Customer Segmentation Report

By: Ruhul Fatima Abdi
Date: 01/02/2025

Introduction

Customer segmentation is essential for understanding buying behaviour and improving business strategies. In this analysis, we applied **K-Means Clustering** to segment customers based on their **Total Spending** and **Purchase Frequency**.

To determine the optimal number of clusters (K), we evaluated multiple clustering metrics:

- **Davies-Bouldin Index (DBI)** is a metric used to evaluate the quality of clustering algorithms. It is the average of the maximum ratio of the within-cluster distance to the between-cluster distance for each cluster. The ratio is defined as the similarity between the clusters. *(Lower is better, indicating well-separated clusters).*
- **Silhouette Score** is a metric that measures how well data points are clustered together. It measures how similar a data point is to the other data points in its cluster *(Higher is better, indicating well-defined clusters).*
- **Calinski-Harabasz Index,** also known as the **Variance Ratio Criterion**(**VRC**), is a metric used to evaluate the clustering algorithms. It measures the ratio of between-cluster dispersion to within-cluster dispersion *(Higher is better, indicating dense, well-structured clusters).*

After testing **K values from 2 to 10**, the optimal number of clusters was found to be **K = 2** with a **DBI of 0.734**, the lowest among all tested values.

Clustering Performance Metrics

K (Clusters)	Davies-Bouldin Index (↓ Better)	Silhouette Score (↑ Better)	Calinski-Harabasz Index (↑ Better)
2 (Optimal)	**0.734**	**0.487**	**281.87**
3	0.766	0.424	277.30
4	0.859	0.377	266.56
5	0.849	0.390	252.46
6	0.917	0.391	245.54
7	0.876	0.386	252.21
8	0.908	0.371	244.52
9	0.880	0.371	244.83
10	0.855	0.381	248.49

Best K = 2 because:

- **Lowest DBI (0.734) → Best cluster separation**

- **Highest Silhouette Score (0.487) → Well-defined clusters**
- **Highest Calinski-Harabasz Index (281.87) → Strong cluster density**

Business Insights from Segmentation

1. Two Major Customer Segments Identified

With **K = 2**, customers are divided into **two distinct segments**:

- **Segment 1:** Low-spending, infrequent buyers
- **Segment 2:** High-spending, frequent buyers

This helps businesses **tailor marketing strategies** for each group.

2. Low-Spending Customers: Need More Engagement

- The majority of customers **spend less and purchase infrequently**.
- Strategies: **Targeted promotions, loyalty rewards, and personalized discounts** can encourage repeat purchases.

3. High-Spending Customers: VIP Treatment

- A smaller group of customers **spends significantly more and purchases frequently**.
- Strategies: **Exclusive memberships, early access to sales, and personalized recommendations** can retain these high-value customers.

Recommendations

Retarget Low-Spending Customers

- Offer **discounts on their second purchase** to increase repeat orders.
- Use **email reminders** for abandoned carts.

Increase Spending Among Mid-Tier Customers

- Introduce **product bundles or upsell strategies**.
- Suggest **related products based on purchase history**.

Reward High-Spending Customers

- Create a **VIP program** with early access to new products.
- Provide **priority customer support** and exclusive deals.

Conclusion

This clustering analysis helps businesses understand **customer behavior patterns** and improve marketing strategies. By leveraging segmentation insights, businesses can drive **higher revenue, improve customer retention, and optimize marketing efforts**.

The final segmentation results are saved in: **RuhulFatimaAbdi_Optimized_Clustering.csv**