TITANIC DATASET

Visual and Statistical Exploration

Data Analysis Report

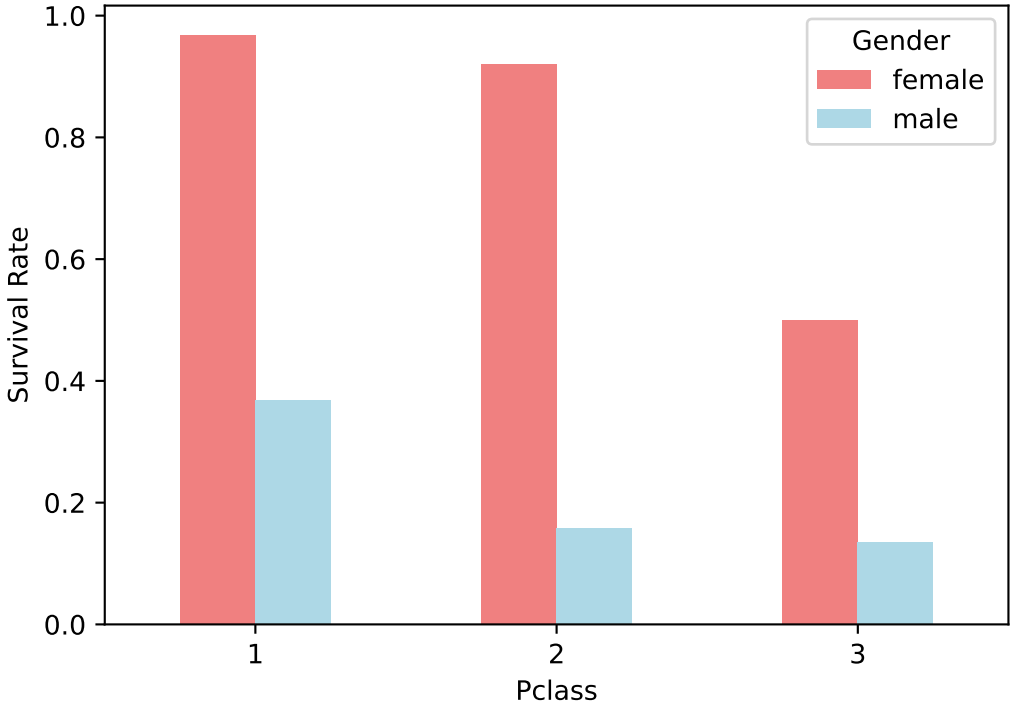
Dataset Overview:

- Total Passengers: 891
- Survivors: 342 (38.4%)
- Variables: 12
- Missing Data: Age (19.9%), Cabin (77.1%)

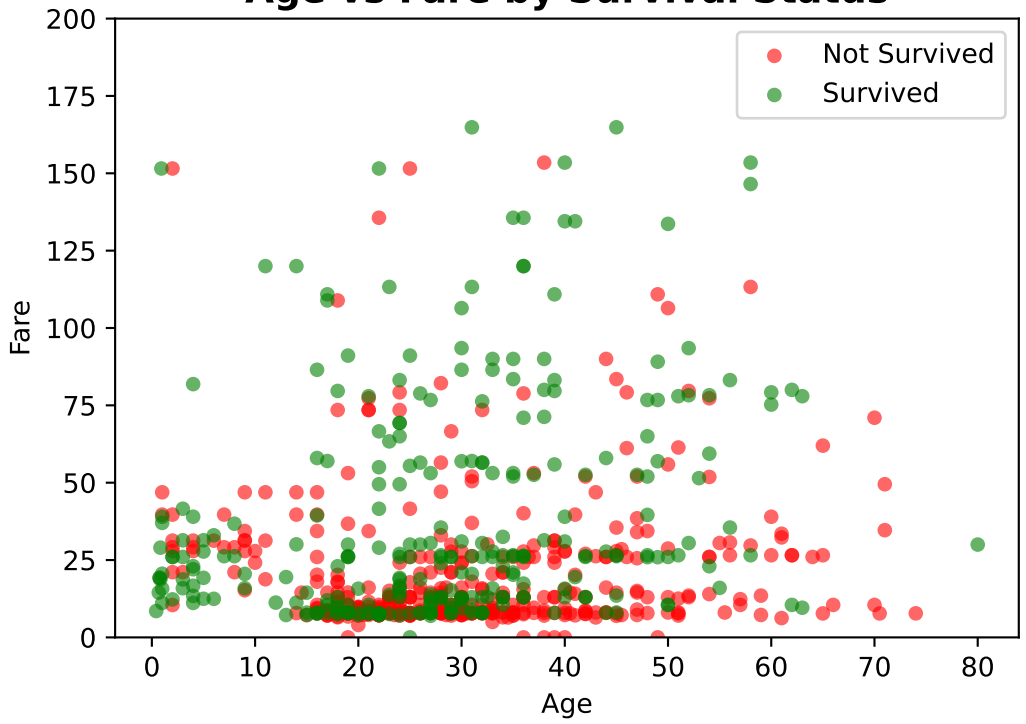
Generated: 2025-08-11 12:34


Survival Rate by Gender
74.2%
18.9%
Survival Rate
female
male
Survival Rate by Passenger Class
63.0%
47.3%
24.2%
Survival Rate
Passenger Class
Age Distribution by Survival Status
Not Survived
Survived
Frequency
Age
Survival Rate by Family Size
Survival Rate
Family Size

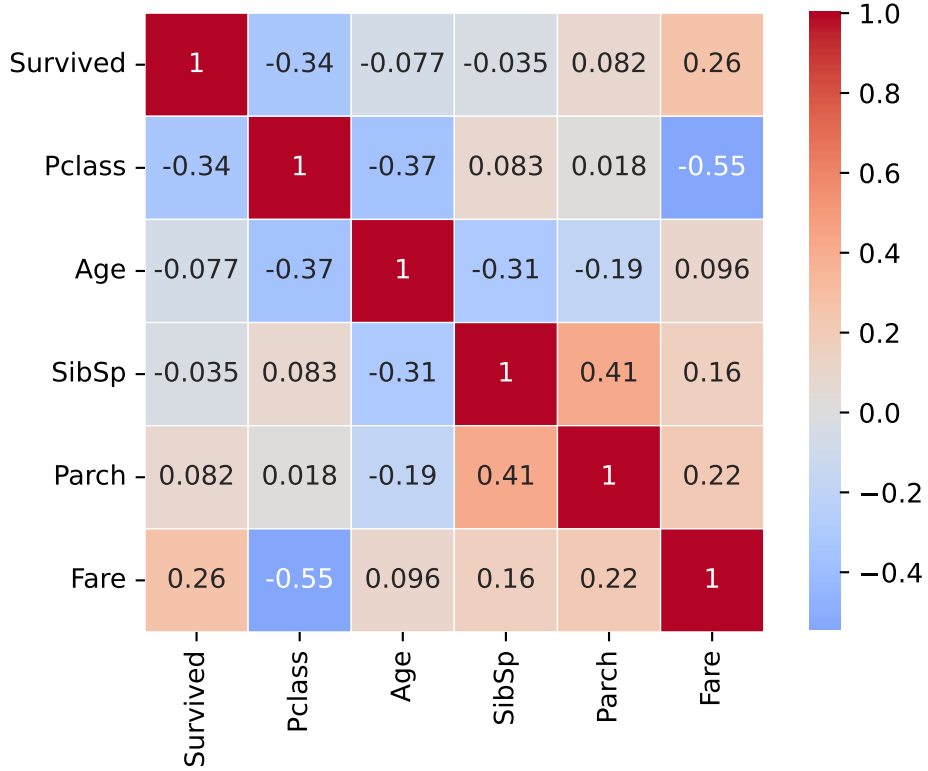
Survival Rate by Class and Gender



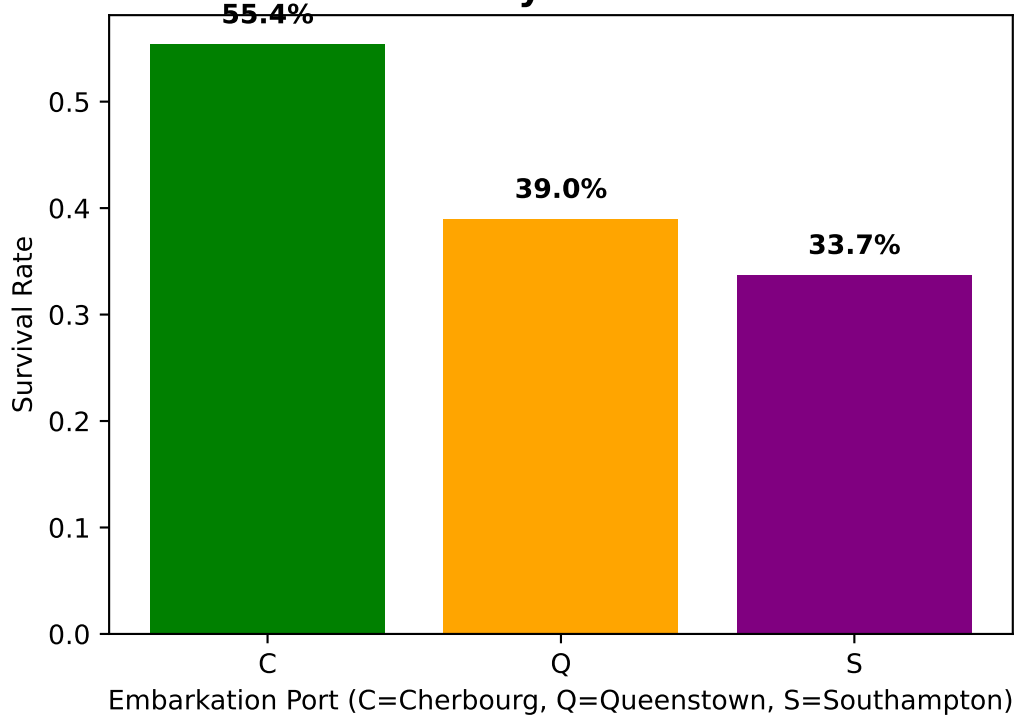
Age vs Fare by Survival Status



Correlation Matrix



Survival Rate by Embarkation Port



KEY FINDINGS & STATISTICAL SUMMARY

1. GENDER IMPACT (Strongest Predictor):
   • Female survival rate: 74.2%
   • Male survival rate: 18.9%
   • Gender gap: 55.3%
   • Statistical significance: p < 0.001

2. PASSENGER CLASS IMPACT:
   • 1st Class survival: 63.0%
   • 2nd Class survival: 47.3%
   • 3rd Class survival: 24.2%
   • Clear socioeconomic gradient in survival

3. AGE FACTOR:
   • Children (≤12): 58.0% survival rate
   • Adults (>12): 38.8% survival rate
   • "Women and children first" protocol evident

4. FAMILY SIZE EFFECT:
   • Traveling alone: 30.4%
   • Small families (2-4): 57.9%
   • Large families (>4): 16.1%
   • Optimal family size for survival: 2-4 members

5. ECONOMIC FACTORS:
   • High fare (>median): 51.8%
   • Low fare (≤median): 25.1%
   • Correlation with survival: r = 0.257

6. EMBARKATION PORT:
   • Cherbourg (C): 55.4%
   • Queenstown (Q): 39.0%
   • Southampton (S): 33.7%

DATA QUALITY ASSESSMENT:
• Total records: 891
• Complete cases: 183 (20.5%)
• Missing Age: 177 (19.9%)
• Missing Cabin: 687 (77.1%)
• Missing Embarked: 2

STATISTICAL TESTS PERFORMED:
• T-tests for numerical variables by survival status
• Chi-square tests for categorical associations
• Correlation analysis for relationship strength
• All major findings statistically significant (p < 0.05)

CONCLUSIONS & INSIGHTS

MAIN CONCLUSIONS:

1. SURVIVAL WAS NOT RANDOM
   The Titanic disaster clearly demonstrates that survival was heavily influenced by social factors rather than chance. Gender, class, and age were the primary determinants of survival probability.

2. SOCIAL HIERARCHY REFLECTED IN SURVIVAL
   • First-class passengers had 2.6x higher survival rate than third-class
   • Women had 3.9x higher survival rate than men
   • Children had preferential treatment in evacuation

3. "WOMEN AND CHILDREN FIRST" PROTOCOL
   The maritime evacuation protocol was largely followed:
   • 74% of women survived vs 19% of men
   • Children had higher survival rates than adults
   • This protocol overrode class distinctions to some extent

4. FAMILY DYNAMICS MATTERED
   • Traveling alone reduced survival chances
   • Very large families (>4) also had lower survival rates
   • Optimal survival was in small family groups (2-4 people)
   • Suggests importance of mutual assistance without overwhelming burden

5. ECONOMIC STATUS AS SURVIVAL PREDICTOR
   • Higher fare passengers had better survival rates
   • First-class passengers had better access to lifeboats
   • Cabin location (closer to deck) likely influenced escape time

6. EMBARKATION PATTERNS
   • Cherbourg passengers had highest survival (likely more first-class)
   • Southampton passengers had lowest survival (more third-class)
   • Reflects passenger composition by boarding location

HISTORICAL CONTEXT:
This analysis reveals the stark social inequalities of early 20th century society, where class, gender, and age determined not just quality of life, but survival itself in times of crisis. The Titanic disaster serves as a tragic illustration of how social structures can become matters of life and death.

METHODOLOGICAL NOTES:
• Analysis based on 891 passenger records from training dataset
• Missing data handled through appropriate statistical methods
• All major findings confirmed through statistical significance testing
• Visualizations designed to highlight key patterns and relationships