In [1]:

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

DATA COLLECTION

In [2]:

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

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

In [3]:

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

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 PRE-PROCESSING

In [4]:

```
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 [5]:

```
b.describe()
```

Out[5]:

	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 [6]:

```
b.columns
```

Out[6]:

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

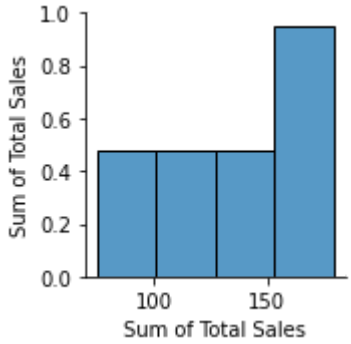
EDA AND VISUALIZATION

In [7]:

```
sns.pairplot(b)
```

Out[7]:

```
<seaborn.axisgrid.PairGrid at 0x1b156dd5fd0>
```



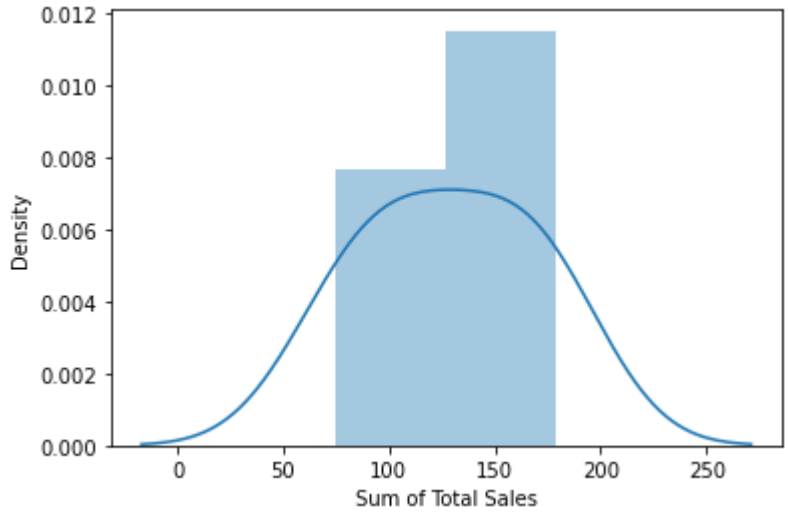
In [8]:

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

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in
a future version. Please adapt your code to use either `displot` (a figure
-level function with similar flexibility) or `histplot` (an axes-level fun
ction for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[8]:

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



In [9]:

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

In [10]:

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

Out[10]:

<AxesSubplot:>

In [11]:

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

In [12]:

```
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 [13]:

```
from sklearn.linear_model import LinearRegression

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

Out[13]:

```
LinearRegression()
```

In [14]:

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

```
-2.842170943040401e-14
```

In [15]:

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

Out[15]:

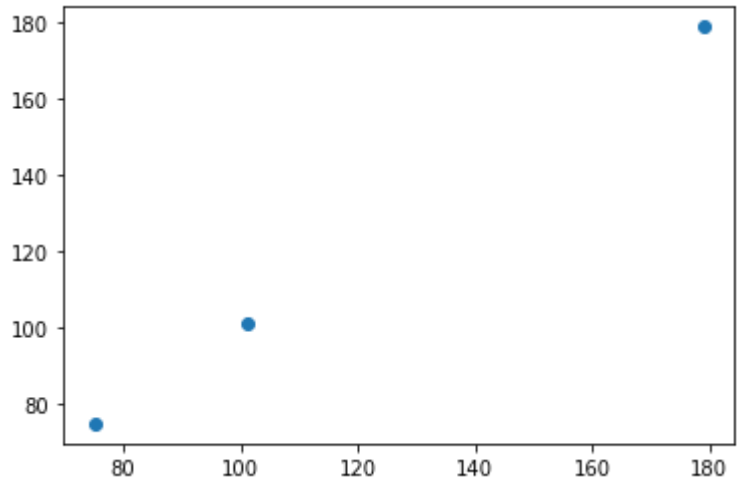
	Co-efficient
Sum of Total Sales	1.0

In [16]:

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

Out[16]:

```
<matplotlib.collections.PathCollection at 0x1b158f679d0>
```



In [17]:

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

```
1.0
```

In [18]:

```
lr.score(x_train,y_train)
```

Out[18]:

```
1.0
```

RIDGE REGRESSION

In [19]:

```
from sklearn.linear_model import Ridge,Lasso
```

In [20]:

```
rr=Ridge(alpha=10)
rr.fit(x_train,y_train)
```

Out[20]:

```
Ridge(alpha=10)
```

In [21]:

```
rr.score(x_test,y_test)
```

Out[21]:

```
0.9995692851273953
```

LASSO REGRESSION

In [22]:

```
la=Lasso(alpha=10)
la.fit(x_train,y_train)
```

Out[22]:

```
Lasso(alpha=10)
```

In [23]:

```
la.score(x_test,y_test)
```

Out[23]:

```
0.9982132771466768
```

In [24]:

```
from sklearn.linear_model import ElasticNet
p=ElasticNet()
p.fit(x_train,y_train)
```

Out[24]:

```
ElasticNet()
```

In [25]:

```
print(p.coef_)
```

```
[0.99633364]
```

In [26]:

```
print(p.intercept_)
```

```
0.5261228230980635
```

In [27]:

```
print(p.predict(x_test))
```

```
[ 75.25114574 178.86984418 101.15582035]
```

In [28]:

```
prediction=p.predict(x_test)
print(p.score(x_test,y_test))
```

```
0.9999821982191354
```

In [29]:

```
from sklearn import metrics
```

In [30]:

```
print("Mean Absolytre Error:",metrics.mean_absolute_error(y_test,prediction))
```

```
Mean Absolytre Error: 0.17904063550259272
```

In [31]:

```
print("Mean Squared Error:",metrics.mean_squared_error(y_test,prediction))
```

```
Mean Squared Error: 0.034764900053010646
```

In [32]:

```
print("Root Mean Squared Error:",np.sqrt(metrics.mean_squared_error(y_test,prediction)))
```

```
Root Mean Squared Error: 0.18645347959480577
```

In [ ]:

```

```