

**Project Title:** Exploratory Data Analysis (EDA) on Titanic Dataset

**Submitted by:** Sourav Banerjee (Data Analyst)

**Tools Used:** Python, Pandas, Matplotlib, Seaborn, Jupyter Notebook

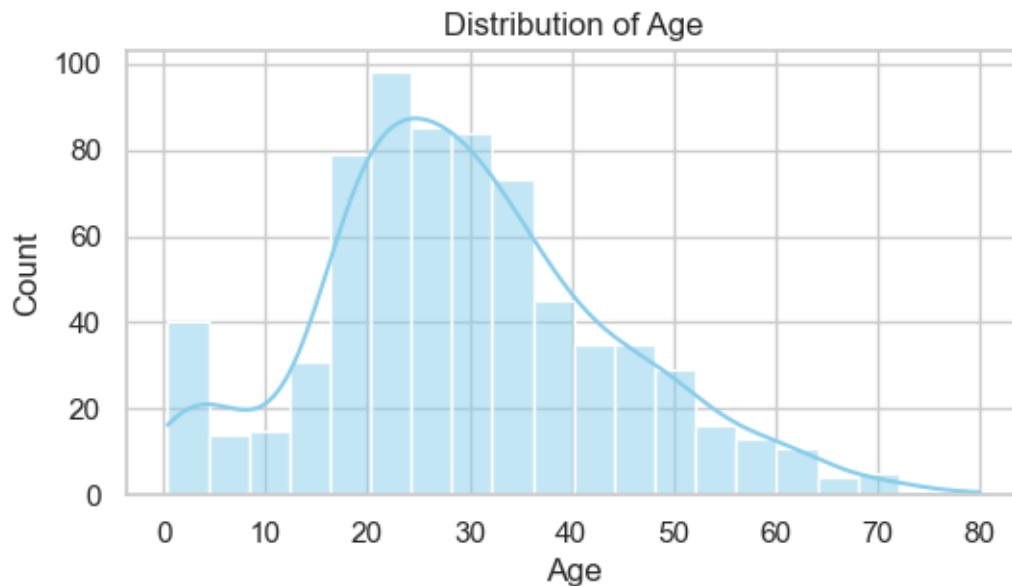
**Data Source:** [Titanic - Machine Learning from Disaster | Kaggle](#)

## 1. Dataset Overview

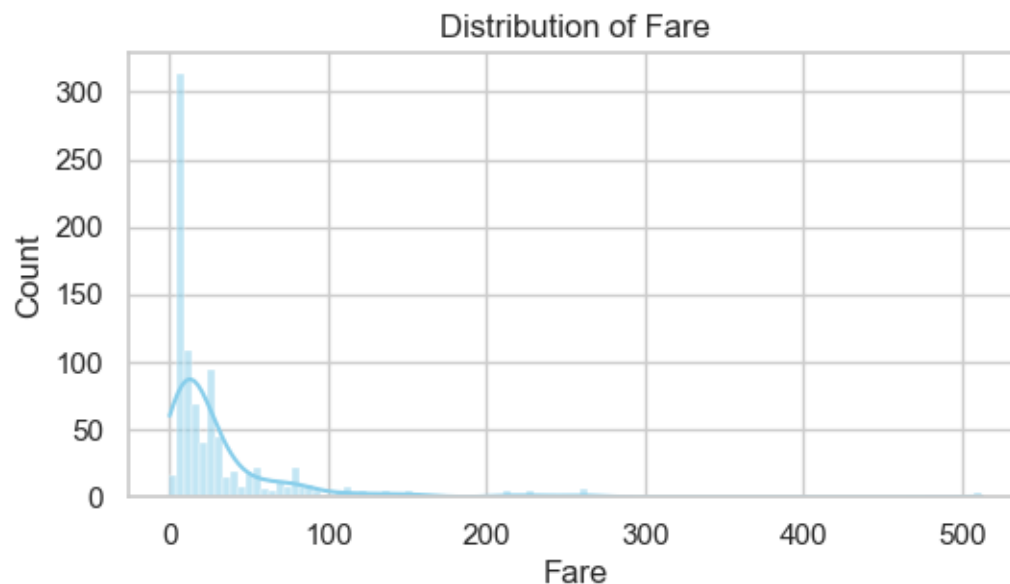
- The dataset contains passenger information such as **Age**, **Sex**, **Passenger Class (Pclass)**, **Fare**, **Embarked**, and **Survival status**.
  - Total records: 891 (approx.)
  - About 38% of passengers survived.
  - Missing data was mainly in **Age** and **Cabin** columns.
- 

## 2. Univariate Analysis

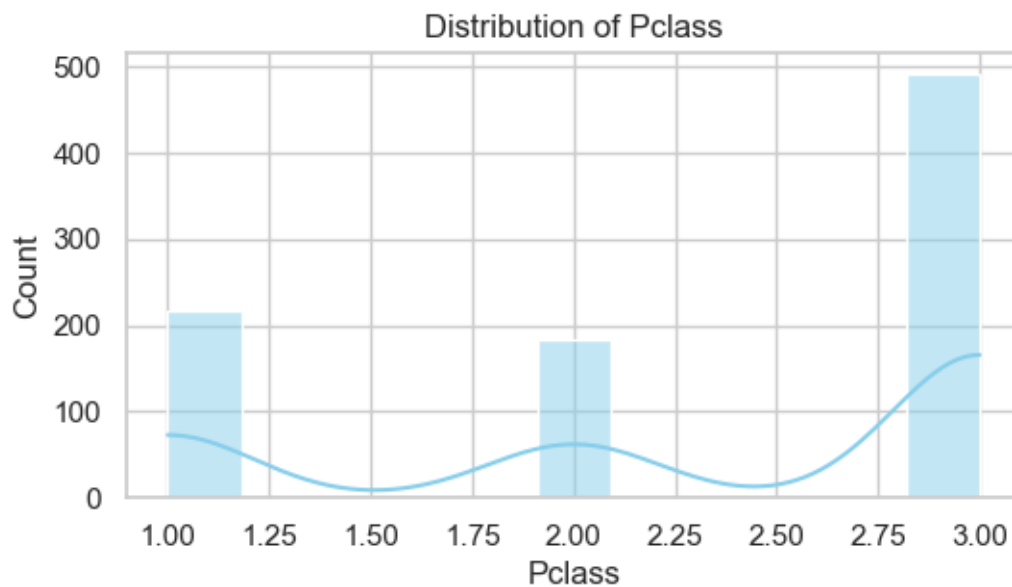
- **Age Distribution:** Most passengers were between **20–40 years**, indicating a young adult majority.



- **Fare Distribution:** Fares are **right-skewed**, with a few very high-paying passengers.



- **Pclass:** Majority of passengers traveled in **3rd class**, followed by 1st and 2nd.



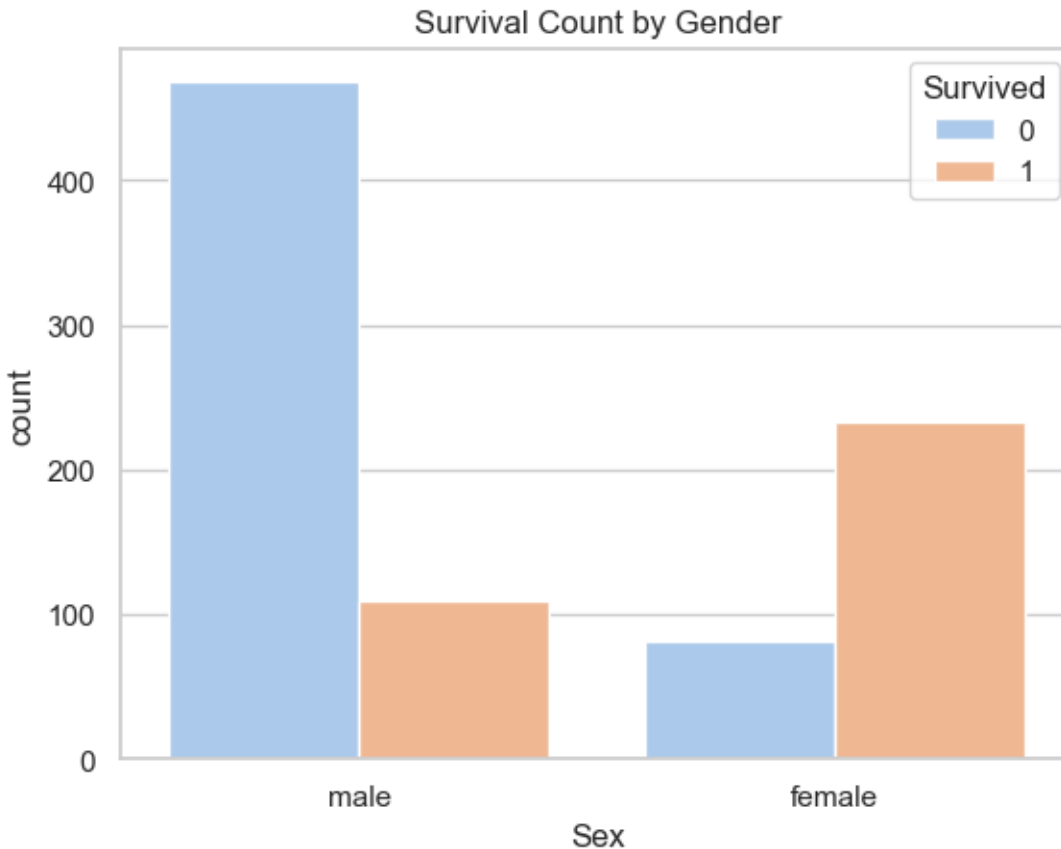
*Observation:* The passenger profile suggests that the majority were low-fare, 3rd-class males in their 20s–30s.

---

### 3. Bivariate Analysis:

- **Survival vs. Gender:**

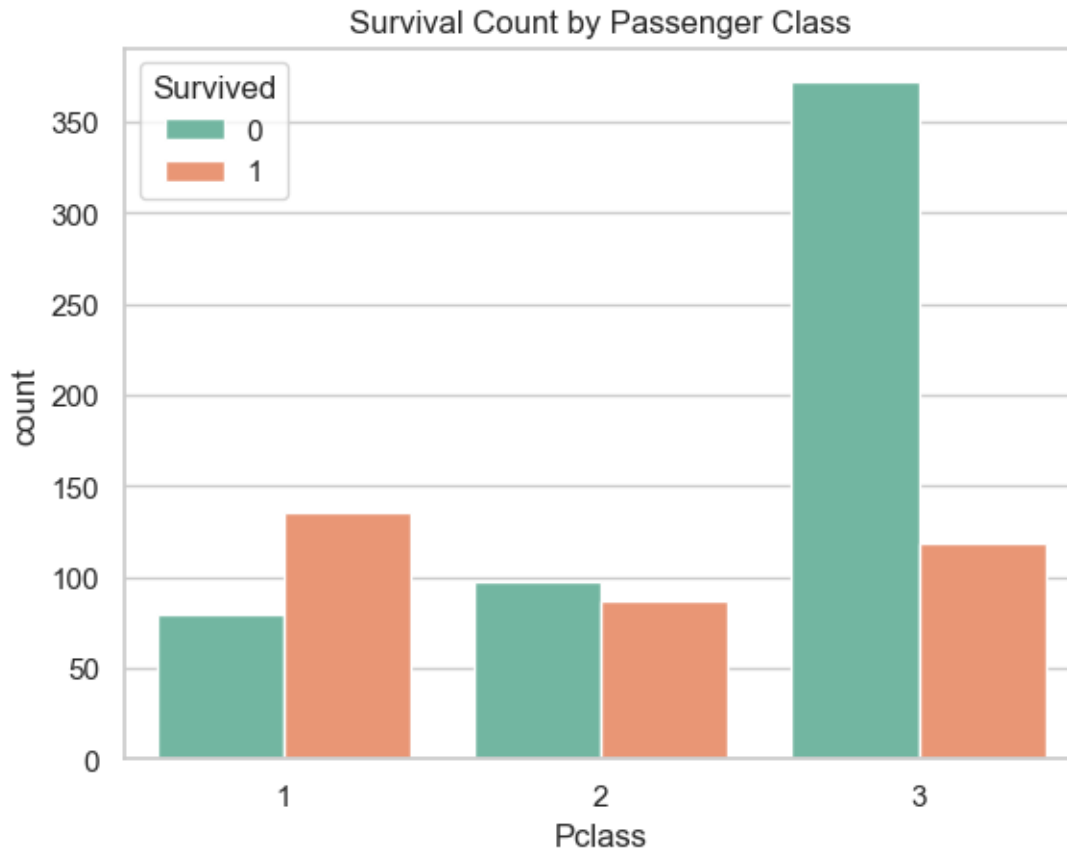
Females had a much higher survival rate than males (around **74% vs 19%**).



*Insight:* Gender played a major role in survival — supporting the “women and children first” policy.

- **Survival vs. Pclass:**

1st-class passengers had the highest survival rate (~63%), while 3rd-class had the lowest (~24%)



*Insight:* Socio-economic status strongly influenced survival chances.

- **Survival vs. Age:**

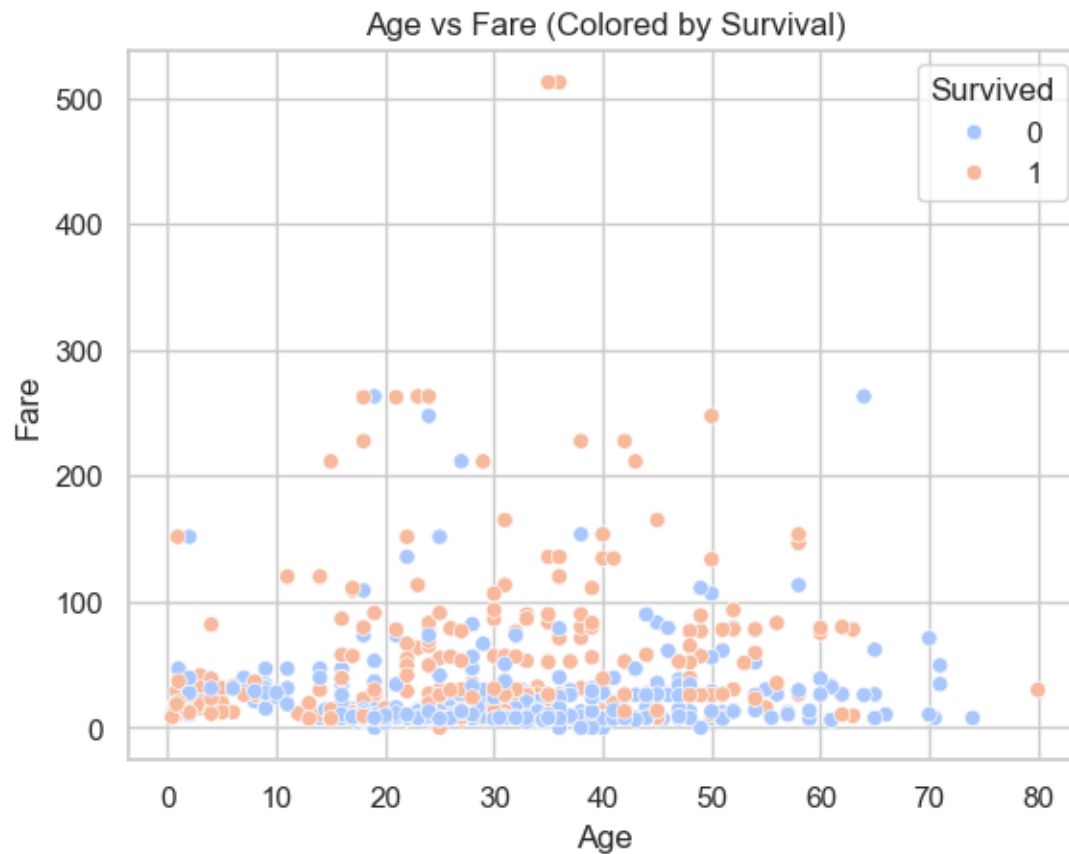
Younger passengers (especially children) had higher survival rates compared to older passengers.

*Insight:* Children were prioritized during rescue operations.

- **Survival vs. Fare:**

Passengers who paid **higher fares** were more likely to survive.

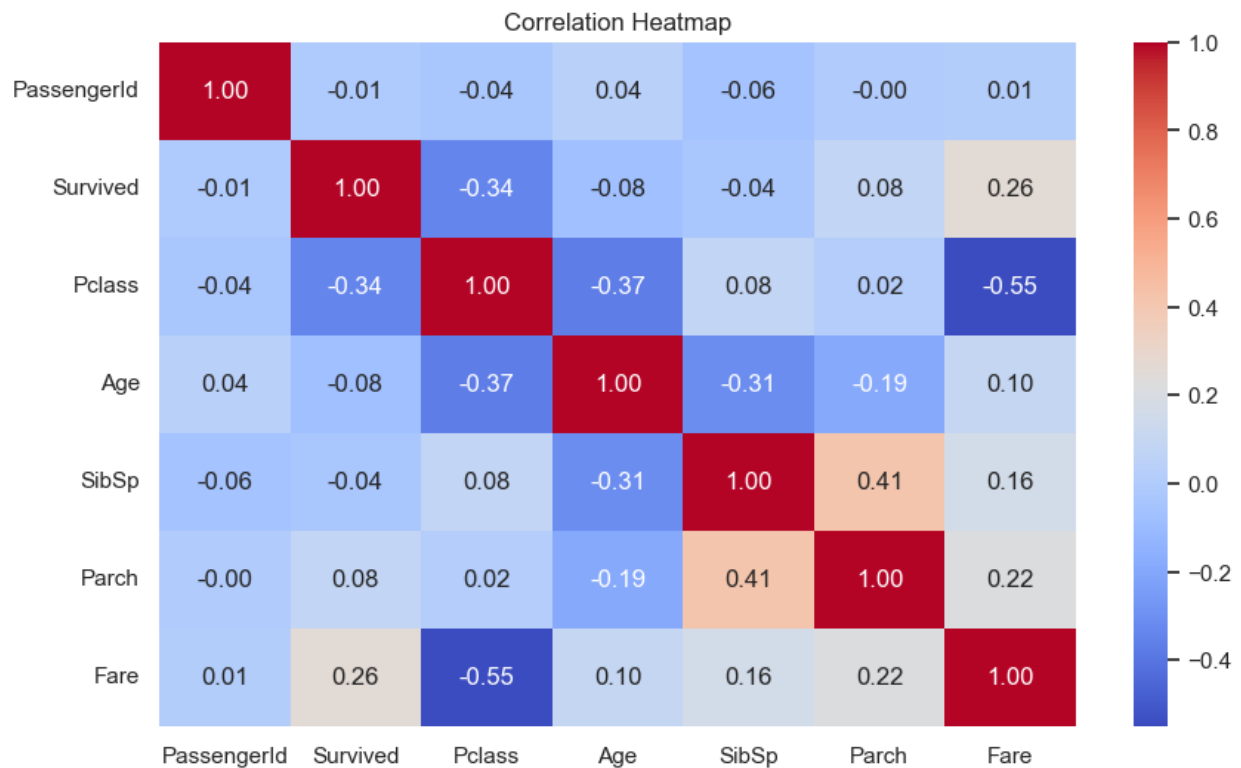
*Insight:* Fare correlates positively with survival, indirectly reflecting class advantage.



#### 4. Correlation Analysis

- Strong negative correlation between **Pclass** and **Fare** ( $-0.55$ ).

- **Survived** shows positive correlation with **Fare** and negative with **Pclass**.
- Weak or no correlation between **Age** and **Survival**.



*Insight:* Wealthier passengers in higher classes had better chances of survival.

## 5. Pairplot Observations

- Distinct clusters visible for **1st-class survivors**.
- Overlap among **3rd-class non-survivors**, indicating higher density of loss in that group.



## 6. Summary of Key Insights

- Overall survival rate: **~38%**.
- **Gender:** Females survived at a significantly higher rate than males.



- **Class:** 1st-class passengers were most likely to survive; 3rd-class least.
  - **Age:** Children had better survival chances than adults.
  - **Fare:** Higher fares → higher survival probability.
  - **Embarkation:** Passengers from Cherbourg had slightly better survival rates. **7.**
- 

## 8. Conclusion

- The EDA reveals that **socio-economic status, gender, and age** were the strongest determinants of survival aboard the Titanic. Wealthier, first-class passengers—especially women and children—had the highest chances of survival, reflecting clear patterns of social inequality and emergency prioritization during the disaster.
-