

Task 3: Customer Segmentation / Clustering

Step-by-Step Plan:

1. Load the Data:

Load the [Customers.csv](#), [Products.csv](#), and [Transactions.csv](#) files into dataframes.

2. Prepare the Data:

Merge the [Customers.csv](#) and [Transactions.csv](#) datasets to create a comprehensive dataset with customer profiles and transaction history.

Aggregate transaction data to create features for clustering (e.g., total transaction value, total quantity purchased).

3. Feature Engineering:

Create features that represent customer profiles and transaction history.

Normalize the features if necessary.

4. Perform Clustering:

Choose a clustering algorithm (e.g., K-means, DBSCAN) and determine the optimal number of clusters.

Calculate clustering metrics, including the DB Index.

Visualize the clusters using relevant plots.

Explanation of the Code:

Load the Data:

- The [Customers.csv](#), [Products.csv](#), and [Transactions.csv](#) files are loaded into dataframes.

Merge Datasets:

- The transactions dataframe is merged with the customers dataframe on CustomerID.
- Aggregate Transaction Data:
- Transaction data is aggregated to create features for clustering, such as total transaction value and total quantity purchased.

Feature Engineering:

- Features representing customer profiles and transaction history are created.
- The features are normalized using StandardScaler.

Perform Clustering:

- K-means clustering is performed with 5 clusters.
- The Davies-Bouldin Index is calculated to evaluate the clustering performance.
- The clusters are visualized using a scatter plot.

Save Clustering Results:

- The clustering results are saved to a CSV file (FirstName_LastName_Clustering.csv).