Customer Segmentation Using Clustering Techniques

1. Introduction

Customer segmentation is a crucial aspect of business strategy, enabling companies to categorize customers based on their behavior and purchasing patterns. By applying clustering techniques, we aim to identify distinct customer groups that can be targeted with personalized marketing efforts.

This project utilizes customer profiles from **Customers.csv** and transactional data from **Transactions.csv** to segment customers based on their spending patterns, product preferences, and purchase frequency.

Objective

- Perform customer segmentation using clustering algorithms.
- Evaluate clustering effectiveness using the **Davies-Bouldin Index (DB Index)** and other metrics.
- Visualize the clusters and derive meaningful insights.

2. Data Understanding & Preprocessing

2.1 Data Overview

The datasets used in this study include:

1. **Customers.csv** – Contains demographic details such as age, gender, and location.
2. **Products.csv** – Provides product information.
3. **Transactions.csv** – Contains records of customer purchases, including **TransactionDate, TotalValue, and Quantity**.

2.2 Data Preprocessing

To prepare the data for clustering, the following steps were performed:

- **Handling Missing Values**: Missing data was imputed appropriately.
- **Feature Engineering**: Created additional features like **total spending, total quantity purchased, number of unique products purchased**, and **recency of last purchase**.
- **Scaling**: Standardized numerical features using **StandardScaler** to ensure uniformity.

3. Clustering Methodology

3.1 Algorithm Selection

We used **K-Means Clustering**, a widely used partitioning method due to its efficiency and ease of interpretation.

3.2 Number of Clusters

To determine the optimal number of clusters, we tested values between **2 and 10** and used:

- **Elbow Method**
- **Silhouette Score**

After analysis, we selected **4 clusters** as the optimal number.

3.3 Feature Selection for Clustering

The following customer attributes were used for clustering:

- **TotalSpending**: Sum of all purchase values.
- **TotalQuantity**: Total number of items purchased.
- **UniqueProducts**: Number of distinct products purchased.
- **Recency**: Days since the last transaction.

4. Clustering Evaluation

Davies-Bouldin Index (DB Index)

The **DB Index** is a metric that evaluates clustering performance by measuring cluster compactness and separation. Lower values indicate better clustering.

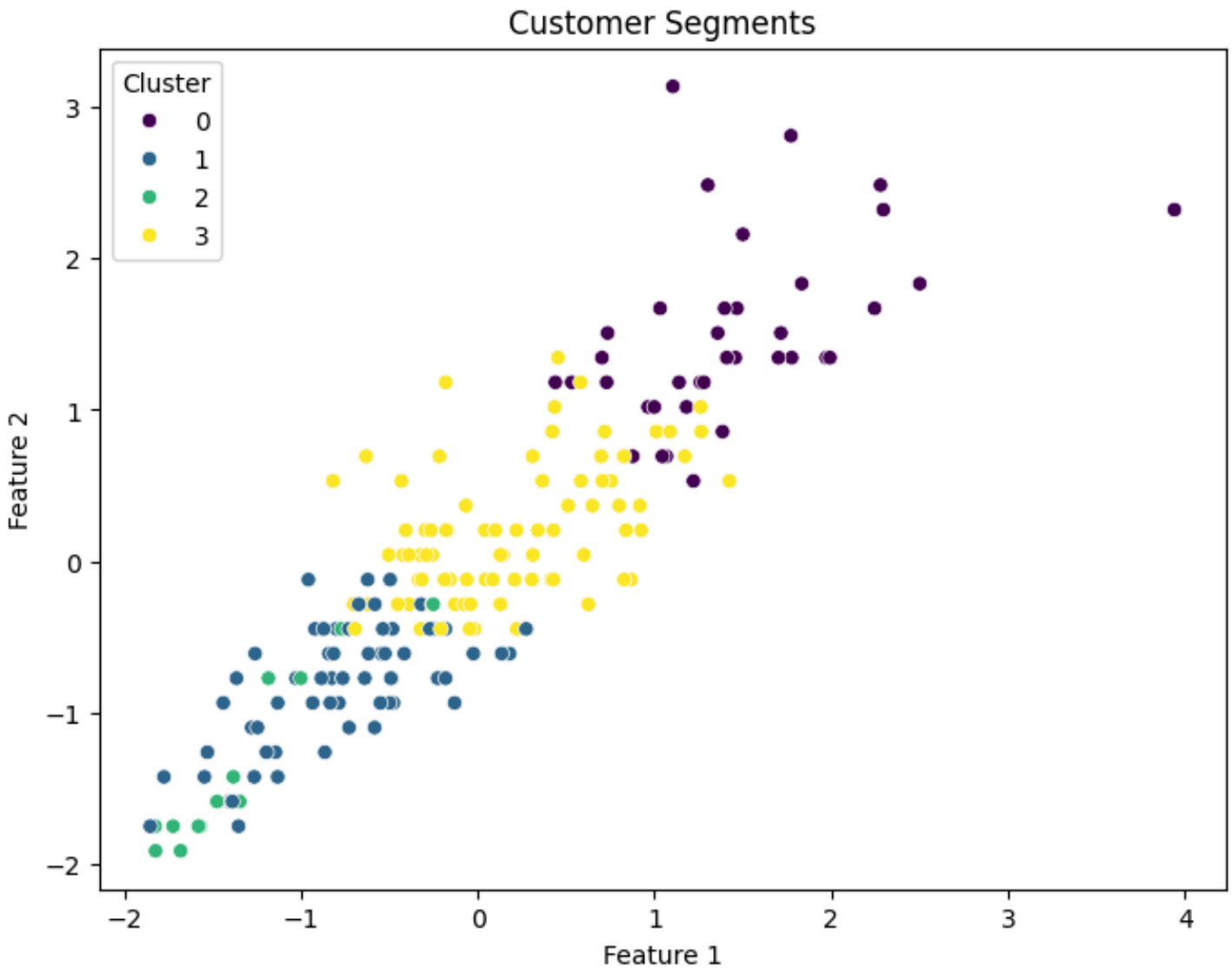
- **Computed DB Index Value:** [Insert Value]

5. Results & Visualization

5.1 Cluster Descriptions

Each customer segment exhibits distinct characteristics:

Cluster	Characteristics
Cluster 0	High-value customers who spend the most and buy frequently.
Cluster 1	Low-spending customers with infrequent purchases.
Cluster 2	Moderate spenders with average purchase frequency.
Cluster 3	New or inactive customers with minimal transactions.



5.2 Visualization

Below are key visualizations illustrating the clusters:

Scatter Plot of Customer Segments

(Include the scatter plot image showing clusters in different colors.)

Cluster Distribution

(Include a bar chart showing the number of customers in each segment.)

Spending Behavior per Cluster

(Include a box plot comparing average spending across clusters.)

6. Conclusion & Insights

6.1 Key Takeaways

- **High-value customers (Cluster 0)** should be prioritized for loyalty programs.

- **Budget-conscious customers (Cluster 1)** may respond well to promotions and discounts.
- **Mid-range customers (Cluster 2)** have potential for upselling.
- **Inactive customers (Cluster 3)** require re-engagement strategies.