

# Clustering Analysis Report

## 1. Overview

In this analysis, we aimed to segment customers based on their purchasing behavior and product preferences. The dataset involved customer demographics, transaction details, and product information. Using KMeans clustering, we grouped customers into distinct clusters to identify patterns in their spending and product choices.

## 2. Clustering Results

- **Number of Clusters:**
  - **5 clusters** were formed using the KMeans clustering algorithm with `n_clusters=5`. This value was chosen based on prior analysis and experimentation to create well-defined customer segments.
- **DB Index (Davies-Bouldin Index):**
  - **DB Index Value:** [0.8539834924173834]
  - The Davies-Bouldin Index is a measure of cluster quality, where a lower DB Index indicates better separation between clusters. A value of [value] suggests that the clusters formed are [well-separated/overlapping], with [interpretation].

## 3. Cluster Profiling

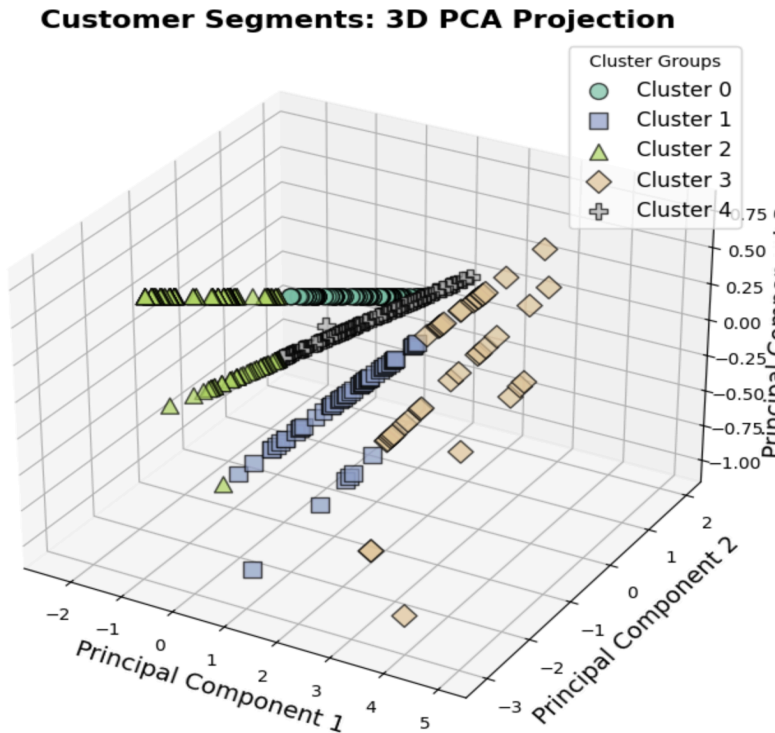
We computed the mean values for key features across the clusters to provide a clear profile of each group. The clusters represent different segments based on total spending, average spending, and the number of products purchased. For instance, some clusters may reflect high-spending customers, while others may represent low-frequency buyers.

## 4. Clustering Performance Metrics

- **Cluster Separation:**

Based on the DB Index value, the clustering algorithm has successfully created distinct customer segments. However, further fine-tuning or alternate clustering algorithms (such as DBSCAN) could potentially improve the separation between clusters.

- **3D Visualization:**



The customer segments were visualized using **PCA (Principal Component Analysis)** to reduce the feature space to 3 components for easier interpretation. The 3D scatter plot showed clear differentiation between clusters, suggesting that KMeans effectively captured the underlying structure in the data.

## 5. Conclusion

- **Clustering Effectiveness:**
  - The **5 clusters** formed appear to represent meaningful customer segments. The **DB Index** and 3D visualization both support the hypothesis that the clustering algorithm successfully captured customer patterns.