Data analysis process

overview of the data analysis process:

- **Define the problem**: Identify the business problem or question that needs to be addressed through data analysis.
- **Collect the data**: Gather the relevant data from various sources, such as databases, APIs, or files.
- **Data cleaning and preparation**: Clean and preprocess the data to remove errors, inconsistencies, or missing values, and transform it into a format suitable for analysis.
- **Data exploration and analysis**: **Explore the data** using various techniques, such as visualizations, statistical analysis, or machine learning, to gain insights into the data and identify patterns, trends, or relationships.
- **Interpret the results**: Interpret the results of the analysis and **draw conclusions** based on the insights gained.
- **Implement the solution**: Take action based on the insights gained and implement the solution to address the business problem.
- **Communicate the results**: Present the findings of the analysis in a clear and understandable way to stakeholders, using visualizations, reports, or dashboards.

problems that data analysts solve:

- Sales and marketing analysis
- Financial analysis
- Operational analysis
- Risk analysis
- Product development
- Business strategy

Data analysts use data to help organizations **make informed decisions** that can improve performance, increase profitability, **reduce costs**, and mitigate risks. They analyze large and complex data sets using various tools and techniques to identify patterns, trends, and relationships that can inform business decisions. Data analysts work across various industries and functions, including healthcare, finance, retail, and technology, among others.

Asking questions

Asking questions is a fundamental part of the data analysis process:

- It helps to clarify the problem being addressed, identify the relevant data, and generate insights that can inform decision-making.
- To ask the right questions, data analysts should start with **a clear problem statement** that defines the scope of the analysis and the desired outcomes.
- They should use data-driven questions that are **specific, measurable, and relevant to the problem being addressed.**
- Data analysts should consider the wider context of the problem, including the **business goals**, the industry, and the competitive landscape.
- Finally, data analysts should work closely with stakeholders to ensure that the questions being asked are relevant and aligned with the business goals.

Data wrangling VS exploratory data analysis (EDA)

They are two important steps in the data analysis process. Here's an overview of each step:

Data Wrangling:

Data wrangling, also known as data cleaning and preparation, involves transforming raw data into a format suitable for analysis.

This includes tasks such as removing duplicates, handling missing values, transforming data types, and merging data sets.

Data wrangling is important because it ensures that the data is accurate, complete, and ready for analysis.

Exploratory Data Analysis (EDA):

EDA is the process of analyzing and summarizing the main characteristics of a dataset to gain insights and identify patterns.

This step includes tasks such as visualizing the data using charts and graphs, calculating summary statistics, and identifying outliers.

EDA is important because it gives information , which can help in the subsequent steps in the data analysis process, such as modeling and hypothesis testing.