mk 27-07-2023

In [ ]:
```python
# import libaries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [104]:
```python
x=pd.read_csv(r"C:\Users\user\Downloads\4_drug200 - 4_drug200.csv")
```

Out[104]:

	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 [105]: `x=x.head(100)`

Out[105]:

	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
...	...	...	...	...	...	...
95	36	M	LOW	NORMAL	11.424	drugX
96	58	F	LOW	HIGH	38.247	drugY
97	56	F	HIGH	HIGH	25.395	drugY
98	20	M	HIGH	NORMAL	35.639	drugY
99	15	F	HIGH	NORMAL	16.725	drugY

100 rows × 6 columns

In [106]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 6 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Age          100 non-null    int64
 1   Sex          100 non-null    object
 2   BP           100 non-null    object
 3   Cholesterol  100 non-null    object
 4   Na_to_K      100 non-null    float64
 5   Drug         100 non-null    object
dtypes: float64(1), int64(1), object(4)
memory usage: 4.8+ KB
```

In [107]:

Out[107]: `Index(['Age', 'Sex', 'BP', 'Cholesterol', 'Na_to_K', 'Drug'], dtype='object')`

In [109]:
```python
d=x[['Age', 'Sex', 'BP', 'Cholesterol', 'Na_to_K', 'Drug']]
```

Out[109]:

	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
...	...	...	...	...	...	...
95	36	M	LOW	NORMAL	11.424	drugX
96	58	F	LOW	HIGH	38.247	drugY
97	56	F	HIGH	HIGH	25.395	drugY
98	20	M	HIGH	NORMAL	35.639	drugY
99	15	F	HIGH	NORMAL	16.725	drugY

100 rows × 6 columns

In [110]:
```python
d.describe()
```

Out[110]:

	Age	Na_to_K
count	100.000000	100.000000
mean	43.770000	16.823000
std	16.367531	7.257723
min	15.000000	7.285000
25%	30.500000	11.031250
50%	43.000000	15.025500
75%	58.000000	20.020250
max	74.000000	38.247000

In [111]:

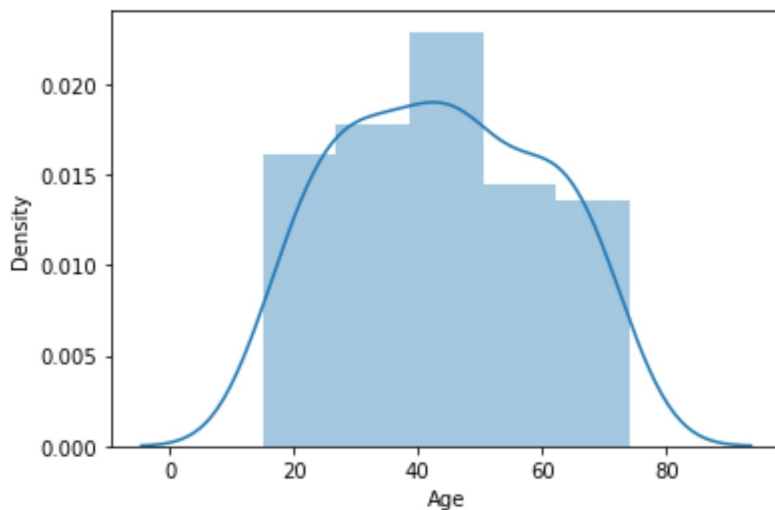Out[111]: <seaborn.axisgrid.PairGrid at 0x2602cb536a0>
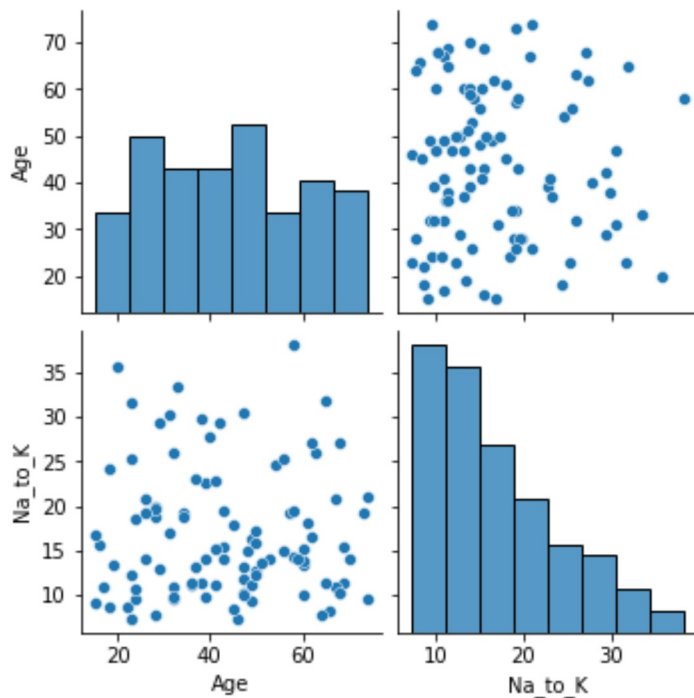
In [113]:

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: Fut
ureWarning: `distplot` is a deprecated function and will be removed in a futu
re version. Please adapt your code to use either `displot` (a figure-level fu
nction with similar flexibility) or `histplot` (an axes-level function for hi
stograms).
  warnings.warn(msg, FutureWarning)
```
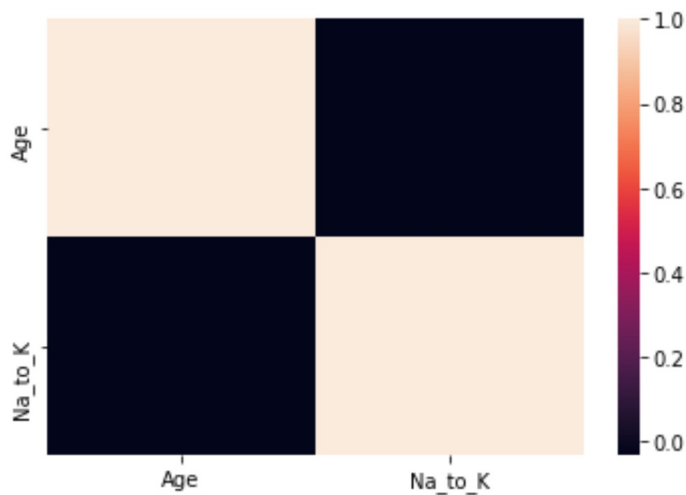
Out[113]: <AxesSubplot:xlabel='Age', ylabel='Density'>

In [114]: x1=x[['Age', 'Sex', 'BP', 'Cholesterol', 'Na_to_K', 'Drug']]

In [115]:

Out[115]: <AxesSubplot:>

In [131]: `x=x1[['Age','Na_to_K']]`

In [132]:
```python
# to split my dataset into traning and test date

from sklearn.model_selection import train_test_split
```

In [133]:
```python
from sklearn.linear_model import LinearRegression

lr=LinearRegression()
```
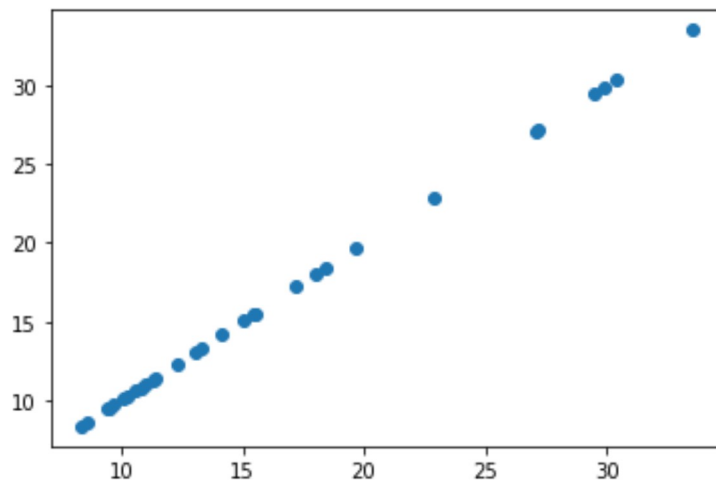
Out[133]: LinearRegression()

In [134]:

1.0658141036401503e-14

In [135]: `coeff=pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])`

Out[135]:

	Co-efficient
Age	6.566793e-18
Na_to_K	1.000000e+00

In [136]:
```python
prediction=lr.predict(x_test)
```

Out[136]: <matplotlib.collections.PathCollection at 0x2602cb2a880>

In [137]:

Out[137]: 1.0

In [ ]: