



What is Exploratory Data Analysis (EDA) and How Does it Work?

Introduction :

Exploratory Data Analysis (EDA) is a process of describing the data by means of statistical and visualization techniques in order to bring important aspects of that data into focus for further analysis. This involves inspecting the dataset from many angles, describing & summarizing it without making any assumptions about its contents.

“Exploratory Data analysis is an attitude, a state of flexibility, a willingness to look for those things that we believe are not there, as well as those we believe to be there” – John W. Tukey

EDA is a significant step to take before diving into statistical modeling or machine learning, to ensure the data is really what it is claimed to be and that there are no obvious errors. It should be part of data science projects in every organization.

What is Exploratory Data Analysis?

1. **Look at the Data:** You start by gathering information about the data you have. How many rows and columns are there? What kind of information does each column contain?
2. **Clean the Data:** Sometimes, data can be messy. There might be missing values, or some values might be wrong. You clean up the data by fixing these issues.
3. **Make Summaries:** You summarize the data to get a general idea of what's in it. You might find out things like the average value, the most common value, or how the values are spread out.
4. **Visualize the Data:** It's helpful to see the data in graphs or charts. This way, you can spot trends or patterns more easily. Python has libraries like Pandas, NumPy, and Matplotlib that are commonly used for this purpose in exploratory data analysis Python.
5. **Ask Questions:** As you explore the data, you might come up with questions. Why is one part of the data different from the rest? Are there any relationships between different parts of the data?
6. **Find Answers:** You try to answer the questions you've asked by digging deeper into the data. This might involve doing more analysis or creating models.

Types of EDA

▶ Here are five types of EDA techniques:

1. **Univariate Analysis:** In EDA Analysis, univariate analysis examines individual variables to understand their distributions and summary statistics.
2. **Bivariate Analysis:** This aspect of EDA explores the relationship between two variables, uncovering patterns through techniques like scatter plots and correlation analysis.
3. **Visualization Techniques:** EDA relies heavily on visualization methods to depict data distributions, trends, and associations using various charts and graphs.
4. **Outlier Detection:** EDA involves identifying outliers within the data, anomalies that deviate significantly from the rest, employing tools such as box plots and z-score analysis.
5. **Statistical Tests:** EDA often includes performing statistical tests to validate hypotheses or discern significant differences between groups, adding depth to the analysis process.

Project explanation:

- ▶ In our assignment, we have done all these things like cleaning the data, removing outliers and removing null values which are more than 40%.
- ▶ Data is well explained such as comparison of categorical data and numerical data.
- ▶ Numerical and numerical comparisons .
- ▶ And also comparisons of categorical and categorical .
- ▶ To understand the data better, we have a good explanation in our Jupyter Notebook.