Assignment

Dataset

Take any dataset of your choice that contains multiple numerical and categorical variables representing different attributes. The dataset may include missing values, duplicate entries, and potential outliers, which need to be addressed in this assignment.

Tasks to be Performed

1. Data Cleaning:

- Load the dataset and inspect its structure.
- Handle missing values using imputation or removal techniques.
- Identify and remove duplicate records.
- Detect and treat outliers using appropriate statistical methods.
- Standardize categorical values (e.g., fixing typos or formatting inconsistencies).

2. Exploratory Data Analysis (EDA):

Univariate Analysis (Single-Variable Exploration)

- Summary statistics (mean, median, mode, variance, skewness, etc.).
- Frequency distributions for categorical variables.
- Histograms and box plots to visualize distributions.

Bivariate Analysis (Two-Variable Exploration)

- Correlation matrix to identify relationships between numerical variables.
- Scatter plots for continuous variable relationships.
- Bar plots, violin plots, and box plots to compare categorical and numerical variables.

Multivariate Analysis (Multiple Variables Exploration)

- Pair plots to analyze multiple relationships simultaneously.
- Heatmaps to visualize correlations among multiple variables.
- Grouped comparisons to identify combined effects of multiple features.

Expected Outcome

By the end of this assignment, students should be able to:

- Apply data cleaning techniques to prepare a dataset for analysis.
- Understand and visualize distributions of individual variables.
- Identify relationships between variables and interpret their significance.
- Use multivariate techniques to analyze complex interactions in the data.
- Prepare an elaborate report about your analysis.

Submission Process:

Google form has been created and shared to upload the github link that will contain all these information.

https://docs.google.com/forms/d/e/1FAIpQLSeY1S1QtlSyRI1S95-LoytSY5Jjh2dDkFdUTrYWnmtWI4Xi5Q/viewform?usp=header