**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

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





