

Exploratory Data Analysis (EDA) - Report of Findings

1. Import Libraries

Imported essential libraries: pandas, numpy, matplotlib, seaborn.

2. Load Dataset

Data loaded using pandas. Common formats include CSV, Excel, etc.

3. Basic Information

Checked dataset shape, column types, previewed top rows, and summarized numerical statistics.

4. Missing Values

Identified and visualized missing data. Options considered: dropping or imputing values.

5. Handle Missing Data

Filled missing values using appropriate methods (mean, median, mode) or dropped rows.

6. Explore Categorical Variables

Used value counts and count plots to understand category distributions.

7. Explore Numerical Variables

Used histograms and boxplots to analyze distributions and detect skewness or spread.

8. Variable Relationships

Used correlation matrix, heatmaps, and scatter plots to explore inter-feature relationships.

9. Detect Outliers

Detected outliers using boxplots and Z-score method.

10. Feature Engineering

Exploratory Data Analysis (EDA) - Report of Findings

Created new meaningful features by combining or transforming existing columns.

11. Summary & Insights

Key findings were noted, such as outliers, missing values, strong correlations, and potential features for modeling.