Customer Segmentation Analysis Report

# 1. Introduction

This project aims to perform customer segmentation using unsupervised machine learning techniques on the Online Retail dataset. The main goal is to identify different customer groups based on their purchase behaviors using RFM (Recency, Frequency, and Monetary) analysis. These insights can help businesses implement more personalized and effective marketing strategies.

# 2. Data Preprocessing

The original dataset was cleaned by:  
- Removing null and missing CustomerIDs  
- Filtering out cancelled transactions (Invoices with 'C')  
- Converting InvoiceDate to datetime format  
- Calculating total revenue per transaction (Quantity × UnitPrice)  
- Keeping only positive Quantity and UnitPrice values

# 3. RFM Analysis

RFM analysis is a marketing technique used to quantify customer value:  
- \*\*Recency\*\*: Days since the customer's last purchase  
- \*\*Frequency\*\*: Number of purchases made  
- \*\*Monetary\*\*: Total amount spent by the customer  
Each customer's RFM score was computed and later scaled for clustering.

# 4. Clustering Using KMeans

KMeans clustering was used to group customers based on their RFM scores. The Elbow Method was applied to determine the optimal number of clusters. Customers were segmented into distinct groups with similar purchasing behaviors.

# 5. Segment Summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Segment | Recency | Frequency | Monetary | Customer Count |
| 0.0 | 245.58 | 1.57 | 0.0 | 970.0 |
| 1.0 | 42.8 | 3.59 | 0.0 | 2732.0 |
| 2.0 | 1.0 | 116.83 | 0.0 | 6.0 |
| 3.0 | 11.21 | 21.71 | 0.0 | 213.0 |

# 6. Visual Analysis & Observations

The distribution of segments across Recency, Frequency, and Monetary values revealed:  
- Some segments showed high spending but low frequency (high-value infrequent buyers).  
- Other segments had frequent buyers with moderate spending.  
- Recency helped distinguish currently active customers from dormant ones.

# 7. Business Recommendations

Based on the segments identified:  
- \*\*Loyal Customers\*\*: Offer exclusive deals and VIP programs.  
- \*\*At-Risk Customers\*\*: Use email marketing or discounts to re-engage.  
- \*\*Big Spenders\*\*: Provide premium services or early product access.  
- \*\*Low-Value Customers\*\*: Use educational onboarding emails to increase activity.  
- \*\*New Customers\*\*: Send welcome offers and nurture with email sequences.

# 8. Conclusion

Unsupervised clustering based on RFM values provides actionable insights into customer behavior. By segmenting customers and analyzing their characteristics, businesses can develop more targeted campaigns, increase retention, and boost revenue.