

```
In [3]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
drug = pd.read_csv ('drug200.csv')
drug
```

```
Out[3]:
```

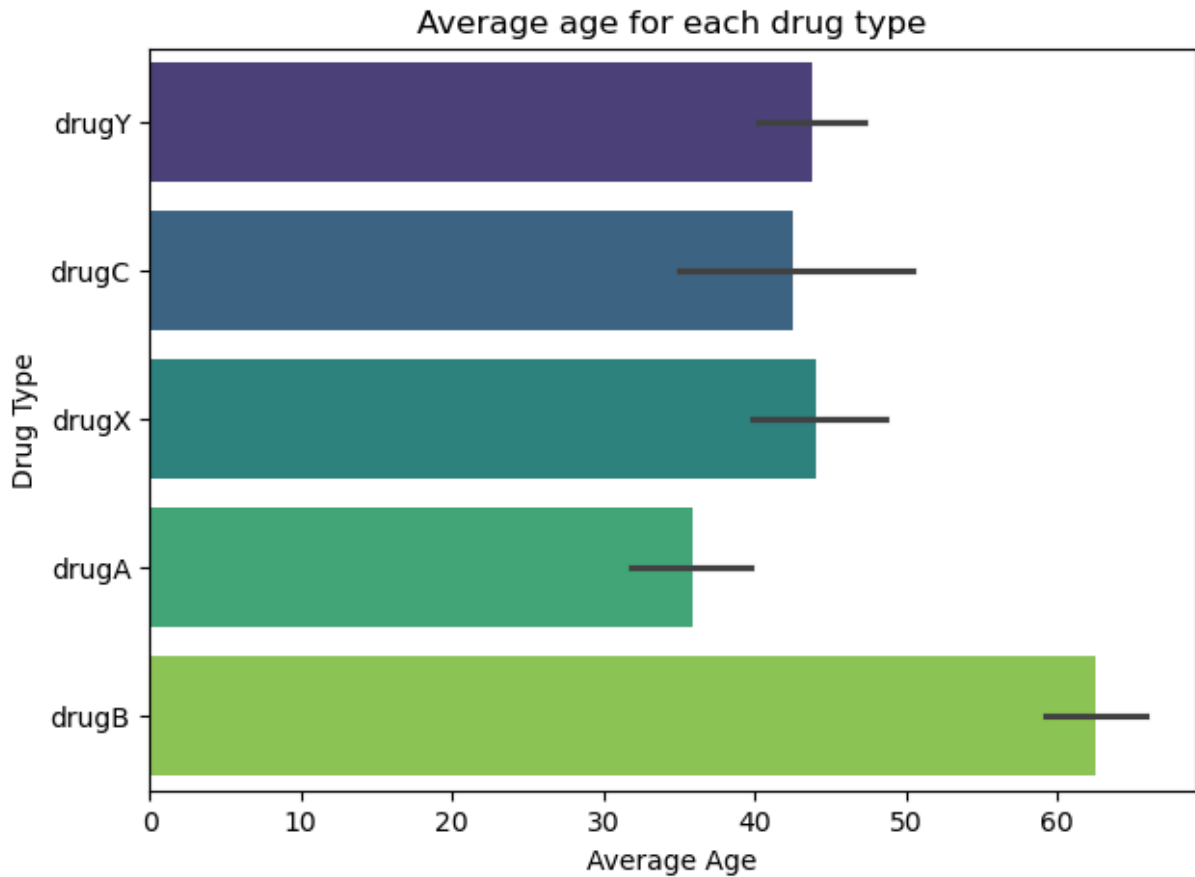
	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY
...	...	...	...	...	...	...
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894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

200 rows × 6 columns

### Q1: What is the average age for each drug type?

```
In [4]: avg_age_per_drug = drug.groupby("Drug")["Age"].mean()
print(avg_age_per_drug)
plt.figure(figsize=(7,5))
sns.barplot(data=drug, x="Age", y="Drug", hue="Drug", palette="viridis")
plt.title("Average age for each drug type")
plt.xlabel("Average Age")
plt.ylabel("Drug Type")
plt.show()
```

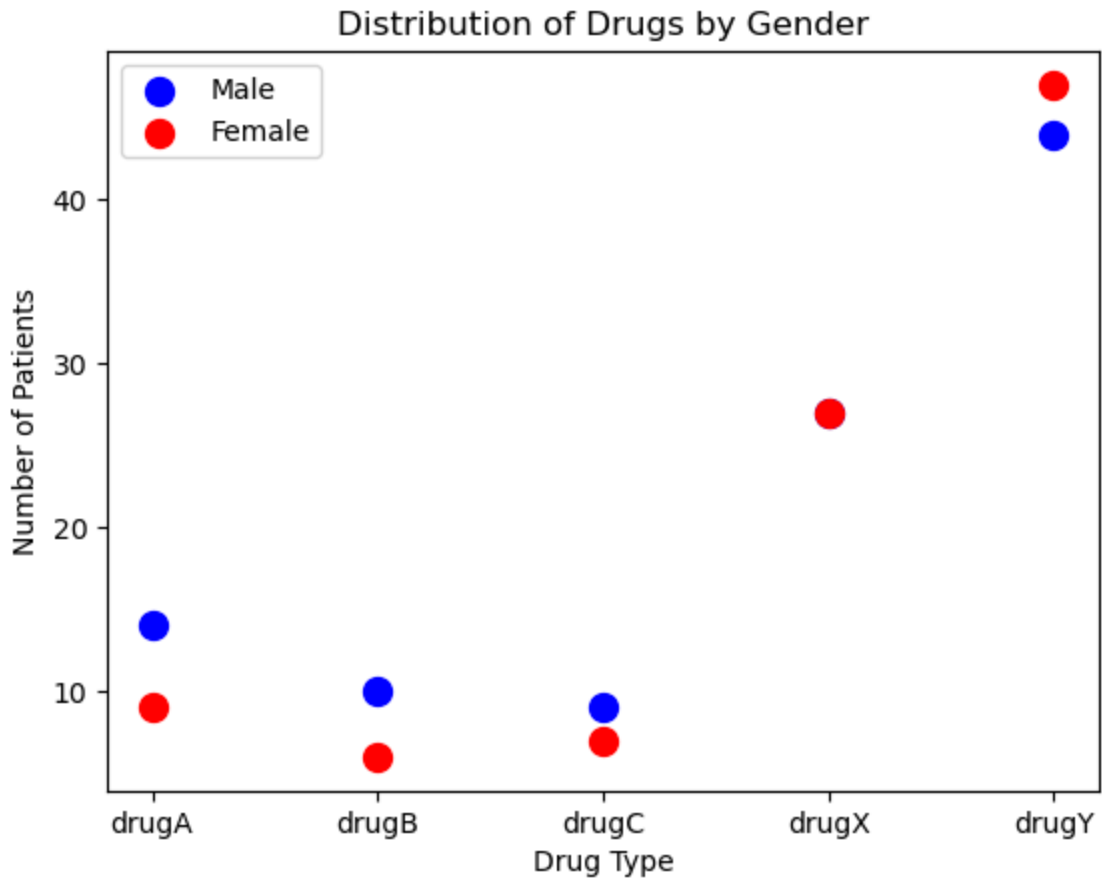
```
Drug
drugA    35.869565
drugB    62.500000
drugC    42.500000
drugX    44.018519
drugY    43.747253
Name: Age, dtype: float64
```



## Q2: What is the distribution of drugs by gender (Male/Female)?

```
In [26]: counts = drug.groupby(["Drug", "Sex"]).size().unstack()
print (counts)
plt.scatter(counts.index, counts["M"], color="blue", label="Male", s=100)
plt.scatter(counts.index, counts["F"], color="red", label="Female", s=100)
plt.title("Distribution of Drugs by Gender")
plt.xlabel("Drug Type")
plt.ylabel("Number of Patients")
plt.legend()
plt.show()
```

Sex	F	M
Drug		
drugA	9	14
drugB	6	10
drugC	7	9
drugX	27	27
drugY	47	44



**Q3: Who are the top 5 patients with the highest Na\_to\_K values, and what drugs are prescribed to them ?**

```
In [7]: drug.sort_values(['Na_to_K'], ascending=False).head()
```

```
Out[7]:
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
96	58	F	LOW	HIGH	38.247	drugY
184	18	F	HIGH	HIGH	37.188	drugY
98	20	M	HIGH	NORMAL	35.639	drugY
188	65	M	HIGH	NORMAL	34.997	drugY
194	46	F	HIGH	HIGH	34.686	drugY

**Q4: Find patients older than 50 with LOW blood pressure and their drugs**

```
In [8]: drug[(drug['Age'] > 50) & (drug['BP'] == 'LOW')]
```

Out[8]:

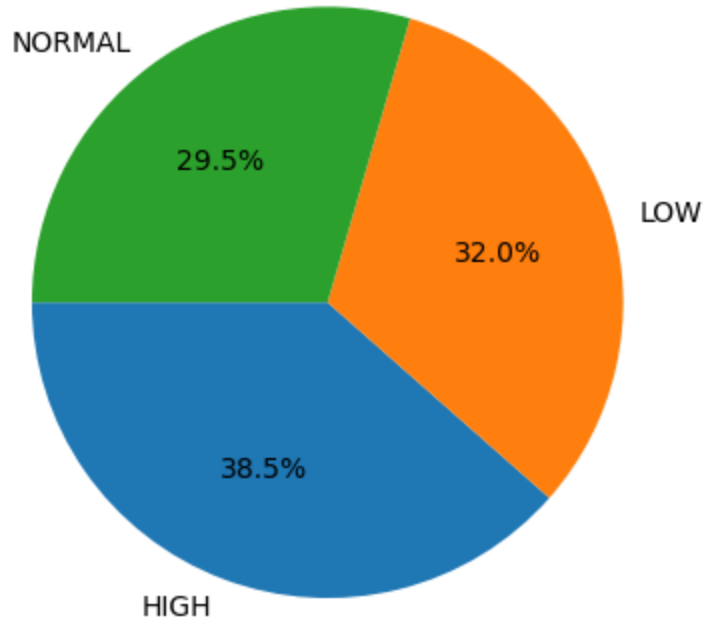
	Age	Sex	BP	Cholesterol	Na_to_K	Drug
4	61	F	LOW	HIGH	18.043	drugY
13	74	F	LOW	HIGH	20.942	drugY
16	69	M	LOW	NORMAL	11.455	drugX
20	57	M	LOW	NORMAL	19.128	drugY
47	68	M	LOW	HIGH	10.291	drugC
52	62	M	LOW	NORMAL	27.183	drugY
62	67	M	LOW	NORMAL	20.693	drugY
87	69	M	LOW	HIGH	15.478	drugY
94	56	M	LOW	HIGH	15.015	drugY
96	58	F	LOW	HIGH	38.247	drugY
113	65	F	LOW	NORMAL	13.769	drugX
131	52	M	LOW	NORMAL	32.922	drugY
135	74	M	LOW	NORMAL	11.939	drugX
141	64	F	LOW	NORMAL	25.741	drugY
148	61	F	LOW	NORMAL	7.340	drugX
153	72	F	LOW	NORMAL	14.642	drugX
157	53	M	LOW	HIGH	22.963	drugY
158	59	F	LOW	HIGH	10.444	drugC
166	58	F	LOW	HIGH	26.645	drugY
168	51	F	LOW	NORMAL	23.003	drugY
192	72	M	LOW	HIGH	16.310	drugY
193	72	M	LOW	HIGH	6.769	drugC
195	56	F	LOW	HIGH	11.567	drugC

### Q5: How many patients are there for each blood pressure category (LOW / NORMAL / HIGH)?

```
In [23]: bp_counts = drug["BP"].value_counts()
print(bp_counts)
plt.pie(bp_counts, labels=bp_counts.index, autopct='%1.1f%%', startangle=180)
plt.title("Patients by Blood Pressure Category")
plt.show()
```

BP  
HIGH 77  
LOW 64  
NORMAL 59  
Name: count, dtype: int64

Patients by Blood Pressure Category



### Q6: How does Cholesterol level (HIGH vs NORMAL) relate to drug choice?

```
In [9]: plt.figure(figsize=(7,5))
sns.countplot(data=drug, x='Cholesterol', hue='Drug')
plt.title("Drug Distribution by Cholesterol Level")
plt.xlabel("Cholesterol Level")
plt.ylabel("Number of Patients")
plt.legend(title="Drug")
plt.show()
```

