Registration Number: 19BCE2119

Name: Gaurav Kumar Singh

Course: CSE3024 Web Mining

K Means Clustering

PROCEDURE

1) Import necessary libraries:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

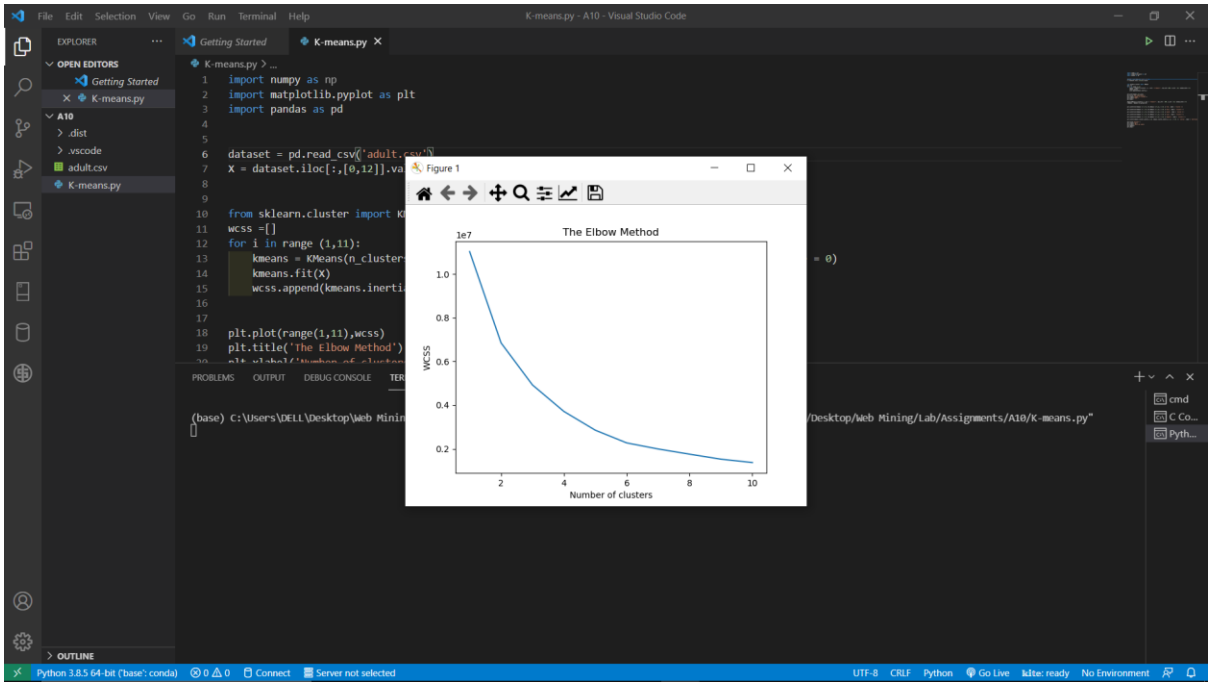
2) Import dataset as dataframe using pandas
3) Select column whose cluster relation we want to visualize.
4) Using sklearn library and its k means function, form the clusters
5) Using matplotlib library plot these clusters.

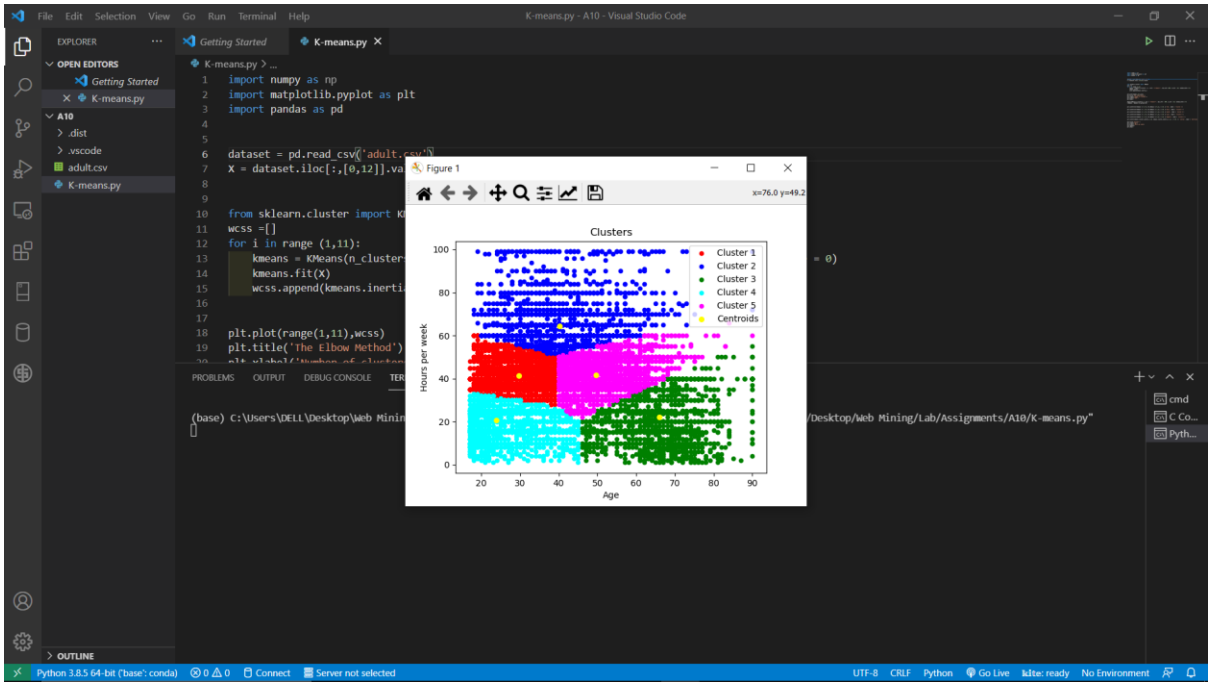
CODE

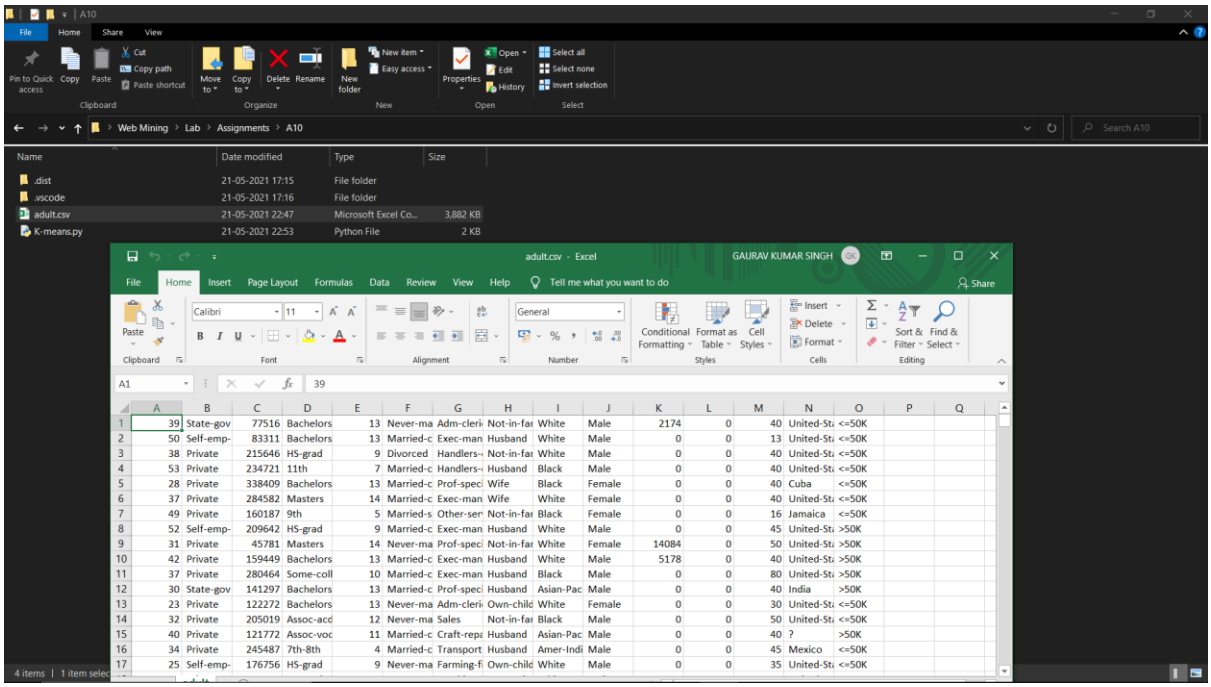
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd


dataset = pd.read_csv('adult.csv')
X = dataset.iloc[:,[0,12]].values


from sklearn.cluster import KMeans
wcss =[]
for i in range (1,11):
    kmeans = KMeans(n_clusters = i, init = 'k-
means++', max_iter =300, n_init = 10, random_state = 0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)


plt.plot(range(1,11),wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()

kmeans=KMeans(n_clusters= 5, init = 'k-
means++', max_iter = 300, n_init = 10, random_state = 0)
Y_Kmeans = kmeans.fit_predict(X)
```

```
plt.scatter(X[Y_Kmeans == 0, 0], X[Y_Kmeans == 0,1],s = 20, c='red', label = '
Cluster 1')

plt.scatter(X[Y_Kmeans == 1, 0], X[Y_Kmeans == 1,1],s = 20, c='blue', label =
'Cluster 2')

plt.scatter(X[Y_Kmeans == 2, 0], X[Y_Kmeans == 2,1],s = 20, c='green', label =
 'Cluster 3')

plt.scatter(X[Y_Kmeans == 3, 0], X[Y_Kmeans == 3,1],s = 20, c='cyan', label =
'Cluster 4')

plt.scatter(X[Y_Kmeans == 4, 0], X[Y_Kmeans == 4,1],s = 20, c='magenta', label
 = 'Cluster 5')

plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], s = 30
, c = 'yellow', label = 'Centroids')

plt.title('Clusters')
plt.xlabel('Age')
plt.ylabel('Hours per week')
plt.legend()
plt.show()
```

OUTPUT




K-means.py - A10 - Visual Studio Code
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
dataset = pd.read_csv('adult.csv')
X = dataset.iloc[:,[0,12]].va
from sklearn.cluster import K
wcss =[]
for i in range (1,11):
kmeans = KMeans(n_cluster
kmeans.fit(X)
wcss.append(kmeans.inerti
plt.plot(range(1,11),wcss)
plt.title('The Elbow Method')
Figure 1
Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Centroids
Hours per week
Age
(base) C:\Users\DELL\Desktop\Web Minin
/Desktop/Web Mining/Lab/Assignments/A10/K-means.py"


A10
Web Mining › Lab › Assignments › A10
Search A10
Name
Date modified
Type
Size
.dist
21-05-2021 17:15
File folder
.vscode
21-05-2021 17:16
File folder
adult.csv
21-05-2021 22:47
Microsoft Excel Co...
3,882 KB
K-means.py
21-05-2021 22:53
Python File
2 KB
adult.csv - Excel
GAURAV KUMAR SINGH
4 items