

The Titanic dataset provides detailed information about passengers aboard the RMS Titanic, including their demographic characteristics, ticket class, fare, and survival status. The goal of this Exploratory Data Analysis (EDA) was to uncover meaningful patterns, trends, and relationships that influenced passenger survival during the sinking.

Dataset Overview

The dataset contains variables such as Passenger Class (Pclass), Sex, Age, Fare, Embarked, and Survival. A significant amount of missing values were found in the Age and Cabin columns, requiring careful handling during analysis.

Key Insights

1. Gender Influence:

One of the strongest predictors of survival was **Sex**. Female passengers had a much higher survival rate compared to male passengers. This reflects the “women and children first” evacuation policy followed during the disaster.

2. Class Influence:

Passenger Class (Pclass) showed a clear trend:

- **1st Class passengers survived the most**, followed by 2nd Class, while 3rd Class had the lowest survival rate.

Higher socioeconomic status appears strongly linked with better chances of survival.

3. Age Patterns:

Age distribution showed that most passengers were between 20–40 years old.

Younger passengers (children) had comparatively better survival rates than older adults.

4. Fare and Socioeconomic Factors:

Fare displayed a positive correlation with survival. Passengers who paid higher fares—typically in 1st Class—had better rescue access and opportunities.

5. Correlation Insights:

The heatmap confirmed strong relationships between Pclass, Fare, and Survival.

There were no strong correlations among most other numeric variables, but categorical factors (Sex, Class) played a major role.

Visual Findings

Charts such as histograms, boxplots, barplots, and heatmaps revealed patterns like skewed fare distribution, distinct survival differences by sex and class, and variation in age and fare across passenger groups.

Conclusion

The EDA demonstrates that **Sex**, **Passenger Class**, and **Fare** are the most influential factors affecting survival on the Titanic. The data clearly indicates that females and passengers from higher classes had significantly better chances of survival. These insights can be further used for predictive modeling and deeper statistical analysis.
