

In [1]:

```
print("Name: Jas")
print("We will be cleaning heart disease data, and conclude which age group has high risk o
print("We will also find which gender has the most not normal platelets count in blood, and
```

Name: Jas

We will be cleaning heart disease data, and conclude which age group has high risk of heart stroke as per diabetes and high blood pressure level

We will also find which gender has the most not normal platelets count in blood, and plot a pie chart around it

## Task 1 - Find the diabetic and high blood pressure patients across all age groups, and conclude the risk heart stroke is higher in which age group

In [2]:

```
#Import Libraries
import pandas as pd
import matplotlib.pyplot as plt

#read the csv
df = pd.read_csv('heart_disease.csv')
df
```

Out[2]:

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure
0	75.0	0	582	0	20	1
1	55.0	0	7861	0	38	0
2	65.0	0	146	0	20	0
3	50.0	1	111	0	20	0
4	65.0	1	160	1	20	0
...	...	...	...	...	...	...
294	62.0	0	61	1	38	1
295	55.0	0	1820	0	38	0
296	45.0	0	2060	1	60	0
297	45.0	0	2413	0	38	0
298	50.0	0	196	0	45	0

299 rows × 7 columns

In [4]:

```
#Filter by diabetes(condition will be who has diabetes) and create new dataframe

diabetic_patient=df.loc[df['diabetes']==1]
diabetic_patient
```

Out[4]:

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure
4	65.0	1	160	1	20	0
7	60.0	1	315	1	60	0
19	48.0	1	582	1	55	0
21	65.0	1	128	1	30	1
23	53.0	0	63	1	60	0
...	...	...	...	...	...	...
290	45.0	0	615	1	55	0
292	52.0	0	190	1	38	0
293	63.0	1	103	1	35	0
294	62.0	0	61	1	38	1
296	45.0	0	2060	1	60	0

125 rows × 13 columns



In [5]:

```
#On this new data frame perform group operation as per age and create new dataframe
groupby_age_diabetes=diabetic_patient.groupby('age')['diabetes'].count().reset_index()
groupby_age_diabetes
```

Out[5]:

	age	diabetes
0	40.000	2
1	42.000	2
2	44.000	2
3	45.000	11
4	46.000	1
5	48.000	2
6	49.000	1
7	50.000	12
8	51.000	2
9	52.000	2
10	53.000	3
11	54.000	1
12	55.000	4
13	56.000	1
14	58.000	5
15	59.000	3
16	60.000	19
17	60.667	2
18	61.000	2
19	62.000	3
20	63.000	4
21	65.000	15
22	66.000	1
23	68.000	2
24	69.000	1
25	70.000	9
26	72.000	2
27	73.000	1
28	75.000	3
29	80.000	3
30	82.000	2
31	90.000	1

	age	diabetes
32	94.000	1

In [7]:

```
#Filter by high_blood_pressure(condition will be who has high_blood_pressure) and create new variable patient_high_bp
patient_high_bp=df.loc[df['high_blood_pressure']==1]
patient_high_bp
```

Out[7]:

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure
0	75.0	0	582	0	20	1
5	90.0	1	47	0	40	1
9	80.0	1	123	0	35	1
10	75.0	1	81	0	38	1
11	62.0	0	231	0	25	1
...	...	...	...	...	...	...
255	52.0	1	191	1	30	1
257	58.0	0	132	1	38	1
270	44.0	0	582	1	30	1
275	45.0	0	582	0	38	1
294	62.0	0	61	1	38	1

105 rows × 7 columns



In [8]:

```
#On this new data frame perform group operation as per age and create new dataframe
groupby_age_bp=patient_high_bp.groupby('age')['high_blood_pressure'].count().reset_index()
groupby_age_bp
```

Out[8]:

	age	high_blood_pressure
0	40.000	1
1	44.000	2
2	45.000	5
3	46.000	2
4	48.000	1
5	49.000	3
6	50.000	10
7	51.000	1
8	52.000	1
9	53.000	3
10	54.000	1
11	55.000	4
12	57.000	1
13	58.000	3
14	59.000	2
15	60.000	9
16	60.667	1
17	61.000	2
18	62.000	3
19	63.000	2
20	65.000	8
21	66.000	2
22	68.000	2
23	70.000	13
24	72.000	5
25	73.000	2
26	75.000	7
27	77.000	1
28	79.000	1
29	80.000	3
30	82.000	1
31	90.000	1

	age	high_blood_pressure
32	94.000	1
33	95.000	1

In [20]:

```
#plot the scatter graph to show which age group is more prone to diabetes
```

```
plt.figure(figsize=(25,25))
```

```
diabetes=groupby_age_diabetes['diabetes']  
age=groupby_age_diabetes['age']
```

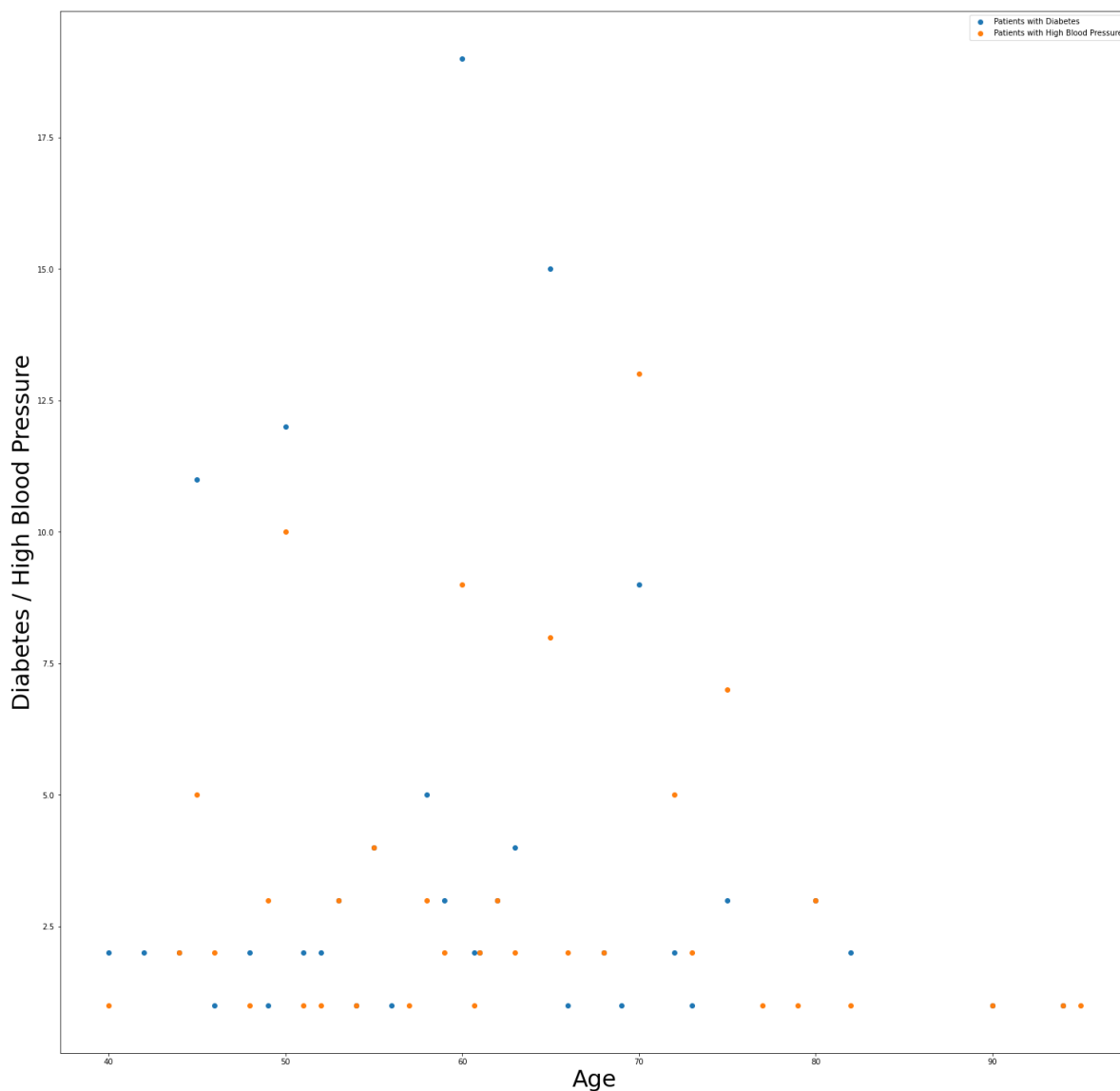
```
plt.scatter(age,diabetes,label="Patients with Diabetes")
```

```
bp=groupby_age_bp['high_blood_pressure']  
age2=groupby_age_bp['age']
```

```
plt.scatter(age2,bp,label="Patients with High Blood Pressure")
```

```
plt.xlabel("Age",size=30)  
plt.ylabel("Diabetes / High Blood Pressure",size=30)
```

```
plt.legend()  
plt.show()
```



Conclusion - When you are 60 years old you are most prone to diabetes and when you are 70 years old you are most prone to High Blood Pressure.

## **Task 2 - Find as per gender who has not normal platelets level in blood**

In [22]:

```
#Filter by platelets(condition lesser then 150000 OR greater then 450000) and create new da
platelets=df.loc[(df['platelets']<150000.00)|(df['platelets']>450000.00)]
platelets
```

Out[22]:

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure
<b>6</b>	75.0	1	246	0	15	0
<b>7</b>	60.0	1	315	1	60	0
<b>12</b>	45.0	1	981	0	30	0
<b>15</b>	82.0	1	379	0	50	0
<b>19</b>	48.0	1	582	1	55	0
<b>25</b>	80.0	0	148	1	38	0
<b>47</b>	60.0	0	582	1	38	1
<b>48</b>	80.0	1	553	0	20	1
<b>55</b>	95.0	1	371	0	30	0
<b>65</b>	60.0	0	68	0	20	0
<b>69</b>	65.0	0	113	1	25	0
<b>71</b>	58.0	0	582	1	35	0
<b>73</b>	65.0	0	224	1	50	0
<b>102</b>	80.0	0	898	0	25	0
<b>105</b>	72.0	1	328	0	30	1
<b>109</b>	45.0	0	292	1	35	0
<b>117</b>	85.0	1	102	0	60	0
<b>123</b>	60.0	1	582	0	30	1
<b>163</b>	50.0	1	2334	1	35	0
<b>167</b>	59.0	0	66	1	20	0
<b>176</b>	69.0	0	1419	0	40	0
<b>177</b>	49.0	1	69	0	50	0
<b>194</b>	45.0	0	582	0	20	1
<b>200</b>	63.0	1	1767	0	45	0
<b>211</b>	50.0	0	582	0	62	1
<b>212</b>	78.0	0	224	0	50	0
<b>223</b>	47.0	0	582	0	25	0
<b>224</b>	58.0	0	582	1	25	0
<b>227</b>	55.0	1	2794	0	35	1
<b>230</b>	60.0	0	166	0	30	0
<b>240</b>	70.0	0	81	1	35	1
<b>264</b>	61.0	0	582	1	38	0



	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure
267	56.0	1	135	1	38	0
277	70.0	0	582	1	38	0
279	55.0	0	84	1	38	0
281	70.0	0	582	0	40	0
287	45.0	0	582	1	55	0
291	60.0	0	320	0	35	0
296	45.0	0	2060	1	60	0
297	45.0	0	2413	0	38	0

In [23]:

```
#On this new dataframe perform group operation as per gender and create new dataframe

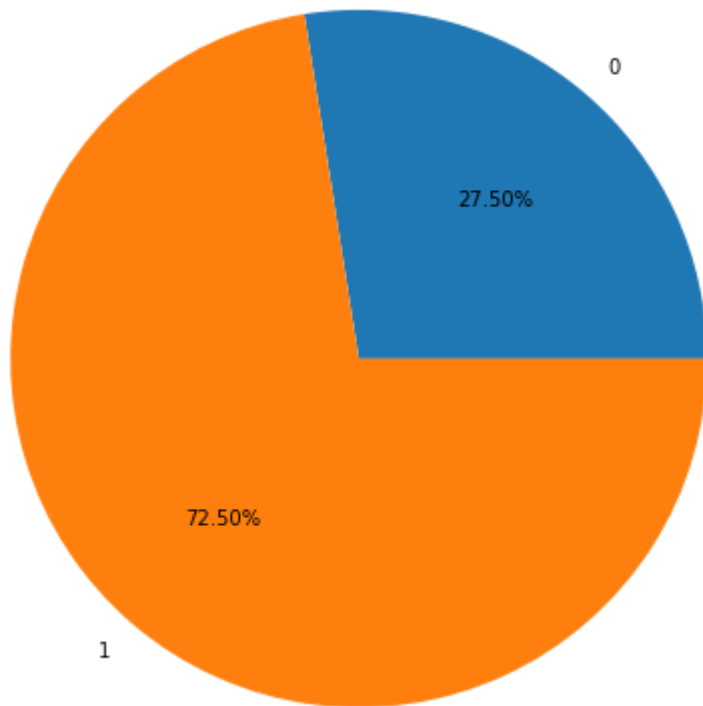
groupby_gender=platelets.groupby('gender')['platelets'].count().reset_index()
groupby_gender
```

Out[23]:

	gender	platelets
0	0	11
1	1	29

In [24]:

```
#Plot a pie chart as per the gender to show the percentage of male and female who has not n  
value=groupby_gender['platelets']  
label=groupby_gender['gender']  
  
plt.pie(value,labels=label,autopct='%0.2f%%',radius=2)  
plt.show()
```



Conclusion - The gender 1 has more not normal platelets than gender 0.

In [ ]: