Customer Segmentation Clustering Report

This report summarizes the customer segmentation analysis using clustering techniques. The dataset includes customer profiles and transaction data. We used the K-Means clustering algorithm and evaluated clustering performance using the Davies-Bouldin Index (DB Index).

1. Data Overview

Customers.csv: Contains customer profile information (CustomerID, Name, Region, SignupDate).- Transactions.csv: Contains transaction details (CustomerID, ProductID, TotalValue, Quantity, TransactionDate).- Features derived for clustering:

* TotalValue: Total transaction value per customer.
* AverageTransactionValue: Average value of each transaction.
* TotalQuantity: Total quantity of products purchased.
* CustomerTenureDays: Duration between signup and the most recent transaction.

2. Clustering Process

We applied K-Means clustering on the scaled feature set. The number of clusters was varied from 2 to 10, and the optimal number of clusters was determined using the Davies-Bouldin Index.

3. Results and Metrics

Optimal Number of Clusters: 3 (based on DB Index).- Final DB Index: 0.93 (lower values indicate better clustering performance).- Silhouette Score: 0.65 (indicates good cluster separation).

4. Visualizations

Cluster visualizations were created using Principal Component Analysis (PCA) to reduce the features to 2D. The scatter plot shows clear separation among clusters.

5. Conclusion

The clustering analysis successfully segmented customers into distinct groups. These clusters can help in understanding customer behavior and tailoring marketing strategies. Future improvements could include experimenting with advanced clustering algorithms like DBSCAN or Agglomerative Clustering.