

Step 04

# DATA EXPLORATION

- What is data exploration
- Need of data exploration
- Various tasks performed in it



## What is data exploration?

- Data exploration is the process of analyzing and understanding the characteristics of a dataset.
- It involves the application of various statistical and visualization techniques to uncover patterns, trends, anomalies, and relationships among the data.

## What is the need for it?

Data exploration helps in

- Identifying patterns
- Understanding the distribution of the data
- Understand relationships
- Determine feature importance
- Improve data quality
- Selecting the appropriate machine learning algorithms and techniques.

Here are some of the most important tasks that should be performed during data exploration,

### **Univariate Analysis**

- This involves analyzing individual variables in the dataset to gain insights into their distributions, central tendencies, and spread.
- By using techniques such as histograms, box plots, and density plots, we can identify any outliers, skewness, or other anomalies that may impact our modeling approach.

### **Bivariate Analysis**

- This involves analyzing the relationship between two variables in the dataset.
- By using techniques such as scatter plots and correlation matrices, we can identify any patterns or trends in the data and determine which variables may be useful for our modeling approach.

## **Multivariate Analysis**

- This involves analyzing the relationship between multiple variables in the dataset.
- By using techniques such as heat maps and pair plots, we can identify any complex relationships or interactions between variables that may impact our modeling approach.

## **Dimensionality Reduction**

- This involves reducing the number of variables in the dataset while preserving the most important information.
- By using techniques such as principal component analysis (PCA), we can identify any redundant or irrelevant variables that may be removed from our modeling approach.

## **Data Visualization**

- This involves creating visual representations of the data that can help us gain insights and communicate our findings to others.
- By using techniques such as scatter plots, bar charts, and heat maps, we can identify patterns and trends in the data and communicate our findings effectively.

By performing these tasks, we can gain a deeper understanding of the data, identify important variables, and improve the accuracy of our model.

Finally, Data exploration is a secret weapon that can make all the difference in our machine learning journey.

Happy learning!😊

# THANK YOU

@ Gangadhar Neelam