

Assignment -1

Medical insurance costs

Import the libraries

```
In [36]: import warnings
warnings.simplefilter('ignore')
```

import numpy and panda

```
In [37]: import pandas as pd
import numpy as np

In [38]: import matplotlib.pyplot as plt
%matplotlib inline
```

import the dataset

```
In [6]: read_file = pd.read_csv (r'insurance.csv')
read_file.to_csv (r'insurance.csv', index=None)

In [52]: dataset = pd.read_csv('insurance.csv')

In [53]: dataset

Out[53]:
```

| | 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 [54]: dataset = dataset.drop(['sex','smoker','region','children','bmi'],axis=1)
```

```
In [55]: dataset

Out[55]:
```

| | age | charges |
|------|-----|-------------|
| 0 | 19 | 16884.92400 |
| 1 | 18 | 1725.55230 |
| 2 | 28 | 4449.46200 |
| 3 | 33 | 21984.47061 |
| 4 | 32 | 3866.85520 |
| ... | ... | ... |
| 1333 | 50 | 10600.54830 |
| 1334 | 18 | 2205.98080 |
| 1335 | 18 | 1629.83350 |
| 1336 | 21 | 2007.94500 |
| 1337 | 61 | 29141.36030 |

1338 rows × 2 columns

reshaping of data

```
In [56]: x=dataset.iloc[:,0].values.reshape(-1,1)

In [57]: x.shape

Out[57]: (1338, 1)

In [58]: x

Out[58]: array([[19],
 [18],
 [28],
 ...,
 [18],
 [21],
 [61]], dtype=int64)

In [59]: y=dataset.iloc[:,-1].values.reshape(-1,1)

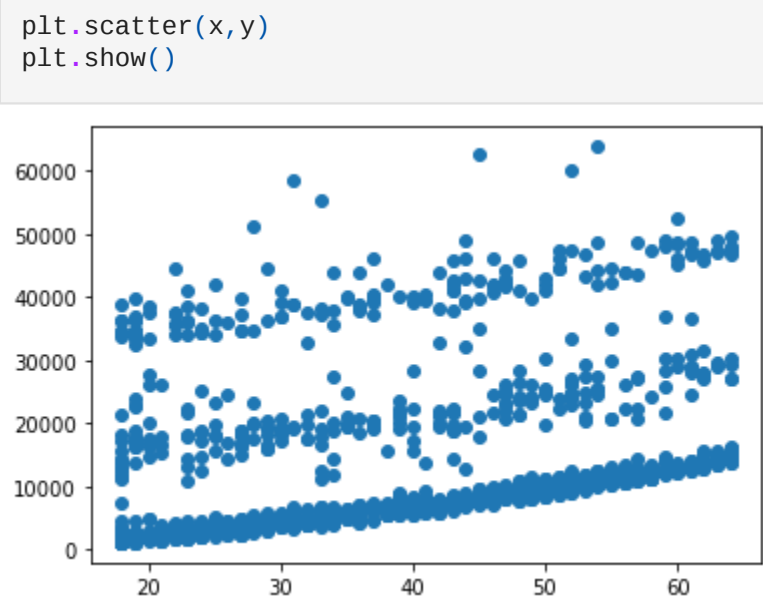
In [60]: y.shape

Out[60]: (1338, 1)

In [61]: y

Out[61]: array([[16884.924 ],
 [ 1725.5523],
 [ 4449.462 ],
 ...,
 [ 1629.8335],
 [ 2007.945 ],
 [29141.3603]])
```

Scatter plot



Dividing the dataset into training and testing set

```
In [63]: from sklearn.model_selection import train_test_split

In [64]: x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state =0)

In [65]: x_train.shape

Out[65]: (1070, 1)

In [66]: x_test.shape

Out[66]: (268, 1)

In [67]: y_train.shape

Out[67]: (1070, 1)

In [68]: y_test.shape

Out[68]: (268, 1)
```

import the linear regression function

```
In [69]: from sklearn.linear_model import LinearRegression

In [70]: lm=LinearRegression()
```

Train the model - perform linear regression

```
In [71]: lm.fit(x_train,y_train)

Out[71]: LinearRegression()

In [72]: y_pred=lm.predict(x_test)
```

visulize a regression line

