Exploratory Data Analysis (EDA) Report: Titanic Dataset

1■■ Pairplot: Relationships among Age, Fare, Family Size, and Survival

Survivors generally paid higher fares. Age distribution is similar for both groups, but slightly more survivors are seen in younger age ranges. Family size shows that people with very large families had lower survival chances.

2■■ Feature Correlation Heatmap

Fare is positively correlated with survival. Pclass shows negative correlation. Family size has weak negative correlation, indicating larger families slightly reduced survival chances.

3■■ Survival Count by Sex

Female passengers had a much higher survival count compared to males, indicating gender was a strong factor in survival due to the 'women and children first' policy.

4■■ Survival Rate by Sex

Average survival rate for females is around 70–75%, whereas for males it is below 20%.

5■■ Survival Rate by Sex and Passenger Class

Females in 1st and 2nd class had the highest survival rates. Males in 3rd class had extremely low survival chances.

6■■ Distribution of Numerical Features

Age: Most passengers were 20–40 years old. Fare: Right-skewed distribution. Family size: Mostly small families (0–3 members).

7■■ Boxplot of Age by Survival

Median age is slightly lower for survivors. Non-survivors include older passengers more frequently.

8■■ Boxplot of Fare by Survival

Survivors had significantly higher median fares. Indicates higher socioeconomic status led to higher survival chances.

9■■ Scatterplot of Age vs Fare by Survival

Survivors cluster in high-fare regions. Non-survivors paid lower fares across all age groups.

■ *Impact of Passenger Class on Survival*

1st class passengers had the highest survival rate (~60–65%), while 3rd class had the lowest (~25%).

Summary of Findings

1. Gender was the most influential factor — females survived at much higher rates.
2. Fare and class were strong socioeconomic indicators of survival.
3. Age had a mild influence — younger passengers slightly more likely to survive.
4. Family size affected survival negatively for very large families.
5. No strong multicollinearity between numerical features.
6. Survival probability was highest for females, 1st-class, and higher-fare passengers.

Conclusion

Socioeconomic status and gender were the dominant determinants of survival on the Titanic. Females and 1st-class passengers had the highest chances of survival. The findings align with historical accounts, confirming the validity of the dataset and analysis.

Prepared by: Nived K M