Basic Analysis using numpy and pandas

Instagram dataset

To import library

In [1]:

```
import numpy as np
import pandas as pd
```

To import dataset

In [2]:

```
d=pd.read_csv(r"C:\Users\user\Downloads\5_Instagram data.csv")
d
```

Out[2]:

	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
118	36919	13473	4176	16444	2547	653	5	26	443	611

119 rows × 13 columns

To get top 10 record

In [3]:

```
d.head(10)
```

Out[3]:

	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
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

To get last 10

In [4]:

```
d.tail(10)
```

Out[4]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits
109	17713	2449	2141	12389	561	504	3	23	308	70
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	61

To describe statistics Analysis

In [5]:

```
d.describe()
```

Out[5]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	C
count	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	11
mean	5703.991597	2475.789916	1887.512605	1078.100840	171.092437	153.310924	
std	4843.780105	1489.386348	1884.361443	2613.026132	289.431031	156.317731	
min	1941.000000	1133.000000	116.000000	0.000000	9.000000	22.000000	
25%	3467.000000	1945.000000	726.000000	157.500000	38.000000	65.000000	
50%	4289.000000	2207.000000	1278.000000	326.000000	74.000000	109.000000	
75%	6138.000000	2602.500000	2363.500000	689.500000	196.000000	169.000000	
max	36919.000000	13473.000000	11817.000000	17414.000000	2547.000000	1095.000000	1

To get rows and columns

In [6]:

```
np.shape(d)
```

Out[6]:

```
(119, 13)
```

To get number of elements

In [7]:

```
np.size(d)
```

Out[7]:

```
1547
```

To get the missing value

In [8]:

```
d.isna()
```

Out[8]:

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

In [9]:

```
d.dropna(axis=1,how='any')
```

Out[9]:

	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
118	36919	13473	4176	16444	2547	653	5	26	443	611

119 rows × 13 columns

In [10]:

```
d["Impressions"]
```

Out[10]:

```
0       3920
1       5394
2       4021
3       4528
4       2518
       ...
114    13700
115     5731
116     4139
117    32695
118    36919
Name: Impressions, Length: 119, dtype: int64
```

In [11]:

```
data=d[['Impressions','From Other']]
data
```

Out[11]:

	Impressions	From Other
0	3920	56
1	5394	78
2	4021	533
3	4528	73
4	2518	37
...	...	...
114	13700	77
115	5731	65
116	4139	33
117	32695	170
118	36919	2547

119 rows × 2 columns

In [12]:

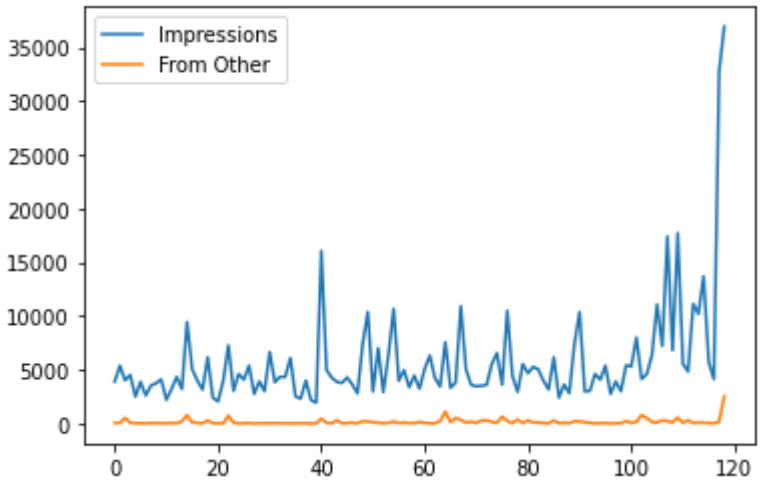
```
import matplotlib.pyplot as pp
```

In [13]:

```
data.plot.line()
```

Out[13]:

<AxesSubplot:>

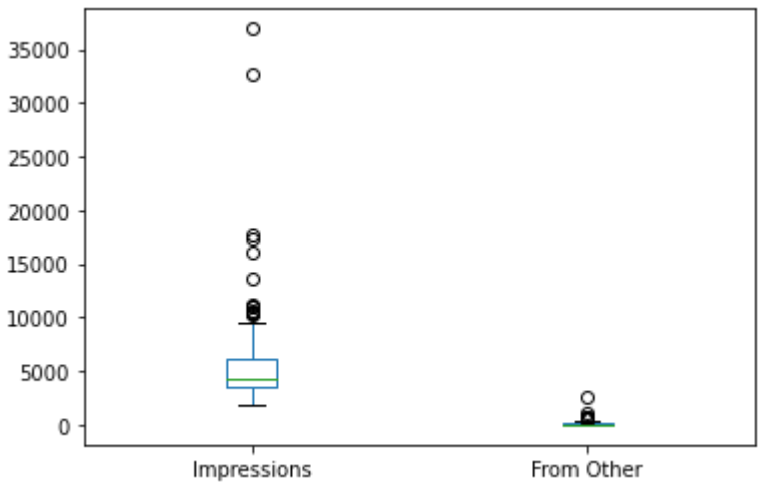


In [14]:

```
data.plot.box()
```

Out[14]:

<AxesSubplot:>

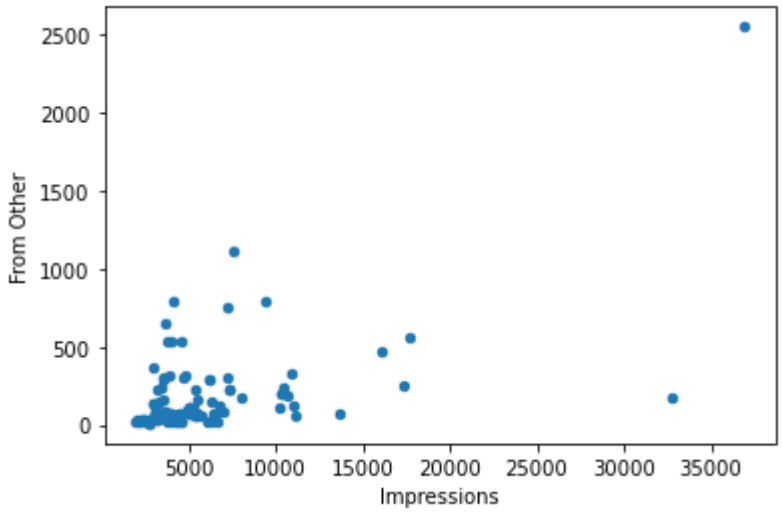


In [15]:

```
data.plot.scatter(x="Impressions",y="From Other")
```

Out[15]:

```
<AxesSubplot:xlabel='Impressions', ylabel='From Other'>
```

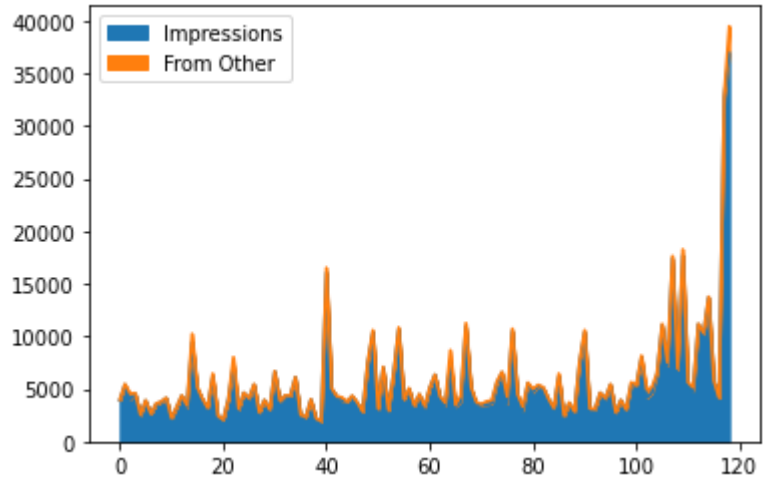


In [16]:

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

Out[16]:

```
<AxesSubplot:>
```

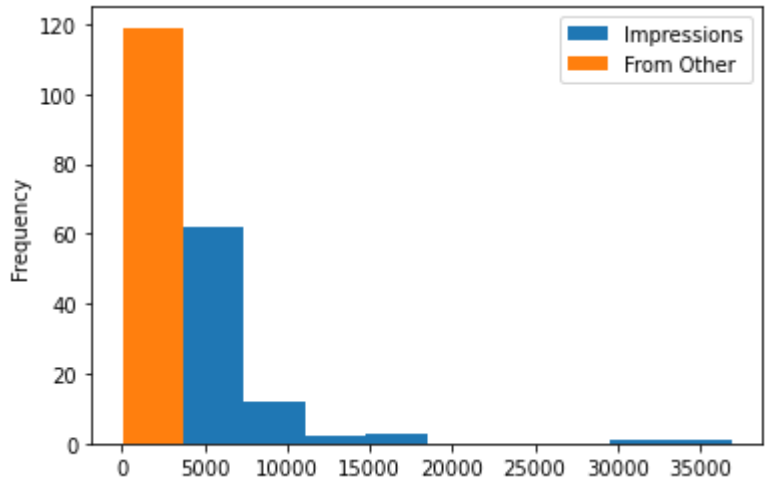


In [17]:

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

Out[17]:

```
<AxesSubplot:ylabel='Frequency'>
```

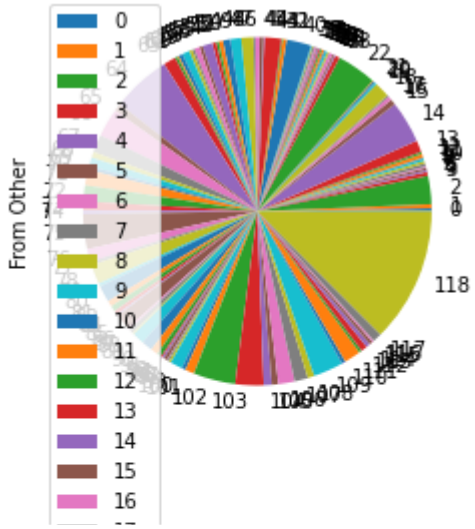


In [18]:

```
d.plot.pie(y="From Other")
```

Out[18]:

```
<AxesSubplot:ylabel='From Other'>
```



In [ ]: