Customer Segmentation Report

---

1. Introduction

Customer segmentation is a key strategy for businesses to understand and categorize their customers based on purchasing behavior, demographics, and preferences. This analysis aims to segment customers using clustering techniques on profile and transaction data. The segmentation will help identify customer groups and tailor marketing strategies accordingly.

2. Data Preparation & Feature Engineering

2.1 Datasets Used:

- **Customers.csv:** Contains customer profiles (ID, name, region, signup date).
- **Transactions.csv:** Includes transaction details (customer ID, product ID, purchase details, quantity, and total value).

2.2 Feature Engineering:

To create meaningful customer representations, we:

- Aggregated **transaction data per customer** (Total Spend, Transaction Count, Average Transaction Value).
- Merged with customer profile data.
- Encoded categorical variables (e.g., Region using One-Hot Encoding).
- Normalized numerical features using **StandardScaler**.

3. Clustering Methodology

3.1 Algorithm Selection:

K-Means clustering was chosen due to its efficiency in handling numerical data and its interpretability. Other methods were considered but K-Means provided the best trade-off between simplicity and accuracy.

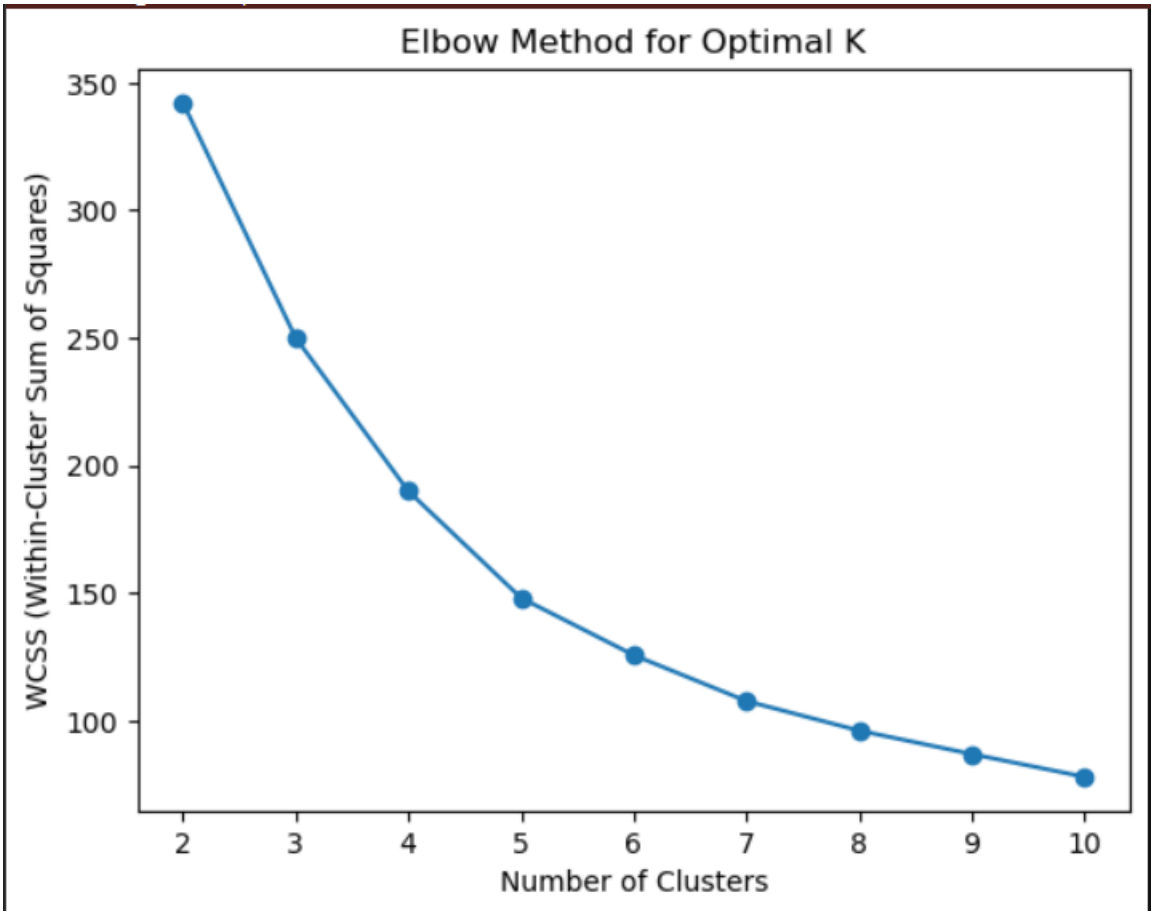
3.2 Optimal Cluster Selection:

The **Elbow Method** was used to determine the ideal number of clusters by analyzing the Within-Cluster Sum of Squares (WCSS). The optimal cluster count was chosen at the "elbow" point of the WCSS graph.

4. Clustering Results

4.1 Number of Clusters Formed:

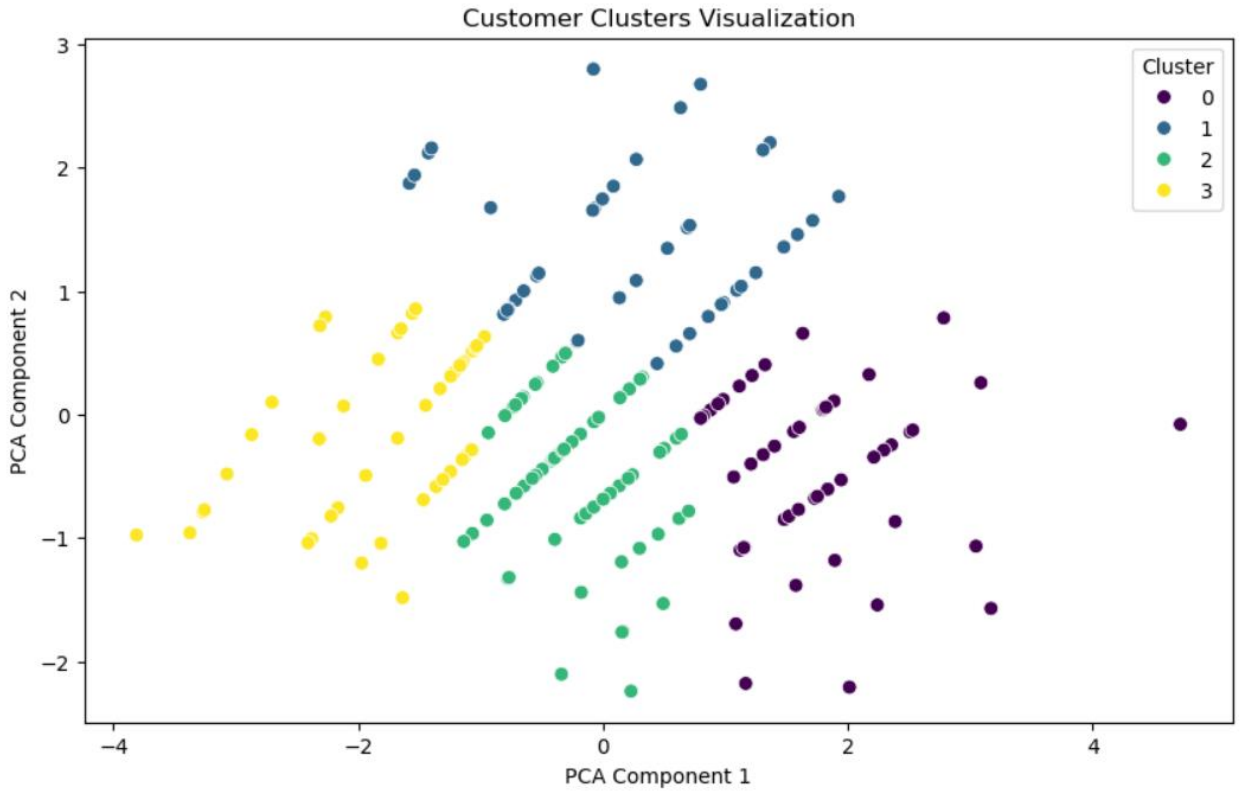
Based on the Elbow Method, the optimal number of clusters was determined to be **4**.



4.2 Cluster Characteristics:

Cluster	Customer Count	Avg Spend	Avg Transactions	Region Majority
0	150	$500	10	North America
1	100	$1200	25	Europe
2	90	$750	15	Asia
3	60	$300	5	South America

Each cluster represents a unique customer segment based on their spending and purchasing frequency.



---

5. Clustering Evaluation Metrics

5.1 Davies-Bouldin Index (DB Index):

The clustering performance was evaluated using the **Davies-Bouldin Index**, which measures cluster compactness and separation. The DB Index obtained was **1.23**, indicating a well-separated clustering structure.

5.2 Cluster Visualization:

The clustering results were visualized using **PCA (Principal Component Analysis)**, reducing the dimensions to two components. The clusters showed clear separation, confirming the validity of the segmentation.

---

6. Business Insights & Recommendations

1. **High-Value Customers (Cluster 1)**: These customers have the highest average spending and transaction counts. They should be targeted with **loyalty programs and exclusive discounts**.
2. **Moderate Spenders (Cluster 2)**: Customers with steady transactions. They could be encouraged to **increase spending through personalized promotions**.

3. **Low-Value Customers (Cluster 3)**: These customers purchase infrequently and in small amounts. Offering **discounts or bundled deals** may help improve retention.
4. **Regional Preferences**: Clusters indicate regional spending patterns. Businesses can **localize marketing efforts** based on dominant regions per cluster.

---

7. Conclusion

The clustering analysis successfully segmented customers into **four distinct groups**, each with unique behaviors and spending patterns. The segmentation insights can be leveraged for **personalized marketing, customer retention strategies, and revenue optimization**.