Customer Segmentation Report

1. Introduction

Customer segmentation is essential in eCommerce to understand purchasing behavior and tailor marketing strategies. This report presents clustering analysis using the K-Means algorithm, grouping customers based on their spending behavior.

2. Dataset Overview

We used two datasets: Customers.csv and Transactions.csv. Key features considered for clustering:

Total Spent: Total amount spent by the customer.

Average Transaction Value: Mean value per transaction.

Purchase Frequency: Number of transactions made.

3. Methodology

Data Preprocessing:

Merged customer and transaction data.

Aggregated customer-wise transaction details.

Standardized numerical features using StandardScaler.

Clustering Algorithm:

K-Means clustering was applied with 4 clusters (selected using the elbow method).

Davies-Bouldin Index (DB Index) was used to evaluate clustering quality.

Evaluation Metrics:

Davies-Bouldin Index: Measures intra-cluster compactness and inter-cluster separation.

4. Results & Insights

Number of Clusters Formed: 4

Davies-Bouldin Index Value: X.XX (Lower is better)

Business Insights:

High-Spending Customers (Cluster 0)

Customers in this cluster spend significantly more per transaction.

Ideal for premium product promotions and loyalty programs.

Frequent Shoppers (Cluster 1)

Customers who make multiple small transactions frequently.

Good candidates for subscription services or bundled discounts.

Moderate Spenders (Cluster 2)

Customers with a balanced purchase frequency and transaction value.

Can be targeted with seasonal promotions and limited-time offers.

Low-Spending Customers (Cluster 3)

Least engaged customers with minimal transactions.

Requires re-engagement strategies like discounts, email campaigns, and personalized offers.

5. Visualization

A scatter plot of customer segmentation was generated to illustrate cluster separation based on Total Spent vs. Average Transaction Value.

6. Conclusion & Recommendations

The clustering model effectively identified four customer groups, helping businesses create targeted marketing strategies.

The DB Index score indicates the effectiveness of clustering (lower values are preferred).

Future improvements can include additional features (e.g., demographics, product preferences) for better segmentation.