Data Analytics for Accounting

SECOND EDITION

Vernon J. Richardson
University of Arkansas, Baruch College

Ryan A. Teeter
University of Pittsburgh

Katie L. Terrell
University of Arkansas

Mc Graw Hill

DATA ANALYTICS FOR ACCOUNTING, SECOND EDITION

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 LWI 24 23 22 21 20

ISBN 978-1-260-83783-4 (bound edition)
MHID 1-260-83783-1 (bound edition)
ISBN 978-1-260-90431-4 (loose-leaf edition)
MHID 1-260-90431-8 (loose-leaf edition)

Portfolio Manager: *Steve Schuetz*
Product Developers: *Sarah Wood; Alexandra Kukla*
Marketing Manager: *Claire McLemore*
Content Project Managers: *Fran Simon; Angela Norris*
Buyer: *Susan K. Culbertson*
Design: *Egzon Shaquiri*
Content Licensing Specialists: *Lori Hancock*
Cover Image: *Rawpixel.com/Shutterstock*
Compositor: *SPi Global*

All credits appearing on page or at the end of the book are considered to be an extension of the copyright page.

Library of Congress Cataloging-in-Publication Data

Names: Richardson, Vernon J., author. | Teeter, Ryan, author. | Terrell, Katie, author.
Title: Data analytics for accounting / Vernon J. Richardson, University of Arkansas, Xi'an Jiaotong Liverpool University, Ryan A. Teeter, University of Pittsburgh, Katie L. Terrell, University of Arkansas.
Description: Second Edition. | New York : McGraw-Hill Education, 2020. | Revised edition of the authors' Data analytics for accounting, [2019]
Identifiers: LCCN 2019034761 | ISBN 9781260837834 (paperback)
Subjects: LCSH: Accounting–Data processing.
Classification: LCC HF5679 .R534 2020 | DDC 657.0285–dc23
LC record available at https://lccn.loc.gov/2019034761

mheducation.com/highered

Dedications

**My wonderful daughter, Melissa,
for your constant love, encouragement and support.**

–Vern Richardson

**My wife, Erin, and children, Sylvia
and Theodore.**

–Ryan Teeter

To my co-author, friend, and colleague, Vernon Richardson. Thank you for inviting me to be on this textbook journey. And thank you for your guidance and patience–I'm thrilled to be a part of your team!

–Katie Terrell

Preface

Data Analytics is changing the business world—data simply surrounds us! So much data is available to businesses about each of us—how we shop, what we read, what we buy, what music we listen to, where we travel, whom we trust, where we invest our time and money, etc. Accountants can create value by addressing fundamental business and accounting questions using data analytics.

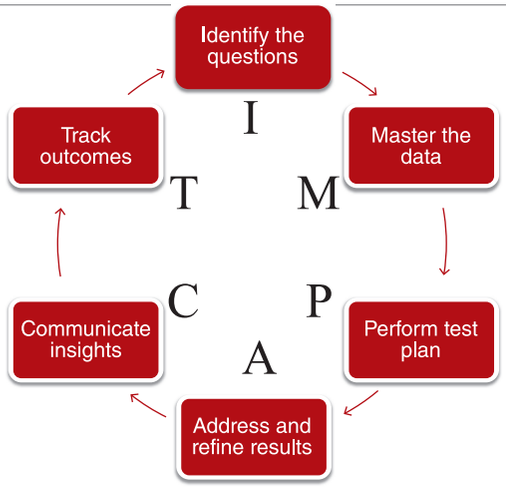
All accountants must develop data analytic skills to address the needs of the profession in the future. *Data Analytics for Accounting, 2e* recognizes that accountants don't need to become data scientists—they may never need to build a data repository or do the real hard-core Data Analytics or learn how to program a computer to do machine learning. However, there are seven skills that analytic-minded accountants must have to be prepared for a data-filled world, including:

1. An analytics mindset—recognize when and how Data Analytics can address accounting questions.
2. Data scrubbing and data preparation—comprehend the process needed to extract (query), clean and prepare the data before analysis.
3. Data quality—recognize what is meant by data quality, be it completeness, reliability, or validity.
4. Descriptive data analysis—perform basic analysis to understand the quality of the underlying data and their ability to address the business question.
5. Data analysis through data manipulation—demonstrate ability to sort, rearrange, merge, and reconfigure data in a manner that allows enhanced analysis.
6. Problem solving through statistical data analysis—identify and implement an approach that will use statistical data analysis to draw conclusions and make recommendations on a timely basis.
7. Data visualization and data reporting—report results of analysis in an accessible way to each varied decision maker and his or her specific needs.

Consistent with these skills, it's important to recognize that Data Analytics is a process. The process begins by identifying business questions that can be addressed with data, extracting and testing the data, refining our testing, and finally, communicating those findings to management. *Data Analytics for Accounting,* 2e describes this process by relying on an established data analytics model called the IMPACT cycle[1]

1. **I**dentify the question.
2. **M**aster the data.
3. **P**erform test plan.
4. **A**ddress and refine results.
5. **C**ommunicate insights.
6. **T**rack outcomes

[1]Jean Paul Isson and Jesse S. Harriott, *Win with Advanced Business Analytics: Creating Business Value from Your Data* (Hoboken, NJ: Wiley, 2013).



Adapted from *Win with Advanced Business Analytics: Creating Business Value from Your Data*, by Jean Paul Isson and Jesse S. Harriott.

The IMPACT cycle is described in the first four chapters and then the process is illustrated in audit, managerial accounting, financial accounting and tax in Chapters 5-9, adding an all-new tax chapter to *Data Analytics for Accounting,* 2e. In response to instructor feedback, *Data Analytics for Accounting,* 2e now also includes two new project chapters, giving students a chance to practice the full IMPACT model with multiple labs that build on each other.

Data Analytics for Accounting, 2e emphasizes hands-on practice. Students are provided with hands-on instruction (e.g., click-by-click instructions, screenshots, etc.) on datasets within the chapter; within the end-of-chapter materials; and in the labs at the end of each chapter. Throughout the text, students identify questions, extract and download data, perform testing, and then communicate the results of that testing.

The use of real-world data is highlighted by using data from **LendingClub**, **College Scorecard**, **Dillard's**, the **State of Oklahoma**, as well as other data from our labs. In particular, we emphasize the rich data from **Dillard's** sales transactions that we use in more than fifteen of the labs throughout the text (including Chapter 11).

Data Analytics for Accounting, 2e also emphasizes the various data analysis tools students will use throughout the rest of their career—Microsoft Excel, Microsoft Access (including SQL), Tableau (free student license), IDEA (free student license), and Weka (free student license). Using multiple tools allows students to learn which tool is best suited for the necessary data analysis, data visualization, and communication of the insights gained—for example, which tool is easiest for internal controls testing, which is best for analysis or querying (using SQL) big datasets, which is best for data visualizations, and so on.

About the Authors

Vernon J. Richardson

Vernon J. Richardson is a Distinguished Professor of Accounting and the G. William Glezen Chair in the Sam M. Walton College of Business at the University of Arkansas and a Visiting Professor at Baruch College. He received his BS, Master of Accountancy, and MBA from Brigham Young University and a PhD in accounting from the University of Illinois at Urbana–Champaign. He has taught students at the University of Arkansas, Baruch College, University of Illinois, Brigham Young University, Aarhus University, and University of Kansas and internationally at the China Europe International Business School (Shanghai), Xi'an Jiaotong Liverpool University, and the University of Technology Sydney.

Dr. Richardson is a member of the American Accounting Association. He has served as president of the American Accounting Association Information Systems section. He previously served as an editor of *The Accounting Review* and is currently an editor at *Accounting Horizons.* He has published articles in *The Accounting Review, Journal of Information Systems, Journal of Accounting and Economics, Contemporary Accounting Research, MIS Quarterly, International Journal of Accounting Information Systems, Journal of Management Information Systems, Journal of Operations Management,* and *Journal of Marketing.* Dr. Richardson is also the author of McGraw-Hill's *Accounting Information Systems* textbook.

Ryan A. Teeter

Ryan A. Teeter is a Clinical Assistant Professor of Accounting in the Katz Graduate School of Business at the University of Pittsburgh. He teaches accounting information systems, auditing, and accounting data analytics. Prior to receiving his PhD in accounting information systems from Rutgers University, he worked at Google in Mountain View, California. He has since worked with internal audit organizations at Siemens, Procter & Gamble, Alcoa/Arconic, and FedEx, helping to develop robotic process automation programs and data analytic solutions.

Dr. Teeter is a member of the American Accounting Association and has published articles in the *Journal of Strategic Technologies in Accounting* and *Issues in Accounting Education.* He has received grant funding for data analytics research from PwC.

Katie L. Terrell

Katie L. Terrell is an instructor in the Sam M. Walton College of Business at the University of Arkansas. She received her BA degrees in English literature and in the Spanish language from the University of Central Arkansas and her MBA from the University of Arkansas. She expects a doctoral degree by 2020. She has taught students at the University of Arkansas; Soochow University (Suzhou, China); the University College Dublin (Ireland); and Duoc UC, a branch of the Catholic University of Chile (Vina del Mar, Chile).

She is a member of the American Accounting Association and has published a *Statement on Management Accounting* for the Institute of Management Accountants on managing organizational change in operational change initiatives. Terrell was named the 2019 Business Professional of the Year (Education) by the national Beta Alpha Psi organization. She has recently been recognized for her innovative teaching by being the recipient of the Mark Chain/FSA Teaching Award for innovative graduate-level accounting teaching practices in 2016. She has worked with Tyson Foods, where she held various information system roles, focusing on business analysis, project management for ERP implementations and upgrades, and organizational change management.

Acknowledgments

Our sincere thanks to all who helped us on this project.

Our biggest thanks to the awesome team at McGraw-Hill Education, including Steve Schuetz, Tim Vertovec, Allie Kukla, Fran Simon, Kevin Moran, and Sarah Wood.

Our thanks also to each of the following:

The Walton College Enterprise Team (Paul Cronan, Ron Freeze, Michael Gibbs, Michael Martz, Tanya Russell) for their work helping us get access to the Dillard's data.

Shane Lunceford from LendingClub for helping gain access to LendingClub data.

Marcia Watson, University of North Carolina–Charlotte; Ryan Baxter, Boise State University; Antoinette Smith, Florida International University; and Lorrie Metzger, University at Buffalo for their accuracy check and review of the manuscript.

In addition, the following reviewers and classroom testers who provided ideas and insights for this edition. We appreciate their contributions.

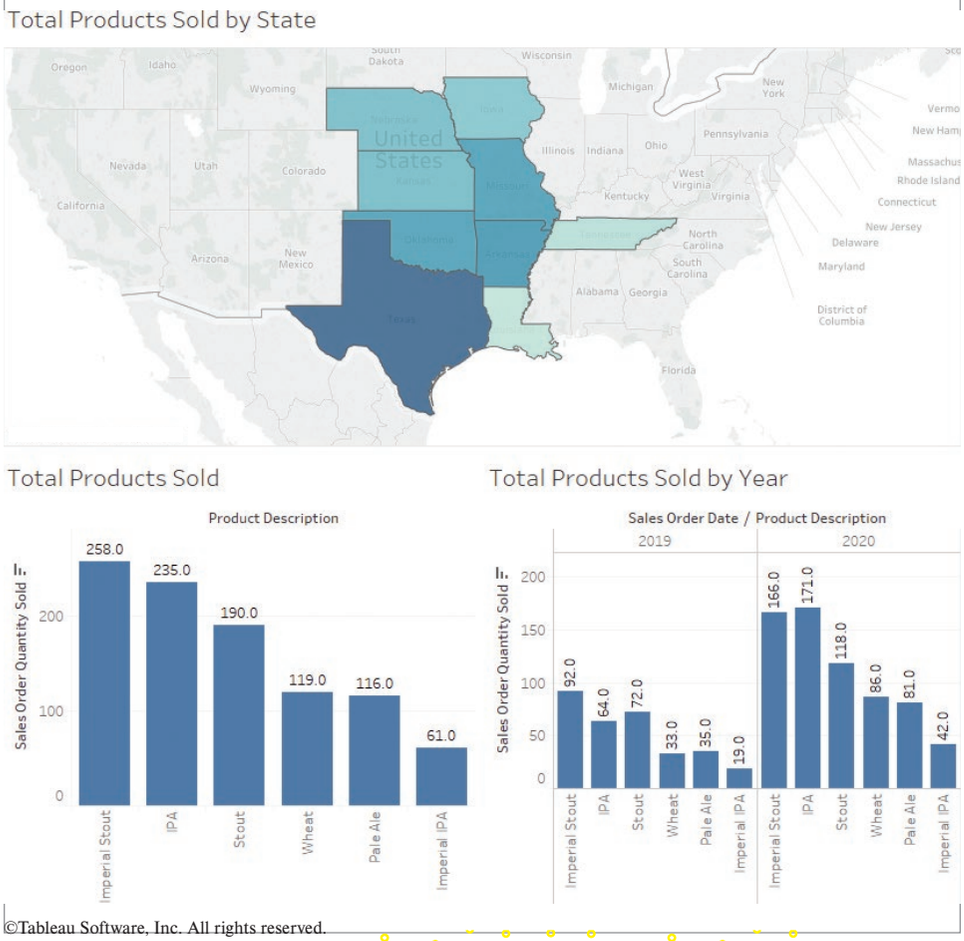
Amelia Annette Baldwin
University of South Alabama
Dereck Barr-Pulliam
University of Wisconsin–Madison
Heather Carrasco
Texas Tech University
Elizabeth Felski
State University of New York at Geneseo
Chris C. Hsu
York College, City University of New York
Venkataraman Iyer
University of North Carolina at Greensboro
Andrea S. Kelton
Middle Tennessee State University
Brandon Lock
Baruch College, CUNY
Sharon M. Lightner
National University
Margarita Maria Lenk
Colorado State University
Partha Mohapatra
California State University, Sacramento
Uday Murthy
University of South Florida

Vernon Richardson
Ryan Teeter
Katie Terrell

Kalana Malimage
University of Wisconsin–Whitewater
Bonnie Morris
Duquesne University
Kathy Nesper
University at Buffalo
Ali Saeedi
University of Minnesota Crookston
Karen Schuele
John Carroll University
Drew Sellers
Kent State University
Joe Shangguan
Robert Morris University
Vincent J. Shea
St. John's University
Marcia Watson
University of North Carolina at Charlotte
Liu Yang
Southeast Missouri State University
Zhongxia Ye
University of Texas, San Antonio
Qiongyao (Yao) Zhang
Robert Morris University

Key Features

- **Emphasis on Skills:** Working through the IMPACT cycle framework, students will learn problem assessment, data preparation, data analysis, data visualization, control contesting, and more.
- **Emphasis on Hands-On Practice:** Students will be provided hands-on learning (click-by-click instructions with screenshots) on datasets within each chapter, within the end-of-chapter materials, and in the labs and comprehensive cases.
- **Emphasis on Datasets:** To illustrate data analysis techniques and skills, multiple practice datasets (audit, financial, and managerial data) will be used in every chapter. Students gain real-world experience working with data from **LendingClub, Dillard's, College Scorecard,** the **State of Oklahoma,** as well as financial statement data (via XBRL) from *Fortune* 100 companies.
- **Emphasis on Tools:** Students will learn how to conduct data analysis using Excel Access (including SQL), Tableau (free student license), IDEA (free student license), and Weka (free student license). Students will compare and contrast the different tools to determine which are best suited for basic data analysis and data visualization, which are easiest for internal controls testing, which are best for SQL queries, and so on.



Main Text Features

Chapter Maps

These maps provide a guide of what we're going to cover in the chapter as well as a guide of what we've just learned and what's coming next.

Chapter-Opening Vignettes

Because companies are facing the new and exciting opportunities with their use of Data Analytics to help with accounting and business decisions, we detail what they're doing and why in our chapter-opening vignettes.

Wichy/Shutterstock

We are lucky to live in a world in which data are abundant. However, even with rich sources of data, when it comes to being able to analyze data and turn them into useful information and insights, very rarely can an analyst hop right into a dataset and begin analyzing. Datasets almost always need to be cleaned and validated before they can be used. Not knowing how to clean and validate data can, at best, lead to frustration and poor insights and, at worst, lead to horrible security violations. While this text takes advantage of open source datasets, these datasets have all been scrubbed not only for accuracy, but also to protect the security and privacy of any individual or company whose details were in the original dataset.

In 2015, a pair of researchers named Emil Kirkegaard and Julius Daugbjerg Bjerrekaer scraped data from **OkCupid**, a free dating website, and provided the data onto the "Open Science Framework," a platform researchers use to obtain and share raw data. While the aim of the Open Science Framework is to increase transparency, the researchers in this instance took that a step too far—and a step into illegal territory. Kirkegaard and Bjerrekaer did not obtain permission from **OkCupid** or from the 70,000 **OkCupid** users whose identities, ages, genders, religions, personality traits, and other personal details maintained by the dating site were provided to the public without any work being done to anonymize or sanitize the data. If the researchers had taken the time to not just validate that the data were complete, but also to sanitize them to protect the individuals' identities, this would not have been a threat or a news story. On May 13, 2015, the Open Science Framework removed the **OkCupid** data from the platform, but the damage of the privacy breach had already been done.[1]

OBJECTIVES

After reading this chapter, you should be able to:

- **LO 2-1** Understand how data are organized in an accounting information system.
- **LO 2-2** Understand how data are stored in a relational database.
- **LO 2-3** Explain and apply extraction, transformation, and loading (ETL) techniques.

Learning Objectives

We feature learning objectives at the beginning of each chapter. Having these learning objectives provides students with an overview of the concepts to be taught in the chapter and the labs.

Progress Checks

Periodic progress check questions are posed to the students throughout each chapter. These checks provoke the student to stop and consider the concepts presented.

PROGRESS CHECK

1. Referring to Exhibit 2-1, locate the relationship between the Supplier and Purchase Order tables. What is the unique identifier of each table? (The unique identifier attribute is called the primary key—more on how it's determined in the next learning objective.) Which table contains the attribute that creates the relationship? (This attribute is called the foreign key—more on how it's determined in the next learning objective.)
2. Referring to Exhibit 2-1, review the attributes in the Purchase Order table. There are two foreign keys listed in this table that do not relate to any of the tables in the diagram. Which tables do you think they are? What type of data would be

End-of-Chapter Materials

Answers to Progress Checks

Allow students to evaluate if they are on track with their understanding of the materials presented in the chapter.

ANSWERS TO PROGRESS CHECKS

1. The unique identifier of the Supplier table is [Supplier ID], and the unique i
   Purchase Order table is [PO No.]. The Purchase Order table contains the t
2. The foreign key attributes in the Purchase Order table that do not relate
   in the view are EmployeeID and CashDisbursementID. These attributes p
   to the Employee table (so that we can tell which employee was respon
   Purchase Order) and the Cash Disbursement table (so that we can tell i
   Orders have been paid for yet, and if so, on which check). The Employee t
   a complete listing of each Employee, as well containing the details about e
   (for example, phone number, address, etc.). The Cash Disbursement tab
   listing of the payments the company has made.

Multiple Choice Questions

Quickly assess student's knowledge of chapter content.

Multiple Choice Questions

1. Mastering the data can also be described via the ETL process. Th
   stands for:
   a. extract, total, and load data.
   b. enter, transform, and load data.
   c. extract, transform, and load data.
   d. enter, total, and load data.
2. Which of the following describes part of the goal of the ETL process:
   a. identify which approach to data analytics should be used.
   b. load the data into a relational database for storage.
   c. communicate the results and insights found through the analysis.

Discussion Questions

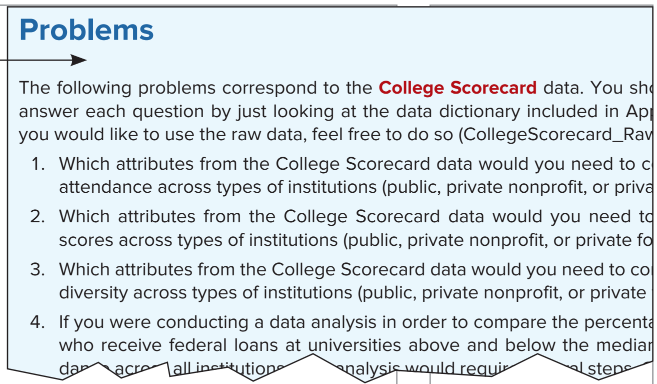
Provide questions for group discussion.

Discussion Questions

1. The advantages of a relational database include limiting the amount of r
   that are stored in a database. Why is this an important advantage? What
   when redundant data are stored?
2. The advantages of a relational database include integrating business p
   is it preferable to integrate business processes in one information syste
   store different business process data in separate, isolated databases?
3. Even though it is preferable to store data in a relational database, stori
   separate tables can make data analysis cumbersome. Describe three rea
   the trouble to store data in a relational database.
4. Among the advantages of using a relational database is enforcing busine
   on your understanding of how the structure of a relational database help
   redundancy and other advant[illegible]es, how does the prima[illegible]ey/foreign k

Problems

Challenge the student's ability to see relationships in the learning objectives by employing higher-level thinking and analytical skills.

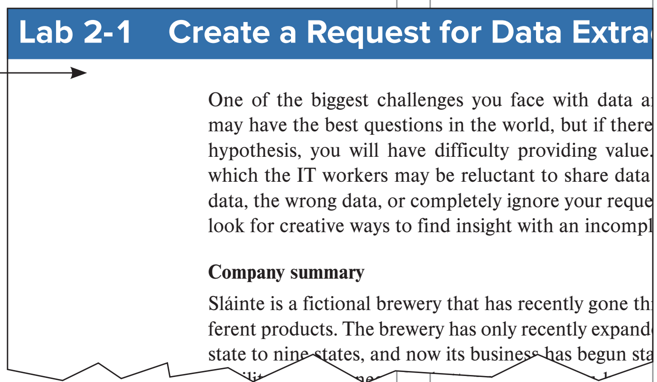

Problems

The following problems correspond to the **College Scorecard** data. You sh
answer each question by just looking at the data dictionary included in Ap
you would like to use the raw data, feel free to do so (CollegeScorecard_Raw

1. Which attributes from the College Scorecard data would you need to c
attendance across types of institutions (public, private nonprofit, or priva
2. Which attributes from the College Scorecard data would you need to
scores across types of institutions (public, private nonprofit, or private fo
3. Which attributes from the College Scorecard data would you need to co
diversity across types of institutions (public, private nonprofit, or private
4. If you were conducting a data analysis in order to compare the percenta
who receive federal loans at universities above and below the median

Labs

Give students hands-on experience working with different types of data and the tools used to analyze them. Students will conduct data analysis using Excel, Access (including SQL), Tableau, IDEA, XBRL, and Weka.


Lab 2-1 Create a Request for Data Extra

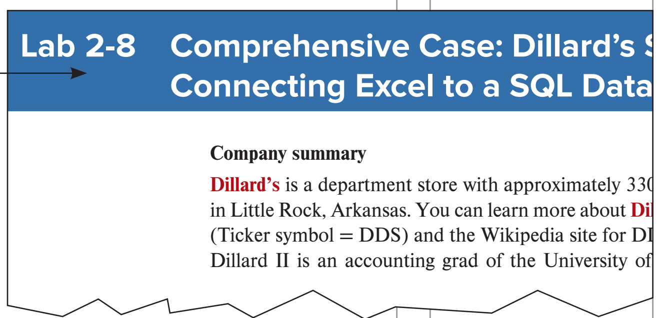
One of the biggest challenges you face with data a
may have the best questions in the world, but if there
hypothesis, you will have difficulty providing value.
which the IT workers may be reluctant to share data
data, the wrong data, or completely ignore your reque
look for creative ways to find insight with an incompl

Company summary

Sláinte is a fictional brewery that has recently gone th
ferent products. The brewery has only recently expand
state to nine states, and now its business has begun sta

Comprehensive Cases

Use a real-life Big Data set based on **Dillard's** actual company data. This dataset allows students to build their skills and test their conclusions across concepts covered in each chapter. The Comprehensive Cases can be followed continuously from the first chapter or picked up at any later point in the book; enough information is provided to ensure students can get right to work.


**Lab 2-8 Comprehensive Case: Dillard's S
Connecting Excel to a SQL Data**

Company summary

Dillard's is a department store with approximately 33(
in Little Rock, Arkansas. You can learn more about Di
(Ticker symbol = DDS) and the Wikipedia site for DI
Dillard II is an accounting grad of the University of

Data Analytics for Accounting, 2e Content Updates

General Updates for the 2nd Edition

- Added additional End-of-Chapter Multiple Choice Questions and Problems throughout the text.
- Significantly revised many End-of-Chapter Problems for availability and auto-grading within Connect.
- Revised and added many new Discussion Questions in most chapters.

Chapter by Chapter Updates

Specific chapter changes for *Data Analytics for Accounting,* 2nd Edition, are as follows:

Chapter 1

- Updated the opening vignette and statistics on Alibaba sales and use of e-commerce.
- Updated the statistics and screenshots for Lending Club Analysis.
- Revised Connect questions for problems and labs.

Chapter 2

- Improved and clarified the discussion of relational databases, including updated figures.
- Expanded the discussion of different RDBMS (Access, SQLite, and SQL Server).
- Improved discussion of Excel and SQL. The brief introduction to how to use SQL now has its own place in a dedicated appendix at the end of the text, and it has been vastly expanded to teach beginners how to write queries.
- Expanded the discussion on data quality.
- Added a brief discussion of ETL v. ELT.
- Improved labs for clarity and a better learning experience, particularly Labs 2-1, 2-2, and 2-4.

Chapter 3

- Reorganized chapter structure to follow the descriptive, diagnostic, predictive, and prescriptive approaches to Data Analytics.
- New exhibits and examples to illustrate analytics approaches.
- Removed previous edition flowchart for model selection.
- Additional explanation and examples of each of the methods and approaches.
- Improved labs for clarity.

Chapter 4

- Updated the opening vignette.
- Improved the discussion on the differences between qualitative and quantitative data and the discussion of the normal distribution.
- Improved and clarified how to select a visualization based on the four chart types (qualitative vs. quantitative and declarative vs. exploratory).

- Updated the discussion on the Gartner Quadrant to take into account Gartner's January 2019 analysis of BI tools (focusing on Excel and Tableau).
- Extended the discussion on written and spoken communication.
- Added a lab to work with visualizing data and creating dashboards in Power BI to interactively compare the tool with Tableau.

Chapter 5

- Expanded discussion on the modern data environment.
- Included additional examples of the Audit Data Standard.
- Improved and clarified content to match the focus on descriptive, diagnostic, predictive, and prescriptive analytics.
- New labs (5-1 and 5-2) that have students transform data using a common data model.
- Improved existing labs.

Chapter 6

- Clarified chapter content to match the focus on descriptive, diagnostic, predictive, and prescriptive analytics.
- Improved labs.

Chapter 7

- Clarified chapter content and provided additional new exhibits and examples, such as variance analysis.
- Improved labs.

Chapter 8

- Reorganized chapter content to focus on financial statement analysis using descriptive, diagnostic, predictive, and prescriptive approaches.
- Added new content on common size and ratio analysis.
- Improved discussion of XBRL data.
- Improved XBRL dataset (in Lab 8-4), accessible via Microsoft Access and included options to do analysis in Excel.

Chapter 9

- All-new chapter on tax analytics, including examples of tax data, tax analysis, tax planning, and tax visualizations.

Chapter 10

- All-new basic project chapter that explores the order-to-cash and procure-to-pay cycles from different user perspectives.

Chapter 11

- All-new advanced project chapter, estimating sales returns at **Dillard's** with three question sets highlighting descriptive and exploratory analysis, hypothesis testing, and predictive analytics.

Appendixes

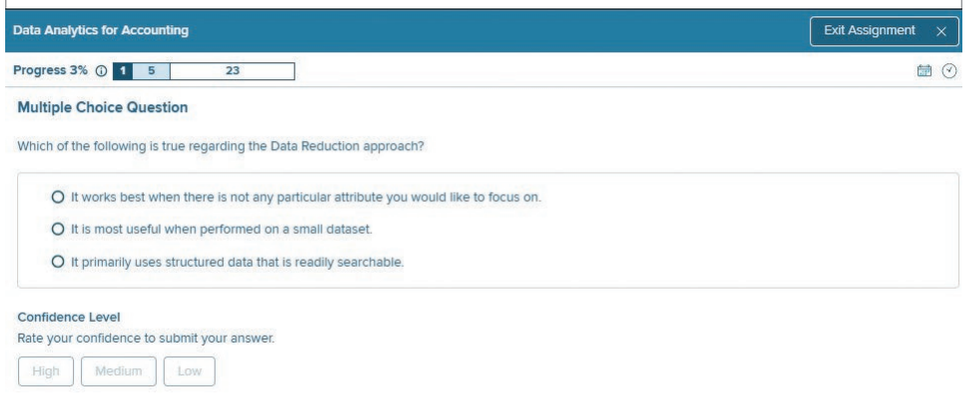
Several all-new appendixes have been added to ease the lab experience and introduce tools used or mentioned throughout the text:

- Appendix A: Basic Statistics Tutorial.
- Appendix B: Accessing the Excel Data Analysis Toolpak.
- Appendix C: Excel (Formatting, Sorting, Filtering, and PivotTables).
- Appendix D: SQL Part 1. This tutorial introduces the SQL language for extracting data and explains the following SQL syntax: SELECT, FROM, INNER JOIN, ON, WHERE, GROUP BY, HAVING, ORDER BY.
- Appendix E: SQLite. We have added SQLite files as an option for each lab that uses Microsoft Access. This lab explains how to download SQLite and how to use the tool.
- Appendix F: Power Query. This appendix contains a short tutorial on transforming data using Power Query. How to access data files on the University of Arkansas' remote desktop is also discussed.
- Appendix G: Tableau.
- Appendix H: SQL Part 2: On the heels of learning Tableau, students learn about more complex joins–LEFT and RIGHT.
- Appendix I: Power BI.
- Appendix J: Dillard's ER Diagram.
- Appendix K: Data Dictionaries.

Connect for Data Analytics for Accounting

With **Connect** for Data Analytics in Accounting, your students receive proven study tools and hands-on assignment materials as well as an adaptive eBook. All of the following assets are assignable in Connect.

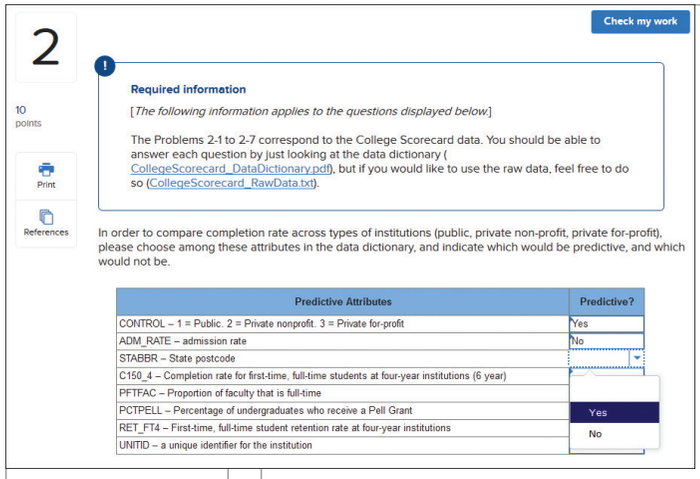
SmartBook 2.0: A personalized and adaptive learning tool used to maximize the learning experience by helping students study more efficiently and effectively. Smartbook 2.0 highlights where in the chapter to focus, asks review questions on the materials covered and tracks the most challenging content for later review recharge. Smartbook 2.0 is available both online and offline.



Orientation Videos: Video-based tutorial assignments are designed to train students via an overview video followed by a quiz for each of the assignment types they will find in Connect.

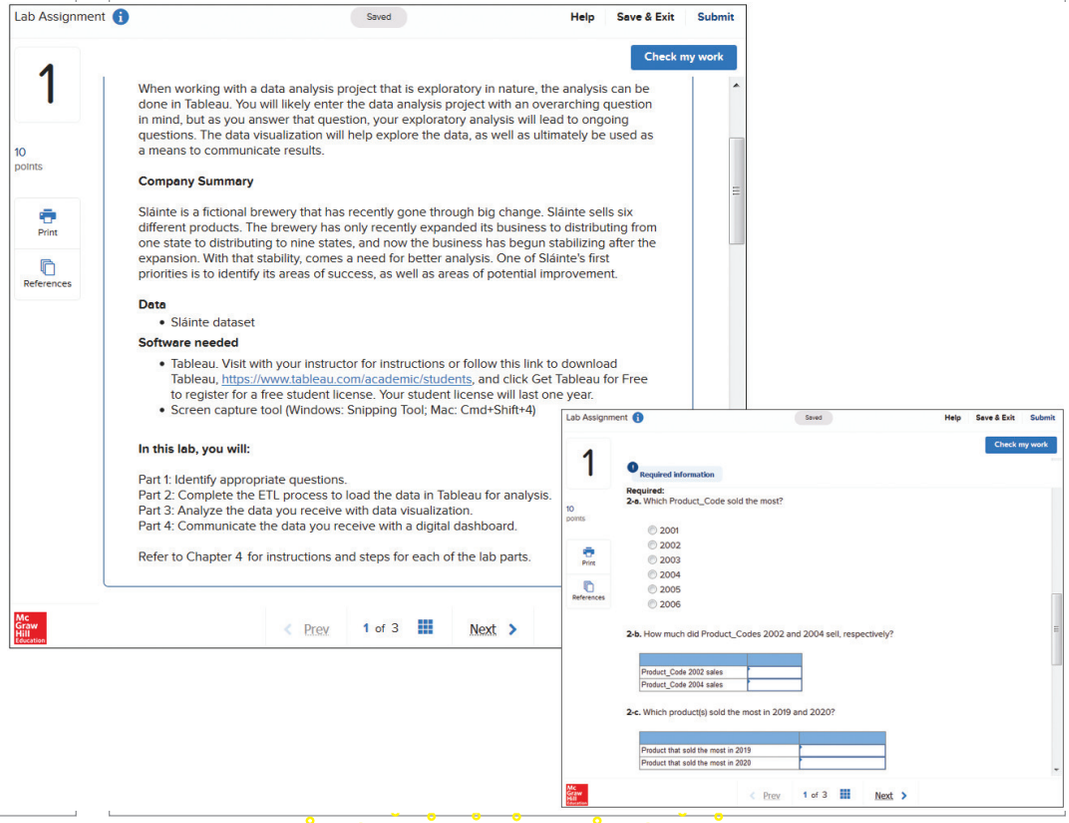
Multiple Choice Questions: The multiple choice questions from the end-of-chapter materials are assignable in Connect, providing students with instant feedback on their answers.

Problems: Select problems from the text are available for assignment in Connect to ensure students are building an analytical skill set.



Labs: Select labs are assignable in Connect but will require students to work outside of Connect to complete the lab. Once completed, students go back into Connect to answer questions designed to ensure they completed the lab and understood the key skills and outcomes from their lab work.

Comprehensive Cases: Select comprehensive labs/cases are assignable in Connect but will require students to work outside of Connect to complete the lab using the **Dillard's** real-world Big Data set. Once students complete the comprehensive lab, they will go back into Connect to answer questions designed to ensure they completed the lab and understood the key skills and outcomes from their lab work.



Lab Walkthrough Videos: Get the help you need, when you need it. These author-led videos will explain how to access and use the tools needed to complete processes essential to the labs.

Author Lecture Videos: Lecture Videos teach each chapter's core learning objectives and concepts through an author-developed, hands-on presentation, bringing the text content to life. The videos have the touch and feel of a live lecture, rather than a canned presentation, so you can learn at your own pace.

Test Bank: The test bank includes auto-graded multiple choice and true/false assessment questions. It is available in Connect and Test Builder.

Mc Graw Hill connect®

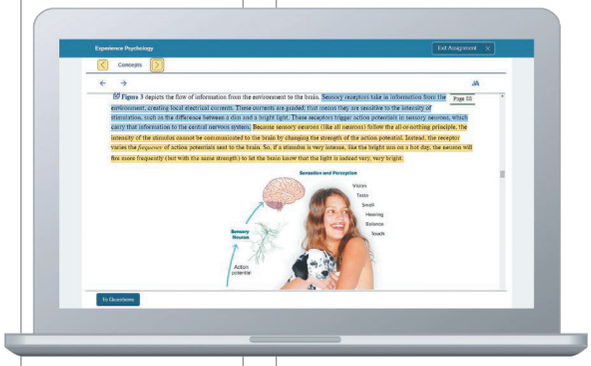
FOR INSTRUCTORS

You're in the driver's seat.

Want to build your own course? No problem. Prefer to use our turnkey, prebuilt course? Easy. Want to make changes throughout the semester? Sure. And you'll save time with Connect's auto-grading too.

65%

Less Time Grading



Laptop: McGraw-Hill; Woman/dog: George Doyle/Getty Images

They'll thank you for it.

Adaptive study resources like SmartBook® 2.0 help your students be better prepared in less time. You can transform your class time from dull definitions to dynamic debates. Find out more about the powerful personalized learning experience available in SmartBook 2.0 at **www.mheducation.com/highered/connect/smartbook**

Make it simple, make it affordable.

Connect makes it easy with seamless integration using any of the major Learning Management Systems—Blackboard®, Canvas, and D2L, among others—to let you organize your course in one convenient location. Give your students access to digital materials at a discount with our inclusive access program. Ask your McGraw-Hill representative for more information.

Padlock: Jobalou/Getty Images

Solutions for your challenges.

A product isn't a solution. Real solutions are affordable, reliable, and come with training and ongoing support when you need it and how you want it. Our Customer Experience Group can also help you troubleshoot tech problems—although Connect's 99% uptime means you might not need to call them. See for yourself at **status.mheducation.com**

Checkmark: Jobalou/Getty Images

SUPPORT AT every step

FOR STUDENTS

Effective, efficient studying.

Connect helps you be more productive with your study time and get better grades using tools like SmartBook 2.0, which highlights key concepts and creates a personalized study plan. Connect sets you up for success, so you walk into class with confidence and walk out with better grades.

Study anytime, anywhere.

Download the free ReadAnywhere app and access your online eBook or SmartBook 2.0 assignments when it's convenient, even if you're offline. And since the app automatically syncs with your eBook and SmartBook 2.0 assignments in Connect, all of your work is available every time you open it. Find out more at **www.mheducation.com/readanywhere**

"I really liked this app—it made it easy to study when you don't have your textbook in front of you."

- Jordan Cunningham, Eastern Washington University

Calendar: owattaphotos/Getty Images

No surprises.

The Connect Calendar and Reports tools keep you on track with the work you need to get done and your assignment scores. Life gets busy; Connect tools help you keep learning through it all.

Learning for everyone.

McGraw-Hill works directly with Accessibility Services Departments and faculty to meet the learning needs of all students. Please contact your Accessibility Services office and ask them to email accessibility@mheducation.com, or visit **www.mheducation.com/about/accessibility** for more information.

Top: Jenner Images/Getty Images, Left: Hero Images/Getty Images, Right: Hero Images/Getty Images

Brief Table of Contents

Detailed TOC

Chapter 1

Data Analytics for Accounting and Identifying the Questions

A Look at This Chapter

Data Analytics is changing the business world. In this chapter, we define it and explain its impact on business and the accounting profession, noting that the value of Data Analytics is in the insights it provides. We also describe how to develop an analytics mindset. We describe the Data Analytics Process using the IMPACT cycle model and explain how this process is used to address both business and accounting questions. We specifically emphasize the importance of identifying appropriate questions that Data Analytics might be able to address.

A Look Ahead

Chapter 2 provides a description of how data are prepared and scrubbed to be ready for analysis to answer business questions. We explain how to extract, transform, and load data and then how to validate and normalize the data. In addition, we explain how data standards are used to facilitate the exchange of data between senders and receivers.

Shutterstock

The Chinese e-commerce company **Alibaba** is perhaps the biggest online commerce company in the world. Using its three main websites, Taobao.com, Tmall.com, and Alibaba.com, it hosts millions of businesses and hundreds of millions of users with $345 billion in 2018 sales last year (more than **eBay** and **Amazon** combined!). With so many transactions and so many users, Alibaba has worked to capture fraud signals directly from its extensive database of user behaviors and its network. It then analyzes them in real time using machine learning to accurately sort the potentially fraudulent users from the good ones. Alibaba has developed five stages of fraud detection for each user: (1) account check, (2) device check, (3) activity check, (4) risk strategy, and (5) manual review. These stages all combine to develop a risk score for each user. This fraud risk prevention score is so valuable to Alibaba and others that Alibaba shares and sells it to external customers. What will Data Analytics do next?

Sources: J. Chen, Y. Tao, H. Wang, and T. Chen, "Big Data Based Fraud Risk Management at Alibaba," *Journal of Finance and Data Science* 1, no. 1 (2015), pp. 1–10; and K. Pal, "How to Combat Financial Fraud by Using Big Data," 2016, http://www.kdnuggets.com/2016/03/combat-financial-fraud-using-big-data.html.

OBJECTIVES

After reading this chapter, you should be able to:

- **LO 1-1** Define Data Analytics.
- **LO 1-2** Understand why Data Analytics matters to business.
- **LO 1-3** Explain why Data Analytics matters to accountants.
- **LO 1-4** Describe the Data Analytics Process using the IMPACT cycle.
- **LO 1-5** Describe the skills needed by accountants.
- **LO 1-6** Explain how to translate common business questions into fields and values.

LO 1-1

Define Data Analytics.

DATA ANALYTICS

Data surrounds us! By the year 2020, about 1.7 megabytes of new information will be created every second for every human being on the planet. In fact, more data have been created in the last 2 years than in the entire previous history of the human race.[1] With so much data available about each of us (i.e., how we shop, what we read, what we've bought, what music we listen to, where we travel, whom we trust, etc.), arguably, there is the potential for analyzing those data in a way that can answer fundamental business questions and create value.

We define **Data Analytics** as the process of evaluating data with the purpose of drawing conclusions to address business questions. Indeed, effective Data Analytics provides a way to search through large structured and unstructured data to discover unknown patterns or relationships.[2] In other words, Data Analytics often involves the technologies, systems, practices, methodologies, databases, statistics, and applications used to analyze diverse business data to give organizations the information they need to make sound and timely business decisions.[3] That is, the process of Data Analytics aims to transform raw data into knowledge to create value.

Big Data refers to datasets that are too large and complex for businesses' existing systems to handle utilizing their traditional capabilities to capture, store, manage, and analyze these datasets. Another way to describe Big Data is by use of 3 Vs: its volume (the sheer size of the dataset), velocity (the speed of data processing), and variety (the number of types of data). While sometimes *Data Analytics* and *Big Data* are terms used interchangeably, we will use the term *Data Analytics* throughout and focus on the ability to turn data into knowledge and knowledge into value.

PROGRESS CHECK

1. How does having more data around us translate into value for a company?
2. Banks know a lot about us, but they have traditionally used externally generated credit scores to assess creditworthiness when deciding whether to extend a loan. How would you suggest a bank use Data Analytics to get a more complete view of its customers' creditworthiness? Assume the bank has access to a customer's loan history, credit card transactions, deposit history, and direct deposit registration. How could it assess whether a loan might be repaid?

LO 1-2

Understand why Data Analytics matters to business.

HOW DATA ANALYTICS AFFECTS BUSINESS

There is little question that the impact of data analytics on business is overwhelming. In fact, in **PwC**'s 18th Annual Global CEO Survey, 86 percent of chief executive officers (CEOs) say they find it important to champion digital technologies and emphasize a clear vision of using technology for a competitive advantage, while 85 percent say they put a high value on Data Analytics. In fact, per **PwC**'s 6th Annual Digital IQ survey

[1] http://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#2a3289006c1d (accessed March 2019).
[2] Roger S. Debreceny and Glen L. Gray, "IT Governance and Process Maturity: A Multinational Field Study," *Journal of Information Systems* 27, no. 1 (Spring 2013), pp. 157–88.
[3] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence Research," *MIS Quarterly* 34, no. 1 (2010), pp. 201–3.

of more than 1,400 leaders from digital businesses, the area of investment that tops CEOs' list of priorities is business analytics.[4]

A recent study from McKinsey Global Institute estimates that Data Analytics could generate up to $3 trillion in value per year in just a subset of the total possible industries affected.[5] Data Analytics could very much transform the manner in which companies run their businesses in the near future because the real value of data comes from Data Analytics. With a wealth of data on their hands, companies use Data Analytics to discover the various buying patterns of their customers, investigate anomalies that were not anticipated, forecast future possibilities, and so on. For example, with insight provided through Data Analytics, companies could execute more directed marketing campaigns based on patterns observed in their data, giving them a competitive advantage over companies that do not use this information to improve their marketing strategies. Patterns discovered from past archives also enable businesses to identify opportunities and risks and better plan for the future. In addition to producing more value externally, studies show that Data Analytics affects internal processes, improving productivity, utilization, and growth.[6]

PROGRESS CHECK

3. Let's assume a brand manager at **Samsung** identifies that an older demographic might be concerned with the use of a Samsung Galaxy smartphone and the radiation impact it might have on the brain. How might **Samsung** use Data Analytics to assess if this is a problem?
4. How might Data Analytics assess the higher cost of paying employees to work overtime? Consider how Data Analytics might be helpful in reducing a company's overtime direct labor costs in a manufacturing setting.

HOW DATA ANALYTICS AFFECTS ACCOUNTING

LO 1-3

Explain why Data Analytics matters to accountants.

Data Analytics is expected to have dramatic effects on auditing and financial reporting as well as tax and managerial accounting. We detail how we think this might happen in each of the following sections.

Auditing

Data Analytics plays an increasingly critical role in the future of audit. In a recent *Forbes Insights*/KPMG report, "Audit 2020: A Focus on Change," the vast majority of survey respondents believe both that:

1. Audit must better embrace technology.
2. Technology will enhance the quality, transparency, and accuracy of the audit.

Indeed, "As the business landscape for most organizations becomes increasingly complex and fast-paced, there is a movement toward leveraging advanced business analytic techniques to

[4]"Data Driven: What Students Need to Succeed in a Rapidly Changing Business World," PwC, https://www.pwc.com/us/en/faculty-resource/assets/pwc-data-driven-paper-feb2015.pdf, February 2015 (accessed March 20, 2019).
[5]"Open Data: Unlocking Innovation and Performance with Liquid Information," McKinsey Global Institute, http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information, October 2013 (accessed September 7, 2015).
[6]Joseph Kennedy, "Big Data's Economic Impact," https://www.ced.org/blog/entry/big-datas-economic-impact, December 3, 2014 (accessed January 9, 2016).

refine the focus on risk and derive deeper insights into an organization."[7] Many auditors believe that auditor data analytics will, in fact, lead to deeper insights that will enhance audit quality. This sentiment of the impact of Data Analytics on the audit has been growing for several years now and has given many public accounting firms incentives to invest in technology and personnel to capture, organize, and analyze financial statement data to provide enhanced audits, expanded services, and added value to their clients. As a result, Data Analytics is expected to be the next innovation in the evolution of the audit and professional accounting industry.

Given the fact that operational data abound and are easier to collect and manage, combined with CEOs' desires to utilize these data, the accounting firms may now approach their engagements with a different mindset. No longer will they be simply checking for errors, material misstatements, fraud, and risk in financial statements or merely be reporting their findings at the end of the engagement. Instead, audit professionals will now be collecting and analyzing the company's data similar to the way a business analyst would to help management make better business decisions. This means that, in many cases, external auditors will stay engaged with clients beyond the audit. This is a significant paradigm shift. The audit process will be changed from a traditional process toward a more automated one, which will allow audit professionals to focus more on the logic and rationale behind data queries and less on the gathering of the actual data.[8] As a result, audits will not only yield important findings from a financial perspective, but also information that can help companies refine processes, improve efficiency, and anticipate future problems.

> "It's a massive leap to go from traditional audit approaches to one that fully integrates big data and analytics in a seamless manner."[9]

Data Analytics also expands auditors' capabilities in services like testing for fraudulent transactions and automating compliance-monitoring activities (like filing financial reports to the SEC or to the IRS). This is possible because Data Analytics enables auditors to analyze the complete dataset, rather than the sampling of the financial data done in a traditional audit. Data Analytics enables auditors to improve its risk assessment in both its substantive and detailed testing.

Financial Reporting

Data Analytics also potentially has an impact on financial reporting. With the use of so many estimates and valuations in financial accounting, some believe that employing Data Analytics may substantially improve the quality of the estimates and valuations. Data from within an enterprise system and external to the company and system might be used to address many of the questions that face financial reporting. Many financial statement accounts are just estimates and so accountants often ask themselves questions like this to evaluate those estimates:

1. How much of the accounts receivable balance will ultimately be collected? What should the allowance for loan losses look like?
2. Is any of our inventory obsolete? Should our inventory be valued at market or cost (applying the lower-of-cost-or-market rule)? When will it be out of date? Do we need to offer a discount on it now to get it sold?

[7]Deloitte, "Adding Insight to Audit: Transforming Internal Audit Through Data Analytics." Accessed January 10, 2016. http://www2.deloitte.com/content/dam/Deloitte/ca/Documents/audit/ca-en-audit-adding-insight-to-audit.pdf.

[8]PwC, "Data Driven: What Students Need to Succeed in a Rapidly Changing Business World," http://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf, February 2015 (accessed January 9, 2016).

[9]EY, "How Big Data and Analytics Are Transforming the Audit." Accessed January 27, 2016. https://eyo-iis-pd.ey.com/ARC/documents/EY-reporting-ssue-9.pdf,posted April 2015.

3. Has our goodwill been impaired due to the reduction in profitability from a recent merger? Will it regain value in the near future?
4. How should we value contingent liabilities like warranty claims or litigation? Do we have the right amount?

Data Analytics may also allow an accountant or auditor to assess the probability of a goodwill write-down, warranty claims or the collectability of bad debts based on what customers, investors, and other stakeholders are saying about the company in blogs and in social media (like Facebook and Twitter). This information might help the firm determine both its optimal response to the situation and appropriate adjustment to its financial reporting.

It may be possible to use Data Analytics to scan the environment—that is, scan Google searches and social media (such as Instagram and Facebook) to identify potential risks and opportunities to the firm. For example, in a data analytics sense, it may allow a firm to monitor its competitors and its customers to better understand opportunities and threats around it. For example, are its competitors, customers, or suppliers facing financial difficulty, etc., that might affect the company's interactions with them and/or open up new opportunities that otherwise it wouldn't have considered?

Tax

Traditionally, tax work dealt with compliance issues based on data from transactions that have already taken place. Now, however, tax executives must develop sophisticated tax planning capabilities that assist the company with minimizing its taxes in such a way to avoid or prepare for a potential audit. This shift in focus makes tax data analytics valuable for its ability to help tax staffs to predict what will happen rather than reacting to what just did happen. Arguably, one of the things that Data Analytics does best is predictive analytics—predicting the future! An example of how tax data analytics might be used is the capability to predict the potential tax consequences of a potential international transaction, R&D investment, or proposed merger or acquisition.

One of the issues of performing predictive Data Analytics is the efficient organization and use of data stored across multiple systems on varying platforms that were not originally designed for the tax department. Organizing tax data into a data warehouse to be able to consistently model and query the data is an important step toward developing the capability to perform tax data analytics. This issue is exemplified by the 29 percent of tax departments that find the biggest challenge in executing an analytics strategy is integration with the IT department and available technology tools.[10]

PROGRESS CHECK

5. How could the use of internal audit data analytics find the pattern that one accountant enters the majority of the journal entries each quarter? How might this data be used to check if segregation of duties was appropriately maintained? Why might this be an issue that would need addressing?
6. How specifically will Data Analytics change the way a tax staff does its taxes?

[10]Deloitte, "The Power of Tax Data Analytics," http://www2.deloitte.com/us/en/pages/tax/articles/top-ten-things-about-tax-data-analytics.html (accessed October 12, 2016).

LO 1-4

Describe the Data Analytics Process using the IMPACT cycle.

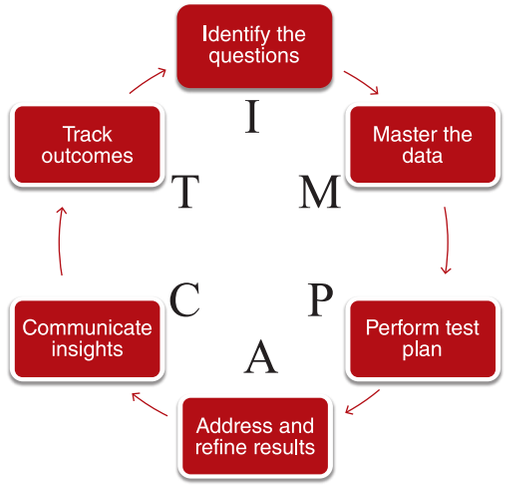
THE DATA ANALYTICS PROCESS USING THE IMPACT CYCLE

Data Analytics is a process to identify business questions and problems that can be addressed with data. We start to describe our Data Analytics Process by using an established Data Analytics model called the IMPACT cycle by Isson and Harriott (as shown in Exhibit 1-1).

We explain the full IMPACT cycle briefly here, but in more detail in later in Chapters 2, 3, and 4. We use its approach throughout this textbook.

EXHIBIT 1-1
The IMPACT Cycle

Source: Isson, J. P., and J. S. Harriott. *Win with Advanced Business Analytics: Creating Business Value from Your Data.* Hoboken, NJ: Wiley, 2013.



Step 1: Identify the Questions (Chapter 1)

It all begins with understanding a business problem that needs addressing. Questions can arise from many sources, including how to better attract customers, how to price a product, or how to find errors or fraud. Having a concrete, specific question that is potentially answerable by Data Analytics is an important first step.

Accountants and auditors might be interested in questions like the following:

- Are employees circumventing internal controls over payments?
- Are there any suspicious travel and entertainment expenses?
- How can we increase the amount of add-on sales of additional goods to our customers?
- Are our customers paying us in a timely manner?
- How can we predict the allowance for loan losses for our bank loans?
- How can we find transactions that are risky in terms of accounting issues?
- Who authorizes checks above $100,000?
- How can errors be identified?

Step 2: Master the Data (Chapter 2)

Mastering the data requires one to know what data are available and whether those data might be able to help address the business problem. We need to know everything about the data, including how to access, availability, reliability (if there are errors), and what time periods are covered to make sure the data coincide with the timing of our business problem, etc.

In addition, to give us some idea of the data questions, we may want to consider the following:

- Review data availability in a firm's internal systems (including those in the financial reporting system or enterprise systems that might occur in its accounting processes—financial, procure-to-pay, production, order-to-cash, human resources).
- Review data availability in a firm's external network, including those that might already be housed in an existing data warehouse.
- Data dictionaries and other contextual data—to provide details about the data.
- Extraction, transformation, and loading.
- Data validation and completeness—to provide a sense of the reliability of the data.
- Data normalization—to reduce data redundancy and improve data integrity.
- Data preparation and scrubbing—Data Analytics professionals estimate that they spend between 50 and 90 percent of their time cleaning data so the data can be analyzed.[11]

Step 3: Perform Test Plan (Chapter 3)

After mastering the data and after the data are ready (in step 2), we are prepared for analysis. With the data ready for analysis, we need to think of the right approach to the data to be able to answer the question.

In Data Analytics, we work to extract knowledge from the data to address questions and problems. Using all available data, we see if we can identify a relationship between the **response or dependent variables** and those items that affect the response (also called **predictor**, **explanatory**, or **independent variables**). To do so, we'll generally make a model, or a simplified representation of reality, to address this purpose.

An example might be helpful here. Let's say we are trying to predict each of your classmates' performance on their next intermediate accounting exam. The response or dependent variable will be the score on the next exam. What helps predict the performance of each exam will be our predictor, explanatory, or independent variables. Variables such as study time, score on last exam, IQ, and standardized test scores (ACT, SAT, etc.), as well as student enjoyment of accounting, might all be considered. Perhaps given your experience you can name other predictor variables to include in our model predicting exam performance.

The research question, the model, the data availability, and the expected statistical inference may all suggest the use of different data approaches. Provost and Fawcett[12] detail eight different approaches to data analytics depending on the question. We will discuss the most applicable ones to accounting more formally in Chapter 3 and highlight accounting questions that they might address. The eight different approaches include the following:

- **Classification**—An attempt to assign each unit (or individual) in a population into a few categories. An example classification might be, of all the loans this bank has offered, which are most likely to default? Or which loan applications are expected to be approved? Or which transactions would a credit card company flag as potentially being fraudulent and deny payment?
- **Regression**—A data approach used to predict a specific dependent variable value based on independent variable inputs using a statistical model. An example regression analysis might be, given a balance of total accounts receivable held by a firm, what is the appropriate level of allowance for doubtful accounts for bad debts?

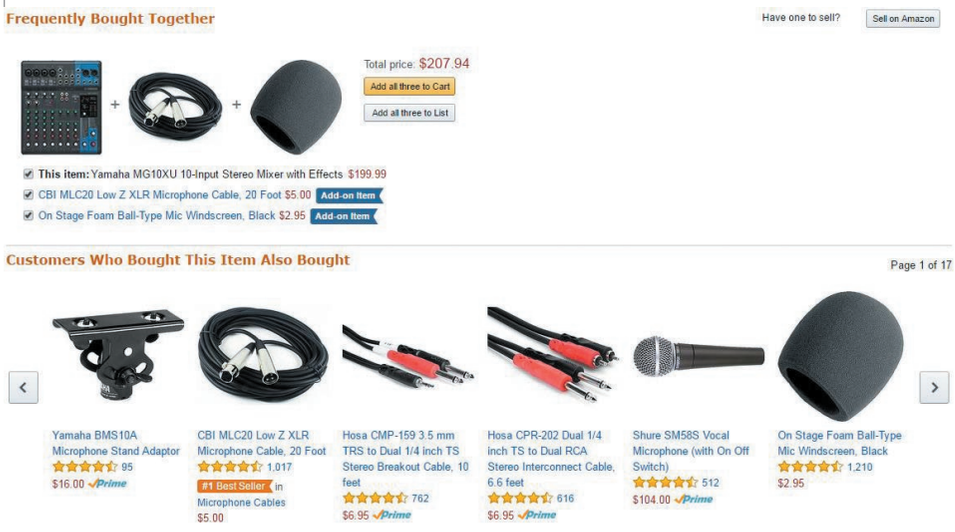
[11] "One-Third of BI Pros Spend Up to 90% of Time Cleaning Data," http://www.eweek.com/database/one-third-of-bi-pros-spend-up-to-90-of-time-cleaning-data.html, posted June 2015 (accessed March 15, 2016).

[12] Foster Provost and Tom Fawcett, *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking* (Sebastopol, CA: O'Reilly Media, 2013).

- **Similarity matching**—An attempt to identify similar individuals based on data known about them. The opening vignette mentioned **Alibaba** and its attempt to identify seller and customer fraud based on various characteristics known about them to see if they were similar to known fraud cases.
- **Clustering**—An attempt to divide individuals (like customers) into groups (or clusters) in a useful or meaningful way. In other words, identifying groups of similar data elements and the underlying drivers of those groups. For example, clustering might be used to segment a customer into a small number of groups for additional analysis and marketing activities.
- **Co-occurrence grouping**—An attempt to discover associations between individuals based on transactions involving them. **Amazon** might use this to sell another item to you by knowing what items are "frequently bought together" or "Customers who bought this item also bought . . ." as shown in Exhibit 1-2.

EXHIBIT 1-2
Example of Co-occurrence Grouping on Amazon .com

Source: Amazon Inc.



- **Profiling**—An attempt to characterize the "typical" behavior of an individual, group, or population by generating summary statistics about the data (including mean, standard deviations, etc.). By understanding the typical behavior, we'll be able to more easily identify abnormal behavior. When behavior departs from that typical behavior—which we'll call an anomaly—then further investigation is warranted. Profiling might be used in accounting to identify fraud or just those transactions that might warrant some additional investigation (e.g., travel expenses that are three standard deviations above the norm).
- **Link prediction**—An attempt to predict a relationship between two data items. This might be used in social media. For example, because an individual might have 22 mutual Facebook friends with me and we both attended Brigham Young University, is there a chance we would like to be Facebook friends as well? Exhibit 1-3 provides an example of this used in Facebook. Link prediction in an accounting setting might work to use social media to look for relationships between related parties that are not otherwise disclosed.
- **Data reduction** —A data approach that attempts to reduce the amount of information that needs to be considered to focus on the most critical items (i.e., highest cost,

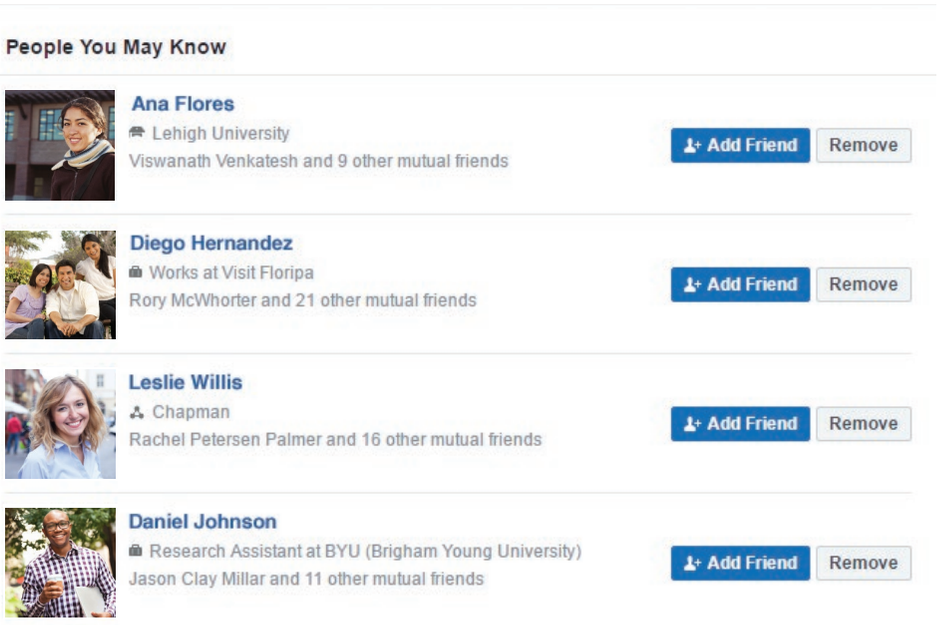


EXHIBIT 1-3
Example of Link Prediction on Facebook

Source: Facebook Inc.; Exactostock/SuperStock (Ana); Michael DeLeon/ Getty Images (Diego); Daniel Ernst/Getty Images (Leslie); Sam Edwards/ Glow Images (Daniel)

highest risk, largest impact, etc.). It does this by taking a large set of data (perhaps the population) and reducing it with a smaller set that has the vast majority of the critical information of the larger set. An example might include the potential to use these techniques in auditing. While auditing has employed various random and stratified sampling over the years, Data Analytics suggests new ways to highlight which transactions do not need the same level of vetting as other transactions.

Step 4: Address and Refine Results (Chapter 3)

After the data have been analyzed (in step 3 of the IMPACT cycle), the fourth step is to address and refine results. Data analysis is iterative. We slice and dice the data, find correlations, ask ourselves further questions, ask colleagues what they think, and revise and rerun the analysis. But once that is complete, we have the results ready to communicate to interested stakeholders.

Steps 5 and 6: Communicate Insights and Track Outcomes (Chapter 4 and each chapter thereafter)

Once the results have been determined (in step 4 of the IMPACT cycle), insights are formed by decision makers and are communicated (the "C" in the IMPACT cycle) and some outcomes will be continuously tracked (the "T" in the IMPACT cycle).

Chapter 4 discusses ways to communicate results, including the use of static reports, digital dashboards, and data visualizations. Data Analytics is especially interested in reporting results that help decision makers see the data in an all-new way to develop insights that help answer business questions. Digital dashboards and data visualizations are particularly helpful in communicating results.

Back to Step 1

Of course, the IMPACT cycle is iterative, so once insights are gained and outcomes are tracked, new questions emerge and the IMPACT cycle begins anew.

PROGRESS CHECK

7. Let's say we are trying to predict how much money college students spend on fast food each week. What would be the response, or dependent, variable? What would be examples of independent variables?
8. How might a data reduction approach be used in auditing to spend time and effort on the most important items?

LO 1-5

Describe the skills needed by accountants.

DATA ANALYTIC SKILLS NEEDED BY ANALYTIC-MINDED ACCOUNTANTS

While we don't believe that accountants need to become data scientists—they may never need to build a data repository or do the real, hardcore Data Analytics—they must know how to do the following:

- Clearly articulate the business problem the company is facing.
- Communicate with the data scientists about specific data needs and understand the underlying quality of the data.
- Draw appropriate conclusions to the business problem based on the data and make recommendations on a timely basis.
- Present their results to individual members of management (CEOs, audit managers, etc.) in an accessible manner to each member.

Consistent with that, in this text, we emphasize seven skills that analytic-minded accountants should have:

1. Develop an analytics mindset—recognize when and how data analytics can address business questions.
2. Data scrubbing and data preparation—comprehend the process needed to clean and prepare the data before analysis.
3. Data quality—recognize what is meant by data quality, be it completeness, reliability, or validity.
4. Descriptive data analysis—perform basic analysis to understand the quality of the underlying data and its ability to address the business question.
5. Data analysis through data manipulation—demonstrate ability to sort, rearrange, merge and reconfigure data in a manner that allows enhanced analysis.
6. Define and address problems through statistical data analysis—identify and implement an approach that will use statistical data analysis to draw conclusions and make recommendations on a timely basis.
7. Data visualization and data reporting—report results of analysis in an accessible way to each varied decision maker and his or her specific needs.

We address these seven skills throughout the first four chapters in the text in hopes that the analytic-minded accountant will develop and practice these skills to be ready to address business questions. We then demonstrate these skills in the labs and hands-on analysis throughout the rest of the book.

HANDS-ON EXAMPLE OF THE IMPACT MODEL

LO 1-6

Explain how to translate common business questions into fields and values.

Here we provide a complete, hands-on example of the IMPACT model to show how it could be implemented for a specific situation.

Let's suppose I am trying to get a loan to pay off some credit card debt and my friend has told me about a new source of funds that doesn't involve a bank. In recent years, facilitated by the Internet, peer-to-peer lenders allow individuals to both borrow and lend money to each other. While there are other peer-to-peer lenders, in this case, we will specifically consider the **LendingClub.**

My question is whether I will be able to get a loan, given my prior loan history (poor), credit score, and the like. According to our approaches mentioned above, this would be an example of a classification approach because we are attempting to predict whether a person applying for a loan will be approved and funded or whether she will be denied a loan.

Identify the Questions

Stated specifically, our question is, "Given my borrower profile, can I expect the **LendingClub** to extend a loan to me?"

Master the Data

LendingClub is a U.S.-based, peer-to-peer lending company, headquartered in San Francisco, California. **LendingClub** facilitates both borrowing and lending by providing a platform for unsecured personal loans between $1,000 and $35,000. The loan period is for either 3 or 5 years. There is information available that allows potential investors to search and browse the loan listings on the **LendingClub** website and select loans in which they would like to invest. The available information includes information supplied about the borrower, amount of the loan, loan grade (and related loan interest rate), and loan purpose. Investors invest in the loans and make money from interest. **LendingClub** makes money by charging borrowers an origination fee and investors a service fee. Since 2007, hundreds of thousands of borrowers have obtained more than $44 billion in loans via **LendingClub**.[13]

Some basic lending statistics are included on the LendingClub Statistics website (Exhibit 1-4). Each bar represents the volume of loans each quarter during its respective year.

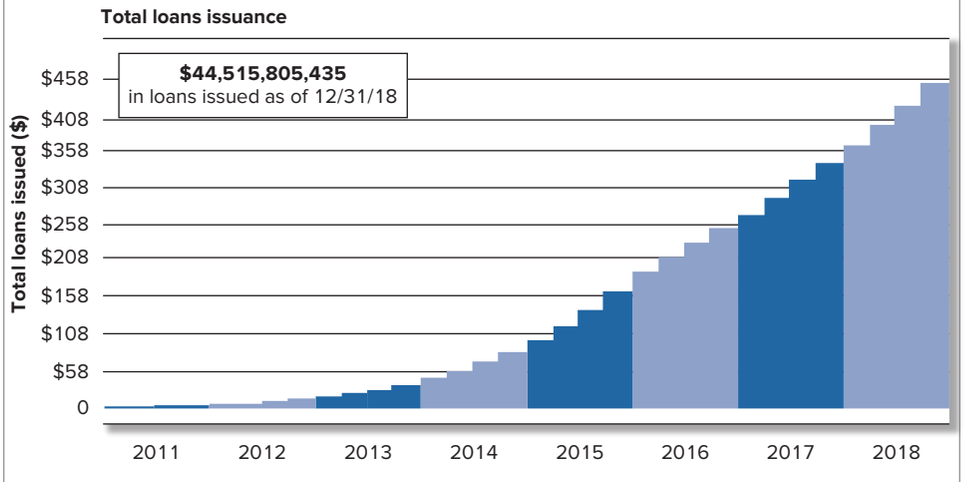


EXHIBIT 1-4
LendingClub Statistics

Source: Accessed March, 2019. https://www.lendingclub.com/info/statistics.action.

[13]https://www.lendingclub.com/ (accessed September 29, 2016).

Borrowers borrow money for a variety of reasons, including refinancing other debt and paying off credit cards, as well as borrowing for other purposes (Exhibit 1-5).

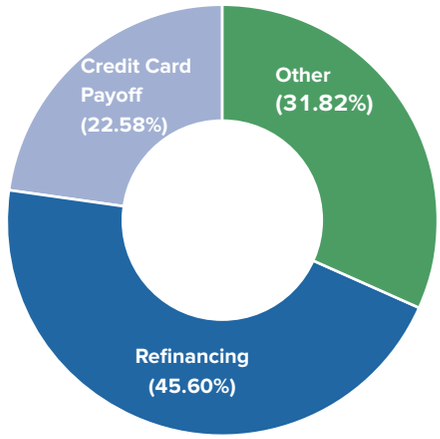


EXHIBIT 1-5
LendingClub Statistics by Reported Loan Purpose
68.18% of LendingClub borrowers report using their loans to refinance existing loans or pay off their credit cards as of 12/31/18.

Source: Accessed March, 2019. https://www.lendingclub.com/info/statistics.action.

LendingClub actually provides datasets: data on the loans they approved and funded as well as data for the loans that were declined. In this chapter, we will emphasize the rejected loans and the reasons why they were rejected.

The datasets and the data dictionary are available at https://www.lendingclub.com/info/download-data.action.

As we learn about the data, it is important to know what is available to us. To that end, there is a **data dictionary** that provides descriptions for all of the data attributes of the dataset. A cut-out of the data dictionary for the rejected stats file (i.e., the statistics about those loans rejected) is shown in Exhibit 1-6.

EXHIBIT 1-6
2007–2012 LendingClub Data Dictionary for Declined Loan Data

Source: Accessed March, 2019. Available at https://www.lendingclub.com/info/download-data.action.

RejectStats File	Description
Amount Requested	Total requested loan amount
Application Date	Date of borrower application
Loan Title	Loan title
Risk_Score	Borrower risk (FICO) score
Dept-To-Income Ratio	Ratio of borrower total monthly debt payments divided by monthly income.
Zip Code	The first 3 numbers of the borrower zip code provided from loan application.
State	Two digit State Abbreviation provided from loan application.
Employment Length	Employment length in years, where 0 is less than 1 and 10 is greater than 10.
Policy Code	policy_code=1 if publicly available. policy_code=2 if not publicly available

We could also take a look at the data files available for the funded loan data. However, for our analysis in the rest of this chapter, we use the Excel file "RejectStatsA Ready,"

which has rejected loan statistics from 2007 to 2012. It is a cleaned-up, transformed file ready for analysis. We'll learn more about data scrubbing in Chapter 2.

Exhibit 1-7 provides a cut-out of the 2007–2012 "Approved Loan" dataset provided.

EXHIBIT 1-7
2007–2012 Declined Loan Applications (RejectStatsA) Dataset

Source: Microsoft Excel, 2016.

Amount Requested	Application Date	Loan Title	Risk_Score
2175	12/19/2012	major_purchase	850
35000	8/13/2012	other	850
10000	9/19/2012	major_purchase	850
10000	11/9/2012	car	850
3000	11/27/2012	vacation	850
5000	5/20/2012	Lower Rate	850
20000	9/8/2012	Home loan	850
8000	10/22/2012	Loan is for new kitch	850
18500	7/19/2012	bussiness loan	850
10000	7/11/2012	car	850
25000	10/6/2010	debt_consolidation	849
1000	9/9/2012	Hospital expenses	849
35000	5/26/2012	small_business	849
6800	7/13/2012	Be my Own Boss	849
25000	8/10/2012	home_improvement	849
35000	12/2/2012	debt_consolidation	848
1500	11/13/2012	other	848

Perform Test Plan

Considering our question, "Will I receive a loan from **LendingClub**?" and the available data, we will do three analyses to predict whether we will receive a loan, including:

1. The debt-to-income ratios and number of rejected loans.
2. The length of employment and number of rejected loans.
3. The credit (or risk) score and number of rejected loans.

Because **LendingClub** collects this information, we believe it will give **LendingClub** an idea if the borrower will be able to pay back the loan and give us an idea if our loan will be approved or rejected.

The first analysis we perform considers the *debt-to-income ratio.* That is, how big is the debt compared to the size of the annual income earned?

To consider the debt-to-income ratio in our analysis, three buckets (labeled DTI bucket) are constructed for each grouping of the debt-to-income ratio. These three buckets include the following:

1. High (debt is greater than 20 percent of income).
2. Medium ("mid") (debt is between 10 and 20 percent of income).
3. Low (debt is less than 10 percent of income).

Once those buckets are constructed, we are ready to analyze the breakdown of rejected loan applications by the debt-to-income ratio.

The Excel PivotTable is an easy way to make comparisons between the different levels of DTI. When we run a PivotTable analysis, we highlight the loans, which count the number of loans applied for and rejected, and the DTI bucket (see Exhibit 1-8). The PivotTable counts the number of loan applications in each of the three DTI buckets: high, medium (mid), and low. This suggests that because the high DTI bucket has the highest number of

EXHIBIT 1-8
LendingClub Declined Loan Applications by DTI (Debt-to-Income)
DTI bucket includes high (debt > 20 percent of income), medium ("mid") (debt between 10 and 20 percent of income), and low (debt < 10 percent of income). (PivotTable shown here required manually sorting rows to get in proper order.)

Source: Microsoft Excel, 2016.

Row Labels	Count of Rejected Loans
High	312,986
Low	171,228
Mid	161,200
Grand Total	**645,414**

PivotTable Fields

Choose fields to add to report:

- ☑ **Rejected Loans**
- ☐ Amount Requested
- ☐ Application Date
- ☐ Loan Title
- ☐ Risk_Score
- ☐ Risk Score Bucket
- ☐ Debt-To-Income Ratio
- ☑ **DTI Bucket**
- ☐ Zip Code
- ☐ State
- ☐ Employment Length
- ☐ EmploymentNum

loan applications, perhaps the applicant asked for a loan that was too big given his or her income. **LendingClub** might have seen that as too big of a risk and chosen to not extend the loan to the borrower using the debt-to-income ratio as an indicator.

The second analysis was on the length of employment and its relationship with rejected loans (see Exhibit 1-9). Arguably, the longer the employment, the more stable of a job and income stream you will have to ultimately repay the loan. **LendingClub** reports the number of years for each of the rejected applications. The PivotTable analysis lists the number of loans by the length of employment. Almost 77 percent (495,109 out of 645,414) out of the total rejected loans had worked at a job for less than 1 year, suggesting potentially an important reason for rejecting the requested loan. Perhaps some had worked a week, or just a month, and still want a big loan?

EXHIBIT 1-9
LendingClub Declined Loan Applications by Employment Length (Years of Experience)
DTI bucket includes high (debt > 20 percent of income), medium ("mid") (debt between 10 and 20 percent of income), and low (debt < 10 percent of income). (PivotTable shown here required manually sorting rows to get in proper order.)

Source: Microsoft Excel, 2016.

Row Labels	Count of Rejected Loans
< 1 year	495,109
1 year	20,732
2 years	21,987
3 years	17,487
4 years	13,848
5 years	12,865
6 years	9,829
7 years	7,221
8 years	6,652
9 years	5,083
10+ years	34,601
Grand Total	**645,414**

PivotTable Fields

Choose fields to add to report:

- ☑ **Rejected Loans**
- ☐ Amount Requested
- ☐ Application Date
- ☐ Loan Title
- ☐ Risk_Score
- ☐ Risk Score Bucket
- ☐ Debt-To-Income Ratio
- ☐ DTI Bucket
- ☐ Zip Code
- ☐ State
- ☑ **Employment Length**
- ☐ EmploymentNum

The third analysis we perform is to consider the credit or risk score of the applicant. As noted in Exhibit 1-10, risk scores are typically classified in this way with those in the excellent and very good category receiving the lowest possible interest rates and best terms with a credit score above 750. On the other end of the spectrum are those with very bad credit (with a credit score less than 600).

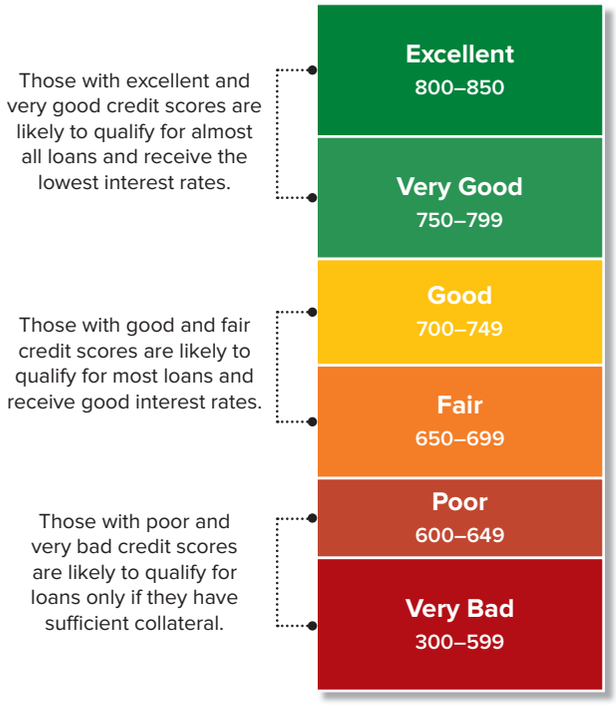


EXHIBIT 1-10
Breakdown of Customer Credit Scores (or Risk Scores)

Source: Cafecredit.com.

Another predictor of loan repayment is the credit score that the borrower has. We classify the sample according to this breakdown into excellent, very good, good, fair, poor, and very bad credit according to their credit score noted in Exhibit 1-10.

Address and Refine Results

After performing a PivotTable analysis (as seen in Exhibit 1-11), we count the number of rejected loan applications by credit (risk) score. We'll note in the rejected loans that nearly 82 percent [(167,379 + 151,716 + 207,234)/645,414] of the applicants have either very bad, poor, or fair credit ratings, suggesting this might be a good reason for a loan rejection. We also note that only 0.3 percent (2,494/645,414) of those rejected loan applications had excellent credit.

So, if these are the applications that were all rejected, the question is how many of these that might apply for a loan not only had excellent credit, but also had worked more than 10 years and had asked for a loan that was less than 10 percent of their income (in the low DTI bucket)? Use of a PivotTable (as shown in Exhibit 1-12) allows us to consider this three-way interaction and provides an answer of 365 out of 645,414 (0.057 percent of the total). This might suggest that the use of these three metrics is reasonable at predicting loan rejection because the number who have excellent credit, worked more than 10 years, and requested a loan that was less than 10 percent of their income was such a small percentage of the total.

Row Labels	Count of Rejected Loans
Excellent	2,494
Very Good	20,036
Good	96,555
Fair	207,234
Poor	151,716
Very Bad	167,379
Grand Total	**645,414**

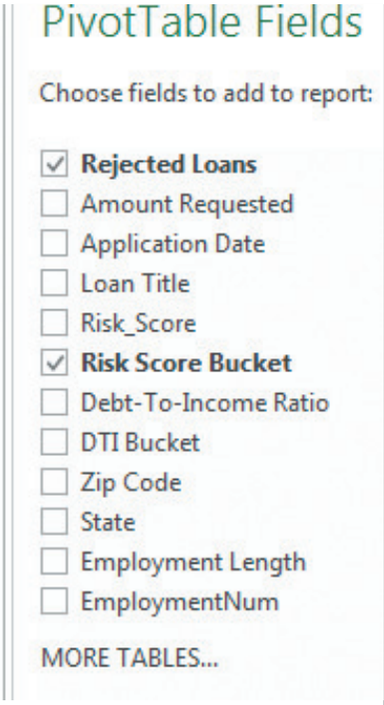


EXHIBIT 1-11 **The Count of LendingClub Rejected Loan Applications by Credit or Risk Score Classification Using PivotTable Analysis** (PivotTable shown here required manually sorting rows to get in proper order.)

Source: Microsoft Excel, 2016.

Row Labels	Count of Rejected Loans
Excellent	**2,494**
High	**762**
< 1 year	543
1 year	14
10+ years	89
2 years	26
3 years	16
4 years	16
5 years	14
6 years	20
7 years	10
8 years	7
9 years	7
Low	**1339**
< 1 year	457
1 year	38
10+ years	365
2 years	68

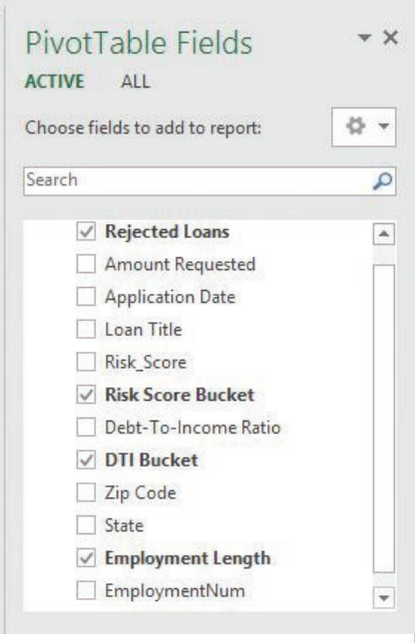
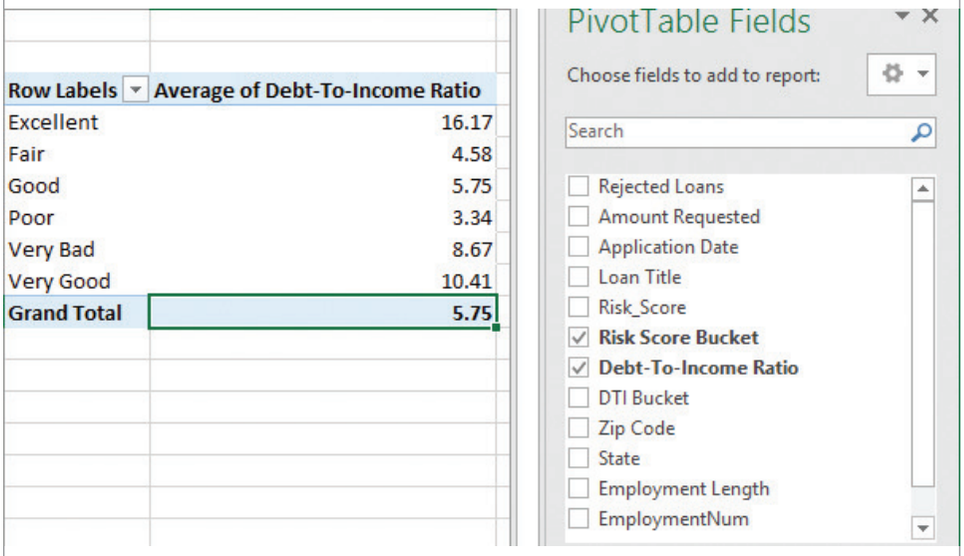


EXHIBIT 1-12 **The Count of LendingClub Declined Loan Applications by Credit Score, Debt-to-Income, and Employment Length Using PivotTable Analysis (highlighting added)**

Source: Microsoft Excel, 2016.

Perhaps those with excellent credit just asked for too big of a loan given their existing debt and that is why they were rejected. Exhibit 1-13 shows the PivotTable analysis. The analysis shows those with excellent credit asked for a larger loan (16.2 percent of income) given the debt they already had as compared to any of the others, suggesting a reason even those potential borrowers with excellent credit were rejected.



Row Labels	Average of Debt-To-Income Ratio
Excellent	16.17
Fair	4.58
Good	5.75
Poor	3.34
Very Bad	8.67
Very Good	10.41
Grand Total	5.75

EXHIBIT 1-13
The Average Debt-to-Income Ratio (shown as a percentage) by Credit (Risk) Score for LendingClub Declined Loan Applications Using PivotTable Analysis

Source: Microsoft Excel, 2016.

Communicate Insights

Certainly further and more sophisticated analysis could be performed, but at this point we have a pretty good idea of what **LendingClub** uses to decide whether to extend a loan. We can communicate these insights either by showing the PivotTables or stating what three of the determinants are.

Track Outcomes

There are a wide variety of outcomes that could be tracked. But in this case, it might be best to see if we could predict future outcomes. For example, the data we analyzed was from 2007–2012. We could make our predictions for subsequent years based on what we had found in the past and then test and see how accurate we are with those predictions. We could also change our prediction model when we learn new insights and additional data become available.

In this chapter, we discussed how businesses and accountants derive value from Data Analytics. We gave some specific examples of how Data Analytics is used in business, auditing, managerial accounting, financial accounting, and tax accounting.

We introduced the IMPACT model and explained how it is used. And then we talked specifically about the importance of identifying the question. We walked through the first few steps of the IMPACT model and introduced eight data approaches. We also discussed the data analytic skills needed by analytic-minded accountants.

We followed this up by looking at the case of why **LendingClub** rejected loans for a set of its customers using the IMPACT model. We performed this analysis using various filtering and PivotTable tasks.

PROGRESS CHECK

9. Doing your own analysis, download the rejected loans dataset titled "RejectStatsA Ready" and perform an Excel PivotTable analysis by state and figure out the number of rejected applications for the state of California. That is, count the loans by state and see what percentage of the rejected loans came from California. How close is that to the relative proportion of the population of California as compared to that of the United States?
10. Doing your own analysis, download the rejected loans dataset titled "RejectStatsA Ready" and run an Excel PivotTable by risk (or credit) score classification and DTI bucket to determine the number of rejected loans requested by those rated as having an excellent credit score.

Summary

- With data all around us, businesses and accountants are looking at Data Analytics to extract the value that the data might possess.
- Data Analytics is changing the audit and the way that accountants look for risk. Now, auditors can consider 100 percent of the transactions in their audit testing. It is also helpful in finding anomalous or unusual transactions. Data Analytics is also changing the way financial accounting, managerial accounting, and taxes are done at a company.
- The IMPACT cycle is a means of doing Data Analytics that goes all the way from identifying the question, to mastering the data, to performing data analyses and communicating results. It is recursive in nature, suggesting that as questions are addressed, new important questions may emerge that can be addressed in a similar way.
- Eight data approaches address different ways of testing the data: classification, regression, similarity matching, clustering, co-occurrence grouping, profiling, link prediction, and data reduction. These are explained in more detail in Chapter 3.
- Data analytic skills needed by analytic-minded accountants are specified and are consistent with the IMPACT cycle, including the following:
  - Develop an analytics mindset.
  - Data scrubbing and data preparation.
  - Data quality.
  - Descriptive data analysis.
  - Data analysis through data manipulation.
  - Define and address problems through statistical data analysis.
  - Data visualization and data reporting.

Key Words

Big Data (***4***) Datasets that are too large and complex for businesses' existing systems to handle utilizing their traditional capabilities to capture, store, manage, and analyze these datasets.

classification (***9***) A data approach that attempts to assign each unit in a population into a few categories potentially to help with predictions.

clustering (***10***) A data approach that attempts to divide individuals (like customers) into groups (or clusters) in a useful or meaningful way.

co-occurrence grouping (***10***) A data approach that attempts to discover associations between individuals based on transactions involving them.

Data Analytics (***4***) The process of evaluating data with the purpose of drawing conclusions to address business questions. Indeed, effective Data Analytics provides a way to search through large structured and unstructured data to identify unknown patterns or relationships.

data dictionary (***14***) Centralized repository of descriptions for all of the data attributes of the dataset.

data reduction (***10***) A data approach that attempts to reduce the amount of information that needs to be considered to focus on the most critical items (i.e., highest cost, highest risk, largest impact, etc.)

link prediction (***10***) A data approach that attempts to predict a relationship between two data items.

profiling (***10***) A data approach that attempts to characterize the "typical" behavior of an individual, group or population by generating summary statistics about the data (including mean, standard deviations, etc.).

predictor (or independent or explanatory) variable (***9***) A variable that predicts or explains another variable.

response (or dependent) variable (***9***) A variable that responds to, or is dependent, on another.

regression (***9***) A data approach used to predict a specific dependent variable value based on independent variable inputs using a statistical model.

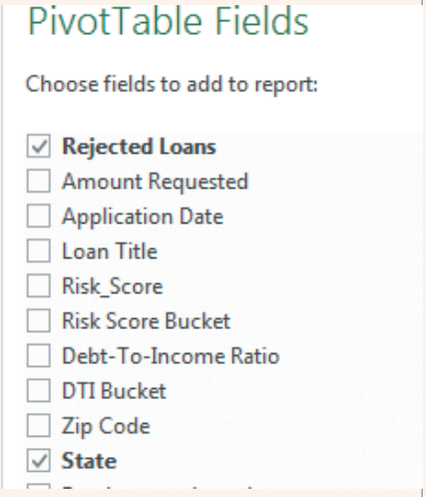
similarity matching (***10***) A data approach that attempts to identify similar individuals based on data known about them.

ANSWERS TO PROGRESS CHECKS

1. The plethora of data alone does not necessarily translate into value. However, if we carefully use the data to help address critical business problems and questions, the data may create value.
2. Banks could certainly use credit scores from companies like **Experian**, **TransUnion**, and **Equifax**, but if they have access to all of the banking information of their clients, arguably they could make more informed decisions. Banks would know how much money they have and how they spend it. Banks would know if they had prior loans and if they were paid in a timely manner. Banks would know where they work and their monthly income via the direct deposits. All of these combined, in addition to a credit score, might be used to assess creditworthiness to gain a better evaluation of customers' creditworthiness when they would like a loan. It might also give us needed information for a marketing campaign to target potential creditworthy customers.
3. The brand manager at **Samsung** might use Data Analytics to see what is being said about **Samsung's** phones on social media websites (e.g., **Snapchat**, **Instagram**, and **Facebook**), particularly those that attract an older demographic. This will help the manager assess if there is a problem with the perceptions of its phones.
4. Data Analytics might be used to collect information on the amount of overtime. Who worked overtime? What were they working on? Do we actually need more full-time employees to reduce the level of overtime (and its related costs to the company and to the employees)? All of these questions could be addressed by looking at recent records explaining the use of overtime records.
5. Data Analytics could tabulate the number of journal entries by an accountant to see who entered the most journal entries. This might be an issue if there was a perception of a problem in risk, such as segregation of duties in having one person enter so many journal entries or just how the accounting workload is distributed across accounting staff.

6. The tax staff would become much more adept at efficiently organizing data from multiple systems across an organization and performing Data Analytics to help with tax planning to structure transactions in a way that might minimize taxes.
7. The dependent variable could be the amount of money spent on fast food. Independent variables could be proximity of the fast food, ability to cook own food, discretionary income, socioeconomic status, etc.
8. The data reduction approach might help auditors spend more time and effort on the riskiest transactions or on those that might be anomalous in nature. This will help them more efficiently spend their time on items that may well be of highest importance.
9. An analysis of the rejected loans suggests that 85,793 of the total 645,414 rejected loans were from the state of California. That represents 13.29 percent of the total rejected loans. This is greater than the relative population of California to the United States as of the 2010 census, of 12.1 percent (37,253,956/308,745,538).

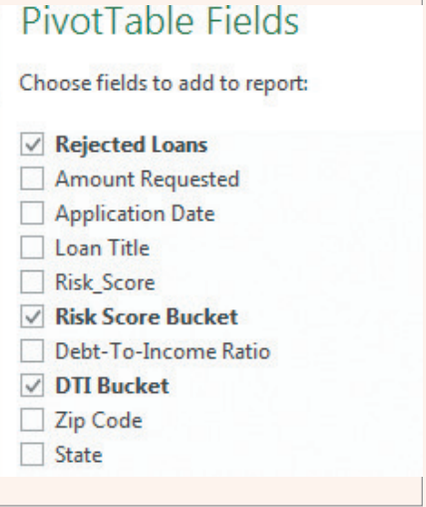
Row Labels	Count of Rejected Loans
AK	1,608
AL	10,959
AR	7,868
AZ	13,830
CA	85,793
CO	11,697
CT	10,589
DC	1,524
DE	2,287
FL	49,620
GA	23,774
HI	3,715



Source: Microsoft Excel, 2016.

10. A PivotTable analysis of the rejected loans suggests that more than 30.6 percent (762/2,494) of those in the excellent risk/credit score range asked for a loan with a debt-to-income ratio of more than 20 percent.

Row Labels	Count of Rejected Loans
Excellent	**2,494**
High	762
Mid	393
Low	1,339
Very Good	**20,036**
High	9,593
Mid	4,022
Low	6,421
Good	**96,555**
High	59,241
Mid	18,291
Low	19,023



Source: Microsoft Excel, 2016.

connect

Multiple Choice Questions

1. Big Data is often described by the three Vs, or
   a. volume, velocity, and variability.
   b. volume, velocity, and variety.
   c. volume, volatility, and variability.
   d. variability, velocity, and variety.
2. Which approach to Data Analytics attempts to assign each unit in a population into a small set of classes (or groups) where the unit best fits?
   a. Regression
   b. Similarity matching
   c. Co-occurrence grouping
   d. Classification
3. Which approach to Data Analytics attempts to identify similar individuals based on data known about them?
   a. Classification
   b. Regression
   c. Similarity matching
   d. Data reduction
4. Which approach to Data Analytics attempts to predict relationship between two data items?
   a. Profiling
   b. Classification
   c. Link prediction
   d. Regression
5. Which of these terms is defined as being a central repository of descriptions for all of the data attributes of the dataset?
   a. Big Data
   b. Data warehouse
   c. Data dictionary
   d. Data Analytics
6. Which skills were *not* emphasized that analytic-minded accountants should have?
   a. Develop an analytics mindset
   b. Data scrubbing and data preparation
   c. Classification of test approaches
   d. Define and address problems through statistical data analysis
7. Which skills were *not* emphasized that analytic-minded accountants should have?
   a. Data quality
   b. Descriptive data analysis
   c. Data visualization
   d. Data and systems analysis and design
8. The IMPACT cycle includes all *except* the following process:
   a. perform test plan.
   b. visualize the data.
   c. master the data.
   d. track outcomes.

9. The IMPACT cycle includes all *except* the following process:
   a. data preparation.
   b. communicate insights.
   c. address and refine results.
   d. perform test plan.
10. By the year 2020, about 1.7 megabytes of new information will be created every:
   a. week.
   b. second.
   c. minute.
   d. day.

Discussion Questions

1. Define Data Analytics and explain how a university might use its techniques to recruit and attract potential students.
2. Give an example of how Data Analytics creates value for businesses.
3. Give an example of how Data Analytics creates value for accounting.
4. How might Data Analytics be used in financial reporting? And how might it be used in doing tax planning?
5. Describe the IMPACT cycle. Why does its order of the processes and its recursive nature make sense?
6. Why is identifying the question such a critical first step in the IMPACT process cycle?
7. What is included in mastering the data as part of the IMPACT cycle described in the chapter?
8. In the chapter, we mentioned eight different data approaches. Which data approach was used by **Alibaba,** as mentioned in the chapter-opening vignette?
9. What data approach mentioned in the chapter might be used by Facebook to find friends?
10. Auditors will frequently use the data reduction approach when considering potentially risky transactions. Provide an example of why focusing on a portion of the total number of transactions might be important for auditors to assess risk.
11. Which data approach might be used to assess the appropriate level of the allowance for doubtful accounts?
12. Why might the debt-to-income attribute included in the declined loans dataset considered in the chapter be a predictor of declined loans? How about the credit (risk) score?
13. To address the question "Will I receive a loan from **LendingClub**?" we had available data to assess the relationship among (1) the debt-to-income ratios and number of rejected loans, (2) the length of employment and number of rejected loans, and (3) the credit (or risk) score and number of rejected loans. What additional data would you recommend to further assess whether a loan would be offered? Why would it be helpful?

Problems

1. Navigate to the Additional Student Resources page on Connect. Under Chapter 1 Data Files, download and consider the **LendingClub** data dictionary file "LCDataDictionary" specifically the LoanStats tab. This represents the data dictionary for the loans that were funded. Choosing some of the data attributes listed there, which attributes do you think might predict which loans will go delinquent and which will ultimately be fully repaid? How could we test that?

2. Download and consider the rejected loans dataset of **LendingClub** data titled "RejectStatsA Ready." Given the analysis performed in the chapter, what three items do you believe would be most useful in predicting loan acceptance or rejection? What additional data do you think could be solicited either internally or externally that would help you predict loan acceptance or rejection?
3. Download the rejected loans dataset of **LendingClub** data titled "RejectStatsA Ready" from the Connect website and do an Excel PivotTable by state; then figure out the number of rejected applications for the state of Arkansas. That is, count the loans by state and compute the percentage of the total rejected loans in the USA that came from Arkansas. How close is that to the relative proportion of the population of Arkansas as compared to the overall U.S. population (per 2010 census)?
4. Download the rejected loans dataset of **LendingClub** data titled "RejectStatsA Ready" from the Connect website and do an Excel PivotTable by state; then figure out the number of rejected applications for each state. Reorder these and make a graph ordering the states and the number of rejected loans from highest to lowest. Is there a lot of variability among states?

   For Problems 5, 6, and 7, we will be cleaning a data file in preparation for subsequent analysis.

   The analysis performed on **LendingClub** data in the chapter was for the years 2007–2012. For this and subsequent problems, please download the declined loans table for 2013–2014 from the Connect website.
5. Consider the 2013 declined loan data from **LendingClub** titled "RejectStatsB2013" from the Connect website. Similar to the analysis done in the chapter, let's scrub the risk score data. First, because our analysis requires risk scores, debt-to-income data, and employment length, we need to make sure each of them has valid data.
   a. Open the file in Excel.
   b. Sort the file based on risk score and remove those observations (the complete row or record) that have a missing score or a score of zero, if needed.
   c. Assign each risk score to a risk score bucket similar to the chapter. That is, classify the sample according to this breakdown into excellent, very good, good, fair, poor, and very bad credit according to their credit score noted in Exhibit 1-10. Classify those with a score greater than 850 as "Excellent." Consider using nested if–then statements to complete this. Or sort by risk score and manually input into appropriate risk score buckets.
   d. Run a PivotTable analysis that shows the number of loans in each risk score bucket. Which group had the most rejected loans (biggest count)? Which group had the least rejected loans (smallest count)? This is the deliverable. Is it similar to Exhibit 1-11 performed on years 2007–2012?
6. Consider the 2013 declined loan data from **LendingClub** titled "RejectStatsB2013." Similar to the analysis done in the chapter, let's scrub the debt-to-income data. Because our analysis requires risk scores, debt-to-income data, and employment length, we need to make sure each of them has valid data.
   a. Sort the file based on debt-to-income and remove those observations (the complete row or record) that have a missing score, a score of zero, or a negative score.
   b. Assign each valid debt-to-income ratio into three buckets (labeled DTI bucket) by classifying each debt-to-income ratio into high (>20.0 percent), medium (10.0–20.0 percent), and low (<10.0 percent) buckets. Consider using nested if-then statements to complete this. Or sort the row and manually input.
   c. Run a PivotTable analysis that shows the number of loans in each DTI bucket. Any interpretation of why these loans were declined based on debt-to-income ratios?
7. Consider the 2013 declined loan data from **LendingClub** titled "RejectStatsB2013." Similar to the analysis done in the chapter, let's scrub the employment length. Because our analysis requires risk scores, debt-to-income data, and employment length, we need to make sure each of them has valid data.

a. Sort the file based on employment length and remove those observations (the complete row or record) that have a missing score ("NA"). Note that we are including the employment lengths of zero, different than the analysis in the chapter text.

b. Sort the file based on debt-to-income and remove those observations (the complete row or record) that have a missing score, a score of zero, or a negative score, similar to that done in Problem 1-6.

c. Sort the file based on risk score and remove those observations (the complete row or record) that have a missing score or a score of zero, similar to that done in Problem 1-5.

d. There should now be 669,993 observations. Any thoughts on what biases are imposed when we remove observations? Is there another way to do this?

e. Run a PivotTable analysis to show the number of Excellent Risk Scores but High DTI Bucket loans in each Employment year bucket. Any interpretation of why these loans were declined?

Lab 1-0 How to Complete Labs in This Text

The labs in this book will provide valuable hands-on experience in generating and analyzing accounting problems. Each lab will provide a company summary with relevant facts, techniques that you will use to complete your analysis, software that you'll need, and an overview of the lab steps.

When you've completed your lab, you will submit a lab report showing your thought process with written responses and validating that you've completed specific checkpoints by taking screenshots along the way. This lab will demonstrate how to use basic lab tools.

In this lab, you will:

Part 1: Create a Word document on OneDrive.

Part 2: Take a screenshot of your document.

Part 3: Add another screenshot and submit your document.

Submit two screenshots.

Part 1: Create a New Word Document on OneDrive

On Office.com

1. Open your web browser and go to www.office.com.
2. Click All Microsoft and then click OneDrive and log in using your university or personal email address and password.
3. Click + **New** > **OneDrive Word document.** A new window will open with a new blank document.
4. Type "Lab 1-0 Data Analytics Lab Overview [Your name] [Your university email address]" in the first line (e.g., *Lab 1-0 Data Analytics Lab Overview Ryan Teeter rteeter@pitt.edu*).
5. Click **File**> **Save As** > **Save As** and name the document "*Lab 1-0 Data Analytics Lab Overview Ryan Teeter rteeter@pitt.edu."* (You can also click the document name in the title bar (e.g., Document2) and change it there.
6. Because your document is in the cloud, changes are saved automatically and you won't lose your document when you log out of a lab computer.
7. Keep your document open and go to the next part of the lab.

Part 2: Take a Screenshot of Your Document

In Windows

1. Click the **Start** button and Search for "Snipping Tool."
2. Click **New** (**Rectangular Snip**) and draw a rectangle across your screen that includes your entire window.
3. A preview window with your screenshot will appear.
4. Press **Ctrl** + **C** to copy your screenshot.
5. Go to your Word document and press **Ctrl** + **V** to paste the screenshot into your document.
6. Keep your document open and go the next part of the lab.

On a Mac

1. Press **Cmd** + **Shift** + **4** and draw a rectangle across your screen that includes your entire window.
2. Your screenshot will be saved in your Desktop folder.
3. Drag the screenshot file into your Word document.
4. Keep your document open and go the next part of the lab.

Part 3: Add Another Screenshot and Submit Your Document

1. Open a new web browser window and go to mhhm.com.
2. Take a screenshot of your results (label it 1-0A) of the page and paste it into your lab document.
3. Save your document and submit it to your instructor. To download your document for OneDrive, click **File > Save As > Download a Copy.**

End of Lab

Lab 1-1 Data Analytics in Financial Accounting

Let's see how we might perform some simple Data Analytics. The purpose of this lab is to help you identify relevant questions that may be answered using Data Analytics.

Company summary

You were just hired as an analyst for a credit rating agency that evaluates publicly listed companies in the United States. The company already has some Data Analytics tools that it uses to evaluate financial statements and determine which companies have higher risk and which companies are growing quickly. The company uses these analytics to provide ratings that will allow lenders to set interest rates and determine whether to lend money in the first place. As a new analyst, you're determined to make a good first impression.

Technique

- Some experience with spreadsheets and basic formulas is helpful here.

Software needed

- Word processor
- Web browser
- Screen capture tool (Windows: Snipping Tool; Mac: Cmd + Shift + 4)

In this lab, you will:

Part 1: Identify appropriate questions, and develop a hypothesis for each question.

Part 2: Master the data.

Part 3: Perform a simple analysis.

Part 1: Identify the Questions

Think about ways that you might analyze data from a financial statement. You could use a horizontal analysis to view trends over time, a vertical analysis to show account proportions, or ratios to analyze relationships.

1. Create a new word processing document and name the file "**Lab 1-1 Data Analytics in Financial Accounting Lab–[Your name] [Your email address].**"
2. Use what you know about financial statement analysis (or search the web if you need a refresher) to generate three different metrics for evaluating financial performance.

For example, if you wanted to evaluate a company's profit margin from one year to the next your question might be, "Has [Company X's] gross margin increased in the last three years?" **Type your three questions in your document.**

3. Next to each question **generate a hypothetical answer to the question** to help you identify what your expected output would be. You may use some insight or intuition or search for industry averages to inform your hypothesis. For example: "Hypothesis: **Apple Inc**'s gross margin has increased slightly in the past 3 years."
4. Save your document.

Part 2: Master the Data

To answer your questions, you'll need to evaluate specific account values or financial statement paragraphs. As an analyst, you have access to the Security and Exchange Commission's (SEC's) EDGAR database of XBRL financial statements as well as a list of XBRL tags from the Financial Accounting Standards Board (FASB). XBRL stands for eXtensible Business Reporting Language and is used to make the data in financial statements machine-readable. Public companies have been preparing XBRL reports since 2008. While there are some issues with XBRL data, such data have become a useful means for comparing and analyzing financial statements. Every value, date, and paragraph is "tagged" with a label that identifies what each specific value represents, similar to assigning attributes in a database. Because companies tag their financial statements with XBRL tags, you can use those tags to identify specific data that you need to answer your questions from Part 1.

Analyze your questions:

5. **Evaluate each question from Part 1.** There are specific data attributes that will help you find the answer you're looking for. For example, if your question was "Has [Company X's] gross margin increased in the last 3 years?" and the expected answer is "**Apple Inc**'s gross margin has increased slightly in the past 3 years," this tells you what attributes (or fields) to look for: *company name, gross margin* (*sales revenues – cost of goods sold*), *year.*
6. For each of your questions, **identify the account or data attribute you need** to answer your question. Then use FASB's XBRL taxonomy (see next section for instructions) to identify the specific XBRL tags that represent those accounts. For example:

   Company name = EntitySectorIndustryClassificationPrimary

   Gross margin = GrossProfit

   Sales revenues = SalesRevenueNet

   Cost of goods sold = CostOfGoodsAndServicessold

   Year = DocumentPeriodEndDate
7. Save your document.

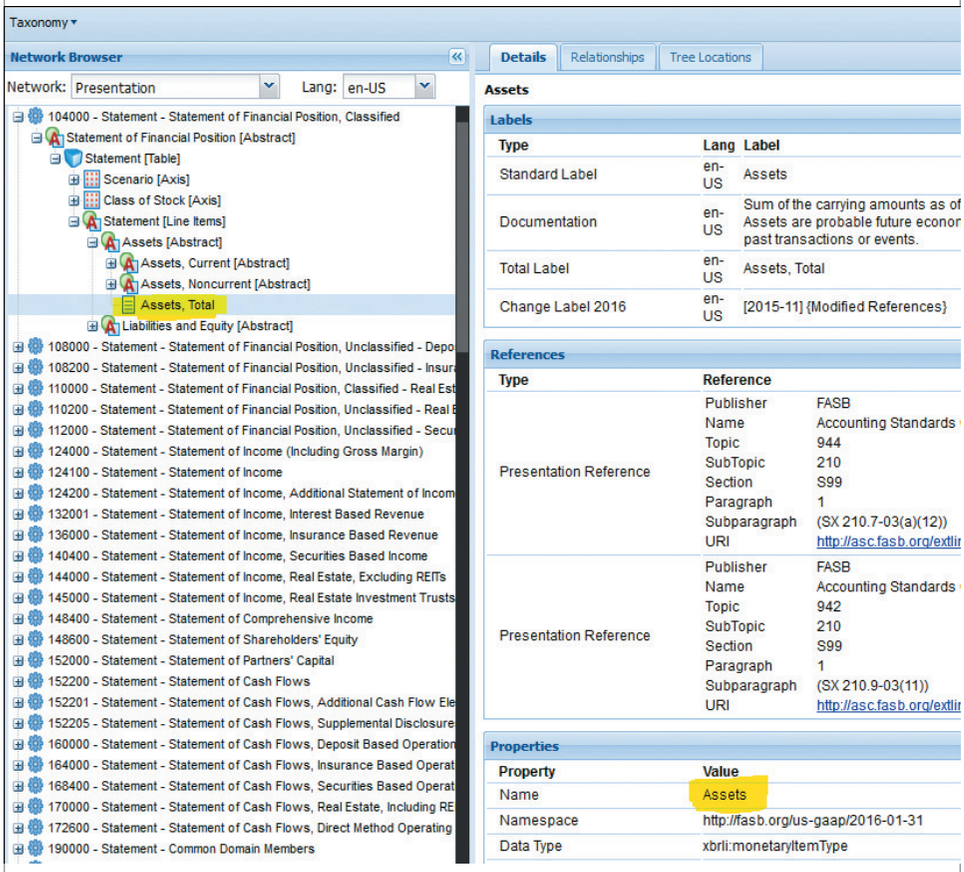
Identify XBRL tags from the FASB's taxonomy:

8. Open a web browser, and go to xbrlview.fasb.org.
9. Click the + next to US GAAP (2019-01-31).
10. Click the **ALL (Main/Entire)** option, and then click **Open** to load the taxonomy.
11. Navigate through the financial statements to determine which accounts you need to answer your questions from Part 1. The name of the XBRL tag is found in the properties pane next to "Name." For example, the tag for Total Assets can be found by clicking + Statement of Financial Position [Abstract], + Statement [Table], + Statement [Line Items], + Assets [Abstract], + Assets, Total, as shown in Lab Exhibit 1-1A. You may also use the search function.

Note: Be careful when you use the search function. The tag you see in the results may appear in the wrong statement. Double-click the tag to expand the tree and show where the account appears.

LAB EXHIBIT 1-1A
Browse the XBRL Taxonomy for Financial Fact Tags Needed for Your Analysis

Source: Google.



Part 3: Perform the Analysis

Now that you've identified your questions and the data sources, you can build your model and perform your analysis. Because XBRL data are dynamic, we'll use a tool that pulls live data based on your inputs.

12. In your web browser, click on the eBook via Connect to locate the table of contents where you will find **Additional Student Resources > Financial Statement Analysis**.
13. Log into your Google Account.
14. Click **File > Make a Copy. . .**
15. In your new document, **add your tags** from Part 2 under the Financial Facts header, similar to Exhibit 1-1B.

Financial Statement Analysis

Company 1 Ticker	AAPL	(e.g. MSFT)	EntitySectorIndustryClassificationPrimary
Most Recent Year	2017	(e.g. 2014)	DocumentPeriodEndDate
Period	FY	(e.g. FY or Q2)	
Round to	100,000	(e.g. 100,000)	

Apple Inc

Financial Facts	2014	2015	2016	2017
SalesRevenueNet	$ 1,827,950	$ 2,337,150	$ 2,156,390	$ 2,292,340
CostofGoodsAndServicesSold	$ 1,122,580	$ 1,400,890	$ 1,313,760	$ 1,410,480
GrossProfit	$ 705,370	$ 936,260	$ 842,630	$ 881,860
[Add XBRL Tag Here]	#ERROR!	#ERROR!	#ERROR!	#ERROR!
[Add XBRL Tag Here]	#ERROR!	#ERROR!	#ERROR!	#ERROR!
Analysis				
Return on Asset		74.20%	62.08%	64.74%
Profit Margin		40.06%	39.08%	38.47%
Asset Turnover		1.85	1.59	1.68

LAB EXHIBIT 1-1B
Add Your Tags to Perform a Simple Analysis Using XBRL Data

Source: Google.

16. Under the Analysis header, **use formulas to create your analysis** from Part 1. You may enhance your output by using conditional formatting or other visualizations that will be covered in Chapter 4.
17. Take a screenshot (label it 1-1A) of your analysis and paste it into your lab document.
18. Save your document and submit it to your instructor.

End of Lab

Lab 1-2 Data Analytics in Managerial Accounting

Let's see how we might use customer data to understand some simple data analytics. The purpose of this lab is to help you identify relevant questions that may be answered using data analytics.

Company summary

LendingClub is a U.S.-based, peer-to-peer lending company, headquartered in San Francisco, California. **LendingClub** facilitates both borrowing and lending by providing a platform for unsecured personal loans between $1,000 and $35,000. The loan period is for either 3 or 5 years. You have been brought in to help managers improve their loan application process.

Technique

- Some critical and creative thinking is helpful here.

Software needed

- Word processor

In this lab, you will:

Part 1: Identify appropriate questions and develop a hypothesis for each question.

Part 2: Identify fields and values in a database that are relevant to your questions.