# Exploratory Data Analysis on Titanic Dataset

Submitted by: Hitesh Suthar

Date: 19-07-2025

## 1. Introduction

The Titanic dataset is a classic machine learning dataset that contains information about the passengers on the Titanic. This report explores the data using statistical analysis and visualizations to uncover insights related to survival rates, demographics, class-wise distribution, and fare patterns.

## 2. Dataset Overview & Statistics

Dataset Shape: (887, 8)
Data Types:
Survived      int64
Pclass        int64
Name          object
Sex           object
Age           float64
SibSp         int64
Parch         int64
Fare          float64
dtype: object
Missing Values:
Survived   0
Pclass     0
Name       0
Sex        0
Age        0
SibSp      0
Parch      0
Fare       0
dtype: int64
Summary Statistics:

|        | Survived   | Pclass     | Name                   | Sex  | Age        |
|--------|------------|------------|------------------------|------|------------|
| count  | 887.000000 | 887.000000 | 887                    | 887  | 887.000000 |
| unique | NaN        | NaN        | 887                    | 2    | NaN        |
| top    | NaN        | NaN        | Mr. Owen Harris Braund | male | NaN        |
| freq   | NaN        | NaN        | 1                      | 573  | NaN        |
| mean   | 0.385569   | 2.305524   | NaN                    | NaN  | 29.471443  |
| std    | 0.487004   | 0.836662   | NaN                    | NaN  | 14.121908  |
| min    | 0.000000   | 1.000000   | NaN                    | NaN  | 0.420000   |
| 25%    | 0.000000   | 2.000000   | NaN                    | NaN  | 20.250000  |
| 50%    | 0.000000   | 3.000000   | NaN                    | NaN  | 28.000000  |
| 75%    | 1.000000   | 3.000000   | NaN                    | NaN  | 38.000000  |
| max    | 1.000000   | 3.000000   | NaN                    | NaN  | 80.000000  |

|        | SibSp      | Parch      | Fare      |
|--------|------------|------------|-----------|
| count  | 887.000000 | 887.000000 | 887.00000 |
| unique | NaN        | NaN        | NaN       |

---

```
top         NaN      NaN      NaN
freq        NaN      NaN      NaN
mean    0.525366  0.383315  32.30542
std     1.104669  0.807466  49.78204
min     0.000000  0.000000   0.00000
25%     0.000000  0.000000   7.92500
50%     0.000000  0.000000  14.45420
75%     1.000000  0.000000  31.13750
max     8.000000  6.000000 512.32920
```

Survival Rate by Pclass:

```
Survived        0          1
Pclass
1       37.037037  62.962963
2       52.717391  47.282609
3       75.564682  24.435318
```

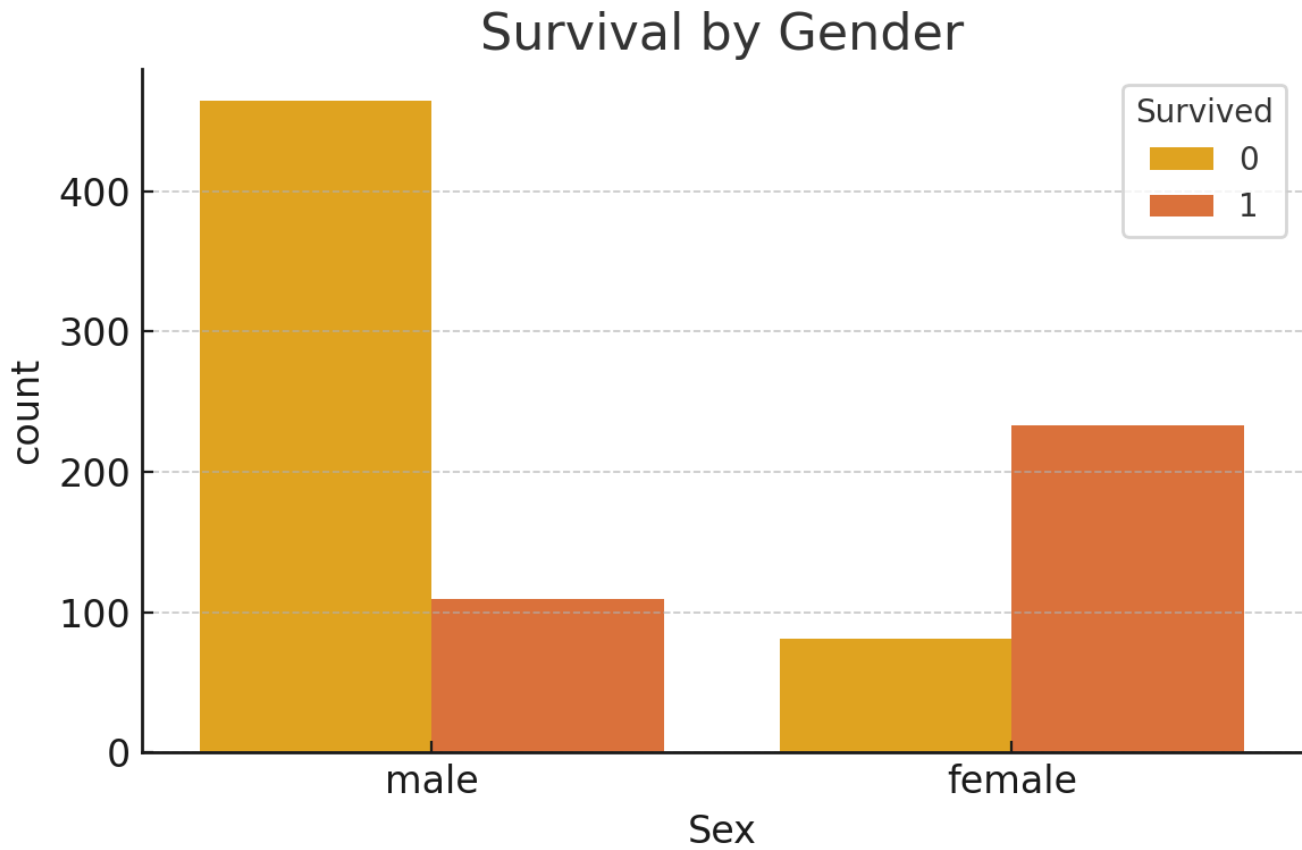Fare Outliers Found: 20

## 3. Age Distribution & Fare by Class
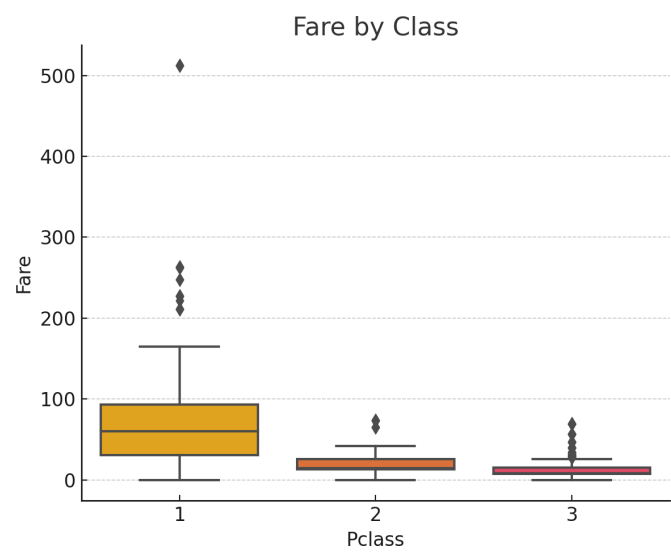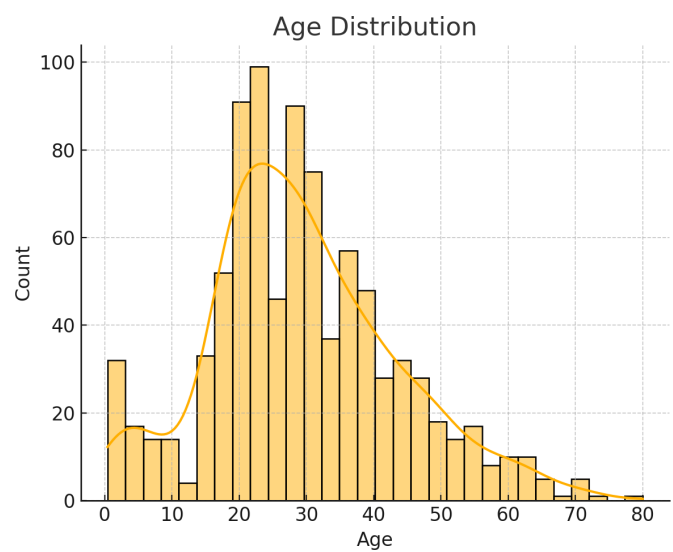


## 4. Survival by Gender

---


Survival by Gender

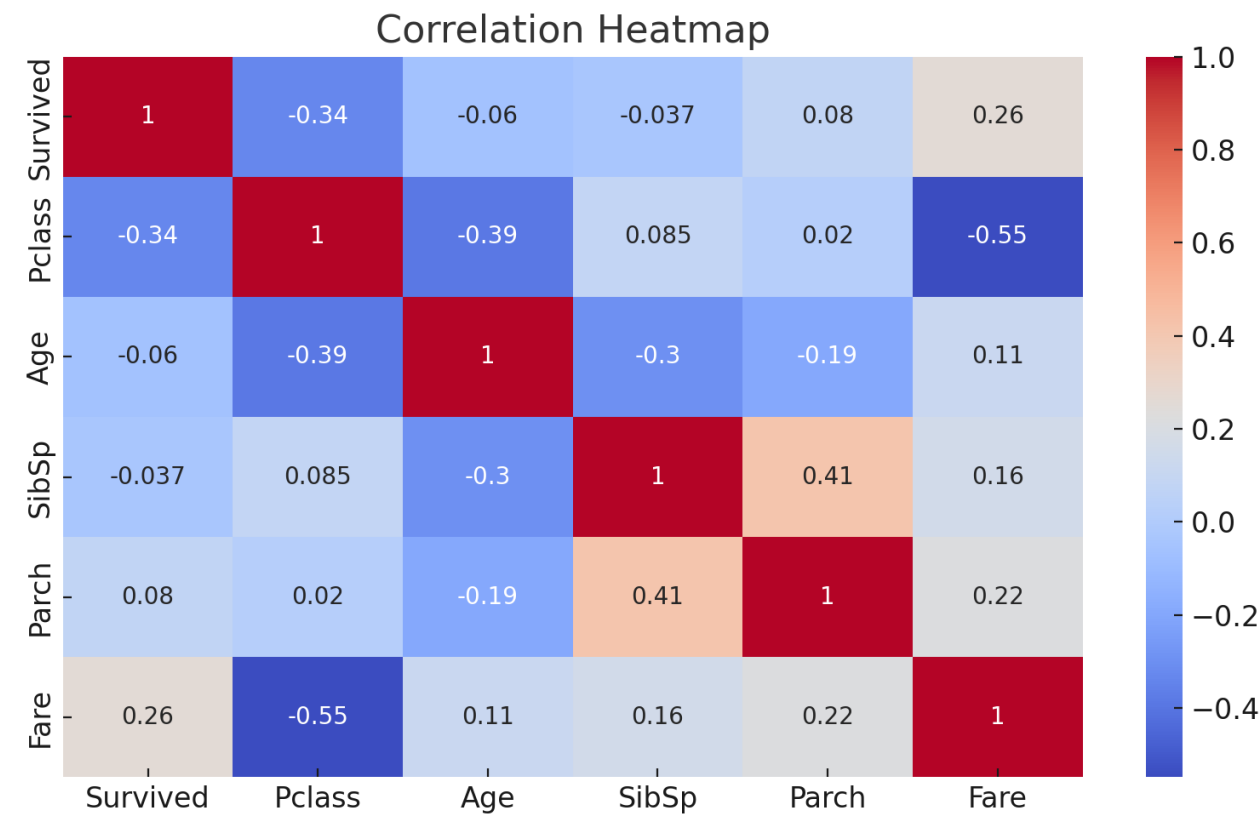## 5. Correlation Heatmap

# Exploratory Data Analysis on Titanic Dataset

Submitted by: Hitesh Suthar

Date: 19-07-2025

---

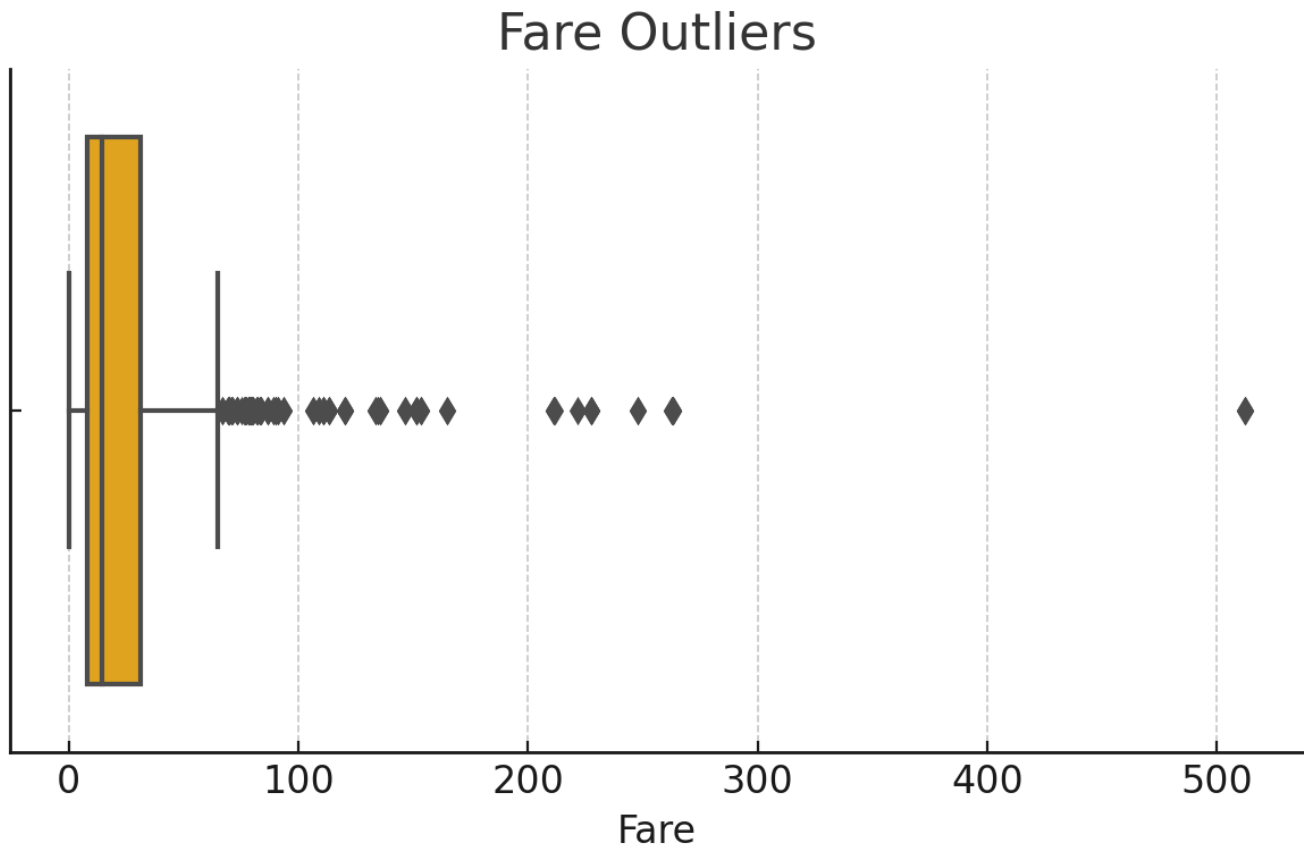## Correlation Heatmap



## 6. Fare Outlier Detection

# Exploratory Data Analysis on Titanic Dataset

Submitted by: Hitesh Suthar

Date: 19-07-2025

---

## Fare Outliers



## 7. Conclusion

The analysis shows a strong correlation between gender, class, and survival. Women and first-class passengers had significantly higher survival rates. There are visible outliers in fare data, mostly due to luxury suites. Future exploration could include feature engineering and machine learning modeling.