

EDA-Exploitory Data Analysis -Instagram Dataset

import libraries

```
In [1]: import pandas as pd
import numpy as np
import matplotlib as pp
```

import dataset

```
In [2]: data=pd.read_csv(r"E:\154\5_Instagram data.csv")
```

```
In [3]: display(data)
```

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
0	3920	2586	1028	619	56	98	9	5	162	35	2
1	5394	2727	1838	1174	78	194	7	14	224	48	10
2	4021	2085	1188	0	533	41	11	1	131	62	12
3	4528	2700	621	932	73	172	10	7	213	23	8
4	2518	1704	255	279	37	96	5	4	123	8	0
...

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
114	13700	5185	3041	5352	77	573	2	38	373	73	80
115	5731	1923	1368	2266	65	135	4	1	148	20	18
116	4139	1133	1538	1367	33	36	0	1	92	34	10
117	32695	11815	3147	17414	170	1095	2	75	549	148	214
118	36919	13473	4176	16444	2547	653	5	26	443	611	228

119 rows × 13 columns

To display top 10 rows

In [4]:

data.head()

Out[4]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
0	3920	2586	1028	619	56	98	9	5	162	35	2
1	5394	2727	1838	1174	78	194	7	14	224	48	10

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
2	4021	2085	1188	0	533	41	11	1	131	62	12
3	4528	2700	621	932	73	172	10	7	213	23	8
4	2518	1704	255	279	37	96	5	4	123	8	0

To display last 5 rows

In [5]:

```
data.tail()
```

Out[5]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
114	13700	5185	3041	5352	77	573	2	38	373	73	80
115	5731	1923	1368	2266	65	135	4	1	148	20	18
116	4139	1133	1538	1367	33	36	0	1	92	34	10
117	32695	11815	3147	17414	170	1095	2	75	549	148	214

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

In [6]:

data.dtypes

Out[6]:

Impressions	int64
From Home	int64
From Hashtags	int64
From Explore	int64
From Other	int64
Saves	int64
Comments	int64
Shares	int64
Likes	int64
Profile Visits	int64
Follows	int64
Caption	object
Hashtags	object
dtype:	object

To view statistical summary

In [7]:

data.describe()

Out[7]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	
count	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000
mean	5703.991597	2475.789916	1887.512605	1078.100840	171.092437	153.310924	6.663866	
std	4843.780105	1489.386348	1884.361443	2613.026132	289.431031	156.317731	3.544576	
min	1941.000000	1133.000000	116.000000	0.000000	9.000000	22.000000	0.000000	
25%	3467.000000	1945.000000	726.000000	157.500000	38.000000	65.000000	4.000000	
50%	4289.000000	2207.000000	1278.000000	326.000000	74.000000	109.000000	6.000000	
75%	6138.000000	2602.500000	2363.500000	689.500000	196.000000	169.000000	8.000000	
max	36919.000000	13473.000000	11817.000000	17414.000000	2547.000000	1095.000000	19.000000	

To Print no of elements

In [8]:

data.size

```
Out[8]: 1547
```

```
In [9]: data.ndim
```

```
Out[9]: 2
```

To print no of rows and columns

```
In [10]: data.shape
```

```
Out[10]: (119, 13)
```

To find missing values

```
In [11]: data.isna()
```

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...
114	False	False	False	False	False	False	False	False	False	False	False
115	False	False	False	False	False	False	False	False	False	False	False
116	False	False	False	False	False	False	False	False	False	False	False
117	False	False	False	False	False	False	False	False	False	False	False
118	False	False	False	False	False	False	False	False	False	False	False

119 rows × 13 columns

To drop null values with constatns

```
In [12]: data.fillna(5)
```

Out[12]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
0	3920	2586	1028	619	56	98	9	5	162	35	2
1	5394	2727	1838	1174	78	194	7	14	224	48	10
2	4021	2085	1188	0	533	41	11	1	131	62	12
3	4528	2700	621	932	73	172	10	7	213	23	8
4	2518	1704	255	279	37	96	5	4	123	8	0
...
114	13700	5185	3041	5352	77	573	2	38	373	73	80
115	5731	1923	1368	2266	65	135	4	1	148	20	18
116	4139	1133	1538	1367	33	36	0	1	92	34	10
117	32695	11815	3147	17414	170	1095	2	75	549	148	214

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
118	36919	13473	4176	16444	2547	653	5	26	443	611	228

119 rows × 13 columns

In [13]:

`data.dropna()`

Out[13]:

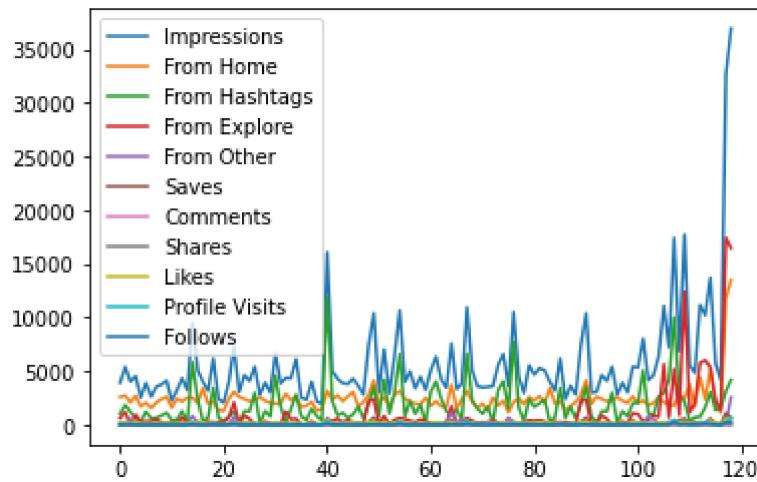
	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
0	3920	2586	1028	619	56	98	9	5	162	35	2
1	5394	2727	1838	1174	78	194	7	14	224	48	10
2	4021	2085	1188	0	533	41	11	1	131	62	12
3	4528	2700	621	932	73	172	10	7	213	23	8
4	2518	1704	255	279	37	96	5	4	123	8	0
...
114	13700	5185	3041	5352	77	573	2	38	373	73	80

119 rows × 13 columns

Line Plot

In [14]: `data.plot.line()`

Out[14]: <AxesSubplot:>

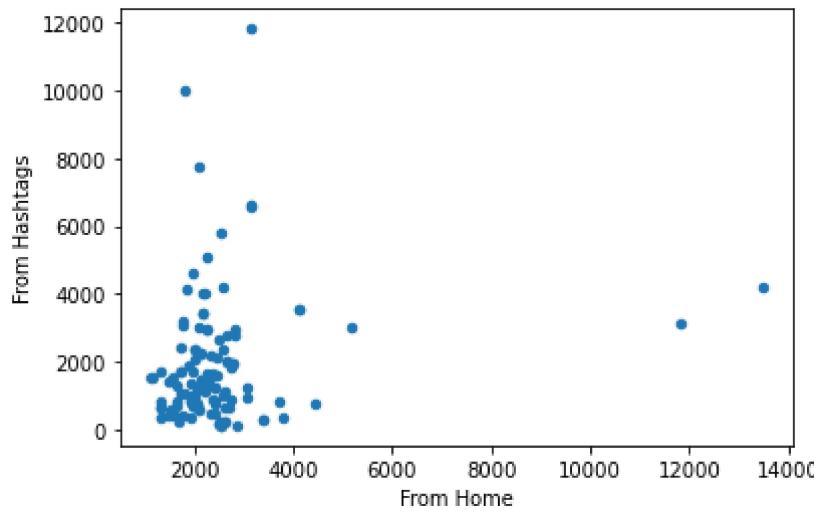


Scatter Plot

In [15]:

```
data.plot.scatter(x='From Home',y='From Hashtags')
```

Out[15]: <AxesSubplot:xlabel='From Home', ylabel='From Hashtags'>

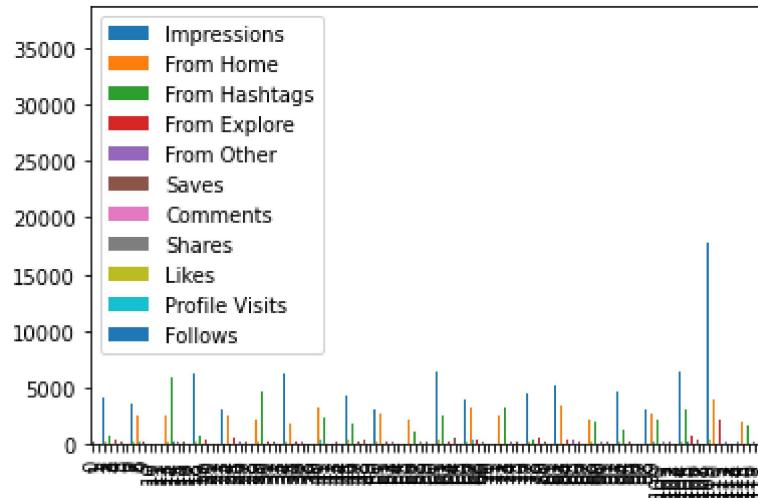


Bar Chart

In [16]:

```
data.plot.bar()
```

Out[16]: <AxesSubplot:>

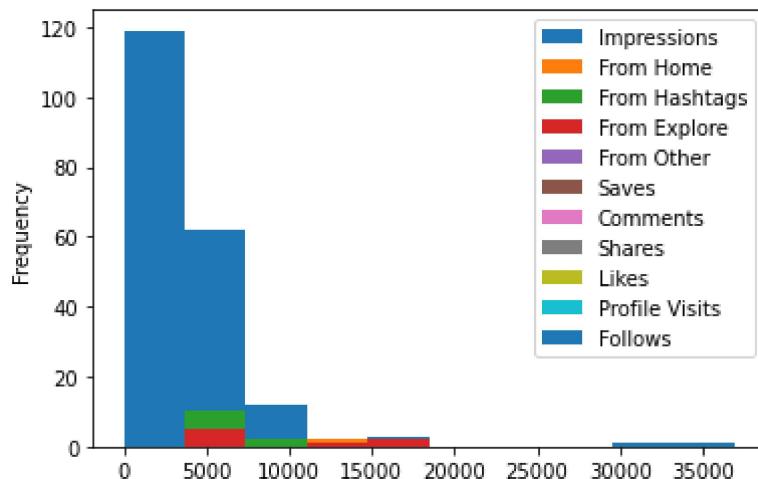


Histogram

In [17]:

```
data.plot.hist()
```

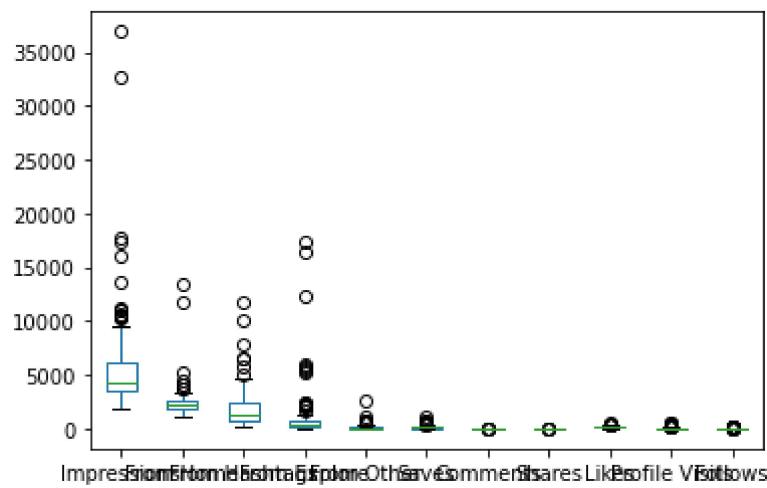
Out[17]: <AxesSubplot:ylabel='Frequency'>



Box Plot

```
In [18]: data.plot.box()
```

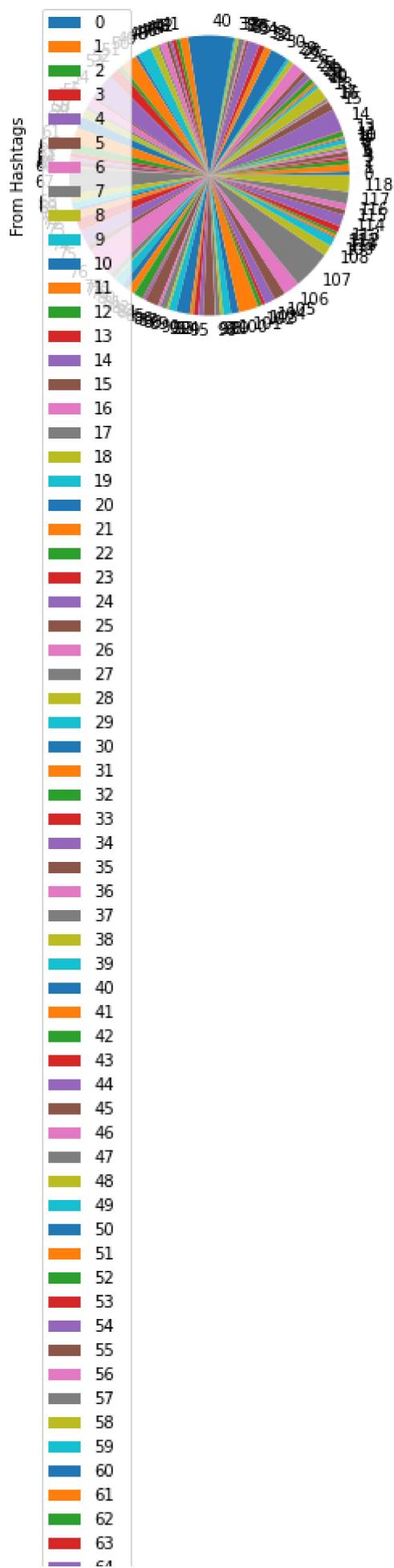
```
Out[18]: <AxesSubplot:>
```

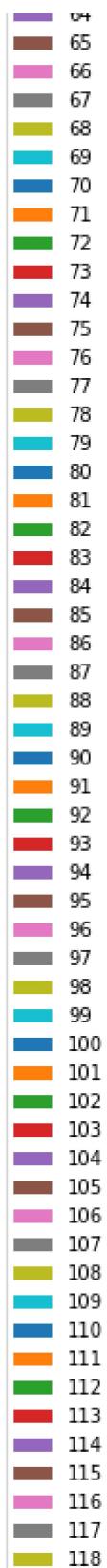


Pie Chart

```
In [19]: data.plot.pie(y="From Hashtags")
```

```
Out[19]: <AxesSubplot:ylabel='From Hashtags'>
```



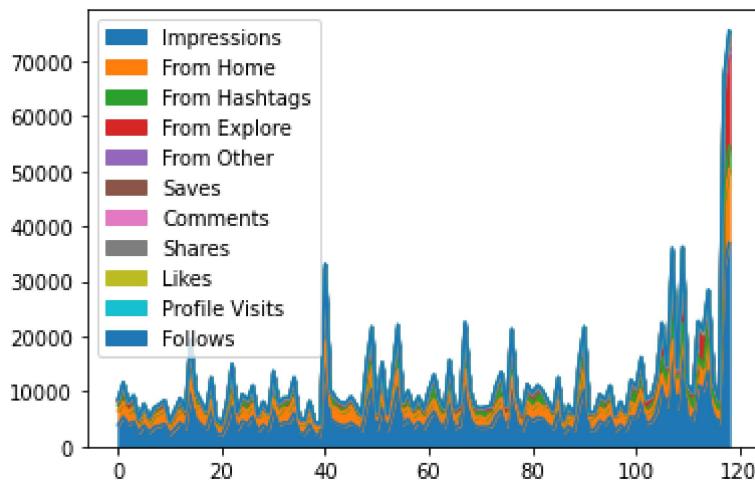


Area

In [20]:

```
data.plot.area()
```

Out[20]: <AxesSubplot:>



To Find Mean

In [21]: `data.mean()`

Out[21]:

Impressions	5703.991597
From Home	2475.789916
From Hashtags	1887.512605
From Explore	1078.100840
From Other	171.092437
Saves	153.310924
Comments	6.663866
Shares	9.361345
Likes	173.781513
Profile Visits	50.621849
Follows	20.756303

dtype: float64

To Find Median

In [22]: `data.median()`

Out[22]:

Impressions	4289.0
From Home	2207.0
From Hashtags	1278.0
From Explore	326.0
From Other	74.0
Saves	109.0
Comments	6.0
Shares	6.0
Likes	151.0
Profile Visits	23.0
Follows	8.0

dtype: float64

To Find Mode

In [23]: `data.mode()`

Out[23]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows	Created At
0	5394.0	1975.0		116	45.0	34.0	40.0		6.0	3.0	114.0	19.0
1	NaN	NaN		201	84.0	NaN	135.0		NaN	NaN	151.0	21.0
2	NaN	NaN		278	NaN	NaN	144.0		NaN	NaN	NaN	NaN
3	NaN	NaN		362	NaN	NaN	NaN		NaN	NaN	NaN	NaN
4	NaN	NaN		411	NaN	NaN	NaN		NaN	NaN	NaN	NaN
5	NaN	NaN		583	NaN	NaN	NaN		NaN	NaN	NaN	NaN
6	NaN	NaN		655	NaN	NaN	NaN		NaN	NaN	NaN	NaN
7	NaN	NaN		707	NaN	NaN	NaN		NaN	NaN	NaN	NaN
8	NaN	NaN		771	NaN	NaN	NaN		NaN	NaN	NaN	NaN
9	NaN	NaN		794	NaN	NaN	NaN		NaN	NaN	NaN	NaN
10	NaN	NaN		1248	NaN	NaN	NaN		NaN	NaN	NaN	NaN
11	NaN	NaN		1260	NaN	NaN	NaN		NaN	NaN	NaN	NaN
12	NaN	NaN		1278	NaN	NaN	NaN		NaN	NaN	NaN	NaN
13	NaN	NaN		1693	NaN	NaN	NaN		NaN	NaN	NaN	NaN
14	NaN	NaN		1938	NaN	NaN	NaN		NaN	NaN	NaN	NaN
15	NaN	NaN		2351	NaN	NaN	NaN		NaN	NaN	NaN	NaN
16	NaN	NaN		2975	NaN	NaN	NaN		NaN	NaN	NaN	NaN
17	NaN	NaN		3450	NaN	NaN	NaN		NaN	NaN	NaN	NaN
18	NaN	NaN		3551	NaN	NaN	NaN		NaN	NaN	NaN	NaN



Describe

In [24]:

data.describe()

Out[24]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows	Created At
count	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000
mean	5703.991597	2475.789916	1887.512605	1078.100840	171.092437	153.310924	6.663866	1.0	1.0	1.0	1.0	1.0

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments
std	4843.780105	1489.386348	1884.361443	2613.026132	289.431031	156.317731	3.544576
min	1941.000000	1133.000000	116.000000	0.000000	9.000000	22.000000	0.000000
25%	3467.000000	1945.000000	726.000000	157.500000	38.000000	65.000000	4.000000
50%	4289.000000	2207.000000	1278.000000	326.000000	74.000000	109.000000	6.000000
75%	6138.000000	2602.500000	2363.500000	689.500000	196.000000	169.000000	8.000000
max	36919.000000	13473.000000	11817.000000	17414.000000	2547.000000	1095.000000	19.000000

Sum

In [25]: `data.sum()`

Out[25]:

Impressions	678775
From Home	294619
From Hashtags	224614
From Explore	128294
From Other	20360
Saves	18244
Comments	793
Shares	1114
Likes	20680
Profile Visits	6024
Follows	2470
Caption	Here are some of the most important data visua...
Hashtags	#finance◆#money◆#business◆#investing◆#investme...
dtype: object	

Cumulative Sum

In [26]: `data.cumsum()`

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
0	3920	2586	1028	619	56	98	9	5	162	35	2
1	9314	5313	2866	1793	134	292	16	19	386	83	12
2	13335	7398	4054	1793	667	333	27	20	517	145	24

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
3	17863	10098	4675	2725	740	505	37	27	730	168	32
4	20381	11802	4930	3004	777	601	42	31	853	176	32
...
114	599291	266275	214385	90803	17545	16325	782	1011	19448	5211	2000
115	605022	268198	215753	93069	17610	16460	786	1012	19596	5231	2018
116	609161	269331	217291	94436	17643	16496	786	1013	19688	5265	2028
117	641856	281146	220438	111850	17813	17591	788	1088	20237	5413	2242
118	678775	294619	224614	128294	20360	18244	793	1114	20680	6024	2470

119 rows × 13 columns

Minimum Values

In [27]: `data.min()`

```
Out[27]: Impressions          1941
From Home           1133
From Hashtags       116
From Explore         0
From Other            9
Saves                  22
Comments                 0
Shares                  0
Likes                   72
Profile Visits        4
Follows                  0
Caption      170 Python Projects with Source Code solved an...
Hashtags      #career◆#job◆#jobs◆#jobsearch◆#education◆#busi...
dtype: object
```

Maximum Values

In [28]: `data.max()`

```
Out[28]: Impressions          36919
From Home           13473
From Hashtags       11817
From Explore         17414
From Other            2547
Saves                  1095
Comments                 19
Shares                  75
Likes                   549
Profile Visits        611
Follows                  260
Caption      You must have seen the news divided into categ...
Hashtags      #timeseries◆#time◆#statistics◆#datascience◆#bi...
dtype: object
```

Correlation

In [29]: `from scipy.stats import spearmanr
print(spearmanr(data['From Home'], data['From Hashtags']))`

```
SpearmanResult(correlation=0.11752786942921449, pvalue=0.203031655807403)
```

Covariance

In [30]: `from scipy.stats import pearsonr
print(pearsonr(data['From Home'], data['From Hashtags']))`

```
(0.17751565433098784, 0.053434143091160374)
```

Count

```
In [31]: data.count()
```

```
Out[31]: Impressions      119
From Home          119
From Hashtags     119
From Explore       119
From Other         119
Saves              119
Comments           119
Shares              119
Likes               119
Profile Visits    119
Follows             119
Caption             119
Hashtags           119
dtype: int64
```