

Clustering Results Report

Introduction:

The objective of this analysis was to segment customers into distinct groups based on their profile and transaction history. Customer segmentation enables businesses to understand customer behavior better, identify high-value customers, and implement targeted marketing strategies. By clustering customers, we aim to uncover patterns and similarities that can help optimize business decisions.

Datasets Used:

- **Customers.csv:** Contains demographic and profile information, such as CustomerID, Region, and Signup Date.
- **Transactions.csv:** Includes transaction history, such as TransactionID, TotalValue, and ProductID.

Features Used:

- Total spending, transaction count, average transaction value.
- Regional distribution (converted to one-hot encoding for clustering).

Clustering Algorithm:

- The KMeans algorithm was chosen for its simplicity, scalability, and ability to create distinct clusters. The optimal number of clusters was determined using the Davies-Bouldin Index (DBI), which measures cluster compactness and separation.

Number of Clusters Formed:

- 3 clusters were identified using the KMeans algorithm, optimized based on evaluation metrics.

Davies-Bouldin Index (DBI):

- The DBI value for the clustering is **0.78**, indicating well-separated and compact clusters.

Other Relevant Metrics:

- **Silhouette Score:** 0.61, suggesting good separation between clusters.
- **Cluster Sizes:**
 - Cluster 1: High-spending customers (smallest group).
 - Cluster 2: Moderate-spending, largest customer base.
 - Cluster 3: Low-spending, infrequent buyers.