Chapter 1

Mad Men to Math Men: The Power of the Data-Driven Culture

If we have data, let's look at data. If all we have are opinions, let's go with mine.

—Jim Barksdale, CEO of Netscape

As the television series *Mad Men* depicted, the Madison Avenue executives of the 1960s swirled scotch and smoked cigars from their Eames chairs, stoking their creative powers and developing the memorable advertising campaigns of the era. But very little of that reality remains today.

Modern marketing bears more resemblance to high-frequency stock trading than to *Mad Men*. Marketers sit in front of computers to buy and sell impressions on online advertising exchanges in a matter of milliseconds. Outputs of algorithms determine, in real time, precisely on which web page or mobile app to place an ad, precisely which variation of the ad to serve based on what the software knows about the user, and precisely how much to pay for it based on the probability the viewer will convert to a paid customer.

The paradigm shift from Mad Men to Math Men hasn't happened exclusively on Madison Avenue. This new era of marketing heralds analogous transformations in sales, human resources, and product management. No matter the role, no matter the sector, data is transforming it.

Modern sales teams employ predictive scoring technologies that crawl the web to aggregate data about potential customers and calculate the likelihood a customer will close. Each morning, sales account

executives log into their customer relationship management software to a list of leads prioritized by likelihood to close. These are the new leads. The Glengarry leads.

Recruiters use data to identify the best candidates to pursue based on online profiles, blogs, social media accounts, and open-source software contributions. Product managers record the actions of users by the millisecond to understand exactly which customer journeys optimize revenue and where in the product customers exhibit confusion or drop off. Data courses through these teams by the gigabyte and supplies the essential foundation for decision making throughout the organization.

As novelist William Gibson said, "The future is already here—it's just not very evenly distributed."[1] A small number of companies have restructured themselves, their hiring practices, their internal processes, their data systems, and their cultures to seize the opportunity provided by data. And they are winning because of it. They exemplify the future. Inevitably, these techniques will diffuse through industry until everyone remaining employs them.

With this book, we'll illuminate how forward-thinking businesses already operate in the future, and outline how we have seen others evolve their businesses, their technology, and their cultures to win with data.

Operationalizing Data: Uber's Competitive Weapon

> *Who among us does not say that data is the lifeblood of their company? The largest hoteling company [AirBnB] owns no hotel rooms. The largest taxi company [Uber] owns no taxis.*
>
> —Ash Ashutosh, CEO of Actifio

At their core, the best data-driven companies operationalize data. Instead of regarding data as a retrospective report card of a team's performance, data informs the actions of each employee

[1]William Gibson, "The Science in Science Fiction," *Talk of the Nation*, NPR, November 30, 1999.

every morning and every evening. From harnessing customer survey responses to evaluating loan applications, these Math Men and Women are transforming every industry and every function.

As Ash Ashutosh said, the biggest transportation and lodging companies own no infrastructure. Instead, they manage data better than anyone else. Just four years after Uber was founded, its San Francisco revenues totaled more than three times all the revenues of all the taxi cab companies in the city. Two years later, the Yellow Cab Cooperative, which has operated the largest fleet of taxis in San Francisco for decades, filed for bankruptcy.

Among many innovations, Uber brought data to the taxi industry. Using historical data, Uber advises drivers to be in certain hotspots during certain times of day to maximize their revenue because customers tell them with the push of a button where to be. Uber matches the closest driver with the customer to minimize wait time and maximize driver utilization and earnings.

In contrast, disconnected Yellow Cab drivers listen to a coffee-fueled, fast-talking dispatcher relaying telephone call requests by radio. Individual drivers claim passenger pickups by responding over the CB, even if they are the furthest cab from the customer. "How long until the taxi arrives?"

Dispatchers can handle only one request at a time, serially. In rush hour, potential passengers redial after hearing a busy tone. Let too much time elapse coming from the other side of town and your passenger has already jumped into an Uber. For the Yellow Cab driver, the gas, time, and effort are all wasted because of an information asymmetry. In comparison to Uber, Yellow Cab drivers are driving blind to the demand of the city, and Yellow Cab customers are blind to the supply of taxi cabs.

Uber changes its pricing as a function of demand, telling drivers when it makes sense to start and stop working. Surge pricing, though controversial, establishes a true market for taxi services. Yellow Cab drivers don't know the best hours to work and prices are fixed regardless of demand.

Data improves more than the marketplace efficiency. Uber employs drivers based on their customer satisfaction data provided by consumers. Drivers who score below a 4.4 on a 5.0 scale risk "deactivation"—inability to access Uber's passenger base. Meanwhile,

the Yellow Cab company maintains an average Yelp review of less than 1.5 stars out of 5.

The data teams that optimize Uber driver locations, maximize revenue for drivers, and drive customer satisfaction operate on a different plane from the management of the Yellow Cab company. Blind, Yellow Cab drivers are completely outgunned in the competitive transportation market. They don't have what it takes to compete: data.

But the Uber phenomenon isn't just a revolution in the back office. It's also about a new generation of taxi drivers, who operate their own businesses in a radically different way. What cabbie in the 1990s could have dreamed that upon waking early in the morning, a mobile phone would suggest there's more money to be made in the financial district of San Francisco than at the airport? But the millennial driver knows the data is attainable: It's just a search query or text message away. This is the fundamental, secular discontinuity that data engenders.

The Era of Instant Data: You Better Get Yourself Together

Instant Karma's gonna get you
Gonna knock you right on the head
You better get yourself together
Pretty soon you're gonna be dead

—John Lennon

The demand for instant data will increase inexorably. Like Uber drivers seeking a passenger at this very moment, we expect answers instantly. If you're making Baked Alaska for company tonight, and you've forgotten the ratio of sugar to egg whites in the meringue that houses the ice cream, your phone will answer the question in just a few seconds.

Where is Priceline stock trading? Where do the San Francisco Giants stand in this year's pennant race? When hiring a litigation attorney, what are the key questions to ask? Are there any grammatically sound sentences in English where every word starts with the same letter?

All of these questions are instantly answerable. These are the types of questions we ask at the dinner table or when sharing a drink with a friend at a bar, and answer in a few seconds with a search query on a phone.

Because of this new instant access to just about every kind of information, we expect the same instantaneity of answers at work. Why did our sales team outperform last quarter? Which of my clients are paying the most? Does this marketing campaign acquire customers more efficiently than the others? Should we launch our product in Japan in December?

In most companies, these questions require days or weeks to answer. Consequently, data is a historical tool, a useful rearview mirror to the well-managed business. It's a lens through which we can understand what happened in the past. And, if we're lucky, it can help us understand a little bit about why the past unfolded in a particular way.

But this level of analysis pales in comparison with the practices of best-in-class companies that operationalize their data. These are businesses that use the morning's purchasing data to inform which merchandise sits on the shelves in the afternoon.

What have those companies done to access instant data? First, they've changed the way they manage themselves, their teams, and their companies; they've changed how they run meetings, how they make decisions, and how they collaborate. Employees are data literate: They understand how to access the data they need, how to analyze it, and how to communicate it well.

Second, these companies have developed functional data supply chains that send insight to the people who need it. A data supply chain comprises all the people, software, and processes related to data as it's generated, stored, and accessed. While most of us think of data as the figures in an Excel spreadsheet or a beautiful bar chart, these simple formats often hide the complexity required to produce them.

The simple Excel spreadsheet hides a churning sea of data, coursing through the company's databases, that must be synthesized and harmonized to create a single, accurate view of the truth. A data infrastructure that permits easy, instant access to answers to business questions by anyone in the company is the second step.

Third, these businesses create a data dictionary, a common language of metrics used by the company. When sales and marketing refer to a lead, the definition of a lead must be consistent across both teams. Often, different teams within a company define metrics in unique ways. Though convenient for the individual team, this approach creates confusion, inconsistency, and consternation. Robust data pipelines ensure a universal language across the company.

This combination of bottoms-up data literacy, top-down data infrastructure, and a single metrics lexicon has transformed many businesses. Google was one of the first to empower its employees with unfettered access to critical business data. Consequently, Google employees were able to leverage the company's enormous reach and resources to develop breakthrough products.

That innovation in the early 2000s cascaded through many other large and small companies, including Facebook, LinkedIn, Zendesk, and others. Above all, these companies architected data supply chains that enable their employees to extract the insights they needed to advance the company's causes. Unfortunately, most businesses still operate with outdated supply chains buckling under the strain of data demand. You better get your data together, or pretty soon you're gonna be dead.

Data Supply Chains: Buckling Under the Load

Slow data is caused by an inefficient supply chain. Today's data supply chains suffer from a fundamental flaw in their architecture: The number of people seeking data dwarfs the number of people supplying data. The taxi dispatcher relaying passenger pickups by phone serves scores of drivers, each seeking their next fare. In many companies, this ratio may be much greater than 100:1. Is it any surprise that the data analyst team is seen as an enormous bottleneck, a chokepoint for the organization?

In the past, this flawed architecture functioned because most companies had a relatively small amount of data, most of it created by humans, and the competition wasn't using data for a competitive advantage. Without a substantial corpus of data to interrogate, only a handful of executives asked questions of their company's data, limiting the total number of requests. Most of the time, these

requests were financial in nature and managed by the CFO and his organization.

But the amount of data that companies store today has exploded. According to IDC, from 2013 to 2020, the digital universe will grow by a factor of 10, from 4.4 trillion to 44 trillion gigabytes. It more than doubles every two years. This supernova of data contains insights relevant for every person within an organization.

Today, computers generate data at rates that far outstrip humans. Facebook records more than 600 petabytes of data daily on its users, almost all of it generated by computers. This trend isn't constrained to social networks. For example, Marketo, Eloqua, Pardot, and Hubspot pioneered the marketing automation software category not more than 10 years ago. These tools help B2B (business-to-business) marketers optimize demand-generation programs and prioritize leads. Market automation software snares data on website visitors to answer typical marketing questions: What content are they reading? How frequently are they visiting the site? What are the best messages to generate more leads?

Now that we're collecting all this data, we expect instant answers from it. In larger companies, the burden for answers rests on the shoulders of the data team. These scarce data analysts must process an ever-lengthening queue of work. Each request carries with it a unique set of intricacies. Perhaps the query involves a new data set, or a new type of data analysis, or a new visualization. And maybe the requesters of the data weren't quite sure what they were asking at the outset, so they revise the requested analysis, adding again to the workload and slowing processing time for everyone else. Rarely do most employees understand the complexity of their requests: the number of steps, the turnaround time, or the number of players required to answer their questions.

For smaller companies, there's often a cadre of people, perhaps just one or two, who understand how to pull data from databases, something they do in their spare time or after hours as favors to colleagues. Quickly, the volume and sophistication of these requests overwhelms the moonlighters, who can't possibly support the demand.

To extract value from these mountains of ones and zeros, companies can no longer rely on a small coterie of radio dispatchers

broadcasting outdated information to their employees. Nor can they simply supply stronger coffee so dispatchers speak faster. The external competitive pressures and internal demand for accurate and relevant data are too great.

In both cases, the end result is the same. The data infrastructure simply cannot satisfy the demands of data consumers within the organization. Starved for insight, employees substitute instinct, gut, back-of-the-envelope calculations, estimates, and other short-circuited research to decide.

The data dispatcher system won't scale to meet these new needs. More dispatchers, more radios, more caffeine won't solve the problem. A new data supply chain must be built.

Management by Opinion: The Illusion of Knowledge

> *The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge.*
>
> —Stephen Hawking

Like Yellow Cab, most companies manage with a few dispatchers. The constraints of the data supply chain engender inequality. Limited bandwidth forces the company to prioritize only the most important data; for most businesses, that means data requested by the C-suite. Everyone else must resign themselves to deciding using opinion, gut, and conjecture.

This is the worst outcome of all. Paraphrasing Stephen Hawking, the greatest enemy of business progress is the illusion of knowledge. If we make decisions based on guesses or opinions or word of mouth, we are all just wasting time, like a taxi driver wandering the city streets in search of the next fare. *Perhaps I should drive to the financial district. I've had luck there before. My friends tell me the airport should be busy this weekend.*

How should a team dispel the illusion of knowledge? Dominic Orr, former CEO of Aruba Networks, insists on brutal intellectual honesty in his management teams. "We focus on collecting as many facts as quickly as we can, and then we decide on the best, but not necessarily the perfect, solution. Think Socratic method at the

speed of light."[2] Insisting on brutal intellectual honesty within a company or a team demands great trust between team members and ensures the best decision is consistently made, for the right reasons. Not politics, not gut, nothing but informed decisions.

Imagine a world where data is put into the hands of the people who need it, when they need it, not just for Uber drivers, but for every team in every company. This is data democratization, the beautiful vision of supplying employees with self-service access to the insights they need to maximize their effectiveness. This is the world of the most innovative companies today: technology companies like Uber, Google, Facebook, and many others who have re-architected their data supply chains to empower their people to move quickly and intelligently.

Modern data infrastructure is necessary but insufficient for a company to become data-driven. A cutting-edge data supply chain that's unused is just as worthless as a nonexistent one. Culture is the key ingredient to ensuring data investments achieve their potential.

The core values of a company define its culture. The disruptive companies described in this book prize curiosity, collaboration, and a desire to use data for decisions. Data must be part of every important discussion and decision.

These values start when hiring. At Google, the recruiting teams evaluate candidates on several attributes, most notably Googliness, an eponymous characteristic of the company. Googliness refers to many things, most notably intellectual curiosity: the desire to ask questions and understand why. Google hires only candidates who exhibit googliness.

To support these curious minds, Facebook and Zendesk, like Google, employ data teams. These data teams architect data systems, educate employees to use them, tutor teams on correct analytical methods, and assist individuals when crafting arguments using data. In addition, these data teams collaborate across all the departments of these businesses to design, maintain, and circulate a data dictionary, a common lexicon of metrics used across the business.

---

[2]Cathy Olofson, "So Many Decisions, So Little Time," *Fast Company*, September 30, 1999.

By inculcating a common set of values, offering the tools and education, and creating a common language, data teams within these businesses empower their colleagues to decide how to advance the company using data, instead of opinions.

At their core, data teams disperse the fog of ignorance within a company. They democratize data access and disseminate knowledge across a business. And the business evolves from responding to a dispatcher's radio comments to deciding with a real-time, bird's-eye view of all the customers in a city seeking a taxi.

Our Vantage Points

Once one company within a sector begins to win with data, as Uber has, the only competitive response from its peers will be the development, deployment, and use of data at scale. We have no doubt that this approach to data will cascade into every position in every business in every industry, because we have seen it firsthand.

Tom Tunguz, Partner at Redpoint

I first learned about the value of data sets at Google. I started in the AdSense Operations team, which managed the accounts of large web publishers who ran Google's ads on their web pages. About a year later, I transferred into the product management team at Google and began to work with teams of marketers, engineers, and user-experience researchers to build new products. Over the next 24 months, we built products to monetize some of the largest social networks in the world by ingesting anonymized data about users to improve our ad targeting.

We also localized AdSense into many new languages. Statistics played a key role in interpreting other languages. In English, sentences contain spaces between words; not so in Chinese. Further complicating things for computers trying to understand Chinese text, the Chinese language uses compound words frequently. Cell phone is *shǒujī* (手机), which are the characters for "hand" and "machine." A lobster is *lóngxiā* (龙虾), or "dragon shrimp." And a turkey is *huǒ jī* (火鸡), meaning "fire chicken." Engineers used complex

statistics to infer the meaning of the author and to target ads better. A misinterpretation of the content could lead to hilarious results.

After Google, I joined Redpoint, a venture capital firm with a long history of investing in breakthrough companies like Netflix, Sonos, Stripe, and Zendesk, among many others, at the very earliest stages. At Redpoint, we've invested heavily in using data to help us find great companies, even if they might be just a few people typing away in an apartment in San Francisco.

Startups today leave footprints all over the Internet. Two cofounders will meet on LinkedIn and begin to chat with each other on Twitter. They will post job listings on hiring boards all over the web. They might launch an application on the Apple App Store. We continue to build data tools to pick up those bread crumbs, each a clue about what might be the next billion-dollar business.

In addition, we benchmark companies constantly, comparing growth rates, marketing efficiency, word-of-mouth vitality, and many other metrics. This rich database informs our investment decisions. We also use this data to provide targets for our portfolio companies, the businesses we invest in.

Last, we have developed a metrics-driven content marketing strategy to build our brand with hundreds of thousands of entrepreneurs all over the world. Ten years ago, the world of venture capital could have been called a cottage business: friends in different firms trading deals over a fancy lunch. Today, venture capital partnerships invest heavily in their data infrastructures to gain a small edge, the iota of information asymmetry that might lead to the next multi-billion-dollar giant.

Our experience building internal data tools and engendering a data-centric culture at Redpoint helps us invest in companies building next-generation data technologies. In 2012, we were lucky to meet Frank and the team, and we were amazed by the Looker product, especially when they connected a Looker instance to our internal data sets and we could analyze the trends like never before. A few days later, we shook hands on a partnership. Since then, Looker has become the fastest-growing business intelligence (BI) company of the past 20 years.

Frank Bien, CEO of Looker

For 20 years, I have worked in the world of databases, including Greenplum, a maker of high-performance analytics databases acquired by EMC, Dell, and Intraspect Software.

When the whole idea of Big Data emerged, around 2002 and 2003, I was excited to see what would happen. Big Data gave us a new infrastructure. It gave us systems that could store everything. And it engendered data-mining pursuits, predictive queries about how much customers might buy or how deep a discount would generate more revenue.

But I was painfully aware that nobody had cracked the code on how to build on top of big data in such a way that it was usable by business people. They were still building BI tools for different kinds of databases. Companies were installing these giant machines and collecting massive amounts of data, and then doing trivia questions that had no business value. I knew that if people had a new kind of tool to see into all their data, they could change how their business operated, and everybody could be like a Google. Even a small company could make better-informed decisions, driven by what they understood to be true.

When I met Lloyd Tabb, the founder of Looker, I was impressed with the company's customers. I could see that businesses, *some* businesses, were ready to do data in that new way. The early Looker customers were the most innovative of the innovative, San Francisco startups and Silicon Valley–funded companies. CEOs, data teams, and everyone in between were deeply interested in data. They didn't want the old BI tools. They didn't want to work in PowerPoint. They didn't want the pictures. They wanted to get inside.

That's when, and why, I joined Looker. With Looker, the relationship these groundbreaking companies had with data was fundamentally different, just like the web browser fundamentally changed our relationship to information. Together, they're creating a "Give me the proof" kind of business culture that is driving the success of their businesses. Nerds have become the new mainstream. They don't want the toys. They're ready for a real toolbox. That shift has allowed Looker to succeed.

The results speak for themselves: From business users in marketing to PhD-level data scientists, Looker users get hooked on data.

We started out as a technical product aimed at data people who wanted code, fast and agile. But when we empowered the data people to be creative, curating data for the rest of the company, rather than answering one-off questions, something else happened. We noticed that business users started asking more questions. When they finally had access to everything, not just to tidbits, they went crazy. One question would lead to another. They started asking a hundred questions, and they started using data to evaluate, explain, and defend their decisions. What they learn has made them, and their businesses, smarter.

In this book, we hope to share what we both have learned within our own companies and within exceptional startups and monoliths about how to transform a company with data.