

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')

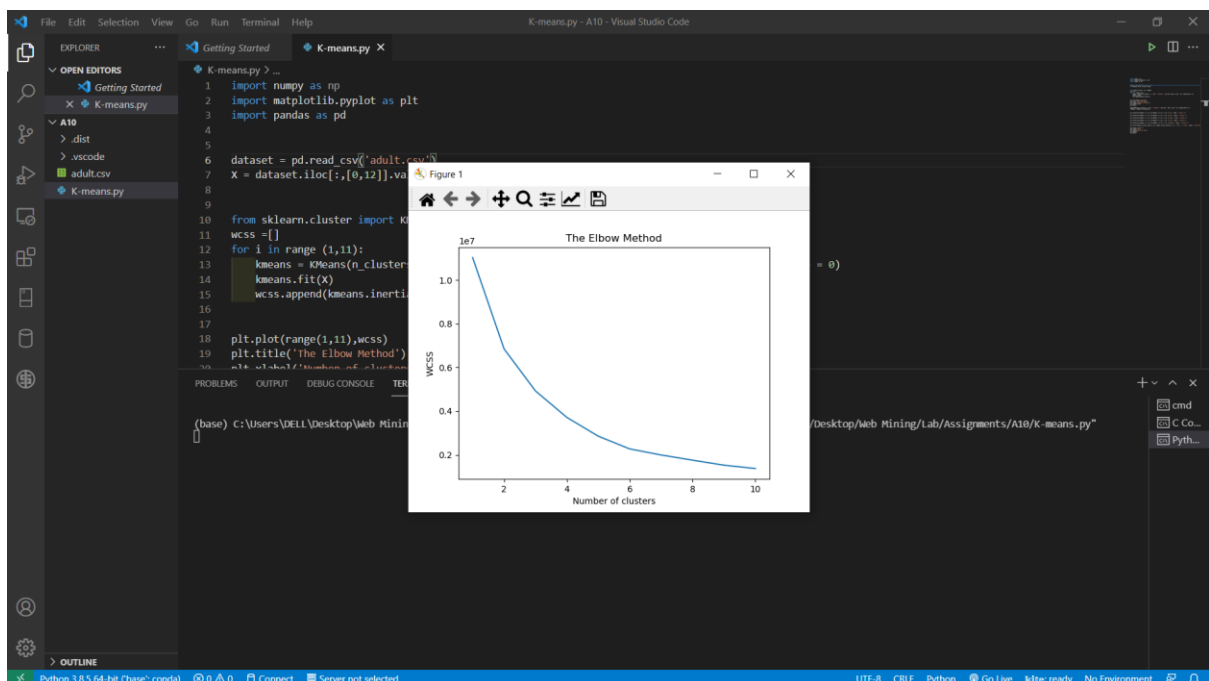
plt.scatter(X[Y_Kmeans == 4, 0], X[Y_Kmeans == 4,1],s = 20, c='magenta', label = 'Cluster 5')

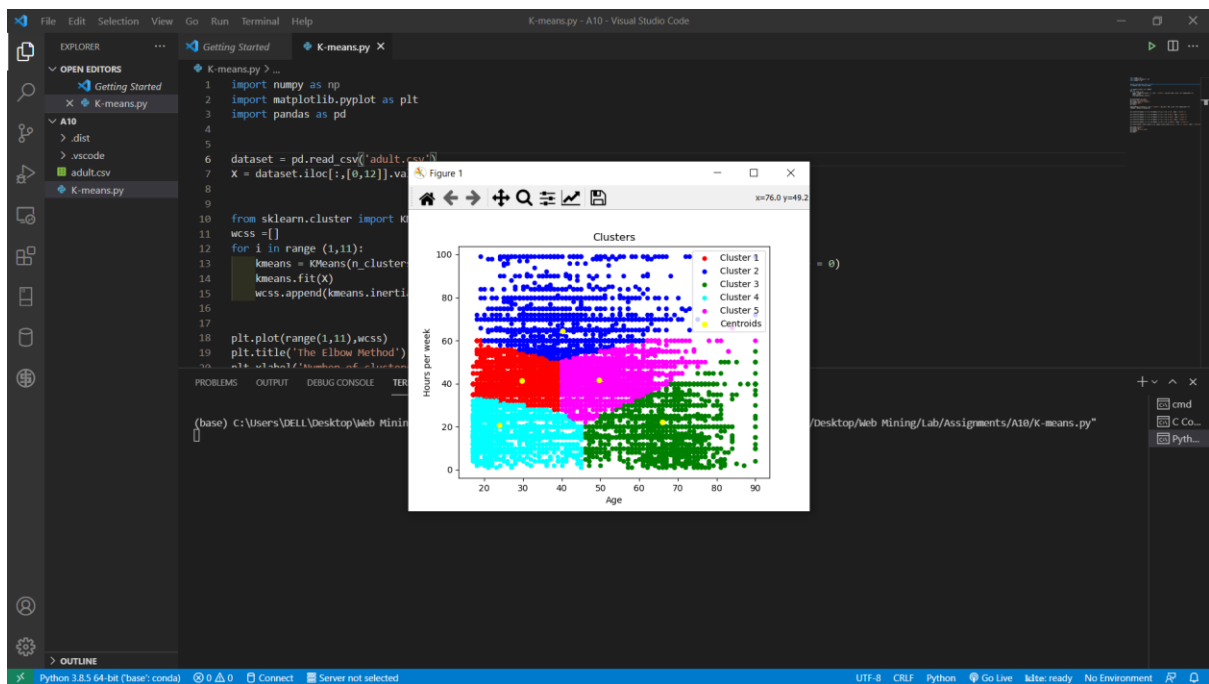
plt.scatter(kmeans.cluster_centers[:,0], kmeans.cluster_centers[:,1], s = 30, c = 'yellow', label = 'Centroids')

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

```

OUTPUT





File Explorer view showing the directory structure for the K-means project:

- A10
 - .dist
 - .vscode
 - adult.csv
 - K-means.py

Excel spreadsheet view showing the data from adult.csv:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerk	Not-in-far	White	Male	2174	0	40	United-States	<=50K		
2	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K		
3	38	Private	215646	HS-grad	9	Divorced	Handlers-support	Not-in-far	White	Male	0	0	40	United-States	<=50K		
4	53	Private	234721	11th	7	Married-civ-spouse	Handlers-support	Husband	Black	Male	0	0	40	United-States	<=50K		
5	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K		
6	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K		
7	49	Private	160187	9th	5	Married-civ-spouse	Other-service	Not-in-far	Black	Female	0	0	16	Jamaica	<=50K		
8	52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45	United-States	>50K		
9	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-far	White	Female	14084	0	50	United-States	>50K		
10	42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	5178	0	40	United-States	>50K		
11	37	Private	280464	Some-col-higher	10	Married-civ-spouse	Exec-managerial	Husband	Black	Male	0	0	80	United-States	>50K		
12	30	State-gov	141297	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	Asian-Pac-Islander	Male	0	0	40	India	>50K		
13	23	Private	122272	Bachelors	13	Never-married	Adm-clerk	Own-child	White	Female	0	0	30	United-States	<=50K		
14	32	Private	205019	Assoc-voc	12	Never-married	Sales	Not-in-far	Black	Male	0	0	50	United-States	<=50K		
15	40	Private	121772	Assoc-voc	11	Married-civ-spouse	Craft-repair	Husband	Asian-Pac-Islander	Male	0	0	40	?	>50K		
16	34	Private	245487	7th-8th	4	Married-civ-spouse	Transportation	Husband	Amer-Indian-Alaska-Native	Male	0	0	45	Mexico	<=50K		
17	25	Self-emp-not-inc	176756	HS-grad	9	Never-married	Farming-fishing	Own-child	White	Male	0	0	35	United-States	<=50K		