TITANIC DATASET

ANALYSIS

Presented by: Lavender Echessa

OVERVIEW

01
Introduction

02
Problem Statement

03
Goals

04
Data Understandingand Preparation

05
EDA

06
Conclusion

07
Recommendation

INTRODUCTION

The Titanic dataset is a well-known dataset in the field of data science and serves as a valuable resource for exploring and understanding various data analysis. This dataset contains information about the passengers aboard the RMS Titanic, including whether they survived or not, passenger class, age, gender, fare, and more. The dataset is widely used for predictive modeling, hypothesis testing, and exploratory data analysis.

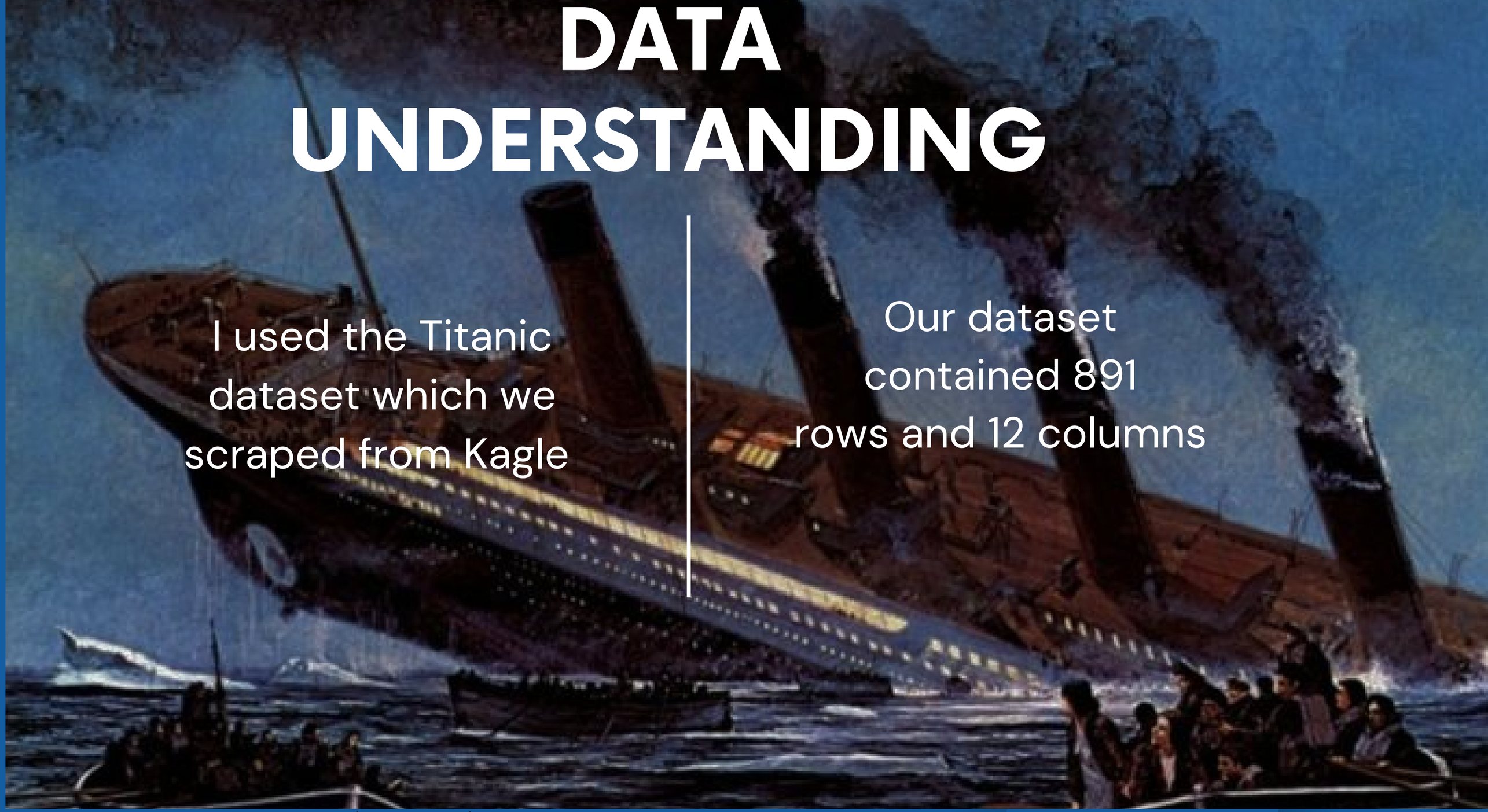
PROBLEM STATEMENT

Perform data cleaning and exploratory data analysis (EDA) on Titanic dataset from Kaggle. Explore the relationships between variables and identify patterns and trends in the data.

GOALS

1. Explore the Titanic dataset with the goal of gaining insights into passenger demographics, survival rates, and the factors that contributed to survival.
2. perform data cleaning to ensure the dataset is ready for analysis and conduct exploratory data analysis (EDA) to visualize and understand the data better.
3. investigate relationships between variables and identify patterns and trends that may shed light on the events that unfolded during the Titanic's ill-fated voyage.

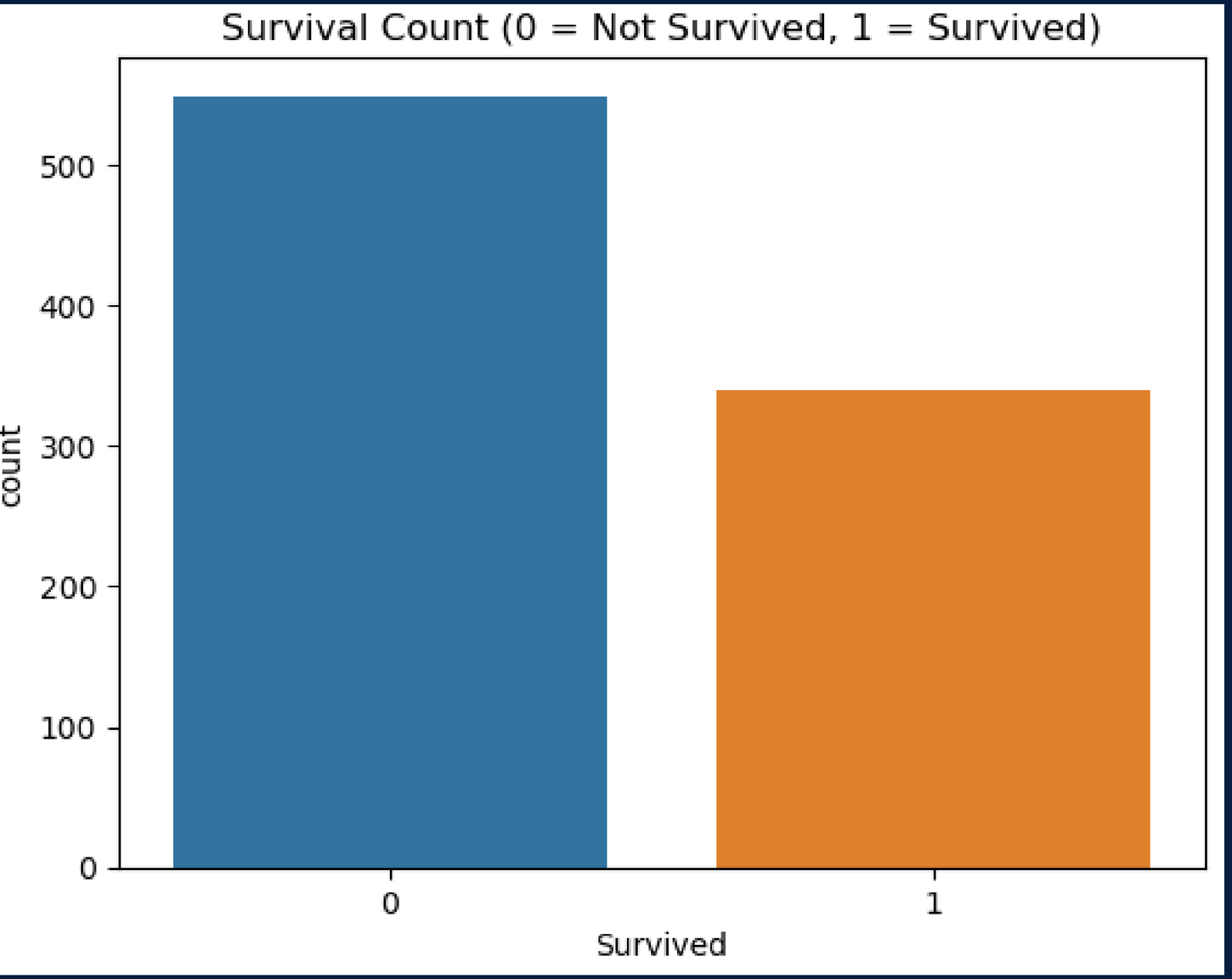
DATA UNDERSTANDING



I used the Titanic dataset which we scraped from Kagle

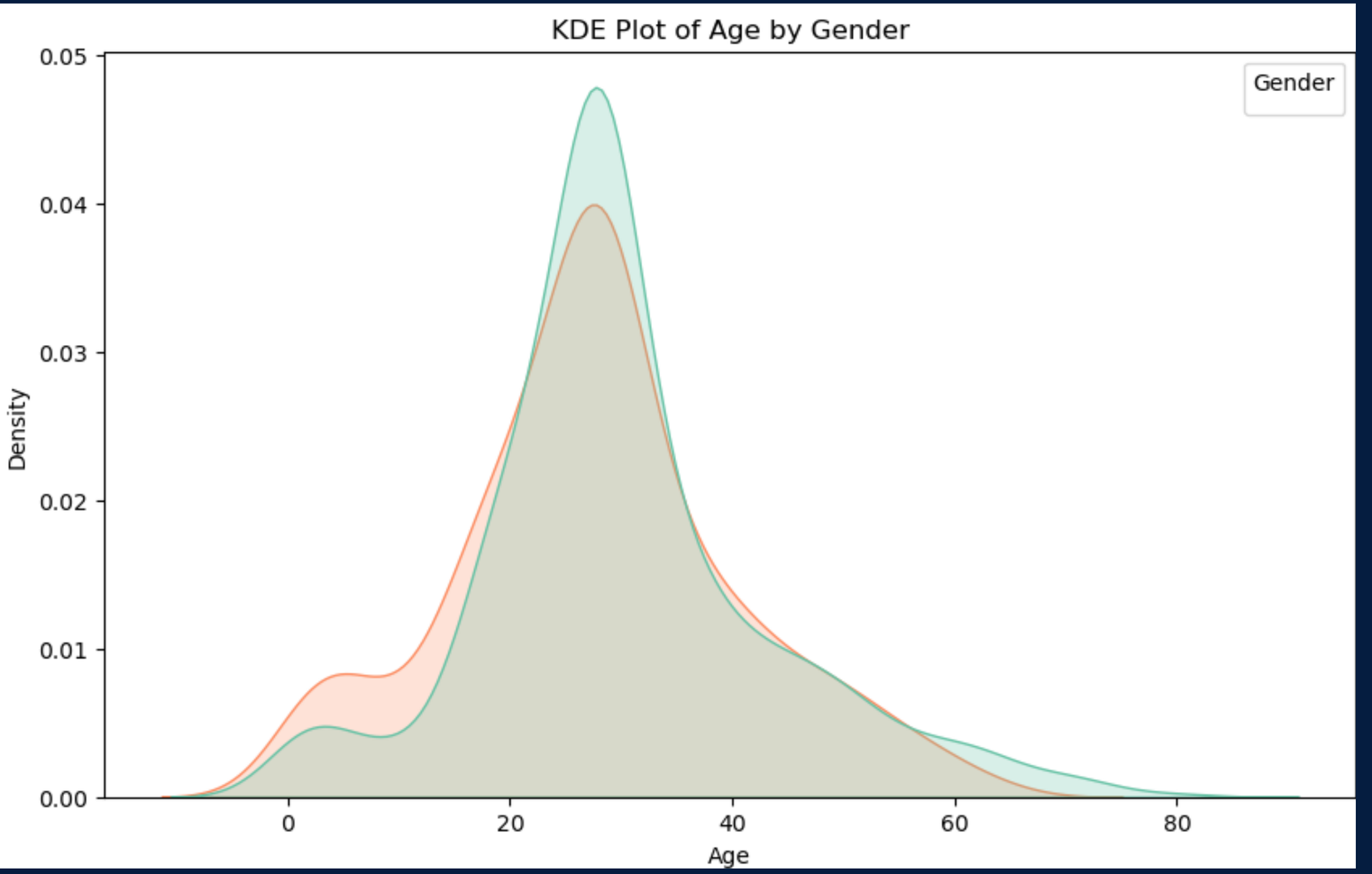
Our dataset contained 891 rows and 12 columns

SURVIVAL ANALYSIS



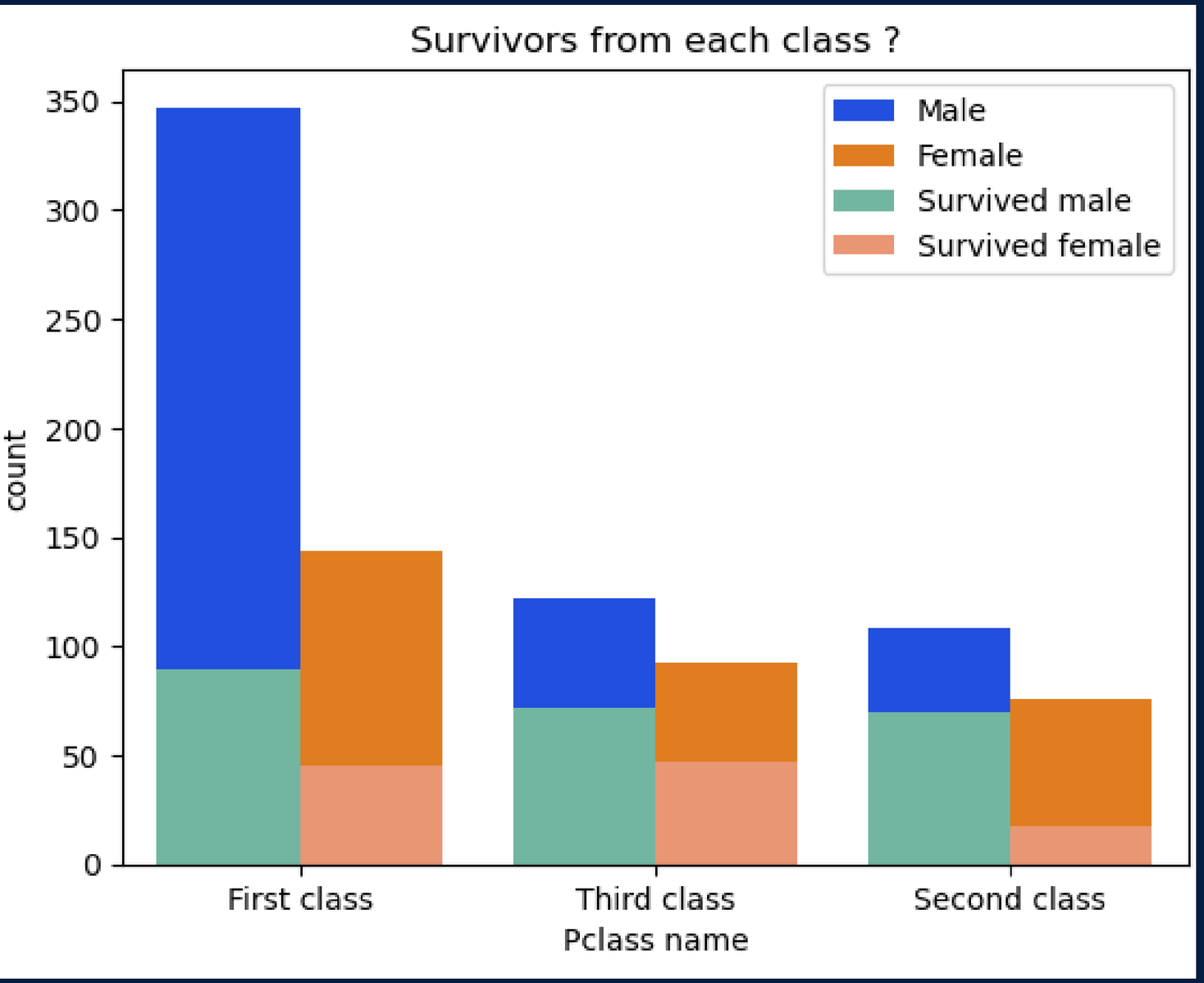
here we can see that 424 people did not survived (died) and 288 people survived

SURVIVAL BY GENDER ANALYSIS



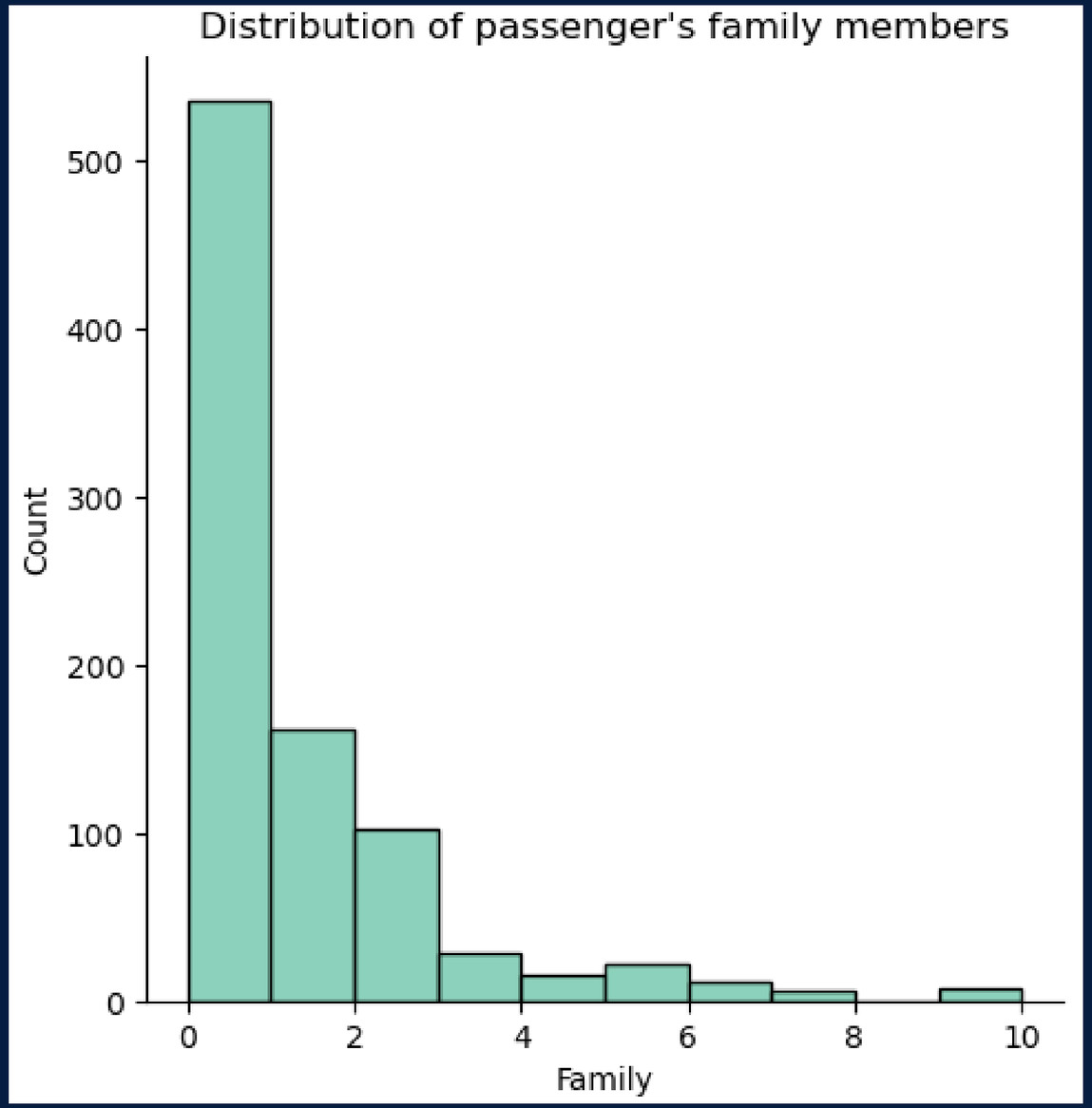
The peaks in each shaded area indicate where the age distribution is most concentrated for each gender. for males, there is a peak around 25-30 years old, while for females, there is a peak around 20-25 years old. In age ranges, between 20 and 30 years old, there is overlap, indicating that passengers of both genders are present in similar age groups. there is a broader distribution of ages among males, while the female age distribution appears to be slightly more concentrated.

SURVIVAL BY CLASS ANALYSIS



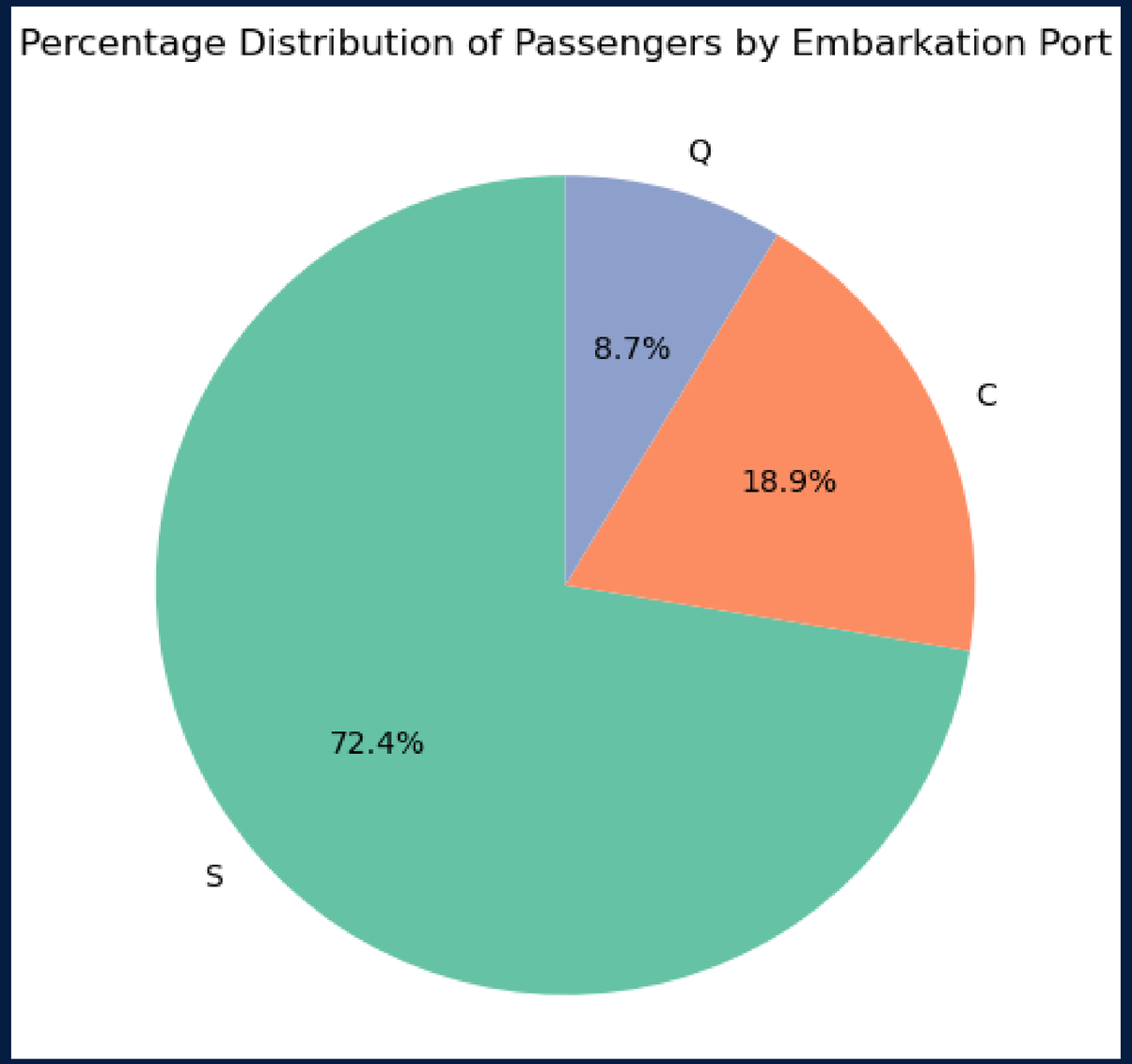
In all class, it appears that more males survived compared to females. Higher-class passengers (1st class) seem to have higher survival rates compared to lower-class passengers (3rd class).

FAMILYDISTRIBUTION ANALYSIS



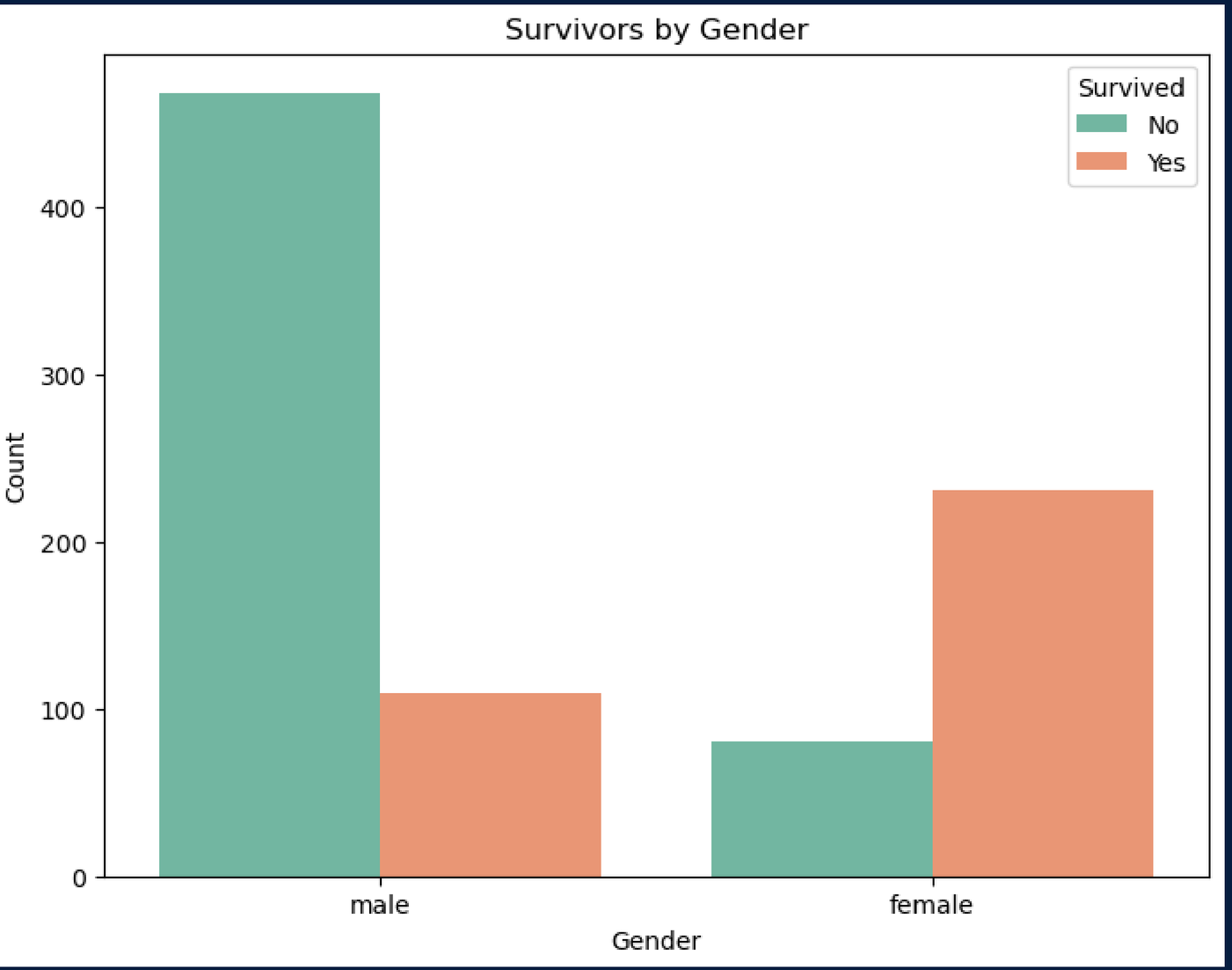
This visualization provides an overview of the family size distribution among passengers. It indicates that a substantial portion of passengers traveled without immediate family members.

EMBARKED DISTRIBUTION ANALYSIS



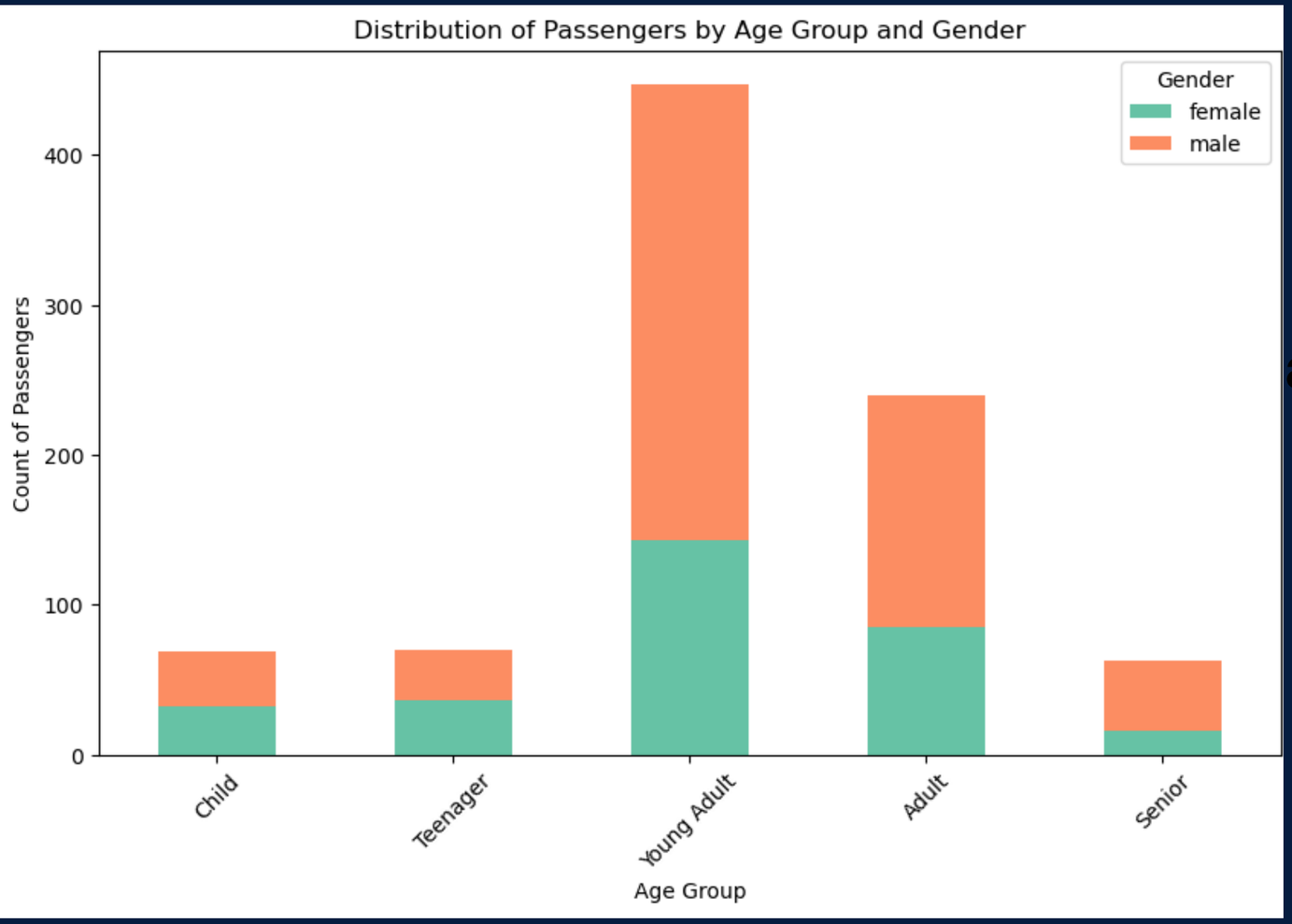
The pie chart shows that the majority of passengers embarked from Southampton ('S'), followed by Cherbourg ('C'). Queenstown ('Q') had the smallest percentage of passengers.

SURVIVAL BY GENDER ANALYSIS



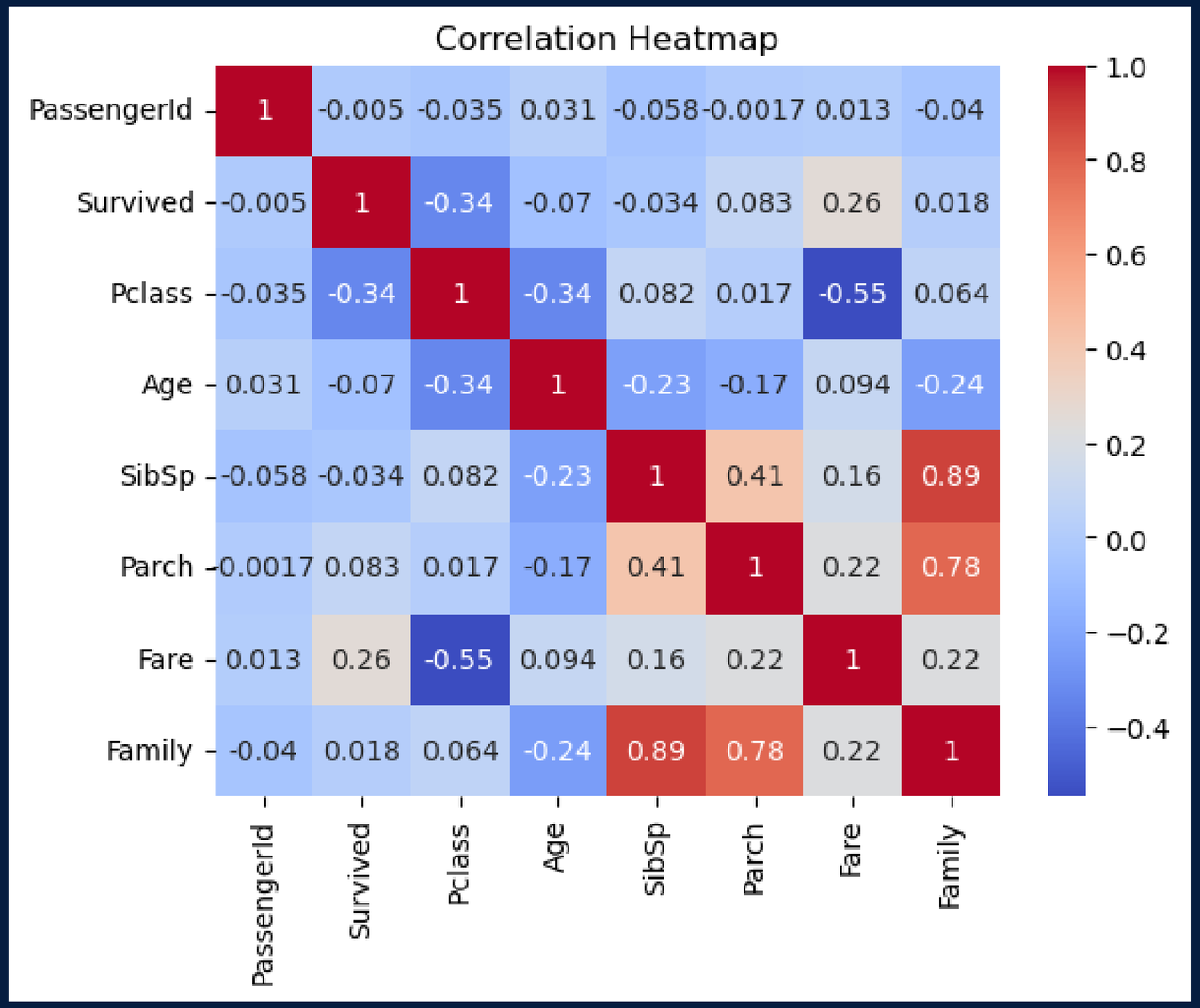
The chart shows survival rates between genders. It is evident that a higher proportion of females survived compared to males. From this visualization, one can infer that gender played a crucial role in determining survival chances on the Titanic. Females had a notably higher survival rate compared to males.

DISTRIBUTION OF PASSENGERS BY AGE GROUP AND GENDER



The chart shows the gender distribution within each age group. it shows how many males and females fall into each age category. In some age groups, you can observe a gender imbalance. For instance, in the 'Child' age group, there appears to be a relatively balanced distribution between males and females. However, in other age groups, there may be more pronounced gender differences.

CORRELATION



Positive correlation values (closer to 1) indicate that when one variable increases, the other tends to increase as well. Negative correlation values (closer to -1) indicate that when one variable increases, the other tends to decrease. Correlation values closer to 0 suggest little to no linear relationship between the variables.

CONCLUSSION

Gender-Based Survival Discrepancy: Females had a notably higher survival rate compared to males, reflecting the adherence to the "Women and Children First" principle during evacuation.

Passenger Class and Survival:First-class passengers had higher survival rates compared to those in lower classes, emphasizing the socio-economic disparities present during the time of the Titanic's voyage.

Age Group and Survival: While age did not show as stark a difference as gender and class, there were patterns in survival rates among different age groups. Younger passengers, especially children, had higher survival rates.

Family Size Impact: Our analysis revealed that passengers traveling with larger families (spouse, siblings, parents, or children) had varying survival rates, suggesting complex dynamics in evacuation decisions.

RECOMMENDATIOS

Based on our analysis, I offer the following recommendations:

Safety and Evacuation Training: Future travel and maritime safety protocols should consider incorporating gender-neutral safety training to ensure a fair distribution of survival resources during emergencies.

Economic Equity: Efforts to promote economic equity should be a priority in the design and allocation of passenger class facilities and resources on modern transportation vessels.

Family-Centric Planning: Emergency planning should account for the presence of families and their unique needs during evacuation scenarios, ensuring coordinated efforts and resources.

Fare Transparency: Travel companies should adopt transparent pricing strategies that reflect fare variations based on age, class, and other factors

contact
email- lavenderechessa@gmail.com
github-https://github.com/lavEche
linkedIn-
https://www.linkedin.com/in/lavender-echessa-6a423520a/