Customer Segmentation using Data Science

Phase 4: Development Part 2

What is Customer Segmentation?

- Customer Segmentation is the process of division of customer base into several groups of individuals that share a similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits.
- The technique of customer segmentation is dependent on several key differentiators that divide customers into groups to be targeted. Data related to demographics, geography, economic status as well as behavioral patterns play a crucial role in determining the company direction towards addressing the various segments.

Dataset:

The dataset is acquired from kaggle and the link is given below:

https://www.kaggle.com/datasets/akram24/mall-customers

The dataset consists of following five features of 200 customers:

- Customer ID: Unique ID assigned to the customer
- Gender: Gender of the customer
- Age: Age of the customer
- Annual Income (k$): Annual Income of the customer
- Spending Score (1-100): Score assigned by the mall based on customer behavior and spending nature.

Training the KMeans Algorithm:

Optimum Number of Clusters = 5

Training the k-Means Clustering Model

```
[ ]  kmeans = KMeans(n_clusters=5, init='k-means++', random_state=0)

    # return a label for each data point based on their cluster
    Y = kmeans.fit_predict(X)

    print(Y)
```

```
[3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3
 1 3 1 3 1 3 0 3 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 2 4 2 0 2 4 2 4 2 0 2 4 2 4 2 4 2 4 2 0 2 4 2 4 2
 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4
 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2]
```

5 Clusters - 0, 1, 2, 3, 4

Visualizing the clusters:

Visualizing all the Clusters

```
# plotting all the clusters and their Centroids

plt.figure(figsize=(8,8))
plt.scatter(X[Y==0,0], X[Y==0,1], s=50, c='green', label='Cluster 1')
plt.scatter(X[Y==1,0], X[Y==1,1], s=50, c='red', label='Cluster 2')
plt.scatter(X[Y==2,0], X[Y==2,1], s=50, c='yellow', label='Cluster 3')
plt.scatter(X[Y==3,0], X[Y==3,1], s=50, c='violet', label='Cluster 4')
plt.scatter(X[Y==4,0], X[Y==4,1], s=50, c='blue', label='Cluster 5')

# plot the centroids
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], s=100, c='cyan', label='Centroids')

plt.title('Customer Groups')
plt.xlabel('Annual Income')
plt.ylabel('Spending Score')
plt.show()
```

Observations:

- The age of the customers ranges from 18-70. This shows that the mall attracts shops and things which suit all age group people.
- The average age of customers is 39.
- The average income of customers is 60 K$.
- The average spending score of customers is 50.

Conclusion:

Using k-means clustering we have managed to form different clusters based on different features. Mall management can target the clusters with average spending score to increase their profit and should also maintain good relationship with premium customers with high spending score. They should also work on coming up with new innovative ideas to upgrade the customers with low spending scores.