**Titanic Analysis Report — Part E: Insights & Reporting**

**I. Introduction**

*This analysis investigates survival patterns among Titanic passengers using the Kaggle train.csv dataset. By applying exploratory data analysis (EDA), visualizations, and probability concepts, the report examines key demographic and socioeconomic factors—like gender, class, age, and fare—and how these influenced survival outcomes. The central argument: survival aboard the Titanic was not random but significantly impacted by variables such as sex, class, and age.*

**II. Methodology**

*1. Data Cleaning: We loaded the dataset using Pandas, displayed initial rows and summary statistics, handled missing values by imputing missing Age with the mean and missing Embarked with the mode.*

*2. Analysis Tools: Employed Pandas and NumPy for numerical and categorical summarization; used Matplotlib and Seaborn for data visualizations (bar charts, histograms, countplots, boxplots, violinplots, and heatmaps).*

*3. Probability Concepts: Computed empirical probabilities (e.g., overall survival rate, survival by gender/class/age group) and compared these with theoretical assumptions (e.g., 50% survival if random, equal across genders and classes).*

*4. Limitations: The dataset is a subset (891 passengers), may not include all Titanic passengers or crew, and imputation (mean/mode) can introduce bias.*

**III. Results**

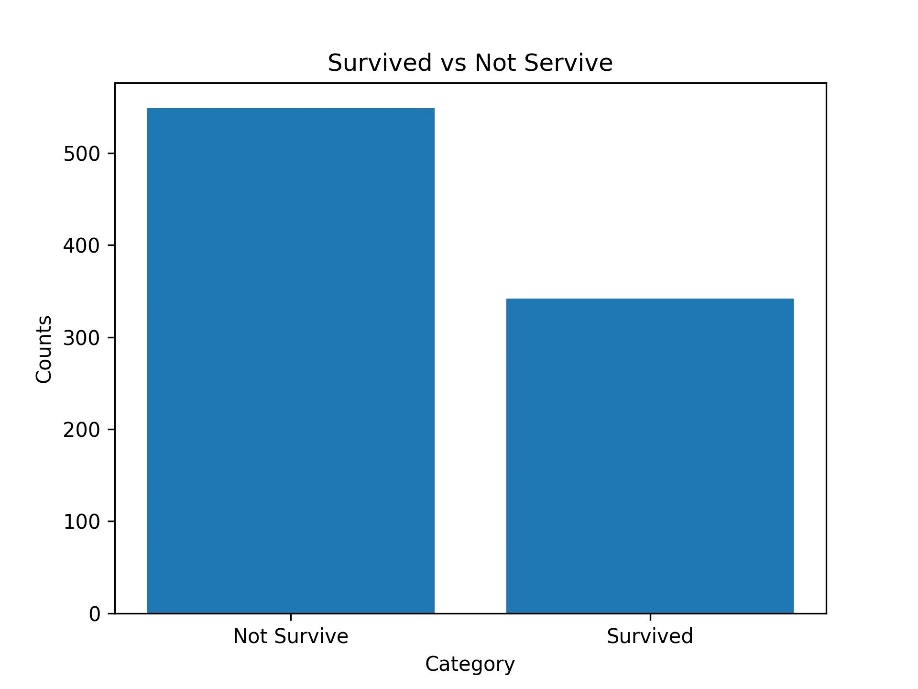
**1. Basic Statistics**

*Total passengers analyzed: 891*

*Average Age ≈ 29.7; Average Fare ≈ 32.2; Male ≈ 577; Female ≈ 314.*

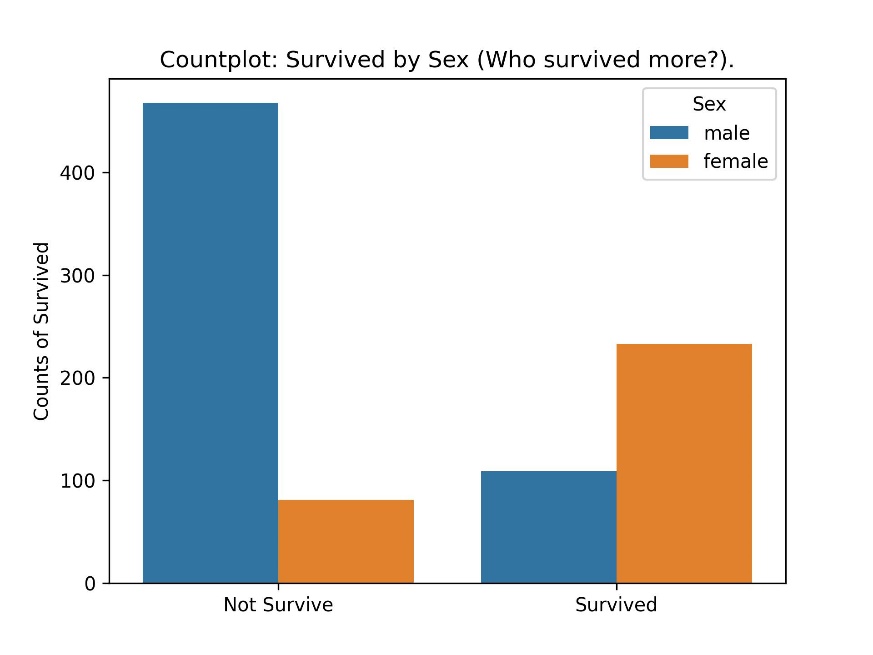
**2. Overall Survival**

*Survival count: e.g., 342 survived → empirical survival rate ≈ 38.4%.*



**3. Survival by Gender**

*Female survival ~ 74% vs. male ~ 20%*

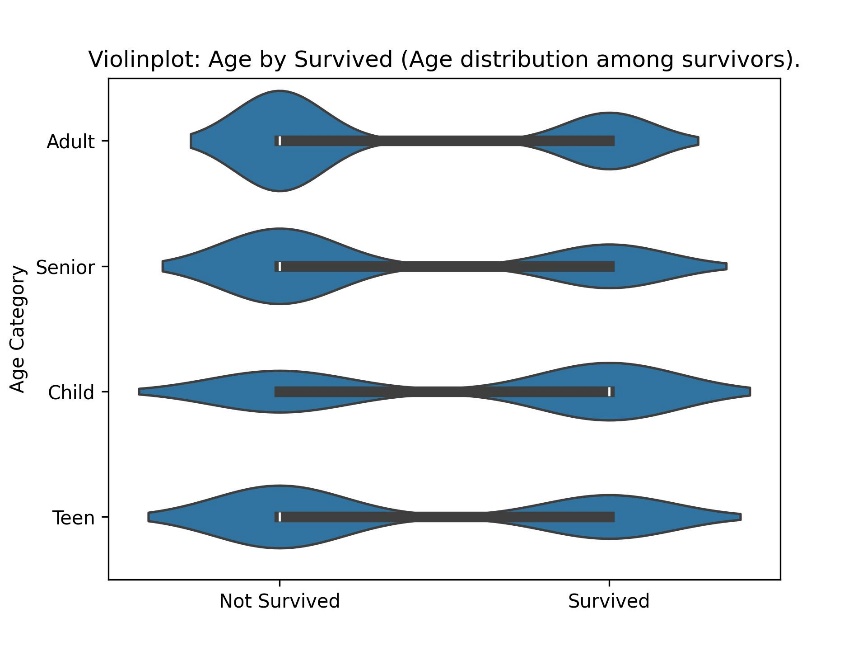


**4. Survival by Class**

*First-class survival notably higher than other classes (e.g., around 61–63%), while third-class significantly lower (around 23–24%)*

**5. Survival by Age Group**

*Children (especially ages 1–10) had highest survival (~59–83%), adults and especially older seniors had lower rates*



**6. Fare vs Survival**

*Higher fare correlated with higher survival chances, aligning with class-based advantages*

**7. Correlation Overview**

*Feature importance (correlation with survival): Sex (0.53) > Fare (0.26) > Age (~0.08)*

**8. Subgroup Highlights**

*First-class women had extremely high survival (e.g., 97%), whereas second-class men had lowest survival (8–14%)*

**IV. Discussion**

* *Gender Impact: Women had much higher survival rates than men, supporting the “women and children first” norm*
* *Class Influence: First-class passengers had significantly better survival outcomes than lower classes—likely due to proximity to lifeboats and access privileges.*
* *Age Effects: Children, especially young ones, were prioritized in rescue efforts; older passengers saw lower survival.*
* *Fare as Proxy: Higher fare belonged to wealthier passengers, correlated with higher survival probability.*
* *Empirical vs. Theoretical Probability: Real survival rates deviate greatly from a theoretical 50% random survival assumption—showing clear demographic biases.*
* *Intersectional Disadvantages: Second-class men suffered extreme mortality (~8–14% survival), highlighting compounded disadvantage due to both class and gender*

**V. Conclusion**

1. *Survival on the Titanic was far from random; gender, socioeconomic status, and age were critical determinants.*
2. *Women and children benefitted most, followed by wealthier passengers.*
3. *Real-world disasters like this illustrate how demographics and resources affect survival, urging the need for fair policies in safety and planning.*
4. *Empirical findings clearly contradict naive theoretical assumptions of equal probabilities.*

**5–7 Summary Bullet Points (for quick inclusion):**

* *Female survival rate (74%) far exceeded male (20%)*
* *First-class passengers survived at rates around 61–63%, while third-class fared worst (23–24%)*
* *Children (especially ages 1–10) had highest survival (up to ~83% for some age groups), older adults much lower*
* *Higher fare was linked to better survival odds—age and fare patterns mirrored class advantages*
* *Empirical survival rates diverged strongly from the theoretical 50% random model.*
* *Intersectional disadvantage: second-class men had one of the lowest survival rates (~8–14%)*
* *Overall conclusion: survival was influenced by sex, class, fare, and age—not random.*