

Customer Cluster Analysis Report

Introduction

This report will analyze the customer segmentation using the K-Means clustering algorithm. It will try to find the unique customer groups that have distinct purchasing behaviors and tenure, thus enabling businesses to use strategies more precisely, enhance customer engagement, and eventually improve profitability. The analysis is conducted on the transaction data with behavioral attributes such as total spending, purchase count, average purchase value, total quantity, and customer tenure for the purpose of segmentation of the customer base into actionable clusters.

Methodology

1. Data Sources

The analysis employed the following datasets:

Customers.csv: It has customer-level information, including CustomerID, Name, Region, and SignupDate.

Products.csv: It contains product information, such as ProductID, Name, Category, and Price.

Transactions.csv: It captures transactional data, including CustomerID, ProductID, Quantity, and timestamps of transactions.

2. Data Preprocessing

Missing Values: Missing data points were handled using appropriate imputation methods.

Normalization: Numerical variables like Total Spent, Purchase Count, and Total Quantity were scaled to give equal weight in clustering.

Feature Engineering:

Total Spent: The total amount of money spent by a customer on all purchases.

Purchase Count: The number of transactions a customer has made.

Average Purchase Value: Total spending divided by the number of transactions.

Total Quantity: Total number of items purchased.

Customer Tenure: Time since the customer's signup date.

3. Clustering Technique

The K-Means algorithm was chosen for clustering because it is simple and effective for segmentation.

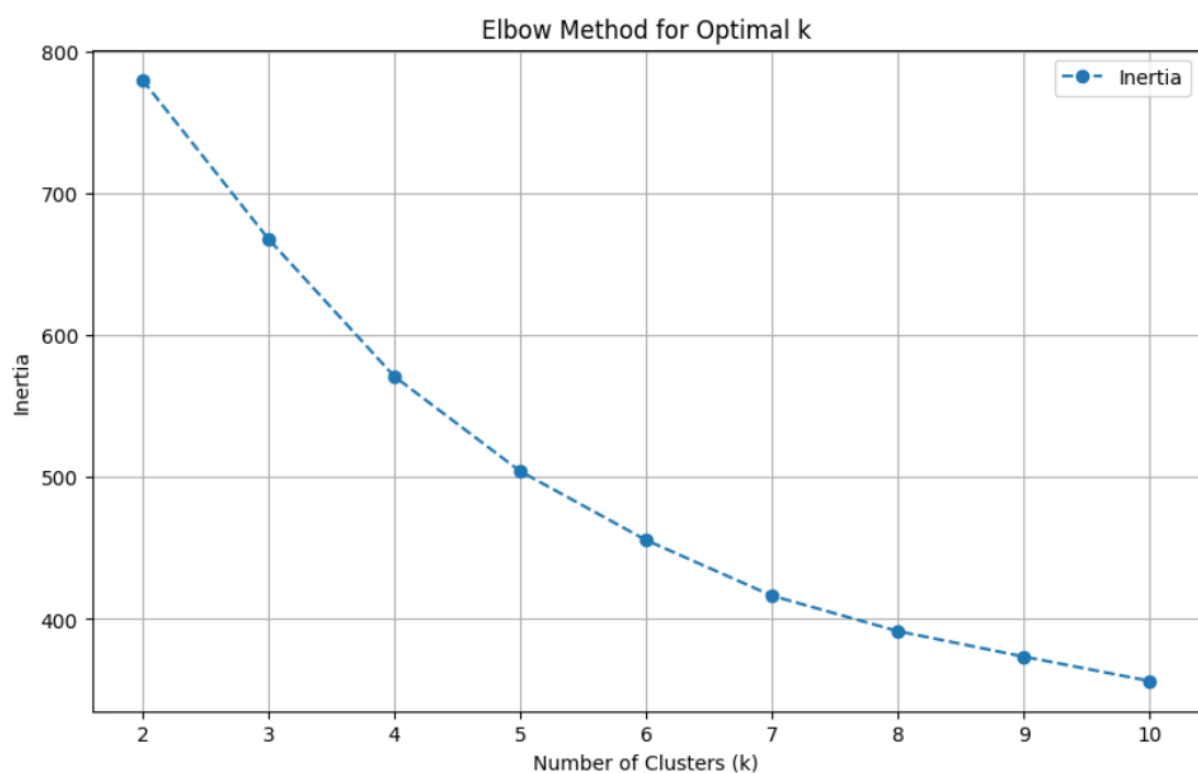
To find the optimal number of clusters, k , the following methods were used:

Elbow Method: This measures the reduction in inertia as the number of clusters increases.

Silhouette Score: This measures the separation and compactness of clusters.

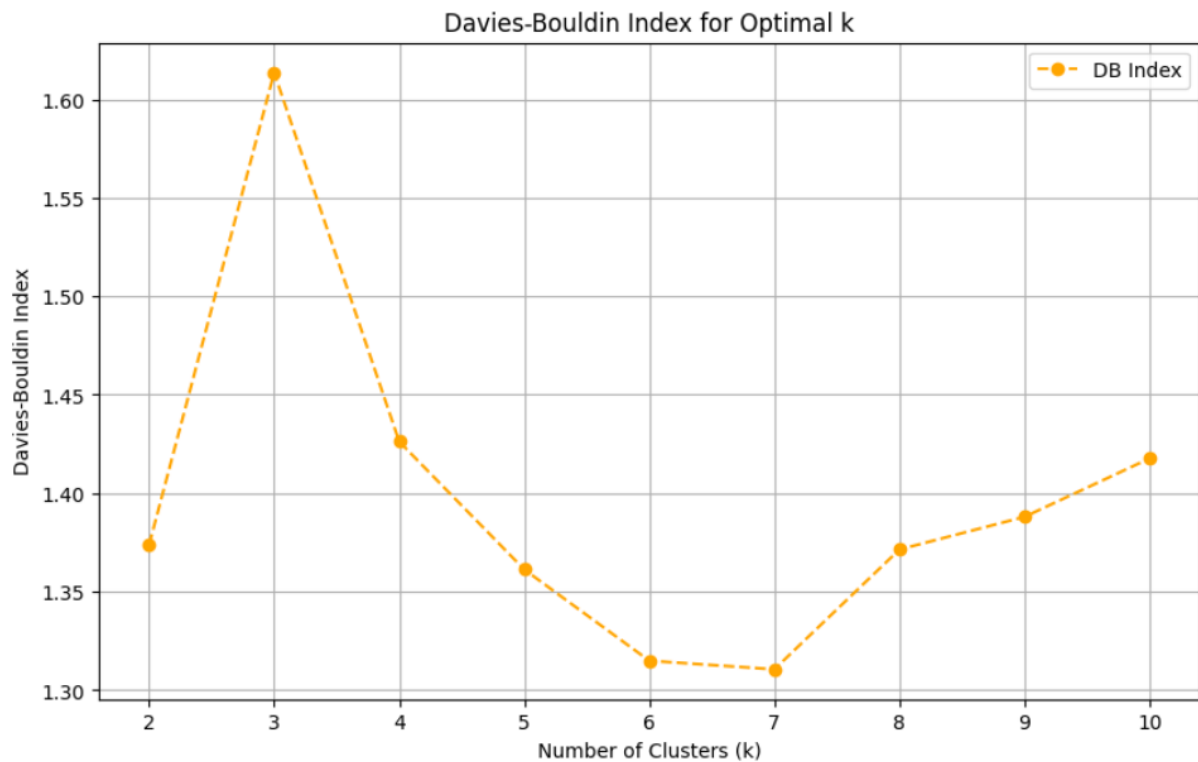
Davies-Bouldin Index (DBI): This evaluates clustering quality by comparing inter-cluster and intra-cluster distances.

1. Number of Clusters to be Optimized



Elbow Method

In the Elbow Method plot, the elbow is seen to exist at a point of a sharp decrease in inertia while k increases with the rate of further decrease gradually lowering. Thus, the optimum between complexity and clustering quality can be found at $k = 4$ or $k = 5$.



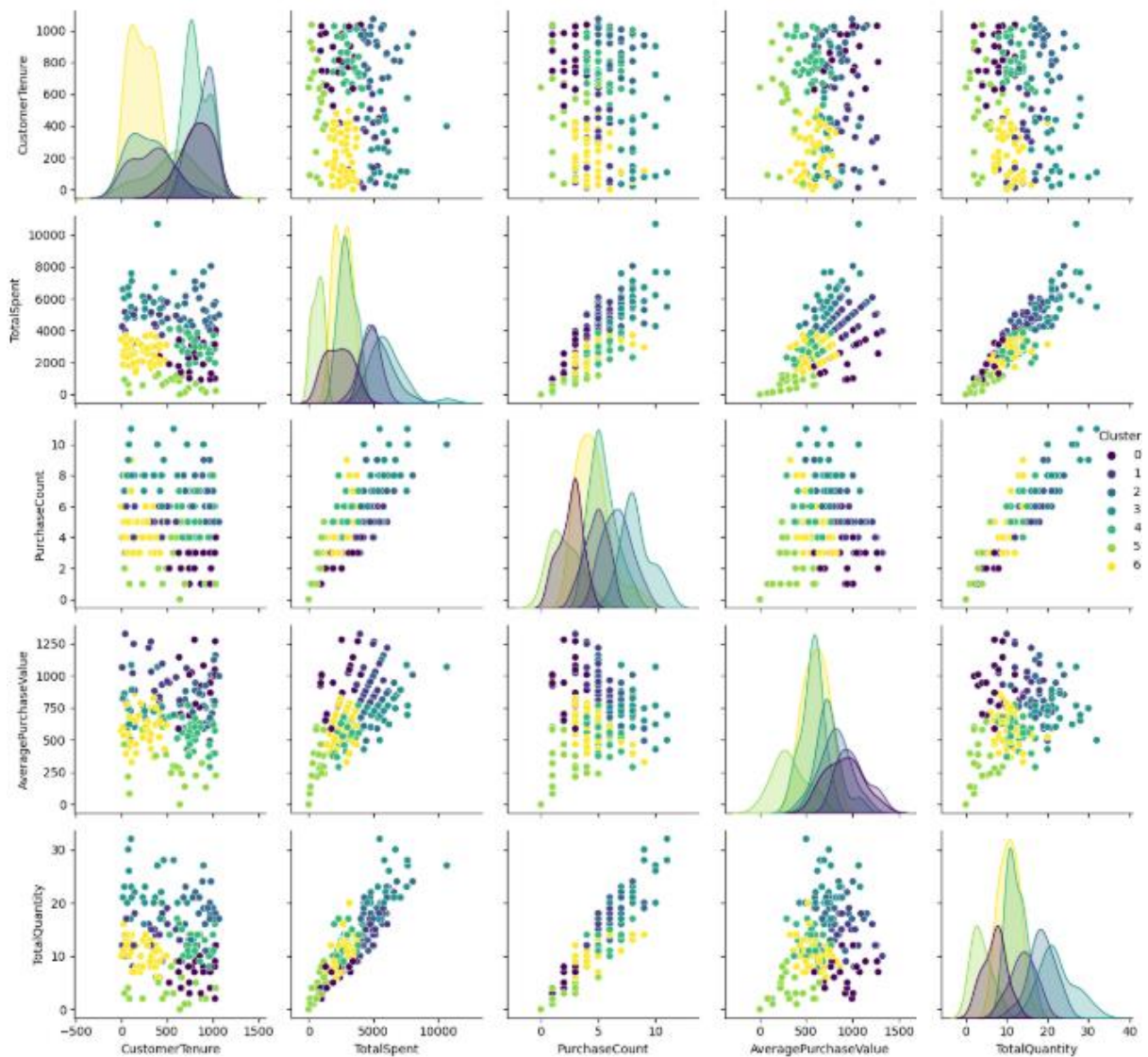
Davies-Bouldin Index

The DBI analysis indicates that the best clustering performance is achieved with $k = 7$, because at this value, the index is minimized, reflecting better separation and compactness of clusters.

2. Quality of Clustering

- Silhouette Score: A silhouette score of 0.218 means moderate quality of clustering. Although the clusters are not overlapping completely, it is still a meaningful segmentation for actionable insights.
- Davies-Bouldin Index: A DBI score of 1.31 means well-defined clusters with enough separation.

Pairplot of Customer Data with 7 Clusters



1. Customer Tenure

Distribution:

- Some clusters have customers with lower tenure (e.g., purple cluster), indicating newer customers.
- Other clusters (e.g., yellow and green) show customers with high tenure, representing long-term loyalty.

Insights:

- High-tenure clusters often correlate with higher total spending and purchase count.

Actionable Steps:

- Focus retention strategies (e.g., loyalty rewards) on clusters with high tenure to maintain long-term relationships.
- Create onboarding programs or promotions for low-tenure customers to increase their engagement.

2. Total Spent

Distribution:

- The expenditure patterns are very different in the clusters, with yellow and green clusters having much higher expenditure levels than the others.

Pairwise Correlation

- Total Spent and Purchase Count are positively correlated. That is, customers who make purchases more often also have a tendency to spend more.
- Some clusters reflect customers who spend a lot but make relatively fewer purchases. The visible outliers in green and purple clusters are some examples.

Actionable Steps:

- Target high-spending customers in specific clusters for premium offers or exclusive memberships.
- Identify and engage outliers to maintain their spending behaviors.

3. Purchase Count

Distribution:

- Some clusters (for instance, green and yellow) have a lot of purchases, while others (for instance, purple) have fewer transactions.

Insights:

- High-frequency buyers are good revenue streams. Such customers are usually found in clusters with high total spend and high tenure.

Actionable Steps:

- For frequent buyers, create bundle offers or multi-buy discounts.
- For low-frequency clusters, determine the reasons for fewer purchases and design specific campaigns to encourage repeat purchases.

4. Average Purchase Value

Distribution:

- Some clusters (purple) have high average purchase values but lower transaction counts.
- Others (yellow) are volume buyers with middle-of-the-road average purchase values.

Pairwise Correlation:

- In some clusters, there is an obvious trade-off: customers who make more frequent purchases have lower average purchase values.

Actionable Steps:

- Target high-margin goods to customers with high average purchase values.
- Personalize upselling for bulk buyers.

5. Total Quantity

Distribution:

- Clusters vary significantly as some are focused on high-quantity purchases-yellow and green clusters.

Insight

- Total quantity highly correlates with total spending. Bulk-buying clusters typically mean bulk revenue.

Action

- Promote quantity-based incentives such as "Buy More, Save More" promotions for high-quantity clusters.

Yellow Cluster:

- High-spending customers with good tenures and frequently buy.
- Probably a long-term customer, generating a high amount of revenue.

Green Cluster:

- They are frequently purchasing with a moderate spend. Therefore, they may be a bulk buyer or repeat customer.

Purple Cluster:

- They are low frequency in purchase but high average purchase value, which indicates customers like expensive, high-margin products.

Other Clusters:

- It includes varied customer behavior such as one-time buyers or new customers.

Opportunities:

High-Spending Clusters (Yellow/Green):

- Design exclusive loyalty programs to retain such profitable segments.
- Encourage sales of high-margin products or services.

Low-Spending Clusters (Purple):

- Barriers - perhaps variety and pricing
- Work these by selective advertising.
- Interplay between Attributes

Tenure vs Total Spent:

- There's a positive association but high and positively trending; retaining the customers has positive implications to increase revenue
- Customer is purchasing longer and more and thus the potential candidate for up sale.

Purchase Count vs Total Spent:

- Good association but that suggests buying again will add on well to revenues.
- Low frequency, high spenders: one-time buyers or infrequent big spenders

Average Purchase Value and Total Quantity

- Bulk buyer: low average purchase value; high total quantity
- Luxury shopper: high average purchase value; low total quantity.

Relationships Between Attributes

Tenure and Total Spent:

- here is a good positive relationship, which will emphasize the need to retain the customer for higher revenue.
- Long-term customers have a tendency to spend much more; hence, they are productive for upselling.

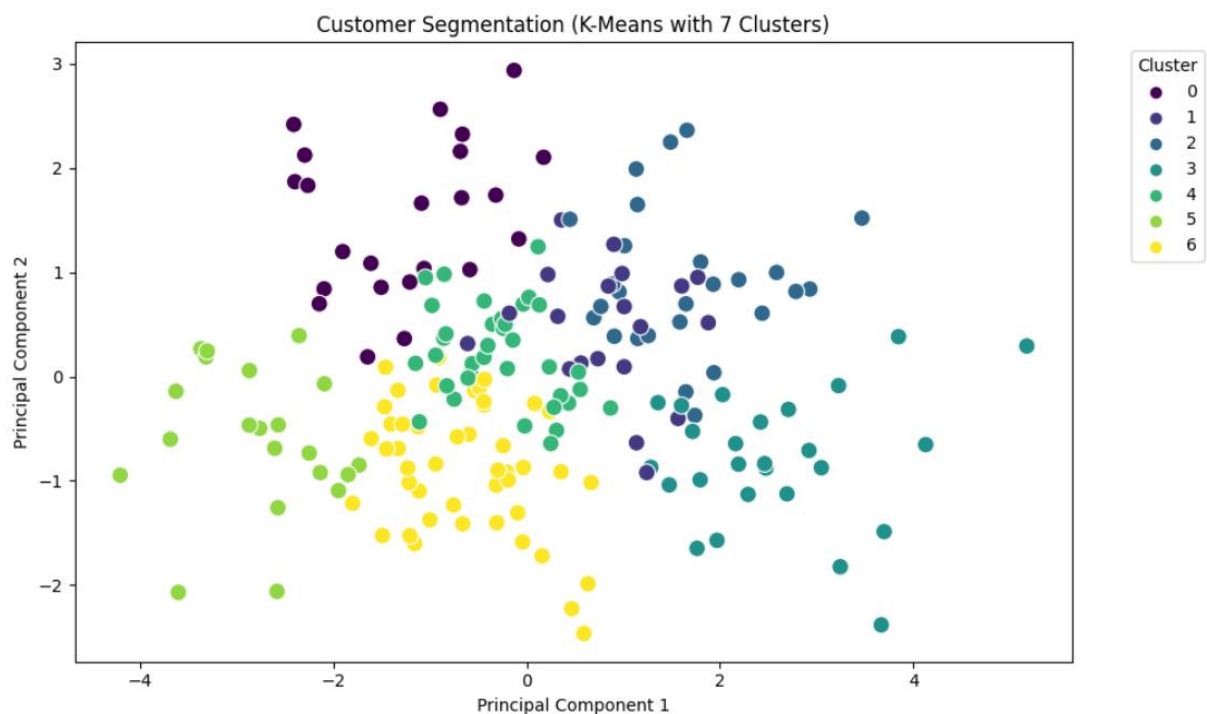
Purchase Count and Total Spent:

There is high correlation; frequent buyers are consistent revenue generators, while low purchase frequency but high spending in clusters may point toward one-time buyers or occasional big spenders

Average Purchase Value and Total Quantity:

Bulk buyers are opposite to luxury shoppers, who tend to have high average purchase values but low total quantity.

Visualization of Clusters:



1. PCA-Based Scatter Plot

- A 2D PCA visualization highlights the strong separation of clusters, with some overlap reflecting the natural diversity of real-world customers.



2. Behavioral Scatter Plot

A Total Spent vs. Purchase Count plot shows:

- Revenue is concentrated in high-spending, high-frequency clusters (Yellow, for example).
- Low-frequency, high-value clusters represent opportunities for targeted marketing efforts (Purple, for example).