Task 5: Exploratory Data Analysis (EDA)

Objective: Extract insights using visual and statistical exploration.

Tools Used: Python (Pandas, Matplotlib, Seaborn)

1. Data Overview

The dataset was loaded and examined using .info(), .describe(), and .isnull().sum(). Basic data structure, data types, and missing values were identified.

2. Univariate Analysis

Distribution plots (histograms, boxplots) were created for each numerical variable. Skewness and outliers were noted for continuous variables.

3. Bivariate Analysis

Relationships between two variables were explored using scatterplots and boxplots. Patterns between numerical and categorical features were analyzed.

4. Multivariate Analysis

A correlation heatmap and pairplot were generated using Seaborn. Strong positive and negative correlations were identified among variables.

5. Outlier Detection

Boxplots helped visualize the presence of outliers in numerical features. Extreme values were considered for potential removal or treatment.

6. Insights & Summary

The EDA revealed data distributions, missing patterns, correlations, and possible data quality issues. Key trends and anomalies were summarized to guide further modeling steps.

Conclusion:

The Exploratory Data Analysis provided a clear understanding of the dataset structure, variable relationships, and potential issues. Patterns and insights identified here will be essential for building predictive models and making data-driven decisions.