Detailed Clustering Report

Overview

Customer segmentation was performed using clustering techniques on customer profiles and transaction data. The primary evaluation metric was the **Davies-Bouldin Index (DB Index)**, with the **Silhouette Score** providing additional insights into the quality of the clustering. The results across different cluster configurations (2 to 10 clusters) are presented below.

Clustering Results

The clustering results are summarized in the following table:

Number of Clusters	DB Index	Silhouette Score
2	1.6994	0.2094
3	1.5014	0.2290
4	1.3767	0.2667
5	1.3756	0.2553
6	1.3357	0.2518
7	1.4268	0.2380
8	1.3147	0.2671
9	1.3408	0.2629
10	1.3017	0.2613

Key Insights

1. Number of Clusters Formed

The analysis explored clustering solutions ranging from **2 to 10 clusters**. The **optimal number of clusters** was determined to be **10**, as it resulted in the lowest DB Index value of **1.3017**, indicating the best balance of cluster compactness and separation.

2. Davies-Bouldin Index (DB Index)

The DB Index steadily decreased as the number of clusters increased, reaching a minimum at 10 clusters:
- Interpretation: Lower DB Index values indicate that clusters are compact and well-separated.
- The consistent reduction in DB Index suggests that increasing the number of clusters improves clustering quality, though marginal gains diminish after a certain point.

3. Silhouette Score

The Silhouette Score peaked at **0.2671 (8 clusters)** and slightly decreased with 9 and 10 clusters:
- Interpretation: A higher score indicates better-defined clusters.
- While the score decreased slightly with 10 clusters, the reduction was minimal, making the tradeoff for a lower DB Index acceptable.

4. Cluster Stability and Trends

- **Clusters 2-4**: Initial clusters show poorer compactness (higher DB Index) and weak separation (lower Silhouette Scores).
- **Clusters 5-7**: Clusters become more distinct, with the DB Index decreasing significantly and Silhouette Scores stabilizing.
- **Clusters 8-10**: Further refinement improves compactness (DB Index), though the Silhouette Score slightly fluctuates.

Visualization Insights

Scatterplot (PCA-Reduced Visualization)

The scatterplot shows clusters visualized in 2D space using PCA (Principal Component Analysis):

- **Distinct Clusters**: Several clusters are well-separated, particularly those in the top and bottom regions of the plot.
- **Overlap**: A few clusters exhibit mild overlap, consistent with the moderate Silhouette Scores.
- **Cluster Sizes**: The visualization highlights varying cluster sizes, with smaller clusters potentially representing niche groups and larger ones dominating the customer base.

Recommendations

1. Optimal Number of Clusters

Based on the results, **10 clusters** provide the best segmentation, achieving a balance between compactness and separation.

2. Next Steps

- **Customer Profiling**: Use the clusters to analyze customer behavior and identify actionable insights (e.g., top-spending clusters, regional trends).
- **Business Validation**: Validate the clusters against business objectives to ensure they align with key strategies.
- **Feature Engineering**: Consider incorporating additional features (e.g., customer lifetime value or seasonal patterns) to refine clustering quality further.

3. Applications

- **Marketing Campaigns**: Target specific clusters with tailored campaigns based on their profiles and behaviors.
- **Product Recommendations**: Use segmentation to recommend products that align with cluster preferences.
- **Resource Allocation**: Focus resources on high-value clusters for maximum ROI.

Conclusion

The clustering analysis revealed **10 well-defined customer segments**, with a **Davies-Bouldin Index of 1.3017** and a moderate **Silhouette Score of 0.2613**. These clusters can be leveraged to gain deeper insights into customer behavior, enabling data-driven decision-making for marketing, sales, and customer retention strategies.