

# Task 3: Customer Segmentation / Clustering

## Customer Segmentation and Clustering Report

This report outlines the results of customer segmentation using clustering techniques to group customers based on their transaction profiles. The primary aim of this analysis is to segment customers into distinct groups for better-targeted marketing strategies and personalized services.

## Methodology

The first step involved merging two datasets: **Customers.csv** and **Transactions.csv**. The customers' profile information, such as their region and signup date, was merged with their transaction history, including the total spent, number of purchases, and transaction values. After merging the datasets, key features were engineered to represent customer behavior:

1. **TotalSpent**: The total value of all transactions made by the customer.
2. **NumPurchases**: The number of purchases made by the customer.
3. **AvgTransactionValue**: The average value per transaction.

The next step was to standardize these features to ensure that differences in scales across features (e.g., total spent and number of purchases) did not bias the clustering process. Standardization was critical to allow the clustering algorithm to give equal weight to all features.

For clustering, **K-Means** was chosen as the algorithm of choice. The K-Means algorithm assigns each customer to one of several clusters based on similarity, aiming to minimize intra-cluster variance and maximize inter-cluster separation. In this case, we performed clustering with **4 clusters**, as it provided the most meaningful segmentation based on our evaluation metrics.

## Results

After running the K-Means algorithm, we obtained four distinct customer segments. These segments grouped customers with similar purchasing behaviors, which could help personalize marketing campaigns, offer targeted discounts, and improve overall customer satisfaction.

The clustering quality was evaluated using the **Davies-Bouldin (DB) Index**, which measures the compactness and separation of clusters. A lower DB Index indicates better clustering quality, as it reflects that the clusters are compact and well-separated. In our case, the DB Index value was found to be **1.13**, which suggests that the clusters are reasonably well-separated and compact.

## Conclusion

The K-Means clustering technique successfully segmented the customers into four distinct groups based on their transaction behaviors. These segments offer valuable insights into customer profiles and will serve as a foundation for more effective marketing strategies. The

DB Index value indicates that the clusters formed are of good quality and can be used to identify target audiences for various business initiatives.