ANALYTICS THE RIGHT WAY

A BUSINESS LEADER'S GUIDE TO PUTTING DATA TO PRODUCTIVE USE

TIM WILSON
JOE SUTHERLAND

Analytics the Right Way

Analytics the Right Way

A Business Leader's Guide to Putting Data to Productive Use

TIM WILSON
JOE SUTHERLAND

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

ISBN: 978-1-394-26449-0 (paperback)
ISBN: 978-1-394-26451-3 (ePDF)
ISBN: 978-1-394-26450-6 (epub)
LCCN: 2024946991

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Cover Image: © Paul M. Lyren
Cover Design: Wiley

To Julie and Sarah

Contents

Acknowledgments

Together, we'd like to thank our publishing and editorial team at Wiley—Jim Minatel, Cathleen Small, Pete Gaughan, and Sara Deichman—without whom this book wouldn't have been possible. Paul Lyren provided both comical wit and a talented hand in illustrating our book, taking our goofy ideas of what passed for "comedy" and making them actually funny; the herculean task of illustrating no fewer than 11 chapters worth of content about analytics—yuck!—is a feat, and we deeply appreciate him doing so. Mike Gustafson and our many colleagues at Search Discovery deserve credit for bringing the two of us together, and for supporting our harebrained ideas, for whatever reason—it was Mike's light touch that led us to build a friendship and point of view for how the analytics business is and should be.

Tim: I would like to thank everyone from whom I learned something that worked its way into the ideas and examples in this book. But then the editors said that would be several orders of magnitude beyond the appropriate word count, so this list is tragically abbreviated. Matt Gershoff is quoted in the book, and I've learned that "minimization of regret" means jumping at every opportunity to have a drink, a meal, or a walk with him—be the topic reinforcement learning, decision-making at the margin, the cost of reducing uncertainty, or the relative merits of European versus American restaurant payment systems, it's always a delight. Another Matt—Matt Coen—is responsible for giving me the language of "the two magic questions." I was loosely applying the idea, but it was during our relatively brief work together when he articulated them (it was purely my cheekiness that then branded them as "magic"). John Lovett—a coworker twice over—provided an on-the-job master class in how to listen to business partners and clients effectively, and he provided encouragement and advice ever since this book was the faintest of glimmers in my eye back in 2017. Val Kroll and I have been collaborating for more than a decade, although we didn't realize we were codeveloping a shared point of view until we became coworkers several years ago, and she's now my cofounder (and podcast cohost), so that collaboration continues. Alison Ferguson, Jim Sterne, Eric Peterson, and Matty Wishnow all took leaps of faith (twice in Matty's case!) that moved my career and my professional

growth forward—giving me the opportunity to develop and practice many of the ideas included in this book. The on-stage practice of those ideas—at industry conferences and in meetings with clients—went much better than they could have thanks to concepts and techniques I learned from Lea Pica, and her friendship and encouragement to tackle the authorship of this book at all was a valuable boost.

The Analytics Power Hour is a podcast that has given me more opportunities to have deep discussions with smart people—the cohosts and the guests—for a decade (and counting) as of the publication of the book, and I don't think the book would have been possible without those discussions. I'm indebted to Jim "Analysts Don't Do Anything" Cain for being the spark that got that rolling. Michael Helbing—the smoothest voice in explicit analytics podcasts—has been my cohost, my mentor, my manager, and my sounding board and has calmly talked me down from countless temper tantrums in all of those roles, so most importantly, he is my dear friend. Moe Kiss joined the show near the start, and her brilliant perspective, friendship, and willingness to hash out just about any topic on or off the mic has made me a better analyst and a better consultant. Julie Hoyer was a coworker long before she became a cohost, but her ability to nail the weaknesses in any idea, as well as to build on any idea that has some underlying merit, and to do so instantaneously terrifies me just a little bit, but she contributed directly to expanding and solidifying several of the techniques described in the book.

Joe: I'd like to thank my family and friends for providing ample leeway and grace during the work on this manuscript. The late nights, rain checks, and vacation days spent working on this book can add up, and I appreciate their kindness and support during the process. My wife, Sarah, deserves special appreciation for giving me the space to peck away at the manuscript while we gleefully anticipated the arrival of our son.

Some of the themes in this book, which fed into the mind meld Tim and I achieved in producing this book and wouldn't have been developed if not for our partnership, were the culmination of practical ideas from my two-decade journey through software engineering, scientific inference, and executive service. It was a public library in my youth (and its staff and patrons) that I must thank for setting me on a path to fully learn and appreciate the power of engineering, machine learning, and artificial intelligence in an applied business context (it all started with a book on *How to Learn BASIC* when I was a kid). My Washington University in St. Louis, Columbia, Princeton, Johns Hopkins, and Emory colleagues helped me develop entirely new perspectives through which to view data, analytics, and business. The ideas of counterfactual reasoning and evidentiary weight owe homage to my work with, and learning from, Don Green, Andy Gelman, Greg Wawro, Bob Shapiro, and Suresh

Naidu. I credit Jon Rogowski, Andrew Reeves, and Dan Butler for launching me on my academic odyssey, which has proven enjoyable and rewarding.

It is impossible to recount all who in the course of my business inspired and guided the ideas that are reflected in this work; we stand on the shoulders of giants. Evan Schnidman and Bill Macmillan encouraged me to use data and analysis in what was an entirely new way at the time: to forecast Federal Reserve policymaking from unstructured texts. My colleagues and clients at Peachtree AI, Search Discovery, Cisco, and Amazon, lent me the experiences needed to operationally enable these ideas in the business context.

Finally, thank you to Emory University, the Department of Quantitative Theory and Methods, the Center for AI Learning, the Weidenbaum Center, and the Woodruff Library, for enabling me to work on this book.

About the Authors

TIM WILSON

Tim started his career in architecture, pivoted quickly to technical writing, and then found himself working in marketing communications just as the internet was starting to become a thing that businesses took seriously in the early 2000s as a means of finding, acquiring, and keeping customers. In retrospect, he realizes he was in the right place at the right time to have this weird and wonderful thing called "web analytics" land in his lap—a primitive but useful set of technologies for collecting and analyzing the behavior of visitors to the website of the high tech B2B company where he was working at the time. He went on to head up the business intelligence department at that same company before stepping into the agency and consulting world: creating and growing the analytics practices at three different agencies that worked with a range of large consumer brands; consulting with the analytics teams at various Fortune 500 companies on the their strategies, processes, and tactics for effectively putting analytics to actionable use; and then cofounding a consultancy, facts & feelings (factsandfeelings.io), dedicated to helping organizations productively employ research, analytics, and experimentation to drive growth. Tim is a long-time creator of pragmatic content for analysts and marketers, including, in 2008, cofounding the still-running Data and Analytics Wednesday monthly meetup in Columbus, Ohio, and cohosting the biweekly Analytics Power Hour podcast (analyticshour.io) since 2015. He has been a regular conference speaker across multiple continents on a wide range of topics: data visualization, data storytelling, the R programming language, marketing and digital analytics, and, of course, many of the concepts and techniques addressed in this book. While Tim is physically based in Columbus, his heart and soul maintain joint custody with Austin, Texas. He holds a bachelor of science from the Massachusetts Institute of Technology, an MBA from the University of Texas at Austin, and a Certificate of Amazement from Joe that a hillbilly raised in Sour Lake, Texas, can, indeed, learn the fundamentals of causal inference.

DR. JOE SUTHERLAND

With a career spanning the White House, the Dow Jones 30, and America's top universities, Dr. Joe Sutherland has served as executive, public service leader, and educator. Sutherland is the founding director of the Emory Center for AI Learning, where he serves as lead principal investigator of the Emory branch of the US AI Safety Institute Consortium, associate faculty of the Empathetic AI for Health Institute at Emory Healthcare, and fellow of the Weidenbaum Center on the Economy, Public Policy, and Government at Washington University in St. Louis. He is a professor in Emory's Department of Quantitative Theory & Methods. Sutherland's professional experience spans public service in the White House, technology entrepreneurship, executive roles including as chief executive officer of an AI company and at Amazon and Cisco, and academic positions at Columbia, Johns Hopkins, and Princeton. Sutherland founded two startups that were later acquired: Peachtree AI, a professional services firm specializing in artificial intelligence integrations, and Prattle, a fintech company that uses natural language processing to forecast both the Federal Reserve's monetary policy decisions and the performance of publicly traded companies. From 2011 to 2013, he served in the White House Office of Scheduling and Advance for President Barack Obama, traveling with the president in support of various policy initiatives.

His research exploring the utilization of machine learning and AI in a wide variety of topics is published in top peer-reviewed journals, and his work has been featured on *FOX 5 Good Day Atlanta, Atlanta Journal Constitution, Forbes, Georgia Trend, Government Technology, MIT Sloan Management Review*, and many other venues. In 2017, the National Science Foundation recognized his work in state politics and policy with Honorable Mention, considered a national honor. Sutherland earned his PhD, MPhil, and master's degrees in political science from Columbia University and his bachelor's degree in political science from Washington University in St. Louis. He lives in Historic Brookhaven, Atlanta, Georgia with his family, where he enjoys playing golf and tennis.

CHAPTER 1

Is This Book Right for You?

You picked up this book, which means you're thinking that something about the way you and your organization use data and analytics is not "right." Time and again, the executives, managers, and new hires who make up our clients, colleagues, and friends have expressed to us their anxieties related to how they and their teams are using data and analytics:

> "We have plenty of data, but the actionable insights we get from it are few and far between."
>
> "Our team consistently invests in the latest data tools and platforms to ensure we're collecting and storing all the data we might need, but the recommendations we generate from those data never really increase in quality or volume."
>
> "We work with agencies and consultancies that do a lot of reporting on the results they're delivering for us. Those tend to be lengthy presentations with a ton of charts, but I often feel like I'm just having data thrown at me that may or may not be representing real business value being delivered."
>
> "I never feel comfortable investing the millions we invest in paid media; it's unclear if we're actually getting the returns our agencies report, or if they just tortured the data until it confessed a positive answer."
>
> "We have talented analytics and data science teams, but it feels like we're talking past each other when I interact with them. I really need them to generate insights and recommendations, and they seem frustrated when I tell them that that's not what they're providing."
>
> "My data engineers over-promise what their machine learning and AI techniques can do for our stakeholders; it tanks our credibility when we promise magic but don't understand the nuts and bolts well enough to do it right."
>
> "My product teams build these exotic proofs-of-concept using the latest and greatest AI tools. But to scale them up is way too expensive, and the production engineers tasked with doing so can't understand the opaque mathematical techniques being used."
>
> "Our technology platform partners sell us licenses to their latest technology and their latest AI or machine learning, and they share

> eye-popping stories for how effective they are. But when we dig into the pilots, the platforms don't offer anything more than what we're already doing. I wish I could see through these sales pitches earlier."
>
> "We have a ton of automated dashboards, and I understand most of the data that they include, but I still struggle to figure out how I should be using that data to make decisions. Where do I start?"

If any of these quotes feel familiar, then this book is for you. We've heard these frustrations in every data-related function in nearly every industry, ranging from pharmaceuticals to health care, retail, financial services, and consumer packaged goods. And we've worked with clients in all of these industries to shift their approaches. Putting your data to use can be productive, profitable, and even fun! That's why we wrote this book: to guide business leaders who want to *use their data effectively*.

THE DIGITAL AGE = THE DATA AGE

A common theme across all of the frustrations we hear from organizations about their struggles to effectively and consistently extract meaningful business value from their investments in data and analytics is that, well, *there's just so much data*. Our instincts have long been that more data is better, but the shifting of all aspects of our lives from analog to digital over the past three decades has wrought such an extreme version of "more" that it has left many managers questioning those instincts. The origins of the internet are often traced back to the mid-1960s and the creation of ARPANET as a distributed control computer network funded by the US Department of Defense. It was not until 1989, though, that Tim Berners-Lee at CERN conceived of an easier-to-use evolution of what had become "the internet" that would become the "World Wide Web." Within four years, Marc Andreesen, a student at the University of Illinois Urbana-Champaign created the Mosaic web browser while working with the National Center for Supercomputing Applications (NCSA), and the internet was on its way to catching mainstream fire. From the several hundred websites that existed by the end of 1993, to the more than 20,000 in 1995, to 17 million in 2000,[1] the growth of digital content was exponential.

[1] Internet Live Stats. (2009). *Total number of websites—internet live stats* [online]. Internetlivestats.com. Available at: https://www.internetlivestats.com/total-number-of-websites/.

Organizations began transitioning every aspect of their businesses to digital formats. Digital bits and bytes trumped paper on countless fronts: storability (a room full of file cabinets was replaced with a thumb drive), searchability (leafing through those file cabinets pulling out folder after folder and scanning the pages within those folders was replaced by a rectangle on a computer screen into which keywords could be typed), portability (traipsing to the library or the records room or a coworker's office was replaced by launching a browser from any device connected to the internet, and seemingly *every* device is connected to the internet). At a macro scale, global life began going through an analog-to-digital conversion:

- Rather than sending a letter, we could send an email.
- Rather than going to a brick-and-mortar establishment to buy a book, or leafing through a publisher's quarterly catalog, we could search for one online and order it immediately.
- Rather than receiving a book in the mail, we could read it instantaneously in a digital format.
- Rather than advertising on billboards, in magazines and newspapers, or with direct mail, we could advertise on the personalized screens that consumers were spending more and more time looking at, by running ads on websites and search engines.
- Rather than staffing a customer service representative to help prospects find what they need, we could use data science to offer our customers personalized recommendations in real time.

As early as 1994, *BusinessWeek* reported, "Companies are collecting mountains of information about you, crunching it to predict how likely you are to buy a product, and using that knowledge to craft a marketing message precisely calibrated to get you to do so [...] Many companies were too overwhelmed by the sheer quantity of data to do anything useful with the information [...] Still, many companies believe they have no choice but to brave the database-marketing frontier."[2] The digital data revolution was in full swing.

For companies, perhaps the most exciting aspect of this pervasive transformation to a digital-first world was the increased scale and fidelity of the data that could be collected along the way. Ask a retailer how their customers walk through one of their physical stores, and they would have to hire a set of observers to position themselves in the store and take

[2] Berry, J. (1994, Sept. 4). Database marketing. *BusinessWeek*.

copious notes. And they would only have data for the periods when those observers were on site. *And* they would run the risk of affecting their customers' behavior in the process, the so-called "observer effect." Ask a retailer how their customers navigate their *website*, though, and they are just a few clicks away from being able to pull up a report in a digital analytics platform like Google Analytics.

Expectations were high. With *all of this data*, it seemed obvious that *amazing things were possible*! And amazing things *can* be done with data. But over the last 25 years, businesses have slid into what Matt Gershoff, the chief executive officer of Conductrics, refers to as a "big table mentality." They have begun the never-ending and ever-increasing pursuit of gathering "all" the data—striving to clean, store, integrate, and maintain all of the data has become a goal in and of itself. "We can predict, discover, and engineer *anything*, if only we can observe *everything*," the philosophy suggests. "We're going to be truly scientific with all of this data" is the *idea*, but a misunderstanding of scientific principles and their application leads to ineffective and frustrating results rather than the "actionable truths" that we expected.