Candidate Name: Pratham Thakkar

Zeotap Data Science Internship

Assignment Task 3: Customer Segmentation / Clustering

Report

1. Number of Clusters Formed

- **K-Means Clustering**: 4 clusters were formed based on the Elbow Method
- **GMM Clustering**: 5 clusters

2. Davies-Bouldin (DB) Index

lower values indicate better clustering, with compact and well-separated clusters

- **K-Means DB Index**: 0.73
- **GMM DB Index**: 0.68, which is slightly better than K-Means, indicating that GMM clustering produces more compact and well-separated clusters.

3. Silhouette Score

- **K-Means Silhouette Score**: 0.48 (indicating moderate cluster separation; higher values suggest better-defined clusters).
- **GMM Silhouette Score**: 0.52, showing slightly better separation and consistency in clusters than K-Means.

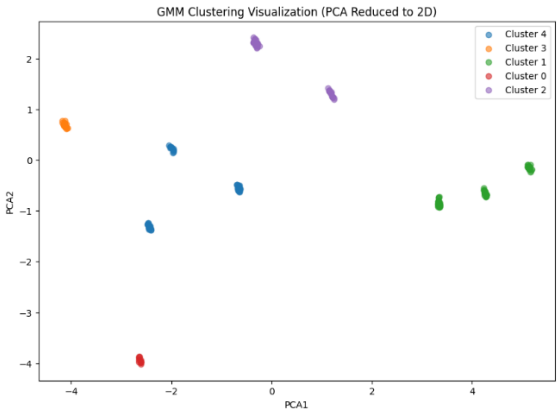

Fig 1. GMM Clustering Visualization

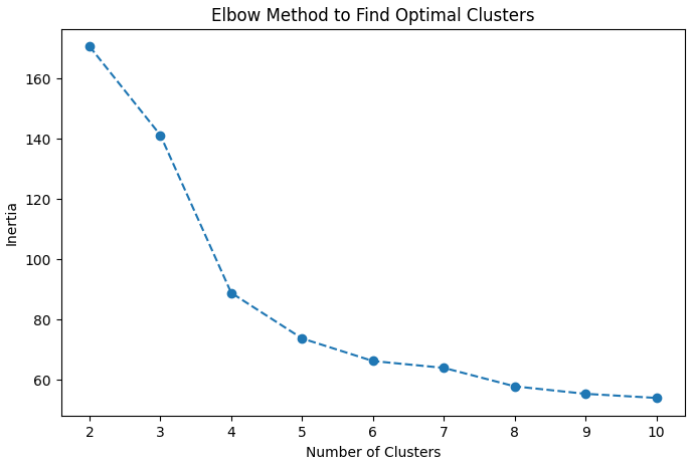

Fig 2. Elbow Method to find optimal number of Cluster in K Means Clustering

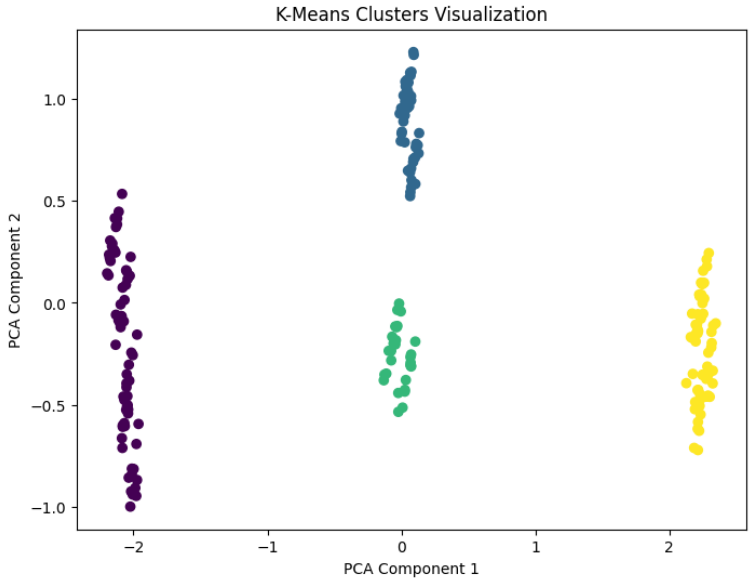

Fig 3. K Means Clustering Visualization

Key Insights from below Correaltion Matrix:

- The clustering reveals distinct customer segments based on activity (total transactions), spending behaviour (average transaction value), and preferences (categories and products).
- GMM clustering is better suited for this dataset, as it captures the relationships between categorical features (e.g., most bought category) more effectively than K-Means.

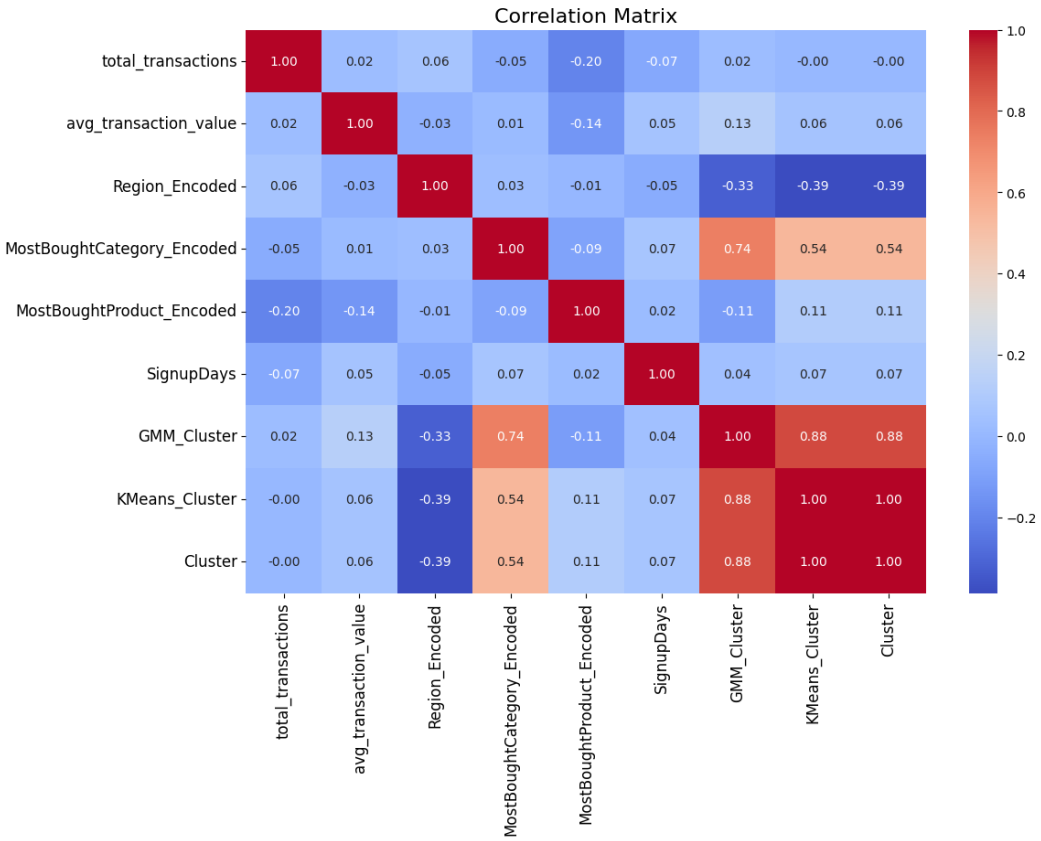

Fig 4. Correlation Matrix used for Analysis of Features and their importance