CUSTOMER SEGMENTATION REPORT

Author: Minh Quan

Contents

Introduction

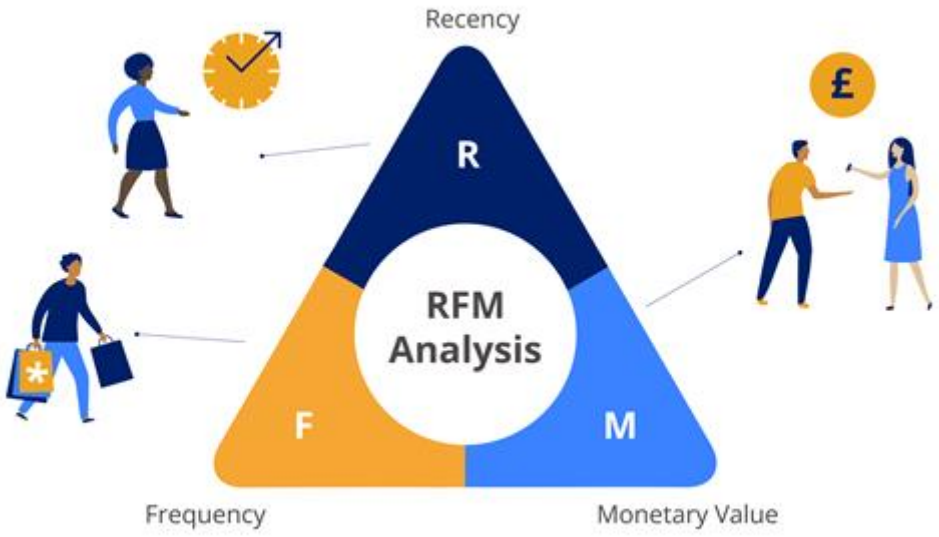
Companies cannot satisfy all potential buyers of their products or services. They should choose a group of customers that they can serve better than their competitors. In order to do this, it is necessary to segment customers and create their profiles. Customer segmentation is a tool to communicate effectively with customers. By means of the partitioning process, the characteristics of the hidden customer groups are defined in the data. Profiling is performed after customer segmentation. The main purpose of creating a customer profile is to make effective marketing strategies to target the most profitable customers. In this report, I will use the RFM method to segment the customers of a telecommunication company based on transaction data to help the company create customized marketing campaign for each segment.

Chapter I: Overview of RFM

1. What is RFM?

Recency, frequency, monetary value (RFM) is a model that segments a company's consumer base by their purchasing patterns or habits. The RFM model is based on three quantitative factors:

1. **Recency**: How recently a customer has made a purchase
2. **Frequency**: How often a customer makes a purchase
3. **Monetary value**: How much money a customer spends on purchases



RFM analysis numerically ranks a customer in each of these three categories. The "best" customer would receive a top score in every category.

2. Why is RFM analysis helpful?

Knowing a person's RFM score means that you can assign communications or automation capabilities appropriate to their relationship with you. RFM could, for example, support you in identifying a 'VIP' segment that you may wish to treat differently to a low-value customer segment, as the score identifies who your best, and worst, customers are. With this information, you can ensure that your VIP customers stay VIPs by providing them with early release stock or products they can't get as part of the wider mailing list. On the other hand, you could attempt to prevent customer churn and increase the length of a customer relationship with your organization by sharing promotional offers with customers who have a low RFM score. To sum it up, RFM helps optimize your marketing strategy budget, improve customer interactions, and increase revenue.

Chapter II: Methodology and analysis

To implementation RFM model for segmentation and profiling of the customer according to their value, the transaction data is obtained from a telecommunications company. Data is contained in two tables Customer_Registered and Customer_Transaction.

Table Customer_Registered records customers information about customer ID (ID), contract name (Contract), contract signing date (created_date),..

Column Name	#	Type	Length	Scale	Precision	Not Null	Identity	Default
ID	1	bigint	8		19	[ ]	[ ]	
Contract	2	varchar	100			[ ]	[ ]	
LocationID	3	int	4		10	[ ]	[ ]	
BranchCode	4	tinyint	1		3	[ ]	[ ]	
Status	5	tinyint	1		3	[ ]	[ ]	
created_date	6	datetime	8	3	23	[ ]	[ ]	
stopdate	7	datetime	8	3	23	[ ]	[ ]	

Customer_Registered table

Table Customer_Transaction contains information about transactions which includes transaction ID (ID), purchase date (Purchase_Date), transaction value (GMV), customer ID (CustomerID)

Column Name	#	Type	Length	Scale	Precision	Not Null	Identity	Default
ID	1	bigint	8		19	[ ]	[ ]	
Purchase_Date	3	datetime	8	3	23	[ ]	[ ]	
GMV	4	bigint	8		19	[ ]	[ ]	
CustomerID	2	varchar	200			[ ]	[ ]	

Customer_Transaction table

The dataset consists of the 1,048,056 customers' frequency of service registration, last register date and total spend amount until September 1, 2022. The last register date is used for Recency, register frequency is used as Frequency and total spent amount is used for Monetary Value.

To calculate RFM score of each customer, RFM data is encoded in 4 equal portions. Customer's register date that is sorted in descending order of timeliness case. The top 25% segment is coded as 4, while the next 25% segment is coded as 3 and so forth. The frequency of customer are also sorted in descending order. All customers conferred by 444, 443,...111. The customers are now divided into 64(4*4*4) groups. Customers who have the highest RFM score are the most profitable customers.

1. Calculate Recency, Frequency, Monetary score

Process:

- Join two tables Customer_Registered and Customer_Transaction using JOIN function to get information.
- Calculate Recency: Total days since the customer's last purchase to September 1, 2022

- Calculate Frequency: Total purchase times/contract age
- Calculate Monetary: Total spent amount of a customer

	CustomerID	Recency	Frequency	Monetary
1	900290	1,450	1	568,182
2	1014520	1,356	1	568,182
3	570380	1,749	0.8	568,182
4	238554	2,343	0.67	568,182
5	272252	2,282	0.67	568,182
6	824132	1,512	1.5	518,182
7	651068	1,669	1	508,182
8	218218	2,381	0.67	508,182
9	585209	1,732	0.8	503,182
10	843859	1,494	1.25	503,182
11	706353	1,609	1	488,182
12	784317	1,541	1	488,182

Customers with their RFM

- Based on IQR (Interquartile Range), after dividing the amount of R-F-M data into 4 equal parts, scores of Recency, Frequency, Monetary will be evaluated on a scale of 1 to 4 with 4 being the highest score. The higher the value, the higher the score. For Recency, it's the opposite cause it indicates that customers have not registered for a long time.

Score	1	2	3	4
Recency	>=1975	1975 - 1663	1663 - 1475	1475 - 1333
Frequency	0.14 – 0.2	0.2 – 0.25	0.25	0.25 – 1.5
Monetary	<=75,000	75,000 - 85000	85,000 – 105,000	105,000 – 568,182

	customerid	Recency	Frequency	Monetary	rec_score	fre_score	mon_score
4	156041	2,493	0.14	0	1	1	1
5	167518	2,469	0.14	0	1	1	1
6	170846	2,464	0.14	0	1	1	1
7	178884	2,449	0.14	0	1	1	1
8	180263	2,446	0.14	0	1	1	1
9	180274	2,446	0.14	0	1	1	1
10	180279	2,446	0.14	0	1	1	1
11	180255	2,446	0.14	0	1	1	1
12	180314	2,446	0.14	0	1	1	1
13	184355	2,439	0.14	0	1	1	1
14	188049	2,432	0.17	0	1	1	1
15	188058	2,432	0.17	0	1	1	1

Group of each customer based on thei RFM score

2. Customer segmentation based on RFM score

Customer Segmentation	RFM groups	Characteristic
VIP	444, 443, 434, 433, 344, 343, 334	A group of customers who use the service regularly and spend a lot
LOYAL	441, 431, 424, 423, 414, 413, 342, 332, 323, 324, 331, 243, 244, 234, 432, 442, 333	Customers who have used the service recently, spend a decent amount of money and have multiple purchases
POTENTIAL	441, 431, 424, 423, 414, 413, 342, 332, 323, 324, 331, 243, 244, 234, 432, 442, 333	Customers who have recently purchased, spend not too much money, do not use service frequently
LOST	111, 112, 113, 114, 121, 122, 123, 124, 131, 132, 133, 134, 141, 142, 143, 144, 211, 212, 213, 214, 221, 222, 223, 224, 231, 232, 233, 241, 242	Customers have used the service but have not returned for a long time

3. Recommendation for each segment

VIP and LOYAL: maintaining their loyalty and maximizing their value to the business

- Give them special benefits such as special offers or early access to the product
- Implement membership campaigns
- Keep in touch with VIP customers regularly through newsletters, personalized emails, or phone calls.
- Introducing new products

POTENTIAL: to convert them into loyal clients

- Create limited special promotions
- Implement campaigns to increase engagement
- Send follow-up emails or messages after initial contact to keep the conversation going.

LOST: reconnect with lost customer by

- Solve their problem while using our products
- Create free trial policies
- Suggestions based on products they have purchased

Code Reference

```
-- Cac chi so Recency, Frequency, Monetary

WITH abc as(

SELECT CustomerID,

        DATEDIFF(day,MAX(cast(created_date as date)), '2022-9-1') as Recency,

        ROUND(CAST(COUNT(Purchase_date) AS FLOAT)/CAST(DATEDIFF(YEAR,
CAST(created_date AS DATE), '2022-09-01') AS FLOAT), 2) AS Frequency,

        SUM(GMV) as Monetary,

        ROW_NUMBER () over(order by DATEDIFF(day,MAX(cast(created_date as
date)), '2022-9-1') desc) as recency_order,

        ROW_NUMBER () over(order by ROUND(CAST(COUNT(Purchase_date) AS
FLOAT)/CAST(DATEDIFF(YEAR, CAST(created_date AS DATE), '2022-09-01') AS FLOAT), 2)) as
frequency_order,

        ROW_NUMBER () over(order by SUM(GMV)) as monetary_order

FROM Customer_Transaction ct

JOIN Customer_Registered cr

ON ct.CustomerID = cr.ID

WHERE ct.CustomerID != 0

GROUP BY CustomerID, created_date

)

-- Tinh diem RFM dua tren row_number duoc sap theo gia tri cua cac chi so R,F,M

SELECT customerid, Recency, Frequency, Monetary,

CASE

        when recency_order <= (SELECT max(recency_order)*0.25 FROM abc) then 1

        when recency_order <= (SELECT max(recency_order)*0.5 FROM abc) then 2

        when recency_order <= (SELECT max(recency_order)*0.75 FROM abc) then 3

        else 4

END as rec_score,

CASE
```

```
        when frequency_order <= (SELECT max(frequency_order)*0.25 FROM abc) then 1
        when frequency_order <= (SELECT max(frequency_order)*0.5 FROM abc) then 2
        when frequency_order <= (SELECT max(frequency_order)*0.75 FROM abc) then 3
        else 4
END as fre_score,
CASE
        when monetary_order <= (SELECT max(monetary_order)*0.25 FROM abc) then 1
        when monetary_order <= (SELECT max(monetary_order)*0.5 FROM abc) then 2
        when monetary_order <= (SELECT max(monetary_order)*0.75 FROM abc) then 3
        else 4
END as mon_score
into rfm
From abc
-- Phan loai khach hang theo tung nhom RFM
SELECT *,CONCAT(rec_score, fre_score, mon_score) as RFM,
CASE
        when CONCAT(rec_score, fre_score, mon_score) in
('444','443','434','433','344','343','334') then 'VIP'
        when CONCAT(rec_score, fre_score, mon_score) in
        ('441','431','424','423','414','413','342','332','323','324','331','331','243','244','234','432'
,'442','333') then 'LOYAL'
        when CONCAT(rec_score, fre_score, mon_score) in
        ('422','421','412','411','311','312','313','314','321','322','341') then 'POTENTIAL'
        else 'LOST'
END as Segmentation
FROM rfm
```