Rationale for Data Cleaning Steps

1. **Handling Missing Values**:

   Age: Filled with the median to retain age information for age-based analysis without affecting due to null values.

   Embarked: Filled with the mode, the most common value, to maintain port-related data.

   Cabin: Dropped due to a high volume of missing values, which would add noise to the analysis.

2. **Removing Duplicates**:

   Ensures each passenger is unique, avoiding duplicate counts that could affect results.

3. **Standardizing Date Columns**:

   There were no date-related columns so it was not applicable.

4. **Creating Age Group Column**:

   Groups ages into "Child", "Adult", and "Senior" for easier comparison in survival analysis.

Visualizations and Statistical Findings

Gender Distribution Bar Plot:

The bar plot shows that there were more males on board than females. This distribution provides insight into the gender composition of passengers on the Titanic.

Age Distribution Histogram:

The histogram of age distribution reveals that most passengers were between 20-30 years old, with a peak of around 25 years. This indicates that the Titanic

had many adults, which may have influenced survival trends in specific age groups.

Survival Rate by Gender and Class:

The survival rate plot indicates that females had a higher survival rate compared to males, particularly in First class.

Statistical Findings

1. **Fare Analysis**:

   The mean fare is 32.20, which is higher than the median fare of 14.45, indicating a right-skewed distribution. This suggests that while most passengers paid around 14.45, some paid significantly more, raising the average fare.

2. **Age Analysis**:

   The mean age is 29.36, close to both the median and mode age of 28. This alignment suggests that the age distribution is relatively symmetrical around the average, with most passengers being around 28 years old.

3. **T-Test for Survival Rate by Gender**:

   The t-test statistic of -19.30, combined with a p-value of 1.41e-69, indicates a highly significant difference in survival rates between genders. With a p-value much smaller than the conventional significance level of 0.05, we can conclude that gender had a statistically significant impact on survival. Specifically, this suggests that one gender (likely females) had a higher survival rate, which could reflect the "women and children first" policy during the evacuation.