

Customer Segmentation using Clustering

The objective of this task is to segment customers based on their profile information and transaction history using clustering techniques. The goal is to identify distinct customer groups that exhibit similar purchasing behaviors, which can help in targeted marketing and personalized recommendations.

Data Description: The dataset consists of three files:

1. **Customers.csv** - contains information on customers, including CustomerID, Name, Region, and Signup Date
2. **Products.csv** - lists product details, including ProductID, Name, Category, and Price
3. **Transactions.csv** - records transaction details, including TransactionID, CustomerID, ProductID, Transaction Date, Quantity, and Total Transaction Value

For clustering, we used data from **Customers.csv** and **Transactions.csv**, focusing on:

- **Total Spent:** Total amount spent by a customer across all transactions.
- **Transaction Count:** Number of transactions made by the customer.
- **Recency:** Days since the last transaction.
- **Region:** Encoded categorical variable representing the customer's region.

1. Data Preprocessing:

- Merged customer data with transaction data.
- Computed key features (Total Spent, Transaction Count, Recency).
- Encoded categorical variables and scaled numerical features using StandardScaler.

2. Determining the Optimal Number of Clusters:

- Used the **Elbow Method** to determine the optimal number of clusters by analyzing the inertia curve.
- Selected **K=4** based on the Elbow Method.

3. Clustering Algorithm:

- Applied **K-Means Clustering** with K=4.
- Assigned each customer to a cluster.

4. Evaluation Metrics:

- **Davies-Bouldin Index (DB Index):** evaluates cluster compactness and separation.
- **Silhouette Score:** measures how well-separated the clusters are.

5. Visualization:

- Applied **PCA (Principal Component Analysis)** to reduce dimensionality.
- Visualized clusters in a 2D scatter plot.

RESULTS:

- **Number of Clusters Formed: 4**
- **Davies-Bouldin Index (DB Index): 1.0185** (lower values indicate better clustering)
- **Silhouette Score: 0.3103** (higher values indicate better-defined clusters)

Conclusion: The clustering analysis successfully segmented customers into four distinct groups based on their purchasing behavior. The **Davies-Bouldin Index of 1.0185** indicates a reasonably good separation between clusters, while the **Silhouette Score of 0.3103** suggests that the clustering structure is moderately well-defined.

This segmentation can help in:

- Identifying high-value customers who make frequent purchases.
- Targeting less active customers with promotions to increase engagement.
- Understanding customer spending patterns for better marketing strategies.