

An Analytical Study of Survival Factors Aboard the RMS Titanic

By Antanya Lindsay

1. Introduction

The sinking of the RMS Titanic on April 15, 1912, remains one of the most significant maritime disasters in recorded history, claiming over 1,500 lives out of approximately 2,200 passengers and crew. Beyond its historical significance, the disaster provides an opportunity to examine human behavior under extreme conditions, particularly how survival was influenced by factors such as socioeconomic status, gender, age, family structure, and embarkation patterns.

This study aims to uncover the primary factors contributing to survival outcomes aboard the Titanic. By analyzing passenger demographics, ticket information, and survival rates, we can discern patterns and draw data-driven conclusions. The analysis uses Kaggle's Titanic dataset, consisting of 891 passenger records with attributes such as survival status, gender, age, class, fare, and port of embarkation. The objectives of this study are as follows:

1. Determine the significance of **gender** and **class** in influencing survival rates.
2. Analyze the impact of **age** and **family size** on survival outcomes.
3. Investigate the relationship between **ticket fare** and survival.
4. Examine how the **port of embarkation** influenced survival patterns.
5. Reflect on the implications of missing data, particularly in **Age** and **Cabin**.

The findings not only provide insight into the Titanic disaster but also demonstrate the application of data analysis techniques to uncover meaningful patterns.

2. Data Preprocessing and Cleaning

To ensure the integrity and reliability of the analysis, the dataset underwent rigorous preprocessing:

1. Handling Missing Values:

- The **Age** column, with 177 missing values, was imputed using the **median age** (28). This approach was chosen to mitigate the influence of outliers while preserving the distribution.
- The **Embarked** column, with two missing entries, was filled with the most frequent port, **Southampton (S)**, as it was the primary departure location.
- The **Cabin** column, which had 687 missing values (~77%), was dropped from the analysis due to its sparsity.

2. Feature Engineering:

- A new variable, **FamilySize**, was created to reflect the total number of family members aboard. It was calculated as:
$$\text{FamilySize} = \text{SibSp} + \text{Parch} + 1$$

$$\text{FamilySize} = \text{SibSp} + \text{Parch} + 1$$
 This feature allowed us to analyze the influence of family structure on survival.

3. Encoding Categorical Variables:

- The **Sex** column was encoded as a binary variable: Male = 0, Female = 1.
- The **Embarked** column was mapped numerically: C (Cherbourg) = 0, Q (Queenstown) = 1, S (Southampton) = 2.

4. Validation:

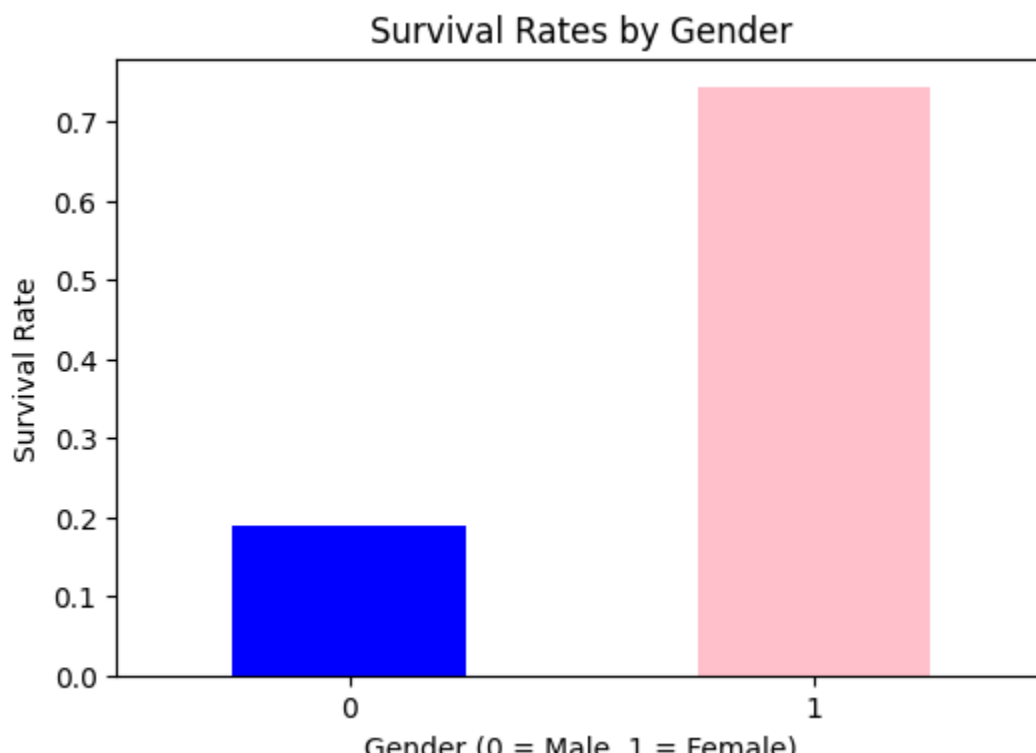
- The dataset was reviewed to confirm the absence of missing values and ensure numerical consistency for analysis.

The preprocessed dataset provided a robust foundation for exploratory data analysis and visualization.

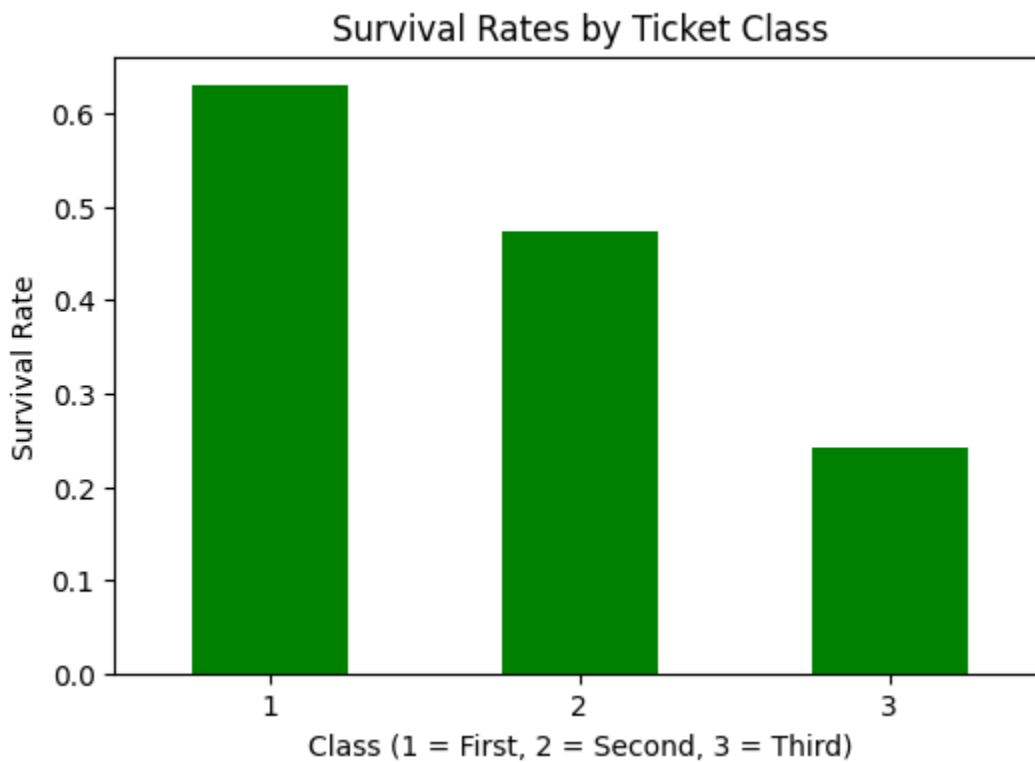
3. Exploratory Data Analysis and Results

3.1 Gender and Class: The Primary Determinants of Survival

A striking observation from the analysis was the impact of **gender** on survival. Female passengers had a survival rate of **74.2%**, compared to only **18.9%** for males. This discrepancy reflects the prioritization of "women and children first" during evacuation efforts.



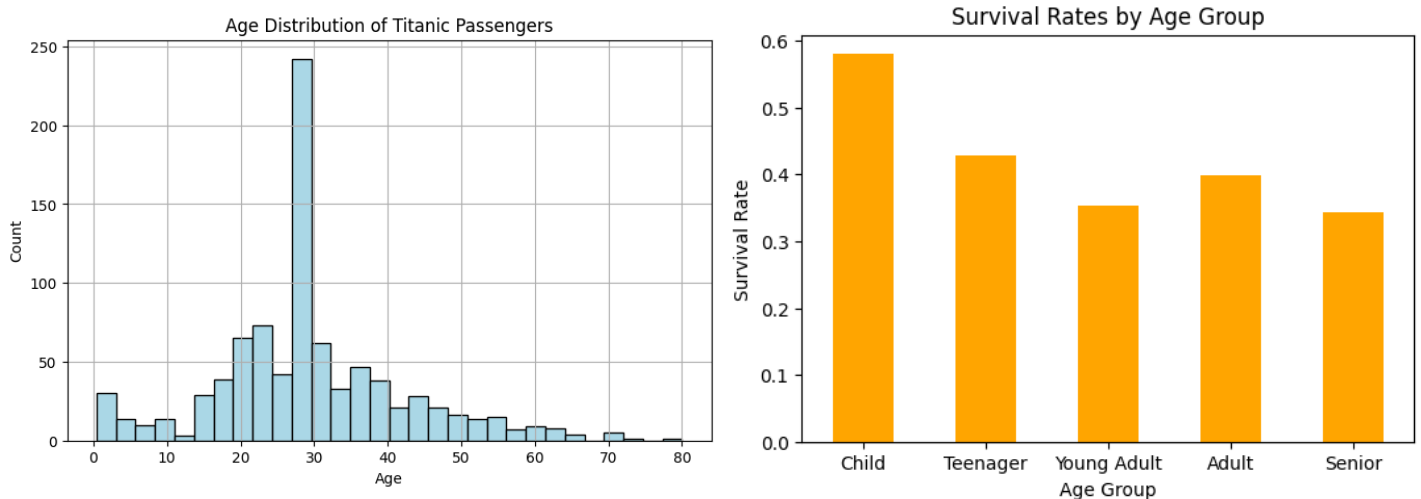
Similarly, **ticket class** strongly influenced survival rates. First-class passengers experienced a survival rate of **62.6%**, significantly higher than second-class (**47.3%**) and third-class (**24.2%**) passengers. The disparity highlights how socioeconomic privilege provided greater access to lifeboats and safer deck areas during the crisis.



3.2 Age: Prioritization of Children

Age emerged as another key determinant of survival. Passengers were grouped into age categories to identify trends:

Children had the highest survival rate (57.8%), underscoring their prioritization during evacuation. Conversely, survival rates declined for young adults and seniors, possibly due to their concentration in lower-class cabins.



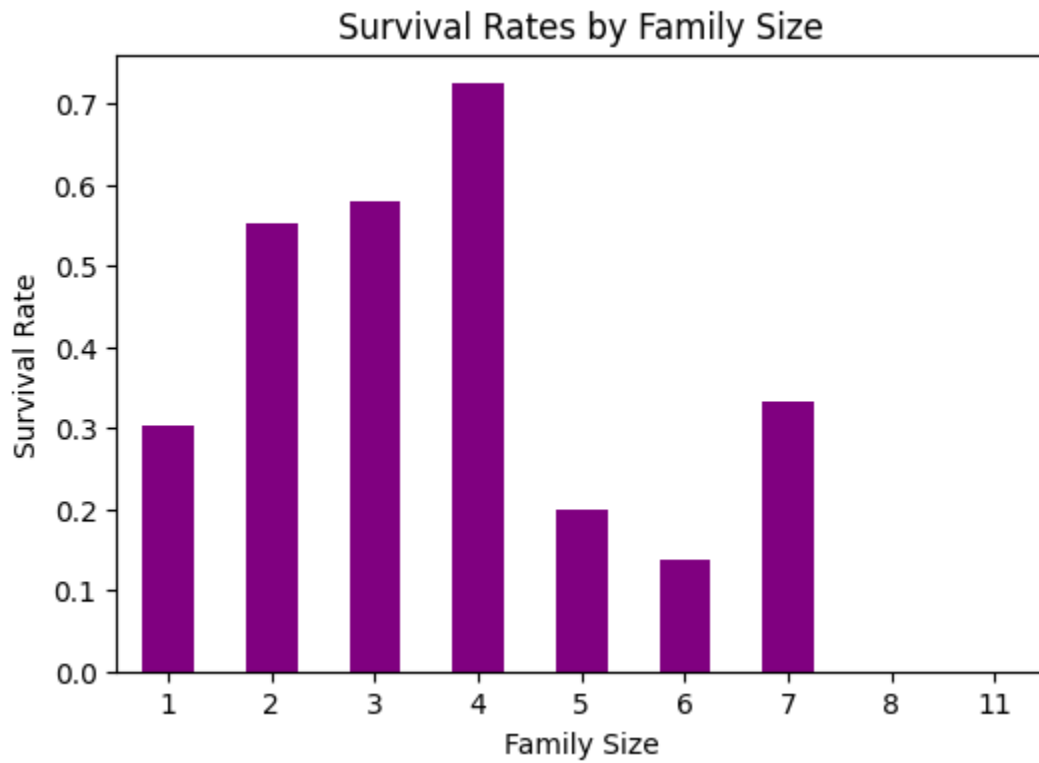
The age distribution revealed a peak around **30 years**, reflecting the predominance of young adults on board.

3.3 Family Size: The Advantage of Small Families

Family relationships also played a role in survival:

- Passengers with **small family sizes (2-4 members)** had the highest survival rates, reaching up to **72%** for families of 4.
- Individuals traveling **alone** had a survival rate of just **30.2%**.

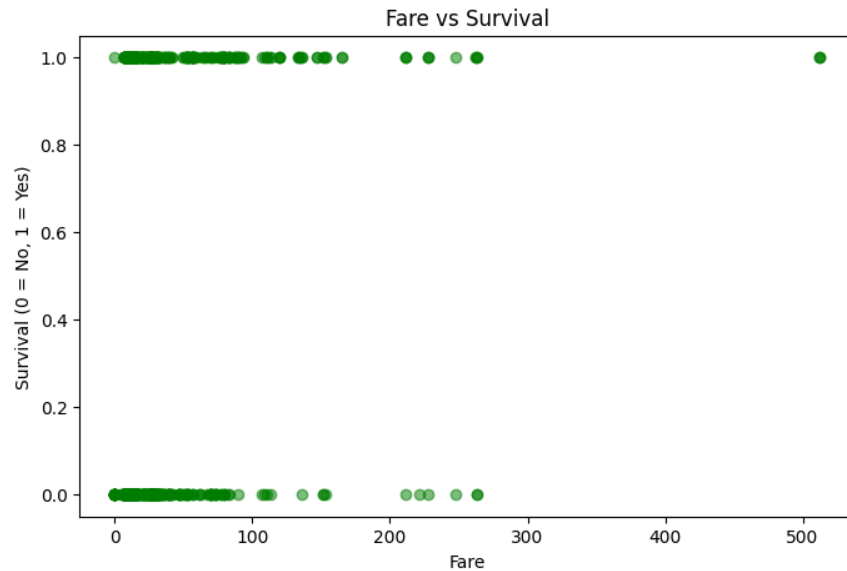
- Large families (5+ members) faced reduced survival rates, possibly due to challenges in coordinating evacuation efforts.



This suggests that small family units could provide mutual assistance without overwhelming limited resources.

3.4 Economic Factors: Fare and Survival

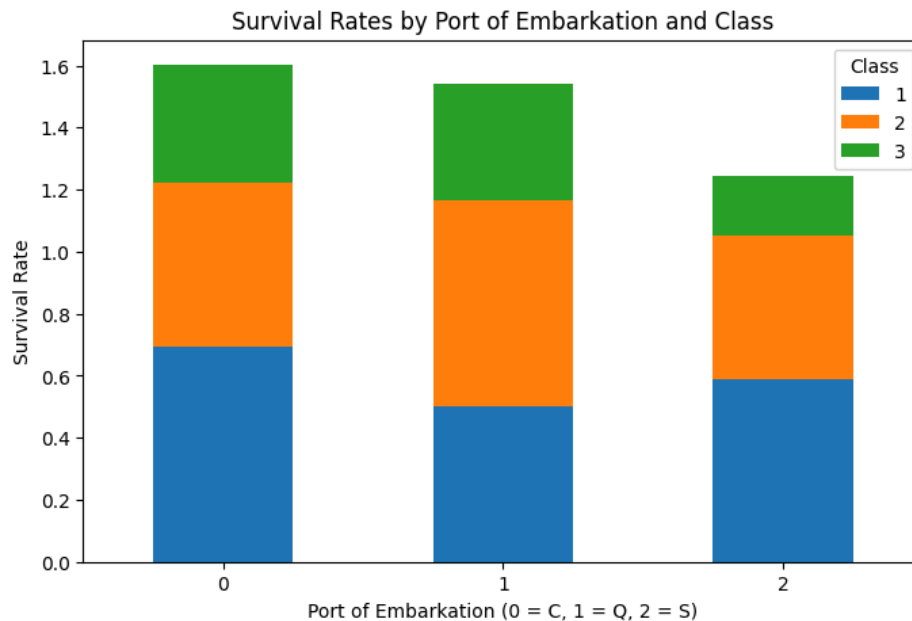
The relationship between ticket **fare** and survival was positive, with higher fares correlating with better outcomes. This trend aligns with the higher survival rates observed in first-class passengers, who paid significantly more for their tickets.



3.5 Port of Embarkation

The **port of embarkation** influenced survival patterns:

- Passengers embarking from **Cherbourg (C)** had the highest survival rates, likely due to a larger proportion of first-class passengers.
- Southampton (S) and Queenstown (Q) passengers had lower survival rates, reflecting the concentration of second- and third-class ticket holders.

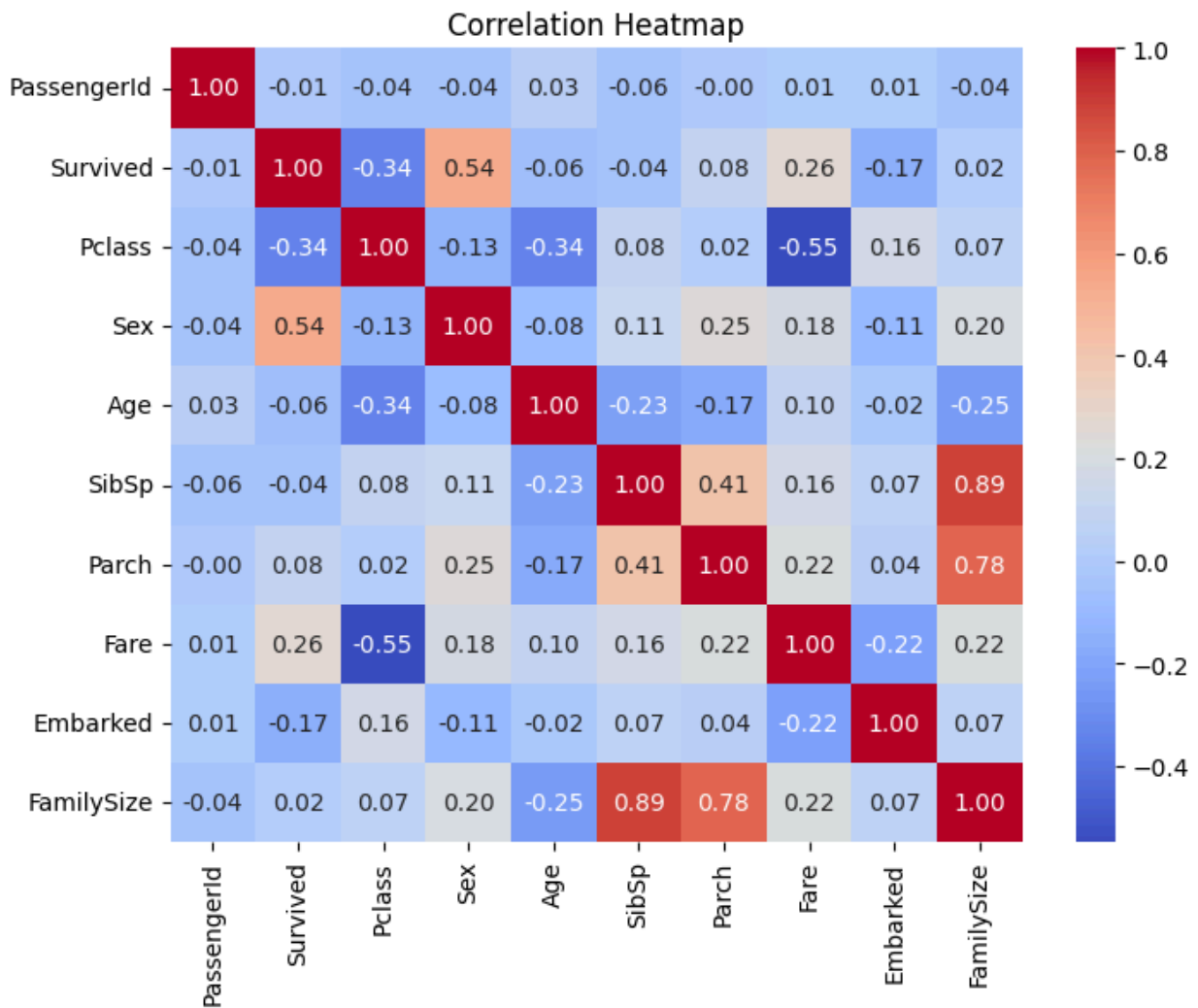


3.6 Correlation Analysis

The correlation matrix provided a holistic view of relationships between features. Key insights include:

- A **negative correlation** between **Pclass** and **Survived** (-0.34): Lower classes were less likely to survive.

- A **positive correlation** between **Fare** and **Survived** (0.26): Higher fares increased survival chances.
- Weak correlations between survival and family size (**SibSp**, **Parch**).



6. Conclusion

The analysis of Titanic passenger data offers a compelling glimpse into the factors that dictated survival during one of history's most devastating maritime tragedies. Through methodical data exploration and visualization, we uncovered patterns that align with societal norms, economic realities, and human behavior under crisis.

Gender and **socioeconomic class** emerged as the most significant determinants of survival.

Female passengers, prioritized during evacuation efforts, had survival rates nearly four times higher than their male counterparts. Similarly, passengers in **first class**—positioned closer to safety and possessing greater resources—were far more likely to survive compared to those in third class. These findings underscore how deeply embedded systems of privilege and protection shaped the outcomes of the disaster.

Demographically, **children** were prioritized over adults and seniors, reflecting the principle of safeguarding the vulnerable. Family dynamics also played a role—small families had a survival advantage, likely due to their ability to coordinate efforts without overwhelming limited resources. In contrast, individuals traveling alone and large family groups faced greater challenges during evacuation. Economically, survival was directly tied to **fare price**, a stark reminder of how wealth often translates into opportunity, even in matters of life and death.

The influence of **port of embarkation** highlighted regional disparities, with passengers boarding from Cherbourg (C) faring better due to the concentration of first-class ticket holders.

Meanwhile, the exploration of correlations reinforced these observations, revealing significant relationships between class, fare, and survival outcomes.

This study goes beyond numbers and graphs; it tells a story of humanity under duress. It reveals how societal hierarchies, resource allocation, and interpersonal bonds shaped the fates of those aboard the Titanic. While grounded in a historical context, these findings resonate with contemporary discussions on privilege, access, and equity in crisis situations.

Ultimately, this project demonstrates the power of data analysis in uncovering hidden patterns and bringing clarity to complex narratives. By transforming raw data into meaningful insights, we not only better understand a pivotal event in history but also reinforce the importance of analytical skills in solving real-world problems.