

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

Read Heart.csv

```
In [4]: data = pd.read_csv('Heart.csv')
```

```
In [5]: data
```

```
Out[5]:
```

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak
0	1	63	1	typical	145	233	1	2	150	0	2.3
1	2	67	1	asymptomatic	160	286	0	2	108	1	1.5
2	3	67	1	asymptomatic	120	229	0	2	129	1	2.6
3	4	37	1	nonanginal	130	250	0	0	187	0	3.5
4	5	41	0	nontypical	130	204	0	2	172	0	1.4
...
298	299	45	1	typical	110	264	0	0	132	0	1.2
299	300	68	1	asymptomatic	144	193	1	0	141	0	3.4
300	301	57	1	asymptomatic	130	131	0	0	115	1	1.2
301	302	57	0	nontypical	130	236	0	2	174	0	0.0
302	303	38	1	nonanginal	138	175	0	0	173	0	0.0

303 rows × 15 columns



Check the shape of Dataset

```
In [6]: data.shape
```

```
Out[6]: (303, 15)
```

Check For the Null Values

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

Out[7]:

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...
298	False	False	False	False	False	False	False	False	False	False	False
299	False	False	False	False	False	False	False	False	False	False	False
300	False	False	False	False	False	False	False	False	False	False	False
301	False	False	False	False	False	False	False	False	False	False	False
302	False	False	False	False	False	False	False	False	False	False	False

303 rows × 15 columns



In [8]: `data.isna().sum()`

Out[8]:

```

Unnamed: 0      0
Age             0
Sex             0
ChestPain       0
RestBP          0
Chol            0
Fbs            0
RestECG         0
MaxHR           0
ExAng           0
Oldpeak         0
Slope           0
Ca              4
Thal            2
AHD             0
dtype: int64

```

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

Out[9]: 6

Finding The datatype of the each column

In [10]: data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 15 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   Unnamed: 0   303 non-null    int64
 1   Age          303 non-null    int64
 2   Sex          303 non-null    int64
 3   ChestPain    303 non-null    object
 4   RestBP       303 non-null    int64
 5   Chol         303 non-null    int64
 6   Fbs          303 non-null    int64
 7   RestECG      303 non-null    int64
 8   MaxHR        303 non-null    int64
 9   ExAng        303 non-null    int64
10   Oldpeak      303 non-null    float64
11   Slope        303 non-null    int64
12   Ca           299 non-null    float64
13   Thal         301 non-null    object
14   AHD          303 non-null    object
dtypes: float64(2), int64(10), object(3)
memory usage: 35.6+ KB
```

In [11]: data.dtypes

```
Out[11]: Unnamed: 0      int64
Age                int64
Sex                int64
ChestPain          object
RestBP             int64
Chol               int64
Fbs                int64
RestECG            int64
MaxHR              int64
ExAng              int64
Oldpeak            float64
Slope              int64
Ca                 float64
Thal               object
AHD                object
dtype: object
```

Finding out the Zero's

```
In [12]: (data==0).sum(axis=0)
```

```
Out[12]: Unnamed: 0      0
         Age          0
         Sex         97
         ChestPain    0
         RestBP       0
         Chol         0
         Fbs        258
         RestECG     151
         MaxHR        0
         ExAng       204
         Oldpeak      99
         Slope        0
         Ca          176
         Thal         0
         AHD          0
         dtype: int64
```

```
In [14]: #Number of zero's in row
         (data==0).sum(axis=1)
```

```
Out[14]: 0      2
         1      1
         2      1
         3      4
         4      4
         ..
         298    4
         299    2
         300    2
         301    4
         302    4
         Length: 303, dtype: int64
```

Finding out the mean age of patients

```
In [15]: data["Age"].mean()
```

```
Out[15]: 54.43894389438944
```

Extracting only the necessary Columns

```
In [16]: df = data[["Age", "Sex", "ChestPain", "RestBP", "Chol"]]
```

In [17]: df

Out[17]:

	Age	Sex	ChestPain	RestBP	Chol
0	63	1	typical	145	233
1	67	1	asymptomatic	160	286
2	67	1	asymptomatic	120	229
3	37	1	nonanginal	130	250
4	41	0	nontypical	130	204
...
298	45	1	typical	110	264
299	68	1	asymptomatic	144	193
300	57	1	asymptomatic	130	131
301	57	0	nontypical	130	236
302	38	1	nonanginal	138	175

303 rows × 5 columns

Dividing data into train amd test data set

In [18]: `from sklearn.model_selection import train_test_split`
`x_train,x_test,y_train,y_test = train_test_split(df,data,test_size=0.25)`

In [19]: x_train

Out[19]:

	Age	Sex	ChestPain	RestBP	Chol
140	59	1	nontypical	140	221
191	51	1	asymptomatic	140	298
221	54	0	nonanginal	108	267
243	61	1	typical	134	234
102	57	0	asymptomatic	128	303
...
115	41	1	nontypical	135	203
40	65	0	asymptomatic	150	225
179	53	1	nonanginal	130	246
244	60	0	nonanginal	120	178
241	41	0	nontypical	126	306

227 rows × 5 columns

In [20]: `y_train`

Out[20]:

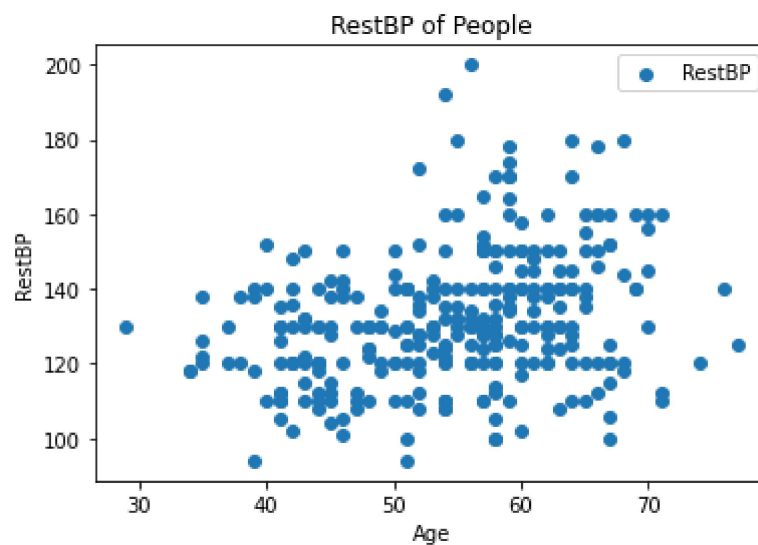
	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak
140	141	59	1	nontypical	140	221	0	0	164	1	0.0
191	192	51	1	asymptomatic	140	298	0	0	122	1	4.2
221	222	54	0	nonanginal	108	267	0	2	167	0	0.0
243	244	61	1	typical	134	234	0	0	145	0	2.6
102	103	57	0	asymptomatic	128	303	0	2	159	0	0.0
...
115	116	41	1	nontypical	135	203	0	0	132	0	0.0
40	41	65	0	asymptomatic	150	225	0	2	114	0	1.0
179	180	53	1	nonanginal	130	246	1	2	173	0	0.0
244	245	60	0	nonanginal	120	178	1	0	96	0	0.0
241	242	41	0	nontypical	126	306	0	0	163	0	0.0

227 rows × 15 columns

Data Visualization

In [21]: `import matplotlib.pyplot as plt`

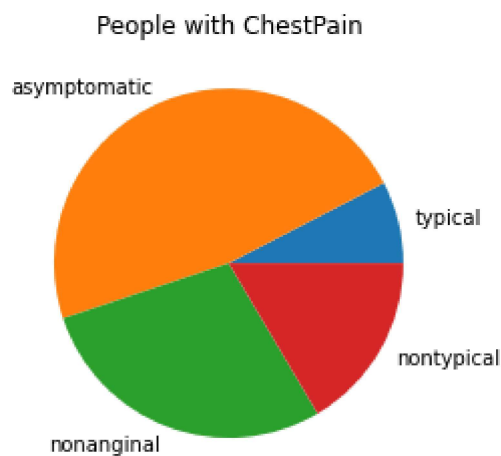
In [24]: `plt.scatter(data.Age, data.RestBP, label = "RestBP")
plt.xlabel("Age")
plt.ylabel("RestBP")
plt.title("RestBP of People")
plt.legend()
plt.show()`



```
In [28]: plt.scatter(data.Age[data.Sex == 1],data.Chol[data.Sex == 1],color='green',label='Male')
plt.scatter(data.Age[data.Sex == 0],data.Chol[data.Sex == 0],color='orange',label='Female')
plt.xlabel("Age")
plt.ylabel("Chol")
plt.title("Cholestrol of people")
plt.legend(loc="upper left")
plt.show()
```



```
In [32]: data1 = [df.ChestPain[df.ChestPain == 'typical'].count(),df.ChestPain[df.ChestPain == 'asymptomatic'].count(),df.ChestPain[df.ChestPain == 'nonanginal'].count(),df.ChestPain[df.ChestPain == 'nontypical'].count()]
label1 = ['typical','asymptomatic','nonanginal','nontypical']
plt.pie(data1,labels=label1);
plt.title("People with ChestPain")
plt.show()
```



```
In [33]: plt.hist(data.MaxHR,bins=10,color='orange')  
plt.title("Max HR")  
plt.show()
```



```
In [ ]:
```