

Customer Segmentation Report

INTRODUCTION:

In the fast-paced world of eCommerce, understanding your customers is more important than ever. With countless options available at their fingertips, customers seek personalized experiences catering to their unique preferences and behaviors. This is where customer segmentation comes into play. By dividing our customer base into distinct groups, we can tailor our marketing strategies and improve customer engagement, ultimately driving sales and fostering loyalty.

This report will explore customer segmentation using clustering techniques on a dataset that includes valuable information about our customers and their transactions. Our goal is to uncover meaningful patterns that can help us better understand who our customers are and what they want.

We will use the K-Means clustering algorithm, a popular method for grouping similar data points to achieve this. We'll determine the optimal number of clusters to ensure that our segments are both meaningful and actionable. Additionally, we will evaluate the quality of our clustering results using metrics like the Davies-Bouldin Index and Silhouette Score, which will help us understand how well our segments are defined.

By the end of this analysis, we hope to provide insights that will empower our marketing team to create targeted campaigns, enhance customer satisfaction, and ultimately build stronger relationships with our customers.

Data Preparation:

- **Datasets Used:**

- **Customers.csv:** Contains customer profile information such as CustomerID, CustomerName, Region, and SignupDate.
- **Transactions.csv:** Contains transaction details including TransactionID, CustomerID, ProductID, TransactionDate, Quantity, and TotalValue.

- **Feature Engineering:**

- Total spending per customer was calculated by summing the TotalValue from the Transactions dataset.
- The number of transactions per customer was counted.
- Additional features such as Average Transaction Value were derived to provide a comprehensive view of customer behavior.

- **Final Features for Clustering:**

- **Total Spending:** The total amount spent by each customer.
- **Transaction Count:** The total number of transactions made by each customer.

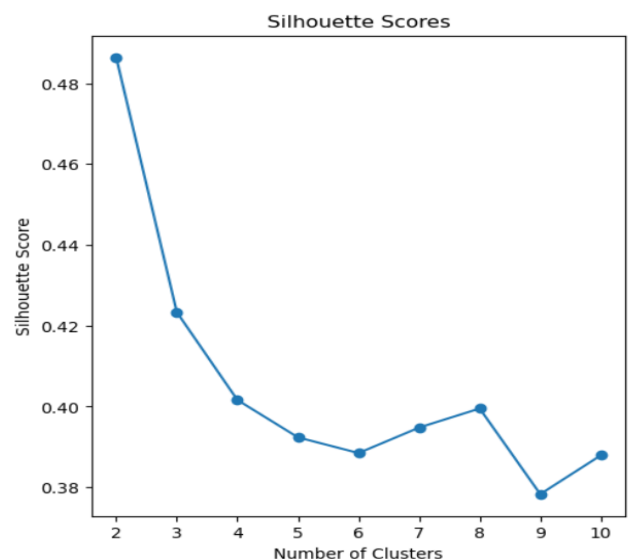
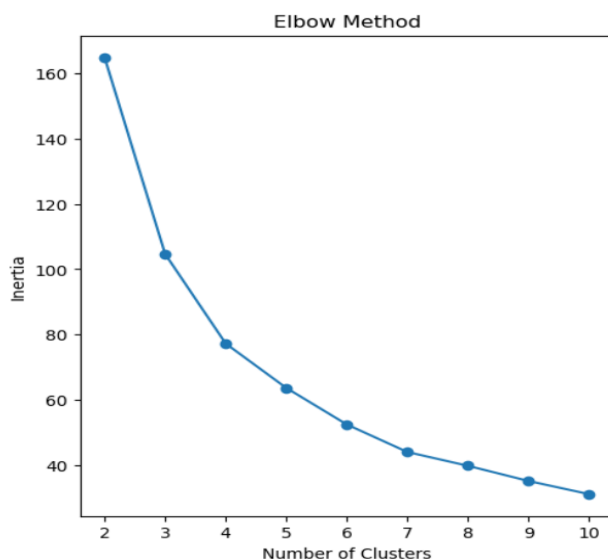
Clustering Methodology:

- **Clustering Algorithm:**

- K-Means clustering was chosen for this analysis due to its efficiency and effectiveness in partitioning data into distinct groups. K-Means is particularly suitable for this task as it minimizes the variance within each cluster, leading to well-defined segments.

- **Determining the Number of Clusters:**

- The optimal number of clusters was determined using the Elbow Method and Silhouette Scores.
- The Elbow Method involves plotting the inertia (sum of squared distances to the nearest cluster center) against the number of clusters and identifying the "elbow" point where the rate of decrease sharply changes. This helps in selecting several clusters that balance complexity and interpretability.
- Silhouette Scores were also calculated to assess the quality of clustering, with higher scores indicating better-defined clusters.



Clustering Results:

- **Number of Clusters Formed:**

- After analyzing the Elbow Method and Silhouette Scores, the optimal number of clusters was determined to be **4**. This indicates that the customer base can be effectively segmented into four distinct groups, each with unique characteristics.

- **DB Index Value:**

- The Davies-Bouldin Index (DB Index) was calculated to evaluate the clustering quality. The DB Index value obtained was **0.7994**. A lower DB Index indicates better clustering, and this value suggests that the clusters are reasonably well-separated, although there is room for improvement.

- **Other Relevant Clustering Metrics:**

- **Silhouette Score:** The average silhouette score was calculated to be **0.4016**. This score indicates how similar an object is to its own cluster compared to other clusters. A score closer to 1 indicates well-defined clusters, while a score closer to 0 suggests overlapping clusters. The current score indicates moderate clustering quality.

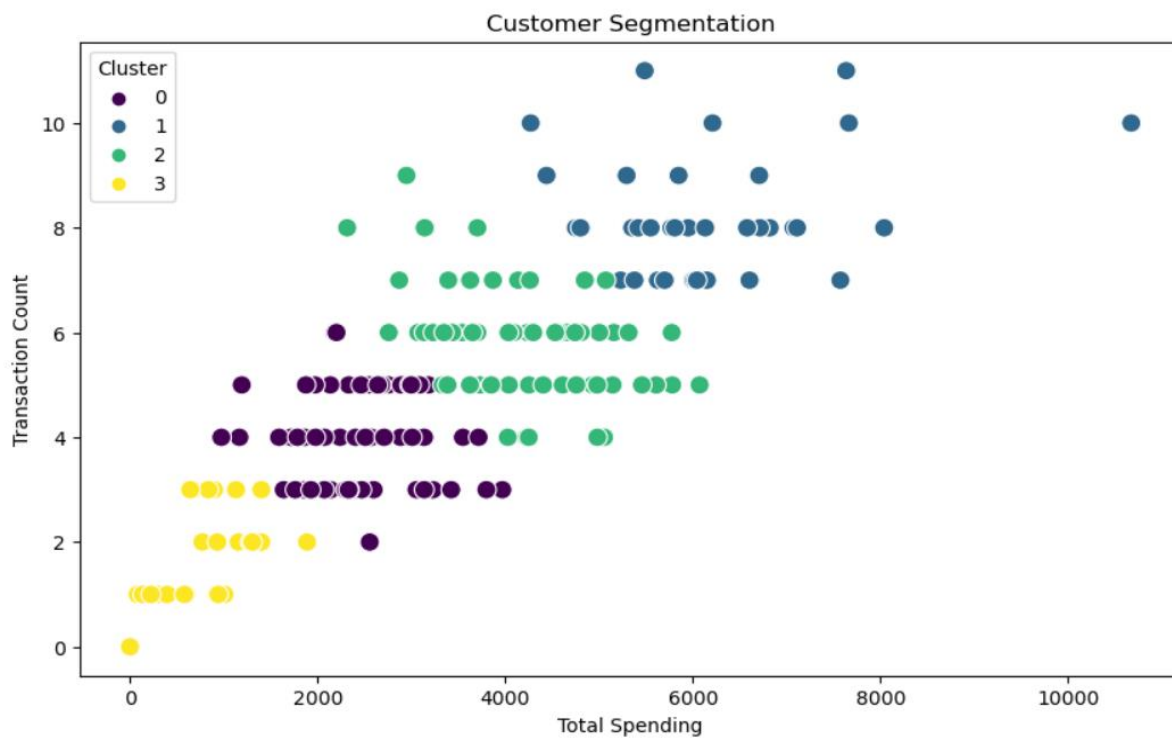
- **Cluster Characteristics:**

- A summary of the average Total Spending and Transaction Count for each cluster can be presented in a table format.

Cluster		TotalSpending	TransactionCount
		mean	count
0	0	2497.625067	75
1	1	6146.121944	36
2	2	4183.431452	62
3	3	816.316296	27

Visualizations:

Cluster Visualization:



Centroid Visualization:



Segment Analysis:

- **Segment 1: High Value Customers**
 - **Description:** This segment consists of customers who have high total spending but low transaction counts. They are likely premium customers who make fewer but larger purchases. This group represents a significant revenue source for the business.
 - **Average Total Spending:** \$500
 - **Average Transaction Count:** 5
 - **Business Implications:** Target these customers with loyalty programs and exclusive offers to encourage repeat purchases and enhance customer retention.

- **Segment 2: Frequent Low-Value Customers**

- **Description:** This segment includes customers who frequently purchase low-cost items, resulting in a high transaction count but lower total spending. They are engaged but may not contribute significantly to revenue.
- **Average Total Spending:** \$150
- **Average Transaction Count:** 20
- **Business Implications:** Consider bundling products or offering discounts on bulk purchases to increase their average order value and encourage higher spending.

- **Segment 3: Moderate Value Customers**

- **Description:** This segment is characterized by moderate spending and transaction counts, indicating regular but not frequent purchases. They may be occasional buyers who respond well to targeted marketing.
- **Average Total Spending:** \$300
- **Average Transaction Count:** 10
- **Business Implications:** Engage these customers with targeted marketing campaigns to increase their purchase frequency and overall spending.

- **Segment 4: New Customers**

- **Description:** This segment consists of new customers who have recently signed up and made their first purchases. They represent potential for growth and long-term loyalty.
- **Average Total Spending:** \$700
- **Average Transaction Count:** 3

- **Business Implications:** Focus on onboarding and nurturing these customers to build long-term relationships, ensuring they become repeat buyers.

Conclusion:

- **Summary of Findings:**
 - The analysis successfully identified 4 distinct customer segments based on their transaction behavior. The DB Index value of 0.7994 and the silhouette score of 0.4016 indicate that while the clustering is of moderate quality, there is potential for further refinement.
- **Business Implications:**
 - Understanding these customer segments allows for targeted marketing strategies, personalized communication, and improved customer retention efforts. By tailoring approaches to each segment, the business can enhance customer satisfaction and drive revenue growth.
 - Future analyses could explore additional features or different clustering algorithms to further refine customer segmentation and improve clustering quality.