In [ ]:
```
kaviyadevi 20106064
```

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

In [27]:
```
#to import dataset
data=pd.read_csv(r"C:\Users\user\Downloads\5_Instagram data - 5_Instagram data.cs
data
```

Out[27]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits
0	3920	2586	1028	619	56	98	9	5	162	35
1	5394	2727	1838	1174	78	194	7	14	224	48
2	4021	2085	1188	0	533	41	11	1	131	62
3	4528	2700	621	932	73	172	10	7	213	23
4	2518	1704	255	279	37	96	5	4	123	8
...	...	...	...	...	...	...	...	...	...	...
114	13700	5185	3041	5352	77	573	2	38	373	73
115	5731	1923	1368	2266	65	135	4	1	148	20
116	4139	1133	1538	1367	33	36	0	1	92	34
117	32695	11815	3147	17414	170	1095	2	75	549	148

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits
118	36919	13473	4176	16444	2547	653	5	26	443	611

119 rows × 13 columns

In [28]:
```
#to display top 5 rows
data.head()
```

Out[28]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follo
0	3920	2586	1028	619	56	98	9	5	162	35	
1	5394	2727	1838	1174	78	194	7	14	224	48	
2	4021	2085	1188	0	533	41	11	1	131	62	
3	4528	2700	621	932	73	172	10	7	213	23	
4	2518	1704	255	279	37	96	5	4	123	8	

DATA CLEANING AND PREPROCESSING

In [29]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119 entries, 0 to 118
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   Impressions     119 non-null    int64 
 1   From Home       119 non-null    int64 
 2   From Hashtags   119 non-null    int64 
 3   From Explore    119 non-null    int64 
 4   From Other      119 non-null    int64 
 5   Saves           119 non-null    int64 
 6   Comments        119 non-null    int64 
 7   Shares          119 non-null    int64 
 8   Likes           119 non-null    int64 
 9   Profile Visits  119 non-null    int64 
 10  Follows         119 non-null    int64 
 11  Caption         119 non-null    object
 12  Hashtags        119 non-null    object
dtypes: int64(11), object(2)
memory usage: 12.2+ KB
```

In [30]:
```
#to display summary of statistics
data.describe()
```

Out[30]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Commer
count	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.0000
mean	5703.991597	2475.789916	1887.512605	1078.100840	171.092437	153.310924	6.6638
std	4843.780105	1489.386348	1884.361443	2613.026132	289.431031	156.317731	3.5445
min	1941.000000	1133.000000	116.000000	0.000000	9.000000	22.000000	0.0000
25%	3467.000000	1945.000000	726.000000	157.500000	38.000000	65.000000	4.0000
50%	4289.000000	2207.000000	1278.000000	326.000000	74.000000	109.000000	6.0000
75%	6138.000000	2602.500000	2363.500000	689.500000	196.000000	169.000000	8.0000
max	36919.000000	13473.000000	11817.000000	17414.000000	2547.000000	1095.000000	19.0000

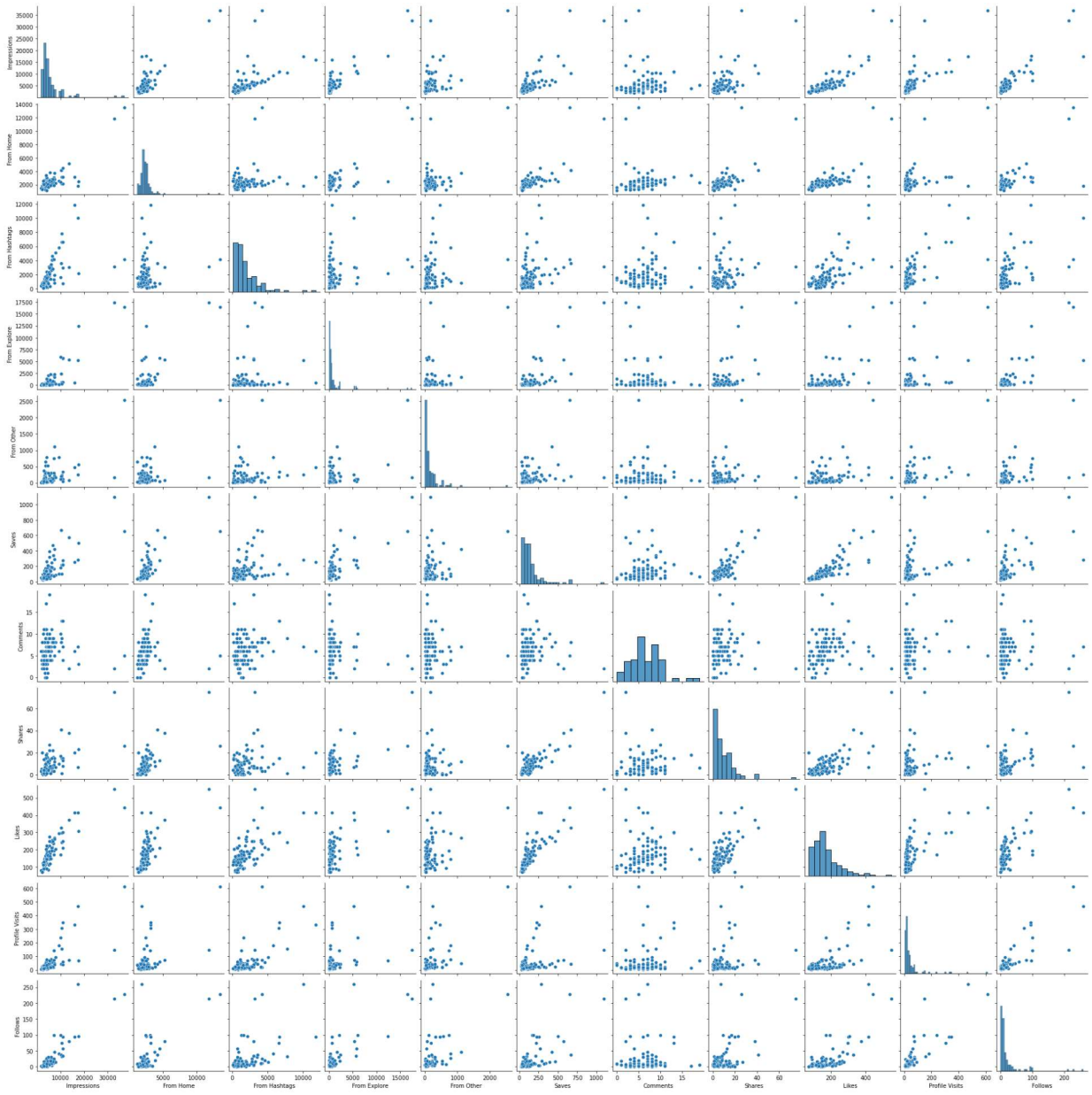
In [31]:
```
#to display the column heading
data.columns
```

Out[31]:
```
Index(['Impressions', 'From Home', 'From Hashtags', 'From Explore',
       'From Other', 'Saves', 'Comments', 'Shares', 'Likes', 'Profile Visits',
       'Follows', 'Caption', 'Hashtags'],
      dtype='object')
```

EDA and DATA VISUALIZATION

In [32]:
```
sns.pairplot(data)
```

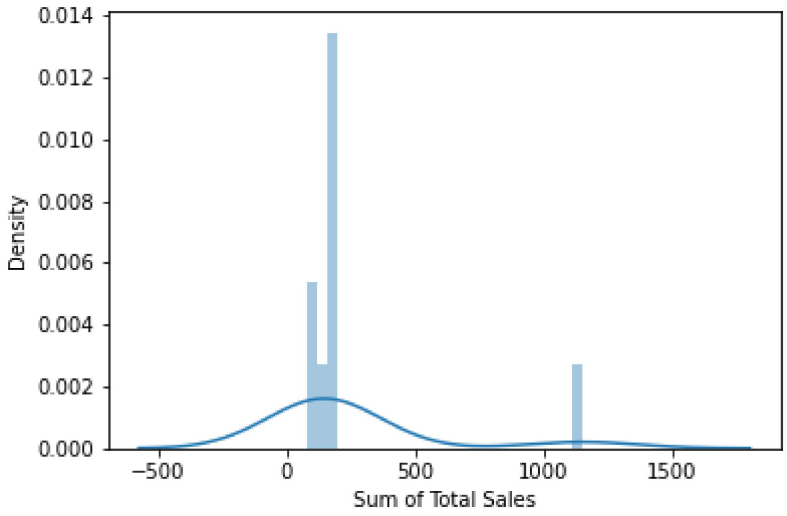
Out[32]:
```
<seaborn.axisgrid.PairGrid at 0x1e5abfb4d30>
```



In [9]:
```
sns.distplot(data["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 function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[9]: <AxesSubplot:xlabel='Sum of Total Sales', ylabel='Density'>



In [33]:
```
df=data[['Impressions', 'From Home', 'From Hashtags', 'From Explore',
       'From Other', 'Saves', 'Comments', 'Shares', 'Likes', 'Profile Visits',
       'Follows', 'Caption', 'Hashtags']]
```

In [11]:
```
sns.heatmap(df.corr())
```

Out[11]: <AxesSubplot:>

MODEL TRAINNING

In [42]:
```
x=df[['From Home', 'From Hashtags', 'From Explore','From Other', 'Saves', 'Commer
y=df['Impressions']
```

In [43]:
```
#to split my dataset into trainning and test

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 [44]:
```
from sklearn.linear_model import LinearRegression

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

Out[44]: LinearRegression()

In [45]:
```
#to find intercept
print(lr.intercept_)
```

```
117.33230332429048
```

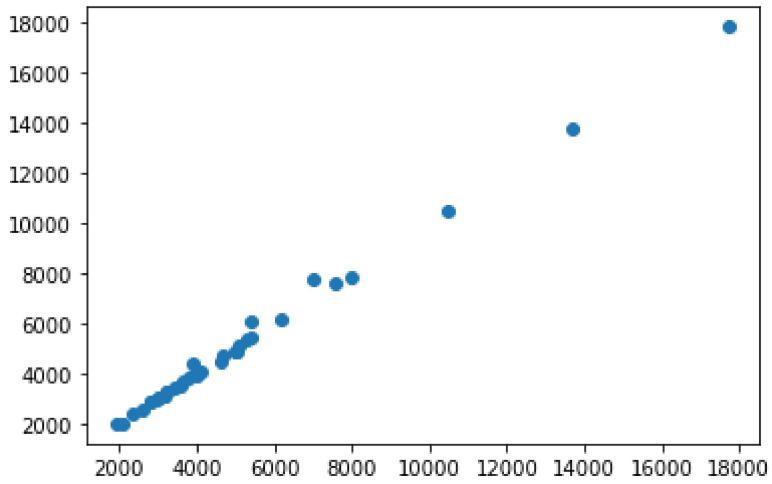
In [46]:
```
coeff = pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])
coeff
```

Out[46]:

	Co-efficient
From Home	0.979445
From Hashtags	0.996322
From Explore	1.008960
From Other	1.077264
Saves	0.226221
Comments	0.643832
Shares	-0.558899
Likes	-0.010052
Profile Visits	0.145518
Follows	-0.440303

In [47]:
```
prediction = lr.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[47]:
```
<matplotlib.collections.PathCollection at 0x1e5b066c7f0>
```



In [48]:
```
print(lr.score(x_test,y_test))
```

```
0.9959710875840057
```