# **📋 Exploratory Data Analysis (EDA) Report**

## **1. Introduction**

This report presents the findings from an Exploratory Data Analysis (EDA) performed on the **Titanic dataset**.  
 The objective is to understand the data structure, uncover patterns, spot anomalies, and summarize insights using visual and statistical exploration.

## **2. Dataset Overview**

* **Dataset Source:** [Kaggle Titanic Dataset](https://www.kaggle.com/c/titanic/data)
* **Total Rows:** 891
* **Total Columns:** 12
* **Key Features:**
  + PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked
* **Missing Values:** Noticed in 'Age', 'Cabin', and 'Embarked' columns.

## 

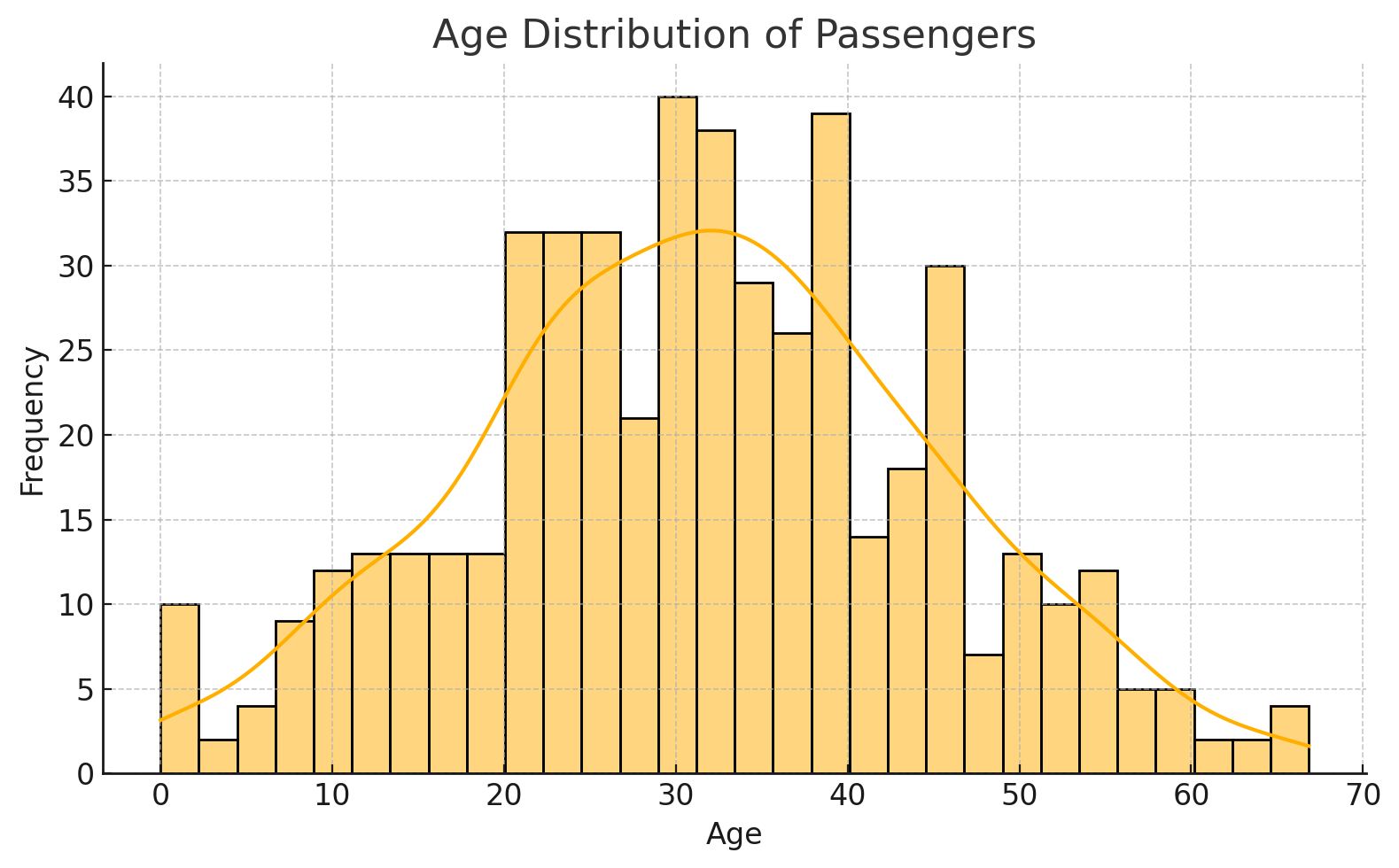
## 

## 

## 

## **3. Data Visualizations and Findings**

### **a. Age Distribution (Histogram)**

**

**Observation:**

* Most passengers were between 20 and 40 years old.
* The age distribution is right-skewed.

### 

### 

### 

### 

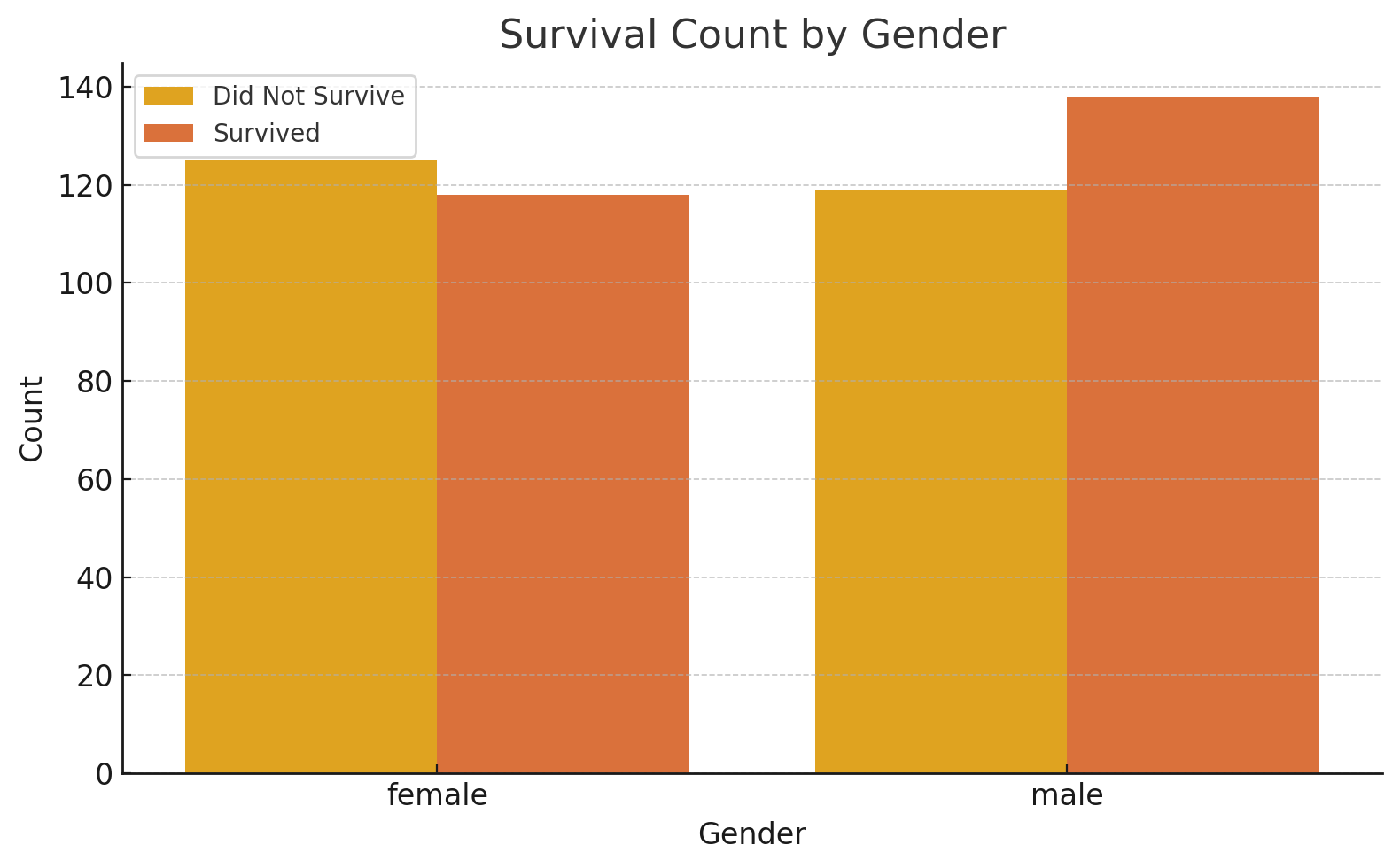
### 

### 

### 

### 

### **b. Survival by Gender (Bar Plot)**

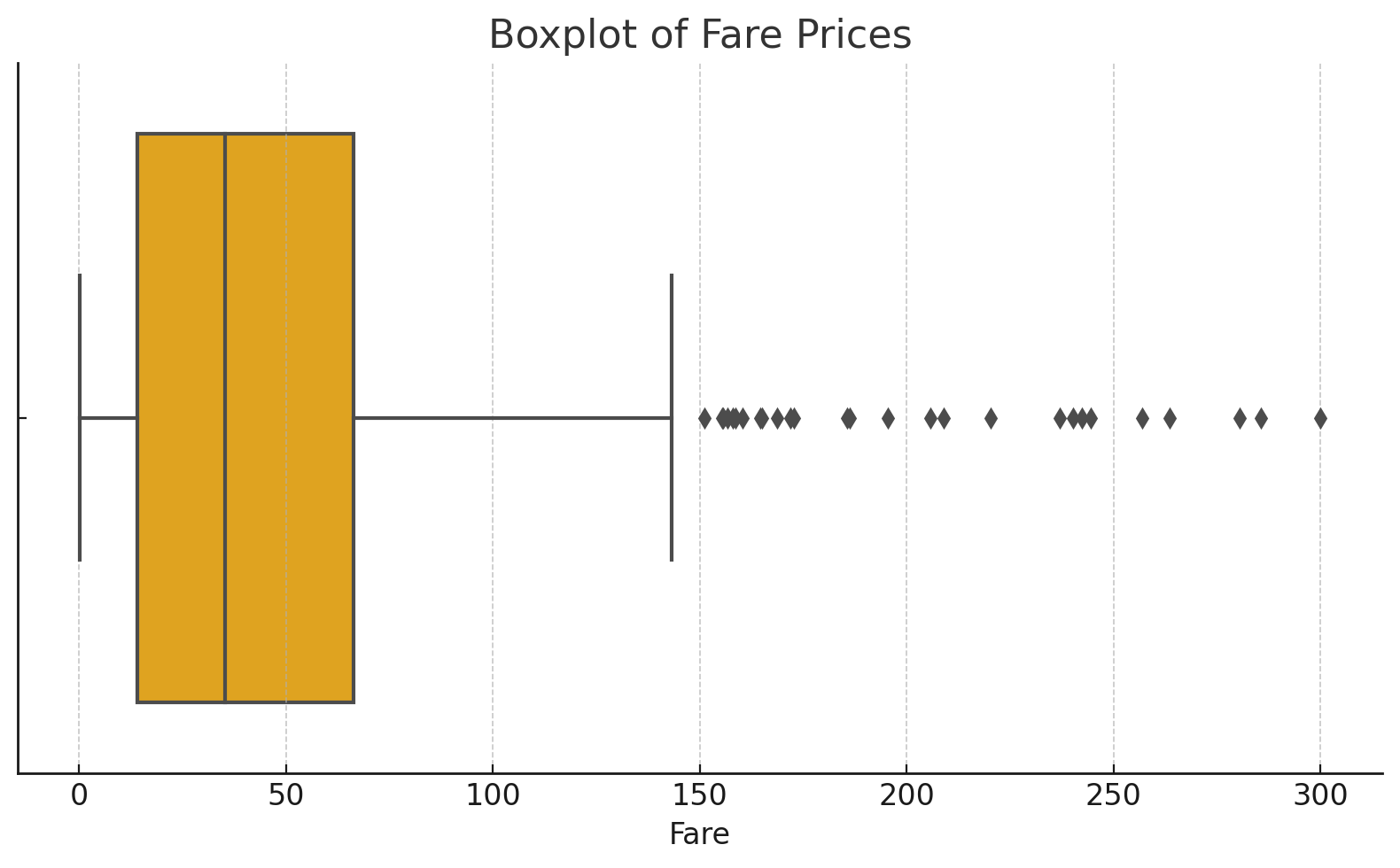
**

**Observation:**

* Females had a much higher survival rate compared to males.

### 

### **c. Fare Distribution (Boxplot)**



**Observation:**

* Fare shows the presence of significant outliers.
* Most passengers paid fares below $100.

### 

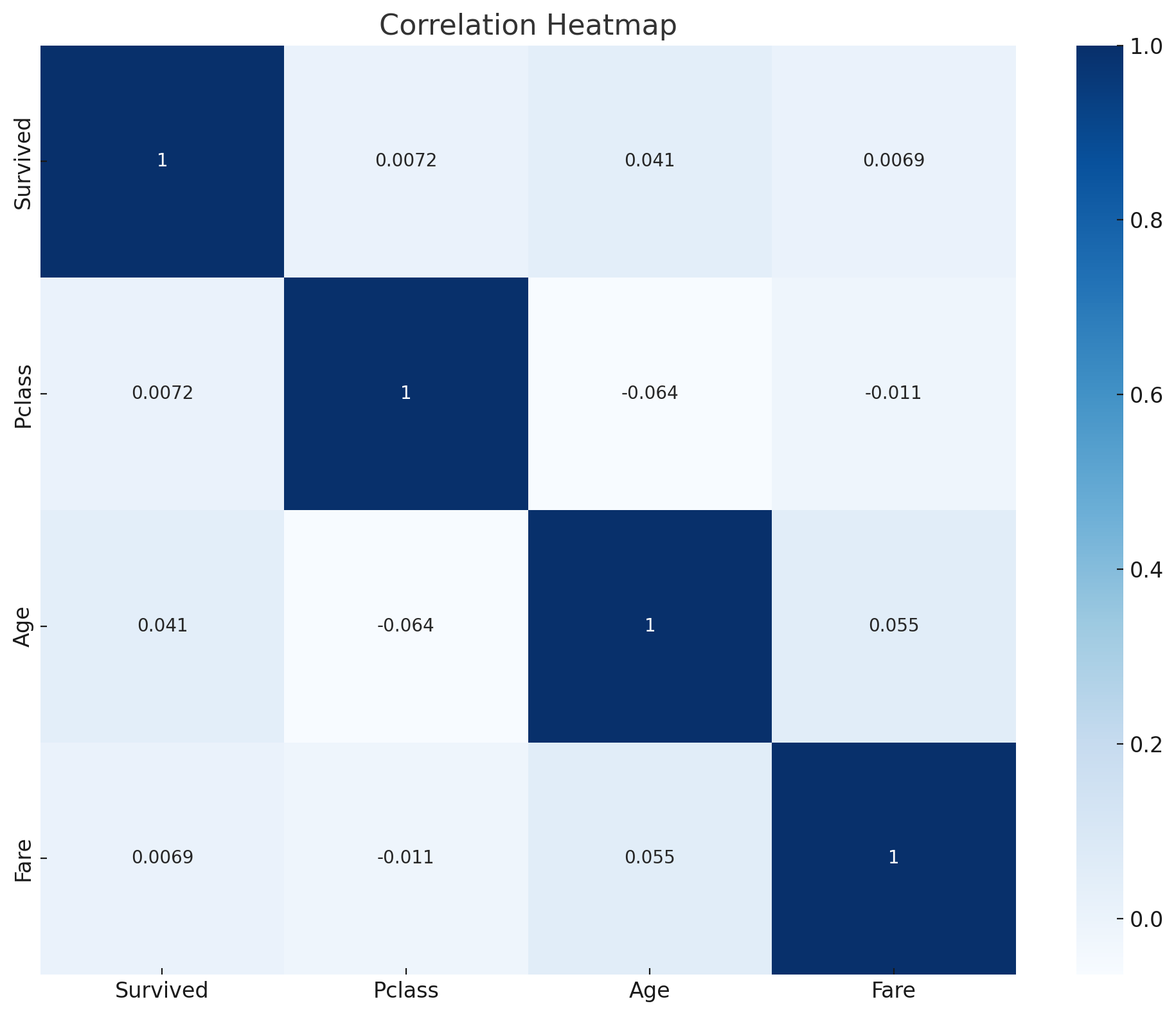
### 

### 

### 

### 

### **d. Correlation Heatmap**

**

**Observation:**

* Positive correlation between Fare and Survival.
* Negative correlation between Pclass and Survival.

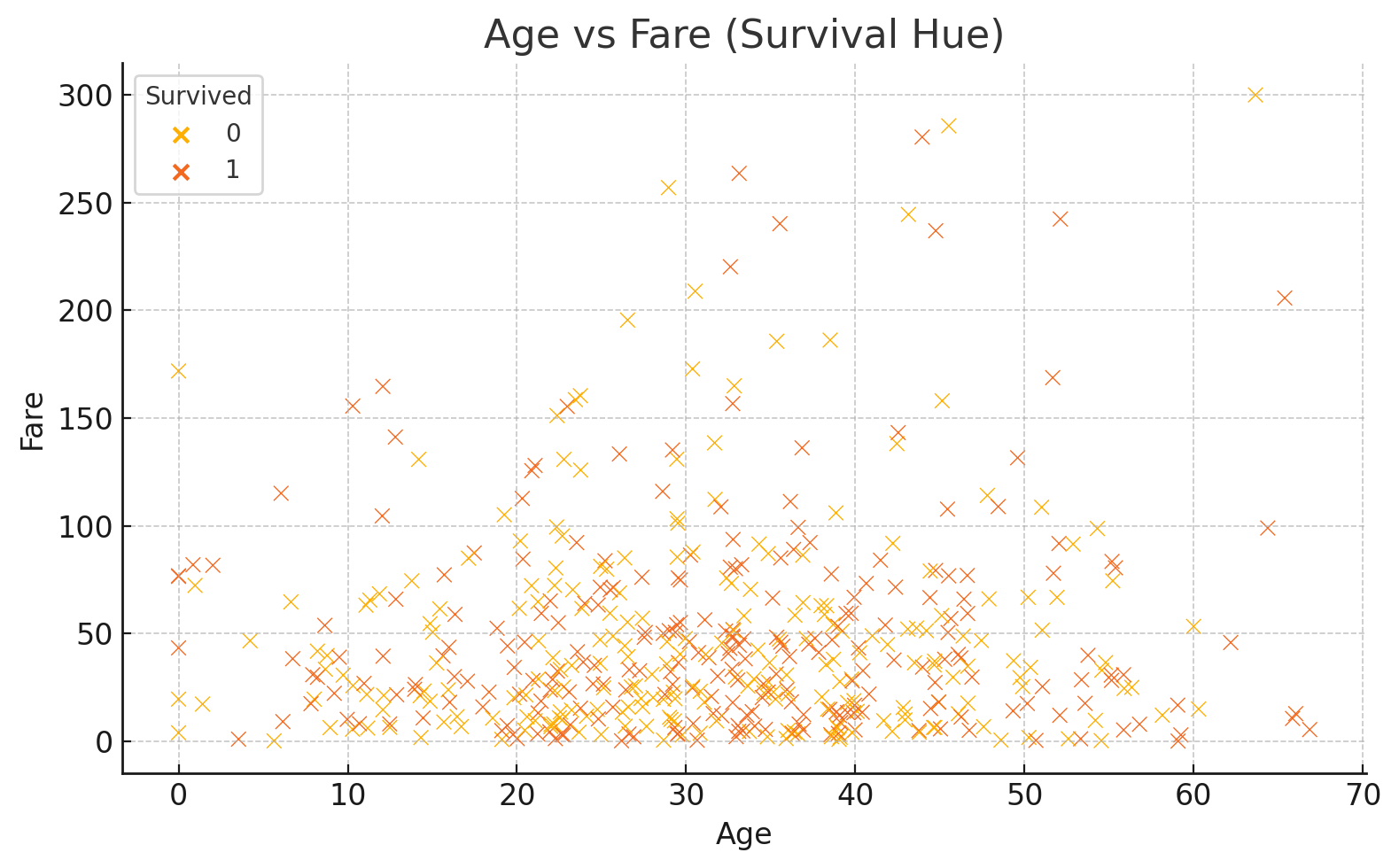
### 

### 

### 

### 

### **e. Scatterplot (Age vs Fare)**

**

**Observation:**

* Higher fares were paid by a few younger and middle-aged passengers.
* No strong linear relationship observed.

## **4. Key Insights**

* Women and children had higher chances of survival.
* Passengers in 1st class were more likely to survive than those in lower classes.
* Missing values in 'Cabin' are significant and may need imputation.
* Outliers present in Fare should be considered during further modeling.
* Younger age slightly correlated with higher survival rates.

## **5. Conclusion**

Exploratory Data Analysis allowed a better understanding of the Titanic dataset, revealing critical patterns and data issues.  
 These insights will help in feature selection, data cleaning, and building better predictive models.

**Prepared by:** Rishabh Singh  
 **Date:28/04/2025**