

# Customer Segmentation Report on eCommerce Transactions Dataset

---

## Overview

The objective of this analysis was to perform customer segmentation using clustering techniques to identify distinct groups of customers based on their transaction behavior and profile information. The insights from this segmentation can be used for personalized marketing strategies, customer retention, and business growth. This analysis utilized the K-Means algorithm, with a range of clusters evaluated using the Davies-Bouldin Index (DBI), and visualized the clusters for interpretability.

---

## Clustering Methodology

### 1. Data Preparation:

- **Datasets Used:** Customers.csv and Transactions.csv.
- **Features Engineered:**
  - **Total Purchase Value:** Aggregated total transaction values for each customer.
  - **Transaction Frequency:** Count of transactions made by each customer.
  - **Region Encoding:** One-hot encoded categorical data for the region feature.
- **Normalization:** Standardized all numerical features to ensure fair weight in clustering.

### 2. Clustering Algorithm:

- **Algorithm Used:** K-Means clustering.
- **Cluster Range:** Explored cluster counts from 2 to 10 to identify the optimal number.
- **Evaluation Metric:** Davies-Bouldin Index (DBI) was used to evaluate the compactness and separation of clusters. Lower DBI values indicate better clustering quality.

### 3. Optimal Cluster Selection:

- **Optimal Clusters:** The analysis identified **9 clusters** as optimal.
  - **Optimal DBI Value:** The minimum DB Index value achieved was **0.72**, indicating well-formed clusters.
- 

## Clustering Results

### 1. Number of Clusters Formed:

- **Optimal Clusters:** 9 distinct clusters were identified.

## 2. Cluster Evaluation Metrics:

- **Davies-Bouldin Index (DBI):**
  - Achieved DBI for 9 clusters: **0.72**.
- **Inertia:** A measure of within-cluster variance, minimized with optimal clusters.
- **Silhouette Analysis:** Optional metric that could further validate cluster cohesion and separation.

## 3. Cluster Characteristics:

- Each cluster reflects unique patterns in customer behavior:
    - **Cluster 0:** Customers with high transaction frequency but moderate total purchase value.
    - **Cluster 1:** High-value customers with fewer transactions.
    - **Cluster 2:** Customers concentrated in specific regions with average transaction activity.
    - **Cluster 3:** Customers with balanced transaction frequency and purchase value.
    - **Other Clusters (4-8):** Reflect additional nuances in customer behavior, such as region-specific traits or outliers.
- 

## Visual Representation

### 1. PCA Visualization:

- Dimensionality reduction using Principal Component Analysis (PCA) was performed.
- A 2D scatter plot illustrated the cluster distribution, with distinct colors representing each cluster.
- The clusters appear well-separated, confirming the effectiveness of the clustering process.

### 2. Cluster Distribution:

- The distribution of customers across clusters indicates meaningful segmentation.
- No significant imbalance was observed, ensuring robust group formation.

### 3. Additional Visuals:

- **Box Plots:** To highlight differences in transaction value and frequency across clusters.
  - **Heatmaps:** Show correlations between features within clusters.
  - **Cluster Size Bar Chart:** Displaying the number of customers per cluster.
-

## Conclusion

- The clustering process successfully segmented customers into **9 meaningful groups**.
- The achieved DB Index value (**0.72**) confirms the compactness and separation of clusters.
- These segments reveal distinct patterns in customer transaction behavior and profiles, which can drive personalized business strategies.

## Key Insights:

- High-value customers (Cluster 1) can be targeted with exclusive loyalty programs.
  - Frequent buyers with moderate spending (Cluster 0) may respond well to bundle offers.
  - Region-specific clusters (e.g., Cluster 2) suggest opportunities for localized marketing campaigns.
- 

## Next Steps

1. **Validation:**
    - Apply additional metrics like the Silhouette Score and CH Index to confirm clustering quality.
  2. **Business Application:**
    - Use cluster insights to design targeted campaigns, retention strategies, and cross-sell opportunities.
  3. **Model Extension:**
    - Explore advanced clustering techniques such as DBSCAN or hierarchical clustering.
    - Incorporate temporal features like seasonal trends in customer transactions.
-