

Exploratory Data Analysis

Exploratory data analysis (EDA) is used by data scientists to study and explore data sets and describe their primary properties, frequently incorporating data visualisation techniques. In order to find trends, outliers, and verify hypotheses with the use of summary statistics and visual representations, exploratory data analysis is an essential first step. The goal of exploratory data analysis is to get a thorough familiarity with the data and become familiar with its many properties. It helps you acquire a better sense of your data and discover actionable patterns within it. EDA simplifies data modelling by making it easy to understand the structure of a dataset. EDA's major objective is to eliminate redundant information from data sets. It helps find outliers in the data so they can be eliminated and the data cleaned up. It also helps us understand the interplay between the variables, giving us a more complete picture of the information and opening up opportunities for further research. It also helps in assessing the statistical metrics of the dataset. Inaccuracies in machine learning models may be caused by outliers, or unusual events, in a dataset. The data collection may also include some values that are duplicated or missing entirely. All of the undesirable characteristics of the dataset may be fixed or removed with the help of EDA.

DIFFERENT STEPS IN EXPLORATORY DATA ANALYSIS (EDA)

- Data Collection
- Data Cleaning
- Univariate Non-graphical
- Multivariate Non-graphical
- Univariate graphical
- Multivariate graphical
- Check for duplicate values.
- Unique values in the dataset.
- Visualize the unique counts.
- Know the datatypes
- Find the null values.
- Replace the null values
- Check for correlation of variables