In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:

```
#-------------------read data---------------------------#
df = pd.read_csv('../DSBDAFINAL/heart_desease_data.csv')
```

In [3]:

```
df
```

Out[3]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [4]:

```
df.head()
```

Out[4]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |

In [5]:

```
df.tail()
```

Out[5]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

In [6]:

```
df.shape
```

Out[6]:

```
(303, 14)
```

In [7]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   age       303 non-null    int64  
 1   sex       303 non-null    int64  
 2   cp        303 non-null    int64  
 3   trestbps  303 non-null    int64  
 4   chol      303 non-null    int64  
 5   fbs       303 non-null    int64  
 6   restecg   303 non-null    int64  
 7   thalach   303 non-null    int64  
 8   exang     303 non-null    int64  
 9   oldpeak   303 non-null    float64
 10  slope     303 non-null    int64  
 11  ca        303 non-null    int64  
 12  thal      303 non-null    int64  
 13  target    303 non-null    int64  
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

In [8]:
```
df.describe()
```

Out[8]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 303.000000 | 303.000000 | 303.000000 | 303.000000 | 303.000000 | 303.000000 | 303.000000 | 303.000000 | 303.000000 | 303.000000 | 303.000000 |
| **mean** | 54.366337 | 0.683168 | 0.966997 | 131.623762 | 246.264026 | 0.148515 | 0.528053 | 149.646865 | 0.326733 | 1.039604 | 1.399340 |
| **std** | 9.082101 | 0.466011 | 1.032052 | 17.538143 | 51.830751 | 0.356198 | 0.525860 | 22.905161 | 0.469794 | 1.161075 | 0.616226 |
| **min** | 29.000000 | 0.000000 | 0.000000 | 94.000000 | 126.000000 | 0.000000 | 0.000000 | 71.000000 | 0.000000 | 0.000000 | 0.000000 |
| **25%** | 47.500000 | 0.000000 | 0.000000 | 120.000000 | 211.000000 | 0.000000 | 0.000000 | 133.500000 | 0.000000 | 0.000000 | 1.000000 |
| **50%** | 55.000000 | 1.000000 | 1.000000 | 130.000000 | 240.000000 | 0.000000 | 1.000000 | 153.000000 | 0.000000 | 0.800000 | 1.000000 |
| **75%** | 61.000000 | 1.000000 | 2.000000 | 140.000000 | 274.500000 | 0.000000 | 1.000000 | 166.000000 | 1.000000 | 1.600000 | 2.000000 |
| **max** | 77.000000 | 1.000000 | 3.000000 | 200.000000 | 564.000000 | 1.000000 | 2.000000 | 202.000000 | 1.000000 | 6.200000 | 2.000000 |

In [9]:
```
df
```

Out[9]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [10]:
```
#----------------------------Data cleaning----------------------------#
```

In [12]:
```
#Drop the rows of missing values
df.dropna(inplace=True)
```

In [13]:
```
df
```

Out[13]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [14]:

```
#Removing duplication
df.drop_duplicates(inplace=True)
```

In [15]:

```
df
```

Out[15]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

302 rows × 14 columns

In [16]:

```
#data type conversion
#convert columns to the appropriate data types needed
df['age'] = df['age'].astype(int)
df['chol']=df['chol'].astype(float)
```

In [17]:

```
df.to_csv("cleaned_heart_diseases_data.csv",index=False)
```

In [18]:

```
df1=pd.read_csv("cleaned_heart_diseases_data.csv")
```

In [19]:

```
df1
```

Out[19]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233.0 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250.0 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204.0 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236.0 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354.0 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **297** | 57 | 0 | 0 | 140 | 241.0 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **298** | 45 | 1 | 3 | 110 | 264.0 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **299** | 68 | 1 | 0 | 144 | 193.0 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **300** | 57 | 1 | 0 | 130 | 131.0 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **301** | 57 | 0 | 1 | 130 | 236.0 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

302 rows × 14 columns

In [20]:

```
#-------------------------Data Integration---------------------------#
```

In [21]:

```
df=pd.read_csv("../DSBDAFINAL/heart_desease_data.csv")
```

In [22]:

```
df
```

Out[22]:

|  | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [23]:

```
#create subset of the data based on specific condition
subset1 = df[df['age'] < 45]
subset2 = df[df['age'] >= 45]
```

In [26]:

```
subset1
```

Out[26]:

|  | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **7** | 44 | 1 | 1 | 120 | 263 | 0 | 1 | 173 | 0 | 0.0 | 2 | 0 | 3 | 1 |
| **18** | 43 | 1 | 0 | 150 | 247 | 0 | 1 | 171 | 0 | 1.5 | 2 | 0 | 2 | 1 |
| **21** | 44 | 1 | 2 | 130 | 233 | 0 | 1 | 179 | 1 | 0.4 | 2 | 0 | 2 | 1 |
| **22** | 42 | 1 | 0 | 140 | 226 | 0 | 1 | 178 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **24** | 40 | 1 | 3 | 140 | 199 | 0 | 1 | 178 | 1 | 1.4 | 2 | 0 | 3 | 1 |
| **30** | 41 | 0 | 1 | 105 | 198 | 0 | 1 | 168 | 0 | 0.0 | 2 | 1 | 2 | 1 |
| **32** | 44 | 1 | 1 | 130 | 219 | 0 | 0 | 188 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **44** | 39 | 1 | 2 | 140 | 321 | 0 | 0 | 182 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **46** | 44 | 1 | 2 | 140 | 235 | 0 | 0 | 180 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **53** | 44 | 0 | 2 | 108 | 141 | 0 | 1 | 175 | 0 | 0.6 | 1 | 0 | 2 | 1 |
| **58** | 34 | 1 | 3 | 118 | 182 | 0 | 0 | 174 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **63** | 41 | 1 | 1 | 135 | 203 | 0 | 1 | 132 | 0 | 0.0 | 1 | 0 | 1 | 1 |
| **65** | 35 | 0 | 0 | 138 | 183 | 0 | 1 | 182 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **68** | 44 | 1 | 1 | 120 | 220 | 0 | 1 | 170 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **72** | 29 | 1 | 1 | 130 | 204 | 0 | 0 | 202 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **74** | 43 | 0 | 2 | 122 | 213 | 0 | 1 | 165 | 0 | 0.2 | 1 | 0 | 2 | 1 |
| **80** | 41 | 1 | 2 | 112 | 250 | 0 | 1 | 179 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **84** | 42 | 0 | 0 | 102 | 265 | 0 | 0 | 122 | 0 | 0.6 | 1 | 0 | 2 | 1 |
| **98** | 43 | 1 | 2 | 130 | 315 | 0 | 1 | 162 | 0 | 1.9 | 2 | 1 | 2 | 1 |
| **100** | 42 | 1 | 3 | 148 | 244 | 0 | 0 | 178 | 0 | 0.8 | 2 | 2 | 2 | 1 |
| **103** | 42 | 1 | 2 | 120 | 240 | 1 | 1 | 194 | 0 | 0.8 | 0 | 0 | 3 | 1 |
| **113** | 43 | 1 | 0 | 110 | 211 | 0 | 1 | 161 | 0 | 0.0 | 2 | 0 | 3 | 1 |
| **115** | 37 | 0 | 2 | 120 | 215 | 0 | 1 | 170 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **116** | 41 | 1 | 2 | 130 | 214 | 0 | 0 | 168 | 0 | 2.0 | 1 | 0 | 2 | 1 |
| **122** | 41 | 0 | 2 | 112 | 268 | 0 | 0 | 172 | 1 | 0.0 | 2 | 0 | 2 | 1 |
| **124** | 39 | 0 | 2 | 94 | 199 | 0 | 1 | 179 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **125** | 34 | 0 | 1 | 118 | 210 | 0 | 1 | 192 | 0 | 0.7 | 2 | 0 | 2 | 1 |
| **132** | 42 | 1 | 1 | 120 | 295 | 0 | 1 | 162 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **133** | 41 | 1 | 1 | 110 | 235 | 0 | 1 | 153 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **134** | 41 | 0 | 1 | 126 | 306 | 0 | 1 | 163 | 0 | 0.0 | 2 | 0 | 2 | 1 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **141** | 43 | 1 | 0 | 115 | 303 | 0 | 1 | 181 | 0 | 1.2 | 1 | 0 | 2 | 1 |
| **142** | 42 | 0 | 2 | 120 | 209 | 0 | 1 | 173 | 0 | 0.0 | 1 | 0 | 2 | 1 |
| **146** | 44 | 0 | 2 | 118 | 242 | 0 | 1 | 149 | 0 | 0.3 | 1 | 1 | 2 | 1 |
| **148** | 44 | 1 | 2 | 120 | 226 | 0 | 1 | 169 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **149** | 42 | 1 | 2 | 130 | 180 | 0 | 1 | 150 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **154** | 39 | 0 | 2 | 138 | 220 | 0 | 1 | 152 | 0 | 0.0 | 1 | 0 | 2 | 1 |
| **157** | 35 | 1 | 1 | 122 | 192 | 0 | 1 | 174 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **162** | 41 | 1 | 1 | 120 | 157 | 0 | 1 | 182 | 0 | 0.0 | 2 | 0 | 2 | 1 |
| **163** | 38 | 1 | 2 | 138 | 175 | 0 | 1 | 173 | 0 | 0.0 | 2 | 4 | 2 | 1 |
| **164** | 38 | 1 | 2 | 138 | 175 | 0 | 1 | 173 | 0 | 0.0 | 2 | 4 | 2 | 1 |
| **175** | 40 | 1 | 0 | 110 | 167 | 0 | 0 | 114 | 1 | 2.0 | 1 | 0 | 3 | 0 |
| **178** | 43 | 1 | 0 | 120 | 177 | 0 | 0 | 120 | 1 | 2.5 | 1 | 0 | 3 | 0 |
| **185** | 44 | 1 | 0 | 112 | 290 | 0 | 0 | 153 | 0 | 0.0 | 2 | 1 | 2 | 0 |
| **189** | 41 | 1 | 0 | 110 | 172 | 0 | 0 | 158 | 0 | 0.0 | 2 | 0 | 3 | 0 |
| **200** | 44 | 1 | 0 | 110 | 197 | 0 | 0 | 177 | 0 | 0.0 | 2 | 1 | 2 | 0 |
| **212** | 39 | 1 | 0 | 118 | 219 | 0 | 1 | 140 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **215** | 43 | 0 | 0 | 132 | 341 | 1 | 0 | 136 | 1 | 3.0 | 1 | 0 | 3 | 0 |
| **227** | 35 | 1 | 0 | 120 | 198 | 0 | 1 | 130 | 1 | 1.6 | 1 | 0 | 3 | 0 |
| **239** | 35 | 1 | 0 | 126 | 282 | 0 | 0 | 156 | 1 | 0.0 | 2 | 0 | 3 | 0 |
| **251** | 43 | 1 | 0 | 132 | 247 | 1 | 0 | 143 | 1 | 0.1 | 1 | 4 | 3 | 0 |
| **259** | 38 | 1 | 3 | 120 | 231 | 0 | 1 | 182 | 1 | 3.8 | 1 | 0 | 3 | 0 |
| **280** | 42 | 1 | 0 | 136 | 315 | 0 | 1 | 125 | 1 | 1.8 | 1 | 0 | 1 | 0 |
| **283** | 40 | 1 | 0 | 152 | 223 | 0 | 1 | 181 | 0 | 0.0 | 2 | 0 | 3 | 0 |
| **294** | 44 | 1 | 0 | 120 | 169 | 0 | 1 | 144 | 1 | 2.8 | 0 | 0 | 1 | 0 |

In [27]:

```
subset2
```

Out[27]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **5** | 57 | 1 | 0 | 140 | 192 | 0 | 1 | 148 | 0 | 0.4 | 1 | 0 | 1 | 1 |
| **6** | 56 | 0 | 1 | 140 | 294 | 0 | 0 | 153 | 0 | 1.3 | 1 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

247 rows × 14 columns

In [ ]:

In [24]:

```
merged_data =pd.concat([subset1, subset2],ignore_index=True)
```

In [25]:

```
merged_data
```

Out[25]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **1** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **2** | 44 | 1 | 1 | 120 | 263 | 0 | 1 | 173 | 0 | 0.0 | 2 | 0 | 3 | 1 |
| **3** | 43 | 1 | 0 | 150 | 247 | 0 | 1 | 171 | 0 | 1.5 | 2 | 0 | 2 | 1 |
| **4** | 44 | 1 | 2 | 130 | 233 | 0 | 1 | 179 | 1 | 0.4 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [30]:

```
merged_data.to_csv("integrated_heart_desease_data", index=False)
```

In [31]:

```
df2=pd.read_csv("integrated_heart_desease_data")
```

In [32]:

```
df2
```

Out[32]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **1** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **2** | 44 | 1 | 1 | 120 | 263 | 0 | 1 | 173 | 0 | 0.0 | 2 | 0 | 3 | 1 |
| **3** | 43 | 1 | 0 | 150 | 247 | 0 | 1 | 171 | 0 | 1.5 | 2 | 0 | 2 | 1 |
| **4** | 44 | 1 | 2 | 130 | 233 | 0 | 1 | 179 | 1 | 0.4 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [33]:

```
#-----------------------Data transformation------------------#
```

In [34]:

```
df=pd.read_csv('../DSBDAFINAL/heart_desease_data.csv')
```

In [35]:

```
df
```

Out[35]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [37]:

```
#Renaming columns
df=df.rename(columns={'age':'Age','sex':'Gender','cp':'chestpain'})
```

In [38]:

```
df
```

Out[38]:

| | Age | Gender | chestpain | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [41]:
```
#mapping the categorial values
df['Gender']=df['Gender'].map({0:'Female', 1:'Male'})
```

In [42]:
```
df
```

Out[42]:

| | Age | Gender | chestpain | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | Male | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | Male | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | Female | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | Male | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | Female | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | Female | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | Male | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | Male | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | Male | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | Female | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [53]:
```
#encoding categorial values
df=pd.get_dummies(df,columns=['exang'])
```

In [57]:
```
df
```

Out[57]:

| | Age | Gender | trestbps | chol | fbs | restecg | thalach | oldpeak | slope | ca | thal | target | chestpain_0 | chestpain_1 | chestpain_2 | chestpain_3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | Male | 145 | 233 | 1 | 0 | 150 | 2.3 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| **1** | 37 | Male | 130 | 250 | 0 | 1 | 187 | 3.5 | 0 | 0 | 2 | 1 | 0 | 0 | 1 | 0 |
| **2** | 41 | Female | 130 | 204 | 0 | 0 | 172 | 1.4 | 2 | 0 | 2 | 1 | 0 | 1 | 0 | 0 |
| **3** | 56 | Male | 120 | 236 | 0 | 1 | 178 | 0.8 | 2 | 0 | 2 | 1 | 0 | 1 | 0 | 0 |
| **4** | 57 | Female | 120 | 354 | 0 | 1 | 163 | 0.6 | 2 | 0 | 2 | 1 | 1 | 0 | 0 | 0 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | .. |
| **298** | 57 | Female | 140 | 241 | 0 | 1 | 123 | 0.2 | 1 | 0 | 3 | 0 | 1 | 0 | 0 | 0 |
| **299** | 45 | Male | 110 | 264 | 0 | 1 | 132 | 1.2 | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 1 |
| **300** | 68 | Male | 144 | 193 | 1 | 1 | 141 | 3.4 | 1 | 2 | 3 | 0 | 1 | 0 | 0 | 0 |
| **301** | 57 | Male | 130 | 131 | 0 | 1 | 115 | 1.2 | 1 | 1 | 3 | 0 | 1 | 0 | 0 | 0 |
| **302** | 57 | Female | 130 | 236 | 0 | 0 | 174 | 0.0 | 1 | 1 | 2 | 0 | 0 | 1 | 0 | 0 |

303 rows × 18 columns

In [56]:
```
df.to_csv("transformed_heart_desease_data",index=False)
```

In [58]:
```
df3=pd.read_csv("transformed_heart_desease_data")
```

In [59]:
```
df3
```

Out[59]:

| | Age | Gender | trestbps | chol | fbs | restecg | thalach | oldpeak | slope | ca | thal | target | chestpain_0 | chestpain_1 | chestpain_2 | chestpain_3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | Male | 145 | 233 | 1 | 0 | 150 | 2.3 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| **1** | 37 | Male | 130 | 250 | 0 | 1 | 187 | 3.5 | 0 | 0 | 2 | 1 | 0 | 0 | 1 | 0 |
| **2** | 41 | Female | 130 | 204 | 0 | 0 | 172 | 1.4 | 2 | 0 | 2 | 1 | 0 | 1 | 0 | 0 |
| **3** | 56 | Male | 120 | 236 | 0 | 1 | 178 | 0.8 | 2 | 0 | 2 | 1 | 0 | 1 | 0 | 0 |
| **4** | 57 | Female | 120 | 354 | 0 | 1 | 163 | 0.6 | 2 | 0 | 2 | 1 | 1 | 0 | 0 | 0 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | Female | 140 | 241 | 0 | 1 | 123 | 0.2 | 1 | 0 | 3 | 0 | 1 | 0 | 0 | 0 |
| **299** | 45 | Male | 110 | 264 | 0 | 1 | 132 | 1.2 | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 1 |
| **300** | 68 | Male | 144 | 193 | 1 | 1 | 141 | 3.4 | 1 | 2 | 3 | 0 | 1 | 0 | 0 | 0 |
| **301** | 57 | Male | 130 | 131 | 0 | 1 | 115 | 1.2 | 1 | 1 | 3 | 0 | 1 | 0 | 0 | 0 |
| **302** | 57 | Female | 130 | 236 | 0 | 0 | 174 | 0.0 | 1 | 1 | 2 | 0 | 0 | 1 | 0 | 0 |

303 rows × 18 columns

In [60]:

```
#---------------------------Error correction----------------------------------#
```

In [61]:

```
df=pd.read_csv('../DSBDAFINAL/heart_desease_data.csv')
```

In [62]:

```
df
```

Out[62]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [63]:

```
#handling missing values
value=3
df['thal'].fillna(value, inplace=True)
```

In [64]:

```
df
```

Out[64]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [65]:

```
#data type correction
df['thal']=df['thal'].astype(float)
```

In [66]:

```
df
```

Out[66]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1.0 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2.0 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2.0 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2.0 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2.0 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3.0 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3.0 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3.0 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3.0 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2.0 | 0 |

303 rows × 14 columns

In [67]:

```
df.to_csv("Corrected_heart_disease_data",index=False)
```

In [68]:

```
df4=pd.read_csv("Corrected_heart_disease_data")
```

In [69]:

```
df4
```

Out[69]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1.0 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2.0 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2.0 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2.0 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2.0 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3.0 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3.0 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3.0 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3.0 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2.0 | 0 |

303 rows × 14 columns

In [70]:

```
#---------------------Model handling----------------------------#
```

In [71]:

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

In [72]:

```
df=pd.read_csv('../DSBDAFINAL/heart_desease_data.csv')
```

In [73]:

```
df
```

Out[73]:

| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| **1** | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| **2** | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| **3** | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| **4** | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **298** | 57 | 0 | 0 | 140 | 241 | 0 | 1 | 123 | 1 | 0.2 | 1 | 0 | 3 | 0 |
| **299** | 45 | 1 | 3 | 110 | 264 | 0 | 1 | 132 | 0 | 1.2 | 1 | 0 | 3 | 0 |
| **300** | 68 | 1 | 0 | 144 | 193 | 1 | 1 | 141 | 0 | 3.4 | 1 | 2 | 3 | 0 |
| **301** | 57 | 1 | 0 | 130 | 131 | 0 | 1 | 115 | 1 | 1.2 | 1 | 1 | 3 | 0 |
| **302** | 57 | 0 | 1 | 130 | 236 | 0 | 0 | 174 | 0 | 0.0 | 1 | 1 | 2 | 0 |

303 rows × 14 columns

In [74]:

```
#splitting into features and target variables
```

In [74]:
```
#splitting into features and target variables
X=df.drop('target',axis=1)
Y=df['target']
```

In [78]:
```
#splitting into training and testing
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.2,random_state=2)
```

In [79]:
```
#Logistic Regression model
model = LogisticRegression()
model.fit(X_train,Y_train)
```

```
C:\Users\Kasturi Pramod Desai\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:814: ConvergenceWar
ning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

Out[79]:
```
LogisticRegression()
```

In [80]:
```
#prediction of the test cases
y_pred=model.predict(X_test)
```

In [81]:
```
#calculate the accuracy of the model
```

In [82]:
```
accuracy=accuracy_score(Y_test,y_pred)
print("Accuracy",accuracy)
```

```
Accuracy 0.9016393442622951
```

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js