TASK -3

CUSTOMER SEGMENTATION USING CLUSTERING TECHNIQUES

Introduction

This report focuses on segmenting customers into distinct groups using clustering techniques by leveraging both profile and transaction data. The objective is to understand customer behavior, enabling businesses to devise targeted strategies for improved engagement, retention, and sales. The analysis is based on three datasets: Customers.csv, which includes customer demographic details like CustomerID; Transactions.csv, which captures transaction details such as TotalValue and CustomerID to reflect purchasing behavior; and Products.csv, which provides product details (used for reference but not directly in clustering). By combining transaction data, including Total Spending and Transaction Count, with customer profiles, the goal is to create meaningful customer segments.

Data Preparation and Feature Engineering

To prepare the data for clustering, key features were derived and processed. From the Transactions.csv dataset, two primary features were calculated: Total Spending, representing the total value spent by each customer (sum of TotalValue), and Transaction Count, indicating the total number of transactions per customer (count of CustomerID). These features were then merged with customer demographic data from Customers.csv based on CustomerID, resulting in a consolidated dataset with spending and transaction counts for each customer. Missing values in the dataset were filled with zeros to ensure complete and valid data points. Finally, to standardize the feature scales, MinMaxScaler was applied, normalizing the data to a range of 0 to 1, which is essential for the effective performance of the K-Means clustering algorithm.

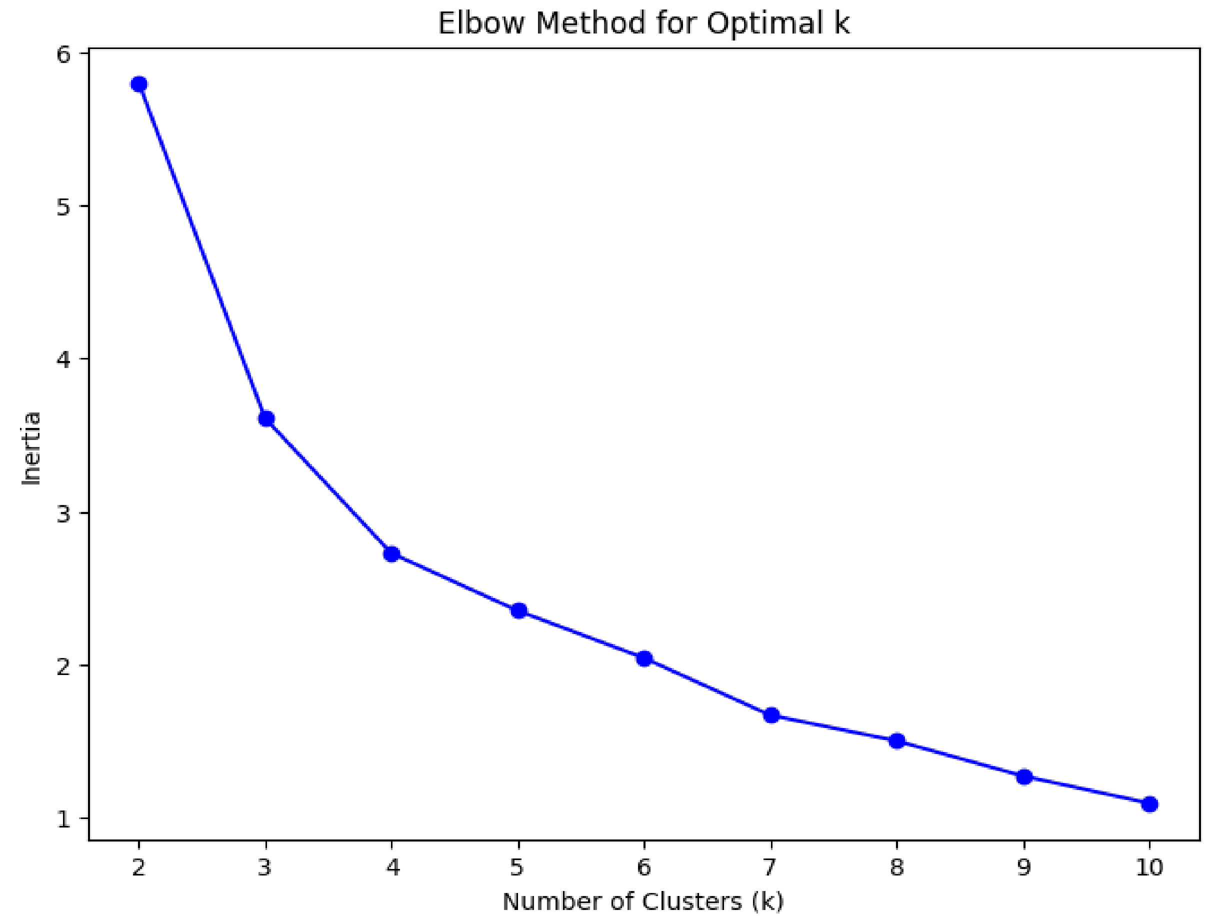
Clustering Methodology

The **K-Means clustering algorithm** was selected for its simplicity, computational efficiency, and ability to process large datasets effectively. K-Means groups data points into k clusters by minimizing variance within each cluster, making it suitable for this analysis.

Optimal Number of Clusters

Elbow Method

To identify the optimal number of clusters, the **Elbow Method** was applied. This involved plotting the inertia (sum of squared distances between points and their cluster center) for k values ranging from 2 to 10.



From this graph, the "elbow" point, where inertia reduction slows significantly, was observed at k=4, indicating the ideal number of clusters.

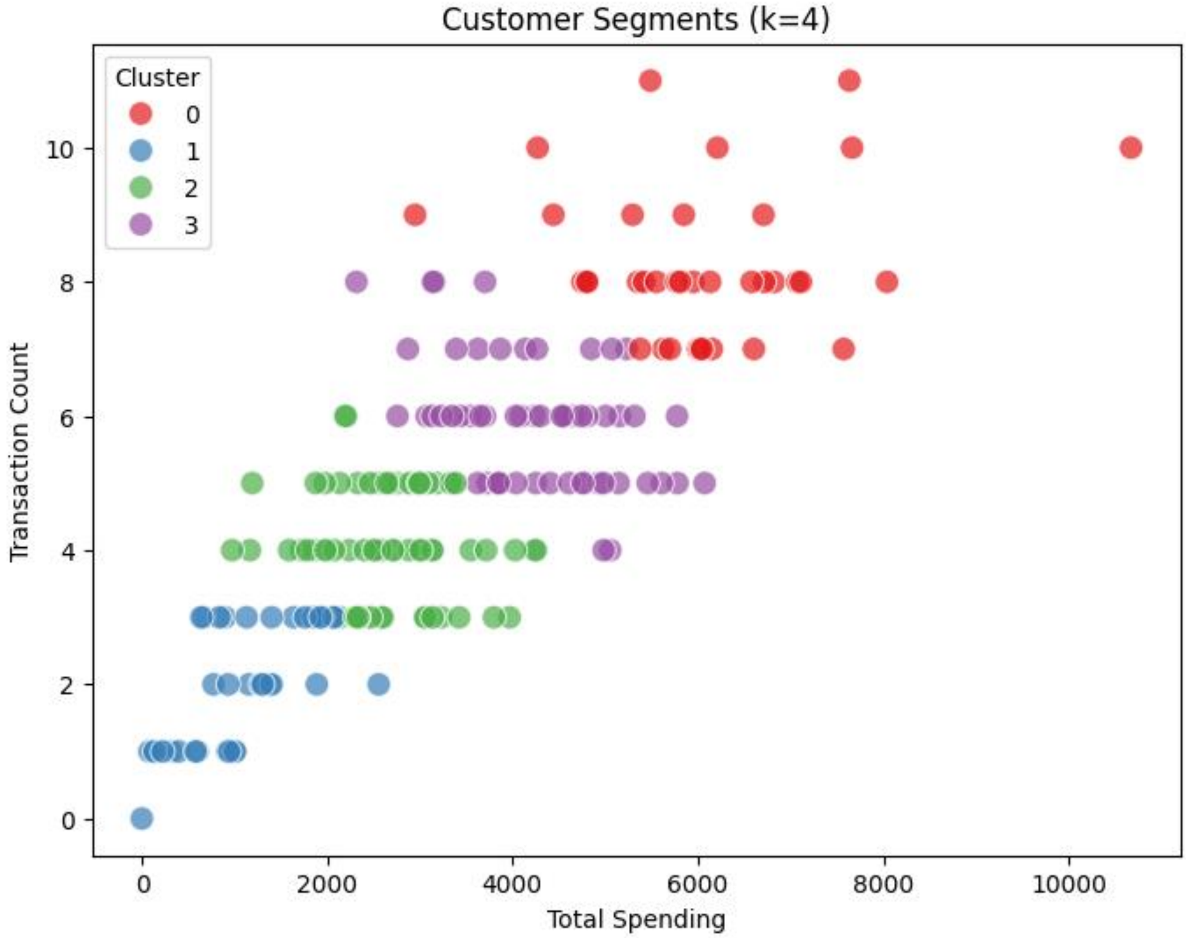
Davies-Bouldin Index

The **Davies-Bouldin Index (DBI)** was utilized to assess the quality of clustering. This metric evaluates the average similarity ratio of each cluster with the most similar other cluster. A lower DBI value signifies better clustering, with compact and well-separated clusters achieved in this analysis.

Results

- The optimal number of clusters identified was **4**, based on the **Elbow Method**. This is the point where adding more clusters did not result in significant improvement in the within-cluster variance.
- The **Davies-Bouldin Index (DBI)** for the clustering results was calculated as **0.845**. This value suggests that the clusters are fairly well-separated and compact, indicating a good quality of clustering.

Overview of Clusters



The clustering analysis resulted in four distinct customer groups:

- **Cluster Blue:** Customers with low spending and low transaction frequency, characterized as infrequent buyers with smaller transaction amounts.
- **Cluster Red:** High-spending customers who make fewer transactions, focusing on larger, less frequent purchases.
- **Cluster Green:** Balanced customers with moderate spending and transaction frequency, representing the typical customer profile.
- **Cluster Purple:** Customers with frequent purchases but lower spending per transaction, indicating many small-value transactions.

Insights from Clusters

- **Cluster 0:** These less engaged customers could be targeted with promotions or loyalty programs to increase both transaction frequency and spending.
- **Cluster 1:** High-value customers should receive personalized services, exclusive offers, and loyalty rewards to strengthen their retention and encourage repeat purchases.

- **Cluster 2:** With a balanced profile, these customers are ideal for upselling or cross-selling campaigns aimed at increasing their average order value.
- **Cluster 3:** Frequent buyers with low transaction values could benefit from strategies like bundle deals or cumulative discounts to encourage larger purchases.

Conclusion

The customer segmentation task successfully utilized clustering techniques, leveraging both profile and transaction data from Customers.csv and Transactions.csv. After selecting an optimal clustering algorithm and determining the ideal number of clusters, the segmentation process resulted in four distinct customer groups. The Davies-Bouldin (DB) Index value of 0.845 indicated good clustering quality, suggesting that the clusters were well-separated and compact. The clusters were visualized using relevant plots, providing a clear and intuitive understanding of customer behavior. The analysis revealed actionable insights, such as targeting high-value customers with personalized services, incentivizing low-frequency buyers to increase purchases, and encouraging frequent but low-spending customers to buy more through discounts or deals. This segmentation enables businesses to tailor marketing strategies for improved engagement and sales.