

## Task 3: Customer Segmentation/Clustering

### Methodology:

#### 1. Data Preparation:

- Combined customer data from Customers.csv and transaction data from Transactions.csv.
- Aggregated transaction data to calculate customer-level metrics: total transactions, total quantity purchased, total spending, and average order value.
- Normalized the numerical features to ensure they are on the same scale for clustering.

#### 2. Clustering Algorithm:

- Used **K-Means Clustering** with 4 clusters (determined experimentally).
- Evaluated the clustering quality using the **Davies-Bouldin Index**.

#### 3. Evaluation:

- **Davies-Bouldin Index:** 0.93 (indicating well-separated and compact clusters).
- Visualized clusters using PCA for dimensionality reduction to 2D.

#### 4. Visualization:

- The clusters were plotted in 2D PCA space to observe the separation and distribution of customer groups.

---

### Clustering Results:

#### 1. Number of Clusters:

- 4 clusters were formed, representing distinct customer segments.

#### 2. Davies-Bouldin Index:

- Achieved a value of 0.93. Lower values indicate better clustering performance (compact and well-separated clusters).

#### 3. Cluster Characteristics:

- Each cluster has unique traits based on spending habits, transaction frequency, and average order value:
    - **Cluster 0:** High-frequency buyers with moderate spending.
    - **Cluster 1:** Low-frequency buyers with high average order value.
    - **Cluster 2:** Moderate buyers with balanced metrics.
    - **Cluster 3:** High-spending customers with large quantities purchased.
-

## **Key Insights:**

### **1. Customer Segments:**

- The segmentation reveals distinct groups, which can be targeted with personalized marketing strategies.

### **2. High-Value Customers:**

- Customers in Cluster 3 contribute significantly to revenue and should be prioritized for retention programs.

### **3. Low-Engagement Customers:**

- Cluster 1 consists of customers with low transaction frequency. Incentives or promotions can increase engagement.

### **4. Balanced Cluster:**

- Cluster 2 represents an average customer base, which can be nurtured to transition into high-value customers.

### **5. Operational Strategy:**

- Use cluster insights to allocate resources effectively for customer retention, acquisition, and engagement.
-