

Name - Gayatri Bhakare
Roll No - 10
Class - 3rd year
Section - CSE

```
In [1]: import pandas as pd  
import os
```

```
In [2]: os.getcwd()
```

```
Out[2]: 'C:\\Users\\sarth\\Data Analytics projects\\College Proj'
```

```
In [4]: os.chdir("D:\\Users\\SARTHAK\\Picture\\Desktop")
```

```
In [5]: df = pd.read_csv("framingham.csv")
```

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

```
Out[6]:
```

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

```
In [7]: df.tail()
```

```
Out[7]:
```

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
4235	0	48	2.0	1	20.0	NaN	0	
4236	0	44	1.0	1	15.0	0.0	0	
4237	0	52	2.0	0	0.0	0.0	0	
4238	1	40	3.0	0	0.0	0.0	0	
4239	0	39	3.0	1	30.0	0.0	0	

```
In [9]: df.shape
```

```
Out[9]: (4240, 16)
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4240 entries, 0 to 4239
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   male                  4240 non-null   int64  
1   age                   4240 non-null   int64  
2   education              4135 non-null   float64
3   currentSmoker          4240 non-null   int64  
4   cigsPerDay             4211 non-null   float64
5   BPMeds                 4187 non-null   float64
6   prevalentStroke        4240 non-null   int64  
7   prevalentHyp           4240 non-null   int64  
8   diabetes               4240 non-null   int64  
9   totChol                4190 non-null   float64
10  sysBP                  4240 non-null   float64
11  diaBP                  4240 non-null   float64
12  BMI                    4221 non-null   float64
13  heartRate              4239 non-null   float64
14  glucose                3852 non-null   float64
15  TenYearCHD             4240 non-null   int64  
dtypes: float64(9), int64(7)
memory usage: 530.1 KB
```

```
In [12]: df.isnull().sum()
```

```
Out[12]: male                0
age                0
education          105
currentSmoker      0
cigsPerDay         29
BPMeds             53
prevalentStroke    0
prevalentHyp       0
diabetes           0
totChol            50
sysBP              0
diaBP              0
BMI                19
heartRate          1
glucose            388
TenYearCHD         0
dtype: int64
```

```
In [16]: column_name_changes = {'currentSmoker': 'is_smoking', 'male': 'sex'}
df = df.rename(columns=column_name_changes)
```

```
In [17]: df.dropna(axis=0,inplace=True)
df.isnull().sum()
```

```
Out[17]: sex                0
age                0
education          0
is_smoking         0
cigsPerDay         0
BPMeds             0
prevalentStroke    0
prevalentHyp       0
diabetes           0
totChol            0
sysBP              0
diaBP              0
BMI                0
heartRate          0
glucose            0
TenYearCHD         0
dtype: int64
```

```
In [ ]:
```