Deena 20104016

Basic Analysis using Numpy and Pandas

Import Libraries

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from numpy import cov
from scipy.stats import pearsonr
from scipy.stats import spearmanr
```

Importing Dataset

In [2]:
```
df=pd.read_csv("5_Instagram data.csv")
df
```

Out[2]:

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

119 rows × 13 columns

To display first 10 rows

In [3]:
```
df.head(10)
```

Out[3]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Foll
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	
5	3884	2046	1214	329	43	74	7	10	144	9	
6	2621	1543	599	333	25	22	5	1	76	26	
7	3541	2071	628	500	60	135	4	9	124	12	
8	3749	2384	857	248	49	155	6	8	159	36	

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Foll
9	4115	2609	1104	178	46	122	6	3	191	31	

To display last 5 rows

In [4]:
```
df.tail(5)
```

Out[4]:

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

Satistical Summary

In [5]:
```
df.mean()
```

Out[5]:
```
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
```

In [6]:
```
df.median()
```

Out[6]:
```
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
```

In [7]:
```
df.mode()
```

Out[7]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Fo
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	

In [8]:
```
df.cumsum()
```

Out[8]:

	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	9314	5313	2866	1793	134	292	16	19	386	83
2	13335	7398	4054	1793	667	333	27	20	517	145
3	17863	10098	4675	2725	740	505	37	27	730	168
4	20381	11802	4930	3004	777	601	42	31	853	176
...	...	...	...	...	...	...	...	...	...	...
114	599291	266275	214385	90803	17545	16325	782	1011	19448	5211
115	605022	268198	215753	93069	17610	16460	786	1012	19596	5231
116	609161	269331	217291	94436	17643	16496	786	1013	19688	5265
117	641856	281146	220438	111850	17813	17591	788	1088	20237	5413

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits
118	678775	294619	224614	128294	20360	18244	793	1114	20680	6024

119 rows × 13 columns

In [9]:
```
df.count()
```

Out[9]:
```
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
```

In [10]:
```
df.max()
```

Out[10]:
```
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
```

In [11]:
```
df.min()
```

Out[11]:
```
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
```

In [12]:
```
df.sum()
```

Out[12]:
```
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
```

In [13]:
```
cov(df['From Home'],df['From Hashtags'])
```

Out[13]:
```
array([[2218271.69277881,  498205.17639937],
       [ 498205.17639937, 3550818.04856858]])
```

In [14]:
```
df.describe()
```

Out[14]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comm
count	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.00
mean	5703.991597	2475.789916	1887.512605	1078.100840	171.092437	153.310924	6.66
std	4843.780105	1489.386348	1884.361443	2613.026132	289.431031	156.317731	3.54
min	1941.000000	1133.000000	116.000000	0.000000	9.000000	22.000000	0.00
25%	3467.000000	1945.000000	726.000000	157.500000	38.000000	65.000000	4.00
50%	4289.000000	2207.000000	1278.000000	326.000000	74.000000	109.000000	6.00
75%	6138.000000	2602.500000	2363.500000	689.500000	196.000000	169.000000	8.00
max	36919.000000	13473.000000	11817.000000	17414.000000	2547.000000	1095.000000	19.00

pearsonr

In [15]:
```
pearsonr(df['From Home'],df['From Hashtags'])
```

Out[15]:
```
(0.17751565433098784, 0.053434143091160374)
```

spearmanr

In [16]:
```
spearmanr(df['From Home'],df['From Hashtags'])
```

Out[16]:
```
SpearmanrResult(correlation=0.11752786942921449, pvalue=0.203031655807403)
```

To find shape and size

In [17]:
```
df.shape
```

Out[17]:
```
(119, 13)
```

In [18]:
```
df.size
```

Out[18]:
```
1547
```

To fill the null values

In [19]:
```
df.isna()
```

Out[19]:

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

119 rows × 13 columns

To fill missing values

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

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
...	...	...	...	...	...	...	...	...	...	...

coloumns

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

Out[21]:
```
Index(['Impressions', 'From Home', 'From Hashtags', 'From Explore',
       'From Other', 'Saves', 'Comments', 'Shares', 'Likes', 'Profile Visit
s',
       'Follows', 'Caption', 'Hashtags'],
      dtype='object')
```

to print a particular coloumn

In [22]:
```
data=df[['From Home', 'From Hashtags']]
data
```

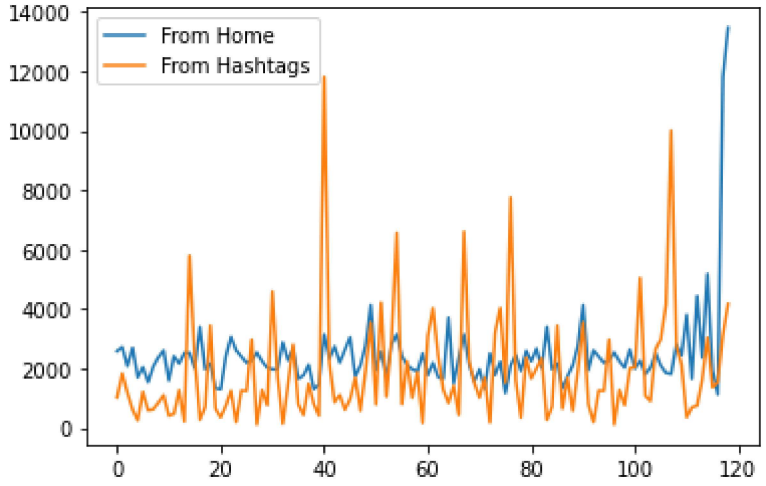
	From Home	From Hashtags
0	2586	1028
1	2727	1838
2	2085	1188
3	2700	621
4	1704	255
...	...	...
114	5185	3041
115	1923	1368
116	1133	1538
117	11815	3147
118	13473	4176

119 rows × 2 columns

line plot

In [23]:
```
data.plot.line()
```

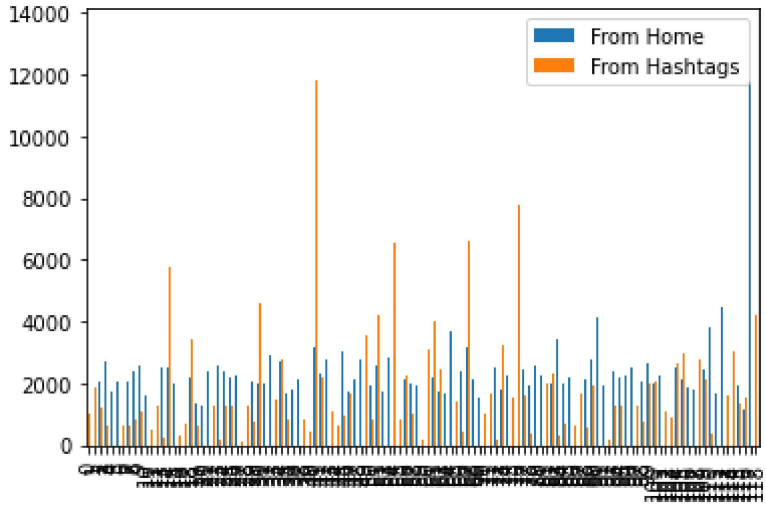
Out[23]: <AxesSubplot:>



bar plot

In [24]:
```
data.plot.bar()
```

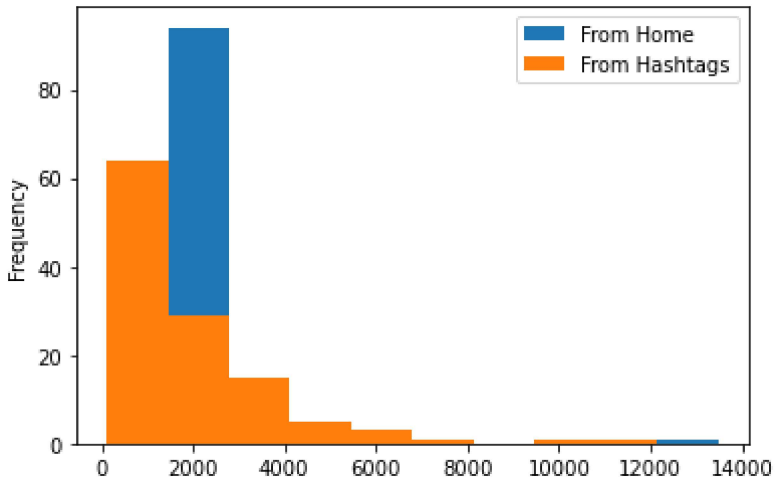
Out[24]: <AxesSubplot:>



hist plot

In [25]:
```
data.plot.hist()
```

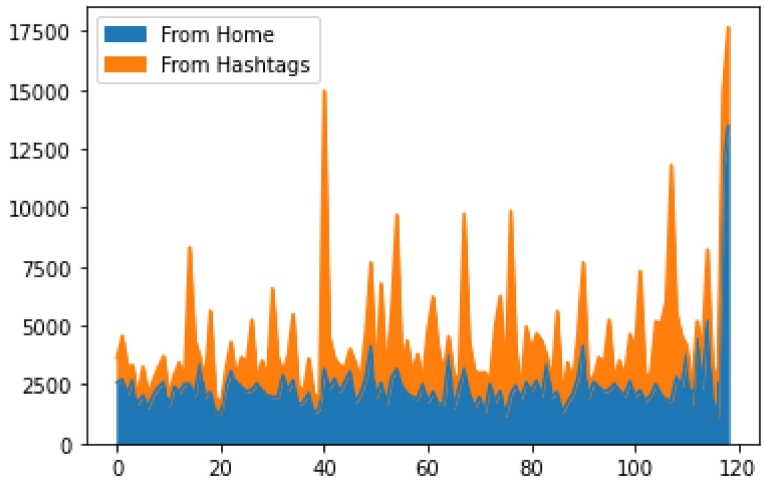
Out[25]:
```
<AxesSubplot:ylabel='Frequency'>
```



Area plot

In [26]:
```
data.plot.area()
```

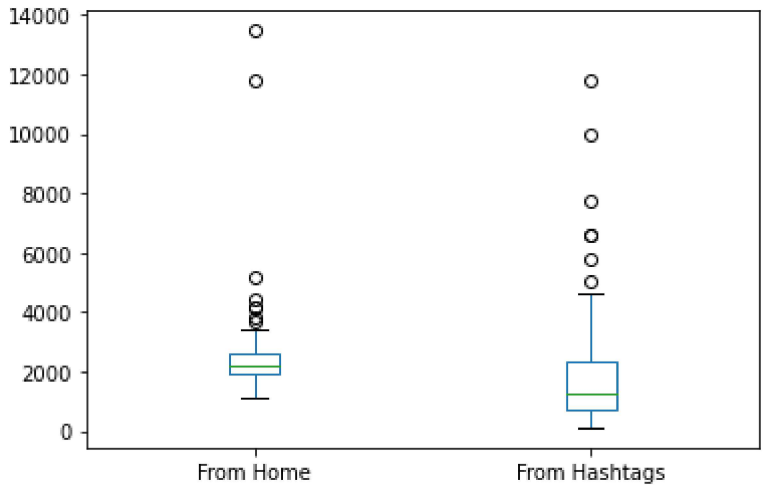
Out[26]:
```
<AxesSubplot:>
```



Box plot

In [27]:
```
data.plot.box()
```

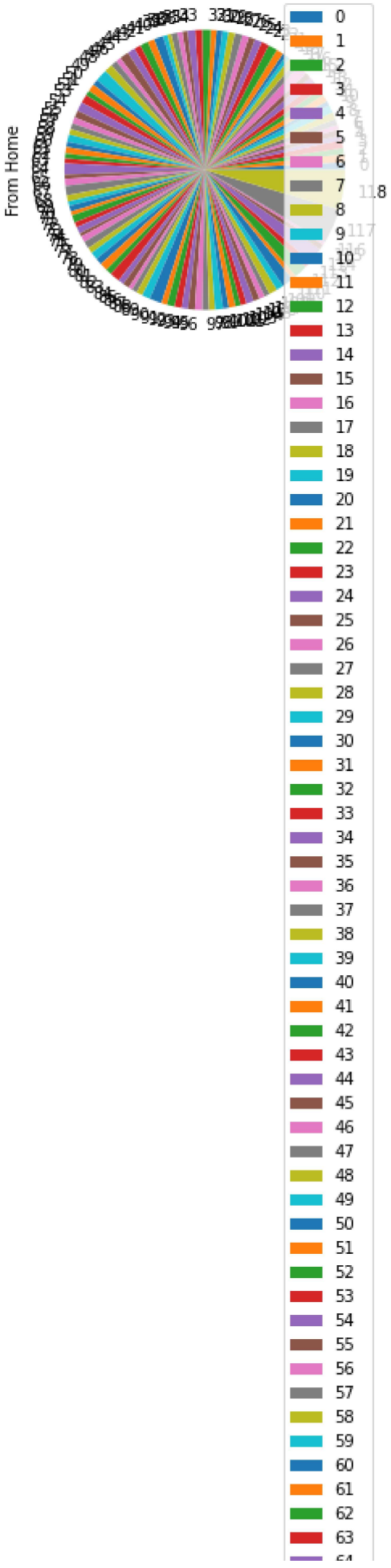
Out[27]: <AxesSubplot:>

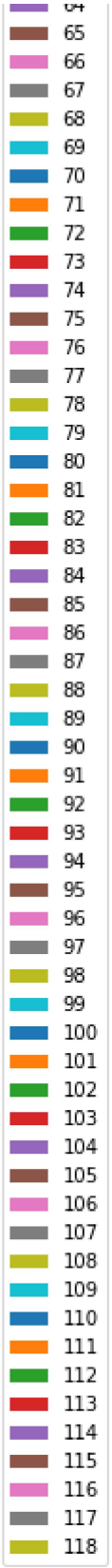


pie plot

In [28]:
```
data.plot.pie(y='From Home')
```

Out[28]:
```
<AxesSubplot:ylabel='From Home'>
```


From Home
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63


64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118

In [29]:
```
data.plot.scatter(x= 'From Hashtags',y='From Home')
```

Out[29]:
```
<AxesSubplot:xlabel='From Hashtags', ylabel='From Home'>
```