National University of Computer and Emerging Sciences

LAB K-MEAN CLUSTERING

You own a supermarket mall and through membership cards, you have some basic data about your customers like Customer ID, age, gender, annual income and spending score. You want to understand the customers like who are the target customers so that the sense can be given to the marketing team and plan the strategy accordingly. Solve the given problem by using the dataset given on the GCR.

1. What are the features used in this dataset for customer segmentation?
2. What is the distribution of the 'Age' feature in the dataset?
3. Which feature has the highest correlation with the 'Spending Score (1-100)' feature?
4. What is the optimal number of clusters for customer segmentation according to the Elbow Method?
5. What is the average annual income of customers in the dataset?
6. What is the average spending score of male customers in the dataset?
7. Which cluster has the highest average income and spending score?
8. What is the percentage of customers in Cluster 1?
9. What is the most frequent age group in Cluster 2?
10. What is the average income of customers in Cluster 3?
11. Show the count of value in each cluster.
12. Show customers from each cluster.
13. Make a visualization of the clusters.
14. Now Take the 3 features (["Age","Annual Income (k$)","Spending Score (1-100)"]) and repeat steps from 11 to 13.