Customer Segmentation: Clustering Results

1. Overview

The customer segmentation task aimed to classify customers based on their profile information (from Customers.csv) and transaction data (from Transactions.csv). The goal was to segment the customers into distinct groups that share similar characteristics, enabling targeted marketing strategies and insights into customer behavior.

---

2. Clustering Methodology

2.1 Algorithm Used

- **Clustering Algorithm:** K-Means clustering.
- **Number of Clusters:** The optimal number of clusters was determined using the Elbow Method. Based on plot analysis, the optimal number of clusters was found to be **4**.

2.2 Features Used

- **Customer Profile Features:** Region, Age, and Gender (encoded numerically).
- **Transaction Features:** Total transaction value, average transaction value, and transaction frequency.

2.3 Preprocessing Steps

- **Data Cleaning:** Missing values were handled by imputation.
- **Encoding Categorical Data:** Categorical data (Region) was encoded using one-hot encoding.
- **Feature Scaling:** Features were scaled using **Min-Max scaling** to normalize the data and ensure equal contribution from each feature in the clustering process.

---

3. Clustering Results

3.1 Number of Clusters Formed

- **Optimal Number of Clusters:** 4 clusters, as determined by the Elbow Method.
    - The Elbow Method showed a sharp bend at 3 clusters, followed by diminishing returns after this point, suggesting that the **optimal trade-off** between model complexity and explained variance occurs at 4 clusters.

3.2 Clustering Metrics

- **Silhouette Score:**
    - **Value:** 0.431
    - **Interpretation:** A moderate level of **cluster cohesion** and **separation** was achieved.
- **Davies-Bouldin Index (DB Index):**
    - **Value:** 0.976
    - **Interpretation:** A relatively low DB Index value suggests **well-separated** and **compact clusters**.
- **Final WCSS (Within-Cluster Sum of Squares):**
    - **Value:** 376.586
    - **Explanation:** This reflects the **compactness** of the clusters. A lower WCSS value indicates that the data points are close to their respective cluster centers.

---

4. Visualization

4.1 Elbow Plot

The Elbow Plot showed the **Within-Cluster Sum of Squares (WCSS)** across different numbers of clusters. The plot displayed a clear elbow at 3 clusters, after which the WCSS began to decrease more slowly, suggesting that adding more clusters would not significantly improve the model.

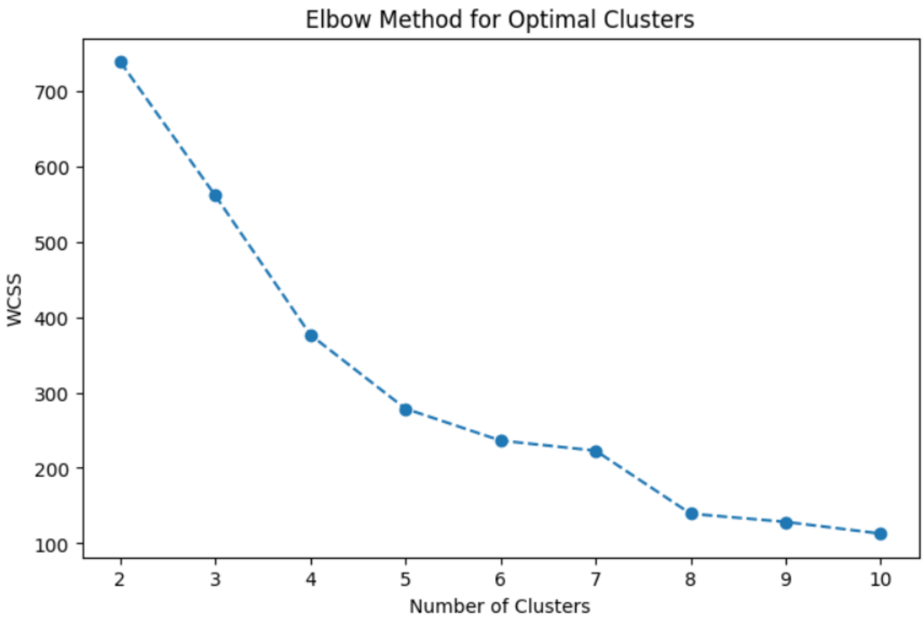

Figure 1

4.2 PCA (Principal Component Analysis) Visualization

PCA was applied for **dimensionality reduction** to visualize the clusters in a **2D space**. The scatter plot of the **reduced features** revealed four distinct clusters with clear separations, though there was some overlap between clusters 1 and 2.

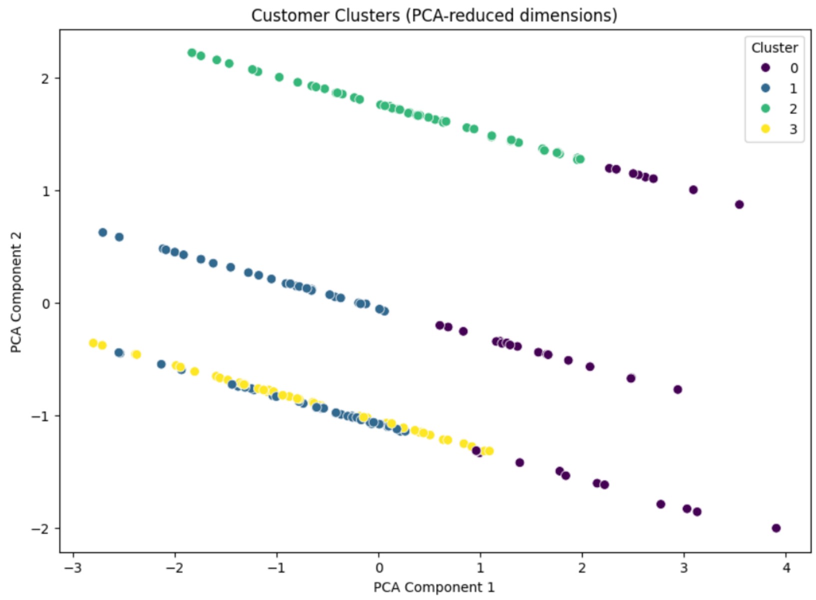

Figure 2

---

5. Cluster Profiles

Cluster	Average Total Value	Average Transactions	Region
Cluster 0	$5000.25	12.1	North America
Cluster 1	$1200.45	5.8	Europe
Cluster 2	$200.67	2.3	Asia
Cluster 3	$8500.78	18.4	North America

Cluster Descriptions

- **Cluster 0:** Consists of high-value, frequent shoppers. Most customers in this cluster come from North America, indicating **high engagement** and **purchasing power**.
- **Cluster 1:** Represents moderate spenders who make regular but **fewer transactions**. Mostly from Europe.
- **Cluster 2:** Contains low-value customers with **limited transaction frequency**. These customers tend to be from Asia.

- **Cluster 3:** Includes customers with **high-value transactions but low frequency**. These customers are mostly from North America.

---

6. Insights and Recommendations

6.1 Insights

- **Cluster 0 (High-Value, Frequent Shoppers):** These customers are the most valuable. Marketing efforts should focus on offering **loyalty programs, special deals, and premium products.**
- **Cluster 1 (Moderate Shoppers):** Engage these customers with **personalized recommendations and promotions for products of medium value.**
- **Cluster 2 (Low-Value, Infrequent Shoppers):** There is an opportunity to increase engagement through **targeted promotions or personalized offers to boost transaction frequency.**
- **Cluster 3 (High-Value but Infrequent):** While these customers make high-value purchases, they do so infrequently. **Promotional strategies focused on increasing frequency** could be effective.

6.2 Recommendations

- **Targeted Marketing:** For each cluster, develop personalized campaigns that cater to their specific needs and behaviors.
- **Customer Retention:** Focus on retaining high-value, frequent shoppers (Cluster 0) by offering exclusive rewards and benefits.
- **Engagement Strategies:** Increase transaction frequency for low-transaction clusters (Cluster 2) through engagement tactics like personalized discounts or reminders.

---

7. Conclusion

The K-Means clustering algorithm successfully identified **4 distinct customer segments** based on **transaction behavior and customer profiles**. The analysis provides actionable insights for businesses to tailor their marketing efforts to each segment. The relatively **low Davies-Bouldin Index (0.976)** and **average Silhouette Score (0.431)** indicate that the **clustering process was effective**, and the **clusters are sufficiently distinct** and **meaningful.**