

- » Discovering the value of data
- » Complying with regulations
- » Protecting customer privacy
- » Predicting expected actions with data
- » Changing plans to control outcome

Chapter **1**

Driving Business with Data and Analytics

In the twenty-first century, personalization is king — and data makes personalization possible. A good friend can pick out a much more personal gift for you than a stranger because that friend knows what you like and dislike. Marketers have known for decades that establishing a connection with someone can dramatically increase the chances that the person will become a customer. Organizations' desire to attract customers and increase sales drives the pursuit of meeting consumers' needs.

Consumers demand personal attention and have come to expect a high level of individualized customer service, online or when physically shopping in a bricks-and-mortar store. Due to advances in consumer interaction sophistication, the bar is high for all types of organizations. For example, it isn't good enough for web searches to return a general list of responses. Consumers expect their searches to be personalized and filtered based on their preferences. Today's search engines, and most shopping sites, suggest responses before you even finish typing. It's almost as if the search function knows you and what you're about to ask.

The capability to guess what a user is likely to ask or find interesting is based on data. Humans are creatures of habit and most processes (and even natural events) tend to be cyclic. The repetitive nature of behavior means that if you have enough historical data, you should be able to predict what comes next. Expending effort to

collect, maintain, and analyze data related to your organization's operation can help to reduce costs, limit exposure to fines and lawsuits, and lead to increased revenue.

In short, learning how to use your data helps you learn how to make your organization more profitable. In this chapter, you learn about ways that data can provide value to organizations.

Deriving Value from Data

The increased trend toward personalized offerings both depends on data and exposes data's importance to business operations. Data is no longer simply a consequence of engaging in transactions — data is necessary to increase the volume of transactions. Organizations are learning how valuable data is to their capability to conduct and expand operations. If you want to stay competitive in today's economy, you'll have to provide an experience that's responsive and personal. Data from previous transactions makes it possible to anticipate subsequent activity and tailor offerings to customer and partner preferences.

For example, the items you've bought online in the past give online shopping sites such as Amazon.com enough of your background to be able to make suggestions for additional purchases. Using past data to recommend future purchase or actions is a common way to derive value from data. In this section, I introduce three ways organizations can identify data with the greatest potential value.

Monetizing data

Over the past two decades, many organizations have come to view data as the primary fuel of the information age. Since the dawn of the twenty-first century, many organizations with data as their central business driver either started or expanded rapidly. Amazon relies on customer data to make additional purchase suggestions, while companies such as Facebook and Google rely on data as their primary product to drive advertising revenue. All these organizations found ways to turn data into revenue.

As data becomes more directly associated with revenue, data giants Google, Facebook, and Amazon control a growing demand for access to that data. Users have long been encouraged to share their personal data and activities, with little or no compensation. In the beginning, the perception was that sharing personal data was harmless and had little value.

However, a growing number of consumers and business partners realize that their data has value. Legislative bodies have recognized the importance of personal data and are passing new levels of privacy protection legislation each year. Data not only has value in and of itself but, when linked to other related personal data, can also provide valuable insight into personal behavior.

The realization that personal data has value has resulted in a game of sorts. Organizations that value consumer data attempt to acquire as much data as possible, while consumers are becoming more willing to deny free access to their personal data or demand compensation. Compensation often takes the form not of a direct monetary payment but of other perks or discounts.

Exchanging data

As organizations realize the increasing value of consumer and partner data, the more they explore ways to leverage that value. When consumers interact with any organization, or organizations interact with partners, a trail of data artifacts is left behind. Artifacts that document transaction timing and contents, as well as any changes to data, describe how entities interact with organizations. As more interactions with all types of organizations become more automated, the quantity and frequency of data artifacts increases.

Organizations that collect data artifacts find that not all are useful — at least not to that organization. However, as data becomes more and more valuable, many organizations have expanded the scope of data they collect with the intention of selling that data to other organizations. As data becomes a source of both direct and indirect revenue, data collection and management moves from a supporting role to a strategic planning concern.

For example, political campaigns routinely spend large sums of money to purchase demographic information on customers who have purchased specific types of products. Political candidates who strongly support environmental issues find value in identifying people who purchase green products because these customers are likely potential supporters. The identities can then be used to solicit campaign donations.



TIP

The overuse of data selling has led to concern and frustration over personal privacy. Most people come to the eventual realization that online activity has consequences. Every time you provide your email address or telephone number to anyone, your data will likely end up being used by some other organization (or probably multiple organizations). Always be careful about what data you allow others to use.

Sharing and exchanging data isn't always bad. In some cases, you want your data to be shared among businesses and organizations. For example, sharing the

complete service history for your car could make getting service easier and more reliable. With shared service data, you could take your car to any service provider and not have to remember the last time you had the oil changed or tires rotated. Techniques that support beneficial and responsible data sharing among organizations can be valuable to business and consumers.

Verifying data

One of the obstacles to realizing the full value of data is the dependence on its quality. Quality data is valuable, while incomplete or untrusted data is often worthless. What's worse, low-quality data may require more budget to clean than it will potentially generate in revenue. The only way to realize data's true value is to ensure that the data is valid and represents entities in the real world.

Verifying data has long been one of the highest costs associated with collecting and using data. Campaigns that depend on physical or email addresses will have little effect if the target addresses are largely incorrect. Bad data can come from many sources, including mischievous data submission, sloppy data collection, or even malicious data modification. An important aspect of relying on data is putting controls in place that verify the source of any collected data, along with that data's adherence to collection requirements.

A simple approach to verifying data in a distributed environment is to carry out a simple validation at the source and again at the server as the data is stored in a repository. While validating data at least twice may seem excessive, the practice makes user errors easier to catch and ensures that data received by the server is clean.



TIP

Validating data twice makes it possible for client applications to quickly catch errors, such as too many digits in a phone number or a missing field, while the server handles more complex validation tasks. A server may need access to other related data to ensure that data is valid before storing it in a repository. Server validation could include things such as verifying that order quantities are available in a warehouse and that data wasn't changed by a malicious agent during transmission from the client.

One of the reasons data verification is so important is that organizations are relying more and more on their data to direct business efforts. Aligning business activities with expectations based on faulty data leads to undesirable results. In other words, decisions are only as good as the data on which those decisions are based. The “garbage in, garbage out” adage still holds true.

Understanding and Satisfying Regulatory Requirements

The information age offers many new opportunities and just as many (if not more) challenges. The vast amount of data available to organizations of all types empowers advanced decision-making and raises new questions of privacy and ethics. Consumer protection groups have long been voicing concerns about how personal data is being used. In response to discovered abuses and the recognition of potential future abuses, governing bodies around the world have passed regulations and legislation to limit how data is collected and used.

Although collecting a few pieces of information about a customer may seem innocent, it doesn't take long for accumulated data to paint a picture of an individual's personal characteristics and behavior. Knowing the past behavior of someone makes it relatively easy to predict the person's future actions and choices. Predicting actions has value for marketing but also poses a danger to an individual's privacy.

Classifying individuals

The concern is that personal data has been, and will continue to be, used to classify individuals based on their past behavior. Classifying individuals can be great for marketing and sales purposes. For example, any retailer that can identify engaged couples can target them with ads and coupons for wedding-related items. This type of targeted advertising is generally more productive than general marketing. Advertising budget can be focused on target markets that provide the greatest ROI.

On the other hand, knowing too much about individuals may violate a person's privacy. One instance of a privacy violation was a result of the Target Corporation's astute data analysis. Target's analysts were able to identify expectant mothers early in their pregnancy based on their changing purchasing habits. When a new expectant mother was identified, Target would send unsolicited coupons for baby-related items. In one case, the coupons arrived in the mail before the mother had shared that she was pregnant; her family found out about the pregnancy from a retailer. Privacy is such a difficult issue because legitimate actions can violate a person's privacy.

Identifying criminals

Another aspect of privacy is when criminals, or other individuals who deliberately want to operate anonymously, hide their identities from exposure. Privacy may be important to the general population, but it's a necessity for criminal activity. The ability to deny, or repudiate, some action is crucial in avoiding discovery and

capture, and to any subsequent defense. Money laundering and fraud are two activities in which privacy and anonymity are desired to obfuscate illegal activity.

On the other hand, law enforcement needs the ability to associate actions with individuals. That's why laws exist that protect the general public but allow law enforcement to conduct investigations and identify alleged perpetrators.

Protecting the privacy of law-abiding individuals while identifying criminals has become important across a spectrum of organizations. To enable law enforcement to deal with online privacy issues, legislative bodies have passed various laws to address those issues directly.

Examining common privacy laws

Here are a few of the most important privacy-related laws you'll likely encounter and may be compelled to satisfy:

- » **Children's Online Privacy Protection Act (COPPA):** Passed in 1998, COPPA requires parental or guardian consent before collecting or using private information about children under the age of 13.
- » **Health Insurance Portability and Accountability Act (HIPAA):** Passed in 1996, HIPAA modernized the flow of healthcare information and contains specific stipulations on protecting the privacy of personal health information (PHI).
- » **Family Educational Rights and Privacy Act (FERPA):** Passed in 1974, FERPA protects access to educational information, including protection for the privacy of student records.
- » **General Data Protection Regulation (GDPR):** Passed in 2016 (and implemented in 2018), GDPR is a comprehensive regulation from the European Union (EU) protecting the private data of EU citizens. Every organization, regardless of location, must comply with GDPR to conduct business with EU citizens. The EU citizen must retain control over his or her own data, its collection, and its use.
- » **California Consumer Protection Act (CCPA):** Passed in 2018, CCPA has been called "GDPR lite" to imply that it includes many of the requirements of GDPR. CCPA requires any organization that conducts business to protect consumer data privacy.
- » **Anti-Money Laundering Act (AML):** AML is a set of laws and regulations that assists law enforcement investigations by requiring financial transactions to be associated with validated identities. AML imposes requirements and

procedures on financial institutions that essentially make it very difficult to transfer money without leaving a clear audit trail.

- » **Know Your Customer (KYC):** KYC laws and regulations work with AML to ensure that businesses expend reasonable effort to verify the identity of each customer and business partner. KYC helps to discourage money laundering, bribery, and other financial-based criminal activities that rely on anonymity.

Predicting Future Outcomes with Data

Data can unlock lots of secrets. Data you collect through regular interactions with your customers and business partners can help you understand them and better meet their needs and wants. Assuming you have taken measures to protect individual privacy and have permission to collect and use the data, analyzing that data can benefit your organization and your customers (and partners, too).

A common way to use data is to build analytics models that help to explain the data, uncover hidden information, and even predict future behavior. Data analytics is all about using formal methods to unlock secrets that your data is hiding. These secrets aren't hidden on purpose — they just get lost in the mountains of data you collect. Without a structured approach to examining your data, you might miss some of its value that can lead to increased revenue.

Classifying entities

An *entity* is any object that your data describes, such as a customer, a vendor, a product, an order, or anything else that has characteristics data items can describe. In traditional database terms, an entity would correspond to a record or a row. The concept of a row maps to a spreadsheet concept as well. Think of a spreadsheet of customers. Each row would contain all the data that describes a single customer. Figure 1-1 shows a collection of customers in a table format.



These customers are stored in a comma-separated value (CSV) text file named `customer.csv`, and displayed in Visual Studio Code using the Edit as CSV extension. To learn more about Visual Studio Code and its extensions, see Chapter 4.

Note that each customer has a set of characteristics, such as name, address, and contact, stored in separate columns. Data analytics models use these different characteristics, also called *features*, to examine how different entities are related.

Customer ID	Country	Name	Address	City	State	Postal Code	Contact	
1	USA	Lift Tours	276 North Drive	Burlington	MA	01730	Gloria Shepley	
2	Finland	Upton Frisbee	Rantapolku 3	Oulu	Uusimaa	45321	Urho Lappajoki	
3	USA	Hopps	Suite 415	40 Grove St.	Atlanta	GA	02132	Michael Trahser
4	United Kingdom	Go Fishing Ltd	Unit 2	83 Ponders End Rd	Harrow	Middlesex	HA8 8AB	Alan Proghinok
5	USA	Match Point Tennis	600 Homer Pl	Address 2	Boston	MA	02134	Robert Dior
6	United Kingdom	Fanatical Athletics	60 Bloop Bridge Rd	Leyton	Montgomery	AL	48855	Andrew Strong
7	Finland	Aerobics valine Ry	Peltolantie 3		Uusimaa	4567	Elli Imanen	
8	USA	Game Set Match	Box 60	Deatsville	AL	36114	Tove Bruak	
9	Finland	Pöytäpöytä Pyry	Puikolantie 2	Riihimäki	Etelä-Savo	44900	Markku Ventoilinen	
10	United Kingdom	Just Soggers Limited	Fairview Trading Est	Shoe Lane	Rainnabottom	Sussex	BL0 9ND	George Lacey
11	Finland	Kelkku ja Biljardi	Veturinmenkijä 8 A		Helsinki	Uusimaa	230	Riitta Hammermäki
12	Finland	Suuri Lautavälikassat	Venemestarinkatu 35		Salo	Varsinais-Suomi	24240	Veikko Lahtio
13	Finland	Biljardi ja tennis	Ulkohäätä 1		Monsala	Uusimaa	4600	Tuomo A.L. Mattinen
14	France	Paris St Germain	113, avenue du Stade	BP 338	Paris	Saône	75012	Philippe Daiglt
15	USA	Hoopla Basketball	87 Colummet St	Egg Harbor	NJ	08215	Robert Johns	
16	USA	Thundering Surf Inc.	354 Market St	College City	TX	75763	Fran Horan	
17	United Kingdom	High Tide Sailing	178 Schooner Hill	South Killingholme	Birmingham	Humber-side	DN48 3LU	Angela Buckle
18	Finland	Astin Metsästysseura	Vanhainkirkkokuja 1		Syväsky	Keiski-Suomi	19700	Antti Oymrinmaa
19	USA	Buffalo Shuffleboard	155 Carolina Ave		Buffalo	NY	13142	George Barfkus
20	Finland	Espeon Pöytäpelikassat	Sirkkälöntie 18		Espos	Uusimaa	1200	Pekka Mäkeläjänt
21	USA	Pedal Power Cycles	11 Perkins St		Boston	MA	02145	Alicia Primes
22	Finland	Kespeet Kallikka	Kokkilaantie 4		Forssa	Etelä-Häme	30100	Pentti Koskinen
23	Finland	Jazz Futei Kauppa	Yrjontkatu 10		Pori	Satakunta	28100	Markku Vaa-Iedoja
24	France	La Boule Lyonnais	rué de la gare	Z.A. Bellecour	Lyon	Rhône	69010	Mir Roulemussaa
25	United Kingdom	Bulls Eye Sports	Highgrove House	180 Goodshot Street	Leeds	West Yorkshire	LS17 4TS	Kevin Batten
26	Austria	Bumm Bumm Tennis	Wäldermarkt 17		Wien	Öst	A-1020	He. Ballner
27	Finland	Luisin ja Pöytä Oy	Ruututie 7 A		Laiha	Etelä-Pohjanmaa	66400	Jarkko J. Scott
28	USA	Bug in a Rug-by	120 Henry St		Yamhill	OR	97148	Chip Clark
29	USA							

FIGURE 1-1:
Customer entities
presented as a
table.

One type of analysis is to examine the features of different entities to see if some features can help group entities or imply some relationship. For example, suppose you asked a group of people to name their favorite baseball team. You would expect that most people who answered “the Colorado Rockies” most likely live near Colorado. However, you can’t always make such simple associations. If you asked the same question in the 1990s, not everyone who answered “the Atlanta Braves” lived in Georgia. During the 1990s, cable TV was becoming popular and Turner Broadcasting System, whose owner also owned the Braves, broadcast all Braves games nationally. Many people who didn’t live in Georgia became Braves fans.

The Braves example shows that analytics models cannot be trusted unconditionally. Data analytics can provide tremendous value but also requires care and diligence to build models that return results that hold true over time.

Assuming that you invest sufficiently to build good models, classification models can help to identify entities that are similar. Similarity information helps organizations develop targeted marketing campaigns and services to give customers and partners the sense of being treated individually. You learn about several classification models in Chapter 7 and build a few in Chapter 10.

Predicting behavior

Although the capability to classify entities to identify groups of similarity can be valuable, analytics can also make predictions. Past behavior is a strong indication of future behavior. Humans tend to repeat actions and decisions, so you can use models that identify patterns to predict future actions. The capability to predict future actions can have tremendous value to organizations. If an organization can determine items that tend to be purchased together frequently, it can use that information to make additional purchase suggestions.

You've undoubtedly seen frequent item analysis results when you shop online. When your favorite website recommends that you purchase an additional item, and that item makes sense, it's because other people have bought that same item set in the past. How does the website know that? It used analytics.

One of the common analytics models you learn about in Chapter 7 and build in Chapter 11 is regression. Don't worry about the name right now (or the math). Regression is kind of like calculating the slope of a line on steroids. A *regression model* basically examines your data and figures out a line (or a curve) that matches the data you've seen. After you can graph your data, you can use that graph to guess what will happen based on new input data.

Let's see how that can help. Figure 1-2 shows a linear regression model built on audition data and resulting score data. This example comes from an example you use to build this model in Chapter 11.

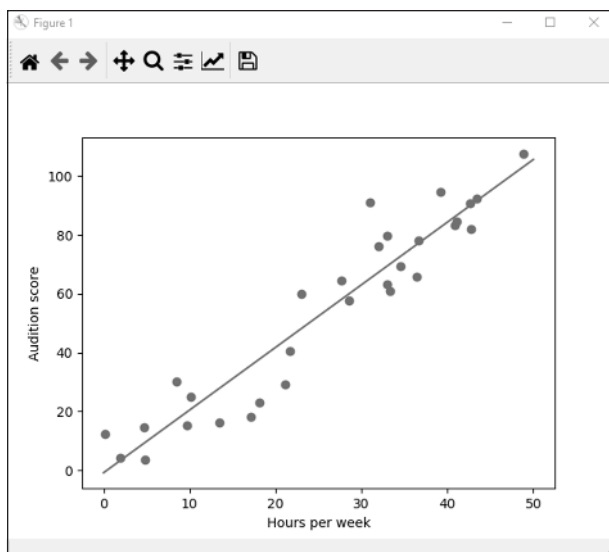


FIGURE 1-2: Linear regression model using hours practiced and audition scores data.

Here's the explanation you see again in Chapter 11: Suppose you're helping student musicians prepare for honor band tryouts. You've collected historical data on how many hours a week each student practiced, whether the student was accepted in the honor band, and what audition score each student earned. As you would expect, a linear correlation exists between hours of practice and audition score: The more a student practiced each week, the better score that student earned at his or her audition. A linear regression model can predict any student's audition score if you know how many hours that student practices each week. If you have a

student who practices 30 hours per week, you could expect that student to earn a score of about 60 on the audition.

Regression models can help to accurately predict future actions. Using data to know what's next can be worth its weight in gold when making business decisions. (Yeah, I know data doesn't have weight, but you get the point.)

Making decisions based on models

Analytics models can help organizations make astounding decisions and gain lots of money. They can also lead organizations to make dumb decisions and lose lots of money. The trick is in knowing how good your models are.

This book is about building analytics models using blockchain data. You learn about blockchain technology and data in Chapters 2 and 3, but don't forget that although the quality of your data is important, building the right model is crucial to getting quality output. Never rely on your first choice of a model or on a single model. Always compare model types and configurations to find the right combination to return the highest quality results.



REMEMBER

If you take only one thing away from this book, I hope that it is to demand measurable verification from every model you build. You should be able to provide metrics for each model indicating its accuracy and that it actually works. Never release a model to your business unit without exhaustive verification. Your organization will use your models to make big decisions. Do your best to give it good tools.

Changing Business Practices to Create Desired Outcomes

Classifying your customers or building models to predict what comes next can help your organization be more responsive to needs. You can use analytics to help plan better and be ready for whatever comes next. But with some additional work, you can do far more with analytics results. Instead of just getting ready for what might happen next, you can use analytics results to alter today's activities and affect future outcome.

Predictive analytics predicts what future results may be. The next step in analytics maturity is *prescriptive analytics*. With *prescriptive analytics*, the model identifies changes you can make now to achieve a desired outcome. For example, prescriptive analytics can tell you how many tables to set out in a restaurant or which

register lanes to open in a grocery store to meet sales goals. Prescriptive analytics gives organizations the leverage to make operational changes based on their understanding of data that leads to satisfying their goals.

Defining the desired outcome

In the preceding section, you learned about using analytics models to make predictions of future outcomes. There can be tremendous value in prediction, but you can use analytics also to set the outcome and tell you how to get there. Think about it. It's one thing to predict next week's sales, but wouldn't it be cool to set your next week's sales goals and let your analytics models tell you how to get there? With good analytics models, it's possible.

Predictive analytics basically gives you an equation: $y = mx + b$ (yes, that's a simple one and the same as the point-slope form of a line). Your model provides values for m and b . Your data provides a value for x and you solve for y . Simple algebra.

Prescriptive analytics is a little different. Prescriptive analytics ask the question: "If I choose a value of y , what value of x will get me there?" In other words, you choose a value of y (maybe your goal for next week's sales), and then solve for x . After you know x (perhaps x represents the number of prospect calls you need to make), you know what it will take to reach y (your sales goal). At its core, it's still simple algebra.

Even though the algebra is simple, putting prescriptive analytics into practice can be tricky. In algebra, equality is reflexive, which means you can read left-to-right or right-to-left. Technically, models should work the same way, but they don't always work that simply. Prescriptive analytics can provide some guidance on reaching goals, but you always have to take that guidance with a grain of salt. Try your model's recommendations, and then evaluate the results. Fine-tune your changes, and then try it again. The best use of prescriptive analytics is as a good suggestion, not a surefire approach to reaching goals.

Building models for simulation

One of the challenges in prescriptive analytics is the iterative and flexible nature of using models this way. Predictive analytics is pretty straightforward. You can determine future outcomes within a known range of error. When turning that model around and using it for prescriptive purposes, you can never be sure that your model is taking into account all the influences that affect outcome. The outcomes your predictive model measures may include unsampled features (characteristics) that happen even though you don't measure them. If this is the case, just changing one feature may not have the effect you expect.

Because prescriptive analytics is more than just turning a predictive model backwards, you'll have to run your model multiple times over your dataset, changing a single feature at a time. Building a model that is flexible enough to respond to multiple feature changes is the basis of simulation. You're simulating the nature of reality, which encompasses multiple features that change and some level of unmeasured uncertainty.

Investing the effort to build a good simulation can more than pay for itself. A solid simulation is flexible enough to change as new input shows different trends and still provide output that you can trust. A simulation that tells you how to reach your goals is even better than knowing the future.

Aligning operations and assessing results

The best response to having good analytics models is to change operations based on your model output. Whether your focus is understanding your business, predicting tomorrow's environment, or using your models to direct decisions, analytics will add value only if you make changes based on what you learn.

Using analytics models, especially those built on data from a blockchain, is the purpose of this book. As you work through each chapter, you should gain an appreciation of the rich data available to you, and how you can use that data to enhance your organization. Enjoy the journey.