

Titanic Using Machine Learning

Step 1: Import Libraries and Load Data

```
In [38]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore", category=FutureWarning)

# Load the dataset
titanic = pd.read_csv("C:\\Users\\Ankit\\Desktop\\Data Science\\Titanic\\Dataset\\train.csv")
```

Step 2: Explore the Data

```
In [39]: print(titanic.head())
```

	PassengerId	Survived	Pclass	
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

	Name	Sex	Age	SibSp	
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
4	Allen, Mr. William Henry	male	35.0	0	

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S

```
In [40]: # Check for missing values
print(titanic.isnull().sum())
```

PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0
Parch	0
Ticket	0
Fare	0
Cabin	687
Embarked	2

dtype: int64

Step 3: Data Cleaning

```
In [41]: # Fill missing values in 'Age' with the median
titanic['Age'].fillna(titanic['Age'].median(), inplace=True)

In [42]: # Fill missing values in 'Embarked' with the most frequent value
titanic['Embarked'].fillna(titanic['Embarked'].mode()[0], inplace=True)

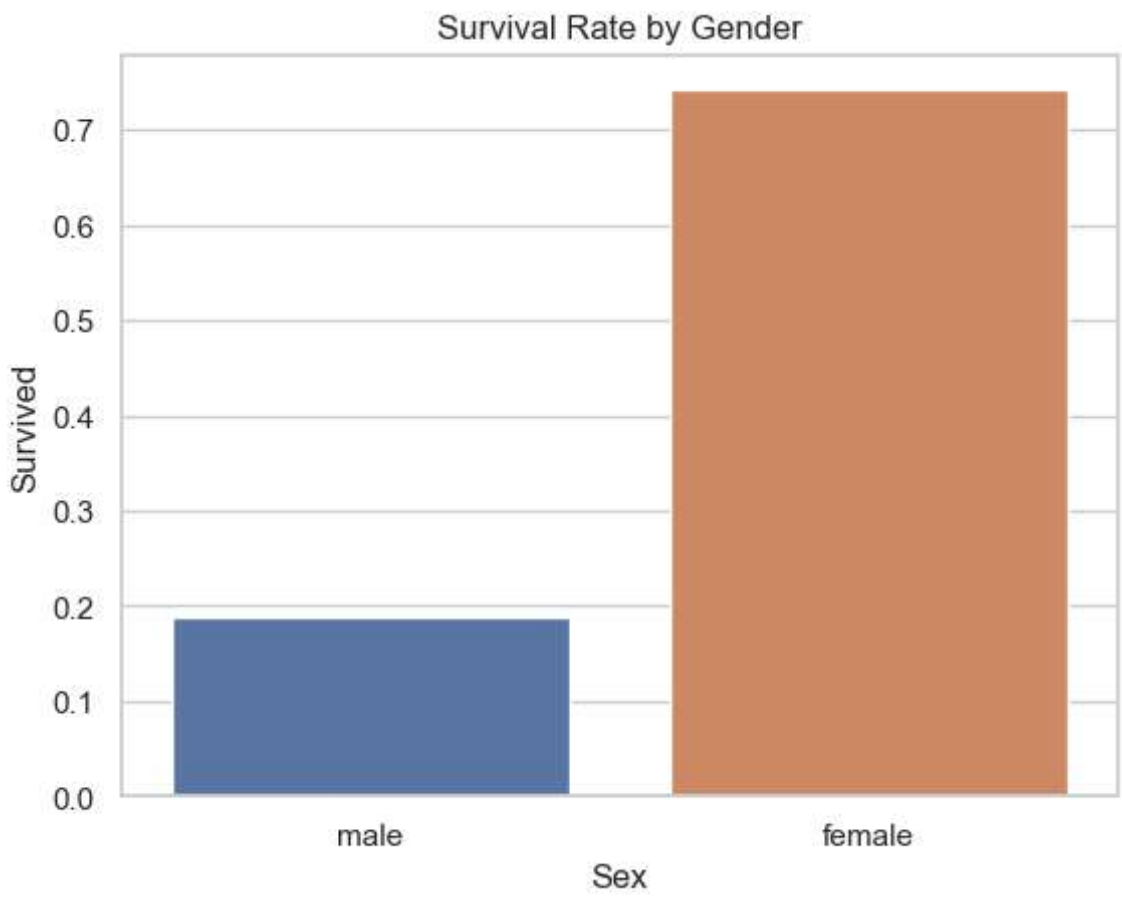
In [43]: # Drop 'Cabin' column due to too many missing values
titanic.drop('Cabin', axis=1, inplace=True)

In [44]: # Check again for missing values
print(titanic.isnull().sum())
```

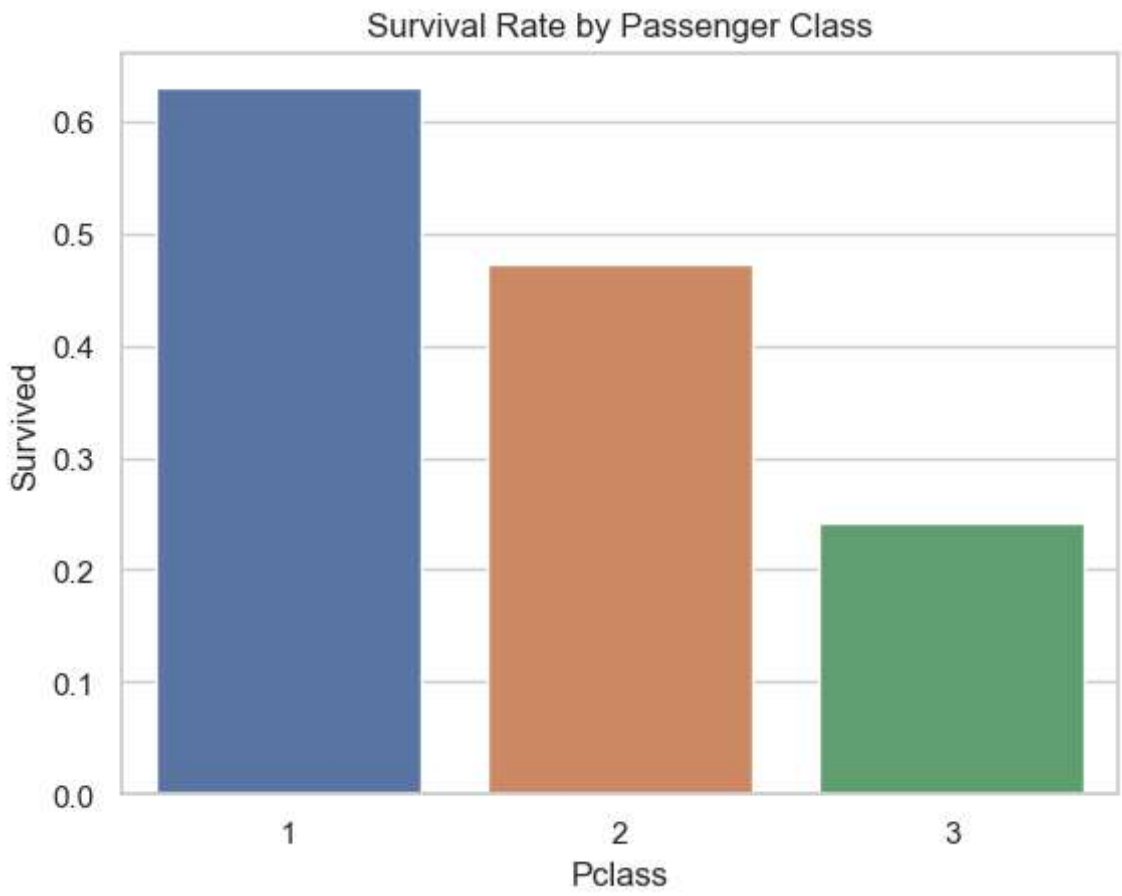
```
PassengerId    0
Survived        0
Pclass          0
Name            0
Sex             0
Age            0
SibSp           0
Parch           0
Ticket          0
Fare            0
Embarked        0
dtype: int64
```

Step 4:Explore Relationships and Patterns

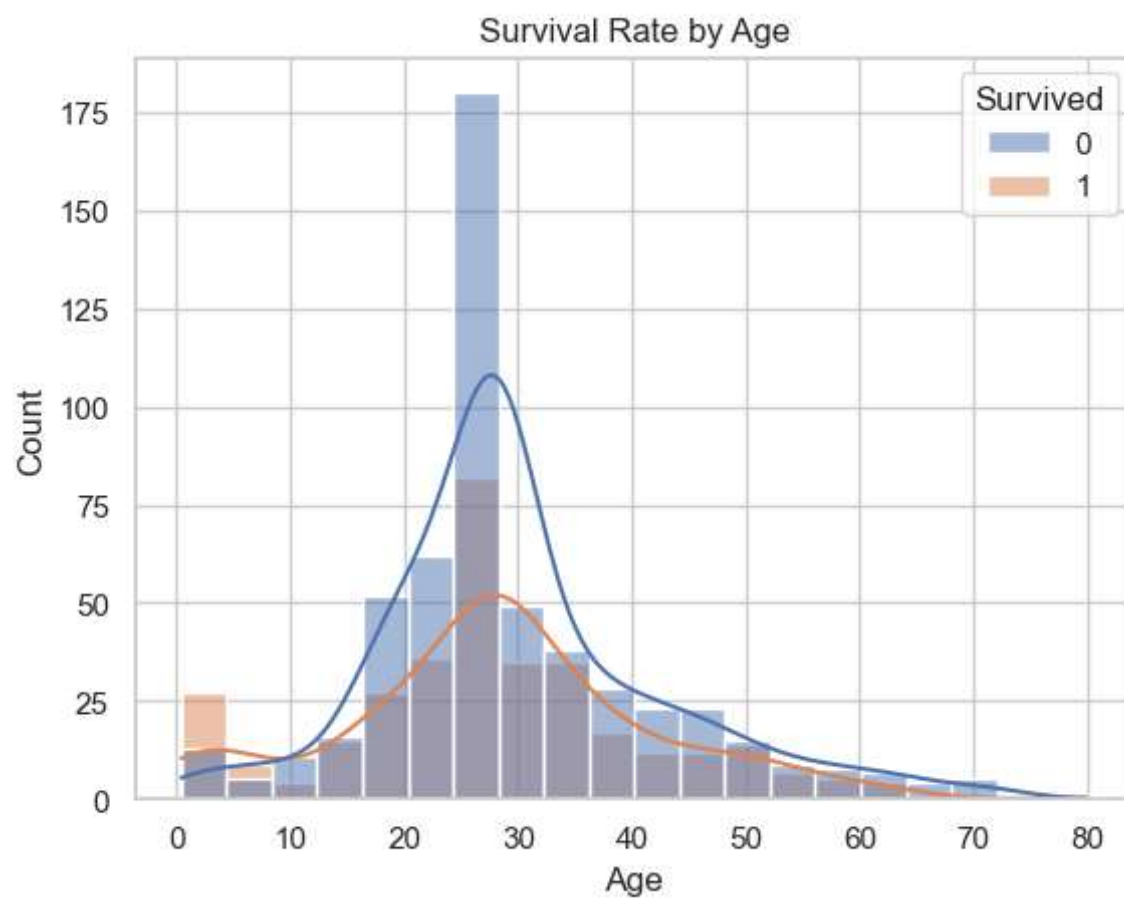
```
In [45]: ## a.Survival Rate by Gender
sns.set(style="whitegrid")
sns.barplot(x='Sex', y='Survived', data=titanic, errorbar=None)
plt.title('Survival Rate by Gender')
plt.show()
```



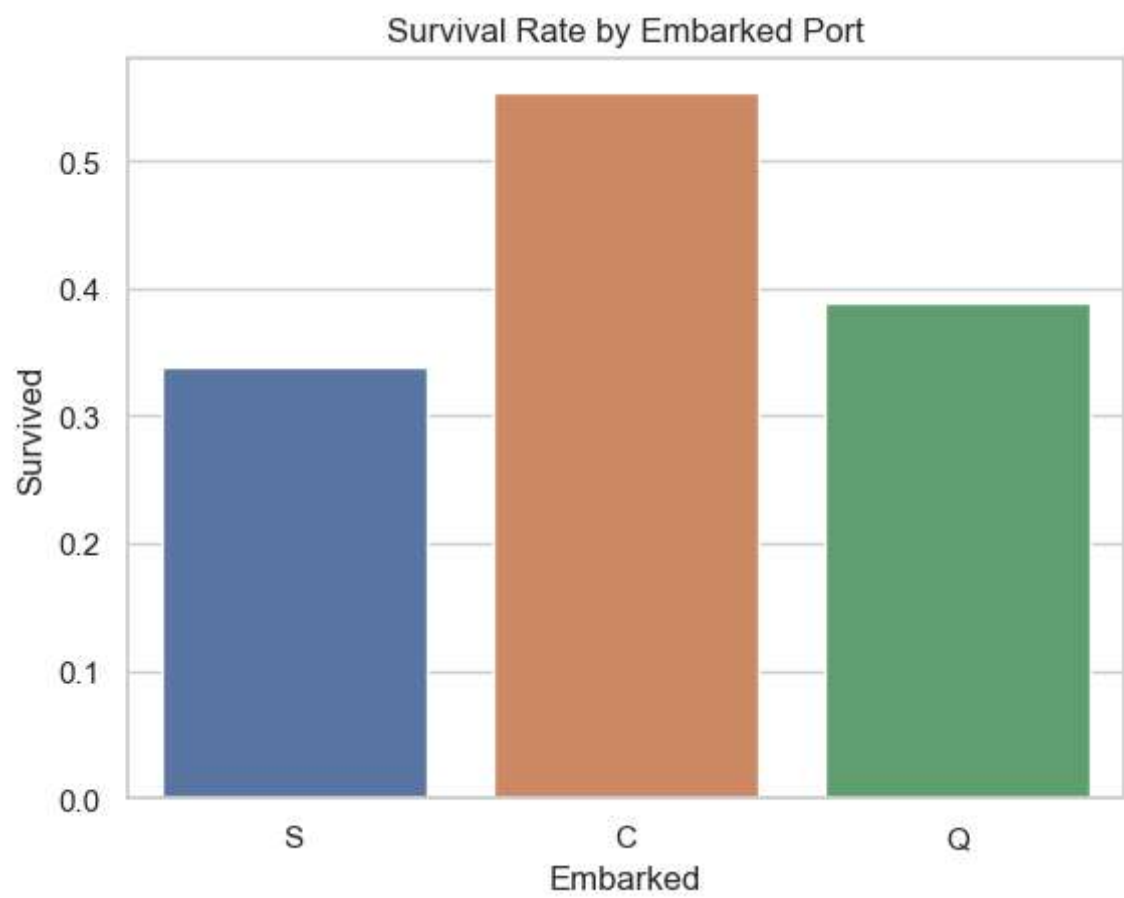
```
In [46]: ##Survival Rate by Passenger Class
sns.barplot(x='Pclass', y='Survived', data=titanic, errorbar=None)
plt.title('Survival Rate by Passenger Class')
plt.show()
```



```
In [47]: ## Survival Rate by Age:
sns.histplot(x='Age', hue='Survived', data=titanic, bins=20, kde=True)
plt.title('Survival Rate by Age')
plt.show()
```

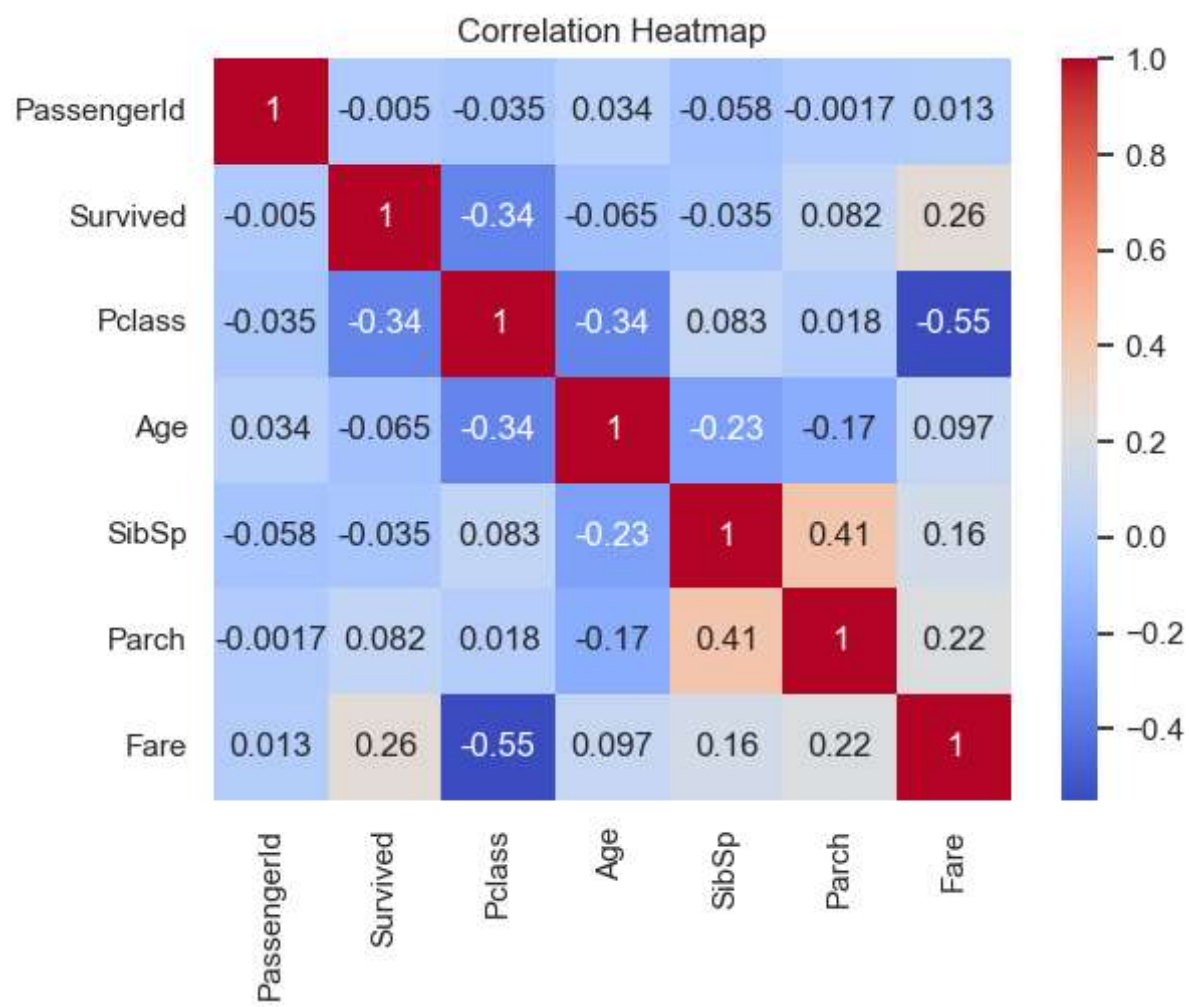


```
In [48]: ## Survival Rate by Embarked Port
sns.barplot(x='Embarked', y='Survived', data=titanic, errorbar=None)
plt.title('Survival Rate by Embarked Port')
plt.show()
```

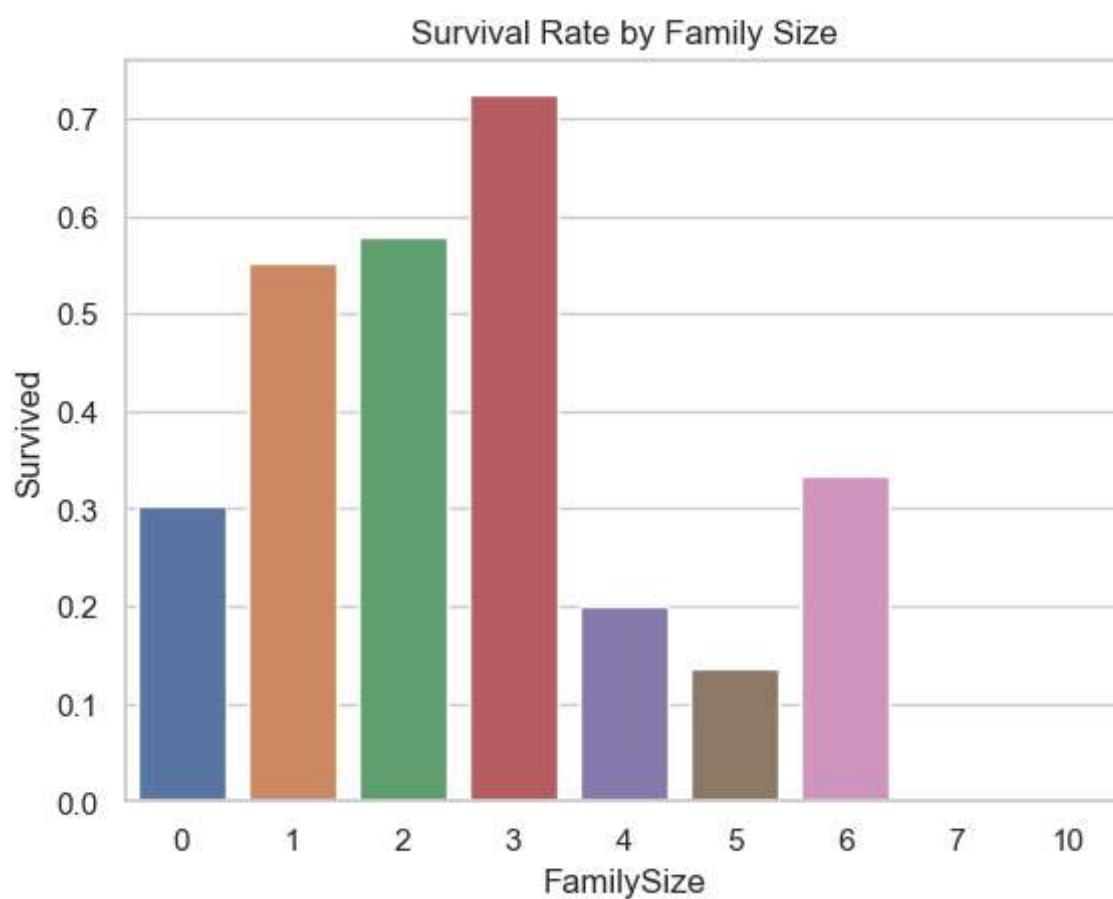


Additional Exploration

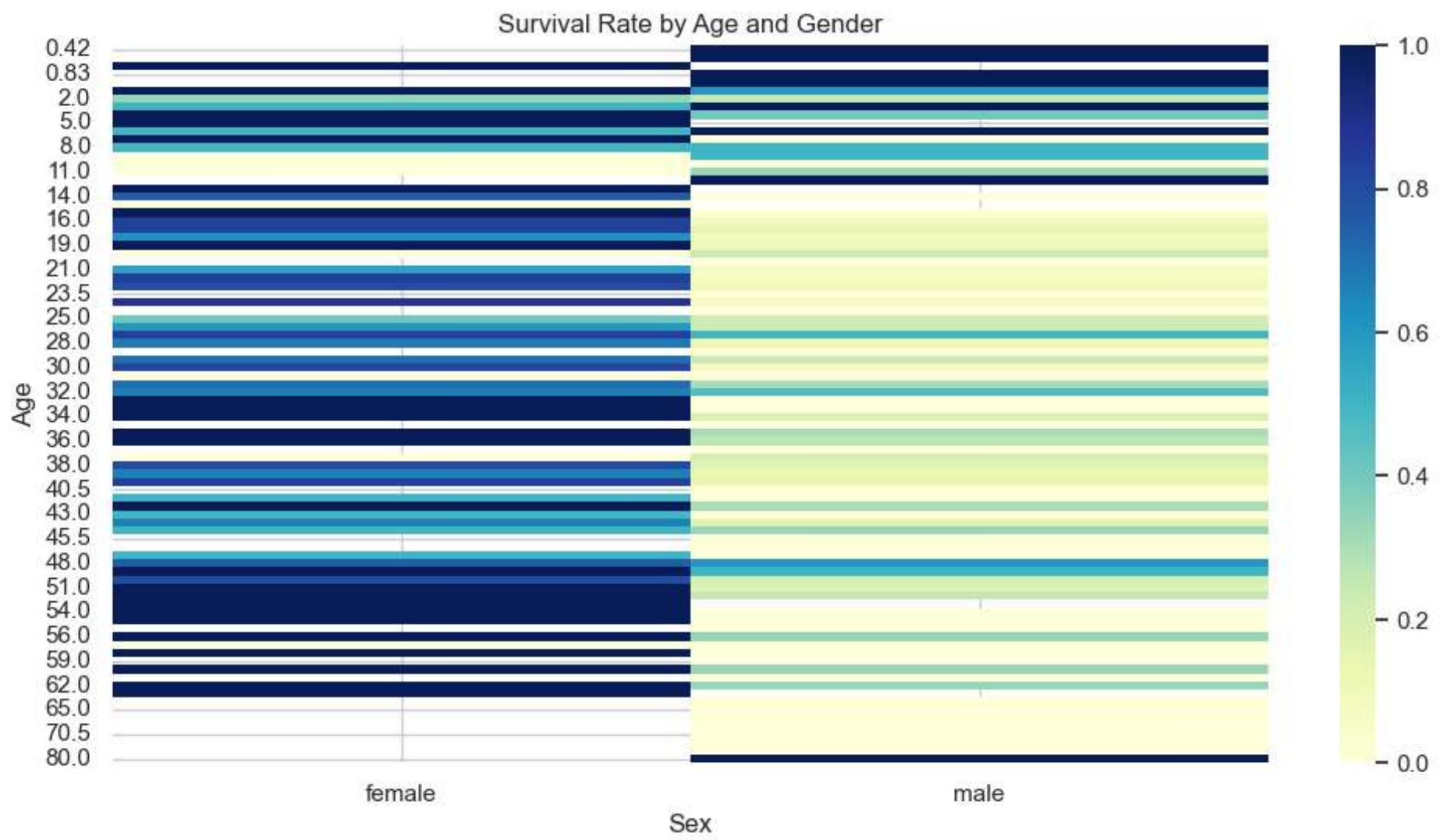
```
In [49]: ## Correlation Heatmap:
correlation_matrix = titanic.corr()
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm")
plt.title('Correlation Heatmap')
plt.show()
```



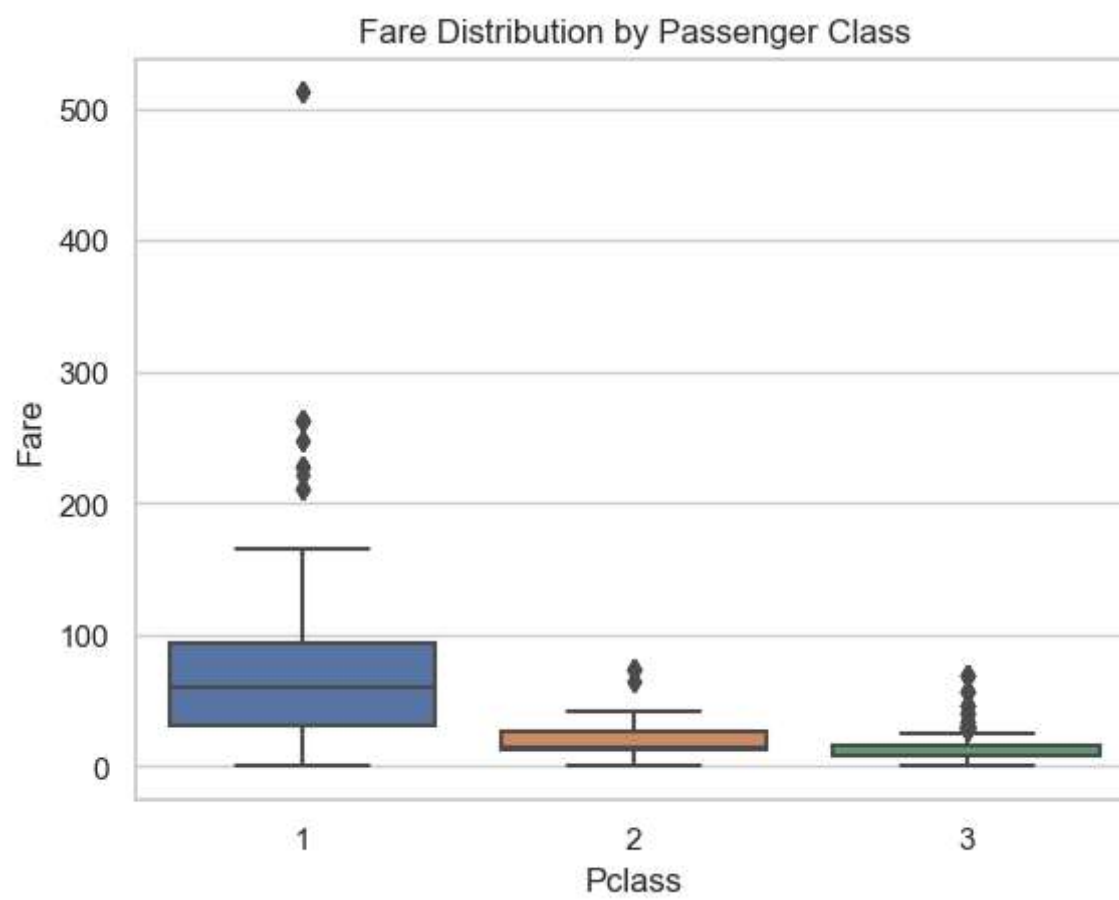
```
In [50]: ##survival rate by family size
titanic['FamilySize'] = titanic['SibSp'] + titanic['Parch']
sns.barplot(x='FamilySize', y='Survived', data=titanic, ci=None)
plt.title('Survival Rate by Family Size')
plt.show()
```



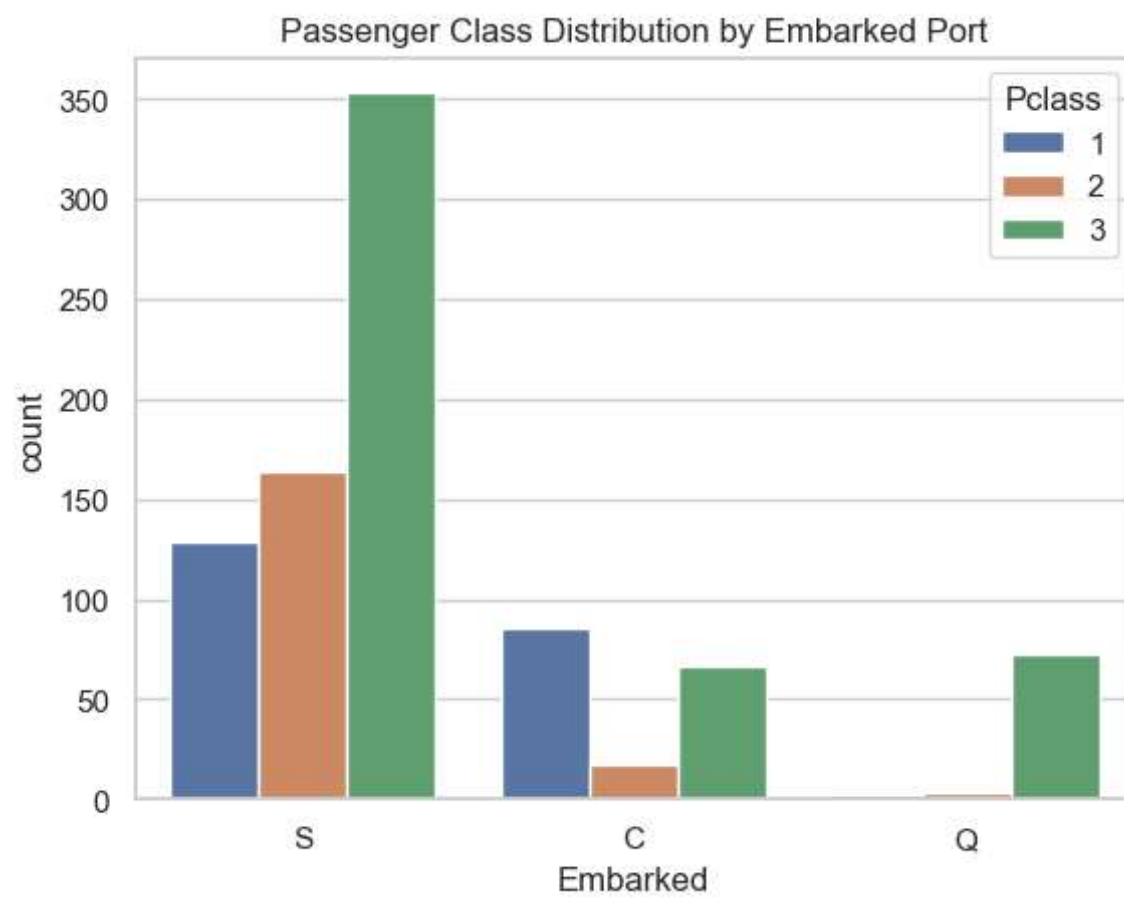
```
In [51]: ## Survival Rate by Age and Gender:
age_gender_survival = titanic.groupby(['Age', 'Sex'])['Survived'].mean().unstack()
plt.figure(figsize=(12, 6))
sns.heatmap(age_gender_survival, cmap="YlGnBu")
plt.title('Survival Rate by Age and Gender')
plt.show()
```



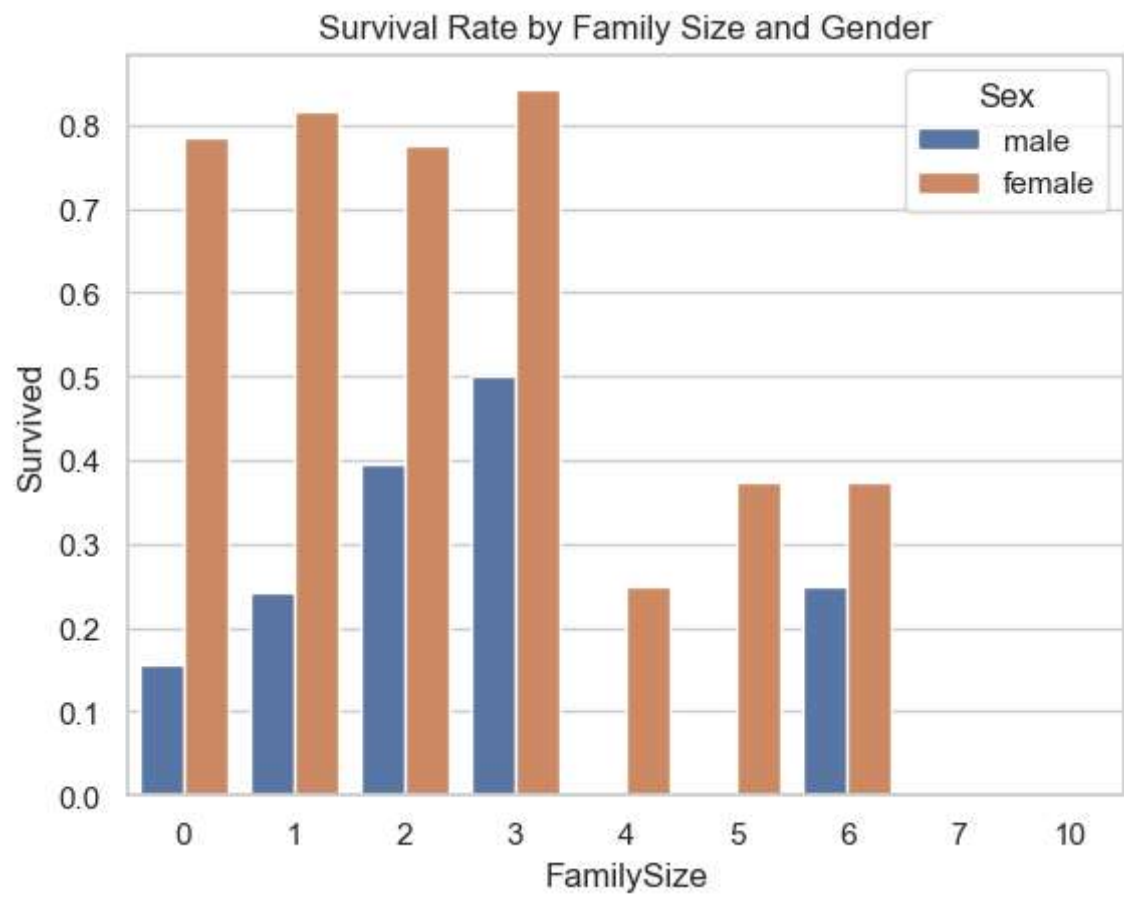
```
In [52]: ## Survival Rate by Fare and Passenger Class:
sns.boxplot(x='Pclass', y='Fare', data=titanic)
plt.title('Fare Distribution by Passenger Class')
plt.show()
```



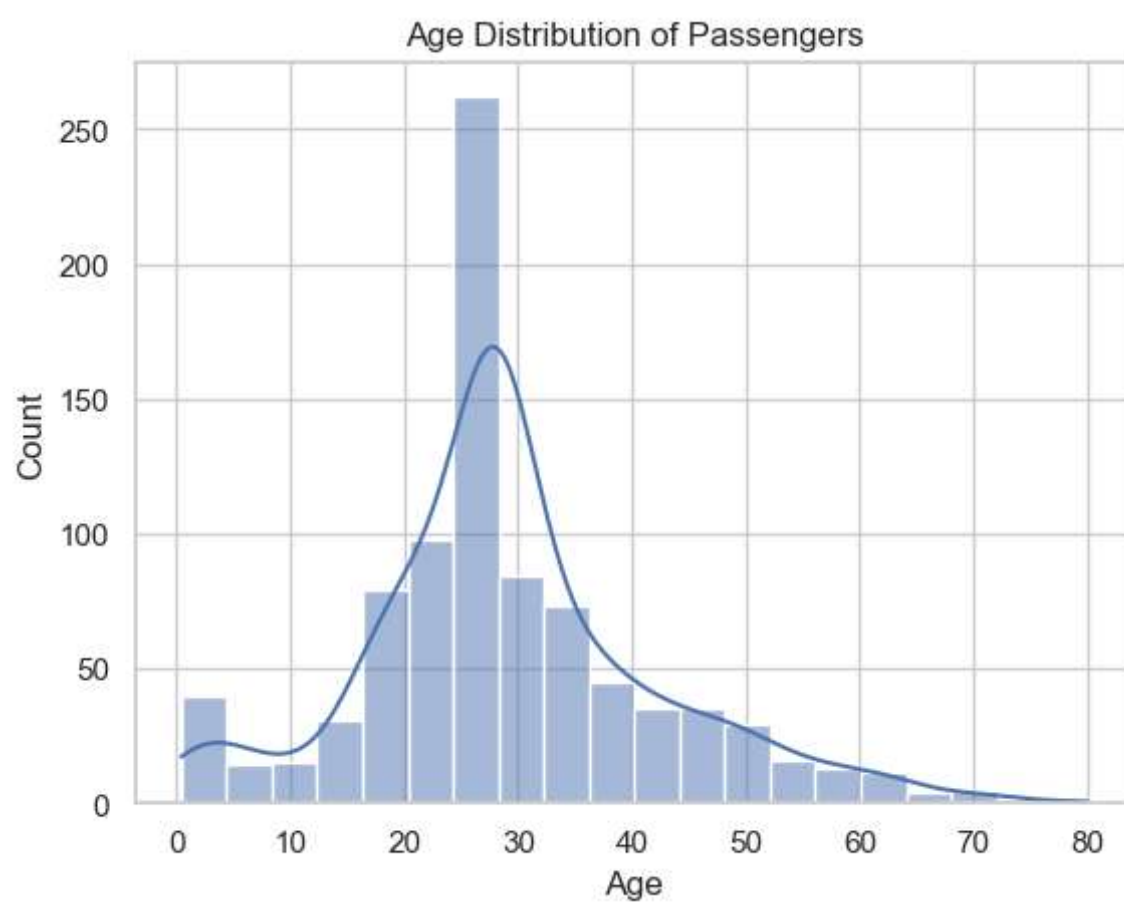
```
In [53]: ## Survival Rate by Embarked Port and Passenger Class
sns.countplot(x='Embarked', hue='Pclass', data=titanic)
plt.title('Passenger Class Distribution by Embarked Port')
plt.show()
```



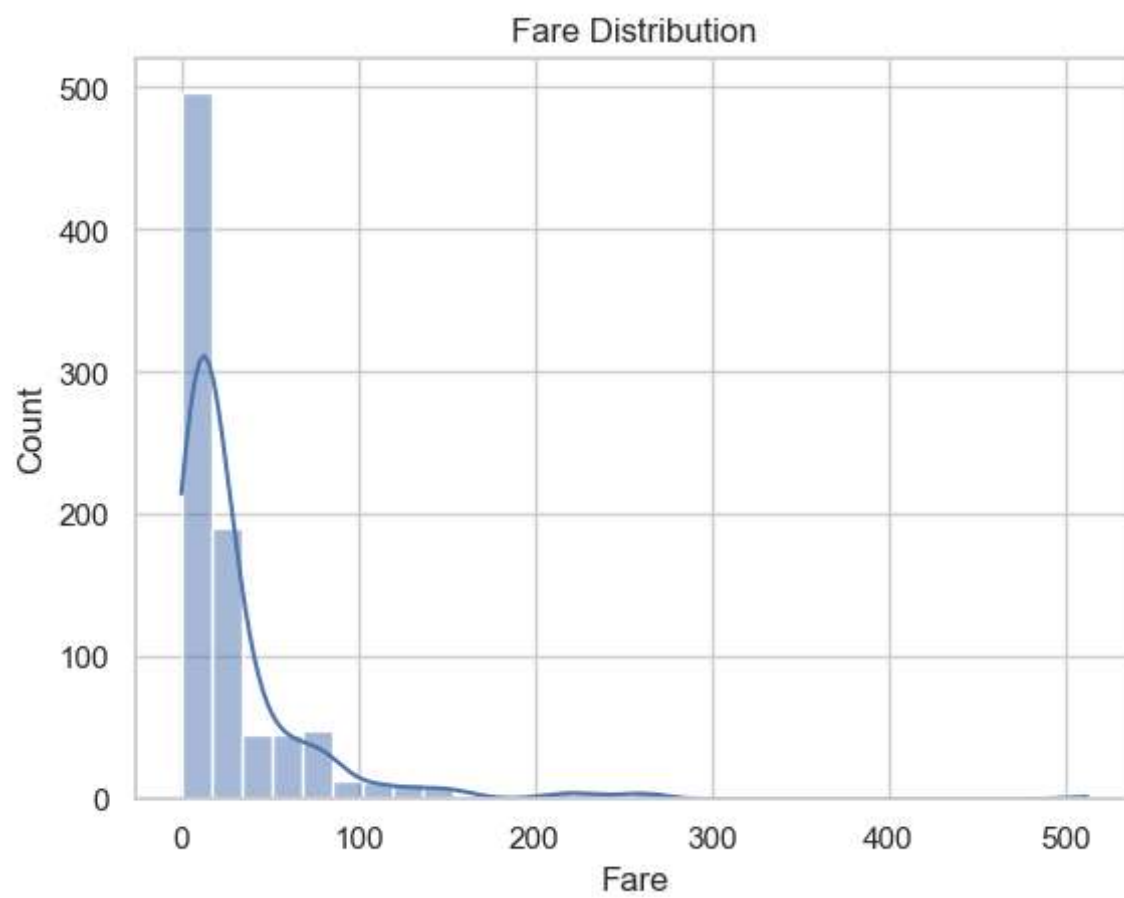
```
In [55]: ##Survival Rate by Family Size and Gender:
sns.barplot(x='FamilySize', y='Survived', hue='Sex', data=titanic, ci=None)
plt.title('Survival Rate by Family Size and Gender')
plt.show()
```



```
In [56]: ## Age Distribution of Passengers
sns.histplot(x='Age', data=titanic, bins=20, kde=True)
plt.title('Age Distribution of Passengers')
plt.show()
```



```
In [57]: ##Fare Distribution:
sns.histplot(x='Fare', data=titanic, bins=30, kde=True)
plt.title('Fare Distribution')
plt.show()
```



```
In [ ]:
```