

Titanic Dataset - EDA Report

Step-by-Step Exploratory Data Analysis (EDA) Summary

Total rows: 891, Columns: 12

Missing values - Age: 177, Cabin: 687, Embarked: 2

Survived counts: 0 = 549, 1 = 342

Overall survival rate: 38.38%

Survival rate by Sex: female = 74.2%, male = 18.89%

Survival rate by Pclass: 1 = 62.96%, 2 = 47.28%, 3 = 24.24%

Observations:

- Strong survival difference by Sex (females higher).
- Strong survival difference by Pclass (wealthier passengers survived more).
- Age has moderate effect; children more likely to survive.
- Fare distribution is right-skewed; high fares linked with survival.
- Cabin has high missingness, use derived features like 'HasCabin'.

Recommendations:

- Impute Age (median by Pclass/Sex or predictive model).
- Fill Embarked with mode.
- Engineer features (FamilySize, Title, IsAlone, HasCabin).
- Transform skewed Fare values.