

Zeotap Assignment Task – 3

Customer Segmentation Report

Objective

The goal of this task was to perform customer segmentation using clustering techniques based on customer profile information (from Customers.csv) and transaction details (from Transactions.csv). The evaluation was based on the Davies-Bouldin Index (DBI), clustering logic, and visual representation.

Methodology

1. Data Preprocessing:
 - Merged customer profile and transaction datasets.
 - One-hot encoded categorical variables (e.g., Region).
 - Scaled all features using MinMaxScaler for normalization.
2. Clustering Algorithm:
 - Used the K-Means Clustering algorithm for segmentation.
 - Determined the optimal number of clusters (k) using the Davies-Bouldin Index (DBI) and Elbow Method.
 - Chose 4 clusters as the optimal value for k.
3. Clustering Metrics:
 - Calculated the Davies-Bouldin Index (DBI) to evaluate cluster compactness and separation.

- Computed the Silhouette Score to measure how well samples were assigned to clusters.

4. Cluster Analysis:

- Aggregated and analyzed numeric features for each cluster to identify distinct customer groups.

Results

1. Clustering Metrics:

- **Number of Clusters: 4**
- **Davies-Bouldin Index (DBI): 0.431**
 - **Lower DBI indicates well-defined clusters.**
- **Silhouette Score: 0.727**
 - **High score suggests good cluster assignments.**

2. Cluster Characteristics:

- Key numeric features (e.g., Total Value, Transaction Count) were analysed for each cluster:
 - Cluster 0: High-value customers with frequent transactions.
 - Cluster 1: Low-value customers with sporadic transactions.
 - Cluster 2: Moderate-value customers, balanced transaction frequency.
 - Cluster 3: Regional clusters with distinct product preferences.

3. Visualization:

- Visualized clusters using:

- Scatter plots: Displayed clusters using PCA for dimensionality reduction.
- Bar charts: Highlighted mean feature values for each cluster.

Conclusion

The clustering successfully segmented customers into 4 groups, each with distinct behaviours and profiles. This segmentation can be used to:

- Personalize marketing strategies.
- Target high-value customers.
- Identify and retain at-risk customer groups.

Recommendations:

- Use the segmentation to offer cluster-specific promotions or loyalty programs.
- Explore other clustering algorithms (e.g., DBSCAN or Agglomerative) for comparison.
- Regularly update clustering models with new data for accuracy.