In [2]:
```
import numpy as np
```

In [3]:
```
import pandas as pd
```

Pre-processing

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

Out[4]:

	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

In [13]:
```
df=data.head(10)
df
```

Out[13]:

	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	43	M	LOW	NORMAL	19.368	drugY

In [6]:
```
data.describe()
```

Out[6]:

	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 [15]:
```
df1=data.tail(10)
df1
```

Out[15]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
190	58	M	HIGH	HIGH	18.991	drugY
191	23	M	HIGH	HIGH	8.011	drugA
192	72	M	LOW	HIGH	16.310	drugY
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

In [6]:
```
data.mean()
```

Out[6]:
```
Age        44.315000
Na_to_K    16.084485
dtype: float64
```

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

Out[7]:
```
Age        45.0000
Na_to_K    13.9365
dtype: float64
```

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

Out[8]:

	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

In [9]:
```
data.sum()
```

Out[9]:
```
Age                                                          8863
Sex            FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMMFMFFMMFF...
BP             HIGHLOWLOWNORMALLOWNORMALNORMALLOWNORMALLOWLOW...
Cholesterol    HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH...
Na_to_K                                                  3216.897
Drug           drugYdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...
dtype: object
```

In [10]:
```
data.cumsum()
```

Out[10]:

	Age	Sex	
0	23	F	
1	70	FM	
2	117	FMM	
3	145	FMMF	
4	206	FMMFF	
...	...	...	
195	8732	FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMMFMFFMMFF...	HIGHLOWLOWNO
196	8748	FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMMFMFFMMFF...	HIGHLOWLOWNO
197	8800	FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMMFMFFMMFF...	HIGHLOWLOWNO
198	8823	FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMMFMFFMMFF...	HIGHLOWLOWNO
199	8863	FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMMFMFFMMFF...	HIGHLOWLOWNO

200 rows × 6 columns

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

Out[11]:
```
Age               74
Sex                M
BP            NORMAL
Cholesterol   NORMAL
Na_to_K       38.247
Drug           drugY
dtype: object
```

In [12]:
```
data.min()
```

Out[12]:
```
Age               15
Sex                F
BP              HIGH
Cholesterol     HIGH
Na_to_K        6.269
Drug           drugA
dtype: object
```

In [17]:
```
data.cov()
```

Out[17]:

	Age	Na_to_K
Age	273.714347	-7.543752
Na_to_K	-7.543752	52.185533

In [18]:
```
data.corr()
```

Out[18]:

	Age	Na_to_K
Age	1.000000	-0.063119
Na_to_K	-0.063119	1.000000

In [19]:
```
data.count()
```

Out[19]:
```
Age            200
Sex            200
BP             200
Cholesterol    200
Na_to_K        200
Drug           200
dtype: int64
```

In [ ]:
```

```