Clustering Report Summary

VASU GAUR

25/01/2025

1. Number of Clusters:

KMeans: The optimal number of clusters identified for KMeans clustering was 10 based on the Davies-Bouldin Index (DBI).

Gaussian Mixture Model (GMM): The optimal number of clusters identified for GMM clustering was 7, based on the lowest DBI.

2. Davies-Bouldin Index (DBI):

KMeans DBI: The DBI for KMeans clustering was 0.9481. A lower DBI indicates better clustering quality, suggesting the KMeans clusters are relatively well-separated and compact.

GMM DBI: The DBI for GMM clustering was 1.0537, which is slightly higher, suggesting that the GMM clusters have more overlap or are less distinct compared to KMeans.

KMeans: The clustering with 10 clusters captured more distinct customer segments, and the separation between clusters appears clearer in the visualizations (such as the scatter plots).

GMM: While the GMM clustering resulted in 7 clusters, the overlap between some clusters was slightly higher, as seen in the DBI value.

3. Cluster Visualizations:

KMeans clusters were visualized in both the original feature space (Total Spend vs Total Quantity) and in the reduced 2D PCA space. The KMeans clusters appeared more distinct in both visualizations.

GMM clustering was also visualized similarly, showing that the GMM clusters are less distinct but still provide useful customer segmentation insights.

4. Customer Segmentation:

The customer segmentation results from both KMeans and GMM can be used for targeted marketing and analysis of customer behavior based on transaction data and customer profiles.

Conclusion:

KMeans clustering produced better-defined clusters based on the Davies-Bouldin Index, with 10 clusters offering more separation and a lower DBI score.

- **GMM clustering**, while useful for modeling soft clusters, showed a slightly higher DBI and fewer clusters (7), indicating that KMeans might be more suitable for this dataset in terms of separation and compactness.