

A Report on

**CUSTOMER SEGMENTATION –  
- CLUSTERING**

Submitted By:

**Alaina Umme Hafiza**

B.Tech in Computer Science Engineering

CMR University

Submitted To:

**Zeotap India Pvt. Ltd.**



# Introduction

Customer segmentation is a powerful technique that allows businesses to categorize their customers into distinct groups based on their purchasing behavior and demographic characteristics. This report presents an analysis of customer segmentation using clustering techniques on an eCommerce transactions dataset. The dataset consists of three files: Customers.csv, Products.csv, and Transactions.csv.

The objective of this study is to identify meaningful customer groups using clustering methods. The analysis includes preprocessing the data, applying the K-Means clustering algorithm, evaluating the clustering performance using the Davies-Bouldin Index, and visualizing the results.

## Data Description

The dataset consists of the following files:

- **Customers.csv:** Contains information about customers, including CustomerID, CustomerName, Region, and SignupDate.
- **Products.csv:** Contains details about the products, including ProductID, ProductName, Category, and Price.
- **Transactions.csv:** Contains transaction details such as TransactionID, CustomerID, ProductID, TransactionDate, Quantity, and TotalValue.

## Data Preprocessing

1. **Merging Datasets:** The transaction data was merged with customer details to include demographic attributes.
2. **Feature Engineering:**
  - Total transaction value per customer.
  - Total quantity of products purchased per customer.
  - Number of unique transaction dates per customer (purchase frequency).
3. **Standardization:** The numerical features were standardized using the StandardScaler to ensure uniform scaling.

## Clustering Methodology

- **Algorithm Used:** K-Means clustering was chosen due to its efficiency and interpretability. The number of clusters was set to 4.
- **Cluster Assignment:** Each customer was assigned to one of the four clusters based on the K-Means results.
- **Evaluation Metric:** The Davies-Bouldin Index (DB Index) was used to measure the clustering performance.

## Results and Analysis

- **Number of Clusters Formed:** 4
- **Davies-Bouldin Index:** 0.865

## Cluster Characteristics

1. **High-Value Customers:** Customers with high transaction values and frequent purchases.
2. **Moderate Buyers:** Customers with medium transaction values and moderate purchase frequency.
3. **Occasional Shoppers:** Customers who purchase infrequently but with varying total values.
4. **Low-Value Customers:** Customers with minimal purchases and low total transaction values.

## Visualization

To better understand the clusters, PCA was used to reduce the dimensionality of the data, and a scatter plot was created to visualize the clusters. The plot illustrates distinct groups, validating the segmentation results.

## Conclusion and Business Insights

- **Personalized Marketing:** Different marketing strategies should be applied to each cluster. High-value customers may benefit from loyalty programs, while occasional shoppers may need targeted promotions.
- **Inventory Management:** Understanding purchasing patterns can help optimize inventory levels.
- **Customer Retention:** Identifying customers with declining purchase frequency can enable proactive engagement strategies.
- **Regional Insights:** Regional segmentation may reveal purchasing trends, allowing targeted regional promotions.