Practical - 8

AIM : Implementation of K-Means clustering algorithms using the Mall_Customers.csv dataset. The dataset contains customer information with attributes including CustomerID, Genre, Age, Annual Income (k$), and Spending Score (1-100). Apply a clustering technique to segment customers based on their characteristics, compare the results, and derive insights into customer behavior patterns. Visualize the clusters and evaluate the effectiveness of each clustering method in identifying distinct customer groups within the mall's customer base.

Code : import numpy as np import matplotlib.pyplot as plt import pandas as pd
from sklearn.cluster import KMeans

Code :
dataset = pd.read_csv("/content/drive/MyDrive/DATASET/Mall_Customers.csv")
dataset.head() Output :

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Code :
dataset.shape Output :

```
(200, 5)
```

Code :
X = dataset.iloc[:,[3,4]].values
X Output :

```
array([[ 15,  39],
       [ 15,  81],
       [ 16,   6],
       [ 16,  77],
       [ 17,  40],
       [ 17,  76],
       [ 18,   6],
       [ 18,  94],
       [ 19,   3],
       [ 19,  72],
       [ 19,  14],
       [ 19,  99],
       [ 20,  15],
       [ 20,  77],
       [ 20,  13],
       [ 20,  79],
       [ 21,  35],
       [ 21,  66],
       [ 23,  29],
```

Code : 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)

Name: DHRUV SHERE
Enrollment No: 23012022021
Batch: 4IT-B-2

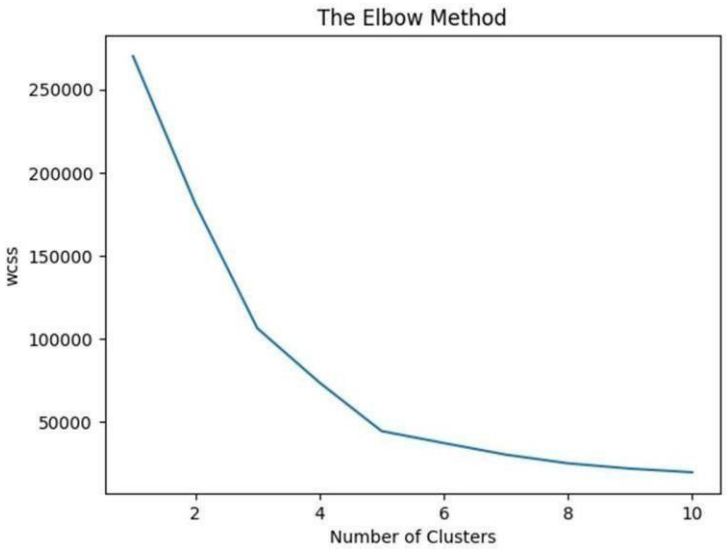
wcss.append(kmeans.inertia_) print(wcss)

Output :

```
[269981.28000000014]
[269981.28000000014, 181363.59595959607]
[269981.28000000014, 181363.59595959607, 106348.37306211119]
[269981.28000000014, 181363.59595959607, 106348.37306211119, 73679.78903948837]
[269981.28000000014, 181363.59595959607, 106348.37306211119, 73679.78903948837, 44448.45544793369]
[269981.28000000014, 181363.59595959607, 106348.37306211119, 73679.78903948837, 44448.45544793369, 37265.8
[269981.28000000014, 181363.59595959607, 106348.37306211119, 73679.78903948837, 44448.45544793369, 37265.8
```

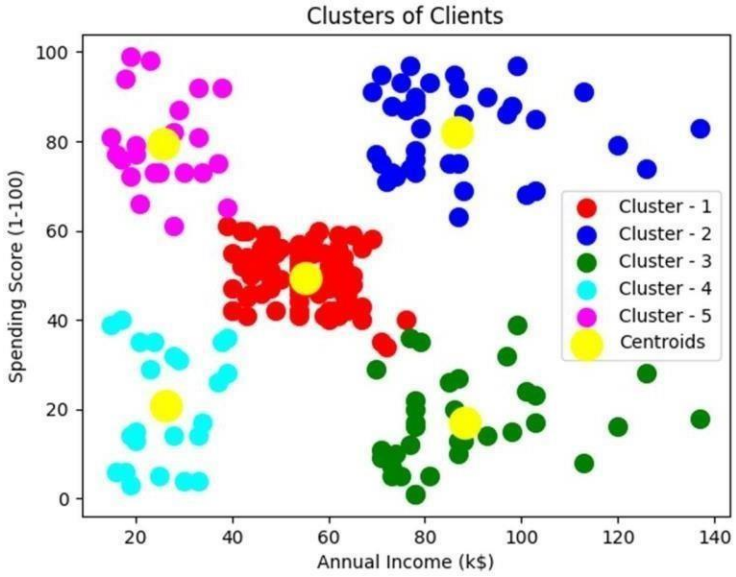
Code : plt.plot(range(1,11), wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of
Clusters') plt.ylabel('wcss') plt.show()

Output :



Code : 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 = 100, c = 'red', label = 'Cluster - 1')
plt.scatter(X[y_kmeans==1, 0], X[y_kmeans==1, 1], s = 100, c = 'blue', label = 'Cluster - 2')
plt.scatter(X[y_kmeans==2, 0], X[y_kmeans==2, 1], s = 100, c = 'green', label = 'Cluster - 3')
plt.scatter(X[y_kmeans==3, 0], X[y_kmeans==3, 1], s = 100, c = 'cyan', label = 'Cluster - 4')
plt.scatter(X[y_kmeans==4, 0], X[y_kmeans==4, 1], s = 100, c = 'magenta', label = 'Cluster - 5')
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 300, c='yellow', label =
'Centroids')
plt.title('Clusters of Clients')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.legend() plt.show() Output :



Name: DHRUV SHERE
Enrollment No: 23012022021
Batch: 4IT-B-2