Task 5 PDF Report

Sofia Parveen

June 2025

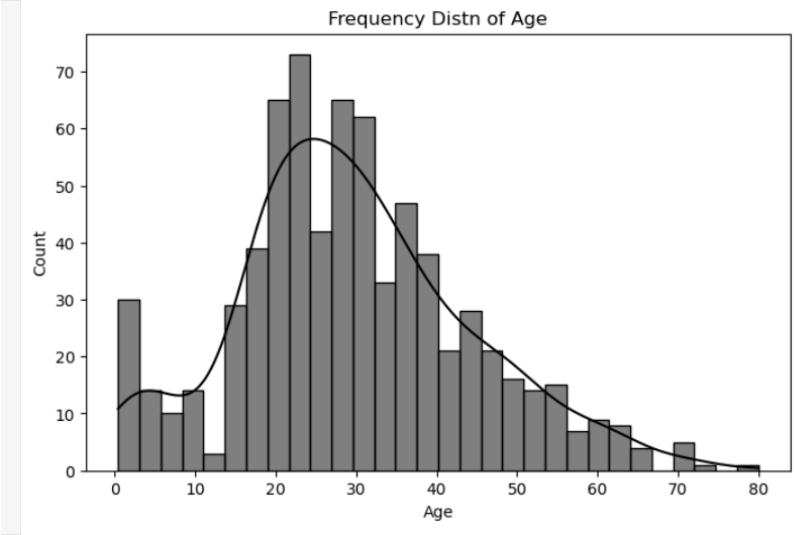

Figure 1: Age Distribution

This histogram shows that most passengers were between 20 and 40 years old. There are fewer elderly passengers, and many children were on board as well. The distribution is slightly right-skewed, and missing values exist for some passengers.

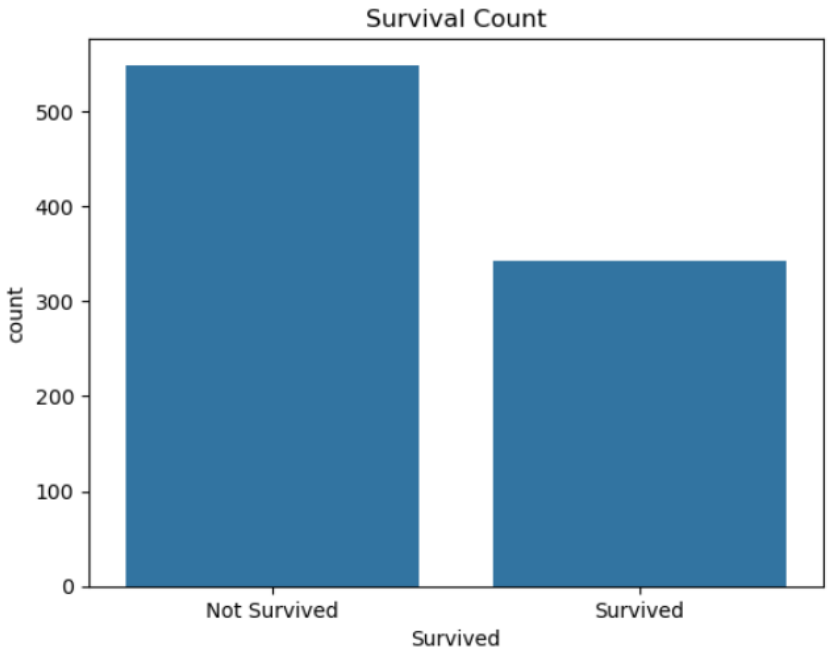

Figure 2: Survival Count

This bar plot shows that more passengers died (label 0) than survived (label 1). This imbalance gives insight into the severity of the disaster and sets a baseline for comparing survival across other features.

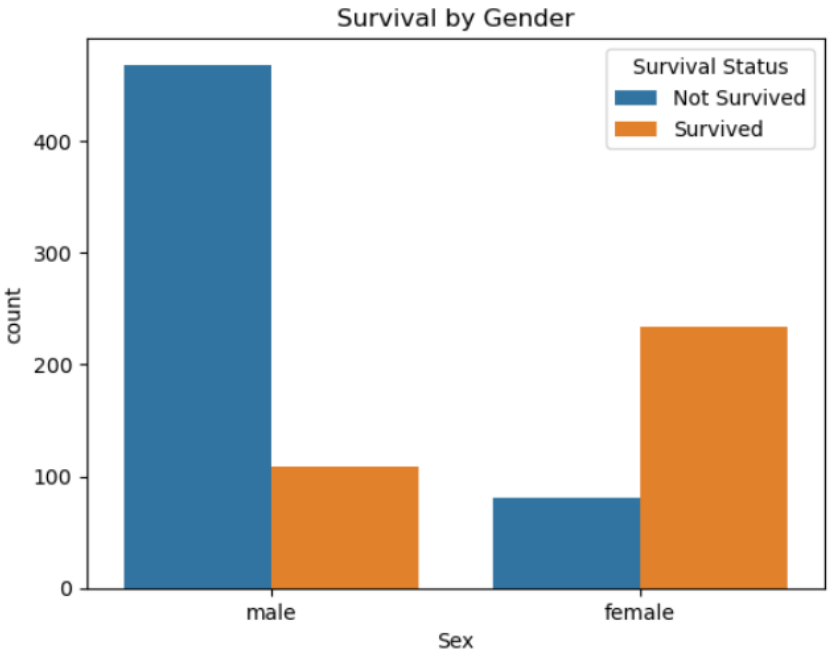

Figure 3: Survival by Gender

This plot clearly shows that females had a much higher survival rate compared to males. Most male passengers did not survive, while the majority of female passengers did. This reflects the "women and children first" evacuation policy followed during the disaster.

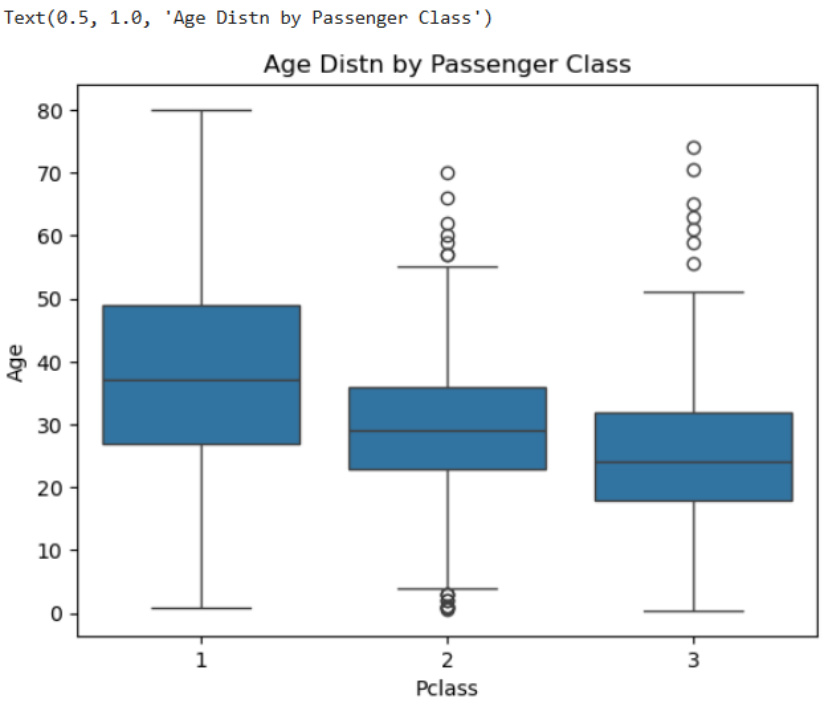

Figure 4: Age Distribution by Passenger Class

Passengers in 1st class tend to be older, while 3rd class passengers skew younger. This may be due to socioeconomic differences in who could afford higher-class tickets. The median age increases with class level.

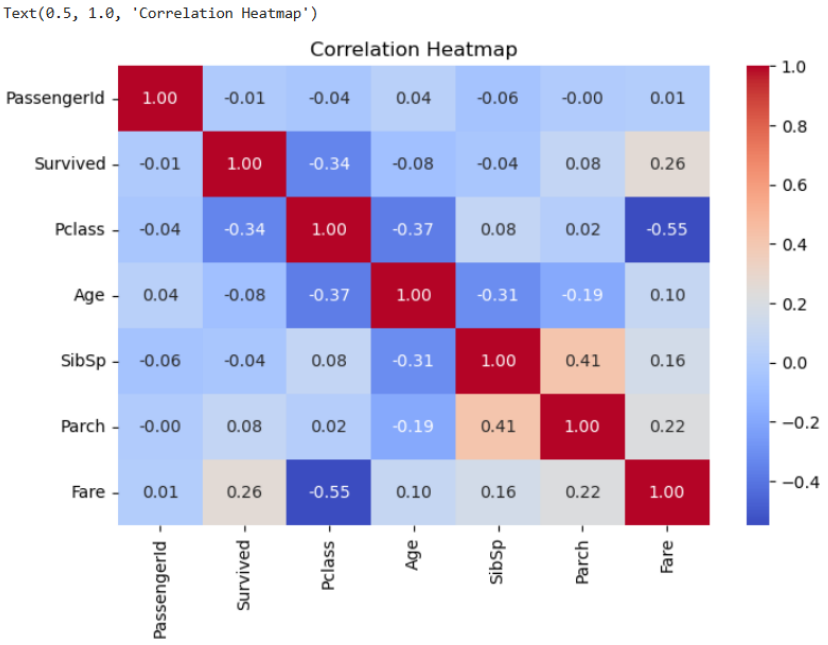

Figure 5: Correlation Matrix

Fare and Pclass show moderate correlation with Survived. Age and SibSp (siblings/spouses aboard) have weak correlations. Strong multicollinearity is not evident, so most features can be considered independent.

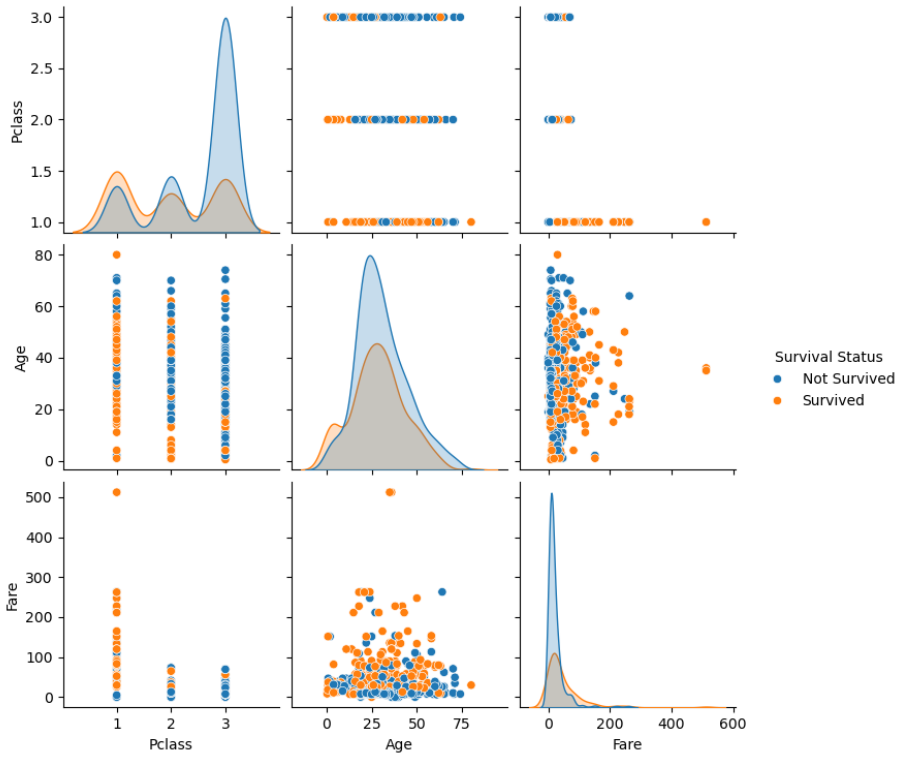

Figure 6: Pairplot

This pairplot visualizes the relationships among passenger class (Pclass), age, and fare, by survival status. The following patterns emerge:

- **Pclass vs Fare:** Survivors generally paid higher fares and were clustered in 1st class, indicating a strong correlation between ticket class and survival.
- **Age distribution:** Survivors span all age groups, but a slight density shift toward younger passengers is observed.
- **Fare distribution:** The fare distribution is highly right-skewed. Most passengers paid less than $100, but some outliers paid significantly more, and many of those survived.
- **Pclass vs Age:** Younger passengers are mostly in 3rd class, while 1st class passengers tend to be older.

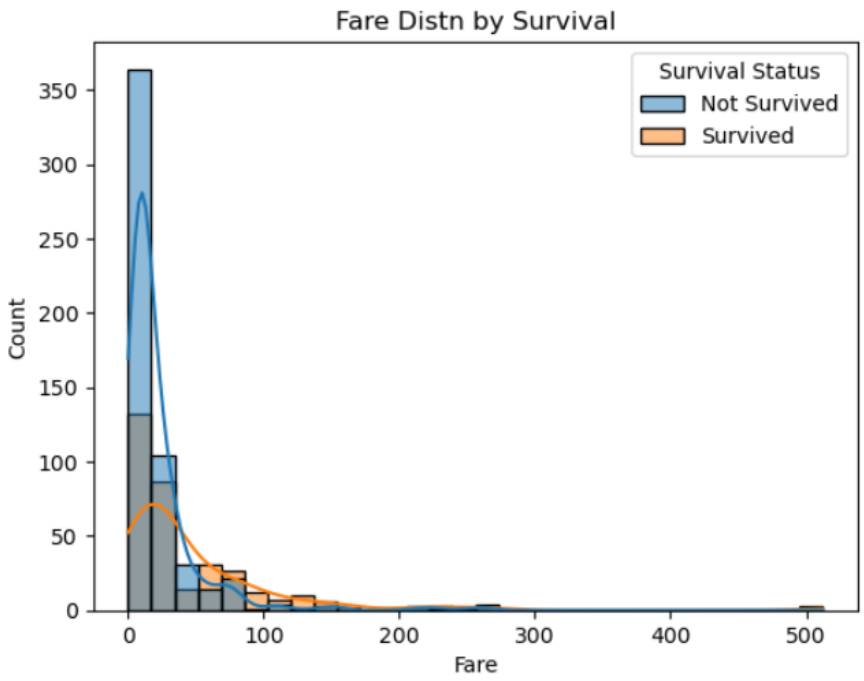

Figure 7: Fare Distribution by Survival

Survivors paid, on average, higher fares than non-survivors. Many who died paid lower fares, indicating they were likely traveling in lower classes. The distribution is right-skewed due to a few very high-ticket fares in 1st class.

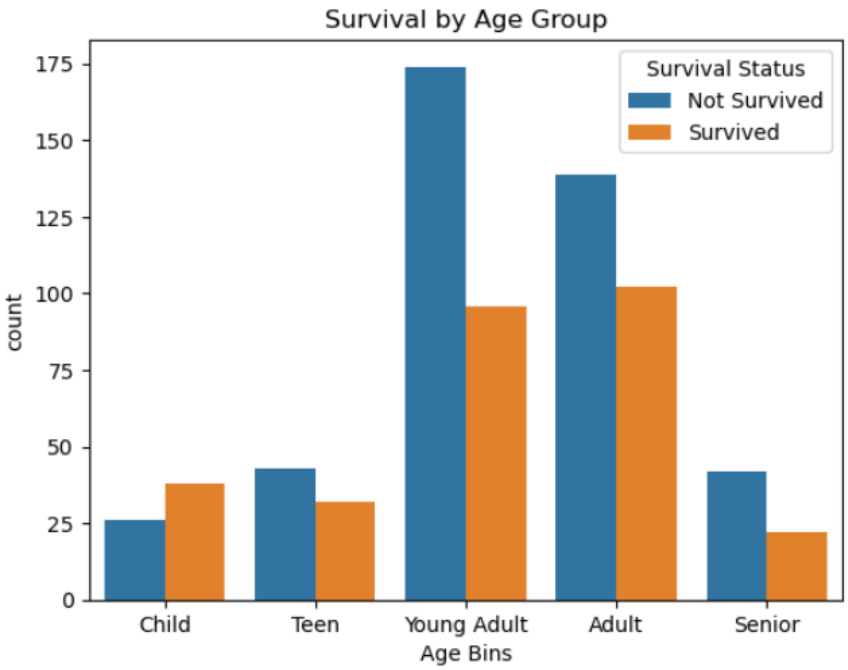

Figure 8: Survival by Age Group

Children and young adults had relatively higher survival rates, especially when compared to adults and seniors. This supports the observation that younger passengers were given priority during evacuation. However, the highest non-survival rates are among young adults and the lowest non-survival rates are among the seniors. This suggests that during evacuation, seniors and children were given priority over young adults and adults.

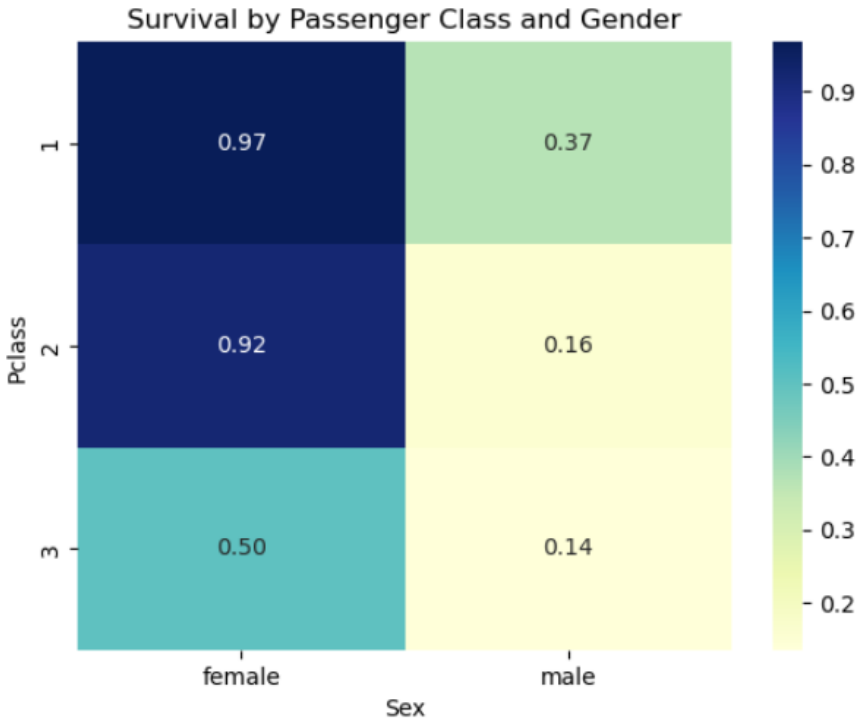

Figure 9: Survival by Passenger Class and Gender

1st class females have the highest survival rate, followed by 2nd class females. Males across all classes have relatively low survival. This clearly shows gender and class together were strong predictors of survival.

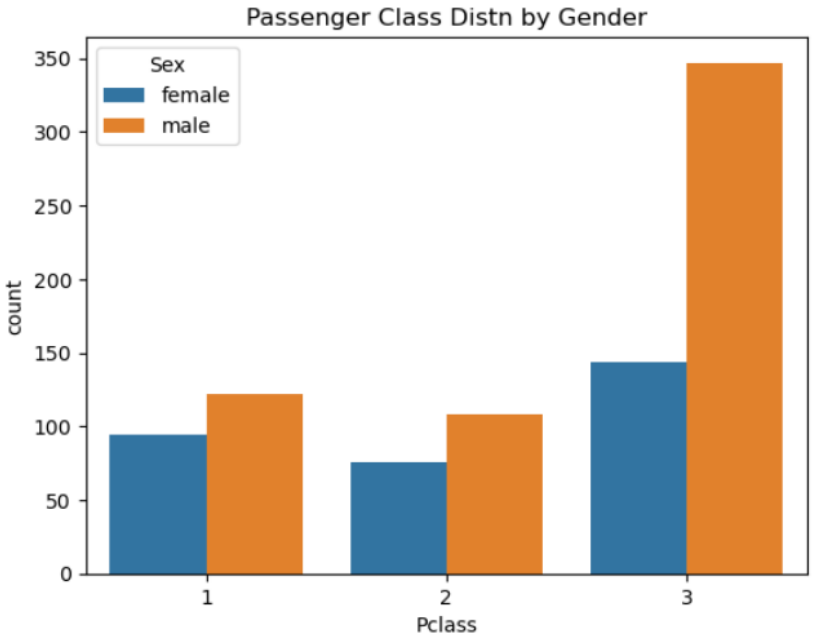

Figure 10: Passenger Class Distn by Gender

A higher number of females traveled in 1st and 2nd class, whereas most males were in 3rd class. This distribution likely contributed to gender-based survival differences observed earlier.

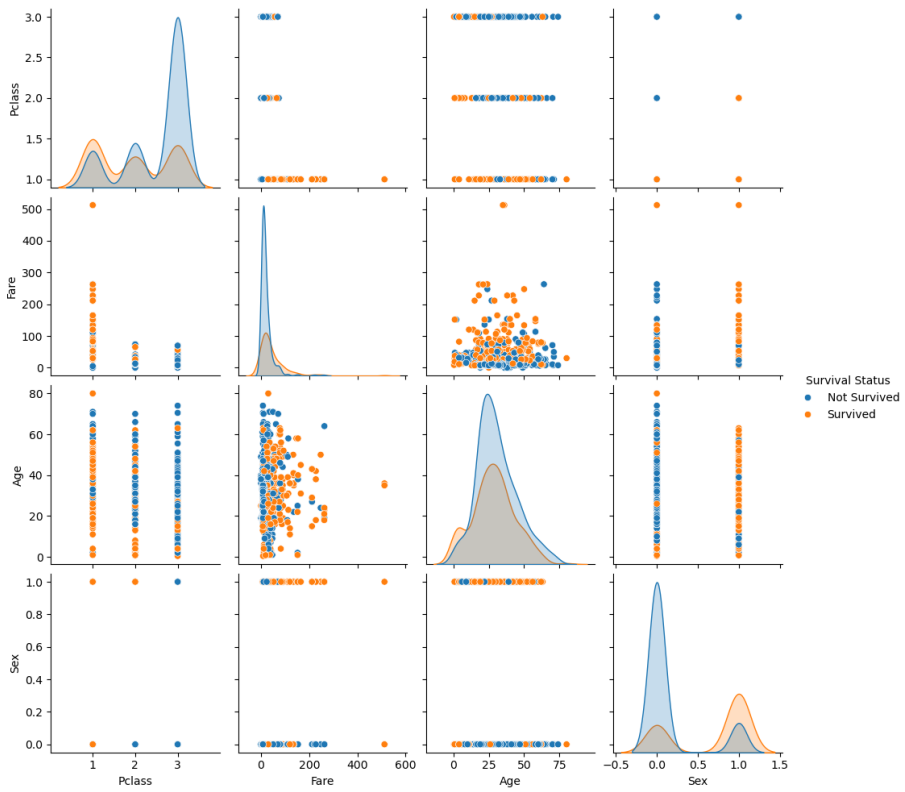

Figure 11: Pairplot

This pairplot presents the relationships among passenger class (Pclass), fare, age, and sex (numerically encoded), differentiated by survival status. The following insights can be listed:

- **Sex vs Survival:** Most survivors are from the group where Sex = 1 (females), while non-survivors are predominantly from Sex = 0 (males). This confirms that gender was a major factor in survival.
- **Fare vs Pclass:** As expected, higher ticket fares are strongly associated with 1st class passengers, and these passengers show a higher survival rate.
- **Age vs Fare:** Survivors are scattered across age groups but tend to have paid higher fares, especially within the younger age brackets.
- **Sex vs Pclass:** Many female survivors were in 1st and 2nd class, while males were more concentrated in 3rd class, where survival was rare.
- The diagonal KDE plots show that the ditribution of survivors are skewed towards females, 1st class and high fare paying passengers.

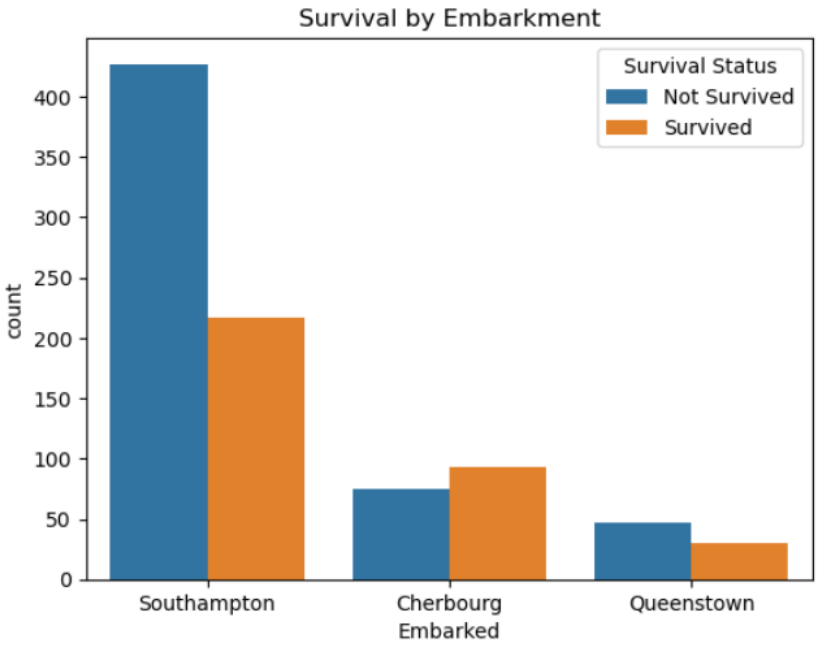

Figure 12: Survival by Embarkment

Passengers who embarked from Cherbourg (C) had a relatively higher chance of survival compared to Queenstown and Southampton. The highest number of non-survivors are from Southampton.