

1.IMPORTING LIBRARIES

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

2.Importing dataset

```
In [2]: data=pd.read_csv(r"C:\Users\user\Downloads\4_drug200 - 4_drug200.csv")  
data
```

```
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

3.head

```
In [3]: data.head(8)
```

```
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

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
5	22	F	NORMAL	HIGH	8.607	drugX
6	49	F	NORMAL	HIGH	16.275	drugY
7	41	M	LOW	HIGH	11.037	drugC

4.tail

In [4]:

```
data.tail(7)
```

Out[4]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
193	72	M	LOW	HIGH	6.769	drugC
194	46	F	HIGH	HIGH	34.686	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

5.describe()

In [5]:

```
data.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

6.shape()

In [6]:

```
np.shape(data)
```

Out[6]: (200, 6)

7.size()

In [7]: np.size(data)

Out[7]: 1200

8.isna()

In [8]: data.isna()

	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

9.dropna

In [9]: data.dropna()

	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

	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

200 rows × 6 columns

10.selecting specific column

```
In [10]: da=data[["Age","Na_to_K"]]
da
```

Out[10]:

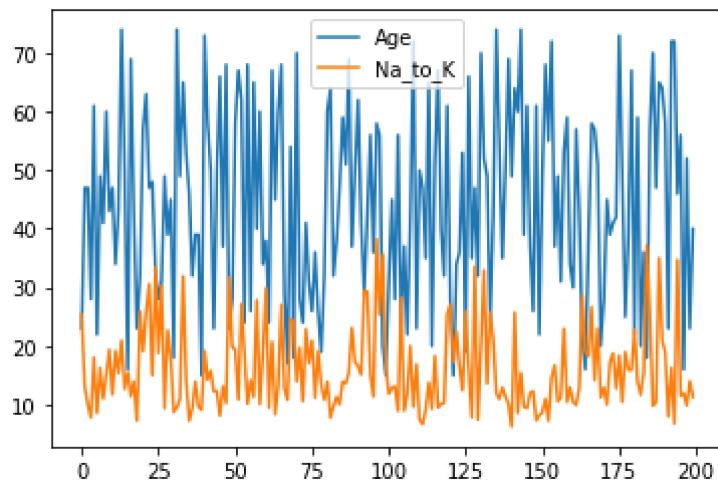
	Age	Na_to_K
0	23	25.355
1	47	13.093
2	47	10.114
3	28	7.798
4	61	18.043
...
195	56	11.567
196	16	12.006
197	52	9.894
198	23	14.020
199	40	11.349

200 rows × 2 columns

11.line plot

```
In [11]: data.plot.line()
```

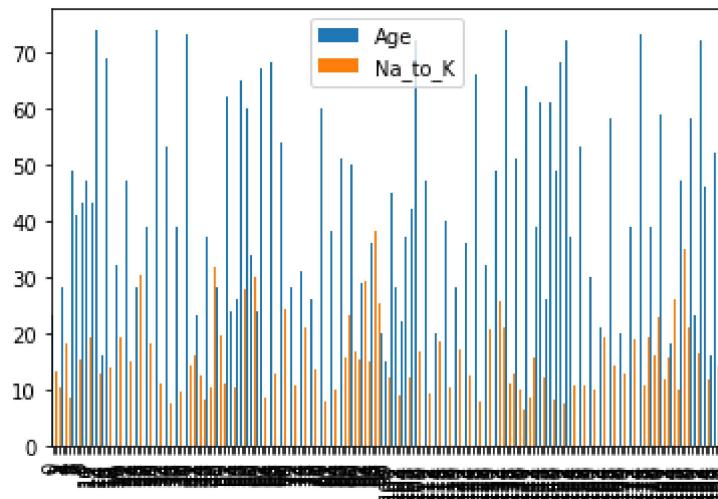
Out[11]: <AxesSubplot:>



12.bar plot

```
In [12]: data.plot.bar()
```

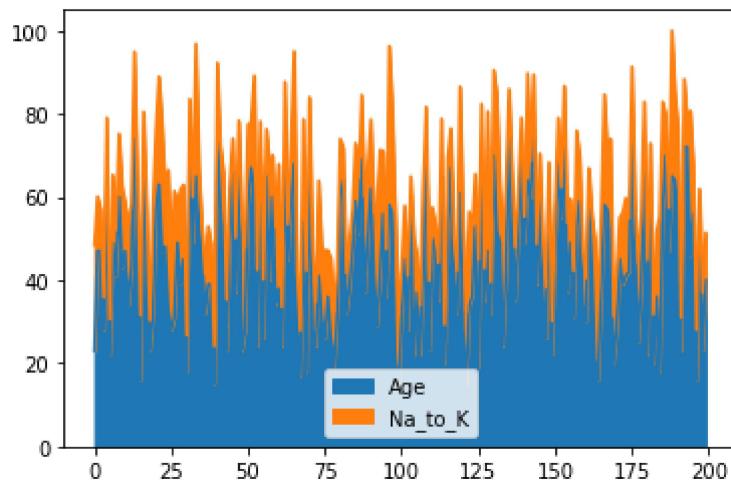
```
Out[12]: <AxesSubplot:
```



13.area plot

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

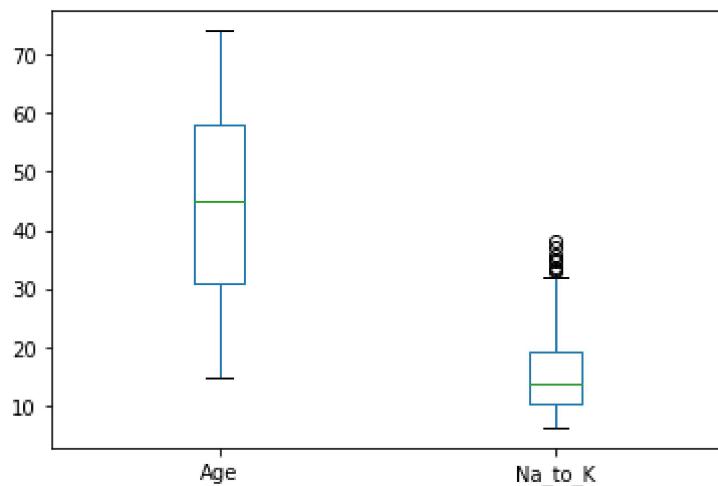
```
Out[13]: <AxesSubplot:
```



14.box plot

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

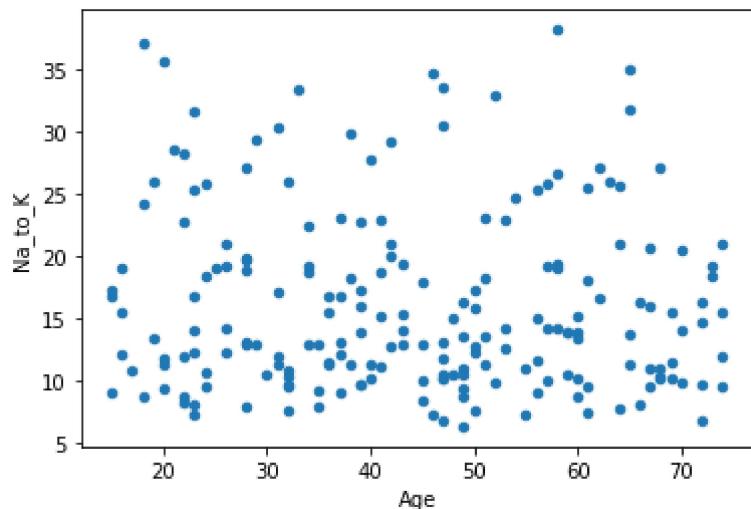
```
Out[14]: <AxesSubplot:
```



15.scatter plot

```
In [15]: data.plot.scatter("Age", "Na_to_K")
```

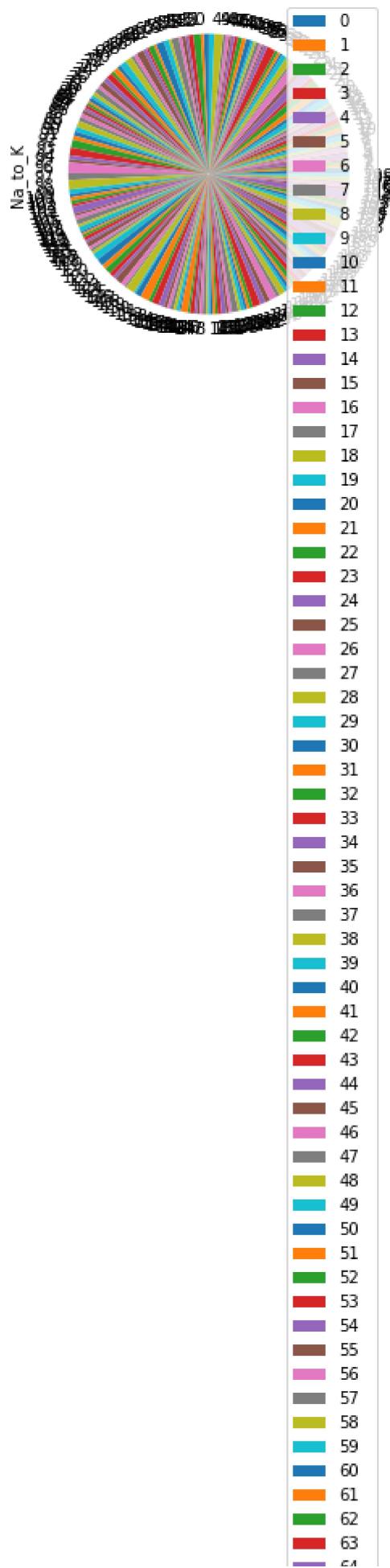
```
Out[15]: <AxesSubplot:xlabel='Age', ylabel='Na_to_K'>
```



16.pie plot

In [16]:
`data.plot.pie(x="Age",y="Na_to_K")`

Out[16]: <AxesSubplot:ylabel='Na_to_K'>

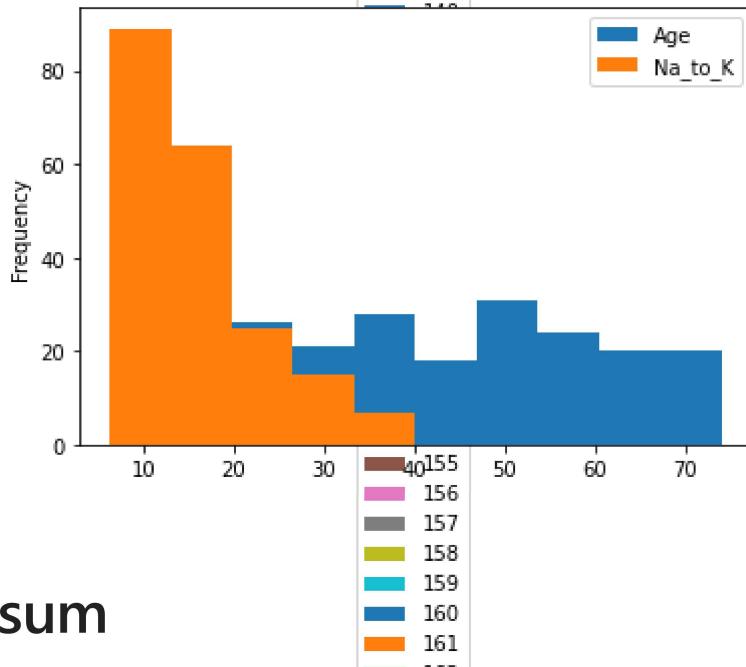


u+	64
■	65
■	66
■	67
■	68
■	69
■	70
■	71
■	72
■	73
■	74
■	75
■	76
■	77
■	78
■	79
■	80
■	81
■	82
■	83
■	84
■	85
■	86
■	87
■	88
■	89
■	90
■	91
■	92
■	93
■	94
■	95
■	96
■	97
■	98
■	99
■	100
■	101
■	102
■	103
■	104
■	105
■	106
■	107
■	108
■	109
■	110
■	111
■	112
■	113
■	114
■	115
■	116
■	117
■	118
■	119
■	120
■	121
■	122
■	123
■	124
■	125
■	126
■	127
■	128
■	129

17.histogram

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

Out[17]: <AxesSubplot:ylabel='Frequency'>



sum

In [18]: `print(data.sum())`

Age	165	8863
Sex	166	FMMFFFFM...FFMMF...FFMMF...FFMMF...FFMMF...FFMMF...
BP	167	HIGHLOWLOWNORMAL...LOWNORMALNORMALLOWNORMALLOWLOW...
Cholesterol	168	HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH...
Na_to_K	169	3216.897
Drug	170	drugYdrugCdrugYdrugXdrugYdrugXdrugCdrugYd...
dtype: object	171	

mean

In [19]: `print(data.mean())`

Age	44.315000
Na_to_K	16.084485
dtype: float64	

median

In [20]: `print(data.median())`

Age	45.0000
Na_to_K	13.9365
dtype: float64	

mode

In [21]:

```
print(data.mode())
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	47.0	M	HIGH	HIGH	12.006	drugY
1	NaN	NaN	NaN	NaN	18.295	NaN

min

In [22]:

```
print(data.min())
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
	15	F	HIGH	HIGH	6.269	drugA

dtype: object

max

In [23]:

```
print(data.max())
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
	74	M	NORMAL	NORMAL	38.247	drugY

dtype: object

count

In [24]:

```
print(data.count())
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
	200	200	200	200	200	200

dtype: int64

cumsum

In [25]:

```
print(data.cumsum())
```

	Age	Sex
0	23	F

```

1      70                               FM
2     117                              FMM
3     145                              FMMF
4     206                             FMMFF
..    ...
195   8732  FMMFFFFMFFFMMMFMMMF... FMFFMFMMFMMMF... FMMFF...
196   8748  FMMFFFFMFFFMMMFMMMF... FMFFMFMMFMMMF... FMMFF...
197   8800  FMMFFFFMFFFMMMFMMMF... FMFFMFMMFMMMF... FMMFF...
198   8823  FMMFFFFMFFFMMMFMMMF... FMFFMFMMFMMMF... FMMFF...
199   8863  FMMFFFFMFFFMMMFMMMF... FMFFMFMMFMMMF... FMMFF...

                                BP \
0          HIGH
1        HIGHLOW
2      HIGHLOWLOW
3    HIGHLOWLOWNORMAL
4  HIGHLOWLOWNORMALLOW
..
195 HIGHLOWLOWNORMALLOWNORMALNORMALLOWLOW...
196 HIGHLOWLOWNORMALLOWNORMALNORMALLOWLOW...
197 HIGHLOWLOWNORMALLOWNORMALNORMALLOWLOW...
198 HIGHLOWLOWNORMALLOWNORMALNORMALLOWLOW...
199 HIGHLOWLOWNORMALLOWNORMALNORMALLOWLOW...

                                Cholesterol  Na_to_K \
0           HIGH      25.355
1         HIGHHIGH  38.448
2      HIGHHIGHHIGH 48.562
3    HIGHHIGHHIGHHIGH 56.360
4  HIGHHIGHHIGHHIGHHIGH 74.403
..
195 HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH... 3169.628
196 HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH... 3181.634
197 HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH... 3191.528
198 HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH... 3205.548
199 HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH... 3216.897

                                Drug
0       drugY
1   drugYdrugC
2 drugYdrugCdrugC
3 drugYdrugCdrugCdrugX
4 drugYdrugCdrugCdrugXdrugY
..
195 drugYdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...
196 drugYdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...
197 drugYdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...
198 drugYdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...
199 drugYdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...

```

[200 rows x 6 columns]

standard deviation

In [26]:

```
print(data.std())
```

```

Age      16.544315
Na_to_K  7.223956
dtype: float64

```

describe

```
In [27]: print(data.describe())
```

	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

variance

```
In [28]: print(data.var())
```

Age	273.714347
Na_to_K	52.185533
dtype:	float64

covariance

```
In [29]: print(data.cov())
```

	Age	Na_to_K
Age	273.714347	-7.543752
Na_to_K	-7.543752	52.185533

correlation

```
In [30]: from scipy.stats import spearmanr  
from scipy.stats import pearsonr
```

pearson correlation

```
In [31]: print(pearsonr(data["Age"], data["Na_to_K"]))
```

(-0.06311949726772592, 0.3745756399034559)

spearman correlation

```
In [32]: print(spearmanr(data["Age"], data["Na_to_K"]))
```

SpearmanResult(correlation=-0.047273882688479915, pvalue=0.5062200581387418)