

# Basic Analysis using Numpy and Pandas

## import libraries

```
In [4]: import pandas as pd
import numpy as np
import matplotlib as pp
```

## import dataset

```
In [6]: data=pd.read_csv(r"E:\154\4_drug200.csv")
```

```
In [7]: display(data)
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY
...	...	...	...	...	...	...
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

200 rows × 6 columns

## To display top 10 rows

```
In [8]: data.head()
```

Out[8]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY

To display last 5 rows

In [9]:

```
data.tail()
```

Out[9]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

In [10]:

```
data.dtypes
```

Out[10]:

Age int64  
Sex object  
BP object  
Cholesterol object  
Na\_to\_K float64  
Drug object  
dtype: object

To view statistical summary

In [11]:

```
data.describe()
```

Out[11]:

	Age	Na_to_K
count	200.000000	200.000000
mean	44.315000	16.084485
std	16.544315	7.223956
min	15.000000	6.269000
25%	31.000000	10.445500
50%	45.000000	13.936500
75%	58.000000	19.380000
max	74.000000	38.247000

```
In [12]: data.size
```

Out[12]: 1200

```
In [13]: data.ndim
```

Out[13]: 2

## To print no of rows and columns

```
In [14]: data.shape
```

Out[14]: (200, 6)

## To find missing values

```
In [15]: data.isna()
```

Out[15]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
195	False	False	False	False	False	False
196	False	False	False	False	False	False
197	False	False	False	False	False	False
198	False	False	False	False	False	False
199	False	False	False	False	False	False

200 rows × 6 columns

## To drop nulll values with constatns

```
In [16]: data.fillna(5)
```

Out[16]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
--	-----	-----	----	-------------	---------	------

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY
...	...	...	...	...	...	...
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

200 rows × 6 columns

In [17]:

data.dropna()

Out[17]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY
...	...	...	...	...	...	...
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

200 rows × 6 columns

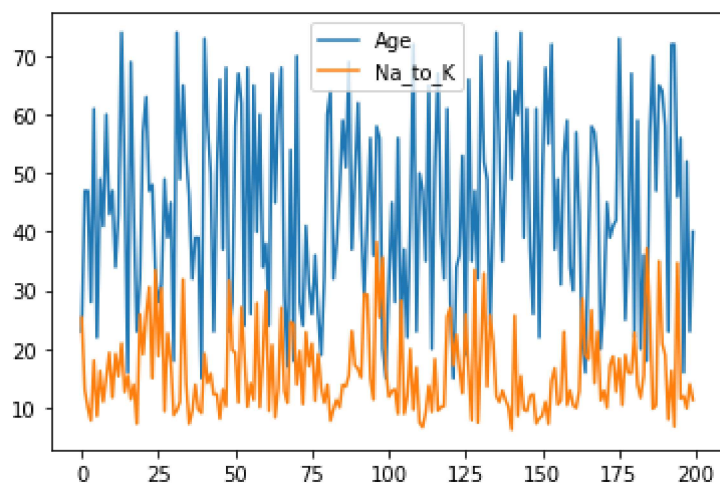
# Line Plot

In [18]:

data.plot.line()

Out[18]: <AxesSubplot:>

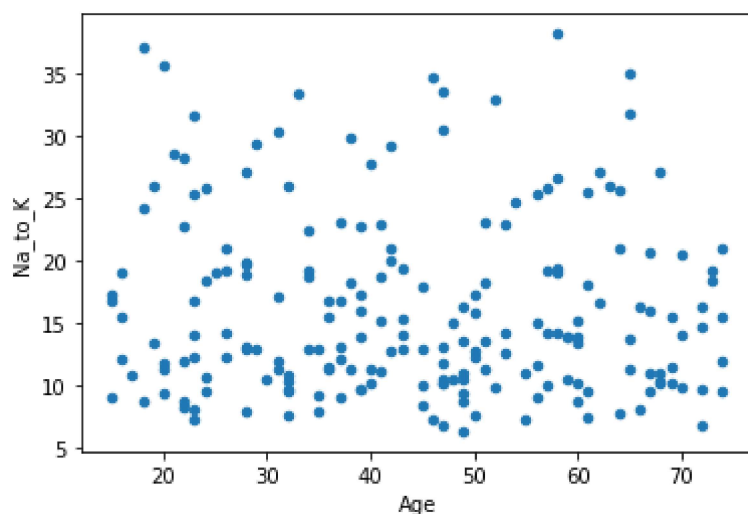
Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js



## Scatter Plot

```
In [19]: data.plot.scatter(x='Age',y='Na_to_K')
```

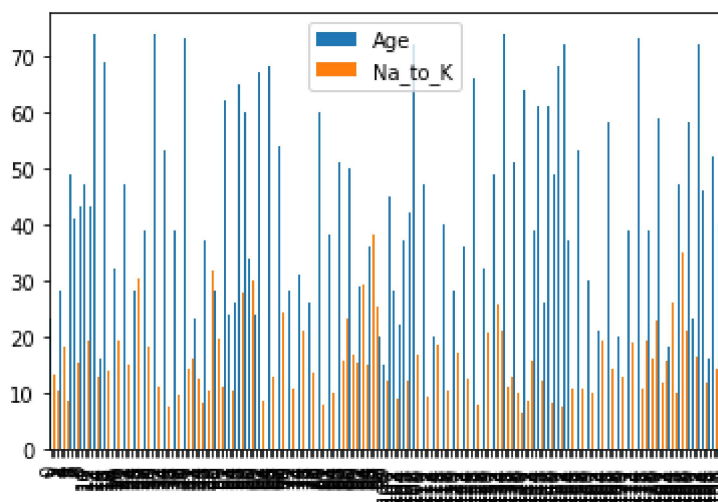
```
Out[19]: <AxesSubplot:xlabel='Age', ylabel='Na_to_K'>
```



## Bar Chart

```
In [20]: data.plot.bar()
```

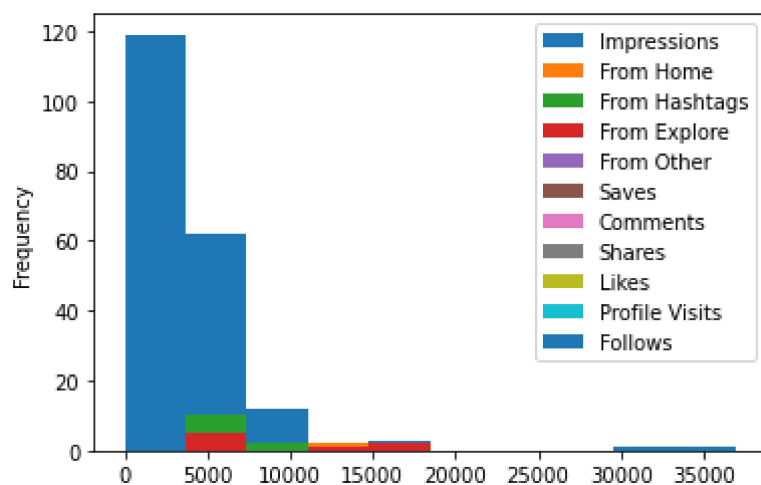
```
Out[20]: <AxesSubplot:>
```



## Histogram

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

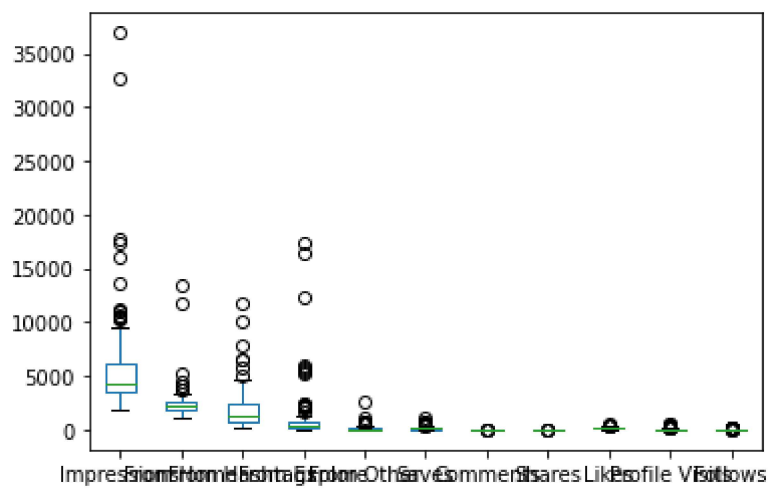
```
Out[18]: <AxesSubplot:ylabel='Frequency'>
```



## Box Plot

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

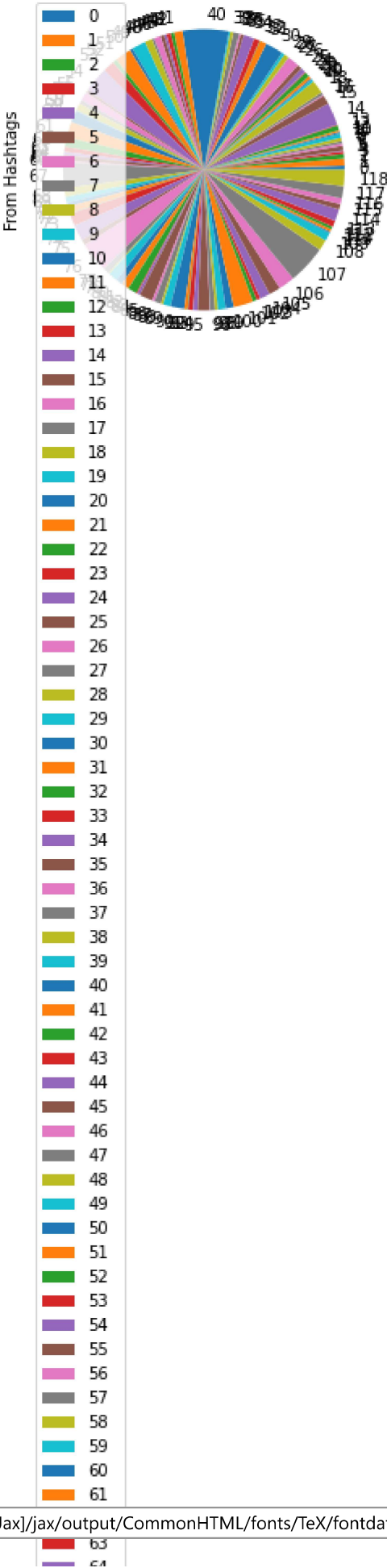
```
Out[19]: <AxesSubplot:>
```



## Pie Chart

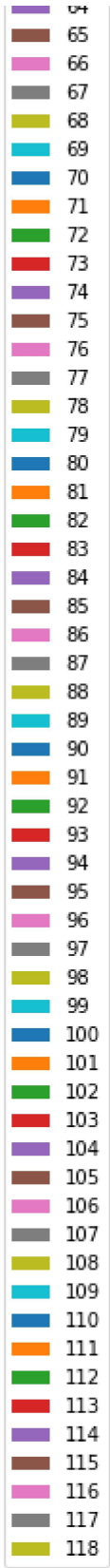
```
In [20]: data.plot.pie(y="From Hashtags")
```

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



Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

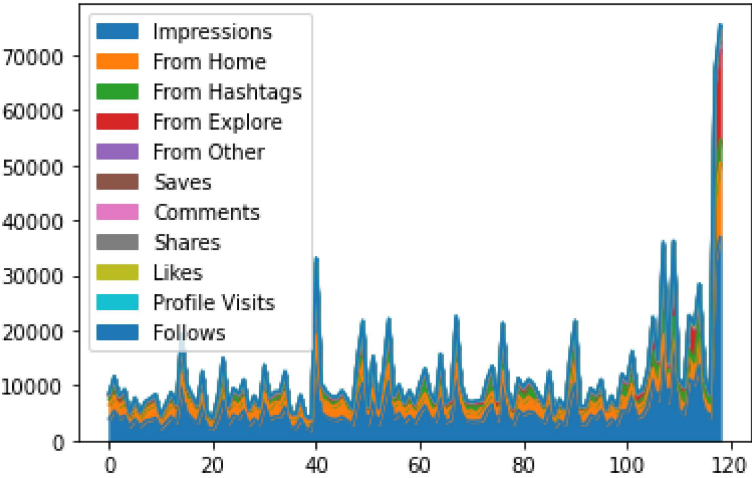




# Area

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

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js



```
In [ ]:
```