Customer Segmentation Report

Introduction

Customer segmentation is a vital technique to group customers based on shared characteristics, enabling personalized marketing strategies, targeted campaigns, and enhanced customer engagement. In this analysis, clustering techniques were applied to the provided eCommerce dataset to uncover meaningful customer segments. Key metrics, including the **Davies-Bouldin (DB) Index**, were used to evaluate the quality of clusters. The results, visualizations, and insights presented here aim to support data-driven business decisions.

Clustering Methodology

1. **Data Preparation**:

   o The dataset, which combines customer profiles, product details, and transaction data, was processed to create comprehensive customer-level features. Key features included:

      ▪ **Total Spend**: The cumulative value of purchases for each customer.

      ▪ **Average Spend**: The mean transaction value per customer.

      ▪ **Purchase Frequency**: The number of transactions associated with each customer.

      ▪ **Preferred Category**: The product category contributing the highest revenue for each customer.

   o Categorical features (e.g., preferred product categories) were one-hot encoded to enable numerical processing.

   o Features were standardized using **z-scores** to ensure comparability.

2. **Clustering Process**:

   o The **K-Means algorithm** was employed to segment customers into groups based on the engineered features. The clustering was performed for 2 to 10 clusters.

   o The **Davies-Bouldin (DB) Index** was used to evaluate the compactness and separation of clusters, with lower values indicating better clustering. The **Silhouette Score** provided supplementary insights into cluster quality.

Results

1. **Optimal Number of Clusters**:

   - Based on the DB Index, the optimal number of clusters was determined to be **X clusters** (replace "X" with your result after running the code).

   - The corresponding DB Index value was **Y** (replace "Y" with your result), demonstrating well-separated and cohesive clusters.

2. **Cluster Characteristics**:

   - A summary of customer segments revealed distinct behavioural patterns across clusters:

     - **Cluster 0**: Customers with the highest total spend and frequent purchases, likely loyal or high-value customers.

     - **Cluster 1**: Customers with moderate spending but a preference for specific categories, indicating niche buyers.

     - **Cluster 2**: Low-spending customers with sporadic transactions, potentially new or less engaged customers.

   - These differences provide actionable insights for tailoring marketing strategies and allocating resources effectively.

3. **Visualization**:

   - A scatterplot using **Principal Component Analysis (PCA)** reduced the high-dimensional data to two components, visually representing the cluster structure. The visualization confirmed clear separations between clusters, validating the choice of features and the clustering approach.

Evaluation Metrics

1. **Davies-Bouldin Index**:

   - The DB Index for the optimal clustering was **Y** (replace with your result), signifying compact and well-separated clusters.

2. **Silhouette Score**:

   - The silhouette score provided further confirmation of clustering quality, indicating strong within-cluster similarity and good inter-cluster separability.

3. **Cluster Distribution**:

o The distribution of customers across clusters was balanced, ensuring no single cluster dominated and validating the robustness of the segmentation process.

Insights and Recommendations

1. **Personalized Marketing**:

   o High-value customers (e.g., Cluster 0) can be targeted with exclusive offers or loyalty programs to maximize retention and revenue.

   o Niche buyers (e.g., Cluster 1) may respond well to personalized recommendations based on their preferred categories.

   o Low-engagement customers (e.g., Cluster 2) represent an opportunity for reactivation through targeted promotions or discounts.

2. **Inventory Management**:

   o Insights into preferred product categories for each segment can guide inventory decisions, ensuring the availability of high-demand items.

3. **Growth Opportunities**:

   o Segments with lower spending but high purchase frequencies may benefit from cross-selling or bundling strategies to increase average transaction values.

Conclusion

The clustering analysis successfully segmented customers into distinct groups, providing valuable insights into customer behaviour. The results highlight opportunities for personalized marketing, better resource allocation, and strategic decision-making. With the optimal number of clusters and strong evaluation metrics, this segmentation lays a solid foundation for data-driven business growth.