

Deena 20104016

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as pp
import seaborn as sns
```

Problem Statement

LINEAR REGRESSION

```
In [2]: a = pd.read_csv("Fitness.csv")
```

Out[2]:

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	A	5.62%	7.73%	6.16%	75
1	B	4.21%	17.27%	19.21%	160
2	C	9.83%	11.60%	5.17%	101
3	D	2.81%	21.91%	7.88%	127
4	E	25.28%	10.57%	11.82%	179
5	F	8.15%	16.24%	18.47%	167
6	G	18.54%	8.76%	17.49%	171
7	H	25.56%	5.93%	13.79%	170
8	Grand Total	100.00%	100.00%	100.00%	1150

HEAD

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

Out[3]:

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	A	5.62%	7.73%	6.16%	75
1	B	4.21%	17.27%	19.21%	160
2	C	9.83%	11.60%	5.17%	101
3	D	2.81%	21.91%	7.88%	127
4	E	25.28%	10.57%	11.82%	179

Data Cleaning and Preprocessing

```
In [4]: df.head()
```

```
Out[4]:
```

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	A	5.62%	7.73%	6.16%	75
1	B	4.21%	17.27%	19.21%	160
2	C	9.83%	11.60%	5.17%	101
3	D	2.81%	21.91%	7.88%	127
4	E	25.28%	10.57%	11.82%	179

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

```
Out[5]:
```

	Sum of Total Sales
count	9.000000
mean	255.555556
std	337.332963
min	75.000000
25%	127.000000
50%	167.000000
75%	171.000000
max	1150.000000

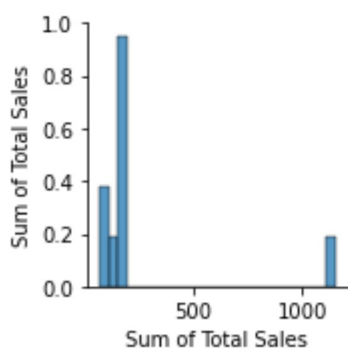
To display heading

```
In [6]: df.columns
```

```
Out[6]: Index(['Row Labels', 'Sum of Jan', 'Sum of Feb', 'Sum of Mar',  
              'Sum of Total Sales'],  
              dtype='object')
```

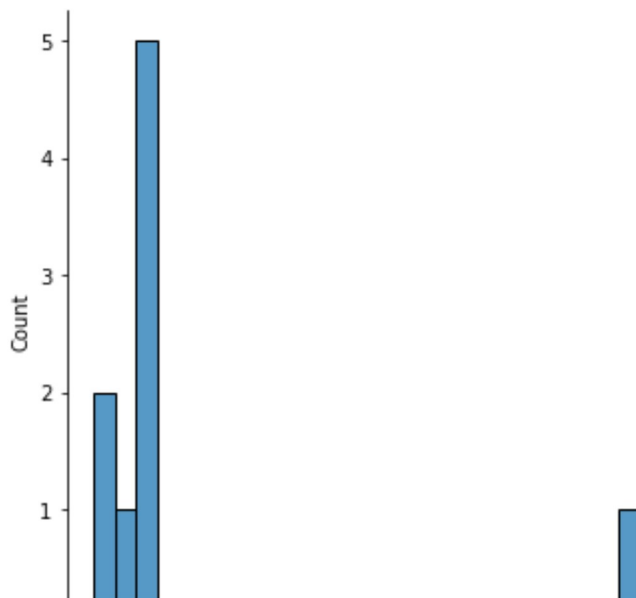
```
In [7]: sns.pairplot(df)
```

```
Out[7]: <seaborn.axisgrid.PairGrid at 0x1aaf40e6e20>
```



In [8]: `sns.distplot(a['Sum of Total Sales'])`

Out[8]: `<seaborn.axisgrid.FacetGrid at 0x1aaf3fc7b20>`



In [9]: `sns.heatmap(a.corr())`

Out[9]: `<AxesSubplot:>`



TO TRAIN THE MODEL - MODEL BUILDING

In [10]: `x = a[['Sum of Total Sales']]`

In [11]: `# to split my dataset into training and test data`
`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)`

```
In [12]: from sklearn.linear_model import LinearRegression  
lr = LinearRegression()  
lr.fit(x_train, y_train)
```

Out[12]: LinearRegression()

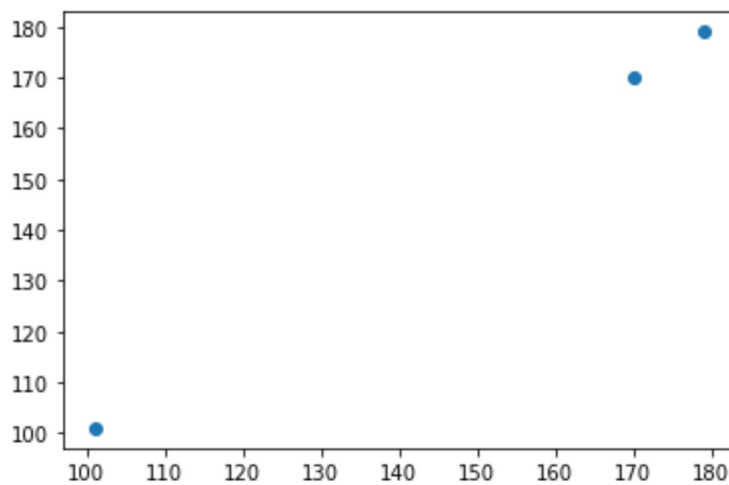
```
In [13]: coeff = pd.DataFrame(lr.coef_, x.columns, columns=['Co-efficient'])  
coeff
```

Out[13]:

	Co-efficient
Sum of Total Sales	1.0

```
In [14]: prediction = lr.predict(x_test)
```

Out[14]: <matplotlib.collections.PathCollection at 0x1aaf62a32e0>



```
In [15]: lr.score(x_test, y_test)
```

Out[15]: 1.0