

Observations for each visual

1-Survival Count



Observation:

- Around 62% passengers died (label 0) and only 38% survived (1).
- There is a clear class imbalance — more deaths than survivals.

2-Passenger Class Distribution



Observation:

- Majority of passengers belonged to 3rd class, followed by 1st and 2nd.
- Indicates most travelers were from lower socio-economic backgrounds.

3-Gender Distribution



Observation:

- About 65% of passengers were male, and 35% were female.
- Shows a gender imbalance aboard the Titanic.

4 – Embarkation Port Distribution



Observation:

- Most passengers (over 70%) boarded at Southampton (S).
- Fewer passengers boarded at Cherbourg (C) and Queenstown (Q).

5- Age Distribution



Observation:

- Most passengers were between 20 and 40 years old.
- A few passengers were infants or elderly, but they were in the minority.

6- Fare Distribution (Boxplot)



Observation:

- Fare distribution is right-skewed with many outliers.
- Most passengers paid under \$100, but a few paid over \$500.

7- Survival by Gender



Observation:

- A much higher proportion of females survived compared to males.
- Shows strong bias in survival rate based on gender.

8- Survival by Passenger Class



Observation:

- 1st class passengers had the highest survival rate.
- 3rd class passengers had the lowest — indicating class privilege influenced survival.

9- Survival by Embarkation Port



Observation:

- Passengers from Cherbourg (C) had a higher chance of survival than those from Southampton (S).
 - Could indicate a correlation between boarding port and passenger class.
-

10- Age vs Survival (Boxplot)



Observation:

- Younger passengers, especially children, had a slightly better chance of survival.
- Median age of survivors was lower than that of non-survivors.

11- Correlation Heatmap



Observation:

- Fare and Pclass show strong negative correlation.
- Survival positively correlates with Fare and Parch, and negatively with Pclass.
- Age and Survival have a weak negative correlation.

12- Pairplot



Observation:

- Clear separations seen in Fare vs Pclass vs Survival.
- Survivors cluster in lower Pclass and higher Fare.

Summary of Findings – Titanic Dataset EDA

1. Survival Rate

- **38% of passengers survived**, while **62% did not**.
- The dataset is **imbalanced**, with more non-survivors than survivors.

2. Gender Impact

- Females had a much higher survival rate than males.
- This suggests "**women and children first**" protocol was followed during evacuation.

3. Class and Socioeconomic Status

- **1st class passengers** had the **highest survival rate**.
- **3rd class passengers** had the **lowest survival rate**.
- **Pclass is strongly negatively correlated** with survival ($r \approx -0.34$), indicating class privilege.

4. Age Distribution

- Most passengers were aged between **20 and 40**.
- Children had **slightly higher survival chances**.
- Age has **weak negative correlation** with survival.

5. Fare and Survival

- Higher ticket fares were generally associated with **higher survival probability**.
- Fare and Pclass are **strongly negatively correlated** ($r \approx -0.55$), implying richer people traveled in higher classes.

6. Family Size (SibSp & Parch)

- Small families (1-2 members) had a **better survival rate**.
- Large families or those traveling alone had **lower survival rates**.

7. Embarkation Port

- Most passengers boarded at **Southampton (S)**.
- Passengers from **Cherbourg (C)** had **higher survival rates**, likely because many 1st class passengers boarded there.

8. Missing Values

- **Cabin** column had too many missing values and was dropped.
- **Age** and **Embarked** columns were imputed using median and mode respectively.

9. Key Influential Features for Survival

Based on the correlation and visual analysis, the following features significantly influenced survival:

- **Sex**
- **Pclass**
- **Fare**
- **Embarked**
- **Family size (SibSp, Parch)**

Conclusion

The analysis highlights **strong survival bias based on gender, class, and fare**. These insights are valuable for predictive modeling and understanding the human factors in survival scenarios.