# STASTICAL DESCRIPTION ON DATA

In [1]:

```
#Aim: To Perform Stastical Description on Data
#Exp no:3
#Name:Khushi Chandrashekhar Satpute
#Sec:3rd B
#Roll no:43
#Sub:ET-1
#Date:19/07/2024
```

In [3]:

```
import pandas as pd
```

In [4]:

```
import os
```

In [7]:

```
os.getcwd()
```

Out[7]:

```
'C:\\Users\\asus\\Downloads'
```

In [9]:

```
os.chdir("C:\\Users\\asus\\Desktop")
```

In [11]:

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

In [13]:

```
df.head()
```

Out[13]:

| | male | age | education | currentSmoker | cigsPerDay | BPMeds | prevalentStroke | prevalentHyp | diabetes |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 39 | 4.0 | 0 | 0.0 | 0.0 | 0 | 0 | 0 |
| **1** | 0 | 46 | 2.0 | 0 | 0.0 | 0.0 | 0 | 0 | 0 |
| **2** | 1 | 48 | 1.0 | 1 | 20.0 | 0.0 | 0 | 0 | 0 |
| **3** | 0 | 61 | 3.0 | 1 | 30.0 | 0.0 | 0 | 1 | 0 |
| **4** | 0 | 46 | 3.0 | 1 | 23.0 | 0.0 | 0 | 0 | 0 |

In [15]:

```
df.head(100)
```

Out[15]:

| | male | age | education | currentSmoker | cigsPerDay | BPMeds | prevalentStroke | prevalentHyp | diabete |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 39 | 4.0 | 0 | 0.0 | 0.0 | 0 | 0 | |
| **1** | 0 | 46 | 2.0 | 0 | 0.0 | 0.0 | 0 | 0 | |
| **2** | 1 | 48 | 1.0 | 1 | 20.0 | 0.0 | 0 | 0 | |

| | male | age | education | currentSmoker | cigsPerDay | BPMeds | prevalentStroke | prevalentHyp | diabete |
|---|---|---|---|---|---|---|---|---|---|
| **3** | 0 | 61 | 3.0 | 1 | 30.0 | 0.0 | 0 | 1 | |
| **4** | 0 | 46 | 3.0 | 1 | 23.0 | 0.0 | 0 | 0 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | . |
| **95** | 0 | 65 | 3.0 | 0 | 0.0 | 0.0 | 0 | 0 | |
| **96** | 0 | 63 | 4.0 | 1 | 20.0 | 0.0 | 0 | 0 | |
| **97** | 0 | 40 | 2.0 | 0 | 0.0 | 0.0 | 0 | 0 | |
| **98** | 0 | 56 | 1.0 | 0 | 0.0 | 0.0 | 0 | 1 | |
| **99** | 0 | 56 | 1.0 | 1 | 15.0 | 0.0 | 0 | 0 | |

100 rows × 16 columns

In [17]:

```
df.tail()
```

Out[17]:

| | male | age | education | currentSmoker | cigsPerDay | BPMeds | prevalentStroke | prevalentHyp | diabe |
|---|---|---|---|---|---|---|---|---|---|
| **4233** | 1 | 50 | 1.0 | 1 | 1.0 | 0.0 | 0 | 1 | |
| **4234** | 1 | 51 | 3.0 | 1 | 43.0 | 0.0 | 0 | 0 | |
| **4235** | 0 | 48 | 2.0 | 1 | 20.0 | NaN | 0 | 0 | |
| **4236** | 0 | 44 | 1.0 | 1 | 15.0 | 0.0 | 0 | 0 | |
| **4237** | 0 | 52 | 2.0 | 0 | 0.0 | 0.0 | 0 | 0 | |

In [19]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4238 entries, 0 to 4237
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   male             4238 non-null   int64  
 1   age              4238 non-null   int64  
 2   education        4133 non-null   float64
 3   currentSmoker    4238 non-null   int64  
 4   cigsPerDay       4209 non-null   float64
 5   BPMeds           4185 non-null   float64
 6   prevalentStroke  4238 non-null   int64  
 7   prevalentHyp     4238 non-null   int64  
 8   diabetes         4238 non-null   int64  
 9   totChol          4188 non-null   float64
 10  sysBP            4238 non-null   float64
 11  diaBP            4238 non-null   float64
 12  BMI              4219 non-null   float64
 13  heartRate        4237 non-null   float64
 14  glucose          3850 non-null   float64
 15  TenYearCHD       4238 non-null   int64  
dtypes: float64(9), int64(7)
memory usage: 529.9 KB
```

In [21]:

```
df.shape
```

Out[21]:

```
(4238, 16)
```

In [23]:

```
df.size
```

Out[23]:

```
67808
```

In [25]:

```
df.tail(10)
```

Out[25]:

| | male | age | education | currentSmoker | cigsPerDay | BPMeds | prevalentStroke | prevalentHyp | diabe |
|---|---|---|---|---|---|---|---|---|---|
| **4228** | 0 | 50 | 1.0 | 0 | 0.0 | 0.0 | 0 | 1 | |
| **4229** | 0 | 51 | 3.0 | 1 | 20.0 | 0.0 | 0 | 1 | |
| **4230** | 0 | 56 | 1.0 | 1 | 3.0 | 0.0 | 0 | 1 | |
| **4231** | 1 | 58 | 3.0 | 0 | 0.0 | 0.0 | 0 | 1 | |
| **4232** | 1 | 68 | 1.0 | 0 | 0.0 | 0.0 | 0 | 1 | |
| **4233** | 1 | 50 | 1.0 | 1 | 1.0 | 0.0 | 0 | 1 | |
| **4234** | 1 | 51 | 3.0 | 1 | 43.0 | 0.0 | 0 | 0 | |
| **4235** | 0 | 48 | 2.0 | 1 | 20.0 | NaN | 0 | 0 | |
| **4236** | 0 | 44 | 1.0 | 1 | 15.0 | 0.0 | 0 | 0 | |
| **4237** | 0 | 52 | 2.0 | 0 | 0.0 | 0.0 | 0 | 0 | |

In [27]:

```
df.ndim
```

Out[27]:

```
2
```

In [29]:

```
df.describe()
```

Out[29]:

| | male | age | education | currentSmoker | cigsPerDay | BPMeds | prevalentStrok |
|---|---|---|---|---|---|---|---|
| **count** | 4238.000000 | 4238.000000 | 4133.000000 | 4238.000000 | 4209.000000 | 4185.000000 | 4238.0000 |
| **mean** | 0.429212 | 49.584946 | 1.978950 | 0.494101 | 9.003089 | 0.029630 | 0.0058 |
| **std** | 0.495022 | 8.572160 | 1.019791 | 0.500024 | 11.920094 | 0.169584 | 0.0765 |
| **min** | 0.000000 | 32.000000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0000 |
| **25%** | 0.000000 | 42.000000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0000 |
| **50%** | 0.000000 | 49.000000 | 2.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0000 |
| **75%** | 1.000000 | 56.000000 | 3.000000 | 1.000000 | 20.000000 | 0.000000 | 0.0000 |
| **max** | 1.000000 | 70.000000 | 4.000000 | 1.000000 | 70.000000 | 1.000000 | 1.0000 |

In [ ]:

```

```