# Customer Segmentation Report

## Objective:

The goal of this task was to perform customer segmentation using clustering techniques, leveraging both customer profile data (from Customers.csv) and transaction data (from Transactions.csv). The analysis aimed to identify meaningful customer segments and evaluate the clustering results using the Davies-Bouldin Index (DB Index).

## Approach:

1.Data Preparation and Merging:

- Merged the Customers.csv and Transactions.csv datasets using the CustomerID as the key.
- Aggregated transaction data for each customer to compute total spending (TotalValue) and the total number of items purchased (Quantity).
- Combined these aggregated values with customer profile information, such as Region.

2.Data Preprocessing:

- Encoded categorical variables (e.g., Region) using Label Encoding for compatibility with clustering algorithms.
- Standardized the dataset using StandardScaler to ensure all features contributed equally to the clustering process.

3.Clustering Methodology:

- Applied the KMeans clustering algorithm with the number of clusters set to 5, as determined during the analysis.
- Clustering was performed on the processed features, including transaction and profile attributes.

4.Evaluation Metrics:

- The Davies-Bouldin Index (DB Index) was calculated to evaluate the quality of the clustering. A lower DB Index indicates better-defined clusters.
- Final DB Index Value: 0.9039, indicating well-separated and compact clusters.

5.Visualization:

- Generated scatter plots to visualize the clusters based on selected features, providing insights into customer segmentation patterns.
- Created a heatmap showing the correlation between numeric features to understand feature relationships better.

## Results and Analysis

Number of Clusters Formed:

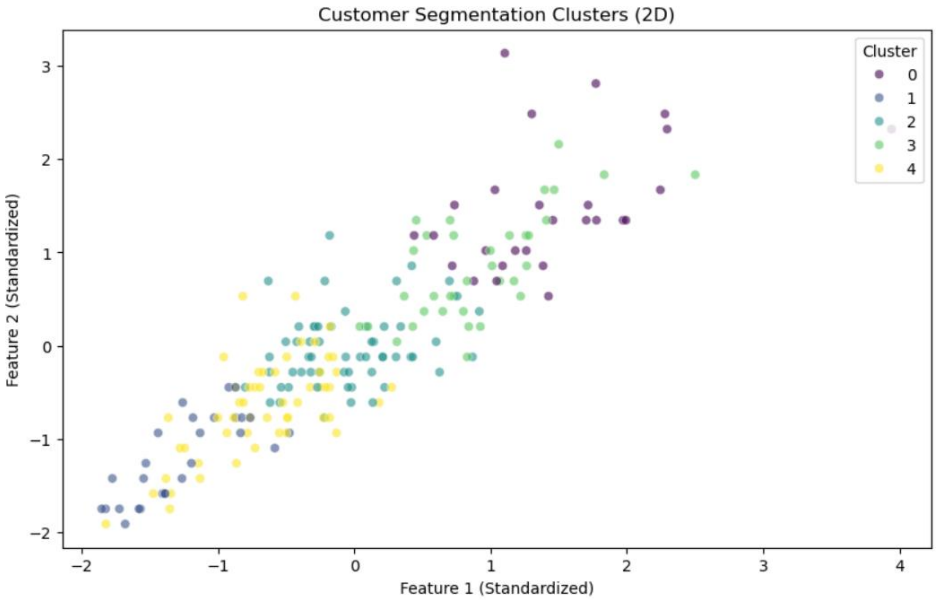
- 5 clusters were formed, representing distinct customer segments based on purchasing patterns and demographics.

Clustering Evaluation (DB Index):

- The calculated DB Index for the clustering was 0.9039, indicating well-defined clusters with minimal overlap.

Cluster Visualizations:

- To better understand and communicate the clustering results, the following visualizations were generated:




- ➢ A scatter plot visualizing the distribution of customers across the 5 clusters, plotted using key numeric features like TotalValue and Quantity. This provides a clear representation of cluster boundaries and densities.


- A correlation heatmap illustrating relationships between numeric features used in the clustering process (e.g., TotalValue, Quantity, Age). This supports the feature selection process and highlights interactions between variables.