

# Customer Segmentation Analysis Report

Prepared For: Business Stakeholders

Date: 27-01-2025

## 1. Introduction

**Objective:** Identify distinct customer groups using transaction history and profile data to enable targeted marketing strategies.

**Data Sources:**

- `Customers.csv` : Customer profiles (Region, Signup Date)
- `Transactions.csv` : Purchase records (Dates, Quantities, Prices)

## 2. Methodology

**Why K-Means Clustering?**

Reason	Explanation
Simplicity	Easy to implement and explain to non-technical teams
Speed	Processes 1,000+ customers efficiently
Feature Handling	Works well with numerical data after scaling
RFM Alignment	Naturally groups customers by Recency-Frequency-Monetary patterns

## 3. Key Technical Terms

Term	Meaning	Why It Matters
DB Index	Measures how distinct clusters are (Lower = Better)	Tells us if groups are well-separated
Silhouette Score	Measures similarity within clusters (0-1, Higher = Better)	Shows if customers in same group behave alike
PCA	Reduces complex data to 2D visuals	Helps "see" clusters intuitively
RFM	Recency (last purchase), Frequency (transactions), Monetary (spending)	Industry-standard customer value metrics

## 4. Step-by-Step Approach

### Feature Creation

- Calculated customer loyalty ( **Tenure** = Last Purchase Date – Signup Date)
- Measured engagement ( **Recency** = Days since last purchase)
- Aggregated transaction patterns:

### Data Preparation

- Scaled numerical features to equal importance
- Encoded regions (Asia/Europe/etc.) as binary flags

### Cluster Optimization

- Tested 2-10 clusters using **Elbow Method**
- Selected **4 clusters** based on:
  - **DB Index**: 1.48 (Best balance)
  - **Silhouette Score**: 0.20 (Moderate grouping)

### Validation

- Visualized clusters in 2D using PCA

---

## 5. Cluster Profiles

(Averages Per Group)

Cluster	Tenure (Days)	Last Purchase (Days Ago)	Transactions	Total Spent	Avg Price	Dominant Region
0	485	46	7.65	\$5,707	\$274	South America
1	502	54	4.63	\$2,416	\$241	South America
2	199	256	1.43	\$779	\$229	Asia
3	406	86	3.98	\$3,783	\$339	Europe

---

## 6. Business Interpretation

### Cluster 0: Loyal Power Users (South America)

- **What:** Frequent buyers spending \$5,707 on average
- **Action:** Launch loyalty rewards program

#### **Cluster 1: Casual Shoppers (South America)**

- **What:** Moderate spenders (\$2,416) with recent activity
- **Action:** Recommend frequently bought items

#### **Cluster 2: At-Risk Customers (Asia)**

- **What:** No purchases in 256 days (longest gap)
- **Action:** Send "We miss you" discounts

#### **Cluster 3: Premium Buyers (Europe)**

- **What:** High price tolerance (\$339/transaction)
  - **Action:** Promote luxury/premium products
- 

## **7. Limitations & Next Steps**

#### **Challenges:**

- Moderate silhouette score (0.20) → Some customer overlap
- Regional bias in clusters

#### **Improvement Plan:**

1. Add purchase time patterns (weekday/weekend habits)
  2. Test DBSCAN algorithm for outlier detection
  3. Merge with customer survey data for deeper insights
- 

## **8. Conclusion**

This analysis successfully identified **4 customer types** using transaction history and signup data. The South American market shows the highest engagement, while Asian customers need re-engagement efforts. European customers represent an opportunity for premium product marketing.

#### **Recommended Actions:**

- Prioritize Cluster 0 retention with exclusive offers
- Reactivate Cluster 2 with targeted discounts
- Analyze European buying patterns for premium upsells