Clustering Report

The dataset initially included the following attributes: Price, Quantity, TotalValue, SignupDate, TransactionID, ProductID, TransactionDate, CustomerName and Region. I calculated the difference between the SignupDate and TransactionDate to create a new feature called DaysToFirstPurchase in order to improve the analysis. Furthermore, by calculating the quantity of transactions linked to every customer, I developed an additional feature named Number of Transactions per customers.

I created two more features—Total Spending, which adds together all of the money spent by each customer and Average Spending per Customer, which determines the average transaction value—in order to learn more about how much people spend. Region, DaysToFirstPurchase, Number of Transactions, Total Spending, and Average Spending were the final feature set that resulted from this transformation. I used Label Encoding to transform the Region column's categorical data into a numerical representation.

I used MinMaxScaler to normalise the data through feature scaling. After that, I used K-Means clustering with cluster values ranging from 2 to 10. I determined that the ideal number of clusters might be either four or five by examining the elbow plot. Five clusters offered the best customer segmentation, according to my visual evaluation of the clustering results. I created several 2D and 3D scatter plots to support the clusters and show how various features influenced the clustering outcomes.

I calculated the Davies-Bouldin Index (DBI), which evaluates the compactness and separation of clusters, in order to evaluate the model. My model's DBI score is 1.051; a lower number denotes higher clustering quality. Furthermore, the clusters' degree of

separation, measured by the Silhouette Score was 0.372. The final measure of how closely the points within each cluster are concentrated around their individual centroids is the Inertia score (Within-Cluster Sum of Squares), which came out at 26.91.

Customers were successfully divided into five significant segments by the clustering technique, which gave a more thorough comprehension of customer behaviour based on spending trends and history of transactions. Personalised suggestions, client retention initiatives and focused marketing techniques can all benefit from these insights.