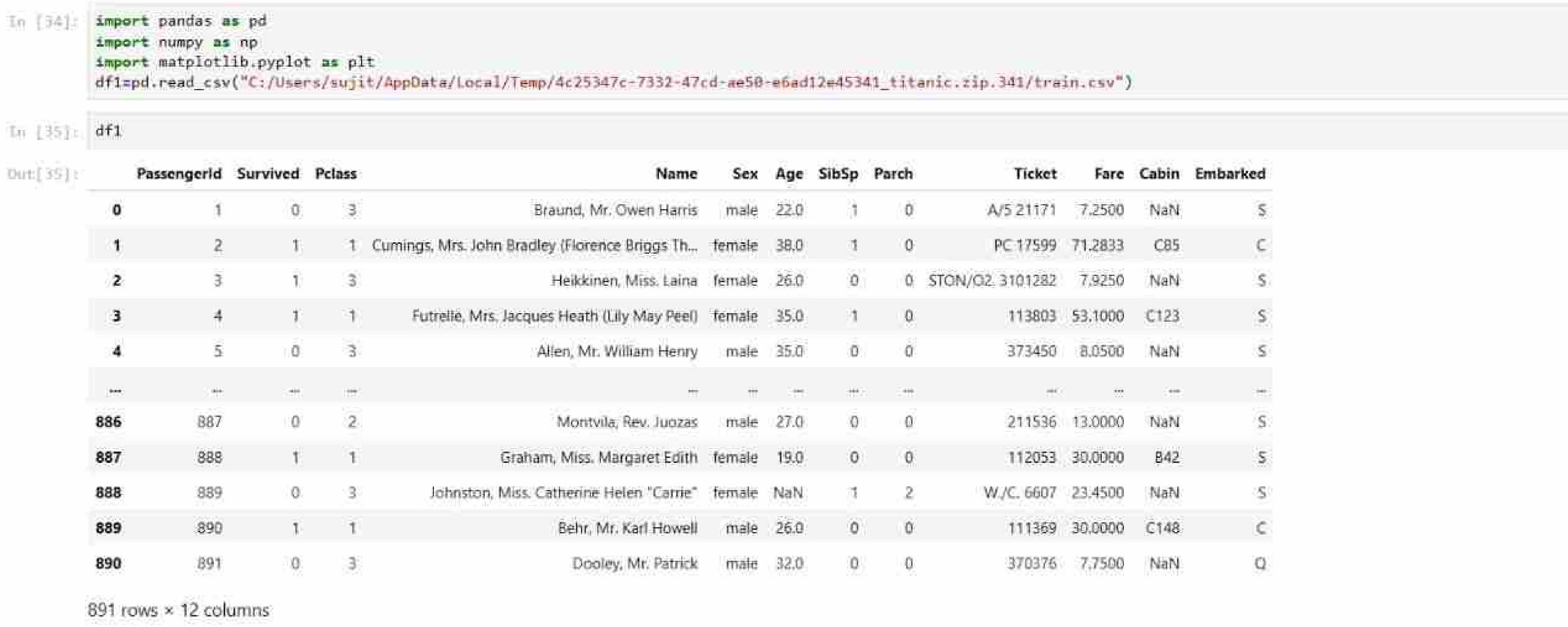


In [34]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df1=pd.read_csv("C:/Users/sujit/AppData/Local/Temp/4c25347c-7332-47cd-ae50-e6ad12e45341_titanic.zip.341/train.csv")
```

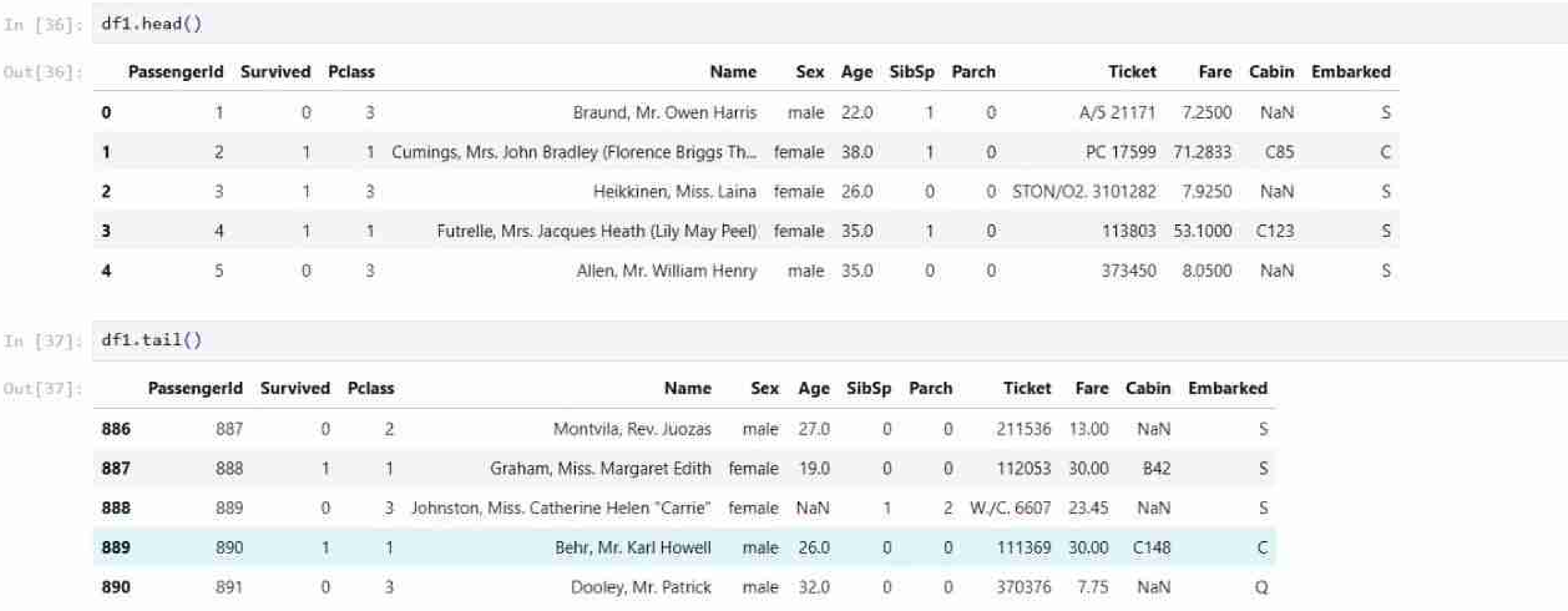
In [35]:

```
df1
```

Out[35]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns



```
In [36]: df1.head()
```

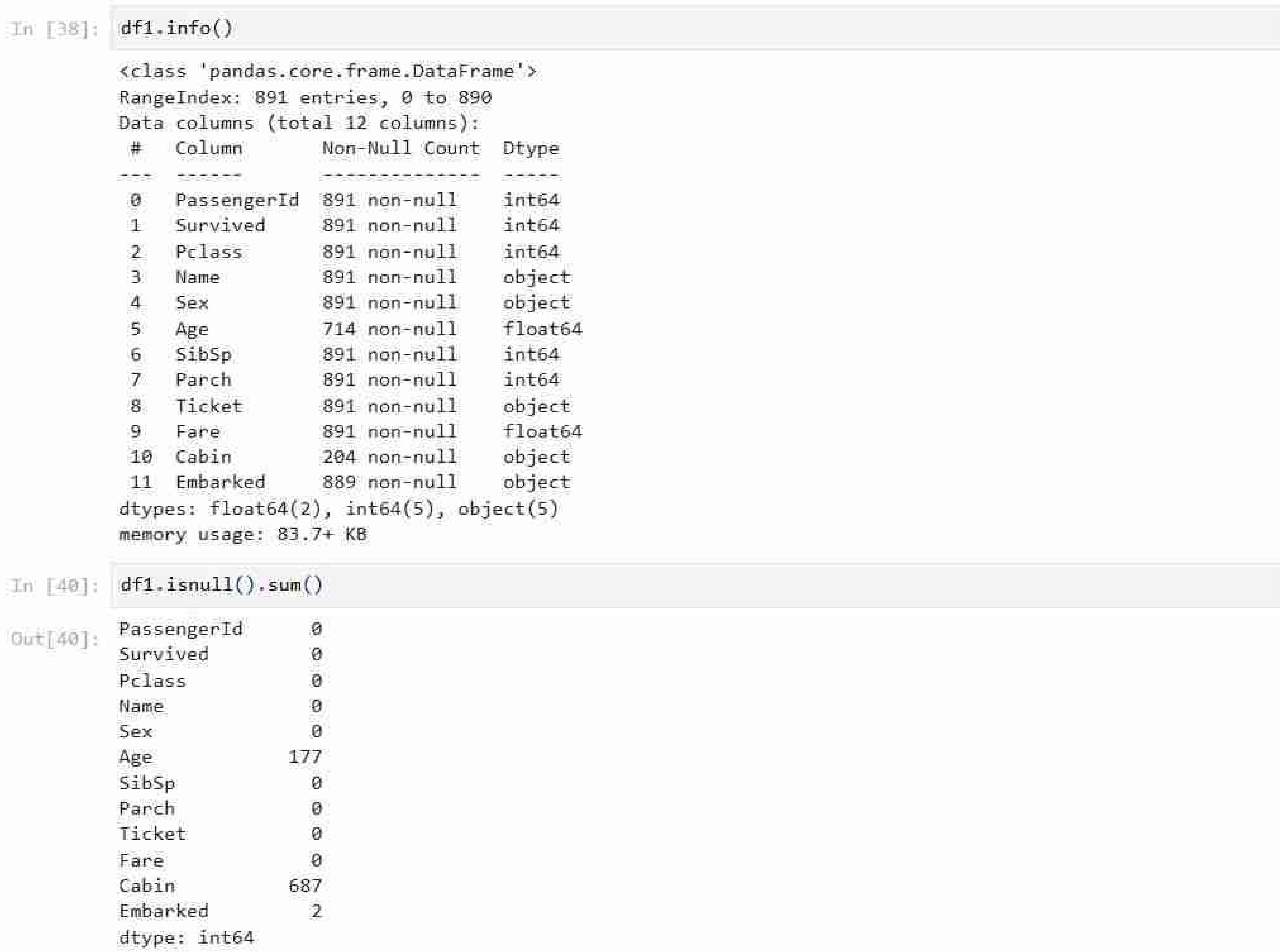
Out[36]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

```
In [37]: df1.tail()
```

Out[37]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q



In [38]:

```
df1.info()
```

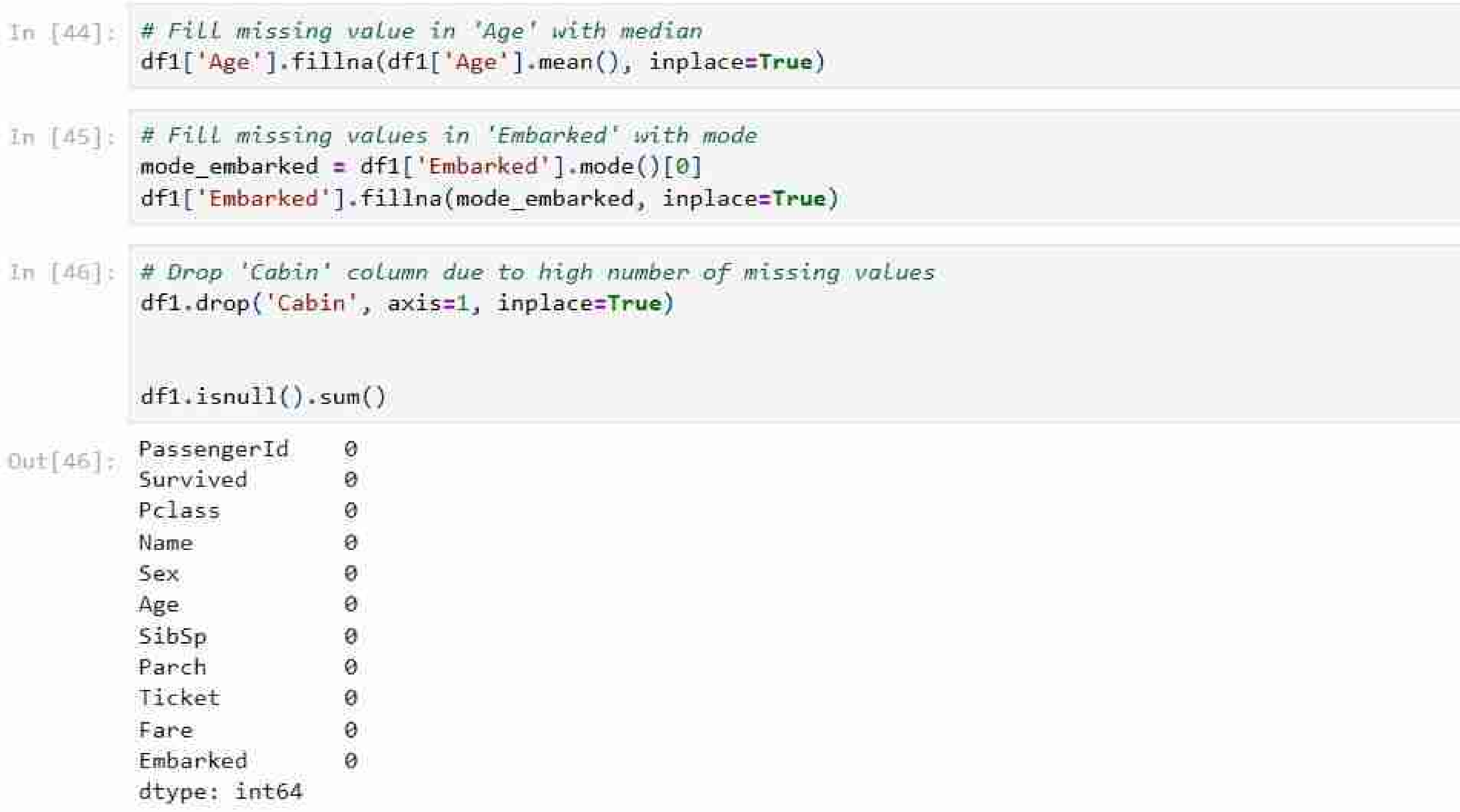
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 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  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

In [40]:

```
df1.isnull().sum()
```

Out[40]:

```
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
```



In [44]:

```
# Fill missing value in 'Age' with median
df1['Age'].fillna(df1['Age'].mean(), inplace=True)
```

In [45]:

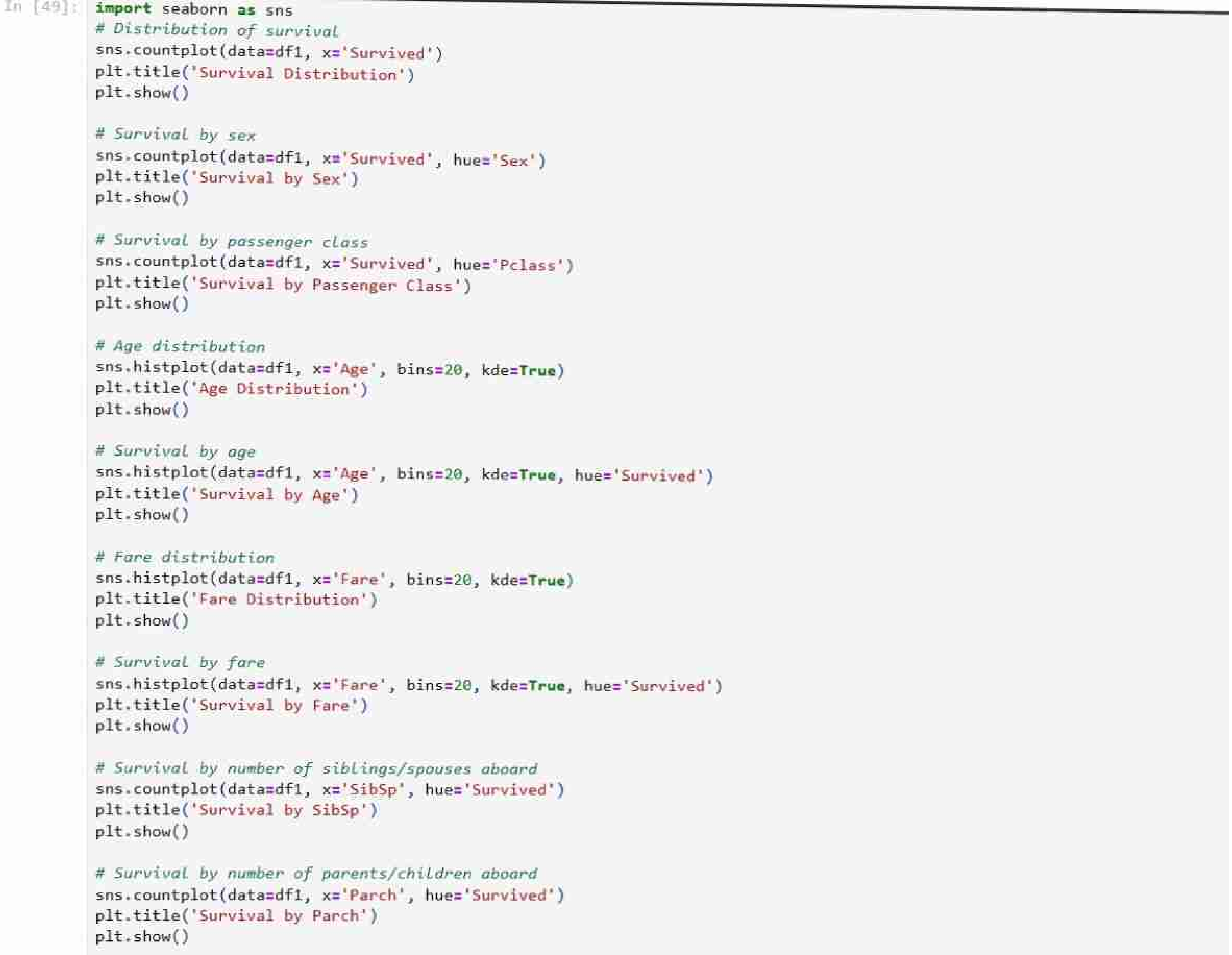
```
# Fill missing values in 'Embarked' with mode
mode_embarked = df1['Embarked'].mode()[0]
df1['Embarked'].fillna(mode_embarked, inplace=True)
```

In [46]:

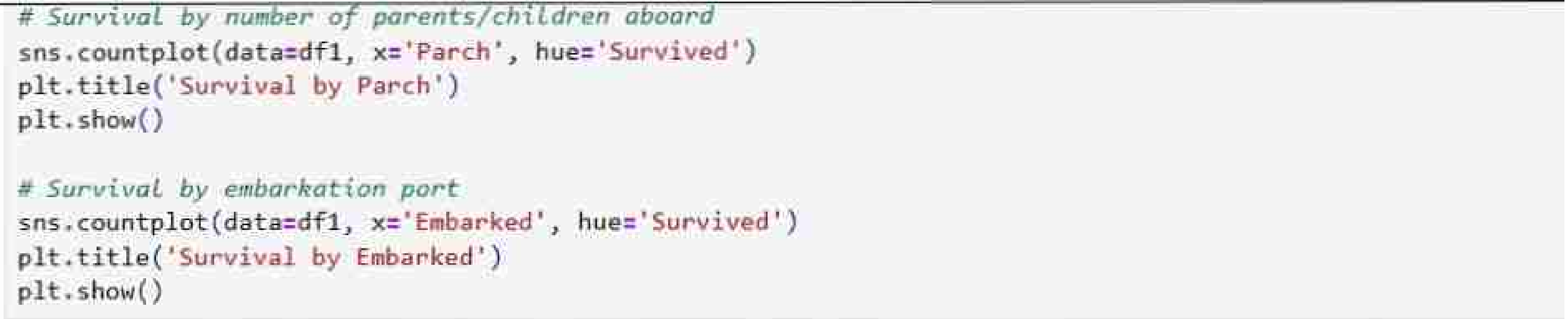
```
# Drop 'Cabin' column due to high number of missing values
df1.drop('Cabin', axis=1, inplace=True)



df1.isnull().sum()
```

Out[46]:

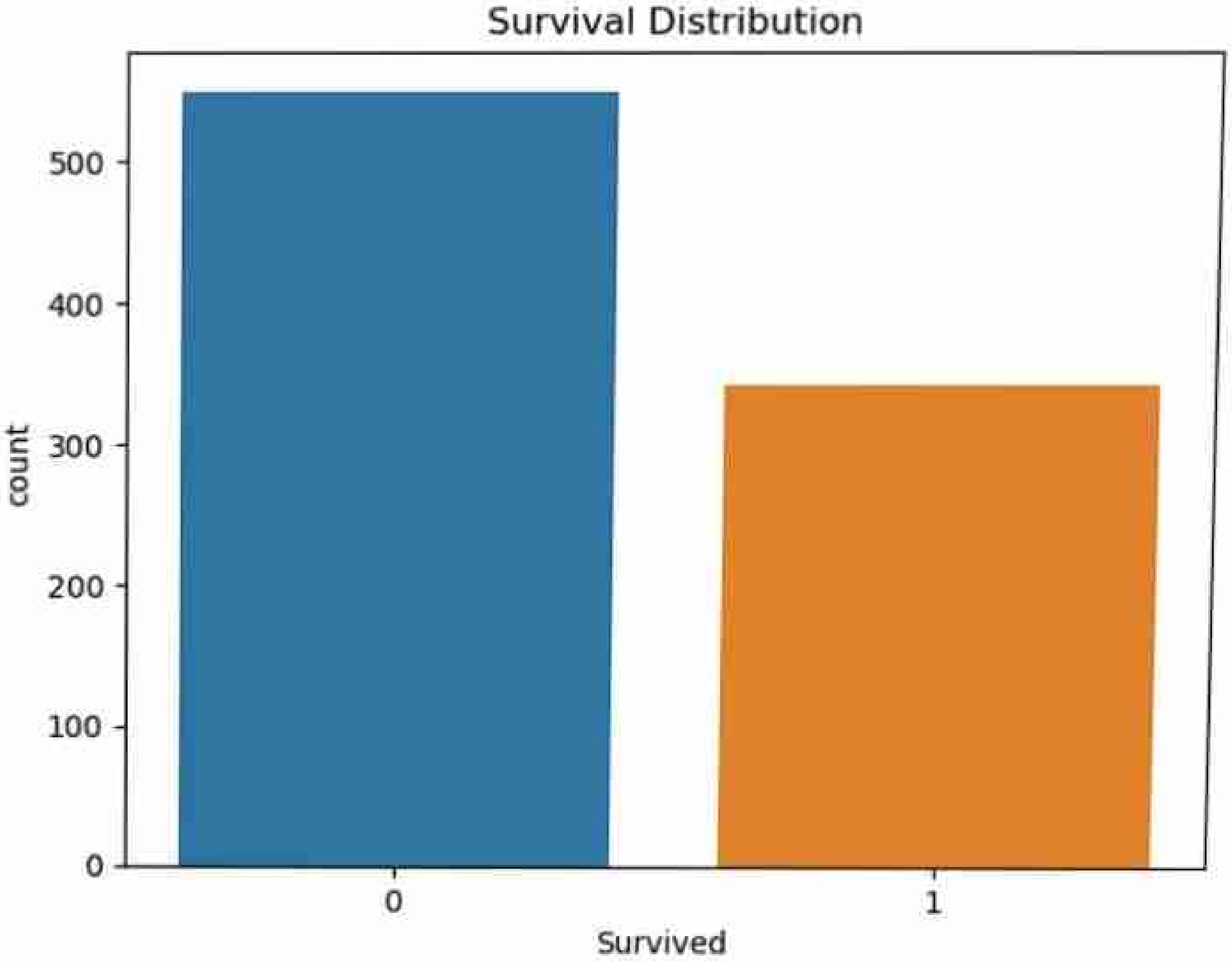
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

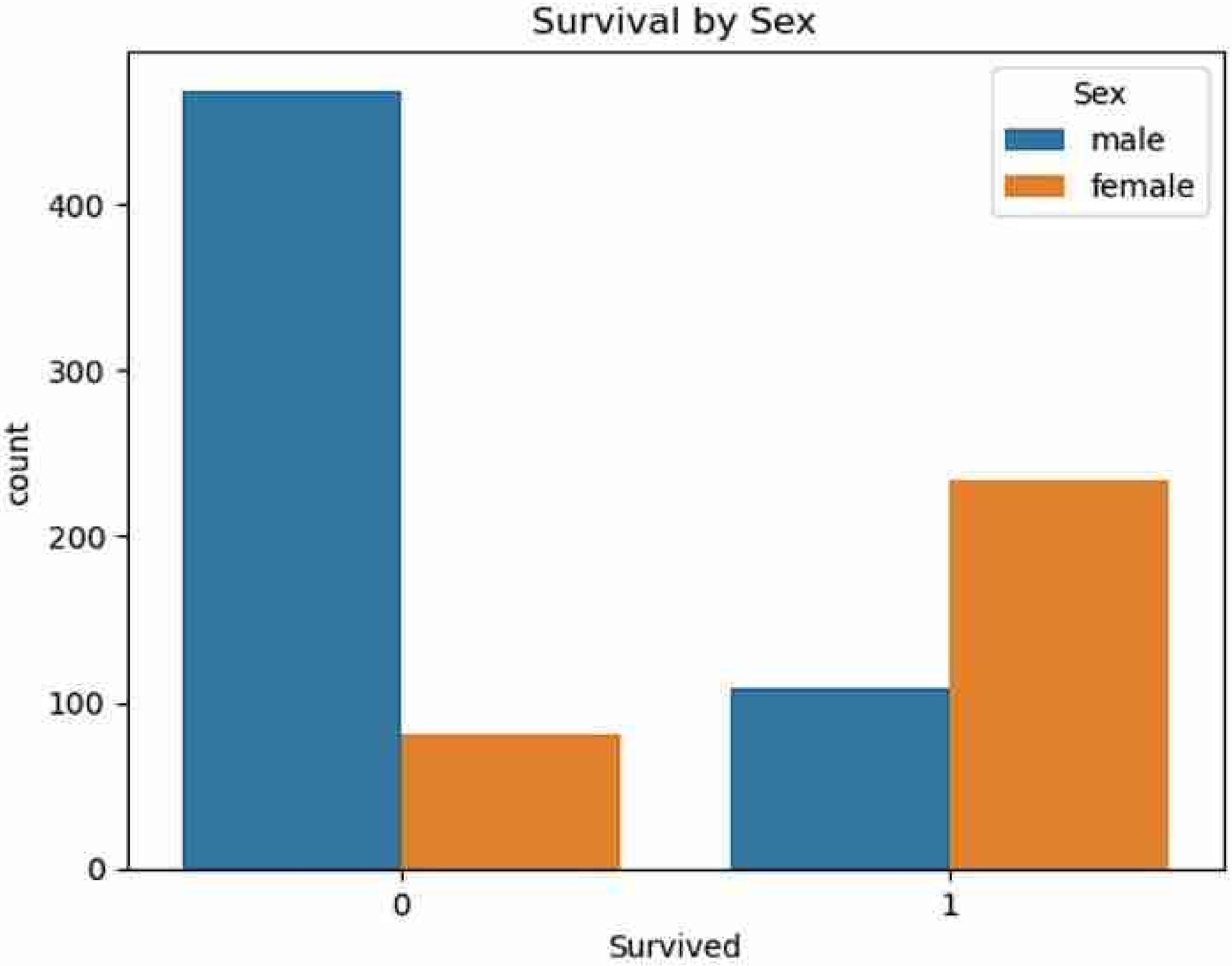


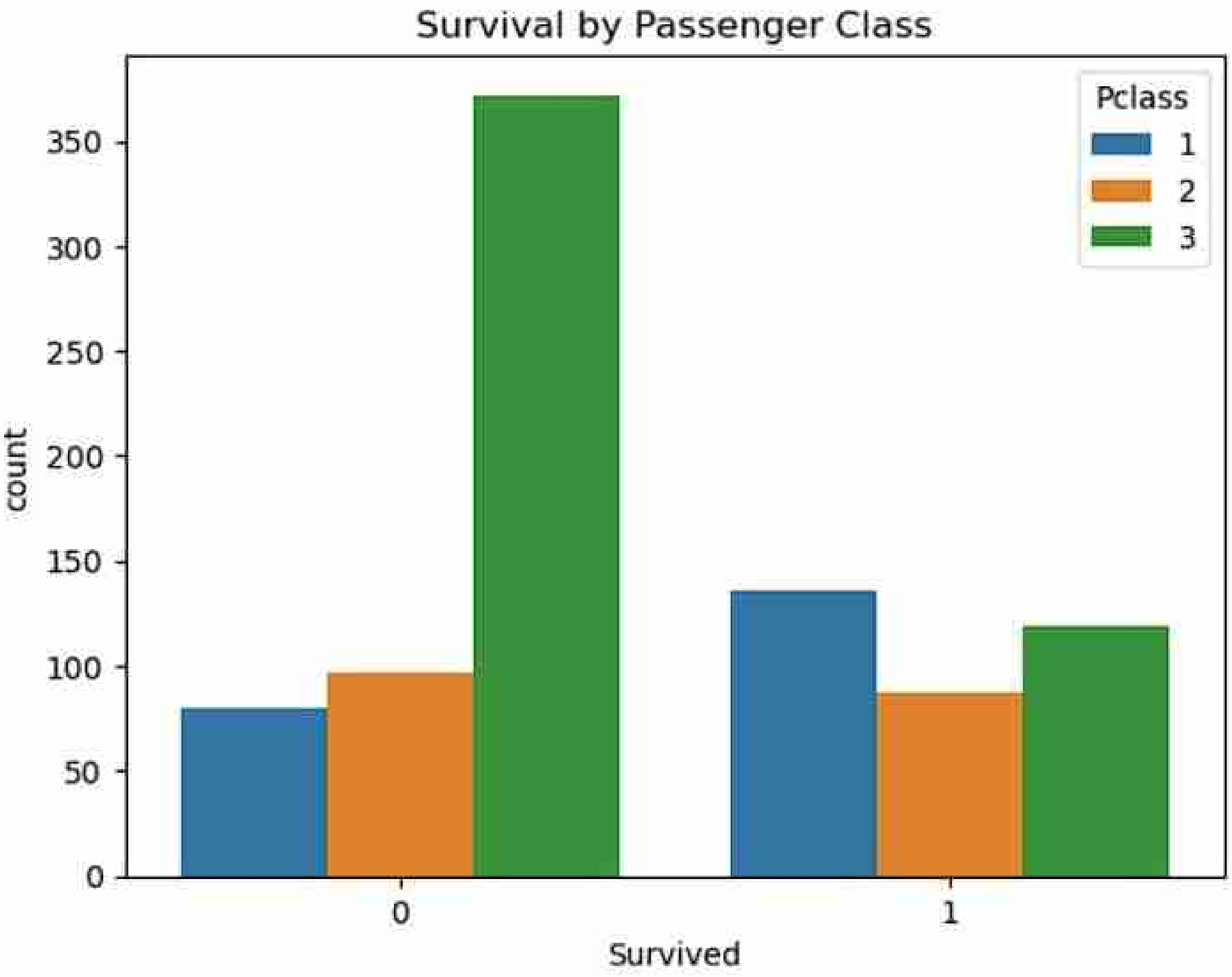
```
import seaborn as sns
# Distribution of survival
sns.countplot(data=df1, x='Survived')
plt.title('Survival Distribution')
plt.show()

# Survival by sex
sns.countplot(data=df1, x='Survived', hue='Sex')
plt.title('Survival by Sex')
plt.show()

# Survival by passenger class
sns.countplot(data=df1, x='Survived', hue='Pclass')
plt.title('Survival by Passenger Class')
plt.show()

# Age distribution
sns.histplot(data=df1, x='Age', bins=20, kde=True)
plt.title('Age Distribution')
plt.show()

# Survival by age
sns.histplot(data=df1, x='Age', bins=20, kde=True, hue='Survived')
plt.title('Survival by Age')
plt.show()

# Fare distribution
sns.histplot(data=df1, x='Fare', bins=20, kde=True)
plt.title('Fare Distribution')
plt.show()

# Survival by fare
sns.histplot(data=df1, x='Fare', bins=20, kde=True, hue='Survived')
plt.title('Survival by Fare')
plt.show()

# Survival by number of siblings/spouses aboard
sns.countplot(data=df1, x='SibSp', hue='Survived')
plt.title('Survival by SibSp')
plt.show()

# Survival by number of parents/children aboard
sns.countplot(data=df1, x='Parch', hue='Survived')
plt.title('Survival by Parch')
plt.show()
```

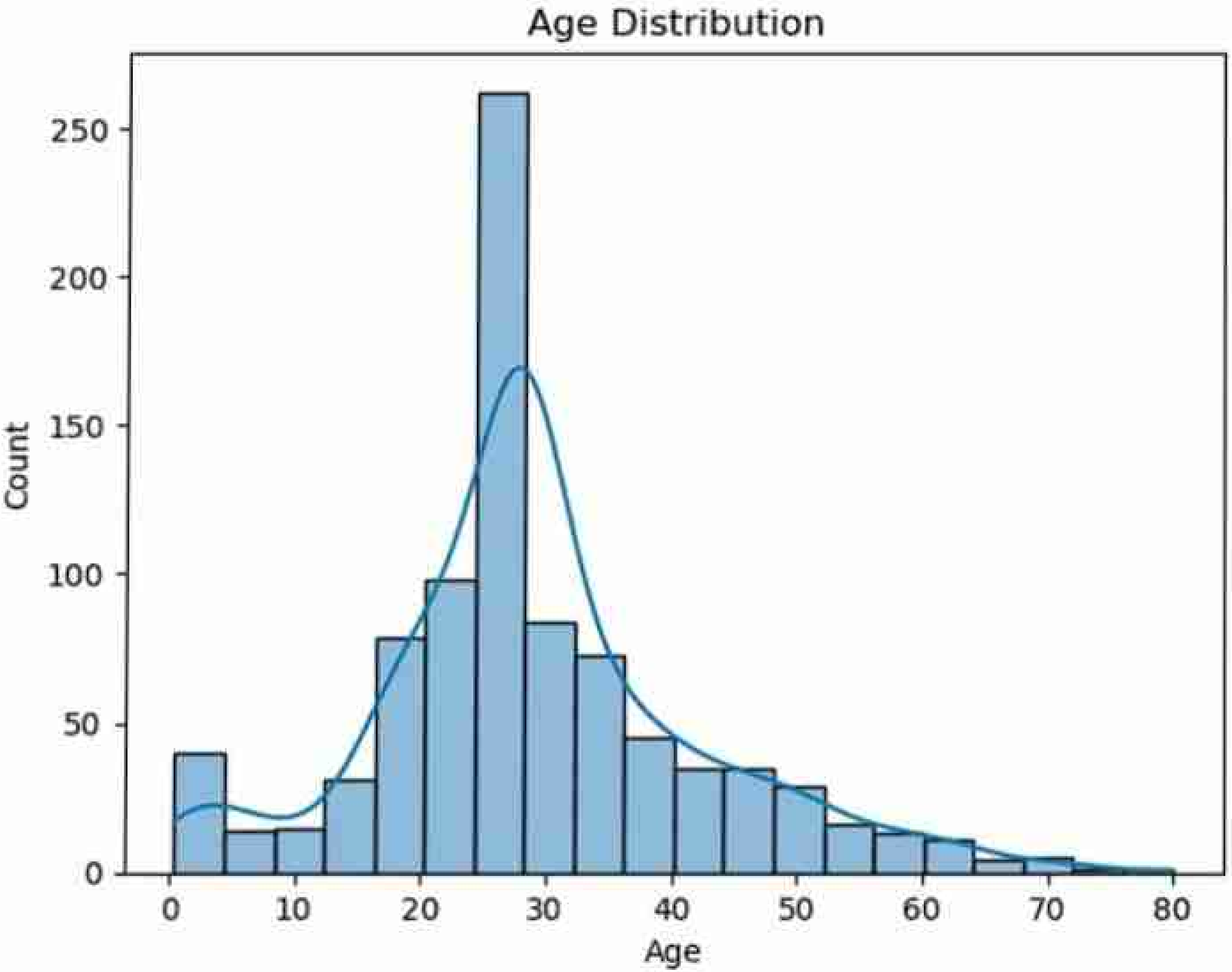


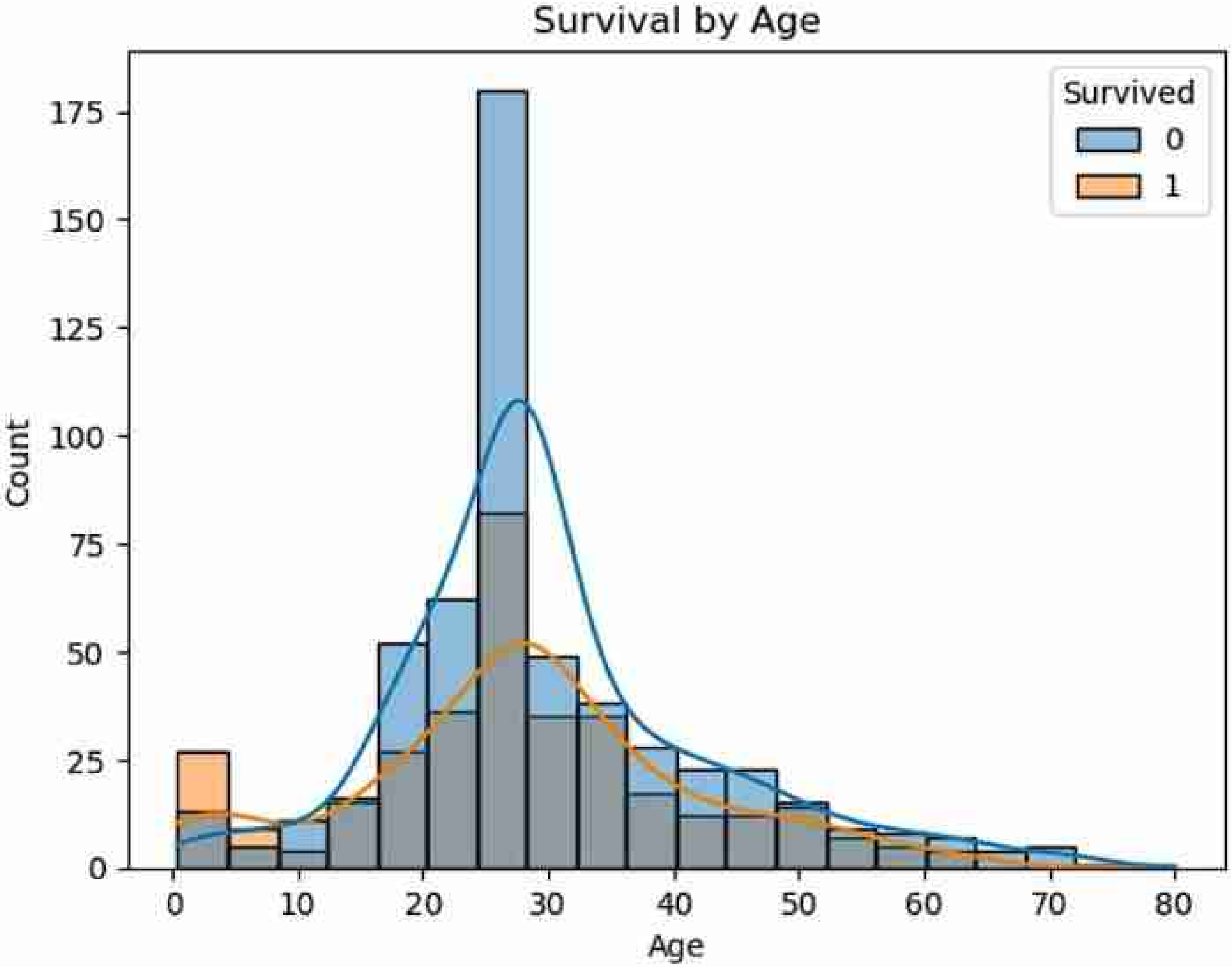
```
# Survival by number of parents/children aboard
sns.countplot(data=df1, x='Parch', hue='Survived')
plt.title('Survival by Parch')
plt.show()

# Survival by embarkation port
sns.countplot(data=df1, x='Embarked', hue='Survived')
plt.title('Survival by Embarked')
plt.show()
```

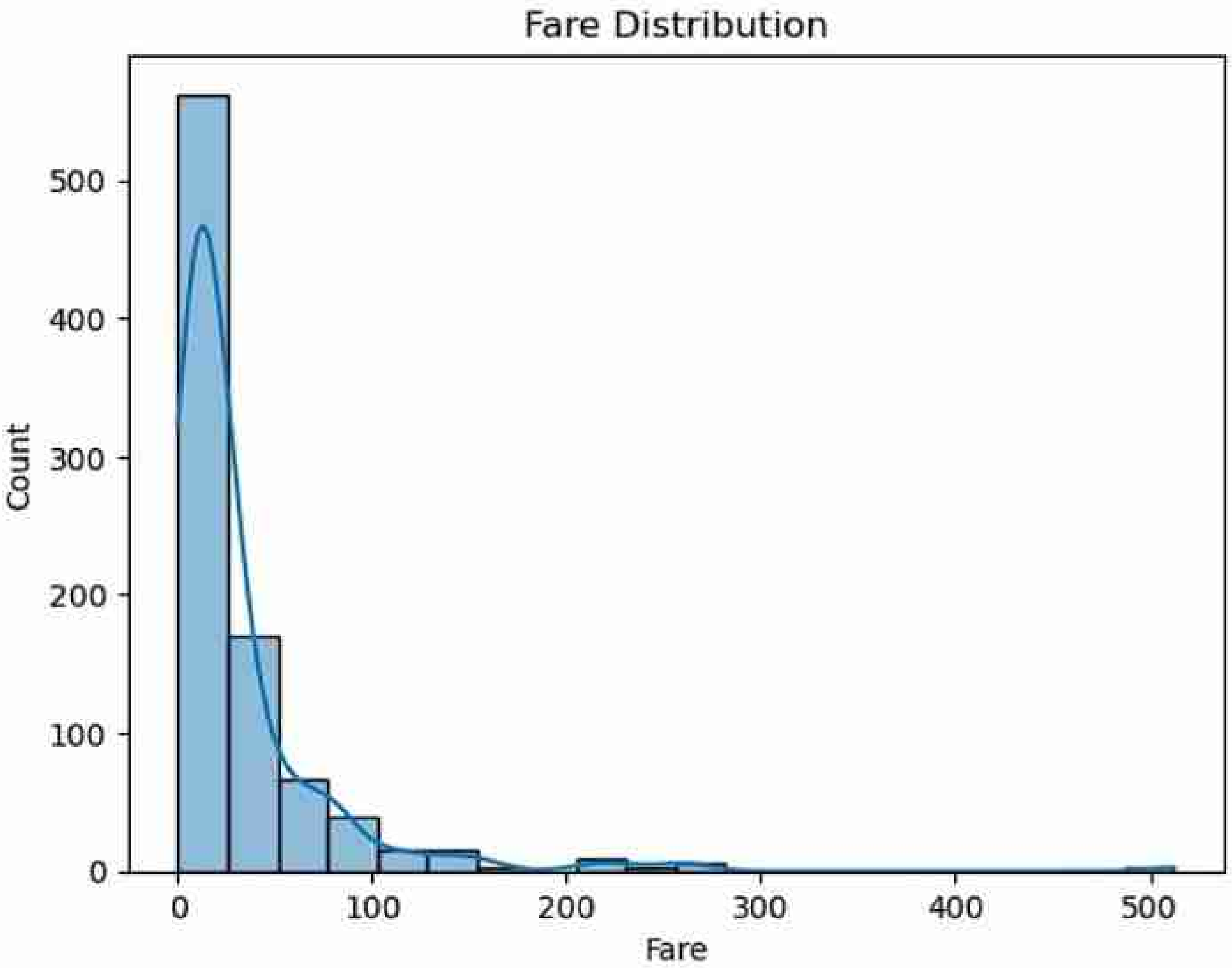


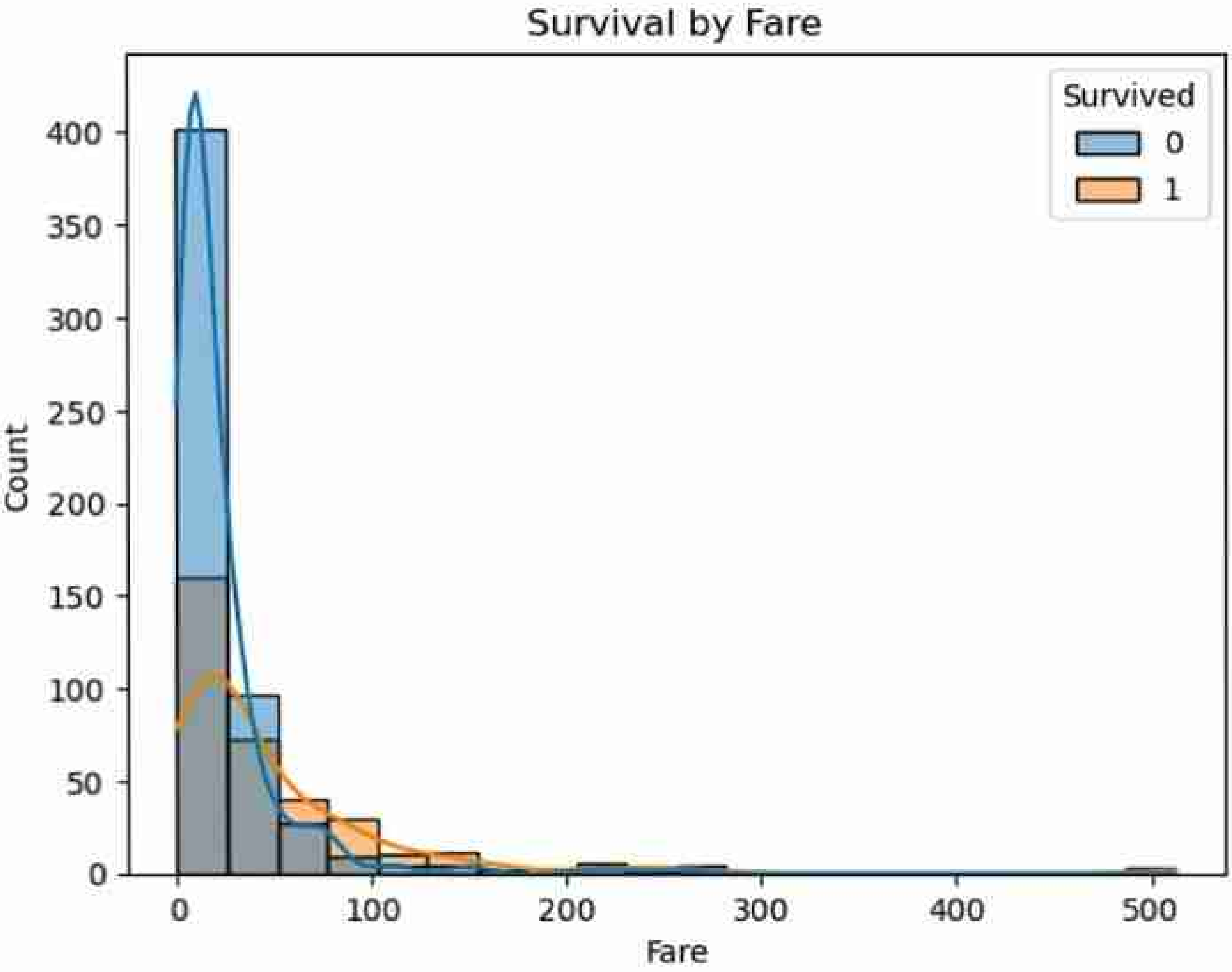

Survival by Sex
Sex
male
female
count
400
300
200
100
0
0
1
Survived

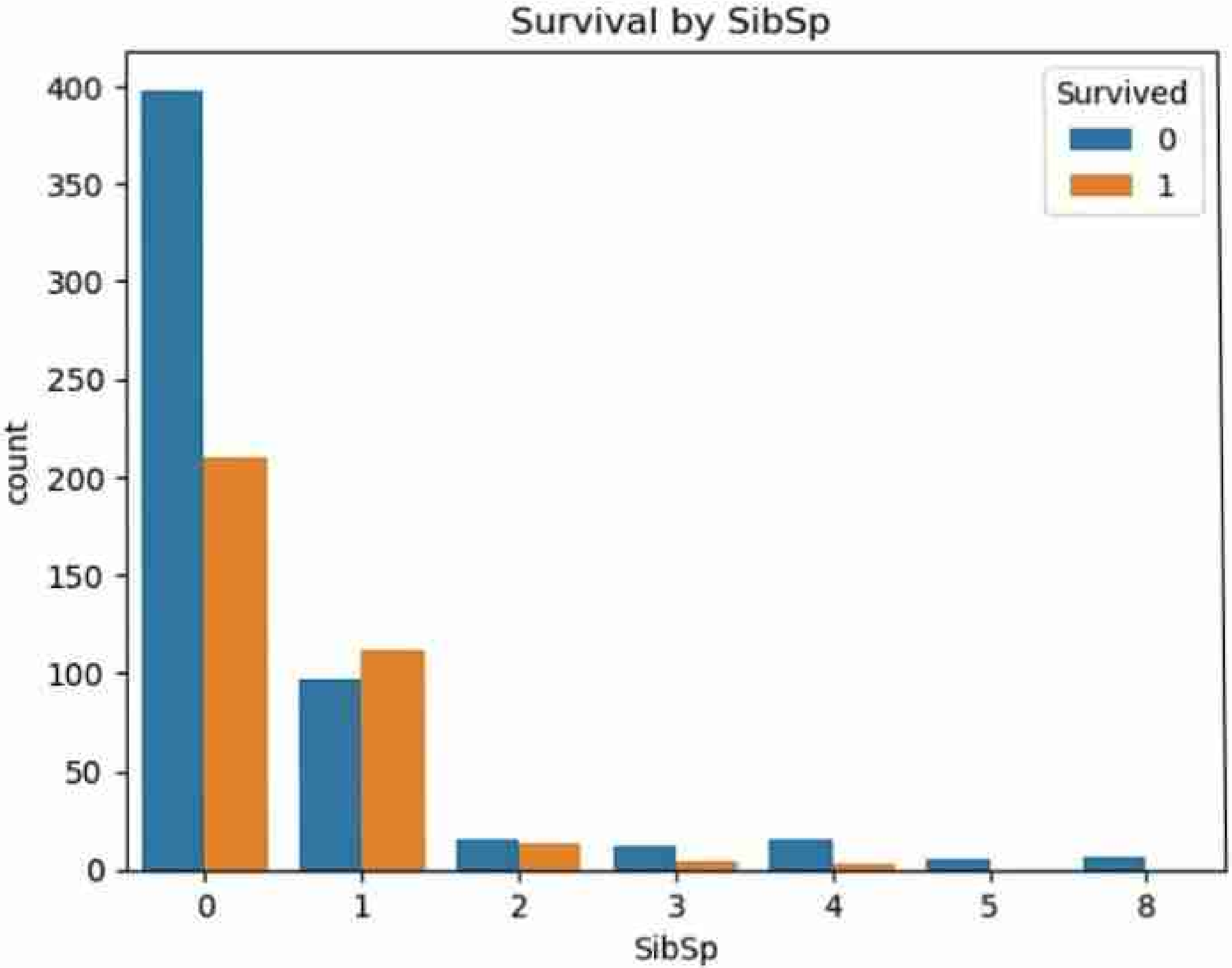

Survival by Passenger Class
Pclass
1
2
3
count
0
50
100
150
200
250
300
350
0
1
Survived

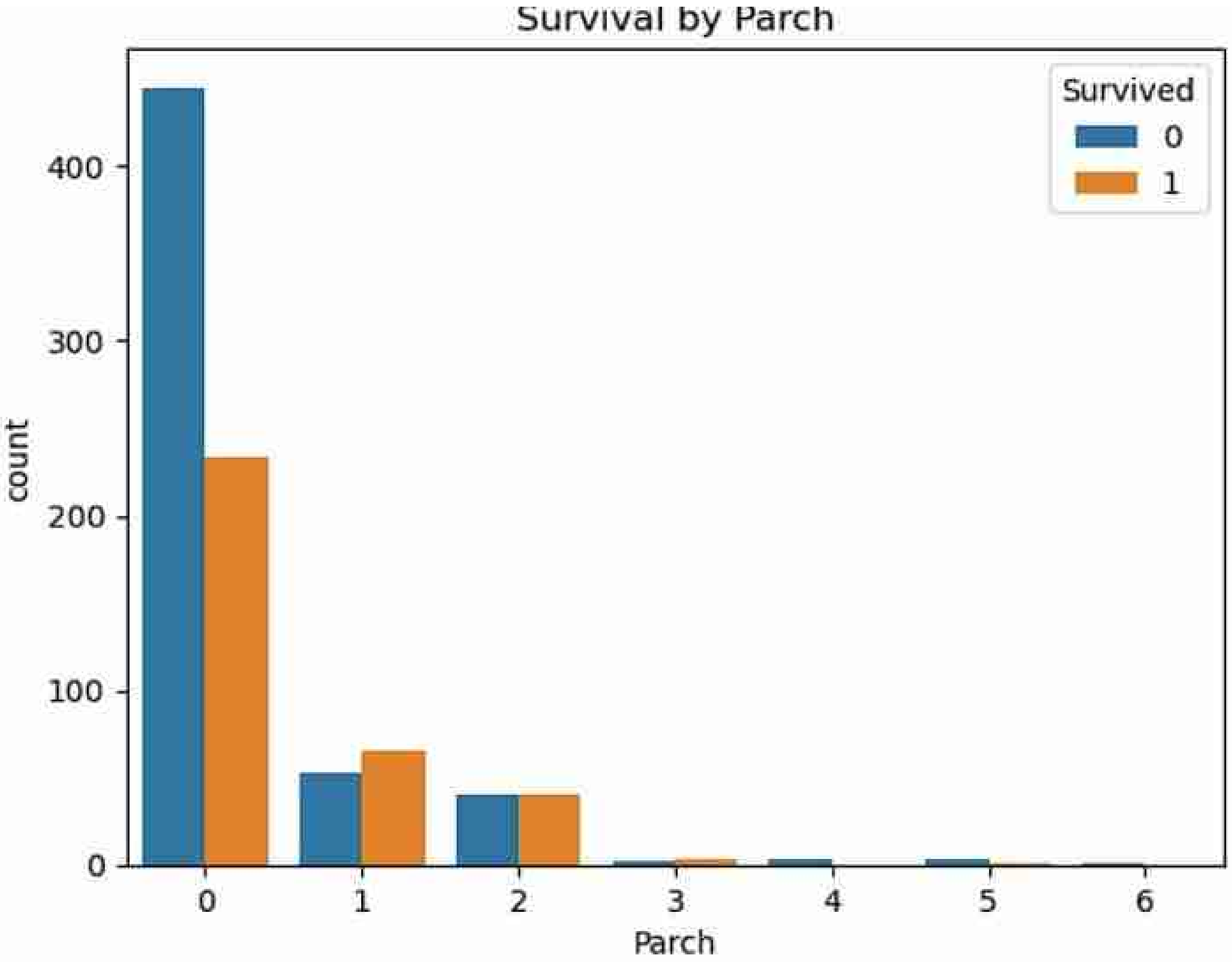

Age Distribution
Count
250
200
150
100
50
0
0
10
20
30
40
50
60
70
80
Age

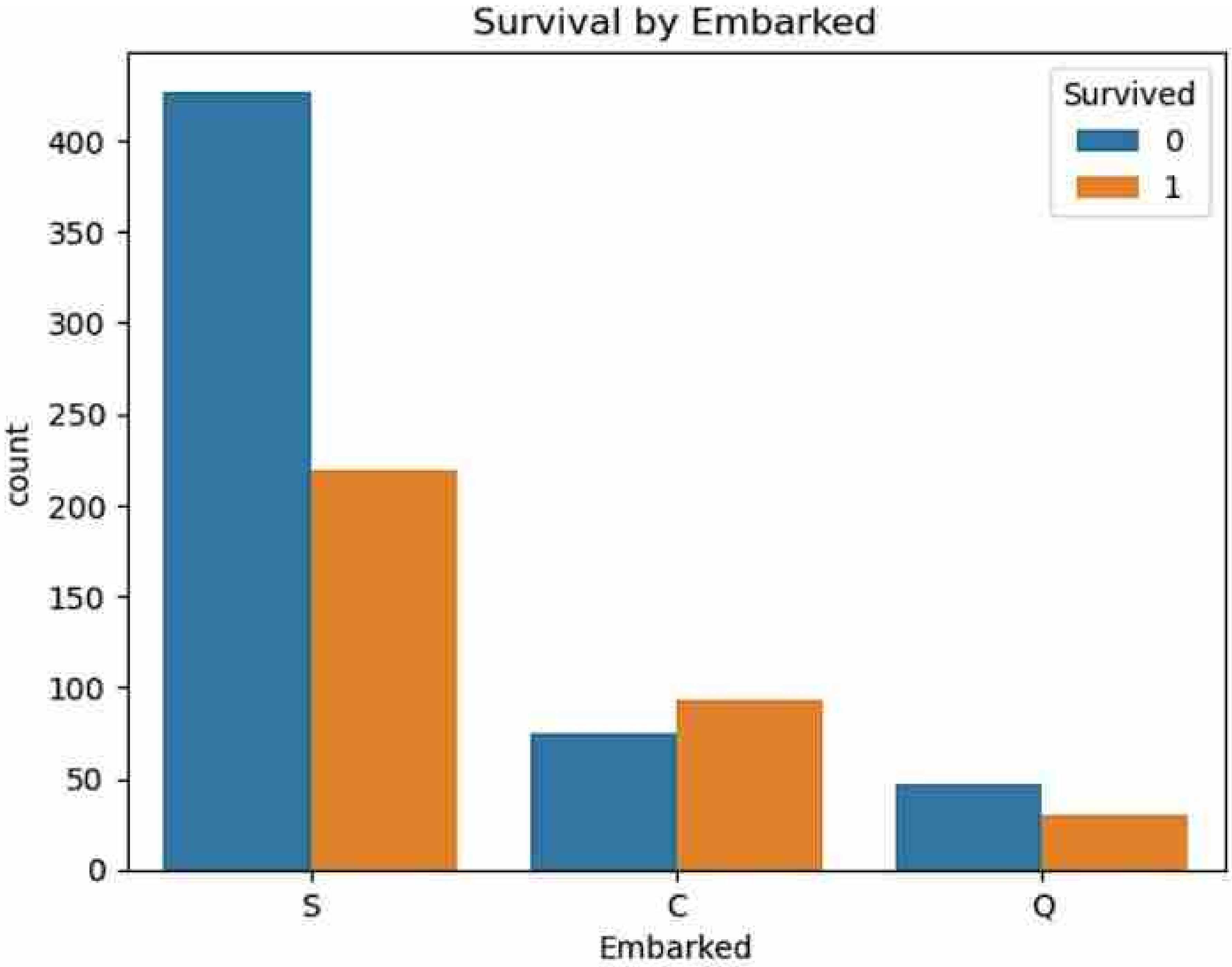

Survival by Age
Survived
0
1
Count
175
150
125
100
75
50
25
0
0
10
20
30
40
50
60
70
80
Age


Fare Distribution
Count
0
100
200
300
400
500
Fare
0
100
200
300
400
500


Survival by Fare
Survived
0
1
Count
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
Fare


Survival by SibSp
Survived
0
1
count
400
350
300
250
200
150
100
50
0
SibSp
0
1
2
3
4
5
8


Survival by Parch
Survived
0
1
400
300
200
100
0
count
0
1
2
3
4
5
6
Parch


Survival by Embarked
Survived
🟦 0
🟧 1
count
0
50
100
150
200
250
300
350
400
S
C
Q
Embarked

[ ]: