Clustering Results of Customer Dataset

Summary

This report summarizes the clustering results of a customer dataset based on their transaction data. The goal was to group customers into clusters with similar transactional behaviors to derive actionable insights.

Key Metrics

- Number of clusters: 4
- Davies-Bouldin Index: 1.4370 (lower values indicate better-defined clusters)
- Silhouette Score: 0.2221 (indicating moderate separation among clusters)

Cluster Sizes

The number of customers in each cluster is as follows:
- Cluster 0: 57 customers
- Cluster 1: 77 customers
- Cluster 2: 23 customers
- Cluster 3: 42 customers

Cluster Profiles

The following normalized values highlight the characteristics of each cluster:
- **Cluster 0:** High transaction count, high total value, moderate quantity, and high price variation.
- **Cluster 1:** Low transaction count, moderate total value, lower quantity, and minimal price variation.
- **Cluster 2:** Very low transaction count, very low total value, low quantity, and significant price variation.
- **Cluster 3:** Low transaction count, moderate total value, high price mean, and moderate recency.

Visualizations

The analysis included the following visualizations:
1. **Elbow Method:** Used to determine the optimal number of clusters based on inertia.
2. **Davies-Bouldin Score:** Showcased the cluster compactness for different k values.
3. **Silhouette Score:** Highlighted the separation quality for each k.
4. **Customer Segments Visualization:** Illustrated the clusters in a 2D space using PCA.

Conclusion

The clustering analysis identified 4 distinct customer groups with varying transaction characteristics. These insights can be used for targeted marketing, personalized recommendations, and inventory planning. The moderate silhouette score indicates potential overlap in some clusters, warranting further refinement.

Elbow Method

Davies-Bouldin Score

Silhouette Score

Customer Segments Visualization