# What is EDA?

Exploratory Data Analysis also known as EDA refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

It is a good practice to understand the data first and try to gather as many insights as possible from it. The primary goal is to make sense of the data before getting them dirty and ready for conclusions and descriptions.

The objectives of EDA are to:

- Suggest hypotheses about the cause of observed phenomena
- Assess assumptions on which statistical inference will be based
- Support the selection of appropriate statistical tools and techniques
- Provide a basis for further data collection through surveys or experiments

Many EDA techniques have been adopted into data mining. They are also being taught to young students as a way to introduce them to statistical thinking.

## Technique and Tools

There are a number of tools that are useful for EDA, but EDA is characterized more by the attitude taken than by particular techniques.

Typical graphical techniques used in EDA are:

- Box plot
- Histogram
- Multi-vari chart
- Run chart
- Pareto chart
- Scatter plot
- Stem-and-leaf plot
- Parallel coordinates

References

https://en.wikipedia.org/wiki/Exploratory_data_analysis

https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15