Heart Attack

Import Libraries

In [5]:

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
dt=pd.read_csv("C:\ML_projectNew\HeartAttack.csv",na_values='?')   #the data set
dt.head(10)
```

Out[5]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
0	28	1	2	130.0	132.0	0.0	2.0	185.0	0.0	0.0	NaN	NaN	NaN	0
1	29	1	2	120.0	243.0	0.0	0.0	160.0	0.0	0.0	NaN	NaN	NaN	0
2	29	1	2	140.0	NaN	0.0	0.0	170.0	0.0	0.0	NaN	NaN	NaN	0
3	30	0	1	170.0	237.0	0.0	1.0	170.0	0.0	0.0	NaN	NaN	6.0	0
4	31	0	2	100.0	219.0	0.0	1.0	150.0	0.0	0.0	NaN	NaN	NaN	0
5	32	0	2	105.0	198.0	0.0	0.0	165.0	0.0	0.0	NaN	NaN	NaN	0
6	32	1	2	110.0	225.0	0.0	0.0	184.0	0.0	0.0	NaN	NaN	NaN	0
7	32	1	2	125.0	254.0	0.0	0.0	155.0	0.0	0.0	NaN	NaN	NaN	0
8	33	1	3	120.0	298.0	0.0	0.0	185.0	0.0	0.0	NaN	NaN	NaN	0
9	34	0	2	130.0	161.0	0.0	0.0	190.0	0.0	0.0	NaN	NaN	NaN	0

In [6]:

```python
dt.tail(10)
```

Out[6]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
284	49	1	4	128.0	212.0	0.0	0.0	96.0	1.0	0.0	NaN	NaN	NaN	1
285	49	1	4	150.0	222.0	0.0	0.0	122.0	0.0	2.0	2.0	NaN	NaN	1
286	50	1	4	140.0	231.0	0.0	1.0	140.0	1.0	5.0	2.0	NaN	NaN	1
287	50	1	4	140.0	341.0	0.0	1.0	125.0	1.0	2.5	2.0	NaN	NaN	1
288	52	1	4	140.0	266.0	0.0	0.0	134.0	1.0	2.0	2.0	NaN	NaN	1
289	52	1	4	160.0	331.0	0.0	0.0	94.0	1.0	2.5	NaN	NaN	NaN	1
290	54	0	3	130.0	294.0	0.0	1.0	100.0	1.0	0.0	2.0	NaN	NaN	1
291	56	1	4	155.0	342.0	1.0	0.0	150.0	1.0	3.0	2.0	NaN	NaN	1
292	58	0	2	180.0	393.0	0.0	0.0	110.0	1.0	1.0	2.0	NaN	7.0	1
293	65	1	4	130.0	275.0	0.0	1.0	115.0	1.0	1.0	2.0	NaN	NaN	1

information of the dataset

In [7]:

```python
dt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 294 entries, 0 to 293
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   age         294 non-null    int64  
 1   sex         294 non-null    int64  
 2   cp          294 non-null    int64  
 3   trestbps    293 non-null    float64
 4   chol        271 non-null    float64
 5   fbs         286 non-null    float64
 6   restecg     293 non-null    float64
 7   thalach     293 non-null    float64
 8   exang       293 non-null    float64
 9   oldpeak     294 non-null    float64
 10  slope       104 non-null    float64
 11  ca          3 non-null      float64
 12  thal        28 non-null     float64
 13  num         294 non-null    int64  
dtypes: float64(10), int64(4)
memory usage: 32.3 KB
```

In [8]:

```python
dt.isnull().sum()
```

Out[8]:

```
age             0
sex             0
cp              0
trestbps        1
chol           23
fbs             8
restecg         1
thalach         1
exang           1
oldpeak         0
slope         190
ca            291
thal          266
num             0
dtype: int64
```

In [9]:

```python
dt.describe()
```

Out[9]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
count	294.000000	294.000000	294.000000	293.000000	271.000000	286.000000	293.000000	293.000000	293.000000	294.000000	104.000000	3.0	28.000000	294.000000
mean	47.826531	0.724490	2.982993	132.583618	250.848708	0.069930	0.218430	139.129693	0.303754	0.586054	1.894231	0.0	5.642857	0.360544
std	7.811812	0.447533	0.965117	17.626568	67.657711	0.255476	0.460868	23.589749	0.460665	0.908648	0.338995	0.0	1.615074	0.480977
min	28.000000	0.000000	1.000000	92.000000	85.000000	0.000000	0.000000	82.000000	0.000000	0.000000	1.000000	0.0	3.000000	0.000000
25%	42.000000	0.000000	2.000000	120.000000	209.000000	0.000000	0.000000	122.000000	0.000000	0.000000	2.000000	0.0	5.250000	0.000000
50%	49.000000	1.000000	3.000000	130.000000	243.000000	0.000000	0.000000	140.000000	0.000000	0.000000	2.000000	0.0	6.000000	0.000000
75%	54.000000	1.000000	4.000000	140.000000	282.500000	0.000000	0.000000	155.000000	1.000000	1.000000	2.000000	0.0	7.000000	1.000000
max	66.000000	1.000000	4.000000	200.000000	603.000000	1.000000	2.000000	190.000000	1.000000	5.000000	3.000000	0.0	7.000000	1.000000

In [10]:

```python
dt.head()
```

Out[10]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
0	28	1	2	130.0	132.0	0.0	2.0	185.0	0.0	0.0	NaN	NaN	NaN	0
1	29	1	2	120.0	243.0	0.0	0.0	160.0	0.0	0.0	NaN	NaN	NaN	0
2	29	1	2	140.0	NaN	0.0	0.0	170.0	0.0	0.0	NaN	NaN	NaN	0
3	30	0	1	170.0	237.0	0.0	1.0	170.0	0.0	0.0	NaN	NaN	6.0	0
4	31	0	2	100.0	219.0	0.0	1.0	150.0	0.0	0.0	NaN	NaN	NaN	0

In [11]:

```python
    dt=dt.drop(columns=["slope","ca","thal"],axis=1)
```

In [12]:

```python
dt.head()
```

Out[12]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	num
0	28	1	2	130.0	132.0	0.0	2.0	185.0	0.0	0.0	0
1	29	1	2	120.0	243.0	0.0	0.0	160.0	0.0	0.0	0
2	29	1	2	140.0	NaN	0.0	0.0	170.0	0.0	0.0	0
3	30	0	1	170.0	237.0	0.0	1.0	170.0	0.0	0.0	0
4	31	0	2	100.0	219.0	0.0	1.0	150.0	0.0	0.0	0

In [13]:

```python
dt=dt.dropna()
```

In [14]:

```python
dt.head()
```

Out[14]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	num
0	28	1	2	130.0	132.0	0.0	2.0	185.0	0.0	0.0	0
1	29	1	2	120.0	243.0	0.0	0.0	160.0	0.0	0.0	0
3	30	0	1	170.0	237.0	0.0	1.0	170.0	0.0	0.0	0
4	31	0	2	100.0	219.0	0.0	1.0	150.0	0.0	0.0	0
5	32	0	2	105.0	198.0	0.0	0.0	165.0	0.0	0.0	0

In [15]:

```python
dt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 261 entries, 0 to 293
Data columns (total 11 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   age         261 non-null    int64  
 1   sex         261 non-null    int64  
 2   cp          261 non-null    int64  
 3   trestbps    261 non-null    float64
 4   chol        261 non-null    float64
 5   fbs         261 non-null    float64
 6   restecg     261 non-null    float64
 7   thalach     261 non-null    float64
 8   exang       261 non-null    float64
 9   oldpeak     261 non-null    float64
 10  num         261 non-null    int64  
dtypes: float64(7), int64(4)
memory usage: 24.5 KB
```

In [16]:

```python
dt.isnull().sum()
```

Out[16]:

```
age           0
sex           0
cp            0
trestbps      0
chol          0
fbs           0
restecg       0
thalach       0
exang         0
oldpeak       0
num           0
dtype: int64
```

In [17]:

```python
dt["sex"].value_counts()
```

Out[17]:

```
1    192
0     69
Name: sex, dtype: int64
```

In [18]:

```python
dt["cp"].value_counts()
```

Out[18]:

```
4    113
2     92
3     46
1     10
Name: cp, dtype: int64
```

In [19]:

```python
dt["restecg"].value_counts()
```

Out[19]:

```
0.0    208
1.0     47
2.0      6
Name: restecg, dtype: int64
```

In [20]:

```python
dt["fbs"].value_counts()
```

Out[20]:

```
0.0    242
1.0     19
Name: fbs, dtype: int64
```

Dummies using Pandas to convert categorical value to one-hot encoding

In [21]:

```python
dt=pd.get_dummies(dt,columns=["cp","restecg"])
```

In [22]:

```python
dt.head()
```

Out[22]:

	age	sex	trestbps	chol	fbs	thalach	exang	oldpeak	num	cp_1	cp_2	cp_3	cp_4	restecg_0.0	restecg_1.0	restecg_2.0
0	28	1	130.0	132.0	0.0	185.0	0.0	0.0	0	0	1	0	0	0	0	1
1	29	1	120.0	243.0	0.0	160.0	0.0	0.0	0	0	1	0	0	1	0	0
3	30	0	170.0	237.0	0.0	170.0	0.0	0.0	0	1	0	0	0	0	1	0
4	31	0	100.0	219.0	0.0	150.0	0.0	0.0	0	0	1	0	0	0	1	0
5	32	0	105.0	198.0	0.0	165.0	0.0	0.0	0	0	1	0	0	1	0	0

In [23]:

```python
dt.columns
```

Out[23]:

```
Index(['age', 'sex', 'trestbps', 'chol', 'fbs', 'thalach', 'exang', 'oldpeak',
       'num        ', 'cp_1', 'cp_2', 'cp_3', 'cp_4', 'restecg_0.0',
       'restecg_1.0', 'restecg_2.0'],
      dtype='object')
```

In [24]:

```python
dt=dt.rename(columns = {"num       ":"target"})
```

In [32]:

```python
dt.head()
```

Out[32]:

	age	sex	trestbps	chol	fbs	thalach	exang	oldpeak	target	cp_1	cp_2	cp_3	cp_4	restecg_0.0	restecg_1.0	restecg_2.0
0	28	1	130.0	132.0	0.0	185.0	0.0	0.0	0	0	1	0	0	0	0	1
1	29	1	120.0	243.0	0.0	160.0	0.0	0.0	0	0	1	0	0	1	0	0
3	30	0	170.0	237.0	0.0	170.0	0.0	0.0	0	1	0	0	0	0	1	0
4	31	0	100.0	219.0	0.0	150.0	0.0	0.0	0	0	1	0	0	0	1	0
5	32	0	105.0	198.0	0.0	165.0	0.0	0.0	0	0	1	0	0	1	0	0

In [29]:

```python
numerical_cols=["age","trestbps","chol","thalach","oldpeak"]
cat_cols=list(set(dt.columns) - set(numerical_cols)-{"target"})
```

In [33]:

```python
numerical_cols
```

Out[33]:

```
['age', 'trestbps', 'chol', 'thalach', 'oldpeak']
```

In [34]:

```python
cat_cols
```

Out[34]:

```
['restecg_2.0',
 'cp_1',
 'fbs',
 'cp_4',
 'exang',
 'restecg_0.0',
 'restecg_1.0',
 'cp_3',
 'cp_2',
 'sex']
```

In [ ]: