GLOBAL EDITION

Analytics, Data Science, & Artificial Intelligence

Systems for Decision Support

ELEVENTH EDITION

Ramesh Sharda • Dursun Delen • Efraim Turban

ELEVENTH EDITION

GLOBAL EDITION

ANALYTICS, DATA SCIENCE, & ARTIFICIAL INTELLIGENCE

SYSTEMS FOR DECISION SUPPORT

Ramesh Sharda
Oklahoma State University

Dursun Delen
Oklahoma State University

Efraim Turban
University of Hawaii

Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • Sao Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Pearson Education Limited
KAO Two
KAO Park
Hockham Way
Harlow
Essex
CM17 9SR
United Kingdom

and Associated Companies throughout the world

Visit us on the World Wide Web at: www.pearsonglobaleditions.com

Authorized adaptation from the United States edition, entitled Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support, *11th Edition, ISBN 978-0-13-519201-6 by Ramesh Sharda, Dursun Delen, and Efraim Turban, published by Pearson Education © 2021.*

Acknowledgments of third-party content appear on the appropriate page within the text.

ISBN 10: 1-292-34155-6
ISBN 13: 978-1-292-34155-2
eBook ISBN 13: 978-1-292-34160-6

eBook formatted by Integra Software Service Pvt. Ltd.

BRIEF CONTENTS

CONTENTS

PREFACE

Analytics has become the technology driver of this decade. Companies such as IBM, Oracle, Microsoft, and others are creating new organizational units focused on analytics that help businesses become more effective and efficient in their operations. Decision makers are using data and computerized tools to make better decisions. Even consumers are using analytics tools directly or indirectly to make decisions on routine activities such as shopping, health care, and entertainment. The field of business analytics (BA)/data science (DS)/decision support systems (DSS)/business intelligence (BI) is evolving rapidly to become more focused on innovative methods and applications to utilize data streams that were not even captured some time back, much less analyzed in any significant way. New applications emerge daily in customer relationship management, banking and finance, health care and medicine, sports and entertainment, manufacturing and supply chain management, utilities and energy, and virtually every industry imaginable.

The theme of this revised edition is analytics, data science, and AI for enterprise decision support. In addition to traditional decision support applications, this edition expands the reader's understanding of the various types of analytics by providing examples, products, services, and exercises by means of introducing AI, machine-learning, robotics, chatbots, IoT, and Web/Internet-related enablers throughout the text. We highlight these technologies as emerging components of modern-day business analytics systems. AI technologies have a major impact on decision making by enabling autonomous decisions and by supporting steps in the process of making decisions. AI and analytics support each other by creating a synergy that assists decision making.

The purpose of this book is to introduce the reader to the technologies that are generally and collectively called *analytics* (or *business analytics*) but have been known by other names such as decision support systems, executive information systems, and business intelligence, among others. We use these terms interchangeably. This book presents the fundamentals of the methods, methodologies, and techniques used to design and develop these systems. In addition, we introduce the essentials of AI both as it relates to analytics as well as a standalone discipline for decision support.

We follow an EEE approach to introducing these topics: **Exposure, Experience**, and **Explore**. The book primarily provides **exposure** to various analytics techniques and their applications. The idea is that a student will be inspired to learn from how other organizations have employed analytics to make decisions or to gain a competitive edge. We believe that such **exposure** to what is being done with analytics and how it can be achieved is the key component of learning about analytics. In describing the techniques, we also introduce specific software tools that can be used for developing such applications. The book is not limited to any one software tool, so the students can **experience** these techniques using any number of available software tools. Specific suggestions are given in each chapter, but the student and the professor are able to use this book with many different software tools. Our book's companion Web site will include specific software guides, but students can gain **experience** with these techniques in many different ways. Finally, we hope that this **exposure** and **experience** enable and motivate readers to **explore** the potential of these techniques in their own domain. To facilitate such **exploration**, we include exercises that direct them to Teradata University Network and other sites as well that include team-oriented exercises where appropriate. In our own teaching experience, projects undertaken in the class facilitate such **exploration** after the students have been **exposed** to the myriad of applications and concepts in the book and they have **experienced** specific software introduced by the professor.

This edition of the book can be used to offer a one-semester overview course on analytics, which covers most or all of the topics/chapters included in the book. It can also be used to teach two consecutive courses. For example, one course could focus on the overall analytics coverage. It could cover selective sections of Chapters 1 and 3–9. A second course could focus on artificial intelligence and emerging technologies as the enablers of modern-day analytics as a subsequent course to the first course. This second course could cover portions of Chapters 1, 2, 6, 9, and 10–14. The book can be used to offer managerial-level exposure to applications and techniques as noted in the previous paragraph, but it also includes sufficient technical details in selected chapters to allow an instructor to focus on some technical methods and hands-on exercises.

Most of the specific improvements made in this eleventh edition concentrate on three areas: reorganization, content update/upgrade (including AI, machine-learning, chatbots, and robotics as enablers of analytics), and a sharper focus. Despite the many changes, we have preserved the comprehensiveness and user friendliness that have made the textbook a market leader in the last several decades. We have also optimized the book's size and content by eliminating older and redundant material and by adding and combining material that is parallel to the current trends and is also demanded by many professors. Finally, we present accurate and updated material that is not available in any other text. We next describe the changes in the eleventh edition.

The book is supported by a Web site (**www.pearsonglobaleditions.com**). We provide links to additional learning materials and software tutorials through a special section of the book Web site.

WHAT'S NEW IN THE ELEVENTH EDITION?

With the goal of improving the text and making it current with the evolving technology trends, this edition marks a major reorganization to better reflect on the current focus on analytics and its enabling technologies. The last three editions transformed the book from the traditional DSS to BI and then from BI to BA and fostered a tight linkage with the Teradata University Network (TUN). This edition is enhanced with new materials paralleling the latest trends in analytics including AI, machine learning, deep learning, robotics, IoT, and smart/robo-collaborative assisting systems and applications. The following summarizes the major changes made to this edition.

- ***New organization.*** The book is now organized around two main themes: (1) presentation of motivations, concepts, methods, and methodologies for different types of analytics (focusing heavily on predictive and prescriptive analytic), and (2) introduction and due coverage of new technology trends as the enablers of the modern-day analytics such as AI, machine learning, deep learning, robotics, IoT, smart/robo-collaborative assisting systems, etc. Chapter 1 provides an introduction to the journey of decision support and enabling technologies. It begins with a brief overview of the classical decision making and decision support systems. Then it moves to business intelligence, followed by an introduction to analytics, Big Data, and AI. We follow that with a deeper introduction to artificial intelligence in Chapter 2. Because data is fundamental to any analysis, Chapter 3 introduces data issues as well as descriptive analytics including statistical concepts and visualization. An online chapter covers data warehousing processes and fundamentals for those who like to dig deeper into these issues. The next section covers predictive analytics and machine learning. Chapter 4 provides an introduction to data mining applications and the data mining process. Chapter 5 introduces many of the common data mining techniques: classification, clustering, association mining, and so forth. Chapter 6 includes coverage of deep learning and cognitive computing. Chapter 7 focuses on text mining applications as well as Web analytics, including social media analytics, sentiment analysis, and other related topics. The following section brings the "data science" angle to a further depth. Chapter 8 covers prescriptive analytics including optimization and simulation. Chapter 9 includes more details of Big Data analytics. It also includes introduction to cloud-based analytics as well as location analytics. The next section covers Robotics, social networks, AI, and the Internet of Things (IoT). Chapter 10 introduces robots in business and consumer applications and also studies the future impact of such devices on society. Chapter 11 focuses on collaboration systems, crowdsourcing, and social networks. Chapter 12 reviews personal assistants, chatbots, and the exciting developments in this space. Chapter 13 studies IoT and its potential in decision support and a smarter society. The ubiquity of wireless and GPS devices and other sensors is resulting in the creation of massive new databases and unique applications. Finally, Chapter 14 concludes with a brief discussion of security, privacy, and societal dimensions of analytics and AI.

We should note that several chapters included in this edition have been available in the following companion book: *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*, 4th Edition, Pearson (2018) (Hereafter referred to as BI4e). The structure and contents of these chapters have been updated somewhat before inclusion in this edition of the book, but the changes are more significant in the chapters marked as new. Of course, several of the chapters that came from BI4e were not included in previous editions of this book.

- ***New chapters.*** The following chapters have been added:

Chapter 2 "Artificial Intelligence: Concepts, Drivers, Major Technologies, and Business Applications" This chapter covers the essentials of AI, outlines its benefits, compares it with humans' intelligence, and describes the content of the field. Example applications in accounting, finance, human resource management, marketing and CRM, and production-operation management illustrate the benefits to business (100% new material)

Chapter 6, "Deep Learning and Cognitive Computing" This chapter covers the generation of machine learning technique, deep learning as well as the increasingly more popular AI topic, cognitive computing. It is an almost entirely new chapter (90% new material).

Chapter 10, "Robotics: Industrial and Consumer Applications" This chapter introduces many robotics applications in industry and for consumers and concludes with impacts of such advances on jobs and some legal ramifications (100% new material).

Chapter 12, "Knowledge Systems: Expert Systems, Recommenders, Chatbots, Virtual Personal Assistants, and Robo Advisors" This new chapter concentrates on different types of knowledge systems. Specifically, we cover new generations of expert systems and recommenders, chatbots, enterprise chatbots, virtual personal assistants, and robo-advisors (95% new).

Chapter 13, "The Internet of Things as a Platform for Intelligent Applications" This new chapter introduces IoT as an enabler to analytics and AI applications. The following technologies are described in detail: smart homes and appliances, smart cities (including factories), and autonomous vehicles (100% new).

Chapter 14, "Implementation Issues: From Ethics and Privacy to Organizational and Societal Impacts" This mostly new chapter deals with implementation issues of intelligent systems (including analytics). The major issues covered are protection of privacy, intellectual property, ethics, technical issues (e.g., integration and security) and administrative issues. We also cover the impact of these technologies on organizations and people and specifically deal with the impact on work and jobs. Special attention is given to possible unintended impacts of analytics and AI (robots). Then we look at relevant technology trends and conclude with an assessment of the future of analytics and AI (85% new).

- ***Streamlined coverage.*** We have optimized the book size and content by adding a lot of new material to cover new and cutting-edge analytics and AI trends and technologies while eliminating most of the older, less-used material. We use a dedicated Web site for the textbook to provide some of the older material as well as updated content and links.
- ***Revised and updated content.*** Several chapters have new opening vignettes that are based on recent stories and events. In addition, application cases throughout the book are new or have been updated to include recent examples of applications of a specific technique/model. These application case stories now include suggested questions for discussion to encourage class discussion as well as further exploration of the specific case and related materials. New Web site links have been added throughout the book. We also deleted many older product links and references. Finally, most chapters have new exercises, Internet assignments, and discussion questions throughout. The specific changes made to each chapter are as follows: Chapters 1, 3–5, and 7–9 borrow material from BI4e to a significant degree.

Chapter 1, "Overview of Business Intelligence, Analytics, Data Science, and Artificial Intelligence: Systems for Decision Support" This chapter includes some material from DSS10e Chapters 1 and 2, but includes several new application cases, entirely new material on AI, and of course, a new plan for the book (about 50% new material).

Chapter 3, "Nature of Data, Statistical Modeling, and Visualization"
- 75% new content.
- Most of the content related to nature of data and statistical analysis is new.
- New opening case.
- Mostly new cases throughout.

Chapter 4, "Data Mining Process, Methods, and Algorithms"
- 25% of the material is new.
- Some of the application cases are new.

Chapter 5, "Machine Learning Techniques for Predictive Analytics"
- 40% of the material is new.
- New machine-learning methods: naïve Bayes, Bayesian networks, and ensemble modeling.
- Most of the cases are new.

Chapter 7, "Text Mining, Sentiment Analysis, and Social Analytics"
- 25% of the material is new.
- Some of the cases are new.

Chapter 8, "Prescriptive Analytics: Optimization and Simulation"
- Several new optimization application exercises are included.
- A new application case is included.
- 20% of the material is new.

Chapter 9, "Big Data, Cloud Computing, and Location Analytics: Concepts and Tools" This material has bene updated substantially in this chapter to include greater coverage of stream analytics. It also updates material from Chapters 7 and 8 from BI4e (50% new material).

Chapter 11, "Group Decision Making, Collaborative Systems, and AI Support" The chapter is completely revised, regrouping group decision support. New topics include collective and collaborative intelligence, crowdsourcing, swarm AI, and AI support of all related activities (80% new material).

We have retained many of the enhancements made in the last editions and updated the content. These are summarized next:

- ***Links to Teradata University Network (TUN).*** Most chapters include new links to TUN (**teradatauniversitynetwork.com**). We encourage the instructors to register and join **teradatauniversitynetwork.com** and explore the various content available through the site. The cases, white papers, and software exercises available through TUN will keep your class fresh and timely.
- ***Book title.*** As is already evident, the book's title and focus have changed.
- ***Software support.*** The TUN Web site provides software support at no charge. It also provides links to free data mining and other software. In addition, the site provides exercises in the use of such software.

THE SUPPLEMENT PACKAGE: WWW.PEARSONGLOBALEDITIONS.COM

A comprehensive and flexible technology-support package is available to enhance the teaching and learning experience. The following instructor and student supplements are available on the book's Web site, **www.pearsonglobaleditions.com**:

- ***Instructor's Manual.*** The Instructor's Manual includes learning objectives for the entire course and for each chapter, answers to the questions and exercises at the end of each chapter, and teaching suggestions (including instructions for projects). The Instructor's Manual is available on the secure faculty section of **www.pearsonglobaleditions.com**
- ***Test Item File and TestGen Software.*** The Test Item File is a comprehensive collection of true/false, multiple-choice, fill-in-the-blank, and essay questions. The questions are rated by difficulty level, and the answers are referenced by book page number. The Test Item File is available in Microsoft Word and in TestGen. Pearson Education's test-generating software is available from **www.pearsonglobaleditions.com**. The software is PC/MAC compatible and preloaded with all of the Test Item File questions. You can manually or randomly view test questions and drag-and-drop to create a test. You can add or modify test-bank questions as needed. Our TestGens are converted for use in BlackBoard, WebCT, Moodle, D2L, and Angel. These conversions can be found on **www.pearsonglobaleditions.com**. The TestGen is also available in Respondus and can be found on **www.respondus.com**.
- ***PowerPoint slides.*** PowerPoint slides are available that illuminate and build on key concepts in the text. Faculty can download the PowerPoint slides from **www.pearsonglobaleditions.com**.

ACKNOWLEDGMENTS

Many individuals have provided suggestions and criticisms since the publication of the first edition of this book. Dozens of students participated in class testing of various chapters, software, and problems and assisted in collecting material. It is not possible to name everyone who participated in this project, but our thanks go to all of them. Certain individuals made significant contributions, and they deserve special recognition.

First, we appreciate the efforts of those individuals who provided formal reviews of the first through eleventh editions (school affiliations as of the date of review):

Robert Blanning, Vanderbilt University
Ranjit Bose, University of New Mexico

Warren Briggs, Suffolk University
Lee Roy Bronner, Morgan State University
Charles Butler, Colorado State University
Sohail S. Chaudry, University of Wisconsin–La Crosse
Kathy Chudoba, Florida State University
Wingyan Chung, University of Texas
Woo Young Chung, University of Memphis
Paul “Buddy” Clark, South Carolina State University
Pi’Sheng Deng, California State University–Stanislaus
Joyce Elam, Florida International University
Kurt Engemann, Iona College
Gary Farrar, Jacksonville University
George Federman, Santa Clara City College
Jerry Fjermestad, New Jersey Institute of Technology
Joey George, Florida State University
Paul Gray, Claremont Graduate School
Orv Greynholds, Capital College (Laurel, Maryland)
Martin Grossman, Bridgewater State College
Ray Jacobs, Ashland University
Leonard Jessup, Indiana University
Jeffrey Johnson, Utah State University
Jahangir Karimi, University of Colorado Denver
Saul Kassicieh, University of New Mexico
Anand S. Kunnathur, University of Toledo
Shao-ju Lee, California State University at Northridge
Yair Levy, Nova Southeastern University
Hank Lucas, New York University
Jane Mackay, Texas Christian University
George M. Marakas, University of Maryland
Dick Mason, Southern Methodist University
Nick McGaughey, San Jose State University
Ido Millet, Pennsylvania State University–Erie
Benjamin Mittman, Northwestern University
Larry Moore, Virginia Polytechnic Institute and State University
Simitra Mukherjee, Nova Southeastern University
Marianne Murphy, Northeastern University
Peter Mykytyn, Southern Illinois University
Natalie Nazarenko, SUNY College at Fredonia
David Olson, University of Nebraska
Souren Paul, Southern Illinois University
Joshua Pauli, Dakota State University
Roger Alan Pick, University of Missouri–St. Louis
Saeed Piri, University of Oregon
W. “RP” Raghupaphi, California State University–Chico
Loren Rees, Virginia Polytechnic Institute and State University
David Russell, Western New England College
Steve Ruth, George Mason University
Vartan Safarian, Winona State University
Glenn Shephard, San Jose State University
Jung P. Shim, Mississippi State University
Meenu Singh, Murray State University
Randy Smith, University of Virginia

James T. C. Teng, University of South Carolina
John VanGigch, California State University at Sacramento
David Van Over, University of Idaho
Paul J. A. van Vliet, University of Nebraska at Omaha
B. S. Vijayaraman, University of Akron
Howard Charles Walton, Gettysburg College
Diane B. Walz, University of Texas at San Antonio
Paul R. Watkins, University of Southern California
Randy S. Weinberg, Saint Cloud State University
Jennifer Williams, University of Southern Indiana
Selim Zaim, Sehir University
Steve Zanakis, Florida International University
Fan Zhao, Florida Gulf Coast University
Hamed Majidi Zolbanin, Ball State University

Several individuals contributed material to the text or the supporting material. For this new edition, assistance from the following students and colleagues is gratefully acknowledged: Behrooz Davazdahemami, Bhavana Baheti, Varnika Gottipati, and Chakradhar Pathi (all of Oklahoma State University). Prof. Rick Wilson contributed some examples and new exercise questions for Chapter 8. Prof. Pankush Kalgotra (Auburn University) contributed the new streaming analytics tutorial in Chapter 9. Other contributors of materials for specific application stories are identified as sources in the respective sections. Susan Baskin, Imad Birouty, Sri Raghavan, and Yenny Yang of Teradata provided special help in identifying new TUN content for the book and arranging permissions for the same.

Many other colleagues and students have assisted us in developing previous editions or the recent edition of the companion book from which some of the content has been adapted in this revision. Some of that content is still included this edition. Their assistance and contributions are acknowledged as well in chronological order. Dr. Dave Schrader contributed the sports examples used in Chapter 1. These will provide a great introduction to analytics. We also thank INFORMS for their permission to highlight content from *Interfaces*. We also recognize the following individuals for their assistance in developing Previous edition of the book: Pankush Kalgotra, Prasoon Mathur, Rupesh Agarwal, Shubham Singh, Nan Liang, Jacob Pearson, Kinsey Clemmer, and Evan Murlette (all of Oklahoma State University). Their help for BI 4e is gratefully acknowledged. The Teradata Aster team, especially Mark Ott, provided the material for the opening vignette for Chapter 9. Dr. Brian LeClaire, CIO of Humana Corporation led with contributions of several real-life healthcare case studies developed by his team at Humana. Abhishek Rathi of vCreaTek contributed his vision of analytics in the retail industry. In addition, the following former PhD students and research colleagues of ours have provided content or advice and support for the book in many direct and indirect ways: Asil Oztekin, University of Massachusetts-Lowell; Enes Eryarsoy, Sehir University; Hamed Majidi Zolbanin, Ball State University; Amir Hassan Zadeh, Wright State University; Supavich (Fone) Pengnate, North Dakota State University; Christie Fuller, Boise State University; Daniel Asamoah, Wright State University; Selim Zaim, Istanbul Technical University; and Nihat Kasap, Sabanci University. Peter Horner, editor of *OR/MS Today*, allowed us to summarize new application stories from *OR/MS Today* and *Analytics Magazine*. We also thank INFORMS for their permission to highlight content from *Interfaces*. Assistance from Natraj Ponna, Daniel Asamoah, Amir Hassan-Zadeh, Kartik Dasika, and Angie Jungermann (all of Oklahoma State University) is gratefully acknowledged for DSS 10th edition. We also acknowledge Jongswas Chongwatpol (NIDA, Thailand) for the material on SIMIO software, and Kazim Topuz (University of Tulsa) for his contributions to the Bayesian networks section in

Chapter 5. For other previous editions, we acknowledge the contributions of Dave King (a technology consultant and former executive at JDA Software Group, Inc.) and Jerry Wagner (University of Nebraska–Omaha). Major contributors for earlier editions include Mike Goul (Arizona State University) and Leila A. Halawi (Bethune-Cookman College), who provided material for the chapter on data warehousing; Christy Cheung (Hong Kong Baptist University), who contributed to the chapter on knowledge management; Linda Lai (Macau Polytechnic University of China); Lou Frenzel, an independent consultant whose books *Crash Course in Artificial Intelligence and Expert Systems* and *Understanding of Expert Systems* (both published by Howard W. Sams, New York, 1987) provided material for the early editions; Larry Medsker (American University), who contributed substantial material on neural networks; and Richard V. McCarthy (Quinnipiac University), who performed major revisions in the seventh edition.

Previous editions of the book have also benefited greatly from the efforts of many individuals who contributed advice and interesting material (such as problems), gave feedback on material, or helped with class testing. These include Warren Briggs (Suffolk University), Frank DeBalough (University of Southern California), Mei-Ting Cheung (University of Hong Kong), Alan Dennis (Indiana University), George Easton (San Diego State University), Janet Fisher (California State University, Los Angeles), David Friend (Pilot Software, Inc.), the late Paul Gray (Claremont Graduate School), Mike Henry (OSU), Dustin Huntington (Exsys, Inc.), Subramanian Rama Iyer (Oklahoma State University), Elena Karahanna (The University of Georgia), Mike McAulliffe (The University of Georgia), Chad Peterson (The University of Georgia), Neil Rabjohn (York University), Jim Ragusa (University of Central Florida), Alan Rowe (University of Southern California), Steve Ruth (George Mason University), Linus Schrage (University of Chicago), Antonie Stam (University of Missouri), Late Ron Swift (NCR Corp.), Merril Warkentin (then at Northeastern University), Paul Watkins (The University of Southern California), Ben Mortagy (Claremont Graduate School of Management), Dan Walsh (Bellcore), Richard Watson (The University of Georgia), and the many other instructors and students who have provided feedback.

Several vendors cooperated by providing development and/or demonstration software: Dan Fylstra of Frontline Systems, Gregory Piatetsky-Shapiro of **KDNuggets.com**, Logic Programming Associates (UK), Gary Lynn of NeuroDimension Inc. (Gainesville, Florida), Palisade Software (Newfield, New York), Jerry Wagner of Planners Lab (Omaha, Nebraska), Promised Land Technologies (New Haven, Connecticut), Salford Systems (La Jolla, California), Gary Miner of StatSoft, Inc. (Tulsa, Oklahoma), Ward Systems Group, Inc. (Frederick, Maryland), Idea Fisher Systems, Inc. (Irving, California), and Wordtech Systems (Orinda, California).

Special thanks to the Teradata University Network and especially to Hugh Watson, Michael Goul, and Susan Baskin, Program Director, for their encouragement to tie this book with TUN and for providing useful material for the book.

Many individuals helped us with administrative matters and editing, proofreading, and preparation. The project began with Jack Repcheck (a former Macmillan editor), who initiated this project with the support of Hank Lucas (New York University). Jon Outland assisted with the supplements.

Finally, the Pearson team is to be commended: Executive Editor Samantha Lewis who orchestrated this project; the copyeditors; and the production team, Faraz Sharique Ali at Pearson, and Gowthaman and staff at Integra Software Services, who transformed the manuscript into a book.

We would like to thank all these individuals and corporations. Without their help, the creation of this book would not have been possible. We want to specifically acknowledge the contributions of previous coauthors Janine Aronson, David King, and T. P. Liang, whose original contributions constitute significant components of the book.

R.S.

D.D.

E.T.

GLOBAL EDITION ACKNOWLEDGMENTS

Pearson would like to thank the following people for their work on the Global Edition:

Contributor

Stefania Paladini, Birmingham City University

Reviewers

Xavier Pierron, Edinburgh Napier University
Qizhang Liu, National University of Singapore
Krish Saha, Birmingham City University

Note that Web site URLs are dynamic. As this book went to press, we verified that all the cited Web sites were active and valid. Web sites to which we refer in the text sometimes change or are discontinued because companies change names, are bought or sold, merge, or fail. Sometimes Web sites are down for maintenance, repair, or redesign. Most organizations have dropped the initial "www" designation for their sites, but some still use it. If you have a problem connecting to a Web site that we mention, please be patient and simply run a Web search to try to identify the new site. Most times, the new site can be found quickly. Some sites also require a free registration before allowing you to see the content. We apologize in advance for this inconvenience.

ABOUT THE AUTHORS

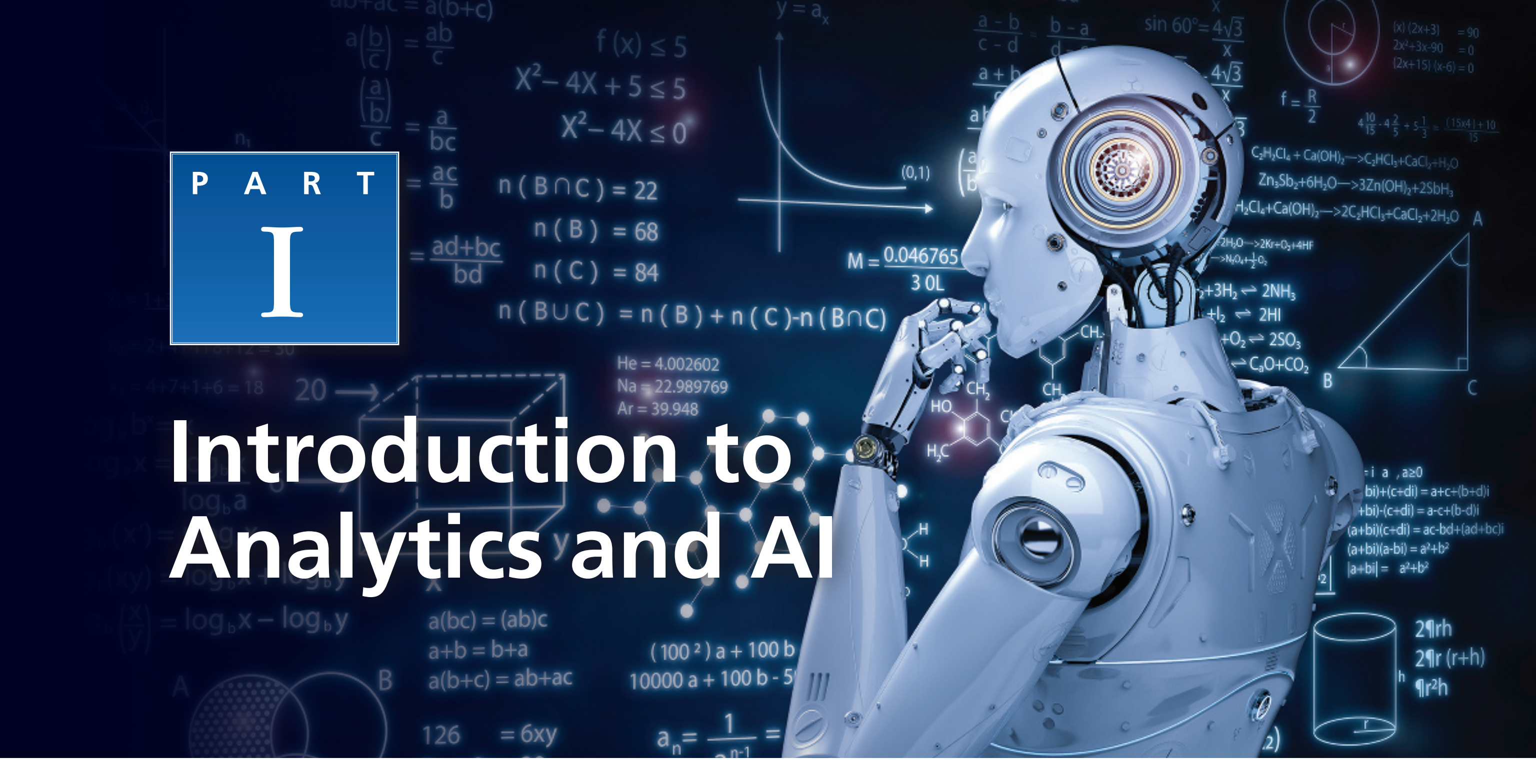
Ramesh Sharda (M.B.A., Ph.D., University of Wisconsin–Madison) is the Vice Dean for Research and Graduate Programs, Watson/ConocoPhillips Chair and a Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University. His research has been published in major journals in management science and information systems including *Management Science, Operations Research, Information Systems Research, Decision Support Systems, Decision Sciences Journal, EJIS, JMIS, Interfaces, INFORMS Journal on Computing, ACM Data Base*, and many others. He is a member of the editorial boards of journals such as the *Decision Support Systems, Decision Sciences*, and *ACM Database*. He has worked on many sponsored research projects with government and industry, and has also served as consultants to many organizations. He also serves as the Faculty Director of Teradata University Network. He received the 2013 INFORMS Computing Society HG Lifetime Service Award, and was inducted into Oklahoma Higher Education Hall of Fame in 2016. He is a Fellow of INFORMS.

Dursun Delen (Ph.D., Oklahoma State University) is the Spears and Patterson Chairs in Business Analytics, Director of Research for the Center for Health Systems Innovation, and Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University (OSU). Prior to his academic career, he worked for a privately owned research and consultancy company, Knowledge Based Systems Inc., in College Station, Texas, as a research scientist for five years, during which he led a number of decision support and other information systems–related research projects funded by federal agencies such as DoD, NASA, NIST, and DOE. Dr. Delen's research has appeared in major journals including *Decision Sciences, Decision Support Systems, Communications of the ACM, Computers and Operations Research, Computers in Industry, Journal of Production Operations Management, Journal of American Medical Informatics Association, Artificial Intelligence in Medicine, Expert Systems with Applications*, among others. He has published eight books/textbooks and more than 100 peer-reviewed journal articles. He is often invited to national and international conferences for keynote addresses on topics related to business analytics, Big Data, data/text mining, business intelligence, decision support systems, and knowledge management. He served as the general co-chair for the 4th International Conference on Network Computing and Advanced Information Management (September 2–4, 2008, in Seoul, South Korea) and regularly serves as chair on tracks and mini-tracks at various business analytics and information systems conferences. He is the co-editor-in-chief for the *Journal of Business Analytics*, the area editor for Big Data and Business Analytics on the Journal of Business Research, and also serves as chief editor, senior editor, associate editor, and editorial board member on more than a dozen other journals. His consultancy, research, and teaching interests are in business analytics, data and text mining, health analytics, decision support systems, knowledge management, systems analysis and design, and enterprise modeling.

Efraim Turban (M.B.A., Ph.D., University of California, Berkeley) is a visiting scholar at the Pacific Institute for Information System Management, University of Hawaii. Prior to this, he was on the staff of several universities, including City University of Hong Kong; Lehigh University; Florida International University; California State University, Long

Beach; Eastern Illinois University; and the University of Southern California. Dr. Turban is the author of more than 110 refereed papers published in leading journals, such as *Management Science*, *MIS Quarterly*, and *Decision Support Systems*. He is also the author of 22 books, including *Electronic Commerce: A Managerial Perspective* and *Information Technology for Management*. He is also a consultant to major corporations worldwide. Dr. Turban's current areas of interest are Web-based decision support systems, digital commerce, and applied artificial intelligence.

This page is intentionally left blank



PART I

Introduction to Analytics and AI

CHAPTER

1

Overview of Business Intelligence, Analytics, Data Science, and Artificial Intelligence: Systems for Decision Support

LEARNING OBJECTIVES

- Understand the need for computerized support of managerial decision making
- Understand the development of systems for providing decision-making support
- Recognize the evolution of such computerized support to the current state of analytics/data science and artificial intelligence
- Describe the business intelligence (BI) methodology and concepts
- Understand the different types of analytics and review selected applications
- Understand the basic concepts of artificial intelligence (AI) and see selected applications
- Understand the analytics ecosystem to identify various key players and career opportunities

The business environment (climate) is constantly changing, and it is becoming more and more complex. Organizations, both private and public, are under pressures that force them to respond quickly to changing conditions and to be innovative in the way they operate. Such activities require organizations to be agile and to make frequent and quick strategic, tactical, and operational decisions, some of which are very complex. Making such decisions may require considerable amounts of relevant data, information, and knowledge. Processing these in the framework of the needed decisions must be done quickly, frequently in real time, and usually requires some computerized support. As technologies are evolving, many decisions are being automated, leading to a major impact on knowledge work and workers in many ways.

This book is about using business analytics and artificial intelligence (AI) as a computerized support portfolio for managerial decision making. It concentrates on the theoretical and conceptual foundations of decision support as well as on the commercial tools and techniques that are available. The book presents the fundamentals of the techniques and the manner in which these systems are constructed and used. We follow an EEE (*exposure, experience*, and *exploration*) approach to introducing these topics. The book primarily provides exposure to various analytics/AI techniques and their applications. The idea is that students will be inspired to learn from how various organizations have employed these technologies to make decisions or to gain a competitive edge. We believe that such exposure to what is being accomplished with analytics and that how it can be achieved is the key component of learning about analytics. In describing the techniques, we also give examples of specific software tools that can be used for developing such applications. However, the book is not limited to any one software tool, so students can experience these techniques using any number of available software tools. We hope that this *exposure* and *experience* enable and motivate readers to *explore* the potential of these techniques in their own domain. To facilitate such exploration, we include exercises that direct the reader to Teradata University Network (TUN) and other sites that include team-oriented exercises where appropriate. In our own teaching experience, projects undertaken in the class facilitate such exploration after students have been exposed to the myriad of applications and concepts in the book and they have experienced specific software introduced by the professor.

This introductory chapter provides an introduction to analytics and artificial intelligence as well as an overview of the book. The chapter has the following sections:

1.1 OPENING VIGNETTE: How Intelligent Systems Work for KONE Elevators and Escalators Company

KONE is a global industrial company (based in Finland) that manufactures mostly elevators and escalators and also services over 1.1 million elevators, escalators, and related equipment in several countries. The company employs over 50,000 people.

THE PROBLEM

Over 1 billion people use the elevators and escalators manufactured and serviced by KONE every day. If equipment does not work properly, people may be late to work, cannot get home in time, and may miss important meetings and events. So, KONE's objective is to minimize the downtime and users' suffering.

The company has over 20,000 technicians who are dispatched to deal with the elevators anytime a problem occurs. As buildings are getting higher (the trend in many places), more people are using elevators, and there is more pressure on elevators to handle the growing amount of traffic. KONE faced the responsibility to serve users smoothly and safely.

THE SOLUTION

KONE decided to use IBM Watson IoT Cloud platform. As we will see in Chapter 6, IBM installed cognitive abilities in buildings that make it possible to recognize situations and behavior of both people and equipment. The Internet of Things (IoT), as we will see in Chapter 13, is a platform that can connect millions of "things" together and to a central command that can manipulate the connected things. Also, the IoT connects sensors that are attached to KONE's elevators and escalators. The sensors collect information and data about the elevators (such as noise level) and other equipment in real time. Then, the IoT transfers to information centers via the collected data "cloud." There, analytic systems (IBM Advanced Analytic Engine) and AI process the collected data and predict things such as potential failures. The systems also identify the likely causes of problems and suggest potential remedies. Note the predictive power of IBM Watson Analytics (using machine learning, an AI technology described in Chapters 4–6) for finding problems before they occur.

The KONE system collects a significant amount of data that are analyzed for other purposes so that future design of equipment can be improved. This is because Watson Analytics offers a convenient environment for communication of and collaboration around the data. In addition, the analysis suggests how to optimize buildings and equipment operations. Finally, KONE and its customers can get insights regarding the financial aspects of managing the elevators.

KONE also integrates the Watson capabilities with Salesforce's service tools (Service Cloud Lightning and Field Service Lightning). This combination helps KONE to immediately respond to emergencies or soon-to-occur failures as quickly as possible, dispatching some of its 20,000 technicians to the problems' sites. Salesforce also provides superb customer relationship management (CRM). The people–machine communication, query, and collaboration in the system are in a natural language (an AI capability of Watson Analytics; see Chapter 6). Note that IBM Watson analytics includes two types of analytics: *predictive*, which predicts when failures may occur, and *prescriptive*, which recommends actions (e.g., preventive maintenance).

THE RESULTS

KONE has minimized downtime and shortened the repair time. Obviously, elevators/escalators users are much happier if they do not have problems because of equipment downtime, so they enjoy trouble-free rides. The prediction of "soon-to-happen" can save many problems for the equipment owners. The owners can also optimize the schedule of their own employees (e.g., cleaners and maintenance workers). All in all, the decision makers at both KONE and the buildings can make informed and better decisions. Some day in the future, robots may perform maintenance and repairs of elevators and escalators.

Note: This case is a sample of IBM Watson's success using its cognitive buildings capability. To learn more, we suggest you view the following YouTube videos: (1) **youtube.com/watch?v=6UPJHyiJft0** (1:31 min.) (2017); (2) **youtube.com/watch?v=EVbd3ejEXus** (2:49 min.) (2017).

Sources: Compiled from J. Fernandez. (2017, April). "A Billion People a Day. Millions of Elevators. No Room for Downtime." IBM developer Works Blog. **developer.ibm.com/dwblog/2017/kone-watson-video/** (accessed September 2018); H. Srikanthan. "KONE Improves 'People Flow' in 1.1 Million Elevators with IBM Watson IoT." Generis. **https://generisgp.com/2018/01/08/ibm-case-study-kone-corp/** (accessed September 2018); L. Slowey. (2017, February 16). "Look Who's Talking: KONE Makes Elevator Services Truly Intelligent with Watson IoT." IBM Internet of Things Blog. **ibm.com/blogs/internet-of-things/kone/** (accessed September 2018).

QUESTIONS FOR THE OPENING VIGNETTE

1. It is said that KONE is embedding intelligence across its supply chain and enables smarter buildings. Explain.
2. Describe the role of IoT in this case.
3. What makes IBM Watson a necessity in this case?
4. Check IBM Advanced Analytics. What tools were included that relate to this case?
5. Check IBM cognitive buildings. How do they relate to this case?

WHAT CAN WE LEARN FROM THIS VIGNETTE?

Today, intelligent technologies can embark on large-scale complex projects when they include AI combined with IoT. The capabilities of integrated intelligent platforms, such as IBM Watson, make it possible to solve problems that were economically and technologically unsolvable just a few years ago. The case introduces the reader to several of the technologies, including advanced analytics, sensors, IoT, and AI that are covered in this book. The case also points to the use of "cloud." The cloud is used to centrally process large amounts of information using analytics and AI algorithms, involving "things" in different locations. This vignette also introduces us to two major types of analytics: predictive analytics (Chapters 4–6) and prescriptive analytics (Chapter 8).

Several AI technologies are discussed: machine learning, natural language processing, computer vision, and prescriptive analysis.

The case is an example of *augmented intelligence* in which people and machines work together. The case illustrates the benefits to the vendor, the implementing companies, and their employees and to the users of the elevators and escalators.

1.2 CHANGING BUSINESS ENVIRONMENTS AND EVOLVING NEEDS FOR DECISION SUPPORT AND ANALYTICS

Decision making is one of the most important activities in organizations of all kind—probably the most important one. Decision making leads to the success or failure of organizations and how well they perform. Making decisions is getting difficult due to internal and external factors. The rewards of making appropriate decisions can be very high and so can the loss of inappropriate ones.

Unfortunately, it is not simple to make decisions. To begin with, there are several types of decisions, each of which requires a different decision-making approach. For example, De Smet et al. (2017) of McKinsey & Company management consultants classify organizational decision into the following four groups:

- Big-bet, high-risk decisions.
- Cross-cutting decisions, which are repetitive but high risk that require group work (Chapter 11).
- Ad hoc decisions that arise episodically.
- Delegated decisions to individuals or small groups.

Therefore, it is necessary first to understand the nature of decision making. For a comprehensive discussion, see (De Smet et al. 2017).

Modern business is full of uncertainties and rapid changes. To deal with these, organizational decision makers need to deal with ever-increasing and changing data. This book is about the technologies that can assist decision makers in their jobs.

Decision-Making Process

For years, managers considered decision making purely an art—a talent acquired over a long period through experience (i.e., learning by trial and error) and by using intuition. Management was considered an art because a variety of individual styles could be used in approaching and successfully solving the same types of managerial problems. These styles were often based on creativity, judgment, intuition, and experience rather than on systematic quantitative methods grounded in a scientific approach. However, recent research suggests that companies with top managers who are more focused on persistent work tend to outperform those with leaders whose main strengths are interpersonal communication skills. It is more important to emphasize methodical, thoughtful, analytical decision making rather than flashiness and interpersonal communication skills.

Managers usually make decisions by following a four-step process (we learn more about these in the next section):

1. Define the problem (i.e., a decision situation that may deal with some difficulty or with an opportunity).
2. Construct a model that describes the real-world problem.
3. Identify possible solutions to the modeled problem and evaluate the solutions.
4. Compare, choose, and recommend a potential solution to the problem.

A more detailed process is offered by Quain (2018), who suggests the following steps:

1. Understand the decision you have to make.
2. Collect all the information.
3. Identify the alternatives.
4. Evaluate the pros and cons.
5. Select the best alternative.
6. Make the decision.
7. Evaluate the impact of your decision.

We will return to this process in Section 1.3.

The Influence of the External and Internal Environments on the Process

To follow these decision-making processes, one must make sure that sufficient alternative solutions, including good ones, are being considered, that the consequences of using these alternatives can be reasonably predicted, and that comparisons are done properly. However, rapid changes in internal and external environments make such an evaluation process difficult for the following reasons:

- Technology, information systems, advanced search engines, and globalization result in more and more alternatives from which to choose.
- Government regulations and the need for compliance, political instability and terrorism, competition, and changing consumer demands produce more uncertainty, making it more difficult to predict consequences and the future.
  - **Political factors.** Major decisions may be influenced by both external and internal politics. An example is the 2018 trade war on tariffs.
  - **Economic factors.** These range from competition to the genera and state of the economy. These factors, both in the short and long run, need to be considered.

- **Sociological and psychological factors regarding employees and customers.** These need to be considered when changes are being made.
- **Environment factors.** The impact on the physical environment must be assessed in many decision-making situations.

Other factors include the need to make rapid decisions, the frequent and unpredictable changes that make trial-and-error learning difficult, and the potential costs of making mistakes that may be large.

These environments are growing more complex every day. Therefore, making decisions today is indeed a complex task. For further discussion, see Charles (2018). For how to make effective decisions under uncertainty and pressure, see Zane (2016).

Because of these trends and changes, it is nearly impossible to rely on a trial-and-error approach to management. Managers must be more sophisticated; they must use the new tools and techniques of their fields. Most of those tools and techniques are discussed in this book. Using them to support decision making can be extremely rewarding in making effective decisions. Further, many tools that are evolving impact even the very existence of several decision-making tasks that are being automated. This impacts future demand for knowledge workers and begs many legal and societal impact questions.

Data and Its Analysis in Decision Making

We will see several times in this book how an entire industry can employ analytics to develop reports on what is happening, predict what is likely to happen, and then make decisions to make the best use of the situation at hand. These steps require an organization to collect and analyze vast stores of data. In general, the amount of data doubles every two years. From traditional uses in payroll and bookkeeping functions, computerized systems are now used for complex managerial areas ranging from the design and management of automated factories to the application of analytical methods for the evaluation of proposed mergers and acquisitions. Nearly all executives know that information technology is vital to their business and extensively use these technologies.

Computer applications have moved from transaction-processing and monitoring activities to problem analysis and solution applications, and much of the activity is done with cloud-based technologies, in many cases accessed through mobile devices. Analytics and BI tools such as data warehousing, data mining, online analytical processing (OLAP), dashboards, and the use of cloud-based systems for decision support are the cornerstones of today's modern management. Managers must have high-speed, networked information systems (wired or wireless) to assist them with their most important task: making decisions. In many cases, such decisions are routinely being fully automated (see Chapter 2), eliminating the need for any managerial intervention.

Technologies for Data Analysis and Decision Support

Besides the obvious growth in hardware, software, and network capacities, some developments have clearly contributed to facilitating the growth of decision support and analytics technologies in a number of ways:

- **Group communication and collaboration.** Many decisions are made today by groups whose members may be in different locations. Groups can collaborate and communicate readily by using collaboration tools as well as the ubiquitous smartphones. Collaboration is especially important along the supply chain, where partners—all the way from vendors to customers—must share information. Assembling a group of decision makers, especially experts, in one place can be costly. Information systems can improve the collaboration process of a group and enable its members to be at different locations (saving travel costs). More critically, such supply chain collaboration permits manufacturers to know about the changing patterns of demand in near real time and thus react to marketplace changes faster. For a comprehensive coverage and the impact of AI, see Chapters 2, 10, and 14.

- **Improved data management.** Many decisions involve complex computations. Data for these can be stored in different databases anywhere in the organization and even possibly outside the organization. The data may include text, sound, graphics, and video, and these can be in different languages. Many times it is necessary to transmit data quickly from distant locations. Systems today can search, store, and transmit needed data quickly, economically, securely, and transparently. See Chapters 3 and 9 and the online chapter for details.
- **Managing giant data warehouses and Big Data.** Large data warehouses (DWs), like the ones operated by Walmart, contain huge amounts of data. Special methods, including parallel computing and Hadoop/Spark, are available to organize, search, and mine the data. The costs related to data storage and mining are declining rapidly. Technologies that fall under the broad category of Big Data have enabled massive data coming from a variety of sources and in many different forms, which allows a very different view of organizational performance that was not possible in the past. See Chapter 9 for details.
- **Analytical support.** With more data and analysis technologies, more alternatives can be evaluated, forecasts can be improved, risk analysis can be performed quickly, and the views of experts (some of whom may be in remote locations) can be collected quickly and at a reduced cost. Expertise can even be derived directly from analytical systems. With such tools, decision makers can perform complex simulations, check many possible scenarios, and assess diverse impacts quickly and economically.This, of course, is the focus of several chapters in the book. See Chapters 4–7.
- **Overcoming cognitive limits in processing and storing information.** The human mind has only a limited ability to process and store information. People sometimes find it difficult to recall and use information in an error-free fashion due to their cognitive limits. The term *cognitive limits* indicates that an individual's problem-solving capability is limited when a wide range of diverse information and knowledge is required. Computerized systems enable people to overcome their cognitive limits by quickly accessing and processing vast amounts of stored information. One way to overcome humans' cognitive limitations is to use AI support. For coverage of cognitive aspects, see Chapter 6.
- **Knowledge management.** Organizations have gathered vast stores of information about their own operations, customers, internal procedures, employee interactions, and so forth through the unstructured and structured communications taking place among various stakeholders. Knowledge management systems (KMS) have become sources of formal and informal support for decision making to managers, although sometimes they may not even be called *KMS*. Technologies such as text analytics and IBM Watson are making it possible to generate value from such knowledge stores. (See Chapters 6 and 12 for details.)
- **Anywhere, anytime support.** Using wireless technology, managers can access information anytime and from any place, analyze and interpret it, and communicate with those using it. This perhaps is the biggest change that has occurred in the last few years. The speed at which information needs to be processed and converted into decisions has truly changed expectations for both consumers and businesses. These and other capabilities have been driving the use of computerized decision support since the late 1960s, especially since the mid-1990s. The growth of mobile technologies, social media platforms, and analytical tools has enabled a different level of information systems (IS) to support managers. This growth in providing

data-driven support for any decision extends not just to managers but also to consumers. We will first study an overview of technologies that have been broadly referred to as BI. From there we will broaden our horizons to introduce various types of analytics.

- **Innovation and artificial intelligence.** Because of the complexities in the decision-making process discussed earlier and the environment surrounding the process, a more innovative approach is frequently need. A major facilitation of innovation is provided by AI. Almost every step in the decision-making process can be influenced by AI. AI is also integrated with analytics, creating synergy in making decisions (Section 1.8).

SECTION 1.2 REVIEW QUESTIONS

1. Why is it difficult to make organizational decisions?
2. Describe the major steps in the decision-making process.
3. Describe the major external environments that can impact decision making.
4. What are some of the key system-oriented trends that have fostered IS-supported decision making to a new level?
5. List some capabilities of information technologies that can facilitate managerial decision making.

1.3 DECISION-MAKING PROCESSES AND COMPUTERIZED DECISION SUPPORT FRAMEWORK

In this section, we focus on some classical decision-making fundamentals and in more detail on the decision-making process. These two concepts will help us ground much of what we will learn in terms of analytics, data science, and artificial intelligence.

Decision making is a process of choosing among two or more alternative courses of action for the purpose of attaining one or more goals. According to Simon (1977), managerial decision making is synonymous with the entire management process. Consider the important managerial function of planning. Planning involves a series of decisions: What should be done? When? Where? Why? How? By whom? Managers set goals, or plan; hence, planning implies decision making. Other managerial functions, such as organizing and controlling, also involve decision making.

Simon's Process: Intelligence, Design, and Choice

It is advisable to follow a systematic decision-making process. Simon (1977) said that this involves three major phases: intelligence, design, and choice. He later added a fourth phase: implementation. Monitoring can be considered a fifth phase—a form of feedback. However, we view monitoring as the *intelligence phase* applied to the *implementation phase*. Simon's model is the most concise and yet complete characterization of rational decision making. A conceptual picture of the decision-making process is shown in Figure 1.1. It is also illustrated as a decision support approach using modeling.

There is a continuous flow of activity from intelligence to design to choice (see the solid lines in Figure 1.1), but at any phase, there may be a return to a previous phase (feedback). Modeling is an essential part of this process. The seemingly chaotic nature of following a haphazard path from problem discovery to solution via decision making can be explained by these feedback loops.

The decision-making process starts with the **intelligence phase**; in this phase, the decision maker examines reality and identifies and defines the problem. *Problem ownership* is established as well. In the **design phase**, a model that represents the system is constructed. This is done by making assumptions that simplify reality and by writing down

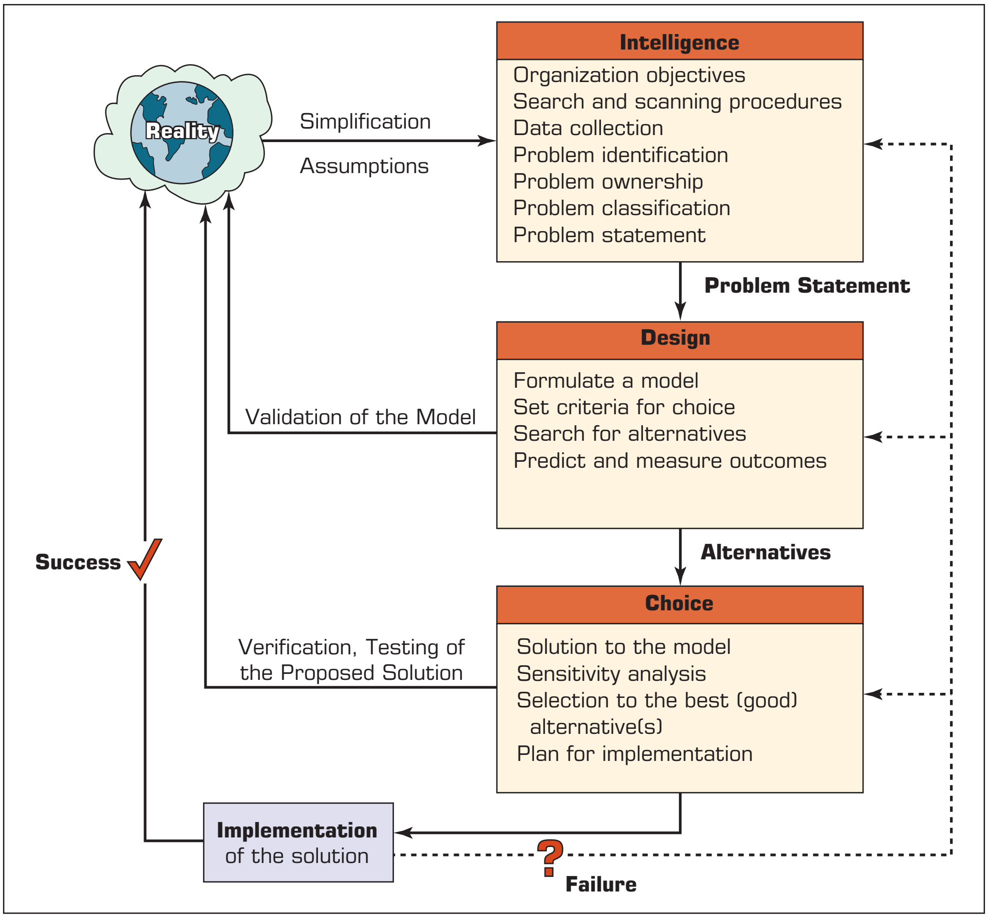


FIGURE 1.1 **The Decision-Making/Modeling Process.**

the relationships among all the variables. The model is then validated, and criteria are determined in a principle of choice for evaluation of the alternative courses of action that are identified. Often, the process of model development identifies alternative solutions and vice versa.

The **choice phase** includes the selection of a proposed solution to the model (not necessarily to the problem it represents). This solution is tested to determine its viability. When the proposed solution seems reasonable, we are ready for the last phase: implementation of the decision (not necessarily of a system). Successful implementation results in solving the real problem. Failure leads to a return to an earlier phase of the process. In fact, we can return to an earlier phase during any of the latter three phases. The decision-making situations described in the opening vignette follow Simon's four-phase model, as do almost all other decision-making situations.

The Intelligence Phase: Problem (or Opportunity) Identification

The intelligence phase begins with the identification of organizational goals and objectives related to an issue of concern (e.g., inventory management, job selection, lack of or incorrect Web presence) and determination of whether they are being met. Problems occur because of dissatisfaction with the status quo. Dissatisfaction is the result of a difference between what people desire (or expect) and what is occurring. In this first phase, a decision maker attempts to determine whether a problem exists, identify its symptoms, determine its magnitude, and explicitly define it. Often, what is described as a problem (e.g., excessive costs) may be only a symptom (i.e., measure) of a problem (e.g., improper inventory levels). Because real-world problems are usually complicated by many interrelated factors, it is sometimes difficult to distinguish between the symptoms and the real problem. New opportunities and problems certainly may be uncovered while investigating the causes of symptoms.

The existence of a problem can be determined by monitoring and analyzing the organization's productivity level. The measurement of productivity and the construction of a model are based on real data. The collection of data and the estimation of future data are among the most difficult steps in the analysis.

ISSUES IN DATA COLLECTION The following are some issues that may arise during data collection and estimation and thus plague decision makers:

- Data are not available. As a result, the model is made with and relies on potentially inaccurate estimates.
- Obtaining data may be expensive.
- Data may not be accurate or precise enough.
- Data estimation is often subjective.
- Data may be insecure.
- Important data that influence the results may be qualitative (soft).
- There may be too many data (i.e., information overload).
- Outcomes (or results) may occur over an extended period. As a result, revenues, expenses, and profits will be recorded at different points in time. To overcome this difficulty, a present-value approach can be used if the results are quantifiable.
- It is assumed that future data will be similar to historical data. If this is not the case, the nature of the change has to be predicted and included in the analysis.

When the preliminary investigation is completed, it is possible to determine whether a problem really exists, where it is located, and how significant it is. A key issue is whether an information system is reporting a problem or only the symptoms of a problem. For example, if reports indicate that sales are down, there is a problem, but the situation, no doubt, is symptomatic of the problem. It is critical to know the real problem. Sometimes it may be a problem of perception, incentive mismatch, or organizational processes rather than a poor decision model.

To illustrate why it is important to identify the problem correctly, we provide a classical example in Application Case 1.1.

Application Case 1.1 Making Elevators Go Faster!

This story has been reported in numerous places and has almost become a classic example to explain the need for problem identification. Ackoff (as cited in Larson, 1987) described the problem of managing complaints about slow elevators in a tall hotel tower. After trying many solutions for reducing the complaint—staggering elevators to go to different floors, adding operators, and so on—the management determined that the real problem was not about the *actual* waiting time but rather the *perceived* waiting time. So the solution was to install full-length mirrors on elevator doors on each floor. As Hesse and Woolsey (1975) put it, "The women would look at themselves in the mirrors and make adjustments, while the men would look at the women, and before they knew it, the elevator was there." By reducing the perceived waiting time, the problem went away. Baker and Cameron (1996)

(*Continued*)

Application Case 1.1 (Continued)

give several other examples of distractions, including lighting and displays, that organizations use to reduce perceived waiting time. If the real problem is identified as *perceived* waiting time, it can make a big difference in the proposed solutions and their costs. For example, full-length mirrors probably cost a whole lot less than adding an elevator!

Sources: Based on J. Baker and M. Cameron. (1996, September). "The Effects of the Service Environment on Affect and Consumer Perception of Waiting Time: An Integrative Review and Research Propositions," *Journal of the Academy of Marketing Science, 24,* pp. 338–349; R. Hesse and G. Woolsey (1975). *Applied Management Science: A Quick and Dirty Approach*. Chicago, IL: SRA Inc; R. C. Larson. (1987, November/December). "Perspectives on Queues: Social Justice and the Psychology of Queuing." *Operations Research, 35*(6), pp. 895–905.

Questions for Case 1.1

1. Why this is an example relevant to decision making?
2. Relate this situation to the intelligence phase of decision making.

PROBLEM CLASSIFICATION Problem classification is the conceptualization of a problem in an attempt to place it in a definable category, possibly leading to a standard solution approach. An important approach classifies problems according to the degree of structuredness evident in them. This ranges from totally structured (i.e., programmed) to totally unstructured (i.e., unprogrammed).

PROBLEM DECOMPOSITION Many complex problems can be divided into subproblems. Solving the simpler subproblems may help in solving a complex problem. Also, seemingly poorly structured problems sometimes have highly structured subproblems. Just as a semistructured problem results when some phases of decision making are structured whereas other phases are unstructured, and when some subproblems of a decision-making problem are structured with others unstructured, the problem itself is semistructured. As a decision support system is developed and the decision maker and development staff learn more about the problem, it gains structure.

PROBLEM OWNERSHIP In the intelligence phase, it is important to establish problem ownership. A problem exists in an organization only if someone or some group takes the responsibility for attacking it and if the organization has the ability to solve it. The assignment of authority to solve the problem is called *problem ownership*. For example, a manager may feel that he or she has a problem because interest rates are too high. Because interest rate levels are determined at the national and international levels and most managers can do nothing about them, high interest rates are the problem of the government, not a problem for a specific company to solve. The problem that companies actually face is how to operate in a high interest-rate environment. For an individual company, the interest rate level should be handled as an uncontrollable (environmental) factor to be predicted.

When problem ownership is not established, either someone is not doing his or her job or the problem at hand has yet to be identified as belonging to anyone. It is then important for someone to either volunteer to own it or assign it to someone.

The intelligence phase ends with a formal problem statement.

The Design Phase

The design phase involves finding or developing and analyzing possible courses of action. These include understanding the problem and testing solutions for feasibility. A model of the decision-making problem is constructed, tested, and validated. Let us first define a model.

MODELS A major characteristic of computerized decision support and many BI tools (notably those of business analytics) is the inclusion of at least one model. The basic idea is to perform the analysis on a model of reality rather than on the real system. A *model* is a simplified representation or abstraction of reality. It is usually simplified because reality is too complex to describe exactly and because much of the complexity is actually irrelevant in solving a specific problem.

Modeling involves conceptualizing a problem and abstracting it to quantitative and/or qualitative form. For a mathematical model, the variables are identified and their mutual relationships are established. Simplifications are made, whenever necessary, through assumptions. For example, a relationship between two variables may be assumed to be linear even though in reality there may be some nonlinear effects. A proper balance between the level of model simplification and the representation of reality must be obtained because of the cost–benefit trade-off. A simpler model leads to lower development costs, easier manipulation, and a faster solution but is less representative of the real problem and can produce inaccurate results. However, a simpler model generally requires fewer data, or the data are aggregated and easier to obtain.

The Choice Phase

Choice is the critical act of decision making. The choice phase is the one in which the actual decision and the commitment to follow a certain course of action are made. The boundary between the design and choice phases is often unclear because certain activities can be performed during both of them and because the decision maker can return frequently from choice activities to design activities (e.g., generate new alternatives while performing an evaluation of existing ones). The choice phase includes the search for, evaluation of, and recommendation of an appropriate solution to a model. A solution to a model is a specific set of values for the decision variables in a selected alternative. Choices can be evaluated as to their viability and profitability.

Each alternative must be evaluated. If an alternative has multiple goals, they must all be examined and balanced against each other. Sensitivity analysis is used to determine the robustness of any given alternative; slight changes in the parameters should ideally lead to slight or no changes in the alternative chosen. What-if analysis is used to explore major changes in the parameters. Goal seeking helps a manager determine values of the decision variables to meet a specific objective. These topics are addressed in Chapter 8.

The Implementation Phase

In *The Prince*, Machiavelli astutely noted some 500 years ago that there was "nothing more difficult to carry out, nor more doubtful of success, nor more dangerous to handle, than to initiate a new order of things." The implementation of a proposed solution to a problem is, in effect, the initiation of a new order of things or the introduction of change. And change must be managed. User expectations must be managed as part of change management.

The definition of *implementation* is somewhat complicated because implementation is a long, involved process with vague boundaries. Simplistically, the **implementation phase** involves putting a recommended solution to work, not necessarily implementing a computer system. Many generic implementation issues, such as resistance to change, degree of support of top management, and user training, are important in dealing with information system–supported decision making. Indeed, many previous technology-related waves (e.g., business process reengineering [BPR] and knowledge management) have faced mixed results mainly because of change management challenges and issues. Management of change is almost an entire discipline in itself, so we recognize its importance and encourage readers to focus on it independently. Implementation also includes a thorough understanding of project management. The importance of project management goes far beyond analytics, so the last few years have witnessed a major growth in certification programs for project managers. A very popular certification now is the Project Management Professional (PMP). See **pmi.org** for more details.

Implementation must also involve collecting and analyzing data to learn from the previous decisions and improve the next decision. Although analysis of data is usually conducted to identify the problem and/or the solution, analytics should also be employed in the feedback process. This is especially true for any public policy decisions. We need to be sure that the data being used for problem identification is valid. Sometimes people find this out only after the implementation phase.

The decision-making process, though conducted by people, can be improved with computer support, which is introduced next.

The Classical Decision Support System Framework

The early definitions of decision support system (DSS) identified it as a system intended to support managerial decision makers in semistructured and unstructured decision situations. DSS was meant to be an adjunct to decision makers, extending their capabilities but not replacing their judgment. DSS was aimed at decisions that required judgment or at decisions that could not be completely supported by algorithms. Not specifically stated but implied in the early definitions was the notion that the system would be computer based, would operate interactively online, and preferably would have graphical output capabilities, now simplified via browsers and mobile devices.

An early framework for computerized decision support includes several major concepts that are used in forthcoming sections and chapters of this book. Gorry and Scott-Morton created and used this framework in the early 1970s, and the framework then evolved into a new technology called *DSS*.

Gorry and Scott-Morton (1971) proposed a framework that is a 3-by-3 matrix, as shown in Figure 1.2. The two dimensions are the degree of structuredness and the types of control.

DEGREE OF STRUCTUREDNESS The left side of Figure 1.2 is based on Simon's (1977) idea that decision-making processes fall along a continuum that ranges from highly structured (sometimes called *programmed*) to highly unstructured (i.e., *non-programmed*) decisions. Structured processes are routine and typically repetitive problems for which standard solution methods exist. *Unstructured processes* are fuzzy, complex problems for which there are no cut-and-dried solution methods.

An *unstructured problem* is one where the articulation of the problem or the solution approach may be unstructured in itself. In a *structured problem*, the procedures for obtaining the best (or at least a good enough) solution are known. Whether the problem involves finding an appropriate inventory level or choosing an optimal investment strategy, the objectives are clearly defined. Common objectives are cost minimization and profit maximization.

Semistructured problems fall between structured and unstructured problems, having some structured elements and some unstructured elements. Keen and Scott-Morton (1978) mentioned trading bonds, setting marketing budgets for consumer products, and performing capital acquisition analysis as semistructured problems.

TYPES OF CONTROL The second half of the Gorry and Scott-Morton (1971) framework (refer to Figure 1.2) is based on Anthony's (1965) taxonomy, which defines three broad categories that encompass all managerial activities: *strategic planning*, which involves defining long-range goals and policies for resource allocation; *management control*, the

	Type of Control		
Type of Decision	**Operational Control**	**Managerial Control**	**Strategic Planning**
Structured	1 Monitoring accounts receivable Monitoring accounts payable Placing order entries	2 Analyzing budget Forecasting short-term Reporting on personnel Making or buying	3 Managing finances Monitoring investment portfolio Locating warehouse Monitoring distribution systems
Semistructured	4 Scheduling production Controlling inventory	5 Evaluating credit Preparing budget Laying out plant Scheduling project Designing reward system Categorizing inventory	6 Building a new plant Planning mergers and acquisitions Planning new products Planning compensation Providing quality assurance Establishing human resources policies Planning inventory
Unstructured	7 Buying software Approving loans Operating a help desk Selecting a cover for a magazine	8 Negotiating Recruiting an executive Buying hardware Lobbying	9 Planning research and development Developing new technologies Planning social responsibility

FIGURE 1.2 **Decision Support Frameworks.**

acquisition and efficient use of resources in the accomplishment of organizational goals; and *operational control*, the efficient and effective execution of specific tasks.

THE DECISION SUPPORT MATRIX Anthony's (1965) and Simon's (1977) taxonomies are combined in the nine-cell decision support matrix shown in Figure 1.2. The initial purpose of this matrix was to suggest different types of computerized support to different cells in the matrix. Gorry and Scott-Morton (1971) suggested, for example, that for making *semistructured decisions* and *unstructured decisions*, conventional management information systems (MIS) and management science (MS) tools are insufficient. Human intellect and a different approach to computer technologies are necessary. They proposed the use of a supportive information system, which they called a *DSS*.

Note that the more structured and operational control-oriented tasks (such as those in cells 1, 2, and 4 of Figure 1.2) are usually performed by lower-level managers, whereas the tasks in cells 6, 8, and 9 are the responsibility of top executives or highly trained specialists.

COMPUTER SUPPORT FOR STRUCTURED DECISIONS Since the 1960s, computers have historically supported structured and some semistructured decisions, especially those that involve operational and managerial control. Operational and managerial control decisions are made in all functional areas, especially in finance and production (i.e., operations) management.

Structured problems, which are encountered repeatedly, have a high level of structure, as their name suggests. It is therefore possible to abstract, analyze, and classify them into specific categories. For example, a make-or-buy decision is one category. Other examples of categories are capital budgeting, allocation of resources, distribution, procurement, planning, and inventory control decisions. For each category of decision, an easy-to-apply prescribed model and solution approach have been developed, generally as quantitative formulas. Therefore, it is possible to use a *scientific approach* for automating portions of managerial decision making. Solutions to many structured problems can be fully automated (see Chapters 2 and 12).

COMPUTER SUPPORT FOR UNSTRUCTURED DECISIONS Unstructured problems can be only partially supported by standard computerized quantitative methods. It is usually necessary to develop customized solutions. However, such solutions may benefit from data and information generated from corporate or external data sources. Intuition and judgment may play a large role in these types of decisions, as may computerized communication and collaboration technologies, as well as cognitive computing (Chapter 6) and deep learning (Chapter 5).

COMPUTER SUPPORT FOR SEMISTRUCTURED PROBLEMS Solving semistructured problems may involve a combination of standard solution procedures and human judgment. Management science can provide models for the portion of a decision-making problem that is structured. For the unstructured portion, a DSS can improve the quality of the information on which the decision is based by providing, for example, not only a single solution, but also a range of alternative solutions along with their potential impacts. These capabilities help managers to better understand the nature of problems and, thus, to make better decisions.

DECISION SUPPORT SYSTEM: CAPABILITIES The early definitions of DSS identified it as a system intended to support managerial decision makers in semistructured and unstructured decision situations. DSS was meant to be an adjunct to decision makers, extending their capabilities but not replacing their judgment. It was aimed at decisions that required judgment or at decisions that could not be completely supported by algorithms. Not specifically stated but implied in the early definitions was the notion that the system would be computer based, would operate interactively online, and preferably would have graphical output capabilities, now simplified via browsers and mobile devices.

A DSS Application

A DSS is typically built to support the solution of a certain problem or to evaluate an opportunity. This is a key difference between DSS and BI applications. In a very strict sense, **business intelligence (BI)** systems monitor situations and identify problems and/or opportunities using analytic methods. Reporting plays a major role in BI; the user generally must identify whether a particular situation warrants attention and then can apply analytical methods. Again, although models and data access (generally through a data warehouse) are included in BI, a DSS may have its own databases and is developed to solve a specific problem or set of problems and are therefore called DSS applications.

Formally, a DSS is an approach (or methodology) for supporting decision making. It uses an interactive, flexible, adaptable computer-based information system (CBIS) especially developed for supporting the solution to a specific unstructured management problem. It uses data, provides an easy user interface, and can incorporate the decision maker's own insights. In addition, a DSS includes models and is developed (possibly by

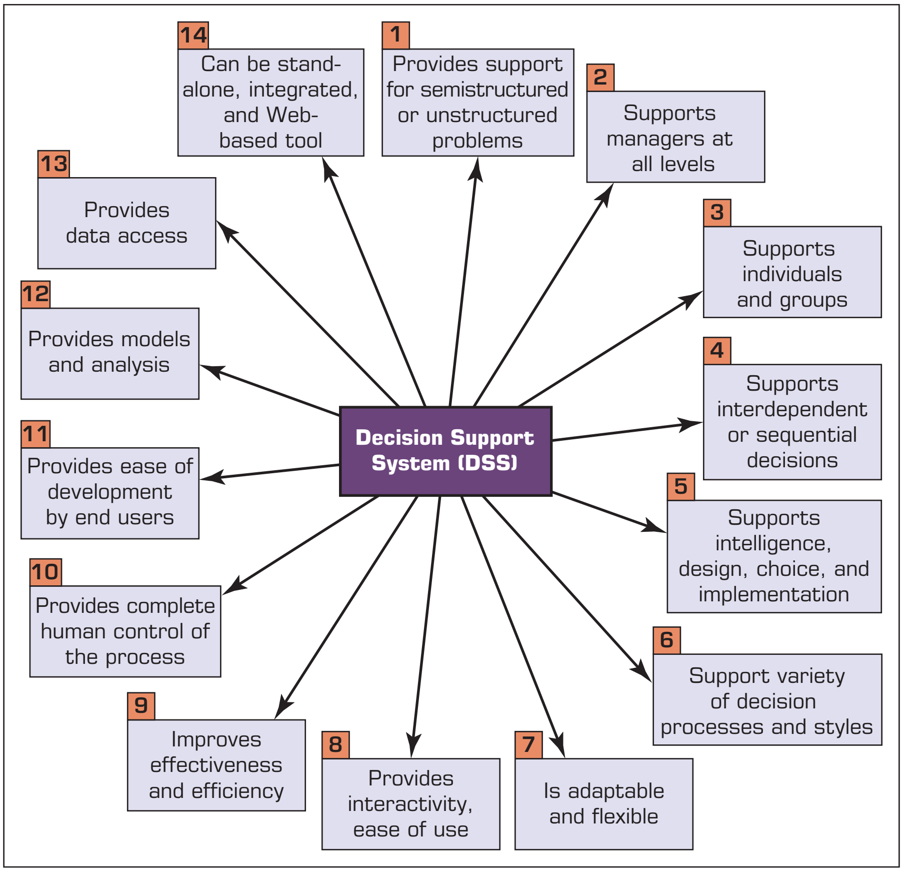


FIGURE 1.3 **Key Characteristics and Capabilities of DSS.**

end users) through an interactive and iterative process. It can support all phases of decision making and may include a knowledge component. Finally, a DSS can be used by a single user or can be Web based for use by many people at several locations.

THE CHARACTERISTICS AND CAPABILITIES OF DSS Because there is no consensus on exactly what a DSS is, there is obviously no agreement on the standard characteristics and capabilities of DSS. The capabilities in Figure 1.3 constitute an ideal set, some members of which are described in the definitions of DSS and illustrated in the application cases.

The key characteristics and capabilities of DSS (as shown in Figure 1.3) are as follows:

1. Supports decision makers, mainly in semistructured and unstructured situations, by bringing together human judgment and computerized information. Such problems cannot be solved (or cannot be solved conveniently) by other computerized systems or through use of standard quantitative methods or tools. Generally, these problems gain structure as the DSS is developed. Even some structured problems have been solved by DSS.
2. Supports all managerial levels, ranging from top executives to line managers.
3. Supports individuals as well as groups. Less-structured problems often require the involvement of individuals from different departments and organizational levels or even from different organizations. DSS supports virtual teams through collaborative Web tools. DSS has been developed to support individual and group work as well

as to support individual decision making and groups of decision makers working somewhat independently.

4. Supports interdependent and/or sequential decisions. The decisions may be made once, several times, or repeatedly.
5. Supports all phases of the decision-making process: intelligence, design, choice, and implementation.
6. Supports a variety of decision-making processes and styles.
7. Is flexible, so users can add, delete, combine, change, or rearrange basic elements. The decision maker should be reactive, able to confront changing conditions quickly, and able to adapt the DSS to meet these changes. It is also flexible in that it can be readily modified to solve other, similar problems.
8. Is user-friendly, has strong graphical capabilities, and a natural language interactive human-machine interface can greatly increase the effectiveness of DSS. Most new DSS applications use Web-based interfaces or mobile platform interfaces.
9. Improves the effectiveness of decision making (e.g., accuracy, timeliness, quality) rather than its efficiency (e.g., the cost of making decisions). When DSS is deployed, decision making often takes longer, but the decisions are better.
10. Provides complete control by the decision maker over all steps of the decision-making process in solving a problem. A DSS specifically aims to support, not to replace, the decision maker.
11. Enables end users to develop and modify simple systems by themselves. Larger systems can be built with assistance from IS specialists. Spreadsheet packages have been utilized in developing simpler systems. OLAP and data mining software in conjunction with data warehouses enable users to build fairly large, complex DSS.
12. Provides models that are generally utilized to analyze decision-making situations. The modeling capability enables experimentation with different strategies under different configurations.
13. Provides access to a variety of data sources, formats, and types, including GIS, multimedia, and object-oriented data.
14. Can be employed as a stand-alone tool used by an individual decision maker in one location or distributed throughout an organization and in several organizations along the supply chain. It can be integrated with other DSS and/or applications, and it can be distributed internally and externally, using networking and Web technologies.

These key DSS characteristics and capabilities allow decision makers to make better, more consistent decisions in a timely manner, and they are provided by major DSS components,

Components of a Decision Support System

A DSS application can be composed of a data management subsystem, a model management subsystem, a user interface subsystem, and a knowledge-based management subsystem. We show these in Figure 1.4.

The Data Management Subsystem

The data management subsystem includes a database that contains relevant data for the situation and is managed by software called the database management system (DBMS). *DBMS* is used as both singular and plural (*system* and *systems*) terms, as are many other acronyms in this text. The data management subsystem can be interconnected with the corporate data warehouse, a repository for corporate relevant decision-making data.

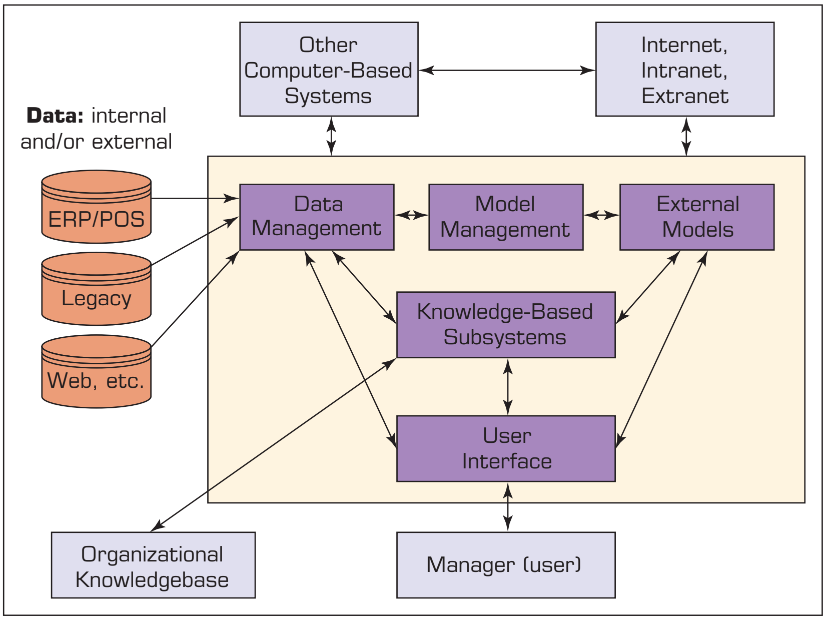


FIGURE 1.4 **Schematic View of DSS.**

Usually, the data are stored or accessed via a database Web server. The data management subsystem is composed of the following elements:

- DSS database
- Database management system
- Data directory
- Query facility

Many of the BI or descriptive analytics applications derive their strength from the data management side of the subsystems.

The Model Management Subsystem

The model management subsystem is the component that includes financial, statistical, management science, or other quantitative models that provide the system's analytical capabilities and appropriate software management. Modeling languages for building custom models are also included. This software is often called a model base management system (MBMS). This component can be connected to corporate or external storage of models. Model solution methods and management systems are implemented in Web development systems (such as Java) to run on application servers. The model management subsystem of a DSS is composed of the following elements:

- Model base
- MBMS
- Modeling language
- Model directory
- Model execution, integration, and command processor

Because DSS deals with semistructured or unstructured problems, it is often necessary to customize models, using programming tools and languages. Some examples of these are .NET Framework languages, C++, and Java. OLAP software may also be used to work with models in data analysis. Even languages for simulations such as Arena and

Application Case 1.2

SNAP DSS Helps OneNet Make Telecommunications Rate Decisions

Telecommunications network services to educational institutions and government entities are typically provided by a mix of private and public organizations. Many states in the United States have one or more state agencies that are responsible for providing network services to schools, colleges, and other state agencies. One example of such an agency is OneNet in Oklahoma. OneNet is a division of the Oklahoma State Regents for Higher Education and operated in cooperation with the Office of State Finance.

Usually agencies such as OneNet operate as an enterprise-type fund. They must recover their costs through billing their clients and/or by justifying appropriations directly from the state legislatures. This cost recovery should occur through a pricing mechanism that is efficient, simple to implement, and equitable. This pricing model typically needs to recognize many factors: convergence of voice, data, and video traffic on the same infrastructure; diversity of user base in terms of educational institutions and state agencies; diversity of applications in use by state clients from e-mail to videoconferences, IP telephoning, and distance learning; recovery of current costs as well as planning for upgrades and future developments; and leverage of the shared infrastructure to enable further economic development and collaborative work across the state that leads to innovative uses of OneNet.

These considerations led to the development of a spreadsheet-based model. The system, SNAP-DSS, or Service Network Application and Pricing (SNAP)-based DSS, was developed in Microsoft Excel 2007 and used the VBA programming language.

The SNAP-DSS offers OneNet the ability to select the rate card options that best fit the preferred pricing strategies by providing a real-time, user-friendly, graphical user interface (GUI). In addition, the SNAP-DSS not only illustrates the influence of the changes in the pricing factors on each rate card option but also allows the user to analyze various rate card options in different scenarios using different parameters. This model has been used by OneNet financial planners to gain insights into their customers and analyze many what-if scenarios of different rate plan options.

Source: Based on J. Chongwatpol and R. Sharda. (2010, December). "SNAP: A DSS to Analyze Network Service Pricing for State Networks." *Decision Support Systems, 50*(1), pp. 347–359.

statistical packages such as those of SPSS offer modeling tools developed through the use of a proprietary programming language. For small- and medium-sized DSS or for less complex ones, a spreadsheet (e.g., Excel) is usually used. We use Excel for several examples in this book. Application Case 1.2 describes a spreadsheet-based DSS.

The User Interface Subsystem

The user communicates with and commands the DSS through the user interface subsystem. The user is considered part of the system. Researchers assert that some of the unique contributions of DSS are derived from the intensive interaction between the computer and the decision maker. A difficult user interface is one of the major reasons that managers do not use computers and quantitative analyses as much as they could, given the availability of these technologies. The Web browser provided a familiar, consistent GUI structure for many DSS in the 2000s. For locally used DSS, a spreadsheet also provides a familiar user interface. The Web browser has been recognized as an effective DSS GUI because it is flexible, user-friendly, and a gateway to almost all sources of necessary information and data. Essentially, Web browsers have led to the development of portals and dashboards, which front end many DSS.

Explosive growth in portable devices, including smartphones and tablets, has changed the DSS user interfaces as well. These devices allow either handwritten input or typed input from internal or external keyboards. Some DSS user interfaces utilize natural language input (i.e., text in a human language) so that the users can easily express themselves in a meaningful way. Cell phone inputs through short message service (SMS) or chatbots are becoming more common for at least some consumer DSS-type applications. For example, one can send an SMS request for search on any topic to GOOGL (46645). Such capabilities are most useful in locating nearby businesses, addresses, or phone numbers, but it can also be used for many other decision support tasks. For example, users can find definitions of words by entering the word "define" followed by a word, such as "define extenuate." Some of the other capabilities include

- Price lookups: "Price 64GB iPhone X."
- Currency conversions: "10 US dollars in euros."
- Sports scores and game times: Just enter the name of a team ("NYC Giants"), and Google SMS will send the most recent game's score and the date and time of the next match.

This type of SMS-based search capability is also available for other search engines such as Microsoft's search engine Bing.

With the emergence of smartphones such as Apple's iPhone and Android smartphones from many vendors, many companies are developing *apps* to provide purchasing-decision support. For example, Amazon's app allows a user to take a picture of any item in a store (or wherever) and send it to **Amazon.com**. **Amazon.com's** graphics-understanding algorithm tries to match the image to a real product in its databases and sends the user a page similar to **Amazon.com's** product info pages, allowing users to perform price comparisons in real time. Millions of other apps have been developed that provide consumers support for decision making on finding and selecting stores/restaurants/service providers on the basis of location, recommendations from others, and especially from your own social circles. Search activities noted in the previous paragraph are also largely accomplished now through apps provided by each search provider.

Voice input for these devices and the new smart speakers such as Amazon Echo (Alexa) and Google Home is common and fairly accurate (but not perfect). When voice input with accompanying speech-recognition software (and readily available text-to-speech software) is used, verbal instructions with accompanied actions and outputs can be invoked. These are readily available for DSS and are incorporated into the portable devices described earlier. An example of voice inputs that can be used for a general-purpose DSS is Apple's Siri application and Google's Google Now service. For example, a user can give her or his zip code and say "pizza delivery." These devices provide the search results and can even place a call to a business.

The Knowledge-Based Management Subsystem

Many of the user interface developments are closely tied to the major new advances in their knowledge-based systems. The knowledge-based management subsystem can support any of the other subsystems or act as an independent component. It provides intelligence to augment the decision maker's own or to help understand a user's query so as to provide a consistent answer. It can be interconnected with the organization's knowledge repository (part of a KMS), which is sometimes called the *organizational knowledge base*, or connect to thousands of external knowledge sources. Many artificial intelligence methods have been implemented in the current generation of learning systems and are easy to integrate into the other DSS components. One of the most widely publicized knowledge-based DSS is IBM's Watson, which was introduced in the opening vignette and will be described in more detail later.

This section has covered the history and progression of Decision Support Systems in brief. In the next section we discuss evolution of this support to business intelligence, analytics, and data science.

SECTION 1.3 REVIEW QUESTIONS

1. List and briefly describe Simon's four phases of decision making.
2. What is the difference between a problem and its symptoms?
3. Why is it important to classify a problem?
4. Define *implementation*.
5. What are structured, unstructured, and semistructured decisions? Provide two examples of each.
6. Define *operational control, managerial control*, and *strategic planning*. Provide two examples of each.
7. What are the nine cells of the decision framework? Explain what each is for.
8. How can computers provide support for making structured decisions?
9. How can computers provide support for making semistructured and unstructured decisions?

1.4 EVOLUTION OF COMPUTERIZED DECISION SUPPORT TO BUSINESS INTELLIGENCE/ANALYTICS/DATA SCIENCE

The timeline in Figure 1.5 shows the terminology used to describe analytics since the 1970s. During the 1970s, the primary focus of information systems support for decision making focused on providing structured, periodic reports that a manager could use for decision making (or ignore them). Businesses began to create routine reports to inform decision makers (managers) about what had happened in the previous period (e.g., day, week, month, quarter). Although it was useful to know what had happened in the past, managers needed more than this: They needed a variety of reports at different levels of granularity to better understand and address changing needs and challenges of the business. These were usually called *management information systems (MIS)*. In the early 1970s, Scott-Morton first articulated the major concepts of DSS. He defined DSS as "interactive computer-based systems, which help decision makers utilize *data* and *models* to solve unstructured problems" (Gorry and Scott-Morton, 1971). The following is another classic DSS definition provided by Keen and Scott-Morton (1978):

> Decision support systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. It is a computer-based support system for management decision makers who deal with semistructured problems.

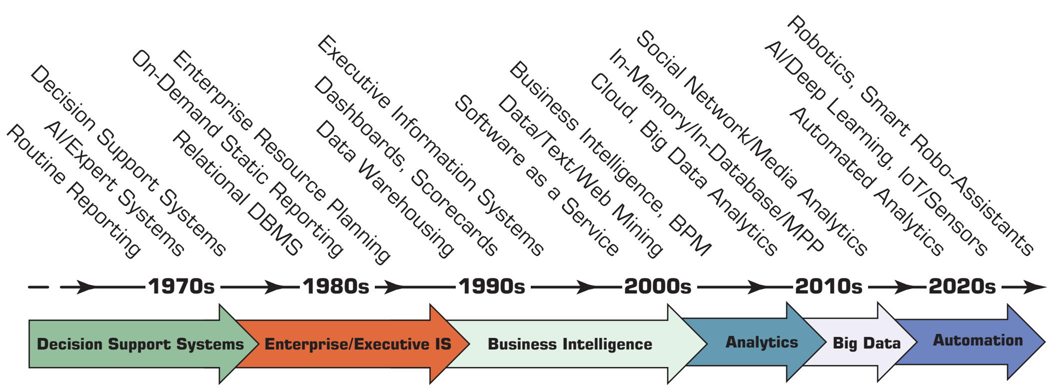


FIGURE 1.5 Evolution of Decision Support, Business Intelligence, Analytics, and AI.

Note that the term *decision support system*, like *management information system* and several other terms in the field of IT, is a content-free expression (i.e., it means different things to different people). Therefore, there is no universally accepted definition of DSS.

During the early days of analytics, data were often obtained from the domain experts using manual processes (i.e., interviews and surveys) to build mathematical or knowledge-based models to solve constrained optimization problems. The idea was to do the best with limited resources. Such decision support models were typically called operations research (OR). The problems that were too complex to solve optimally (using linear or nonlinear mathematical programming techniques) were tackled using heuristic methods such as simulation models. (We will introduce these as prescriptive analytics later in this chapter).

In the late 1970s and early 1980s, in addition to the mature OR models that were being used in many industries and government systems, a new and exciting line of models had emerged: rule-based expert systems (ESs). These systems promised to capture experts' knowledge in a format that computers could process (via a collection of if-then-else rules or heuristics) so that these could be used for consultation much the same way that one would use domain experts to identify a structured problem and to prescribe the most probable solution. ESs allowed scarce expertise to be made available where and when needed, using an "intelligent" DSS.

The 1980s saw a significant change in the way organizations captured business-related data. The old practice had been to have multiple disjointed information systems tailored to capture transactional data of different organizational units or functions (e.g., accounting, marketing and sales, finance, manufacturing). In the 1980s, these systems were integrated as enterprise-level information systems that we now commonly call *enterprise resource planning (ERP)* systems. The old mostly sequential and nonstandardized data representation schemas were replaced by relational database management (RDBM) systems. These systems made it possible to improve the capture and storage of data as well as the relationships between organizational data fields while significantly reducing the replication of information. The need for RDBM and ERP systems emerged when data integrity and consistency became an issue, significantly hindering the effectiveness of business practices. With ERP, all the data from every corner of the enterprise is collected and integrated into a consistent schema so that every part of the organization has access to the single version of the truth when and where needed. In addition to the emergence of ERP systems, or perhaps because of these systems, business reporting became an on-demand, as-needed business practice. Decision makers could decide when they needed to or wanted to create specialized reports to investigate organizational problems and opportunities.

In the 1990s, the need for more versatile reporting led to the development of executive information systems (EISs; DSS designed and developed specifically for executives and their decision-making needs). These systems were designed as graphical dashboards and scorecards so that they could serve as visually appealing displays while focusing on the most important factors for decision makers to keep track of the key performance indicators. To make this highly versatile reporting possible while keeping the transactional integrity of the business information systems intact, it was necessary to create a middle data tier known as a DW as a repository to specifically support business reporting and decision making. In a very short time, most large- to medium-sized businesses adopted data warehousing as their platform for enterprise-wide decision making. The dashboards and scorecards got their data from a DW, and by doing so, they were not hindering the efficiency of the business transaction systems mostly referred to as ERP systems.

In the 2000s, the DW-driven DSS began to be called *BI systems*. As the amount of longitudinal data accumulated in the DWs increased, so did the capabilities of hardware and software to keep up with the rapidly changing and evolving needs of the decision makers. Because of the globalized competitive marketplace, decision makers needed current information in a very digestible format to address business problems and to take advantage of market opportunities in a timely manner. Because the data in a DW are updated periodically, they do not reflect the latest information. To elevate this information latency problem, DW vendors developed a system to update the data more frequently, which led to the terms *real-time data warehousing* and, more realistically, *right-time data warehousing*, which differs from the former by adopting a data-refreshing policy based on the needed freshness of the data items (i.e., not all data items need to be refreshed in real time). DWs are very large and feature rich, and it became necessary to “mine” the corporate data to “discover” new and useful knowledge nuggets to improve business processes and practices, hence, the terms *data mining* and *text mining*. With the increasing volumes and varieties of data, the needs for more storage and more processing power emerged. Although large corporations had the means to tackle this problem, small- to medium-sized companies needed more financially manageable business models. This need led to service-oriented architecture and software and infrastructure-as-a-service analytics business models. Smaller companies, therefore, gained access to analytics capabilities on an as-needed basis and paid only for what they used, as opposed to investing in financially prohibitive hardware and software resources.

In the 2010s, we are seeing yet another paradigm shift in the way that data are captured and used. Largely because of the widespread use of the Internet, new data generation mediums have emerged. Of all the new data sources (e.g., radio-frequency identification [RFID] tags, digital energy meters, clickstream Web logs, smart home devices, wearable health monitoring equipment), perhaps the most interesting and challenging is social networking/social media. These unstructured data are rich in information content, but analysis of such data sources poses significant challenges to computational systems from both software and hardware perspectives. Recently, the term *Big Data* has been coined to highlight the challenges that these new data streams have brought on us. Many advancements in both hardware (e.g., massively parallel processing with very large computational memory and highly parallel multiprocessor computing systems) and software/algorithms (e.g., Hadoop with MapReduce and NoSQL, Spark) have been developed to address the challenges of Big Data.

The last few years and the upcoming decade are bringing massive growth in many exciting dimensions. For example, streaming analytics and the sensor technologies have enabled the IoT. Artificial Intelligence is changing the shape of BI by enabling new ways of analyzing images through deep learning, not just traditional visualization of data. Deep learning and AI are also helping grow voice recognition and speech synthesis, leading to new interfaces in interacting with technologies. Almost half of U.S. households already have a smart speaker such as Amazon Echo or Google Home and have begun to interact with data and systems using voice interfaces. Growth in video interfaces will eventually enable gesture-based interaction with systems. All of these are being enabled due to massive cloud-based data storage and amazingly fast processing capabilities. And more is yet to come.

It is hard to predict what the next decade will bring and what the new analytics-related terms will be. The time between new paradigm shifts in information systems and particularly in analytics has been shrinking, and this trend will continue for the foreseeable future. Even though analytics is not new, the explosion in its popularity is very new. Thanks to the recent explosion in Big Data, ways to collect and store these data and intuitive software tools, data-driven insights are more accessible to business professionals than ever before. Therefore, in the midst of global competition, there is a huge opportunity to make better managerial decisions by using data and analytics to increase revenue while decreasing costs by building better products, improving customer experience, and catching fraud before it happens, improving customer engagement through targeting and customization, and developing entirely new lines of business, all with the power of analytics and data. More and more companies are now preparing their employees with the know-how of business analytics to drive effectiveness and efficiency in their day-to-day decision-making processes.

The next section focuses on a framework for BI. Although most people would agree that BI has evolved into analytics and data science, many vendors and researchers still use that term. So the next few paragraphs pay homage to that history by specifically focusing on what has been called BI. Following the next section, we introduce analytics and use that as the label for classifying all related concepts.

A Framework for Business Intelligence

The decision support concepts presented in Sections 1.2 and 1.3 have been implemented incrementally, under different names, by many vendors that have created tools and methodologies for decision support. As noted in Section 1.2, as the enterprise-wide systems grew, managers were able to access user-friendly reports that enabled them to make decisions quickly. These systems, which were generally called EISs, then began to offer additional visualization, alerts, and performance measurement capabilities. By 2006, the major *commercial* products and services appeared under the term *business intelligence (BI)*.

DEFINITIONS OF BI *Business intelligence (BI)* is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies. It is, like DSS, a content-free expression, so it means different things to different people. Part of the confusion about BI lies in the flurry of acronyms and buzzwords that are associated with it (e.g., business performance management [BPM]). BI's major objective is to enable interactive access (sometimes in real time) to data, to enable manipulation of data, and to give business managers and analysts the ability to conduct appropriate analyses. By analyzing historical and current data, situations, and performances, decision makers get valuable insights that enable them to make more informed and better decisions. The process of BI is based on the *transformation* of data to information, then to decisions, and finally to actions.

A BRIEF HISTORY OF BI The term *BI* was coined by the Gartner Group in the mid-1990s. However, as the history in the previous section points out, the concept is much older; it has its roots in the MIS reporting systems of the 1970s. During that period, reporting systems were static, were two dimensional, and had no analytical capabilities. In the early 1980s, the concept of EISs emerged. This concept expanded the computerized support to top-level managers and executives. Some of the capabilities introduced were dynamic multidimensional (ad hoc or on-demand) reporting, forecasting and prediction, trend analysis, drill-down to details, status access, and critical success factors. These features appeared in dozens of commercial products until the mid-1990s. Then the same capabilities and some new ones appeared under the name BI. Today, a good BI-based enterprise information system contains all the information that executives need. So, the original concept of EIS was transformed into BI. By 2005, BI systems started to include *artificial intelligence* capabilities as well as powerful analytical capabilities. Figure 1.6 illustrates the various tools and techniques that may be included in a BI system. It illustrates the evolution of BI as well. The tools shown in Figure 1.6 provide the capabilities of BI. The most sophisticated BI products include most of these capabilities; others specialize in only some of them.

The Architecture of BI

A BI system has four major components: a *DW*, with its source data; *business analytics*, a collection of tools for manipulating, mining, and analyzing the data in the DW; *BPM* for monitoring and analyzing performance; and a *user interface* (e.g., a **dashboard**). The relationship among these components is illustrated in Figure 1.7.

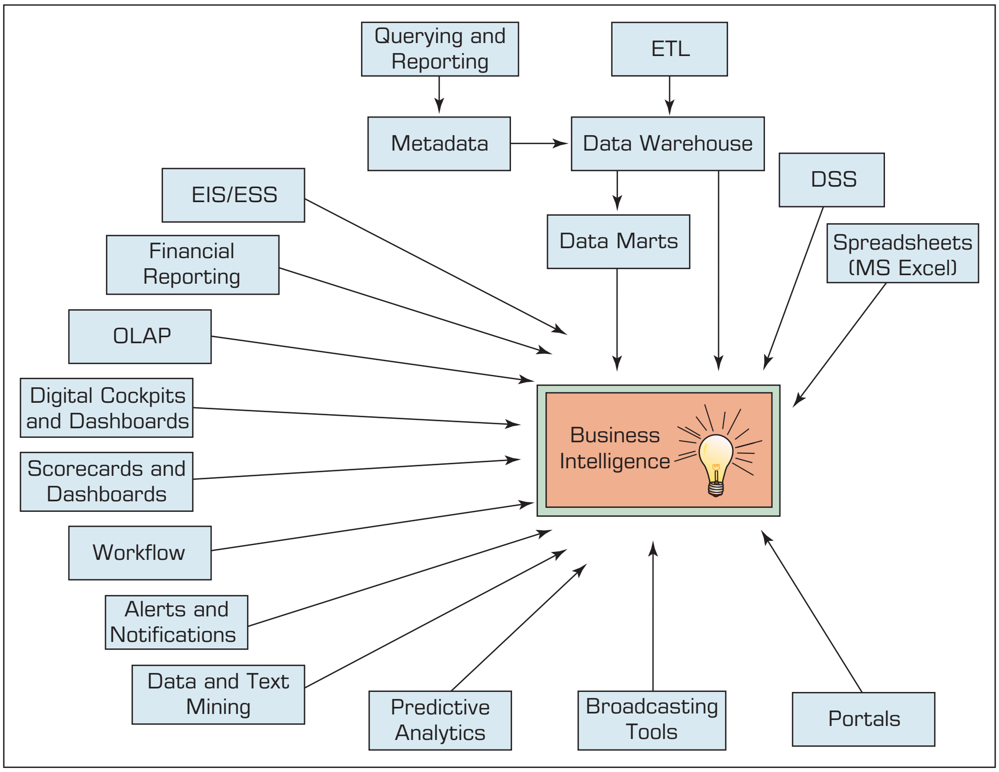


FIGURE 1.6 Evolution of Business Intelligence (BI).

The Origins and Drivers of BI

Where did modern approaches to DW and BI come from? What are their roots, and how do those roots affect the way organizations are managing these initiatives today? Today's investments in information technology are under increased scrutiny in terms of their bottom-line impact and potential. The same is true of DW and the BI applications that make these initiatives possible.

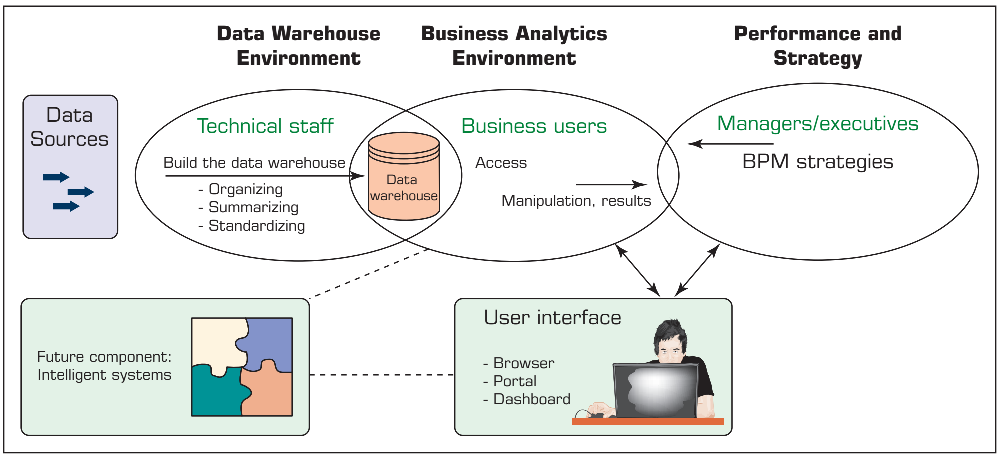


FIGURE 1.7 A High-Level Architecture of BI. *Source:* Based on W. Eckerson. (2003). *Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions.* Seattle, WA: The Data Warehousing Institute, p. 32, Illustration 5.

Organizations are being compelled to capture, understand, and harness their data to support decision making to improve business operations. Legislation and regulation (e.g., the Sarbanes-Oxley Act of 2002) now require business leaders to document their business processes and to sign off on the legitimacy of the information they rely on and report to stakeholders. Moreover, business cycle times are now extremely compressed; faster, more informed, and better decision making is, therefore, a competitive imperative. Managers need the *right information* at the *right time* and in the *right place*. This is the mantra for modern approaches to BI.

Organizations have to work smart. Paying careful attention to the management of BI initiatives is a necessary aspect of doing business. It is no surprise, then, that organizations are increasingly championing BI and under its new incarnation as analytics.

Data Warehouse as a Foundation for Business Intelligence

BI systems rely on a DW as the information source for creating insight and supporting managerial decisions. A multitude of organizational and external data is captured, transformed, and stored in a DW to support timely and accurate decisions through enriched business insight. In simple terms, a *DW* is a pool of data produced to support decision making; it is also a repository of current and historical data of potential interest to managers throughout the organization. Data are usually structured to be available in a form ready for analytical processing activities (i.e., OLAP, data mining, querying, reporting, and other decision support applications). A DW is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process.

Whereas a DW is a repository of data, data warehousing is literally the entire process. Data warehousing is a discipline that results in applications that provide decision support capability, allows ready access to business information, and creates business insight. The three main types of data warehouses are data marts (DMs), operational data stores (ODS), and enterprise data warehouses (EDW). Whereas a DW combines databases across an entire enterprise, a DM is usually smaller and focuses on a particular subject or department. A DM is a subset of a data warehouse, typically consisting of a single subject area (e.g., marketing, operations). An operational data store (ODS) provides a fairly recent form of customer information file. This type of database is often used as an interim staging area for a DW. Unlike the static contents of a DW, the contents of an ODS are updated throughout the course of business operations. An EDW is a large-scale data warehouse that is used across the enterprise for decision support. The large-scale nature of an EDW provides integration of data from many sources into a standard format for effective BI and decision support applications. EDWs are used to provide data for many types of DSS, including CRM, supply chain management (SCM), BPM, business activity monitoring, product life-cycle management, revenue management, and sometimes even KMS.

In Figure 1.8, we show the DW concept. Data from many different sources can be extracted, transformed, and loaded into a DW for further access and analytics for decision support. Further details of DW are available in an online chapter on the book's Web site.

Transaction Processing versus Analytic Processing

To illustrate the major characteristics of BI, first we will show what BI is not—namely, transaction processing. We are all familiar with the information systems that support our transactions, like ATM withdrawals, bank deposits, and cash register scans at the grocery store. These *transaction processing* systems are constantly involved in handling updates to what we might call *operational databases*. For example, in an ATM withdrawal transaction, we need to reduce our bank balance accordingly; a bank deposit adds to an account; and a grocery store purchase is likely reflected in the store's calculation of total sales for the day, and it should reflect an appropriate reduction in the store's inventory for the items we bought, and so on. These **online transaction processing (OLTP)** systems handle a

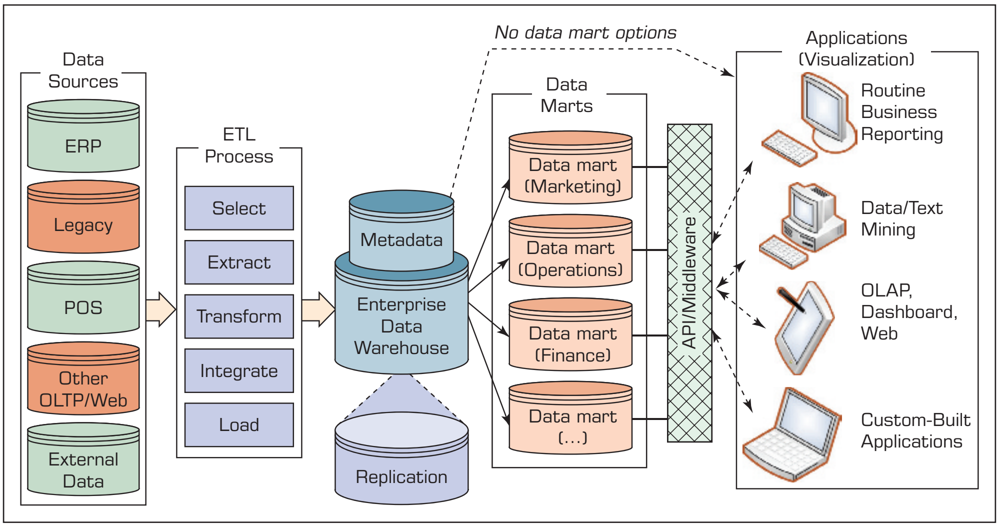


FIGURE 1.8 **Data Warehouse Framework and Views.**

company's routine ongoing business. In contrast, a DW is typically a distinct system that provides storage for data that will be used for *analysis*. The intent of that analysis is to give management the ability to scour data for information about the business, and it can be used to provide tactical or operational decision support whereby, for example, line personnel can make quicker and/or more informed decisions. DWs are intended to work with informational data used for **online analytical processing (OLAP)** systems.

Most operational data in ERP systems—and in their complementary siblings like *SCM* or *CRM*—are stored in an OLTP system, which is a type of computer processing where the computer responds immediately to user requests. Each request is considered to be a *transaction*, which is a computerized record of a discrete event, such as the receipt of inventory or a customer order. In other words, a transaction requires a set of two or more database updates that must be completed in an all-or-nothing fashion.

The very design that makes an OLTP system efficient for transaction processing makes it inefficient for end-user ad hoc reports, queries, and analysis. In the 1980s, many business users referred to their mainframes as "black holes" because all the information went into them, but none ever came back. All requests for reports had to be programmed by the IT staff, whereas only "precanned" reports could be generated on a scheduled basis, and ad hoc real-time querying was virtually impossible. Although the client/server-based ERP systems of the 1990s were somewhat more report friendly, they have still been a far cry from a desired usability by regular, nontechnical end users for things such as operational reporting and interactive analysis. To resolve these issues, the notions of DW and BI were created.

DWs contain a wide variety of data that present a coherent picture of business conditions at a single point in time. The idea was to create a database infrastructure that was always online and contained all the information from the OLTP systems, including historical data, but reorganized and structured in such a way that it was fast and efficient for querying, analysis, and decision support. Separating the OLTP from analysis and decision support enables the benefits of BI that were described earlier.

A Multimedia Exercise in Business Intelligence

TUN includes videos (similar to the television show *CSI*) to illustrate concepts of analytics in different industries. These are called "BSI Videos (Business Scenario Investigations)." Not only are these entertaining, but they also provide the class with some questions for discussion. For starters, please go to **https://www.teradatauniversitynetwork.com/Library/Items/BSI-The-Case-of-the-Misconnecting-Passengers/** or **www.youtube.**

com/watch?v=NXEL5F4_aKA. Watch the video that appears on YouTube. Essentially, you have to assume the role of a customer service center professional. An incoming flight is running late, and several passengers are likely to miss their connecting flights. There are seats on one outgoing flight that can accommodate two of the four passengers. Which two passengers should be given priority? You are given information about customers' profiles and relationships with the airline. Your decisions might change as you learn more about those customers' profiles.

Watch the video, pause it as appropriate, and answer the questions on which passengers should be given priority. Then resume the video to get more information. After the video is complete, you can see the slides related to this video and how the analysis was prepared on a slide set at **www.slideshare.net/teradata/bsi-how-we-did-it-the-case-of-the-misconnecting-passengers**.

This multimedia excursion provides an example of how additional available information through an enterprise DW can assist in decision making.

Although some people equate DSS with BI, these systems are not, at present, the same. It is interesting to note that some people believe that DSS is a part of BI—one of its analytical tools. Others think that BI is a special case of DSS that deals mostly with reporting, communication, and collaboration (a form of data-oriented DSS). Another explanation (Watson, 2005) is that BI is a result of a continuous revolution, and as such, DSS is one of BI's original elements. Further, as noted in the next section onward, in many circles, BI has been subsumed by the new terms *analytics* or *data science*.

APPROPRIATE PLANNING AND ALIGNMENT WITH THE BUSINESS STRATEGY First and foremost, the fundamental reasons for investing in BI must be aligned with the company's business strategy. BI cannot simply be a technical exercise for the information systems department. It has to serve as a way to change the manner in which the company conducts business by improving its business processes and transforming decision-making processes to be more data driven. Many BI consultants and practitioners involved in successful BI initiatives advise that a framework for planning is a necessary precondition. One framework, proposed by Gartner, Inc. (2004), decomposed planning and execution into *business, organization, functionality*, and *infrastructure* components. At the business and organizational levels, strategic and operational objectives must be defined while considering the available organizational skills to achieve those objectives. Issues of organizational culture surrounding BI initiatives and building enthusiasm for those initiatives and procedures for the intra-organizational sharing of BI best practices must be considered by upper management—with plans in place to prepare the organization for change. One of the first steps in that process is to assess the IS organization, the skill sets of the potential classes of users, and whether the culture is amenable to change. From this assessment, and assuming there are justification and the need to move ahead, a company can prepare a detailed action plan. Another critical issue for BI implementation success is the integration of several BI projects (most enterprises use several BI projects) among themselves and with the other IT systems in the organization and its business partners.

Gartner and many other analytics consulting organizations promoted the concept of a BI competence center that would serve the following functions:

- A center can demonstrate how BI is clearly linked to strategy and execution of strategy.
- A center can serve to encourage interaction between the potential business user communities and the IS organization.
- A center can serve as a repository and disseminator of best BI practices between and among the different lines of business.
- Standards of excellence in BI practices can be advocated and encouraged throughout the company.
- The IS organization can learn a great deal through interaction with the user communities, such as knowledge about the variety of types of analytical tools that are needed.

- The business user community and IS organization can better understand why the DW platform must be flexible enough to provide for changing business requirements.
- The center can help important stakeholders like high-level executives see how BI can play an important role.

Over the last 10 years, the idea of a BI competence center has been abandoned because many advanced technologies covered in this book have reduced the need for a central group to organize many of these functions. Basic BI has now evolved to a point where much of it can be done in "self-service" mode by the end users. For example, many data visualizations are easily accomplished by end users using the latest visualization packages (Chapter 3 will introduce some of these). As noted by Duncan (2016), the BI team would now be more focused on producing curated data sets to enable self-service BI. Because analytics is now permeating across the whole organization, the BI competency center could evolve into an analytics community of excellence to promote best practices and ensure overall alignment of analytics initiatives with organizational strategy.

BI tools sometimes needed to be integrated among themselves, creating synergy. The need for integration pushed software vendors to continuously add capabilities to their products. Customers who buy an all-in-one software package deal with only one vendor and do not have to deal with system connectivity. But they may lose the advantage of creating systems composed from the "best-of-breed" components. This led to major chaos in the BI market space. Many of the software tools that rode the BI wave (e.g., Savvion, Vitria, Tibco, MicroStrategy, Hyperion) have either been acquired by other companies or have expanded their offerings to take advantage of six key trends that have emerged since the initial wave of surge in business intelligence:

- Big Data.
- Focus on customer experience as opposed to just operational efficiency.
- Mobile and even newer user interfaces—visual, voice, mobile.
- Predictive and prescriptive analytics, machine learning, artificial intelligence.
- Migration to cloud.
- Much greater focus on security and privacy protection.

This book covers many of these topics in significant detail by giving examples of how the technologies are evolving and being applied, and the managerial implications.

SECTION 1.4 REVIEW QUESTIONS

1. List three of the terms that have been predecessors of analytics.
2. What was the primary difference between the systems called MIS, DSS, and Executive Information Systems?
3. Did DSS evolve into BI or vice versa?
4. Define *BI*.
5. List and describe the major components of BI.
6. Define *OLTP*.
7. Define *OLAP*.
8. List some of the implementation topics addressed by Gartner's report.
9. List some other success factors of BI.

1.5 ANALYTICS OVERVIEW

The word *analytics* has largely replaced the previous individual components of computerized decision support technologies that have been available under various labels in the past. Indeed, many practitioners and academics now use the word *analytics* in place of BI. Although many authors and consultants have defined it slightly differently, one can view **analytics** as the process of developing actionable decisions or recommendations for actions based on insights generated from historical data. According to the Institute for Operations Research and Management Science (INFORMS), analytics represents the combination of computer technology, management science techniques, and statistics to solve real problems. Of course, many other organizations have proposed their own interpretations and motivations for analytics. For example, SAS Institute Inc. proposed eight levels of analytics that begin with standardized reports from a computer system. These reports essentially provide a sense of what is happening with an organization. Additional technologies have enabled us to create more customized reports that can be generated on an ad hoc basis. The next extension of reporting takes us to OLAP-type queries that allow a user to dig deeper and determine specific sources of concern or opportunities. Technologies available today can also automatically issue alerts for a decision maker when performance warrants such alerts. At a consumer level, we see such alerts for weather or other issues. But similar alerts can also be generated in specific settings when sales fall above or below a certain level within a certain time period or when the inventory for a specific product is running low. All of these applications are made possible through analysis and queries of data being collected by an organization. The next level of analysis might entail statistical analysis to better understand patterns. These can then be taken a step further to develop forecasts or models for predicting how customers might respond to a specific marketing campaign or ongoing service/product offerings. When an organization has a good view of what is happening and what is likely to happen, it can also employ other techniques to make the best decisions under the circumstances.

This idea of looking at all the data to understand what is happening, what will happen, and how to make the best of it has also been encapsulated by INFORMS in proposing three levels of analytics. These three levels are identified as descriptive, predictive, and prescriptive. Figure 1.9 presents a graphical view of these three levels of analytics. It suggests that these three are somewhat independent steps and one type of analytics

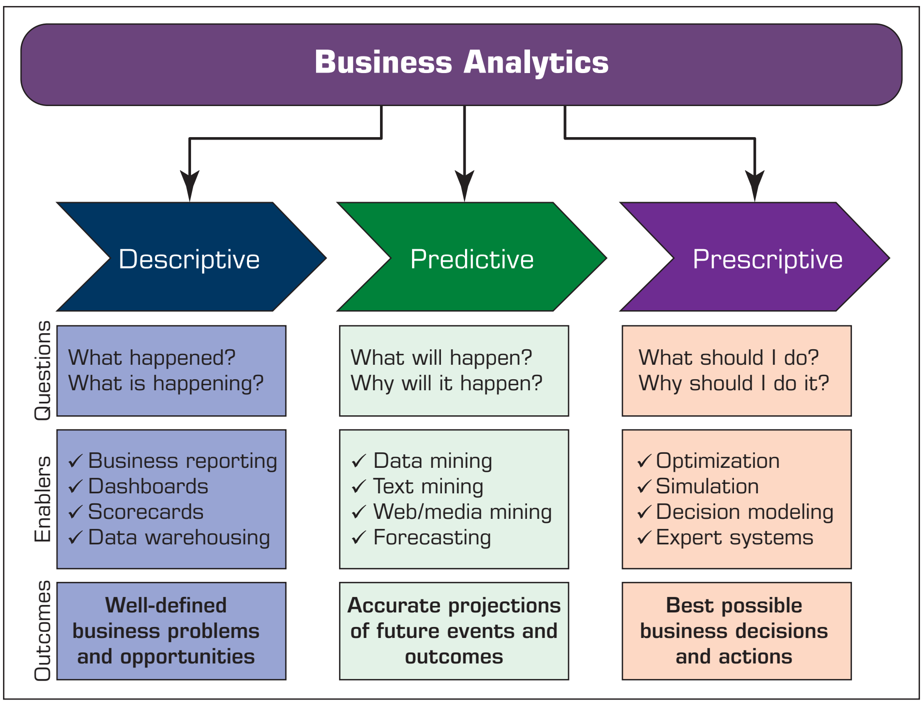


FIGURE 1.9 Three Types of Analytics.

applications leads to another. It also suggests that there is actually some overlap across these three types of analytics. In either case, the interconnected nature of different types of analytics applications is evident. We next introduce these three levels of analytics.

Descriptive Analytics

Descriptive (or reporting) analytics refers to knowing what is happening in the organization and understanding some underlying trends and causes of such occurrences. First, this involves the consolidation of data sources and availability of all relevant data in a form that enables appropriate reporting and analysis. Usually, the development of this data infrastructure is part of DWs. From this data infrastructure, we can develop appropriate reports, queries, alerts, and trends using various reporting tools and techniques.

A significant technology that has become a key player in this area is visualization. Using the latest visualization tools in the marketplace, we can now develop powerful insights in the operations of our organization. Two examples of the application of advanced visualization tools in business operations, available to both private and public companies, are analyzed in the Application Cases 1.3 and 1.4.

Application Case 1.3

An Post and the Use of Data Visualization in Daily Postal Operations

An Post, the state-owned corporation that manages postal services in the Republic of Ireland, presents an interesting and well-documented case of successful innovation in the public sector, which is often plagued by inefficiency and lackluster performance. Established in 1984, An Post is now one of Ireland's largest public employers. For most of its history, An Post's operations did not meet the breakeven point, let alone make a profit. This only changed in 2014, after eight years of losses.

This is why it came as a surprise when An Post made the headlines in 2018 as one of the rare public services to have successfully adopted advanced technological solutions to address customers' orders and feedback and make them significantly more manageable and streamlined.

Interaction with customers can be particularly tricky and time-consuming, and this is where platform innovation and visualization tools can help. For An Post, this service is provided by the Oracle Analytics Cloud platform and An Post's own Customer Experience Management portal. Oracle Analytics Cloud is one of the most comprehensive cloud analytics platforms, offering self-service visualization, powerful inline data preparation to enterprise reporting, and self-learning analytics.

Today, An Post's portal allows a select few corporate customers, including Amazon, Sky TV, and the Inditex Group (which owns brands such as Massimo Dutti and Zara), to visualize their parcel traffic on An Post's cloud service. There are plans to make this data visualization tool available to more of An Post's customers.

The move to data visualization services was so successful that the state-owned public service was honored with the 2018 Analytics Innovation—Global Excellence Award, celebrated in October 2018 at the Oracle Open World ceremony in San Francisco.

Sources: An Post Media centre. "Oracle Says 'An Post You're a Winner'." **https://www.anpost.com/Media-Centre/News/Oracle-says-An-Post-you-re-a-winner.** Oracle.com. "Oracle Analytics Cloud Overview." **https://www.oracle.com/technetwork/middleware/oac/overview/index.html.**

Questions for Case 1.3

1. Why was Oracle's system important for An Post?
2. What additional challenges does a state-owned service face when adopting innovative solutions?

What We Can Learn from This Application Case

While the adoption of innovation technology might at times appear daunting, it is always advantageous to streamline processes. As the case shows, public services can also benefit from adopting them in their operations.

Application Case 1.4

Siemens Reduces Cost with the Use of Data Visualization

Siemens is a German company headquartered in Berlin, Germany. It is one of the world's largest companies focusing on the areas of electrification, automation, and digitalization. It has an annual revenue of 76 billion euros.

The visual analytics group of Siemens is tasked with end-to-end reporting solutions and consulting for all of Siemens internal BI needs. This group was facing the challenge of providing reporting solutions to the entire Siemens organization across different departments while maintaining a balance between governance and self-service capabilities. Siemens needed a platform that could analyze its multiple cases of customer satisfaction surveys, logistic processes, and financial reporting. This platform should be easy to use for their employees so that they could use these data for analysis and decision making. In addition, the platform should be easily integrated with existing Siemens systems and give employees a seamless user experience.

Siemens started using Dundas BI, a leading global provider of BI and data visualization solutions. It allowed Siemens to create highly interactive dashboards that enabled it to detect issues early and thus save a significant amount of money. The dashboards developed by Dundas BI helped Siemens global logistics organization answer questions like how different supply rates at different locations affect the operation, thus helping the company reduce cycle time by 12 percent and scrap cost by 25 percent.

Questions for Case 1.4

1. What challenges were faced by Siemens visual analytics group?
2. How did the data visualization tool Dundas BI help Siemens in reducing cost?

What We Can Learn from This Application Case

Many organizations want tools that can be used to analyze data from multiple divisions. These tools can help them improve performance and make data discovery transparent to their users so that they can identify issues within the business easily.

Sources: **Dundas.com**. "How Siemens Drastically Reduced Cost with Managed BI Applications." **https://www.dundas.com/Content/pdf/siemens-case-study.pdf** (accessed September 2018); Wikipedia.org. "SIEMENS." **https://en.wikipedia.org/wiki/Siemens** (accessed September 2018); **Siemens.com**. "About Siemens." **http://www.siemens.com/about/en/** (accessed September 2018).

Predictive Analytics

Predictive analytics aims to determine what is likely to happen in the future. This analysis is based on statistical techniques as well as other more recently developed techniques that fall under the general category of **data mining**. The goal of these techniques is to be able to predict whether the customer is likely to switch to a competitor ("churn"), what and how much the customer would likely buy next, what promotions the customer would respond to, whether the customer is a creditworthy risk, and so forth. A number of techniques are used in developing predictive analytical applications, including various classification algorithms. For example, as described in Chapters 4 and 5, we can use classification techniques such as logistic regression, decision tree models, and neural networks to predict how well a motion picture will do at the box office. We can also use clustering algorithms for segmenting customers into different clusters to be able to target specific promotions to them. Finally, we can use association mining techniques (Chapters 4 and 5) to estimate relationships between different purchasing behaviors. That is, if a customer buys one product, what else is the customer likely to purchase? Such analysis can assist a retailer in recommending or promoting related products. For example, any product search on **Amazon.com** results in the retailer also suggesting similar products that a customer may be interested in. We will study these techniques and their applications in Chapters 3 through 6. Application Case 1.5 illustrates how Asian companies have made use of predictive analytics.

Application Case 1.5 SagaDigits and the Use of Predictive Analytics

Predictive analytics is widely held to be the most actionable form of business intelligence. As IBM famously stated, if business can be considered a "numbers game," predictive analytics is the way this game is best played and won.

Many companies in China and Hong Kong are increasingly using data mining and predictive analytics to better cater to their customers' needs. This has led to the growth, in the last ten years, of enterprises specializing in these IT solutions. Among them is the award-winning SagaDigits Group.

Incorporated in 2016 in Hong Kong, the group consists of two sister organizations that work together to offer services to Asian businesses. SagaDigits Limited provides data mining, cleansing, extraction, and analytics services based on a series of methods, including natural language processing, big data, and AI technologies. Compathnion Technology Limited specializes instead in data collection, visual recognition, predictive analytics, and statistical modeling for indoor and outdoor uses.

One of the most interesting scalable solutions that SagaDigits offers its customers is Smart Box, an original, highly configurable AI solution for online and offline retail stores. When a company adopts this product for its retail business, it gets a real-time zone detection service in selected retail stores to predict the number of shoppers in those specific areas and learn their preferences based on the spatial information collected.

To successfully predict customer behavior, Smart Box employs a mixed set of indicators and information tools, including advanced visual recognition. Its sensors can detect consumers' gender, emotion, and approximate age group with a high level of accuracy. Finally, based on its own behavioral models' predictions and the historical records of similar customers' transaction histories, Smart Box provides automatic recommendations for advertisement and product selection for display.

Smart Box is one among many of SagaDigits' solutions that use predictive analytics. Another system is Smart User Pro, which uses a variety of public data and internal data to predict various marketing trends in retail and marketing for consumer goods.

Sources: Eric Seigel (2015). "Seven Reasons You Need Predictive Analytics Today." Prediction Impact, Inc. **https://www.ibm.com/downloads/cas/LKMPR8AJ** (accessed October 2019). Saga Digits. **https://sagadigits.com/about** (accessed October 2019). **https://sagadigits.com** (accessed October 2019).

Questions for Case 1.5

1. Why is predictive analytics becoming increasingly common?
2. What is the most interesting feature of Smart Box?
3. To which kind of corporate organization is Smart Box targeted?
4. Describe alternative uses of predictive analytics that Saga Digits has developed solutions for.

What We Can Learn from This Application Case

Innovative IT solutions and sophisticated tools such as predictive analytics are being increasingly used across the world. East Asia was one of the first regions to adopt them, especially places such as Hong Kong, which has many links with North American universities focused on technology and computer science. Saga Digits is one of many companies that offer predictive analytics for consumers' behaviors and future marketing trends.

Prescriptive Analytics

The third category of analytics is termed **prescriptive analytics**. The goal of prescriptive analytics is to recognize what is going on as well as the likely forecast and make decisions to achieve the best performance possible. This group of techniques has historically been studied under the umbrella of OR or management sciences and is generally aimed at optimizing the performance of a system. The goal here is to provide a decision or a recommendation for a specific action. These recommendations can be in the form of a specific yes/no decision for a problem, a specific amount (say, price for a specific item or airfare to charge), or a complete set of production plans. The decisions may be presented to a decision maker in a report or may be used directly in an automated decision rules system (e.g., in airline pricing systems). Thus, these types of analytics can also be termed **decision or normative analytics**. Application Case 1.6 gives an example of such prescriptive analytic applications. We will learn about some aspects of prescriptive analytics in Chapter 8.

ANALYTICS APPLIED TO DIFFERENT DOMAINS Applications of analytics in various industry sectors have spawned many related areas or at least buzzwords. It is almost fashionable to attach the word *analytics* to any specific industry or type of data. Besides the general category of text analytics—aimed at getting value out of text (to be studied in Chapter 7)—or Web analytics—analyzing Web data streams (also in

Application Case 1.6

A Specialty Steel Bar Company Uses Analytics to Determine Available-to-Promise Dates

This application case is based on a project that involved one of the coauthors. A company that does not wish to disclose its name (or even its precise industry) was facing a major problem of making decisions on which inventory of raw materials to use to satisfy which customers. This company supplies custom configured steel bars to its customers. These bars may be cut into specific shapes or sizes and may have unique material and finishing requirements. The company procures raw materials from around the world and stores them in its warehouse. When a prospective customer calls the company to request a quote for the specialty bars meeting specific material requirements (composition, origin of the metal, quality, shapes, sizes, etc.), the salesperson usually has just a little bit of time to submit such a quote including the date when the product can be delivered, prices, and so on. It must make available-to-promise (ATP) decisions, which determine in real time the dates when the salesperson can promise delivery of products that customers requested during the quotation stage. Previously, a salesperson had to make such decisions by analyzing reports on available inventory of raw materials. Some of the available raw material may have already been committed to another customer's order. Thus, the inventory in stock might not really be inventory available. On the other hand, there may be raw material that is expected to be delivered in the near future that could also be used for satisfying the order from this prospective customer. Finally, there might even be an opportunity to charge a premium for a new order by repurposing previously committed inventory to satisfy this new order while delaying an already committed order. Of course, such decisions should be based on the cost–benefit analyses of delaying a previous order. The system should thus be able to pull real-time data about inventory, committed orders, incoming raw material, production constraints, and so on.

To support these ATP decisions, a real-time DSS was developed to find an optimal assignment of the available inventory and to support additional what-if analysis. The DSS uses a suite of mixed-integer programming models that are solved using commercial software. The company has incorporated the DSS into its enterprise resource planning system to seamlessly facilitate its use of business analytics.

Questions for Case 1.6

1. Why would reallocation of inventory from one customer to another be a major issue for discussion?
2. How could a DSS help make these decisions?

Source: M. Pajouh Foad, D. Xing, S. Hariharan, Y. Zhou, B. Balasundaram, T. Liu, & R. Sharda, R. (2013). "Available-to-Promise in Practice: An Application of Analytics in the Specialty Steel Bar Products Industry." *Interfaces, 43*(6), pp. 503–517. **http://dx.doi.org/10.1287/inte.2013.0693** (accessed September 2018).

Chapter 7)—many industry- or problem-specific analytics professions/streams have been developed. Examples of such areas are marketing analytics, retail analytics, fraud analytics, transportation analytics, health analytics, sports analytics, talent analytics, behavioral analytics, and so forth. For example, we will soon see several applications in *sports analytics*. The next section will introduce health analytics and market analytics broadly. Literally, any systematic analysis of data in a specific sector is being labeled as "(fill-in-blanks)" analytics. Although this may result in overselling the concept of analytics, the benefit is that more people in specific industries are aware of the power and potential of analytics. It also provides a focus to professionals developing and applying the concepts of analytics in a vertical sector. Although many of the techniques to develop analytics applications may be common, there are unique issues within each vertical segment that influence how the data may be collected, processed, analyzed, and the applications implemented. Thus, the differentiation of analytics based on a vertical focus is good for the overall growth of the discipline.

ANALYTICS OR DATA SCIENCE? Even as the concept of analytics is receiving more attention in industry and academic circles, another term has already been introduced and is becoming popular. The new term is *data science*. Thus, the practitioners of data science are data scientists. D. J. Patil of LinkedIn is sometimes credited with creating the term *data science*. There have been some attempts to describe the differences between data analysts and data scientists (e.g., see "Data Science Revealed," 2018) (**emc.com/collateral/about/news/emc-data-science-study-wp.pdf**). One view is that *data analyst* is just another term for professionals who were doing BI in the form of data compilation, cleaning, reporting, and perhaps some visualization. Their skill sets included Excel use, some SQL knowledge, and reporting. You would recognize those capabilities as descriptive or reporting analytics. In contrast, data scientists are responsible for predictive analysis, statistical analysis, and use of more advanced analytical tools and algorithms. They may have a deeper knowledge of algorithms and may recognize them under various labels—data mining, knowledge discovery, or machine learning. Some of these professionals may also need deeper programming knowledge to be able to write code for data cleaning/analysis in current Web-oriented languages such as Java or Python and statistical languages such as R. Many analytics professionals also need to build significant expertise in statistical modeling, experimentation, and analysis. Again, our readers should recognize that these fall under the predictive and prescriptive analytics umbrella. However, prescriptive analytics also includes more significant expertise in OR including optimization, simulation, and decision analysis. Those who cover these fields are more likely to be called *data scientists* than *analytics professionals*.

Our view is that the distinction between analytics professional and data scientist is more of a degree of technical knowledge and skill sets than functions. It may also be more of a distinction across disciplines. Computer science, statistics, and applied mathematics programs appear to prefer the data science label, reserving the analytics label for more business-oriented professionals. As another example of this, applied physics professionals have proposed using *network science* as the term for describing analytics that relate to groups of people—social networks, supply chain networks, and so forth. See **http://barabasi.com/networksciencebook/** for an evolving textbook on this topic.

Aside from a clear difference in the skill sets of professionals who only have to do descriptive/reporting analytics versus those who engage in all three types of analytics, the distinction between the two labels is fuzzy at best. We observe that graduates of our analytics programs tend to be responsible for tasks that are more in line with data science professionals (as defined by some circles) than just reporting analytics. This book is clearly aimed at introducing the capabilities and functionality of all analytics (which include data science), not just reporting analytics. From now on, we will use these terms interchangeably.

WHAT IS BIG DATA? Any book on analytics and data science has to include significant coverage of what is called **Big Data analytics**. We cover it in Chapter 9 but here is a very brief introduction. Our brains work extremely quickly and efficiently and are versatile in processing large amounts of all kinds of data: images, text, sounds, smells, and video. We process all different forms of data relatively easily. Computers, on the other hand, are still finding it hard to keep up with the pace at which data are generated, let alone analyze them quickly. This is why we have the problem of Big Data. So, what is Big Data? Simply put, Big Data refers to data that cannot be stored in a single storage unit. Big Data typically refers to data that come in many different forms: structured, unstructured, in a stream, and so forth. Major sources of such data are clickstreams from Web sites, postings on social media sites such as Facebook, and data from traffic, sensors, or weather. A Web search engine such as Google needs to search and index billions of Web pages to give you relevant search results in a fraction of a second. Although this is not done in real time, generating an index of all the Web pages on the Internet is not an easy task. Luckily for Google, it was able to solve this problem. Among other tools, it has employed Big Data analytical techniques.

There are two aspects to managing data on this scale: storing and processing. If we could purchase an extremely expensive storage solution to store all this at one place on one unit, making this unit fault tolerant would involve a major expense. An ingenious solution was proposed that involved storing these data in chunks on different machines connected by a network—putting a copy or two of this chunk in different locations on the network, both logically and physically. It was originally used at Google (then called the Google File System) and later developed and released by an Apache project as the Hadoop Distributed File System (HDFS).

However, storing these data is only half of the problem. Data are worthless if they do not provide business value, and for them to provide business value, they must be analyzed. How can such vast amounts of data be analyzed? Passing all computation to one powerful computer does not work; this scale would create a huge overhead on such a powerful computer. Another ingenious solution was proposed: Push computation to the data instead of pushing data to a computing node. This was a new paradigm and gave rise to a whole new way of processing data. This is what we know today as the MapReduce programming paradigm, which made processing Big Data a reality. MapReduce was originally developed at Google, and a subsequent version was released by the Apache project called *Hadoop MapReduce*.

Today, when we talk about storing, processing, or analyzing Big Data, HDFS and MapReduce are involved at some level. Other relevant standards and software solutions have been proposed. Although the major toolkit is available as an open source, several companies have been launched to provide training or specialized analytical hardware or software services in this space. Some examples are HortonWorks, Cloudera, and Teradata Aster.

Over the past few years, what was called Big Data changed more and more as Big Data applications appeared. The need to process data coming in at a rapid rate added velocity to the equation. An example of fast data processing is algorithmic trading. This uses electronic platforms based on algorithms for trading shares on the financial market, which operates in microseconds. The need to process different kinds of data added variety to the equation. Another example of a wide variety of data is sentiment analysis, which uses various forms of data from social media platforms and customer responses to gauge sentiments. Today, Big Data is associated with almost any kind of large data that have the characteristics of volume, velocity, and variety. As noted before, these are evolving quickly to encompass stream analytics, IoT, cloud computing, and deep learning–enabled AI. We will study these in various chapters in the book.

SECTION 1.5 REVIEW QUESTIONS

1. Define *analytics*.
2. What is descriptive analytics? What are the various tools that are employed in descriptive analytics?
3. How is descriptive analytics different from traditional reporting?
4. What is a DW? How can DW technology help enable analytics?
5. What is predictive analytics? How can organizations employ predictive analytics?
6. What is prescriptive analytics? What kinds of problems can be solved by prescriptive analytics?
7. Define *modeling* from the analytics perspective.
8. Is it a good idea to follow a hierarchy of descriptive and predictive analytics before applying prescriptive analytics?
9. How can analytics aid in objective decision making?
10. What is Big Data analytics?
11. What are the sources of Big Data?
12. What are the characteristics of Big Data?
13. What processing technique is applied to process Big Data?

1.6 ANALYTICS EXAMPLES IN SELECTED DOMAINS

You will see examples of analytics applications throughout various chapters. That is one of the primary approaches (exposure) of this book. In this section, we highlight three application areas—sports, healthcare, and retail—where there have been the most reported applications and successes.

Sports Analytics—An Exciting Frontier for Learning and Understanding Applications of Analytics

The application of analytics to business problems is a key skill, one that you will learn in this book. Many of these techniques are now being applied to improve decision making in all aspects of sports, a very hot area called *sports analytics*. It is the art and science of gathering data about athletes and teams to create insights that improve sports decisions, such as deciding which players to recruit, how much to pay them, who to play, how to train them, how to keep them healthy, and when they should be traded or retired. For teams, it involves business decisions such as ticket pricing as well as roster decisions, analysis of each competitor's strengths and weaknesses, and many game-day decisions.

Indeed, sports analytics is becoming a specialty within analytics. It is an important area because sport is a big business, generating about $145 billion in revenues each year plus an additional $100 billion in legal and $300 billion in illegal gambling, according to Price Waterhouse ("Changing the Game: Outlook for the Global Sports Market to 2015" (2015)). In 2014, only $125 million was spent on analytics (less than 0.1 percent of revenues). This is expected to grow at a healthy rate to $4.7 billion by 2021 ("Sports Analytics Market Worth $4.7B by 2021" (2015)).

The use of analytics for sports was popularized by the *Moneyball* book by Michael Lewis in 2003 and the movie starring Brad Pitt in 2011. It showcased Oakland A's general manager Billy Beane and his use of data and analytics to turn a losing team into a winner. In particular, he hired an analyst who used analytics to draft players who were able to get on base as opposed to players who excelled at traditional measures like runs batted in or stolen bases. These insights allowed the team to draft prospects overlooked by other teams at reasonable starting salaries. It worked—the team made it to the playoffs in 2002 and 2003.

Now analytics are being used in all parts of sports. The analytics can be divided between the front office and back office. A good description with 30 examples appears in Tom Davenport's survey article (). Front-office business analytics include analyzing fan behavior ranging from predictive models for season ticket renewals and regular ticket sales to scoring tweets by fans regarding the team, athletes, coaches, and owners. This is very similar to traditional CRM. Financial analysis is also a key area such as when salary cap (for pros) or scholarship (for colleges) limits are part of the equation.

Back-office uses include analysis of both individual athletes and team play. For individual players, there is a focus on recruitment models and scouting analytics, analytics for strength and fitness as well as development, and PMs for avoiding overtraining and injuries. Concussion research is a hot field. Team analytics include strategies and tactics, competitive assessments, and optimal roster choices under various on-field or on-court situations.

The following representative examples illustrate how two sports organizations use data and analytics to improve sports operations in the same way that analytics have improved traditional industry decision making.

Example 1: The Business Office

Dave Ward works as a business analyst for a major pro baseball team, focusing on revenue. He analyzes ticket sales, both from season ticket holders and single-ticket buyers. Sample questions in his area of responsibility include why season ticket holders renew (or do not renew) their tickets as well as what factors drive last-minute individual seat ticket purchases. Another question is how to price the tickets.

Some of the analytical techniques Dave uses include simple statistics on fan behavior such as overall attendance and answers to survey questions about likelihood to purchase again. However, what fans say versus what they do can be different. Dave runs a survey of fans by ticket seat location ("tier") and asks about their likelihood of renewing their season tickets. But when he compares what they say versus what they do, he discovers big differences. (See Figure 1.10.) He found that 69 percent of fans in Tier 1 seats who said on the survey that they would "probably not renew" actually did. This

Tier	Highly Likely	Likely	Maybe	Probably Not	Certainly Not
1	92	88	75	69	45
2	88	81	70	65	38
3	80	76	68	55	36
4	77	72	65	45	25
5	75	70	60	35	25

FIGURE 1.10 **Season Ticket Renewals—Survey Scores.**

is useful insight that leads to action—customers in the green cells are the most likely to renew tickets and so require fewer marketing touches and dollars to convert compared to customers in the blue cells.

However, many factors influence fan ticket purchase behavior, especially price, which drives more sophisticated statistics and data analysis. For both areas, but especially single-game tickets, Dave is driving the use of dynamic pricing—moving the business from simple static pricing by seat location tier to day-by-day up-and-down pricing of individual seats. This is a rich research area for many sports teams and has huge upside potential for revenue enhancement. For example, his pricing takes into account the team's record, who they are playing, game dates and times, which star athletes play for each team, each fan's history of renewing season tickets or buying single tickets, and factors such as seat location, number of seats, and real-time information like traffic congestion historically at game time and even the weather. See Figure 1.11.

Which of these factors are important and by how much? Given his extensive statistics background, Dave builds regression models to pick out key factors driving these historic behaviors and create PMs to identify how to spend marketing resources to drive revenues. He builds churn models for season ticket holders to create segments of customers who will renew, will not renew, or are fence-sitters, which then drives more refined marketing campaigns.

In addition, Dave does sentiment scoring on fan comments such as tweets that help him segment fans into different loyalty segments. Other studies about single-game attendance drivers help the marketing department understand the impact of giveaways like bobble-heads or T-shirts or suggestions on where to make spot TV ad buys.

Beyond revenues, there are many other analytical areas that Dave's team works on, including merchandising, TV and radio broadcast revenues, inputs to the general manager on salary negotiations, draft analytics especially given salary caps, promotion effectiveness including advertising channels, and brand awareness, as well as partner analytics. He's a very busy guy!

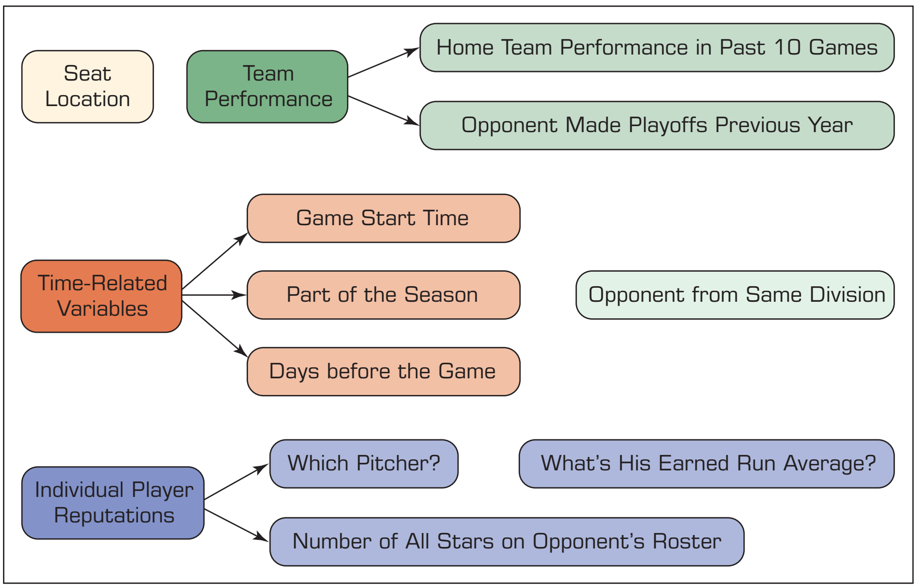


FIGURE 1.11 Dynamic Pricing Previous Work—Major League Baseball. *Source:* Based on C. Kemper and C. Breuer, "How Efficient is Dynamic Pricing for Sports Events? Designing a Dynamic Pricing Model for Bayern Munich", *Intl. Journal of Sports Finance*, 11, pp. 4–25, 2016.

Example 2: The Coach

Bob Breedlove is the football coach for a major college team. For him, everything is about winning games. His areas of focus include recruiting the best high school players, developing them to fit his offense and defense systems, and getting maximum effort from them on game days. Sample questions in his area of responsibility include: Whom do we recruit? What drills help develop their skills? How hard do I push our athletes? Where are opponents strong or weak, and how do we figure out their play tendencies?

Fortunately, his team has hired a new team operations expert, Dar Beranek, who specializes in helping the coaches make tactical decisions. She is working with a team of student interns who are creating opponent analytics. They used the coach's annotated game film to build a cascaded decision tree model (Figure 1.12) to predict whether the next play will be a running play or passing play. For the defensive coordinator, they have built heat maps (Figure 1.13) of each opponent's passing offense, illustrating their tendencies to throw left or right and into which defensive coverage zones. Finally, they built some time-series analytics (Figure 1.14) on explosive plays (defined as a gain of more than 16 yards for a passing play or more than 12 yards for a run play). For each play, they compare the outcome with their own defensive formations and the other team's offensive formations, which help Coach Breedlove react more quickly to formation shifts during a game. We explain the analytical techniques that generated these figures in much more depth in Chapters 3–6 and Chapter 9.

New work that Dar is fostering involves building better high school athlete recruiting models. For example, each year the team gives scholarships to three students who are wide receiver recruits. For Dar, picking out the best players goes beyond simple

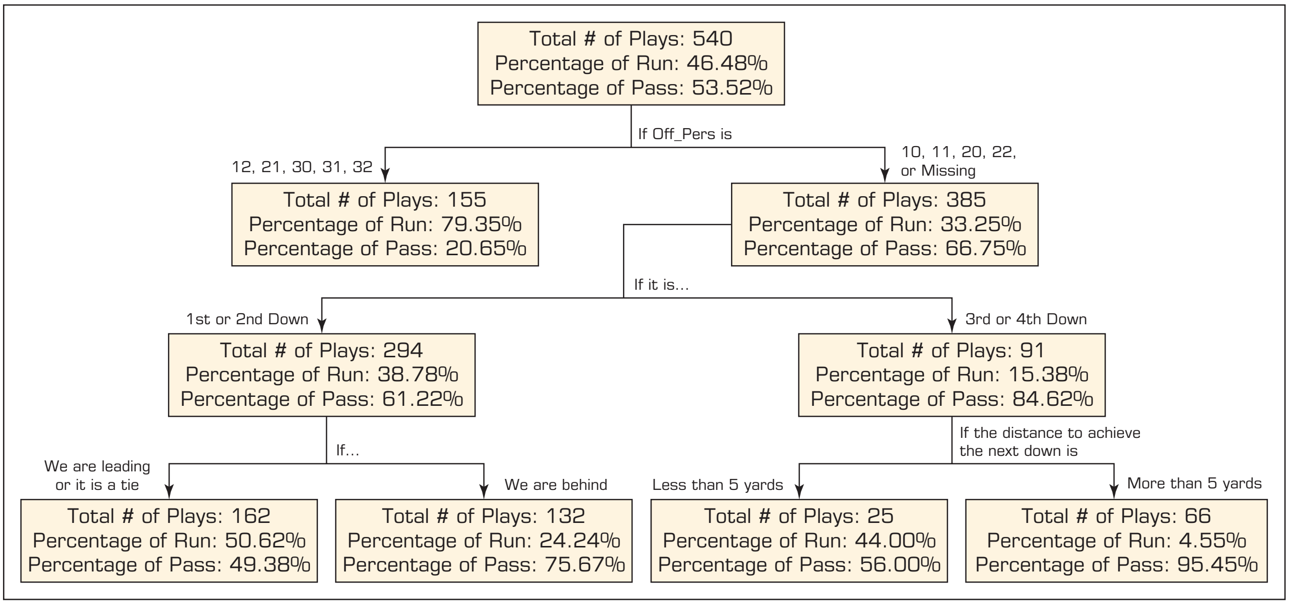


FIGURE 1.12 Cascaded Decision Tree for Run or Pass Plays. *Source:* Contributed by Dr. Dave Schrader, who retired after 24 years in advanced development and marketing at Teradata. He has remained on the Board of Advisors of the Teradata University Network, where he spends his retirement helping students and faculty learn more about sports analytics. Graphics by Peter Liang and Jacob Pearson, graduate students at Oklahoma State University, as part of a student project in the spring of 2016 in Prof. Ramesh Sharda's class under Dr. Dave Schrader's coaching.

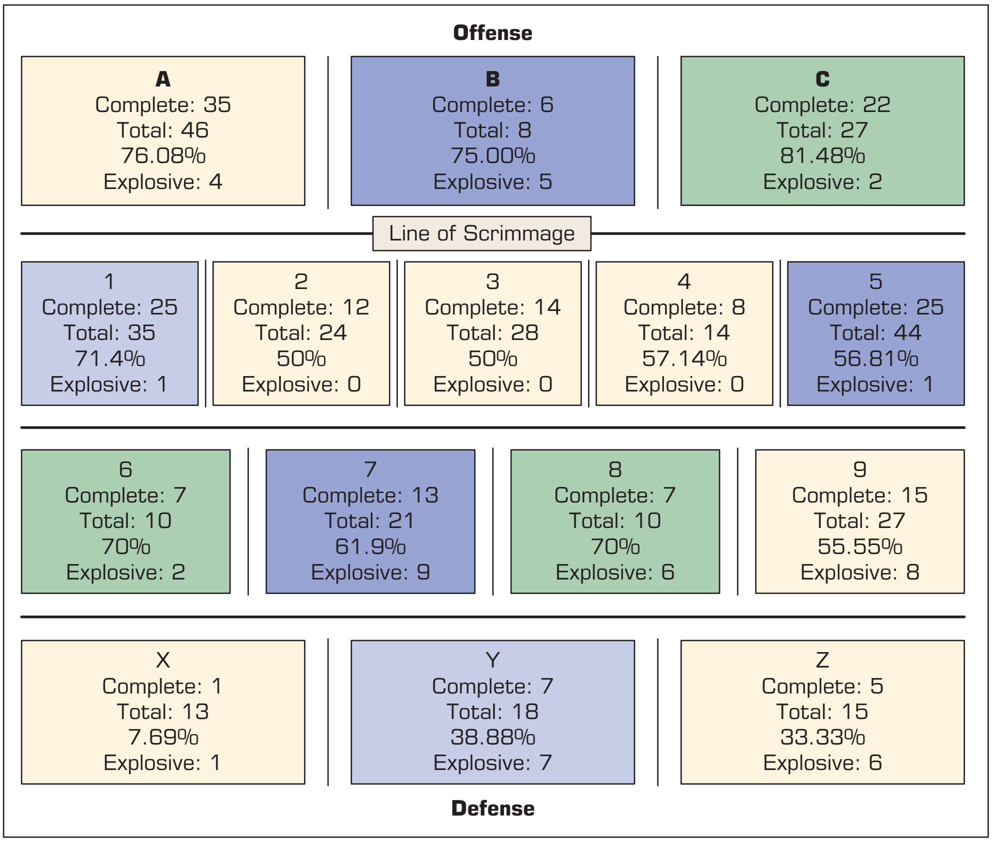


FIGURE 1.13 Heat Map Zone Analysis for Passing Plays. *Source:* Contributed by Dr. Dave Schrader, who retired after 24 years in advanced development and marketing at Teradata. He has remained on the Board of Advisors of the Teradata University Network, where he spends his retirement helping students and faculty learn more about sports analytics. Graphics by Peter Liang and Jacob Pearson, graduate students at Oklahoma State University, as part of a student project in the spring of 2016 in Prof. Ramesh Sharda's class under Dr. Dave Schrader's coaching.

measures like how fast athletes run, how high they jump, or how long their arms are to newer criteria like how quickly they can rotate their heads to catch a pass, what kinds of reaction times they exhibit to multiple stimuli, and how accurately they run pass routes. Some of her ideas illustrating these concepts appear on the TUN Web site; look for the Business Scenario Investigation (2015) "The Case of Precision Football."

WHAT CAN WE LEARN FROM THESE EXAMPLES? Beyond the front-office business analysts, the coaches, trainers, and performance experts, there are many other people in sports who use data, ranging from golf groundskeepers who measure soil and turf conditions for PGA tournaments to baseball and basketball referees who are rated on the correct and incorrect calls they make. In fact, it is hard to find an area of sports that is *not* being impacted by the availability of more data, especially from sensors.

Skills you will learn in this book for business analytics will apply to sports. If you want to dig deeper into this area, we encourage you to look at the Sports Analytics section of the TUN, a free resource for students and faculty. On its Web site, you will find descriptions of what to read to find out more about sports analytics, compilations of places where you can find publically available data sets for analysis, as well as examples

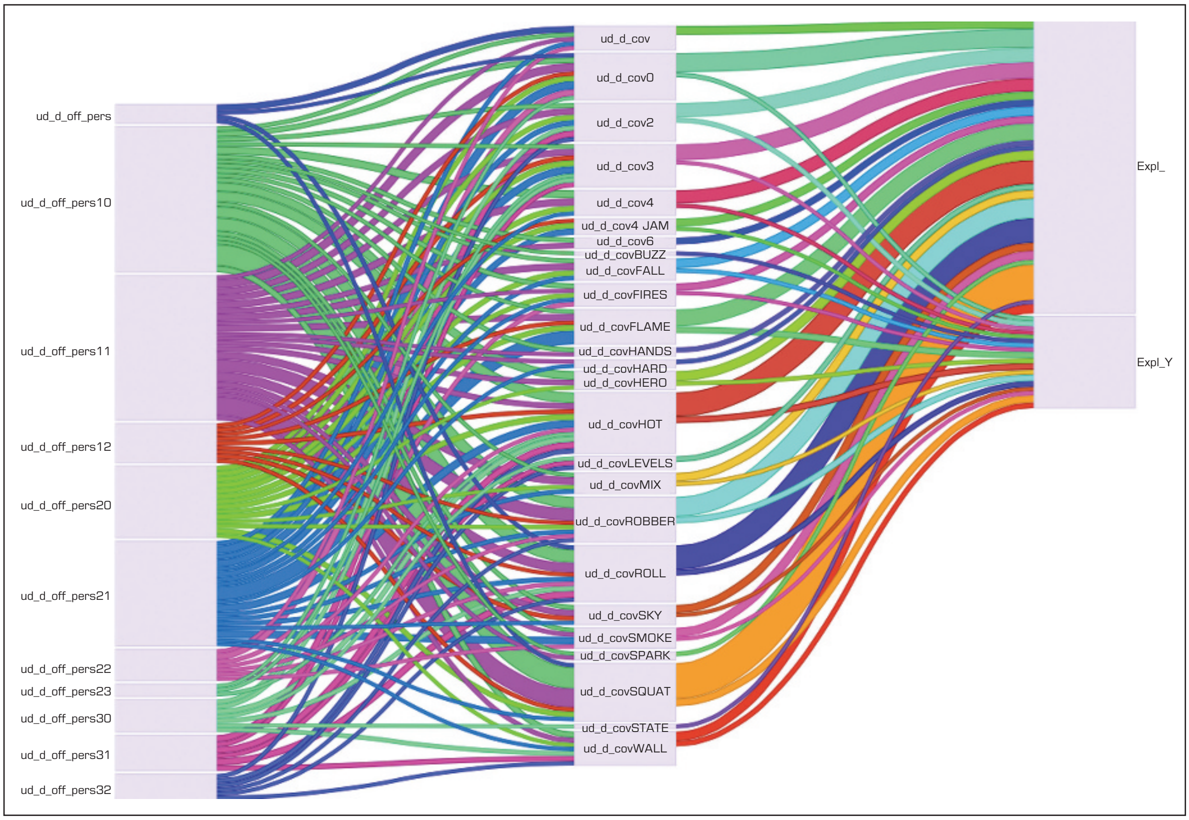


FIGURE 1.14 **Time-Series Analysis of Explosive Plays.**

of student projects in sports analytics and interviews of sports professionals who use data and analytics to do their jobs. Good luck learning analytics!

Analytics Applications in Healthcare—Humana Examples

Although healthcare analytics span a wide variety of applications from prevention to diagnosis to efficient operations and fraud prevention, we focus on some applications that have been developed at a major health insurance company in the United States, Humana. According to its Web site, "The company's strategy integrates care delivery, the member experience, and clinical and consumer insights to encourage engagement, behavior change, proactive clinical outreach and wellness. . . ." Achieving these strategic goals includes significant investments in information technology in general and analytics in particular. Brian LeClaire is senior vice president and CIO of Humana. He has a PhD in MIS from Oklahoma State University. He has championed analytics as a competitive differentiator at Humana—including cosponsoring the creation of a center for excellence in analytics. He described the following projects as examples of Humana's analytics initiatives, led by Humana's chief clinical analytics officer, Vipin Gopal.

Humana Example 1: Preventing Falls in a Senior Population—An Analytic Approach

Accidental falls are a major health risk for adults age 65 years and older with one-third experiencing a fall every year.[1] The costs of falls pose a significant strain on the U.S. healthcare system; the direct costs of falls were estimated at $34 billion in 2013 alone.[1]

With the percent of seniors in the U.S. population on the rise, falls and associated costs are anticipated to increase. According to the Centers for Disease Control and Prevention (CDC), "Falls are a public health problem that is largely preventable" (**www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html**).[1] Falls are also the leading factor for both fatal and nonfatal injuries in older adults with injurious falls increasing the risk of disability by up to 50 percent (Gill et al., 2013).[2] Humana is the nation's second-largest provider of Medicare Advantage benefits with approximately 3.2 million members, most of whom are seniors. Keeping its senior members well and helping them live safely at their homes is a key business objective of which prevention of falls is an important component. However, no rigorous methodology was available to identify individuals most likely to fall, for whom falls prevention efforts would be beneficial. Unlike chronic medical conditions such as diabetes and cancer, a fall is not a well-defined medical condition. In addition, falls are usually underreported in claims data as physicians typically tend to code the consequence of a fall such as fractures and dislocations. Although many clinically administered assessments to identify fallers exist, they have limited reach and lack sufficient predictive power (Gates et al., 2008).[3] As such, there is a need for a prospective and accurate method to identify individuals at greatest risk of falling so that they can be proactively managed for fall prevention. The Humana analytics team undertook the development of a Falls Predictive Model in this context. It is the first comprehensive PM reported that utilizes administrative medical and pharmacy claims, clinical data, temporal clinical patterns, consumer information, and other data to identify individuals at high risk of falling over a time horizon.

Today, the Falls PM is central to Humana's ability to identify seniors who could benefit from fall mitigation interventions. An initial proof-of-concept with Humana consumers, representing the top 2 percent of those at the highest risk of falling, demonstrated that the consumers had increased utilization of physical therapy services, indicating consumers are taking active steps to reduce their risk for falls. A second initiative utilizes the Falls PM to identify high-risk individuals for remote monitoring programs. Using the PM, Humana was able to identify 20,000 consumers at a high risk of falling who benefited from this program. Identified consumers wear a device that detects falls and alerts a 24/7 service for immediate assistance.

This work was recognized by the Analytics Leadership Award by Indiana University Kelly School of Business in 2015, for innovative adoption of analytics in a business environment.

Contributors: Harpreet Singh, PhD; Vipin Gopal, PhD; Philip Painter, MD.

Humana Example 2: Humana's Bold Goal—Application of Analytics to Define the Right Metrics

In 2014, Humana, Inc. announced its organization's Bold Goal to improve the health of the communities it serves by 20 percent by 2020 by making it easy for people to achieve their best health. The communities that Humana serves can be defined in many ways, including geographically (state, city, neighborhood), by product (Medicare Advantage, employer-based plans, individually purchased), or by clinical profile (priority conditions including diabetes, hypertension, congestive heart failure [CHF], coronary artery disease [CAD], chronic obstructive pulmonary disease [COPD], or depression). Understanding the health of these communities and how they track over time is critical not only for the evaluation of the goal, but also in crafting strategies to improve the health of the whole membership in its entirety.

A challenge before the analytics organization was to identify a metric that captures the essence of the Bold Goal. Objectively measured traditional health insurance metrics such as hospital admissions or emergency room visits per 1,000 persons would not capture the spirit of this new mission. The goal was to identify a metric that captures health and its improvement in a community and was relevant to Humana as a business. Through rigorous analytic evaluations, Humana eventually selected “Healthy Days,” a four-question, quality-of-life questionnaire originally developed by the CDC to track and measure Humana’s overall progress toward the Bold Goal.

It was critical to make sure that the selected metric was highly correlated to health and business metrics so that any improvement in Healthy Days resulted in improved health and better business results. Some examples of how “Healthy Days” is correlated to metrics of interest include the following:

- Individuals with more unhealthy days (UHDs) exhibit higher utilization and cost patterns. For a five-day increase in UHDs, there are (1) an $82 increase in average monthly medical and pharmacy costs, (2) an increase of 52 inpatient admits per 1,000 patients, and (3) a 0.28-day increase in average length of stay (Havens, Peña, Slabaugh, Cordier, Renda, & Gopal, 2015).[1]
- Individuals who exhibit healthy behaviors and have their chronic conditions well managed have fewer UHDs. For example, when we look at individuals with diabetes, UHDs are lower if they obtained an LDL screening (−4.3 UHDs) or a diabetic eye exam (−2.3 UHDs). Likewise, if they have controlled blood sugar levels measured by HbA1C (−1.8 UHDs) or LDL levels (−1.3 UHDs) (Havens, Slabaugh, Peña, Haugh, & Gopal 2015).[2]
- Individuals with chronic conditions have more UHDs than those who do not have (1) CHF (16.9 UHDs), (2) CAD (14.4 UHDs), (3) hypertension (13.3 UHDs), (4) diabetes (14.7 UHDs), (5) COPD (17.4 UHDs), or (6) depression (22.4 UHDs) (Havens, Peña, Slabaugh et al., 2015; Chiguluri, Guthikonda, Slabaugh, Havens, Peña, & Cordier, 2015; Cordier et al., 2015).[1,3,4]

Humana has since adopted Healthy Days as their metric for the measurement of progress toward Bold Goal (Humana, **http://populationhealth.humana.com/wp-content/uploads/2016/05/BoldGoal2016ProgressReport_1.pdf**).[5]

Contributors: Tristan Cordier, MPH; Gil Haugh, MS; Jonathan Peña, MS; Eriv Havens, MS; Vipin Gopal, PhD.

Humana Example 3: Predictive Models to Identify the Highest Risk Membership in a Health Insurer

The 80/20 rule generally applies in healthcare; that is, roughly 20 percent of consumers account for 80 percent of healthcare resources due to their deteriorating health and chronic conditions. Health insurers like Humana have typically enrolled the highest-risk enrollees in clinical and disease management programs to help manage the chronic conditions the members have.

Identification of the correct members is critical for this exercise, and in the recent years, PMs have been developed to identify enrollees with high future risk. Many of these PMs were developed with heavy reliance on medical claims data, which results from the medical services that the enrollees use. Because of the lag that exists in submitting and processing claims data, there is a corresponding lag in identification of high-risk members for clinical program enrollment. This issue is especially relevant when new members join a health insurer as they would not have a claims history with an insurer. A claims-based PM could take on average of 9–12 months after enrollment of new members to identify them for referral to clinical programs.

In the early part of this decade, Humana attracted large numbers of new members in its Medicare Advantage products and needed a better way to clinically manage this membership. As such, it became extremely important that a different analytic approach be developed to rapidly and accurately identify high-risk new members for clinical management, to keep this group healthy and costs down.

Humana's Clinical Analytics team developed the New Member Predictive Model (NMPM) that would quickly identify at-risk individuals soon after their new plan enrollments with Humana rather than waiting for sufficient claim history to become available for compiling clinical profiles and predicting future health risk. Designed to address the unique challenges associated with new members, NMPM developed a novel approach that leveraged and integrated broader data sets beyond medical claims data such as self-reported health risk assessment data and early indicators from pharmacy data, employed advanced data mining techniques for pattern discovery, and scored every Medicare Advantage (MA, a specific insurance plan) consumer daily based on the most recent data Humana has to date. The model was deployed with a cross-functional team of analytics, IT, and operations to ensure seamless operational and business integration.

Since NMPM was implemented in January 2013, it has been rapidly identifying high-risk new members for enrollment in Humana's clinical programs. The positive outcomes achieved through this model have been highlighted in multiple senior leader communications from Humana. In the first quarter 2013 earnings release presentation to investors, Bruce Broussard, CEO of Humana, stated the significance of "improvement in new member PMs and clinical assessment processes," which resulted in 31,000 new members enrolled in clinical programs, compared to 4,000 in the same period a year earlier, a 675 percent increase. In addition to the increased volume of clinical program enrollments, outcome studies showed that the newly enrolled consumers identified by NMPM were also referred to clinical programs sooner with over 50 percent of the referrals identified within the first three months after new MA plan enrollments. The consumers identified also participated at a higher rate and had longer tenure in the programs.

Contributors: Sandy Chiu, MS; Vipin Gopal, PhD.

These examples illustrate how an organization explores and implements analytics applications to meet its strategic goals. You will see several other examples of healthcare applications throughout various chapters in the book.

ANALYTICS IN THE RETAIL VALUE CHAIN The retail sector is where you would perhaps see the most applications of analytics. This is the domain where the volumes are large but the margins are usually thin. Customers' tastes and preferences change frequently. Physical and online stores face many challenges to succeed. And market dominance at one time does not guarantee continued success. So investing in learning about your suppliers, customers, employees, and all the stakeholders that enable a retail value chain to succeed and using that information to make better decisions has been a goal of the analytics industry for a long time. Even casual readers of analytics probably know about Amazon's enormous investments in analytics to power their value chain. Similarly, Walmart, Target, and other major retailers have invested millions of dollars in analytics for their supply chains. Most of the analytics technology and service providers have a major presence in retail analytics. Coverage of even a small portion of those applications to achieve our exposure goal could fill a whole book. So this section highlights just a few potential applications. Most of these have been fielded by many retailers and are available through many technology providers, so in this section, we will take a more general view rather than point to specific cases. This general view has been proposed by Abhishek Rathi, CEO of **vCreaTek.com**. vCreaTek, LLC is a boutique analytics software and service company that has offices in India, the United States, the United Arab Emirates (UAE), and Belgium. The company develops applications in multiple domains, but retail analytics is one of its key focus areas.

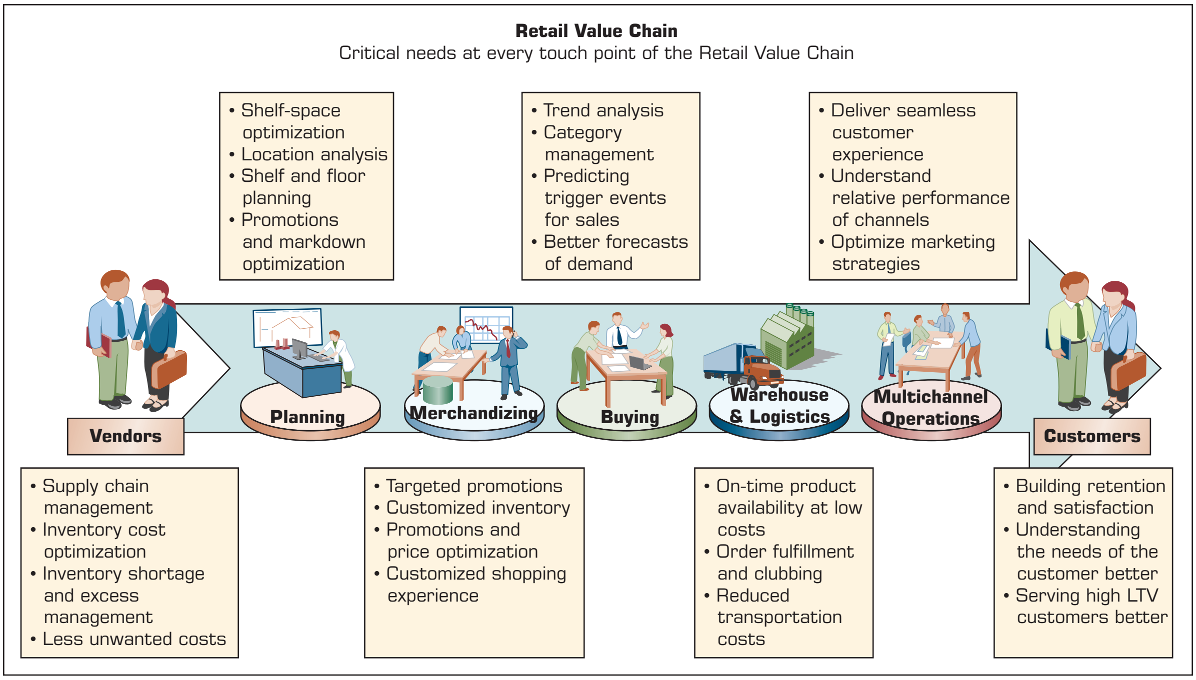


FIGURE 1.15 Example of Analytics Applications in a Retail Value Chain. *Source:* Contributed by Abhishek Rathi, CEO, **vCreaTek.com**.

Figure 1.15 highlights selected components of a retail value chain. It starts with suppliers and concludes with customers but illustrates many intermediate strategic and operational planning decision points where analytics—descriptive, predictive, or prescriptive—can play a role in making better data-driven decisions. Table 1.1 also illustrates some of the important areas of analytics applications, examples of key questions that can be answered through analytics, and of course, the potential business value derived from fielding such analytics. Some examples are discussed next.

An online retail site usually knows its customer as soon as the customer signs in, and thus they can offer customized pages/offerings to enhance the experience. For any retail store, knowing its customer at the store entrance is still a huge challenge. By combining the video analytics and information/badge issued through its loyalty program, the store may be able to identify the customer at the entrance itself and thus enable an extra opportunity for a cross-selling or up-selling. Moreover, a personalized shopping experience can be provided with more customized engagement during the customer's time in the store.

Store retailers invest lots of money in attractive window displays, promotional events, customized graphics, store decorations, printed ads, and banners. To discern the effectiveness of these marketing methods, the team can use shopper analytics by observing closed-circuit television (CCTV) images to figure out the demographic details of the in-store foot traffic. The CCTV images can be analyzed using advanced algorithms to derive demographic details such as age, gender, and mood of the person browsing through the store.

Further, the customer's in-store movement data when combined with shelf layout and planogram can give more insight to the store manager to identify the hot-selling/profitable areas within the store. Moreover, the store manager also can use this information to plan the workforce allocation for those areas for peak periods.