**K-Means Clustering Implementation Report**

**Data Preprocessing**

Loaded datasets: Customers.csv and Transactions.csv.

Merged the datasets on CustomerID.

**Feature Engineering Attempts**

Several feature engineering techniques were applied to improve clustering performance:

Created a new feature to calculate the difference between the signup date and the last transaction.

Constructed a new dataframe containing CustomerID, total money spent, transaction count, total items purchased, and customer region.

Applied one-hot encoding for the categorical variable Region and scaled numerical variables.

**Clustering**

Performed dimensionality reduction to reduce the number of features to three.

Used the reduced feature set to determine the optimal number of clusters.

Implemented Agglomerative Clustering for final clustering.

**Challenges and Further Work**

The clustering results are not satisfactory, as indicated by the Davies-Bouldin Index (DBI) and Silhouette Score.

To improve cluster separation, additional product-related features should be incorporated.

Focus needed on refining feature selection, exploring alternative clustering methods, and optimizing hyperparameters to enhance clustering performance.