

Introduction

Data — A Collection of facts, can include numbers, pictures, videos, words, measurements, observations, & more. More & more data is being created every second.

Data Analysis — The collection, transformation, & organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

Data Analyst — Someone who collects, transforms, & organizes data in order to help make informed decisions. Data analysts can tap into the power of data to do all kinds of amazing things. With data, they can gain valuable insights, verify their theories or assumptions, better understand opportunities & challenges, support an objective, help make a plan, & much more.

Phases of the Data Analysis Process:

- Ask
- Prepare
- Process
- Analyze
- Share
- Act

Why do Businesses need Data Analysis?

- Improve processes
- Identify opportunities & trends
- Launch new products
- Serve customers
- Make thoughtful decisions

For businesses to be on top of their competition they need to have a good sense of their data, hence they hire data analysts to control waves of data they collect every day, make sense of it, & then draw conclusions or make predictions.

This is the process of turning data into insights, & it's how analysts help businesses put all their data to good use.

Data Ecosystem — Made up of various elements that interact with one another in order to produce, manage, store, organize, analyze, & share data. These elements include hardware & software tools, & the people who use them.

Cloud — A place to keep data online, rather than a computer hard drive. Here data is accessed over the internet.

Data Science — Creating new ways of modeling & understanding the unknown by using raw data.

Data Analytics — Science of Data. Umbrella term for data, data analysis & data ecosystem.

Myth: Someone who works in data should know the everything of data

This is untrue, the universe of data is expanded well far, It's really difficult for one person to know & be the everything of data. Hence there is an important need to specialize. And one must pick their specialization based on which type of impact suits their personality. Data science — the discipline of making data useful, is an umbrella term that encompasses three disciplines. They are separated by how many decisions you know you want to make before you begin with them:

- **Machine learning** — Performance is the excellence of the machine learning & AI engineer.
- **Statistics** — The excellence of statistics is rigor. Statisticians are essentially philosophers, epistemologists. They are very, very careful about protecting decision-makers from coming to the wrong conclusion.
- **Analytics** — The excellence of an analyst is speed. How quickly you surf through vast amounts of data to explore it & discover the gems, the beautiful potential insights that are worth knowing about & bringing to your decision-maker.

Use Data to make Effective Decisions

Data Driven Decision Making — Using facts to guide business strategy. More likely to lead to successful outcomes.

The first step in data driven decision making is figuring out the business need, aka the problem that needs to be solved.

Once the problem is defined, a data analyst finds data, analyzes it & uses it to uncover trends patterns & relationships.

Example 1 — A music or movie streaming service.

How do these companies know what people want to watch or listen to, & how do they provide it? Using data-driven decision making they gather information about what their customers are currently listening to, analyze it, then use the insights they've gained to make suggestions for things people will most likely enjoy in future. This keeps customers happy & will come back for more, which in turn means more revenue for the company.

Example 2 — Rise of E-Commerce, over time data showed people's preferences were changing.

To get the most out of data-driven decision making, it's important to include insights from people who are familiar with the business problem aka **Subject Matter Experts**, they have the ability to look at the results of data analysis & identify any inconsistencies, make sense of gray areas, & eventually validate choices being made.

As a data analyst, you play a key role in empowering these organizations to make data-driven decisions, which is why it's so important for you to understand how data plays a part in the decision making process.

Skills & Characteristics key to a Career as a Data Analyst

Analytical Skills — Qualities & characteristics associated with solving problems using facts. There are a lot of aspects of analytical skills, but here are the 5 essential skills of a data analyst:

1. **Curiosity** — Wanting to learn something, seek out new challenges & experiences, leads to knowledge.

2. **Understanding context** — Context is the condition in which something exists or happens. Basically, finding patterns from context & able to identify things out of context.
3. **Having a technical mindset** — Ability to break things down into smaller steps or pieces & work with them in an orderly & logical way.
4. **Data design** — How you organize information, has to do with actual databases.
5. **Data strategy** — The management of the people, process & tools used in data analysis.

How these 5 Skills help you tap into all the Potential of Data-driven Decision Making:

1. **Curiosity** — The more you learn about the power of data, the more curious you become, you start to see patterns & relationships in everyday.
2. **Context** — By making predictions, research answers, & eventually draw conclusions about what they've discovered.
3. **Having a technical mindset** — Pure Gut feelings, Data analysts train themselves to build on those feelings & use a more technical approach to explore them. They do this by always seeking out the facts, putting them to work through analysis, & using the insights they gain to make informed decisions.
4. **Data design** — Designing your data so that is organized in a logical way makes it easy for analysts to access, understand, & make the most of available information.
5. **Data strategy** — Manage people by making sure they know how to use the right data to find solutions to the problem you're working on. For processes, it's about making sure the path to that solution is clear & accessible. For tools, you make sure the right technology is being used for the job.

Analytical Thinking — Identifying & defining a problem & then solving it by using data in an organized, step-by-step manner.

5 Aspects of Analytical Thinking:

1. **Visualization** — Graphical representation of information, it's important because visuals can help data analysts understand & explain information more effectively.
2. **Strategy** — With so much data on hand, having a strategic mindset is the key to staying focused & on track. Strategy also helps improve the quality & usefulness of data we collect.
3. **Problem-orientation** — Data analysts use a problem oriented approach in order to identify, describe & solve problems. Keeping the problem on top of the mind throughout the entire project.
4. **Correlation** — Being able to identify a correlation between two or more pieces of data. Correlation does not equal causation, i.e. just because two pieces of data are trending in the same direction, that does not necessarily mean they are related.
5. **Big-picture & detail-oriented thinking** — Being able to see the big picture as well as the details. Ex. Jigsaw-puzzle. Detail oriented thinking involves figuring out all the aspects that will help you execute a plan. Ex. the pieces that make up your puzzle. Most of us are naturally better at one or the other, but we can always develop skills to fit both pieces together.

Why is it important to think in different ways as a Data Analyst?

Versatile thinking, analytical thinking, Creative & Critical thinking. The more ways you think, the easier to think outside the box & come up with fresh ideas. In data analysis, solutions are almost never in front of you. You need to think critically to find out the right questions to ask, but you also need to think creatively to get new unexpected answers.

Some of the Questions Data Analysts ask when they are on a Hunt for a Solution

1> What is the root cause of the problem?

Root cause — The reason why the problem occurs. Ask, “why?” five times to reveal the root cause. By the time you reach the 5th “why” you would have most likely reveal the root cause.

2> Where are the Gaps in our Process?

Gap Analysis — A method for examining & evaluating how a process works currently in order to get where you want to be in future. The general approach to gap analysis is understanding where you are now compared to where you want to be. Then you can identify the gaps that exist between a current & future state & determine how to bridge them.

3> What did we not consider before?

This is a great way to think about what information or procedure might be missing from a process, so you can identify ways to make better decisions moving forward.

The way that a data analyst thinks & asks questions plays a big part in how businesses make decisions. That’s why analytical thinking & understanding the right questions can have such a huge impact on the overall success of a business.

Data Life Cycle

1. **Plan** — Happens well before the start of an analysis project. Business decides what kind of data it needs, how it will be managed throughout its life cycle, who will be responsible for it, & the optimal outcomes.
2. **Capture** — Data is collected from a variety of different sources & brought into the organization. Usually surveys. databases or datasets.
3. **Manage** — How we care for our data, how & where it’s stored, tools used to keep it safe & secure, & the actions taken to make sure that it’s maintained properly.
4. **Analyze** — In this phase, data is used to solve problems, make decisions, & support business goals.
5. **Archive** — Storing data in a place where it’s still available, but may not be used anymore.
6. **Destroy** — Unwanted data stored on multiple hard drives are wiped out, paper files are shredded. This is important to protect the company’s private information as well as private data about it’s customers.

Phases of the Data Analysis Process

It is the process of analyzing data. Categorized into 6 phases:

1. **Ask** — We define the problem to be solved & we make sure that we fully understand stakeholder expectations. **Stakeholders** hold a 'stake' in the project, people who have invested time & resources into a project & are interested in the outcome. Communicating with your stakeholders is key in making sure you stay engaged & on track throughout the project.
2. **Prepare** — This is where data analysts collect & store data they'll use for the upcoming analysis process.
3. **Process** — Here, data analyst's find & eliminate any errors & inaccuracies that can get in the way of results. This usually means data cleaning data, transforming it into a more useful format, combining two or more datasets to make information more complete & removing outliers (any data points that could skew the information). Also fix typos, inconsistencies, or missing data & inaccurate data.
4. **Analyze** — Analyzing the data you've collected involves using tools to transform & organize that information so that you can draw useful conclusions, make predictions & drive informed decision making. Tools data analysts use — Spreadsheets, Structured Query Languages, etc.
5. **Share** — Here, data analyst's interprets results & share them with others to help stakeholders make effective data-driven decisions. In this phase visualization is our best friend.
6. **Act** — Final phase when the business takes all insights you provided & puts them to work in order to solve the original business problem.

Data Analysis Tools

- **Spreadsheets** — Examples: Microsoft Excel, Google Sheets, Libre Office, etc. Spreadsheet is a digital worksheet. It stores, organizes, & sorts data. The usefulness of your data depends on how well it's structured. When data is put into a spreadsheet you can see patterns, group information & easily find the information you need. They also have some really useful features called formulas & functions. A **formula** is a set of instructions that perform a specific calculation using the data in a spreadsheet (Ex. +, -, *, /, Average, etc.) A **function** is a preset command that automatically performs a specific task using the data in a spreadsheet. i.e. a function is a simpler more efficient way of doing something that would normally take a lot of time. (Ex. MAX, MIN, COUNT, etc.)

A good, short tutorial to get started with spreadsheets in general.

- **Query Languages for Databases** — A computer programming language that allows you to retrieve & manipulate data from a database. Ex. Structured Query Language (SQL). Database is a collection of data stored in a computer.
- **Data Visualization tools** — Data visualization is the graphical representation of information. Ex. Graphs, Charts, Tables, etc. Most people process visuals more easily, hence visualization helps data analysts communicate their insights to others, in a more effective & compelling way. Makes it easier for stakeholders to draw conclusions, make decisions & come up with strategies. Some popular data visualization tools are Tableau, Power BI, Looker, etc.

Yes, I'm aware of the fact that we are looking at other resources, it's just faster to understand.

Fairness in Analysis

Fairness — Ensuring that your analysis doesn't create or reinforce bias. As a data analyst, you want to create systems that are fair & inclusive to everyone.

Sometimes conclusions based on data can be true & unfair. Let's say a company has a male dominant employees & there aren't many women employees. This company wants to see which employees are doing well so they start gathering data on employee performance. The data shows that women just aren't succeeding as often as men as their company. The conclusion, that they should hire fewer women.

But that's not a fair conclusion for the following reasons:

- It doesn't even consider all of the available data on company culture, so it paints an incomplete picture.
- It doesn't think about the other surrounding factors that impact the data, i.e., the conclusion doesn't include the difficulties women have trying to navigate a toxic work environment. If the company only looks at this conclusion, they won't acknowledge & address how harmful their culture is, they won't understand why women are set up to fail within it.

The conclusion that women aren't succeeding in this company is true, but it ignores the other systemic factors that are contributing to this problem.

An ethical data analyst could look at the data gathered, & conclude that the company culture is preventing women from succeeding, & the company needs to address these problems to boost performance. This conclusion paints a much more complete & fair picture.

As a data analyst, it's your responsibility to make sure your analysis is fair, & factors in the complicated social context that could create bias in your conclusion.