Exploratory Data Analysis Report: Titanic Survival Dataset

1. Introduction

This report provides exploratory data analysis on the Titanic survival dataset. The goal is to understand patterns and identify key features influencing survival such as gender, class, age, and fare.

2. Dataset Overview

The cleaned dataset contains 708 rows and 11 columns. There are no missing values. Categorical features include Sex, Embarked, and Pclass. Numerical features include Age and Fare.

3. Key Observations

- Most passengers were male (Sex=1), and most embarked from Southampton.
- Majority belonged to Pclass 3.
- Age distribution is right-skewed; most were in their 20s and 30s.
- Fare values vary widely, with a few high-paying outliers.
- Correlation matrix shows that Sex, Pclass, and Fare influence survival.
- Female passengers (Sex=0) had a significantly higher survival rate.
- Passengers in 1st class had higher chances of survival than those in 2nd and 3rd classes.

4. Summary of Findings

- Gender and Pclass were the strongest indicators of survival.
- Women and passengers in higher classes had better survival chances.
- Age had some influence; younger passengers had slightly better survival.
- Fare correlated positively with survival, suggesting socio-economic impact.
- These findings confirm historical rescue biases and class-based survival outcomes.