

# CLUSTERING REPORT

## Clustering Methodology:

The steps involved in the clustering process included:

### 1. Data Preparation:

- Loaded the Customers.csv and Transactions.csv datasets.
- Merged the datasets on CustomerID to create a comprehensive dataset.
- Created features such as total transaction value, average transaction value, total quantity of items purchased, total number of transactions, and customer tenure in days.

### 2. Normalization:

- Applied Standard Scaler to normalize features for consistent scaling, as clustering algorithms like K-Means are sensitive to differences in magnitude.

### 3. Clustering:

- Configured the K-Means algorithm for 5 clusters (n\_clusters=5) based on data exploration.
- Evaluated the clustering performance using the Davies-Bouldin Index (DB Index), which measures cluster compactness and separation.

### 4. Visualization:

- Used PCA for dimensionality reduction to project high-dimensional data into two dimensions.
- Visualized clusters in a 2D scatterplot with clusters represented by different colors.

## DB Index Value

- Davies-Bouldin Index: 1.68

A lower DB Index indicates better-defined clusters. The calculated value suggests reasonably good clustering quality.

## Cluster Insights

### 1. Cluster 0:

- Represents customers with the highest transaction value and transaction count.
- These customers are the most profitable and might be key accounts.
- Strategic focus: Personalized marketing and exclusive loyalty programs.

## **2. Cluster 1:**

- Indicates mid-tier customers with moderate transaction frequency and value.
- Strategic focus: Incentives to encourage higher spending (e.g., upsell opportunities).

## **3. Cluster 2:**

- Contains customers with low transaction value and frequency.
- These customers may include occasional buyers or new customers.
- Strategic focus: Engage with promotional offers and targeted campaigns.

## **4. Cluster 3:**

- Comprised of customers with relatively high tenure but below-average spending.
- These customers might have the potential for growth but require encouragement to increase transaction frequency.
- Strategic focus: Retention campaigns with value-oriented incentives.

## **5. Cluster 4:**

- Encompasses customers with short tenure and moderate transactions.
- Likely represents newer customers who are still exploring the services/products.
- Strategic focus: Building long-term loyalty through introductory offers and enhanced customer support.

## **Recommendations**

### **1. High-Value Customers (Cluster 0):**

- Implement retention strategies to maintain engagement.
- Introduce exclusive memberships to maximize profitability.

## **2. Mid-Tier Customers (Cluster 1):**

- Offer special discounts to convert them into high-value customers.

## **3. Low-Engagement Customers (Cluster 2):**

- Address these with re-engagement strategies like promotions, bundle offers, and personalized discounts.

## **4. Customers with Growth Potential (Cluster 3 & 4):**

- For long-tenure customers with low spending (Cluster 3), analyze reasons for inactivity and improve service quality.
- For newer customers (Cluster 4), focus on customer onboarding to build trust and engagement.

## **Final Deliverables**

### **1. Clusters Visualization:**

- The 2D scatter plot effectively distinguishes different customer segments.
- Each color represents a unique cluster.

### **2. Output Files:**

- Clustering results were saved to Clustering\_Results.csv for further analysis.

Overall, this analysis aids in designing tailored strategies for each customer segment, fostering improved business outcomes.