

Customer Segmentation & Clustering Report

Prepared by: Riya Bose

Contents

1	Introduction	2
2	Clustering Methodology	2
2.1	Dataset Description	2
2.2	Preprocessing Steps	2
2.3	Clustering Algorithm	2
3	Results & Analysis	2
3.1	Number of Clusters Formed	2
3.2	Evaluation Metrics	2
3.3	Cluster Descriptions	3
3.4	Visualizations	3
4	Conclusion & Recommendations	3
4.1	Key Findings	3
4.2	Recommendations	3
4.3	Next Steps	4

1. Introduction

This report focuses on customer segmentation using clustering techniques. The goal is to group customers based on their transaction behavior and profile information, enabling data-driven marketing and operational strategies. Clustering evaluation metrics, such as the Davies-Bouldin Index (DBI), are used to assess clustering performance.

2. Clustering Methodology

2.1 Dataset Description

- **Customers.csv:** Contains customer profile information, including region, signup date, and demographic details.
- **Transactions.csv:** Includes transaction history, such as transaction dates, values, and product categories.

2.2 Preprocessing Steps

- **Data Cleaning:** Handled missing values and ensured data consistency.
- **Encoding:** Categorical variables, such as *Region* and *TopCategory*, were one-hot encoded.
- **Scaling:** Applied standard scaling to normalize the features.

2.3 Clustering Algorithm

- **Algorithm Used:** K-Means Clustering.
- **Number of Clusters:** Determined using the Elbow Method.
- **Evaluation Metrics:** Davies-Bouldin Index (DBI) and Silhouette Score.

3. Results & Analysis

3.1 Number of Clusters Formed

Based on the Elbow Method, the optimal number of clusters was determined to be **4**.

3.2 Evaluation Metrics

- **Davies-Bouldin Index:** 1.46
- **Silhouette Score:** 0.29

3.3 Cluster Descriptions

- **Cluster 0:** Low spenders with infrequent transactions.
- **Cluster 1:** High spenders with frequent transactions (VIP customers).
- **Cluster 2:** Mid-tier spenders with consistent purchasing behavior.
- **Cluster 3:** Occasional bulk buyers or outliers.

3.4 Visualizations

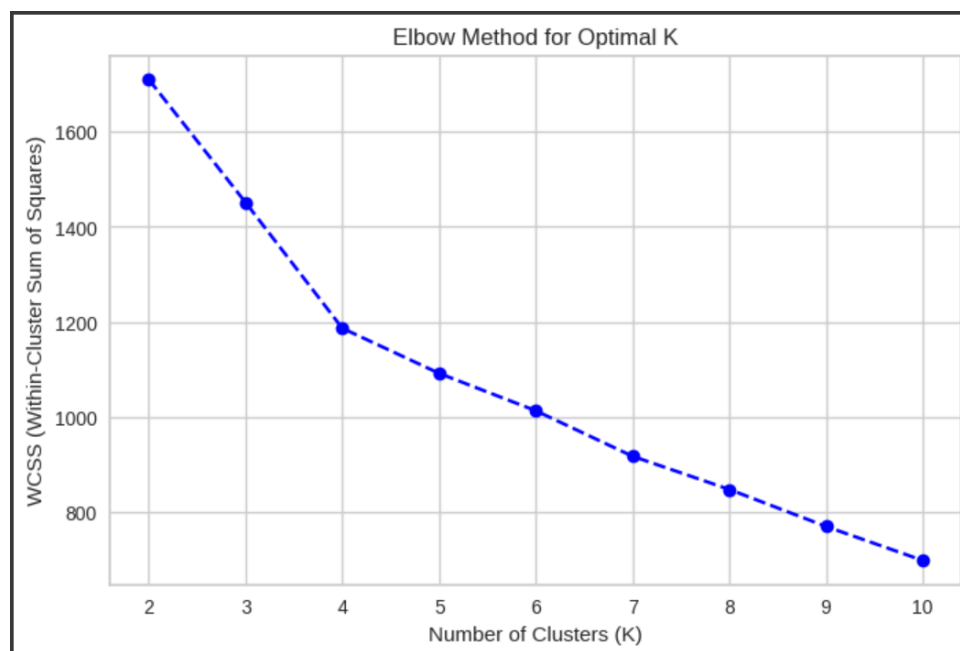


Figure 1: Elbow Method to Determine Optimal Clusters

4. Conclusion & Recommendations

4.1 Key Findings

- The clustering analysis identified four distinct customer groups.
- High spenders (Cluster 1) contribute disproportionately to revenue.
- Low spenders (Cluster 0) represent an opportunity for targeted re-engagement.

4.2 Recommendations

- **VIP Customers (Cluster 1):** Offer exclusive loyalty programs and personalized discounts to retain these customers.
- **Low Spenders (Cluster 0):** Design targeted marketing campaigns to increase engagement and transactions.

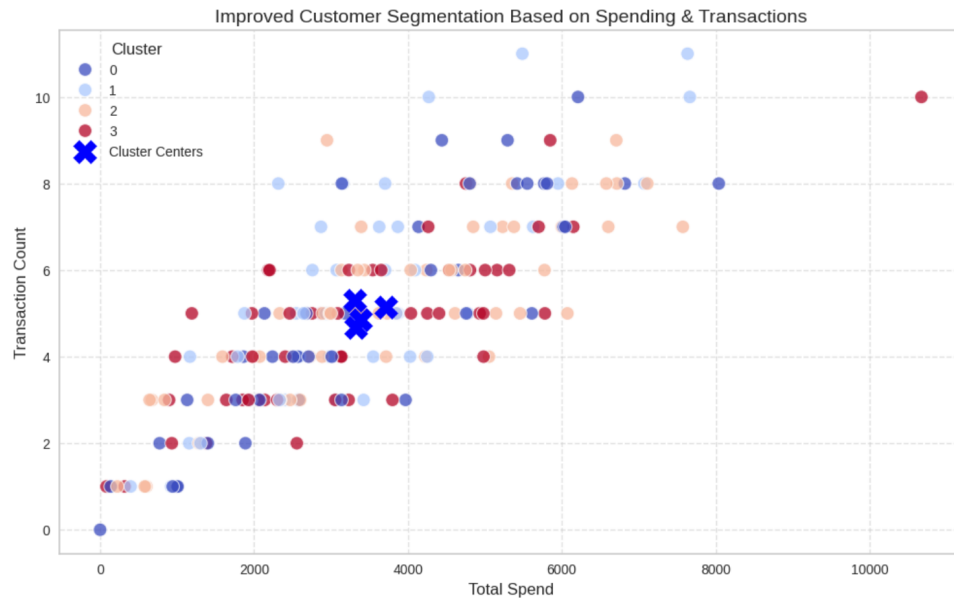


Figure 2: Cluster Visualization Based on Spending and Transactions

- **Mid-Tier Buyers (Cluster 2):** Promote subscription plans and upselling opportunities to encourage higher spending.
- **Bulk Buyers (Cluster 3):** Analyze bulk purchasing patterns to identify specific needs and preferences.

4.3 Next Steps

- Implement targeted marketing strategies based on cluster insights.
- Monitor cluster performance and update segmentation periodically.