O'REILLY

Analytical Skills for AI & Data Science

Building Skills for an AI-driven Enterprise

Early Release
RAW & UNEDITED

Daniel Vaughan

Analytical Skills for AI and Data Science

Daniel Vaughan

Analytical Skills for AI and Data Science

by Daniel Vaughan

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://oreilly.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Acquisitions Editor: Jonathan Hassell

Developmental Editor: Michele Cronin

Production Editor: Daniel Elfanbaum

Interior Designer: David Futato

Cover Designer: Karen Montgomery

Illustrator: Rebecca Demarest

July 2020: First Edition

Revision History for the Early Release

- 2020-02-04: First Early Release
- 2020-03-20: Second Early Release

See http://oreilly.com/catalog/errata.csp?isbn=9781492060949 for release details.

978-1-492-06087-1

Preface

> **A NOTE FOR EARLY RELEASE READERS**
>
> With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.
>
> This will be the preface of the final book. Please note that the GitHub repo will be made active later on.
>
> If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the author at analyticalthinkingbook@gmail.com.

Why analytical skills for AI

Judging from the headlines and commentary in social media during the second half of the 2010s, the age of artificial intelligence has finally arrived with its promises of automation and value creation. Not too long ago, a similar promise came with the big data revolution that started around 2005. And while it is true that some selected companies have been able to disrupt industries through AI- and data-driven business models, many have yet to realize the promises.

There are several explanations for this lack of measurable results — all with some validity, surely — , but the one put forward in this book is the general lack of analytical skills that are *complementary* to these new technologies.

The central premise of the book is that value at the enterprise is created by *making decisions*, not with data or predictive technologies alone. Nonetheless, we can piggyback on the big data and AI revolutions and start making better choices in a systematic and scalable way, by transforming our companies into modern AI- and data-driven decision-making enterprises.

To make better decisions we need first to ask the right questions, forcing us to move from descriptive and predictive analyses to *prescriptive* courses of action. I will devote the first few chapters on clarifying these concepts and learning how to ask better business questions suitable for this type of analysis. I will then delve into the anatomy of decision-making, starting with the consequences or outcomes we want to achieve, moving backwards to the actions we can make, and discussing the problems and opportunities created by intervening uncertainty and causality. Finally we will learn how to pose and solve prescriptive problems.

Use-case-driven approach

Since my aim is to help practitioners to create value from AI and data science using this analytical skillset, in each chapter I will show how each skill works with the help of a collection of use cases. I selected them from my own experience, because many companies face them and are thus advertised by consulting companies without providing alternative solutions, because students found them interesting or because they are building blocks for more complex problems that are found in the industry. But in the end this choice was subjective and depending on your industry they may be more or less relevant.

What this book isn’t

This book isn’t about artificial intelligence or machine learning. This book is about the extra skills needed to be successful at creating value from these predictive technologies.

I have provided an introduction to machine learning in the Appendix for the purpose of being self-contained, but it isn’t a detailed presentation of machine learning related material nor was it planned as one. For that you can check many of the great books out there (some mentioned in the Suggested Readings section of the Appendix).

Who is this book for

This book is for anyone wanting to create value from machine learning. I've used parts of the material with business students, data scientists and business people alike.

The most advanced material deals with decision-making under uncertainty and optimization, so having a background on probability, statistics or calculus should help. For readers without this background I've tried to make the presentation self-contained. On a first pass, you may just skip the technical details and focus on developing an intuition and an understanding of the main messages for each chapter.

- If you're a *business person* with no interest whatsoever in doing machine learning yourself, this book should at least help redirect the questions you want your data scientists to answer. Business people have great ideas but have difficulties expressing what they want to more technical types. If you want to start using AI in your own line of work, this book will help you formulate and translate the questions so that others can work on the solution. My hope is that it will also serve as inspiration to solve new problems you didn't think were attainable.
- If you're a *data scientist*, this book will provide a holistic view on how you can approach your stakeholders and generate ideas to apply your technical knowledge. In my experience, data scientists become really good at solving predictive problems, but many times have difficulties delivering prescriptive courses of action. The result is that your work doesn't create as much value as you want and expect. If you've felt frustrated because your stakeholders don't understand the relevance of machine learning, this

book could help you transform the question you're solving to take it "closer to the business".

- If you're a *developer interested in data science* this book will take you closer to the business and provide an understanding of how data science creates value. You may already have other more technical readings on your path to deep learning and the like, so this may feel just right when you want to read something more "businessy" without completely losing the more formal and technical foundations. It should also serve as a North Star to remind you that it's not about technical knowledge but about value creation.

What's needed

I wrote this book in a style that is supposed to be readable for very different audiences. I do *not* expect the reader to have any prior knowledge of probability or statistics, machine learning, economics or the theory of decision making.

Readers with such backgrounds will find the more technical material introductory, and that's actually great. In my opinion, the key to creating value through these techniques is to not focus on the technical side but on the business. I hope that by focusing on the use cases they can find many new ways to solve the problems they're facing.

For readers with no background in these topics I've tried to provide a very minimal introduction to the key themes that I need to develop each of the use cases. If you're interested in going deeper I've also

provided a list of references that I’ve found useful, but I’m sure you can find many more on the internet. If you’re not interested in going deeper, that’s fine too. My advice is to focus on the broader picture and intuition. That way you’ll be able to ask the right questions to the right people at your companies.

What’s really needed to get the most value from this book is curiosity. And if you’ve reached this paragraph most likely you’re good on this.

Conventions Used in This Book

The following typographical conventions are used in this book:

Italic

Indicates new terms, URLs, email addresses, filenames, and file extensions.

`Constant width`

Used for program listings, as well as within paragraphs to refer to program elements such as variable or function names, databases, data types, environment variables, statements, and keywords.

`Constant width bold`

Shows commands or other text that should be typed literally by the user.

`Constant width italic`

Shows text that should be replaced with user-supplied values or by values determined by context.

TIP

This element signifies a tip or suggestion.

NOTE

This element signifies a general note.

WARNING

This element indicates a warning or caution.

Using Code Examples

Supplemental material (code examples, exercises, etc.) is available for download at *https://github.com/oreillymedia/title_title*.

This book is here to help you get your job done. In general, if example code is offered with this book, you may use it in your programs and documentation. You do not need to contact us for permission unless you're reproducing a significant portion of the code. For example, writing a program that uses several chunks of code from this book does not require permission. Selling or distributing a CD-ROM of examples from O'Reilly books does require permission. Answering a question by citing this book and quoting example code does not require permission. Incorporating a

significant amount of example code from this book into your product's documentation does require permission.

We appreciate, but do not require, attribution. An attribution usually includes the title, author, publisher, and ISBN. For example: "*Book Title* by Some Author (O'Reilly). Copyright 2012 Some Copyright Holder, 978-0-596-xxxx-x."

If you feel your use of code examples falls outside fair use or the permission given above, feel free to contact us at *permissions@oreilly.com*.

O'Reilly Online Learning

> **NOTE**
>
> For almost 40 years, *O'Reilly Media* has provided technology and business training, knowledge, and insight to help companies succeed.

Our unique network of experts and innovators share their knowledge and expertise through books, articles, conferences, and our online learning platform. O'Reilly's online learning platform gives you on-demand access to live training courses, in-depth learning paths, interactive coding environments, and a vast collection of text and video from O'Reilly and 200+ other publishers. For more information, please visit *http://oreilly.com*.

How to Contact Us

Please address comments and questions concerning this book to the publisher:

O'Reilly Media, Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

800-998-9938 (in the United States or Canada)

707-829-0515 (international or local)

707-829-0104 (fax)

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at *http://oreilly.com/catalog/errata.csp?isbn=9781492060949*.

To comment or ask technical questions about this book, send email to *bookquestions@oreilly.com*.

For more information about our books, courses, conferences, and news, see our website at *http://www.oreilly.com*.

Find us on Facebook: *http://facebook.com/oreilly*

Follow us on Twitter: *http://twitter.com/oreillymedia*

Watch us on YouTube: *http://www.youtube.com/oreillymedia*

Acknowledgments

This book had three sources of inspiration. First, it has been the backbone in a Big Data for Managers course at the Tecnologico de Monterrey, in Mexico City. As such I’m grateful to the university and the EGADE Business School specifically; they have provided a great place to think, discuss and lecture on these ideas. Each cohort of students helped improve on the material, presentation and use cases. To them I’m infinitely grateful.

My second source of inspiration came from my work as Head of Data Science at Telefonica Movistar Mexico and the wonderful team of data scientists that were there during my tenure. They helped create a highly energetic atmosphere where we could think out of the box and propose new projects to our business stakeholders.

I’m finally indebted to the different business people that I’ve encountered during my career, and especially during my tenure at Telefonica Movistar Mexico. It was never easy to sell these ideas, and the constant challenge helped improve my understanding of how they view the business, forcing me to build bridges between these two seemingly unrelated worlds.

I’m grateful to my family and friends for their support from the beginning. Finally, I’m infinitely grateful to my dogs Matilda and Domingo. They were the perfect companions to the many long hours working on the book, always willing to cheer me up. We’ll finally have more time to go to the park now.

Last but not least, I’m deeply grateful to my editor, Michele Cronin. Her suggestions dramatically helped improve the presentation of the book. Any mistakes that remain are my own, of course.

Chapter 1. Analytical thinking and the AI-driven enterprise

A NOTE FOR EARLY RELEASE READERS

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

This will be the 1st chapter of the final book. Please note that the GitHub repo will be made active later on.

If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the author at analyticalthinkingbook@gmail.com.

It is March 2020 and the world is in the middle of a very serious global pandemic caused by Covid-19 with confirmed cases in the hundreds of thousands and deaths in the thousands. If you'd searched online for `AI coronavirus` you could've found some very prestigious media and academic outlets highlighting the role that artificial intelligence (AI) can play in the battle against the pandemic (Figure 1-1).

Google

coronavirus AI

All | Images | News | Videos | Maps | More | Settings | Tools

About 1,330,000,000 results (0.44 seconds)

Top stories

Coronavirus: How can AI help fight the pandemic?

BBC · 9 hours ago

Inside the company that used AI to create a coronavirus test

CNN.com · 14 hours ago

→ More for coronavirus AI

www.bbc.com › news › technology-51851292 ▾

Coronavirus: How can AI help fight the pandemic? - BBC News

9 hours ago - A World Health Organization report last month said **AI** and big data were a key part of China's response to the virus. Sharing data. Data on ...

www.technologyreview.com › how-baidu-is-bringing-ai-to-the-fight-... ▾

How Baidu is bringing AI to the fight against coronavirus - MIT ...

1 day ago - Scientific and medical communities worldwide are using **AI** to understand and contain Covid-19, treat infected patients, and ultimately develop ...

towardsdatascience.com › how-to-fight-the-coronavirus-with-ai-and-d... ▾

How to Fight the Coronavirus with AI and Data Science

How to Fight the **Coronavirus** with **AI** and Data Science. WHO, BlueDot Global, and Harvard Medical School leading the way to disease prevention.

Figure 1-1. AI and the coronavirus

What makes me uncomfortable with these headlines is that they dress AI with a superhero suit that has become rather common, overstretching the limits of what can be achieved with AI today.

What is AI

If I had to divide the world according to their understanding of the term I'd say there are four types of people.

On one end of the spectrum are those who've never heard of the term. Since AI has become part of the popular folklore and is now a common theme in movies, TV shows, books, magazines, talk shows and the like, I guess that this group is rather small.

Most people belong to a second group that believes that AI is closer to what practitioners call *Artificial General Intelligence* (AGI) or human-like intelligence. In their view AI are humanoid-like machines that are able to complete the same tasks and make decisions like humans. For them AI is no longer in the realm of science fiction as almost every day they find some type of media coverage on how AI is changing our lives.

A third group, the practitioners, actually dislike the term and prefer to use the less sexy machine learning (ML) label to describe what they do. ML is mainly concerned about making accurate predictions with the use of powerful algorithms and vast amounts of data. There are many such algorithms, but the darling of ML techniques is known as

deep learning (short for deep neural networks) and is pretty much responsible for all the media attention the field gets nowadays.

To be sure, deep learning are also predictive algorithms that have proven quite powerful in tackling problems that a few years ago were only accessible to humans, specifically in the domains of image recognition and natural language processing (think Facebook automatically labelling your friends in a photo or virtual assistants like Alexa smoothing out your purchases on Amazon and turning on and off your lights or any other device connected to the internet at home).

I don't want to distract your attention with technical details so if you want to learn more about these topics please consult the Appendix. The only thing I want to highlight here is that practitioners think "ML" when they hear or read "AI", and in their minds this really just means *prediction algorithms*.

The fourth and final group is what I'll call "the experts", those very few individuals that are doing research, and thus, advancing the field of AI. These days most funds are directed towards advancing the field of Deep Learning, but in some cases they are doing significant research on other topics that aim at achieving AGI.

In this book I'll use AI and ML interchangeably since it has become the standard in the industry, but keep in mind that there are other topics different from prediction that are part of the AI research arena.

Difficulties with current AI

The trouble with AI starts with the name itself as it inevitably makes us think about machines with human-like intelligence. But the difficulty comes not only from a misnomer but also from comments coming from within, as some recognized leaders in the field have reinforced expectations that will be hard to accomplish in the short term. One such leader claimed in 2016 that "(p)retty much anything that a normal person can do in <1 sec, we can now automate with AI".[1] Others may be more cautious, but their firm conviction that deep neural networks are fundamental building blocks for achieving AGI provides the media with juicy headlines.

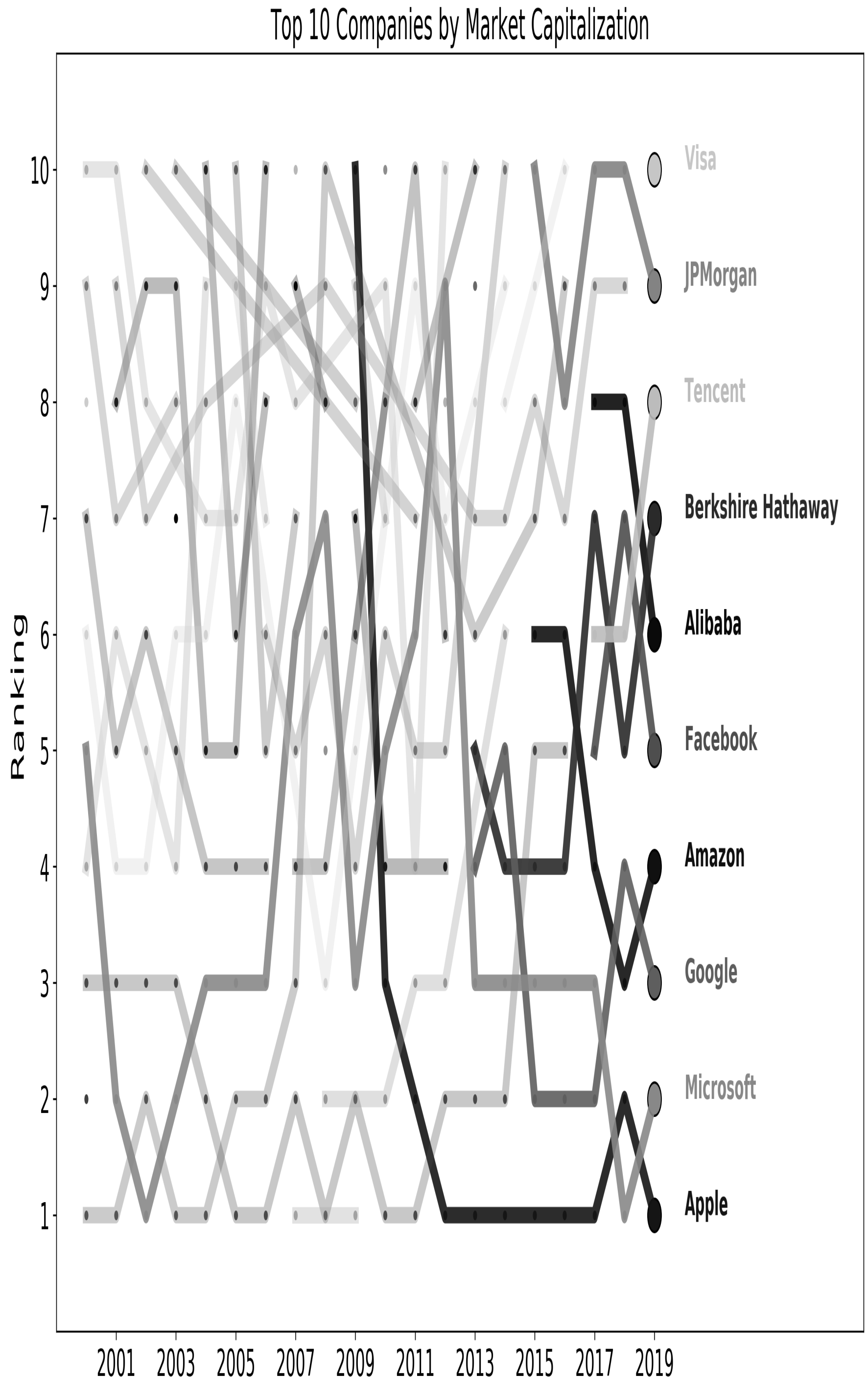
But I digress: what really matters for the purpose of this book is how this hype has affected the way we run our businesses. It is not uncommon to hear Chief Executive Officers and other high-ranking executives say that they are disrupting their industries with AI. While they may not be fully aware of what the term entails, they are nonetheless backed by vendors and consultants that are very happy to share the riches before the bubble pops.

Hypes are risky because a natural response to unfulfilled expectations is to cut all funds and organizational focus.[2] The aim of this book is to show that while we may be far from creating human-like intelligence, we can still generate substantial value for our businesses by using AI as an input to make better decisions.

Before that let's understand how we got here, as this will help showcase some of the difficulties in the current approach and the opportunities that are already achievable.

How did we get here

Figure 1-2 shows the evolution of the top 10 global companies by market capitalization. With the probable exception of Berkshire Hathaway — Warren Buffett's conglomerate — , Visa and JP Morgan, all of the remaining companies are in the technology sector and all have embraced the data and AI revolutions.[3] At face value this would suggest that if this worked for them it must work for any other company. But is this the case?



Top 10 Companies by Market Capitalization
Ranking
10
9
8
7
6
5
4
3
2
1
2001
2003
2005
2007
2009
2011
2013
2015
2017
2019
Visa
JPMorgan
Tencent
Berkshire Hathaway
Alibaba
Facebook
Amazon
Google
Microsoft
Apple

Figure 1-2. Evolution of market capitalization top-10 ranking. Companies that left the ranking before 2018 are not labeled.

Behind these successes there are two stories that only converged until recently. One has to do with the evolution of AI and the other with the Big Data revolution.

The Data Revolution

Not so long ago barely anyone talked about AI. On the contrary, as The Economist claimed in 2017, big data was the new oil.[4]

The year is 2004 and Google published their famous MapReduce paper that enabled companies to distribute computation of large chunks of data (that wouldn't fit in a single computer) across different machines.[5] Later, Yahoo! made and open sourced their own version of the Google algorithm, marking the beginning of the data revolution.

It took a couple of years for technology commentators and consulting firms to start claiming that data would provide companies with endless opportunities for value creation. At the beginning this revolution was built around one pillar: having more, diverse and fastly-accessible data. As the hype matured two more pillars we are added: predictive algorithms and a data-driven culture.

THE 3 V'S

The first pillar involved the now well-known three Vs: *volume*, *variety* and *velocity*. The internet transformation had provided companies with ever-increasing volumes of data. One 2018 estimate

claims that in the previous two years, 90% of the data created in the history of human kind had been generated, and many such calculations abound.[6] Technology had to adapt if we wanted to analyze this apparently unlimited supply of information: we not only had to store and process larger amounts of data, but also needed to deal with new unstructured types of data such as text, images, videos and recordings that were not easily stored or processed with the data infrastructure available at the time.

STRUCTURED AND UNSTRUCTURED DATA

The second V, *variety*, emphasizes the importance to analyze all type of data, not only structured data. If you have never heard of this distinction think of your favorite worksheet (excel, google sheets, etc.). These organize the information in tabular arrangements of rows and columns, that provide a lot of structure so that we can efficiently process information within a friendly user interface. This a simple example of structured data: anything you can store and *analyze* using rows and columns belongs to this class.

Have you ever copied and pasted an image in Excel? If you had then you know it can be done, as well as pasting entire texts, images and videos. But the fact you can paste them doesn't mean you can *analyze*. And storage isn't efficient either: you can save a lot of space on disk by using some type of compression or efficient formats. *Unstructured* datasets are not efficiently stored or analyzed using tabular formats, and these include all type of multimedia (images, videos, tweets, etc.). Now, these provide *a lot* of valuable information for companies, so why should we not use them?

After the innovations were made, consultants and vendors came up with new ways to market these new technologies. Before the age of big data, the Enterprise Data Warehouse was used to store and analyze structured data. The new age needed something equally new

and thus the *Data Lake* was born with the promise of providing flexibility and computational power to store and analyze big data.

Flexibility came in two flavors: thanks to “linear scalability”, if twice the work needed to be done, we would just have to install twice the computing power to meet the same deadlines. Similarly, for a given task, we could cut the current time by half by doubling the amount of infrastructure. Computing power could be easily added by way of commodity hardware, efficiently operated by open-source software readily available for us to use. But the data lake also allowed for quick access to the larger variety of data sources.

Once we tackled the volume and variety problems, velocity was the next frontier, and our objective had to be the reduction of time-to-action and time-to-decision. We were now able to store and process large amounts of very diverse data in real-time or near real-time if necessary. The three Vs were readily achievable for any company willing to invest in the technology and the know-how. Nonetheless, the riches were not at sight yet so two new pillars were added — prediction and data-driven culture — along with a recipe for success.

DATA MATURITY MODELS

Since data alone was not creating the value that was promised we needed some extra guidance; this is where maturity models entered with the promise of helping companies navigate through the turbulent waters created by the revolution. One such model is depicted in Figure 1-3, which I will explain now.

Hierarchy of Value Creation

Figure 1-3. A possible data maturity model showing a hierarchy of value creation

Descriptive stage

Starting from the left, one thing was apparent from the outset: having more, better and timely data could provide a more granular view of our businesses' performance. And our ability to react quickly would certainly allow us to create some value. A health analogy may help to understand why.

Imagine you install sensors in your body, either externally through wearables or by means of other soon-to-be-invented internal devices, that provide you with more, better and timely data on your health. Since you may now know when your heart rate or your blood pressure increases above some critical level, you can take whatever measures are needed to take them back to normality. Similarly, you can track your sleeping patterns or sugar levels and adjust your daily habits accordingly. If we react fast enough, this newly-available data may even save our lives. This kind of descriptive analysis of past data may provide some insights about your health, and the creation of value depends critically on our ability to react quickly enough.

Predictive stage

But more often than not it's too late when we react. Can we do better? One approach would be to replace reaction by predictive action. As long as predictive power is high enough, this layer should buy us time to find better actions, and thus, new opportunities to create value.

This new stage allowed us to develop new *data products* such as recommendation engines (think Netflix), but it also gave rise to the age of data monetization. The online advertising business was thus born, marking an important inflection point in our story. The dream of marketers came to life with the promise of *selling the right product to the right person at the right time*, all this thanks to the data and predictions created with it.

IMPORTANCE OF ONLINE ADVERTISING

Most of the riches created by big data were the product of the success of online advertising. The online advertising business is huge and highly lucrative. One source estimates that more than $500 billion will be spent during 2023 across the globe (*https://www.emarketer.com/content/global-digital-ad-spending-2019*). If that figure doesn't say much, it is close to Belgium's Gross Domestic Product (*https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)*).

The two main players in this business are Google and Facebook. They have built their businesses largely funded by the revenues from this profitable industry, and thanks to the riches that came with them they have been able to fund the fast recent development on the AI arena (many times through acquisitions).

So it seems fair to say that the success of big data in online advertising has played an important role in facilitating the current AI hype.

Prescriptive stage

The top rank in this hierarchy of value creation is taken by our ability to automate and design intelligent systems. We are now at the *prescriptive* layer: once you have enough predictive power you can start finding the *best* actions for your business objectives. This is the layer where firms move from prediction to optimization, the throne in the data olympus, and interestingly enough, this is the least explored step in most maturity models.

Understanding what failed

In less than 15 years we've lived through two hypes — the big data revolution followed by the current AI stage — so you may wonder why the promises have yet to be fulfilled.

I'm not a big fan of data maturity models but I believe the answer lies within them: *most companies have yet to arrive to the prescriptive stage*. Big data was all about the descriptive stage, and as we've mentioned, AI is primarily concerned about prediction. Since everything has been laid out for us in the last few years, it begs the question of what's behind our apparent inability to move forward.

I'm convinced that market forces are an important factor, meaning that once a hype begins, market players want to reap the benefits until completely exhausted before moving on to the next big thing. Since we're still in that phase there are no incentives to move forward yet.

But it is also true that to become prescriptive we need to acquire a new set of analytical skills. As of today, with the current technology, this stage is done by humans, so we need to to prepare humans to pose and solve prescriptive problems. This book aims at taking us closer to that objective.

Analytical skills for the modern AI-driven enterprise

Tom Davenport's now classic *Competing on Analytics* pretty much equates analytical thinking with what later became to be known as

data-drivenness: “By analytics we mean the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions.” One alternative definition can be found in Albert Rutherford’s *The Analytical Mind*: “Analytical skills are, simply put, problem-solving skills. They are characteristics and abilities that allow you to approach problems in a logical, rational manner in an effort to sort out the best solution.”

In this book I will define *analytical reasoning* as the ability to translate business problems into *prescriptive solutions*. This ability entails both being data-driven, and being able to solve problems rationally and logically, so the definition is in fact in accordance with the two described above.

To make things practical, I will equate business *problems* with business *decisions*. Other problems that are purely informative and do not entail actions may have intrinsic value for some companies, but I will not treat them here, as my interest is in creating value through analytical *decision-making*. Since most decisions are made without knowing the actual consequences, AI will be our weapon to embrace this intrinsic uncertainty. Notice that under this approach, prediction technologies are important *inputs* into our decision-making process but not the *end*. Improvements in the quality of predictions can have first- or second-order effects depending on whether we are already making near-to-optimal choices.

Key takeaways

- **Most companies haven't been able to create value through data or AI in a sustainable and systematic way**: nonetheless, many have already embarked on their own efforts just to reach a wall of disappointment.
- **Today's AI is about prediction**: AI is overhyped, not only because of its deceiving name but also because there is so much one achieve through better prediction. These days AI most commonly refers to deep learning. Deep neural networks are highly nonlinear prediction algorithms that have shown remarkable success in the areas of image recognition and natural language processing.
- **Before AI we had the big data revolution**: the data revolution preceded the current hype and also came with the promise to generate outstanding business results. It was built around the 3Vs — volume, variety and velocity — and later complemented with prediction algorithms and data-driven culture.
- **But data and prediction cannot create sustainable value by themselves**: maturity models suggest that value is created by making optimal decisions in a data-driven way. For this, we need data and prediction as inputs in our decision-making process.
- **We need a new set of analytical skills to be successful in this prescriptive stage**: current technology precludes us from automating the process of translating business problems into prescriptive solutions. Since humans need to be involved all along the way, we need to upscale our skillset to capture all the value from data and AI-driven decision-making.

Futher Reading

2019 and 2020 witnessed a very interesting debate on the limits on what can be achieved through AI. You can see one such debate in the discussion that Gary Marcus and Joshua Bengio had in Montreal (*https://www.youtube.com/watch?v=EeqwFjqFvJA*). If you prefer reading, Gary Marcus' and Ernst Davis' *Rebooting AI: Building Artificial Intelligence We Can Trust* will provide many of the details on why many are critical about deep learning being the way to achieve AGI.

On the topic of how AI will affect businesses I highly recommend Ajay Agrawal's, Joshua Gans' and Avi Goldfarb's, *Prediction Machines. The Simple Economics of Artificial Intelligence*. Written by three economists and AI strategists, they provide a highly-needed, away-from-the-hype, down-to-earth account of current AI. Their key takeaway is that thanks to current developments, the cost of predictive solutions within the firm has considerably fallen while quality has kept increasing, providing great opportunities for companies to transform their business models. Also written by economists Andrew McAffe and Erik Brynjolfsson, *Machine Platform Crowd. Harnessing Our Digital Future* discusses how the data, artificial intelligence and digital transformations are affecting our businesses, the economy and society as a whole.

Data maturity models appear on several books: you can check Thomas Davenport's and Jeane Harris' *Competing on Analytics, Big Data at Work: Dispelling the Myths, Uncovering the Opportunities* also by Tom Davenport or Bill Schmarzo's *Big Data: Understanding How Data Powers Big Business*.

If you're interested in learning more about our quest to achieve AGI, Nick Bostrom's *Superintelligence. Paths, Dangers, Strategies* discusses at great length and depth what intelligence is and how superintelligence could emerge, as well as the dangers from this development and how it can affect society. Similar discussions can be found in Max Tegmark's *Life 3.0. Being Human in the Age of Artificial Intelligence*.

Finally, on the podcast side, I recommend following Lex Fridman's Artificial Intelligence (*https://lexfridman.com/ai/*). There are many great interviews with leaders in the field that will provide much more context on the current state of affairs.

---

1 See here *https://twitter.com/andrewyng/status/788548053745569792?lang=en*

2 The field of AI knows very well about this risk as it has lived at least two "winters" where funding was almost entirely denied to any researcher.

3 Data from *https://en.wikipedia.org/wiki/List_of_public_corporations_by_market_capitalization*. Retrieved March 2020. In the plot I use the information corresponding to the last quarter only.

4 *https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data*

5 *https://static.googleusercontent.com/media/research.google.com/es//archive/mapreduce-osdi04.pdf*

6 *https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#4cd02c7d60ba*

Chapter 2. Intro to Analytical Thinking

A NOTE FOR EARLY RELEASE READERS

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

This will be the 2nd chapter of the final book. Please note that the GitHub repo will be made active later on.

If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the author at analyticalthinkingbook@gmail.com.

In the last chapter I defined *analytical thinking* as the ability to translate business problems into prescriptive solutions. There is a lot to unpack from this definition, and this will by our task in this chapter.

To really understand the power of prescriptive solutions, I will start by precisely defining each of the three stages present in any analysis of business decisions, these are the descriptive, predictive and prescriptive steps we have already mentioned in Chapter 1.

Since one crucial skill in our analytical toolbox will be formulating the right business questions from the outset, I will provide a first glimpse into this topic. Spoiler alert: we will only care about business questions that entail business decisions. We will then disect decisions

into levers, consequences and business results. The link between levers and consequences is intermediated by *causation* so I will devote quite a bit of time talking about this topic. Finally, I will talk about the role that uncertainty plays in business decisions. Each of these topics is tied to one skill that will developed throughout the book.

Descriptive, Predictive and Prescriptive Questions

In Chapter 1 we saw that data maturity models usually depict a nice, smooth road that starts at the descriptive stage, goes through the predictive plateau to finally ascend to the predictive summit. But why is this the case? Let’s start by understanding what these mean and then we can discuss why commentators and practitioners alike believe that this is the natural ascension in the data evolution.

In a nutshell, *descriptive* relates to how things are, *predictive* to how we believe things will be, and *prescriptive* to how things ought to be. Take Tyrion Lannister´s quote in Game of Thrones’ “The Dance of the Dragons” episode: “It’s easy to confuse what is with what *ought* to be, especially when *what is* has worked out in your favor” (my emphasis). Tyrion seems to be claiming that when the outcome of a decision turns out to be positive, we think that this was the best we could do (a type of confirmation bias). Incidentally, when the outcome is negative, our tendency is to think that this was the worst possible result and attribute our fate to some version of Murphy´s Law.

In any case, as this discussion shows, the prescriptive stage is a place where we can rank different options so that words like “best” or “worst” make any sense at all. It follows that the prescriptive layer can never be inferior to the descriptive one, as in the former we can always make the best decision.

But what about prediction? To start, its intermediate ranking is at least problematic, since description and prescription relate to the *quality of decisions*, and prediction is an input to make decisions, which may or may not be optimal or even good. The implicit assumption in all maturity models is that the quality of decisions can be improved when we have better predictions about the underlying uncertainty in the problem; that good predictions allow us to plan ahead and move proactively, instead of reacting to the past with little or no room for maneuver.

When is predictive analysis powerful: the case of cancer detection

Let’s take an example where better prediction can make a huge difference: cancer detection.[1] Oncologists usually use some type of visual aid such as X-rays or the more advanced CT scans for early detection of different pathologies. In the case of lung cancer an X-ray or a CT scan is a description of the patient’s current health status. Unfortunately, visual inspection is highly ineffective unless the disease has already reached a late stage, so description here, by itself, may not provide enough time to react quickly enough. AI has shown remarkable prowess in predicting the existence of lung cancer from inspecting CT scans, by identifying spots that will eventually show to

be malignant.[2] But prediction can only take us so far. A doctor should then recommend the right course of action for the patient to fully recover. AI provides the predictive muscle, but humans prescribe the treatment.

Descriptive Analysis: the case of customer churn

Let’s run a somewhat typical descriptive analysis of a use case that most companies have dealt with: customer churn or attrition. We will see that without guidance from our business objectives, this type of analysis might take us to a dead end.

HOW HAS CUSTOMER CHURN EVOLVED IN THE RECENT PAST?

Suppose that your boss wants to get churn under control. As a first step, she may ask you to diagnose the magnitude of the problem. After wrangling with the data you come up with the following two plots (Figure 2-1). The left plot shows a time series of daily churn rates. Confidently, you state two things: after having a relatively stable beginning of the year, churn is now on the rise. Second, there is a clear seasonal pattern, with weekends having lower than average churn. In the right panel you show that municipalities with higher average income also have higher churn rates, which of course, is a cause of concern since your most valuable customers may be switching to other companies.

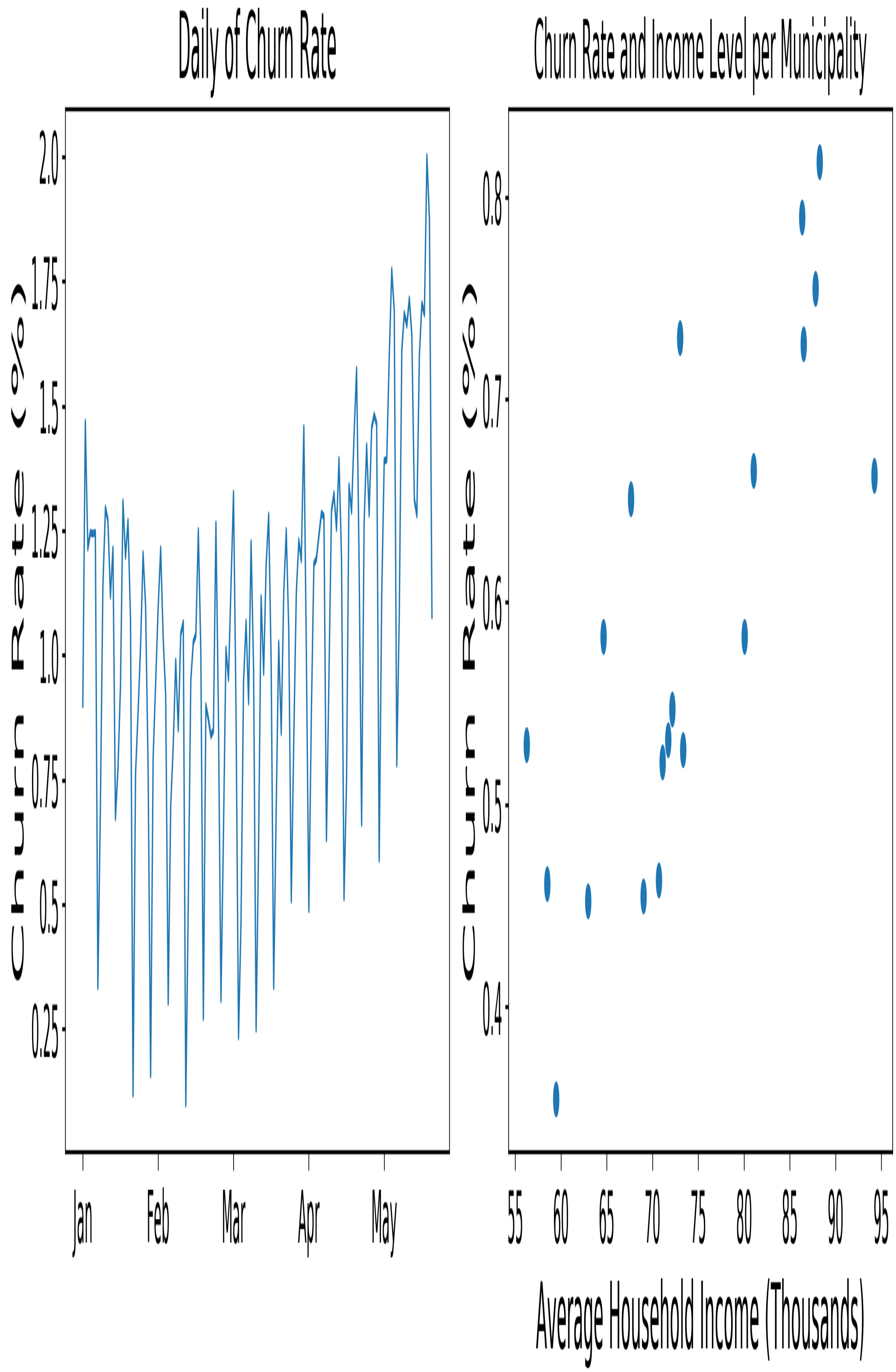

Daily of Churn Rate
Churn Rate (%)
2.0
1.75
1.5
1.25
1.0
0.75
0.5
0.25
Jan
Feb
Mar
Apr
May
Churn Rate and Income Level per Municipality
Churn Rate (%)
0.8
0.7
0.6
0.5
0.4
55
60
65
70
75
80
85
90
95
Average Household Income (Thousands)

Figure 2-1. Descriptive analysis of our company's churn rate

This is a great example of what can be achieved with descriptive analysis. Here we have a relatively granular, up-to-date picture of the problem, and unfortunately, news are not good for the company: your boss was right to ask for this analysis, since churn is on the rise, having reached a yearly record at the end of May with no signs of going back to previous levels. Moreover, our remarkable ability to recognize patterns allows us to clearly identify three patterns in the data: changes in the trend and the existence of seasonal effects in our time series, and positive correlation between two variables in the scatterplot.

There are problems and risks with this type of analysis, however. As you've probably heard elsewhere, *correlation does not imply causation*, a topic that will be discussed at length later in this chapter. For instance, one plausible recommendation could be that the company should pull away from richer municipalities, as measured by average household income. This follows from a causal interpretation going from household incomes to churn rates, albeit an incorrect one.

Moreover, the question remains as to how much value was provided by this detailed snapshot of the current state of affairs. We now know that churn is rising (admitedly, this is better than not knowing) but since we don't know the root causes, it will be hard, if not impossible, to devise some guidance for improvement. You may argue that if we further inspected the data we may find out what's behind this upward trend, and this is the right way to proceed: formulate hypothesis that

guide our analysis of the data. Inspecting data without advancing some plausible explanations is the perfect recipe for making your analytics and data science teams waste valuable time.

THE TRAP OF FINDING *ACTIONABLE* INSIGHTS

One common catch phrase among consultants and vendors of big data solutions is that once given enough data your data analysts and data scientists will be able to find *actionable insights*.

This is a common trap among business people and novice data practitioners: the idea that given some data, if we inspect it long enough, these actionable insights will emerge, almost magically. I've seen teams spending weeks waiting for the actionable insights to appear, without any luck.

Experienced practitioners reverse engineer the problem: start with the question, formulate hypotheses, and use your descriptive analysis to find evidence against or in favor of these hypotheses. Note the difference: under this approach we actively search of actionable insights by first deciding where to look for them, as opposed to waiting for them to emerge from chaos.

PREDICTING CHURN

As a next step, your boss may ask you to *predict* churn in the future. How should you proceed? It really depends and what you want to achieve with this analysis. If you work in finance, for example, and you're interested in forecasting the income statement for the next quarter, you'd be happy to predict aggregate churn rates into the future. If you are in the marketing department, however, you may want to predict which customers are at risk of leaving the company, possibly because you may try using different retention campaigns.

WHAT IS THE BEST LEVER WE CAN PULL TO PREVENT CUSTOMERS FROM CHURNING

Finally, suppose that your boss asks you to recommend alternative courses of action to *reduce* the rate of customer churn. This is where the prescriptive toolkit becomes quite handy and where the impact of making good decisions can be most appreciated. You may then pose a cost-benefit analysis for customer retention and come up with a rule such as the following: retain a customer whenever the benefit (future revenue stream that could've been lost) is higher than the cost from the retention lever. If you have several levers at your disposal, the recommended course of action is to use the one with higher incremental impact for the company. If you only have one, the course of action is to pull it only with customers where the campaign leaves a positive margin and let go the remaining ones.

We will have the opportunity to go into greater detail on this use case, but let me just single out two characteristics of any prescriptive analysis: as opposed to the two previous analyses, here we actively recommend courses of action that can improve our position, by way of incentivizing a likely-to-leave customer to stay longer with us. Second, prediction is used as an input in the decision-making process, helping us calculate *expected* savings and costs. AI will help us better estimate these quantities, necessary for our proposed decision rule. But it is this decision rule that creates value, not prediction itself.

One of the objectives of this book is to prepare us on how to translate business questions into prescriptive solutions, so don't worry if it's not obvious yet. We will have time to go through many step-by-step examples.

Business questions and KPIs

One foundational idea in the book is that value is created from making *decisions*. As such, prediction in the form of machine learning is just an input to create value. In this book, whenever we talk about business questions, we will always have in mind business decisions. Surely, there are business questions that are purely informative and no actions are involved. But since our aim is to systematically create value, we will only consider actionable questions. As a matter of fact, one byproduct of this book is that we will learn to look for actionable insights in an almost automatic fashion.

It begs the question, then, of *why* we have to make a decision. Only by answering this question will we be able to know how to measure the appropriateness or not of the choices we make. Decisions that cannot be judged in the face of any relevant evidence are to be discarded. As such, we will have to learn how to select the right metrics to track our performance. Many data science projects and business decisions fail not because of the logic used but because the metrics were just not right for the problem.

There is a whole literature on how to select the right key performance indicators (KPIs), and I believe I have little to add on this topic. The two main characteristics I look for are *relevance* and *measurability*. A KPI is relevant when it allows us to clearly assess the results from our decisions *with respect* to the business question. Notice that this doesn't have to do with how pertinent the business question is, but rather, on whether we are able to evaluate if the decision worked or

not, and by how much. It follows that a good KPI should be measurable and this better be with little or no delay with respect to the time when the decision was made. Not only is there an opportunity cost of delayed measurement, but it may also be harder to identify the root cause.

SMART KPIS

For us, a good KPI has to be relevant and measurable. Compare this with the now classic SMART definition of good KPIs. The acronym stands for: Specific, Measurable, Achievable, Relevant and Timely. We have already mentioned the time dimension, and it's hard to argue against specificity: there is a long distance between "improving our company's state" and "increasing our profit margins". The later is quite specific and the former is so abstract that it can't be actionable.

In my opinion, however, the property of being achievable sounds closer to a definition of a goal than to a performance indicator.

KPIs to measure the success of a loyalty program

Let's briefly discuss one example. Suppose that our Chief Marketing Officer asks us to evaluate the creation of a loyalty program for the company. Since the question starts with an action (creating or not the loyalty program) it immediately classifies for us as a business problem. What metrics should we track? To answer this let's start the sequence of *why* questions.

- Create a loyalty program. *Why?*
- Because you want to reward loyal customers. *Why?*

- Because you want to incentivize customers to stay longer with the company. *Why?*
- Because you want to increase your revenues in the longer term. *Why?*

THE SEQUENCE OF *WHY* QUESTIONS

This example is showcasing a technique that I call *the sequence of why questions*. It is used to identify the business metric that we want to optimize.

It works by starting with what you, your boss or colleagues may think they want to achieve and question the reasons for focusing on such objective. Move one step above and repeat. It terminates when you're satisfied with the answer. Just in passing, recall that to be satisfied you must have a relevant and measurable KPI to quantify the business outcome you will focus on.

And of course, the list can go on. The important thing is that the final answer to these questions will usually let you clearly identify what KPI is relevant for the problem at hand, and any intermediate metrics that may provide useful; if it's also measurable then you have found the right metric for your problem.

Consider the second question, for example. Why would anyone want to reward loyal customers? They are already loyal, without the need for any extrinsic motivation, so this strategy may even backfire. But putting aside the underlying reasoning, why is loyalty meaningful and how would you go about measuring the impact of the reward? I argue that loyalty by itself is not meaningful: we prefer loyal customers to not-so-loyal customers because they represent a more stable stream of revenues in the future. If you're not convinced, think about those

loyal but *unprofitable* customers. Do you still rank their loyalty as high as before? Don't feel bad if your answer is negative: it just means that you are doing business because you want to make a decent living. If loyalty per se is not what you're pursuing then you should keep going down the sequence of *why* questions, since it appears that we are aiming at the wrong objective.

Just for the sake of the discussion, suppose that you still want to reward loyal customers. How do we measure if our the program worked, or put differently, what is a good KPI for this? One commonly used method is to directly ask our customers, as done with the Net Promoter Score (NPS). To calculate the NPS we first ask our customers how likely they would recommend us as a company on a scale from 0 to 10. We than classify them into *Promoters* (9 to 10), *Detractors* (0 to 6) and *Passive* (7,8). Individual answers are finally aggregated into the NPS by subtracting the percentage of detractors from the percentage of promoters.

On the bright side, this is a pretty *direct* assessment: we just go and ask our customers if they value the reward. It can't get more straightforward than that. The problem here is that humans act upon motivations so we generally can't tell if the answer is truthful or if there's some other underlying motive and are trying to game our system. This type of strategic considerations matter when we assess the impact of our decisions.

An alternative is to let the customers indirectly *reveal* their level of satisfaction through their actions, say from the amount, frequency or ticket in their recent transactions, or through a lower churn rate for

those who receive the reward relative to a well-designed control group.[3] Companies will always have customer surveys, and they should be treated as a potentially rich source of information. But a good practice is to always check if what they *say* is supported by their actions.

An Anatomy of a Decision: a simple decomposition

Figure 2-2 shows the general framework we will use to decompose and understand business decisions. Starting from the right, it is useful to repeat one more time that we *always start with the business* — possibly making use of the sequence of *why* questions described above, that allow us to precisely pinpoint what we wish to accomplish. If your objective is unclear or fuzzy, most likely the decision shouldn't be made at all. Companies tend to have a bias for action, so fruitless decisions are sometimes made; this not only may have unintended negative consequences on the business side but may also take a toll on employees' energy and morale. Moreover, we now take for granted that our business objective can be measured through relevant KPIs. This is not to say that metrics arise naturally: they must be carefully chosen by us, the humans, as will be shown below with an example.

Decomposing Decisions

Figure 2-2. Decomposing decisions

It is generally the case that we can't simply manipulate those business objectives ourselves (remember Enron[4]), so we need to take some

actions or pull some levers in order to try to generate results. Actions themselves map to a set of consequences that directly affect our business objective. To be sure: *we* pull the levers, and our business objective depend on consequences that arise when the “environment” reacts. The environment can be humans or technology, as we will see later.

Even if the mapping is straightforward (most times it isn’t) it’s still mediated by uncertainty, since at the time of the decision it is impossible to know exactly what the consequences will be. We will use the powers of AI to embrace this underlying uncertainty, allowing us to make better decisions. But make no mistake: *value is derived from the decision and prediction is an input to make better decisions*.

To sum up, in our daily lives and in the business, we generally pursue well-chosen, measurable objectives. Decision-making is the act of choosing among competing actions to attain these objectives. Data-driven decision-making is acting upon evidence to assess alternative courses of action. Prescriptive decision-making is the science of choosing the action that produces the best results for us; we must therefore be able to rank our choices relative to a measurable and relevant KPI.

An example: why did you buy this book?

One example should illustrate how this decomposition works for *every* decision we make (Figure 2-3). Take your choice to purchase this book. This is an action you already made, but, surely, you could have decided otherwise. Since we always start with the business

problem, let me imagine what type of problem you were trying to solve.

Decomposing Decisions

Figure 2-3. Decomposing your decision to buy this book

Since this book is published by O’Reilly Media, most likely your objective is to advance your career and not just to have a nice, pleasurable Friday night read.[5] This sounds like a medium-to-long run goal, and one possible metric is the number of interviews you get once you master the material (or at least write it down on your résumé, update your LinkedIn profile or the like). If you don’t want to change jobs, but rather be more effective at your current position, alternative metrics could be the number of end-to-end delivery of data science projects, number of ideas for new projects, or even the incremental dollar value these projects generated for your company. Notice how we must adjust the KPIs to different objectives. For now, let me just assume that your goal is to be more effective at work, as it might be easier to measure.

The set of possible levers you can pull is now larger than just “buying” this book or “not”. You could have, for instance, adopted the “seven habits of highly effective people”, enrolled on an online course, keep improving your technical skills, improve your interpersonal skills, bought other books, or just do nothing. The advantage of starting with the business problem — as opposed to a set of specific actions like “buy” or “not buy" — is that your menu of options usually gets enlarged.

To simplify even more, let us assume that we only consider two actions: buy or not buy. If you don't buy it (but please do) your productivity may keep increasing at the current rate. This is not the only possible consequence, of course. It could be that you get a sudden burst of inspiration and surprisingly start understanding all the intricacies of your job, positively and dramatically increasing your productivity. Or the opposite could happen, of course. The universe is full of examples where symmetry dominates. Nonetheless, let's appeal to Occam's razor and consider the only consequence that seems likely to occur: no impact on your productivity.

OCCAM'S RAZOR

When there are many plausible explanations to a problem, the principle known as Occam's razor appeals for the simplest one. Similarly, in statistics, when we have many possible models to explain an outcome, if we apply this principle we would attempt at using the most parsimonious one.

We will devote a whole chapter on the skill of simplification.

If you do buy and read this book, now we have at least three likely consequences: the book works and improves your analytical skills, it does nothing, or it worsens your skills. Contrary to the previous analysis, in this case the latter is likely and should survive Occam's razor: I could be presenting some really bad practices that you haven't heard of and that you misguidedly end up adopting. Now, at the time of making the decision you don't really know the actual consequence so you may have to resort to finding additional information, read reviews or use heuristics to assess the likelihood of

each possible outcome. This is the underlying uncertainty in this specific problem.

To sum up, notice how a simple action helped us to clearly and logically find the problem being solved, a set of levers, their consequences, and the underlying uncertainty. This is in general a good practice that applies to any decision you make: if you are already making choices or decisions, think back to what specific problem you are attempting to solve — you can even try answering the sequence of *why* questions — and then reverse engineer a set of possible actions, consequences and uncertainty. Once we set up the decision problem we are ready to find the best course of action.

A primer on causation

We will devote a chapter on each of the "stages" in the decomposition, so there will be enough time to understand where these levers come from and how they map to consequences. It is important, though, to stop now, and recognize that this mapping is mediated by *causal* forces.

Going back to the saying that "correlation does not imply causation", no matter how many times we've heard about it, it is still very common to get the two terms confused. Our human brain evolved to become a powerful pattern recognizing machine, but we are not so well equipped to distinguish causation from correlation. To be fair, even after taking into account this apparent impairment, we are by far the most sophisticated causal creatures that we know of, and infinitely superior to machines (since at the time of the writing they

completely lack the ability and it is not even clear when this ability may be achieved or if it's achievable at all).

Defining correlation and causation

Strictly speaking, correlation is the presence or absence of any linear dependencies in two or more variables. Though technically accurate, we can dispense of the "linear" part and be concerned about general relationships between variables. For instance, the scatterplot in Figure 2-1, showed that average household income in each municipality was positively correlated with churn: they tend to move in the same direction, so that on average, higher (lower) churn in a municipality is associated with higher (lower) average income.

Causality is harder to define, so let us take the shortcut followed by almost everyone: a relation of causality is one of cause and effect. X (partially) causes Y if Y is (partially) an effect of X. The "partial" qualifier is used because rarely is one factor the unique source of a relationship. To provide an alternative, less circular, definition let us think in terms of counterfactuals: *had X not taken place, is it true that Y had been observed?* If the answer is positive then it is unlikely that a causal relationship from X to Y exists. Again, the qualifier "unlikely" is important and related to the previous "partial" qualifier: there are causal relations that only occur if the right combination of causes is present. One example is whether our genes determine our behavior: it has been found that our genetic makeup by itself is generally not the unique cause of our behavior; instead, the right combination of genes and environmental conditions are needed for behavior to arise.

Going back to the scatterplot example, our brain immediately recognizes a pattern of positive correlation. How do we even start thinking about causality? It is common to analyze each of the two possible causal directions and see whether one, the other or both, make sense with respect to our understanding of the world. Is it possible that the high churn rates *cause* the higher average income in each municipality? Since household income usually depends on more structural economic forces — such as the education levels of the members of the household, their occupations and employment status — this direction of causality seems dubious, to say the least. We can easily imagine a counterfactual world where we lower the churn rates (say, by aggressively giving retention discounts) without changing the recipient household's income in a significant way.

What about the other direction? Can higher income be the cause of the higher rates of churn. It is plausible that higher income customers — paying higher prices — also expect higher quality, on average. If the quality of our product doesn't match their expectations they may be more likely to switch companies. How would the counterfactual work? Imagine we could artificially increase some of our customers' household incomes. Would their churn rate increase? This ability to create counterfactuals is fundamental to even have a conversation about causality.

Understanding Causality: some examples

To fully appreciate the difficulty in identifying causality from data let's look at some examples.

SIMULATED DATA

Let’s start by analyzing the data in Figure 2-4. Here, again, we immediately identify a very strong positive correlation between variables Y and X. Can it be that two variables move together in such a strong manner, and there is no causal relationship between them? One thing should be clear from the outset: there is no way we can device causal stories without having some context, that is, without knowing what X and Y are and how they relate to the world.

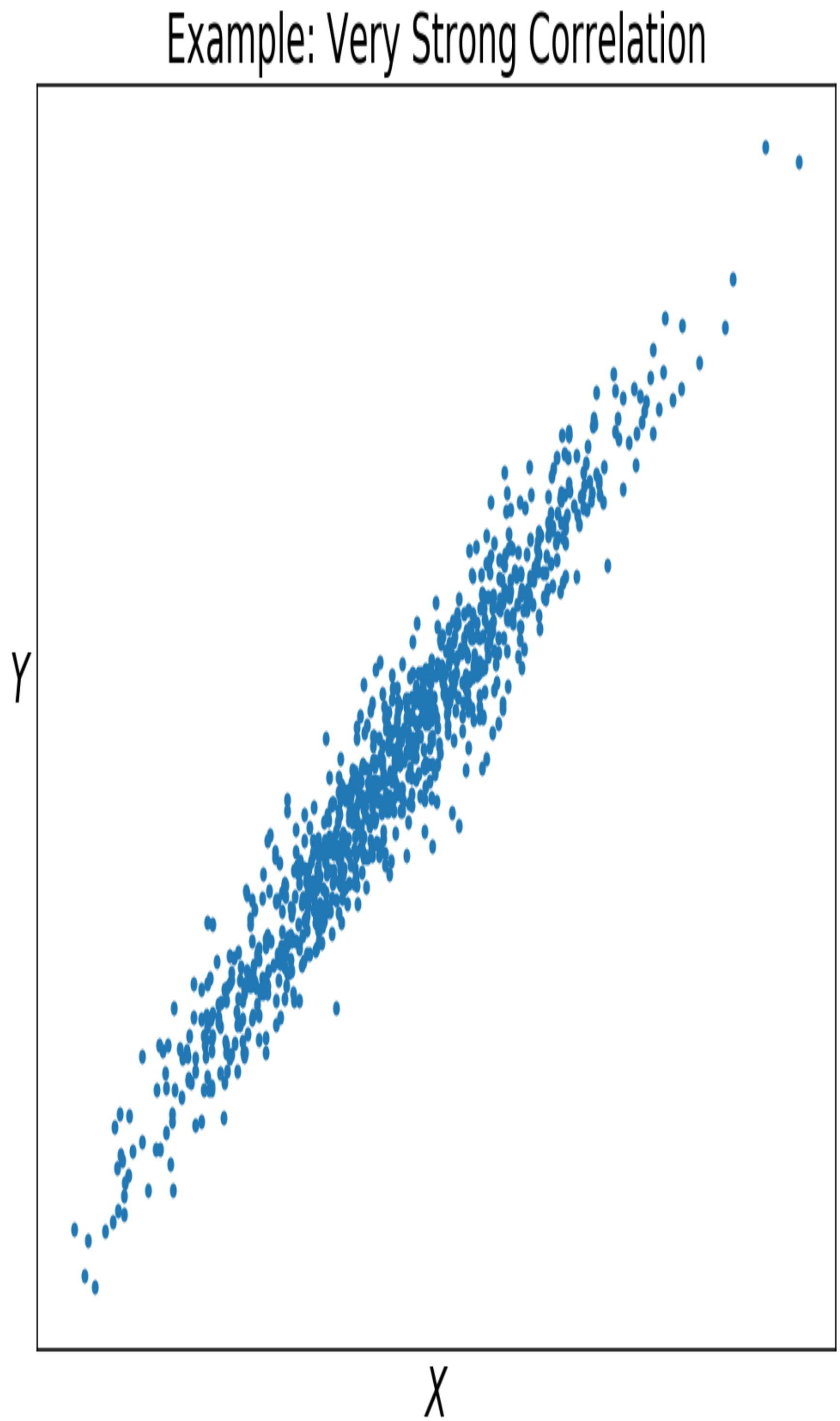

Example: Very Strong Correlation
Y
X

Figure 2-4. A simulation of two highly correlated variables

This is an example of a *spurious* correlation, the case when two variables falsely appear to be related. The source of this deception is the presence of a *third variable* Z that affects both variables, $Z \implies X$ and $Z \implies Y$; if we don't *control* for this third variable, the two will appear to move together even when they are not related at all. I know that this is the case because the following Python code was used to simulate the data.

Example 2-1. Simulating the effect of a third unaccounted variable on the correlation of other two

```
# fix a seed for our random number generator and number of observations to simulate
np.random.seed(422019)
nobs = 1000
# our third variable will be standard normal
z = np.random.randn(nobs,1)
# let's say that z --> x and z--> y
# Notice that x and Y are not related!
x = 0.5 + 0.4*z + 0.1*np.random.randn(nobs,1)
y  = 1.5 + 0.2*z + 0.01*np.random.randn(nobs,1)
```

A simulation was used to unequivocally show the dangers of a third variable that is not taken into account in our analysis, so you may wonder if this is something we should worry about in your day-to-day work. Unfortunately, spurious correlations abound in the real world, so we should better learn to identify them and find workarounds. The effect of misrepresenting causation on the quality of a decision will not only lead to ineffective decision-making but also to a loss of valuable time when developing the predictive algorithms.

CHURN AND INCOME

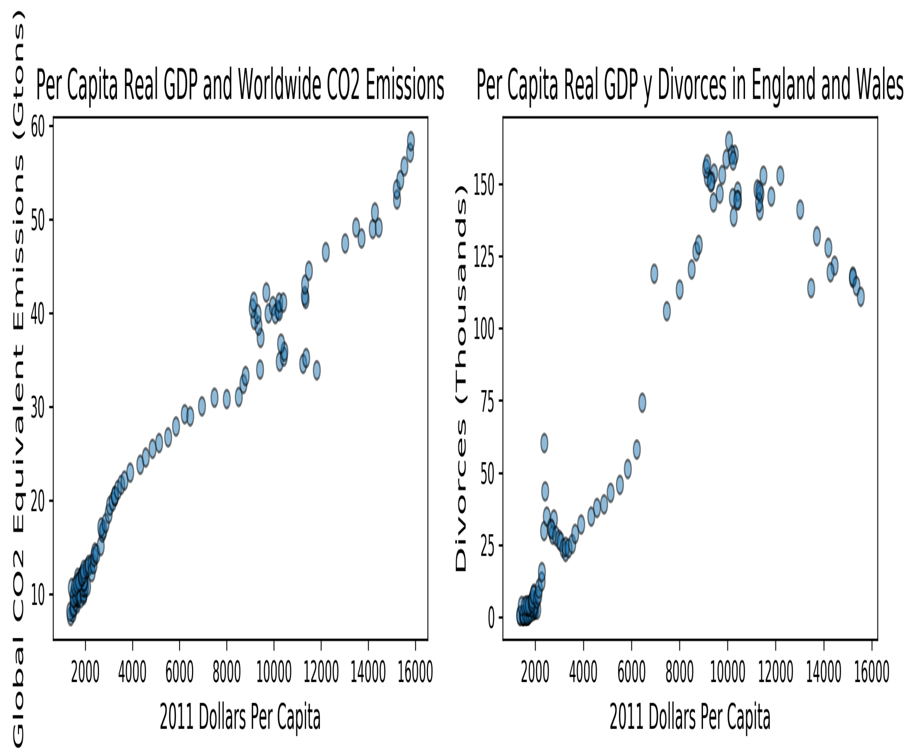
Let us quickly revisit the positive association found between churn rates and households' average income per municipality (Figure 2-1) and imagine that a strong competitor has entered the market with an aggressive pricing strategy targeted at customers in the medium-to-high income segments. It will be the case, then, that this third variable explains the positive correlation: more of your higher income customers will churn across all municipalities, but the effect will be higher in those where their relative size is also larger.

CAN DIVORCES IN ENGLAND EXPLAIN POLLUTION IN MEXICO?

Consider now the examples in Figure 2-5. The top left panel plots a measure of global CO2 emissions and per capita real gross domestic product (GDP) in Mexico for the period 1900-2016. The top right panel plots the number of divorces in Wales and England against Mexican GDP for 1900-2014. The bottom panel plots the three time series, indexed so that the 1900 observation is 100.[6]

Inspecting the first scatter plot, we find a very strong, almost linear relationship between the state of the Mexican economy, as measured by per capita real GDP, and global CO2 emissions. How can this be? Let's explore causality in both directions: it is unlikely that CO2 emissions cause Mexican economic growth (to the best of my knowledge, CO2 is not an important input for any production processes in the Mexican economy). Since the Mexican economy isn't that large on a global scale, it is also unlikely that Mexico's economic growth has had such an effect on global contaminants. One can imagine that fast-growing economies like China and India (or the US and Great Britain during the 19th and 20th Centuries) would be

responsible for a big part of global CO_2 emissions, but this is unlikely for the case of Mexico.


Per Capita Real GDP and Worldwide CO2 Emissions
Global CO2 Equivalent Emissions (Gtons)
2011 Dollars Per Capita
Per Capita Real GDP y Divorces in England and Wales
Divorces (Thousands)
2011 Dollars Per Capita

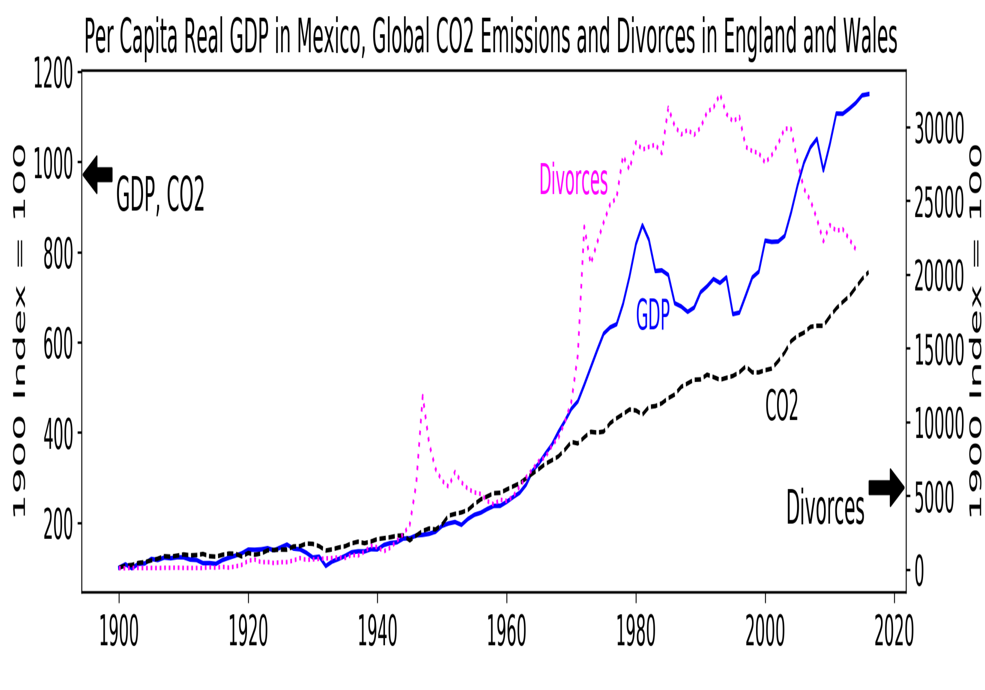

Per Capita Real GDP in Mexico, Global CO2 Emissions and Divorces in England and Wales
1900 Index = 100
GDP, CO2
Divorces
GDP
CO2
Divorces
1900 Index = 100

Figure 2-5. Top left panel plots global CO2 emissions against real per capita Gross Domestic Product (GDP) for Mexico for the period 1900-2016. Top right panel does the same, replacing CO2 emission with the number of divorces in Wales and England during 1900-2014. Bottom plot shows the time series for each of these variables.

The second scatterplot shows an even more striking relationship: per capita real GDP in Mexico is positively related to the number of divorces in England and Wales, but up to a certain point (close to $10K dollars per person); after reaching that level the relationship becomes negative. Causal stories in this case become rather convoluted. Just for illustration, one such story — from economic growth in Mexico to divorce rates in the UK — could be that as the Mexican economy developed, more English and Welch people migrated to the North American country to find jobs and share the pie of economic prosperity. This could have created broken homes and a high prevalence of divorce. This story is plausible, but highly unlikely, so there must be some other explanation.

As before, there is a third variable that explains the very strong but spurious correlations found in the data. This is what statisticians and econometricians call a *time trend*, that is, the tendency of a time series to increase (or decrease) over time. The bottom plot depicts the three time series over time. Observe first the evolution of per capita GDP and CO2 emissions. The two time series evolve hand-by-hand until the late 1960s and beginning of the 1970s, thereafter maintaining different, but still positive, trends or growth rates. A similar comment can be made for the number of divorces. These trends are the third variable that is common to the three time series, creating strong but spurious correlations. For this reason, practitioners always start by *detrending* or controlling for the

common trend among different time series allowing them to extract more information from this noisy data.

HIGH VALUE CUSTOMERS HAVE LOWER NET PROMOTER SCORES

Let’s give some examples closer to the business now starting with the risks of making customer satisfaction comparisons across customer segments. Recall that the NPS is a metric commonly used to track customer satisfaction, so it is natural to compare it across segments as in Figure 2-6.

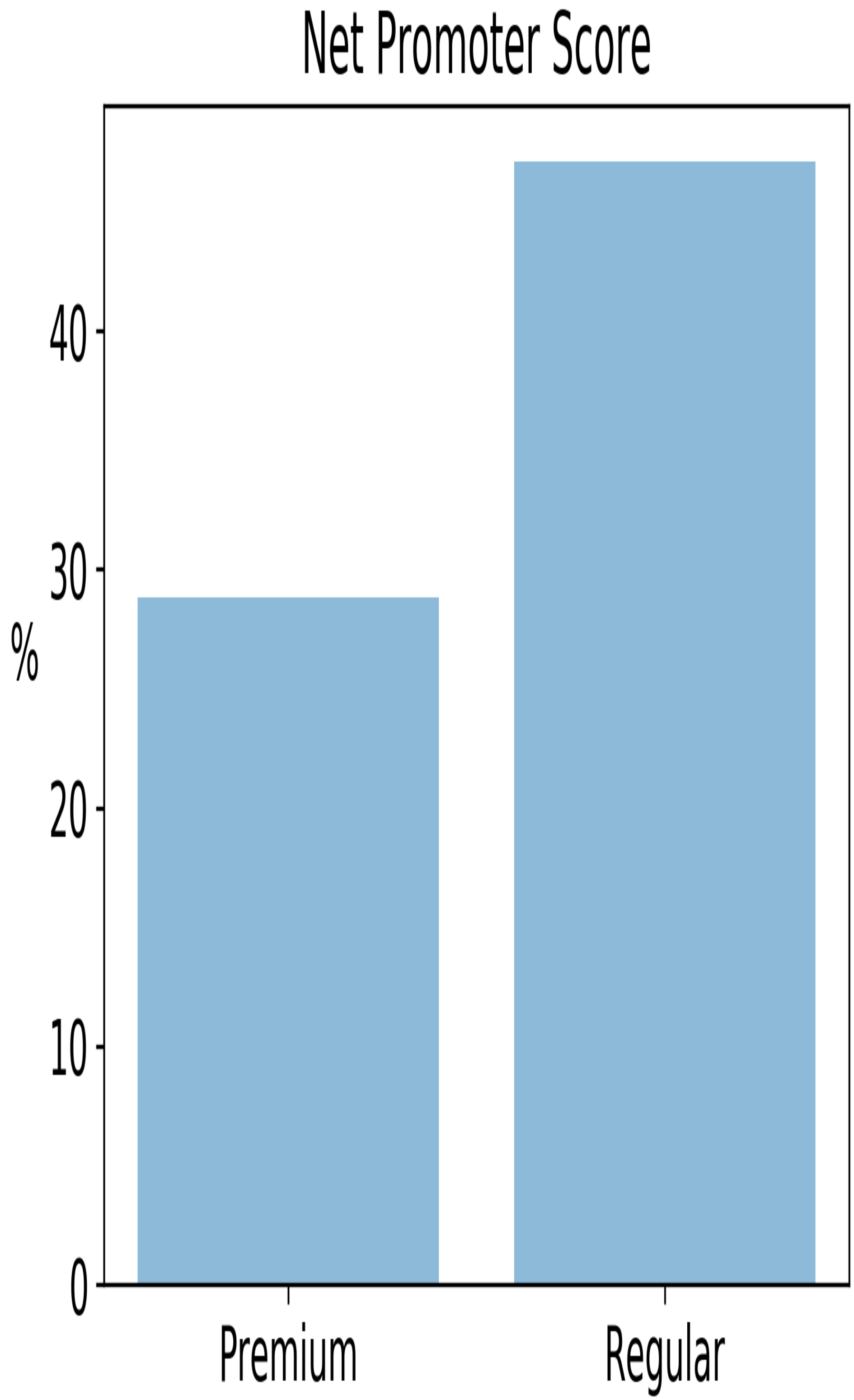

Net Promoter Score
%
0
10
20
30
40
Premium
Regular

Figure 2-6. Net Promoter Score (NPS) for two different value segments

The plot depicts average NPS for two different customer value segments, *Premium* and *Regular*, corresponding to customers with high and low customer lifetime values (CLV), respectively. The bars show that NPS is *negatively* correlated with the value of the customer, as measured by the CLV. Since this is a customer-centric company, one possible recommendation could be to focus only on lower-value customers (since they are the most satisfied). This is a causal interpretation that goes from value to satisfaction, and a customer's value is treated as a lever. It could be, however, that a third variable is affecting both the NPS and the CLV, and that once we control for this intervening variable the relationship disappears. One such possible third value is our customers' socioeconomic level, possibly capturing the higher quality expectations that we described in the churn example.

CUSTOMER LIFETIME VALUE (CLV)

How should we value our customers? One approach is to assign the current value derived from each one of them. The problem with this short-term view is that companies invest in their customers all the time, from acquisition to retention, marketing, etc, so to value those investments we also need the long-run view from the revenues side.

Several decades ago people started looking at customers as assets (*https://www0.gsb.columbia.edu/mygsb/faculty/research/pubfiles/721/gupta_customers.pdf*), and under this approach the right metric is the *stream* of profits derived from them. One difficulty with the stream approach is that at any time our customers may decide to change companies, so we need to incorporate an uncertain time window into the analysis.

The CLV measures the discounted present value of all profits obtained from a relationship with one customer along his or her expected duration with the company.

For instance, assuming a monthly discount rate of 1%, a new customer who will remain purchasing our goods and services for the next 11 months, leaving a monthly profit of $1 dollar, will have a CLV of $\$1 + \$1/(1.01) + \$1/(1.01)^2 + \cdots + \$1/(1.01)^{10}$ = $10.4 dollars. In practice, to compute the CLV we need an estimate of the expected duration of a customer with us, as well as an estimate of how profits change in time.

SELECTION EFFECTS AND THE HEALTH STATUS OF OUR EMPLOYEES

Another important reason why we cannot immediately identify causality when analyzing our data are *selection effects*.[7] Suppose that our Chief Human Resources Officer is considering saving costs by eliminating the company's on-site medical service, since in his opinion most employees use the company-provided off-site services. Since this is a data-driven company, he decides to apply an anonymous survey asking the following two questions to the employees:

1. How is your health in general? Please provide an answer from 1 to 5, where 1 is *Very Bad*, 2 is *Bad*, 3 is *Fair*, 4 is *Good* and 5 is *Excellent*.

2. In the last 2 months, have you been treated by our on-site doctors?

Figure 2-7 shows the average self-reported health status for groups of employees that declare having used the medical service or not. The CHRO was deeply concerned with the result presented by the Analytics Unit: treated employees report having worse health than those not treated. It seems that our on-site medical service is generating the exact opposite result it was designed for.

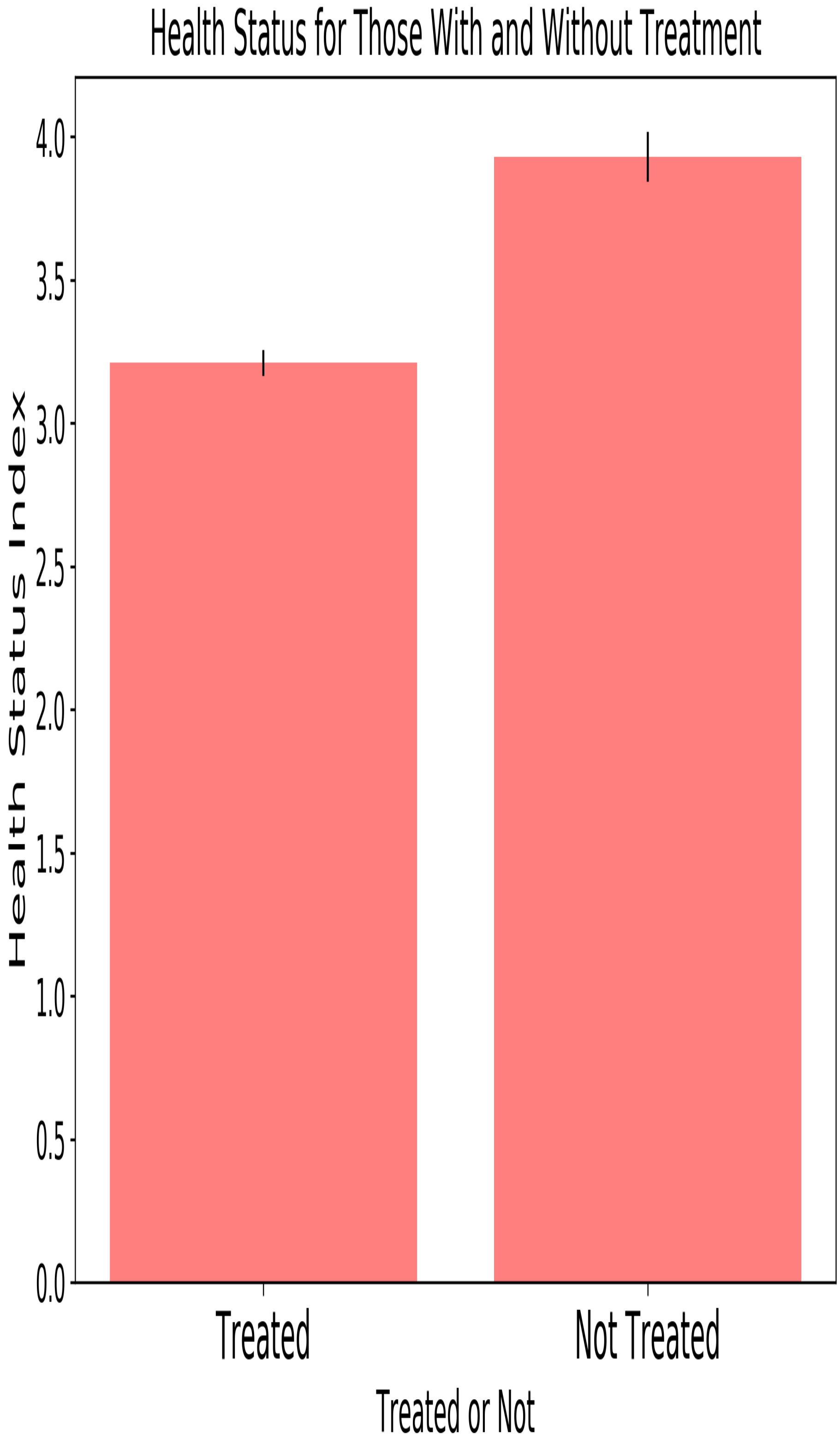

Health Status for Those With and Without Treatment
Health Status Index
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Treated
Not Treated
Treated or Not

Figure 2-7. Self-reported health status for employees treated not treated on-site. Vertical lines correspond to confidence intervals.

But is this analysis sound? One data scientist pointed out that it could be the result of selection effects. Her rationale is that employees that feel sick are most likely *self-selecting* into using the on-site service, so the results are evidence of the higher likelihood of going to the doctor when you're sick and *not* of a negative causal effect of providing on-site assistance on the employees' health.

Figure 2-8 shows the two directions of causality. Self-selection (1) implies that sick employees are more likely to use the on-site service. They get treated according to standard medical practices, on average improving their conditions (2). The causal effect that the CHRO expected to find was given by (2), but unfortunately the selection effect was strong enough to counter the positive impact that the company's doctors have.

health status and selection effects

Figure 2-8. Self-selection explains why treated employees show worse health conditions than those that have not attended the on-site medical service

CAN INVESTING IN INFRASTRUCTURE INCREASE CUSTOMER CHURN?

Another example from a capital intensive industry like telecommunications will show the dangers of decision-making in the presence of selection effects. Telcos have very large capital expenditures (CAPEX) since they need to constantly invest in building and maintaining a network to provide high quality communication services to their customers. Suppose that our Chief Financial Officer must decide where to focus our investing efforts

during the next quarter. After looking at the data they plot churn rates in cities with no CAPEX last year and those where there was positive investment (Figure 2-9).

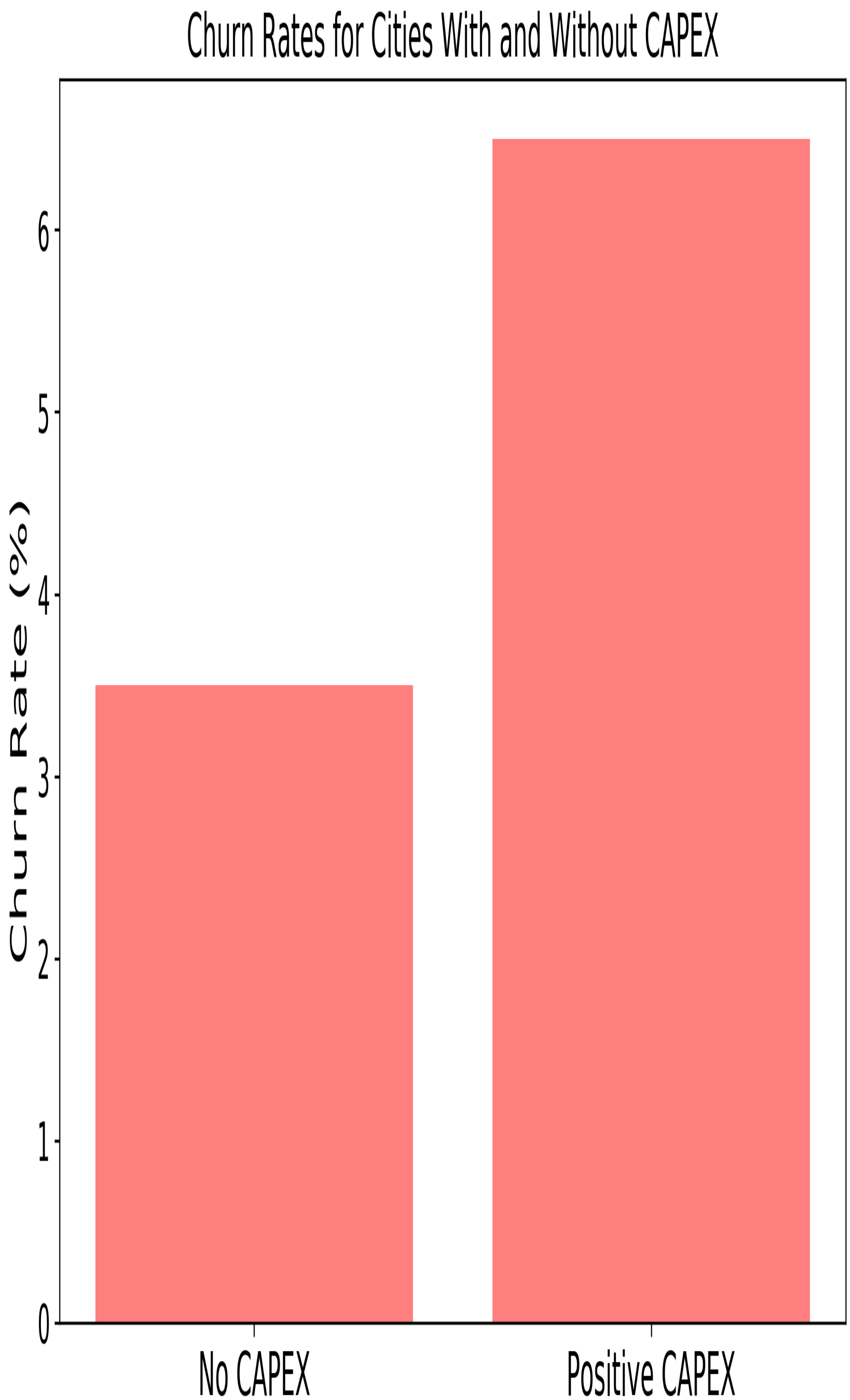

Churn Rates for Cities With and Without CAPEX
Churn Rate (%)
0
1
2
3
4
5
6
No CAPEX
Positive CAPEX

Figure 2-9. Average churn rates in cities with and without CAPEX during the previous year

The results are both surprising and frustrating: it appears that CAPEX has had unintended consequences as churn is higher in cities where they invested last year relative to those without CAPEX. Could it be, maybe, that competitors reacted strategically investing even more heavily in those cities and capturing an increasing market share? That is likely but the most plausible explanation are selection effects. Last year they focused the investment efforts in cities that were lagging in terms of customer churn and satisfaction. The result is exactly like the one in the medical condition: the patients (the cities) still haven't recovered fully so the data still shows the disadvantaged initial conditions.

MARKETING MIX AND CHANNEL OPTIMIZATION: THE CASE OF ONLINE ADVERTISING

Our Chief Marketing Officer asks us to estimate the revenue impact of advertising spend on different channels, with a specific focus on our digital channels. After devoting considerable effort getting the data, we find some very pleasant news: the online advertising Return on Investment (ROI) was 12% for the year, the second consecutive year with double digits. However, members from the data science team raised concerns about us overestimating the real impact. Their logic was as follows (Figure 2-10).

Our retargeting partner usually waits for internet users to visit our webpage. When that happens, they place a cookie so that we can track their online behavior. Some time later, an ad is served on a publisher's website with the hope of converting this lead. Some of

these customers end up buying our products on our website, so it seems that our advertising investment has done wonders for us.

selection effects for retargeting

Figure 2-10. Selection effects explains high digital marketing ROI

The problem here is conceptual, though: ideally, advertising should be done to convert a user who was *not* planning to buy from us into a buyer. But those who end up buying from us had already self-selected themselves by visiting our webpage, hence showing interest on our products. *Had we not placed the ad, would they have purchased anyway?* If the answer to this counterfactual is affirmative, then there is a case that the ROI might be overestimated; it could even be negative! To get a reasonable estimate we must get rid of our customers' self-selection.

Some difficulties in estimating causal effects

Estimating the causal impact on outcome Y of pulling a lever $X \implies Y$ is paramount since we are trying to engineer optimal decision-making. The analogy is not an accident: like the engineer who has to understand the laws of physics to build skyscrapers, bridges, cars or planes, the analytical leaders of today must have some level of understanding of the causal laws mediating our own actions and the consequences to make the best possible decisions. And this is something that humans must do; AI will help us later in the decision-making process, but we must first overcome the causal hurdles.

PROBLEM 1: WE CAN'T OBSERVE COUNTERFACTUALS

As discussed in the previous sections, there are several problems that make our identification of causal effects much harder. The first one is that we only observe the facts so we must imagine alternative *counterfactual* scenarios. In each of the previous examples, we knew that direct causal interpretation was problematic since we were able to imagine alternative universes with different outcomes. It is an understatement that one of the most important skills analytical thinkers must develop is to question the initial interpretation given to empirical results, and to come up with counterfactual alternatives to be tested. Would the consequences be different, had we pulled different levers, or the same levers but under different conditions?

Let's stop briefly to discuss what this question entails. Suppose we want to increase lead conversion in our telemarketing campaigns. Tom, a junior analyst who took one class in college on Freudian psychoanalysis suggests that female call center representatives should have higher conversion rates, so they decide to make all outbound calls for a day with their very capable group of women representatives. The next day they meet to review the results: lead conversion went from the normal 5% to an outstanding 8.3%. It appears that Freud was right, or better, that Tom's decision to take the class had finally proven correct. Or does it?

To get the right answer, we need to imagine a customer receiving one call from the female representative in one universe, and the *exact same* call from a male representative in a parallel universe (Figure 2-11). Exact customer, exact timing, exact mood and exact message; everything is the same in the two scenarios: we only change the tone of voice from that of a male to a female. Needless to say, putting in

practice such counterfactual sounds impossible. Later in this chapter we will describe how we can simulate these impossible counterfactuals through well-designed randomized experiments or A/B tests.

call center counterfactual

Figure 2-11. Counterfactual analysis of lead conversion rates in a call center

PROBLEM 2: HETEROGENEITY

A second problem is *heterogeneity*. Humans are intrinsically different, each and every one the product of both our genetic makeup and lifetime experiences, creating unique world visions and behaviors. Our task is not only to estimate how behavior changes when we choose to pull a specific lever — the causal effect — but we must also take care of the fact that different customers react differently. An influencer recommending our product will have different effects on you and me: I may now be willing to try it while you may choose to remain loyal to your favorite brand. How do we even measure heterogenous effects?

Figure 2-12 shows the famous bell curve, the normal distribution, the darling of statistical aficionados. I'm using it here to represent the natural variation we may encounter when analyzing our customers' response when our influencer recommends our product. Some of his followers, like me, will accept the cue and react positively — represented as an action right of the vertical dashed line, the average response across all followers, followers' followers and so on. Some will have no reaction whatsoever, and some may even react negatively; that's the beauty of human behavior, we sometimes get

the full spectrum of possible actions and reactions. The shape of the distribution has important implications, and in reality, our responses may not be as symmetric; we may have longer left or right tails and reactions may be skewed towards the positive or the negative. The important thing here is that people react differently, making things even more difficult for us when we try to estimate a causal effect.

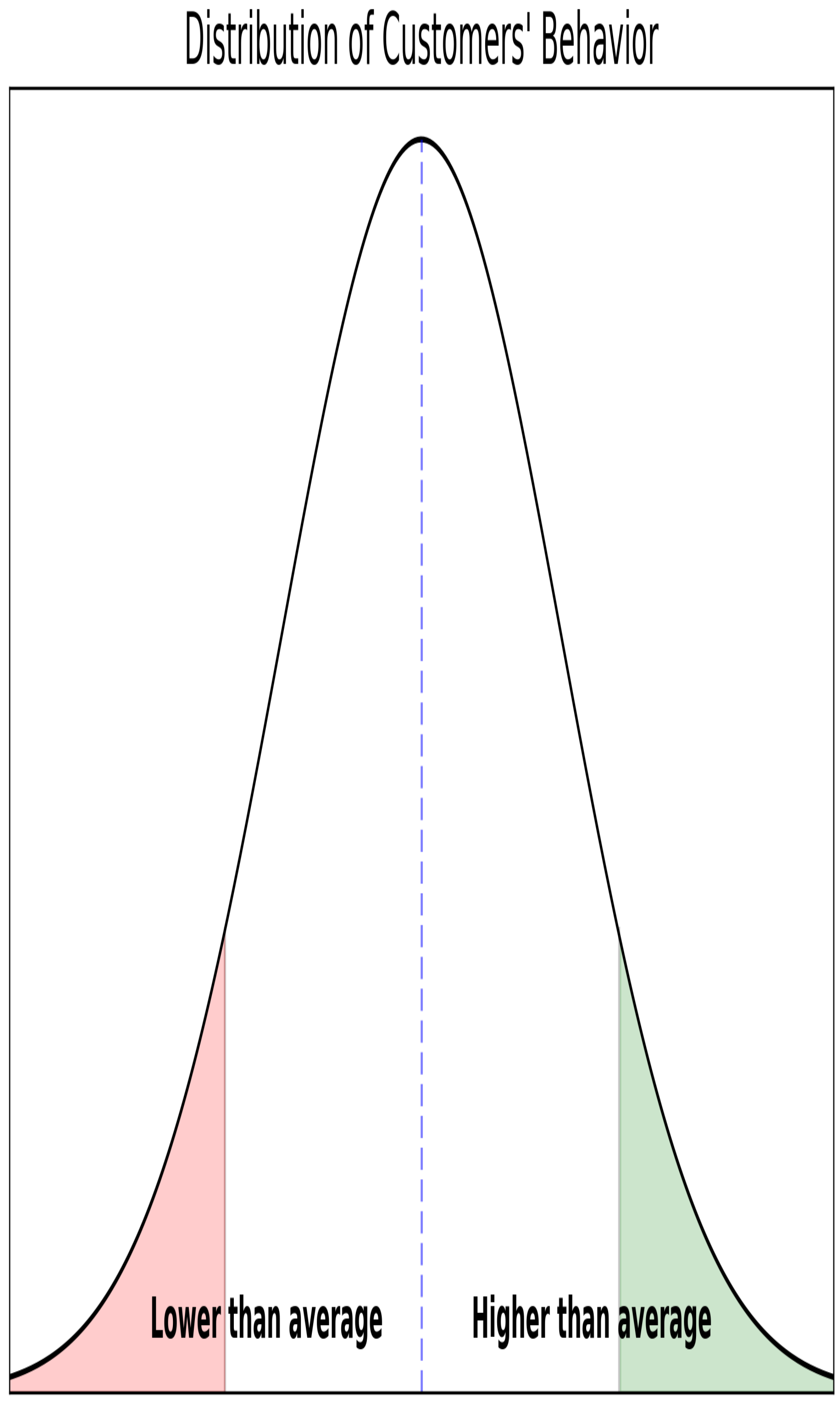

Distribution of Customers' Behavior
Lower than average
Higher than average

Figure 2-12. A normal distribution as a way to think about customer heterogeneity

The way we usually deal with heterogeneity is by dispensing of it by estimating a unique response, usually given by the average or the mean (the vertical line in Figure 2-12). The mean, however, is overly sensitive to extreme observations, so we may sometimes replace it by the median, having the property that 50% of responses are lower (to the left) and 50% higher (to the right); with bell-shaped distributions the mean and the median are conveniently the same.

PROBLEM 3: SELECTION EFFECTS

One final problem covered in detail in the previous section is the prevalence of selection effects. This usually arise because we we choose the customer segments we want to act upon, or customers self-select themselves, or both. An important result in causal inference is that if we wish to estimate the causal effect from a treatment by comparing the average outcomes of two groups we need to find a way to eliminate selection bias.[8]

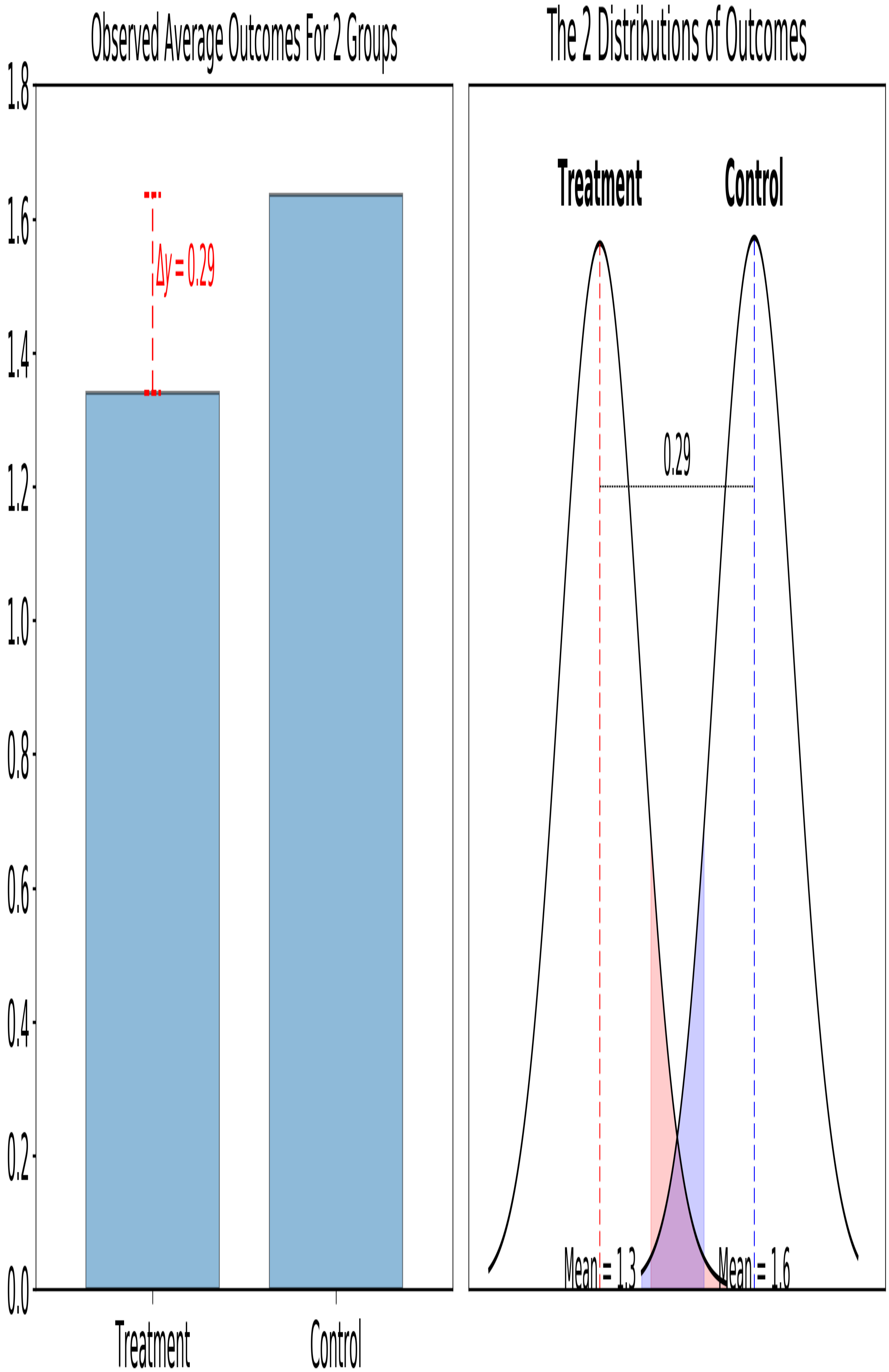
SELECTION BIAS AND CAUSAL EFFECTS

Because of selection bias we may over or under estimate a causal effect when we just take the difference in average outcomes across treated and control groups.

$$\text{Observed Difference in Means} = \text{Causal Effect} + \text{Selection Bias}$$

It is standard practice to plot average outcomes as in the left panel of Figure 2-13. In this case, the outcome for the control is 0.29 units

(say hundreds of dollars) higher than for those exposed to our action or lever. This number corresponds to the left-hand side of the previous equation. The right panel shows the corresponding distributions of outcomes. Using the mean to calculate differences is standard practice, but it is useful to remember that there are a full spectrum of responses, in some cases with a clear overlap between the two groups: the shaded areas show responses from customers in the two groups that are indistinguishable from each other.



Observed Average Outcomes For 2 Groups
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Δy = 0.29
Treatment
Control
The 2 Distributions of Outcomes
Treatment
Control
0.29
Mean = 1.3
Mean = 1.6

Figure 2-13. Left panel plots the observed differences in average outcomes for treatment and control groups. Right panel shows the actual distributions of outcomes.

In any case, the difference in observed outcomes (left-hand side) is not enough for us since we already know that it is potentially biased by selection effects; since our interest is in estimating the causal effect we must therefore device a method to cancel this pervasive effect.

Statisticians and econometricians, not to mention philosophers and scientists, have been thinking about this problem for centuries now. Since it is physically impossible to get an exact copy of each of our customers, is there a way to assign our treatments and circumvent the selection bias? It was Ronald A. Fisher, the famous 20th century statistician and scientist who put on firm grounds the method of experimentation, the most prevalent among practitioners when we want to estimate causal effects. The idea is simple enough to describe without making use of technical jargon.

A primer on A/B testing

While we may not be able to get exact copies of our customers, we may still be able to simulate such copying device using *randomization*, that is, by randomly assigning customers to two groups: those who receive the treatment and those who don't (the control group). Note that the choice of two groups is done for ease of exposition, as the methodology applies for more than two treatments.

We know that customers in each group are different, but by correctly using a random assignment we dispose of any selection bias: our

customers were selected by chance, and chance is thought to be unbiased. In practical terms before our customers get a call from our call center representatives, customers in the female treatment are, *on average*, ex-ante the same as those in the male treatment. Luckily, we can always check if random assignment created groups that are, on average, ex-ante equal.

RISKS WHEN RUNNING RANDOMIZED TRIALS

We have noted that randomization is unbiased in the sense that the result of a random draw is obtained by chance. In practice we simulate pseudorandom numbers that have the look-and-feel of a random outcome, but are in fact computed with a deterministic algorithm. For instance, in Excel, you can use the `=RAND()` funtion to simulate a pseudo-random draw from a uniform distribution.

It is important to remember, however, that using randomization does not necessarily eliminate selection bias. For example, even though the probability of happening may be *extremely* low, by pure chance, we may end up with a group of male customers on the male representative group and female customers on the control (female reps) group, so our random assignment ended up selecting by gender, potentially biasing our results. It's a good practice to check if random assignment passes the ex-post test by checking differences in means on observable variables.

Last but not least, there may be ethical concerns since in practice we are potentially affecting the outcomes of one group of customers. One should always checklist any ethical considerations we might have before running an experiment.

You may be wondering what it means for two groups to be indistinguishable before making the random assignment (ex-ante equal). Think about how you would tell two people apart: start

checking, one by one, each and every observable characteristic and see if they match. If there's something where they look different then they are not indistinguishable. We do the same for two different groups of people: list all observable characteristics and check if their group averages are the same, after taking into account the natural random variation. For instance, if customers in the female and male representative groups are on average 23 and 42 years old respectively, we should repeat the randomization to make them indistinguishable in terms of all observables, including age.

A/B TESTING IN PRACTICE

In the industry, the process of randomizing to assign different treatments is called A/B testing. The name comes from the idea that we want to test an alternative B to our default action A, the one we commonly use. As opposed to many of the techniques in the machine learning toolbox, A/B testing can be performed by anyone without a strong technical background. We may need, however, to guarantee that our test satisfies a couple of technical statistical properties, but these are relatively easy to understand and put in practice. The process usually goes as follows:

1. Select an actionable hypothesis you want to test: for example, call center female representatives have a higher conversion rate than men. This is a crisp hypothesis that is falsifiable.
2. Choose a relevant and measurable KPI to quantify the results from the test; in the example we choose conversion rates as our outcome. If average conversion for female reps *isn't* "significantly larger" than that for men, we can't conclude

that the treatment worked, so we keep running the business as usual. It is standard practice to use the concept of statistical significance to have a precise definition of what *larger* means.

3. Select the number of customers that will be participating in the test: this is the first technical property that must be carefully selected and will be discussed below.
4. Randomly assign the customers to both groups and check that randomization produced groups that satisfy the ex-ante indistinguishable property.
5. After the test is performed, measure the difference in average outcomes. We should take care of the rather technical detail of whether a difference is generated by pure chance or not (statistical significance).

If randomization was done correctly, we have eliminated the selection bias, and the difference in average outcomes provides an estimate of the causal effect.

UNDERSTANDING POWER AND SIZE CALCULATIONS

Step 3, selecting the number of customers, is what practioners call *power and size calculations*, and unfortunately there are key trade-offs we must face. Recall that one common property of statistical estimation is that the larger the sample size the lower the uncertainty we have about our estimate. We can always estimate the average outcome for groups of 5, 10 or 1000 customers assigned to the B group, but our estimate will be more precise for the latter than for the former. From a strictly statistical point of view, we prefer having large experiments or tests.

From a business perspective, however, testing with large groups may not be desirable. First, our assignment must be respected until the test comes to an end, so there is the opportunity cost of trying other potentially more profitable treatments, or even our control or base scenario. Because of this, it is not uncommon that the business stakeholders want to finish the test as quickly as possible. In our call center example, it could very much have been the case that conversion rates were *lower* with the group of female reps, so during a full day we operated suboptimally which may take an important toll on the business (and our colleagues' bonuses). We simply can't know at the outset (but a well designed experiment should include some type of analysis of this happening).

Because of this trade-off we usually select the *minimum* number of customers that satisfies two statistical properties: experiments should have the right statistical size and power so that we can conclude with enough confidence if it was a success or not. This takes us to the topic of false positives and false negatives.

FALSE POSITIVES AND FALSE NEGATIVES

In our call center example, suppose that contrary to Tom's assumption, male and female representatives have the exact same conversion efficiency. In an ideal scenario we would find no difference between the two cases, but in practice this is *always* non-zero, even if small. How do we know if the difference in average outcomes is due to random noise or if it is showing a real, but possibly small difference? Here's where statistics enter the story.

There is a *false positive* when we mistakenly conclude that there is a difference in average outcomes across groups and we therefore conclude that the treatment had an effect. We choose the *size of the test* to minimize the probability of this happening.

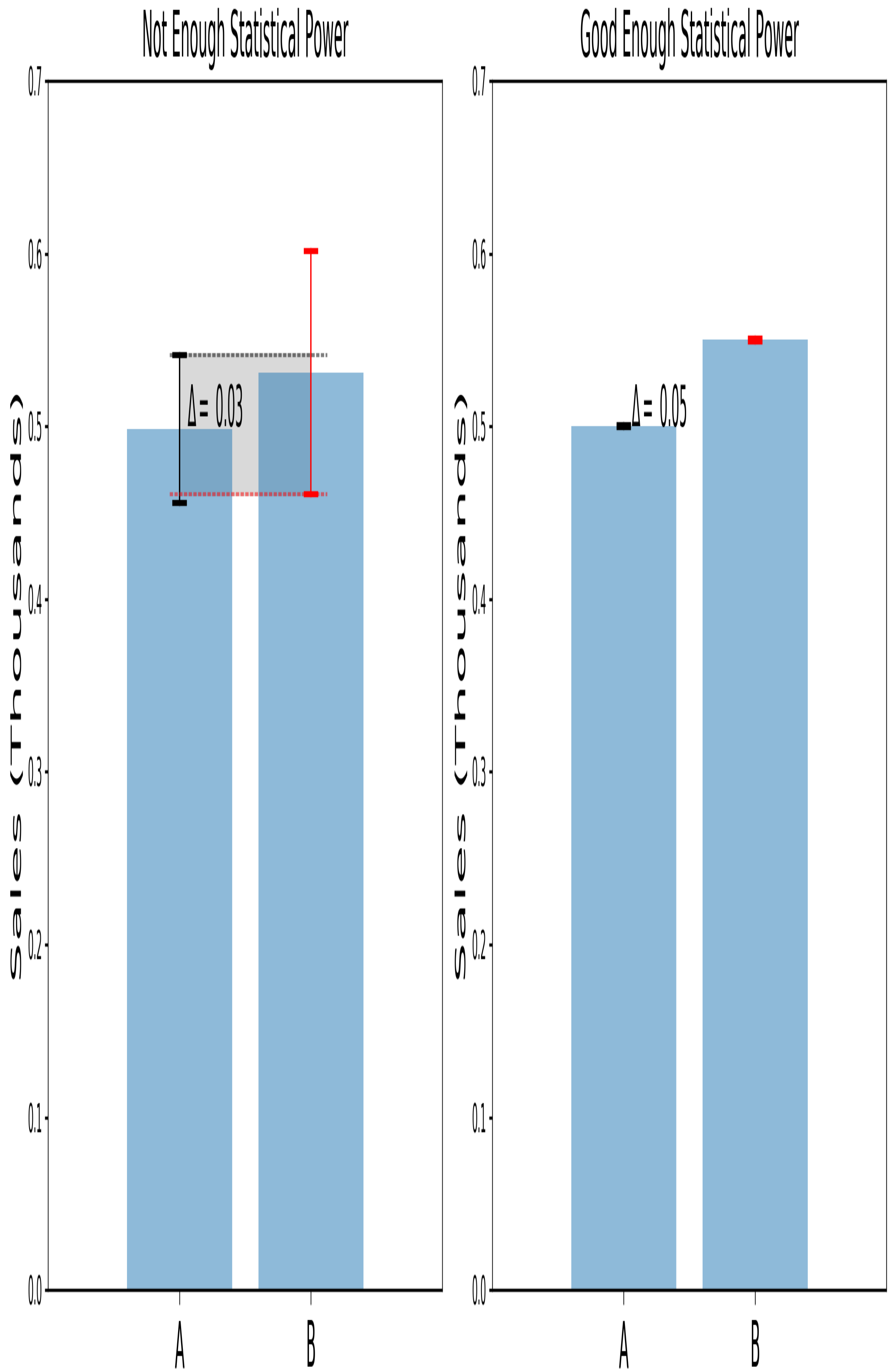
On the other hand, it could be that the treatment actually worked but we may not be able to detect the effect with enough confidence. This usually happens when the number of participants in the experiment is relatively small. The result is that we end up with an *underpowered* test. In our call center example, we may falsely conclude that representatives' productivity is the same across genders when indeed one has higher conversion rates.

STATISTICAL SIZE AND POWER

Somewhat loosely speaking, the *size* of a statistical test is the probability of encountering a false positive. The *power* of the test is the probability of correctly finding a difference between treatment and control.

The left panel in Figure 2-14 shows the case of an underpowered test. The alternative B treatment creates 30 additional sales, but because of the small sample sizes, this difference is estimated with insufficient precision (as seen by the wide and overlapping confidence intervals represented by the vertical lines).

The right panel shows the case where the real difference is close to 50 extra sales, and we were able to precisely estimate the averages and their differences (since confidence intervals are so small that they don't even look like intervals).



Not Enough Statistical Power
Good Enough Statistical Power
Sales (Thousands)
Sales (Thousands)
Δ= 0.03
Δ= 0.05
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
A
B

Figure 2-14. Left panel shows the result of an underpowered test: there is a difference in the average outcomes for the treated and untreated but the small sample sizes for each group cannot estimate this effect with enough precision. Right panel shows the ideal result where there is a difference and we can correctly conclude this is the case.

Let’s briefly talk about the costs of false positives and false negatives in the context of A/B testing. For this, recall what we wanted to achieve with the experiment to begin with: we are currently pulling a lever and want to know if an alternative is superior for a given metric that impacts our business. As such, there are two possible outcomes: we either continue pulling our A lever, or we substitute it with the B alternative. In the case of a false positive, the outcome is making a subpar substitution. Similarly, with a false negative we mistakenly continue pulling the A lever, which also impacts our results. In this sense both are kind of symmetric (in both cases we have an uncertain long-term impact), but it is not uncommon to treat them asymmetrically, by setting the probability of a false positive at 5% or 10% (size), and the probability of a false negative at 20% (one minus the power).

There is however the opportunity cost of designing and running the experiment, so we’d better run it assuming the best-case scenario that the alternative has an effect. That’s why most practitioners tend to fix the size of a test and find the minimum sample size that allows us to detect some minimum effect.

SELECTING THE SAMPLE SIZE

In tests where we only compare two alternatives, it is common to encounter the following relationship between the variables of interest:

$$MDE = (t_\alpha + t_{1-\beta})\sqrt{\frac{\mathrm{Var}(\mathrm{Outcome})}{NP(1-P)}}$$

Here t_k is critical value to reject a hypothesis with probability k according to a t distribution, α and $1-\beta$ are the size and power of test (that you can replace to calculate corresponding critical values), MDE the minimum detectable effect of the experiment, N the number of customers in the test, P is the fraction assigned to the treatment group, and $\mathrm{Var}(\mathrm{Outcome})$ is the variance of the outcome metric you're using to decide if the test is successful or not.

As you can see from this formula, for a given MDE the larger the variance of your outcome the larger sample you will need. This is standard in A/B testing: noisy metrics will require larger experiments. Also, remember that our objective is to have a small enough MDE that allows us to detect incremental changes caused by the treatment, putting even more pressure on the size of the experiment.

The next snippet shows how to calculate the sample size for your experiment with Python.

```
# Example: calculating the sample size for an A/B test
from scipy import stats
def calculate_sample_size(var_outcome, size, power, MDE):
    '''
    Function to calculate the sample size for an A/B test
    MDE = (t_alpha +
t_oneminusbeta)*np.sqrt(var_outcome/(N*P*(1-P)))
    df: degrees of freedom when estimating the variance
    of the outcome (if sample size is large df is also
    large so I artificially set it at 1000)
    '''
    df = 1000
    t_alpha = stats.t.ppf(1-size, df)
    t_oneminusbeta = stats.t.ppf(power, df)
    # same number of customers in treatment and control group
    P = 0.5
    # solve for the minimum sample size
    N = ((t_alpha + t_oneminusbeta)**2 * var_outcome)/(MDE**2
```

```
* P * (1-P))
    return N
```

```python
# parameters for the example below
var_y = 4500
size = 0.05
power = 0.8
MDE = 10
sample_size_for_experiment = calculate_sample_size(var_y,
size, power, MDE)
print('We need at least {0} customers in experiment'.format(
   np.around(sample_size_for_experiment),decimals=0))
```

In practice, we start by setting the power and size of the test and then choose an MDE. One way to think about it is that it is the minimum change on our outcome metric that makes the experiment worthwhile from a business standpoint. We can finally reverse engineer the sample size we need from the formula.

To see this in practice, suppose that we want to run an A/B test to see if we can increase our average customer spend or ticket by way of a price discount. In this price elasticity experiment, the treatment group will get the new lower price, and the control will keep paying the regular price. Because of those very high spend customers, the variance in monthly spend is 4500 (standard deviation is about $67). As a benchmark we choose standard values for size and power (5% and 80%). Finally, our business stakeholders convince us that from their perspective it only makes sense to try the new alternative if we find a minimum effect (MDE) of 10 dollars (or 15% of one standard deviation). We run our size calculator and find that we need at least 1115 participants in the experiment. Since our contact rate is around 2%, we should send emails to around 1115/0.02 = 55.2K customers.

Uncertainty

We have now talked about each of the stages in the decomposition: starting with the business we reverse engineer the actions or levers that impact our objective and corresponding KPIs, mediated by some consequences. However, since decisions are made under uncertainty, this mapping from actions to consequences is not known to us at the time of the decision. But by now we already know that uncertainty is not our enemy and that we can embrace it thanks to the advances in predictive power of AI.

But why do we have uncertainty? Let us first discuss what this uncertainty is not, and then we can talk about what it is. Think about flipping a coin. We know that with a balanced coin the chances it falls on heads are 50% and that the final outcome cannot be fully anticipated from the outset. Since we have played heads and tails for most of our lifetimes this is an example of randomness that is quite close and natural to us.

This is not, however, the type of uncertainty we have when we are making decisions, and that is good news for us. The fact that ours is not pure randomness allows us to use powerful predictive algorithms, combined with our knowledge of the problem to select input variables or features to create a prediction. With pure randomness, the best thing we can do is learn or model the distribution of outcomes and derive some theoretical properties that allows us to make smart choices or predictions.[9]

The four main sources of uncertainty *when we make decisions* are our need to simplify, heterogeneity, complex and strategic behavior arising from social interactions and pure ignorance about the phenomenon, each of which will be described in turn. Note that as analytical thinkers we should always know where uncertainty comes from, but it is not uncommon that we end up being taken by surprise.

Uncertainty from simplification

Albert Einstein has many great quotes, but one my favorites is “everything should be made as simple as possible. But not simpler.” In the same vein, statistician George Box famously said that “all models are wrong, but some are useful”. Models are simplifications, metaphors that help us understand the workings of the highly complex world we live in.

I cannot exaggerate enough the importance that learning to simplify has for the modern analytical thinker. We will have enough time in [Link to Come] to exercise our analytical muscle through some well-known techniques, but we should now discuss the toll that simplification has.

As analytical thinkers and decision-makers we constantly face the trade-off between getting a good-enough answer or devoting more time to develop a more realistic picture of the problem at hand. We must decide how much uncertainty we’re comfortable with and how much we are willing to accept, in order to get a timely solution. But this calibration takes practice, as Einstein succinctly puts it in the first quote.

One clear example of the powers and dangers of simplification are maps. Figure 2-15 shows a section of the official Transit for London (Tfl) London's tube map on the left and a more realistic version on the right also by the transportation authority.[10] With the objective of making our transportation decisions fast an easy, a map trades-off realism for ease-of-use. As users of the map, we now face uncertainty about the geography, distances, angles and even the existence of possible relevant venues such as parks or museums. But to a first approximation we feel comfortable with this choice of granularity since our first objective is being able to get from our origin to a destination. We can later take care of the remaining parts of the problem.

maps are simplifications, models

Figure 2-15. Sections of the London underground maps. Left panel corresponds to the official tube map. Right panel shows a more realistic version of the same section.

This last point takes me to another related issue: one common simplification technique is to divide a complex problem into simpler subproblems that can each be tackled independently; something that computer scientists call the *Divide and Conquer* technique. When each of these subproblems gives rise to some uncertainty, nothing guarantees that the resulting uncertainty after aggregation becomes more tractable (unless we impose some simplifying assumptions to start with).

The moral of this story is that we should always remember that simplifying a problem usually brings additional uncertainty to the table. As Box, the statistician, complemented "(…) the approximate nature of the model must always be born in mind".[11]

Uncertainty from heterogeneity

One important source of uncertainty when making business decisions comes from the fact that our customers react in very different ways. This large variety of behaviors, tastes and responses can be modelled with the use of distributions since that's how we generally deal with uncertainty (recall Figure 2-12). By doing so we can dispense of the nitty-gritty details of how and why outcomes are so diverse, and just focus on how uncertainty affects our final outcomes. This modelling approach is quite handy and forces us to know some basic properties about distributions.

Take the case of the *uniform distribution*. While it is most commonly assumed for simplification purposes it can also be used if there's no reason to believe that outcomes will tend to accumulate. To give a concrete example, think about how people waiting for a train during peak hours end up being distributed across the platform. If their goal is to find a sit and enter the train as quickly as possible it is most natural that they end up distributing uniformly.

We have already encountered the *normal distribution* which is quite pervasive in the sciences. It is sometimes used for simplification purposes as it has some highly desirable properties (linearity, additivity) but it also arises naturally in many settings. For instance, we may appeal to a version of the Central Limit Theorem, that states that under certain conditions, the distribution of averages or sums of numbers end up being close enough to a normal.[12]

Other commonly used distributions are power-law (or heavy tailed) distributions, that, contrary to the Gaussian distribution, have longer tails.[13] For instance, when modelling the reach or just the number of followers that your influencer has, we may resort to a power-law distribution, but there are many other examples where these distributions arise most naturally.[14]

Figure 2-16 shows the results of drawing one million observations from uniform, normal and power-law distributions.

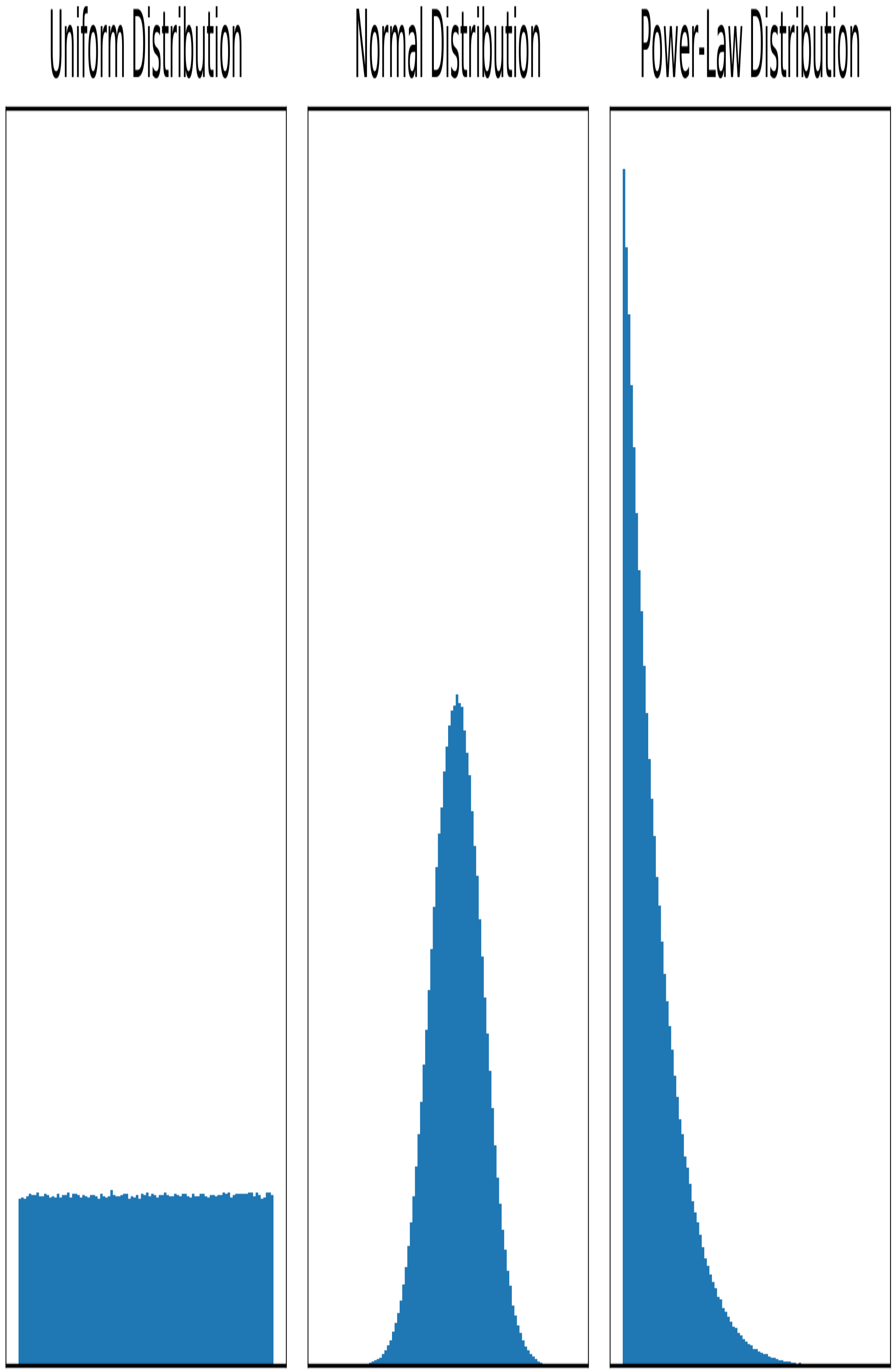

Uniform Distribution
Normal Distribution
Power-Law Distribution

Figure 2-16. Histograms for the results of drawing one million observations from a uniform (left), normal (center) and power-law (right) distribution

Uncertainty from social interactions

Another source of uncertainty arises from the simple fact that we are social animals continuously interacting with each other. While this has been taking place for hundreds of thousands of years, the explotion of interactions with modern social networks has made it even more salient and prevalent.

A first source of uncertainty comes from the strategic nature of our interactions with our customers and workforce, just to give two examples. With customer retention offers, for instance, it is not uncommon that they understand our workings and motivations and end up gaming our system. Similarly, compensation schemes are quite commonly gamed by our sales executives giving rise to somewhat unexpected results like delayed sales when goals have been or are unlikely to be reached.

But uncertainty may also arise from nonstrategic and very simple decision rules. One well-studied example is John Conway's Game of Life that evolves in a two-dimensional grid such as the one depicted on Figure 2-17.[15] At any given time, each colored pixel can only interact with its immediate neighbors thereby creating three possible outcomes: it lives, dies or multiplies. There are only three simple rules of interaction and depending on the initial conditions you can get completely different outcomes that appear to be random to any observer.

different distributions

Figure 2-17. John Conway's Game of Life. A plethora of aggregate phenomena arises from three simple rules of how each cell or pixel interact with its neighbors.

You may wonder if this is something worth your time and attention or if it's just an intellectual curiosity. As a starter it should serve as a cautionary tale that even simple rules of behavior can create complex outcomes so we don't really need sophisticated consumers trying to game our systems. But social scientists have also been using these tools to make sense of human behavior so, at the minimum, they ought to be useful for us when making decisions in our businesses.

Uncertainty from ignorance

The last source of uncertainty is pure ignorance as many times we simply don't know what will happen when a lever is pulled and we are also unaware of the the likely distribution of outcomes. In this cases it is not uncommon to start by assuming that outcomes follow a uniform or a normal distribution, later improving our knowledge by some sort of experimentation.

A company's ability to scale testing at the organizational level can create a rich knowledge base to innovate and create value in the medium-to-long term. But there is always a trade-off: we may need to sacrifice short-term profits for medium term value and market leadership. That's why we need a new brand of analytical decision makers in our organizations.

Key takeaways

- **Analytical thinking**: is the ability to identify and translate business questions into prescriptive solutions.
- **Value is created by making decisions**: we create value for our companies by making better decisions. Prediction is only one input necessary in our decision-making process.
- **Stages in the analysis of decisions**: there are generally three stages when we analyze a decision: we first gather, understand and interpret the facts (descriptive stage). We then may wish to predict the outcomes of interest. Finally, we choose the levers to pull to make the best possible outcome (prescriptive stage).
- **Prescriptive decision-making**: decision-making is the act of choosing among competing actions to attain these objectives. *Data-driven* decision-making is acting upon evidence to assess alternative courses of action. *Prescriptive* decision-making is the science of choosing the action that produces the best results for us.
- **Anatomy of a decision**: we choose an action that may have one or several consequences that impact our business outcomes. Since generally we don’t know which consequence will result, this choice is made under conditions of uncertainty. The link between actions and consequences is mediated by causality.
- **Start with the business**: since our aim is to find the best course of action we’d better be optimizing for the right question. So start with the business. One side benefit is that we usually enlarge the menu of levers available to us.
- **As important as asking the right question is the selection of the metrics to measure the impact of our decision-making**: Many data science projects fail not because of the logic used but because we used the wrong set of metrics to

measure the impact for our business question. Good metrics should be relevant and measurable.

- **Estimating causal effects has several important difficulties**: selection biases abound, so directly estimating the causal effect of a lever is generally not possible. We also need to master the use of counterfactual thinking and dealing with heterogenous effects.

Further Reading

Almost every book on data science or big data describes the distinction between descriptive, predictive and prescriptive analysis. You may check Thomas Davenport’s now classic *Competing on Analytics* (or any of the sequels) or Bill Schmarzo’s *Big Data: Understanding How Data Powers Big Business* (or any of the prequels and sequels).

The anatomy of decisions used here follows that literature and is quite standard. We will come back to this topic in [Link to Come] where I will provide enough references.

My favorite treatments of causality can be found in the books by Joshua Angrist and Jörn-Steffen Pischke *Mostly Harmless Econometrics* and the most recent *Mastering ‘Metrics’: The Path from Cause to Effect*. If you are interested you can find there the mathematical derivation of the equality between difference in observed outcomes and causal effects plus selection bias. They also present alternative methods to identify causality from *observational*

data, that is, from data that was not obtained through a well-designed test.

A substantially different approach to causal reasoning can be found in Judea Pearl's and Dana Mackenzie's *The Book of Why. The new science of cause and effect*. Scott Cunningham's *Causal Inference: the mixtape* provides a great bridge between the two approaches, focusing mostly on the first literature (econometrics of causal inference) but devoting a chapter and several passages to Pearl's approach using causal graphs and diagrams. At the time of the writing of this book it's also free to download from *https://www.scunning.com/cunningham_mixtape.pdf*.

There are many treatments of A/B testing, starting with Dan Siroker's and Pete Koomen's *A/B Testing: The Most Powerful Way to Turn Clicks into Customers*. Peter Bruce's and Andrew Bruce's *Practical Statistics for Data Scientists* from O'Reilly Media provides an accessible introduction to statistical foundations, including power and size calculations. Carl Andersen's *Creating a Data-Driven Organization*, also from O'Reilly, briefly discusses some best practices in A/B testing emphasizing its role on data- and analytics-driven organizations. Ron Kohavi (previously at Microsoft and now at Airbnb) has been forcefully advancing the use of experimentation in the industry. You can find some great material in his (and others') ExP Experimentation Platform (*https://exp-platform.com/*), including an online version of a book coauthored with Diane Tang and Ya Xu *Advanced Topics in Experimentation* (*https://exp-platform.com/advanced-topics-in-online-experiments/*).

My discussion of uncertainty follows many ideas in Scott E. Page's *The Model Thinker: What you need to know to make data work for you*. This is a great place to start thinking about simplification and modelling, and provides many examples where distinct distributions, complex behavior and network effects appear in real life.

---

1 *https://www.theguardian.com/society/2014/sep/22/cancer-late-diagnosis-half-patients*

2 *https://www.nytimes.com/2019/05/20/health/cancer-artificial-intelligence-ct-scans.html*

3 We will talk about designing experiments or A/B tests later in this chapter.

4 *https://www.investopedia.com/updates/enron-scandal-summary/*

5 Not that it couldn't be used like that, of course.

6 Sources: GDP data comes from *https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2018*. CO2 emissions from *https://www.co2.earth/images/data/2100-projections_climate-scoreboard_20* 15-1027.xlsx. Divorce rates from *https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/birthsdeathsandmarriages/divorce/datasets/divorcesinenglandandwales/2014/divorceta* bles2014.xls.

7 This use case is motivated by the opening example in the book Mostly Harmless Econometrics. See the references at the end of the chapter.

8 Hereafter I will use the term "treated" or "those who receive a treatment" refering to those customers that are exposed to our action or lever. This jargon is common in the statistical analysis of experiments and it is no coincidence that we have already encountered it discussing the case of our employees health status, as it was first used in the analysis of medical trials.

9 In the coin tossing example, for instance, after observing the outcomes we may end up modelling the distribution as Bernoulli trials, and predict a theoretically derived expected value (number for trials times the estimated probability of heads, say).

10 *https://www.timeout.com/london/blog/tfl-has-secretly-made-a-geographically-accurate-tube-map-091515*

11 *https://en.wikipedia.org/wiki/All_models_are_wrong*

12 *https://en.wikipedia.org/wiki/Central_limit_theorem*

13 The Normal distribution accumulates 99% of the possible outcomes within 2.57 standard deviations from the mean and 99.9% within almost 3.3 standard deviations.

14 Other examples and applications of power-law distributions in business can be found in *http://www.hermanaguinis.com/JBV2015.pdf*

15 You can "play" the game yourself at *https://playgameoflife.com/* and marvel at the rich diversity of outcomes that can be generated by simple deterministic rules. See also *https://en.wikipedia.org/wiki/Conway%27s_Game_of_Life*

Chapter 3. Learning to ask good business questions

> **A NOTE FOR EARLY RELEASE READERS**
>
> With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.
>
> This will be the 3rd chapter of the final book. Please note that the GitHub repo will be made active later on.
>
> If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the author at analyticalthinkingbook@gmail.com.

Chapter 2 provided a quick overview of the general framework we'll be developing in the upcoming chapters. Since our ultimate objective is to translate business problems into prescriptive solutions, we should better start learning how to ask the *right* questions. I hope it shouldn't come as a suprise that learning to frame the questions can have an impact comparable in magnitude to adopting the techniques that will follow.

We also introduced a very simple technique that I've found quite useful to understand what we really want to accomplish: *the sequence of why questions*.[1]. It starts by questioning what you think you are trying to accomplish, move up one level or stop when you are convinced that the business objective is in fact just right. In our voyage to find prescriptive solutions it is of outmost importance to

guarantee that we are tackling the right objectives. One nice byproduct that will be quite handy in Chapter 4 is that it usually enlarges the set of possible actions or levers we have. This is usually the case when we start by questioning an action and the procedure ends up taking us to the metrics we really want to affect. It is almost natural, then, to question if there are other actions that can be used to affect the same objective.

In this chapter we will delve a bit more into some of the better practices when asking good business questions, the difference between descriptive, predictive and prescriptive questions, and we will finish with some examples from common use cases. These were selected from my own experience, from other use cases I have discussed with students in class and colleagues, and because they are good to present and understand the methods. But first we should better understand where business questions come from (Figure 3-1).

start from the results

Figure 3-1. Start with the business

From business objectives to business questions

Traditionally, companies have been organized by clearly separating each area's responsibilities or objectives (Figure 3-2). In the past few years, however, the *agile* movement has helped many companies to break the functional silos and organize into cross-functional teams. The outcome being that each team has very clearly delimited business objectives and metrics to pursue.[2]

functional organization

Figure 3-2. An example of a company organized by functional divisions. From left to rigth, the acronyms correspond to Chief Officers in Finance, Marketing, Human Resources, Data, Information, Analytics, Sales and Operations, respectively. There are many more such acronyms.

This is good news for us, since our business objectives are usually well-defined and, supposedly, relatively easy to evaluate through well-defined KPIs. It is our task, however, to ask the necessary business questions to achieve these objectives. In general, for any business objective there are multiple business questions that can be asked, and for each of these there are different actions or levers.

HARD AND SOFT KPIS

Even though there isn’t an accepted defitinion, it not uncommon hear about *hard* and *soft* KPIs. Hard metrics are thought of being relatively straightforward to objectively measure, like financial KPIs, for instance. On the other hand, soft metrics like brand awareness, customer satisfaction or service quality are more difficult to measure in an accurate, objective manner.

The distinction isn’t obvious, and there will always be ground for debate, but in these examples there is a sense that the former rest on firmer ground and are more easily and precisely measurable, which as discussed in Chapter 2, is one of the properties of good KPIs to track our business objetives and decisions.

How do we formulate good business questions? Since for our purposes a business question is always *actionable* it is necessary first to understand the business objectives we want to affect as well as the metrics used to assess the results, and to at least have some idea of some candidate levers we can pull. If you have not identified any actions you can take, either the question is not actionable, or you

haven't thought through the problem. Otherwise we are on the right track. We now need to distinguish between descriptive, predictive and prescriptive questions.

Descriptive, Predictive and Prescriptive Questions

In their article "*What is the question*?", Jeff Leek and Roger Peng describe six types of questions that you may want to answer with data: descriptive, exploratory, inferential, predictive, causal and mechanistic.[3]. Data analysis usually mirrors our analytical processes, so these map somewhat neatly to the threefold classification used here: descriptive, predictive and prescriptive.

In Chapter 2 I described the three types of analysis, so here I'll just repeat that *descriptive* analysis generally looks at the past, *predictive* at the future and *prescriptive* finds the best actions we can make today to change the future.

One of the motivations to write this book was the casual finding that most people tend to ask descriptive questions and have trouble finding the right place to use predictive and prescriptive analysis. Later in this chapter I'll provide enough examples to eliminate any confusion you may still have about these concepts.

Always start with the business question and work backwards

One of the preferred catch phrases in the data world is that practitioners create value by finding *actionable insights*. While there's nothing wrong about this assertion there is a risk of spending hours, days or weeks in search for the million-dollar insight.

At some point in my career I did something similar: I found that it is relatively easy to look at the tails of the different distributions — those with lower probability to arise — and find unseen business opportunities on these microsegments. Since most models focus on the average customer (thereby neglecting the tails), this was a relatively straightforward way to help my employer make some money. That was the definition of low-hanging fruit. There were two problems, however: it was not scalable, and it was a highly manual and time-expensive process.

In general, a better practice is to always start with the business question and move backwards to the data. This process leads to faster actionable insights, since, well, you have already started with the actionable insights you want to find from the beginning![4] The process described in this book will help you discipline the analysis and hopefully you won't waist your or your team's valuable time in the search for the promised actionable insights.

Further deconstructing the business questions

The sequence of why questions helps us move from specific to more general questions, in the quest to find the metric that we really want to impact. The risk is that this final metric may be too general to be

actionable (the highest level is almost always something like “increase profits”). We should remember, however, that our own business objectives act as a natural constraint, so there’s usually an upper bound in the sequence. Furthermore, there are techniques that allow us to do just the oppposite and start decomposing questions in order to find just the right level where we can clearly identify intermediate objectives that are also actionable.

For instance, consider the problem of finding the best actions to get the highest conversion rate possible for your outbound marketing campaigns. Notice that I have already framed the question as a prescriptive one on purpose: the business metric is well defined (conversion rate) and if we find suitable actions then we can (in principle) choose the best ones for our purposes.

DECOMPOSING CONVERSION RATES

Any ratio can be decomposed by multiplying and dividing by different metrics. Here we start with the ratio of sales to leads — the conversion rate — and first multiply and divide by the number of reached customers. We then repeat with the number of customers that we actually called (dialed). In the end we reorganize the equation so that each of the parts represents a relevent metric in its own right.

$$\underbrace{\frac{Sales}{Leads}}_{CR} = \underbrace{\frac{Sales}{Reached}}_{A} \times \underbrace{\frac{Reached}{Dialed}}_{B} \times \underbrace{\frac{Dialed}{Leads}}_{C}$$

Conversion rates can be easily decomposed, leading to more directly actionable questions. In this case, the conversion rates CR are the product of three different ratios, each with different possible levers to pull, and with possibly different accountabilities.[5] Starting from the rightmost ratio (C), if out of 100 leads you only tried to contact 15 by dialing their numbers, it could mean that your telemarketing team

is in a low productivity valley, and you'd better talk with their lead to find actions or at least understand what's happening.

Similarly, if you have already dialed each of the phone numbers and were only able to reach a low fraction of them (B) you may want to search for variables that allow you to predict the best time to contact your customers: this may now be a job you assign to your company's data scientists.

Finally, if your sales team is only able to convert a small fraction of those who were reached (A), it could be that the predictive models should be improved to generate higher quality leads, that your compensation scheme needs to be adjusted, or your product-market fit is not right yet.

Notice how the decomposition immediately allows us to find intermediate metrics or questions, with their corresponding actions, to increase conversion rates. This trick can be easily applied to most conversion funnels. Let's take the example of an archetypical *two-sided platform*.

Example with a two-sided platform

Two-sided platforms, or marketplaces, generally try to match users in one side with users on the other side. Facebook, for instance, matches companies that want to place ads (in order to make sales) with the right customers (users of the social network). Amazon tries to match distributors or sellers of goods with the right buyers, Uber matches drivers with passengers and so on.

Imagine you start your own dating platform. Here the two sides are users that want to find their perfect match. Most of these dating apps allow users to communicate with each other. For simplicity let's say that the rules of the game allow only one message per user; the more general case will only make the decomposition longer.

If they like each other they can take it to another place (a coffee shop, a bar or a restaurant). Your team of data scientists wants to improve the app's matching efficiency, measured by the ratio of converted matches. For the sake of the argument let's say that users always provide feedback to the app so that we can always know whether two users met.[6]

We have a data set of all users, their interactions (Message 1 and Message 2) and the final outcome (Met or Didn't Meet). The matching efficiency (ME) can then be decomposed as follows:

DECOMPOSING THE MATCHING EFFICIENCY FOR A DATING APP

We display images of users in a dating app with the hope that these are high-quality potential matches for other users. Each user may decide to start a conversation by sending a first message (message 1) which may be replied by the second user (message 2). After this they either decide to meet elsewhere or stop the conversation.

$$\underbrace{\frac{\text{Met}}{\text{Displayed}}}_{ME} = \underbrace{\frac{\text{Met}}{\text{Message2}}}_{A} \times \underbrace{\frac{\text{Message2}}{\text{Message1}}}_{B} \times \underbrace{\frac{\text{Message1}}{\text{Displayed}}}_{C}$$

In this equation each term denotes the number of occurrences for each event. For instance, *Met* denotes the number of people that ended up meeting and *Message1* and *Message2* denote the numbers of first messages sent and the number of replies, respectively. Also,

each ratio should be less than one since the number in the numerator is a count for a subset of the event in the denominator. This is always the case when decomposing conversion funnels.

Notice what the decomposition buys for us: if we want users to match we need them to exchange messages, which can be represented by three ratios: once a user finds someone displayed on the app, she can send a first message (Message1). This ratio (C) shows if the algorithm is being efficient from the point of view of user 1: if the app displayed 10 candidates and all were of high-quality, then she would message all of them.[7] User 2 may now reply or not: if she does, it may signal that the algorithm is doing also a good job for her (B). Finally, after the second message is delivered they either meet or not (A).

But not everything depends on the algorithm's accuracy: a decision to start a conversation (message 1) or reply (message 2) depends also on each user's attention, say, because of delays in communication: dating apps are fast-moving platforms, so if any user takes too long to reply, the other user may lose interest and continue searching for potential dates. We can then device methods to incentivize faster communication (emails, push notifications or pop-ups reminding that someone is waiting for a reply). Bumble, for example, does just that: the first contact for each side must be withing the first 24 hours or the match is lost.

The takeaway here is that some business questions can be further decomposed to find the right actions, so we may need to reconsider

affecting intervening KPIs to achieve our objectives. We will now go through some real-life common use cases.

Learning to ask business questions: examples from common use cases

We will now go through a selection of examples, starting with what I ´ve seen is the standard way to frame the business question, and posing the corresponding descriptive, predictive and prescriptive counterparts. Recall that a good prescriptive question should always find ways to pull some levers so that we get the *best possible* outcome in terms of the business objective we have chosen. I will further develop some of these examples in subsequent chapters, to the point of providing what I think is a good-enough prescriptive solution; I'll let you find ways to improve on that. For now remember that our purpose in this chapter is just to learn to translate business questions.

Lowering churn

In all companies we need customers in order to generate revenues. We start by *acquiring* customers and then part of our job is to keep them loyal for the longest time possible. The rate at which customers leave — the churn rate — is the ratio of the number of customers we lost in a fixed period of time relative to the overall customer base in that same period. Since acquisition costs can be relatively large compared to retention costs, most companies have specialized areas with the specific objective of safeguarding as much as possible their current base.

This is one standard use case in most companies, so it provides a great way to start applying the techniques (Figure 3-3).

questions for churn

Figure 3-3. Different questions asked for the churn use case

DEFINING THE BUSINESS QUESTION

Let us start with the business question most companies face: how can we lower the churn rate? This is an example where we start with an action and not with the business objective, so we can apply the sequence of why questions, and most likely we'll end up with the simple fact that customers are our main source of recurring revenues. It seems straightforward, but this simple fact takes us to the main KPI we want to maximize: it's not the churn rate that we want to make as small as possible, it is revenues that we want to be high. Or is it? You can always give away everything to keep your customers, thereby increasing our costs. It follows that this is *not* the right metric we want to impact either: it is profits, measured as the difference between revenues and retention costs.

DESCRIPTIVE QUESTIONS

At the most descriptive level we want to do several things. Of course, we start by asking whether our churn rate is abnormally high and how it has evolved in the past. We may start at the most aggregate level by looking at time trends and patterns of seasonality giving us a sense of our current health status. But data has the power to go deeper and tell us *who* are the customers that have already left. Are they high- or low-value customers? What is their tenure with us? Have they reached us in the past showing their dissatisfaction? Are they

geographically located in specific areas? What are some of their sociodemographic characteristics, such as age and gender? What are their usage or consumption patterns?

We can get as granular as our data and time allows. But you get the idea: this is just a photo, and hopefully I have convinced you by now that no matter how high-definition it is, it’s hard to get more value out of it. At this point it has mainly been informative. The real value from this descriptive analysis is its ability to take us further in our quest to find the best decisions we can, in order to achieve our ultimate objective.

PREDICTIVE QUESTIONS

AI and machine learning can help us find answers to the predictive question: can we anticipate which customers are more likely to leave? Thanks to the richness of our descriptive analysis, we have hopefully now found some of the primary drivers that explain our current churn rate. But data alone can only take us so far. The best data scientists are those who *understand* and *hypothesize* why customers are leaving. In this way they can create more specific predictors in a process called *feature engineering* and it is the best way to get really good predictive power. Knowing what to include or not in our models is the holy grail in the construction of good models, even more than, say, choosing the ever-more powerful available algorithms.

How much value does the predictive Q&A provide? In Figure 3-3 I suggest is higher than the descriptive step, but it could be null. It all depends on how you use the predictive results and many times they

are not used at all, possibly because the original question was not actionable.

PRESCRIPTIVE QUESTIONS

Finally, we have arrived at the prescriptive question: what levers should we pull if we want to maximize our profits from our retention campaigns? But are we thinking of *short-term profits*? Will customers learn our strategy and start gaming our retention system thereby increasing longer term costs? Most mature companies prefer to use the Customer Lifetime Value (CLV) we introduced in Chapter 2 and I agree that this is indeed a better picture of the long-term net value of our customers. But this choice of a metric comes with its own set of difficulties: the future is hard to predict, as Yogi Berra famously said, and even harder is understanding the longer-term effects of our actions.

We will talk about levers in Chapter 4 but suffices to say here that for the case of customer retention, we can always give away something at least in the form of discounts. What, then, are the right discounts for each customer? The CLV provides an upper bound on how much we should reasonably give away, but we always want to find the action with the lowest cost that guarantees retention. This takes us closer to the personalization of levers.

The prescriptive ideal is one where we choose the *right* action, at the *right* time, for the *right* customer. Too many *right stuff*: prescriptive analysis is complicated so most times we will try to simplify our lives. Otherwise we might never do anything! I will talk about the power of simplification in [Link to Come]. But at least we have

already framed the question in a way that, by design, can potentially generate the highest achievable value. Recall that this chapter is about learning how to frame questions. In [Link to Come] I will go into the details of one possible solution to this use case.

Cross-selling: next-best offer

Most companies sell more than one product or offer more than one service. Economists call this natural advantage that a company may have when offering products that can benefit from similar production processes *economies of scope*. It is thus natural for most of us to look for ways to deepen our relationship with our customers by trying to do some cross-selling. In the consulting jargon it has been relabeled as the now famous *next-best offer* which already takes us to the prescriptive terrain.[8]

questions for crosssell

Figure 3-4. Different questions asked for the case of cross-selling

DEFINING THE BUSINESS QUESTION

The business question here is straightforward (Figure 3-4): *what should I offer now to my customers?* If you wonder why would you even want to do such a thing (the sequence of why questions) the answer is not as clear as with customer churn. The difference here is that cross-selling has two effects. The direct effect is the usual channel of higher revenues and profits. But the indirect channel is more interesting and complex: customers who buy more from us tend to be more loyal thereby increasing the time they remain as customers. Because of this, many times we may consider cross-selling at a discount just because the long-term *overall* profits are

higher, even when the transaction of an individual product is made at a loss for the company. It appears, again, that CLV is the right KPI to optimize.

DESCRIPTIVE QUESTIONS

On the descriptive terrain, the type of questions one would normally explore are things like the patterns of consumption for different customers. Specifically, it is natural to explore if certain *sequences* of products arise more naturally with different customers. Think of a bank, for example: most customers start at a young age with a relatively simple product like a credit card. With time, and with their incomes increasing with job experience, they tend to move to more sophisticated credit and investment opportunities: you may first get a mortgage, move to life insurance and so on. With sequences, the order in which each product is purchased matters, so it is standard to start by looking for those patterns in the data.

PREDICTIVE QUESTIONS

Now, since each customer has already purchased something, it seems natural to ask if we can *predict* what they are most likely to purchase given their patterns of consumption to date. We could then move proactively and not wait and see if they purchase with us or our competitors. But should we offer the good with the largest profits for us, or something else with a higher likelihood of being purchased? Going back to the bank example, you may want your customers to accept a mortgage loan (because of its large returns) but for college students and young professionals it may be highly unlikely that they will accept. This takes us to one of the most interesting tradeoffs in

next-best offer analysis: likelihood of purchase vs. increase in value. Which in turn brings us to the prescriptive question.

PRESCRIPTIVE QUESTIONS

Since we can offer several items to each customer, which one should we select so that we can capture the highest value? As mentioned above, since we are dealing with sequences and time, the right metric is most likely the CLV. In a truly customer centric sense, the prescriptive ideal would take us, again, to find the *right* product, for the *right* customer, at the *right* price and the *right* time. We'll see later an approach to try to tackle this highly complex question.

CAPEX optimization

The automotive, oil and gas, telecommunications and airlines are examples of industries that are capital intensive: in order to operate they need to allocate large amounts of resources in building and maintaining the factories and plants, towers, planes and any other physical assets that depreciate in time. This type of investment is called capital expenditure or CAPEX and is common to all industries, not only the four cited above.[9]

One natural question that CFOs and other executives have in any company is how to allocate CAPEX, say, across functional areas or geographical locations (Figure 3-5). Since it may represent a large part of a company's cash flow we even have specific KPIs to measure its impact, such as the Return on Investment (ROI) or Return on Capital Employed (ROCE). Nonetheless, we should always proceed and question why we need to allocate CAPEX and what exactly are

we trying to accomplish: for instance, where is Income in the ROI numerator coming from?[10]

questions for capex optimization

Figure 3-5. Different questions asked for the case of CAPEX optimization

At a descriptive level, we could start by finding correlations between different CAPEX allocations and revenues across geographies. This exploits the variation in previous investments with the key metric that we believe should be impacted: if capital expenditures do not affect our revenues why are we even doing it? Another possibility is to exploit the variation in time and plot aggregate series in search of any preliminary hints of a relationship between CAPEX allocations and revenues.

The main problem we have when considering any investment is that we *do not know* what the returns will be, so it would be great if we could perfectly predict them. Optimal allocation could then just be a matter of rankings: if I have one dollar to invest and know the returns of all candidate allocations I would put it on the one with highest returns. But can we trust the correlations in our descriptive analysis? Is the effect we find really causal? As usual, the hard part is to find reliable causal predictions and that's what our data scientists will try to find with the use of their machine learning toolkit.

But assuming we have achieved reliable and accurate predictions, the prescriptive part is almost done for us: allocate your budget in different geographies ranked by their returns. Later I will show you one example of how this can be done but for now all we need to learn is the framing of the question.

Stores locations

One of my favorite use cases is where to open a store, and since we have already talked about CAPEX optimization we immediately see that this is just an instance of the same problem. We have a budget to strengthen our commercial presence and ideally we would just open a store where we will have the largest possible return (Figure 3-6). A natural KPI is the net present value (NPV) of the store's profits, or is it?

Just to show the complexity of the problem consider opening a store that is already very close to another one (have you ever wondered why there are so many Starbucks in one specific block or neighborhood?). You could capture extra revenues and profits but only at the expense of profits in nearby stores. So a more reliable KPI would be the aggregate level of profits, at least at a local (neighborhood, street or even city) level.

questions for store location

Figure 3-6. Different questions asked for the case of where to open a new store

Descriptively I would start by looking for patterns of variations in profits across different spatial locations: are there any of our own stores in a vicinity? What about the competition? Do we have data to approximate the number of potential customers that enter different stores? What about the average income in the neighborhood? Is it a residential neighborhood? Many questions that we may pose in order to find the patterns that explain variations in profits.

Just like with capital expenditure allocation, if we could perfectly predict the NPV of overall profits we are almost done: allocate all of your budget in opening stores ranked by this KPI. Of course, I’m assuming here that you have a finite budget and that you won’t invest in opening stores that have negative returns.

Who should I hire

It is an understatement that our employees make our company great or not so great. So one of the most important decisions we constantly make is who to hire and human resources units spend considerable efforts in having a robust and reliable recruitment process (Figure 3-7). The main problem we face in hiring is that some of the KPIs may not be as easy to measure. Consider productivity, for example. If you are a salesperson we can clearly measure your own productivity with the number of sales in a fixed period. But for many other positions it is harder to measure productivity or even their impact on revenues.

questions for hiring

Figure 3-7. Different questions asked for the case of hiring decisions

Suppose we can reliably measure productivity like in the case of our sales force. Is that the only KPI that matters? What about tenure? You may not want to hire a superstar sales person if she changes jobs a month later, as it may not compensate for the hiring and training costs. Ideally we would like to measure something like the customer lifetime value, so let’s use the analogous term — the employee lifetime value: the net present value of our individual contribution to profits. That way we can include the expected duration and the monetary impact.

Let us imagine we have a dataset for all of our sales people in the past 24 months. As with customer churn, we need our dataset to include active employees with different tenures as well as those who have already left. This variation would allow us to start finding patterns. It would also be ideal to have some of their hard performance metrics (monthly sales) and also softer metrics like their 360 survey results. Finally, it would be great to have some of the data we actually get when hiring: their CVs, studies, previous experience, psychometric studies, gender, age, etc. We can then search for correlations in the data to have a sense of the kind of variables that may predict performance.

The predictive question is simple enough: with that ex-ante information — anything we collect before making the hiring decision – can we predict the candidates' performance? If we could, the problem would be at least solvable, but as before, there are complexities. The main issue here is one related to all search problems (think of the problem of finding a couple): should we keep looking for new candidates? Is this the best we can find? If we keep searching for another month, say, will we be able to find someone better? We will later talk about the *explore-exploit* tradeoff common to most search problems, but for now just notice that we can either hire and get the best of our new employee — that's the exploit part, but it is not related to labor exploitation, of course—or should we keep exploring the market for better candidates.

One final word of caution when using AI in problems like these is that our predictive algorithms are very sensitive to biases in our data and our data scientists should go through the trouble of searching for

them and try to find ways to correct them. Imagine that your data set shows that most female sales people are very productive but quit after a month. Your prediction model might end up showing that the employees lifetime value for women is considerably lower and you will end up hiring mostly male candidates. But why are women leaving so quickly? Can it be that our company has a terribly misogynist manager? We should better fire that person first, and then hire more women. But the moral here is that unless we have debiased our data the best that we can, our predictive models will be highly deficient: as we say in the data world, “garbage in, garbage out”.

Delinquency rates

Many companies provide ways to finance their customers, for instance, with store-specific credit cards that are most commonly found with large retailers. Even better if we can leave that job to specialized firms (banks), but many times we do the funding ourselves. The business question is how to provide lending without increasing the delinquency rates (Figure 3-8).

questions for lending

Figure 3-8. Different questions asked for the case of lending decisions

The descriptive questions are similar to the cases we have already discussed but let me just reinforce two ideas: if you have correctly defined the business question and framed it as a prescriptive one you should be looking for patterns in the data that guide that objective, and not the other way around. Also, we should try to exploit variations in the underlying characteristics of the problem in order to predict the outcome we care about. So I would go about looking for

variation across geographies and customers and correlate them with delinquency rates and delinquency outcomes to set up the predictive problem: can we predict if a customer will default on a loan? If he does, can we recover any fraction of it possibly through an aggressive collection strategy?

We will talk more about ethical problems later, but again, it is important to mention that biases in our data can pervasively affect the outcomes we want to pursue. With loans we should be careful not to affect minorities or groups that are underrepresented because of the way we have lent in the past.

To set up the prescriptive question start with the metric we want to affect: it's not whether a customer defaults — their default probability — , but rather the expected benefits net of costs from the loan. We will deal a lot with expected values later, but for now it suffices to say that to formulate and provide answers to all prescriptive questions we need to have a clear understanding of the costs and benefits from our actions. What are our levers? We are certainly able to determine the amount of the loan, and of course, the decision to do it or not. Banks also have the ability to set up to interest rate, but because of regulatory issues, this lever may not be available to other companies.

Stock or inventory optimization

A very common problem for most companies is how many units of each product that we sell should be in each store's inventory. In a similar vein, banks constantly decide how much cash to have in their

ATMs. Let us start backwards this time, starting with the prescriptive question (Figure 3-9).

questions for stock optimization

Figure 3-9. Different questions asked for the case of stock optimization

Consider the costs and benefits from over or understocking one particular item. At a first level, if there are not enough units we will hinder our sales in a given day, so the cost is reduced revenues. Understocking can also increase transportation and logistics costs that may be considerable enough to include in our analysis. What about overstocking? With some probability the value of those items may decrease, either by depreciation, by the risk of mismanagement or theft, or just because tomorrow a new and better alternative arises and there will be no demand for the old stuff.

At the prescriptive stage we would want to find the right amount of each item to minimize the *expected* sum of these costs. Later I will delve in the details of how this can be done. What is the underlying uncertainty of the problem? We will use AI to help us deal with it in the predictive stage.

First of all, we do not know how many units we will sell of each product. If each day we always sell the same amount, say 100 units, at the very least we should always have those 100 units. A dissatisfied customer will not only represent the current foregone sales opportunity but possibly many in the future: she and her acquaintances may not come back to the store, so let’s hope she’s not an influencer! Therefore, we should better start by predicting demand in a fixed period of time. But what is that period? A day? A week? It

depends on all other costs: if transportation is cheap, relative to other risks such as theft or depreciation, you can stock again tomorrow without a problem (that is the case with ATMs, for example). Otherwise you may need to predict the likelihood of these taking place. Again, we are in the arena of expected values, a topic to be discussed in a later chapter, but yes, as you can see optimization can be very hard.

So now we can guide the descriptive analysis: how do sales vary across time and geographical locations? Are there seasonal effects? What about theft and robberies? If your items are durable goods (cars, fridges, cell phones, laptops or the like) how do their values fall with time? You can see the picture now.

Stores Staffing

Our final example has to do with the problem of choosing the number of sales people in a store (Figure 3-10). In a sense it is a similar problem to the stocking problem: what are the cost and benefits of over or understaffing?

questions for staffing decisions

Figure 3-10. Different questions asked for the case of staffing decisions

If we do not have enough people we will certainly have lower revenues: those customers who wait a long time will leave the store and buy with the competition. Or they will stay and buy today but their lower satisfaction will likely result in higher customer churn affecting our revenues in the future. On the other hand, overstaffing creates unnecessary and foreseeable costs. Therefore, expected

profits — revenues from sales minus the staffing cost — seem to be a reasonable KPI to optimize. The customer churn effect may be just too much to begin with, so let us start by trying to find the right number of salespeople in our stores to have the highest possible expected profits in a day. If we tackle this already difficult problem we can proceed to optimize the longer-term version (recall the value of simplification).

What do we need to know in order to solve this problem? How would we proceed if there is no underlying uncertainty? Ideally I would like to know the number of customers coming to our store at any given time in a day, say, each hour or in time periods of thirty minutes. We will need to predict this flow of customers in each of our stores. This will naturally lead us to waiting times given the size of our sales staff. We may now need to decide what is a reasonable waiting time as the limit of no waiting time may just be too costly, especially since there are peak hours where we have many customers and valleys where it appears that we have overstaffed. And waiting times affect our profits today.

What should we look for in the data then? Variation across stores in demand and staffing is what we need to exploit, but also the outcomes we want to predict: sales, profits, waiting times and customer satisfaction are four that immediately pop up. The descriptive stage should be set up to search for these correlations.

Key takeaways

- **Business objectives are usually already defined**: but we must learn to ask the right business questions to achieve these objectives.
- **Always start with the business and move backwards**: for any decision you're planning or have already made, think about the business objective you wanted to achieve. You can then move backwards to figure out the set of possible levers and how these create consequences that affect the business.
- **The sequence of why questions can help define the right business objective you want to achieve**: this bottom-up approach generally helps identifying business objectives and enlarging the set of actions we can make. But other times you can also use a top-down approach similar to the decomposition of conversion rates.
- **Descriptive, predictive and prescriptive questions**: descriptive questions relate to the current state of the business objective; predictive questions are about its future state. Prescriptive questions help us choose the right levers to attain the best possible future scenario.

Further Reading

I haven't found any books on how to ask *good* business questions in the context of decision-making. This is not to say that I'm the first practitioner that suggests that the way we frame our business problems can make a big difference in the results. Almost any book on data science methodology will at least mention the topic. You can go back to the references in Chapter 1 or check Foster Provost's and Tom Fawcett's *Data Science for Business*.

In my opinion the literature has at least two shortcomings: most data scientists rarely care about solving the prescriptive problem and rather focus on providing high-quality predictive solutions. Also, the literature directed to business people hasn't been able to provide end-to-end views of decision problems that can be tackled with AI and analytical thinking.

Several of the use cases covered here can be found elsewhere, but I'm not sure at what level. You can search online for white papers written by consulting firms and they will provide some possibly interesting insights (but of course, consulting firms make money from developing those use cases so don't expect too many details).

On two-sided platforms I enjoyed reading Geoffrey G. Parker, Marshall W. Van Alstyne and Sangeet Paul Choudary, *Platform Revolution: How Networked Markets Are Transforming the Economyand How to Make Them Work for You* and David S. Evans and Richard Schmalensee, *Matchmakers: The New Economics of Multisided Platforms*.

The topic of ethical concerns in machine learning is an important one, but I will provide references in [Link to Come].

asdfasf

---

1 The sequence of *why* questions

2 On different organizational structures see *http://www.informit.com/articles/article.aspx?p=2931568&seqNum=2* or *https://www.mckinsey.com/business-functions/organization/our-insights/the-five-trademarks-of-agile-organizations*.

3 See *https://science.sciencemag.org/content/347/6228/1314.full.pdf+html*

4 Cassie Kozyrkov, Chief Decision Scientist at Google, has presented a similar view *https://towardsdatascience.com/hypothesis-testing-decoded-for-movers-and-shakers-bfc2bc34da41* or *https://hbr.org/2019/06/the-first-thing-great-decision-makers-do* .

5 In case you didn't notice, I'm multiplying and dividing by the same metric so that the equality is always preserved.

6 Some dating apps actually incentivize users to provide this formal feedback but other times there are indirect ways to measure the matching efficiency.

7 This is not to say that there may be different strategies: a user may start only messaging the top candidate displayed and see where that takes her. In that sense, you may want to restrict the analysis to a decomposition that excludes the last ratio.

8 Almost, the *best offer* refers to choosing actions (offers) without making reference to the metric with respect these are best. We'll see that this is not immediate in the next couple of paragraphs.

9 Compare this with operating expenditure or OPEX that includes, among many other things, the salaries payed to your employees.

10 Recall that $ROI = \frac{\text{Income from Investment - Cost of Investment}}{\text{Cost of Investment}}$

Chapter 4. Actions, levers and decisions

> **A NOTE FOR EARLY RELEASE READERS**
>
> With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.
>
> This will be the 4th chapter of the final book. Please note that the GitHub repo will be made active later on.
>
> If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the author at analyticalthinkingbook@gmail.com.

Chapter 3 was all about learning how to translate business problems into prescriptive questions that, in our case, must always be actionable. But what is actionable? Or even better, is *everything* actionable? We now turn to this question, in our quest to find levers that take us closer to the prescriptive ideal.

One word of caution is in place: to find levers we need *to know our business*. This is not to say that you must have spent many years in one specific industry. That of course helps as you must've developed strong intuitions about why things work and when they don't. But it is also true that many times having a non-expert, even naive view, can help think out of the box and expand our menu of options.

Going back to our decomposition, we will now move from the outer right side where business outcomes live to the outer left side where the levers we pull are Figure 4-1. As we’ve already mentioned, this is the natural and healthy sequence to adopt: we start with the business, and then ask how we can achive the best results by pulling the right set of levers.

our actions and choices

Figure 4-1. Identifying the levers we want to pull

Understanding what is actionable

The hard truth about life and business-making is that most of our objectives can only be achieved indirectly, through actions we take. For instance, we can’t increase our sales, our productivity or customer satisfaction or reduce our costs just because we say so. These intervening factors (human or technological) restrict our ability to do the absolute best we wish we could accomplish.

The impact that our decisions have on our business objectives is mediated by the rules of cause and effect, and it usually takes a lot of experimentation and domain knowledge to understand what works and what doesn’t for our businesses.

WHAT IS A LEVER

In the context of this book, “levers” is synonymous for “actions” or “decisions, so whenever we say that “we want to pull some lever to obtain a business outcome” this just means that we are looking for suitable actions or decisions.

In general, we can divide levers into two types: those that depend mostly on the rules of the physical world to create consequences and those that arise from human behavior. As you would expect, each of them has their own sets of complexities and difficulties. Levers of the physical type depend on our understanding of the laws of nature and on technological advance. Human levers depend on our understanding of human behavior.

Physical Levers

As it turns out, the original use of the word “lever” is of a physical nature: you take a beam and a fulcrum, pull the beam down, and you are now able to move objects that were too heavy to lift by yourself. This use notwithstanding, physical levers have become a landmark of the modern economies: the rapid growth during the industrial revolution, the invention of the microchip and the current internet revolution, just to mention a few, were vastly facilitated by this class of levers.

Thanks to Henry Ford’s assembly line, for instance, the production of cars was greatly improved. It *only* took a complete redesign of the production process, but once you pulled that “lever” you were able to produce more cars in less time, with the consequent reduction in production costs.

Engineering advances generate physical levers that we may not be conscious about. For instance, changing the height or angle of an antenna in a cell site improves the quality of the calls we make or the speed at which we can transfer data in our day-to-day mobile

communications. Similarly, better software configuration may improve your ability to work on the cloud or on premise. Physical levers require technical expertise that may be costly to acquire or hire, but since modern economies are built around the technological revolution, having at least some general knowledge of what can be achieved can take us very far if we want to be more productive or have more satisfied customers.

physical levers: queues

Figure 4-2. Queues as physical levers: left hand shows a multiple-line, multiple server design. By moving to a single-line, multiple server design (right) we may impact waiting time, so this change is a lever when we want to have an effect on customer satisfaction.

Let’s consider the design of queues as a final example. Figure 4-2 shows two possible designs: multiple-line, multiple-server on the left and a single line, multiple servers on the right.[1] This is not the place to even try and delve into the technicalities, but let’s just mention that under certain conditions it can be proved that the average waiting time for the design on the left is longer than for the case a single line. If these conditions are satisfied at your workplace and your objective is to improve general customer satisfaction (measured by the time they spend waiting in line), you can just make a redesign of your queues and you may meet your goals.

PHYSICAL AND PSYCHOLOGICAL LEVERS IN WAITING LINES

In Figure 4-2 I also claim that the *perception* of waiting times may also be positively affected by switching to the design on the right, but this would take as to the terrain of human levers where psychological laws operate. We will address this topic shortly, but you can check *https://www.nytimes.com/2012/08/19/opinion/sunday/why-waiting-in-line-is-torture.html* for some evidence on the psychology of waiting in line.

Human Levers

Just as the design and use of physical levers requires considerable technical expertise, human levers entail a thorough understanding of how humans behave. Humans, as opposed to materials, have a very specific set of complications of their own. Let´s discuss the most important briefly.

The most obvious one is that we can’t force others to behave the way we want: we have to *incentivize* them. You can’t force potential customers to buy your products, or your employees to work more or be more productive: you need to create the conditions that will lead them to act in ways that are favorable to our objectives out of their own self-interest.

Moreover, we are *heterogenous and diverse* beings: even identical twins that share all of their genetic material behave in different ways. We also have a sense of *agency*: we have intentions and these vary from individual to individual and throughout our lifetimes.

To add one more layer of complexity, we are *social animals* and our behavior may vary drastically if we make choices surrounded by people or alone. We also learn from experience, a process common for toddlers, the elderly and everyone in between. Finally, we make errors: we may regret some of our previous decisions, but these may not be easily predictable.

Why do we behave the way we do

I will set out on an ambitious agenda and try to condense why humans behave into three categories that I believe cover a big part of the reasons behind our behavior. I was trained as an economist, so you may see a bias in this enterprise, but hopefully other social scientists won't disagree that much.

I will claim that most of our behavior is driven by our *preferences* or values, our *expectations* and the *restrictions* we face. These map neatly to the economists' portrait of a rational being but rationality has little to do with this characterization.[2]

Think about why you bought this book: my guess is that you wanted to learn about AI and how to use it to make better decisions, but since you weren't sure of the quality of the material, you took a leap of faith and hoped for the best. Nonetheless, you could be doing anything else right now: you could be reading some other book, technical or not, watching a movie, sleeping or spending time with your beloved ones. You must have valued reading this book (at least expected it to be the case). At the same time, you were able to afford

it and have the time to do it, two of the most basic restrictions we generally face.

Does this generalize to any other choices? I believe it does with most choices we make, if not all. In a sense, the claim is almost tautological: ask anyone why they just behaved as they did and they could easily say “because I wanted to”.

Now, preferences come in at least two flavors: we have individual and social preferences, and this distinction allow us to account for the differences in choices when we are surrounded by others and when we are alone.

We will now discuss in detail each of these.

Levers from restrictions

Let’s start with the pricing lever, arguably one of the most common actions we take to achieve the specific business objective of increasing our revenues. It is one of our favorite levers, since it directly affects our revenues — price times sales volume or $P \times Q$.

Interestingly, revenues depend on price in a way that makes the choice to pull the lever not obvious at all. The difficulty comes from what economists call the “Law of Demand”: when we *increase* our price, our sales generally *fall*. Since sales depend on the price we charge, revenues should better be expressed as $P \times Q(P)$ to make clear that our choice of the pricing lever has two effects on our revenues: a positive, direct effect coming from the first term, and a

negative indirect effect from to latter term. The overall effect depends on the sensitivity of demand to changes in prices.

THE LAW OF DEMAND

Figure 4-3 shows how demand (Q) (horizontal axis) varies with price (vertical axis). Don't be confused by the choices of the axes: for historical reasons this is how economists depict a demand function, even though prices — our lever — would most naturally be depicted on the horizontal axis. A better term is *inverted* demand function, but the term never stuck.

law of demand

Figure 4-3. Purchases fall as we increase prices

The important thing to recognize is that as the price falls consumers purchase more of our product. Notice that reducing the price $10 from $100 to $90 generates an increase in purchases of 7 thousand units. Compare this with Case A in Figure 4-4, where in order to generate the same increase in volume we just need a reduction in price of $1.

law of demand

Figure 4-4. Case A shows a relatively price-sensitive demand function. Case B depicts the case of a demand function that violates the Law of Demand.

The differences in the two is what economists call the *price elasticity* or our customers' price sensitivity. There are many determinants of price elasticity, most importantly the availability of close alternatives or substitutes and the proportion of our income devoted to the consumption of a specific product.

The right panel in Figure 4-4 shows the case of a product for which the demand increases as prices *increase*, thereby violating the Law of Demand. Are there real-life examples of Giffen goods, as economists refer to them? Take the case of fine wine or jewelry or any other premium goods. For some people their demand will increase when the price is higher, as it may signal better quality or premium status. Say that such a wine and customers exist: if we now decrease the price of each bottle, will they consume *less*? If what they value is the price, it may well be the case, but if they like the wine, irrespective of the price, it is unlikely.

Figure 4-5 shows a somewhat standard relationship between revenues and our pricing lever. It should now be clear that if we want to pull the price lever we'd better know if we're to the right or left of the vertical line: our company will be better off if we increase prices in range **A** since a price increase generates *higher* revenues. The opposite will happen in range **B**. The math notwithstanding, the

intuition should be clear: if our customers are not too price sensitive, an increase in price, say by one dollar, will *decrease* demand less than proportionally, thereby generating an overall positive impact on our revenues. This sort of calibration is standard when we are doing price and revenue optimization, one area where prescriptive analysis has been most successful and that we will revisit in further detail in [Link to Come].

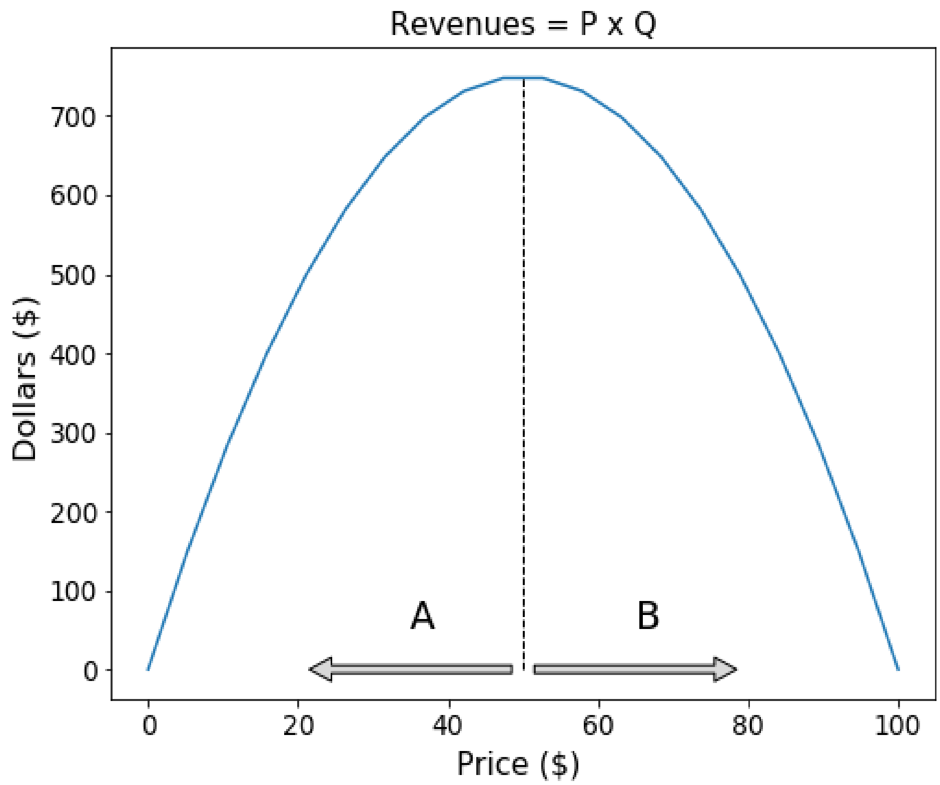


Figure 4-5. How revenues change with our pricing choices

I hope this example convinces you that the choice of the price lever is far from obvious, but in my opinion it's one of the most interesting and successful cases of prescriptive analysis. If we are considering giving away discounts, it'd better be that demand increases

proportionally faster than the falling prices.[3] Otherwise we should better look for other levers.

PRICING LEVERS AS RESTRICTIONS

You may wonder why I chose to classify the pricing lever as a constraint. One important reason why customers generally follow the Law of demand — the negative relation between purchases and prices prices — is that by changing prices we affect their budget constraints.

Interestingly, this effect operates not only with our customers but also with other prospective ones that haven't started buying since current prices may be too high.

But is this always true? The discussion of Case B in Figure 4-5 notwithstanding, most of the time we follow the Law of Demand so most people just take it for granted.

TIME RESTRICTIONS

It is not a coincidence that two of the main restrictions we face are time and money. We have already discussed the budget constraint, but what about time constraints? Do companies leverage time restrictions as they do with budget constraints?

Consider digital banking (and the digital transformation, in general). I don't know about you, but most people I know can't stand going to a branch, as it feels like a waist of our valuable time. One of the best cases for a better user experience is that we relax our customers' time constraints, and give more time back to other activities they value.

If you’re not convinced by the banking example, think about this: in your case, is there something you would engage into if the effort (time) was reduced? Imagine that by cutting by half the time in the gym, say from sixty to thirty minutes, you would get the same results. All of those infomercials that promise the perfect abs in just 10 minutes a day are pulling this lever. People value their time as much as their money, because as they say, “time is money”.

Levers that affect our preferences

We will now consider some of the different determinants of why we like and value what we do. As we will see, *all* of them are actionable and are constantly used by companies around the world.

GENETICS

How much of our behavior is determined by our genetic makeup and how much by our social upbring? This *Nature vs. Nurture* debate is one of the most important and controversial in the social and behavioral sciences, since it is very difficult to disentangle empirically their relative importance on our behavior (Figure 4-6). For instance, if you enjoy a glass of red wine like your parents, is it because of your genes? Could it be that you were raised watching them enjoying red wine which itself created a positive, but *social*, effect on you?

genes and environment

Figure 4-6. Genes and environment both shape our preferences and choices.

Let’s take the most accepted view that both genes and the environment matter, and that some behavior is most likely to arise

when certain genes are exposed to certain environments. We can now ask ourselves if we could leverage this knowledge to attain our business objectives.

At the outset it seems clear that we cannot change our customers DNA but in the foreseeable future it is possible that with further advances in behavioral genetics, we´ll eventually have a thorough understanding of how to leverage the exposure to certain environments to specific customers.[4] Many stores already do some *very* basic and crude genetic leveraging by changing the aromas present in the store when we are buying. But imagine the case of genetic profiling: a person enters a store, we have knowledge of some genetic markers that matter for our product, and offer a complete sensorial experience that makes her more likely to purchase.

I’ll leave this here but keep in mind that this topic raises all kinds of ethical issues. I’ll have time to discuss some of these later in the book.

SOCIAL REASONS: LEARNING

The truth about our choices is that many times we don’t know what we want or what we like, in contrast with the view of rational and consistent choices put forward by most decision theorists and economists. Some people are more prone to try new things and explore the variety and diversity of their tastes. At the other end of the spectrum, other people have had terrible experiences when trying new things and just stick to the same dietary routine they already know and feel comfortable about.

In any case, the fact that preferences are not fixed and consistent, and that most of us like to try new things to at least some degree, should help us find levers to achieve some of our business objectives. This is especially true when a a company launches a new product; since customers are reluctant to pay for something they haven't tried the company typically gives out free samples. This reduces the real and perceived cost of trying the product and is done in the hope that the customer will be willing to pay full price the next time.

You may have noticed that I'm refering to social learning as opposed to individual learning. In the former, when others change their behavior these changes spread throughout the social environment in a process similar to the contagion of diseases. This is generally what happens with the spread of ideas and knowledge: if some new technique dominates we quickly switch to the better and new one. In this case we have a second possible lever: influential people can help us spread the use of new product without the need to give it away for free.

SOCIAL REASONS: STRATEGIC EFFECTS

Imagine we see something like the behavior in Figure 4-7: here, a newcomer to the group brings new ideas or behavior. She first convinces one member of the group who starts behaving similarly. And then another one, and then several others.

What can we conclude as external observers? One thing appears to be clear: an outsider to a group started spreading his behavior, but why could this be? And most importantly for our current discussion, can we leverage this type of social effects to achieve our business

objectives? In the previous section we sketched one possible reason — social learning — and discussed two possible levers (price and influential people).

social spread

Figure 4-7. Social contagion

But it could also be that there are strategic effects that explain these dynamics. Think about two-sided platforms like Airbnb, Uber, WhatsApp, Facebook, Google, operating systems like iOS or Windows, etc.[5]. Going back to Figure 4-7, imagine that one person in our group of friends comes from Europe telling us that they are using the latest messaging app. At the beginning only her best friend downloads it to try it and, of course, to chat with her. But now two other people try it, because, yes, they want to know what their friends keep talking about! The more people that join, the larger our incentives to join: this is a first type of network effect that operates in two-sided networks.

The second type has to do with the other side of the network. Think about Uber: if more drivers join, the easier to find rides for passengers, so now more customers join. But the larger demand makes joining for the drivers also more profitable: you can now see why platforms generate these huge positive feedback loops.

It is common to refer to these as “strategic effects” since our behavior depends on the choices of others, and vice versa. This feedback loop creates quite interesting social phenomena and game theorists keep searching for *equilibria* to these games. The nice thing about equilibrium is that once reached, no one has incentives to deviate. But

from an empirical point of view, they also provide predictions we can test.

Can we use this as a lever to attain our business objectives? Most certainly: one of the most popular levers for two-sided markets is to subsidize the side of the market that is *most* price sensitive by way of discounts or lower fees. This will generate the two positive feedback loops we just described, and by choosing the most price sensitive side we reduce the cost of the lever. For instance, Uber subsidizes the value of each trip by dropping prices for the customers and Google gives away the use of their search engine, but auctions ad space to the other side.

SOCIAL REASONS: CONFORMITY AND PEER EFFECTS

Many times we change our behavior in response to our social network attitudes because we just want to conform or belong. Though plausible, the problem with conformity as a behavioral hypothesis is that it's quite hard to find conclusive empirical evidence that supports it. Think about social learning or strategic effects, and go back to Figure 4-7: you can always claim that social contagion is caused by our desire "to belong". How can we differentiate one from the other? And does it matter?

The short answer to the last question is that it matters because if our hypothesis about how people behave is wrong, we may pull a lever without finding the desired effects.

Conformity is most easily illustrated with influencers. Why would we buy a swimsuit worn by Selena Gomez or Cristiano Ronaldo on Instagram? It could be that we *learned* that it actually looks good on us only after seeing it on them, instead of appeling to a need or desire to *belong*.

Note that conformity may arise from strategic effects: peer and group pressure creates a burden on me, so I may find it on my best interest to do what everyone else is doing. The same reasoning applies to my friends and peers giving rise to what is sometimes refered as herding behavior.

To sum up, this discussion isn't purely academic: it affects and enlarges our set of levers, especially with certain demographic groups such as teenagers. It may not be that effective with other demographic groups, or at least I haven't seen credible empirical evidence showing that we should care about it.

As a final note, let's discuss the case of corporate culture, a common use case where conformity might play an important role. Most people believe that a positive culture will make employees happier and more productive, and a negative culture can produce really bad outcomes such as robbery, corruption and the like. Precisely because we think it matters, it is generally the CEO's and Chief Human Resources Officer task to find ways to create and grow a favorable corporate culture. The desire to conform is but one of the reasons why new cultures arise, so one lever is to find some people that could serve as corporate influencers. Who better than the CEO herself and her whole executive committee?

FRAMING EFFECTS

Let’s now move to the terrain of behavioral economics, the systematic study of “irrational” or “inconsistent” behavior. We’ll see that there is a lot of consistency in our inconsistent behavior that can be used to achieve our objectives.

Suppose that given a choice between your product and your competitors, your average customer chooses yours in some circumstances and your competitors in others. This inconsistency of choice is troubling since it suggests that nothing intrinsic about your product (or your competitor’s) explains the choice, but rather, that something external like the decision context may be the cause of the final outcome.

framing effects

Figure 4-8. Framing effects

Consider Figure 4-8 where three alternative TVs are portrayed with respect to two different attributes, size and price. The problem here is that these attribute compete against each other: I prefer a larger TV, but unfortunately it comes at a cost so I have to trade-off one for the other. Brand A has the smallest screen and thus, it is also the cheapest one. Brand B isn’t too different from A (especially when compared with C), and finally C is the best in terms of size, but you’d have to pay some extra dollars to get it. Which one would you choose?

If you’re like most people then you would’ve chosen B. It appears to be the reasonable choice in terms of the two attributes, especially since C is considerably more expensive. Marketers have been

studying these effects for a long time, so they usually pull the framing lever to direct our choices to whatever they want to sell. Let me repeat what I just said to make clear the point: they want to sell alternative B from the outset and to do so they decide to pull a "framing lever". They carefully select the two alternatives they want to display so that we "naturally" choose B.

Consider Figure 4-9 now, and imagine your objective is to buy a new laptop where we only care about two attributes: the amount of memory (RAM) and the speed of the processor (CPU). Case A shows two alternatives that clearly trade off both attributes: you either have a lot of memory but low CPU (A) or vice versa (B). What should we choose? This type of choice makes us pretty uncomfortable, since there is no clear winner with respect to all attributes we care about, and life is so much easier when we don't have to make sacrifices.

framing effects

Figure 4-9. Buying a computer: another case of framing effects

Wouldn't it be nice if we could find a reason to choose one or the other? This takes us to Case B, where our retailer now presents a third alternative that is clearly dominated by laptop A (C has less memory and computing power). Why would he do that? Alternative C acts as a reference points that helps us find undisputed arguments to choose A now.

Note that the lever here is *the way we present or frame the choice situation*. That's great! If this type of levers work (and they do many times) we don't have to give price discounts to increase our sales. Just frame the decision problem correctly.

PATH DEPENDENCE OR ANCHORING

Suppose you recently moved from New York to Bogota, Colombia. You were used to paying close to 4500 dollars for rent each month for that amazing loft in SOHO. In Bogota you can find something really big and fancy for the same price, so much that you may not have anything to do with the extra space! It seems to me that for *the same space*, you can actually spend *less*. Is this what actually happens?

Most people *anchor* their current choices to what they did previously, and in this case this means that at least at the beginning they do not change how much they are willing to pay. Why on earth could this happen? Remember that many times we don't know what we want, or even worse, what we are willing to pay for what we want. For instance, can you clearly state what is the maximum price you're willing to pay for this book?[6]

A general principle that applies to many choice situations is that people like to understand why a choice is made, even if this only happens afterwards. Would you expect this behavior to remain in the future? Not really. Eventually we may realize that the price per square foot is just too high, and we end up adjusting our willingness to pay to the new reality. And importantly, we are able to explain why this happened, that is, we can rationalize our choices. This is called anchoring, and you may leverage it in some circumstances.

One such circumstance is in negotiation situations. Many great negotiators pull the anchoring lever by starting with a really strong, low offer that serves as anchor to the counterpart, with the end result of tilting the balance in their favor.

LOSS AVERSION

Our final example of levers that may affect your customers preferences is known as *loss aversion*. As its name suggests, the idea is that the worth of something changes if we own it or not, or put differently, if the choice situation is framed as a loss. Figure 4-10 shows one such example where the vertical axis denotes how much we value having more or less of dollars. To the right of zero we gain money and to the left we lose. In the absence of loss aversion the worth in absolute value terms of gaining or losing 25 cents, say, should be the same, but as the figure shows this is not the case.

Is this something we can use to attain our objective? It may not come as a surprise by now, but yes, the way you communicate with your customers can make a difference. What the theory of loss aversion suggests — again, backed by tons of experimental evidence — is that framing choices as losses can make a difference.

loss aversion

Figure 4-10. Loss aversion: wining 5 cents is judged as a lesser option compared to losing the same amount

Suppose you want to sell the latest version of your product. If you give some credit to this theory, you may try AB-testing something like these two alternative messages:

Alternative A: “Buy our amazing new product!”

Alternative B: “Don’t miss the opportunity to buy our new product! It’s a one-in-a-lifetime opportunity!”

Since alternative B frames the communication as a loss, we should expect to have a higher conversion rate on it relative to A. This may sound crazy, but since testing is *relatively* cheap, why not try it? Recall that our aim is to sell more without having to give out our products at a discount.

Loss aversion may be also partly responsible for the success of a strategy commonly used by infomercials. Many times we decide to order because they make pretty clear that if we're not satisfied we can always send it back. But if loss aversion is at play, once you have the product we may be less willing to return it even when the company covers the extra shipping and handling cost.

Levers that change your expectations

We've now covered preferences and restrictions. Preferences guide our choices and behavior and restrictions force us to chose between competing alternatives. What role do *expectations* have, then?

Most of our decisions are made without us knowing the outcome of our choices. Should you date or marry that person? Should you buy coffee or tea? Should you accept that job? If you think about it, all of these choices are made under conditions of *uncertainty*.

Our brain is a powerful pattern recognition machine that allows us to make relatively good predictions many times. But how do we do it? Do we have hardwired the laws of probability in our DNA?

The work of psychologist and economics Nobel prize winner Daniel Kahneman and his coauthor, the late Amos Tversky (and their many

students), has taught us that our brain simplifies many of the computations needed to survive in a world where uncertainty is queen. Two of the most important heuristics or shortcuts that we make are availability and representativeness. And by understanding how they work we can find new levers that affect choices and our business objectives.

THE AVAILABILITY AND REPRESENTATIVENESS HEURISTICS

Recall that heuristics are shortcuts or approximations to computationally hard problems like making decisions under uncertainty. Some times quick and dirty — though vaguely approximate — is better than no decision. That's probably how our brain evolved into a powerful patter recognition machine.

Quantifying beliefs requires gathering evidence and this may be costly. With the availability heuristic we simplify this process by taking whatever evidence is most readily available and use it to approximate likely scenarios. With representativeness we use whatever evidence, even if scant, and extrapolate it. Note that these are shortcuts: if we had more time and resources we could have collected more and better evidence to form our beliefs.

Let's put these heuristics in practice by thinking about the choice to date someone you just met online. Should you take it one step ahead and meet in person? Say that the last time you dated someone from Tinder it didn't go well at all. This is the most recent evidence you have — it is most readily available — and therefore conclude that it is

highly unlikely that the date will end up well and end up staying home.

But you then remember your friend Tom who met his husband using Bumble. If it happened to Tom, why wouldn't it happen to you? You then extrapolate this superb dating experience possibly neglecting the fact that, on average, only a small fraction of dates end up like it did with Tom.

That's the problem with heuristics. Some times they work and some times they don't.

BACK TO OUR BUSINESS OBJECTIVES

Recall that our objective here is to find levers that we can pull with the hope that they positively affect our business objectives. Take the case of advertising. To be sure, the lever is to advertise or not and how much and where.

Most of our potential customers may not need our products now, but when they do they will need to assess the uncertain quality. In this scenario, advertising works by biasing their beliefs about our products, most likely through the availability heuristic.

What about representativeness. If your first product was really good your customers may also be willing to buy the second one. You have built a good reputation that is extrapolated to the second or third product. Or think about issues of corporate governance: if you have created a reputation for disrespecting the most basic ethical standards, customers may extrapolate that to the quality of your product. Choice

heuristics abound so we should use them in our favor (and be extra careful not to put them against us).

Revisiting our use cases

The last pages presented a lot of material. My goal was to point at different sources of inspiration to find levers to achieve our business objectives. This will be more apparent now when we revisit the use cases from Chapter 3.

Customer churn

Just as a reminder, we first framed the business question as a prescriptive one and we now want to start looking for levers to achieve such objective. In Chapter 3 we concluded that our aim is not to minimize churn, but rather to maximize profits from our retention campaign. In this case it may be optimal to let go some customers if it's just too costly to keep them loyal to our brand.

What actions can we take to achieve this business objective? Well, think about it: why on earth would someone want to be our customer instead of our competitors? So let's go back to basics.

Customers generally want three things from the companies they buy from: good-quality products, an affordable price and good customer service in case they need support. Moreover, they'll likely be willing to trade off one or some of these, at least to some degree. Now that we have identified some likely drivers behind our customers' actions

we may now start exploring which levers to pull. This inevitably takes us to the terrain of preferences, restrictions and expectations.

With a price discount we may be willing to sacrifice short term profitability if the long-term impact is positive and incremental. But this is not the only lever that we have. We can also create the perception that switching is costly, creating a loyalty program or highliting some of the least favorable attributes of our competitors products. At least our quality is known to our customers. That of our competitors may be uncertain and we can take that and use it in our avantage (recall detergent commercials).

On a final note, what about some of the consistently inconsistent behavior we mentioned above? Most economists believe these will work only temporarily and eventually your customers (or some competitor) will realize that they're being framed. You may exploit these short-term rents but be careful if you're considering making them an integral part of your business model.

Cross-selling

In cross-selling we are looking for the next-best offer for each of our customers so that we maximize their customer lifetime value. In this sense, our main lever is to offer, or not, each of our products to each of our customers. Note that we may want to include the "no-offer" option as a lever, since we may lose customers by making undesired offers, thereby reducing their lifetime values.

This said, you can use some of the techniques we have described as second-order levers, that is, as levers used to accomplish the actual cross-sale. For instance, the way we communicate and frame our offers can always be used in our favor.

Capital Expenditure (CAPEX) optimization

In this case our levers have already been spelled out from the outset by posing the question in prescriptive terms: the problem of optimizing CAPEX is about choosing where and how much to allocate our budget. At a first level we can choose between investing or not in different projects. At a second level we may wish to fine tune the actual amounts. If feasible we may even disinvest.

Think about opening a new factory. Should we expand our production in one of our current markets, if yes where and by how much, if not, should we enter new markets? Can we close one plant to move capital resources to more profitable venues? These are the natural levers we consider with optimal allocation business problems.

Stores location

Similar to the problem of CAPEX optimization, the problem of choosing where to open new stores can be enlarged: we can play with geographical locations, with the decision to open or not a store and to close existing ones, or even to grow existing locations if physically possible. Since the objective is to maximize our budget’s ROI, all of these are competing levers that we could consider pulling.

Who should I hire

Recall that our objective here is to maximize the incremental returns from hiring. For this we must have a good understanding of the employees impact on our business, which, as we'll see in detail later in [Link to Come] is not obvious at all. But assuming we have this piece of information our decision is then to hire or not, and at what cost. Again, we have a binary lever (hire or not) and one that we can fine tune more granularly (the salary, benefits, emotional salary, work environment and all other levers used by recruiters).

Delinquency rates

The business problem is to maximize the ROI from lending resources to one customer. As such, the three natural levers for this use case are the size of the loan (zero inclusive), the time horizon or maturity, and the interest rate if regulation permits. At this point we can forget about the complexity to optimize all three: we must first start understanding what is the menu of levers we have at our disposal.

But we can be way more creative and test behavioral levers. What if we print children's photos on credit cards? Will that make our customers *more* likely to pay on time their debts? Or talking about communication strategies, can we *nudge* better payment behavior just by sending an SMS with a happy emoticon? Again, testing is relatively cheap: we just need a working hypotheses, the ability to think out of the box and stakeholder's buyin to find less costly levers.

Stock optimization

At the most basic level, we want to leverage how many units of each item we should have in stock. Levers then are just a number, that

could be positive (we need to have stock), zero (current amount is just right) or even negative (move some of these items to other stores since we will never be able to sell them at this location).

Stores staffing

The choice of levers in this problem is again restricted by physical and operational constraints. For instance, is it operationally feasible to make staffing decisions for any given hour of any given day? What about every half hour? Recall that we should have the right number of sales people in each store, in order to maximize our profits or customer satisfaction. But this depends on how many customers we have at any given moment, so depending on the granularity we may always be under or overstaffed.

If we are willing to think out of the box, we might even consider relaxing these operational constraints by “uberizing” our staff: hire people only when demand is high enough.

Key takeaways

- **Once we define a business objective we must go back and consider if it’s actionable**: most times our problems are actionable, but we may have to think out of the box.
- **The problem of choosing levers is one of causality**: we want to make decisions that impact our business objectives so there must be a causal relation from levers to consequences.
- **To understand the relation between actions and consequences we must construct hypotheses**: most times

we don't need to rediscover the wheel, as there's plenty of knowledge out there about how things work or how humans behave. I provided a very quick and incomplete overview of some findings that I've found useful to think about these problems.

- **Hypotheses fail but embrace the process:** many times we start with a theory about causes and consequences only to see it fail during testing. That's fine. It's part of the process. Embrace it and guarantee your team learns from these failures.

Further Reading

Physical levers are problem-specific so my suggestion here is to ask your more technical colleagues for some reading suggestions that can help you gain at last some general knowledge of what can be achieved or not.

The human levers discussed here have been thoroughly studied by social scientists including economists, psychologists and sociologists. I'd recommend starting with an introductory textbook on microeconomics since most of our business decisions have some economic foundation. Search online for good reviews you trust, and just in case you need some extra help, I agree with most of the recomendations here, except for the Mas-Colell, Winston and Green suggestion (too much math and less intuition that I would want at an introductory level). Any of the other books will provide a detailed account of the role that our preferences, expectations and restrictions have always in the context of rational choice. I personally enjoy how

David Kreps, Professor at Stanford GSB, explains these very technical topics to the more general public.

Books on behavioral economics will give you some extra background on less-than-rational choices. One of my favorites is Dan Ariely's *Predictably Irrational* but a safe choice will also be Daniel Kahneman's *Thinking, Fast and Slow*. I certainly recommend Ariely's book as he provides many examples of levers we might pull to improve our business that we would never have expected to work. If you want to learn to think out of the box this is the type of literature I'd recommend.

Becker and Murphy's *Social Economics* is still a good reference on choice under the social umbrella but another classic, full of intuition is Thomas Schelling's *Micromotives and Macrobehavior*. A highly technical and encyclopedic treatment of the economics of social networks can be found in Matthew Jacksons's *Social and Economic Networks*. The *Handbook of Social Economics* edited by my Ph.D. adviser Alberto Bisin and other experts in the field provide possibly outdated literature reviews on most topics explored in these areas, as well as on the economics of how culture is transmitted across and within generations.[7]

Finally, strategic behavior and game theory are topics of their own. You can start with an introductory textbook that is less technical but provides a lot of intuition. Dixit and Nalebuf's *The Art of Strategy* or an introductory textbook by Ken Binmore's such as *Fun and Games* or *Playing for Real* might do the job.

1 In this setting *server* is the person or machine responsible for serving each customer (like a cashier) and not a computer server.

2 Rationality has to do with consistency of choices which I will not use or claim at all.

3 Unless you want to capture market share that would supposedly increase long-term revenues, but this is a different business objective. In this case you may consider operating at a loss in the short-term but you must then be optimizing the net present value for your longer-term profits.

4 See *https://www.theglobeandmail.com/news/national/time-to-lead/why-your-dna-is-a-gold-mine-for-marketers/article6293064/* for an example.

5 We also talked about two-sided platforms in Chapter 3.

6 I only hope it's not negative! I would then have to pay you to read it.

7 Handbooks in Economics provide up-to-date reviews at the time of publication so by now some of those reviews may be outdated.