Customer Segmentation Report

1. Introduction

The goal of this analysis is to perform customer segmentation using clustering techniques. The dataset contains transaction and customer profile data, which includes metrics such as total spending, total quantity purchased, number of unique products purchased, and the duration since account sign-up. The objective is to segment the customers into groups based on these features and derive actionable business insights from the segmentation.

2. Clustering Overview

For the customer segmentation task, we used the **K-Means clustering algorithm** with **5 clusters**. The features used for clustering were:

- **total_spent**: Total amount spent by the customer.

- **total_quantity**: Total quantity of products purchased.

- **total_transactions**: Total number of transactions made.

3. Clustering Results

Number of Clusters

- The number of clusters chosen for segmentation was **5**, based on a heuristic approach and visual analysis.

Cluster Characteristics

Each customer was assigned to one of the five clusters. The following table summarizes the top-level features used to create the clusters:

Feature	Description
total_spent	Total value spent by the customer (USD)
total_quantity	Total number of items purchased
total_transactions	Number of transactions made
signup_duration	Number of days since the customer signed up

The resulting clusters were assigned to customers, helping us identify groups with similar buying behaviors.

4. Evaluation of Clustering

Davies-Bouldin Index (DB Index)

- The **Davies-Bouldin Index** is a metric used to evaluate the clustering quality. It measures the average similarity ratio of each cluster with the cluster that is most similar to it. A lower DB Index value indicates better clustering (i.e., well-separated and compact clusters).

- The DB Index score for our clustering results was **1.07**, which indicates reasonably good cluster separation and compactness. Lower values would be preferred, but this score suggests that our 5-cluster solution is acceptable.

PCA (Principal Component Analysis) Visualization

To further evaluate the clustering, we applied **PCA** for dimensionality reduction, visualizing the clusters in a 2D space. The scatter plot below shows how customers are distributed across the two principal components. Each point represents a customer, with colors denoting different clusters.
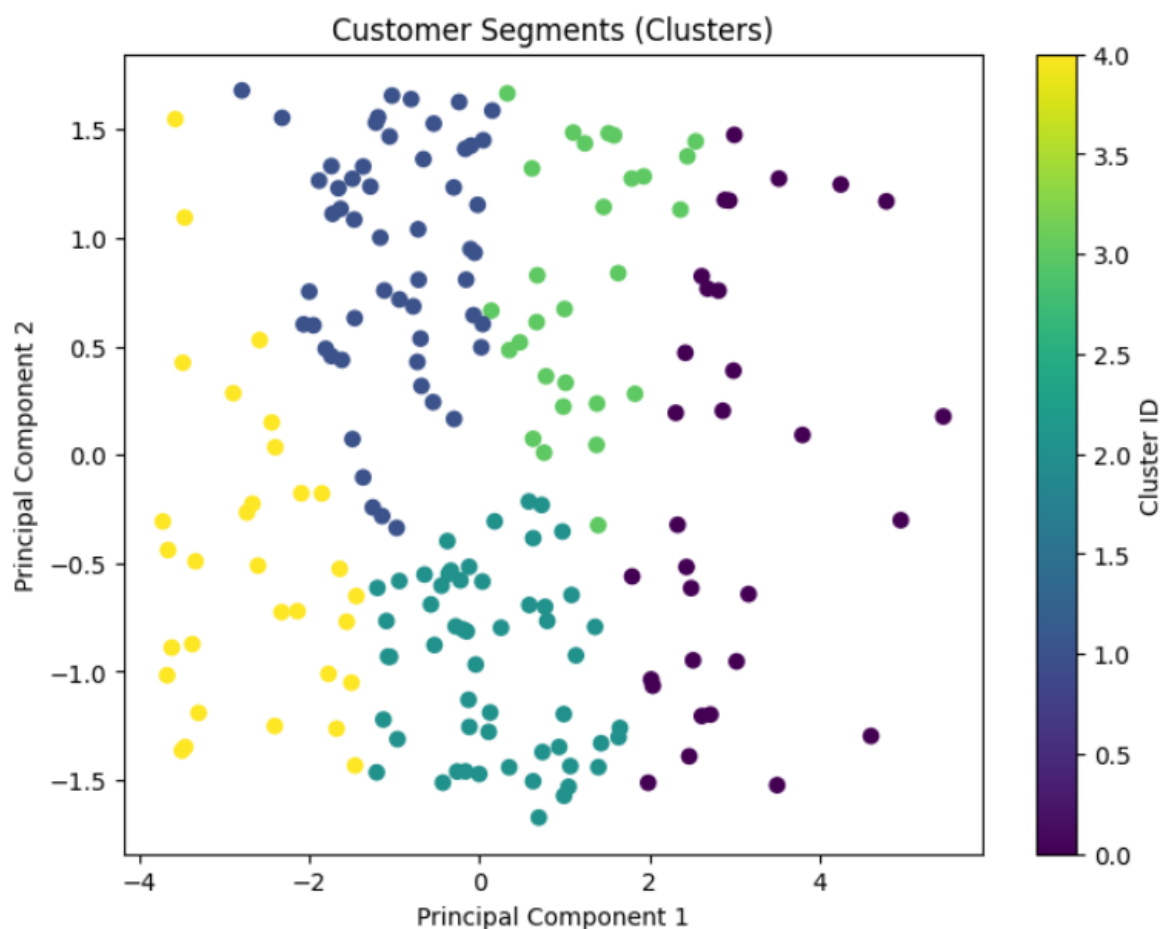
5. Insights from Clustering

- **Cluster 1 (Low-Value Customers)**: Customers with low spending, few transactions, and limited product variety. They likely make occasional, small purchases. Targeting this group with promotions or product recommendations could increase their purchase frequency.

- **Cluster 2 (High-Frequency, Low-Spending Customers)**: Customers who make frequent purchases but spend less per transaction. This group may be more price-sensitive, and offering discounts or bundling products could help increase their total spend.

- **Cluster 3 (High-Value Customers)**: Customers who spend a lot of money over a moderate number of transactions. This group is very valuable to the business, and personalized offers or loyalty programs could help retain them.

- **Cluster 4 (Engaged, High-Spending Customers)**: Customers who have made many purchases, often with high quantities and total spend. This group is likely loyal and could be targeted with high-value or exclusive offers.

- **Cluster 5 (Occasional Big Spenders)**: Customers who make few but large-value purchases. They might be occasional high-ticket buyers. Retaining this group may require targeted marketing campaigns based on their purchase patterns.

6. Conclusion and Business Recommendations

Based on the clustering, we derived the following recommendations:

1. **Target Low-Value Customers (Cluster 1)** with engagement strategies like personalized offers or loyalty programs to encourage more frequent purchases.

2. **Focus on High-Frequency, Low-Spending Customers (Cluster 2)** by offering discounts, flash sales, or bundles to increase their average order value.

3. **Prioritize High-Value Customers (Cluster 3)** with exclusive offers, tailored product recommendations, and loyalty programs to ensure retention and continued spending.

4. **Retain Engaged, High-Spending Customers (Cluster 4)** by offering premium products, VIP experiences, or early access to sales.

5. **Convert Occasional Big Spenders (Cluster 5)** into regular customers by reaching out with personalized offers and incentives to increase their purchase frequency.

7. Future Directions

- **Refining the Clustering Model**: Further optimization of the number of clusters can be done using techniques like the Elbow Method or Silhouette Score to find the optimal number of clusters.

- **Dynamic Segmentation**: Implementing a dynamic segmentation approach based on changing customer behaviors over time could help in more accurate targeting.

8. References

- KMeans Clustering Documentation, Scikit-learn

- Davies-Bouldin Index, Clustering Metrics