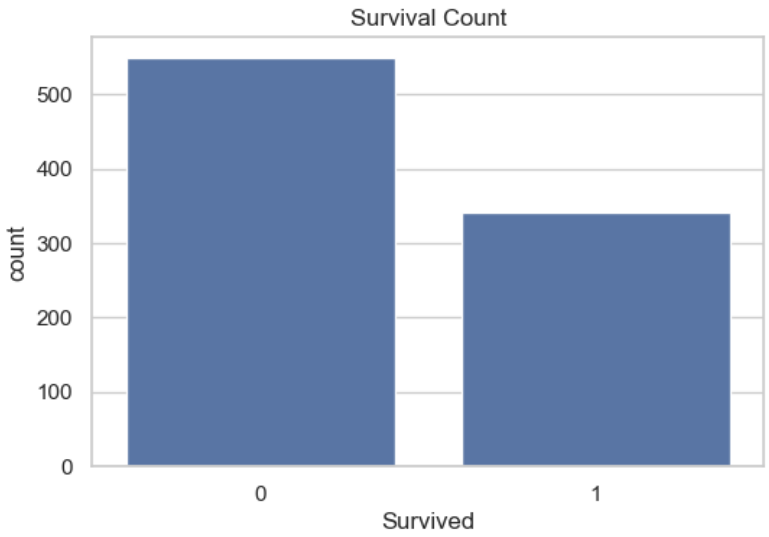
REPORT OF ANALYSIS

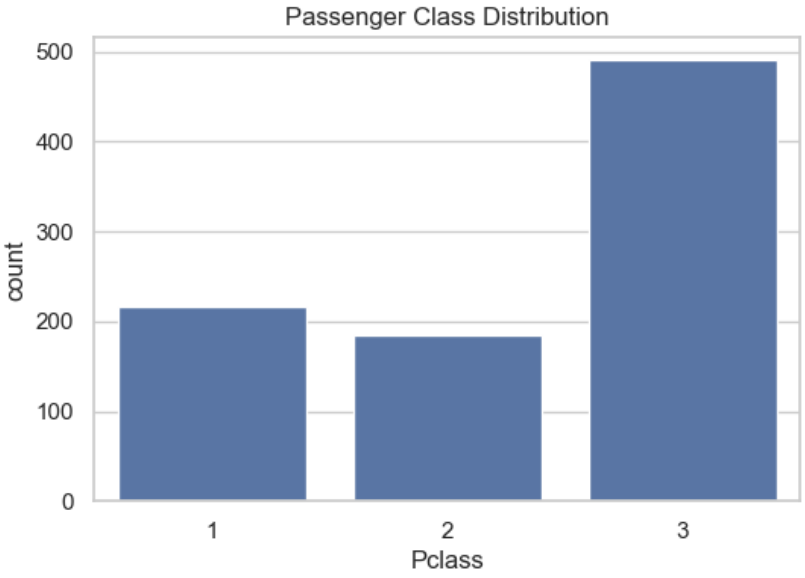
ANALYSIS

PLOT 1 - Survival Count



- More passengers died than survived: The number of non-survivors (Survived = 0) is significantly higher than the number of survivors (Survived = 1), indicating that survival was less common on the Titanic.
- Class imbalance: The survival outcome is imbalanced, which is important to consider in predictive modeling, as it may affect model performance and require techniques like resampling or class weighting.

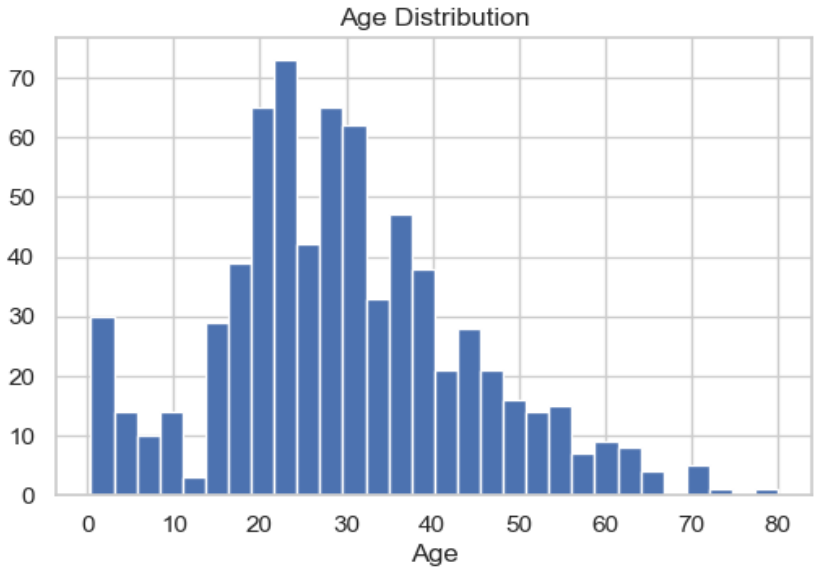
PLOT 2 - Passenger Class Distribution



- Most passengers were in 3rd class: The largest group of passengers traveled in 3rd class, indicating that lower-income individuals made up the majority on the Titanic.

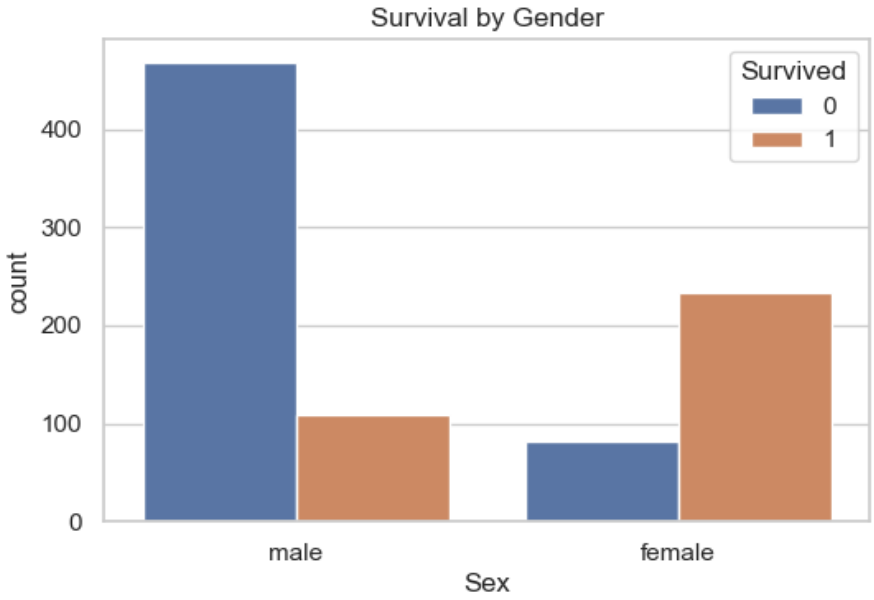
- Class distribution is uneven: There is a clear imbalance across passenger classes, which may influence survival rates and socioeconomic analysis in further exploration.

PLOT 3 - Age Distribution



- Majority of passengers were young adults: Most passengers were between the ages of 20 and 40, with a noticeable peak in the early 20s.
- Age distribution is right-skewed: Fewer passengers were older (above 60), and there's a gradual decline in count with increasing age, indicating a younger population overall.

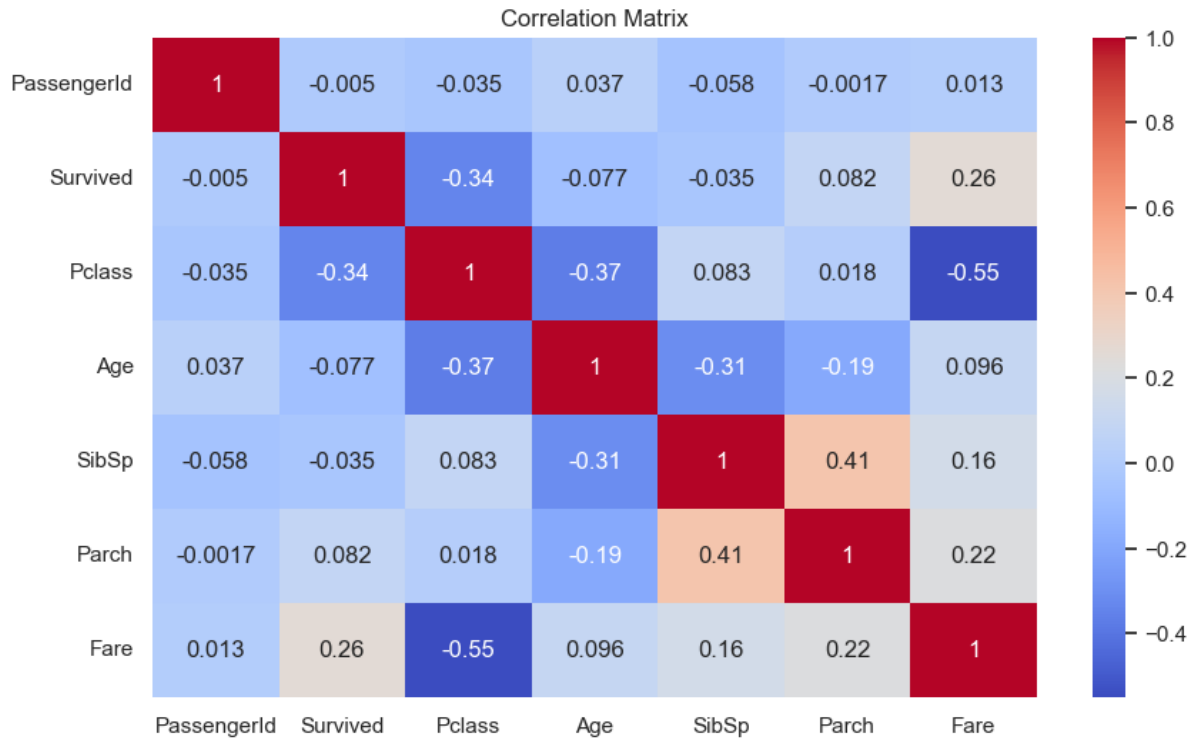
PLOT 4 - Survival by Gender



- Females had a higher survival rate: Most female passengers survived, while most males did not, indicating gender had a strong influence on survival.

- "Women and children first" policy likely applied: The clear gender-based survival disparity suggests that safety protocols prioritized women during the evacuation.

PLOT 5 – Correlation Matrix



- Survival is moderately correlated with Fare and Pclass: Passengers who paid higher fares and those in higher classes (lower Pclass number) were more likely to survive.
- Some multicollinearity exists: There's a moderate positive correlation between SibSp and Parch (0.41), suggesting that family-related features might overlap and should be handled carefully in modeling.

CONCLUSION

1. Female passengers had higher survival rates than males.
2. Passengers in 1st class were more likely to survive.
3. Younger passengers (especially children) had better survival chances.
4. Higher fare seems correlated with higher survival likelihood.
5. Significant missing values in 'Cabin'; may need to drop or impute.
6. Age has missing values that may need imputation.
7. Most passengers embarked from 'S'.