

Insights from Naman Mishra Clustering.ipynb file

1) Optimal Number of Clusters:

- Using the Elbow method and Silhouette scores, the optimal number of clusters was decided to be 4.

2) Cluster Metrics:

- DB Index (Davies-Bouldin Index):

DB Index: 0.5160644633752305

The calculated DB Index is low, that means here well-separated clusters.

- Silhouette Score: 0.8326502307852304

A silhouette score between 0.5 and 1 showing that the clusters are compact and well-separated.

3) Customer Segmentation Features:

- Clustering done while utilizing RFM metrics (NormRecency, LogFrequency, LogMonetary).
- Scaling and Dimensionality reduction for improved clustering.

4) Cluster Characteristics:

- Customers from different clusters showing variations in spending patterns from different regions.

- These insights could help to prepare marketing strategies for different segments.

5) Visualization:

- Principal Component Analysis (PCA) was used to perform the reducing of the dimensionality of the data to 2D for visualization.
- The clusters are visually distinct in the scatterplot, confirming their separation.

