

Task-3: Customer Segmentation / Clustering

Objective:

The goal was to segment customers using clustering techniques based on both customer profile and transaction data.

Data Overview:

Data was sourced from:

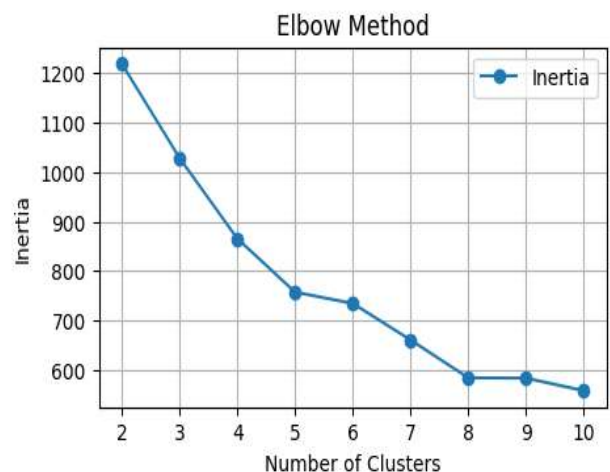
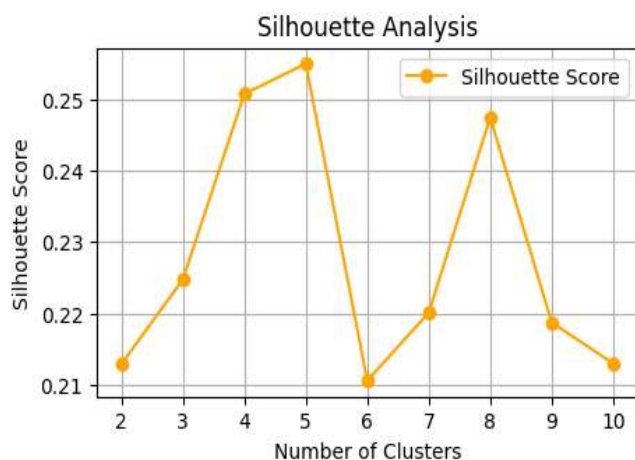
- **Customer Profiles:** Includes CustomerID, Region, and SignupDate (from Customers.csv).
- **Transaction History:** Includes TransactionID, Quantity, and TotalValue (from Transactions.csv).

To prepare the data for clustering, the transaction data was aggregated at the customer level. The resulting dataset, has the following features:

- **TotalQuantity:** Total number of items each customer bought.
- **TotalSpending:** Total amount spent by each customer.
- **AvgSpending:** Average amount spent per transaction.
- **TransactionCount:** Total number of transactions made by each customer.

Clustering Methodology:

- **Algorithm:** K-Means clustering
- **Number of Clusters:** 4 clusters were selected based on optimal performance using the Elbow Method and Silhouette Analysis.



Evaluation Metrics:

- **Davies-Bouldin Index:** The DB Index was 1.36, indicating reasonable cluster separation.
- **Silhouette Score:** The score of 0.25 suggests a moderate level of overlap, but still reflects meaningful grouping.

Dimensionality Reduction:

In order to visualize the clusters in two dimensions while maintaining variance, PCA was used to reduce the data to two dimensions (PC1 and PC2).

Cluster Profiles:

Cluster Summary:				
Cluster	TotalSpending		TransactionCount	TotalQuantity
	mean	sum	mean	mean
0	5770.18	300049.52	7.38	20.31
1	2905.60	127846.20	4.45	10.86
2	2316.06	143595.54	3.61	8.71
3	2890.35	118504.30	4.78	11.29

Cluster 0:

- **Behavior:** Customers in this cluster are high spenders with a long customer tenure. They are distributed across multiple regions, with a strong presence in Europe and South America.
- **Business Insight:** These customers are highly valuable and should be targeted with premium offers and loyalty programs to maximize their lifetime value.

Cluster 1:

- **Behavior:** Customers in this cluster are lower spenders with fewer transactions. They are concentrated in South America and have a long customer tenure.
- **Business Insight:** These customers may benefit from targeted promotions and discounts to encourage increased spending and engagement.

Cluster 2:

- **Behavior:** Customers in this cluster are moderate spenders with a moderate number of transactions. They are mostly located in Europe and have a long customer tenure.
- **Business Insight:** These customers are ideal candidates for cross-selling and upselling strategies to increase their average transaction value.

Cluster 3:

- **Behavior:** Customers in this cluster are moderate spenders with a shorter customer tenure. They are based solely in North America and exhibit distinct behavior compared to other clusters.
- **Business Insight:** These customers may require re-engagement strategies, such as personalized recommendations or special offers, to increase their loyalty and spending.

Visual Representation:

The clusters were visualized using PCA, with the following scatter plot showing customer distribution:



Conclusion:

- Clusters Formed: 4 clusters were formed.
- DB Index: 1.36, reflecting solid cluster separation.
- Silhouette Score: 0.25, indicating reasonable clustering cohesion.