

To meet the deliverables for your clustering results, We'll follow a structured approach for customer segmentation using clustering and evaluating the results based on various metrics like the Davies-Bouldin (DB) Index and visualizing the clusters effectively.

### Steps for Deliverables:

#### 1. Clustering Setup:

- Prepare data by aggregating transaction metrics for each customer (e.g., total spending, quantity purchased).
- Normalize features (using StandardScaler) to ensure that features with different scales do not dominate the clustering algorithm.

#### 2. Clustering Algorithm:

- Use **KMeans clustering** to segment customers into distinct groups based on their transaction behavior.
- The optimal number of clusters will be determined, but we will initially set the number of clusters to 5. This can be adjusted using the **Elbow Method** or **Silhouette Score**.

#### 3. Clustering Metrics:

- **Davies-Bouldin Index (DB Index)**: Measures the average similarity ratio of each cluster with the cluster that is most similar to it. A lower DB index indicates better clustering.
- **Silhouette Score** (optional): Provides insight into how well each data point lies within its cluster.
- **Within-cluster sum of squares (WCSS)**: The sum of squared distances between each point and the centroid of its cluster.

#### 4. Visualization:

- Use a scatter plot to visualize how the customers are distributed across clusters.
- Highlight the clusters using color coding and display the cluster centers.

#### 5. Reporting:

- Summarize the clustering results, the number of clusters, DB index, and other relevant metrics.
- Provide insights into the characteristics of each cluster based on the transaction behavior.

