Cluster Analysis Report

Overview

The purpose of this analysis is to separate customers into different groups based on their profile information and transaction data. In doing so, we aim to analyze customer behavior and develop strategies for personalized marketing and operational excellence.

1. Number of groups created:

- As a result of analyzing the data using the K-Means clustering algorithm, 4 main clusters were identified. The quality of the clusters was determined using the bracket, which measures the number of changes explained by each additional cluster. - These categories represent different types of customers:
- Cluster 0: Mostly clothes related.
- Cluster 1: Mostly electronics related.
- Cluster 2: Focus on home improvement.
- Cluster 3: Related books.

2. Davis-Boulding Index (DB Index):

- The DB performance index of the cluster is 1.634. This value indicates the performance of the cluster, as the DB Performance index measures the compactness and separation of the cluster (a lower value indicates a better cluster).

Other measurement methods

1. Inertia

- The elbow method uses inertia measurements to measure differences within a group. Inertia values are plotted for groups of 2 to 10. The elbow point of 4 groups shows the best balance between model complexity and the cost of explaining variance.

2. PCA Visualization

- Use Principal Component Analysis (PCA) to visualize the clustering results and reduce the remaining data to 2 components for easier understanding.
- Scatter plot of reduced features showing different groups, showing the performance of the segmentation.

3. CLASS FEATURES

- Each category reflects its unique features and business models, maximizing value through customer distribution across regions and products.

Customer Segment Analysis

1. Cluster 0: *Clothing*

- North America: 14 divisions (maximum).
- South America: 9 divisions (minimum).
- Europe: 10 divisions.
- Asia: 13 divisions.

- Key Findings: North America is the region with the highest interest in apparel, indicating regional market potential. South America has the lowest participation, indicating growth opportunity.

2. Cluster 1: *Electronics*

- Regional Distribution:
- South America: 16 divisions (maximum).
- North America: 9 divisions
- Europe: 9 divisions
- Asia: 11 divisions
- Key Findings: South America leads in energy participation, while interest is lower in North America and Europe. Marketing campaigns can focus on growing the business in these areas.

3. Cluster 2: *Home Decor*

- Europe: 14 divisions (maximum).
- South America: 13 divisions
- Asia: 8 divisions (lowest)
- North America: 7 divisions
- Key Findings: Europe has the highest interest in home improvement, with lower participation in North America and Asia. These regions could benefit from awareness projects.

4. Cluster 3: *Books*

- South America: 21 divisions (maximum)
- Europe: 17 divisions
- North America: 16 divisions.
- Asia: 13 divisions (lowest).
- Key Findings: Books are most popular in South America, followed by Europe and North America. Asia has the lowest participation, indicating a potential market for digital or other types of content.
-

In conclusion

- The successful collaboration process divided the customers into 4 different groups and each group represented a specific behavior and profile change.
- The DB index is 1.634 and visual inspection with PCA shows that the clusters are compact and well separated.
- This assessment provides actionable insights into marketing plans, enabling personalization of the approach to customer interaction.