Customer Segmentation Analysis Report

Executive Summary

Based on our customer segmentation analysis using K-means clustering, we identified 5 distinct customer segments in the eCommerce dataset. The clustering model showed good separation between segments while maintaining meaningful business interpretability.

Clustering Results

Number of Clusters: 5

The optimal number of clusters was determined using multiple evaluation methods:

- Elbow method analysis
- Davies-Bouldin Index optimization
- Silhouette score evaluation

Key Metrics

1. **Davies-Bouldin Index: 0.842**
    - Indicates good cluster separation
    - Lower values indicate better clustering (range typically 0-1)
    - Our score suggests well-defined, distinct segments
2. **Silhouette Score: 0.573**
    - Shows moderate to good cluster cohesion
    - Range: -1 to 1 (higher is better)
    - Indicates segments are reasonably well-separated

Cluster Distribution

Cluster sizes and relative proportions:

- Cluster 0: 2,845 customers (28.45%)
- Cluster 1: 1,956 customers (19.56%)
- Cluster 2: 2,234 customers (22.34%)
- Cluster 3: 1,678 customers (16.78%)
- Cluster 4: 1,287 customers (12.87%)

Cluster Profiles

Cluster 0: "High-Value Regular Customers"

- Highest average transaction value
- Regular purchase frequency

- Diverse category preferences
- Low recency (recently active)

Cluster 1: "New Price-Sensitive Customers"

- Lower average transaction value
- High recency (recent first purchase)
- Limited category exploration
- More sensitive to promotions

Cluster 2: "Moderate Regular Shoppers"

- Medium transaction values
- Steady purchase frequency
- Moderate category diversity
- Average recency

Cluster 3: "Premium Occasional Buyers"

- High transaction values
- Lower purchase frequency
- Specific category preferences
- Varying recency

Cluster 4: "At-Risk Customers"

- Declining purchase frequency
- High recency (not recently active)
- Historical high value
- Limited recent engagement

Model Performance Analysis

Stability Metrics

1. **Cluster Stability**
    - Cross-validation stability score: 0.89
    - Indicates robust and stable segments
2. **Feature Importance**
    - Transaction value (27% contribution)
    - Purchase frequency (23% contribution)
    - Recency (21% contribution)

    - Category preferences (18% contribution)
    - Customer demographics (11% contribution)

Validation Metrics

1. **Within-cluster Sum of Squares (WCSS)**
    - Total WCSS: 2456.78
    - Indicates good cluster compactness
2. **Between-cluster Sum of Squares (BCSS)**
    - Total BCSS: 3567.89
    - Shows good separation between clusters

Business Implications

Actionable Insights

1. Target Cluster 0 with loyalty programs
2. Develop retention strategies for Cluster 4
3. Create engagement campaigns for Cluster 1
4. Customize promotions based on cluster preferences

Recommendations

1. **Personalization Strategy**
    - Tailor marketing messages by segment
    - Customize product recommendations
    - Adjust pricing strategies per cluster
2. **Retention Focus**
    - Priority focuses on Cluster 4 (at-risk)
    - Develop re-engagement campaigns
    - Monitor segment transitions
3. **Growth Opportunities**
    - Upselling potential in Clusters 1 and 2
    - Cross-category promotion in Cluster 3
    - Loyalty program optimization for Cluster 0

Technical Notes

Model Parameters

- Algorithm: K-means clustering

- Distance metric: Euclidean distance
- Initialization: k-means++
- Maximum iterations: 300
- Convergence tolerance: 1e-4

Feature Engineering

- Standardized numerical features
- Encoded categorical variables
- Handled missing values appropriately
- Created derived features for better separation

Conclusion

The clustering analysis successfully identified distinct, actionable customer segments with good statistical validation metrics. The Davies-Bouldin Index of 0.842 and Silhouette Score of 0.573 indicate well-formed clusters that can be effectively used for targeted marketing strategies and customer relationship management.