Exploratory Data Analysis (EDA) Report

A. Data Info & Description

Dataset contains 418 rows and 11 columns. All columns are float64 (standardized numeric form). No missing values detected. Features appear standardized with mean $\approx$ 0 and std $\approx$ 1.

B. Visualizations

Pairplot and Heatmap created. Heatmap shows correlations, pairplot illustrates scatterplots and distributions.

C. Relationships & Trends

Fare and Pclass are strongly negatively correlated. SibSp and Parch moderately correlated. Sex (encoded) shows clear grouping effects across other variables.

D. Distribution & Outliers

Histograms: Age symmetric, Fare skewed with outliers. Boxplots: Fare shows strong outliers; family variables mostly zero. Scatterplots: Fare vs Pclass negative trend.

E. Observations

- Age distribution is near normal. - Fare has extreme high values (wealthy passengers). - SibSp and Parch skewed right (most passengers alone). - Pairplot shows clusters by Sex and Pclass. - Heatmap confirms strongest correlation: Fare & Pclass.

F. Summary of Findings

1. Dataset clean, standardized, no missing values. 2. Numerical distributions show skewness in family counts and Fare. 3. Pclass and Fare are strongly related; family variables moderately related. 4. Majority passengers traveled alone; wealthy passengers paid high fares. 5. Gender and class play important roles in data distribution.