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

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.

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).

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."

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.

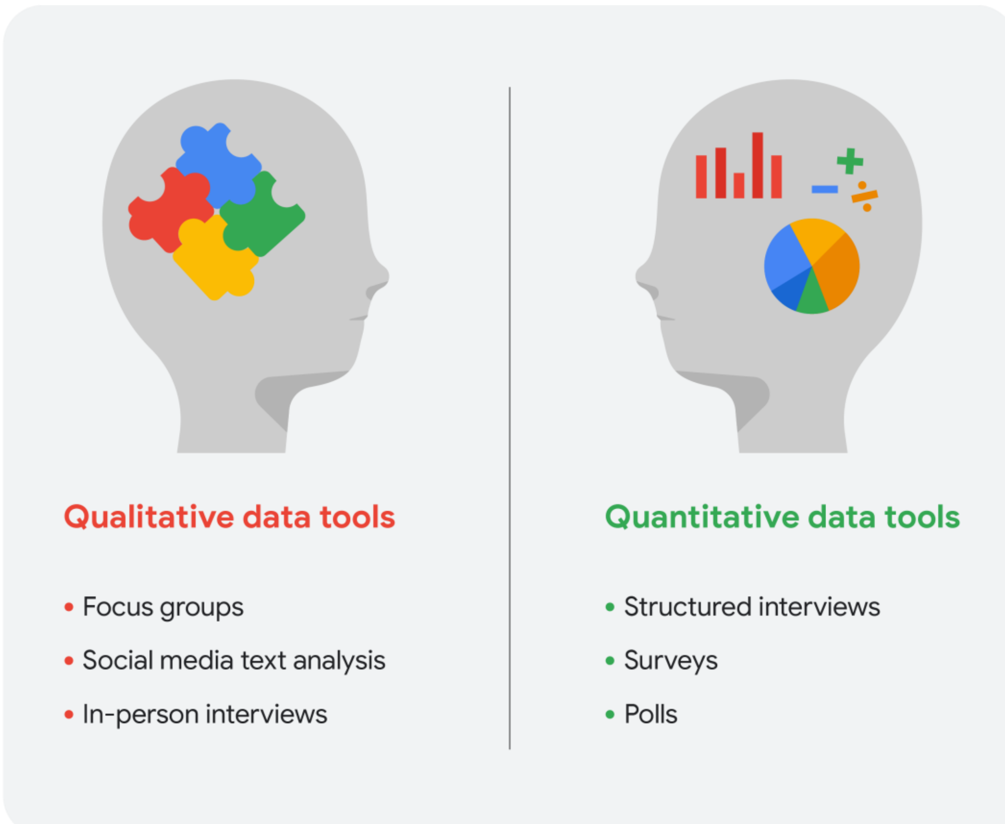
No one realized there was a problem until the Orbiter burst into flames in the Martian atmosphere. Later, a NASA review board investigating the cause of the problem discovered the issue was in the software that controlled the thrusters. One program calculated the thrusters' force in pounds; another program 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.

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.



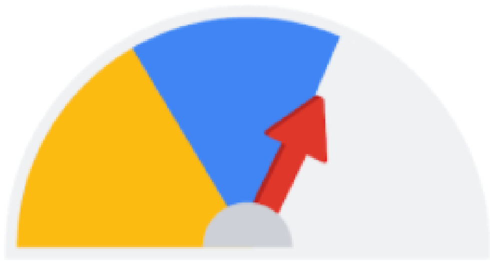
Qualitative data tools: focus groups, social media text analysis, and in-person interviews
Quantitative data tools: structured interviews, surveys, and polls

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 show times.

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)

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

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.

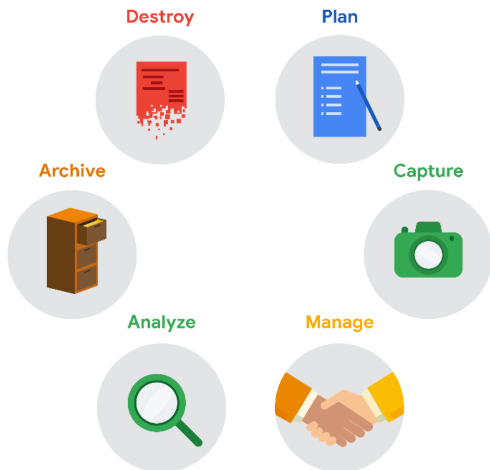
Two pie charts show an even distribution of 4 parts of a whole. The first pie chart is more traditional, appearing as a solid circle. The second pie chart is styled to show the same data in a doughnut shape.

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.

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.

Limitations of data

Data is powerful, but it has its limitations. Has someone’s personal opinion found its way into the numbers? Is your data telling the whole story? Part of being a great data analyst is knowing the limits of data and planning for them. This reading explores how you can do that.

The case of incomplete (or nonexistent!) data

If you have incomplete or nonexistent data, you might realize during an analysis that you don't have enough data to reach a conclusion. Or, you might even be solving a different problem altogether! For example, suppose you are looking for employees who earned a particular certificate but discover that certification records go back only two years at your company. You can still use the data, but you will need to make the limits of your analysis clear. You might be able to find an alternate source of the data by contacting the company that led the training. But to be safe, you should be up front about the incomplete dataset until that data becomes available.

Don’t miss misaligned data

If you're collecting data from other teams and using existing spreadsheets, it is good to keep in mind that people use different business rules. So one team might define and measure things in a completely different way than another. For example, if a metric is the total number of trainees in a certificate program, you could have one team that counts every person who registered for the training, and another team that counts only the people who completed the program. In cases like these, establishing how to measure things early on standardizes the data across the board for greater reliability and accuracy. This will make sure comparisons between teams are meaningful and insightful.

Deal with dirty data

Dirty data refers to data that contains errors. Dirty data can lead to productivity loss, unnecessary spending, and unwise decision-making. A good data cleaning effort can help you avoid this. As a quick reminder, data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When you find and fix the errors - while tracking the changes you made - you can avoid a data disaster. You will learn how to clean data later in the training.

Tell a clear story

Below are some of the best practices he recommends for good data storytelling:

- **Compare the same types of data**: Data can get mixed up when you chart it for visualization. Be sure to compare the same types of data and double check that any segments in your chart definitely display different metrics.
- **Visualize with care**: A 0.01% drop in a score can look huge if you zoom in close enough. To make sure your audience sees the full story clearly, it is a good idea to set your Y-axis to 0.
- **Leave out needless graphs:** If a table can show your story at a glance, stick with the table instead of a pie chart or a graph. Your busy audience will appreciate the clarity.
- **Test for statistical significance:** Sometimes two datasets will look different, but you will need a way to test whether the difference is real and important. So remember to run statistical tests to see how much confidence you can place in that difference.
- **Pay attention to sample size**: Gather lots of data. If a sample size is small, a few unusual responses can skew the results. If you find that you have too little data, be careful about using it to form judgments. Look for opportunities to collect more data, then chart those trends over longer periods.

Be the judge

In any organization, a big part of a data analyst’s role is making sound judgments. When you know the limitations of your data, you can make judgment calls that help people make better decisions supported by the data. Data is an extremely powerful tool for decision-making, but if it is incomplete, misaligned, or hasn’t been cleaned, then it can be misleading. Take the necessary steps to make sure that your data is complete and consistent. Clean the data before you begin your analysis to save yourself and possibly others a great amount of time and effort.