In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
df=pd.read_csv('../pr1/Heart.csv')
```

In [4]:
```
df.shape
```

Out[4]:
```
(303, 15)
```

In [5]:
```
df.isnull()
```

Out[5]:

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

303 rows × 15 columns

In [6]:
```
df.head
```

Out[6]:
```
<bound method NDFrame.head of      Unnamed: 0  Age  Sex     ChestPain  RestBP  Chol
Fbs  RestECG  MaxHR  \
0             1   63    1       typical     145   233    1        2    150
1             2   67    1  asymptomatic     160   286    0        2    108
2             3   67    1  asymptomatic     120   229    0        2    129
3             4   37    1    nonanginal     130   250    0        0    187
4             5   41    0    nontypical     130   204    0        2    172
..          ...  ...  ...           ...     ...   ...  ...      ...    ...
298         299   45    1       typical     110   264    0        0    132
299         300   68    1  asymptomatic     144   193    1        0    141
300         301   57    1  asymptomatic     130   131    0        0    115
301         302   57    0    nontypical     130   236    0        2    174
302         303   38    1    nonanginal     138   175    0        0    173

     ExAng  Oldpeak  Slope   Ca        Thal  AHD
0        0      2.3      3  0.0       fixed   No
1        1      1.5      2  3.0      normal  Yes
2        1      2.6      2  2.0  reversable  Yes
3        0      3.5      3  0.0      normal   No
4        0      1.4      1  0.0      normal   No
..     ...      ...    ...  ...         ...  ...
298      0      1.2      2  0.0  reversable  Yes
299      0      3.4      2  2.0  reversable  Yes
300      1      1.2      2  1.0  reversable  Yes
301      0      0.0      2  1.0      normal  Yes
302      0      0.0      1  NaN      normal   No

[303 rows x 15 columns]>
```

In [7]:
```
df.columns
```

Out[7]:
```
Index(['Unnamed: 0', 'Age', 'Sex', 'ChestPain', 'RestBP', 'Chol', 'Fbs',
       'RestECG', 'MaxHR', 'ExAng', 'Oldpeak', 'Slope', 'Ca', 'Thal', 'AHD'],
      dtype='object')
```

In [8]:
```
df.dtypes
```

Out[8]:
```
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
```

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
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 [10]:
```
df["Age"].mean()
```

Out[10]:
```
54.43894389438944
```

In [11]:
```
from sklearn.model_selection import train_test_split
```

In [16]:
```
df=df[['Age','Sex','ChestPain','RestBP','Chol']]
X=df.drop('Chol',axis=1)
y=df['Chol']
print(X)
```

```
     Age  Sex     ChestPain  RestBP
0     63    1       typical     145
1     67    1  asymptomatic     160
2     67    1  asymptomatic     120
3     37    1    nonanginal     130
4     41    0    nontypical     130
..   ...  ...           ...     ...
298   45    1       typical     110
299   68    1  asymptomatic     144
300   57    1  asymptomatic     130
301   57    0    nontypical     130
302   38    1    nonanginal     138

[303 rows x 4 columns]
```

In [17]:
```
print(y)
```

```
0      233
1      286
2      229
3      250
4      204
      ...
298    264
299    193
300    131
301    236
302    175
Name: Chol, Length: 303, dtype: int64
```

In [18]:
```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.25,random_state=0)
```

In [19]:
```
X_train.shape
```

Out[19]:
```
(227, 4)
```

In [20]:
```
y_test.shape
```

Out[20]:
```
(76,)
```

In [ ]:
```

```