HubSpot

Introduction to Data Analytics

Table of Contents

HubSpot

Introduction

What is data analytics, and why is it important?

Data is much more than a collection of numbers. But if you don't do anything with it, it's just that—a nice collection taking up space on your hard drive. Instead of filing your data away, never to be seen again, put it to good use and analyze your data to help tell the story of your company's growth.

Data analytics is a science. It's the process of turning raw data into meaningful metrics companies can use to help make informed decisions. Think of it this way: the raw data is trying to tell a story, but it's jumbled and needs to be deciphered. That's what data analytics does. Data analytics uses tools, algorithms, and artificial intelligence to identify patterns and trends over a specific data set.

The goal of data analytics is to answer questions about possible outcomes. However, identifying possible outcomes is one of many reasons companies invest so much time in data analytics. Depending on your analysis, the data can tell you things like valuable market insights or if your company uses its resources efficiently.

There are many reasons why you should seek to analyze and understand your data. For one, having a firm grasp of your company's metrics can help you optimize your processes and procedures. With optimal strategies, you're more likely to see a more significant return on investments and be more efficient.

A second reason? Businesses that succeed are known for taking risks– but they don't usually take a chance without having an idea of the outcome. Data analytics helps lower the dangers of high-stakes decisions by allowing you to review carefully calculated outcomes.

The results of data analysis can also help with customer retention. Analyzing your customers' trends, patterns, and behaviors can help your team better market to those individuals. When you understand what your customers want according to the data, you can provide them with precisely what they need.

So, with data analytics, you can uncover your company's patterns and trends and then make better, more accurate assumptions, predictions, and conclusions for your team.

Ready to learn how to analyze your data? Let's take a look at the fundamentals of data analytics.
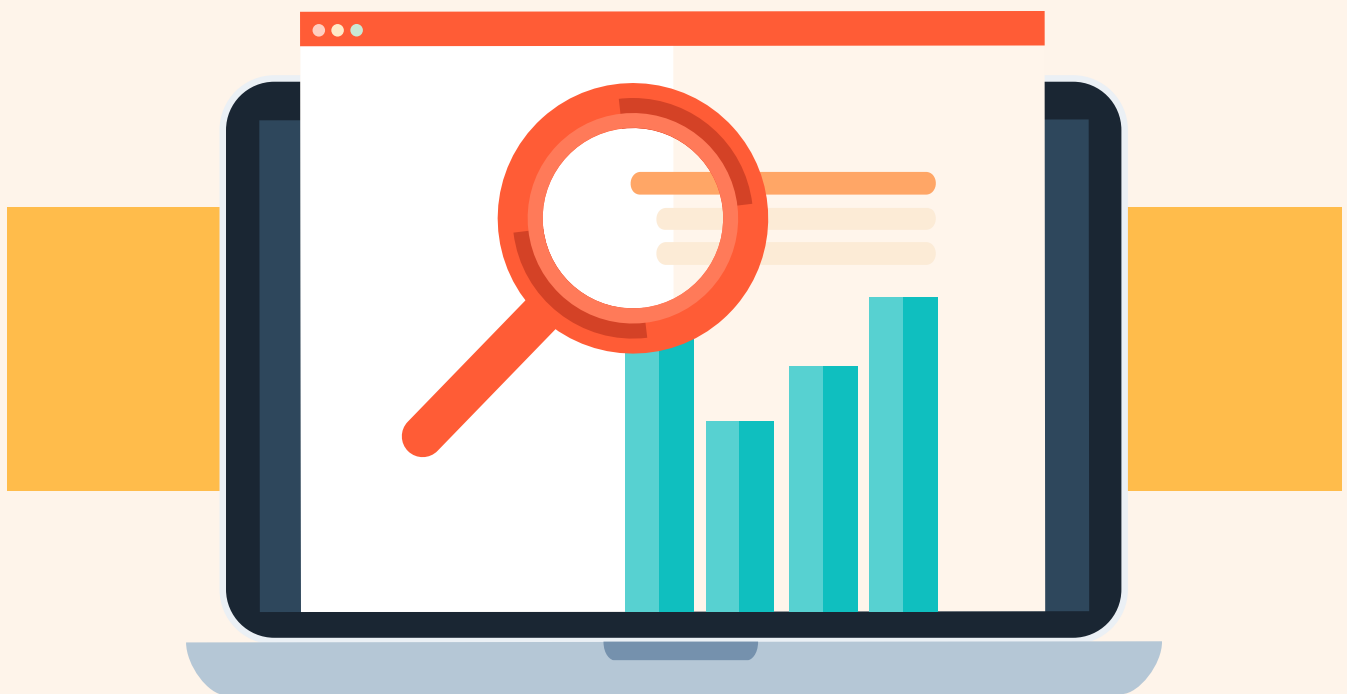
Fundamentals of Data Analytics

Data analytics is a science. Simply put, it is the collection and processing of data to gain insights and draw conclusions. Data analysis is vital to any business, no matter the industry, as the insights gained can help support you and your team when making crucial business decisions.

It's essential to understand precisely what data is. Most likely, numbers and figures come to mind when you hear "data." But it's so much more than that. Text snippets, images, and videos are also classified as data. This is important because it means you have options regarding the type of data you can collect. For example, you can run a text or sentiment analysis on social media comments to gain insight into your customers' thoughts and feelings regarding your product or service. In other words, you can count it as data if it's trackable.

Numbers, like dollars earned or lost, are essential to understanding how well your business is doing. However, metrics like customer satisfaction are also valuable to help you understand what your business is doing right and how you can improve.

Before we continue with how to implement data analytics in your business, we need to define some useful terms. Use this section as a reference guide.

- **Data collection:** This is the process of collecting data from various sources, including but not limited to spreadsheets, social media, and data sensors.

- **Data cleansing:** Data cleansing is reviewing raw data to ensure critical information is not missing. If an essential variable is missing, that data piece should be removed from the data set. During the data cleansing process, you'll also complete data **normalization** and **data transformation**. This means you're putting your collected data into usable form for analysis.

- **Data storage** of **Data silos:** This refers to the places where your data is stored. Think hard drives or warehouses.

- **Data ethics:** Did you receive prior consent from participants in your data sets? Do your data collection methods align with government policies like GDPR and CPPA? It's essential you can answer "yes" to both these questions before continuing with your analysis. Otherwise, you could end up in serious legal trouble.

- **Sampling:** This term refers to a technique and a data set. As a technique, sampling is done when there is a large data set. You'll choose a small sample set, or a sampling, of data for analysis.

- **Classification:** This is done when you set parameters for your data. You likely see classification happening throughout your workday. If you have spam filters turned on for your email inbox, this is classification. Spam filters are great at detecting spam and separating it from essential emails.

- **Clustering:** This is the process of grouping data sets together. We'll talk about this later.

- **Bias and variance:** Bias is a term that refers to system errors, while variance refers to errors that randomly occur. To get the best result from your data analysis, bias and variance are best balanced.

- **Correlation and causation:** Describes the events or variables that caused a specific outcome.

- **Overlifting and underfitting:** Overlifting occurs when the results of your analysis match the data sets too closely. While this might not seem bad at first, it is. It means somehow your training and testing sets combined and created an error. You'll need **cross-validation to solve this**, ensuring your testing data and training data are separate sets.

    Underlifting describes the phenomenon of oversimplifying the data. This is bad for your analysis because it will only represent part of the picture.

- **Machine learning:** Machine learning uses algorithms and, now, artificial intelligence to learn your data, classify it, build models, and make predictions. Machine learning can automate data analysis, saving your analysts time during the workday.

HubSpot

- **Deep learning:** A subset of machine learning that uses neural networks to function like the human brain.

- **Noise:** Corrupt or meaningless data.

As a business, you likely collect data daily (probably more like every minute!). The data you collect tells a hidden story of your company's performance. But you'll never get the full story without critically examining your data and understanding what it means.

Data analytics can be a manual process, and typically, this means a data analyst spends hours manually crunching data sets. While there is nothing wrong with this approach, let's be honest: we are human and sometimes make mistakes. Manual data analysis leaves room for error, which could change the outcome of what the data shows.

However, with the rise of new and faster technology, better algorithms, and the introduction of AI, data analysis can also be an automated process. Automated analysis reduces the chance of human error and saves your analysts time, leaving them with more time to do what they do best— make predictions and assumptions for your company.
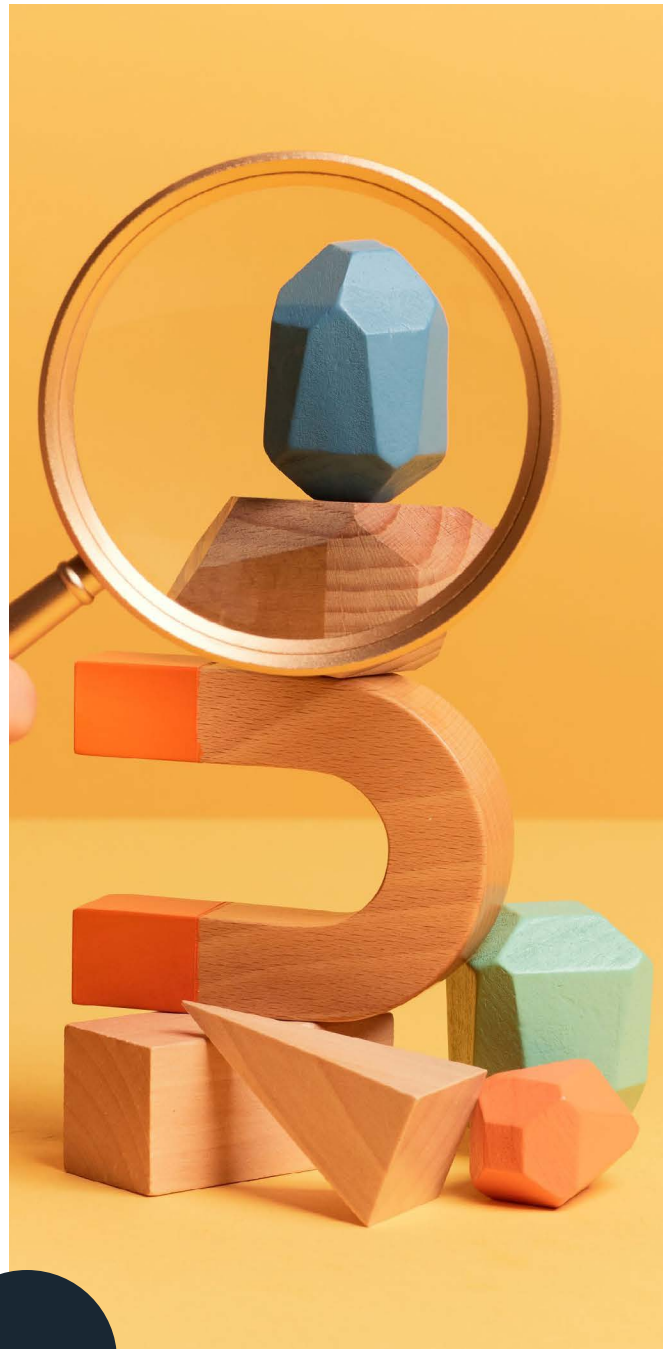
Using AI for Data Analysis

Any data analyst will tell you that algorithms are essential to their work. Most analysts will also tell you that using an algorithm to crunch numbers manually takes a significant amount of time. So, it is no surprise to anyone that artificial intelligence would make its way into data analytics.

Artificial intelligence, or AI, is more than a trend– a rapidly evolving technology that's here to stay. AI is a computer science powered by a machine whose goal is to mimic human intelligence. This human-like thinking allows the computer to detect patterns, make predictions, and problem-solve. When introduced to data analytics, AI can help data analysts quickly and efficiently run analyses on any given dataset.

As more and more businesses attempt to establish themselves as a data-driven company, it's a good idea to consider implementing AI into business practices. Although AI is an intelligent technology, it should not replace your data analysts. Instead, AI is a valuable tool to help data analysts compile a comprehensive overview of your company's processes and metrics. Think of AI as a secret weapon. It can help your business stand out amongst the competition and help you and your team better understand the markets and your customers.

Applications of AI techniques in data analysis

There are several reasons to use AI in your business practices, particularly for data analytics. Data analytics uses your business's data to tell your company's story. AI, though, helps tell the complete data-driven story, allowing you and your team to understand better what happened, why it happened, what's happening, what's likely to happen, and what could happen.

AI can be applied to help:

- Provide and explore insights
- Create datasets for training purposes
- Create dashboards and reports
- Predict market outcomes
- Understand markets and customer behaviors

- Forecast demand
- Make informed decisions
- Improve production and efficiency
- Monitor business performance

Artificial intelligence lends an extremely helping hand when attempting to make predictions with your data. We'll talk more about this type of analysis later, but predictive analysis and the above list aren't an exhaustive list of the uses of AI. Artificial intelligence can help you and your team with data analytics in hundreds of different ways. Before you jump into implementing AI into your process, let's look at some of the benefits and challenges of AI.

Benefits and challenges of incorporating AI in data analytics

You might think AI is not for you or your company. And you might be right. However, it makes your data analyst's job easier, primarily if your company collects large amounts of data, and it's worth considering this technology.

One of the reasons a company chooses to implement data analytics into its processes is to help them make decisions. Without AI, this responsibility lies solely on the analyst to look at the data, compute the numbers, and present the options. With large datasets, this can be challenging and time-consuming. There's always the chance that a formula is miscalculated or an essential piece of data is missing.

AI helps alleviate those problems. Artificial intelligence can quickly parse immense volumes of data. This dramatically improves the accuracy and efficiency of data review, leaving your analysts with more time to review results and consider what the data says. AI can help with decision-making, too, as it can easily predict outcomes depending on the analysis and model you choose. Those are just some of the benefits for your analysts. Let's not forget about your customers. AI technologies can help learn customer data and predict products and services your customers will like based on past purchases.

AI can be a fantastic tool for your business operations. However, there are a few drawbacks. AI and its algorithms are only as good as your datasets, meaning insufficient data will lead to inaccurate results. You and your team will need to ensure your data is ready for your applications, which could take significant time. AI is also not great at detecting bias in a dataset, so you must ensure your data accurately represents your customers.
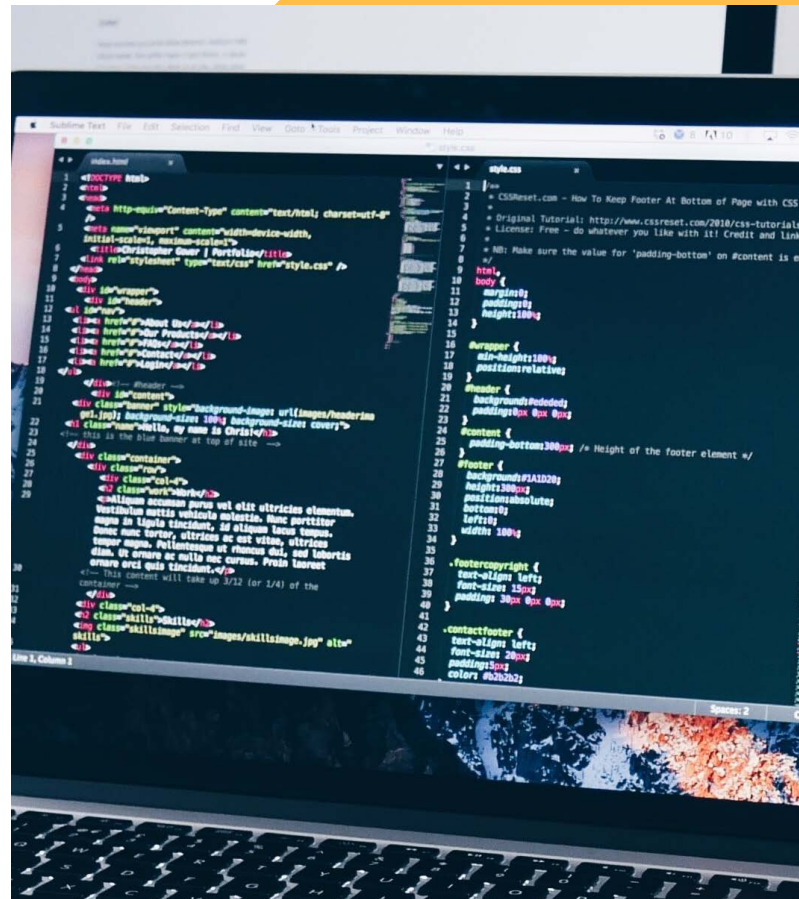
Because AI is an ever-changing technology, you'll need to ensure your teams continually keep up with trends, understand the complex algorithms, and are trained for the technology. AI will also require your company to work across departments, as you'll need to team up with your analytics team, your IT team, your infrastructure team, and any others who play an integral role in data collection and storage. AI can be costly, but its benefits greatly outweigh its cons.

With the help of AI, data analytics is a smart investment for any company, big or small. But before you decide to run any kind of analysis, you should consider several important factors, like which kind of analysis to run on your data. There are various types of data analysis to help you discover the big picture of your company.

Let's take a look at them.

Managing Data for Analytics

Before you can begin analyzing data, you must collect it. Because data can come from anywhere, your business is likely generating data every minute of the day. However, data collection becomes a problem if you do not have the proper management tools and systems. That's why you need to implement data management into your business operations. Data management is a critical part of data analytics.

Data management is data collection, organization, processing, and storage. Normally, data is managed by a data management team that consists of IT professionals, data scientists, and data administrators. It's important to create a team of professionals responsible for data management. The role of this team is to ensure the data collection methods comply with governing policies, like GDPR (or General Data Protection Regulations). They also determine how your data is defined and stored, and they help monitor the integrity of data and conduct any necessary security updates, data recovery, backups, and software installations.

You'll likely need to assign a member of each department as a data manager, too, but they'll work to maintain data on a smaller scale. This person can access the necessary data relevant to their department. They'll also be able to work closely with the data management to become their department's point of contact for anything data-related.

Let's take a moment to look at the necessary components of data management to ensure your data quality is top-notch and ready for analysis.

Sources of data collection

It's helpful to think of data as a life cycle. The first step of the cycle is data generation. Data is generated from various sources, and each source may have relevance to your business operations. There are three main types of data sources: first-party sources, second-party sources, and third-party sources.

First-party sources are sources of information that your company generates itself. These are sources of data where the data relates directly to your business operations. Social media interactions, transactions and receipts, observations, cookies, and customer survey results are considered first-party sources. Each source relates directly to your business and how your customers interact with your websites, products, and services.

Second-party sources are necessary data, too. Although this is not data your company generates, it's likely data other businesses in your field generate that can be useful. Secondary sources include published interviews, online databases, and government or institutional records. This data is likely in the public domain, and you can use it to train your algorithms before you test your data.

The last source of important data is called third-party data. Third-party data is collected from sources outside of your organization and, sometimes, industry. Normally, this data is bought, sold, or rented. Be wary of the validity of this data, though, because it may not have been collected according to government and industry standards. You'll need to ensure the data is trustworthy before you use it for any reason.
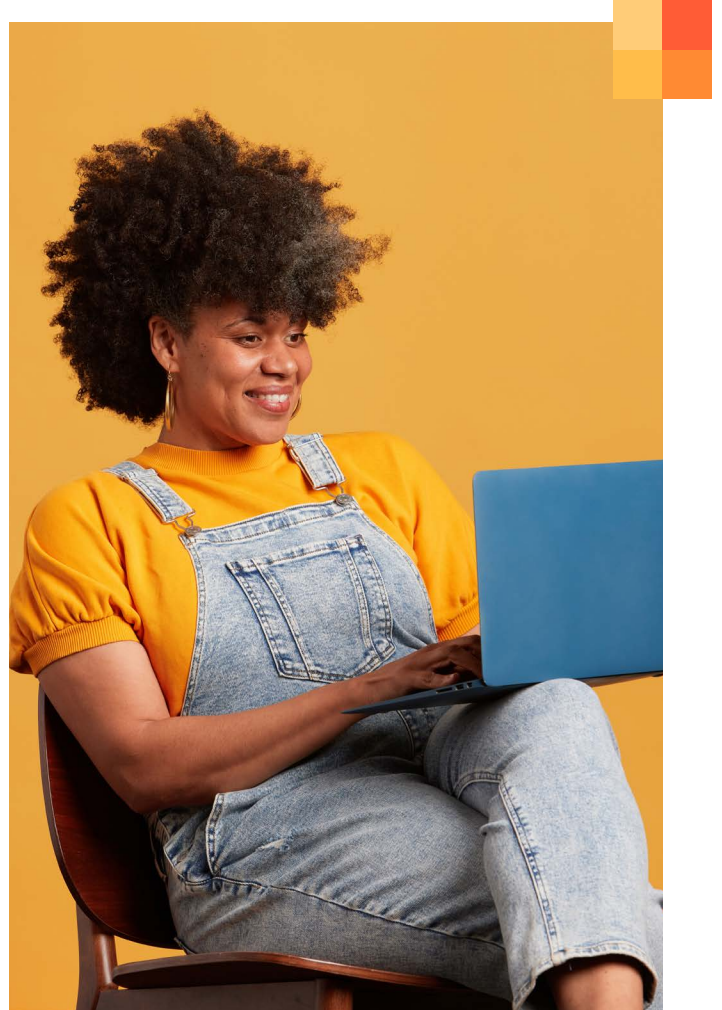
Reprocessing and data quality assurance

Once your data has been identified and collected, you or your data scientist should spend some time preprocessing it. Raw data, or the data you've collected directly from your sources, is not usually in a usable or readable form. It must be translated into a language your data storage system can understand. Plus, raw data likely contains errors or missing information. Data cleansing is an important part of data quality assurance. It's okay to throw out data that is missing or incomplete. Leaving flawed data in the dataset can cause significant issues and skewed results later.

Be sure to keep a watchful eye on the data pipeline, too. If you notice a large number of insignificant data, it could be that something in the data pipeline is broken, causing data points to be left out or corrupted before reaching their destination. If something is broken, you should fix it as soon as possible to ensure the quality of your data.

HubSpot

Data quality assurance also includes validation. This means you should continually be aware of collection methods to ensure they always comply with data policies and rules. If not, unethical or illegal collection methods can land your business in hot water with the federal government.

After the data has been assured for quality, preprocessed, and translated, the next step is to input the data into the data management system. How you store and organize your data is a key determining factor of what you can do with it later on. So, if you haven't already built or implemented a data management system, pay particular attention to the next section. In the next section, we'll cover the different types of data storage and how your storage methods can determine your analysis methods.

Data storage and organization

When thinking about data storage and organization, it's helpful to imagine a building. To construct the building, you first need software and a database. The database is the foundation of your building that allows for the construction of rooms. Inside each of these rooms, there is a place to store your data. Some rooms may be standard columns for numerical data, others may be components of graphs, and some rooms might be pools of unorganized raw data, like text, images, or sounds. When it's time to analyze something, you will just go to the particular room, extract the data, and send it for analysis. This is a simplified version of data storage. However, it provides a decent visual of your data infrastructure and how it functions.

There are two main types of databases we need to discuss. Those are SQL and NoSQL databases. An SQL database is a structured, relational database that requires data to be translated into a readable language. This means the data is stored and organized in a table or connected tables. This database allows for easy analysis and modeling because the data is likely already translated to a language an algorithm can read.

SQL databases are popular amongst data scientists because they follow the ACID criteria well. Each acronym letter describes four criteria components necessary for data integrity about how data moves throughout the system. Let's define ACID before we continue:

HubSpot

1) Atomicity:

This term describes data transactions. It means that each transaction of a dataset is counted as its transaction. If, for some reason, the transaction fails, it is not applied to the data. Instead, the data is reverted to its original state.

2) Consistency:

This property ensures that transactions remain consistent across the database. It also maintains data integrity and ensures the absence of data corruption.

3) Isolation:

Data scientists can quickly encounter problems if multiple data transactions coincide. Isolation ensures transactions do not interfere with one another.

4) Durability:

Durability refers to the security of a transaction. In other words, once changes are made to a database during a transaction, those changes are permanent and stored in the database.

Data stored in SQL databases is normally stored in a data warehouse. Data warehouses organize data into neat boxes, making queries easy. These boxes also determine the type of analysis you can run because it mostly collects historical data.

NoSQL databases are slightly different from SQL databases and have different purposes.. For some datasets, a structured database is not the best choice of storage and can't be immediately organized. That's where NoSQL databases come into play. NoSQL databases lack defined boundaries, models, and schemas. This means that data can be stored in large pools on the framework. These large pools are called data lakes. If you plan to run predictive analytics using AI technologies, particularly machine learning or deep learning, data lakes are imperative because they can manage a continuous data stream.

Let's say, though, that the data you collect can be stored in both a data warehouse and a lake. Instead of building two separate data management systems, you can combine your storage options and use what is known as a data lake house. Data lake houses are flexible, rigid when needed, can be easily queried, and allow for scalability.

The type of data storage system you implement directly determines the types of data analytics you can perform. Static datasets, or data in a data warehouse, are perfect for exploratory and descriptive data analytics, while data stored in a lake is necessary for predictive analytics.

Let's look at each type of analysis to understand better what each one can tell you about datasets.
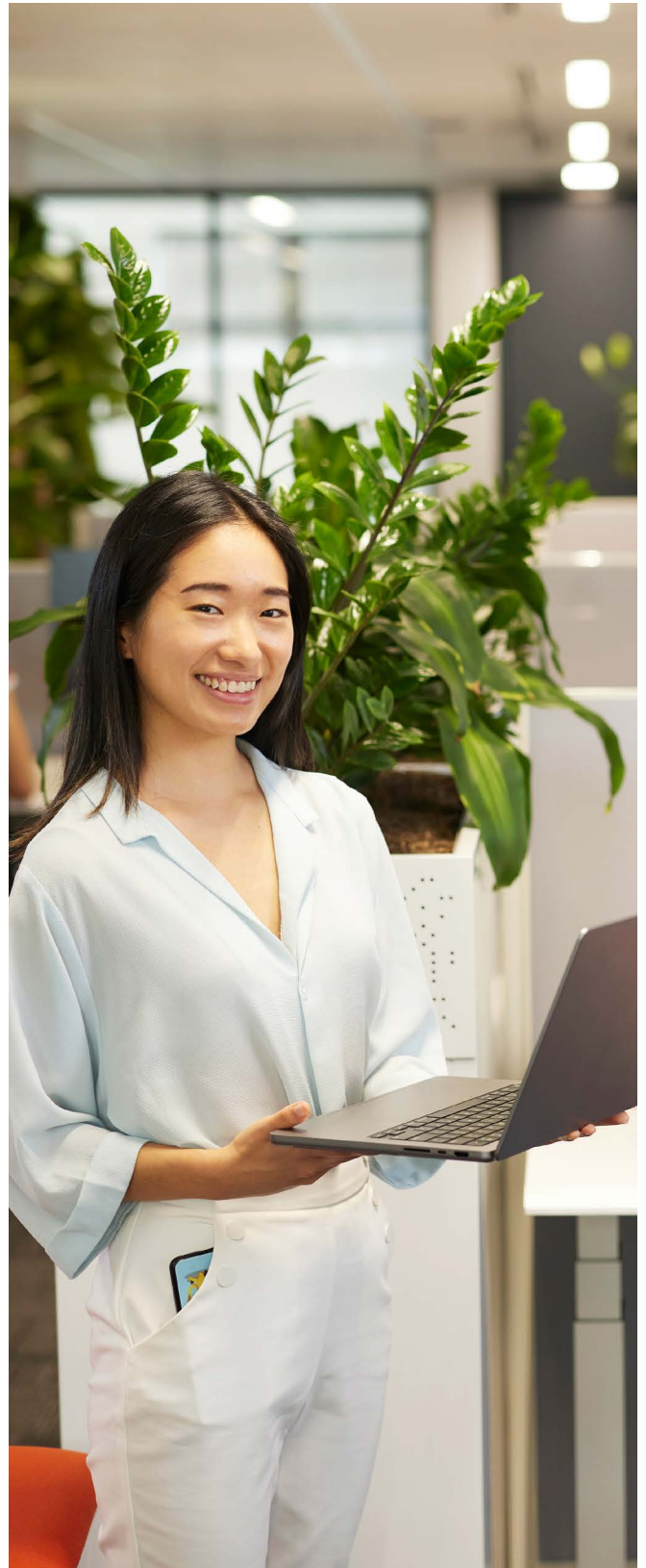
Types of Data Analysis

The term "data analytics" is broadly used to describe the process of data collection and analysis. Analyzing data can help you look at the "here and now," but it is also helpful to understand why your business is where it is today, how it got there, and where it's going.

The answers to "what happened?", "Why did it happen?", "What will happen?" and "What should we do?" are not all answered by the same type of analysis, though. That's why there are various types of data analysis, and it's essential to choose the correct type of analysis before you begin answering any of your questions or making predictions.

Let's first look at exploratory data analysis.

Exploratory data analysis

Let's pretend you have a complete data set and have no idea where to start with data analytics. Before you do anything else with your data, like putting it in an algorithm for forecasting or projections, you need to make sense of it. That's where exploratory data analysis comes in. Simply put, exploratory data analysis means looking at the data to search for patterns and trends.

Exploratory data analysis was developed in **1977 by mathematician John Tukey** to identify dominant characteristics in a data set. If you think of data analytics as a series of steps, exploratory data analysis is always the first step, no matter which type of analysis you plan to complete next.

There are two main parts of exploratory data analysis: collecting data and visualizing it. During the data collection phase, you'll need to gather your data, review it, and clean it to ensure it's usable. Data cleansing might mean you must throw out some pieces of data entirely because they are incomplete. Doing so will help ensure you have good data. Good, clean data will give you the most accurate analysis results.

Once your data is in a usable format, it's time to sort and visualize it. Data visualization refers to organizing your data in a way that makes it easy to see and understand. Think of bar graphs and pie charts. These are easy ways to understand what the data says quickly.

Let's use an example, though, of how to visualize data and what you can gain from exploratory data analysis. Let's pretend you have the ages of your customers and you want to create a customer persona. Before creating customer personas, you'll need information about your current customers to create an "average customer." This is why collecting customer data is essential— to help better market to your client base in the future.

Let's say you've collected the ages of twenty of your customers. To get a quick view of the ages of your clients, you'll first want to organize the ages from oldest to youngest. Then, draw a graph or use the Excel or Google Sheets software to create a bar graph or scatter plot. These graphs will give you a better idea of your customers' ages than looking at a list of out-of-order numbers.
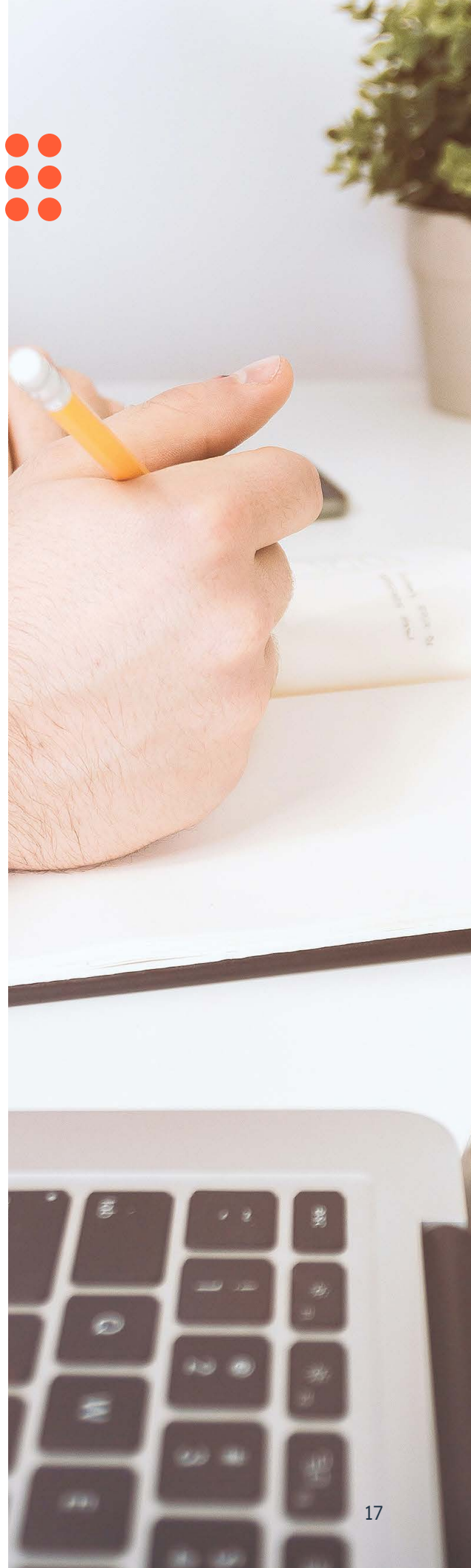
Data visualization can help you determine other essential information, too, like the mean, median, mode, and range of your data set. The mean of a data set refers to the average of your data. So, to find the mean, you'll add the data set together and divide the number by the total number of data points. The value you calculate represents the average of your data set, or for the example of customers' ages, the average age of customers who purchase your products.
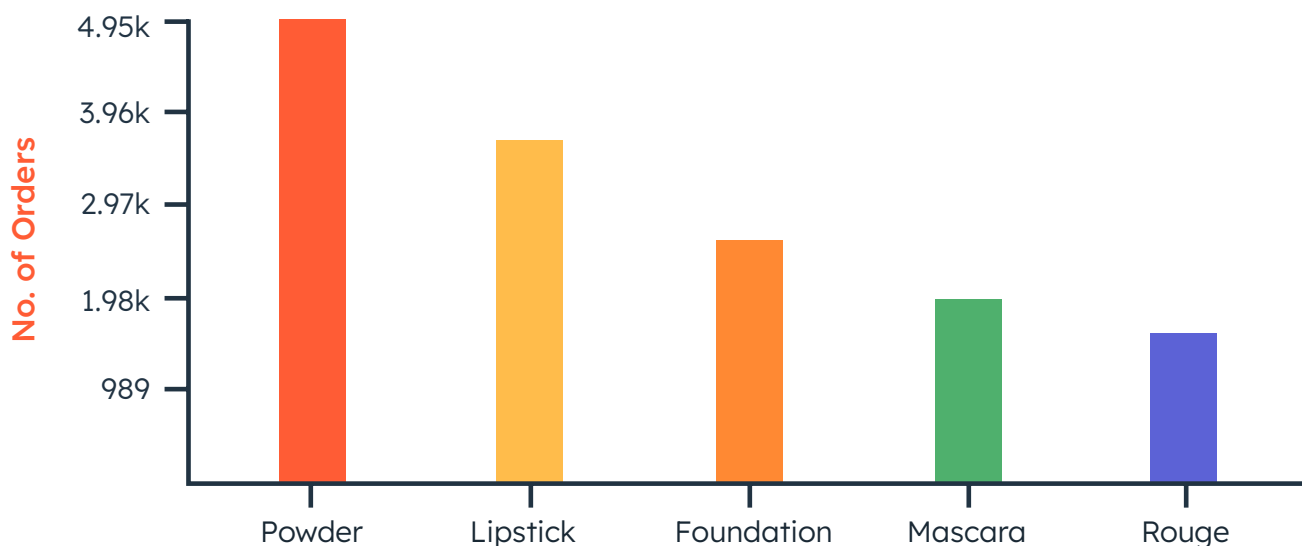
You can determine the median of the data set by putting your variables in order from least to greatest. The median is the value directly in the middle of the data set.

The mode simply refers to the value that is the most common. Continuing with the customers' age example, if twelve of twenty customers are 26 years old and the other eight vary in age, then 26 is the mode because it is the most common value.

Mean, median, and mode all describe the middle of the data set, but in different ways. The range, however, represents the span between the lowest and highest values. You can easily find the range of a data set by subtracting the lowest value from the highest.
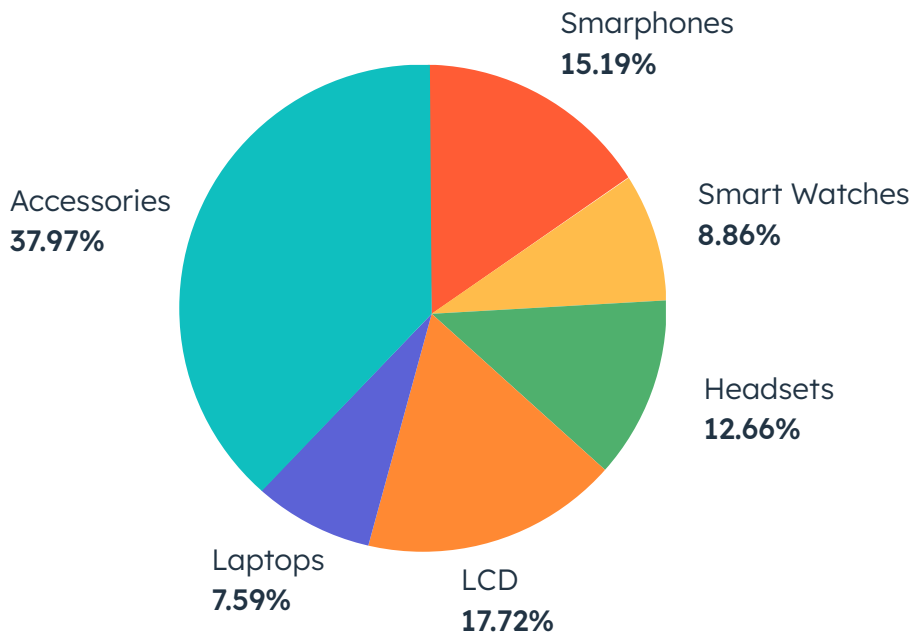
Understanding the mean, median, mode, and range of your data set is helpful, as it will provide the basis for what your company can and should expect. However, this is not the only information you can gather through exploratory data analysis. You'll want to look at the outliers or data points that don't necessarily align with the rest and calculate the standard deviation to determine how much your data points differ from the average. This is helpful because it will give you a solid understanding of your client base in the case of the age example.
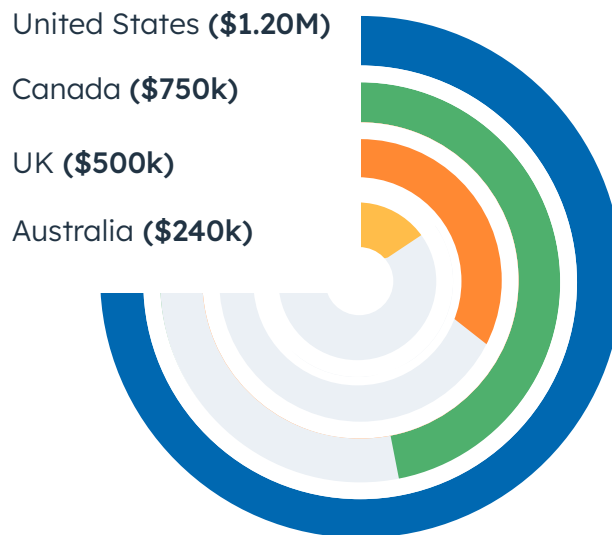
Beauty Products Sales Order Analysis

Image Source

Unit Sold by product category

Smarphones
15.19%

Smart Watches
8.86%

Accessories
37.97%

Headsets
12.66%

Laptops
7.59%

LCD
17.72%

Budget Consumption

Image Source

United States **($1.20M)**

Canada **($750k)**

UK **($500k)**

Australia **($240k)**

Looking at and graphing your data set for exploratory data analysis can also help understand correlation and causation. Studying the correlation and causation of your data set can help you understand which variables need to be in place for something else to occur. For example, if you are looking at data related to new customer subscriptions, you might notice more signups due to a one-day sale.

If exploratory data analysis is completed correctly, you'll likely have more questions than answers. Use the results of your exploratory data analysis to help form hypotheses for further analysis of your data.
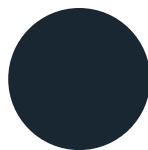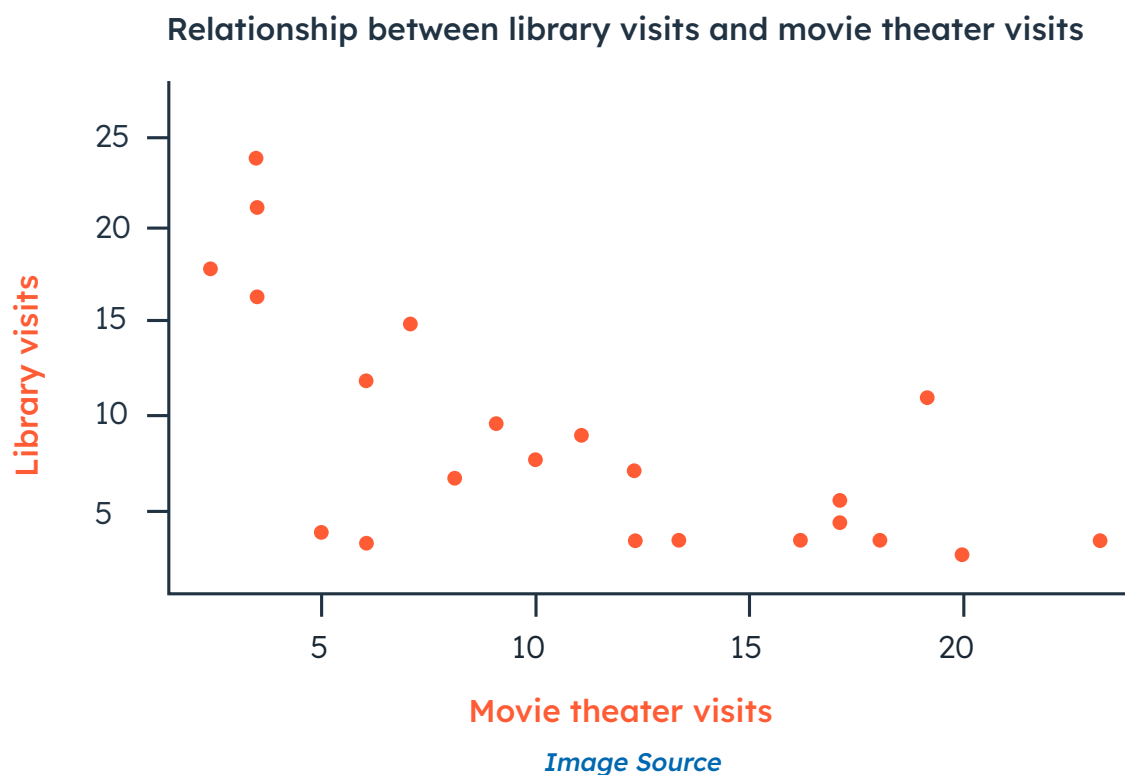
Descriptive data analysis

If exploratory data analysis is the first step of data analytics, descriptive data analysis is the second. If you've been in business for any length of time, you've seen the results of descriptive analytics in actions. Year-over-year increases and month-to-month revenue changes are examples of descriptive analysis results.

Descriptive analysis aims to answer the question, "What happened?" This kind of analysis is helpful because it uncovers patterns and trends hidden in historical data. It's important to note that results from this kind of analysis should not be used to predict future outcomes (predictions and forecasts are made in a different type of analysis that we'll cover later). Instead, descriptive analysis is designed to help make sense of past operations so we can understand current business operation models.

When conducting descriptive analysis, it is necessary to complete all of the same steps that you would do for exploratory data analysis. (Do you see why exploratory data analysis is the first step? It's the foundation of data analytics!) You'll need to take the necessary steps to gather your data from internal or external sources and clean it to ensure it is usable. However, before you visualize your data, take a moment to explore it.

Data exploration is a vital step of descriptive analysis. This is part of the process where you will plug your data into a spreadsheet (if it's not already in one), run statistical equations, or review it for its apparent characteristics, like trends or patterns. It's helpful to use tools, like artificial intelligence or built-in software for your spreadsheet, to help you analyze your data. Then, when you have a solid understanding of the data, you can visualize it, summarize it, and present your findings to your team. Hopefully, with your team members' input, you can begin to interpret what the data is telling you.

Relationship between library visits and movie theater visits

Image Source

With the various tools available, descriptive data analysis is a simple process. If you've taken the time to ensure your data is good, your analysis's results should accurately describe your business's past metrics. Descriptive data analysis is helpful, too, because you can keep track of key performance indicators or KPIs.

The downside to descriptive data analytics is that it does not answer "why?" It just describes what has happened. However, you can use the results to understand what and what is not working for your company. Many stakeholders use the results of descriptive data analytics to help determine what to do with their investments in your company, as this kind of analysis is known to reveal red or green flags. This could be problematic if you have slightly less-than-perfect numbers and skittish stakeholders.

Don't let the stakeholder's interest dissuade you from conducting descriptive data analysis– this analysis is necessary for any business. It is helpful to know and understand metrics like year-on-year growth, sales revenue and income reporting, shipping logistics, and sales trends. You've likely seen descriptive analytics in action, too, in other ways. Think social media engagement reporting and web traffic analysis. These are all data points you can gather from descriptive analytics to help you understand what variables were in place for your current business operations to exist.

If you want to use data analytics to get an idea of potential data projections and forecasts, you'll want to implement predictive data analysis. Let's take a look at it now.

HubSpot

Predictive data analysis

Predictive data analysis is an advanced technology that uses data, algorithms, machine learning, and deep learning to study a data set and predict future events or outcomes based on historical data. Weather forecasters use predictive analytics regularly to help forecast storms and their projected paths. If you've ever wondered if you can use the same technology for your company, the answer is a resounding yes.

Before we get into how to use predictive analytics and its benefits, let's take a minute to review machine learning and deep learning. Machine learning is an artificial intelligence technology that uses algorithms and models to make predictions based on the collected data. Depending on the type of machine learning you use, you may need to program certain algorithms for your specific data set. Some machine learning technologies do not need to be programmed and can run as is.

Deep learning is a machine learning type that processes data similarly to how a human brain processes information. This type of learning uses neural networks, or connected neurons that resemble the brain, to recognize complicated patterns and trends that might have been missed during descriptive data analysis. Deep learning can review text, pictures, video, or sounds to make predictions and provide valuable insight.

Think of predictive data analysis as a crystal ball. It combines machine learning and deep learning to analyze patterns and trends in a data set, allowing you and your team to gain insight into potential outcomes if you change or manipulate variables.

HubSpot
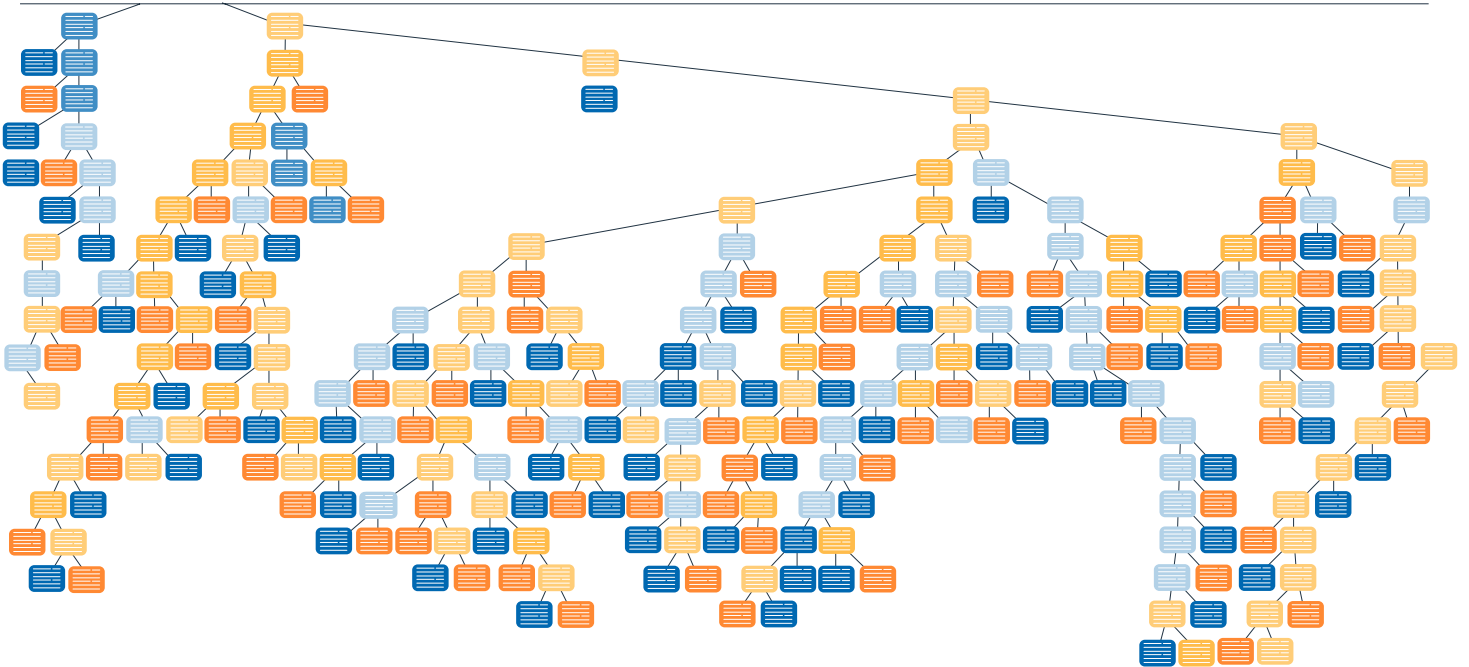
Image Source

It's important to understand that you shouldn't feed data directly into the algorithm without cleaning it. Missing information can have a significant outcome on the accuracy of your predictions. You wouldn't want to use predictions made on misleading information as it could negatively impact your business, defeating the whole purpose of using predictive analytics in the first place.

Data fed into algorithms for predictive analysis must also be separated into two groups: a testing group and a training group. The testing group should contain as much information about your data set as possible.

For example, let's say you own a restaurant and notice a slight uptick in soup sales on cloudy or rainy days. Because this is purely anecdotal, you want to be sure of your findings and decide to use predictive data analytics to estimate future sales. To do this, you should provide the algorithm with the number of bowls of soup sold and the weather conditions for a set time, including sunny and cloudy or rainy days. Because you already know how many bowls of soup were sold on cloudy days, you should be able to run the analysis on the training set and compare results with your true historical data.

Once the algorithm is trained on your data, you can feed new data into the algorithm, like the following week's weather forecast, and get an idea of how many bowls of soup you might sell in the next week. This lowers the risk of making extra soup that doesn't sell because now you have a fairly accurate prediction of projected sales based on your past numbers.

Lowering and mitigating risks is just one example of how predictive data analysis can help your business. Predictive data analysis can do much more than just project potential sales; it works best with real-time data. Many companies use this kind of analysis for customer retention. Predictive data analysis can help pinpoint potential churn when run with real-time data. With the analysis results, you and your team can take the appropriate measures to stop churn before customers reach that point.

If you are an ecommerce business, predictive analytics can help you recommend new products or services to your customers. Simply provide the algorithm with your customers' past behaviors and purchases. The algorithm will match your clients to products or services based on what other customers with similar behaviors bought in the past. It can also help prevent fraud by detecting suspicious user activity in your operation systems, thus keeping your data secure.

While predictive data analysis is suitable for forecasting, using the results to change your business operations is always a risk because there is a chance variables could change, or an unexpected hiccup could occur. Prescriptive data analytics considers all the likelihood of variables changing and can help make the best recommendations based on your data.

Prescriptive data analysis

If you've ever wanted someone to help you make a business decision, you need to consider using prescriptive data analysis. Unlike predictive analytics, which answers the question "what could happen?" prescriptive data analytics helps you understand what you should do and the outcomes you would face if you followed its recommendation. Prescriptive data analytics is the most advanced stage of data analytics, helping you take the guesswork out of your decisions.

Prescriptive data analytics is complex. First, you'll need to define the question you want to understand. Then, you'll need to link the AI-powered algorithm with your data storage system. This kind of analysis requires continuous historical, real-time, and internal and external data to give you the most accurate outcomes.

HubSpot
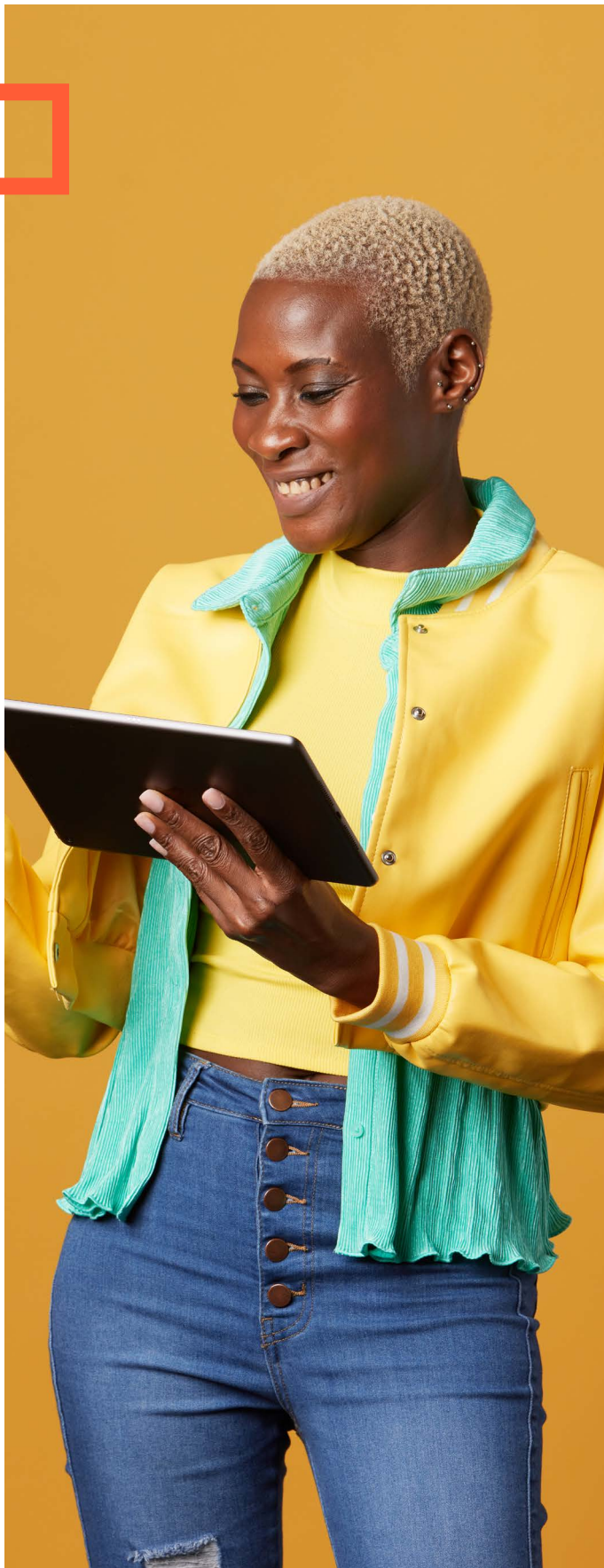
And unless you are comfortable building your own models, you'll likely need an analyst or a technician with AI and algorithm experience to get your models up and running. This person can also nail down any necessary tweaks before deploying any models.

Could a human sift through the data and help you decide what to do? Theoretically, sure. But with large, never-ending data sets, your analyst will waste time sifting through data, potentially miss something important, and give biased recommendations.

Prescriptive analytics removes human bias and the temptation to make decisions based on human intuition. The results of prescriptive data analysis take all potential scenarios based on all relevant variables into consideration and offer the best recommendations and potential outcomes, even the worst-case scenarios. This is helpful because it gives you ample time to create a contingency plan to combat adverse outcomes.

Decision support and decision automation are two of the most significant benefits of prescriptive data analysis. While the results of prescriptive analytics are backed by research and data, you should always have a human review of the results before implementing any of the recommendations. As brilliant as artificial intelligence is, it does not outweigh the opinions of those running the business.

Popular Data Analysis Methods

No matter which type of analysis you choose for your data, you must analyze the results to understand them. There are four popular data analysis methods to help you understand and make sense of your analysis results.

Regression analysis

Regression analysis is a statistical model that depicts the relationship between two variables, an independent variable, and a dependent variable. Regression analysis models are often considered the "go-to method" for data analytics because they explain the relationship between the dependent and independent variables. Plus, we can use it to predict future sales based on our historical data.

Regression analysis models give you an idea of what's happening to reduce the likelihood of assumption. It's essential that you choose independent variables that matter; otherwise, your models will be filled with insignificant data points. So, be as specific as possible when determining your independent variables.

The mathematical equation that best represents regression analysis is

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Let's define what each of those symbols means:
   Y: the dependent variable that you want to predict
   X: the independent variable, or the known variable
   B_0: the intercept value, or the value of Y when X is 0
   B_1: the coefficient of the X variable
 $Ɛ$: The error term, or how much error you should expect in your data

This looks like a lot of math and if math isn't your strong suit, don't worry. That's why you've hired data analysts and used statistical programs. It's helpful to understand these terms, though, as it will give you a better understanding of model predictions. Let's look at a real-world example to understand better the equation and how to graph the data.

Let's look at the soup sales again. Except this time, as a business owner, you want to explore if the time of day impacts the number of soup sales. You love to have a bowl of soup at lunch, and you want to assume that your customers do, too. However, you paid attention in stats class and know you should never assume without looking at the data first.

So, you first collect several weeks' worth of data and track the number of soup sales throughout the day. For example, one particular day, you might find that you do not sell any soup the first hour your restaurant is open. But, in the fifth hour, you sell two bowls of soup. Track this data across several weeks before running a regression analysis model. As always, the most accurate data will produce the best results. Bad data will lead to an inaccurate analysis.

In this example, the number of soup sales is the dependent variable (or Y). It's called the dependent variable because it depends on the value of the independent variable (or X), which is the time of day.
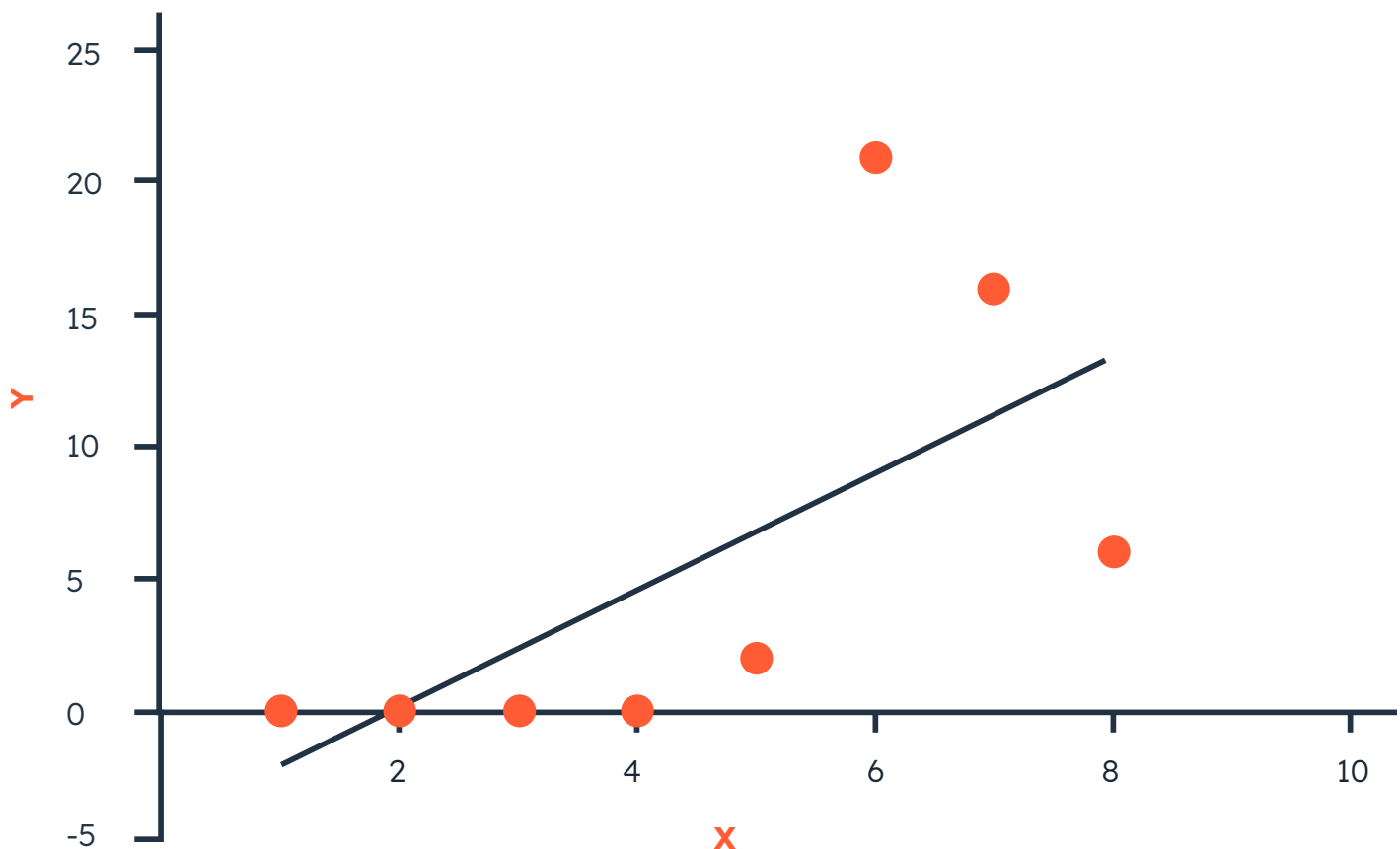
Image Source

Armed with weeks worth of historical data, including soup sales and the time of day, you should plot those points on a graph. The graph's x-axis, or the horizontal axis, represents the number of soup sales, and the y-axis represents the time of day.

Now that you have a visual representation of your sales look to see if there is a linear pattern in your data. This graph shows a positive relationship between the time of day and soup sales. Your data analyst or the program you're using can determine the regression line. The regression line shows the line of best fit for your data. It's important to remember that there may be a small chance of error in the regression line. The error term mentioned above acts as an extra layer of insurance for estimating sales. The smaller the error term, the more you can rely on the accuracy of the estimation.

Linear regression models, like the one pictured above, are among the most common regression analysis models. But if a linear regression model doesn't seem to represent your data fully, there are other types of models to run. Those are:

⭐ Nonlinear regression models:

These models do not depict a clear relationship between independent and dependent variables.

⭐ Multiple regression models:

These models are just like linear regression models; instead of focusing on one independent variable, multiple independent variables are added to the graph. This makes it easier to see the impact of independent variables on your dependent variable.

⭐ Polynomial regression models:

This model represents the relationship between independent and dependent variables that are not necessarily linear. Instead, other equations, like the quadratic formula or cubic equation, best fit the line of regression.

⭐ Logistic regression models:

If your dependent variables are binary, meaning other "yes or no" or "0 or 1", a logistic regression model is the best model to fit your data. It helps explain the probability of something occurring based on certain factors, or independent variables.

HubSpot

There are a few important things we must consider when it comes to regression analysis models. The first is that correlation does not always mean causation. In the example of the soup sales, the time of day does not always determine the number of soup sales. Although the time of day definitely influences when customers are most likely to have soup, you also have to consider the hunger levels of customers and the type of soup you are offering that day.

The second thing to remember is that you should be as specific as possible when choosing your independent variables. Too broad of an independent variable will result in inconsistent or useless results. The more accurate the data is, the better chance of an accurate regression analysis.

Cluster analysis

The cluster analysis method, or clustering, involves grouping data points based on similarities. This means that your data set might not have any target values, but with the help of algorithms, you can sort your data into groups that make sense.

It's helpful to think of a box of used crayons to understand how clustering works. Each crayon is a different color; some may be broken, others half-used, or some brand new. The point is no two crayons are the same. To group these crayons, you might decide to sort them by colors, meaning every shade of green, including forest green, lime green, and yellow-green, go into the same pile. Or, you could sort them based on their use, so you'll have a cluster of new crayons, used and broken crayons.
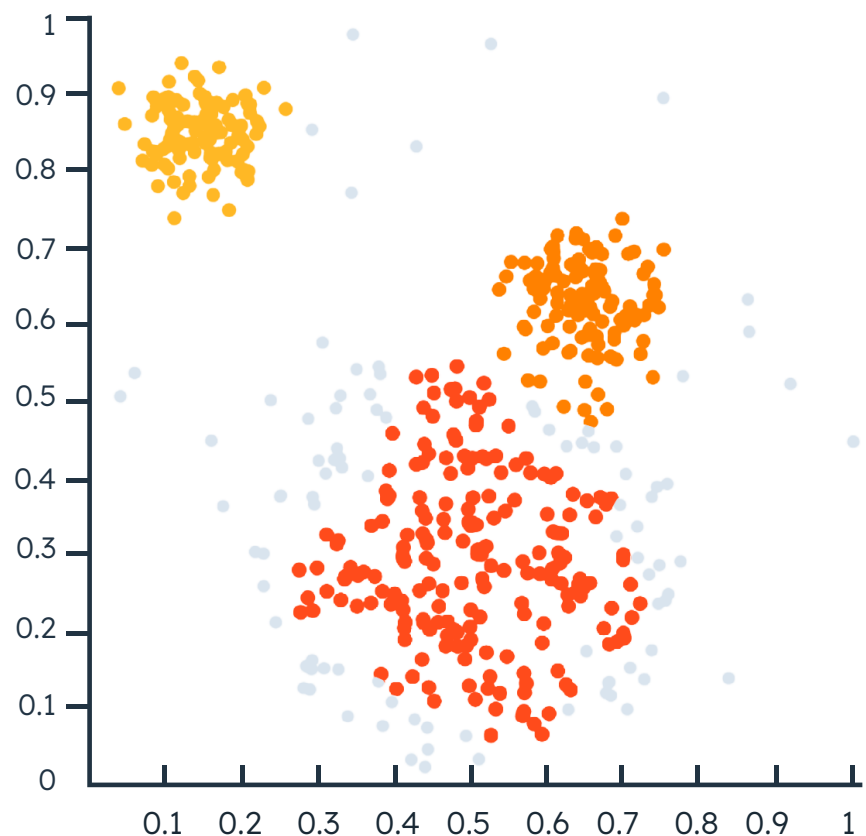
Image Source

The same concept can be applied to your data sets. However, instead of manually sorting your data to look for often hidden similarities, following various cluster models and using the accompanying algorithm is helpful.

There are six different methods of clustering. Let's take a look at each of them and their algorithm.

Connectivity-based clustering

Connectivity-based clustering, also known as hierarchical clustering, centers around the idea that each piece of data is connected to its neighbor based on its relationship, or proximal distance, to its neighbor. If you use an algorithm to compute a connectivity-based cluster, your results will be shown in a dendrogram.

Image Source
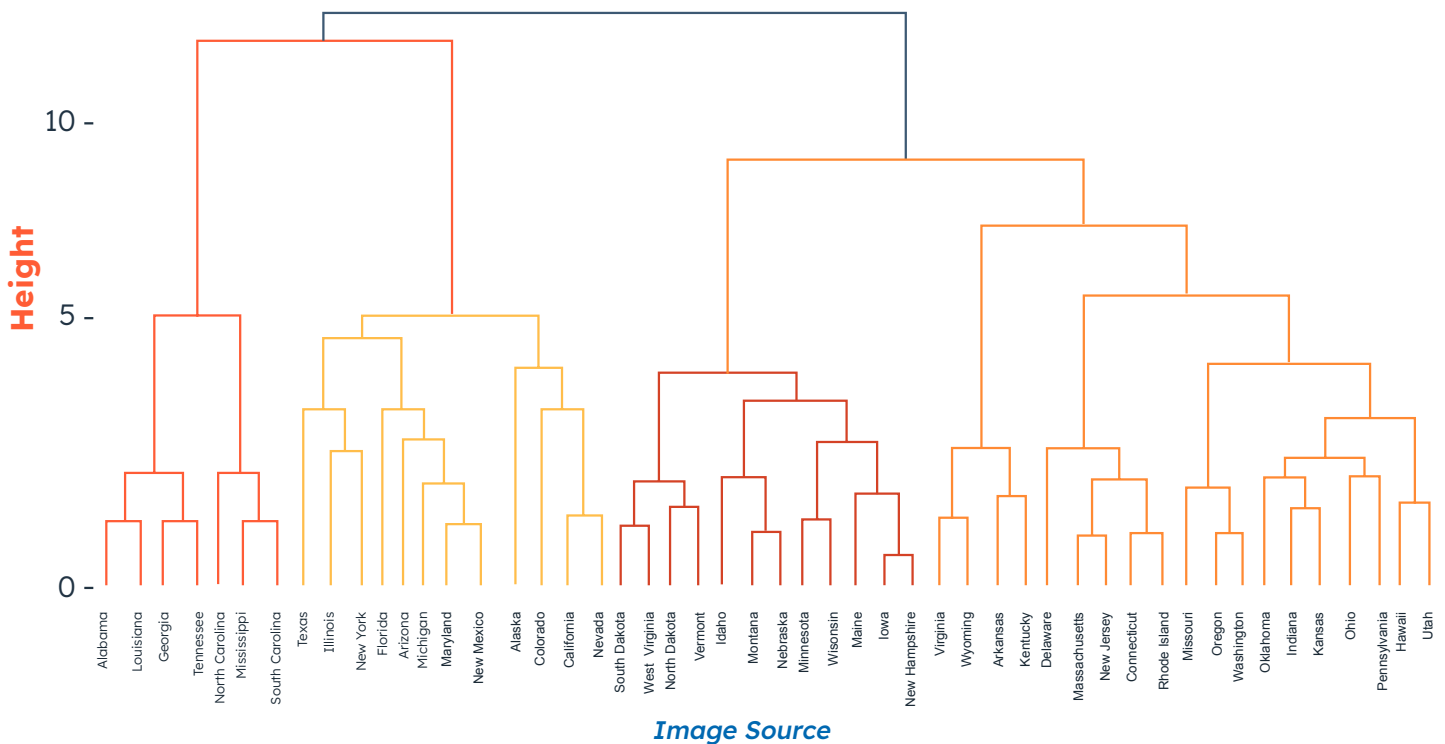
Looking at the above example, you'll notice that the overall data is split into several different groups. Each data point within the group is then divided into another similar subgroup. The x-axis describes the clusters that do not merge, while the y-axis represents the distance between each cluster.

The rule of thumb for connectivity-based clustering is that if the data is similar to an established cluster, it is sorted into that group. If dissimilar, it goes elsewhere or farther away from the established cluster and can form its cluster if needed.

There are two main approaches to connectivity-based clustering: divisive and agglomerative approaches. The divisive approach filters data from the top down. This means all data is filed into one cluster and sorted into smaller clusters based on specific termination criteria. Agglomerative approaches, on the other hand, assume each data point is an individual cluster. Once the data is established as an individual, it is grouped into a cluster it most closely resembles. In other words, it sorts data from the bottom up.

The algorithm you should use for this kind of clustering is called the BIRCH algorithm, or Balanced Iterative Reducing and Clustering Using Hierarchies. Running this algorithm is quick and efficient and works best with large data sets. Unlike other algorithms we'll discuss later, this algorithm only makes one pass through the data and needs a few set parameters to run well. Before running the algorithm, define the CF tree and its threshold. A CF tree consists of each subgroup or leaf cluster, and each leaf cluster can only get as big as the threshold allows. A new leaf is formed once the threshold reaches the maximum number of data points.

Centroid Clustering

The Centroid Clustering method is the easiest of all clustering methods, making it the most commonly used clustering technique. The most difficult part of this clustering model is choosing the number of clusters, or k, you want your data set divided into and assigning those clusters a vector value. Vectr value simply refers to a collection of values within a group. After those parameters are set, your data is sorted into the given set of clusters based on how closely it matches the vector value.
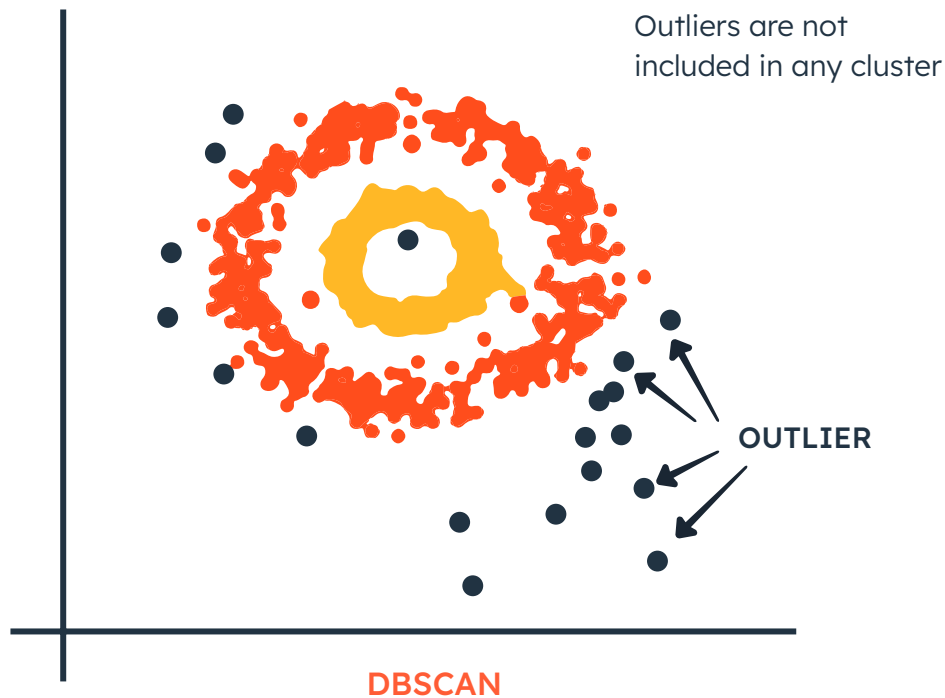
Ideal Clustering

Image Source

This clustering method relies heavily on the K-means clustering algorithm. This algorithm sorts data into groups according to the predefined k-cluster. Each time the algorithm is run, the center value, or the centroid, of k may change. The algorithm is run enough times so that an optimal k is discovered. Optimal k values should be the average of all of the centroid points.

Density-based Clustering

Unlike other clustering models, density-based clustering considers density concentration over the distance between points. These models look noticeably different as they pair data points into areas of high or low concentration. This model also considers that all data sets will contain noise and outliers. Instead of throwing the noise and outliers out, they are plotted in the model and recognized as low-concentration areas. And, unlike other types of modeling, it does not conform to a certain geometrical shape. The data determines the shape of the graph.
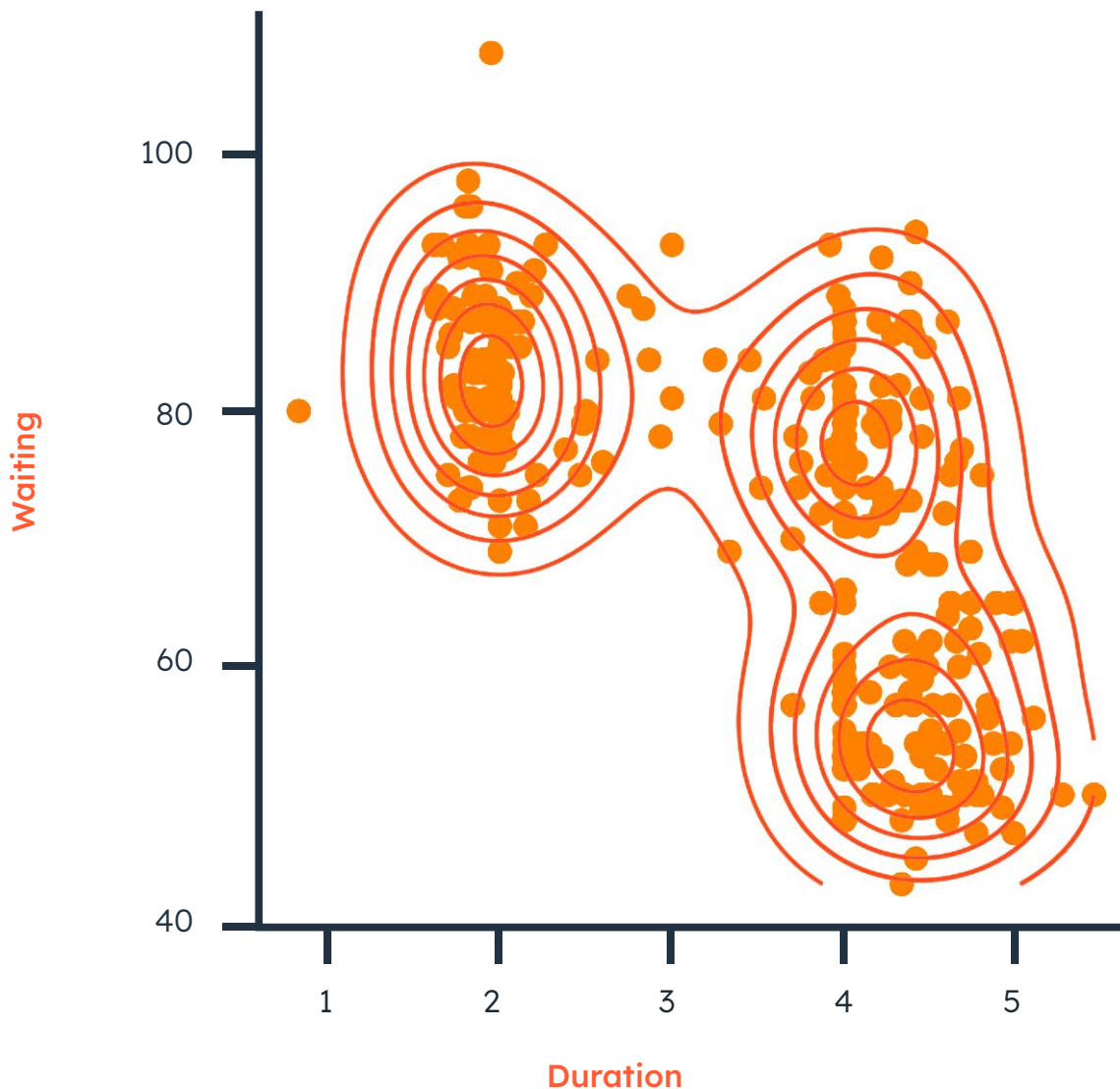

Image Source

DBSCAN, or Density-based Spatial Clustering of Applications with Noise, is the most efficient algorithm to use with this model. This algorithm can find hidden similarities within data sets. Unlike the BIRCH algorithm, which is a "one and done" kind of algorithm, this algorithm combs the data set until each piece of data is correctly classified as a cluster or noise.

Distribution-Based Clustering

The distribution-based clustering model takes an entirely different approach to clustering compared to the previous models we have discussed. This model categorizes data into groups based on the likelihood of a piece of data belonging to that group. Distribution-based clustering works if there are predetermined central points. Once those points are identified, it is put into that cluster if the data looks like it might belong.
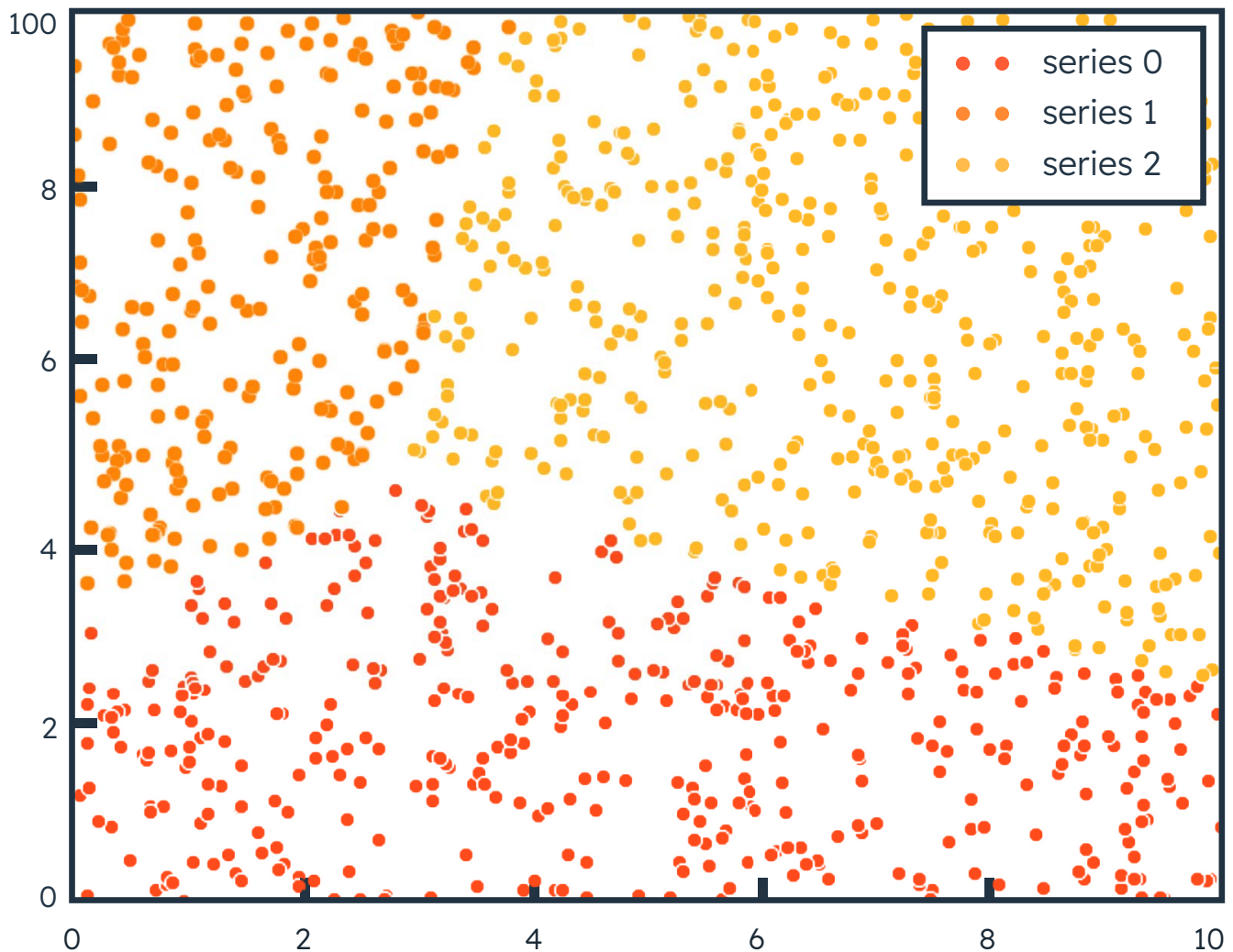

Image Source

There are several algorithms that deal with distribution-based clustering, including K-means clustering and the DBSCAN.

Fuzzy Clustering

What do you do if your dataset can be classified into multiple clusters? That's where fuzzy clustering comes in. Fuzzy clustering describes a model where data points are categorized first based on similarities to the central point. On the second test run, data that has not yet been categorized is grouped based on the probability of belonging.

Random points classified according to known centers

Image Source

The most popular algorithm associated with fuzzy clustering is called the Fuzzy C-Means algorithm. This algorithm assigns each data point a membership value representing the standard deviation from the central value of the cluster.

Constraint-based Clustering

Algorithms and clustering methods are great for helping your analysts identify hidden patterns within a data set. However, there are times when your analysts might already expect how the data should be sorted. In these cases, the constraint-based clustering model works best. This model allows the analyst to set parameters for the data, including the number of clusters, the number of allowed data points in the cluster, and its allowed dimensions.
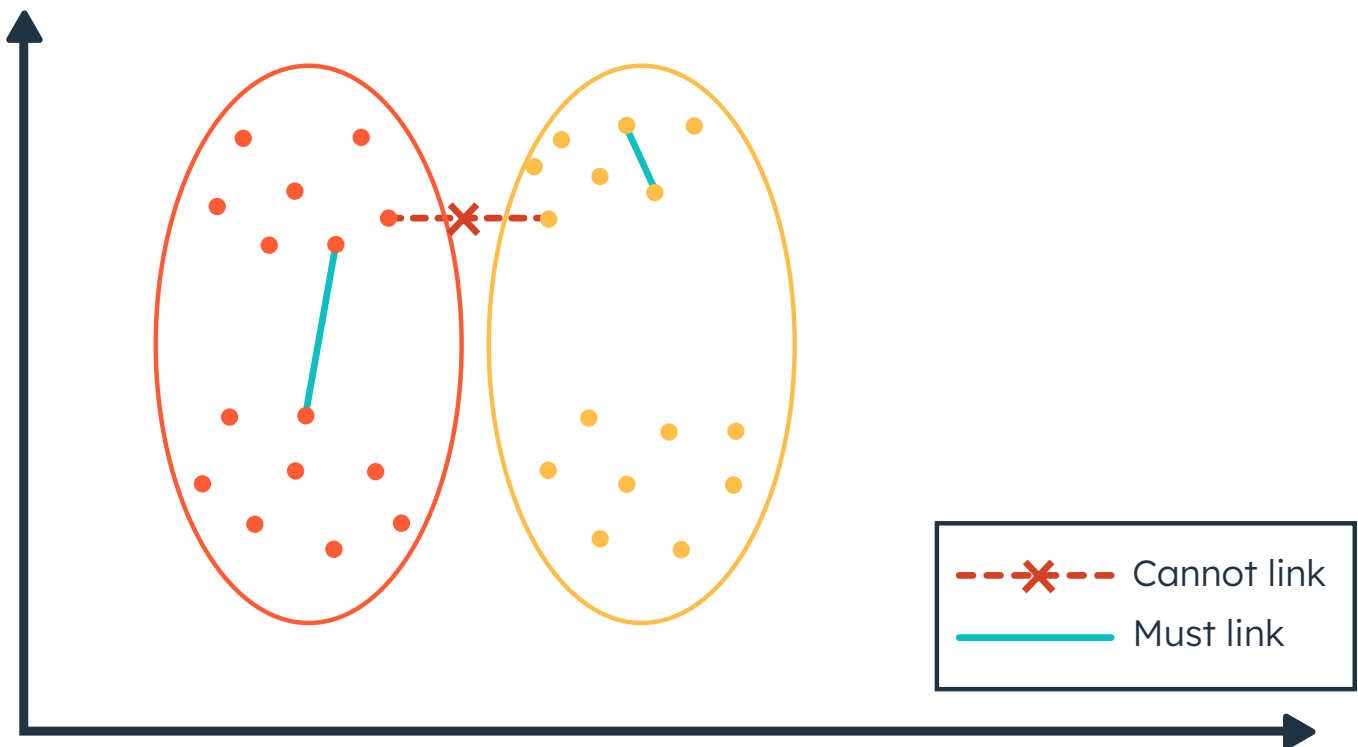
Image Source

Several algorithms work for constraint-based clustering. It's important to note that the algorithms mentioned in this section are not an exhaustive list of possible algorithms, and various algorithms work for multiple types of data modeling. If you use any software to help plot your data, your software will likely suggest programmed algorithms to help you best sort your data points, making it much more efficient for your analysts.
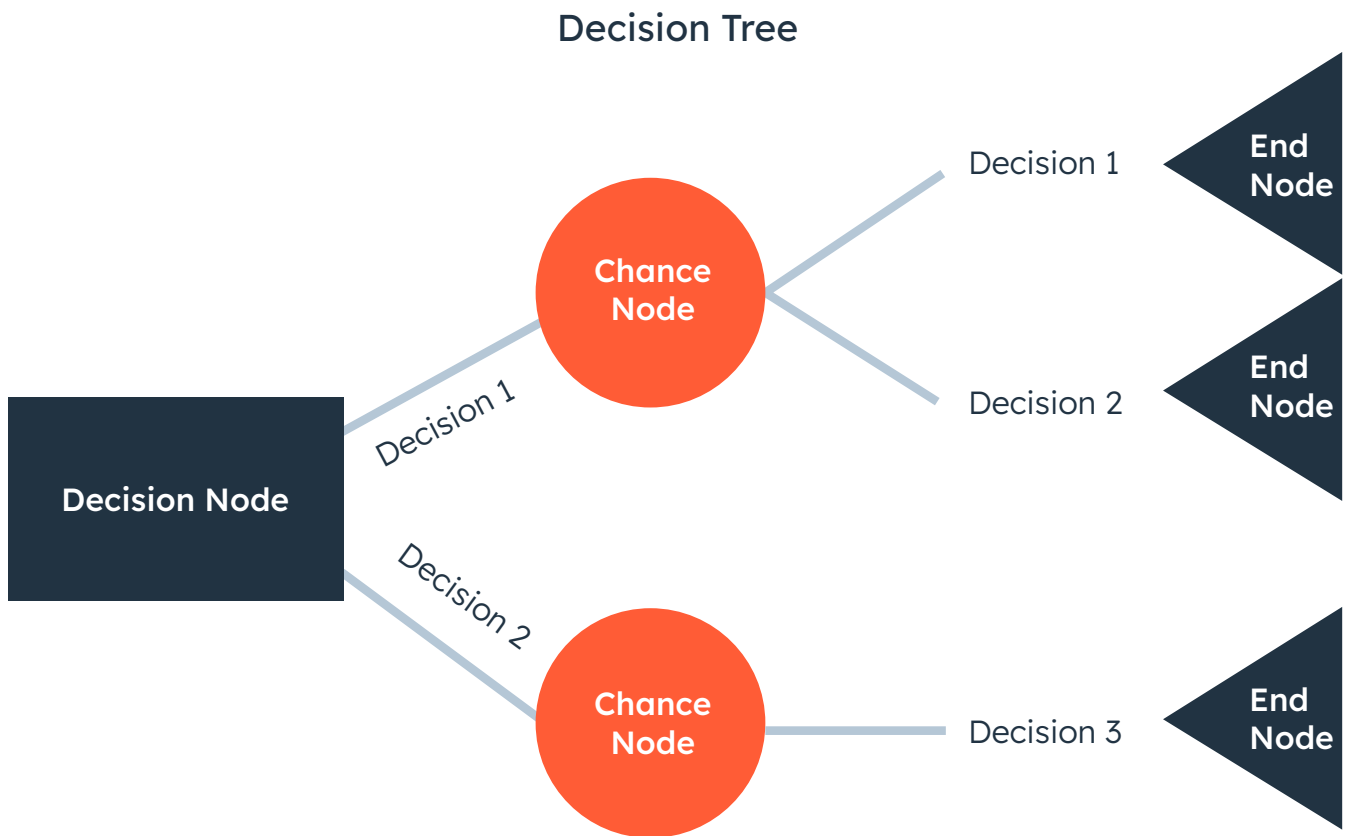
Decision tree analysis

While seeking answers in a crystal ball is, unfortunately, fiction, with the help of software, you can use decision tree models to manipulate your data to help you find answers and outcomes. These models use supervised machine learning to compute numerical or categorical data to outline all possible options and their outcomes, meaning you have the best shot of taking the most appropriate course of action for your business based on your data.

"supervised machine learning" means the model needs to be trained and tested on your data. Based on the outcomes of the training set, the algorithm can make predictions based on the rest of your data. It's important to note that these models do not tolerate large sets of data well, and outliers and noise can throw off the outcomes. While you do not necessarily need to clean your data first, you might want to look it over and throw out missing or incorrect data. This will help produce more accurate results later on.

Simply put, a decision tree model consists of one question that points to multiple options. Normally, the questions have binary answers, like yes or no. As you move further down the tree, more questions and options may present themselves. The beauty of a decision tree is you can weigh all of the outcomes against the risks and rewards.

There are two decision tree types: categorical variable decision trees and continuous variable decision trees. A categorical variable decision tree is a simple model. It categorizes data based on the question provided. Continuous variable decision trees, though, do not always provide a simple answer. These models are called regression trees because the outcome depends on previous and sometimes multiple variables.

Like a living tree, you can cut out branches of your decision tree based on inaccurate data (like noise or outliers). Decision trees are easy to understand, but large data sets can quickly become complicated. Therefore, you must ensure an appropriate sample size before running a decision tree model on your data.

Decision Tree

Image Source

Decision tree models are great for evaluating business outcomes, and you can also employ them in your systems to help suggest customer recommendations.

Time series analysis

Time series analysis, as the name suggests, is an analytical technique used to understand data collected over a series of specific intervals. Many data analysts run time series analyses to detect and understand how time affects seasonality. This kind of analysis is also helpful to identify patterns, trends, and behaviors and make forecasted predictions.

This analytical method assumes that time is an independent variable and all other variables, regardless of what they are, depend upon the continuation of time. Large amounts of data are collected over a series of evenly spaced intervals to get the most accurate results from this kind of analysis. The massive volume of data helps to ensure the consistency and reliability of your results, and it also cuts through noise and ensures any detected trends or patterns are not influenced by outliers.

Time series data analysis is a simplistic idea that can quickly become complicated, depending on how you want to run your data and the metrics you are looking for in your dataset. This type of analysis is broken into two important classifications: stock time series data and flow time series data. Think of stock time series data as a snapshot of the collected data. The data within that snapshot is measured and assessed for its patterns and trends. While stock time series data is only one period, flow time series data refers to a continuous data flow over a predetermined time.
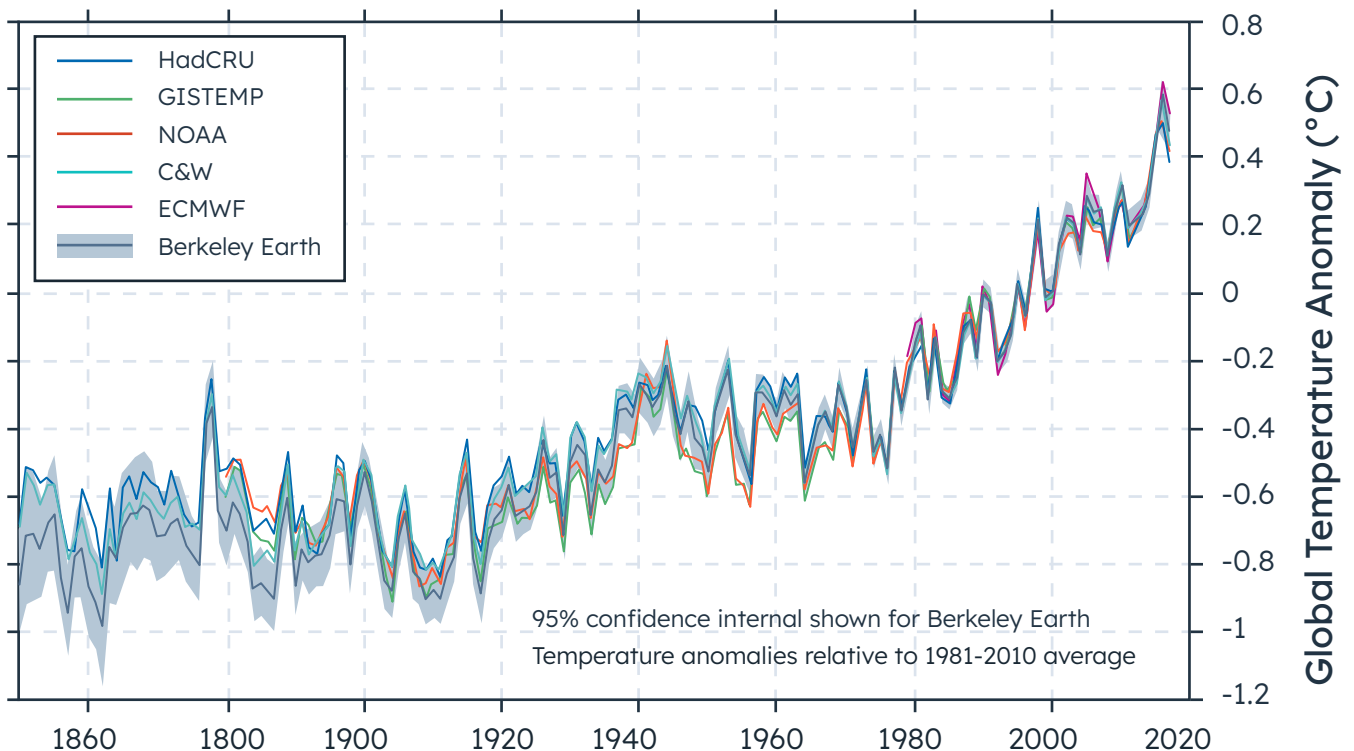

95% confidence internal shown for Berkeley Earth
Temperature anomalies relative to 1981-2010 average

Image Source

There are also general variations in this kind of analysis, too. For instance, if you want to look at notable trends, you'll want to run a functional analysis. If you wanted to see if the pattern flowed in one direction, you would run a trend analysis. And, if you wanted to see if the data is consistent on a seasonal basis, you'd run a seasonal variation analysis.

No matter which kind of analysis you choose, a few key indicators concerning possible patterns are essential to understand. When a trend is revealed, the data will most likely follow a specific pattern, either in an increasing or decreasing direction. Some patterns reveal seasonality, which means the pattern is regular and repeats at specified intervals, like days or weeks. The data could also reveal a cyclic pattern, meaning the fluctuations in the data do not follow a designated time. And finally, the pattern could be irregular, completely random, and unpredictable.

There are several benefits to running a time series analysis. Besides revealing patterns and trends dependent on time, this analysis often provides better data visualization. Depending on the analysis model you choose, either the Box Jenkins ARIMA Model or the Box Jenkins Multivariate Model, you can plot a singular variable or multiple variables. Be careful, though, with the temptation to plot all your known variables at once, as too many variables can quickly become too complicated to understand, making it difficult to spot any trend.

With the help of algorithms and machine learning, there are multiple ways to categorize and explore your data. You can use machine learning for:

- **Classification:** Or identifying and classifying data according to its trends and behaviors

- **Curve Fitting:** Curve fitting refers to plotting data along a curve to make comparisons and understand the relationships in the dataset.

- **Descriptive analysis:** Reviewing the historical data to make connections within the dataset.

- **Explanatory analysis:** Explanatory analysis means that data is used to explain why an event occurred.

- **Exploratory analysis:** Exploring the dataset to predict outcomes based on the historical data.

- **Forecasting:** Forecasting involves studying the dataset to predict future events.

- **Intervention:** Or determining how a single event changes the outcome of the data.

- **Segmentation:** Segmentation means that data is split into digestible segments to understand its underlying properties and the data's source.

Time series analysis is useful and can be used in various business functions. Demand forecasting, financial analysis, resource and inventory management, and risk management can all be determined with this analysis.

Now that you understand the different types of analysis you can do on your data, let's look at some best practices to ensure the best data analytics results.

HubSpot
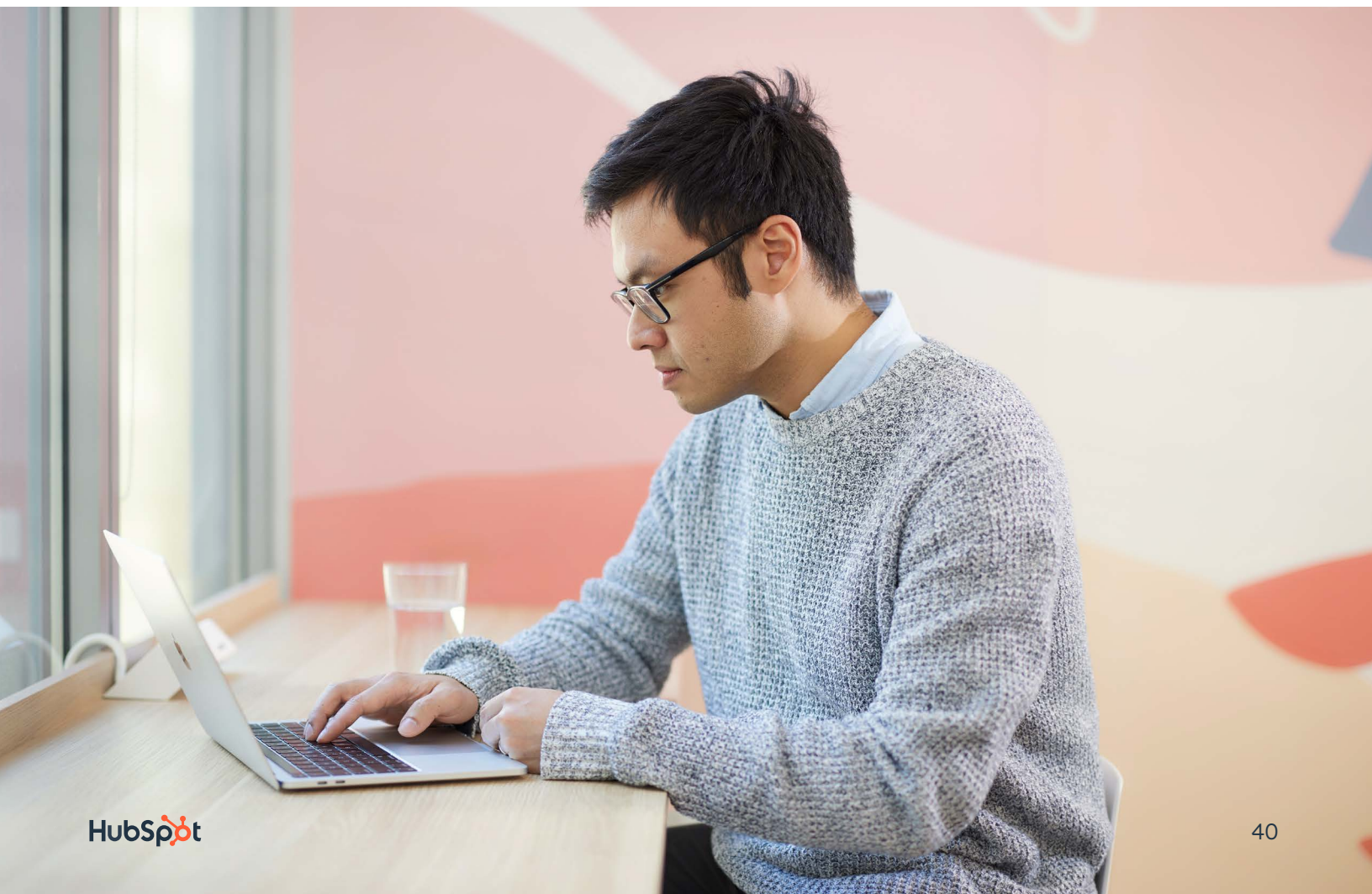
Best Practices in Data Analytics

As we've learned, there are several moving parts surrounding data analytics. Data analytics is a huge undertaking and an even bigger responsibility. Data points are more than just numbers (or texts, images, and sounds). Most of the time, the data represent individuals, and it's important to treat your data carefully. Implementing best practices for data analytics is a smart choice.

Let's look at how you can establish and follow best practices for each facet of data analytics.

Establishing Data-Driven Decision-Making Processes

The goal of conducting data analytics is to help your company better understand key metrics indicated by the data. Those metrics can help you and your team make better choices and decisions for the direction of your company. However, before your company's leaders make any decisions, it's best to establish and define some best practices for the decision-making process. This can help reduce confusion and conflict later on.

⭐ Involve Your Stakeholders

It goes without saying your stakeholders are essential to your company's success. Those stakeholders should be included and involved in the decision-making process, whether they are investors, directors, or key team members. Make a list of important stakeholders and do what you can to ensure their opinions and concerns are heard in discussions regarding important data-driven decisions.

⭐ Define Roles and Responsibilities

Data analytics is certainly a cross-department collaboration. Although several departments have a role in how data is collected, stored, and processed, not even a team needs a seat at the table when company decisions are made. This can lead to too many kitchen cooks and not enough chefs. So, before you begin with your data analytics endeavor, take a moment to clearly define and communicate each department and team member's roles and responsibilities. This allows for consistency, transparency, and understanding. Plus, open communication makes teamwork easier.

HubSpot

⭐ Define the Context of a Decision

It's easy to make a decision on a whim or at the drop of a hat. But when decisions are made under these circumstances, there usually is not a lot of thought regarding the impact of the decision or its possible consequences. You should avoid this situation in the business world because it could be a costly mistake.

Data analytics helps you understand the risks and possible solutions to some of your business questions. However, you and the team must take the time to understand the decision, the relevant data, the analysis results, and the criteria for making a decision. Clearly defining the context of a decision allows for thorough analysis and understanding of the impact before executing new business initiatives.

⭐ Document Your Decisions

Record-keeping is necessary for any business, and keeping records concerning any business decisions is important. This document should contain a summary of the initial question, any relevant information concerning the data and how it was processed, an overview of the potential outcomes and risks, a brief detailing the decision made, and a list of key people involved. This might sound like a lot of information for one document, but it is helpful for compliance reasons and for reviewing past decisions in the future.

Defining Clear KPIs

It's tough to check data analysis without understanding what you are looking for and attempting to monitor. This is why it's important to determine key performance indicators or KPIs, before you begin an analysis. Like defining your decision-making process, you should follow a few best practices when choosing your KPIs.

⭐ Choose Relevant KPIs

It can be tempting to track every performance indicator relevant to your business. But the truth is, while most performance indicators have some weight and can tell you something about your company's performance, not everyone is essential to track and understand.

Before running an analysis, take time to brainstorm with your team and make a list of relevant KPIs. Your key metrics should be SMART indicators. SMART means that the metric is specific, measurable, achievable, relevant, and time-specific. You should also consider creating a balance of leading and lagging indicators. Leading indicators will help predict future performance and lagging indicators will help you understand past performance.

Taking relevant KPIs into account and tracking them will help you and your team effectively make more intelligent decisions with your data.

⭐ Conduct Regular Audits

When you first choose your KPIs, it's not uncommon to discover that the metric is less relevant than you thought. And this is okay; it happens. That's why it's essential to conduct regular audits of your KPIs. You and your team might find that there are other, more helpful metrics available.

Ultimately, your KPIs help you make decisions for improved performance. Continually auditing your metrics is a good practice to ensure you have the best data in your hands.

⭐ Clearly Communicate KPIs to Your Teams

Communication is essential for smooth business operations. As a best practice, you should communicate which KPIs you and your team are tracking. That way, everyone is on the same page, and those with the power to influence your KPIs (think sales teams and customer service reps) can better understand what they can do to help improve the numbers. Clear communication also helps your stakeholders to make sense of the metrics.

Applying Data Governance and Ethics Principles

Have you ever thought about the huge responsibility attached to data collection and analysis? Data, particularly how it's collected and what a company does with it, is tricky, and it's why there are entire government departments devoted to regulating data collection practices. Not following government regulations and guidelines can open your business to fines and lawsuits. Save your company from scandal by following best practices regarding data governance and ethics.

⭐ Establish Policies and Guidelines

Obviously, you'll want to make government policies, like GDPR, the center for any governance framework you and your team create. But you should consider going further and enacting your policies for an extra layer of protection.

It's a good idea to involve relevant stakeholders, including lawyers and policy analysts, to help you create your documents. These policies should define what your company plans to do with the data, who can review it, and outline protections for whistleblowers if it occurs. You should also consider enacting a policy to help mitigate bias and ensure fairness.

⭐ Ensure Data Privacy Compliance

Your data is sensitive and often represents individuals. Even though data is easily generated, you need to acquire informed consent from your users before you add their data to your database. Take the time to create an opt-in form that clearly describes your intent for user data and have your users consent to it. It goes without saying, but for any user who declines to consent, their data is off limits, and you should have the appropriate measures to ensure their data does not enter your system.

Again, you'll want to be transparent with your users' private data. Plus, this adds a layer of public accountability for compliance.

⭐ Ethical AI Practices

Artificial intelligence is a smart technology that can quickly become complicated and hard to understand. Not understanding the AI technology you choose for your business operations is a big no-no. If you cannot clearly explain your AI to another person, it can seem like you are hiding a major part of your business practices. Hiding your business practices, whether you intend to or not, is unethical and should be avoided.

Have you heard the phrase, "Explain it to me like I'm five?" This just means explaining a topic on a level a five-year-old can understand. Keep this phrase in mind when choosing an AI technology for your business. Having the ability to clearly explain the technologies you use leads to transparency and trust.

Ensuring Data Security and Privacy

It cannot be stressed enough that the data your company collects is private and confidential. And it goes without saying data encryption should be a top priority. Depending on your business, you may collect information containing medical records, financial statements, or other sensitive data. Data security and privacy are essential to running your business well. Follow the below best practices to help ensure you are protecting your data.

⭐ Collect Only Necessary Information

Let's be honest: there is a lot of information your company could collect, but only a small amount of it is relevant to the success of your company. You can reduce security risks and ensure privacy by only collecting necessary data critical to your business. If you need to collect data from other parties, take the time to ensure the privacy and security of those lists before adding them to your own data warehouse.

⭐ Limit Access to Data

One of the easiest ways to ensure your data is safe is to limit who has access to it. Determine key team members who need access to your data management systems and give necessary permissions to only those members. It's also important that these members are provided security training to refresh their knowledge base of security and privacy practices consistently.

⭐ Create an Incident Response Plan

Unfortunately, there are bad actors out there who are determined to hack data management systems. Hopefully, this will not happen to your organization, but you must be prepared if it does. Brainstorm with your team members and create an incident response plan that outlines the exact steps you and your team will take to mitigate any incidents if they occur. Be sure to outline each critical role and their responsibilities for shutting down an incident so there is no confusion if something were to happen. Ensure Regulatory Compliance

Regulatory compliance extends to all facets of data analytics. Like ensuring your data collection methods align with government regulations, you also need to confirm your privacy measures align with government policies. This includes managing the data lifecycle from start to finish. Ensure your data deletion policies are founded on government rules to comply with regulatory standards.

Data Visualization and Data Analytics

If data analytics is your company's bread, data visualization is the butter. Remember how we talked about the different types of analytics and how certain types uncover patterns and trends? That information is usually visualized in a chart or a graph. The term "**data visualization**" refers to how you can display numbers, statistics, and other data in a diagram or graph to make it easier to understand and present.

Download HubSpot's ebook, **_Introduction to Data Visualization._**

Importance and Benefits of Data Visualization

One of the primary purposes of data analytics is to help spot trends and patterns within your business. Any appropriate algorithm can detect these trends, but it's helpful to visualize a trend rather than read about it. This is why data visualization is essential. Creating a chart or a graph can help analysts quickly spot trends, patterns, and outliers.

A chart or a graph can aid your analysts in communicating the data to others, too. Instead of talking about what they've learned by analyzing data, they can quickly and effectively highlight insights on a graph or a chart to help provide a visual aid for their audience. Plus, if they're using this information to present to new stakeholders, they can develop a diagram or a graph in your business colors to impress your audience with your company branding.

If you are using data analytics to help plan or set goals, showing business projections on a graph can be beneficial. A negative or positive trend is easily recognizable. Depending on the direction, the positive or negative movement can help your entire team understand what your goals intend to solve.

Principles of Effective Data Visualisation

Most algorithms and analytical software have built-in features to make graphs or charts. However, there are a few things you can do and a few tweaks you can make to ensure your data is effectively visualized.

⭐ Keep it Clear and Concise

A graph or chart that is overly complicated is not a good visual at all. A complicated visual often leads to more questions than answers. An excellent visual seeks to answer just one specific question and features different colors representing various data segments. It also keeps the data in context, meaning the graphic does not make invalid claims.

When creating a visual, also keep your audience in mind. How your audience will view the data will help you understand the best way to present it.

⭐ Choose the Most Appropriate Visual for Your Data

Effective data visualization relies on choosing the right type of visual to represent your information. There are all kinds of graphs and charts you can use, including, but not limited to, a line graph, bar graph, pie chart, histogram, or scatter plot. While each one essentially performs the same function and is a visual representation, not every graph or chart is an excellent match for your data. For example, it wouldn't make sense to represent revenue growth on a scatter plot. Instead, growth is best represented on a line graph or bar graph.

Because data analysis helps tell your company's story, you should choose the best visual to represent your data accurately.

⭐ Techniques and Tools for Visualizing Data

If you're worried about visualizing data by hand, don't sweat it. You can use numerous techniques and tools to help make data visualization easier for everyone.

For starters, you will need some analyzed data before attempting any type of visualization. Once you have your analyzed data, you can enter it into a spreadsheet, like Google Sheets or Microsoft Excel, and use the built-in functions to create a graph or chart. If you use Excel, grab a copy of our free **Excel Graph Generator Template** to simplify the process. Some software, like your CRM or analytical software, also have built-in features to graph the data it analyzes quickly and easily.

But you have options if you want to make your graphs by hand. Programs like Canva are simple to use. Canva's graphics also feature various customizable graphs and charts. Not only can you make the visual fit your data, but you can also brand it to your company's colors. If you need a heat map, input your data into Hotjar and let the program do the rest.

For an in-depth look at data visualization, check out the free ebook, **An Introduction to Data Visualization.**

HubSpot

Closing

At first glance, analyzing your data can seem daunting. However, with the right data management system, a responsive algorithm, and an appropriate method of analysis, you can dive into the world of data analytics and uncover patterns and trends within your collected data.

With the drive to move towards more data-driven business approaches, implementing data analytics is more important than ever. The results of data analysis can help you make better, more informed business decisions, reduce the risks associated with making a decision, project future performance, and help enhance your customer experience. With data analysis in your hands, there is no telling what your company can achieve!