

# Basic Analysis using Numpy and Pandas

## insta Dataset

To import library

```
In [1]: import numpy as np
```

```
In [2]: import pandas as pd
```

To import dataset

```
In [3]: data=pd.read_csv(r"C:\Users\user\Downloads\insta.csv")  
data
```

Out[3]:

	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 get Top 20 record

In [4]: `data.head(20)`

Out[4]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Fo
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	
9	4115	2609	1104	178	46	122	6	3	191	31	
10	2218	1597	411	162	15	28	6	3	81	29	
11	3234	2414	476	185	75	122	8	14	151	15	
12	4344	2168	1274	673	40	119	7	11	162	8	

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Fo
13	3216	2524	212	201	223	121	5	5	142	20	
14	9453	2525	5799	208	794	100	6	10	294	181	
15	5055	2017	2351	298	108	101	7	11	159	17	
16	4002	3401	278	128	73	111	17	18	205	16	
17	3169	1979	707	341	32	106	8	1	121	21	
18	6168	2177	3450	153	296	82	6	6	151	77	
19	2407	1338	655	276	39	40	8	20	72	10	

To get last 20 record

In [5]: `data.tail(20)`



Out[5]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	F
99	5409	2643	2006	1068	230	393	10	27	275	38	
100	5323	2000	2053	988	74	152	6	13	192	42	
101	8001	2261	5055	300	172	83	8	7	203	92	
102	4150	1807	1085	463	792	74	4	2	145	75	
103	4609	2032	911	857	532	94	1	12	168	21	
104	6348	2517	2660	737	154	188	4	15	194	26	
105	11068	2099	2986	5634	122	214	7	8	250	39	
106	7231	1855	4156	703	309	73	8	3	171	74	
107	17396	1817	10008	5192	251	285	7	7	416	467	
108	6814	2816	2769	900	128	469	8	22	249	24	
109	17713	2449	2141	12389	561	504	3	23	308	70	

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	F
110	5563	3813	362	1135	76	149	5	8	163	22	
111	4842	1658	694	2036	310	55	6	4	86	46	
112	11149	4439	747	5762	53	273	4	13	210	61	
113	10206	2371	1624	6000	117	182	10	17	172	237	
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	

Statistical Analysis

```
In [6]: data.describe()
```

```
Out[6]:
```

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



To get row and column

```
In [7]: print(np.shape(data))
```

```
(119, 13)
```

Find Number of Elements

```
In [8]: np.size(data)
```

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

Find Missing values

```
In [9]: data.isna()
```

Out[9]:

	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 drop the missing values

```
In [10]: data.dropna(axis=1,how="any")
```

Out[10]:

	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

```
In [11]: data["From Home"]
```

```
Out[11]: 0      2586
          1      2727
          2      2085
          3      2700
          4      1704
          ...
         114     5185
         115     1923
         116     1133
         117    11815
         118    13473
          Name: From Home, Length: 119, dtype: int64
```

```
In [12]: data1=data[['From Home','Likes']]
          data1
```

```
Out[12]:
```

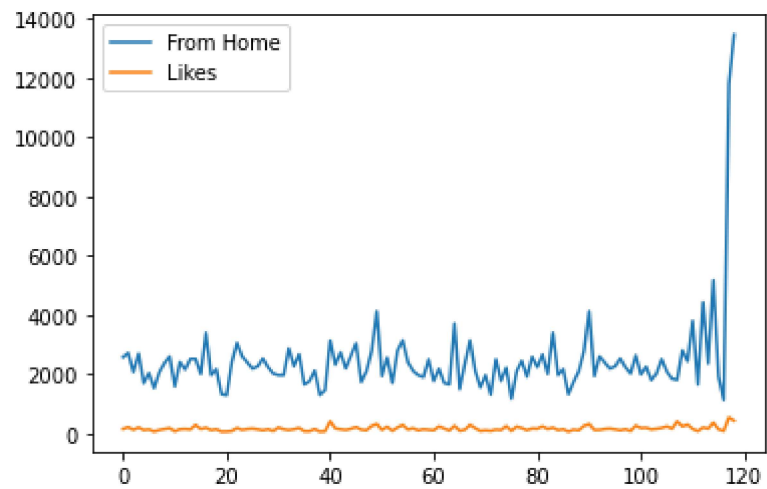
	From Home	Likes
0	2586	162
1	2727	224
2	2085	131
3	2700	213
4	1704	123
...	...	...
114	5185	373
115	1923	148
116	1133	92
117	11815	549
118	13473	443

119 rows × 2 columns

```
In [13]: import matplotlib as mp
```

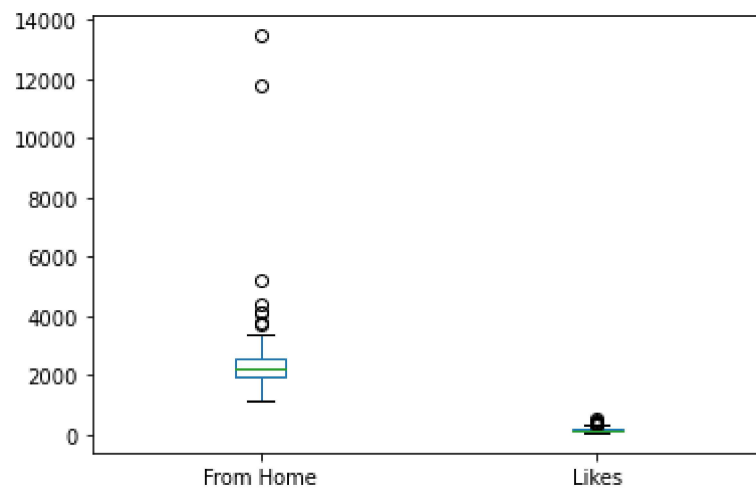
```
In [14]: data1.plot.line()
```

```
Out[14]: <AxesSubplot:>
```



```
In [15]: data1.plot.box()
```

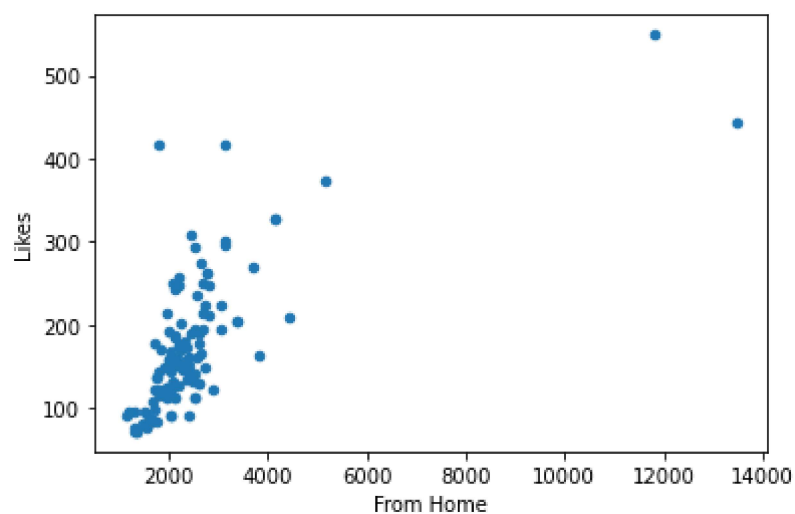
```
Out[15]: <AxesSubplot:>
```





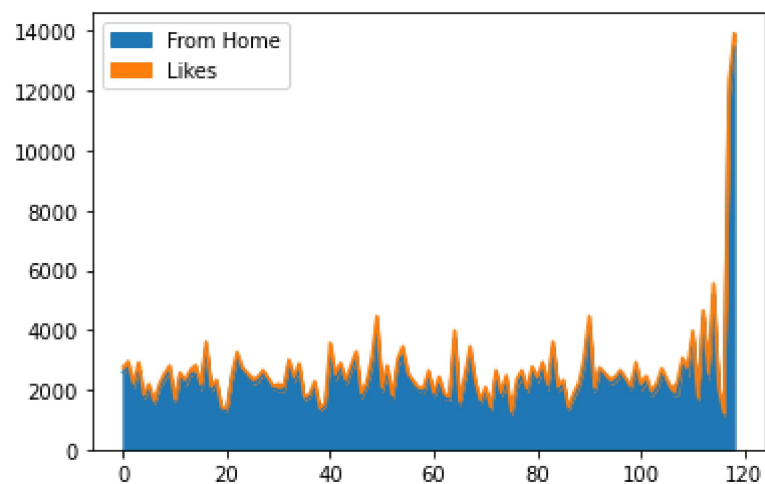
```
In [16]: data1.plot.scatter(x='From Home',y='Likes')
```

```
Out[16]: <AxesSubplot:xlabel='From Home', ylabel='Likes'>
```



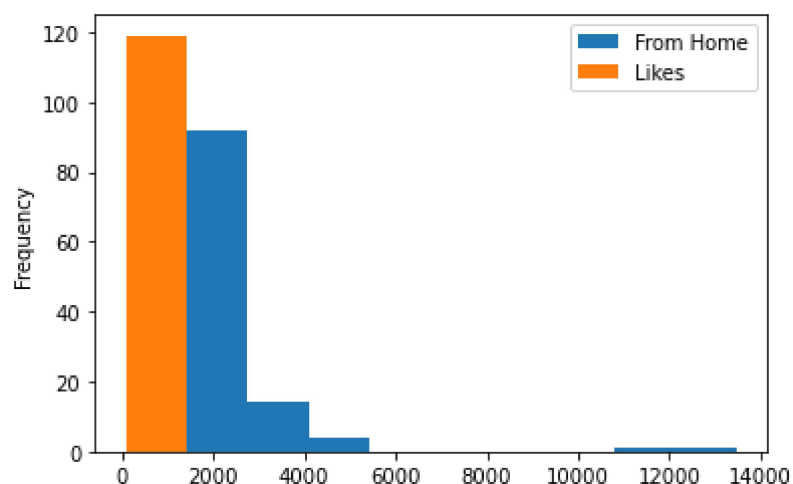
```
In [17]: data1.plot.area()
```

```
Out[17]: <AxesSubplot:>
```



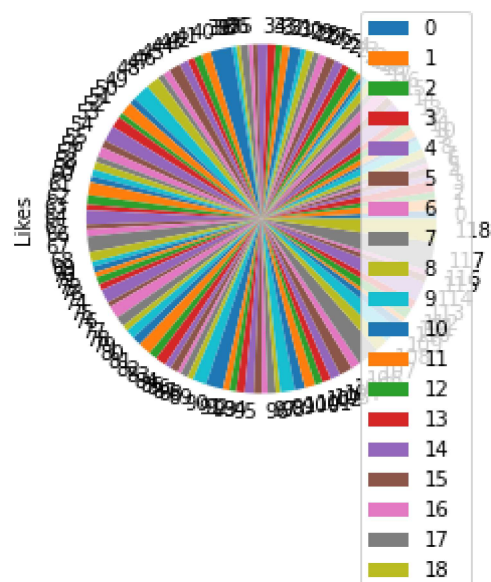
```
In [18]: data1.plot.hist()
```

```
Out[18]: <AxesSubplot:ylabel='Frequency'>
```



```
In [19]: data1.plot.pie(y="Likes")
```

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



```
In [ ]:
```