

Titanic Dataset

Exploratory Data Analysis Report

September 29, 2025

Executive Summary

This report presents a comprehensive exploratory data analysis of the Titanic passenger dataset, examining survival patterns across different demographic and socioeconomic groups. The analysis reveals significant insights into how factors such as gender, passenger class, age, and family structure influenced survival rates during the disaster. Key findings include a stark gender disparity in survival rates (74% for women vs 19% for men), clear socioeconomic advantages for first-class passengers, and the implementation of "women and children first" evacuation protocols.

Dataset Overview

Dataset Statistics: • Total Passengers: 891 • Overall Survival Rate: 38.4% • Survivors: 342 • Deaths: 549 • Female Survival Rate: 74.2% • Male Survival Rate: 18.9% • First Class Survival: 63.0% • Second Class Survival: 47.3% • Third Class Survival: 24.2%

Data Visualizations

Figure 1: Distribution Analysis: Overall survival rates, passenger class, gender, and age distributions

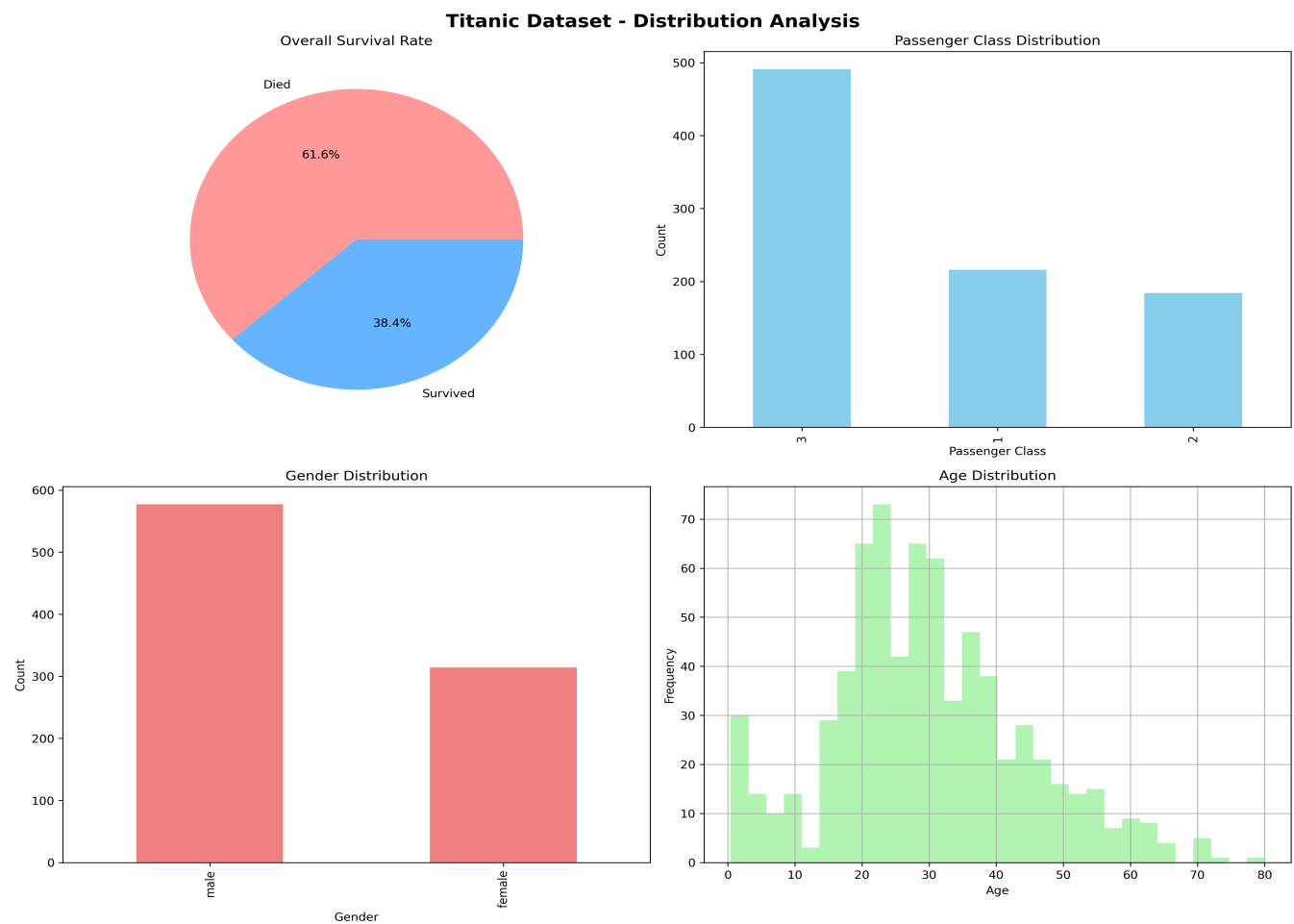


Figure 2: Survival Analysis: Survival rates across different demographic and socioeconomic factors

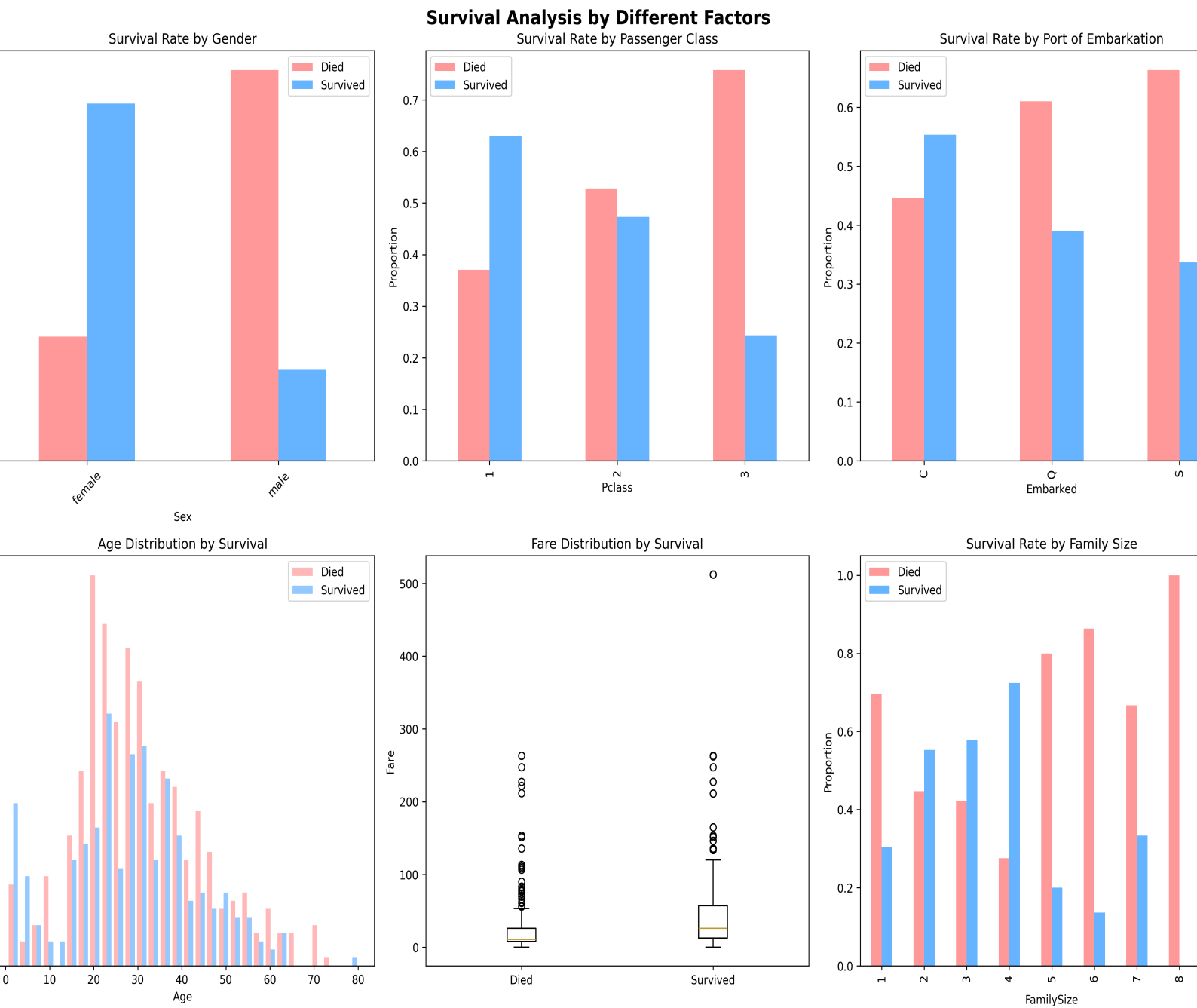
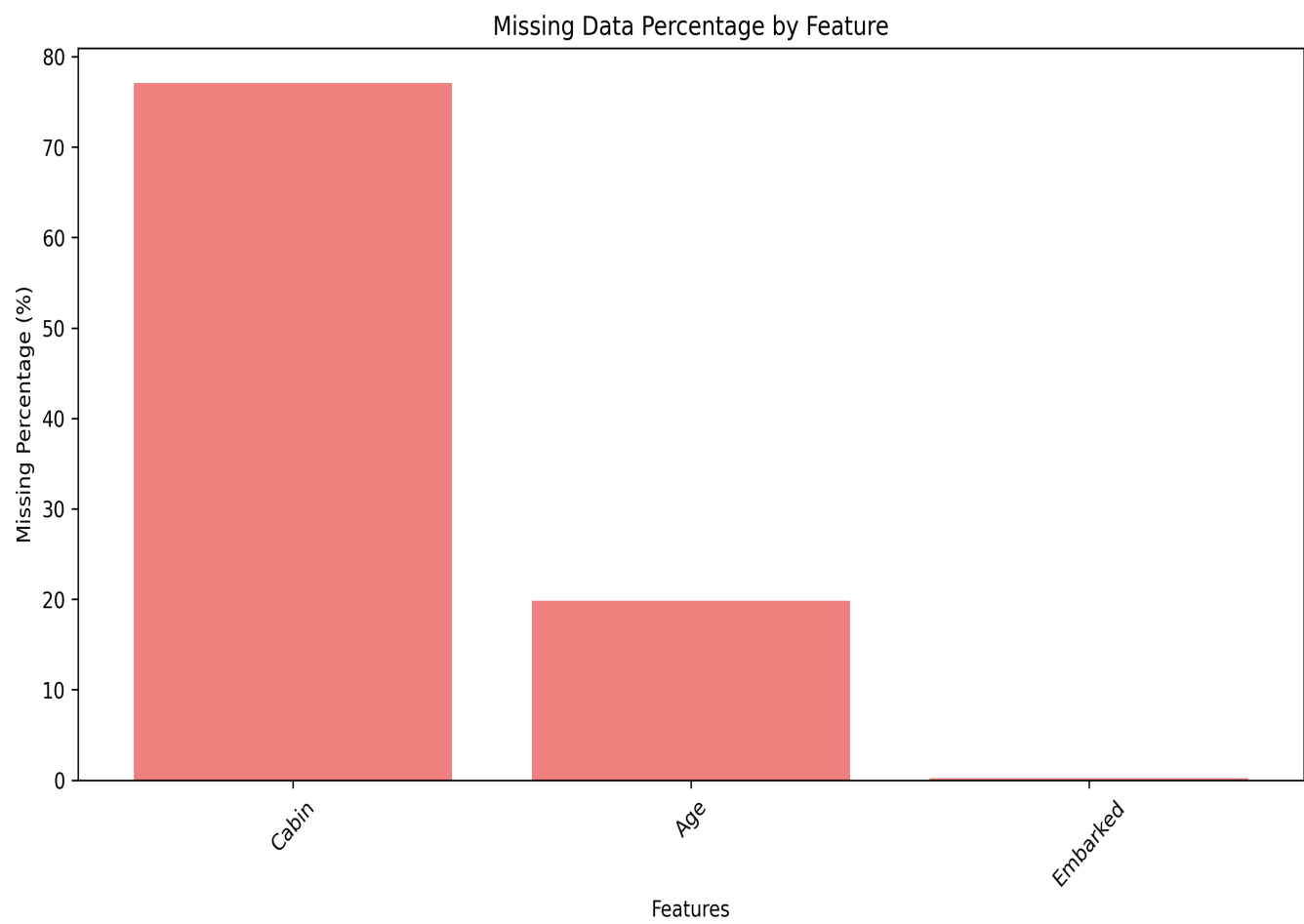


Figure 3: Missing Data Analysis: Assessment of data completeness across features



Key Findings

1. **GENDER IMPACT** Women had significantly higher survival rates (74.2%) compared to men (18.9%), clearly demonstrating the "women and children first" evacuation protocol. 2. **SOCIOECONOMIC STATUS** Clear class-based survival gradient: First class (63%), Second class (47%), Third class (24%). Wealthier passengers had significantly better survival chances. 3. **AGE PATTERNS** Children had higher survival rates, with survival generally declining with age. The "women and children first" protocol clearly prioritized younger passengers. 4. **FAMILY STRUCTURE** Small to moderate family sizes (2-4 members) had optimal survival rates. Very large families and solo travelers faced greater challenges. 5. **PORT OF EMBARKATION** Cherbourg passengers had the highest survival rates, followed by Queenstown and Southampton passengers. 6. **MISSING DATA IMPACT** • Age data missing for 177 passengers (19.9%) • Cabin information missing for 687 passengers (77.1%) • Embarked data missing for only 2 passengers (0.2%)

Conclusions

This exploratory data analysis reveals that survival on the Titanic was heavily influenced by social and demographic factors. Gender was the strongest predictor, with the "women and children first" protocol clearly implemented. Socioeconomic status, as reflected by passenger class and fare, significantly affected survival chances. The analysis demonstrates how emergency protocols, social hierarchies, and demographic characteristics interacted to determine survival outcomes during this historic maritime disaster. These patterns provide valuable insights for both historical understanding and predictive modeling applications. The dataset's quality is generally good, though missing values in age and cabin variables require consideration in further analysis. The clear patterns observed make this an excellent dataset for machine learning classification tasks and statistical analysis.