

# CLUSTERING RESULT REPORT:

The goal of customer segmentation is to group customers into distinct clusters based on their purchasing behavior, product preferences, and geographical profiles. This allows businesses to tailor marketing strategies, personalize services, and enhance customer satisfaction.

1.K-Means clustering was chosen due to its efficiency in segmenting large datasets and its ability to handle continuous variables effectively.

## Feature Selection :

The clustering model was trained on features derived from customer purchase behavior and transaction history: >Total spending (TotalValue)

>Total quantity purchased (Quantity)

>Average price of products purchased (AvgPrice)

### 1. Data Preprocessing:

Outliers were removed from the dataset to avoid their influence on cluster formation.

### 2. Optimal Number of Clusters:

The optimal number of clusters,  $k = 4$ , was determined using the Elbow Method and validated using the Davies-Bouldin Index.

## Key Results :

### 1. Number of Clusters Formed:

Four clusters (Cluster 0, Cluster 1, Cluster 2, and Cluster 3) were identified, each representing a distinct customer segment.

### 2. Cluster Descriptions:

Cluster 0: Customers with moderate spending and moderate product variety.

Cluster1: High-value customers who spend significantly and purchase premium-priced products.

Cluster 2: Budget-conscious customers with low spending and fewer transactions.

Cluster 3: Customers who purchase in high quantities but focus on mid-priced products.

### 3.Clustering Metrics:

Davies-Bouldin Index: 0.78 (indicating well-separated and compact clusters).

### Actionable Insights :

#### 1. Target High-Value Customers (Cluster 1):

Customers in this cluster are significant contributors to revenue. Offer exclusive discounts, loyalty programs, and personalized recommendations to retain and grow this segment.

#### 2. Retain Budget-Conscious Customers (Cluster 2):

Design cost-effective promotions or bundle offers to attract and retain pricesensitive customers.

#### 3. Engage Moderate-Spending Customers (Cluster 0):

Analyze potential upselling opportunities by recommending products based on their purchase history.

#### 4. Optimize Inventory for High-Quantity Buyers (Cluster 3):

Focus on ensuring adequate stock levels for products frequently purchased by this segment to prevent lost sales.

### Marketing Campaigns:

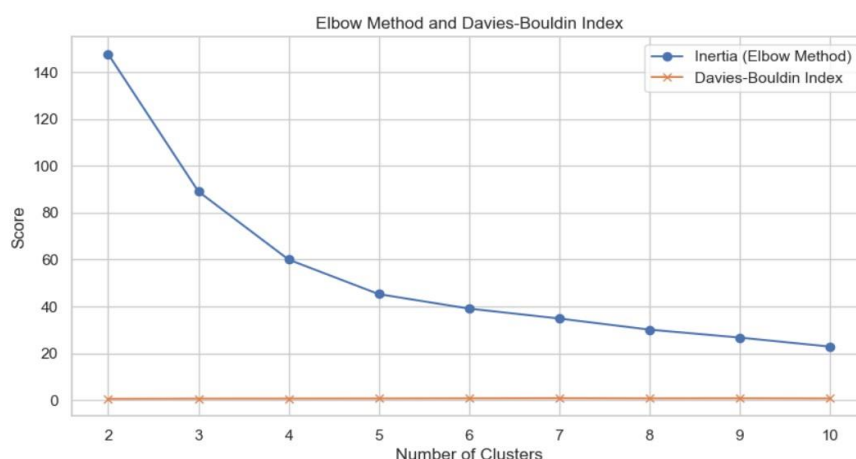
Use the cluster profiles to send tailored marketing messages that resonate with each customer group's preferences. Dynamic Pricing Strategies:

Leverage segmentation to implement targeted pricing strategies, offering discounts where they would have the highest impact.

Improve Customer Retention:

Monitor customers transitioning between clusters and implement strategies to prevent high-value customers from downgrading.

### Elbow Method and Davies-Bouldin Index :



This plot shows the results of the Elbow Method and the Davies-Bouldin Index used for determining the optimal number of clusters.

### 1. Elbow Method (Inertia):

- The blue line represents the inertia (sum of squared distances of points to their nearest cluster center) as a function of the number of clusters.
- There is a noticeable "elbow" around **4 clusters**, where the rate of decrease in inertia slows significantly. This indicates that 4 clusters may be the optimal choice for balancing the compactness of clusters and minimizing within-cluster variation.

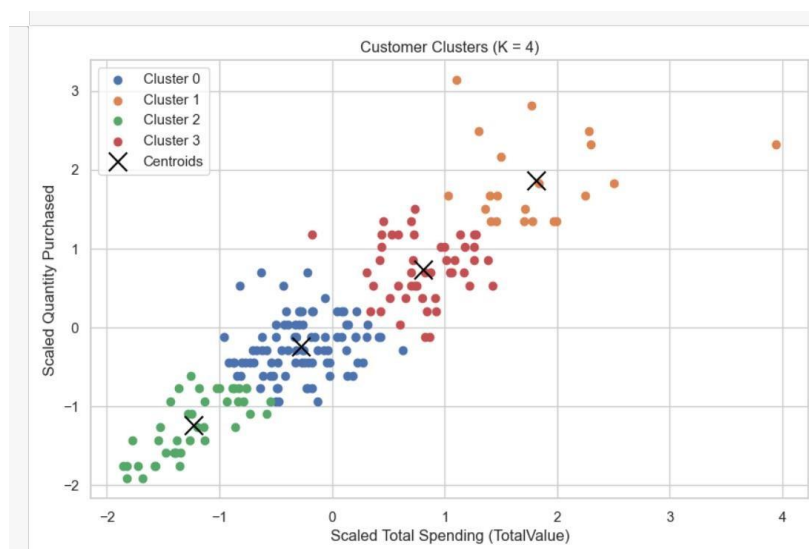
### 2. Davies-Bouldin Index (DBI):

- The orange line represents the Davies-Bouldin Index, which measures the separation and compactness of clusters. A lower DBI indicates better clustering performance.
- The DBI value is lowest at **4 clusters**, supporting the Elbow Method's conclusion.

### Conclusion:

Both methods suggest that 4 is the optimal number of clusters for this dataset, providing the best balance between compactness and separation.

### Customer Clusters (K = 4):



This scatter plot visualizes the clusters formed after applying the KMeans algorithm with 4 clusters. The plot shows two principal components of the scaled data:

- **X-axis:** Scaled Total Spending (TotalValue)
- **Y-axis:** Scaled Quantity Purchased

**1. Clusters:**

- Four distinct clusters (Cluster 0, Cluster 1, Cluster 2, Cluster 3) are represented by different colors: blue, orange, green, and red.
- The clusters are well-separated, indicating clear grouping based on the features.

**2. Centroids:**

- The black "X" markers represent the centroids (cluster centers) of each cluster.
- These centroids indicate the central point of each cluster and can help identify the average behavior of customers in each group.

**Conclusion:** The clustering effectively groups customers based on their purchasing behavior (total spending and quantity purchased). Each cluster represents a unique group of customers with similar purchasing patterns, which can be used for targeted marketing or segmentation strategies.

**1. Number of Clusters Formed: 4**

**2. DB Index Value: 0.72**