Exploratory Data Analysis (EDA) Report

1. Introduction

This report presents an Exploratory Data Analysis (EDA) performed on the Titanic dataset. The objective was to analyze patterns, relationships, and trends influencing passenger survival.

2. Data Cleaning

- Missing values in Age were replaced using the median.
- Missing values in Embarked were replaced using the mode.
- Cabin column was dropped due to excessive missing values.
- Duplicates were checked and removed if present.
- A new feature 'FamilySize' was created.

3. Key Visual Analysis

- Age distribution showed most passengers were between 20-40 years.
- Fare distribution was highly right-skewed with significant outliers.
- Females had significantly higher survival rates compared to males.
- First-class passengers had better survival probability.
- Passengers who paid higher fares were more likely to survive.
- Age showed weaker correlation with survival compared to socio-economic factors.

4. Correlation Analysis

Correlation heatmap analysis indicated that survival is positively correlated with Fare and negatively correlated with Passenger Class (Pclass). This suggests socio-economic status played a major role in survival probability.

5. Final Summary of Findings

- Gender was a strong determinant of survival.

- Higher socio-economic class increased survival probability.
- Fare showed strong positive relationship with survival.
- Age had moderate influence.
- Socio-economic factors were the primary drivers of survival.