

# Customer Segmentation Analysis Report

## Objective

This analysis is based on customer transaction data. Therefore, the purpose of the report is to segment customers on-upon identification of groups that are significant from a business standpoint. The aim is to educate us with insight towards high-value customer groups and develop strategies for marketing efforts, promotional exercises, or customer retention.

## Data Overview

Focusing on the two datasets that were used to conduct this analysis, they are listed below:

1. Customers Dataset (Contains customer's details like CustomerID and SignupDate)
2. Transactions Dataset (Contains transaction's details like CustomerID, TransactionID, TransactionDate, TotalValue (amount spend), and Quantity)

## Approach

The process involving customer segmentation involves:

### 1. Data Preprocessing

- Date Handling: We had converted SignupDate and TransactionDate into datetime formats for purposes of calculation of the relevant metrics such as the days since signup and active purchase duration.
- Missing Data: Working with customers without any transactions, missing values were coded to zero for the metric derived itself.

### 2. Feature Engineering

- We calculated different metrics out of the raw data available, which described customer behavior:
- Days Active: Number of days since a customer signed up.
- Transaction Count: Total transactions by each customer.
- Total Spend: Total money spent by the customer overall.
- Average Transaction Value: The average value spent on each transaction (Total Spend / Transaction Count).
- Purchase Frequency: Transactions made monthly.
- Purchase Duration: Time between the first and last transaction.
- This category of metrics actually offers an insight into recency, frequency, and monetary values.

These metrics provided a comprehensive view of customer behavior, including recency (Days Active), frequency (Transaction Count and Purchase Frequency), and monetary value (Total Spend and Avg Transaction Value).

### 3. Data Scaling

The calculated features had varying scales (e.g., Days Active was in days, Total Spend in dollars). To ensure fair comparison during clustering, we used `StandardScaler` to normalize the data.

### 4. Clustering

To group customers into segments:

1. **Model Selection:** We chose the K-Means algorithm due to its simplicity and effectiveness for numeric data.
2. **Optimal Cluster Selection:** We evaluated clusters ranging from 2 to 10 using the following metrics:
  - **Davies-Bouldin Index (DBI):** Measures the compactness and separation of clusters. Lower values indicate better-defined clusters.
  - **Silhouette Score:** Measures how well samples are clustered. Higher values indicate better-defined clusters.
  - Based on the results, the optimal number of clusters was determined to be **8**.

### 5. Visualization

Two visualizations were created to interpret the clustering results:

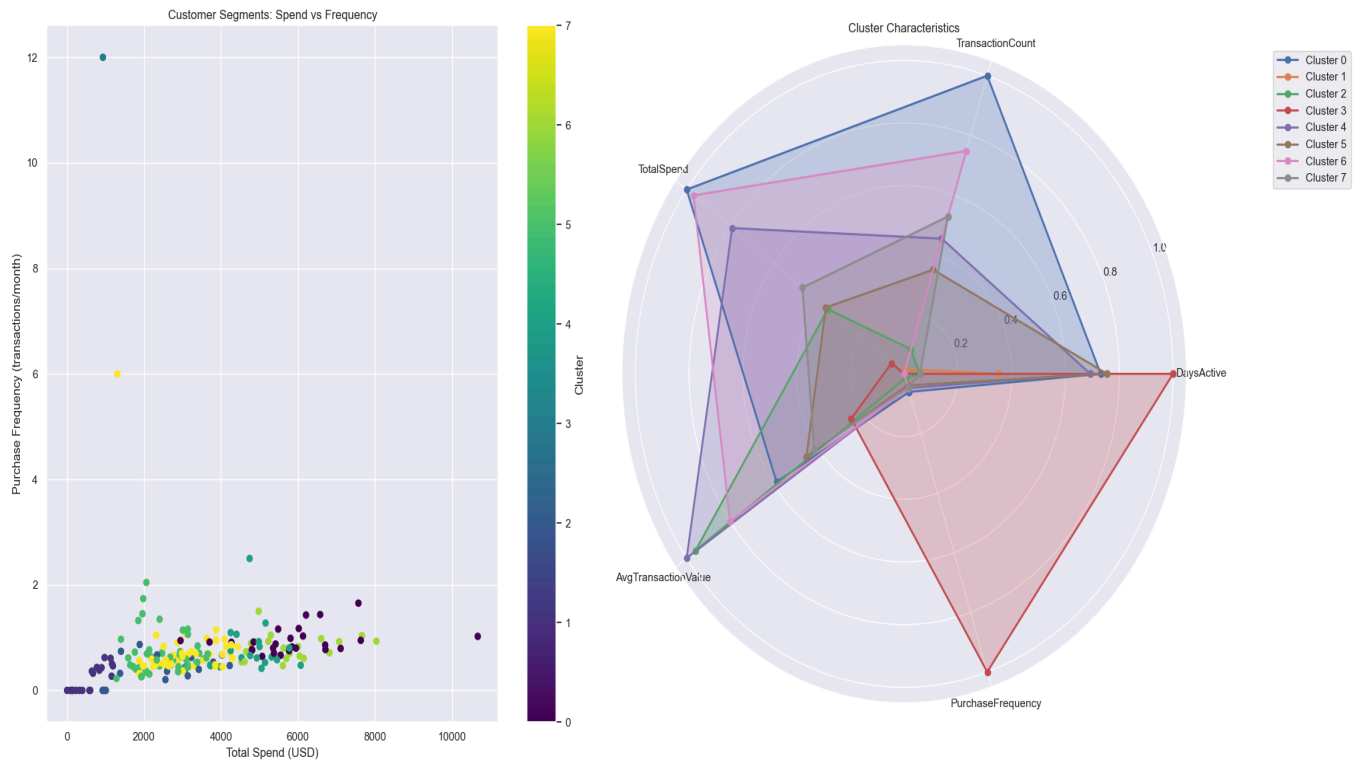
1. **Scatter Plot:** A 2D scatter plot of Total Spend vs. Purchase Frequency, colored by cluster labels, to observe customer distribution and segmentation.
2. **Radar Chart:** Showed normalized cluster centers for all features, highlighting differences in behavior between clusters.

## Results

### Clustering Evaluation

The evaluation metrics for the clustering model were as follows:

- **Optimal Number of Clusters:** 8
- **Davies-Bouldin Index:** 0.9116 (lower is better, indicating compact clusters with good separation).
- **Silhouette Score:** 0.2924 (indicating moderate cluster cohesion and separation).



## Cluster Profiles

Below are the key characteristics of each customer cluster:

### 1. Cluster 0:

- **Size:** 26 customers
- **Behavior:** High spenders with moderate transaction counts and frequency.
- **Average Spend:** \$5,861.11 per customer
- **Average Transactions:** 8.38
- **Frequency:** 0.95 transactions/month
- **Average Transaction Value:** \$706.65
- **Insights:** These customers are valuable, making fewer but high-value purchases.

### 2. Cluster 1:

- **Size:** 19 customers
- **Behavior:** Low spenders with minimal transactions and low frequency.
- **Average Spend:** \$627.26
- **Average Transactions:** 2.11
- **Frequency:** 0.23 transactions/month
- **Average Transaction Value:** \$295.05
- **Insights:** These customers are infrequent buyers with low overall engagement.

### 3. Cluster 2:

- **Size:** 19 customers
- **Behavior:** Moderate spenders with low transaction counts but high average transaction value.
- **Average Spend:** \$2,462.81

- **Average Transactions:** 2.53
  - **Frequency:** 0.41 transactions/month
  - **Average Transaction Value:** \$968.23
  - **Insights:** High-value buyers with limited but significant transactions.
4. **Cluster 3:**
- **Size:** 1 customer
  - **Behavior:** Unique customer with very high transaction frequency (12/month).
  - **Average Spend:** \$931.83
  - **Average Transactions:** 2.0
  - **Frequency:** 12.00 transactions/month
  - **Average Transaction Value:** \$465.92
  - **Insights:** Possibly an anomaly or a very specific customer with extremely frequent low-value purchases.
5. **Cluster 4:**
- **Size:** 19 customers
  - **Behavior:** High-value, moderate-frequency customers.
  - **Average Spend:** \$4,767.65
  - **Average Transactions:** 4.89
  - **Frequency:** 0.80 transactions/month
  - **Average Transaction Value:** \$996.64
  - **Insights:** A valuable group with steady purchasing habits.
6. **Cluster 5:**
- **Size:** 48 customers
  - **Behavior:** Average spenders with consistent but moderate purchasing.
  - **Average Spend:** \$2,512.05
  - **Average Transactions:** 4.23
  - **Frequency:** 0.69 transactions/month
  - **Average Transaction Value:** \$610.20
  - **Insights:** Stable customers with potential for upselling.
7. **Cluster 6:**
- **Size:** 22 customers
  - **Behavior:** High spenders with moderate transaction counts and frequency.
  - **Average Spend:** \$5,696.99
  - **Average Transactions:** 6.77
  - **Frequency:** 0.79 transactions/month
  - **Average Transaction Value:** \$856.64
  - **Insights:** Another high-value group with consistent engagement.
8. **Cluster 7:**
- **Size:** 46 customers
  - **Behavior:** Moderate spenders with steady purchasing habits.
  - **Average Spend:** \$3,075.35
  - **Average Transactions:** 5.37
  - **Frequency:** 0.77 transactions/month
  - **Average Transaction Value:** \$585.12
  - **Insights:** A middle-tier group with balanced behavior.

## Conclusion

The segmentation identified 8 distinct customer groups with varying spending, frequency, and transaction behaviors. Key takeaways include:

- **Cluster 0, 4, and 6:** These are high-value customers who should be prioritized for retention strategies or premium offers.
- **Cluster 1 and 3:** Low-value or unique customers who may require re-engagement or customized offers.
- **Clusters 5 and 7:** Moderate spenders with potential for upselling or cross-selling opportunities.

This segmentation provides actionable insights that can help tailor marketing efforts and improve customer lifetime value.