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


df = pd.read_csv('Titanic.csv')


df.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
```

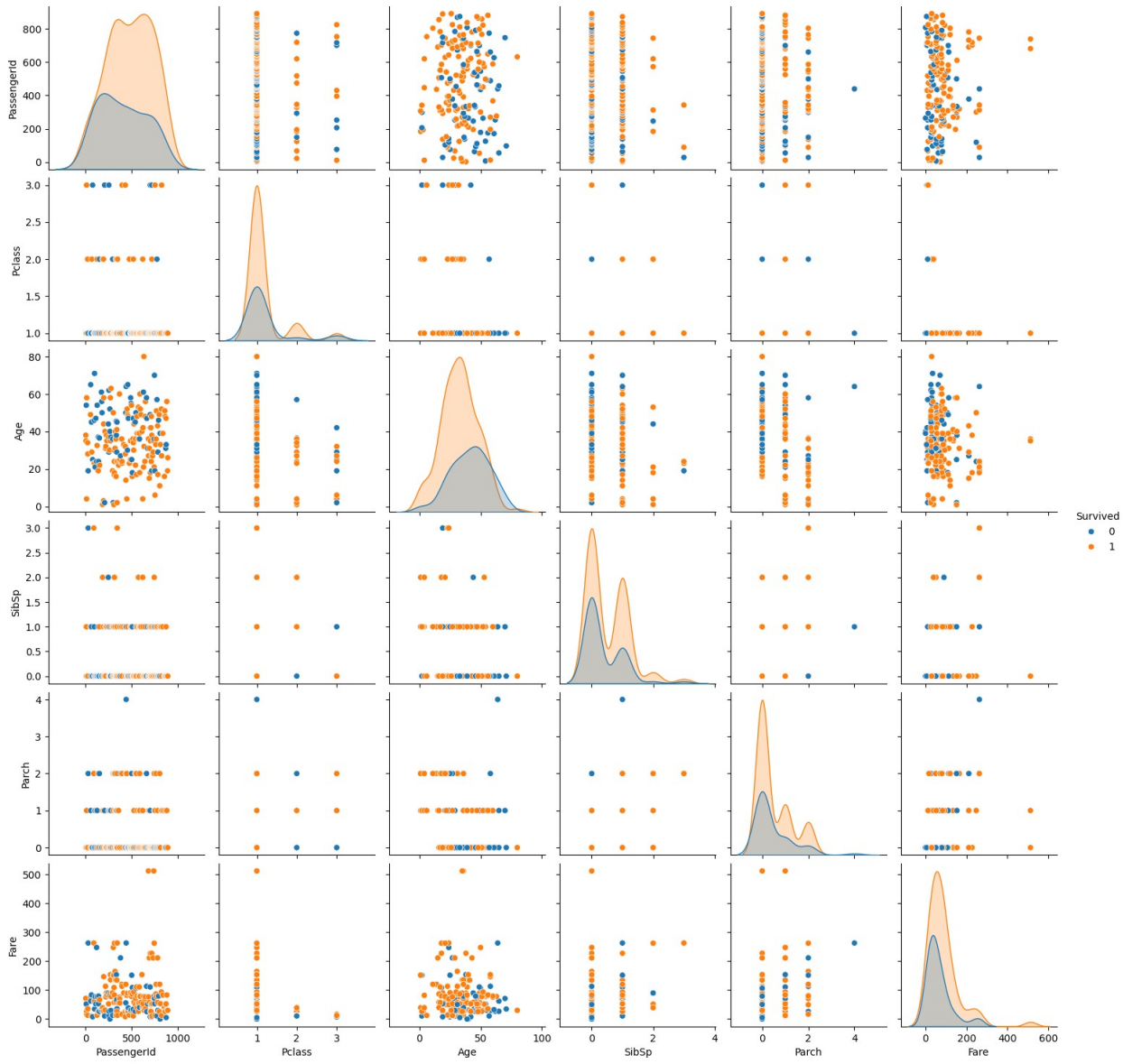
```
df.info()


df.describe(include='all')
df.isnull().sum()

for col in df.select_dtypes(include='object').columns:
    print(f"\n{col}:\n", df[col].value_counts())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
```

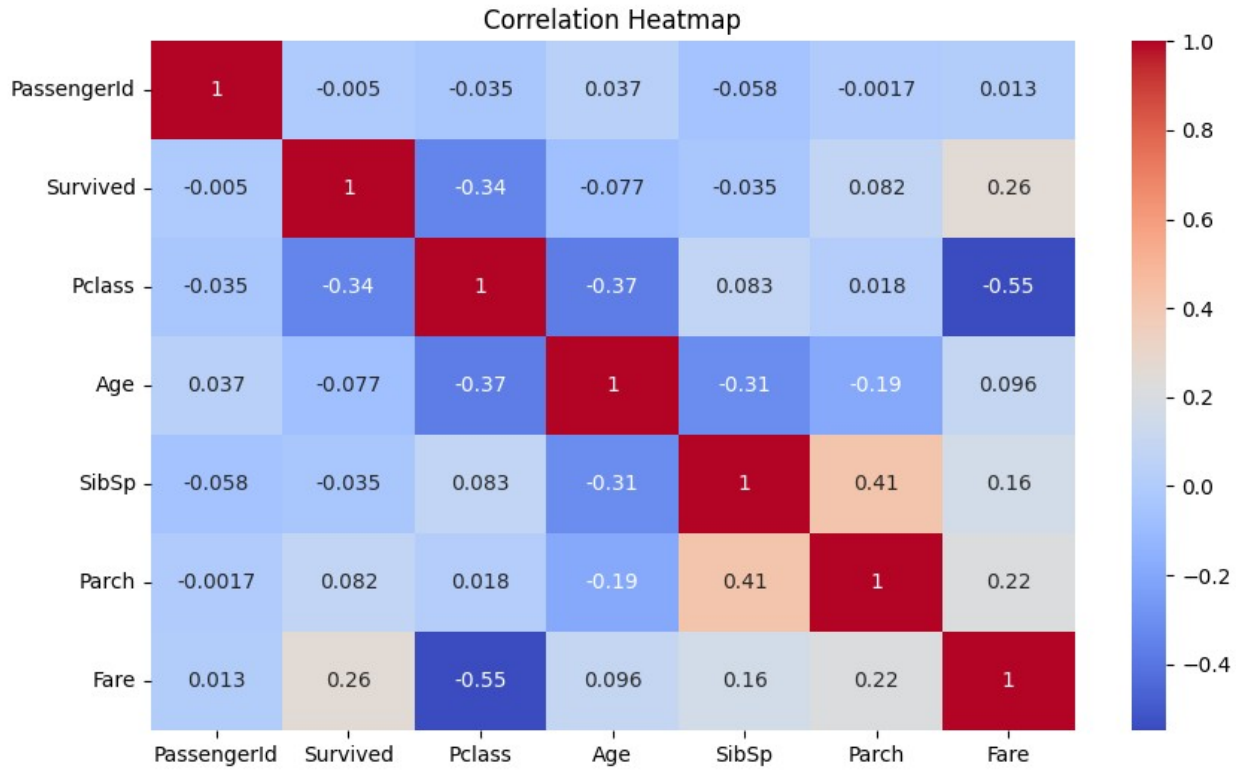
```
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Survived     891 non-null    int64
 2   Pclass       891 non-null    int64
 3   Name         891 non-null    object
 4   Sex          891 non-null    object
 5   Age          714 non-null    float64
 6   SibSp        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

Name:
 Name
Dooley, Mr. Patrick                                   1
Braund, Mr. Owen Harris                               1
Cumings, Mrs. John Bradley (Florence Briggs Thayer)   1
Heikkinen, Miss. Laina                                1
Futrelle, Mrs. Jacques Heath (Lily May Peel)          1
                                                     ..
Hewlett, Mrs. (Mary D Kingcome)                       1
Vestrom, Miss. Hulda Amanda Adolfina                  1
Andersson, Mr. Anders Johan                           1
Saundercock, Mr. William Henry                        1
Bonnell, Miss. Elizabeth                              1
Name: count, Length: 891, dtype: int64

Sex:
 Sex
male      577
female    314
Name: count, dtype: int64

Ticket:
 Ticket
347082              7
1601                7
CA. 2343            7
3101295             6
CA 2144             6
                   ..
PC 17590            1
17463               1
330877              1
373450              1
STON/O2. 3101282    1
```

```
Name: count, Length: 681, dtype: int64

Cabin:
 Cabin
G6             4
C23 C25 C27    4
B96 B98        4
F2             3
D              3
              ..
E17            1
A24            1
C50            1
B42            1
C148           1
Name: count, Length: 147, dtype: int64

Embarked:
 Embarked
S    644
C    168
Q     77
Name: count, dtype: int64
```

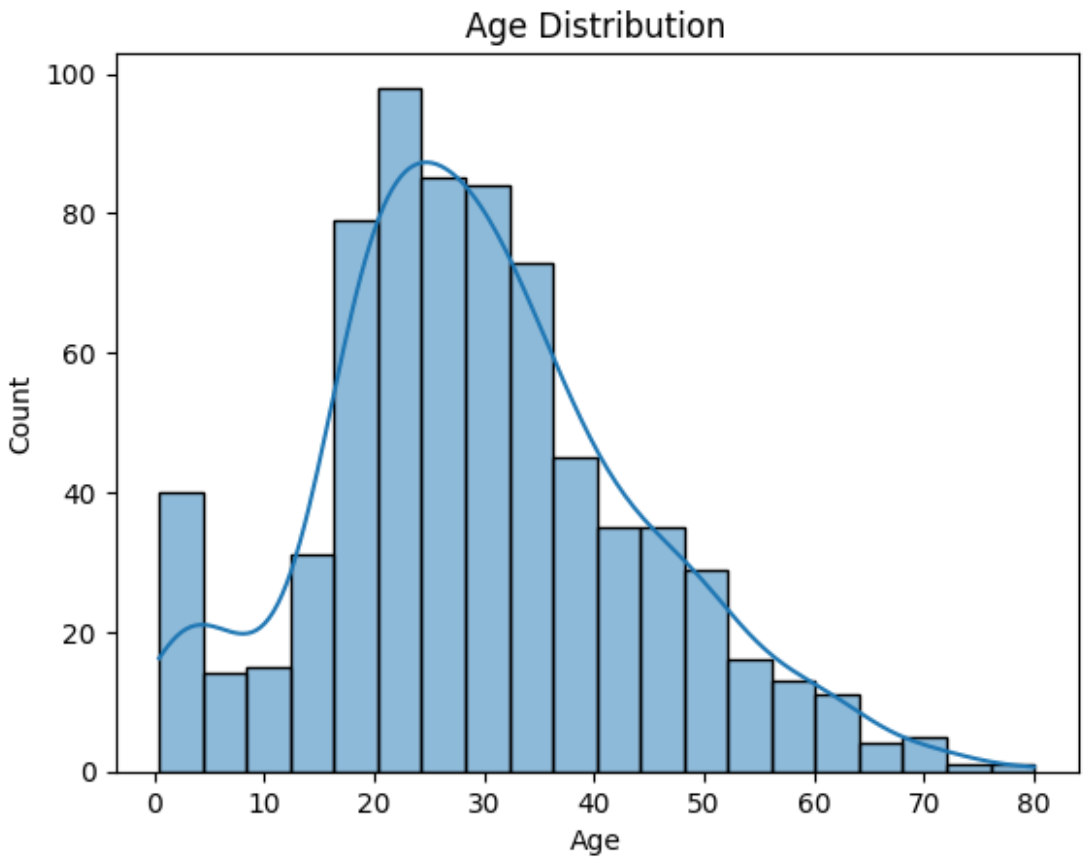
```
sns.pairplot(df.dropna(), hue='Survived')
plt.show()
```



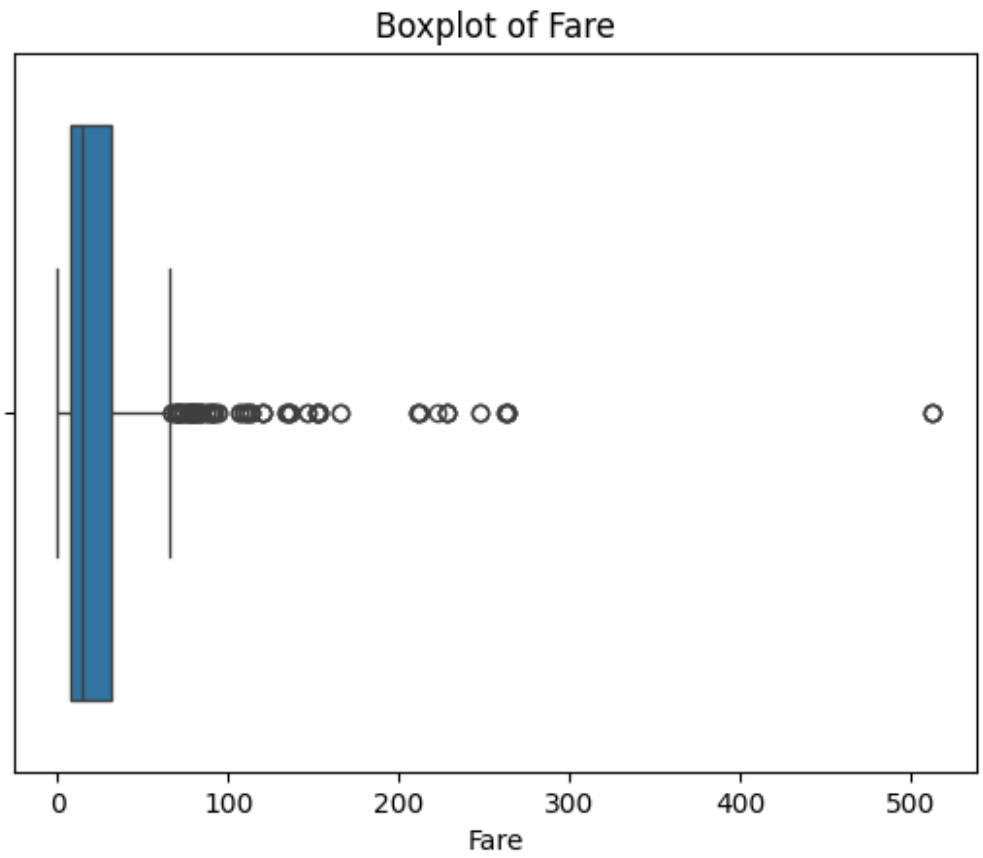

```
numeric_df = df.select_dtypes(include='number')
plt.figure(figsize=(10, 6))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()
```



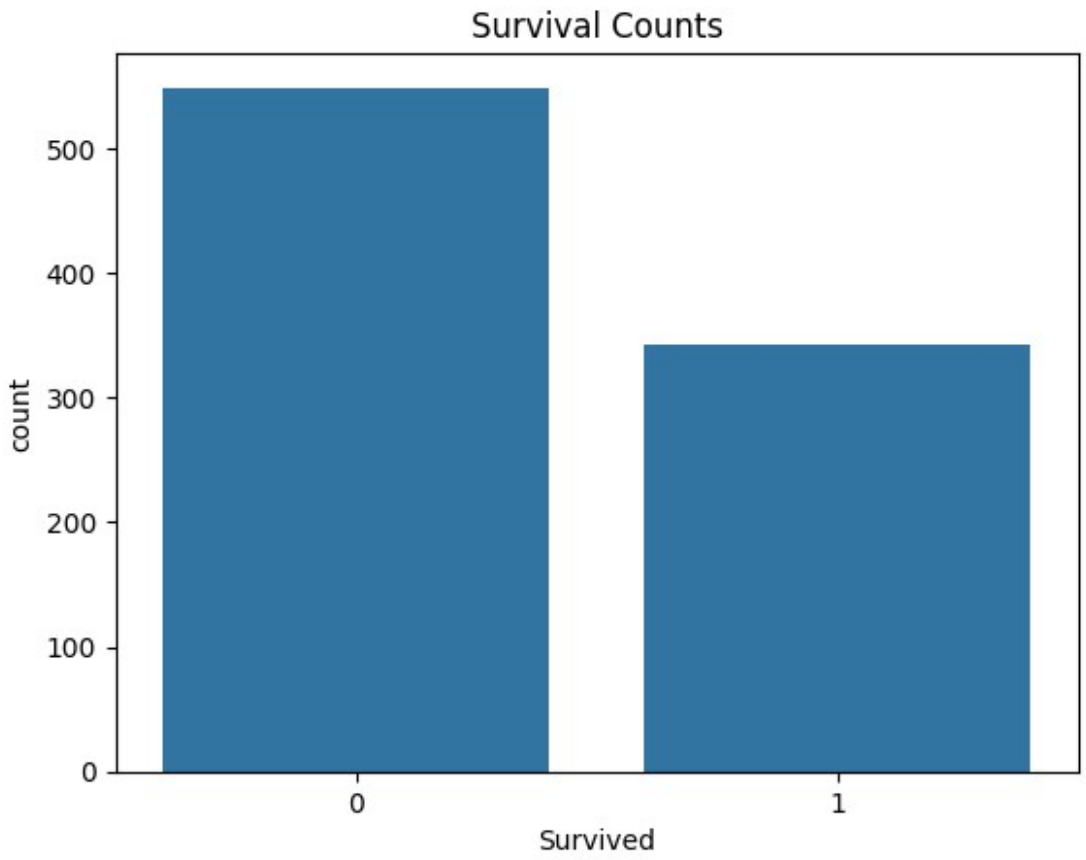

```
sns.histplot(df['Age'].dropna(), kde=True)
plt.title("Age Distribution")
plt.show()
```



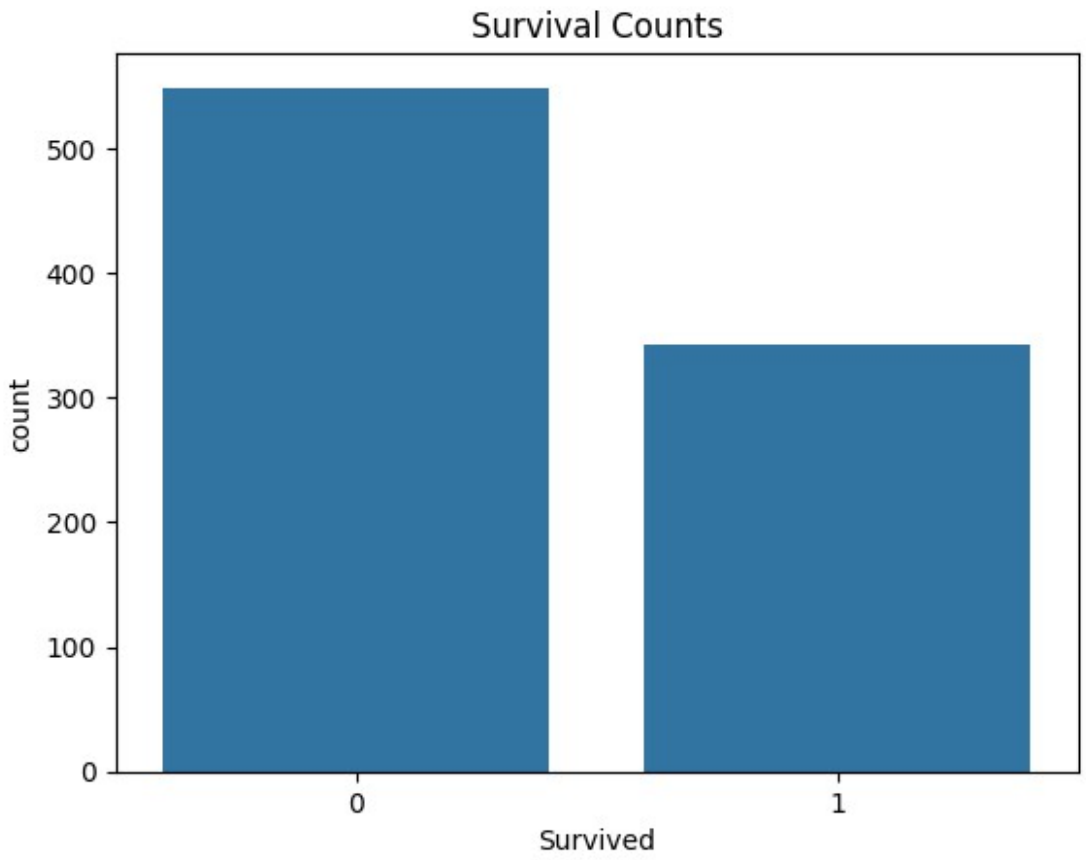

```
sns.boxplot(x='Fare', data=df)
plt.title("Boxplot of Fare")
plt.show()
```



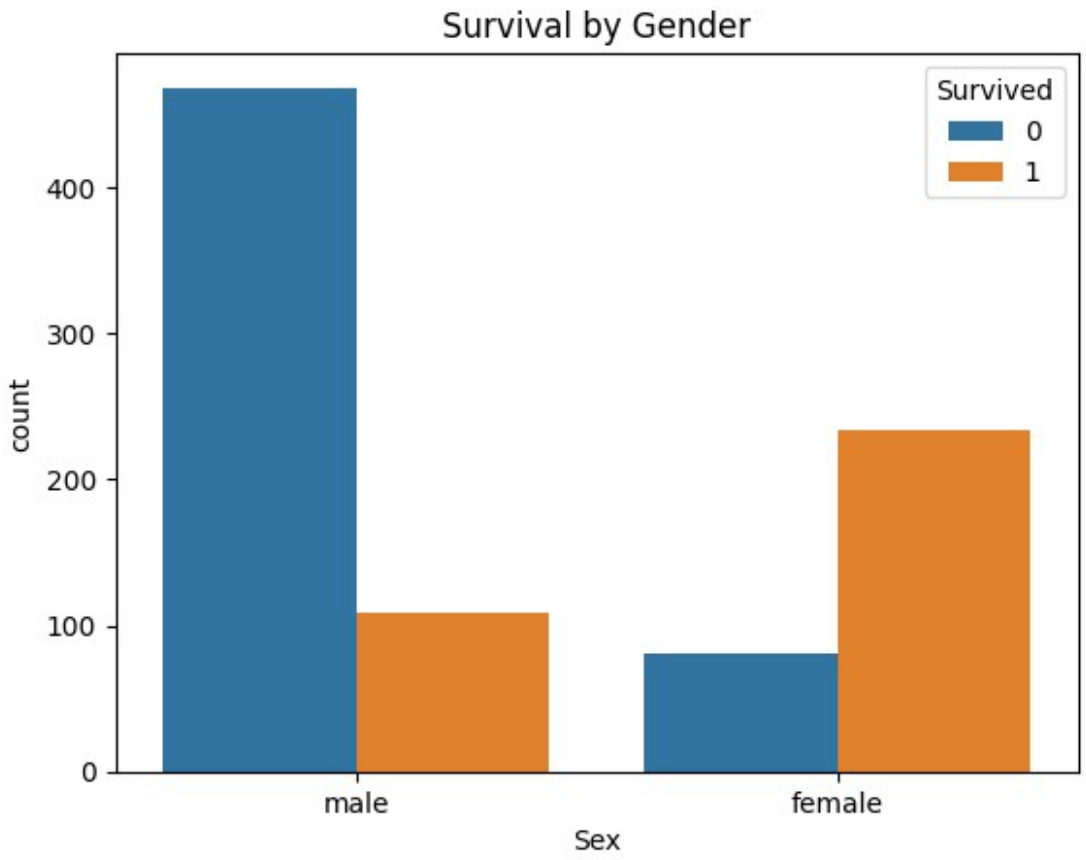

```
sns.countplot(x='Survived', data=df)
plt.title("Survival Counts")
plt.show()
```



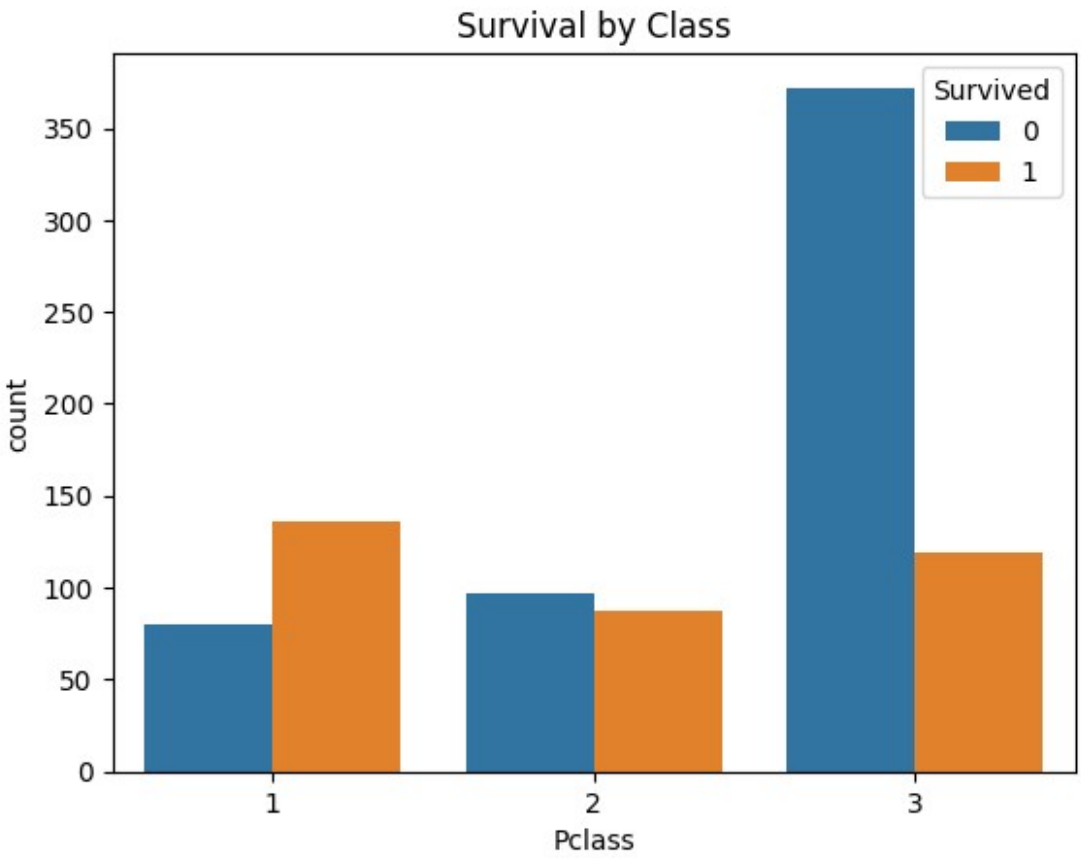

```
sns.countplot(x='Survived', data=df)
plt.title("Survival Counts")
plt.show()
```




```
sns.countplot(x='Sex', hue='Survived', data=df)
plt.title("Survival by Gender")
plt.show()
```




```
sns.countplot(x='Pclass', hue='Survived', data=df)
plt.title("Survival by Class")
plt.show()
```


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