International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 3, May-June 2025

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Application of Recency, Frequency, Monetary Analysis in Customer Segmentation

Hai Hoang Thanh*, Nguyet Mong Thi**

*(Thai Nguyen University of Economics and Business Administration, Viet Nam

Email: hoangthanhhai03091988@gmail.com)

** (Thai Nguyen High School, Viet Nam

Email: nguyetmt@tnue.edu.vn)

--************************----------------------------------

Abstract:

Customer segmentation is the process of dividing customers into disjoint groups based on shared characteristics. This helps companies to better understand customer needs or behaviors and tailor marketing strategies, products or services to meet specific segment requirements. RFM analysis (Recency, Frequency, Monetary) is a common customer segmentation technique used in marketing to evaluate and segment customers based on their transaction history. It is a simple yet powerful segmentation method, which supplies deep insights in customer data and then sharpens specialized marketing strategies. In this paper, we apply RFM method to segment customers of an online retail company. We thoroughly analysis each group and recommend suitable marketing strategies accordingly.

Keywords **— Put your keywords here, keywords are separated by comma.**

--************************----------------------------------

I. INTRODUCTION

Today, customer segmentation has become increasingly popular in business as companies recognize the value of understanding and addressing the unique needs of different customer groups. In a competitive and data-driven market, businesses are leveraging advanced analytics and digital tools to divide their customers into meaningful segments based on demographics, behaviour, preferences, or purchasing habits. This allows them to deliver personalized marketing, improve product offerings, and enhance customer experiences. For example, an online retailer might segment customers based on past purchases to recommend relevant products, while a telecom company may tailor pricing plans for different age groups or usage patterns. As a results, customer segmentation is no longer just a marketing tactic but a core strategy for companies to allocate resources more efficiently, boost customer satisfaction, and gain a competitive edge in the market.

There are several types of customer segmentation. For example, demographic segmentation groups customers according to shared characteristics, such as gender, age, marital status, educational level, occupation. For geographic segmentation, we divide customers according to geography. Psychographic segmentation is based on customer interests and personality traits, whereas behavioural segmentation considers customer purchase history, response to marketing campaigns and product or feature usage patterns when grouping [1].

The RFM method is a widely used approach in customer segmentation that analyzes customer behaviour based on three key metrics: Recency, Frequency, and Monetary value. It helps businesses identify and prioritize their most valuable customers by examining how recently a customer made a purchase (Recency), how often they purchase (Frequency), and how much they spend (Monetary). By scoring and categorizing customers using these dimensions, companies can tailor marketing strategies, improve customer retention, and increase profitability [2]. RFM is especially popular because it is simple, data-driven, and highly effective in revealing actionable insights from transactional data.

Our main objectives in this work are:

1) To segment customers of an online retail company into meaningful segments;
2) To analysis customer behaviour in each group;
3) To recommend data-drive marketing strategies suitable for specific segment.

II. METHODS

In this paper, we use RFM method for customer segmentation and exploratory data analysis (EDA) to explore customer behaviours in different groups and the relationship between customers' features and company profit.

2.1 Dataset

We use a retail dataset from the UCI machine learning repository. This is a transactional data set which consists of all the transactions from 01/12/2010 to 09/12/2011 of a UK-based and non-store online retail [3]. The major products are unique all-occation gifts. Many of its customers are wholesalers.

The data contains 541,909 instances and eight different variables detailed below.

TABLE I
ONLINE RETAIL DATASET DESCRIPTION

Variable Name	Role	Type	Description (Units)
InvoiceNo	ID	Categorical	a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation
StockCode	ID	Categorical	a 5-digit integral number uniquely assigned to each distinct product
Description	Feature	Categorical	product name
Quantity	Feature	Integer	the quantities of each product (item) per transaction
InvoiceDate	Feature	Date	the day and time when each transaction was generated
UnitPrice	Feature	Continuous	product price per unit (sterling)
CustomerID	Feature	Categorical	a 5-digit integral number uniquely assigned to each customer
Country	Feature	Categorical	the name of the country where each customer resides

The dataset is chosen due to its size and popularity in researches about retail customer segmentation.

2.2 RFM

RFM analysis is a method that categorizes customers into segments based on their purchasing behaviour: how recently, how often, and how much did a spender buy.

- **Recency (R):** How recent was the customer's last shopping? Lower recency value indicates higher engagement and greater likelihood of responding to marketing programs. Higher recency value may suggest the customer is becoming inactive or even lost.
- **Frequency (F):** F value reflects how often a client makes purchases in a given time period. Higher frequency means the customer is regular and repeatedly engages with the brand, lower one may indicate one-time buyers or infrequent shoppers who require re-engagement.
- **Monetary (M):** M captures the total amount a customer has spent. Lower monetary value may indicate budget-conscious or low-value customers, while higher value suggests a high-value customer who contributes more revenue.

The R, F, M values are assigned on a scale from 1 to 5, with higher scores indicating more valuable customers. Using quintiles, we attain corresponding score per metric. The RFM score is made by combining the R, F, and M scores. For example, a shopper with R = 1, F = 5, M = 3 will have the RFM score = 153. The RFM scores have $5^3 = 125$ segments in total, which is too large to work with. In this paper, we divide customers into 10 segments based on R and F values. Then, we explore the total amount customers have spent within these segments.

TABLE II
CUSTOMER SEGMEN DISCRIPTION

Segment	R, F score	Description
Champions	R = 5 F = 4, 5	Bought recently, buy often the most
Loyal Customers	R = 3, 4 F = 4, 5	Buy on a regular basis. Responsive to promotions.
Potential Loyalist	R = 4, 5 F = 2, 3	Recent customers with average frequency.
Recent Customers	R = 5 F = 1	Bought most recently, but not often.
Promising	R = 4 F = 1	Recent shoppers, but haven't spent much.
Customers Needing Attention	R = 3 F = 3	Above average recency, frequency values. May not have bought very recently though.
About To Sleep	R = 3 F = 1, 2	Below average recency and frequency. Will lose them if not reactivated
At Risk	R = 1, 2 F = 3, 4	Purchased often but a long time ago. Need to bring them back!
Can't Lose Them	R = 1, 2 F = 5	Used to purchase frequently but haven't returned for a long time
Hibernating	R = 1, 2 F = 1, 2	Last purchase was long back and low number of orders. May be lost.

2.3 Data Pre-Processing

Data cleaner is an essential step to ensure a reliable analysis. Firstly, the null values were identified in the "Description" and "CustomerID" features. The dataset after removal of blank CustormerID instances has a size of 406,829. Then, we calculate the R, F, and M values of each customer as following:

1) **R value:** R value is computed based on InvoiceDate feature. The last purchase day is 09/12/2011, we choose a reference to calculate R value which is one day after the last purchase. Then, for each customer

$$R = \text{Reference Day} - \text{Last Invoice Day}$$

2) **F value:** In the dataset, each transaction has a unique invoice number. Based on CustomerID feature, we count the numbers of transactions per customer (total number of the customer's unique invoice numbers). Table III shows F values of 5 customers

TABLE III
CUSTOMER SAMPLE DATA F VALUES

CustomerID	F value
12346.0	2
12347.0	7
12348.0	4
12349.0	1
12350.0	1

3) **M value:** In the dataset, each instance is a transaction of a product that a customer has ordered. A shopper can buy multiple products, corresponding to several instances in the data. Based on Quantity and UnitPrice features, we make a new feature Price to get the revenue of each product per transaction by the formula $\text{Price} = \text{Quantity} \times \text{Unit Price}$. The M value is computed as the sum of Price of the customer's transactions.

Calculate the R, F, and M scores: Based on the R, F, M values of 3 metrics, we use quintiles to divide the value range of each metric into 5 segments, each contains 20% values of each metric. The value belonging to ith segment will get i score, $i = 1,2,3,4,5$. Finally, we attain the table of Recency, Frequency, Monetary values and R, F, M, RFM scores per customer as depicted in Table IV.

TABLE IV
RFM SCORES

CustomerID	Recency	Frequency	Monetary	R	F	M	RFM	Segment
12346.0	326	2	0.00	1	2	1	121	hibernating
12347.0	2	7	4310.00	5	4	5	545	champions
12348.0	75	4	1797.24	2	3	4	234	at risk
12349.0	19	1	1757.55	4	1	4	414	promising
12350.0	310	1	334.40	1	1	2	112	hibernating

III. RESULTS

Segment Monetary. We group customers into 10 segments based on R, F scores. We explore each segment with respect to 3 criteria: total number of its elements, the total revenue and the mean revenue the segment has contributed to the company. The results is shown in Table V.

TABLE V
SEGMENT MONETARY

Segment	No. of customers	Total Monetary	Mean Monetary
Champions	1244	486408.41	391.00
Loyal Customers	633	1982012.54	3131.14
Potential Loyalist	610	531024.96	870.53
Recent Customers	561	4190867.90	7470.35
Promising	440	516048.91	1172.83
Customers Needing Attention	431	196305.01	455.46
About To Sleep	190	190321.69	1001.69
At Risk	153	45956.36	300.36
Can't Lose Them	59	20639.61	349.82
Hibernating	51	140480.42	2754.51

Segment Size.

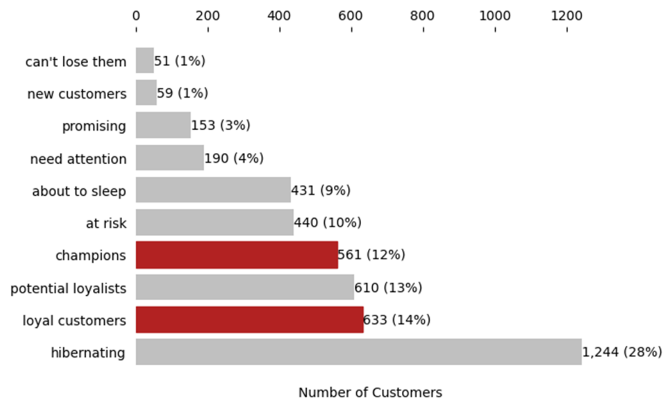

Fig. 1 Number of Customers per Segment

Figure 1 shows the number of shoppers in each segment. The largest customer group is the "hibernating" segment, accounting for 1,244 customers (28%), suggesting a significant portion of the customer base is inactive and potentially disengaged. On the other hand, high-value segments such as "loyal customers" and "champions" represent 14% (633 customers) and 12% (561 customers) respectively, indicating a strong core of engaged and potentially profitable customers. The "potential loyalists" also form a considerable share at 13% (610 customers), showing promise for future growth in loyalty. Segments such as "at risk" (10%), "about to sleep" (9%), and "need attention" (4%) highlight areas where customer retention strategies could be vital. Smaller segments like "can't lose them" (1%), "new customers" (1%), and "promising" (3%) suggest there are limited but valuable groups that may require focused engagement to foster loyalty or prevent churn. Overall, while the company has a healthy base of loyal customers, a significant portion is at risk of churn or already inactive, indicating an opportunity to implement targeted re-engagement and retention strategies.

Segment Revenue.

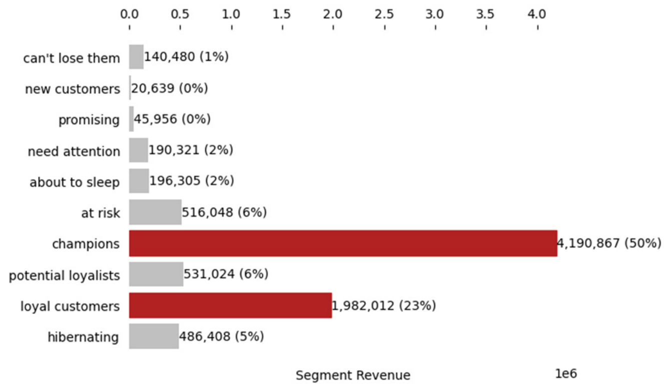

Fig. 2 Revenue per Segment

The bar plot in Figure 2 illustrates the revenue contribution of different customer segments. Despite comprising only 12% of the total customers, the "champions" segment generates the highest revenue at 4,190,867, accounting for a significant 50% of total revenue. This underlines their exceptional value and importance to the business. The "loyal customers" segment follows, contributing 1,982,012 (23% of total revenue) with 14% of the customer base, highlighting another critical group for sustained profitability. In contrast, the largest segment by customer count, "hibernating" (28%), contributes only 5% of revenue (486,408), indicating low spending and engagement. Similarly, segments such as "can't lose them," "new customers," and "promising" have negligible revenue impact despite their strategic importance, each contributing less than 1% of total revenue. Segments like "at risk" and "potential loyalists" contribute modestly at 6% each, suggesting they could offer revenue growth opportunities if properly nurtured. Overall, the data underscores the need to prioritize and retain high-value segments like "champions" and "loyal customers" while seeking to activate underperforming groups with potential.

IV. CONCLUSIONS

In this paper, we use RFM method to segment customers into meaningful and distinct groups based on behavioural metrics, enabling a clear understanding of the customer base. We conduct an in-depth analysis of customer behaviour within each segment, examining factors such as purchasing frequency, monetary value, and engagement levels. Finally, the study provides data-driven marketing strategy recommendations tailored to each segment, aiming to improve customer retention, boost revenue, and enhance the overall effectiveness of marketing efforts.

REFERENCES

[1] Christiana Jolaoso, *Customer Segmentation: The Ultimate Guide,* Forbes, Available: https://www.forbes.com/advisor/business/customer-segmentation/

[2] John JM, Shobayo O, Ogunleye B. An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics*. 2023; 2(4):809-823. https://doi.org/10.3390/analytics2040042

[3] Chen, D, *Online Retail [Dataset].* UCI Machine Learning Repository., 2015, Available: https://doi.org/10.24432/C5BW33.

[4] Guillaume Martin, RFM segmentation with Python, Available: https://guillaume-martin.github.io/rfm-segmentation-with-python.html

[5] Avinash Navlani, *Introduction to Customer Segmentation in Python*, DataCamp, 2018, Available: https://www.datacamp.com/tutorial/introduction-customer-segmentation-python