

# Titanic Dataset – Exploratory Data Analysis

## Internship Task 5 – Data Analyst

### 1. Introduction

The Titanic dataset contains passenger details such as age, sex, ticket class, fare, and survival outcome.

The objective of this analysis is to perform Exploratory Data Analysis (EDA) to identify survival patterns, trends, and anomalies.

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

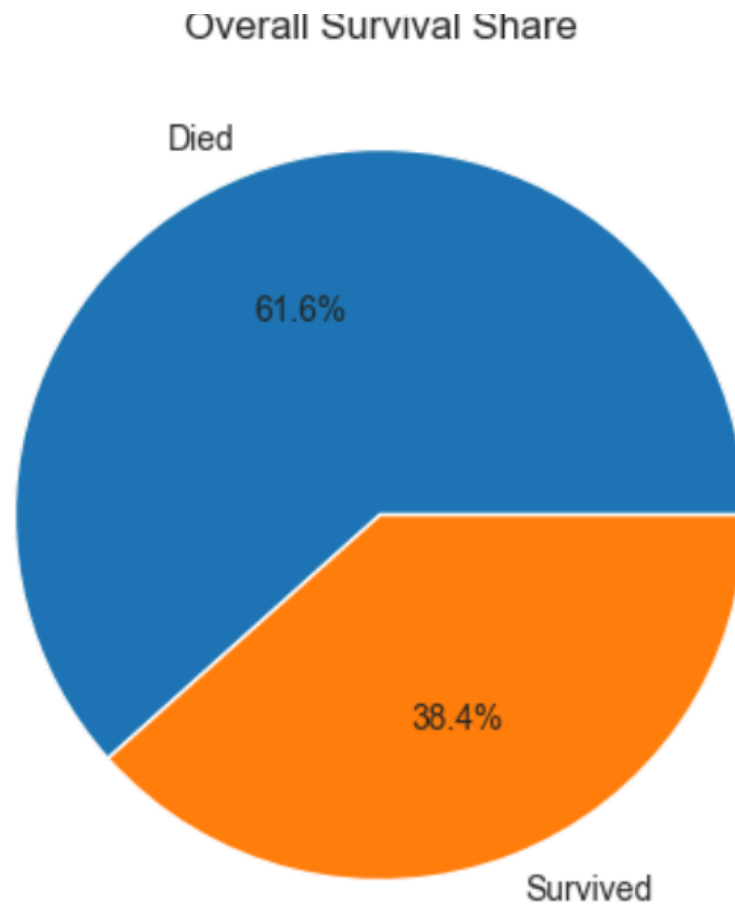
### 2. Dataset Overview

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

#### Findings:

- Dataset has 891 rows and 12 columns.
- Missing values in Age , Cabin , and Embarked .
- Fare has extreme outliers (>500).
- Pclass is numeric but represents categories.

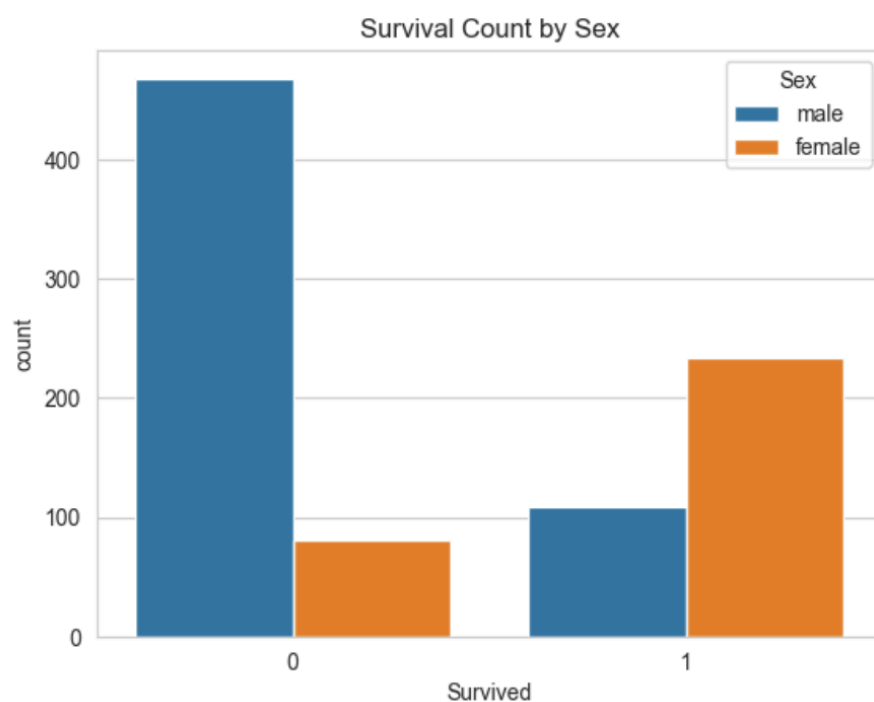
### 3. Survival Distribution



## Findings:

- About 62% of passengers did not survive.
- Only ~38% of passengers survived.

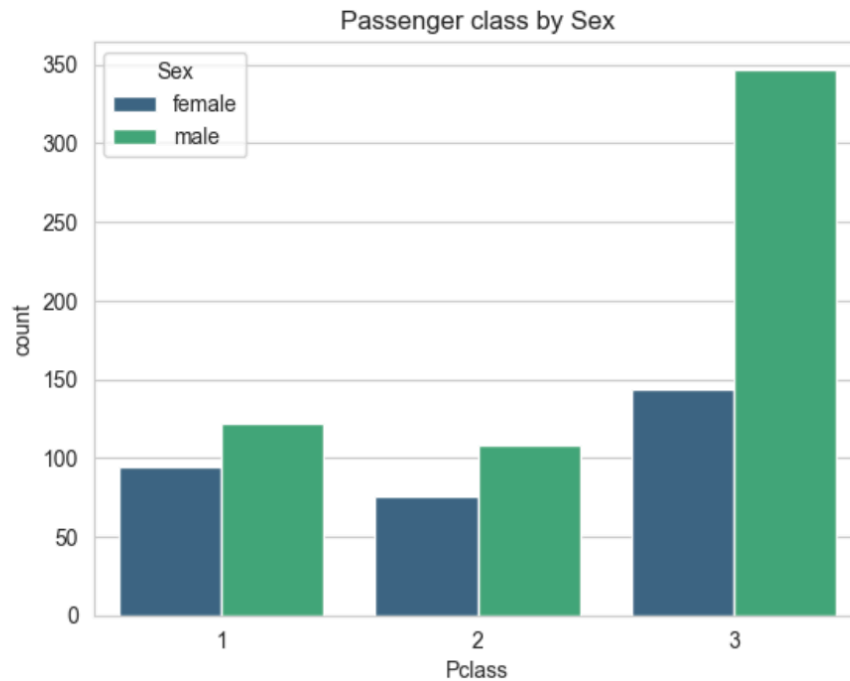
## 4. Survival by Gender



## Findings:

- Female survival rate ~74%, male survival rate ~19%.
  - Gender was a key factor in survival.
- 

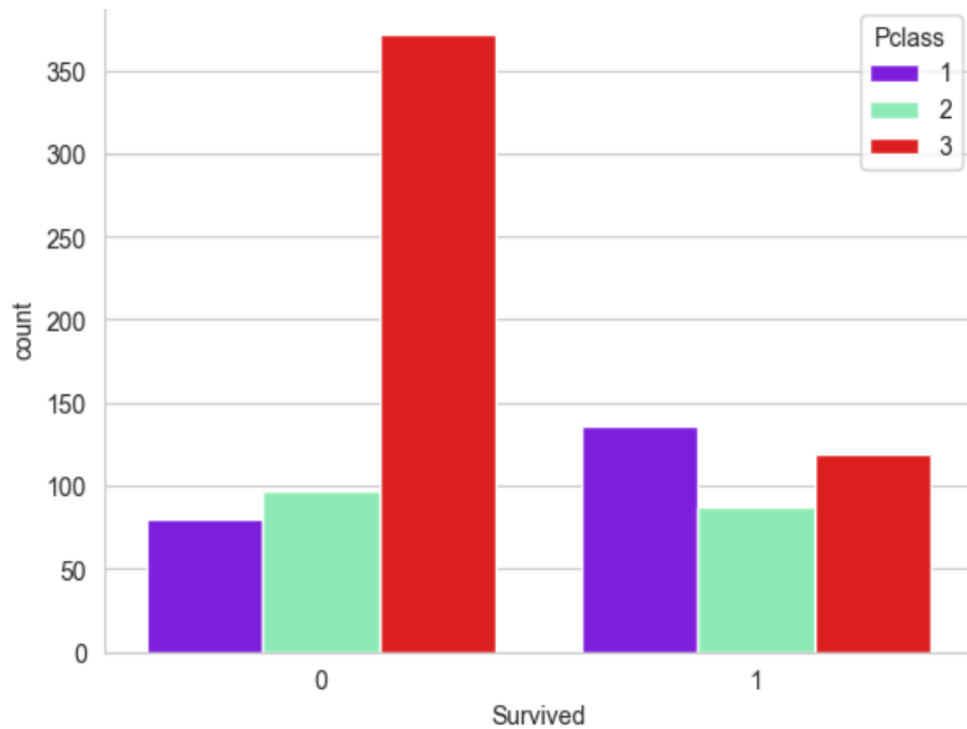
## 5. Passenger Class Distribution



## Findings:

- Most passengers were in 3rd class.
  - 1st and 2nd class had fewer passengers.
- 

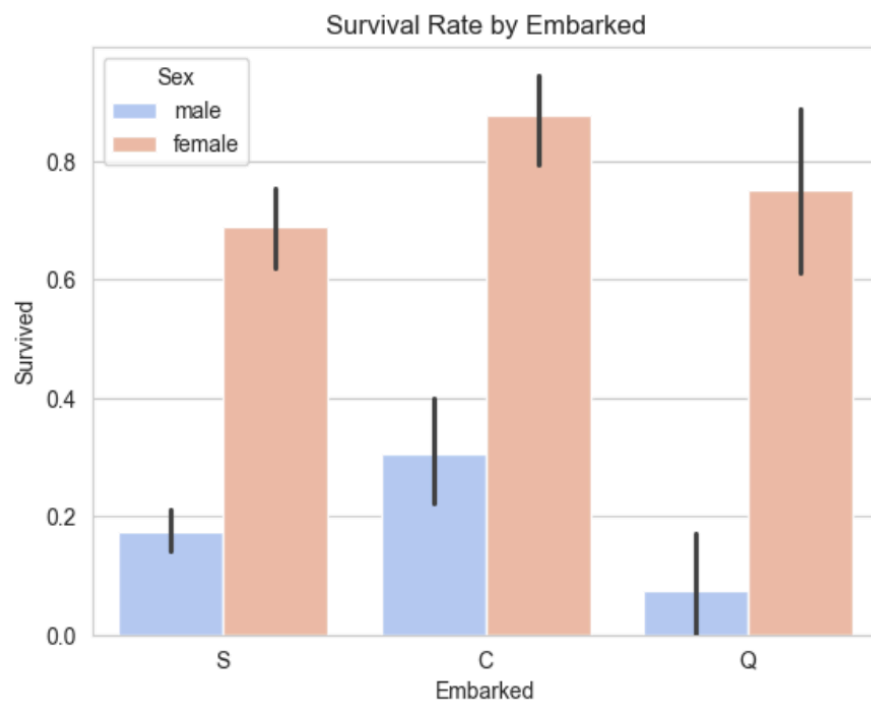
## 6. Survival by Class



### Findings:

- 1st class survival ~63%, 2nd ~47%, 3rd ~24%.
- Higher class passengers survived more.

## 7. Survival by Embarked Port



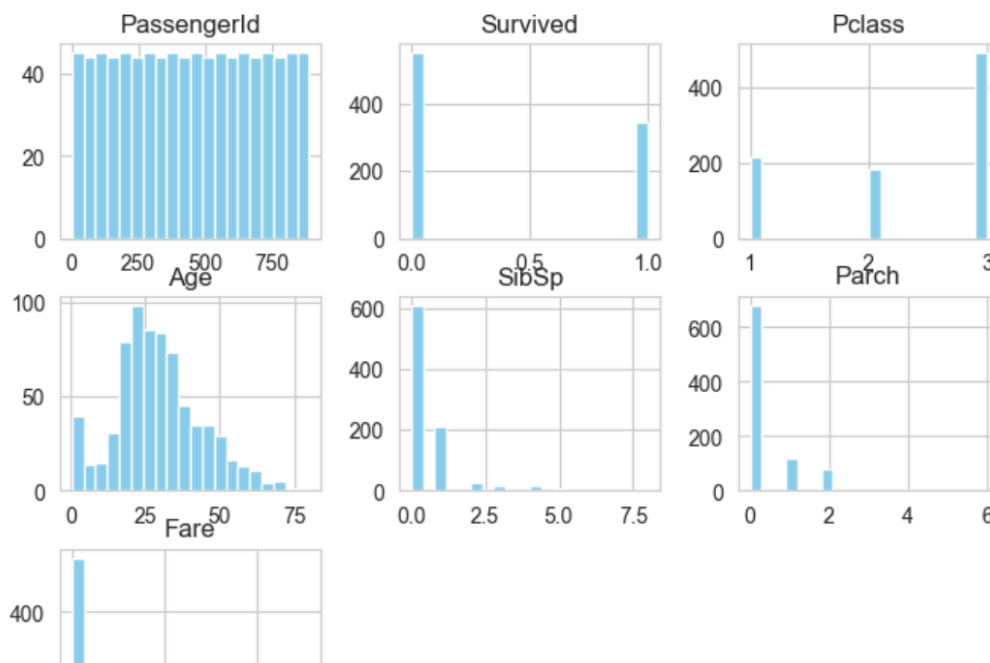
### Findings:

- Cherbourg passengers had the highest survival.

- Southampton passengers had the lowest.

## 9. Histograms of Numeric Features

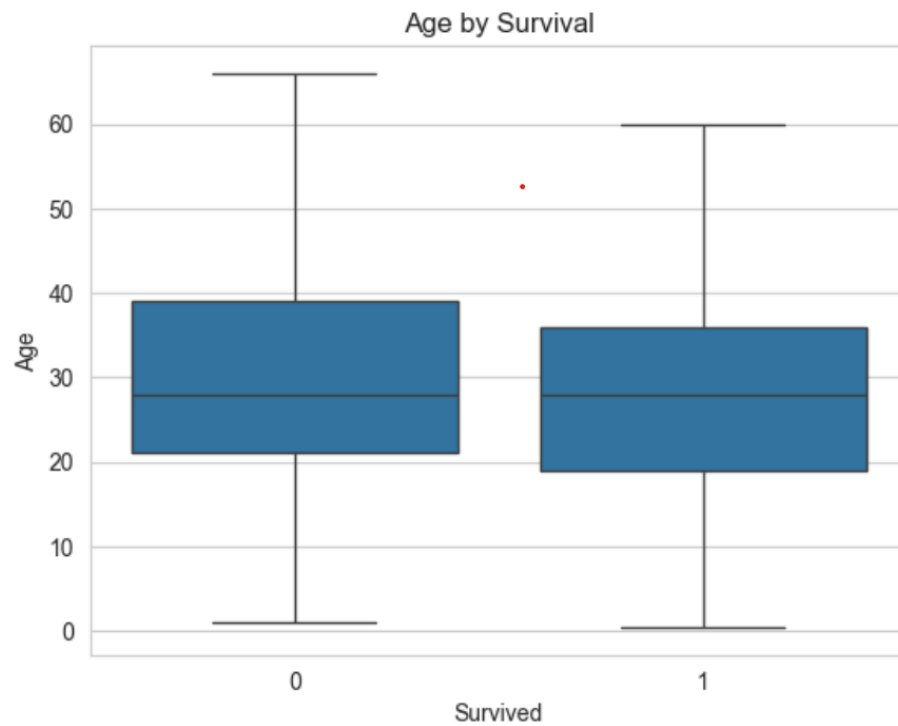
Histograms of Numeric Features



### Findings:

- **Age**: Most passengers were 20–40 years old.
- **Fare**: Positively skewed with extreme outliers.
- **SibSp**: Majority traveled alone (0 siblings/spouses).
- **Parch**: Most had no parents/children aboard.

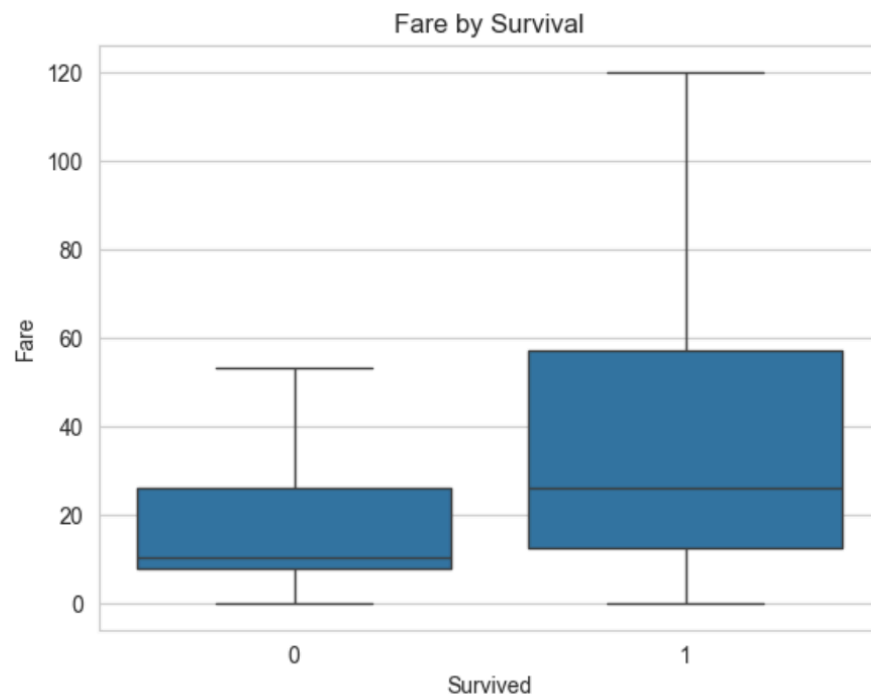
## 9. Age vs Survival



### Findings:

- Children (<10) had higher survival chances.
- Survival decreases with increasing age.

## 10. Fare Distribution vs Survival

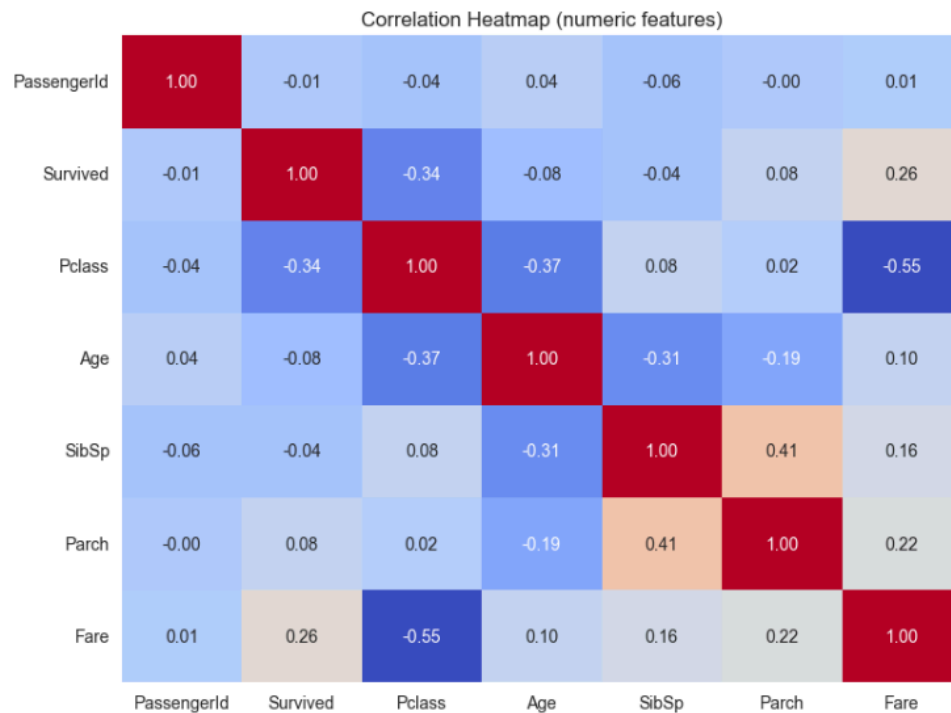


### Findings:

- Survivors generally paid higher fares.

- Wealthier passengers had better survival chances.

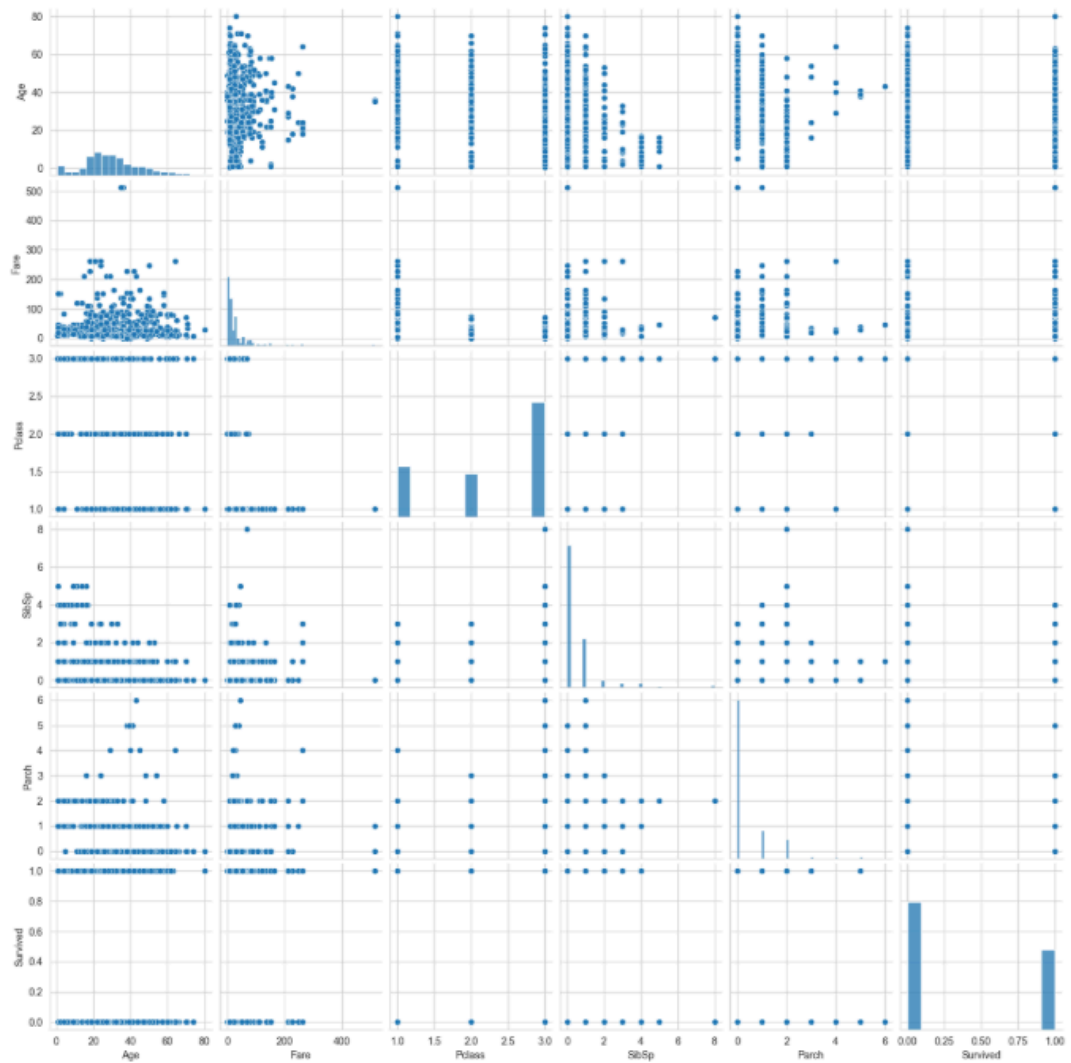
## 11. Correlation Heatmap



### Findings:

- `Survived` moderately correlated with `Sex` and `Pclass`.
- `Fare` negatively correlated with `Pclass`.
- Other features show weak correlation.

## 12. Pairplot

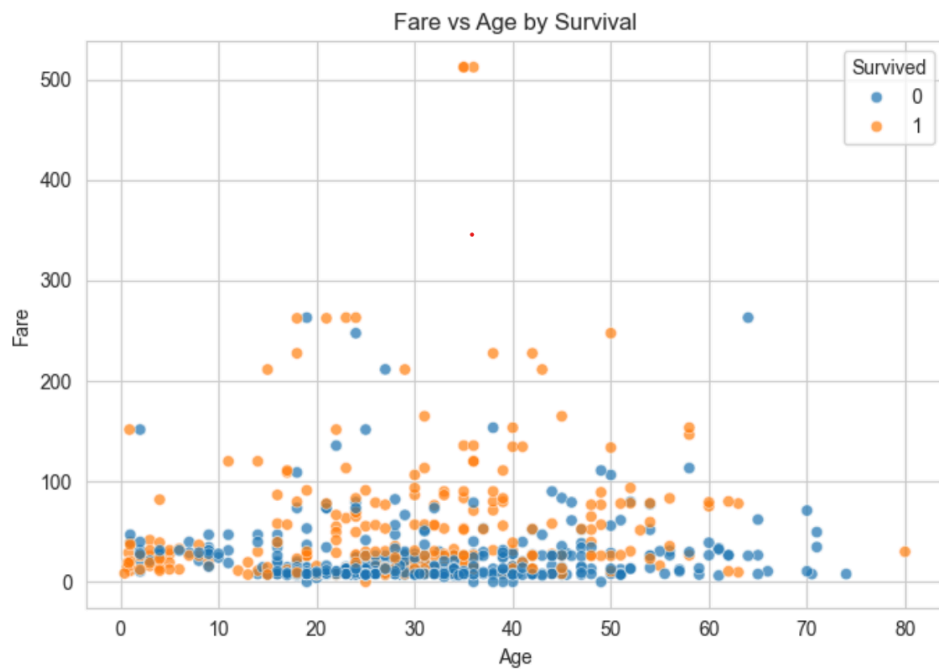


## Findings:

- Survivors cluster in 1st class, higher fares, and female group.
- 3rd class males dominate the non-survivor group.

## 13. Outlier Analysis





## Findings:

- Age is mostly normal, with few elderly outliers.
- Fare has significant extreme outliers (>500).

## 14. Final Summary

### Final Insights:

- Women and children had much higher survival rates.
- 1st class passengers survived significantly more than 2nd and 3rd.
- Cherbourg port passengers had better survival compared to Southampton.
- Survival was influenced strongly by **gender, class, and fare**.
- Data quality issues: missing values in Age, Cabin, and Embarked.

In [ ]: