

## importing the required libraries

```
In [41]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

## reading the csv file in data variable

```
In [42]: data=pd.read_csv(r"C:\Users\lpooj\Downloads\Heart.csv")
```

## display first five rows of dataset

```
In [43]: data.head()
```

```
Out[43]:
```

	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	

## display last ten rows of dataset

```
In [44]: data.tail(10)
```

```
Out[44]:
```

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	
293	294	63	1	asymptomatic	140	187	0	2	144	1	4.0	
294	295	63	0	asymptomatic	124	197	0	0	136	1	0.0	
295	296	41	1	nontypical	120	157	0	0	182	0	0.0	
296	297	59	1	asymptomatic	164	176	1	2	90	0	1.0	
297	298	57	0	asymptomatic	140	241	0	0	123	1	0.2	
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	

## display the shape of the dataset i.e (row,columns)

```
In [45]: data.shape
```

```
Out[45]: (303, 15)
```

## describes the dataset

```
In [46]: data.describe()
```

```
Out[46]:
```

	Unnamed: 0	Age	Sex	RestBP	Chol	Fbs	RestECG	Mi
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	152.000000	54.438944	0.679868	131.689769	246.693069	0.148515	0.990099	149.600000
std	87.612784	9.038662	0.467299	17.599748	51.776918	0.356198	0.994971	22.800000
min	1.000000	29.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000
25%	76.500000	48.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000
50%	152.000000	56.000000	1.000000	130.000000	241.000000	0.000000	1.000000	153.000000
75%	227.500000	61.000000	1.000000	140.000000	275.000000	0.000000	2.000000	166.000000
max	303.000000	77.000000	1.000000	200.000000	564.000000	1.000000	2.000000	202.000000

## display the information about dataset

```
In [47]: 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
```

## find the missing values in dataset

In [48]: `data.isnull()`

Out[48]:

	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

## find the sum of missing values in the dataset

In [49]: `data.isnull().sum()`

```
Out[49]: 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
```

## display the data types of the dataset

```
In [50]: data.dtypes
```

```
Out[50]: 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
```

## finds the minimum value of the particular row in the dataset

```
In [51]: data['Age'].min()
```

```
Out[51]: 29
```

## finds the maximum value of the particular row in the dataset

```
In [52]: data['Sex'].max()
```

```
Out[52]: 1
```

## finds the unique value of the particular row

## in the dataset

```
In [53]: data['ChestPain'].unique()
```

```
Out[53]: array(['typical', 'asymptomatic', 'nonanginal', 'nontypical'],
      dtype=object)
```

## used to rename the particular column

```
In [54]: data['MaxHR'].rename('MaxHRnew')
```

```
Out[54]: 0      150
1      108
2      129
3      187
4      172
...
298    132
299    141
300    115
301    174
302    173
Name: MaxHRnew, Length: 303, dtype: int64
```

## finding out the zeros in the dataset

```
In [55]: for column_name in data.columns:
      column=data[column_name]
      count=(column==0).sum()
      print("count of zeros in columns",column_name,"is : ",count)
```

```
count of zeros in columns Unnamed: 0 is : 0
count of zeros in columns Age is : 0
count of zeros in columns Sex is : 97
count of zeros in columns ChestPain is : 0
count of zeros in columns RestBP is : 0
count of zeros in columns Chol is : 0
count of zeros in columns Fbs is : 258
count of zeros in columns RestECG is : 151
count of zeros in columns MaxHR is : 0
count of zeros in columns ExAng is : 204
count of zeros in columns Oldpeak is : 99
count of zeros in columns Slope is : 0
count of zeros in columns Ca is : 176
count of zeros in columns Thal is : 0
count of zeros in columns AHD is : 0
```

```
In [56]: data['AHD']
```

```
Out[56]: 0      No
          1      Yes
          2      Yes
          3      No
          4      No
          ...
          298    Yes
          299    Yes
          300    Yes
          301    Yes
          302    No
          Name: AHD, Length: 303, dtype: object
```

## mapping 'yes' and 'no' by '1' and '0' in the AHD column of the dataset

```
In [57]: data['AHD']=data['AHD'].map({'Yes':1,'No':0})
```

```
In [58]: data['AHD']
```

```
Out[58]: 0      0
          1      1
          2      1
          3      0
          4      0
          ..
          298    1
          299    1
          300    1
          301    1
          302    0
          Name: AHD, Length: 303, dtype: int64
```

## display the names of the columns in the dataset

```
In [59]: data.columns
```

```
Out[59]: Index(['Unnamed: 0', 'Age', 'Sex', 'ChestPain', 'RestBP', 'Chol', 'Fbs',
                'RestECG', 'MaxHR', 'ExAng', 'Oldpeak', 'Slope', 'Ca', 'Thal', 'AHD'],
                dtype='object')
```

```
In [ ]:
```