In [1]:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

DATA COLLECTION

In [3]:

```python
a=pd.read_csv(r"C:\Users\user\Downloads\Fitness (2).csv")
a
```

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
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

In [4]:

```python
b=a.head(5)
b
```

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

DATA CLEANING AND PRE-PROCESSING

In [5]:

```
b.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 5 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Row Labels          5 non-null      object
 1   Sum of Jan          5 non-null      object
 2   Sum of Feb          5 non-null      object
 3   Sum of Mar          5 non-null      object
 4   Sum of Total Sales  5 non-null      int64
dtypes: int64(1), object(4)
memory usage: 328.0+ bytes
```

In [6]:

```
b.describe()
```

Out[6]:

	Sum of Total Sales
count	5.000000
mean	128.400000
std	42.317845
min	75.000000
25%	101.000000
50%	127.000000
75%	160.000000
max	179.000000

In [7]:

```
b.columns
```

Out[7]:

```
Index(['Row Labels', 'Sum of Jan', 'Sum of Feb', 'Sum of Mar',
       'Sum of Total Sales'],
      dtype='object')
```
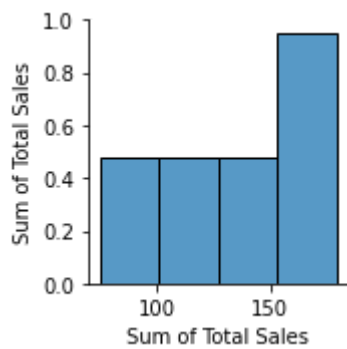
EDA AND VISUALIZATION

In [8]:

```python
sns.pairplot(b)
```
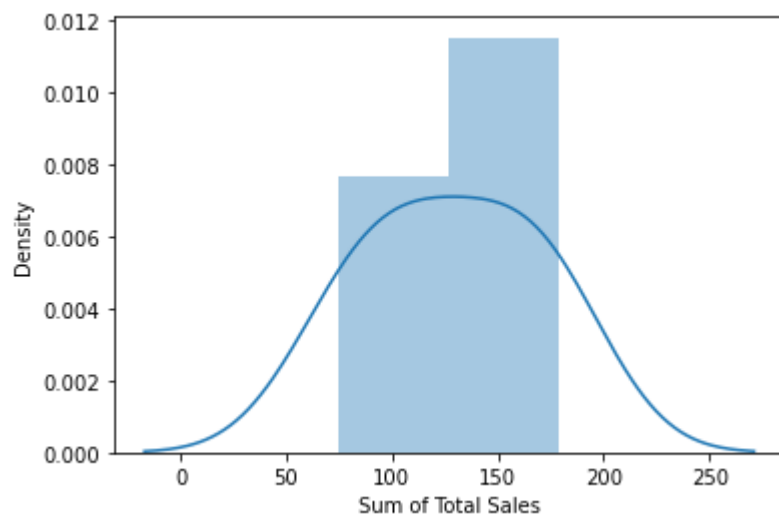
Out[8]:

```
<seaborn.axisgrid.PairGrid at 0x22519228fa0>
```


In [11]:

```python
sns.distplot(b['Sum of Total Sales'])
```

Out[11]:

```
<AxesSubplot:xlabel='Sum of Total Sales', ylabel='Density'>
```


In [13]:

```python
f=b[['Sum of Jan', 'Sum of Feb', 'Sum of Mar',
     'Sum of Total Sales']]
```

In [14]:

```
sns.heatmap(f.corr())
```

Out[14]:

```
<AxesSubplot:>
```


In [19]:

```
x=f[['Sum of Total Sales']]
y=f['Sum of Total Sales']
```

In [20]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.5)
```

In [21]:

```
from sklearn.linear_model import LinearRegression

lr=LinearRegression()
lr.fit(x_train,y_train)
```

Out[21]:

```
LinearRegression()
```

In [22]:

```
print(lr.intercept_)
```

```
1.4210854715202004e-14
```

In [23]:

```
r=pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])
r
```

Out[23]:

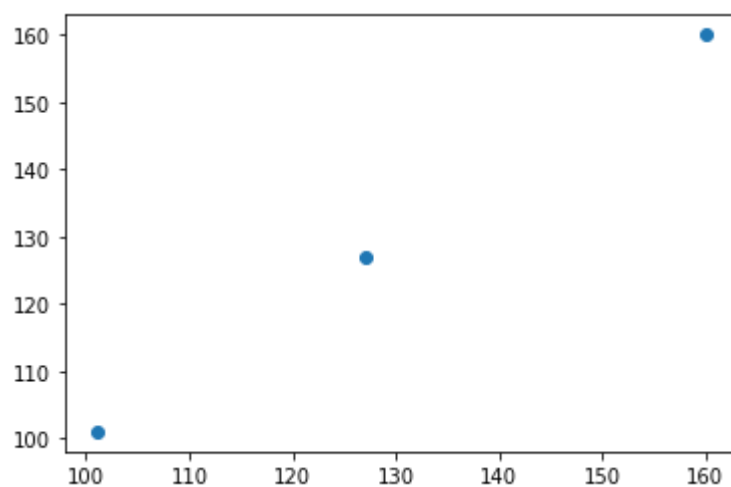|	Co-efficient
Sum of Total Sales	1.0

In [24]:

```
u=lr.predict(x_test)
plt.scatter(y_test,u)
```

Out[24]:

```
<matplotlib.collections.PathCollection at 0x2251b622ac0>
```


In [25]:

```
print(lr.score(x_test,y_test))
```

```
1.0
```

In [ ]: