In [83]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import math
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
```

In [84]:
```
filename = 'titanic_data'
path = 'E:/desktop/ML/Logistic/{}.csv'.format(filename)
titanic_data = pd.read_csv(path)
titanic_data.head()
```

Out[84]:

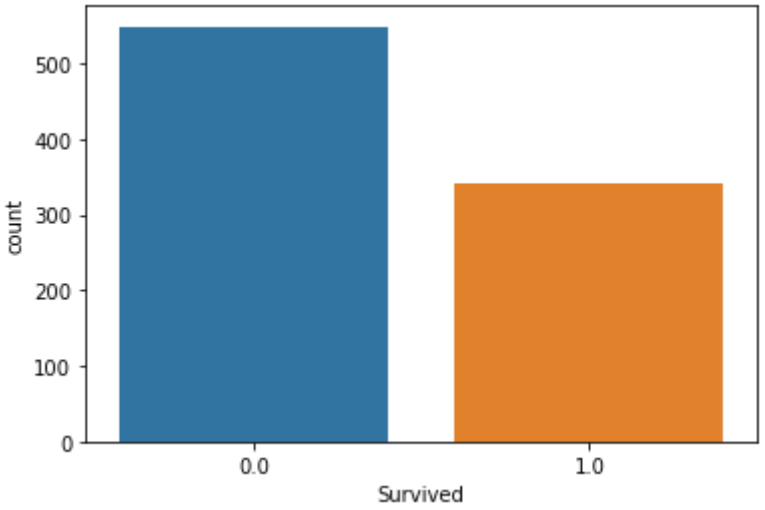
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	...	Embarked	WikiId	Name_wiki	Age_wiki	Ho
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	...	S	691.0	Braund, Mr. Owen Harris	22.0	B
1	2	1.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	...	C	90.0	Cumings, Mrs. Florence Briggs (née Thayer)	35.0	New Y
2	3	1.0	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	...	S	865.0	Heikkinen, Miss Laina	26.0	J
3	4	1.0	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	...	S	127.0	Futrelle, Mrs. Lily May (née Peel)	35.0	Massa
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	...	S	627.0	Allen, Mr. William Henry	35.0	Birr West I

5 rows × 21 columns

In [85]:
```
## Analyse Data
```

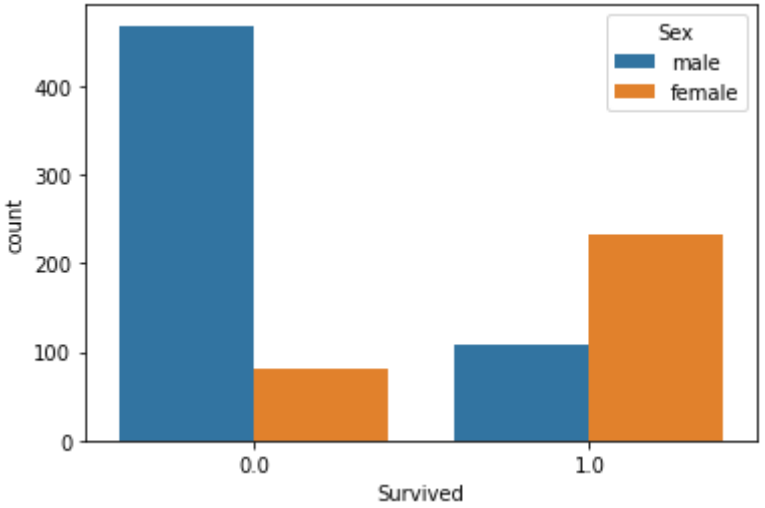
In [86]:
```
sns.countplot(x='Survived',data=titanic_data)
```

Out[86]: <AxesSubplot:xlabel='Survived', ylabel='count'>



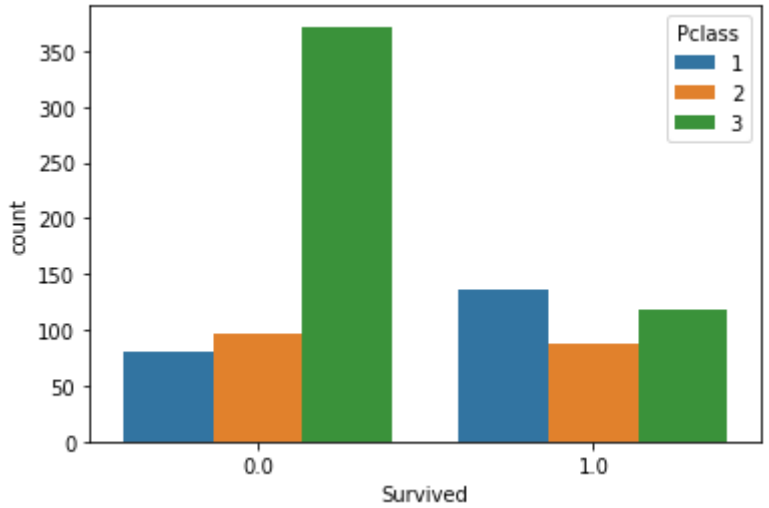
In [87]:
```
sns.countplot(x='Survived',hue='Sex',data=titanic_data)
```

Out[87]: <AxesSubplot:xlabel='Survived', ylabel='count'>



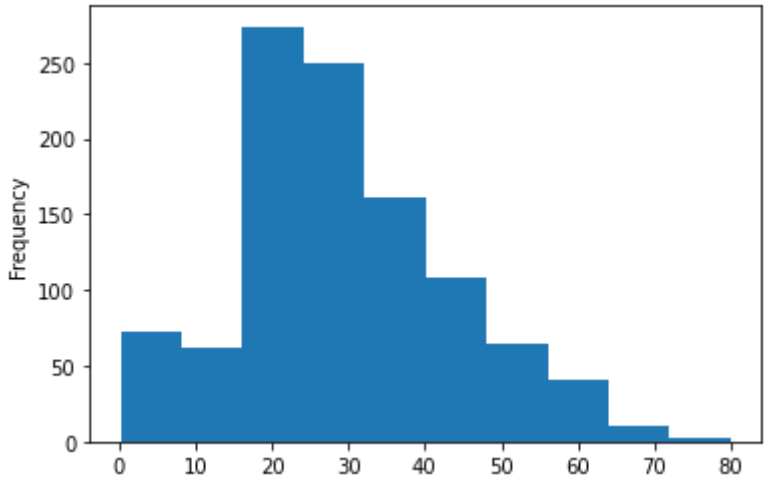
In [88]:
```
sns.countplot(x='Survived',hue='Pclass',data=titanic_data)
```

Out[88]: <AxesSubplot:xlabel='Survived', ylabel='count'>



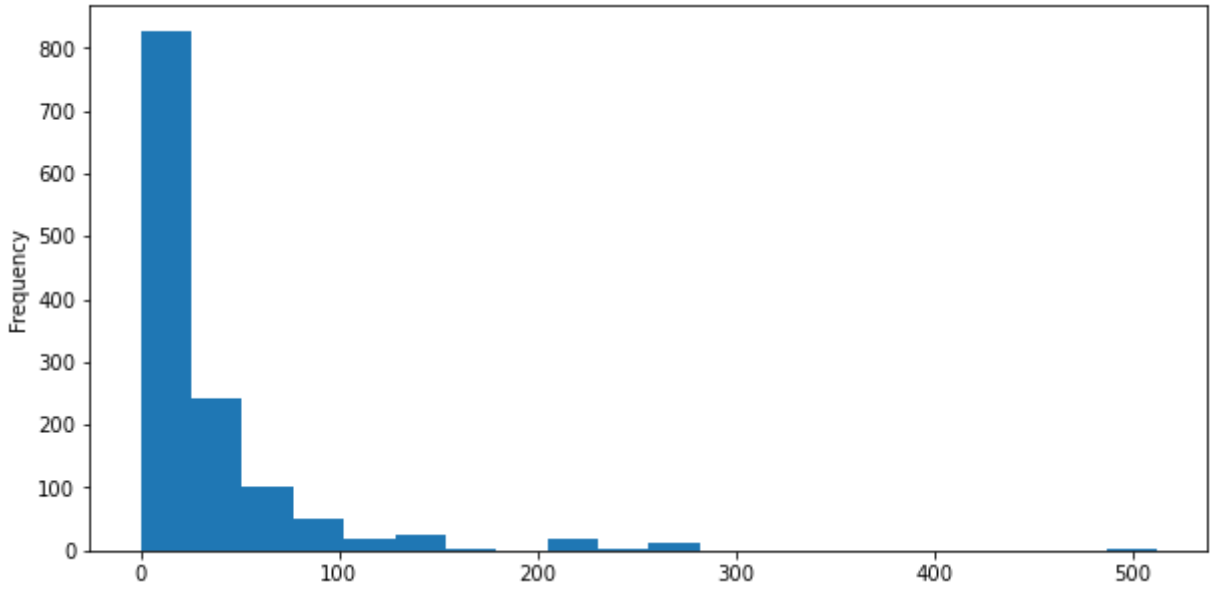
In [89]:
```
titanic_data['Age'].plot.hist()
```

Out[89]: <AxesSubplot:ylabel='Frequency'>



In [90]:
```
titanic_data['Fare'].plot.hist(bins=20,figsize=(10,5))
```

Out[90]: <AxesSubplot:ylabel='Frequency'>



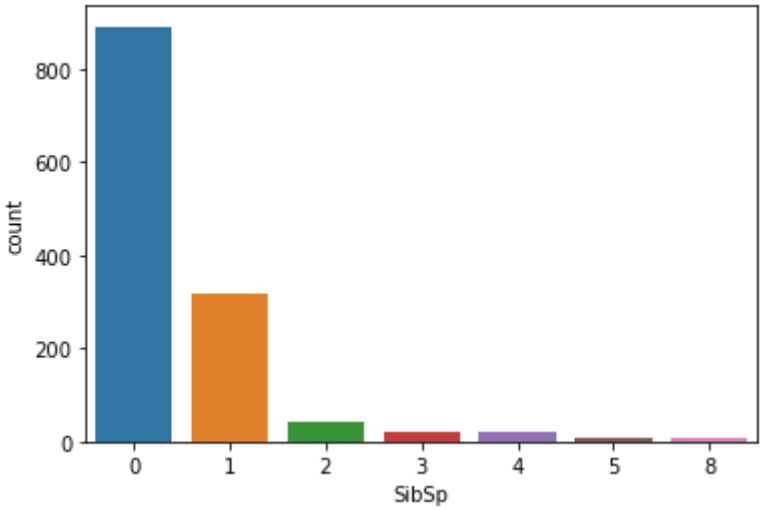
In [91]:
```
titanic_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 21 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  1309 non-null   int64  
 1   Survived     891 non-null    float64
 2   Pclass       1309 non-null   int64  
 3   Name         1309 non-null   object 
 4   Sex          1309 non-null   object 
 5   Age          1046 non-null   float64
 6   SibSp        1309 non-null   int64  
 7   Parch        1309 non-null   int64  
 8   Ticket       1309 non-null   object 
 9   Fare         1308 non-null   float64
 10  Cabin        295 non-null    object 
 11  Embarked     1307 non-null   object 
 12  WikiId       1304 non-null   float64
 13  Name_wiki    1304 non-null   object 
 14  Age_wiki     1302 non-null   float64
 15  Hometown     1304 non-null   object 
 16  Boarded      1304 non-null   object 
 17  Destination  1304 non-null   object 
 18  Lifeboat     502 non-null    object 
 19  Body         130 non-null    object 
 20  Class        1304 non-null   float64
dtypes: float64(6), int64(4), object(11)
memory usage: 214.9+ KB
```

In [92]:
```
#Age of child in titanic
```

In [93]:
```
sns.countplot(x='SibSp',data=titanic_data)
```

Out[93]: <AxesSubplot:xlabel='SibSp', ylabel='count'>



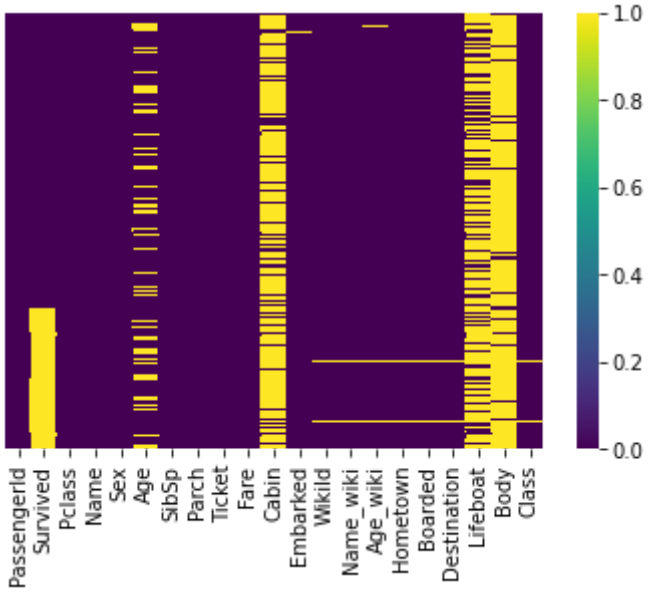
In [94]:
```
# data Wrangling : removing Nan values or unnecessary values
```

In [95]:
```
titanic_data.isnull().sum()
```

Out[95]:
```
PassengerId       0
Survived        418
Pclass            0
Name              0
Sex               0
Age             263
SibSp             0
Parch             0
Ticket            0
Fare              1
Cabin          1014
Embarked          2
WikiId            5
Name_wiki         5
Age_wiki          7
Hometown          5
Boarded           5
Destination       5
Lifeboat        807
Body           1179
Class             5
dtype: int64
```

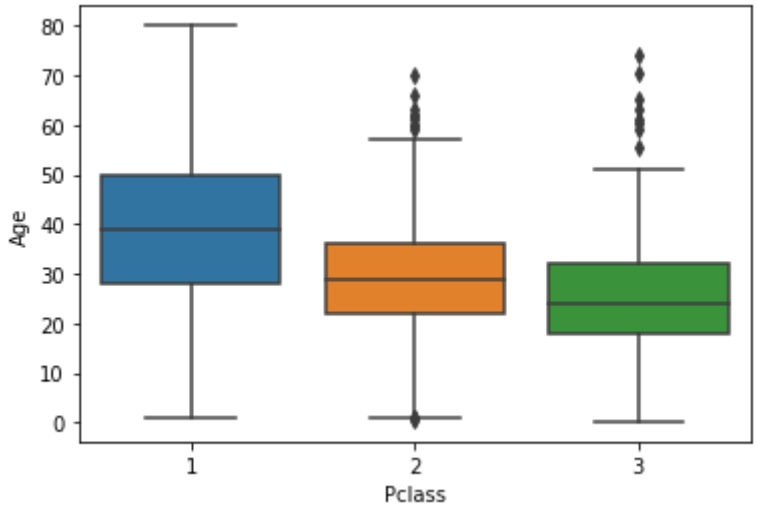
In [96]:
```
sns.heatmap(titanic_data.isnull(),yticklabels=False,cmap='viridis')
```

Out[96]: <AxesSubplot:>



In [97]:
```
sns.boxplot(x='Pclass',y='Age',data=titanic_data)
```

Out[97]:
```
<AxesSubplot:xlabel='Pclass', ylabel='Age'>
```



In [98]:
```
titanic_data.isnull()
```

Out[98]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	...	Embarked	WikiId	Name_wiki	Age_wiki	Hometown
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1304	False	True	False	False	False	True	False	False	False	False	...	False	False	False	False	False
1305	False	True	False	False	False	False	False	False	False	False	...	False	False	False	False	False
1306	False	True	False	False	False	False	False	False	False	False	...	False	False	False	False	False
1307	False	True	False	False	False	True	False	False	False	False	...	False	False	False	False	False
1308	False	True	False	False	False	True	False	False	False	False	...	False	False	False	False	False

1309 rows × 21 columns

In [99]:
```
# titanic_data.dropna(inplace=True)
titanic_data.head(5)
```

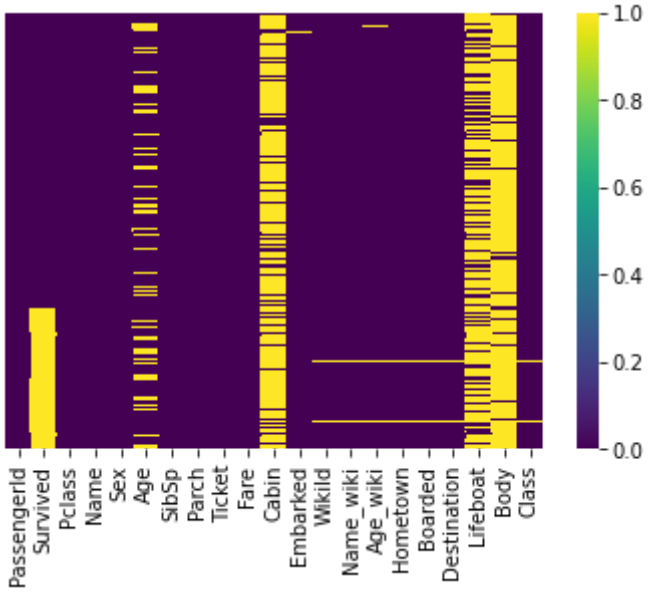
Out[99]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	...	Embarked	WikiId	Name_wiki	Age_wiki	Ho
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	...	S	691.0	Braund, Mr. Owen Harris	22.0	B
1	2	1.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	...	C	90.0	Cumings, Mrs. Florence Briggs (née Thayer)	35.0	New Y
2	3	1.0	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	...	S	865.0	Heikkinen, Miss Laina	26.0	J
3	4	1.0	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	...	S	127.0	Futrelle, Mrs. Lily May (née Peel)	35.0	Massa
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	...	S	627.0	Allen, Mr. William Henry	35.0	Birr West I

5 rows × 21 columns

In [100]:
```
sns.heatmap(titanic_data.isnull(),yticklabels=False,cmap='viridis')
```

Out[100]: <AxesSubplot:>



In [101]:
```
titanic_data.drop(['Cabin','Lifeboat','Body'],axis=1,inplace=True)
titanic_data.dropna(inplace=True)
titanic_data.isnull().sum()
```

Out[101]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
WikiId         0
Name_wiki      0
Age_wiki       0
Hometown       0
Boarded        0
Destination    0
Class          0
dtype: int64
```

In [102]:
```
titanic_data.head(2)
```

Out[102]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	WikiId	Name_wiki	Age_wiki	Hometown	
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S	691.0	Braund, Mr. Owen Harris	22.0	Bridgerule, Devon, England	S
1	2	1.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C	90.0	Cumings, Mrs. Florence Briggs (née Thayer)	35.0	New York, New York, US	

In [103]:
```
sex=pd.get_dummies(titanic_data['Sex'],drop_first=True)
sex.head(5)
```

Out[103]:

	male
0	1
1	0
2	0
3	0
4	1

In [104]:
```
embark=pd.get_dummies(titanic_data['Embarked'],drop_first=True)
embark
```

Out[104]:

	Q	S
0	0	1
1	0	0
2	0	1
3	0	1
4	0	1
...	...	...
885	1	0
886	0	1
887	0	1
889	0	0
890	1	0

712 rows × 2 columns

In [105]:
```
pc1=pd.get_dummies(titanic_data['Pclass'],drop_first=True)
pc1.head(5)
```

Out[105]:

	2	3
0	0	1
1	0	0
2	0	1
3	0	0
4	0	1

In [106]:
```
titanic_data=pd.concat([titanic_data,sex,embark,pc1],axis=1)
titanic_data.head(5)
```

Out[106]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	...	Age_wiki	Hometown	Boarded	Desti
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	...	22.0	Bridgerule, Devon, England	Southampton	Qu'A Saskatch C
1	2	1.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	...	35.0	New York, New York, US	Cherbourg	New Yorl Yc
2	3	1.0	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	...	26.0	Jyväskylä, Finland	Southampton	New Yc
3	4	1.0	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	...	35.0	Scituate, Massachusetts, US	Southampton	Sc Massach
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	...	35.0	Birmingham, West Midlands, England	Southampton	New Yc

5 rows × 23 columns

In [107]:
```
titanic_data.drop(['Class','Boarded','Age_wiki','WikiId','Sex','Embarked','Pclass','PassengerId','Name','Ticket','Name_wiki','Hometown','Destination'],axis=1,inplace=True)
titanic_data.head(5)
```

Out[107]:

	Survived	Age	SibSp	Parch	Fare	male	Q	S	2	3
0	0.0	22.0	1	0	7.2500	1	0	1	0	1
1	1.0	38.0	1	0	71.2833	0	0	0	0	0
2	1.0	26.0	0	0	7.9250	0	0	1	0	1
3	1.0	35.0	1	0	53.1000	0	0	1	0	0
4	0.0	35.0	0	0	8.0500	1	0	1	0	1

Train dataset

In [108]:
```
X = titanic_data.drop('Survived',axis=1)
y = titanic_data['Survived']
```

In [116]:
```
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.3,random_state=1)
```

In [131]:
```
logmodel = LogisticRegression(solver='lbfgs')
```

In [132]:
```
logmodel.fit(X_train,y_train)
```

```
c:\users\surajc\appdata\local\programs\python\python38\lib\site-packages\sklearn\linear_model\_logistic.py:762: Conve
rgenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

Out[132]:
```
LogisticRegression()
```

In [134]:
```
predictions=logmodel.predict(X_test)
```

In [137]:
```
classification_report(y_test,predictions)
```

Out[137]:
```
'              precision    recall  f1-score   support\n\n         0.0       0.80      0.81      0.81       126\n
1.0       0.72      0.72      0.72        88\n\n    accuracy                           0.77       214\n   macro avg
0.76      0.76      0.76       214\nweighted avg       0.77      0.77      0.77       214\n'
```

In [139]:
```
confusion_matrix(y_test,predictions)
```

Out[139]:
```
array([[102,  24],
       [ 25,  63]], dtype=int64)
```

In [141]:
```
accuracy_score(y_test,predictions)
```

Out[141]:
```
0.7710280373831776
```

Accuracy of the fitted model is 77%