

OBSERVATION of VISUALS and REPORT FINDINGS:

1. Histogram of Numeric Features

Features: Age, Fare, SibSp, Parch

Observations:

Age: Right skewed, most passengers are between 20 and 40 years old.

Fare: Highly skewed with long right tail. Most fares are below ₹100, but some passengers paid very high amounts.

SibSp & Parch: Most passengers were traveling with few or no siblings/spouses or parents/children.

2. Boxplots

a. Survived vs Age

Observation:

Median age is slightly lower for survivors.

More outliers (very old passengers) among non-survivors.

Children had a higher chance of survival.

b. Survived vs Fare

Observation:

Survivors generally paid higher fares.

The spread of fares is wider among survivors, indicating high-class (expensive) ticket holders had better survival odds.

There are extreme fare outliers among survivors.

3. Pairplot of Age, Fare, Pclass, and Survived

Observations:

Survivors cluster in higher Fare and lower Pclass regions (i.e., higher-class cabins).

Passengers in 1st class paid higher fares and had better survival outcomes.

Age and Fare show a loose positive relationship.

4. Heatmap of Correlation (Numerical Features Only)

Observations:

Survived is:

Negatively correlated with Pclass ($r \approx -0.34$) – lower class = less chance of survival.

Positively correlated with Fare ($r \approx 0.26$) – higher fare = better chance of survival.

Fare and Pclass are also negatively correlated – higher class, higher fare.

5. Countplot of Sex vs Survived

Observation:

Survival rate is significantly higher among females.

Very few males survived compared to females.

Confirms the "Women and Children First" policy during evacuation.

6. Countplot of Embarked vs Survived

Observation:

Passengers from Cherbourg (C) had the highest survival rate.

Those from Southampton (S) had the lowest.

Embarkation location may relate to cabin class or deck location.

7. Barplot of Pclass vs Survival Rate

Observation:

1st Class: Highest survival rate (~63%)

2nd Class: Moderate survival rate (~47%)

3rd Class: Lowest survival rate (~24%)

Indicates clear privilege by class during evacuation.

8. Barplot of Average Fare by Pclass

Observation:

1st Class: Average fare is significantly higher than other classes.

3rd Class: Most economical.

Fare is a strong proxy for socio-economic status, which affected survival.

Summary of Findings:

Age Distribution: Most passengers were between 20–40 years old.

Fare: Highly right-skewed, with a few passengers paying very high fares.

Sex vs Survival: Females had a much higher survival rate.

Class vs Survival: Passengers in 1st class had higher chances of survival.

Correlation: Fare and Pclass show a moderate inverse correlation.

Prepared by: Joseph R Pakhuongte. (MSc Data Science)