Exploratory Data Analysis (EDA) – Report

Objective

To analyze the dataset using statistical summaries and visual exploration to understand patterns, relationships, and anomalies.

1. Data Overview

Loaded dataset using Pandas

Reviewed structure using:

df.info() → Checked data types and missing values

df.describe() → Summary statistics for numerical features

df.isnull().sum() → Identified columns with missing data

Key Findings

Dataset contains both numerical and categorical variables.

Several columns have missing values that require imputation or removal.

2. Univariate Analysis

Histograms were plotted for Age, Fare, and other numerical columns.

Value counts were used for categorical features.

Observations:

Numerical data shows skewness in some variables.

Categorical variables show imbalanced distributions.

3. Bivariate Analysis

Scatterplots and pair plots (sns.pairplot) were used to understand relationships.

Heatmap (sns.heatmap) used to check correlation strength.

Observations:

Some variables show moderate correlation.

Outliers detected in numerical features (Age, Fare, etc.).

Certain features show clear separation patterns.

4. Statistical Insights

Mean, median, spread, and distribution shape were extracted.

Detected variance differences between groups.

Identified anomalies/outliers from boxplots.

5. Key Patterns Identified

Strong correlation between selected variables.

Patterns visible in distributions (e.g., skewed pricing, clustered ranges).

Certain categories influence outcomes more significantly.

6. Summary of Findings

Dataset requires missing-value treatment before modeling.

Clear trends exist between major numeric variables.

Outliers should be handled for better model performance.

Visual analysis exposed meaningful group differences.

conclusion:

The EDA helped us understand the data better and find important patterns.

We found missing values, outliers, and some useful relationships between features.

Before building any model, the data needs cleaning and proper preparation.