

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

```
d=pd.read_csv('/content/drive/MyDrive/Colab Notebooks/Z-Alizadeh sani dataset (1).csv')
df=pd.DataFrame(d)
df
```

| | Age | Weight | Length | Sex | BMI | DM | HTN | Current Smoker | EX-Smoker | FH | ... | K | Na | WBC | Lymph | Neut | PLT | EF-TTE | Region RWMA | |
|-----|-----|--------|--------|-------|-----------|-----|-----|----------------|-----------|-----|-----|-----|-----|-------|-------|------|-----|--------|-------------|---|
| 0 | 53 | 90 | 175 | Male | 29.387755 | 0 | 1 | 1 | 0 | 0 | ... | 4.7 | 141 | 5700 | 39 | 52 | 261 | 50 | 0 | |
| 1 | 67 | 70 | 157 | Fmale | 28.398718 | 0 | 1 | 0 | 0 | 0 | ... | 4.7 | 156 | 7700 | 38 | 55 | 165 | 40 | 4 | |
| 2 | 54 | 54 | 164 | Male | 20.077335 | 0 | 0 | 1 | 0 | 0 | ... | 4.7 | 139 | 7400 | 38 | 60 | 230 | 40 | 2 | |
| 3 | 66 | 67 | 158 | Fmale | 26.838648 | 0 | 1 | 0 | 0 | 0 | ... | 4.4 | 142 | 13000 | 18 | 72 | 742 | 55 | 0 | |
| 4 | 50 | 87 | 153 | Fmale | 37.165193 | 0 | 1 | 0 | 0 | 0 | ... | 4.0 | 140 | 9200 | 55 | 39 | 274 | 50 | 0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | | |
| 298 | 58 | 84 | 168 | Male | 29.761905 | 0 | 0 | 0 | 0 | 0 | ... | 4.8 | 146 | 8500 | 34 | 58 | 251 | 45 | 0 | |
| 299 | 55 | 64 | 152 | Fmale | 27.700831 | 0 | 0 | 0 | 0 | 0 | ... | 4.0 | 139 | 11400 | 16 | 80 | 377 | 40 | 0 | |
| 300 | 48 | 77 | 160 | Fmale | 30.078125 | 0 | 1 | 0 | 0 | 0 | 1 | ... | 4.0 | 140 | 9000 | 35 | 55 | 279 | 55 | 0 |
| 301 | 57 | 90 | 159 | Fmale | 35.599858 | 1 | 0 | 0 | 0 | 0 | 0 | ... | 3.8 | 141 | 3800 | 48 | 40 | 208 | 55 | 0 |
| 302 | 56 | 85 | 170 | Fmale | 29.411765 | 0 | 1 | 1 | 0 | 0 | 0 | ... | 4.4 | 147 | 6000 | 32 | 55 | 302 | 55 | 0 |

303 rows × 56 columns

```
df.head()
```

| | Age | Weight | Length | Sex | BMI | DM | HTN | Current Smoker | EX-Smoker | FH | ... | K | Na | WBC | Lymph | Neut | PLT | EF-TTE | Region RWMA | |
|---|-----|--------|--------|-------|-----------|----|-----|----------------|-----------|----|-----|-----|-----|------|-------|------|-----|--------|-------------|---|
| 0 | 53 | 90 | 175 | Male | 29.387755 | 0 | 1 | 1 | 0 | 0 | ... | 4.7 | 141 | 5700 | 39 | 52 | 261 | 50 | 0 | |
| 1 | 67 | 70 | 157 | Fmale | 28.398718 | 0 | 1 | 0 | 0 | 0 | ... | 4.7 | 156 | 7700 | 38 | 55 | 165 | 40 | 4 | |
| 2 | 54 | 54 | 164 | Male | 20.077335 | 0 | 0 | 1 | 0 | 0 | 0 | ... | 4.7 | 139 | 7400 | 38 | 60 | 230 | 40 | 2 |
| 3 | 66 | 67 | 158 | Fmale | 26.838648 | 0 | 1 | 0 | 0 | 0 | 0 | ... | 4.4 | 142 | 13000 | 18 | 72 | 742 | 55 | 0 |
| 4 | 50 | 87 | 153 | Fmale | 37.165193 | 0 | 1 | 0 | 0 | 0 | 0 | ... | 4.0 | 140 | 9200 | 55 | 39 | 274 | 50 | 0 |

5 rows × 56 columns

```
df.tail()
```

| | Age | Weight | Length | Sex | BMI | DM | HTN | Current Smoker | EX-Smoker | FH | ... | K | Na | WBC | Lymph | Neut | PLT | EF-TTE | Region RWMA | |
|-----|-----|--------|--------|-------|-----------|----|-----|----------------|-----------|----|-----|-----|-----|-----|-------|------|-----|--------|-------------|---|
| 298 | 58 | 84 | 168 | Male | 29.761905 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 4.8 | 146 | 8500 | 34 | 58 | 251 | 45 | 0 |
| 299 | 55 | 64 | 152 | Fmale | 27.700831 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 4.0 | 139 | 11400 | 16 | 80 | 377 | 40 | 0 |
| 300 | 48 | 77 | 160 | Fmale | 30.078125 | 0 | 1 | 0 | 0 | 0 | 1 | ... | 4.0 | 140 | 9000 | 35 | 55 | 279 | 55 | 0 |
| 301 | 57 | 90 | 159 | Fmale | 35.599858 | 1 | 0 | 0 | 0 | 0 | 0 | ... | 3.8 | 141 | 3800 | 48 | 40 | 208 | 55 | 0 |
| 302 | 56 | 85 | 170 | Fmale | 29.411765 | 0 | 1 | 1 | 0 | 0 | 0 | ... | 4.4 | 147 | 6000 | 32 | 55 | 302 | 55 | 0 |

5 rows × 56 columns

```
df.shape
```

(303, 56)

```
df.info
```

```
<bound method DataFrame.info of
   Age    Weight    Length     Sex      BMI     DM     HTN  Current Smoker \
0      53        90       175   Male  29.387755      0      1
1      67        70       157  Fmale  28.398718      0      1
2      54        54       164   Male  20.077335      0      0
3      66        67       158  Fmale  26.838648      0      1
4      50        87       153  Fmale  37.165193      0      1
```

[303 rows x 56 columns]>

```
df.describe
```

| <bound method NDFrame.describe of | Age | Weight | Length | Sex | BMI | DM | HTN | Current Smoker | \ | | | |
|-----------------------------------|--------|--------|---------|-----------|-------|-------|------|----------------|--------|--------|--------|---|
| 0 53 | 90 | 175 | Male | 29.387755 | 0 | 1 | | 1 | | | | |
| 1 67 | 70 | 157 | Fmale | 28.398718 | 0 | 1 | | 0 | | | | |
| 2 54 | 54 | 164 | Male | 20.077335 | 0 | 0 | | 1 | | | | |
| 3 66 | 67 | 158 | Ffemale | 26.838648 | 0 | 1 | | 0 | | | | |
| 4 50 | 87 | 153 | Ffemale | 37.165193 | 0 | 1 | | 0 | | | | |
| | ... | ... | ... | ... | ... | ... | ... | ... | | | | |
| 298 58 | 84 | 168 | Male | 29.761905 | 0 | 0 | | 0 | | | | |
| 299 55 | 64 | 152 | Ffemale | 27.700831 | 0 | 0 | | 0 | | | | |
| 300 48 | 77 | 160 | Ffemale | 30.078125 | 0 | 1 | | 0 | | | | |
| 301 57 | 90 | 159 | Ffemale | 35.599858 | 1 | 0 | | 0 | | | | |
| 302 56 | 85 | 170 | Ffemale | 29.411765 | 0 | 1 | | 1 | | | | |
| | | | | | | | | | | | | |
| EX-Smoker | FH | ... | K | Na | WBC | Lymph | Neut | PLT | EF-TTE | Region | RWMA | \ |
| 0 0 | 0 | ... | 4.7 | 141 | 5700 | 39 | 52 | 261 | 50 | | 0 | |
| 1 0 | 0 | ... | 4.7 | 156 | 7700 | 38 | 55 | 165 | 40 | | 4 | |
| 2 0 | 0 | ... | 4.7 | 139 | 7400 | 38 | 60 | 230 | 40 | | 2 | |
| 3 0 | 0 | ... | 4.4 | 142 | 13000 | 18 | 72 | 742 | 55 | | 0 | |
| 4 0 | 0 | ... | 4.0 | 140 | 9200 | 55 | 39 | 274 | 50 | | 0 | |
| | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 298 0 | 0 | ... | 4.8 | 146 | 8500 | 34 | 58 | 251 | 45 | | 0 | |
| 299 0 | 0 | ... | 4.0 | 139 | 11400 | 16 | 80 | 377 | 40 | | 0 | |
| 300 0 | 1 | ... | 4.0 | 140 | 9000 | 35 | 55 | 279 | 55 | | 0 | |
| 301 0 | 0 | ... | 3.8 | 141 | 3800 | 48 | 40 | 208 | 55 | | 0 | |
| 302 0 | 0 | ... | 4.4 | 147 | 6000 | 32 | 55 | 302 | 55 | | 0 | |
| | | | | | | | | | | VHD | Cath | \ |
| 0 N | Cad | | | | | | | | | | | |
| 1 N | Cad | | | | | | | | | mild | Cad | \ |
| 2 Severe | Normal | | | | | | | | | | | |
| 3 Severe | Normal | | | | | | | | | Severe | Normal | \ |
| 4 Severe | Normal | | | | | | | | | | | |
| | ... | | | | | | | | | N | Cad | \ |
| 298 N | Cad | | | | | | | | | | | |
| 299 mild | Normal | | | | | | | | | N | Normal | \ |
| 300 N | Normal | | | | | | | | | | | |
| 301 N | Normal | | | | | | | | | N | Cad | \ |
| 302 N | Cad | | | | | | | | | | | |

[303 rows x 56 columns]>

```
df.columns
```

```
Index(['Age', 'Weight', 'Length', 'Sex', 'BMI', 'DM', 'HTN', 'Current Smoker',
       'EX-Smoker', 'FH', 'Obesity', 'CRF', 'CVA', 'Airway disease',
       'Thyroid Disease', 'CHF', 'DLP', 'BP', 'PR', 'Edema',
       'Weak Peripheral Pulse', 'Lung rales', 'Systolic Murmur',
       'Diastolic Murmur', 'Typical Chest Pain', 'Dyspnea', 'Function Class',
       'Atypical', 'Nonanginal', 'Exertional CP', 'LowTH Ang', 'Q Wave',
       'St Elevation', 'St Depression', 'Tinversion', 'LVH',
       'Poor R Progression', 'BBB', 'FBS', 'CR', 'TG', 'LDL', 'HDL', 'BUN',
       'ESR', 'HB', 'K', 'Na', 'WBC', 'Lymph', 'Neut', 'PLT', 'EF-TTE',
       'Region RWMA', 'VHD', 'Cath'],
      dtype='object')
```

```
df.isnull().sum()
```

| | |
|-----------------------|---|
| Age | 0 |
| Weight | 0 |
| Length | 0 |
| Sex | 0 |
| BMI | 0 |
| DM | 0 |
| HTN | 0 |
| Current Smoker | 0 |
| EX-Smoker | 0 |
| FH | 0 |
| Obesity | 0 |
| CRF | 0 |
| CVA | 0 |
| Airway disease | 0 |
| Thyroid Disease | 0 |
| CHF | 0 |
| DLP | 0 |
| BP | 0 |
| PR | 0 |
| Edema | 0 |
| Weak Peripheral Pulse | 0 |
| Lung rales | 0 |
| Systolic Murmur | 0 |
| Diastolic Murmur | 0 |
| Typical Chest Pain | 0 |
| Dyspnea | 0 |
| Function Class | 0 |
| Atypical | 0 |
| Nonanginal | 0 |
| Exertional CP | 0 |
| LowTH Ang | 0 |
| Q Wave | 0 |
| St Elevation | 0 |
| St Depression | 0 |
| Tinversion | 0 |
| LVH | 0 |
| Poor R Progression | 0 |
| BBB | 0 |
| FBS | 0 |
| CR | 0 |
| TG | 0 |
| LDL | 0 |
| HDL | 0 |
| BUN | 0 |
| ESR | 0 |
| HB | 0 |
| K | 0 |
| Na | 0 |
| WBC | 0 |
| Lymph | 0 |
| Neut | 0 |
| PLT | 0 |
| EF-TTE | 0 |
| Region RWMA | 0 |
| VHD | 0 |
| Cath | 0 |

```
dtype: int64
```

```
df.dtypes
```

| | |
|----------------|---------|
| Age | int64 |
| Weight | int64 |
| Length | int64 |
| Sex | object |
| BMI | float64 |
| DM | int64 |
| HTN | int64 |
| Current Smoker | int64 |

```

EX-Smoker           int64
FH                int64
Obesity          object
CRF              object
CVA              object
Airway disease   object
Thyroid Disease  object
CHF              object
DLP              object
BP                int64
PR                int64
Edema            int64
Weak Peripheral Pulse object
Lung rales        object
Systolic Murmur  object
Diastolic Murmur object
Typical Chest Pain int64
Dyspnea          object
Function Class   int64
Atypical         object
Nonanginal        object
Exertional CP    object
LowTH Ang        object
Q Wave           int64
St Elevation     int64
St Depression    int64
Tinversion       int64
LVH              object
Poor R Progression object
BBB              object
FBS              int64
CR                float64
TG                int64
LDL              int64
HDL              float64
BUN              int64
ESR              int64
HB                float64
K                 float64
Na                int64
WBC              int64
Lymph            int64
Neut             int64
PLT              int64
EF-TTE           int64
Region RWMA     int64
VHD              object
Cath              object
dtype: object

```

```
pd.value_counts(df['Sex'])
```

```

Male      176
Female    127
Name: Sex, dtype: int64

```

```
sns.countplot(x=df['Sex'])
```

```
<Axes: xlabel='Sex', ylabel='count'>
```

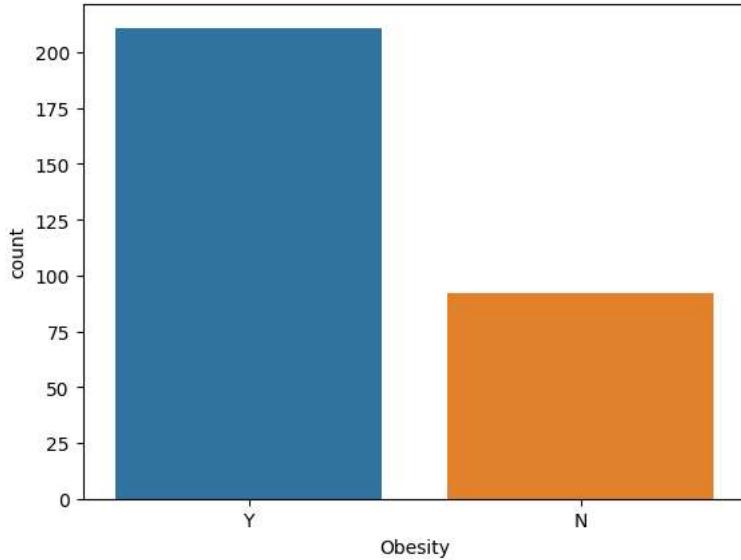
```
pd.value_counts(df['Obesity'])
```

```
Y    211  
N    92  
Name: Obesity, dtype: int64
```

```
125 |
```

```
sns.countplot(x=df['Obesity'])
```

```
<Axes: xlabel='Obesity', ylabel='count'>
```

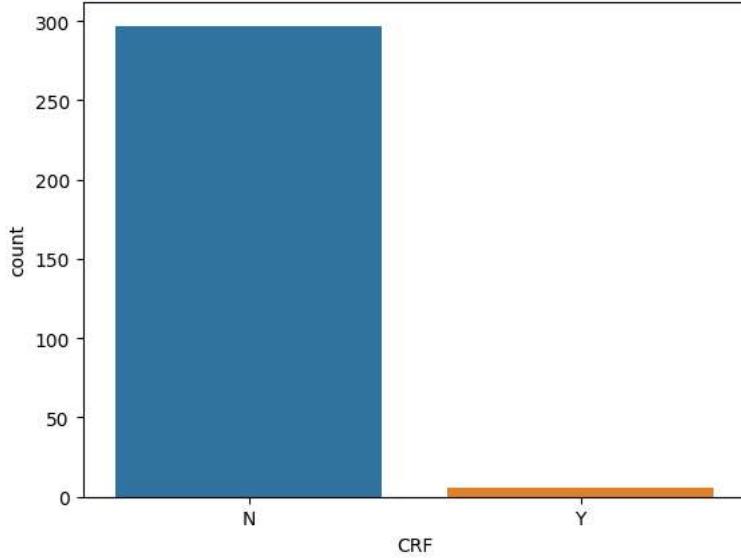


```
pd.value_counts(df['CRF'])
```

```
N    297  
Y     6  
Name: CRF, dtype: int64
```

```
sns.countplot(x=df['CRF'])
```

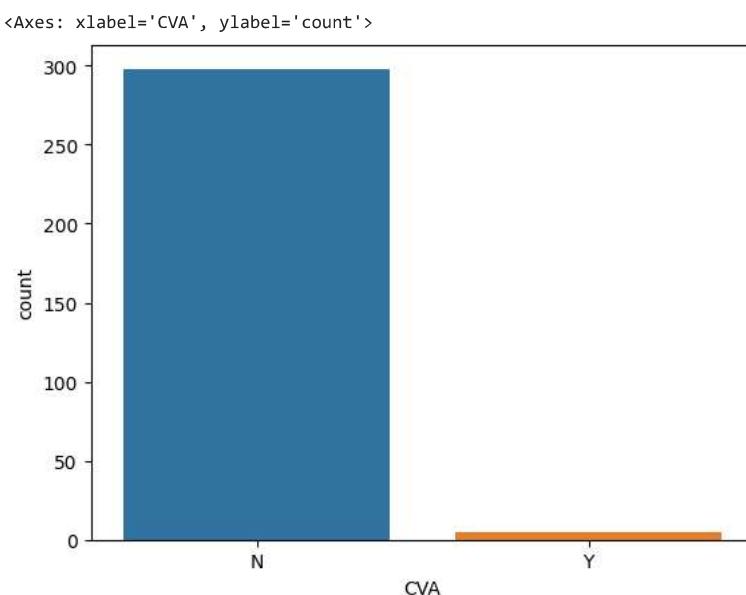
```
<Axes: xlabel='CRF', ylabel='count'>
```



```
pd.value_counts(df['CVA'])
```

```
N    298  
Y     5  
Name: CVA, dtype: int64
```

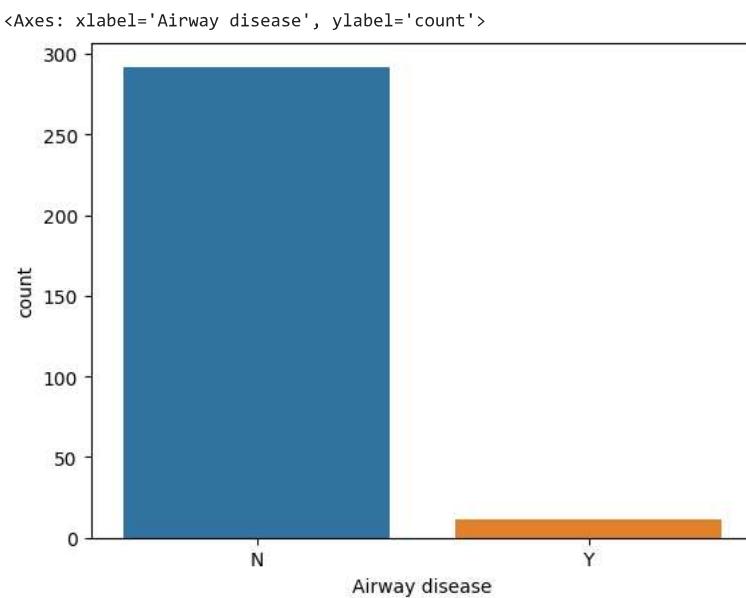
```
sns.countplot(x=df['CVA'])
```



```
pd.value_counts(df['Airway disease'])
```

```
N    292  
Y     11  
Name: Airway disease, dtype: int64
```

```
sns.countplot(x=df['Airway disease'])
```

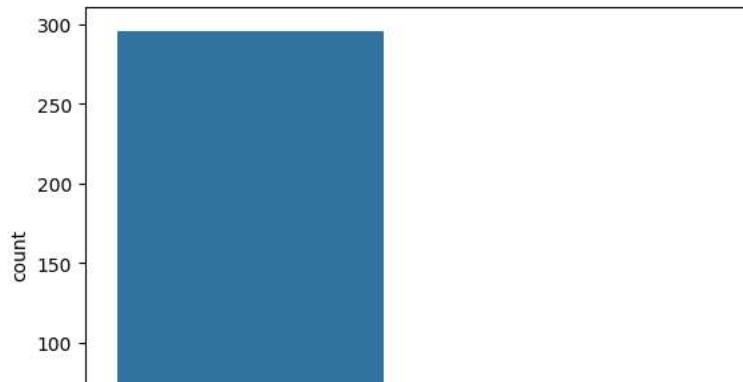


```
pd.value_counts(df['Thyroid Disease'])
```

```
N    296  
Y      7  
Name: Thyroid Disease, dtype: int64
```

```
sns.countplot(x=df['Thyroid Disease'])
```

```
<Axes: xlabel='Thyroid Disease', ylabel='count'>
```

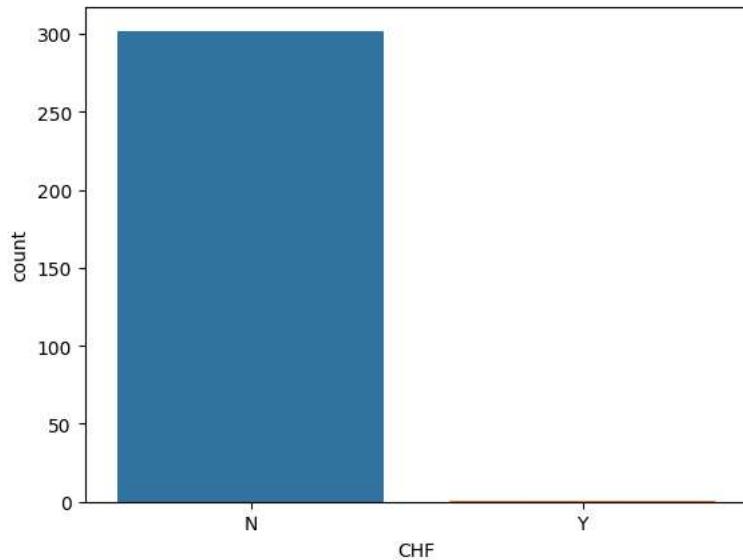


```
pd.value_counts(df['CHF'])
```

```
N      302  
Y       1  
Name: CHF, dtype: int64
```

```
sns.countplot(x=df['CHF'])
```

```
<Axes: xlabel='CHF', ylabel='count'>
```



```
pd.value_counts(df['DLP'])
```

```
N      191  
Y      112  
Name: DLP, dtype: int64
```

```
sns.countplot(x=df['DLP'])
```

```
<Axes: xlabel='DLP', ylabel='count'>
```

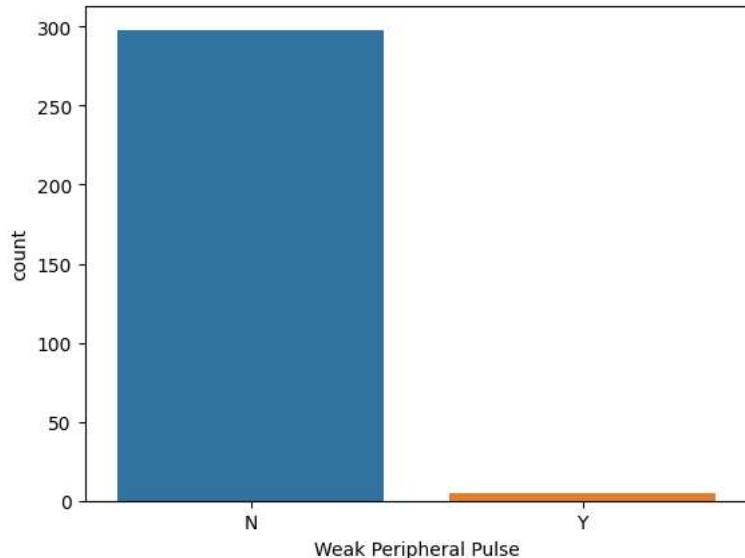


```
pd.value_counts(df['Weak Peripheral Pulse'])
```

```
N    298  
Y     5  
Name: Weak Peripheral Pulse, dtype: int64
```

```
sns.countplot(x=df['Weak Peripheral Pulse'])
```

```
<Axes: xlabel='Weak Peripheral Pulse', ylabel='count'>
```

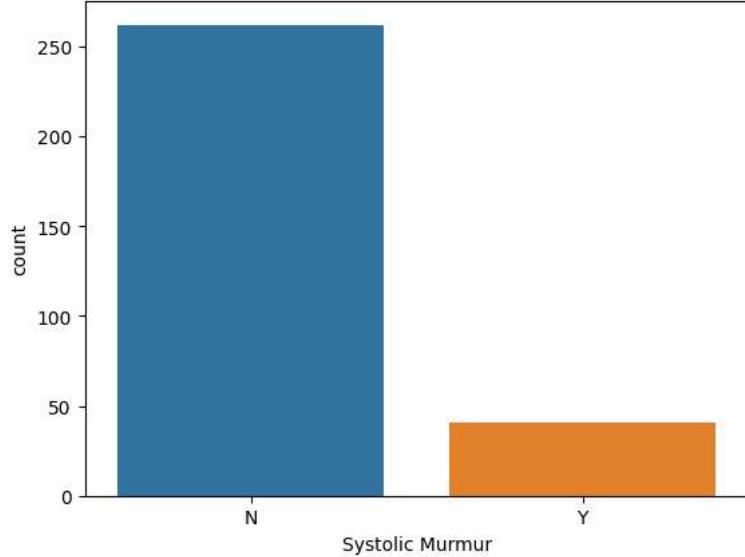


```
pd.value_counts(df['Systolic Murmur'])
```

```
N    262  
Y    41  
Name: Systolic Murmur, dtype: int64
```

```
sns.countplot(x=df['Systolic Murmur'])
```

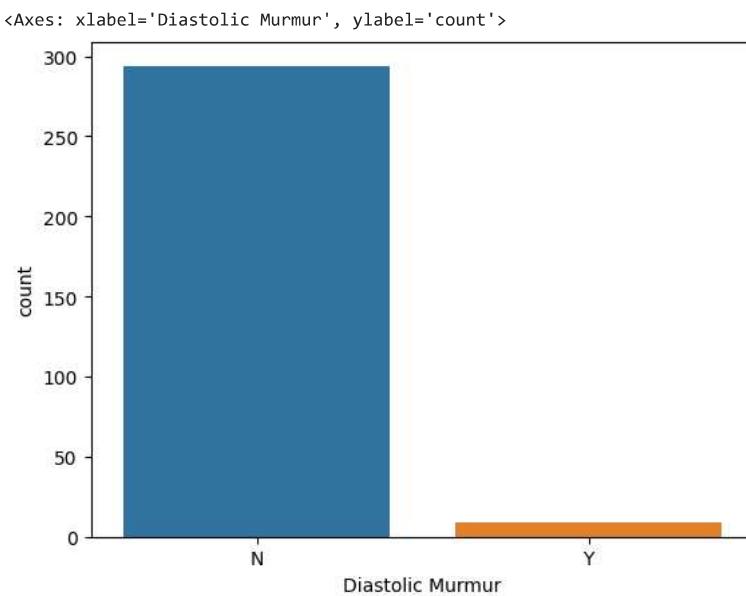
```
<Axes: xlabel='Systolic Murmur', ylabel='count'>
```



```
pd.value_counts(df['Diastolic Murmur'])
```

```
N    294  
Y     9  
Name: Diastolic Murmur, dtype: int64
```

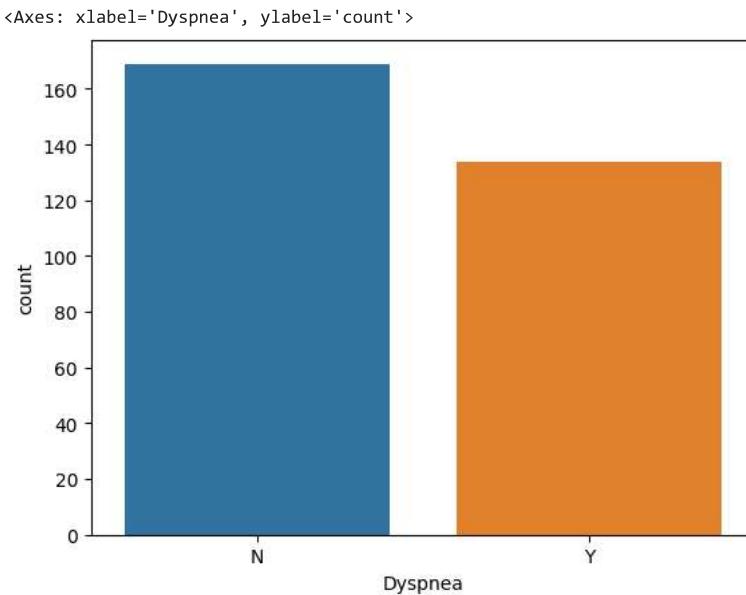
```
sns.countplot(x=df['Diastolic Murmur'])
```



```
pd.value_counts(df['Dyspnea'])
```

```
N    169  
Y    134  
Name: Dyspnea, dtype: int64
```

```
sns.countplot(x=df['Dyspnea'])
```

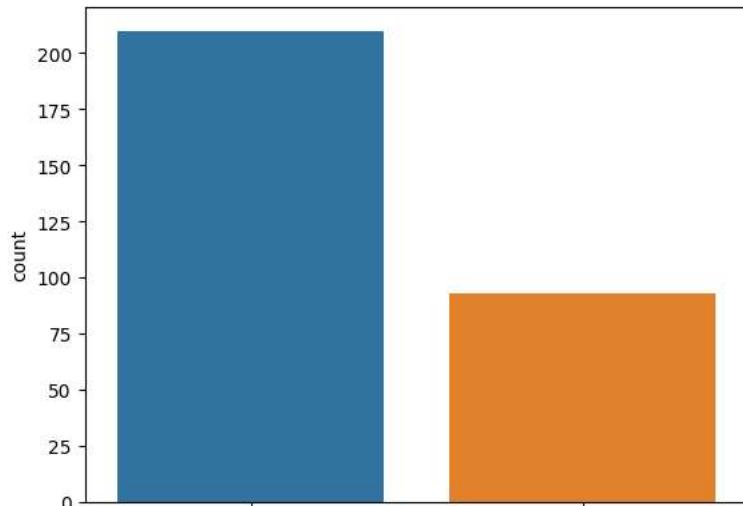


```
pd.value_counts(df['Atypical'])
```

```
N    210  
Y    93  
Name: Atypical, dtype: int64
```

```
sns.countplot(x=df['Atypical'])
```

```
<Axes: xlabel='Atypical', ylabel='count'>
```

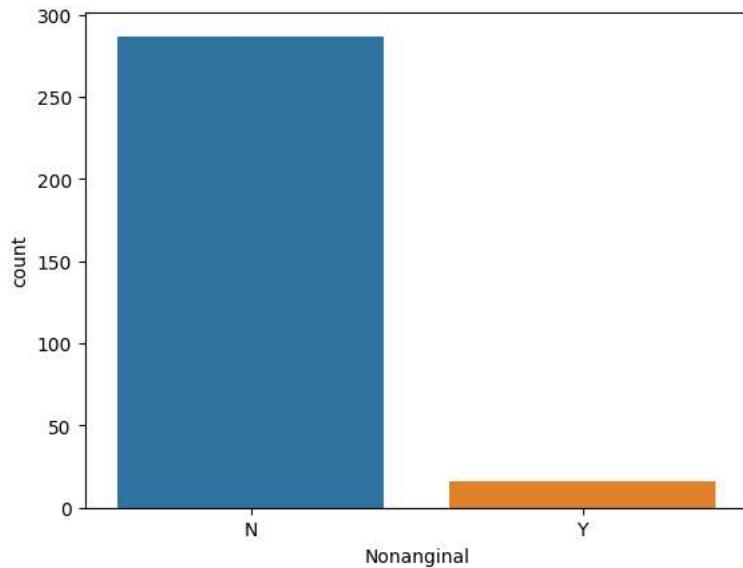


```
pd.value_counts(df['Nonanginal'])
```

```
N    287  
Y     16  
Name: Nonanginal, dtype: int64
```

```
sns.countplot(x=df['Nonanginal'])
```

```
<Axes: xlabel='Nonanginal', ylabel='count'>
```



```
pd.value_counts(df['Exertional CP'])
```

```
N    303  
Name: Exertional CP, dtype: int64
```

```
sns.countplot(x=df['Exertional CP'])
```

```
<Axes: xlabel='Exertional CP', ylabel='count'>
```

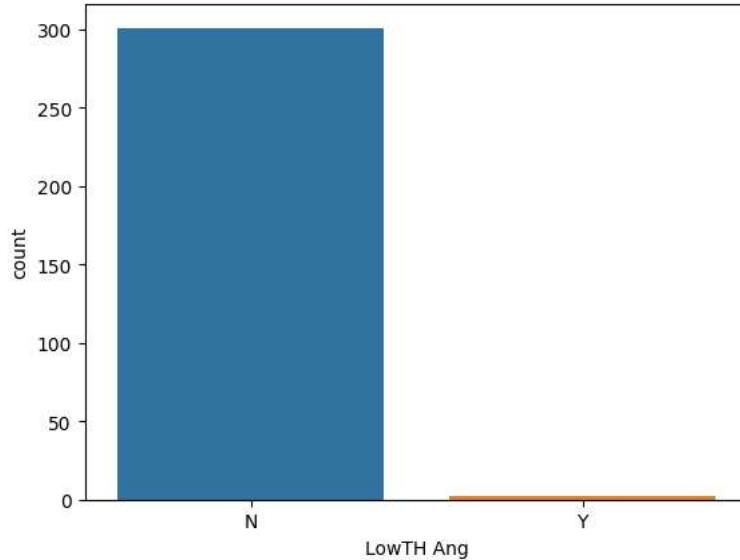


```
pd.value_counts(df['LowTH Ang'])
```

```
N    301  
Y     2  
Name: LowTH Ang, dtype: int64
```

```
sns.countplot(x=df['LowTH Ang'])
```

```
<Axes: xlabel='LowTH Ang', ylabel='count'>
```

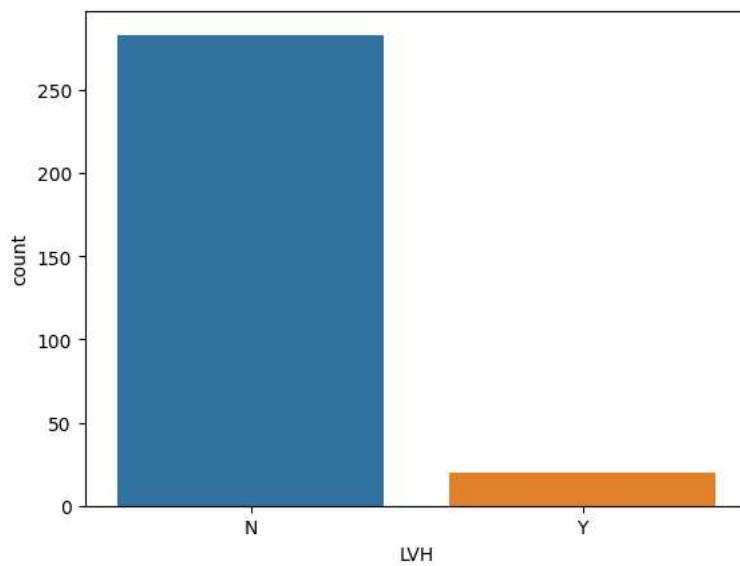


```
pd.value_counts(df['LVH'])
```

```
N    283  
Y     20  
Name: LVH, dtype: int64
```

```
sns.countplot(x=df['LVH'])
```

```
<Axes: xlabel='LVH', ylabel='count'>
```

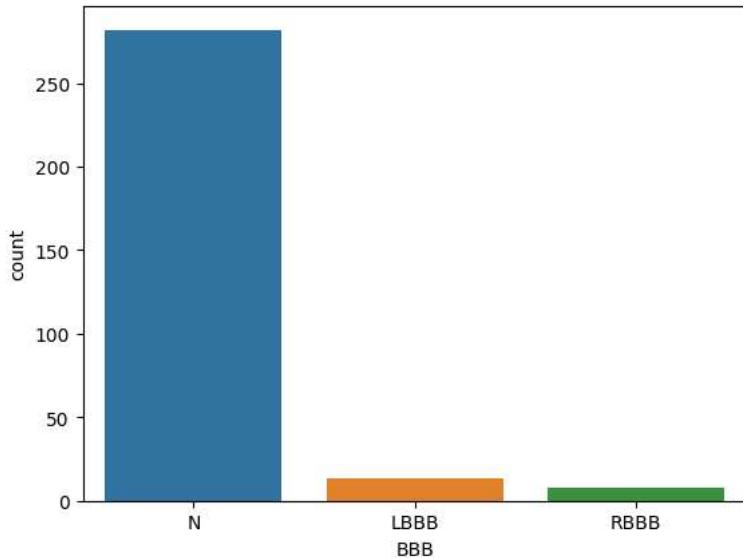


```
pd.value_counts(df['BBB'])
```

| | |
|-------------------------|-----|
| N | 282 |
| LBBB | 13 |
| RBBB | 8 |
| Name: BBB, dtype: int64 | |

```
sns.countplot(x=df['BBB'])
```

<Axes: xlabel='BBB', ylabel='count'>

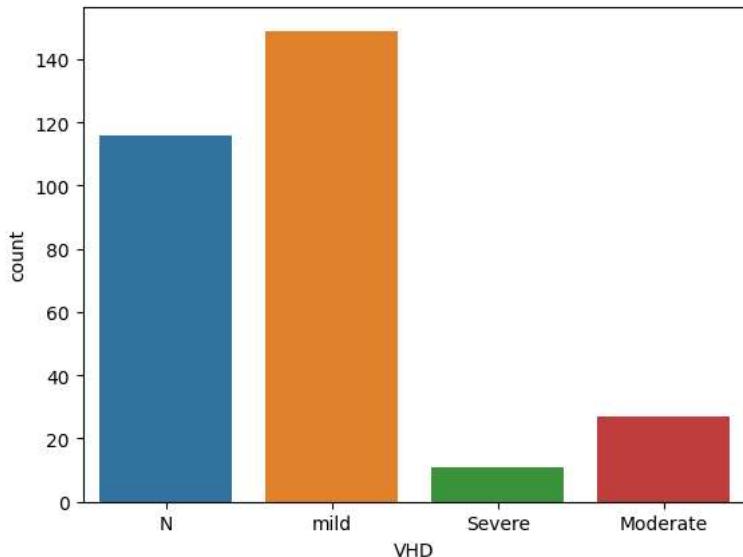


```
pd.value_counts(df['VHD'])
```

| | |
|-------------------------|-----|
| mild | 149 |
| N | 116 |
| Moderate | 27 |
| Severe | 11 |
| Name: VHD, dtype: int64 | |

```
sns.countplot(x=df['VHD'])
```

<Axes: xlabel='VHD', ylabel='count'>



```
df1=pd.get_dummies(df[['VHD']],drop_first=True)
df1
```

| | VHD_N | VHD_Severe | VHD_mild |
|-----|-------|------------|----------|
| 0 | 1 | 0 | 0 |
| 1 | 1 | 0 | 0 |
| 2 | 0 | 0 | 1 |
| 3 | 0 | 1 | 0 |
| 4 | 0 | 1 | 0 |
| ... | ... | ... | ... |
| 298 | 1 | 0 | 0 |
| 299 | 0 | 0 | 1 |
| 300 | 1 | 0 | 0 |
| 301 | 1 | 0 | 0 |
| ... | - | - | - |

```
col=['Sex','Obesity','CRF','CVA','Airway disease','Thyroid Disease','CHF','DLP','Weak Peripheral Pulse','Lung rales','Systolic Murmur','Diastole','Hypotension','Hypertension','DM','HTN','Smoker','FH','K','Na','WBC','Lymph','Neut','PLT','EF-TTE','Region','RWMA','VHD','Cath']
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
for i in col:
    df[i]=le.fit_transform(df[i])
df
```

| | Age | Weight | Length | Sex | BMI | DM | HTN | Current Smoker | EX-Smoker | FH | ... | K | Na | WBC | Lymph | Neut | PLT | EF-TTE | Region RWMA | VHD | Cath | |
|-----|-----|--------|--------|-----|-----------|-----|-----|----------------|-----------|-----|-----|-----|-----|-------|-------|------|-----|--------|-------------|--------|--------|--------|
| 0 | 53 | 90 | 175 | 1 | 29.387755 | 0 | 1 | 1 | 0 | 0 | ... | 4.7 | 141 | 5700 | 39 | 52 | 261 | 50 | 0 | N | Cad | |
| 1 | 67 | 70 | 157 | 0 | 28.398718 | 0 | 1 | 0 | 0 | 0 | ... | 4.7 | 156 | 7700 | 38 | 55 | 165 | 40 | 4 | N | Cad | |
| 2 | 54 | 54 | 164 | 1 | 20.077335 | 0 | 0 | 1 | 0 | 0 | ... | 4.7 | 139 | 7400 | 38 | 60 | 230 | 40 | 2 | mild | Cad | |
| 3 | 66 | 67 | 158 | 0 | 26.838648 | 0 | 1 | 0 | 0 | 0 | ... | 4.4 | 142 | 13000 | 18 | 72 | 742 | 55 | 0 | Severe | Normal | |
| 4 | 50 | 87 | 153 | 0 | 37.165193 | 0 | 1 | 0 | 0 | 0 | ... | 4.0 | 140 | 9200 | 55 | 39 | 274 | 50 | 0 | Severe | Normal | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 298 | 58 | 84 | 168 | 1 | 29.761905 | 0 | 0 | 0 | 0 | 0 | ... | 4.8 | 146 | 8500 | 34 | 58 | 251 | 45 | 0 | N | Cad | |
| 299 | 55 | 64 | 152 | 0 | 27.700831 | 0 | 0 | 0 | 0 | 0 | ... | 4.0 | 139 | 11400 | 16 | 80 | 377 | 40 | 0 | mild | Normal | |
| 300 | 48 | 77 | 160 | 0 | 30.078125 | 0 | 1 | 0 | 0 | 0 | 1 | ... | 4.0 | 140 | 9000 | 35 | 55 | 279 | 55 | 0 | N | Normal |
| 301 | 57 | 90 | 159 | 0 | 35.599858 | 1 | 0 | 0 | 0 | 0 | 0 | ... | 3.8 | 141 | 3800 | 48 | 40 | 208 | 55 | 0 | N | Normal |
| 302 | 56 | 85 | 170 | 0 | 29.411765 | 0 | 1 | 1 | 0 | 0 | ... | 4.4 | 147 | 6000 | 32 | 55 | 302 | 55 | 0 | N | Cad | |

302 rows x 22 columns

| | VHD_N | VHD_Severe | VHD_mild | Age | Weight | Length | Sex | BMI | DM | HTN | ... | K | Na | WBC | Lymph | Neut | PLT | EF-TTE | Region RWMA | VHD |
|-----|-------|------------|----------|-----|--------|--------|-----|-----------|-----|-----|-----|-----|-----|-------|-------|------|-----|--------|-------------|--------|
| 0 | 1 | 0 | 0 | 53 | 90 | 175 | 1 | 29.387755 | 0 | 1 | ... | 4.7 | 141 | 5700 | 39 | 52 | 261 | 50 | 0 | N |
| 1 | 1 | 0 | 0 | 67 | 70 | 157 | 0 | 28.398718 | 0 | 1 | ... | 4.7 | 156 | 7700 | 38 | 55 | 165 | 40 | 4 | N |
| 2 | 0 | 0 | 1 | 54 | 54 | 164 | 1 | 20.077335 | 0 | 0 | ... | 4.7 | 139 | 7400 | 38 | 60 | 230 | 40 | 2 | mild |
| 3 | 0 | 1 | 0 | 66 | 67 | 158 | 0 | 26.838648 | 0 | 1 | ... | 4.4 | 142 | 13000 | 18 | 72 | 742 | 55 | 0 | Severe |
| 4 | 0 | 1 | 0 | 50 | 87 | 153 | 0 | 37.165193 | 0 | 1 | ... | 4.0 | 140 | 9200 | 55 | 39 | 274 | 50 | 0 | Severe |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 298 | 1 | 0 | 0 | 58 | 84 | 168 | 1 | 29.761905 | 0 | 0 | ... | 4.8 | 146 | 8500 | 34 | 58 | 251 | 45 | 0 | N |
| 299 | 0 | 0 | 1 | 55 | 64 | 152 | 0 | 27.700831 | 0 | 0 | ... | 4.0 | 139 | 11400 | 16 | 80 | 377 | 40 | 0 | mild |
| 300 | 1 | 0 | 0 | 48 | 77 | 160 | 0 | 30.078125 | 0 | 1 | ... | 4.0 | 140 | 9000 | 35 | 55 | 279 | 55 | 0 | N |
| 301 | 1 | 0 | 0 | 57 | 90 | 159 | 0 | 35.599858 | 1 | 0 | ... | 3.8 | 141 | 3800 | 48 | 40 | 208 | 55 | 0 | N |
| 302 | 1 | 0 | 0 | 56 | 85 | 170 | 0 | 29.411765 | 0 | 1 | ... | 4.4 | 147 | 6000 | 32 | 55 | 302 | 55 | 0 | N |

302 rows x 20 columns

```
df2.dtypes
```

```
VHD_mild          uint8
Age              int64
Weight           int64
Length           int64
Sex              int64
BMI              float64
DM               int64
HTN              int64
Current Smoker   int64
EX-Smoker        int64
FH               int64
Obesity          int64
CRF              int64
CVA              int64
Airway disease   int64
Thyroid Disease  int64
CHF              int64
DLP              int64
BP               int64
PR               int64
Edema            int64
Weak Peripheral Pulse int64
Lung rales       int64
Systolic Murmur  int64
Diastolic Murmur int64
Typical Chest Pain int64
Dyspnea          int64
Function Class   int64
Atypical         int64
Nonanginal        int64
Exertional CP    int64
LowTH Ang        int64
Q Wave           int64
St Elevation     int64
St Depression    int64
Tinversion        int64
LVH              int64
Poor R Progression int64
BBB              int64
FBS              int64
CR               float64
TG               int64
LDL              int64
HDL              float64
BUN              int64
ESR              int64
HB               float64
K                float64
Na               int64
WBC              int64
Lymph            int64
Neut             int64
PLT              int64
EF-TTE           int64
Region RWMA     int64
VHD              object
Cath             object
dtype: object
```

```
df2=df2.drop(['CVA','Function Class','Neut','VHD'],axis=1)
df2
```

| VHD_N | VHD_Severe | VHD_mild | Age | Weight | Length | Sex | BMI | DM | HTN | ... | ESR | HB | K | Na | WBC | Lymph | PLT | EF-TTE | Region-RWMA | |
|-------|------------|----------|-----|--------|--------|-----|-----|-----------|-----|-----|-----|----|------|-----|-----|-------|-----|--------|-------------|---|
| 0 | 1 | 0 | 0 | 53 | 90 | 175 | 1 | 29.387755 | 0 | 1 | ... | 7 | 15.6 | 4.7 | 141 | 5700 | 39 | 261 | 50 | 0 |

```
df2.corr()
```

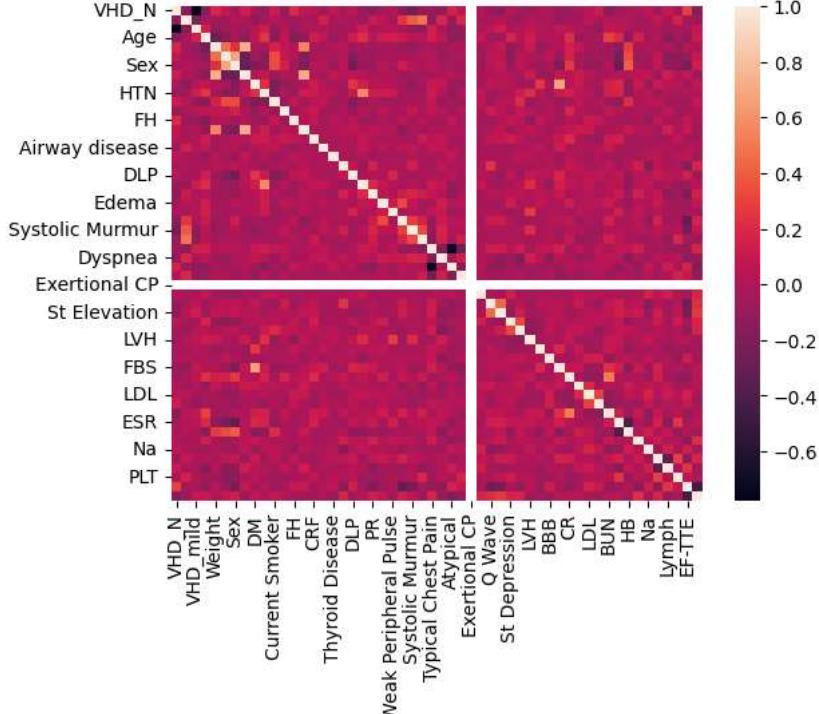
| | VHD_N | VHD_Severe | VHD_mild | Age | Weight | Length | Sex | BMI | DM | HTN | ... | BUN |
|------------------------|-----------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----|-----------|
| VHD_N | 1.000000 | -0.152867 | -0.774713 | -0.166309 | 0.180712 | 0.085980 | -0.046504 | 0.128441 | -0.021626 | -0.103953 | ... | -0.151715 |
| VHD_Severe | -0.152867 | 1.000000 | -0.190914 | -0.035501 | -0.048873 | -0.052822 | 0.021834 | -0.011517 | -0.126164 | -0.053769 | ... | 0.026629 |
| VHD_mild | -0.774713 | -0.190914 | 1.000000 | 0.149683 | -0.114694 | -0.077774 | -0.034088 | -0.066342 | 0.039622 | 0.133951 | ... | 0.091491 |
| Age | -0.166309 | -0.035501 | 0.149683 | 1.000000 | -0.264585 | -0.163753 | -0.045769 | -0.161414 | 0.072543 | 0.246690 | ... | 0.300663 |
| Weight | 0.180712 | -0.048873 | -0.114694 | -0.264585 | 1.000000 | 0.460631 | 0.234529 | 0.725005 | -0.003531 | -0.028532 | ... | -0.057670 |
| Length | 0.085980 | -0.052822 | -0.077774 | -0.163753 | 0.460631 | 1.000000 | 0.700279 | -0.269356 | -0.052318 | -0.153668 | ... | -0.071229 |
| Sex | -0.046504 | 0.021834 | -0.034088 | -0.045769 | 0.234529 | 0.700279 | 1.000000 | -0.284088 | -0.194348 | -0.149278 | ... | 0.009351 |
| BMI | 0.128441 | -0.011517 | -0.066342 | -0.161414 | 0.725005 | -0.269356 | -0.284088 | 1.000000 | 0.045360 | 0.091652 | ... | -0.011139 |
| DM | -0.021626 | -0.126164 | 0.039622 | 0.072543 | -0.003531 | -0.052318 | -0.194348 | 0.045360 | 1.000000 | 0.217864 | ... | 0.144394 |
| HTN | -0.103953 | -0.053769 | 0.133951 | 0.246690 | -0.028532 | -0.153668 | -0.149278 | 0.091652 | 0.217864 | 1.000000 | ... | 0.152895 |
| Current Smoker | 0.115128 | 0.030995 | -0.113548 | -0.143879 | 0.157385 | 0.335248 | 0.336330 | -0.089398 | -0.208458 | -0.169000 | ... | -0.061596 |
| EX-Smoker | -0.107497 | -0.035857 | 0.040002 | 0.076608 | 0.068977 | 0.079034 | 0.156932 | 0.005016 | -0.120087 | 0.041045 | ... | 0.066455 |
| FH | 0.197552 | -0.084208 | -0.101320 | -0.183900 | 0.021963 | 0.004488 | -0.071098 | 0.014045 | -0.064434 | -0.098467 | ... | -0.014419 |
| Obesity | 0.062328 | -0.025328 | -0.010898 | -0.126190 | 0.547267 | -0.171962 | -0.211808 | 0.712501 | 0.020839 | 0.136482 | ... | -0.014311 |
| CRF | 0.034259 | -0.027587 | 0.002346 | 0.126980 | -0.025713 | -0.033826 | 0.024718 | 0.009386 | 0.114975 | 0.118299 | ... | 0.119344 |
| Airway disease | -0.080267 | 0.150996 | 0.020850 | 0.069941 | -0.057719 | 0.004021 | 0.021834 | -0.062781 | 0.028291 | 0.053887 | ... | 0.105386 |
| Thyroid Disease | 0.059664 | -0.029848 | -0.019434 | -0.095886 | 0.033370 | -0.042495 | -0.091986 | 0.068834 | -0.051885 | 0.038634 | ... | -0.045902 |
| CHF | -0.045322 | -0.011169 | -0.056602 | -0.021618 | 0.029659 | 0.014113 | 0.048881 | 0.020205 | -0.037405 | -0.069137 | ... | 0.070411 |
| DLP | 0.057981 | -0.038966 | -0.014714 | 0.128147 | -0.080068 | -0.173428 | -0.277911 | 0.046570 | 0.250364 | 0.108948 | ... | 0.014585 |

sns.heatmap(df2.corr())

<ipython-input-57-28890f610699>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version

sns.heatmap(df2.corr())

<Axes: >



df2.dtypes

```

VHD_N          uint8
VHD_Severe    uint8
VHD_mild      uint8
Age           int64
Weight         int64
Length         int64
Sex            int64
BMI            float64
DM             int64
HTN            int64
Current Smoker int64
EX-Smoker     int64
FH             int64
Obesity        int64
CRF            int64
Airway disease int64
Thyroid Disease int64
CHF            int64
DLP            int64
BP             int64
PR             int64
Edema          int64
Weak Peripheral Pulse int64
Lung rales     int64
Systolic Murmur int64
Diastolic Murmur int64
Typical Chest Pain int64
Dyspnea         int64
Atypical        int64
Nonanginal      int64
Exertional CP   int64
LowTH Ang      int64
Q Wave          int64
St Elevation    int64
St Depression   int64
Tinversion      int64
LVH             int64
Poor R Progression int64
BBB             int64
FBS             int64
CR              float64
TG              int64
LDL             int64
HDL             float64
BUN             int64
ESR             int64
HB              float64
K               float64
Na              int64
WBC             int64
Lymph           int64
PLT             int64
EF-TTE          int64
Region RWMA    int64
Cath            object
dtype: object

```

```

x=df2.drop(['Cath'],axis=1)
x

```

```

y=df2['Cath']
y

0      Cad
1      Cad
2      Cad
3  Normal
4  Normal
...
298     Cad
299  Normal
300  Normal
301  Normal
302     Cad
Name: Cath, Length: 303, dtype: object

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.30,random_state=42)
302    1      0      0   56     85    170    0  29.411765    0    1 ...   16   13  14.7   4.4   147   6000    32   302    55

from sklearn.feature_selection import mutual_info_classif
mutual_info=mutual_info_classif(x_train,y_train)
mutual_info

array([0.04183737, 0.08950845, 0.          , 0.04302064, 0.00254775,
       0.          , 0.          , 0.          , 0.02121726, 0.06405332,
       0.01470585, 0.          , 0.          , 0.01580906, 0.00158258,
       0.03835314, 0.          , 0.          , 0.03163234, 0.02091254,
       0.02494314, 0.          , 0.04452742, 0.          , 0.01919112,
       0.00717074, 0.13134688, 0.          , 0.12904161, 0.          ,
       0.          , 0.          , 0.03655832, 0.00189247, 0.          ,
       0.0138027, 0.          , 0.          , 0.          , 0.06481592,
       0.          , 0.          , 0.02681995, 0.02460756, 0.02243692,
       0.00333635, 0.          , 0.          , 0.00772809, 0.          ,
       0.04879725, 0.          , 0.06537088, 0.05811669])

mutual_info=pd.Series(mutual_info)
mutual_info.index=x_train.columns
mutual_info.sort_values(ascending=False)

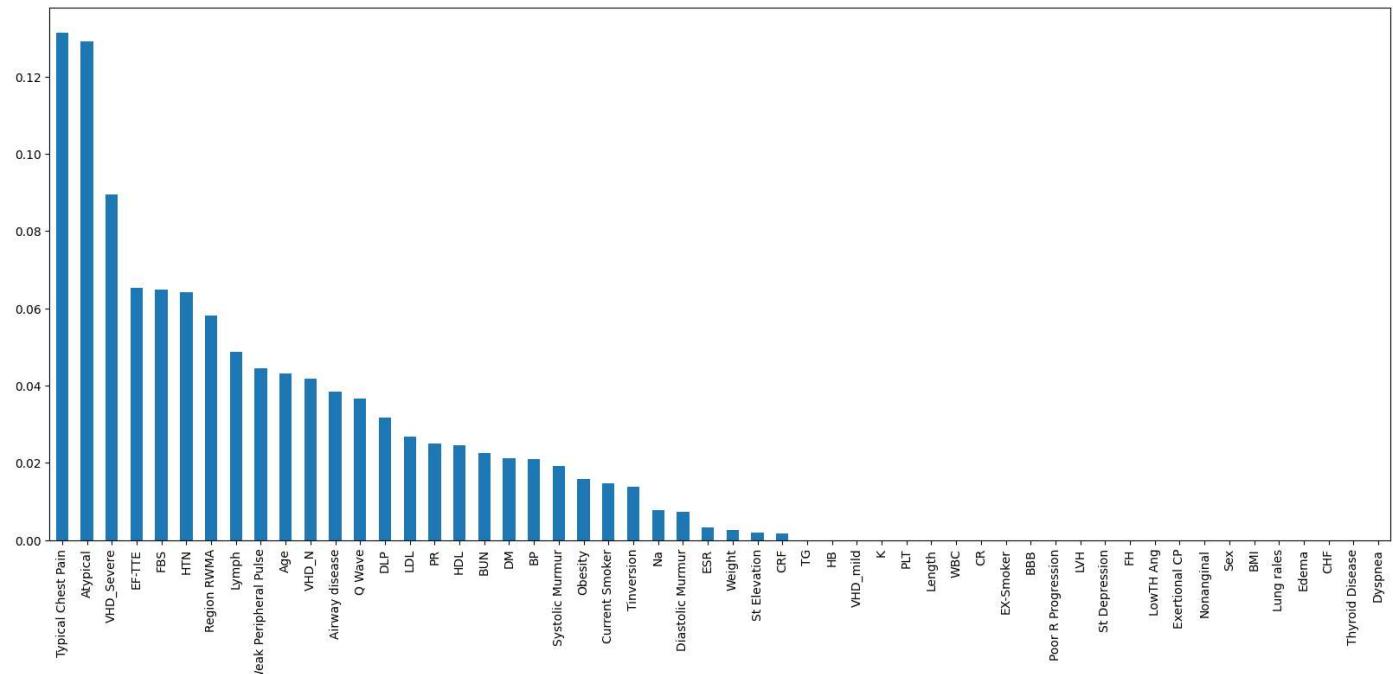
Typical Chest Pain      0.131347
Atypical                0.129042
VHD_Severe               0.089508
EF-TTE                  0.065371
FBS                      0.064816
HTN                      0.064053
Region RWMA              0.058117
Lymph                    0.048797
Weak Peripheral Pulse    0.044527
Age                      0.043021
VHD_N                   0.041837
Airway disease            0.038353
Q Wave                   0.036558
DLP                      0.031632
LDL                      0.026820
PR                       0.024943
HDL                      0.024608
BUN                      0.022437
DM                       0.021217
BP                       0.020913
Systolic Murmur           0.019191
Obesity                  0.015809
Current Smoker             0.014706
Tinversion                 0.013803
Na                       0.007728
Diastolic Murmur           0.007171
ESR                      0.003336
Weight                   0.002548
St Elevation              0.001892
CRF                      0.001583
TG                       0.000000
HB                       0.000000
VHD_mild                 0.000000
K                        0.000000
PLT                      0.000000
Length                   0.000000
WBC                      0.000000
CR                       0.000000
EX-Smoker                 0.000000
BBB                      0.000000

```

```
Poor R Progression      0.000000
LVH                    0.000000
St Depression          0.000000
FH                     0.000000
LowTH Ang              0.000000
Exertional CP          0.000000
Nonanginal             0.000000
Sex                    0.000000
BMI                   0.000000
Lung rales             0.000000
Edema                 0.000000
CHF                   0.000000
Thyroid Disease        0.000000
Dyspnea               0.000000
dtype: float64
```

```
mutual_info.sort_values(ascending=False).plot.bar(figsize=(20,8))
```

<Axes: >



```
from sklearn.feature_selection import SelectKBest
```

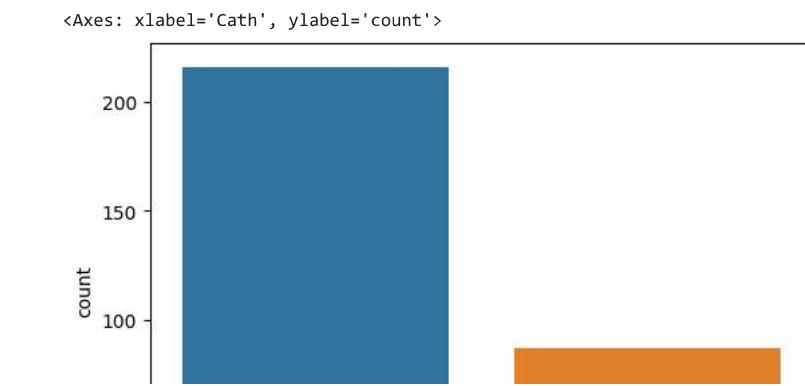
```
sel_five_cols=SelectKBest(mutual_info_classif,k=5)
sel_five_cols.fit(x_train.fillna(0),y_train)
x_train.columns[sel_five_cols.get_support()]
```

```
Index(['Typical Chest Pain', 'Atypical', 'Q Wave', 'FBS', 'Region RWMA'], dtype='object')
```

```
pd.value_counts(df2['Cath'])
```

```
Cad      216
Normal   87
Name: Cath, dtype: int64
```

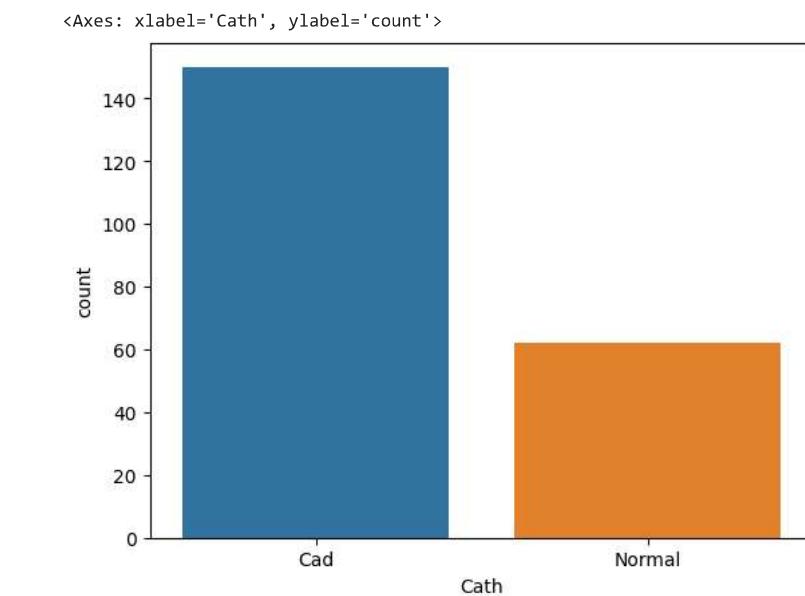
```
sns.countplot(x=df2['Cath'])
```



```
pd.value_counts(y_train)
```

```
Cad      150
Normal    62
Name: Cath, dtype: int64
```

```
sns.countplot(x=y_train)
```



```
# Balancing dataset
```

```
from imblearn.over_sampling import SMOTE
```

```
# Applying SMOTE to the train set
```

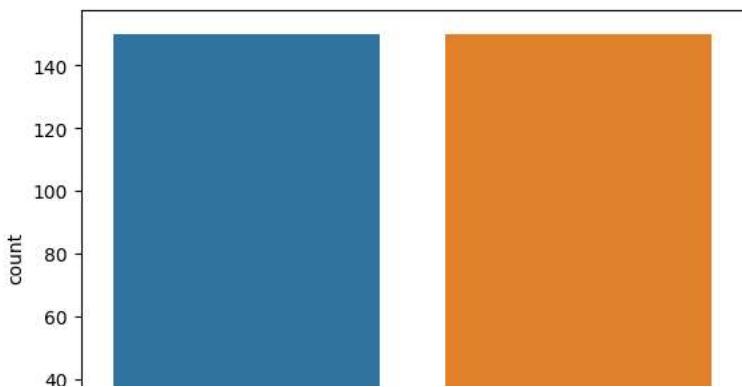
```
smote=SMOTE(random_state=27)
smote_x_train,smote_y_train=smote.fit_resample(x_train,y_train)
```

```
from imblearn.over_sampling._smote.base import Counter
print('Before',Counter(y_train))
print('After',Counter(smote_y_train))
```

```
Before Counter({'Cad': 150, 'Normal': 62})
After Counter({'Cad': 150, 'Normal': 150})
```

```
sns.countplot(x=smote_y_train)
```

```
<Axes: xlabel='Cath', ylabel='count'>
```



```
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
smote_x_train=scaler.fit_transform(smote_x_train)
x_test=scaler.fit_transform(x_test)
```

```
from sklearn.neighbors import KNeighborsClassifier  
model=KNeighborsClassifier(n_neighbors=5)  
model.fit(smote_x_train,smote_y_train)  
y_pred=model.predict(x_test)  
y_pred
```

```
array(['Normal', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad', 'Cad', 'Cad',
       'Normal', 'Cad', 'Cad', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad',
       'Cad', 'Cad', 'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad',
       'Cad', 'Normal', 'Normal', 'Normal', 'Cad', 'Normal', 'Cad', 'Cad',
       'Cad', 'Normal', 'Normal', 'Normal', 'Cad', 'Normal', 'Normal', 'Cad',
       'Normal', 'Normal', 'Normal', 'Normal', 'Cad', 'Normal', 'Normal', 'Normal',
       'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal'],
      dtype=object)
```

```
from sklearn.metrics import confusion_matrix,accuracy_score,classification_report,ConfusionMatrixDisplay
```

```
result=confusion_matrix(y_test,y_pred)  
result
```

```
array([[54, 12],  
       [ 5, 20]])
```

```
score=accuracy_score(y_test,y_pred)  
score
```

0.8131868131868132

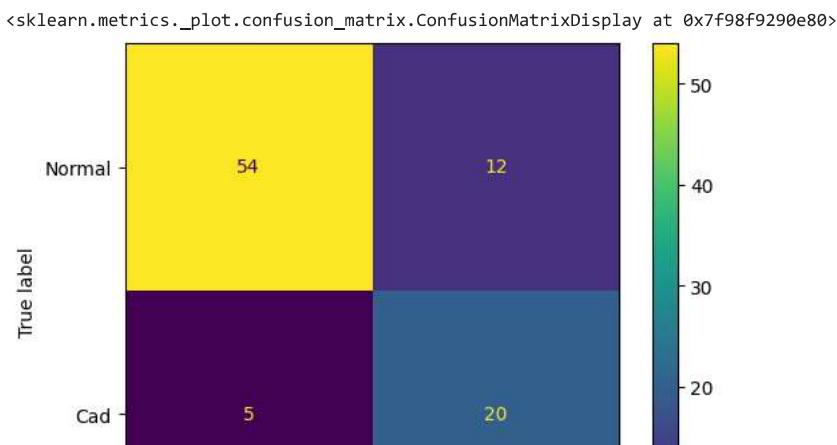
```
report=classification_report(y_test,y_pred)  
report
```

```

precision    recall   f1-score   support\n\n
          Cad      0.92      0.82      0.86      66\n
2    0.80      0.70      25\n\naccuracy
                           0.81      91\n  macro avg      0.77      0.81      0.78
91\nweighted avg      0.84      0.81      0.82      91\n

```

```
ic=['Normal','Cad']
display=ConfusionMatrixDisplay(result,display_labels=ic)
display.plot()
```



Model Creation using Naive Bayes

```
from sklearn.naive_bayes import MultinomialNB
model=MultinomialNB()
model.fit(smote_x_train,smote_y_train)
y_pred=model.predict(x_test)
y_pred

array(['Normal', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad', 'Cad',
       'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad',
       'Cad', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad',
       'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad',
       'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad',
       'Cad', 'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad',
       'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Normal', 'Normal',
       'Cad', 'Cad', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal',
       'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal'],
      dtype='|<U6')

from sklearn.metrics import confusion_matrix,accuracy_score,classification_report,ConfusionMatrixDisplay

result=confusion_matrix(y_test,y_pred)
result

array([[62,  4],
       [ 6, 19]])

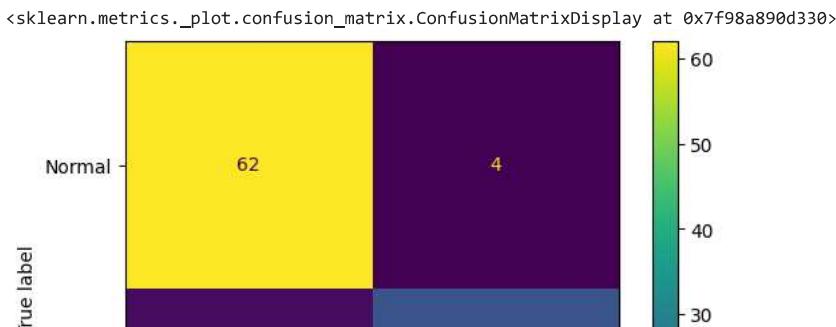
score=accuracy_score(y_test,y_pred)
score

0.8901098901098901

report=classification_report(y_test,y_pred)
report

          precision    recall  f1-score   support\n\n
  mal       0.83     0.76     0.79      25\n\n      accuracy
  0.87       0.85     0.86     0.91\nweighted avg       0.89      0.89      0.91\n\n

ic=['Normal','Cad']
display=ConfusionMatrixDisplay(result,display_labels=ic)
display.plot()
```



Model Creation using SVM

```
from sklearn.svm import SVC
model=SVC()
model.fit(smote_x_train,smote_y_train)
y_pred=model.predict(x_test)
y_pred

array(['Normal', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad', 'Cad',
       'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad', 'Cad',
       'Cad', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad', 'Normal',
       'Normal', 'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal',
       'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal',
       'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal'],
      dtype=object)
```

```
from sklearn.metrics import confusion_matrix,accuracy_score,classification_report,ConfusionMatrixDisplay
```

```
result=confusion_matrix(y_test,y_pred)
result
```

```
array([[60,  6],
       [ 6, 19]])
```

```
score=accuracy_score(y_test,y_pred)
score
```

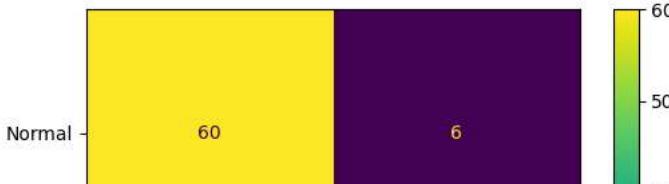
```
0.8681318681318682
```

```
report=classification_report(y_test,y_pred)
report
```

```
'          precision    recall   f1-score   support\n\n          Cad      0.91      0.91      0.91   66\\n\n 91\\nweighted avg     0.87      0.87      0.87     91\\n'
```

```
ic=['Normal','Cad']
display=ConfusionMatrixDisplay(result,display_labels=ic)
display.plot()
```

```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f98a89c27d0>
```



Model Creation using Decision Tree

```
from sklearn.tree import DecisionTreeClassifier
model=DecisionTreeClassifier()
model.fit(smote_x_train,smote_y_train)
y_pred=model.predict(x_test)
y_pred

array(['Normal', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad', 'Normal',
       'Normal', 'Cad', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Normal', 'Normal', 'Cad', 'Normal', 'Cad',
       'Cad', 'Cad', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal',
       'Normal', 'Normal', 'Normal', 'Normal', 'Normal', 'Normal'],
      dtype=object)

from sklearn.metrics import confusion_matrix,accuracy_score,classification_report,ConfusionMatrixDisplay

result=confusion_matrix(y_test,y_pred)
result

array([[54, 12],
       [11, 14]])

score=accuracy_score(y_test,y_pred)
score

0.7472527472527473

report=classification_report(y_test,y_pred)
report

          precision    recall  f1-score   support\n\n
  mal      0.54      0.56      0.55     25\n\n      accuracy
  0.68      0.69      0.69      0.75     91\nweighted avg
                           0.75      0.75     91\n'

ic=['Normal','Cad']
display=ConfusionMatrixDisplay(result,display_labels=ic)
display.plot()
```



Model creation using Random Forest Classifier

```
from sklearn.ensemble import RandomForestClassifier
model=RandomForestClassifier(n_estimators=10,criterion='entropy')
model.fit(smote_x_train,smote_y_train)
y_pred=model.predict(x_test)
y_pred
```

```
array(['Normal', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad', 'Cad',
       'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad',
       'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Cad', 'Normal',
       'Normal', 'Cad', 'Cad', 'Normal', 'Normal', 'Cad', 'Cad', 'Cad',
       'Normal', 'Normal', 'Cad', 'Cad', 'Normal', 'Normal', 'Cad', 'Normal',
       'Normal', 'Cad', 'Cad', 'Cad', 'Cad', 'Normal', 'Cad', 'Normal',
       'Cad', 'Cad', 'Cad', 'Normal', 'Cad', 'Normal', 'Cad', 'Cad',
       'Cad', 'Normal', 'Cad', 'Cad', 'Normal', 'Cad', 'Normal'],
      dtype=object)
```

```
result=confusion_matrix(y_test,y_pred)
result
```

```
array([[59,  7],
       [10, 15]])
```

```
score=accuracy_score(y_test,y_pred)
score
```

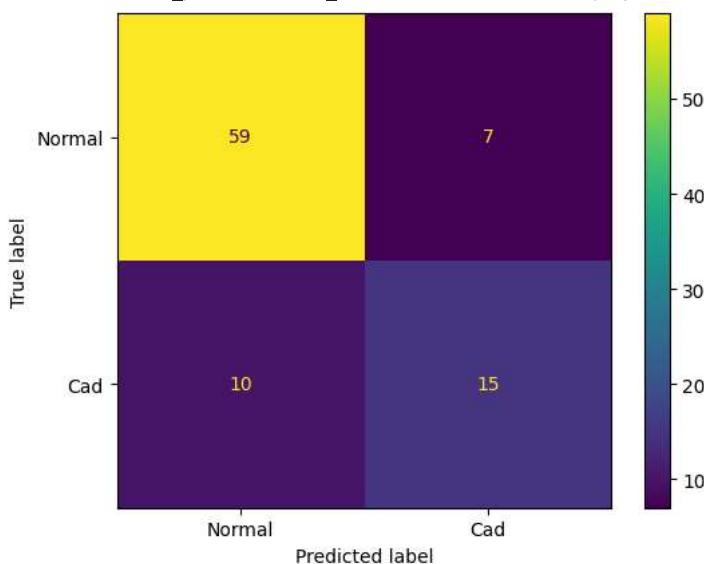
```
0.8131868131868132
```

```
report=classification_report(y_test,y_pred)
report
```

| | precision | recall | f1-score | support | | | | | |
|----|--------------|--------|----------|---------|----------|------|------|------|-----|
| 8 | 0.60 | 0.64 | 0.62 | 25 | accuracy | 0.81 | 0.89 | 0.87 | 66 |
| 91 | weighted avg | 0.81 | 0.81 | 0.81 | n | 0.77 | 0.75 | 0.76 | 0.6 |

```
ic=['Normal','Cad']
display=ConfusionMatrixDisplay(result,display_labels=ic)
display.plot()
```

<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f98a875d000>



```
import time
from xgboost import XGBClassifier
```

https://colab.research.google.com/drive/1J49zXxE6Dljhd0FykUP0PNQbvfjfj_1j#scrollTo=Lfk6wA9aNHo&printMode=true