In [60]:
```
import pandas as pd

heart_data = pd.read_csv('Heart.csv')

heart_data.head()
```

Out[60]:

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	1	63	1	typical	145	233	1	2	150	0	2.3	3	0.0	fixed	No
1	2	67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3.0	normal	Yes
2	3	67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2.0	reversable	Yes
3	4	37	1	nonanginal	130	250	0	0	187	0	3.5	3	0.0	normal	No
4	5	41	0	nontypical	130	204	0	2	172	0	1.4	1	0.0	normal	No

In [61]:
```
heart_data.tail()
```

Out[61]:

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
298	299	45	1	typical	110	264	0	0	132	0	1.2	2	0.0	reversable	Yes
299	300	68	1	asymptomatic	144	193	1	0	141	0	3.4	2	2.0	reversable	Yes
300	301	57	1	asymptomatic	130	131	0	0	115	1	1.2	2	1.0	reversable	Yes
301	302	57	0	nontypical	130	236	0	2	174	0	0.0	2	1.0	normal	Yes
302	303	38	1	nonanginal	138	175	0	0	173	0	0.0	1	NaN	normal	No

In [62]:
```
#Dropping the unknown column
df=heart_data.drop(labels=['Unnamed: 0'],axis=1)
df
```

Out[62]:

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	63	1	typical	145	233	1	2	150	0	2.3	3	0.0	fixed	No
1	67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3.0	normal	Yes
2	67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2.0	reversable	Yes
3	37	1	nonanginal	130	250	0	0	187	0	3.5	3	0.0	normal	No
4	41	0	nontypical	130	204	0	2	172	0	1.4	1	0.0	normal	No
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	45	1	typical	110	264	0	0	132	0	1.2	2	0.0	reversable	Yes
299	68	1	asymptomatic	144	193	1	0	141	0	3.4	2	2.0	reversable	Yes
300	57	1	asymptomatic	130	131	0	0	115	1	1.2	2	1.0	reversable	Yes
301	57	0	nontypical	130	236	0	2	174	0	0.0	2	1.0	normal	Yes
302	38	1	nonanginal	138	175	0	0	173	0	0.0	1	NaN	normal	No

303 rows × 14 columns

In [63]:
```
#Shape of the data

df.shape
```

Out[63]:(303, 14)

In [64]:
```
#No of missing value
#df.isna()
#df.notna()
df.isnull().sum()
```

Out[64]:
```
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 [65]:
```
# Find datatype of each column

df.dtypes
```

Out[65]: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 [66]:
```
# Find no of zeros in each column

(df==0).sum(axis=0)
```

Out[66]: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 [67]:
```
# Find mean age of the patients

df['Age'].mean()
```

Out[67]:54.43894389438944

In [71]:
```
# Divide the dataset into 75% and 25%
#data1 = (0.75 * 1025)
#print(data1)
data1 = df.drop(columns = 'AHD')
data1 = data1.sample(frac=0.75)
data1
```

Out[71]:

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal
144	58	1	nonanginal	105	240	0	2	154	1	0.6	2	0.0	reversable
216	46	0	nontypical	105	204	0	0	172	0	0.0	1	0.0	normal
254	43	1	asymptomatic	115	303	0	0	181	0	1.2	2	0.0	normal
58	54	1	nonanginal	125	273	0	2	152	0	0.5	3	1.0	normal
146	57	1	asymptomatic	165	289	1	2	124	0	1.0	2	3.0	reversable
...	...	...	...	...	...	...	...	...	...	...	...	...	...
214	52	1	asymptomatic	112	230	0	0	160	0	0.0	1	1.0	normal
288	56	1	nontypical	130	221	0	2	163	0	0.0	1	0.0	reversable
106	59	1	asymptomatic	140	177	0	0	162	1	0.0	1	1.0	reversable
65	60	1	asymptomatic	145	282	0	2	142	1	2.8	2	2.0	reversable
163	58	0	asymptomatic	100	248	0	2	122	0	1.0	2	0.0	normal

227 rows × 13 columns

In [72]:
```
#data2 = (0.25 * 1025)
#print(data2)
data2 = df.drop(columns='AHD')
data2 = data2.sample(frac=0.25)
data2
```

Out[72]:

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal
42	71	0	nontypical	160	302	0	0	162	0	0.4	1	2.0	normal
13	44	1	nontypical	120	263	0	0	173	0	0.0	1	0.0	reversable
172	59	0	asymptomatic	174	249	0	0	143	1	0.0	2	0.0	normal
300	57	1	asymptomatic	130	131	0	0	115	1	1.2	2	1.0	reversable
189	69	1	nonanginal	140	254	0	2	146	0	2.0	2	3.0	reversable
...	...	...	...	...	...	...	...	...	...	...	...	...	...
18	48	0	nonanginal	130	275	0	0	139	0	0.2	1	0.0	normal
104	49	1	nonanginal	120	188	0	0	139	0	2.0	2	3.0	reversable
160	46	1	nontypical	101	197	1	0	156	0	0.0	1	0.0	reversable
76	60	1	asymptomatic	125	258	0	2	141	1	2.8	2	1.0	reversable
276	66	0	nonanginal	146	278	0	2	152	0	0.0	2	1.0	normal

76 rows × 13 columns

In [73]:*#Extracting the required column*

```
sub_data = df[['Age','Sex','ChestPain','RestBP','Chol']]
sub_data
```

Out[73]:

	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

In [74]:
```
# Dividing the above the data into 75%
sub_data1 = sub_data.sample(frac= 0.75)
sub_data1.count()
```

Out[74]:
```
Age         227
Sex         227
ChestPain   227
RestBP      227
Chol        227
dtype: int64
```

In [75]:
```
#Dividing the above data into 25%
sub_data2 = sub_data.sample(frac=0.25)
sub_data2.count()
```

Out[75]:
```
Age         76
Sex         76
ChestPain   76
RestBP      76
Chol        76
dtype: int64
```

In [76]:
```
#Replacing the value of Yes,No with 1,0
df['AHD'] = df['AHD'].map({'Yes':1,'No':0})
df
```

Out[76]:

	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	63	1	typical	145	233	1	2	150	0	2.3	3	0.0	fixed	0
1	67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3.0	normal	1
2	67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2.0	reversable	1
3	37	1	nonanginal	130	250	0	0	187	0	3.5	3	0.0	normal	0
4	41	0	nontypical	130	204	0	2	172	0	1.4	1	0.0	normal	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	45	1	typical	110	264	0	0	132	0	1.2	2	0.0	reversable	1
299	68	1	asymptomatic	144	193	1	0	141	0	3.4	2	2.0	reversable	1
300	57	1	asymptomatic	130	131	0	0	115	1	1.2	2	1.0	reversable	1
301	57	0	nontypical	130	236	0	2	174	0	0.0	2	1.0	normal	1
302	38	1	nonanginal	138	175	0	0	173	0	0.0	1	NaN	normal	0

303 rows × 14 columns

In [77]:
```
#importing train_test_split from sklearn
from sklearn.model_selection import train_test_split
X = df.drop(columns='AHD')
y = df['AHD']
X_train,X_test, y_train,y_test = train_test_split(X, y, train_size=0.8)
print(X_train)
print(X_test)
print(y_train)
print(y_test)
```

```
     Age  Sex     ChestPain  RestBP  Chol  Fbs  RestECG  MaxHR  ExAng  \
249   62    1    nontypical     128   208    1        2    140      0
176   52    1  asymptomatic     108   233    1        0    147      0
51    65    1  asymptomatic     120   177    0        0    140      0
131   51    1    nonanginal      94   227    0        0    154      1
293   63    1  asymptomatic     140   187    0        2    144      1
..   ...  ...           ...     ...   ...  ...      ...    ...    ...
229   66    1  asymptomatic     112   212    0        2    132      1
97    60    0  asymptomatic     150   258    0        2    157      0
231   55    0  asymptomatic     180   327    0        1    117      1
108   61    1  asymptomatic     120   260    0        0    140      1
45    58    1    nonanginal     112   230    0        2    165      0

     Oldpeak  Slope   Ca        Thal
249      0.0      1  0.0      normal
176      0.1      1  3.0  reversable
51       0.4      1  0.0  reversable
131      0.0      1  1.0  reversable
293      4.0      1  2.0  reversable
..       ...    ...  ...         ...
229      0.1      1  1.0      normal
97       2.6      2  2.0  reversable
231      3.4      2  0.0      normal
108      3.6      2  1.0  reversable
45       2.5      2  1.0  reversable

[242 rows x 13 columns]
     Age  Sex     ChestPain  RestBP  Chol  Fbs  RestECG  MaxHR  ExAng  \
159   68    1    nonanginal     118   277    0        0    151      0
279   58    0  asymptomatic     130   197    0        0    131      0
255   42    0    nonanginal     120   209    0        0    173      0
250   57    1  asymptomatic     110   201    0        0    126      1
221   54    0    nonanginal     108   267    0        2    167      0
..   ...  ...           ...     ...   ...  ...      ...    ...    ...
291   55    0    nontypical     132   342    0        0    166      0
137   62    1    nontypical     120   281    0        2    103      0
65    60    1  asymptomatic     145   282    0        2    142      1
262   60    0       typical     150   240    0        0    171      0
5     56    1    nontypical     120   236    0        0    178      0

     Oldpeak  Slope   Ca        Thal
159      1.0      1  1.0  reversable
279      0.6      2  0.0      normal
255      0.0      2  0.0      normal
250      1.5      2  0.0       fixed
221      0.0      1  0.0      normal
..       ...    ...  ...         ...
291      1.2      1  0.0      normal
137      1.4      2  1.0  reversable
```

65 2.8 2 2.0 reversable
262 0.9 1 0.0 normal
5 0.8 1 0.0 normal

[61 rows x 13 columns]
249 0
176 0
51 0
131 0
293 1
..
229 1
97 1
231 1
108 1
45 1
Name: AHD, Length: 242, dtype: int64
159 0
279 0
255 0
250 0
221 0
..
291 0
137 1
65 1
262 0
5 0
Name: AHD, Length: 61, dtype: int64
In [78]:print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
(242, 13) (61, 13) (242,) (61,)
In [ ]: