In [123]:
```
import pandas as pd
```

In [124]:
```
data=pd.read_csv("/home/placement/Downloads/Titanic Dataset.csv")
```

In [125]:
```
data
```

Out[125]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [126]:
```
data.describe()
```

Out[126]:

	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 [127]:
```
data.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 [128]:
```
data.isna()
```

Out[128]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	False	False	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	True	False
3	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	True	False
...	...	...	...	...	...	...	...	...	...	...	...	...
886	False	False	False	False	False	False	False	False	False	False	True	False
887	False	False	False	False	False	False	False	False	False	False	False	False
888	False	False	False	False	False	True	False	False	False	False	True	False
889	False	False	False	False	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False	False	False	True	False

891 rows × 12 columns

In [129]:
```
data.Pclass.unique()
```

Out[129]:
```
array([3, 1, 2])
```

In [130]:
```
data.Survived.unique()
```

Out[130]:
```
array([0, 1])
```

In [131]:
```
data.SibSp.unique()
```

Out[131]:
```
array([1, 0, 3, 4, 2, 5, 8])
```

In [132]:
```
data.Parch.unique()
```

Out[132]:
```
array([0, 1, 2, 5, 3, 4, 6])
```

In [133]:
```
data.Age.unique()
```

Out[133]:
```
array([22.  , 38.  , 26.  , 35.  ,   nan, 54.  ,  2.  , 27.  , 14.  ,
        4.  , 58.  , 20.  , 39.  , 55.  , 31.  , 34.  , 15.  , 28.  ,
        8.  , 19.  , 40.  , 66.  , 42.  , 21.  , 18.  ,  3.  ,  7.  ,
       49.  , 29.  , 65.  , 28.5 ,  5.  , 11.  , 45.  , 17.  , 32.  ,
       16.  , 25.  ,  0.83, 30.  , 33.  , 23.  , 24.  , 46.  , 59.  ,
       71.  , 37.  , 47.  , 14.5 , 70.5 , 32.5 , 12.  ,  9.  , 36.5 ,
       51.  , 55.5 , 40.5 , 44.  ,  1.  , 61.  , 56.  , 50.  , 36.  ,
       45.5 , 20.5 , 62.  , 41.  , 52.  , 63.  , 23.5 ,  0.92, 43.  ,
       60.  , 10.  , 64.  , 13.  , 48.  ,  0.75, 53.  , 57.  , 80.  ,
       70.  , 24.5 ,  6.  ,  0.67, 30.5 ,  0.42, 34.5 , 74.  ])
```

In [134]:
```
data1=data.drop(['PassengerId','Cabin','Name','Ticket','SibSp','Parch'],axis=1)
```

In [135]:
```
data1
```

Out[135]:

	Survived	Pclass	Sex	Age	Fare	Embarked
0	0	3	male	22.0	7.2500	S
1	1	1	female	38.0	71.2833	C
2	1	3	female	26.0	7.9250	S
3	1	1	female	35.0	53.1000	S
4	0	3	male	35.0	8.0500	S
...	...	...	...	...	...	...
886	0	2	male	27.0	13.0000	S
887	1	1	female	19.0	30.0000	S
888	0	3	female	NaN	23.4500	S
889	1	1	male	26.0	30.0000	C
890	0	3	male	32.0	7.7500	Q

891 rows × 6 columns

In [136]:
```
data1.shape
```

Out[136]:
```
(891, 6)
```

In [137]:
```
data1['Sex']=data['Sex'].map({'male':1,'female':0})
```

In [138]:
```
data2=data1.fillna(data.median())
```

```
/tmp/ipykernel_11163/1290514040.py:1: FutureWarning: The default value of numeric_only in DataFrame.median
is deprecated. In a future version, it will default to False. In addition, specifying 'numeric_only=None' i
s deprecated. Select only valid columns or specify the value of numeric_only to silence this warning.
  data2=data1.fillna(data.median())
```

In [139]:
```
data2.isna().sum()
```

Out[139]:
```
Survived    0
Pclass      0
Sex         0
Age         0
Fare        0
Embarked    2
dtype: int64
```

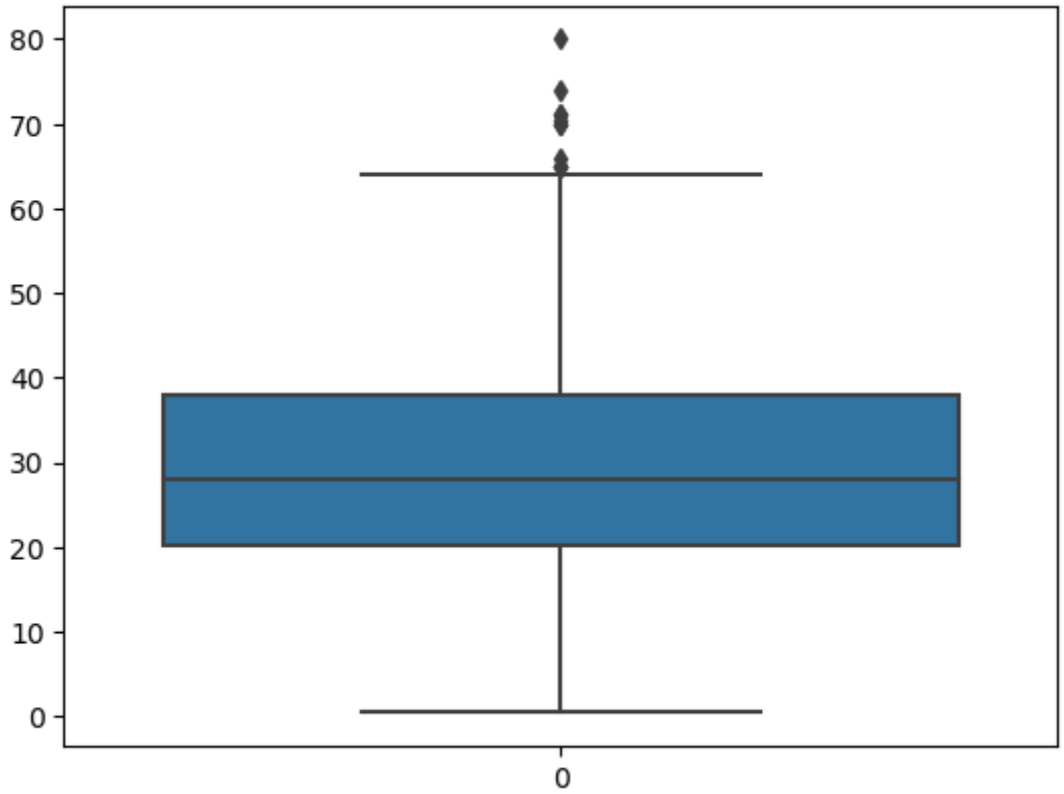
In [140]:
```
data.head()
```

Out[140]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

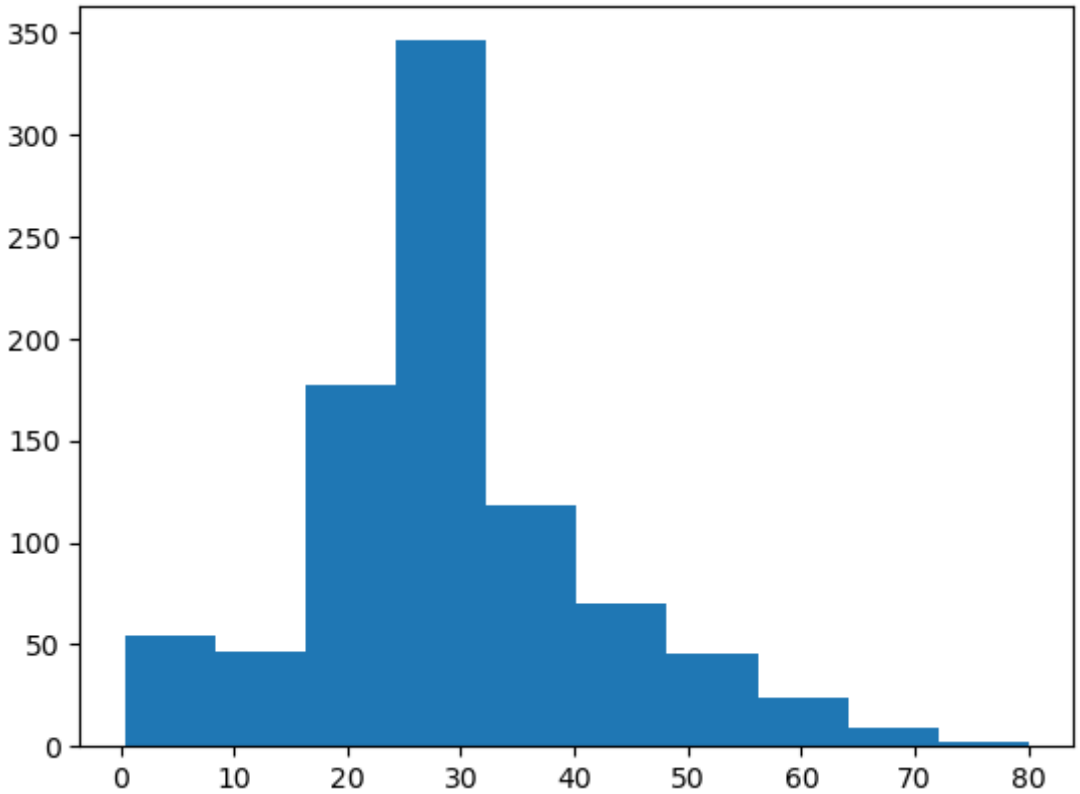
In [141]:
```
import seaborn as sns
import matplotlib.pyplot as plt
sns.boxplot(data.Age)
```

Out[141]: <Axes: >



In [142]:
```
plt.hist(data2['Age'])
```

Out[142]:
```
(array([ 54.,  46., 177., 346., 118.,  70.,  45.,  24.,   9.,   2.]),
 array([ 0.42 ,  8.378, 16.336, 24.294, 32.252, 40.21 , 48.168, 56.126,
        64.084, 72.042, 80.   ]),
 <BarContainer object of 10 artists>)
```



In [143]:
```
data.describe()
```

Out[143]:

	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 [144]:
```
data['Age'].unique()
```

Out[144]:
```
array([22.  , 38.  , 26.  , 35.  ,   nan, 54.  ,  2.  , 27.  , 14.  ,
        4.  , 58.  , 20.  , 39.  , 55.  , 31.  , 34.  , 15.  , 28.  ,
        8.  , 19.  , 40.  , 66.  , 42.  , 21.  , 18.  ,  3.  ,  7.  ,
       49.  , 29.  , 65.  , 28.5 ,  5.  , 11.  , 45.  , 17.  , 32.  ,
       16.  , 25.  ,  0.83, 30.  , 33.  , 23.  , 24.  , 46.  , 59.  ,
       71.  , 37.  , 47.  , 14.5 , 70.5 , 32.5 , 12.  ,  9.  , 36.5 ,
       51.  , 55.5 , 40.5 , 44.  ,  1.  , 61.  , 56.  , 50.  , 36.  ,
       45.5 , 20.5 , 62.  , 41.  , 52.  , 63.  , 23.5 ,  0.92, 43.  ,
       60.  , 10.  , 64.  , 13.  , 48.  ,  0.75, 53.  , 57.  , 80.  ,
       70.  , 24.5 ,  6.  ,  0.67, 30.5 ,  0.42, 34.5 , 74.  ])
```

In [145]:
```
data.groupby(['Age']).count()
```

Out[145]:

Age	PassengerId	Survived	Pclass	Name	Sex	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0.42	1	1	1	1	1	1	1	1	1	0	1
0.67	1	1	1	1	1	1	1	1	1	0	1
0.75	2	2	2	2	2	2	2	2	2	0	2
0.83	2	2	2	2	2	2	2	2	2	0	2
0.92	1	1	1	1	1	1	1	1	1	1	1
...	...	...	...	...	...	...	...	...	...	...	...
70.00	2	2	2	2	2	2	2	2	2	1	2
70.50	1	1	1	1	1	1	1	1	1	0	1
71.00	2	2	2	2	2	2	2	2	2	1	2
74.00	1	1	1	1	1	1	1	1	1	0	1
80.00	1	1	1	1	1	1	1	1	1	1	1

88 rows × 11 columns

In [146]:
```
data2['place']=data2['Pclass'].map({1:'F',2:'S',3:'third'})
```

In [147]:
```
data2.head(10)
```

Out[147]:

	Survived	Pclass	Sex	Age	Fare	Embarked	place
0	0	3	1	22.0	7.2500	S	third
1	1	1	0	38.0	71.2833	C	F
2	1	3	0	26.0	7.9250	S	third
3	1	1	0	35.0	53.1000	S	F
4	0	3	1	35.0	8.0500	S	third
5	0	3	1	28.0	8.4583	Q	third
6	0	1	1	54.0	51.8625	S	F
7	0	3	1	2.0	21.0750	S	third
8	1	3	0	27.0	11.1333	S	third
9	1	2	0	14.0	30.0708	C	S

In [148]:
```
data2=pd.get_dummies(data2)
data2.head()
```

Out[148]:

	Survived	Pclass	Sex	Age	Fare	Embarked_C	Embarked_Q	Embarked_S	place_F	place_S	place_third
0	0	3	1	22.0	7.2500	0	0	1	0	0	1
1	1	1	0	38.0	71.2833	1	0	0	1	0	0
2	1	3	0	26.0	7.9250	0	0	1	0	0	1
3	1	1	0	35.0	53.1000	0	0	1	1	0	0
4	0	3	1	35.0	8.0500	0	0	1	0	0	1

In [149]:
```
cor_mat=data2.corr()
cor_mat
```

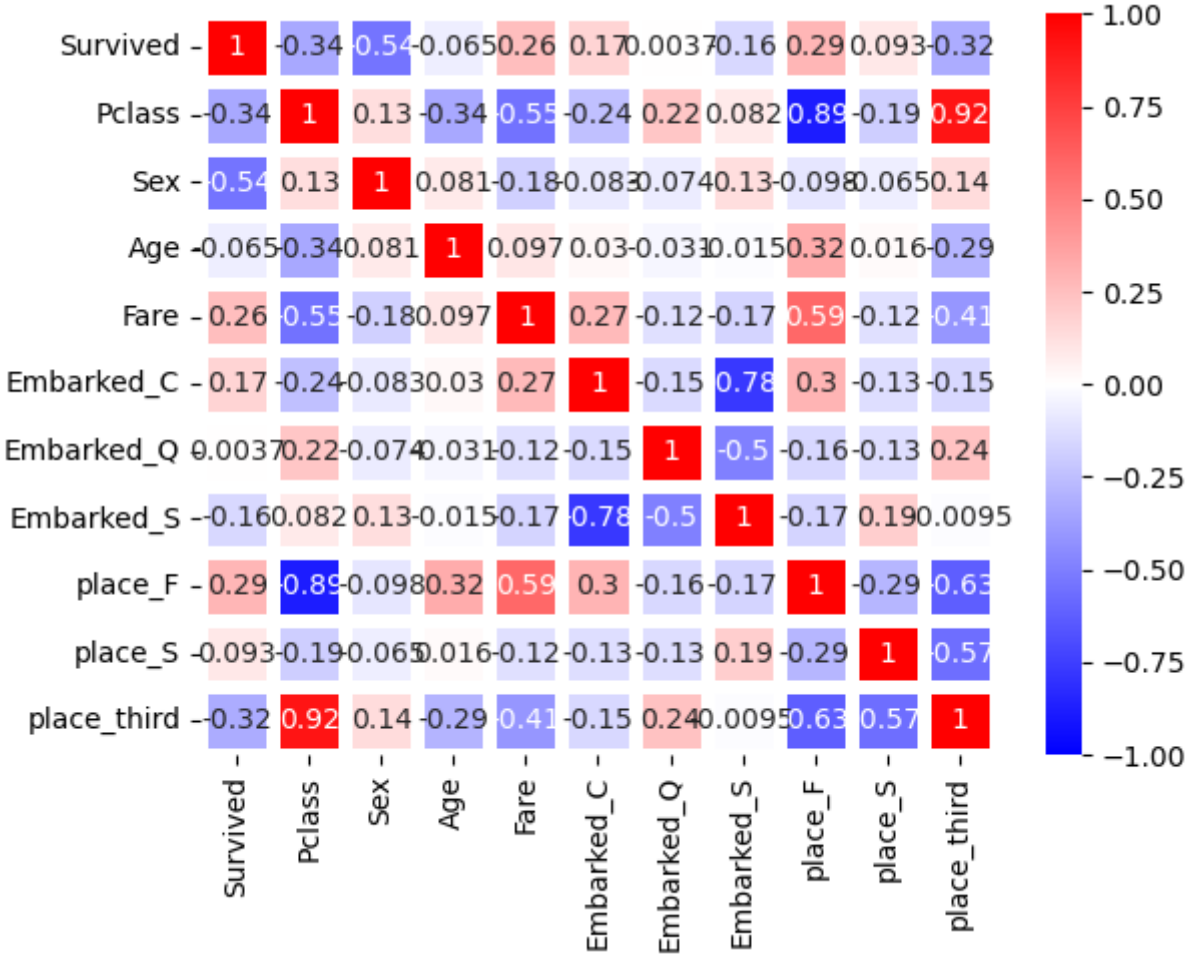
Out[149]:

	Survived	Pclass	Sex	Age	Fare	Embarked_C	Embarked_Q	Embarked_S	place_F	place_S	place_third
Survived	1.000000	-0.338481	-0.543351	-0.064910	0.257307	0.168240	0.003650	-0.155660	0.285904	0.093349	-0.322308
Pclass	-0.338481	1.000000	0.131900	-0.339898	-0.549500	-0.243292	0.221009	0.081720	-0.885924	-0.188432	0.916673
Sex	-0.543351	0.131900	1.000000	0.081163	-0.182333	-0.082853	-0.074115	0.125722	-0.098013	-0.064746	0.137143
Age	-0.064910	-0.339898	0.081163	1.000000	0.096688	0.030248	-0.031415	-0.014665	0.323896	0.015831	-0.291955
Fare	0.257307	-0.549500	-0.182333	0.096688	1.000000	0.269335	-0.117216	-0.166603	0.591711	-0.118557	-0.413333
Embarked_C	0.168240	-0.243292	-0.082853	0.030248	0.269335	1.000000	-0.148258	-0.778359	0.296423	-0.125416	-0.153329
Embarked_Q	0.003650	0.221009	-0.074115	-0.031415	-0.117216	-0.148258	1.000000	-0.496624	-0.155342	-0.127301	0.237449
Embarked_S	-0.155660	0.081720	0.125722	-0.014665	-0.166603	-0.778359	-0.496624	1.000000	-0.170379	0.192061	-0.009511
place_F	0.285904	-0.885924	-0.098013	0.323896	0.591711	0.296423	-0.155342	-0.170379	1.000000	-0.288585	-0.626738
place_S	0.093349	-0.188432	-0.064746	0.015831	-0.118557	-0.125416	-0.127301	0.192061	-0.288585	1.000000	-0.565210
place_third	-0.322308	0.916673	0.137143	-0.291955	-0.413333	-0.153329	0.237449	-0.009511	-0.626738	-0.565210	1.000000

In [150]:
```
import seaborn as sns
```

In [153]:
```
sns.heatmap(cor_mat,vmax=1,vmin=-1,annot=True,linewidth=5,cmap='bwr')
```

Out[153]: <Axes: >



In [154]:
```
data2.groupby('Survived').count()
```

Out[154]:

Survived	Pclass	Sex	Age	Fare	Embarked_C	Embarked_Q	Embarked_S	place_F	place_S	place_third
0	549	549	549	549	549	549	549	549	549	549
1	342	342	342	342	342	342	342	342	342	342

In [156]:
```
y=data1['Survived']
x=data2.drop('Survived',axis=1)
```

In [157]:
```
from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test=train_test_split(x,y,test_size=0.33,random_state=42)
```

In [160]:

```
/home/placement/anaconda3/lib/python3.10/site-packages/sklearn/linear_model/_logistic.py:458: Convergence
Warning: 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 (https://scikit-learn.org/stable/modules/p
reprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.o
rg/stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
```

Out[160]:
```
LogisticRegression()
```
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [161]:
```
y_pred=classifier.predict(X_test)
```

In [162]:
```
y_pred
```

Out[162]:
```
array([0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1,
       0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1,
       0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0,
       1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0,
       0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1,
       0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0,
       0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0,
       1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0,
       0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1,
       0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 1, 1, 0])
```

In [164]:
```
from sklearn.metrics import confusion_matrix
confusion_matrix(Y_test,y_pred)
```

Out[164]:
```
array([[154,  21],
       [ 37,  83]])
```

In [165]:
```
from sklearn.metrics import accuracy_score
accuracy_score(Y_test,y_pred)
```

Out[165]:
```
0.8033898305084746
```

In [166]:
```
y
```

Out[166]:
```
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 [167]:
```
154+83
```

Out[167]:
```
237
```

In [168]:
```
237/(237+21+37)
```

Out[168]:
```
0.8033898305084746
```

In [ ]:
```

```