Roll No: 31440 ¶

Assignment-8

Importing libraries

In [1]:

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

Dataset Loading ,observation and preprocessing

In [2]:

```
df=pd.read_csv("titanic.csv")
df
```

Out[2]:

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

891 rows × 12 columns

In [3]:

```
df.isnull().sum()
```

Out[3]:

```
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 [4]:

```
df['Age'].value_counts()
```

Out[4]:

```
24.00    30
22.00    27
18.00    26
19.00    25
28.00    25
         ..
36.50     1
55.50     1
0.92      1
23.50     1
74.00     1
Name: Age, Length: 88, dtype: int64
```

In [5]:

```
df['Age'].mode()
```

Out[5]:

```
0    24.0
dtype: float64
```

In [6]:

```
df['Embarked'].value_counts()
```

Out[6]:

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

In [7]:

```
df["Age"].fillna(df["Age"].mean(),inplace=True)
df["Cabin"].fillna(0, inplace=True)
df["Embarked"].fillna("S", inplace=True)
df
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.000000	1	0	A/5 21171	7
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.000000	1	0	PC 17599	71
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O2. 3101282	7
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	113803	53
4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	373450	8
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.000000	0	0	211536	13
887	888	1	1	Graham, Miss. Margaret Edith	female	19.000000	0	0	112053	30
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	29.699118	1	2	W./C. 6607	23
889	890	1	1	Behr, Mr. Karl Howell	male	26.000000	0	0	111369	30
890	891	0	3	Dooley, Mr. Patrick	male	32.000000	0	0	370376	7

891 rows × 12 columns

In [8]:

```
df.isnull().sum()
```

Out[8]:

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

Heatmap

In [9]:

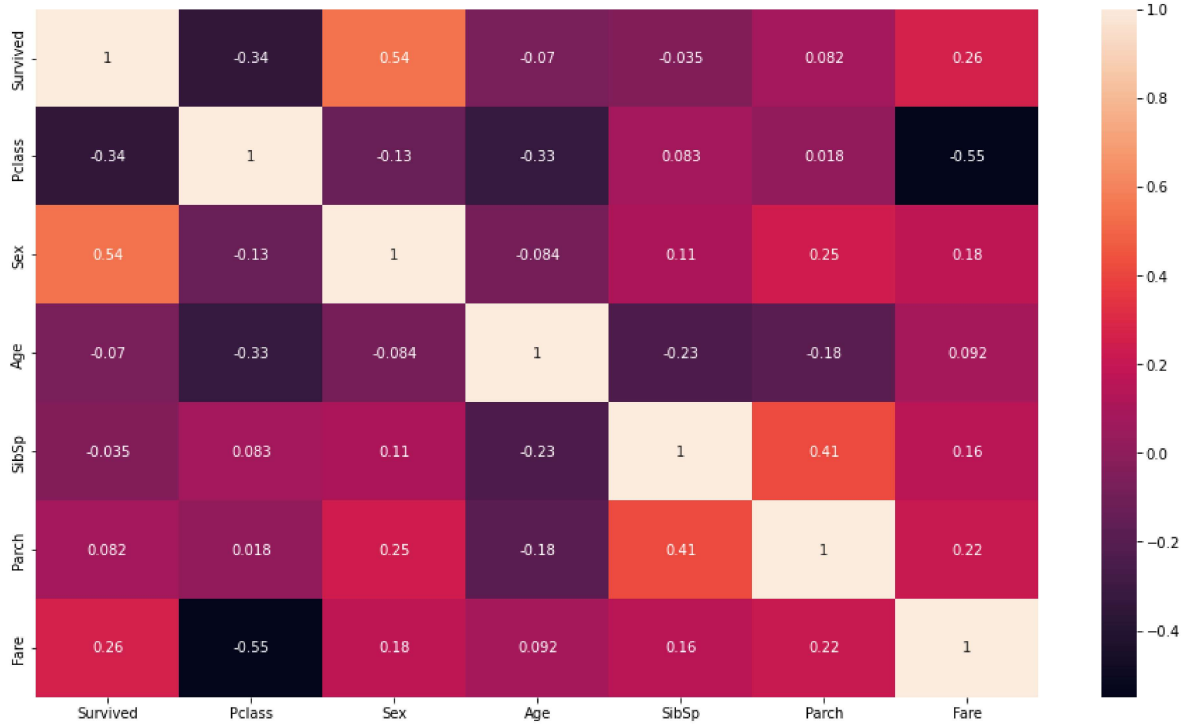
```
df_map=df.drop(columns=["PassengerId","Name","Ticket","Cabin","Embarked"])
df_map["Sex"].replace(["male","female"],[0,1],inplace=True)
```

In [10]:

```
plt.figure(figsize=(16,9))
sns.heatmap(df_map.corr(),annot=True)
```

Out[10]:

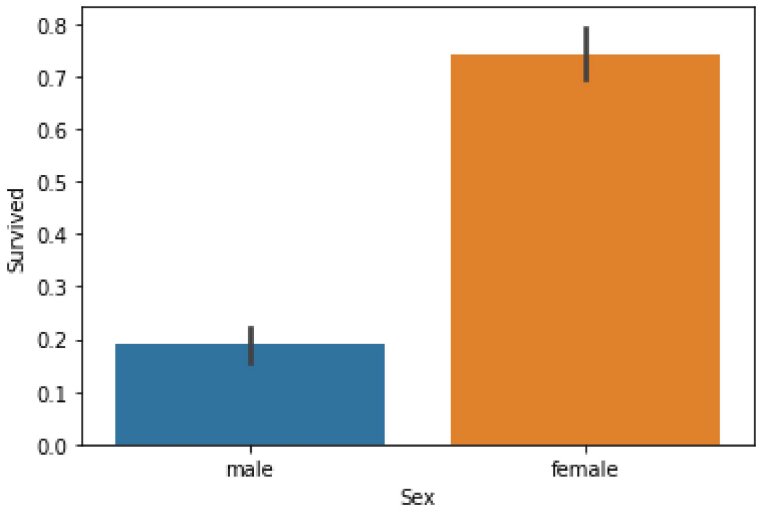
```
<AxesSubplot:>
```



Barplot

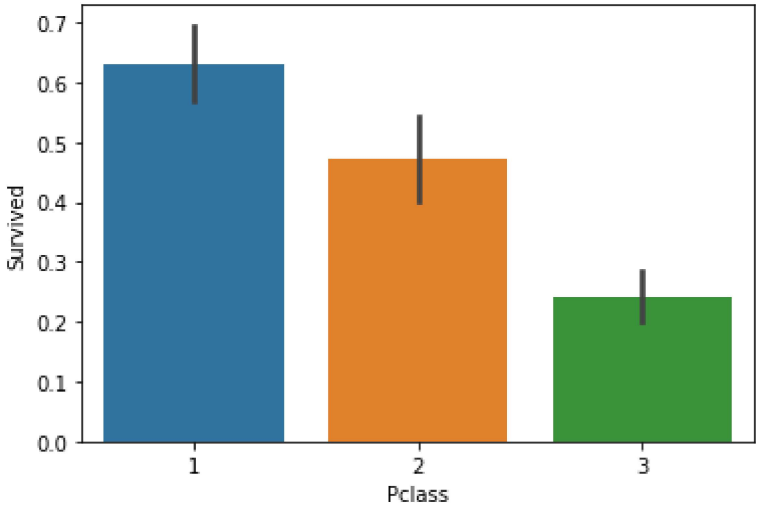
In [11]:

```
sns.barplot(x=df["Sex"],y=df['Survived'])
plt.show()
```



In [12]:

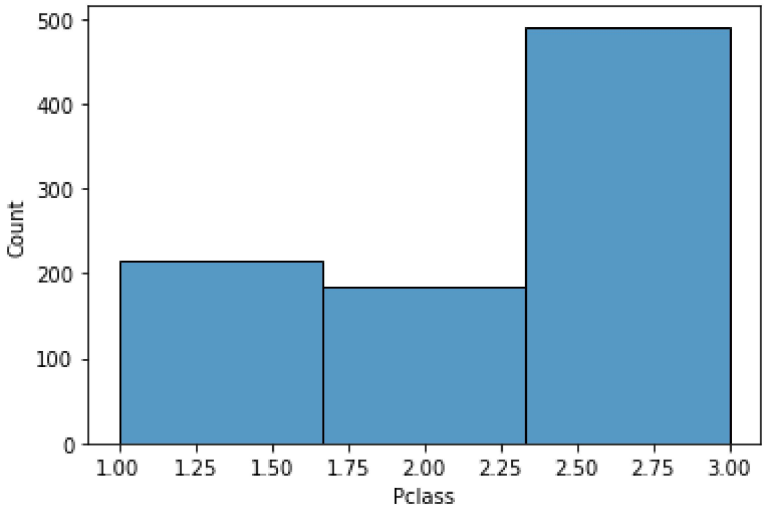
```
sns.barplot(x=df["Pclass"],y=df['Survived'])
plt.show()
```



Histogram

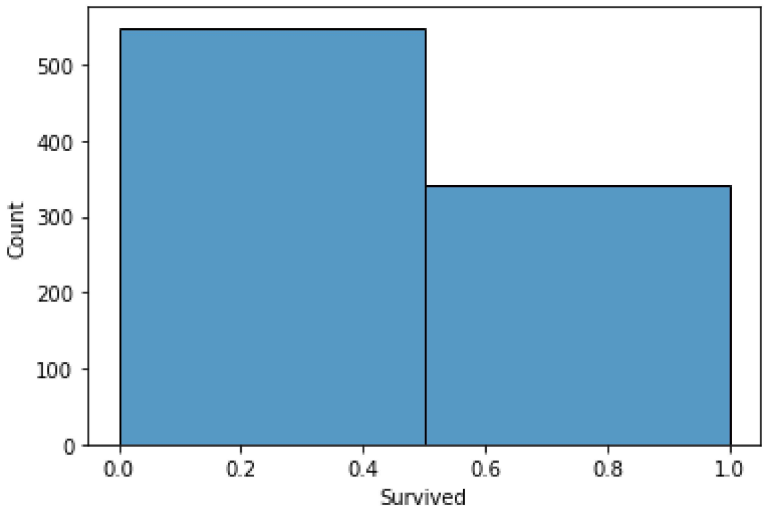
In [13]:

```
sns.histplot(df['Pclass'],bins=3,kde=False)
plt.show()
```



In [14]:

```
sns.histplot(df['Survived'],bins=2,kde=False)
plt.show()
```

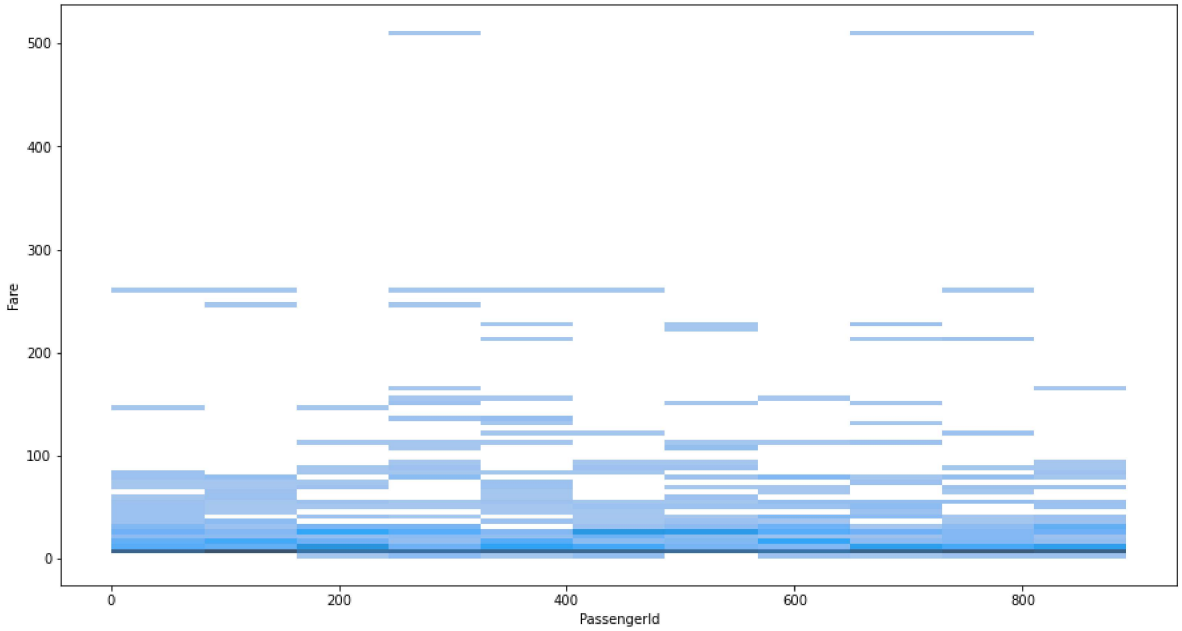


In [15]:

```
plt.figure(figsize=(15,8))
sns.histplot(x=df["PassengerId"],y=df["Fare"])
```

Out[15]:

```
<AxesSubplot:xlabel='PassengerId', ylabel='Fare'>
```

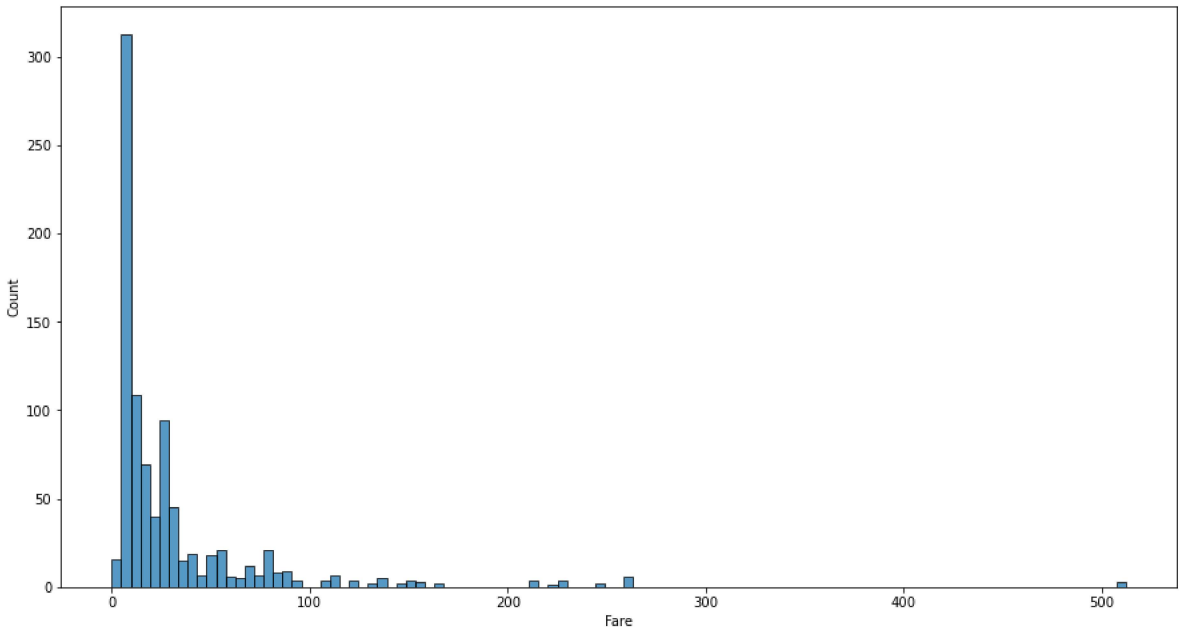


In [16]:

```
plt.figure(figsize=(15,8))
sns.histplot(x=df["Fare"])
```

Out[16]:

```
<AxesSubplot:xlabel='Fare', ylabel='Count'>
```



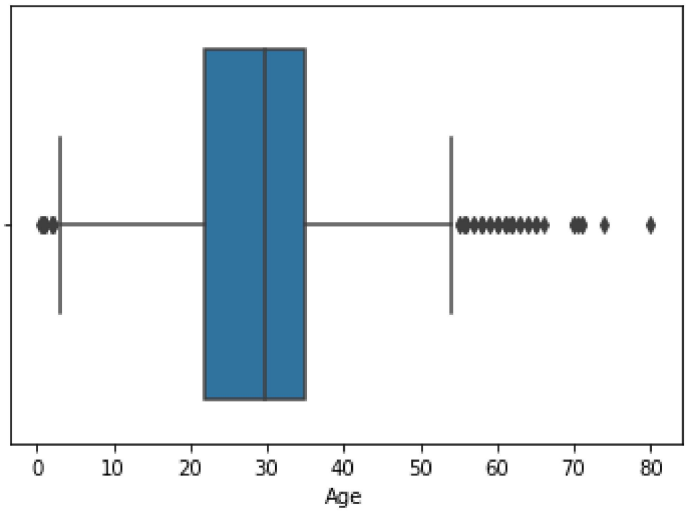
Boxplot

In [17]:

```
sns.boxplot(x=df["Age"])
```

Out[17]:

```
<AxesSubplot:xlabel='Age'>
```

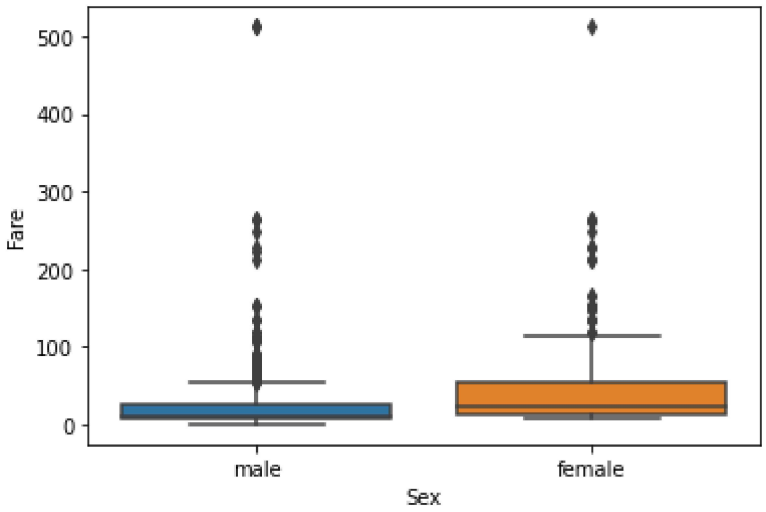


In [18]:

```
sns.boxplot(y=df["Fare"],x=df['Sex'])
```

Out[18]:

```
<AxesSubplot:xlabel='Sex', ylabel='Fare'>
```



In [19]:

```
df[df["Fare"]>500]
```

Out[19]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Ca
258	259	1	1	Ward, Miss. Anna	female	35.0	0	0	PC 17755	512.3292	
679	680	1	1	Cardeza, Mr. Thomas Drake Martinez	male	36.0	0	1	PC 17755	512.3292	
737	738	1	1	Lesurer, Mr. Gustave J	male	35.0	0	0	PC 17755	512.3292	B

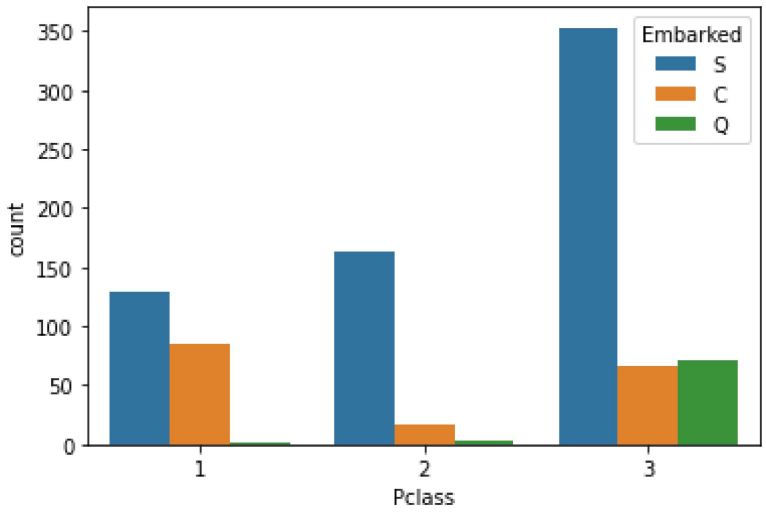
CountPlot

In [20]:

```
sns.countplot(x=df["Pclass"],hue=df["Embarked"])
```

Out[20]:

```
<AxesSubplot:xlabel='Pclass', ylabel='count'>
```

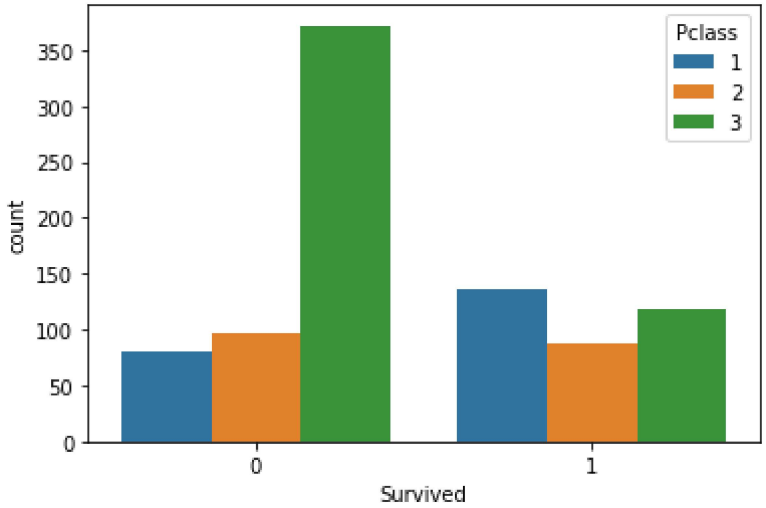


In [21]:

```
sns.countplot(x=df['Survived'],hue=df["Pclass"])
```

Out[21]:

```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```

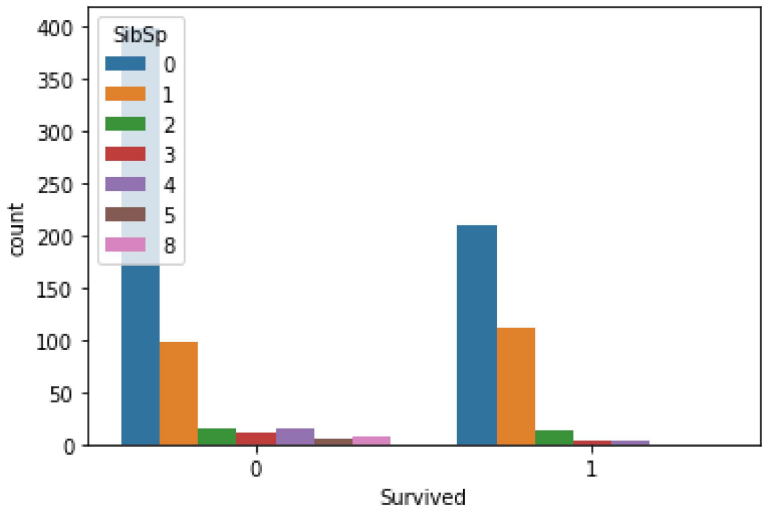


In [22]:

```
sns.countplot(x=df['Survived'],hue=df["SibSp"])
```

Out[22]:

```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```



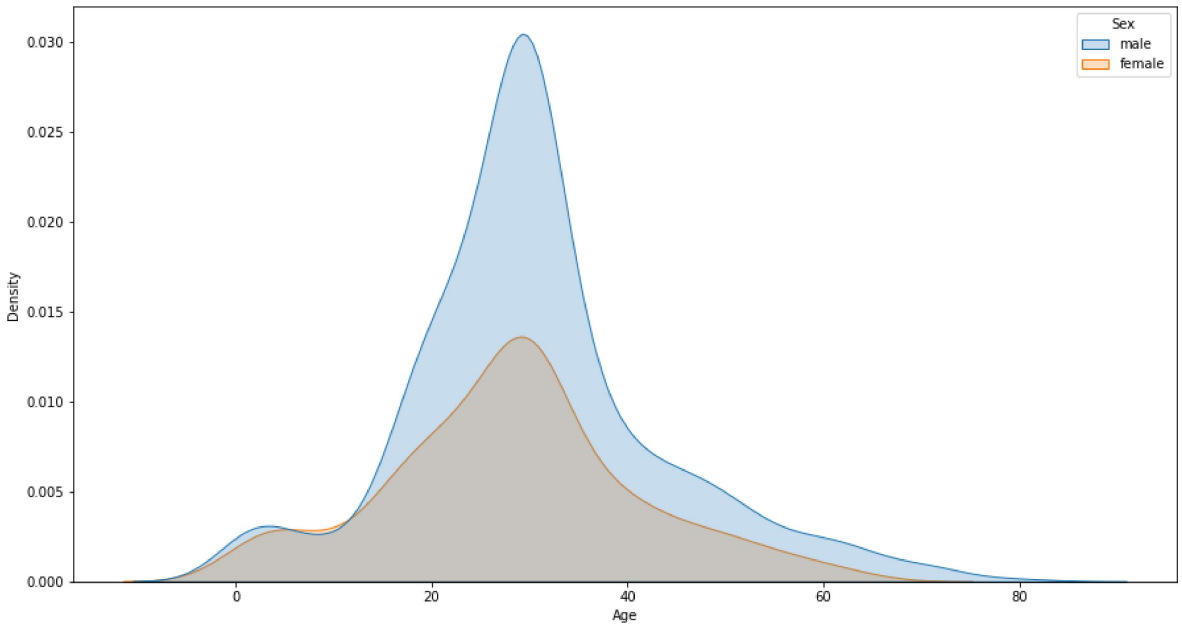
KDE plot

In [23]:

```
plt.figure(figsize=(15,8))
sns.kdeplot(x=df["Age"],hue=df["Sex"],shade=True)
```

Out[23]:

```
<AxesSubplot:xlabel='Age', ylabel='Density'>
```

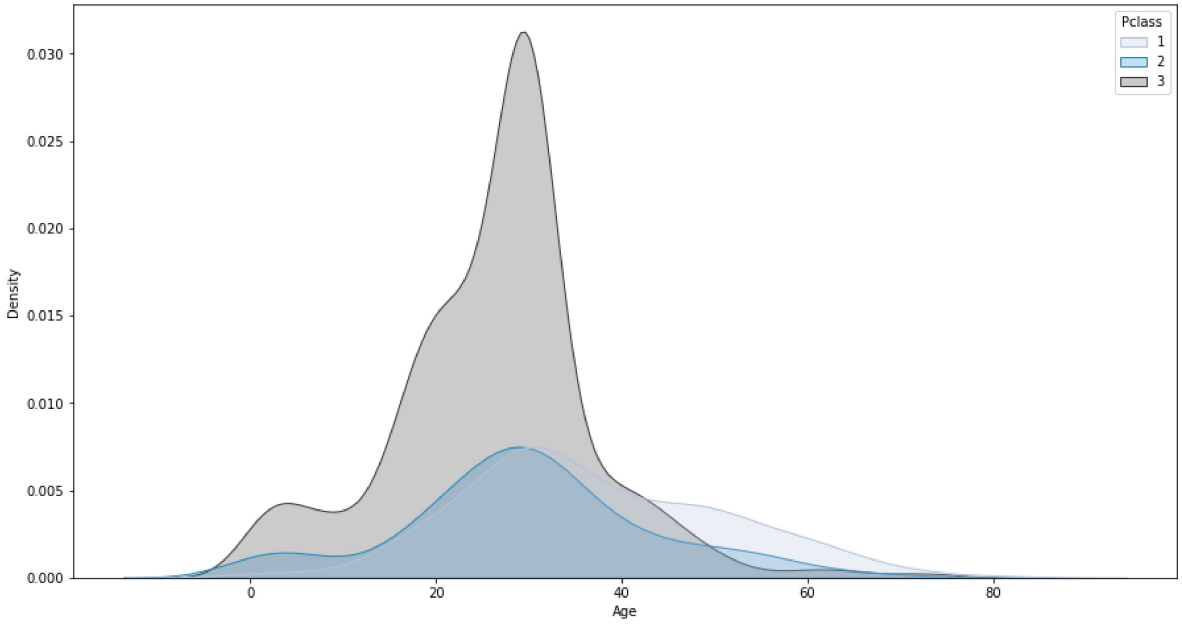


In [24]:

```
plt.figure(figsize=(15,8))
sns.kdeplot(x=df["Age"],hue=df["Pclass"],palette="PuBu_d",shade=True)
```

Out[24]:

```
<AxesSubplot:xlabel='Age', ylabel='Density'>
```



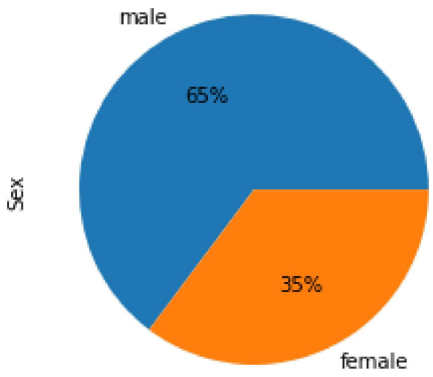
Pie chart

In [25]:

```
df["Sex"].value_counts().plot(kind="pie",autopct='%1.0f%%')
```

Out[25]:

```
<AxesSubplot:ylabel='Sex'>
```

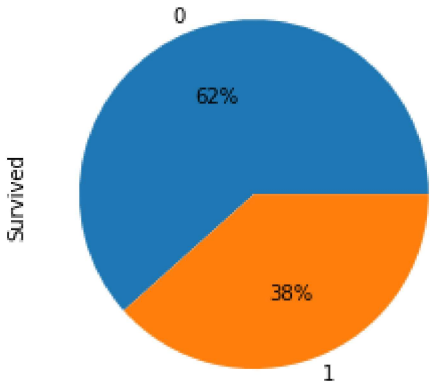


In [26]:

```
df['Survived'].value_counts().plot(kind="pie",autopct='%1.0f%%')
```

Out[26]:

```
<AxesSubplot:ylabel='Survived'>
```



Violin Plot

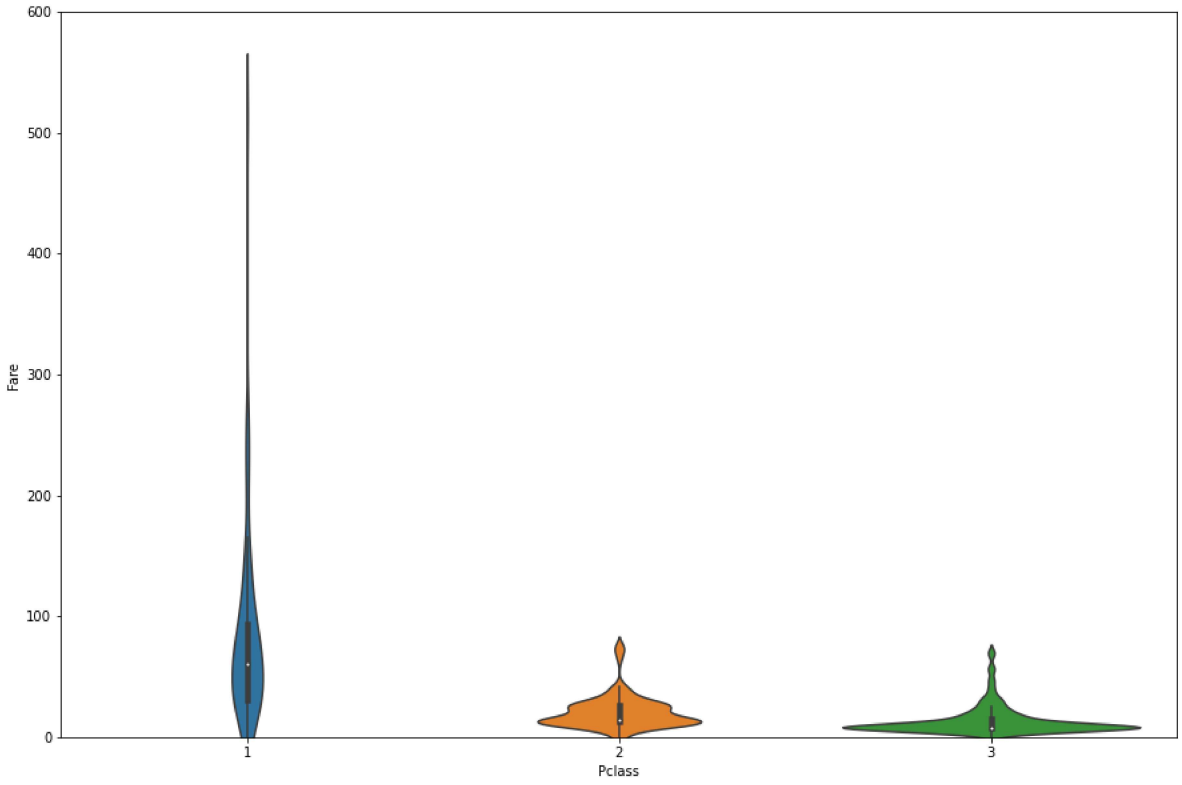
In [27]:

```
plt.figure(figsize=(15,10))
sns.violinplot(df["Pclass"],df["Fare"]).set_ylim(0,600)
```

```
C:\Users\Dell\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureW
arning: Pass the following variables as keyword args: x, y. From version 0.1
2, the only valid positional argument will be `data`, and passing other argu
ments without an explicit keyword will result in an error or misinterpretati
on.
  warnings.warn(
```

Out[27]:

```
(0.0, 600.0)
```



In [ ]: