```python
import numpy as np
import pandas as pd
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
```

In [37]:
```python
df=pd.read_csv(r"C:\Users\MY HOME\Downloads\drug200.csv")
df
```

Out[37]:

	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 [ ]:
```python
df.info()
```

In [15]: ▶| `df.describe()`

Out[15]:

	Age	Sex	BP	Cholesterol	Na_to_K
count	200.000000	200.000000	200.000000	200.000000	200.000000
mean	44.315000	0.480000	0.975000	1.485000	16.084485
std	16.544315	0.500854	0.785788	0.501029	7.223956
min	15.000000	0.000000	0.000000	1.000000	6.269000
25%	31.000000	0.000000	0.000000	1.000000	10.445500
50%	45.000000	0.000000	1.000000	1.000000	13.936500
75%	58.000000	1.000000	2.000000	2.000000	19.380000
max	74.000000	1.000000	2.000000	2.000000	38.247000

In [16]: ▶| `df.shape`

Out[16]: `(200, 6)`

In [17]: ▶| `df['Sex'].value_counts()`

Out[17]:
```
Sex
0    104
1     96
Name: count, dtype: int64
```

In [18]: ▶| `df['BP'].value_counts()`

Out[18]:
```
BP
1    77
0    64
2    59
Name: count, dtype: int64
```

In [19]:
```python
df['Cholesterol'].value_counts()
```

Out[19]:
```
Cholesterol
1    103
2     97
Name: count, dtype: int64
```

In [20]:
```python
convert={"Sex":{"M":0,"F":1}}
df=df.replace(convert)
df
```

Out[20]:

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

200 rows × 6 columns

In [21]:

```python
convert={"BP":{"HIGH":1,"LOW":0,"NORMAL":2}}
df=df.replace(convert)
df
```

Out[21]:

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

200 rows × 6 columns

In [22]:
```python
convert={"Cholesterol":{"HIGH":1,"LOW":0,"NORMAL":2}}
df=df.replace(convert)
df
```

Out[22]:

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

200 rows × 6 columns

In [32]:
```python
convert={"Drug":{"drugX":3,"drugA":0,"drugB":1,"drugC":2,"drugY":4}}
df=df.replace(convert)
df
```

Out[32]:

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

200 rows × 6 columns

In [33]:
```python
x=["Sex","BP","Cholesterol"]
y=[0,1,2,3,4]
all_inputs=df[x]
all_classes=df["Drug"]
```

In [34]:
```python
x_train,x_test,y_train,y_test=train_test_split(all_inputs,all_classes,test_size=0.25)
```

In [35]: ▶| ```python
clf=DecisionTreeClassifier(random_state=0)
```

In [36]: ▶| ```python
clf.fit(x_train,y_train)
```

Out[36]:   ▼   DecisionTreeClassifier
           DecisionTreeClassifier(random_state=0)

In [38]: ▶| ```python
score=clf.score(x_test,y_test)
print(score)
```

0.34