About the Exploratory Data Analysis (EDA)

The initial phase in the data analysis process is EDA-Exploratory Data Analysis. It enables us to comprehend the data we are dealing with by explaining and summarizing the dataset's essential properties, frequently using visual approaches like as bar and pie charts, histograms, boxplots, scatterplots, heat maps, and many more.

What is the significance of EDA?

- Gain maximum understanding into a dataset (listen to your data)
- Discover the underlying structure/patterns
- Identify outliers and abnormalities
- Extract and choose critical variables
- Improve computational efficiency
- Put underlying assumptions (such as business intuition) to the test.

Furthermore, Descriptive Statistics may help us explore and explain the dataset's properties with all of its attributes, providing insights and efficient numeric summaries of the data.

Data exploration is the earliest stage of data analysis and uses statistical methods and data visualization tools to identify the characteristics and general patterns of the data collection.

To visually explore data sets, look for similarities, patterns, and outliers, as well as to determine the links between various variables, raw data is often analyzed during exploration using a combination of manual workflows and automated data-exploration approaches.

Exploratory data analysis, a statistical approach used to examine raw data sets in search of their general features, is another name for this that is occasionally used.

Being visual learners, humans can process visual information considerably more quickly than numerical information. As a result, it can be difficult for data scientists to examine thousands of rows of data points and inference can be obtained without assistance.

It is possible to identify relationships or anomalies through the use of data visualization tools and elements including colors, forms, lines, graphs, and angles.

Few examples of tools for data exploration:

Software for data visualization and business intelligence platforms like Microsoft Power BI, Qlik, and Tableau are examples of tools for data exploration. Businesses can combine diverse, disparate data sources for quicker data exploration by integrating regression functionality, data profiling, and visualization capabilities into open source data exploration tools that are readily available.Knime, OpenRefine, NodeXL, Pentaho, R programming, and RapidMiner are a few popular open source (Jacqueline Biscobing, Katie Terrell Hanna, 2021) technologies.

