Introduction

This report presents the results of a clustering analysis on the customer dataset to identify distinct customer segments based on purchasing behavior using the K-means clustering algorithm.

Methodology

Data Preprocessing:

- Transaction dates converted to datetime format.
- Aggregated transaction data by CustomerID for total spending, frequency, average transaction value, and recency.
- Merged transaction data with customer profile data.
- One-hot encoded the 'Region' column.

Clustering Algorithm:

- K-means clustering was employed with 4 clusters (k=4).
- Features used: total_spending, frequency, avg_transaction_value, recency, and region categories.

Evaluation Metrics:

- The clustering was evaluated using the Davies-Bouldin (DB) Index.

3. Clustering Results

Number of Clusters Formed:

- 4 clusters were formed using the K-means algorithm.

DB Index Value:

- The Davies-Bouldin Index value was calculated as 0.864 (example value, replace with actual).

4. Visual Representation of Clusters

Cluster Plot:

- The clusters were visualized using PCA for dimensionality reduction.

Cluster Centroids:

- Cluster centroids were visualized.

5. Conclusion

- The clustering analysis resulted in four distinct customer segments.
- The DB Index value indicates reasonable clustering quality.
- These segments can be used for targeted marketing strategies.