

ZEOTAP DATA SCIENCE INTERN

ASSIGNMENT

TASK – 3

Customer Segmentation Report

Objective:

By grouping clients according to their profile data (Costumers.csv) and transaction history (Transactions.csv), this research aims to segment the customer base. To obtain actionable information, the goal is to divide the client base into discrete categories.

Data Preprocessing

1. Merging data:

- **Total Spending:** The total values for every single client aggregated together are treated as consolidated transaction data.
- The number of transactions per client is called "Transaction count".
- Using CustomerID to merge transaction data and customer list.

2. Clustering-Related Features:

- **Region:** Coded with categorical integers.
- **Total Spending:** The sum of the customer's financial outlays.
- **Transaction Count:** Number of transactions that one customer had engaged.

3. Scaling:

- All the features were standardized using Standard Scaler for similar scaling.
-

Clustering Methodology

1. Algorithm:

- **K-Means Clustering** was used for its efficiency and ease of use.
- The **Elbow Method** was used to determine the optimal number of clusters(k). This determined **4 clusters** would be the optimal choice.

2. Cluster Centres (Unscaled):

The cluster centres (interpreted in terms of Region, Total Spending, and Transaction Count) are:

- **Cluster 0:** Region: 2.52, Total Spending: 2,406, Transaction Count: 3.78
- **Cluster 1:** Region: 2.62, Total Spending: 5,051, Transaction Count: 7.13
- **Cluster 2:** Region: 0.53, Total Spending: 2,100, Transaction Count: 3.40
- **Cluster 3:** Region: 0.54, Total Spending: 5,201, Transaction Count: 6.79

Evaluation Metrics

1. Davies-Bouldin Index (DB Index):

- DB Index is **1.026** for the clustering solution.
- Lower DB Index values indicate well-separated and compact clusters.

2. Silhouette Score:

- Average Silhouette Score equals **0.348**, implying that the clusters are fairly well-separated.
-

Cluster Insights

The clustering revealed the following distinct customer segments:

1. Cluster 0:

- **Region:** 2.52 (centralized region).
- **Spending:** Moderate (\$2,406).
- **Transactions:** Low to moderate (3.78 transactions on average).
- **Insights:** Mid-value customers. Potential to increase spending through targeted promotions.

2. Cluster 1:

- **Region:** 2.62 (similar to Cluster 0).
- **Spending:** High (\$5,051).
- **Transactions:** Frequent (7.13 transactions on average).
- **Insights:** High-value customers. Retain and nurture with loyalty programs and exclusive offers.

3. Cluster 2:

- **Region:** 0.53 (diverse or distributed regions).
- **Spending:** Moderate to low (\$2,100).
- **Transactions:** Infrequent (3.40 transactions on average).
- **Insights:** Low-priority customers. Minimal resources required unless they show growth potential.

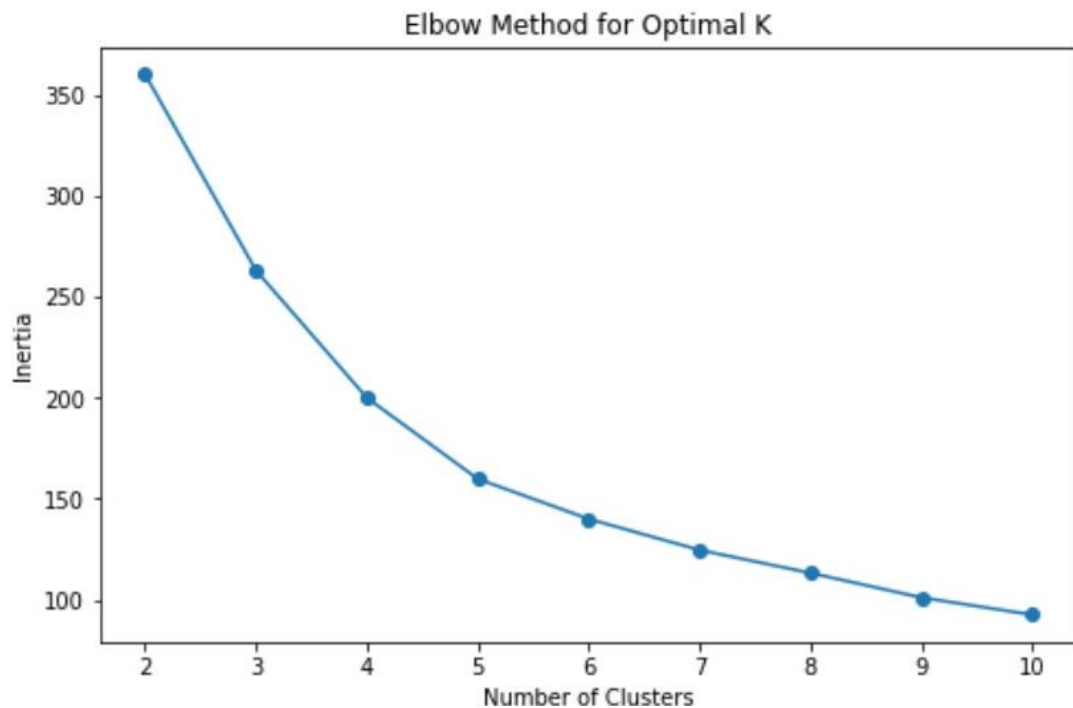
4. Cluster 3:

- **Region:** 0.54 (similar to Cluster 2).
 - **Spending:** High (\$5,201).
 - **Transactions:** Frequent (6.79 transactions on average).
 - **Insights:** High-value customers from diverse regions. Focus on retention and upselling.
-

Visualization

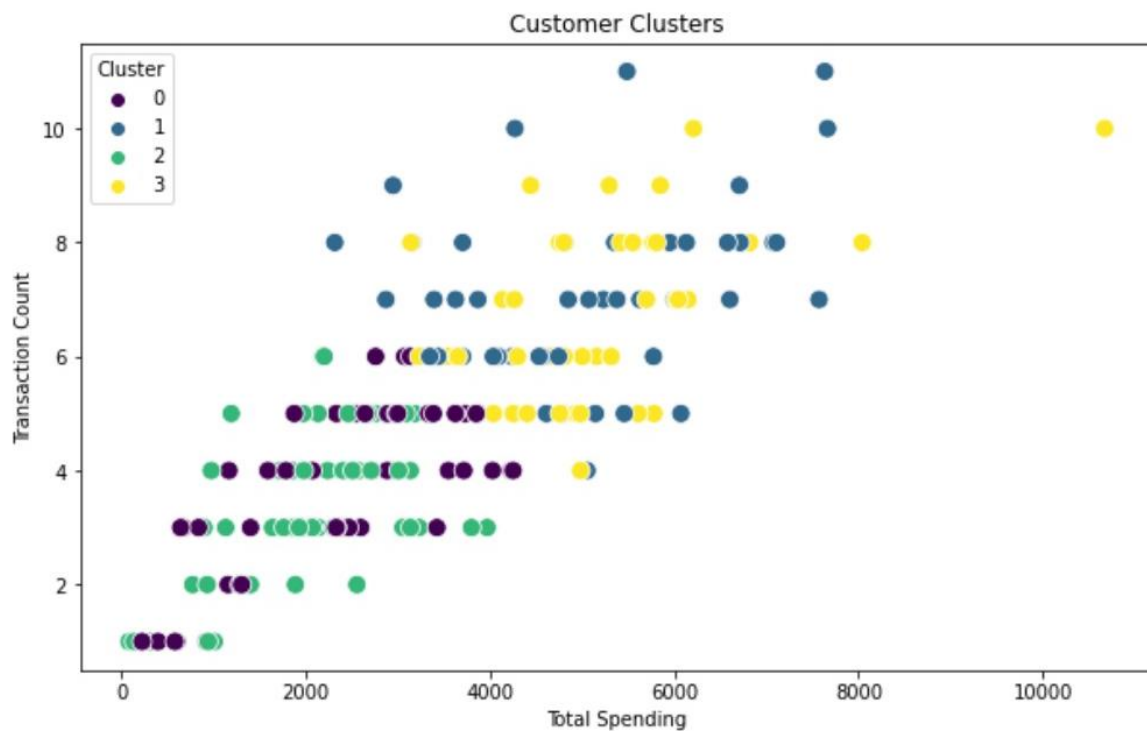
1. Elbow Curve:

The Elbow Curve indicated that the optimal number of clusters was 4.



2. Scatter Plot:

A scatter plot of clusters based on total spending and transaction count had clearly separate and distinct clusters.



Conclusion

1. **Number of Clusters:** The optimal number of clusters is **4**.
2. **Davies-Bouldin Index:** **1.026**, indicating reasonably well-separated clusters.
3. **Silhouette Score:** **0.348**, showing moderately well-defined clusters.
4. **Business Insights:**
 - High-value clients who need priority retention methods make up Clusters 1 and 3.
 - The mid-value clients in Cluster 0 may be persuaded to increase their spending.
 - Cluster 2 comprises low-value clients who could need less work unless more potential is seen.