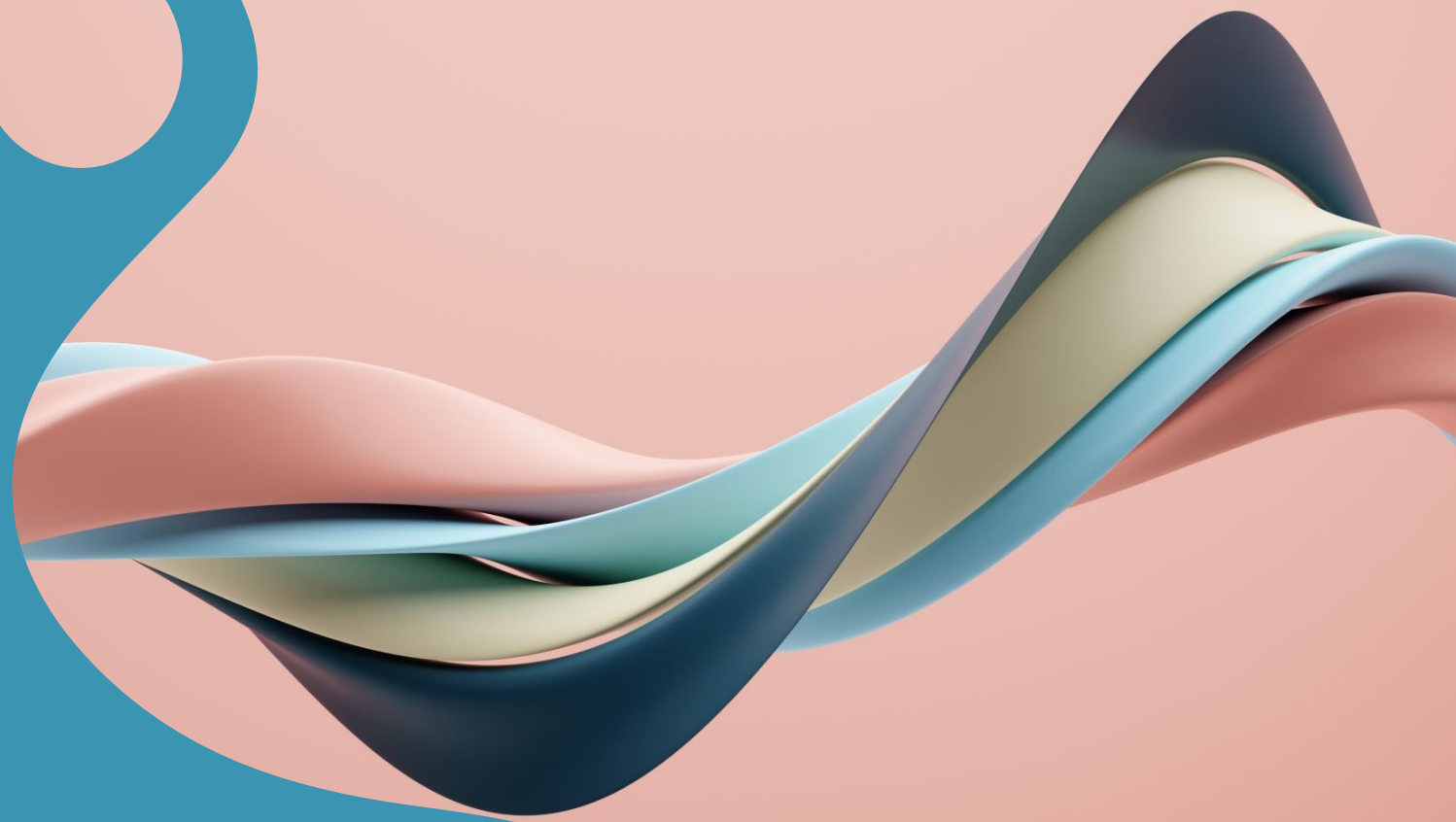
Titanic Data Analysis



By: Sean Bunk

Below is a link to the data set from Kaggle, as well as a link to my GitHub repository so you can see my data cleaning and my python code.

Kaggle link : https://www.kaggle.com/competitions/titanic/data

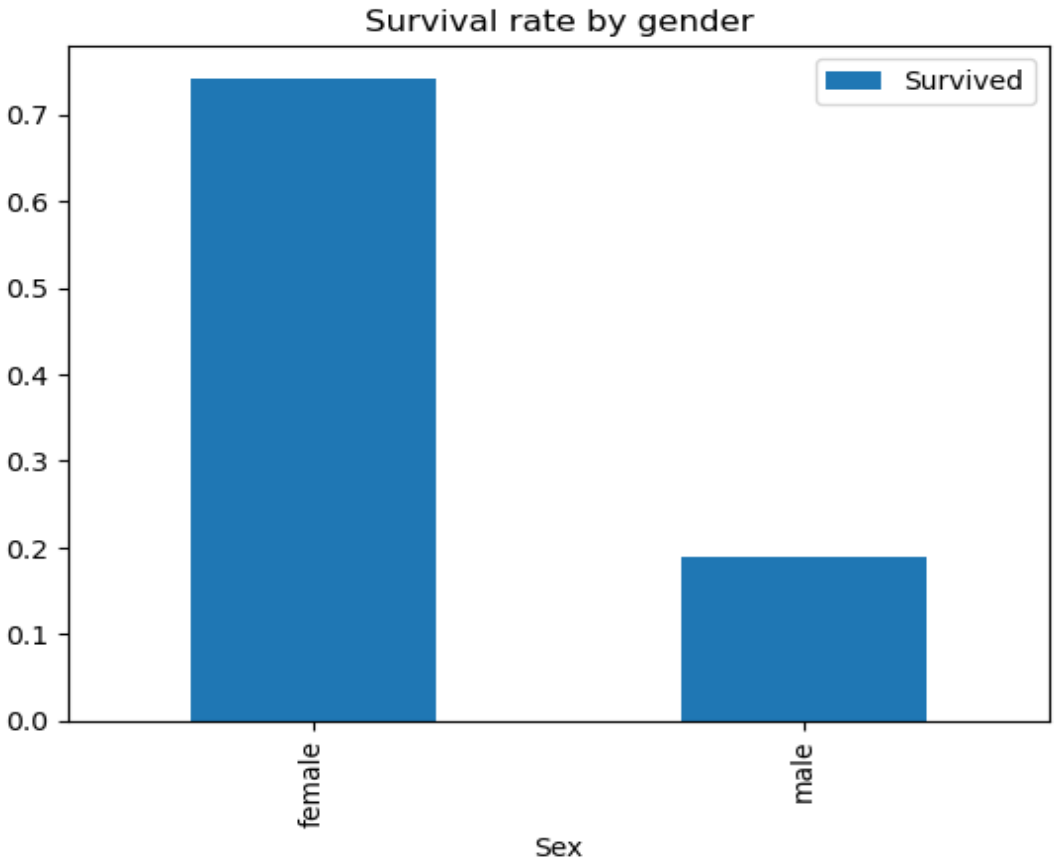
GitHub link: https://github.com/seanbunk/Capstone_1.git

Hypothesis

The survival rate of passengers on the Titanic can be predicted based on factors such as age, sex, passenger class, fare, embarked port, siblings and spouses, and ticket class. I expect that those who have paid for the high-class tickets, women, children, and the elderly would be the first evacuated from the ship. I also expect those who embarked from different ports might have easier access to life-boats than others.

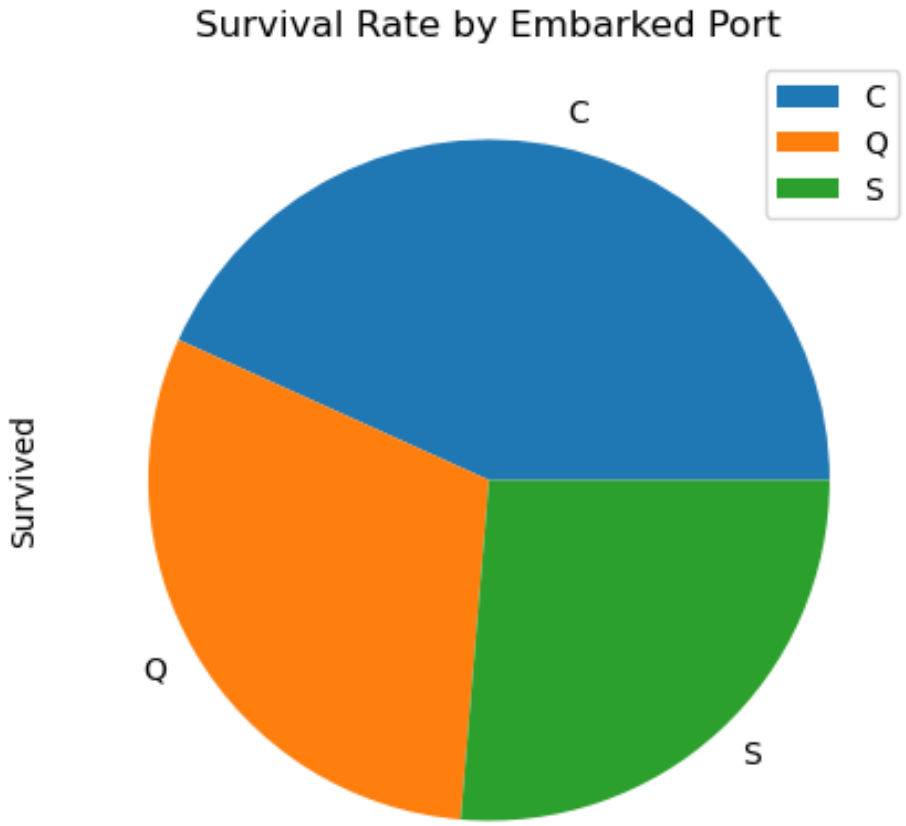
First let's start by looking at the survival rate by gender. Looking at the bar chart below we can see that females have a significantly higher survival rate than males.

In [48]:



Survived	
Sex	
female	0.742038
male	0.188908

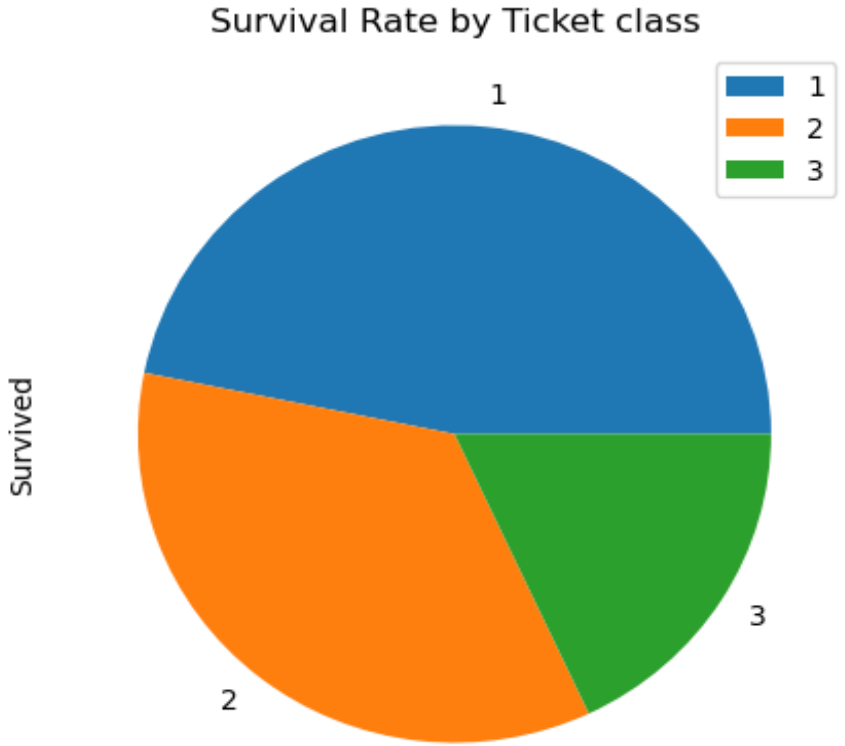
Next, we will look at survival rate by embarked port. The three ports passengers were traveling from are Cherbourg, Queenstown, and Southampton.

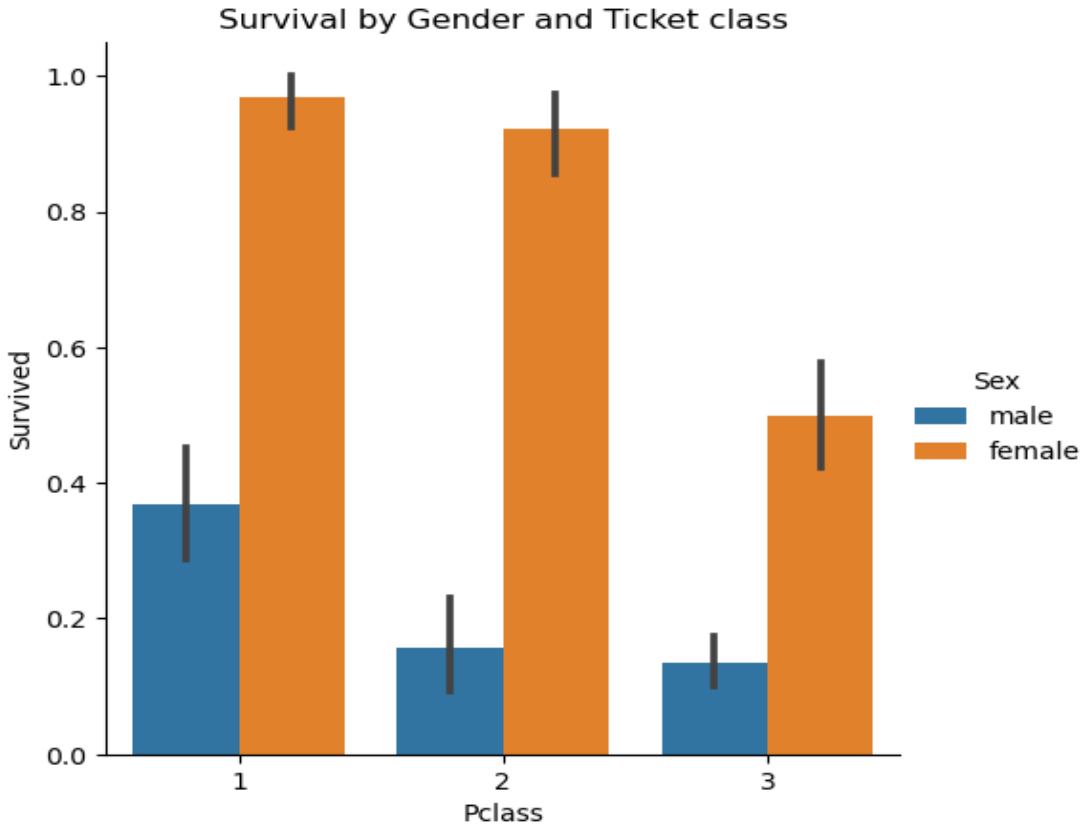


Embarked	Survived
C	0.553571
Q	0.389610
S	0.336957

Looking at the pie chart on the left we can clearly see that those embarked from Cherbourg had a higher survival rate , with Queenstown being a close 2nd and Southampton having the lowest survival rate.

This pie chart represents the survival rate by ticket class. Those who purchased first class tickets have the highest survival rate of all ticket classes.

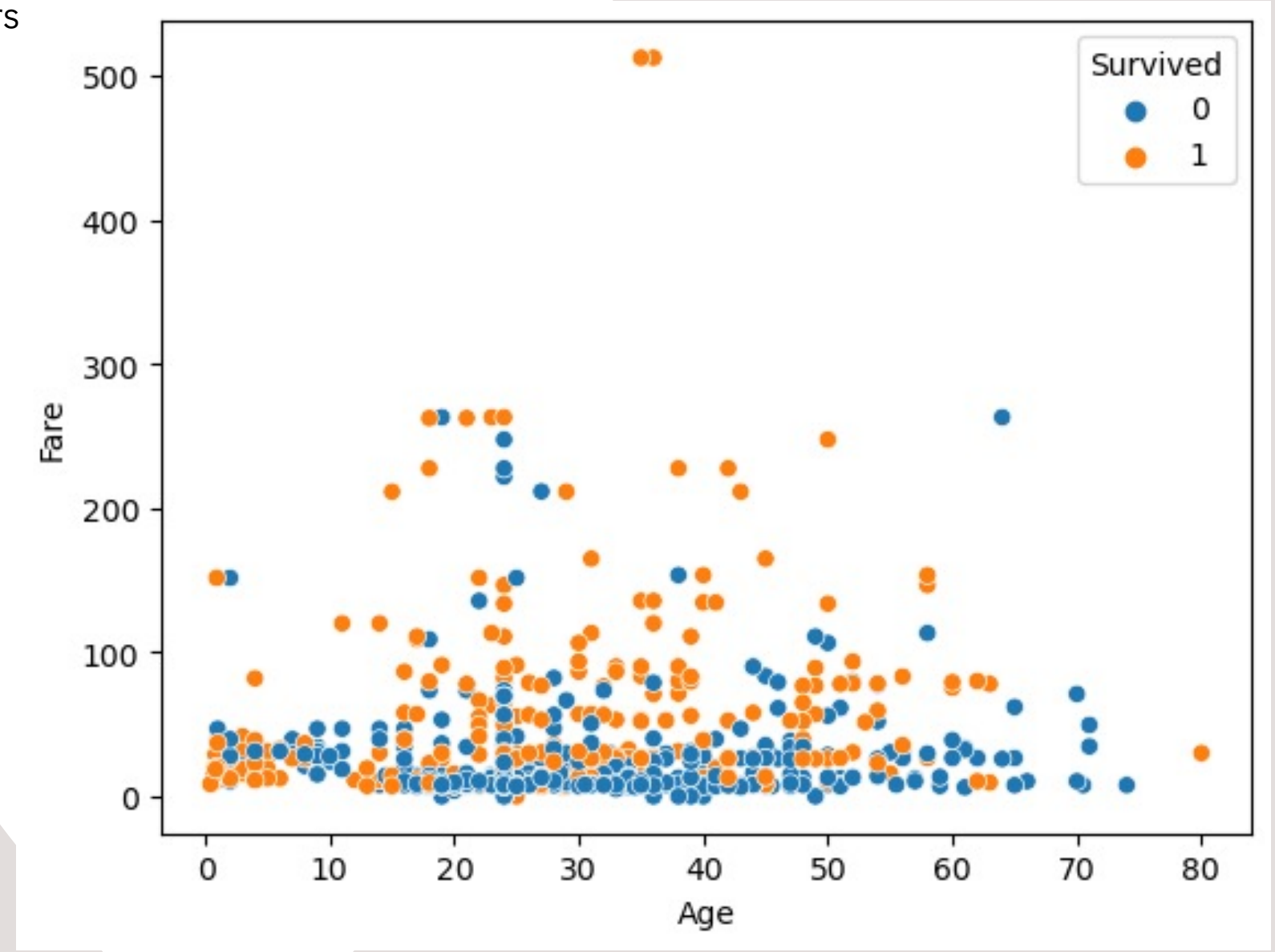


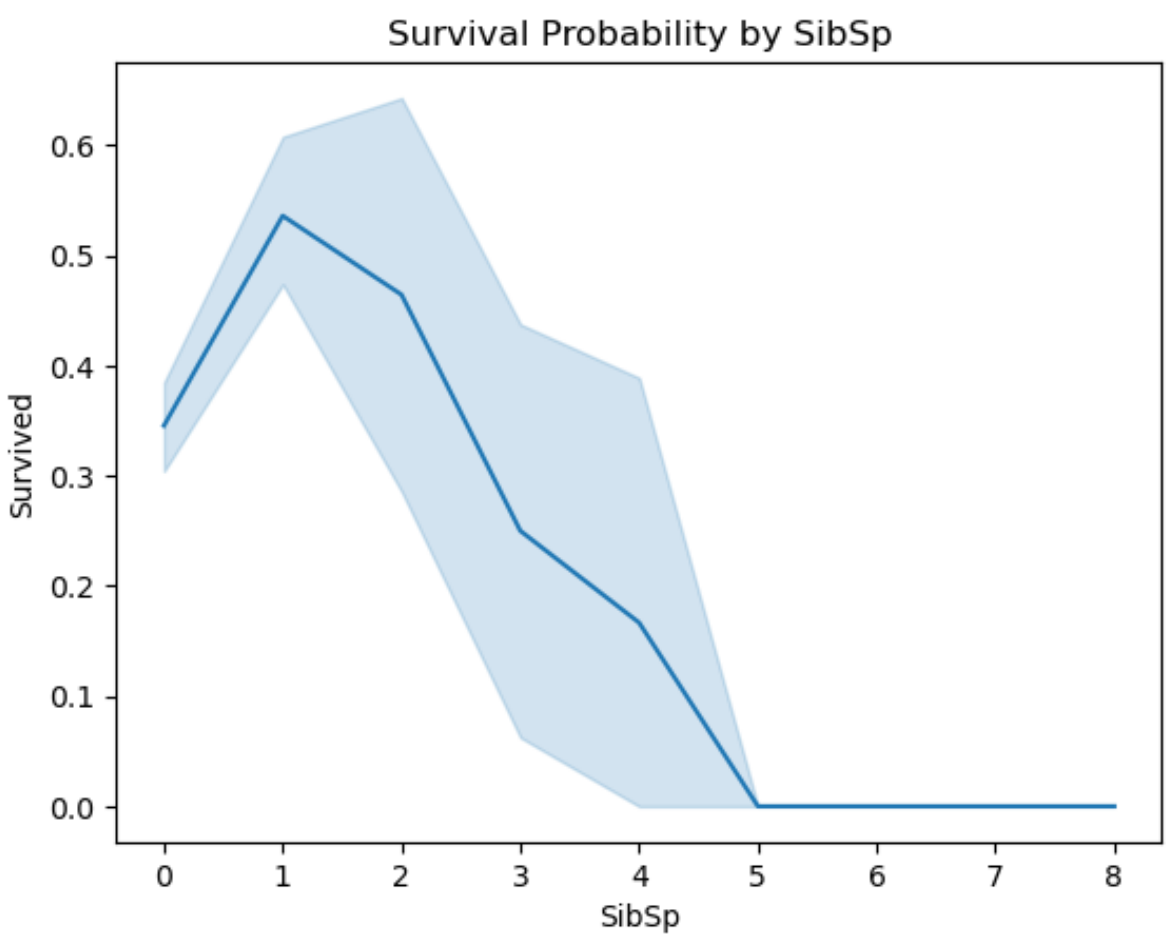


The bar chart shows males in ticket class 1 had the highest survival rate amongst men, and the females in both Ticket class 1 and 2 have the higher survival rate then those in ticket class 3.

The **average** age of the passengers aboard is 24 years old.

The scatterplot on the right shows that passengers who spent more money on their tickets and those who were below the age of 50 years had a higher survival rate.





Lastly, we will look at those who were parents to children as well as those with spouses.

SibSp	Survived
0	0.345395
1	0.535885
2	0.464286
3	0.250000
4	0.166667
5	0.000000
8	0.000000

Passengers with 2 or less siblings and or spouses
Have a higher survival rate than those that have 3 or more. Those
With 5 or more have a 0% survival rate.

Conclusion

After analyzing this data set for the Titanic, I came the conclusion that first, females have an extremely higher survival rate than males. Those who embarked from the port of Cherbourg had a significantly higher chance to survive. Men who spent the money for first class tickets had the highest survival rate out of all the ticket classes, as well as the women in class 1 and 2. The average age of passengers was 24 and those who were below the age of 50 were more likely to survive, I was wrong on this in my hypothesis. Lastly, passengers with 2 or less children and or spouses had the highest survival rate then those who had 3 or more.