In [ ]:

In [2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [ ]:

In [3]:
```
df = pd.read_csv("titanic.csv")
```

In [4]:
```
df.shape
```

Out[4]:
```
(712, 7)
```

In [5]:
```
df.head()
```

Out[5]:

	pclass	gender	age	sibling	fare	embark_town	survived
0	3	male	22.0	1	7.2500	Southampton	no
1	1	female	38.0	1	71.2833	Cherbourg	yes
2	3	female	26.0	0	7.9250	Southampton	yes
3	1	female	35.0	1	53.1000	Southampton	yes
4	3	male	35.0	0	8.0500	Southampton	no

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 712 entries, 0 to 711
Data columns (total 7 columns):
pclass         712 non-null int64
gender         712 non-null object
age            712 non-null float64
sibling        712 non-null int64
fare           712 non-null float64
embark_town    712 non-null object
survived       712 non-null object
dtypes: float64(2), int64(2), object(3)
memory usage: 39.0+ KB
```

In [7]:
```
df.describe()
```

Out[7]:

	pclass	age	sibling	fare
count	712.000000	712.000000	712.000000	712.000000
mean	2.240169	29.642093	0.514045	34.567251
std	0.836854	14.492933	0.930692	52.938648
min	1.000000	0.420000	0.000000	0.000000
25%	1.000000	20.000000	0.000000	8.050000
50%	2.000000	28.000000	0.000000	15.645850
75%	3.000000	38.000000	1.000000	33.000000
max	3.000000	80.000000	5.000000	512.329200

In [ ]:

In [10]:
```
gender = pd.get_dummies(df['gender'], drop_first=True)
gender.head()
```

Out[10]:

	male
0	1
1	0
2	0
3	0
4	1

In [11]:
```
city = pd.get_dummies(df['embark_town'], drop_first=True)
city.head()
```

Out[11]:

	Queenstown	Southampton
0	0	1
1	0	0
2	0	1
3	0	1
4	0	1

In [12]:
```
target = pd.get_dummies(df['survived'], drop_first=True)
target.head()
```

Out[12]:

	yes
0	0
1	1
2	1
3	1
4	0

In [ ]:
```

```

In [13]:
```
df.drop(["gender", "embark_town", "survived"], axis = 1, inplace = True)
```

In [14]:
```
df.head()
```

Out[14]:

	pclass	age	sibling	fare
0	3	22.0	1	7.2500
1	1	38.0	1	71.2833
2	3	26.0	0	7.9250
3	1	35.0	1	53.1000
4	3	35.0	0	8.0500

In [15]:
```
df_new = pd.concat([df, gender, city, target], axis = 1)
```

In [17]:

```
df_new.head()
```

Out[17]:

	pclass	age	sibling	fare	male	Queenstown	Southampton	yes
0	3	22.0	1	7.2500	1	0	1	0
1	1	38.0	1	71.2833	0	0	0	1
2	3	26.0	0	7.9250	0	0	1	1
3	1	35.0	1	53.1000	0	0	1	1
4	3	35.0	0	8.0500	1	0	1	0

In [ ]:

In [ ]:

In [19]:

```
# input
#x = df_new.drop("yes", axis = 1, inplace = True)
x = df_new.drop("yes", axis = 1)

# output
y = df_new['yes']
```

In [ ]:

In [ ]:

In [21]:

```
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics
```

In [ ]:

In [22]:

```
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size = 0.3, random_st
ate = 100)
```

In [ ]:

In [84]:

```
model = LogisticRegression()
model.fit(xtrain, ytrain)
y_predict = model.predict(xtest)
```

```
C:\Users\Nitish\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:4
32: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Specify
a solver to silence this warning.
  FutureWarning)
```

In [ ]:

In [89]:

```
print(metrics.confusion_matrix(ytest, y_predict))
```

```
[[111  13]
 [ 27  63]]
```

In [88]:
```
print(metrics.classification_report(ytest, y_predict))
```

```
              precision    recall  f1-score   support

           0       0.80      0.90      0.85       124
           1       0.83      0.70      0.76        90

    accuracy                           0.81       214
   macro avg       0.82      0.80      0.80       214
weighted avg       0.81      0.81      0.81       214
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [57]:
```
import random
```

In [68]:
```
random.random()
```

Out[68]:
```
0.15838287025480557
```

In [82]:
```
random.seed(2)
random.random()
```

Out[82]:
```
0.9560342718892494
```

In [ ]:

In [ ]:

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```