

# Titanic Survival Probability Analysis Report

---

## I. Introduction

The sinking of the Titanic is one of history's most well-known maritime disasters. This report analyzes survival data from the Titanic dataset to understand the factors that influenced survival chances. The study examines variables such as **gender**, **passenger class**, **age**, and **fare**. Additionally, it compares empirical probabilities (calculated from actual data) with theoretical probabilities (based on assumptions of equal chance).

**Thesis Statement:** Survival on the Titanic was not random but strongly influenced by **gender**, **social class**, and **other factors**, which challenges the assumption of equal probability of survival for all passengers.

## II. Methodology

- **Dataset Used:** Titanic dataset from Kaggle (**Titanic-Dataset.csv**).
- **Data Cleaning:** Missing values in Age were replaced with the mean; missing values in Embarked were replaced with the most frequent port.
- **Analysis Tools:** Python (Pandas, Numpy, Matplotlib, Seaborn) was used for data exploration, visualization, and probability calculations.
- **Probability Approach:**
  - \* Empirical Probability: Calculated from actual passenger outcomes.
  - \* Theoretical Probability: Assumed equal survival chances across gender and passenger class.
- **Limitations:**
  - \* Dataset only covers ~891 passengers (not the full Titanic list).
  - \* Assumptions in theoretical probabilities may not reflect real-world conditions.

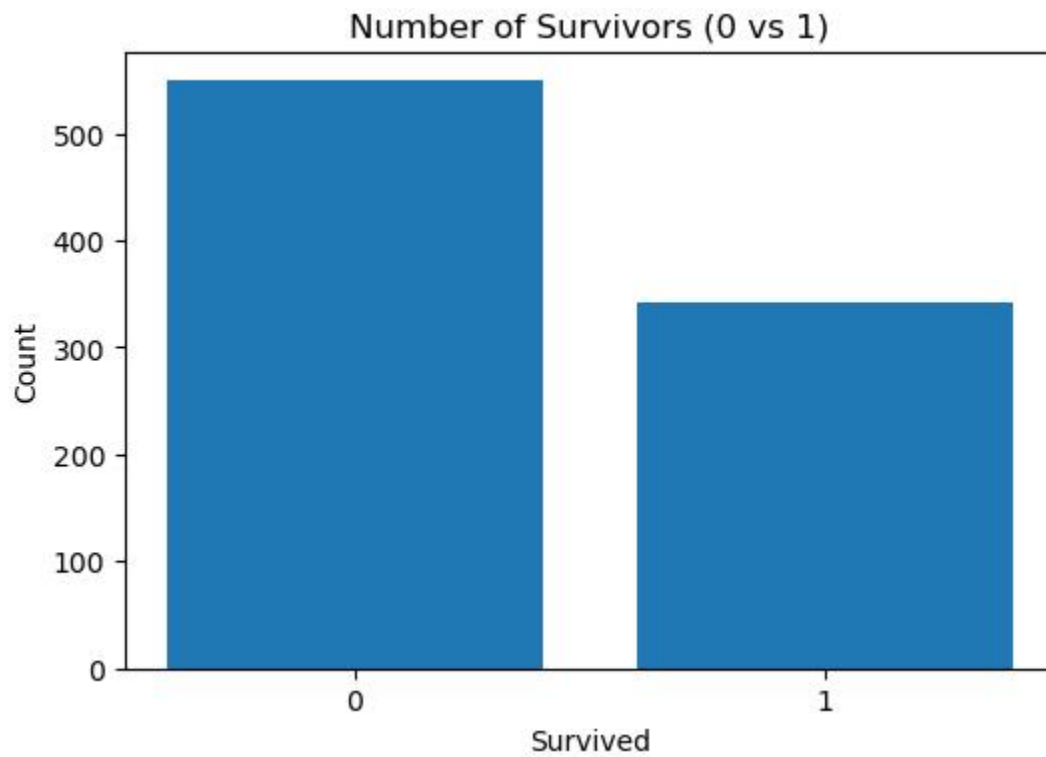
## III. Results

### Key Findings:

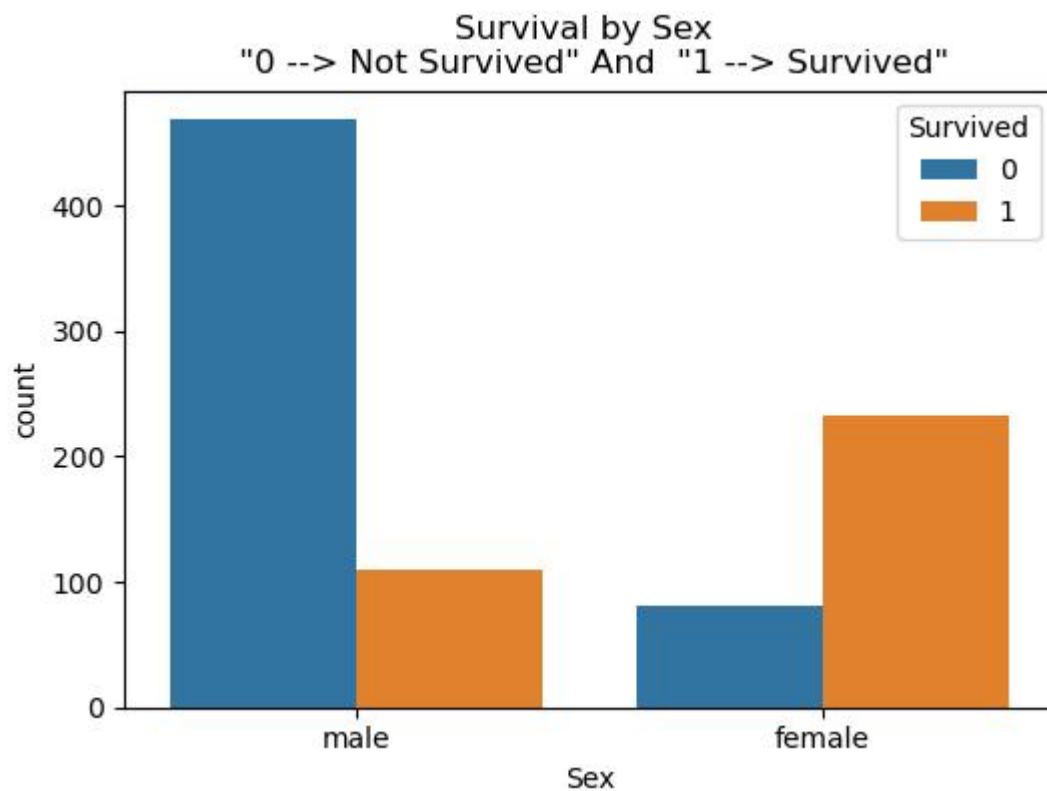
- **Survival Distribution:** About **38%** survived, while **62%** did not survive.
- **Gender and Survival:** Females had a much higher survival rate (~**74%**) compared to males (~**19%**).
- **Class and Survival:**
  - \* 1st Class → ~**63%** survival
  - \* 2nd Class → ~**47%** survival
  - \* 3rd Class → ~**24%** survival
- **Age Groups and Survival:** Children (0–12) had better chances compared to adults and seniors.
- **Fare and Survival:** Higher ticket fares correlated with higher survival probability.

### Visuals:

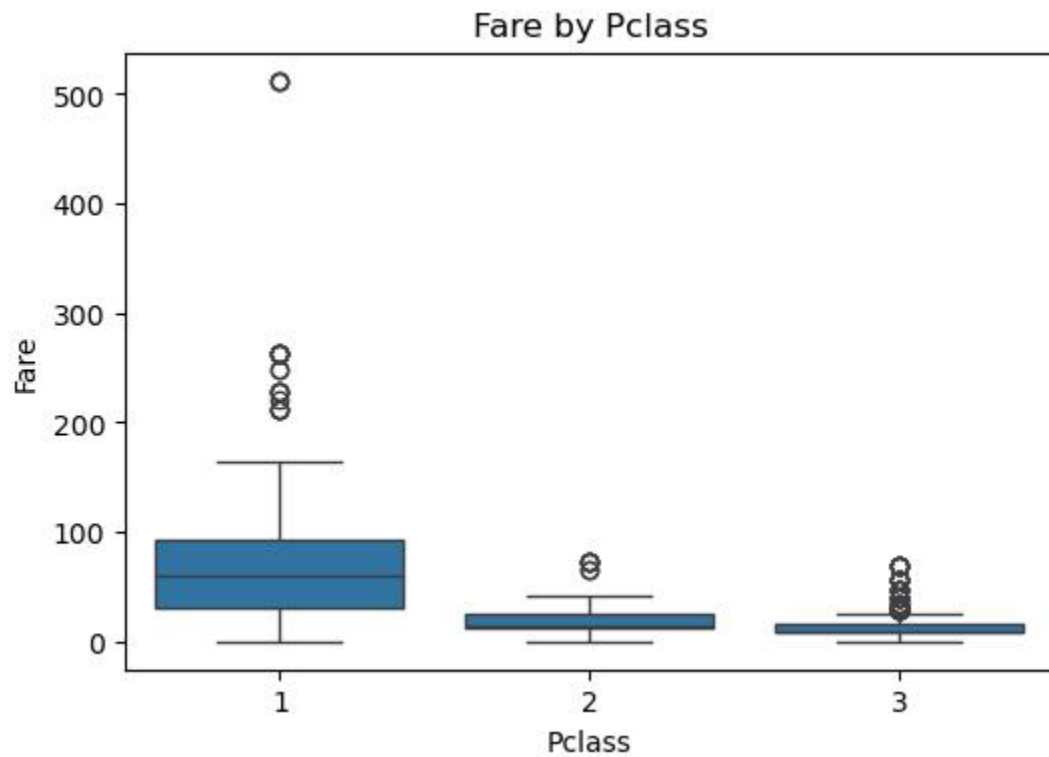
- Bar chart of survivors (Survived->1 vs Not Survived->0).



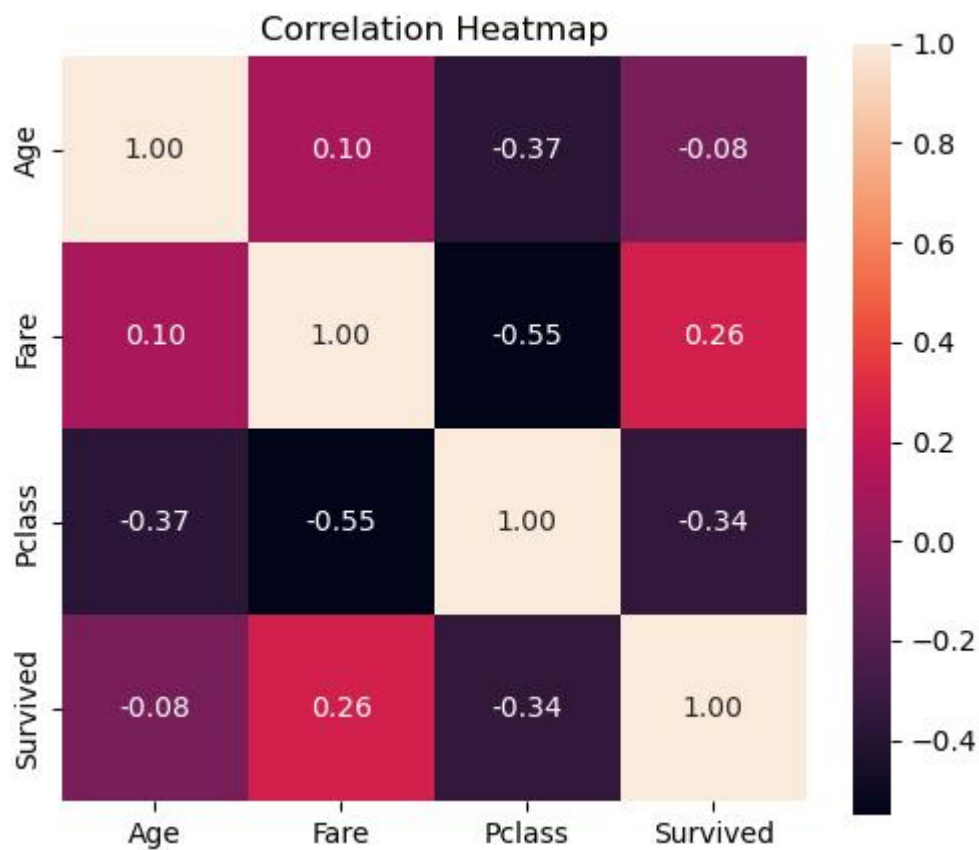
- Countplot of survival by gender.



- Boxplot of fare by passenger class.



- Heatmap showing correlation (Fare, Age and Pclass linked with survival).



## IV. Discussion

- **Women Survived More:** Data strongly supports the 'women and children first' evacuation policy.
- **Class Privilege:** Wealthier passengers (1st class) had better access to lifeboats and thus higher survival chances.
- **Age Influence:** Children had higher survival rates than adults, especially in lower classes.
- **Fare Connection:** Passengers who paid more (higher fare) were often in better cabins and had greater survival chances.
- **Empirical vs Theoretical Probability:**
  - \* Empirical: Shows survival was unequal and depended on gender/class.
  - \* Theoretical: Assumed equal chance (50% survival, 1/3 per class), which clearly doesn't align with reality.

## V. Conclusion

This analysis shows that Titanic survival was not random. Gender, passenger class, and **ticket price** significantly impacted chances of survival. **Women** and **children**, as well as **wealthier passengers**, had far better survival rates. The difference between empirical and theoretical probabilities highlights the importance of data-driven analysis over assumptions.

**Final Thought:** If survival were truly random (theoretical model), outcomes would be equal across groups. However, the Titanic tragedy demonstrates how social structures and inequalities can influence life-or-death outcomes.