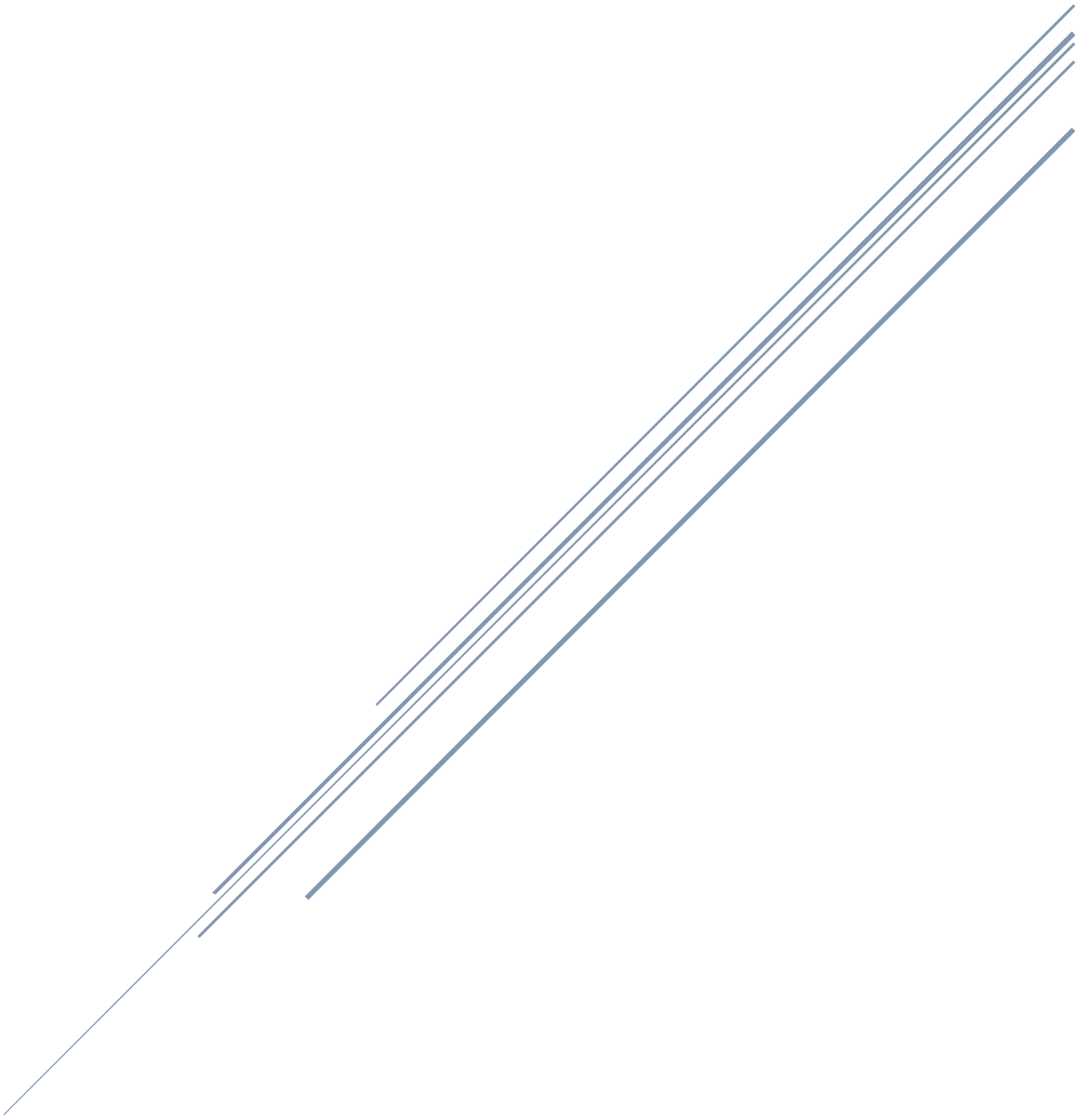
DATA MINING

CUSTOMER SEGMENTATION



1-The DataSet link :

https://www.kaggle.com/aryashah2k/customer-segmentation-hac-kmeans-rfm-matrix

2-DataSet Description and Target :

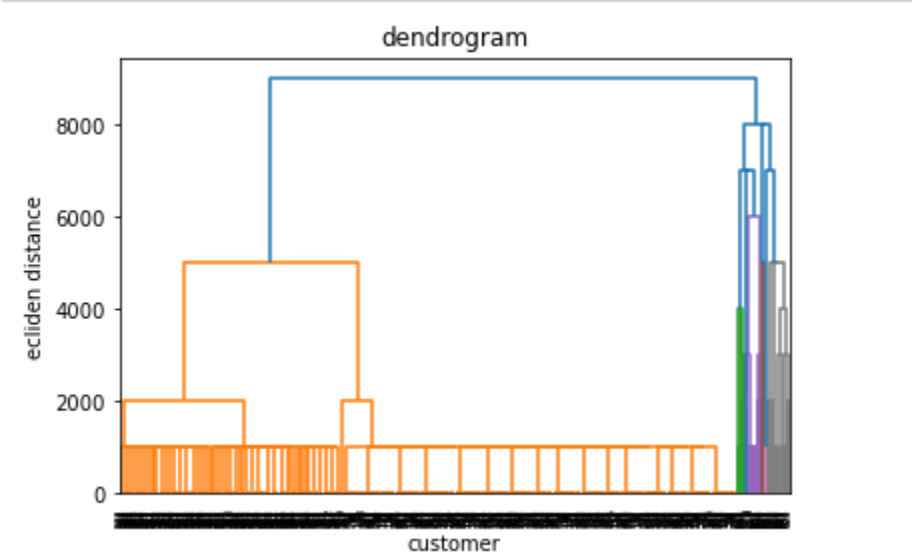
A Customer Credit Card Information Dataset which can be used for Identifying Loyal Customers, Customer Segmentation, Targeted Marketing and other such use cases in the Marketing Industry.

A few tasks that can be performed using this dataset is as follows:

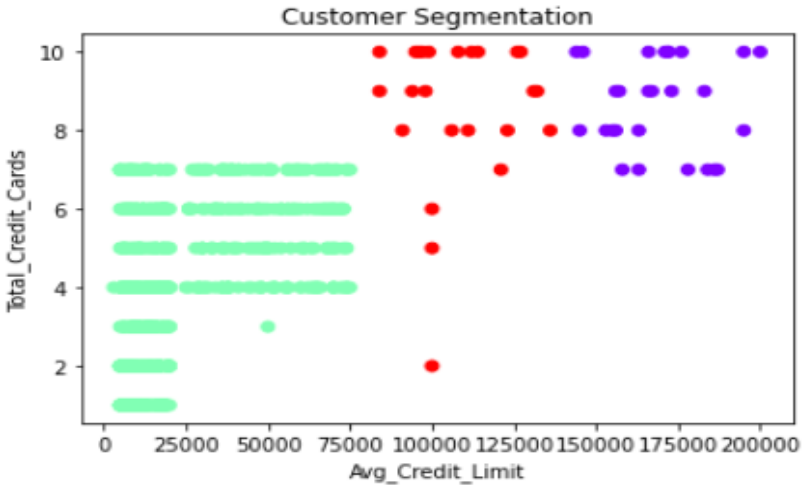
- Perform Data-Cleaning,Preprocessing,Visualizing and Feature Engineering on the Dataset.
- Implement Heirarchical Clustering, K-mediods Clustering models.

3-Plotted Graphs:

3-1 Hierarchical: Dendrogram:



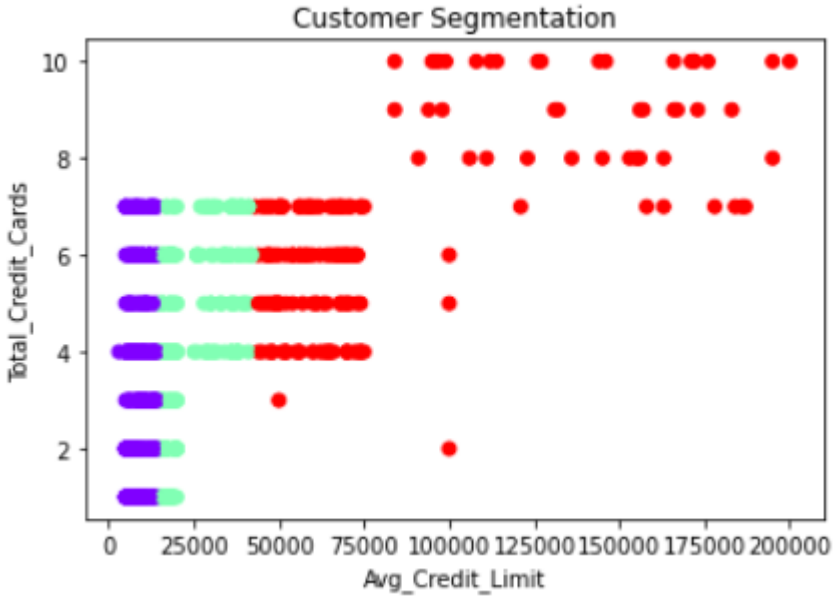
Scatter :



Customers with Avg credit limit >100000 And Customers with Avg credit limit < 75000 and cu Avg credit limit >130000

3-2 k-mediods:

Scatter:



The majority of the customers with Avg credit limit >50000 and Avg credit limit with<50000and the rest is <25000

5-Comparison between hierarchial clustering and k-mediods

Overall for this hierarchial dataset has given better results. The silhoutte score of
k-mediods is 0.09368221010034446
Hierarchical Cluster is 0.5255869852401159
respectively

Therefore hierarchial has higher silhoutte score.

Since This measure has a range of [-1, 1].

Silhouette coefficients (as these values are referred to as) near +1 indicate that the sample is far away from the neighboring clusters. A value of 0 indicates that the sample is on or very close to the decision boundary between two neighboring clusters and negative values indicate that those samples might have been assigned to the wrong cluster.