

Clustering Results Report

Clustering was performed on customer transaction data to identify distinct segments using K-Means and Agglomerative Clustering. Various metrics were evaluated, including the **Davies-Bouldin Index (DBI)** and **Silhouette Score**, to assess the performance and suitability of the clusters.

Key Results

1. Number of Clusters Formed:

- Both K-Means and Agglomerative Clustering were applied with **5 clusters**, which offered a balance between compactness and separation of clusters.

2. Evaluation Metrics:

- **K-Means Clustering:**
 - **Davies-Bouldin Index (DBI):** 0.5254 (lower values indicate better-defined clusters).
 - **Silhouette Score:** 0.5366 (values closer to 1 indicate well-separated clusters).
- **Agglomerative Clustering:**
 - **Davies-Bouldin Index (DBI):** 0.5349.
 - **Silhouette Score:** 0.5298.

Interpretation: Both methods yielded similar results, but K-Means slightly outperformed Agglomerative Clustering in cluster compactness and separation.

3. Cluster Characteristics:

- **Cluster Sizes:**
 - K-Means: Cluster sizes were relatively balanced, with no extreme dominance by any one cluster.
 - Agglomerative: Slightly imbalanced cluster distribution was observed.
- **Features Influencing Clusters:**
 - Total spending, total quantity, and transaction count were the most significant factors driving clustering.
 - Regional preferences influenced segmentation.

4. Dimensionality Reduction:

- **PCA** was applied to visualize the clusters in 2D space.
- Clear separation was observed among clusters, confirming the robustness of the segmentation.

5. DBI Across Different Cluster Sizes:

- For K-Means:
 - DBI was lowest for 7 clusters (0.4493) and second lowest for 5 clusters (0.5254), suggesting **5 to 7 clusters** as optimal.
- For Agglomerative Clustering:
 - DBI was lowest for 10 clusters (0.4220), but practical interpretability may reduce at higher cluster counts.