Customer Segmentation Report

1. Introduction

The goal of this report is to present the customer segmentation results using clustering techniques. The datasets used for this task are "Customers.csv" (profile information) and "Transactions.csv" (transaction data).

The K-Means clustering algorithm was chosen for this task, and various metrics such as the Davies-Bouldin Index (DB Index) were computed to evaluate the clustering results.

2. Data Preparation

Data preprocessing involved the following steps:

- Merging Datasets: The customer and transaction data were merged using CustomerID as the key.
- Handling Missing Data: Missing values were handled by filling them with the mean of the respective column.
- Feature Selection: Only relevant numerical features were selected for clustering, such as spending habits and transaction frequency.
- Data Normalization: The data was scaled using Standard Scaler to ensure each feature contributes equally to the clustering process.

3. Clustering

The K-Means clustering algorithm was applied with the following process:

- Choosing the Algorithm: K-Means was chosen for its simplicity and effectiveness in partitioning data into distinct groups.
- Determining the Optimal Number of Clusters: The Elbow Method was used to determine the optimal value for k. The "elbow" point

  appeared at k=3, indicating that 3 clusters would be most suitable for segmentation.

- Fitting K-Means: The K-Means algorithm was then applied with k=3, and the clustering labels were assigned to each customer.

4. Evaluation

The clustering performance was evaluated using the following metrics:

- Davies-Bouldin Index: The calculated DB Index was 0.98.

5. Cluster Visualization

PCA (Principal Component Analysis) was used to reduce the dimensionality of the data for 2D visualization.

A scatter plot was generated to visualize the customer segments. The clusters are color-coded and represent distinct groups based

on customer behavior and transaction patterns.

6. Key Findings

The clusters revealed the following insights:

- Cluster 1: High-value customers with significant spending but low frequency of purchases, indicating they are occasional large spenders.
- Cluster 2: Frequent but smaller-value customers, suggesting a regular buying pattern with lower transaction amounts.
- Cluster 3: Low-value, low-frequency customers, likely casual buyers or new customers

These clusters provide valuable insights into customer behavior that can inform targeted marketing and sales strategies.

Future work may include testing other clustering algorithms such as DBSCAN or exploring additional features for clustering.