

Survival Was Not an Accident

A Data Investigation into the Sinking of the Titanic



The Central Question: What Determined Who Survived?

On 15 April 1912, the RMS Titanic sank, resulting in the deaths of over 1,500 people. With limited lifeboats, survival was a brutal lottery. But was it truly random?

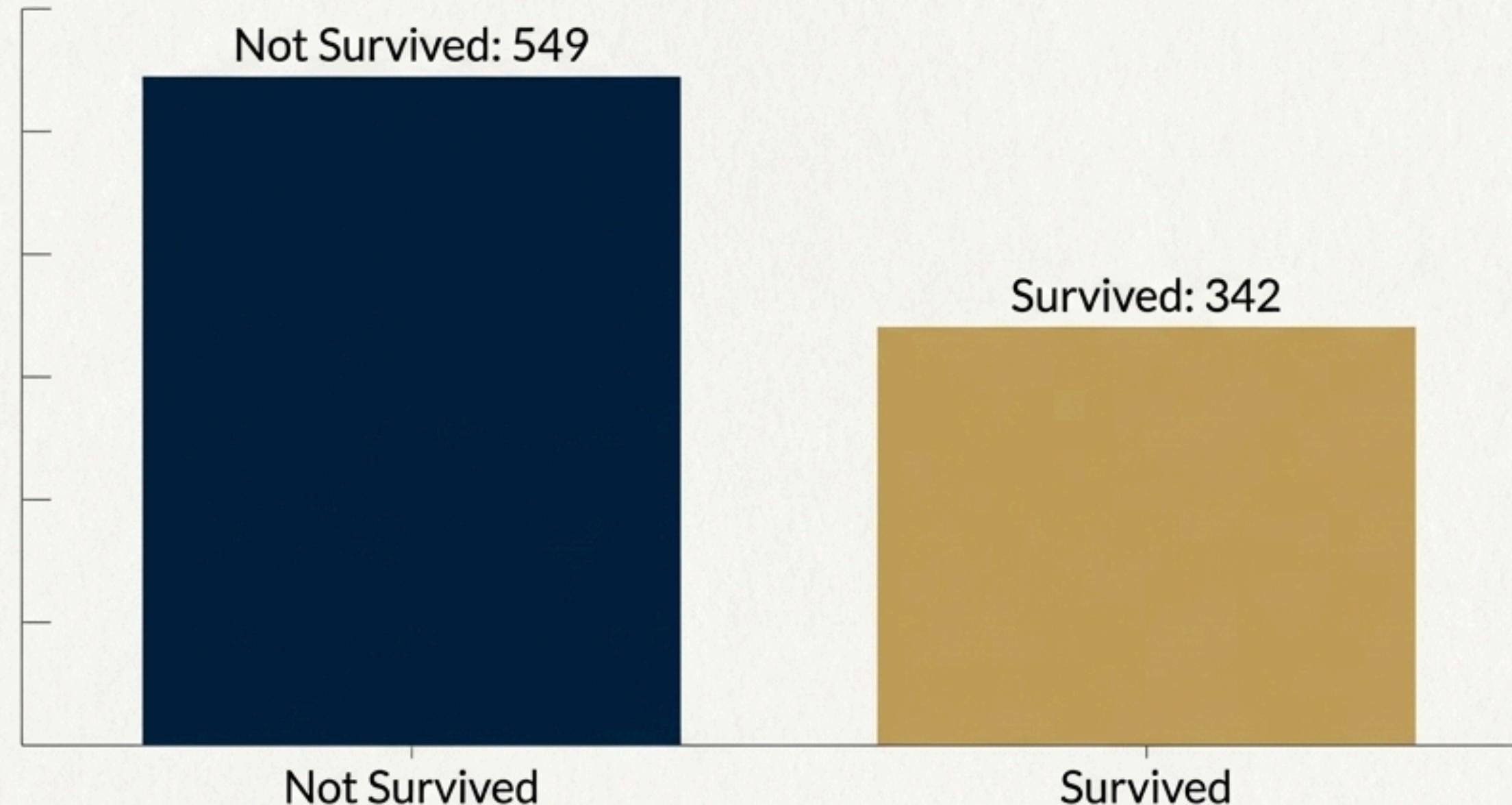
This analysis uses passenger data to uncover the hidden patterns of survival and loss, exploring the factors that separated those who lived from those who died.

891

Passengers in this Analysis

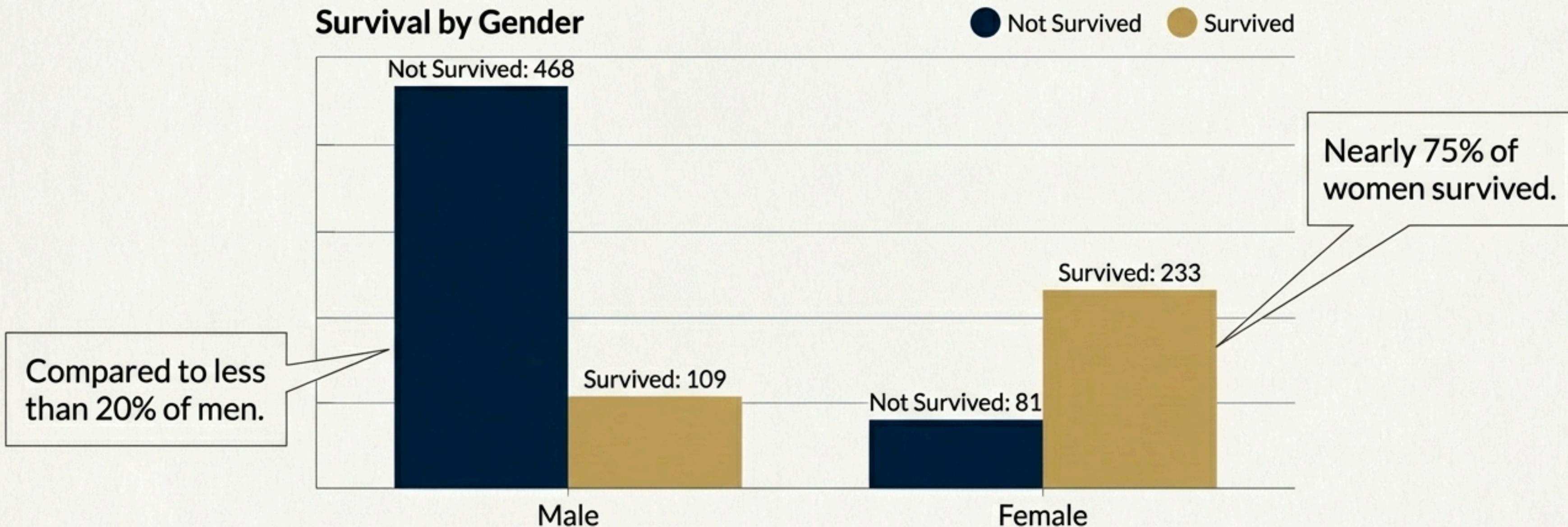
The Odds Were Stacked Against Everyone

The first look at the data reveals a stark truth: a majority of passengers did not survive the disaster. Out of 891 passengers in this dataset, 549 perished.



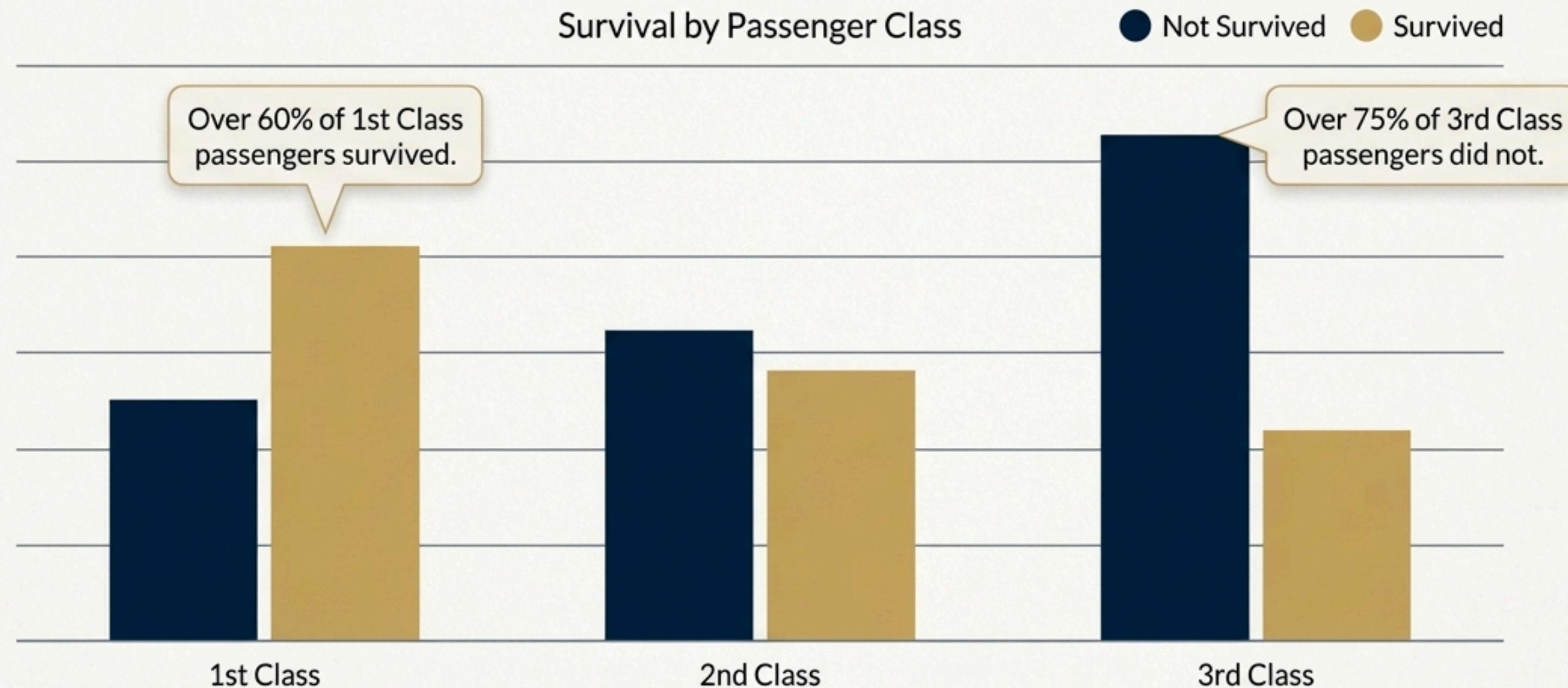
A Woman's Chance: Gender Was the Strongest Predictor of Survival

The long-held maritime tradition of “women and children first” was not just a saying; it was a stark reality. The data shows that female passengers had a dramatically higher chance of survival than males.



Wealth Bought Safety: First Class Held a Decisive Advantage

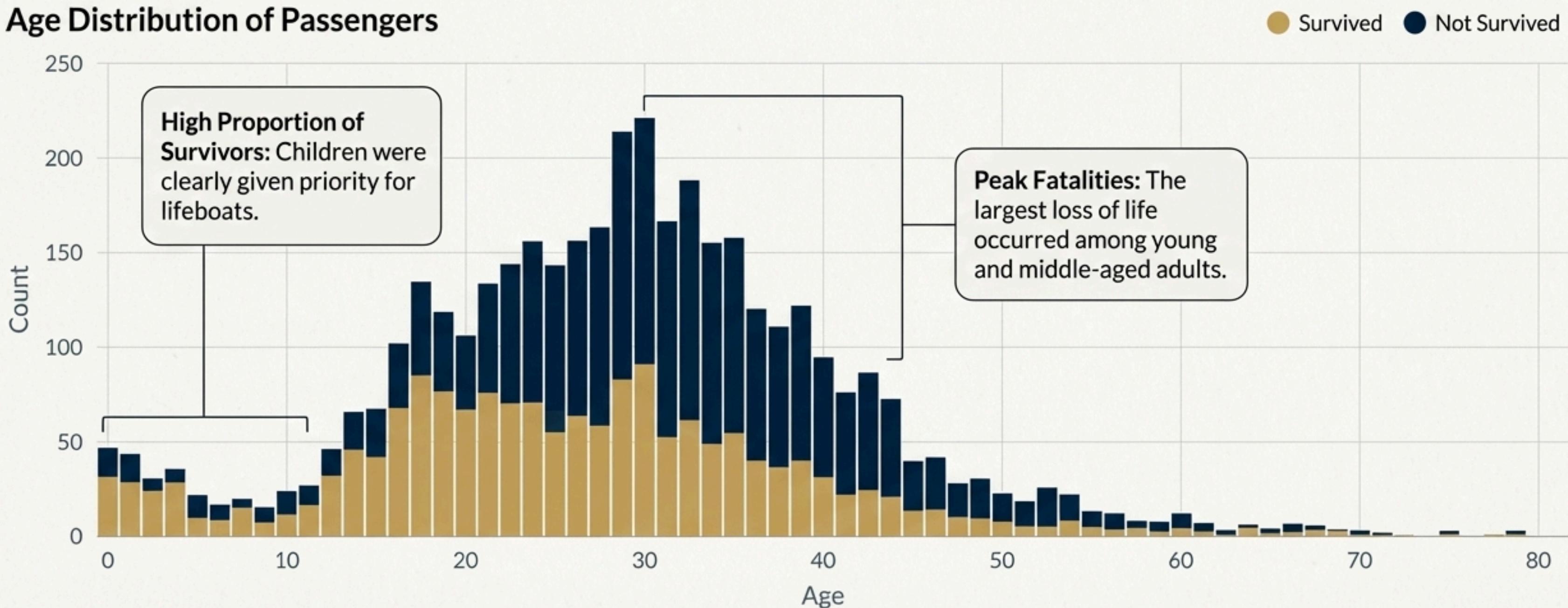
A passenger's ticket class was a strong determinant of their fate. First-class passengers had a significantly higher survival rate than those in second class, and an even greater advantage over those in third class, who faced the worst odds.



The Vulnerable and the Valued: Children Were Prioritised, While Adults in Their Prime Perished

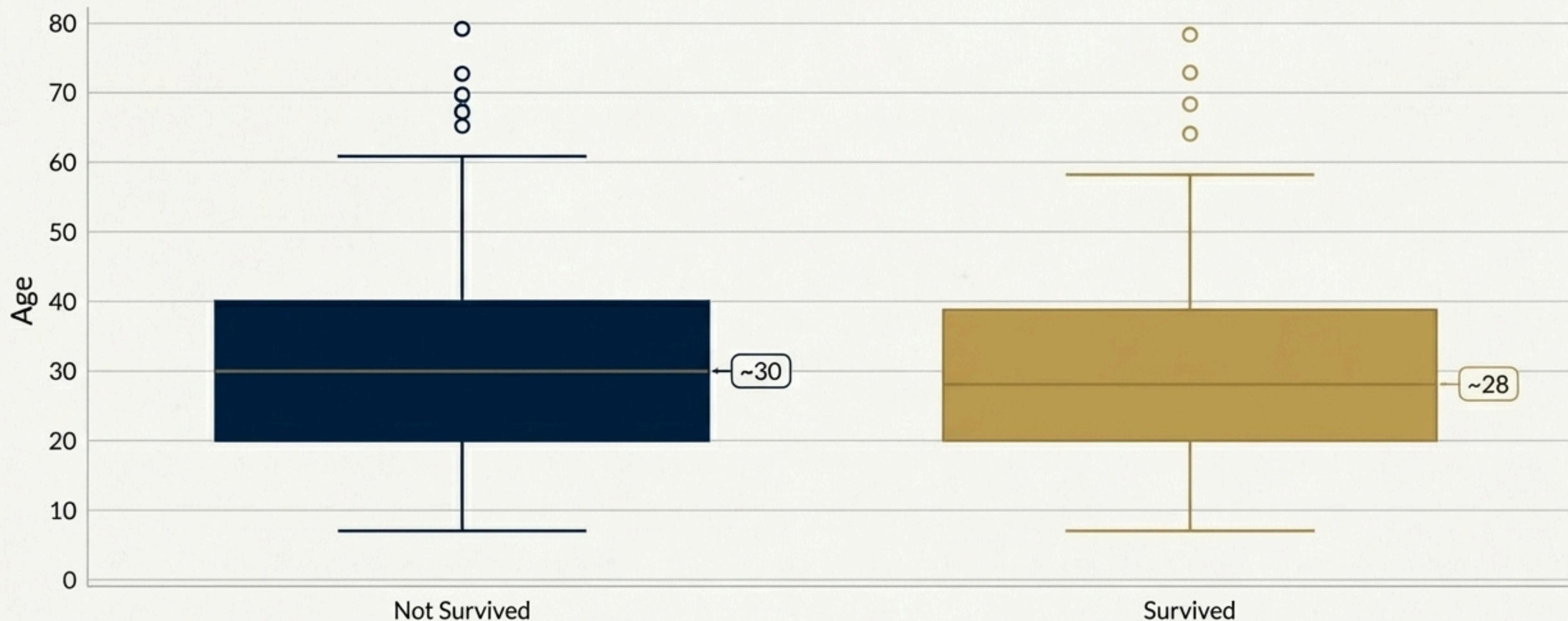
Age played a complex role. The data shows a higher survival rate among children. Conversely, the largest group of fatalities was **adults between the ages of 20 and 45**, the most common age group on the ship.

Age Distribution of Passengers



A Statistical View: Survivors Were, on Average, Younger

A box plot comparison provides a clear statistical summary. The median age of survivors is visibly lower than the median age of those who did not survive, confirming that youth was an advantage.



The Data-Driven Profile of a Titanic Survivor

Combining the key factors from our investigation, a clear picture emerges of the passenger with the highest probability of survival. This profile is not a guarantee, but a reflection of the powerful biases at play.



Female. Gender was the single most important factor.



First Class. Travelled with the highest-priced ticket.



Child. Or was an adult responsible for children.

Survival on the Titanic Was a Story of Class, Gender, and Age

The data is unequivocal. The fate of a passenger on the Titanic was not left to chance alone. It was profoundly influenced by the prevailing social structures of the era. A person's position in society—defined by their wealth, gender, and age—ultimately determined their odds of stepping onto a lifeboat.

The Foundation of the Analysis

A compelling data story rests on rigorous and transparent methodology. The following provides a brief overview of the data preparation steps taken to ensure the integrity and accuracy of this investigation.

Data Preparation in Brief

Handling Missing Data

Age

Missing values (177) were imputed using the median age of all passengers to maintain the overall distribution.



Embarked

The two missing port of embarkation values were filled with the mode (the most common port).



Cabin

The 'Cabin' column was dropped due to an excessive number of missing values (687), making it unreliable for analysis.



Feature Selection & Cleaning

Dropped Columns

'PassengerId', 'Name', and 'Ticket' were removed as they are unique identifiers with no predictive power for this analysis.



Encoding

Categorical variables ('Embarked', 'Sex') were converted into a numerical format suitable for analysis.



Labels

The 'Survived' column (0/1) was mapped to descriptive strings ('Not Survived'/'Survived') for clearer visualisations.

