

PHÂN LOẠI KHÁCH HÀNG

SỬ DỤNG

MÔ HÌNH

RFM

Tác giả:

Thái Trung Kiên

kienthai2711@gmail.com

0945613721



Mục lục

1. Đặt vấn đề.....	3
2. Kiến thức sử dụng.....	4
2.1. Customer 360.....	4
2.1.1. Giới thiệu về Customer 360.....	4
2.1.2. Lợi ích của Customer 360.....	5
2.2. Mô hình RFM.....	5
2.2.1. Các khái niệm cơ bản trong phân tích RFM Customer 360.....	5
Recency (Thời gian mua hàng gần nhất).....	7
Frequency (Tần suất mua hàng).....	7
Monetary (Giá trị tiền mỗi lần mua hàng).....	8
2.2.2. Lợi ích của phân tích RFM Customer 360.....	9
2.3. Tính điểm RFM sử dụng từ phân vị trong xác suất thống kê.....	9
2.4. Phân khúc khách hàng.....	11
3. Các bước thực hiện.....	13
3.1. Quy trình tổng quan.....	13
3.2. Data schema.....	13
3.3. Làm sạch dữ liệu.....	14
3.4. Phân nhóm RFM.....	14
3.5. Phân nhóm khách hàng.....	16
3.6. Visualize và phân tích.....	16
3.7. SQL Query.....	16
4. Visualize và phân tích.....	23
4.1. Thông tin chung.....	23
4.2. Phân tích chi tiết.....	24
5. Kết luận.....	27

1. Đặt vấn đề

Phân loại khách hàng là một phần không thể thiếu trong bất kỳ chiến lược kinh doanh nào của doanh nghiệp. Phân loại khách hàng giúp doanh nghiệp hiểu rõ hơn về nhu cầu, sở thích và hành vi mua hàng của từng nhóm khách hàng. Điều này giúp doanh nghiệp phát triển chiến lược marketing và bán hàng linh hoạt và hiệu quả nhất, đồng thời tăng cường hiệu quả vận hành doanh nghiệp và tăng cường lợi nhuận. Hơn nữa, phân loại khách hàng giúp tổ chức tập trung nỗ lực vào việc phục vụ nhóm khách hàng có tiềm năng cao nhất để đạt hiệu suất kinh doanh cao nhất. Việc này giúp doanh nghiệp xây dựng mối quan hệ chặt chẽ và lâu dài với khách hàng, đồng thời tạo ra trải nghiệm mua hàng tốt nhất cho họ, từ đó giúp doanh nghiệp nắm bắt và phát triển thị trường một cách hiệu quả và bền vững.

Do đó, mục tiêu của báo cáo này là đề xuất các phân nhóm khách hàng dựa trên lịch sử giao dịch của khách hàng khi sử dụng dịch vụ của doanh nghiệp bằng cách sử dụng mô hình phân tích RFM và các đánh giá, quan sát dựa theo tình hình thực tế của doanh nghiệp. Các phần còn lại của báo cáo được trình bày như sau:

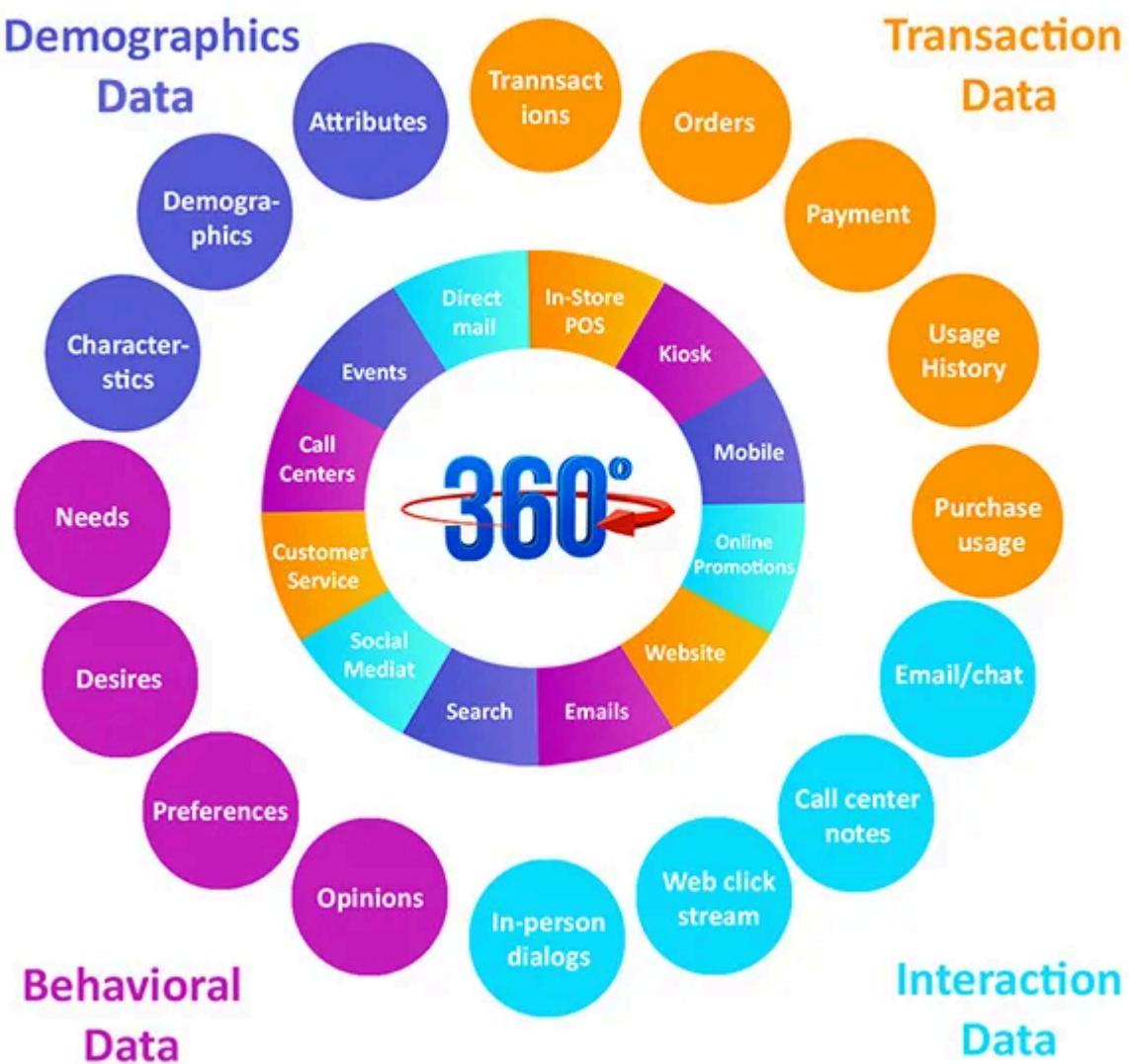
- 1. Phần 2: Mô tả các kiến thức chung sử dụng làm tiền đề để phân tích dữ liệu khách hàng.**
- 2. Phần 3: Mô tả các bước thực hiện áp dụng các kiến thức chung để tìm ra phân nhóm khách hàng.**
- 3. Phần 4: Phân tích kết quả quan sát được dựa trên dữ liệu và đưa ra định hướng hành động cho tổ chức.**
- 4. Phần 5: Kết luận.**

2. Kiến thức sử dụng

2.1. Customer 360

2.1.1. Giới thiệu về Customer 360

Customer 360 là chiến lược phân tích dựa trên dữ liệu giao dịch (Transaction Data), dữ liệu tương tác trên website hoặc ứng dụng (Interaction Data), hành vi tiêu dùng và nhu cầu của khách hàng (Behavioral Data), và dữ liệu về độ tuổi và nhân khẩu học của khách hàng (Demographics Data).



Hình 1. Mô hình Customer 360.

Trong mô hình Customer 360, có nhiều loại dữ liệu khác nhau được tích hợp từ nhiều nguồn để tạo ra một cái nhìn toàn diện về khách hàng. Hình 1 mô tả 4 nhóm dữ liệu chính trong mô hình này. Bốn nhóm dữ liệu đó bao gồm:

Khía cạnh dữ liệu	Mô tả
Demographic Data	Dữ liệu nhân khẩu học như vị trí địa lý, thông tin cá nhân, và thông tin nhạy cảm.
Transaction Data	Dữ liệu giao dịch liên quan đến các hoạt động mua bán của khách hàng.
Interaction Data	Dữ liệu tương tác của khách hàng với các nền tảng trực tuyến, thường là big data.
Behavior Data	Dữ liệu hành vi của khách hàng, bao gồm phản hồi và đánh giá dịch vụ.

2.1.2. Lợi ích của Customer 360

Customer 360 cung cấp cái nhìn toàn diện về khách hàng và các tương tác của họ với công ty:

- Tạo ra các chiến dịch bán hàng và tiếp thị chiến lược hơn:**

Một cái nhìn 360 độ về khách hàng mở rộng khả năng của đội ngũ bán hàng và tiếp thị. Bằng cách tận dụng dữ liệu hiện có, các đội ngũ này có thể thiết kế các chiến dịch chiến lược hơn.

- Hiểu rõ khách hàng và hành vi của họ:**

Với cái nhìn 360 độ, doanh nghiệp có thể hiểu sâu về sở thích và hành vi của khách hàng.

- Cung cấp trải nghiệm mua sắm cá nhân hóa hơn:**

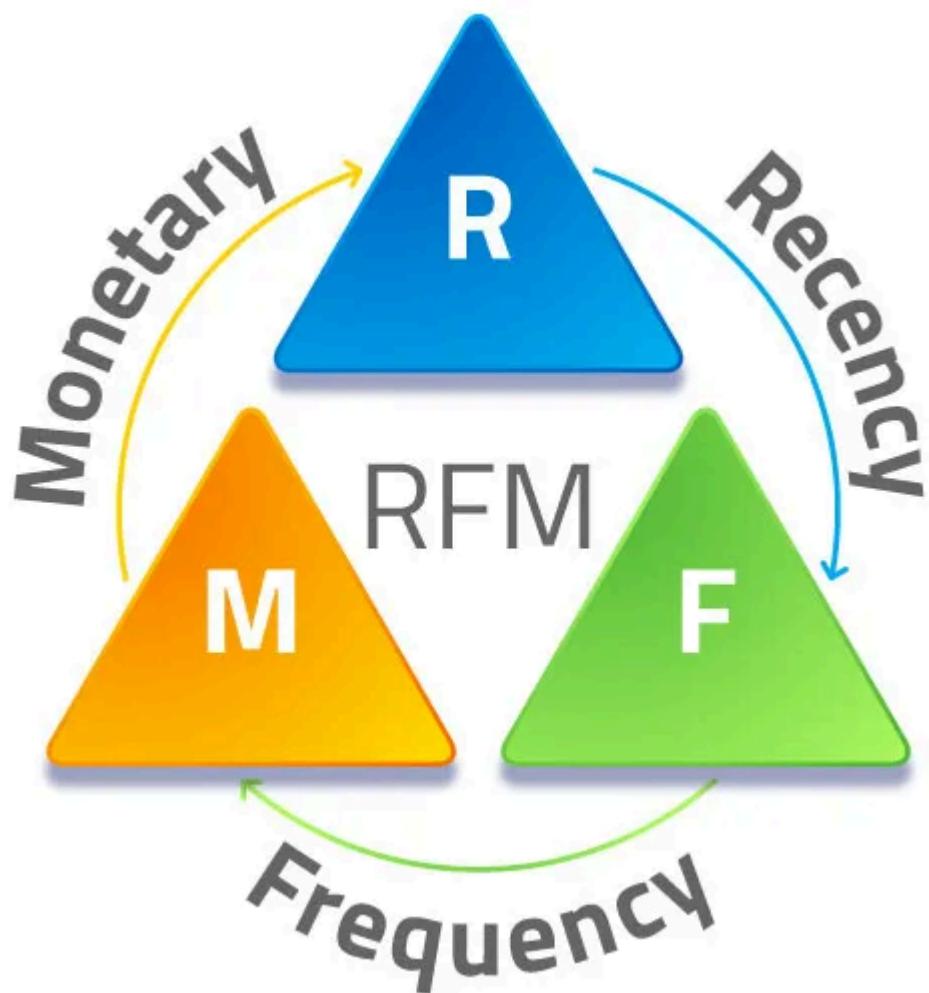
Lợi ích quan trọng nhất là khả năng cung cấp trải nghiệm mua sắm cá nhân hóa.

2.2. Mô hình RFM

2.2.1. Các khái niệm cơ bản trong phân tích RFM Customer 360

RFM Customer 360 là phương pháp phân tích khách hàng dựa trên mô hình RFM (Recency - Frequency - Monetary). Phương pháp này giúp các doanh nghiệp đánh giá giá trị của từng khách hàng dựa trên mức độ tương tác của họ với doanh nghiệp và đưa ra các chiến lược tiếp cận khác nhau để tối ưu hóa việc chăm sóc khách hàng. Mô hình RFM đánh giá các khía cạnh quan trọng của hành vi mua hàng của khách hàng dựa trên 3 yếu tố:

- Recency (Thời gian gần nhất mua hàng)
- Frequency (Tần suất mua hàng)
- Monetary (Giá trị doanh thu trên mỗi lần mua hàng)



Hình 2. Mô hình phân tích RFM.

Recency (Thời gian mua hàng gần nhất)

Recency được **đo thời gian từ lần mua hàng gần nhất của khách hàng**.

Một khách hàng được xem là có giá trị cao hơn nếu lần mua hàng gần nhất của họ xảy ra gần đây hơn. Ví dụ, một khách hàng mới mua hàng trong tháng trước sẽ có mức Recency cao hơn so với một khách hàng mua hàng cách đây một năm. Recency được sử dụng để phân loại khách hàng thành các nhóm khác nhau như:

- **Khách hàng mới:** Lần mua hàng gần nhất xảy ra gần đây và có thể được coi là tiềm năng cho việc tương tác và tiếp thị tiếp theo.
- **Khách hàng thường xuyên:** Lần mua hàng gần nhất xảy ra không quá lâu và đánh dấu một mức độ tương tác liên tục với doanh nghiệp.
- **Khách hàng có khả năng rời bỏ:** Lần mua hàng gần nhất xảy ra cách đây lâu và đây có thể là tín hiệu cho sự suy giảm quan tâm của khách hàng hoặc mất dần liên hệ với doanh nghiệp.
- **Khách hàng không hoạt động:** Lần mua hàng gần nhất xảy ra rất lâu và khách hàng không có hoạt động mới. Nhóm này thường đại diện cho khách hàng không còn tương tác hoặc mất quan tâm đến doanh nghiệp.

Bằng cách xác định thời gian gần nhất mua hàng, doanh nghiệp có thể tập trung vào việc tương tác và tiếp cận khách hàng theo cách phù hợp, nhằm tăng cường quan hệ và gia tăng giá trị từ mỗi khách hàng.

Frequency (Tần suất mua hàng)

Frequency **đếm số lần mua hàng của khách hàng trong một khoảng thời gian nhất định, đánh giá mức độ tương tác lặp lại của khách hàng với doanh nghiệp**. Một khách hàng được coi là có giá trị cao hơn nếu họ mua hàng thường xuyên hơn. Số lần mua hàng của khách hàng trong một khoảng thời gian cụ thể sẽ quyết định mức độ Frequency của họ. Ví dụ, một khách hàng mua hàng hàng tháng sẽ có mức Frequency cao hơn so với một khách hàng chỉ mua

hàng một lần mỗi năm. Frequency được sử dụng để phân loại khách hàng thành các nhóm khác nhau như:

- **Khách hàng thường xuyên:** Mua hàng với tần suất cao trong một khoảng thời gian nhất định. Đây là nhóm khách hàng quan trọng và có tiềm năng mang lại doanh thu lớn.
- **Khách hàng định kỳ:** Mua hàng với tần suất ổn định trong một khoảng thời gian nhất định, không quá thường xuyên. Đây là nhóm khách hàng có thể tiềm năng phát triển và đạt được mức độ thường xuyên hơn.
- **Khách hàng ngẫu nhiên:** Mua hàng không có một mô hình đều đặn, không có tần suất cụ thể. Đây là nhóm khách hàng khá khó dự đoán và cần quan tâm để tăng cường tương tác.
- **Khách hàng không thường xuyên:** Mua hàng rất ít lần hoặc không mua hàng trong một khoảng thời gian dài. Đây là nhóm khách hàng không phải ưu tiên tiếp thị, và có thể yêu cầu nỗ lực đặc biệt để thúc đẩy họ tăng tần suất mua hàng.

Monetary (Giá trị tiền mỗi lần mua hàng)

Monetary đo **giá trị tiền hoặc giá trị đơn hàng mà khách hàng đã chi tiêu trong mỗi lần mua hàng**, thể hiện **mức độ giá trị mà khách hàng mang lại cho doanh nghiệp**. Tùy vào nhu cầu phân tích dữ liệu của doanh nghiệp cho từng chiến dịch marketing, giá trị tiền được đo bằng tổng số tiền mà khách hàng đã chi trả hoặc tổng giá trị đơn hàng mỗi lần giao dịch. Monetary được sử dụng để phân loại khách hàng thành các nhóm khác nhau như:

- **Khách hàng có giá trị cao:** Chi tiêu nhiều tiền trong mỗi lần mua hàng, đóng góp lớn vào doanh thu của doanh nghiệp. Đây là nhóm khách hàng quan trọng và cần được tạo điều kiện thuận lợi để họ có thể duy trì mức chi tiêu cao.

- **Khách hàng có giá trị trung bình:** Chi tiêu một số tiền vừa phải trong mỗi lần mua hàng. Đây là nhóm khách hàng có tiềm năng phát triển và có thể tăng giá trị bằng cách tăng cường tương tác và tiếp thị.
- **Khách hàng có giá trị thấp:** Chi tiêu ít tiền trong mỗi lần mua hàng. Đây là nhóm khách hàng có thể không mang lại lợi nhuận lớn cho doanh nghiệp và có thể cần nỗ lực để tăng cường giá trị mỗi giao dịch.

2.2.2. Lợi ích của phân tích RFM Customer 360

- **Hiểu rõ hơn về khách hàng:** Phân tích RFM giúp doanh nghiệp có cái nhìn tổng quan về khách hàng của họ.
- **Phân loại khách hàng và tối ưu hóa chiến dịch tiếp thị:** Sau khi xếp hạng từng khách hàng theo tiêu chí RFM, doanh nghiệp có thể tối ưu hóa chiến lược tiếp thị.
- **Giảm chi phí tiếp thị:** Phân tích RFM giúp doanh nghiệp tập trung vào các nhóm khách hàng tiềm năng.

2.3. Tính điểm RFM sử dụng từ phân vị trong xác suất thống kê

Từ phân vị trong xác suất thống kê là cách chia dữ liệu thành bốn phần bằng nhau dựa trên thứ tự của chúng. Cụ thể, từ phân vị chia dữ liệu thành bốn phần có cùng số lượng quan sát, tức là mỗi phần chứa 25% tổng số quan sát.

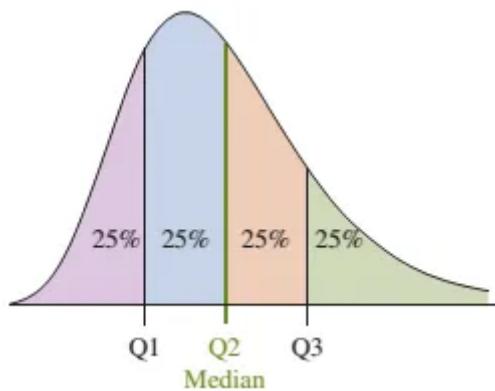
Tứ phân vị có 3 giá trị, đó là tứ phân vị thứ nhất (Q1), thứ hai (Q2) và thứ ba (Q3). Ba giá trị này chia một tập hợp dữ liệu (đã sắp xếp dữ liệu theo trật tự từ bé đến lớn) thành 4 phần có số lượng quan sát đều nhau.

Giả sử tập dữ liệu có n quan sát.

Tứ phân vị thứ nhất được tính bằng công thức $Q1 = 25 * (n+1) / 100$.

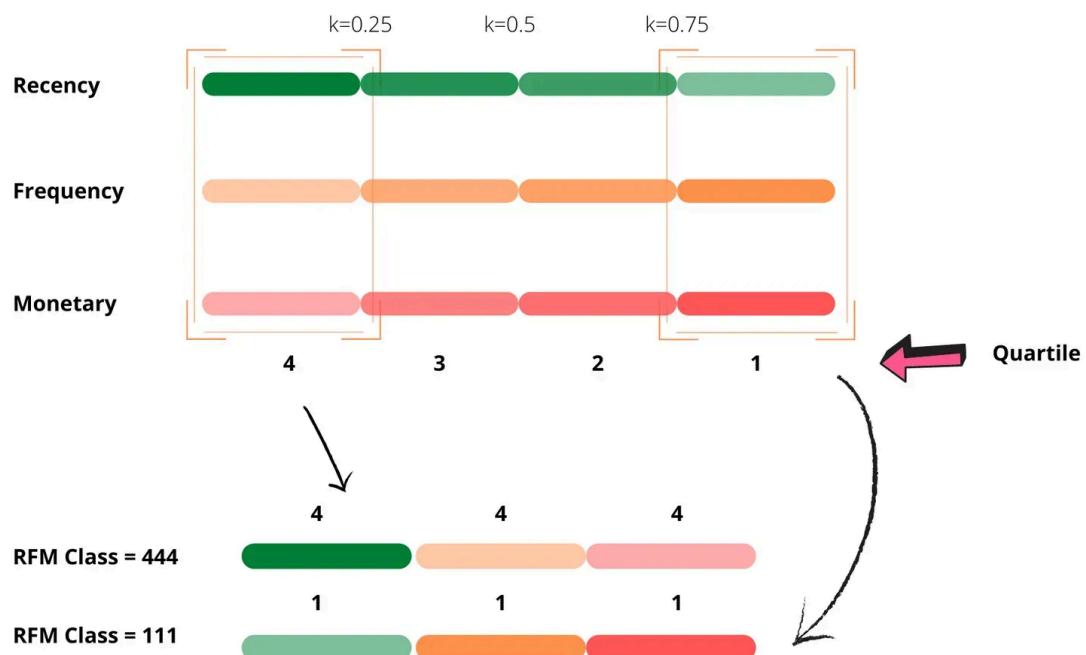
Tứ phân vị thứ hai là giá trị trung vị $Q2 = (n+1)/2$

Tứ phân vị thứ ba được tính bằng công thức $Q3 = 75 * (n+1) / 100$.



Hình 3. Tứ phân vị.

Với mỗi biến (R, F, M), chúng ta chia thành các nhóm từ 1 đến 4, với nhóm 4 đại diện cho các giá trị cao nhất và nhóm 1 đại diện cho các giá trị thấp nhất. Riêng với R, điểm 4 có nghĩa là khách hàng mới mua hàng gần đây, và 1 là khách hàng đã rất lâu chưa có phát sinh giao dịch với doanh nghiệp.



Hình 4. Quy tắc tính điểm RFM.

2.4. Phân khúc khách hàng

Dựa trên điểm RFM, khách hàng của một doanh nghiệp có thể được chia thành các phân khúc khác nhau như sau:

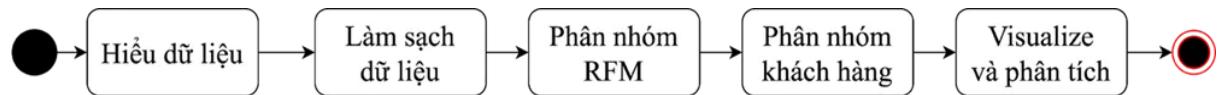
Nhóm khách hàng	Mô tả	Nhóm RFM
Champions (nhóm khách hàng VIP)	Đây là nhóm khách hàng tốt nhất, họ đã mua hàng gần đây nhất, mua hàng thường xuyên nhất và chi tiêu nhiều tiền nhất.	444, 443, 434, 344
Loyal Customer (nhóm khách hàng trung thành)	Khách hàng trung thành với thương hiệu; có thể họ không chi tiêu nhiều bằng các khách hàng VIP, nhưng họ thường xuyên sử dụng dịch vụ và sử dụng dịch vụ trong thời gian gần đây	442, 441, 432, 431, 433, 343, 342, 341
Potential Loyalist (nhóm khách hàng tiềm năng)	Khách hàng có tiềm năng trở thành các khách hàng trung thành; họ sử dụng dịch vụ ít thường xuyên hơn, tuy nhiên có mức chi tiêu cao	424, 423, 324, 323, 413, 414, 343, 334
Promising (nhóm khách hàng hứa hẹn)	Khách hàng chi tiêu ít hơn khách hàng trung thành, nhưng sử dụng dịch vụ khá thường xuyên và khá gần đây	333, 332, 331, 313
New Customer (nhóm khách hàng mới)	Khách hàng mới sử dụng dịch vụ gần đây; họ chưa có nhiều cơ hội mua hàng nên chi tiêu chưa nhiều và chưa thường xuyên	422, 421, 412, 411, 311, 321, 312, 322
Price Sensitive (nhóm khách hàng cần chú ý về giá)	Khách hàng chi tiêu thường xuyên, tuy nhiên mức chi tiêu thấp, có thể do họ khó tính hoặc tiết kiệm trong việc chi tiêu	131, 132, 141, 142, 231, 232, 241, 242

Needs Attention (nhóm khách hàng cần tiếp cận lại)	Khách hàng đã chi tiêu nhiều và sử dụng dịch vụ thường xuyên trong quá khứ, tuy nhiên đã 1 thời gian rồi chưa sử dụng dịch vụ	244, 234, 243, 233, 224, 214, 213, 134, 144, 143, 133
At Risk Customers (nhóm khách hàng có nguy cơ rời bỏ)	Khách hàng đã chi tiêu nhiều và sử dụng dịch vụ thường xuyên trong quá khứ, tuy nhiên đã 1 thời gian rồi chưa sử dụng dịch vụ	244, 234, 243, 233, 224, 214, 213, 134, 144, 143, 133
Lost Customer (Nhóm khách hàng đã rời bỏ)	Khách hàng đã sử dụng dịch vụ tuy nhiên không còn quay lại nữa	111, 112, 113, 114, 121, 122, 123, 221, 211, 222

3. Các bước thực hiện

3.1. Quy trình tổng quan

Hình 5 mô tả các bước thực hiện để phân nhóm khách hàng dựa trên dữ liệu lịch sử.



Hình 5. Quy trình tổng quan.

3.2. Data schema

Trong trường hợp này, có hai bảng cần phân tích: bảng **Customer_Transaction** và bảng **Customer_Registered**.

- Bảng **Customer_Register** chứa thông tin khi đăng ký của khách hàng: Bảng **Customer_Registered** ghi nhận thông tin về các khách hàng đã đăng ký thẻ thành viên.

STT	Trường	Mô tả	Kiểu dữ liệu
1	ID	Mã khách hàng	Int
2	Contract	Mã hợp đồng	Varchar
3	LocationID	Mã vị trí	Int
4	BranchCode	Mã chi nhánh	Int
5	Status	Trạng thái	Int
6	created_date	Ngày đăng ký	Datetime
7	stop_date	Ngày huỷ	Datetime

- Bảng **Customer_Transaction** chứa thông tin giao dịch của khách hàng: Bảng **Customer_Transaction** ghi lại các giao dịch của khách hàng từ tháng 6/2022 đến tháng 8/2022. Mỗi hàng trong bảng tương ứng với một nhiệm vụ giao dịch. Mỗi giao dịch ghi nhận thông tin về khách hàng, ngày mua hàng, và số tiền.

STT	Trường	Mô tả	Kiểu dữ liệu
1	ID	Mã giao dịch	Int
2	CustomerID	Mã khách hàng	Int
3	Purchase_Date	Ngày giao dịch	DateTime
4	GMV	Gross Monetary Value - Số tiền giao dịch	Int

3.3. Làm sạch dữ liệu

Trong bước này, chúng ta loại bỏ các giá trị có thể làm nhiễu dữ liệu như sau:

- Loại bỏ các khách hàng có **ID=0** vì đây là account test của hệ thống
- Loại bỏ các dữ liệu không có thông tin về **created_date** hoặc ngày **created_date rỗng**
- Loại bỏ các KH có **stop_date != NULL** vì các khách hàng này đã ngừng sử dụng dịch vụ hiện tại
- Loại bỏ các giao dịch không có phát sinh doanh thu, tức **GMV bằng 0**.

3.4. Phân nhóm RFM

- Công cụ: MS SQL SERVER
- Quy trình:
 - Xử lý dữ liệu OLTP thành OLAP từ bảng **Customer_Transaction**.

- Sử dụng **JOIN** để kết hợp hai bảng **Customer_Registered** và **Customer_Transaction**.
- Sử dụng các hàm **SUM**, **DATEDIFF**, **WINDOWFUNCTION**, **CASE WHEN**.
- Để phân nhóm RFM, chúng ta thực hiện theo các bước như sau:
 - **Bước 1:** Tính các giá trị R, F, M đối với từng khách hàng
 - **R:** Thời gian mua hàng gần nhất, được tính bằng số ngày giữa ngày hiện tại và lần mua gần nhất của khách hàng.
 - **F:** Tần suất mua hàng. Được tính bằng tổng số đơn hàng khách hàng mua đã mua theo tuổi hợp đồng. Tuổi hợp đồng được tính từ ngày khách hàng đăng ký đến ngày hiện tại. Việc tính theo tuổi hợp đồng đảm bảo các khách hàng có thời gian sử dụng khác nhau không bị thiên vị.
 - **M:** Số tiền khách đã chi khi mua sắm các sản phẩm hoặc sử dụng dịch vụ của doanh nghiệp. Chỉ số này cũng được tính theo tuổi hợp đồng của khách hàng.
 - **Bước 2:** Tính điểm R, F, M theo tứ phân vị.
 - Chia dữ liệu R-F-M thành 4 phần bằng nhau (quartiles), được đánh số từ 1 đến 4.
 - Trong trường hợp này, chúng ta quy ước rằng số càng lớn, giá trị càng tốt (Số 4 là tốt nhất).
 - Đối với Recency, giá trị càng cao thì càng kém, chỉ ra rằng khách hàng đã không quay lại mua hàng trong một thời gian dài.
 - **Bước 3:** Tính nhóm RFM bằng cách ghép các điểm R, F, M lại với nhau.

3.5. Phân nhóm khách hàng

Sau khi có điểm RFM, thực hiện phân nhóm khách hàng theo bảng map về phân khúc khách hàng đã trình bày ở mục 2.4.

3.6. Visualize và phân tích

Sử dụng **PowerBI** để trình bày dashboard và phân tích kết quả thu được. Chi tiết về kết quả phân tích được thực hiện trong phần 4.

3.7. SQL Query

```
-- Tính giá trị RFM

SELECT
    CustomerID AS customer_id,
    MIN(CAST(created_date AS DATE)) AS created_date,
    COUNT(DISTINCT(CAST(Purchase_Date AS DATE))) AS total_purchases,
    DATEDIFF(YEAR, MIN(CAST(created_date AS DATE)), '2022-09-01') + 1 AS contract_year,
    DATEDIFF(DAY, MAX(CAST(Purchase_Date AS DATE)), '2022-09-01') AS recency,
    1.0 * COUNT(DISTINCT(CAST(Purchase_Date AS DATE))) / (DATEDIFF(YEAR, MIN(CAST(created_date AS DATE)), '2022-09-01') + 1) AS frequency,
    SUM(CAST(GMV AS BIGINT)) / (DATEDIFF(YEAR, MIN(CAST(created_date AS DATE)), '2022-09-01') + 1) AS monetary,
    ROW_NUMBER() OVER (ORDER BY DATEDIFF(DAY, MAX(CAST(Purchase_Date AS DATE)), '2022-09-01')) AS rn_recency,
```

```

        ROW_NUMBER() OVER (ORDER BY 1.0 *)
COUNT(DISTINCT(CAST(Purchase_Date AS DATE))) /
(DATEDIFF(YEAR, MIN(CAST(created_date AS DATE)), '2022-09-01') +
1)) AS rn_frequency,
        ROW_NUMBER() OVER (ORDER BY SUM(CAST(GMV AS BIGINT)) /
(DATEDIFF(YEAR, MIN(CAST(created_date AS DATE)), '2022-09-01') +
1)) AS rn_monetary
INTO #rfm_raw
FROM Customer_Registered cr
JOIN Customer_Transaction ct ON cr.ID = ct.CustomerID
WHERE created_date < '2022-09-01'
    AND created_date != "
    AND created_date IS NOT NULL
    AND stopdate != "
    AND stopdate IS NOT NULL
    AND cr.ID != 0
    AND ct.GMV > 0
GROUP BY CustomerID
HAVING SUM(GMV) > 0;

-- Tính toán điểm RFM
SELECT
    *,

```

CASE

WHEN recency >= (SELECT recency FROM #rfm_raw WHERE rn_recency = 1)

AND recency < (SELECT recency FROM #rfm_raw WHERE rn_recency = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.25 AS INT) FROM #rfm_raw))

THEN '4'

WHEN recency >= (SELECT recency FROM #rfm_raw WHERE rn_recency = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.25 AS INT) FROM #rfm_raw))

AND recency < (SELECT recency FROM #rfm_raw WHERE rn_recency = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.5 AS INT) FROM #rfm_raw))

THEN '3'

WHEN recency >= (SELECT recency FROM #rfm_raw WHERE rn_recency = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.5 AS INT) FROM #rfm_raw))

AND recency < (SELECT recency FROM #rfm_raw WHERE rn_recency = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.75 AS INT) FROM #rfm_raw))

THEN '2'

ELSE '1'

END AS R,

CASE

WHEN frequency >= (SELECT frequency FROM #rfm_raw WHERE

```

rn_frequency = 1)

        AND frequency < (SELECT frequency FROM #rfm_raw WHERE
rn_frequency = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.25
AS INT) FROM #rfm_raw))

        THEN '1'

        WHEN frequency >= (SELECT frequency FROM #rfm_raw WHERE
rn_frequency = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.25
AS INT) FROM #rfm_raw))

        AND frequency < (SELECT frequency FROM #rfm_raw WHERE
rn_frequency = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.5 AS
INT) FROM #rfm_raw))

        THEN '2'

        WHEN frequency >= (SELECT frequency FROM #rfm_raw WHERE
rn_frequency = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.5 AS
INT) FROM #rfm_raw))

        AND frequency < (SELECT frequency FROM #rfm_raw WHERE
rn_frequency = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.75
AS INT) FROM #rfm_raw))

        THEN '3'

    ELSE '4'

END AS F,
CASE

    WHEN monetary >= (SELECT monetary FROM #rfm_raw WHERE
rn_monetary = 1)

        AND monetary < (SELECT monetary FROM #rfm_raw WHERE

```

```

rn_monetary = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.25 AS
INT) FROM #rfm_raw))

THEN '1'

WHEN monetary >= (SELECT monetary FROM #rfm_raw WHERE
rn_monetary = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.25 AS
INT) FROM #rfm_raw))

AND monetary < (SELECT monetary FROM #rfm_raw WHERE
rn_monetary = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.5 AS
INT) FROM #rfm_raw))

THEN '2'

WHEN monetary >= (SELECT monetary FROM #rfm_raw WHERE
rn_monetary = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.5 AS
INT) FROM #rfm_raw))

AND monetary < (SELECT monetary FROM #rfm_raw WHERE
rn_monetary = (SELECT CAST(COUNT(DISTINCT customer_id) * 0.75 AS
INT) FROM #rfm_raw))

THEN '3'

ELSE '4'

END AS M

INTO #rfm_score

FROM #rfm_raw;

-- Kết hợp điểm R, F và M

SELECT

```

```

*,  

CONCAT(R, F, M) AS rfm  

INTO all_rfm_score  

FROM #rfm_score;  

-- Phân loại khách hàng dựa trên RFM score  

CREATE VIEW rfm_statistic AS  

SELECT  

*,  

CASE  

WHEN rfm IN ('444', '443', '434', '344') THEN 'Champions'  

WHEN rfm IN ('442', '441', '432', '431', '433', '343', '342', '341') THEN  

'Loyal Customer'  

WHEN rfm IN ('424', '423', '324', '323', '413', '414', '343', '334') THEN  

'Potential Loyalist'  

WHEN rfm IN ('333', '332', '331', '313') THEN 'Promising'  

WHEN rfm IN ('422', '421', '412', '411', '311', '321', '312', '322') THEN  

>New Customer'  

WHEN rfm IN ('131', '132', '141', '142', '231', '232', '241', '242') THEN  

'Price Sensitive'  

WHEN rfm IN ('244', '234', '243', '233', '224', '214', '213', '134', '144',  

'143', '133') THEN 'Need Attention'  

WHEN rfm IN ('244', '234', '243', '233', '224', '214', '213', '134', '144',

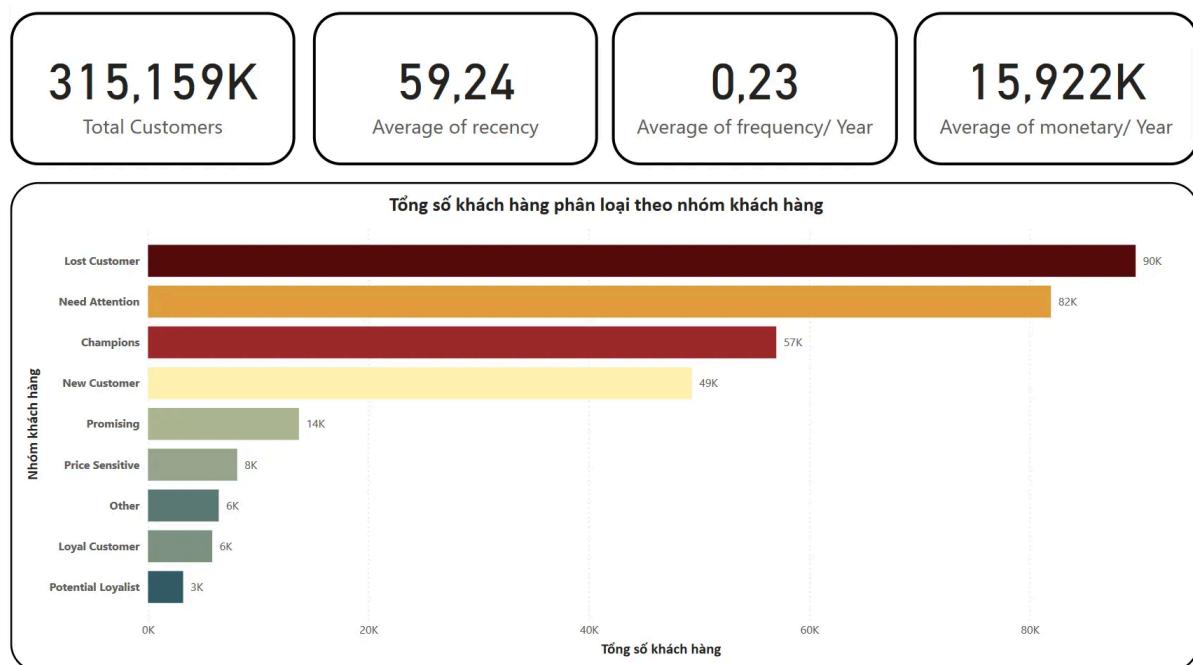
```

```
'143', '133') THEN 'At Risk Customers'  
WHEN rfm IN ('111', '112', '113', '114', '121', '122', '123', '221', '211',  
'222') THEN 'Lost Customer'  
ELSE 'Other'  
END AS customer_type  
FROM all_rfm_score;  
  
-- Gọi VIEW  
SELECT * FROM rfm_statistic;
```

4. Visualize và phân tích

4.1. Thông tin chung

Dữ liệu được sử dụng trong báo cáo bao gồm thông tin giao dịch của **315.159 khách hàng**, với thời gian gần nhất sử dụng dịch vụ trung bình là 59 ngày, tần suất sử dụng dịch vụ là **0,23 lần/năm** và số tiền trung bình khách hàng chi mỗi năm là **15.922 ngàn đồng/năm**.



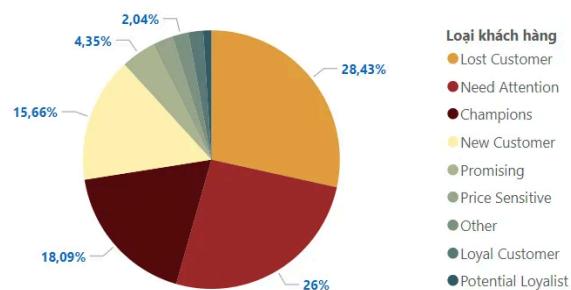
Dựa vào phân phối của tập dữ liệu này, chúng ta tính điểm RFM như sau:

Điểm	R Ngày giao dịch gần nhất	F Tần suất giao dịch/ năm	M Số tiền giao dịch/ năm
1	Từ 92 ngày trở lên	0,00 - 0,13	1 - 9.999
2	Từ 62 - 91 ngày	0,14 - 0,19	10.000 - 13.322
3	Từ 31 - 61 ngày	0,20 - 0,24	13.333 - 17.467
4	Dưới 31 ngày	0,25 - 2,00	17.500 - 832.000

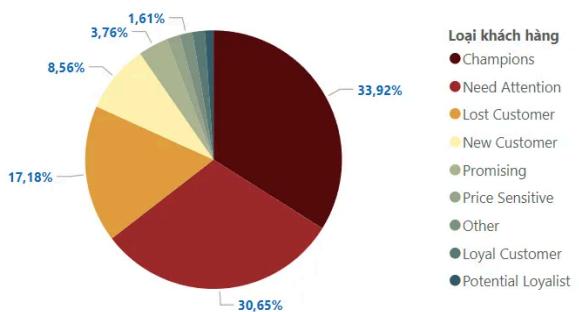
4.2. Phân tích chi tiết

Phân khúc khách hàng	Tổng số khách hàng	Tổng giá tiền đơn hàng	Tần suất mua hàng trung bình
Lost Customer	89609	861934704	0,14
Need Attention	81927	1537982367	0,27
Champions	57008	1702030049	0,44
New Customer	49348	429396687	0,13
Promising	13702	188777029	0,20
Price Sensitive	8100	83661763	0,29
Other	6429	80853761	0,13
Loyal Customer	5837	79069767	0,25
Potential Loyalist	3199	54339993	0,18
Total	315159	5018046120	0,23

Biểu đồ phân khúc khách hàng theo doanh thu



Biểu đồ phân khúc khách hàng theo doanh thu



Qua biểu đồ, chúng có thể thấy rằng **tỷ lệ khách hàng rời bỏ (Lost Customer)** đang ở **mức đáng quan tâm**, khi nhóm khách hàng này có số lượng lớn nhất với **89.609 khách hàng**, chiếm tỷ lệ **28,43% trên tổng số khách hàng** và chiếm tới **17,18% doanh thu**. Để hiểu rõ hơn về lý do mà khách hàng rời bỏ, chúng ta cần phân tích các yếu tố gây ra hiện tượng này. Điều này có thể bao gồm chất lượng sản phẩm/dịch vụ, trải nghiệm khách hàng, chiến lược giá cả, cũng như các yếu tố cạnh tranh từ các đối thủ trong ngành. Bằng cách này, chúng ta có thể phát triển các chiến lược cụ thể để giảm tỷ lệ khách hàng rời bỏ và tăng cường sự trung thành của khách hàng hiện tại, từ đó tối ưu hóa doanh thu và lợi nhuận của doanh nghiệp.

Phân khúc khách hàng thứ 2 mà doanh nghiệp cần quan tâm là nhóm **khách hàng cần chú ý (Needs Attention)**. Nhóm này chiếm tỷ lệ đông đảo **26%** với nguồn doanh thu mang lại tương đối lớn **30,65%** chỉ sau nhóm khách

hàng VIP. Họ là những người mua hàng với tần suất cao và số lượng lớn, tuy nhiên đã lâu chưa ghé thăm và thực hiện giao dịch. Việc này có thể là dấu hiệu của sự mất mát các khách hàng trung thành tiềm năng. Điều này có thể xuất phát từ nhiều nguyên nhân như sự cạnh tranh từ các đối thủ, chất lượng sản phẩm/dịch vụ không đáp ứng được mong đợi của họ, hoặc đơn giản là doanh nghiệp đã không thực hiện các chiến lược tiếp thị hoặc khuyến mãi hiệu quả đủ để duy trì sự quan tâm của họ. Để giữ chân nhóm này và tăng cường tương tác với họ, doanh nghiệp có thể áp dụng các chiến lược như chương trình khuyến mãi và ưu đãi đặc biệt dành riêng cho họ, tạo ra các chương trình thúc đẩy tái mua hàng, cũng như cải thiện trải nghiệm mua sắm và dịch vụ khách hàng. Đồng thời, việc nắm bắt được phản hồi và ý kiến phản hồi từ nhóm này cũng rất quan trọng để điều chỉnh và cải thiện chiến lược kinh doanh của doanh nghiệp.

Phân khúc **khách hàng VIP** (Champion) chiếm tỷ trọng khoảng **18,09%** và là nhóm khách hàng mang lại **doanh thu cao nhất 33,92%**. Mặc dù chỉ chiếm một phần nhỏ trong tổng số khách hàng, nhưng đây lại đóng góp một tỷ trọng lớn vào doanh thu của doanh nghiệp. Điều này xuất phát từ việc họ thường xuyên mua hàng với số lượng lớn, thực hiện các giao dịch có giá trị cao và đặc biệt là sự trung thành với thương hiệu. Doanh nghiệp cần đặc biệt chú ý và duy trì mối quan hệ tốt đẹp với nhóm khách hàng này bằng cách cung cấp các ưu đãi đặc biệt, các dịch vụ cá nhân hóa và tiếp tục nâng cao trải nghiệm mua sắm của họ.

Trong tệp khách hàng này, **nhóm khách hàng mới** (New Customer) có **số lượng chỉ 15,66%**, nhưng nhóm này mang lại **một lượng doanh thu tương đối 8,56%**. Điều này có thể là dấu hiệu tích cực, cho thấy doanh nghiệp đang thu hút được sự quan tâm và sự chú ý của nhiều đối tượng mới. Để tận dụng tốt cơ hội này, cần có chiến lược tiếp thị và chăm sóc khách hàng đặc biệt dành cho nhóm này để chuyển đổi họ thành khách hàng trung thành và đóng góp vào sự phát triển lâu dài của doanh nghiệp.

Nhóm **khách hàng trung thành** (Loyal Customer) chỉ chiếm một tỷ lệ nhỏ khoảng **1,9%** với doanh thu mang lại khoảng **1,6%**. Mặc dù nhóm khách hàng trung thành chiếm tỷ lệ nhỏ trong tệp khách hàng và không đóng góp một phần lớn vào doanh thu, nhưng họ lại có giá trị lớn đối với doanh nghiệp. Khách hàng trung thành thường là những người ủng hộ và quảng bá thương hiệu một cách tích cực. Chú trọng vào việc duy trì mối quan hệ với nhóm này, cung cấp dịch vụ chất lượng và ưu đãi đặc biệt có thể giúp gia tăng độ trung thành và tạo ra hiệu ứng lan tỏa tích cực trong cộng đồng khách hàng.

5. Kết luận

Như vậy, qua báo cáo, chúng ta nhận thấy rằng mỗi nhóm khách hàng đều có vai trò quan trọng và ảnh hưởng đáng kể đến doanh thu và lợi nhuận của doanh nghiệp. Các nhóm khách hàng VIP và khách hàng cần chú ý thường đem lại doanh thu lớn nhất, nhưng đòi hỏi chiến lược chăm sóc khách hàng đặc biệt để duy trì sự trung thành và tăng cường sự tương tác. Nhóm khách hàng mới cung cấp tiềm năng phát triển, trong khi nhóm khách hàng trung thành đóng vai trò quan trọng trong việc quảng bá thương hiệu và duy trì ổn định cho doanh nghiệp. Từ những nhận định này, doanh nghiệp có thể phát triển các chiến lược tiếp thị và chăm sóc khách hàng phù hợp để tối ưu hóa lợi ích từ mỗi nhóm khách hàng, từ đó thúc đẩy sự phát triển và bền vững của doanh nghiệp trong thị trường cạnh tranh ngày nay.