

Customer Segmentation Report

1. Introduction

The objective of this project is to segment customers based on their purchasing behavior using clustering techniques. By applying RFM (Recency, Frequency, Monetary) analysis and clustering algorithms, the goal is to identify distinct customer groups for targeted marketing and strategic decision-making.

2. Data Preparation

The analysis is based on a retail customer dataset. The key steps in data preparation include:

- **RFM Analysis:** Recency, Frequency, and Monetary metrics are calculated for each customer. These metrics help assess the purchasing behavior of customers.
 - **Recency:** The number of days since the customer's most recent purchase.
 - **Frequency:** The total number of purchases made by the customer.
 - **Monetary:** The total value of purchases made by the customer.
- **Normalization:** To ensure the features are comparable, the RFM values are normalized using **Min-Max Scaling**, which transforms the data to a 0-1 scale.

The RFM dataset before and after scaling is shown below:

RFM Dataset (Before Scaling)

CustomerID	Recency	Frequency	Monetary
	y	y	
12346	326	2	0.00
12347	2	182	4310.00
12348	75	31	1797.24
12349	19	73	1757.55
12350	310	17	334.40

RFM Dataset (After Scaling)

CustomerID	Recency	Frequency	Monetary
	y	y	

12346	0.87	0.000125	0.015
12347	0.0027	0.0227	0.030
12348	0.198	0.0038	0.021
12349	0.048	0.009	0.021
12350	0.828	0.002	0.016

3. Clustering Analysis

K-Means Clustering

We applied **K-Means Clustering** to the normalized RFM dataset to identify customer segments. The optimal number of clusters was determined using the **Elbow Method** and **Silhouette Score**.

- **Elbow Method:** The optimal number of clusters (**k=3**) was chosen by identifying the "elbow" point on the inertia graph (Elbow Curve).
- **Silhouette Score:** This score further confirmed that **k=3** was the most appropriate number of clusters for the dataset.

DBSCAN Clustering

We also experimented with **DBSCAN** (Density-Based Spatial Clustering of Applications with Noise) as a density-based clustering technique. DBSCAN was able to identify dense regions of points but also detected some points as noise (outliers).

Cluster Summary

The following table summarizes the average Recency, Frequency, and Monetary values for each K-Means cluster:

KMeans Cluster	Avg Recency	Avg Frequency	Avg Monetary
0	10	50	1200.00
1	200	5	300.00
2	30	80	5000.00

- **Cluster 0:** These customers are frequent purchasers with a low recency score, suggesting that they make regular, ongoing purchases.
- **Cluster 1:** Customers in this cluster have made few purchases recently, with low monetary value.
- **Cluster 2:** High-value customers with a moderate recency score and frequent purchases.

Clustering Visualization

Below are visualizations of the clusters:

- **Elbow Method:** This plot helps to determine the optimal number of clusters ($k=3$).
 - **Silhouette Analysis:** This graph shows how well-separated the clusters are.
 - **PCA 2D Visualization:** This plot visualizes the 2D distribution of the customers in the clusters.
 - **t-SNE 3D Visualization:** This 3D plot shows how the clusters are distributed based on the t-SNE algorithm.
-

4. Recommendations for Marketing Strategies

Cluster 0: High Engagement Customers

- **Marketing Strategy:**
 - These customers have high engagement levels (high frequency and monetary). Implement **loyalty programs**, **exclusive offers**, or **early access to new products** to further increase retention and maximize revenue.

Cluster 1: Dormant Customers

- **Marketing Strategy:**
 - Customers in this cluster have low recency and frequency. Focus on **reactivation campaigns** using personalized offers, discounts, or email marketing to encourage these customers to make a purchase.

Cluster 2: High-Value Customers

- **Marketing Strategy:**
 - These customers spend a lot but may not be as frequent in their purchases. Offer **VIP or premium customer programs**, **personalized product recommendations**, and **high-end loyalty benefits** to retain them and increase purchase frequency.

5. Conclusion

This segmentation analysis helps identify key customer groups based on their purchasing behavior. By implementing tailored marketing strategies for each cluster, businesses can increase engagement, retention, and overall revenue. The K-Means and DBSCAN clustering algorithms proved effective in categorizing customers into meaningful groups, offering actionable insights for future marketing initiatives.

6. Deliverables

1. Preprocessed RFM dataset (before and after scaling).
2. Code for K-Means and DBSCAN clustering algorithms.
3. Visualizations (Elbow Method, Silhouette Analysis, PCA 2D, t-SNE 3D).
4. Cluster summary table and marketing recommendations.