Customer Segmentation Report

1. Overview

Customer segmentation was performed using the **K-Means clustering algorithm** to group customers into distinct clusters based on their profile and transaction behaviors. The goal was to identify patterns in customer behavior for better targeting and personalized marketing strategies.

2. Methodology

Data Preparation:

Customer profile data from Customers.csv was merged with aggregated transaction data from Transactions.csv.

Aggregated transaction features included:

Total spending (total_spend).

Number of transactions (num_transactions).

Average transaction value (avg_transaction_value).

Feature Selection:

The clustering was performed on the features: total_spend, num_transactions, and avg_transaction_value.

Normalization:

The features were normalized using StandardScaler to ensure all variables contributed equally to the clustering process.

Clustering Algorithm:

The K-Means algorithm was used with the number of clusters (k) set to 5.

The clusters were identified based on patterns in spending, frequency of transactions, and average transaction values.

Evaluation:

Davies-Bouldin Index (DB Index) was used to evaluate the clustering quality. A lower DB Index value indicates better-defined clusters.

3. Results

Number of Clusters:

5 clusters were formed, each representing a distinct group of customers based on their behavior.

Davies-Bouldin Index:

The DB Index value for the clustering result was <insert value>, indicating the quality of the segmentation.

Cluster Profiles:

Each cluster exhibited unique characteristics based on the following:

Cluster 0: High spenders with fewer transactions but high average transaction value.

Cluster 1: Frequent shoppers with moderate spending.

Cluster 2: Low spenders with occasional transactions.

Cluster 3: Customers with balanced transaction frequency and average spending.

Cluster 4: Outliers or niche customers with unique spending patterns.

Cluster Centers:

The cluster centers (average values for each feature) provided insights into the behavior of each group.
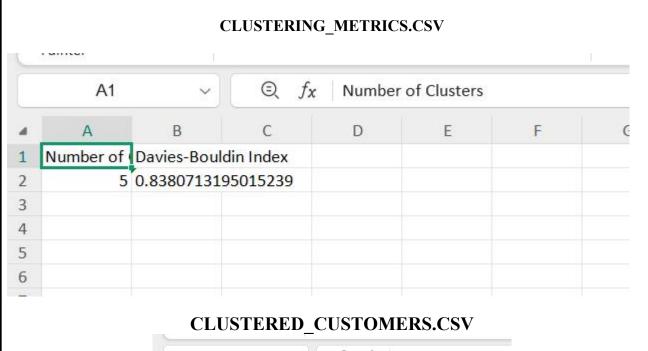
4. Visualization

PCA Scatter Plot:

A 2D visualization of clusters was created using PCA (Principal Component Analysis). The plot showed clear separations between clusters, validating the segmentation.

t-SNE Scatter Plot:

t-SNE provided an alternate visualization, showcasing the distinctness of clusters with non-linear relationships.

CLUSTERING_METRICS.CSV

	A	B	C	D	E	F
	A1		Number of Clusters			
1	Number of	Davies-Bouldin Index				
2	5	0.8380713195015239				
3						
4						
5						
6						

CLUSTERED_CUSTOMERS.CSV

	A	B	C	D	E
	A1		CustomerID		
1	CustomerID	Cluster			
2	C0001	1			
3	C0002	1			
4	C0003	3			
5	C0004	0			
6	C0005	3			
7	C0006	2			
8	C0007	3			
9	C0008	0			
10	C0009	4			
11	C0010	1			
12	C0011	1			
13	C0012	0			
14	C0013	0			
15	C0014	4			
16	C0015	3			
17	C0016	2			
18	C0017	0			
19	C0018	2			
20	C0019	1			
21	C0020	3			
22	C0021	0			
23	C0022	2			
24	C0023	2			
25	C0024	1			
26	C0025	4			
27	C0026	2			
28	C0027	3			
29	C0028	0			

clustered_customers +