Summary of Findings:
Titanic Data Exploration

📊Data Overview

- The dataset contains information on 891 passengers from the Titanic, with 11 features after preprocessing.
- Key variables include survival status, passenger class, age, gender, fare, and embarkation point.

🔧Preprocessing Steps

1. **Handled Missing Values:**
    a. Age: 177 missing values filled with the mean age (30 years)
    b. Cabin: Dropped the column due to 687 missing values (77% missing)
    c. Embarked: 2 missing values filled with the mode ('S')
2. **Data Type Adjustments:**
    a. Converted Age from float to integer after imputation

📈Key Statistics

- Survival Rate: 38.4% survived (342 passengers), 61.6% did not (549 passengers)
- Gender Distribution:
    - Male: 577 passengers (65%)
    - Female: 314 passengers (35%)
- Passenger Classes:
    - 1st: 24%
    - 2nd: 21%
    - 3rd: 55%
- Age Range: 0 to 80 years (mean = 30, median = 28)
- Embarkation Points:
    - Southampton (S): 72%
    - Cherbourg (C): 19%
    - Queenstown (Q): 9%

🔭Notable Observations

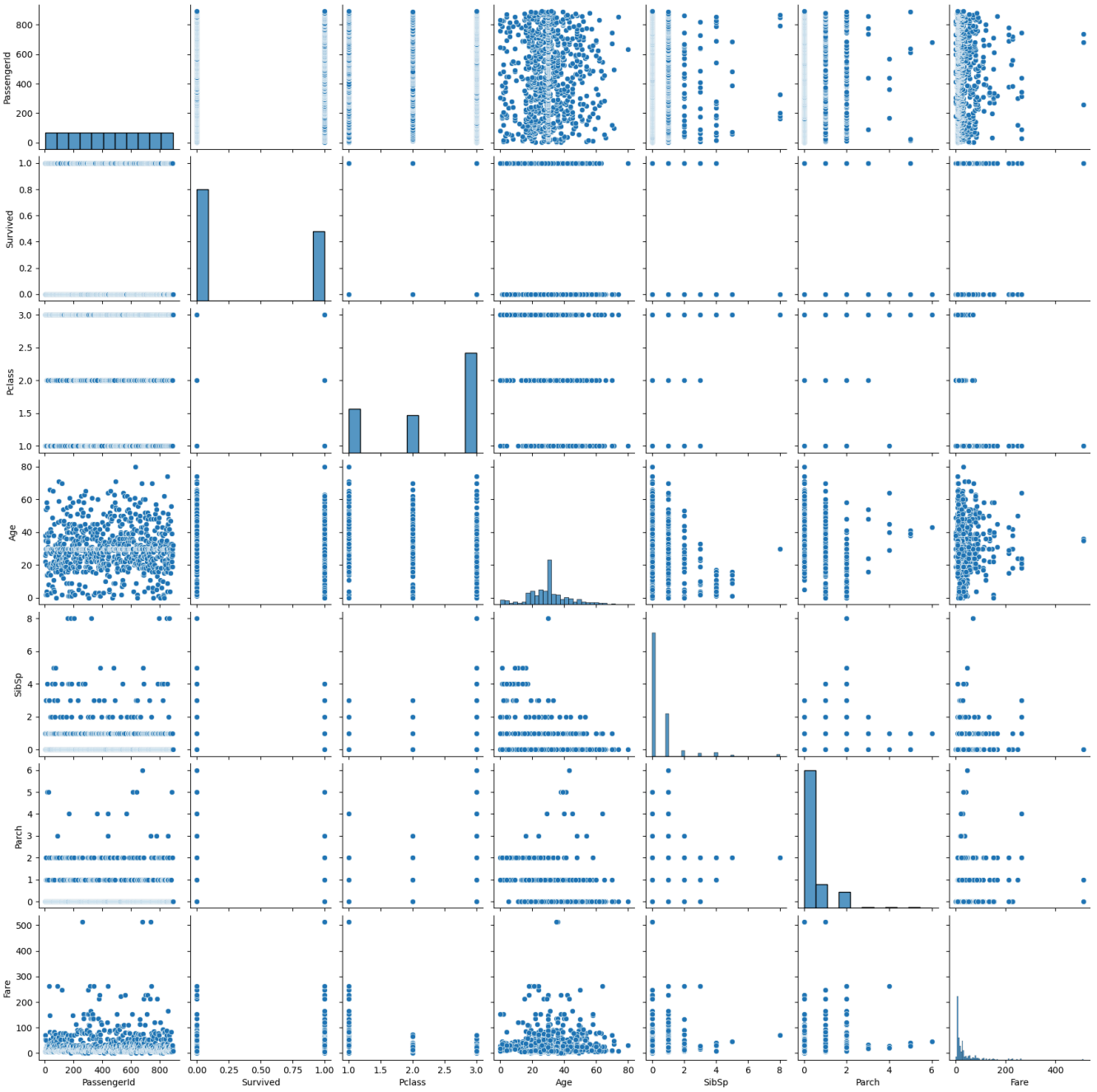
1. Age Distribution:
    a. Most common ages were 24 (30 passengers), 22 (27), and 18 (26)
    b. Many children (ages 0-10) were present in the dataset

2. Family Relationships:
    a. Most passengers traveled without siblings/spouses (SibSp=0) or parents/children (Parch=0)
    b. Some large families present (max SibSp=8, max Parch=6)
3. Fare Variation:
    a. Wide range from
    b. 0to
    c. 0 to 512.33
    d. Median fare was $14.45, indicating many lower-cost tickets

The dataset has been cleaned and prepared for further analysis or modeling, with all missing values addressed and appropriate data types assigned. The initial exploration reveals interesting demographic patterns that could be explored further to understand survival factors.

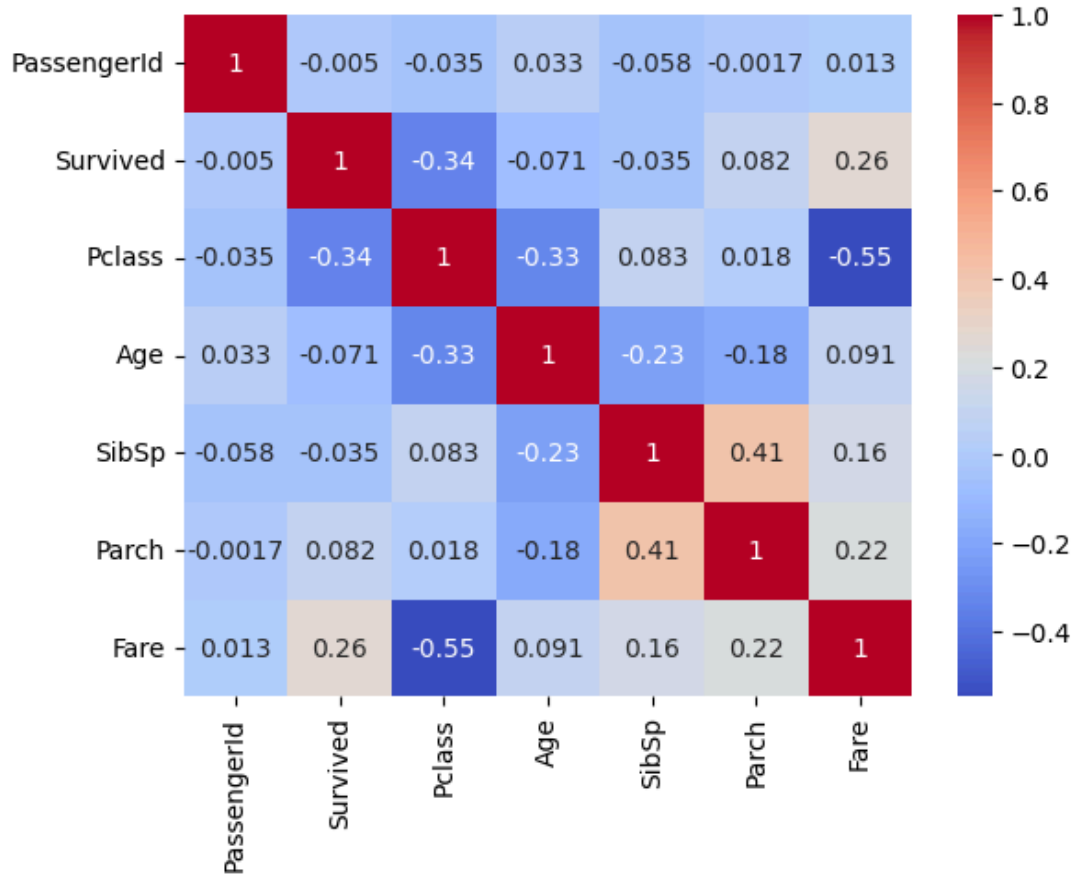
Visual Analysis

A. Pairplot of Numerical Features



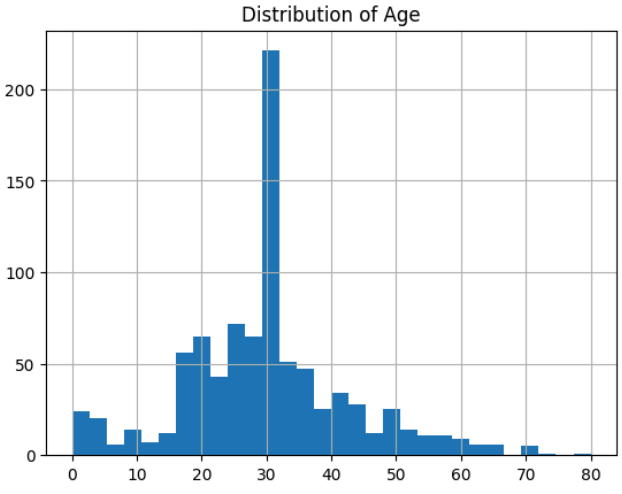
This plot shows pairwise relationships between numerical variables. Look for clusters and separation patterns by features like Age, Fare, etc.

B. Correlation Heatmap



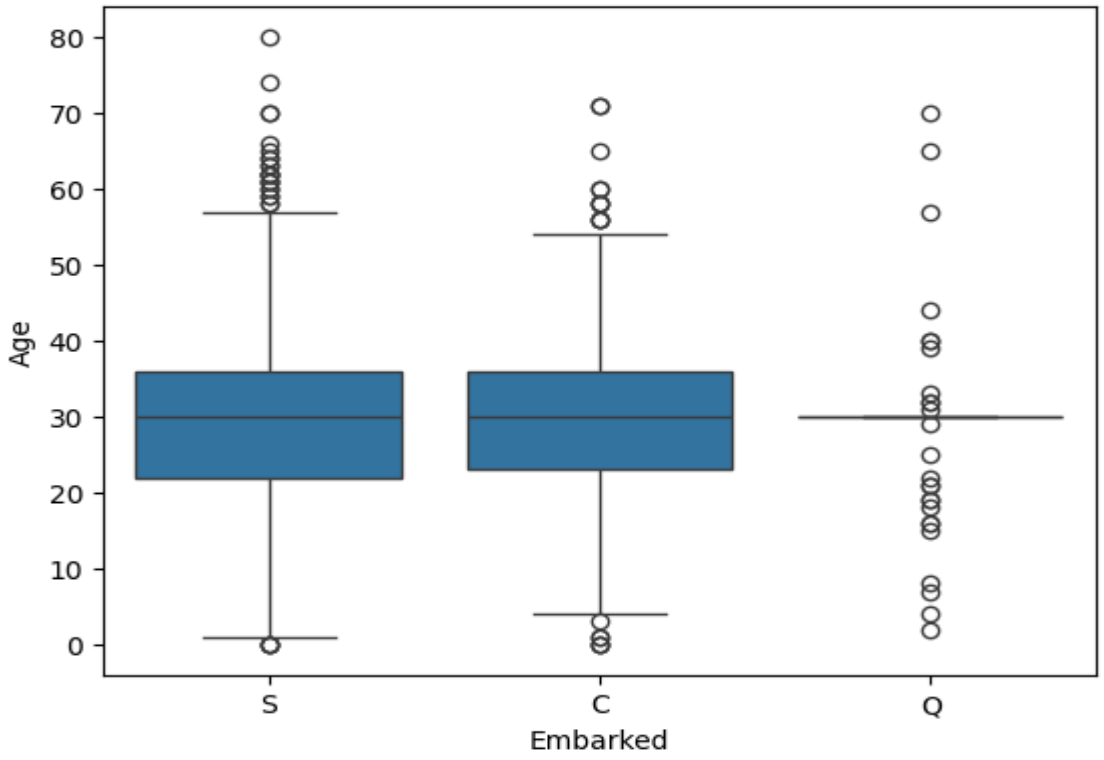
- Shows correlation coefficients among numeric features.
- High correlation between Fare and Pclass observed

C. Distribution of Age



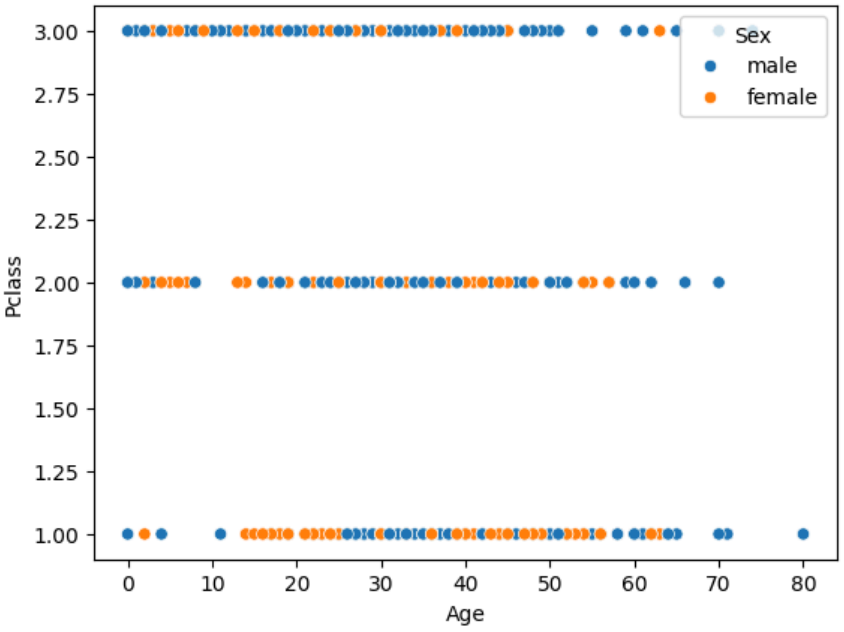
Age is right-skewed

D. Boxplot : Age vs Embarked Port



- Median age is similar across all ports (~30 years), but Queenstown (Q) has a narrower age range, indicating less variability.
- Outliers are present in all groups, especially from Southampton (S) and Cherbourg (C), suggesting a wider age distribution in those ports.

E. Scatterplot of Age vs Passenger Class by Sex



- Most younger passengers (under 15) appear across all classes, but are more frequent in 3rd class, especially among males.
- 1st class passengers are generally older, and both males and females are fairly represented across all age groups and classes.

Final Summary

1. The Titanic dataset shows that survival chances were higher for females and 1st class passengers, with strong links to fare and class.

2. Younger passengers were more concentrated in 3rd class, and age distributions varied slightly by embarkation port.

3. Key patterns and outliers were identified through visual tools like boxplots, scatterplots, pairplots, and heatmaps, enabling a clearer understanding of passenger demographics and relationships.