EDA PROJECT REPORT

1. Introduction

This report presents an end-to-end Exploratory Data Analysis (EDA) process on a sample dataset. It includes data cleaning, handling missing values, univariate and bivariate analysis, correlation study, and insights.

2. Data Loading

- Loaded dataset using pandas.

- Displayed head, shape, and data types.

3. Data Cleaning

• Missing Values Handling

- Numerical columns filled using median.

- Categorical columns filled using mode.

- Removed columns with more than 30% missing values.

• Duplicate Removal

- Duplicate rows removed.

4. Univariate Analysis

- Histograms plotted for numerical features.

- Countplots created for categorical features.

5. Bivariate Analysis

- Scatterplots to examine feature relationships.

- Boxplots to study target vs categorical features.

6. Correlation Analysis

- Generated correlation matrix.

- Heatmap used to visualize correlations.

7. Outlier Detection

- Detected outliers using boxplots.

- Applied IQR-based removal.

8. Feature Encoding

- Applied One-Hot Encoding for categorical columns.

9. Conclusion

The dataset was cleaned, analyzed, and prepared for modeling. Insights were extracted, and all preprocessing steps were completed.