Customer Segmentation Analysis Report

1. Clustering Overview

The analysis resulted in the identification of 8 distinct customer segments using K-means clustering algorithm. This segmentation was performed using a combination of RFM (Recency, Frequency, Monetary) metrics and geographical information.

2. Clustering Metrics

- Davies-Bouldin Index: 1.1093
    - This relatively low DB index indicates good cluster separation
    - The score suggests that the clusters are well-defined and distinct from each other.
    - Lower DB index values indicate better clustering (range typically 0 to 2+)
- Additional Validation Metrics: Silhouette Score: 0.3227 -
    - Indicates moderate cluster cohesion and separation
    - Range is from -1 to 1, with higher values indicating better-defined clusters
- Calinski-Harabasz Score: 46.2685
    - Suggests good cluster density and separation.
    - Higher values indicate better-defined clusters

3. Cluster Characteristics

Cluster Size Distribution:

a) Largest clusters: Clusters 2 and 3 (19% and 19.5% of customers)
b) Smallest clusters: Cluster 0 and 5 (5% and 6% of customers)
c) Moderate-sized clusters: Clusters 1, 4, 6, and 7 (ranging from 9% to 17.5%)

Cluster Profiles:

1) High-Value Active Customers (Cluster 0)

a) Highest average monetary value ($6,550.72)

b) Highest frequency (9.2 transactions)
c) Most recent customers (23.6 days)
d) Predominantly North American

2) Mid-Value South American Segment (Cluster 1)
   a) Moderate monetary value ($2,510.84)
   b) Average frequency (4.17 transactions)
   c) Moderate recency (69.4 days)

3) North American Core Segment (Cluster 2)
   a) Moderate monetary value ($2,992.20)
   b) Average frequency (4.84 transactions)
   c) Less recent activity (85.1 days)

4) European Mid-Market (Cluster 3)
   a) Above-average monetary value ($3,383.85)
   b) Average frequency (4.59 transactions)
   c) Moderate recency (73.9 days)

5) Asian High-Value Segment (Cluster 4)
   a) High monetary value ($5,379.35)
   b) High frequency (7.5 transactions)
   c) Good recency (58.4 days)

6) Low-Value European Segment (Cluster 5)
   a) Lowest monetary value ($691.48)
   b) Lowest frequency (2.17 transactions)
   c) Least recent activity (150.25 days)

7) Asian Value Segment (Cluster 6)
   a) Lower monetary value ($2,263.99)
   b) Low frequency (3.33 transactions)
   c) Moderate recency (74 days)

8) South American High-Value (Cluster 7)
   a) High monetary value ($5,478.05)
   b) High frequency (6.58 transactions)
   c) Good recency (53.75 days)

4. Key Findings

1) Geographic Patterns:
   a) Clear regional differentiation in customer behaviour
   b) Each major region has both high-value and low-value segments

2) Value Segmentation:
   a) Three distinct value tiers identified:
   b) High-value (Clusters 0, 4, 7)
   c) Mid-value (Clusters 1, 2, 3)
   d) Low-value (Clusters 5, 6)

3) Engagement Levels:
   a) Strong correlation between frequency and monetary value
   b) Recency patterns vary significantly across clusters

5. Validation of Clustering Quality

The combination of metrics (DB Index, Silhouette Score, and Calinski-Harabasz Score) suggests that the 8-cluster solution provides a robust and meaningful segmentation of the customer base. The relatively low Davies-Bouldin Index (1.1093) particularly indicates well-separated clusters, making this segmentation suitable for targeted marketing strategies and customer management.