# Clustering Report

## Objective:

Perform customer segmentation using clustering techniques. Use both profile information (from Customers.csv) and transaction information (from Transactions.csv).

● You have the flexibility to choose any clustering algorithm and any number of clusters in between(2 and 10)

● Calculate clustering metrics, including the DB Index(Evaluation will be done on this).

● Visualise your clusters using relevant plots.

## Data Overview:

Two datasets were used for this analysis:

1. Customers.csv: Contained demographic and profile information (e.g., CustomerID, Region, Signup Date).

2. Transactions.csv: Included transaction details (e.g., CustomerID, Transaction Value, Transaction Date). The datasets were merged on CustomerID to create a unified view, enabling analysis of both customer profiles and transaction behavior.

## Feature Engineering:

Key features were engineered from the merged dataset:

• TotalValue: Total value of transactions per customer.

• Quantity: Total number of Quantity per customer.

• SignupDate: Average SignupDate value per customer.

• ProductID: ProductID for each customer.

• Category: Categorical variable indicating the customer's region (one-hot encoded for clustering).

These features were normalized to ensure equal weighting during clustering.

# Clustering Approach:

1. Clustering Algorithm: K-Means clustering was used due to its simplicity and efficiency for this type of segmentation.

2. Range of Clusters Tested: Clusters were evaluated for $k \in [2,10]$k \in [2, 10]. 3. Evaluation Metric: The Davies-Bouldin Index (DB Index) was used to determine cluster quality. Lower DB Index values indicate better cluster compactness and separation.

# Results:

Optimal Number of Clusters The DB Index was calculated for each $k \in [2,10]$k \in [2, 10].

The lowest DB Index value was achieved at k = 4, suggesting that 4 clusters provide the most meaningful segmentation.
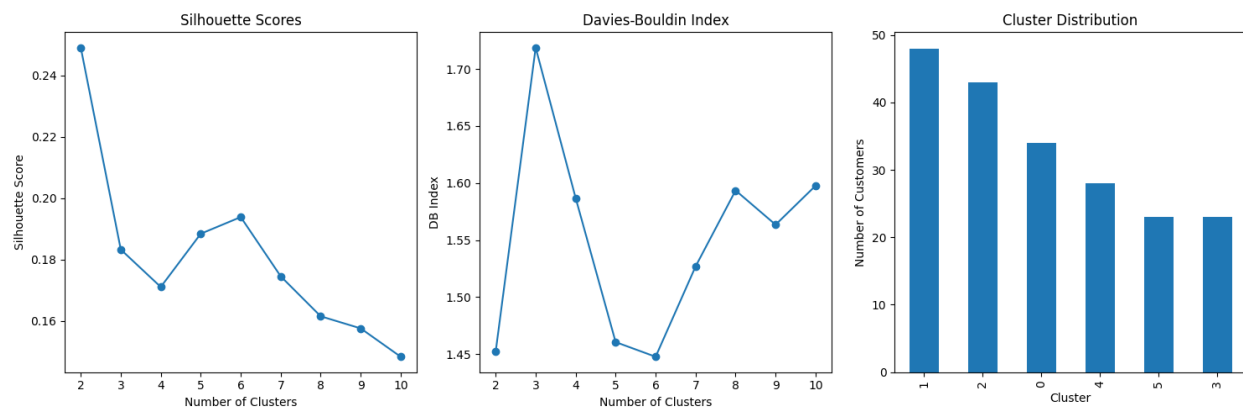


# Cluster characteristics :

Best Number of Clusters: 6

# Recommendations:

 1. Marketing Strategies: o Personalized promotions based on cluster profiles. o Retention programs for high-value customers.
2. Further Analysis:  Use additional algorithms (e.g., hierarchical clustering or DBSCAN) to validate results. Segment based on other features, such as product preferences or demographics.
3. Business Applications: o Allocate resources effectively for customer acquisition and retention. o Monitor the performance of clusters over time to refine strategies.