

BÁO CÁO PHÂN TÍCH KHÁCH HÀNG

DỰA TRÊN MÔ HÌNH

RFM

Nguyễn Hoàng Phương Linh

MỤC LỤC

TỔNG QUAN

MÔ HÌNH RFM 1

ĐỘ TRẢI GIỮA 3
(IQR - INTERQUARTILE RANGE)

PHÂN NHÓM KHÁCH HÀNG

DỮ LIỆU 4

NHÓM ĐIỂM R-F-M 5

PHÂN NHÓM KHÁCH HÀNG 7
THEO ĐIỂM RFM

PHÂN TÍCH NHÓM KHÁCH HÀNG 8

TỔNG KẾT 12

MYSQL QUERY 13

TỔNG QUAN

MÔ HÌNH RFM

Mô hình RFM là gì?

Mô hình RFM là một kỹ thuật phân khúc khách hàng được sử dụng để xác định và thấu hiểu các nhóm khách hàng chính dựa trên hành vi mua hàng của họ. RFM là viết tắt của ba yếu tố:

- **Recency** (Tính Mới Nhất): Đo lường khoảng thời gian kể từ lần mua hàng cuối cùng của khách hàng.
- **Frequency** (Tần Suất): Đo lường số lần khách hàng đã mua hàng trong một khoảng thời gian nhất định.
- **Monetary** (Giá Trị): Đo lường tổng số tiền mà khách hàng đã chi tiêu.

Lợi ích của mô hình RFM

Mô hình RFM giúp các doanh nghiệp hiểu rõ hơn về hành vi mua sắm của khách hàng và phân loại khách hàng thành các nhóm khác nhau. Việc này giúp tối ưu hóa các chiến dịch marketing và tăng cường hiệu quả kinh doanh thông qua:

- **Phân Khúc Khách Hàng:** Phân loại khách hàng thành các nhóm khác nhau dựa trên hành vi mua sắm của họ.
- **Xác Định Khách Hàng Quan Trọng:** Xác định những khách hàng có giá trị cao nhất để có thể tập trung vào việc giữ chân họ
- **Chiến Dịch Marketing Cá Nhân Hóa:** Tạo ra các chiến dịch marketing phù hợp với từng nhóm khách hàng cụ thể.

TỔNG QUAN

MÔ HÌNH RFM

Các Bước Thực Hiện Mô Hình RFM

1. Thu Thập Dữ Liệu

- Dữ liệu giao dịch: Bao gồm các thông tin về khách hàng như ID khách hàng, ngày giao dịch, số lượng giao dịch, giá trị giao dịch, v.v.
- Dữ liệu khách hàng: Bao gồm các thông tin cá nhân của khách hàng như tên, địa chỉ, ngày đăng ký, v.v.

2. Chuẩn Bị Dữ Liệu

- Xử lý dữ liệu: Loại bỏ các bản ghi trùng lặp, xử lý các giá trị bị thiếu, và định dạng lại dữ liệu khi cần thiết.
- Chọn khung thời gian: Xác định khoảng thời gian dùng để phân tích.

3. Tính Toán Điểm RFM

Mỗi khách hàng sẽ được gán một điểm số cho từng yếu tố Recency, Frequency, Monetary.

Điểm số này thường được xác định bằng cách xếp hạng các khách hàng khi so sánh với nhau và chia thành các nhóm theo từng yếu tố trên.
(ví dụ: nhóm 1-4, với nhóm 4 là cao nhất tương đương 4 điểm).

Cuối cùng, các thứ hạng R, F và M riêng lẻ của khách hàng được kết hợp để tạo thành điểm RFM tổng hợp. Dựa trên điểm RFM, khách hàng sẽ được phân loại thành các phân khúc hoặc nhóm khác nhau.

TỔNG QUAN

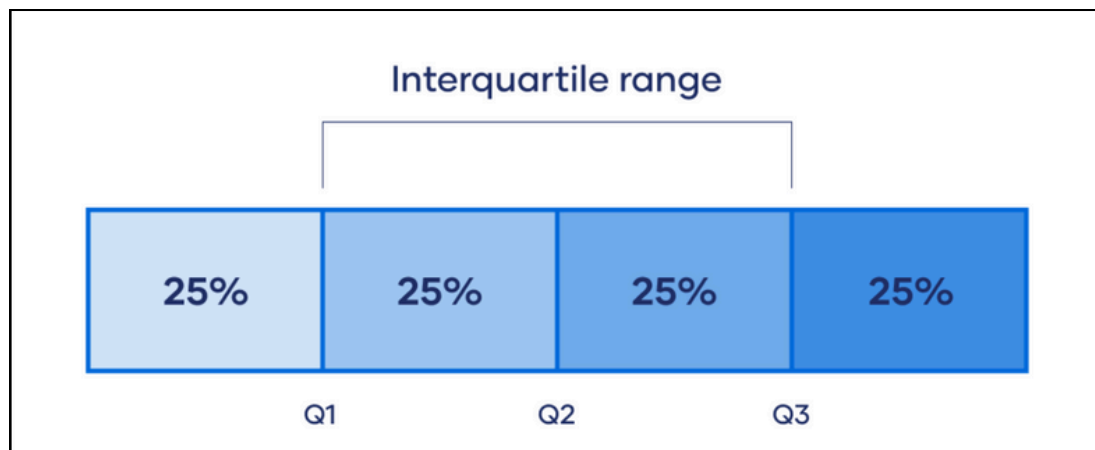
ĐỘ TRẢI GIỮA (IQR - INTERQUARTILE RANGE)

IQR là gì?

IQR là một thước đo thống kê dùng để đo lường sự trải rộng của dữ liệu. IQR giúp ta hiểu được mức độ phân tán của 50% dữ liệu trung tâm trong một tập dữ liệu và loại bỏ ảnh hưởng của những giá trị quá cao hoặc quá thấp (giá trị ngoại lai).

Phương pháp IQR chia dữ liệu thành 4 phần bằng nhau với 3 điểm:

- **Q1** (Phần tư thứ nhất): Phân vị thứ 25, 25% dữ liệu nằm dưới giá trị này.
- **Q2** (Phần tư thứ hai hoặc Median): Phân vị thứ 50, 50% dữ liệu nằm dưới giá trị này. Đây cũng chính là trung vị của dữ liệu.
- **Q3** (Phần tư thứ ba): Phân vị thứ 75, 75% dữ liệu nằm dưới giá trị này.



PHÂN NHÓM KHÁCH HÀNG

DỮ LIỆU

Báo cáo này phân tích tập dữ liệu bao gồm hơn **113,800 khách hàng** có ký hợp đồng và thực hiện giao dịch tính đến ngày **01-09-2022** với tổng doanh thu là **2,39 tỷ đô la**.

Dữ liệu giao dịch được thu thập từ **bảng Customer_transaction**

Tên cột	loại dữ liệu	Giải nghĩa
Transaction_ID	int	Mã giao dịch
CustomerID	int	Mã khách hàng
GMV	int	Tổng doanh thu trên một giao dịch
New_Purchase_Date	date	Ngày giao dịch

123 Transaction_ID	123 CustomerID	123 GMV	🕒 new_Purchase_Date
0	1,327,813	95,000	2022-06-01
1	1,157,830	75,000	2022-06-01
2	873,915	95,000	2022-07-01

Dữ liệu khách hàng được thu thập từ **bảng Customer_registered**

Tên cột	loại dữ liệu	Giải nghĩa
ID	int	Mã khách hàng
Contract	varchar(50)	Mã hợp đồng
LocationID	int	Mã vị trí
BranchCode	int	Mã chi nhánh
Status	int	Trạng thái
Stopdate	date	Ngày chấm dứt hợp đồng
New_Created_date	date	Ngày đăng kí

123 ID	ABC Contract	123 LocationID	123 BranchCode	123 Status	ABC stopdate	🕒 new_Created_date
58	SGD186201	8	5	0		2009-03-01
59	HPD005291	31	0	0		2009-03-01
60	SGD238513	8	3	0		2009-03-01

PHÂN NHÓM KHÁCH HÀNG

NHÓM ĐIỂM R-F-M

Theo mỗi yếu tố **Recency (R)**, **Frequency (F)**, **Monetary (M)** khách hàng được chia thành 4 nhóm từ 1 đến 4 tương ứng số điểm từ 1 đến 4 với 4 điểm là cao nhất.

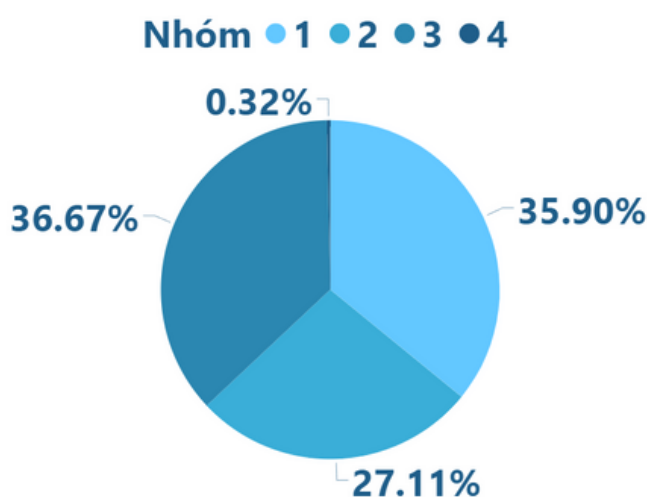
Giới hạn của từng nhóm được xác định bằng thước đo thống kê **IQR** :

GIÁ TRỊ Q1-Q2-Q3 THEO R-F-M

Type	Q1	Q2	Q3
R	31	62	92
F	0.1901	0.2280	0.2572
M	15,646.0502	19,194.3626	23,848.0055

NHÓM ĐIỂM CHI TIẾT

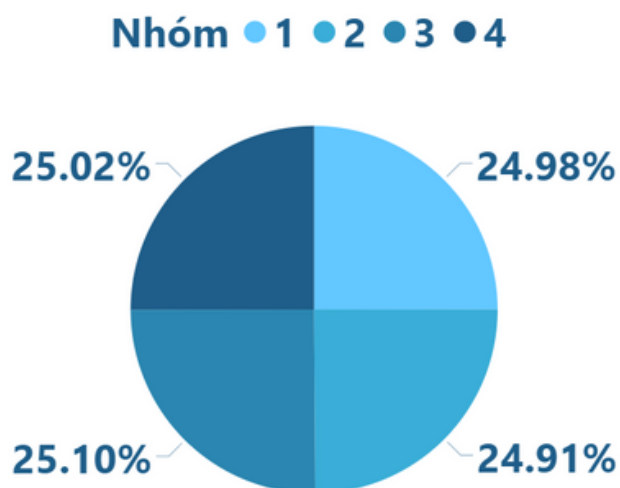
Recency (R) Khoảng cách từ ngày tiêu dùng cuối cùng đến 01-09-2022	
Nhóm 1	$R \geq 92$ ngày
Nhóm 2	$62 \text{ ngày} \leq R < 92 \text{ ngày}$
Nhóm 3	$31 \text{ ngày} \leq R < 62 \text{ ngày}$
Nhóm 4	$R < 31$ ngày



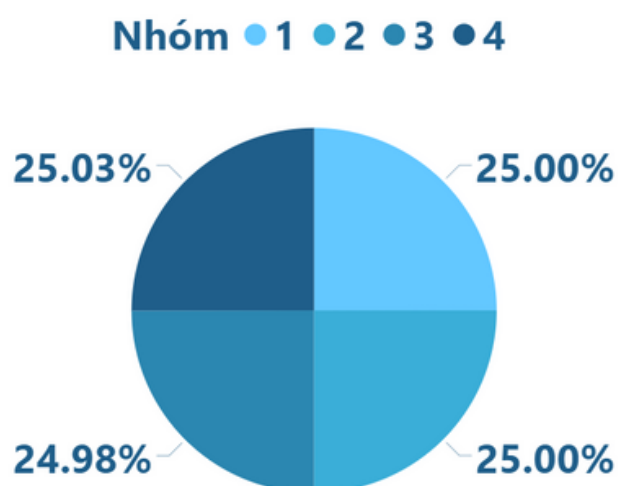
PHÂN NHÓM KHÁCH HÀNG

NHÓM ĐIỂM R-F-M

Frequency (F) (Tổng số ngày mua hàng) / (Khoảng cách từ ngày ký hợp đồng đến ngày 01-09-2022)	
Nhóm 1	$F < 0.19$
Nhóm 2	$0.19 \leq F < 0.23$
Nhóm 3	$0.23 \leq F < 0.26$
Nhóm 4	$F \geq 0.26$



Monetary (M) (Tổng số tiền khách hàng đã chi tiêu) / (Khoảng cách từ ngày ký hợp đồng đến ngày 01-09-2022)	
Nhóm 1	$M < 15,646$
Nhóm 2	$15,646 \leq M < 19,194$
Nhóm 3	$19,194 \leq M < 23,848$
Nhóm 4	$M \geq 23,848$



PHÂN NHÓM KHÁCH HÀNG

PHÂN NHÓM KHÁCH HÀNG THEO TỔ HỢP ĐIỂM RFM

Mỗi khách hàng được gán một tổ hợp điểm RFM, bao gồm điểm của ba yếu tố: Recency (R), Frequency (F) và Monetary (M). Dựa trên tổ hợp điểm này, khách hàng được phân loại vào các nhóm theo bảng mô tả dưới đây:

Nhóm khách hàng	Mô tả	Các tổ hợp điểm RFM
Khách hàng VIP	Khách hàng chi tiêu nhiều, sử dụng dịch vụ thường xuyên và có sử dụng dịch vụ thời gian gần	334, 343, 344, 434, 443, 444
Khách hàng trung thành	Khách hàng chi tiêu ít, sử dụng dịch vụ thường xuyên và có sử dụng dịch vụ thời gian gần	331, 332, 333, 341, 342, 431, 432, 433, 441, 442
Khách hàng tiềm năng	Khách hàng chi tiêu nhiều, số lần sử dụng dịch vụ chưa cao và có sử dụng dịch vụ thời gian gần	313, 314, 323, 324, 413, 414, 423, 424
Khách hàng mới	Khách hàng chi tiêu ít, số lần sử dụng dịch vụ chưa cao và có sử dụng dịch vụ thời gian gần	311, 312, 321, 322, 411, 412, 421, 422
Khách hàng sắp rời bỏ	Khách hàng sử dụng dịch vụ thường xuyên nhưng không sử dụng dịch vụ trong khoảng thời gian tương đối lâu	131, 132, 133, 134, 141, 142, 143, 144, 233, 234, 243, 244, 231, 232, 241, 242
Khách hàng vắng lai	Khách hàng sử dụng dịch vụ tương đối ít và không sử dụng dịch vụ trong thời gian gần	121, 122, 123, 124, 221, 222, 223, 224
Khách hàng đã rời bỏ	Khách hàng có sử dụng dịch vụ nhưng không quay lại	111, 112, 113, 114, 211, 212, 213, 214

PHÂN TÍCH NHÓM KHÁCH HÀNG

Ma Trận Giá Trị Chỉ Tiêu Theo Điểm Recency và Frequency (R-F)

R_score F_score	1	2	3	4
1	\$170,331,689	\$208,333,316	\$245,063,218	\$150,649,107
2	\$126,049,807	\$151,839,292	\$187,899,575	\$130,782,697
3	\$125,910,984	\$151,553,733	\$185,472,164	\$551,242,744
4	\$901,750	\$980,072	\$1,247,036	\$5,306,094

Ma Trận Tổng Số Khách Hàng Theo Điểm Recency và Frequency (R-F)

R_score F_score	1	2	3	4
1	11453	11500	11296	6592
2	8415	8411	8630	5391
3	8479	8372	8562	16307
4	74	57	64	171

Ma Trận ARPU Theo Điểm Recency và Frequency (R-F)

R_score F_score	1	2	3	4
1	14,872.23	18,115.94	21,694.69	22,853.32
2	14,979.18	18,052.47	21,772.84	24,259.45
3	14,849.74	18,102.45	21,662.25	33,804.06
4	12,185.81	17,194.25	19,484.94	31,029.79

Các ma trận làm nổi bật các phân khúc giá trị và xu hướng hành vi quan trọng bằng cách minh họa mối quan hệ giữa điểm Recency, điểm Frequency và các yếu tố:

1. Tổng doanh thu (tổng Monetary).
2. Lượng khách hàng.
3. Doanh thu trung bình trên mỗi khách hàng (ARPU).

(Màu đậm - nhóm khách hàng giá trị cao. Màu nhạt - nhóm khách hàng giá trị thấp)

Các khu vực có **màu đậm đang ở vị trí R=1,2,3 và F=2,3** (khách hàng gần đây không giao dịch và có tần xuất trung bình) cho thấy phần lớn **khách hàng có xu hướng rời bỏ**. Khu có **màu nhạt nhất ở R=4** (khách có giao dịch gần nhất) cho thấy số khách hàng mới không nhiều và doanh thu mang lại cũng không cao, **sản phẩm/dịch vụ hiện tại không thu hút được khách**.

Nhóm R=3, F=4 (khách hàng giao dịch gần đây và thường xuyên) có **giá trị cao nhất** về cả số lượng và doanh thu, **đạt hơn 551 triệu USD**. Đây là nhóm chính mang lại doanh thu ổn định cho doanh nghiệp. Nhóm **R=4, F=4** có **tổng giá trị và lượng khách hàng rất thấp nhưng có ARPU rất cao**, đây là nhóm khách tiềm năng để doanh nghiệp phát triển doanh thu.

Các nhóm khách hàng có **cả 3 yếu tố đều cao nằm ở khu vực R=1, F=2 và R=1,2,3 và F=3**. Đây là khu vực trọng điểm cần được doanh nghiệp theo dõi chặt chẽ vì tuy họ hiện vẫn mang lại giá trị đáng kể, nhưng dấu hiệu về **Recency thấp**, cùng với Nhóm **R=1,F=1** có **lượng khách hàng cao thứ 2** cho thấy **nguy cơ rời bỏ rất cao**.

PHÂN TÍCH NHÓM KHÁCH HÀNG

TỔNG LƯỢNG KHÁCH HÀNG VÀ DOANH THU THEO NHÓM KHÁCH HÀNG



Biểu đồ cột minh họa rõ ràng hơn đặc điểm hành vi của khách hàng sau khi được phân loại vào các nhóm RFM:

- **Nhóm KH SẮP RỜI BỎ** chiếm gần 30% tổng chi tiêu ($\approx 714\text{M USD}$) và 28% số lượng khách hàng, là **nhóm lớn nhất cả về quy mô lẫn giá trị**. Điều này cho thấy công ty đang đối diện nguy cơ sụt giảm doanh thu đáng kể nếu không triển khai biện pháp giữ chân.
- **Nhóm KH VẪNG LẠI** và **nhóm KH ĐÃ RỜI BỎ** có sự tương đồng và cùng đóng góp giá trị đáng kể (≈ 360 triệu USD với khoảng 20.000 khách và ≈ 300 triệu với 20.000 khách). Đây là tín hiệu rõ ràng cho thấy một **tỷ lệ lớn khách hàng chỉ giao dịch ngắn hạn hoặc một lần duy nhất**.
- **Nhóm KH VIP** đứng thứ hai về chi tiêu (≈ 610 triệu USD, 25%) nhưng chỉ chiếm 16% số khách, cho thấy **giá trị trung bình trên mỗi khách rất cao**. Đây là nhóm cần ưu tiên chăm sóc, nuôi dưỡng để duy trì và phát triển doanh thu ổn định.
- Các nhóm còn lại (MỚI, TRUNG THÀNH, TIỀM NĂNG) đóng góp nhỏ hơn nhiều.

PHÂN TÍCH NHÓM KHÁCH HÀNG

BANNER	TỔNG GIÁ TRỊ CHI TIÊU	% GIÁ TRỊ CHI TIÊU	TỔNG SỐ KHÁCH HÀNG	% TỔNG SỐ KHÁCH HÀNG	GIÁ TRỊ / KHÁCH HÀNG
KH SẮP RỜI BỎ	\$714,394,597.72	29.85%	31,909	28.05%	22,388.50
133	\$88,538,687.42	3.70%	4,140	3.64%	21,386.16
134	\$84,999,951.03	3.55%	3,094	2.72%	27,472.51
143	\$70,266,115.12	2.94%	3,468	3.05%	20,261.28
144	\$70,206,982.29	2.93%	2,525	2.22%	27,804.75
244	\$68,698,710.53	2.87%	2,235	1.96%	30,737.68
132	\$68,697,755.57	2.87%	3,795	3.34%	18,102.18
233	\$66,393,692.99	2.77%	3,093	2.72%	21,465.79
234	\$65,739,127.98	2.75%	2,350	2.07%	27,974.10
243	\$53,660,644.58	2.24%	2,646	2.33%	20,279.91
232	\$53,362,360.08	2.23%	2,947	2.59%	18,107.35

NHÓM KH SẮP RỜI BỎ:

Hai nhóm RFM có **doanh thu cao nhất là 133 và 134**. sau 2 nhóm cao nhất, các tổ hợp khác có doanh thu cao nhưng ở mức tương đồng, doanh nghiệp có thể lựa chọn các nhóm có **giá trị ARPU cao** đặt làm trọng tâm như:

144 (27 nghìn), 233 (21 nghìn), 234 (28 nghìn), 244 (31 nghìn).

Đây đều là khách hàng từng thuộc nhóm VIP và TRUNG THÀNH trước đây (tần suất giao dịch cao, giá trị lớn) nhưng đã ngừng mua hàng trong thời gian dài. Vì vậy cần xác định nhóm KH SẮP RỜI BỎ là nhóm trọng điểm để triển khai các chiến dịch giữ chân khẩn cấp, trước khi họ chuyển hẳn sang trạng thái rời bỏ. Các tổ hợp RFM khác cũng cần được chăm sóc nhưng 6 nhóm trên nên được ưu tiên.

BANNER	TỔNG GIÁ TRỊ CHI TIÊU	% GIÁ TRỊ CHI TIÊU	TỔNG SỐ KHÁCH HÀNG	% TỔNG SỐ KHÁCH HÀNG	GIÁ TRỊ / KHÁCH HÀNG
KH VIP	\$610,117,126.09	25.49%	18,088	15.90%	33,730.49
344	\$479,568,905.51	20.04%	12,681	11.15%	37,817.91
334	\$64,059,502.67	2.68%	2,309	2.03%	27,743.40
343	\$61,068,844.19	2.55%	2,962	2.60%	20,617.44
444	\$4,024,927.38	0.17%	88	0.08%	45,737.81
434	\$804,195.56	0.03%	20	0.02%	40,209.78
443	\$590,750.78	0.02%	28	0.02%	21,098.24
Total	\$610,117,126.09	25.49%	18,088	15.90%	33,730.49

NHÓM KH VIP:

Nhóm KH VIP mang lại doanh thu cao nhất cho doanh nghiệp nên cần được chăm sóc chuyên biệt để duy trì sự trung thành. Trong đó, tổ hợp RFM **344 nổi bật với doanh thu vượt trội và số lượng khách hàng cao nhất trong phân khúc VIP**, nhóm này có thể được sử dụng làm “khách hàng mẫu” để phân tích hành vi và xây dựng chân dung khách hàng tương tự trong các chiến dịch tiếp thị. **Tổ hợp 444 và 434 có giá trị ARPU cao nhất toàn bộ dữ liệu với số lượng khách hàng cực kỳ ít**, nhóm này có thể được chọn làm mục tiêu cho các chiến dịch tiếp thị mang tính cá nhân hóa.

PHÂN TÍCH NHÓM KHÁCH HÀNG

BANNER	TỔNG GIÁ TRỊ CHI TIÊU	% GIÁ TRỊ CHI TIÊU	TỔNG SỐ KHÁCH HÀNG	% TỔNG SỐ KHÁCH HÀNG	GIÁ TRỊ / KHÁCH HÀNG
KH VANG LAI	\$360,172,607.46	15.05%	19,911	17.50%	18,089.13
122	\$76,259,705.06	3.19%	4,518	3.97%	16,879.08
123	\$59,325,372.71	2.48%	2,843	2.50%	20,867.17
222	\$54,142,452.32	2.26%	3,212	2.82%	16,856.31
121	\$45,946,103.92	1.92%	3,183	2.80%	14,434.84
223	\$43,466,686.99	1.82%	2,089	1.84%	20,807.41
221	\$34,714,491.07	1.45%	2,429	2.13%	14,291.68
124	\$26,802,133.94	1.12%	956	0.84%	28,035.70
224	\$19,515,661.46	0.82%	681	0.60%	28,657.36
Total	\$360,172,607.46	15.05%	19,911	17.50%	18,089.13

NHÓM KH VẮNG LAI:

Tương tự nhóm KH SẮP RỜI BỎ, doanh nghiệp có thể lựa chọn một số tổ hợp RFM có giá trị **ARPU cao** làm mục tiêu cho các chiến dịch tiếp thị: **123 - 223 (20 nghìn)** và **124 - 224 (28 nghìn)**.

Đây là 4 nhóm cần được ưu tiên triển khai chương trình thúc đẩy tương tác và khuyến khích tái giao dịch, với mục tiêu chuyển đổi dần sang nhóm KH TIỀM NĂNG. Các tổ hợp còn lại vẫn cần theo dõi nhưng nên dành nguồn lực tập trung cho 4 nhóm trọng điểm trên để tối ưu hóa giá trị.

BANNER	TỔNG GIÁ TRỊ CHI TIÊU	% GIÁ TRỊ CHI TIÊU	TỔNG SỐ KHÁCH HÀNG	% TỔNG SỐ KHÁCH HÀNG	GIÁ TRỊ / KHÁCH HÀNG
KH DA RƠI BỎ	\$296,381,496.42	12.38%	19,868	17.46%	14,917.53
111	\$100,068,032.98	4.18%	7,719	6.78%	12,963.86
211	\$72,501,874.53	3.03%	5,611	4.93%	12,921.38
112	\$44,307,337.60	1.85%	2,585	2.27%	17,140.17
212	\$33,350,105.98	1.39%	1,950	1.71%	17,102.62
113	\$18,424,867.53	0.77%	878	0.77%	20,985.04
213	\$11,852,369.18	0.50%	565	0.50%	20,977.64
214	\$8,345,457.75	0.35%	289	0.25%	28,877.02
114	\$7,531,450.87	0.31%	271	0.24%	27,791.33
Total	\$296,381,496.42	12.38%	19,868	17.46%	14,917.53

NHÓM KH ĐÃ RỜI BỎ:

Khách hàng trong nhóm này chủ yếu thuộc tổ hợp RFM 111 và 211. Đặc biệt, nhóm **111 có số lượng khách lớn thứ hai toàn bộ dữ liệu, nhưng điểm chung là họ chỉ thực hiện duy nhất một giao dịch trước khi rời bỏ và giá trị ARPU không cao**. Đây là tín hiệu cảnh báo quan trọng về vấn đề duy trì sau lần mua đầu tiên. Kiến nghị tiến hành khảo sát để xác định nguyên nhân cụ thể, từ đó đề xuất giải pháp cải thiện hành trình khách hàng và giảm tỷ lệ rời bỏ trong tương lai. Tuy nhiên, các tổ hợp có giá trị ARPU cao chỉ chiếm tỷ lệ nhỏ về số lượng và tổng doanh thu cho thấy doanh nghiệp vẫn có khả năng ngăn chặn doanh thu sụt giảm nếu thực hiện các biện pháp giữ chân khách hàng kịp thời.

TỔNG KẾT

Kết quả phân tích RFM cho thấy **tỷ lệ khách hàng rời bỏ đang ở mức cao, khả năng thu hút và tái kích hoạt khách hàng mới thấp**, điều này phản ánh rủi ro thất thoát doanh thu nếu không có hành động kịp thời. Do đó, cần tập trung chiến lược chăm sóc khách hàng theo mức độ ưu tiên sau:

1.NHÓM KH SẮP RỜI BỎ

Chiếm gần 30% tổng chi tiêu (≈ 714 triệu USD) và 28% số khách hàng - **nhóm có doanh thu và lượng khách cao nhất**. Đặc biệt, phần lớn khách hàng giá trị thuộc các tổ hợp RFM như 133, 134, 144, 233, 234, 244. Đây là những khách đã từng nằm trong nhóm VIP hoặc TRUNG THÀNH nhưng **đã lâu không quay lại**. Do đó, cần ưu tiên triển khai các chiến dịch giữ chân và tái kích hoạt ngay để tránh mất nguồn doanh thu lớn.

2.NHÓM KH VIP

Đóng góp khoảng 610 triệu USD (25%), **đứng thứ 2 về doanh thu** nhưng chỉ chiếm 16% số khách hàng. **Giá trị trung bình trên mỗi khách hàng rất cao (ARPU)**, nhóm RFM 444, 434 có giá trị ARPU cao nhất toàn bộ khách hàng, nhóm RFM 344 vừa đông vừa mang lại giá trị vượt trội. Doanh nghiệp cần có chiến lược chăm sóc chuyên biệt, áp dụng ưu đãi cá nhân hóa, đồng thời tận dụng nhóm này làm “khách hàng mẫu” để phân tích hành vi và mở rộng sang phân khúc tương tự.

3.NHÓM KH VẮNG LẠI

Chiếm khoảng 360 triệu USD (15%) với khoảng 20.000 khách hàng, **đứng thứ 2 về quy mô, khách thường chỉ giao dịch ngắn hạn hoặc mua ít lần. Một số phân nhóm có giá trị ARPU rất cao** như RFM 124 (28.035 USD), 224 (28.657 USD), 123 (20.867 USD) và 223 (20.807 USD). Doanh nghiệp nên ưu tiên tập trung vào các phân nhóm này bằng các chương trình khuyến khích tái mua nhằm từng bước chuyển đổi sang Nhóm KH TIỀM NĂNG và KH TRUNG THÀNH.

4.NHÓM KH ĐÃ RỜI BỎ

Chiếm khoảng 285 triệu USD tổng doanh thu (12%) với khoảng 20.000 khách hàng, **đứng thứ 3 về quy mô**. Đáng chú ý, tổ hợp RFM 111 có số lượng khách hàng lớn thứ hai toàn bộ dữ liệu nhưng **chỉ thực hiện duy nhất một giao dịch**. Điều này phản ánh hạn chế trong việc giữ chân khách hàng sau lần mua đầu tiên. Doanh nghiệp nên khảo sát nguyên nhân rời bỏ từ đó tìm giải pháp cải thiện hành trình khách hàng và giảm tỷ lệ rời bỏ trong tương lai.

5.NHÓM KH TRUNG THÀNH

Quy mô nhỏ và mức đóng góp chưa đáng kể so với các nhóm trên. Tuy vậy, nhóm này có tiềm năng duy trì sự ổn định lâu dài nếu được nuôi dưỡng đúng cách. Doanh nghiệp chỉ cần duy trì các hoạt động chăm sóc cơ bản để giữ mối quan hệ, tránh tình trạng họ chuyển sang nhóm rủi ro.

MYSQL QUERY

Trong quá trình phân nhóm khách hàng theo RFM, thay vì áp dụng cách chia đều theo N-tile, báo cáo này sử dụng truy vấn SQL tùy chỉnh kết hợp với phương pháp xác định ngưỡng theo IQR (Interquartile Range). Lý do:

- Phân nhóm bằng N-tile chỉ đảm bảo số lượng khách hàng chia đều, không phản ánh đúng đặc điểm phân bố dữ liệu. Trong thực tế, nhiều khách hàng có cùng giá trị R, F hoặc M nhưng lại bị phân tách vào các nhóm khác nhau, gây ra sự thiếu công bằng trong đánh giá.
- SQL tùy chỉnh cho phép chia nhóm dựa trên giá trị thực tế của dữ liệu, nhờ đó những khách hàng có hành vi tương đồng sẽ được xếp cùng nhóm, đảm bảo kết quả phân tích phản ánh chính xác hơn.
- Việc xác định ngưỡng bằng IQR giúp giảm ảnh hưởng của các giá trị ngoại lệ (outlier), đồng thời phản ánh trung thực sự khác biệt giữa các nhóm khách hàng.

```
SET @endday = '2022-09-01';
CREATE TEMPORARY TABLE `RFM_BASE` AS
WITH RFM_VALUES AS (
    -- Tính Recency, Frequency, Monetary
    SELECT CustomerID, LocationID, BranchCode,
        DATEDIFF( @endday, MAX( new_Purchase_Date ) ) AS Recency,
        COUNT( DISTINCT DATE( new_Purchase_Date ) ) / ( DATEDIFF( @endday , new_Created_date ) / 365 ) AS Frequency,
        SUM( gmV ) / ( DATEDIFF( @endday , new_Created_date ) / 365 ) AS Monetary
    FROM
        customer_transaction ct JOIN customer_registered cr ON ct.CustomerID = cr.ID
    WHERE
        GMV IS NOT NULL AND GMV > 0
    AND
        new_Purchase_Date IS NOT NULL AND new_Purchase_Date < @endday
    AND
        New_Created_date IS NOT NULL AND New_Created_date < @endday
    GROUP BY CustomerID
),
RFM_SORTED AS (
    SELECT *,
        ROW_NUMBER () OVER (ORDER BY recency) AS R_sorted,
        ROW_NUMBER () OVER (ORDER BY frequency) AS F_sorted,
        ROW_NUMBER () OVER (ORDER BY monetary) AS M_sorted
    FROM RFM_VALUES
),
TOTALROWS AS (
    SELECT *, COUNT(*) OVER () AS total_rows
    FROM RFM_SORTED
)
SELECT *,
    total_rows * 0.25 AS Q1,
    total_rows * 0.50 AS Q2,
    total_rows * 0.75 AS Q3
FROM TOTALROWS;
```

MYSQL QUERY

⇒ -- Xác định giá trị Q1-3

```
CREATE TEMPORARY TABLE Q_VALUES AS
```

```
SELECT
    MAX(CASE WHEN R_sorted <= Q1 THEN Recency END) AS Q1R,
    MAX(CASE WHEN R_sorted <= Q2 THEN Recency END) AS Q2R,
    MAX(CASE WHEN R_sorted <= Q3 THEN Recency END) AS Q3R,
    MAX(CASE WHEN F_sorted <= Q1 THEN frequency END) AS Q1F,
    MAX(CASE WHEN F_sorted <= Q2 THEN frequency END) AS Q2F,
    MAX(CASE WHEN F_sorted <= Q3 THEN frequency END) AS Q3F,
    MAX(CASE WHEN M_sorted <= Q1 THEN monetary END) AS Q1M,
    MAX(CASE WHEN M_sorted <= Q2 THEN monetary END) AS Q2M,
    MAX(CASE WHEN M_sorted <= Q3 THEN monetary END) AS Q3M
FROM RFM_BASE;
```

-- tạo bảng tổng hợp các giá trị Q của R-F-M

```
WITH T AS (
```

```
    SELECT s.*, v.line
    FROM Q_VALUES AS s
    JOIN (
        SELECT 1 AS line
        UNION ALL
        SELECT 2
        UNION ALL
        SELECT 3
    ) AS v
    ON 1=1
)
```

```
SELECT
```

```
    CASE line WHEN 1 THEN 'R' WHEN 2 THEN 'F' ELSE 'M' END AS Type,
    CASE line WHEN 1 THEN Q1R WHEN 2 THEN Q1F ELSE Q1M END AS Q1,
    CASE line WHEN 1 THEN Q2R WHEN 2 THEN Q2F ELSE Q2M END AS Q2,
    CASE line WHEN 1 THEN Q3R WHEN 2 THEN Q3F ELSE Q3M END AS Q3
```

```
FROM T;
```

-- Chấm điểm RFM và phân nhóm khách hàng

```
WITH RFM_SCORE AS (
```

```
    SELECT customerID, LocationID, BranchCode, recency, frequency, monetary,
    CASE
        WHEN recency < Q1R THEN '4'
        WHEN recency >= Q1R AND recency < Q2R THEN '3'
        WHEN recency >= Q2R AND recency < Q3R THEN '2'
        ELSE '1'
    END AS R_score,
    CASE
        WHEN frequency < Q1F THEN '1'
        WHEN frequency >= Q1F AND frequency < Q2F THEN '2'
        WHEN frequency >= Q2F AND frequency < Q3F THEN '3'
        ELSE '4'
    END AS F_score,
    CASE
        WHEN monetary < Q1M THEN '1'
        WHEN monetary >= Q1M AND monetary < Q2M THEN '2'
        WHEN monetary >= Q2M AND monetary < Q3M THEN '3'
        ELSE '4'
    END AS M_score
    FROM RFM_BASE, Q_VALUES
    GROUP BY customerID
```

```
),
```

```
RFM_SEGMENT AS (
```

```
    SELECT *,
    CONCAT(R_score, F_score, M_score) AS RFM_group
    FROM RFM_SCORE
```

```
)
```

```
SELECT *,
```

```
    CASE
```

```
        WHEN RFM_GROUP IN ('111','112','113','114','211','212','213','214') THEN 'KH DA ROI BO'
        WHEN RFM_GROUP IN ('121','122','123','124','221','222','223','224') THEN 'KH VANG LAI'
        WHEN RFM_GROUP IN ('131','132','133','134','141','142','143','144','233','234','243','244','231','232','241','242') THEN 'KH SAP ROI BO'
        WHEN RFM_GROUP IN ('311','312','321','322','411','412','421','422') THEN 'KH MOI'
        WHEN RFM_GROUP IN ('313','314','323','324','413','414','423','424') THEN 'KH TIEM NANG'
        WHEN RFM_GROUP IN ('331','332','333','341','342','431','432','433','441','442') THEN 'KH TRUNG THANH'
        WHEN RFM_GROUP IN ('334','343','344','434','443','444') THEN 'KH VIP'
```

```
    END AS BANNER
```

```
FROM RFM_SEGMENT;
```

```
DROP TEMPORARY TABLE `Q_VALUES`;
```

```
DROP TEMPORARY TABLE `RFM_BASE`;
```