

# Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an approach in data analysis that focuses on summarizing, visualizing, and understanding the main characteristics and patterns of a dataset. It involves examining the data from multiple angles, generating hypotheses, and identifying potential patterns or relationships between variables.

In short, EDA reveals the true nature of data.



## The primary goal of EDA

- To gain insights into the data,
- Uncover hidden patterns,
- Detect outliers or anomalies,
- Determine the most appropriate statistical techniques or models for further analysis.



EDA is typically performed in the early stages of a data analysis project to get a preliminary understanding of the data before applying more complex statistical methods.

## STEP's OF EDA

- **Data Collection:** Gather the relevant dataset that you want to analyze.
- **Data Familiarization:** Get familiar with the dataset by understanding its structure and variables. This step involves reading the data into an appropriate software environment (e.g., Python, R, or Excel).
- **Data Cleaning:** Clean the data to handle any issues such as missing values, outliers & much more.
- **Data Visualization:** Create visual representations of the data using various plots, charts, or graphs. This step helps to identify patterns, trends, and relationships between variables. Common visualization techniques include histograms, scatter plots, box plots, line plots, and heatmaps.
- **Descriptive Statistics:** Calculate summary statistics such as mean, median, mode, standard deviation, and range to describe the central tendencies, dispersion, and distribution of the variables. This step provides a numerical understanding of the data.



- **Exploratory Data Analysis Techniques:** Apply specific EDA techniques based on the characteristics of the dataset and research questions.
- **Hypothesis Generation:** Formulate initial hypotheses or research questions based on the insights gained from the EDA process. These hypotheses can guide further analysis and modeling.
- **Iterative Analysis:** Perform iterative analysis by repeating the previous steps as necessary. This includes refining data cleaning procedures, exploring different visualizations, and conducting additional analyses to gain deeper insights into the data.
- **Documentation:** Document the findings, insights, and decisions made during the EDA process.