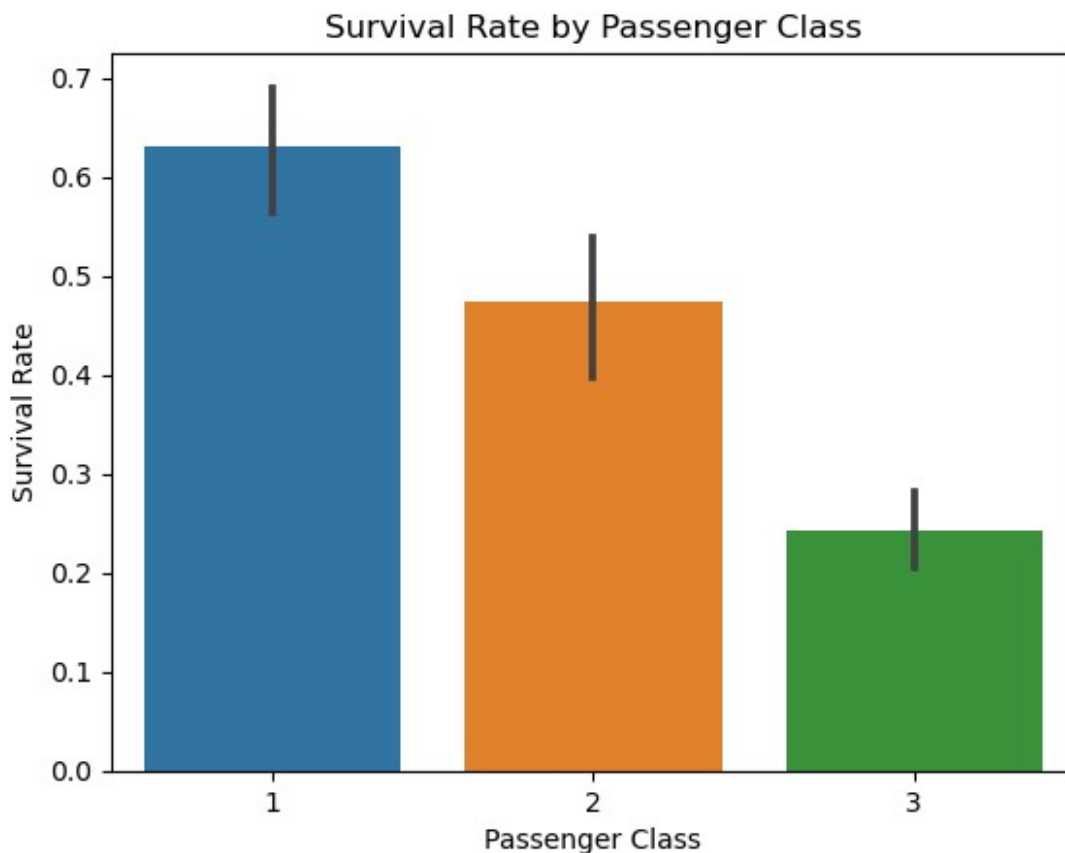


```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

tit = pd.read_csv('titanic.csv')
```

Hypothesis 1: Determine if the survival rate is associated with the class of the passenger.

```
sns.barplot(x='Pclass', y='Survived', data=tit)
plt.title('Survival Rate by Passenger Class')
plt.xlabel('Passenger Class')
plt.ylabel('Survival Rate')
plt.show()
```

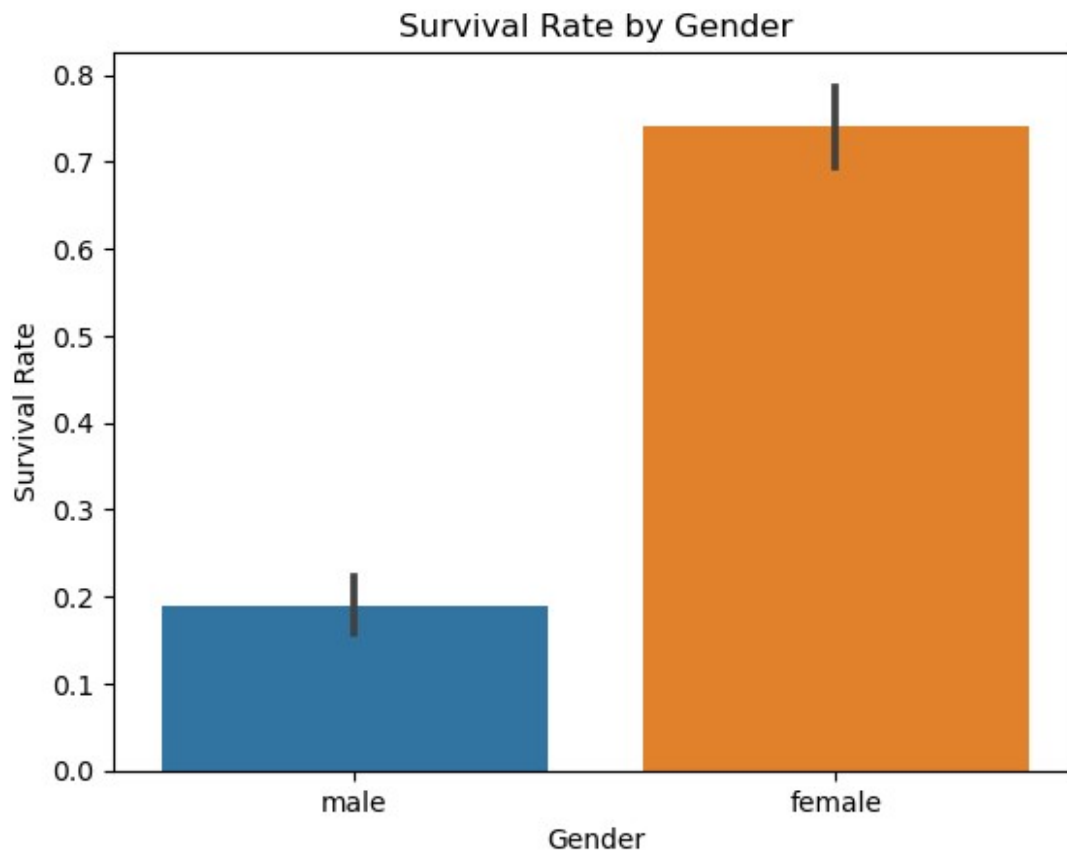


By analysing the graph we can infer that the survival rate and passenger class are related, the higher the passenger class, bigger are the passenger's survival rate

Hypothesis 2: Determine if the survival rate is associated with gender.

```
sns.barplot(x='Sex', y='Survived', data=tit)
plt.title('Survival Rate by Gender')
plt.xlabel('Gender')
```

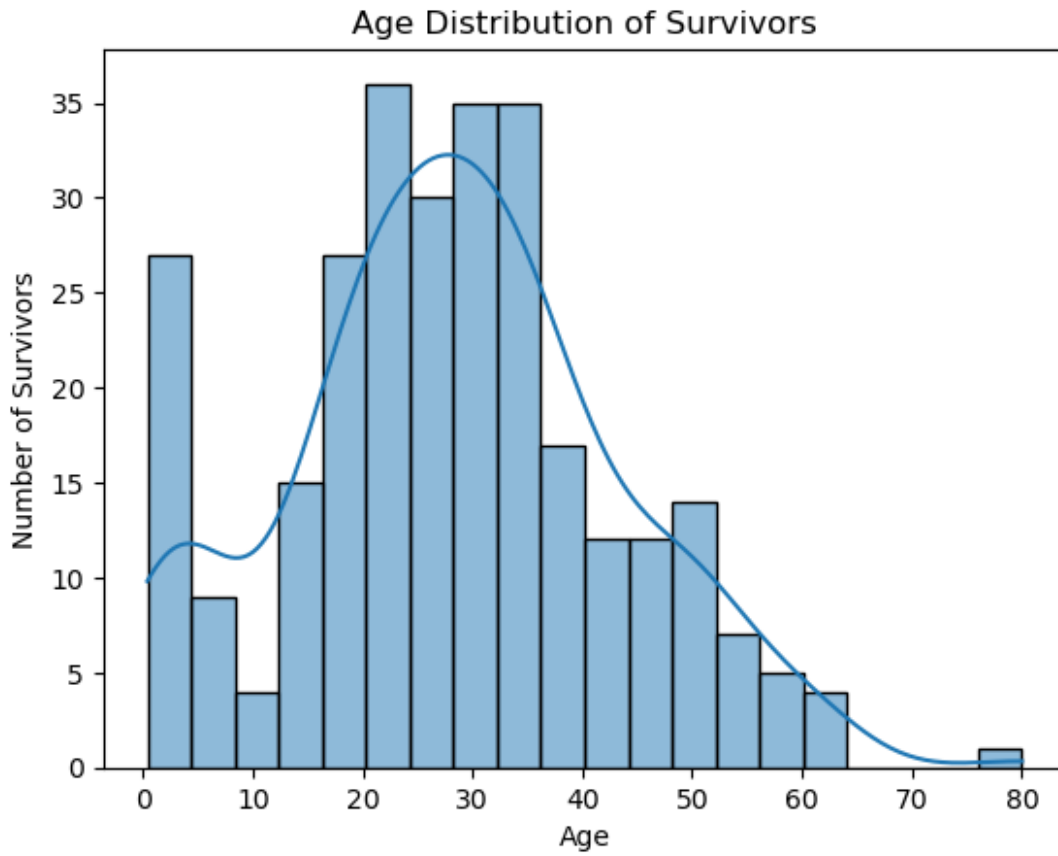
```
plt.ylabel('Survival Rate')  
plt.show()
```



By analysing the graph we can infer that females had a bigger survival rate, almost 400% superior

Hypothesis 3: Determine if the survival rate is associated with age.

```
survived_data = tit[tit['Survived'] == 1]  
sns.histplot(survived_data['Age'], bins=20, kde=True)  
plt.title('Age Distribution of Survivors')  
plt.xlabel('Age')  
plt.ylabel('Number of Survivors')  
plt.show()
```



This graph shows the number of survivors distributed across different age groups. The bars represent the count of survivors in each age bin, and the curved line is a smoothed representation of the distribution.

The biggest number of survivors is in the age range around 20 to 30 years. This is the mode of the data.

The youngest survivors are as young as 0, and the oldest as old as 80.

The number of survivors is relatively high for the age group below 10, suggesting that children had a relatively good survival rate.

After the peak in the 20-30 age range, there's a decline as age increases. The number of survivors between the ages of 30 and 50 is moderately distributed, with a gradual decrease.

Beyond 50 years of age, the number of survivors drops more significantly, with very few survivors in the 70-80 age range.

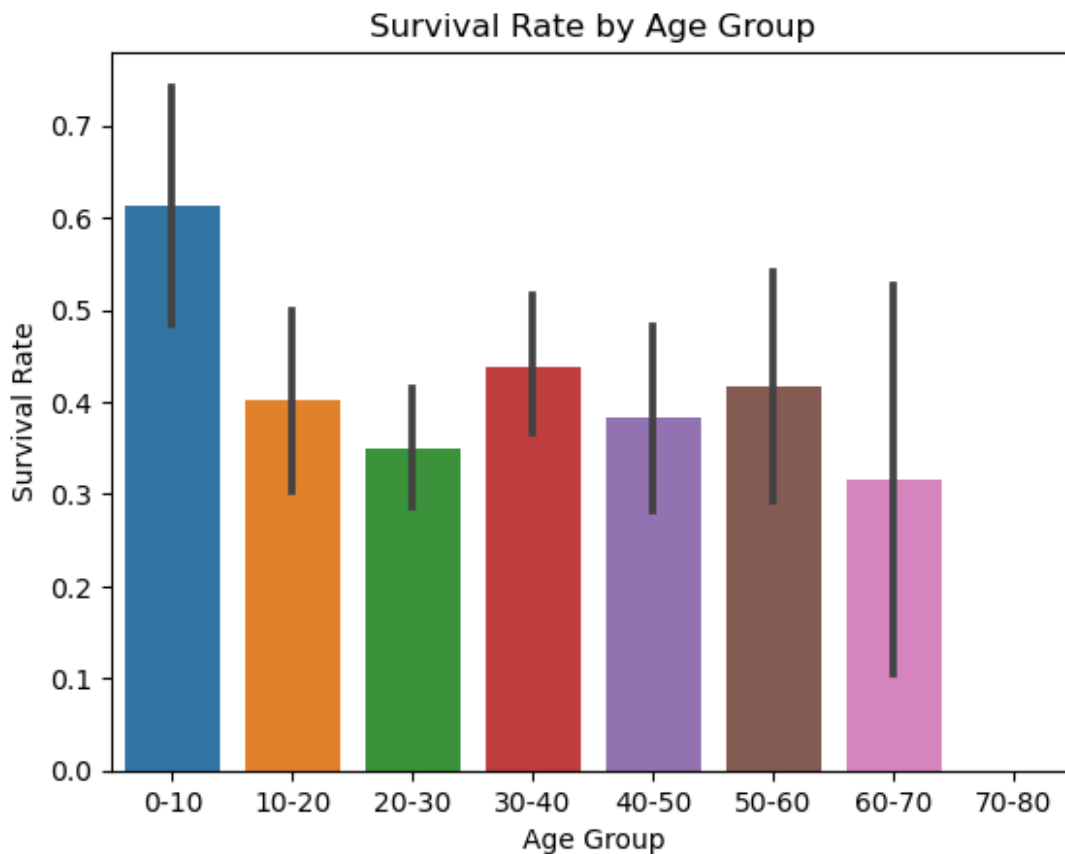
The smooth curve provides an overview of the distribution pattern. It peaks at the 20-30 age range and gradually decreases on either side, with a sharper decline after the age of 50.

It seems that the majority of survivors were young adults, with children also having a good survival rate. Elderly individuals had the least survival rate, especially those above 70.

```

bins = [0, 10, 20, 30, 40, 50, 60, 70, 80]
labels = ['0-10', '10-20', '20-30', '30-40', '40-50', '50-60', '60-70', '70-80']
tit['AgeGroup'] = pd.cut(tit['Age'], bins=bins, labels=labels,
right=False)
sns.barplot(x='AgeGroup', y='Survived', data=tit)
plt.title('Survival Rate by Age Group')
plt.xlabel('Age Group')
plt.ylabel('Survival Rate')
plt.show()

```



From this graph, we can observe how survival rate varies with different age groups. Children in the '0-10' age group have a higher survival rate compared to the other age groups. This could corroborate with the general belief that women and children had priority during the evacuation.

## Summary

With the analysis run, we can observe that sex, age, and passenger class may be associated with survival rate.