**CUSTOMER SEGMENTATION REPORT**

**OBJECTIVE**

This was a customer segmentation clustering task. By utilizing the customer profile data (for example, region, signup date) and transaction data (for example, total spending, transaction frequency), the aim is to divide customers into distinct groups for appropriate targeted marketing and improved business strategies through this customer segmentation.

**CLUSTERING METHODOLOGY**

1.DATA PREPARATION:

* Merged the customer profile data (Customers.csv) with transaction data (Transactions.csv) into a single data set.
* Engineered key features from the data include:
* Total Spend: Sum of transaction values by customer.
* Transaction Count: Count of transactions per customer.
* Avg. Transaction Value: Mean of all transactions per customer.
* Categorical variables such as Region were one-hot encoded
* Normalized all the numerical features with StandardScaler.

2.CLUSTERING ALGORITHM:

* K-Means clustering was applied because it is simple and efficient for moderately sized datasets.
* The optimal number of clusters was determined using the Elbow Method, balancing between WCSS (Within-Cluster Sum of Squares) and interpretability.

3.EVALUATION METRICS:

* Davies-Bouldin Index (DB Index): Measures the compactness and separation of clusters (lower values are better).
* Silhouette Score: Assesses how well-separated clusters are (ranges from -1 to 1; higher values are better).

**RESULTS**

1.NUMBER OF CLUSTERS:

* The Elbow Method showed that the number of clusters is 4.

2.EVALUATION METRICS:

* Davies-Bouldin Index: 1.1926 (well-separated clusters with moderate compactness).
* Silhouette Score: 0.3197 (moderately cohesive clusters with some overlap).

3.CLUSTER CHARACTERISTICS: Every cluster is made up of a set of customers with different profiles, depending on their behaviour and demographics:

Cluster 0: High-value customers with frequent transactions.

Cluster 1: Average spenders with average transaction frequency.

Cluster 2: Low-frequency customers with minimal spending.

Cluster 3: Customers with infrequent but high-value transactions.

**CLUSTER VISUALIZATION**

A 2D PCA plot is created to view the clusters below:

* PCA reduces the dimensionality of the dataset while retaining most of the variance.
* The scatterplot shows distinct groups, with overlapping regions indicating customers with similar profiles.

**BUSINESS INSIGHTS**

1.TARGET HIGH-VALUE CUSTOMERS:

* Cluster 0 includes customers with high spend and frequent transactions. Premium offers or loyalty programs can be offered to maximize lifetime value for such customers.

2.INCREASING ACTIVITY:

* Cluster 2 is characterized by low-frequency customers. Focused engagement campaigns, like discounts or reminders, can help increase their activity.

3.UPSELL OPPORTUNITIES:

* Cluster 3 includes customers who have made sporadic but high-value purchases. Such customers can be targeted through upselling strategies or exclusive deals.

4.RESOURCE OPTIMIZATION:

* By segmenting customers, businesses can allocate marketing resources effectively, prioritizing high-value and medium-engagement clusters.

**CONCLUSION**  
The segmentation was successful in clearly identifying four well-defined customer groups according to their profile and transaction behaviour. Evaluation metrics, particularly the Davies-Bouldin Index of 1.1926, indicate that the clustering is of good quality. These insights may help the business tailor marketing efforts for better customer satisfaction and higher profitability.