In [1]:
```
import os
os.getcwd()
```

Out[1]:
```
'/home/sapatevaibhav/Documents/ML'
```

In [2]:
```
import pandas as pd
```

In [5]:
```
df = pd.read_csv('Heart.csv')
```

In [6]:
```
df.head()
```

Out[6]:

	ID	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

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

Out[7]:
```
(303, 15)
```

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

Out[8]:

	ID	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	O
0	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	
298	False	False	False	False	False	False	False	False	False	False	
299	False	False	False	False	False	False	False	False	False	False	
300	False	False	False	False	False	False	False	False	False	False	
301	False	False	False	False	False	False	False	False	False	False	
302	False	False	False	False	False	False	False	False	False	False	

303 rows × 15 columns

In [10]:
```
df.isnull().sum()
```

Out[10]:
```
ID          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
```

In [11]:
```
df.count()
```

Out[11]:
```
ID          303
Age         303
Sex         303
ChestPain   303
RestBP      303
Chol        303
Fbs         303
RestECG     303
MaxHR       303
ExAng       303
Oldpeak     303
Slope       303
Ca          299
Thal        301
AHD         303
dtype: int64
```

In [12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 15 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   ID         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
```

In [15]:
```
df.dtypes
```

Out[15]:
```
ID             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
```

In [16]:
```
df == 0
```

Out[16]:

	ID	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	O
0	False	False	False	False	False	False	False	False	False	True	
1	False	False	False	False	False	False	True	False	False	False	
2	False	False	False	False	False	False	True	False	False	False	
3	False	False	False	False	False	False	True	True	False	True	
4	False	False	True	False	False	False	True	False	False	True	
...	...	...	...	...	...	...	...	...	...	...	
298	False	False	False	False	False	False	True	True	False	True	
299	False	False	False	False	False	False	False	True	False	True	
300	False	False	False	False	False	False	True	True	False	False	
301	False	False	True	False	False	False	True	False	False	True	
302	False	False	False	False	False	False	True	True	False	True	

303 rows × 15 columns

In [17]:
```
df[df == 0]
```

Out[17]:

	ID	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	0.0	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	0.0	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	0.0	0.0	NaN	0.0	NaN
4	NaN	NaN	0.0	NaN	NaN	NaN	0.0	NaN	NaN	0.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...
298	NaN	NaN	NaN	NaN	NaN	NaN	0.0	0.0	NaN	0.0	NaN
299	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	NaN	0.0	NaN
300	NaN	NaN	NaN	NaN	NaN	NaN	0.0	0.0	NaN	NaN	NaN
301	NaN	NaN	0.0	NaN	NaN	NaN	0.0	NaN	NaN	0.0	0.0
302	NaN	NaN	NaN	NaN	NaN	NaN	0.0	0.0	NaN	0.0	0.0

303 rows × 15 columns

In [19]:
```
df [df == 0].count()
```

Out[19]:
```
ID             0
Age            0
Sex           97
ChestPain      0
RestBP         0
Chol           0
Fbs          258
RestECG      151
MaxHR          0
ExAng        204
Oldpeak       99
Slope          0
Ca           176
Thal           0
AHD            0
dtype: int64
```

In [20]:
```
df.columns
```

Out[20]:
```
Index(['ID', 'Age', 'Sex', 'ChestPain', 'RestBP', 'Chol', 'Fbs', 'RestECG',
       'MaxHR', 'ExAng', 'Oldpeak', 'Slope', 'Ca', 'Thal', 'AHD'],
      dtype='object')
```

In [21]:
```
df['Age'].mean()
```

Out[21]:
```
54.43894389438944
```

In [22]:
```
df[['Age', 'Sex', 'ChestPain', 'RestBP', 'Chol']]
```

Out[22]:

	Age	Sex	ChestPain	RestBP	Chol
0	63	1	typical	145	233
1	67	1	asymptomatic	160	286
2	67	1	asymptomatic	120	229
3	37	1	nonanginal	130	250
4	41	0	nontypical	130	204
...	...	...	...	...	...
298	45	1	typical	110	264
299	68	1	asymptomatic	144	193
300	57	1	asymptomatic	130	131
301	57	0	nontypical	130	236
302	38	1	nonanginal	138	175

303 rows × 5 columns

In [23]:
```
from sklearn.model_selection import train_test_split
```

In [25]:
```
train, test = train_test_split(df, random_state=0, test_size=0.25)
```

In [26]:
```
train.shape
```

Out[26]: (227, 15)

In [27]:
```
test.shape
```

Out[27]: (76, 15)

In [ ]:
```

```