In [1]:
```
import pandas as pd
import pickle
import warnings
warnings.filterwarnings("ignore")
```

In [2]:
```
data=pd.read_csv("C:\\Users\\reshma_koduri\\OneDrive\\Documents\\Titanic Dataset crt
data
```

Out[2]:

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

891 rows × 12 columns

In [3]:
```
data.describe()
```

Out[3]:

	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 [4]:
```
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 [5]:
```
data.shape
```

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

In [6]:
```
data.head(10)
```

Out[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley	female	38.0	1	0	PC 17599	71.2833	C85

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
				(Florence Briggs Th...							
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN

In [7]:
```
data.tail(10)
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7.8958	Na
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	Na
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000	Na

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500	Na
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	Na
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	Na
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B4
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	Na
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C14
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	Na

In [8]:
```
data.isnull().sum()
```

Out[8]:
```
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 [9]:
```
data.fillna(32,inplace=True)
```

In [10]:
```
data.isnull().sum()
```

Out[10]:
```
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
```

In [11]:
```
data['Age'].unique()
```

Out[11]:
```
array([22.  , 38.  , 26.  , 35.  , 32.  , 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.  , 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 [12]:
```
data['Pclass'].unique()
```

Out[12]:
```
array([3, 1, 2], dtype=int64)
```

In [13]:
```
data.groupby('Sex').count()
```

Out[13]:

Sex	PassengerId	Survived	Pclass	Name	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
female	314	314	314	314	314	314	314	314	314	314	314
male	577	577	577	577	577	577	577	577	577	577	577

In [14]:
```
data1=data.drop(['PassengerId','SibSp','Parch','Name','Ticket','Cabin'],axis=1)
data1
```

Out[14]:

	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	32.0	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 [15]:
```
data2=pd.get_dummies(data1,dtype=int)
data2
```

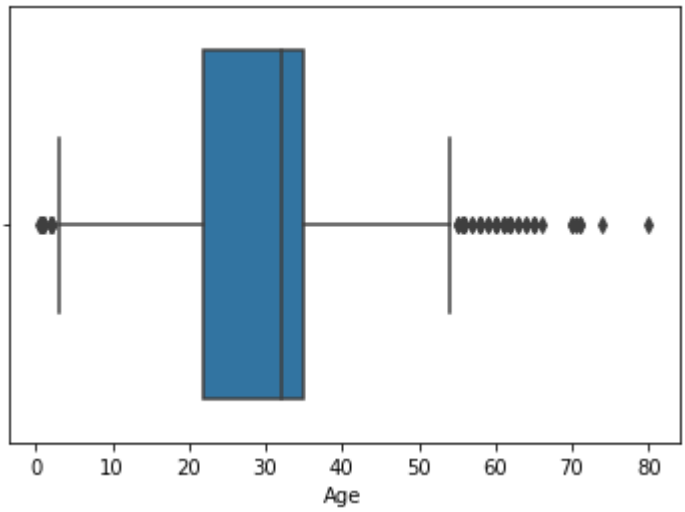
Out[15]:

	Survived	Pclass	Age	Fare	Sex_female	Sex_male	Embarked_32	Embarked_C	Embarked_Q
0	0	3	22.0	7.2500	0	1	0	0	0
1	1	1	38.0	71.2833	1	0	0	1	0
2	1	3	26.0	7.9250	1	0	0	0	0
3	1	1	35.0	53.1000	1	0	0	0	0
4	0	3	35.0	8.0500	0	1	0	0	0
...	...	...	...	...	...	...	...	...	...
886	0	2	27.0	13.0000	0	1	0	0	0
887	1	1	19.0	30.0000	1	0	0	0	0
888	0	3	32.0	23.4500	1	0	0	0	0
889	1	1	26.0	30.0000	0	1	0	1	0
890	0	3	32.0	7.7500	0	1	0	0	1

891 rows × 10 columns

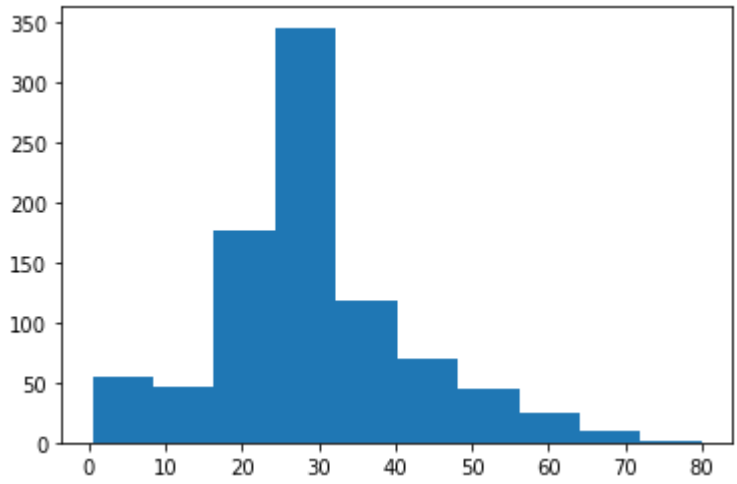
In [16]:
```
import seaborn as sb
import matplotlib.pyplot as plt
sb.boxplot(data1.Age)
```

Out[16]:
```
<AxesSubplot:xlabel='Age'>
```



In [17]:
```
import matplotlib.pyplot as plt
plt.hist(data1['Age'])
```

Out[17]:
```
(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 [18]:
```
data1.head(15)
```

Out[18]:

	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
5	0	3	male	32.0	8.4583	Q
6	0	1	male	54.0	51.8625	S
7	0	3	male	2.0	21.0750	S
8	1	3	female	27.0	11.1333	S
9	1	2	female	14.0	30.0708	C
10	1	3	female	4.0	16.7000	S
11	1	1	female	58.0	26.5500	S
12	0	3	male	20.0	8.0500	S
13	0	3	male	39.0	31.2750	S
14	0	3	female	14.0	7.8542	S

In [19]:
```
#data['Sex']=data['Sex'].map({'male':1,'female':0})
#data
```

In [20]:
```
data['Sex'].unique()
```

Out[20]:
```
array(['male', 'female'], dtype=object)
```

In [21]:
```
data1['Sex']=data1['Sex'].map({'male':1,'female':0})
data1
```

Out[21]:

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

891 rows × 6 columns

In [22]:
```
data.shape
```

Out[22]:
```
(891, 12)
```

In [23]:
```
data1.shape
```

Out[23]:
```
(891, 6)
```

In [24]:
```
data2.shape
```

Out[24]:
```
(891, 10)
```

In [25]:
```
data1.describe()
```

Out[25]:

	Survived	Pclass	Sex	Age	Fare
count	891.000000	891.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	0.647587	30.156195	32.204208
std	0.486592	0.836071	0.477990	13.034420	49.693429
min	0.000000	1.000000	0.000000	0.420000	0.000000
25%	0.000000	2.000000	0.000000	22.000000	7.910400
50%	0.000000	3.000000	1.000000	32.000000	14.454200
75%	1.000000	3.000000	1.000000	35.000000	31.000000
max	1.000000	3.000000	1.000000	80.000000	512.329200

In [26]:
```
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 6 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   Survived  891 non-null    int64  
 1   Pclass    891 non-null    int64  
 2   Sex       891 non-null    int64  
 3   Age       891 non-null    float64
 4   Fare      891 non-null    float64
 5   Embarked  891 non-null    object 
dtypes: float64(2), int64(3), object(1)
memory usage: 41.9+ KB
```

In [27]:
```
data1.groupby(['Age']).count()
```

Out[27]:

Age	Survived	Pclass	Sex	Fare	Embarked
0.42	1	1	1	1	1
0.67	1	1	1	1	1
0.75	2	2	2	2	2
0.83	2	2	2	2	2
0.92	1	1	1	1	1
...	...	...	...	...	...
70.00	2	2	2	2	2
70.50	1	1	1	1	1
71.00	2	2	2	2	2
74.00	1	1	1	1	1
80.00	1	1	1	1	1

88 rows × 5 columns

In [28]:
```
data1['Pclass']=data1['Pclass'].map({1:'f',2:'s',3:'t'})
```

In [29]:
```
data1.isna().sum()
```

Out[29]:
```
Survived    0
Pclass      0
Sex         0
Age         0
Fare        0
Embarked    0
dtype: int64
```

In [30]:
```
data1.fillna(data1['Pclass'],inplace=True)
```

In [31]:
```
data1.isna().sum()
```

Out[31]:
```
Survived    0
Pclass      0
Sex         0
Age         0
Fare        0
Embarked    0
dtype: int64
```

In [32]:
```
data1
```

Out[32]:

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

891 rows × 6 columns

In [33]:
```
data1.shape
```

Out[33]:
```
(891, 6)
```

In [34]:
```
cor_mat=data2.corr()
cor_mat
```

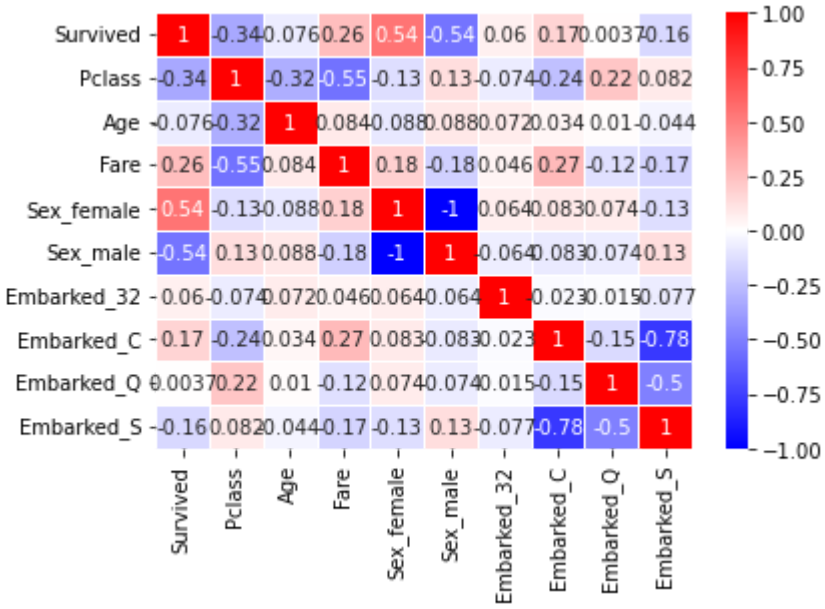
Out[34]:

	Survived	Pclass	Age	Fare	Sex_female	Sex_male	Embarked_32	Embar
Survived	1.000000	-0.338481	-0.076132	0.257307	0.543351	-0.543351	0.060095	0.1
Pclass	-0.338481	1.000000	-0.318328	-0.549500	-0.131900	0.131900	-0.074282	-0.2
Age	-0.076132	-0.318328	1.000000	0.084242	-0.087835	0.087835	0.072251	0.0
Fare	0.257307	-0.549500	0.084242	1.000000	0.182333	-0.182333	0.045646	0.2
Sex_female	0.543351	-0.131900	-0.087835	0.182333	1.000000	-1.000000	0.064296	0.0
Sex_male	-0.543351	0.131900	0.087835	-0.182333	-1.000000	1.000000	-0.064296	-0.0
Embarked_32	0.060095	-0.074282	0.072251	0.045646	0.064296	-0.064296	1.000000	-0.0
Embarked_C	0.168240	-0.243292	0.034289	0.269335	0.082853	-0.082853	-0.022864	1.0
Embarked_Q	0.003650	0.221009	0.009957	-0.117216	0.074115	-0.074115	-0.014588	-0.1

	Survived	Pclass	Age	Fare	Sex_female	Sex_male	Embarked_32	Embar
Embarked_S	-0.155660	0.081720	-0.043852	-0.166603	-0.125722	0.125722	-0.076588	-0.7

In [35]:
```
import seaborn as sb
sb.heatmap(cor_mat,vmax=1,vmin=-1,annot=True,linewidth=.5,cmap="bwr")
```

Out[35]:
```
<AxesSubplot:>
```



In [36]:
```
data2.groupby(['Survived']).count()
```

Out[36]:

Survived	Pclass	Age	Fare	Sex_female	Sex_male	Embarked_32	Embarked_C	Embarked_Q	Embark
0	549	549	549	549	549	549	549	549	
1	342	342	342	342	342	342	342	342	

In [37]:
```
y=data2['Survived']
x=data2.drop(['Survived'],axis=1)
```

In [38]:
```
y
```

Out[38]:
```
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 [39]:
```
x
```

Out[39]:

	Pclass	Age	Fare	Sex_female	Sex_male	Embarked_32	Embarked_C	Embarked_Q	Embarke
0	3	22.0	7.2500	0	1	0	0	0	
1	1	38.0	71.2833	1	0	0	1	0	
2	3	26.0	7.9250	1	0	0	0	0	
3	1	35.0	53.1000	1	0	0	0	0	
4	3	35.0	8.0500	0	1	0	0	0	
...	...	...	...	...	...	...	...	...	
886	2	27.0	13.0000	0	1	0	0	0	
887	1	19.0	30.0000	1	0	0	0	0	
888	3	32.0	23.4500	1	0	0	0	0	
889	1	26.0	30.0000	0	1	0	1	0	
890	3	32.0	7.7500	0	1	0	0	1	

891 rows × 9 columns

In [40]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30,random_state=66)
```

In [41]:
```
x_train
```

Out[41]:

	Pclass	Age	Fare	Sex_female	Sex_male	Embarked_32	Embarked_C	Embarked_Q	Embark
594	2	37.0	26.0000	0	1	0	0	0	
624	3	21.0	16.1000	0	1	0	0	0	
341	1	24.0	263.0000	1	0	0	0	0	
792	3	32.0	69.5500	1	0	0	0	0	
767	3	30.5	7.7500	1	0	0	0	1	
...	...	...	...	...	...	...	...	...	
122	2	32.5	30.0708	0	1	0	1	0	
51	3	21.0	7.8000	0	1	0	0	0	
631	3	51.0	7.0542	0	1	0	0	0	
828	3	32.0	7.7500	0	1	0	0	1	
20	2	35.0	26.0000	0	1	0	0	0	

623 rows × 9 columns

In [42]:
```
y_train
```

Out[42]:
```
594    0
624    0
341    1
792    0
767    0
      ..
122    0
51     0
631    0
828    1
20     0
Name: Survived, Length: 623, dtype: int64
```

In [43]:
```
x_test
```

Out[43]:

	Pclass	Age	Fare	Sex_female	Sex_male	Embarked_32	Embarked_C	Embarked_Q	Embarke
28	3	32.0	7.8792	1	0	0	0	1	
652	3	21.0	8.4333	0	1	0	0	0	
141	3	22.0	7.7500	1	0	0	0	0	
669	1	32.0	52.0000	1	0	0	0	0	
200	3	28.0	9.5000	0	1	0	0	0	
...	...	...	...	...	...	...	...	...	
529	2	23.0	11.5000	0	1	0	0	0	
849	1	32.0	89.1042	1	0	0	1	0	
463	2	48.0	13.0000	0	1	0	0	0	
515	1	47.0	34.0208	0	1	0	0	0	
430	1	28.0	26.5500	0	1	0	0	0	

268 rows × 9 columns

In [44]:
```
from sklearn.linear_model import LogisticRegression
classifier=LogisticRegression()
classifier.fit(x_train, y_train)
```

Out[44]:
```
LogisticRegression()
```

In [45]:
```
ypred=classifier.predict(x_test)
ypred
```

Out[45]:
```
array([1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1,
       0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0,
       1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0,
       1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1,
       0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1,
       0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0,
       0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1,
       1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0,
```

```
       0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0,
       1, 0, 0, 0], dtype=int64)
```

In [46]:
```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,ypred)
```

Out[46]:
```
array([[146,  29],
       [ 28,  65]], dtype=int64)
```

In [47]:
```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,ypred)
```

Out[47]:
```
0.7873134328358209
```

In [ ]:

In [ ]: