In [1]:

```python
#Roll No-33238
import pandas as pd
import numpy as np
import matplotlib as plt
```

In [2]:

```python
data=pd.read_csv('Heart(1).csv')
```

In [3]:

```python
data.head()
```

Out[3]:

	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

In [4]:

```python
data.describe()
```

Out[4]:

	Unnamed: 0	Age	Sex	RestBP	Chol	Fbs	RestECG	
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	30
mean	152.000000	54.438944	0.679868	131.689769	246.693069	0.148515	0.990099	14
std	87.612784	9.038662	0.467299	17.599748	51.776918	0.356198	0.994971	2
min	1.000000	29.000000	0.000000	94.000000	126.000000	0.000000	0.000000	7
25%	76.500000	48.000000	0.000000	120.000000	211.000000	0.000000	0.000000	13
50%	152.000000	56.000000	1.000000	130.000000	241.000000	0.000000	1.000000	15
75%	227.500000	61.000000	1.000000	140.000000	275.000000	0.000000	2.000000	16
max	303.000000	77.000000	1.000000	200.000000	564.000000	1.000000	2.000000	20

In [5]:

```python
p=list(data.columns)
```

In [6]:

```python
print(p)
```

```
['Unnamed: 0', 'Age', 'Sex', 'ChestPain', 'RestBP', 'Chol', 'Fbs', 'Re
stECG', 'MaxHR', 'ExAng', 'Oldpeak', 'Slope', 'Ca', 'Thal', 'AHD']
```

In [7]:

```python
data.shape
```

Out[7]:

```
(303, 15)
```

In [8]:

```python
for i in range(len(p)):
    a=p[i]
    print(p[i],data[a].isna().sum())
```

```
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
```

In [9]:

```python
data.isna().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
```

```
data.dtypes
```

```
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
```

```
np.mean(data['Age'])
```

```
54.43894389438944
```

```
(data==0).sum()
```

```
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 [13]:

```python
(data==0).sum().sum()
```

Out[13]:

985

In [14]:

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

In [15]:

```python
df.head()
```

Out[15]:

	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

In [16]:

```python
from sklearn.model_selection import train_test_split
```

In [17]:

```python
train,test= train_test_split(df, test_size=0.25, random_state=42)
```

In [18]:

```python
train.head()
```

Out[18]:

	Age	Sex	ChestPain	RestBP	Chol
287	58	1	nontypical	125	220
282	55	0	asymptomatic	128	205
197	45	0	asymptomatic	138	236
158	60	1	asymptomatic	140	293
164	48	1	nonanginal	124	255

In [19]:

```python
train.shape
```

Out[19]:

(227, 5)

In [20]:

```
test.shape
```

Out[20]:

```
(76, 5)
```

In [ ]: