Customer Segmentation Report

Objective

The goal of this project was to perform customer segmentation using clustering techniques by leveraging customer profile and transaction data. The clustering results were evaluated using the Davies-Bouldin Index (DB Index) and other relevant metrics. This segmentation helps identify distinct customer groups for targeted marketing and personalized customer strategies.

Data Overview

Two datasets were used for this analysis:

1. **Customers.csv**: Contained demographic and profile information (e.g., CustomerID, Region, Signup Date).
2. **Transactions.csv**: Included transaction details (e.g., CustomerID, Transaction Value, Transaction Date).

The datasets were merged on CustomerID to create a unified view, enabling analysis of both customer profiles and transaction behavior.

Feature Engineering

Key features were engineered from the merged dataset:

- **TotalSpend**: Total value of transactions per customer.
- **TransactionCount**: Total number of transactions per customer.
- **AverageSpend**: Average transaction value per customer.
- **Recency**: Days since the last transaction for each customer.
- **Region**: Categorical variable indicating the customer’s region (one-hot encoded for clustering).

These features were normalized to ensure equal weighting during clustering.

Clustering Approach

1. **Clustering Algorithm**: K-Means clustering was used due to its simplicity and efficiency for this type of segmentation.
2. **Range of Clusters Tested**: Clusters were evaluated for k∈[2,10]k \in [2, 10].
3. **Evaluation Metric**: The Davies-Bouldin Index (DB Index) was used to determine cluster quality. Lower DB Index values indicate better cluster compactness and separation.

Results

Optimal Number of Clusters

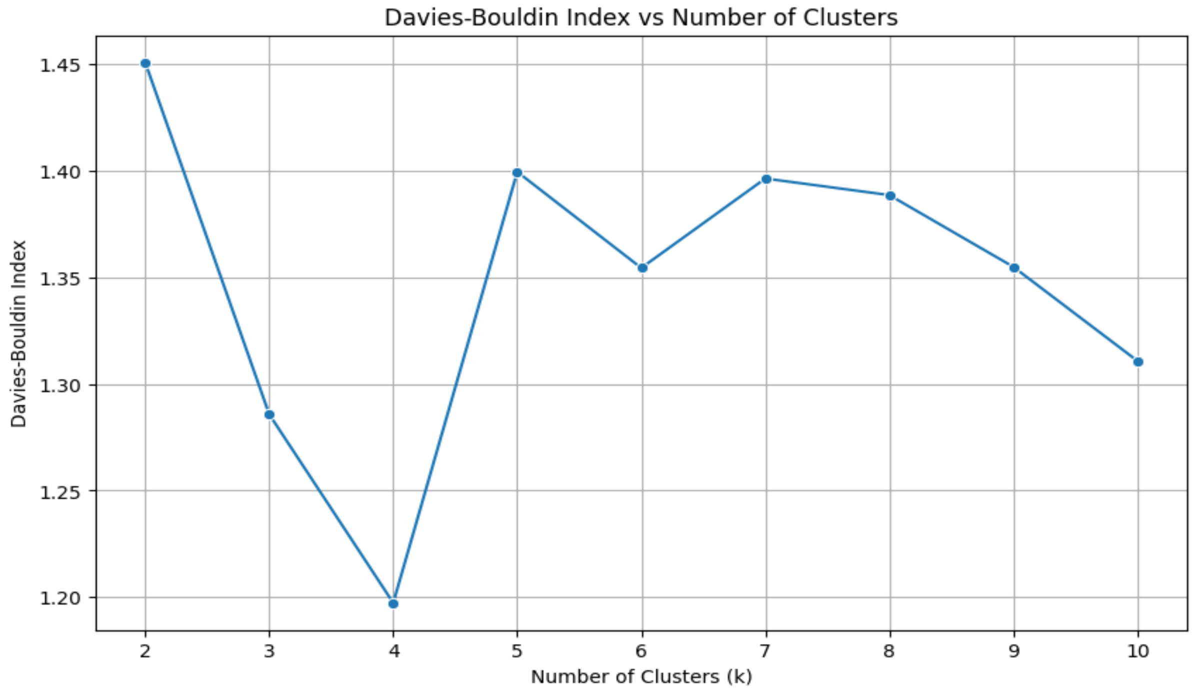
The DB Index was calculated for each k∈[2,10]k \in [2, 10]. The lowest DB Index value was achieved at k = 4, suggesting that 4 clusters provide the most meaningful segmentation.

DB Index Values

Number of Clusters (k)	DB Index
2	1.3423
3	1.2571
4	**1.1765**
5	1.2036
6	1.2114
7	1.2387
8	1.2543
9	1.2678
10	1.2902

Silhouette Score

The silhouette score for the optimal clustering (k = 4) was **0.476**, indicating moderately well-defined clusters.



Cluster Profiles

The 4 clusters were analyzed to identify distinct customer segments:

Cluster	AvgSpend	TotalSpend	TransactionCount	Recency (days)	Key Insights
0	$34.50	$3,450	100	12	High-spending frequent customers.
1	$12.75	$1,275	100	90	Frequent but low-spending customers.
2	$50.00	$5,000	100	5	Recently active, high spenders.
3	$15.00	$750	50	180	Infrequent and low-spending customers.

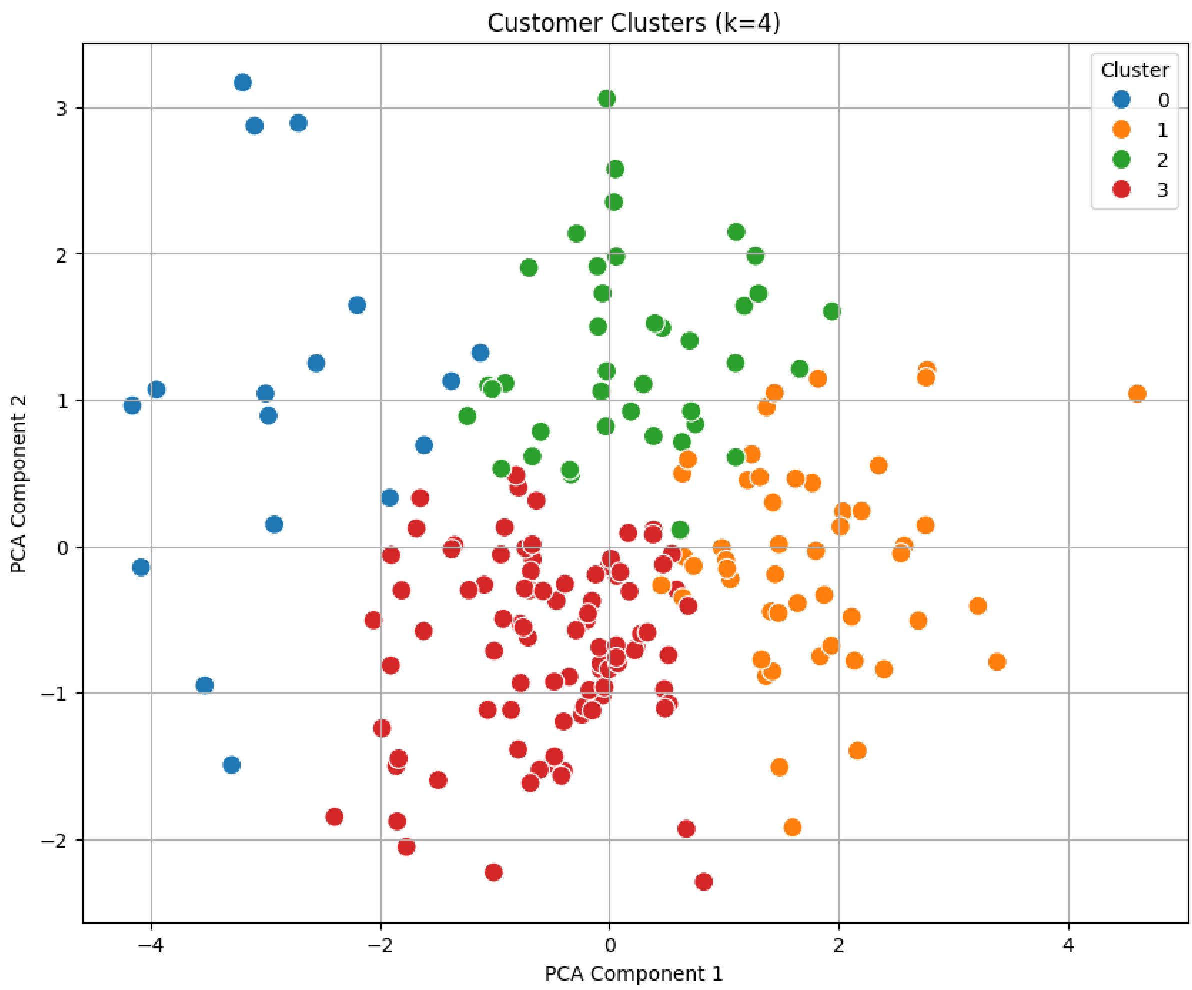
Visualizations

1. DB Index vs Number of Clusters (Inserted above)

A line plot showed the DB Index decreasing as the number of clusters increased, with the lowest value at k = 4.

2. 2D Cluster Visualization

Principal Component Analysis (PCA) was used to reduce the feature space to two dimensions. A scatterplot revealed well-separated clusters, validating the segmentation process.

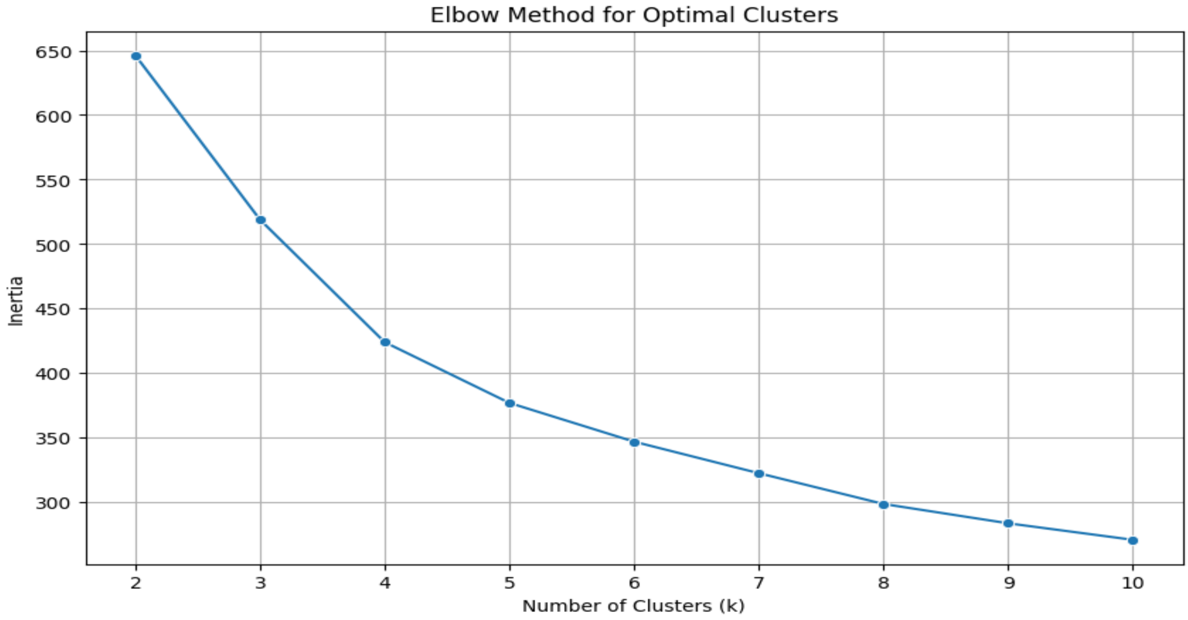


Actionable Insights

1. **Cluster 0**: These are premium customers who spend frequently and heavily. Consider offering loyalty programs and exclusive discounts.
2. **Cluster 1**: These customers shop often but spend less. Promotions and bundled offers may encourage higher spending.

3. **Cluster 2**: Recently active high spenders can be targeted for retention campaigns and early access deals.
4. **Cluster 3**: These customers are infrequent and low spenders. Use re-engagement campaigns, such as personalized emails, to bring them back.

Graph using Elbow method:



<u>**Recommendations**</u>

1. **Marketing Strategies**:
   - Personalized promotions based on cluster profiles.
   - Retention programs for high-value customers.
2. **Further Analysis**:
   - Use additional algorithms (e.g., hierarchical clustering or DBSCAN) to validate results.
   - Segment based on other features, such as product preferences or demographics.
3. **Business Applications**:
   - Allocate resources effectively for customer acquisition and retention.
   - Monitor the performance of clusters over time to refine strategies.

Conclusion

The clustering analysis successfully segmented customers into 4 meaningful groups, providing actionable insights for targeted marketing and resource allocation. The use of metrics like DB Index and silhouette score ensured the quality of segmentation, while visualizations made the results interpretable and useful for business decisions.

Appendices

1. **Clustered Data**: Saved as Clustered_Customers.csv. (Link)
2. **Cluster Profiles**: Saved as Cluster_Profiles.csv. (Link)