Titanic Dataset - Visual & Statistical Exploration Report

1. Overview

The Titanic dataset contains 891 passenger records with features like survival, class, sex, age, fare, and embarkation port. The survival rate was around 38.4%. Missing data is significant in the 'Age' (~20% missing) and 'Cabin' (~77% missing) columns.

2. Key Insights

- Survival by Gender: Females had a ~74% survival rate compared to ~19% for males.
- Survival by Class: 1st class ~63%, 2nd class ~47%, 3rd class ~24%.
- Survival by Age: Children (<18) ~50%, Adults ~35-40%, Elderly (~23-26%).
- Correlation: Survival is most strongly associated with Sex (-0.54), Pclass (-0.34), and Fare (+0.26).
- Missing Data: 'Cabin' data is largely missing; 'Age' has substantial gaps; 'Embarked' missing only 2 records.

3. Other Observations

- Passengers from Cherbourg had higher survival (~55%) than those from Queenstown (~39%) or Southampton (~34%).
- Small families had slightly better survival rates than solo travelers or large families.
- Titles extracted from names show survival patterns: women and boys (Master, Miss, Mrs) fared better than men (Mr).

4. Conclusions

Women, children, and wealthier passengers had a clear survival advantage. First-class status and higher fares correlated with higher survival. These trends are consistent with the "women and children first" evacuation policy.