Customer Segmentation Clustering Report

1. Overview

This report presents the findings of customer segmentation based on transactional data. The clustering was performed to identify patterns in customer purchasing behavior and categorize customers into distinct groups for better-targeted marketing strategies. The analysis was conducted using K-Means clustering.

2. Number of Clusters Formed

After evaluating multiple clustering models, the optimal number of clusters was determined to be 9. This was based on the Davies-Bouldin Index (DB Index), which measures the compactness and separation of clusters. A lower DB Index indicates better clustering.

3. Evaluation Metrics

• Optimal Number of Clusters: 9

• Davies-Bouldin Index: 1.0846 (indicating well-separated and compact clusters)

Other Metrics:

• Inertia:

Measures the sum of squared distances of samples to their nearest cluster center. o Observed decreasing inertia values with an increase in clusters, ensuring convergence.

• Silhouette Score: Though not computed directly, visual inspection indicates that clusters have clear separations in the feature space.

4. Clustering Features

The clustering was based on the following standardized features:

• Total Spending (TotalValue): Total amount spent by a customer.

• Number of Transactions (TransactionID): Total transactions made by a customer.

• Total Quantity Purchased (Quantity): Total number of items purchased.

• Average Price per Item (Price): Average price of items purchased.

• Region (Region): Encoded customer region.

5. Cluster Analysis

• Cluster 0: Low spenders with minimal quantity purchases and infrequent transactions.

• Cluster 1: Medium spenders with moderate transactions and quantities purchased.

• Cluster 2: High spenders with frequent transactions and larger quantities purchased.

• Cluster 3: Customers who exhibit an average spending pattern with a specific preference for medium-priced items.

• Cluster 4: High spenders in specific regions with unique purchasing behaviors.

• Cluster 5: Customers with low spending but high transaction frequency.

• Cluster 6: Customers with moderate spending and high preference for certain categories.

• Cluster 7: Low to medium spenders with infrequent purchases.

• Cluster 8: Premium customers with significant spending and high quantities purchased.

6. Visualization

A scatter plot was generated to visualize customer segmentation:

• X-axis: Total Spending (USD)

• Y-axis: Total Quantity Purchased

• Clusters: Represented using distinct colors for clear visualization.

7. Applications and Insights

The clustering results can be used for:

• Targeted Marketing: Tailoring promotions and offers for specific customer segments.

• Customer Retention: Identifying low-spending customers and strategizing ways to increase their engagement

. • Resource Allocation: Focusing resources on high-value customers (e.g., Cluster 8).

• Regional Strategies: Developing region-specific marketing campaigns based on customer preferences.

8. Limitations and Future Improvements

• Feature Selection: Additional features like customer demographics or seasonal trends could improve the clustering quality.

• Validation: Incorporating additional clustering validation metrics like the Silhouette Score.

• Dynamic Analysis: Analyzing temporal data to identify seasonal trends and changing customer behaviors.

9. Conclusion

The customer segmentation analysis successfully identified 9 distinct clusters with a DB Index of 1.0846. These clusters provide valuable insights into customer behavior and can guide data-driven decision-making for targeted marketing and customer relationship management.