

Task 3: Summary

The code focuses on customer segmentation using clustering. Here's a brief summary

1. Imports and Preprocessing:

- I. Libraries like pandas, numpy, scikit-learn, and visualization tools are imported.
- II. Data is loaded from Customers.csv and Transactions.csv.

2. Data Preparation:

- I. Merging datasets based on CustomerID.
- II. Derived features include TotalAmount and TransactionFrequency.

3. Clustering Process:

- I. Feature scaling using StandardScaler.
- II. Iterative K-Means clustering (testing clusters from 2 to 10).
- III. Metrics like Davies-Bouldin (DB) Index and Silhouette Score are calculated for evaluation.

4. Cluster Visualization:

- I. PCA reduces dimensions for visualizing clusters in a 2D scatterplot.

5. Metrics and Optimal Clusters:

- I. Identifies the optimal number of clusters based on the DB Index.
- II. Visualizes clusters and prints the DB Index and Silhouette Score.