Clustering Report

1. Introduction

This report summarizes the customer segmentation task using clustering techniques. The objective is to group customers based on their purchasing behaviors, leveraging both profile and transaction data. Key outputs include clustering metrics, visualizations, and an evaluation of the clusters formed.

2. Methodology

2.1 Data Preparation

Data from Customers.csv and Transactions.csv were merged on CustomerID. Aggregated features for each customer included:

- **Total Quantity Purchased** (Quantity)
- **Total Value of Transactions** (TotalValue)
- **Transaction Count** (TransactionDate renamed to TransactionCount)

2.2 Clustering Algorithm

The K-Means clustering algorithm was employed, with the optimal number of clusters determined using the Elbow Method.

3. Results

3.1 Optimal Number of Clusters

Using the Elbow Method, the optimal number of clusters ("k") was found to be **4**, where the rate of decline in inertia began to level off.

3.2 Davies-Bouldin Index

The Davies-Bouldin Index (DB Index) for the clustering result was calculated as:

- **DB Index:** 0.5495184341182848

A lower DB Index indicates better-defined clusters.

3.3 Clustering Visualizations

The clusters were visualized using a scatter plot of Total Quantity Purchased vs. Total Value of Transactions, color-coded by cluster.

Visualization: Elbow Method

Visualization: Customer Segmentation Clusters

4. Summary of Clusters

Cluster ID	Average Quantity	Average Total Value	Average Transaction Count
0	21.93	6525.94	7.72
1	10.68	2670.68	4.48
2	4.85	969.09	2.27
3	15.64	4507.83	6.04

5. Conclusion

The clustering analysis identified 4 distinct customer segments. These insights can help tailor marketing strategies, improve customer retention, and optimize sales efforts. Future work could include exploring different clustering algorithms and incorporating additional features for enhanced segmentation.

6. Deliverables

1. Clustering Metrics:
    - Optimal Clusters: 4
    - Davies-Bouldin Index: 0.5495184341182848
2. Visualizations:
    - Elbow Method for Optimal k
    - Customer Segmentation Clusters