Titanic Dataset EDA Report

Total Rows: 891
Total Columns: 12

Columns:

PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked

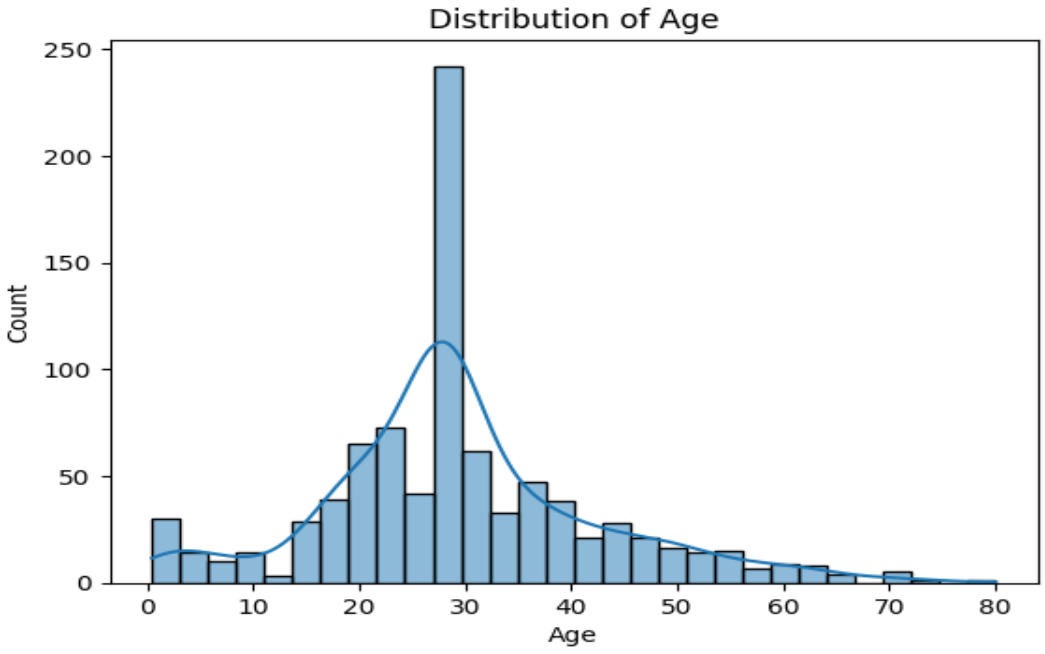
Missing Values:

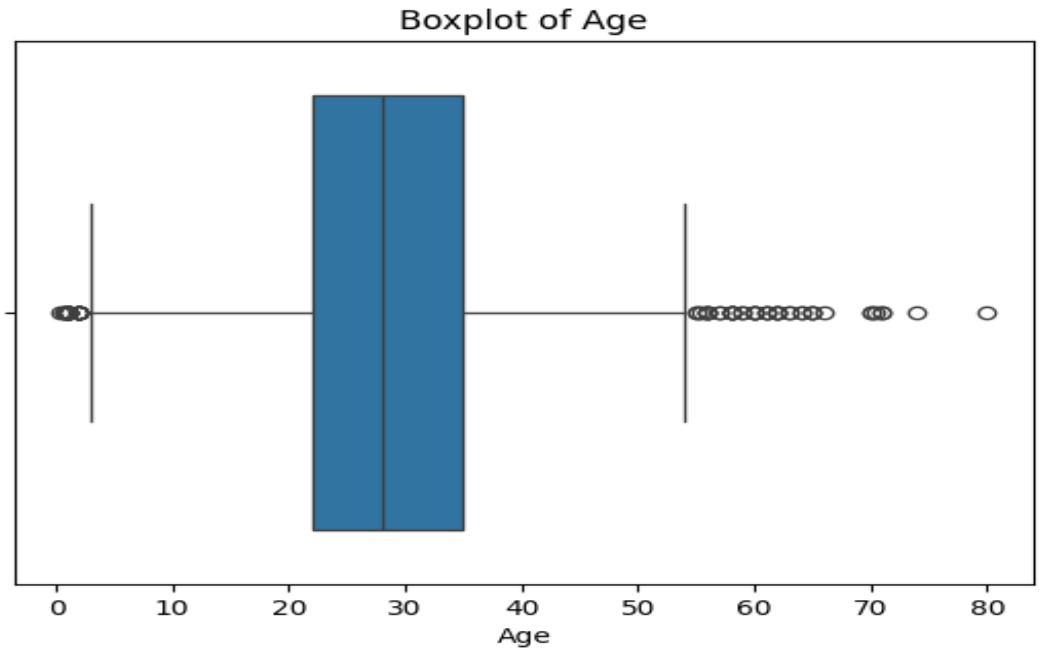
Missing Count PassengerId 0 Survived 0 Pclass 0 Name 0 Sex 0 Age 177 SibSp 0 Parch 0 Ticket 0 Fare 0 Cabin 687 Embarked 2

Summary Statistics:

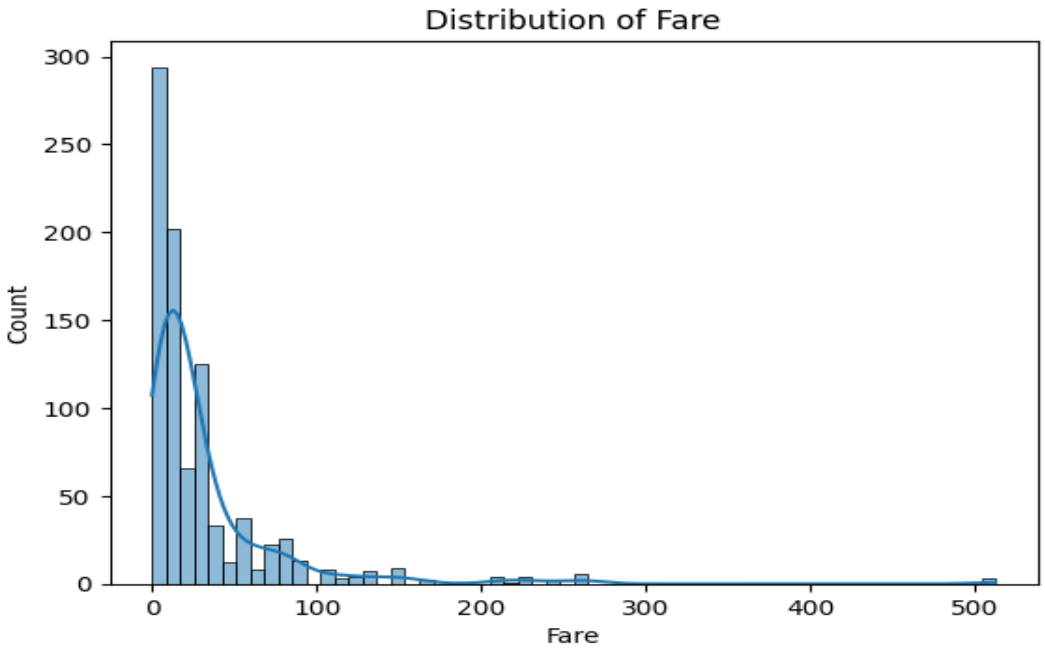
PassengerId Survived Pclass Age SibSp Parch Fare count 891.00 891.00 891.00 714.00 891.00 891.00 891.00 mean 446.00 0.38 2.31 29.70 0.52 0.38 32.20 std 257.35 0.49 0.84 14.53 1.10 0.81 49.69 min 1.00 0.00 1.00 0.42 0.00 0.00 0.00 25% 223.50 0.00 2.00 20.12 0.00 0.00 7.91 50% 446.00 0.00 3.00 28.00 0.00 0.00 14.45 75% 668.50 1.00 3.00 38.00 1.00 0.00 31.00 max 891.00 1.00 3.00 80.00 8.00 6.00 512.33

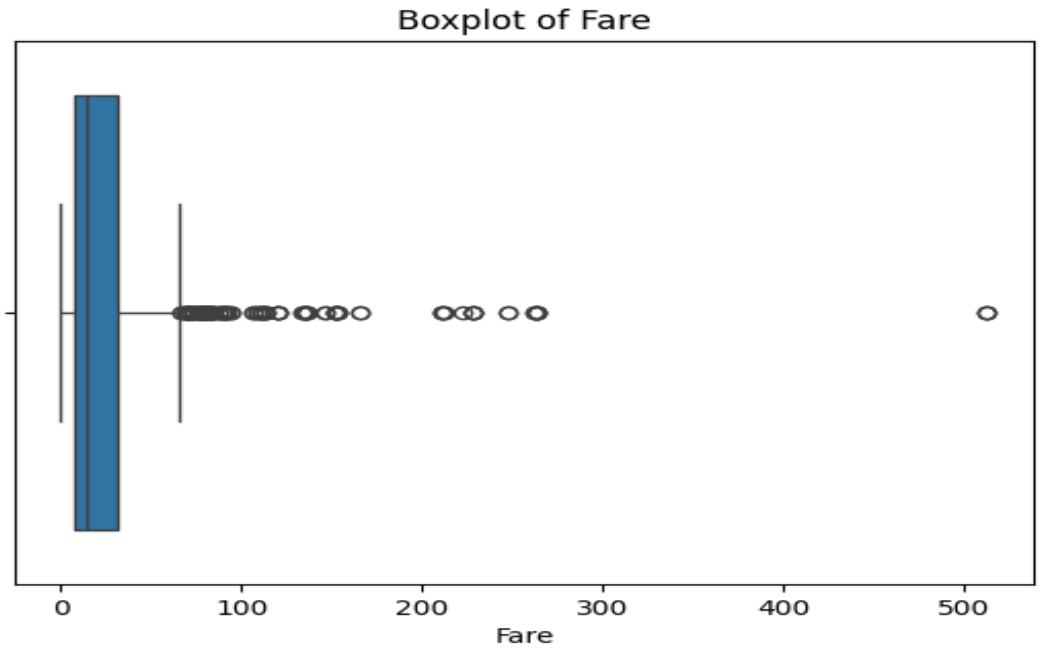
Distribution of Age



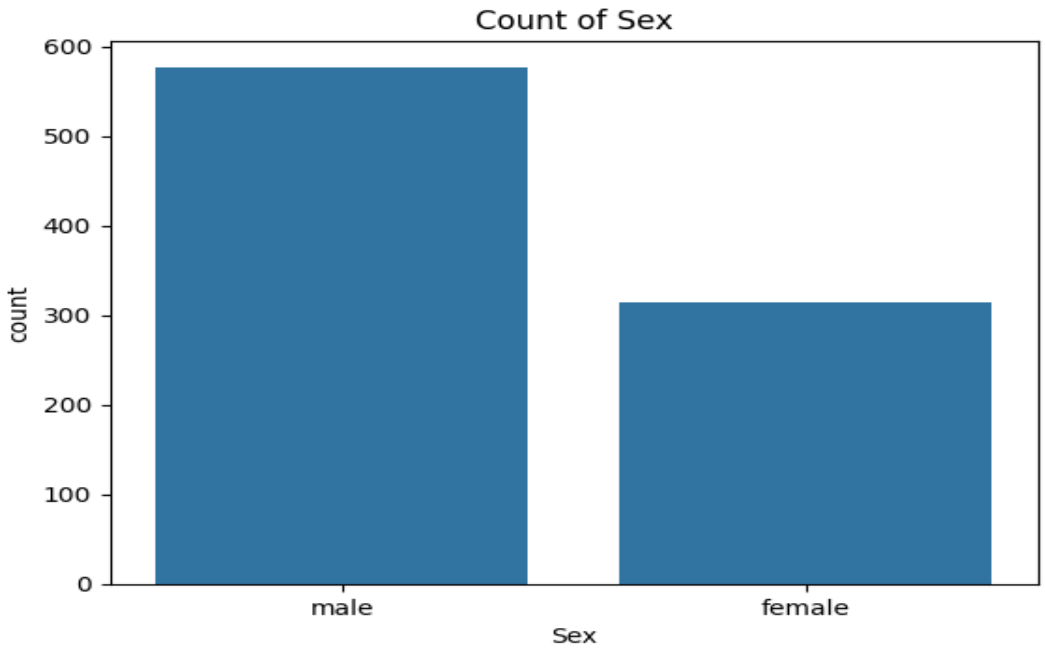


Distribution of Fare

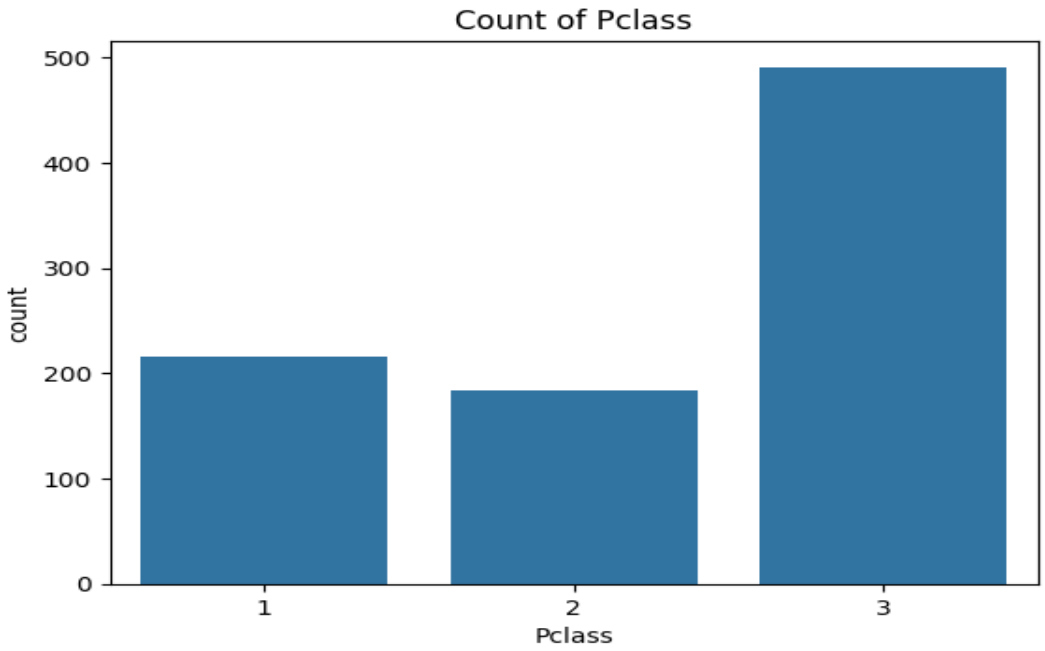




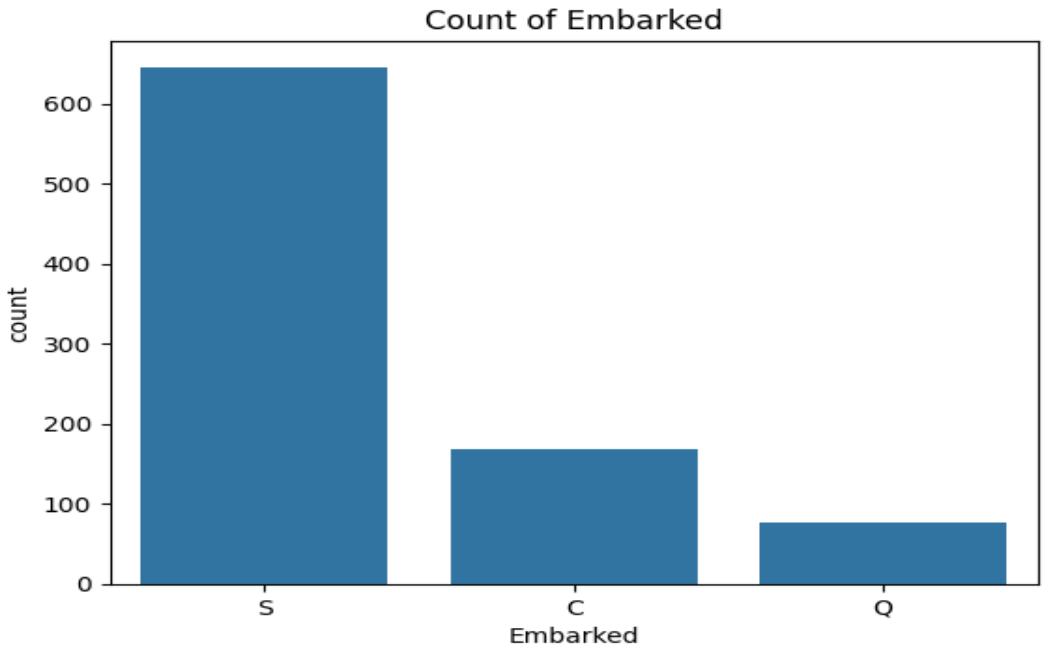
Count of Sex



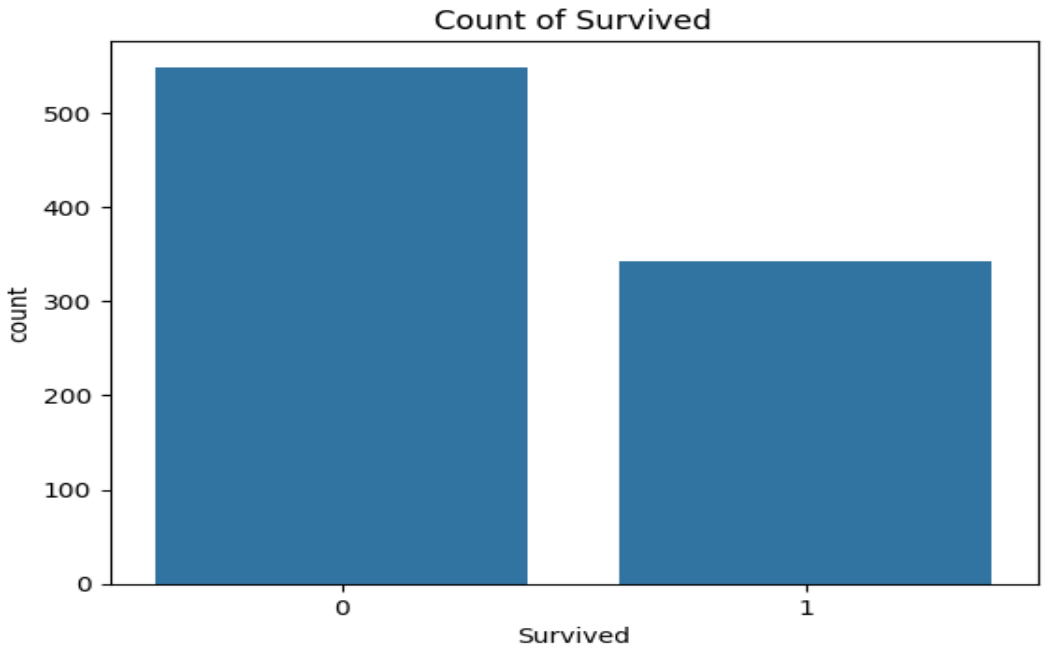
Count of Pclass



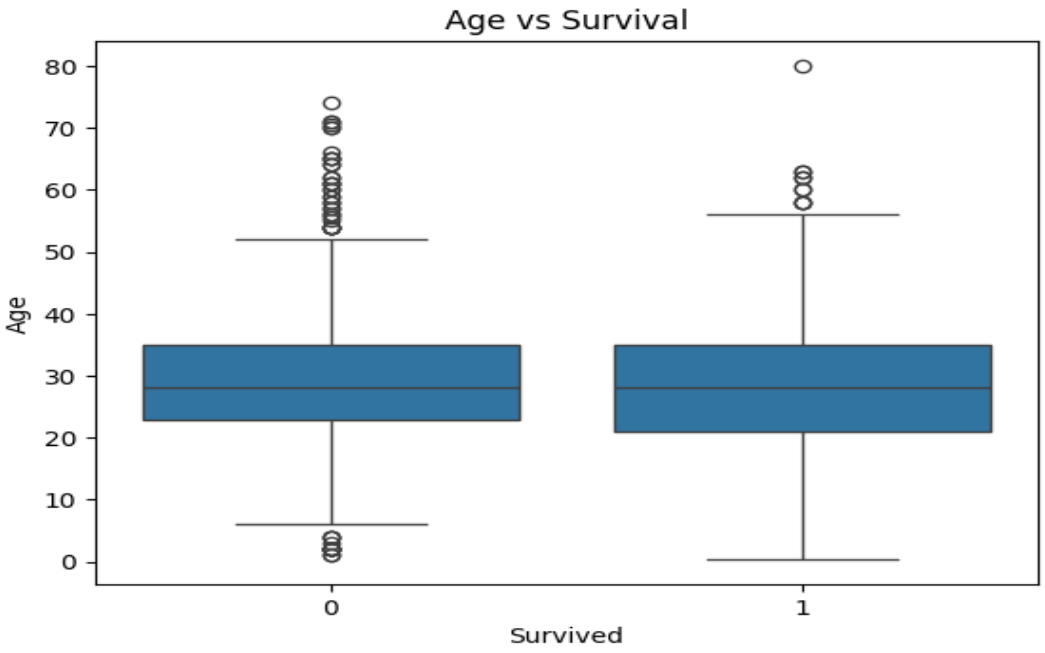
Count of Embarked

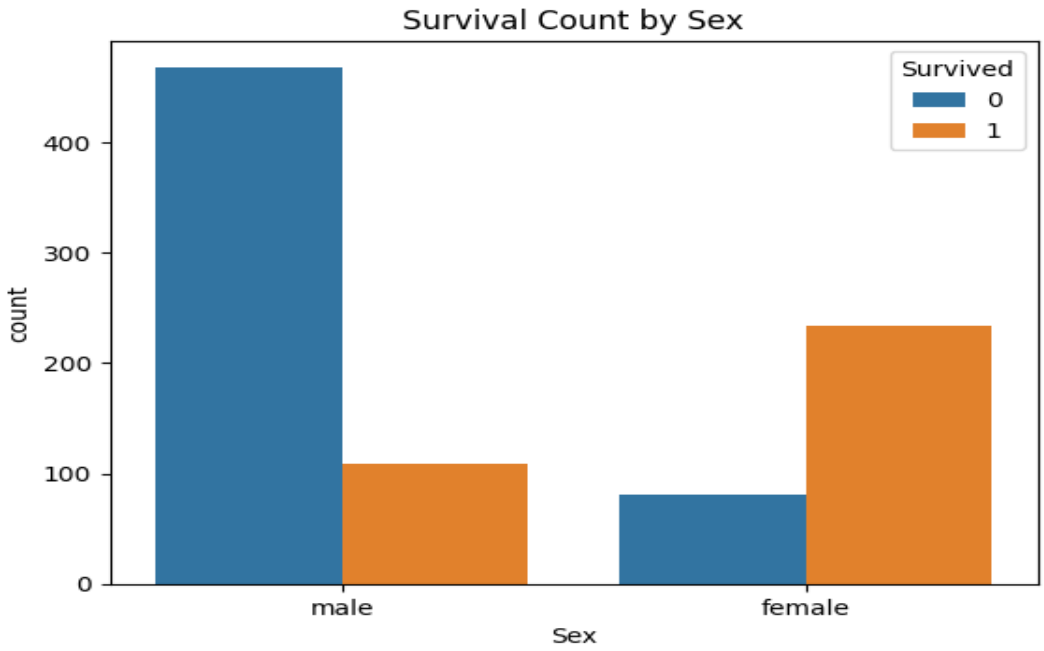


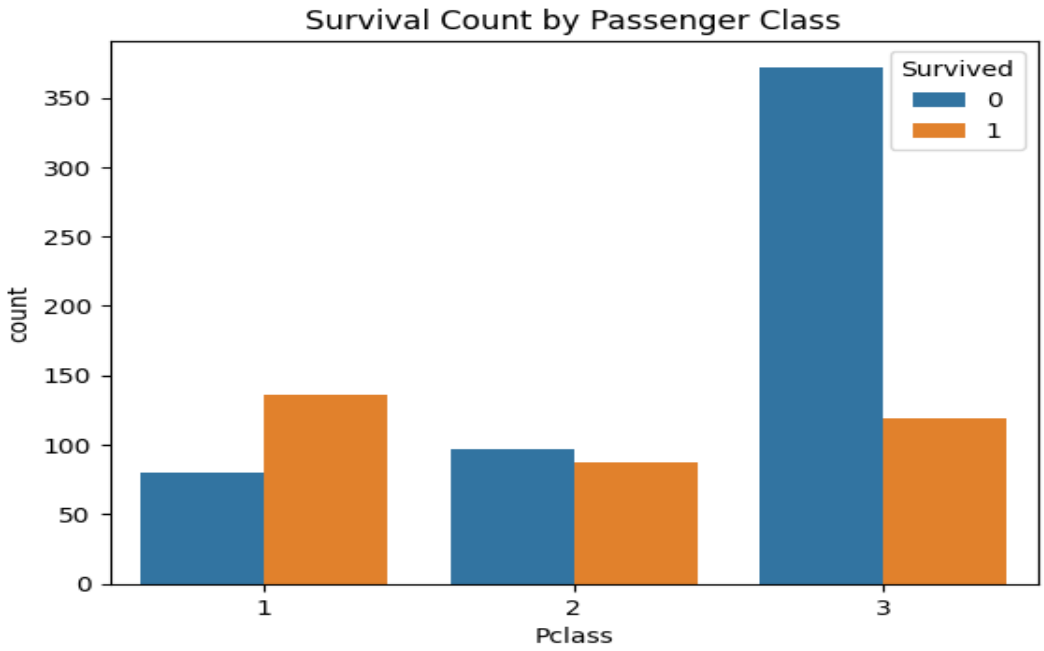
Count of Survived



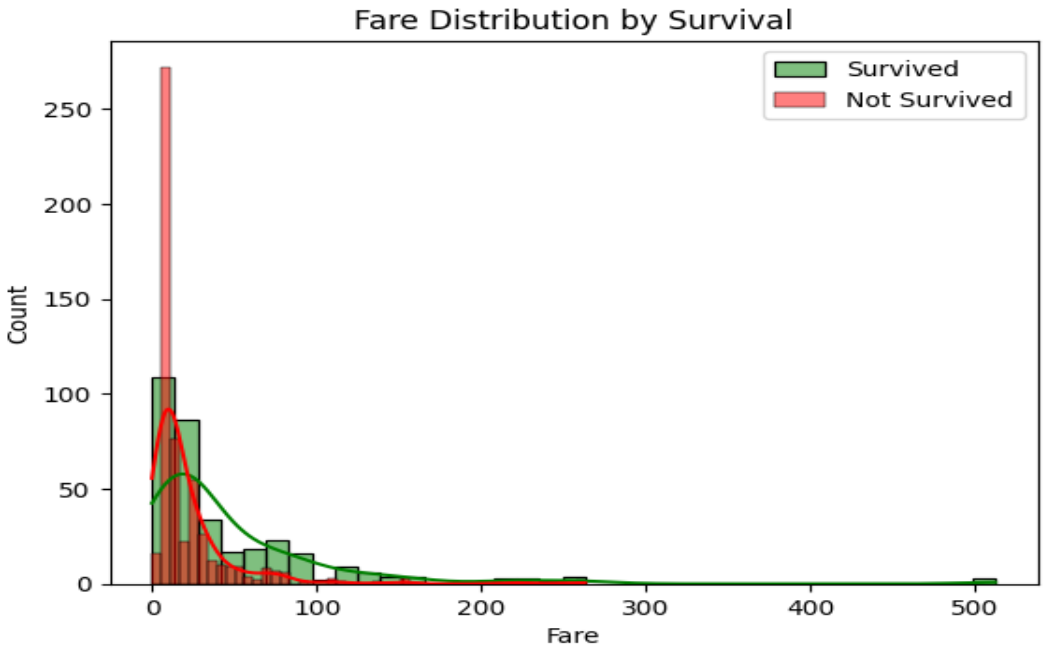
Age vs Survival



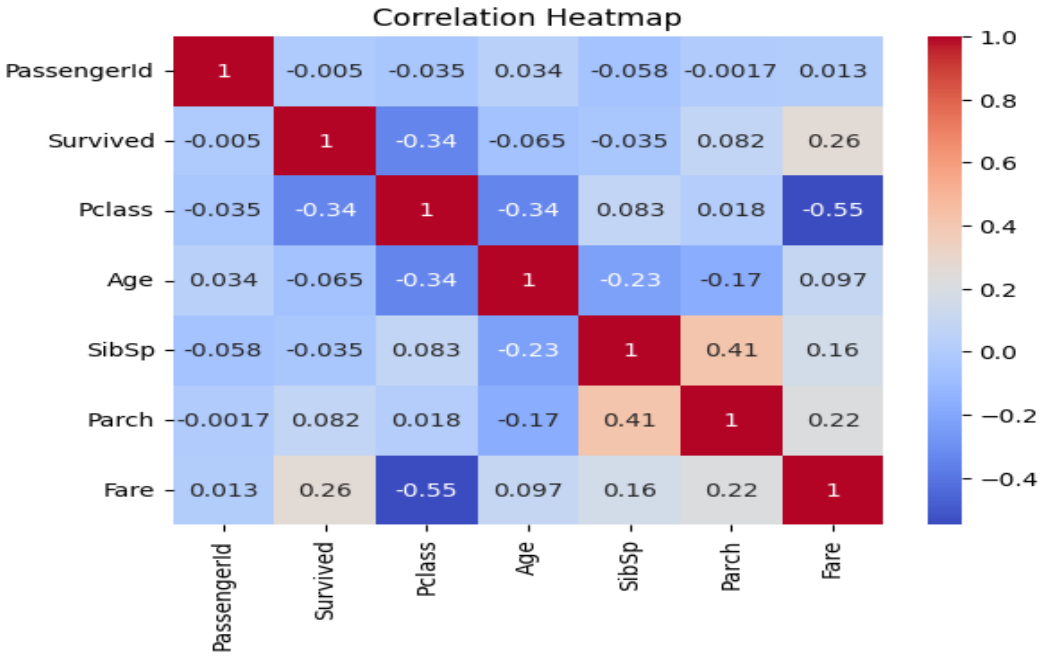




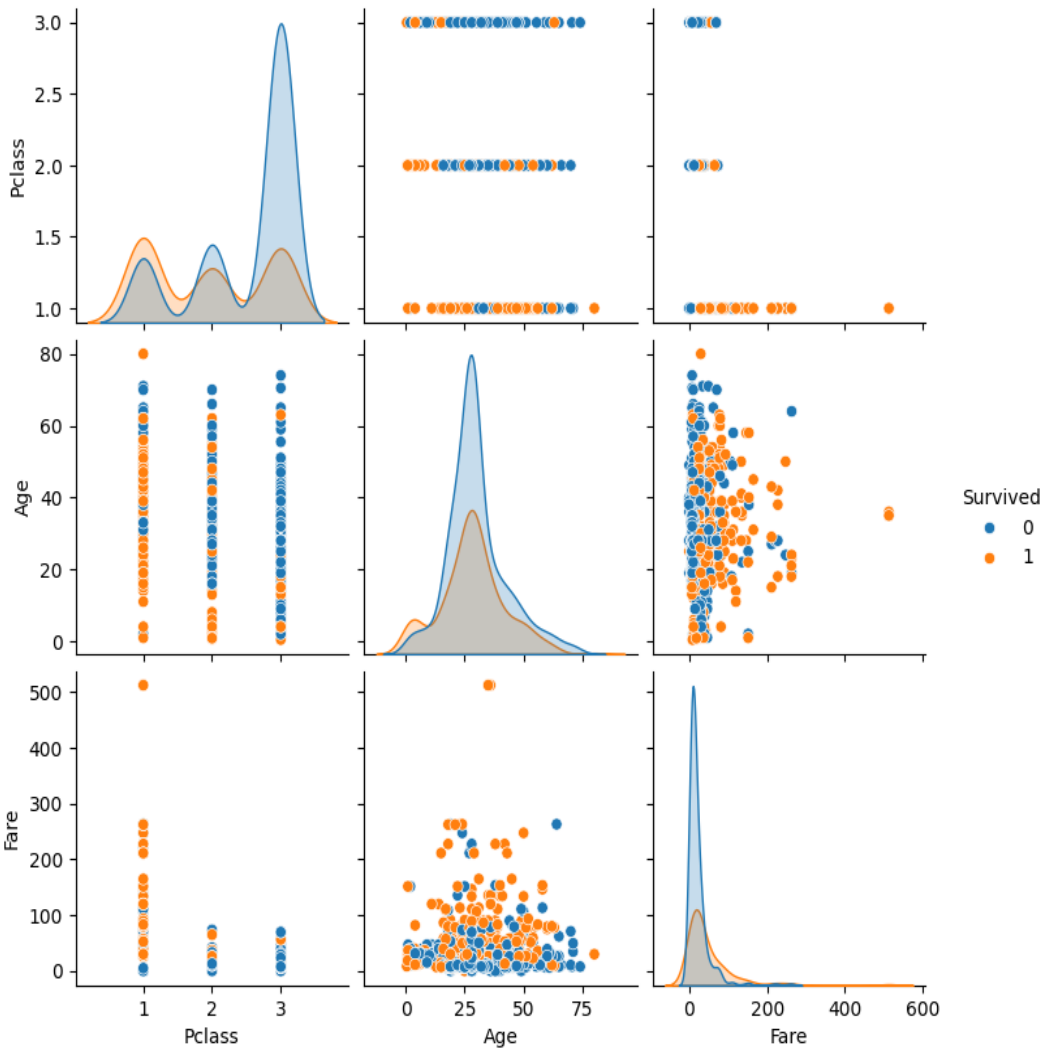
Fare vs Survival



Correlation Heatmap



Pairplot of Selected Features



Summary of Insights

1. Most passengers were in 3rd class.
2. Majority of passengers were male, but survival rate was higher for females.
3. Younger passengers had slightly better survival chances.
4. Higher fares were associated with higher survival rates.
5. Pclass is strongly correlated with survival — 1st class had better survival rates.
6. Age distribution is slightly right-skewed.
7. Fare distribution has a long tail with outliers (very high fares).