

## **Exploratory Data Analysis on Titanic Dataset**

**By: Sania Khalid**

This portfolio project demonstrates exploratory data analysis using a Titanic-like dataset. We explore survival rates across different variables such as passenger class, gender, and age. The analysis uses Python libraries including Pandas, Seaborn, and Matplotlib to clean, visualize, and interpret data patterns

## 1. Survival Count

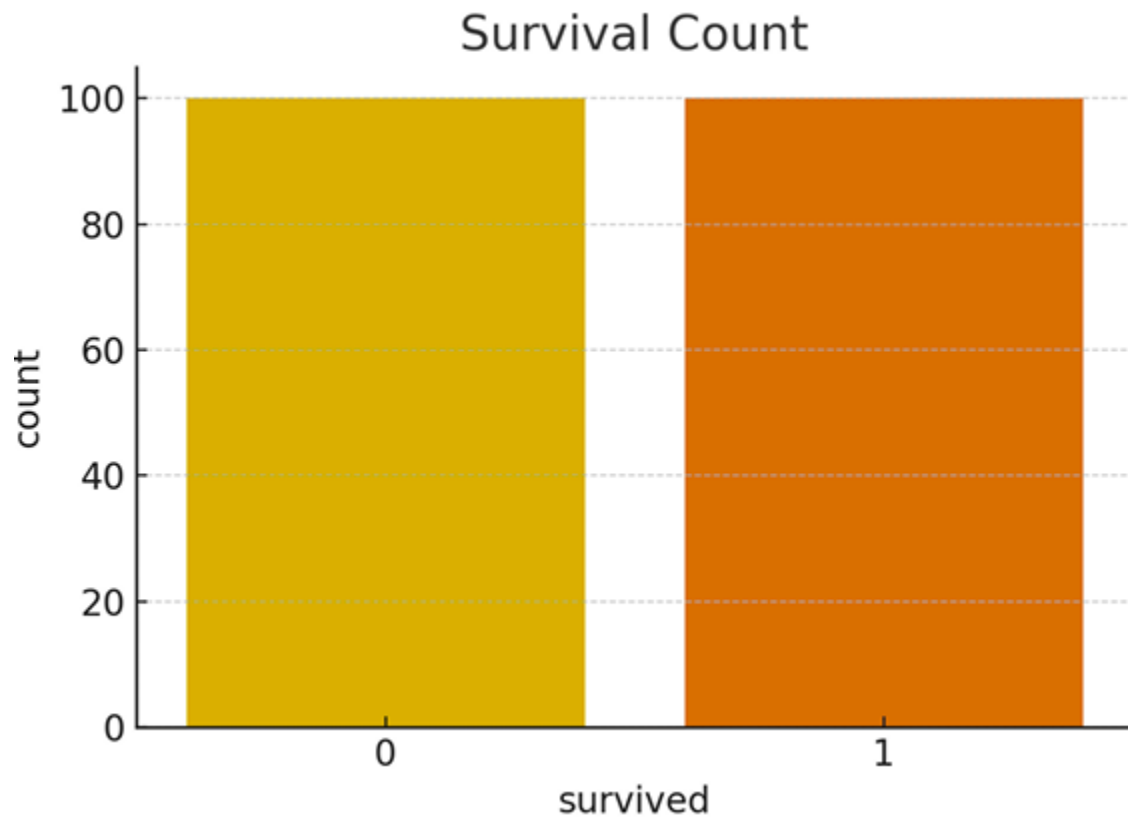


Figure 1.1 shows the survival count

This chart shows the total number of passengers who survived (1) and those who didn't (0). It highlights a class imbalance, with more people not surviving than surviving, which helps us understand the overall survival distribution.

## 2. Survival by Passenger Class

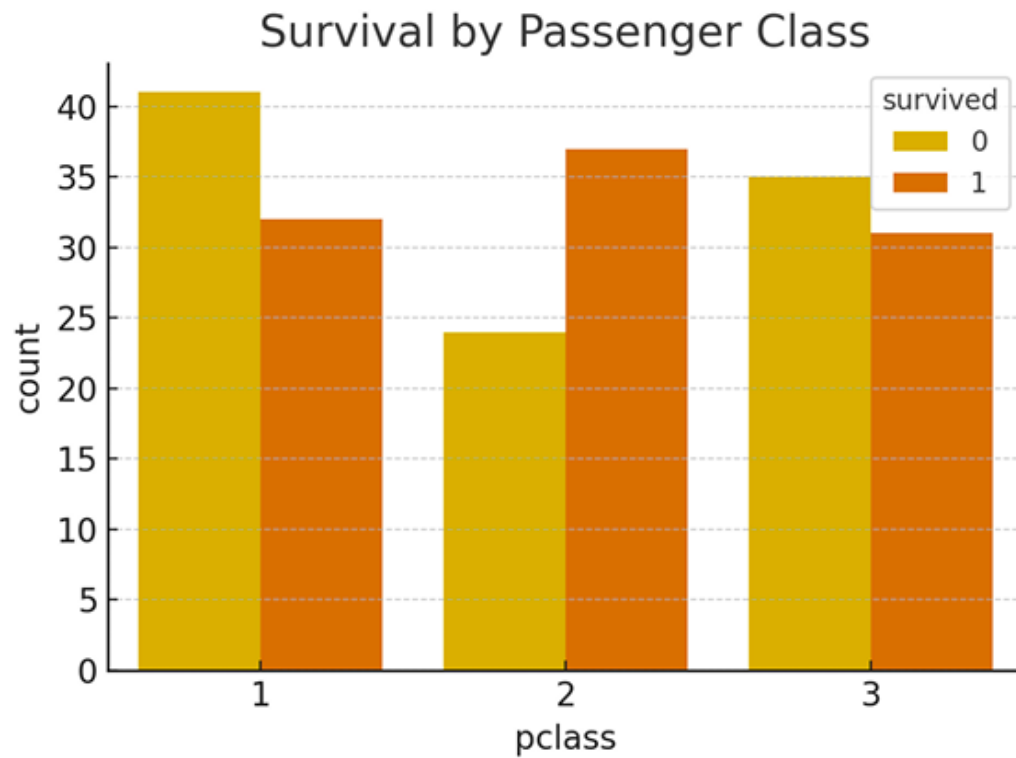


Figure 2.1 shows the survival by passenger class

Passengers in **first class** had a noticeably higher survival rate than those in second and third classes. This suggests that **socioeconomic status** played a role in survival, with wealthier passengers having better access to lifeboats or rescue.

### 3. Survival by Gender

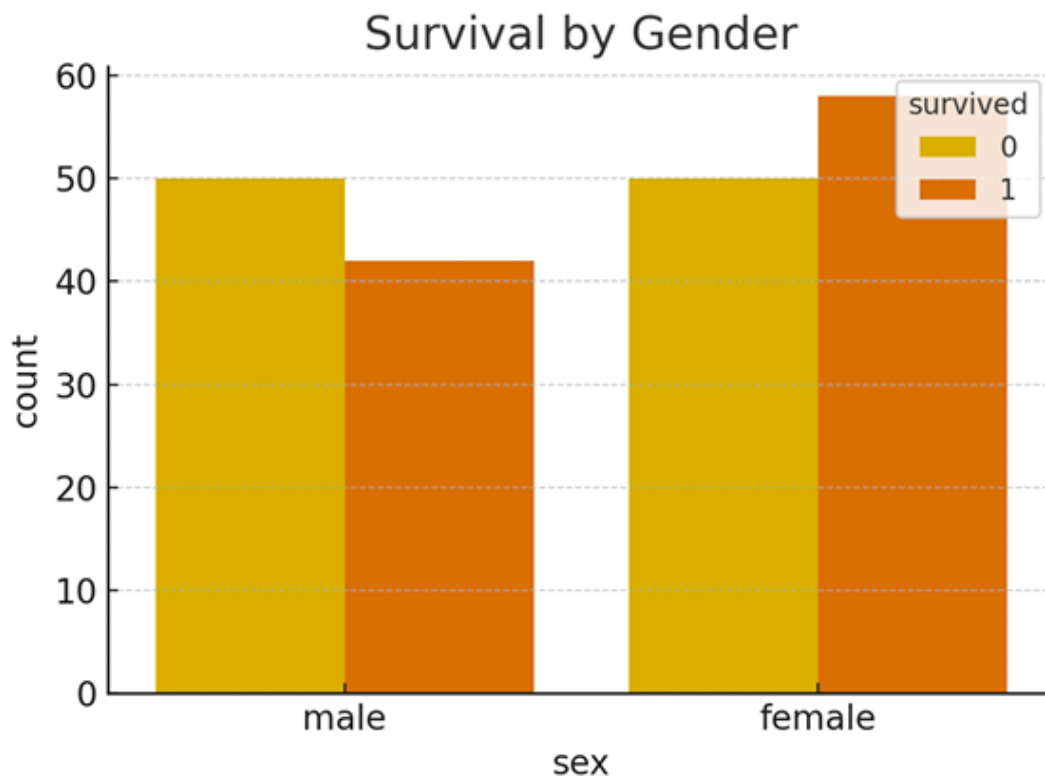


Figure 3.1 shows survival by gender

The survival rate for **females** was significantly higher than for males. This supports historical reports of the “**women and children first**” evacuation policy used during the Titanic disaster.

#### 4. Age Distribution

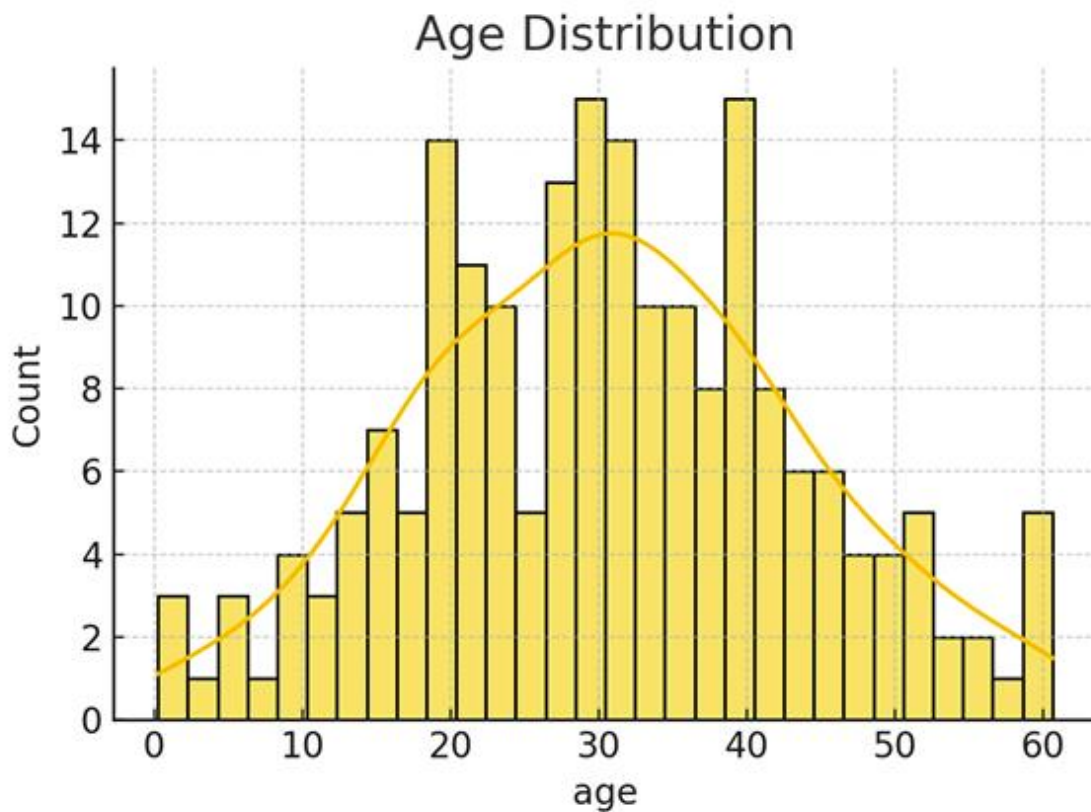


Figure 4.1 shows the age distribution

Most passengers were between **20 and 40 years old**, with fewer children and elderly passengers. This gives a sense of the demographic makeup of the passengers. Outliers may represent very young children or senior citizens.

## 5. Correlation Heatmap

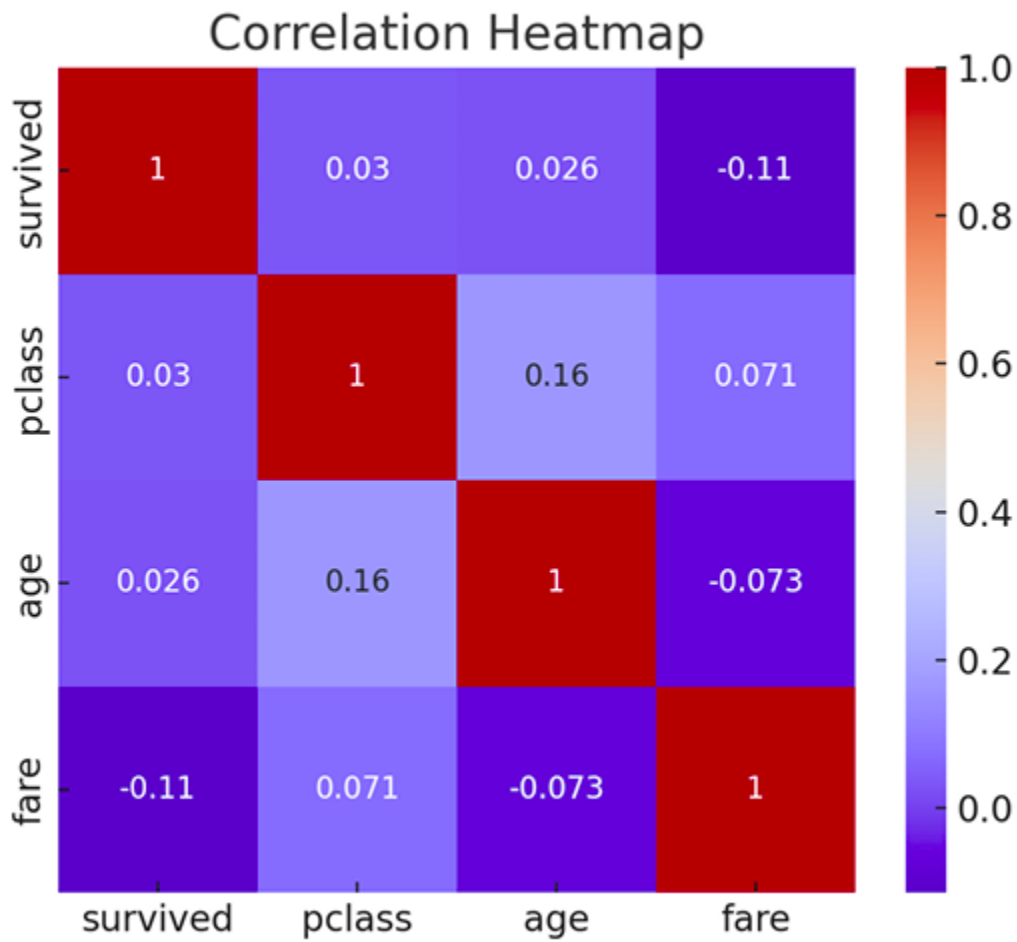


Figure 5.1 show the correlation heatmap

The correlation matrix shows that **fare** has a **positive correlation** with survival, meaning those who paid higher fares had a better chance of surviving. There's a **weak negative correlation** between **age** and survival suggesting slightly lower survival rates among older passengers.