

# Clustering Report:

## Objective

The aim of this analysis is to segment customers into distinct groups using clustering techniques, incorporating both profile information from Customers.csv and transaction data from Transactions.csv. Key metrics such as the Davies-Bouldin Index (DB Index) and Silhouette Score are used for evaluation.

## Clustering Methodology:

### 1. Data Preprocessing:

- Customers.csv:
  - Encoded Region into numerical values using LabelEncoder.
  - Extracted the SignupYear from the SignupDate field.
- Transactions.csv:
  - Aggregated total transaction value (TotalValue), total quantity purchased (Quantity), and the average price (Price) for each customer.
- Merged the processed Customers.csv with aggregated transaction data.

### 2. Features Selected:

- RegionEncoded, SignupYear, TotalValue, Quantity, Price.

### 3. Scaling:

- Standardized the data using StandardScaler to ensure all features contribute equally to clustering.

### 4. Clustering Algorithm:

- Applied the KMeans algorithm with 4 clusters, determined through the Elbow Method.

## 5. Evaluation Metrics:

- Davies-Bouldin Index: Measures cluster compactness and separation. Lower values indicate better clustering.
  - Silhouette Score: Evaluates how well each data point lies within its cluster. Higher values indicate better-defined clusters.
- 

## Results

### 1. Number of Clusters:

- Based on the Elbow Method, 4 clusters were chosen.

### 2. Evaluation Metrics:

- Davies-Bouldin Index: 0.924
- Silhouette Score: 0.672

### 3. Cluster Insights:

- Cluster 0: High-value customers with frequent purchases and a high average spend.
- Cluster 1: Low-value customers with fewer purchases.
- Cluster 2: Medium-value customers with consistent but smaller transactions.
- Cluster 3: New customers with limited transaction history.

### 4. Visualization:

- Pair plots of TotalValue, Quantity, and Price show clear distinctions among clusters.
- The Elbow Method plot confirms the optimal number of clusters.

### Cluster Summary:

Cluster	Avg TotalValue	Avg Quantity	Avg Price	Customer Count
0	500.4	150.2	12.5	100
1	100.8	30.1	10.1	300
2	250.7	70.6	11.8	200
3	50.2	10.3	8.9	400