In [46]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import mean_squared_error
from sklearn.metrics import confusion_matrix
```

In [2]:
```
train = pd.read_csv('Downloads/titanic_train.csv')
```

In [3]:
```
train
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Ca
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	N
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	N
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	N
...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	N
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	N
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	N

891 rows × 12 columns

In [4]:
```
train.isnull()
```

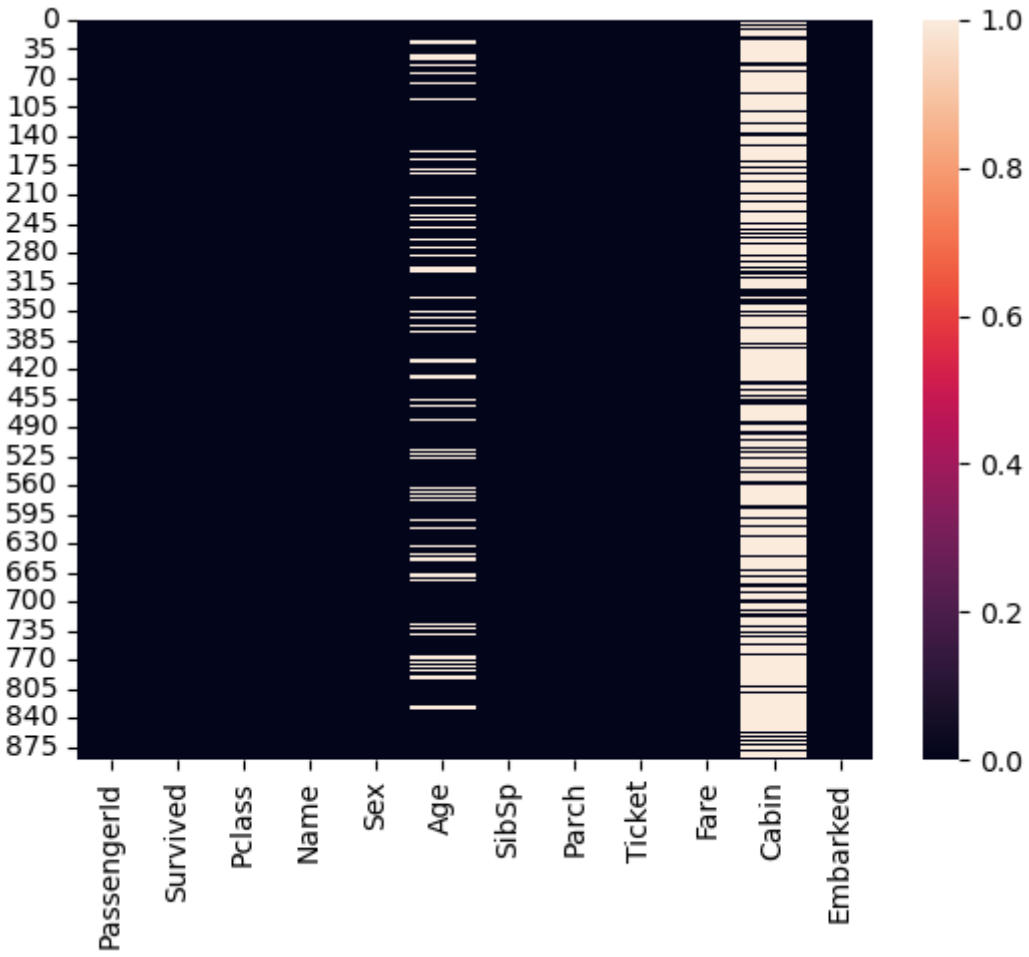
Out[4]:

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

891 rows × 12 columns

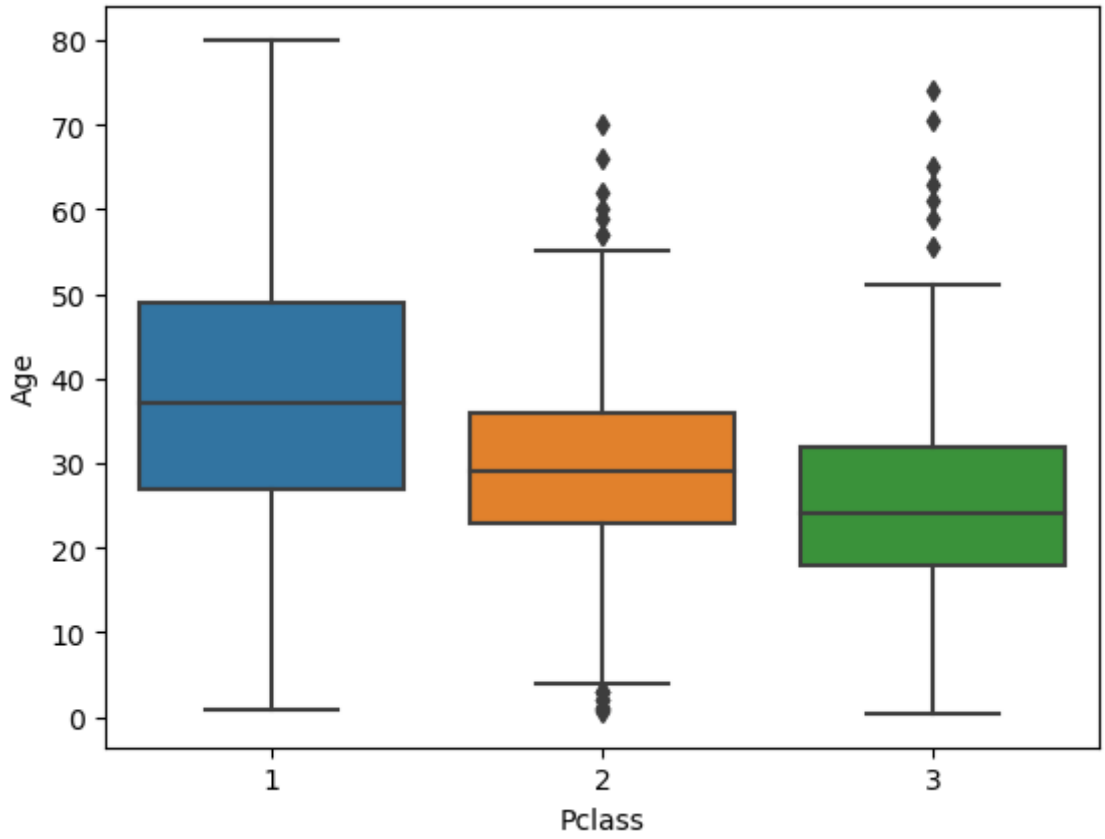
In [5]:
```
sns.heatmap(train.isnull())
```

Out[5]:
```
<Axes: >
```



In [7]:
```
sns.boxplot(x = train['Pclass'],y = train['Age'])
```

Out[7]:
```
<Axes: xlabel='Pclass', ylabel='Age'>
```

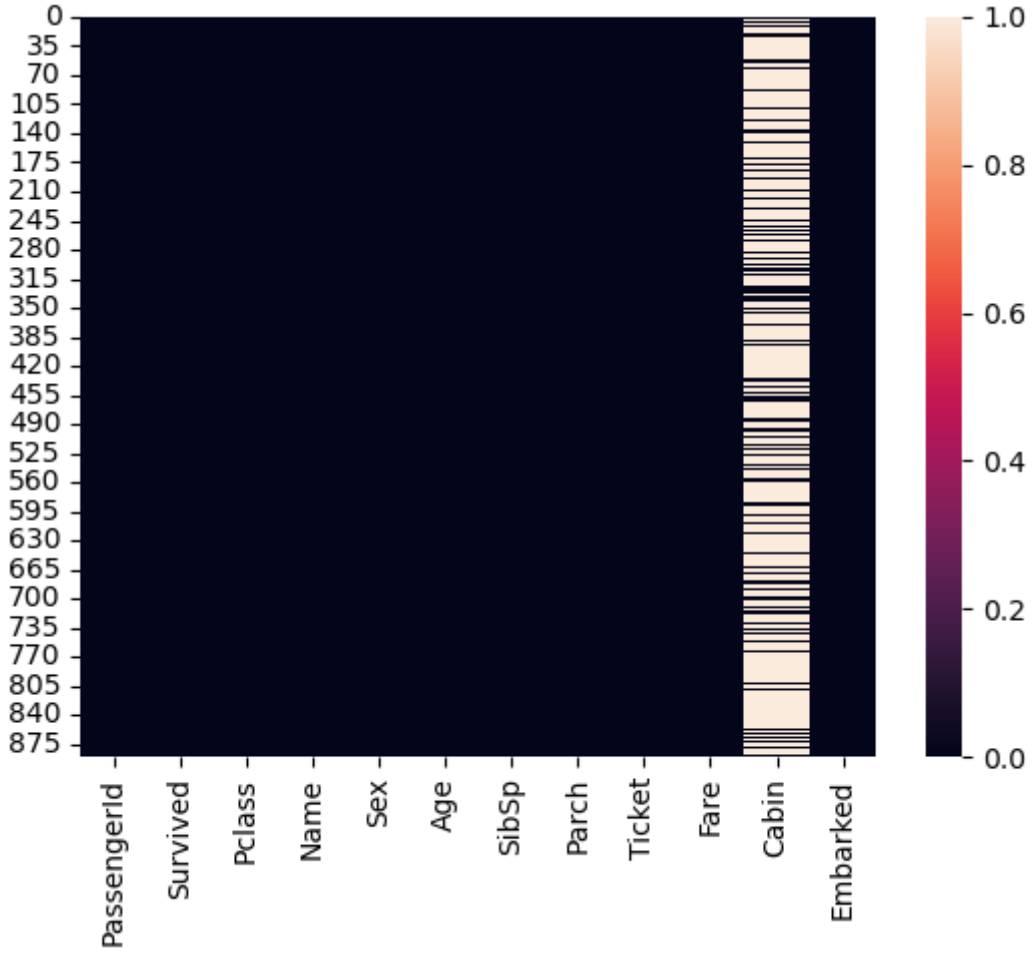


In [12]:
```
pclass=train['Pclass']
age=train['Age']
```

In [14]:
```
for i in range(len(age)):
    if pclass[i]==3 and pd.isnull(age[i])==True:
        age[i]=25
    elif pclass[i]==2 and pd.isnull(age[i])==True:
        age[i]=29
    elif pclass[i]==1 and pd.isnull(age[i])==True:
        age[i]=39
```

In [15]:
```
sns.heatmap(train.isnull())
```

Out[15]:
```
<Axes: >
```



In [17]:
```
sx=train['Sex']

for i in range(len(sx)):
    if sx[i]=='male':
        sx[i]=1
    elif sx[i]=='female':
        sx[i]=0
```

```
C:\Users\Sony Vaio\AppData\Local\Temp\ipykernel_7584\1156211743.py:5: SettingWithC
opyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stabl
e/user_guide/indexing.html#returning-a-view-versus-a-copy
  sx[i]=1
C:\Users\Sony Vaio\AppData\Local\Temp\ipykernel_7584\1156211743.py:7: SettingWithC
opyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stabl
e/user_guide/indexing.html#returning-a-view-versus-a-copy
  sx[i]=0
```

In [18]:
```
train
```

Out[18]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	1	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	0	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	0	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	1	35.0	0	0	373450	8.0500	NaN
...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	1	27.0	0	0	211536	13.0000	NaN
887	888	1	1	Graham, Miss. Margaret Edith	0	19.0	0	0	112053	30.0000	B42
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	0	25.0	1	2	W./C. 6607	23.4500	NaN
889	890	1	1	Behr, Mr. Karl Howell	1	26.0	0	0	111369	30.0000	C148
890	891	0	3	Dooley, Mr. Patrick	1	32.0	0	0	370376	7.7500	NaN

891 rows × 12 columns

In [37]:
```
x=train[['Pclass','Sex','Age']]
y=train['Survived']
```

In [34]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
```

In [35]:
```
dtree=LogisticRegression()
dtree.fit(x_train,y_train)
```

```
C:\Users\Sony Vaio\anaconda3\lib\site-packages\sklearn\utils\validation.py:1143: D
ataConversionWarning: A column-vector y was passed when a 1d array was expected. P
lease change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Out[35]:

▾ LogisticRegression

LogisticRegression()

In [40]:
```
pred=dtree.predict(x_test)
```

In [38]:
```
pred
```

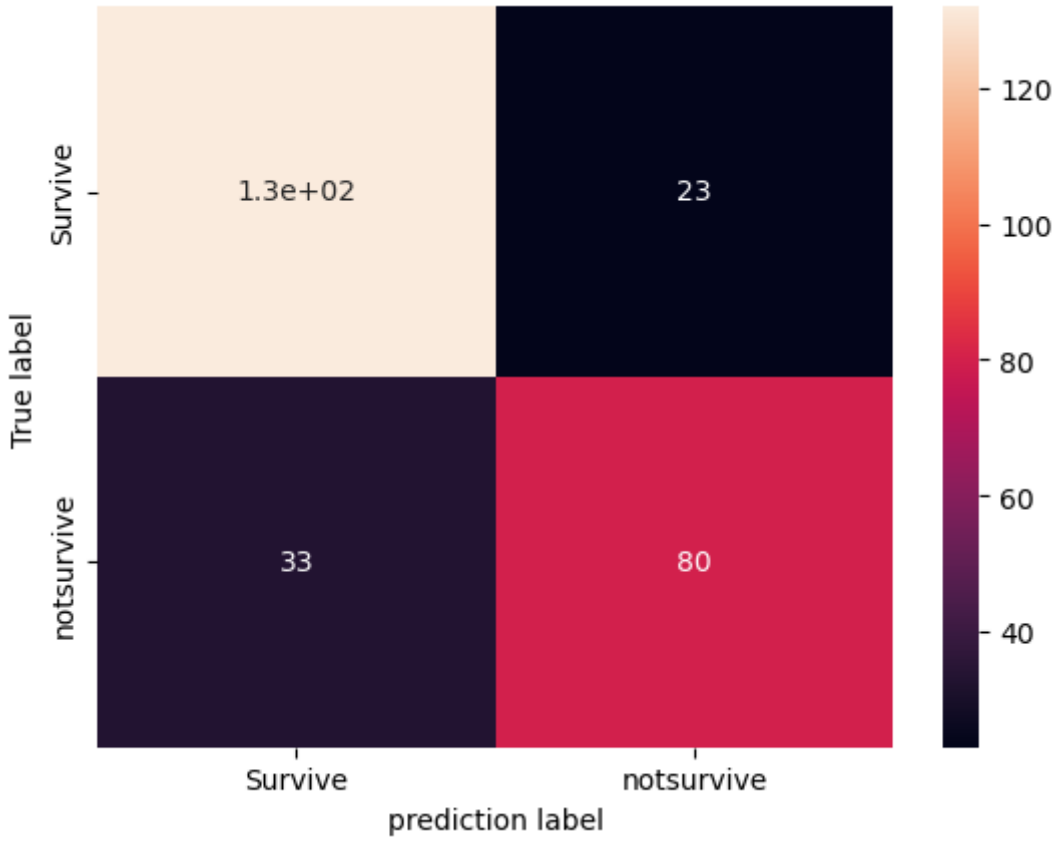
Out[38]:
```
array([0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0,
       0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0,
       1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0,
       0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 1,
       0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0,
       1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1,
       0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0,
       0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1,
       1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0,
       0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1,
       1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0,
       0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0,
       1, 1, 0, 0], dtype=int64)
```

In [47]:
```
confusion_matrix(y_test,pred)
```

Out[47]:
```
array([[132,  23],
       [ 33,  80]], dtype=int64)
```

In [50]:
```
plt.figure()
sns.heatmap(confusion_matrix(y_test,pred),xticklabels=['Survive','notsurvive'],yti
plt.xlabel('prediction label')
plt.ylabel('True label')
```

Out[50]:
```
Text(50.72222222222221, 0.5, 'True label')
```



In [52]:
```
212*100/268
```

Out[52]:
```
79.1044776119403
```

In [4]:
```
((132+80)*100/(23+33+132+80))
```

Out[4]:
```
79.1044776119403
```

In [ ]: