Titanic Dataset

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

```python
In [2]:  df = pd.read_csv("D:\\Kaggle Notebook\\train.csv")
```

```python
In [3]:  df.head()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
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	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	

In [4]: `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 [5]: `df.shape`

Out[5]: `(891, 12)`

In [6]: `df.nunique()`

Out[6]:
```
PassengerId    891
Survived         2
Pclass           3
Name           891
Sex              2
Age             88
SibSp            7
Parch            7
Ticket         681
Fare           248
Cabin          147
Embarked         3
dtype: int64
```

In [7]: `df.dtypes`

Out[7]:
```
PassengerId        int64
Survived           int64
Pclass             int64
Name              object
Sex               object
Age              float64
SibSp              int64
Parch              int64
Ticket            object
Fare             float64
Cabin             object
Embarked          object
dtype: object
```
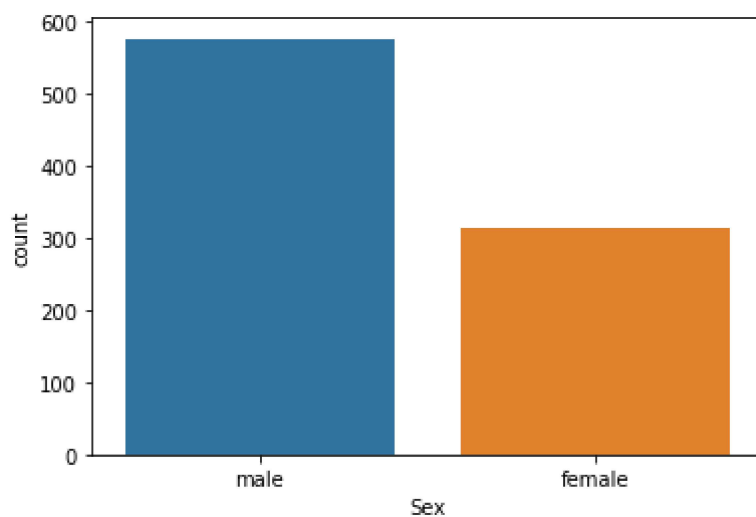
In [8]:
```python
sns.countplot(data=df,x="Sex")
M,F = df["Sex"].value_counts()
print("Number of Males:",M)
print("Number of Females:",F)
plt.show()
```

```
Number of Males: 577
Number of Females: 314
```


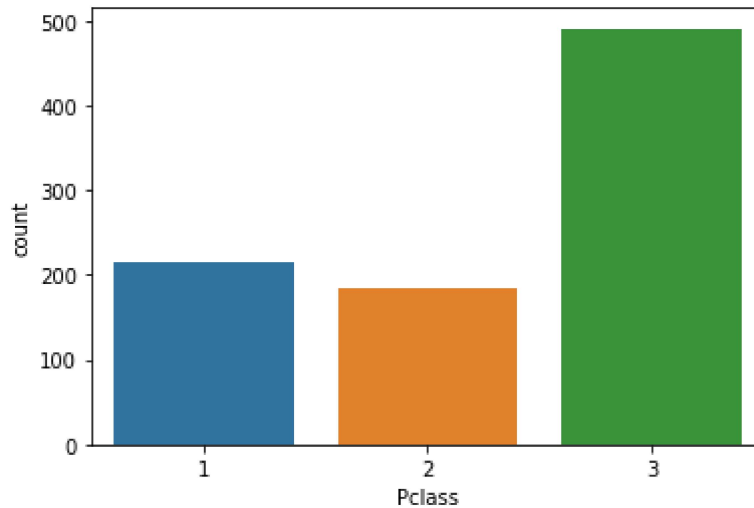In [9]: `df["Pclass"].value_counts()`

Out[9]:
```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

```
In [10]: sns.countplot(data = df, x="Pclass")
         A,B,C =  df["Pclass"].value_counts()
         print("Number of Passenger in Class_1:",C)
         print("Number of Passenger in Class_2:",B)
         print("Number of Passenger in class_3:",A)
         plt.show()
```

```
Number of Passenger in Class_1: 184
Number of Passenger in Class_2: 216
Number of Passenger in class_3: 491
```



```
In [11]: df["SibSp"].value_counts()
```

```
Out[11]: 0    608
         1    209
         2     28
         4     18
         3     16
         8      7
         5      5
         Name: SibSp, dtype: int64
```

In [12]:
```python
sns.countplot(data=df,x="SibSp")
plt.show()
```
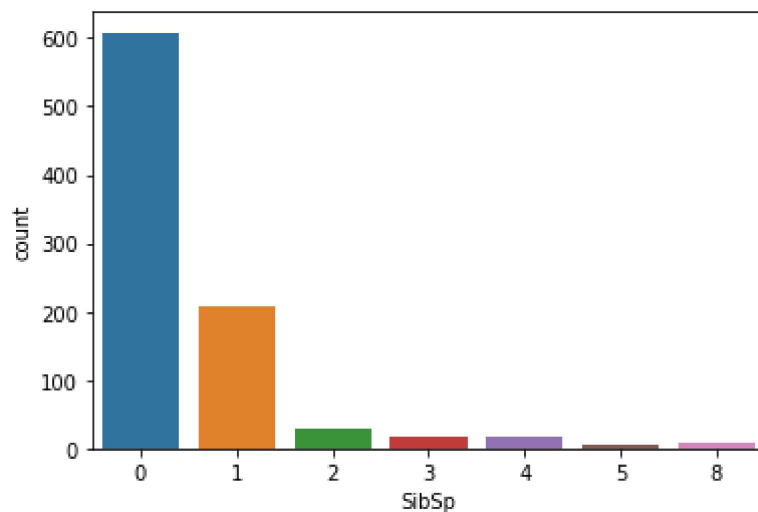

In [13]:
```python
df["Parch"].value_counts()
```

Out[13]:
```
0    678
1    118
2     80
5      5
3      5
4      4
6      1
Name: Parch, dtype: int64
```

In [14]:
```python
sns.countplot(data=df,x="Parch")
plt.show()
```
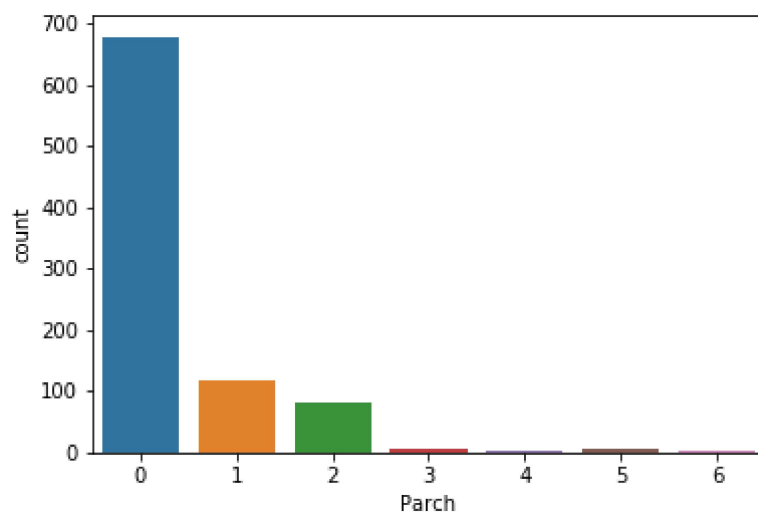
In [15]:
```python
df["Embarked"].value_counts()
```

Out[15]:
```
S    644
C    168
Q     77
Name: Embarked, dtype: int64
```

In [16]:
```python
sns.countplot(data=df,x="Embarked")
plt.title("Number of People Embarked")

plt.show()
```
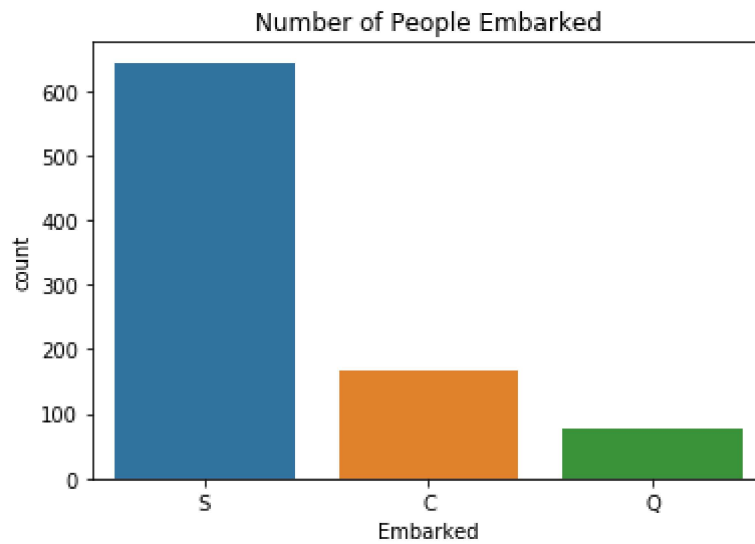

In [17]:
```python
sns.countplot(data=df,x="Pclass",hue="Survived")
```

Out[17]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a563047188>
```


In [ ]:

In [18]:
```python
df["Age"].mean()
```

Out[18]:
```
29.69911764705882
```

```python
In [19]: from sklearn.impute import SimpleImputer
         mean_imputer = SimpleImputer(strategy = "mean")
         df["Age"] = pd.DataFrame(mean_imputer.fit_transform(df[["Age"]]))
```

```python
In [20]: df.isnull().sum()
```

```
Out[20]: PassengerId      0
         Survived         0
         Pclass           0
         Name             0
         Sex              0
         Age              0
         SibSp            0
         Parch            0
         Ticket           0
         Fare             0
         Cabin          687
         Embarked         2
         dtype: int64
```
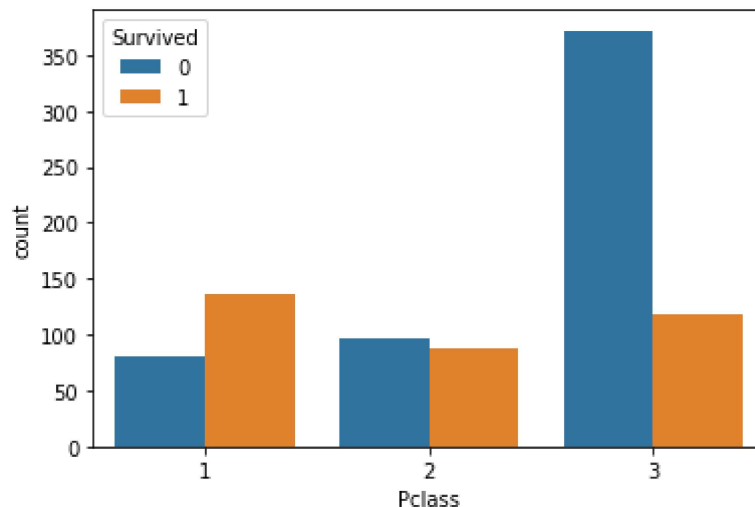
```python
In [21]: df["Cabin"]
```

```
Out[21]: 0        NaN
         1        C85
         2        NaN
         3       C123
         4        NaN
                 ...
         886      NaN
         887      B42
         888      NaN
         889     C148
         890      NaN
         Name: Cabin, Length: 891, dtype: object
```

In [22]: `df.head(10)`

Out[22]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	1	0	3	Braund, Mr. Owen Harris	male	22.000000	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.000000	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	373450	8.050
5	6	0	3	Moran, Mr. James	male	29.699118	0	0	330877	8.458
6	7	0	1	McCarthy, Mr. Timothy J	male	54.000000	0	0	17463	51.862
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.000000	3	1	349909	21.075
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.000000	0	2	347742	11.133
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.000000	1	0	237736	30.070

In [23]: `df.shape`

Out[23]: `(891, 12)`

```
In [24]: df.drop(["Cabin"],axis=1)
```

Out[24]:

	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	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.000000	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	373450	8.0
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.000000	0	0	211536	13.0
887	888	1	1	Graham, Miss. Margaret Edith	female	19.000000	0	0	112053	30.0
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	29.699118	1	2	W./C. 6607	23.4
889	890	1	1	Behr, Mr. Karl Howell	male	26.000000	0	0	111369	30.0
890	891	0	3	Dooley, Mr. Patrick	male	32.000000	0	0	370376	7.7

891 rows × 11 columns

```
In [25]: df1 = pd.get_dummies(df.Sex)
```
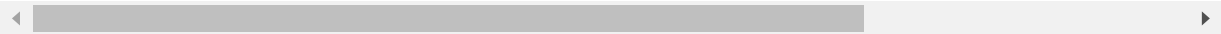
```
In [26]: df_2 = pd.concat([df,df1],axis="columns")
```

In [27]: `df_2`

Out[27]:

	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	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.000000	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	373450	8.0
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.000000	0	0	211536	13.0
887	888	1	1	Graham, Miss. Margaret Edith	female	19.000000	0	0	112053	30.0
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	29.699118	1	2	W./C. 6607	23.4
889	890	1	1	Behr, Mr. Karl Howell	male	26.000000	0	0	111369	30.0
890	891	0	3	Dooley, Mr. Patrick	male	32.000000	0	0	370376	7.7

891 rows × 14 columns

In [28]: `df_2.drop(["Sex","Cabin"],axis="columns",inplace=True)`

In [29]: `df_2`

Out[29]:

	PassengerId	Survived	Pclass	Name	Age	SibSp	Parch	Ticket	Fare	E
0	1	0	3	Braund, Mr. Owen Harris	22.000000	1	0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	38.000000	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	26.000000	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	35.000000	1	0	113803	53.1000	
4	5	0	3	Allen, Mr. William Henry	35.000000	0	0	373450	8.0500	
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	27.000000	0	0	211536	13.0000	
887	888	1	1	Graham, Miss. Margaret Edith	19.000000	0	0	112053	30.0000	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	29.699118	1	2	W./C. 6607	23.4500	
889	890	1	1	Behr, Mr. Karl Howell	26.000000	0	0	111369	30.0000	
890	891	0	3	Dooley, Mr. Patrick	32.000000	0	0	370376	7.7500	

891 rows × 12 columns
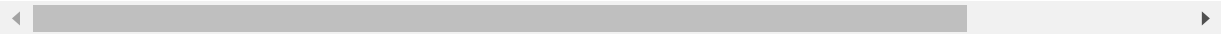
In [30]: `df_3 = pd.get_dummies(df_2.Embarked)`

In [31]: `df_3`

Out[31]:

|	C	Q	S
0	0	0	1
1	1	0	0
2	0	0	1
3	0	0	1
4	0	0	1
...	...	...	...
886	0	0	1
887	0	0	1
888	0	0	1
889	1	0	0
890	0	1	0

891 rows × 3 columns

In [32]:
```python
df_4 = pd.concat([df_2,df_3],axis="columns")
```

In [33]: `df_4`

Out[33]:

	PassengerId	Survived	Pclass	Name	Age	SibSp	Parch	Ticket	Fare	Er
0	1	0	3	Braund, Mr. Owen Harris	22.000000	1	0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	38.000000	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	26.000000	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	35.000000	1	0	113803	53.1000	
4	5	0	3	Allen, Mr. William Henry	35.000000	0	0	373450	8.0500	
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	27.000000	0	0	211536	13.0000	
887	888	1	1	Graham, Miss. Margaret Edith	19.000000	0	0	112053	30.0000	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	29.699118	1	2	W./C. 6607	23.4500	
889	890	1	1	Behr, Mr. Karl Howell	26.000000	0	0	111369	30.0000	
890	891	0	3	Dooley, Mr. Patrick	32.000000	0	0	370376	7.7500	

891 rows × 15 columns
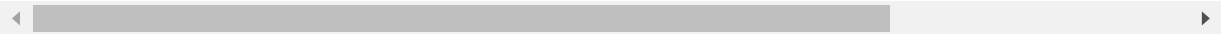
In [34]: `df_5=df_4.drop(["Embarked"],axis="columns")`

In [35]: `df.corr()`

Out[35]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
PassengerId	1.000000	-0.005007	-0.035144	0.033207	-0.057527	-0.001652	0.012658
Survived	-0.005007	1.000000	-0.338481	-0.069809	-0.035322	0.081629	0.257307
Pclass	-0.035144	-0.338481	1.000000	-0.331339	0.083081	0.018443	-0.549500
Age	0.033207	-0.069809	-0.331339	1.000000	-0.232625	-0.179191	0.091566
SibSp	-0.057527	-0.035322	0.083081	-0.232625	1.000000	0.414838	0.159651
Parch	-0.001652	0.081629	0.018443	-0.179191	0.414838	1.000000	0.216225
Fare	0.012658	0.257307	-0.549500	0.091566	0.159651	0.216225	1.000000

In [36]: `df_5.head(10)`

Out[36]:

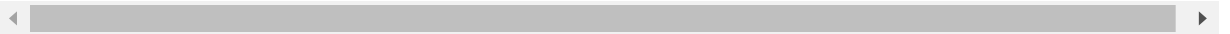|	PassengerId	Survived	Pclass	Name	Age	SibSp	Parch	Ticket	Fare	fema
0	1	0	3	Braund, Mr. Owen Harris	22.000000	1	0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	38.000000	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	26.000000	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	35.000000	1	0	113803	53.1000	
4	5	0	3	Allen, Mr. William Henry	35.000000	0	0	373450	8.0500	
5	6	0	3	Moran, Mr. James	29.699118	0	0	330877	8.4583	
6	7	0	1	McCarthy, Mr. Timothy J	54.000000	0	0	17463	51.8625	
7	8	0	3	Palsson, Master. Gosta Leonard	2.000000	3	1	349909	21.0750	
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	27.000000	0	2	347742	11.1333	
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	14.000000	1	0	237736	30.0708	

In [37]: `df_6 = df_5.drop(["Ticket","Name"],axis = "columns")`

In [38]: df_6

Out[38]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	female	male	C	Q	S
0	1	0	3	22.000000	1	0	7.2500	0	1	0	0	1
1	2	1	1	38.000000	1	0	71.2833	1	0	1	0	0
2	3	1	3	26.000000	0	0	7.9250	1	0	0	0	1
3	4	1	1	35.000000	1	0	53.1000	1	0	0	0	1
4	5	0	3	35.000000	0	0	8.0500	0	1	0	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	27.000000	0	0	13.0000	0	1	0	0	1
887	888	1	1	19.000000	0	0	30.0000	1	0	0	0	1
888	889	0	3	29.699118	1	2	23.4500	1	0	0	0	1
889	890	1	1	26.000000	0	0	30.0000	0	1	1	0	0
890	891	0	3	32.000000	0	0	7.7500	0	1	0	1	0

891 rows × 12 columns

In [39]: df_6.shape

Out[39]: (891, 12)

In [40]: X = df_6.drop("Survived",axis = 1)
y = df["Survived"]

In [41]: X

Out[41]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	female	male	C	Q	S
0	1	3	22.000000	1	0	7.2500	0	1	0	0	1
1	2	1	38.000000	1	0	71.2833	1	0	1	0	0
2	3	3	26.000000	0	0	7.9250	1	0	0	0	1
3	4	1	35.000000	1	0	53.1000	1	0	0	0	1
4	5	3	35.000000	0	0	8.0500	0	1	0	0	1
...	...	...	...	...	...	...	...	...	...	...	...
886	887	2	27.000000	0	0	13.0000	0	1	0	0	1
887	888	1	19.000000	0	0	30.0000	1	0	0	0	1
888	889	3	29.699118	1	2	23.4500	1	0	0	0	1
889	890	1	26.000000	0	0	30.0000	0	1	1	0	0
890	891	3	32.000000	0	0	7.7500	0	1	0	1	0

891 rows × 11 columns

In [42]: `y`

Out[42]:
```
0      0
1      1
2      1
3      1
4      0
      ..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64
```

In [43]:
```python
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test  = train_test_split(X,y,test_size=0.2,random_sta
te=101)
```

In [44]:
```python
X_train.shape,X_test.shape
```

Out[44]: `((712, 11), (179, 11))`

In [45]:
```python
import warnings
warnings.filterwarnings('ignore')
```

In [46]:
```python
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(X_train,y_train)
```

Out[46]:
```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, l1_ratio=None, max_iter=100,
                   multi_class='auto', n_jobs=None, penalty='l2',
                   random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
                   warm_start=False)
```

In [47]:
```python
model.coef_
```

Out[47]:
```
array([[ 9.42117517e-04, -5.98687413e-01, -2.44178194e-02,
        -2.56730140e-01, -2.62823891e-01,  1.15450639e-02,
         1.73709762e+00, -1.06566847e+00,  4.41601658e-01,
         2.45708750e-02,  1.70205690e-01]])
```

In [48]:
```python
y_pred = model.predict(X_test)
```

In [49]:
```python
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

Out[49]: `0.8044692737430168`

```
In [50]: from sklearn.metrics import confusion_matrix,classification_report
         confusion_matrix(y_test,y_pred)
```

```
Out[50]: array([[89, 10],
                [25, 55]], dtype=int64)
```

```
In [51]: print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.78	0.90	0.84	99
1	0.85	0.69	0.76	80
accuracy			0.80	179
macro avg	0.81	0.79	0.80	179
weighted avg	0.81	0.80	0.80	179

```
In [52]: from sklearn.metrics import plot_confusion_matrix
         plot_confusion_matrix(model,X_test,y_test)
         plt.show()
```
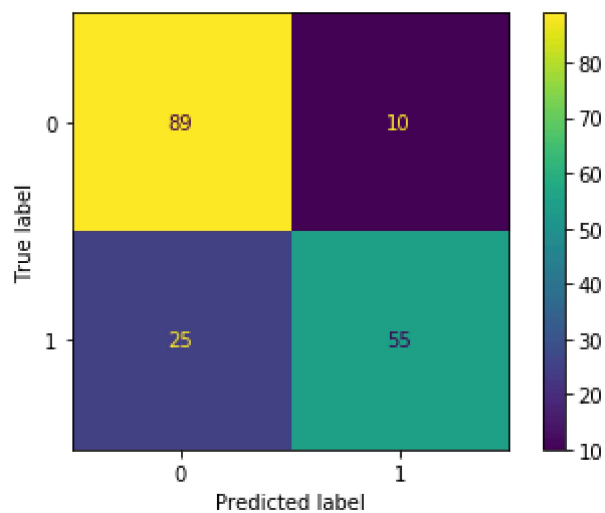
In [53]:
```python
from sklearn.metrics import plot_roc_curve
plot_roc_curve(model,X_test,y_test)
plt.show()
```
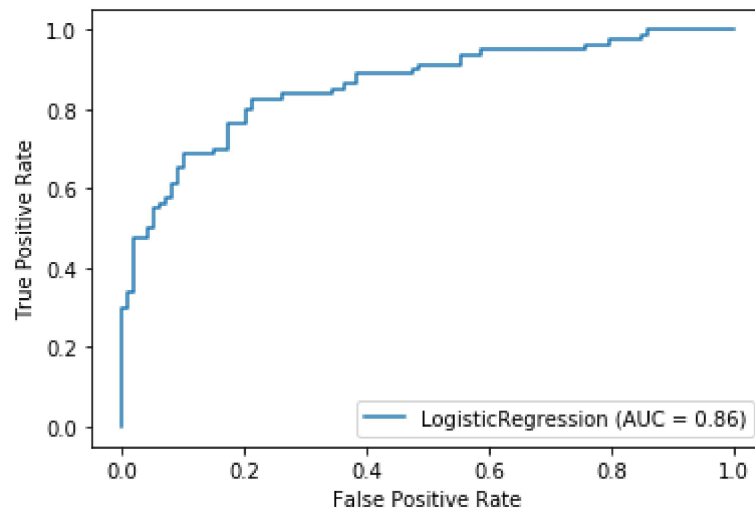

In [54]:
```python
df_t = pd.read_csv("D:\\Kaggle Notebook\\test.csv")
```

In [55]: `df_t`

Out[55]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN
...	...	...	...	...	...	...	...	...	...	...
413	1305	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN
414	1306	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105
415	1307	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN
416	1308	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN
417	1309	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN

418 rows × 11 columns

In [56]: `df_t1=df_t.drop(["Name","Ticket"],axis = "columns")`

In [57]: `df_t1`

Out[57]:

	PassengerId	Pclass	Sex	Age	SibSp	Parch	Fare	Cabin	Embarked
0	892	3	male	34.5	0	0	7.8292	NaN	Q
1	893	3	female	47.0	1	0	7.0000	NaN	S
2	894	2	male	62.0	0	0	9.6875	NaN	Q
3	895	3	male	27.0	0	0	8.6625	NaN	S
4	896	3	female	22.0	1	1	12.2875	NaN	S
...	...	...	...	...	...	...	...	...	...
413	1305	3	male	NaN	0	0	8.0500	NaN	S
414	1306	1	female	39.0	0	0	108.9000	C105	C
415	1307	3	male	38.5	0	0	7.2500	NaN	S
416	1308	3	male	NaN	0	0	8.0500	NaN	S
417	1309	3	male	NaN	1	1	22.3583	NaN	C

418 rows × 9 columns

In [58]: `df_t2 = pd.get_dummies(df_t1.Sex)`

In [59]: `df_t2`

Out[59]:

	female	male
0	0	1
1	1	0
2	0	1
3	0	1
4	1	0
...	...	...
413	0	1
414	1	0
415	0	1
416	0	1
417	0	1

418 rows × 2 columns

In [60]: `df_t3 = pd.concat([df_t1,df_t2],axis = "columns")`

In [61]: `df_t3`

Out[61]:

	PassengerId	Pclass	Sex	Age	SibSp	Parch	Fare	Cabin	Embarked	female	mal
0	892	3	male	34.5	0	0	7.8292	NaN	Q	0	
1	893	3	female	47.0	1	0	7.0000	NaN	S	1	
2	894	2	male	62.0	0	0	9.6875	NaN	Q	0	
3	895	3	male	27.0	0	0	8.6625	NaN	S	0	
4	896	3	female	22.0	1	1	12.2875	NaN	S	1	
...	...	...	...	...	...	...	...	...	...	...	.
413	1305	3	male	NaN	0	0	8.0500	NaN	S	0	
414	1306	1	female	39.0	0	0	108.9000	C105	C	1	
415	1307	3	male	38.5	0	0	7.2500	NaN	S	0	
416	1308	3	male	NaN	0	0	8.0500	NaN	S	0	
417	1309	3	male	NaN	1	1	22.3583	NaN	C	0	

418 rows × 11 columns

In [62]: `df_t4 = df_t3.drop(["Cabin","Sex"],axis="columns")`

In [63]: `df_t4`

Out[63]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	Embarked	female	male
0	892	3	34.5	0	0	7.8292	Q	0	1
1	893	3	47.0	1	0	7.0000	S	1	0
2	894	2	62.0	0	0	9.6875	Q	0	1
3	895	3	27.0	0	0	8.6625	S	0	1
4	896	3	22.0	1	1	12.2875	S	1	0
...	...	...	...	...	...	...	...	...	...
413	1305	3	NaN	0	0	8.0500	S	0	1
414	1306	1	39.0	0	0	108.9000	C	1	0
415	1307	3	38.5	0	0	7.2500	S	0	1
416	1308	3	NaN	0	0	8.0500	S	0	1
417	1309	3	NaN	1	1	22.3583	C	0	1

418 rows × 9 columns

In [64]: `df_t4.isnull().sum()`

Out[64]:
```
PassengerId      0
Pclass           0
Age             86
SibSp            0
Parch            0
Fare             1
Embarked         0
female           0
male             0
dtype: int64
```

In [65]:
```python
from sklearn.impute import SimpleImputer
mean_imputer = SimpleImputer(strategy = "mean")
df_t4["Age"] = pd.DataFrame(mean_imputer.fit_transform(df_t4[["Age"]]))
```

In [66]: `df_t4.isnull().sum()`

Out[66]:
```
PassengerId      0
Pclass           0
Age              0
SibSp            0
Parch            0
Fare             1
Embarked         0
female           0
male             0
dtype: int64
```

In [67]: `df_t4["Fare"].mean()`

Out[67]: 35.6271884892086

In [68]:
```python
from sklearn.impute import SimpleImputer
mean_imputer = SimpleImputer(strategy = "mean")
df_t4["Fare"]=  pd.DataFrame(mean_imputer.fit_transform(df_t4[["Fare"]]))
```

In [69]: `df_t4.isnull().sum()`

Out[69]:
```
PassengerId      0
Pclass           0
Age              0
SibSp            0
Parch            0
Fare             0
Embarked         0
female           0
male             0
dtype: int64
```

In [70]: `df_t5=pd.get_dummies(df_t4.Embarked)`

In [71]: `df_t5.head(10)`

Out[71]:

	C	Q	S
0	0	1	0
1	0	0	1
2	0	1	0
3	0	0	1
4	0	0	1
5	0	0	1
6	0	1	0
7	0	0	1
8	1	0	0
9	0	0	1

In [72]: `df_t6 = pd.concat([df_t4,df_t5],axis= "columns")`

In [73]: `df_test=df_t6.drop(["Embarked"],axis = "columns")`

In [74]: `df_test`

Out[74]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	female	male	C	Q	S
0	892	3	34.50000	0	0	7.8292	0	1	0	1	0
1	893	3	47.00000	1	0	7.0000	1	0	0	0	1
2	894	2	62.00000	0	0	9.6875	0	1	0	1	0
3	895	3	27.00000	0	0	8.6625	0	1	0	0	1
4	896	3	22.00000	1	1	12.2875	1	0	0	0	1
...	...	...	...	...	...	...	...	...	...	...	...
413	1305	3	30.27259	0	0	8.0500	0	1	0	0	1
414	1306	1	39.00000	0	0	108.9000	1	0	1	0	0
415	1307	3	38.50000	0	0	7.2500	0	1	0	0	1
416	1308	3	30.27259	0	0	8.0500	0	1	0	0	1
417	1309	3	30.27259	1	1	22.3583	0	1	1	0	0

418 rows × 11 columns

In [79]:
```python
model.predict(df_test)
```

Out[79]:
```
array([0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0,
       1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1,
       1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1,
       1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1,
       1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1,
       0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1,
       1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
       0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0,
       1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1,
       1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1,
       0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0,
       0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0,
       0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0,
       1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0,
       1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0,
       1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1,
       0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0],
      dtype=int64)
```

In [ ]: