# Titanic Data Analysis

## Overview

The sinking of the Titanic is one of the most infamous shipwrecks in history.
On April 15, 1912, during her maiden voyage, the widely considered “unsinkable” RMS Titanic sank after colliding with an iceberg.
Unfortunately, there weren’t enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew.
While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

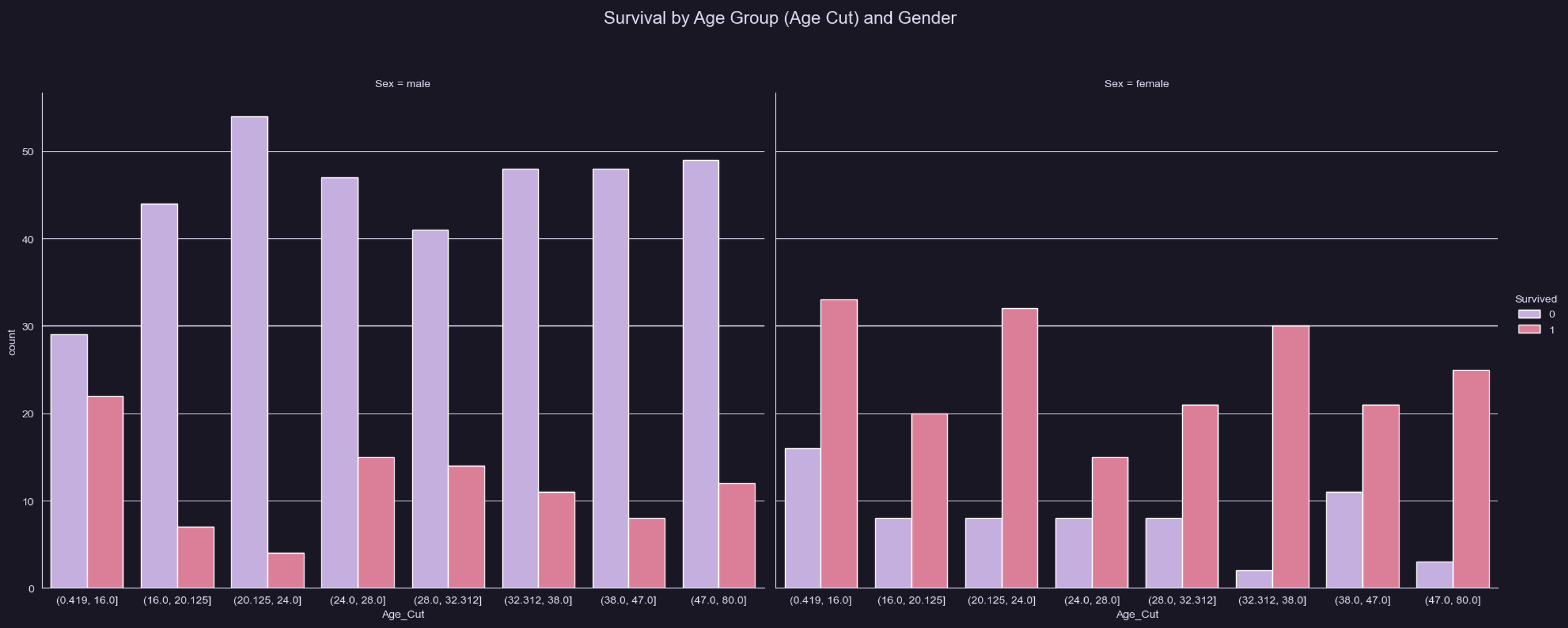
## Objective

To take a deep dive into the Titanic dataset, clean the raw data to ensure its quality and relevance.
Perform a thorough data analysis to identify key factors influencing survival rates.
Build a predictive model using machine learning techniques to accurately predict whether a passenger survived the disaster based on their attributes.

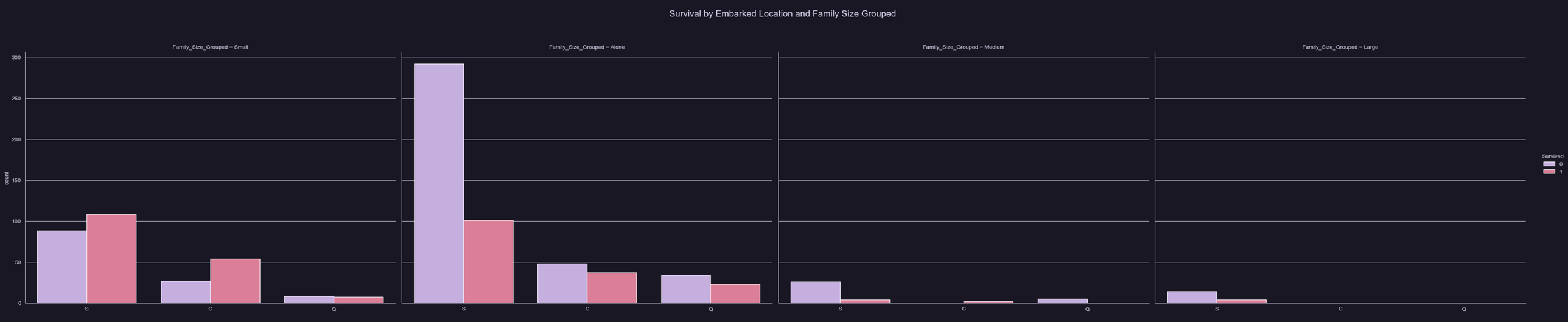
Explore the Titanic Survival Prediction Project!
Scan the QR code to view the complete analysis, code, and insights on GitHub.

project by Sarvamm

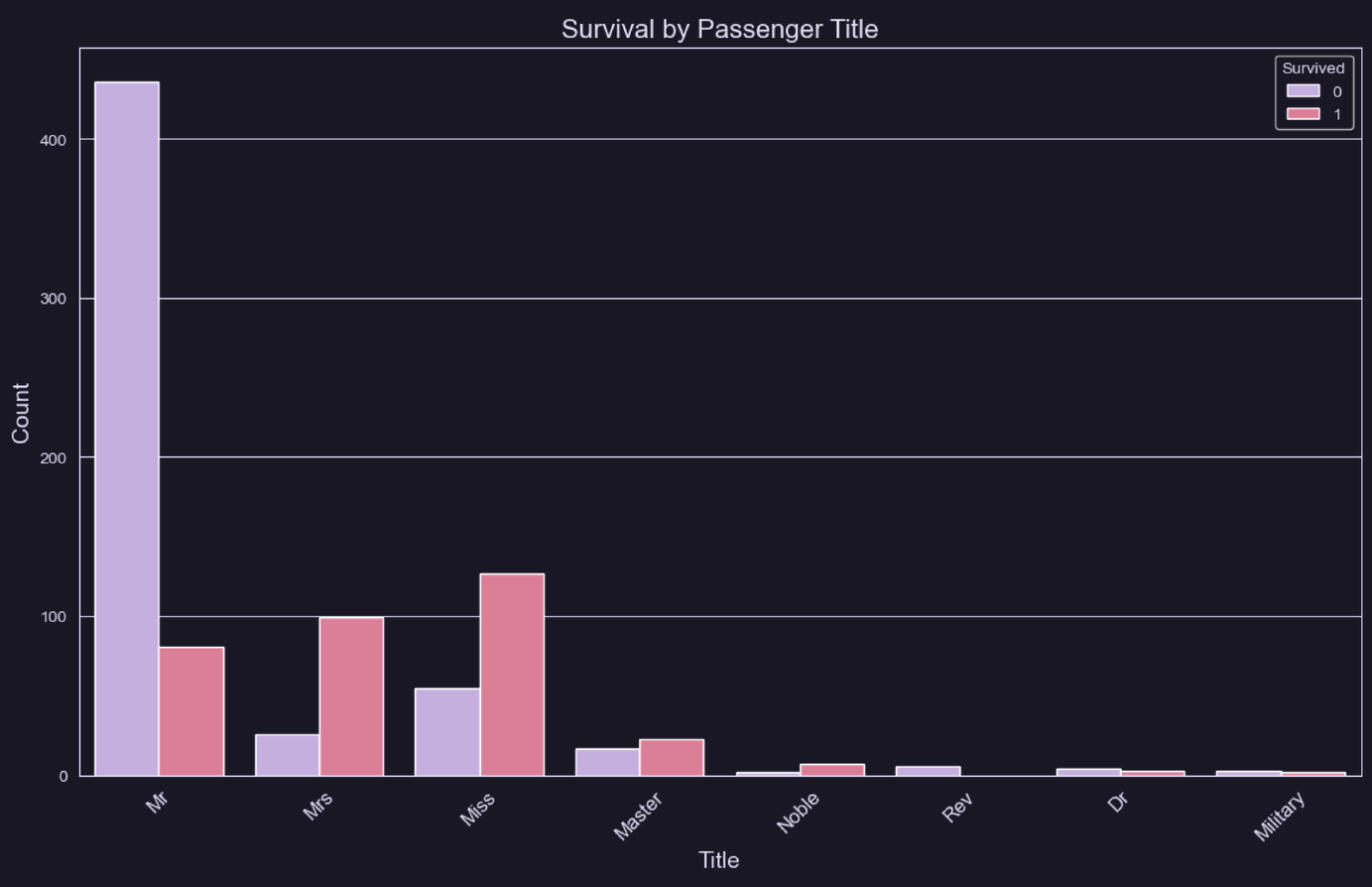
# The Data




The graph indicates that children and women, especially younger ones, had the highest survival rates on the Titanic, reflecting the "women and children first" protocol. Older men had the lowest survival rates, while women in middle age (30-50 years) were among the likeliest to survive.

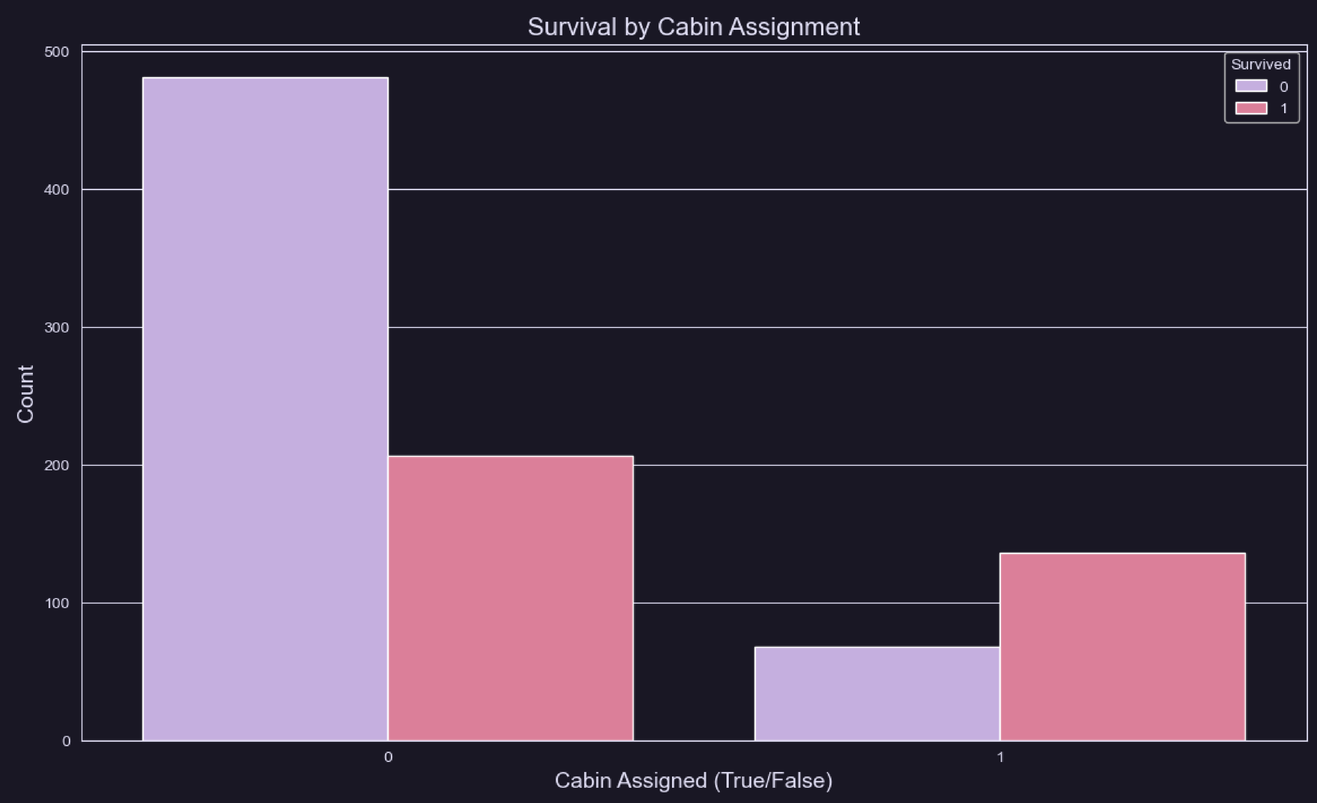


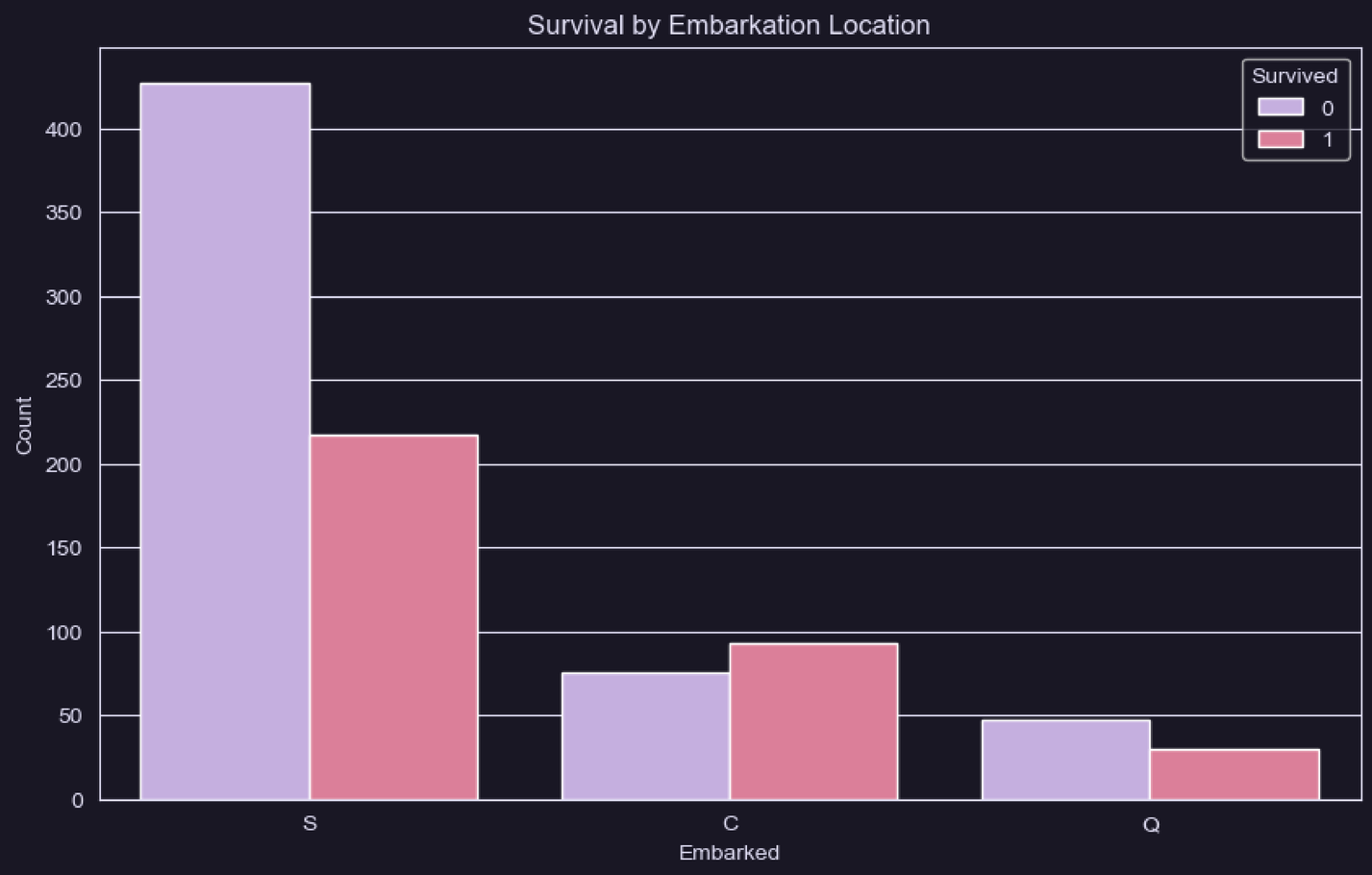

This chart gives us a clear look at how survival rates on the Titanic varied based on family size and where people boarded. Smaller families and lone travelers from Cherbourg had a higher survival rate, while larger families from Southampton faced lower chances of survival.



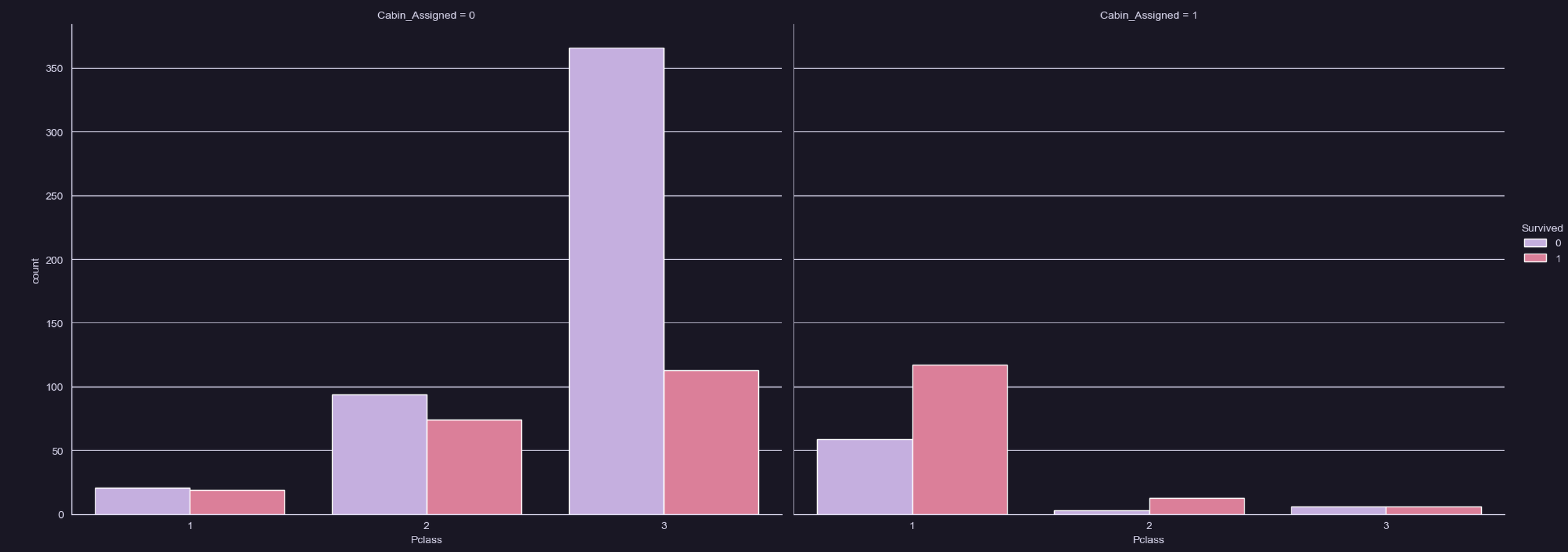

Survival by Passenger Title: This bar chart shows the count of Titanic passengers with different titles (Mr, Mrs, Miss, Master, Noble, Rev, Dr, Military) and their survival status. Men with the title "Mr" faced the highest non-survival rate, whereas titles like "Mrs" and "Miss" had higher survival rates. Fewer passengers had titles like "Master," "Noble," "Rev," "Dr," and "Military," with varying survival rates.

The bar chart shows the survival rates based on cabin assignment on the Titanic. Those assigned a cabin had higher survival rates compared to those without a cabin.

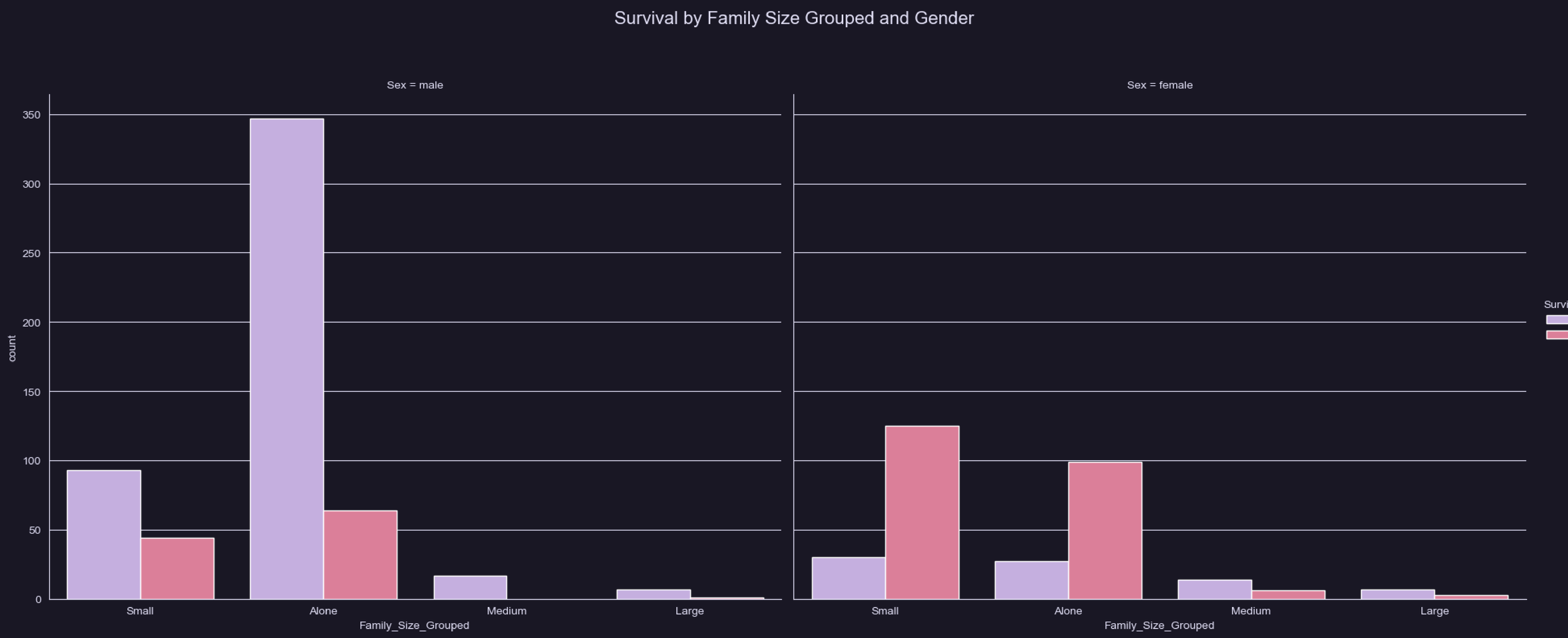


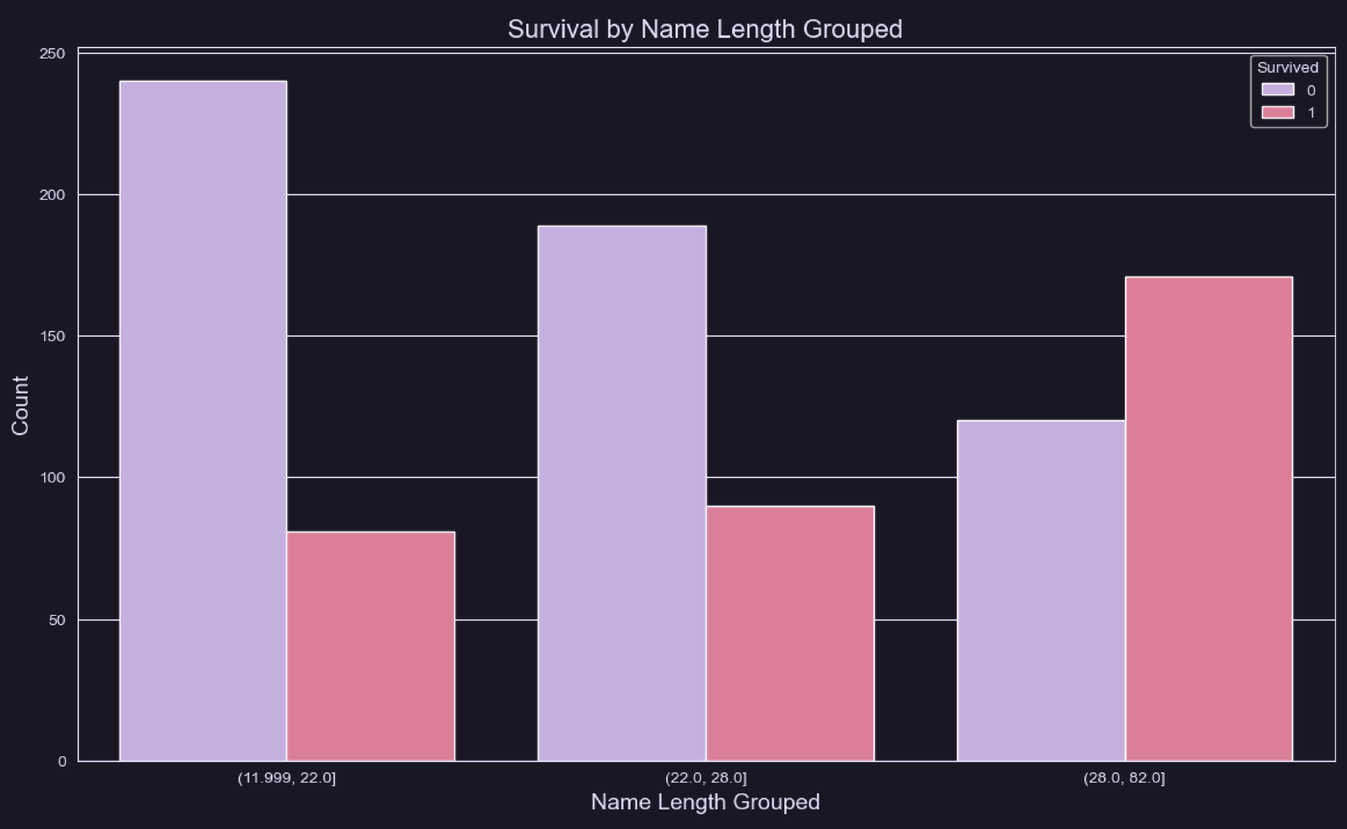

Passengers who embarked from Cherbourg had a slightly higher survival rate than those from Southampton and Queenstown. Southampton had the highest count of non-survivors, while Cherbourg had the most survivors




- Class Disparity: Passengers in first class had a much higher survival rate, especially if they were assigned a cabin.
- Cabin Influence: Having a cabin significantly increased survival chances across all classes.
- Third Class Struggle: Most third-class passengers, especially those without cabins, faced the lowest survival rates.

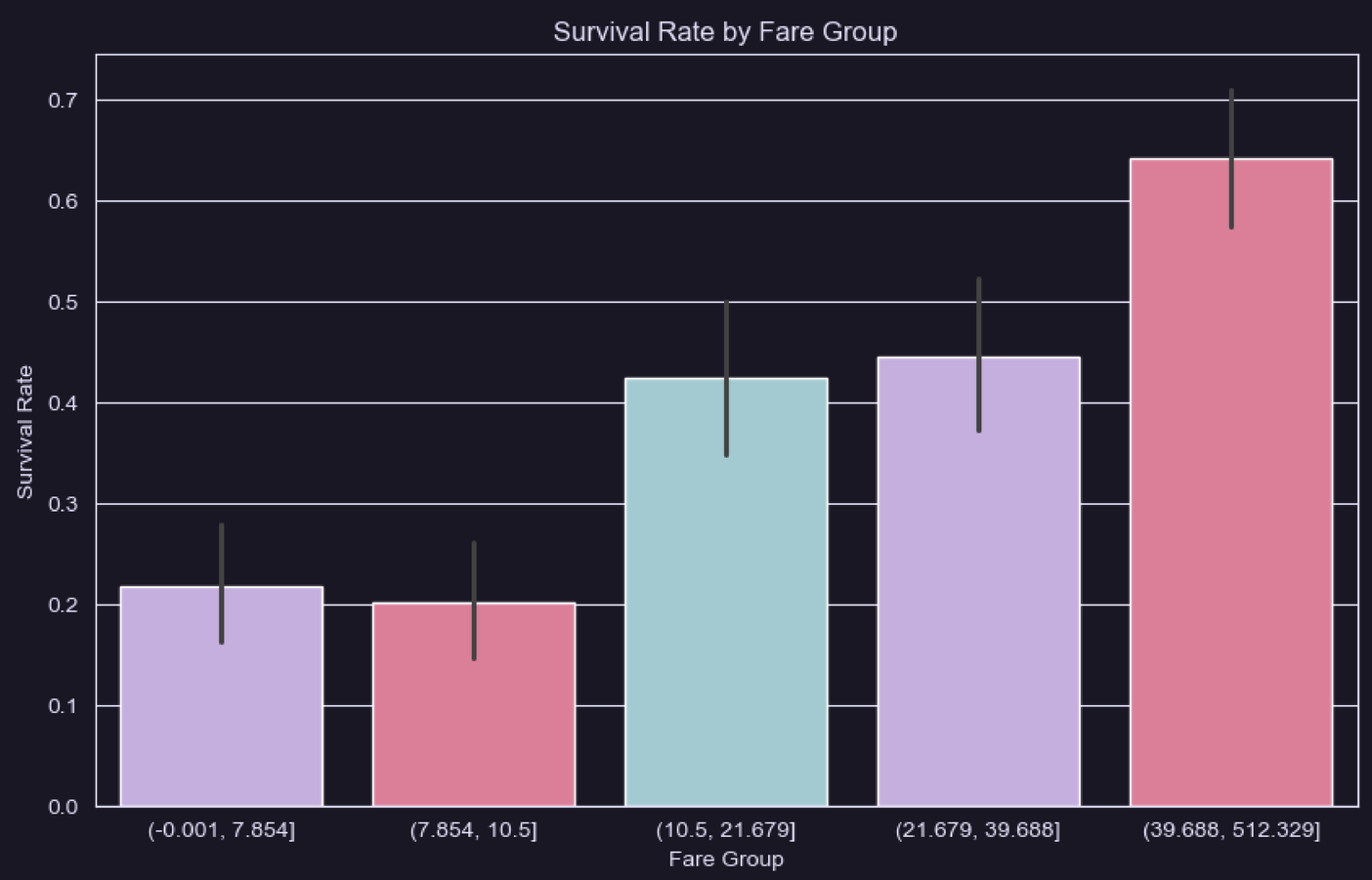


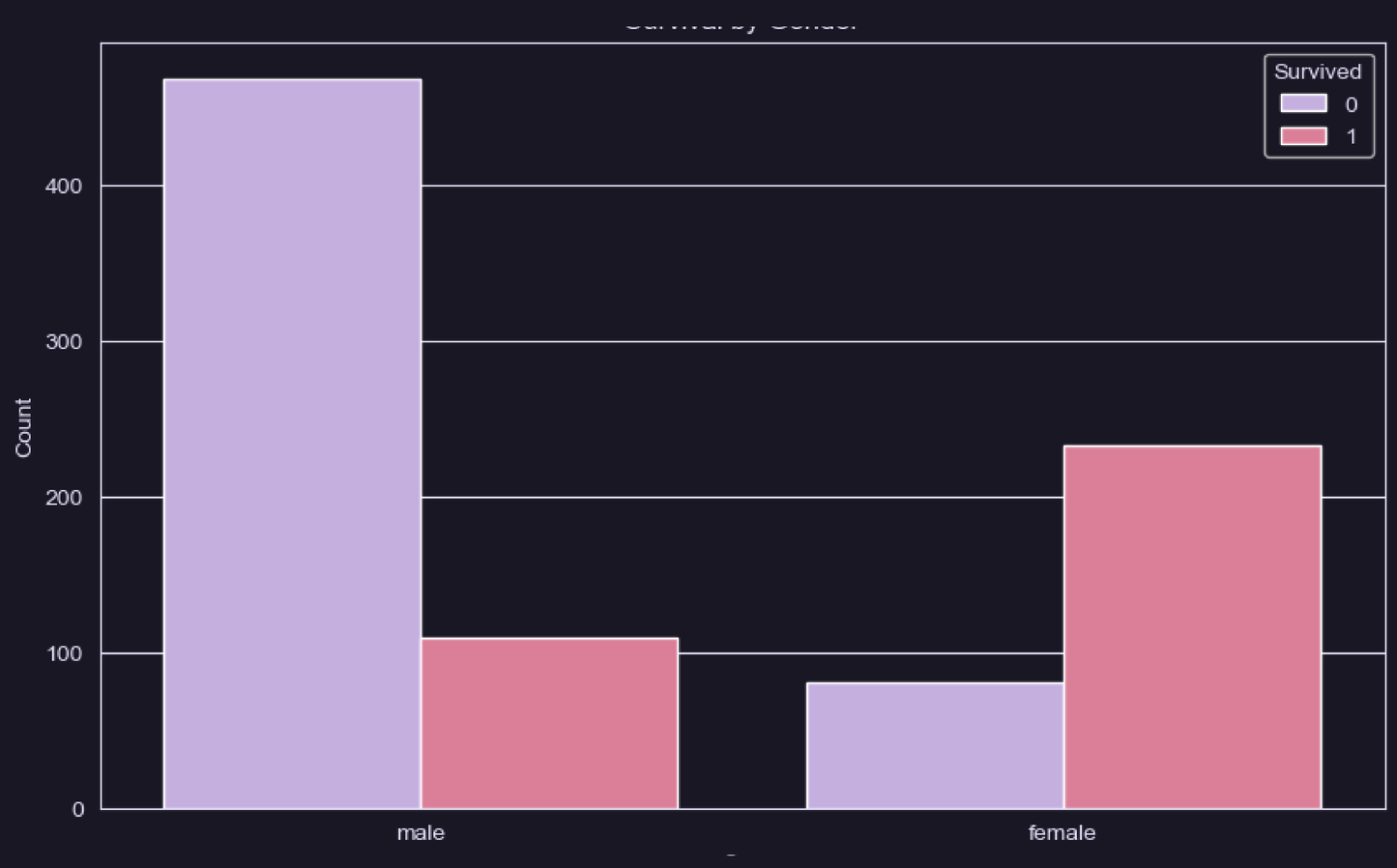

The bar chart reveals that males traveling alone had the highest count of non-survivors, while females with small families had the highest count of survivors. Both males and females in medium and large family sizes saw lower counts overall, indicating a lower survival rate in these groups.




The bar chart shows that passengers with shorter names (12-22 characters) had a higher count of non-survivors, while those with longer names (28-82 characters) had a higher count of survivors this is because people of importance and wealth had longer names

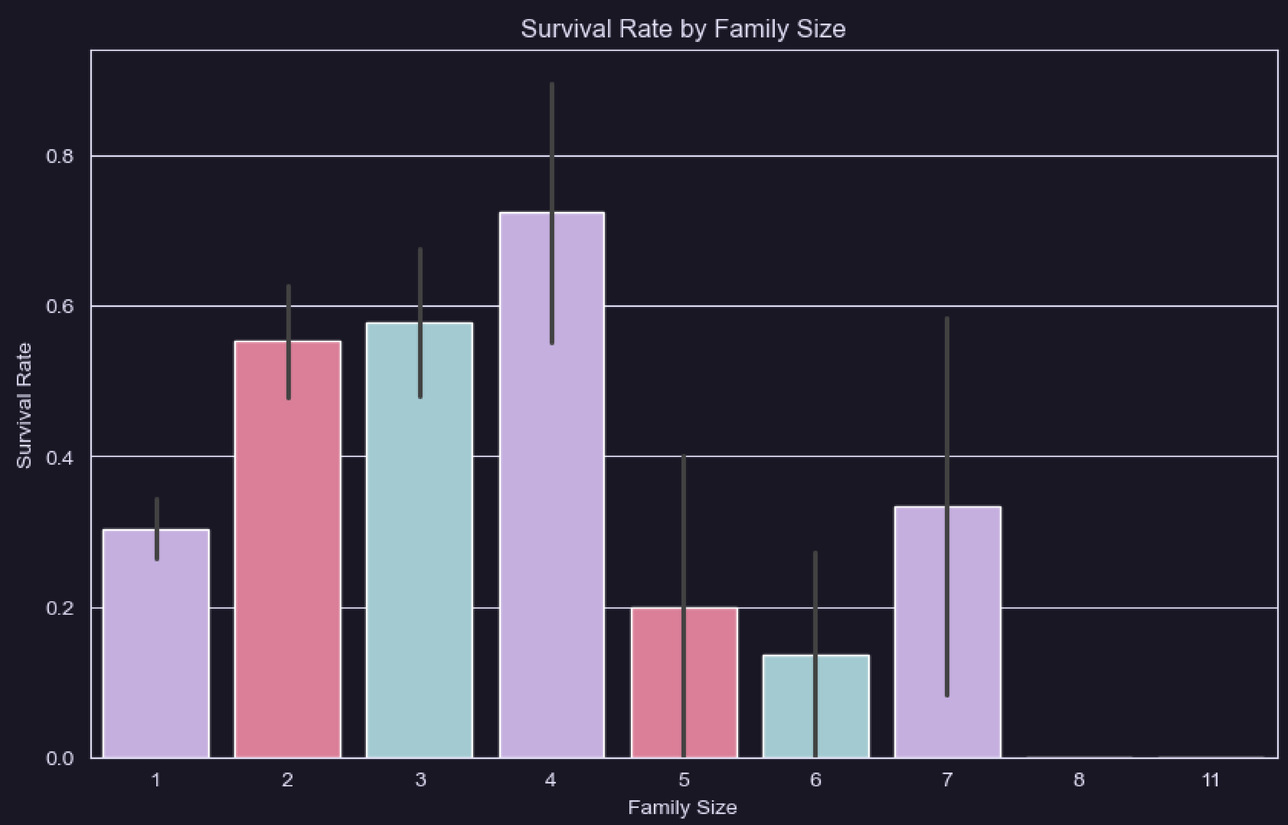
Passengers who paid higher fares had a better chance of survival. Those in the lowest fare groups had survival rates below 25%, while the highest fare group exceeded 60%. The data highlights a positive correlation between fare amount and survival rate.

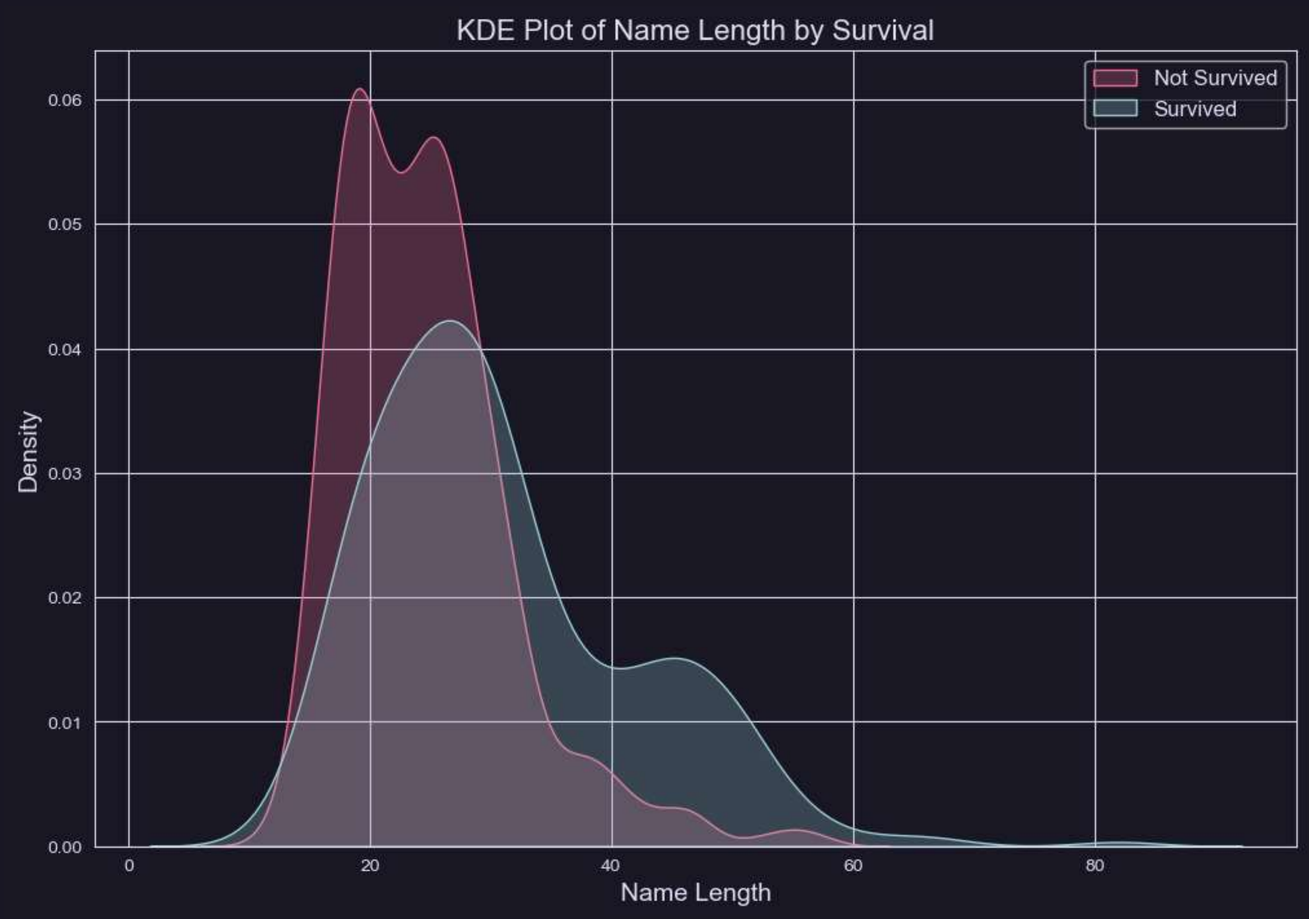



While a significantly higher number of males did not survive, females had a considerably higher survival rate. This emphasizes the "women and children first" protocol followed during the evacuation. The graph clearly shows the prioritization of women over men during the rescue efforts.

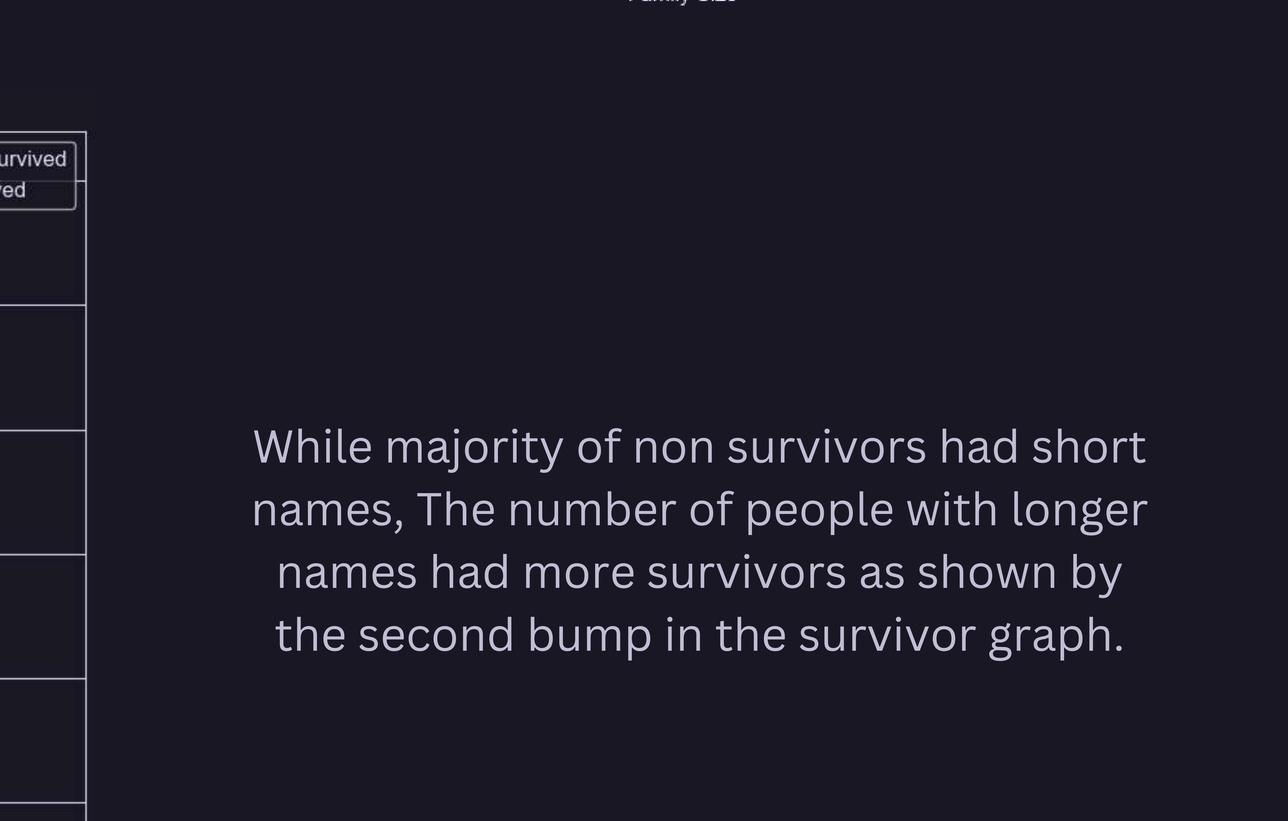
Passengers with smaller families (2-4 members) had the highest chances of survival, with family sizes of 4 showing the highest survival rate of around 75%. Meanwhile, larger families (5-6 members) faced significantly lower survival rates, all below 20%. Interestingly, family sizes of 1 and 7 showed moderate survival rates around 30-40%.

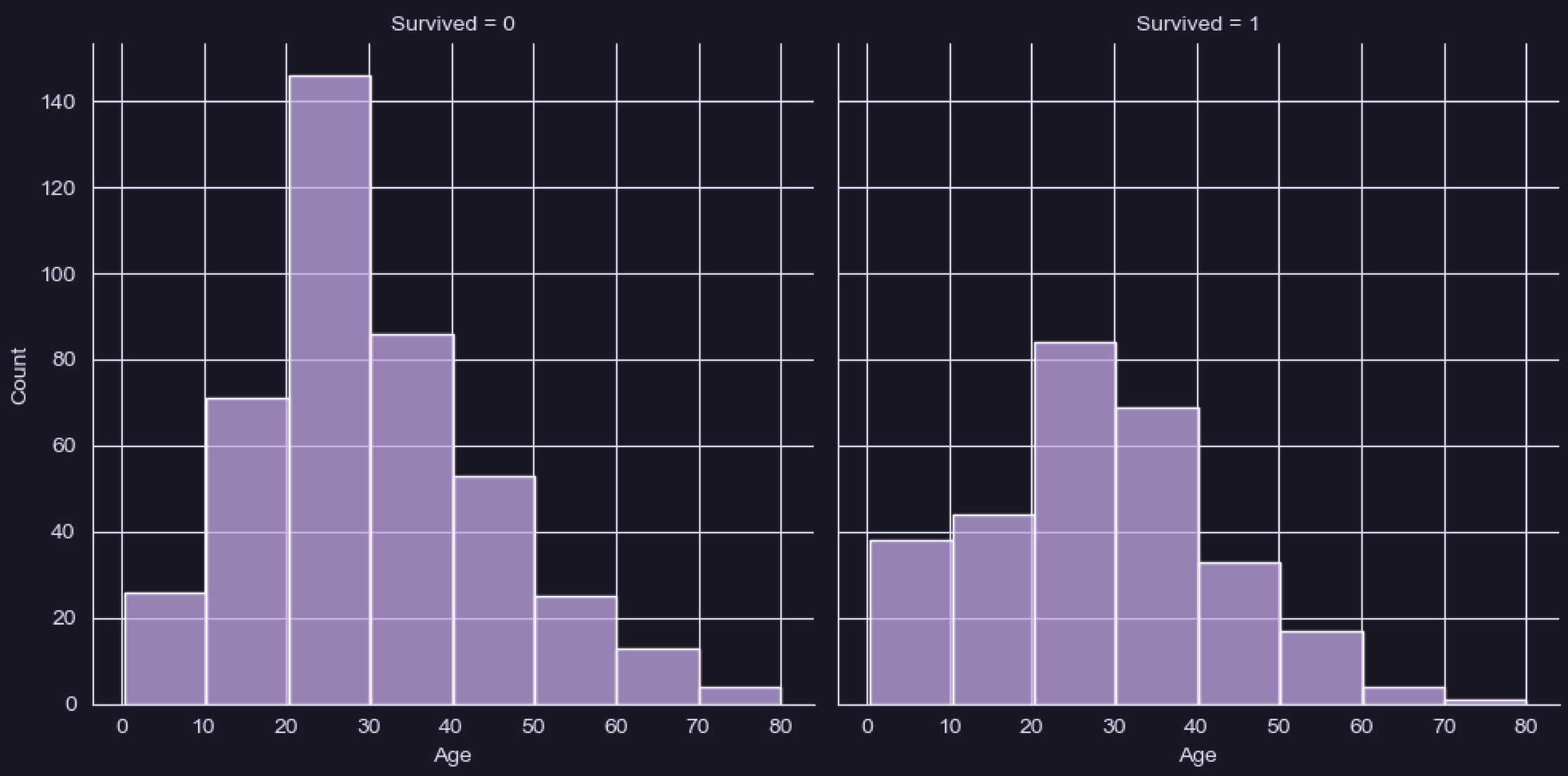


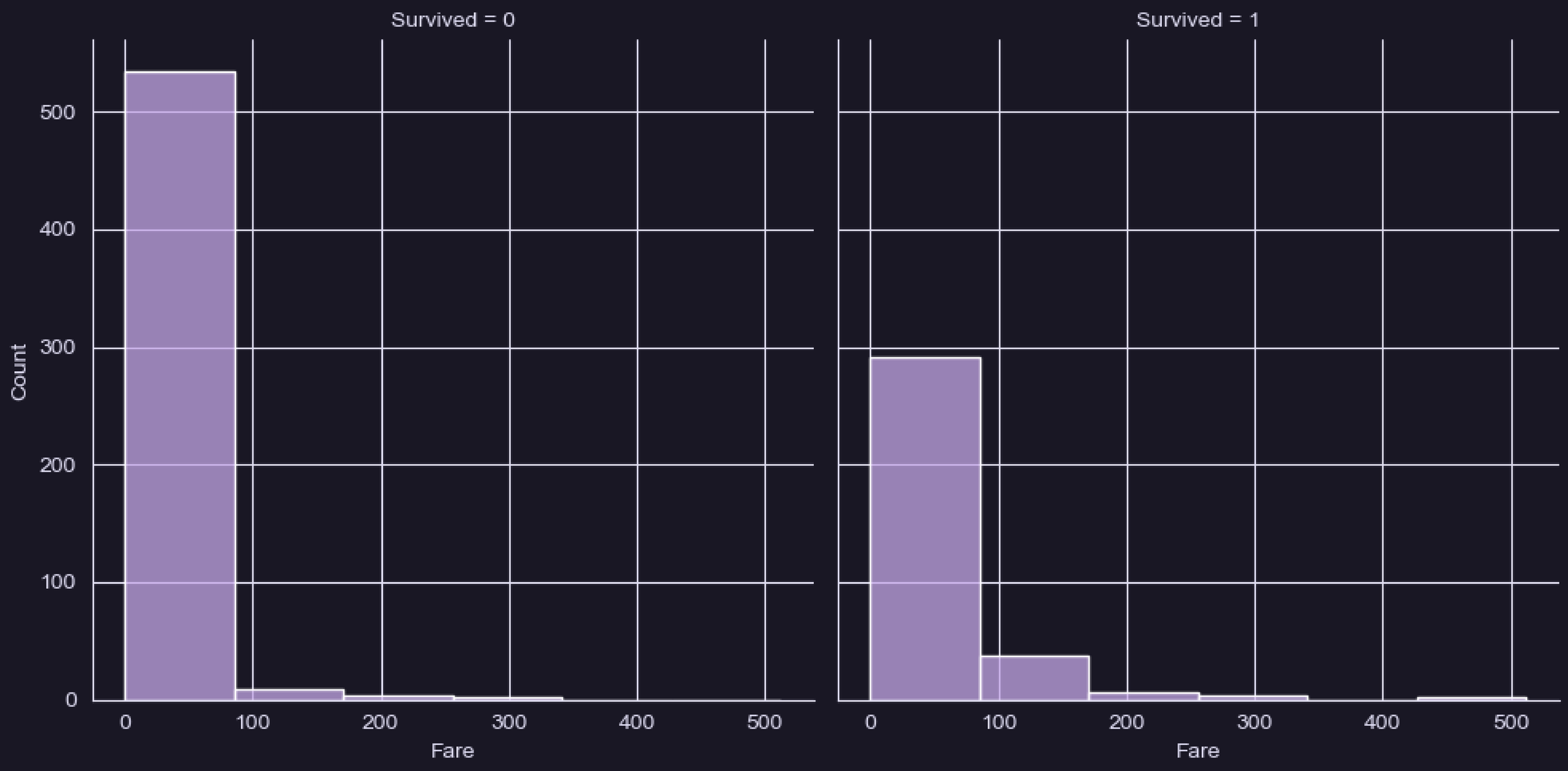

While majority of non survivors had short names, The number of people with longer names had more survivors as shown by the second bump in the survivor graph.

The highest survival rate is observed for passengers with ticket number count 3, while the lowest survival rate is seen for those with ticket number count 1. This suggests that passengers with certain ticket number counts had a better chance of surviving the disaster.






The two histograms display the age distributions of Titanic passengers who survived and those who didn’t. The left histogram, labeled "Survived = 0," shows that most non-survivors were in their 20s and 30s. On the right, "Survived = 1" indicates that survivors were also mainly in their 20s and 30s but with a more even spread across other age groups.




The two histograms show the fare distributions for Titanic passengers. Most passengers, both those who survived and those who didn’t, paid fares below 100. Interestingly, there’s a slight increase in survival rate among those who paid higher fares.

Individuals who were alone had the highest number of non-survivors, while small families had a higher survival rate. Medium and large families faced lower survival rates, especially large families, which had no survivors.




---

# Key Takeaways

Analyzing the survival data from the Titanic has unveiled some compelling patterns and social truths. The evidence highlights a stark gender disparity, with women and children having much higher survival rates, reflecting the "women and children first" protocol. Age played a significant role, with younger passengers, particularly children, having better survival chances. Economic status influenced outcomes, as passengers with higher fares were more likely to survive, pointing to socio-economic inequalities. Smaller families had higher survival rates, indicating that larger family groups faced more challenges during the evacuation. Additionally, embarkation points influenced survival chances, with passengers boarding from Cherbourg faring better than those from Southampton or Queenstown.

github.com/sarvamm

instagram.com/sarvamm_

linkedin.com/in/sarvamm