

Heamnath

20104028

Basic Analysis using Numpy and Pandas

Importing libraries

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

importing datasets

```
In [2]: df=pd.read_csv("4_drug200.csv")
df
```

```
Out[2]:
```

	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 first 10 rows

```
In [3]: df.head(10)
```

```
Out[3]:
```

	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
5	22	F	NORMAL	HIGH	8.607	drugX
6	49	F	NORMAL	HIGH	16.275	drugY
7	41	M	LOW	HIGH	11.037	drugC
8	60	M	NORMAL	HIGH	15.171	drugY
9	42	M	LOW	NORMAL	19.368	drugY

To display last 5 rows

```
In [4]: df.tail(5)
```

```
Out[4]:
```

	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

Statistical Summary

```
In [5]: df.describe()
```

```
Out[5]:
```

	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

To print no of rows and columns

```
In [6]: df.shape
```

```
Out[6]: (200, 6)
```

To print total no of elements

```
In [7]: df.size
```

```
Out[7]: 1200
```

To find the null value

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

```
Out[8]:
```

	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 fill the missing value

```
In [9]: df.fillna(value=0)
```

```
Out[9]:
```

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

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
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

Print column names

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

```
Out[10]: Index(['Age', 'Sex', 'BP', 'Cholesterol', 'Na_to_K', 'Drug'], dtype='object')
```

To print particular column names

```
In [11]: data=df[["Age","Na_to_K"]]  
data
```

```
Out[11]:
```

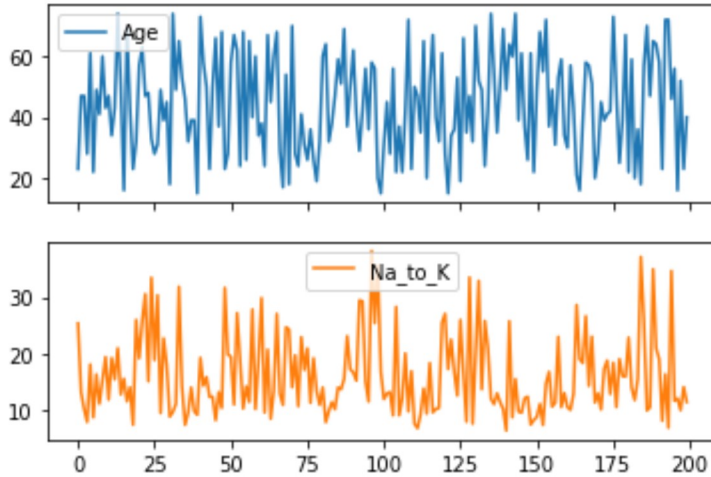
	Age	Na_to_K
0	23	25.355
1	47	13.093
2	47	10.114
3	28	7.798
4	61	18.043
...
195	56	11.567
196	16	12.006
197	52	9.894
198	23	14.020
199	40	11.349

200 rows × 2 columns

Line chart with subplots

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

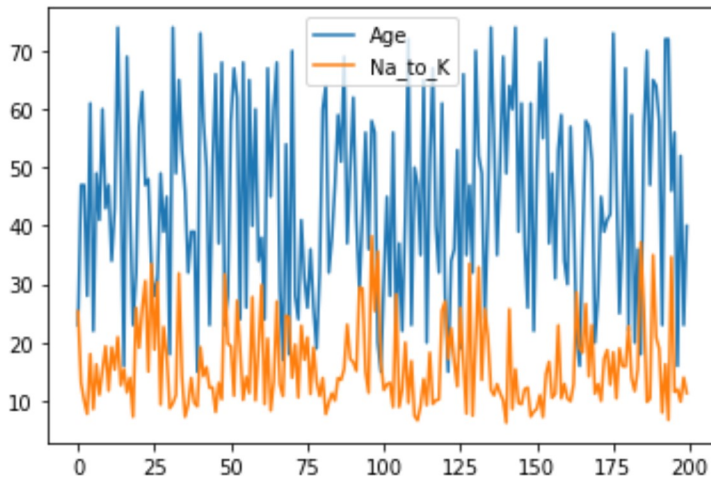
```
Out[12]: array([<AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



Line chart

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

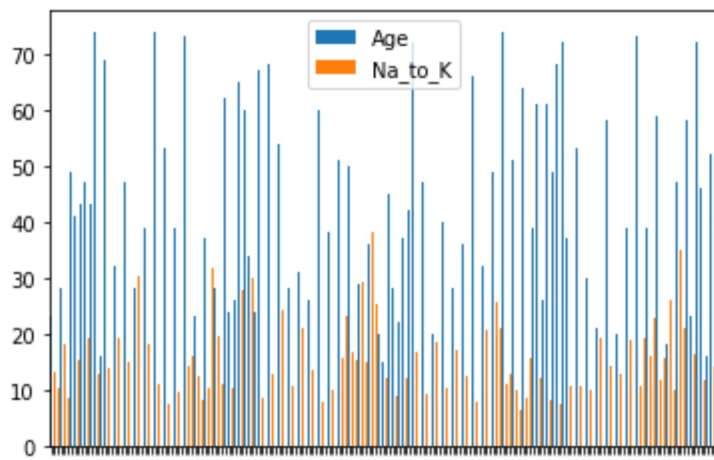
```
Out[13]: <AxesSubplot:>
```



Bar chart

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

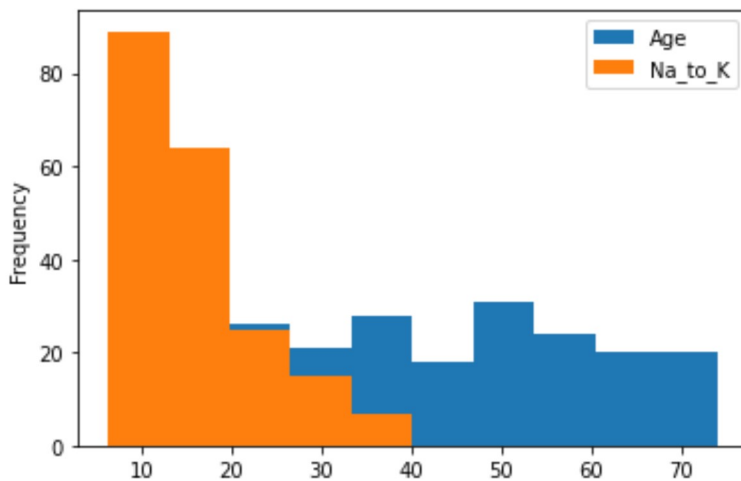
```
Out[14]: <AxesSubplot:>
```



Histogram

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

```
Out[15]: <AxesSubplot:ylabel='Frequency'>
```



Area chart

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

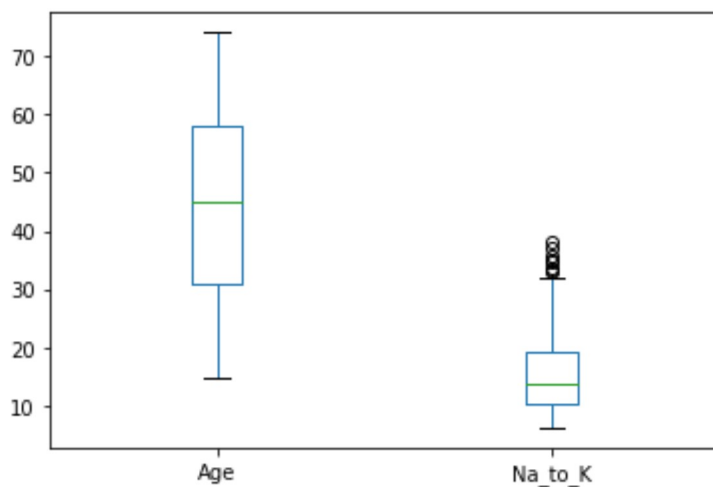
```
Out[16]: <AxesSubplot:>
```



Box chart

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

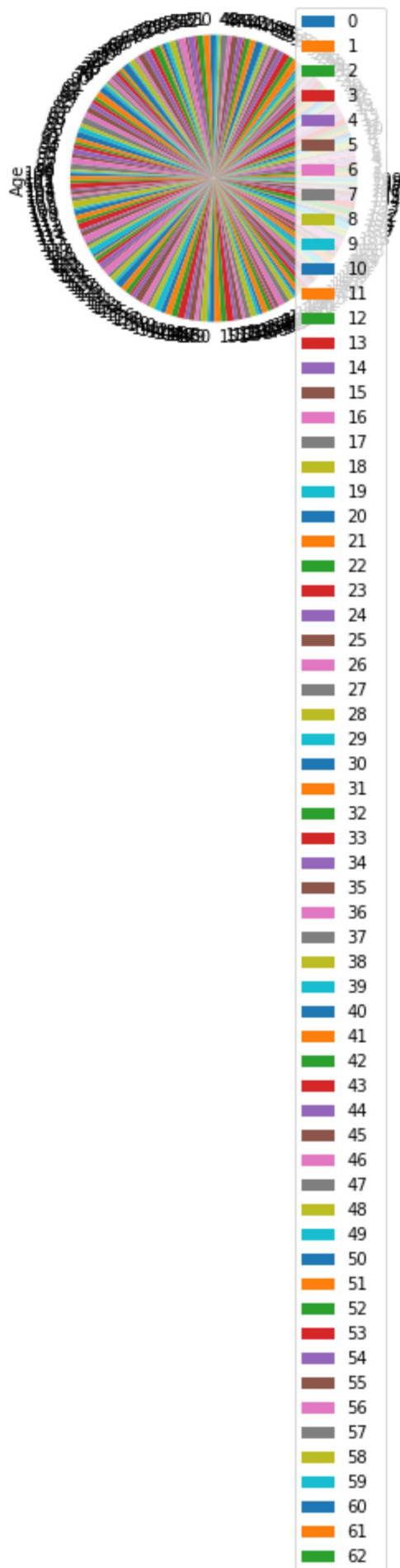
```
Out[17]: <AxesSubplot:>
```



































































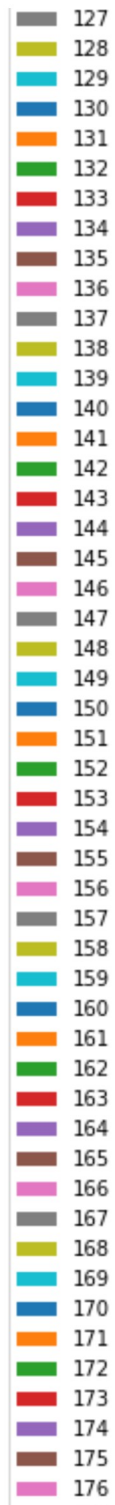
Pie chart

```
In [18]: data.plot.pie(y='Age')
```

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



	63
	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
	119
	120
	121
	122
	123
	124
	125
	126



Scatter chart

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

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

