# Titanic Data Analysis: EDA Report

Introduction

This report presents an exploratory data analysis (EDA) of the Titanic dataset, focusing on the socio-demographic factors influencing passenger survival. The analysis aims to uncover patterns and insights related to passenger characteristics such as class, age, family size, and gender, which may have impacted their chances of survival during the tragic sinking of the Titanic. By visualizing and interpreting the data, we seek to enhance our understanding of the factors contributing to survival outcomes in this historical event.

1. Univariate Analysis

**Age Distribution (with histogram and boxplot)**
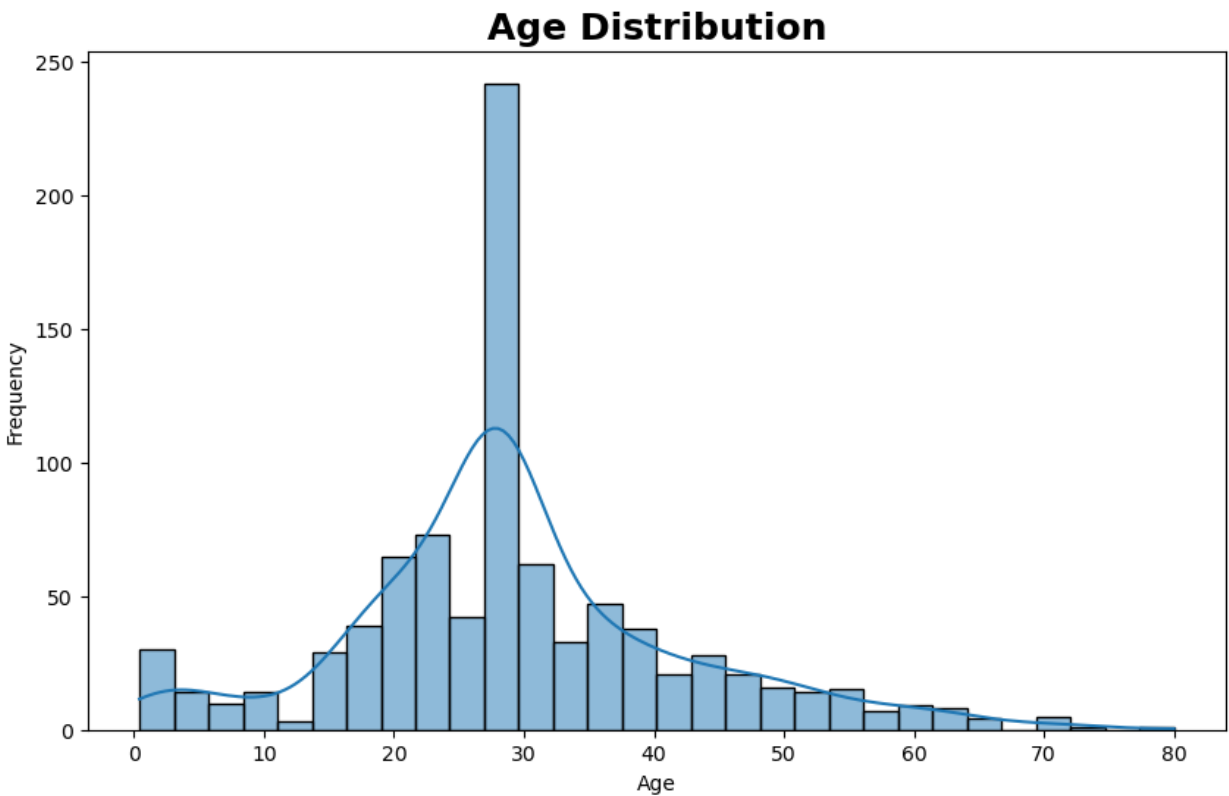
This chart shows the age distribution of a population. The bars represent the number of people in each age group, and the line graph shows the overall trend. The highest frequency is in the 20-30

range, meaning most people in the dataset are in their 20s. The frequencies decrease as you move towards younger and older ages.

Age Distribution With Boxplot
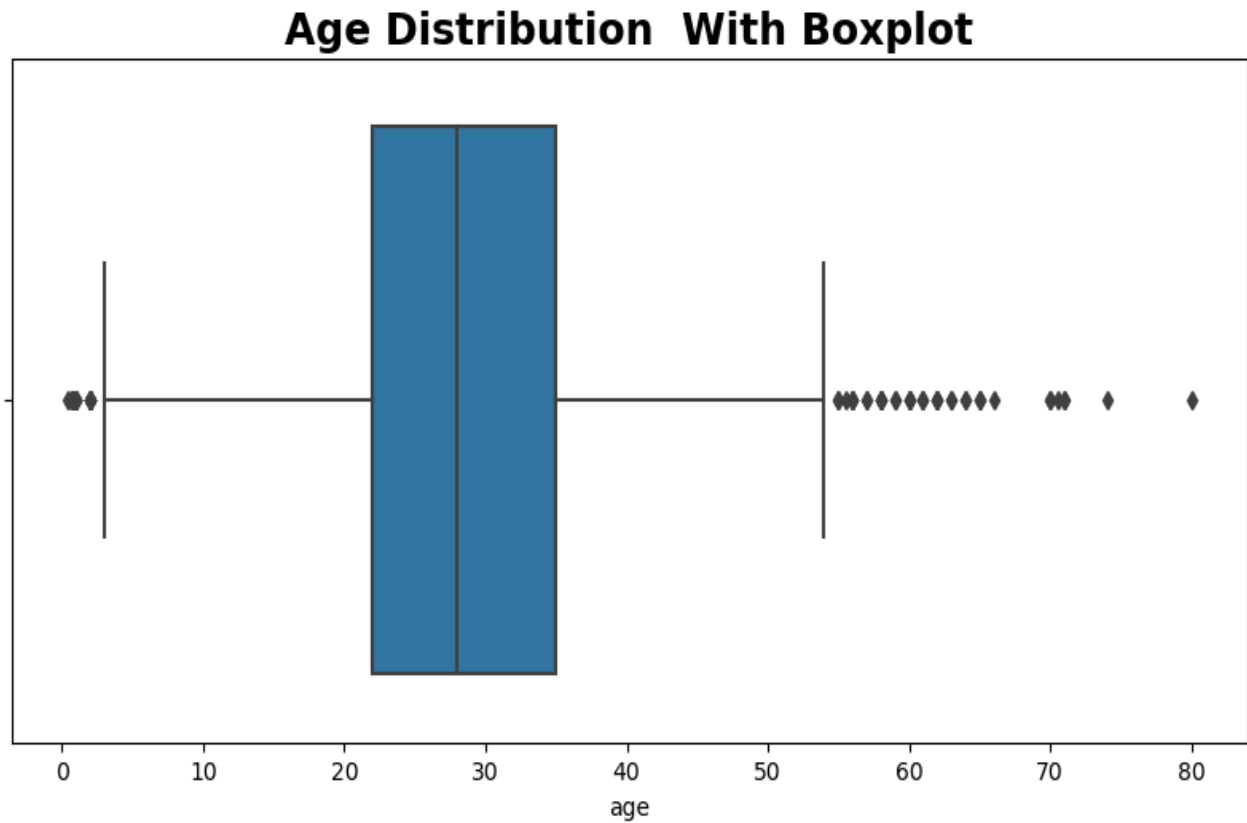
The blue box represents the middle 50% of the data (interquartile range), with the line inside the box indicating the median age. The "whiskers" extend to the smallest and largest values within 1.5 times the interquartile range. Any dots outside the whiskers are outliers. This plot helps you quickly see the spread and central tendency of the ages, as well as any unusual values.
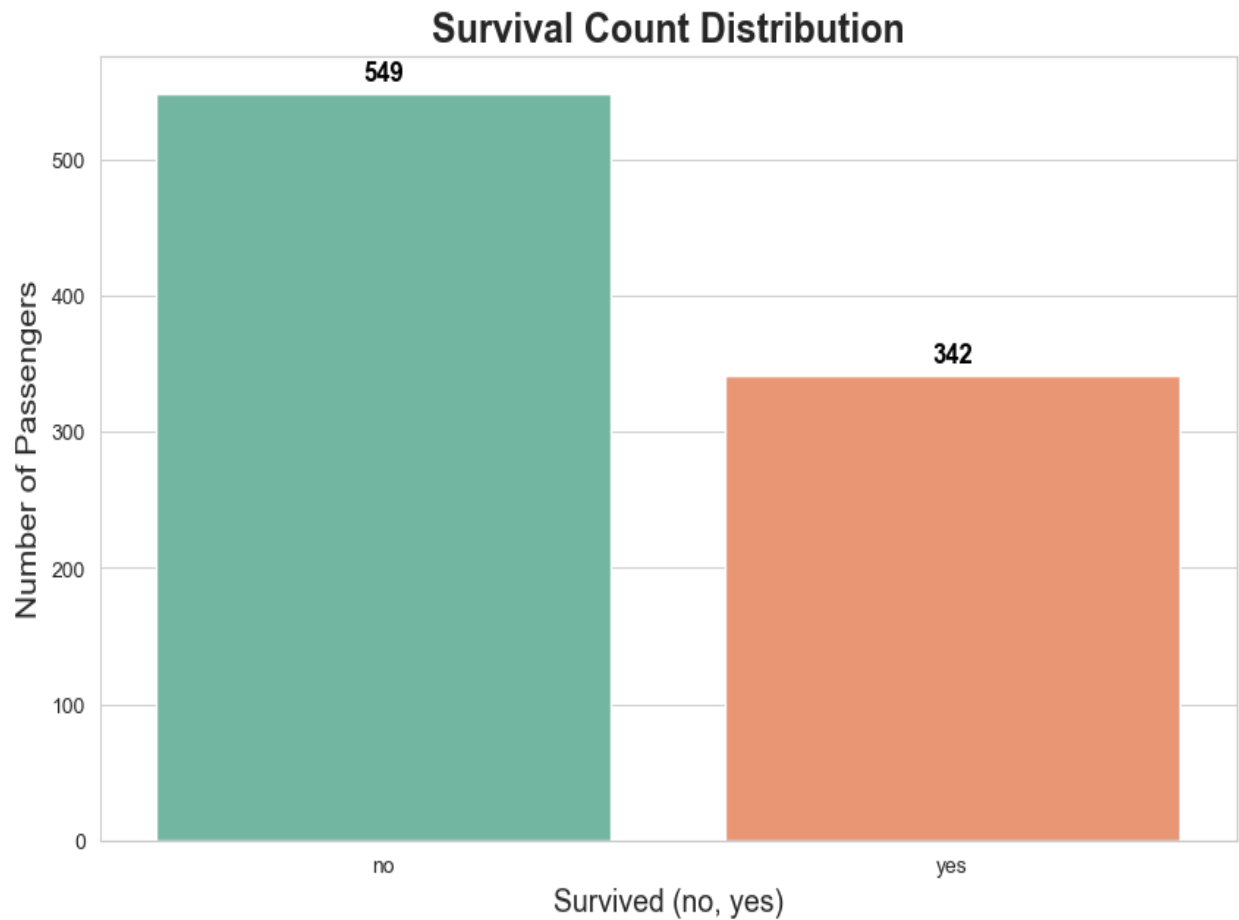
Passenger Class Distribution
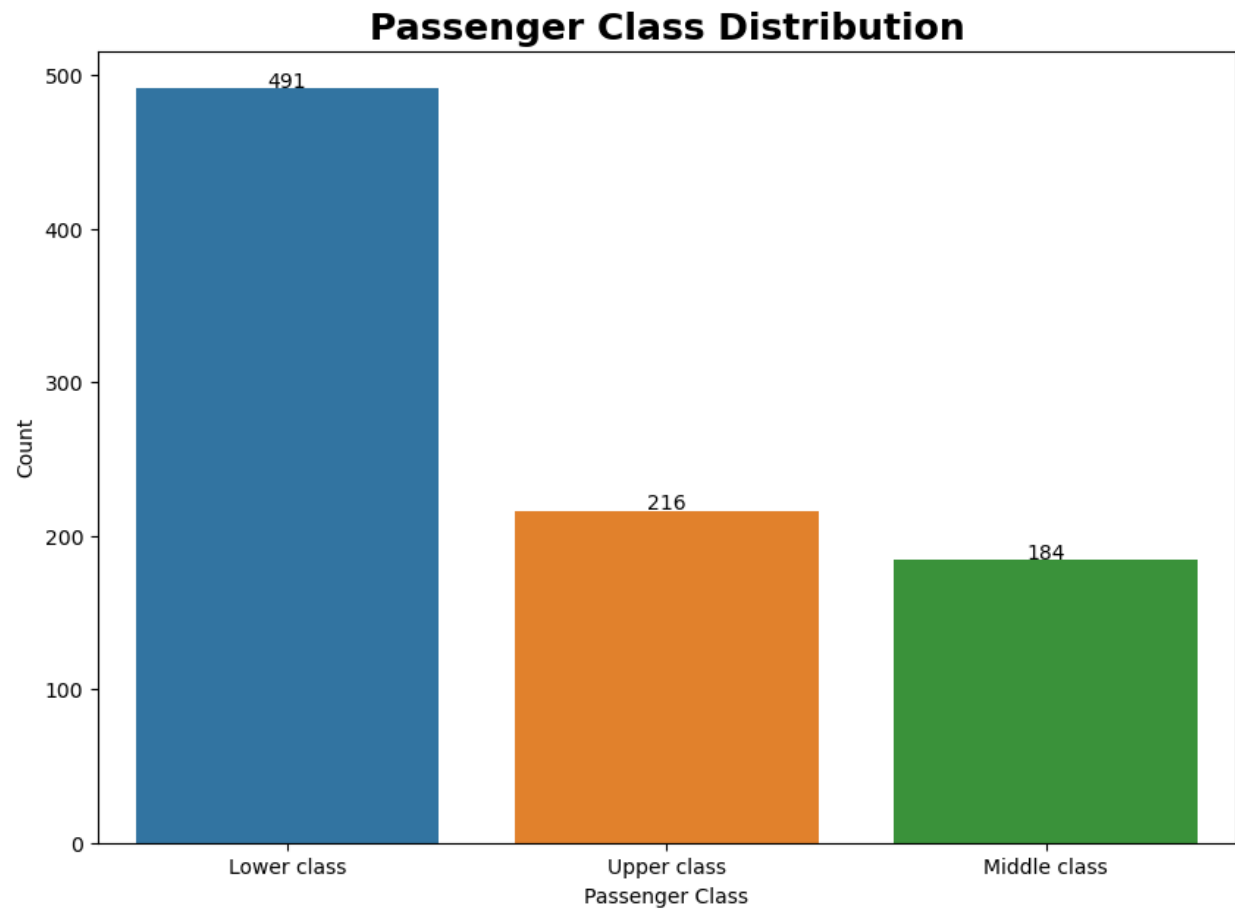
The passenger class distribution reveals that the majority of passengers belonged to the lower class (491), followed by the upper class (216) and middle class (184). This suggests that lower-class passengers were more numerous on board the Titanic.

Family Size Distribution

Family Size Distribution On The Basis of Family Size Column
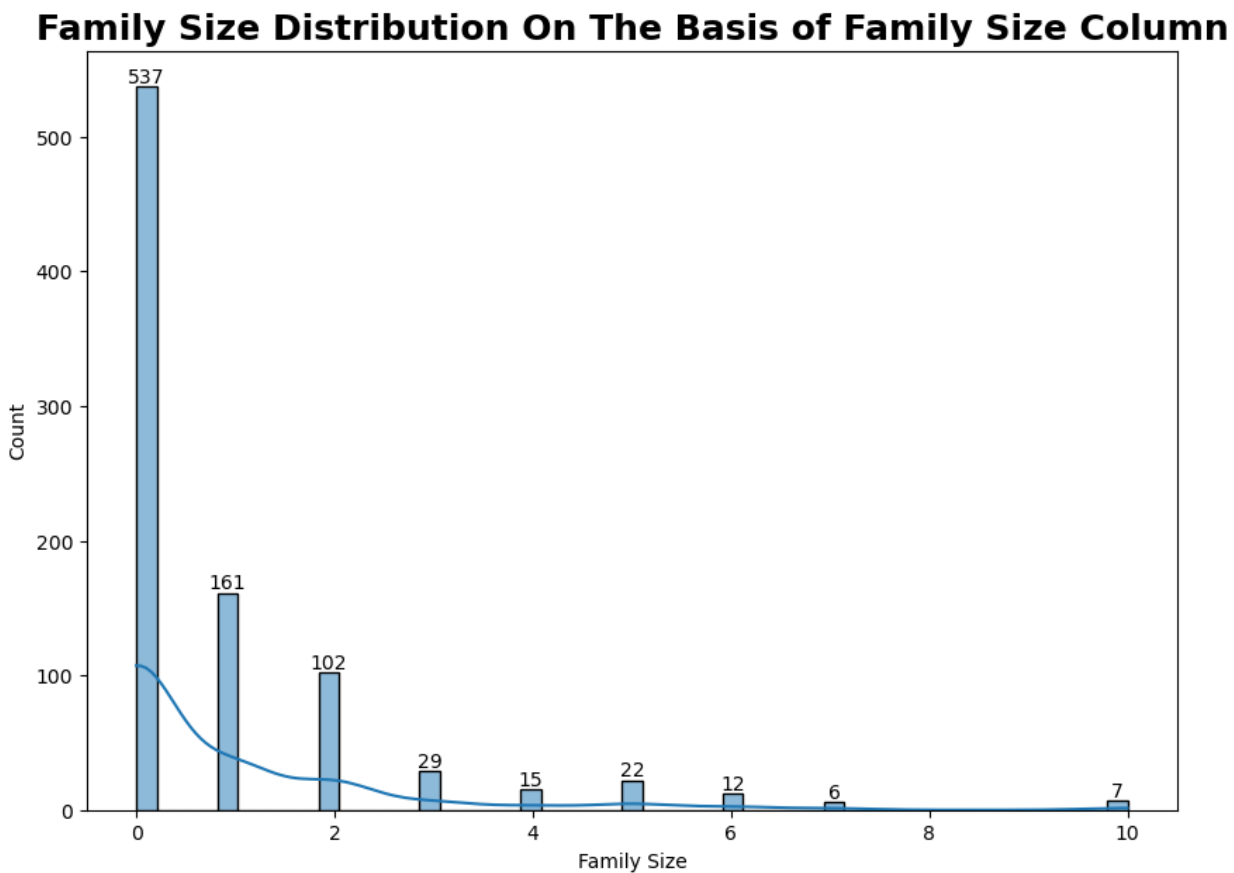
The family size distribution indicates that most passengers traveled alone, with a family size of 0 (537), while only a few passengers traveled with larger families, as evidenced by the lower counts for family sizes greater than 1.
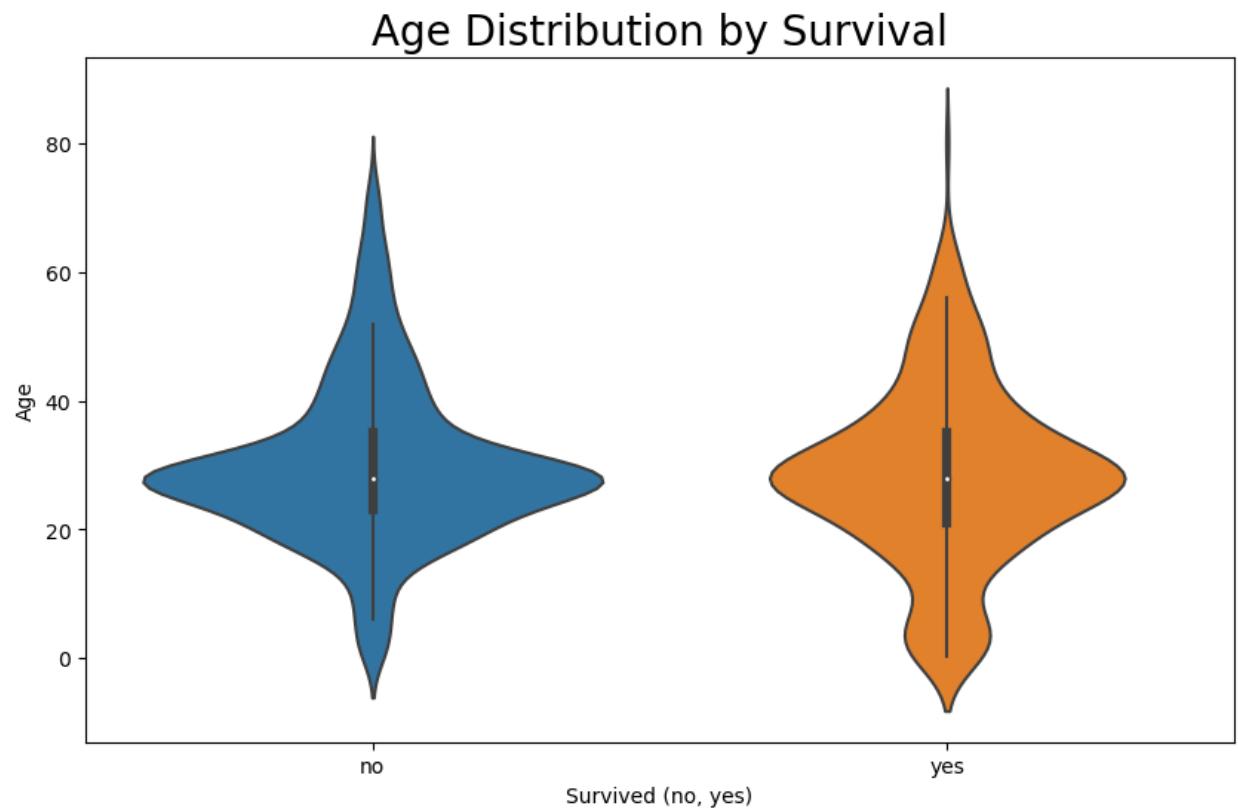
Survival Count Distribution

The survival count distribution shows that 549 passengers did not survive, while 342 passengers survived. This stark contrast highlights the challenges faced by many during the disaster.

2. Bivariate Analysis

Age Distribution for Survival

Age Distribution by Survival

This violin plot shows the age distribution of individuals based on their survival statu. Blue violin (no): Represents the age distribution of those who did not survive. Orange violin (yes): Represents the age distribution of those who survived.

In summary, younger individuals were more likely to not survive, while older individuals had a higher chance of survival

Survival vs Passenger Class

**Survival Rate by Passenger Class**
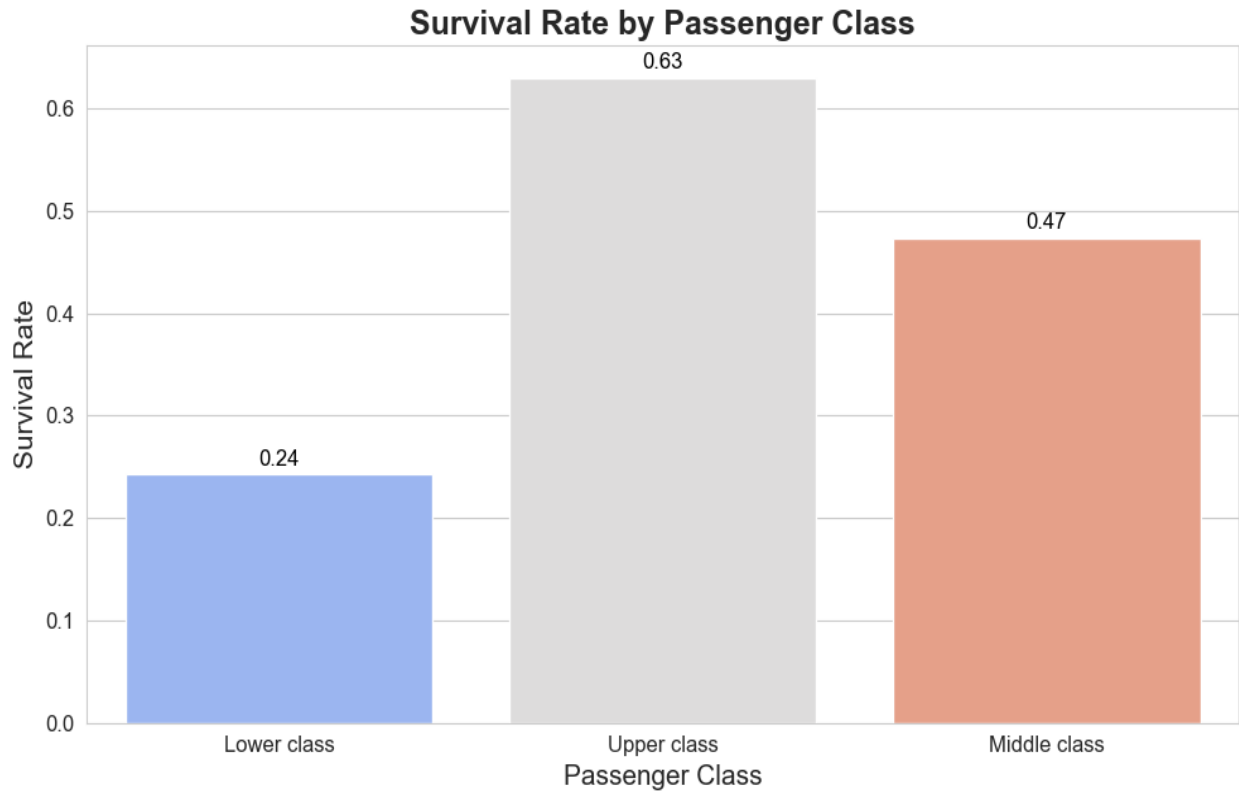
The bar plot indicates that survival rates varied significantly by passenger class, with 24% survival in the lower class, 63% in the upper class, and 47% in the middle class. This suggests that passengers in higher classes had better chances of survival.

Survival vs. Age

Survival vs Age
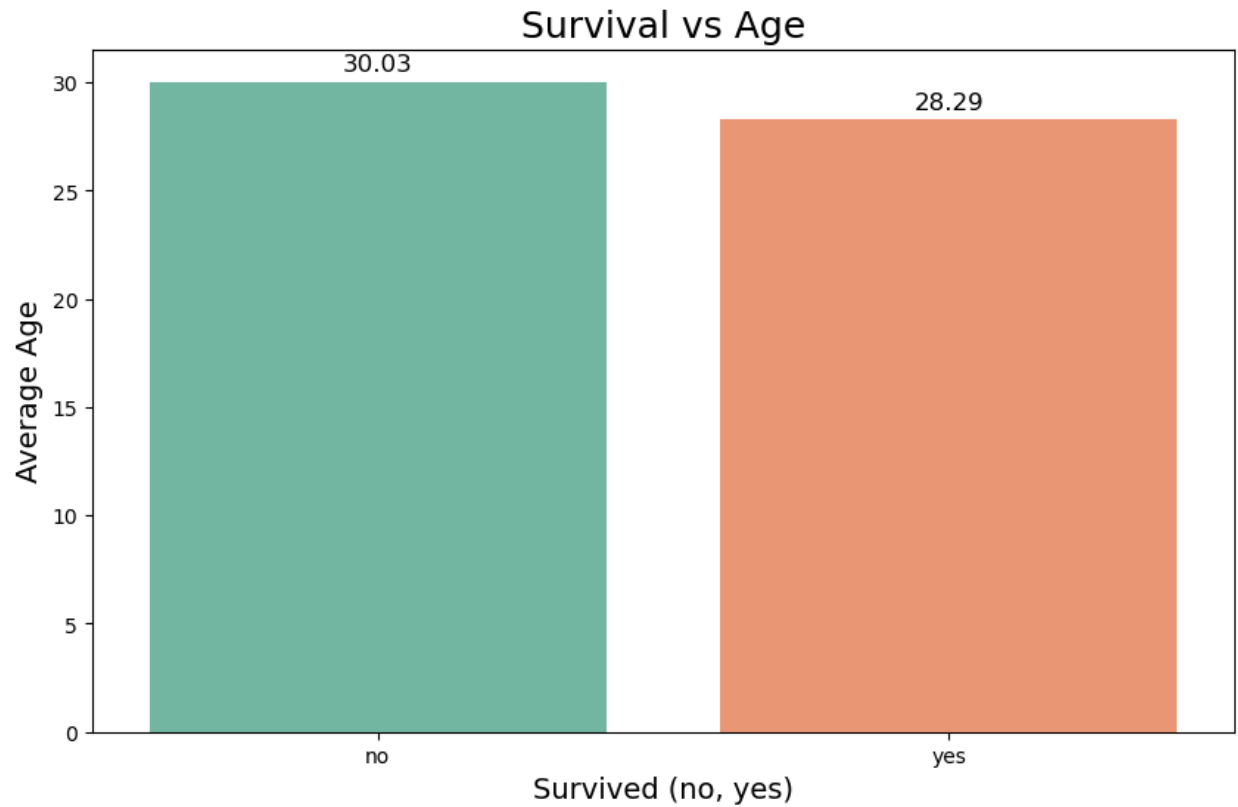
The violin plot shows that younger passengers had higher survival rates, particularly among children, suggesting that age played a crucial role in survival outcomes.


Survival Rate by Family Size
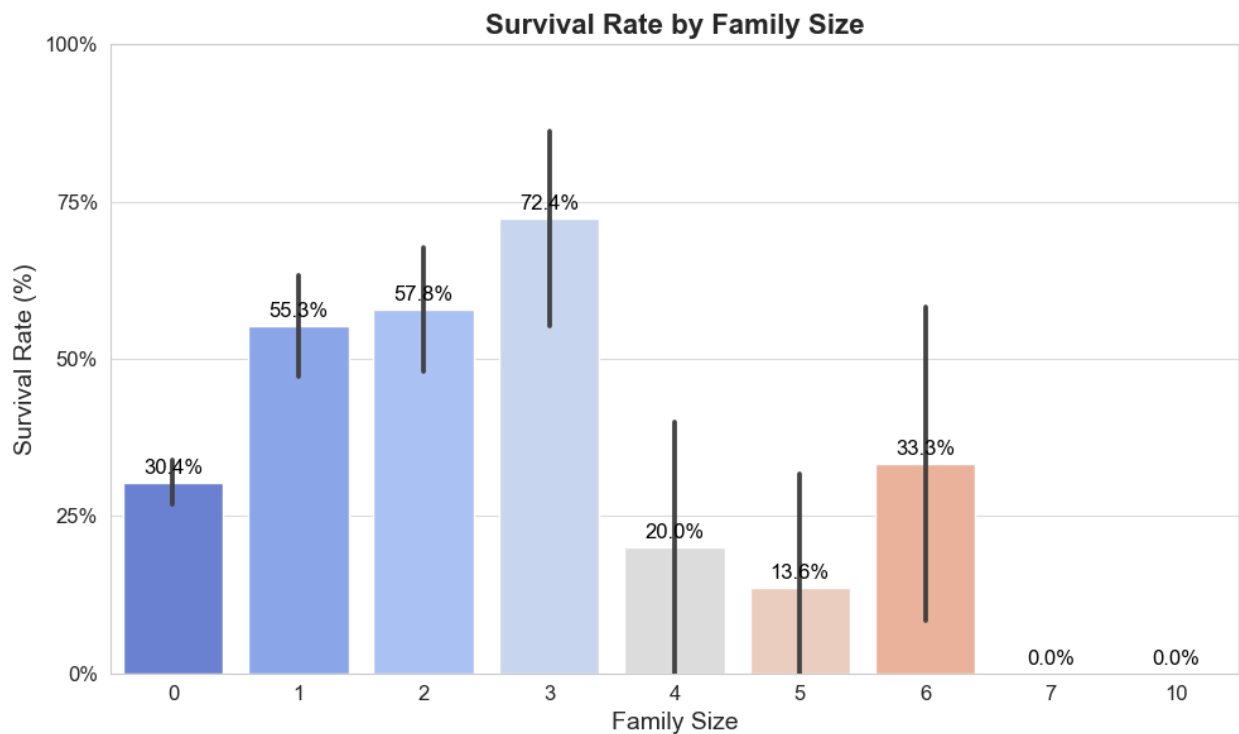
This plot shows that: Family size of 3 has the highest survival rate at around 72.4%. Family sizes of 2 and 1 also have relatively high survival rates at 57.8% and 55.3%, respectively. Family size of 0 (individuals alone) has a survival rate of 30.4%. Larger family sizes (4, 5, 6, and 7) have much lower survival rates, with family sizes of 6 and 7 having 0% survival. This suggests that smaller family sizes, particularly around 3, had better survival rates, while being alone or in very large families was less favorable for survival.
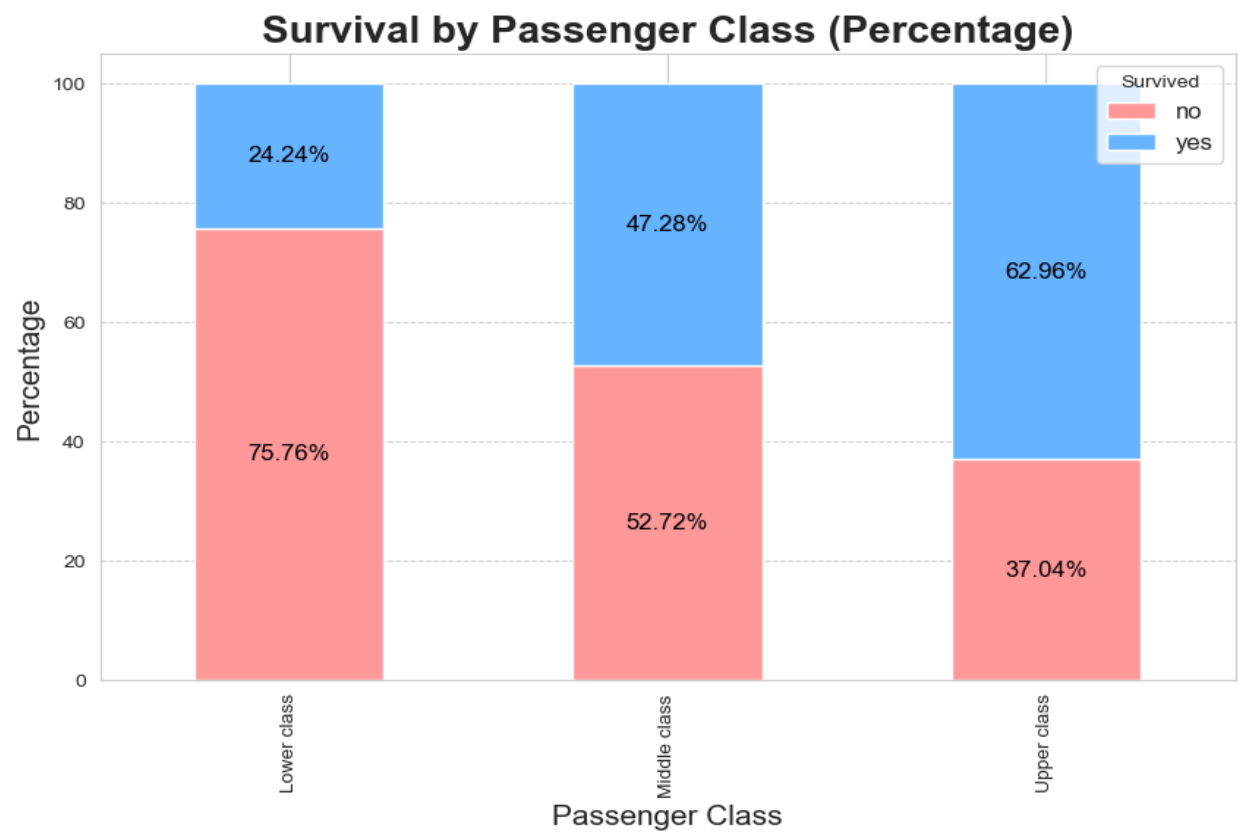
3. Multivariate Analysis
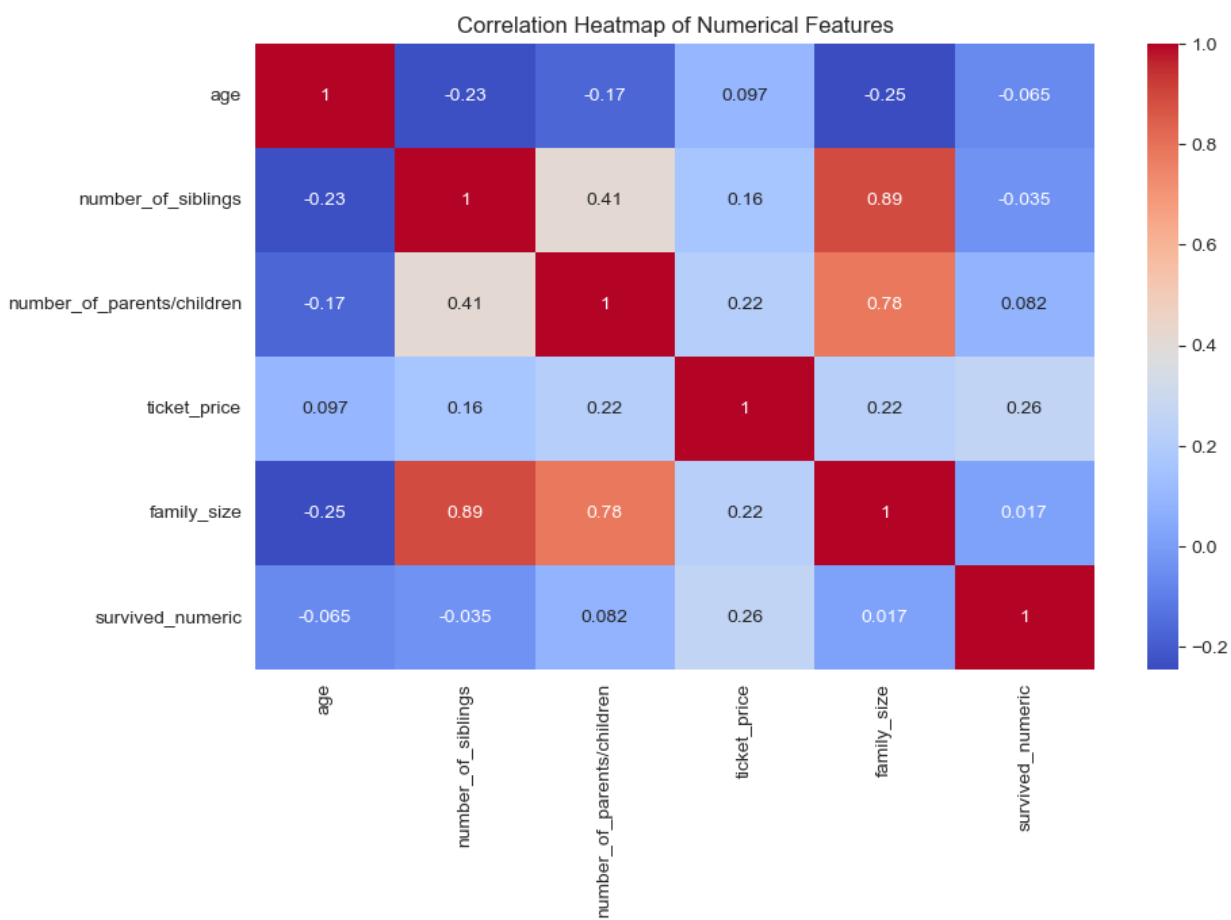
Survival by Age and Passenger Class
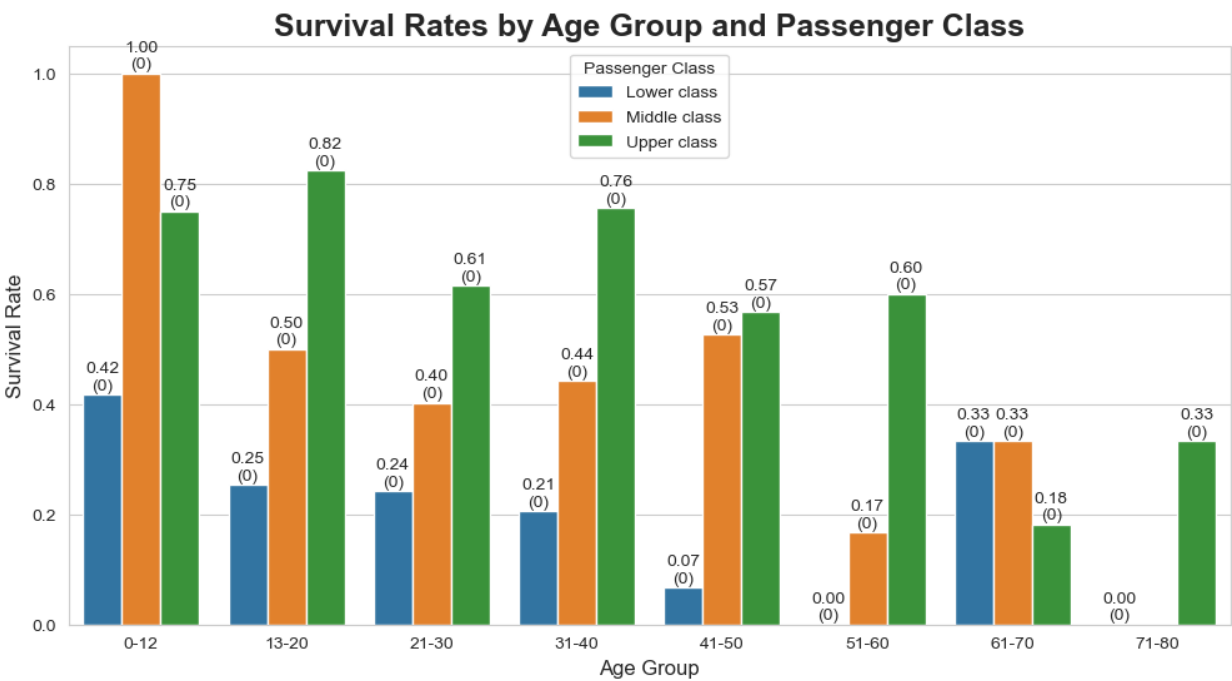
This plot illustrates the interaction between age and passenger class regarding survival, indicating that younger individuals in higher classes had better survival rates.

Survival by Age and Class (Violin Plot)

Survival by Age and Class

Survival by Passenger Class (Percentage)

Correlation Heatmap

The correlation heatmap reveals a positive correlation between age and survival, indicating that younger passengers tended to survive more frequently than older passengers.

Conclusion

This analysis highlights that passenger class, age, and family size significantly influenced survival rates on the Titanic. The findings suggest that individuals in higher socio-economic classes and younger age groups had better survival chances. Understanding these patterns can provide valuable insights into the socio-demographic factors that played a role in this historical tragedy.

Recommendations

➤ **Improved Safety Protocols**: Based on the findings that younger passengers had higher survival rates, it may be beneficial to implement safety protocols that prioritize children and vulnerable passengers in emergency situations.

➤ **Class Awareness**: The analysis indicates that passengers from lower classes had significantly lower survival rates. Future policies should consider equitable access to safety measures for all passengers, regardless of socio-economic status.

➤ **Family Safety**: Given that family size influenced survival outcomes, it is recommended that safety procedures be designed to facilitate family evacuations, ensuring that family members can stay together during emergencies.

➤ **Education and Training**: Increased training for crew members on how to assist passengers effectively during evacuations could improve overall survival rates, especially for those in vulnerable groups.

➤ **Further Research**: Conduct further analyses on additional variables (such as ticket price, cabin location, etc.) to gain deeper insights into survival outcomes and improve future maritime safety standards.