

Exploratory Data Analysis – Titanic Dataset

**Name: Peddi Praharshita
(Data Analyst Internship – Elevate Labs) -
Task_5**

Objective

- To perform exploratory data analysis on the Titanic dataset.
- To identify patterns, trends, and relationships affecting survival.
- To understand feature distributions and correlations.

Dataset Overview

- Total Observations: 891 passengers
- Total Features (after dropping unnecessary columns): 9
- Target Variable: Survived (0 = Did Not Survive, 1 = Survived)
- Dataset Type: Binary Classification Dataset

Key Features:

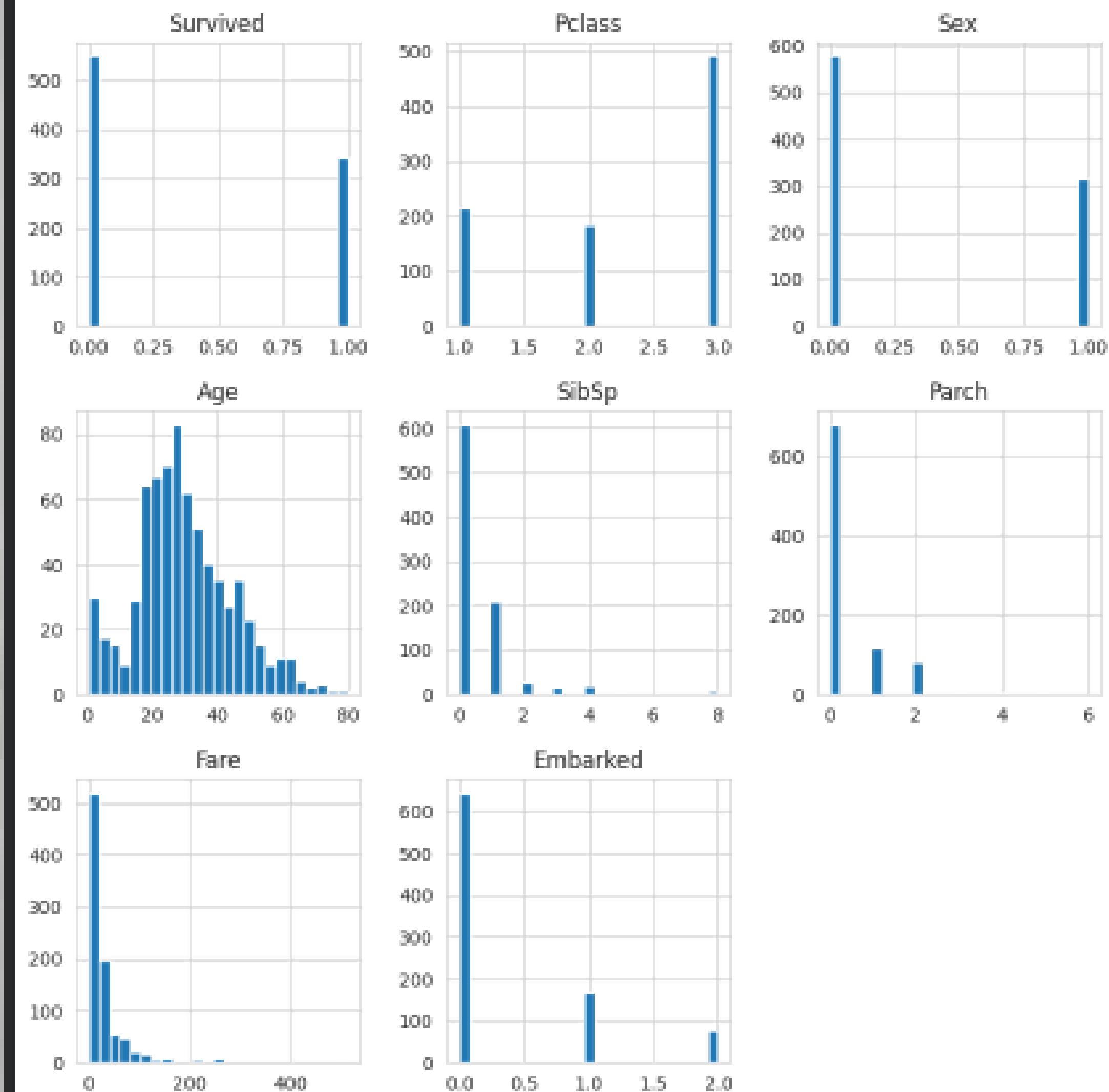
Pclass (Passenger Class), Sex, Age – Age in years, SibSp (Number of siblings/spouses aboard), Parch (Number of parents/children aboard), Fare – Ticket fare, Embarked – Port of boarding, Cabin – Cabin number

Statistical Summary Highlights:

Average Age ≈ 29.7 years, Average Fare ≈ 32.20, Survival Rate ≈ 38.38%

Histograms

- Survival rate is around 38%.
- Majority passengers were from 3rd class and were male.
- Most passengers were aged 20–40 years.
- Most traveled alone (SibSp = 0, Parch = 0).
- Fare is highly positively skewed with noticeable outliers.
- Most passengers embarked from Southampton (S).



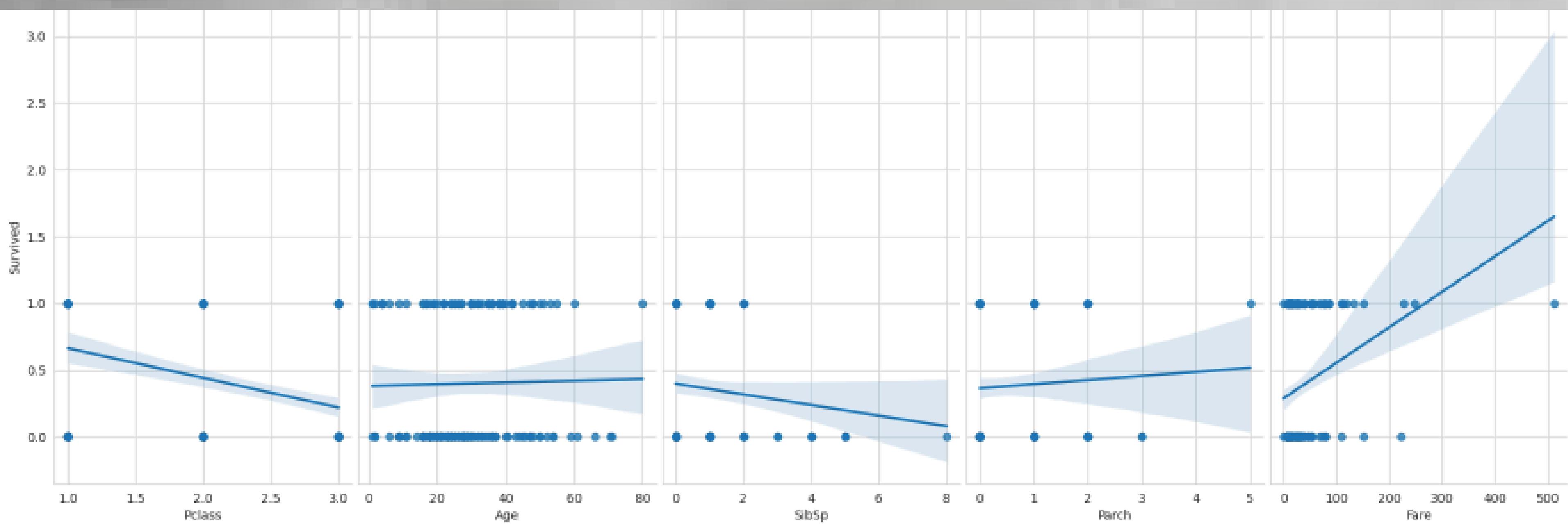
HeatMap

- Sex has strong positive correlation with Survival (~0.54).
- Pclass has moderate negative correlation with Survival (~ -0.34).
- Fare shows positive correlation with Survival (~0.26).
- Pclass and Fare are strongly negatively correlated (~ -0.55).
- SibSp and Parch show moderate positive correlation (~0.41).
- Age has weak correlation with Survival (~ -0.07).

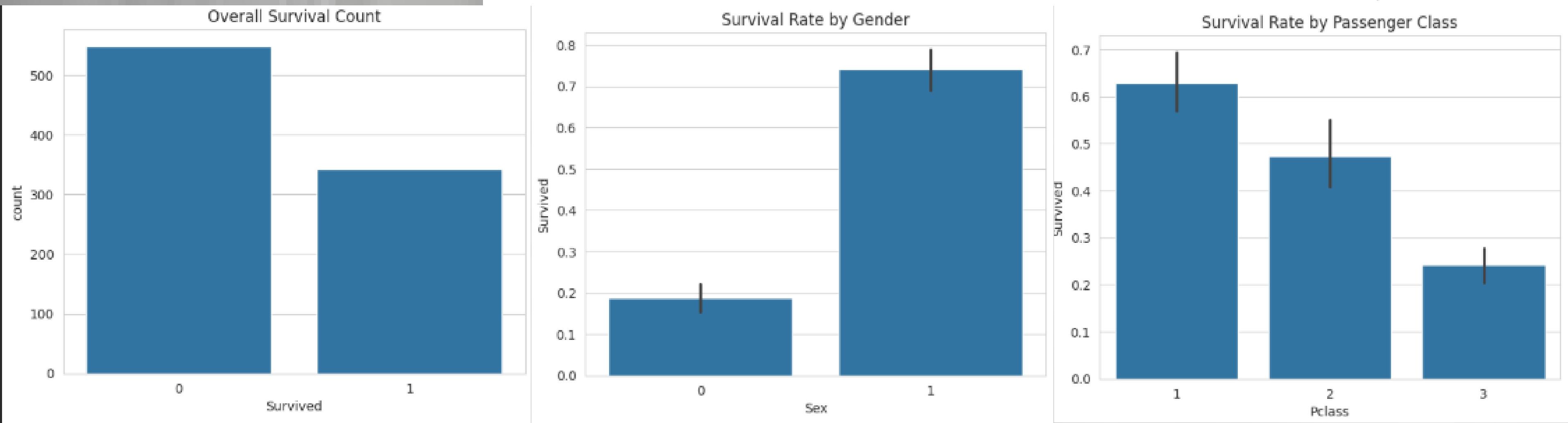
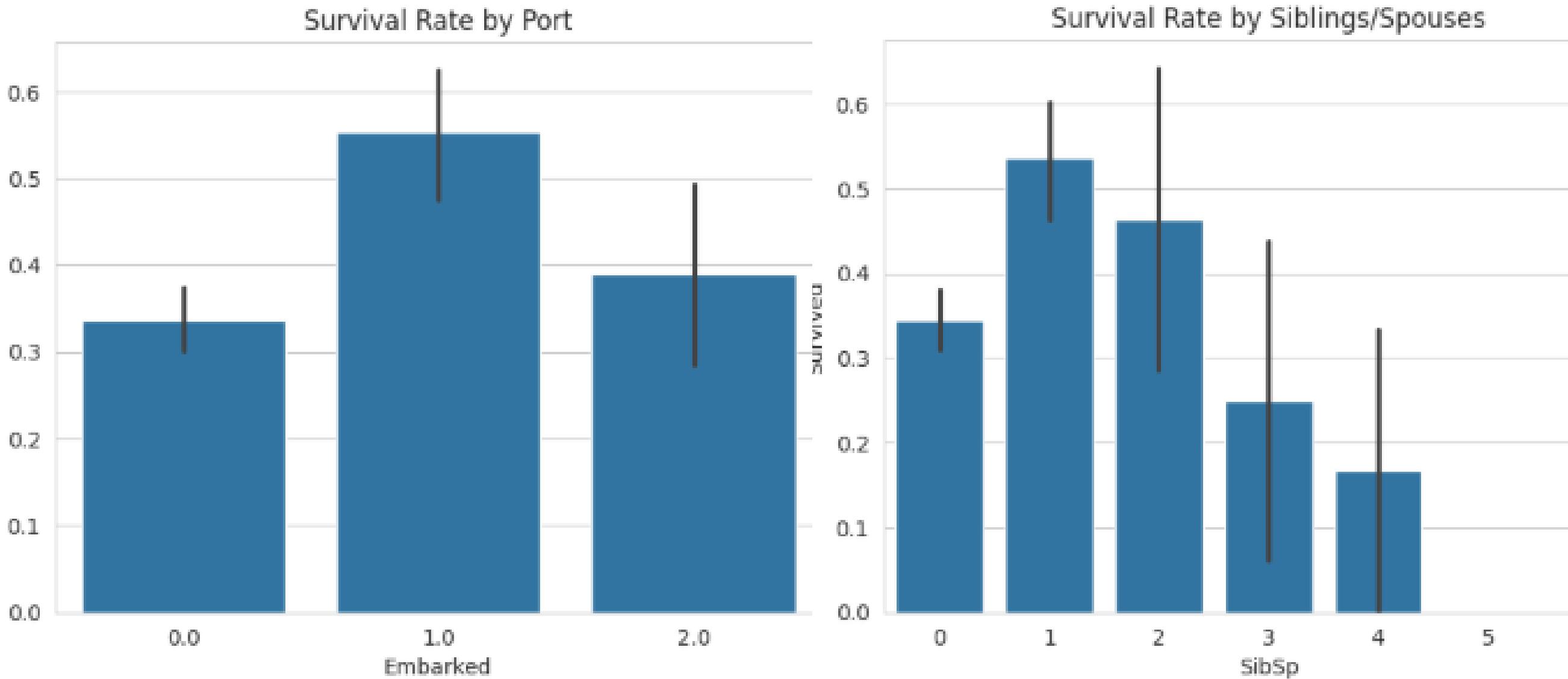


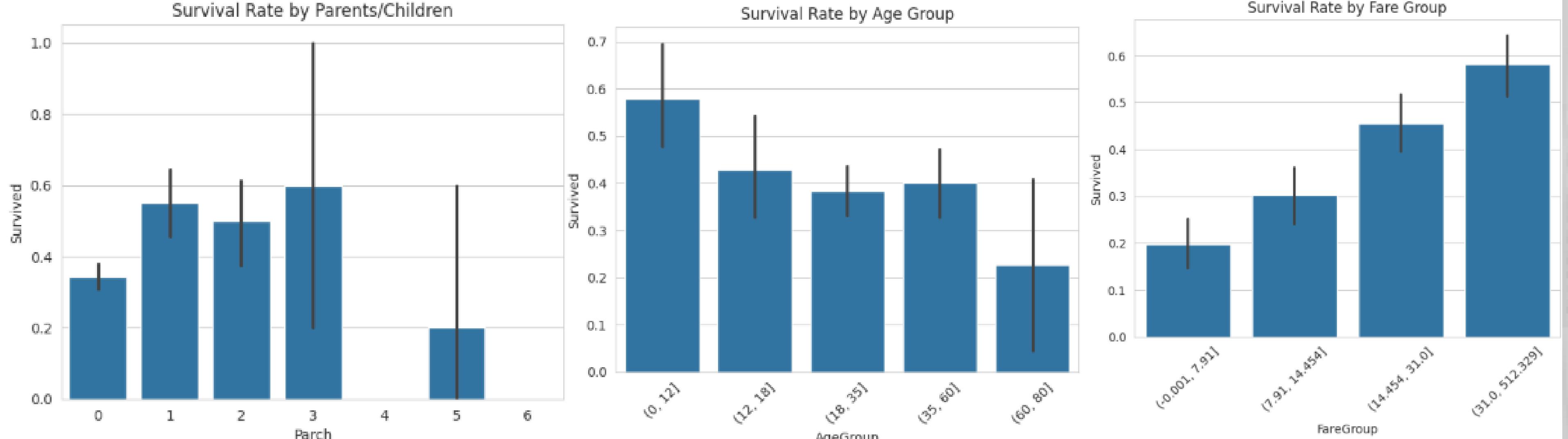
Scatter plot

- Survival is higher for 1st class and high-fare passengers, while age and family size show weak influence on survival.



Survival Count Plot, Target vs Feature section





THANK YOU!