Basic Analysis using Numpy and Pandas

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\3_Fitness-1.csv")
```

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

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	A	5.62%	7.73%	6.16%	75
1	B	4.21%	17.27%	19.21%	160
2	C	9.83%	11.60%	5.17%	101
3	D	2.81%	21.91%	7.88%	127
4	E	25.28%	10.57%	11.82%	179
5	F	8.15%	16.24%	18.47%	167
6	G	18.54%	8.76%	17.49%	171
7	H	25.56%	5.93%	13.79%	170
8	Grand Total	100.00%	100.00%	100.00%	1150

To display top 10 rows

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

Out[4]:

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	A	5.62%	7.73%	6.16%	75
1	B	4.21%	17.27%	19.21%	160
2	C	9.83%	11.60%	5.17%	101
3	D	2.81%	21.91%	7.88%	127
4	E	25.28%	10.57%	11.82%	179

To display last 5 rows

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

Out[5]:

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
4	E	25.28%	10.57%	11.82%	179
5	F	8.15%	16.24%	18.47%	167
6	G	18.54%	8.76%	17.49%	171
7	H	25.56%	5.93%	13.79%	170
8	Grand Total	100.00%	100.00%	100.00%	1150

In [6]:
```
data.dtypes
```

Out[6]:
```
Row Labels            object
Sum of Jan            object
Sum of Feb            object
Sum of Mar            object
Sum of Total Sales     int64
dtype: object
```

To view statistical summary

In [7]:
```
data.describe()
```

Out[7]:

	Sum of Total Sales
count	9.000000
mean	255.555556
std	337.332963
min	75.000000
25%	127.000000
50%	167.000000
75%	171.000000
max	1150.000000

To Print no of elements

In [8]:
```
data.size
```

Out[8]: 45

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

Out[9]: 2

To print no of rows and columns

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

Out[10]: (9, 5)

To find missing values

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

Out[11]:

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
5	False	False	False	False	False
6	False	False	False	False	False
7	False	False	False	False	False
8	False	False	False	False	False

To drop nulll values with constatns

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

Out[12]:

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	A	5.62%	7.73%	6.16%	75
1	B	4.21%	17.27%	19.21%	160
2	C	9.83%	11.60%	5.17%	101
3	D	2.81%	21.91%	7.88%	127
4	E	25.28%	10.57%	11.82%	179
5	F	8.15%	16.24%	18.47%	167
6	G	18.54%	8.76%	17.49%	171

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
7	H	25.56%	5.93%	13.79%	170
8	Grand Total	100.00%	100.00%	100.00%	1150

In [13]:
```
data.dropna()
```

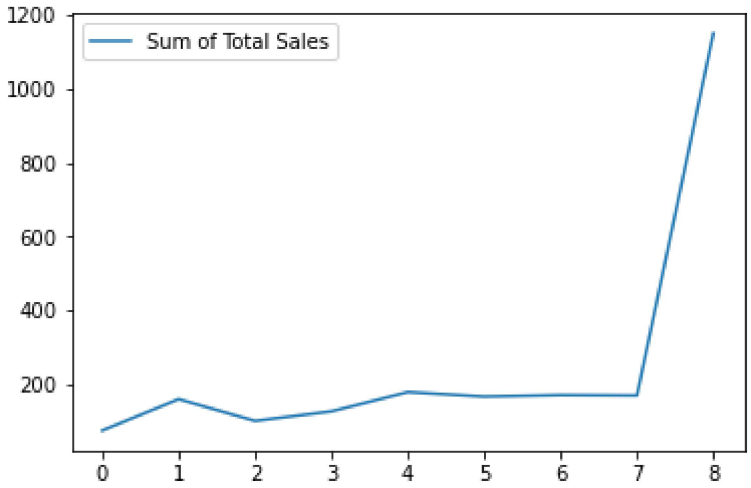
Out[13]:

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	A	5.62%	7.73%	6.16%	75
1	B	4.21%	17.27%	19.21%	160
2	C	9.83%	11.60%	5.17%	101
3	D	2.81%	21.91%	7.88%	127
4	E	25.28%	10.57%	11.82%	179
5	F	8.15%	16.24%	18.47%	167
6	G	18.54%	8.76%	17.49%	171
7	H	25.56%	5.93%	13.79%	170
8	Grand Total	100.00%	100.00%	100.00%	1150

Line Plot

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

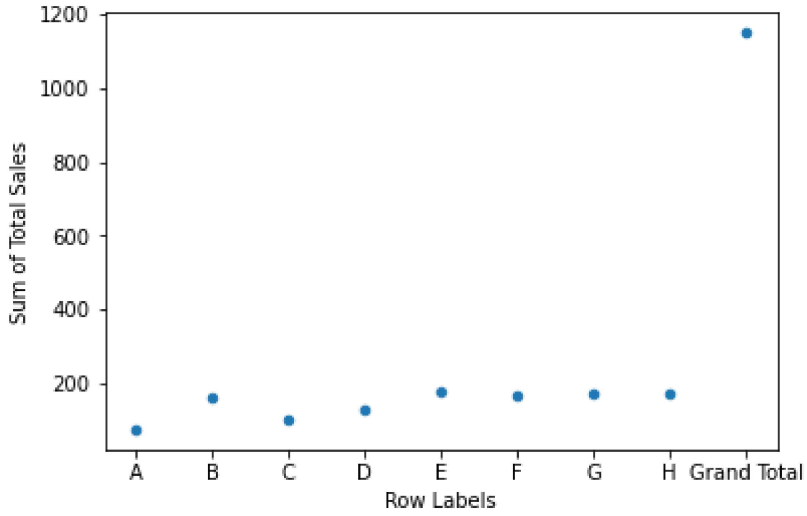
Out[14]: <AxesSubplot:>



Scatter Plot

In [16]:
```
data.plot.scatter(x='Row Labels',y='Sum of Total Sales')
```

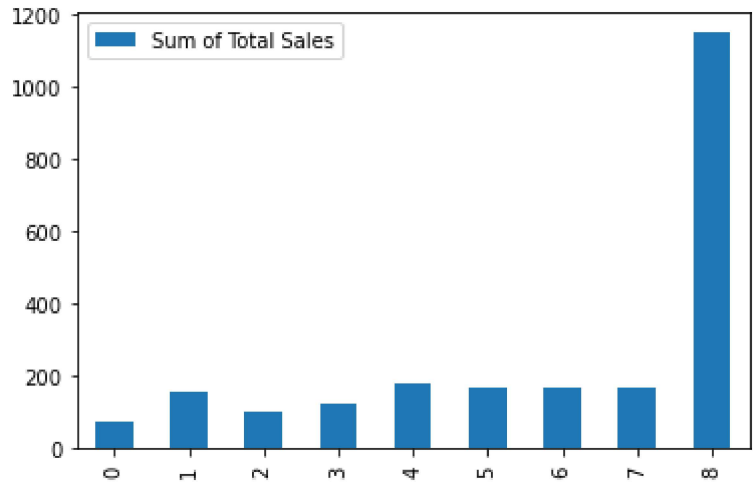
Out[16]: <AxesSubplot:xlabel='Row Labels', ylabel='Sum of Total Sales'>



Bar Chart

In [17]:
```
data.plot.bar()
```

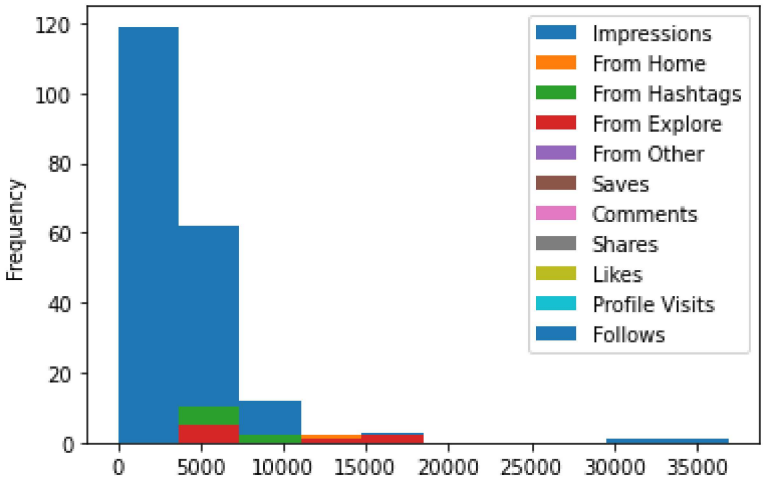
Out[17]: <AxesSubplot:>



Histogram

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

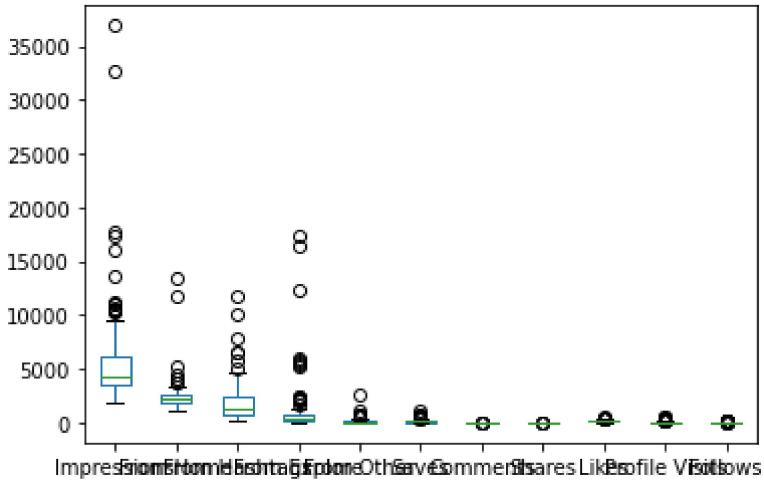
Out[18]: <AxesSubplot:ylabel='Frequency'>



Box Plot

In [19]:
```
data.plot.box()
```

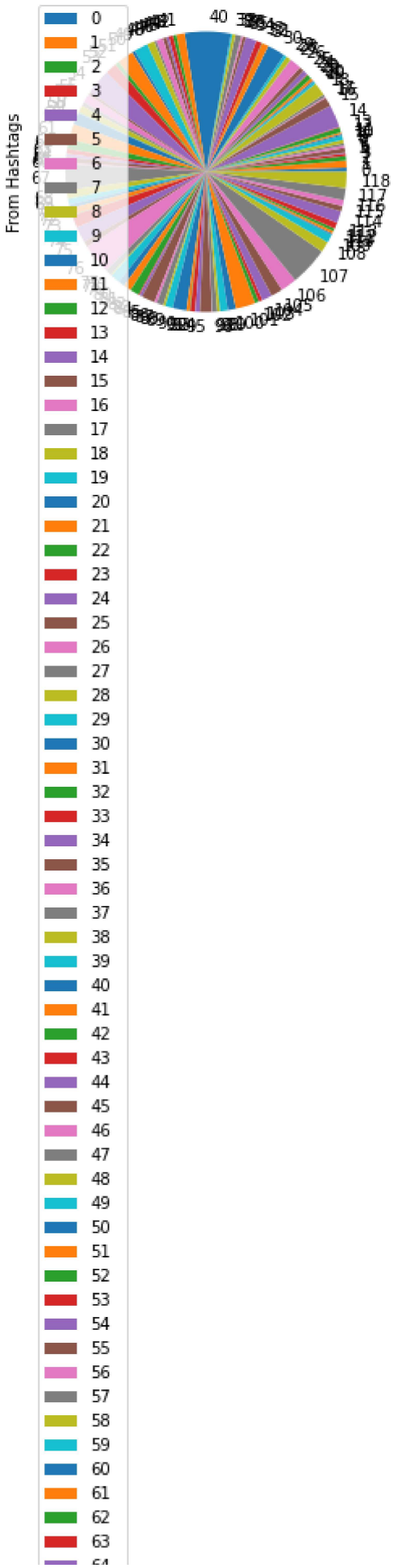
Out[19]: <AxesSubplot:>

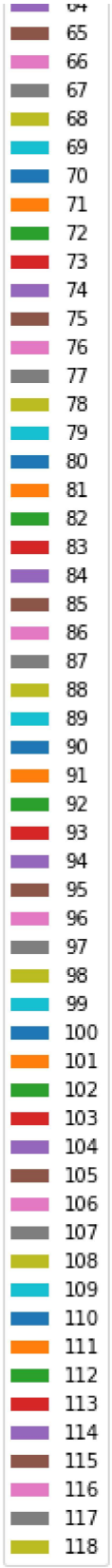


Pie Chart

In [20]:
```
data.plot.pie(y="From Hashtags")
```

Out[20]: <AxesSubplot:ylabel='From Hashtags'>

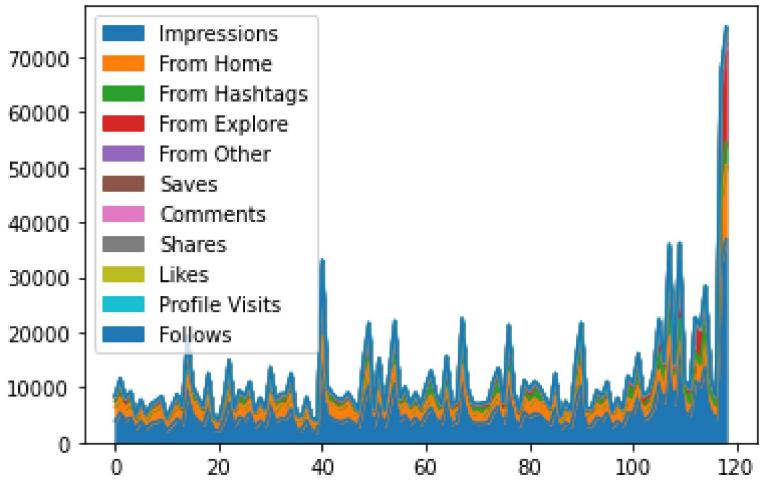

From Hashtags
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63



Area

In [21]:
```
data.plot.area()
```

Out[21]: <AxesSubplot:>


Impressions
From Home
From Hashtags
From Explore
From Other
Saves
Comments
Shares
Likes
Profile Visits
Follows
70000
60000
50000
40000
30000
20000
10000
0
0
20
40
60
80
100
120

In [ ]: