EXPT. NO. | NAME | M T W T F S S | Page No.: 13 | Date: 11/2/25 | YOUVA

EXPERIMENT NO. 5

AIM:

THEORY:

Exploratory Data Analysis (EDA)

Our practical session centered around performing a comprehensive exploratory data analysis on a campus placement dataset. EDA involves thoroughly investigating the dataset's structure, patterns and relationships before embarking on formal modeling. This exploratory phase helps identify potential data quality issues, formulate hypotheses and make informed decisions about feature engineering & model selection.

Theoretical Framework of EDA:

It leverages the principles of descriptive statistics and data visualization to gain insights into the data. Descriptive statistics provide a quantitative summary of the data's characteristics. Data visualization, through techniques like histograms, scatter plots and box plots, enables a visual

understanding of the data's distribution, relationship and potential anomalies.

1. Data Acquisition and Preliminary Examination:

pd.read_csv(): To import the dataset from a csv file into a pandas DataFrame, enabling efficient manipulation and analysis.

data.head(): To examine the first few rows of the DataFrame, providing a preliminary understanding of its structure, data types and content.

2. Descriptive Statistics and Data Cleansing:

data.describe(): To generate descriptive statistics including measures of central tendency, dispersion and qualities for each numerical feature in the dataset, providing a comprehensive overview of the data's characteristics.

data.drop(): To remove irrelevant columns, such as 'sl-no' from the DataFrame, ensuring that the analysis focuses on relevant features and improving computational efficiency.

Teacher's Signature:

3. Univariate Analysis : Feature-wise Explanation:

sns.countplot() : To visualize the distribution & frequency of categorical features like 'status' (employment status), providing insights into their prevalence & potential imbalances.

sns.displot() : To examine the distribution of numerical features like 'ssc-p' (ssc percentage), identifying potential outliers, skewness and other distributional characteristics, aiding in understanding the data's underlying patterns.

4. Bivariate Analysis : Feature Relationships :

sns.catplot() : To visualize the relationship between numerical features like 'ssc-p' and 'status', revealing patterns and potential correlations, helping understand the influence of numerical features on the target variable.

sns.countplot(hue=) : To explore the relationship between categorical features like 'gender' and 'status', providing insights into potential dependencies and associations between categorical variables and the target variable.

Teacher's Signature:

EXPT. NO. | NAME | Page No.: 16 | Date: 11/12/25 | M T W T F S S | YOUVA

5. Multivariate Analysis : Feature interactions :

sns.pairplot() : To generate a matrix of scatter plot, showcasing the relationships between numerical features like 'ssc-p', 'hsc-p' and 'degree-p', while considering the target variable 'status' through colour-coding, providing a deeper understanding of the interplay between features and their potential impact on the target variable.

6. Feature Engineering and selection : Data Refinement :

data.map() : To map categorical features like 'gender' and 'workex' to numerical values, preparing the data for potential modeling tasks and ensuring compatibility with various machine learning algorithms.

CONCLUSION :

Our EDA journey involved a systematic and technically rigorous process of data inspection, visualization and transformation. Utilizing functions, we gained valuable insights into the campus placement dataset, identified key features, uncovered relationships between variables and prepared the data for further analysis and model building.

Teacher's Signature: