Customer Segmentation Report

1. Overview

Customer segmentation is a critical task for identifying patterns in customer behaviour and tailoring strategies for engagement. Using clustering techniques, we grouped customers based on similar transaction patterns and profiles. For this analysis, we selected the K-Means algorithm due to its simplicity, scalability, and ease of interpretation. These properties make it particularly effective for analysing structured data like customer and transaction records. The results will guide personalized marketing efforts, customer retention strategies, and targeted promotions.

2. Methodology

- **Data Preparation:**
  Customer and transaction datasets were merged to create a unified view of customer behaviour. Key features used in the clustering process included:

    - Total spending by each customer

    - Average transaction value over time

    - Total number of transactions

    - Average quantity purchased per transaction
      To ensure fairness among features, the data was standardized. This step was essential to prevent features with larger scales from dominating the clustering process.

- **Clustering Algorithm:**
  K-Means was chosen as the clustering algorithm. It divides data into non-overlapping groups based on similarities. We tested cluster sizes ranging from 2 to 10 and evaluated their quality using the Davies-Bouldin (DB) Index. A lower DB Index indicates compact and well-separated clusters, which represent meaningful groupings of customers.

3. Results

- **Optimal Number of Clusters:**
  After testing, the optimal number of clusters was determined to be 5, achieving a DB Index of 1.05. This configuration offered a good balance between model complexity and cluster quality.

- **Description:**

  Cluster 0: High-frequency buyers who have medium levels of spending. These customers are actively engaged but not necessarily the highest spenders.

Cluster 1: This segment comprises low-frequency and low-spending customers, representing occasional buyers with minimal engagement.

Cluster 2: These high-value customers are characterized by significant spending and moderate engagement. They constitute a critical segment for premium product offerings.

Cluster 3: This group of new or less active customers demonstrates very little activity, suggesting a need for reactivation campaigns to re-engage them.

Cluster 4: Customers in this group are moderate spenders who purchase frequently, exhibiting consistent engagement and presenting opportunities for increasing their basket size.

- **Visualization:**
  Dimensionality reduction using PCA helped visualize the clusters in a 2D space. The scatter plot highlighted clear separations between clusters, validating the effectiveness of the segmentation.

4. Why K-Means?
K-Means was selected for several reasons:

- **Efficiency:** It efficiently handles large datasets with multiple features.
- **Practicality:** The algorithm is intuitive and provides distinct, easy-to-interpret clusters.
- **Evaluation:** Using metrics like the DB Index ensures that clusters are meaningful and not arbitrary.
- **Relevance:** K-Means works well with standardized numerical data, aligning with the features derived for this analysis.

5. Recommendations:

Cluster 0: These actively engaged customers could be further motivated with customer loyalty programs or personalized promotional offers.

Cluster 1: Providing basic incentives or reminders may re-engage these occasional buyers and convert them into more frequent customers.

Cluster 2: This high-value customer segment should be targeted with premium services, upselling strategies, or exclusive product offerings.

Cluster 3: Less active customers necessitate personalized reactivation campaigns. Tailoring offers or reminders based on their past purchasing behaviour may encourage their return.

Cluster 4: Moderate spenders could be incentivized to increase their average order value through discounts on complementary or bundled products.

6. Future Scope

- The DB Index score of 1.05 suggests the clustering results are acceptable, but future research could investigate alternative techniques like hierarchical clustering or DBSCAN.

- Adding additional data sources, such as personal information (like age, income) or psychographic information (e.g. preferences, behaviour), could enable more granular and insightful customer segmentation.

- Studying how clusters evolve over time could uncover trends and improve the company's ability to predict customer behaviour.