Titanic Survival EDA Report

Objective

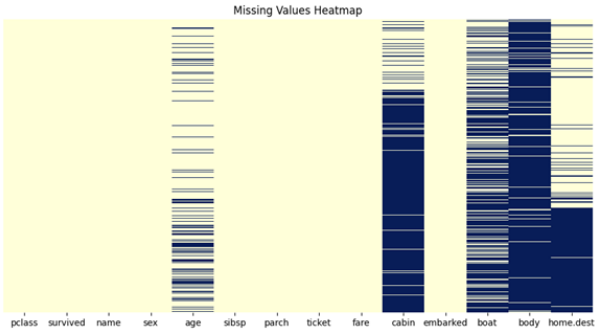
To analyze the Titanic dataset and uncover factors affecting passenger survival using statistical methods and visual exploration.

Dataset Overview

- Total Passengers: 1309
- Features: 14 (e.g., age, gender, class, fare, family, survival)
- Target Variable: survived (1 = survived, 0 = not survived)

Data Cleaning Summary

Missing Values:



· age: 263 missing → filled using median (~28 years)

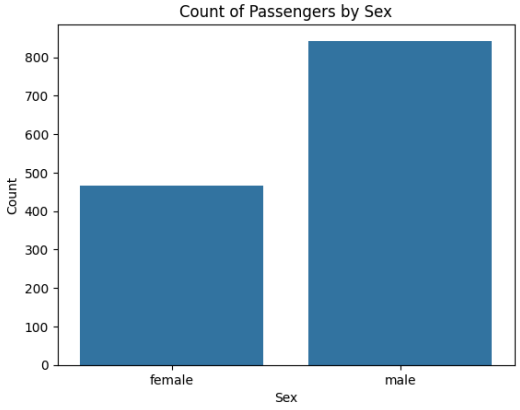
· fare: 1 missing → filled using median (~£14.45)

· embarked: 2 missing → filled using mode ('S')

Dropped Columns: cabin, boat, body, home.dest (too many nulls, not useful for EDA)

Univariate Insights

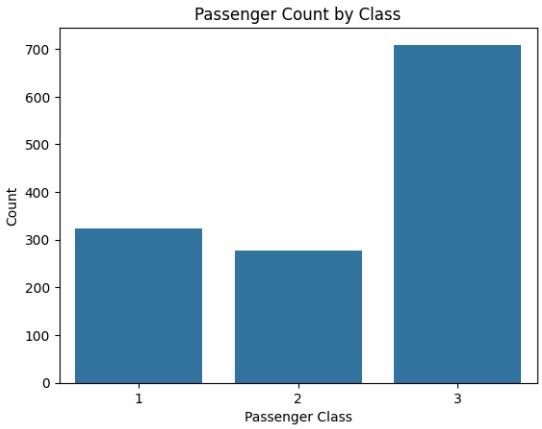
1. Gender



- Males: ~840
- Females: ~470

There were significantly more male passengers on board.

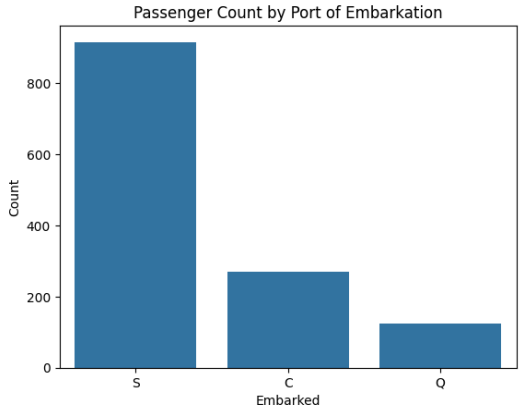
2. Passenger Class



- 1st Class: ~320
- 2nd Class: ~280
- 3rd Class: ~710

Most passengers traveled in 3rd class, reflecting economic diversity.

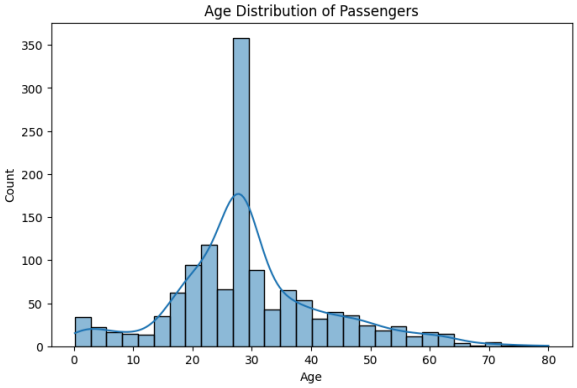
3. Embarkation Port



- Southampton (S): 914
- Cherbourg (C): 270
- Queenstown (Q): 123

Majority of passengers boarded at Southampton.

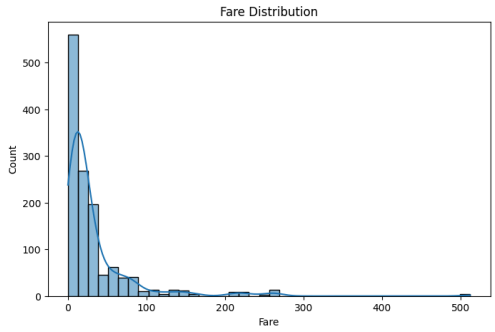
4. Age Distribution



- Median Age: ~28
- Range: 0.17 to 80

🔍*Most passengers were aged between 20 and 40, with some infants and elderly.*

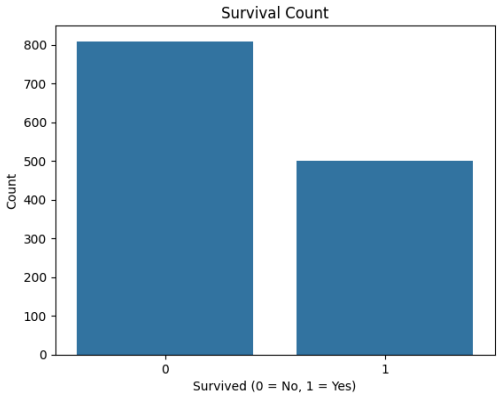
5. Fare



- Median Fare: ~£14.45
- Max Fare: £512.33

Fare distribution is right-skewed — few passengers paid very high prices.

6. Survival Distribution

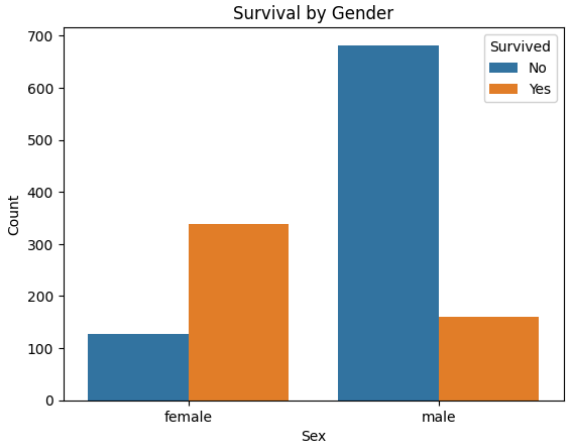


- Survived: 500 (~38%)
- Not Survived: 809 (~62%)

Dataset is imbalanced; majority of passengers did not survive.

Bivariate Insights (with Survival)

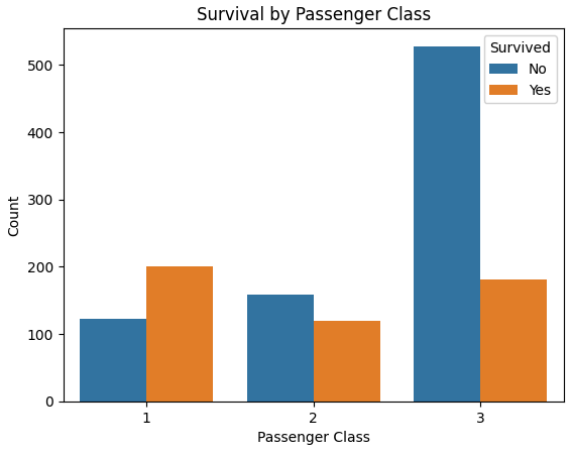
7. Gender vs Survival



- Female: ~325 survived, ~125 died → ~72% survived
- Male: ~180 survived, ~680 died → ~21% survived

Women had far higher survival rates — likely due to "women and children first" evacuation.

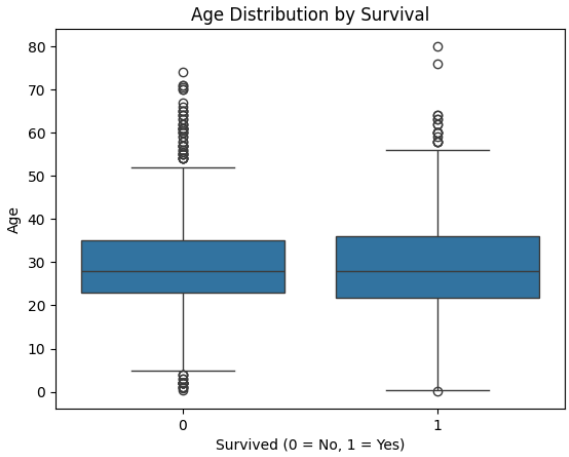
8. Class vs Survival



- 1st Class: ~195 survived, ~120 died → ~62% survived
- 2nd Class: ~120 survived, ~160 died → ~43% survived
- 3rd Class: ~180 survived, ~510 died → ~26% survived

Higher class meant better access to lifeboats and early evacuation — a clear class divide.

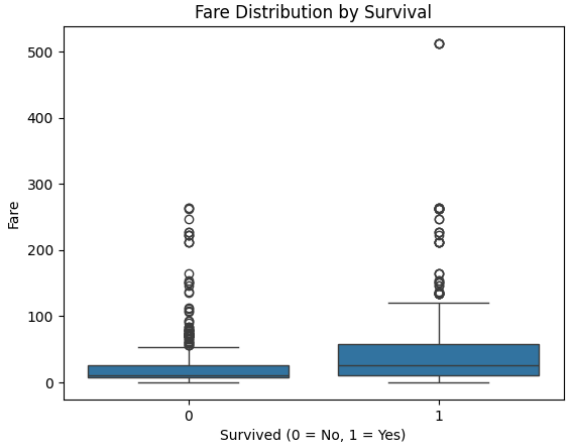
9. Age vs Survival



- Children (<15): Majority survived
- Adults (20–40): Mixed outcomes
- Elderly (>60): Mostly did not survive

Young children were prioritized; elderly possibly had reduced mobility.

10. Fare vs Survival



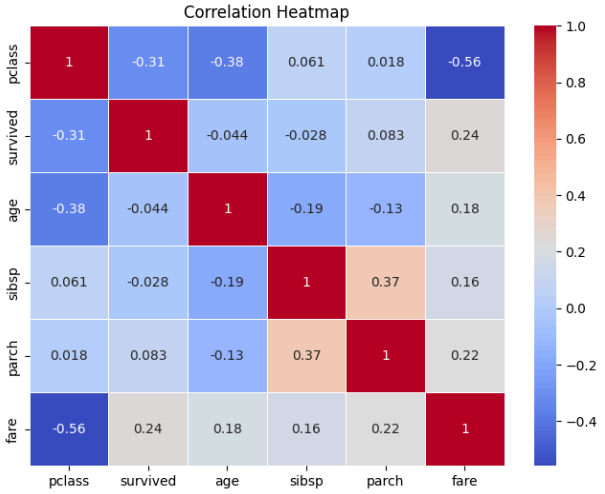
- Fare > £50: High survival density
- Fare < £20: Mostly non-survivors

Fare correlates with both wealth and class — passengers who paid more had better chances.

11. Family Size vs Survival (SibSp & Parch)

- SibSp = 1–2: Best survival rate (~50%)
- SibSp > 3 or alone (0): Survival rate drops below 30%
- Parch = 1–2: Higher survival
- Parch > 3: Very low survival*Small families helped each other escape; large families were harder to evacuate together.*

Correlation Analysis

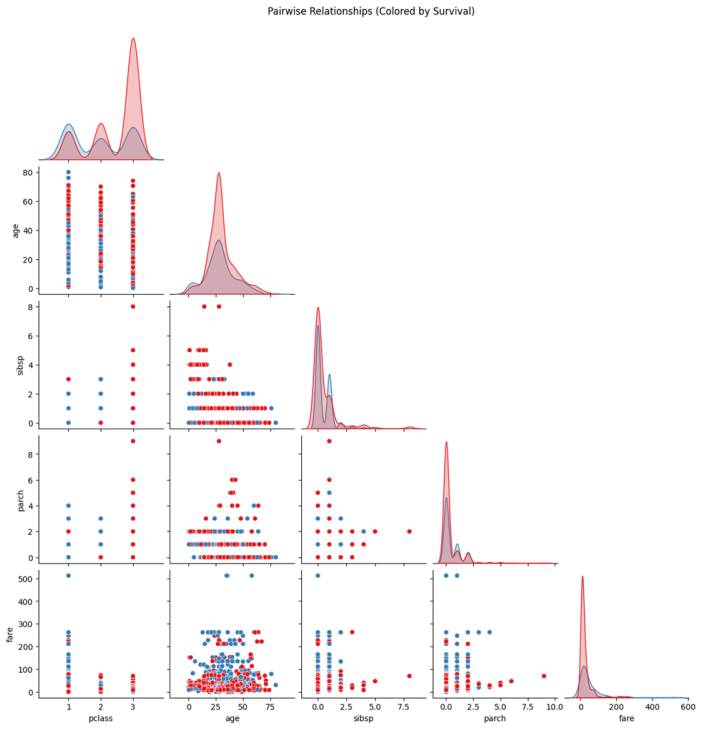


Fare and Class had the strongest influence on survival, while age and family size had nuanced effects.

Feature	***Correlation with Survival***	***Interpretation***
Fare	+0.24	Higher fare → more likely to survive
Pclass	–0.31	Lower class → less chance of survival
Age	–0.07	Elderly less likely to survive
Parch	+0.08	Having 1–2 children/parents helped

SibSp	–0.03	Large groups lowered survival

Pairplot Interpretation



<u>Survivors were clustered in:</u>

- High Fare + Low Pclass (1st class)
- Younger Age (especially children)
- Small Family Groups (1–2 relatives)

<u>Non-survivors:</u>

- Concentrated in 3rd class
- Low Fare, often alone or in large families
- Older adults and males had lowest survival

Summary

- Women, children, 1st class passengers had the highest survival chances.
- Fare and Pclass were the two strongest predictors.
- Family helped — 1–2 relatives onboard improved survival odds.
- Elderly, men, and 3rd class passengers were most vulnerable.
- Survival outcomes reveal clear socio-economic and policy-driven patterns.