CLUSTERING

The data was grouped into **5 clusters** using the **KMeans clustering algorithm.**

Clustering Metrics

- **Davies-Bouldin Index:** The Davies-Bouldin Index is a metric used to evaluate the quality of clustering by measuring the average similarity ratio of each cluster to its most similar one. A lower score indicates better clustering quality.
- Davies-Bouldin Index: 0.9384521529981805

- **Silhouette Score:** The silhouette score measures how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The score ranges from -1 (poor clustering) to +1 (well-clustered), with scores closer to +1 indicating better-defined clusters.
- Silhouette Score: 0.40077093834633815

Clustering Overview

Based on the clustering of customer data, the algorithm identified 5 distinct customer segments. These segments were formed based on various factors, including:

- Total Revenue: The sum of transaction values for each customer.
- Total Quantity: The total number of items purchased.
- Total Transactions: The count of distinct products purchased.
- Region: The geographical categorization of customers.

The scatter plot of the first two features of the normalized customer profile shows a clear distribution of customers into distinct clusters. Each customer is colored according to their assigned cluster.The resulting clusters represent different segments of customers based on their purchasing behavior. By analyzing the characteristics of each cluster businesses can tailor marketing strategies, product recommendations, and customer service offerings to meet the needs of each group.