**Titanic Train Dataset – Exploratory Data Analysis (EDA) Report**

**1. Objective**

The purpose of this analysis is to explore the Titanic train dataset using visual and statistical methods to uncover patterns, trends, and anomalies related to passenger demographics, ticket fares, and survival rates.

**2. Dataset Information**

* **Source**: Titanic Train Dataset (Kaggle)
* **Number of Rows**: *891*
* **Number of Columns**: *12*
* **Key Features**:
  + PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked
* **Target Variable**: Survived (1 = survived, 0 = did not survive)

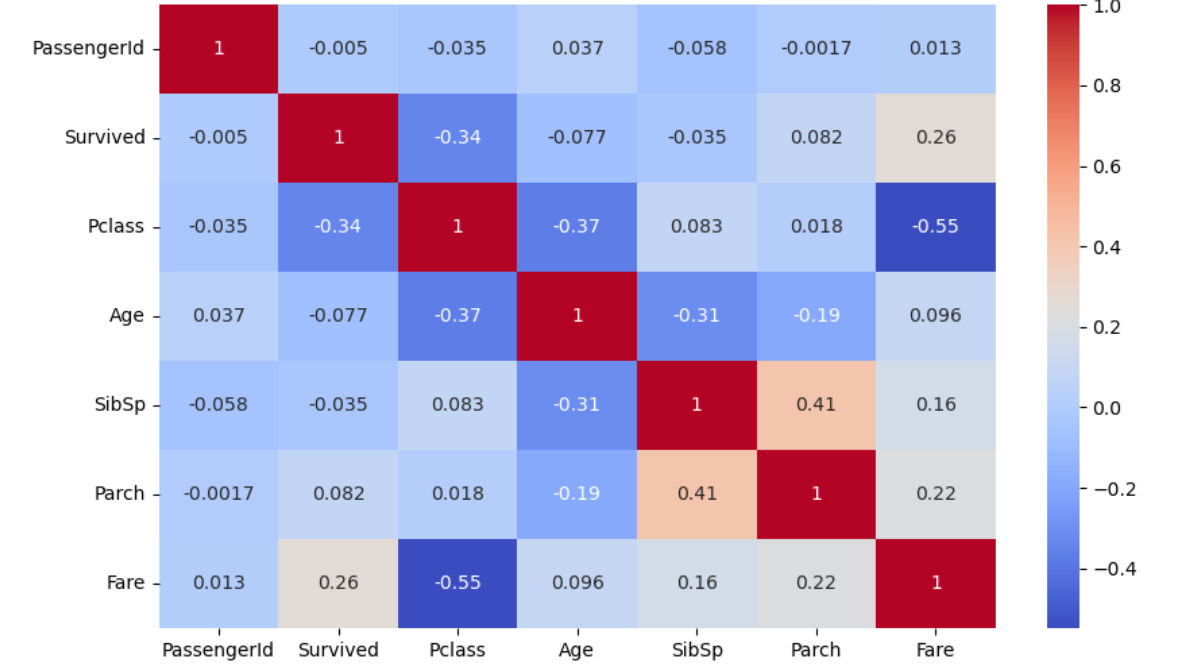
**3. Summary Statistics**

* Average Age: ~29.7 years
* Median Fare: ~14.45
* Majority Class: Pclass 3 (lower class)
* Gender Distribution: More males than females
* Missing Values: Age (~177 missing), Cabin (~687 missing), Embarked (~2 missing)

**4. Visual Analysis & Observations**

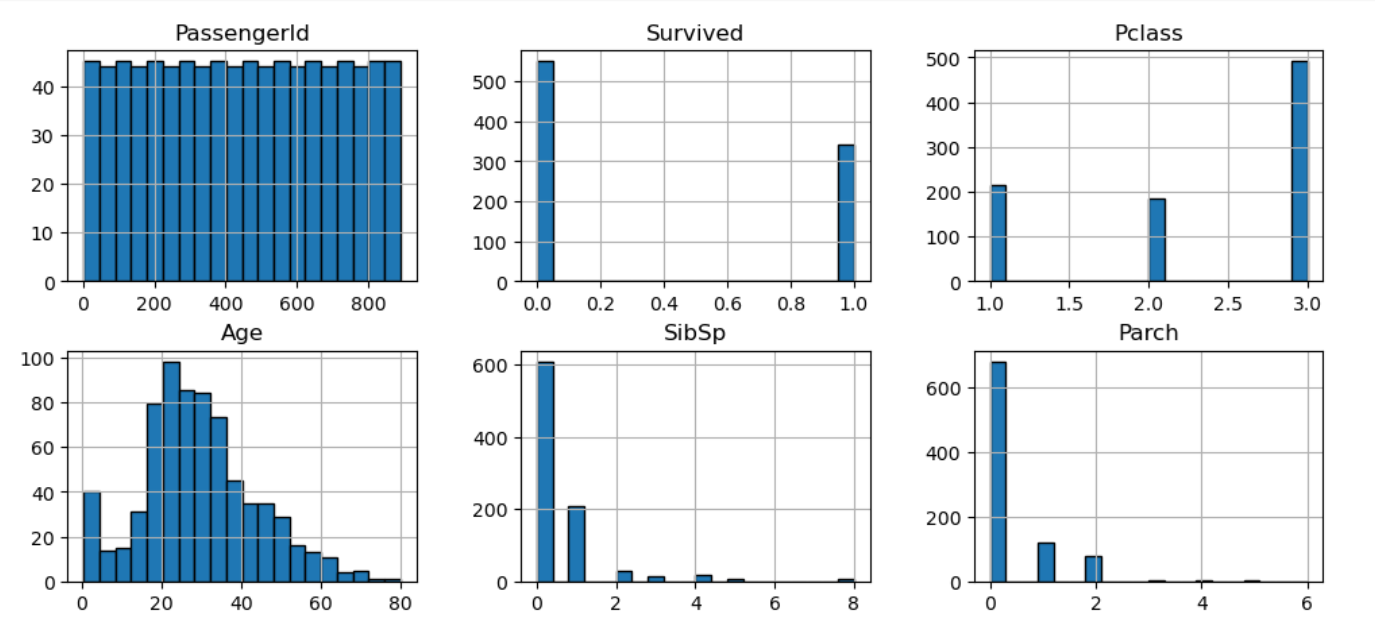
**a. Correlation Heatmap**

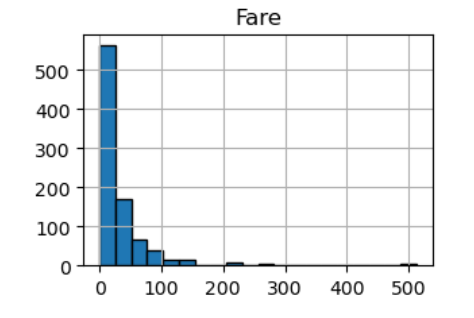
* Shows a strong negative correlation between Pclass and Fare.
* Survived has a positive correlation with Fare and a negative correlation with Pclass.



**b. Histograms**

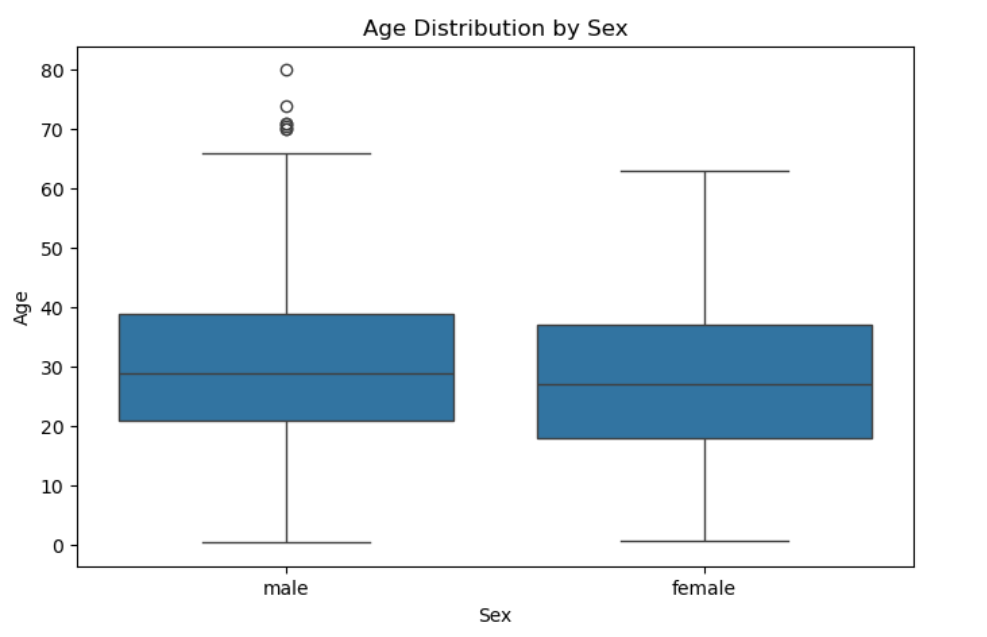
* Age is right-skewed, with more passengers aged 20–40.
* Fare distribution is highly skewed with some high-value outliers.



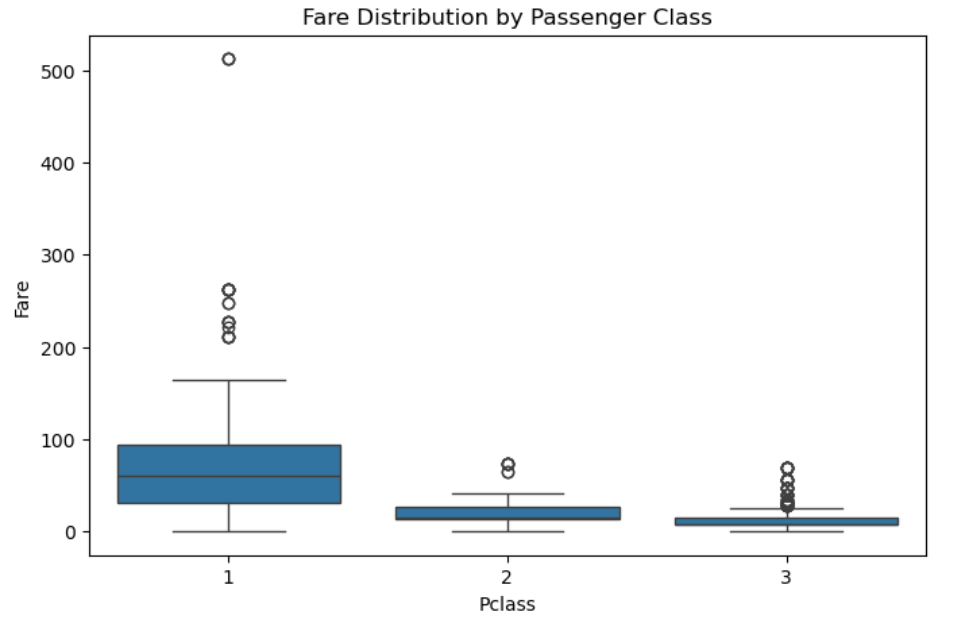


**c. Boxplots**

* **Age by Sex**: Males have a slightly lower median age compared to females.

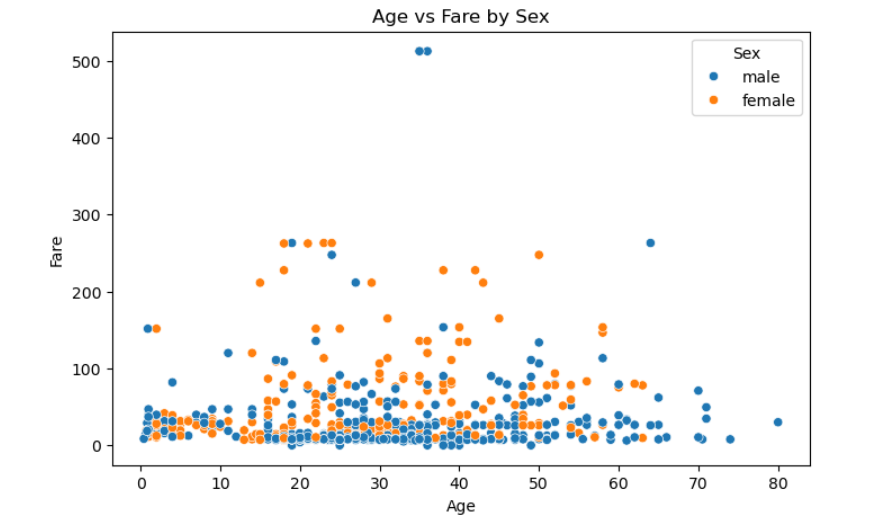


* **Fare by Class**: First-class passengers pay significantly higher fares than lower classes.



**d. Scatterplot (Age vs Fare)**

* Higher fares are generally associated with younger passengers in higher classes.



**5. Key Observations**

1. Majority of passengers are male.
2. Most passengers belong to the third class (Pclass 3).
3. Fare distribution is skewed; few passengers paid very high amounts.
4. Younger passengers appear to have slightly higher survival chances.
5. Survival rate is higher among females and first-class passengers.

**6. Summary of Findings**  
The analysis of the Titanic dataset aimed to uncover patterns and relationships between passenger demographics, ticket classes, and survival rates. The dataset provided details such as passenger age, gender, ticket class, fare, and survival status, enabling a comprehensive exploration of the factors influencing survival during the tragedy.

**Key Observations:**

1. **Overall Survival Rate:**  
   Approximately **38%** of passengers survived. This shows that survival was not evenly distributed across all passengers and may have been influenced by social and demographic factors.
2. **Impact of Gender:**  
   A strong gender disparity was observed — **female passengers had a significantly higher survival rate (~74%)** compared to male passengers (~19%). This supports the “women and children first” evacuation principle that appears to have been applied.
3. **Impact of Passenger Class (Pclass):**  
   First-class passengers had the highest survival rate (~63%), followed by second class (~47%), while third-class passengers had the lowest (~24%). This suggests that higher-class passengers had better access to lifeboats and assistance during evacuation.
4. **Age Distribution and Survival:**  
   Children under 10 years had a survival rate of around **60%**, indicating a priority for younger passengers during rescue operations. However, survival rates decreased steadily for older age groups.
5. **Fare Influence:**  
   Passengers who paid higher fares generally had a better survival rate, which aligns with the higher survival rates of first-class passengers.
6. **Embarkation Point:**  
   Passengers who boarded from **Cherbourg (C)** had a higher survival rate compared to those from **Southampton (S)** and **Queenstown (Q)**. This may be related to the proportion of first-class passengers embarking from each location.