Clustering Results Report

1. Introduction

- **Objective:** Provide an overview of the task, i.e., customer segmentation using clustering techniques based on both customer profile and transaction data.

- **Dataset Description:**

    o **Customers.csv:** Includes CustomerID, CustomerName, Region, SignupDate.

    o **Transactions.csv:** Includes TransactionID, CustomerID, ProductID, TransactionDate.

- **Purpose:** To identify meaningful customer segments to enable targeted marketing strategies.

2. Methodology

- **Data Preprocessing:**

    o Data cleaning steps (e.g., handling missing values, encoding categorical variables).

    o Scaling or normalization applied to the features.

- **Features Used:**

    o A list of features derived from the datasets (e.g., Region, SignupDate, frequency of transactions).

- **Clustering Algorithm:**

    o Specify the clustering algorithm used (e.g., K-Means, DBSCAN, or Agglomerative Clustering).

    o Justify why this algorithm was chosen.

- **Number of Clusters:**

    o Describe how the optimal number of clusters was determined (e.g., Elbow Method, Silhouette Analysis).

3. Results

- **Number of Clusters Formed:**

    o the final number is 4 clusters.

- **Cluster Characteristics:**

    o the characteristics of each cluster (e.g., high transaction frequency, specific regions, or signup behavior).

    o Include a table summarizing key metrics for each cluster.

4. Evaluation Metrics

- **Davies-Bouldin Index (DB Index):**

  - Present the calculated DB Index value (e.g., **0.95**) and interpret its significance.

- **Other Clustering Metrics:**

  - Silhouette Score I have obtained is 0.33.