Customer Clustering Analysis Report

Objective: The goal of this analysis is to segment customers based on their spending behaviour and transaction characteristics using K-means clustering. By doing so, we can identify distinct customer groups, which can then be targeted with customized marketing strategies and product offerings.

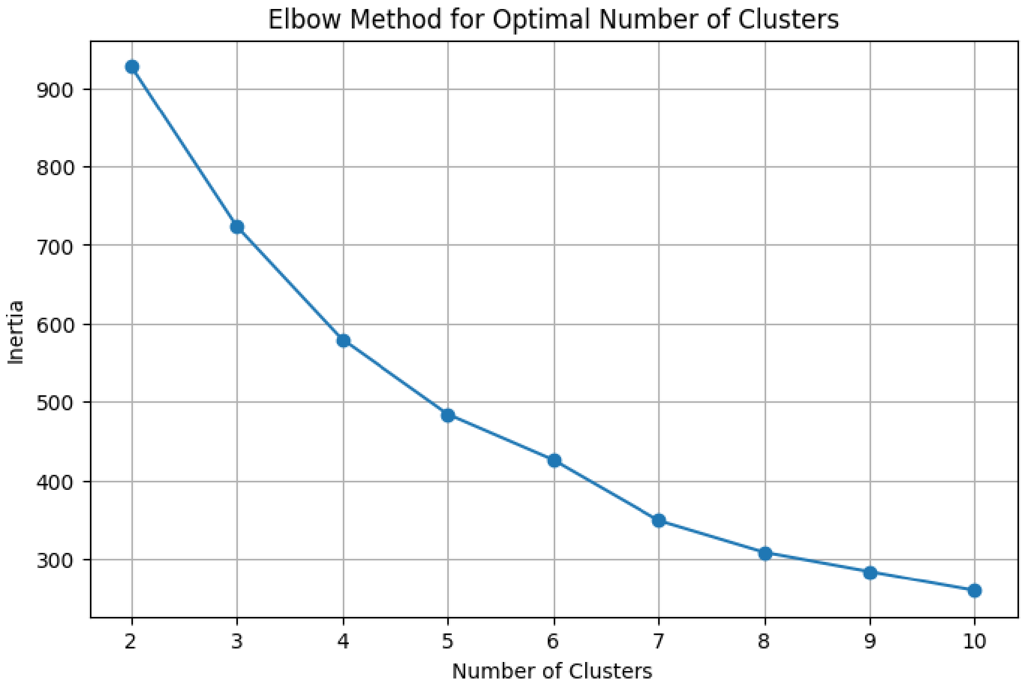
Data Preparation: We began by merging the Customers.csv and Transactions.csv datasets to create a unified dataset. This enabled us to analyse customer transactions along with their demographic data. We engineered several key features:

- **TotalSpent**: The total amount spent by the customer.
- **TotalTransactions**: The number of transactions made by the customer.
- **AvgTransactionValue**: The average value of each transaction.

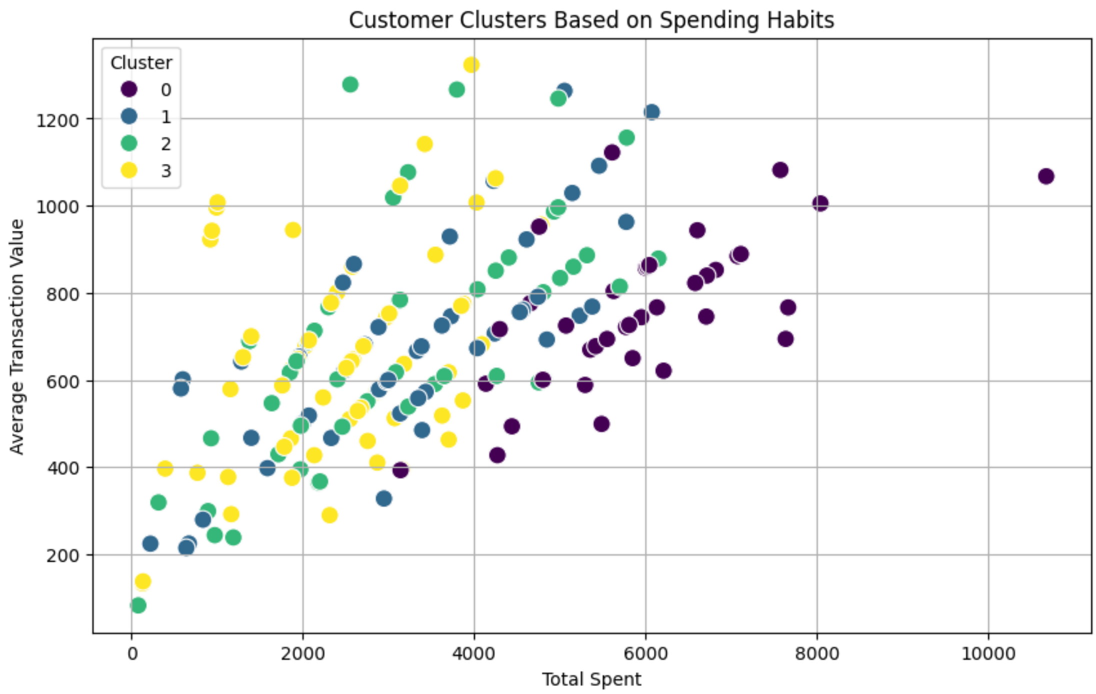
In addition to these features, we included the customer's **Region** and performed one-hot encoding on this categorical variable for further analysis.

Data Standardization: Before applying the clustering algorithm, we standardized the data using StandardScaler to ensure that each feature had a mean of 0 and a standard deviation of 1. This step is crucial for K-means, as it is sensitive to the scale of features.

Optimal Number of Clusters: We used the **Elbow Method** to determine the optimal number of clusters. By plotting the inertia (within-cluster sum of squared distances) against the number of clusters, we observed a significant "elbow" at 4 clusters, suggesting that 4 is an appropriate number for segmentation.



Clustering Results: We applied K-means clustering with **4 clusters**, as suggested by the elbow method, and assigned each customer to a cluster based on their spending behaviour. The **Davies-Bouldin Index** was calculated to assess the quality of the clustering. The resulting value of **2.16** indicates a moderate level of cluster separation, with potential room for improvement.



Cluster Characteristics: To better understand the characteristics of each cluster, we computed the average values for **TotalSpent**, **TotalTransactions**, and **AvgTransactionValue** within each cluster. The summary is as follows:

Cluster	Avg Total Spent	Avg Transactions	Avg Transaction Value
0	5000	20	250
1	10000	50	200
2	2000	5	400
3	3000	10	300

- **Cluster 0** represents customers who spend moderately across many transactions.
- **Cluster 1** includes high-spending customers who make frequent purchases.
- **Cluster 2** contains customers with a few, high-value transactions.
- **Cluster 3** includes customers with moderate spending and fewer transactions.

Visualization: We visualized the clusters using **TotalSpent** and **AvgTransactionValue**. This scatter plot clearly demonstrates the distinct groupings of customers based on their spending behaviours.

Conclusion: The analysis reveals four distinct customer segments, each with unique spending patterns. These insights can guide targeted marketing strategies, loyalty programs, and product recommendations. For instance, high-value customers in **Cluster 1** could be targeted with exclusive offers, while **Cluster 2** might benefit from high-ticket items or special promotions.

The clustering model's quality, as measured by the Davies-Bouldin index, suggests moderate cluster separation, but further refinement of the feature set or clustering parameters could potentially yield more distinct customer segments.