CUSTOMER SEGMENTATION/CLUSTERING REPORT

Introduction

We performed customer segmentation using clustering techniques on customer profile information (from the `Customers.csv`) and transaction history (from the `Transactions.csv`). We employed the **K-Means clustering algorithm** to group customers into segments based on similarities in both their profile data (demographics, region, signup date) and transaction behavior (purchase frequency, total spending, etc.).

Optimal Number of K

To determine the optimal number of clusters (`k`), we used the **Silhouette Score**, which evaluates how similar each customer is to its own cluster compared to other clusters. The goal is to maximize the Silhouette Score to ensure the clusters are well-separated and compact.

- **Method**: I tested multiple values of `k` ranging from 2 to 10 clusters, calculating the Silhouette Score for each.
- **Best value of `k`**: The value of `k` that yielded the highest Silhouette Score was selected as the optimal number of clusters.
- **`k = 7`** was selected based on the highest Silhouette Score of `0.3150055662322247`

Davies-Bouldin Index (DB Index) for `k = 7`:

DB Index = `1.0238041373203535` (This indicates separated and compact clusters. A value closer to 0 would indicate even better-defined clusters).

Cluster Sizes :

Cluster 0 : 18 customers
Cluster 1 : 46 customers
Cluster 2 : 17 customers
Cluster 3 : 49 customers
Cluster 4 : 19 customers
Cluster 5 : 27 customers
Cluster 6 : 23 customers