Data Analytics for Accounting

THIRD 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, THIRD EDITION

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 LWI 27 26 25 24 23 22

ISBN 978-1-264-44490-8 (bound edition)
MHID 1-264-44490-7 (bound edition)
ISBN 978-1-264-45732-8 (loose-leaf edition)
MHID 1-264-45732-4 (loose-leaf edition)

Executive Portfolio Manager: *Steve Schuetz*
Product Developer: *Michael McCormick*
Marketing Manager: *Claire McLemore*
Lead Content Project Managers: *Christine Vaughan; Angela Norris*
Senior Buyer: *Susan K. Culbertson*
Design: *Laurie Entringer*
Senior Content Licensing Specialist: *Lori Hancock*
Cover Image: *sasirin pamai/Shutterstock*
Compositor: *Straive*

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, Baruch College, Ryan A. Teeter, University of Pittsburgh, Katie L. Terrell, University of Arkansas.
Description: Third edition. | New York, NY : McGraw Hill LLC, [2023] | Includes index.
Identifiers: LCCN 2021045784 (print) | LCCN 2021045785 (ebook) | ISBN 9781264444908 (paperback ; alk. paper) | ISBN 9781264460571 (ebook)
Subjects: LCSH: Accounting—Data processing.
Classification: LCC HF5679 .R534 2023 (print) | LCC HF5679 (ebook) | DDC 657.0285—dc23
LC record available at https://lccn.loc.gov/2021045784
LC ebook record available at https://lccn.loc.gov/2021045785

mheducation.com/highered

Dedications

My wonderful daughter, Rachel, for your constant love, encouragement, and support. You always make me laugh and smile!

—Vern Richardson

To my three wonderful little Teeter tots, who keep me on my toes.

—Ryan Teeter

To the Mustache Running Club. Over many miles you all have learned more about accounting data analytics than you ever hoped for! Thanks for all of your support—on and off the trail.

—Katie Terrell

Preface

Data Analytics is changing the business world—data simply surround us! So many data are 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, and so on. Accountants create value by addressing fundamental business and accounting questions using Data Analytics.

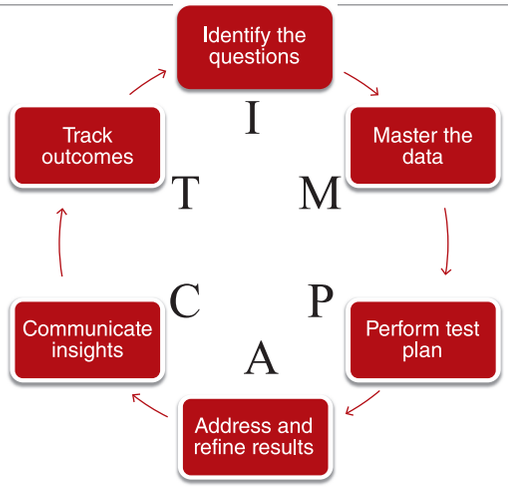
All accountants must develop data analytic skills to address the needs of the profession in the future—it is increasingly required of new hires and old hands. *Data Analytics for Accounting, 3e* recognizes that accountants don't need to become data scientists—they may never need to build a data repository or do the real hardcore 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. Developed analytics mindset—know 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 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. This may include diagnostic, predictive, or prescriptive analytics to appropriately analyze the data.
6. Statistical data analysis competency—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 an iterative 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*, *3e* describes this process by relying on an established Data Analytics model called the IMPACT cycle:[1]

1. **I**dentify the questions.
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 auditing, managerial accounting, financial accounting, and taxes in Chapters 5 through 9. In response to instructor feedback, *Data Analytics for Accounting, 3e* now also includes two new project chapters, giving students a chance to practice the full IMPACT model with multiple labs that build on one another.

Data Analytics for Accounting, 3e emphasizes hands-on practice with real-world data. 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 **Avalara**, **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 15 of the labs throughout the text (including Chapter 11).

Data Analytics for Accounting, 3e also emphasizes the various data analysis tools students will use throughout the rest of their career around two tracks—the Microsoft track (Excel, Power BI) and a Tableau track (Tableau Prep and Tableau Desktop—available with 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, Chinese University of Hong Kong–Shenzhen, 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 an author of McGraw Hill's *Accounting Information Systems* and *Introduction to Data Analytics for Accounting* textbooks.

Ryan A. Teeter

Ryan A. Teeter is a Clinical Associate 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. Dr. Teeter is also an author of McGraw Hill's *Introduction to Data Analytics for Accounting* textbook.

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 2021. 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. Terrell is also an author of McGraw Hill's *Introduction to Data Analytics for Accounting* textbook.

Acknowledgments

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

Our biggest thanks to the awesome team at McGraw Hill, including Steve Schuetz, Tim Vertovec, Rebecca Olson, Claire McLemore, Michael McCormick, Christine Vaughan, Kevin Moran, Angela Norris, and Lori Hancock.

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.

Joy Caracciolo, Will Cocker, and Tommy Morgan from Avalara for their help to grant permissions usage of the Avalara data.

Bonnie Klamm, North Dakota State University, and Ryan Baxter, Boise State University, for their accuracy check review of the manuscript and Connect content.

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
Ryan Baxter
Boise State University
Cory Campbell
Indiana State University
Heather Carrasco
Texas Tech University
Curtis Clements
Abilene Christian University
Elizabeth Felski
State University of New York at Geneseo
Amber Hatten
The University of Southern Mississippi
Jamie Hoeischer
Southern Illinois University, Edwardsville
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
Bonnie Klamm
North Dakota State University
Gregory Kogan
Long Island University, Brooklyn
Hagit Levy
Baruch College, CYNY
Brandon Lock
Baruch College, CUNY
Sharon M. Lightner
National University
Kalana Malimage
University of Wisconsin–Whitewater
Partha Mohapatra
California State University, Sacramento
Bonnie Morris
Duquesne University
Uday Murthy
University of South Florida
Kathy Nesper
University at Buffalo
Kamala Raghavan
Texas Southern University
Marie Rice
West Virginia University
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
Jacob Shortt
Virginia Tech
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

Vernon Richardson
Ryan Teeter
Katie Terrell

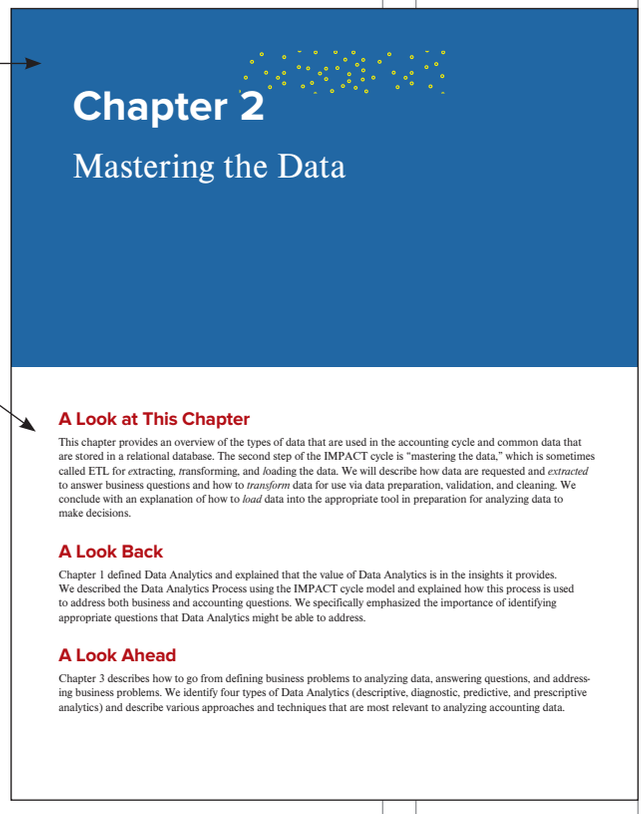
Key Features

- **NEW! Color Coded Multi-Track Labs:** Instructors have the flexibility to guide students through labs using the Green Track: Microsoft tools (including Excel, Power Query, and Power BI); Blue Track: Tableau tools (including Tableau Prep Builder and Tableau Desktop); or both. Each track is clearly identified and supported with additional resources.
- **NEW! Lab Example Outputs:** Each lab begins with an example of what students are expected to create. This provides a clear reference and guide for student deliverables.
- **NEW! Auto-Graded Problems:** The quantity and variety of auto-graded problems that are assignable in McGraw Hill Connect have been expanded.
- **NEW! Discussion and Analysis:** Now available as manually graded assignments in McGraw Hill Connect.
- **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 **Avalara**, **LendingClub**, **Dillard's**, **College Scorecard**, the **State of Oklahoma**, as well as financial statement data (via XBRL) from S&P100 companies.
- **Emphasis on Tools:** Students will learn how to conduct data analysis using Microsoft and Tableau tools. 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

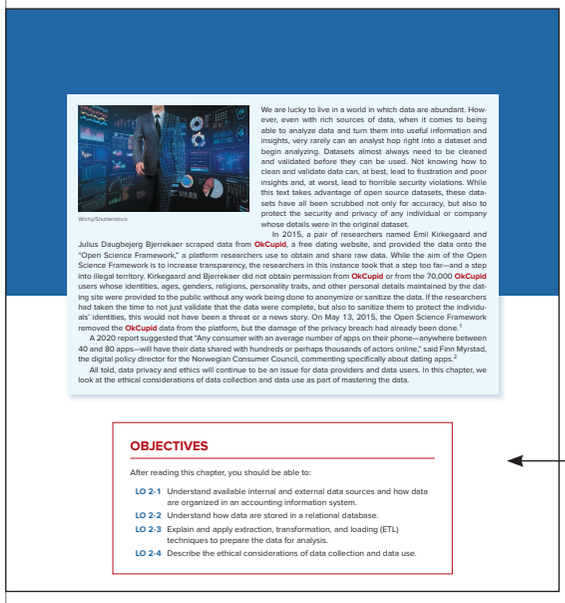
Chapter Maps

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 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.

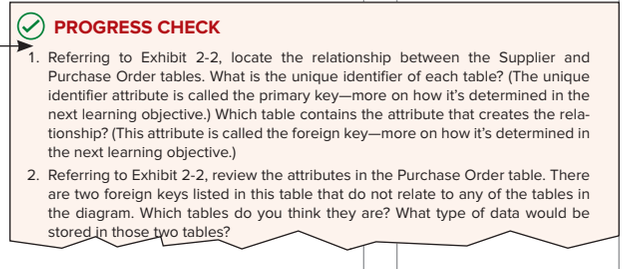


Learning Objectives

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.



End-of-Chapter Materials

Answers to Progress Checks

The answers 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 identifier
   Purchase Order table is [PO Number]. The Purchase Order table contains the foreig
2. The foreign key attributes in the Purchase Order table that do not relate to any t
   in the view are EmployeeID and CashDisbursementID. These attributes probably
   to the Employee table (so that we can tell which employee was responsible for
   Purchase Order) and the Cash Disbursement table (so that we can tell if the Pur
   Orders have been paid for yet, and if so, on which check). The Employee table woul
   complete listing of each employee, as well as containing the details about each emp
   (for example, phone number, address, etc.). The Cash Disbursement table would be
   ing of the payments the company has made.

Multiple Choice Questions

The multiple choice questions quickly assess student's knowledge of chapter content.

Multiple Choice Questions

1. **(LO 2-3)** Mastering the data can also be described via the ETL process. The E
   cess 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.

Discussion and Analysis—Now in Connect!

This feature provides questions for group discussion and analysis. Now available as manually graded assignments in McGraw Hill Connect!

Discussion and Analysis

1. **(LO 2-2)** The advantages of a relational database include limiting the amount of re
   dant data that are stored in a database. Why is this an important advantage? Wha
   go wrong when redundant data are stored?
2. **(LO 2-2)** The advantages of a relational database include integrating business
   cesses. Why is it preferable to integrate business processes in one information sys
   rather than store different business process data in separate, isolated databases?
3. **(LO 2-2)** Even though it is preferable to store data in a relational database, storing
   across separate tables can make data analysis cumbersome. Describe three reas
   is worth the trouble to store data in a relational database.
4. **(LO 2-2)** Among the advantages of using a relational database is enforcing busi
   rules. Based on your understanding of how the structure of a relational database h
   prevent data redundancy and other advantages, how does the primary key/fo
   key relationship structure help enforce a business rule that indicates that a com
   shouldn't process any purchase orders from suppliers who don't exist in the datab

Problems

The problems challenge the student's ability to see relationships in the learning objectives with analysis options that employ higher-level thinking and analytical skills. The quantity of auto-graded problems has been expanded. The manually graded analysis problems are also now assignable in McGraw Hill Connect.

Problems

1. **(LO 2-2)** Match the relational database function
   term:
   - Composition primary key
   - Descriptive attribute
   - Foreign key
   - Primary key
   - Relational database

NEW! Color Coded Multi-Track Labs

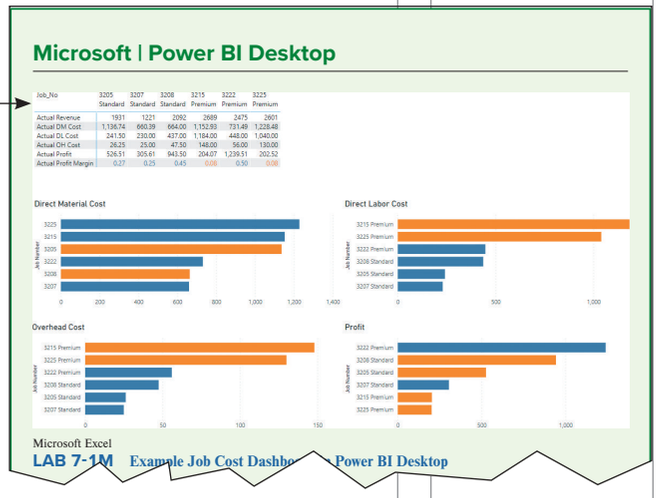
The labs give students hands-on experience working with different types of data and the tools used to analyze them. Students complete labs using the instructor-led track and answer common questions. Clear step-by-step directions help model the expected output of each lab exercise.

The Green Track—Microsoft/ Power BI: Example Output

The Green Track—Microsoft / Power BI: Easy to Follow Step-by-Step Lab Instruction

Microsoft | Power BI Desktop

1. Open Power BI Desktop and connect to your data:
   a. Click **Home > Get Data > Excel.**
   b. Browse to find the **Lab 7-1 Slainte Job Costs.xlsx** file and click **O**
   c. Check all of the tables and click **Load.**
   d. Click **Modeling > Manage relationships** to verify that the tables l correctly. For example, if you see an issue with the relationship b

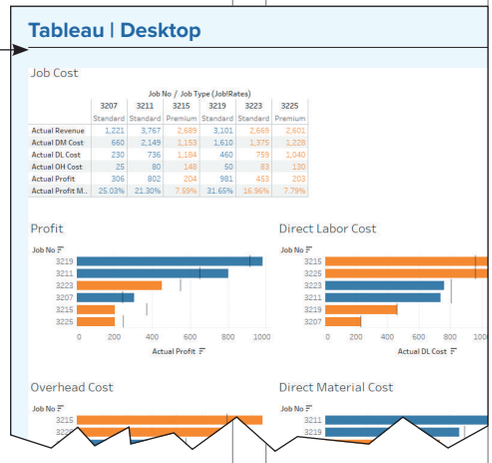


The Blue Track—Tableau: Example Output

The Blue Track—Tableau: Easy to Follow Step-by-Step Lab Instruction

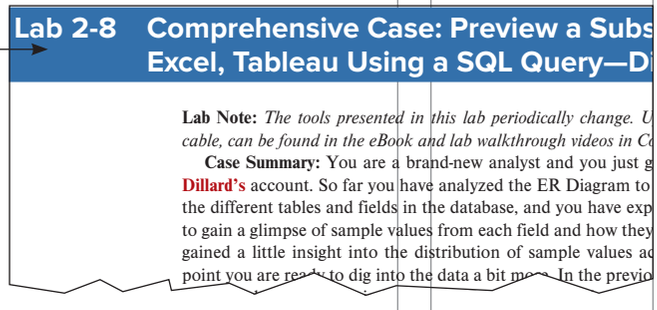
Tableau | Desktop

1. Open Tableau Desktop and connect to your data:
   a. Click **Connect to Data > Microsoft Excel.**
   b. Browse to find the **Lab 7-1 Slainte Job Costs.xlsx** file and click **Open.**
   c. Drag the **Job_Orders** table to the data model panel, then connect the **Customers, Time_Record, Material_Requisition,** and **Job_Rates** tables to right of it.
   d. Finally, drag the **Employees** table to the right of the **Time_Record** table.



Comprehensive Case Labs

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.



Data Analytics for Accounting, 3e
Content Updates

General Updates for the 3rd Edition

- Color coded multi-track labs now emphasize two tracks: The green Microsoft Track (including Excel, Power Query, and Power BI) and blue Tableau Track (including Tableau Prep Builder and Tableau Desktop).
- Added additional End-of-Chapter Multiple Choice Questions throughout the text that are auto-graded in Connect.
- Significantly revised many End-of-Chapter Problems for availability and auto-grading within Connect. Analysis Problems in Connect are manually graded.
- Linked chapter content to lab content using Lab Connections within the chapter content.

Chapter by Chapter Updates

Specific chapter changes for *Data Analytics for Accounting, 3e* are as follows:

Chapter 1

- Added new opening vignette regarding a recent IMA survey of finance and accounting professionals and their use of Big Data and Data Analytics.
- Added discussion on how analytics are used in auditing, tax, and management accounting.
- Included introduction to the variety of analytics tools available and explanation of dual tracks for labs including Microsoft Track and Tableau Track.
- Added "Data Analytics at Work" box feature: What Does an Analyst Do at a Big Four Accounting Firm.
- Added six new Connect-ready problems.
- Implemented lab changes:
  - All-new tool connections in Lab 1-5.
  - Revised Labs 1-0 to 1-4.

Chapter 2

- Edited opening vignette to include current examples regarding data privacy and ethics.
- Added a discussion on ethical considerations related to data collection and use.
- Added exhibit with potential external data sources to address accounting questions.
- Expanded the data extraction section to first include data identification, including the use of unstructured data.
- Added "Data Analytics at Work" box feature: Jump Start Your Accounting Career with Data Analytics Knowledge.
- Added six new Connect-ready problems.
- Implemented lab changes:
  - Revised Labs 2-1 to 2-8.

Chapter 3

- Refined the discussion on diagnostic analytics.
- Improved the discussion on the differences between qualitative and quantitative data and the discussion of the normal distribution.
- Refined the discussion on the use of regression as an analytics tool.
- Added examples of time series analysis in the predictive analytics section.
- Added "Data Analytics at Work" box feature: Big Four Invest Billions in Tech, Reshaping Their Identities as Professional Services Firm with a Technology Core.
- Added six new Connect-ready problems.
- Implemented lab changes:
  - All-new cluster analysis in Lab 3-2.
  - Revised Labs 3-1, 3-3 to 3-6.

Chapter 4

- Added discussion of statistics versus visualizations using Anscombe's quartet.
- Updated explanations of box plots and Z-scores.
- Added "Data Analytics at Work" box feature: Data Visualization: Why a Picture Can Be Worth a Thousand Clicks.
- Added six new Connect-ready problems.
- Implemented lab changes:
  - All-new dashboard in Lab 4-3.
  - Revised Labs 4-1, 4-2, 4-4, 4-5.

Chapter 5

- Improved and clarified content to match the focus on descriptive, diagnostic, predictive, and prescriptive analytics.
- Added "Data Analytics at Work" box feature: Citi's $900 Million Internal Control Mistake: Would Continuous Monitoring Help?
- Added six new Connect-ready problems.
- Implemented lab changes:
  - Revised Labs 5-1 to 5-5.

Chapter 6

- Clarified chapter content to match the focus on descriptive, diagnostic, predictive, and prescriptive analytics.
- Added "Data Analytics at Work" box features: Do Auditors Need to Be Programmers?
- Added six new Connect-ready problems.
- Implemented lab changes:
  - Major revisions to Labs 6-1 to 6-5.

Chapter 7

- Added new exhibit and discussion that maps managerial accounting questions to data approaches.
- Added "Data Analytics at Work" box feature: Maximizing Profits Using Data Analytics
- Added five new Connect-ready problems.
- Implemented lab changes:
  - All-new job cost, balanced scorecard, and time series dashboards in Lab 7-1, 7-2, 7-3.
  - Revised Lab 7-4, 7-5.

Chapter 8

- Added new exhibit and discussion that maps financial statement analysis questions to data approaches.
- Added four new Connect-ready problems.
- Implemented lab changes:
  - All-new sentiment analysis in Lab 8-4.
  - Revised Labs 8-1 to 8-3.

Chapter 9

- Added new exhibit and discussion that maps tax questions to data approaches.
- Added four new Connect-ready problems.
- Implemented lab changes:
  - Revised Labs 9-1 to 9-5.

Chapter 10

- Updated project chapter that evaluates different business processes, including the order-to-cash and procure-to-pay cycles, from different user perspectives with a choice to use the Microsoft track, the Tableau track, or both.
- Added extensive, all-new set of objective and analysis questions to assess analysis and learning.

Chapter 11

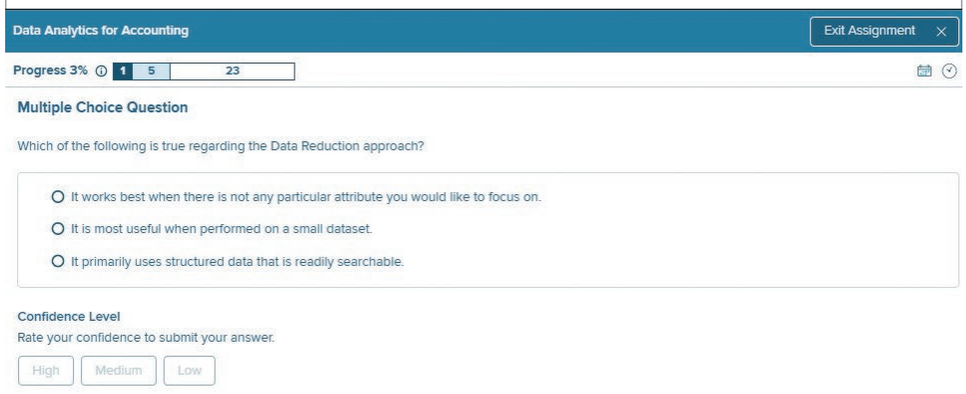
- Updated project chapter, estimating sales returns at **Dillard's** with three question sets highlighting descriptive and exploratory analysis, hypothesis testing, and predictive analytics with a choice to use the Microsoft track, the Tableau track, or both.
- Added extensive, all-new set of objective and analysis questions to assess analysis and learning.

Connect for *Data Analytics for Accounting*

With **McGraw Hill Connect** for *Data Analytics for Accounting,* your students receive proven study tools and hands-on assignment materials, as well as an adaptive eBook. Here are some of the features and assets available with Connect.

Proctorio: New remote proctoring and browser-locking capabilities, hosted by Proctorio within Connect, provide control of the assessment environment by enabling security options and verifying the identity of the student. Seamlessly integrated within Connect, these services allow instructors to control students' assessment experience by restricting browser activity, recording students' activity, and verifying students are doing their own work. Instant and detailed reporting gives instructors an at-a-glance view of potential academic integrity concerns, thereby avoiding personal bias and supporting evidence-based claims.

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.

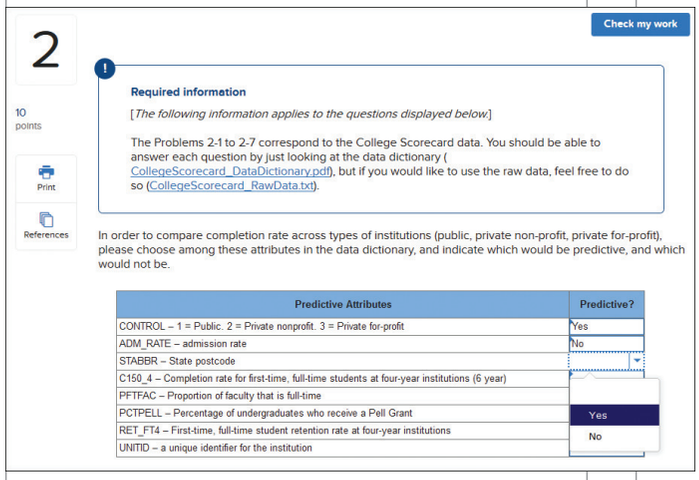


Orientation Videos: Video-based tutorial assignments are designed to train students via an overview video followed by a quiz for each of the assignment types they will find in McGraw Hill Connect.

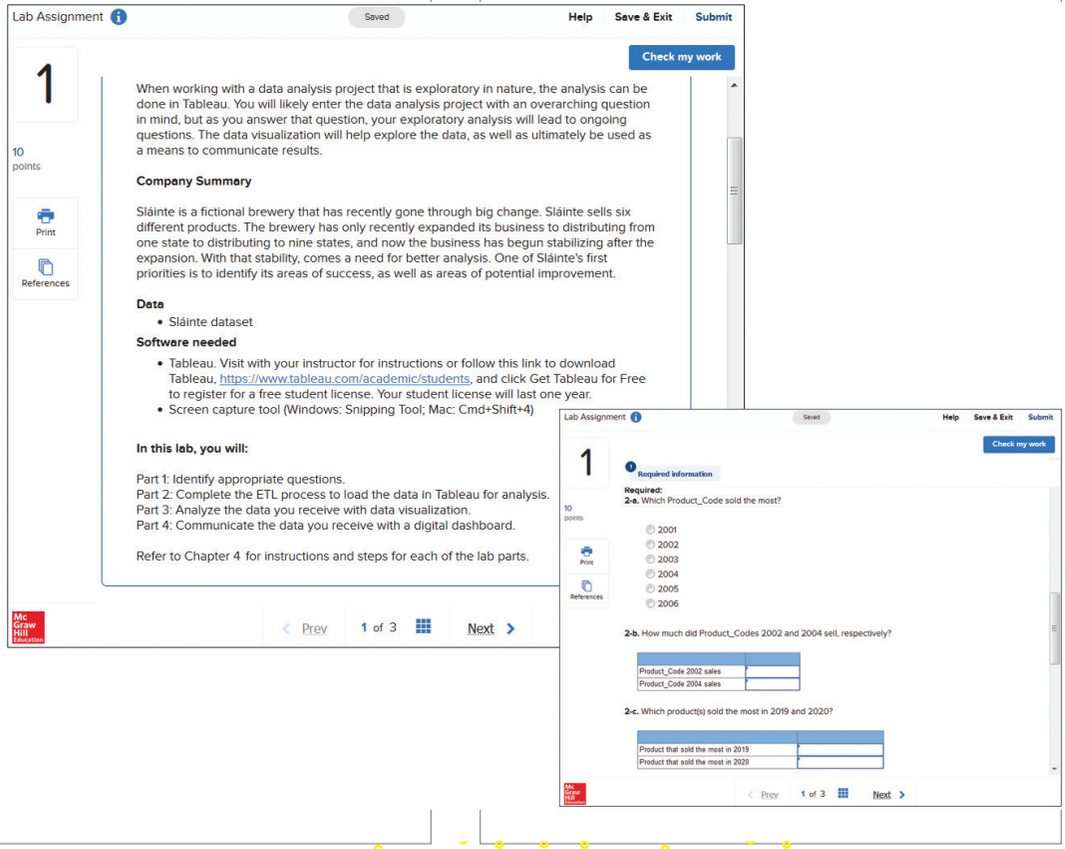
Multiple Choice Questions: The multiple choice questions from the end-of-chapter materials are assignable and auto-gradable in McGraw Hill Connect, with the option to provide students with instant feedback on their answers and performance.

Discussion and Analysis Questions: We have added the Discussion and Analysis questions into McGraw Hill Connect as manually graded assignments for convenience of assignment organization. These can be utilized for small group or in-class discussion.

Problems: Select problems from the text are auto-graded in McGraw Hill Connect. Manually graded analysis problems are also now available to ensure students are building an analytical skill set.



Color Coded Multi-Track Labs: Labs are assignable in McGraw Hill Connect as the green Microsoft Track (including Excel, Power Query, and Power BI) and blue Tableau Track (including Tableau Prep Builder and Tableau Desktop).



Students complete their lab work outside of Connect in the lab track selected by their professor. Students answer assigned lab questions designed to ensure they understood the key skills and outcomes from their lab work. Both auto-graded lab objective questions and manually graded lab analysis questions are assignable in Connect.

Comprehensive Cases: Comprehensive case labs are assignable in McGraw Hill Connect. Students 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.

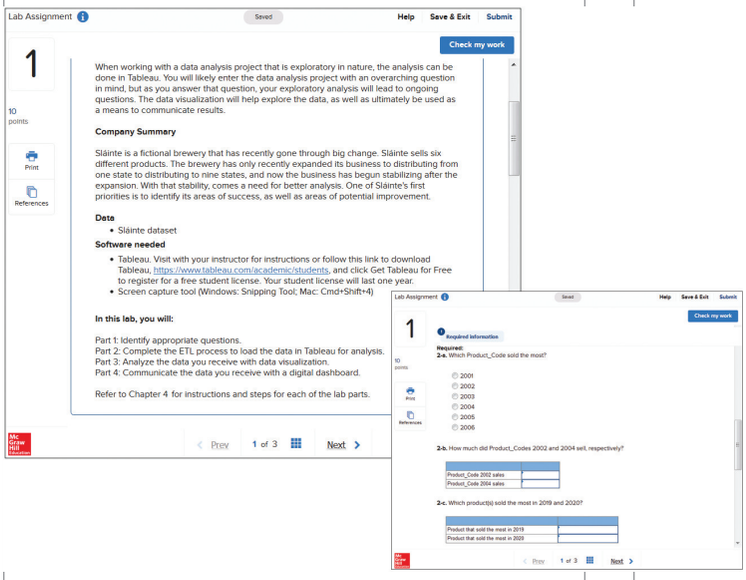
Lab 3-4 Comprehensive Case: Descriptive Analytics: Generate Summary Statistics—Dillard's

Lab Note: *The tools presented in this lab periodically change. Updated instructions, if applicable, can be found in the eBook and lab walkthrough videos in Connect.*

Case Summary: You are a brand new analyst and you just got assigned to work on the Dillard's account. So far you have analyzed the ER Diagram to gain a bird's eye view of all of the different tables and fields in the database, and you have explored the data in each table to gain a glimpse at sample values from each field and how they are all formatted. You also gained a little insight into the distribution of sample values across each field, but at this point you are ready to dig into the data a bit more.

Data: Dillard's sales data is only available on the University of Arkansas Remote Desktop (waltonlab.uark.edu). See your instructor for login credentials.

Lab Walkthrough Videos: These author-led lab videos in McGraw Hill Connect explain how to access and use the tools needed to complete the processes essential to the labs. Lab videos improve student success and minimize student questions!



Author Lecture Videos: Lecture Videos assignable in McGraw Hill Connect 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.

Writing Assignment: The Writing Assignment tool delivers a learning experience to help students improve their written communication skills and conceptual understanding. As an instructor you can assign, monitor, grade, and provide feedback on writing more efficiently and effectively in McGraw Hill Connect.

Test Bank: The test bank includes auto-graded multiple choice and true/false assessment questions. The test bank can be assigned directly within McGraw Hill Connect or exported from Test Builder.

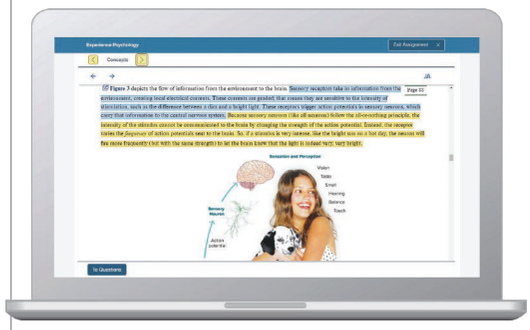
Mc Graw Hill connect®

Instructors: Student Success Starts with You

Tools to enhance your unique voice

Want to build your own course? No problem. Prefer to use an OLC-aligned, 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

Study made personal

Incorporate adaptive study resources like SmartBook® 2.0 into your course and help your students be better prepared in less time. Learn more about the powerful personalized learning experience available in SmartBook 2.0 at **www.mheducation.com/highered/connect/smartbook**

Affordable solutions, added value

Make technology work for you with LMS integration for single sign-on access, mobile access to the digital textbook, and reports to quickly show you how each of your students is doing. And with our Inclusive Access program you can provide all these tools at a discount to your students. 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. Visit **www.supportateverystep.com** for videos and resources both you and your students can use throughout the semester.

Checkmark: Jobalou/Getty Images

SUPPORT AT every step

Students: Get Learning that Fits You

Effective tools for efficient studying

Connect is designed to help you be more productive with simple, flexible, intuitive tools that maximize your study time and meet your individual learning needs. Get learning that works for you with Connect.

Study anytime, anywhere

Download the free ReadAnywhere app and access your online eBook, SmartBook 2.0, or Adaptive Learning Assignments when it's convenient, even if you're offline. And since the app automatically syncs with your Connect account, 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 text-book in front of you."

- Jordan Cunningham,
Eastern Washington University

Calendar: owattaphotos/Getty Images

Everything you need in one place

Your Connect course has everything you need—whether reading on your digital eBook or completing assignments for class, Connect makes it easy to get your work done.

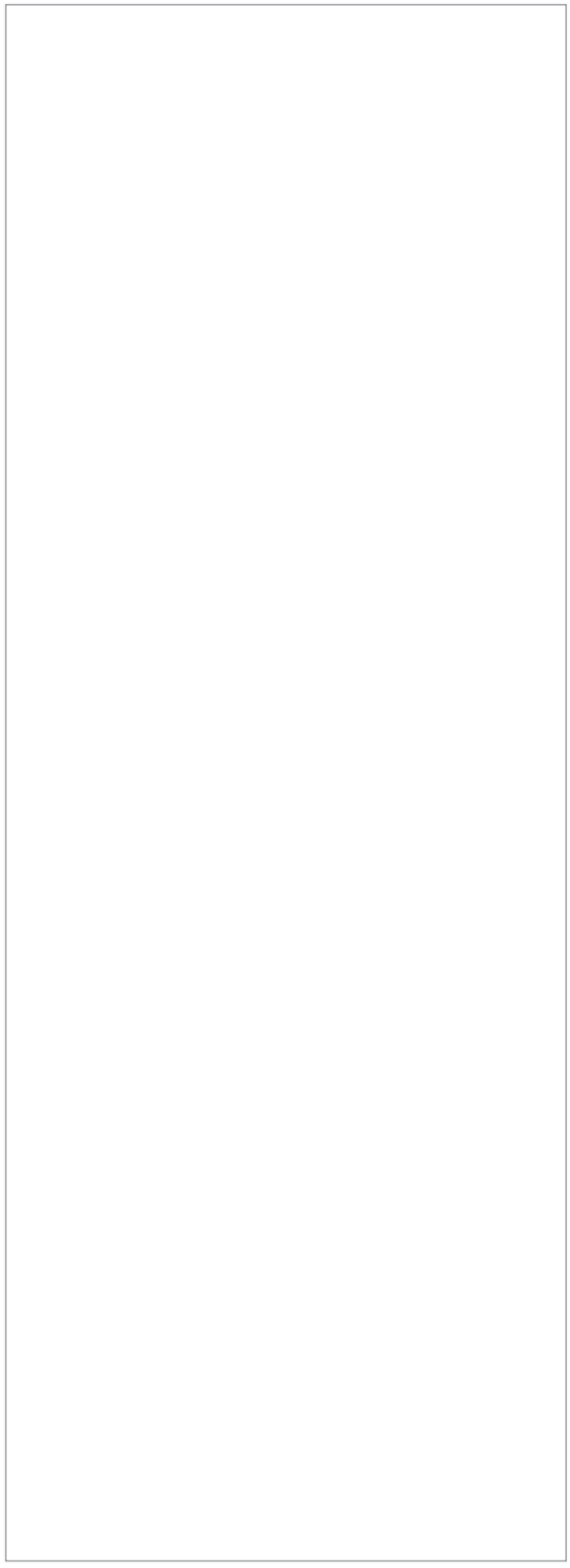
Learning for everyone

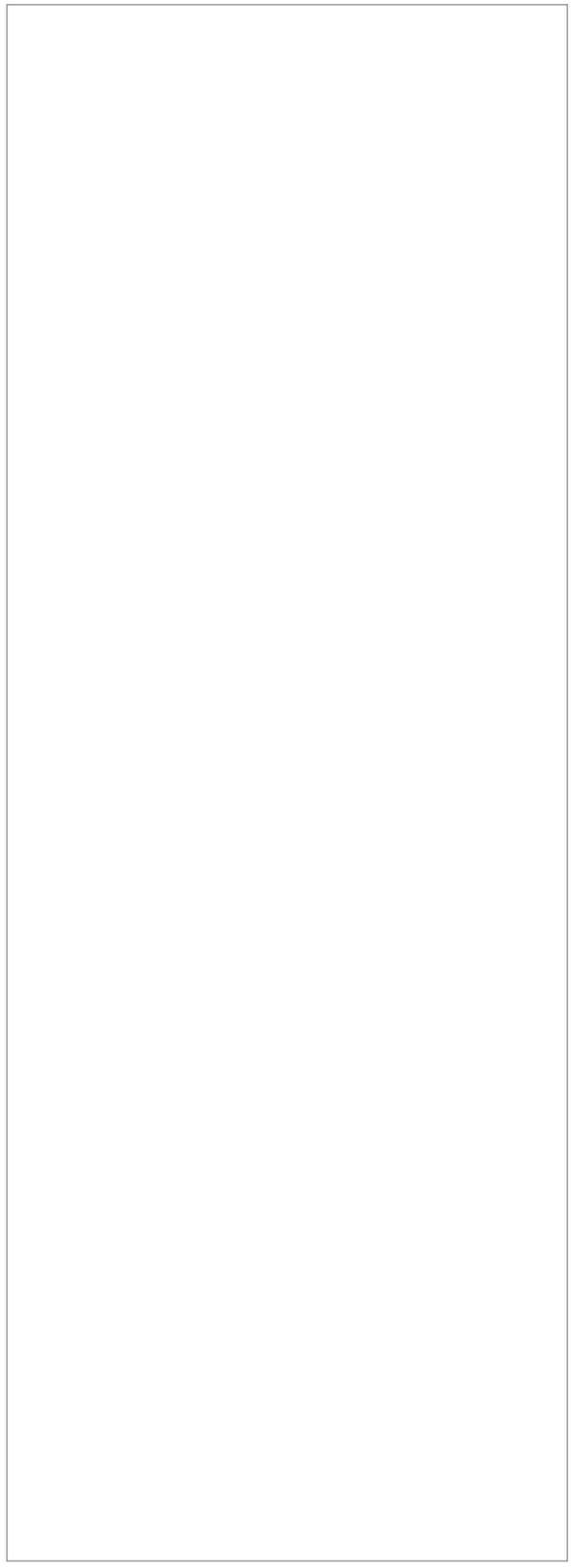
McGraw Hill works directly with Accessibility Services Departments and faculty to meet the learning needs of all students. Please contact your Accessibility Services Office and ask them to email accessibility@mheducation.com, or visit **www.mheducation.com/about/accessibility** for more information.

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

Brief Table of Contents

Detailed TOC





Data Analytics for Accounting

Chapter 1

Data Analytics for Accounting and Identifying the Questions

A Look at This Chapter

Data Analytics is changing both business and accounting. In this chapter, we define Data Analytics and explain its impact on business and the accounting profession, noting that the value of Data Analytics is derived from the insights it provides. We also describe the need for an analytics mindset in the accounting profession. We next describe the Data Analytics Process using the IMPACT cycle and explain how this process is used to address both business and accounting questions. We then emphasize the skills accountants need as well as the tools available for their use. In this chapter, we specifically emphasize the importance of identifying appropriate accounting 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 address accounting 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 data sender and receiver. We finalize the chapter by emphasizing the need for ethical data collection and data use to maintain data privacy.

Cobalt S-Elinoi/Shutterstock

As the access to accounting data proliferates and tools and accountant skills advance, accountants are relying more on Big Data to address accounting questions. Whether those questions relate to audit, tax or other accounting areas, increasingly value will be created by performing Data Analytics. In this chapter, we introduce you to the need for Data Analytics in accounting, and how accounting professionals are increasingly asked to develop an analytics mindset for any and all accounting roles.

Technology such as Data Analytics, artificial intelligence, machine learning, blockchain, and robotic process automation will be playing a greater role in the accounting profession this year, according to a recent report from the Institute of Management Accountants.

The report indicates that finance and accounting professionals are increasingly implementing Big Data in their business processes, and the pattern is likely to continue in the future. The IMA surveyed its members for the report and received 170 responses from CFOs and other management accountants. Many of the CFOs are predicting big changes for 2020 in their businesses.

Sources: M. Cohn, "Accountants to Rely More on Big Data in 2020," *Accounting Today,* January 4, 2020, https://www.accountingtoday.com/news/accountants-to-rely-more-on-big-data-in-2020 (accessed December 2020).

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 the IMPACT model may be used to address a specific business question.

LO 1-1

Define Data Analytics.

DATA ANALYTICS

Data surround us! By the year 2024, it is expected that the volume of data created, captured, copied, and consumed worldwide will be 149 zettabytes (compared to 2 zettabytes in 2010 and 59 zettabytes in 2020).[1] In fact, more data have been created in the last 2 years than in the entire previous history of the human race.[2] With so much data available about each of us (e.g., how we shop, what we read, what we've bought, what music we listen to, where we travel, whom we trust, what devices we use, 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 data** (data that adheres to a predefined data model in a tabular format) and **unstructured data** (data that does not adhere to a predefined data format) to discover unknown patterns or relationships.[3] 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.[4] 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 (or frankly any available data source) is by use of four Vs: its volume (the sheer size of the dataset), velocity (the speed of data processing), variety (the number of types of data), and veracity (the underlying quality of the data). While sometimes *Data Analytics* and *Big Data* are terms used interchangeably, we will use the term *Data Analytics* throughout and focus on the possibility of turning data into knowledge and that knowledge into insights that create value.

PROGRESS CHECK

1. How does having more data around us translate into value for a company? What must we do with those data to extract value?
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 and 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

[1] Statista, https://www.statista.com/statistics/871513/worldwide-data-created/ (accessed December 2020).
[2] Bernard Marr, "Big Data: 20 Mind-Boggling Facts Everyone Must Read," *Forbes,* September 30, 2015, at http://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#2a3289006c1d (accessed March 2019).
[3] 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.
[4] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence Research," *MIS Quarterly* 34, no. 1 (2010), pp. 201–3.

a high value on Data Analytics. In fact, per **PwC**'s 6th Annual Digital IQ survey of more than 1,400 leaders from digital businesses, the area of investment that tops CEOs' list of priorities is business analytics.[5]

A recent study from **McKinsey** Global Institute estimates that Data Analytics and technology could generate up to $2 trillion in value per year in just a subset of the total possible industries affected.[6] 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. By pairing structured data with unstructured data, patterns could be discovered that create new meaning, creating value and competitive advantage. In addition to producing more value externally, studies show that Data Analytics affects internal processes, improving productivity, utilization, and growth.[7]

And increasingly, data analytic tools are available as self-service analytics allowing users the capabilities to analyze data by aggregating, filtering, analyzing, enriching, sorting, visualizing, and dashboarding for data-driven decision making on demand.

PwC notes that while data has always been important, executives are more frequently being asked to make data-driven decisions in high-stress and high-change environments, making the reliance on Data Analytics even greater these days![8]

PROGRESS CHECK

3. Let's assume a brand manager at **Procter and Gamble** identifies that an older demographic might be concerned with the use of Tide Pods to do their laundry. How might **Procter and Gamble** use Data Analytics to assess if this is a problem?
4. How might Data Analytics assess the decision to either grant overtime to current employees or hire additional employees? Specifically, 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.

[5] "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).
[6] "The Trillion-Dollar Opportunity for the Industrial Sector: How to Extract Full Value from Technology," McKinsey Global Institute, https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/the-trillion-dollar-opportunity-for-the-industrial-sector#, November 2018 (accessed December 2018).
[7] Joseph Kennedy, "Big Data's Economic Impact," https://www.ced.org/blog/entry/big-datas-economic-impact, December 3, 2014 (accessed January 9, 2016).
[8] "What's Next for Tech for Finance? Data-Driven Decision Making," PwC, https://www.pwc.com/us/en/cfo-direct/accounting-podcast/data-driven-decision-making.html, October 2020 (accessed December 2020).

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. Audits 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 refine the focus on risk and derive deeper insights into an organization."[9] Many auditors believe that audit 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 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 is changing 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.[10] 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."[11]

Data Analytics also expands auditors' capabilities in services like testing for fraudulent transactions and automating compliance-monitoring activities (like filing financial reports to the U.S. Securities and Exchange Commission [SEC] or to the Internal Revenue Service [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.

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

[10] 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).

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

We address auditing questions and Data Analytics in Chapters 5 and 6.

Lab Connection

Lab 1-3 has you explore questions auditors would answer with Data Analytics.

Management Accounting

Of all the fields of accounting, it would seem that the aims of Data Analytics are most akin to management accounting. Management accountants (1) are asked questions by management, (2) find data to address those questions, (3) analyze the data, and (4) report the results to management to aid in their decision making. The description of the management accountant's task and that of the data analyst appear to be quite similar, if not identical in many respects.

Whether it be understanding costs via job order costing, understanding the activity-based costing drivers, forecasting future sales on which to base budgets, or determining whether to sell or process further or make or outsource its production processes, analyzing data is critical to management accountants.

As information providers for the firm, it is imperative for management accountants to understand the capabilities of data and Data Analytics to address management questions.

We address management accounting questions and Data Analytics in Chapter 7.

Lab Connection

Lab 1-2 and **Lab 1-4** have you explore questions managers would answer with Data Analytics.

Financial Reporting and Financial Statement Analysis

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?
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 to and opportunities for the firm. For example, in a data analytic 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 that might affect the company's interactions with them and/or open up new opportunities that otherwise it wouldn't have considered?

We address financial reporting and financial statement analysis questions and Data Analytics in Chapter 8.

Lab Connection

Lab 1-1 has you explore questions financial accountants would answer with Data Analytics.

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 predict what will happen rather than react 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 in one of their most value-adding tasks, that of tax planning!

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 use in 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 integrating the strategy with the IT department and gaining access to available technology tools.[12]

We address tax questions and Data Analytics in Chapter 9.

PROGRESS CHECK

5. Why are management accounting and Data Analytics considered similar in many respects?
6. How specifically will Data Analytics change the way a tax staff does its taxes?

[12]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).

THE DATA ANALYTICS PROCESS USING THE IMPACT CYCLE

LO 1-4

Describe 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 later in Chapters 2, 3, and 4. We use its approach for thinking about the steps included in Data Analytics throughout this textbook, all the way from carefully identifying the question to accessing and analyzing the data to communicating insights and tracking outcomes.[13]

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, how to reduce costs, or how to find errors or fraud. Having a concrete, specific question that is potentially answerable by Data Analytics is an important first step.

Indeed, accountants often possess a unique skillset to improve an organization's Data Analytics by their ability to ask the right questions, especially since they often understand a company's financial data. In other words, "Your Data Won't Speak Unless You Ask It the Right Data Analysis Questions."[14] We could ask any question in the world, but if we don't ultimately have the right data to address the question, there really isn't much use for Data Analytics for those questions.

Additional attributes to consider might include the following:

- Audience: Who is the audience that will use the results of the analysis (internal auditor, CFO, financial analyst, tax professional, etc.)?
- Scope: Is the question too narrow or too broad?
- Use: How will the results be used? Is it to identify risks? Is it to make data-driven business decisions?

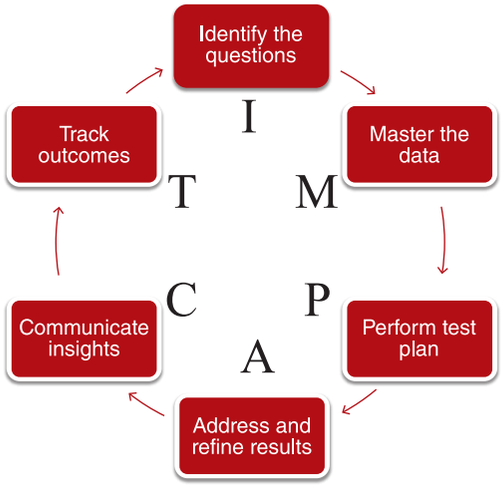


EXHIBIT 1-1
The IMPACT Cycle

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

[13] We also note our use of the terms *IMPACT cycle* and *IMPACT model* interchangeably throughout the book.

[14] M. Lebied, "Your Data Won't Speak Unless You Ask It the Right Data Analysis Questions," *Datapine,* June 21, 2017, https://www.datapine.com/blog/data-analysis-questions/ (accessed December 2020).

Here is an example of potential questions accountants might address using Data Analytics:

- Are employees circumventing internal controls over payments?
- What are appropriate cost drivers for activity-based costing purposes?
- To minimize taxes, should we have our company headquarters in Dublin, Ireland, or in Chicago?
- Are our customers paying us in a timely manner? Are we paying our suppliers in a timely manner?
- How can we more accurately 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 made in journal entries be identified?
- Should we outsource our products to Indonesia, or produce them ourselves?

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 or missing data), frequency of updates, what time periods are covered to make sure the data coincide with the timing of our business problem, and so on.

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.
- Examine data dictionaries and other contextual data—to provide details about the data.
- Evaluate and perform the ETL (extraction, transformation, and loading) processes and assess the time required to complete.
- Assess data validation and completeness—to provide a sense of the reliability of the data.
- Evaluate and perform data normalization—to reduce data redundancy and improve data integrity.
- Evaluate and perform 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.[15]

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

[15] "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).

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.

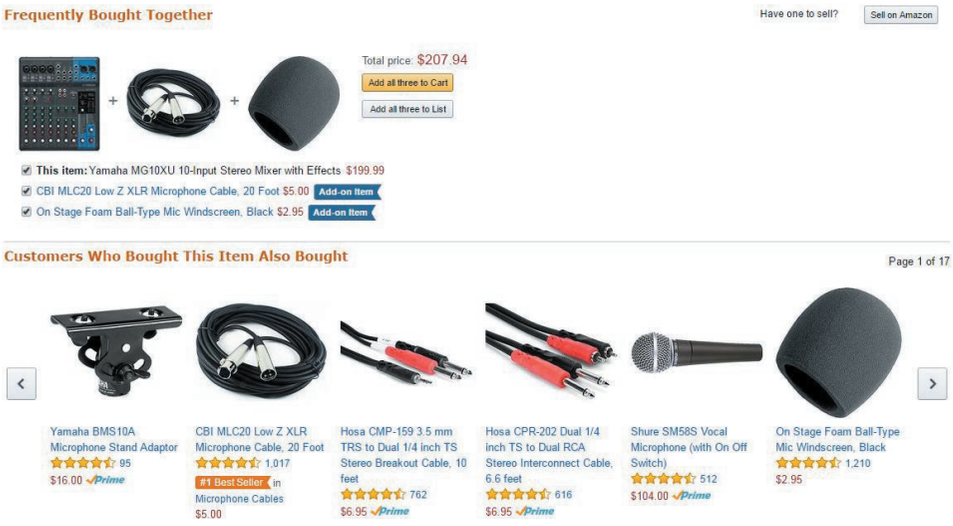
The research question, the model, the data availability, and the expected statistical inference may all suggest the use of different data approaches. Provost and Fawcett[16] detail eight different approaches to Data Analytics depending on the question. We will discuss the most applicable ones to accounting 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 of 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? Which companies are most likely to go bankrupt in the next two years?
- **Regression–**A data approach used to predict a specific dependent variable value based on independent variable inputs using a statistical model. Regression analysis might be used to assess the relationship between an investment in R&D and subsequent operating income. Another example would be the use of regression to identify an appropriate cost driver to allocate overhead as part of activity-based costing.
- **Similarity matching–**An attempt to identify similar individuals based on data known about them. A company may use similarity matching to find new customers that may closely resemble their best customers (in hopes that they find additional profitable customers).
- **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 loyalty card customers into groups based on buying behavior related to shopping frequency or purchasing volume, 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. . . ." Exhibit 1-2 shows us an **Amazon** search for the Yamaha MG10XU stereo mixer provides several related item suggestions to the customer.
- **Profiling–**An attempt to characterize the "typical" behavior of an individual, group, or population by generating summary statistics about the data (including mean, median, minimum, maximum, and standard deviation). 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).

[16]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).

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

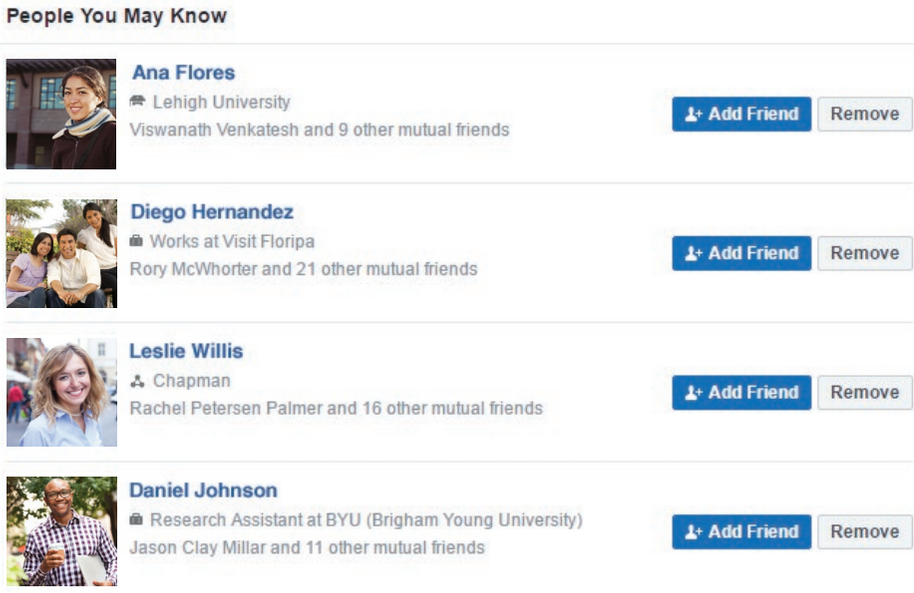
Amazon Inc.



- **Link prediction**—An attempt to predict connections 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 in the same year, 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.

EXHIBIT 1-3
Example of Link Prediction on Facebook

Michael DeLeon/Getty Images; Sam Edwards/Glow Images; Daniel Ernst/Getty Images; Exactostock/SuperStock; McGraw HIll



- **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 (e.g., highest cost, 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 additional vetting (such as substantive testing) 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, dice, and manipulate the data; find correlations; test hypotheses; ask ourselves further, hopefully better questions; ask colleagues what they think; and revise and rerun the analysis potentially multiple times. But once that is complete, we have the results ready to communicate to interested stakeholders that hopefully directly addresses their questions.

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 executive summaries, 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, recognizing that different users consume deliverables in a potentially different way. Increasingly, digital dashboards and data visualizations are particularly helpful in communicating insights and tracking outcomes.

Back to Step 1

Since the IMPACT cycle is iterative, once insights are gained and outcomes are tracked, new more refined questions emerge that may use the same or different data sources with potentially different analyses and thus, 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 allow the auditor to spend more time and effort on the most important (e.g., most risky, largest dollar volume, etc.) items?

DATA ANALYTIC SKILLS AND TOOLS NEEDED BY ANALYTIC-MINDED ACCOUNTANTS

LO 1-5

Describe the skills and tools needed by accountants.

While we don't believe that accountants need to become data scientists—they may never need to build a database from scratch or perform 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 skills that analytic-minded accountants should have in the following seven areas:

1. Developed analytics mindset—know 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. This may include diagnostic, predictive, or prescriptive analytics to appropriately analyze the data.
6. Statistical data analysis competency—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.

Data Analytics at Work

What Does a Data Analyst Do at a Big Four Accounting Firm?

Data Sources: We extract financial data from a number of different ERP systems including SAP, Abacus, Sage, and Microsoft Navision (among others).

Data Scrubbing and Data Preparation: A huge part of our time goes into data cleaning and data transformation.

Tools Used: Excel, Unix commands, SQL, and Python are used to automate large chunks of our work.

Knowledge Needed: Basic Excel, programming skills (SQL, Python), and audit knowledge such as understanding journal entries and trial balances are needed.

Source: "Data Analyst at a Big 4—What Is It Like? My Opinion Working as a Data Analyst at a Big Four," https://cryptobulls.info/data-analyst-at-a-big-4-what-is-it-like-pros-cons-ernst-young-deloitte-pwc-kpmg, posted February 29, 2020 (accessed January 2, 2021).

Choose the Right Data Analytics Tools

In addition to developing the right skills, it is also important to be familiar with the right Data Analytics tools for each task. There are many tools available for Data Analytics

preparation, modeling, and visualization. Gartner annually assesses a collection of these tools and creates the "magic quadrant" for business intelligence, depicted in Exhibit 1-4. The magic quadrant can provide insight into which tools you should consider using.

EXHIBIT 1-4
Gartner Magic Quadrant for Business Intelligence and Analytics Platforms

Source: Sallam, R. L., C. Howson, C. J. Idoine, T. W. Oestreich, J. L. Richardson, and J. Tapadinhas, "Magic Quadrant for Business Intelligence and Analytics Platforms," Gartner RAS Core Research Notes, Gartner, Stamford, CT (2020).

Based on Gartner's magic quadrant, it is easy to see that Tableau and Microsoft provide innovative solutions. While there are other tools that are popular in different industries, such as Qlik and TIBCO, Tableau and Microsoft tools are the ones you will most likely encounter because of their position as leaders in the Data Analytics space. For this reason, each of the labs throughout this textbook will give you or your instructor the option to choose either a Microsoft Track or a Tableau Track to help you become proficient in those tools. The skills you learn as you work through the labs are transferrable to other tools as well.

The Microsoft Track

Microsoft's offerings for Data Analytics and business intelligence (BI) include Excel, Power Query, Power BI, and Power Automate. It is likely that you already have some familiarity with Excel as it is used for everything from recording transactions to running calculations and preparing financial reports. These tools are summarized in Exhibit 1-5.

Excel is the most ubiquitous spreadsheet software and most commonly used for basic data analysis. It allows for the creation of tables, advanced formulas to do quick or complex calculations, and the ability to create PivotTables and basic charts and graphs. One major issue with using Excel for analysis of large datasets is the 1,048,576 row limit due to memory constraints. It is available on Windows and Mac as well as through the Microsoft

EXHIBIT 1-5
Microsoft Data Analytics Tools

Tool	Excel	Power Query	Power BI	Power Automate
Good for	Small datasets Data tables PivotTables Basic visualization	Large datasets Data joining Data cleaning Data transformation	Large datasets Advanced visualization Dashboards Presentation	Collect data from multiple sources Robotics process automation
Platform	Windows/Mac/Online	Windows	Windows/Online	Online

365 online service for simple collaboration and sharing, although the most complete set of features and compatibility with advanced plug-ins is available only on Windows.

Power Query is a tool built into Excel and Power BI Desktop on Windows that lets Excel connect to a variety of different data sources, such as tables in Excel, databases hosted on popular platforms like SQL Server, or through open database connectivity (ODBC) connections. Power Query makes it possible to connect, manipulate, clean, and join data so you can pull them into your Excel sheet or use them in Power BI to create summary reports and advanced visualizations. Additionally, it tracks each step you perform so you can apply the same transformations to new data without recreating the work from scratch.

Power BI is an analytic platform that enables generation of simple or advanced Data Analytics models and visualizations that can be compiled into dashboards for easy sharing with relevant stakeholders. It builds on data from Excel or other databases and can leverage models created with Power Query to quickly summarize key data findings. Microsoft provides Power BI Desktop for free only on Windows or through a web-based app, though the online version does not have all of the features of the desktop version and is primarily used for sharing.

Power Automate is a tool that leverages robotics process automation (RPA) to automate routine tasks and workflows, such as scraping and collecting data from nonstructured sources, including emails and other online services. These can pull data from relevant sources based on events, such as when an invoice is generated. Power Automate is a web-based subscription service with a tool that works only on Windows to automate keystrokes and mouse clicks.

The Tableau Track

In previous years, Tableau was ranked slightly higher than Microsoft on its ability to execute and it continues to be a popular choice for analytics professionals. Tableau's primary offerings include Tableau Prep for data preparation and Tableau Desktop for data visualization and storytelling. Tableau has an advantage over Microsoft in that its tools are available for both Windows and Mac computers. Additionally, Tableau offers online services through Tableau Server and Tableau Online with the same, complete feature set as their apps. These are summarized in Exhibit 1-6.

EXHIBIT 1-6
Tableau Data Analytics Tools

Tool	Tableau Prep Builder	Tableau Desktop	Tableau Public
Good for	Large datasets Data summarization Data joining Data cleaning Data transformation	Large datasets Advanced visualization Dashboards Presentation	Analyze and share public datasets
Platform	Windows/Mac/Online	Windows/Mac/Online	Windows/Mac/Online

Tableau Prep is primarily used for data combination, cleaning, manipulation, and insights. It enables users to interact with data and quickly identify data quality issues with

a clear map of steps performed so others can review the cleaning process. It is available on Windows, Mac, and Tableau Online.

Tableau Desktop can be used to generate basic to advanced Data Analytics models and visualizations with an easy-to-use drag-and-drop interface.

Tableau Public is a free limited edition of Tableau Desktop that is specifically tailored to sharing and analyzing public datasets. It has some significant limitations for broader analysis.

PROGRESS CHECK

9. Given the "magic quadrant" in Exhibit 1-4, why are the software tools represented by the Microsoft and Tableau tracks considered innovative?
10. Why is having the Tableau software tools fully available on both Windows and Mac computers an advantage for Tableau over Microsoft?

HANDS-ON EXAMPLE OF THE IMPACT MODEL

LO 1-6

Explain how the IMPACT model may be used to address a specific business question.

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, 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, "What are some characteristics of rejected loans?"

Master the Data

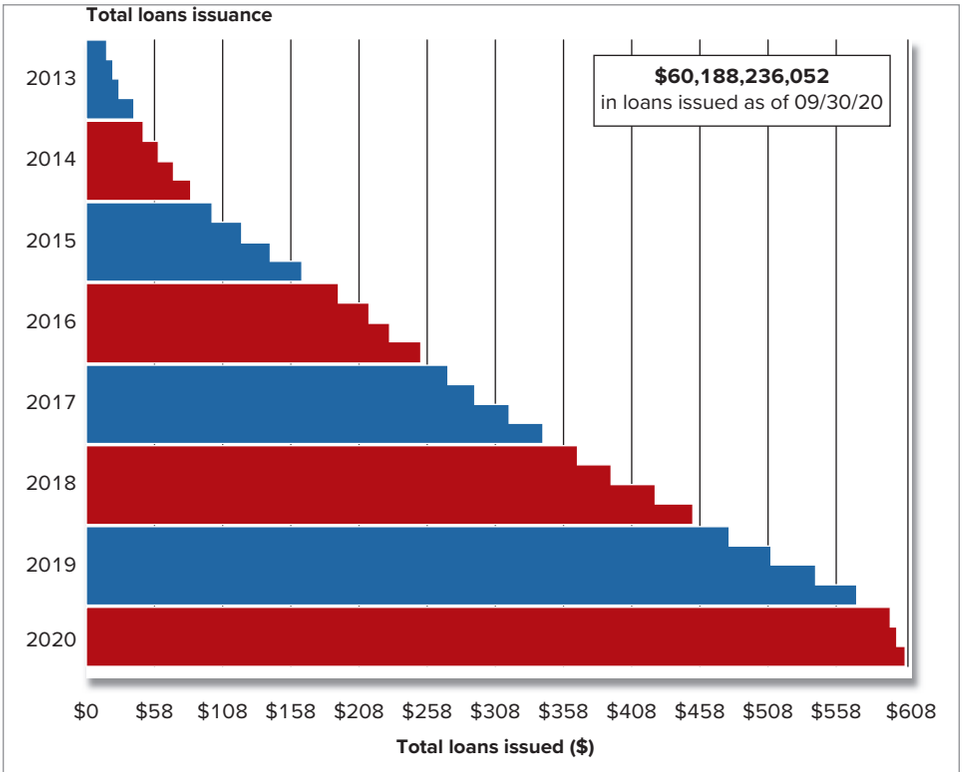
LendingClub is a U.S.-based, peer-to-peer lending company, headquartered in San Francisco, California. **LendingClub** facilitates both borrowing and lending by providing a platform for unsecured personal loans between $1,000 and $35,000. The loan period is for either 3 or 5 years. There is information 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 $60 billion in loans via **LendingClub**.[17]

Some basic lending statistics are included on the LendingClub Statistics website (Exhibit 1-7). Each bar represents the volume of loans each quarter during its respective year.

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

EXHIBIT 1-7
LendingClub Statistics

Source: Accessed December 2020, https://www.lendingclub.com/info/statistics.action

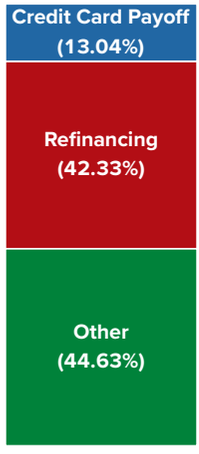


Borrowers borrow money for a variety of reasons, including refinancing other debt and paying off credit cards, as well as borrowing for other purposes (Exhibit 1-8).

EXHIBIT 1-8
LendingClub Statistics by Reported Loan Purpose

42.33% of **LendingClub** borrowers report using their loans to refinance existing loans as of September 30, 2020.

Source: Accessed December 2020, https://www.lendingclub.com/info/statistics.action



LendingClub provides datasets on the loans it approved and funded as well as data for the loans that were declined. To address the question posed, "What are some characteristics of rejected loans?," we'll use the dataset of rejected loans.

The rejected loan datasets and related data dictionary are available from your instructor or from Connect (in Additional Student Resources).

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 1-9.

EXHIBIT 1-9
2007–2012 Lending-Club Data Dictionary for Declined Loan Data

Source: LendingClub

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 "DAA Chapter 1-1 Data" that has rejected loan statistics from **LendingClub** for the time period of 2007 to 2012. It is a cleaned-up, transformed file ready for analysis. We'll learn more about data scrubbing and preparation of the data in Chapter 2.

Exhibit 1-10 provides a cut-out of the 2007–2012 "Declined Loan" dataset provided.

EXHIBIT 1-10
2007–2012 Declined Loan Applications (DAA Chapter 1-1 Data) Dataset

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, "What are the characteristics of rejected loans (at **LendingClub**)?," and the available data, we will do three analyses to assess what is considered in rejecting/accepting 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 these three loan characteristics, we believe it will provide **LendingClub** with the data needed to assess whether the potential 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* of the potential borrower. That is, before adding this potential loan, how big is the potential borrower's debt compared to the size of the potential borrower's annual income?

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.

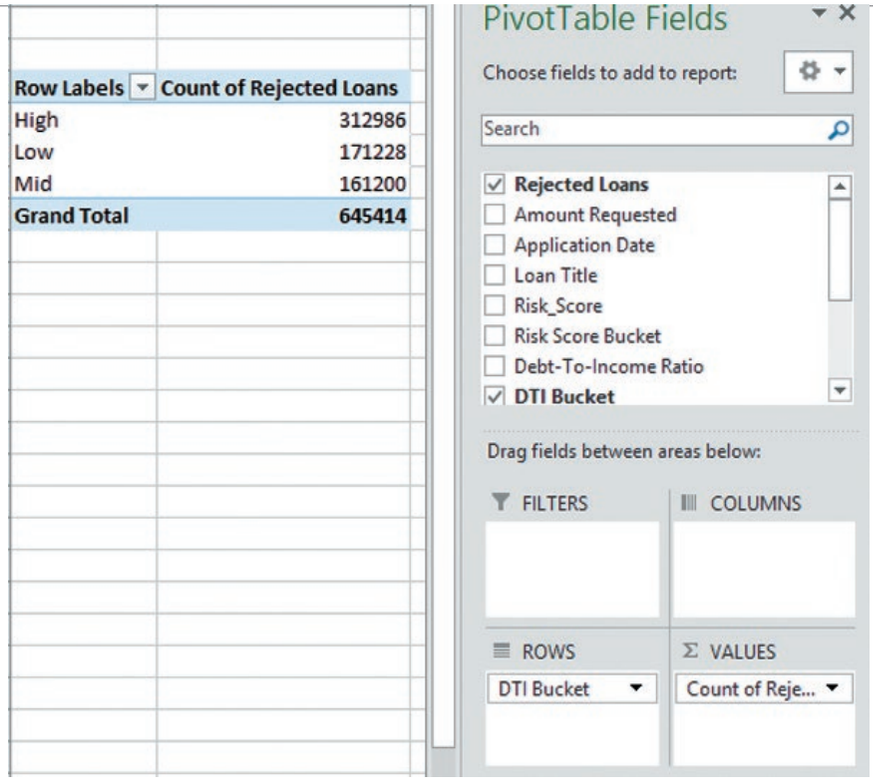
The Excel PivotTable is an easy way to make comparisons between the different levels of DTI. When we run a PivotTable analysis, we highlight the loans, which count the number of loans applied for and rejected, and the DTI bucket (see Exhibit 1-11). The PivotTable counts the number of loan applications 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 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 is on the length of employment and its relationship with rejected loans (see Exhibit 1-12). 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 of employment 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?

The third analysis we perform is to consider the credit or risk score of the applicant. As noted in Exhibit 1-13, 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).

We will 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 1-13.

As part of the analysis of credit score and rejected loans, we again perform PivotTable analysis (as seen in Exhibit 1-14) by counting 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.



Row Labels	Count of Rejected Loans
High	312986
Low	171228
Mid	161200
Grand Total	645414

PivotTable Fields

Choose fields to add to report:

Search

- [x] Rejected Loans
- [ ] Amount Requested
- [ ] Application Date
- [ ] Loan Title
- [ ] Risk_Score
- [ ] Risk Score Bucket
- [ ] Debt-To-Income Ratio
- [x] DTI Bucket

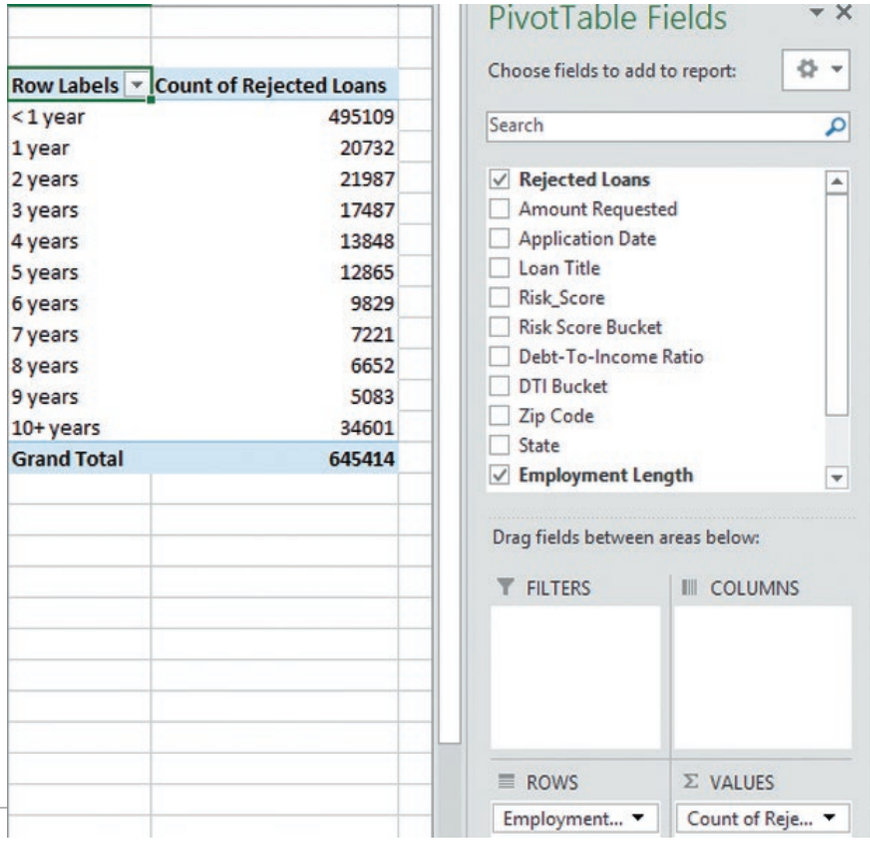
Drag fields between areas below:

FILTERS | COLUMNS

ROWS: DTI Bucket | VALUES: Count of Reje...

EXHIBIT 1-11
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).

Microsoft Excel, 2016



Row Labels	Count of Rejected Loans
< 1 year	495109
1 year	20732
2 years	21987
3 years	17487
4 years	13848
5 years	12865
6 years	9829
7 years	7221
8 years	6652
9 years	5083
10+ years	34601
Grand Total	645414

PivotTable Fields

Choose fields to add to report:

Search

- [x] Rejected Loans
- [ ] Amount Requested
- [ ] Application Date
- [ ] Loan Title
- [ ] Risk_Score
- [ ] Risk Score Bucket
- [ ] Debt-To-Income Ratio
- [ ] DTI Bucket
- [ ] Zip Code
- [ ] State
- [x] Employment Length

Drag fields between areas below:

FILTERS | COLUMNS

ROWS: Employment... | VALUES: Count of Reje...

EXHIBIT 1-12
LendingClub Declined Loan Applications by Employment Length (Years of Experience)

Microsoft Excel, 2016

EXHIBIT 1-13
Breakdown of Customer Credit Scores (or Risk Scores)

Source: Cafecredit.com

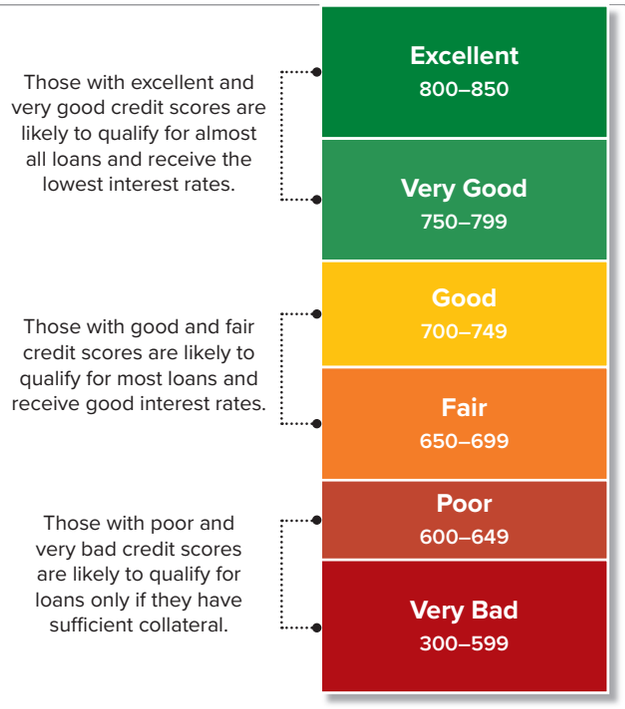
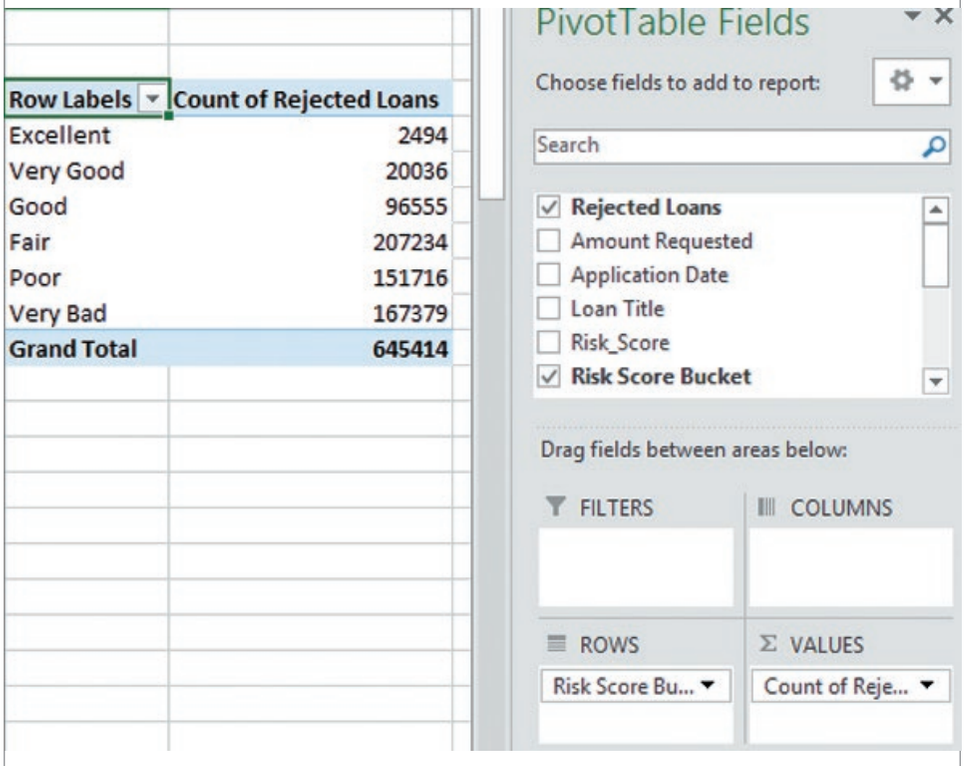


EXHIBIT 1-14
The Count of Lending-Club Rejected Loan Applications by Credit or Risk Score Classification Using PivotTable Analysis
(PivotTable shown here required manually sorting rows to get in proper order.)

Microsoft Excel, 2016

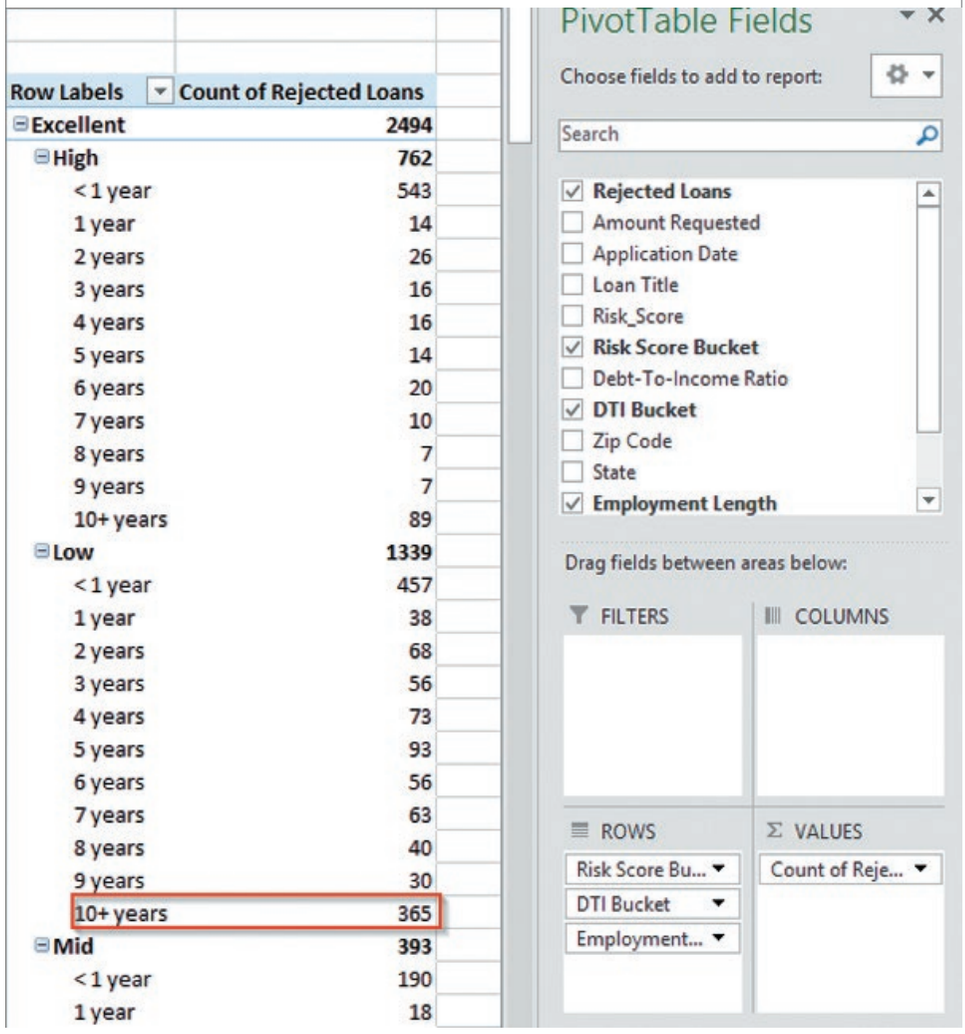


Row Labels	Count of Rejected Loans
Excellent	2494
Very Good	20036
Good	96555
Fair	207234
Poor	151716
Very Bad	167379
Grand Total	645414

Address and Refine Results

Now that we have completed the basic analysis, we can refine our analysis for greater insights. An example of this more refined analysis might be a further investigation of the rejected loans. For example, 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 1-15) 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.

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 1-16 shows the PivotTable analysis. The analysis shows those with excellent credit asked for a larger loan (16.2 percent of income)



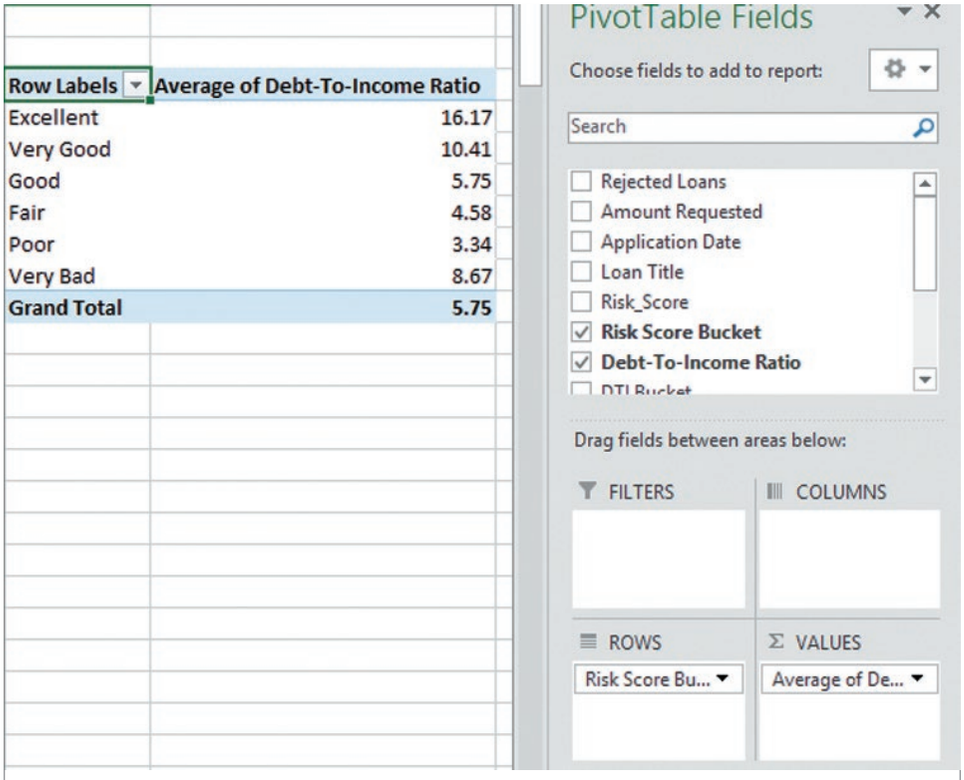
Row Labels	Count of Rejected Loans
Excellent	2494
High	762
< 1 year	543
1 year	14
2 years	26
3 years	16
4 years	16
5 years	14
6 years	20
7 years	10
8 years	7
9 years	7
10+ years	89
Low	1339
< 1 year	457
1 year	38
2 years	68
3 years	56
4 years	73
5 years	93
6 years	56
7 years	63
8 years	40
9 years	30
10+ years	365
Mid	393
< 1 year	190
1 year	18

EXHIBIT 1-15
The Count of LendingClub Declined Loan Applications by Credit (or Risk Score), Debt-to-Income (DTI Bucket), and Employment Length Using PivotTable Analysis (Highlighting Added)

Microsoft Excel, 2016

EXHIBIT 1-16
The Average Debt-to-Income Ratio (Shown as a Percentage) by Credit (Risk) Score for LendingClub Declined Loan Applications Using PivotTable Analysis

Microsoft Excel, 2016



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.

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 or reject a loan to a potential borrower. We can communicate these insights either by showing the PivotTables or simply stating what three of the determinants are. What is the most effective communication? Just showing the PivotTables themselves, showing a graph of the results, or simply sharing the names of these three determinants to the decision makers? Knowing the decision makers and how they like to receive this information will help the analyst determine how to communicate insights.

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 were from 2007 to 2012. We could make our predictions for subsequent years based on what we had found in the past and then test to 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.

PROGRESS CHECK

11. Lenders often use the data item of whether a potential borrower rents or owns their house. Beyond the three characteristics of rejected loans analyzed in this section, do you believe this data item would be an important determinant of rejected loans? Defend your answer.
12. Performing your own analysis, download the rejected loans dataset titled "DAA Chapter 1-1 Data" and perform an Excel PivotTable analysis by state (including the District of Columbia) 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?
13. Performing your own analysis, download the rejected loans dataset titled "DAA Chapter 1-1 Data" and run an Excel PivotTable by risk (or credit) score classification and DTI bucket to determine the number of (or percentage of) rejected loans requested by those rated as having an excellent credit score.

Summary

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 to address accounting questions. 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 that might be used to address different accounting questions. We also discussed the data analytic skills needed by analytic-minded accountants.

We followed this up using a hands-on example of the IMPACT model, namely what are the characteristics of rejected loans at **LendingClub**. We performed this analysis using various filtering and PivotTable tasks.

- With data all around us, businesses and accountants are looking at Data Analytics to extract the value that the data might possess. **(LO 1-1, 1-2, 1-3)**
- 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. **(LO 1-3)**
- The IMPACT cycle is a means of performing Data Analytics that goes all the way from identifying the question, to mastering the data, to performing data analyses and communicating and tracking results. It is recursive in nature, suggesting that as questions are addressed, new, more refined questions may emerge that can be addressed in a similar way. **(LO 1-4)**
- 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. **(LO 1-4)**
- Data analytic skills needed by analytic-minded accountants are specified and are consistent with the IMPACT cycle, including the following: **(LO 1-5)**
  - Developed analytics mindset.
  - Data scrubbing and data preparation.

- Data quality.
- Descriptive data analysis.
- Data analysis through data manipulation.
- Statistical data analysis competency.
- Data visualization and data reporting.

- We showed an example of the IMPACT cycle using **LendingClub** data regarding rejected loans to illustrate the steps of the IMPACT cycle. **(LO 1-6)**

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

clustering (***11***) A data approach that attempts to divide individuals (like customers) into groups (or clusters) in a useful or meaningful way.

co-occurrence grouping (***11***) 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 (***19***) Centralized repository of descriptions for all of the data attributes of the dataset.

data reduction (***12***) 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 (***12***) A data approach that attempts to predict a relationship between two data items.

predictor (or independent or explanatory) variable (***11***) A variable that predicts or explains another variable, typically called a predictor or independent variable.

profiling (***11***) 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.).

regression (***11***) A data approach that attempts to estimate or predict, for each unit, the numerical value of some variable using some type of statistical model.

response (or dependent) variable (***10***) A variable that responds to, or is dependent on, another.

similarity matching (***11***) A data approach that attempts to identify similar individuals based on data known about them.

structured data (***4***) Data that are organized and reside in a fixed field with a record or a file. Such data are generally contained in a relational database or spreadsheet and are readily searchable by search algorithms.

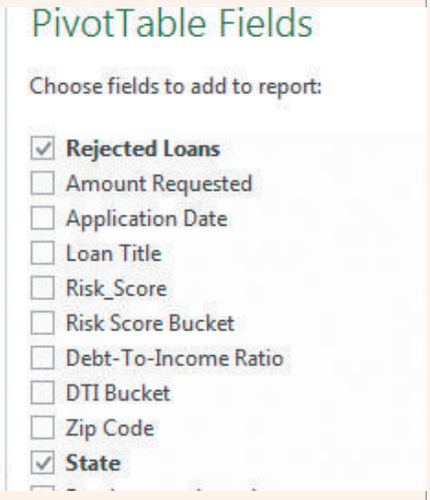
unstructured data (***4***) Data that do not adhere to a predefined data model in a tabular format.

ANSWERS TO PROGRESS CHECKS

1. The plethora of data alone does not necessarily translate into value. However, if we carefully analyze the data to help address critical business problems and questions, the data have the potential to create value.

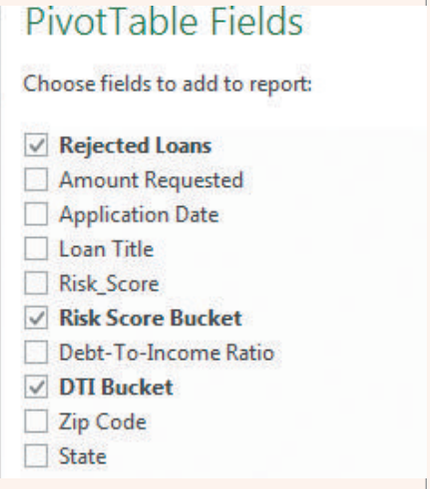
2. Banks frequently use credit scores from outside sources like **Experian**, **TransUnion**, and **Equifax** to evaluate creditworthiness of its customers. However, if they have access to all of their customer's banking information, Data Analytics would allow them to evaluate their customers' creditworthiness. 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 the size and stability of monthly income via the direct deposits. All of these combined, in addition to a credit score, might be used to assess creditworthiness if customers desire a loan. It might also give banks needed information for a marketing campaign to target potential creditworthy customers.
3. The brand manager at **Procter and Gamble** might use Data Analytics to see what is being said about **Procter and Gamble's** Tide Pods product on social media websites (e.g., Snapchat, Twitter, 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 laundry detergent products.
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)? Would it be cost-effective to just hire full-time employees instead of paying overtime? How much will costs increase just to pay for fringe benefits (health care, retirement, etc.) for new employees versus just paying existing employees for their overtime. All of these questions could be addressed by analyzing recent records explaining the use of overtime.
5. Management accounting and Data Analytics both (1) address questions asked by management, (2) find data to address those questions, (3) analyze the data, and (4) report the results to management. In all material respects, management accounting and Data Analytics are similar, if not identical.
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, and so on.
8. The data reduction approach might help auditors spend more time and effort on the most risky 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. According to the "magic quadrant," the software tools represented by the Microsoft and Tableau Tracks are considered innovative because they lead the market in the "ability to execute" and "completeness of vision" dimensions.
10. Having Tableau software tools available on both the Mac and Windows computers gives the analyst needed flexibility that is not available for the Microsoft Track, which are fully available only on Windows computers.
11. The use of the data item whether a potential borrower owns or rents their house would be expected to complement the risk score, debt levels (DTI bucket), and length of employment, since it can give a potential lender additional data on the financial position and financial obligations (mortgage or rent payments) of the borrower.
12. 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).
13. 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
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



Microsoft Excel, 2016

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



Microsoft Excel, 2016

Multiple Choice Questions

1. **(LO 1-1)** Big Data is often described by the four Vs, or
   a. volume, velocity, veracity, and variability.
   b. volume, velocity, veracity, and variety.
   c. volume, volatility, veracity, and variability.
   d. variability, velocity, veracity, and variety.
2. **(LO 1-4)** Which data approach 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. **(LO 1-4)** Which data approach attempts to identify similar individuals based on data known about them?
   a. Classification
   b. Regression
   c. Similarity matching
   d. Data reduction
4. **(LO 1-4)** Which data approach attempts to predict connections between two data items?
   a. Profiling
   b. Classification
   c. Link prediction
   d. Regression
5. **(LO 1-6)** 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. **(LO 1-5)** Which skills were *not* emphasized that analytic-minded accountants should have?
   a. Developed an analytics mindset
   b. Data scrubbing and data preparation
   c. Classification of test approaches
   d. Statistical data analysis competency
7. **(LO 1-5)** In which areas were skills *not* emphasized for analytic-minded accountants?
   a. Data quality
   b. Descriptive data analysis
   c. Data visualization and data reporting
   d. Data and systems analysis and design
8. **(LO 1-4)** The IMPACT cycle includes all *except* the following steps:
   a. perform test plan.
   b. visualize the data.
   c. master the data.
   d. track outcomes.
9. **(LO 1-4)** The IMPACT cycle specifically includes all *except* the following steps:
   a. data preparation.
   b. communicate insights.
   c. address and refine results.
   d. perform test plan.
10. **(LO 1-1)** By the year 2024, the volume of data created, captured, copied, and consumed worldwide will be 149 _____.
    a. zettabytes
    b. petabytes
    c. exabytes
    d. yottabytes