Harpartap Singh - Clustering Report

Customer Segmentation Report

1. Overview:

Customer segmentation is crucial for understanding customer behavior patterns and improving business strategy. By analyzing purchase history and demographic data, we aimed to identify distinct groups of customers with similar characteristics and spending habits.

2. Data Preparation:

The segmentation analysis was conducted using the merged customer and transaction data. Key features considered for clustering include total spending, quantity of purchases, and geographical region.

To enhance the quality of clustering, the following preprocessing steps were applied:

- Data cleaning to remove inconsistencies and missing values.

- Feature scaling using StandardScaler to normalize numerical attributes.

- One-hot encoding of categorical variables to ensure compatibility with clustering algorithms.

3. Optimal Number of Clusters Determination:

The optimal number of clusters was determined using the Elbow Method and Davies-Bouldin Index (DB Index).

- Elbow Method:

The sum of squared distances from each point to its assigned cluster center was plotted against different values of k. The optimal k value is identified where the curve starts to flatten, indicating diminishing returns.

- Davies-Bouldin Index Calculation:

The DB Index evaluates the quality of clustering by measuring intra-cluster dispersion and inter-cluster separation. A lower DB Index suggests better clustering. The steps involved are:

1. Compute the average distance between each point in a cluster and the cluster centroid (intra-cluster distance).

2. Calculate the distance between centroids of different clusters (inter-cluster distance).

3. Compute the DB Index as the average of the maximum ratio of intra-cluster to inter-cluster distances across all clusters.

The final calculated DB Index for the optimal cluster configuration (k=8) is: 0.632

4. Clustering Insights:

- High-Value Customers: A cluster was identified comprising customers with high total purchase value, suggesting a focus on premium product offerings.
- Budget-Conscious Customers: Another segment showed a preference for lower-cost products, providing opportunities for discount-focused marketing strategies.
- Regional Preferences: Distinct clusters emerged based on region, highlighting potential for localized marketing strategies.

5. Recommendations:

Based on the clustering results, the following strategic recommendations are proposed:

- Develop targeted marketing campaigns to cater to high-value and budget-conscious customer segments.
- Customize product offerings based on regional preferences and trends observed from cluster analysis.
- Introduce loyalty programs for high-value customers to enhance retention and increase lifetime value.

6. Conclusion:

Customer segmentation has provided actionable insights into different purchasing behaviors and regional preferences. These insights can guide data-driven business decisions to optimize revenue and customer engagement.