Customer Segmentation Analysis Report

Key Metrics

- Number of clusters formed: 10
- Davies-Bouldin Index: 1.1727
- Silhouette Score: 0.4347

Clustering Quality Assessment

Validation Metrics Analysis

1. Davies-Bouldin Index (DB):

- Value of 1.1727 indicates relatively good cluster separation
- The DB Index plot shows a clear elbow at 4 clusters, with diminishing improvements afterward
- Lower DB Index values indicate better clustering, and our score suggests moderately well-defined clusters

2. Silhouette Score:

- Score of 0.4347 indicates moderate cluster cohesion and separation
- The score has been improving with increasing cluster numbers, suggesting that the higher number of clusters better captures the data's natural groupings
- The silhouette plot shows steady improvement from 2 to 10 clusters, with the steepest improvements between 4-8 clusters

Cluster Characteristics

Based on the heatmap, several distinct customer segments emerge:

1. High-Value Cluster (Column 4):

- Highest total spends (5794.05)
- Highest purchase timespan (280.36 days)
- Largest total quantity (21.13)
- Above-average transaction value (751.73)

2. New Customer Cluster (Column 9):

- Youngest customer age (261.45 days)
- Predominantly South American customers
- Lower than average transaction values
- Moderate purchase frequency

3. Frequent Purchaser Cluster (Column 3):

- Highest purchase frequency (1.00)
- Lower total spends (528.41)
- No variation in transaction values (std_transaction_value = 0)
- Shortest purchase timespan (0.00)

4. Regional Segments:

- Asia-focused cluster (Column 6)
- Europe-focused cluster (Column 5)
- North America-focused cluster (Column 7)
- Each showing distinct purchasing patterns

Clustering Performance Analysis

The clustering solution shows several strengths:

1. Good separation between clusters as indicated by the DB Index
2. Improving silhouette score with increased cluster numbers
3. Clear differentiation in customer behaviors across clusters
4. Meaningful regional segmentation

Areas for consideration:

1. Some clusters show similar patterns, suggesting possible over-segmentation
2. Moderate silhouette score indicates some overlap between clusters
3. High number of clusters (10) may make practical implementation more complex

Conclusion

The 10-cluster solution provides a detailed segmentation of the customer base, balancing granularity with cluster quality. While the validation metrics suggest acceptable clustering quality, business users should consider whether a simpler solution with fewer clusters might be more actionable for practical purposes.