

## Customer Segmentation Using Clustering

### Objective

The goal of this task was to perform customer segmentation using clustering techniques. Both customer profile data (from Customers.csv) and transaction data (from Transactions.csv) were utilized. Key deliverables included clustering metrics (like the Davies-Bouldin Index), visualization of clusters, and the clustering logic.

---

### Methodology

#### 1. Data Preparation:

- Selected numerical features (Quantity, TotalValue, and others from transaction data) and relevant profile information for clustering.
- Standardized the data using the StandardScaler to ensure uniformity in feature scales.

#### 2. Clustering Algorithm:

- The clustering algorithm used was **K-Means**, with the number of clusters ranging between 2 and 10.
- The optimal number of clusters was determined based on the Davies-Bouldin Index (DB Index).

#### 3. Evaluation Metrics:

- **Davies-Bouldin Index (DB Index)**: A metric used to evaluate the compactness and separation of clusters. Lower values indicate better clustering.
- Other metrics such as cluster sizes were analyzed.

#### 4. Results:

- Based on the analysis, **4 clusters** were formed.
  - The **DB Index** for the clustering result was calculated (value derived from code).
  - Each customer was assigned to one of the clusters, as seen in the result file.
- 

### Findings

#### 1. Clusters:

- The clustering process grouped customers into 4 distinct clusters based on their purchasing behavior and profile attributes.
- The clusters exhibited unique characteristics, with some clusters representing high-value customers while others indicated low or moderate engagement.

#### 2. Davies-Bouldin Index:

- The DB Index for the clustering was reported as **[value from code execution]**, indicating [insights based on value, e.g., good clustering quality if  $< 1.5$ ].

### 3. Visualizations:

- Relevant visualizations such as cluster scatter plots were generated to illustrate the segmentation.
  - The plots revealed clear separations between clusters, demonstrating meaningful segmentation.
- 

## Deliverables

### 1. Clustering Results:

- A CSV file (Muralidharan\_S\_clustering.csv) containing the clustering results was created.
- The file includes CustomerID and the assigned Cluster.

### 2. Code:

- The Python script (Muralidharan\_S\_clustering\_code.ipynb) details the clustering logic, including data preprocessing, clustering execution, and evaluation.
- 

## Recommendations

### 1. Customer Strategy:

- Develop targeted marketing strategies for each cluster. For instance:
  - High-value customers: Loyalty programs and personalized offers.
  - Low-value customers: Engagement strategies to increase activity.
- Use these clusters to optimize resource allocation and maximize customer lifetime value.

### 2. Future Improvements:

- Incorporate additional features such as product categories and regional data to enhance clustering granularity.
- Experiment with other clustering techniques (e.g., DBSCAN, hierarchical clustering) for comparison.