D10 ¶

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

In [2]:
```
df=pd.read_csv(r"C:\Users\user\Downloads\3_Fitness-1.csv")
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(10)
```

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 [18]:
```
df["Sum of Jan"]=df["Sum of Jan"].replace("%","",regex=True).astype(float)
df["Sum of Feb"]=df["Sum of Feb"].replace("%","",regex=True).astype(float)
df["Sum of Mar"]=df["Sum of Mar"].replace("%","",regex=True).astype(float)
df
```

Out[18]:

	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 [19]:
```
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      float64
 2   Sum of Feb          9 non-null      float64
 3   Sum of Mar          9 non-null      float64
 4   Sum of Total Sales  9 non-null      int64  
dtypes: float64(3), int64(1), object(1)
memory usage: 488.0+ bytes
```

In [20]:
```
df.describe()
```

Out[20]:

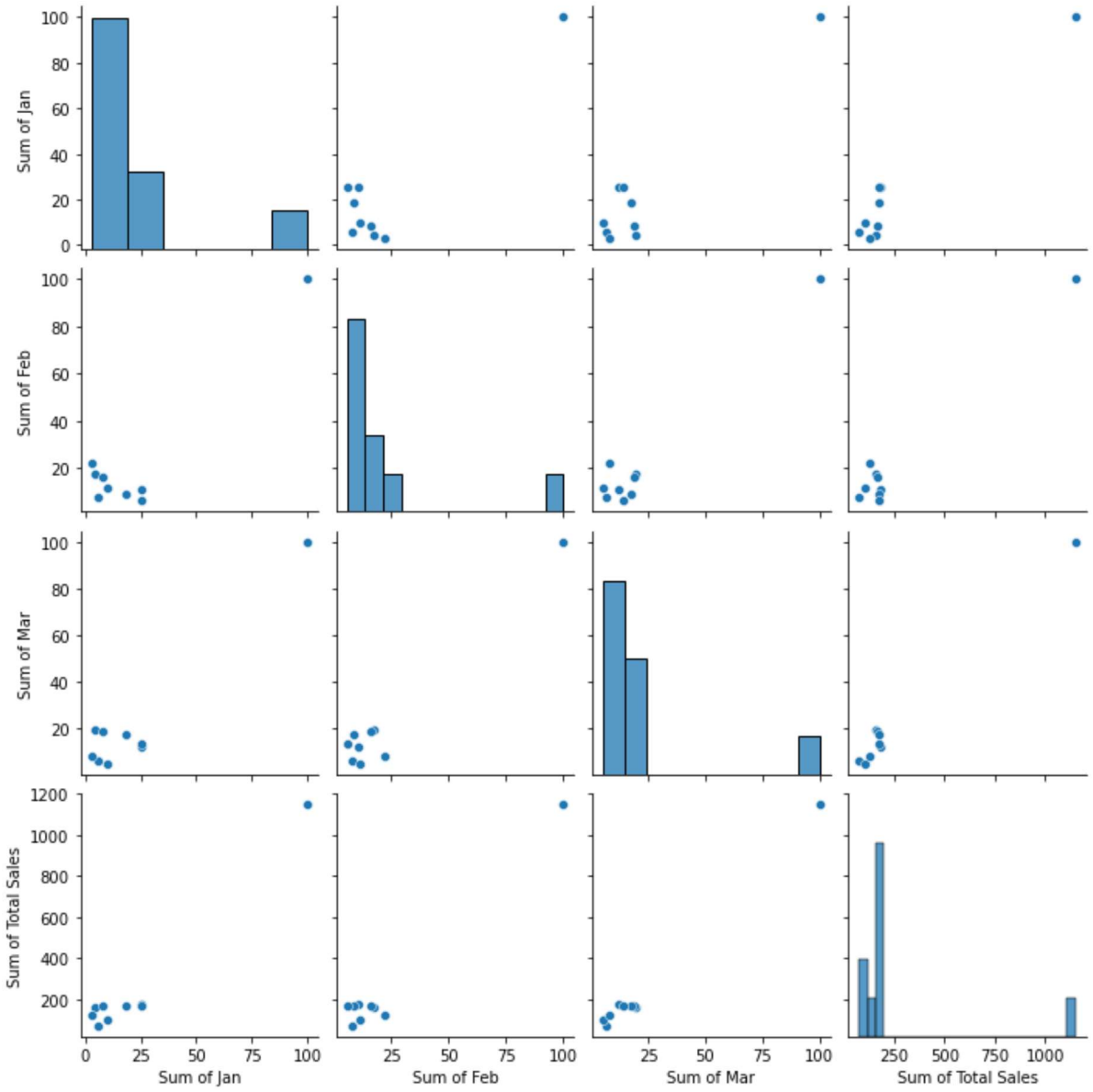
	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
count	9.000000	9.000000	9.000000	9.000000
mean	22.222222	22.223333	22.221111	255.555556
std	30.438329	29.612265	29.640999	337.332963
min	2.810000	5.930000	5.170000	75.000000
25%	5.620000	8.760000	7.880000	127.000000
50%	9.830000	11.600000	13.790000	167.000000
75%	25.280000	17.270000	18.470000	171.000000
max	100.000000	100.000000	100.000000	1150.000000

In [21]:
```
df.columns
```

Out[21]:
```
Index(['Row Labels', 'Sum of Jan', 'Sum of Feb', 'Sum of Mar',
       'Sum of Total Sales'],
      dtype='object')
```

In [22]:
```
sns.pairplot(df)
```

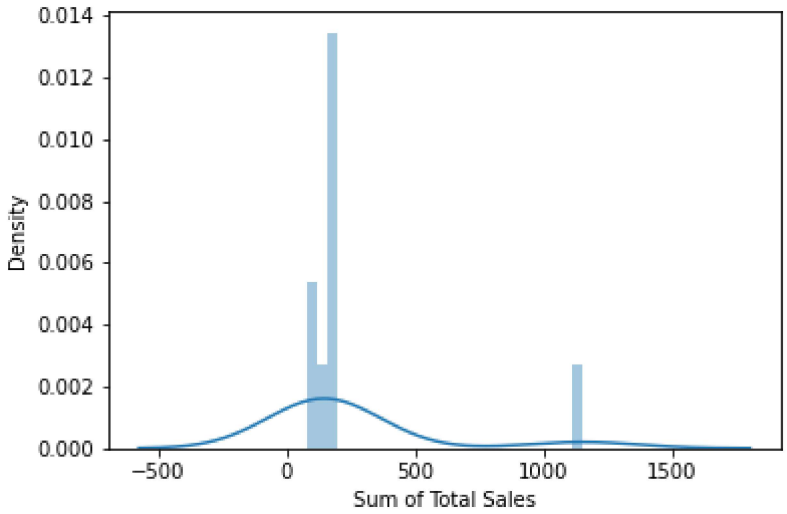
Out[22]:
```
<seaborn.axisgrid.PairGrid at 0x239b591a130>
```



In [23]:
```
sns.distplot(df["Sum of Total Sales"])
```

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

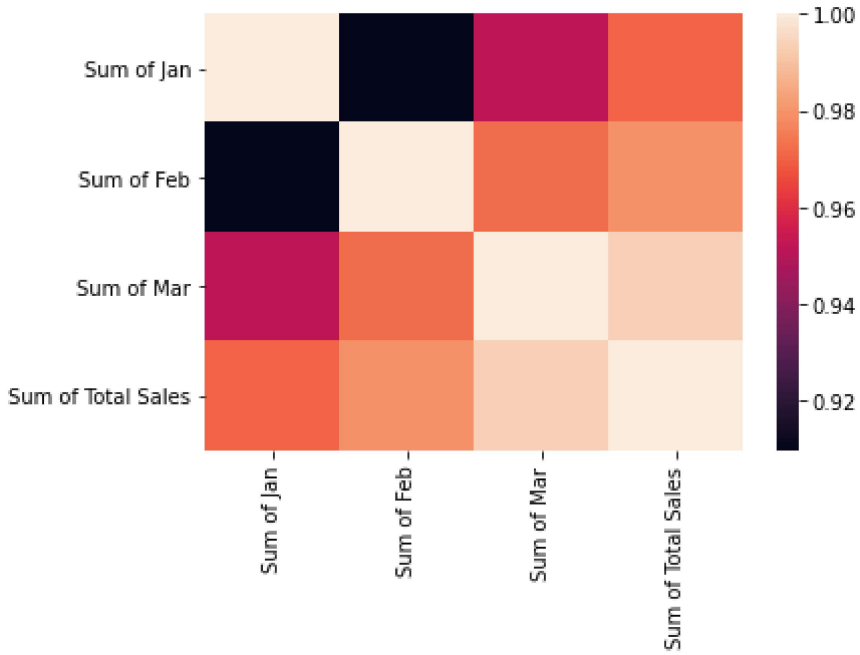
Out[23]: <AxesSubplot:xlabel='Sum of Total Sales', ylabel='Density'>



In [33]:
```
df1=df[['Sum of Jan', 'Sum of Feb', 'Sum of Mar',
       'Sum of Total Sales']]
```

In [34]:
```
sns.heatmap(df1.corr())
```

Out[34]: <AxesSubplot:>



In [36]:
```
x=df1[['Sum of Jan', 'Sum of Feb', 'Sum of Mar']]
y=df1['Sum of Total Sales']
```

In [37]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
```

In [38]:
```
from sklearn.linear_model import LinearRegression
lr=LinearRegression()
lr.fit(x_train,y_train)
```

Out[38]:
```
LinearRegression()
```

In [39]:
```
print(lr.intercept_)
```

```
-0.005149663362601586
```

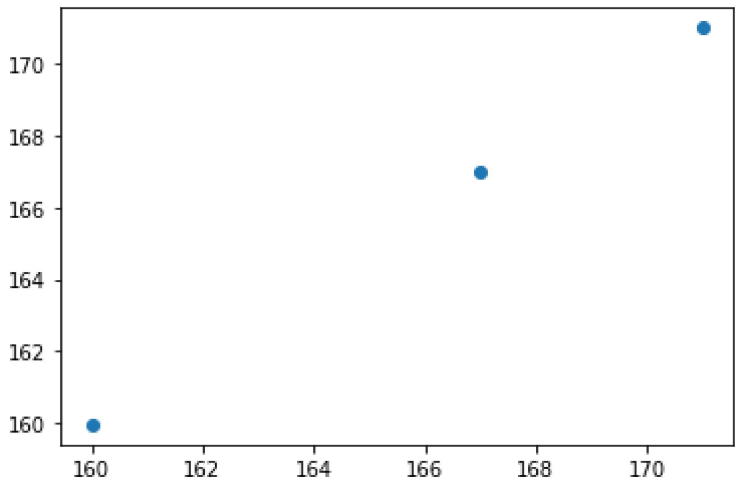
In [40]:
```
coeff = pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])
coeff
```

Out[40]:

	Co-efficient
Sum of Jan	3.561011
Sum of Feb	3.880201
Sum of Mar	4.058839

In [41]:
```
prediction=lr.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[41]:
```
<matplotlib.collections.PathCollection at 0x239b68b89d0>
```



In [42]:
```
print(lr.score(x_test,y_test))
```

```
0.9999832128877887
```

In [43]:
```
from sklearn.linear_model import Ridge,Lasso
```

In [44]:
```
rr=Ridge(alpha=10)
rr.fit(x_train,y_train)
```

Out[44]:
```
Ridge(alpha=10)
```

In [45]:
```
rr.score(x_test,y_test)
```

Out[45]:
```
0.9901349522324193
```

In [46]:
```
la=Lasso(alpha=10)
la.fit(x_train,y_train)
```

Out[46]:
```
Lasso(alpha=10)
```

In [47]:
```
la.score(x_test,y_test)
```

Out[47]:
```
0.9977817969271068
```

In [ ]:
```

```