In [1]:
```
import numpy as np                                    #Notebook make by 181B226
import pandas as pd
```

In [2]:
```
df = pd.read_csv(r'C:\Users\Admin\Downloads\processed.cleveland.data')
df.head()
```

Out[2]:

	63.0	1.0	1.0.1	145.0	233.0	1.0.2	2.0	150.0	0.0	2.3	3.0	0.0.1	6.0	0
0	67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0	2
1	67.0	1.0	4.0	120.0	229.0	0.0	2.0	129.0	1.0	2.6	2.0	2.0	7.0	1
2	37.0	1.0	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	3.0	0.0	3.0	0
3	41.0	0.0	2.0	130.0	204.0	0.0	2.0	172.0	0.0	1.4	1.0	0.0	3.0	0
4	56.0	1.0	2.0	120.0	236.0	0.0	0.0	178.0	0.0	0.8	1.0	0.0	3.0	0

In [3]:
```
print(df.info())
print(df.shape)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 302 entries, 0 to 301
Data columns (total 14 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   63.0    302 non-null    float64
 1   1.0     302 non-null    float64
 2   1.0.1   302 non-null    float64
 3   145.0   302 non-null    float64
 4   233.0   302 non-null    float64
 5   1.0.2   302 non-null    float64
 6   2.0     302 non-null    float64
 7   150.0   302 non-null    float64
 8   0.0     302 non-null    float64
 9   2.3     302 non-null    float64
 10  3.0     302 non-null    float64
 11  0.0.1   302 non-null    object 
 12  6.0     302 non-null    object 
 13  0       302 non-null    int64  
dtypes: float64(11), int64(1), object(2)
memory usage: 33.2+ KB
None
(302, 14)
```

In [4]:
```
df.columns=['Age','Sex','CP','restBP','Chol','FBS','RestCG','Thalach','Exang',
'Oldpeal','Slope','CA','Thal','HD']
df.head()
```

Out[4]:

	Age	Sex	CP	restBP	Chol	FBS	RestCG	Thalach	Exang	Oldpeal	Slope	CA	Thal	HD
0	67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0	2
1	67.0	1.0	4.0	120.0	229.0	0.0	2.0	129.0	1.0	2.6	2.0	2.0	7.0	1
2	37.0	1.0	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	3.0	0.0	3.0	0
3	41.0	0.0	2.0	130.0	204.0	0.0	2.0	172.0	0.0	1.4	1.0	0.0	3.0	0
4	56.0	1.0	2.0	120.0	236.0	0.0	0.0	178.0	0.0	0.8	1.0	0.0	3.0	0

In [5]:
```
df.isnull().any()
```

Out[5]:
```
Age         False
Sex         False
CP          False
restBP      False
Chol        False
FBS         False
RestCG      False
Thalach     False
Exang       False
Oldpeal     False
Slope       False
CA          False
Thal        False
HD          False
dtype: bool
```

In [6]:
```
df.isnull().sum()                #No missing value(null)
```

Out[6]:
```
Age         0
Sex         0
CP          0
restBP      0
Chol        0
FBS         0
RestCG      0
Thalach     0
Exang       0
Oldpeal     0
Slope       0
CA          0
Thal        0
HD          0
dtype: int64
```

In [7]:
```
df.dtypes
```

Out[7]:
```
Age         float64
Sex         float64
CP          float64
restBP      float64
Chol        float64
FBS         float64
RestCG      float64
Thalach     float64
Exang       float64
Oldpeal     float64
Slope       float64
CA           object
Thal         object
HD            int64
dtype: object
```

In [8]:
```
#here age column is in float type but it should be integer
df['Age']=df['Age'].astype(int)
df['Sex']=df['Sex'].astype(int)
df.dtypes
```

Out[8]:
```
Age           int32
Sex           int32
CP          float64
restBP      float64
Chol        float64
FBS         float64
RestCG      float64
Thalach     float64
Exang       float64
Oldpeal     float64
Slope       float64
CA           object
Thal         object
HD            int64
dtype: object
```

In [9]:
```
df.head()
```

Out[9]:

	Age	Sex	CP	restBP	Chol	FBS	RestCG	Thalach	Exang	Oldpeal	Slope	CA	Thal	HD
0	67	1	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0	2
1	67	1	4.0	120.0	229.0	0.0	2.0	129.0	1.0	2.6	2.0	2.0	7.0	1
2	37	1	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	3.0	0.0	3.0	0
3	41	0	2.0	130.0	204.0	0.0	2.0	172.0	0.0	1.4	1.0	0.0	3.0	0
4	56	1	2.0	120.0	236.0	0.0	0.0	178.0	0.0	0.8	1.0	0.0	3.0	0

In [10]:
```
df.info()              #check garbage value other than null value like '?'
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 302 entries, 0 to 301
Data columns (total 14 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   Age      302 non-null    int32  
 1   Sex      302 non-null    int32  
 2   CP       302 non-null    float64
 3   restBP   302 non-null    float64
 4   Chol     302 non-null    float64
 5   FBS      302 non-null    float64
 6   RestCG   302 non-null    float64
 7   Thalach  302 non-null    float64
 8   Exang    302 non-null    float64
 9   Oldpeal  302 non-null    float64
 10  Slope    302 non-null    float64
 11  CA       302 non-null    object 
 12  Thal     302 non-null    object 
 13  HD       302 non-null    int64  
dtypes: float64(9), int32(2), int64(1), object(2)
memory usage: 30.8+ KB
```

In [16]:
```
df['CP']
```

Out[16]:
```
0      4.0
1      4.0
2      3.0
3      2.0
4      2.0
      ... 
297    1.0
298    4.0
299    4.0
300    2.0
301    3.0
Name: CP, Length: 302, dtype: float64
```

In [19]:
```
df['CA'][5]       #values are in string but its type is int
```

Out[19]:
```
'2.0'
```

In [21]:
```
# to see ?
len(df.loc[df['CA']=='?'])
```

Out[21]:
```
4
```

In [22]:
```
len(df.loc[df['Slope']=='?'])
```

Out[22]:
```
0
```

In [23]:
```
len(df.loc[df['HD']=='?'])
```

Out[23]:
```
0
```

In [24]:
```
len(df.loc[df['Thal']=='?'])     # to see ?
```

Out[24]:
```
2
```

In [28]:
```
df=df.loc[df['CA']!='?']
df=df.loc[df['Thal']!='?']       #remove ? because it is very less
```

In [29]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 296 entries, 0 to 300
Data columns (total 14 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   Age      296 non-null    int32  
 1   Sex      296 non-null    int32  
 2   CP       296 non-null    float64
 3   restBP   296 non-null    float64
 4   Chol     296 non-null    float64
 5   FBS      296 non-null    float64
 6   RestCG   296 non-null    float64
 7   Thalach  296 non-null    float64
 8   Exang    296 non-null    float64
 9   Oldpeal  296 non-null    float64
 10  Slope    296 non-null    float64
 11  CA       296 non-null    object 
 12  Thal     296 non-null    object 
 13  HD       296 non-null    int64  
dtypes: float64(9), int32(2), int64(1), object(2)
memory usage: 32.4+ KB
```

In [30]:
```
print(len(df.loc[df['CA']=='?']))
print(len(df.loc[df['Thal']=='?']))
```

```
0
0
```

In [31]:
```
print(len(df.loc[df['CP']=='?']))
```

```
0
```

In [ ]:
```

```