In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

In [2]:
```
df=pd.read_csv(r"C:\Users\user\Downloads\archive (4).zip")
df
```

Out[2]:

	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
...	...	...	...	...	...	...	...	...
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 rows × 16 columns

In [3]:
```
df.head()
```

Out[3]:

cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	BMI	hea
0.0	0.0	0	0	0	195.0	106.0	70.0	26.97	
0.0	0.0	0	0	0	250.0	121.0	81.0	28.73	
20.0	0.0	0	0	0	245.0	127.5	80.0	25.34	
30.0	0.0	0	1	0	225.0	150.0	95.0	28.58	
23.0	0.0	0	0	0	285.0	130.0	84.0	23.10	

In [4]:

```
df.tail()
```

Out[4]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
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 [5]:

```
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 [6]:
```
df.describe()
```

Out[6]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevaler
count	4238.000000	4238.000000	4133.000000	4238.000000	4209.000000	4185.000000	4238
mean	0.429212	49.584946	1.978950	0.494101	9.003089	0.029630	0
std	0.495022	8.572160	1.019791	0.500024	11.920094	0.169584	0
min	0.000000	32.000000	1.000000	0.000000	0.000000	0.000000	0
25%	0.000000	42.000000	1.000000	0.000000	0.000000	0.000000	0
50%	0.000000	49.000000	2.000000	0.000000	0.000000	0.000000	0
75%	1.000000	56.000000	3.000000	1.000000	20.000000	0.000000	0
max	1.000000	70.000000	4.000000	1.000000	70.000000	1.000000	1

In [7]:
```
df.shape
```

Out[7]:
```
(4238, 16)
```

In [8]:
```
df.isnull().sum()
```

Out[8]:
```
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 [11]:
```
df['diabetes'].value_counts()
```

Out[11]:
```
diabetes
0    4129
1     109
Name: count, dtype: int64
```

In [12]:

```
df['BMI'].value_counts()
```

Out[12]:

```
BMI
22.19    18
22.54    18
23.48    18
22.91    18
23.09    16
         ..
34.13     1
23.21     1
29.13     1
19.87     1
43.67     1
Name: count, Length: 1363, dtype: int64
```

In [10]:

```
x=df.drop(columns='TenYearCHD',axis=1)
y=df['TenYearCHD']
```

In [13]:
```
print(x)
```

```
      male  age  education  currentSmoker  cigsPerDay  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   
...    ...  ...        ...            ...         ...     ...   
4233     1   50        1.0              1         1.0     0.0   
4234     1   51        3.0              1        43.0     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   

      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
...               ...           ...       ...      ...    ...    ...    ...
4233                0             1         0    313.0  179.0   92.0  25.97
4234                0             0         0    207.0  126.5   80.0  19.71
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

      heartRate  glucose  
0          80.0     77.0  
1          95.0     76.0  
2          75.0     70.0  
3          65.0    103.0  
4          85.0     85.0  
...         ...      ...  
4233       66.0     86.0  
4234       65.0     68.0  
4235       84.0     86.0  
4236       86.0      NaN  
4237       80.0    107.0  

[4238 rows x 15 columns]
```

In [14]:
```
print(y)
```

```
0       0
1       0
2       0
3       1
4       0
       ..
4233    1
4234    0
4235    0
4236    0
4237    0
Name: TenYearCHD, Length: 4238, dtype: int64
```

In [ ]: