Data Acquisition

In [1]:
```
#Name:janhavi Nitin warghade
#Sec:3C
#Roll no.:69
#Sub.:E.T.1
```

In [1]:
```
# Aim:perform operation on data aquisition
```

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

In [3]:
```
import os
```

In [4]:
```
os.getcwd
```

Out[4]:
```
<function nt.getcwd()>
```

In [5]:
```
os.chdir("C:\\Users\DELL\\OneDrive\\Desktop")
```

In [6]:
```
data=pd.read_csv("framingham.csv")
```

In [7]:
```
data.tail(12)
```

Out[7]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
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
4238	1	40	3.0	0	0.0	0.0	0	1
4239	0	39	3.0	1	30.0	0.0	0	0

In [8]:
```
data.head(14)
```

Out[8]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	
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	
3	0	61	3.0	1	30.0	0.0	0	1	
4	0	46	3.0	1	23.0	0.0	0	0	
5	0	43	2.0	0	0.0	0.0	0	1	
6	0	63	1.0	0	0.0	0.0	0	0	
7	0	45	2.0	1	20.0	0.0	0	0	
8	1	52	1.0	0	0.0	0.0	0	1	
9	1	43	1.0	1	30.0	0.0	0	1	
10	0	50	1.0	0	0.0	0.0	0	0	
11	0	43	2.0	0	0.0	0.0	0	0	
12	1	46	1.0	1	15.0	0.0	0	1	
13	0	41	3.0	0	0.0	1.0	0	1	

In [9]:
```
data.info
```

Out[9]:
```
<bound method DataFrame.info of       male  age  education  currentSmoker  ci
gsPerDay  BPMeds  \
0           1   39        4.0              0         0.0     0.0
1           0   46        2.0              0         0.0     0.0
2           1   48        1.0              1        20.0     0.0
3           0   61        3.0              1        30.0     0.0
4           0   46        3.0              1        23.0     0.0
...       ...  ...        ...            ...         ...     ...
4235        0   48        2.0              1        20.0     NaN
4236        0   44        1.0              1        15.0     0.0
4237        0   52        2.0              0         0.0     0.0
4238        1   40        3.0              0         0.0     0.0
4239        0   39        3.0              1        30.0     0.0

      prevalentStroke  prevalentHyp  diabetes  totChol  sysBP  diaBP    BMI
\
0                   0             0         0    195.0  106.0   70.0  26.97
1                   0             0         0    250.0  121.0   81.0  28.73
2                   0             0         0    245.0  127.5   80.0  25.34
3                   0             1         0    225.0  150.0   95.0  28.58
4                   0             0         0    285.0  130.0   84.0  23.10
...               ...           ...       ...      ...    ...    ...    ...
4235                0             0         0    248.0  131.0   72.0  22.00
4236                0             0         0    210.0  126.5   87.0  19.16
4237                0             0         0    269.0  133.5   83.0  21.47
4238                0             1         0    185.0  141.0   98.0  25.60
4239                0             0         0    196.0  133.0   86.0  20.91

      heartRate  glucose  TenYearCHD
0          80.0     77.0           0
1          95.0     76.0           0
2          75.0     70.0           0
3          65.0    103.0           1
4          85.0     85.0           0
...         ...      ...         ...
4235       84.0     86.0           0
4236       86.0      NaN           0
4237       80.0    107.0           0
4238       67.0     72.0           0
4239       85.0     80.0           0

[4240 rows x 16 columns]>
```

In [10]:
```
data.describe()
```

Out[10]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevale
count	4240.000000	4240.000000	4135.000000	4240.000000	4211.000000	4187.000000	424
mean	0.429245	49.580189	1.979444	0.494104	9.005937	0.029615	
std	0.495027	8.572942	1.019791	0.500024	11.922462	0.169544	
min	0.000000	32.000000	1.000000	0.000000	0.000000	0.000000	
25%	0.000000	42.000000	1.000000	0.000000	0.000000	0.000000	
50%	0.000000	49.000000	2.000000	0.000000	0.000000	0.000000	
75%	1.000000	56.000000	3.000000	1.000000	20.000000	0.000000	
max	1.000000	70.000000	4.000000	1.000000	70.000000	1.000000	

In [ ]:

In [ ]: