

# Clustering Results Report

## Overview:

This model involves segmenting customers into groups based on their transaction and spending behavior. Using profile information from Customers.csv and transaction information from Transactions.csv, we employed clustering methods to determine useful customer segments. These segments enable companies to have a better understanding of their customers and allow them to target their marketing efforts.

Data Preparation: We blended customer profile information with purchase information to estimate:

Total Spend: The total spent by each customer.

Transaction Frequency: How many transactions each of the customers made.

Region: Encoded using one-hot encoding to handle categorical data.

Missing values were imputed to provide clean data for analysis.

Clustering: We implemented the K-Means clustering algorithm, which clusters customers according to the similarities in their spending and transactional patterns.

The number of clusters was determined to be 3. This implies that customers were separated into three groups based on their behavior.

Assessment: To quantify how good the clustering was, we computed the Davies-Bouldin (DB) Index, which is used to measure the quality of clustering results. The smaller the DB Index, the better the performance of the clustering.

The DB Index for this project was [Your DB Index Value Here], which meant [Your Interpretation, e.g., "moderately well-separated clusters"].

Visualization: In order to make the clusters more interpretable, we transformed the data into two dimensions with Principal Component Analysis (PCA) and visualized the clusters in a 2D scatter plot.

## Results:

Number of Clusters: We have found 3 clusters:

Cluster 0: Customers who are high spenders and make frequent transactions.

Cluster 1: Moderately spending customers with average frequency of transactions.

Cluster 2: Low spend and low frequency customers.

DB Index: The DB Index reading was [Insert Value], indicating that the clusters are [Insert Interpretation, e.g., "reasonably well-separated"] by their spending and transaction behavior.

Cluster Visualization: The scatter plot clearly shows three distinct groups of customers. Each group has unique characteristics, making it easier to target specific clusters with tailored marketing strategies.

### **Observations:**

Cluster 0 is characterized by high-value consumers who can be expected to be responsive to premium promotions or loyalty schemes.

Cluster 1 consists of average customers who can be offered discounts or upselling.

Cluster 2 is made up of low-spending customers who could need re-engagement campaigns to have their spending or frequency of transactions boosted.

### **Conclusion:**

This cluster analysis is highly informative in terms of customer behavior, which helps companies develop focused marketing strategies. Targeting high-value customers (Cluster 0) and re-targeting low-value customers (Cluster 2), companies are able to maximize their marketing opportunities and enhance customer satisfaction.

The paper proves the efficiency of applying clustering methods for customer behavior understanding and emphasizes the significance of clean data preparation and appropriate evaluation metrics such as the DB Index. The

results of clustering and visualizations can be the starting point for further decision-making and analysis.