

Titanic Dataset – Exploratory Data Analysis (Detailed Insights Report)

This report presents a structured analysis of the Titanic dataset, focusing on demographic patterns, fare behaviour, and variables that influenced survival outcomes. All insights are supported by the six visuals you provided.

1. Univariate Analysis

1.1 Age Distribution (Histogram)

Insight:

- The **majority of passengers were between ages 20 and 40**, with the **peak around 28–30 years**.
- Very few passengers were above **60 years** or below **5 years**, indicating that the ship carried relatively **fewer children and elderly**.
- The distribution is slightly **right-skewed**, meaning younger adults were more common than older adults.

Business / Analytical Interpretation:

- Demographically, most travellers were working-age adults, suggesting Titanic passengers were mainly individuals capable of long voyages.
 - This age imbalance also plays a role in understanding survival patterns: younger adults had different survival probabilities depending on gender and class.
-

1.2 Gender Distribution (Bar Chart)

Insight:

- The dataset contains **significantly more males (~580)** than females (~310).
- Males represent nearly **65% of total passengers**, showing a clear demographic imbalance.

Business Interpretation:

- This imbalance directly affects survival outcomes — since survival rates vary heavily by gender, having more males results in a lower overall survival percentage.
 - The overrepresentation of males could reflect the travel norms of the early 1900s where men commonly travelled alone for work/trade.
-

✓ 2. Bivariate Analysis

2.1 Survival by Gender (Grouped Bar Chart)

Observations (Numerical + Insightful):

- **~75% of females survived**, compared to **only ~20% of males**.
- Male non-survivors (~470) drastically outweigh male survivors (~110).
- Female survivors (~230) outnumber female non-survivors (~80).

Business Interpretation:

- This confirms the “**Women and Children First**” evacuation policy during the sinking.
 - Gender was the **strongest predictor** of survival in this dataset — stronger than age or fare.
 - FEMALES had **3 to 4 times higher** survival probability than males.
-

2.2 Fare vs Survival (Box Plot)

Insights:

- Survivors generally paid **higher fares** than non-survivors.
- Median fare for survivors \approx **26–30 units** vs **~10–12 units** for non-survivors.
- Numerous high-fare outliers exist among survivors (some paying **>150–300 units**).

Business Interpretation:

- Passengers in **higher socioeconomic classes** (represented by higher fares) had better access to life-saving resources (lifeboats).
 - This reinforces that **class inequality** significantly influenced survival.
-

✓ 3. Multivariate Analysis

3.1 Correlation Heatmap (Numerical Features)

Key Findings with Numeric Evidence:

- **Fare & Survived: Positive correlation (~0.26)**
→ Higher fare = Higher chance of survival
- **Pclass & Survived: Negative correlation (~-0.34)**
→ Lower class number (1st class) = Higher survival
- **Age & Survived: Very weak correlation (~-0.07)**
→ Age was not a major factor in survival
- **SibSp/Parch correlations are near zero**
→ Family size had little influence on overall survival

Business Interpretation:

- **Passenger class (Pclass)** is the strongest numerical predictor of survival.

- **Fare**, as a proxy for wealth, plays a significant role in survival chances.
 - Age and family size are **not** strong standalone predictors.
-

3.2 Age vs Fare Coloured by Survival (Pastel Scatter Plot)

Insights:

- Survivors (pastel blue) are more concentrated in the **higher fare** range.
- Non-survivors (pastel pink) are mostly in the **lower fare** ranges.
- Across all ages, higher fare consistently aligns with higher survival.
- No meaningful pattern between age and survival — variations are spread out.

Business Interpretation:

- Fare (wealth) was far more influential than age in determining survival.
 - Even younger or older passengers with lower fares had reduced survival probability.
 - The plot visually confirms that **socioeconomic status > age** as a survival determinant.
-

4. Summary of Key Insights (Professional & Business-Oriented)

1. Passenger Demographics

- Majority were **20–40 years old**, suggesting the ship carried mainly working-age adults.
 - Gender imbalance: **65% male**, which affects overall survival ratios.
-

✦ 2. Survival Predictors (Ranked by Importance)

Based on patterns across all visuals:

1️⃣ Gender — strongest factor

- Female survival rate: **~75%**
- Male survival rate: **~20%**

2️⃣ Socioeconomic Status (Fare & Class)

- Survivors paid **~2x higher fares** than non-survivors.
- First-class passengers dominated survival counts.

3️⃣ Age — weak predictor

- Minimal correlation with survival (-0.07).
 - Young adults and older adults had similar probabilities after adjusting for gender/class.
-

✦ 3. Wealth Inequality in Outcomes

- Passengers who paid higher fares had disproportionately higher survival, indicating **clear class-based advantage** during evacuation.
-

✦ 4. Data-Driven Business Understanding

If this were a modern customer or safety analysis project, findings imply:

- **VIP / Premium customers** receive significantly better outcomes in crisis situations.

- **Customer demographics strongly influence service accessibility.**
 - **Gender-based evacuation policies drastically shift outcome patterns.**
-

5. Final Executive Summary (Short & Business Ready)

The Titanic dataset reveals that survival was **not random**. Instead, it was heavily influenced by **gender and socioeconomic status**. Females and 1st-class passengers had significantly higher survival chances, while males and lower-class passengers experienced disproportionately higher casualties. Age played a minimal role, and wealthy passengers (high fare payers) were dramatically more likely to survive.