In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
df = pd.read_csv("titanic.csv", sep='\t').dropna()
```

In [3]:
```
df.head()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S

In [4]:
```
y = df['Survived']
```

In [5]:
```
x = df.drop(['PassengerId','Survived','Name','Ticket','Cabin','Embarked'], axis=1)
```

In [6]:
```
x.head()
```

Out[6]:

	Pclass	Sex	Age	SibSp	Parch	Fare
1	1	female	38.0	1	0	71.2833
3	1	female	35.0	1	0	53.1000
6	1	male	54.0	0	0	51.8625
10	3	female	4.0	1	1	16.7000
11	1	female	58.0	0	0	26.5500

In [7]:
```
y.head()
```

Out[7]:
```
1     1
3     1
6     0
10    1
11    1
Name: Survived, dtype: int64
```

In [8]:
```
from sklearn.preprocessing import LabelEncoder
```

In [9]:
```
le = LabelEncoder()
```

In [10]:
```
le.fit(x['Sex'])
```

Out[10]:
```
LabelEncoder()
```

In [11]:
```
le.classes_
```

Out[11]:
```
array(['female', 'male'], dtype=object)
```

In [12]:
```
S = le.transform(x['Sex'])
```

In [13]:
```
x['Sex'] = S
```

In [14]:
```
x
```

Out[14]:

	Pclass	Sex	Age	SibSp	Parch	Fare
1	1	0	38.0	1	0	71.2833
3	1	0	35.0	1	0	53.1000
6	1	1	54.0	0	0	51.8625
10	3	0	4.0	1	1	16.7000
11	1	0	58.0	0	0	26.5500
21	2	1	34.0	0	0	13.0000
23	1	1	28.0	0	0	35.5000
27	1	1	19.0	3	2	263.0000
52	1	0	49.0	1	0	76.7292
54	1	1	65.0	0	1	61.9792
62	1	1	45.0	1	0	83.4750
66	2	0	29.0	0	0	10.5000
75	3	1	25.0	0	0	7.6500
88	1	0	23.0	3	2	263.0000
92	1	1	46.0	1	0	61.1750
96	1	1	71.0	0	0	34.6542
97	1	1	23.0	0	1	63.3583
102	1	1	21.0	0	1	77.2875
110	1	1	47.0	0	0	52.0000
118	1	1	24.0	0	1	247.5208
123	2	0	32.5	0	0	13.0000
124	1	1	54.0	0	1	77.2875
136	1	0	19.0	0	2	26.2833
137	1	1	37.0	1	0	53.1000
139	1	1	24.0	0	0	79.2000
148	2	1	36.5	0	2	26.0000
151	1	0	22.0	1	0	66.6000

In [15]:
```
x.head()
```

Out[15]:

	Pclass	Sex	Age	SibSp	Parch	Fare
1	1	0	38.0	1	0	71.2833
3	1	0	35.0	1	0	53.1000
6	1	1	54.0	0	0	51.8625
10	3	0	4.0	1	1	16.7000
11	1	0	58.0	0	0	26.5500

In [16]:
```
from sklearn.model_selection import train_test_split
```

In [17]:
```
xTrain, xTest, yTrain, yTest = train_test_split(x, y, test_size = 0.2, random_state = 5)
```

In [18]:
```
from sklearn.linear_model import LinearRegression
```

In [19]:
```
ln = LinearRegression()
ln.fit(xTrain,yTrain)
```

Out[19]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)
```

In [20]:
```
ln.score(xTest,yTest)
```

Out[20]:
```
0.8175051234403508
```

In [21]:
```
#Shift + Double Tab for details
```