçlar doğurabilir. Fikirler bazen yavaş yavaş, bazen de virüs gibi jet hızında yayılır. İnsanlar teşviklere kestirilemeyen şekillerde yanıt verirler.

Bu bağlantılar ve ilişkiler, bu alçalışlar ve yükselişler küçük çaplı araştırmalarla ya da geleneksel veri yöntemleriyle incelenemez. Kısacası, dünya az sayıda veriyle anlaşılamayacak kadar karmaşık ve zengindir.

Kötücül İkizlerimiz

2009 Haziran'ında David "Big Papi" Ortiz'in kariyeri bitmiş görünüyordu. Boston beş yıldır, bu Dominik doğumlu, aralık dişli ve güzel gülüşlü şutöre hayrandı.

---

* Bu hikâye kötü görünen bir şeyin, daha kötü bir şeyi engellediğinde iyi olabileceğinin bir göstergesidir. Stanford'da eğitim almış eski bir top yakalayıcı olan Ed McCaffrey dört çocuğunun hepsinin Amerikan futbolu oynamasına izin vermesini meşrulaştırmak için bu argümana başvurur: "Bu çocukların içinde bir enerji var. Futbol oynamasalar, kaykaya biniyorlar, ağaca tırmanıyorlar, arka bahçede ebelemece oynuyorlar, paintball oynuyorlar. Demem o ki orada öylece oturup hiçbir şey yapmayacaklarını düşünemeyiz. Dolayısıyla benim meseleye bakışım şu şekilde: En azından futbol bir spor olduğundan kuralları var... Benim çocuklarımın yere çakıldıkları için, bisiklet kazası geçirdikleri için, kaykay yaptıklarından ötürü veya ağaçtan düştükleri için acile gitmişlikleri var. Siz sorun ben söyleyeyim... Evet, bu şiddet içeren bir spor . Ama öte yandan afacan çocuklarım artık sincaplar gibi dallarda dolaşmıyorlar ya da benzeri delilikler yapmıyorlar. Yani buna tertipli bir saldırganlık diyebiliriz, sanırım." McCaffrey'in *The Herd with Colin Cowherd*'deki mülakatında dile getirdiği bu argümanı daha önce duymamıştım. Dahl ile DellaVigna'nın yazısını okuduktan sonra bu argümanı ciddiye aldım. Laboratuvarlardaki verilerden ziyade gerçek dünyadaki devasa veri kümelerinin bir üstünlüğü de bu tür sonuçları yakalayabilmesidir.

Beş yıl üst üste All-Star (Yıldızlar Karması) maçına seçilmiş, MVP (En Değerli Oyuncu) ödülünü kazanmış ve Boston'ın seksen altı yıllık şampiyonluk hasretine son vermişti. Ama 32 yaşına girdiği 2008 yılında, ortalamaları düşmeye başlamıştı. Vuruş ortalaması 68 puan, oyunda kalma yüzdesi 76 puan, vuruş yüzdesi de 114 puan gerilemişti. 2009 sezonunun başında Ortiz'in ortalamaları düşmeye devam ediyordu.

Spor yazarı ve koyu bir Boston Red Sox taraftarı olan Bill Simmons 2009 sezonunun başlarında olan biteni şöyle anlatıyordu: "David Ortiz'in artık beyzbolda mükemmel düzeyde olmadığı açık... İri yarı, vuruşları sert bir oyuncu porno yıldızlarına, güreşçilere, NBA pivotlarına veya yaşlı adamlarla evlendirilen genç kızlara benzer: Gitti mi gider." Koyu taraftarlar kendi gözlerine güvenirler ve Simmons'ın gözleri Ortiz'in işinin bittiğini söylüyordu. Simmons'ın tahmini Ortiz'in yakında yedeğe çekileceği ya da serbest bırakılacağı yönündeydi.

Gerçekten de Ortiz'in kariyeri bitmiş miydi? Eğer siz 2009'da Boston'ın genel menajeri olsaydınız, Ortiz'i takımdan çıkarır mıydınız? Daha genel bir soru sorarsak, bir beyzbol oyuncusunun ileride nasıl performans göstereceğini nasıl tahmin edebiliriz? Beyzbolu bir kenara bırakıp daha da genel bir soru soracak olursak, insanların gelecekte ne yapacaklarını tahmin etmek için Büyük Veri'den nasıl yararlanabiliriz?

Veri biliminde bizi daha ileri götürecek olan şöyle bir teori var: Sabermetrik (beyzbol ölçüm) uzmanlarının yaptıklarına bakın ve bunun veri biliminin diğer alanlarına yayılmasını umut edin. Beyzbol hemen her şeyle ilgili kapsamlı veri kümelerinin ve hayatlarını bu verileri anlamlandırmaya adamış akıllı insanlar ordusunun bulunduğu ilk alanlardan biridir. Şimdiyse hemen her alanda bu aşamaya gelinmiştir. Beyzbol başı çekmiş, diğerleri peşinden gelmiştir. Sabermetrik dünyaya hükmetmektedir.

Bir beyzbol oyuncusunun geleceğini tahmin etmenin en kolay yolu şu anda olduğu gibi performans göstereceğini varsaymaktır. Eğer bir oyuncu son bir buçuk yıldır kötü oynuyorsa, sonraki bir buçuk yılda da kötü oynayacağını tahmin edebilirsiniz.

Bu metodolojiye göre, Boston'ın David Ortiz'i kadro dışı bırakması gerekirdi.

Ne var ki bazı bilgiler konu hakkında daha iyi fikir verebilir. Çoğu insanın sabermetrik biliminin kurucusu olarak gördüğü Bill James 1980'lerde yaşın önemini vurgulamıştı. James'in bulgularına göre, beyzbol oyuncuları erken yaşta, kabaca 27 yaşında zirveye varıyordu. Takımlar oyuncuların yaşlandıkça ne kadar gerilediğini genellikle görmezden geliyor ve yaşlı oyunculara değerinden fazla para ödüyorlardı.

Boston bu daha gelişkin metodolojiye göre de Ortiz'i kesinlikle kadro dışı bırakmalıydı.

Ama bu yaş analizi bir şeyleri gözden kaçırabilir. Her oyuncunun gelişim çizgisi aynı yolu takip etmez. Bazı oyuncular yirmi üçünde, bazıları otuz ikisinde zirveye ulaşır. Kısa oyuncular uzun oyunculardan, şişman oyuncular cılız oyunculardan daha farklı yaşlanabilir. Beyzbol istatistikçileri farklı şekillerde yaşlanan oyuncu türleri olduğunu tespit ettiler. Bu hikâye Ortiz için daha da beter bir sonuca işaret ediyordu: "İri yarı, vuruşları sert oyuncular", gerçekten de, ortalamaya vurulduğunda, erken yaşta zirveye ulaşıp, otuzunu geçtikten kısa süre sonra çöküyordu.

Boston son yıllardaki performansını, yaşını ve fiziğini dikkate alsaydı, hiç kuşkusuz David Ortiz'i serbest bırakması gerekirdi.

Sonra, 2003'te, istatistikçi Nate Silver PECOTA adını verdiği ve oyuncu performansını kestirmeye yarayan yeni bir model geliştirdi. Bunun mevcut modellerin hepsinden daha iyi ve daha havalı olduğu ispatlandı. Silver oyuncuların kötücül ikizlerini arıyordu. Yöntem şöyle işliyor. Gelmiş geçmiş tüm MLB oyuncularının, yani 18 binden fazla beyzbolcunun yer aldığı bir veri tabanı oluşturun. Bu oyuncular hakkında bildiğiniz her şeyi kümeye dahil edin: Boyları, yaşları, oynadıkları mevki; kariyerlerinin her bir yılındaki *home run*'ları (sayı turu), vuruş ortalamaları, ıskaları ve *strikeout*'ları (oyun dışı). Sonra, kariyerlerinde bu noktaya gelene kadar Ortiz'e en fazla benzeyen yirmi beyzbol oyuncusunu bulun: 24, 25, 26, 27, 28, 29, 30, 31, 32 ve 33 yaşındayken onun gibi oynamış oyuncuları. Son olarak da, Ortiz'in ruh ikizlerinin kariyerlerinin nasıl ilerlediğine bakın.

Ruh ikizi araması, meseleye yakından bakmanın başka bir örneğidir. Belli bir insana en fazla benzeyen insanların oluşturduğu küçük altkümeye odaklanır. Ve her odaklanmada olduğu gibi, elinizde ne kadar veri olursa o kadar iyi sonuç alırsınız. Görünen o ki Ortiz'in ruh ikizleri Ortiz'in geleceği için çok farklı bir tahminde bulunmuştu. Ortiz'in ruh ikizleri arasında Jorge Posada ve Jim Thome vardı. Bu oyuncular kariyerlerine biraz yavaş başlamış, yirmilerinin sonunda müthiş bir patlama gerçekleştirip dünya çapında birer oyuncu haline gelmiş ve otuzlarının başında kötü oynamışlardı.

Silver böylece bu ruh ikizlerinin daha sonra ne yaptığına bakarak Ortiz'in geleceğini tahmin edebilecekti. Ulaştığı sonuç şuydu: Sonra yeniden yükselişe geçmişlerdi. Simmons'ın tespiti yaşlı adamlarla evlendirilen genç kızlar için geçerli olabilirdi: Gitti mi gider. Ama Ortiz'in ruh ikizleri gitseler de geri dönmüşlerdi.

Beyzbol oyuncuklarının performansını tahmin etmek için bu zamana kadar kullanılmış en iyi metodoloji olan ruh ikizi aramaları, Boston'ın Ortiz konusunda

sabırlı olması gerektiğini söylüyordu. Nitekim Boston yaşlanan oyuncusuna müsamaha gösterdi. 2010'da Ortiz'in ortalaması .270'e yükseldi. 32 *home run* yaptı ve Yıldızlar Karması'na seçildi. Ortiz sonraki üç sene de Yıldızlar Karması'na seçildi. Otuz yedi yaşına bastığı 2013'te, kadroda geleneksel olarak üçüncü sırada vuran Ortiz .688 ortalama tuttурdu ve Boston, Final serisinde St. Louis'i 4-2 yendi. Ortiz finallerin en değerli oyuncusu seçildi.*

Nate Silver'ın oyuncuların kariyerlerinin seyrini tahmin etme yaklaşımını okuduktan sonra, benim de bir ruh ikizim olup olmadığını düşünmeye başladım.

Ruh ikizi aramaları sadece sporda değil, birçok alanda umut vaat edicidir: İlgi alanlarımızın en çok örtüştüğü kişiyi bulabilir miydim? Belki bana en uygun kişiyi bulsam birlikte takılabilirdik. Seveceğimiz bir restorana giderdik. Beni seveceğimi hiç düşünmeyeceğim şeylerle tanıştırabilirdi.

Ruh ikizi aramaları bireylerin özelliklerine odaklanır. Ayrıca her odaklanma gibi, elinizdeki veriler arttıkça daha da kesin bir hal alır. Ruh ikizimi on civarı insanın yer aldığı bir veri kümesinde aradığımı varsayalım. Benim gibi kitaplara ilgi duyan birini bulabilirim. Bin küsur kişi arasında aradığımı varsayalım. Popüler fizik kitaplarına ilgi duyan birini bulabilirdim. Ama ruh ikizimi milyonlarca insanın dahil edildiği bir veri kümesinde aradığımı varsayalım. O zaman, hakikaten bana benzeyen birini bulabilirdim.

Bir gün sosyal medyada ruh ikizimi aramaya giriştim. Twitter profillerinin tamamından yararlanarak, ilgi alanlarımın en çok örtüştüğü insanları aradım.

Twitter hesabımdan kimleri takip ettiğime bakarak ilgi alanlarım hakkında kesin şeyler söyleyebilirsiniz. Genele bakıldığında, 250 kadar insanı takip ediyorum; spora, siyasete, komediye, bilime ve Yahudi halk şarkıcılarına ilgi duyduğum görülüyor.

O halde, evrende bu hesapların 250'sini de takip eden biri, yani benim Twitter ikizim var mıydı? Elbette yoktu. Ruh ikizleri bizimle aynı değil, benzerdir. Zaten takip ettiğim hesapların 200'ünü birden takip eden de yoktu, hatta 150'sini bile yoktu.

Gelgelelim sonunda benim takip ettiğim hesapların –inanılmaz ama– 100'ünü de takip eden birini buldum: Country Music Radio Today. O ne yahu? Country

---

* Kitabın bu bölümüne kadar iyi hikâyelere genellikle alaycı bir tarzda yaklaştığımı gördünüz. Burada iyi bir hikâye olsun istedim, dolayısıyla alaycılığımı dipnota bıraktım. Bence PECOTA Ortiz'in steroid kullandığını, steroid kullanmayı bıraktığını ve yeniden kullanmaya başlayacağını öğrenmişti. Tahmin açısından bakıldığında, PECOTA'nın bunu tespit edebilmesi çok havalıdır; fakat bu durum elbette hikâyenin etkileyiciliğine gölge düşürür.

Music Radio Today "takibe takip" ilkesi uyarınca 750 bin Twitter profilini takip eden ve artık kullanılmayan bir bot hesaptı!

Eski kız arkadaşlarımdan biri duysa bu garip durumdan kesin kendine bir pay çıkarırdı. Bir keresinde bana insandan çok robota benzediğimi söylemişti.

Şaka bir yana, ruh ikizimin rastgele 750 bin hesabı takip eden bir bot hesap olduğunu gösteren ilk bulgularım ruh ikizi aramalarındaki çok önemli bir hususu vurguluyor. Bir ruh ikizi aramasının gerçekten doğru olması için, salt sizinle aynı şeyleri seven birini bulmak doğru yöntem değildir. Aynı zamanda, sizin sevmediğiniz şeyleri sevmeyen birini de bulmanız gerekir.

Benim ilgi alanlarım sadece takip ettiğim alanlardan değil, takip etmemeyi seçtiğim alanlardan da anlaşılabilir. Ben sporla, siyasetle, komediyle ve bilimle ilgileniyorum, ama yemek, moda ya da tiyatroyla ilgilenmiyorum. Takiplerim Bernie Sanders'ı sevdiğimi ama Elizabeth Warren'ı sevmediğimi, Sarah Silverman'ı sevdiğimi, ama Amy Schumer'ı sevmediğimi, *New Yorker*'ı sevdiğimi ama *Atlantic*'i sevmediğimi, arkadaşlarım Noah Popp, Emily Sands ve Josh Gottlieb'i sevdiğimi ama arkadaşım Sam Asher'ı sevmediğimi gösteriyor. (Üzgünüm Sam, ama Twitter hesabın içimi bayıyor.)

Twitter'daki toplam 200 milyon insandan benimkine en fazla benzeyen profil hangisiydi? Ruh ikizim meğer *Vox* yazarı Dylan Matthews'muş. Biraz hayal kırıklığına uğradığımı söylemeliyim; zira medya tüketimimi ilerletmek istediğimden zaten Matthews'u Twitter ve Facebook'ta takip ediyor, *Vox*'taki yorumlarını deli gibi okuyordum. Dolayısıyla onun benim ruh ikizim olduğunu öğrenmek hayatımı gerçek manasıyla değiştirmedi. Ama dünyada size en fazla benzeyen kişinin kim olduğunu bilmek çok hoş, özellikle de hayran olduğunuz birisiyse. Bu kitabı bitirip yeniden insan içine karıştığımda, belki Matthews'la takılıp James Surowiecki'nin yazıları hakkında sohbet edebiliriz.

Ortiz'in ruh ikizi araması beyzbol taraftarları için çok netti. Benim ruh ikizi aramamsa eğlenceliydi, en azından benim için. Peki, bu aramalar başka neleri açığa çıkarabilir? Örneğin, ruh ikizi aramaları en büyük internet şirketleri tarafından tekliflerini ve kullanıcı deneyimlerini geliştirmek için kullanılmaktadır. Amazon sevebileceğiniz kitapları önermek için ruh ikizi araması gibi bir şeyden yararlanıyor. Size benzeyen insanların neler seçtiğine bakıyor ve önerilerini buna göre yapıyor.

Pandora da hangi şarkıları dinlemek isteyebileceğinizi belirlemek için aynısını yapıyor. Keza Netflix sevebileceğiniz filmleri böyle belirliyor. Bu yöntemin etkisinin büyüklüğünü şu anekdottan anlayabiliriz: Amazon'un mühendis ekibinden Greg Linden okurların kitap tercihlerini tahmin etmek için ruh ikizi aramalarını

ilk kez kullandığında önerilerdeki ilerleme o kadar iyiymiş ki, Amazon kurucusu Jeff Bezos dizlerinin üstüne çöküp "Ben [Linden'a] layık değilim!" diye bağırmış.

Ama ruh ikizi aramalarının gücünü dikkate aldığımızda, asıl ilgi çekici olan kısım bugün ne kadar yaygın bir şekilde kullanıldıkları değil, ne kadar sık kullanılmadıklarıdır. Hayatın birçok önemli alanında bu tür kişiselleştirme aramalarıyla muazzam bir ilerleme kaydedilebilir. Sözgelimi sağlık alanında.

Harvard'da bilgisayar mühendisi ve tıbbi araştırmacı olan Isaac Kohane bu ilkeyi tıp alanına uygulamaya çalışıyor. Kohane'in amaçladığı şekilde, tüm sağlık bilgilerimiz düzenlendiği takdirde, doktorlar hepsini tek bir torbaya sıkıştıran yaklaşımı bir kenara bırakıp, tam bizim gibi hastaları bulabileceklerdir. Ardından daha kişiselleştirilmiş, daha odaklı tanılar ve tedaviler de mümkün hale gelecektir.

Kohane bunu radikal bir yaklaşım olarak değil, tıp alanının doğal bir uzantısı olarak görüyor. "Tanı nedir?" diye soruyor Kohane. "Tanı aslında önceden incelenmiş hastalarla ortak özelliklerinizin olduğunun belirtilmesidir. Ben size, Allah muhafaza, kalp krizi tanısı koyduğumda, diğer insanlardan öğrendiğim kadarıyla kalp krizi geçirdiğiniz anlamına gelen bir patofizyolojiniz olduğunu kastediyorum."

Tanı, özünde, ilkel türde bir ruh ikizi aramasıdır. Sorun şu ki, doktorların tanı koymak için kullandıkları veri kümeleri küçüktür. Bu günlerde bir tanı doktorun tedavi ettiği hastalarla ilgili deneyimlerine dayalıdır ve belki bir de, diğer araştırmacıların karşılaştıkları küçük popülasyonlardan akademik yazılarla desteklenir. Ama gördüğümüz gibi, bir ruh ikizi aramasının gerçekten iyi olması için çok daha fazla vakayı içermesi gerekir.

İşte size Büyük Veri'nin gerçekten yardımcı olabileceği bir alan. O halde, bu kadar gecikme neden? Neden şimdiden yaygın bir şekilde kullanılmıyor? Sorun veri toplamayla alakalıdır. Tıbbi raporların çoğu hâlâ kâğıt üzerindedir, dosyalara gömülüdür; bilgisayar ortamına aktarılmış olanlarsa çoğu zaman uyumsuz formatları yüzünden sürgülü kapılardan farksızdır. Kohane, sağlığa kıyasla beyzbolda çoğu zaman daha yeterli miktarda veriye sahibiz, diyor. Ama basit ölçümler çok işe yarayacaktır. Kohane sık sık "alt dallardaki meyveler"den bahseder. Sözgelimi salt çocukların boy ve kilolarıyla ve yaşayabilecekleri hastalıklarla ilgili eksiksiz bir veri kümesi oluşturmanın pediatri için devrim niteliğinde sonuçları olacağına inanıyor. O zaman, bütün çocukların büyüme şekli birbiriyle karşılaştırılabilecektir. Bir bilgisayar benzer bir şekilde yol almış çocuklar bulabilir ve sorunlu örüntüleri otomatik şekilde işaretleyebilir. Bir çocuğun boyunun uzamasının vaktinden önce durduğunu tespit edebilir; bu tanı belli durumlarda şu iki sebepten birinden kaynaklanıyor olabilir: tiroit yetmezliği ya da beyin tümörü. Her iki durumda da

erken tanı hayat kurtarır. Kohane'e göre, "Bunlar benzersiz türlerdir, on binde bir görülür. Çocuklar genellikle sağlıklıdır. Kanaatimce çocuklarda görülen hastalıklara daha erken, en az bir yıl daha erken tanı koyabiliriz. Bunu kesinlikle yapabiliriz."

James Heywood tıbbi veriler arasında bağ kurma konusunda yaşanan zorluklara farklı yönden yaklaşan bir müteşebbistir. Heywood'un kurduğu PatientsLikeMe.com adlı sitede bireyler kendi bilgilerini (sağlık durumlarını, tedavilerini ve ilaçların yan etkilerini) paylaşabiliyorlar. Hastalıkların seyrini izleme ve bu konudaki ortak anlayışımızla karşılaştırma konusunda şimdiden epey bir başarı elde etmiştir.

Heywood'un hedefi yeterli sayıda hastayı çekmek, yeterli sayıda sağlık durumunu ele almak ve böylece insanların sağlık alanındaki ruh ikizlerini bulmalarını sağlamaktır. Heywood sizinle aynı yaşta ve cinsiyette olan, aynı hastalıkları yaşamış, sizinkine benzer emareler gösteren insanları bulup onlarda neyin işe yaradığını görebileceğinizi umuyor. Gerçekten de bu çok farklı türde bir tıp olacaktır.

Veri Hikâyeleri

Ben yakından bakma eylemini belli bir çalışmanın tikel verilerine kıyasla birçok açıdan daha değerli görüyorum; çünkü bu yaklaşım yaşama bakmanın, yaşam hakkında konuşmanın yeni bir yolunu sunmaktadır.

İnsanlar benim veri bilimci ve yazar olduğumu öğrendiklerinde, kimi zaman bir olguyu ya da bir araştırmayı dikkatime sunuyorlar. Ben bu verileri çoğu zaman sıkıcı buluyorum: Durağan ve cansızlar. Anlatacak bir hikâyeleri yok.

Aynı şekilde, arkadaşlarımın da onlarla birlikte roman ve biyografi okumamı teklif etmişlikleri var. Ama bunların da benim için ilgi çekici bir yanı yok. Her zaman şu soruları soruyorum: "Başka durumlarda ne olurdu? Geneli kapsayan ilke nedir?" Hikâyeleri temsil edici nitelikten yoksun ve küçük geliyor bana.

Bu kitapta anlatmaya çalıştığım şey, benim açımdan başka hiçbir şeye benzemiyor. Verilere ve rakamlara dayanıyor; örneklendirici ve çok çeşitli sonuçlar barındırıyor. Ama veriler o kadar zengindir ki, insanların bu verilerin altında kaybolduğunu hayal edebilirsiniz. Edmonton'daki su tüketimine dakika dakika odaklandığımızda, kendi adıma ben her periyodun sonunda insanların koltuklarından kalktıklarını görebiliyorum. Philadelphia'dan Miami'ye taşınıp vergi dolandırıcılığı yapmaya başlayan insanlara odaklandığımızda, kendi adıma ben apartmanlarında komşularıyla konuşan ve vergi dalaverecsini öğrenmeye çalışan insanları görebiliyorum. Her yaştan beyzbol taraftarlarına odaklandığımızda, kendi adıma ben kendi çocukluğumu ve kardeşimin çocukluğunu ve de daha sekiz

yaşındayken kalplerini kazanmış olan takım için hâlâ gözyaşı döken milyonlarca yetişkin erkeği görebiliyorum.

Bir kez daha büyük laflar etme pahasına söyleyebilirim ki, bence bu kitapta yer alan iktisatçılar ve veri bilimcileri sadece yeni bir alet değil, aynı zamanda yeni bir yazın türü yaratıyorlar. Bu bölümde ve kitabın çoğunda sunmaya çalıştığım şey, çok büyük ve çok zengin olan, bize çok yakından bakma imkânı veren ve böylece kendimizi temsili niteliği bulunmayan tikel bir insanla sınırlandırmadan karmaşık ve zihin açıcı hikâyeler anlatabilmemizi sağlayan verilerdir.

ALTINCI BÖLÜM

Bütün Dünya Bir Laboratuvardır

Google'ın Mountain View kampüsünde 27 Şubat 2000, sıradan bir gün olarak başladı. Güneş parıldıyordu. Bisikletçiler pedal çeviriyor, masözler masaj yapıyor, çalışanlar salatalık suyuyla susuzluklarını gideriyorlardı. Neden sonra, bu sıradan günde, birkaç Google mühendisinin aklına bugün internetin büyük kısmının devindirici gücü mahiyetindeki bir sırrın perdelerini aralayan bir fikir geldi. Mühendisler sitede tıklamanızı, siteye geri gelmenizi ve sitede kalmanızı sağlayacak en iyi yolu bulmuşlardı.

Ne yaptıklarını tarif etmeden önce, bağıntı ile nedensellik arasındaki ilişkiden bahsetmemiz gerekiyor. Bu ilişki bugün veri analizinde çok büyük bir meseledir ve henüz yeterince ele alınmamıştır.

Medya bizi her gün bağıntı temelli çalışmalarla bombardımana tutuyor. Örneğin tadını kaçırmayacak kadar alkol alan birinin daha sağlıklı olacağı söylenip duruyor. Bu bir bağıntıdır.

Buradan hareket ederek, kararında içki içmenin insanın sağlığını iyiye götüreceği sonucu, yani bir nedensellik (sebep-sonuç ilişkisi) çıkar mı? Sanırım çıkmaz. Sağlıklı olmak insanların kararında içki içmelerine sebep oluyor olabilir. Sosyal bilimciler buna ters nedensellik adını veriyorlar. Veya hem kararında içmeye hem de sağlıklı olmaya sebep olan bağımsız bir etken var olabilir. Belki de insanın dostlarıyla çok zaman geçirmesi kararında alkol tüketmeye ve sağlıklı olmaya yol açar. Sosyal bilimciler buna atlanmış değişkenin yol açtığı sapma adını veriyorlar.

O halde, nedenselliği daha doğru kavramanın yolu nedir? Altın standardı rastgeleleştirilmiş, kontrollü bir deneydir. Nasıl mı oluyor? İnsanları rastgele iki gruba ayırıyorsunuz. Deney grubundan bir şey yapması ya da alması isteniyor. Diğer grup olan kontrol grubundansa istenmiyor. Sonra, her iki grubun da nasıl yanıt verdiğine bakıyorsunuz. İki grubun sonuçları arasındaki fark sizin nedensel etkinizdir.

Örneğin kararında içki içmenin sağlıklı olmaya sebep olup olmadığını test etmek için, rastgele seçtiğiniz kişilerden bir grubun yıl boyunca her gün bir kadeh şarap içmesini, diğerinin de hiç alkol almamasını isteyip, deneyin sonunda iki grubun

sağlık durumunu karşılaştırabilirsiniz. İnsanlar iki gruba rastgele atandığından, gruplardan birinin üyelerinin daha fazla sosyalleşmiş olduğunu ya da sağlığının en başta daha iyi olduğunu düşünmek için bir neden yoktur. Şarabın etkilerinin sebep mahiyetinde olduğuna güvenebilirsiniz. Rastgeleleştirilmiş, kontrollü deneyler her alanda en fazla güvenilen kanıtlardır. Eğer bir hap rastgeleleştirilmiş, kontrollü bir deneyi geçebilirse, halka reçeteyle verilmeye hak kazanabilir. Eğer bu testi geçemezse, eczane raflarında yerini alamaz.

Rastgeleleştirilmiş deneyler sosyal bilimlerde de giderek daha fazla kullanılıyor. MIT'den Fransız iktisatçı Esther Duflo kalkınma iktisadında (dünyadaki en yoksul insanlara yardım etmek için yöntem geliştirmeye çalışan dal) deneylerden daha fazla yararlanılması için yürütülen kampanyaya öncülük ediyor. Duflo'nun meslektaşlarıyla birlikte yürüttüğü şu çalışmayı ele alalım: Ortaokul öğrencilerinin yarısının basit bir cümleyi bile okuyamadığı Hindistan kırsalında eğitim seviyesi nasıl ilerletilebilir? Öğrencilerin bu kadar sıkıntı yaşamalarının potansiyel nedenlerinden biri öğretmenlerin derslere düzenli gelmemeleridir. Hindistan kırsalındaki okulların bazılarında öğretmenlerin yüzde 40'ından fazlasının "okulu kırdığı" görülmektedir.

Duflo'nun testi neydi? Duflo ve meslektaşları okulları rastgele iki gruba ayırdılar. Birinde (deney grubunda), öğretmenlere temel ücretleri haricinde, işe geldikleri her gün için küçük bir miktar (50 rupi, yani 1,15 dolar civarı) ek ödeme yapıldı. Diğer gruba böyle bir ek ödeme yapılmadı. Sonuçlar kayda değerdi. Ek ödeme yapıldığında, öğretmenlerin okulu asma oranları yarı yarıya düşmüştü. Öğrencilerin sınav performansları da ciddi ölçüde düzelmiş, en büyük atılımıysa kız öğrenciler kaydetmişti. Deneyin sonunda öğretmenlere ek ödeme yapılan okullardaki kız öğrencilerin okuma yazma oranı yüzde 7 artmıştı.

New Yorker'daki bir yazıya göre, Bill Gates Duflo'nun çalışmasını duyduğunda o kadar etkilenmiş ki, "Size fon bağlamamız lazım," demiş.

A/B Testinin ABC'si

Rastgeleleştirilmiş deneyler nedenselliği kanıtlamanın altın standardıdır ve bu deneylerden sosyal bilimlerin her alanında yararlanılmaya başlamıştır. Şimdi, 27 Şubat 2000 tarihine ve Google ofisine geri dönüyoruz. Google o gün internet devrimine yol açan ne yapmıştı?

O gün, birkaç mühendis Google'ın sitesinde bir deney yapmaya karar verdi. Kullanıcıları rastgele iki gruba ayırdılar. Deney grubuna arama sonuçları sayfasında yirmi tane bağlantı gösterildi. Kontrol grubunaysa her zamanki gibi on

tane. Ardından, mühendisler iki grubun memnuniyetini Google'a ne kadar sık geri döndüklerine bakarak karşılaştırdılar.

Bu bir devrim midir? O kadar da devrimci görünmüyor. Rastgeleleştirilmiş deneylerin ilaç şirketleri ve sosyal bilimciler tarafından da kullanıldığını belirtmiştim. Onları taklit etmek niye bu kadar abartılıyor ki?

Google mühendislerinin de hemen fark ettikleri kilit husus dijital dünyadaki deneylerin çevrimdışı dünyadakilere kıyasla muazzam bir üstünlüğe sahip olmasıdır. Fiziki dünyada uygulanan rastgeleleştirilmiş deneyler ikna edici olabildiği gibi, aynı zamanda kaynaklara yüklenerek yapılan deneylerdir. Duflo'nun çalışması için okullarla temas kurmak, fon işini ayarlamak, bazı öğretmenlere ödeme yapmak ve tüm öğrencileri sınava sokmak gerekmişti. Bu tarz deneyler binlerce, yüz binlerce dolara mal olabilir ve yürütülmesi aylar, yıllar alabilir.

Dijital dünyadaysa rastgeleleştirilmiş deneyler ucuz ve hızlı olabiliyor. Katılımcılar bulup onlara ödeme yapmanıza gerek yok. Bunun yerine, bir satır kod yazıp bunları rastgele bir gruba atayabilirsiniz. Kullanıcılara anket formu doldurtmanıza gerek yok. Bunun yerine, bilgisayar faresinin hareketlerini ve tıklamalarını ölçebilirsiniz. Elle kodlama yapıp yanıtları incelemenize gerek yok. Geliştirdiğiniz bir programla bunu otomatik olarak yapabilirsiniz. Kimseyle temas kurmanıza gerek yok. Hatta kullanıcılara bir deneyin parçası olduklarını söylemenize bile gerek yoktur.

Büyük Veri'nin dördüncü güçlü yanı da budur: Gerçekten nedensel etkileri saptayabilen rastgeleleştirilmiş deneyler yapar ve bu deneyleri yapmak çok çok daha kolaydır; internete bağlı olduğunuz sürece her zaman, aşağı yukarı her yerde yapabilirsiniz. Büyük Veri çağında tüm dünya bir laboratuvardır.

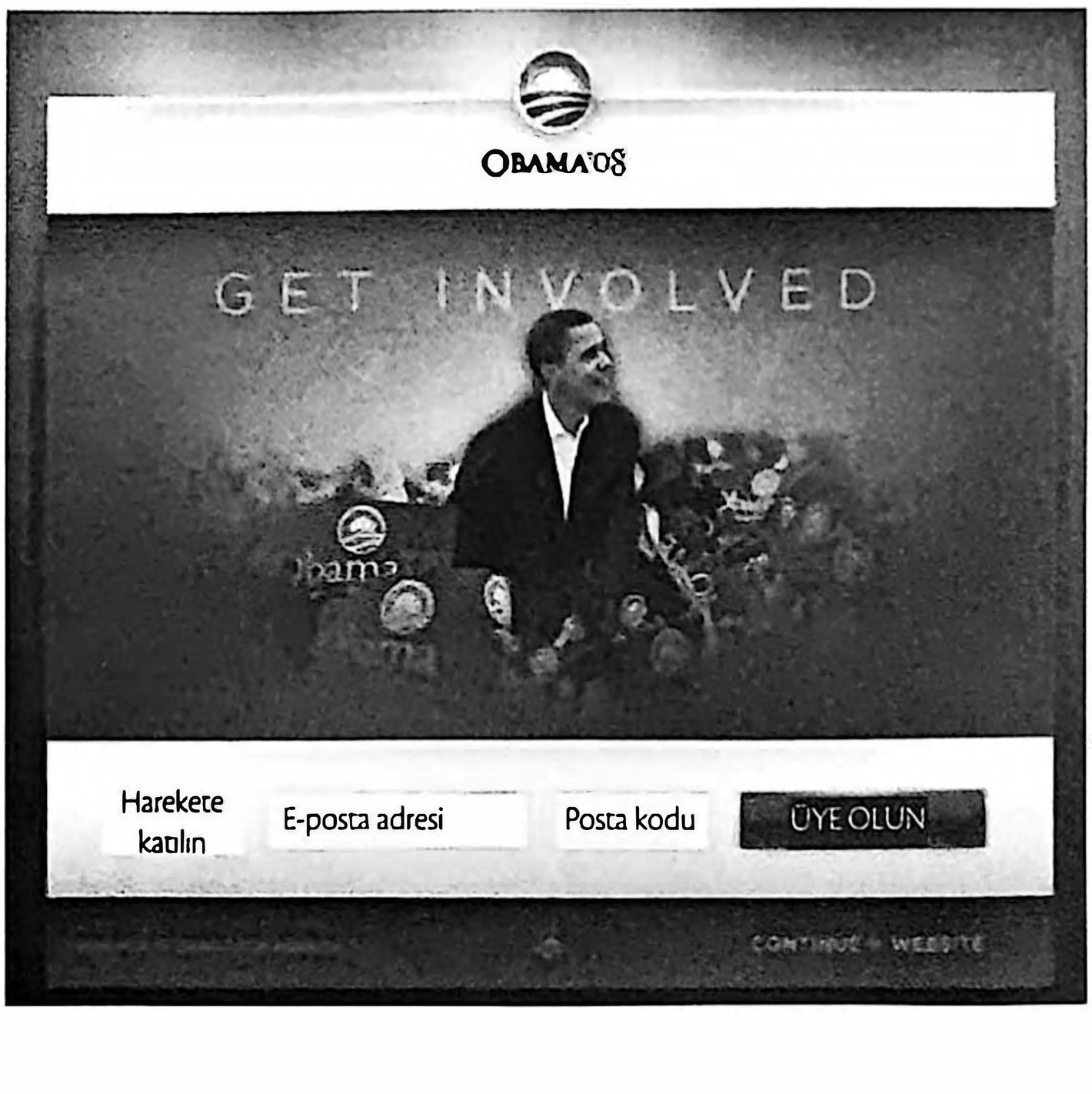
Bu içgörü Google'da ve ardından Silikon Vadisi'nin geri kalanında hızla yayıldı ve rastgeleleştirilmiş kontrollü deneylere bundan böyle "A/B testi" adı verildi. Google mühendisleri yalnızca 2011'de yedi bin A/B testi yaptılar ve bu sayı gün geçtikçe artıyor.

Google, sitedeki reklamlara daha fazla insanın tıklamasını nasıl sağlayabileceğini öğrenmek için, reklamlarda mavinin iki tonunu deneyebiliyor: A grubu için bir tonu, B grubu için başka bir tonu. Ardından, tıklama oranlarını karşılaştırıyor. Elbette bu tür testlerin kolaylığı aşırı kullanıma yol açabilir. Kimi çalışanlar, testlerin uygulanması fazlasıyla zahmetsiz olduğundan, Google'ın haddinden fazla deney yaptığı sonucuna varmıştı. 2009'da bir tasarımcı, Google A/B testinde mavinin 41 farklı çeşidini denedikten sonra işi bırakmıştı. Ama bu tasarımcının takıntılı bir pazar araştırması yerine sanattan yana tavır alması metodolojinin yayılmasını

Bugün Facebook günde bin tane A/B testi yapıyor; bu, Facebook'un küçük araştırmacı grubunun tek bir günde tüm ilaç sanayisinin bir yılda yaptığından daha çok rastgeleleştirilmiş, kontrollü deney yaptığı anlamına gelir.

A/B testi dev teknoloji şirketleriyle sınırlı kalmamıştır. Google'ın eski çalışanlarından Dan Siroker bu metodolojiyi Barack Obama'nın ilk seçim kampanyasında kullanmıştı: Ana sayfa tasarımları, e-posta aralıkları ve bağış formları A/B testinden geçirilmişti. Akabinde Siroker kurumlara hızlı A/B testi yapma imkânı veren Optimizely adında yeni bir şirket kurdu. 2012'de Optimizely hem Obama hem de rakibi Mitt Romney tarafından üyeleri, gönüllüleri ve bağışları azamiye çıkarmak için kullanılıyordu. Bu yöntem Netflix, TaskRabbit ve *New York* dergisi gibi farklı şirketler tarafından da kullanılmaktadır.

Bu testin ne kadar değerli olduğunu, Obama'nın kampanyasına daha fazla insanı katmak amacıyla bu testten nasıl yararlandığına bakarak anlayabiliriz. Obama'nın ana sayfasında adayın bir resmi, altında da insanları "kaydol"maya davet eden bir buton vardı.



İnsanların siteye girdiklerinde görecekleri ilk şey bunlar mı olmalıydı? Obama'nın ekibi Siroker'ın yardımıyla farklı bir resim ve butonun insanları gerçekten kaydolmaya teşvik edip edemeyeceğini test ettiler. Ana sayfada daha ciddi suratlı bir Obama resmi olsa daha fazla insan tıklar mıydı? Butonda "Şimdi Katılın" yazsa daha fazla insan bir sonraki adıma geçer miydi? Obama'nın ekibi kullanıcılara resimlerin ve butonların farklı bileşimlerini gösterip, kaç tanesinin butona tıkladı-

ğını ölçmüştü. Bir bakın bakalım, hangi resmin ve hangi butonun kazandığını bulabiliyor musunuz?

Test Edilen Resimler







Test Edilen Butonlar

BİZE HEMEN KATILIN

DAHA FAZLA BİLGİ

ÜYE OLUN

Kazanan, Obama'nın ailesinin yer aldığı resim ile "Daha Fazla Bilgi" yazan buton oldu. Ve dev bir zafer elde edilmişti. Obama'nın kampanyası bu bileşimi kullanarak, tahminlere göre, katılanların sayısının yüzde 40 oranında artmasını sağlamıştı ki, bu da kampanyaya kabaca 60 milyon dolar ek kaynak anlamına geliyordu.

Kazanan İkili



Bu altın standart testinin çok ucuza ve kolayca yapılabilmesi haricinde çok önemli bir özelliği daha var: Bizi, Birinci Bölüm'de sorunlu yanları olduğu belirtilen bir tutumdan, yani sezgilerimize bel bağlamaktan biraz daha kurtarıyor. A/B testinin önemli olmasının temel nedenlerinden biri, insanların kestirilemez olmasıdır. Sezgilerimiz insanların nasıl yanıt vereceğini çoğu zaman kestiremiyor.

Siz söyleyin, sezgileriniz Obama'nın en iyi internet sitesinin hangisi olduğunu bilebildi mi?

İşte size sezgilerinizi sınamak için birkaç test daha. *The Boston Globe*'un uygulağı A/B testleri hangi haber başlıklarının okurun tıklamasını sağlayacağını tespit etmeyi hedefliyor. Şu çiftlerden hangisinin kazandığını tahmin etmeye çalışın:

Tahminimce yarısından çoğunu bildiniz, muhtemelen "Ben olsaydım hangisine tıklardım acaba?" diye düşünerek. Ama büyük ihtimalle hepsini doğru tahmin edemediniz.

Neden? Neyi kaçırdınız? İnsan davranışlarına dair hangi içgörülerden yoksundunuz? Bu yanlışlarınızdan ne tür dersler çıkarabilirsiniz?

Yanlış tahminlerde bulunduktan sonra genellikle bu tür sorular sorarız.

Ama *Globe* haber başlıklarından genel sonuçlar çıkarmanın ne kadar zor olduğunu da unutmayalım. İlk haber başlığı testinde, tek bir kelimeyi değiştirmek ("bu"

Şu Haber Başlıklarından Biri Tık Alma Konusunda
Diğerinden Çok Daha İyiydi.

	HABER BAŞLIĞI A	HABER BAŞLIĞI B
1.	SnotBot drone'u balinaları kurtarabilir mi?	Bu drone balinaları kurtarmaya yardımcı olabilir mi?
2.	Elbette "patlak toplar" Massachusetts'te en fazla aranan kelimelerdir	Massachusetts'te en fazla Google'lanan yüz kızartıcı kelime
3.	St. Paul tecavüz davasının altından yiyişme yarışı çıktı	Özel okuldaki seks skandalında ceza çıkmadı
4.	Yaşlı kadın nadir bulunan bir beyzbol kartıyla voliyi vurdu	Yaşlı kadın nadir bulunan bir beyzbol kartından 179 bin $ kazandı
5.	MBTA yıllık işlem açığının 2020'de ikiye katlanacağını öngörüyor	Hazır olun: MBTA'nın açığı ikiye katlanmak üzere
6.	Massachusetts doğum kontrol hakkı elde etmenize nasıl yardımcı oldu?	Boston Üniversitesi "iffete saldırı"nın sonlandırılmasında nasıl bir rol oynadı?
7.	Boston'da ilk metronun açıldığı gün	Boston'da ilk metro açıldığında yayınlanan karikatürler
8.	Özel okuldaki tecavüz davasında kurban ve ailesi kültürel çürümeyi suçluyor	Özel okuldaki tecavüz davasının kurbanı ve ailesinden açıklama
9.	Kılık değiştirmiş Miley Cyrus'u tanıyan tek kişi "Free Brady" şapkalı adam oldu	Pats taraftarı kılık değiştirmiş Miley Cyrus'u tanımayı başardı

Tahminlerinizi yaptınız mı? Cevaplar şöyle:

	HABER BAŞLIĞI A	HABER BAŞLIĞI B	KAZANAN?
1.	Kazanan		% 53 daha fazla tık
2.		Kazanan	% 986 daha fazla tık

	HABER BAŞLIĞI A	HABER BAŞLIĞI B	KAZANAN?
3.		Kazanan	% 108 daha fazla tık
4.	Kazanan		% 38 daha fazla tık
5.		Kazanan	% 62 daha fazla tık
6.		Kazanan	% 188 daha fazla tık
7.	Kazanan		% 33 daha fazla tık
8.		Kazanan	% 76 daha fazla tık
9.		Kazanan	% 67 daha fazla tık

yerine "SnotBot") büyük bir başarı getirmiş. Buradan, daha fazla ayrıntının kazandıracağı sonucunu çıkarabilirsiniz. Ama ikinci haber başlığında, "patlak toplar", yani ayrıntı sunan kelimeler kaybeden olmuş. Dördüncü haber başlığında, "voliyi vurdu" lafı "179 bin $"ı alt etmiş. Buradan da argo tabirlerin kazanacağı sonucu çıkarılabilir. Ama üçüncü başlıktaki "yiyişme yarışı" ifadesi kaybetmiş, ona ne diyeceksiniz?

A/B testinden çıkarılacak en önemli ders, herhalde, genel dersler çıkarmaya temkinli yaklaşmak gerektiğidir. Clark Benson ranker.com adında, haber başlıklarını ve site tasarımını seçmek için büyük oranda A/B testinden yararlanan bir haber ve eğlence sitesinin CEO'sudur. "Nihayetinde, varsayımlarla bir yere varamazsınız. Gerçekten her şeyi test edin," diyor Benson.

Test etmek insan doğasına ilişkin anlayışımızdaki boşlukları dolduruyor. Bu boşluklar her zaman var olacaktır. Eğer kendi yaşam deneyimlerimiz temelinde cevabın ne olacağını bilebilseydik, elbette testlerin bir kıymeti olmazdı. Ama bilemeyiz, bu yüzden testler değerlidir.

A/B testlerinin çok önemli olmasının bir diğer nedeni de görünüşte küçük değişikliklerin büyük sonuçlar doğurabilmesidir. Benson'ın ifade ettiği şekliyle, "Küçük, çok küçük etkenlerin testlerde boyundan büyük sonuçlara yol açabilmesine her zaman şaşırmışımdır."

Google, Aralık 2012'de reklamlarını değiştirdi. Reklamlara bir kutucuğun içinde sağa işaret eden bir ok ekledi.



Bu okun ne kadar tuhaf olduğuna bir bakın. Sağa dönük, ama kesinlikle hiçbir şeye işaret etmiyor. Aslında bu oklar ilk çıktığında, Google'ın birçok müşterisi tarafından eleştirilmişti. Reklamlara neden anlamsız oklar ekliyorlardı ki?

Ortalama Deneyi (Tutmadı)

Best Selling iPad 2 Case
The ZAGGmate™ - Tough Aluminum Case
with build in Bluetooth Keyboard
www.zagg.com

Yeşil Yıldız Deneyi (Tutmadı)

Foster's Hollywood Restaurant **Reviews**, Madrid, Spain ...
www.tripadvisor.co.uk > ... > **Madrid** > **Madrid Restaurants** ▾ TripAdvisor ▾
★★★★★ Rating: 3 - 118 reviews
Foster's Hollywood, Madrid: See 118 unbiased **reviews** of **Foster's Hollywood**, rated 3 of 5 on TripAdvisor and ranked #3647 of 6489 restaurants in Madrid

Yeni Yazı Tipi Deneyi (Tutmadı)

Live Stock Market News
Free Charts, News and Tips from UTVi Experts. Visit us Today!
UTVi.com/Stocks

Google şirket sırlarını saklı tuttuğundan, okların tam olarak ne kadar kazandırdığını bilmiyoruz. Ama bu okların A/B testinden geçtiğini biliyoruz. Google'ın bu okları eklemesinin nedeni, çok daha fazla insandan tık almayı başarmasıydı. Bu küçük, görünüşte anlamsız değişiklik Google'a ve reklam ortaklarına dünyanın parasını kazandırdı.

O halde, muazzam kârlar getiren bu küçük değişiklikleri bulmayı nasıl başarabiliriz? Birçok şeyi, hatta entipüften görünen birçok şeyi test etmeniz gerekir. Google kullanıcıları reklamlarda birçok kez küçük değişiklikler yapıldığını, ama ardından reklamların önceki hallerine geri döndürüldüğünü belirtmişlerdir. Kullanıcılar, farkında olmadan, A/B testlerindeki denek gruplarında yer almışlardır; gerçi bunun tek bedeli küçük çeşitlilikleri görmek olmuştur.

Yukarıdaki değişiklikler hiçbir zaman geniş kitlelere ulaşmadı. Kaybolup gitti. Ama hepsi de kazananı bulma sürecinin parçasıydı. Tıklanabilir bir oka giden yol çirkin yıldızlarla, yanlış konumlandırmalarla ve reklam hilelerine başvuran yazı türleriyle örülüdür.

* * *

İnsanların tıklamasını sağlayan şeyin ne olduğunu tahmin etmek eğlenceli olabilir. Eğer Demokrat partiliyseniz test yapmanın Obama'ya daha fazla para kazandırdığını bilmek hoş olabilir. Ama A/B testinin bir de karanlık yanı var.

Adam Alter *Irresistible* adlı mükemmel kitabında, günümüz toplumunda davranışsal bağımlılıkların yükselişinden bahseder. Birçok insan için internetin birçok yönünden vazgeçmek giderek daha da zorlaşıyor.

Benim en sevdiğim veri kümesi olan Google aramaları bize insanların neyi en fazla bağımlılık yaratıcı bulduğuna dair bazı ipuçları verebilir. Google'a göre, birçok bağımlılık hâlâ insanların uzun yıllardır mücadele ettikleri şeylerdir; ilaçlar, seks ve alkol gibi. Ama internet de listede varlığını hissettirmeye başlıyor: "Porno" ve "Facebook" artık bağımlılık listesinde ilk onda yer alıyor.

Google'a Göre İlk 10 Bağımlılık, 2016

İlaçlar	Alkol	Kumar
Seks	Şeker	Facebook
Porno	Aşk	

A/B testi internetin böyle lanet bir bağımlılık yaratmasında rol oynuyor olabilir.

"Tasarım etiği" alanında çalışan Tristan Harris'in *Irresistible*'da aktarılan sözleri, insanların internette belli sitelere direnmekte neden zorlandıklarını açıklar: "Ekranın diğer tarafında, işleri sizin kendinizi denetleme çabanızı bozmak olan yüzlerce insan bulunuyor."

Ve bu insanlar A/B testinden yararlanıyorlar.

Facebook bu testler yoluyla belli bir butonu belli bir renkte yapmanın insanların o siteye daha sık geri dönmelerini sağladığını anlayabiliyor. Böylece buton o renge çevriliyor. Ardından belli bir yazı tipinin insanların siteye daha sık geri dönmelerini sağladığını anlayabiliyorlar. Böylece metin o yazı tipine çevriliyor. Ardından insanlara belli aralıklarla e-posta yollamanın siteye daha sık geri dönmelerini sağladığını anlayabiliyorlar. Böylece o doğru anda insanlara e-posta yolluyorlar.

Facebook kısa sürede insanların geçirdikleri zamanı azamiye çıkaracak şekilde en fazla verim alınan sitelerden biri haline gelmiştir. Başka bir deyişle, yeterli sayıda A/B testinde en etkili araçları bulduğunuz takdirde, bağımlılık yaratan bir siteye kavuşursunuz. Sigara şirketlerinin asla sahip olmadığı türde bir geri bildirimdir bu.

A/B testi gitgide oyun endüstrisinin de parçası haline geliyor. Alter'ın belirttiği gibi, World of Warcraft, oyunun çeşitli versiyonlarıyla ilgili A/B testi yapıyor. Görevlerden biri, birini öldürmenizi istemek olabilir. Bir diğerinde sizden bir şeyi korumanız istenebilir. Oyun tasarımcıları oyuncuların farklı görevlerine dair farklı örneklemler sunabilir ve ardından hangilerinin insanların daha fazla oynamasını sağladığını görebilirler. Örneğin sizden bir insanı kurtarmanızı isteyen görevin insanları yüzde 30 oranında daha fazla geri döndürdüğünü öğrenebilirler. Eğer birçok ama birçok görevi test ederlerse, giderek daha çok sayıda etkili aracı tespit ederler. Bu yüzde 30'luk oran giderek artar ve en sonunda birçok yetişkin erkeği ebeveynlerinin bodrum katında gizli kapaklı oyun oynamaya iten bir oyun ortaya çıkar.

Eğer bundan rahatsız olduysanız, ben de hislerinizi paylaştığımı söylemeliyim. Kitabın sonlarına doğru bunun ve Büyük Veri'nin diğer veçhelerinin etik içerimlerinden biraz daha bahsedeceğiz. Ama iyisiyle kötüsüyle, deney yapmak artık veri bilimcilerinin alet çantasının vazgeçilmez bir parçasıdır. Dahası bu alet çantasında başka bir deney biçimi daha bulunuyor. Bu deney, insanlara çeşitli sorular sormak için kullanılıyor; televizyon reklamlarının gerçekten işe yarayıp yaramadığı sorusu da bu teste dahildir.

Doğanın Acımasız Ama Aydınlatıcı Deneyleri

Tarihler 22 Ocak 2012'yi gösteriyor. New England Patriots, AFC şampiyonluk maçında Baltimore Ravens ile karşı karşıya.

Maçın son bir dakikası. Ravens geride, ama top onlarda. Bu son altmış saniye Super Bowl'da hangi takımın yer alacağını belirleyecek. Son altmış saniye oyuncuların gelecekte efsane olmalarında rol oynayacak. Ve maçın bu son bir dakikası, bir iktisatçı için çok daha derin anlamları olan bir şeyi gerçekleştirecek: Bu son altmış saniye nihayet bize reklamların işe yarayıp yaramayacağını kesin olarak söyleyecek.

Reklamların satışları artırdığı düşüncesi ekonomimiz açısından kuşkusuz hayati bir role sahiptir. Ama bunu kanıtlamak insanı delirtecek kadar zordur. Aslında bağıntı ile nedensellik arasında ayrım yapmanın tam olarak ne kadar zor olduğuna dair de ders kitaplarında yer alacak bir örnektir.

En fazla reklamı yapılan ürünlerin en çok satışı yaptığına kuşku yoktur. 20th Century Fox *Avatar* filmini pazarlamak için 150 milyon dolar harcadı ve *Avatar* gelmiş geçmiş en yüksek gişe hasılatına ulaştı. Peki, *Avatar*'ın 2,7 milyar dolarlık gişe hasılatının ne kadarı bu yoğun pazarlama sayesindeydi? 20th Century Fox'un reklama bu kadar para harcamasının bir sebebi de muhtemelen arzulanan bir ürüne sahip olduğunu bilmesiydi.

Şirketler reklamlarının ne kadar etkili olduğunu bildiklerine inanırlar. İktisatçılar ise gerçekten bildiklerine şüpheyle yaklaşırlar. Chicago Üniversitesi iktisat profesörü Steven Levitt bir elektronik şirketiyle çalışırken, şirketin "Reklamlarımızın ne kadar işe yaradığını biliyoruz!" tavrından ötürü epey sıkıntı yaşamıştı. Levitt, "Bu özgüvenin kaynağı ne?" diye merak ediyordu.

Şirketin açıklaması, her yıl Babalar Günü'nden önce televizyon reklamlarına harcadıkları reklamları artırdıkları yönündeydi. Kuşkusuz her yıl Babalar Günü'nden önceki günler en fazla satış yaptıkları zamandır. Ah, ama belki de bunun nedeni birçok çocuğun, özellikle de Babalar Günü'nde, reklamlardan bağımsız olarak elektronik eşya satın almasıdır.

"Sebep-sonuç ilişkisini tamamen tersine çevirmişler," diyor Levitt bir konferansında. Belki de öyledir. Bilmiyoruz. "Gerçekten zor bir problem," diye ekliyor.

Bu problemi çözmek zor olmasına karşın, şirketler sıkı deneyler yapmaya yaşmıyorlar. Levitt bu elektronik şirketini televizyon reklamlarının ne kadar etkili olduğunu anlamak için rastgeleleştirilmiş, kontrollü bir deney yapmaya ikna etmeye çalışmıştı. A/B testlerini televizyonda yapmak henüz mümkün olmadığından, bazı alanlarda reklam verilmediğinde ne olduğuna bakmak gerekiyordu.

Şirket şöyle yanıt vermişti: "Deli misin? Yirmi ayrı pazarda reklam vermememiz mümkün değil. CEO bizi öldürür." Böylece Levitt bu şirketle çalışmayı bıraktı.

Şimdi Patriots-Ravens maçına dönelim. Bir Amerikan futbolu maçının sonuçları reklamcılıktaki sebep-sonuç ilişkisini belirlememize nasıl yardımcı olabilir?

Doğrusu, belli bir şirketin belli bir reklam kampanyasının sonuçlarını bu yolla öğrenemeyiz. Ama birçok büyük şirketin reklamlarının ortalama sonuçlarına dair kanıtlar bulabiliriz.

Bunun gibi maçlarda gizli yürütülen bir reklam deneyi olduğu anlaşılıyor. Deney şöyle işliyor: Bu şampiyonluk maçları oynandığı vakit, şirketler Super Bowl reklamlarını satın almış ve üretmiş oluyorlar. İşletmeler hangi reklamları oynatacaklarına karar verdiklerinde, maçta hangi takımların karşı karşıya geleceğini bilmiyor oluyorlar.

Oysa *playoff*'ların sonuçları Super Bowl'u kimlerin izleyeceğine devasa bir etkide bulunacaktır. Nihayetinde bu finale kalacak olan iki takım muazzam sayıda izleyici getirecektir. Eğer maçlarını Boston yakınlarında oynayan New England kazanırsa, Super Bowl'u Boston'dan izleyenlerin sayısı Baltimore'dakinden çok daha fazla olacaktır. Keza tersi de geçerlidir.

Şirketler açısından Baltimore'da ya da Boston'da binlerce fazladan insanın reklamlarını izlemesini neredeyse yazı-turaya denk bir şey belirlemektedir, üstelik para havaya, yayın süreleri satın alınıp reklamlar çekildikten sonra atılmaktadır.

Şimdi sahaya geri dönelim. CBS'ten Jim Nantz mikrofonda, bu deneyin nihai sonuçlarını açıklıyor.

Billy Cundiff skoru eşitlemek için sahneye çıkıyor. Büyük ihtimalle maç uzatmaya gidecek. Son iki yılda saha içi şutlarında on altıda on altı yaptı. Otuz iki yardlık şutla beraberlik şansı. Vuruş! Veeee... İsabetsiz! Patriots, Indianapolis yolcusu. 46. Super Bowl'da oynamaya hak kazanan Patriots oluyor.

İki hafta sonra Super Bowl XLVI, Boston'da 60,3 seyirci payına, Baltimore'daysa 50,2 paya sahip olacaktı. 2012 reklamlarını Boston'dan 60 bin daha fazla insan seyredecekti.

Ertesi yıl aynı iki takım AFC şampiyonluğu için bir kez daha karşı karşıya geldi. Bu kez kazanan Baltimore oldu. 2013 Super Bowl reklamlarını Baltimore'dan izleyenlerin sayısı daha yüksekti.

	2012 SUPER BOWL REYTİNGLERİ (BOSTON OYNADI)	2013 SUPER BOWL REYTİNGLERİ (BALTIMORE OYNADI)
Boston	56,7	48,0
Baltimore	47,9	59,6

Google'ın baş iktisatçısı Hal Varian, Carnegie Mellon'dan iktisatçı Michael D. Smith ve ben bu iki maçtan ve 2004-2013 arasındaki diğer Super Bowl maçlarından yararlanarak, Super Bowl reklamlarının işe yarayıp yaramadığını ve yarıyorsa ne

kadar yaradığını test ettik. Özel olarak, bir yapım şirketi Super Bowl'da bir film için reklam verdiğinde o maçın izlenme oranlarının yüksek olduğu şehirlerde bilet satışlarında büyük bir artış olup olmadığına baktık.

Gerçekten de artış oluyordu. Super Bowl'a katılmaya hak kazanan takımların yer aldığı şehirlerdeki insanlar Super Bowl sırasında reklamı yapılan filmlere, katılamayan takımların yer aldığı şehirlerdeki insanlardan çok daha fazla gidiyorlar. Bu şehirlerde daha fazla insan reklamları görüyordu. Bu şehirlerde daha fazla insan filme gitmeye karar veriyordu.

Alternatif açıklamalardan birine göre, eğer takımınız Super Bowl'da oynayacaksa, filme gitmeniz daha muhtemeldir. Buna karşılık, Super Bowl takımlarının yer aldığı şehirlerde benzer bütçeleri olan ve benzer zamanlarda gösterime giren ama Super Bowl'da reklamı yapılmayan bazı filmlerin seyirci sayısında artış yaşanmadığını tespit ettik.

Pekâlâ, tahmin edebileceğiniz üzere, reklamlar işe yarıyor. Bu pek de şaşırtıcı değil.

Ama mesele sadece işe yaramaları değil. Reklamlar acayip etkili oluyordu. Aslında sonuçları ilk gördüğümüzde, doğru olup olmadığını iki kez, üç kez, dört kez kontrol ettik, çünkü çok yüksek bir kazanç vardı. Örneklemimizdeki ortalama bir film Super Bowl reklamlarına yaklaşık 3 milyon dolar ödemiş, karşılığında bilet satışlarından 8,3 milyon dolar elde etmişti: 1'e 2,8 veren bir yatırım.

Bu sonucu doğrulayan bağımsız araştırmacılar Wesley R. Hartman ile Daniel Klapper da benzer bir görüş ileri sürmüştü. Bu iktisatçılar Super Bowl sırasında verilen bira ve meşrubat reklamlarını incelemiş ve aynı zamanda finale kalan takımların bulunduğu şehirlerde reklamı izleyenlerin sayısındaki artıştan yararlanmışlardı. 1'e 2,5'lik bir kazanç oranı tespit etmişlerdi. Her iki araştırmanın sonuçlarına bakarak, Super Bowl reklamlarının pahalı olmakla birlikte, talebi artırmakta çok başarılı olduğunu ve dolayısıyla şirketlerin aslında ederinden daha az ödeme yaptığını söyleyebiliriz.

Peki, tüm bunlar Levitt'in çalıştığı elektronik şirketindeki arkadaşlarımız açısından ne anlam ifade ediyor? Super Bowl reklamlarının diğer reklam biçimlerine kıyasla daha düşük maliyetli olması mümkündür. Ama en azından bizim çalışmamız Babalar Günü reklamlarının muhtemelen iyi bir fikir olduğunu söylüyor.

* * *

Süper Bowl deneyinin olumlu yanlarından biri, kimsenin kasıtlı olarak deney grubuna ya da kontrol grubuna seçilmesini gerektirmemesiydi. Bir futbol ma-

çındaki topun isabetine dayalıydı. Başka bir deyişle, doğal olarak gerçekleşmişti. Bunu neden üstünlük sayıyoruz? Doğal olmayan, rastgele kontrollü deneyler çok etkilidir ve dijital çağda gerçekleştirilmesi daha kolaydır, ama yine de her zaman mümkün değildir.

Kimi zaman eylemlerimizi zamansal açıdan ortaklaştıramayız. Kimi zaman reklam kampanyasında deney yapmak istemeyen elektronik şirketinde olduğu gibi, bunu test edemeyecek kadar fazla yatırım yapmış oluruz.

Kimi zaman deneylerin uygulanması olanaksızdır. Bir ülkenin liderini kaybettiği takdirde nasıl yanıt vereceğini öğrenmek istediğimizi varsayalım. Savaşa mı gider? Ekonomisi durma noktasına mı gelir? Fazla bir değişiklik olmaz mı? Malumunuz, kayda değer sayıda devlet başkanını ve başbakanı öldürüp ne olduğuna bakmamız mümkün değildir. Bu yalnızca olanaksız değil, etik açıdan da yanlış bir tutum olacaktır. Üniversiteler önerilen bir deneyin etik olup olmadığını belirlemek için kurumsal değerlendirme kurulları (IRB) oluşturmuşlardır.

Dolayısıyla belli bir senaryodaki nedensel etkileri bilmek istiyorsak ve bir deneyi yapmak etik değilse ya da başka açılardan uygun değilse, ne yapabiliriz? Deneyin doğal alanını Amerikan futbolu maçlarını kapsayacak kadar geniş bir şekilde tanımlayarak, iktisatçıların doğal deneyler adını verdikleri şeylerden yararlanabiliriz.

İyisiyle kötüsüyle (tamam kabul, kötü olduğu açık), yaşam muazzam bir rastgelelik sergiler. Evrenden neyin ya da kimin sorumlu olduğunu hiç kimse kesin olarak bilmez. Ama bir şey açıktır: İşin başında kim varsa (kuantum mekaniği yasaları mı, Tanrı mı, bilgisayarında evren simülasyonu yapan yaramaz bir pijamalı çocuk mu, artık her neyse ve kimse) onlar ya da o, IRB'den icazet almıyor.

Doğa her an bizim üzerimizde deneyler yapıyor. İki insan vuruluyor. Mermilerden biri hayati organlarından birini sıyırıyor. Diğeri sıyırmıyor. Bu kötü hadiseler hayatı adaletsiz kılan şeylerdir. Teselli olur mu bilmem ama kötü hadiseler, hayatı, iktisatçıların çalışma yürütmeleri bakımından biraz daha kolay hale getiriyor. İktisatçılar hayatın keyfiliğinden, nedensel etkileri bulmak için yararlanırlar.

43 ABD başkanından 16'sı ciddi suikast girişimine maruz kalmış, 4'ü öldürülmüştür. Bazılarının hayatta kalmış olması tamamen şans eseridir.

John F. Kennedy ile Ronald Reagan'ı karşılaştırın. İki başkana da vücutlarının en hayati bölgelerini hedef alan silahlı saldırılar düzenlenmiştir. JFK'in yediği mermi beynini parçalamış, kısa sürede ölümüne yol açmıştır. Reagan'a isabet eden kurşun ise kalbini birkaç santimetre ile sıyırmış, bu sayede doktorlar hayatını kurtarabilmiştir. Reagan hayatta kalmış, JFK ölmüştür, ne bir nedeni var ne de başka bir şey, sırf şans.

Liderlerin hayatına kastetme girişimleri ve hayatta kalıp kalmayacaklarının belirsizliği tüm dünyada görülen bir şeydir. Çeçenistanlı Ahmet Kadirov ile Alman Adolf Hitler'i karşılaştırın. İkisi de kusursuz şekilde yerleştirilmiş bombalardan birkaç metre ötedeydi. Kadirov öldü. Hitler ise takvimini değiştirmesi sayesinde treni yakalamak için bubi tuzaklı odadan birkaç dakika erken ayrıldı ve hayatta kaldı.

Doğanın soğukkanlı keyfiliğinden (Kennedy'yi öldürüp Reagan'ı öldürmemesinden) yararlanarak, bir ülkenin lideri suikasta uğradığında ne olduğunu kabaca görebiliriz. Benjamin F. Jones ve Benjamin A. Olken adlı iki iktisatçı tam da bunu yaptılar. Burada kontrol grubu, birkaç sene önce kıl payı atlatılan bir suikastın yaşandığı herhangi bir ülke, örneğin 1980'lerin ortalarındaki ABD. Denek grubu ise, başarıya ulaşmış bir suikastın yaşandığı bir ülkedir, örneğin 1960'ların ortalarındaki ABD.

Peki, bir ülkenin liderinin öldürülmesinin sonuçları nelerdir? Jones ve Olken başarılı suikastların dünya tarihini çarpıcı şekilde değiştirdiğini, ülkeleri alabildiğine farklı yollara sokmuğunu tespit ettiler. Yeni bir lider, öncesinde barışçıl olan ülkeleri savaşa, öncesinde savaşan ülkeleri de barışa götürüyordu. Yeni bir lider ekonomik açıdan yükselişteki ülkeleri dibe vurdururken, dibe çakılmakta olan ülkeleri ekonomik açıdan yükselişe geçiriyordu.

Aslında bu suikast temelli doğal deneyin sonuçları ülkelerin nasıl ayakta kaldığına dair geleneksel düşünüşün çökmesine yol açmıştır. Birçok iktisatçı, liderlerin dış kuvvetlerin kölesi haline gelmiş güçsüz kuklalar olduğu görüşüne yakın duruyordu. Jones ve Olken'ın doğal deney analizleri böyle olmadığını gösteriyor.

Birçok insan dünya liderlerine yönelik suikast girişimlerine dair bu deneyi Büyük Veri'nin bir örneği olarak görmeyecektir. Çalışmanın dikkate aldığı suikasta uğrayan ya da suikasttan kıl payı kurtulan lider sayısı çok azdı, tıpkı akabinde patlak veren ya da vermeyen savaş sayısı gibi. Bir ekonominin yörüngesini nitelendirmek için zorunlu olan ekonomik veri kümeleri büyüktü, ama büyük oranda dijital çağ öncesine aitti.

Gelgelelim bu tür doğal deneyler –bugün artık neredeyse sadece iktisatçılar tarafından kullanılsa da– güçlüdür ve daha fazla, daha iyi ve daha büyük veri kümelerinin olduğu bir çağda giderek daha fazla önem kazanacaktır. Veri bilimlerinin uzun süre uzak kalamayacakları bir araçtır.

Ve evet, buraya kadar anlamış olacağınız gibi, iktisatçılar veri biliminin gelişiminde çok büyük bir rol oynuyorlar. En azından ben böyle düşünmek istiyorum, zira bu rahlei tedristen geçtim.

* * *

Doğal deneyleri, başka bir deyişle olayların rastgele seyrinin insanları denek ve kontrol grupları şeklinde ayrıştırdığı durumları başka nerede bulabiliriz?

En bariz örnek piyangodur, tam da bu nedenle iktisatçılar piyangoları severler: Yanlış anlamayın, oynamayı değil, zira bize göre akıldışıdır, sadece çalışmayı. Eğer üzerinde üç olan bir pinpon topu tepeye çıkarsa, Bay Jones zengin olacaktır. Eğer onun yerine altı gelirse, Bay Johnson zengin olacaktır.

Kişinin başına talih kuşu konmasıyla gelen paraların nedensel etkilerini test etmek isteyen iktisatçılar, piyango kazanan kişileri bilet alan ama kazanamayan kişilerle karşılaştırırlar. Bu çalışmalar genellikle piyango kazanmanın insanı kısa vadede mutlu etmediğini, ama uzun vadede gerçekten mutlu kıldığını gösteriyor.*

İktisatçılar piyangoların rastgeleliğinden yararlanarak, komşusuna piyango çıkan birinin hayatının nasıl değiştiğini de inceleyebilirler. Veriler piyango kazanan komşunuzun sizin hayatınıza etkisi olabileceğini gösteriyor. Eğer komşunuz piyangodan büyük ikramiye kazanırsa, siz de büyük ihtimalle pahalı bir araba, örneğin bir BMW alıyorsunuz. Neden? İktisatçıların iddiasına göre, bunun sebebi zengin komşunuz pahalı bir araba alınca kıskançlık krizine girmenizdir. İnsan doğasının bileşenlerine eklensin lütfen: Eğer Bay Johnson Bay Jones'un son model bir BMW kullandığını görürse, Bay Johnson da aynısından almak isteyecektir.

Ne yazık ki Bay Johnson'ın BMW alacak parası yoktur, işte bu nedenle iktisatçılar piyango kazanan kişilerin komşularının iflas etme ihtimalinin çok yüksek olduğunu tespit etmişlerdir. Bu noktada Jones'larla aşık atmaya çalışmak imkânsızdır.

Ama doğal deneylerin piyangolar gibi açıkça rastgele olması gerekmiyor. Rastgeleliği aradığınız takdirde her yerde görmeye başlar, ardından bunu dünyanın nasıl işlediğini anlamak için kullanabilirsiniz.

Doktorlar doğal bir deneyin parçasıdır. Hükümet, hekimlere Medicare sigortasına sahip hastaların parasını geri ödemek için kullandığı formülü, özünde keyfi nedenlerden ötürü belirli aralıklarla değiştiriyor. Bazı bölgelerdeki doktorlar belli işlemlerde aldıkları paranın arttığını görürken, bazı bölgelerde tam tersi oluyor.

Eskiden sınıf arkadaşlarım olan Jeffrey Clemens ve Joshua Gottlieb bu keyfi değişikliğin etkilerini test ettiler. Doktorlar hastalara her zaman aynı şekilde mi bakıyorlar, en zorunlu gördükleri tedaviyi mi uyguluyorlar? Yoksa parasal saikler de rol oynuyor mu?

---

* 1978 yılına ait ünlü bir yazıda ortaya atılan "piyango kazanmanın insanı mutlu etmediği" iddiası büyük oranda çürütülmüştür.

Veriler doktorların parasal saiklerle hareket edebildiğini açıkça ortaya koyuyor. Geri ödemelerin daha yüksek olduğu bölgelerde bazı doktorlar geri ödemesi daha yüksek olan prosedürleri çok daha fazla kullanıyorlar: Mesela daha fazla katarakt ameliyatı ve kolonoskopi yapıp daha fazla MR istiyorlar.

O halde, büyük soru şudur: Hastaları bu fazladan bakım sonucunda daha sağlıklı hale geliyor mu? Bunların hastaların sağlığına sadece "küçük etkiler"i olduğunu belirten Clemens ve Gottlieb, ölüm oranına istatistiksel açıdan kayda değer bir etkisi olmadığını tespit etmişler. Bu doğal deneyin bize söylediği şudur: Doktorlara belli prosedürlere başvurmaları için daha güçlü finansal teşvikte bulunulduğu takdirde, bazı doktorlar hastaların sağlığı açısından büyük bir fark yaratmayan ve yaşam sürelerini uzattığı görülmeyen prosedürlere daha fazla başvuracaktır.

Doğal deneyler hayat memat meselelerinin cevaplanmasına yardımcı olabilir. Ayrıca bazı gençler açısından hayat memat meselesi gibi görünen sorulara da yardımcı olabilir.

* * *

"Stuy" diye bilinen Stuyvesant Lisesi, Aşağı Manhattan'da Dünya Ticaret Merkezi'nden birkaç blok ötede bulunan, Hudson Nehri'ne nazır, 150 milyon dolarlık on katlı bir tuğla binadır. Stuy tek kelimeyle etkileyicidir. 55 tane ileri seviye sınıfı, yedi dil dersi ve Yahudi tarihi, bilimkurgu ve Asya-Amerika edebiyatı seçmeli dersleri vardır. Mezunlarının kabaca dörtte biri Sarmaşık Birliği'ne (Ivy League) ya da benzer prestije sahip üniversitelere kabul edilmektedir. Harvardlı fizik profesörü Lisa Randall, Obama'nın stratejisti David Axelrod, Oscar ödüllü aktör Tim Robbins ve romancı Gary Shteyngart Stuy'da eğitim görmüştür. Açılış konuşmacıları arasında Bill Clinton, Kofi Annan ve Conan O'Brien yer almaktadır.

Stuyvesant'ın sunduğu imkânlardan ve mezunlarından daha kayda değer olan bir şey varsa o da ücretidir: Sıfır dolar. Devlet okuludur ve muhtemelen ülkenin en iyisidir. Gerçekten de 300 bin öğrenci ve ebeveynden toplam 27 milyon değerlendirmeyi dikkate alarak yapılan bir çalışmada ABD'deki bütün liseler sıralanmış ve Stuy birinci sırayı almıştır. Dolayısıyla New York'taki hırslı, orta sınıf ailelerin ve onların bir o kadar hırslı çocuklarının Stuy markası konusunda takıntılı olmalarına şaşmamak gerek.

Queens'te yaşayan bir sigortacı ile öğretmenin oğlu olan Ahmed Yılmaz,* "lise dendi mi Stuy'u tek geçer"di.

---

* İsmi ve birkaç küçük ayrıntıyı değiştirdim.

"İşçi sınıfından ve göçmen aileler Stuy'u kurtuluş olarak görüyorlar. Eğer çocuğunuz Stuy'a giderse, ilk yirmideki iyi üniversitelerden birine girecektir. Aile de kurtulacaktır," diyor Yılmaz.

Peki, Stuyvesant Lisesi'ne girmenin yolu nedir? New York'un beş ilçesinden birinde yaşamanız ve kabul sınavlarında belli bir puanın üstüne çıkmanız gerekiyor. Hepsi bu. Ne bir tavsiye mektubu, ne bir yazı, ne bir sicil yoklaması. Tek bir gün, tek bir sınav, tek bir puan. Eğer puanınız belli bir barajı geçerse, işlem tamam.

Her yıl Kasım ayında yaklaşık 27 bin New Yorklu genç bu sınava giriyor. Amansız bir rekabet var. Sınava girenlerin yüzde beşinden azı Stuy'da okumaya hak kazanıyor.

Yılmaz, "Annem canını dişine taktı ve elindeki azıcık parayı sınava hazırlanmam için harcadı," diyor. Yılmaz hafta içleri öğleden sonralarını, hafta sonlarının ise tamamını sınav hazırlığına harcadıktan sonra, Stuy'a gireceğine emindi. Sonuçların yazılı olduğu zarfı aldığı günü hâlâ unutamıyor. İki soruyla direkten dönmüştü.

Bunun nasıl bir his olduğunu sordum. "Ortaokuldayken dünyan başına yıkılırsa ne hissedersin?" diye yanıt verdi.

Teselli ödülü de hiç de kötü değildi: Yine sayılı, üst sıralardaki devlet okullarından biri olan Bronx Fen Lisesi. Ama Yılmaz, Bronx Fen Lisesi'nin teknik meselelere ilgi duyan gençlere yönelik bir okul olduğunu düşünüyordu. Dört yıl sonra, Princeton'dan ret cevabı aldı. Tufts'ta okudu ve kariyerinde farklı yollardan ilerledi. Bugün bir teknoloji şirketinde makul derecede başarılı bir çalışan olmasına karşın, yaptığı işin "zihinsel açıdan hissizleştirici" olduğunu ve yaptığı işin karşılığını alamadığını düşünüyor.

Üzerinden on küsur yıl geçmiş olmasına karşın, Yılmaz eğer Stuy'a gitseydim hayatım ne kadar farklı bir mecrada gelişirdi, diye merak etmekten kendini alamıyor. "Her şey farklı olurdu. Gerçek anlamıyla tanıdığım herkes farklı olurdu," diyor. Stuyvesant Lisesi'ne gitseydi SAT'de daha yüksek puan alabilir miydi, Princeton ya da Harvard gibi Tufts'tan çok daha iyi olduğunu düşündüğü kalburüstü okullara gidebilir miydi, daha yüksek maaşlı ya da tatmin edici bir işte çalışabilir miydi? Bu soruları kendine sormadan edemiyor.

İnsanların varsayımlara kafa yormaları eğlenceli olabileceği gibi, kendine eziyete de dönüşebilir. Eğer şu kıza ya da oğlana açılsaydım hayatım nasıl olurdu? Eğer şu işi kapsaydım? Eğer şu okula gitseydim? Ama bu "eğer"lerin bir cevabı yok gibi duruyor. Hayat bir video oyunu değil ki! Farklı senaryolarda istediğiniz sonucu alana kadar başa dönemezsiniz.

Çek yazar Milan Kundera'nın *Varolmanın Dayanılmaz Hafifliği** adlı kitabında buna ilişkin özlü bir sözü var: "İnsan hayatı ancak bir defa yaşanır ve kararlarımızın hangilerinin doğru hangilerinin yanlış olduğunu kestirememizin nedeni, verili bir durumda ancak bir tek karar verebilecek olmamızdır; ikinci, üçüncü ya da dördüncü bir yaşamımız yok ki çeşitli kararlarımızı birbirleriyle karşılaştıralım."

Yılmaz o sınavda iki puan daha almayı bir şekilde başardığı bir hayatı asla yaşayamayacak.

Ama belki de çok sayıda Stuyvesant Lisesi mezunuyla ilgili yapacağımız bir araştırmayla Yılmaz'ın hayatının ne kadar farklı olabileceğine dair bir içgörüye ulaşabiliriz.

En kaba ve naif metodoloji Stuyvesant'a gitmiş öğrenciler ile gitmemiş öğrencilerin hepsini karşılaştırmak olacaktır. AP (Advanced Placement – İleri Düzey Yerleştirme) sınavlarında ve SAT'de (Scholastic Aptitude Test – Eğitim Yetenek Testi) kaç puan aldıklarını ve hangi üniversitelere kabul edildiklerini inceleyebiliriz. Bunu yaptığımızda, Stuyvesant'a gitmiş öğrencilerin standart sınavlarda çok daha fazla puan aldıklarını ve çok daha iyi üniversitelere kabul edildiklerini tespit ederiz. Ama yukarıda gördüğümüz üzere, bu tür bir kanıt başlı başına ikna edici değildir. Belki de Stuyvesant öğrencilerinin çok daha iyi puanlar almalarının sebebi, Stuy'un en baştan zaten iyi öğrencileri bünyesinde topluyor olmasıdır. Burada bağıntı nedenselliği kanıtlamaz.

Stuyvesant Lisesi'nin nedensel etkilerini test etmek için neredeyse benzeş olan iki grubu karşılaştırmamız gerekiyor: Stuy tedrisinden geçmiş olanlar ile geçmemiş olanlar. Doğal bir deneye ihtiyacımız var. Ama nasıl yapabiliriz?

Cevap: Yılmaz gibi Stuyvesant'ın kıyısından dönecek kadar puan almış öğrencilerle.† Stuy'u kıl payı kaçıran öğrenciler kontrol grubudur; sınavı kazananlar ise deney grubudur.

İki gruptan birindeki öğrencilerin yetenek ya da azim konusunda büyük farkları olduğunu düşünmek için pek bir neden yok. Nihayetinde, bir insanın sınavda bir ya da iki puan fazla almasının sebebi ne olabilir ki? Belki daha az puan alan öğrenci on dakika daha az uyumuştur ya da besin değeri daha düşük bir kahvaltı

---

* Çev. Fatih Özgüven, İstanbul: İletişim, 1980, s. 226.

† Aldığı puanla Stuyvesant'ı kıl payı kaçıranları ararken, gençlik yıllarındaki bu sınav deneyimini hatırlayan ve kıl payı kaçırmış olmasını çarpıcı sözcüklerle anlatan bir dolu insan (yirmili yaşlardan ellili yaşlara kadar) olduğunu öğrenmek beni çok şaşırttı. Bu kişiler arasında Kongre üyeliği yapmış ve New York belediye başkanlığına adaylığını koymuş olan Anthony Weiner da vardı. Dediğine göre, Stuy'u bir puanla kaçırmış. Telefon görüşmemizde biraz sitemliydi: "Beni almadılar."

yapmıştır. Belki yüksek puan olan öğrenci sınavdaki bilhassa zor bir kelimeyi üç yıl önce büyükannesiyle yaptığı bir konuşmadan hatırlıyordur.

Bu doğal deneyler kategorisi o kadar güçlüdür ki, iktisatçılar arasında nam salmıştır: Regresyon süreksizliği. İnsanları iki farklı gruba ayıran kesin bir rakam (bir süreksizlik) söz konusu olduğunda, iktisatçılar sınıra çok ama çok yakın olan insanların akıbetlerini karşılaştırma (regresyon) imkânına sahipler.

M. Keith Chen ve Jese Shapiro adlı iki iktisatçı kötü hapishane koşullarının gelecekte işlenecek suçlara etkilerini test etmek için federal hapishanelerin kullandığı keskin bir eşik kıstasından yararlanmıştı. ABD'de federal hapishanelerdeki mahkûmlara suçlarının doğasına ve suç sicillerine bakılarak belli bir puan veriliyor. Bu puan hapishanede kaldıkları koşulları belirliyor. Yeterince yüksek puan alan mahkûmlar yüksek güvenlikli bir ıslahevine giderler; bu şu anlama geliyor: diğer insanlarla daha az temas, daha az hareket serbestisi ve muhtemelen gardiyanlardan ya da diğer mahkûmlardan daha fazla şiddet.

Burada da, yüksek güvenlikli hapishanelere giden tüm mahkûmları düşük güvenlikli hapishanelere giden tüm mahkûmlarla karşılaştırmak adil olmayacaktır. Yüksek güvenlikli hapishanelerde daha fazla katil ve tecavüzcü tutulurken, düşük güvenlikli hapishanelerde daha fazla uyuşturucu satıcısı ve gaspçı bulunur.

Ama bu kesin sayısal eşiğin hemen üstünde ya da hemen altında olanların suç sicili ve geçmişleri neredeyse aynıdır. Gelgelelim tek bir puan farkı çok farklı bir hapishane deneyimine yol açmıştır.

Sonuç? Deneyi yapan iki iktisatçı, zor koşullara mahkûm edilen mahpusların salıverildikten sonra yeni suçlara karışmalarının daha muhtemel olduğunu tespit ettiler. Zor hapishane koşulları, suça bulaşmaya tövbe ettireceğine, onları daha da hırslandırmakta ve dış dünyaya döndüklerinde daha fazla şiddete bulaşmalarına yol açmaktadır.

O halde bu tür bir "regresyon süreksizliği" Stuyvesant Lisesi için neyi gösterir? MIT ve Duke'ta çalışan bir grup iktisatçı (Atila Abdulkadiroğlu, Joshua Angrist ve Parag Pathak) bu çalışmaya imza attı. Eşiğin her iki tarafındaki öğrencilerin akıbetlerini karşılaştırdılar. Başka bir deyişle, bu iktisatçılar Yılmaz gibi Stuyvesant'ı bir-iki soruyla kaçırmış olan yüzlerce öğrenciye baktılar. Onları daha iyi bir sınav çıkarmış ve Stuy'a girmeye bir-iki soruyla hak kazanmış olan öğrencilerle karşılaştırdılar. Başarı kıstasları AP puanları, SAT puanları ve nihayetinde gittikleri üniversitelerin sıralamadaki yerleriydi.

Elde ettikleri şaşırtıcı sonuçlar, yazılarının başlığında açıkça görülüyordu: "Elit Yanılsaması". Stuyvesant Lisesi'nin etkisi var mıydı? Hayır. Hiç. Sıfır. Eşiğin iki

tarafındaki öğrenciler de aynı AP puanlarını ve aynı SAT puanlarını almış, eşit derecede prestijli üniversitelere gitmişlerdi.

Araştırmacılara göre, Stuy öğrencilerinin hayatta Stuy'lu olmayan öğrencilerden daha başarılı olmalarının tek sebebi, her şeyden önce Stuy'a zaten daha iyi öğrencilerin gitmesiymiş. Stuy sizin AP sınavlarında daha iyi puan almanızı, SAT'de daha başarılı olmanızı ya da daha iyi bir üniversiteye gitmenizi sağlamaz.

Deneyi yapan iktisatçılar, "Çok sayıda öğrenciye yönelik ileri düzey eğitim, okula giriş sınavında yaşanan yoğun rekabetin gerekçesi olamaz," sonucuna varmışlardı.

Hangi okula gittiğinizin neden bir önemi yok? Birkaç hikâye daha cevabı bulmanıza yardımcı olabilir. Sarah Kauhmann ve Jessica Eng adlı iki öğrenciyi düşünün. Bu iki New Yorklu genç de küçük yaştan beri Stuy'a gitmenin hayalini kuruyordu. Kaufmann'ın puanı, sınavı kazanmasına bir soruyla yetmişti. Kaufmann o günü, "Bu kadar heyecan verici bir şey daha yaşayacağımı düşünmüyorum," diye hatırlıyor. Eng'in puanıysa tam sınırdaydı; bir soruyla kaçırmıştı. Kaufmann hayalindeki okula gitmiş, Eng gidememişti.

Peki, hayatları nasıl şekillendi? İkisi de sonrasında başarılı bir kariyer sahibi oldu, sınavlarda yüzde beşe giren New Yorkluların çoğu gibi. İşin ironik yanı, Eng lise yıllarını daha güzel anlatıyor. Eng'in gittiği Bronx Fen Lisesi, Holokost müzesi olan yegâne liseydi. Eng küratörlüğü sevdiğini fark etmiş ve Cornell'de antropoloji okumuştu.

Kaufmann ise Stuy'da kendini biraz kaybettiğini düşünüyordu. Notlara çok odaklanmıştı ve öğretmeye değil, sınavlara önem verildiğini düşünüyordu. Deneyimini, "kesinlikle karmakarışık bir şey" olarak nitelendirmişti. Ama öğretici bir deneyim olmuştu. Üniversite seçimine sıra geldiğinde, öğretmeye daha fazla vurgu yapan sosyal bilimler bölümlerine ve okullara başvurmuştu. Hayallerindeki okul olan Wesleyan Üniversitesi'ne kabul edilmişti. Orada başkalarına yardım etmenin zevkine varmış ve bir kamu avukatı olmuştu.

İnsanlar deneyimlerine uyum sağlayan varlıklardır ve başarılı olacak insan her durumda kendisine yarar sağlayacak şeyleri bulup çıkartır. Sizi başarılı kılan şey yeteneğiniz ve azminizdir; mezuniyet konuşmanız ya da en prestijli okulların sunduğu diğer üstünlükler değildir.

Bu yalnızca tek bir çalışma ve muhtemelen Stuyvesant'ı kıl payı kaçıran öğrencilerin çoğunun yine iyi bir okula girmiş olmaları çalışmanın zayıf kalan yönüdür. Ama iyi bir okula gitmek önemli olsa da, mümkün olan en iyi okula gitmenin pek de kazanç sağlamadığına dair kanıtlar giderek artmaktadır.

Üniversiteyi ele alalım. Harvard gibi dünya üzerindeki en iyi üniversitelerden birine ya da Penn State gibi sağlam bir okula gitmeniz önemli midir?

Bir kez daha, okulunuzun sıralamadaki yeri ile kazandığınız para arasında net bir bağıntı vardır. Ortalama bir Harvard mezunu kariyerinin onuncu yılında 123 bin dolar kazanıyor, ortalama bir Penn State mezunu ise 87.800 dolar.

Ama bu bağıntı nedensellik bağı olduğu anlamına gelmez.

Stacy Dale ve Alan B. Krueger adlı iki iktisatçı, elit üniversitelerin mezunlarının gelecekteki kazançlarına nedensel etkisini test etmenin dâhice bir yöntemini buldular. Lise öğrencileriyle ilgili, hangi üniversiteye başvurduklarından hangi üniversiteye kabul edildiklerine, hangi üniversiteye gittiklerinden aile geçmişlerine ve ilerideki gelirlerine kadar bir dolu bilginin izini süren büyük bir veri kümesi oluşturdular.

Dale ve Krueger, deney grubu ile kontrol grubunu oluşturmak için benzer kökenlerden gelen ve aynı okullara kabul edilen ama farklı okulları seçen öğrencileri karşılaştırdılar. Harvard'a kabul edilen öğrencilerden bazıları Penn State'e gitmişti; kim bilir, belki sevgilisine daha yakın olmak için ya da ders almak istediği bir profesör orada çalıştığı için. Başka bir deyişle, bu öğrenciler kabul komitelerine göre Harvard'a giden öğrenciler kadar yetenekliydi. Ama farklı bir eğitim deneyimi yaşamışlardı.

Peki, benzer kökenlerden gelen iki öğrencinin ikisi de Harvard'a kabul edilirse ama biri Penn State'i seçerse, ne oluyordu? Araştırmacıların sonuçları Stuyvesant Lisesi araştırmacılarının ulaştıkları kadar şaşırtıcıydı. Bu öğrenciler ileride az çok aynı gelire sahip olmuşlardı. Eğer gelecekteki maaşınız ölçüyse, benzer derecede prestijli okullara seçilen ama farklı okullara gitmeyi tercih eden benzer öğrenciler ileride aynı yere varmışlardı.

Gazetelerimizde Sarmaşık Birliği okullarına giden müthiş başarılı insanlarla ilgili yazılardan geçilmez: Microsoft'un kurucusu Bill Gates ile Facebook'un kurucuları Mark Zuckerberg ve Dustin Moskovitz gibi isimlerin hepsi de Harvard'da okumuştur. (Kabul, ama hepsi de okulu bırakmış ve dolayısıyla Sarmaşık Birliği eğitiminin değerini daha da şüpheli hale getirmişlerdir.)

Keza Sarmaşık Birliği'ne kabul edilecek kadar yetenekli olup daha az prestijli bir okula gitmeyi tercih eden ve son derece başarılı bir kariyer sahibi olan insanlar da vardır: Pennsylvania Üniversitesi'nde Wharton Fakültesi (Sarmaşık Birliği'ne üye bir işletme fakültesi) ile başlayıp sonradan naklini Nebraska-Lincoln Üniversitesi'ne aldıran (çünkü bu okul daha ucuzdu, Philadelphia'dan nefret ediyordu ve Wharton'daki dersleri sıkıcı buluyordu) Warren Buffett gibi isimler. Veriler gösteriyor ki, en

azından gelir bakımından, daha az prestijli bir okula gitmeyi tercih etmek Buffett ve diğerleri için iyi bir karar olmuştur.

* * *

Bu kitaba *Bana Yalan Söylediler* adını verdim. Bunu derken, esasen, insanların kendilerini daha iyi göstermek için –arkadaşlarına, anketlere ve kendilerine– yalan söylediklerini kastediyorum.

Ama dünya da hatalı, yanıltıcı veriler sunarak bize yalan söylüyor. Dünya bir sürü başarılı Harvard mezununu gözümüze gözümüze sokuyor, ama aynısını çok daha az sayıda Penn State mezunu başarılı isim için yapıyor; böylece biz de Harvard'a gitmenin müthiş bir üstünlük sağladığını varsayıyoruz.

Doğanın deneylerinden bir anlam çıkararak, dünyanın verilerini doğru bir şekilde anlamlandırıp neyin gerçekten yararlı olup neyin olmadığını bulabiliriz.

* * *

Doğal deneylerin önceki bölümle de bağı var. Çoğu zaman bunlara da yakından bakmak gerekiyor: Deney grubuna ve kontrol grubuna, yani Super Bowl deneyindeki şehirlere, Medicare fiyatlandırma deneyindeki coğrafi bölgelere, Stuyvesant deneyinde eşiğe yakın olan öğrencilere. Önceki bölümde ele alındığı üzere, yakından bakma çoğu zaman büyük, kapsamlı veri kümeleri gerektirir: Dünyanın dijitalleşmesiyle birlikte giderek daha da erişilebilir hale gelen türde verilerdir bunlar. Doğanın ne zaman deney yapmaya girişeceğini bilmediğimizden, sonuçları değerlendirmek için küçük bir araştırma ya da anket yapamayız. Bu müdahalelerden dersler çıkarmak için bir sürü mevcut veriye ihtiyacımız var. Büyük Veri'ye ihtiyacımız var.

Bu bölümde ayrıntılarıyla ele alınan deneyler (ister bizim yaptığımız deneyler ister doğanınkiler) hakkında vurgulanması gereken önemli bir husus daha var. Bu kitabın büyük kısmında dünyayı anlamaya odaklandık: Irkçılık Obama'ya nelere mal oldu, gerçekte kaç gey erkek vardır, erkekler ve kadınlar bedenlerine ne kadar güveniyorlar? Ama bu kontrollü ya da doğal deneylerin pratik bir yanı da bulunuyor. Karar alma yetimizi geliştirmeyi, işe yarayan ve yaramayan müdahaleleri öğrenmemize yardımcı olmayı amaçlıyor.

Şirketler daha fazla müşteri elde etmenin yollarını öğrenebilirler. Hükümetler doktorları en iyi şekilde motive etmek için geri ödemelerden nasıl yararlanılacağına bakıp bunlardan dersler çıkarabilirler. Öğrenciler hangi okulların ileride en fazla yarar sağlayacağını anlayabilirler. Bu deneyler, tahminlerin, geleneksel düşünme biçiminin ve kaba bağıntıların yerine gerçekten işe yarar şeyleri koyabilme konusunda Büyük Veri'nin sunduğu potansiyeli nedensel olarak gösteriyor.

Üçüncü Kısım

Büyük Veri: Özenle Ele Alınız

YEDİNCİ BÖLÜM

Büyük Veri, Büyük Şema mı? Elinden Gelmeyenler

Bir gün, "Seth, Lawrence Summers seninle görüşmek istiyor," şeklinde biraz şifreli bir e-posta aldım. Tez danışmanlarımdan Lawrence Katz atmıştı. Katz, Summers'ın çalışmamla neden ilgilendiğini yazmamıştı, ama sonradan öğrendim ki Katz yapıp ettiklerimi başından beri takip ediyormuş.

Summers'ın ofisinin bekleme odasında oturdum. Biraz gecikmenin ardından ABD'nin sabık Hazine Bakanı, Harvard eski dekanı ve iktisatta en önemli ödüllerden bazılarının sahibi olan Summers beni içeri buyur etti.

Summers ırkçılığın Obama'ya etkilerini ele aldığım yazımı okuyarak karşıladı beni. Sekreteri yazının çıktısını almıştı. Summers hızlı okuyan biriydi. Okurken dilini sık sık dışarı çıkarıp sağa atıyor, gözlerini de hızla soldan sağa ve aşağıya doğru kaydırıyordu. Summers'ın sosyal bilimler alanından bir yazı okuyor olmasını, büyük bir piyanistin bir sonatı icra etmesine benzettim. O kadar odaklanmış haldeydi ki, geri kalan her şeyle bağını koparmıştı. Otuz sayfalık yazımı okuması beş dakika sürmedi.

"Google'daki 'nigger' aramalarının ırkçılığın göstergesi olduğunu söylüyorsunuz," dedi Summers. "Akla yatkın geliyor. Obama'nın Kerry'den daha az destek aldığı yerleri tahmin etme imkânı sunuyor. İlginçmiş. Obama ile Kerry'yi gerçekten aynı görebilir miyiz?"

"Siyaset bilimcilerin gözünde ikisi benzer ideolojilere sahiptir," yanıtını verdim. "Ayrıca ırkçılık ile Başkanlık seçimindeki oylar arasında bir bağıntı bulunmuyor. Demografi, kiliseye gitme alışkanlığı ve silah sahibi olma gibi etkenleri eklediğimizde bile sonuç aynı." Biz iktisatçılar böyle konuşuruz. Galeyana gelmiştim.

Summers durup yüzüme baktı. Bir an ofisindeki televizyonuna döndü, CNBC açıktı, sonra bir daha bana baktı, sonra bir daha televizyona, sonra tekrar bana baktı. "Güzel, tuttum bu yazıyı. Başka neler çalışıyorsunuz?" diye sordu.

Sonraki altmış dakika entelektüel açıdan hayatımın en güzel anları olabilir. Summers'la faiz oranları ve enflasyon hakkında, polisin icraatları ve suç oranları, iş dünyası ve yardım kampanyaları hakkında konuştuk. Summers'la tanışan birçok

insanın heyecanlanması boşuna değil. Ömrüm boyunca birçok akıllı insanla tanışıp konuşma fırsatım oldu; Summers içlerinde en akıllısıydı. Fikirler konusunda her şeyden daha takıntılı biri; görünen o ki bu yüzden sık sık başı belaya giriyor. Bilim alanında kadınların sayıca az olmasının bir nedeni de erkeklerin IQ'larının daha fazla çeşitlilik içermesi olabilir, dediği için Harvard'daki görevinden istifa etmek zorunda kalmıştı. Summers ilgi çekici bir fikir bulursa, birilerini rahatsız etmek pahasına da olsa sözünü sakınmıyor.

Görüşmemizin tahmini bitiş süresini yarım saat aşmıştık. Mest edici bir sohbetti, ama yine de neden orada olduğumu da, ne zaman gitmem gerektiğini de bilmiyordum; hatta ne zaman gitmem gerektiğini nasıl anlayacağımı dahi kestiremiyordum. O an, Summers'ın da bu görüşmeyi niye ayarladığını unutmuş olabileceği hissine kapıldım.

Ardından milyon dolarlık, belki de milyar dolarlık soruyu sordu: "Bu verilerle borsayı öngörebileceğinizi düşünüyor musunuz?"

Aha! Nihayet ağzındaki baklayı çıkarmıştı.

Summers bana bu soruyu soran ilk insan değildi. Babam benim bu geleneksel olmayan araştırma tarzımı ve ilgi alanlarımı genel anlamda hep desteklemiştir. Ama bir keresinde konuyu buraya getirmişti. "Irkçılık, çocuk istismarı, kürtaj iyi hoş da, bu uzmanlığınla para kazanamıyor musun?" Dostlarım ve diğer aile fertlerim de bu konuyu gündeme getirmiştir. Keza mesai arkadaşlarım ve internetteki tanımadığım insanlar da öyle. Anlaşılan o ki, Google aramalarından ya da diğer Büyük Veri bilgilerinden yararlanarak borsadan para kaldırıp kaldıramayacağım herkesin içine dert olmuştu. Şimdi de sıra sabık Hazine Bakanı'na gelmişti. Bu daha ciddiydi.

Yeni Büyük Veri kaynakları borsanın seyrini doğru tahmin edebilir mi? Kısa ve net bir cevap vereyim: Hayır.

Önceki bölümlerde Büyük Veri'nin dört güçlü yanından bahsetmiştik. Bu bölümdeyse Büyük Veri'nin sınırlı yanlarından bahsedeceğiz: Hem bununla neler yapamayacağımızdan hem de ara ara ne yapmamamız gerektiğinden. Summers'ın ve benim piyasaların altını üstüne getiremeyişimizin hikâyesini anlatmak münasip bir kalkış noktası olabilir.

Üçüncü Bölüm'de, bir alanda mevcut araştırmaların zayıf olması durumunda yeni verilerin büyük ihtimalle büyük getirileri olacağını belirtmiştik. Irkçılık, çocuk istismarı ya da kürtaj konusunda yeni içgörüler elde etmenin bir şirketin işleyişi konusunda yeni, kârlı bir içgörü elde etmekten çok daha kolay olması dünyanın talihsiz bir hakikatidir. Bunun sebebi, ticari performans ölçümünde en ufak bir

üstünlük elde etmek için bile muazzam kaynaklar ayrılıyor olmasıdır. Finans dünyasında kıyasıya bir rekabet var. Bu daha en baştan bizim için büyük bir darbeydi.

Zaten başka insanların zekâsından kolay kolay etkilenmemesiyle tanınan Summers, serbest yatırım fonlarının bizden çok ileride olduğundan emindi. Konuşmamız sırasında onlara çok saygı duyduğunu gördüm, ayrıca önerilerimden çoğunun onların bizden üstün olduklarını gösterdiğini düşünüyordu; açıkçası bunlar beni epey bir şaşırttı. Google Trends'ten daha eksiksiz veri temin etmemi sağlayan bir algoritma geliştirdiğimi gururla anlattım. Zekice olduğunu söyledi. Nitelikli bir serbest yatırım fonu olan Rönesans'ın bu algoritmayı çözüp çözemeyeceğini sorduğumda, kıkırdayıp, "Yani, elbette çözerler" dedi.

Summers'la birlikte piyasaları alt etmek için yeni ve büyük veri kümelerinden yararlanırken karşılaştığımız yegâne temel problem serbest yatırım fonlarıyla aşık atmanın zorluğu değildi.

Boyutsallık Laneti

Borsada başarılı tahminler yapma stratejinizin şanslı bir madeni para bulmak olduğunu varsayalım, ama bu madeni para dikkatli bir test sonucunda bulunacaktır. İşte size metodoloji: Bin tane madeni parayı 1'den 1000'e kadar etiketliyorsunuz. İki yıl boyunca her sabah her birini havaya atıp, yazı mı tura mı geldiğini yazıyorsunuz ve ardından Standard & Poors Endeksi o gün yükselmiş mi düşmüş mü not ediyorsunuz. Tüm verilerinizi didik didik ediyorsunuz. İşte zafer! Bir şey buldunuz. 391 numaralı paranın yazı geldiği örneklerin yüzde 70,3'ünde S&P Endeksi yükselmişti. İlişki istatistiksel açıdan önemlidir, kesinlikle öyledir! Dolayısıyla şanslı paranızı buldunuz, gözünüz aydın!

Her sabah 391 numaralı madeni parayı havaya atın ve ne zaman yazı gelirse hisse alın. Şanslı tişörtler ve Japon eriştesi yeme günleriniz geride kaldı. 391 numaralı madeni para güzel bir hayatın anahtarıdır!

Ya da değildir.

"Boyutsallık laneti"nin en şeytani yönlerinden birinin yeni kurbanı oldunuz. Çok sayıda değişkenin –ya da "boyut"un– (bu örnekte bin tane madeni paranın), o kadar fazla gözlemin (bu örnekte iki yılda 504 ticari işlem gününün) izini süremediği her durumda bu lanet kapınızı çalabilir. Bu boyutlardan biri (bu örnekte 391 numaralı madeni para) muhtemelen şanslı çıkacaktır. Değişkenlerin sayısını azaltın (sadece yüz madeni parayı havaya atın), o zaman bunlardan birinin şanslı para olma ihtimali çok daha azalacaktır. Gözlemlerin sayısını artırın (S&P Endeksi'ni yirmi yıl boyunca tahmin etmeye çalışın), yazı turayla gidişatı yakalamakta zorlanırsınız.

Boyutsallık laneti Büyük Veri için temel bir sorundur; zira yeni veri kümeleri bize sık sık geleneksel veri kaynaklarından çok çok daha fazla değişken (her arama terimi, her tweet kategorisi vb.) sunar. Bir Büyük Veri kaynağından yararlanarak piyasaları tahmin ettiğini iddia eden birçok insan bu lanetin tuzağına düşmüştür; zira aslında tek yaptıkları 391 numaralı madeni paranın muadilini bulmaktır.

Örneğin insanların ne tür tweet'ler attığına bakarak piyasaların seyrini doğru tahmin edebileceklerini iddia eden Indiana Üniversitesi ile Manchester Üniversitesi'nden bir grup bilgisayar mühendisini ele alalım. Dünyanın günlük ruh halini tweet'ler temelinde şifreleyecek bir algoritma geliştirmişlerdi. Üçüncü Bölüm'de ele alınan duygu analizine benzer tekniklerden yararlanmışlardı. Gelgelelim sadece tek bir ruh halini değil, birçok ruh halini kodlamışlardı: Mutluluk, öfke, nezaket ve daha fazlası. Sakinlik ifade eden tweet'lerin ağır basmasının ("sakin hissediyorum" gibi) Dow Jones Sınai Ortalaması'nın altı gün sonra yükseleceğinin habercisi olduğunu tespit etmişlerdi. Bulgularından yararlanmak üzere bir serbest yatırım fonu oluşturuldu.

Burada sorun neydi?

Temel sorun çok fazla şeyi test etmiş olmalarıydı. Eğer yeterince şeyi test ederseniz, bunlardan biri sırf şansına, istatistiksel açıdan kayda değer olacaktır. Bu mühendisler birçok duyguyu test etmişlerdi. Ayrıca tahmin etmeye çalıştıkları borsa davranışından bir gün önce, iki gün önce, üç gün önce ve yedi gün önceye kadar her bir duyguyu test etmişlerdi. Tüm bu değişkenler Dow Jones'un sadece birkaç aylık yükseliş ve düşüşlerini açıklamaya çalışmak için kullanılmıştı.

Altı gün önceki sakinlik borsayı doğru tahmin etmek için doğru bir etken değildi. Altı gün önceki sakinlik bizim farazi 391 numaralı madeni paramızın Büyük Veri muadiliydi. Tweet temelli serbest yatırım fonu, kayda değer bir getirileri sağlamadığından dolayı bir ay sonra rafa kaldırıldı.

Piyasaları tweet'lerle ölçmeye çalışan serbest yatırım fonları boyutsallık lanetinin yegâne mağdurları değildir. Varlığımızın genetik anahtarlarını bulmaya çalışan sayısız bilimci de bu tuzağa düşmüştür.

İnsan Genomu Projesi sayesinde, bugün artık insanların DNA'sını eksiksiz bir şekilde toplayıp analiz edebiliyoruz. Bu projenin potansiyelleri muazzam görünüyordu.

Belki de ileride şizofreniye neden olan geni bulabiliriz. Belki de Alzheimer, Parkinson ve ALS hastalığına neden olan geni keşfedebiliriz. Belki de –heyhat– zekâya neden olan geni bulabiliriz. İnsanın IQ puanını yükselten tek bir gen var mıdır?

Seçkin davranışsal genetikçilerden Robert Plomin 1998'de cevabı bulduğunu iddia etti. Yüzlerce öğrencinin DNA ve IQ'sunun yer aldığı bir veri kümesi oluşturdu. "Dâhiler"in DNA'sını (160 ya da daha yüksek IQ'su olanları) ortalama IQ'su olanlarla karşılaştırdı.

Bu iki grubun DNA'larında çarpıcı bir fark tespit etti: 6. kromozomun küçük bir köşesinde yer alıyordu, beynin metabolizmasında kullanılan muğlak ama güçlü bir gendi. Bu genin IGF2r adı verilen bir versiyonu dâhilerde iki kat daha yaygındı.

New York Times, "Yüksek Zekâyla İlişkilendirilecek İlk Genin Bulunduğu Söyleniyor" başlığını atmıştı.

Plomin'in bulgusunun birçok açıdan etik sorunlar doğurduğunu söyleyebilirsiniz. Aileler çocuklarının IGF2r'sinin görüntülenmesine izin vermeli midir? IQ'su düşük olan bir bebeği aldırma hakları olmalı mıdır? İnsanları genetik olarak değiştirip yüksek IQ'lu yapmalı mıyız? IGF2r ile ırk arasında bağıntı var mıdır? Bu sorunun cevabını bilmek istiyor muyuz? IQ genetiğiyle ilgili araştırmalar devam etmeli midir?

Biyoetikçilerin bu zorlu sorulara el atmasından da önce, genetikçilerin önünde (Plomin dahil) daha temel bir soru vardı. Sonuç doğru muydu? IGF2r gerçekten de IQ'yu öngörebiliyor muydu? Dâhilerin bu genin belli bir çeşidine sahip olma ihtimalleri gerçekten iki kat daha fazla mıydı?

Hayır. Plomin ilk çalışmasından birkaç yıl sonra, başka bir insan grubun IQ'larını ve DNA'larını inceleme şansı buldu. Bu kez IGF2r ile IQ arasında bağıntıya rastlamadı. Plomin –iyi bir bilimciye yakışır şekilde– iddiasını geri çekti.

Aslında genetik ve IQ araştırmalarında genel bir örüntüden bahsediyoruz. İlkin, bilimciler IQ'yu tahmin eden bir gen türü bulduklarını söylüyorlar. Ardından yeni veriler elde edip ilk iddianın yanlış olduğunu keşfediyorlar.

Örneğin yakın dönemde yayımlanan bir yazıda, Christopher Chabris'in başını çektiği bir grup bilimci, IQ ile bağlantılı genetik çeşitler hakkındaki on iki asal iddiayı inceledi. On bin insanın verilerini masaya yatırdıktan sonra bu on ikisinden hiçbiriyle bağıntı bulamadılar.

Tüm bu iddiaların sorunlu yanı nedir? Boyutsallık laneti. İnsan genomu, bilimcilerin artık bildiği üzere, milyonlarca şekilde farklılık gösterir. Kısacası, test edilecek çok fazla gen vardır.

Eğer borsayla bağıntısı olup olmadığını görmek için yeterince tweet'i test ederseniz, kazara bağıntısı olan bir tweet bulursunuz. Eğer IQ'yla bağıntısı olup olmadığını görmek için yeterince gen çeşidini test ederseniz, kazara bağıntısı olan bir gen bulursunuz.

Boyutsallık lanetinden nasıl kurtulabiliriz? Çalışmanız konusunda biraz alçakgönüllü olmanız ve sonuçlarınızın cazibesine fazla kapılmamanız gerekir. Bu sonuçları başka testlerden de geçirmeniz gerekir. Örneğin bütün paranızı 391 numaralı paraya yatırmadan önce, sonuçların birkaç yıl boyunca nasıl geliştiğini görmeniz gerekir. Sosyal bilimciler buna "örnekleme dışı" test diyorlar. Ne kadar çok değişken denerseniz, o kadar alçakgönüllü olmanız gerekir. Ne kadar çok değişken denerseniz, örnekleme dışı test o kadar zorlaşır. Ayrıca yaptığınız her testin izini sürmeniz de hayati önemdedir. O zaman lanete kurban gitme ihtimalinizin ne kadar güçlü olduğunu ve sonuçlarınıza ne kadar şüpheyle yaklaşmanız gerektiğini bilirsiniz. Bu da bizi Larry Sumerrs'la olan hikâyeme geri getiriyor. İşte piyasaların altını üstüne getirme çabamızın özeti.

Summers'ın ilk fikri iPhone gibi kilit ürünlerin gelecekteki satış rakamlarını tahmin etmek için aramalardan yararlanmak ve böylece Apple gibi bir şirketin borsadaki performansına ışık tutmaktı. "iPhone" aramaları ile iPhone satışları arasında gerçekten de bir bağıntı vardı. İnsanlar Google'da çok fazla "iPhone" araması yaptığında, daha fazla satış yapılacağını tahmin edebiliriz. Gelgelelim bu bilgi Apple'ın borsadaki fiyatına zaten yansımıştır. Çok açık ki, "iPhone" konusunda çok sayıda Google araması yapıldığında, serbest yatırım fonları büyük bir satış rakamına ulaşılacağını anlıyordu (arama verilerinden mi yoksa başka bir kaynaktan mı yararlandıklarından bağımsız olarak).

Summers'ın bir sonraki fikri, gelişmekte olan ülkelere gelecekte yapılacak yatırımları öngörmekti. Eğer Brezilya ya da Meksika gibi ülkelere yakın gelecekte çok sayıda yatırımcı para akıtacaksa, o zaman bu ülkelerdeki şirketlerin hisseleri kuşkusuz yükselecektir. Kim bilir, belki kilit Google aramalarıyla (mesela "Meksika'ya yatırım" ya da "Brezilya'da yatırım fırsatları" gibi) yatırımlardaki yükselişi öngörebilirdik. Bunun da bir çıkmaz sokak olduğunu anladık. Sorun neydi? Aramaların çok nadir olmasıydı. Bu arama verisi ortaya anlamlı örüntüler çıkaracağına, her şeyi daha da karıştırıyordu.

Tekil hisseler için aramaları denedik. Belki de insanlar "GOOG" araması yapıyorlarsa, buradan Google hisseleri almak üzere oldukları anlamı çıkardı. Bu aramalar hisselerin çok fazla satışa konu olacağını tahmin ediyor gibi görünüyordu. Ama hisseler yükselecek mi alçalacak mı, onu tahmin etmiyordu. Temel sorunlardan biri bu aramaların aramayı yapan kişilerin alan tarafta mı satan tarafta mı olacağını öngörememesiydi.

Bir gün, Summers'a yeni fikrimi heyecanla gösterdim: Geçmişteki "altın al" aramaları altının gelecekteki fiyatıyla bağıntılı görünüyordu. Summers bunun tu-

tarlı bir şekilde doğru çıkıp çıkmayacağını test etmem gerektiğini söyledi. Sanırım bir serbest yatırım fonu da aynı sonuca ulaşmıştı; zira bu örüntü bir süre sonra işlememeye başladı.

Nihayet, birkaç aylık uğraşın sonucunda, testlerimizde faydalı bir sonuca varamadık. Kuşkusuz, milyarlarca Google aramasının her birinde piyasa performansıyla bir bağıntı aramış olsaydık, zayıf da olsa işe yarayan bir tane bulabilirdik. Ama bu da muhtemelen bizim 391 numaralı madeni paramız olurdu.

Ölçülebilir Olana Aşırı Vurgu Yapmak

Yale'de pazarlama profesörü olan Zoë Chance, Mart 2012'de Connecticut'a bağlı New Haven'da bulunan ofisinin posta kutusunda beyaz renkli, küçük bir adım ölçer buldu. Chance, gün içinde attığınız adımları ölçüp sonunda size bir puan veren bu aletin nasıl daha fazla egzersiz yapma itkisi yarattığını incelemeye karar verdi.

Sonrasına olanlar, kendisinin bir TEDx sohbetinde anlattığına göre, tam bir Büyük Veri kâbusuydu. Chance puanını artırma konusunda öyle takıntılı, öyle saplantılı bir hale gelmişti ki, mutfağa, yatak odasına, oturma odasına, kilere, ofise kadar her yere yürümeye başladı. Sabah erken saatte yürüyor, geceleri geç saatte yürüyor, günün neredeyse her saatinde yürüyordu: Herhangi günde yirmi bin adım atıyordu. Adım ölçerini gün içinde yüzlerce kez kontrol ediyordu ve insanlarla artık neredeyse sadece tek bir iletişimi vardı, o da diğer adım ölçer kullanıcılarıyla internet üzerinden sohbet edip puanlarını nasıl artırabileceklerine göre stratejiler belirlemek üzerindeydi. Puanını daha yukarıya çekme konusunda öyle takıntılı hale gelmişti ki, üç yaşındaki kızına bile adım ölçer takmıştı.

Chance, puanını en yükseğe çıkarma konusundaki takıntısı nedeniyle tüm perspektifini kaybetmişti. Bir insanın neden adım puanını yukarıya çekmek isteyeceğini düşünemiyordu: Kızına birkaç adım attırmak değil, egzersiz yapmaktı mesele. Nitekim adım ölçerle ilgili akademik bir araştırma yapmayı da başaramadı. Nihayet bir gece yarısı yorgun argın birkaç adım daha atmaya çalışırken yere kapaklanınca bu aletten kurtulma kararı aldı. Mesleği itibariyle verilerin peşinden giden bir araştırmacı olmasına karşın, bu deneyim onu derinden etkilemişti. "Fazladan veriye sahip olmanın her zaman iyi bir şey olup olmadığı konusuna şüpheyle yaklaşmaya başladım" diyor Chance.

Bu, uç bir örnek. Ama karar almak için verilerden yararlanan insanlar için potansiyel bir probleme işaret ediyor. Sayılar baştan çıkarıcı olabilir. Sayılar konusunda saplantılı hale gelebilir ve bunu yaparken daha önemli etkenleri gözden kaçırabiliriz. Zoë Chance hayatının geri kalanını az çok gözden kaçırmaya başlamıştı.

Sayılara –bu kadar saplantılı derecede olmasa bile– meftun olmak sıkıntılar doğurabilir. 21. yüzyılda Amerikan okullarında sınav ve testlere yapılan vurguyu ve öğretmenlerin öğrencilerini puanlarına göre değerlendirilmesini düşünün. Sınıflarda ne olup bittiğini daha nesnel şekilde ölçme arzusu meşru olmakla birlikte, sayılarla yakalanamayacak birçok şey vardır. Ayrıca tüm bu testler birçok öğretmen üzerinde testlere yönelik şeyler öğretme baskısı doğurmuştur. Daha kötüsü de var. Az sayıda da olsa bazı öğretmenler, Brian Jacob ve Steven Levitt'in kaleme aldıkları yazıda kanıtlandığı üzere, bu testlerde alenen hile yapmışlardır.

Sorun şudur: Ölçebildiğimiz şeyler çoğu zaman tam olarak önem verdiğimiz şeyler değildir. Öğrencilerin çoktan seçmeli sorulardaki başarısını ölçebiliriz. Eleştirel düşünceyi, merakı ya da kişisel gelişimi ölçmekse öyle kolay değildir. Ölçmesi kolay tek bir sayıyı (test puanları ya da gün içinde atılan adım sayısını) artırmaya çalışmak hedeflediğimiz şeyin gerçekleşmesine her zaman yardımcı olmaz.

Facebook da sitesini geliştirmeye çalışırken bu problemle karşı karşıya kalıyor. Şirketin elinde insanların bu siteden nasıl yararlandıklarına dair tonlarca veri var. Belli bir İçerik Akışı hikâyesinin beğenilip beğenilmediğini, tıklanıp tıklanmadığını, altına yorum yazılıp yazılmadığını ya da paylaşılıp paylaşılmadığını görmek kolaydır. Ama Facebook'un bu konularda yazıştığım veri bilimcilerinden biri olan Alex Peysakhovich'e göre, bunlardan hiçbiri çok daha kusursuz bir sorunun yerini tutamaz: Bu siteye girmek nasıl bir deneyimdi? Haber ya da hikâye, kullanıcıyı arkadaşlarına bağladı mı? Dünya hakkında bilgilenmesini sağladı mı? Onu güldürdü mü?

Veya 1990'larda beyzbolda yaşanan veri devrimini düşünün. Birçok takım karar alırken eskisi gibi menajerlere kulak vermek yerine giderek daha karmaşıklaşan istatistiklerden yararlanmaya başladı. Hücumu ve atışı ölçmek kolaydı, ama topu karşılama için aynısı geçerli değildi; bu nedenle bazı takımlar savunmaya gerektiği kadar önem vermez oldu. Nate Silver *The Signal and the Noise* adlı kitabında *Moneyball*'da anlatılan veri temelli takımlardan Oakland A'nın 90'ların ortalarında yılda sekiz-on maçı kötü savunma yaptığı için kaybettiğini söyler.

Çözüm her zaman Büyük Veri'yi daha da artırmak değildir. Büyük Veri'nin en iyi şekilde işlemesi için çoğu zaman özel bir sos gerekir: İnsanların ve küçük araştırmaların, küçük veri diyebileceğimiz şeylerin yargısı. A'nın o dönemki genel menajeri ve *Moneyball*'daki ana karakter olan Billy Beane, Silver'la bir mülakatında yaptıkları esas değişikliğin oyuncu avcılığı bütçesini artırmak olduğunu söyler.

Facebook da devasa veri havuzundaki boşlukları doldurmak amacıyla eski moda bir yaklaşım benimsemek zorunda kaldı: İnsanlara ne düşündüklerini sormak. Her

gün İçerik Akışı'nı yüklerken yüzlerce Facebook kullanıcısına orada gördükleri hikâyeler ya da haberlerle ilgili sorular soruluyor. Başka bir deyişle, Facebook'un otomatik olarak toplanan veri kümeleri (beğeniler, tıklamalar, yorumlar) küçük verilerle ("Bu mesajı İçerik Akışı'nızda görmek ister misiniz?", "Neden?" gibi) takviye ediliyor. Evet, Facebook gibi muazzam başarılı bir Büyük Veri örgütü bile bu kitapta çok küçümsenen bir bilgi kaynağına (küçük bir araştırmaya) bazen başvurabiliyor.

Gerçekten de Facebook'un veri takımı, temeli (devasa tıklama, beğeni ve yorum derlemeleri) takviye edecek küçük verilere duyulan bu ihtiyaçtan ötürü, düşündüğünüzden daha farklı görünebilir. Facebook, sosyal psikologları, antropologları ve sosyologları tam da rakamların gözden kaçırabileceği şeyleri yakalamak için istihdam ediyor.

Bazı eğitimciler de Büyük Veri'deki kör noktalara karşı giderek daha uyanık hale geliyorlar. Kitlesel testleri küçük verilerle takviye etme çabası ülke çapında yayılıyor. Öğrenci araştırmaları iyice arttı. Keza ebeveyn araştırmaları ve öğretmen gözlemleri (diğer deneyimli eğitimcilerin ders sırasında öğretmeni izlemeleri) de öyle.

Harvard'dan eğitim profesörü Thomas Kane, "Okullar sadece test puanlarına odaklanmamaları gerektiğini fark ediyorlar," diyor. Bill & Melinda Gates Vakfi'nın yürüttüğü üç yıllık bir araştırma, eğitimde hem büyük hem de küçük verilerin değerini kanıtlıyor. Yazarlar, hangi öğretmenin öğrencilerin daha iyi öğrenmesini sağladığını ölçmekte test puanı bazlı modellerin mi, öğrenci anketlerinin mi, yoksa öğretmen gözlemlerinin mi en etkili olduğunu analiz etmişlerdi. Üç ölçümü tek bir rakama ulaşacak şekilde birleştirdiklerinde, en iyi sonuçları elde ediyorlardı: "Her ölçüm değerli bir şey katıyor," yargısına ulaşılmıştı.

Gerçekten de birçok Büyük Veri işleminin boşlukları doldurmak için küçük verilerden yararlandığını öğrenmeye başladığım dönemde, Jeff Seder'le buluştum. Hatırlayacak olursanız, kendisi Amerikan Firavunu'nun başarısını tahmin etmek için devasa bir veri kümesinin sağladığı derslerden yararlanan Harvard çıkışlı bir at gurusuydu.

Seder tüm bilgisayar dosyalarını ve matematiğini benimle paylaştıktan sonra başka bir silahı daha olduğunu kabul etmişti: Patty Murray.

Seder gibi Murray de zeki, referansları sağlam biriydi: Bryn Mawr mezunuydu. Ayrıca taşrada yaşamak için New York'tan ayrılmıştı. "Ben atları insanlardan daha çok seviyorum," diyor Bayan Murray. Ama kendisi atları değerlendirirken biraz geleneksel bir tavırla hareket eder. Birçok at menajeri gibi o da atları bizzat inceliyor, nasıl yürüdüklerine bakıyor, yara bere var mı kontrol ediyor ve sahipleriyle konuşuyor.

Ardından Murray ile Seder hangi atları önereceklerini belirlemek üzere istişarede bulunuyorlar. Murray atlardaki problemleri (Seder'ın verilerinin, atlar konusunda gelmiş geçmiş en yenilikçi ve önemli veri kümesi olmasına karşın yine de gözden kaçırdığı problemleri) ortaya çıkarıyor.

Ben Büyük Veri'nin ortaya çıkardıkları temelinde bir devrim gerçekleşeceğini tahmin ediyorum. Ama buradan her soruna salt verilerle yaklaşabileceğimiz sonucu çıkmaz. Ayrıca Büyük Veri insanların yüzlerce yıldır dünyayı anlamak için geliştirdikleri diğer yöntemlere olan ihtiyacı ortadan kaldırmaz. Bunlar birbirini tamamlamaktadır.

SEKİZİNCİ BÖLÜM

Daha Fazla Veri, Daha Fazla Sorun mu? Ne Yapmamalı?

Bazen Büyük Veri'nin gücü o kadar etkileyici bir hal alır ki, göz korkutur. Etik soruları gündeme getirir.

Etki Alanı Genişletilen Şirketlerin Oluşturduğu Tehlike

Geçenlerde üç iktisatçı, Columbia'dan Oded Netzer ve Alain Lemaire ile Delaware Üniversitesi'nden Michal Herzenstein, denkler arası borç verme sitesi olan Prosper'ın verilerinden yararlanarak, borç alan birinin borcunu ödeyip ödemeyeceğini tahmin etmenin yollarını araştırdı. Prosper'a başvuran potansiyel borçlanıcılar paraya neden ihtiyaç duyduklarını ve bu paranın ne işlerine yarayacağını anlatan kısa bir açıklama yazıyorlar, potansiyel borç verenler de parayı verip vermeyeceklerini bildiriyorlar. Borcunu ödeyemeyenlerin toplama oranı yüzde 13 civarıdır.

Potansiyel borçlanıcıların kullandıkları dilin geri ödeyip ödemeyeceklerinin güçlü bir işareti olduğu görüldü. Borç verenlerin bu potansiyel borçlularla ilgili elde edebildikleri kredi notu ve gelir gibi kritik bilgiler elinizde bulunsa bile, kullandıkları dil önemli bir işarettir.

Aşağıda araştırmacıların borç isteyenlerin en çok kullandıklarını tespit ettikleri on tabir sıralanmıştır. Bunlardan beşi borcun ödenmesiyle pozitif bağıntılı, beşiyse negatif bağıntılıdır. Başka bir deyişle, beşi güvenebileceğiniz insanların, beşiyse güvenemeyeceğiniz insanların kullandığı tabirlerdir. Ayırt edebilecek misiniz, bir bakın bakalım.

Tanrı	düşük faiz oranı	vergi sonrası
söz	ödeyeceğim	hastane
borçsuz	mezun	
asgari ödeme	teşekkür ederim	

Nazik, dindar bir insanın söz verdiğinde büyük ihtimalle borcunu ödeyeceğini düşünebilirsiniz ya da en azından umut edebilirsiniz. Ama gerçekte durum böyle

değildir. Veriler, bu tür insanların borçlarını ödeme ihtimallerinin ortalamanın altında olduğunu gösteriyor.

İşte borcun ödenmesi ihtimalinin düşük ya da yüksek olmasına göre değişen tabirler:

Borcunu Ödemesi En Muhtemel İnsanların Borç İsterken Kullandıkları Tabirler		
borçsuz	vergi sonrası	mezun
düşük faiz oranı	asgari ödeme	

Borcunun Üzerine Yatması En Muhtemel İnsanların Borç İsterken Kullandıkları Tabirler		
Tanrı	ödeyeceğim	hastane
söz	teşekkür ederim	

Bu çalışmanın etik içerimlerini ele almadan önce, yazarlarının yardımıyla bu çalışmanın insanlar hakkında neler anlattığına bakalım. Farklı kategorilerdeki bu kelimelerden ne anlam çıkarmamız gerekir?

Birincisi, bir insanın borcunu ödeme ihtimalinin yüksek olduğunu gösteren kelimelere bakalım. "Düşük faiz oranı" veya "vergi sonrası" gibi tabirler borç alan kişinin finansal açıdan belli bir incelik gösterdiğine işaret ediyor, dolayısıyla borcunu geri ödemesi muhtemel biriyle bağıntılı olmasında şaşılacak bir durum yoktur. Ayrıca bir üniversite "mezun"u olmak ve "borçsuz" olmak gibi olumlu koşullardan bahsediyorsa, borcunu ödeme ihtimali de yüksektir.

Ardından, borcunu ödeme ihtimali düşük olan birinin kullandığı dili düşünelim. Genel olarak, eğer birisi size borcunu ödeyeceğini söylerse, ödemeyecektir. Bir söz ne kadar iddialıysa, tutulmama ihtimali o kadar yüksektir. Eğer birisi, "Söz veriyorum ödeyeceğim, Tanrı aşkına lütfen yardım edin," diye yazıyorsa, ödeme ihtimali en düşük olan kişilerden birine denk geldiniz demektir. Sizin insafınıza seslenmesi ("hastane"de akrabası olduğu için paraya ihtiyaç duyduğunu söylemesi) de borcunu ödemeyeceği anlamına gelir. Aile fertlerinden (karı, koca, anne-baba ya da çocuk) bahsetmek o kişinin borcunu geri ödemeyeceğine işarettir. Borcun

üzerine yatılacağını gösteren bir diğer kelime de "açıklama"dır: Eğer insanlar neden borcu ödeyebileceklerini açıklamaya çalışıyorlarsa, muhtemelen ödemeyeceklerdir.

Araştırmayı yürüten yazarların, teşekkür etmenin neden borcun ödenemeyebileceğinin işareti olduğuna dair bir teorileri yoktu.

Özetlersek, bu araştırmacılara göre ödemenizi nasıl yapacağınıza dair ayrıntılı bir plan sunmak ve geçmişte sözünüzü tuttuğunuzdan bahsetmek borcun ödeneceğinin kanıtlarıdır. Sözler vermek ve birilerinin insafına sığınmak ise ödenmeyeceğinin açık bir işaretidir. Nedenler ne olursa olsun (ya da sözler vermenin bir insanın aslında bir şeyi yapmayacağının kesin işareti olmasının insan doğası hakkında bize ne anlattığından bağımsız olarak) araştırmacılar bu testin borcun ödenmeyeceğini kestirme konusunda son derece değerli bir bilgi olduğunu tespit ettiler. Tanrı'dan bahseden birinin borcunun üzerine yatma ihtimali 2,2 kat daha fazlaydı. Birinin borcunu ödemeyeceğine dair en güçlü işaretlerden biri Tanrı'nın adını anmasıydı.

Yazarlar, çalışmalarının etik soruları da gündeme getirdiğine inanıyorlar. Bu sadece akademik bir çalışma olmakla birlikte, bazı şirketler borçları onaylarken internetteki verilerden yararlandıklarını açıkça belirtiyorlar. Böyle bir şey kabul edilebilir mi? Şirketlerin, bir insanın borcuna sadık olup olmayacağını tahmin etmek için, borç isterken kullandığı kelimelere baktıkları bir dünyada yaşamak istiyor muyuz? Bu, en hafif tabirle, rahatsız edici ve korkutucudur.

Borç para almak isteyen bir tüketici yakın gelecekte sadece finansal geçmişiyle değil, internet faaliyetleriyle de değerlendirileceğini bildiğinden, endişeleri katlanabilir. Dahası abes görünen etkenlerle yargılanabilir: Mesela, "Teşekkür ederim," diyor mu ya da "Tanrı"dan dem vuruyor mu? Ayrıca gerçekten de hastanedeki kardeşine yardım etmesi gereken ve sonrasında borcunu kesinlikle ödeyecek olan bir kadın ne olacak? Ortalamaya vurulduğunda, hastane masrafları için yardım istediğini iddia eden insanların genellikle yalan söylüyor olmalarından ötürü böyle birinin cezalandırılması korkunç bir şey olurdu. İşlerin bu şekilde yürüdüğü bir dünya distopik bir hal alacaktır.

Etik soru şudur: Şirketlerin, sundukları hizmetlere uygun olup olmadığımızı bu hizmetlerle doğrudan ilintili olmayan soyut ama istatistiki açıdan öngörücü kıstaslara dayandırma hakları var mıdır?

Finans dünyasını bir kenara bırakıp, örneğin işe alma pratikleri açısından barındırdığı tehlikelere bakalım. Patronlar çalışan adaylarını değerlendirirken sosyal medya profillerine artık daha fazla bakıyorlar. Eğer müstakbel çalışanlarının önceki patronlarını kötülediğine ya da önceki patronlarının sırlarını afişe ettiğine dair kanıtlar arıyorlarsa buradan etik sorular doğmaz, hatta Facebook ya da Instagram

paylaşımları, aşırı alkol aldığı anlaşılan birinin işe alınmamasını meşrulaştırabilir bile. Peki, ya önem verdikleri bir şeyle bağıntısı olan, görünüşte zararsız bir işaret bulurlarsa?

Cambridge Üniversitesi'nden ve Microsoft'tan araştırmacılar ABD'deki 58 bin Facebook kullanıcısına kişilikleri ve zekâlarıyla ilgili çeşitli testler yaptılar. Facebook beğenilerinin çoğu zaman IQ ile, dışadönüklükle ve vicdanla bağıntılı olduğunu tespit ettiler. Örneğin Facebook'ta Mozart'ı, gökgürültülü fırtınaları ve kıvırcık patates kızartmasını beğenenlerin IQ'su genellikle daha yüksek oluyor. Harley-Davidson motosikletleri, country müzik grubu Lady Antebellum'u ya da "Anne Olmayı Seviyorum" sayfasını sevenlerinse IQ'su genellikle daha düşük oluyor. Bu bağıntılardan bazıları boyutsallık lanetinden kaynaklanıyor olabilir. Eğer yeterince test yapılsa, bazılarının tesadüfen bağıntılı olduğu anlaşılacaktır. Ama bazı ilgi alanları gerçekten de IQ ile bağıntılı olabilir.

Gelgelelim eğer Harleyleri seven akıllı bir insan farkında olmadan düşük zekâ belirtisi gösterdiği için yetenekleriyle uyumlu bir işe giremiyorsa, burada adaletsiz bir durum olduğu söylenebilir.

Aslında bu tümüyle yeni bir problem de değildir. İnsanlar uzun zamandır doğrudan iş performanslarıyla alakalı olmayan etkenlerle değerlendiriliyorlar: Ne kadar sıkı tokalaştıkları, kıyafetlerinin düzgün olup olmadığı vb. Ama veri devriminin tehlikeli yanlarından biri, hayatımızın daha büyük kısmı nicelleştirilirken, bu ikame yargıların bir yandan daha üstü kapalı ama diğer yandan daha mütecaviz bir hal alabilmesi. Daha iyi tahminler daha incelikli ve daha çirkin ayrımcılıklara yol açabilir.

Öte yandan, daha iyi veriler başka bir tür ayrımcılığa daha, iktisatçıların fiyat ayrımcılığı adını verdikleri şeye yol açabilir. İşletmeler çoğu zaman sundukları hizmet ya da mal karşılığında hangi fiyatı talep etmeleri gerektiğini bulmaya çalışırlar. İdeal haliyle müşterilerinin ödemek istediğinin azamisini almak isterler. Bu yolla mümkün olan azami kârı elde edeceklerdir.

Çoğu işletme genellikle herkesin ödediği tek bir fiyatta karar kılar. Ama bazen belli bir grubun üyelerinin, ortalamaya vurulduğunda, daha fazla ödeyeceğinin farkındadırlar. Tam da bu nedenle sinema salonları kazanma gücünün zirvesindeki orta yaşlı müşterilere; öğrencilere ya da yaşlılara olduğundan daha fazla fiyat çekerler veya havayolları son dakikada bilet alanlardan çoğu zaman daha fazla para talep ederler. Fiyatla ayrımcılık yaparlar.

Büyük Veri işletmelere müşterilerin ne fiyat ödemeye hazır olduğunu çok daha iyi öğrenmelerinde ve dolayısıyla belli gruptan insanları değerlendirmelerinde yar-

dımcı olabilir. Optimal Kararlar Grubu veri biliminden yararlanarak müşterilerin sigorta için ne kadar ödemeye hazır olduklarını tahmin etme konusunda öncü bir kuruluştu. Bunu nasıl yapıyorlardı? Bu kitapta daha önce ele aldığımız bir metodolojiden yararlanarak. Halihazırda sigorta yaptırmak isteyen müşterilere en çok benzeyen eski müşterileri buluyor ve prim tutarının ne kadar üstünde ödemeye hazır olduklarına bakıyorlardı. Başka bir deyişle, bir ruh ikizi araması yapıyorlardı. Bir ruh ikizi araması eğer bir beyzbol oyuncusunun eski şaşaalı günlerine dönüp dönemeyeceğini tahmin etmemize yardımcı olursa eğlencelidir. Bir ruh ikizi araması, eğer birinin hastalığını tedavi etmemize yardımcı olursa müthiştir. Peki, ya bir ruh ikizi araması bir şirketin cebinizdeki son kuruşu nasıl alacağını öğrenmesine yardımcı olursa? Bu pek de hoş olmaz.

Savurgan kardeşime şayet benim gibi cebinde akrep olan birinden daha fazla fiyat çekseler şikâyet etmeye hakkı olurdu.

Kumar, müşterilere yakından bakabilmenin potansiyel açıdan tehlikeli olduğu alanlardan biridir. Büyük kumarhaneler müşterilerini daha iyi anlamak için bir tür ruh ikizi aramasından yararlanıyorlar. Hedefleri nedir? Mümkün olan azami kârı elde etmek, cebinizdeki parayı kasalarına aktarabilmek.

Mekanizma şöyle işliyor. Kumarhaneler her kumarbazın bir "acı eşiği" olduğuna inanırlar. Acı eşiği, kişinin uzunca bir süre kumara tövbe etmesini sağlayacak kadar korkutan kayıp miktarıdır. Örneğin Helen'ın "acı eşiği"nin 3 bin dolar olduğunu varsayalım. Bu demektir ki, eğer Helen 3 bin dolar kaybederse, kumarhane bir müşterisini kaybetmiş olur. Eğer Helen 2.999 dolar kaybederse, mutlu olmayacaktır. Sonuçta kim para kaybetmek ister ki? Ama morali ertesi gece yeniden gelmesini engelleyecek kadar bozulmayacaktır.

Bir an kumarhane işletmecisi olduğunuzu hayal edin. Helen da oyun makinelerinde kumar oynamaya gelmiş olsun. Optimal sonuç nedir? Kuşkusuz, Helen'ın acı eşiğini geçmeden bu noktaya olabildiğince yaklaşmasını istersiniz. Onun 2.999 dolar kaybetmesini ve büyük kârlar elde etmeyi ama bir süre ortadan kaybolmasına yol açacak kadar kaybetmemesini istersiniz.

Bunu nasıl yapabilirsiniz? Helen'ın belli bir miktar kaybettiği anda oynamayı bırakmasını sağlamanın yolları vardır. Örneğin ona yemek ısmarlayabilirsiniz. Bu teklifi yeterince çekici kılarsanız, makinenin başından kalkıp yemeğe oturabilir.

Ama bu yaklaşımın da bir zorluğu vardır. Helen'ın "acı eşiği"ni nereden bileceksiniz? Sorun şu ki, insanların acı eşikleri farklı farklıdır. Helen için 3 bin dolarken, John için 2 bin dolar olabilir. Scott için ise 26 bin dolardır belki. Eğer Helen'ı 2 bin dolar kaybettiğinde kumar oynamaktan vazgeçirirseniz, kârı masada bırakmış

olursunuz. Eğer çok fazla (3 bin dolardan fazla kaybedene kadar) beklerseniz, onu bir süreliğine elinizden kaçırabilirsiniz. Ayrıca Helen size acı eşiğini söylemek istemeyebilir, hatta kendi de bu eşiği bilmiyor olabilir.

O halde, ne yapmalı? Eğer kitapta bu noktaya kadar geldiyseniz, muhtemelen cevabı biliyorsunuz. Veri biliminden yararlanırsınız. Bazı müşterileriniz hakkında öğrenebileceğiniz her şeyi öğrenirsiniz: Yaşları, cinsiyetleri, posta kodları, kumar alışkanlıkları. Ve bu kumar alışkanlığından (kazançları, kayıpları, geliş ve gidişlerinden) hareketle onların "acı eşiği"ni belirlersiniz.

Helen hakkında bildiğiniz tüm bilgileri toplar ve ona benzeyen kumarbazları, onun az çok ruh ikizi olan kumarbazları bulursunuz. Ardından, bu kişilerin acıya ne kadar dayanıklı olduklarını bulursunuz. Muhtemelen Helen'ınkiyle aynıdır. Nitekim Harrah'nın kumarhanesinin bir Büyük Veri şirketi olan Terabyte'tan yardım alarak yaptığı da budur.

Terabyte'ın genel müdürü olan Scott Gnau *Super Crunchers* adlı muhteşem kitapta, kumarhane yöneticilerinin, acı eşiğine yaklaşmakta olan düzenli bir müşteri gördüklerinde ne yaptıklarını şöyle açıklıyor: "Gelip şöyle derler: 'Gördüğüm kadarıyla kötü bir gün geçiriyorsunuz. Bizim biftekçiyi çok sevdiğinizi biliyorum. Alın şunu eşinizi yemeğe götürün, bizdensiniz'."

"Alicenaplığın da bu kadarı" diyebilirsiniz: Bedavaya biftek ziyafeti ha! Ama aslında bu ikramla kendisine iyilik yapmış olur. Kumarhane tam da müşterisinin uzun süre kumarhaneye ayak basmamasına yol açacak kadar kaybetmesini engellemek için oyunu bırakmasını sağlamaya çalışır. Başka bir deyişle, yönetim, müşterilerinden uzun vadede olabildiğince fazla para koparabilmek için incelikli bir veri analizinden yararlanır.

İnternet verilerinden her geçen gün daha da verimli bir biçimde yararlanılmasının, kumarhanelerin, sigorta şirketlerinin, kredi kuruluşlarının ve diğer şirketlerin bize karşı çok daha fazla yetki ve güç sahibi olmalarına yol açmasından korkmamız gayet doğal.

Ama diğer yandan, Büyük Veri müşterilere fazla fiyat çeken ya da kötü ürünleri kakalayan işletmelere darbe vurma imkânı da sunuyor.

Önemli silahlardan biri, Yelp gibi restoranlarla ve başka hizmet sektörleriyle ilgili değerlendirme ve yorumları yayınlayan sitelerdir. Harvardlı iktisatçı Michael Luca'nın yakın dönemde yayımlanan çalışması, işletmelerin Yelp değerlendirmelerinin insafına kaldığını gösteriyor. Bu değerlendirmeleri Washington eyaletindeki satış verileriyle karşılaştıran Luca, Yelp'te bir yıldız az olmasının restoranın gelirlerinde yüzde 5 ila 9 düşüşe yol açtığını tespit etmiş.

Tüketiciler alışveriş sitelerini (Kayak ve Booking.com gibi) karşılaştırarak da işletmelerle olan mücadelelerinde ellerini güçlendiriyorlar. *Freakonomics*'te ele alındığı üzere, bir internet sitesi farklı şirketlerin hayat sigortası için istediği fiyatları yayınlamaya başladığında, bu fiyatlar çarpıcı şekilde düşüyordu. Eğer bir sigorta şirketi fazla fiyat çekiyorsa, müşteriler bunu biliyor ve başka bir şirkete başvuruyorlar. Müşteriler için toplam tasarruf miktarı nedir? Yılda bir milyar dolar.

Başka bir deyişle, internetteki veriler işletmelere hangi müşterilerden kaçınıp hangilerini sömürebileceklerini gösterebilir. Ama müşterilere de hangi işletmelerden uzak durmaları gerektiğini ve hangilerinin onları sömürmeye çalıştığını söyleyebilir. Büyük Veri bu zamana dek tüketiciler ile şirketler arasındaki mücadelede iki tarafa da yardımcı olmuştur. Bunun adil bir kavga olarak devam etmesini sağlamamız gerekiyor.

Hükümetlere Ek Yetkiler Tanımanın Barındırdığı Tehlike

Adriana Donato, eski erkek arkadaşı bir doğum günü partisine geldiğinde, onun üzgün olduğunu biliyordu. Eski aşkının ciddi psikolojik sorunları olduğunu biliyordu. Kaç zamandır bunalımda olduğunu biliyordu. Adam onu arabayla gezmeye davet ettiğinde, zooloji bölümünde okuyan bu yirmi yaşındaki kadının bilmediği bir şey vardı. Eski erkek arkadaşı, yirmi iki yaşındaki James Stoneham'ın son üç haftadır internette birinin nasıl öldürüleceğiyle ve cinayet suçunun cezalarıyla ilgili aramalar yaptığını ve ara ara da Donato'yla ilgili aramalar yaptığını bilmiyordu.

Eğer bunu bilseydi, muhtemelen o arabaya binmezdi ve muhtemelen o akşam bıçaklanarak öldürülmezdi.

Azınlık Raporu adlı filmde psikologlar suçları gerçekleşmeden önce durdurmak için polisle ortaklaşa hareket ederler. Büyük Veri polisin suçlar gerçekleşmeden önce harekete geçmesini sağlamak için polise verilmeli midir? Donato'ya en azından eski erkek arkadaşının tehlikeli aramalarıyla ilgili bir uyarıda bulunulabilir miydi? Polisin Stoneham'ı sorgulaması gerekir miydi?

Öncelikle, suç faaliyetleriyle alakalı Google aramaları ile suç faaliyetleri arasında bağıntı olduğuna dair kanıtların giderek arttığını kabul etmek gerekir. Christine Ma-Kellams, Flora Or, Ji Hyun Baek ve Ichiro Kawachi, intiharla alakalı Google aramalarının eyalet düzeyindeki intihar oranlarıyla güçlü bir bağıntısı olduğunu göstermiştir. Ayrıca Evan Soltas ile ben, haftalık İslam düşmanı aramaların ("Müslümanlardan nefret ediyorum" ya da "Müslümanlara ölüm" gibi), o hafta işlenen Müslümanlara yönelik nefret suçlarıyla bağıntısı olduğunu gösterdik. Bir şey yapmak istediğine dair arama yapan kişi sayısı, o şeyi yapacak olanların sayısını yansıtmaktadır.

O halde, bu bilgiyi nasıl değerlendirmeliyiz? Basit ve herkesçe kabul edildiği söylenebilecek bir fikir: Bölgesel düzeydeki verilerden kaynak aktarımı için yararlanabiliriz. Eğer bir şehirde intiharla alakalı aramalarda büyük bir artış varsa, bu şehirdeki intihar farkındalığını güçlendirebiliriz. Şehirdeki yöneticiler ya da kâr amacı gütmeyen kuruluşlar, örneğin insanların nereden yardım alabileceğine dair reklamlar hazırlayabilirler. Benzer şekilde, eğer bir şehirde "Müslümanlara ölüm" aramalarında müthiş bir artış varsa, polis sokaklarda devriye gezme tarzını değiştirebilir. Örneğin bölgedeki camileri korumak için daha fazla polis görevlendirilebilir.

Ama bir adım var ki, o adımı atmaya hiç yanaşmamamız gerekir: Bir suç işlenmeden önce bireylerin peşine düşmek. Öncelikle bu durum özel hayatın ihlali kapsamına girmektedir. Hükümetin binlerce, yüz binlerce insanın arama verilerine sahip olması ile bir bireyin arama verilerinin polisin elinde olması arasında etik açıdan devasa bir uçurum vardır. Bölgedeki bir camiyi korumak ile birinin evine baskın yapmak arasında etik açıdan devasa bir uçurum vardır. İntiharı engellemek için reklam yayınlamak ile birini rızası hilafına akıl hastanesine kapamak arasında etik açıdan devasa bir uçurum vardır.

Gelgelelim bireysel düzeydeki verilerden yararlanırken son derece ihtiyatlı olma zorunluluğunun nedeni etiğin de ötesine geçmektedir. Bunun verisel bir nedeni de vardır. Veri biliminin bir şehirdeki eylemleri kestirmeye çalışması ile bir bireyin eylemlerini kestirmeye çalışması arasında büyük bir uçurum vardır.

İntihar meselesine geri dönelim. Google'da her yıl ABD sınırları içinde intiharla ilgili 3,5 milyon arama yapılıyor ve bu aramaların çoğunluğu "intihar eğilimli", "intihar etmek", "nasıl intihar edilir?" gibi intiharla alakalı fikirlerdir. Başka bir deyişle, ABD'de her ay yüz Amerikalıdan en az birinin intihar araması yaptığını görüyoruz. Bu durum filozof Friedrich Nietzsche'nin bir sözünü akla getiriyor: "İntihar düşüncesi müthiş bir tesellidir: İnsan bu sayede birçok karanlık geceyi atlatabilir." Google arama verileri bunun ne kadar doğru olduğunu, intihar düşüncesinin ne kadar yaygın olduğunu gösteriyor. Gelgelelim her ay ABD'de işlenen intihar sayısı dört binden azdır. İntihar fikri gerçekten de inanılmaz yaygındır, ama intiharın kendisi değildir. Dolayısıyla internette beynini patlatmak istediğine dair epey bir gürültü koparan herkesin kapısına bir polis dikmenin pek bir anlamı yoktur – zaten polisin başka işi mi yok?

Veya şu iğrenç İslamofobik aramaları düşünelim. 2015'te ABD'de "Müslümanlara ölüm" aramalarının sayısı kabaca 12 bindi. Nefret suçu kapsamında öldürülen Müslüman sayısıysa 12'ydi. Belli ki bu korkunç aramaları yapan insanların ezici çoğunluğu bu aramanın gereği olan eylemi gerçekleştirmiyor.

Bir bireyin davranışını kestirmek ile bir şehirdeki davranışları kestirmek arasındaki farkı açıklayan bir matematik var. İşte size basit bir düşünce deneyi. Bir şehirde bir milyon insan ve bir tane cami olduğunu varsayalım. Eğer birisi "Müslümanlara ölüm" araması yapmıyorsa, onun bir camiye saldırma ihtimalinin 100 milyonda bir olduğunu varsayalım. Eğer birisi "Müslümanlara ölüm" araması yapıyorsa, bu ihtimalin katlanarak artıp 10 binde 1'e yükseldiğini varsayalım. İslamofobinin ışık hızıyla yayıldığını ve "Müslümanlara ölüm" aramalarının 100'den 1.000'e çıktığını varsayalım.

Bu durumda, matematik bize bir camiye saldırı ihtimalinin yaklaşık beş kat artıp yüzde 2 dolaylarından yüzde 10'a yükseldiğini gösteriyor. Ama "Müslümanlara ölüm" araması yapan bir bireyin gerçekten bir camiye saldırı düzenleme ihtimali hâlâ 10 binde 1'dir.

Bu durumda doğru yanıt "Müslümanlara ölüm" araması yapan herkesi hapse tıkmak değildir. Keza evlerinde arama yapmak da doğru olmaz. Bu insanlardan herhangi birinin bir suça bulaşma ihtimali az da olsa vardır. Fakat doğru yanıt, şu an saldırıya maruz kalma ihtimali yüzde 10 olan o camiyi korumak olacaktır.

Şurası açık ki, çok sayıda korkunç arama asla korkunç eylemlere yol açmıyor.

Diğer yandan, şunu da belirtmek lazım, bazı aramaların akabinde korkunç eylemlerin gerçekleşme olasılığının son derece yüksek olması da en azından teorik açıdan mümkündür. Örneğin veri bilimcilerin gelecekte Stoneham'ın Donato'yla ilgili aramalarının ciddi bir kaygı sebebi olduğunu ortaya koyabilecek bir model inşa etmeleri en azından teorik açıdan mümkündür.

2014'te "Kız arkadaşımı nasıl öldürürüm?" şeklinde birebir aynı olan 6 bin kadar arama yapıldı ve 400 kadın cinayeti işlendi. Eğer bu katillerin hepsi öncesinde birebir aynı aramayı yaptıysa, bu demektir ki, "Kız arkadaşımı nasıl öldürürüm?" şeklindeki her 15 aramadan 1'i eyleme dönüşmüştür. Elbette kız arkadaşını öldüren erkeklerin çoğu muhtemelen birebir aynı aramayı yapmamıştır. Dolayısıyla bu aramanın cinayete dönüşme olasılığı gerçekte daha düşüktür, muhtemelen çok daha düşüktür.

Ama eğer veri bilimcileri belli bir bireye karşı tehdidin sözgelimi 100'de 1 olduğunu gösteren bir model inşa edebilirlerse, bu bilgiyle bir şeyler yapabiliriz. En azından tehdit altındaki insanın falanca kişi tarafından öldürülme ihtimalinin 100'de 1 de olsa var olduğunu bilmek o kişinin hakkıdır.

Gelgelelim toplama vurulduğunda, bireysel düzeyde suç tahmininde bulunmak için arama verilerinden yararlanmakta çok ihtiyatlı davranmamız gerekir. Veriler birçok ama birçok korkunç aramanın nadiren korkunç bir suça vardığını açıkça

gösteriyor. Ayrıca hükümetin sırf bu verileri inceleyerek korkunç bir eylemi yüksek bir olasılıkla kestirebileceğini gösteren kanıtlara henüz sahip değiliz. Dolayısıyla hükümetin arama verileri temelinde bireylere müdahale etme hakkına izin verme konusunda gerçekten ihtiyatlı olmamız gerekir: Sadece etik ya da hukuki nedenlerden ötürü değil, en azından şimdilik veri bilimiyle alakalı nedenlerden ötürü de böyle olmalıdır.

SONUÇ

Kaç İnsan Başladığı Kitabı Bitirebiliyor?

Kitabımın sözleşmesini imzaladığımda, yapısının nasıl olacağını kafamda net bir şekilde canlandırabiliyordum. Başlara doğru, hatırlarsanız, ailemle Şükran Günü yemeğinden bir sahneyi anlatmıştım. Aile fertlerim benim deli olup olmadığımı tartışıyor ve otuz üç yaşıma gelmeme rağmen hâlâ neden doğru kızı bulamadığımı anlamaya çalışıyorlardı.

Dolayısıyla bu kitabın sonuç kısmı kendiliğinden kucağıma düşmüş oldu. O kızla tanışıp evlenecektim. Hatta ve hatta, doğru kızla tanışmak için Büyük Veri'den yararlanacaktım. Belki de flört kısmında nasıl hamleler yaptığımı parça parça anlatırdım. Sonra kitabın sonuç kısmında mutlu sona bağlar, düğünümü anlatır, kitabı da bir tür aşk mektubu gibi eşime takdim ederdim.

Ne yazık ki hayat benim fantezime izin vermedi. Eve kapanıp kitap yazmak için dış dünyayla bağları koparmam, sanırım, duygusal hayatımı olumsuz etkiledi. Ve ne yazık ki hâlâ bir hayat arkadaşı bulamadım. En önemlisi, yeni bir sonuç kısmı yazmam gerekiyor.

Müthiş bir sonuç kısmını nasıl yazacağımı bulmak için en sevdiğim kitaplara göz gezdirdim. Kitap boyunca zaten bahsedip durduğum, yüzeyin altında bekleyen şeyi üste çıkarmanın en iyi sonuç olduğu sonucuna vardım. Bu kitabın büyük meselesi şudur: Sosyal bilimler gerçek bir bilim haline geliyor. Ve bu yeni gerçek bilim hayatlarımızı iyileştirmeye hazır ve aç halde.

İkinci Kısım'ın başında Karl Popper'ın Sigmund Freud'a eleştirisini ele almıştım. Popper'ın Freud'un eksantrik dünya anlayışını bilimsel bulmadığını belirtmiştim. Ama Popper'ın eleştirisinde bir kısımdan bahsetmedim, zira söyledikleri aslında Freud'a yönelik bir saldırıdan çok daha fazlasıydı. Popper hiçbir sosyal bilimciyi tam anlamıyla bilimsel bulmuyordu. Popper, bu sözde bilimcilerin yapıp ettikleri şeylerin kesinliğinden etkilenen biri değildi.

Popper'ın bu taarruzunun arkasındaki saik neydi? Popper çağının en iyi aydınlarıyla (fizikçiler, tarihçiler, psikologlarla) sohbet ettiğinde, çarpıcı bir fark tespit etmişti. Fizikçiler konuştuklarında, Popper onların yapıp ettiklerine inanıyordu.

Kuşkusuz yanlış yaptıkları da oluyordu. Kuşkusuz bazen bilinçdışı önyargılarının esiri olup çuvallıyorlardı. Ama fizikçiler dünyanın derin hakikatlerini bulan ve nihayetinde Einsetin'ın İzafiyet Teorisi'yle taçlanan bir sürecin içindeydiler. Oysa dünyanın en ünlü sosyal bilimcisi konuştuğunda, Popper salak saçma laflar dinlediğini düşünüyordu.

Popper bu ayrımı yapan tek insan değildir. Hemen herkes fizikçilerin, biyologların ve kimyacıların gerçek bilimciler olduğunu kabul eder. Maddi dünyanın nasıl işlediğini bulmak için sıkı deneyler yaparlar. Halbuki birçok insan iktisatçıların, sosyologların ve psikologların inandırıcı olmak için anlamsız kelimeler kullanan yumuşak bilimciler olduğunu düşünür.

Bunun doğru olduğunu varsaysak bile, Büyük Veri devrimi her şeyi değiştirmiştir. Eğer Karl Popper bugün yaşasa ve Raj Chetty, Jesse Shapiro, Esther Duflo ya da benim (aman ne komiğim!) bir sunumumuza katılsaydı, o zamanki tepkisini vereceğini hiç sanmıyorum. Dürüst olmak gerekirse, günümüzün büyük sicim teorisyenleri gerçekten bilimsel mi yoksa kendilerini tatmin etmek için zihin jimnastiği mi yapıyorlar, bence muhtemelen bunu sorgulardı.

Eğer bir şehirde bir şiddet filmi gösterime giriyorsa, suç sayısı artar mı azalır mı? Eğer daha çok insan reklama maruz kalırsa, belli bir ürünün satışları artar mı azalır mı? Eğer bir beyzbol takımı bir oğlan yirmi yaşındayken şampiyon olursa, çocuğun kırkına geldiğinde o takımı tutma ihtimali daha yüksek midir? Bunlar çok açık sorulardır, cevapları da çok açıktır, evet ya da hayır! Ve dürüst veriler dağında bu cevapları bulabiliriz.

Bu sahte bilim değil, gerçek bilimdir.

Buradan sosyal bilimlerdeki devrimin basit, tarihüstü yasalar şeklinde geleceği sonucu çıkmaz.

MIT'den müteveffa bilimci Marvin Minsky yapay zekânın mümkün olup olmadığını araştırmış ilk isimlerden biriydi ve psikolojinin fiziği taklit etmeye çalışarak yolunu kaybettiğini savunmuştu. Fizik her zaman her yerde geçerli olan basit yasalar bulmayı başarmıştı.

Minsky insan beyninin bu tür yasalara tabi olamayabileceğini söylemişti. Beyin daha ziyade karmaşık bir kod kırma sistemi gibiydi: Parçalardan biri diğer parçaların yanlışlarını düzeltiyordu. İktisat ve siyasal sistem de benzer şekilde karmaşık olabilir.

Bu nedenle sosyal bilimlerdeki devrimin $E = MC^2$ şeklinde kesin formüller biçimine bürünmesi muhtemel değildir. Aslında eğer birisi sosyal bilimlerdeki devrimin kesin bir formüle dayanacağını iddia ederse, şüpheyle yaklaşın derim.

Bu devrim parça parça, çalışma çalışma, bulgu bulgu gerçekleşecek. İnsan zihninin ve toplumunun karmaşık sistemlerini yavaş yavaş daha iyi anlamayı başaracağız.

* * *

Düzgün bir sonuç yazısı özet mahiyetindedir, ama daha fazlasına da işaret eder.

Bu kitapta bunu yapmak kolay. Burada ele aldığım veri kümeleri devrimcidir, ama henüz çok az irdelenmiştir. Bunlardan öğrenilecek daha çok şey var. Açık konuşayım, akademisyenlerin ezici çoğunluğu dijital çağın getirdiği veri patlamasını görmezden geliyor. Dünyanın en ünlü seks araştırmacıları denenmiş ve doğru olana tutunmakta ısrar ediyorlar. Birkaç yüz deneğe arzularını soruyorlar; PornHub gibi sitelerden veri istemiyorlar. Dünyanın en ünlü dilcileri tek tek metinleri analiz ediyorlar; milyarlarca kitapta kendini gösteren örüntüleri büyük oranda göz ardı ediyorlar. Psikolojide, siyaset biliminde ve sosyolojide lisans öğrencilerine öğretilen metodolojiler, genel itibariyle, dijital devrimden etkilenmemiştir. Veri patlamasının açtığı ve büyük oranda hâlâ bakir olan alan az sayıdaki ileri görüşlü profesöre, dik kafalı lisans öğrencisine ve amatöre terk edilmiştir.

Bu durum er geç değişecek.

Bu kitapta bahsettiğim her fikre karşı, ele alınmayı bekleyen bir o kadar önemli yüz fikir var. Burada ele alınan araştırma, buzdağının görünen ucunun görünen ucudur, o kadar.

O halde, bizi neler bekliyor?

Bir defa, gelmiş geçmiş en başarılı kamu sağlığı çalışmalarından birinde kullanılmış olan metodoloji daha farklı ve geniş bir ölçekte kullanılacak. 19. yüzyılın ortasında Britanyalı hekim John Snow Londra'da kolera salgının nedenini bulmaya çalışıyordu.

Dâhiyane fikri şuydu: Şehirdeki her bir kolera vakasının haritasını çıkarmıştı. Bunu yaparken, hastalığın bilhassa bir su pompasının etrafında toplandığını tespit etti. Demek ki salgın mikroplu su yoluyla yayılıyordu ve dolayısıyla kötü hava yoluyla yayıldığını iddia eden o dönemki geleneksel görüş çürütülmüş oluyordu.

Büyük Veri (ve onun sağladığı yakından bakma olanağı) bu tür bir çalışmayı bugün çok kolaylaştırıyor. İstediğimiz salgını Google arama verileriyle ya da diğer dijital sağlık verileriyle araştırabiliriz. Dünyanın herhangi bir bölgesindeki küçük bir alanda bu salgının olağandışı yükseklikte ya da düşüklükte olup olmadığını bulabiliriz. O zaman bu yerlerin ortak özelliğini görebiliriz. Havada mı bir şey var? Su mu? Toplumsal normlar mı?

Bunu migren için de yapabiliriz. Böbrek taşları için de yapabiliriz. Anksiyete, depresyon, Alzheimer, pankreas kanseri, yüksek kan basıncı, sırt ağrısı, kabızlık, burun kanaması için de yapabiliriz. Her şey için yapabiliriz. Snow'un bir kez yaptığı analizi biz bugün dört yüz kez yapabiliriz (bunu yazarken halihazırda bir tanesine başladım).

Basit bir yöntemi alıp bir analizi kısa sürede birkaç yüz kez yapmak için Büyük Veri'den yararlanarak, buna ölçekli bilim adını verebiliriz. Evet, sosyal ve davranışsal bilimler kesinlikle ölçeklenecek. Sağlık koşullarına yakından bakmak bu bilimlerin ölçeklenmesine yardımcı olacaktır. Ölçeklenmeye yardımcı olacak başka bir şey daha: A/B testleri. A/B testini işletmelerin kullanıcılarını başlıklara ve reklamlara tıklattırma çabası bağlamında ele almıştık; bugüne kadar söz konusu metodolojinin en yaygın kullanıldığı alan budur. Ama A/B testleri insanları bir reklama tıklamaya sevk eden ok işaretinden çok daha temel ve toplumsal açıdan değerli bir şeyleri açığa çıkarmak için kullanılabilir.

Northwestern Üniversitesi'nden iktisatçı Benjamin F. Jones, çocukların öğrenmelerine yardımcı olmak için A/B testlerinden yararlanmaya çalışıyor. Jones, okulların farklı ders müfredatlarını rastgele test etmelerini sağlayan EDU STAR adlı platformun oluşturulmasına yardımcı olmuştur.

Birçok şirket, eğitim yazılımı işine el atmış durumda. Öğrenciler EDU STAR'la bir bilgisayara giriyorlar ve karşılarına rastgele farklı ders müfredatları çıkıyor. Ardından, malzemeyi ne kadar iyi öğrendiklerini görmek için kısa testlere giriyorlar. Başka bir deyişle, okullar öğrencilerin malzemeyi kavramasına yardımcı olması için en iyi yazılımın hangisi olduğunu öğreniyorlar.

EDU STAR şimdiden, diğer bütün büyük A/B test platformları gibi, şaşırtıcı sonuçlar elde ediyor. Birçok eğitimciyi çok heyecanlandıran ders müfredatlarından biri, öğrencilere kesirleri öğretmek için oyunlardan yararlanan bir yazılıma dayanmaktadır. Kuşkusuz, eğer matematiği oyuna çevirirseniz, öğrenciler daha çok eğlenir, daha çok öğrenir ve testlerde daha iyi sonuçlar alırlar, öyle değil mi? Hayır! Kesirleri bir oyun üzerinden öğrenen öğrenciler, daha geleneksel tarzda öğrenen öğrencilerden daha az puan almışlardı.

Çocuklara daha fazla şey öğretmek Silikon Vadisi'nin insanların reklamlara daha fazla tıklamasını sağlamak için öncülük ettiği testin heyecan verici ve toplumsal açıdan faydalı bir kullanımıdır. Keza insanların daha fazla uyumasını sağlamak da öyle. ABD'de ortalama bir birey günde 6,7 saat uyuyor. Birçok Amerikalı daha fazla uyumak istiyor. Ama gece 11 olmuş, spor programı başlamış ya da YouTube seni çağırıyor. Haydi uyu uyuyabilirsen. Yüz binlerce müşterisi olan bir giyilebilir

teknoloji şirketi olan Jawbone müşterilerinin yapmak istedikleri şeyi (yatağa erken gitmek) yapmalarına yardımcı olacak müdahaleler bulmaya çalışan binlerce test yapıyor.

Jawbone iki aşamalı bir hedeften yararlanarak büyük bir zafer elde etti. İlkin, müşterilerinizden o kadar da hevesli olmadıkları bir hedefi seçmelerini isteyin. Onlara şöyle bir mesaj yollayın: "Anlaşılan son üç gündür fazla uyumuyorsun. Neden bu gece yatağa 11:30'da gitmiyorsun? Normalde sabahları 8'de uyandığını biliyoruz." Ardından kullanıcılara "varım" seçeneğine tıklama seçenekleri sunun.

İkincisi, saat gece 10:30'a geldiğinde, Jawbone başka bir mesaj daha yolluyor: "11:30'da uyumanı kararlaştırmıştık. Saat şu an 10:30. Neden şimdi uyumuyorsun?"

Jawbone bu stratejinin fazladan yirmi üç dakika uyku kazandırdığını tespit etti. Müşterilerinin 10:30'da yatağa gitmelerini sağlamıyordu belki, ama erken gitmelerini sağlıyordu.

Elbette bu stratejinin her bir parçası çok sayıda deneyle en iyi hale getirilmeliydi. İlk hedefi çok erken başlatın (kullanıcılara yatağa 11'de gitmelerini söyleyin), çok azının kabul edeceğini görürsünüz. Kullanıcılara gece yarısı yatmalarını söyleyin, o zaman da çok az şey kazanırsınız.

Jawbone, Google'ın sağa işaret eden okunun uyku muadilini bulmak için A/B testlerinden yararlandı. Ama Google'ın reklam ortaklarına fazladan birkaç tık kazandıracağına, pestili çıkmış Amerikalılara fazladan birkaç dakika daha istirahat kazandırdı.

Aslında bir bütün olarak psikoloji disiplini araştırmalarını çarpıcı derecede iyileştirmek için Silikon Vadisi'nin araçlarından yararlanabilir. Birkaç lisans öğrencisiyle yapılan birkaç deneyi ayrıntılarıyla ortaya koymak yerine, yüzlerce hızlı A/B testinin sonuçlarını gösteren ilk psikoloji yazısını dört gözle bekliyorum.

Aylarını az sayıda lisans öğrencisinin tek bir testi yapmalarını beklemekle geçiren akademisyenlerin cefası bitecek. Bunun yerine akademisyenler birkaç yüz ya da birkaç bin fikri sadece birkaç saniyede test etmek için dijital verilerden yararlanacaklar. Çok daha az sürede çok daha fazla şey öğreneceğiz.

Veri olarak metin bize çok şey öğretecek. Fikirler nasıl yayılır? Yeni kelimeler nasıl yaratılır? Kelimeler nasıl kaybolup gider? Şakalar nasıl oluşur? Neden bazı kelimeler eğlenceliyken diğerleri değildir? Lehçeler nasıl gelişir? İddiaya girerim, yirmi yıl içinde tüm bu sorular hakkında derin içgörülere sahip olacağız.

Kanaatimce çocukların nasıl öğrendiklerini ve geliştiklerini görmek için, geleneksel testlere takviye olarak, internetteki davranışlarından yararlanabiliriz (elbette

bu bilgilerin anonim kalması kaydıyla). Telaffuzları nasıl? Disleksi belirtileri gösteriyorlar mı? Olgun, entelektüel ilgi alanları geliştiriyorlar mı? Arkadaşları var mı? Çocukların her gün klavye başında geçirdikleri onca zamanda yapıp ettikleri tüm bu soruların ipuçlarını içeriyor.

Keza çok daha fazla içgörünün kaynağı olan, hiç de entipüften görülemeyecek bir alan daha var.

Rolling Stones'un "Shattered" şarkısında Mick Jagger, New York'u, nam-ı diğer Big Apple'ı (Büyük Elma) böyle sihirli yapan her şeyi öyle güzel anlatır ki. Kahkaha. Zevk. Yalnızlık. Sıçanlar. Tahtakuruları. Gurur. Açgözlülük. Kâğıt torbalarla dolaşan insanlar. Ama Jagger şehri gerçekten özel yapan şeyi defalarca tekrarlar: "Seks ve seks ve seks ve seks."

Büyük Elma için neyse Büyük Veri için de aynısı geçerlidir. Dijital devrim sayesinde sağlık alanından içgörüler geliyor. Uyuma. Öğrenme. Psikoloji. Dil. Artı, seks ve seks ve seks ve seks ve seks.

Şu an irdelediğim sorulardan biri: Cinselliğin kaç boyutu vardır? Bir insanı genellikle ya gey ya da heteroseksüel diye görürüz. Oysa geyler ve heteroseksüel bireyler arasında da türler vardır: Sözgelimi bazı erkekler sarışın severken bazıları esmer sever. Bu tercihler cinsiyet tercihleri kadar güçlü olabilir mi? Cevap aradığım başka bir soru: Cinsel tercihlerin kaynağı nedir? Nasıl ki beyzbol taraftarlığı ya da siyasal görüşleri belirleyen kilit yılları saptayabiliyorsak, aynı şekilde bugün yetişkin cinsel tercihlerini belirleyen kilit yılları da saptayabiliriz. Bu cevapları bulmak için *Herkes Hâlâ Yalan Söylüyor* adını vereceğim bir sonraki kitabımı da almanız gerekecek.

Pornonun varlığı ve pornoyla birlikte gelen veriler insan cinselliği biliminde devrim niteliğinde bir gelişmedir.

Doğa bilimlerinin yaşamlarımızı değiştirmeye başlaması, penisilini, uyduları ve bilgisayarları yaratması zaman almıştır. Büyük Veri'nin sosyal bilimlerde ve davranış bilimlerinde nasıl sevdiğimiz, yaşadığımız ve öğrendiğimiz konularında önemli atılımlar gerçekleştirmesi için de zamana ihtiyacımız olabilir. Ama bu tür ilerlemelere yaklaştığımıza inanıyorum. Bu gelişmeleri en azından ana hatlarıyla bu kitapta görebilmiş olduğunuzu umuyorum. Umarım gerçek hayatta bu kitabı okuyan bazıları bu atılımların yaratılmasına da yardımcı olur.

* * *

Bir yazar düzgün bir sonuç yazısına imza atabilmek için, öncelikle kitabı neden yazdığını düşünmelidir. Hangi hedefe ulaşmaya çalışmaktadır?

Benim bu kitabı yazmamın en önemli nedeninin hayatımın gelişim evresindeki deneyimlerimden biri olduğunu düşünüyorum. *Görünmeyen Ekonomi** kitabı on küsur yıl önce yayımlanmıştı. Şaşırtıcı bir şekilde çok satan bu kitap Chicago Üniversitesi'nden ödüllü bir iktisatçının, yukarıda da sık sık bahsi geçen Steven Levitt'in araştırmasını anlatıyordu. Levitt, "yaramaz bir iktisatçı"ydı. Kırk tilki dolaşan aklının sormayı akıl edebileceği her soruyu verilerden yararlanarak cevaplayabiliyor gibiydi: Sumo güreşçileri hile yapar mı? Yarışma programlarındaki yarışmacılara ayrımcılık yapılıyor mu? Emlakçılar ev alırken sizinle aynı fiyata mı alıyorlar?

Üniversiteyi yeni bitirmiştim; felsefe eğitimi almıştım, ama hayatta kendime nasıl bir yön çizeceğime dair dair en ufak bir fikrim yoktu. *Görünmeyen Ekonomi*'yi okuduktan sonra artık biliyordum. Steven Levitt'in yaptığını yapmak istiyordum. Veri dağlarına dalmak ve dünyanın gerçekte nasıl işlediğini bulmak istiyordum. Onun izinden gidip iktisat mastırı yapmaya karar verdim

Geçen on iki yılda o kadar çok şey değişti ki. Levitt'in bazı çalışmalarında kodlama hataları olduğu tespit edildi. Ayrıca Levitt küresel ısınma konusunda siyasi açıdan yanlış birtakım laflar etti. *Görünmeyen Ekonomi* entelektüel çevrelerde gözden düştü.

Ama birkaç yanlışı bir tarafa, yıllar Levitt'i haklı çıkardı. Levitt merak, yaratıcılık ve veriler bir araya geldiğinde dünyaya ilişkin kavrayışımızın çarpıcı şekilde ilerleyeceğini söylüyordu. Verilerde anlatılmayı bekleyen hikâyeler saklıydı; bunun doğruluğu defalarca kanıtlanmıştır.

Bu kitabın okurları üzerindeki etkisinin de *Görünmeyen Ekonomi*'nin benim üzerindeki etkisi kadar güçlü olmasını umuyorum. Umarım hayatta ne yapması gerektiği konusunda kafası karışık bir genç şu an bu kitabı okuyordur. Eğer istatistiğe biraz yeteneğiniz, epey bir yaratıcılığınız ve merakınız varsa, veri analizi işine girin.

Bu kitap aslında –cüretkârlığımı mazur görün– *Görünmeyen Ekonomi*'nin bir üst aşaması olarak görülebilir. *Görünmeyen Ekonomi*'de ele alınan çalışmalar ile bu kitapta ele alınan çalışmalar arasındaki en temel farklardan biri hırstır. Levitt'in isminin duyulduğu 1990'larda elde çok fazla veri yoktu. Levitt verilerden mahrum şekilde alengirli soruların peşine düşerek isim yapmıştı. Verilerin olmadığı büyük soruları bir kenara bırakmıştı. Bugün ise hemen her konuda elimizde bu kadar çok veri varken insan olmanın ne anlama geldiği sorusunun özüne inen büyük, derin soruların peşine düşmek anlamlıdır.

---

* Çev. S.E. Türkeli, İstanbul: Boyner Yayınları, 2006.

Veri analizinin geleceği parlak. Yeni neslin Kinsey'i, hiç şüphem yok, bir veri bilimcisi olacak. Bir sonraki Foucault bir veri bilimcisi olacak. Bir sonraki Freud bir veri bilimcisi olacak. Bir sonraki Marx bir veri bilimcisi olacak. Bir sonraki Salk da pekâlâ bir veri bilimcisi olabilir.

* * *

Neyse, düzgün bir sonuç kısmı yazmak için bir şeyler yapmaya çalışıyorum. Ama gerçekten önemli sonuç yazılarının çok daha fazlasını yaptığına hükmettim. Çok daha fazlasını. Etkili bir sonuç yazısı ironik olmalıdır. Harekete geçirici olmalıdır. Etkili bir sonuç yazısı derin ve oyunbaz olmalıdır. Derin, mizahi ve hüzünbaz olmalıdır. Etkili bir sonuç yazısı, tek bir cümleyle (ya da iki olsun) öncesinde gelmiş ve sonrasında gelecek her şeyi özetleyen bir hususa parmak basmalıdır. Bunu benzersiz, yeni bir hususta yapmalıdır. Etkili bir kitap akıllıca, komik ve kışkırtıcı bir vuruşla bitmelidir.

Şimdi biraz da benim yazma sürecimden bahsetsek iyi olur. Ben bilhassa ağzı laf yapan bir yazar değilim. Bu kitap, böyle zengin bir konu için biraz kısa sayılır: Yaklaşık 75 bin kelime.

Ama derinlikten yana olan açığımı takıntılılığımla kapatıyorum. *New York Times*'taki ilk cinsellik yazıma iki bin kelime için beş ay harcadım, kırk yedi taslak yazdım. Bu kitaptaki bazı bölümlerin altmış kadar taslağı var. Bir dipnottaki bir cümlede doğru bir kelimeyi bulmak için saatler harcayabilirim.

Son yılımı çoğunlukla keşiş gibi geçirdim. Sadece ben ve bilgisayarım. New York'un çok hareketli bir bölgesinde yaşadığım halde, neredeyse hiç dışarı çıkmadım. Bence bu benim şaheserim, hayatımda sahip olup olabileceğim en iyi fikir. Bu fikri doğru ifade edebilmek için ne fedakârlık gerekiyorsa yapmaya hazırdım. Bu kitaptaki her bir kelimeyi savunabilmeyi istiyordum. Telefonum cevap vermeyi unuttuğum e-postalarla, hiç açmadığım davetiyelerle, bakmadığım Bumble mesajlarıyla dolu.*

On üç aylık sıkı çalışmanın ardından nihayet tam denebilecek bir taslak çıkartabilmiştim. Fakat eksik olan bir şey vardı: Sonuç.

---

* Madem herkes yalan söylüyor, siz de benim bu söylediklerimi sorgulayabilirsiniz. Belki de ben çalışma konusunda takıntılı bir adam değilimdir. Belki de bu kitap için olağanüstü sıkı çalışmadım. Belki de birçok insan gibi çalışmalarımla ilgili gerçekleri abartıyorumdur. Belki de on üç aylık "sıkı çalışmam" sırasında bazı aylar yan gelip yatmışımdır. Belki de keşiş hayatı yaşamamışımdır. Belki de Facebook profilime baktıysanız, sözde keşişlik dönemimde arkadaşlarımla fotoğraflarımı görmüşsünüzdür. Veya belki de keşiş olmasına keşiştim, ama isteyerek olmadı. Belki de birçok gecemi tek başıma, çalışamadan, birisi beni arasın diye bekleyerek geçirdim. Belki de hiç kimse bana davetiye göndermiyor. Belki de hiç kimse bana Bumble'dan mesaj atmıyor. Her anlatıcı güvenilmezdir.

Editörüm Denise'e bunun için birkaç aya daha ihtiyacım olacağını söyledim. Büyük ihtimalle altı ay gerekeceğini izah ettim. Kanımca bir kitabın en önemli kısmı sonuç kısmıdır. Ve ben etkili bir sonuç bölümünün nasıl yazılacağını daha yeni öğreniyordum. Söylememe gerek yok, Denise bu durumdan hoşlanmamıştı.

Sonra bir gün, bir arkadaşım, Jordan Ellenberg'in bir çalışmasını yolladı. Wisconsin Üniversitesi'nden matematikçi Ellenberg kaç insanın başladığı kitabı sonuna kadar okuduğunu merak ediyordu. Büyük Veri'den yararlanarak bunu öğrenmenin dâhiyane bir yolunu bulmuştu. Ellenberg kitapların başındaki alıntıların altının, sonundaki alıntılara kıyasla çok daha fazla çizildiğini fark etti. Bu bize okurların kitapları sonuna kadar okuyup okumama eğilimlerine dair kaba bir fikir verebilirdi. Ellenberg'in hesaplamalarına göre, okurların yüzde 90'ından fazlası Donna Tartt'ın *Saka Kuşu** romanını bitirmişti. Oysa Nobel ödüllü iktisatçı Daniel Kahneman'ın şaheseri *Hızlı ve Yavaş Düşünme*'yi† okurların yalnızca yüzde 7'si sonuna kadar okumuştu. Bu kaba metodolojiyle yapılan tahmine göre, iktisatçı Thomas Piketty'nin çok tartışılan ve övülen kitabı *Yirmi Birinci Yüzyılda Kapital*'i‡ okurların yüzde üçü bile sonuna kadar okumamıştı. Başka bir deyişle, insanlar genellikle iktisatçıların kitaplarını sonuna kadar okumuyorlar.

Bu kitabın vurguladığı konulardan biri Büyük Veri'nin peşine düşüp bizi nereye götürürse götürsün gitmek ve buna göre hareket etmektir. Çoğu okurun her kelimemi dikkatle okuyacağını ve son sayfaları önceki sayfalara bağlayan örüntüleri tespit etmeye çalışacağını umuyorum. Ama yazı tarzımı cilalamak için ne kadar çok çalışsam da, birçok insan ilk elli sayfayı okuyup, birkaç not alıp sonra işine bakacak.

Dolayısıyla bu kitabı uygun olan tek bir şekilde bitireceğim: İnsanların ne söylediklerine değil, verilere, yani insanların ne yaptıklarına bakacağım. Bu lanet olası sonuç kısmını yazmaya çalışmayı bırakacağım ve birkaç arkadaşımla çıkıp iki bira içeceğim. Büyük Veri bana çok azınızın hâlâ okumakta olduğunu söylüyor.

---

* Çev. M.S. Ilgın, İstanbul: Pegasus, 2016.

† Çev. F. Deniztekin ve O.Ç. Deniztekin, İstanbul: Varlık Yayınları, 2015.

‡ Çev. H. Koçak, İstanbul: İş Bankası Kültür Yayınları, 2015.

Teşekkür

Bu kitap bir takım çalışmasıydı.

Bu fikirleri Harvard'da öğrenciyken, Google'da veri bilimcisiyken ve *New York Times*'ta yazarken geliştirdim.

Google'da teşriki mesaide bulunduğum Hal Varian'ın bu kitaptaki fikirlere etkisi çok büyüktür. Bence Hal her zaman çağının yirmi yıl ötesindedir. Carl Shapiro'yle birlikte kaleme aldıkları *Information* kitabı resmen geleceği öngörmüştür. Keza Hyunyoung Choi ile birlikte yazdıkları "Predicting The Present" [Şimdiyi Öngörmek] adlı yazının, bu kitapta anlatılan sosyal bilimlerdeki Büyük Veri devrimini başlattığını söylemek yanlış olmaz. Ayrıca kendisi harikulade ve kibar bir akıl hocasıdır, kendisiyle birlikte çalışan birçok insan şahittir. Hal daima, birlikte yazdığınız yazı için sizden çok çalışmasına rağmen, önce sizin isminizin yazılmasında ısrar eder. Hal'ın dehasıyla cömertliği nadiren karşılaştığım bir bileşimdir.

Yazılarım ve fikirlerim, *New York Times*'taki her bir makalemin editörü olan Aaron Retica'nın kanatları altında gelişti. Aaron bir allamedir. Nasıl yapıyor bilmiyorum, ama müzik, spor, siyaset, sosyoloji, iktisat konusunda ve daha kim bilir hangi konularda bilgi sahibi. *New York Times*'ta yayımlanan yazılarımda iyi ve güzel olan birçok şeyi ona borçluyum. Bu çalışmalardaki takım oyuncularından diğerleri: Grafikleri beni benden almaya devam eden Bill Marsh, Kevin McCarthy ve Gita Daneshjoo. Bu kitapta mezkûr yazılardan parçalar da –gerekli izinler alındıktan sonra– yayımlanmıştır.

Önsöz yazma nezaketini gösteren Steven Pinker'ı uzun zamandır idolüm olarak görüyorum. Sosyal bilimlerde modern bir kitabın ne olması gerektiği konusunda çıtayı o belirlemiştir: İnsan doğasının temel özelliklerini alabildiğine araştırmak, çok çeşitli disiplinlerden en iyi araştırmaları alıp yorumlamak. Bu çıtaya ulaşmak için ömrüm boyunca çabalayacağım.

Bu kitabın temelini oluşturan tezim, mükemmel ve sabırlı danışmanlarım Alberto Alesina, David Cutler, Ed Glaeser ve Lawrence Katz'ın gözetiminde yazıldı.

Denise Oswald müthiş bir editör. İşini ne kadar iyi yaptığını bilmek istiyorsanız, bu son taslağımı ilk taslağımla bir kıyaslayın: Aslında bunu yapamazsınız; çünkü o utanç verici ilk taslağı kimseye göstermeyeceğim. Keza HarperCollins'teki ekibimin

diğer üyelerine de teşekkür etmek istiyorum: Michael Barrs, Lynn Grady, Lauren Janiec, Shelby Meizlik ve Amber Oliver.

Menajerim Eric Lupfer daha en baştan bu projede bir ışık gördü, teklifin oluşturulmasında kritik rol oynadı ve nihayete erdirilmesine yardımcı oldu.

Veri doğrulamada çıkardığı muhteşem işten ötürü Melvis Acosta'ya müteşekkirim.

Mesleki ve akademik yaşamımda çok şey öğrendiğim diğer insanlar arasında Susan Athey, Shlomo Benartzi, Jason Bordoff, Danielle Bowers, David Broockman, Bo Cowgill, Steven Delpome, John Donohue, Bill Gale, Claudia Goldin, Suzanne Greenberg, Shane Greenstein, Steve Grove, Mike Hoyt, David Laibson, A.J. Magnuson, Dana Maloney, Jeffrey Oldham, Peter Orszag, David Reiley, Jonathan Rosenberg, Michael Schwarz, Steve Scott, Rich Shavelson, Michael D. Smith, Lawrence Summers, Jon Vaver, Michael Wiggins ve Qing Wu'yu saymalıyım.

Tim Requarth ve NeuWrite'a metni geliştirmeme yardımcı oldukları için teşekkür ediyorum.

Çalışmaları yorumlama konusundaki yardımlarından ötürü Christopher Chabris, Raj Chetty, Matt Gentzkow, Solomon Messing ve Jesse Shapiro'ya müteşekkirim.

Emma Pierson ve Katia Sobolski'ye kitabımdaki bir bölümle ilgili tavsiye verip vermeyeceklerini sordum. Nedenini hâlâ anlamamış olsam da, kitabın tamamını okumaya karar verdiler ve her bir paragrafla ilgili bilgece tavsiyelerde bulundular.

Annem Esther Davidowitz kitabı birkaç kez baştan sona okudu ve daha iyi hale getirilmesine ciddi katkılarda bulundu. Ayrıca bana sonu nereye varırsa varsın meraklarımın peşinden gitmem gerektiğini lafta değil örnek olarak gösterdi. Bir akademik iş için girdiğim mülakatta, profesörün biri beni sorguya çekmişti: "Anneniz yaptığınız bu işe ne diyor?" Benim seks ve diğer tabu konuları araştırıyor olmamdan annemin utanıyor olabileceğini ima etmişti. Ama sonu nereye varırsa varsın meraklarımın peşinden gittiğim için annemin benimle gurur duyduğunu her zaman biliyordum.

Kitabın belli kısımlarını okuyup değerli yorumlar sunan birçok insan var. Eduardo Acevedo, Coren Apicella, Sam Asher, David Cutler, Stephen Dubner, Christopher Glazek, Jessica Goldberg, Lauren Goldman, Amanda Gordon, Jacob Leshno, Alex Peysakhovich, Noah Popp, Ramon Roullard, Greg Sobolski, Evan Soltas, Noah Stephens-Davidowitz, Lauren Stephens-Davidowitz ve Jean Yang'a müteşekkirim. Bu kitabı yazma sürecimde Jean en iyi dostum oldu, ona bu yüzden de teşekkür borçluyum.

Veri toplama konusundaki yardımlarından ötürü MindGeek'ten Brett Goldenberg, James Rogers ve Mike Williams'a ve Baseball Prospectus'ten Rob McQuown ile Sam Miller'a teşekkür ediyorum.

Alfred Sloan Vakfı'na mali desteğinden ötürü minnettarım.

Bu kitabı yazarken bir noktada derin bir açmaza girdim, kendimi kaybettim ve projeyi tümden bırakmanın eşiğine geldim. Sonra babam Mitchell Stephens'la köye gittik. Bir haftada babam beni toparladı. Birlikte yürüyüşlere çıkıp aşktan, ölümden, başarıdan, mutluluktan ve yazmaktan bahsettik. Sonra beni masanın başına oturttu, tek tek her bir cümlenin üzerinden geçmeme yardımcı oldu. O olmasaydı bu kitabı bitiremezdim.

Geriye kalan hataların hepsi elbette bana aittir.

Notlar

GİRİŞ

11 Amerikalı seçmenlerin çoğu Barack Obama'nın siyah olmasını umursamıyordu:

Katie Fretland, "Gallup: Race Not Important to Voters," The Swamp, *Chicago Tribune,* Haziran 2008.

11 Berkeley ... didiklemeye başladılar:

Alexandre Mas ve Enrico Moretti, "Racial Bias in the 2008 Presidential Election," *American Economic Review* 99, no. 2 (2009).

12 ırkçılık sonrası bir toplumda:

12 Kasım 2009'da Lou Dobbs, programının bir parçasında "partizanlığın, ırkçılığın geride kaldığı bir toplumda" yaşadığımızı söylemişti. 27 Ocak 2010'daki programında Chris Matthews Başkan Obama'nın "tüm görünüşü bakımından ırk ötesi" olduğunu söylemişti. Diğer örnekler için, bkz. Michael C. Dawson ve Lawrence D. Bobo, "One Year Later and the Myth of a Post-Racial Society," *Du Bois Review: Social Science Research on Race* 6, no. 2 (2009).

13 General Social Survey'in verilerini tahlil ettim.

Tüm bu hesaplamalarla ilgili ayrıntıları internet sitem sethsd.com'da "Seks Verileri" başlığı altında bulabilirsiniz. General Social Survey'den veriler için, bkz. http://gss.norc.org/.

13 her yıl satılan kondom sayısı 600 milyondan azdır:

Yazarla paylaşılan veriler.

15 Stormfront'a üyelikler normalin on kat üstündeydi:

Yazarın Google Trends verilerine dair analizinden. Ayrıca Seth Stephens-Davidowitz, "The Data of Hate"te (*New York Times*, 13 Temmuz 2014) ele alınan Stormfront'un tüm üyelerine dair verileri de masaya yatırdım. İlgili veriler sethsd.com'daki "Stormfront" başlıklı veri kısmından indirilebilir.

15 "zenci başkan" araması "ilk siyah başkan"dan daha fazlaydı:

Yazarın Google Trends verilerine dair analizinden. Bunun doğru olduğu eyaletler arasında Kentucky, Louisiana, Arizona ve Kuzey Carolina vardır.

16 beş akademik dergi tarafından reddedildi:

Yazı en sonunda Seth Stephens-Davidowitz, "The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data" adıyla yayınlandı, *Journal of Public Economics* 118 (2014). Araştırma hakkında daha fazla ayrıntı orada bulunabilir. Ayrıca veriler kendi internet sitem sethsd.com'da "Irkçılık" başlıklı veri kısmında bulunabilir.

17 En iyi bağıntı:

"Trump desteği için bulduğum en güçlü bağıntı Google'daki "*n-word*" (aşağılayıcı manada "zenci" için sansürlü kullanım) aramalarıydı. Başkaları da bundan bahsetmiştir" (28 Şubat 2016, tweet). Ayrıca bkz. Nate Cohn, "Donald Trump's Strongest Supporters: A New Kind of Democrat," *New York Times*, 31 Aralık 2015, A3.

20 "*Nigger*" kelimesinin yer aldığı Google aramalarının yüzdesini gösteriyor bu. Ölçüm Google aramalarının bir yüzdesi olarak yapıldığından, nüfusu yüksek olan ya da çok fazla arama yapılan yerlerde keyfi şekilde yüksek olmadığını belirtelim. Ayrıca bu haritadaki ve Trump desteği haritasındaki farklardan bazılarının bariz açıklamaları olduğunu söylemek lazım. Trump Teksas'ta ve Arkansas'da popülerliğini yitirdi, çünkü bu eyaletler rakiplerinin, biri Ted Cruz'un diğeri Mike Huckabee'nin memleketiydi.

20 Bu veriler Civis Analytics'in Aralık 2015 tarihli araştırma verileridir. Gerçek oy verilerinin burada yararı daha azdır; zira ön eleme seçimlerinin gerçekleştiği zamandan ve oy formatından fazlasıyla etkilenmektedir. Haritalar *New York Times*'ın izniyle basılmıştır.

21 2,5 trilyon bayt veri yaratıyorlar:

"Bringing Big Data to the Enterprise," IBM, https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

22 giderek büyüyen bir samanlıkta iğne aramaya benzemesi:

Nassim M. Taleb, "Beware the Big Errors of 'Big Data'", *Wired*, 8 Şubat 2013, http://www.wired.com/2013/02/big-data-means-big-errors-people.

23 ne ırkçı aramalar ne de Stormfront üyelikleri:

İnternet ırkçılığının 2008 krizine en çok maruz kalan ve en az maruz kalan bölgelerde nasıl değişiklik gösterdiğini inceledim. Hem Google'daki "*nigger*" arama oranlarına hem de Stormfront üyeliğine baktım. İlgili veriler sethsd.com'daki "Irksal Nefret" ve "Stormfront" başlıklı kısımlardan indirilebilir.

23 Google'daki anksiyeteyle alakalı aramalara:

Seth Stephens-Davidowitz, "Fifty States of Anxiety," *New York Times*, 7 Ağustos 2016, SR2 Google aramaları çok daha büyük örneklemler sunmakla birlikte, bu örüntü araştırma kanıtlarıyla tutarlıdır. Örneğin bkz. William C. Reeves vd., "Mental Illness Surveillance Among Adults in the United States," *Morbidity and Mortality Weekly Report Supplement* 60, no. 3 (2011).

23 Şaka aramaları:

Buna dair bir değerlendirme için, bkz. Seth Stephens-Davidowitz, "Why Are You Laughing?" *New York Times*, 15 Mayıs 2016, SR9. İlgili veriler sethsd.com'daki "Şakalar" başlıklı kısımdan indirilebilir.

24 "kocamın en çok istediği şey onu emzirmem":

Buna dair değerlendirme için, bkz. Seth Stephens-Davidowitz, "What Do Pregnant Women Want?" *New York Times*, 17 Mayıs 2014, SR6.

24 kadınların erkekleri emzirdiği porno aramaları:

Yazarın PornHub verilerine dair analizi.

24 Kadınların cinsel organlarıyla ilgili kaygı belirten aramaları:

Buna dair bir değerlendirme için, bkz. Seth Stephens- Davidowitz, "Searching for Sex," *New York Times*, 25 Ocak 2015, SR1.

24 "poemas para mi esposa embarazada":

Stephens-Davidowitz, "What Do Pregnant Women Want?"

24 Friedman ... der:

Jerry Friedman'la telefon görüşmesi, 27 Ekim 2015.

25 Tüm verilerinin örneklemeleri:

Hal R. Varian, "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives* 28, no. 2 (2014).

1. BÖLÜM

30 Aslında en iyi veri bilimi şaşırtıcı derecede sezgiseldir:

Veri analizinin bildiğim kısmından bahsediyorum: İnsan davranışını açıklamaya ve tahmin etmeye çalışan veri bilimi. Sözgelimi araba sürmeye çalışan yapay zekâdan bahsetmiyorum. Bu metodolojiler insan beyninde keşfedilen aletlerden yararlanmakla birlikte, anlaşılması daha zor şeylerdir.

31 pankreas kanserinin habercisi:

John Paparrizos, Ryan W. White ve Eric Horvitz, "Screening for Pancreatic Adenocarcinoma Using Signals from Web Search Logs: Feasibility Study and Results," *Journal of Oncology Practice* (2016).

33 Kış mevsimi:

Bu araştırma hakkında değerlendirme için, bkz. Seth Stephens-Davidowitz, "Dr. Google Will See You Now," *New York Times*, 11 Ağustos 2013, SR12.

33 insan ilişkileri hakkında toplanmış en büyük veri kümesini analiz etti:

Lars Backstrom ve Jon Kleinberg, "Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook," *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* içinde (2014).

34 insanlar tutarlı bir şekilde:

Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011).

34 astım yaklaşık yetmiş kat daha fazla ölüme sebep olmaktadır:

1979-2010 arasında her yıl ortalama 55,81 Amerikalı kasırgadan, 4216,53 Amerikalıysa astımdan ötürü hayatını kaybetmiştir. Bkz. Annual U.S. Killer Tornado Statistics, National Weather Service, http://www.spc.noaa.gov/climo/torn/fatalmap.php ve Trends in Asthma Morbidity and Mortality, American Lung Association, Epidemiology and Statistics Unit.

34 Patrick Ewing:

En sevdiğim Ewing videoları: "Patrick Ewing's Top 10 Career Plays," YouTube videosu, yayınlanma tarihi 18 Eylül 2015, https://www.youtube.com/

watch?v=Y29gMuYymv8 ile "Patrick Ewing Knicks Tribute," YouTube videosu, yayınlanma tarihi 12 Mayıs 2006, https://www.youtube.com/watch?v=8T2l5Emzu-I.

35 "basketbola hayat memat gözüyle bakıyorlar":

S. L. Price, "Whatever Happened to the White Athlete?" *Sports Illustrated*, 8 Aralık 1997.

35 bir internet araştırması:

Bu benim 22 Ekim 2013'te yürüttüğüm Google Tüketici Anketi'ydi. "Sizce NBA oyuncularının çoğunluğu nerede doğmuştur?" sorusunu yönelttim. İki seçenek vardı: Ya "yoksul mahallelerde" ya da "orta sınıf mahallelerde". Katılımcıların yüzde 59,7'si "yoksul mahallelerde" yanıtını verdi.

36 bir Siyahinin ilk adının sosyo-ekonomik kökenine dair bir işaret olması:

Roland G. Fryer Jr. ve Steven D. Levitt, "The Causes and Consequences of Distinctively Black Names," *Quarterly Journal of Economics* 119, no. 3 (2004).

37 1980'lerde doğup en fazla sayı atmış olan yüz Afrikalı-Amerikalı NBA oyuncusuyla:

Centers for Disease Control and Prevention, "Health, United States, 2009," Tablo 9, Nonmarital Childbearing, Detailed Race and Hispanic Origin of Mother, and Maternal Age: United States, Selected Years 1970–2006.

37 Chris Bosh . . . Chris Paul:

"Not Just a Typical Jock: Miami Heat Forward Chris Bosh's Interests Go Well Beyond Basketball," Palm BeachPost.com, 15 Şubat 2011, http://www.palmbeachpost.com/news/sports/basketball/not-just-a-typical-jock-miami-heat-forward-chris-b/nLp7Z/; Dave Walker, "Chris Paul's Family to Compete on 'Family Feud,' nola.com, 31 Ekim 2011, http://www.nola.com/tv/index.ssf/2011/10/chris_pauls_family_to_com pete.html.

37 on santimetre daha uzundur:

"Why Are We Getting Taller as a Species?" *Scientific American*, http://www.scientificamerican.com/article/why-are-we-getting-taller/. İşin ilginci, Amerikalıların uzaması durmuştur. Amanda Onion, "Why Have Americans Stopped Growing Taller?" ABC News, 3 Temmuz 2016, http://abcnews.go.com/Technology/story?id=98438&page=1. NBA'de oynayan yabancı oyuncu sayısındaki artışının nedenlerinden biri olarak, diğer ülkelerin boy ortalamasının ABD'yi yakalamış olmasıyla bağ kurdum. NBA'deki 2,13'lük Amerikalıların sayısı, Amerikalıların genelinin boy atmasına paralel olarak 1946-1980 arasında on altı kat artmıştır. Sonrasında artış olmamıştır; zira Amerikalıların uzaması da durmuştur. Diğer yandan, diğer ülkelerdeki 2,13'lüklerin sayısı ciddi ölçüde artmıştır. Benim bulabildiğim kadarıyla, yabancı oyunculardaki en büyük artış Türkiye, İspanya ve Yunanistan gibi son yıllarda çocuk sağlığı ve yetişkin boyu konusunda gözle görülür artışın yaşandığı ülkelerin son derece uzun oyuncularından gelmiştir.

37 yoksul ailelerden gelen Amerikalıların:

Carmen R. Isasi vd., "Association of Childhood Economic Hardship with Adult Height and Adult Adiposity among Hispanics/Latinos: The HCHS/SOL Socio-Cultural Ancillary Study," *PloS One* 11, no. 2 (2016); Jane E. Miller ve Sanders Korenman, "Poverty and Children's Nutritional Status in the United States," *American*

Journal of Epidemiology 140, no. 3 (1994); Harry J. Holzer, Diane Whitmore Schanzenbach, Greg J. Duncan ve Jens Ludwig, "The Economic Costs of Childhood Poverty in the United States," *Journal of Children and Poverty* 14, no. 1 (2008).

38 Amerikalı ortalama bir erkeğin boyu 1.76'dır:

Cheryl D. Fryar, Qiuping Gu ve Cynthia L. Ogden, "Anthropometric Reference Data for Children and Adults: United States, 2007–2010," *Vital and Health Statistics Series* 11, no. 252 (2012).

38 beşte bir gibi bir orandan bahsediyoruz:

Pablo S. Torre, "Larger Than Real Life," *Sports Illustrated*, 4 Temmuz 2011.

38 orta sınıfa mensup iki ebeveynli aileler:

Tim Kautz, James J. Heckman, Ron Diris, Bas Ter Weel ve Lex Borghans, "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success," *National Bureau of Economic Research Working Paper* 20749, 2014.

38 Wrenn kadar yükseğe zıplayabilen:

Desmond Conner, "For Wrenn, Sky's the Limit," *Hartford Courant*, 21 Ekim 1999.

38 Ama Wrenn:

Doug Wrenn'in hikâyesi Percy Allen, "Former Washington and O'Dea Star Doug Wrenn Finds Tough Times"ta anlatılmaktadır, *Seattle Times*, 29 Mart 2009.

39 "Doug Wrenn öldü": a.g.e.

39 Jordan muhtemelen zor bir çocuktu:

Melissa Isaacson, "Portrait of a Legend," ESPN.com, 9 Eylül 2009, http://www.espn.com/chicago/columns/story?id=4457017&columnist=isaacson_melissa. İyi bir Jordan biyografisi için, bkz. Roland Lazenby, *Michael Jordan: The Life* (Boston: Back Bay Books, 2015).

39 Jordan'ın babası:

Barry Jacobs, "High-Flying Michael Jordan Has North Carolina Cruising Toward Another NCAA Title," *People*, 19 Mart 1984.

39 Jordan'ın yaşamı ailesinin onu büyük, rekabetçi bir yeteneğin düşebileceği tuzaklardan uzaklaştırdığı hikâyelerle doludur:

Isaacson, "Portrait of a Legend."

39 Basketbol Şöhretler Kadrosu'na (Hall of Fame) dahil edilmesi:

Michael Jordan'ın Basketbol Hall of Fame'ine Giriş Konuşması, YouTube videosu, 21 Şubat 2012, https://www.youtube.com/watch?v=XLzBMGXfK4c. Jordan'ın konuşmasının en ilgi çekici yanı anne-babasını övgülere boğması değil; hâlâ daha kariyerinin başlarındaki ufak yanlışlara dikkat çekme ihtiyacı duymasıdır. Belki de gelmiş geçmiş en büyük basketbolcu olmak için, küçük yanlışları ömür boyu takıntı haline getirmek zorunludur.

40 LeBron James mülakatı:

"I'm LeBron James from Akron, Ohio," YouTube videosu, 20 Haziran 2013'te yayınlandı, https://www.youtube.com/watch?v=XceMbPVAggk.

2. BÖLÜM

43 **Fallus şekilli bir yiyecek:**
Enine değil boyuna büyük olan yiyecekleri fallus şekilli olarak kodladım. Salatalık, mısır, havuç, patlıcan, kabak ve muz bu kapsamdadır. Veriler ve kod sethsd.com'da bulunabilir.

45 Microsoft araştırmacıları tarafından derlenen:

Veri kümesi https://www.microsoft.com/en-us/download/details.aspx?id=52418 adresinden indirilebilir. Araştırmacılar Amazon Mechanical Turk kullanıcılarından görüntüleri tarif etmelerini istediler. Klavye kayıtlarını analiz edip birinin bir kelimeyi düzelttiği her durumda not aldılar. Daha fazla ayrıntı için bkz. Yukino Baba ve Hisami Suzuki, "How Are Spelling Errors Generated and Corrected? A Study of Corrected and Uncorrected Spelling Errors Using Keystroke Logs," Proceedings of the Fiftieth Annual Meeting of the Association for Computational Linguistics, 2012. Veriler, kod ve bu araştırmaya dair daha fazla bilgi sethsd.com adresinde bulunabilir.

47 "Falancamla seks yapmak istiyorum":

	"... ile seks yapmak istiyorum" BU KELİMELERLE YAPILAN GOOGLE ARAMALARI (AYLIK)
annemle	720
oğlumla	590
kız kardeşimle	590
kuzenimle	480
babamla	480
erkek arkadaşımla	480
erkek kardeşimle	320
kızımla	260
arkadaşımla	170
kız arkadaşımla	140

47 porno animasyon:

Örneğin "porno", aşağıda görüldüğü üzere, son derece popüler olan çeşitli animasyon programlarla ilgili Google aramalarında en sık kullanılan kelimelerden biridir.

Animasyonlar ve porno

(çeşitli animasyonlarla ilgili en yaygın google aramaları)

family guy pornosu	simpsons izle	futurama pornosu	scooby doo oyunları
family guy bölümleri	simpsons pornosu	futurama leela	scooby doo filmi
family guy ücretsiz	simpsons	online futurama bölümleri	scooby doo pornosu
family guy izle	simpsons filmi	futurama online	scooby doo velma

47 bebek bakıcıları

Yazarın hesaplamalarına göre, erkeklerin porno aramalarındaki en popüler kadın meslekleri şöyledir (erkeklerin yaşına göre):

Erkeklerin Porno Aramalarındaki Kadın Meslekleri, Erkeklerin Yaşına Göre

	18-24	25-64	65+
1.	Bebek bakıcısı	Bebek bakıcısı	Bebek bakıcısı
2.	Öğretmen	Yoga Eğitmeni	Ponpon kız
3.	Yoga Eğitmeni	Öğretmen	Doktor
4.	Ponpon kız	Ponpon kız	Öğretmen
5.	Doktor	Emlakçı	Emlakçı
6.	Fahişe	Doktor	Hemşire
7.	Emlakçı	Fahişe	Yoga Eğitmeni
8.	Hemşire	Sekreter	Sekreter
9.	Sekreter	Hemşire	Fahişe

3. BÖLÜM

51 algoritmalar:

Matthew Leising, "HFT Treasury Trading Hurts Market When News Is Released," Bloomberg Markets, 16 Aralık 2014; Nathaniel Popper, "The Robots Are Coming for Wall Street," *New York Times Magazine*, 28 Şubat 2016, MM56; Richard Finger, "High Frequency Trading: Is It a Dark Force Against Ordinary Human Traders and Investors?" *Forbes*, 30 Eylül 2013, http://www.forbes.com/sites/richardfinger/2013/09/30/high-frequency-trading-is-it-a-dark-force-against-ordinary-human-traders-and-investors/#50875fc751a6.

52 Alan Krueger:

Alan Krueger ile 8 Mayıs 2015'te telefonda görüştüm.

52 gribin yayılma hızı:

İlk yazı Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski ve Larry Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature* 457, no. 7232 (2009) idi. Özgün modeldeki kusurlar David Lazer, Ryan Kennedy, Gary King ve Alessandro Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis"te ele alınmıştır *Science* 343, no. 6176 (2014). Düzeltilmiş bir model Shihao Yang, Mauricio Santillana ve S. C. Kou, "Accurate Estimation of Influenza Epidemics Using Google Search Data Via ARGO"da sunulmuştur, *Proceedings of the National Academy of Sciences* 112, no. 47 (2015).

52 konut fiyatları:

Seth Stephens-Davidowitz ve Hal Varian, "A Hands-on Guide to Google Data," mimeo, 2015. Ayrıca bkz. Marcelle Chauvet, Stuart Gabriel ve Chandler Lutz, "Mortgage Default Risk: New Evidence from Internet Search Queries," *Journal of Urban Economics* 96 (2016).

54 Bill Clinton:

Sergey Brin ve Larry Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," Seventh International World-Wide Web Conference, 14-18 Nisan1998, Brisbane, Avustralya.

55 porno siteleri:

John Battelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture* (New York: Penguin, 2005).

55 kullanıcıların görüşlerinden içerik oluşturmak:

Buna dair iyi bir değerlendirme için, bkz. Steven Levy, In *the Plex: How Google Thinks, Works, and Shapes Our Lives* (New York: Simon & Schuster, 2011).

56 "Evinizi satın":

Bu alıntı Joe Drape, "Ahmed Zayat's Journey: Bankruptcy and Big Bets"te de yer almıştı, *New York Times*, 5 Haziran 2015, A1. Gelgelelim, yazı alıntıyı Seder'e atfetme yanlışına düşüyor. Oysa takımın başka bir üyesi tarafından söylenmişti.

57 Seder'la tanışmam:

Jeff Seder ve Patty Murray ile Ocala, Florida'da 12-14 Haziran 2015 tarihleri arasında görüştük.

57 Kabaca üçte biri:

Yarış atlarının hüsrana yol açmalarının nedenleri bu işte geçirdiği yıllara dayanarak Jeff Seder'in kaba tahminleridir.

57 yüzlerce at:

Supplemental Tables of Equine Injury Database Statistics for Thoroughbreds, http://jockeyclub.com/pdfs/ eid_7_year_tables.pdf.

57 çoğunlukla da bacakları kırıldığı için:

"Postmortem Examination Program," California Animal Health and Food Laboratory System, 2013.

58 Yine de, dörtte üçünden daha fazlası:

Avalyn Hunter, "A Case for Full Siblings," *Bloodhorse*, 18 Nisan 2014, http://www.bloodhorse.com/horse-racing/articles/115014/a-case-for-full-siblings.

58 Earvin Johnson III:

Melody Chiu, "E. J. Johnson Loses 50 Lbs. Since Undergoing Gastric Sleeve Surgery," *People*, 1 Ekim 2014.

58 LeBron James'in annesinin boyu sadece 1.64'tür:

Eli Saslow, "Lost Stories of LeBron, Part 1," ESPN.com, 17 Ekim 2013, http://www.espn.com/nba/story/_/id/9825052/how-lebron-james-life-changed-fourth-grade-espn-magazine.

59 Yeşil Maymun:

Bkz. Sherry Ross, "16 Million Dollar Baby," *New York Daily News*, 12 Mart 2006 ve Jay Privman, "The Green Monkey, Who Sold for $16M, Retired," ESPN.com, 12 Şubat 2008, http://www.espn.com/sports/horse/news/story?id=3242341. Müzayedenin videosu için, bkz. "$16 Million Horse," YouTube videosu, 1 Kasım 2008'de yayınlandı, https://www.youtube.com/watch?v=EyggMC85Zsg.

61 Google'ın arama verilerinden yararlanarak grip hastalığını tahmin etme:

Sharad Goel, Jake M. Hofman, Sébastien Lahaie, David M. Pennock ve Duncan J. Watts, "Predicting Consumer Behavior with Web Search," *Proceedings of the National Academy of Sciences* 107, no. 41 (2010).

62 Çilekli Pop-Tart:

Constance L. Hays, "What Wal-Mart Knows About Customers' Habits," *New York Times*, 14 Kasım 2004.

62 Orley Ashenfelter:

Orley Ashenfelter ile 27 Ekim 2016'da telefonda görüştüm.

67 Leonard Cohen bir keresinde:

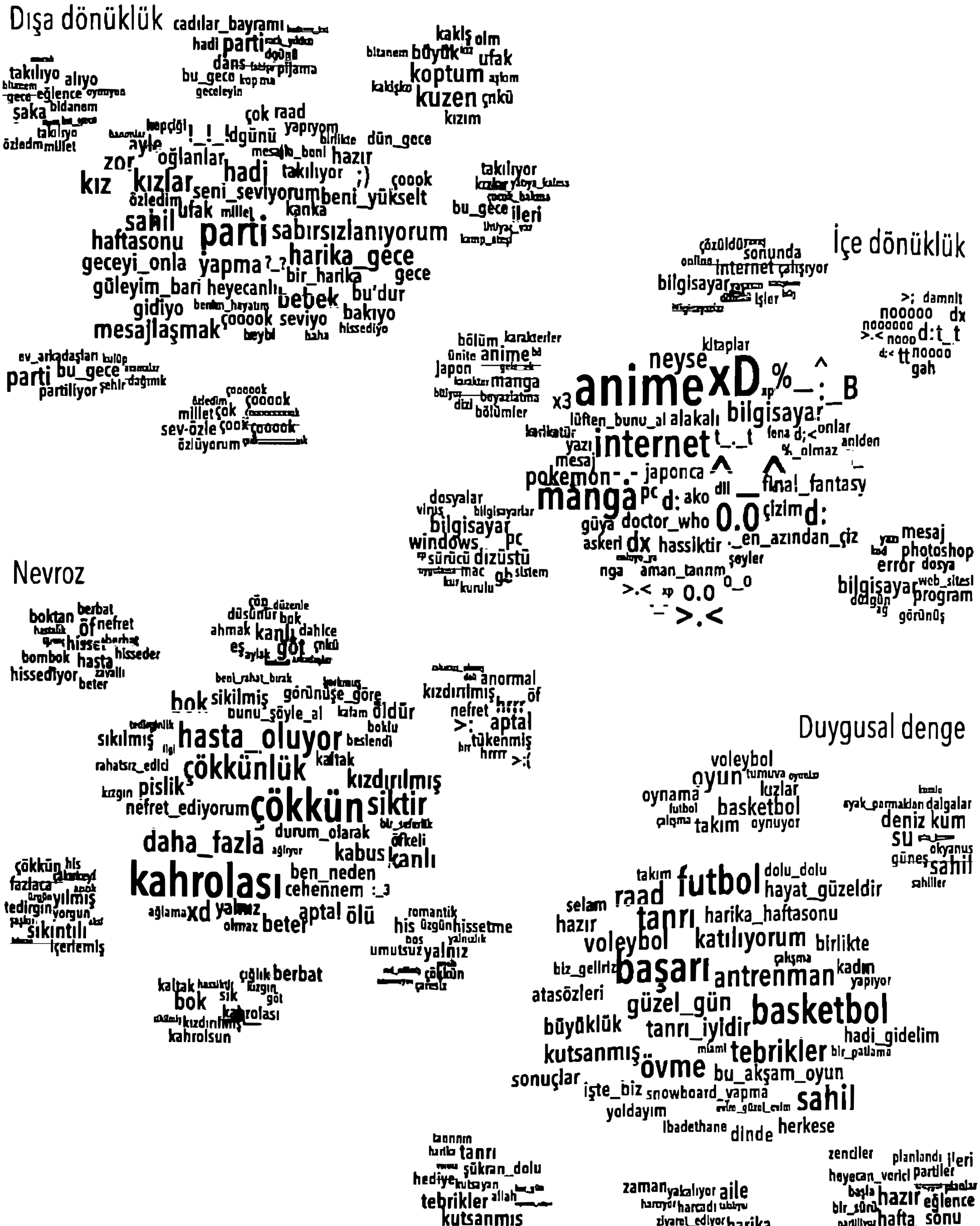
Jonathan Greenberg, "What I Learned From My Wise Uncle Leonard Cohen," *Huffington Post*, 11 Kasım 2016.

69 Yüzlerce heteroseksüel Speeddate kullanıcısı:

Daniel A. McFarland, Dan Jurafsky ve Craig Rawlings, "Making the Connection: Social Bonding in Courtship Situations," *American Journal of Sociology* 118, no. 6 (2013).

69 yüz binlerce Facebook gönderisinde:

H. Andrew Schwartz vd., "Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach," *PloS One* 8, no. 9 (2013). Yazıda kişilik testlerinde kaç puan aldıklarına bağlı olarak insanların nasıl konuştukları da inceleniyor. İşte bulgular:



74 binlerce kitap ve film senaryosunun metni:

Andrew J. Reagan, Lewis Mitchell, Dilan Kiley, Christopher M. Danforth ve Peter Sheridan Dodds, "The Emotional Arcs of Stories Are Dominated by Six Basic Shapes," *EPJ Data Science* 5, no. 1 (2016).

76 hangi tür haberlerin paylaşıldığı:

Jonah Berger ve Katherine L. Milkman, "What Makes Online Content Viral?" *Journal of Marketing Research* 49, no. 2 (2012).

79 Neden bazı gazeteler sola meylederken:

Araştırma şu yazıda ete kemiğe büründü: Matthew Gentzkow ve Jesse M. Shapiro, "What Drives Media Slant? Evidence from U.S. Daily Newspapers," *Econometrica* 78, no. 1 (2010). Gentzkow ve Shapiro bu proje başladığında sadece birer mastır öğrencisi olmalarına karşın, artık yıldız iktisatçılardır. Şu an Stanford'da profesör olan Gentzkow, kırk yaşın altındaki en iyi iktisatçılara verilen 2014 John Bates Clark Madalyası'nı kazandı. Brown'da profesör olan Shapiro ise şu anda prestijli *Journal of Political Economy* dergisinin editörlerinden biridir. Ortaklaşa kaleme aldıkları medya taraflılığı hakkındaki yazı ikisinin de en çok zikredilen yazıları arasındadır.

79 Rupert Murdoch:

Murdoch'un muhafazakâr *New York Post* gazetesinin sahibi olması, New York çok büyük olduğundan çeşitli bakış açılarına sahip gazeteleri barındırabiliyor olmasıyla açıklanabilir. Gelgelelim bu gazetenin düzenli olarak para kaybettiği açık görünmektedir. Örneğin bkz. Joe Pompeo, "How Much Does the 'New York Post' Actually Lose?" *Politico*, 30 Ağustos 2013, http://www.politico.com/media/story/2013/08/how-much-does-the-new-york-post-actually-lose-001176.

80 Shapiro bana:

Matt Gentzkow ve Jesse Shapiro ile 16 Ağustos 2015'te Royal Sonesta Boston'da mülakat yaptım.

81 Amerikan yıllıkları:

Kate Rakelly, Sarah Sachs, Brian Yin ve Alexei A. Efros, "A Century of Portraits: A Visual Historical Record of American High School Yearbooks", tebliğ, International Conference on Computer Vision, 2015. Fotoğraflar yazarların izniyle basılmıştır.

81 Fotoğraflardaki özneler:

Örneğin bkz. Christina Kotchemidova, "Why We Say 'Cheese': Producing the Smile in Snapshot Photography," *Critical Studies in Media Communication* 22, no. 1 (2005).

82 Bu ülkelerde geceleri ne kadar ışık yakıldığına bakarak GSYH'nin ölçülmesine:

J. Vernon Henderson, Adam Storeygard ve David N. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review* 102, no. 2 (2012).

83 tahmini GSYH yüzde 90:

Kathleen Caulderwood, "Nigerian GDP Jumps 89% as Economists Add in Telecoms, Nollywood," *IBTimes*, 7 Nisan 2014, http://www.ibtimes.com/nigerian-gdp-jumps-89-economists-add-telecoms-nollywood-1568219.

83 Reisinger ... dedi:

Joe Reisinger ile 10 Haziran 2015'te telefonda görüştüm.

84 50 milyon dolar:

Leena Rao, "SpaceX and Tesla Backer Just Invested $50 Million in This Startup," *Fortune*, 24 Eylül 2015.

4. BÖLÜM

85 1950 yılında önemli bir yazıda:

Hugh J. Parry ve Helen M. Crossley, "Validity of Responses to Survey Questions," *Public Opinion Quarterly* 14, 1 (1950).

86 Maryland Üniversitesi mezunlarına üniversite deneyimleri hakkında çeşitli sorular soruldu:

Frauke Kreuter, Stanley Presser ve Roger Tourangeau, "Social Desirability Bias in CATI, IVR, and Web Surveys," *Public Opinion Quarterly* 72(5), 2008.

86 anketlerin bilememesinde:

Yalan söylemenin Trump'a olan desteği tahmin etmeye çalışmak açısından bir problem olabileceğini savunan bir makale için bkz. Thomas B. Edsall, "How Many People Support Trump but Don't Want to Admit It?" *New York Times*, 15 Mayıs 2016, SR2. Fakat bunun büyük bir etken olmadığını savunanlar da vardır, örneğin bkz. Andrew Gelman, "Explanations for That Shocking 2% Shift", *Statistical Modeling, Causal Inference, and Social Science*, 9 Kasım 2016, http://andrewgelman.com/2016/11/09/explanations-shocking-2-shift/.

86 Tourangeau diyor:

Roger Tourangeau ile 5 Mayıs 2015'te telefonda görüştüm.

87 neden bu kadar çok insanın ortalamanın üstünde olduğunu söylediğini:

Buna dair değerlendirme için, bkz. Adam Grant, *Originals: How Non-Conformists Move the World* (New York: Viking, 2016). Özgün kaynak: David Dunning, Chip Heath ve Jerry M. Suls, "Flawed Self-Assessment: Implications for Health, Education, and the Workplace," *Psychological Science in the Public Interest* 5 (2004).

87 anketlere gıcık olmaktır:

Anya Kamenetz, "'Mischievous Responders' Confound Research on Teens", *nprED*, 22 Mayıs 2014, http://www.npr.org/sections/ed/2014/05/22/313166161/mischievous-responders-confound-research-on-teens. Bu makalenin ele aldığı özgün araştırma Joseph P. Robinson-Cimpian, "Inaccurate Estimation of Disparities Due to Mischievous Responders"tır, *Educational Researcher* 43, no. 4 (2014).

88 Amerikalılar "porno" aramasını "hava durumu" aramasından daha çok yapıyorlar:

https:// www.google.com/trends/explore?date=all&geo=US&q=porn,weather.

88 [illegible]rno izlediğini kabul ediyor:

Amanda Hess, "How Many Women Are Not Admitting to Pew That They Watch Porn?", *Slate*, 11 Ekim 2013, http://www.slate.com/blogs/xx_factor/2013/10/11/pew_online_viewing_study_percentage_of_women_who_watch_online_porn_is_growing.html.

88 "çük", "sikmek" ve "[illegible]rno":

Nicholas Diakopoulus, "Sex, Violence, and Autocomplete Algorithms," *Slate*, 2 Ağustos 2013, http://www.slate.com/articles/technology/future_tense/2013/08/words_banned_from_bing_and_google_s_autocomplete_algorithms.html.

89 Google'a söyleme ihtimalleri çocuksuz yetişkinlerinkinden 3,6 kat fazladır:

Çeşitli ifade tarzlarıyla hesapladığımda, ABD'de her ay, çocuk yaptığına pişman olduğunu açıkça söyleyen 1.730 Google araması yapılmaktadır. Çocuk sahibi olmadığına pişman olanların sayısı ise sadece 50 civarındadır. Çocuk sahibi olmayan 45 yaşın üstündekilerin sayısı 15,9 milyondur. Çocuk sahibi Amerikalıların sayısıysa 152 milyon civarıdır. Bu demektir ki, çocuk sahibi insanların Google'da pişmanlık belirtme oranı çocuk sahibi olmayanlarındakinden 3,6 kat daha fazladır. Metinde belirtilmiş olsa da tekrar vurgulanmaya değer: Bu Google itirafları sadece az sayıda insan tarafından yapılmaktadır, bunlar muhtemelen Google'ın kendilerine bu konuda yardımcı olamayacağını bir an unutacak kadar pişmanlık hissinin ağır bastığı kişilerdir.

90 gey evliliğine desteğin en yüksek olduğu eyalet olan Rhode Island'da:

Bu tahminler Nate Silver'a aittir, "How Opinion on Same-Sex Marriage Is Changing, and What It Means," *FiveThirtyEight*, 26 Mart 2013, http://fivethirtyeight.blogs.nytimes.com/2013/03/26/how-opinion-on-same-sex-marriage-is-changing-and-what-it-means/?_r=0.

91 erkek Facebook kullanıcılarının yaklaşık yüzde 2,5'i erkeklerle:

Yazarın Facebook reklam verilerine dair analizinden. "Erkeklerle ve kadınlarla" ilgilendiğini söyleyen Facebook kullanıcılarını dikkate almadım. Analizim kadınlarla ve erkeklerle ilgilendiğini söyleyen azımsanmayacak sayıda kullanıcının sorunu bir aşk ilişkisi olarak değil, arkadaşlık ilişkisi olarak yorumladığını savunuyor.

92 erkek porno aramalarının yaklaşık yüzde beşi gey erkek pornosu aramaktadır:

Belirtildiği üzere, Google Trends aramaları cinsiyete göre ayırmıyor. Google AdWords çeşitli kategorilerden sayfa görüntülemeleri cinsiyete göre ayırıyor. Gelgelelim bu veri o kadar kesin değildir. Aramaları cinsiyete göre tahmin etmek için, öncelikle, arama verilerini gey pornosu arama yüzdesine dair eyalet çapında bir tahmine ulaşmak için kullanıyorum. Ardından bu veriyi Google AdWords toplumsal cinsiyet verilerine göre normalleştiriyorum. Cinsiyete özgü veriler elde etmenin bir diğer yolu da PornHub verilerinden yararlanmaktır. Gelgelelim PornHub son derece seçilmiş bir örneklem olabilir, zira birçok gey birey muhtemelen sadece gey pornosuna odaklanan siteleri kullanıyor. PornHub erkekler arasında gey pornosu kullanımının Google aramalarının söylediğinden daha düşük olduğunu söylüyor. Gelgelelim eşcinselliğe hoşgörü ile erkek gey pornosu kullanımı arasında güçlü bir ilişki olduğunu doğruluyor. Tüm bu veriler ve diğer notlar kendi internet sitem sethsd.com'da "Seks" başlığı altında bulunabilir.

92 Bu öğrencilerin yaklaşık yüzde 4'ü gey olduğunu Facebook'ta açıkça:

Yazarın Facebook reklam verilerine dair analizinden: 8 Şubat 2017, San Francisco-Oakland-San Jose Facebook ortamındaki kabaca 300 erkek lise öğrencisi erkeklere ilgi duyduğunu söylemişti. Kabaca, 7.800'üyse kadınlara ilgi duyduğunu söylemişti.

95 "Bizde öyle şeyler olmaz":

"'We Don't Have Any Gays in Iran,' Iranian President Tells Ivy League Audience," DailyMail.com, 25 Eylül 2007, http://www.dailymail.co.uk/news/article-483746/We-dont-gays-Iran-Iranian-president-tells-Ivy-League-audience.html.

95 "Bizde onlardan bulunmaz,":

Brett Logiurato, "Sochi Mayor Claims There Are No Gay People in the City," *Sports Illustrated*, 27 Ocak 2014.

95 Soçi'de ve İran'da gey pornosuna epey bir ilgi olduğunu ortaya koyuyor:

Google AdWords'e göre, her yıl on binlerce "гей порно" (gey pornosu) araması yapılmaktadır. Gey pornosu için porno aramalarının yüzdesi Soçi ile ABD'de kabaca aynıdır. Google AdWords'te İran verileri yer almıyor. PornHub da İran verilerini kullanmıyor. Gelgelelim PornMD onların arama verilerini incelemiş ve İran'da ilk ondaki beş arama teriminin gey pornosu olduğunu belirtmiştir. Bunlar arasında "baba sevgisi" ve "hoteldeki işadamı" da vardır ve Joseph Patrick McCormick'in "Survey Reveals Searches for Gay Porn Are Top in Countries Banning Homosexuality" haberinde anlatılmaktadır: *PinkNews*, http://www.pinknews.co.uk/2013/03/13/survey-reveals-searches-for-gay-porn-are-top-in-countries-banning-homosexuality/. Google Trends'e göre, İran'da porno aramalarının yüzde 2 kadarı gey pornosudur ki, ABD'den daha düşük olmasına karşın yine de yaygın bir ilgiye işaret eder.

97 Sıra sekse geldiğinde:

Stephens-Davidowitz, "Searching for Sex." Bu kısımla ilgili veriler internet sitem sethsd.com'daki "Seks" başlıklı kısımda bulunabilir.

97 yüzde 11'i:

"Current Contraceptive Status Among Women Aged 15–44: United States, 2011–2013", Centers for Disease Control and Prevention, http://www.cdc.gov/nchs/data/data briefs/db173_table.pdf#1.

97 her ay yüzde onunun hamile kalması gerektiğini düşünüyor:

David Spiegelhalter, "Sex: What Are the Chances?", *BBC News*, 15 Mart 2012, http://www.bbc.com/future/story/20120313-sex-in-the-city-or-elsewhere.

97 çocuk doğurma yaşına gelmiş kadınlarda 113'te 1:

Her yıl kabaca 6,6 milyon hamilelik vakası görülmektedir ve 15-44 yaşları arasındaki kadın sayısı 62 milyondur.

101 partnerlerine nasıl daha iyi oral seks yapabileceği:

Belirtildiği üzere, Google araştırmacılarının cinsiyetini bilmiyorum. Kadın cinsel organına oral seksin nasıl yapılacağını öğrenmeye çalışan aramaların ezici çoğunluğunun erkekler olduğunu, erkek cinsel organına oral seksin yapılacağını öğrenmeye çalışan aramaların ezici çoğunluğunun kadınlar olduğunu varsayıyorum. Bunun nedeni hem insanların ezici çoğunluğunun heteroseksüel olması hem de hemcins partnerine nasıl oral seks yapılacağını öğrenme ihtiyacının daha az olacağını düşünmemdir.

101 en çok kullanılan beş olumsuz kelime:

Yazarın Google AdWords verilerine dair analizi.

102 Müslümanları öldürmek:

Evan Soltas ve Seth Stephens-Davidowitz, "The Rise of Hate Search", *New York Times*, 13 Aralık 2015, SR1. Veriler ve daha fazla ayrıntı kendi internet sitem sethsd.com'daki "İslamofobi" başlıklı kısımda bulunabilir.

104 on yedi kat:

Yazarın Google AdWords verilerine dair analizi.

104 Martin Luther King Jr. Günü:

Yazarın Google AdWords verilerine dair analizi.

105 siyahlar ile beyazlar arasındaki ücret farklılığıyla bağıntıya sahiptir:

Ashwin Rode ve Anand J. Shukla, "Prejudicial Attitudes and Labor Market Outcomes," mimeo, 2013.

106 Ebeveynleri:

Seth Stephens-Davidowitz, "Google, Tell Me. Is My Son a Genius?" *New York Times*, 19 Ocak 2014, SR6. Kesin aramalarla ilgili veriler Google AdWords kullanarak bulunabilir. Google Trends'ten yararlanarak, "yetenekli" ve "oğlum" ile "yetenekli" ve "kızım" kelimelerine dair aramaları karşılaştırarak da tahminlerde bulunulabilir. Örneğin karş. https://www.google.com/trends/explore?date=all&geo=US&q=gifted%20son,gifted%20daughter ve https://www.google.com/trends/explore?date=all&geo=US&q=overweight%20son,overweight%20daughter. Genel bir örüntü olarak, erkek evlatların beyinleri, kız evlatlarınsa bedenleri hakkında daha fazla soru sorulmasının tek istisnası "şişman oğul" aramasının "şişman kız" aramasından daha fazla olmasıdır. Bu durum yukarıda ele alınan ensest pornosunun popülerliğiyle alakalı görünmektedir. "Şişman" ve "oğul" kelimeleriyle yapılan aramaların kabaca yüzde 20'sinde "porno" kelimesi de vardır.

106 kızların yetenek gerektiren programlarda yer alma ihtimalleri oğlan çocuklarına göre yüzde 9 daha fazladır:

"Gender Equity in Education: A Data Snapshot," Medeni Haklar Dairesi, ABD Eğitim Bakanlığı, Haziran 2012, http://www2.ed.gov/about/offices/list/ocr/docs/gender-equity-in-education.pdf.

107 Kız çocuklarının yaklaşık yüzde 28'i fazla kiloluyken, oğlan çocuklarında bu oran yüzde 35'tir:

Data Resource Center for Child and Adolescent Health, http://www.childhealthdata.org/browse/survey/results?q=2415 &g=455&a=3879&r=1.

107 Stormfront profilleri:

Stephens-Davidowitz, "The Data of Hate." İlgili veriler sethsd.com'daki "Stormfront" başlıklı kısımdan indirilebilir.

109 Donald Trump'ın adaylığı sırasında Stormfront'a ilgi:

Google'da Stormfront aramasına duyulan ilgi Ekim 2016'da, Ekim 2015'tekine benzerdi. Oysa Obama'nın ilk kez başkan seçildiği gün çok farklıydı. Ekim 2008'de Stormfront aramasına duyulan ilgi bir önceki yılın aynı ayına göre neredeyse yüzde 60 artmıştı. Obama seçildikten sonraki gün Google'daki Stormfront aramaları kabaca on kat artmıştı. Trump seçildikten sonraki gün ise Stormfront aramaları yaklaşık iki buçuk kat artmıştı. Bu oran, George W. Bush'un 2004'te seçildikten sonraki artış

oranıyla hemen hemen aynıdır ve siyasetle ilgilenmeyen kesimlerin ilgisini yansıttığı söylenebilir.

110 internette siyasal kutuplaşma:

Matthew Gentzkow ve Jesse M. Shapiro, "Ideological Segregation Online and Offline," *Quarterly Journal of Economics* 126, no. 4 (2011).

112 Facebook arkadaşları:

Eytan Bakshy, Solomon Messing ve Lada A.Adamic, "Exposure to Ideologically Diverse News and Opinion on Facebook," *Science* 348, no. 6239 (2015). Aktif Facebook kullanıcılarından ideolojisini açıkça belirten yüzde 9'luk kesim içinde, arkadaşlarından ideolojisini belirtenlerin yüzde 23'ü karşıt ideolojiyi savunmaktadır ve Facebook'ta gördükleri haberlerin yüzde 28,5'i karşıt ideolojidendir. Bu rakamları kutuplaşma rakamlarıyla doğrudan karşılaştıramayız çünkü ideolojisini belirten az sayıdaki Facebook kullanıcısını baz almaktadır. Muhtemelen bu kullanıcılar siyasi açıdan aktiftir ve aynı ideolojiye sahip olan siyaseten aktif diğer kullanıcılarla bağlantılıdır. Eğer bu doğruysa, tüm kullanıcılar arasındaki çeşitlilik çok daha fazla olacaktır.

112 can alıcı bir nedeni bulunuyor:

Sosyal medyayı şaşırtıcı derecede çeşitli kılan bir diğer etken fevkalade popüler ve çok paylaşılan makalelere, siyasi eğilimi ne olursa olsun, büyük prim yaptırmasıdır. Bkz. Solomon Messing and Sean Westwood, "Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online," 2014.

113 Facebook'ta gerçek hayatta olduğundan daha çok arkadaş:

Bkz. Ben Quinn, "Social Network Users Have Twice as Many Friends Online as in Real Life," *Guardian*, 8 Mayıs 2011. Bu makale Kistik Fibrozis Derneği'nin 2011 yılına ait bir çalışmasını ele alıyor. Buna göre ortalama bir sosyal ağ kullanıcısının 55 gerçek dünyadan arkadaşı varken 121 sanal dünyadan arkadaşı vardır. 2014'teki bir Pew Research çalışmasına göre, ortalama Facebook kullanıcısının 300'den fazla arkadaşı vardır. Bkz. Aaron Smith, "6 New Facts About Facebook," February 3, 2014, http://www.pewre search.org/fact-tank/2014/02/03/6-new-facts-about-facebook/.

113 zayıf bağlar:

Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic, "The Role of Social Networks in Information Diffusion," *Proceedings of the 21st International Conference on World Wide Web*, 2012.

113 "Bittik, mahvolduk tahminleri doğru çıkmadı":

"Study: Child Abuse on Decline in U.S.," Associated Press, 12 Aralık 2011.

114 çocuk istismarı gerilemiş:

Bkz. Seth Stephens-Davidowitz, "How Googling Unmasks Child Abuse," *New York Times*, 14 Temmuz 2013, SR5 ve Seth Stephens-Davidowitz, "Unreported Victims of an Economic Downturn," mimeo, 2013.

114 uzun süreler beklemek zorunda kalıp vazgeçen:

"Stopping Child Abuse: It Begins With You," *The Arizona Republic*, 26 Mart 2016.

115 gebeliklerini sonlandırmak için kayıt dışı yolları:

Seth Stephens-Davidowitz, "The Return of the D.I.Y. Abortion," *New York Times*, 6 Mart 2016, SR2. Veriler ve daha fazla ayrıntı internet sitem sethsd.com'daki "Self-Induced Abortion" başlıklı kısımda bulunabilir.

117 satış rakamıyla benzer ortalamalara:

Alliance for Audited Media, Consumer Magazines, http://abcas3.auditedmedia.com/ecirc/magtitle search.asp.

117 Facebook'ta:

Yazarın hesaplamaları, 4 Ekim 2016, Facebook'un Reklam uygulamasından yararlanarak.

117 en fazla ziyaret edilen on internet sitesinden:

"List of Most Popular Websites," Wikipedia. İnternette dolaşma davranışlarının izini süren Alexa'ya göre, 4 Eylül 2016 itibariyle en popüler porno sitesi XVideos'tu ve en popüler siteler sıralamasında da 57. sıradaydı. SimilarWeb'e göre, 4 Eylül 2016 itibariyle en popüler porno sitesi XVideos'tu ve en popüler siteler sıralamasında da 17. sıradaydı. Alexa'ya göre ilk on: Google, YouTube, Facebook, Baidu, Yahoo!, Amazon, Wikipedia, Tencent QQ, Google India ve Twitter'dır.

119 Facebook 5 Eylül 2006'da:

Bu hikâye David Kirkpatrick imzalıdır, *The Facebook Effect: The Inside Story of the Company That Is Connecting the World* (New York: Simon & Schuster, 2010).

120 büyük işletmelerin sırlar üzerine kurulu olduğunu söyler:

Peter Thiel ve Blake Masters, *Zero to One: Notes on Startups, or How to Build the Future* (New York: The Crown Publishing Group, 2014).

122 Xavier Amatriain:

Xavier Amatriain ile 5 Mayıs 2015'te telefonda görüştüm.

123 Obama'nın 2014:

Yazarın Google Trends verilerine dair analizi.

125 bu kez bir camide:

"The President Speaks at the Islamic Society of Baltimore," YouTube videosu, 3 Şubat 2016'da yayınlandı, https://www.youtube.com/watch?v=LRRVdVqAjdw.

126 Müslüman aleyhtarı nefret dolu, öfkeli aramalar:

Yazarın Google Trends verilerine dair analizi. "Müslümanlara ölüm" aramaları bir hafta önceki karşılaştırılabilir döneme göre daha düşüktü. Ayrıca "Müslümanlar"ın ve bu grupla ilgili olumsuz beş kelimeden birinin yer aldığı aramalar daha düşüktü.

. BÖLÜM

128 hangi beyzbol takımını tuttuğunuzu:

Seth Stephens-Davidowitz, "They Hook You When You're Young," *New York Times*, 20 Nisan 2014, SR5. Bu çalışmayla ilgili veriler ve kod, internet sitem sethsd.com'daki "Bevzbol" baslıklı kısımda bulunabilir.

Yair Ghitza ve Andrew Gelman, "The Great Society, Reagan's Revolution, and Generations of Presidential Voting," yayımlanmamış taslak.

132 Chetty şöyle açıklıyor:

Raj Chetty ile 30 Temmuz 2015'te telefonda görüştüm.

135 ölümden kaçamaz:

Raj Chetty vd., "The Association Between Income and Life Expectancy in the United States, 2001–2014," *JAMA* 315, no. 16 (2016).

136 bulaşıcı davranış olabileceğini:

Julia Belluz, "Income Inequality Is Chipping Away at Americans' Life Expectancy," vox.com, 11 Nisan 2016.

136 neden bazı insanların vergi konusunda dolandırıcılık yaparken:

Raj Chetty, John Friedman ve Emmanuel Saez, "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review 103*, no. 7 (2013).

138 Wikipedia bilgilerini indirmeye karar verdim:

Kaynak: Seth Stephens-Davidowitz, "The Geography of Fame," *New York Times*, 23 Mart 2014, SR6. Veriler internet sitem sethsd.com'daki "Wikipedia Birth Rate, by County" başlıklı kısımda bulunabilir. Wikipedia'daki her bir ismin doğum yerinin indirilmesi ve kodlanmasındaki yardımından ötürü Noah Stephens-Davidowitz'e teşekkür ederim.

140 büyük bir şehir:

Şehirlerin değeri hakkında daha fazla bilgi için, bkz. Ed Glaeser, *Triumph of the City* (New York: Penguin, 2011). (Glaeser üniversitede danışmanımdı.)

145 gerçek hayatın sanatı taklit ettiği:

David Levinson (haz.), *Encyclopedia of Crime and Punishment* (Thousand Oaks, CA: SAGE, 2002).

145 şiddet filmine maruz kalan deneklerin:

Craig Anderson vd., "The Influence of Media Violence on Youth," *Psychological Science in the Public Interest* 4 (2003).

146 Popüler bir şiddet filminin:

Gordon Dahl ve Stefano DellaVigna, "Does Movie Violence Increase Violent Crime?" *Quarterly Journal of Economics* 124, no. 2 (2009).

147 Google aramaları dakika dakika da analiz edilebilir:

Seth Stephens-Davidowitz, "Days of Our Digital Lives," *New York Times*, 5 Temmuz 2015, SR4.

148 Alkolün suça ciddi bir katkısı olduğunu biliyorlardı:

Anna Richardson ve Tracey Budd, "Young Adults, Alcohol, Crime and Disorder," *Criminal Behaviour and Mental Health* 13, no. 1 (2003); Richard A. Scribner, David P. MacKinnon ve James H. Dwyer, "The Risk of Assaultive Violence and Alcohol Availability in Los Angeles County," *American Journal of Public Health* 85, no. 3 (1995); Dennis M. Gorman, Paul W. Speer, Paul J. Gruenewald ve Erich W.

Labouvie, "Spatial Dynamics of Alcohol Availability, Neighborhood Structure and Violent Crime," *Journal of Studies on Alcohol* 62, no. 5 (2001); Tony H. Grubesic, William Alex Pridemore, Dominique A. Williams ve Loni Philip-Tabb, "Alcohol Outlet Density and Violence: The Role of Risky Retailers and Alcohol-Related Expenditures," *Alcohol and Alcoholism* 48, no. 5 (2013).

149 dört çocuğunun hepsinin Amerikan futbolu:

"Ed McCaffrey Knew Christian McCaffrey Would Be Good from the Start—'The Herd,'" YouTube videosu, 3 Aralık 2015'te yayınlandı, https://www.youtube.com/watch?v=boHMmp7DpX0.

149 veri yığınlarını analiz:

Araştırmacılar küçük zaman artışlarına bölünebilen bu suç verilerinden yararlanarak daha çok şey elde ettiler. Bir örnek verelim: Bir şehrin Amerikan futbolu takımı, kazanması beklenen bir maçı kaybettikten hemen sonra ev içi şiddet şikâyetleri artmaktadır. Bkz. David Card ve Gordon B. Dahl, "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *Quarterly Journal of Economics* 126, no. 1 (2011).

150 Bill Simmons:

Bill Simmons, "It's Hard to Say Good-bye to David Ortiz," ESPN.com, 2 Haziran 2009, http://www.espn.com/espnmag/story?id=4223584.

150 bir beyzbol oyuncusunun ileride nasıl performans göstereceğini:

Buna dair bir değerlendirme için, bkz. Nate Silver, *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't* (New York: Penguin, 2012).

151 "İri yarı, vuruşları sert oyuncular":

Ryan Campbell, "How Will Prince Fielder Age?" 28 Ekim 2011, http://www.fangraphs.com/blogs/how-will-prince-fielder-age/.

151 Ortiz'in ruh ikizleri:

Bu veriyi bana sunma nezaketi gösteren Baseball Prospectus'tan Rob McQuown'a minnettarım.

154 diye soruyor Kohane:

Isaac Kohane'le 15 Haziran 2015'te telefonda görüştüm.

155 James Heywood:

James Heywood 17 Ağustos 2015'te telefonda görüştüm.

BÖLÜM

157 27 Şubat 2000:

Bu hikâyenin ele alındığı birçok yerden biri de, Brian Christian, "The A/B Test: Inside the Technology That's Changing the Rules of Business"tır, *Wired*, 25 Nisan 2012, http://www.wired.com/2012/04/ff_abtesting/.

158 Ek ödeme yapıldığında, öğretmenlerin okulu asma oranları:

Esther Duflo, Rema Hanna ve Stephen P. Ryan, "Incentives Work: Getting Teachers to Come to School," *American Economic Review* 102, no. 4 (2012).

158 Bill Gates:

Ian Parker, "The Poverty Lab," *New Yorker*, 17 Mayıs 2010.

159 Google mühendisleri 7 bin:

Christian, "The A/B Test."

159 mavinin 41 farklı çeşidini:

Douglas Bowman, "Goodbye, Google," stopdesign, 20 Mart 2009, http://stopdesign.com/archive/2009/03/20/goodbye-google.html.

160 Facebook günde bir tane:

Eytan Bakshy, "Big Experiments: Big Data's Friend for Making Decisions," 3 Nisan 2014, https://www.facebook.com/notes/facebook-data-science/big-experiments-big-datas-friend-for-making-decisions/10152160441298859/. Eczacılıkla ilgili çalışmalara dair bilgi kaynakları için, bkz. "How many clinical trials are started each year?" Quora paylaşımı, https://www.quora.com/How-many-clinical-trials-are-started-each-year.

160 Optimizely:

Dan Siroker ile 29 Nisan 2015'te telefonda görüştüm.

162 kabaca 60 milyon dolar:

Dan Siroker, "How Obama Raised $60 Million by Running a Simple Experiment," Optimizely, 29 Kasım 2010, https://blog.optimizely.com/2010/11/29/how-obama-raised-60-million-by-running-a-simple-experiment/.

162 *The Boston Globe* A/B-testleri:

The Boston Globe A/B testleri ve sonuçlar yazara ulaştırıldı. *Globe*'un testleri hakkında bazı ayrıntılar "The Boston Globe: Discovering and Optimizing a Value Proposition for Content"'te bulunabilir, Marketing Sherpa Video Archive, https://www.marketingsherpa.com/video/boston-globe-optimization-summit2. Buna *Globe*'tan Peter Doucette ile MECLABS'tan Pamela Markey arasındaki kayda alınmış bir konuşma da dahildir.

164 diyor Benson:

Clark Benson'la 23 Temmuz 2015'te telefonda görüştüm.

165 kutucuğun içinde sağa işaret eden bir ok:

"Enhancing Text Ads on the Google Display Network," Inside Ad-Sense, 3 Aralık 2012, https://adsense.googleblog.com/2012/12/enhancing-text-ads-on-google-display.html.

165 G gle'ın birçok müşterisi tarafından eleştirilmişti:

Örneğin bkz. "Large arrows appearing in google ads—please remove," DoubleClick Publisher Help Forum, https://productforums.google.com/forum/#!topic/dfp/p_TRMqWUF9s.

166 günümüz toplumunda:

Adam Alter, *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked* (New York: Penguin, 2017).

166 Google aramaları bize insanların neyi en fazla müptelalık:

Yazarın Google Trends verilerine dair analizinden.

168 diyor Levitt bir konferansında:

Buna dair bir değerlendirme için, bkz. Harry Walker Speakers Bureau'nun Freakonomics sayfasında şu an yayında olan bir video, http://www.harrywalker.com/speakers/authors-of-freako nomics/.

169 Super Bowl sırasında verilen bira ve meşrubat:

Wesley R. Hartmann ve Daniel Klapper, "Super Bowl Ads," yayınlanmamış taslak, 2014.

171 yaramaz bir pijamalı çocuk:

Şu an bir bilgisayar simülasyonunda yaşıyor olma ihtimalimizi güçlü bir şekilde savunan görüş için, bkz. Nick Bostrom, "Are We Living in a Computer Simulation?" *Philosophical Quarterly* 53, no. 211 (2003).

171 43 ABD başkanından:

Los Angeles Times çalışanları, "U.S. Presidential Assassinations and Attempts," *Los Angeles Times*, 22 Ocak 2012, http://timelines.latimes.com/us-presi dential-assassinations-and-attempts/.

171 John F. Kennedy ile Ronald Reagan:

Benjamin F. Jones ve Benjamin A. Olken, "Do Assassins Really Change History?" *New York Times*, 12 Nisan 2015, SR12.

172 Kadirov öldü:

Saldırının rahatsız edici görüntüleri için, bkz. "Parade surprise (Chechnya 2004)," YouTube video, 31 Mart 2009'da yayınlandı, https://www.youtube.com/watch?v=fHWhs5QkfuY.

172 Hitler ise takvimini değiştirmiş:

Bu hikâye Jones ve Olken tarafından da, "Do Assassins Really Change History?"de ele alınıyor.

172 bir ülkenin liderinin öldürülmesinin sonuçları:

Benjamin F. Jones ve Benjamin A. Olken, "Hit or Miss? The Effect of Assassinations on Institutions and War," *American Economic Journal: Macroeconomics* 1, no. 2 (2009).

173 piyango kazanmanın:

Bu husus John Tierney tarafından, "How to Win the Lottery (Happily)"de ele alınmaktadır, *New York Times*, 27 Mayıs 2014, D5. Tierney'nin yazısı şu çalışmaları ele alıyor: Bénédicte Apouey ve Andrew E. Clark, "Winning Big but Feeling No Better? The Effect of Lottery Prizes on Physical and Mental Health," *Health Economics* 24, no. 5 (2015); Jonathan Gardner ve Andrew J. Oswald, "Money and Mental Wellbeing: A Longitudinal Study of Medium-Sized Lottery Wins," *Journal of Health Economics* 26, no. 1 (2007); ve Anna Hedenus, "At the End of the Rainbow: Post-Winning Life Among Swedish Lottery Winners," yayınlanmamış taslak, 2011. Tierney'nin yazısı da piyangodan para kazanmanın insanı mutlu kılmadığını savunan 1978'deki ünlü çalışmanın (Philip Brickman, Dan Coates ve Ronnie Janoff-Bulman, "Lottery Winners and Accident Victims: Is Happiness Relative?" *Journal of Personality*

and Social Psychology 36, no. 8 (1978)) küçük bir örneklemeye dayandığına dikkat çeker.

173 komşusuna piyango çıkan birinin:

Bkz. Peter Kuhn, Peter Kooreman, Adriaan Soetevent ve Arie Kapteyn, "The Effects of Lottery Prizes on Winners and Their Neighbors: Evidence from the Dutch Postcode Lottery," *American Economic Review* 101, no. 5 (2011) ve Sumit Agarwal, Vyacheslav Mikhed ve Barry Scholnick, "Does Inequality Cause Financial Distress? Evidence from Lottery Winners and Neighboring Bankruptcies," ön makale, 2016.

173 komşunuza piyangodan büyük ikramiye çıkarsa:

Agarwal, Mikhed ve Scholnick, "Does Inequality Cause Financial Distress?"

173 Doktorlara belli işlemlere başvurmaları için ne kadar güçlü finansal teşvikte bulunursanız:

Jeffrey Clemens ve Joshua D. Gottlieb, "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review* 104, no. 4 (2014). Bu sonuçlar doktorların kötü insanlar olduğu anlamına gelmiyor. Aslında doktorların ek para kazanmak için istediği fazladan işlemler gerçekten yaşam kurtarıyorsa, sonuçlar daha da can sıkıcı olabilir. Eğer durum böyleyse, doktorlara yaşam kurtaran tedavileri hastalara önermeleri için para ödenmesi gerektiği sonucu çıkacaktır. Oysa Clemens ve Gottlieb'in bulguları, doktorların kendilerine ne kadar ek ücret ödendiğinden bağımsız olarak, hayat kurtaran tedavileri zaten isteyeceklerini söylüyor. Çok da işe yaramayan işlemler içinse doktorlara yeterli paranın ödenmesi gerekiyor. Bunu şöyle de ifade edebiliriz: Doktorlar hayatı tehdit eden şeyler konusunda parasal teşviklere fazla önem vermezler; önemsiz şeylerde ise bunu epey önemserler.

174 150 milyon dolar:

Robert D. McFadden ve Eben Shapiro, "Finally, a Face to Fit Stuyvesant: A High School of High Achievers Gets a High-Priced Home," *New York Times*, 8 Eylül 1992.

174 55 tane ileri seviye sınıfı…:

Ders içerikleri Stuy'un internet sitesinde bulunabilir, http://stuy.enschool.org/index.jsp.

174 Mezunlarının kabaca dörtte biri:

Anna Bahr, "When the College Admissions Battle Starts at Age 3," *New York Times*, 29 Temmuz 2014, http://www.nytimes.com/2014/07/30/upshot/when-the-college-admissions-battle-starts-at-age-3.html.

174 Stuy'da eğitim görmüştür:

Sewell Chan, "The Obama Team's New York Ties," *New York Times*, 25 Kasım 2008; Evan T. R. Rosenman, "Class of 1984: Lisa Randall," *Harvard Crimson*, 2 Haziran 2009; "Gary Shteyngart on Stuyvesant High School: My New York," YouTube videosu, 4 Ağustos 2010'da yayınlandı, https://www.youtube.com/watch?v=NQ_phGkC-Tk; Candace Amos, "30 Stars Who Attended NYC Public Schools," *New York Daily News*, 29 Mayıs 2015.

174 Açılış konuşmacıları arasında:

Carl Campanile, "Kids Stuy High Over Bubba: He'll Address Ground Zero School's Graduation," *New York Post*, 22 Mart 2002; Birleşmiş Milletler Basın Bülteni, "Stuyvesant High School's 'Multicultural Tapestry' Eloquent Response to Hatred, Says Secretary-General in Graduation Address," 23 Haziran 2004; "Conan O'Brien's Speech at Stuyvesant's Class of 2006 Graduation in Lincoln Center," YouTube videosu, 6 Mayıs 2012, https://www.youtube.com/watch?v=zAMkUE9Oxnc.

174 Stuy birinci sırayı almıştır:

Bkz. https://k12.niche.com/rankings/public-high-schools/best-overall/.

175 Sınava girenlerin yüzde beşinden azı:

Pamela Wheaton, "8th-Graders Get High School Admissions Results," *Insideschools*, 4 Mart 2016, http://insideschools.org/blog/item/1001064-8th-graders-get-high-school-admissions-results.

177 koşullara mahkûm edilen mahpusların:

M. Keith Chen ve Jesse M. Shapiro, "Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-Based Approach," *American Law and Economics Review* 9, no. 1 (2007).

177 Stuyvesant Lisesi'nin etkileri?

Atila Abdulkadiroğlu, Joshua Angrist ve Parag Pathak, "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica* 82, no. 1 (2014). Aynı sonuca Will Dobbie ve Roland G. Fryer Jr. da ulaşmıştır, "The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools," *American Economic Journal: Applied Economics* 6, no. 3 (2014).

179 Ortalama bir Harvard mezunu:

Bkz. http://www.payscale.com/college-salary-report/bachelors.

179 benzer kökenlerden gelen ve aynı okullara kabul edilen:

Stacy Berg Dale ve Alan B. Krueger, "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables," *Quarterly Journal of Economics* 117, no. 4 (2002).

179 Warren Buffett:

Alice Schroeder, *The Snowball: Warren Buffett and the Business of Life* (New York: Bantam, 2008).

7. BÖLÜM

185 piyasaların seyrini doğru tahmin edebileceklerini iddia eden:

Johan Bollen, Huina Mao ve Xiaojun Zeng, "Twitter Mood Predicts the Stock Market," *Journal of Computational Science* 2, no. 1 (2011).

186 Piyasaları tweet'lerle ölçmeye çalışan serbest yatırım fonları:

James Mackintosh, "Hedge Fund That Traded Based on Social Media Signals Didn't Work Out," *Financial Times*, 25 Mayıs 2012.

187 hiçbiriyle bağıntı bulamadılar:

Christopher F. Chabris vd., "Most Reported Genetic Associations with General Intelligence Are Probably False Positives," *Psychological Science* (2012).

189 Zoë Chance:

Bu hikâye TEDx Talks, "How to Make a Behavior Addictive: Zoë Chance at TEDx Mill River"da ele alınmaktadır, YouTube videosu, 14 Mayıs 2013'te yayınlandı, https://www.youtube.com/watch?v=AHfiKav9fcQ. Hikâyenin bazı ayrıntıları (adım ölçerin rengi gibi), mülakatlarda somutlaştırılmıştı. Chance'la 20 Nisan 2015'te telefonla, 11 Temmuz ve 8 Eylül 2016'daysa e-posta yoluyla görüştüm.

189 Sayılar baştan çıkarıcı olabilir:

Bu kısım Alex Peysakhovich ve Seth Stephens-Davidowitz, "How Not to Drown in Numbers"tan alınmıştır, *New York Times*, 3 Mayıs 2015, SR6.

190 bu testlerde alenen hile yapmışlardır:

Brian A. Jacob ve Steven D. Levitt, "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics* 118, no. 3 (2003).

191 Thomas Kane:

Thomas Kane'le 22 Nisan 2015'te telefonda görüştüm.

191 "Her ölçüm değerli bir şey katıyor":

Bill ve Melinda Gates Vakfı, "Ensuring Fair and Reliable Measures of Effective Teaching," http://k12education.gatesfoundation.org/wp-content/uploads/2015/05/MET_Ensuring_Fair_and_Reliable_Measures_Prac titioner_Brief.pdf.

8. BÖLÜM

193 Geçenlerde üç iktisatçı:

Oded Netzer, Alain Lemaire ve Michal Herzenstein, "When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications," 2016.

193 yüzde 13 civarındadır:

Peter Renton, "Another Analysis of Default Rates at Lending Club and Prosper," 25 Ekim 2012, http://www.lendacademy.com/lending-club-prosper-default-rates/.

196 Facebook beğenilerinin çoğu zaman IQ ile:

Michal Kosinski, David Stillwell ve Thore Graepel, "Private Traits and Attributes Are Predictable from Digital Records of Human Behavior," *PNAS* 110, no. 15 (2013).

198 Yelp değerlendirmelerinin insafına kaldığını gösteriyor:

Michael Luca, "Reviews, Reputation, and Revenue: The Case of Yelp," yayınlanmamış taslak, 2011.

199 İntiharla alakalı Google aramaları:

Christine Ma-Kellams, Flora Or, Ji Hyun Baek ve Ichiro Kawachi, "Rethinking Suicide Surveillance: Google Search Data and Self-Reported Suicidality Differentially Estimate Completed Suicide Risk," *Clinical Psychological Science* 4, no. 3 (2016).

200 3,5 milyon Google aramaları:

Burada internet sitemde kendi kendine kürtaj kısmında ele alınan bir metodolojiden yararlanılıyor. Google'ın "intihar" kategorisindeki aramaları "kravat nasıl bağlanır" aramalarıyla karşılaştırıyorum. 2015'te Google'da 6,6 milyon tane "kravat nasıl bağlanır" araması yapıldı. İntihar kategorisindeki aramalar bunun 6,5 katı daha fazlaydı. 6.5***6.6/12 » 3.5.

20[illegible] Müslüman sayısıysa 12'ydi:

Bridge Initiative Team, "When Islamophobia Turns Violent: The 2016 U.S. Presidential Election," 2 Mayıs 2016, http://bridge.georgetown.edu/when-islamophobia-turns-violent-the-2016-u-s-presidential-elections/.

SONUÇ

203 Popper'ın eleştirisinde:

Karl Popper, *Conjectures and Refutations* (Londra: Routledge & Kegan Paul, 1963).

205 şehirdeki her bir kolera vakası:

Simon Rogers, "John Snow's Data Journalism: The Cholera Map That Changed the World," *The Guardian*, 15 Mart 2013.

206 Benjamin F. Jones:

Benjamin Jones'la 1 Haziran 2015'te telefonda görüştüm. Bu çalışmayı ele alan başka bir yazı için, bkz. Aaron Chatterji ve Benjamin Jones, "Harnessing Technology to Improve K–12 Education," Hamilton Projesi Münazara Makalesi, 2012.

211 iktisatçıların kitaplarını:

Jordan Ellenberg, "The Summer's Most Unread Book Is . . . ," *Wall Street Journal*, 3 Temmuz 2014.

Dizin

127 Saat (film) 75, 76

A

ABD
  hamilelik 142, 143, 144
  ırkçılık 12
Abdulkadiroğlu, Atila 177, 239
A/B testleri 162, 206, 236
Adamic, Lada 112, 232
Adams, John 67, 125
adım ölçer 189
Afrikalı-Amerikalılar 14, 36, 37, 102, 104, 105, 120, 122, 139. Ayrıca bkz. "*nigger*", ırk/ırkçılık çağı
Aiden, Erez 65, 66, 67
Almanya 76
  hamilelik 144
AltaVista (arama motoru) 54
Alter, Adam 166, 167, 236
altın 20, 147, 158, 162, 188
Amatriain, Xavier 122, 233
Amazon 24, 153, 154, 222, 233
Amerikan Firavunu (85 numaralı at) 25, 57, 61, 191
Amerikan İç Savaşı 66, 67
Angrist, Joshua 177, 239
anketler 90
anksiyete 23, 206
Ashenfelter, Orley 62, 63, 225
Asher, Sam 153, 214
Asyalılar 102, 109
aşerme 142
Atlantic dergisi 117, 118, 153
atlar 58, 60, 62, 63, 192
Avatar (film) 168
Avustralya 224
  hamilelik 142, 143
ayrımcılık 105, 139, 196, 209
Azınlık Raporu (film) 199

B

bağımlılık 166, 167
Bakshy, Eytan 112, 232, 236
Baltimore Ravens-New England Patriots maçları 167, 168
Bartleby sendromu 58
basketbol 13, 35, 38, 39, 40
basketbolcular 58
Beane, Billy 190
bedenler, veri olarak 56-63
Beethoven, Ludwig von 145
Benson, Clark 164, 165, 236
Berger, Jonah 76, 227
beyaz 13, 15, 16, 18, 35, 38, 55, 104, 105, 108, 109, 110, 112, 113, 122, 134, 140, 189
beyaz milliyetçiliği 109
beyzbol 131, 155, 197, 208, 233, 235
  bir oyuncunun geleceğini tahmin etmek 150, 151
  çocukluk deneyimlerinin etkileri 127, 128, 129
  ölçülebilirliğe vurgu 154, 190, 204
Bezos, Jeff 154
Big Papi 149

bilim 18, 203, 204, 206
Billings (Montana) *Gazette* 79
Bill ve Melinda Gates Vakfı 240
Bing (arama motoru) ve Columbia Üniversitesi-Microsoft pankreas kanseri araştırması 31, 32
bireyler 95, 155, 208
Black, Don 12, 107, 217, 220
Blink (Gladwell) 32
Bloodstock, Incardo 56
Boehner, John 123
Booking.com 199
borcunu ödeme 193, 194
borç para alma 195
Bosh, Chris 37, 220
Boston Maratonu 23
Boston Red Sox 150
Brezilya 188
Brin, Sergey 54, 55, 84, 224
Britanya, hamilelik 142, 143, 144
Bronx Fen Lisesi (New York City) 175, 178
Buffett, Warren 179, 180, 239
bulaşıcı davranış 136, 234
Bullock, Sandra 141
Bundy, Ted 138
Bush, George W. 59, 231
Büyük Veri 23, 24, 25, 27, 29, 30, 32, 44, 45, 59, 90, 124, 128, 145, 146, 150, 154, 159, 167, 172, 180, 181, 183, 203, 205, 206, 208, 211, 213
  bilim 68, 69
  boyutsallık laneti 185, 186, 187, 188
  değeri/önemi 53
  dijital hakikat serumu 49, 88, 117
  dil 65
  dürüst veriler 49, 204
  Facebook 91, 117
  Google 54, 55, 93, 116
  güçlü yanları 21, 22, 46, 48, 49, 88, 131, 159, 193
  insanların gelecekte yapacaklarını tahmin etmek 150, 199
  ölçülebilir olana aşırı vurgu 189, 190, 191, 192
  şirketleri 25, 186, 198
  şüpheyle yaklaşanlar 132
  tanım 20

C-Ç

Calhoun, Jim 38
Chabris, Christopher 187, 214, 240
Chance, Zoë 189, 240
Chaplin, Charlie 23
Chen, M. Keith 177, 239
Chetty, Raj 132, 133, 134, 135, 136, 137, 141, 204, 214, 234
Churchill, Winston 130
cinayet 108, 146, 199
cinsiyet 106, 108, 138, 208, 229
Clemens, Jeffrey 173, 174, 238
Clinton, Bill 174, 224
Clinton, Bill (aramaları) 54, 55
Clinton, Hillary 17, 18
cnn.com 112, 113
coğrafya 142
Cohen, Leonard 68, 69, 127, 225
Columbia Üniversitesi 31, 130
Country Music Radio 152
Craigslist 93
Cundiff, Billy 169
Cutler, David 136, 213, 214
çocuk
  -larda öğrenme 206, 207
  boy ve kilo verileri 107
  istismarı 20, 113, 114, 122, 124, 184, 232
  pornosu 96

sağlığı 220
yapma kararı 10, 89, 90, 229
zekâ 106
çocukluk 38, 46, 47, 48

D

Dahl, Gordon 146, 148, 149, 234, 235
Dale, Stacy 179, 239
davranış bilimi ve dijital devrim 208
DellaVigna, Stefano 146, 149, 234
Demokratlar (ABD) 16, 77, 78
depresyon 21, 33, 73, 74, 206
dijital dünya 159
dijital hakikat serumu 49, 85, 88, 99, 103, 118, 119, 121, 122, 124, 125
dijital yalanlar 118
dijitalleşme 65, 180
dil 26, 44, 45, 64, 65, 70, 136, 174, 193
din 20, 103, 136
DNA 186, 187
doğal deneyler 171, 172, 177
doğru soruları sormak 25
doktorlar 31, 32, 154, 171, 173, 174
Donato, Adriana 199, 201
Duflo, Esther 158, 159, 204, 235
duygu analizi 73, 75, 76
Dünya Bankası 83, 84
dünyayı anlamak 12, 55, 192
dürüstlük 88

E

ebeveynler 106, 107
Edmonton, su tüketimi 147, 155
EDU STAR 206
eğitim 16, 23, 33, 56, 149, 158, 174, 178, 179, 191, 238
yazılımı 206
Eisenhower, Dwight D. 130, 131
Ellenberg, Jordan 211, 241
Ellerbee, William 35
Eng, Jessica 178
ensest 46, 47, 96, 231
EPCOR 147
EQB 56
ergenler 87
erkekler 13, 14, 22, 24, 30, 35, 36, 37, 46, 47, 48, 56, 67, 68, 69, 70, 85, 135, 138, 148, 180, 184, 201, 208, 223, 229, 230
etik
Büyük Veri 167, 193-199
deneyler 171
IQ-DNA çalışmaları 187
işe alma pratikleri 194, 195, 196
evlilik 14, 29
Ewing, Patrick 34, 219, 220

F

Facebook 20, 21, 24, 33, 69, 73, 74, 82, 91, 92, 93, 112, 113, 116, 117, 118, 119, 120, 128, 129, 131, 153, 160, 166, 167, 179, 190, 191, 195, 196, 210, 219, 226, 229, 232, 233, 236, 240
Facemash 120
fallik semboller 44
Farook, Rizwan 102
fırsat eşitliği 133
Filipinler 84
filmler 121, 122, 146, 148
fizik 152, 174
flört 10, 109, 203
fotoğraflar 82, 84
Frances Kasırgası 62
Freakonomics 199, 237
Freud, Sigmund 7, 43, 44, 46, 47, 48, 49, 96, 203, 210
Friedman, Jerry 24, 25, 219, 234
Fryer, Roland 36, 220, 239
futbol 54, 69, 73, 149, 170

G

Gabriel, Stuart 17, 18, 224
Gana 144
Gangnam Style 118
Garland, Judy 91
Gates, Bill 158, 179, 191, 236, 240
Gelles, Richard 113
Gelman, Andrew 130, 228, 234
genetik 186, 187
Gentzkow, Matt 64, 65, 77, 78, 79, 80, 111, 112, 214, 227, 232
gerçek bilim 203
gey 25, 91, 102, 122, 123, 180, 208, 228, 230
  evlilikleri 64, 77, 90
  İran 95
  kimliğini saklama 91
  sayısı 90, 92, 95
  ve pornografi 92, 93
Ghitza, Yair 130, 234
Ginsberg, Jeremy 52, 224
Gladwell, Malcolm 32
Gnau, Scott 198
Goldman Sachs 51, 54
Google AdWords 92, 99, 229, 230, 231
Google aramaları 14, 15, 16, 18, 47, 53, 54, 61, 65, 97, 101, 117, 122, 125, 147, 166, 199, 218, 234, 236, 240
Google Correlate 52, 53
Google Flu 52, 61, 224
Google Ngrams 65, 66, 67
Google STD 61
Google Trends 12, 13, 14, 185, 217, 229, 230, 231, 233, 236
Gottlieb, Joshua 153, 173, 174, 238
Görünmeyen Ekonomi (Levitt) 209
Great Body, Great Sex, Great Blowjob 118, 119
Grinin Elli Tonu 121
GSYH 82, 83, 227
Guttmacher Enstitüsü 115, 116
Güney Afrika, hamilelik 143

H

hakikat 49, 88, 99, 103, 117, 119, 122, 124, 125, 126
hamilelik 97, 142, 143, 144, 230
Hannibal (film) 146, 148
Harris, Tristan 167
Harry Potter ve Ölüm Yadigârları (Rowling) 74, 75
Hartmann, Wesley R. 237
Harvard Crimson 120, 238
Hata Botu 45
hava durumu 88, 89, 228
Henderson, J. Vernon, 82, 227
Herzenstein, Michal 193, 240
Heywood, James 155, 235
Hıristiyanlar 102
Hızlı ve Yavaş Düşünme (Kahneman) 211
Hindistan 23, 24, 25, 47, 158
  hamilelik 142, 143, 144
Hitler, Adolf 107, 172, 237
Hüseyin, Saddam 78

I-İ

Indiana Üniversitesi 186
Instagram 82, 118, 195
Irak Savaşı 78
ırk 12, 35, 187, 217
iktisat 12, 64, 168, 209, 213
iktisatçılar 38, 110, 156, 170, 172, 173, 177, 178, 183
ilaç şirketleri 159
İnsan Genomu Projesi 186
İnsan Hakları Kampanyası 124
insanlık 22
internet 10, 11, 12, 35, 54, 55, 59, 64, 87,

88, 95, 99, 103, 108, 109, 110, 111, 112, 113, 118, 120, 121, 125, 144, 153, 158, 162, 166, 189, 195, 199, 217, 220, 229, 230, 231, 233, 234, 238, 241
intihar 199, 200, 241
İran 95, 230
İslamofobi 102, 125, 231
İspanya 220
hamilelik 144
işçi hakları 78
işe alma pratikleri 195
işsizlik 51, 52, 53, 109, 114

J

Jacob, Brian 190, 214, 240
James, Bill 35, 37, 40, 58, 67, 150, 153, 155, 199, 215, 221, 225, 234, 235, 239
James, LeBron 35, 37, 40, 58, 67, 150, 153, 155, 199, 215, 221, 225, 234, 235, 239
Jawbone 207
Ji Hyun Baek 199, 240
Jobs, Steve 141
Johnson, Earvin III 58, 131, 173, 225
Johnson, Lyndon B. 58, 131, 173, 225
Jones, Benjamin F. 172, 173, 186, 206, 237, 241
Jordan, Jeffrey 39, 58, 211, 221, 241
Jordan, Marcus 39, 58, 211, 221, 241
Jordan, Michael 39, 58, 211, 221, 241
Jurafsky, Dan 67, 225

K

Kaçak Gelin (film) 146, 148
kadınlar 13, 14, 46, 47, 56, 68, 69, 70, 73, 81, 85, 92, 95, 96, 99, 100, 101, 115, 116, 135, 142, 143, 180, 230
Kadirov, Ahmed 172, 237
Kahneman, Daniel 9, 211, 219
kalça 99
Kane, Thomas 191, 240
kanser 31
Katrina Kasırgası 104
Katz, Lawrence 183, 213
Kaufmann, Sarah 178
Kawachi, Ichiro 199, 240
Kayak 199
Kazanılmış Gelir Vergisi Kredisi 136, 137
kelimeler, veri olarak 64-80
Kennedy, John F. 131, 171, 172, 224, 237
Kerry, John 16, 183
kız çocukları 106
King, Martin Luther Jr. 104, 231
King, William Lyon Mackenzie (Müstear ad) 109
Kinsey, Alfred 90, 210
Kirkpatrick, David 119, 233
kitaplar, dijitalleşme 65, 66, 67, 80
Klapper, Daniel 170, 237
Kodak 82
Kohane, Isaac 154, 155, 235
kolera 205, 241
kondom 13, 14, 97, 100, 217
konut fiyatları 53, 224
Kral John (Shakespeare) 75
Krueger, Alan B. 52, 179, 223, 239
Kubrick, Stanley 145
Ku Klux Klan 18, 107
kumarhaneler 197
Kundera, Milan 176
kutuplaşma 111, 112, 232
kürtaj 20, 115, 116, 122, 184, 241

L

Lemaire, Alain 193, 240
Levitt, Steven 36, 168, 170, 190, 209, 220, 237, 240

liberaller 110, 111
Linden, Greg 153, 154
Luca, Michael 198, 240
Lycos 54

M

Macon County, Alabama 139, 142
Magic 58
Ma-Kellams, Christine 199, 240
Malik, Tashfeen 102
Manchester Üniversitesi 186
Maryland Üniversitesi 86, 228
Matthews, Dylan 153, 217
McCaffrey, Ed 149, 235
McFarland, Daniel 67, 225
McPherson, James 67
medya 64, 79, 103, 108, 112, 118, 124, 153, 195, 227
Meksikalılar 102
merak 29, 62, 64, 77, 138, 168, 175, 209, 211
Messing, Solomon 112, 214, 232
MetaCrawler 54
MIT 141, 158, 177, 204
Michel, Jean-Baptiste 65, 66, 67
Microsoft 31, 32, 45, 179, 196, 222
Milkman, Katherine L. 76, 227
Minsky, Marvin 204
mizah 29, 30
Moore, Julianne 141
Moskovitz, Dustin 179
msnbc.com 112
muhafazakârlar 64, 110, 111
Murdoch, Rupert 79, 227
Murray, Patty 191, 192, 224
mülkiyet hakları 78
Müslümanlar 102, 233
müşteriler 199

N

Nantz, Jim 169
National Enquirer 117, 118
NBA 34, 35, 36, 37, 38, 39, 40, 58, 150, 220
nedensellik 157, 168, 179
nefret 13, 14, 15, 17, 18, 90, 102, 103, 104 107, 110, 119, 122, 126, 127, 141, 145, 179, 199, 233
Netflix 118, 121, 122, 153, 160
Netzer, Oded 193, 240
New Jack City (film) 145
News Corporation 79
newslibrary.com 78
New York City 13, 57, 239
New York dergisi 160
New Yorker 153, 158, 236
New York Mets 127
New York Post 79, 227, 239
New York Times 20, 55, 76, 79, 103, 107, 109, 113, 124, 187, 210, 213, 217, 218, 219, 223, 224, 225, 228, 231, 232, 233, 234, 237, 238, 240
Nietzsche, Friedrich 48, 200
nigger 14, 15, 20, 104, 105, 183, 218
Nijerya 83
  hamilelik 143, 144
Nixon, Richard M. 131

O-Ö

Obama, Barack 11, 12, 15, 16, 17, 18, 19, 20, 52, 103, 104, 105, 109, 123, 12 126, 160, 162, 166, 174, 180, 183, 217, 231, 233, 236, 238
Ocala at mezarı 57, 59, 60
OkCupid 109
Olken, Benjamin A. 172, 237
oral 101, 230
ortalama yaşam süresi 135
Ortiz, David 149, 150, 151, 152, 153, 235

Otomatik Portakal (film) 145
otomatik tamamlama 88, 89
Oyuncak Hikâyesi (film) 146
ölçekli bilim 206
Ölülerin Şafağı (film) 146
ölüm 34, 73, 78, 114, 116, 174, 199, 200, 201, 233
önyargı 12, 104, 106

P

Page, Larry 54, 55, 84, 224
Pandora 153
pankreas kanseri 31, 32, 206
Parks, Rosa 78
Parr, Ben 119
Pathak, Parag 177, 239
PatientsLikeMe.com 155
Paul, Chris 37, 220, 234, 241
penis 24, 43, 45, 98
penistrian 43, 45, 46
Peysakhovich, Alex 190, 214, 240
Philadelphia Daily News 79
Pierson, Emma 124, 214
Piketty, Thomas 211
Pinky Pizwaanski 60
piyango 173, 237, 238
pizza 44
PlentyOfFish 109
Plomin, Robert 187
Popper, Karl 43, 203, 204, 223, 241
Popp, Noah 153, 214
Pop-Tart 62, 225
PornHub 20, 46, 47, 48, 49, 92, 95, 96, 205, 218, 229, 230
pornografi 48, 88, 96
Posada, Jorge 151
proquest.com 78
Prosper 193, 240
psikoloji 12, 207

Q-R

Quantcast 107
Rawlings, Craig 67, 225
Reagan, Andy 74, 76, 171, 172, 227, 234, 237
Reagan, Ronald 74, 76, 171, 172, 227, 234, 237
regresyon süreksizliği 177
Reisinger, Joseph 83, 84, 227
reklamlar 169, 170, 200
  Google 159, 165, 166
  ve satış 168
  ve Super owl maçları 169, 170
  televizyon 167, 168
Renkler (film) 145
resimler, veri olarak 80-4
Rocket Tube 92
Rolling Stones 208
Romney, Mitt 17, 160
Roseau County, Minnesota 142
Rönesans 185
ruh ikizleri 151, 152, 153, 154, 155, 235
Rusya 95
rüyalar ve fallik semboller 43, 44

S-Ş

saatler 9, 126, 210
Sabermetrik 150
sağlık 20, 25, 31, 37, 52, 61, 136, 145, 154, 155, 158, 205, 208
sağlık sigortası 136
Saka Kuşu (Tartt) 211
sanat 122, 138-9, 145, 159, 234
San Bernardino, California 102
Sands, Emily 153
seçimler, 2008 ABD 11, 105, 109
seçimler, 2012 ABD 105

seçimler, 2016 ABD 105
Seder, Jeff 56, 57, 59, 60, 61, 62, 63, 120, 191, 192, 224
seks 10, 13, 14, 22, 45, 47, 92, 94, 95, 96, 97, 101, 119, 123, 145, 166, 205, 208, 214, 222, 230
sezgi 32
Shadow 44
Shakespeare, William 75, 107
Shapiro, Jesse 64, 65, 77, 78, 79, 80, 110, 111, 112, 177, 204, 213, 214, 227, 232, 238, 239
Shattered 208
sigara ekonomisi (Filipinler) 85
Silver, Nate 17, 19, 20, 105, 151, 152, 190, 229, 235
Simmons, Bill 150, 151, 235
Siroker, Dan 160, 236
siyaset 12, 104, 130, 205, 213
Slutload 53
Smith, Michael D. 169, 214, 232
Snow, John 205, 206, 241
Soçi, Rusya 95, 230
Soltas, Evan 103, 199, 214, 231
sorular 64, 86, 100, 101, 123, 143, 144, 162, 167, 191, 195, 207, 228
sosyal bilimler 10, 65, 178, 183
sosyal medya 108, 118, 124, 195
Spider Solitaire 53
Stephens-Davidowitz, Noah 2, 10, 214, 217, 218, 219, 224, 230, 231, 232, 233, 234, 240
Stephens-Davidowitz, Seth 2, 10, 214, 217, 218, 219, 224, 230, 231, 232, 233, 234, 240
Stoneham, James 199, 201
Stone, Oliver 141
Stormfront 15, 20, 23, 107, 108, 109, 110, 112
Stormfront.org 107
Stuyvesant Lisesi 174, 175, 176, 177, 179, 239
suç 96, 134, 142, 146, 148, 183, 199, 200, 201, 204, 235
  hapishane koşulları 177
  ve şiddet filmleri 149
suikastlar 171-2
Summers, Lawrence 183, 184, 185, 188, 214
Sunstein, Cass 110
Super Bowl 168, 169, 170, 180, 237
Super Crunchers (Gnau) 198
Surowiecki, James 153
su tüketimi 147
şakalar 14, 21
şarap 62, 63, 143, 144, 157
şecere 59, 61
şiddet 77, 96, 103, 145, 146, 148, 149, 177, 204, 234, 235
şirketler 49, 160, 168, 169, 195, 199

T

tahminler 185, 196, 229
Taleb, Nassim 22, 218
taraflılık 104, 138
Tartt, Donna 211
TaskRabbit 160
tecavüz 96, 122, 145, 146
televizyon 40, 104, 125, 167, 168
testler 49, 164, 167, 190, 196
The Boston Globe ve A/B testleri 162, 236
The Facebook Effect (Kirkpatrick) 119, 233
The Herd with Colin Cowherd 149
The Signal and the Noise 190, 235
Thiel, Peter 120, 233
Think Progress 103
Thome, Jim 151
Tourangeau, Roger 86, 87, 228

Trump, Donald 10, 11, 15, 16, 17, 18, 19, 20, 86, 105, 109, 140, 218, 228, 231, 232
Tuskegee Üniversitesi 139
tüketiciler 199
Twitter 21, 40, 118, 124, 152, 153, 233, 239

U-Ü

Uncharted (Aiden ve Michel) 66, 67
uyku 51, 207
üniversiteler 142, 171, 174, 175, 176, 177, 178, 179

V

vajina 125
Varian, Hal 52, 169, 213, 219, 224
veri bilimi 29, 30, 32, 33, 34, 43, 59, 92, 125, 150, 172, 197, 198, 200, 202, 219
veri hikâyeleri 155, 156
Vox 153
vurgu 78, 178

W

Walmart 61, 62
Washington Post 64, 78
Washington Times 64, 78
Weil, David N. 82, 227
Weiner, Anthony 176
Wikipedia 20, 138, 139, 140, 141, 142, 233, 234
World of Warcraft 167
Wrenn, Doug 38, 39, 40, 221

Y

Yahoo News 112
Yahudiler 101, 102
yakından bakmak 94, 118, 146, 180, 206
yalan 2, 14, 17, 18, 25, 74, 85, 86, 87, 88, 90, 100, 122, 137, 180, 195, 210
yaratıcılık 209
yaş 16, 47, 71, 72, 73, 87, 97, 108, 116, 127, 129, 130, 131, 151
Yelp 198, 240
yeni bilgiler 13, 125
Yeşil Maymun 59, 225
Yılmaz, Ahmed 174, 175, 176, 177
Yirmi Birinci Yüzyılda Kapital (Piketty) 211
yiyecekler 44, 142
yoksulluk 142
YouTube 118, 206, 219, 220, 221, 225, 233, 235, 237, 238, 239, 240

Z

Zayat, Ahmed 56, 57, 224
zekâ 106, 196
zenginlik 9, 76
Zero to One (Thiel) 120, 233
Zuckerberg, Mark 119, 120, 122, 135, 179



978-605-2116-45-6

Bana Yalan Söylediler

İnternet ve Gerçek Yüzümüz

SETH STEPHENS-DAVIDOWITZ

SOSYOLOJİ | VERİ BİLİM

İnsanlar ne sıklıkta seks yapıyorlar?
Reklam vermek işe yarıyor mu?
Kimler vergi kaçırıyor?
Geylerin sayısı belli mi?
Hangi üniversiteye gittiğimin bir önemi var mı?
Borsayı kandırabilir miyiz?
Medya taraflı haber yapıyor mu?
Çocuk büyütmek için en iyi yerler nereleri?
İlk randevuda nelerden bahsetmeliyim ki ikincisi olsun?

Veri bilimcisi Seth Stephens-Davidowitz, çığır açıcı çalışmasında, insanlara dair kalıplaşmış düşüncelerimizin ne ölçüde hatalı olduğunu ortaya koyuyor. Peki, bunun nedeni ne? Çünkü insanlar dostlarına, sevdiklerine, doktorlarına ve anketlere, hatta kendilerine bile yalan söyler. Bugün artık insanların söylediklerleriyle yetinmek zorunda değiliz. İnternetin sağladığı yeni bilgi türü hakikati ortaya koyuyor. Google'da, sosyal medyada, flört sitelerinde, hatta porno içerikli sitelerde bıraktığımız her iz, hakkımızda bir şeyler söylüyor. Bu dijital altın madenini araştıran uzmanlar insanların gerçekte ne düşündüklerini, istediklerini ve yaptıklarını keşfediyorlar.

Doğru soruları sorduğumuz takdirde, Büyük Veri'den insan doğasına ilişkin öğrenebileceklerimizin sınırı yok. Bu alanın öncüleri arasında yer alan Seth Stephens-Davidowitz'in sunduğu yeni veriler sizi zaman zaman kahkahaya boğarken, kimi yerde dumura uğratacak ya da rahatsız edecek. *Bana Yalan Söylediler*'de karşınıza çıkacak verilerin tamamı sizi hem insanlığa hem de dünyanın günümüzdeki haline dair düşündürecek.

Seth Stephens-Davidowitz Wharton School'da ders veriyor, *The New York Times*'ta yazıları yayımlanıyor.

İngilizceden çeviren: Ferit Burak Aydar

KÜY

22 TL

978-605-2116-45-6