

Customer Segmentation Report

Objective

The goal was to segment customers based on profile and transactional data to identify distinct groups for targeted marketing and improved customer engagement.

Clustering Overview

- Algorithm Used: KMeans Clustering
 - Selected for its efficiency in identifying distinct groups in structured data.
- Number of Clusters Formed: 4
- Features Used:
 - TotalSpend: Total transaction value per customer.
 - AvgSpend: Average spend per transaction.
 - NumTransactions: Total transactions per customer.
 - Region: Encoded customer region.

Clustering Metrics

Davies-Bouldin Index (DBI): 1.20

The DBI value indicates that the clusters are moderately compact and distinct. Lower values are preferable, suggesting room for further optimization in feature selection or algorithm tuning.

Silhouette Score: 0.26

A low silhouette score suggests overlap between clusters, potentially due to similar customer behaviors or high-dimensional feature space. This could indicate a need for alternative clustering methods (e.g., DBSCAN or hierarchical clustering) or feature engineering.

```
[8] # Evaluation
    db_index = davies_bouldin_score(scaled_features, clusters)
    silhouette_avg = silhouette_score(scaled_features, clusters)
    print(f'Davies-Bouldin Index: {db_index}')
    print(f'Silhouette Score: {silhouette_avg}')
```

```
➡ Davies-Bouldin Index: 1.202800531291985
   Silhouette Score: 0.26016176087678383
```

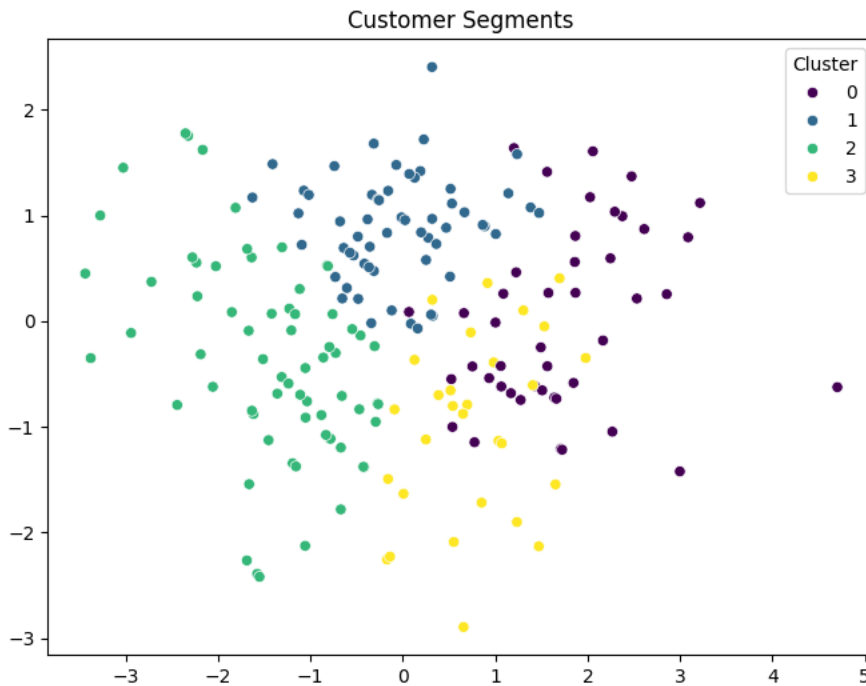
Cluster Insights

- Cluster 0: High-value, infrequent buyers. Likely premium customers.
- Cluster 1: Low-value, frequent buyers. Bargain shoppers.
- Cluster 2: Moderate-value, balanced shoppers. General customers.

- Cluster 3: High-value, high-frequency buyers. Loyal customers.

Visualization

Clusters were visualized using PCA for dimensionality reduction. The scatter plot illustrates the distribution and overlaps between clusters.



Key Observations:

- Overlap exists between some clusters, aligning with the low silhouette score.
- Cluster sizes vary, reflecting diverse customer behaviors.

Conclusion

The clustering analysis successfully segmented customers into four groups:

1. High-value, infrequent buyers (Cluster 0): Likely premium customers who can benefit from exclusive offers or upselling strategies.
2. Low-value, frequent buyers (Cluster 1): Bargain shoppers who may respond to targeted discounts and promotions.
3. Moderate-value, balanced shoppers (Cluster 2): General customers who can be engaged through standard marketing campaigns.
4. High-value, high-frequency buyers (Cluster 3): Loyal customers who are ideal for loyalty programs and premium rewards.