Customer Segmentation Clustering Report

1. Introduction

Customer segmentation is a key step in understanding customer behavior and optimizing marketing strategies. In this report, we summarize the results of clustering analysis performed on customer transaction and profile data.

Objective

The goal is to segment customers into distinct groups based on their transactional and profile features, evaluate the quality of these clusters, and provide insights using clustering metrics and visualizations.

---

2. Data Overview

Datasets Used

- **Customers.csv**:

  Contains customer profile information, including CustomerID, Region, and SignupDate.

- **Transactions.csv**:

  Contains transactional information, including TransactionID, CustomerID, ProductID, Quantity, Price, and TotalValue.

- **Products.csv**:

  Contains product details, including ProductID, Category, and Price.

Key Features used for clustering

- **Total Spending**: Total amount spent by a customer.
- **Number of Transactions**: Frequency of customer transactions.
- **Average Spending per Transaction**: Average value of transactions per customer.
- **Days Since Last Purchase**: Time since the customer's last transaction.
- **Days Since Signup**: Time since the customer registered.

3. Clustering Methodology

Preprocessing

1. Recalculated rows with discrepancies in TotalValue (i.e., when TotalValue != Quantity * Price).
2. Aggregated customer-level features using transaction and profile data.
3. Standardized features using StandardScaler for uniform scaling.

Clustering Algorithm

- **Algorithm Used**: K-Means Clustering
- **Range of Clusters Tested**: 2 to 10 clusters.

Metrics for Evaluation

- **Davies-Bouldin Index (DBI)**: Used to measure the quality of clustering.

---

4. Results

This table shows the number of clusters and their DBIndex

S.No	NumClusters	DBIndex
0	10	1.141978
1	8	1.168368
2	9	1.171877
3	7	1.186953
4	5	1.187702
5	6	1.220782
6	4	1.287697
7	3	1.387530
8	2	1.541226