

Task 05: Exploratory Data Analysis

Name:- Swaraj Gavali

Date:-20/11/2025

Titanic Survival Analysis — Final Summary

The Titanic dataset provides a detailed view of 891 passengers, including attributes such as age, gender, class, fare, and survival status. This analysis aims to understand the key factors that influenced survival during the disaster by performing structured univariate and bivariate exploratory data analysis.

1. Univariate Insights

Initial exploration showed missing values in Age, Cabin, and Embarked. The dataset displayed an imbalanced mix of genders (more males), a higher proportion of passengers in 3rd class, and a right-skewed fare distribution. Age distribution revealed most passengers were between 20 and 40 years old.

2. Bivariate Findings

Multiple variables showed strong relationships with survival:

- Gender Influence

Survival probability was significantly higher for females. Most men did not survive, highlighting the "women first" rescue pattern.

- Passenger Class

Higher-class passengers (especially 1st class) had better survival outcomes. In contrast, 3rd class passengers faced the highest mortality, indicating socioeconomic status heavily affected access to lifeboats and safety.

- Age Impact

Younger passengers—especially children—showed slightly higher survival. Boxplots revealed survivors tended to be somewhat younger than non-survivors, though the relationship was weaker compared to gender and class.

- Fare Relationship

Fare had a positive association with survival. Individuals who paid higher fares generally survived more, again reinforcing the influence of class and access to resources.

- Combined Patterns (Pairplot)

Visual clustering in the pairplot showed survival is strongly tied to fare and class. Survived passengers cluster in lower Pclass values and higher fare values, indicating multivariate dependency.

- Correlation Heatmap

Key correlations observed:

Survival \leftrightarrow Fare: +0.26

Survival \leftrightarrow Pclass: -0.34

Fare \leftrightarrow Pclass: -0.55

These support the earlier findings that wealth and class had significant impact.

3. Conclusion

The analysis concludes that survival on the Titanic was not random. Gender, passenger class, and fare were the strongest determinants of survival, while age showed a moderate effect. These insights collectively highlight the interplay between social norms and socioeconomic status during the disaster.