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

In [3]: d=pd.read_csv(r"C:\Users\user\Downloads\Fitness.csv")

	SALESMAN	JAN	FEB	MAR	APR	MAY	JUN	TOTAL SALES	Unnamed: 8	Unnamed: 9	Unnamed: 10	Unn
0	ANU	70.0	80.0	75.0	60.0	72.0	55.0	412.0	NaN	NaN	NaN	
1	BABU	30.0	48.0	35.0	45.0	25.0	37.0	220.0	NaN	NaN	NaN	Ind Sales
2	CHANDRU	65.0	54.0	49.0	54.0	35.0	65.0	322.0	NaN	NaN	NaN	2. Fi F
3	DAVID	85.0	71.0	68.0	77.0	88.0	73.0	462.0	NaN	NaN	NaN	3. A using to c perce
4	EINSTEIN	55.0	25.0	45.0	50.0	53.0	30.0	258.0	NaN	NaN	NaN	4.
5	FAROOK	35.0	45.0	15.0	45.0	45.0	25.0	210.0	NaN	NaN	NaN	5. R retui rai v
6	GOWTHAM	75.0	66.0	59.0	65.0	56.0	30.0	351.0	NaN	NaN	NaN	
7	HARSHITH	29.0	35.0	49.0	48.0	35.0	55.0	247.0	NaN	NaN	NaN	
8	INIYAN	35.0	35.0	50.0	59.0	67.0	73.0	319.0	NaN	NaN	NaN	
9	JOHN	77.0	85.0	77.0	68.0	56.0	25.0	388.0	NaN	NaN	NaN	
10	MONTHLY SALES	556.0	544.0	522.0	571.0	532.0	468.0	NaN	3193.0	NaN	NaN	
11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3189.0	NaN	NaN	NaN	
	1 2 3 4 5 6 7 8 9	ANU	0 ANU 70.0 1 BABU 30.0 2 CHANDRU 65.0 3 DAVID 85.0 5 FAROOK 35.0 6 GOWTHAM 75.0 7 HARSHITH 29.0 8 INIYAN 35.0 9 JOHN 77.0 10 MONTHLY SALES 556.0	o ANU 70.0 80.0 1 BABU 30.0 48.0 2 CHANDRU 65.0 54.0 3 DAVID 85.0 71.0 4 EINSTEIN 55.0 25.0 5 FAROOK 35.0 45.0 6 GOWTHAM 75.0 66.0 7 HARSHITH 29.0 35.0 8 INIYAN 35.0 35.0 9 JOHN 77.0 85.0 10 MONTHLY SALES 556.0 544.0	ANU 770.0 880.0 75.0 1 BABU 30.0 48.0 35.0 2 CHANDRU 65.0 54.0 49.0 3 DAVID 85.0 71.0 68.0 4 EINSTEIN 55.0 25.0 45.0 5 FAROOK 35.0 45.0 15.0 6 GOWTHAM 75.0 66.0 59.0 7 HARSHITH 29.0 35.0 49.0 8 INIYAN 35.0 35.0 50.0 9 JOHN 77.0 85.0 77.0 10 MONTHLY SALES 556.0 544.0 522.0	ANU 70.0 80.0 75.0 60.0 1 BABU 30.0 48.0 35.0 45.0 2 CHANDRU 65.0 54.0 49.0 54.0 3 DAVID 85.0 71.0 68.0 77.0 4 EINSTEIN 55.0 25.0 45.0 50.0 5 FAROOK 35.0 45.0 15.0 45.0 6 GOWTHAM 75.0 66.0 59.0 65.0 7 HARSHITH 29.0 35.0 49.0 48.0 8 INIYAN 35.0 35.0 50.0 59.0 9 JOHN 77.0 85.0 77.0 68.0 10 MONTHLY SALES 556.0 544.0 522.0 571.0	ANU 70.0 80.0 75.0 60.0 72.0 1 BABU 30.0 48.0 35.0 45.0 25.0 2 CHANDRU 65.0 54.0 49.0 54.0 35.0 3 DAVID 85.0 71.0 68.0 77.0 88.0 4 EINSTEIN 55.0 25.0 45.0 50.0 53.0 5 FAROOK 35.0 45.0 15.0 45.0 45.0 6 GOWTHAM 75.0 66.0 59.0 65.0 56.0 7 HARSHITH 29.0 35.0 49.0 48.0 35.0 8 INIYAN 35.0 35.0 50.0 59.0 67.0 9 JOHN 77.0 85.0 77.0 68.0 56.0 10 MONTHLY MONTHLY SALES 556.0 544.0 522.0 571.0 532.0	ANU 70.0 80.0 75.0 60.0 72.0 55.0 ANU 30.0 48.0 35.0 45.0 25.0 37.0 ANDRU 65.0 54.0 49.0 54.0 35.0 65.0 ANDRU 85.0 71.0 68.0 77.0 88.0 73.0 ANDRU 85.0 71.0 68.0 77.0 88.0 73.0 ANDRU 85.0 25.0 45.0 50.0 53.0 30.0 ANDRU 85.0 45.0 15.0 45.0 53.0 30.0 ANDRU 85.0 45.0 25.0 45.0 50.0 53.0 30.0 ANDRU 45.0 45.0 45.0 45.0 45.0 25.0 ANDRU 45.0 45.0 45.0 45.0 45.0 25.0 ANDRU 45.0 45.0 45.0 45.0 45.0 45.0 25.0 ANDRU 45.0 45.0 45	SALESMAN JAN FEB MAR APR MAY JON SALES 0 ANU 70.0 80.0 75.0 60.0 72.0 55.0 412.0 1 BABU 30.0 48.0 35.0 45.0 25.0 37.0 220.0 2 CHANDRU 65.0 54.0 49.0 54.0 35.0 65.0 322.0 3 DAVID 85.0 71.0 68.0 77.0 88.0 73.0 462.0 4 EINSTEIN 55.0 25.0 45.0 50.0 53.0 30.0 258.0 5 FAROOK 35.0 45.0 15.0 45.0 45.0 25.0 210.0 6 GOWTHAM 75.0 66.0 59.0 65.0 56.0 30.0 351.0 7 HARSHITH 29.0 35.0 49.0 48.0 35.0 55.0 247.0 8 INIYAN 35.0 35.0 50.0 </th <th>SALESMAN JAN FEB MAR APA MAR JON SALES 8 0 ANU 70.0 80.0 75.0 60.0 72.0 55.0 412.0 NaN 1 BABU 30.0 48.0 35.0 45.0 25.0 37.0 220.0 NaN 2 CHANDRU 65.0 54.0 49.0 54.0 35.0 65.0 322.0 NaN 3 DAVID 85.0 71.0 68.0 77.0 88.0 73.0 462.0 NaN 4 EINSTEIN 55.0 25.0 45.0 50.0 53.0 30.0 258.0 NaN 5 FAROOK 35.0 45.0 45.0 45.0 25.0 25.0 NaN 6 GOWTHAM 75.0 66.0 59.0 65.0 56.0 30.0 351.0 NaN 7 HARSHITH 29.0 35.0 49.0 48.0 35.0 55.0</th> <th> Name</th> <th> SALESMAN SAN FEB WAR SAN SAN SALES 8 9 10 </th>	SALESMAN JAN FEB MAR APA MAR JON SALES 8 0 ANU 70.0 80.0 75.0 60.0 72.0 55.0 412.0 NaN 1 BABU 30.0 48.0 35.0 45.0 25.0 37.0 220.0 NaN 2 CHANDRU 65.0 54.0 49.0 54.0 35.0 65.0 322.0 NaN 3 DAVID 85.0 71.0 68.0 77.0 88.0 73.0 462.0 NaN 4 EINSTEIN 55.0 25.0 45.0 50.0 53.0 30.0 258.0 NaN 5 FAROOK 35.0 45.0 45.0 45.0 25.0 25.0 NaN 6 GOWTHAM 75.0 66.0 59.0 65.0 56.0 30.0 351.0 NaN 7 HARSHITH 29.0 35.0 49.0 48.0 35.0 55.0	Name	SALESMAN SAN FEB WAR SAN SAN SALES 8 9 10

In [4]:

d.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 12 entries, 0 to 11 Data columns (total 12 columns): # Column Non-Null Count Dtype SALESMAN object 0 11 non-null float64 1 JAN 11 non-null 2 FEB 11 non-null float64 11 non-null 3 MAR float64 4 11 non-null APR float64 5 11 non-null float64 MAY 6 JUN 11 non-null float64 float64 7 TOTAL SALES 11 non-null float64 8 1 non-null Unnamed: 8 9 float64 Unnamed: 9 0 non-null 10 Unnamed: 10 0 non-null float64 11 Unnamed: 11 6 non-null object dtypes: float64(10), object(2) memory usage: 1.2+ KB d.isna()

In [5]:

Out[5]: **TOTAL Unnamed: Unnamed: Unnamed: Unnamed** CALECMANI ADD BAAV

		SALESMAN	JAN	FEB	MAR	APR	MAY	JUN	SALES	8	9	10	
	0	False	False	False	False	False	False	False	False	True	True	True	٦
	1	False	False	False	False	False	False	False	False	True	True	True	Fi
	2	False	False	False	False	False	False	False	False	True	True	True	Fi
	3	False	False	False	False	False	False	False	False	True	True	True	Fi
	4	False	False	False	False	False	False	False	False	True	True	True	Fi
	5	False	False	False	False	False	False	False	False	True	True	True	Fi
	6	False	False	False	False	False	False	False	False	True	True	True	1
	7	False	False	False	False	False	False	False	False	True	True	True	Fi
	8	False	False	False	False	False	False	False	False	True	True	True	1
	9	False	False	False	False	False	False	False	False	True	True	True	1
1	10	False	False	False	False	False	False	False	True	False	True	True	1
1	11	True	True	True	True	True	True	True	False	True	True	True	1

In [10]:

d.fillna(value=0)

Out[10]: TOTAL Unnamed: Unnamed: Unnamed: Unn JUN **SALESMAN JAN FEB** MAR APR MAY **SALES** 8 9 10 60.0 55.0 0.0 0.0 0.0 0 ANU 70.0 80.0 75.0 72.0 412.0 Ind 1 BABU 30.0 48.0 35.0 45.0 25.0 37.0 220.0 0.0 0.0 0.0 Sales

	SALESMAN	JAN	FEB	MAR	APR	MAY	JUN	TOTAL SALES	Unnamed: 8	Unnamed: 9	Unnamed: 10	Unn
2	CHANDRU	65.0	54.0	49.0	54.0	35.0	65.0	322.0	0.0	0.0	0.0	2. Fi F
3	DAVID	85.0	71.0	68.0	77.0	88.0	73.0	462.0	0.0	0.0	0.0	3. A using to c perce
4	EINSTEIN	55.0	25.0	45.0	50.0	53.0	30.0	258.0	0.0	0.0	0.0	4.
5	FAROOK	35.0	45.0	15.0	45.0	45.0	25.0	210.0	0.0	0.0	0.0	5. R retui rai V
6	GOWTHAM	75.0	66.0	59.0	65.0	56.0	30.0	351.0	0.0	0.0	0.0	
7	HARSHITH	29.0	35.0	49.0	48.0	35.0	55.0	247.0	0.0	0.0	0.0	
8	INIYAN	35.0	35.0	50.0	59.0	67.0	73.0	319.0	0.0	0.0	0.0	
9	JOHN	77.0	85.0	77.0	68.0	56.0	25.0	388.0	0.0	0.0	0.0	
10	MONTHLY SALES	556.0	544.0	522.0	571.0	532.0	468.0	0.0	3193.0	0.0	0.0	
11	0	0.0	0.0	0.0	0.0	0.0	0.0	3189.0	0.0	0.0	0.0	

In [11]:

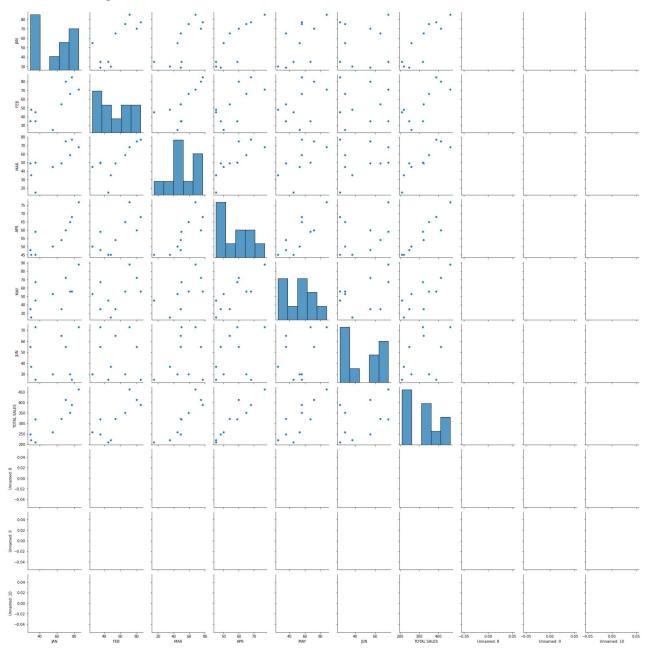
d.describe()

Out[11]:

	JAN	FEB	MAR	APR	MAY	JUN	TOTAL SALES	Unnamed: 8
count	11.000000	11.000000	11.000000	11.000000	11.000000	11.000000	11.000000	1.0
mean	101.090909	98.909091	94.909091	103.818182	96.727273	85.090909	579.818182	3193.0
std	152.263886	148.884153	142.770763	155.277054	145.500578	128.347540	869.142775	NaN
min	29.000000	25.000000	15.000000	45.000000	25.000000	25.000000	210.000000	3193.0
25%	35.000000	40.000000	47.000000	49.000000	40.000000	30.000000	252.500000	3193.0
50%	65.000000	54.000000	50.000000	59.000000	56.000000	55.000000	322.000000	3193.0
75%	76.000000	75.500000	71.500000	66.500000	69.500000	69.000000	400.000000	3193.0
max	556.000000	544.000000	522.000000	571.000000	532.000000	468.000000	3189.000000	3193.0

```
In [12]:
            d.columns
Out[12]: Index(['SALESMAN', 'JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'TOTAL SALES', 'Unnamed: 8', 'Unnamed: 9', 'Unnamed: 10', 'Unnamed: 11'],
                   dtype='object')
In [13]:
            d.index
           RangeIndex(start=0, stop=12, step=1)
In [47]:
            d=d.head(10)
            d
                                                                TOTAL Unnamed:
                                                                                    Unnamed: Unnamed:
                                                                                                            Unnamed
Out[47]:
               SALESMAN JAN
                                 FEB MAR APR MAY JUN
                                                                 SALES
                                                                                 8
                                                                                             9
                                                                                                        10
                                                                                                                    11
           0
                                              60.0
                                                          55.0
                     ANU
                            70.0
                                 80.0
                                        75.0
                                                    72.0
                                                                 412.0
                                                                              NaN
                                                                                          NaN
                                                                                                      NaN
                                                                                                                  NaN
                                                                                                                    1
                                                                                                             Individua
           1
                    BABU
                            30.0 48.0
                                        35.0
                                              45.0
                                                    25.0 37.0
                                                                 220.0
                                                                              NaN
                                                                                          NaN
                                                                                                      NaN
                                                                                                            Sales using
                                                                                                                 Sum(
                                                                                                            2. Find thε
                                                                                                               patterr
                                                                                                                 trenc
                                                    35.0 65.0
                CHANDRU 65.0 54.0
                                        49.0
                                              54.0
                                                                 322.0
                                                                              NaN
                                                                                          NaN
                                                                                                      NaN
                                                                                                                 using
                                                                                                            conditiona
                                                                                                                  fo..
                                                                                                             3. Analyze
                                                                                                            using Pivot
           3
                   DAVID 85.0 71.0
                                        68.0 77.0
                                                    88.0 73.0
                                                                 462.0
                                                                              NaN
                                                                                          NaN
                                                                                                               table as
                                                                                                      NaN
                                                                                                               columr
                                                                                                            percentage
                                                                                                               4. Insert
           4
                 EINSTEIN 55.0 25.0
                                        45.0
                                              50.0
                                                    53.0 30.0
                                                                 258.0
                                                                              NaN
                                                                                          NaN
                                                                                                      NaN
                                                                                                                 Pivo<sup>1</sup>
                                                                                                                 charts
                                                                                                             5. Rank() -
                                                                                                            returns the
           5
                  FAROOK 35.0 45.0
                                        15.0 45.0
                                                    45.0 25.0
                                                                 210.0
                                                                              NaN
                                                                                          NaN
                                                                                                      NaN
                                                                                                              rank of a
                                                                                                                 giver
                                                                                                               value ..
               GOWTHAM 75.0 66.0
                                        59.0
                                              65.0
                                                    56.0 30.0
                                                                 351.0
                                                                              NaN
                                                                                          NaN
                                                                                                      NaN
                                                                                                                  NaN
           7
                HARSHITH 29.0 35.0
                                        49.0
                                              48.0
                                                    35.0
                                                          55.0
                                                                 247.0
                                                                              NaN
                                                                                          NaN
                                                                                                      NaN
                                                                                                                    33
           8
                   INIYAN 35.0 35.0
                                        50.0
                                              59.0
                                                    67.0
                                                          73.0
                                                                  319.0
                                                                              NaN
                                                                                          NaN
                                                                                                      NaN
                                                                                                                  NaN
           9
                    JOHN 77.0 85.0
                                        77.0
                                              68.0
                                                     56.0 25.0
                                                                  388.0
                                                                              NaN
                                                                                          NaN
                                                                                                      NaN
                                                                                                                  NaN
In [48]:
            sns.pairplot(d)
```

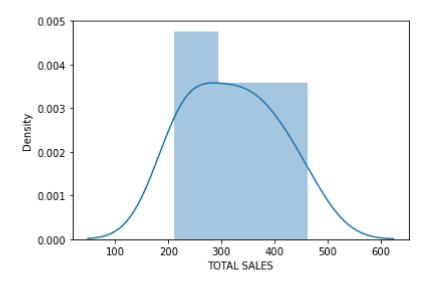
Out[48]: <seaborn.axisgrid.PairGrid at 0x2a760844400>



```
In [49]: sns.distplot(d['TOTAL SALES'])
```

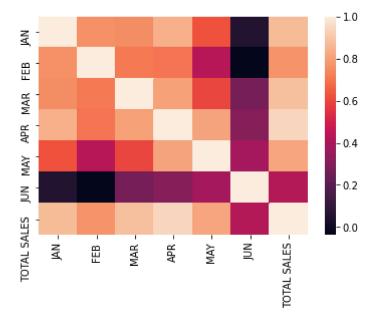
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning:
 distplot` is a deprecated function and will be removed in a future version. Please adap
t your code to use either `displot` (a figure-level function with similar flexibility) o
r `histplot` (an axes-level function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

Out[49]: <AxesSubplot:xlabel='TOTAL SALES', ylabel='Density'>



```
In [50]:
d1=d[['JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'TOTAL SALES']]
sns.heatmap(d1.corr())
```

Out[50]: <AxesSubplot:>



```
In [51]: x=d1[['JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN']]
y=d1[ 'TOTAL SALES']
```

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 [53]: from sklearn.linear_model import LinearRegression

In [54]: lr=LinearRegression()
lr.fit(x_train,y_train)

```
Out[54]: LinearRegression()
In [56]:
           print(lr.intercept_)
          -21.98349503881923
In [57]:
           coeff =pd.DataFrame(lr.coef_,x.columns,columns=["Co-efficient"])
           coeff
                Co-efficient
Out[57]:
           JAN
                  0.623799
           FEB
                  1.374083
          MAR
                  0.863091
          APR
                  1.385802
          MAY
                  1.041095
          JUN
                  1.046403
In [58]:
           prediction =lr.predict(x_test)
           py.scatter(y_test,prediction)
Out[58]: <matplotlib.collections.PathCollection at 0x2a7657b83a0>
          340
          320
          300
          280
          260
          240
                260
                          280
                                   300
                                             320
                                                       340
In [55]:
           print(lr.score(x_test,y_test))
          0.9209143047698785
 In [ ]:
```