

CUSTOMER360

PHÂN LOẠI KHÁCH HÀNG DỰA TRÊN RFM

Báo cáo bởi: **Lê Thanh Tài**

29/03/2024

Mục lục

I. MỞ ĐẦU	3
II. CUSTOMER360	4
III. MÔ HÌNH RFM.....	6
PHÂN LOẠI KHÁCH HÀNG DỰA TRÊN RFM	6
IV. PHÂN TÍCH – KẾT QUẢ	8
A. DỮ LIỆU ĐẦU VÀO	8
B. BẢNG KẾT QUẢ ĐẦU RA.....	8
C. PHÂN LOẠI CỤ THỂ KHÁCH HÀNG THEO NHÓM RFM.....	9
PHỤ LỤC 1: CODE SQL	11
PHỤ LỤC 2: TỨ PHÂN VỊ - QUARTILES	15

I. Mở đầu

A. Giới thiệu

Hiểu rõ khách hàng là yếu tố then chốt cho sự thành công của bất kỳ doanh nghiệp nào. Phân loại khách hàng là một phương pháp hiệu quả để doanh nghiệp có thể hiểu rõ hơn về hành vi và sở thích của khách hàng, từ đó đưa ra những chiến lược marketing và bán hàng phù hợp.

B. Mục tiêu

Trong nền tảng Customer360, báo cáo này nhằm mục tiêu phân loại khách hàng dựa trên mô hình RFM (Recency, Frequency, Monetary), sử dụng ma trận BCG và tứ phương vị (IQR - Interquartile Range).

C. Phạm vi

Báo cáo này sẽ tập trung vào:

- Giới thiệu Customer360, mô hình RFM
- Phân loại khách hàng dựa trên mô hình RFM
- Phân tích đặc điểm của từng nhóm khách hàng

D. Đối tượng

Báo cáo này dành cho:

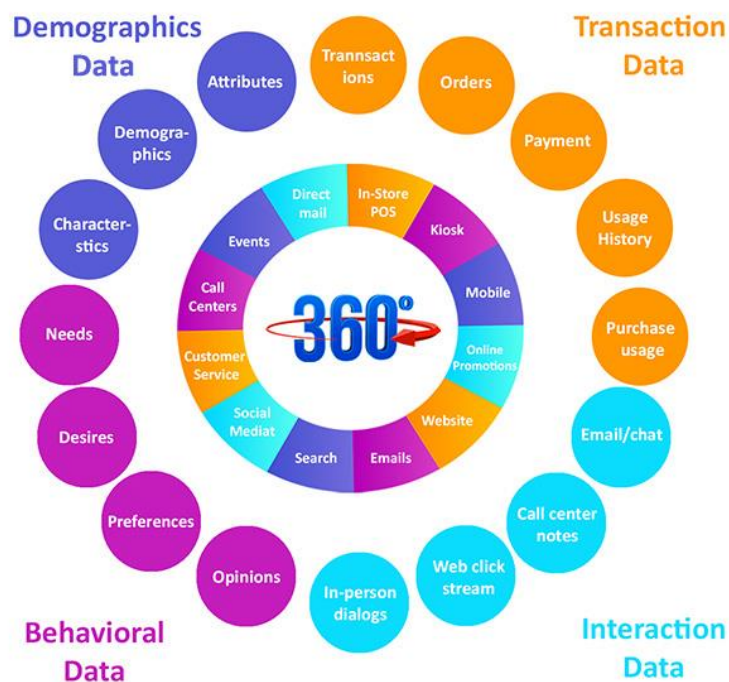
- Ban lãnh đạo doanh nghiệp
- Bộ phận marketing
- Bộ phận bán hàng

II. Customer360

Customer360 là một nền tảng tổng hợp dữ liệu khách hàng giúp doanh nghiệp có cái nhìn toàn diện về hành vi và sở thích của khách hàng. Nền tảng này thu thập dữ liệu từ nhiều nguồn khác nhau như trang web, ứng dụng di động, hệ thống CRM, v.v. và sau đó thống nhất dữ liệu để tạo ra một hồ sơ khách hàng duy nhất.

Với Customer360, doanh nghiệp có thể:

- **Hiểu rõ hơn về khách hàng:** Doanh nghiệp có thể hiểu rõ hơn về hành vi và sở thích của khách hàng, từ đó đưa ra những chiến lược marketing và bán hàng phù hợp.
- **Tăng hiệu quả marketing và bán hàng:** Doanh nghiệp có thể nhắm mục tiêu marketing và bán hàng hiệu quả hơn, dẫn đến tỷ lệ chuyển đổi cao hơn.
- **Cải thiện dịch vụ khách hàng:** Doanh nghiệp có thể cung cấp dịch vụ khách hàng tốt hơn, giải quyết các vấn đề của khách hàng nhanh chóng và hiệu quả hơn.
- **Nâng cao hiệu quả hoạt động:** Doanh nghiệp có thể hiểu rõ hơn về hiệu



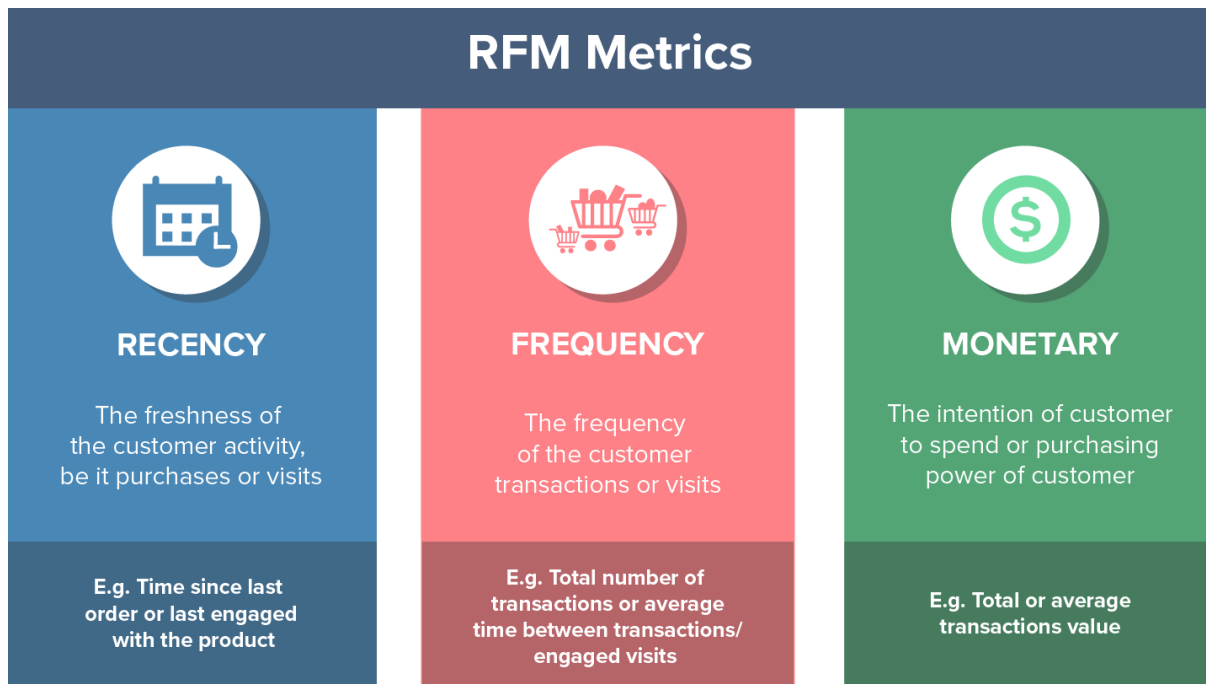
Hình 1: Các loại dữ liệu trong Customer360

Customer360 bao gồm các dữ liệu:

- **Dữ liệu nhân khẩu học (Demographics Data):** Dữ liệu về thông tin cá nhân của khách hàng, bao gồm tên, tuổi, giới tính, địa chỉ, v.v.

- **Dữ liệu giao dịch** (Transaction Data): Dữ liệu về các giao dịch của khách hàng với doanh nghiệp, bao gồm lịch sử mua hàng, thông tin thanh toán, v.v.
- **Dữ liệu tương tác** (Interaction Data): Dữ liệu về cách khách hàng tương tác với doanh nghiệp, bao gồm lượt truy cập trang web, nhấp chuột vào quảng cáo, v.v.
- **Dữ liệu hành vi** (Behavioral Data): Dữ liệu về sở thích và hành vi của khách hàng, bao gồm sở thích sản phẩm, thương hiệu yêu thích, v.v.

III. Mô hình RFM



Hình 2: Ba thành phần của mô hình RFM

Mô hình RFM (Recency, Frequency, Monetary) là một phương pháp phân loại khách hàng dựa trên ba yếu tố:

- **Recency (Thời gian mua hàng gần nhất):** Khoảng thời gian kể từ lần giao dịch gần nhất của khách hàng.
- **Frequency (Tần suất mua hàng):** Số lần giao dịch khách hàng đã thực hiện trong khoảng thời gian nhất định.
- **Monetary (Giá trị giao dịch):** Tổng giá trị các giao dịch của khách hàng trong khoảng thời gian nhất định.

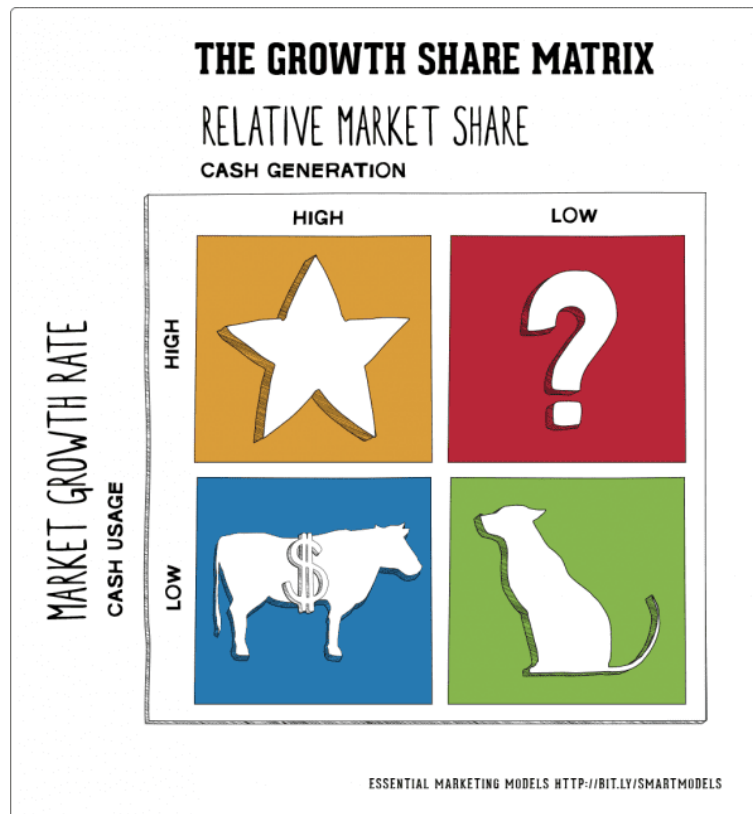
Mô hình RFM được sử dụng để đánh giá mức độ quan trọng của khách hàng và phân chia họ thành các nhóm khác nhau dựa trên hành vi mua hàng.

Phân loại khách hàng dựa trên RFM

Mỗi khách hàng sẽ được tính điểm về 3 tiêu chí R, F, và M, bằng **tứ phân vị quartiles**, trên thang điểm từ 1 đến 4, với 1 là điểm thấp nhất, 4 là điểm cao nhất khi so sánh trong toàn bộ tập dữ liệu khách hàng:

- Recency: 1 điểm – thời điểm giao dịch gần nhất xa; 4 điểm – thời điểm giao dịch gần nhất gần
- Frequency: 1 điểm – tần suất giao dịch thấp; 4 điểm – tần suất giao dịch cao
- Monetary: 1 điểm – tổng giá trị giao dịch thấp; 4 điểm – tổng giá trị giao dịch cao

Ba điểm số này sẽ được gộp thành một bộ chỉ số RFM cho từng khách và được dùng để phân loại khách hàng vào từng nhóm theo **ma trận BCG**.



Hình 3: Ma trận BCG với 4 nhóm khách hàng

Ma trận BCG phân chia khách hàng thành 4 nhóm:

- Star: Nhóm khách hàng có tốc độ tăng trưởng chi tiêu cao và thị phần chi tiêu cao. Đây là nhóm khách hàng VIP và mang lại lợi nhuận lớn cho doanh nghiệp.
- Cash Cow: Nhóm khách hàng có tốc độ tăng trưởng chi tiêu thấp và thị phần chi tiêu cao. Đây là nhóm khách hàng trung thành và mang lại lợi nhuận ổn định cho doanh nghiệp.
- Question Mark: Nhóm khách hàng có tốc độ tăng trưởng chi tiêu cao và thị phần chi tiêu thấp. Đây là nhóm khách hàng tiềm năng nhưng cần được đầu tư để tăng thị phần.
- Dog: Nhóm khách hàng có tốc độ tăng trưởng chi tiêu thấp và thị phần chi tiêu thấp. Đây là nhóm khách hàng ít tiềm năng và mang lại lợi nhuận thấp cho doanh nghiệp.

IV. Phân tích – Kết quả

A. Dữ liệu đầu vào

Dữ liệu đầu vào bao gồm 2 bảng: *Customer_Register* (bảng chứa thông tin đăng ký của khách hàng) và *Customer_Transaction* (bảng chứa thông tin giao dịch của khách hàng).

Tên cột	Kiểu dữ liệu	Ý nghĩa
ID	Bigint	ID của khách hàng
Contract	Varchar(100)	Mã hợp đồng
LocationID	Int	Mã vị trí
BranchCode	Tinyint	Mã chi nhánh
Status	Tinyint	Trạng thái
Created_Date	Datetime	Ngày đăng ký
Stopdate	datetime	Ngày hủy

Bảng 1: Bảng dữ liệu *Customer_Register*

Tên cột	Kiểu dữ liệu	Ý nghĩa
ID	Bigint	ID giao dịch
CustomerID	Varchar(200)	ID của khách hàng
Purchase_Date	Datetime	Ngày giao dịch
GMV	Bigint	Gross Monetary Value – giá trị giao dịch

Bảng 2: Bảng dữ liệu *Customer_Transaction*

B. Bảng kết quả đầu ra

Bảng kết quả đầu ra bao gồm các bộ RFM, số khách hàng có RFM đó và phân loại các bộ RFM này thành các nhóm khách hàng theo ma trận BCG.

Tên cột	Kiểu dữ liệu	Ý nghĩa
Customer_type	Varchar	Phân nhóm khách hàng
Type_count	Bigint	Số khách hàng trong nhóm
Type_percent	Float	Tỉ lệ phần trăm của nhóm khách hàng trên tổng số
Total_GMV_type	Bigint	Tổng giá trị giao dịch của nhóm khách hàng

Percent_GMV_type	Float	Tỉ lệ phần trăm của giá trị giao dịch của nhóm khách hàng trên tổng giá trị giao dịch
-------------------------	-------	---

Bảng 3: Bảng kết quả result

C. Phân loại cụ thể khách hàng theo nhóm RFM

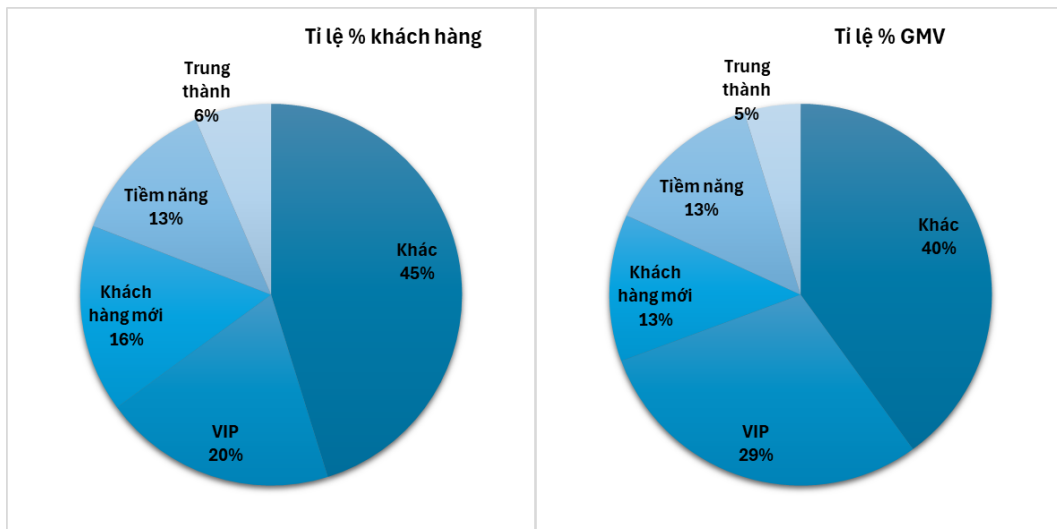
Nhóm khách hàng	Mô tả	Các nhóm điểm RFM
Nhóm khách hàng VIP	Nhóm khách hàng chi tiêu nhiều nhất, sử dụng dịch vụ thường xuyên và có giao dịch gần đây	444, 443, 434, 344, 433, 343, 334, 244, 243
Nhóm khách hàng trung thành	Nhóm khách hàng sử dụng dịch vụ thường xuyên và có giao dịch gần đây, tuy không chi tiêu nhiều như nhóm VIP	442, 441, 432, 431, 342, 341, 332, 331
Nhóm khách hàng tiềm năng	Nhóm khách hàng có mức chi tiêu cao, tuy giao dịch ít thường xuyên	424, 423, 324, 323, 234, 233, 413, 414, 313, 314
Nhóm khách hàng mới	Nhóm khách hàng mới sử dụng dịch vụ gần đây, chưa chi tiêu nhiều và giao dịch không thường xuyên	422, 421, 412, 411, 311, 321, 312, 322
Nhóm khách hàng khác	Nhóm khách hàng vắng lai hoặc không còn sử dụng dịch vụ	Các nhóm điểm RFM còn lại

Bảng 4: Bảng phân loại chi tiết khách hàng theo RFM

Với cách phân loại nhóm khách hàng như trên, ta có tỉ lệ các nhóm khách hàng như sau:

Nhóm khách hàng	Số khách hàng trong nhóm (người)	Tỉ lệ % trong tổng số khách hàng (%)	Tổng chi tiêu của nhóm (VND)	Tỉ lệ % trong tổng chi tiêu (%)
Khác	51,563	45.20	4,450,400,083	40.01
VIP	22,525	19.74	3,264,171,057	29.34
Khách hàng mới	18,164	15.92	1,386,260,678	12.46
Tiềm năng	14,439	12.66	1,494,351,067	13.43
Trung thành	7,390	6.48	529,078,313	4.76

Bảng 5: Số liệu của các nhóm khách hàng



Hình 4: Tỉ lệ % các nhóm khách hàng

Ngoài nhóm *Khách hàng khác* (các khách hàng vắng lai hoặc đã không còn sử dụng dịch vụ), nhóm *Khách hàng VIP* chiếm tỉ lệ cao nhất với **19.74%**, với 18,164 người trên tổng số 114,081 tổng số khách hàng trong tập dữ liệu. Đồng thời nhóm *Khách hàng VIP* chiếm tỉ lệ **29.34%** doanh thu so với tổng doanh thu từ toàn bộ khách hàng. Đây là tệp khách hàng quan trọng nhất và cần được quan tâm nhất trong các chiến dịch marketing của công ty.

Nhóm *Khách hàng trung thành* chiếm tỉ lệ % khách hàng thấp (**6.48%**) và tỉ lệ doanh thu thấp (**4.76%**). Tuy đây là nhóm khách hàng có chi tiêu không cao, nhưng là nhóm khách hàng sử dụng dịch vụ thường xuyên, đem lại nguồn thu ổn định. Cần có chiến lược để kích thích chi tiêu cho tệp khách hàng này.

Một nhóm khách hàng khác có mức chi tiêu khá cao là nhóm *Khách hàng tiềm năng*, chiếm **12.66%** tổng số khách hàng và chiếm **13.43%** tổng doanh thu. Nhóm này có tiềm năng để trở thành *Khách hàng VIP*, do đó cũng cần được quan tâm.

Phụ lục 1: Code SQL

```
# Bang Customer_Registered

SELECT TOP 10 *
FROM dbo.Customer_Registered cr ;


# Bang Customer_Transaction

SELECT TOP 10 *
FROM dbo.Customer_Transaction ct
ORDER BY Purchase_Date DESC;


# ROW_NUMBER cho R, F, M
WITH cal AS (
SELECT CustomerID,
        DATEDIFF(DAY, MAX(Purchase_Date), '2022-09-01') AS recency,
        ROUND(COUNT(DISTINCT CAST(Purchase_Date AS date))*1.00 /
DATEDIFF(year, CAST(created_date AS date), '2022-09-01'), 2) AS frequency,
        (sum(gmv) / DATEDIFF(year, CAST(created_date AS date), '2022-09-01')) AS monetary,
        sum(ct.GMV) AS GMV
FROM dbo.Customer_Transaction ct
JOIN dbo.Customer_Registered cr ON ct.CustomerID = cr.ID
WHERE CustomerID != 0
GROUP BY CustomerID, created_date)
SELECT *,
        ROW_NUMBER() OVER (ORDER BY recency DESC) AS rn_recency,
        ROW_NUMBER() OVER (ORDER BY frequency) AS rn_frequency,
        ROW_NUMBER() OVER (ORDER BY monetary) AS rn_monetary
INTO #calculation
FROM cal;
```

Diem R, F, M dua tren **row_number**

```
SELECT CustomerID,

      CASE

          WHEN rn_recency >= 1 AND rn_recency < (SELECT
cast(COUNT(DISTINCT(CustomerID)) * 0.25 AS int) FROM #calculation)

              THEN '1'

          WHEN rn_recency >= (SELECT
cast(COUNT(DISTINCT(CustomerID)) * 0.25 AS int) FROM #calculation) AND
rn_recency < (SELECT cast(COUNT(DISTINCT(CustomerID)) * 0.5 AS int) FROM
#calculation)

              THEN '2'

          WHEN rn_recency >= (SELECT
cast(COUNT(DISTINCT(CustomerID)) * 0.5 AS int) FROM #calculation) AND
rn_recency < (SELECT cast(COUNT(DISTINCT(CustomerID)) * 0.75 AS int) FROM
#calculation)

              THEN '3'

          ELSE '4'

      END AS R,

      CASE

          WHEN rn_frequency >= 1 AND rn_frequency < (SELECT
cast(COUNT(DISTINCT(CustomerID)) * 0.25 AS int) FROM #calculation)

              THEN '1'

          WHEN rn_frequency >= (SELECT
cast(COUNT(DISTINCT(CustomerID)) * 0.25 AS int) FROM #calculation) AND
rn_frequency < (SELECT cast(COUNT(DISTINCT(CustomerID)) * 0.5 AS int) FROM
#calculation)

              THEN '2'

          WHEN rn_frequency >= (SELECT
cast(COUNT(DISTINCT(CustomerID)) * 0.5 AS int) FROM #calculation) AND
rn_frequency < (SELECT cast(COUNT(DISTINCT(CustomerID)) * 0.75 AS int) FROM
#calculation)

              THEN '3'
```

```

ELSE '4'

END AS F,

CASE

    WHEN rn_monetary >= 1 AND rn_monetary < (SELECT
cast(COUNT(DISTINCT(CustomerID)) * 0.25 AS int) FROM #calculation)

        THEN '1'

        WHEN rn_monetary >= (SELECT
cast(COUNT(DISTINCT(CustomerID)) * 0.25 AS int) FROM #calculation) AND
rn_monetary < (SELECT cast(COUNT(DISTINCT(CustomerID)) * 0.5 AS int) FROM
#calculation)

            THEN '2'

            WHEN rn_monetary >= (SELECT
cast(COUNT(DISTINCT(CustomerID)) * 0.5 AS int) FROM #calculation) AND
rn_monetary < (SELECT cast(COUNT(DISTINCT(CustomerID)) * 0.75 AS int) FROM
#calculation)

                THEN '3'

                ELSE '4'

END AS M,

GMV

INTO #result
FROM #calculation;

# Phan loai khach hang dua tren RFM
SELECT rfm_group,
COUNT(*) AS gr_count,
COUNT(*)*100.00 / (SELECT COUNT(*)*1.00 FROM #result) AS
gr_percent,
CASE
    WHEN rfm_group IN (444, 443, 434, 344, 433, 343, 334, 244, 243)
THEN 'VIP'

    WHEN rfm_group IN (442, 441, 432, 431, 342, 341, 332, 331)
THEN 'Trung thanh'

```

```

        WHEN rfm_group IN (424, 423, 324, 323, 234, 233, 413, 414, 313,
314) THEN 'Tiem nang'

        WHEN rfm_group IN (422, 421, 412, 411, 311, 321, 312, 322)
THEN 'Khach hang moi'

        ELSE 'Khac'

    END AS customer_type,

    Sum(GMV) AS Total_GMV_group,

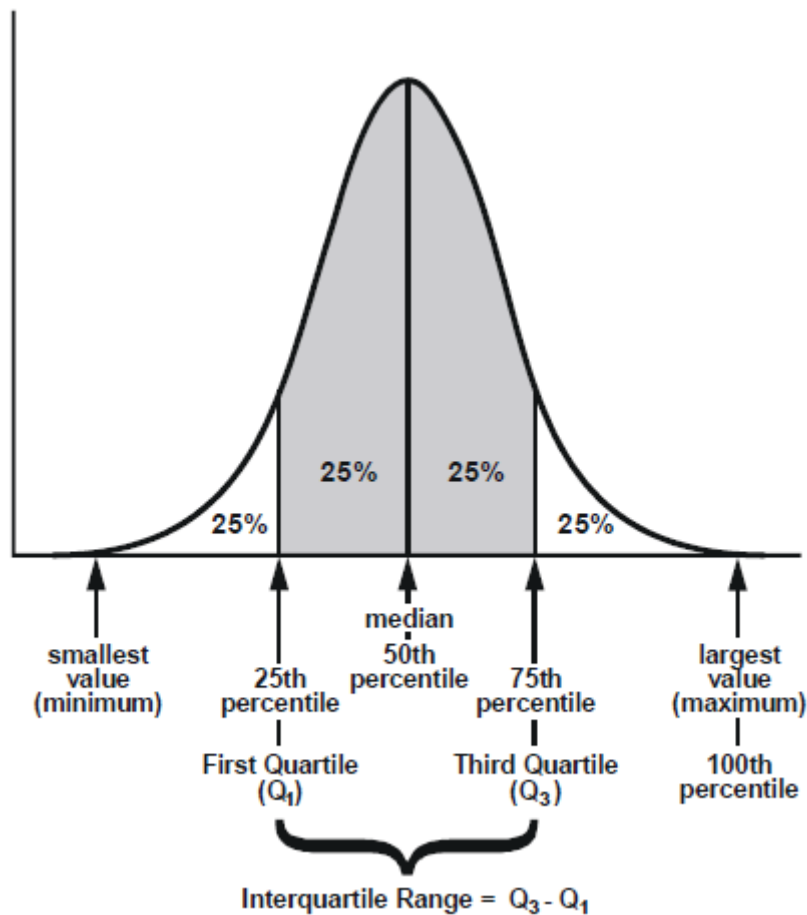
    Sum(GMV)*100.00 / (SELECT SUM(GMV) FROM #RESULT) AS
Percent_GMV_group
INTO #rfm
FROM (
    select *, concat(R,F,M) as rfm_group
    from #result
) AS sub
GROUP BY rfm_group;

# So khach hang, GMV cua moi nhom khach hang
SELECT customer_type,
    sum(gr_count) AS type_count,
    sum(gr_count)*100.00 / (SELECT SUM(gr_count)*1.00 FROM #rfm) AS
type_percent,
    SUM(Total_GMV_group) AS Total_GMV_type,
    SUM(Total_GMV_group)*100.00 / (SELECT SUM(Total_GMV_group) FROM
#rfm) AS Percent_GMV_type
FROM #rfm
GROUP BY customer_type;

# Xoa bang tam
DROP TABLE #result
DROP table #calculation
DROP TABLE #rfm;

```

Phụ lục 2: Tứ phân vị - quartiles



Tứ phân vị, hay quartiles, là một phương pháp thống kê chia tập dữ liệu thành bốn phần bằng nhau, giúp khám phá cấu trúc và phân bố của dữ liệu một cách trực quan và hiệu quả.

Cách thức hoạt động:

1. Sắp xếp dữ liệu theo thứ tự tăng dần.
2. Xác định vị trí của các tứ phân vị:
 - Q_1 : vị trí $(n + 1) / 4$, với n là số lượng dữ liệu.
 - Q_2 : vị trí $n / 2$ (trung vị).
 - Q_3 : vị trí $3(n + 1) / 4$.
3. Dựa vào vị trí tứ phân vị, chia tập dữ liệu thành 4 phần:
 - Phần 1: từ giá trị nhỏ nhất đến Q_1 .
 - Phần 2: từ Q_1 đến Q_2 .
 - Phần 3: từ Q_2 đến Q_3 .

- Phần 4: từ Q3 đến giá trị lớn nhất.

Lợi ích của việc sử dụng tứ phương vị:

- **Phân chia dữ liệu khoa học:** Tứ phân vị giúp chia tập dữ liệu thành các phần có kích thước bằng nhau, tạo nền tảng cho phân tích và so sánh hiệu quả.
- **Khám phá giá trị ngoại lệ:** Các giá trị nằm ngoài phạm vi $Q1 - 1.5 * IQR$ và $Q3 + 1.5 * IQR$ có thể được xem là giá trị ngoại lệ, cần được chú ý và kiểm tra kỹ lưỡng.
- **Đánh giá độ đồng đều:** So sánh độ dài của các phần dữ liệu giúp ta đánh giá mức độ đồng đều của dữ liệu.
- **Hiểu rõ cấu trúc dữ liệu:** Tứ phân vị cung cấp thông tin về sự phân bố của dữ liệu, giúp ta hiểu rõ hơn về cấu trúc và đặc điểm của tập dữ liệu.