From issue to action: The six data analysis phases

There are six data analysis phases that will help you make seamless decisions: ask, prepare, process, analyze, share, and act. Keep in mind, these are different from the data life cycle, which describes the changes data goes through over its lifetime. Going through the steps will help you solve all kinds of business problems that you might face on the job.

Step 1: Ask

It's impossible to solve a problem if you don't know what it is. These are some things to consider:

- Define the problem you're trying to solve
- Make sure you fully understand the stakeholder's expectations
- Focus on the actual problem and avoid any distractions
- Collaborate with stakeholders and keep an open line of communication
- Take a step back and see the whole situation in context

Questions to ask yourself in this step:

1. What are my stakeholders saying their problems are?
2. Now that I've identified the issues, how can I help the stakeholders resolve their questions?

Step 2: Prepare

You will decide what data you need to collect in order to answer your questions and how to organize it so that it is useful. You might use your business task to decide:

- What metrics to measure
- Locate data in your database
- Create security measures to protect that data

Questions to ask yourself in this step:

1. What do I need to figure out how to solve this problem?
2. What research do I need to do?

Step 3: Process

Clean data is the best data and you will need to clean up your data to get rid of any possible errors, inaccuracies, or inconsistencies. This might mean:

- Using spreadsheet functions to find incorrectly entered data
- Using SQL functions to check for extra spaces
- Removing repeated entries
- Checking as much as possible for bias in the data

Questions to ask yourself in this step:

1. What data errors or inaccuracies might get in my way of getting the best possible answer to the problem I am trying to solve?
2. How can I clean my data so the information I have is more consistent?

Step 4: Analyze

You will want to think analytically about your data. At this stage, you might sort and format your data to make it easier to:

- Perform calculations
- Combine data from multiple sources
- Create tables with your results

Questions to ask yourself in this step:

1. What story is my data telling me?
2. How will my data help me solve this problem?
3. Who needs my company's product or service? What type of person is most likely to use it?

Step 5: Share

Everyone shares their results differently so be sure to summarize your results with clear and enticing visuals of your analysis using data via tools like graphs or dashboards. This is your chance to show the stakeholders you have solved their problem and how you got there. Sharing will certainly help your team:

- Make better decisions
- Make more informed decisions
- Lead to stronger outcomes
- Successfully communicate your findings

Questions to ask yourself in this step:

1. How can I make what I present to the stakeholders engaging and easy to understand?
2. What would help me understand this if I were the listener?

Step 6: Act

Now it's time to act on your data. You will take everything you have learned from your data analysis and put it to use. This could mean providing your stakeholders with recommendations based on your findings so they can make data-driven decisions.

Questions to ask yourself in this step:

1. How can I use the feedback I received during the share phase (step 5) to actually meet the stakeholder's needs and expectations?

These six steps can help you to break the data analysis process into smaller, manageable parts, which is called **structured thinking**. This process involves four basic activities:

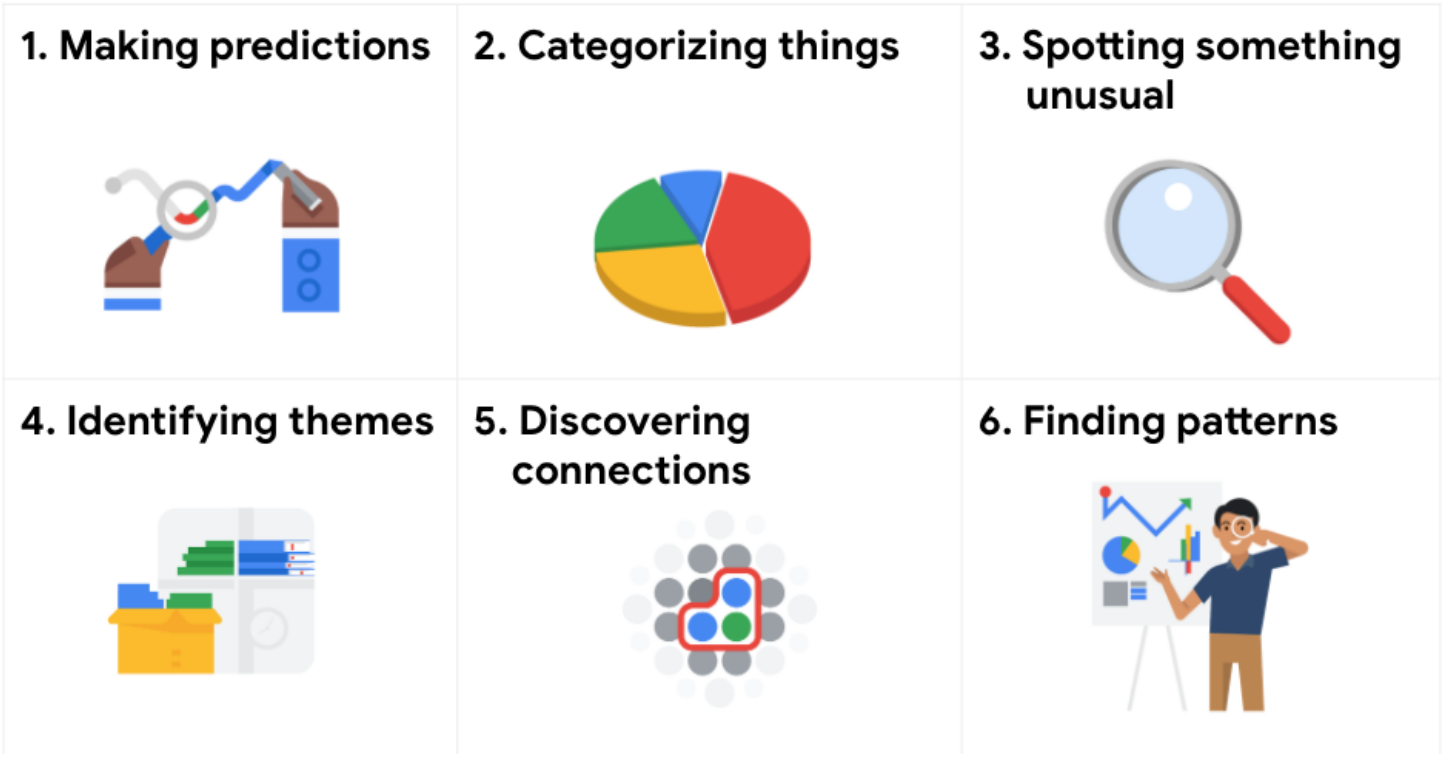
1. Recognizing the current problem or situation
2. Organizing available information
3. Revealing gaps and opportunities
4. Identifying your options

When you are starting out in your career as a data analyst, it is normal to feel pulled in a few different directions with your role and expectations. Following processes like the ones outlined here and using structured thinking skills can help get you back on track, fill in any gaps and let you know exactly what you need.

Six common problem types

Data analytics is so much more than just plugging information into a platform to find insights. It is about solving problems. To get to the root of these problems and find practical solutions, there are lots of opportunities for creative thinking. No matter the problem, the first and most important step is understanding it. From there, it is good to take a problem-solver approach to your analysis to help you decide what information needs to be included, how you can transform the data, and how the data will be used.

Data analysts typically work with six problem types



1. Making predictions 2. Categorizing things 3. Spotting something unusual 4. Identifying themes 5. Discovering connections 6. Finding patterns

A video, Common problem types, introduced the six problem types with an example for each. The examples are summarized below for review.

Making predictions

A company that wants to know the best advertising method to bring in new customers is an example of a problem requiring analysts to make predictions. Analysts with data on location, type of media, and number of new customers acquired as a result of past ads can't guarantee future results, but they can help predict the best placement of advertising to reach the target audience.

Categorizing things

An example of a problem requiring analysts to categorize things is a company's goal to improve customer satisfaction. Analysts might classify customer service calls based on certain keywords or scores. This could help identify top-performing customer service representatives or help correlate certain actions taken with higher customer satisfaction scores.

Spotting something unusual

A company that sells smart watches that help people monitor their health would be interested in designing their software to spot something unusual. Analysts who have analyzed aggregated health data can help product developers determine the right algorithms to spot and set off alarms when certain data doesn't trend normally.

Identifying themes

User experience (UX) designers might rely on analysts to analyze user interaction data. Similar to problems that require analysts to categorize things, usability improvement projects might require analysts to identify themes to help prioritize the right product features for improvement. Themes are most often used to help researchers explore certain aspects of data. In a user study, user beliefs, practices, and needs are examples of themes.

By now you might be wondering if there is a difference between categorizing things and identifying themes. The best way to think about it is: categorizing things involves assigning items to categories; identifying themes takes those categories a step further by grouping them into broader themes.

Discovering connections

A third-party logistics company working with another company to get shipments delivered to customers on time is a problem requiring analysts to discover connections. By analyzing the wait times at shipping hubs, analysts can determine the appropriate schedule changes to increase the number of on-time deliveries.

Finding patterns

Minimizing downtime caused by machine failure is an example of a problem requiring analysts to find patterns in data. For example, by analyzing maintenance data, they might discover that most failures happen if regular maintenance is delayed by more than a 15-day window.

Key takeaway

As you move through this program, you will develop a sharper eye for problems and you will practice thinking through the problem types when you begin your analysis. This method of problem solving will help you figure out solutions that meet the needs of all stakeholders.

More about SMART questions

Companies in lots of industries today are dealing with rapid change and rising uncertainty. Even well-established businesses are under pressure to keep up with what is new and figure out what is next. To do that, they need to ask questions. Asking the right questions can help spark the innovative ideas that so many businesses are hungry for these days.

The same goes for data analytics. No matter how much information you have or how advanced your tools are, your data won’t tell you much if you don’t start with the right questions. Think of it like a detective with tons of evidence who doesn’t ask a key suspect about it. Coming up, you will learn more about how to ask highly effective questions, along with certain practices you want to avoid.

Highly effective questions are SMART questions:

Specific:
Is the question specific? Does it address the problem? Does it have context? Will it uncover a lot of the information you need?

Measurable: Will the question give you answers that you can measure?

Action-oriented: Will the answers provide information that helps you devise some type of plan?

Relevant: Is the question about the particular problem you are trying to solve?

Time-bound: Are the answers relevant to the specific time being studied?

Examples of SMART questions

Here's an example that breaks down the thought process of turning a problem question into one or more SMART questions using the SMART method: **What features do people look for when buying a new car?**

- **Specific**: Does the question focus on a particular car feature?
- **Measurable**: Does the question include a feature rating system?
- **Action-oriented**: Does the question influence creation of different or new feature packages?
- **Relevant**: Does the question identify which features make or break a potential car purchase?
- **Time-bound**: Does the question validate data on the most popular features from the last three years?

Questions should be **open-ended.** This is the best way to get responses that will help you accurately qualify or disqualify potential solutions to your specific problem. So, based on the thought process, possible SMART questions might be:

- On a scale of 1-10 (with 10 being the most important) how important is your car having four-wheel drive? Explain.
- What are the top five features you would like to see in a car package?
- What features, if included with four-wheel drive, would make you more inclined to buy the car?
- How does a car having four-wheel drive contribute to its value, in your opinion?

Things to avoid when asking questions

Leading questions: questions that only have a particular response

- Example: **This product is too expensive, isn't it?**

This is a leading question because it suggests an answer as part of the question. A better question might be, "What is your opinion of this product?" There are tons of answers to that question, and they could include information about usability, features, accessories, color, reliability, and popularity, on top of price. Now, if your problem is actually focused on pricing, you could ask a question like "What price (or price range) would make you consider purchasing this product?" This question would provide a lot of different measurable responses.

Closed-ended questions: questions that ask for a one-word or brief response only

- Example: **Were you satisfied with the customer trial?**

This is a closed-ended question because it doesn't encourage people to expand on their answer. It is really easy for them to give one-word responses that aren't very informative. A better question might be, "What did you learn about customer experience from the trial." This encourages people to provide more detail besides "It went well."

Vague questions: questions that aren't specific or don't provide context

- Example: **Does the tool work for you?**

This question is too vague because there is no context. Is it about comparing the new tool to the one it replaces? You just don't know. A better inquiry might be, "When it comes to data entry, is the new tool faster, slower, or about the same as the old tool? If faster, how much time is saved? If slower, how much time is lost?" These questions give context (data entry) and help frame responses that are measurable (time).

Terms and definitions for Course 2, Module 1

Action-oriented question: A question whose answers lead to change

Cloud: A place to keep data online, rather than a computer hard drive

Data analysis process: The six phases of ask, prepare, process, analyze, share, and act whose purpose is to gain insights that drive informed decision-making

Data life cycle: The sequence of stages that data experiences, which include plan, capture, manage, analyze, archive, and destroy

Leading question: A question that steers people toward a certain response

Measurable question: A question whose answers can be quantified and assessed

Problem types: The various problems that data analysts encounter, including categorizing things, discovering connections, finding patterns, identifying themes, making predictions, and spotting something unusual

Relevant question: A question that has significance to the problem to be solved

SMART methodology: A tool for determining a question's effectiveness based on whether it is specific, measurable, action-oriented, relevant, and time-bound

Specific question: A question that is simple, significant, and focused on a single topic or a few closely related ideas

Structured thinking: The process of recognizing the current problem or situation, organizing available information, revealing gaps and opportunities, and identifying options

Time-bound question: A question that specifies a timeframe to be studied

Unfair question: A question that makes assumptions or is difficult to answer honestly

Data trials and triumphs
Introduction

A data analytics professional's job is to provide the data necessary to inform key decisions. They also need to frame their analysis in a way that helps business leaders make the best possible decisions.

In this reading, you're going to explore the role of data in decision-making and the reasons why data analytics professionals are so important to this process. You'll compare data-driven and data-inspired decisions to understand the difference between them. You'll also check out some examples where projects failed or succeeded based on how the data was applied.

Both data-driven and data-inspired approaches are rooted in the idea that data is inherently valuable for making a decision. Well-curated data can provide information to decision-makers that improves the quality of their decisions. Remember: Data does not make decisions, but it does improve them.

Data-driven decisions

As you've been learning, data-driven decision-making means using facts to guide business strategy. The phrase "data-driven decisions" means exactly that: Data is used to arrive at a decision. This approach is limited by the quantity and quality of readily-available data. If the quality and quantity of the data is sufficient, this approach can far improve decision-making. But if the data is insufficient or biased, this can create problems for decision-makers. Potential dangers of relying entirely on data-driven decision-making can include overreliance on historical data, a tendency to ignore qualitative insights, and potential biases in data collection and analysis

Example of a data-driven decision

A/B testing is a simple example of collecting data for data-driven decision-making. For example, a website that sells widgets has an idea for a new website layout they think will result in more people buying widgets. For two weeks, half of their website visitors are directed to the old site; the other half are directed to the new site. After those two weeks, the analyst gathers the data about their website visitors and the number of widgets sold for analysis. This helps the analyst understand which website layout resulted in more widget sales. If the new website performed better in producing widget sales, then the company can confidently make the decision to use the new layout!

Data-inspired decisions

Data-inspired decisions include the same considerations as data-driven decisions while adding another layer of complexity. They create space for people using data to consider a broader range of ideas: drawing on comparisons to related concepts, giving weight to feelings and experiences, and considering other qualities that may be more difficult to measure. Data-inspired decision-making can avoid some of the pitfalls that data-driven decisions might be prone to.

Example of a data-inspired decision

A customer support center gathers customer satisfaction data (often known as a "CSAT" score). They use a simple 1–10 score along with a qualitative description in which the customer describes their experience. The customer support center manager wants to improve customer experience, so they set a goal to improve the CSAT score. They start by analyzing the CSAT scores and reading each of the descriptions from the customers. Additionally, they interview the people working in the customer support center. From there, the manager formulates a strategy and decides what needs to improve the most in order to raise customer satisfaction scores. While the manager certainly relies on the CSAT data in the decision-making process, input of support center representatives and other qualitative information informs the approach as well.

A data analysis triumph

When data is used strategically, businesses can transform and grow their revenue. Consider the example below.

PepsiCo

Since the days of the New Coke launch, things have changed dramatically for beverage and other consumer packaged goods (CPG) companies.

According to a *Think with Google* article by Shyam Venugopal, PepsiCo "hired analytical talent and established cross-functional workflows around an infrastructure designed to put consumers' needs first. Then [the company] set up the right processes to make critical decisions based on data and technology use cases. Finally, [it] invested in the right technology stack and platforms so that data could flow into a central cloud-based hub. This is critical. When data comes together, we develop a holistic understanding of the consumer and their journeys."

In this data-inspired decision, PepsiCo is not just using its own set of data, but also employing external sources to supplement its datasets and expand its market reach. Learn about how PepsiCo is delivering a more personal and valuable experience to customers using data in How one of the world's biggest marketers ripped up its playbook and learned to anticipate intent.

Data analysis failures

You've been learning why data is such a powerful business tool and how data analysts help their companies make data-driven decisions for great results. Using data to draw accurate conclusions and make good recommendations starts with having complete, correct, and relevant data.

Note: It's important to remember that it's possible to have solid data and still make the wrong choices. It's up to data analysts to interpret the data accurately. When data is interpreted incorrectly, that incorrect interpretation can lead to huge losses. Consider the following.

Coke launch failure

In 1985, New Coke was launched, replacing the classic Coke formula. The company had done taste tests with 200,000 people and found that test subjects preferred the taste of New Coke over Pepsi, which had become a tough competitor. Based on this data alone, classic Coke was taken off the market and replaced with New Coke. The company thought this was the solution to take back the market share that had been lost to Pepsi.

But as it turns out, New Coke was very unpopular—and the company ended up losing tens of millions of dollars. The data seemed correct, but it was incomplete: The data didn't consider how customers would feel about New Coke replacing classic Coke. The company's decision to retire classic Coke was a data-driven decision based on incomplete data.

Mars Orbiter loss

In 1999, NASA lost the $125 million Mars Climate Orbiter even though the teams had good data. The spacecraft burned to pieces because of poor collaboration and communication. The Orbiter's navigation team was using the International System of Units (newtons) for their force calculations, but the engineers who built the spacecraft used the English Engineering Units system (pounds) for force calculations.

No one realized there was a problem until the Orbiter burst into flames in the Martian atmosphere. Later, a NASA review board investigating the cause of the problem discovered the issue was in the software that controlled the thrusters. One program calculated the thrusters' force in pounds; another program working with the data assumed it was in newtons. The software controllers were making data-driven decisions to adjust the thrust based on 100% accurate data, but these decisions were wrong because of inaccurate assumptions when interpreting it. The two teams might have communicated so they picked a single unit of measure, or so the analysts would have known that conversion was a necessary step in the process to prepare the data. A conversion of the data from one system of measurement to the other could have prevented the loss.

There's a difference between making a decision with incomplete data and making a decision with a small amount of data. You learned that making a decision with incomplete data is dangerous. But sometimes accurate data from a small test can help you make a good decision. Stay tuned: You'll learn about how much data to collect later in the program.

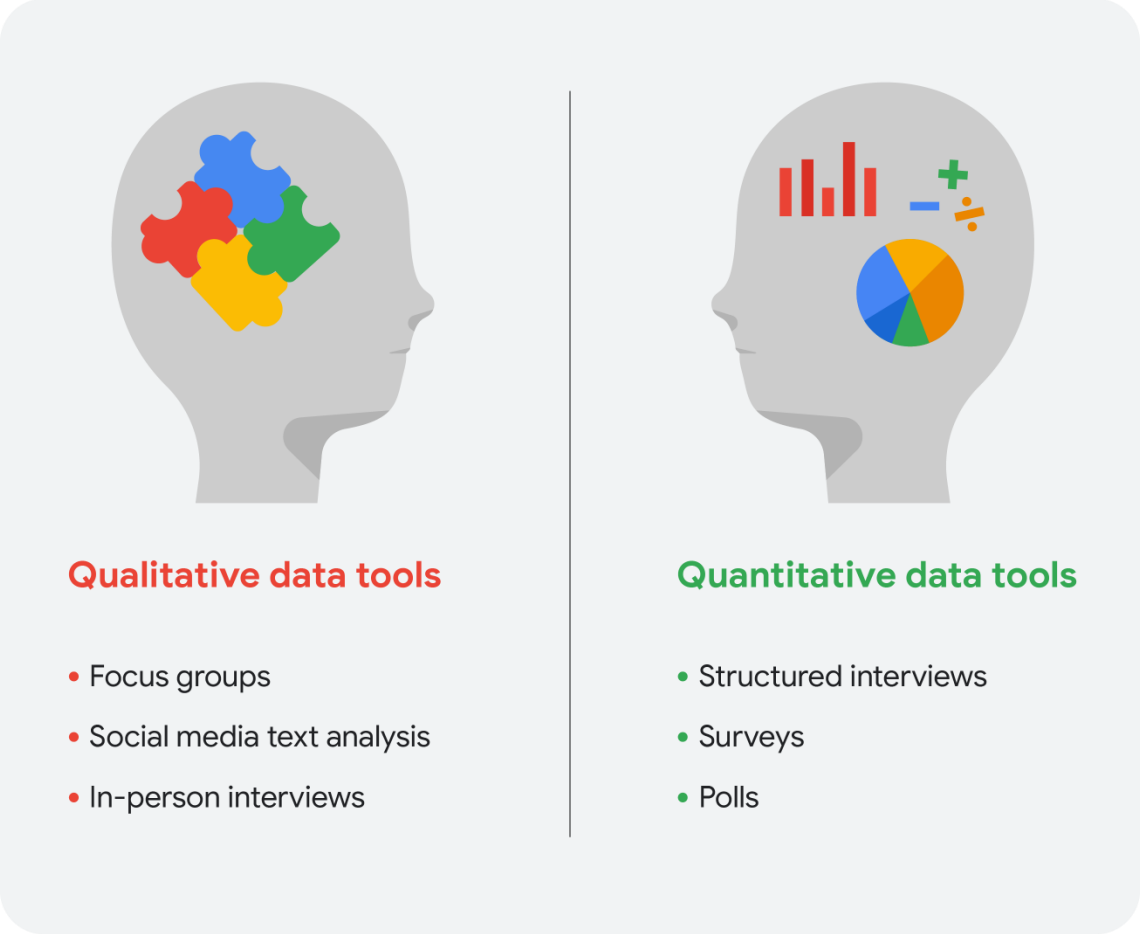
Key takeaways

As a data analyst, you'll rarely need to consider, "Am I being data-driven or data-inspired?" It's helpful to have some context for these two approaches, though your own skills and knowledge will be the most important parts of any analysis project. So, keep a data-driven mindset and ask lots of questions. Experiment with many different possibilities. And use both logic and creativity along the way. Using this approach, you'll be prepared to interpret your data with the highest levels of care and accuracy.

Qualitative and quantitative data in business

This reading further elaborates on the meaning of **qualitative** versus **quantitative**.

As you have learned, there are two types of data: qualitative and quantitative.



Qualitative data tools: focus groups, social media text analysis, and in-person interviews Quantitative data tools: structured interviews, surveys, and polls
Now, take a closer look at the data types and data collection tools. In this scenario, you are a data analyst for a chain of movie theaters. Your manager wants you to track trends in:

- **Movie attendance over time**
- **Profitability of the concession stand**
- **Evening audience preferences**

Assume quantitative data already exists to monitor all three trends.

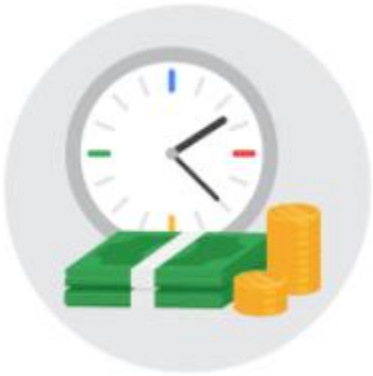
Movie attendance over time

Starting with the historical data the theater has through its loyalty and rewards program, your first step is to investigate what insights you can gain from that data. You look at attendance over the last 3 months. But, because

the last 3 months didn't include a major holiday, you decide it is better to look at a full year's worth of data. As you suspected, the quantitative data confirmed that average attendance was 550 per month but then rose to an average of 1,600 per month for the months with holidays.

The historical data serves your needs for the project, but you also decide that you will resume the analysis again in a few months after the theater increases ticket prices for evening showtimes.

Profitability of the concession stand



Profit is calculated by subtracting cost from sales revenue. The historical data shows that while the concession stand was profitable, profit margins were razor thin at less than 5%. You saw that average purchases totaled $20 or less. You decide that you will keep monitoring this on an ongoing basis.

Based on your understanding of data collection tools, you will suggest an online survey of customers so they can comment on the food at the concession stand. This will enable you to gather even more quantitative data to revamp the menu and potentially increase profits.

Evening audience preferences

Your analysis of the historical data shows that the 7:30 PM showtime was the most popular and had the greatest attendance, followed by the 7:15 PM and 9:00 PM showtimes. You may suggest replacing the current 8:00 PM showtime that has lower attendance with an 8:30 PM showtime. But you need more data to back up your hunch that people would be more likely to attend the later show.

Evening movie-goers are the largest source of revenue for the theater. Therefore, you also decide to include a question in your online survey to gain more insight.

Qualitative data for all three trends plus ticket pricing

Since you know that the theater is planning to raise ticket prices for evening showtimes in a few months, you will also include a question in the survey to get an idea of customers' price sensitivity.

Your final online survey might include these questions for qualitative data:

1. What went into your decision to see a movie in our theater today? (movie attendance)
2. What do you think about the quality and value of your purchases at the concession stand? (concession stand profitability)
3. Which showtime do you prefer, 8:00 PM or 8:30 PM, and why do you prefer that time? (evening movie-goer preferences)
4. Under what circumstances would you choose a matinee over a nighttime showing? (ticket price increase)

Key takeaways

Data analysts will generally use both types of data in their work. Usually, qualitative data can help analysts better understand their quantitative data by providing a reason or more thorough explanation. In other words, quantitative data generally gives you the what, and qualitative data generally gives you the why. By using both quantitative and qualitative data, you can learn when people like to go to the movies and why they chose the theater. Maybe they really like the reclining chairs, so your manager can purchase more recliners. Maybe the theater is the only one that serves root beer. Maybe a later show time gives them more time to drive to the theater from where popular restaurants are located. Maybe they go to matinees because they have kids and want to save money. You wouldn't have discovered this information by analyzing only the quantitative data for attendance, profit, and showtimes.

Tools for visualizing data

In this course, you'll work with Tableau and spreadsheets. Both of these tools have advantages and disadvantages Often, data analysts will discover they need to use multiple tools, even on a single project. What you use will largely be determined by the work you're doing and your goals. This reading explores two of the tools you might use to visualize and present data: spreadsheets and Tableau.

Spreadsheets

Google Workspace and Microsoft Office Suite both offer spreadsheet applications. You've worked with Google Sheets in this course, and it's very similar in function to Microsoft Excel. If you want to compare some of the features of Sheets to Excel, check out the Microsoft video Create a chart from start to finish.

Both Sheets and Excel are go-to choices for creating static charts and graphs. They offer basic data visualization capabilities that are often enough for simple visualizations. In addition, you can use them to clean, sort, and filter data. And both offer a range of chart types, graphing tools, and pivot tables for creating effective data visualizations. These charts are easy to manage; they update when the source data is updated, so they don't require much manual intervention once implemented.

Sheets and Excel are connected to other apps in their product suites. Google Docs and Slides are very similar to Microsoft Word and Powerpoint, for example. You can incorporate data visualizations from Sheets or Excel into reports and documents in Docs and Word. Presentation programs such as Slides and Powerpoint allow you to create engaging presentations that include data visualizations so you can share insights in a presentation format. Learn more about the power of this interconnectivity among Google tools in the article Link a chart, table, or slides to Google Docs or Slides.

Tableau

Tableau is used to create powerful and interactive visualizations, making it an excellent choice for data visualizations such as live dashboards. Tableau also makes it easy to create charts, graphs, and dashboards in a drag-and-drop interface. The application supports a wide range of data sources and provides advanced analytics capabilities. These features allow for in-depth exploration of data trends and patterns.

Tableau is particularly useful for creating visualizations using huge datasets, like in this World Happiness Report by Sustainable Development Solutions which uses global reporting data on different countries' happiness ratings. Likewise, this visualization of Population and Housing State Data from 2020 United States Census Data compares population rates in the United States and available housing.

Tableau is widely known and used for its versatility and power, but it can take quite a bit of time to learn to use Tableau effectively. Soon, you'll begin practicing with Tableau. But if you'd like to check it out now, there is a free environment you can access at Tableau Public.

Key takeaways

There are many visualization tools you will have the opportunity to use as a data professional. Different tools have different advantages and disadvantages. Although Tableau ultimately has more power than a basic spreadsheet application, it's most often used for specific cases and to work with large datasets. Don't underestimate how much you can do with spreadsheets or how powerful interconnectivity between apps can be!

Most of the time, especially for something like a quick report, you're more likely to reach into your toolkit for your spreadsheet app of choice. But your data career will definitely benefit from Tableau, so as you progress take advantage of opportunities to work with it. With so many different data analysis situations, familiarity with all of these tools will help you know which is the best for each situation.

Design compelling dashboards

Dashboards are powerful visual tools that help you tell your data story. A dashboard is a tool that monitors live, incoming data. It organizes information from multiple datasets into one central location, offering huge time savings. Data analysts use dashboards to track, analyze, and visualize data in order to answer questions and solve problems. For a basic idea of what dashboards look like, refer to this article: "Real-world examples of business intelligence dashboards."

The beauty of dashboards

The following table summarizes the benefits of using a dashboard for both data analysts and their stakeholders.

Benefits	For data analysts	For stakeholders
Centralization	Share a single source of data with all stakeholders	Work with a comprehensive view of data, initiatives, objectives, projects, processes, and more
Visualization	Show and update live, incoming data in real time*	Spot changing trends and patterns more quickly
Insightfulness	Pull relevant information from different datasets	Understand the story behind the numbers to keep track of goals and make data-driven decisions
Customization	Create custom views dedicated to a specific person, project, or presentation of the data	Drill down to more specific areas of specialized interest or concern

It's important to remember that changed data is pulled into dashboards automatically only if the data structure is the same. If the data structure changes, you have to update the dashboard design before the data can update live.

Tableau

There are many different visualization tools available. One of the most powerful is Tableau, which supports a range of data sources and has advanced analytics capabilities that allow for in-depth exploration of data trends and patterns. Tableau can handle more data and larger datasets than many other tools and offers real-time data availability.

It does take some time to learn to use Tableau, but your efforts can be well-rewarded, as Tableau visualizations are pleasantly interactive. For a dashboard to be successful, it needs to engage users and help them learn. Tableau has put in a lot of effort to ensure that its users have a great experience and the platform is accessible to everyone.

Create a dashboard

Here's a process you can follow to create a dashboard, whether in Tableau or another visualization tool:

1. Identify the stakeholders who need to see the data and how they will use it

Begin by asking effective questions. Check out this dashboard requirements gathering worksheet to explore a wide range of good questions you can use to identify relevant stakeholders and their data needs. This is a great resource to help guide you through this process again and again.

2. Design the dashboard (what should be displayed)

Use these tips to help make your dashboard design clear and easy to follow:

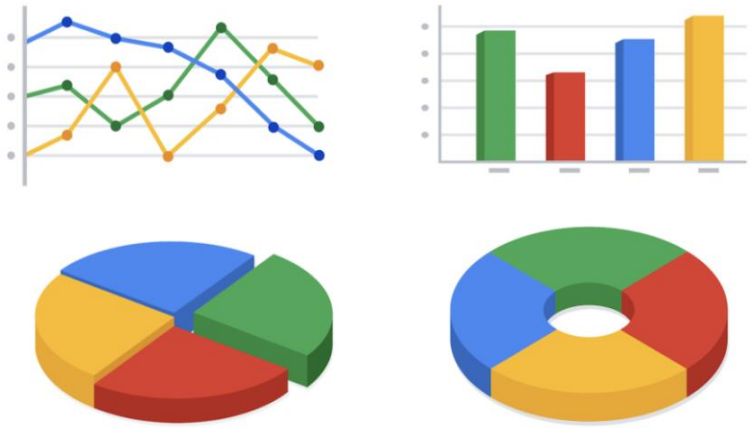
- Use a clear header to label the information.
- Add short text descriptions to each visualization.
- Show the most important information at the top.

3. Create mockups if desired

A mockup is a simple draft of a visualization used for planning a dashboard and evaluating its progress. This is optional, but a lot of data analysts like to sketch out their dashboards before creating them.

4. Select the visualizations

You have a lot of options here. Which visualizations you select depends on the data story you are telling. If you need to show a change in values over time, line charts or bar graphs might be the best choice. If your goal is to show how each part contributes to the whole amount being reported, a pie or donut chart is probably a better choice.



Two pie charts show an even distribution of 4 parts of a whole. The first pie chart is more traditional, appearing as a solid circle. The second pie chart is styled to show the same data in a doughnut shape.
To learn more about choosing the right visualizations, check out Tableau's galleries:

- For more samples of area charts, column charts, and other visualizations, visit the Tableau Dashboard Showcase. This gallery is full of great examples that were created using real data; explore this resource on your own to get some inspiration.
- Explore Tableau's Viz of the Day to check out visualizations curated by the community. These are visualizations created by Tableau users and are a great way to learn more about how other data analysts are using data visualization tools.

5. Create filters as needed

Filters show certain data while hiding the rest of the data in a dashboard. This can be a big help to identify patterns while keeping the original data intact. It's common for data analysts to use and share the same dashboard, but manage their part of it with a filter. To dig deeper into filters and find an example of filters in action, visit Tableau's page on Filter Actions. This is a useful resource to save and come back to when you start practicing using filters in Tableau on your own.

Key takeaways

Just like how the dashboard on an airplane shows the pilot their flight path, your dashboard does the same for your stakeholders. It helps them navigate the path of a project inside the data. If you add clear markers and highlight important points on your dashboard, users will understand where your data story is headed. Then, you can work together to make sure the business gets where it needs to go.

Activity Overview

Previously, you were introduced to the data management tool known as a dashboard. In this self-reflection, you'll examine different kinds of dashboards and consider how they are used by data analysts and their employers.

As a refresher, a dashboard is a single point of access for managing a business's information. It allows analysts to pull key information from data in a quick review by visualizing the data in a way that makes findings easy to understand.

This self-reflection will help you develop insights into your own learning and prepare you to connect your knowledge of dashboards to what you know about business needs. As you answer questions—and come up with questions of your own—you will consider concepts, practices, and principles to help refine your understanding and reinforce your learning. You've done the hard work, so make sure to get the most out of it: This reflection will help your knowledge stick!

Types of dashboards

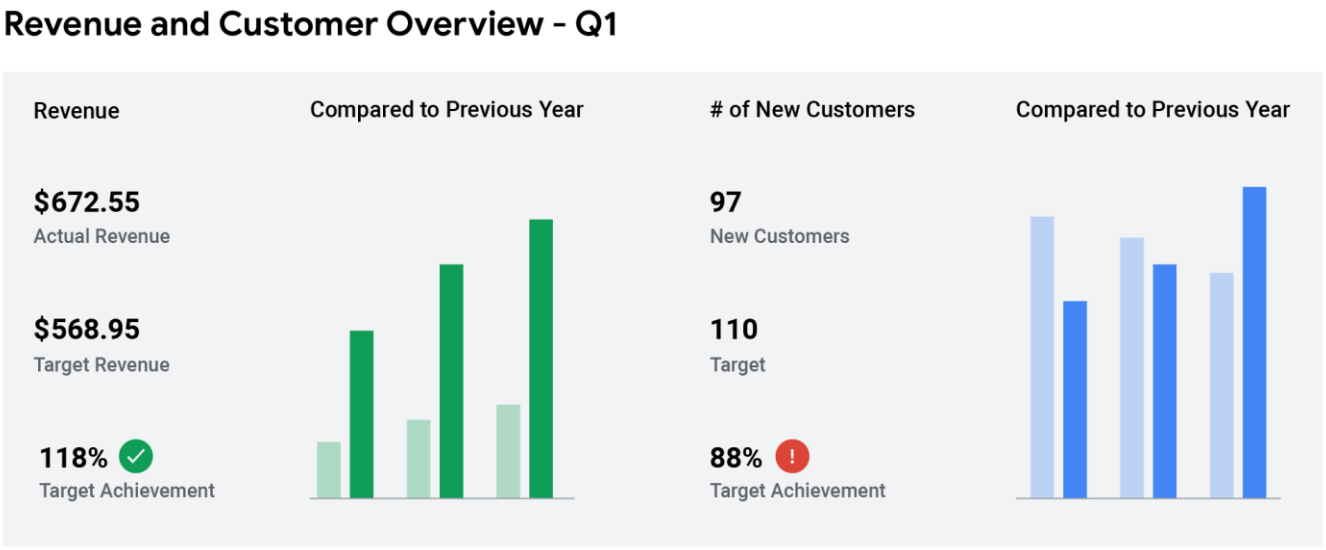
For a refresher, consider the different types of dashboards a business may use. Often, businesses will tailor a dashboard for a specific purpose. The three most common categories are:

- **Strategic**: focuses on long term goals and strategies at the highest level of metrics
- **Operational:** short-term performance tracking and intermediate goals
- **Analytical:** consists of the datasets and the mathematics used in these sets

Strategic dashboards

A wide range of businesses use strategic dashboards when evaluating and aligning their strategic goals. These dashboards provide information over the longest time frame—from a single financial quarter to years.

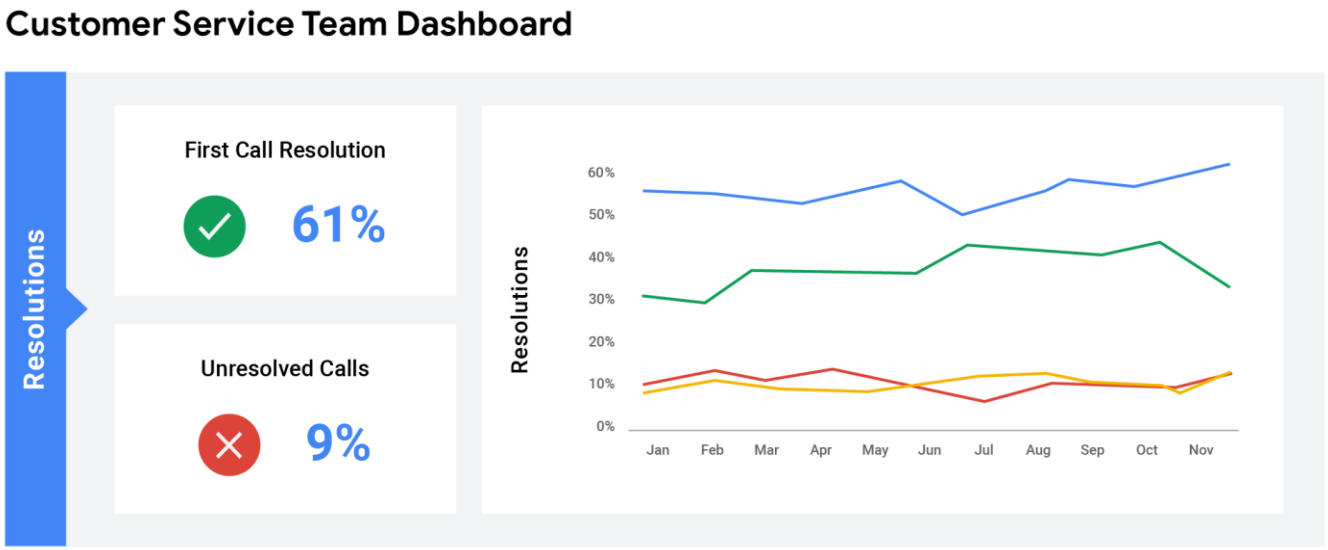
They typically contain information that is useful for enterprise-wide decision-making. Below is an example of a strategic dashboard which focuses on key performance indicators (KPIs) over a year.



Operational dashboards

Operational dashboards are, arguably, the most common type of dashboard. Because these dashboards contain information on a time scale of days, weeks, or months, they can provide performance insight almost in real-time.

This allows businesses to track and maintain their immediate operational processes in light of their strategic goals. The operational dashboard below focuses on customer service.

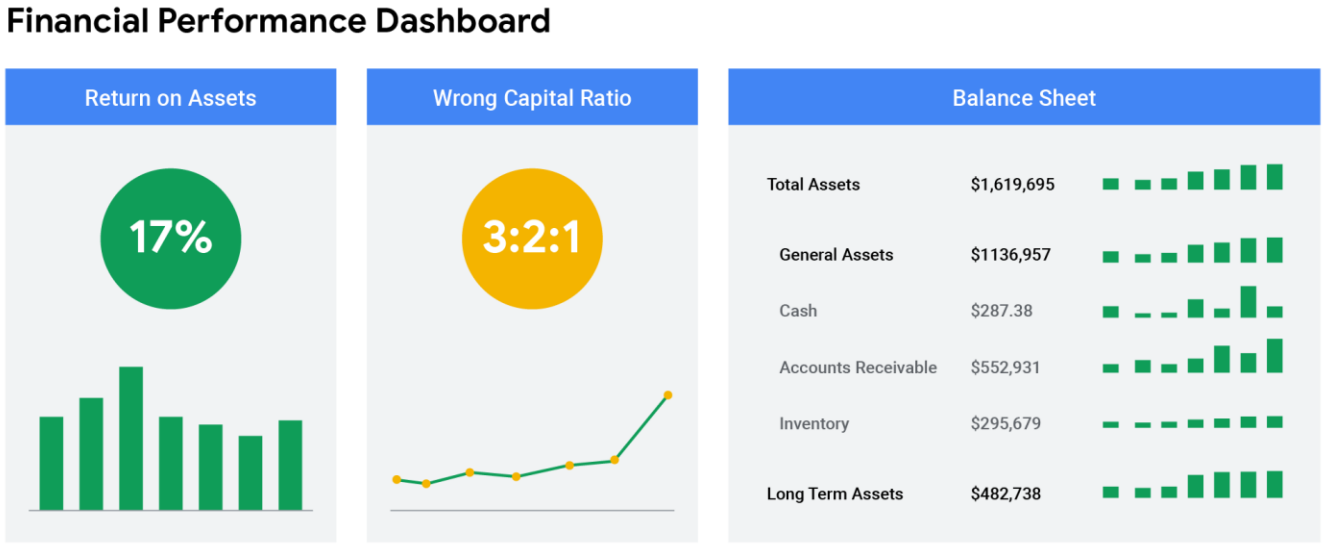


Resolutions are divided between first call resolution (61%) and unresolved calls (9%)

Analytical dashboards

Analytic dashboards contain a vast amount of data used by data analysts. These dashboards contain the details involved in the usage, analysis, and predictions made by data scientists.

Certainly the most technical category, analytic dashboards are usually created and maintained by data science teams and rarely shared with upper management as they can be very difficult to understand. The analytic dashboard below focuses on metrics for a company's financial performance.



Contains Return on assets, wrong capital ratio, and balance sheet

Big and small data

As a data analyst, you will work with data both big and small. Both kinds of data are valuable, but they play very different roles.

Whether you work with big or small data, you can use it to help stakeholders improve business processes, answer questions, create new products, and much more. But there are certain challenges and benefits that come with big data and the following table explores the differences between big and small data.

Small data	Big data
Describes a dataset made up of specific metrics over a short, well-defined time period	Describes large, less-specific datasets that cover a long time period
Usually organized and analyzed in spreadsheets	Usually kept in a database and queried
Likely to be used by small and midsize businesses	Likely to be used by large organizations
Simple to collect, store, manage, sort, and visually represent	Takes a lot of effort to collect, store, manage, sort, and visually represent
Usually already a manageable size for analysis	Usually needs to be broken into smaller pieces in order to be organized and analyzed effectively for decision-making

Challenges and benefits

Here are some **challenges** you might face when working with big data:

- A lot of organizations deal with data overload and way too much unimportant or irrelevant information.
- Important data can be hidden deep down with all of the non-important data, which makes it harder to find and use. This can lead to slower and more inefficient decision-making time frames.
- The data you need isn't always easily accessible.
- Current technology tools and solutions still struggle to provide measurable and reportable data. This can lead to unfair algorithmic bias.
- There are gaps in many big data business solutions.

Now for the good news! Here are some **benefits** that come with big data:

- When large amounts of data can be stored and analyzed, it can help companies identify more efficient ways of doing business and save a lot of time and money.
- Big data helps organizations spot the trends of customer buying patterns and satisfaction levels, which can help them create new products and solutions that will make customers happy.
- By analyzing big data, businesses get a much better understanding of current market conditions, which can help them stay ahead of the competition.
- As in our earlier social media example, big data helps companies keep track of their online presence—especially feedback, both good and bad, from customers. This gives them the information they need to improve and protect their brand.

The three (or four) V words for big data

When thinking about the benefits and challenges of big data, it helps to think about the three Vs: **volume, variety,** and **velocity.** Volume describes the amount of data. Variety describes the different kinds of data. Velocity describes how fast the data can be processed. Some data analysts also consider a fourth V: **veracity.** Veracity

refers to the quality and reliability of the data. These are all important considerations related to processing huge, complex datasets.

Volume	Variety	Velocity	Veracity
The amount of data	The different kinds of data	How fast the data can be processed	The quality and reliability of the data

Glossary terms from module 2
Terms and definitions for Course 2, Module 2

Algorithm: A process or set of rules followed for a specific task

Big data: Large, complex datasets typically involving long periods of time, which enable data analysts to address far-reaching business problems

Dashboard: A tool that monitors live, incoming data

Data-inspired decision-making: The process of exploring different data sources to find out what they have in common

Metric: A single, quantifiable type of data that is used for measurement

Metric goal: A measurable goal set by a company and evaluated using metrics

Pivot chart: A chart created from the fields in a pivot table

Pivot table: A data summarization tool used to sort, reorganize, group, count, total, or average data

Problem types: The various problems that data analysts encounter, including categorizing things, discovering connections, finding patterns, identifying themes, making predictions, and spotting something unusual

Qualitative data: A subjective and explanatory measure of a quality or characteristic

Quantitative data: A specific and objective measure, such as a number, quantity, or range

Report: A static collection of data periodically given to stakeholders

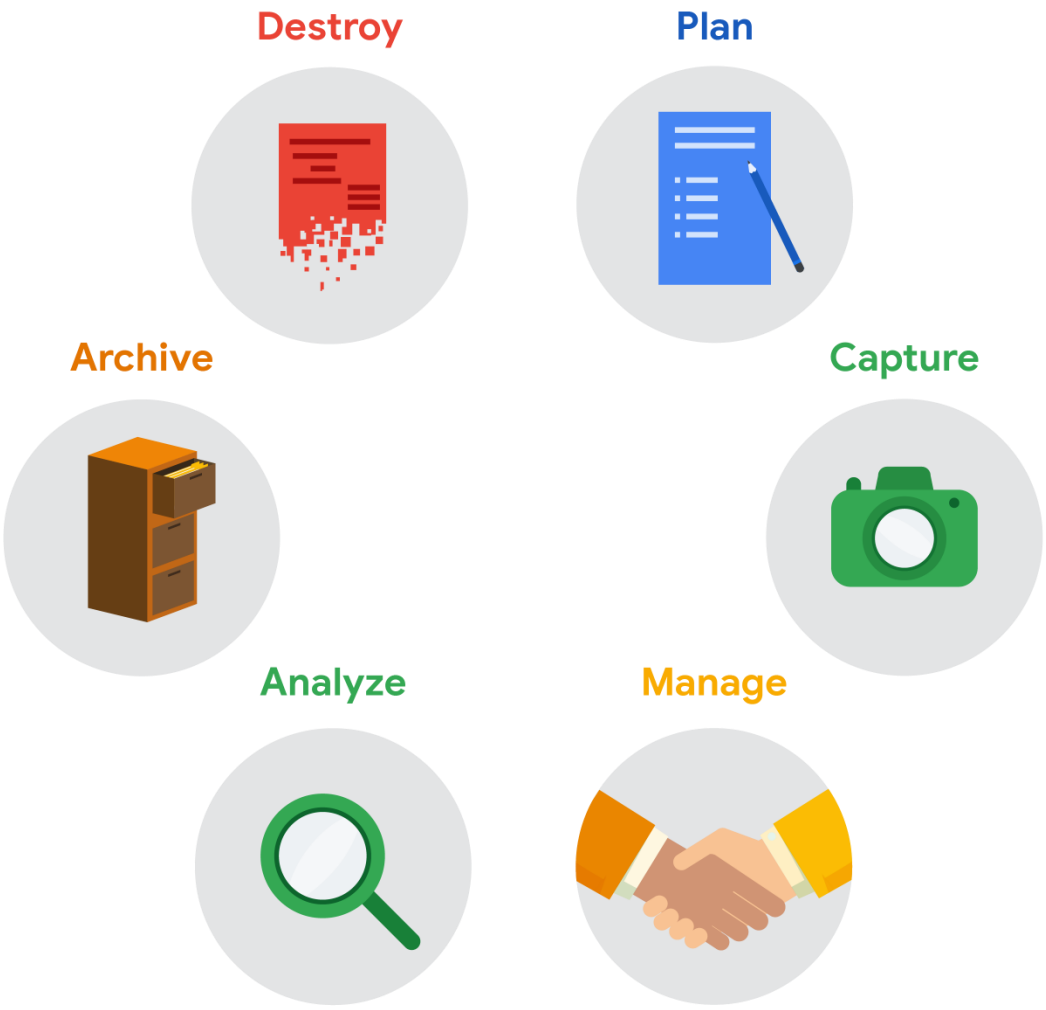
Return on investment (ROI): A formula that uses the metrics of investment and profit to evaluate the success of an investment

Revenue: The total amount of income generated by the sale of goods or services

Small data: Small, specific data points typically involving a short period of time, which are useful for making day-to-day decisions

Spreadsheets and the data life cycle

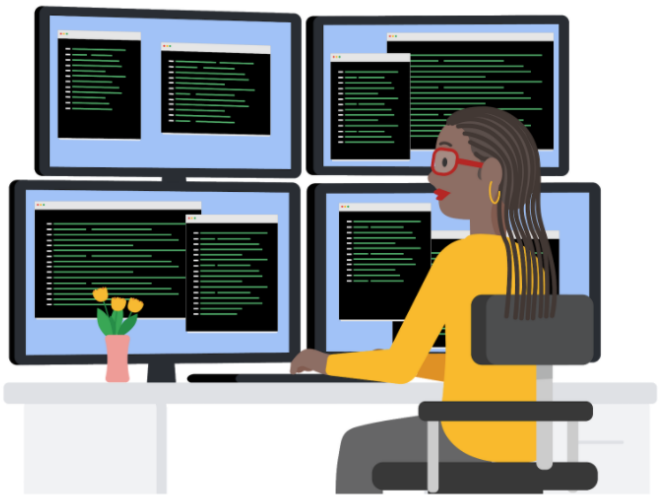
To better understand the benefits of using spreadsheets in data analytics, let's explore how they relate to each phase of the data life cycle: **plan, capture, manage, analyze, archive,** and **destroy**.



- **Plan** for the users who will work within a spreadsheet by developing organizational standards. This can mean formatting your cells, the headings you choose to highlight, the color scheme, and the way you order your data points. When you take the time to set these standards, you will improve communication, ensure consistency, and help people be more efficient with their time.
- **Capture** data by the source by connecting spreadsheets to other data sources, such as an online survey application or a database. This data will automatically be updated in the spreadsheet. That way, the information is always as current and accurate as possible.

- **Manage** different kinds of data with a spreadsheet. This can involve storing, organizing, filtering, and updating information. Spreadsheets also let you decide who can access the data, how the information is shared, and how to keep your data safe and secure.
- **Analyze** data in a spreadsheet to help make better decisions. Some of the most common spreadsheet analysis tools include formulas to aggregate data or create reports, and pivot tables for clear, easy-to-understand visuals.
- **Archive** any spreadsheet that you don't use often, but might need to reference later with built-in tools. This is especially useful if you want to store historical data before it gets updated.
- **Destroy** your spreadsheet when you are certain that you will never need it again, if you have better backup copies, or for legal or security reasons. Keep in mind, lots of businesses are required to follow certain rules or have measures in place to make sure data is destroyed properly.

Quick Reference: Formulas in spreadsheets



You have been learning a lot about spreadsheets and all kinds of time-saving calculations and organizational features they offer. One of the most valuable spreadsheet features is a **formula**. As a quick reminder, a formula is a set of instructions that does a specific calculation using the data in a spreadsheet. Formulas make it easy for data analysts to do powerful calculations automatically, which helps them analyze data more effectively. Below is a quick-reference guide to help you get the most out of formulas.

Formulas

The basics

- When you enter a formula in math, it generally ends with an equal sign (2 + 3 = ?). But with formulas, they always start with one instead (**=A2+A3**). The equal sign tells the spreadsheet that what follows is part of a formula, not just a word or number in a cell.
- After you enter the equal sign, most spreadsheet applications will display an autocomplete menu that lists valid formulas, names, and text strings. This is a great way to create and edit formulas while avoiding typing and syntax errors.
- A fun way to learn new formulas is just by typing an equal sign and a single letter of the alphabet. Choose one of the options that pops up and you will learn what that formula does.

Mathematical operators

- The mathematical operators used in spreadsheet formulas include:
- Subtraction – minus sign (-)

- Addition – plus sign (+)
- Division – forward-slash (/)
- Multiplication – asterisk (*)

Auto-filling

The lower-right corner of each cell has a fill handle. It is a small *green square* in Microsoft Excel and a small *blue circle* in Google Sheets.

- Click the fill handle square or circle for a cell and drag it down a column to auto-fill other cells in the column with the same value or formula in that cell.
- Click the fill handle square or circle for a cell and drag it across a row to auto-fill other cells in the row with the same value or formula in that cell.
- If you want to create a numbered sequence in a column or row, do the following: 1) Fill in the first two numbers of the sequence in two adjacent cells, 2) Select to highlight the cells, and 3) Drag the fill handle square or circle to the last cell to complete the sequence of numbers. For example, to insert 1 through 100 in each row of column A, enter **1** in cell A1 and **2** in cell A2. Then, select to highlight both cells, click the fill handle square or circle in cell A2, and drag it down to cell A100. This auto-fills the numbers sequentially so you don't have to enter them in each cell.

Absolute referencing

- Absolute referencing is marked by a dollar sign (**$**). For example, **=$A$10** has absolute referencing for both the column and the row value
- Relative references (which is what you normally do, e.g. "**=A10**") will change anytime the formula is copied and pasted. They are in relation to where the referenced cell is located. For example if you copied "**=A10**" to the cell to the right it would become "**=B10**". With absolute referencing "**=$A$10**" copied to the cell to the right would remain "**=$A$10**". But if you copied **$A10** to the cell below, it would change to **$A11** because the row value isn't an absolute reference.
- Absolute references will not change when you copy and paste the formula in a different cell. The cell being referenced is always the same.
- To easily switch between absolute and relative referencing in the formula bar, highlight the reference you want to change and press the **F4** key; for example, if you want to change the absolute reference, **$A$10**, in your formula to a relative reference, A10, highlight **$A$10** in the formula bar and then press the **F4** key to make the change.

Data range

- When you click into your formula, the colored ranges let you see which cells are being used in your spreadsheet. There are different colors for each unique range in your formula.
- In a lot of spreadsheet applications, you can press the **F2** (or **Enter**) key to highlight the range of data in the spreadsheet that is referenced in a formula. Click the cell with the formula, and then press the **F2** (or **Enter**) key to highlight the data in your spreadsheet.

Combining with functions

- **COUNTIF()** is a formula and a function. This means the function runs based on criteria set by the formula. In this case, **COUNT** is the formula; it will be executed IF the conditions you create are true. For example, you could use **=COUNTIF(A1:A16, "7")** to count only the cells that contained the number 7. Combining formulas and functions allows you to do more work with a single command.

The importance of context

Context in data analytics is the condition and circumstances that surround and give meaning to the data. Context is important in data analytics because it helps make disorganized data accessible and understood. The fact is, data has little value if it is not paired with context.

Understanding the context behind the data can help us make it more meaningful at every stage of the data analysis process. For example, you might be able to make a few guesses about what you're looking at in the following table, but you couldn't be certain without more context.

2010	28000
2005	18000
2000	23000
1995	10000

On the other hand, if the first column was labeled to represent the years when a survey was conducted, and the second column showed the number of people who responded to that survey, then the table would start to make a lot more sense. Take this a step further, and you might notice that the survey is conducted every 5 years. This added context helps you understand why there are five-year gaps in the table.

Years (Collected every 5 years)	**Respondents**
2010	28000
2005	18000
2000	23000
1995	10000

Context can turn raw data into meaningful information. It is very important for data analysts to contextualize their data. This means giving the data perspective by defining it. To do this, you need to identify:

- Who: The person or organization that created, collected, and/or funded the data collection
- What: The things in the world that data could have an impact on
- Where: The origin of the data
- When: The time when the data was created or collected
- Why: The motivation behind the creation or collection
- How: The method used to create or collect it

Understanding and including the context is important during each step of your analysis process, so it is a good idea to get comfortable with it early in your career. For example, when you collect data, you'll also want to ask questions about the context to make sure that you understand the business and business process. During organization, the context is important for your naming conventions, how you choose to show relationships between variables, and what you choose to keep or leave out. And finally, when you present, it is important to include contextual information so that your stakeholders understand your analysis.

Glossary terms from module 3

Terms and definitions for Course 2, Module 3

AVERAGE: A spreadsheet function that returns an average of the values from a selected range

Borders: Lines that can be added around two or more cells on a spreadsheet

Cell reference: A cell or a range of cells in a worksheet typically used in formulas and functions

COUNT: A spreadsheet function that counts the number of cells in a range that meet a specific criteria

Equation: A calculation that involves addition, subtraction, multiplication, or division (also called a math expression)

Fill handle: A box in the lower-right-hand corner of a selected spreadsheet cell that can be dragged through neighboring cells in order to continue an instruction

Filtering: The process of showing only the data that meets a specified criteria while hiding the rest

Header: The first row in a spreadsheet that labels the type of data in each column

Math expression: A calculation that involves addition, subtraction, multiplication, or division (also called an equation)

Math function: A function that is used as part of a mathematical formula

MAX: A spreadsheet function that returns the largest numeric value from a range of cells

MIN: A spreadsheet function that returns the smallest numeric value from a range of cells

Open data: Data that is available to the public

Operator: A symbol that names the operation or calculation to be performed

Order of operations: Using parentheses to group together spreadsheet values in order to clarify the order in which operations should be performed

Problem domain: The area of analysis that encompasses every activity affecting or affected by a problem

Range: A collection of two or more cells in a spreadsheet

Report: A static collection of data periodically given to stakeholders

Return on investment (ROI): A formula that uses the metrics of investment and profit to evaluate the success of an investment

Revenue: The total amount of income generated by the sale of goods or services

Scope of work (SOW): An agreed-upon outline of the tasks to be performed during a project

Sorting: The process of arranging data into a meaningful order to make it easier to understand, analyze, and visualize

SUM: A spreadsheet function that adds the values of a selected range of cells