Cengage

Business Analytics

Descriptive • Predictive • Prescriptive

Camm Cochran Fry Ohlmann

Business Analytics

Descriptive • Predictive • Prescriptive

Jeffrey D. Camm
Wake Forest University

James J. Cochran
University of Alabama

Michael J. Fry
University of Cincinnati

Jeffrey W. Ohlmann
University of Iowa

Cengage

Australia • Brazil • Canada • Mexico • Singapore • United Kingdom • United States

Cengage

***Business Analytics,* Fifth Edition**
Jeffrey D. Camm, James J. Cochran, Michael J. Fry, Jeffrey W. Ohlmann

VP, Product Management & Marketing, Learning Experiences: Thais Alencarr

Product Director: Joe Sabatino

Senior Portfolio Product Manager: Aaron Arnsparger

Content Manager: Breanna Holmes

Product Assistant: Flannery Cowan

Marketing Manager: Colin Kramer

Senior Learning Designer: Brandon Foltz

Digital Product Manager: Andrew Southwell

Intellectual Property Analyst: Ashley Maynard

Intellectual Property Project Manager: Kelli Besse

Production Service: MPS Limited

Project Manager, MPS Limited: Shreya Tiwari

Art Director: Chris Doughman

Text Designer: Beckmeyer Design

Cover Designer: Beckmeyer Design

Cover Image: iStockPhoto.com/tawanlubfah

Previous Editions: ©2021, ©2019, ©2017

For product information and technology assistance, contact us at
Cengage Customer & Sales Support, 1-800-354-9706
or support.cengage.com.

For permission to use material from this text or product, submit all requests online at **www.copyright.com**.

Library of Congress Control Number: 2023900651

Student Edition:
ISBN: 978-0-357-90220-2

Loose-leaf Edition
ISBN: 978-0-357-90221-9

Cengage
200 Pier 4 Boulevard
Boston, MA 02210
USA

Cengage is a leading provider of customized learning solutions. Our employees reside in nearly 40 different countries and serve digital learners in 165 countries around the world. Find your local representative at: **www.cengage.com.**

To learn more about Cengage platforms and services, register or access your online learning solution, or purchase materials for your course, visit **www.cengage.com.**

Printed in the United States of America
Print Number: 01 Print Year: 2023

Brief Contents

Contents

About the Authors

Jeffrey D. Camm. Jeffrey D. Camm is the Inmar Presidential Chair in Analytics and Senior Associate Dean of Faculty in the School of Business at Wake Forest University. Born in Cincinnati, Ohio, he holds a B.S. from Xavier University (Ohio) and a Ph.D. from Clemson University. Prior to joining the faculty at Wake Forest, he was on the faculty of the University of Cincinnati. He has also been a visiting scholar at Stanford University and a visiting professor of business administration at the Tuck School of Business at Dartmouth College.

Dr. Camm has published over 45 papers in the general area of optimization applied to problems in operations management and marketing. He has published his research in *Science, Management Science, Operations Research, Interfaces*, and other professional journals. Dr. Camm was named the Dornoff Fellow of Teaching Excellence at the University of Cincinnati and he was the recipient of the 2006 INFORMS Prize for the Teaching of Operations Research Practice. A firm believer in practicing what he preaches, he has served as an operations research consultant to numerous companies and government agencies. From 2005 to 2010, he served as editor-in-chief of *INFORMS Journal on Applied Analytics* (formerly *Interfaces*). In 2017, he was named an INFORMS Fellow.

James J. Cochran. James J. Cochran is Professor of Applied Statistics, the Rogers-Spivey Faculty Fellow, and Associate Dean for Faculty and Research at the University of Alabama. Born in Dayton, Ohio, he earned his B.S., M.S., and M.B.A. degrees from Wright State University and his Ph.D. from the University of Cincinnati. He has been at the University of Alabama since 2014 and has been a visiting scholar at Stanford University, Universidad de Talca, the University of South Africa, and Pole Universitaire Leonard de Vinci.

Professor Cochran has published over 50 papers in the development and application of operations research and statistical methods. He has published his research in *Management Science, The American Statistician, Communications in Statistics—Theory and Methods, Annals of Operations Research, European Journal of Operational Research, Journal of Combinatorial Optimization, INFORMS Journal of Applied Analytics, Statistics and Probability Letters*, and other professional journals. He was the 2008 recipient of the INFORMS Prize for the Teaching of Operations Research Practice and the 2010 recipient of the Mu Sigma Rho Statistical Education Award. Professor Cochran was elected to the International Statistics Institute in 2005 and named a Fellow of the American Statistical Association in 2011. He received the Founders Award in 2014 and the Karl E. Peace Award in 2015 from the American Statistical Association. In 2017 he received the American Statistical Association's Waller Distinguished Teaching Career Award and was named a Fellow of INFORMS, and in 2018 he received the INFORMS President's Award. He has twice been recognized as a finalist for the Innovative Applications in Analytics Award.

A strong advocate for effective statistics and operations research education as a means of improving the quality of applications to real problems, Professor Cochran has organized and chaired teaching effectiveness workshops in Montevideo, Uruguay; Cape Town, South Africa; Cartagena, Colombia; Jaipur, India; Buenos Aires, Argentina; Nairobi, Kenya; Buea, Cameroon; Kathmandu, Nepal; Osijek, Croatia; Havana, Cuba; Ulaanbaatar, Mongolia; Chişinău, Moldova; Dar es Salaam, Tanzania; Sozopol, Bulgaria; Tunis, Tunisia; and Saint George's, Grenada. He has served as an operations research consultant to numerous companies and not-for-profit organizations. He served as editor-in-chief of *INFORMS Transactions on Education* from 2006 to 2012 and is on the editorial board of *INFORMS Journal on Applied Analytics* (formerly *Interfaces*), *International Transactions in Operational Research,* and *Significance*.

Michael J. Fry. Michael J. Fry is Professor of Operations, Business Analytics, and Information Systems and Academic Director of the Center for Business Analytics in the Carl H. Lindner College of Business at the University of Cincinnati. Born in Killeen, Texas, he earned a B.S. from Texas A&M University and M.S.E. and Ph.D. degrees from the

University of Michigan. He has been at the University of Cincinnati since 2002, where he was previously Department Head and has been named a Lindner Research Fellow. He has also been a visiting professor at the Samuel Curtis Johnson Graduate School of Management at Cornell University and the Sauder School of Business at the University of British Columbia.

Professor Fry has published more than 25 research papers in journals such as *Operations Research, M&SOM, Transportation Science, Naval Research Logistics, IISE Transactions, Critical Care Medicine,* and *INFORMS Journal on Applied Analytics* (formerly *Interfaces*). His research has been funded by the National Science Foundation and other funding agencies. His research interests are in applying quantitative management methods to the areas of supply chain analytics, sports analytics, and public-policy operations. He has worked with many different organizations for his research, including Dell, Inc., Starbucks Coffee Company, Great American Insurance Group, the Cincinnati Fire Department, the State of Ohio Election Commission, the Cincinnati Bengals, and the Cincinnati Zoo & Botanical Garden. He was named a finalist for the Daniel H. Wagner Prize for Excellence in Operations Research Practice, and he has been recognized for both his research and teaching excellence at the University of Cincinnati.

Jeffrey W. Ohlmann. Jeffrey W. Ohlmann is Associate Professor of Management Sciences and Huneke Research Fellow in the Tippie College of Business at the University of Iowa. Born in Valentine, Nebraska, he earned a B.S. from the University of Nebraska, and M.S. and Ph.D. degrees from the University of Michigan. He has been at the University of Iowa since 2003.

Professor Ohlmann's research on the modeling and solution of decision-making problems has produced more than two dozen research papers in journals such as *Operations Research, Mathematics of Operations Research, INFORMS Journal on Computing, Transportation Science,* the *European Journal of Operational Research,* and *INFORMS Journal on Applied Analytics* (formerly *Interfaces*). He has collaborated with companies such as Transfreight, LeanCor, Cargill, the Hamilton County Board of Elections, and three National Football League franchises. Because of the relevance of his work to industry, he was bestowed the George B. Dantzig Dissertation Award and was recognized as a finalist for the Daniel H. Wagner Prize for Excellence in Operations Research Practice.

Preface

Business Analytics 5E is designed to introduce the concept of business analytics to undergraduate and graduate students. This edition builds upon what was one of the first collections of materials that are essential to the growing field of business analytics. In Chapter 1, we present an overview of business analytics and our approach to the material in this textbook. In simple terms, business analytics helps business professionals make better decisions based on data. We discuss exploring, wrangling, summarizing, visualizing, and understanding data in Chapters 2 through 5. In Chapter 6, we introduce additional models to describe data and the relationships among variables. Chapters 7 through 11 introduce methods for both gaining insights from historical data and predicting possible future outcomes. Chapter 12 covers the use of spreadsheets for examining data and building decision models. In Chapter 13, we demonstrate how to explicitly introduce uncertainty into spreadsheet models through the use of Monte Carlo simulation. In Chapters 14 through 16, we discuss optimization models to help decision makers choose the best decision based on the available data. Chapter 17 is an overview of decision analysis approaches for incorporating a decision maker's views about risk into decision making. In Appendix A we present optional material for students who need to learn the basics of using Microsoft Excel. The use of databases and manipulating data in Microsoft Access is discussed in Appendix B. Appendixes in many chapters illustrate the use of additional software tools such as Tableau, R, and Orange to apply analytics methods.

This textbook can be used by students who have previously taken a course on basic statistical methods as well as students who have not had a prior course in statistics. *Business Analytics* 5E is also amenable to a two-course sequence in business statistics and analytics. All statistical concepts contained in this textbook are presented from a business analytics perspective using practical business examples. Chapters 2, 5, 7, 8, and 9 provide an introduction to basic statistical concepts that form the foundation for more advanced analytics methods. Chapter 2 introduces descriptive statistical measures such as measures of location, measures of variability, methods to analyze distributions of data, and measures of association between variables. Chapter 5 covers material related to probability including discussions of discrete probability distributions and continuous probability distributions. Chapter 7 contains topics related to statistical inference including interval estimation and hypothesis testing. Chapter 8 explores linear regression models including simple linear models and multiple linear regression models, and chapter 9 covers methods for forecasting and time series data. Chapter 3 covers additional data visualization topics and Chapter 4 is an overview of approaches for exploring, cleaning, and wrangling data to make the data more amenable for analysis. These topics are not always covered in traditional business statistics courses, but they are essential in understanding how to analyze data. Chapters 6, 10, and 11 cover additional topics in data mining that are not traditionally part of most introductory business statistics courses, but they are exceedingly important and commonly used in current business environments. Chapter 6 focuses on data mining methods used to describe data and the relationships among variables such as cluster analysis, association rules and text mining. Chapter 6 also contains an overview of dimension reduction techniques such as principal component analysis. Chapters 10 and 11 focus on data mining methods used to make predictions from data. Chapter 10 covers models for regression tasks such as *k*-nearest neighbors, regression trees, and neural networks. Chapter 11 covers models for classification tasks such as logistic regression, *k*-nearest neighbors, classification trees, ensemble methods, and neural networks. Both chapters 10 and 11 cover data sampling methods for prediction models including *k*-fold cross-validation, and also provide an overview of feature selection methods such as wrapper methods, filter methods, and embedded methods (including regularization). Chapter 12 and Appendix A provide the foundational knowledge students need to use Microsoft Excel for analytics applications. Chapters 13 through 17 build upon this spreadsheet knowledge to present additional topics that are used by many organizations that use prescriptive analytics to improve decision making.

Updates in the Fifth Edition

The fifth edition of *Business Analytics* is a major revision that introduces new chapters, new concepts, and new tools. Chapter 4 is a new chapter on data wrangling that covers topics such as how to access and structure data for exploration, how to clean and enrich data to facilitate analysis, and how to validate data. Our coverage of data mining topics has been greatly expanded. Our coverage of descriptive data mining techniques in Chapter 6 now includes a discussion of how to conduct dimension reduction with principal component analysis (PCA), and we have thoroughly updated our coverage of clustering, association rules, and text mining. Our coverage of predictive data mining techniques now includes two separate chapters: Chapter 10 focuses on predicting quantitative outcomes with *k*-nearest neighbors regression, regression trees, ensemble methods, and neural network regression, while Chapter 11 focuses on predicting binary categorical outcomes with *k*-nearest neighbors classification, classification trees, ensemble methods, and neural network classification. In addition, we now include online appendixes that introduce the software package Orange for descriptive and predictive data-mining models. Orange is an open-source machine learning and data visualization software package built using Python. This coverage of Orange and Python complements our existing coverage of R for descriptive and predictive analytics. Our coverage of descriptive analytics methods in Chapter 2 and data visualization in Chapter 3 has been greatly expanded to provide additional depth and introduce new methods related to these topics. We have also increased the size of many data sets in Chapter 8 on linear regression and Chapter 9 on time series analysis and forecasting to better represent realistic data sets that are encountered in practice. We have added additional topics to our coverage of optimization models including a new section in Chapter 16 on heuristic optimization with the Excel Solver Evolutionary method. We also now provide a new online Appendix D that covers the basics of the use of Microsoft Excel Online for statistical analysis. Finally, we have added learning objectives (LOs) to the beginning of each chapter that explain the key concepts covered in each chapter, and we map these LOs to the end-of-chapter problems and cases.

- **New Chapter on Data Wrangling.** Chapter 4 covers methods for accessing and structuring data for exploring, cleaning and enriching data to facilitate analysis, and validating data. Many professionals in analytics and data science spend much of their time preparing data for exploration and analysis. It is essential that students are familiar with these methods. We have also created new online appendixes covering how to implement these methods in R.
- **New Material in Descriptive Data Mining Chapter.** Chapter 6 on descriptive data mining techniques now provides a more in-depth discussion of dimension reduction techniques such as principal component analysis. We present material to help students understand what this approach is doing, as well as how to interpret the output from this method. We have also rewritten many sections in this chapter to provide additional explanation and added context to the data-mining methods introduced.
- **New Material in Predictive Data Mining Chapters.** We have divided the previous single chapter on predictive data mining methods into two different chapters: Chapter 10 covers regression data mining models, and Chapter 11 covers classification data mining models. Splitting this chapter into two chapters allows us to include additional methods and more thoroughly explain the methods and output. Chapter 10 focuses on predicting quantitative outcomes with *k*-nearest neighbors regression, regression trees, and neural network regression. Chapter 11 focuses on predicting binary categorical outcomes with logistic regression, *k*-nearest neighbors classification, classification trees, and neural network classification.
- **New Online Appendixes for Using Python-Based Orange for Data Mining.** We now include online appendixes that introduce the software package Orange for descriptive and predictive data-mining models. Orange is an open-source machine learning and data visualization software package built using Python. It provides an easy-to-use, yet powerful, workflow-based approach to building analytics models. The use of Python by

analytics professionals has increased greatly, and Orange provides an excellent introduction to using Python in an easy-to-learn environment. These appendixes include practice problems for students to solve using Orange.

- **Updates to R Appendixes for Data Mining Chapters.** We continue to include online appendixes for using the popular open-source software package R for descriptive and predictive data mining. We have revised these online appendixes to focus on using script R commands in RStudio rather than the Rattle interface. We have found the use of script R through RStudio to be more robust and more stable than the use of Rattle. To facilitate the teaching of this material, we include complete R script files for all examples used throughout the R appendixes included for this this textbook as well as detailed step-by-step instructions.
- **Practice Problems in Online Appendixes for Orange and R.** Each online appendix for using Orange or R now includes practice problems that can be solved using these popular open-source software packages. The practice problems include specific hints for how to solve these problems using each software. Problem solutions, including the necessary code files, using these software packages are available to instructors for all practice problems.
- **Additional Coverage of Descriptive Analytics and Data Visualization.** We have added additional coverage in Chapters 2 and 3 to provide more depth of our coverage of descriptive analytics methods and data visualization. In Chapter 2, we have rewritten the explanation of histograms and we have added a discussion of frequency polygons to supplement the coverage of histograms as a way of exploring distributions of data. Chapter 3's coverage of data visualization now includes a more comprehensive discussion of best practices in data visualization including an explanation of the use of preattentive attributes and the data-ink ratio to create effective tables and charts. We have also rearranged the chapter and added coverage of table lens, waterfall charts, stock charts, choropleth maps, and cartograms for data visualization.
- **New Material for Optimization.** We have expanded our coverage of optimization models. Chapter 14 on Linear Optimization Models now includes additional models for transportation problems, diet problems, and assignment problems. Chapter 16 on nonlinear optimization models contains a new section on heuristic optimization using Excel's Evolutionary Solver method that can be used to solve complex optimization models. We have also added online appendixes for solving optimization models in R.
- **Increased Size of Data Sets.** We have increased the size of many data sets in Chapter 8 on linear regression and Chapter 9 on time series analysis and forecasting. These larger data sets provide better representations of realistic data sets that are encountered in practice, and they are more useful for conducting inference and examining the underlying assumptions made in regression and time series analysis.
- **Learning Objectives.** We have added Learning Objectives (LOs) to the beginning of each chapter. These LOs explain the key concepts that are covered in each chapter. The LOs are also mapped onto each end-of-chapter problem and case so instructors can easily identify which LOs are covered by each problem.
- **New End-of-Chapter Problems and Cases.** The fifth edition of this textbook includes over 175 new problems and 6 new cases. We have added problems to Chapter 1 so that now each chapter in the textbook includes end-of-chapter problems. We have divided the end-of-chapter problems in Chapters 6, 10, and 11 covering data mining topics into conceptual problems and software application problems. Conceptual problems do not require dedicated software to answer, and we have added many new conceptual problems to each data mining chapter. Software application problems require the use of dedicated software, and allow students to apply the tools covered in these chapters. The software application problems in these chapters are also provided in the online appendixes for the use of Orange and R for data mining applications. The online appendix version of these

problems include specific hints for how to solve the problems in the corresponding software. As we have done in past editions, Excel solution files are available to instructors for many of the problems and cases that require the use of Excel. For software application problems that require the use of software in the data-mining chapters (Chapters 6, 10, and 11), we include solutions for both Orange and R.

Continued Features and Pedagogy

In the fifth edition of this textbook, we continue to offer all of the features that have been successful in the previous editions. Some of the specific features that we use in this textbook are listed below.

- **Integration of Microsoft Excel:** Excel has been thoroughly integrated throughout this textbook. For many methodologies, we provide instructions for how to perform calculations both by hand and with Excel. In other cases where realistic models are practical only with the use of a spreadsheet, we focus on the use of Excel to describe the methods to be used. Excel instructions have been fully updated throughout the textbook to match the latest versions of Excel most likely to be used by students. The textbook assumes the use of desktop versions of Excel for most problems and examples, but the accompanying online Appendix D covers the basics of the use of Microsoft Excel Online.
- **Notes and Comments:** At the end of many sections, we provide Notes and Comments to give the student additional insights about the methods presented in that section. These insights include comments on the limitations of the presented methods, recommendations for applications, and other matters. Additionally, margin notes are used throughout the textbook to provide additional insights and tips related to the specific material being discussed.
- **Analytics in Action:** Each chapter contains an Analytics in Action article. These articles present interesting examples of the use of business analytics in practice. The examples are drawn from many different organizations in a variety of areas including healthcare, finance, manufacturing, marketing, and others.
- **DATAfiles and MODELfiles:** All data sets used as examples and in student exercises are also provided online on the companion site as files available for download by the student. DATAfiles are files that contain data needed for the examples and problems given in the textbook. Typically, the DATAfiles are in .xlsx format for Excel or .csv format for import into other software packages. MODELfiles contain additional modeling features such as extensive use of Excel formulas or the use of Excel Solver, script files for R, or workflow models for Orange.
- **Problems and Cases:** Each chapter, now including Chapter 1, contains an extensive selection of problems to help the student master the material presented in that chapter. The problems vary in difficulty and most relate to specific examples of the use of business analytics in practice. Answers to selected even-numbered problems are provided in an online supplement for student access. With the exception of Chapter 1, each chapter also includes at least one in-depth case study that connects many of the different methods introduced in the chapter. The case studies are designed to be more open-ended than the chapter problems, but enough detail is provided to give the student some direction in solving the cases. We continue to include two cases at the end of the textbook that require the use of material from multiple chapters in the text to better illustrate how concepts from different chapters relate to each other.

MindTap

MindTap is a customizable digital course solution that includes an interactive eBook, autograded exercises from the textbook, algorithmic practice problems with solutions feedback, Excel Online problems, Exploring Analytics visualizations, Adaptive Test Prep, videos, and more! MindTap is also where instructors and users can find the online appendixes for R and Orange. All of these materials offer students better access to resources to understand the materials within the course. For more information on MindTap, please contact your Cengage representative.

WebAssign

Prepare for class with confidence using WebAssign from Cengage. This online learning platform fuels practice, so students can truly absorb what you learn—and are better prepared come test time. Videos, Problem Walk-Throughs, and End-of-Chapter problems and cases with instant feedback help them understand the important concepts, while instant grading allows you and them to see where they stand in class. WebAssign is also where instructors and users can find the online appendixes for R and Orange. Class Insights allows students to see what topics they have mastered and which they are struggling with, helping them identify where to spend extra time. Study Smarter with WebAssign.

Instructor and Student Resources

Additional instructor and student resources for this product are available online. Instructor assets include a Solutions and Answers Guide, Instructor's Manual, PowerPoint® slides, and a test bank powered by Cognero®. Prepared by the authors, solutions for software application problems and cases in chapters 6, 10, and 11 are available using both R and Orange. Student assets include DATAfiles and MODELfiles that accompany the chapter examples and problems as well as solutions to selected even-numbered chapter problems. Sign up or sign in at www.cengage.com to search for and access this product and its online resources.

Acknowledgments

We would like to acknowledge the work of reviewers and users who have provided comments and suggestions for improvement of this text. Thanks to:

Rafael Becerril Arreola
University of South Carolina

Matthew D. Bailey
Bucknell University

Phillip Beaver
University of Denver

M. Khurrum S. Bhutta
Ohio University

Paolo Catasti
Virginia Commonwealth University

Q B. Chung
Villanova University

Elizabeth A. Denny
University of Kentucky

Mike Taein Eom
University of Portland

Yvette Njan Essounga
Fayetteville State University

Lawrence V. Fulton
Texas State University

Tom Groleau
Carthage College

James F. Hoelscher
Lincoln Memorial University

Eric Huggins
Fort Lewis College

Faizul Huq
Ohio University

Marco Lam
York College of Pennsylvania

Thomas Lee
University of California, Berkeley

Roger Myerson
Northwestern University

Ram Pakath
University of Kentucky

Susan Palocsay
James Madison University

Andy Shogan
University of California, Berkeley

Dothan Truong
Embry-Riddle Aeronautical University

Kai Wang
Wake Technical Community College

Ed Wasil
American University

Ed Winkofsky
University of Cincinnati

A special thanks goes to our associates from business and industry who supplied the Analytics in Action features. We recognize them individually by a credit line in each of the articles. We are also indebted to our senior product manager, Aaron Arnsparger; our Senior Content Manager, Breanna Holmes; senior learning designer, Brandon Foltz; subject matter expert, Deborah Cernauskas; digital project manager, Andrew Southwell; and our project manager at MPS Limited, Shreya Tiwari, for their editorial counsel and support during the preparation of this text.

Jeffrey D. Camm
James J. Cochran
Michael J. Fry
Jeffrey W. Ohlmann

Chapter 1

Introduction to Business Analytics

Contents

Learning Objectives

After completing this chapter, you will be able to:

LO 1 Identify strategic, tactical, and operational decisions.

LO 2 Describe the steps in the decision-making process.

LO 3 Identify examples of descriptive, predictive, and prescriptive analytics.

LO 4 Describe applications of analytics for decision making.

In many situations, data are often abundant and can be used to guide decision-making. Suppose you apply for a loan for the first time. How does the bank assess the riskiness of the loan it might make to you? How does Amazon.com know which books and other products to recommend to you when you log in to their web site? How do airlines determine what price to quote to you when you are shopping for a plane ticket?

You may be applying for a loan for the first time, but millions of people around the world have applied for loans before. Many of these loan recipients have paid back their loans in full and on time, but some have not. The bank wants to know whether you are more like those who have paid back their loans or more like those who defaulted. By comparing your credit history, financial situation, and other factors to the vast database of previous loan recipients, the bank can effectively assess how likely you are to default on a loan.

Similarly, Amazon.com has access to data on millions of purchases made by customers on its web site. Amazon.com examines your previous purchases, the products you have viewed, and any product recommendations you have provided. Amazon.com then searches through its huge database for customers who are similar to you in terms of product purchases, recommendations, and interests. Once similar customers have been identified, their purchases form the basis of the recommendations given to you.

Prices for airline tickets are frequently updated. The price quoted to you for a flight between New York and San Francisco today could be very different from the price that will be quoted tomorrow. These changes happen because airlines use a pricing strategy known as revenue management. Revenue management works by examining vast amounts of data on past airline customer purchases and using these data to forecast future purchases. These forecasts are then fed into sophisticated optimization algorithms that determine the optimal price to charge for a particular flight and when to change that price. Revenue management has resulted in substantial increases in airline revenues.

This book is concerned with data-driven decision making and the use of analytical approaches in the decision-making process. Three developments spurred recent explosive growth in the use of analytical methods in business applications. First, technological advances have enabled the ability to track and store large amounts of data. Improved point-of-sale scanner technology, tracking of activity on the internet (e.g,. e-commerce and social networks), sensors on mechanical devices such as aircraft engines, automobiles, and farm machinery through the Internet-of-Things, and personal electronic devices produce incredible amounts of data for businesses. Naturally, businesses want to use these data to improve the efficiency and profitability of their operations, better understand their customers, price their products more effectively, and gain a competitive advantage. Second, ongoing research has resulted in numerous methodological developments to extract knowledge from the data. Examples of these are advances in computational approaches to effectively handle and explore massive amounts of data, faster algorithms for optimization and simulation, and more effective approaches for visualizing data. Third, these methodological developments are paired with an explosion in computing power. Faster computer chips, parallel computing, and cloud computing (the remote use of hardware and software over the Internet) have enabled businesses to solve big problems more quickly and more accurately than ever before.

In summary, the availability of massive amounts of data, improvements in analytic methodologies, and substantial increases in computing power have all come together to result in a dramatic upsurge in the use of analytical methods in business and a reliance on the discipline that is the focus of this text: business analytics. The purpose of this text is to provide students with a sound conceptual understanding of the role that business analytics plays in the decision-making process and to provide a better understanding of the variety of applications in which analytical methods have been used successfully.

1.1 Decision Making

It is the responsibility of managers to plan, coordinate, organize, and lead their organizations to better performance. Ultimately, managers' responsibilities require that they make strategic, tactical, or operational decisions. **Strategic decisions** involve higher-level issues concerned with the overall direction of the organization; these decisions define the organization's overall goals and aspirations for the future. Strategic decisions are usually the domain of higher-level executives and have a time horizon of three to five years. **Tactical decisions** concern how the organization should achieve the goals and objectives set by its strategy, and they are usually the responsibility of midlevel management. Tactical decisions usually span a year and thus are revisited annually or even every six months. **Operational decisions** affect how the firm is run from day to day; they are the domain of operations managers, who are the closest to the customer.

Consider the case of the Thoroughbred Running Company (TRC). Historically, TRC had been a catalog-based retail seller of running shoes and apparel. TRC sales revenues grew quickly as it changed its emphasis from catalog-based sales to Internet-based sales. Recently, TRC decided that it should also establish retail stores in the malls and downtown areas of major cities. This strategic decision will take the firm in a new direction that it hopes will complement its Internet-based strategy. TRC middle managers will therefore have to make a variety of tactical decisions in support of this strategic decision, including how many new stores to open this year, where to open these new stores, how many distribution centers will be needed to supply the new stores, and where to locate these distribution centers. Operations managers in the stores will need to make day-to-day decisions regarding, for instance, how many pairs of each model and size of shoes to order from the distribution centers and how to schedule their sales personnel's work time.

Regardless of the level within the firm, *decision making* can be defined as the following process:

1. Identify and define the problem.
2. Determine the criteria that will be used to evaluate alternative solutions.
3. Determine the set of alternative solutions.
4. Evaluate the alternatives.
5. Choose an alternative.

Step 1 of decision making, identifying and defining the problem, is the most critical. Only if the problem is well-defined, with clear metrics of success or failure (step 2), can a proper approach for solving the problem (steps 3 and 4) be devised. Decision making concludes with the choice of one of the alternatives (step 5).

There are a number of approaches to making decisions: tradition ("We've always done it this way"), intuition ("gut feeling"), and rules of thumb ("As the restaurant owner, I schedule twice the number of waiters and cooks on holidays"). The power of each of these approaches should not be underestimated. Managerial experience and intuition are valuable inputs to making decisions, but what if relevant data are available to help us make more-informed decisions? The vast amounts of data now generated and stored electronically provide tremendous opportunity for businesses to improve their profitability and service to their customers. How can managers convert these data into knowledge they can use to be more efficient and effective in managing their businesses?

1.2 Business Analytics Defined

What makes decision making difficult and challenging? Uncertainty is probably the number one challenge. If we knew how much the demand would be for our product, we could do a much better job of planning and scheduling production. If we knew exactly how long each step in a project would take to be completed, we could better predict the project's cost and completion date. If we knew how stocks would perform, investing would be a lot easier.

Another factor that makes decision making difficult is that we often face such an enormous number of alternatives that we cannot evaluate them all. What is the best combination of stocks to help me meet my financial objectives? What is the best product line for a company that wants to maximize its market share? How should an airline price its tickets so as to maximize revenue?

*Some firms and industries use the simpler term, **analytics**. Analytics is often thought of as a broader category than business analytics, encompassing the use of analytical techniques in the sciences and engineering as well. In this text, we use **business analytics** and **analytics** synonymously.*

Business analytics is the scientific process of transforming data into insight for making better decisions.[1] Business analytics is used for data-driven or fact-based decision making, which is often seen as more objective than other alternatives for decision making.

As we shall see, the tools of business analytics can aid decision making by creating insights from data, by improving our ability to more accurately forecast for planning, by helping us quantify risk, and by yielding better alternatives through analysis and optimization. A study based on a large sample of firms that was conducted by researchers at MIT's Sloan School of Management and the University of Pennsylvania concluded that firms guided by data-driven decision making have higher productivity and market value and increased output and profitability.[2]

1.3 A Categorization of Analytical Methods and Models

Business analytics can involve anything from simple reports to the most advanced optimization techniques (methods for finding the best course of action). Analytics is generally thought to comprise three broad categories of techniques: descriptive analytics, predictive analytics, and prescriptive analytics.

Descriptive Analytics

Descriptive analytics encompasses the set of techniques that describes what has happened in the past. Examples are data queries, reports, descriptive statistics, data visualization including data dashboards, unsupervised learning techniques from data mining, and basic spreadsheet models.

Appendix B, at the end of this book, describes how to use Microsoft Access to conduct data queries.

A **data query** is a request for information with certain characteristics from a database. For example, a query to a manufacturing plant's database might be for all records of shipments to a particular distribution center during the month of March. This query provides descriptive information about these shipments: the number of shipments, how much was included in each shipment, the date each shipment was sent, and so on. A report summarizing relevant historical information for management might be conveyed by the use of descriptive statistics (means, measures of variation, etc.) and data-visualization tools (tables, charts, and maps). Simple descriptive statistics and data-visualization techniques can be used to find patterns or relationships in a large database.

Data dashboards are collections of tables, charts, maps, and summary statistics that are updated as new data become available. Dashboards are used to help management monitor specific aspects of the company's performance related to their decision-making responsibilities. For corporate-level managers, daily data dashboards might summarize sales by region, current inventory levels, and other company-wide metrics; front-line managers may view dashboards that contain metrics related to staffing levels, local inventory levels, and short-term sales forecasts.

Data mining is the use of analytical techniques for better understanding patterns and relationships that exist in large data sets. Data mining includes **unsupervised learning** techniques which are descriptive methods that seek to identify patterns based on notions of similarity (cluster analysis) or correlation (association rules) in different types of data. For

---

[1]We adopt the definition of analytics developed by the Institute for Operations Research and the Management Sciences (INFORMS).

[2]E. Brynjolfsson, L. M. Hitt, and H. H. Kim, "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" Thirty-Second International Conference on Information Systems, Shanghai, China, December 2011.

example, by processing text on social network platforms such as Twitter, similar customer comments can be grouped together into clusters to help Apple better understand how its customers are feeling about the Apple Watch.

Predictive Analytics

Predictive analytics consists of techniques that use models constructed from past data to predict the future or ascertain the impact of one variable on another. For example, past data on product sales may be used to construct a mathematical model to predict future sales. This model can factor in the product's growth trajectory and seasonality based on past patterns. A packaged-food manufacturer may use point-of-sale scanner data from retail outlets to help in estimating the lift in unit sales due to coupons or sales events. Survey data and past purchase behavior may be used to help predict the market share of a new product. All of these are applications of predictive analytics. Traditional statistical methods such as linear regression and time series analysis fall under the banner of predictive analytics.

Data mining includes **supervised learning** techniques which use past data to learn the relationship between an outcome variable of interest and a set of input variables. For example, a large grocery store chain might be interested in developing a targeted marketing campaign that offers a discount coupon on potato chips. By studying historical point-of-sale data, the store may be able to use data mining to predict which customers are the most likely to respond to an offer on discounted potato chips, by purchasing higher-margin items such as beer or soft drinks in addition to the chips, thus increasing the store's overall revenue.

Simulation involves the use of probability and statistics to construct a computer model to study the impact of uncertainty on a decision. For example, banks often use simulation to model investment and default risk in order to stress-test financial models. Simulation is also often used in the pharmaceutical industry to assess the risk of introducing a new drug.

Prescriptive Analytics

Prescriptive analytics differs from descriptive and predictive analytics in that **prescriptive analytics** indicates a course of action to take; that is, the output of a prescriptive model is a decision. Predictive models provide a forecast or prediction, but do not provide a decision. However, a forecast or prediction, when combined with a rule, becomes a prescriptive model. For example, we may develop a model to predict the probability that a person will default on a loan. If we create a rule that says if the estimated probability of default is more than 0.6, we should not award a loan, now the predictive model, coupled with the rule is prescriptive analytics. These types of prescriptive models that rely on a rule or set of rules are often referred to as **rule-based models**.

Other examples of prescriptive analytics are portfolio models in finance, supply network design models in operations, and price-markdown models in retailing. Portfolio models use historical investment return data to determine which mix of investments will yield the highest expected return while controlling or limiting exposure to risk. Supply-network design models provide plant and distribution center locations that will minimize costs while still meeting customer service requirements. Given historical data, retail price markdown models yield revenue-maximizing discount levels and the timing of discount offers when goods have not sold as planned. All of these models are known as **optimization models**, that is, models that give the best decision subject to the constraints of the situation.

Another type of modeling in the prescriptive analytics category is **simulation optimization** which combines the use of probability and statistics to model uncertainty with optimization techniques to find good decisions in highly complex and highly uncertain settings. Finally, the techniques of **decision analysis** can be used to develop an optimal strategy when a decision maker is faced with several decision alternatives and an uncertain set of future events. Decision analysis also employs **utility theory**, which assigns values to outcomes based on the decision maker's attitude toward risk, loss, and other factors.

Table 1.1 Coverage of Business Analytics Topics in This Text

Chapter	Title	Descriptive	Predictive	Prescriptive
1	Introduction	●	●	●
2	Descriptive Statistics	●		
3	Data Visualization	●		
4	Data Wrangling	●		
5	Probability: An Introduction to Modeling Uncertainty	●		
6	Descriptive Data Mining	●		
7	Statistical Inference	●		
8	Linear Regression		●	
9	Time Series & Forecasting		●	
10	Predictive Data Mining: Regression Tasks		●	
11	Predictive Data Mining: Classification Tasks		●	
12	Spreadsheet Models	●	●	●
13	Monte Carlo Simulation		●	●
14	Linear Optimization Models			●
15	Integer Optimization Models			●
16	Nonlinear Optimization Models			●
17	Decision Analysis			●

In this text, we cover all three areas of business analytics: descriptive, predictive, and prescriptive. Table 1.1 shows how the chapters cover the three categories.

1.4 Big Data, the Cloud, and Artificial Intelligence

On any given day, Google receives over 8.5 billion searches, WhatsApp users exchange up to 65 billion messages, and Internet users generate about 2.5 quintillion bytes of data.[3] It is through technology that we have truly been thrust into the data age. Because data can now be collected electronically, the available amounts of it are staggering. The Internet, cell phones, retail checkout scanners, surveillance video, and sensors on everything from aircraft to cars to bridges allow us to collect and store vast amounts of data in real time.

In the midst of all of this data collection, the term *big data* has been created. There is no universally accepted definition of big data. However, probably the most accepted and most general definition is that **big data** is any set of data that is too large or too complex to be handled by standard data-processing techniques and typical desktop software. IBM describes the phenomenon of big data through the four Vs: volume, velocity, variety, and veracity, as shown in Figure 1.1.

Volume

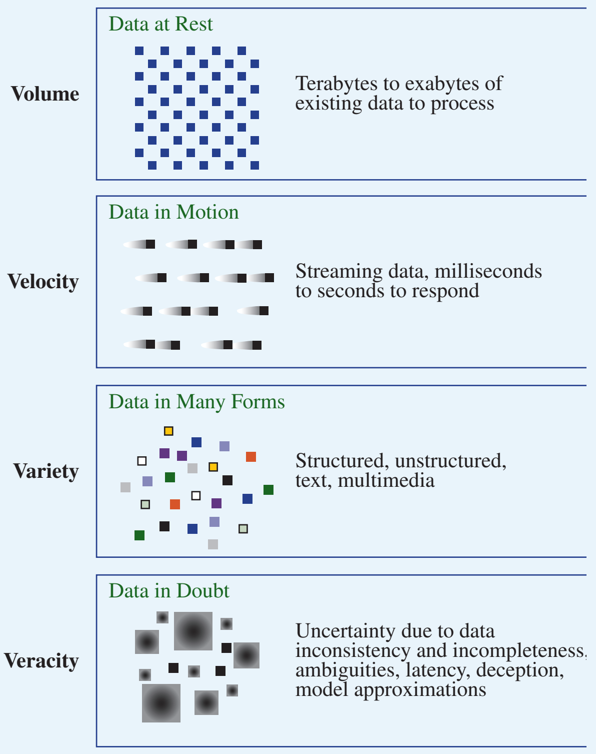
Because data are collected electronically, we are able to collect more of it. To be useful, these data must be stored, and this storage has led to vast quantities of data. Many companies now store in excess of 100 terabytes of data (a terabyte is 1,024 gigabytes).

Velocity

Real-time capture and analysis of data present unique challenges both in how data are stored and the speed with which those data can be analyzed for decision making. For

[3] https://techjury.net/blog/big-data-statistics/#gref

Figure 1.1 The Four Vs of Big Data



Source: IBM.

example, the New York Stock Exchange collects 1 terabyte of data in a single trading session, and having current data and real-time rules for trades and predictive modeling are important for managing stock portfolios.

Variety

In addition to the sheer volume and speed with which companies now collect data, more complicated types of data are now available and are proving to be of great value to businesses. Text data are collected by monitoring what is being said about a company's products or services on social media platforms such as Twitter. Audio data are collected from service calls (on a service call, you will often hear "this call may be monitored for quality control"). Video data collected by in-store video cameras are used to analyze shopping behavior. Analyzing information generated by these nontraditional sources is more complicated in part because of the processing required to transform the data into a numerical form that can be analyzed.

Veracity

Veracity has to do with how much uncertainty is in the data. For example, the data could have many missing values, which makes reliable analysis a challenge. Inconsistencies in units of measure and the lack of reliability of responses also increase the complexity of the data.

Businesses have realized that understanding big data can lead to a competitive advantage. Although big data represents opportunities, it also presents challenges in terms of data storage and processing, security, and available analytical talent.

The four Vs indicate that big data creates challenges in terms of how these complex data can be captured, stored, processed, secured, and then analyzed. Traditional databases typically assume that data fit into nice rows and columns, but that is not always the case with big data. Also, the sheer volume (the first V) often means that it is not possible to store all of the data on a single computer. This has led to technologies such as **Hadoop**—an open-source programming environment that supports big data processing through distributed storage and distributed processing on clusters of computers. Essentially, Hadoop provides a divide-and-conquer approach to handling massive amounts of data, dividing the storage and processing over multiple computers. **MapReduce** is a programming model used within Hadoop that performs the two major steps for which it is named: the map step and the reduce step. The map step divides the data into manageable subsets and distributes it to the computers in the cluster (often termed nodes) for storing and processing. The reduce step collects answers from the nodes and combines them into an answer to the original problem. Technologies such as Hadoop and MapReduce, paired with relatively inexpensive computer power, enable cost-effective processing of big data; otherwise, in some cases, processing might not even be possible.

The massive amounts of available data have led to numerous innovations that help make the data useful for decision making. **Cloud computing**, also known simply as "the cloud," refers to the use of data and software on servers housed external to an organization via the internet. The cloud has made storing and processing massive amounts of data feasible and cost effective. Companies pay a fee to store their big data and use software stored on servers with cloud alternatives such as Amazon Web Services (AWS), Microsoft Azure, and Oracle Cloud.

Storing so much data has raised concerns about the security. Medical records, bank account information, and credit card transactions, for example, are all highly confidential and must be protected from computer hackers. **Data security**, the protection of stored data from destructive forces or unauthorized users, is of critical importance to companies. For example, credit card transactions are potentially very useful for understanding consumer behavior, but compromise of these data could lead to unauthorized use of the credit card or identity theft. Companies such as Target, Anthem, JPMorgan Chase, Yahoo!, Facebook, Marriott, Equifax, and Home Depot have faced major data breaches costing millions of dollars.

In addition to increased interest in cloud computing and data security, big data has accelerated the development of applications of artificial intelligence. Broadly speaking, **artificial intelligence** (AI) is the use of big data and computers to make decisions that in the past would have required human intelligence. Often, AI software mimics the way we understand the human brain functions. AI uses many of the techniques of analytics, but often in real time, with massive amounts of data required to "train" algorithms, to complete an automated task. Applications of AI include for example, facial recognition for security checkpoints and self-driving vehicles.

The complexities of the big data have increased the demand for analysts, but a shortage of qualified analysts has made hiring more challenging. More companies are searching for **data scientists**, who know how to effectively process and analyze massive amounts of data because they are well trained in both computer science and statistics. Next we discuss three examples of how companies are collecting big data for competitive advantage.

Kroger Understands Its Customers Kroger is the largest retail grocery chain in the United States. It sends over 11 million pieces of direct mail to its customers each quarter. The quarterly mailers each contain 12 coupons that are tailored to each household based on several years of shopping data obtained through its customer loyalty card program. By collecting and analyzing consumer behavior at the individual household level, and better matching its coupon offers to shopper interests, Kroger has been able to realize a far higher redemption rate on its coupons. In the six-week period following distribution of the mailers, over 70% of households redeem at least one coupon, leading to an estimated coupon revenue of $10 billion for Kroger. (Source: *Forbes.com*)

MagicBand at Disney The Walt Disney Company offers a wristband to visitors to its Orlando, Florida, Disney World theme park. Known as the MagicBand, the wristband contains technology that can transmit more than 40 feet and can be used to track each visitor's location in the park in real time. The band can link to information that allows Disney to better serve its visitors. For example, prior to the trip to Disney World, a visitor might be asked to fill out a survey on their birth date and favorite rides, characters, and restaurant table type and location. This information, linked to the MagicBand, can allow Disney employees using smartphones to greet you by name as you arrive, offer you products they know you prefer, wish you a happy birthday, have your favorite characters show up as you wait in line or have lunch at your favorite table. The MagicBand can be linked to your credit card, so there is no need to carry cash or a credit card. And during your visit, your movement throughout the park can be tracked and the data can be analyzed to better serve you during your next visit to the park. (Source: *Wired.com*)

Coca-Cola Freestyle Gives Consumers Their Own Personal Soft Drinks Coca-Cola, the largest beverage company in the world, sells its products in over 200 countries. One of Coca-Cola's innovations is Coca-Cola Freestyle, a touch screen self-service soda dispenser that can be found in numerous restaurants such as Subway and Burger King. Coca-Cola Freestyle allows the customer to create their own customized drink by combining existing flavors. For example, a customer might combine Sprite with Hi-C Orange to create a flavor that is not available in stores. But Coca-Cola Freestyle is not only an innovation that better serves customers through mass customization—the machines are also a goldmine of customer preference data. The Freestyle machines collect data on mixes that consumers around the world choose. These data are analyzed to better understand customer preferences and how they differ by country and region. Whereas in the past, marketing analysts would need to rely on expensive focus groups and multiple rounds of market testing to help develop new products, Freestyle machines directly provide data on consumer preferences. For example, a new bottled product, Cherry Sprite, was launched when Freestyle data indicated that demand for that combination of flavors would be strong in the United States. (Source: *buzzfeednews.com*)

Although big data is clearly one of the drivers for the strong demand for analytics, it is important to understand that, in some sense, big data issues are a subset of analytics. Many very valuable applications of analytics do not involve big data, but rather traditional data sets that are very manageable by traditional database and analytics software. The key to analytics is that it provides useful insights and better decision making using the data that are available—whether those data are "big" or not.

1.5 Business Analytics in Practice

Business analytics involves tools as simple as reports and graphs to those that are as sophisticated as optimization, data mining, and simulation. In practice, companies that apply analytics often follow a trajectory similar to that shown in Figure 1.2. Organizations start with basic analytics in the lower left. As they realize the advantages of these analytic techniques, they often progress to more sophisticated techniques in an effort to reap the derived competitive advantage. Therefore, predictive and prescriptive analytics are sometimes referred to as **advanced analytics**. Not all companies reach that level of usage, but those that embrace analytics as a competitive strategy often do.

Analytics has been applied in virtually all sectors of business and government. Organizations such as Procter & Gamble, IBM, UPS, Netflix, Amazon.com, Google, the Internal Revenue Service, and General Electric have embraced analytics to solve important problems or to achieve a competitive advantage. In this section, we briefly discuss some of the types of applications of analytics by application area.

Accounting Analytics

Applications of analytics in accounting are numerous and growing. For example, budget planning relies heavily on analytics. To construct a budget, predictive analytics in the form