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
```

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	

Out[3]:

	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.describe()`

Out[5]:

	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

To find shape and size

In [6]: `df.shape`

Out[6]: `(119, 13)`

In [7]: `df.size`

Out[7]: `1547`

To fill the null values

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

Out[8]:

	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 [9]: `df.dropna()`

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

◀ ▬▬▬▬▬▬▬▬▬▬▬ ▶

coloumns

```
In [10]: df.columns
```

```
Out[10]: 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 [11]: data=df[['From Home', 'From Hashtags']]
         data
```
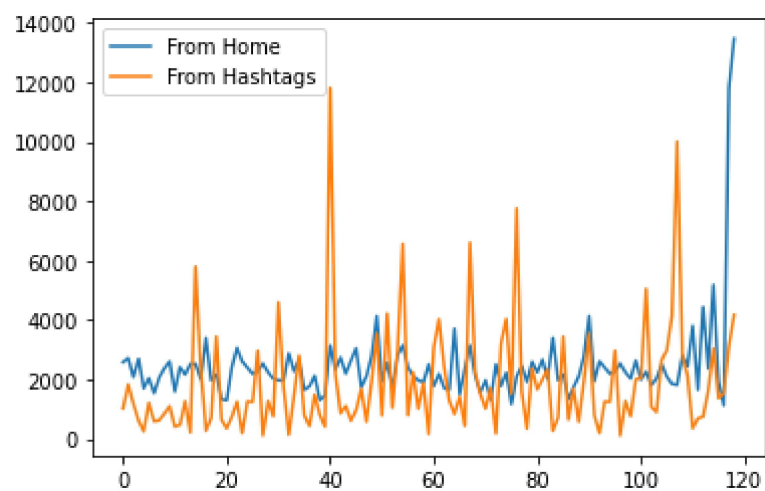
Out[11]:

	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

line plot

In [12]: `data.plot.line()`

Out[12]: `<AxesSubplot:>`

bar plot

In [13]: `data.plot.bar()`

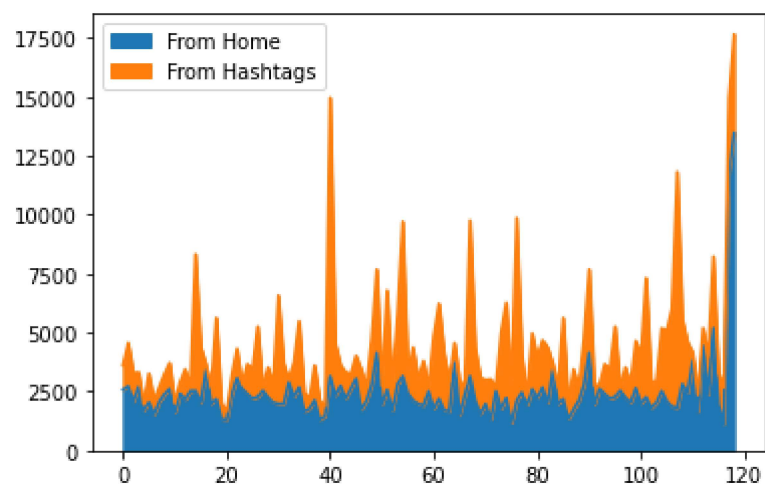Out[13]: `<AxesSubplot:>`
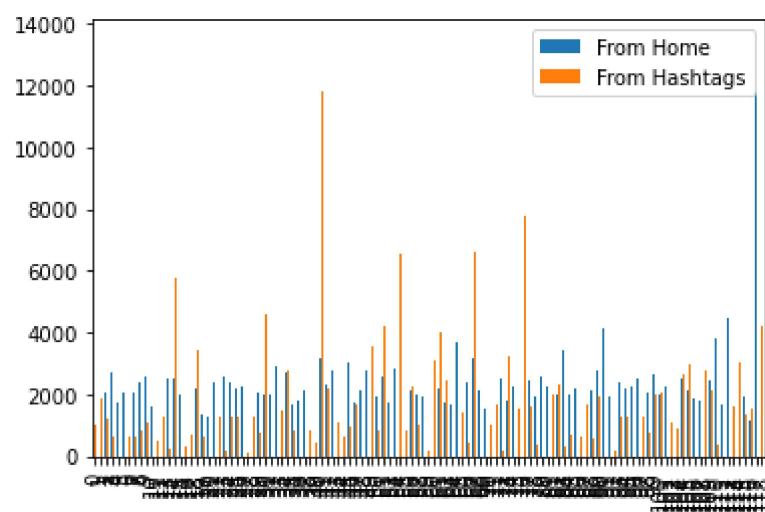
hist plot

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

Out[14]: <AxesSubplot:ylabel='Frequency'>
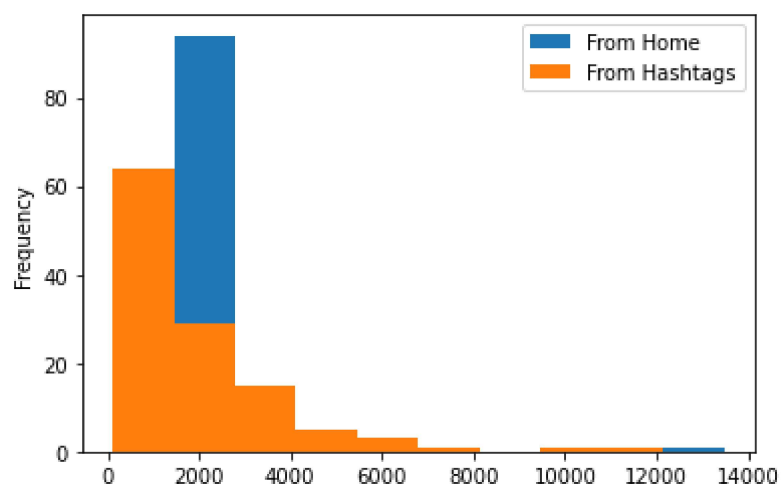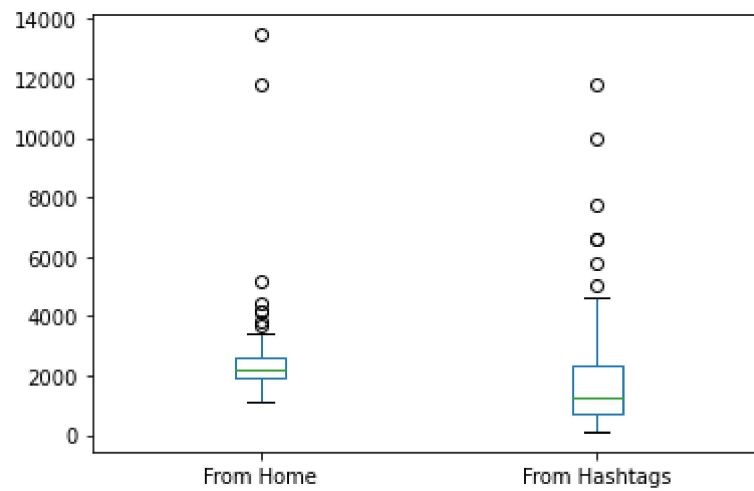

Area plot

In [15]: `data.plot.area()`

Out[15]: <AxesSubplot:>

Box plot

In [16]: `data.plot.box()`

Out[16]: `<AxesSubplot:>`

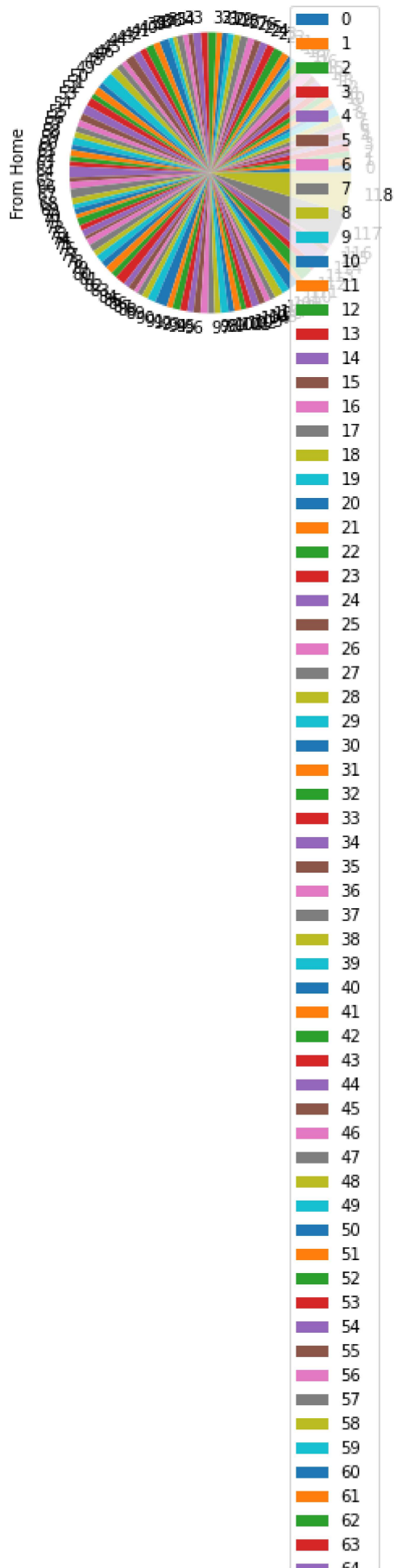
pie plot

In [17]: `data.plot.pie(y='From Home')`

Out[17]: `<AxesSubplot:ylabel='From Home'>`

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

Out[18]: `<AxesSubplot:xlabel='From Hashtags', ylabel='From Home'>`