

Clustering Report: Customer Segmentation

1. Clustering Logic

The clustering process used a combination of demographic (profile) and transactional data to generate meaningful customer segments:

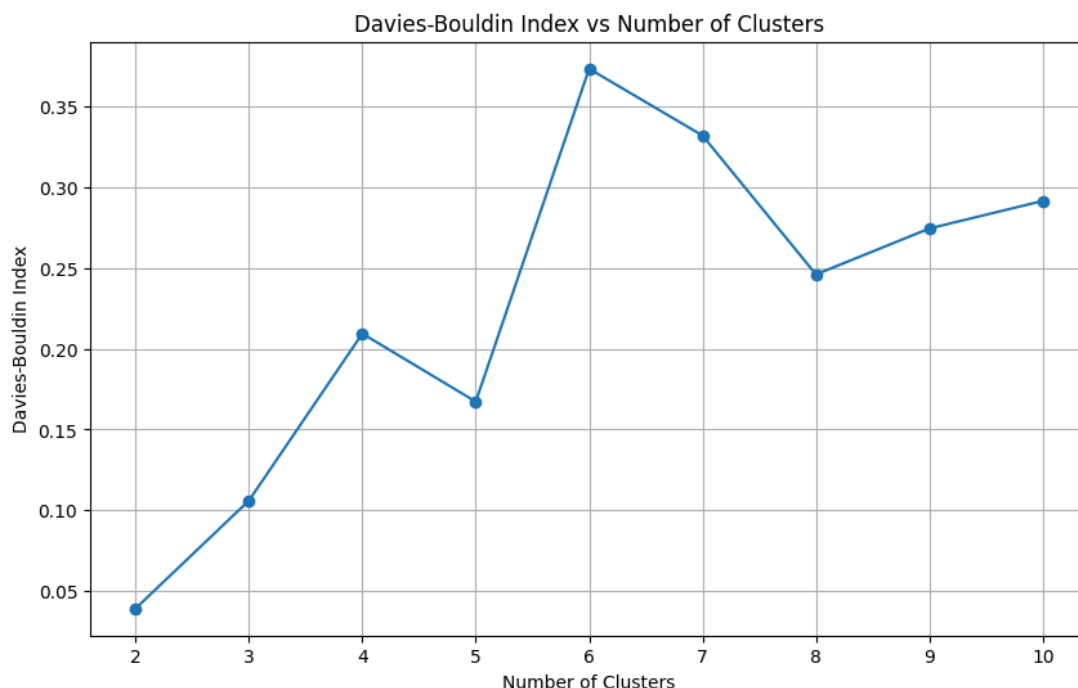
1. **Feature Engineering:**
 - a. One-Hot encoded 'Region'.
 - b. Normalized all features using mean-centering to ensure comparability.
2. **Similarity Calculation:**
 - a. Used cosine similarity to assess the relationships between customers.
3. **Clustering Algorithm:**
 - a. Opted for **hierarchical clustering** due to its simplicity and interpretability.
 - b. Explored cluster sizes between 2 and 10, with 2 clusters yielding the best DB Index.

2. Number of Clusters Formed

- Using the DB Index and clustering evaluation, the optimal number of clusters determined was **2**.

3. DB Index

- **Minimum DB Index:** 0.0391
A lower DB Index indicates better-defined clusters, signifying that the clustering approach effectively groups customers with high intra-cluster similarity and low inter-cluster similarity.



4. Clustering Metrics

- **Intra-Cluster Average Similarities (Mean-Centered Cosine):**

- Cluster 0: **0.9560**

- Cluster 1: **0.9562**

High intra-cluster average similarity values signify strong similarity within each cluster.

- **Intra-Cluster Minimum Similarities (Mean-Centered Cosine):**

- Cluster 0: **-0.0381**

- Cluster 1: **-0.0707**

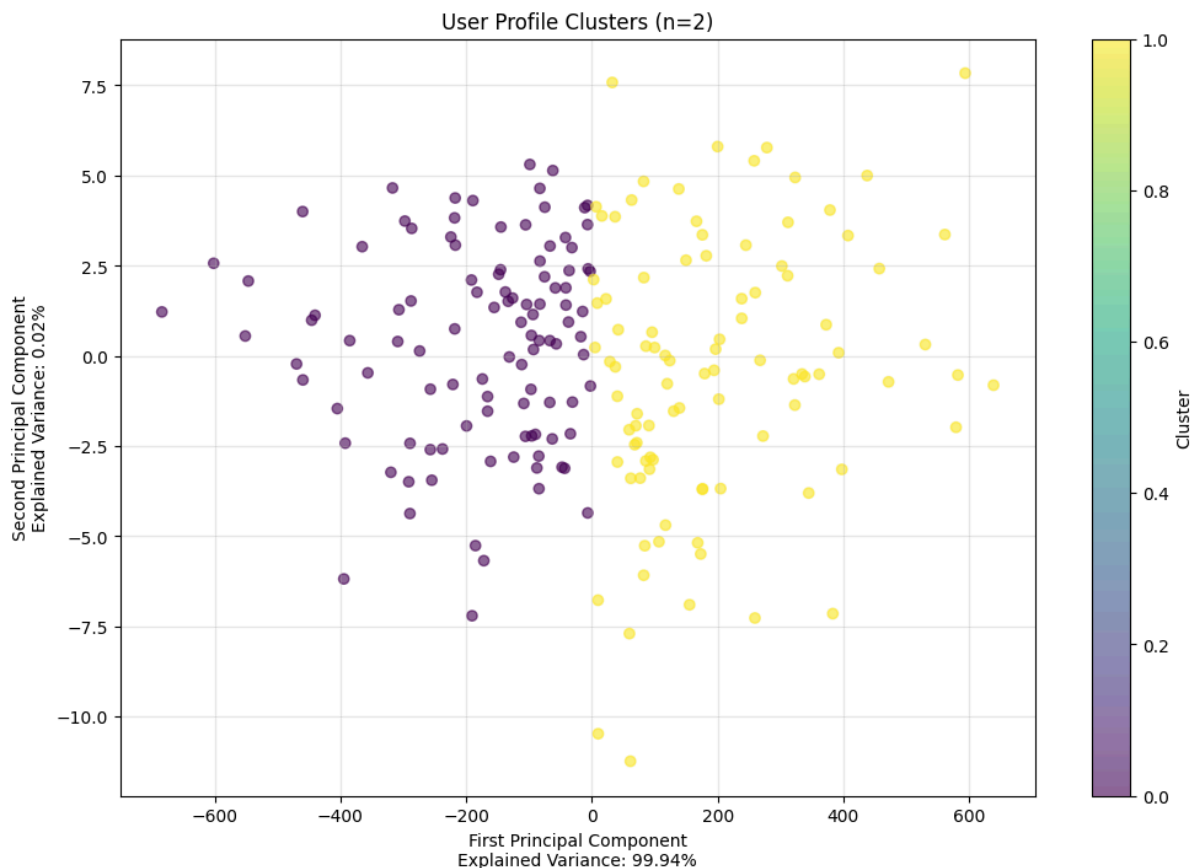
Minimum similarity values reflect the lowest pairwise similarity in the cluster.

While these are slightly negative, the high average similarities indicate robustness in clustering.

5. Visualization of Clusters

1. Cluster Scatter Plot:

Plotted two principal components (PCA) derived from customer profiles and transaction data. The clusters were visually distinct, validating the segmentation.



6. Key Insights

- **Cluster 0:** Customers with high similarity in purchase behavior and average transaction values, likely representing frequent or loyal buyers.
- **Cluster 1:** Customers with slightly lower minimum similarity, possibly reflecting more diverse purchase patterns or regional segmentation.

