

```
In [3]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv("C:/Users/Vaishnavi/Downloads/titanic/train.csv")
```

```
In [4]: df.info()
df.describe()
df.head()
df.isnull().sum()
```

```
<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
```

```
Out[4]: 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 [5]: df.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 [6]: `df.describe()`

Out[6]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
<b>count</b>	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
<b>mean</b>	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
<b>std</b>	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
<b>min</b>	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
<b>25%</b>	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
<b>50%</b>	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
<b>75%</b>	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
<b>max</b>	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [7]: `df['Survived'].value_counts()`

Out[7]:

```

Survived
0      549
1      342
Name: count, dtype: int64

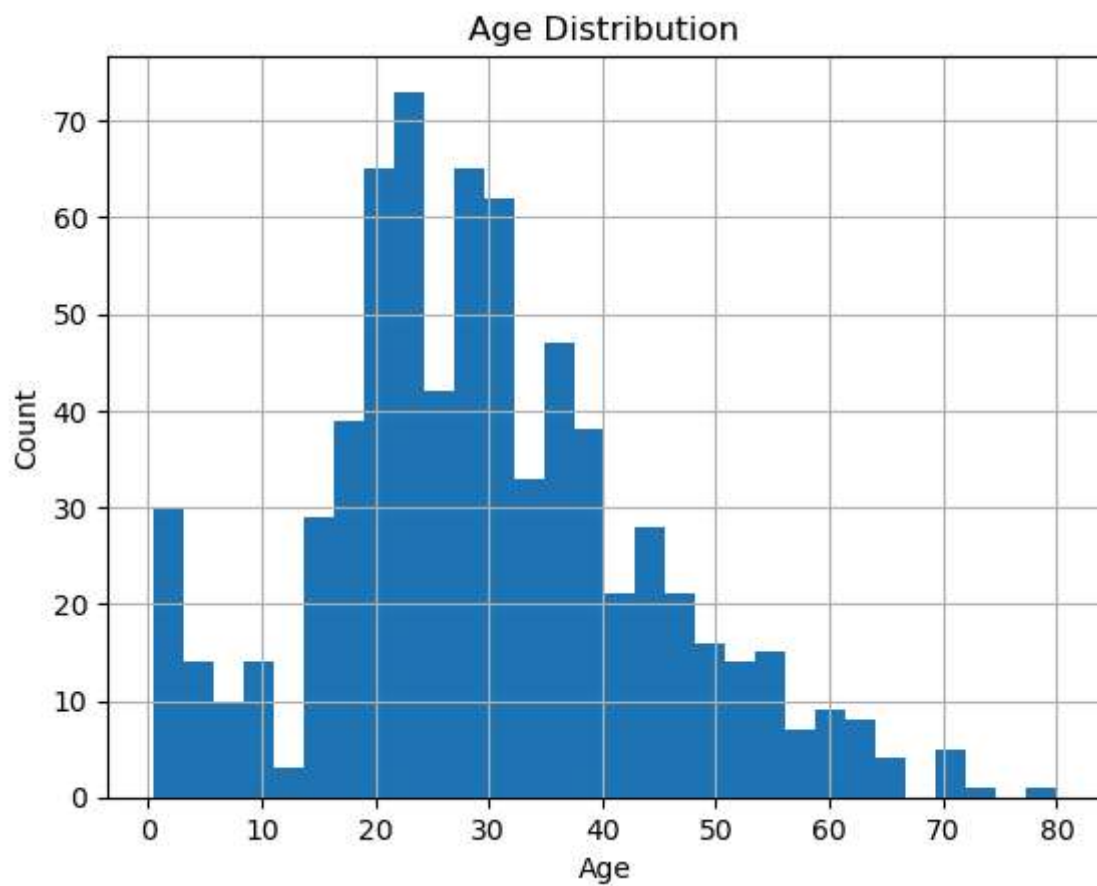
```

In [8]:

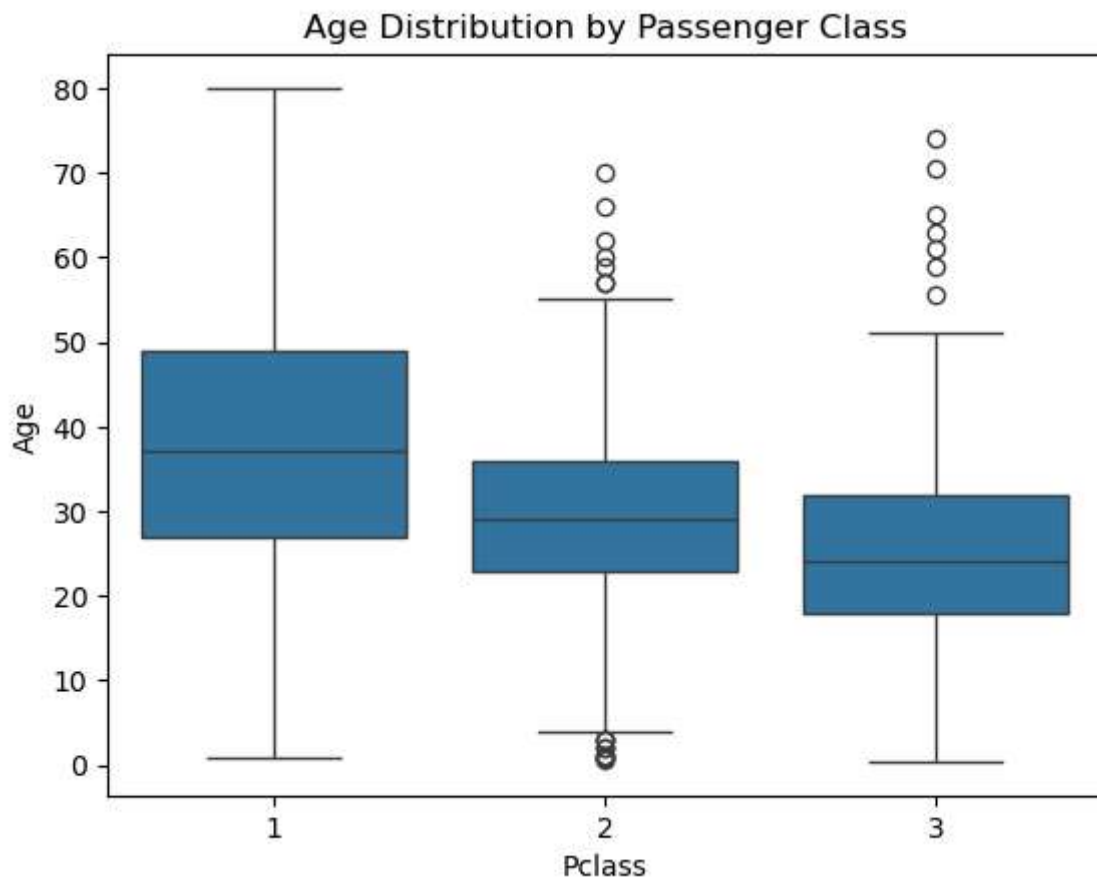
```

df['Age'].hist(bins=30)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()

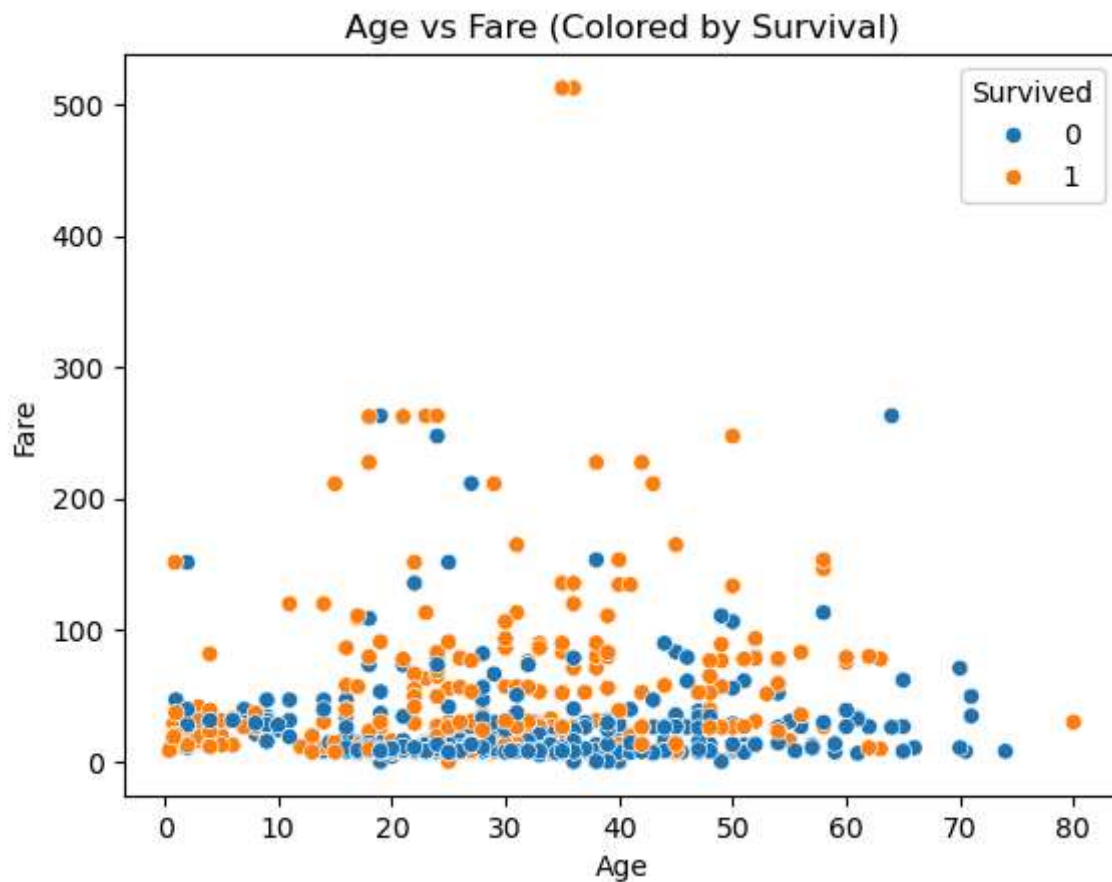
```



```
In [9]: sns.boxplot(data=df, x='Pclass', y='Age')  
plt.title('Age Distribution by Passenger Class')  
plt.show()
```

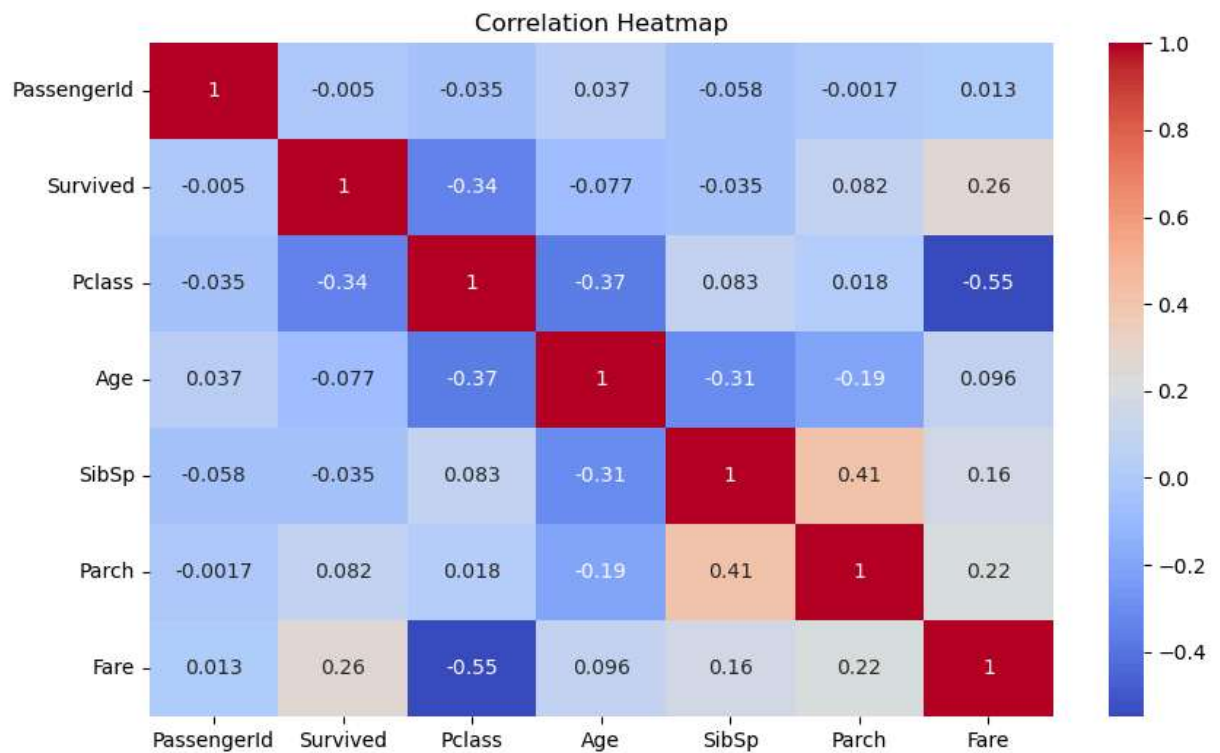


```
In [10]: sns.scatterplot(data=df, x='Age', y='Fare', hue='Survived')
plt.title('Age vs Fare (Colored by Survival)')
plt.show()
```

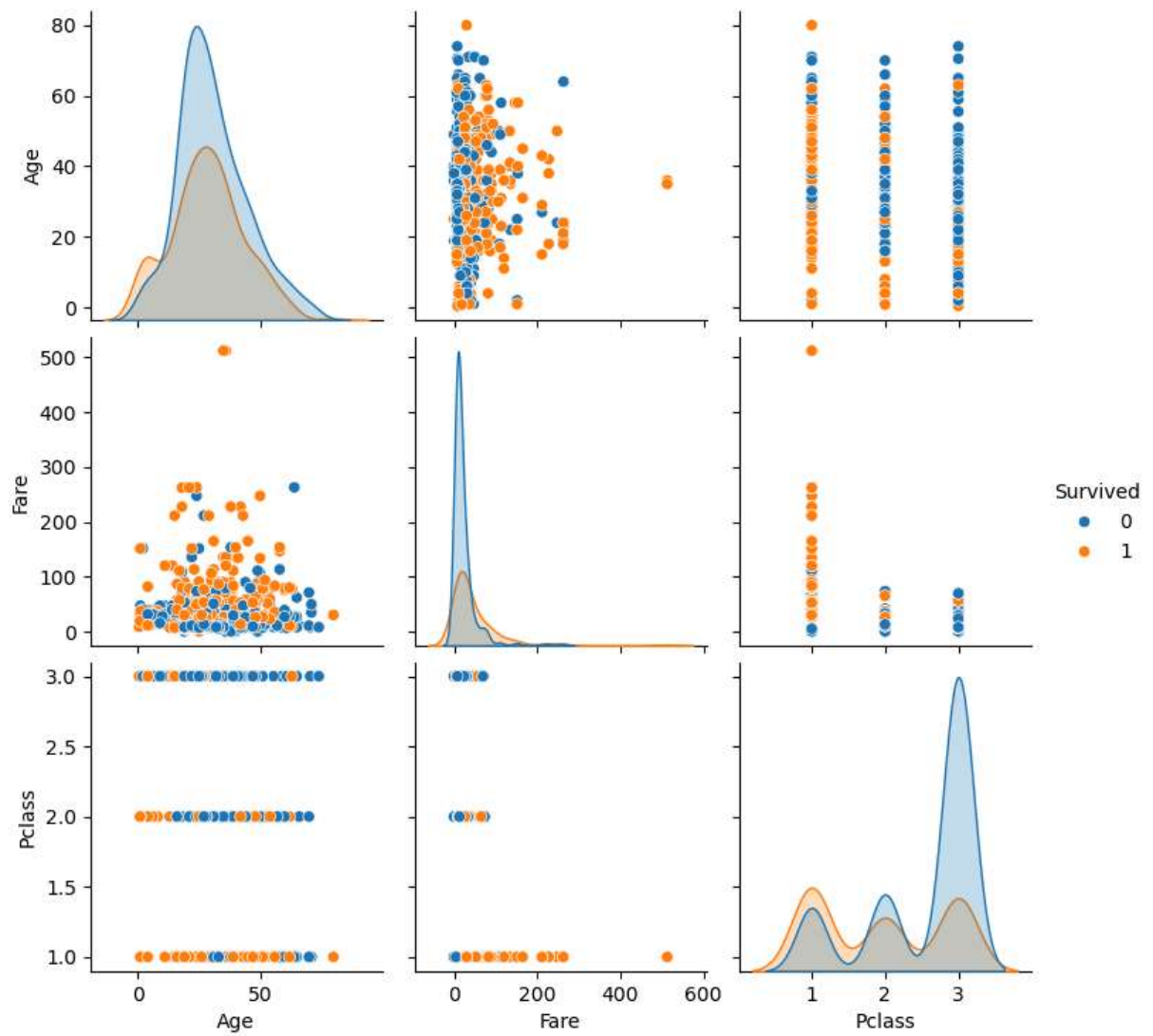


```
In [12]: # Select only numeric columns
numeric_df = df.select_dtypes(include='number')

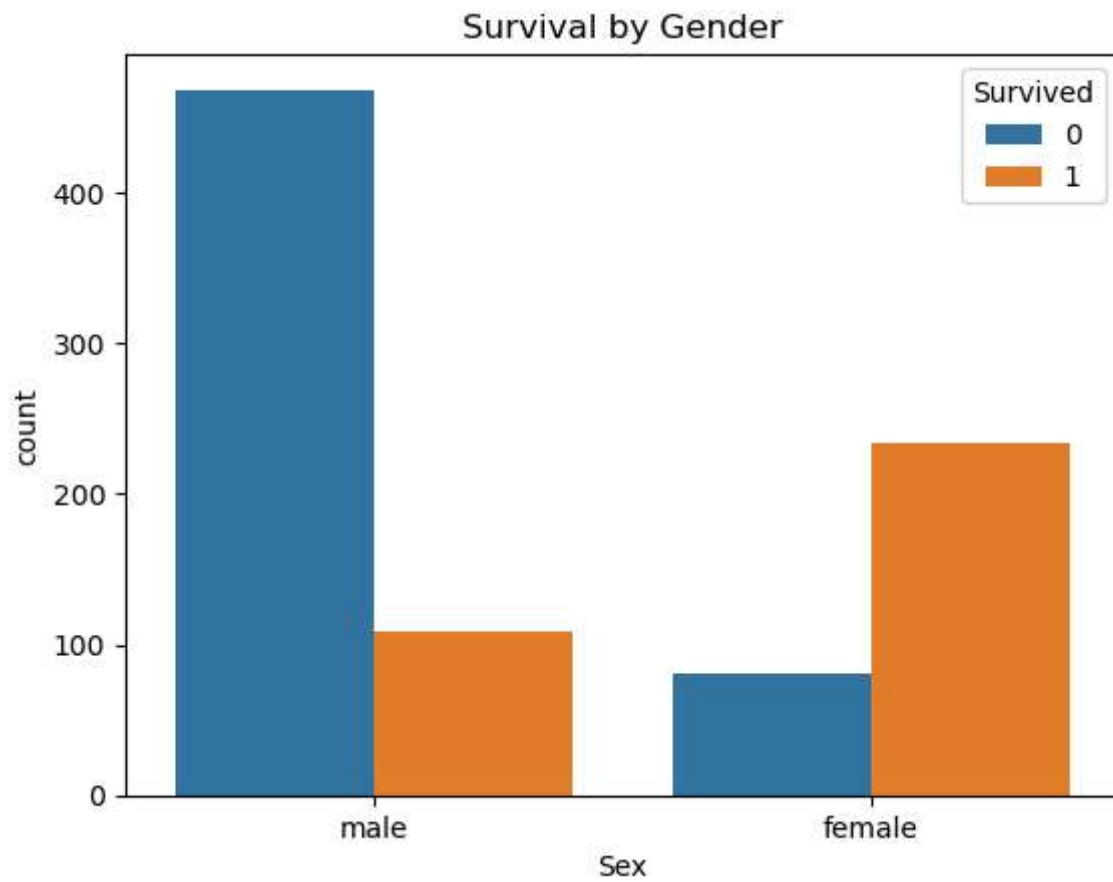
# Correlation heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()
```



```
In [13]: sns.pairplot(df[['Survived', 'Age', 'Fare', 'Pclass']], hue='Survived')  
plt.show()
```

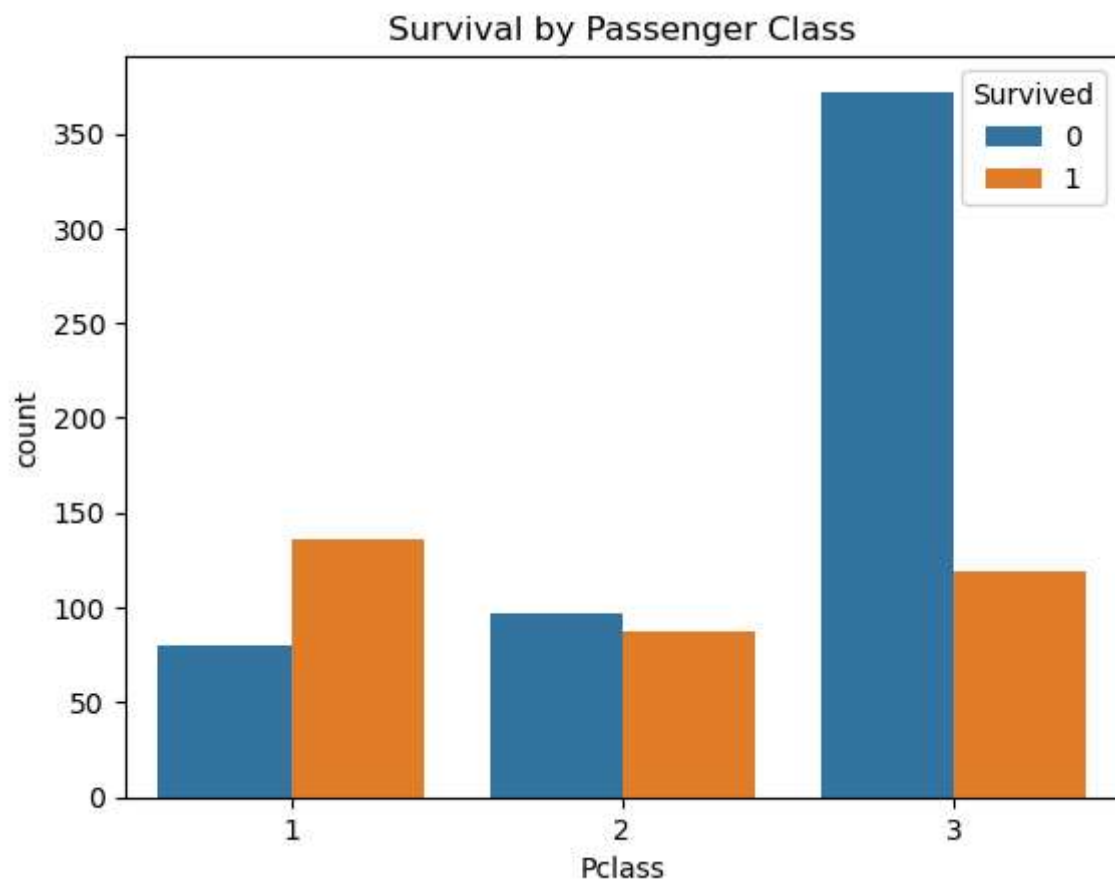


```
In [14]: sns.countplot(data=df, x='Sex', hue='Survived')
plt.title('Survival by Gender')
plt.show()
```



```
In [15]: sns.countplot(data=df, x='Pclass', hue='Survived')  
plt.title('Survival by Passenger Class')  
plt.show()
```





In [ ]: