

Data analyst

The six phases of data analysis

The data analysis process helps analysts break down business problems into a series of manageable tasks:

In the **ask phase**, you'll work to understand the challenge to be solved or the question to be answered. It will likely be assigned to you by stakeholders. As this is the ask phase, you'll ask many questions to help you along the way.

Next, in the **prepare phase**, you'll find and collect the data you'll need to answer your questions. You'll identify data sources, gather data, and verify that it is accurate and useful for answering your questions.

The **process phase** is when you will clean and organize your data. Tasks you perform here include removing any inconsistencies; filling in missing values; and, in many cases, changing the data to a format that's easier to work with. Essentially, you're ensuring the data is ready before you begin analysis.

The **analyze phase** is when you do the necessary data analysis to uncover answers and solutions. Depending on the situation and the data, this could involve tasks such as calculating averages or counting items in categories so you can examine trends and patterns.

Next comes the **share phase**, when you present your findings to decision-makers through a report, presentation, or data visualizations. As part of the share phase, you decide which medium you want to use to share your findings and select the data to include. Tools for presenting data visually include charts made in Google Sheets, Tableau, and R.

Last is the **act phase**, in which you and others in the company put the data insights into action. This could mean implementing a new business strategy, making changes to a website, or any other action that solves the initial problem.

these skills are:

- Curiosity
- Understanding of context
- Technical mindset

- Data design
- Data strategy

The **Five Whys** technique is a problem-solving method used to identify the root cause of an issue by asking "why" multiple times—typically five. Here's how it works:

1. **Identify the Problem:** Start with a clear statement of the problem you want to solve.
2. **Ask Why:** Ask why the problem occurs. Write down the answer.
3. **Repeat:** Take the answer from the previous "why" and ask "why" again. Continue this process for five iterations or until you reach the root cause.
4. **Analyze the Final Answer:** The final answer should reveal the underlying issue that, if addressed, can prevent the problem from recurring.

Example:

- **Problem:** I can't make a blueberry pie.
 - **Why?** There are no blueberries at the store. (1st Why)
 - **Why?** The blueberry bushes don't have enough fruit this season. (2nd Why)
 - **Why?** Birds are eating all the berries. (3rd Why)
 - **Why?** The mulberry bush didn't produce fruit this season, so birds are eating blueberries instead. (4th Why)
 - **Why?** A late frost damaged the mulberry bushes. (5th Why)

You have learned that there are six stages to the data life cycle. Here's a recap:

1. **Plan:** Decide what kind of data is needed, how it will be managed, and who will be responsible for it.
2. **Capture:** Collect or bring in data from a variety of different sources.
3. **Manage:** Care for and maintain the data. This includes determining how and where it is stored and the tools used to do so.

4. **Analyze:** Use the data to solve problems, make decisions, and support business goals.
5. **Archive:** Keep relevant data stored for long-term and future reference.
6. **Destroy:** Remove data from storage and delete any shared copies of the data.

The ask phase

At the start of any successful data analysis, the data analyst:

- Takes the time to fully understand stakeholder expectations
- Defines the problem to be solved
- Decides which questions to answer in order to solve the problem

The prepare phase

In the prepare phase, the emphasis is on identifying and locating data you can use to answer your questions. In an upcoming course, you'll learn more about the different types of data and how to identify which kinds of data are most useful for solving a particular problem. You'll also discover why it's so important that data and results are objective and unbiased. In other words, any decisions made from an analysis should always be based on facts and be fair and impartial.

The process phase

In this phase, the aim is to refine the data. Data analysts find and eliminate any errors and inaccuracies that can get in the way of results. This usually means:

- Cleaning data
- Transforming data into a more useful format
- Combining two or more datasets to make information more complete
- Removing outliers (data points that could skew the information)

The prepare phase

In the prepare phase, the emphasis is on identifying and locating data you can use to answer your questions. In an upcoming course, you'll learn more about the different types of data and how to identify which kinds of data are most useful for solving a particular problem. You'll also discover why it's so important that data and results are objective and unbiased. In other words, any decisions made from an analysis should always be based on facts and be fair and impartial.

The process phase

In this phase, the aim is to refine the data. Data analysts find and eliminate any errors and inaccuracies that can get in the way of results. This usually means:

- Cleaning data
- Transforming data into a more useful format
- Combining two or more datasets to make information more complete
- Removing outliers (data points that could skew the information)

Spreadsheets	Databases
Accessed through a software application	Database accessed using a query language
Structured data in a row and column format	Structured data using rules and relationships
Organizes information in cells	Organizes information in complex collections
Provides access to a limited amount of data	Provides access to huge amounts of data
Manual data entry	Strict and consistent data entry
Generally one user at a time	Multiple users
Controlled by the user	Controlled by a database management system

To name a few others that sound similar but may not be the same role:

- Business analyst—analyzes data to help businesses improve processes, products, or services
- Data analytics consultant—analyzes the systems and models for using data
- Data engineer—prepares and integrates data from different sources for analytical use
- Data scientist—uses expert skills in technology and social science to find trends through data analysis
- Data specialist—organizes or converts data for use in databases or software systems
- Operations analyst—analyzes data to assess the performance of business operations and workflows