

# Report on Clustering Results

In this report, I perform customer segmentation using clustering techniques and use both profile information from `Customers.csv` and transaction information from `Transactions.csv`. The clustering algorithm chosen is K-Means and determine the optimal number of clusters between 2 and 10. Evaluate the clustering results using the Davies-Bouldin Index and other relevant metrics.

## Data Preparation

First load the data from `Customers.csv` and `Transactions.csv` and merge them to create a comprehensive dataset that includes both customer profile information and transaction history.

## Feature Engineering

Extract and create features that represent the characteristics and behaviours of customers. The features including

- Total spend (`TotalSpend`)
- Average transaction value (`AvgTransactionValue`)
- Number of transactions (`NumTransactions`)
- Region (`Region`)

## Data Preprocessing

Encode categorical variables and scale numerical features to prepare the data for clustering. We use `StandardScaler` to scale numerical features and `OneHotEncoder` to encode categorical features.

## Clustering

Choose the K-Means clustering algorithm and determine the optimal number of clusters. We fit the K-Means model to the data and predict the cluster labels.

## Clustering Metrics

Calculate the clustering metrics:

- Davies-Bouldin Index (DB Index)
- Silhouette Score

Visualize the clusters using PCA to reduce the dimensionality of the data and create a scatter plot to visualize the clusters.

## Results

1. Number of Clusters Formed: 4
2. Davies-Bouldin Index: 1.342184356995347
3. Silhouette Score: 0.22808346780803854