CHAPTER

1

An Overview of Business Intelligence, Analytics, and Data Science

LEARNING OBJECTIVES

- Understand the need for computerized support of managerial decision making
- Recognize the evolution of such computerized support to the current state—analytics/data science
- Describe the business intelligence (BI) methodology and concepts
- Understand the different types of analytics 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.

This book is about using business analytics as computerized support 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. This book presents the fundamentals of the techniques and the manner in which these systems are constructed and used. We follow an EEE approach to introducing these topics: **Exposure**, **Experience**, and **Exploration**. 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 give examples of specific software tools that can be used for developing such applications. 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.

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

1.1 OPENING VIGNETTE: 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. Sports analytics 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 sports is a big business, generating about $145B in revenues each year, plus an additional $100B in legal and $300B in illegal gambling, according to Price Waterhouse.[1] In 2014, only $125M was spent on analytics (less than 0.1% of revenues). This is expected to grow at a healthy rate to $4.7B by 2021.[2]

[1] "Changing the Game: Outlook for the Global Sports Market to 2015," Price Waterhouse Coopers Report, appears at https://www.pwc.com/gx/en/hospitality-leisure/pdf/changing-the-game-outlook-for-the-global-sports-market-to-2015.pdf. Betting data from https://www.capcredit.com/how-much-americansspend-on-sports-each-year/.

[2] "Sports Analytics Market Worth $4.7B by 2021," Wintergreen Research Press Release, covered by PR Newswire at http://www.prnewswire.com/news-releases/sports-analytics-market-worth-47-billion-by-2021-509869871.html, June 25, 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 able to get on base as opposed to players who excelled at traditional measures like runs batted in or stolen bases. These insights allowed them to draft prospects overlooked by other teams at reasonable starting salaries. It worked—they 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.[3] 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 customer relationship management (CRM). Financial analysis is also a key area, where salary caps (for pros) or scholarship limits (colleges) are part of the equation.

Back-office uses include analysis of both individual athletes as well as 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 three sports organizations use data and analytics to improve sports operations, in the same way 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 as well as 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 like 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 1.1.) He found that 69% of fans in Tier 1 seats who said on the

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 1.1 Season Ticket Renewals—Survey Scores.

[3]Thomas Davenport, "Analytics in Sports: The New Science of Winning," International Institute for Analytics White paper, sponsored by SAS, February 2014. On the SAS Web site at: http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/iia-analytics-in-sports-106993.pdf. (Accessed July 2016)

survey that they would "probably not renew" actually did. This is useful insight that leads to action—customers in the green cells are the most likely to renew tickets, so require fewer marketing touches and dollars to convert, for example, 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, as well as factors like seat location, number of seats, and real-time information like traffic congestion historically at game time and even the weather. See Figure 1.2.

Which of these factors are important? 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, won't renew, or are fence-sitters, which then drives more refined marketing campaigns.

In addition, he does sentiment scoring on fan comments like 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!

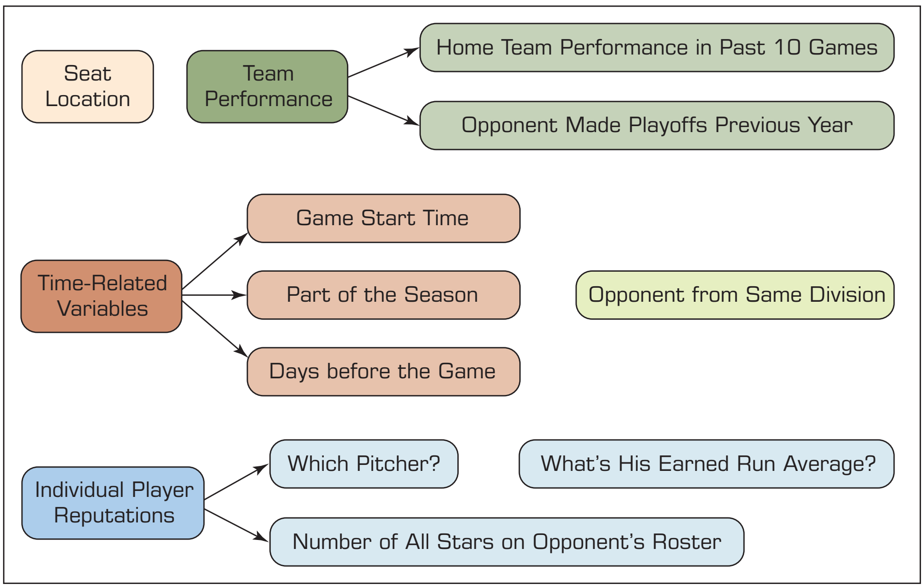

FIGURE 1.2 Dynamic Pricing Previous Work—Major League Baseball. *Source:* Adapted from C. Kemper and C. Breuer, "How Efficient is Dynamic Pricing for Sports Events? Designing a Dynamic Pricing Model for Bayern Munich", *Intl. Journal of Sports Finance*, 11, pp. 4-25, 2016.

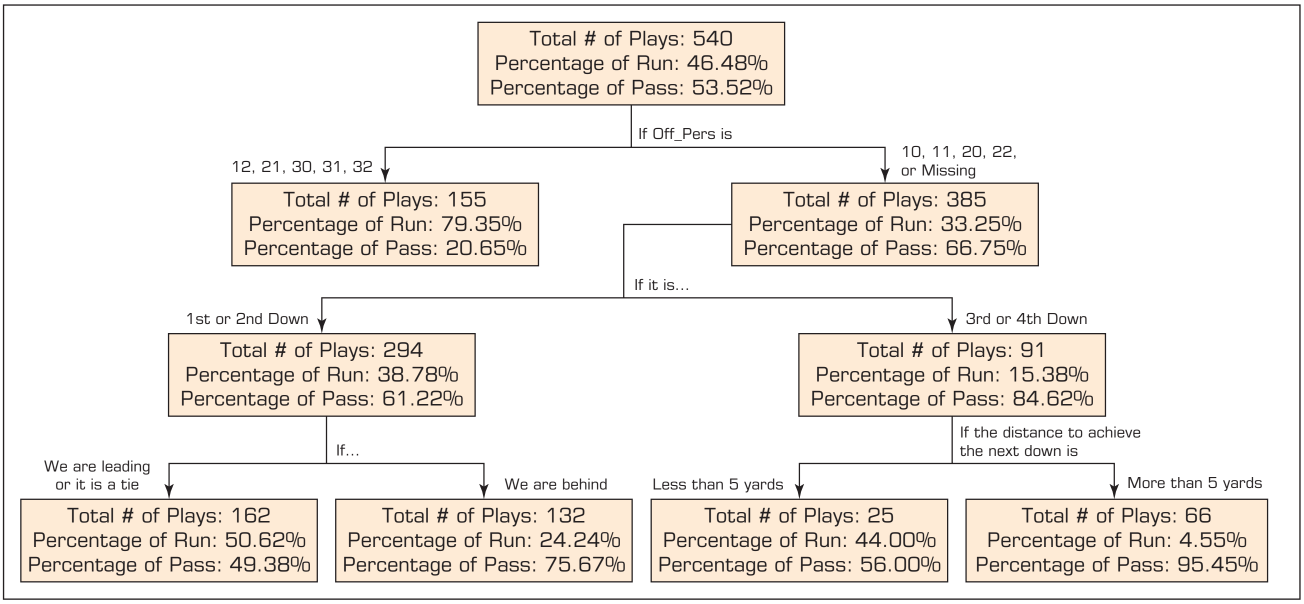

FIGURE 1.3 Cascaded Decision Tree for Run or Pass Plays.

Example 2: The Coach

Bob Breedlove is the football coach for a major college team. For him, it's all 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: Who 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 1.3) to predict whether the next play will be a running play or passing play. For the defensive coordinator, they have built heat maps (Figure 1.4) 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 1.5) 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 helps Coach Breedlove react more quickly to formation shifts during a game. We will explain the analytical techniques that generated these figures in much more depth in Chapters 2–5 and Chapter 7.

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 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 BSI Case of Precision Football.[4]

[4]Business Scenario Investigation BSI: The Case of Precision Football (video). (Fall 2015). Appears on http://www.teradatauniversitynetwork.com/About-Us/Whats-New/BSI–Sports-Analytics—Precision-Football//,Fall 2015. (Accessed September 2016)

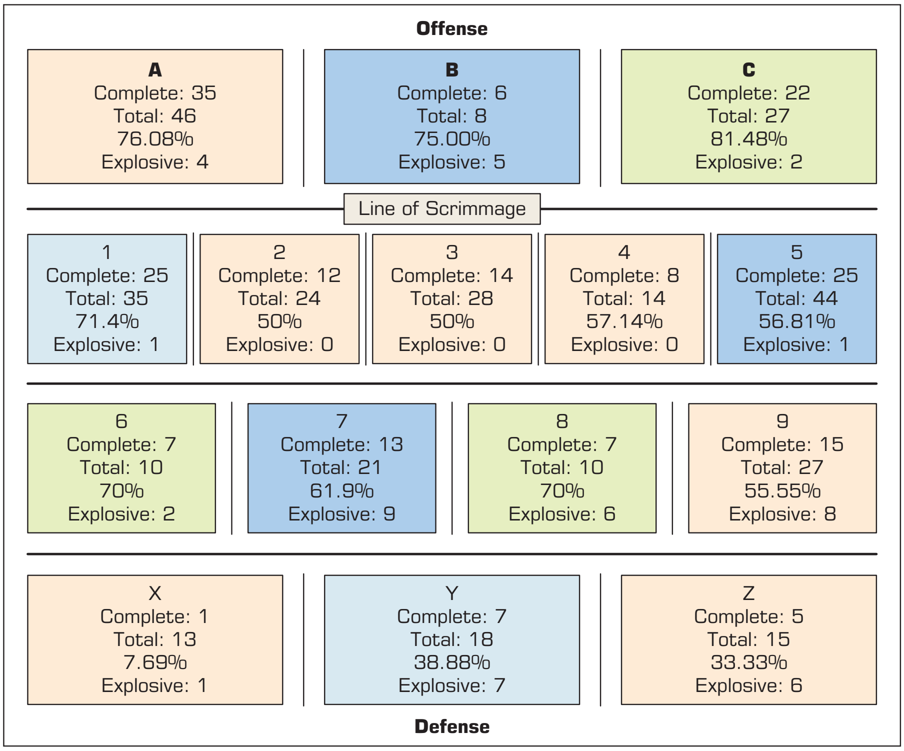

FIGURE 1.4 Heat Map Zone Analysis for Passing Plays.

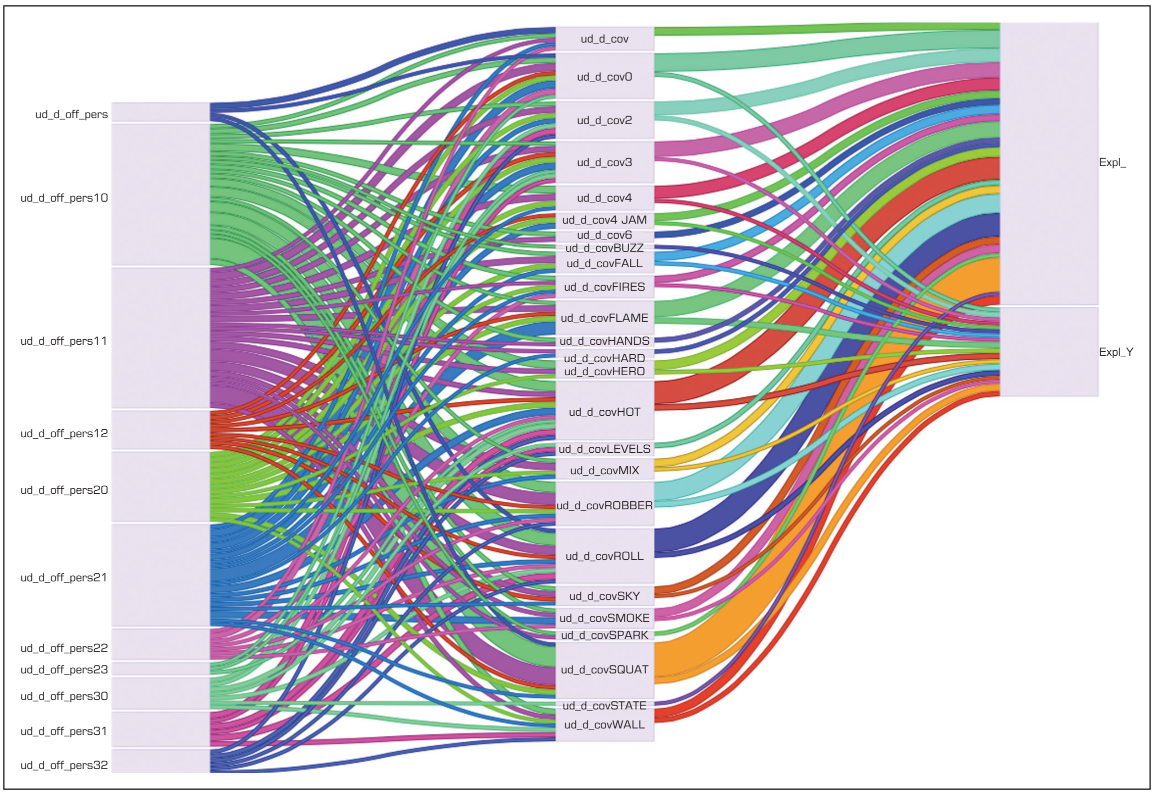

FIGURE 1.5 Time Series Analysis of Explosive Plays.

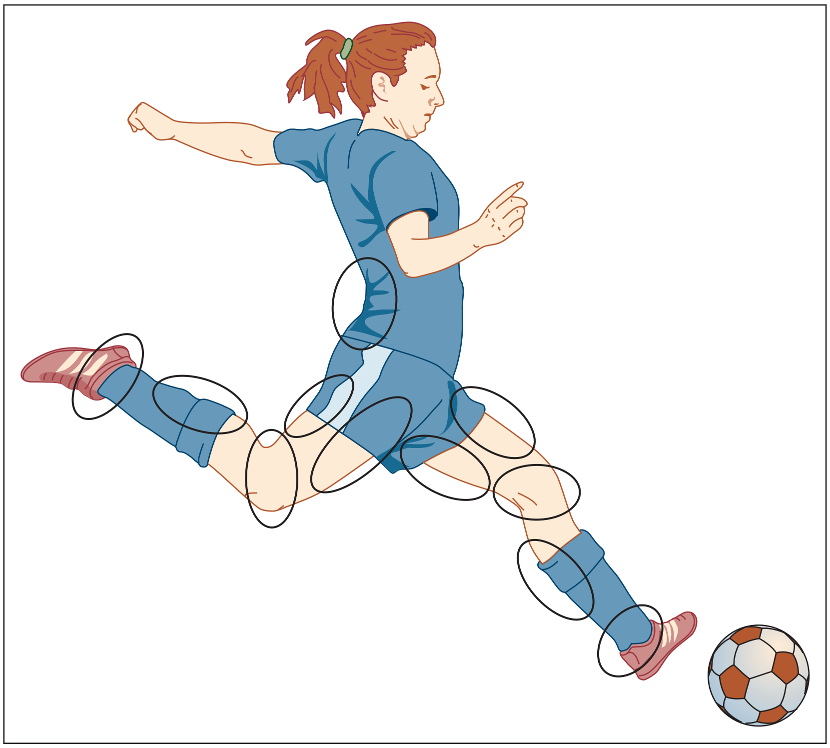

FIGURE 1.6 Soccer Injury Models.[5]

Example 3: The Trainer

Dr. Dan Johnson is the trainer for a women's college soccer team. His job is to help the players stay healthy and to advise the coaches on how much load to put on players during practices. He also has an interest in player well-being, including how much they sleep and how much rest they get between heavy and light practice sessions. The goal is to ensure that the players are ready to play on game days at maximum efficiency.

Fortunately, because of wearables, there is much more data for Dr. Dan to analyze. His players train using vests that contain sensors that can measure internal loads like heartbeats, body temperature, and respiration rates. The vests also include accelerometers that measure external loads like running distances and speeds as well as accelerations and decelerations. He knows which players are giving maximal effort during practices and those who aren't.

His focus at the moment is research that predicts or prevents player injuries (Figure 1.6). Some simple tasks like a Single Leg Squat Hold Test—standing on one foot, then the other—with score differentials of more than 10% can provide useful insights on body core strengths and weaknesses (Figure 1.7). If an athlete is hit hard during a match, a trainer can conduct a sideline test, reacting to a stimulus on a mobile device, which adds to traditional concussion protocols. Sleep sensors show who is getting adequate rest (or who partied all night). He has the MRI lab on campus do periodic brain scans to show which athletes are at risk for brain injury.

[5] "Women's Soccer Injuries," National Center for Catastrophic Sports Injury Research Report, NCAA. NCAA Sport Injury fact sheets are produced by the Datalys Center for Sports Injury Research and Prevention in collaboration with the National Collegiate Athletic Association, and STOP Sports Injuries. Appears at https://www.ncaa.org/sites/default/files/NCAA_W_Soccer_Injuries_WEB.pdf. (Accessed November 2016).

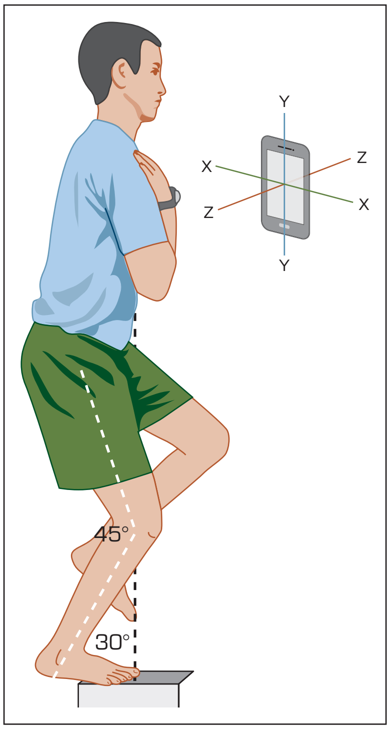

FIGURE 1.7 Single Leg Squat Hold Test–Core Body Strength Test
(*Source:* Figure adapted from Gary Wilkerson and Ashish Gupta).

QUESTIONS ABOUT THESE EXAMPLES

1. What are three factors that might be part of a PM for season ticket renewals?
2. What are two techniques that football teams can use to do opponent analysis?
3. How can wearables improve player health and safety? What kinds of new analytics can trainers use?
4. What other analytics uses can you envision in sports?

What Can We Learn from These Vignettes?

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's 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 Teradata University Network (TUN) a free resource for students and faculty. On this 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 of student projects in sports analytics and interviews of sports professionals who use data and analytics to do their jobs. Good luck learning analytics!

Source and Credits: 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. The football visuals (Figures 1.3–1.5) were constructed by Peter Liang and Jacob Pearson, graduate students at Oklahoma State University, as part of a student project in the spring of 2016. The training visuals (Figures 1.6 and 1.7) are adapted from the images provided by Prof. Gary Wilkerson of the University of Tennessee at Chattanooga and Prof. Ashish Gupta of Auburn University.

1.2 Changing Business Environments and Evolving Needs for Decision Support and Analytics

The opening vignette illustrates how an entire industry can employ analytics to develop reports on what is happening, predict what is likely to happen, and then also 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. From traditional uses in payroll and bookkeeping functions, computerized systems have now penetrated 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 information 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 the cloud-based systems for decision support are the cornerstones of today's modern management. Managers must have high-speed, networked information systems (wireline or wireless) to assist them with their most important task: making decisions. In many cases, such decisions are routinely being automated, eliminating the need for any managerial intervention.

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

- **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.
- **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.
- **Managing giant data warehouses and Big Data.** Large data warehouses (DWs), like the ones operated by Walmart, contain humongous amounts of data. Special

methods, including parallel computing, Hadoop/Spark, and so on, 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 into organizational performance that was not possible in the past.

- **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.
- **Overcoming cognitive limits in processing and storing information.** According to Simon (1977), 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.
- **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 the various stakeholders. Knowledge management systems 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.
- **Anywhere, anytime support.** Using wireless technology, managers can access information anytime and from anyplace, analyze and interpret it, and communicate with those involved. 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, but 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) support for managers. This growth in providing data-driven support for any decision extends to not just the 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.

SECTION 1.2 REVIEW QUESTIONS

1. What are some of the key system-oriented trends that have fostered IS-supported decision making to a new level?
2. List some capabilities of information systems that can facilitate managerial decision making.
3. How can a computer help overcome the cognitive limits of humans?

1.3 Evolution of Computerized Decision Support to Analytics/Data Science

The timeline in Figure 1.8 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 DSSs 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.

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 was 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 and in a bit more detail in Chapter 6.)

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. 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

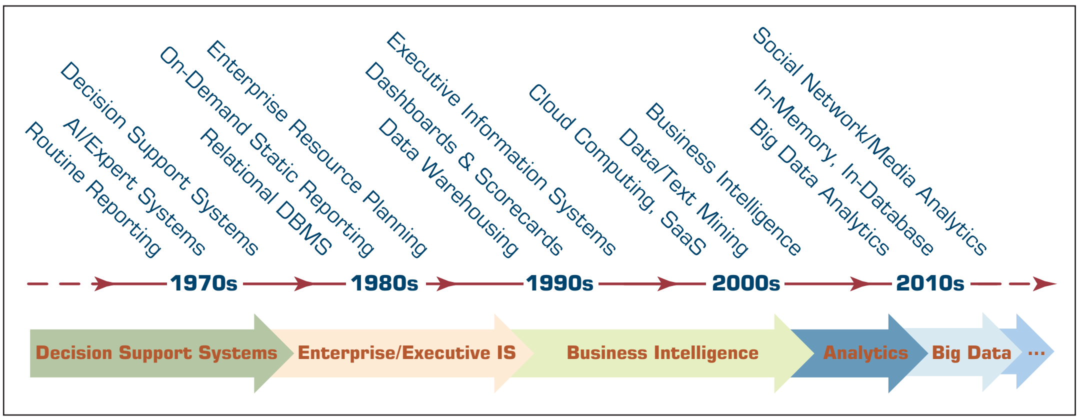

FIGURE 1.8 Evolution of Decision Support, Business Intelligence, and Analytics.

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; DSSs 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 DSSs 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 is updated periodically, it does 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 is 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. This unstructured data is 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) have been developed to address the challenges of Big Data.

It's 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 this 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 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 Section 1.4 pays homage to that history by specifically focusing on what has been called BI. Following the next section, we introduce analytics and will use that as the label for classifying all related concepts.

SECTION 1.3 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 Support Systems?
3. Did DSS evolve into BI or vice versa?

1.4 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 1.3, 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 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 1.9 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 1.9 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 1.10.

The Origins and Drivers of BI

Where did modern approaches to data warehousing 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.

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.

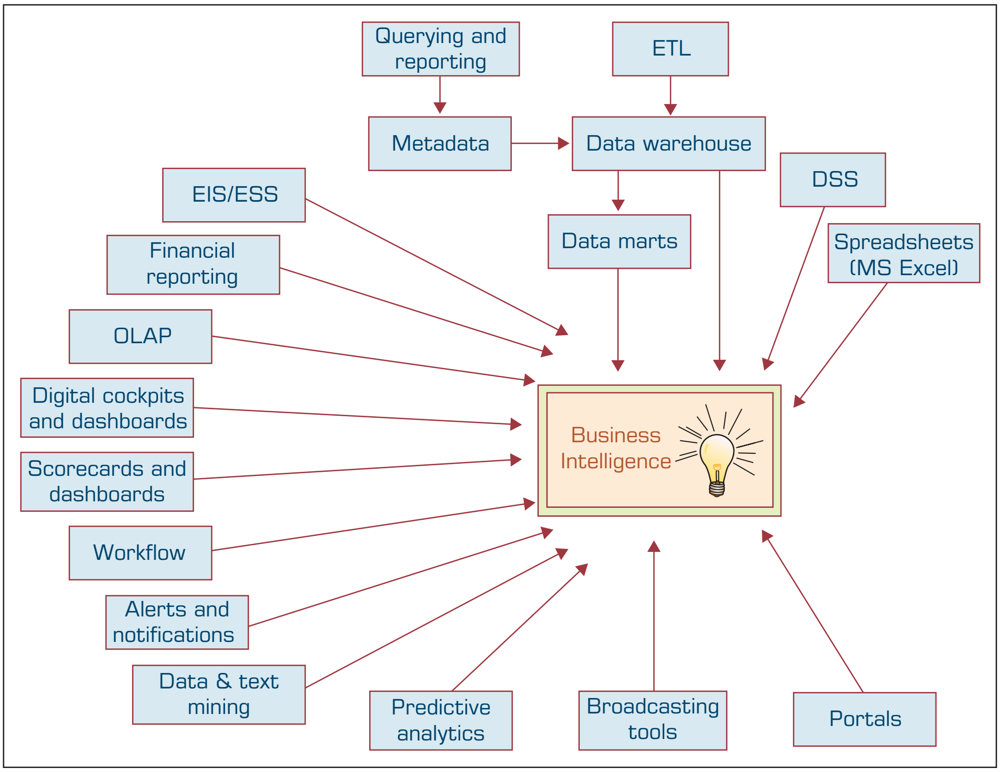

FIGURE 1.9 Evolution of Business Intelligence (BI).

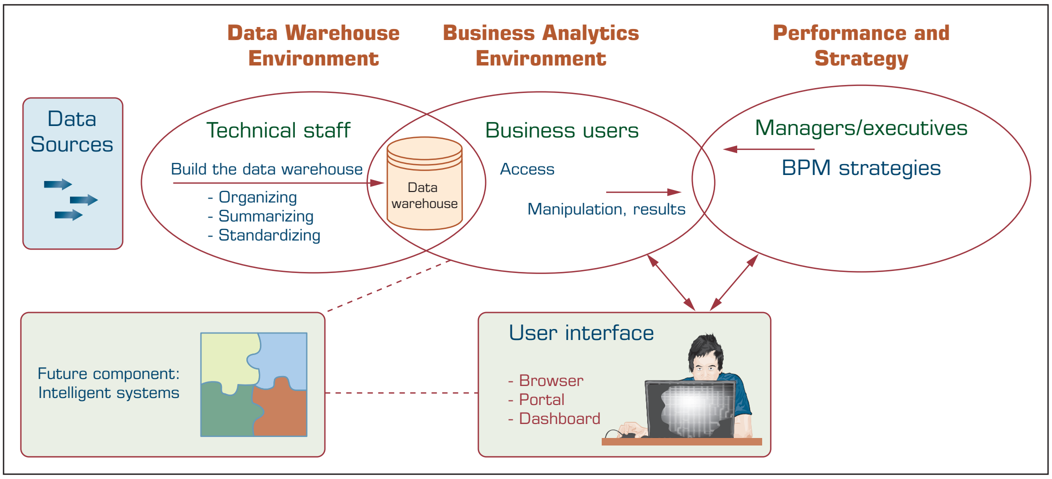

FIGURE 1.10 A High-Level Architecture of BI. (*Source:* Based on W. Eckerson, *Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions.* The Data Warehousing Institute, Seattle, WA, 2003, p. 32, Illustration 5.)

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. Application Case 1.1 illustrates one such application of BI that has helped many airlines as well as, of course, the companies offering such services to the airlines.

Application Case 1.1

Sabre Helps Its Clients Through Dashboards and Analytics

Sabre is one of the world leaders in the travel industry, providing both business-to-consumer services as well as business-to-business services. It serves travelers, travel agents, corporations, and travel suppliers through its four main companies: Travelocity, Sabre Travel Network, Sabre Airline Solutions, and Sabre Hospitality Solutions. The current volatile global economic environment poses significant competitive challenges to the airline industry. To stay ahead of the competition, Sabre Airline Solutions recognized that airline executives needed enhanced tools for managing their business decisions by eliminating the traditional, manual, time-consuming process of aggregating financial and other information needed for actionable initiatives. This enables real-time decision support at airlines throughout the world to maximize their (and in turn Sabre's) return on information by driving insights, actionable intelligence, and value for customers from the growing data.

Sabre developed an Enterprise Travel Data Warehouse (ETDW) using Teradata to hold its massive reservations data. ETDW is updated in near-real time with batches that run every 15 minutes, gathering data from all of Sabre's businesses. Sabre uses its ETDW to create Sabre Executive Dashboards that provide near real-time executive insights using a Cognos BI platform with Oracle Data Integrator and Oracle Goldengate technology infrastructures. The Executive Dashboards offer their client airlines' top-level managers and decision makers a timely, automated, user-friendly solution, aggregating critical performance metrics in a succinct way and providing at a glance a 360-degree view of the overall health of the airline. At one airline, Sabre's Executive Dashboards provide senior management with a daily and intraday snapshot of key performance indicators in a single application replacing the once-a-week, 8-hour process of generating the same report from various data sources. The use of dashboards is not limited to the external customers; Sabre also uses them for their assessment of internal operational performance.

The dashboards help Sabre's customers to have a clear understanding of the data through the visual displays that incorporate interactive drill-down capabilities. It replaces flat presentations and allows for a more focused review of the data with less effort and time. This facilitates team dialog by making the data/metrics pertaining to sales performance available to many stakeholders, including ticketing, seats sold and flown, operational performance including the data on flight movement and tracking, customer reservations, inventory, and revenue across an airline's multiple distribution channels. The dashboard systems provide scalable infrastructure, graphical user interface support, data integration, and aggregation that empower airline executives to be more proactive in taking actions that lead to positive impacts on the overall health of their airline.

With its ETDW, Sabre could also develop other Web-based analytical and reporting solutions that leverage data to gain customer insights through analysis of customer profiles and their sales interactions to calculate customer value. This enables better customer segmentation and insights for value-added services.

Questions for Discussion

1. What is traditional reporting? How is it used in the organization?
2. How can analytics be used to transform the traditional reporting?
3. How can interactive reporting assist organizations in decision making?

What We Can Learn from This Application Case

This case shows that organizations that earlier used reporting only for tracking their internal business activities and meeting the compliance requirements set out by the government are now moving toward generating actionable intelligence from their transactional business data. Reporting has become broader as organizations are now trying to analyze the archived transactional data to understand the underlying hidden trends and patterns that will enable them to make better decisions by gaining insights into problematic areas and resolving them to pursue current and future market opportunities. Reporting has advanced to interactive online reports, which enable the users to pull and build quick custom reports and even present the reports aided by visualization tools that have the ability to connect to the database, providing the capabilities of digging deep into summarized data.

Source: Teradata.com, "Sabre Airline Solutions," Terry, D. (2011), "Sabre Streamlines Decision Making," http://www.teradatamagazine.com/v11n04/Features/Sabre-Streamlines-Decision-Making/ (Accessed July 2016).

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 http://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*.

Transaction Processing versus Analytic Processing

To illustrate the major characteristics of BI, first we will show what BI is not—namely, transaction processing. We're all familiar with the information systems that support our transactions, like ATM withdrawals, bank deposits, cash register scans at the grocery store, and so on. 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 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. We will provide a more technical definition of DW in Chapter 2, but suffice it to say that DWs are intended to work with informational data used for **online analytical processing (OLAP)** systems.

Most operational data in enterprise resources planning (ERP) systems—and in its complementary siblings like *supply chain management* (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, it has still been a far cry from a desired usability by regular, nontechnical, end users for things such as operational reporting, interactive analysis, and so on. 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.

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, developed by Gartner, Inc. (2004), decomposes 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 is 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.

If the company's strategy is properly aligned with the reasons for DW and BI initiatives, and if the company's IS organization is or can be made capable of playing its role in such a project, and if the requisite user community is in place and has the proper motivation, it is wise to start BI and establish a BI Competency Center within the company. The center could serve some or all of the following functions (Gartner, 2004):

- The 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.
- The 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.
- It can help important stakeholders like high-level executives see how BI can play an important role.

Another important success factor of BI is its ability to facilitate a real-time, on-demand agile environment, introduced next.

Real-Time, On-Demand BI Is Attainable

The demand for instant, on-demand access to dispersed information has grown as the need to close the gap between the operational data and strategic objectives has become more pressing. As a result, a category of products called *real-time BI applications* has emerged. The introduction of new data-generating technologies, such as RFID and other sensors is only accelerating this growth and the subsequent need for real-time BI. Traditional BI systems use a large volume of *static* data that has been extracted, cleansed, and loaded into a *DW* to produce reports and analyses. However, the need is not just reporting because users need business monitoring, performance analysis, and an understanding of why things are happening. These can assist users, who need to know (virtually in real time) about changes in data or the availability of relevant reports, alerts, and notifications regarding events and emerging trends in social media applications. In addition, business applications can be programmed to act on what these real-time BI systems discover. For example, an SCM application might automatically place an order for more "widgets" when real-time inventory falls below a certain threshold or when a CRM application automatically triggers a customer service representative and credit control clerk to check a customer who has placed an online order larger than $10,000.

One approach to real-time BI uses the DW model of traditional BI systems. In this case, products from innovative BI platform providers provide a service-oriented, near-real-time solution that populates the DW much faster than the typical nightly *extract/transfer/load* batch update does (see Chapter 3). A second approach, commonly called *business activity management* (BAM), is adopted by pure-play BAM and/or hybrid BAM-middleware providers (such as Savvion, Iteration Software, Vitria, webMethods, Quantive, Tibco, or Vineyard Software). It bypasses the DW entirely and uses **Web services** or other monitoring means to discover key business events. These software monitors (or **intelligent agents**) can be placed on a separate server in the network or on the transactional application databases themselves, and they can use event- and process-based approaches to proactively and intelligently measure and monitor operational processes.

Developing or Acquiring BI Systems

Today, many vendors offer diversified tools, some of which are completely preprogrammed (called *shells*); all you have to do is insert your numbers. These tools can be purchased or leased. For a list of products, demos, white papers, and more current product information, see product directories at tdwi.org. Free user registration is required. Almost all BI applications are constructed with shells provided by vendors who may themselves create a custom solution for a client or work with another outsourcing provider. The issue that companies face is which alternative to select: purchase, lease, or build. Each of these alternatives has several options. One of the major criteria for making the decision is justification and cost–benefit analysis.

Justification and Cost–Benefit Analysis

As the number of potential BI applications increases, the need to justify and prioritize them arises. This is not an easy task due to the large number of intangible benefits. Both direct and intangible benefits need to be identified. Of course, this is where the knowledge of similar applications in other organizations and case studies is extremely useful. For example, The Data Warehousing Institute (tdwi.org) provides a wealth of information about products and innovative applications and implementations. Such information can be useful in estimating direct and indirect benefits.

Security and Protection of Privacy

This is an extremely important issue in the development of any computerized system, especially BI that contains data that may possess strategic value. Also, the privacy of employees and customers needs to be protected.

Integration of Systems and Applications

With the exception of some small applications, all BI applications must be integrated with other systems such as databases, legacy systems, enterprise systems (particularly ERP and CRM), e-commerce (sell side, buy side), and many more. In addition, BI applications are usually connected to the Internet and many times to information systems of business partners.

Furthermore, BI tools sometimes need 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.

SECTION 1.4 REVIEW QUESTIONS

1. Define *BI*.
2. List and describe the major components of BI.
3. Define *OLTP*.
4. Define *OLAP*.
5. List some of the implementation topics addressed by Gartner's report.
6. 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 on 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. These eight levels of analytics are described in more detail in a white paper by SAS (sas.com/news/sascom/analytics_levels.pdf).

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 (informs.org/Community/Analytics) as descriptive, predictive, and prescriptive. Figure 1.11 presents a graphical view of these three levels of analytics. It suggests that these three are somewhat independent steps and one type 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.

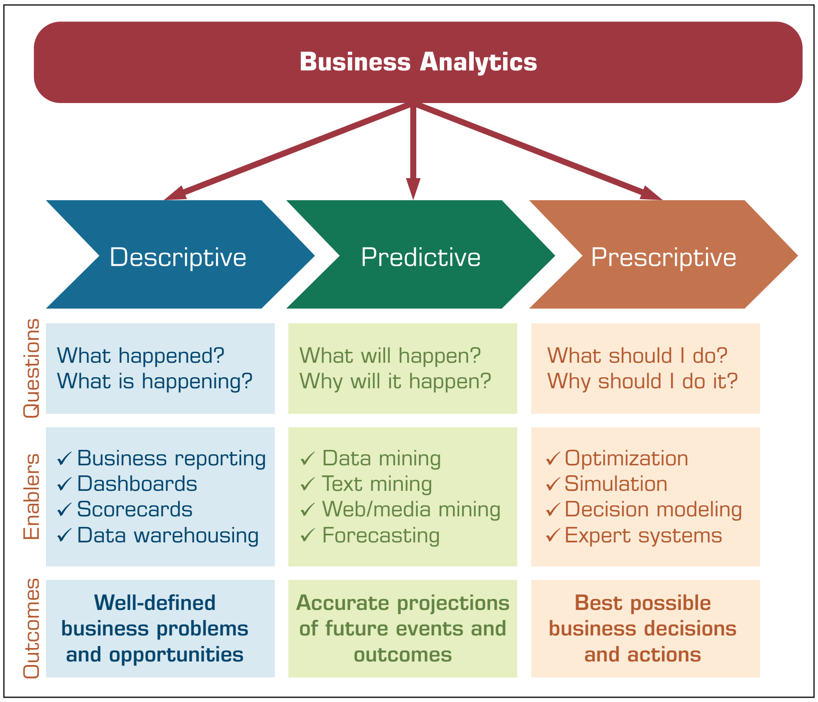

FIGURE 1.11 Three Types of Analytics.

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. Application Cases 1.2 and 1.3 highlight some such applications. Color renderings of visualizations discussed in these applications are available online or the book's companion Web site (dssbibook.com).

Application Case 1.2

Silvaris Increases Business with Visual Analysis and Real-Time Reporting Capabilities

Silvaris Corporation was founded in 2000 by a team of forest industry professionals to provide technological advancement in the lumber and building material sector. Silvaris is the first e-commerce platform in the United States specifically for forest products and is headquartered in Seattle, Washington. It is a leading wholesale provider of industrial wood products and surplus building materials.

Silvaris sells its products and provides international logistics services to more than 3,500 customers. To manage various processes that are involved in a transaction, they created a proprietary online trading platform to track information flow related to transactions between traders, accounting, credit, and logistics. This allowed Silvaris to share its real-time information with its customers and partners. But due to the rapidly changing prices of materials, it became necessary for Silvaris to get a real-time view of data without moving data into a separate reporting format.

Silvaris started using Tableau because of its ability to connect with and visualize live data. Due to dashboards created by Tableau that are easy to understand and explain, Silvaris started using Tableau for reporting purposes. This helped Silvaris in pulling out information quickly from the data and identifying issues that impact their business. Silvaris succeeded in managing online versus offline orders with the help of reports generated by Tableau. Now, Silvaris keeps track of online orders placed by customers and knows when to send renew pushes to which customers to keep them purchasing online. Also, analysts of Silvaris can save time by generating dashboards instead of writing hundreds of pages of reports by using Tableau.

Questions for Discussion

1. What was the challenge faced by Silvaris?
2. How did Silvaris solve its problem using data visualization with Tableau?

What We Can Learn from This Application Case

Many industries need to analyze data in real time. Real-time analysis enables the analysts to identify issues that impact their business. Visualization is sometimes the best way to begin analyzing the live data streams. Tableau is one such data visualization tool that has the capability to analyze live data without bringing live data into a separate reporting format.

Sources: Tableau.com,"Silvaris Augments Proprietary Technology Platform with Tableau's Real-Time Reporting Capabilities," http://www.tableau.com/sites/default/files/case-studies/silvaris-business-dashboards_0.pdf (accessed July 2016); Silvaris.com, "Overview," http://www.silvaris.com/About/ (accessed July 2016).

Application Case 1.3

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 their multiple cases of customer satisfaction surveys, logistic processes, and financial reporting. This platform should be easy to use for their employees so that they can use this 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.

They started using Dundas BI, a leading global provider of BI and data visualization solutions. It allowed Siemens to create highly interactive dashboards that enabled Siemens 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 them to reduce cycle time by 12% and scrap cost by 25%.

Questions for Discussion

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," http://www.dundas.com/resource/getcasestudy?caseStudyName=09-03-2016-Siemens%2FDundas-BI-Siemens-Case-Study.pdf (accessed July 2016); Wikipedia.org, "SIEMENS," https://en.wikipedia.org/wiki/Siemens (accessed July 2016); Siemens.com, "About Siemens," http://www.siemens.com/about/en/ (accessed July 2016).

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 if the customer is likely to switch to a competitor ("churn"), what the customer would likely buy next and how much, what promotions a customer would respond to, whether this 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 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 other similar products that a customer may be interested in. We will study these techniques and their applications in Chapters 3 through 6. Application Case 1.4 illustrates one such application in sports.

Application Case 1.4

Analyzing Athletic Injuries

Any athletic activity is prone to injuries. If the injuries are not handled properly, then the team suffers. Using analytics to understand injuries can help in deriving valuable insights that would enable coaches and team doctors to manage the team composition, understand player profiles, and ultimately aid in better decision making concerning which players might be available to play at any given time.

In an exploratory study, Oklahoma State University analyzed American football-related sports injuries by using reporting and predictive analytics. The project followed the CRISP-DM methodology (to be described in Chapter 4) to understand the problem of making recommendations on managing injuries, understanding the various data elements collected about injuries, cleaning the data, developing visualizations to draw various inferences, building PMs to analyze the injury healing time period, and drawing sequence rules to predict the relationships among the injuries and the various body part parts afflicted with injuries.

The injury data set consisted of more than 560 football injury records, which were categorized into injury-specific variables—body part/site/laterality, action taken, severity, injury type, injury start and healing dates—and player/sport-specific variables—player ID, position played, activity, onset, and game location. Healing time was calculated for each record, which was classified into different sets of time periods: 0–1 month, 1–2 months, 2–4 months, 4–6 months, and 6–24 months.

Various visualizations were built to draw inferences from injury data set information depicting the healing time period associated with players' positions, severity of injuries and the healing time period, treatment offered and the associated healing time period, major injuries afflicting body parts, and so forth.

Neural network models were built to predict each of the healing categories using IBM SPSS Modeler. Some of the predictor variables were current status of injury, severity, body part, body site, type of injury, activity, event location, action taken, and position played. The success of classifying the healing category was quite good: Accuracy was 79.6% percent. Based on the analysis, many business recommendations were suggested, including employing more specialists' input from injury onset instead of letting the training room staff screen the injured players; training players at defensive positions to avoid being injured; and holding practice to thoroughly safety-check mechanisms.

Questions for Discussion

1. What types of analytics are applied in the injury analysis?
2. How do visualizations aid in understanding the data and delivering insights into the data?
3. What is a classification problem?
4. What can be derived by performing sequence analysis?

What We Can Learn from This Application Case

For any analytics project, it is always important to understand the business domain and the current state of the business problem through extensive analysis of the only resource—historical data. Visualizations often provide a great tool for gaining the initial insights into data, which can be further refined based on expert opinions to identify the relative importance of the data elements related to the problem. Visualizations also aid in generating ideas for obscure problems, which can be pursued in building PMs that could help organizations in decision making.

Source: Sharda, R., Asamoah, D., & Ponna, N. (2013). "Research and Pedagogy in Business Analytics: Opportunities and Illustrative Examples." *Journal of Computing and Information Technology, 21*(3), 171–182.

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 are 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 1.5 gives an example of such prescriptive analytic applications. We will learn about some aspects of prescriptive analytics in Chapter 6.

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 5)—or Web analytics—analyzing Web data

Application Case 1.5

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

This application case is based on a project that includes one of us. A company that does not wish to disclose its name (or even the 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 and, of course, prices, and so on. It must make available-to-promise (ATP) decisions, which determine in real time the dates when it 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 may not really be the free 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 Discussion

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: Pajouh Foad, M., Xing, D., Hariharan, S., Zhou, Y., Balasundaram, B., Liu, T., & Sharda, R. (2013). "Available-to-Promise in Practice: An Application of Analytics in the Specialty Steel Bar Products Industry." *Interfaces*, *43*(6), 503–517. http://dx.doi.org/10.1287/inte.2013.0693 (accessed July 2016).

streams (also in Chapter 5)—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, Section 1.1 introduced the phrase *sports analytics*. Application Case 1.1 could also be termed a case study in airline 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 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, some SQL knowledge, and reporting. You would recognize those capabilities as descriptive or reporting analytics. In contrast, a data scientist is responsible for predictive analysis, statistical analysis, and 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, decision analysis, and so on. Those who cover these fields are more likely to be called data scientists than analytics professionals.

Our view is that the distinction between analytics 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 is fuzzy between the two labels, at best. We observe that graduates of our analytics programs tend to be responsible for tasks which 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.

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 data warehousing technology help to 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?

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 two application areas—healthcare and retail, where there have been the most reported applications and successes.

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, Humana. According to the company 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, a major health insurance provider in the United States. 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.

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] 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%.[2] The costs of falls pose a significant strain on

[1]http://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html.

[2]Gill, T. M., Murphy, T. E., Gahbauer, E. A., et al. (2013). Association of injurious falls with disability outcomes and nursing home admissions in community living older persons. *American Journal of Epidemiology, 178*(3), 418–425.

the U.S. healthcare system, with the direct costs of falls 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."[1]

Humana is the nation's second-largest provider of Medicare Advantage benefits with approximately 3.2 million members, most of whom are seniors. Keeping their 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.[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% of 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 falls, 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.

---

[3]Gates, S., Smith, L. A., Fisher, J. D., et al. (2008). Systematic review of accuracy of screening instruments for predicting fall risk among independently living older adults. *Journal of Rehabilitation Research and Development, 45*(8), 1105–1116.

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

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% 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, CHF [congestive heart failure], CAD [coronary artery disease], COPD [chronic obstructive pulmonary disease], 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 ER 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, but was also 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 their overall progress toward the Bold Goal.

It was critical to make sure that the selected metric was highly correlated to health and business metrics, such 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 5-day increase in UHDs, there is (a) an $82 increase in average monthly medical and pharmacy costs, (b) an increase of 52 inpatient admits per 1000 patients, and (c) a 0.28-day increase in average length of stay.[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).[2]
- Individuals with chronic conditions have more UHDs than those who do not have: (a) CHF (16.9 UHDs), (b) CAD (14.4 UHDs), (c) hypertension (13.3 UHDs), (d) diabetes (14.7 UHDs), (e) COPD (17.4 UHDs), or (f) depression (22.4 UHDs).[1,3,4]

Humana has since adopted Healthy Days as their metric for the measurement of progress toward Bold Goal.[5]

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

---

[1]Havens, E., Peña, J., Slabaugh, S., Cordier, T., Renda, A., & Gopal, V. (2015, October). Exploring the relationship between health-related quality of life and health conditions, costs, resource utilization, and quality measures. Podium presentation at the ISOQOL 22nd Annual Conference, Vancouver, Canada.

[2]Havens, E., Slabaugh, L., Peña J., Haugh G., & Gopal, V. (2015, February). Are there differences in Healthy Days based on compliance to preventive health screening measures? Poster presentation at Preventive Medicine 2015, Atlanta, GA.

[3]Chiguluri, V., Guthikonda, K., Slabaugh, S., Havens, E., Peña, J., & Cordier, T. (2015, June). Relationship between diabetes complications and health related quality of life among an elderly population in the United States. Poster presentation at the American Diabetes Association 75th Annual Scientific Sessions. Boston, MA.

[4]Cordier, T., Slabaugh, L., Haugh, G., Gopal, V., Cusano, D., Andrews, G., & Renda, A. (2015, September). Quality of life changes with progressing congestive heart failure. Poster presentation at the 19th Annual Scientific Meeting of the Heart Failure Society of America, Washington, DC.

[5] http://populationhealth.humana.com/wp-content/uploads/2016/05/BoldGoal2016ProgressReport_1.pdf.

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% of consumers account for 80% 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 right members is critical for this exercise, and in the recent years, PMs have been developed to identify enrollees with the 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 MA 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.

Ever 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% 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% of the referrals identified within the first 3 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 in succeeding. 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 just highlights 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 their key focus areas.

Figure 1.12 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.

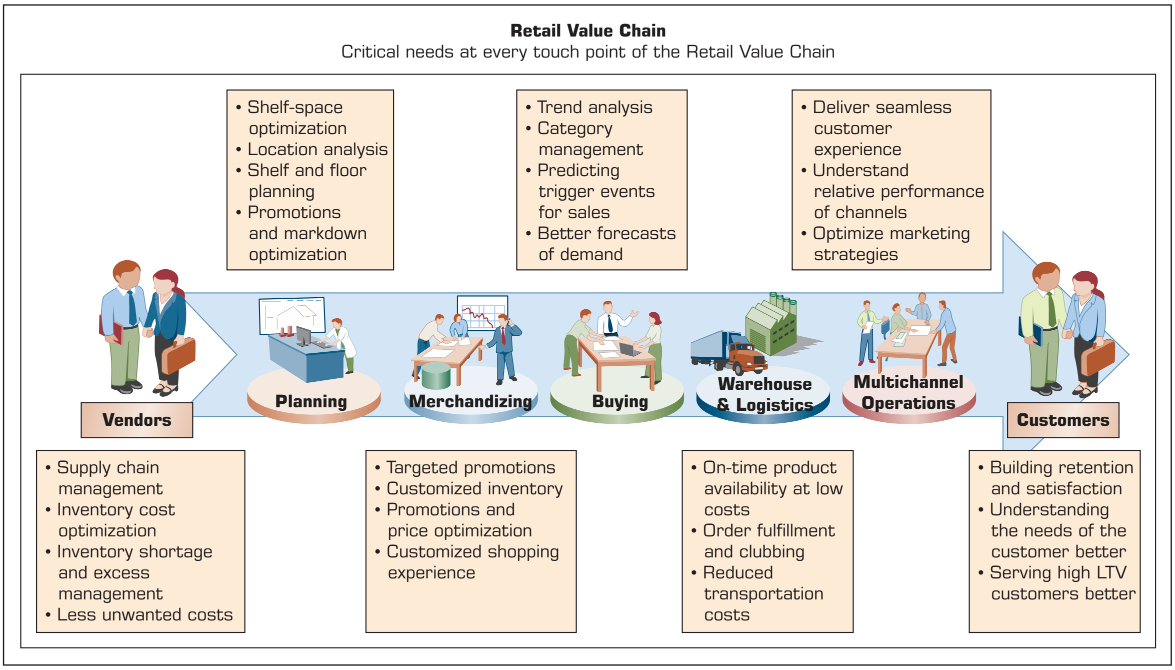

FIGURE 1.12 Example of Analytics Applications in a Retail Value Chain. Contributed by Abhishek Rathi, CEO, vCreaTek.com

TABLE 1.1 Examples of Analytics Applications in the Retail Value Chain

Analytic Application	Business Question	Business Value
Inventory Optimization	1. Which products have high demand? 2. Which products are slow moving or becoming obsolete?	1. Forecast the consumption of fast-moving products and order them with sufficient inventory to avoid a stock-out scenario. 2. Perform fast inventory turnover of slow-moving products by combining them with one in high demand.
Price Elasticity	1. How much net margin do I have on the product? 2. How much discount can I give on this product?	1. Markdown prices for each product can be optimized to reduce the margin dollar loss. 2. Optimized price for the bundle of products is identified to save the margin dollar.
Market Basket Analysis	1. What products should I combine to create a bundle offer? 2. Should I combine products based on slow-moving and fast-moving characteristics? 3. Should I create a bundle from the same category or different category line?	1. The affinity analysis identifies the hidden correlations between the products, which can help in following values: a) Strategize the product bundle offering based on focus on inventory or margin. b) Increase cross-sell or up-sell by creating bundle from different categories or the same categories, respectively.
Shopper Insight	1. Which customer is buying what product at what location?	1. By customer segmentation, the business owner can create personalized offers resulting in better customer experience and retention of the customer.
Customer Churn Analysis	1. Who are the customers who will not return? 2. How much business will I lose? 3. How can I retain them? 4. What demography of customer is my loyal customer?	1. Businesses can identify the customer and product relationships that are not working and show high churn. Thus can have better focus on product quality and reason for that churn. 2. Based on the customer lifetime value (LTV), the business can do targeted marketing resulting in retention of the customer.
Channel Analysis	1. Which channel has lower customer acquisition cost? 2. Which channel has better customer retention? 3. Which channel is more profitable?	1. Marketing budget can be optimized based on insight for better return on investment.
New Store Analysis	1. What location should I open? 2. What and how much opening inventory should I keep?	1. Best practices of other locations and channels can be used to get a jump start. 2. Comparison with competitor data can help to create a differentiator/USP factor to attract the new customers.
Store Layout	1. How should I do store layout for better topline? 2. How can I increase my in-store customer experience?	1. Understand the association of products to decide store layout and better alignment with customer needs. 2. Workforce deployment can be planned for better customer interactivity and thus satisfying customer experience.
Video Analytics	1. What demography is entering the store during the peak period of sales? 2. How can I identify a customer with high LTV at the store entrance so that a better personalized experience can be provided to this customer?	1. In-store promotions and events can be planned based on the demography of incoming traffic. 2. Targeted customer engagement and instant discount enhances the customer experience resulting in higher retention.

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 their loyalty program, the store may be able to identify the customer at the entrance itself and thus enable an extra opportunity for a cross-sell or up-sell. 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 can use this information to also plan the workforce allocation for those areas for peak periods.

Market basket analysis has commonly been used by the category managers to push the sale of the slowly moving SKUs. By using advanced analytics of data available, the product affinity can be done at the lowest level of SKU to drive better ROIs on the bundle offers. Moreover, by using price elasticity techniques, the markdown or optimum price of the bundle offer can also be deduced, thus reducing any loss in the profit margin.

Thus by using data analytics, a retailer can not only get information on its current operations but can also get further insight to increase the revenue and decrease the operational cost for higher profit. A fairly comprehensive list of current and potential retail analytics applications that a major retailer such as Amazon could use is proposed by a blogger at Data Science Central. That list is available at http://www.datasciencecentral.com/profiles/blogs/20-data-science-systems-used-by-amazon-to-operate-its-business. As noted earlier, there are too many examples of these opportunities to list here, but you will see many examples of such applications throughout the book.

SECTION 1.6 REVIEW QUESTIONS

1. Why would a health insurance company invest in analytics beyond fraud detection? Why is it in their best interest to predict the likelihood of falls by patients?
2. What other applications similar to prediction of falls can you envision?
3. How would you convince a new health insurance customer to adopt healthier lifestyles (Humana Example 3)?
4. Identify at least three other opportunities for applying analytics in the retail value chain beyond those covered in this section.
5. Which retail stores that you know of employ some of the analytics applications identified in this section?

1.7 A Brief Introduction to Big Data Analytics

What Is Big Data?

Any book on analytics and data science has to include significant coverage of what is called **Big Data analytics**. We will cover it in Chapter 7 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 is generated, let alone analyze it fast. This is why we have the problem of Big Data. So, what is Big Data? Simply put, Big Data is data that cannot be stored in a single storage unit. Big Data typically refers to data that comes 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 like 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 this 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 as an Apache project as the Hadoop Distributed File System (HDFS).

However, storing this data is only half the problem. Data is worthless if it does not provide business value, and for it to provide business value, it has to 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 has the characteristics of volume, velocity, and variety. Application Case 1.6 illustrates an application of Big Data analytics in the energy industry. We will study Big Data technologies and applications in Chapter 7.

SECTION 1.7 REVIEW QUESTIONS

1. What is Big Data analytics?
2. What are the sources of Big Data?
3. What are the characteristics of Big Data?
4. What processing technique is applied to process Big Data?

Application Case 1.6

CenterPoint Energy Uses Real-Time Big Data Analytics to Improve Customer Service

CenterPoint Energy is a *Fortune* 500 energy delivery company based in Houston, Texas. Its primary business includes electric transmission and distribution, natural gas distribution, and natural gas sales and service. It has over five million metered customers in the United States.

CenterPoint Energy uses smart grids to collect real-time information about the health of various aspects of the grid like meters, transformers, and switches that are used in providing electricity. This real-time power usage information is analyzed with Big Data analytics and allows for a much quicker diagnosis and solution. For example, the data can predict and potentially help prevent a power outage.

In addition, the tool collects weather information allowing historical data to help predict the magnitude of an outage from a storm. This insight will act as a guide for putting the right resources out before a storm occurs to avoid an outage.

Second, to better understand their customers, CenterPoint Energy utilizes sentiment analysis, which examines a customer's opinion by way of emotion (happiness, anger, sadness, etc.). The company segments their customers based on the sentiment and is able to market to these groups in a more personalized way, providing a more valuable customer service experience.

As a result of using Big Data analytics, CenterPoint Energy has saved 600,000 gallons of fuel in the last 2 years by resolving six million service requests remotely. In addition, they have saved $24 million for their customers in this process.

Questions for Discussion

1. How can electric companies predict a possible outage at a location?
2. What is customer sentiment analysis?
3. How does customer sentiment analysis help companies provide a personalized service to their customers?

What We Can Learn from This Application Case

With the use of Big Data analytics, energy companies can better solve customer issues like outages and electric faults within a shorter span of time compared to the earlier process. Also sentiment analysis can help target their customers according to their needs.

Sources: Sap.com, "A 'Smart' Approach to Big Data in the Energy Industry," http://www.sap.com/bin/sapcom/cs_cz/downloadasset.2013-10-oct-09-20.a-smart-approach-to-big-data-in-the-energy-industry-pdf.html (accessed June 2016); centerpointenergy.com, "Electric Transmission & Distribution (T&D)," http://www.centerpointenergy.com/en-us/Corp/Pages/Company-overview.aspx (accessed June 2016); YouTube.com, "CenterPoint Energy Talks Real Time Big Data Analytics," https://www.youtube.com/watch?v=s7CzeSlIEfI (accessed June 2016).

1.8 An Overview of the Analytics Ecosystem

So you are excited about the potential of analytics and want to join this growing industry. Who are the current players, and what to do they do? Where might you fit in? The objective of this section is to identify various sectors of the analytics industry, provide a classification of different types of industry participants, and illustrate the types of opportunities that exist for analytics professionals. Eleven different types of players are identified in an **analytics ecosystem**. An understanding of the ecosystem also gives the reader a broader view of how the various players come together. A secondary purpose of understanding the analytics ecosystem for the BI professional is also to be aware of organizations and new offerings and opportunities in sectors allied with analytics. The section concludes with some observations about the opportunities for professionals to move across these clusters.

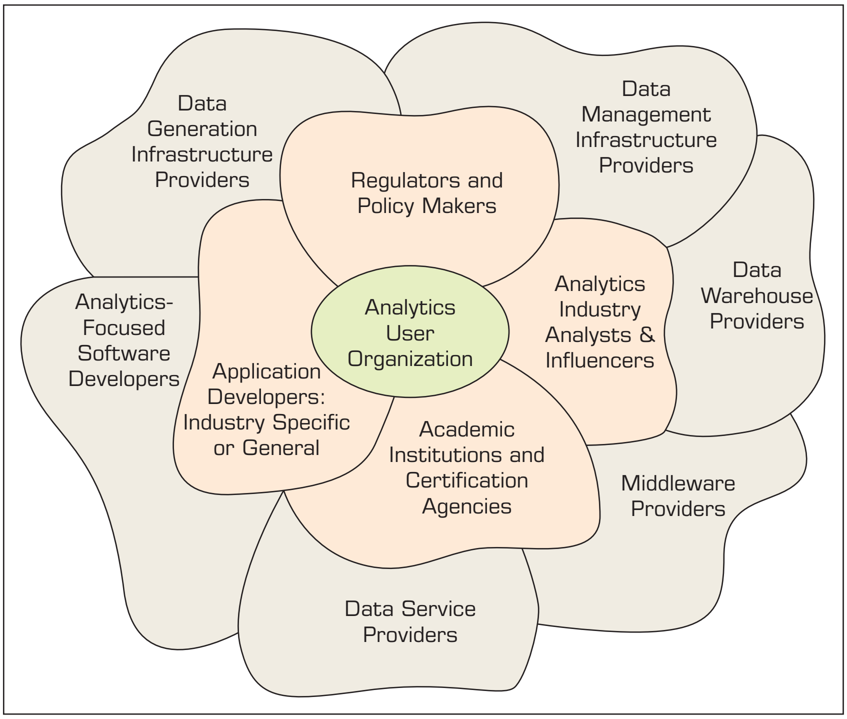

FIGURE 1.13 Analytics Ecosystem.

Although some researchers have distinguished business analytics professionals from data scientists (Davenport and Patil, 2012), as pointed out previously, for the purpose of understanding the overall analytics ecosystem, we treat them as one broad profession. Clearly, skill needs can vary between a strong mathematician to a programmer to a modeler to a communicator, and we believe this issue is resolved at a more micro/individual level rather than at a macro level of understanding the opportunity pool. We also take the widest definition of analytics to include all three types as defined by INFORMS—descriptive/reporting/visualization, predictive, and prescriptive as described earlier.

Figure 1.13 illustrates one view of the analytics ecosystem. The components of the ecosystem are represented by the petals of an analytics flower. Eleven key sectors or clusters in the analytics space are identified. The components of the analytics ecosystem are grouped into three categories represented by the inner petals, outer petals, and the seed (middle part) of the flower.

The outer six petals can be broadly termed as the technology providers. Their primary revenue comes from providing technology, solutions, and training to analytics user organizations so they can employ these technologies in the most effective and efficient manner. The inner petals can be generally defined as the analytics accelerators. The accelerators work with both technology providers and users. Finally, the core of the ecosystem comprises the analytics user organizations. This is the most important component, as every analytics industry cluster is driven by the user organizations.

The metaphor of a flower is well-suited for the analytics ecosystem as multiple components overlap each other. Similar to a living organism like a flower, all these petals grow and wither together. We use the terms *components*, *clusters*, *petals*, and *sectors* interchangeably to describe the various players in the analytics space. We introduce each of the industry sectors next and give some examples of players in each sector. The list of company names included in any petal is not exhaustive. The representative list of companies in each cluster is just to illustrate that cluster's unique offering to describe where analytics talent may be used or hired away. Also, mention of a company's name or its capability in one specific group does not imply that it is the only activity/offering of that organization. The main goal is to focus on the different analytic capabilities within each component of the analytics space. Many companies play in multiple sectors within the analytics industry and thus offer opportunities for movement within the field both horizontally and vertically.

Matt Turck, a venture capitalist with FirstMark has also developed and updates an analytics ecosystem focused on Big Data. His goal is to keep track of new and established players in various segments of the Big Data industry. A very nice visual image of his interpretation of the ecosystem and a comprehensive listing of companies is available through his Web site: http://mattturck.com/2016/02/01/big-data-landscape/ (accessed August 2016). We will also see a similar ecosystem in the context of the Internet of Things (IoT) in the last chapter.

Data Generation Infrastructure Providers

Perhaps the first place to begin identifying the clusters is by noting a new group of companies that enable generating and collection of data that may be used for developing analytical insights. Although this group could include all the traditional point-of-sale systems, inventory management systems, and technology providers for every step in a company's supply/value chain and operations, we mainly consider new players where the primary focus has been on enabling an organization to develop new insights into its operations as opposed to running its core operations. Thus this group includes companies creating the infrastructure for collecting data from different sources.

One of the emerging components of such an infrastructure is the "sensor." Sensors collect a massive amount of data at a faster rate and have been adopted by various sectors such as healthcare, sports, and energy. For example, health data collected by the sensors is generally used to track the health status of the users. Some of the major players manufacturing sensors to collect health information are AliveCor, Google, Shimmer, and Fitbit. Likewise, the sports industry is using sensors to collect data from the players and field to develop strategies and improve team play. Examples of the companies producing sports-related sensors include Sports Sensors, Zepp, Shockbox, and others. Similarly, sensors are used for traffic management. These help in taking real-time actions to control traffic. Some of the providers are Advantech B+B SmartWorx, Garmin, and Sensys Network.

Sensors play a major role in the Internet of Things and are an essential part of smart objects. These make machine-to-machine communication possible. The leading players in the infrastructure of IoT are Intel, Microsoft, Google, IBM, Cisco, Smartbin, SIKO Products, Omega Engineering, Apple, and SAP. This cluster is probably the most technical group in the ecosystem. We will review an ecosystem for IoT in Chapter 8. Indeed, there is an ecosystem around virtually each of the clusters we identify here.

Data Management Infrastructure Providers

This group includes all of the major organizations that provide hardware and software targeting the basic foundation for all data management solutions. Obvious examples of these include all major hardware players that provide the infrastructure for database computing—IBM, Dell, HP, Oracle, and so on; storage solution providers like EMC (recently bought by Dell) and NetApp; companies providing indigenous hardware and software platforms such as IBM, Oracle, and Teradata; and data solution providers offering hardware and platform independent database management systems like the SQL Server family of Microsoft and specialized integrated software providers such as SAP fall under this group. This group also includes other organizations such as database appliance providers, service providers, integrators, developers, and so on, that support each of these companies' ecosystems.

Several other companies are emerging as major players in a related space, thanks to the network infrastructure enabling cloud computing. Companies such as Amazon

(Amazon Web Services), IBM (Bluemix), and Salesforce.com pioneered to offer full data storage and analytics solutions through the cloud, which now have been adopted by several companies listed earlier.

A recent crop of companies in the Big Data space are also part of this group. Companies such as Cloudera, Hortonworks, and many others do not necessarily offer their own hardware but provide infrastructure services and training to create the Big Data platform. This would include Hadoop clusters, MapReduce, NoSQL, Spark, Kafka, Flume, and other related technologies for analytics. Thus they could also be grouped under industry consultants or trainers enabling the basic infrastructure. Full ecosystems of consultants, software integrators, training providers, and other value-added providers have evolved around many of the large players in the data management infrastructure cluster. Some of the clusters listed below will identify these players because many of them are moving to analytics as the industry shifts its focus from efficient transaction processing to deriving analytical value from the data.

Data Warehouse Providers

Companies with a data warehousing focus provide technology and services aimed toward integrating data from multiple sources, thus enabling organizations to derive and deliver value from its data assets. Many companies in this space include their own hardware to provide efficient data storage, retrieval, and processing. Companies such as IBM, Oracle, and Teradata are major players in this arena. Recent developments in this space include performing analytics on the data directly in memory. Another major growth sector has been data warehousing in the cloud. Examples of such companies include Snowflake and Redshift. Companies in this cluster clearly work with all the other sector players in providing DW solutions and services within their ecosystem and hence become the backbone of the analytics industry. It has been a major industry in its own right and, thus, a supplier and consumer of analytics talent.

Middleware Providers

Data warehousing began with a focus on bringing all the data stores into an enterprise-wide platform. Making sense of this data has become an industry in itself. The general goal of the middleware industry is to provide easy-to-use tools for reporting or descriptive analytics, which forms a core part of BI or analytics employed at organizations. Examples of companies in this space include Microstrategy, Plum, and many others. A few of the major players that were independent middleware players have been acquired by companies in the first two groups. For example, Hyperion became a part of Oracle, SAP acquired Business Objects, and IBM acquired Cognos. This sector has been largely synonymous with the BI providers offering dashboarding, reporting, and visualization services to the industry, building on top of the transaction processing data and the database and DW providers. Thus many companies have moved into this space over the years, including general analytics software vendors such as SAS or new visualization providers such as Tableau, or many niche application providers. A product directory at TDWI.org lists 201 vendors just in this category (http://www.tdwidirectory.com/category/business-intelligence-services) as of June 2016, so the sector has been robust. This is clearly also the sector attempting to move to a more data science segment of the industry.

Data Service Providers

Much of the data an organization uses for analytics is generated internally through its operations, but there are many external data sources that play a major role in any organization's decision making. Examples of such data sources include demographic data, weather data,

data collected by third parties that could inform an organization's decision making, and so on. Several companies realized the opportunity to develop specialized data collection, aggregation, and distribution mechanisms. These companies typically focus on a specific industry sector and build on their existing relationships in that industry through their niche platforms and services for data collection. For example, Nielsen provides data sources to their clients on customer retail purchasing behavior. Another example is Experian, which includes data on each household in the United States. Omniture has developed technology to collect Web clicks and share such data with their clients. Comscore is another major company in this space. Google compiles data for individual Web sites and makes a summary available through Google Analytics services. Other examples are Equifax, TransUnion, Acxiom, Merkle, Epsilon, and Avention. This can also include organizations such as ESRI.org, which provides location-oriented data to their customers. There are hundreds of other companies that are developing niche platforms and services to collect, aggregate, and share such data with their clients. As noted earlier, many industry-specific data aggregators and distributors exist and are moving to offer their own analytics services. Thus this sector is also a growing user and potential supplier of analytics talent, especially with specific niche expertise.

Analytics-Focused Software Developers

Companies in this category have developed analytics software for general use with data that has been collected in a DW or is available through one of the platforms identified earlier (including Big Data). It can also include inventors and researchers in universities and other organizations that have developed algorithms for specific types of analytics applications. We can identify major industry players in this space using the three types of analytics: descriptive, predictive, and prescriptive analytics.

REPORTING/DESCRIPTIVE ANALYTICS Reporting or descriptive analytics is enabled by the tools available from the middleware industry players identified earlier, or unique capabilities offered by focused providers. For example, Microsoft's SQL Server BI toolkit includes reporting as well as predictive analytics capabilities. On the other hand, specialized software is available from companies such as Tableau for visualization. SAS also offers a Visual Analytics tool with similar capacity. There are many open source visualization tools as well. Literally hundreds of data visualization tools have been developed around the world, and many such tools focus on visualization of data from a specific industry or domain. Because visualization is the primary way thus far for exploring analytics in industry, this sector has witnessed the most growth. Many new companies are being formed. For example, Gephi, a free and open source software, focuses on visualizing networks. A Google search will show the latest list of such software providers and tools.

PREDICTIVE ANALYTICS Perhaps the biggest recent growth in analytics has been in this category, and there are a large number of companies that focus on predictive analytics. Many statistical software companies such as SAS and SPSS embraced predictive analytics early on, and developed software capabilities as well as industry practices to employ data mining techniques and classical statistical techniques for analytics. IBM-SPSS Modeler from IBM and Enterprise Miner from SAS are some of the examples of tools used for predictive analytics. Other players in this space include KXEN, Statsoft (recently acquired by Dell), Salford Systems, and scores of other companies that may sell their software broadly or use it for their own consulting practices (next group of companies).

Three open source platforms (R, RapidMiner, and KNIME) have also emerged as popular industrial-strength software tools for predictive analytics and have companies that support training and implementation of these open source tools. Revolution Analytics is an example of a company focused on R development and training. R integration is possible with most analytics software. A company called Alteryx uses R extensions for reporting and predictive analytics, but its strength is in shared delivery of analytics solutions processes to customers and other users. Similarly, RapidMiner and KNIME are also examples of open source providers. Companies like Rulequest that sell proprietary variants of Decision Tree software and NeuroDimensions, a Neural Network software company, are examples of companies that have developed specialized software around a specific technique of data mining.

PRESCRIPTIVE ANALYTICS Software providers in this category offer modeling tools and algorithms for optimization of operations usually called management science/operations research software. This field has had its own set of major software providers. IBM, for example, has classic linear and mixed integer programming software. Several years ago, IBM also acquired a company called ILOG, which provides prescriptive analysis software and services to complement their other offerings. Analytics providers such as SAS have their own OR/MS tools—SAS/OR. FICO acquired another company called XPRESS that offers optimization software. Other major players in this domain include companies such as AIIMS, AMPL, Frontline, GAMS, Gurobi, Lindo Systems, Maximal, NGData, Ayata, and many others. A detailed delineation and description of these companies' offerings is beyond the scope of our goals here. Suffice it to say that this industry sector has seen much growth recently.

Of course, there are many techniques that fall under the category of prescriptive analytics, and each has their own set of providers. For example, simulation software is provided by major companies like Rockwell (ARENA) and Simio. Palisade provides tools that include many software categories. Similarly, Frontline offers tools for optimization with Excel spreadsheets, as well as predictive analytics. Decision analysis in multiobjective settings can be performed using tools such as Expert Choice. There are also tools from companies such as Exsys, XpertRule, and others for generating rules directly from data or expert inputs.

Some new companies are evolving to combine multiple analytics models in the Big Data space including social network analysis and stream mining. For example, Teradata Aster includes its own predictive and prescriptive analytics capabilities in processing Big Data streams. Several companies have developed complex event processing (CEP) engines that make decisions using streaming data, such as IBM's Infosphere Streams, Microsoft's StreamInsight, and Oracle's Event Processor. Other major companies that have CEP products include Apache, Tibco, Informatica, SAP, and Hitachi. It is worthwhile to note again that the provider groups for all three categories of analytics are not mutually exclusive. In most cases, a provider can play in multiple components of analytics.

We next introduce the "inside petals" of the analytics flower. These clusters can be called analytics accelerators. Although they may not be involved in developing the technology directly, these organizations have played a key role in shaping the industry.

Application Developers: Industry Specific or General

The organizations in this group use their industry knowledge, analytical expertise, solutions available from the data infrastructure, DW, middleware, data aggregators, and analytics software providers to develop custom solutions for a specific industry. Thus, this industry group makes it possible for analytics technology to be used in a specific industry. Of course, such groups may also exist in specific user organizations. Most major analytics technology providers like IBM, SAS, and Teradata clearly recognize the opportunity to connect to a specific industry or client and offer analytic consulting services. Companies that have traditionally provided application/data solutions to specific sectors are now developing industry-specific analytics offerings. For example, Cerner provides electronic medical records solutions to medical providers, and their offerings now include many analytics reports and visualizations. Similarly, IBM offers a fraud detection engine for the health insurance industry, and is working with an insurance company to employ their famous Watson analytics platform in assisting medical providers and insurance companies with diagnosis and disease management. Another example of a vertical application provider is Sabre Technologies, which provides analytical solutions to the travel industry including fare pricing for revenue optimization and dispatch planning.

This cluster also includes companies that have developed their own domain-specific analytics solutions and market them broadly to a client base. For example, Nike, IBM, and Sportvision develop applications in sports analytics to improve the play and increase the viewership. Acxiom has developed clusters for virtually all households in the United States based on the data they collect about households from many different sources. Credit score and classification reporting companies (FICO, Experian, etc.) also belong in this group. IBM and several other companies offer pricing optimization solutions in the retail industry.

This field represents an entrepreneurial opportunity to develop industry-specific applications. Many emerging in Web/social media/location analytics are trying to profile users for better targeting of promotional campaigns in real time. Examples of such companies and their activities include: YP.com employs location data for developing user/group profiles and targeting mobile advertisements, Towerdata profiles users on the basis of e-mail usage, Qualia aims to identify users through all device usage, and Simulmedia targets advertisements on TV on the basis of analysis of a user's TV watching habits.

The growth of smartphones has spawned a complete industry focused on specific analytics applications for consumers as well as organizations. For example, smartphone apps such as Shazam, Soundhound, or Musixmatch are able to identify a song on the basis of the first few notes and then let the user select it from their song base to play/download/purchase. Waze uses real-time traffic information shared by users, in addition to the location data, for improving navigation. Voice recognition tools such as Siri on the iPhone, Google Now, and Amazon Alexa are leading to many more specialized analytics applications for very specific purposes in analytics applied to images, videos, audio, and other data that can be captured through smartphones and/or connected sensors. Smartphones have also elevated the shared economy providers such as Uber, Lyft, Curb, and Ola. Many of these companies are exemplars of analytics leading to new business opportunities.

Online social media is another hot area in this cluster. Undoubtedly, Facebook is the leading player in the space of online social networking followed by Twitter and LinkedIn. Moreover, the public access to their data has given rise to multiple other companies that analyze their data. For example, Unmetric analyzes Twitter data and provides solutions to their clients. Similarly, there are several other companies that focus on social network analysis.

A trending area in the application development industry is the IoT. Several companies are building applications to make smart objects. For example, SmartBin has developed intelligent remote monitoring systems for the waste and recycling sectors. Several other organizations are working on building smart meters, smart grids, smart cities, connected cars, smart homes, smart supply chains, connected health, smart retail, and other smart objects.

This start-up activity and space is growing and is in major transition due to technology/venture funding and security/privacy issues. Nevertheless, the application developer sector is perhaps the biggest growth industry within analytics at this point. This cluster provides a unique opportunity for analytics professionals looking for more entrepreneurial career options.

Analytics Industry Analysts and Influencers

The next cluster of the analytics industry includes three types of organizations or professionals. The first group is the set of professional organizations that provide advice to the analytics industry providers and users. Their services include marketing analyses, coverage of new developments, evaluation of specific technologies, and development of training/white papers/ and so on. Examples of such players include organizations such as the Gartner Group, The Data Warehousing Institute, Forrester, McKinsey, and many of the general and technical publications and Web sites that cover the analytics industry. Gartner Group's Magic Quadrants are highly influential and are based on industry surveys. Similarly, TDWI.org professionals provide excellent industry overviews and are very aware of current and future trends of this industry.

The second group includes professional societies or organizations that also provide some of the same services but are membership based and organized. For example, INFORMS, a professional organization, has now focused on promoting analytics. Special Interest Group on Decision Support and Analytics, a subgroup of the Association for Information Systems, also focuses on analytics. Most of the major vendors (e.g., Teradata and SAS) also have their own membership-based user groups. These entities promote the use of analytics and enable sharing of the lessons learned through their publications and conferences. They may also provide recruiting services, and are thus good sources for locating talent.

A third group of analytics industry analysts is what we call analytics ambassadors, influencers, or evangelists. These analysts have presented their enthusiasm for analytics through their seminars, books, and other publications. Illustrative examples include Steve Baker, Tom Davenport, Charles Duhigg, Wayne Eckerson, Bill Franks, Malcolm Gladwell, Claudia Imhoff, Bill Inman, and many others. Again, the list is not inclusive. All of these ambassadors have written books (some of them bestsellers!) and/or given many presentations to promote the analytics applications. Perhaps another group of evangelists to include here is the authors of textbooks on BI/analytics who aim to assist the next cluster to produce professionals for the analytics industry. Clearly, it will take some time for an analytics student to become a member of this cluster, but they could be working with members of this cluster as researchers or apprentices.

Academic Institutions and Certification Agencies

In any knowledge-intensive industry such as analytics, the fundamental strength comes from having students who are interested in the technology and choosing that industry as their profession. Universities play a key role in making this possible. This cluster, then, represents the academic programs that prepare professionals for the industry. It includes various components of business schools such as information systems, marketing, management sciences, and so on. It also extends far beyond business schools to include computer science, statistics, mathematics, and industrial engineering departments across the world. The cluster also includes graphics developers who design new ways of visualizing information. Universities are offering undergraduate and graduate programs in analytics in all of these disciplines, though they may be labeled differently. A major growth frontier has been certificate programs in analytics to enable current professionals to retrain and retool themselves for analytics careers. Certificate programs enable practicing analysts to gain basic proficiency in specific software by taking a few critical courses from schools that offer these programs. TUN includes a list of analytics programs. It includes almost 150 programs, and there are likely many more such programs, with new ones being added daily.

Another group of players assists with developing competency in analytics. These are certification programs that award a certificate of expertise in specific software. Virtually every major technology provider (IBM, Microsoft, Microstrategy, Oracle, SAS, Tableau, and Teradata) has their own certification programs. These certificates ensure that potential new hires have a certain level of tool skills. On the other hand, INFORMS offers a Certified Analytics Professional certificate program that is aimed at testing an individual's general analytics competency. Any of these certifications give a college student additional marketable skills.

The growth of academic programs in analytics is staggering. Only time will tell if this cluster is overbuilding the capacity that can be consumed by the other clusters, but at this point, the demand appears to outstrip the supply of qualified analytics graduates, and this is the most obvious place to find at least entry-level analytics hires.

Regulators and Policy Makers

The players in this component are responsible for defining rules and regulations for protecting employees, customers, and shareholders of the analytics organizations. The collection and sharing of the users' data require strict laws for securing privacy. Several organizations in this space regulate the data transfer and protect users' rights. For example, the Federal Communications Commission (FCC) regulates interstate and international communications. Similarly, the Federal Trade Commission (FTC) is responsible for preventing data-related unfair business practices. The International Telecommunication Union (ITU) regulates the access to information and communication technologies (ICTs) to underserved communities worldwide. On the other hand, a nonregulatory federal agency named the National Institute of Standards and Technology (NIST), helps advance the technology infrastructure. There are several other organizations across the globe that regulate the data security and accelerate the analytics industry. This is a very important component in the ecosystem so that no one can misuse consumers' information.

For anyone developing or using analytics applications, it is perhaps crucial to have someone on the team who is aware of the regulatory framework. These agencies and professionals who work with them clearly offer unique analytics talents and skills.

Analytics User Organizations

Clearly, this is the economic engine of the whole analytics industry, and therefore, we represent this cluster as the core of the analytics flower. If there were no users, there would be no analytics industry. Organizations in every industry, regardless of size, shape, and location, are using or exploring the use of analytics in their operations. These include the private sector, government, education, military, and so on. It includes organizations around the world. Examples of uses of analytics in different industries abound. Others are exploring similar opportunities to try and gain/retain a competitive advantage. Specific companies are not identified in this section; rather, the goal here is to see what type of roles analytics professionals can play within a user organization.

Of course, the top leadership of an organization, especially in the information technology group (chief information officer, etc.), is critically important in applying analytics to its operations. Reportedly, Forrest Mars of the Mars Chocolate Empire said that all management boiled down to applying mathematics to a company's operations and economics. Although not enough senior managers subscribe to this view, the awareness of applying analytics within an organization is growing everywhere. A health insurance company executive once told us that his boss (the CEO) viewed the company as an IT-enabled organization that collected money from insured members and distributed it to the providers. Thus efficiency in this process was the premium they could earn over a competitor. This led the company to develop several analytics applications to reduce fraud and overpayment to providers, promote wellness among those insured so they would use the providers less often, generate more efficiency in processing, and thus be more profitable.

Virtually all major organizations in every industry that we are aware of are hiring analytical professionals under various titles. Figure 1.14 is a word cloud of the selected titles of our program graduates at Oklahoma State University from 2013 to 2016. It clearly shows that Analytics and Data Science are popular titles in the organizations hiring graduates of such programs. Other key words appear to include terms such as Risk, Database, Security, Revenue, Marketing, and so on.

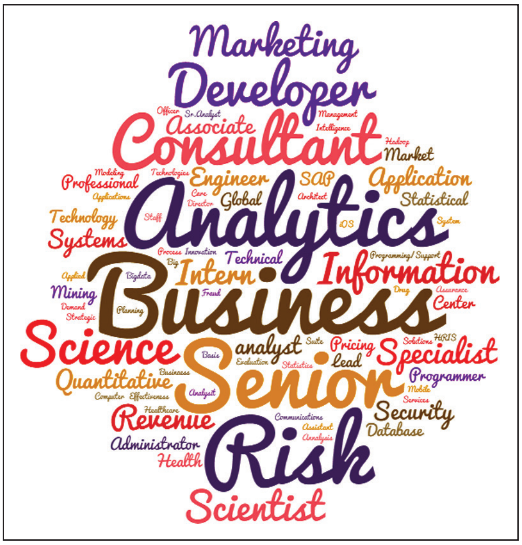

FIGURE 1.14 Word Cloud of Job Titles of Analytics Program Graduates.

Of course, user organizations include career paths for analytics professionals moving into management positions. These titles include project managers, senior managers, and directors, all the way up to the chief information officer or chief executive officer. This suggests that user organizations exist as a key cluster in the analytics ecosystem and thus can be a good source of talent. It is perhaps the first place to find analytics professionals within the vertical industry segment.

The purpose of this section has been to present a map of the landscape of the analytics industry. Eleven different groups that play a key role in building and fostering this industry were identified. More petals/components can be added over time in the analytics flower/ecosystem. Because data analytics requires a diverse skill set, understanding of this ecosystem provides you with more options than you may have imagined for careers in analytics. Moreover, it is possible for professionals to move from one industry cluster to another to take advantage of their skills. For example, expert professionals from providers can sometimes move to consulting positions, or directly to user organizations. Overall, there is much to be excited about the analytics industry at this point.

SECTION 1.8 REVIEW QUESTIONS

1. List the 11 categories of players in the analytics ecosystem.
2. Give examples of companies in each of the 11 types of players.
3. Which companies are dominant in more than one category?
4. Is it better to be the strongest player in one category or be active in multiple categories?

1.9 Plan of the Book

The previous sections have given you an understanding of the need for information technology in decision making, the evolution of BI, and now into analytics and data science. In the last several sections we have seen an overview of various types of analytics and their applications. Now we are ready for a more detailed managerial excursion into these topics, along with some deep hands-on experience in some of the technical topics. Figure 1.15 presents a plan on the rest of the book.

In this chapter, we have provided an introduction, definitions, and overview of DSSs, BI, and analytics, including Big Data analytics and data science. We also gave you an

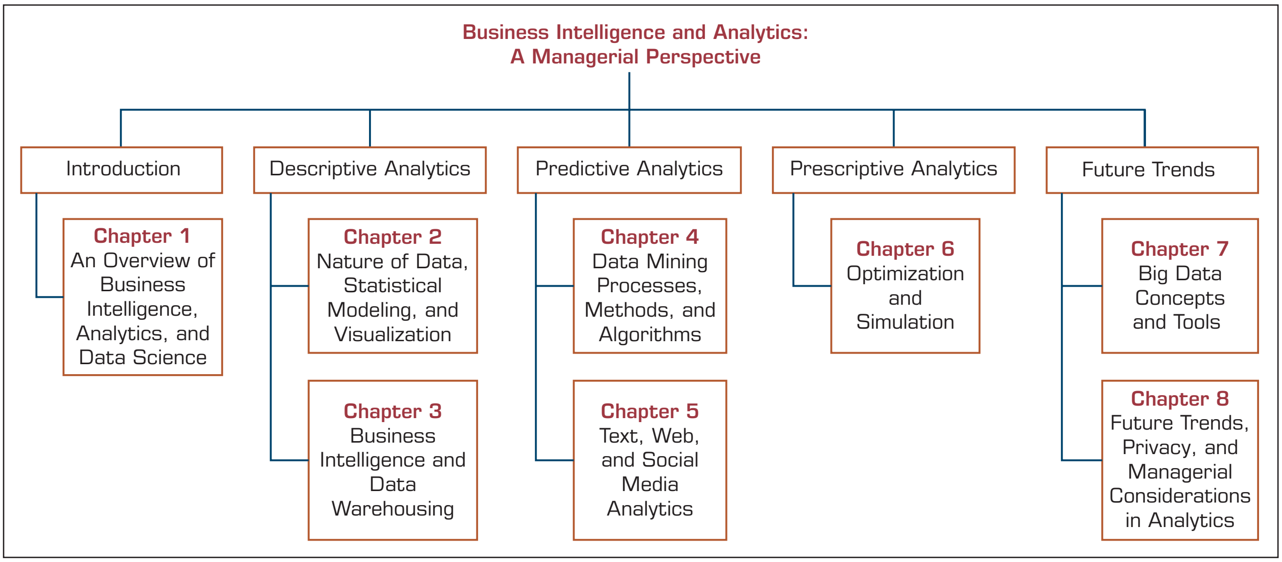

FIGURE 1.15 Plan of the Book.

overview of the analytics ecosystem to have you appreciate the breadth and depth of the industry. Chapters 2 and 3 cover descriptive analytics and data issues. Data clearly form the foundation for any analytics application. Thus we cover an introduction to data warehousing issues, applications, and technologies. This section also covers business reporting and visualization technologies and applications. This is followed by a brief overview of BPM techniques and applications—a topic that has been a key part of traditional BI.

The next section covers predictive analytics. Chapter 4 provides an introduction to predictive analytics applications. It includes many of the common data mining techniques: classification, clustering, association mining, and so forth. Chapter 5 focuses on text mining applications as well as Web analytics, including social media analytics, sentiment analysis, and other related topics. Chapter 6 covers prescriptive analytics. Chapter 7 includes more details of Big Data analytics. Chapter 8 includes a discussion of emerging trends. The ubiquity of wireless and GPS devices and other sensors is resulting in the creation of massive new databases and unique applications. A new breed of analytics companies is emerging to analyze these new databases and create a much better and deeper understanding of customers' behaviors and movements. It is leading to the automation of analytics and has also spanned a new area called the "Internet of Things." The chapter also covers cloud-based analytics, Finally, Chapter 8 also attempts to integrate all the material covered in this book and concludes with a brief discussion of security/privacy dimensions of analytics.

1.10 Resources, Links, and the Teradata University Network Connection

The use of this chapter and most other chapters in this book can be enhanced by the tools described in the following sections.

Resources and Links

We recommend the following major resources and links:

- The Data Warehousing Institute (tdwi.org)
- Data Science Central (datasciencecentral.com)

- DSS Resources (dssresources.com)
- Microsoft Enterprise Consortium (enterprise.waltoncollege.uark.edu/mec.asp)

Vendors, Products, and Demos

Most vendors provide software demos of their products and applications. Information about products, architecture, and software is available at dssresources.com.

Periodicals

We recommend the following periodicals:

- *Decision Support Systems* (www.journals.elsevier.com/decision-support-systems)
- *CIO Insight* (cioinsight.com)

The Teradata University Network Connection

This book is tightly connected with the free resources provided by TUN (see teradata universitynetwork.com). The TUN portal is divided into two major parts: one for students and one for faculty. This book is connected to the TUN portal via a special section at the end of each chapter. That section includes appropriate links for the specific chapter, pointing to relevant resources. In addition, we provide hands-on exercises, using software and other material (e.g., cases) available at TUN.

The Book's Web Site

This book's Web site, pearsonhighered.com/sharda, contains supplemental textual material organized as Web chapters that correspond to the printed book's chapters. The topics of these chapters are listed in the online chapter table of contents.[1]

[1]As this book went to press, we verified that all cited Web sites were active and valid. However, URLs are dynamic. 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. Many 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 possible new site. Most times, you can quickly find the new site through one of the popular search engines. We apologize in advance for this inconvenience.

Chapter Highlights

- The business environment is becoming more complex and is rapidly changing, making decision making more difficult.
- Businesses must respond and adapt to the changing environment rapidly by making faster and better decisions.
- The time frame for making decisions is shrinking, whereas the global nature of decision making is expanding, necessitating the development and use of computerized DSSs.
- DSSs use data, models, and sometimes knowledge management to find solutions for semistructured and some unstructured problems.
- BI methods utilize a central repository called a DW that enables efficient data mining, OLAP, BPM, and data visualization.
- BI architecture includes a DW, business analytics tools used by end users, and a user interface (such as a dashboard).
- Many organizations employ descriptive analytics to replace their traditional flat reporting with interactive reporting that provides insights, trends, and patterns in the transactional data.
- Predictive analytics enable organizations to establish predictive rules that drive the business outcomes through historical data analysis of the existing behavior of the customers.
- Prescriptive analytics help in building models that involve forecasting and optimization techniques based on the principles of OR and management science to help organizations to make better decisions.
- Big Data analytics focuses on unstructured, large data sets that may also include vastly different types of data for analysis.
- Analytics as a field is also known by industry-specific application names, such as sports analytics. It is also known by other related names such as data science or network science.
- Healthcare and retail chains are two areas where analytics applications abound, with much more to come.
- The analytics ecosystem can be first viewed as a collection of providers, users, and facilitators. It can be broken into 11 clusters.

Key Terms

analytics
analytics ecosystem
Big data analytics
business intelligence (BI)
dashboard
data mining
decision or normative analytics
descriptive (or reporting) analytics
intelligent agents
online analytical processing (OLAP)
online transaction processing (OLTP)
predictive analytics
prescriptive analytics
Web services

Questions for Discussion

1. Survey the literature from the past 6 months to find one application each for DSS, BI, and analytics. Summarize the applications on one page, and submit it with the exact sources.
2. Distinguish BI from DSS.
3. Compare and contrast predictive analytics with prescriptive and descriptive analytics. Use examples.
4. Discuss the major issues in implementing BI.

Exercises

Teradata University Network and Other Hands-On Exercises

1. Go to teradatauniversitynetwork.com. Using the site password your instructor provides, register for the site if you have not already previously registered. Log on and learn the content of the site. You will receive assignments related to this site. Prepare a list of 20 items on the site that you think could be beneficial to you.
2. Go to the TUN site. Explore the Sports Analytics page, and summarize at least two applications of analytics in any sport of your choice.
3. Enter the TUN site, and select "Cases, Projects, and Assignments." Then select the case study "Harrah's High Payoff from Customer Information." Answer the following questions about this case:
   - **a.** What information does the data mining generate?
   - **b.** How is this information helpful to management in decision making? (Be specific.)
   - **c.** List the types of data that are mined.
   - **d.** Is this a DSS or BI application? Why?
4. Go to teradatauniversitynetwork.com and find the paper titled "Data Warehousing Supports Corporate Strategy at First American Corporation" (by Watson, Wixom, and Goodhue). Read the paper, and answer the following questions:
   - **a.** What were the drivers for the DW/BI project in the company?
   - **b.** What strategic advantages were realized?
   - **c.** What operational and tactical advantages were achieved?
   - **d.** What were the critical success factors for the implementation?
5. Go to http://analytics-magazine.org/issues/digital-editions and find the January/February 2012 edition titled "Special Issue: The Future of Healthcare." Read the article "Predictive Analytics—Saving Lives and Lowering Medical Bills." Answer the following questions:
   - **a.** What problem is being addressed by applying predictive analytics?
   - **b.** What is the FICO Medication Adherence Score?
   - **c.** How is a prediction model trained to predict the FICO Medication Adherence Score HoH? Did the prediction model classify the FICO Medication Adherence Score?
   - **d.** Zoom in on Figure 4, and explain what kind of technique is applied on the generated results.
   - **e.** List some of the actionable decisions that were based on the prediction results.
6. Go to http://analytics-magazine.org/issues/digital-editions, and find the January/February 2013 edition titled "Work Social." Read the article "Big Data, Analytics and Elections," and answer the following questions:
   - **a.** What kinds of Big Data were analyzed in the article Coo? Comment on some of the sources of Big Data.
   - **b.** Explain the term *integrated system*. What is the other technical term that suits an *integrated system*?
   - **c.** What kinds of data analysis techniques are employed in the project? Comment on some initiatives that resulted from data analysis.
   - **d.** What are the different prediction problems answered by the models?
   - **e.** List some of the actionable decisions taken that were based on the prediction results.
   - **f.** Identify two applications of Big Data analytics that are not listed in the article.
7. Search the Internet for material regarding the work of managers and the role analytics plays. What kinds of references to consulting firms, academic departments, and programs do you find? What major areas are represented? Select five sites that cover one area, and report your findings.
8. Explore the public areas of dssresources.com. Prepare a list of its major available resources. You might want to refer to this site as you work through the book.
9. Go to microstrategy.com. Find information on the five styles of BI. Prepare a summary table for each style.
10. Go to oracle.com, and click the Hyperion link under Applications. Determine what the company's major products are. Relate these to the support technologies cited in this chapter.
11. Go to the TUN questions site. Look for BSI videos. Review the video of the "Case of Retail Tweeters." Prepare a one-page summary of the problem, proposed solution, and the reported results. You can also find associated slides on slideshare.net.
12. Review the Analytics Ecosystem section. Identify at least two additional companies in at least five of the industry clusters noted in the discussion.
13. The discussion for the analytics ecosystem also included several typical job titles for graduates of analytics and data science programs. Research Web sites such datasciencecentral.com and tdwi.org to locate at least three additional similar job titles that you may find interesting for your career.

References

Capcredit.com. (2015). "How Much Do Americans Spend on Sports Each Year?" capcredit.com/how-much-americans-spend-on-sports-each-year/ (accessed July 2016).

CDC.gov. (2015, September 21). "Important Facts about Falls." cdc.gov/homeandrecreationalsafety/falls/adultfalls.html (accessed July 2016).

CenterPointEnergy.com. "Company Overview." centerpoint energy.com/en-us/Corp/Pages/Company-overview.aspx (accessed June 2016).

Chiguluri, V., Guthikonda, K., Slabaugh, S., Havens, E., Peña, J., & Cordier, T. (2015, June). *Relationship between diabetes complications and health related quality of life among an elderly population in the United States*. Poster presentation at the American Diabetes Association 75th Annual Scientific Sessions. Boston, MA.

Cordier, T., Slabaugh, L., Haugh, G., Gopal, V., Cusano, D., Andrews, G., & Renda, A. (2015, September). *Quality of life changes with progressing congestive heart failure*. Poster presentation at the 19th Annual Scientific Meeting of the Heart Failure Society of America, Washington, DC.

Davenport, T., & SAS Institute Inc. (2014, February). Analytics in sports: The new science of winning. sas.com/content/dam/SAS/en_us/doc/whitepaper2/iia-analytics-in-sports-106993.pdf (accessed July 2016).

Davenport, T. H., & Patil, D. J. (2012). Data scientist. *Harvard Business Review*, *90*, 70–76.

Dundas.com. "How Siemens Drastically Reduced Cost With Managed Bi Applications." dundas.com/resource/getcasestudy?caseStudyName=09-03-2016-Siemens%2FDundas-BI-Siemens-Case-Study.pdf (accessed July 2016).

Emc.com. (n.d.). "Data science revealed: A data-driven glimpse into the burgeoning new field". emc.com/collateral/about/news/emc-data-science-study-wp.pdf (accessed July 2016).

Gartner, Inc. (2004). *Using business intelligence to gain a competitive edge. A special report*.

Gates, S., Smith, L. A., Fisher, J. D., et al. (2008). Systematic review of accuracy of screening instruments for predicting fall risk among independently living older adults. *Journal of Rehabilitation Research and Development, 45*(8), 1105–1116.

Gill, T. M., Murphy, T. E., Gahbauer, E. A., et al. (2013). Association of injurious falls with disability outcomes and nursing home admissions in community living older persons. *American Journal of Epidemiology, 178*(3), 418–425.

Gorry, G. A., & Scott-Morton, M. S. (1971). A framework for management information systems. *Sloan Management Review*, *13*(1), 55–70.

Keen, P. G. W., & M. S. Scott-Morton. (1978). *Decision support systems: An organizational perspective*. Reading, MA: Addison-Wesley.

Havens, E., Peña, J., Slabaugh, S., Cordier, T., Renda, A., & Gopal, V. (2015, October). Exploring the relationship between health-related quality of life and health conditions, costs, resource utilization, and quality measures. Podium presentation at the ISOQOL 22nd Annual Conference, Vancouver, Canada.

Havens, E., Slabaugh, L., Peña, J., Haugh, G., & Gopal, V. (2015, February). *Are there differences in Healthy Days based on compliance to preventive health screening measures?* Poster presentation at Preventive Medicine 2015, Atlanta, GA.

Humana. *2016 progress report*. populationhealth.humana.com/wp-content/uploads/2016/05/BoldGoal2016ProgressReport_1.pdf (accessed July 2016).

INFORMS. *Analytics section overview*. informs.org/Community/Analytics (accessed July 2016).

NCAA, National Center for Catastrophic Sports Injury Research Report. NCAA Sport Injury fact sheets are produced by the Datalys Center for Sports Injury Research and Prevention in collaboration with the National Collegiate Athletic Association, and STOP Sports Injuries. *Women's soccer injuries*. ncaa.org/sites/default/files/NCAA_W_Soccer_Injuries_WEB.pdf (accessed July 2016).

Pajouh Foad, M., Xing, D., Hariharan, S., Zhou, Y., Balasundaram, B., Liu, T., & Sharda, R. (2013). Available-to-promise in practice: An application of analytics in the specialty steel bar products industry. *Interfaces*, *43*(6), 503–517. dx.doi.org/10.1287/inte.2013.0693 (accessed July 2016).

Price Waterhouse Coopers Report. (2011, December). *Changing the game: Outlook for the global sports market to 2015*. pwc.com/gx/en/hospitality-leisure/pdf/changing-the-game-outlook-for-the-global-sports-market-to-2015.pdf (accessed July 2016).

Sap.com. (2013, October). A "Smart" Approach to Big Data in the Energy Industry. sap.com/bin/sapcom/cs_cz/downloadasset.2013-10-oct-09-20.a-smart-approach-to-big-data-in-the-energy-industry-pdf.html (accessed June 2016).

Sharda, R., Asamoah, D., & Ponna, N. (2013). Research and pedagogy in business analytics: Opportunities and illustrative examples. *Journal of Computing and Information Technology*, *21*(3), 171–182.

Siemens.com. *About Siemens*. siemens.com/about/en/ (accessed July 2016).

Silvaris.com. *Silvaris overview*. silvaris.com/About/ (accessed July 2016).

Simon, H. (1977). *The New Science of Management Decision*. Englewood Cliffs, NJ: Prentice Hall.

Tableau.com. *Silvaris augments proprietary technology platform with Tableau's real-time reporting capabilities*. tableau.com/sites/default/files/case-studies/silvaris-business-dashboards_0.pdf (accessed July 2016).

TeradataUniversityNetwork.com. (2015, Fall). *BSI: Sports analytics—Precision football* (video). teradatauniversitynetwork.com/About-Us/Whats-New/BSI–Sports-Analytics-Precision-Football/ (accessed July 2016).

Terry, D. (2011), "Sabre Streamlines Decision Making," http://www.teradatamagazine.com/v11n04/Features/Sabre-Streamlines-Decision-Making/ (Accessed July 2016).

Turck, Matt, "Is Big Data Still a Thing? (The 2016 Big Data Landscape)." http://mattturck.com/2016/02/01/big-data-landscape/ (accessed August 2016)

Watson, H. (2005, Winter). Sorting out what's new in decision support. *Business Intelligence Journal*.

Wikipedia.org. *On-base percentage*. en.wikipedia.org/wiki/On_base_percentage (accessed January 2013).

Wikipedia.org. *Sabermetrics*. en.wikipedia.org/wiki/Sabermetrics (accessed January 2013).

Wikipedia.org. *SIEMENS*. en.wikipedia.org/wiki/Siemens (accessed July 2016).

Wintergreen Research Press Release (PR Newswire). (2015, June 25). *Sports analytics market worth $4.7 billion by 2021*. prnewswire.com/news-releases/sports-analytics-market-worth-47-billion-by-2021-509869871.html (accessed July 2016).

YouTube.com. (2013, December 17). *CenterPoint energy talks real time big data analytics*. youtube.com/watch?v=s7CzeSlIEfI (accessed June 2016).