

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
dataset = pd.read_csv(r"C:\Users\kotha\Downloads\insurance.csv")
```

In [3]:

```
dataset
```

Out[3]:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...	...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns

In [4]:

```
dataset.isnull().any()
```

Out[4]:

```
age      False
sex      False
bmi      False
children False
smoker   False
region   False
charges  False
dtype: bool
```

In [5]:

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
dataset["sex"] = le.fit_transform(dataset["sex"])
dataset["smoker"] = le.fit_transform(dataset["smoker"])
dataset["region"] = le.fit_transform(dataset["region"])
```

In [6]:

```
dataset.head(1)
```

Out[6]:

	age	sex	bmi	children	smoker	region	charges
0	19	0	27.9	0	1	3	16884.924

In [7]:

```
x = dataset.iloc[:,0:6].values
y = dataset.iloc[:,6:7].values
```

In [8]:

```
from sklearn.preprocessing import OneHotEncoder
one = OneHotEncoder()
z = one.fit_transform(x[:,5:6]).toarray()
x = np.delete(x,5,axis=1)
x = np.concatenate((z,x),axis=1)
```

In [9]:

```
x.shape
```

Out[9]:

(1338, 9)

In [10]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.2,random_state=0)
```

In [11]:

```
from sklearn.linear_model import LinearRegression
mlr = LinearRegression()
mlr.fit(x_train,y_train)
```

Out[11]:

LinearRegression(copy\_X=True, fit\_intercept=True, n\_jobs=None, normalize=False)

In [12]:

```
x_test
```

Out[12]:

```
array([[ 0.    ,  0.    ,  0.    , ..., 30.2   ,  1.    ,  0.    ],
       [ 0.    ,  0.    ,  1.    , ..., 29.37  ,  1.    ,  0.    ],
       [ 0.    ,  1.    ,  0.    , ..., 40.565 ,  2.    ,  1.    ],
       ...,
       [ 1.    ,  0.    ,  0.    , ..., 40.28  ,  0.    ,  0.    ],
       [ 0.    ,  0.    ,  1.    , ..., 39.05  ,  3.    ,  1.    ],
       [ 1.    ,  0.    ,  0.    , ..., 24.795 ,  3.    ,  0.    ]])
```

In [13]:

```
y_test
```

Out[13]:

```
array([[ 9724.53   ],
       [ 8547.6913 ],
       [45702.02235 ],
       [12950.0712  ],
       [ 9644.2525  ],
       [ 4500.33925 ],
       [ 2198.18985 ],
       [11436.73815 ],
       [ 7537.1639  ],
       [ 5425.02335 ],
       [ 6753.038   ],
       [10493.9458  ],
       [ 7337.748   ],
       [ 4185.0979  ],
       [18310.742   ],
       [10702.6424  ],
       [12523.6048  ],
       [ 3490.5491  ]])
```

In [14]:

```
y_pred = mlr.predict(x_test)
```

In [15]:

```
y_pred
```

Out[15]:

```
array([[1.11699271e+04],
       [9.48670909e+03],
       [3.81811231e+04],
       [1.62663133e+04],
       [6.91464801e+03],
       [3.96348488e+03],
       [1.57939691e+03],
       [1.43852566e+04],
       [9.01257970e+03],
       [7.50846068e+03],
       [4.49176728e+03],
       [1.02795839e+04],
       [8.80129751e+03],
       [3.79802013e+03],
       [2.79262010e+04],
       [1.07151158e+04],
       [1.12889756e+04],
       [6.10501768e+03]])
```

In [16]:

```
from sklearn.metrics import r2_score
accuracy = r2_score(y_test,y_pred)
```

In [17]:

```
accuracy
```

Out[17]:

```
0.7999876970680434
```

In [18]:

```
yp = mlr.predict([[0,0,0,1,20.9,89.9,77.8,90.8,88.5]])
```

In [19]:

```
yp
```

Out[19]:

```
array([[2146175.50313513]])
```

In [20]:

```
x_train
```

Out[20]:

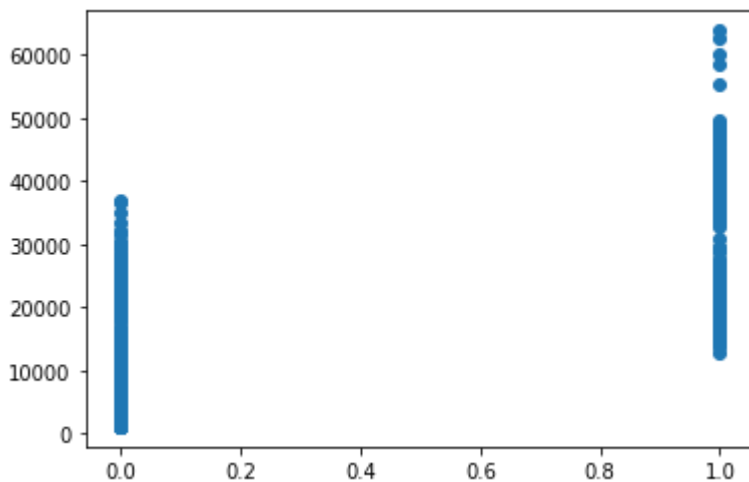
```
array([[ 0. ,  0. ,  0. , ..., 34.1 ,  4. ,  1. ],
       [ 0. ,  0. ,  1. , ..., 34.43,  0. ,  0. ],
       [ 1. ,  0. ,  0. , ..., 36.67,  2. ,  1. ],
       ...,
       [ 0. ,  0. ,  1. , ..., 25.08,  0. ,  0. ],
       [ 0. ,  1. ,  0. , ..., 35.53,  0. ,  0. ],
       [ 0. ,  0. ,  0. , ..., 18.5 ,  1. ,  0. ]])
```

In [21]:

```
import matplotlib.pyplot as plt
plt.scatter(x_train[:,8],y_train)
```

Out[21]:

<matplotlib.collections.PathCollection at 0x15bbbf4148>



In [ ]: