APPLIED BUSINESS ANALYTICS

Integrating Business Process, Big Data, *and* Advanced Analytics

NATHANIEL LIN

"When I first read Nathaniel Lin's Applied Business Analytics, I thought, 'I wish I had written this.' The points Nathaniel makes about analytics deciders hit a unique target—managers who today don't realize what it takes to drive business using data. Having 'analytics deciders' engaged at the 'intersections' makes it possible for innovation to flow throughout the organization, like traffic cops in downtown Manhattan. This is one of those books that you keep on your desk and refer to as you build an analytics team. Here are all the things I loved about this book: the Advanced Analytics Layer (it has immediate applicability); the case studies (which are practical); and the hands-on approach with KNIME (the real-world experience)."

—**Theresa Kushner,** Vice President Enterprise Information Management, VMware and former Sr. Director of Customer Intelligence, Cisco

"Nathaniel has done a terrific job of capturing the key components needed to build high-performing analytics to drive business growth. This is a 'must read' for senior executives involved in analytics for decision-making."

—**John F Carter,** Senior Vice President, Analytics, Insight & Loyalty, Charles Schwab

"Making analytics mainstream: There are scores of books on the application of analytics to organizations, however, Applied Business Analytics is unique in two respects. First, it identifies why organizations fail to truly leverage analytics to achieve transformational outcomes. Second, it proposes a framework in which business processes are modified to accommodate 'analytics' through a continuous dialogue, rather than isolated analytical experiments. Forward-looking business leaders who want to truly leverage Big Data and analytics to transform and lead their organizations would do well to take to heart the core message of this book."

—**Prat Moghe,** CEO of Cazena and former Senior Vice President of IBM Netezza

"Business analytics has developed into an area of intense focus for many organizations, yet its successful application remains more art than science. Nathaniel Lin's Applied Business Analytics represents a significant step forward in remedying that situation by enabling business leaders to create the teams and business processes required to succeed with analytics. Business analytics is by its nature a deeply technical topic, and Dr. Lin does not skimp on the technical details, including meaty yet approachable hands-on exercises. Applied Business Analytics keeps the technical discussion where it belongs: in clear service of well-articulated and relevant business problems that provide the necessary context for analytics projects to create meaningful value. No mere how-to guide, Dr. Lin's book provides a complete picture of the analytical organization, as well as the conditions and processes required to develop and enable analytics deciders, the data-driven change agents who propel organizations to move from gut-based decision making to rational, evidence–based decision making. Packed with numerous real-world case studies and war stories, Nathaniel Lin's Applied Business Analytics is a solid starting point for any executive seeking to transform an organization with analytics."

—**Adam Ferrari,** CTO Crisply and Strategic Advisor Big Data and Vice President of Product Development, Oracle, Former CTO, Endeca Technologies

"Leaders who are establishing a data-driven culture will benefit from the lucid analytical process and methodologies that Nathaniel demonstrates in Applied Business Analytics. The chapter about Analytics Eco-system includes a fantastic comparison of different team structures that exist today."

—**Jane Chen,** Head of Analytics Center of Excellence, Analog Devices

"I am positively impressed by Nathaniel's book. This book has a unique position in the marketplace, standing out as an insightful perspective balancing an 'evangelist book' and 'technical how-to cookbook.' I appreciate Nathaniel's thought leadership and guidance. "

"Applied Business Analytics is 70% business and technical context discussion and 30% of how to be effective in organizational framework. The most challenging part of analytics leaders in today's world is still the conflict or tension between business drivers and technical details. That is why 'analytics deciders' are critical for maximizing analytics' functional impact on businesses. Balancing 'technical details (how)' with 'why analytics is critical to your success' and 'what result you can (expect to) achieve by building analytics capabilities' is challenging. Nathaniel's personal stories, case studies, and real-world success are helpful. I believe this book can be a great reference guide for emerging analytics leaders."

—**Gary Cao,** Vice President, Information Services & Analytics Strategy, Cardinal Health, Inc.

Applied Business Analytics

This page intentionally left blank

Applied Business Analytics

Integrating Business Process, Big Data, and Advanced Analytics

Nathaniel Lin

Editor-in-Chief: Amy Neidlinger
Executive Editor: Jeanne Glasser Levine
Operations Specialist: Jodi Kemper
Cover Designer: Chuti Prasertsith
Managing Editor: Kristy Hart
Project Editor: Deadline Driven Publishing
Copy Editor: Nancy Sixsmith
Proofreader: Deadline Driven Publishing
Indexer: Angie Martin
Compositor: Jake McFarland
Manufacturing Buyer: Dan Uhrig

Upper Saddle River, New Jersey 07458

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact international@pearsoned.com.

Printed in the United States of America

First Printing December

ISBN-10: 0-13-348150-6
ISBN-13: 978-0-13-348150-1

Pearson Education LTD.
Pearson Education Australia PTY, Limited.
Pearson Education Singapore, Pte. Ltd.
Pearson Education Asia, Ltd.
Pearson Education Canada, Ltd.
Pearson Educación de Mexico, S.A. de C.V.
Pearson Education—Japan
Pearson Education Malaysia, Pte. Ltd.

Library of Congress Control Number: 2014952168

To my parents Tah-Fu and Paw-Chu Lin, who have spent years encouraging me to do something special. Even though this book is nothing special, it is a token of my appreciation for their love and nurturing since they brought me into the world 60 years ago. Last, to the one who really matters, I would use the same dedication from my PhD thesis:

"…Jehovah, who stretches forth the heavens and lays the foundations of the earth and forms the spirit of man within him." —Zechariah 12:1

This page intentionally left blank

Contents

Foreword

This book is well-timed for the current business analytics zeitgeist—the spirit of the times—in a variety of ways. Analytics and big data have become accepted tools for business strategy and competition. In surveys, virtually every manager agrees that they are important resources for business. If you are not a believer in data and analytics as guides to decision making, it is becoming embarrassing or politically incorrect to say so. This is all good news, at least from my perspective.

The only problem with this broad-scale consensus is that many managers and organizations still lack the skills and understanding to make analytics work for them. Lin's book is directed at just this issue. As he says, there have been a variety of books that attempt to get managers excited about analytics and textbooks that delve into the details of analytics without much sense of the business context. This book is a bridge between those extremes. It's for managers who already believe that analytics are important, but who don't yet know enough to practice them personally.

This transition is important and, indeed, necessary. In sophisticated companies, it is becoming difficult to draw the line anymore between analytics and the way work is done. Business is analytics, analytics is business. This book is designed to enable regular business people to create and consume analytics, even advanced ones. The detailed modeling and analysis of data may continue to be left up to professional "quants," but the framing, interpreting, and communicating of the analytical work will be a highly collaborative effort among "deciders," as Lin puts it, and quantitative experts.

Another way in which the book fits with the contemporary analytical mood is its focus on business and analytical processes. Lin correctly points out that analytics are increasingly being embedded within key business processes. In that sense, they are becoming somewhat invisible; as I write today, in fact, Gartner named "advanced, pervasive, and invisible analytics" as one of the "Top 10 Strategic Technology Trends for 2015." This invisibility makes the importance of management understanding even greater; if your business is heavily dependent on analytics in the underlying technology and process infrastructure, someone needs to constantly ask intelligent questions about underlying assumptions, methods, and outcomes.

To continue the process orientation, many of the topics and insights in this book are structured in terms of an analytics process that ranges from data preparation to data analysis and interpretation. The idea of a process fits the current analytics environment well. For too long, analytics have been pursued as a slow, one-off, painstaking process that would be difficult to institutionalize. Today, however, many companies are beginning to create "analytical factories" that emphasize speed, scale, and repeatability. The means to do this is to view analytics as a process with clear steps, handoffs, and modules. That is the focus in the book, and it's buttressed by examples using the open source analytics tool KNIME, which is itself structured in terms of processes. If we adopt Lin's model of thinking about analytics in process terms, our analytics will have higher productivity and impact.

A final linkage between this book and the current set of analytics themes is the focus on both big data and small. Most current books are on traditional, small data analytics or big data, but few cross that chasm to address both. Lin moves comfortably between analytics on traditional data (for example, marketing propensity models) and such big data types as social media and text from online product reviews. It's clear to me that before long, we won't use the term "big data," but we will certainly want to deal with unstructured, large-volume data. In any case, you'll be well-prepared for a world that integrates and analyzes all types of data if you read this book. Lin helpfully supplies a list of likely data elements that an organization might employ for typical analyses in different business areas.

This book is not only up-to-the-moment, but extremely comprehensive. There is scarcely an area of business analytics that is not discussed. You may want to read it quickly when you acquire it, and then use it as a reference book as you encounter a variety of analytics situations over the next several years. The ideas are both classical and contemporary, and they won't go out of fashion or relevance for a long time.

Thomas H. Davenport,
Distinguished Professor, Babson College
Author of *Competing on Analytics and Big Data @ Work*

Acknowledgments

I would like to thank the following people: My wife, Lu, for her confidence and encouragement. Alejandro Simkievich, a good friend and fellow Sloanie, who wrote Chapter 6. Any errors or omissions are solely mine. I want to thank also *FT Times*, and Victor for reading the draft and providing valuable suggestions.

About the Author

Dr. Nathaniel Lin is a recognized leader in marketing and business analytics across various industries worldwide. He has over 20 years of frontline experience applying actionable advanced analytics strategies to the world's largest companies in technology, finance, automotive, telecommunications, retail, and across many other businesses, including IBM, Fidelity Investments, OgilvyOne, and Aspen Marketing Analytics.

Nathaniel is currently the Chief Customer Insights Officer of Attract China. He is leading the efforts to develop leading edge Big Data Analytics technology and knowledge assets to deliver unparalleled values to Chinese travelers and U.S. clients. Nathaniel is widely recognized as an expert, teacher, author, and hands-on leader and senior executive in the application of data and advanced analytics in a wide variety of businesses. He is also the Founder and President of Analytics Consult, LLC (www.analyticsconsult.com). He leverages his rich and unique expertise in business analytics to help companies optimize their customer, marketing, and sales strategies. Together with his team, Nathaniel serves as a trusted strategic advisor to senior management teams. He is frequently invited as the keynote speaker in analytics events and advised over 150 CEOs in the U.S. and aboard on analytics and Big Data issues. He was invited by WWW2010 as one of the four expert panelists (together with the heads of Google Analytics, eBay Analytics, and Web Analytics Association) on the Future of Predictive Analytics.

As a recognized analytics expert, Nathaniel has partnered with Professor Tom Davenport to benchmark analytics competencies of major corporations across different industries. He also demonstrates his passion in cultivating future analytics leaders by teaching Strategic CRM and Advanced Business Analytics for MBA students at the Georgia Tech College of Management, Boston College Carroll School of Management, and Quant III Advanced Business Analytics at Bentley University.

Nathaniel holds a PhD in Engineering from Birmingham University (UK) and an MBA from MIT Sloan School of Management.

Preface

I have to admit that this book turned out to be much harder to write than I had envisioned. Having spent decades applying analytics and building and leading analytics teams, I thought it would be simply a summary of what I had learned and done before. Unfortunately, I find myself constantly struggling to find the “Goldilocks” approach—am I covering too much technical detail or too little? The latter, in my opinion, is worse because the lack of details might prevent business readers from doing the hands-on analytics exercises. However, the former with too many technical details would likely discourage business readers from reading the book altogether!

After several rounds of revisions, I hope I have managed to strike a balance to ensure the original objective of writing an analytics hands-on book for business readers has been met. However, in the event you find the content either too technical or not sufficiently technical, I beg for your indulgence. My advice is that you simply skip over the offending sections and go onto the next section. I am sure there are enough examples and business cases in this book that any serious business reader aspiring to be analytical would find relevant and helpful.

Why Another Book on Analytics?

Despite the many titles on analytics, a quick survey reveals that most broadly fall into two categories. One group focuses on promoting how businesses benefit from analytics. These are usually written by "evangelists" promoting the virtues of analytics. At the opposite end are those books written by "quants" for other analytics "quants." These contain too much technical detail for business readers. Few books are written to bridge the gap. There is a need for a book with hands-on applications of business analytics that solve real business problems that business readers can understand, practice, and use. I often hear the laments of business leaders: "I am sold after reading books promoting the power of analytics, but *none can show me how it is done and how to actually apply it in my business* without requiring me to go back to school and getting an advanced degrees in analytics and statistics!" If you share the same sentiments, then this book is for you.

How This Book Is Organized

Before "getting your hands dirty" with hands-on exercises, this book devotes the first four chapters to the basics, that is, the data, analytics tools (from simple to leading edge), and their relationships to the various business processes. The book continues with how to embed analytics and integrate the findings in existing or new strategic processes. With this in mind, the chapters are organized as follows:

- Chapter 1, "Introduction," is an introduction to analytics. It shows how analytics as the refining of data will transform business the way refining of oil transformed the twentieth century. It also addresses definitions, potential areas of applications, some vital lessons learned, how analytics are positioned in the business context, and how analytics are integrated into the business process.
- Chapter 2, "Know Your Ingredients—Data Big and Small," discusses everything you need to know about the basic ingredients of successful business analytics. It defines the differences between Big Data and the various types of data that may be

useful to business analytics. It also describes data formats and the increasing importance of poorly or unstructured data to business insights and how to handle them.

- Chapter 3, "Data Management—Integration, Data Quality, and Governance," provides everything you need to manage and analyze data, from processing diverse data, to the importance of governance on data security and privacy, to good data versus bad data, to quality and latency, and to where to go beyond internal data sources.
- Chapter 4, "Handle the Tools: Analytics Methodology and Tools," discusses the most commonly used analytics models and tools, and provides hands-on exercises for some of the most commonly used analytics modeling techniques so you can see and feel how they work within real business cases. Its coverage spans basic models such as regressions, clustering, decision trees to the more sophisticated Big Data analytics such as text mining, and sentiment analysis on real data including online reviews on Amazon Fine Food and TripAdvisor hotels. A value prediction methodology using Ensemble Regression Tree model is also presented.
- Chapter 5, "Analytics Decision-Making Process and the Analytics Deciders," elucidates the differences between the conventional and the analytics decisioning process. It starts by exploring how conventional decision processes often falls short. It also describes the analytics decisioning process, known in this book as the BAP (Business Analytics Process), and how the right "analytics deciders" are needed to work in the "intersection" between silos to ensure the BAP's success. It ends with notes on how to become, identify, recruit, and retain analytics deciders.
- Chapter 6, "Business Processes and Analytics," discusses the business processes that can benefit from analytics. It also covers

today's ERP (Enterprise Resource Planning) that is fueled by analytics. It includes CRM (Customer Relationship Management), SCM (Supply Chain Management), Financial Management, PLM (Product Cycle Management), and Human Capital Management.

- Chapter 7, "Identifying Business Opportunities by Recognizing Patterns," explains why the current way of dividing customers into "segments" just won't cut it. It tells you how to leverage advanced analytics tools to detect patterns of behaviors, events, trends, topics, preferences, and other critical factors that can impact business. Finally, it discusses pattern detection and how the group behavior of customers may help marketers plan more effective engagements using insights from the Market Basket Analysis and customer segment persona discovered using analytics.
- Chapter 8, "Knowing the 'Unknowable,'" discusses those topics and questions that are often deemed unknowable or unique (UU). These may involve little or limited data, predicting individual behavior in real time, and determining causality and lever settings for complex situations. This chapter provides real-life business cases and actual analytics workflows to illustrate how enlightened businesses can know and leverage analytics in these apparent "unknowable" events, thus, gaining a competitive edge over the competition. A novel methodology is also described to predict customer wallets and how to use that to form effective strategies.
- Chapter 9, "Demonstration of Business Analytics Workflows—Analytics Enterprise," uses the proper ingredients and tools to answer the top business questions except CRM (Customer Relationship Management), which is discussed in Chapter 10. You are encouraged to download and modify the workflows so that you can try to run them and adapt them to your own needs.

- Chapter 10, "Demonstration of Business Analytics Workflows—Analytics CRM," describes how to use the proper ingredients and tools to answer the top business questions surrounding management of customer relations. It provides tips on how to intimately know the customers—who they are, their values, profiles, persona, where and how to find them, and their needs and wants behavior—all through the use of analytics. Important insights such as customer purchase preferences and patterns, loyalty and impending churn, and how new customers may be acquired cost effectively and won back using today's social, mobile, and other media. You are again encouraged to fill out the BAP process involving the analytics modules and to download relevant workflows to try them on your business with your own data.
- Chapter 11, "Analytics Competencies and Ecosystem," discusses your current competency, how to move to the next level, and the importance of a healthy analytics ecosystem. It also provides an example of an enterprise analytics strategic plan, including organization structure, talent management, roles and responsibilities, and tips on how to run an effective analytics organization and to leverage external partners in the analytics ecosystem.
- Chapter 12, "Conclusions and Now What?" provides a summary of learning and suggestions for the next steps you should take toward putting together a strategic roadmap for achieving analytics competency and maturity. A list of recommendations that readers can adopt to achieve analytics competencies in the various areas covered in this book is also included.
- The Appendix contains material that is important but may be too detailed for inclusion in the main body of the book. This includes the basics on the analytics tool KNIME and descriptions of some of the KNIME nodes. You are encouraged to

augment it by consulting the introductory texts on KNIME.com and YouTube.

After Reading and Working Through This Book

After reading this book, you should be able to:

- Possess a sufficient understanding of the entire analytics process.
- Acquire analytics vocabulary and basic knowledge of how analytics models work.
- Explore business analytics models and be able to modify or build real-life sophisticated analytics models.
- Navigate across business and analytics processes. Drill down or aggregate up at any level of the workflow and follow the analytics process without gaps in data, results, and insights.
- Better communicate with the analytics team.
- Be a contributing member of a rapid prototyping analytics team to provide business directions on demand.
- Detect actionable business insights from seemingly insignificant analytics details ignored by data and model "crunchers."
- Ensure business focus and alignment of final deliverables, and ultimately be an effective analytics leader and decider.

Bon voyage!

1

Introduction

Raw Data, the New Oil

"Data is the new oil"[1] has become a common mantra among today's Big Data proponents. However, Clive Humby, the founder of dunnhumby (the well-known retail analytics arm of Tesco Supermarkets in the UK), has also said that "[data] unrefined cannot really be used." For this reason, many skeptics argue that Big Data is not new and is really just a fad. Before you are convinced of the importance or insignificance of Big Data, let's try to use the "data as oil" analogy and show you the remarkable impact oil has had on every aspect of life across the world and (by inference) data.

If you look at a photo taken in New York in the late nineteenth century, you don't see cars and buses. Instead, you see commuters on horseback or in horse-drawn carriages and omnibuses. In fact, New Yorkers living in a city with a million people in the late nineteenth century enjoyed their night life lighted not by using electric or gas lamps, but by burning whale oil. In fact, at its peak, more than 10 million gallons of whale oil per year[2] (see Figure 1-1) were consumed across the United States. Because this usage was one of the main causes of the near-extinction of whales, Samuel M. Kier may be said to have saved the whales by successfully refining the crude oil into usable fuels in 1851. Of course, he also helped the emergence of entrepreneurs such

1 http://ana.blogs.com/maestros/2006/11/data_is_the_new.html
2 http://archive.org/stream/historyofamerica00tow#page/126/mode/2up

as Rockefeller, who started out by selling kerosene as one of the first products of the oil-refining process.

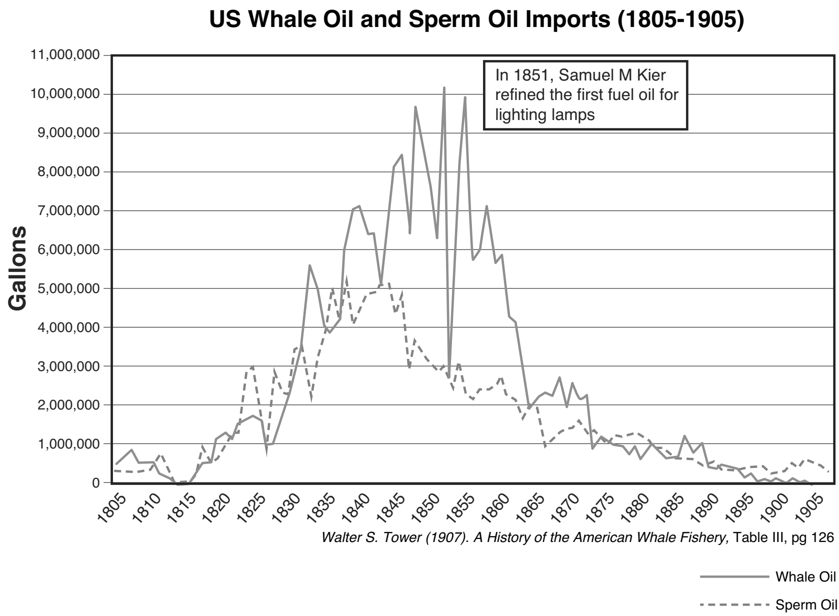

Figure 1-1 U.S. whale oil imports (1805–1905)

Over the next century, the impacts of oil literally reached the sky when U.S. astronauts landed on the moon in 1967. It is fair to say that oil transformed every facet of human culture and existence in the last 100 years. Just as you couldn't ask someone in the early 1900s to predict how oil would transform the world in 100 years, it is just as difficult to imagine today what a transformational role data and analytics might have in the rest of the twenty-first century.

Data Big and Small Is Not New

Although gasoline was new, crude oil was not "new" in the 1900s. In fact, crude oil has been known since prehistoric times. For 7,000 years, humans have been trying to find uses for crude oil.[3] In fact, one of its main uses in the United States in the nineteenth century was as folk medicine. Likewise, data (some may argue even Big [voluminous] Data) has been around for decades. What is new today is our ability to refine it to uncover novel insights and discover new applications.

In this book, I will show how business analytics (BA) is today's way of refining raw data, big or small, into strategic business opportunities. In many schematics for Hadoop and its popular Big Data-related tools such as Hive and others, these Big Data management tools are labeled the "Big-Data Refinery."[4] More appropriately, the Hadoop Distributed File System (HDFS) is more like an oil storage and distribution system than the actual refinery. The actual refining process happens right after the data has been properly stored and then subjected to a refining process, which is defined in this book as the applied business analytics process (BAP).

Definition of Analytics

At this juncture, we should define what the terms "analytics" and "business analytics" mean as they are used in this book. The standard dictionary definition of analytics, which is "the method of logical analysis" with a first-known use in 1590, clearly needs updating. I have also seen that the understanding of analytics in some businesses have not progressed much further. During an interview in 2010 of senior business executives with Professor Tom Davenport in benchmarking analytics competencies of major companies, I was shocked to hear one SVP of analytics consulting asserting that "Everyone in my team is doing analytics as they all work with numbers." Unfortunately, "working with numbers" does not make one an analytics practitioner.

---

3 http://ocw.mit.edu/courses/political-science/17-906-reading-seminar-in-social-science-the-geopolitics-and-geoeconomics-of-global-energy-spring-2007/lecture-notes/lec02_02152007.pdf

4 http://hortonworks.com/blog/big-data-refinery-fuels-next-generation-data-architecture/

A horse carriage has four wheels just like a car, but it is clearly not a car. We did not find any evidence that any "analytics consultants" were doing any analytics as defined here.

In this book, I define "analytics" (or "business analytics") this way:

- **More than just numbers**—Analytics is more than just working with numbers and data to find and report observed correlations and/or statistical distributions.
- **Knowledge and results-centric**—Business analytics is focused on the process of discovery of actionable knowledge and the creation of new business opportunities from such knowledge.
- **Tools-agnostic**—Analytics can use any computation or visualization tools from statistics, computer science, machine learning, and operational research to *describe* and *recognize* patterns, *seek* and *validate* causal relationships and trends, and *predict* and *optimize* outcomes.

To refine data into "business solutions" using the oil analogy, today's analytics "Rockefellers" must be able to do the following:

- **Understand the data**—Know the properties and nature of crude oil.
- **Understand the technology**—Know the different ways of refining crude oil.
- **Understand needs**—Know the human needs that are currently either barely met or unmet by today's technology.
- **Be creative**—Possess creativity in linking newfound properties of refined crude oil to meet known needs, and even create new markets out of new needs such as synthetics in materials, cosmetics, and pharmaceuticals which did not exist before.

Let's take a look at some of these business opportunities for applying analytics.

Top 10 Business Questions for Analytics

In my more than 15 years of direct involvement in applying analytics to business, I have found many important and difficult business questions that only analytics can answer. Today, the list is growing longer as more companies are investing in and applying analytics to new problems.

Although some of these questions can be answered from simply "slicing and dicing" raw data, some are so novel that many people in business might label them as "unknowable." I will give you one such example; you can then come up with your own list of unknowable questions to see whether analytics can help answer them.

One example is to predict the actual wallets of individual customers—and by extension the business share of their wallets. It was commonly viewed by business as unknowable because you can see only the purchases your customers make at your store, not their purchases at other stores.

Other examples of unknowable questions include what online visitors do when they leave your sites, what your customers need when they can't even describe it to themselves, and what to ship to the customers even before they start to consider buying and ordering. The former is the basis of the current ad exchange networks, in which third-party cookies are exchanged to let websites know where the visitors may have been before coming to your site. The last example is actually what Amazon is contemplating launching soon.[5]

Let's look at the first example. With a partial view, the true wallet size each customer possesses is elusive and can be inferred only through surveys. However, such survey data is often unreliable and hard to extrapolate to other individual customers. However, as to be explained in more detail in Chapter 8, I have directly used the best

5 "Amazon says it can ship items before customers order." http://www.usatoday.com/story/money/business/2014/01/18/amazon-anticipates-orders/4637895/

customers' transactions as "lenses" to predict wallets for both business-to-business (B2B) and business-to-customer (B2C) customers. These wallets were successfully tested and deployed in many business problems.

Here is my list of top 10 questions that I feel can best be answered by the use of analytics. The questions have been divided into four sections: Financial Management, Customer Management, HR Management, and Internal Operations.

How analytics can answer the questions might not be apparent at the moment. One of the aims of this book is that you will be able to either find the answers from the examples in this book or be able to customize your own BA workflows for specific industries or business situations. Once comfortable with modules and processes, you might want to go to the list of questions in Chapters 9 and 10 to practice putting together your own analytics workflows to answer the questions.

The following sections discuss my top 10 strategic questions and tips on how analytics can be applied to provide answers.

Financial Management

Are your business and financial goals the right ones?

For a business to set the right goals, it is best to start from the customers and the current market. Because the entire market is rarely addressable due to its product focus or competitive landscape, a business needs to define the markets in which it wants to play. Once decided, analytics can then be used to predict the size of the addressable market and combine with revenues to obtain the respective market share.

Based on customer-level analytics, a business can predict what investment levels are needed to reach specific goals. Again, a business should start at the most granular levels and aggregate up to generate higher-level (business unit [BU] and enterprise) views. Some of the costs may involve realignment of business and product offerings.

Where are the areas of major opportunities?

The conventional top-down approach to market sizing is often not divisible at lower geographic or customer group levels. It has been a formidable exercise, often rife with contention when it was time for the BUs to assign annual sales targets at the various units based on the current market size predicted by the marketing intelligence (MI) team (at IBM during my time there) using the top-down approach. It felt like a prefab house being cut up and retrofitted—often messy and not a pretty sight.

However, it is just the opposite when you start with the smallest building blocks. You can build anything provided that you have a plan, which is the case when the addressable market is produced from predicting individual customer wallets. Armed with the customers' current spends and propensities to spend more, a business can find ways to more appropriately engage and entice different customers to shop more. Summing the wallets over the appropriate business units, the business can size the expected opportunities and identify the major revenue opportunities.

Are your investments adequate?

When the opportunities are identified, the natural next question is whether you are spending enough to win. Before knowing whether the investments are adequate, it is important first to predict the effective "levers" for increasing product wallet shares among the different groups of customers. Once the model results are validated with holdouts, the incremental causal effects of the levers need to be tested in-situ. A properly designed, multifactorial test-and-learn experiment can help determine the optimal combination of levers to give the maximum impacts. With the levers known, the business can then ascertain the costs of implementing the levers.

With the costs and impacts of the various levers determined, the business can put together a pro forma simulator to model the level of investments needed to move the needle (that is, per-unit wallet share gain). Once the optimal return on investment (ROI) has been decided, the requisite costs then need to be checked against investments in the annual strategic plan to ensure that adequate funds have been allocated to achieve the goals.

Customer Management

Do we know enough about our customers?

The conventional way to know customers was to hear directly from the voice of the customers through focus groups and surveys. Unfortunately, focus group and survey results cannot be easily applied to each individual customer. They also take time to organize and are costly to run regularly.

Given the richness of today's customer data, analytics has been shown to be capable of generating detailed individual and actionable customer knowledge. This knowledge includes predicting individual customer's propensities to buy from your brand and product, how much customers buy from you and from your competitors, and how valuable customers are to your business over their lifetime. Analytics can also perform customer multidimensional segmentation (behavior, demographic, attitudinal, and value) and develop segment persona. From the persona, the business can then devise effective strategies to better serve the segment, improve customer experience, and increase customer satisfaction and loyalty.

Lapses can happen despite best efforts, so the business should constantly predict and monitor customers' propensity to churn. The business should try to understand why they churn, what went wrong, and how to fix the problem. Any fixes should also be tested with control groups to make sure they are the right options. During and after implementation, those customers with high churning propensities should be surveyed to ensure that their issues were successfully resolved.

All vital customer knowledge assets should be shared, managed, and leveraged for all marketing, sales, operations, product development, and finance initiatives.

What actionable customer insights do you have?

The chief marketing officer (CMO) of a well-known steakhouse chain client received a customer insight report taken from surveys from a top management consulting firm. It showed that the client has more than 15 million customers that like its food, but they rarely visit the restaurants. It was a great insight, but where and how can they reach and entice such customers to come and dine more frequently? They can deduce from surveys what these customers might look like and devise broadly targeted mass media ads to entice them. They can't be sure whether the strategy would work across their chain when they test at a particular region or city. There are simply too many things at work to determine the effectiveness with sufficient certainty.

To make such insights actionable with analytics, the individual diners' receipts should be linked to the diners—either through a frequent diner program or by their name, credit card, or checking account number. The customer insights developed for the previous question can then be used to formulate campaigns and strategies to test and validate the effectiveness of the solutions derived from analytics insights. With a properly designed test-and-learn methodology, the incremental effects of levers can be ascertained before full-scale rollout. This analytics-driven approach would save costs and also reduce risk of failure.

Are we focused on the right social and mobile issues?

Social and mobile uses and data should be viewed as part of the data analytics value chain. Social and mobile are just part of the media. The customers' social, online, and mobile behavior data must be linked to target behaviors in terms of purchases and conversions. Once integrated, the issues must be germane to the BA process, actionable, aligned with, and have significant impacts on the critical strategic goals.

Is the critical knowledge from analytics properly managed (that is, captured, stored, shared, and reused) as an enterprise asset?

As discussed in Chapter 4, analytics innovation and knowledge tend to occur at the key intersections in which different disciplines, roles, functions, and goals "collide." To ensure that such knowledge can flourish and be captured, stored, shared, and reused as an enterprise asset, a process such as BAP (also see Chapter 4) should be adopted. Beyond the process, key "analytics deciders" who will be defined later must be present at the intersections. For now, an analytics decider is defined as someone who is highly proficient in both business acumen and analytics. A knowledge management IT system and a functioning "analytics sandbox" support the analytics deciders with well-defined rules, ensuring a safe and collaborative environment for creative and productive collisions.

HR Management

Do you have the right strategy for recruiting, managing, and retaining analytics talent?

In my years spent in building and managing analytics teams for big and small companies, I rank organizational and personnel issues as the number-one cause of failure in applying analytics. Many efforts failed because the analytics had either a limiting or conflicting reporting structure; the wrong compositions of talents and skills in the business, data, modeling, and strategy functions; or simply the wrong people (nonanalytics deciders) in key leadership roles within the intersections. Alternatively, companies' analytics efforts failed as they tried to remove the collisions at the intersections by delineating clear lines of responsibility. Unknowingly, these intersections were eliminated as everyone tried hard to stay within their own areas of responsibilities and roles. (Chapter 11 addresses some of these organizational and personnel issues.)

Internal Operations

Are your business processes driven with insights from predictive customer analytics?

Although companies started to use more predictive customer analytics in the past decade, many did not progress beyond tactical campaign targeting. As a result, the effectiveness of analytics can deteriorate over time. Sadly, many pioneers in analytics applications have been stuck doing the same things or have relegated analytics to just one of the support functions. More progressive companies have quietly embedded predictive analytics in their various business processes, and their CEOs rely on analytics for answers to their strategic questions.

For example, during a benchmarking project of companies' analytics competency, I was told that the CEO of one of the major online retailers, while preparing the annual report to the Wall Street analysts, discovered that revenues went up but the number of visitors went down. He wondered whether he should pose it as a positive or negative indication. Instead of consulting with the CFO or anyone within his senior leadership, he instead asked the director of analytics to come up with an answer. The CEO was satisfied with the reply and used the analytics recommendations during the call with the analysts.

I firmly believe that in the next few years, companies will win or lose by how much they can integrate and embed business analytics into their various business processes (as explained in Chapter 6).

Are your sales efforts fueled by analytics insights?

One of my earliest successes in applying analytics was not with marketing efforts, even though my team was part of the IBM MI team; it was to help IBM sales efforts. Since then, I find that sales efforts can be greatly enhanced by analytics insights such as leads generation, leads prioritization, sales force optimization, and telesales/call center performance analytics.

Vital Lessons Learned

Despite the power of BA, my experience in applying analytics in many companies across businesses has its share of frustrations that taught me many vital lessons. These lessons hopefully will aid you as you implement the BA Process for the first time and hopefully avoid wasting valuable time, resources, and opportunities.

Use Analytics

According to a 2013 survey of approximately 500 CMOs,[6] spending in analytics is likely to increase by 66 percent over the next 3 years, yet only 30 percent of the companies surveyed are using analytics. This is a decrease of about 19 percent (from 37 percent) from the previous year! Professor Christine Moorman, who directed the CMO survey, said "I think the mistake that a lot of companies make is that they spend on marketing analytics, but they don't worry enough about how you use marketing analytics." I have witnessed the same phenomenon in many companies.

Reasons Why Analytics Are Not Used

The common causes of senior management's lack of attention to how analytics is used are these:

- **Lack of understanding and trust**—Advanced analytics works, but few executives know how to use it, and fewer have direct experience working with it. As a result, management cannot vouch for or articulate analytics values when tough decisions are to be made among senior executive teams.
- **Overreliance on what they know**—Executives rise through the ranks doing things they know best. Unfortunately, analytics is likely not one of them. So when competing priorities demand their attention, they tend to rely on what they know, not on analytics. When the stakes are high, they often fall back on something they have prior experience with, not analytics (because it represents uncertainty and risk).
- **Lack of strategic vision of analytics' uses**—The true effectiveness of analytics is not in tactical applications such as

6 http://cmosurvey.org/files/2013/02/The_CMO_Survey_Highlights_and_Insights_Feb-2013-Final2.pdf

targeting or reporting. But that is exactly what some companies have been stuck doing, sometimes for more than a decade. Strategic applications cut across organizational silos and demand changes and even transformations of business, which might mean higher risks. Few executives would risk their careers to champion for the use of analytics under such situations.

- **Lost in translation**—Definitions and vernacular often differ between the senior executives and their analytics team, so business focus tends to get lost when traversing through the layers of management and finally reaching the modelers. In the reverse direction, significant customer insights might be ignored or filtered as "bad results" or "results not pretty enough to show the boss" by the analysts or mid-level management who might not interact so are probably not in tune with the senior executives.

The only way to mitigate this, I believe, is to ensure that there is a tight linkage between business and analytics. Hopefully, by taking the executives through the analytics process and enabling them to see analytics at work, they will gain a better understanding and be able to trust and use analytics!

Linking Analytics to Business

Instead of viewing analytics as a tool or a method, this book focuses on analytics as a process.

Business Analytics Value Chain

Rather than focusing on a single stage of the analytics value chain shown in Figure 1-2, this book takes a holistic view on the entire value chain (or value "ring" in this case) for continuous value creation. The focus is on the final business outcomes and their continued improvements using scientific test-and-learn methodology. To achieve sustainable wins over time, it is important that the process be run in a continuous fashion (this is the BAP that will be described in more detail in Chapter 5). Here's a brief description of the five major components:

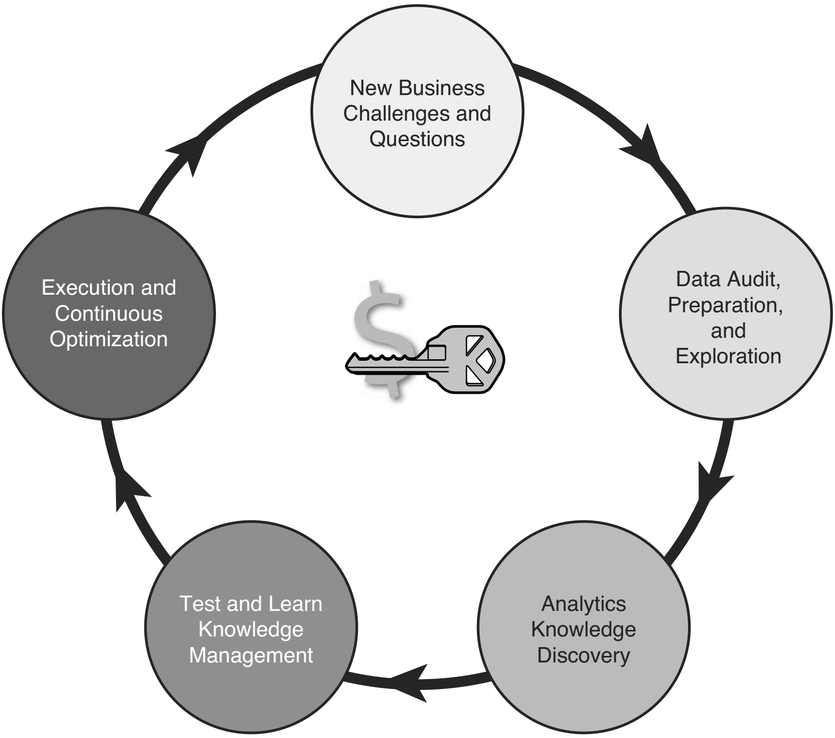

Figure 1-2 Analytics business value chain

- **Start with business questions**—All BA processes must start with a valid and high-value business question or idea. Even when the team is making a decision regarding IT, data, analytics models, and executions, it should not be made in isolation and apart from the business considerations. Even though the original business premise might sometimes need to be modified, it should always focus on business outcomes.
- **Conduct data audit and augmentation**—While the business objective is being set, a quick data audit can help determine

whether the business has the right data to accomplish the objective. If not, additional data has to be acquired or derived to augment the existing data. Sometimes the costs of data augmentation can be too prohibitive in the beginning, and a scaled-back business objective can be created first to validate the idea and value before undertaking a full-scale project.

- **Extract knowledge**—The main goal of the analytics exercises, including simple business intelligence (BI) analysis and advanced analytics modeling, is to extract useful business insights, patterns, and knowledge from the data. This is usually an iterative process. By answering the initial question, a good analytics team often generates several more questions. For example, when you see a group of customers exhibit a certain behavior, your next questions may be these: Who are they? Are they high-value customers? How long has this been the case? Is this a recent phenomenon? By peeling the onion and answering the sequence of questions, you can reach the core of the problem and uncover something that may thus far be hidden and unknown.
- **Test insights and hypotheses and knowledge management**—Once an insight on the predictors' impacts on the business outcomes is obtained, it needs to be tested for causal effects. To do this, the scientific testing protocol known as the design of experiments (DOE) should be used with suitable control groups. No insights should be taken at face value, no matter how intuitive or elegant they are. Once validated, the insights and effects must be saved in a knowledge management system for sharing and reuse.
- **Execution and optimization**—Analytics insights must also take into account how they are to be implemented. The lever settings and their respective effects can be used to help optimize the analytics efforts. The validated insights and optimized

lever settings can then be used to permit the business to ask the next level of questions. This begins another cycle of BAP.

Integrated Approach

Instead of viewing the various aspects of analytics techniques, business problems, and Big Data applications in isolation, I approach this by bringing everything together in the right sequence to produce the desired outcomes. This is why the book is not organized by analytics techniques, but instead by the way business problems are solved through analytics.

Hands-on Exercises

By embedding the entire analytics process with real business data within a powerful and modern advanced analytics platform (KNIME),[7] motivated business readers can then try real analytics on their own data to answer their question.

Note

For questions on installation and use of KNIME, please refer to the Appendix or the excellent KNIME booklet.[8] There is also a Learning Hub,[9] in which many great learning resources are available for all levels of users.

Reasons for Using KNIME Workflows

Based on the following reasons, I use KNIME workflows for all analytics examples and exercises in this book:

7 http://www.knime.com
8 http://www.knime.org/files/bl_sample_s.pdf
9 http://www.knime.org/learning-hub?src=knimeapp#basic

- **Advanced, powerful, and free**—As will be shown later, KNIME is not only powerful but (because it was developed recently) also incorporates the latest advanced analytics models, IT innovations, and user-friendly interfaces (for true click-and-drop analytics model development). Best of all, it is in the public domain and free (except for the server version).
- **Holistic and integrated views**—Some readers might want to see how the results are produced within an entire customer relationship management (CRM) or marketing or sales analytics process; others might want to drill down to a particular stage of the workflow to see how the data is transformed and/or model parameters are set and results produced. More adventurous readers might even want to vary the parameters and see how the changes affect the outcomes. The entire workflow can be viewed at the various levels of aggregation.
- **A single unifying enterprise workflow across silos**—In an ideal case, at the top level, it is strictly an enterprise business workflow. Below it lie the various sub-workflows that can correspond to the analytics workflows relating to IT, finance, marketing, stores, call centers, logistics, supply chain management, and so on. Drilling further down, you can then isolate the various components of the BA process. By embedding the actual analytics and reports within the metanodes of the sub-workflows, various stakeholders can then focus on the particular level and workflows they are interested in and be able to literally trace any questions to their sub-workflows to gain a holistic view of the big picture.

Conclusion

In this chapter, the transformational power of data, whether big or small, was discussed and likened to the way oil transformed the twentieth century. However, the power is not in data alone but also as it is being refined through analytics. Although analytics has become part of the daily vernacular, we have specifically defined BA as "the process of discovery of actionable knowledge and the creation of new business opportunities from such knowledge."

To illustrate the applications of analytics to answer many critical business questions, a list of top 10 questions in every area of business was discussed. Within each question, ways of applying analytics were also suggested. The vital lessons learned in more than a decade of applying analytics in businesses were also shared. To avoid the common trend of investing in analytics without using it, pitfalls were also highlighted based on the author's own experience.

The main lesson is to tightly link analytics to business. To do this, you have to create safe and productive intersections staffed by analytics deciders who are well-versed in both business and analytics, supported by an analytics sandbox as part of the holistic BAP.

This page intentionally left blank

This page intentionally left blank

Index

Q

R