In [1]:

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

In [2]:

```
df = pd.read_csv("Heart.csv")
```

In [3]:

```
df.head(8)
```

Out[3]:

| | Unnamed: 0 | 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 |
| **5** | 6 | 56 | 1 | nontypical | 120 | 236 | 0 | 0 | 178 | 0 | 0.8 |
| **6** | 7 | 62 | 0 | asymptomatic | 140 | 268 | 0 | 2 | 160 | 0 | 3.6 |
| **7** | 8 | 57 | 0 | asymptomatic | 120 | 354 | 0 | 0 | 163 | 1 | 0.6 |

In [5]:

```
df
```

Out[5]:

| | Unnamed: 0 | Age | Sex | ChestPain | RestBP | Chol | Fbs | RestECG | MaxHR | ExAng | Oldpe |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 63 | 1 | typical | 145 | 233 | 1 | 2 | 150 | 0 | |
| **1** | 2 | 67 | 1 | asymptomatic | 160 | 286 | 0 | 2 | 108 | 1 | |
| **2** | 3 | 67 | 1 | asymptomatic | 120 | 229 | 0 | 2 | 129 | 1 | |
| **3** | 4 | 37 | 1 | nonanginal | 130 | 250 | 0 | 0 | 187 | 0 | |
| **4** | 5 | 41 | 0 | nontypical | 130 | 204 | 0 | 2 | 172 | 0 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **298** | 299 | 45 | 1 | typical | 110 | 264 | 0 | 0 | 132 | 0 | |
| **299** | 300 | 68 | 1 | asymptomatic | 144 | 193 | 1 | 0 | 141 | 0 | |
| **300** | 301 | 57 | 1 | asymptomatic | 130 | 131 | 0 | 0 | 115 | 1 | |
| **301** | 302 | 57 | 0 | nontypical | 130 | 236 | 0 | 2 | 174 | 0 | |
| **302** | 303 | 38 | 1 | nonanginal | 138 | 175 | 0 | 0 | 173 | 0 | |

303 rows × 15 columns

In [7]:

```
df.isnull()
```

Out[7]:

| | Unnamed: 0 | Age | Sex | ChestPain | RestBP | Chol | Fbs | RestECG | MaxHR | ExAng | Oldp |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | False | False | False | False | False | False | False | False | False | False | F |
| **1** | False | False | False | False | False | False | False | False | False | False | F |
| **2** | False | False | False | False | False | False | False | False | False | False | F |
| **3** | False | False | False | False | False | False | False | False | False | False | F |
| **4** | False | False | False | False | False | False | False | False | False | False | F |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **298** | False | False | False | False | False | False | False | False | False | False | F |
| **299** | False | False | False | False | False | False | False | False | False | False | F |
| **300** | False | False | False | False | False | False | False | False | False | False | F |
| **301** | False | False | False | False | False | False | False | False | False | False | F |
| **302** | False | False | False | False | False | False | False | False | False | False | F |

303 rows × 15 columns

In [9]:

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

Out[9]:

```
Unnamed: 0    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 [10]:

```
print("Total missing values: ", df.isnull().sum().sum())
```

```
Total missing values:  6
```

In [46]:

```
marks = { "English":[1, 2, 3, 4],
        "Maths":[11, 21, 31, 41],
        "IP":[1, 2, 3, 4]
        }
result = pd.DataFrame(marks, index=["U1", "U2", "U3", "U4"])
```

In [47]:

```
result
```

Out[47]:

| | English | Maths | IP |
|---|---|---|---|
| **U1** | 1 | 11 | 1 |
| **U2** | 2 | 21 | 2 |
| **U3** | 3 | 31 | 3 |
| **U4** | 4 | 41 | 4 |

In [15]:

```
result.to_csv("marks.csv")
```

In [17]:

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

In [18]:

```
df1
```

Out[18]:

| | Unnamed: 0 | English | Maths | IP |
|---|---|---|---|---|
| **0** | U1 | 1 | 11 | 1 |
| **1** | U2 | 2 | 21 | 2 |
| **2** | U3 | 3 | 31 | 3 |
| **3** | U4 | 4 | 41 | 4 |

In [22]:

```
lst = ['Geeks', 'For', 'Geeks']


df = pd.DataFrame(lst,index=["A", "B", "C"], columns=["Portal"])
df
```

Out[22]:

| | Portal |
|---|---|
| **A** | Geeks |
| **B** | For |
| **C** | Geeks |

In [31]:

```
data = {'Name': ['Tom', 'nick', 'krish', 'jack'],
        'Age': [20, 21, 19, 18]}

# Create DataFrame
df = pd.DataFrame(data,index=["S1", "S2", "S3", "S4"])

# Print the output.
df
```

Out[31]:

| | Name | Age |
|---|---|---|
| **S1** | Tom | 20 |
| **S2** | nick | 21 |
| **S3** | krish | 19 |
| **S4** | jack | 18 |

In [33]:

```
data = [10,20,30,40,50,60]

# Create the pandas DataFrame with column name is provided explicitly
df = pd.DataFrame(data, columns=['Numbers'])
# df = pd.DataFrame(data)

# print dataframe.
df
```

Out[33]:

| | Numbers |
|---|---|
| **0** | 10 |
| **1** | 20 |
| **2** | 30 |
| **3** | 40 |
| **4** | 50 |
| **5** | 60 |

In [35]:

```
df2 = pd.read_csv("../33208/Dataset/marks.csv")
df2
```

Out[35]:

| | Unnamed: 0 | English | Maths | IP |
|---|---|---|---|---|
| **0** | U1 | 1 | 11 | 1 |
| **1** | U2 | 2 | 21 | 2 |
| **2** | U3 | 3 | 31 | 3 |
| **3** | U4 | 4 | 41 | 4 |

In [37]:

```
# data = [{'b': 2, 'c': 3}, {'a': 10, 'b': 20, 'c': 30}]
data = [{'a': 10, 'b': 20, 'c': 30},{'b': 2, 'c': 3} ]


df = pd.DataFrame(data, index=['first', 'second'])

df
```

Out[37]:

| | a | b | c |
|---|---|---|---|
| **first** | 10.0 | 20 | 30 |
| **second** | NaN | 2 | 3 |

In [58]:

```
d1 = pd.Series([10, 20, 30, 40],
                        index=['a', 'b', 'c', 'd'])
df5 = pd.DataFrame(d1, columns=["Age"])
df5
```

Out[58]:

| | Age |
|---|---|
| **a** | 10 |
| **b** | 20 |
| **c** | 30 |
| **d** | 40 |

In [54]:

```
df5.to_csv("Age.csv")
```

In [56]:

```
df6 = pd.read_csv("Age.csv")
df6
```

Out[56]:

| | Unnamed: 0 | Age |
|---|---|---|
| **0** | a | 10 |
| **1** | b | 20 |
| **2** | c | 30 |
| **3** | d | 40 |