# Task 3: Customer Segmentation / Clustering

**Clustering Analysis Report**

**1. Introduction** Clustering is a critical unsupervised machine learning technique used to group data into clusters based on their similarities. This report details the clustering results obtained from the analysis of customer transaction data. The analysis focuses on identifying clusters based on transactional and demographic attributes and evaluating the clustering model's performance using metrics such as the Davies-Bouldin Index (DB Index).

---

**2. Clustering Summary**

- **Number of Clusters Formed:** 5
- **Final Davies-Bouldin Index (DB Index):** 1.2435

The DB Index is a measure of cluster compactness and separation, with lower values indicating better clustering quality. A DB Index of 1.2435 suggests that the clusters are reasonably well-separated and compact.

---

**3. Cluster Characteristics** Below are the characteristics of the five clusters identified:

**Cluster 0:**

- **Total Spend:** 774.84
- **Average Transaction Value:** 425.34
- **Standard Deviation of Transaction Value:** 84.91
- **Total Quantity Purchased:** 3.84
- **Average Quantity Per Transaction:** 1.99
- **Transaction Count:** 2.00
- **Days Since Last Purchase:** 147.88
- **Region:** Predominantly South America (32%) and Asia (28%).

**Cluster 1:**

- **Total Spend:** 3927.73
- **Average Transaction Value:** 724.38
- **Standard Deviation of Transaction Value:** 496.77
- **Total Quantity Purchased:** 14.43
- **Average Quantity Per Transaction:** 2.60
- **Transaction Count:** 5.57

- **Days Since Last Purchase:** 53.92
- **Region:** Exclusively Asia (100%).

**Cluster 2:**

- **Total Spend:** 3617.58
- **Average Transaction Value:** 747.41
- **Standard Deviation of Transaction Value:** 500.81
- **Total Quantity Purchased:** 12.76
- **Average Quantity Per Transaction:** 2.61
- **Transaction Count:** 4.89
- **Days Since Last Purchase:** 69.67
- **Region:** Exclusively Europe (100%).

**Cluster 3:**

- **Total Spend:** 4172.61
- **Average Transaction Value:** 751.99
- **Standard Deviation of Transaction Value:** 518.76
- **Total Quantity Purchased:** 14.71
- **Average Quantity Per Transaction:** 2.64
- **Transaction Count:** 5.59
- **Days Since Last Purchase:** 57.20
- **Region:** Exclusively South America (100%).

**Cluster 4:**

- **Total Spend:** 3564.39
- **Average Transaction Value:** 652.57
- **Standard Deviation of Transaction Value:** 432.63
- **Total Quantity Purchased:** 13.88
- **Average Quantity Per Transaction:** 2.51
- **Transaction Count:** 5.69
- **Days Since Last Purchase:** 70.48
- **Region:** Exclusively North America (100%).

---

**4. Key Observations**

1. **Spending Patterns:**

   - Cluster 0 represents low spenders with infrequent transactions and the longest days since the last purchase.
   - Clusters 1 through 4 represent high spenders with varying preferences in transaction count and recency.

2. **Regional Distribution:**

   - Each of Clusters 1, 2, 3, and 4 corresponds to a specific region (Asia, Europe, South America, and North America, respectively).
   - Cluster 0 is more diverse, encompassing customers from multiple regions, primarily South America and Asia.

3. **Transaction Frequency:**

   - Clusters 1, 2, 3, and 4 exhibit high transaction frequencies, suggesting engagement from customers in these regions.
   - Cluster 0, with only two transactions on average, indicates disengaged or low-priority customers.

4. **Recency:**

   - Cluster 0 has customers with the least recent activity (147.88 days since the last purchase), indicating potential churn risk.
   - Customers in Clusters 1 and 3 have the most recent activity, indicating high engagement.

---

## 5. Recommendations

1. **Targeted Marketing Campaigns:**

   - Focus on re-engaging Cluster 0 customers through personalized offers and incentives.
   - Reward loyal and frequent shoppers in Clusters 1 through 4 to maintain their engagement.

2. **Regional Strategies:**

   - Develop region-specific strategies leveraging the unique characteristics of Clusters 1 (Asia), 2 (Europe), 3 (South America), and 4 (North America).

3. **Customer Segmentation:**

   - Use the cluster insights for tailored product recommendations, promotions, and loyalty programs.

---

**6. Conclusion** The clustering analysis provided valuable insights into customer behavior and regional preferences. The DB Index of 1.2435 indicates a satisfactory clustering outcome. By leveraging these results, the organization can enhance customer engagement, improve marketing strategies, and optimize resource allocation for targeted campaigns.