Task 3: Customer Segmentation / Clustering

A report on your clustering results, including:

○ The number of clusters formed.

○ DB Index value.

○ Other relevant clustering metrics.

Customer Segmentation - Clustering Metrics Report

Number of Clusters

The analysis determined the optimal number of clusters through the Davies-Bouldin Index minimization approach, testing cluster ranges from 2 to 10.

- Optimal number of clusters: Variable (determined dynamically by the code)
- Selection method: Davies-Bouldin Index minimization

Clustering Implementation Details

Features Used in Clustering:

- Transaction-based features:
  - Transaction count
  - Total spend
  - Average transaction value
  - Total quantity
  - Average quantity
  - Customer lifetime
- Customer profile features:
  - Days since signup
  - Region (one-hot encoded)

Metrics

Primary Metric

- Davies-Bouldin Index (DB Index)
  - Calculated for cluster ranges 2-10
  - Lower values indicate better cluster separation
  - Final value: Generated dynamically by the code

Supporting Metrics

1. Distribution Metrics
   - Cluster sizes (count of customers per cluster)
   - Mean total spend per cluster
   - Mean transaction count per cluster
   - Mean customer lifetime per cluster
   - Mean days since signup per cluster

2. Standardization
   - All features standardized using StandardScaler
   - Ensures equal feature contribution to clustering

Technical Implementation Notes

- Algorithm: K-means clustering
- Random state: 42 (ensures reproducibility)
- Feature scaling: StandardScaler
- Missing values: Filled with 0
- Data preprocessing:
  - Datetime conversion for temporal features
  - One-hot encoding for categorical variables
  - Aggregation of transaction-level data

Note: The exact DB Index values and cluster distributions should be populated by running the code with your specific dataset.