SUMMARY FINDINGS

Exploratory Data Analysis (EDA) Observations

1. Survival Distribution

- ✓ A majority of passengers did not survive the Titanic disaster.
- ✓ Around 550+ passengers died (Survived = 0) and 340+ survived (Survived = 1).
- ✓ This translates to a survival rate of ~38% and a mortality rate of ~62%.

2. Gender Distribution

- ✓ The dataset contains ~580 male and ~310 female passengers.
- ✓ Males comprised ~65% of the passengers, while females made up ~35%.
- ✓ This gender imbalance is crucial in analyzing survival trends.

3. Passenger Class Distribution (Pclass)

- ✓ Most passengers were from 3rd class (~490), indicating a large number of lower socio-economic travelers.
- ✓ 1st class had ~220 passengers, and 2nd class had ~180.

- ✓ This imbalance suggests Pclass may significantly affect survival outcomes.

4. Age Distribution

- ✓ The distribution is right-skewed, with most passengers between 20 to 35 years.

- ✓ The peak age is around 28.

- ✓ Fewer children and elderly were aboard, and young adults were the dominant age group.

5. Fare Distribution

- ✓ Highly right-skewed — most fares are under $100.

- ✓ A significant number of passengers paid between $0–$50.

- ✓ A few high fares (above $200–$500) likely correspond to 1st class luxury travelers.

6. Survival by Gender

- ✓ Females had a significantly higher survival rate than males.

- ✓ Most males did not survive, whereas most females did survive.

- ✓ This supports the "women and children first" evacuation practice.

- ✓ Gender is a strong predictor of survival and valuable in predictive modeling.

7. Survival by Passenger Class

- ✓ 1st class passengers had a higher survival rate, indicating better access to lifeboats or priority.

- ✓ 2nd class passengers had nearly equal survival and non-survival counts.

- ✓ 3rd class passengers faced the highest mortality, suggesting significant survival disparity by class.

- ✓ Pclass strongly influences survival outcomes.

8. Age vs. Survival

- ✓ The median age of survivors is slightly lower than non-survivors.

- ✓ Children (age <10) had a higher survival rate.

- ✓ Older adults (age >60) were more often among the non-survivors.

- ✓ Age may play a role, especially when considered alongside class and gender.