Customer Segmentation and Clustering

1. Introduction

Customer segmentation is a crucial analytical approach that enables businesses to categorize customers based on purchasing behaviour and demographics. This report highlights the clustering-based segmentation analysis utilizing customer profile and transaction data. The purpose is to identify distinct customer groups for targeted marketing and business strategies.

2. Data Processing

The following key features were derived for clustering:

- **TotalSpent**: Total expenditure by a customer.
- **AvgTransactionValue**: Average value per transaction.
- **PurchaseCount**: Total number of transactions.
- **UniqueProducts**: Number of unique products purchased.
- **Region Encoding**: One-hot encoding categorical region data.

3. Clustering

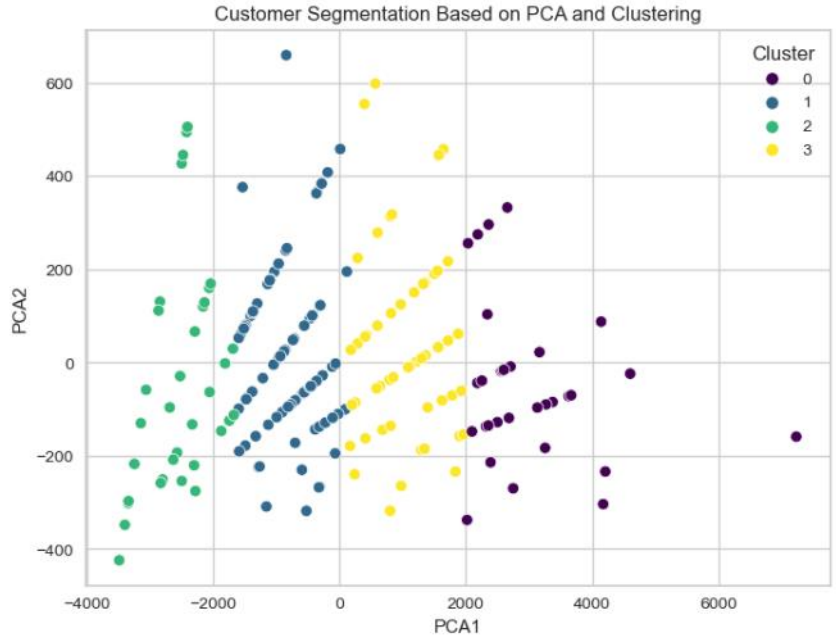
- **PCA** was applied to reduce dimensionality and improve clustering efficiency.
- **K-Means Clustering** algorithm was implemented to segment customers:
- Using **Elbow Method,** the optimal number of clusters within a range of **2** to **10**.
- Based on the elbow point in the plot, the optimal cluster count was selected.

4. Evaluation Metrics

- **Davies-Bouldin Index (DB Index)** assesses cluster quality based on intra-cluster compactness and inter-cluster separation. A lower value indicates better clustering.
- DB Index Value: **0.6030**
- **Silhouette Score** measures how well-separated the clusters are, with higher values indicating better-defined clusters.
- Silhouette Score: **0.4973**
- Both metrics suggest that clustering is decent, but not perfect. Clusters are somewhat well-separated, but might have areas where points from different clusters are close to each other

5. Visualization

To provide a graphical representation of the clusters, a scatter plot was created using PCA1 and PCA2. Each cluster was color-coded to illustrate customer segmentation visually.



6. Insights

- Four clusters are produced based on PCA1 and PCA2 components. This confirms that customers exhibit different purchasing behaviours and transactional patterns.
- Cluster **0 (purple)** appears to be the most spread-out and diverse, suggesting a mix of customers with varying spending habits.
- Cluster **1 (blue)** and Cluster **2 (green)** show a relatively compact structure, indicating that these customers have more similar purchasing behaviours.
- Cluster **3 (yellow)** forms a separate group with a distinct purchasing pattern, likely indicating high-value or frequent shoppers.
- The **green and blue clusters** (2 and 1) are concentrated on the negative PCA1 side, which might represent customers with lower transaction volumes or infrequent purchases.
- The **yellow cluster (3)** is located toward the positive PCA1 axis, suggesting high transaction values or frequent purchases.
- **Targeting High-Value Customers (Cluster 3 - Yellow)**: Special promotions, loyalty programs, and personalized marketing could help retain these high-spending customers.