In [239]:
```
import numpy as np
import pandas as pd
from collections import Counter
import warnings
warnings.filterwarnings('ignore')
import seaborn as sns
import matplotlib.pyplot as plt
```

In [240]:
```
df_Tiatanic=pd.read_csv('datasets_train.csv')
```

In [241]:
```
df_Tiatanic
```

Out[241]:

	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	
...	...	...	...	...	...	...	...	...	...	...	
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	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	

891 rows × 12 columns

In [242]:
```
df_Tiatanic.head(20)
```

Out[242]:

	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
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500
12	13	0	3	Saundercock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.0500
13	14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31.2750
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8542

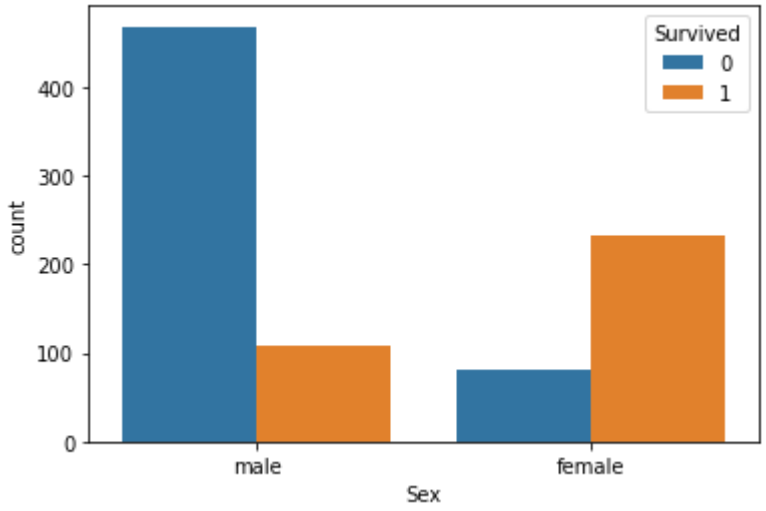
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706	16.0000
16	17	0	3	Rice, Master. Eugene	male	2.0	4	1	382652	29.1250
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0	1	0	345763	18.0000
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2250

In [243]:
```
df_Tiatanic.isnull().sum()
```

Out[243]:
```
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
```

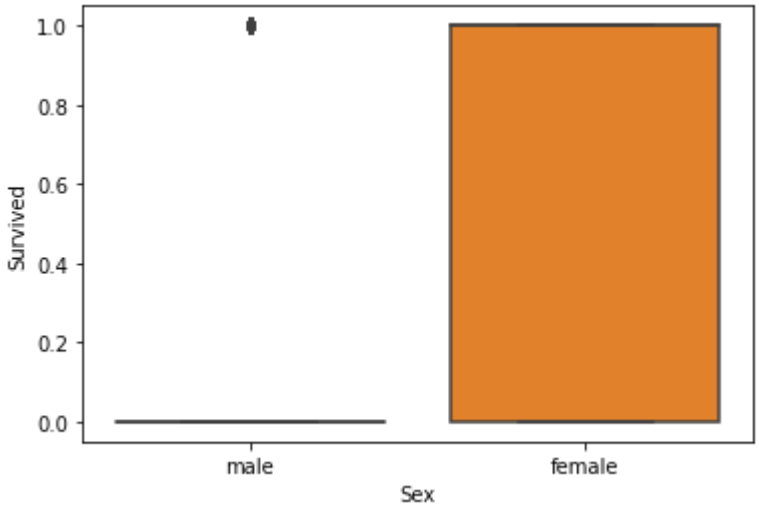
In [244]:
```
sns.countplot('Sex',hue='Survived',data=df_Tiatanic)
```

Out[244]: <AxesSubplot:xlabel='Sex', ylabel='count'>



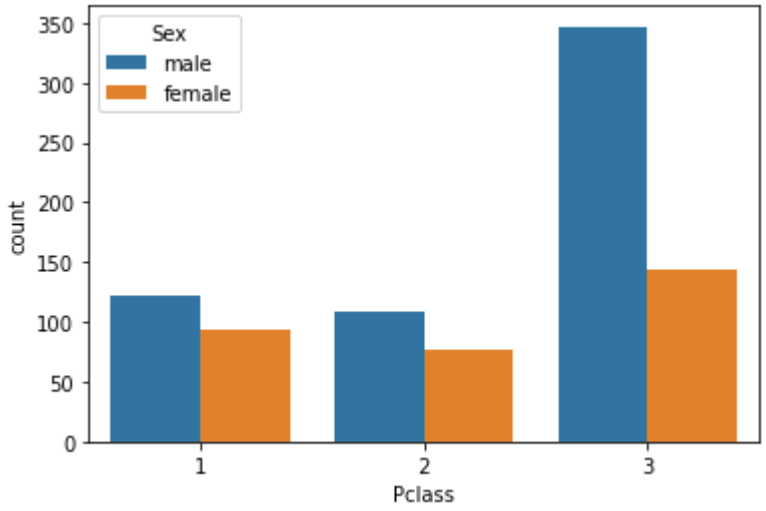
In [245]:
```
sns.boxplot(y='Survived',x='Sex',data=df_Tiatanic)
```

Out[245]: <AxesSubplot:xlabel='Sex', ylabel='Survived'>



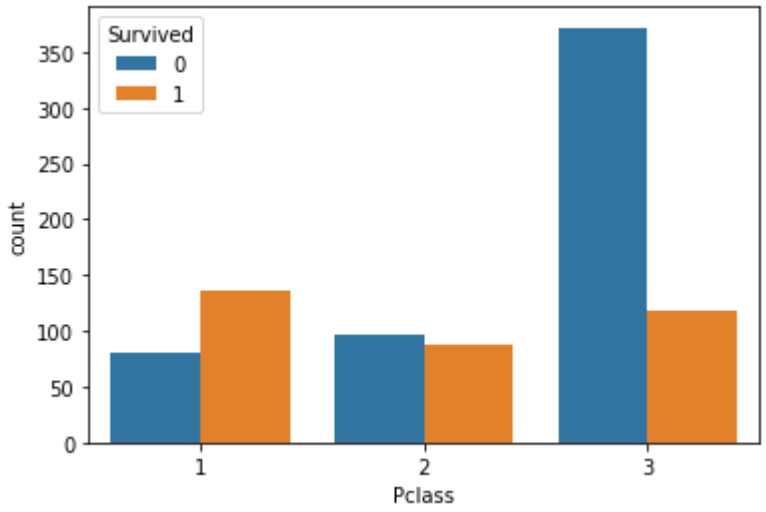
In [246]:
```
sns.countplot('Pclass',hue='Sex',data=df_Tiatanic)
```

Out[246]: <AxesSubplot:xlabel='Pclass', ylabel='count'>



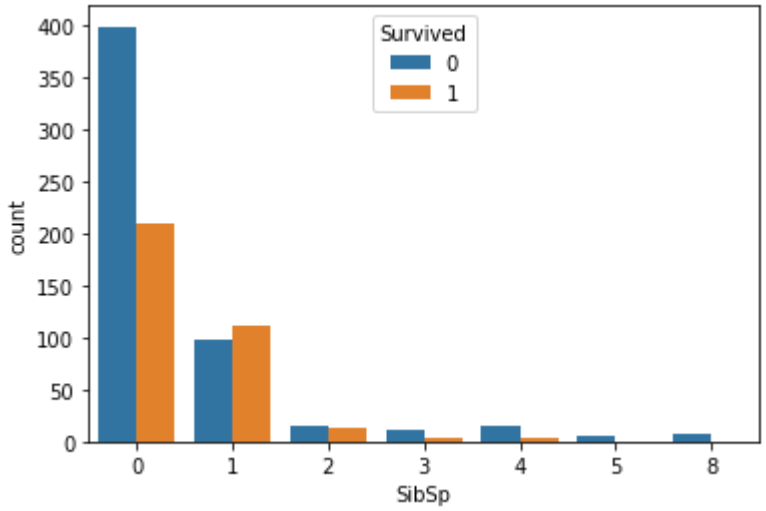
In [247]:
```
sns.countplot('Pclass',hue='Survived',data=df_Tiatanic)
```

Out[247]: <AxesSubplot:xlabel='Pclass', ylabel='count'>



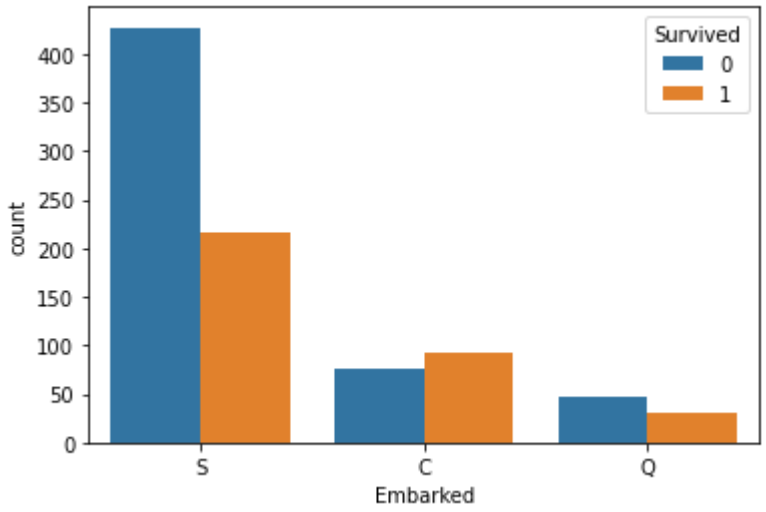
In [248]:
```
sns.countplot('SibSp',hue='Survived',data=df_Tiatanic)
```

Out[248]:
```
<AxesSubplot:xlabel='SibSp', ylabel='count'>
```



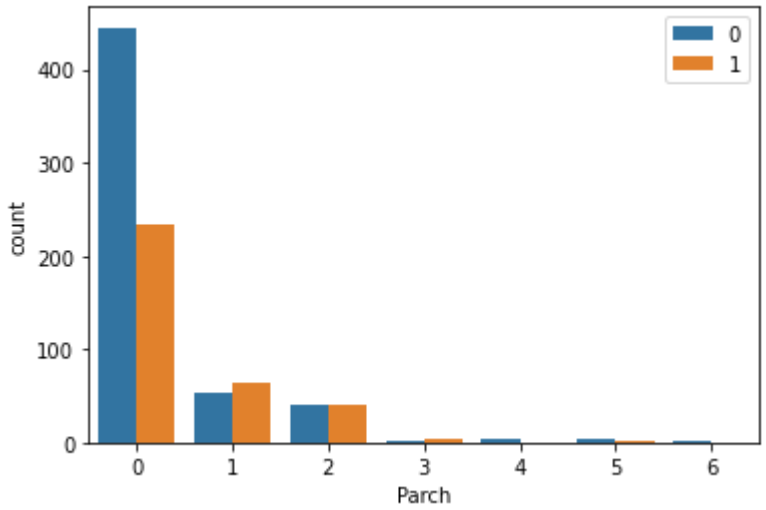
In [249]:
```
sns.countplot('Embarked',hue='Survived',data=df_Tiatanic)
```

Out[249]:
```
<AxesSubplot:xlabel='Embarked', ylabel='count'>
```



In [250]:
```
sns.countplot('Parch',hue='Survived',data=df_Tiatanic,)
plt.legend(loc='upper right')
```

Out[250]:
```
<matplotlib.legend.Legend at 0x135b9196e80>
```



In [251]:
```
df_Tiatanic.describe()
```

Out[251]:

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

In [252]:
```
df_Tiatanic['Age'].value_counts()
```

Out[252]:
```
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 [253]:
```
df_Tiatanic['Sex'].value_counts()
```

Out[253]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [254]:
```
df_Tiatanic['Sex'].replace(['male','female'],[0,1],inplace=True)
```

In [255]:
```
df_Tiatanic
```

Out[255]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
0	1	0	3	Braund, Mr. Owen Harris	0	22.0	1	0	A/5 21171	7.2500	Na
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	1	38.0	1	0	PC 17599	71.2833	C8
2	3	1	3	Heikkinen, Miss. Laina	1	26.0	0	0	STON/O2. 3101282	7.9250	Na
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	35.0	1	0	113803	53.1000	C12
4	5	0	3	Allen, Mr. William Henry	0	35.0	0	0	373450	8.0500	Na
...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	0	27.0	0	0	211536	13.0000	Na
887	888	1	1	Graham, Miss. Margaret Edith	1	19.0	0	0	112053	30.0000	B4
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	1	NaN	1	2	W./C. 6607	23.4500	Na
889	890	1	1	Behr, Mr. Karl Howell	0	26.0	0	0	111369	30.0000	C14
890	891	0	3	Dooley, Mr. Patrick	0	32.0	0	0	370376	7.7500	Na

891 rows × 12 columns

In [256]:
```
print(df_Tiatanic['Embarked'].mode())
```

```
0    S
Name: Embarked, dtype: object
```

In [257]:
```
df_Tiatanic['Sex'].value_counts()
```

Out[257]:
```
0    577
1    314
Name: Sex, dtype: int64
```

In [258]:
```
df_Tiatanic['Embarked'].value_counts()
```

Out[258]:
```
S    644
C    168
Q     77
Name: Embarked, dtype: int64
```

In [259]:
```
df_Tiatanic['Embarked'].replace(['S','C','Q'],[0,1,2],inplace=True)
```

In [260]:
```
df_Tiatanic
```

Out[260]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
0	1	0	3	Braund, Mr. Owen Harris	0	22.0	1	0	A/5 21171	7.2500	Na
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	1	38.0	1	0	PC 17599	71.2833	C8
2	3	1	3	Heikkinen, Miss. Laina	1	26.0	0	0	STON/O2. 3101282	7.9250	Na
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	35.0	1	0	113803	53.1000	C12
4	5	0	3	Allen, Mr. William Henry	0	35.0	0	0	373450	8.0500	Na
...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	0	27.0	0	0	211536	13.0000	Na
887	888	1	1	Graham, Miss. Margaret Edith	1	19.0	0	0	112053	30.0000	B4
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	1	NaN	1	2	W./C. 6607	23.4500	Na
889	890	1	1	Behr, Mr. Karl Howell	0	26.0	0	0	111369	30.0000	C14
890	891	0	3	Dooley, Mr. Patrick	0	32.0	0	0	370376	7.7500	Na

891 rows × 12 columns

In [261]:
```
df_Tiatanic=df_Tiatanic.drop('Cabin',axis=1)
```

In [262]:
```
df_Tiatanic=df_Tiatanic.drop('Ticket',axis=1)
df_Tiatanic=df_Tiatanic.drop('Name',axis=1)
```

In [263]:
```
df_Tiatanic
```

Out[263]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	1	0	3	0	22.0	1	0	7.2500	0.0
1	2	1	1	1	38.0	1	0	71.2833	1.0
2	3	1	3	1	26.0	0	0	7.9250	0.0
3	4	1	1	1	35.0	1	0	53.1000	0.0
4	5	0	3	0	35.0	0	0	8.0500	0.0
...	...	...	...	...	...	...	...	...	...
886	887	0	2	0	27.0	0	0	13.0000	0.0
887	888	1	1	1	19.0	0	0	30.0000	0.0
888	889	0	3	1	NaN	1	2	23.4500	0.0
889	890	1	1	0	26.0	0	0	30.0000	1.0
890	891	0	3	0	32.0	0	0	7.7500	2.0

891 rows × 9 columns

In [264]:
```
df_Tiatanic.isnull().sum()
```

Out[264]:
```
PassengerId      0
Survived         0
Pclass           0
Sex              0
Age            177
SibSp            0
Parch            0
Fare             0
Embarked         2
dtype: int64
```

In [265]:
```
df_Tiatanic['Embarked'].fillna(df_Tiatanic['Embarked'].mean(),inplace=True)
```

In [266]:
```
df_Tiatanic
```

Out[266]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	1	0	3	0	22.0	1	0	7.2500	0.0
1	2	1	1	1	38.0	1	0	71.2833	1.0
2	3	1	3	1	26.0	0	0	7.9250	0.0
3	4	1	1	1	35.0	1	0	53.1000	0.0
4	5	0	3	0	35.0	0	0	8.0500	0.0
...	...	...	...	...	...	...	...	...	...
886	887	0	2	0	27.0	0	0	13.0000	0.0
887	888	1	1	1	19.0	0	0	30.0000	0.0
888	889	0	3	1	NaN	1	2	23.4500	0.0
889	890	1	1	0	26.0	0	0	30.0000	1.0
890	891	0	3	0	32.0	0	0	7.7500	2.0

891 rows × 9 columns

In [267]:
```
df_Tiatanic.isnull().sum()
```

Out[267]:
```
PassengerId      0
Survived         0
Pclass           0
Sex              0
Age            177
SibSp            0
Parch            0
Fare             0
Embarked         0
dtype: int64
```

In [268]:
```
df_Tiatanic['Age'].fillna(df_Tiatanic['Age'].mean(),inplace=True)
```

In [269]:
```
df_Tiatanic
```

Out[269]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	1	0	3	0	22.000000	1	0	7.2500	0.0
1	2	1	1	1	38.000000	1	0	71.2833	1.0
2	3	1	3	1	26.000000	0	0	7.9250	0.0
3	4	1	1	1	35.000000	1	0	53.1000	0.0
4	5	0	3	0	35.000000	0	0	8.0500	0.0
...	...	...	...	...	...	...	...	...	...
886	887	0	2	0	27.000000	0	0	13.0000	0.0
887	888	1	1	1	19.000000	0	0	30.0000	0.0
888	889	0	3	1	29.699118	1	2	23.4500	0.0
889	890	1	1	0	26.000000	0	0	30.0000	1.0
890	891	0	3	0	32.000000	0	0	7.7500	2.0

891 rows × 9 columns

In [270]:
```
df_Tiatanic.isnull().sum()
```

Out[270]:
```
PassengerId    0
Survived       0
Pclass         0
Sex            0
Age            0
SibSp          0
Parch          0
Fare           0
Embarked       0
dtype: int64
```

Train and Test split

In [271]:
```
X=df_Tiatanic.drop('Survived',axis=1)
```

In [272]:
```
y=df_Tiatanic['Survived']
```

In [275]:
```
X
```

Out[275]:

	PassengerId	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	1	3	0	22.000000	1	0	7.2500	0.0
1	2	1	1	38.000000	1	0	71.2833	1.0
2	3	3	1	26.000000	0	0	7.9250	0.0
3	4	1	1	35.000000	1	0	53.1000	0.0
4	5	3	0	35.000000	0	0	8.0500	0.0
...	...	...	...	...	...	...	...	...
886	887	2	0	27.000000	0	0	13.0000	0.0
887	888	1	1	19.000000	0	0	30.0000	0.0
888	889	3	1	29.699118	1	2	23.4500	0.0
889	890	1	0	26.000000	0	0	30.0000	1.0
890	891	3	0	32.000000	0	0	7.7500	2.0

891 rows × 8 columns

In [276]:
```
y
```

Out[276]:
```
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 [278]:
```
from sklearn.model_selection import train_test_split

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3)
```

In [279]:
```
X_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 623 entries, 327 to 79
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  623 non-null    int64  
 1   Pclass       623 non-null    int64  
 2   Sex          623 non-null    int64  
 3   Age          623 non-null    float64
 4   SibSp        623 non-null    int64  
 5   Parch        623 non-null    int64  
 6   Fare         623 non-null    float64
 7   Embarked     623 non-null    float64
dtypes: float64(3), int64(5)
memory usage: 43.8 KB
```

Logistic Regression Algorithm

In [280]:
```
from sklearn.linear_model import LogisticRegression
```

In [285]:
```
logmodel=LogisticRegression()
```

In [286]:
```
logmodel.fit(X_train,y_train)
```

Out[286]:
```
LogisticRegression()
```

In [288]:
```
predictions=logmodel.predict(X_test)
```

In [289]:
```
from sklearn.metrics import classification_report
```

In [290]:
```
print(classification_report(y_test,predictions))
```

```
              precision    recall  f1-score   support

           0       0.85      0.87      0.86       171
           1       0.76      0.73      0.74        97

    accuracy                           0.82       268
   macro avg       0.80      0.80      0.80       268
weighted avg       0.82      0.82      0.82       268
```

In [ ]:
```

```