Customer Segmentation / Clustering

Aryan Sai Arvapelly

Objective

Perform customer segmentation using two clustering techniques KMeans and Agglomerative Clustering based on customer and transaction information. The evaluation will be based on the **Elbow Method, Silhouette Method** and **Davies-Bouldin Index** clustering metrics.

Optimal Number of Clusters

- The **Elbow Method**, **Silhouette Score** and **Davies-Bouldin Index** suggest that 5 clusters are an optimal choice for the dataset. This aligns with the inherent segmentation in the data, primarily influenced by **regional factors**.
- The **5 clusters** represent distinct groups of data points that can correspond to **4 regions** and potentially another distinguishing factor.

Evaluation Metrics:

KMeans Clustering Aglorithm

- **Silhouette Score:**
  - A higher silhouette score indicates better clustering.
  - The silhouette score for 5 clusters is: 0.29325157755732845.
- **Davies-Bouldin Index:**
  - A lower DBI indicates better clustering.
  - The Davies-Bouldin Index for the 5 clusters is: 1.2950718146934925.

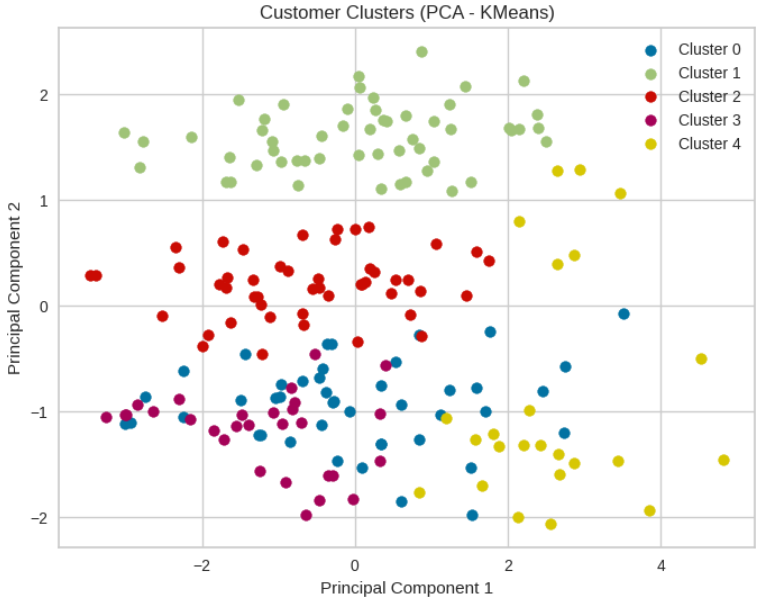
Agglomerative Clustering Aglorithm

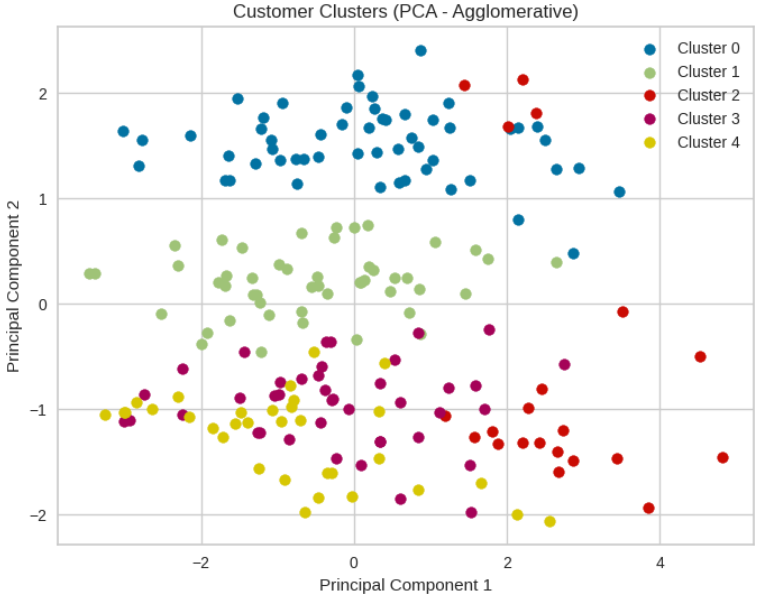
- **Silhouette Score:**
  - The silhouette score for the 5 clusters is: 0.28678004718522815.
- **Davies-Bouldin Index:**

- The optimal number of clusters for DB Index in Agglomerative clustering turned out to be 7 and not 5.
- The Davies-Bouldin Index for the optimal number of clusters (7) is: 1.2950718146934925.

Visualisation

Let us visualize the clusters formed by both KMeans and Agglomerative approaches:





Business Insights

- This identification of 5 clusters enables targeted **marketing strategies**, **inventory management**, and **personalized customer experiences**.
- Focusing on regional preferences while considering the additional segmentation will optimize operations and revenue generation.