**What is Data Analysis?**



Data are everywhere nowadays. And with each passing year, the amount of data we are producing will only continue to increase.

There is a large amount of data available, but what do we do with all that data? How is it all used? And what does all that data mean?

It’s not much use if we just collect and store data in a spreadsheet or database and don't look at it, explor it, or research it.

Data analysts use tools and processes to derive meaning from data. They are responsible for collecting, manipulating, investigating, analyzing, gathering insights, and gaining knowledge from it.

This is one of the reasons data analysts are very high in demand: they play an integral role in business and science.

In this article, I will first go over what data analysis means as a term and explain why it is so important.

I will also break down the data analysis process and list some of the necessary skills required for conducting data analysis.

Here is an overview of what we will cover:

1. [What is data?](https://www.freecodecamp.org/news/what-is-data-analysis/#data-intro)
2. [What is data analysis?](https://www.freecodecamp.org/news/what-is-data-analysis/#data-analysis-intro)
3. [Why is data analysis important?](https://www.freecodecamp.org/news/what-is-data-analysis/#importance)
   1. [Effective customer targeting](https://www.freecodecamp.org/news/what-is-data-analysis/#targeting)
   2. [Measure success and performance](https://www.freecodecamp.org/news/what-is-data-analysis/#performance)
   3. [Problem solving](https://www.freecodecamp.org/news/what-is-data-analysis/#problem-solving)
4. [An overview of the data analysis process](https://www.freecodecamp.org/news/what-is-data-analysis/#overview)
   1. [Step 1: recognising and identifying the questions that need answering](https://www.freecodecamp.org/news/what-is-data-analysis/#step-1)
   2. [Step 2: collecting raw data](https://www.freecodecamp.org/news/what-is-data-analysis/#step-2)
   3. [Step 3: cleaning the data](https://www.freecodecamp.org/news/what-is-data-analysis/#step-3)
   4. [Step 4: analyzing the data](https://www.freecodecamp.org/news/what-is-data-analysis/#step-4)
   5. [Step 5: sharing the results](https://www.freecodecamp.org/news/what-is-data-analysis/#step-5)
5. [What skills are required for data analysis?](https://www.freecodecamp.org/news/what-is-data-analysis/#skills)
   1. [A good grasp of maths and statistics](https://www.freecodecamp.org/news/what-is-data-analysis/#maths)
   2. [Knowledge of SQL and Relational Databases](https://www.freecodecamp.org/news/what-is-data-analysis/#sql)
   3. [Knowledge of a programming language](https://www.freecodecamp.org/news/what-is-data-analysis/#programming-language)
   4. [Knowledge of data visualization tools](https://www.freecodecamp.org/news/what-is-data-analysis/#data-viz)
   5. [Knowledge of Excel](https://www.freecodecamp.org/news/what-is-data-analysis/#excel)

**What Is Data? Meaning and Definition of Data**

Data refers to collections of facts and individual pieces of information.

Data is vital for decision-making, planning, and even telling a story.

There are two broad and general types of data:

* Qualitative data
* Quantitative data

**Qualitative data** is data expressed in non-numerical characters.

It is expressed as images, videos, text documents, or audio.

This type of data can’t be measured or counted.

It is used to determine how people feel about something – it’s about people's feelings, motivations, opinions, perceptions and involves bias.

It is descriptive and aims to answer questions such as ‘Why’, ‘How’, and ‘What’.

Qualitative data is gathered from observations, surveys, or user interviews.

**Quantitative data** is expressed in numerical characters.

This type of data is countable, measurable, and comparable.

It is about amounts of numbers and involves things such as quantity and the average of numbers.

It aims to answer questions such as ‘How much, ‘How many’, ‘How often’, ‘and 'How long’.

The act of collecting, analyzing, and interpreting quantitative data is known as performing statistical analysis.

Statistical analysis helps uncover underlying patterns and trends in data.

**What Is Data Analysis? A Definition For Beginners**

Data analysis is the act of turning raw, messy data into useful insights by cleaning the data up, transforming it, manipulating it, and inspecting it.

The insights gathered from the data are then presented visually in the form of charts, graphs, or dashboards.

The insights discovered can help aid the company’s or organization’s growth. Decision-makers will be able to come to an actionable conclusion and make the right business decisions.

Extracting knowledge from raw data will help the company/organization take steps towards achieving greater customer reach, improving performance, and increasing profit.

At its core, data analysis is about identifying and predicting trends and figuring out patterns, correlations, and relationships in the available data, and finding solutions to complex problems.

**Why Is Data Analysis Important?**

Data equals knowledge.

This means that data analysis is integral for every business.

It can be useful and greatly beneficial for every department, whether it's administration, accounting, logistics, marketing, design, or engineering, to name a few.

Below I will explain why exploring data and giving data context and meaning is really important.

**Data Analysis Improves Customer Targeting**

By analyzing data, you understand your competitors, and you will be able to match your product/service to the current market needs.

It also helps you determine the appropriate audience and demographic best suited to your product or service.

This way, you will be able to come up with an effective pricing strategy to make sure that your product/service will be profitable.

You will also be able to create more targeted campaigns and know what methods and forms of advertising and content to use to reach your audience directly and effectively.

Knowing the right audience for your product or service will transform your whole strategy. It will become more customer-oriented and customized to fit customers' needs.

Essentially, with the appropriate information and tools, you will be able to figure out how your product or service can be of value and high quality.

You'll also be able to make sure that your product or service helps solve a problem for your customers.

This is especially important in the product development phases since it cuts down on expenses and saves time.

**Data Analysis Measures Success and Performance**

By analyzing data, you can measure how well your product/service performs in the market compared to others.

You are able to identify the stronger areas that have seen the most success and desired results. And you will be able to identify weaker areas that are facing problems.

Additionally, you can predict what areas could possibly face problems before the problem actually occurs. This way, you can take action and prevent the problem from happening.

Analyzing data will give you a better idea of what you should focus more on and what you should focus less on going forward.

By creating performance maps, you can then go on to set goals and identify potential opportunities.

**Data Analysis Can Aid Problem Solving**

By performing data analysis on relevant, correct, and accurate data, you will have a better understanding of the right choices you need to make and how to make more informed and wiser decisions.

Data analysis means having better insights, which helps improve decision-making and leads to solving problems.

All the above will help a business grow.

Not analyzing data, or having insufficient data, could be one of the reasons why your business is not growing.

If that is the case, performing data analysis will help you come up with a more effective strategy for the future.

And if your business is growing, analyzing data will help it grow even further.

It will help reach its full potential and meet different goals – such as boosting customer retention, finding new customers, or providing a smoother and more pleasant customer experience.

**An Overview Of The Data Analysis Process**

**Step 1: Recognising and Identifying The Questions That Need Answering**

The first step in the data analysis process is setting a clear objective.

Before setting out to gather a large amount of data, it is important to think of why you are actually performing the data analysis in the first place.

What problem are you trying to solve?

What is the purpose of this data analysis?

What are you trying to do?

What do you want to achieve?

What is the end goal?

What do you want to gain from the analysis?

Why do you even need data analysis?

At this stage, it is paramount to have an insight and understanding of your business goals.

Start by defining the right questions you want to answer and the immediate and long-term business goals.

Identify what is needed for the analysis, what kind of data you would need, what data you want to track and measure, and think of a specific problem you want to solve.

**Step 2: Collecting Raw Data**

The next step is to identify what type of data you want to collect – whether it will be qualitative (non-numerical, descriptive ) or quantitative (numerical).

The way you go about collecting the data and the sources you gather from will depend on whether it is qualitative or quantitative.

Some of the ways you could collect relevant and suitable data are:

* By viewing the results of user groups, surveys, forms, questionnaires, internal documents, and interviews that have already been conducted in the business.
* By viewing customer reviews and feedback on customer satisfaction.
* By viewing transactions and purchase history records, as well as sales and financial figure reports created by the finance or marketing department of the business.
* By using a customer relationship management system (CRM) in the company.
* By monitoring website and social media activity and monthly visitors.
* By monitoring social media engagement.
* By tracking commonly searched keywords and search queries.
* By checking which ads are regularly clicked on.
* By checking customer conversion rates.
* By checking email open rates.
* By comparing the company’s data to competitors using third-party services.
* By querying a database.
* By gathering data through open data sets using web scraping. [Web scraping](https://www.freecodecamp.org/news/how-to-scrape-websites-with-python-2/) is the act of extracting and collecting data and content from websites.

**Step 3: Cleaning The Data**

Once you have gathered the data from multiple sources, it is important to understand the structure of that data.

It is also important to check if you have gathered all the data you needed and if any crucial data is missing.

If you used multiple sources for the data collection, your data will likely be unstructured.

Raw, unstructured data is not usable. Not all data is necessarily good data.

Cleaning data is the most important part of the data analysis process and one on which data analysts spend most of their time.

Data needs to be cleaned, which means correcting errors, polishing, and sorting through the data.

This could include:

* Looking for [outliers](https://www.freecodecamp.org/news/what-is-an-outlier-definition-and-how-to-find-outliers-in-statistics/) (values that are unusually big or small).
* Fixing typos.
* Removing errors.
* Removing duplicate data.
* Managing inconsistencies in the format.
* Checking for missing values or correcting incorrect data.
* Checking for inconsistencies
* Getting rid of irrelevant data and data that is not useful or needed for the analysis.

This step will ensure that you are focusing on and analyzing the correct and appropriate data and that your data is high-quality.

If you analyze irrelevant or incorrect data, it will affect the results of your analysis and have a negative impact overall.

So, the accuracy of your end analysis will depend on this step.

**Step 4: Analyzing The Data**

The next step is to analyze the data based on the questions and objectives from step 1.

There are four different data analysis techniques used, and they depend on the goals and aims of the business:

* **Descriptive Analysis**: This step is the initial and fundamental step in the analysis process. It provides a summary of the collected data and aims to answer the question: “**What** happened?”. It goes over the key points in the data and emphasizes what has already taken place.
* **Diagnostic Analysis**: This step is about using the collected data and trying to understand the cause behind the issue at hand and identify patterns. It aims to answer the question: “**Why** has this happened?”.
* **Predictive Analysis**: This step is about detecting and predicting future trends and is important for the future growth of the business. It aims to answer the question: “**What is likely to happen** in the future?
* **Prescriptive Analysis:** This step is about gathering all the insights from the three previous steps, making recommendations for the future, and creating an actionable plan. It aims to answer the question: “**What needs to be done?**”

**Step 5: Sharing The Results**

The last step is to interpret your findings.

This is usually done by creating reports, charts, graphs, or interactive dashboards using data visualization tools.

All the above will help support the presentation of your findings and the results of your analysis to stakeholders, business executives, and decision-makers.

Data analysts are storytellers, which means having strong communication skills is important.

They need to showcase the findings and present the results in a clear, concise, and straightforward way by taking the data and creating a narrative.

This step will influence decision-making and the future steps of the business.

**What Skills Are Required For Data Analysis?**

**A Good Grasp Of Maths And Statistics**

The amount of maths you will use as a data analyst will vary depending on the job. Some jobs may require working with maths more than others.

You don’t necessarily need to be a math wizard, but with that said, having at least a fundamental understanding of math basics can be of great help.

Here are some math courses to get you started:

* [College Algebra – Learn College Math Prerequisites with this Free 7-Hour Course](https://www.freecodecamp.org/news/learn-algebra-to-improve-your-programming-skills/)
* [Precalculus – Learn College Math Prerequisites with this Free 5-Hour Course](https://www.freecodecamp.org/news/precalculus-learn-college-math-prerequisites-with-this-free-5-hour-course/)
* [Math for Programmers Course](https://www.freecodecamp.org/news/maths-for-programmers/)

Data analysts need to have good knowledge of statistics and probability for gathering and analyzing data, figuring out patterns, and drawing conclusions from the data.

To get started, take an intro to statistics course, and then you can move on to more advanced topics:

* [Learn College-level Statistics in this free 8-hour course](https://www.freecodecamp.org/news/free-statistics-course/)
* [If you want to learn Data Science, take a few of these statistics classes](https://www.freecodecamp.org/news/if-you-want-to-learn-data-science-take-a-few-of-these-statistics-classes-9bbabab098b9#.esdiw8wnk)

**Knowledge of SQL and Relational Databases**

Data analysts need to know how to interact with relational databases to extract data.

A database is an electronic storage localization for data. Data can be easily retrieved and searched through.

A relational database is structured in format and all data items stored have pre-defined relationships with each other.

SQL stands for **S**tructured **Q**uery **L**anguage and is the language used for querying and interacting with relational databases.

By writing SQL queries you can perform CRUD (Create, Read, Update, and Delete) operations on data.

To learn SQL, check out the following resources:

* [SQL Commands Cheat Sheet – How to Learn SQL in 10 Minutes](https://www.freecodecamp.org/news/learn-sql-in-10-minutes/)
* [Learn SQL – Free Relational Database Courses for Beginners](https://www.freecodecamp.org/news/learn-sql-free-relational-database-courses-for-beginners/)
* [Relational Database Certification](https://www.freecodecamp.org/learn/relational-database/)

**Knowledge Of A Programming Language**

To further organize and manipulate databases, data analysts benefit from knowing a programming language.

Two of the most popular ones used in the data analysis field are Python and R.

Python is a general-purpose programming language, and it is very beginner-friendly thanks to its syntax that resembles the English language. It is also one of the most used technical tools for data analysis.

Python offers a wealth of packages and libraries for data manipulation, such as Pandas and NumPy, as well as for data visualization, such as Matplotlib.

To get started, [first see how to go about learning Python as a complete beginner](https://www.freecodecamp.org/news/how-to-learn-python/).

Once you understand the fundamentals, you can move on to learning about Pandas, NumPy, and Matplotlib.

Here are some resources to get you started:

* [How to Get Started with Pandas in Python – a Beginner's Guide](https://www.freecodecamp.org/news/python-pandas-functions/)
* [The Ultimate Guide to the Pandas Library for Data Science in Python](https://www.freecodecamp.org/news/the-ultimate-guide-to-the-pandas-library-for-data-science-in-python/)
* [The Ultimate Guide to the NumPy Package for Scientific Computing in Python](https://www.freecodecamp.org/news/the-ultimate-guide-to-the-numpy-scientific-computing-library-for-python/)
* [Learn NumPy and start doing scientific computing in Python](https://www.freecodecamp.org/news/numpy-python-tutorial/)
* [How to Analyze Data with Python, Pandas & Numpy - 10 Hour Course](https://www.freecodecamp.org/news/how-to-analyze-data-with-python-pandas/)
* [Matplotlib Course – Learn Python Data Visualization](https://www.freecodecamp.org/news/matplotlib-course-learn-python-data-visualization/)
* [Python Data Science – A Free 12-Hour Course for Beginners. Learn Pandas, NumPy, Matplotlib, and More.](https://www.freecodecamp.org/news/python-data-science-course-matplotlib-pandas-numpy/)

R is a language used for statistical analysis and data analysis. That said, it is not as beginner-friendly as Python.

To get started learning it, check out the following courses:

* [R Programming Language Explained](https://www.freecodecamp.org/news/r-programming-language-explained/)
* [Learn R programming language basics in just 2 hours with this free course on statistical programming](https://www.freecodecamp.org/news/r-programming-course/)

**Knowledge of data visualization tools**

Data visualization is the graphical interpretation and presentation of data.

This includes creating graphs, charts, interactive dashboards, or maps that can be easily shared with other team members and important stakeholders.

Data visualization tools are essentially used to tell a story with data and drive decision-making.

One of the most popular data visualization tools used is Tableau.

To learn Tableau, check out the following course:

* [Tableau for Data Science and Data Visualization - Crash Course](https://www.freecodecamp.org/news/tableau-for-data-science-and-data-visualization-crash-course/)

**Knowledge of Excel**

Excel is one of the most essential tools used in Data analysis.

It is used for storing, structuring, and formatting data, performing calculations, summarizing data and identifying trends, sorting data into categories, and creating reports.

You can also use Excel to create charts and graphs.

To learn how to use Excel, check out the following courses:

* [Learn Microsoft Excel - Full Video Course](https://www.freecodecamp.org/news/learn-microsoft-excel/)
* [Excel Classes Online – 11 Free Excel Training Courses](https://www.freecodecamp.org/news/excel-classes-online-free-excel-training-courses/)
* [Data Analysis with Python for Excel Users Course](https://www.freecodecamp.org/news/data-analysis-with-python-for-excel-users-course/)

**Conclusion**

This marks the end of the article – thank you so much for making it to the end!

Hopefully this guide was helpful, and it gave you some insight into what data analysis is, why it is important, and what skills you need to enter the field.

Thank you for reading!