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

In [2]:
```
df=pd.read_csv("/Users/bob/Downloads/3_Fitness-1.csv")
df.fillna(0,inplace=True)
df
```

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]:
```
df.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

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

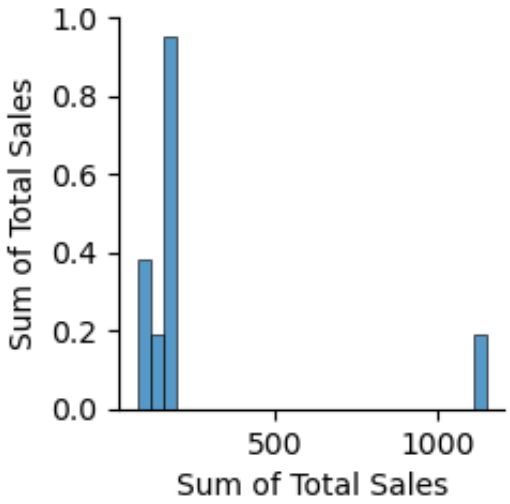
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9 entries, 0 to 8
Data columns (total 5 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   Row Labels          9 non-null      object
 1   Sum of Jan          9 non-null      object
 2   Sum of Feb          9 non-null      object
 3   Sum of Mar          9 non-null      object
 4   Sum of Total Sales  9 non-null      int64 
dtypes: int64(1), object(4)
memory usage: 488.0+ bytes
```

In [5]:
```
df.columns
```

Out[5]:
```
Index(['Row Labels', 'Sum of Jan', 'Sum of Feb', 'Sum of Mar',
       'Sum of Total Sales'],
      dtype='object')
```

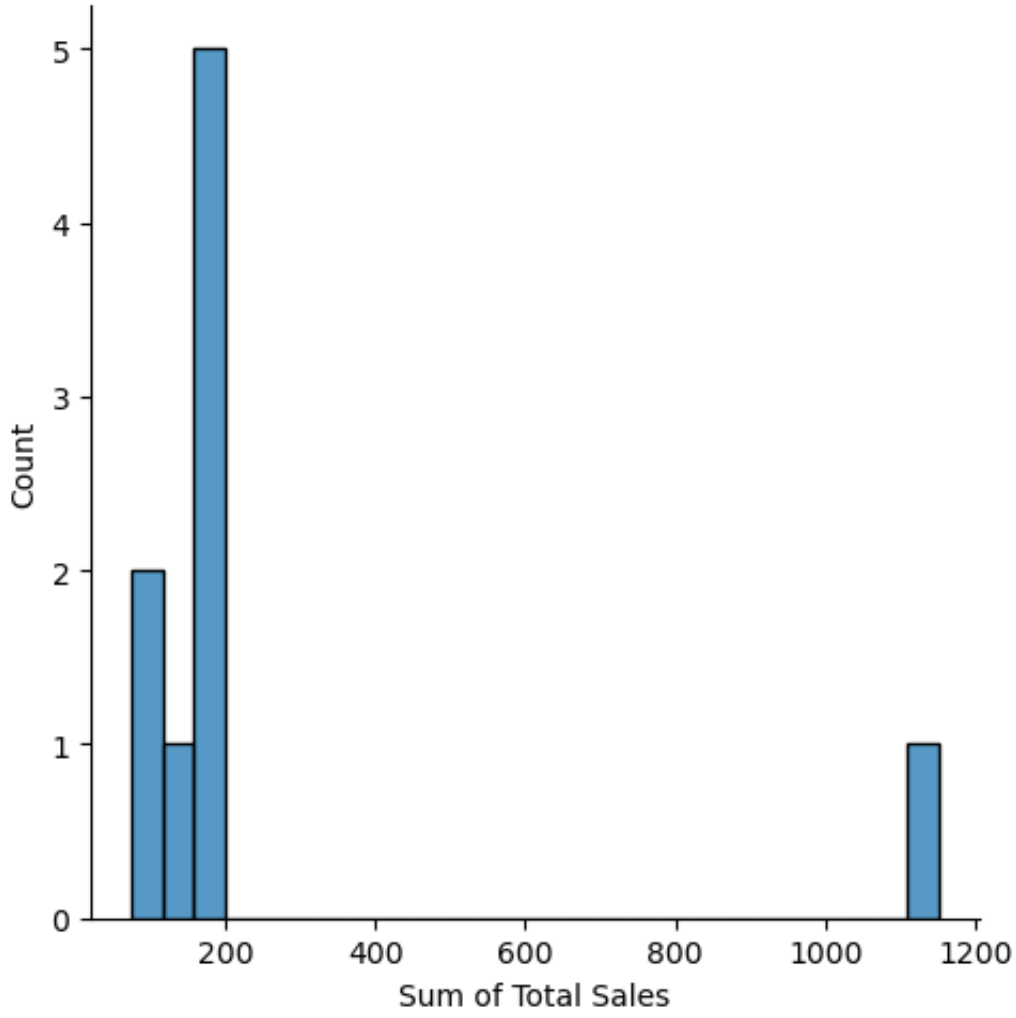
In [6]:
```
sns.pairplot(df)
```

Out[6]:
```
<seaborn.axisgrid.PairGrid at 0x7fc7e2f14970>
```



In [7]:
```
sns.displot(df['Sum of Total Sales'])
```

Out[7]:
```
<seaborn.axisgrid.FacetGrid at 0x7fc7b00ed990>
```



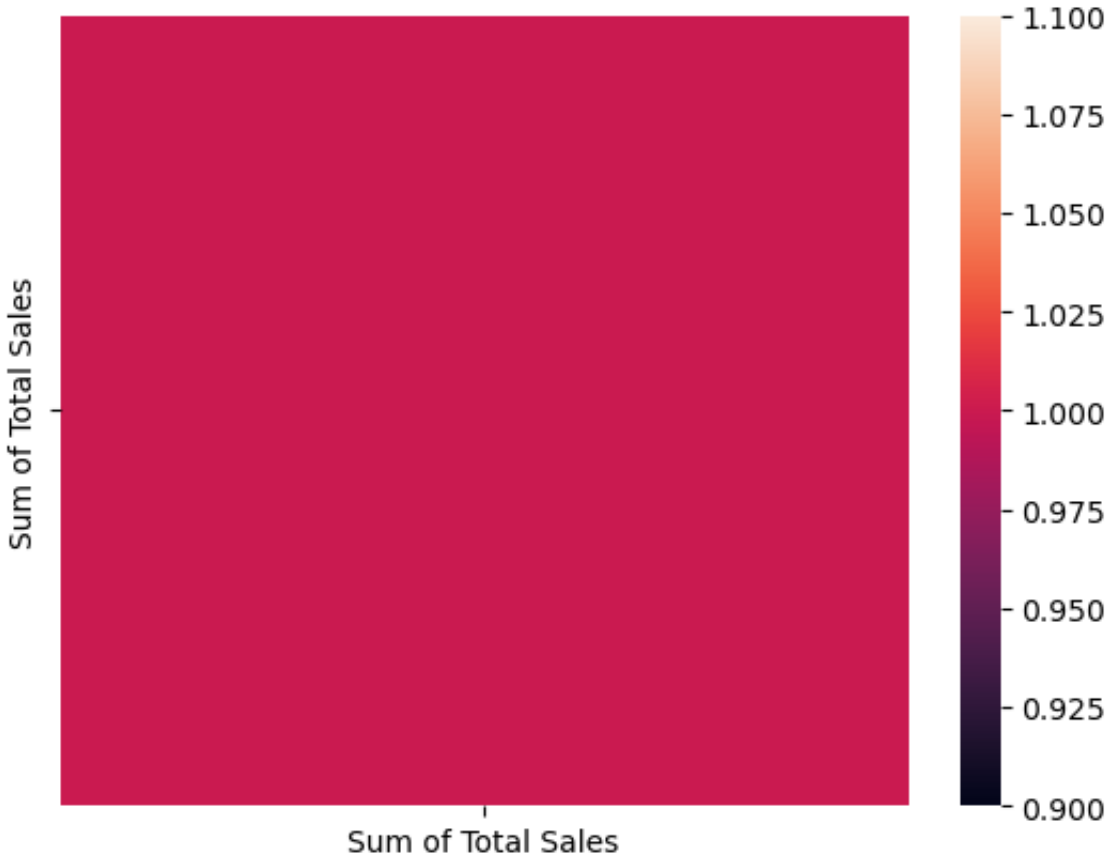
In [8]:
```
df1=df.drop(['Row Labels'],axis=1)
df1
df1=df1.drop(df1.index[1537:])
df1.isna().sum()
```

Out[8]:
```
Sum of Jan            0
Sum of Feb            0
Sum of Mar            0
Sum of Total Sales    0
dtype: int64
```

In [9]:
```
sns.heatmap(df1.corr())
```

```
/var/folders/2n/rrl24lws3pb1nz8_t911srvm0000gn/T/ipykernel_15617/781785195.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  sns.heatmap(df1.corr())
```

Out[9]: <Axes: >



In [10]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [11]:
```
df1.isna().sum()
```

Out[11]:
```
Sum of Jan            0
Sum of Feb            0
Sum of Mar            0
Sum of Total Sales    0
dtype: int64
```

In [18]:
```
y=df1['Sum of Total Sales']
x=df1.drop(['Sum of Jan','Sum of Feb','Sum of Mar'],axis=1)
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
print(x_train)
```

```
   Sum of Total Sales
7                 170
1                 160
0                  75
4                 179
3                 127
5                 167
```

In [19]:
```
model=LinearRegression()
model.fit(x_train,y_train)
model.intercept_
```

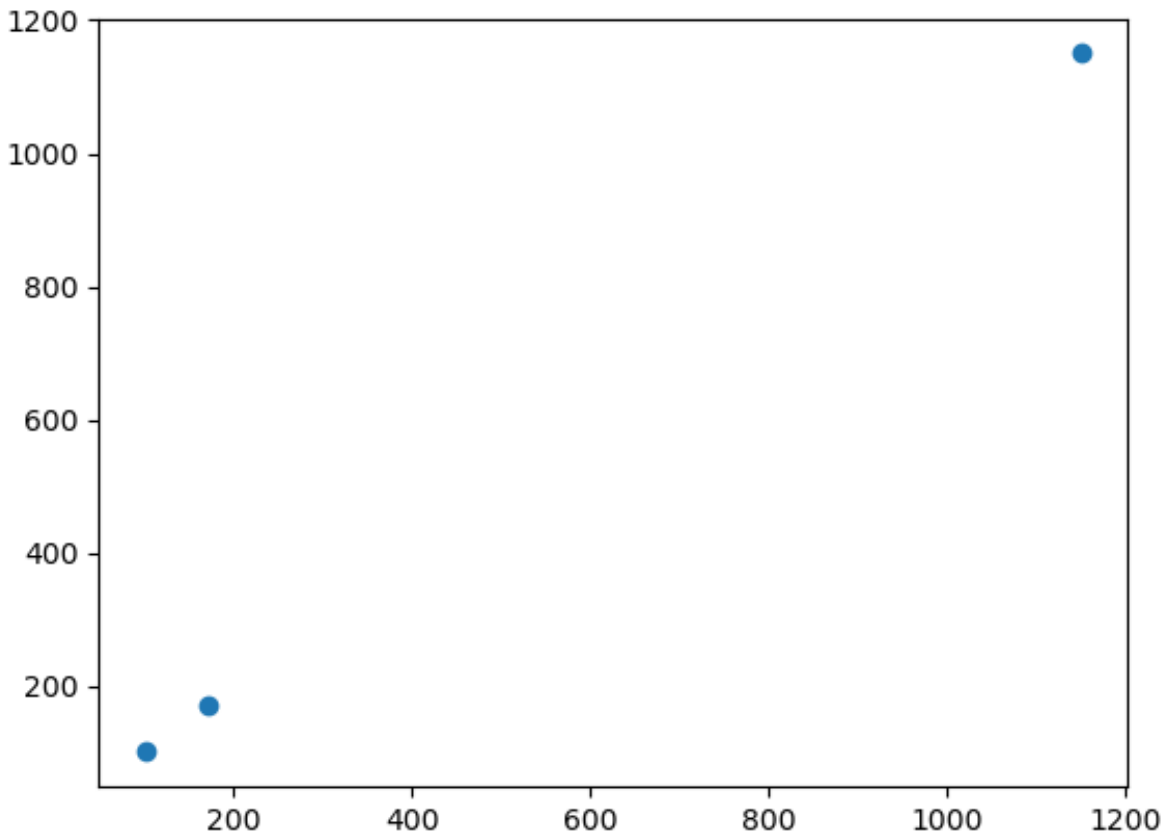
Out[19]:
```
8.526512829121202e-14
```

In [20]:
```
model.coef_
```

Out[20]:
```
array([1.])
```

In [21]:
```
prediction=model.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[21]:
```
<matplotlib.collections.PathCollection at 0x7fc7e3bc6380>
```



In [22]:
```
model.score(x_test,y_test)
```

Out[22]: 1.0

In [ ]:
```

```