

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

```
In [20]: import numpy as np # linear algebra
import os
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt # for data visualization
import seaborn as sns # for statistical data visualization
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

```
In [21]: os.getcwd()
```

```
Out[21]: 'D:\\Users\\SARTHAK\\Picture\\Desktop'
```

```
In [22]: os.chdir('D:\\Users\\SARTHAK\\Picture\\Desktop')
```

```
In [23]: df=pd.read_csv('framingham.csv')
```

```
In [24]: df.head()
```

```
Out[24]:
```

	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 [25]: 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 [26]: df.size

Out[26]: 67840

In [27]: df.shape

Out[27]: (4240, 16)

In [28]: df.describe()

Out[28]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	pre
count	4240.000000	4240.000000	4135.000000	4240.000000	4211.000000	4187.000000	
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 [29]: from sklearn.cluster import KMeans
from sklearn.metrics import adjusted_rand_score
```

```
In [31]: from sklearn.datasets import make_blobs

# Generate sample data
X, _ = make_blobs(n_samples=100, centers=2, n_features=2, random_state=0)

# Now you can proceed with your clustering code
from sklearn.cluster import KMeans

kmeans = KMeans(n_clusters=2, random_state=0)
kmeans.fit(X)
```

Out[31]: KMeans(n_clusters=2, random_state=0)

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.

On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
In [32]: kmeans.cluster_centers_
```

Out[32]: array([[2.14160463, 1.0291066],
 [0.89753417, 4.28323334]])

```
In [33]: kmeans.inertia_
```

Out[33]: 190.86228627180597

```
In [ ]:
```