# siddhu9

April 15, 2025

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from seaborn import load_dataset
```

[2]:
```
data = pd.read_csv("train.csv")
```

[3]:
```
data
```

[3]:
```
     PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
..           …       …     …
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         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
..                                                   …     …   …    …
886                              Montvila, Rev. Juozas    male  27.0      0
887                       Graham, Miss. Margaret Edith  female  19.0      0
888           Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1
889                              Behr, Mr. Karl Howell    male  26.0      0
890                                Dooley, Mr. Patrick    male  32.0      0

     Parch            Ticket     Fare Cabin Embarked
```

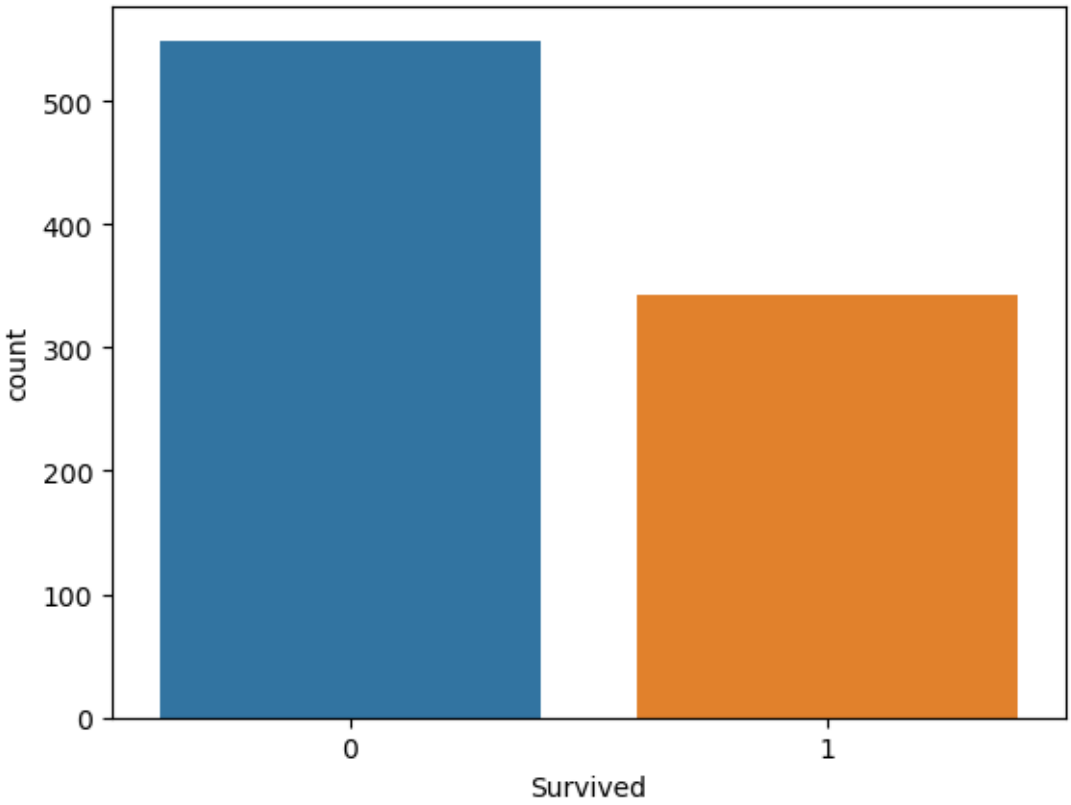
```
0        0         A/5 21171   7.2500   NaN        S
1        0          PC 17599  71.2833   C85        C
2        0  STON/O2. 3101282   7.9250   NaN        S
3        0            113803  53.1000  C123        S
4        0            373450   8.0500   NaN        S
..     …               …       …   …      …
886      0            211536  13.0000   NaN        S
887      0            112053  30.0000   B42        S
888      2        W./C. 6607  23.4500   NaN        S
889      0            111369  30.0000  C148        C
890      0            370376   7.7500   NaN        Q

[891 rows x 12 columns]
```

[4]:
```
tips = load_dataset("tips")
```

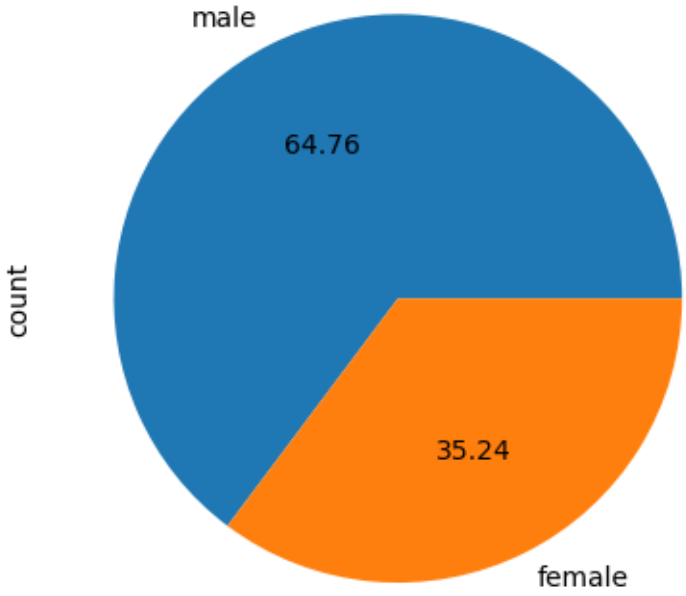
[5]:
```
tips
```

[5]:
```
     total_bill   tip     sex smoker   day    time  size
0         16.99  1.01  Female     No   Sun  Dinner     2
1         10.34  1.66    Male     No   Sun  Dinner     3
2         21.01  3.50    Male     No   Sun  Dinner     3
3         23.68  3.31    Male     No   Sun  Dinner     2
4         24.59  3.61  Female     No   Sun  Dinner     4
..          …   …      …    …   …      …   …
239       29.03  5.92    Male     No   Sat  Dinner     3
240       27.18  2.00  Female    Yes   Sat  Dinner     2
241       22.67  2.00    Male    Yes   Sat  Dinner     2
242       17.82  1.75    Male     No   Sat  Dinner     2
243       18.78  3.00  Female     No  Thur  Dinner     2

[244 rows x 7 columns]
```

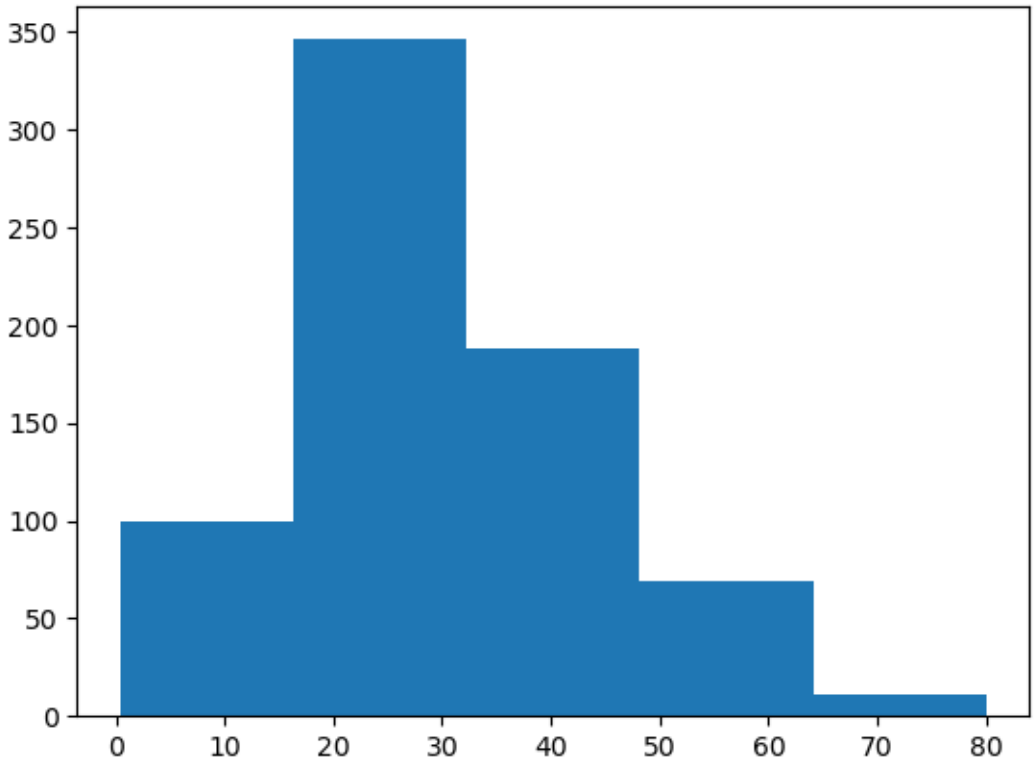
[6]:
```
sns.countplot(x='Survived', data=data)
plt.show()
```




[7]:
```
data['Sex'].value_counts().plot(kind="pie", autopct="%.2f")
plt.show()
```

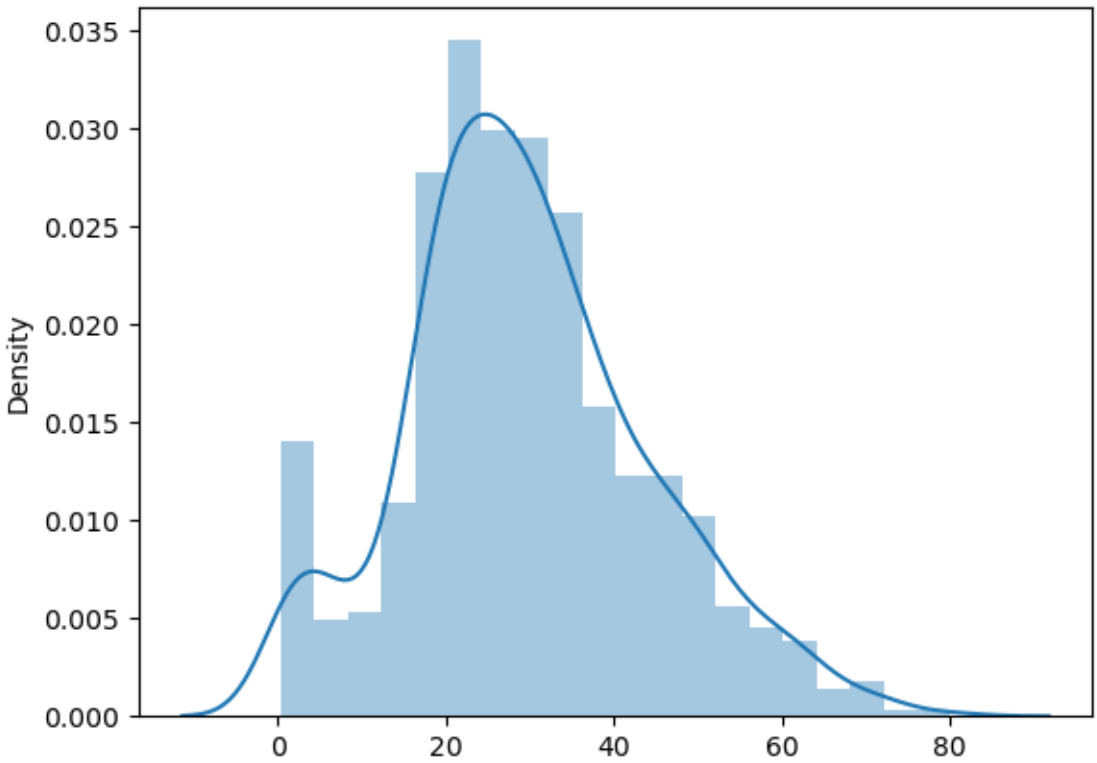



[8]:
```
plt.hist(data['Age'], bins=5)
plt.show()
```



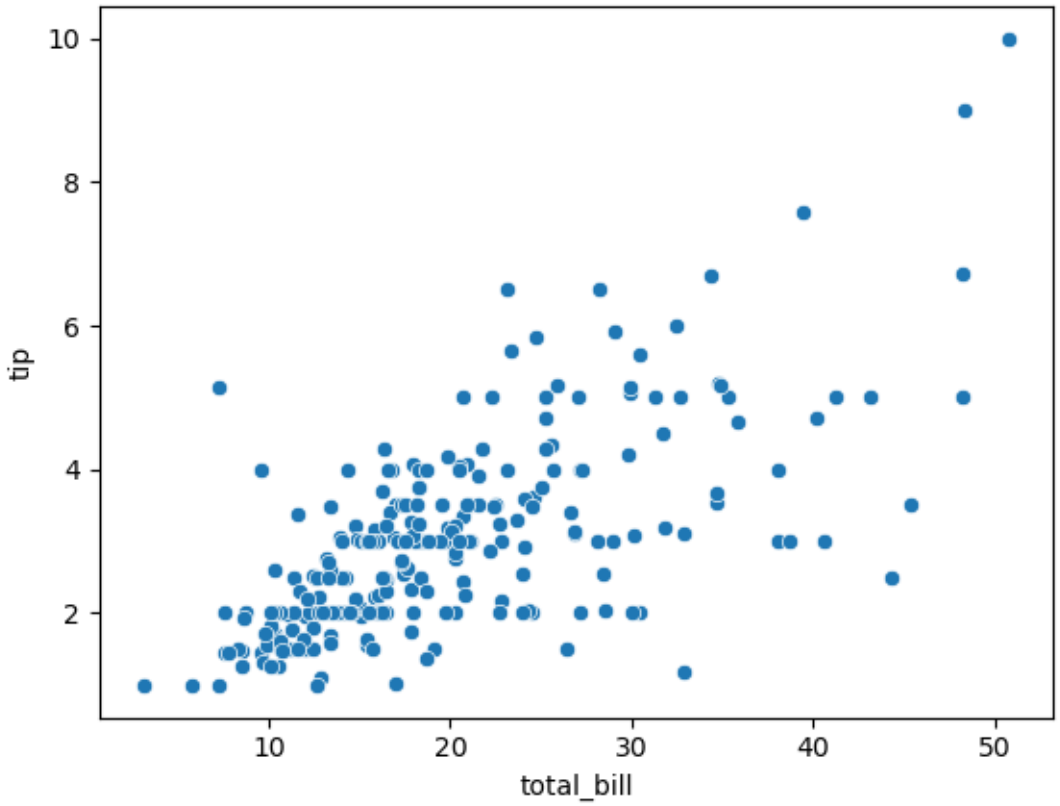

[9]:
```
sns.distplot(x=data['Age'])
plt.show()
```

```
C:\Users\WINDOWS 10\AppData\Local\Temp\ipykernel_10968\1326968689.py:1:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(x=data['Age'])
C:\Users\WINDOWS 10\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed in a
future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

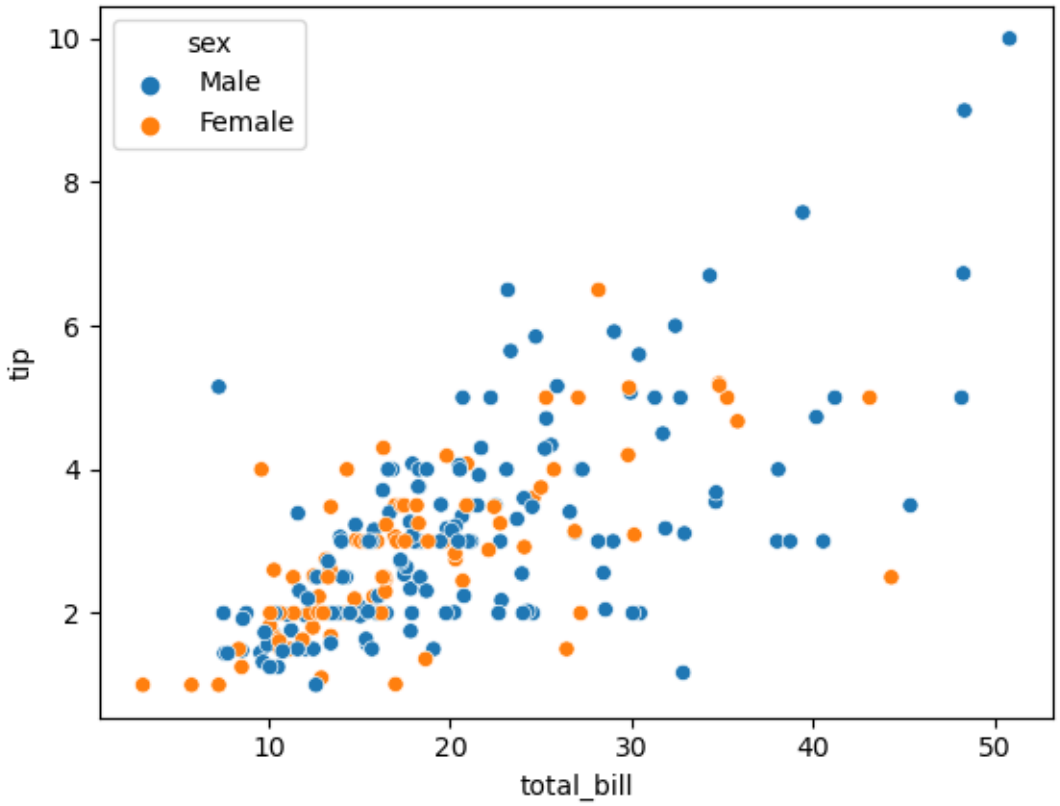



[10]:
```
sns.scatterplot(x=tips["total_bill"], y=tips["tip"])
```

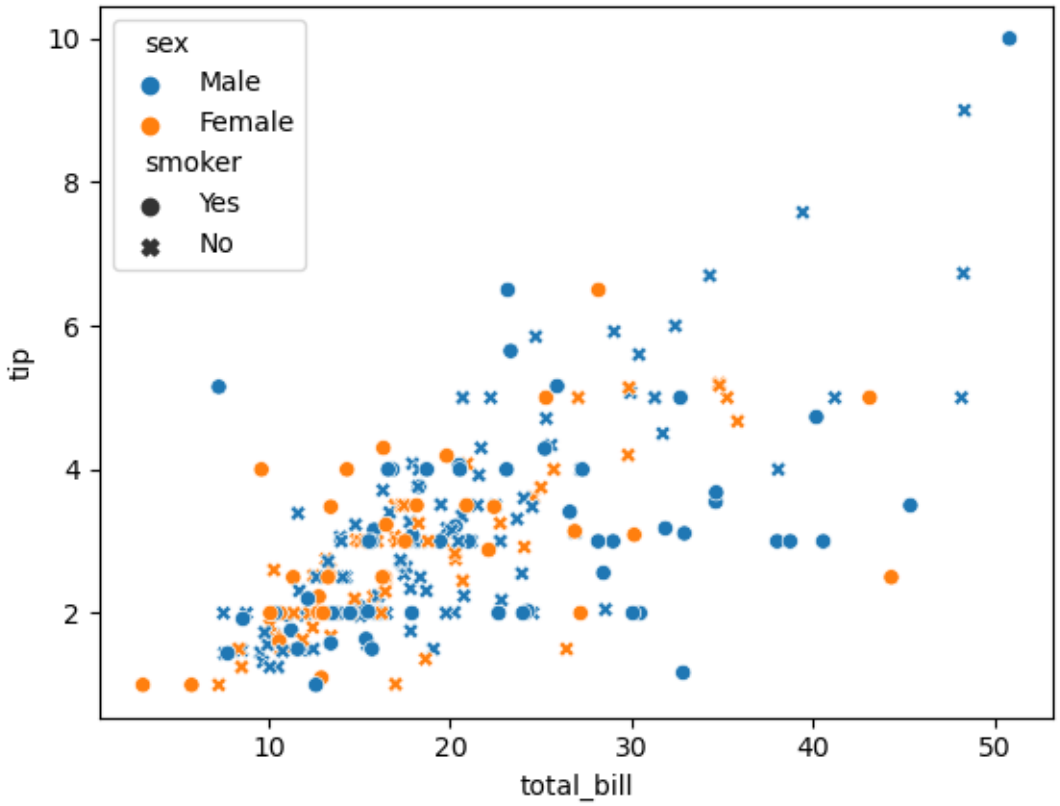
[10]:
```
<Axes: xlabel='total_bill', ylabel='tip'>
```



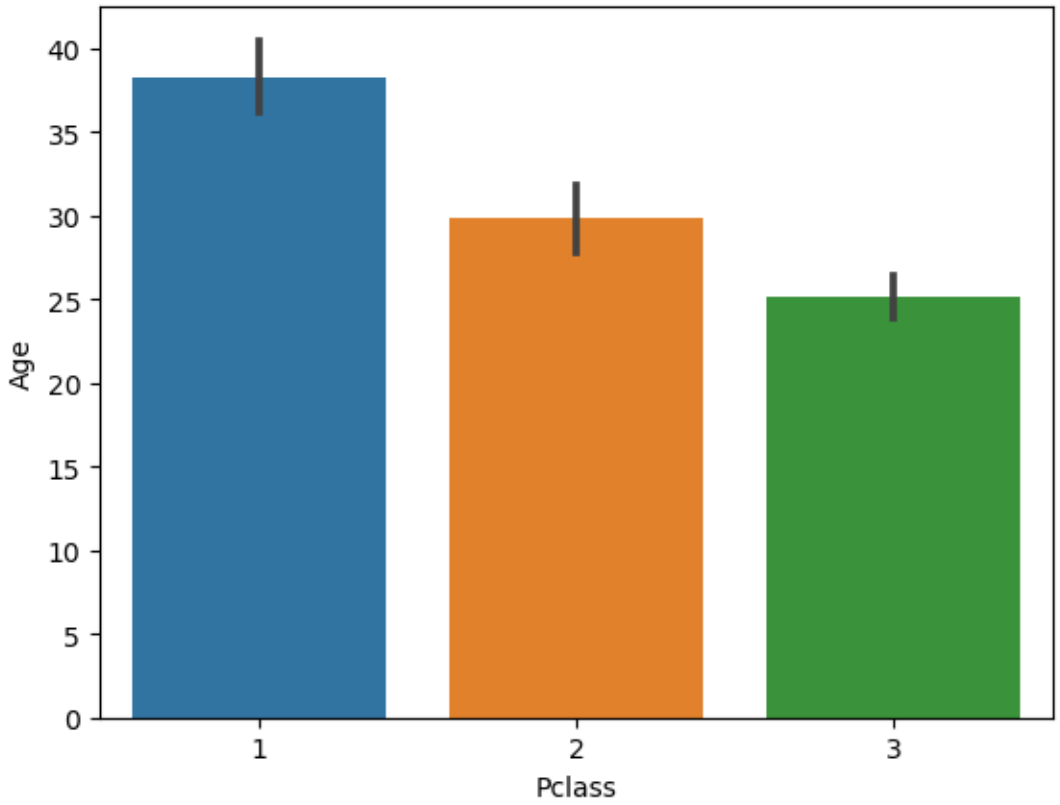

[11]:
```
sns.scatterplot(x="total_bill", y="tip", hue="sex", data=tips)
plt.show()
```



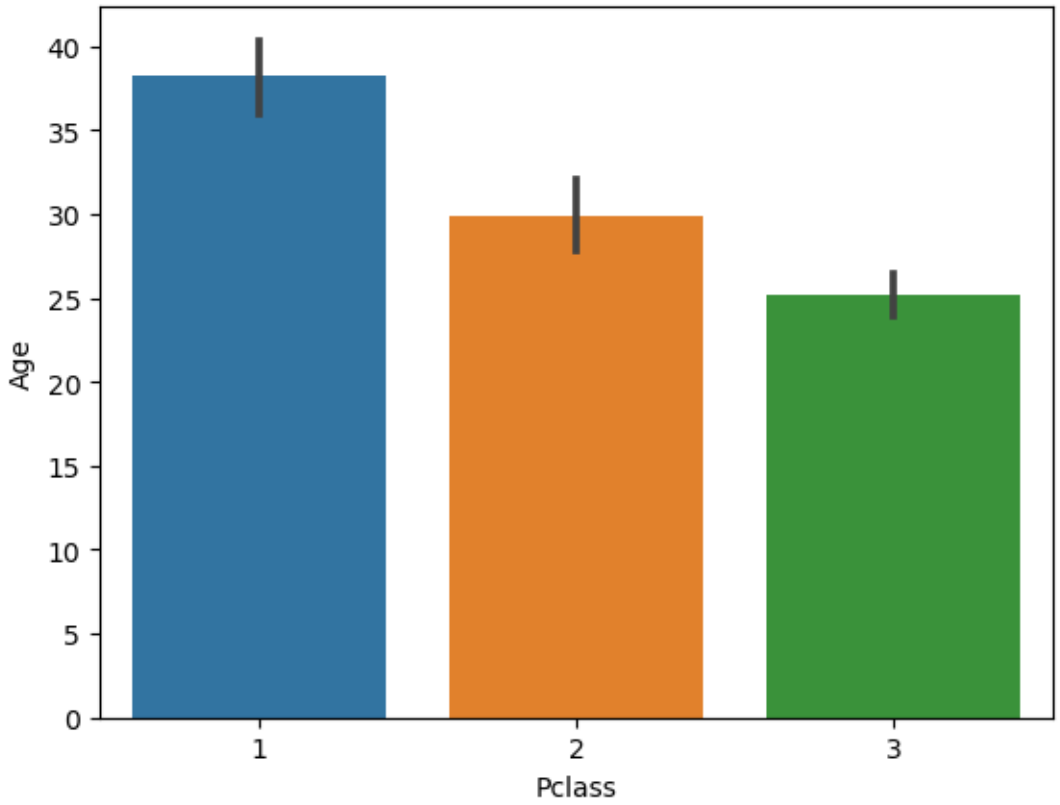

[12]:
```
sns.scatterplot(x=tips["total_bill"], y=tips["tip"],␣
↪hue=tips["sex"],style=tips["smoker"])
plt.show()
```



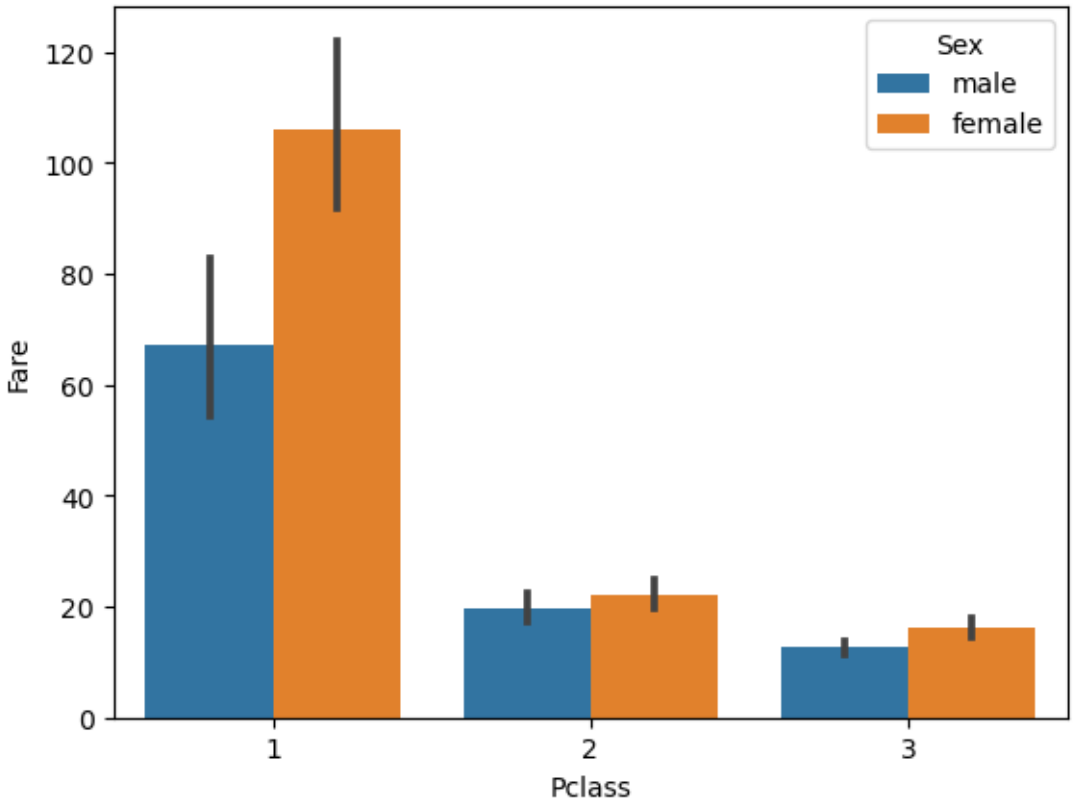

[13]:
```
sns.barplot(x=data['Pclass'], y=data['Age'])
plt.show()
```



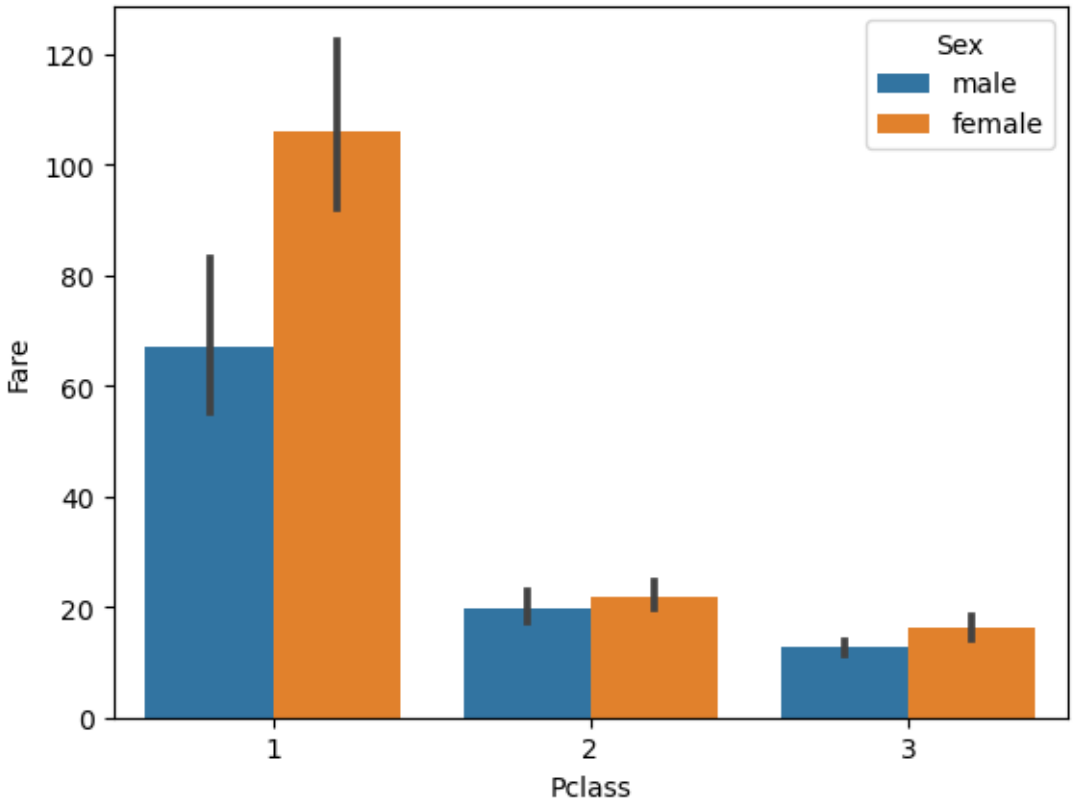

[14]:
```
sns.barplot(x=data['Pclass'], y=data['Age'])
plt.show()
```



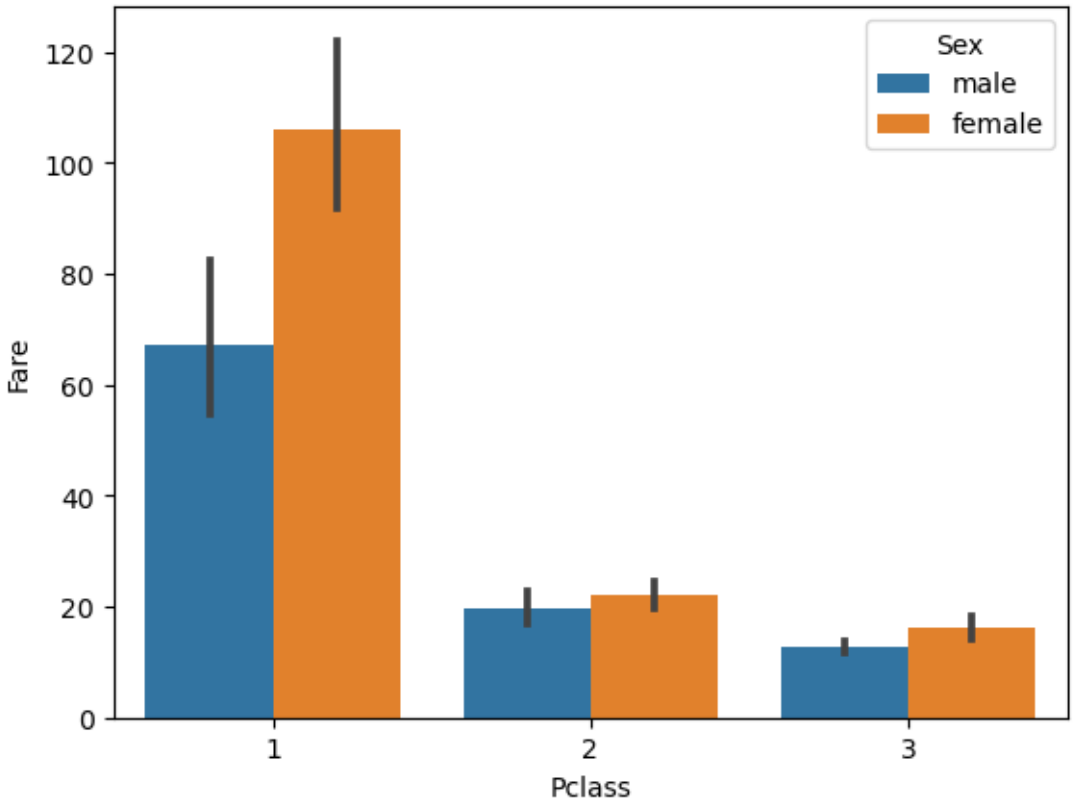

[15]:
```
sns.barplot(x=data['Pclass'], y=data['Fare'], hue = data["Sex"])
plt.show()
```



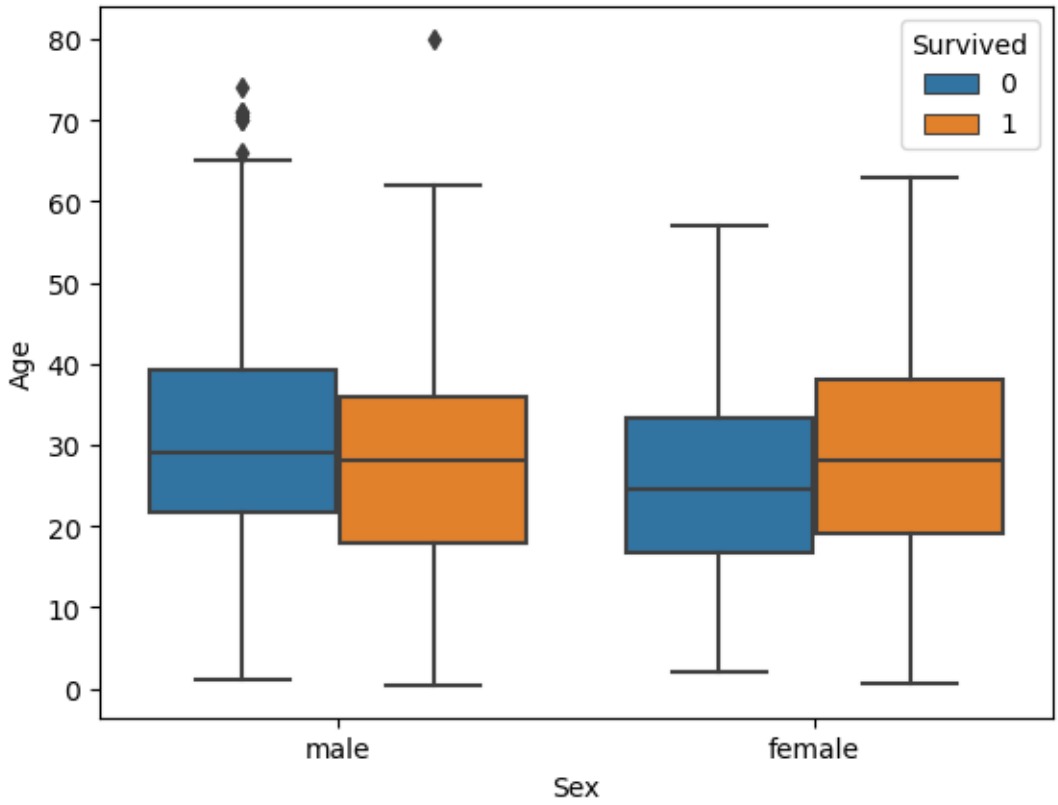

[16]:
```
sns.barplot(x=data['Pclass'], y=data['Fare'], hue = data["Sex"])
plt.show()
```




[17]:
```
sns.barplot(x=data['Pclass'], y=data['Fare'], hue = data["Sex"])
plt.show()
```

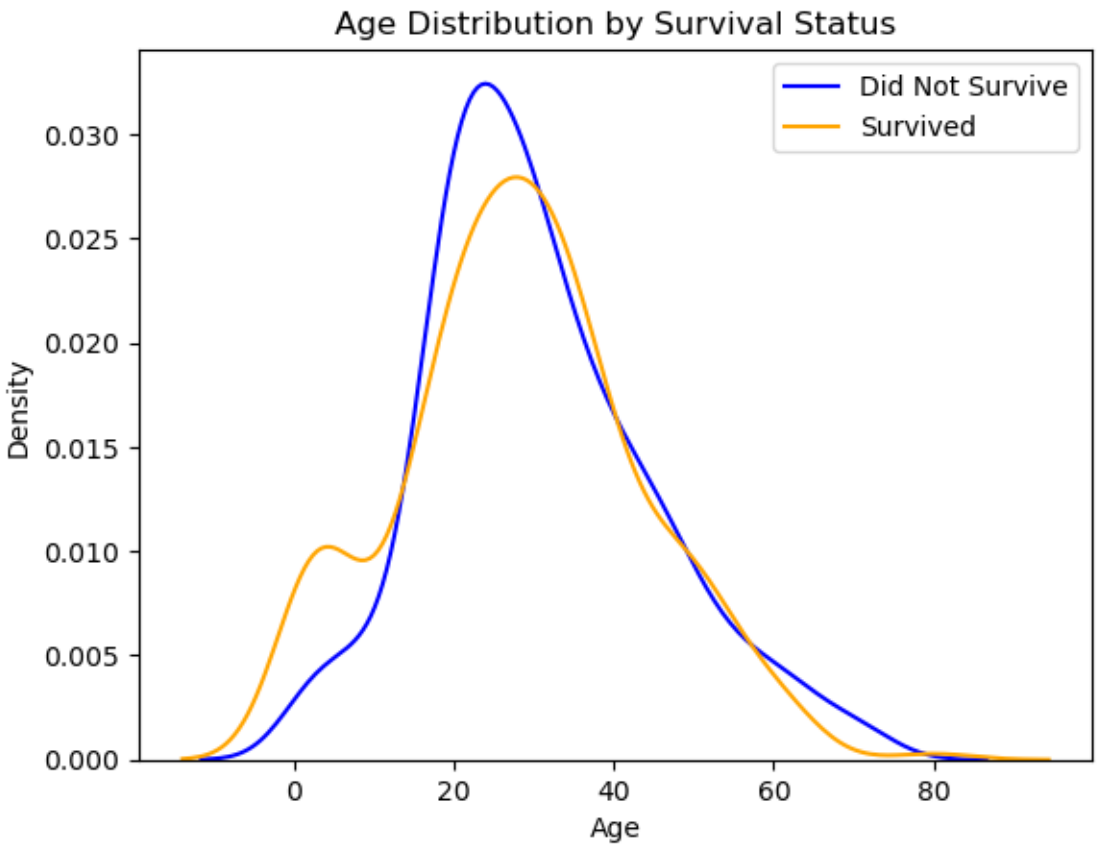



[18]:
```
sns.boxplot(x=data['Sex'], y=data['Age'], hue=data['Survived'])
plt.show()
```




[19]:
```
sns.kdeplot(data=data[data['Survived'] == 0]['Age'], label='Did Not␣
 ↪Survive',color='blue')
sns.kdeplot(data=data[data['Survived'] == 1]['Age'],␣
 ↪label='Survived',color='orange')
plt.xlabel("Age")
plt.title("Age Distribution by Survival Status")
plt.legend()
plt.show()
```

```
C:\Users\WINDOWS 10\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed in a
future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\WINDOWS 10\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed in a
future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

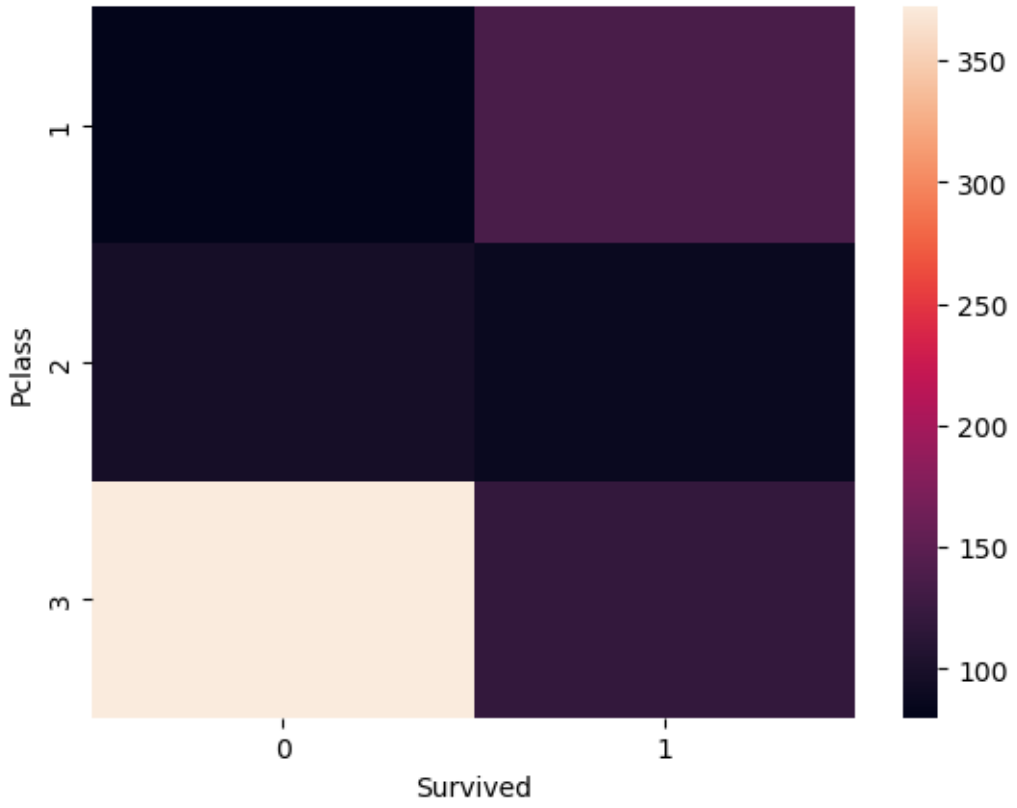



[20]:
```
pd.crosstab(data['Pclass'], data['Survived'])
```

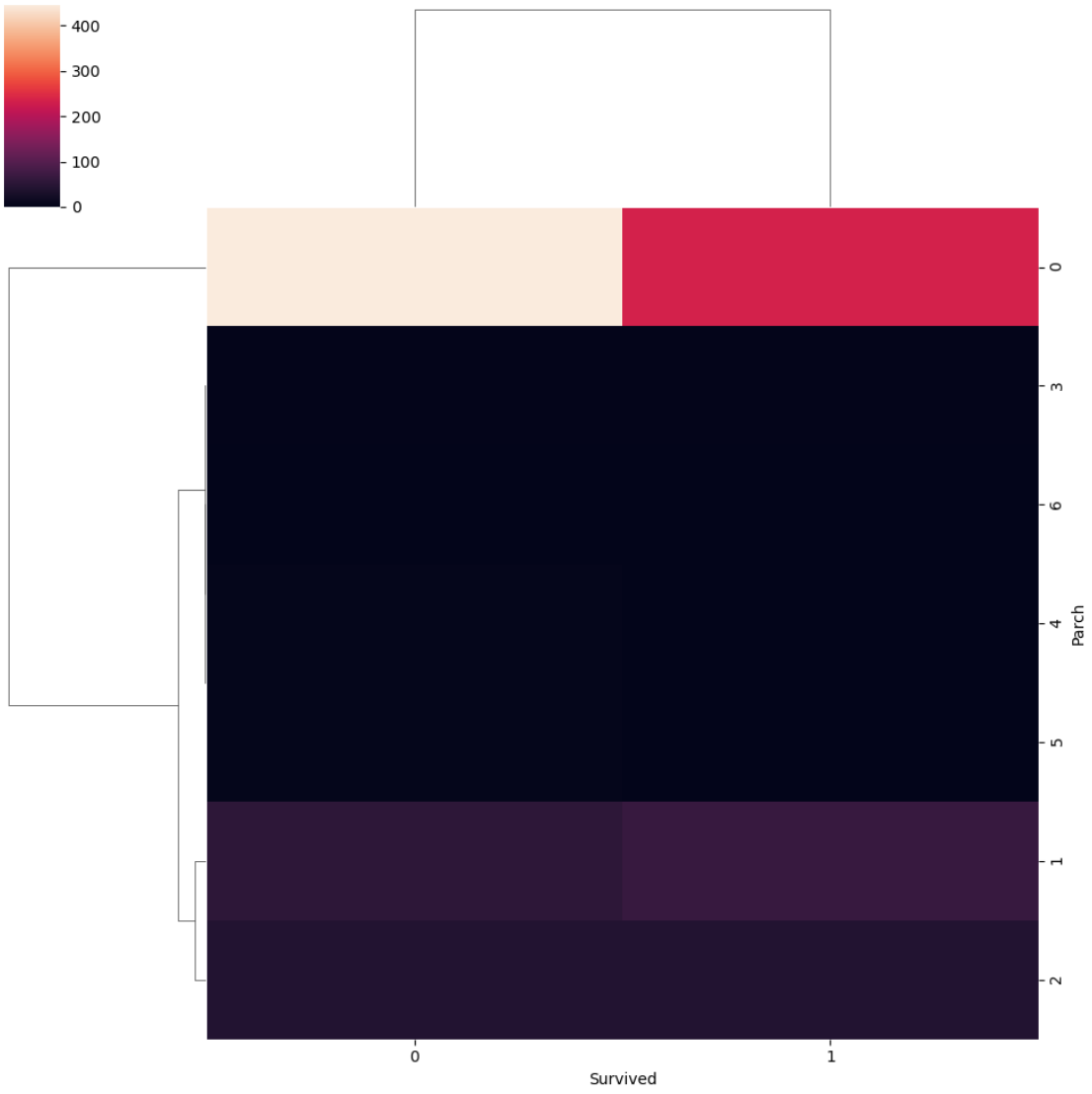
[20]:
```
Survived    0    1
Pclass
1          80  136
2          97   87
3         372  119
```

[21]:
```
sns.heatmap(pd.crosstab(data['Pclass'], data['Survived']))
```

[21]:
```
<Axes: xlabel='Survived', ylabel='Pclass'>
```




[22]:
```
sns.clustermap(pd.crosstab(data['Parch'], data['Survived']))
plt.show()
```


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