Results

1. Number of Clusters:

- The analysis was performed with **5 clusters**.

- Each cluster represents customers with distinct transactional and profile patterns.

2. Davies-Bouldin Index:

- **DB Index:** 0.7451

3. Other Metrics:

- **Silhouette Score:** 0.4807

4. Key Metrics per Cluster:

The clustering process grouped customers based on the following aggregated metrics:

- **Total spending (total_spent):** Sum of all transactions made by a customer.

- **Average transaction value (avg_transaction_value):** Mean purchase value per transaction.

- **Total number of transactions (total_transactions):** Count of transactions made by a customer.

Methodology

1. Data Preparation:

- **Datasets Used:** Customers.csv and Transactions.csv.

- The customer dataset includes profile attributes such as **Region** and **Signup Date**.

- The transaction dataset provides details about purchase behavior.

- Aggregated transaction data into key metrics:

  o **Total number of transactions.**

  o **Total spending.**

  o **Average transaction value.**

- Merged customer and transaction data to prepare features for clustering.

2. Feature Selection and Scaling:

- **Selected Features:**

  o **Region:** Encoded as numeric using LabelEncoder.

  o **total_transactions:** Total number of purchases.

- o **total_spent:** Total value of all purchases.

- o **avg_transaction_value:** Average value per transaction.

- **Scaling:** Applied StandardScaler to standardize numerical features, ensuring equal contribution to clustering.

3. Clustering Algorithm:

- Applied **K-Means clustering** for customer segmentation.

.