

DATA SPECIALIZATION

```
In [1]: #Name:Ankita Gulde  
#Roll no. : 44  
#Section : 3A
```

```
In [2]: #Aim : To Perform Data specialization
```

```
In [3]: import pandas as pd
```

```
In [4]: import os
```

```
In [5]: os.getcwd()
```

```
Out[5]: 'C:\\Users\\HP'
```

```
In [6]: os.chdir("C:\\Users\\HP\\Desktop")
```

```
In [7]: df=pd.read_csv("framingham.csv")
```

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

```
Out[8]: male age education currentSmoker cigsPerDay BPMeds prevalentStroke prevalentHyp diabetes totChol sysBP diaBP BMI  
heartR  
1 0 46 2.0 0 0.0 0.0 0 0 250.0 121.0 81.0 28.73  
2 1 / / 2 / / / / 245.0 127.5 80.0 28.58  
3 0 61 3.0 1 30.0 0.0 0 1 225.0 150.0 95.0 28.58  
4 0 / / 2 / / / / 285.0 130.0 84.0 23.10
```

```
dt.head(100)
```

```
Out[9]:
```

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevailentStroke	prevailentHyp	diabetes	totChol	sysBP	diaBP	BMI	hear
0	1	39	4.0	0	0.0	0.0	0	0	0	195.0	106.0	70.0	26.97	
1	0	46	2.0	0	0.0	0.0	0	0	0	250.0	121.0	81.0	28.73	
2	1	48	1.0	1	20.0	0.0	0	0	0	245.0	127.5	80.0	25.34	
3	0	61	3.0	1	30.0	0.0	0	1	0	225.0	150.0	95.0	28.58	
4	0	46	3.0	1	23.0	0.0	0	0	0	285.0	130.0	84.0	23.10	
...	
95	0	65	3.0	0	0.0	0.0	0	0	0	193.0	123.0	76.5	29.33	
96	0	63	4.0	1	20.0	0.0	0	0	1	239.0	134.0	80.0	26.64	
97	0	40	2.0	0	0.0	0.0	0	0	0	205.0	100.0	60.0	NaN	
98	0	56	1.0	0	0.0	0.0	0	1	0	296.0	180.0	90.0	23.72	
99	0	56	1.0	1	15.0	0.0	0	0	0	269.0	121.0	75.0	22.36	

```
df.tail()
```

```
Out[10]:
```

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevailentStroke	prevailentHyp	diabetes	totChol	sysBP	diaBP	BMI	he
4233	1	50	1.0	1	1.0	0.0	0	1	0	313.0	179.0	92.0	25.97	
4234	1	51	3.0	1	43.0	0.0	0	0	0	207.0	126.5	80.0	19.71	
4235	0	48	2.0	1	20.0	NaN	0	0	0	248.0	131.0	72.0	22.00	
4236	0	44	1.0	1	15.0	0.0	0	0	0	210.0	126.5	87.0	19.16	
4237	0	52	2.0	0	0.0	0.0	0	0	0	269.0	133.5	83.0	21.47	

In [11]:

```
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 [12]: df.shape
```

```
Out[12]: (4238, 16)
```

```
In [13]: df.size
```

```
Out[13]: 67
```

```
In [14]: df.ndim
```

```
Out[14]: 2
```

```
In [15]: df.tail(10)
```

```
Out[15]:   male  age  education  currentSmoker  cigsPerDay  BPMeds  prevalentStroke  prevalentHyp  diabetes  totChol  sysBP  diaBP  BMI  he
4228    0   50       1.0          0        0.0       0.0          0           1        1   260.0  190.0  130.0  43.67
4229    0   51       3.0          1        20.0       0.0          0           1        0   251.0  140.0  80.0   25.60
4230    0   56       1.0          1        3.0       0.0          0           1        0   268.0  170.0  102.0  22.89
4231    1   58       3.0          0        0.0       0.0          0           1        0   187.0  141.0  81.0   24.96
4232    1   68       1.0          0        0.0       0.0          0           1        0   176.0  168.0  97.0   23.14
4233    1   50       1.0          1        1.0       0.0          0           1        0   313.0  179.0  92.0   25.97
4234    1   51       3.0          1        43.0       0.0          0           0        0   207.0  126.5  80.0   19.71
4235    0   48       2.0          1        20.0      NaN          0           0        0   248.0  131.0  72.0   22.00
4236    0   44       1.0          1        15.0       0.0          0           0        0   210.0  126.5  87.0   19.16
4237    0   52       2.0          0        0.0       0.0          0           0        0   269.0  133.5  83.0   21.47
```

```
In [16]: df.describe()
```

```
Out[16]:   male      age  education  currentSmoker  cigsPerDay  BPMeds  prevalentStroke  prevalentHyp  diabetes  totChol  sysBP  diaBP  BMI  he
count  4238.000000  4238.000000  4133.000000  4238.000000  4209.000000  4185.000000  4238.000000  4238.000000  4238.000000  4188.000000
mean   0.429212    49.584946   1.978950    0.494101    9.003089    0.029630    0.005899    0.310524    0.025720    236.72
std    0.495022    8.572160   1.019791    0.500024    11.920094   0.169584    0.076587    0.462763    0.158316    44.59
min    0.000000    32.000000   1.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    107.00
25%    0.000000    42.000000   1.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    206.00
50%    0.000000    49.000000   2.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    234.00
75%    1.000000    56.000000   3.000000    1.000000    20.000000   0.000000    0.000000    1.000000    0.000000    263.00
max    1.000000    70.000000   4.000000    1.000000    70.000000   1.000000    1.000000    1.000000    1.000000    696.00
```

```
In [ ]:
```