

Problem solving with unsupervised learning: Customer Segmentation using Hierarchical Clustering and DBSCAN

Group 2: David Solero Chicano, Fadi Alkhori

16/12/2024

1. Problem Formulation

The goal of this project is to segment customers into meaningful groups using unsupervised machine learning techniques:

Hierarchical Clustering and DBSCAN.

These segments will help provide tailored financial recommendations, such as:

- **Savings plans:** For customers with consistent full payments and high purchase values.
- **Loans:** For customers exhibiting high cash advances.
- **Wealth Management:** For premium customers with high financial activity.
- **Rewards Programs:** For customers with frequent purchases.

2. Data Preparation

Dataset Description

The dataset contains behavioral information for 9000 customers, consisting of 18 numeric features. The column CUST_ID was dropped as it is irrelevant to clustering.

Data Cleaning

- Missing values were imputed using the mean for numeric variables.

3. Model Implementation

Hierarchical Clustering

- **Linkage Method:** Complete linkage was used to compute distances between clusters.

- **Number of Clusters:** The number of clusters was set to $k = 3$ based on dendrogram analysis.
- **Evaluation:** Silhouette Score = **0.85**, indicating strong clustering quality.

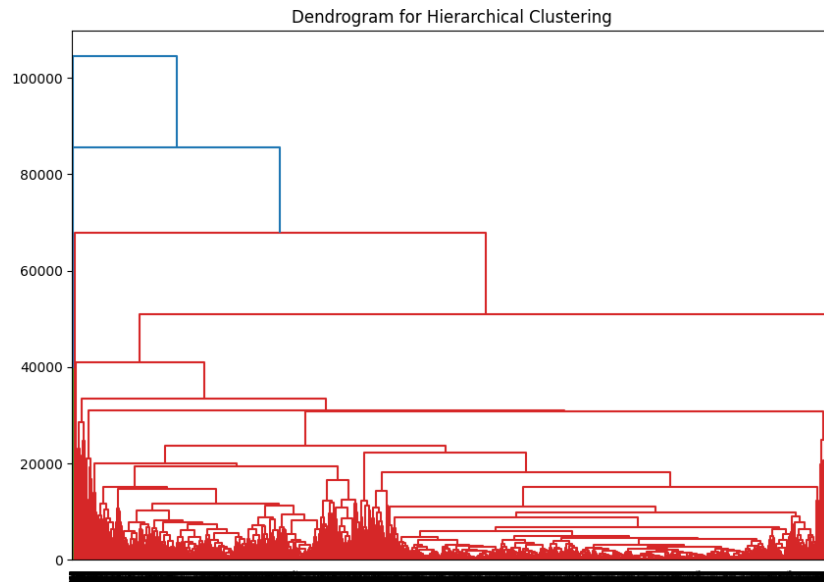


Figure 1: Dendrogram for Hierarchical Clustering

DBSCAN

- **Parameters:** $\text{eps}=5$ and $\text{min_samples}=10$.
- **Results:** DBSCAN identified clusters along with several outliers.
- Silhouette Score indicates moderate clustering due to sparse density.

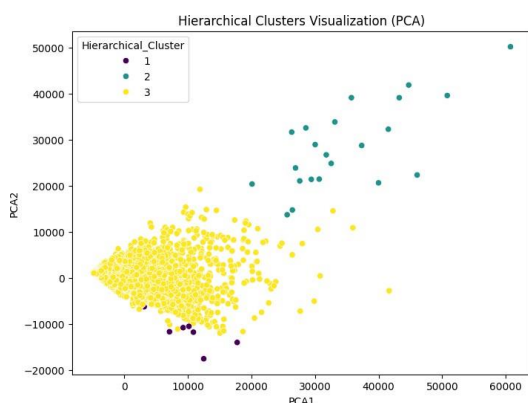


Figure 2: PCA Visualization

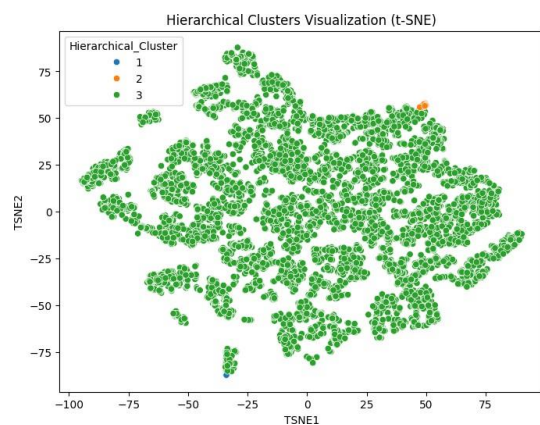


Figure 3: T-SNE Visualization

4. Evaluation and Results

Silhouette Score was used to evaluate clustering quality:

- Hierarchical Clustering: Silhouette Score = **0.85** (strong clustering quality).
- DBSCAN: Moderate Silhouette Score with significant noise detection.

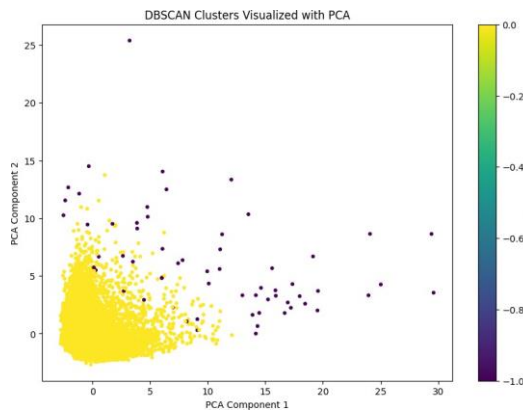


Figure 4: PCA DBSCAN

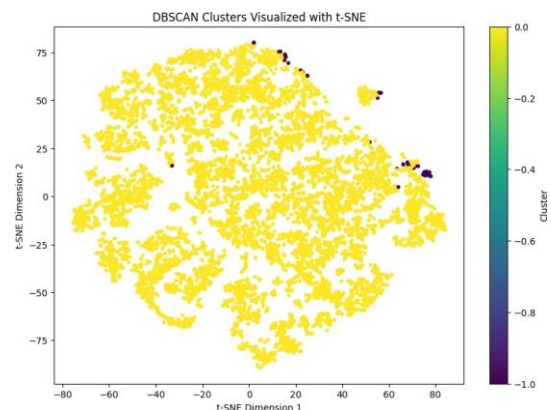


Figure 5: T-SNA DBSCAN

Cluster Insights and Recommendations

Cluster	Behavior	Recommendation
Cluster 1	High CASH_ADVANCE	Offer short-term loans.
Cluster 2	High PRC_FULL_PAYMENT (> 80%)	Suggest wealth management services.
Cluster 3	High PURCHASES_FREQUENCY (> 0.5)	Introduce cashback or rewards programs.
Outliers	Sparse purchase behavior	Engage with personalized offers.

Table 1: Cluster Descriptions and Recommendations

5. Conclusions and Future Work

The analysis demonstrated that **Hierarchical Clustering** effectively grouped customers into interpretable segments, achieving a high Silhouette Score of **0.85**. **DBSCAN** identified outliers and provided insights into sparse customer behaviors but showed sensitivity to hyperparameters.

Recommendations were provided based on cluster behaviors, such as loans for high cash advance users and rewards programs for frequent shoppers. Future work includes automating hyperparameter selection for DBSCAN, exploring advanced clustering techniques such as Gaussian Mixture Models (GMM) for probabilistic clustering, and incorporating domain-specific KPIs for more refined segmentation.