```
 In [1]: import pandas as pd

 In [2]: import numpy as np

 In [3]: import matplotlib.pyplot as plt

 In [4]: import seaborn as sns

 In [5]: # Load the Titanic Dataset from Kaggle

 In [6]: train_df = pd.read_csv('train.csv')

 In [7]: test_df = pd.read_csv('test.csv')

 In [8]: # Data Cleaning

In [48]: print(train_df.info())
         print(train_df.describe())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          891 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Embarked     891 non-null    object 
dtypes: float64(2), int64(5), object(4)
memory usage: 76.7+ KB
None
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  891.000000  891.000000   
mean    446.000000    0.383838    2.308642   29.361582    0.523008   
std     257.353842    0.486592    0.836071   13.019697    1.102743   
min       1.000000    0.000000    1.000000    0.420000    0.000000   
25%     223.500000    0.000000    2.000000   22.000000    0.000000   
50%     446.000000    0.000000    3.000000   28.000000    0.000000   
75%     668.500000    1.000000    3.000000   35.000000    1.000000   
max     891.000000    1.000000    3.000000   80.000000    8.000000   

            Parch        Fare  
count  891.000000  891.000000  
mean     0.381594   32.204208  
std      0.806057   49.693429  
min      0.000000    0.000000  
25%      0.000000    7.910400  
50%      0.000000   14.454200  
75%      0.000000   31.000000  
max      6.000000  512.329200  
```

In [11]:
```
# Fill missing Age values with the median Age
```

In [26]:
```
train_df.loc[:, 'Age'] = train_df['Age'].fillna(1)
```

In [27]:
```
test_df.loc[:, 'Age'] = test_df['Age'].fillna(1)
```

In [28]:
```
# Drop the cabin column since it has too many missing values
```

In [32]:
```
train_df.drop('Cabin', axis=1, inplace=True)
```

In [33]:
```
test_df.drop('Cabin', axis=1, inplace=True)
```

In [34]:
```
# Fill missing embarked values with 'S'
```

In [36]:

```
train_df['Embarked'] = train_df['Embarked'].fillna('S')
```

In [37]:

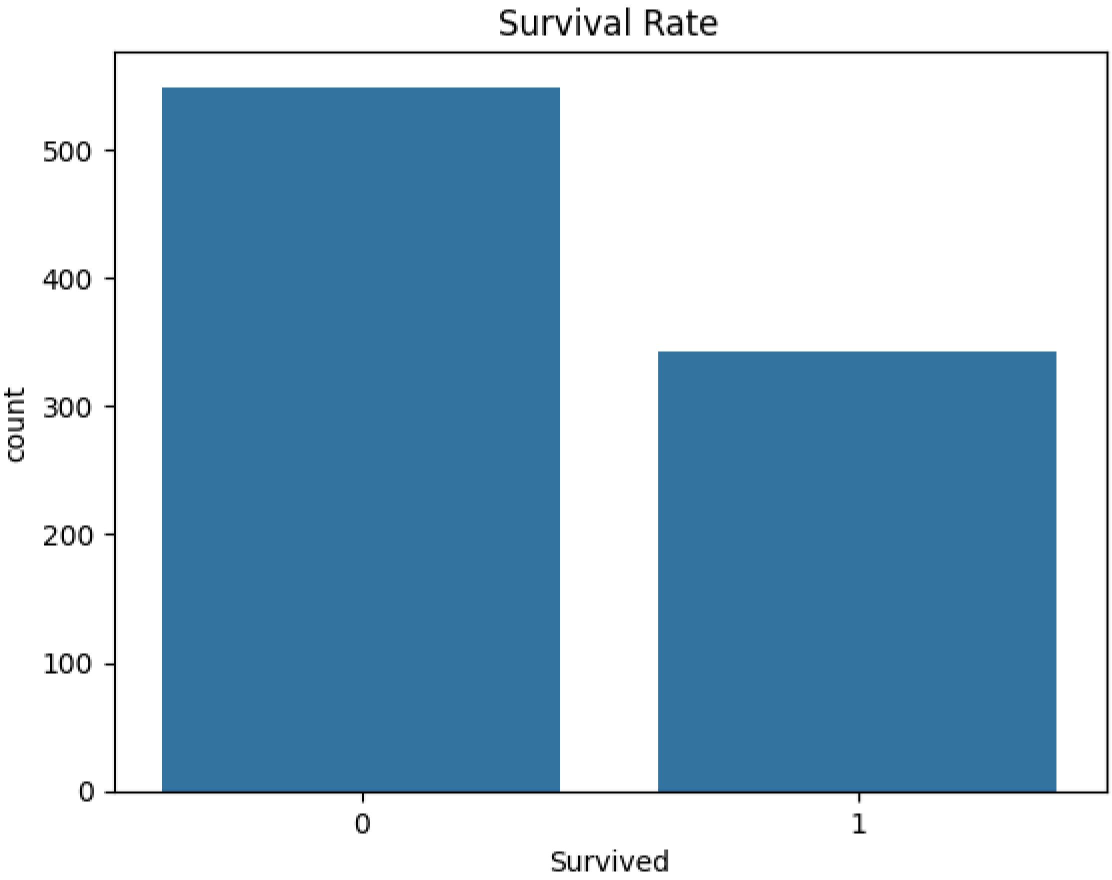
```
# Exploratory Data Analysis (EDA)
```

In [38]:

```
# 1 Survival Rate
```

In [59]:

```
sns.countplot(x='Survived', data=train_df)
plt.title('Survival Rate')
plt.show()
```

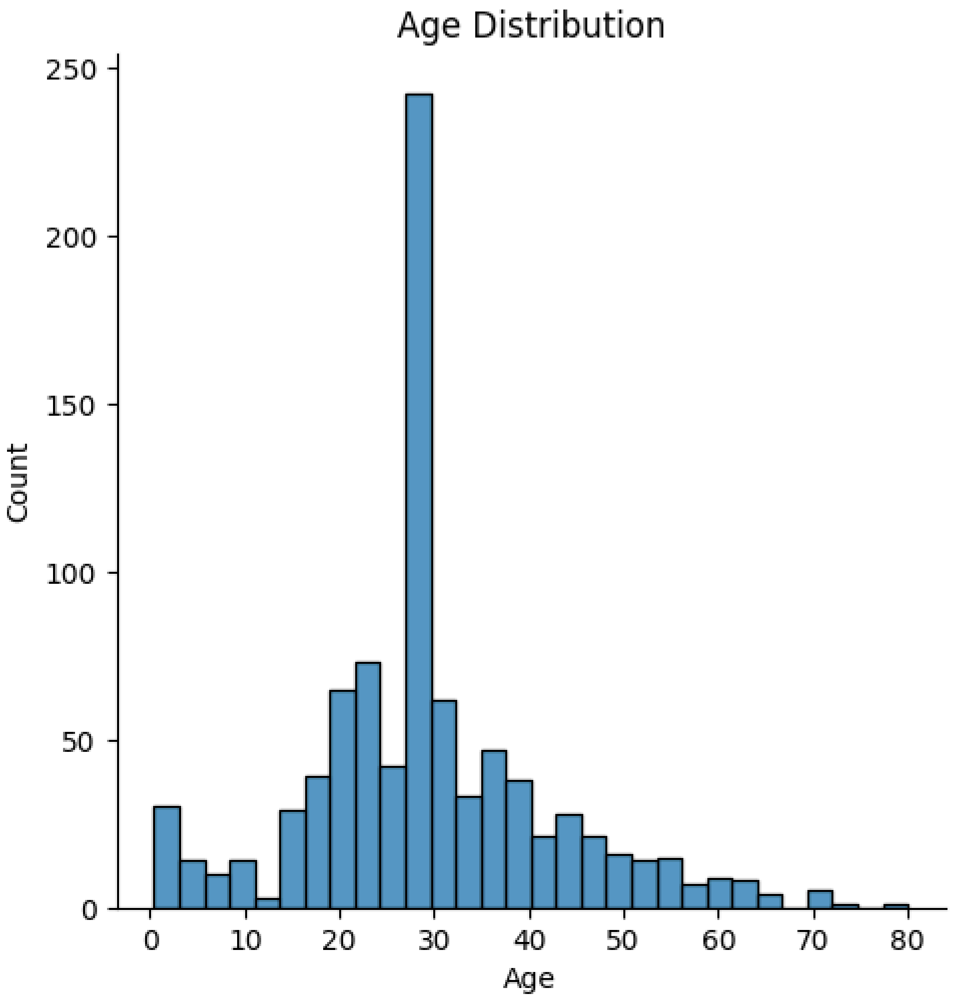


In [43]:

```
# 2 Age Distribution
```

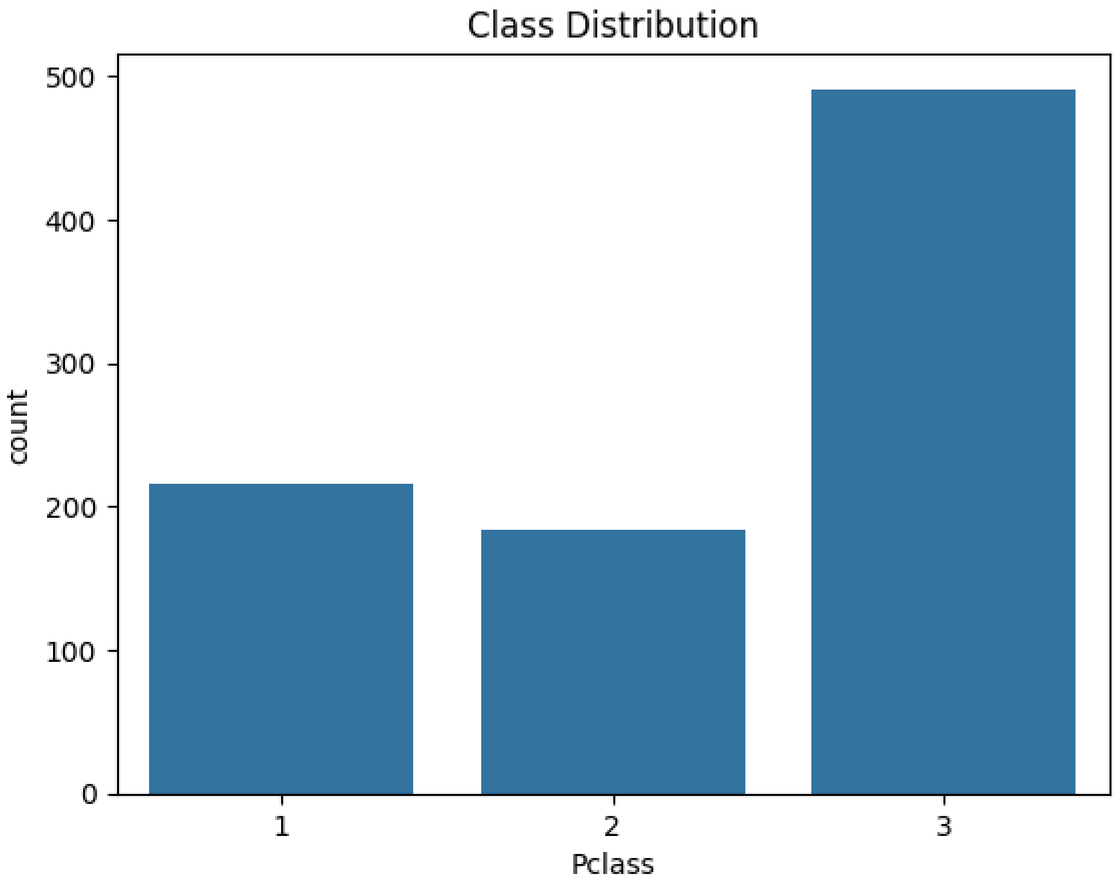
In [58]:

```
sns.displot(train_df['Age'], kde=False)
plt.title('Age Distribution')
plt.show()
```



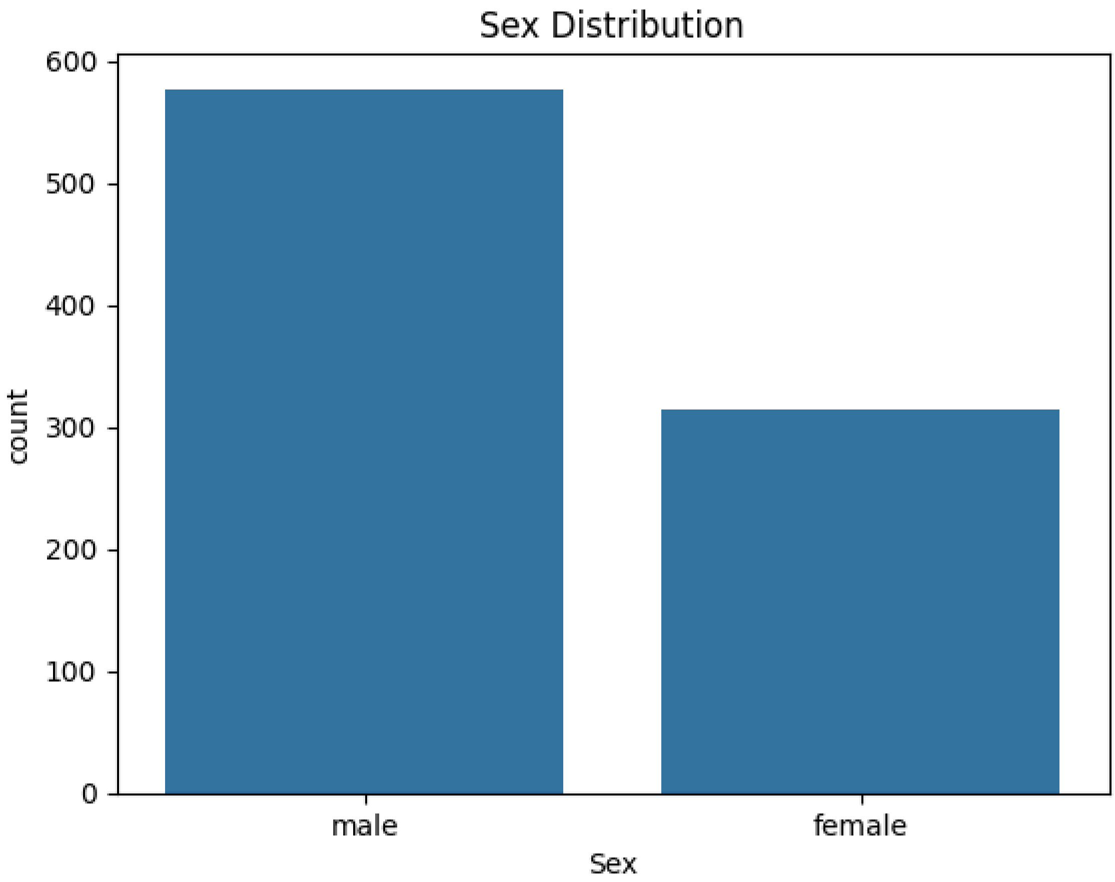
```
In [49]: # 3 Class Distribution
```

```
In [57]: sns.countplot(x='Pclass', data=train_df)
         plt.title('Class Distribution')
         plt.show()
```



```
In [51]: # 4 Sex Distribution
```

```
In [56]: sns.countplot(x='Sex', data=train_df)
         plt.title('Sex Distribution')
         plt.show()
```

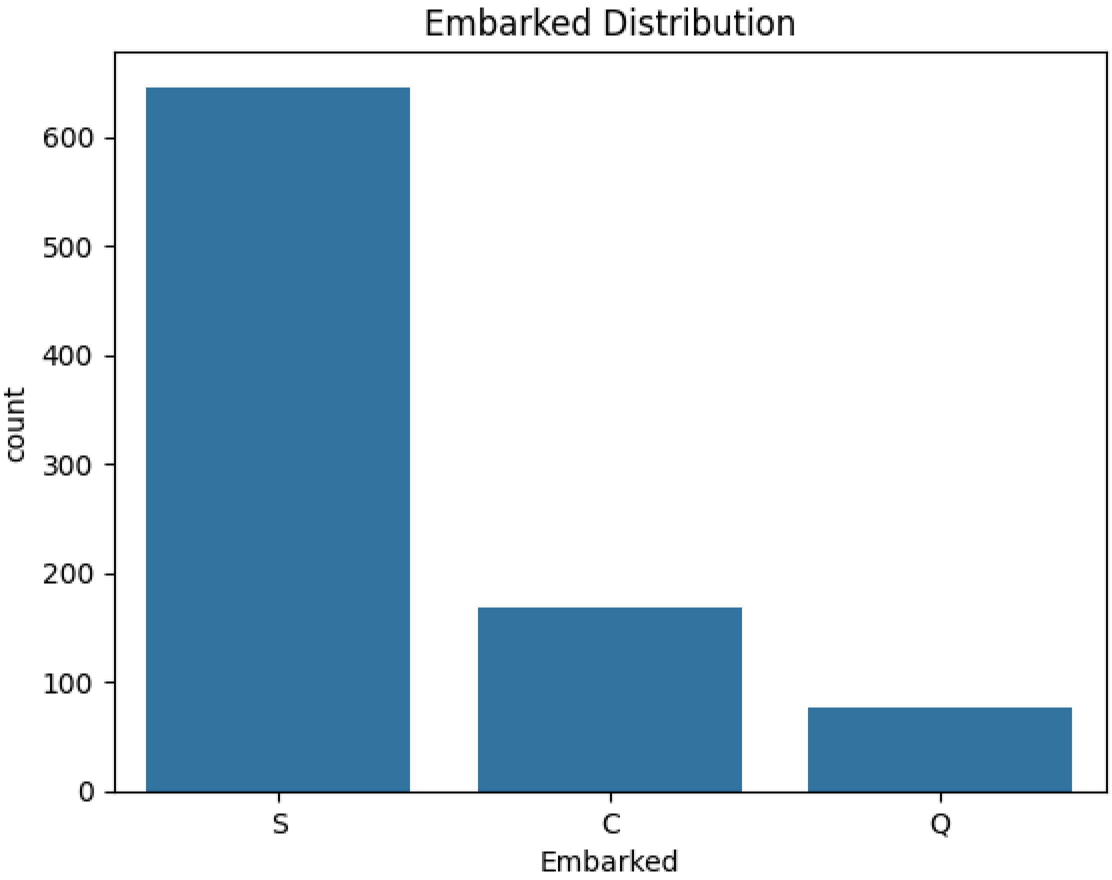


In [53]:

```
# 5 Embarked Distribution
```

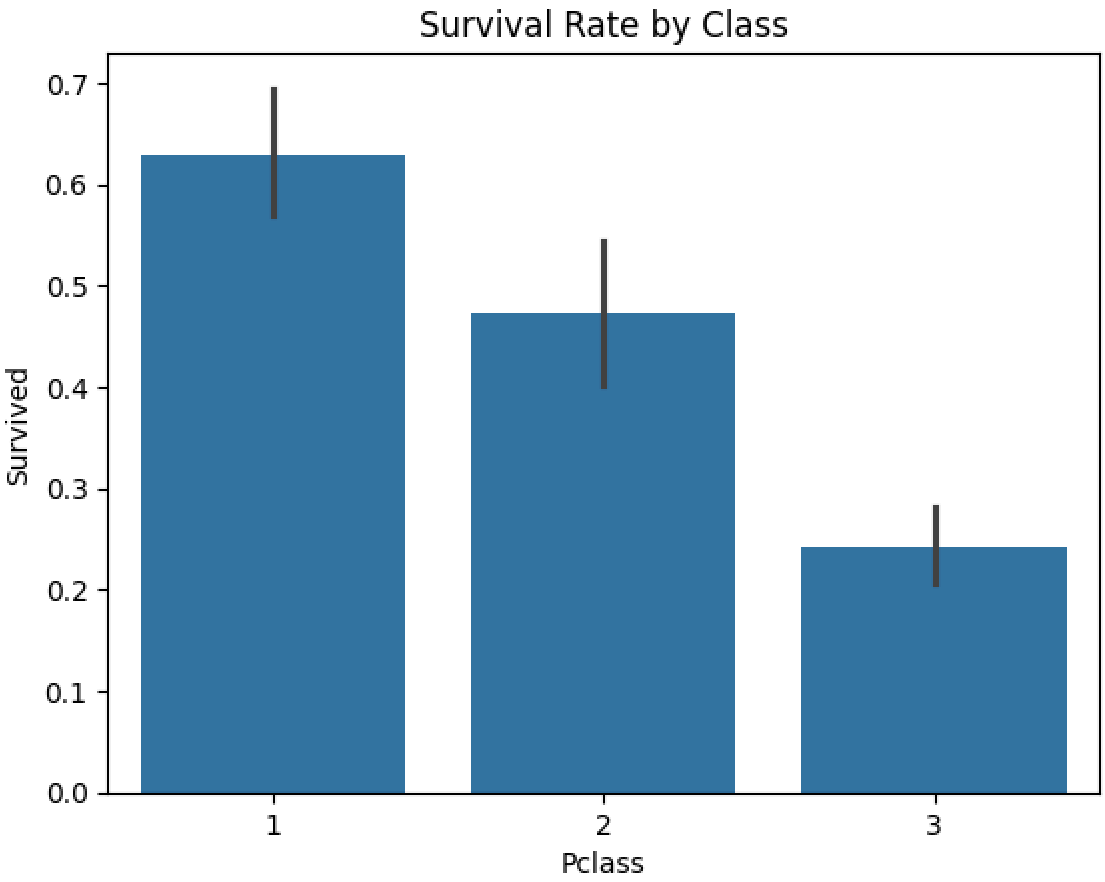
In [55]:

```
sns.countplot(x='Embarked', data=train_df)
plt.title('Embarked Distribution')
plt.show()
```



```
In [60]: # 6 Survival Rate by Class
```

```
In [62]: sns.barplot(x='Pclass', y='Survived', data=train_df)
         plt.title('Survival Rate by Class')
         plt.show()
```

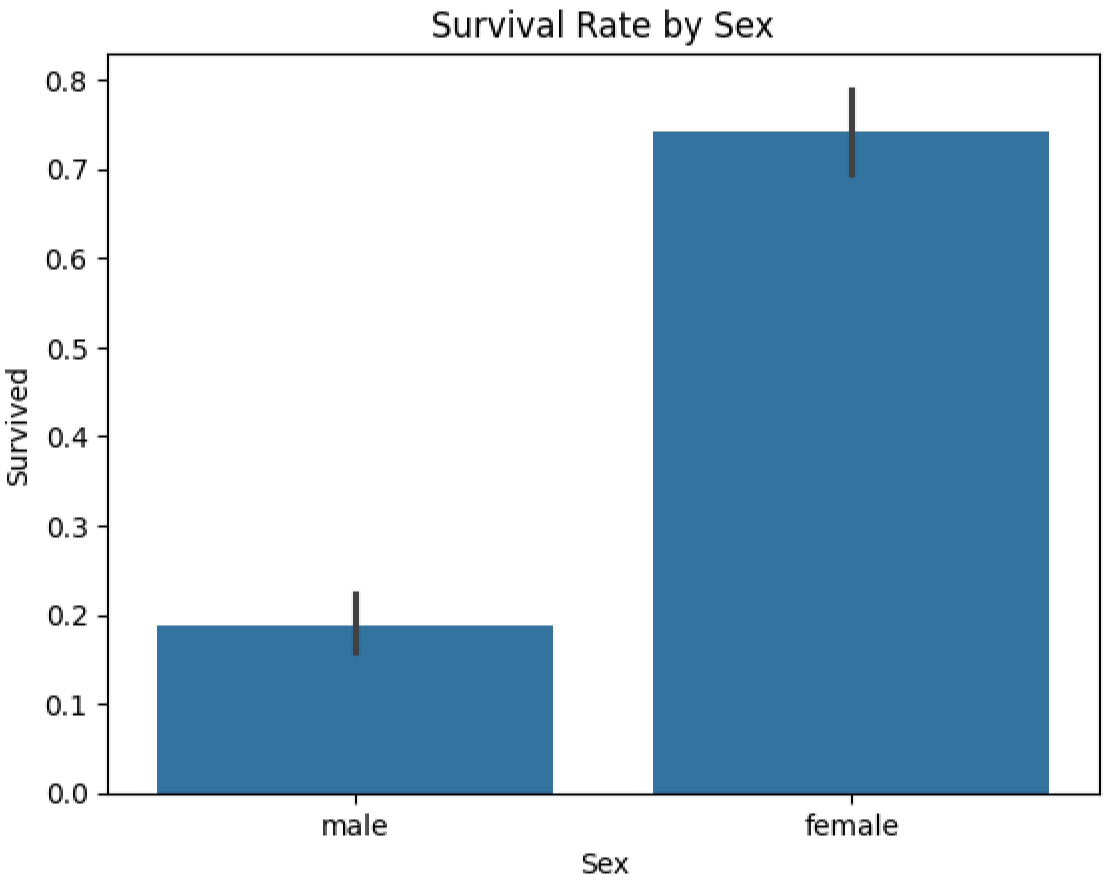


In [63]:

```
# 7 Survival Rate by Sex
```

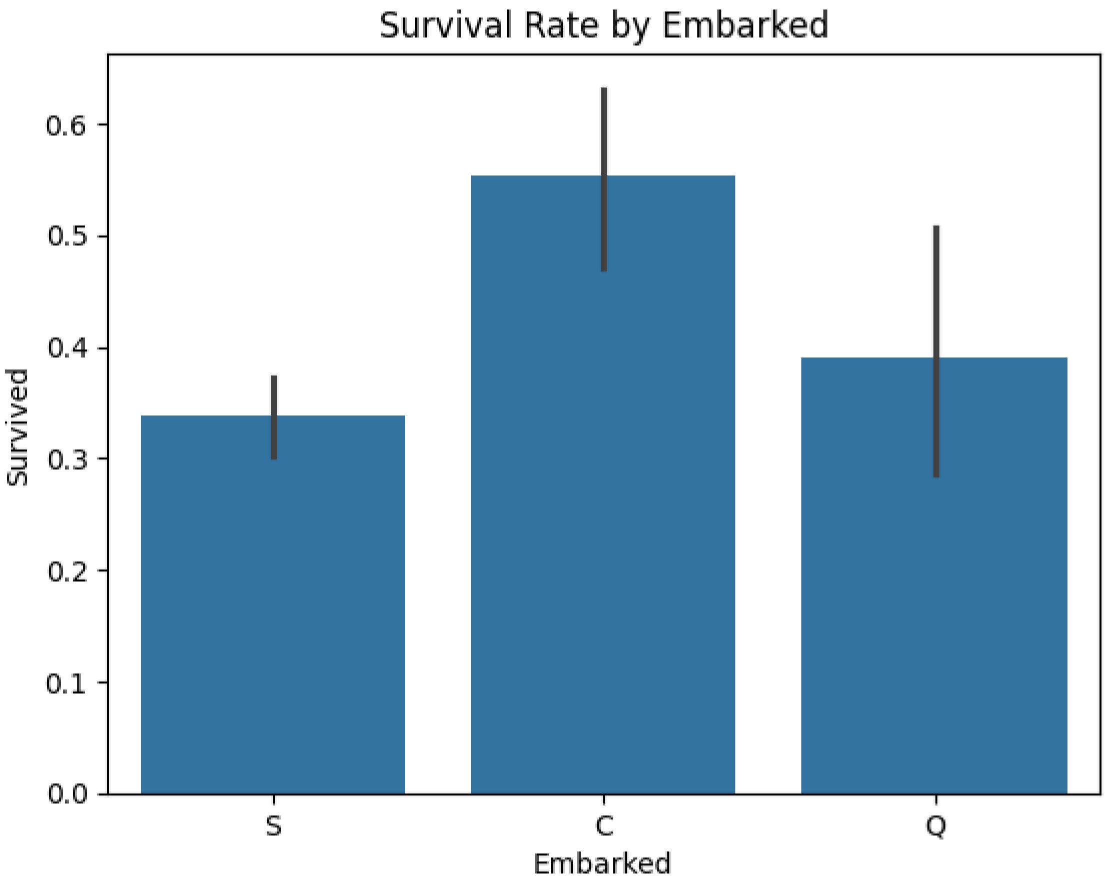
In [64]:

```
sns.barplot(x='Sex', y='Survived', data=train_df)
plt.title('Survival Rate by Sex')
plt.show()
```



```
In [65]: # 8 Survival Rate by Embarked
```

```
In [66]: sns.barplot(x='Embarked', y='Survived', data=train_df)
         plt.title('Survival Rate by Embarked')
         plt.show()
```


Survival Rate by Embarked
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Survived
S
C
Q
Embarked

```
In [ ]:
```