1

Introduction to Analytics

Philip T. Keenan, Jonathan H. Owen, and Kathryn Schumacher

General Motors, Global Research & Development, Warren, MI, USA

1.1 Introduction

We all want to make a difference. We all want our work to enrich the world. As analytics professionals, we are fortunate—this is our time! We live in a world of pervasive data and ubiquitous, powerful computation. This convergence has inspired new applications and accelerated the development of novel analytic techniques and tools, while breathing new life into decades-old approaches that were previously too data- or computation-intensive to be of practical value. The potential for analytics to have an impact has been a call to action for organizations of all types and sizes. Companies are creating new C-level positions and departments to grow analytic capability. A torrent of new start-ups have formed to sell analytics products and services. Even governments have created new high-profile offices to leverage analytics. These changes have driven a surge in demand for analytics professionals, and universities are creating departments, curricula, and new program offerings to fill the gap.

But what exactly do we mean when we say “analytics”? The term is widely used, but has vastly different meanings to different people and communities. A number of well-established disciplines, including statistics, operations research, economics, computer science, industrial engineering, and mathematics, have some claim to “analytics” and interpret it to have specialized meaning within their domains. The popular usage of the term is often comingled with other widely used but equally overloaded terms such as “big data,” “data science,” “machine learning,” “artificial intelligence,” and “cognitive computing.” As a result, this seemingly innocuous term has led to much confusion over the last decade as people using the same language often talk right past each other. In the authors’ own experience, frustration at all levels of an organization is inevitable when well-intentioned and intelligent people believe they have a shared
understanding—on a new project initiative, for example—only to discover weeks or months later that there was a fundamental misunderstanding of what work was to be performed or insights delivered.

In a 2016 article intended to reduce some of this confusion, Robert Rose identified three main usages of the term “analytics” [1]:

1) As a synonym for metrics or summary statistics
2) As a synonym for “data science” (another overloaded term)
3) As a very general term to represent a quantitative approach to organizational decision-making

Our use of the term is closest to the last of these; we consider analytics broadly as a process by which a team of people helps an organization make better decisions (the objective) through the analysis of data (the activity). This chapter gives a brief, high-level introduction to the subject. We first describe a conceptual framework for analytics, and define three primary categories of analytics (descriptive, predictive, and prescriptive). We then discuss considerations for applying analytics within an organization, and briefly discuss the ethical implications of using analytics. Subsequent chapters dive more deeply into each component of the process of applying analytics, including developing a request for a new project, building a cross-functional team, collecting data, analyzing data with a wide variety of mathematical and statistical methods, and communicating results back to the client.

**Interview with Alan Taber**

*Alan Taber, System Engineer with Lockheed Martin Missiles and Fire Control, defines analytics in the following way:*

Analytics is both a mindset and a process. The mindset is that instead of simply reacting to what you perceive your environment to be that you gather data understanding the limits and bounds of that data. You feed it into a model. It can be a very detailed model or a simple model about how situations evolve over time if you do take options A or B or C, or some combination thereof, and then you test that hypothesis. You have the continual feedback loop to say if what you’re doing makes sense and also keep an eye on your surroundings because what may have made sense a year ago or a month ago may no longer make sense. That’s the mindset, to always be paying attention rather than running on autopilot.

The process is to make sure you understand the root problem, figure out if you can frame that as a problem that’s amenable to being solved with data, figure out your data sources, and don’t limit yourself to the data you have on hand and know how to collect. If you need a different data set, go get it. Once you have your data and can run your test, do that. Over and under and around all that, you’re working with your stakeholders so that when you deploy people are
familiar enough with what you’re doing that they’re willing to try it out rather than saying, “I don’t understand the model and therefore I’m busy, I don’t have time to learn, I’m not interested.” If you are overwhelming people with information but not helping them actually solve the problems that they perceive they have, you simply will not get very far. You will have wasted all your time. So that’s the mindset and that’s the process.

This is an excerpt from one of a series of interviews with analytics professionals and educators commissioned by the INFORMS Analytics Body of Knowledge Committee.

1.2 Conceptual Framework

As shown in Figure 1.1, the generic analytics process can be viewed as a continuous cycle where the analysis of data produces insights that inform better decision-making. We use this simple figure to highlight two fundamentally different approaches to analytics: data-centric and decision-centric.

1.2.1 Data-Centric Analytics

The philosophy behind data-centric analysis is to “let the data speak freely.” Working under this philosophy generally involves pulling together as much relevant data as possible, analyzing that data to identify patterns that lead to insight, and serving up those insights to a decision-maker who (hopefully) will make better informed decisions. As shown in Figure 1.2, this follows the natural (clockwise) flow of the analytics process.

Not surprisingly, the data-centric approach has gained popularity with the surge in “big data.” Many of the analytic methodologies employed in this arena—including data mining and classification, machine learning, and artificial intelligence—increase in effectiveness with the volume of data available for analysis. Advocates believe that we are in a new “machine age” that is changing the landscape of business and the world [2–4]. Some argue that the data-centric “big data” paradigm is really about eliminating sampling error; they claim that we are
no longer reliant on small samples since we have storage capacity to hold and computing power to process vast amounts of data [5]. Others have observed that the promised insights have not always materialized, and that the challenge is “to solve new problems and gain new answers—without making the same old statistical mistakes on a grander scale” [6].

1.2.2 Decision-Centric Analytics

*Decision-centric* analytics begins with an understanding of the *decision* that needs to be made and what *insights* would lead to better expected outcomes. Decision-centric models typically encapsulate subject matter expertise (SME) and codify domain knowledge in order to relate decision variables to the target objective. Data requirements are determined by the chosen analytical model; ideally these data already exist in a convenient form, but often they must be extracted from disparate sources or collected through new instrumentation or market research. As summarized in Figure 1.3, this approach starts with the final outcome—the decision—and works backward (counterclockwise) at each step to define and develop needed analysis and data resources.
Decisions are often defined as an “irrevocable allocation of resources” [7]. Improving decision-making requires an understanding of the desired outcome (the objective), alternative actions (decision variables), and boundary conditions (constraints), but also the richer context of possible future conditions (scenarios). It also requires that we answer several softer questions: Who is making the decision? What is her or his scope of control and influence? What information is already available to the decision-maker(s) and where are the gaps? In a decision-centric approach, many of these questions are considered as part of upfront framing activities that look ahead toward operational implementation.

1.2.3 Combining Data- and Decision-Centric Approaches

Analytic practitioners and professional communities are often predisposed to either data-centric or decision-centric approaches. In the authors’ view, this is attributable to different pedagogical perspectives and experiences. Given the centrality of computing and information technologies for handling large amounts of data, it is not surprising that many organizational IT functions are naturally aligned with a data-centric view. Business operations and the analytic teams that support them often have a natural affinity for decision-centric approaches that leverage their deep understanding of key problems and models that support improvements. Table 1.1 summarizes salient features of the two approaches.

Important opportunity arises from combining elements of the two approaches. There is undeniable potential to leverage increasingly pervasive data and computational power associated with data-centric analysis, but contextual knowledge and subject matter expertise provide needed guardrails so that the resulting insights are meaningful.

Acknowledging the natural tendencies of individuals or analytics organizations toward data- or decision-centric approaches may help practitioners to identify growth opportunities. For example, traditionally decision-centric organizations may benefit by expanding the amount of data used in their analyses, including unstructured data sources. Typically, data-centric groups may improve the fit and predictive power of their models by incorporating domain-specific expertise.

Evidence of the benefit of utilizing a combined approach is seen in recent movements to incorporate “thick data” into marketing analytics (see Refs [8,9], for example). Combining thick data, such as ethnographic studies or focus group responses (see Figure 1.5), with big data, such as transaction data, enables a more complete understanding of customers’ preferences and behaviors. Decision-centric framing, domain knowledge, and deep subject matter expertise collectively provide scaffolding that helps big data insights take shape.
### 1.3 Categories of Analytics

A well-known and useful classification scheme for analytics was proposed by Lustig et al., at IBM [10]. Based on their experience with a variety of companies across a diverse set of industries, they defined three broad categories of analytics:

**Table 1.1** Comparison of data-centric and decision-centric approaches.

<table>
<thead>
<tr>
<th></th>
<th>Data-centric analysis</th>
<th>Decision-centric analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mantra</strong></td>
<td>“Start with the data”</td>
<td>“Start with the decision”</td>
</tr>
<tr>
<td><strong>Philosophy</strong></td>
<td>Leverage large amounts of data. Let the data “speak freely” by identifying patterns and revealing implicit (hidden) factor relationships</td>
<td>Leverage domain knowledge and subject matter expertise to model explicit variable relationships</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>More is better, especially for “big data” applications (e.g., speech or image recognition)</td>
<td>Custom collection of curated data sets</td>
</tr>
<tr>
<td><strong>Computing</strong></td>
<td>High-performance computing is often price of entry. Potential need for specialized processors (e.g., GPUs, TPUs) for acceptable execution speeds, especially in contexts requiring real-time analysis</td>
<td>Desktop or server-based computing is typical. Trade-offs between potential benefits of leveraging high-performance computing versus added overhead in development and maintenance</td>
</tr>
</tbody>
</table>
| **Pros**             | • Increasingly automatable  
• Potential to extract weak signals from large, unstructured data sets                                                                                                                                                    | • Causal focus  
• Strategic value beyond historical observations                                                                                                                                                                                        |
| **Cons**             | • Risk of conflating correlation with causation  
• Analysis inferences are limited by history  
• Noisy data with confounded effects                                                                                                                                                    | • Human subject matter expertise required  
• Cost of data acquisition can be high                                                                                                                                                                                                          |
| **Key disciplines**  | • Computer science  
• Data science  
• Machine learning and unstructured data mining  
• Artificial intelligence (AI), deep learning                                                                                                                                          | • Management and decision sciences  
• Operations research  
• Mathematics  
• Classical statistics                                                                                                                                                                                                                     |
| **Example applications** | • Image classification  
• Speech recognition  
• Autonomous vehicle scene recognition                                                                                                                                                                                      | • Supply chain optimization  
• Scenario planning  
• New business model development                                                                                                                                                                                                          |
descriptive, predictive, and prescriptive. As summarized in Figure 1.4, there is a natural progression in the level of insight provided—and potential value—as an organization moves from descriptive to predictive and ultimately to prescriptive analytics. Typically there is also a progression in the mathematical sophistication of the analysis techniques, as well as the organizational maturity required to absorb and act on resulting insights.

1.3.1 Descriptive Analytics

The purpose of descriptive analytics is to reveal and summarize facts about what has happened in the past or, in the case of real-time analysis, what is happening in the present. This is done by examining and synthesizing data collected from a variety of sources. Raw data are captured and recorded in source systems, eventually to be cleaned, retrieved, and normalized such that entities and relationships can be meaningfully understood. The audience for descriptive analytics is broad, potentially reaching all functions and levels of an organization. Descriptive analytics are at the heart of most business intelligence (BI) systems.

Data Modeling

Many organizations have access to vast quantities of data. Useful descriptive analytics generally involves processing the raw facts into higher level abstractions. Data scientists think in terms of entities and relationships. For example, a customer database might contain entities like “Household” and “Product,” linked by relationships like “Purchased,” with data elements...
including the demographics of the households and the price, cost and features of the products.

Sources of data can be highly varied (see Table 1.2 for examples), as can the size and information density of any given data set (see Figure 1.5). There is also high variability in the expense and effort required to collect different types of data. On one end of the spectrum, ethnographic studies require social scientists to spend many hours shopping with or interviewing individual customers, and thus the data are very carefully curated and very expensive to collect. On the other end of the spectrum, “data exhaust” is logged nearly for free, including data generated from smartphones and online activity [11]. Data exhaust is collected without a specific intended purpose and can be especially messy, so substantial cleanup effort is usually necessary before this type of data are usable.

Developing a data model that captures the structure and relationships among the different data elements is a fundamental task. Generic data models are often constructed to efficiently store ingested data, without specific analytic use cases in mind. Although such data models can be useful for general-purpose reporting and data exploration, purpose-built data models are typically needed for efficient

<table>
<thead>
<tr>
<th>Source</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transaction data</td>
<td>Data associated with a transactional event. Example: a purchase transaction with details of the specific item purchased, where and when it was purchased, the price paid and any discounts applied, how the customer paid (e.g., cash, credit card, finance), and other contextually relevant data (e.g., inventory of other items for sale at the same time and location)</td>
</tr>
<tr>
<td>Customer data</td>
<td>Data associated with customers. Examples: detailed demographic or psychographic information on individuals and households, history of interactions (past purchases, Web site visits, customer service requests)</td>
</tr>
<tr>
<td>Sensor data</td>
<td>Data collected through electronic or mechanical instrumentation. Examples: web browser cookies tracking customer activity, electronic sensors monitoring weather conditions, airplane flight data recorder information</td>
</tr>
<tr>
<td>Public data</td>
<td>Open-source data from individuals, organizations, and governments. Example: aggregated census data</td>
</tr>
<tr>
<td>Unstructured</td>
<td>Data without known structure. Examples: text and images from social media, call center recordings, qualitative data from focus groups or ethnographic studies</td>
</tr>
<tr>
<td>Curated data</td>
<td>Data collected for a specific purpose with downstream analysis in mind. Examples: consumer surveys, designed market research experiments</td>
</tr>
</tbody>
</table>
Ethnographic studies
Observe individual behaviors in their typical environments and activities. Provides in-depth, dense information from an extremely small sample. Example: customer needs identification for future product design.

Focus groups
Information gathered through facilitated interactions with small, co-located groups (e.g., gatherings of 5–10 current or potential customers). Example: feedback on potential product design alternatives and feature sets.

Custom market research
Information gathered through specialized, purpose-driven market research. Example: clinic events or customized online surveys to identify attribute preferences and willingness-to-pay for a new product launch.

Syndicated studies
Information gathered through broad market research studies. Example: industry studies on brand perceptions across a broad spectrum of the population.

Primary / business-related data
Business-relevant information, likely gathered through routine data collection. Example: transaction and customer data.

Secondary data and “data exhaust”
All other available data that is not core to the business but can potentially be mined for insight, usually in combination with smaller data sets.

Figure 1.5 Illustration of variability in the size and information density of different data sets.
analysis. Depending on the size of the organization and the speed with which new data arrives, substantial IT support may be required to run systems that capture and record data, clean it, and store it in a warehouse or lake for eventual retrieval and analysis.

**Reporting**
The real value of descriptive analytics comes from putting access to this plethora of data into the hands of analysts who can use it to rapidly answer questions. To this end, Lustig et al. proposed a classification of descriptive analytics into three areas [10]:

1) Standard reporting and dashboards
2) Ad-hoc reporting
3) Analysis/query/drill-down

In our experience, standard reporting and dashboards are useful to a point, but users need to be able to “slice and dice” the data on the fly to gain more meaningful insights, computing summary statistics and visualizing comparisons without being limited to predefined reports.

**Visualization**
Descriptive analytics is often about communication, not math. Authors such as Tufte [12] provide useful guidelines for describing and visualizing data in ways that reduce the cognitive burden on those who must interpret the results. Later chapters will go into more depth on this; however, since the topic is so important, we will elaborate on it later in this chapter (Section 1.4.2) as well when we discuss the communication of project insights.

**Software**
Software for descriptive analytics is plentiful. At the most basic level are ordinary spreadsheets and databases. At the other end of the spectrum are systems designed specifically to support data visualization, exploration, and reporting—such as Cognos, Tableau, and Spotfire. These systems can greatly increase the accessibility of data and basic analytic insights throughout an organization.

### 1.3.2 Predictive Analytics

Descriptive analytics describe the world as it is (or as it recently was). In contrast, *predictive analytics* seek to forecast the likely future state of the world through a deeper understanding of the relationships among data inputs and outcomes. This is a much more demanding goal, so there is much more that can go wrong. Inexperienced analysts and leaders often imagine that once you have a good descriptive model, you can use it to make good forecasts. Not true! Statisticians have long understood that correlation does not imply causation. As a result,
teams that wish to forecast the future need to use more sophisticated modeling approaches and follow more rigorous validation procedures if they want to have confidence that their forecasts make sense.

As a very simple example of the difference between descriptive and predictive analytics, consider television programs that cover the stock market. Every day, talking heads explain why the stock market behaved the way it did the previous day. But can any of them accurately forecast what the market will do tomorrow? Not a one. If they could, they would be billionaires living on a beach, not reading off a teleprompter in a TV studio. Hindsight may be 20–20, but foresight certainly is not.

Data Mining and Pattern Recognition
The starting point for predictive analytics is often mining data to identify meaningful relationships and patterns. As we work with increasingly large and diverse data sets, there is a growing opportunity to identify hidden relationships that relate disparate data. For example, clustering analysis might be used to segment customer populations into groups that go beyond simple demographic or psychographic characteristics. Or we might apply various machine learning techniques to identify objects and trajectories for autonomous vehicle scene recognition and navigation.

The set of available data mining techniques is highly varied, and practitioners need to be adept at selecting appropriate methods based on an understanding of the pros and cons of each within a given application context. Many methods are based on classical statistical models, often to classify populations into distinct groups (e.g., classification and regression trees) or to estimate the impact of a set of descriptor variables on a metric of interest (regression). Machine learning and artificial intelligence techniques can arguably answer a broader set of questions (e.g., image recognition), but trade the transparent simplicity of classical models for a harder-to-explain “black box” capable of representing more complex relationships. Regardless of the methodology, analysts must be alert to the danger of false positives. Given enough computer time and input data, one can always find some sort of “statistically significant” effect that is actually pure noise.¹

Predictive Modeling, Simulation, and Forecasting
Predicting the future requires a model. Simply collecting and reporting data, or identifying interesting patterns about the past and present is not sufficient.

One of the simplest models assumes that the future will behave like the past; for obvious reasons, this is often referred to as a naive model. For an established company, sales next month will likely be similar to sales last month. However, leaders who request analytics projects generally want deeper insights than that!

¹ The reader is encouraged to see https://imgs.xkcd.com/comics/significant.png for a lighthearted cartoon illustrating the dangers of false significance.
The next simplest model is trend extrapolation. If sales were 100 units in January, 110 in February, and 120 in March, it seems plausible to predict that they will be 130 in April and 140 in May. Projecting simple trends can be useful, but it is not always appropriate. Suppose you are selling tax preparation software; this forecast would be inaccurate, as sales in May will instead be close to zero, since most customers will have filed their taxes with the IRS by April 15. In this context, a more advanced model that “seasonally adjusts” the data would be appropriate.

More sophisticated models often include other explanatory variables in addition to time. For example, when trying to predict the number of vehicles the US automotive industry will sell next year, it is often helpful to consider macroeconomic data such as the unemployment rate, interest rates, and inflation. The automotive industry is cyclical—sales fall during recessions and rise during periods of economic expansion. Predicting the timing of the next recession can be almost as challenging as predicting the future course of the stock market. As a result, predictive models generally need to report ranges, or uncertainty bounds, rather than simple point forecasts. Unfortunately, many clients have difficulty consuming range estimates and prefer to pretend that point forecasts suffice. This is one of the many challenges the analytics practitioner faces when trying to communicate results in a form accessible to decision-makers.

Deciding what variables to include in a model can also be challenging. Leave out an important causal factor and the model’s predictions may be seriously wrong. Including extraneous factors can also cause difficulties. For instance, classical regression models can fail if several input variables are closely correlated, an issue known as multicollinearity.

Analysts often attempt to assess the goodness of fit of their proposed model. For example, when fitting a regression model, most software packages report the “R-squared” metric, a measure of how closely the model matches the data. Analysts often construct a variety of models (perhaps using different subsets of variables in each) and pick the one with the highest R-squared. Unfortunately, this technique of “chasing R-squared” is not, in fact, a good approach—it can easily lead to overfitting, which in turn can lead to poor performance when predicting future values.

To avoid this pitfall, analysts can instead divide the data into a “training sample” used for fitting the model, and a “validation sample” used for assessing and comparing models after they have been fitted. Executed properly, this methodology can dramatically reduce the risk of overfitting, so it should be standard operating policy for all analysts whenever sufficient data are available.

Leveraging Expertise

There are a great many methodologies available for building predictive models. Frequentist statistical models have been used for over a century. Bayesian
statistical models became widely used starting around 1995, when faster computers and algorithms made them computationally practical. Machine learning methods have become popular in recent decades, made possible by faster computers and larger data sets. Statistical and machine learning methods work well for analyzing a vast array of situations, but they tend to rely on the computer to discover patterns in the historical data and assume these patterns will repeat in the future. However, sometimes the future is different from the past. For example, when launching a new product, historical sales data are not available. How then to predict future sales?

Potential solutions have been developed for such cases, but they are substantially more complicated and time consuming (i.e., expensive) than methods that make use of existing data. For example, when launching a new product, one such approach is to perform primary market research to test how potential customers react to the new product.

In some situations, a practitioner has abundant knowledge of the structure of the real world, and incorporating that knowledge into the model building process can be extremely valuable. Simulation models are particularly useful in such situations. Simulation is based on the understanding of how some entities—individuals, components, or other actors—behave in isolation, and how their interactions lead to consequences under different scenarios. Simulation techniques can be classified based on what interacts and how the interactions occur. Table 1.3 summarizes key differences between three common types of simulation models: discrete event, agent-based, and system dynamics.

<table>
<thead>
<tr>
<th>Table 1.3</th>
<th>Comparative summary of three common simulation models.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discrete event simulation</strong></td>
<td>Models a system using a central global mechanism, often a network, within which entities interact according to centrally specified rules at discrete points in time (events). Interactions are defined by standardized structures such as queues. Example: call center and discrete manufacturing operations analysis</td>
</tr>
<tr>
<td><strong>Agent-based simulation</strong></td>
<td>Models a system using autonomous agents (representing both individuals and collective groups), each with their own rules for behavior. Interactions are determined by domain-specific rules potentially based on the state of the agents involved and the overall state of the system. The overall system behavior emerges from the interactions of the agents. Example: flight simulation for a flock of birds</td>
</tr>
<tr>
<td><strong>System dynamics</strong></td>
<td>Models a system using stocks and flows. Interactions are defined by feedback loops and control policies. System dynamics is to agent-based simulation as thermodynamics is to molecular simulation, in that it aims to reduce the computational and cognitive burden through aggregation. Example: Bass diffusion model of the impact of advertising</td>
</tr>
</tbody>
</table>
Simulation models require a lot of effort to calibrate to observed history. However, because they model the underlying “physics” (e.g., microeconomics) of the situation, they can incorporate additional data from subject matter experts or market research. Simulation models can be used to evaluate “what-if” scenarios, a capability that is very useful to decision-makers, and is not possible with basic forecasting models.

1.3.3 Prescriptive Analytics

Prescriptive analytics seek to go further than forecasting a future state, to make actionable recommendations about what the decision-maker should do to achieve a particular objective, such as maximize profit. With descriptive and predictive analytics, the analytics team shoulders most of the burden of interpreting the results and developing recommendations for action. With prescriptive analytics, the computer helps with that process by evaluating a large number of potential alternative courses of action and reporting the best ones. The team still needs to apply a level of business judgment in interpreting the answers, since all models are incomplete descriptions of reality. Nonetheless, this sort of analytics has the greatest potential to help decision-makers realize tangible benefits through better decision-making.

However, automating the process of generating actionable recommendations requires a higher standard for defining causal relationships. Consider the following hypothetical example. Suppose you develop a time series model that attempts to forecast US automotive sales using imports of cheese from Mexico as the explanatory variable. You may find that the model fits the data well (it is descriptive). You may well also find that the prediction it makes (more cheese imports correlates with more vehicle sales) also turns out to be accurate year after year into the future (it is predictive). Nevertheless, if you were to then make the prescriptive recommendation that auto manufacturers should lobby Congress to reduce tariffs on Mexican cheese in order to stimulate car sales in the United States, you would be making a very foolish error. The relationship is spurious. There is no causal connection, so reducing tariffs would have no actual effect on vehicle sales. Instead, both cheese sales and vehicle sales are correlated with overall gross domestic product (GDP): when people have more money to spend, they use it for cheese and for cars; when they have less, they defer both kinds of purchases.

The lesson of the tale is clear: you need to first understand how the real-world business situation works, and model it appropriately. One huge risk of “big data” is that analysts will simply throw a huge quantity of data at a machine learning system with no thought about what kinds of relationships are plausible. In some settings this is not an issue (think “people who shopped for X also shopped for Y” recommendation engines). But in other settings, recommending nonsensible actions may destroy credibility.
No one knows the future. What we can hope to achieve with prescriptive analytics is simply to help decision-makers make the best decision possible, given the best data available at the time.

Prescriptive analytics typically require a combination of simulation and optimization. You begin by determining what quantity you wish to maximize—for example, the net present value of operating your business. Next, you list the decision levers available to you, such as investments in advertising, new product development, or price cuts for existing products. Next, you build and calibrate a model that is robust under a wide variety of ways of pulling the levers. This may require something like a system dynamics model, since it may need to capture scenarios in which the future does not look like a simple trend extrapolation of the past. Finally, you embed the simulator inside an optimization loop that evaluates a large number of different ways of setting the decision levers and tells you which one maximizes your objective, for example, is most profitable. The optimizer frequently needs to deal with various sorts of constraints, for instance, some decision levers are discrete, others are continuous, and some economic variables, like price and sales volume, cannot be negative.

Prescriptive models must also consider how entities outside of your control (e.g., competitors) will behave or react to your decisions. These may be “random,” as in Monte Carlo simulation, or “strategic,” as in Game Theory. Real life generally includes both.

For a real-life example, consider “Modeling General Motors and the North American Automobile Market” [13]. The client was the then-President of GM North America. The goal was to maximize future profitability. The team developed a system dynamics simulation model combining internal activities such as engineering, manufacturing, and marketing with external factors such as the competition for consumer purchases in the new and used vehicle marketplaces. Eight groups of automotive manufacturers competed for a decade across 18 vehicle segments, making monthly segment-by-segment decisions about price, volume, and investment in future products. The model included Monte Carlo simulation of random effects, such as how attractive future competitor vehicles turned out to be once they entered the marketplace, and when the next recession would occur. This was then embedded inside an optimization loop that evaluated alternative strategies. Instead of point forecasts, it generated probability distributions on future profitability, as illustrated in Figure 1.6. Ultimately it was able to show that despite future uncertainty, following a particular proposed strategy (B) would produce a probability density shifted to the right (i.e., toward higher profits) as compared to following an initial strategy (A). This supported a prescriptive recommendation to enact strategy B.

Just as with descriptive and predictive models, prescriptive models require substantial amounts of business judgment and work best when the team iterates between analyzing scenarios and discussing them with subject matter experts. No computer model is perfect. The data may contain valuable information, but
inevitably you will get better results if you also incorporate subject matter expertise. At a minimum, this expertise is necessary for qualitatively interpreting the results, and when possible can also be quantitatively incorporated into the model itself.

1.4 Analytics Within Organizations

Suppose you have decided you want to do analytics within your organization. How do you get started?

Until recently, in many large organizations this involved a lot of pushing. Analytically minded employees would see an opportunity, perhaps even build a prototype analysis tool for a particular business challenge, show it to management, and then often watch it die a quiet death at the hands of leaders who did not understand the potential benefits of analytics, or who felt threatened by the thought of being replaced by a computer program.

In the last decade, however, things have changed dramatically. Analytics has become a senior management buzzword and a prominent topic of articles in publications like *Harvard Business Review* and the *McKinsey Quarterly*. These days, it is no longer a question of you, an individual employee, wanting to get more involved. Now the question is: “Your organization has decided it needs to do more analytics. How does it get started?”

The answer is of course unique to each organization, but we will make some general comments, first about the life cycle of an individual analytics project, and then about the alternative ways an organization can implement such projects.
1.4.1 Projects

Analytics projects work best when you have three key ingredients: (1) quantitative analytics professionals who are well-versed in the data and appropriate analytic techniques, collaborating closely with (2) subject matter experts who understand the problem domain, and (3) leadership sponsors in the core business who understand the value of better data-driven decisions and will champion implementation in the organization.

A new analytics project typically begins with a conversation between executives, one with operating responsibility for a difficult business decision and the other with experience doing analytics projects. If they are able to communicate effectively, they will be able to jointly write a framing document: a statement of the problem to be solved that also describes the scope, outputs to be delivered, and a high-level description of the kinds of input data and analytical frameworks that will likely be helpful in creating the desired outputs. The framing document should also include a list of stakeholders whose engagement will be needed to see their project through to implementation.

Next comes a stage we call “invent and pilot.” This is a highly iterative process. The stakeholders assemble a cross-functional team combining analytical experts with business experts. The team gets up to speed on the business problem, obtains samples of available data, tries a variety of methods for analyzing it, discusses the results of each, and eventually settles on an approach that is feasible to execute within the time and resource constraints of the project while also delivering results that make actual business sense to the end clients.

Next comes “productionization.” In a small organization, this could be as simple as providing the client with a spreadsheet. In a large organization, this may be a much longer and more expensive process involving the internal IT organization. Typically IT support is necessary to automate the data feed into the analytical environment, and to provide data security for both the inputs and the results of the analysis. Ideally, IT also provides services such as data cleaning, although often this is beyond their scope and falls to the analytics team instead. This can be a huge undertaking, since a great many real-world data sets have missing values, incorrect values, and are inconsistent with other data sets that are needed for the same project.

IT may also choose to develop some sort of delivery platform, such as a custom app or Web site, in order to simplify the user experience for end client users and to help maintain control of the data for security purposes.

Finally, IT deploys the solution to the client. Typically the analysis team continues to play a major role for the first year or so, conducting ongoing analysis and presenting it to leadership, as well as training people in the client organization to use the system. Often a change management process is required, since the new analytics based method of making decisions may involve a very different process than the one people in the organization are familiar with. It is
Invention
- Modeling, algorithm development, and validation
- Data identification and proof-of-concept integration
- Prototype tool development

Pilot testing
- Testing and validation
- Cycles of learning with modifications as needed

Business use
- Integrate with ongoing operations
- Implement training and support
- Establish and share best practices
- Continuous improvement

Productionize tools and data flows

Training, security, help / support and maintain

Figure 1.7 Life cycle of analytics projects.
best if some members of the client organization were participants in the cross-functional analytics team from the beginning, but at a minimum, some members of the client team must be trained as “superusers”—people who can load data, run the model, and present and interpret results, all without requiring much support from the analytics experts who built the system initially.

Additional activities (e.g., training, security, help/support) are often needed to sustain an analytics capability over time and support ongoing business use. As users become more sophisticated with experience and grow in their ability to leverage insights, new questions arise that require model enhancements. The complete life cycle of a typical analytics project is summarized in Figure 1.7.

<table>
<thead>
<tr>
<th>INTERVIEW WITH ERIC STEPHENS</th>
</tr>
</thead>
</table>

When asked to identify the key skills needed to obtain the problem definition/problem statement, Eric Stephens, Manager of Population Health Analytics at the Vanderbilt University Medical Center, responded as follows:

These aren’t necessarily going to be in any particular order, but first and foremost I think is communication. This means the ability to listen, as well as to speak and write. In fact, listening is probably even more critical in this context than it may be in others because the ability to listen—and to comprehend and understand the situation—is extremely critical to framing the problem properly.

Although it is typically not something an analytics practitioner can influence, the culture of the organization can have a significant effect on the ability to properly define the problem. In my previous organization, there were many cases where I worked very closely with the president. He would frequently call and ask, “I need this data for this time period” or “I need to see this and this,” and that’s all the information he would consider. This is problematic because there may be parameters, circumstances, or other attributes that aren’t stated that could significantly impact the output or the result. I would always have to push back on him a little bit to say, “OK, can we step back just a moment and can you give me a little bit more information about the problem you’re trying to solve? What is it you’re trying to accomplish? What’s the overall objective?” Toward the end of my tenure there things got a little better, but I remember when I tried to initiate this type of conversation early on, it was usually met with something like, “it doesn’t really matter,” “you don’t need to know,” “it’s not important right now,” or “I don’t have time to go into it.” My effort was to try to communicate with him in order to better understand from his perspective what he was trying to accomplish. In situations like this, it’s incumbent upon the analytics professional to convey that he or she is simply trying to provide the executive with the most appropriate solution for their problem.

The communication element is important in terms of being able to
really listen and understand what the situation is; this includes the ability to empathize with the other person. From an analytic standpoint, this means being able to understand what the other person’s overall situation is. For example, they may be under a lot of pressure from the president of the organization. Let’s say that they’re a VP or someone who reports to the senior executive team. Their sales may be significantly down, and they’re trying to understand why so that they can either reorganize their product selection, or hire new salespeople, or whatever the case may be. That person may be thinking such things as “what could this mean in terms of my employment?” or “what impact would this decision have on the overall organization?” Being able to put yourself in another person’s shoes really gives a lot of perspective into what the overall problem is and how it could potentially be addressed with an analytic solution.

Another important skill is the ability to think at the level of the person who is presenting the problem. It goes along with empathy, but it’s really more concrete. In other words, if you are dealing with an executive, then the ability to think from the executive’s perspective in terms of the business implications of the decision is important. It’s not just a problem that you throw some data at and you build some models and that’s it. It is important to be able to think at a higher level: to comprehend and understand the business as executives do. Certainly, it doesn’t mean that every analytics professional needs to have an MBA in business strategy, but the more accomplished or the more adept the analytics professional is at thinking at that level, the more it opens up or exposes additional potential analytic solutions that may not necessarily have come to mind.

All else being equal, being able to communicate with empathy can make all the difference in how successful an analytics professional is in addressing business problems. Consider a situation in which you’ve got Analyst A, who is not able to think or converse at an executive level. They’re mired in the statistical minutia or spend most of their day thinking in computer language rather than in the language of business. This person may be incredibly skilled at developing technical solutions, but has difficulty communicating with those in the business who are requesting their assistance. Contrast that with Analyst B, who is also very adept at building models and at programming whatever tool necessary to do the work that they need to do, but at the same time can switch perspectives so that they can converse with the business owner or executive at their level. Oftentimes, what I see are analytics professionals who can’t bridge that gap, resulting in communication breakdowns at best, and a lack of trust at worst. When this occurs, the executive or businessperson asking the question may feel like the analyst lacks the understanding necessary to be able to deliver effectively. This is definitely not a recipe for analytics success.

This is an excerpt from one of a series of interviews with analytics professionals and educators commissioned by the INFORMS Analytics Body of Knowledge Committee.
1.4.2 Communicating Analytics

The best model in the world is of no value if the team is unable to persuade the decision-maker to act on the recommendation, so clear and transparent communication of recommendations and their rationale is essential. Writing good presentations takes effort. That effort is extremely important, even though it is completely irrelevant to the underlying mathematics. Analysts and executives frequently have very different perspectives and cognitive styles. Analysts are comfortable with mathematical formulae and inherently interested in computation, whereas executives are more focused on people, products, relationships, and results that impact business outcomes.

Junior analysts are prone to presentation pitfalls such as pasting a data table directly into a presentation (complete with six significant digits) and giving the slide a generic topic title like “Future Profit.” Executives look at the mass of numbers and wonder why the analyst is so naive as to believe they can actually distinguish between 10.5678 and 10.5679. Wondering if the analyst is equally naive about other, less obvious issues, the entire analysis is now suspect.

Unfortunately, even experienced analysts can get so caught up in the mathematically interesting details of their work that they neglect to take the time to properly frame their communication. A good presentation uses “sentence titles,” so that a reader who only reads the titles and does not look further into the slide can still follow the gist of the story. Good slides make their point clearly while also looking visually balanced and simple. This requires careful thought. Who is the audience? What is my goal for this meeting? What do I need to tell them to accomplish that goal? Business presentations are not mystery novels: they should lead with the answer and provide supporting details only in backup, for reference in the event they are needed. The analyst has to think about the important themes and illustrate them carefully. This usually means selecting a few key metrics and showing a relevant comparison, such as “benefit if you follow our recommendation versus benefit under the status quo plan.”

1.4.3 Organizational Capability

The sketch of the life cycle of an analytics project in Figure 1.7 highlights some important issues. One is that the analytics experts who build the initial prototype solution tend to be scarce commodities. Whether the organization maintains its own internal pool of analytics talent or hires external consultants for each project, either way these people are expensive and difficult to recruit and retain. That is why it is essential to train a group of “superusers” who can support and maintain the project after the initial stage, so that the analytics specialists can be reassigned to new projects.

This scarcity leaves organizations with two key questions: how to prioritize analytic initiatives, and whether to use internal or external talent.
Prioritizing opportunities should be based on the impact to the organization as a whole. For businesses, this generally means improving the net present value of future free cash flows, or in simpler words, prioritizing the opportunities with the biggest potential bang for the buck.

There are two main ways this can occur: push and pull. In the “push” version, someone—either a central analytics organization or a central planning function—attempts to model the key drivers of business performance and the available levers for influencing those drivers. Applying a sensitivity analysis to this model results in a prioritized list of opportunities for intervention that have the highest potential to improve profitability. The leader of the organization must then “push” this agenda by socializing it with the leaders of the prioritized functions, who may or may not be receptive to the idea that some outsider thinks they can run the area more efficiently or more profitably. However, depending on the culture of the company, some of these leaders will be intrigued by the possibility of improvement, and will champion the initial projects. If those succeed, other leaders will generally become interested as well.

Over the past decade, many organizations have switched from push to pull, as analytics has become more visible in the C-suite. In the “pull” version, the central analytics organization prioritizes requests as they come in from leaders around the business. This version generally works much better than pushing, because the leaders themselves initiate the project and are pulling for it to happen. Someone still needs to set priorities, however, so it is still valuable to model key performance drivers and have a means for estimating the potential impact of each new project. Generally speaking, a project with a billion dollar potential impact requires only modestly more analytics resources than a project with a million dollar impact, so prioritizing based on the estimated size of the impact can be very helpful.

The prioritization decision is closely linked with the question of using internal or external talent. There are pros and cons to both approaches. External consultants can get up to speed quickly, draw upon a deep experience base within their firm, and already have a base of talented analytics professionals available. However, they are expensive. Moreover, “consulting makes the consultant smarter”—unfortunately, the client rarely gets as much of that benefit. Far too often, consultant-based projects turn out to be difficult to productionize without essentially paying the consulting company forever, because only the consultants really understand the analytics process at a deep level. Moreover, despite internal firewalls within consulting companies that keep specific details of competing clients strategies private, once a consulting firm develops a methodology for solving a particular business problem with one client, they are likely to want to leverage that investment by applying the more “generic” elements of that methodology with other clients. Initially, those new clients may indeed be in different industries, but over time the knowledge often diffuses more broadly, with the risk of eventually benefiting competitors of the original client.
As a result, companies that view analytics as a competitive advantage generally prefer to hire their own permanent analytics staff. This strategy too has downsides however, since it may be difficult to attract and retain sufficiently qualified people. Moreover, sometimes internal groups become insular, cut off from the advances in other industries, whereas consultants in a large firm may benefit from seeing many applications across a variety of industries.

If an organization does hire its own analytics staff, where should they fit in the organizational structure? Some companies centralize them under a Chief Analytics Officer, others spread them among a variety of client organizations, and some use a mix of both approaches. Sometimes analytics is viewed as part of the IT function, other times it is separate. Not surprisingly, it is difficult to make one-size-fits-all recommendations—the right answer depends on the size and shape and culture of the organization. For example, if the IT function’s role and culture is primarily to manage infrastructure costs, they will probably not be a good fit for an analytics organization, which by nature is more like a small start-up or internal consulting company. In such cases, a centralized analytics group in conjunction with centers of expertise within client functions may be a good approach.

1.5 Ethical Implications

As analytics become increasingly pervasive, the ethical implications of collecting data and partially or fully automating decision-making become increasingly important. Analytics methods have the potential to provide tremendous value to individual companies and organizations, and to broader society. However, widespread collection of data raises privacy and security concerns. Additionally, broad adoption of algorithms to make decisions may have negative unintended consequences. Analytics professionals should be aware of these potential pitfalls and take actions to ensure that models are deployed in a responsible way.

In many countries, particularly in Europe, laws limit the kind of personal data companies are allowed to collect, store, and share, or provide consumers with the right to have their data erased. Even countries that allow collection of personal information often have laws mandating public notification if the data are inadvertently released or maliciously accessed by hackers. As a result, all organizations that analyze data must now stay informed about the potential legal implications of their data sets and take appropriate security measures to comply with applicable laws.

New technology for collecting data will raise new questions around “who owns data?” For example, who should have access to or be able to sell your Internet search history, your Fitbit health record, or your autonomous vehicle’s sensor data? Similarly, as organizations lean more heavily on automated analytics, who
will bear responsibility for errors, when for instance a driverless car is involved in a crash? There are many open questions that need to be resolved before the potential advantages of these technologies can be fully realized. In the coming decades, conversations regarding these topics will likely continue and will involve policy makers, lawyers, academics, politicians, and analytics professionals. Analytics professionals have a responsibility to honestly represent the capabilities and limitations of these technologies in these discussions, and to work toward solutions that serve the public good.

Algorithms have the potential to make decisions in ways that are more transparent and objective than a human decision-maker. For example, decades ago loan officers explicitly considered applicants’ race when deciding whether to approve their loan applications. Modern credit scoring algorithms explicitly do not consider race as a factor. While not perfect, these algorithms are less discriminatory. However, the predictions or recommendations that come out of a model can be perceived as being completely objective, when in reality they are subject to biases in the data or in the modeling decisions. For example, data collected from smartphone apps are not representative of the whole population, as avid smartphone users skew young, affluent, and urban. Distribution of public services based on smartphone data may potentially exclude individuals who are invisible in the digital data set [14]. Similarly, racial biases in crime data can lead to racial biases in crime predictions, such as those used in predictive policing models [15].

Widespread deployment of certain algorithms can also create self-reinforcing feedback loops. For example, in most states in the United States, auto insurance premiums are substantially higher for people with poor credit [16]. These people then face much higher expenses, increasing the likelihood that their credit remains poor. Cycles like these are not a new phenomenon, but as price discrimination algorithms become more prevalent and more precise, the implications of the cycles become more profound.

Most countries have anti-discrimination laws that forbid discrimination based on factors like race, religion, gender, nationality, disability, and so on. Well-intended modelers who explicitly omit variables representing these categories may still inadvertently discriminate by including variables that correlate with these categories. For example, recidivism models are used to predict the likelihood that criminal defendants will reoffend. While these models do not explicitly use race, they use variables that correlate with race, such as education level and employment history, and thus defendants of different races may receive different risk scores [17].

The potential hazards of using analytics vary widely with the specific application. Sentencing decisions, for example, are fundamentally more morally fraught than decisions regarding which ad to serve on a Web site. Nonetheless, all analytics professionals should be aware of these issues, and should consider the societal consequences of their work. Diakopoulos and Friedler [18] proposed the
following five principles that can guide accountability in the application of analytics:

1) **Responsibility**: Someone should have the authority and resources to deal with adverse consequences. Fully automated decision-making does not require a human in the loop, but a human should be involved to monitor the system and be able to make changes if needed.

2) **Explainability**: Decisions should be explainable to people affected by those decisions. Explaining the outcomes of machine learning models is especially difficult, but efforts are underway to develop interpretable machine learning methods, such as the DARPA Explainable Artificial Intelligence program [19]. In some applications, like speech recognition, explainability may be less important. But when used in contexts that have serious consequences for people’s lives, such as determining who should receive a loan or be released from prison, clear and accessible explanations are essential.

3) **Accuracy**: Sources of error and uncertainty should be identified, logged, and benchmarked. Any model can make inaccurate predictions or misleading recommendations if it is given flawed data.

4) **Auditability**: Just as third parties are often used to identify security vulnerabilities, auditing could be used to identify potential ethical implications. The third party could exist within the same company, to protect proprietary information, but would have a different perspective from the original algorithm designer and could creatively search for potential unintended consequences.

5) **Fairness**: Biases can be “baked in” to existing data, and automated decisions can amplify structural discrimination. Analysts should be aware of this risk, and evaluate for potential discriminatory effects.

Recognizing the increasing risk of unintended consequences in the growing field of analytics, some organizations and professional societies in the area have taken the step of establishing explicit ethics guidelines to heighten awareness and stress the importance of responsible behavior (see Figure 1.8, for example).

1.6 The Changing World of Analytics

The analytics landscape has changed rapidly in recent years, and the pace of change continues to accelerate. Dramatic reductions in the cost to store and transmit data combined with the “Internet of Things” have resulted in much larger and more readily available data sets. Additionally, use of analytics is becoming more widespread, as many influential people and organizations publicize the benefits. Universities are responding to the shortage of trained analysts by developing undergraduate and graduate programs of various levels of rigor, and conferences related to “Big Data and Analytics” abound. At some
Whereas operations research and analytics can have a deep impact on society, with applications ranging from medical decisions to national defense, business strategy, public policy, and many other contexts, we aspire to be:

- Accountable for our professional actions and the impact of our work.
- Forthcoming about our assumptions, interests, sponsors, motivations, limitations, and potential conflicts of interest.
- Honest in reporting our results, even when they fail to yield the desired outcome.
- Objective in our assessments of facts, irrespective of our opinions or beliefs.
- Respectful of the viewpoints and the values of others.
- Responsible for undertaking research and projects that provide positive benefits by advancing our scientific understanding, contributing to organizational improvements, and supporting social good.

Whereas our work influences the success and standing of our organizations (universities, businesses, government, and nonprofit agencies) as well as our constituencies (students, clients, customers, and suppliers), we aspire to be:

- Accurate in our assertions, reports, and presentations.
- Alert to possible unintended or negative consequences that our results and recommendations may have on others.
- Informed of advances and developments in the fields relevant to our work.
- Questioning of whether there are more effective and efficient ways to reach a goal.
- Realistic in our claims of achievable results, and in acknowledging when the best course of action may be to terminate a project.
- Rigorous by adhering to proper professional practices in the development and reporting of our work.

Whereas we are part of the profession of operations research and analytics and have an obligation to help advance the profession and to uphold high standards on behalf of our colleagues and future generations, we aspire to be:

- Cooperative by sharing best practices, information, and ideas with colleagues, young professionals, and students.
- Impartial in our praise or criticism of others and their accomplishments, setting aside personal interests.
- Inclusive of all colleagues, and rejecting discrimination and harassment in any form.
- Tolerant of well-conducted research and well-reasoned results, which may differ from our own findings or opinions.
- Truthful in providing attribution when our work draws from the ideas of others.
- Vigilant by speaking out against actions that are damaging to the profession.

Figure 1.8 Guidelines on ethics from analytics professional society INFORMS [20].
point the hype will diminish, but because the results are real, analytics will not go away. Indeed, we expect organizations will continue to rely even more on analytics-based decision support in the future, as the benefits become increasingly well understood.

Increased volume of data has motivated the rise of parallel and distributed computing systems and the development of new algorithms for efficiently storing and retrieving data in these systems. Although particular vendors and platforms may rise or fall in popularity, the general theme is clear: problems that involve more data than comfortably fits on a single computer can be distributed over many computers in a way that makes answering certain common types of questions very efficient.

Certain kinds of data, such as real-time transaction data, or web browsing data, can be particularly massive, and the future storage requirements will likely grow astronomically. Distributed information systems that store this type of data are particularly well suited to descriptive analytics. It is straightforward to divide a giant database across multiple machines and let each one report back on the subset of data elements that match a given query. This can make report-generating systems run much faster.

Predictive and prescriptive analytics generally require more sophisticated mathematical models that are more difficult to fit into a distributed computing paradigm. This has led to the development of new algorithms for old methods that are better suited to distributed environments, as well as to entirely new methods. For example, “deep learning” methods are a form of machine learning that rely heavily on access to vast quantities of data. Traditional statistical techniques designed for small data situations rely on structure imposed by the analyst. Deep learning is attractive because (in theory) it allows the computer to find structure in the data without the human analyst having to first teach it a great deal. In practice, this depends on having a sufficiently large and rich data set available with a sufficiently high signal within the noise. Deep learning shows particular promise in situations such as voice and image recognition, where defining a structure is especially challenging, and where vast quantities of data are indeed readily available.

Traditional statistical methods are still the preferred choice in many other settings, where there is known structure and the amount of available data are more limited. For example, market research data are integral in many common business decisions, providing considerable value despite being “small” data.

We expect that both “big” and “small” data situations will remain important. Big data often follows the data-centric framing and bottom-up collection process, whereas small data generally start from the top-down decision-centric view. We predict that the most significant advances in applied analytics will come from combining the best of both worlds—leveraging the deep subject matter expertise required for small data applications to make the most of big data opportunities.
Some people are working on approaches to try to automate the analytics process further. At the moment, it is a very labor-intensive process requiring people with significant levels of education and experience. Will it be possible for computers to automate much of that? Perhaps someday, but at least for the foreseeable future, it seems to us that subject matter experts will continue to play a key role in many analytics projects. The real world is infinitely complex. Explaining the world to a computer is not easy. Cleaning data and interpreting results are complex cognitive tasks not easily replaced by current forms of “artificial intelligence.” Applying existing mathematical methods to a problem once the data are clean can be reasonably straightforward, but that is not the time-consuming, rate-limiting step in the analytics process, so automating it will not really solve the problem of scarce talent. Creating new mathematical methods suited to emerging new decision questions will long remain solely the province of human experts.

### 1.7 Conclusion

This chapter broadly defined analytics, using conceptual frameworks (data-centric and decision-centric) and high-level classifications (descriptive, predictive, and prescriptive). We introduced considerations for implementing analytics in organizations, and potential ethical implications. The following chapters will describe in more depth how analytics can be successfully implemented, including how to get started, data and organizational requirements, solution methodologies, and management considerations.

Analytics offers exciting and vast possibilities. The analytics landscape is rapidly evolving, and new methods, data sources, and computing resources create new opportunities. Businesses have opportunities to improve profit by growing revenue or reducing cost. Governments and nonprofits have opportunities to use resources more efficiently and deliver better services. For society more broadly, there are opportunities to improve health outcomes, reduce environmental impact, improve quality of life, and increase transparency and fairness. However, capturing these potential gains is not easy. Effectively implementing analytics requires the right data, the right tools, the right people, and the right systems.

### References


6 Hartford T (2014) Big data: are we making a big mistake? *Financial Times*, March 29.


