PRACTICAL ANALYTICS

APPLIED ANALYTICS CONCEPTS USING MARKET-LEADING SOFTWARE TOOLS

NITIN KALÉ AND NANCY JONES

EPISTEMY PRESS

CHAPTER 1
Data Analytics Overview

LEARNING OBJECTIVES

After completing this chapter, you will be able to:

Describe what data analytics is.

Explain why the study of analytics is important.

Recap examples of analytics in real-world situations, particularly business scenarios.

Describe the structure of the model company and some of its employees who appear in many of the examples in this text.

WHAT IS DATA ANALYTICS?

Before we begin a discussion of data analytics, we need to define what it is. **Data analytics** is a process that involves: (a) identifying the problem, (b) gathering relevant data that frequently are not in a usable form, (c) cleaning up the data to make them usable, (d) loading them into data storage models, (e) manipulating them to discover information that leads to actionable insights, and (f) making decisions based on those insights. From data to insights to decisions, data analytics enables us to answer the following questions:

- What has happened in the past?
- Why did it happen?
- What could happen in the future? With what certainty?
- What actions can we take now to support or prevent certain events from happening in the future?
- Can some of the actions resulting from our insights be automated?
- Can the analytics process be automated?

Data analytics is the process that takes us from data to decision. Figure 1-1 illustrates the steps.

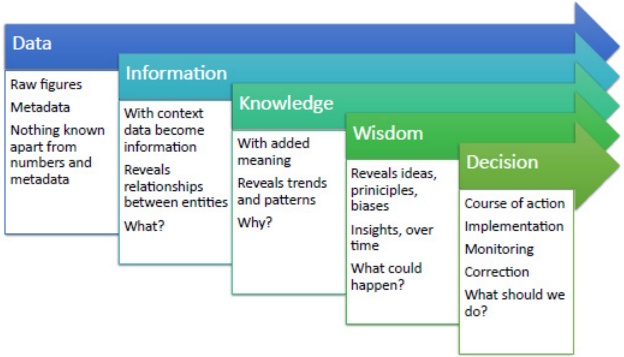

Figure 1-1: Data to Decision

People frequently use the term "data" and "information" interchangeably. In fact, they are distinct concepts. **Data** are the raw figures, numbers, or text that serve as the starting point of analysis. They can be stored digitally or even distributed across the internet. An example of the type of data that businesses typically analyze is sales revenue for each customer. Data become **information** when they reveal the causes or results of the event. For example, we could process sales revenue data to display the average revenue per customer or to reveal which customers did not make any purchases from us within a given time period.

When information is given meaning, we gain knowledge and understanding of the data. Knowledge is created when we learn from information; for example, which customers did not buy from us because of pricing and which customers did not buy from us because of quality? Knowledge provides the answer to why something is the way it is. This knowledge then provides us with the capabilities to gain wisdom. Wisdom, in this context, is acquired when knowledge is gathered over time. Wisdom is the deep understanding of underlying principles and behaviors. For example, by examining the reasons why customers do not buy from us, over time we can acquire the wisdom to identify and implement policies our company should pursue to retain customers. Finally, wisdom coupled with a goal yields a decision. We can implement the decision into action to influence and guide our direction. For instance, if our goal is to retain customers, then we might launch loyalty programs.

As data analytics has moved out of the realm of statistics into business and other fields, the vocabulary of data analytics has evolved as well. What began as a purely mathematical endeavor has evolved to encompass diverse concepts. Many of the terms associated with these concepts have overlapping meaning and scope. Figure 1-2 illustrates the relationship among analytics in three areas: statistics, computer science, and domain knowledge. *Statistics* is a rigorous branch of mathematics that deals with understanding data. It involves the collection or sampling, organization, modeling, analysis, interpretation, and presentation of data. Probability theory plays an important role in statistical inference. *Computer science* is the study of how computers work and the application of theory to improve computing methods and capabilities. Finally, *domain knowledge* relates to the expertise gained by individuals in certain areas or fields. For example, medicine is a domain. Business in general as well as specific types of businesses are also domains. Each domain has its distinctive vocabulary and analytical applications.

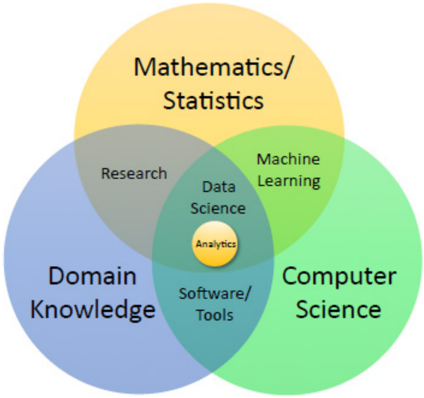

Figure 1-2: The World of Analytics

Data analytics is both supported by and a subset of **data science,** which is the intersection of computer science, statistics and domain knowledge. Data science involves the use of computers to acquire knowledge by analyzing large amounts of data using models and domain expertise. A practitioner of data science is called a **data scientist**. Data scientists are specially trained in mathematics, computer science, and statistics. Data analytics applies the algorithms and models generated from data science. Therefore, analytics is more concerned with applying models than with creating them. Figure 1-3 displays several analytics applications. We discuss each of these applications in the following sections.

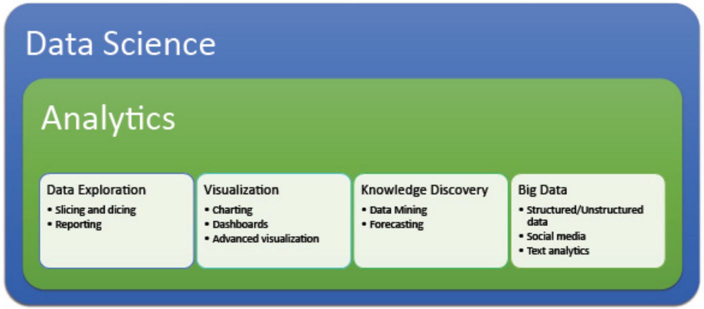

Figure 1-3: Data Analytics is a Subset of Data Science

Business intelligence (BI) has been used to describe analytics in the context of business data. BI focuses primarily on business data, financial data, and marketing data to gain business value, customer loyalty, and other benefits. It is largely descriptive in nature, recounting what has happened in the past. The word "intelligence" was used to describe the insights gained by analyzing business data. "Data analytics" or simply "analytics" is a newer, broader term that encompasses business intelligence. We use "analytics" in this text to describe the techniques used to obtain business intelligence and insights for decision making.

WHY STUDY ANALYTICS?

Analytics is an exciting field that has applications in all walks of life. The demand for people who possess the skills to understand and analyze data has grown to the point that employers are pushing educators to better train students in the fundamentals of data analysis.

[1] *Science Daily* reported in 2013 that 90% of the data available at that time were generated in the preceding two years. That trend continues today, as business professionals, scientists, healthcare providers, public administrators, and athletes, among others, are scrambling to determine how to utilize the available data to obtain greater insights and make better decisions. According to [2] respondents of a 2013 study by the American Management Association (AMA) the top five drivers for better analytics are: accountability for results, the competitive environment, business complexity, the increase in data availability, and risk management. Subsequent studies have echoed these findings.

BUSINESS AND DATA ANALYTICS

In that same AMA study, 99% of the business leaders polled indicated that analytics will be important to their organizations either at the time of the survey or in the five-year period immediately succeeding it. This observation is supported [3] by a 2015 white paper by Price Waterhouse Cooper in which 85% of CEOs in the surveyed companies placed a high value on data analytics and *data mining* (Data mining will be discussed in Section 5.) In fact, they targeted data analytics [4] as their top IT spending priority. A report in *Forbes Tech* disclosed that 89% of the companies surveyed believe that businesses that do not adopt a major data analytics strategy risk losing market share and momentum. Clearly, data analytics is critical in today's business environment.

As stated earlier, businesses are encouraging educators to place greater emphasis on data analysis to address a growing need for workers with these skills. [5] To highlight this need, Accenture and GE surveyed a cross-section of large companies in eight industries regarding their perceptions of *big data* and data analysis, and they published the results in a 2015 report. Roughly 50% of the respondents acknowledged they have talent shortfalls in these fields. Similarly, in the AMA survey cited in the previous paragraph, only 25% of the companies felt they had sufficiently trained personnel to meet their analytics needs.

Let's examine the various levels of data interpretation and utilization skills. These skills can be categorized into the following three areas of responsibility. (1) The data scientist is the most specialized analyst, possessing advanced math and computing skills. (2) Data analysts specialize in analysis, and they come from various educational backgrounds. They typically have an advanced college degree in business, usually with an emphasis in quantitative methods and additional training in data modeling and manipulation. (3) Significantly, most data analysis is being performed by managers and other business users, who may have not received formal education related to data analysis. These relationships are illustrated in Figure 1-4

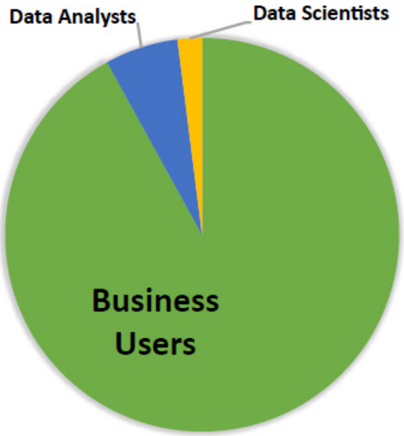

Figure 1-4: Data Analytics Skill Levels

APPLICATIONS OF ANALYTICS

The demand for analytical skills is clearly substantial. But, when you have acquired these skills, what can you do with them? The following are brief examples of how analytics is applied in various fields.

RETAIL:

Analytics is used extensively to assist in pricing, timing of pricing strategies, and amount of discounts; product placement; and up-selling and cross-selling of products.

MANUFACTURING:

A type of data analysis called *demand forecasting* is the core of manufacturing planning.

MARKETING:

Targeted marketing is made possible with predictive analytics. Marketers can analyze each customer's behavior and use these data to predict future behaviors. They can then launch targeted individualized marketing campaigns based on this information. The effectiveness of the campaign is measured based on customer response.

SUPPLY CHAIN:

Selecting suppliers and optimizing distribution costs both utilize data analytics techniques.

CUSTOMER SERVICE/HELP DESK:

Customized service is based on analysis of prior work orders or help tickets, the procedures that succeeded, problem-solving metrics, and data from various sources.

FORECASTING AND BUDGETING:

Most businesses create forecasts and budgets based on historical data and knowledge of the business environment. The more accurate these tools are the more likely management will make appropriate operational decisions. The use of large quantities of data and analytic tools helps improve forecasting and budgeting accuracy.

AUDIT AND ANALYSIS OF INTERNAL CONTROLS:

Data from internal systems are used to analyze risk and determine how well systems comply with management's policy of internal controls.

GOVERNMENT:

Governments also use data analytics. Below are two examples.

The federal government collects census data every 10 years. These data are made available to researchers, policymakers, state and local government agencies, nonprofit organizations, and the general public. The purpose of the census is to gather demographic data that can lead to better resource allocation and assist in the formulation of effective public policies.

Data analysis assists governments in collecting the revenues due them and in the correct amounts. Based on prior tax returns and other factors, governments can use analytical techniques to sift through tax returns or even identify missing returns to red-flag taxpayers for audits and notices of noncompliance.

UTILITIES:

Data analytics assists utility companies in predicting consumer demands for power and managing the supply of power from producers. Grid-tied power from residences and businesses that generate solar or wind power requires the utilities to balance all of the sources and consumption of power that is facilitated via data analytics.

FINANCIAL INVESTORS:

Investors sift through the data of numerous companies to determine which are acceptable investments and which are high risk or unacceptable.

SCIENTISTS:

Scientists and researchers in many fields of study frequently collect vast quantities of data that must be analyzed and interpreted. They gather the data during the course of experimentation, simulations, and samplings, sometimes via sensors. Scientists have been important consumers of data analytics for a considerable period of time.

MEDICINE:

Analytics has a variety of medical applications. One application is to identify risk factors that lead to chronic diseases. Data from a specific population are collected over time: diet, exercise, ethnicity, gender, age, occupation, family history, and many other attributes. Factors that influence (both positively and negatively) the outcome are determined using an analytical technique called data mining. Medical practitioners then utilize this knowledge to make decisions regarding prevention and treatment.

Another use of analytics in medicine is for disease prevention and control. By understanding people's day-to-day habits, exposures to pathogens (disease-causing agents), behaviors of diseases, prior outbreaks, and similar factors, healthcare providers can respond to health threats in a meaningful and timely manner.

SPORTS:

While sports analytics has its beginnings in the mid 1900's, it did not gain common acceptance until relatively recently. A 2011 movie, "Moneyball" featured Sabermetrics (the application of statistical methods to baseball data) and thus propelled baseball analytics into mainstream conversation. Those involved in other types of sports have also seen the value in data analysis resulting in rapid growth in the sports analytics field. Coaches can analyze team and player statistics to develop winning strategies. They can also analyze their players' physiological data to determine whether their performance can be improved.

FRAUD PREVENTION:

Data analytics is particularly valuable for fraud detection and prevention. Analytic techniques enable investigators to flag unusual activities for further investigation. After investigators have identified fraud, they can take corrective actions and implement controls to prevent a reoccurrence. A classic example of this process involves credit card fraud, where an unusual transaction on a cardholder's account can prompt a denial of the transaction or even a hold on the card.

LAW ENFORCEMENT:

Law enforcement uses data analysis to identify patterns of crime in order to allocate resources to the areas and populations that are most impacted by criminal activity which helps reduce crime and the cost of enforcement in those areas.

These are just a few examples of the application of analytics, and most likely you can think of others. By now we're certain you agree that analytics is pervasive in today's society.

Having established the vital role of analytics in today's world, we now turn our attention to an overview of the analytics methodology.

ANALYTICS METHODOLOGY

The analytics process involves various activities, tools, and techniques. The methodology for analytics exists within a framework that provides guidance as to the necessary inputs to the process. In Chapter 14, after you have gained some practical experience, we revisit the analytics process in the context of the decision-making cycle.

Figure 1-5 displays the framework for the analytics methodology. The cycle in the center is the analytics methodology or lifecycle. The surrounding boxes constitute the framework within which the methodology functions. The three main areas of the framework are enablers, benefits, and people.

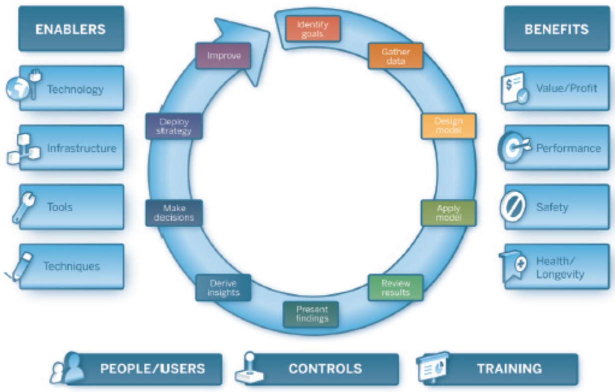

Figure 1-5: Analytics Methodology within a Framework

Enablers are the essential components needed for the methodology to work. They include technology, infrastructure, tools, and techniques. The benefits of analytics are vast and varied. Examples are value/profit, performance, safety, health and longevity of the system, and many others.

People are generally both the creators and the benefactors of analytics activities. User authorizations and internal controls, and training are required within the framework to ensure that analytics is used optimally and securely.

The center of Figure 1-5 displays the 10 key steps in the analytics methodology or lifecycle. These steps can have many sub-steps. We address the sub-steps in more detail in later chapters. We briefly discuss each of the steps below.

1. **Identify goals:** The first step is to define the goals and outcomes of the analytics process. They can be in the form of quantitative measures, business questions, or qualitative descriptions. Examples of outcome goals are: How can I minimize returns from sales deliveries? Which customer brings us the highest profits? Which factors contribute to product misuse?

2. **Gather data:** Data gathering is the next step. In the social sciences, data are often gathered from surveys. In the physical sciences, they are gathered through experiments, simulations, and sensors. In the business world, they are gathered from information systems that record all business activity within a corporation. Chapter 2 describes data acquisition in detail.

3. **Design model:** In this step, analysts determine how best to analyze the data they have acquired. One simple analysis involves dividing the data into more defined segments to better understand what they are informing. The process —called *slicing and dicing*—is covered in detail in Chapter 5

   People who require more insightful analysis than simple slicing and dicing employ data mining models. Data scientists have created a number of powerful models. In most cases an existing data model can be used. If the data or the analysis goal are unique, however, then it might be necessary to design a new model. Doing so requires strong mathematical and analytical skills.

4. **Apply model:** After a model is chosen, it is applied to the dataset. Many easy-to-use tools are available to accomplish this task. These tools frequently provide a drag-and-drop interface that enables users to choose data sources and models simply by dropping them onto the canvas. Users can then configure the options for the chosen model. After the model is run against the dataset, the results are presented to the user, often in the form of a visualization such as a chart.

5. **Review results:** Results are reviewed and checked against test data. Any deviations should fall within acceptable model parameters. If they do not, then the model is redesigned or run again with new data until the desired accuracy is reached.

6. **Present findings:** Findings from the analysis process can be presented in various ways. Displaying the results as columns and rows of numbers is sufficient if the results are not extensive and are easily understood by the viewer. Visualizations are used to display patterns and trends in large datasets. Interactive dashboards bring together multiple visualizations along with end-user controls. Finally, infographics utilize data and images in a story-telling narrative to report the results of the analysis. We discuss all of these visualization techniques in later chapters.

7. **Derive insights:** Insights are drawn using experience and domain knowledge. This is why analytics sits at the intersection of technology, domain expertise, and statistics. Domain knowledge is key to generating insight. End-users obtain insight when they understand or interpret the reported analysis.

8. **Make decision:** The insights lead to decision making. Theoretically, sound analysis complemented by domain knowledge; that is an understanding of the area of study, and analytical skills will lead decision makers to the optimal decision choice.

9. **Deploy strategy:** Decisions are then converted to strategies that are deployed as actions.

10. **Improve:** The outcomes of the actions are measured periodically to assess how well they have met the desired goals. The information obtained from the comparison of goals and actual outcomes is used to improve the process in the future. It also serves as an input at the beginning of the next cycle. After the *improve* step has been completed, we go back to step 1 of the cycle and start the process again.

Now that we have explained what data analytics is and why it is important and have discussed an analytics methodology framework, we introduce you to Global Bikes Inc. and some of its employees, who will appear throughout the remaining chapters. After you become familiar with the material in this chapter you will be ready to study practical analytics.