Simple Linear Regression

Importing Necessary Packages

In [22]:
```
import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import accuracy_score
```

Reading Data

In [23]:
```
data = pd.read_csv ('insurance.csv')
```

In [24]:
```
data.head()
```

Out[24]:

	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

Exploratory Data Analysis

In [25]:
```
data.describe()
```

Out[25]:

	age	bmi	children	charges
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.663397	1.094918	13270.422265
std	14.049960	6.098187	1.205493	12110.011237
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000
75%	51.000000	34.693750	2.000000	16639.912515
max	64.000000	53.130000	5.000000	63770.428010

In [26]:
```
data.info()
```

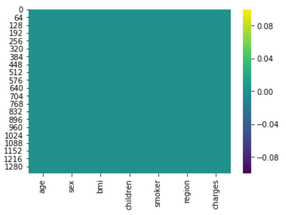
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
age         1338 non-null int64
sex         1338 non-null object
bmi         1338 non-null float64
children    1338 non-null int64
smoker      1338 non-null object
region      1338 non-null object
charges     1338 non-null float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.2+ KB
```

In [27]:
```
data.isnull().sum()
```

Out[27]:
```
age         0
sex         0
bmi         0
children    0
smoker      0
region      0
charges     0
dtype: int64
```

In [28]:
```
sns.heatmap(data.isnull(), cmap = 'viridis')
```

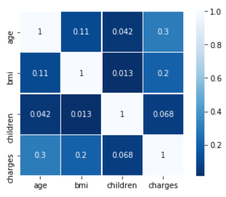
Out[28]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e4465937f0>
```



Finding Co-relations between Feature Columns

In [80]:
```
corr = data.corr()
sns.heatmap (corr , annot = True, square = True , linewidths = .5 ,cmap = 'Blues_r')
```

Out[80]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e44bda1630>
```



Selecting Dependent and Independent Variables

In [66]:
```
x = data.iloc[:,0].values
x = x.reshape(-1,1)
x.shape
```

Out[66]:
```
(1338, 1)
```

In [64]:
```
y = data.iloc[:,6].values
```

Out[64]:
```
(1338,)
```

Building Training and Testing Data

In [67]:
```
x_train,x_test,y_train,y_test = train_test_split (x,y,test_size = 0.3, random_state = 40 )
```

In [68]:
```
x_train.shape
```

Out[68]:
```
(936, 1)
```

In [69]:
```
x_test.shape
```

Out[69]:
```
(402, 1)
```

Implementing Linear Regression Model

slr = simple linear Regression

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

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

Out[71]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

In [72]:
```
y_test.shape
```

Out[72]:
```
(402,)
```

In [73]:
```
y_predict = slr.predict(x_test)
```

In [74]:
```
y_predict.shape
```

Out[74]:
```
(402,)
```

In [75]:
```
plt.scatter(x_train , y_train )
plt.plot (x_train ,slr.predict(x_train),)
```

Out[75]:
```
[<matplotlib.lines.Line2D at 0x1e44b654940>]
```