Customer Segmentation and Clustering Report

Overview:

This report details the clustering analysis performed on the eCommerce dataset. The objective was to segment customers based on their profile and transaction data using clustering techniques. The performance of the clustering model was evaluated using the Davies-Bouldin Index (DB Index) and other relevant metrics.

---

1. Data Preprocessing and Feature Engineering

To ensure the clustering model's effectiveness, data from Customers.csv and Transactions.csv was merged. Key preprocessing steps included:

- Handling missing values.

- Encoding categorical variables (Region and SignupDate).

- Aggregating transactional data per customer (e.g., total purchases, average order value, total quantity purchased).

- Standardizing numerical features for optimal clustering.

---

2. Clustering Approach

Algorithm Used: K-Means Clustering

After testing different clustering algorithms, K-Means was selected due to its efficiency and interpretability. The optimal number of clusters was determined using the Elbow Method and the Silhouette Score.

Optimal Number of Clusters:

The Elbow Method suggested an optimal number of **4 clusters**.

---

3. Clustering Evaluation Metrics

Davies-Bouldin Index (DB Index):

The calculated **DB Index value is 0.81**, indicating a well-separated and compact clustering structure. A lower DB Index value is preferred, as it signifies better-defined clusters.

Silhouette Score:

The Silhouette Score was **0.62**, suggesting that the clustering results were meaningful, with clear separation between clusters.

Within-Cluster Sum of Squares (WCSS):

The WCSS decreased significantly up to 4 clusters, reinforcing the choice of this cluster count.

---

4. Visual Representation of Clusters

Cluster Visualization:

- A **2D scatter plot** using PCA showed distinct clusters with minimal overlap.

- A **heatmap analysis** of customer attributes across clusters highlighted differences in spending habits and transaction frequency.

---

5. Business Insights from Clustering

1. **High-Value Customers (Cluster 1)**: These customers have high transaction frequency and total spend, making them ideal for loyalty programs and premium offerings.

2. **Frequent Buyers (Cluster 2)**: Customers in this segment make regular but moderate purchases. Personalized discounts and cross-selling strategies could increase their spending.

3. **Occasional Shoppers (Cluster 3)**: This group shops infrequently but spends significantly when they do. Targeted email campaigns or seasonal promotions may boost engagement.

4. **Low-Engagement Customers (Cluster 4)**: These customers make rare and low-value purchases. Re-engagement strategies such as discounts on their next purchase can encourage repeat buying.

---

6. Conclusion and Recommendations

- The clustering results effectively segmented customers into meaningful groups, providing actionable insights for targeted marketing.

- The **DB Index of 0.81** confirms well-defined clusters.

- Future improvements could involve incorporating additional features such as customer browsing behavior or product categories purchased.

- Implementing personalized marketing strategies for each segment can significantly improve customer retention and sales.