



CUSTOMER SEGMENTATION REPORT

HOÀNG HIỆP

MỤC LỤC



03

GIỚI THIỆU
CUSTOMER 360

04

MÔ HÌNH RFM

05

TỔNG QUÁT

06

CÁCH LÀM

07

PHÂN LOẠI KHÁCH HÀNG
THEO THANG ĐIỂM RFM

08 & 09

PHÂN TÍCH CHI TIẾT

10

FILE CODE

GIỚI THIỆU CUSTOMER 360

ĐỊNH NGHĨA

- Customer 360 là **hồ sơ dữ liệu toàn diện** về khách hàng. Nó chứa mọi thông tin từ nhân khẩu học, sở thích, lịch sử mua hàng cho tới hành vi tương tác của khách hàng với doanh nghiệp thông qua các nền tảng online.
- Customer 360 cung cấp cho doanh nghiệp **cái nhìn đầy đủ và toàn diện** về những khách hàng của họ.

LỢI ÍCH

- Nâng cao trải nghiệm khách hàng: Phân tích chi tiết Customer 360 giúp doanh nghiệp thấu hiểu khách hàng rõ hơn. Từ đó có thể **phát triển sản phẩm tốt hơn, phù hợp hơn** với khách hàng.
- Đồng nhất dữ liệu khách hàng giữa các bộ phận trong doanh nghiệp: Customer 360 hạn chế tình trạng phân mảnh dữ liệu giữa các bộ phận. Mọi người trong công ty đều có thể **truy cập dễ dàng vào cùng một dữ liệu** về khách hàng.
- Tăng tỷ lệ trung thành của khách hàng: Sử dụng dữ liệu Customer 360 giúp doanh nghiệp dự đoán hành vi và cá nhân hóa trải nghiệm. Từ đó **đề xuất ưu đãi và thông điệp đúng với nhu cầu** của khách hàng.



RFM Metrics



RECENCY

The freshness of the customer activity, be it purchases or visits

E.g. Time since last order or last engaged with the product



FREQUENCY

The frequency of the customer transactions or visits

E.g. Total number of transactions or average time between transactions/engaged visits



MONETARY

The intention of customer to spend or purchasing power of customer

E.g. Total or average transactions value

MÔ HÌNH RFM

ĐỊNH NGHĨA

Đây là một mô hình phân tích và phân khúc khách hàng dựa trên dữ liệu các giao dịch trong lịch sử. Cụ thể, RFM là:

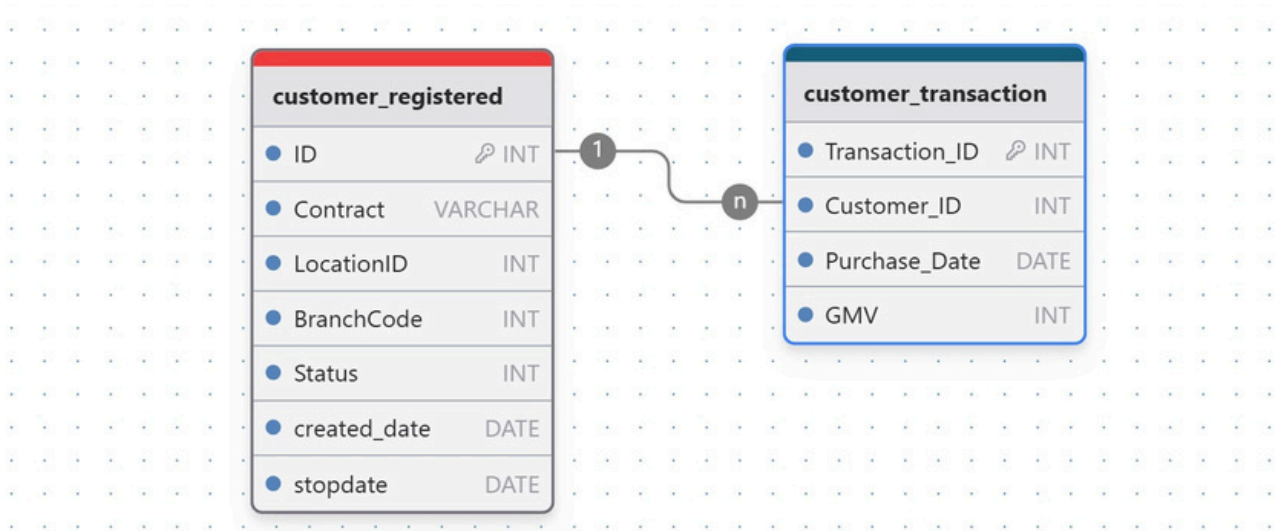
- **Recency:** Khoảng thời gian giữa giao dịch gần nhất và hiện tại.
- **Frequency:** Tần suất mua hàng của khách hàng.
- **Monetary:** Tổng số tiền mà khách hàng đã chi tiêu cho doanh nghiệp trong một khoảng thời gian cụ thể.

LỢI ÍCH

- **Tăng sự trung thành của khách hàng:** RFM giúp doanh nghiệp thực hiện chương trình chăm sóc khách hàng phù hợp.
- **Giảm tỷ lệ khách hàng rời bỏ:** RFM giúp doanh nghiệp dự đoán được các khách hàng này.
- **Tối ưu chi phí marketing:** RFM giúp chọn chính xác đối tượng mục tiêu => Cải thiện đáng kể chi phí.
- **Cải thiện giá trị vòng đời của khách hàng:** RFM giúp upsell và cross-sell cho các phân khúc khách hàng có khả năng phản hồi cao.

TỔNG QUÁT VÀ CÁCH LÀM

Tổng quát



Data model giữa 2 bảng **customer_registered** và **customer_transaction**

	123 Transaction_ID	123 CustomerID	🕒 Purchase_Date	123 GMV
1	0	1,327,813	2022-06-01	95,000
2	1	1,157,830	2022-06-01	75,000
3	2	873,915	2022-07-01	95,000
4	3	3,505,071	2022-07-01	90,000
5	4	2,930,918	2022-07-01	109,091
6	5	899,882	2022-06-01	105,000
7	6	2,248,818	2022-06-01	75,000
8	7	3,331,485	2022-06-01	90,000
9	8	3,497,579	2022-06-01	60,000
10	9	2,636,994	2022-07-01	75,000

Dữ liệu của bảng **customer_transaction** (show 10 row đầu)

	123 ID	📄 Contract	123 LocationID	123 BranchCode	123 Status	🕒 created_date	📅 stopdate
1	1	SGDN00215	8		1	0	2011-11-25 1/5/2012
2	2	SGDN00214	8		1	0	2012-06-14
3	3	SGD374348	8		1	0	2012-11-01
4	4	SGD022064	8		1	2	2011-06-22 5/29/2013
5	5	SGD041015	8		5	2	2011-12-17 11/11/2014
6	6	SGDN00211	[NULL]	[NULL]	2	2015-06-09	9/9/2015
7	7	SGD374348	8		1	3	2012-11-26 12/13/2012
8	8	SGDN00013	0		0	1	2017-12-11
9	9	BEAAA1809	75		0	1	2022-04-04
10	10	BNAAA4298	241		2	1	2022-04-19

Dữ liệu của bảng **customer_registered** (show 10 row đầu)

Cách làm

Bước 1: Tính các chỉ số Recency, Frequency, Monetary

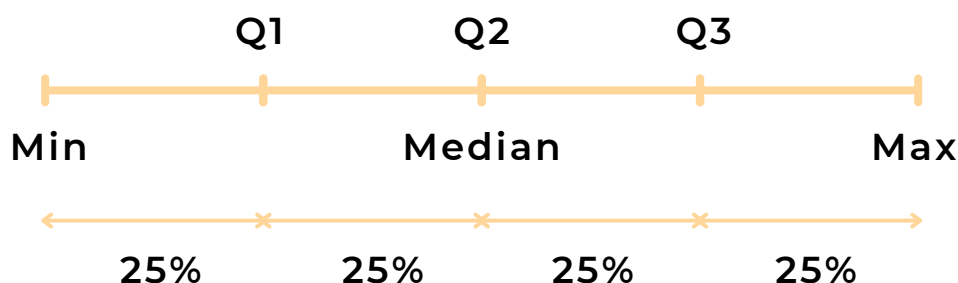
- Recency = Ngày báo cáo - ngày mua hàng gần nhất của khách hàng
- Frequency = Đếm số lần khách mua hàng
- Monetary = Tổng số tiền khách đã chi tiêu

Lưu ý:

+ Frequency và Monetary cần phải chia cho khoảng thời gian (tính theo năm) giữa Ngày báo cáo và Ngày khách tạo tài khoản.

Bước 2: Chấm điểm các chỉ số Recency, Frequency, Monetary dựa trên phương pháp IQR (Interquartile Range)

- Dùng hàm row_number () để đánh số thứ tự xếp hạng cho chỉ số RFM
- Chấm điểm chỉ số RFM từ 1 -> 4 bằng cách tìm tứ phân vị (dùng số thứ tự xếp hạng)



Lưu ý:

+ Frequency và Monetary càng lớn càng tốt -> tỉ lệ thuận với số điểm
+ Recency càng nhỏ càng tốt, vì đây là lần gần nhất khách mua hàng
-> tỉ lệ nghịch với số điểm

Tính điểm	1	2	3	4
R	Từ 92 ngày	62 -> 91 ngày	31 -> 61 ngày	1 -> 30 ngày
F	0,14 -> <0,2	0,2 -> <0,25	0,25	0,25< -> <1,2
M	<15833	15833 -> <18750	18750 -> <23750	23750 -> <142045

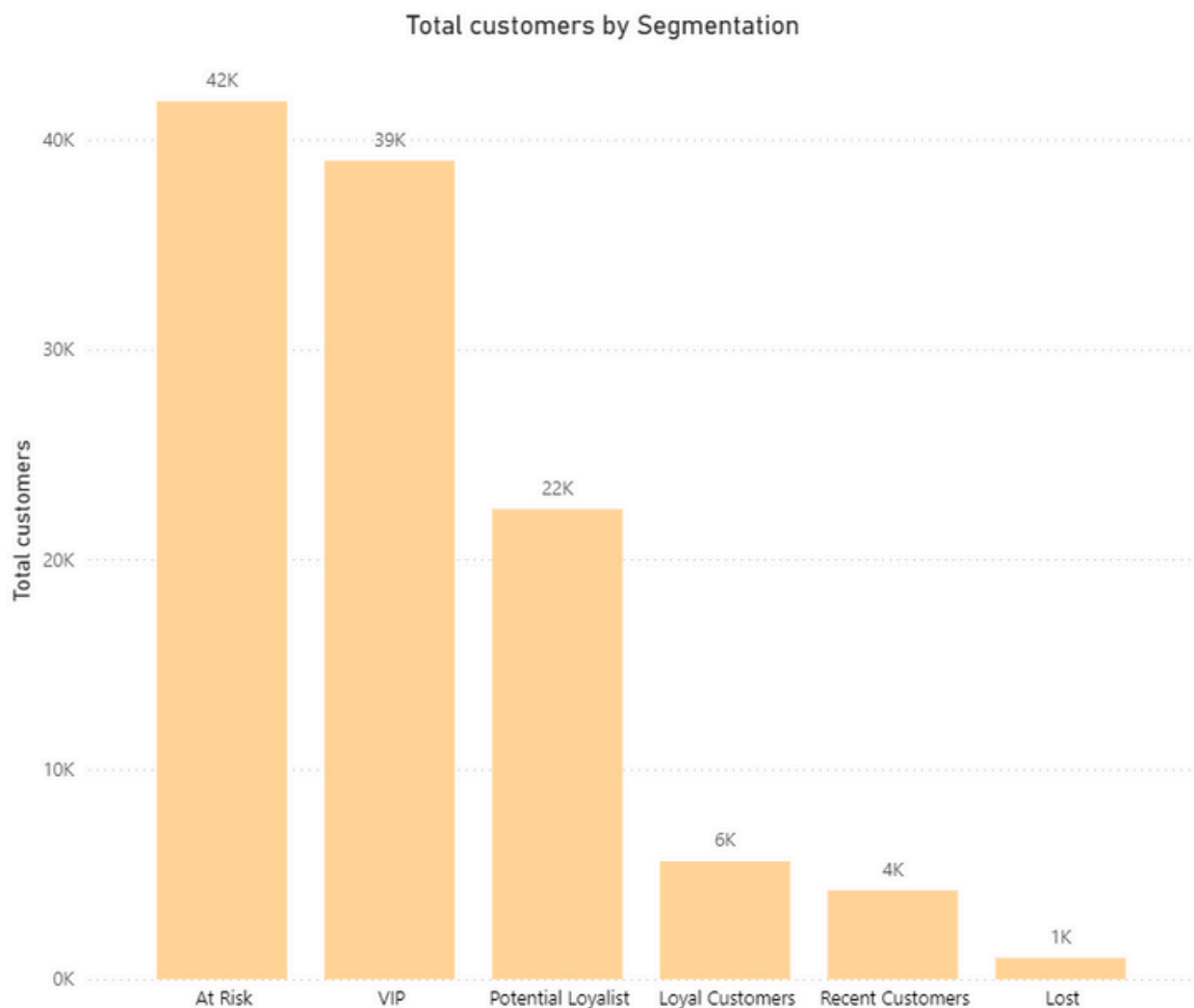
Bước 3: Gộp điểm 3 chỉ số R, F, M lại thành RFM và tính tổng số lượng khách của mỗi tổ hợp

- Dùng hàm concat để gộp điểm 3 chỉ số lại thành RFM
- Dùng hàm count để tính tổng số lượng khách của mỗi tổ hợp RFM
- Phân loại khách hàng thành các phân khúc dựa vào chỉ số RFM

PHÂN LOẠI KHÁCH HÀNG THEO THANG ĐIỂM RFM

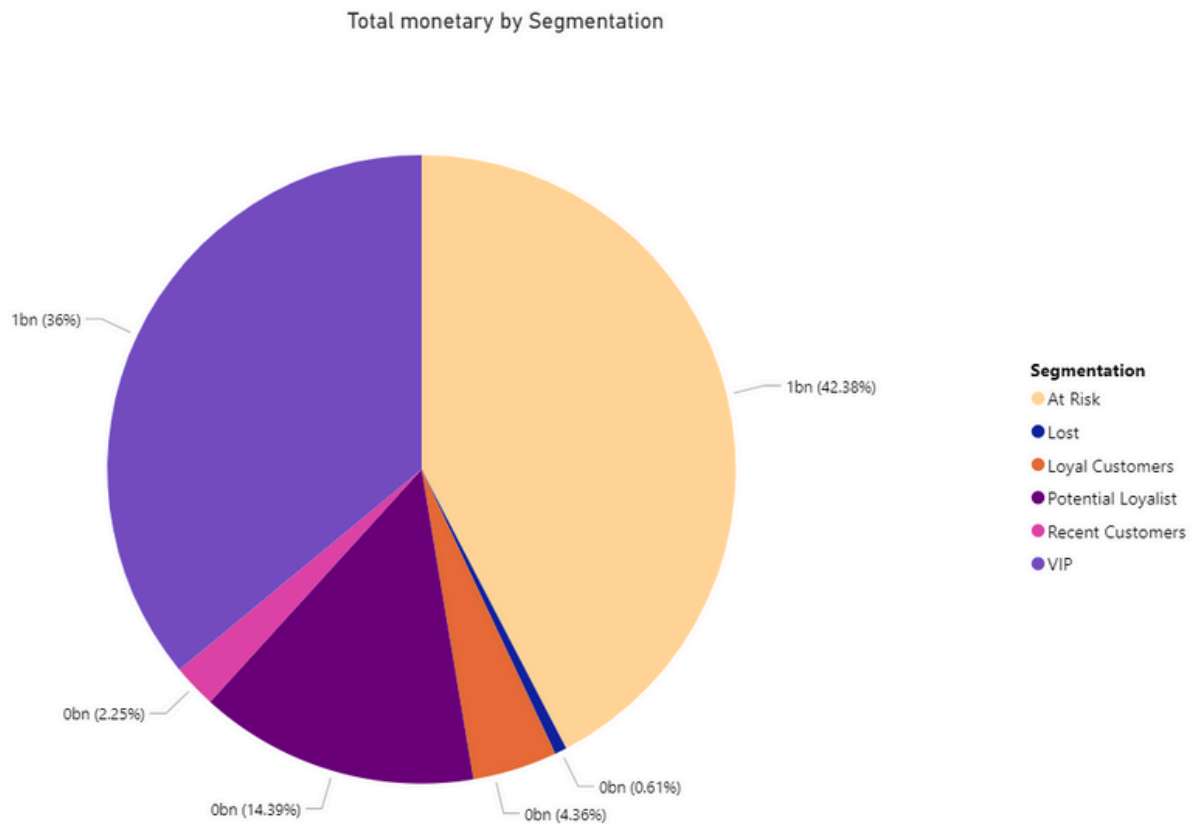
Phân khúc	Mô tả	RFM
VIP	Khách hàng mua gần đây nhất, thường xuyên hoặc khá thường xuyên và có chi tiêu cao	444, 443 344, 343
Loyal Customers	Khách hàng chi tiêu ở mức trung bình - khá nhưng mua hàng thường xuyên	442, 441, 424, 423, 342, 341
Potential Loyalist	Khách hàng mới mua gần đây, chi tiêu trung bình khá	422, 421, 414, 413, 412, 411, 324, 323, 322, 321, 314, 313
Recent Customers	Khách hàng mới mua gần đây nhất, chi tiêu thấp và không mua hàng thường xuyên	312, 311
At Risk	Khách hàng đã khá lâu không quay lại	244, 243, 242, 241, 224, 223, 222, 221, 214, 213, 212, 211
Lost	Khách hàng đã rất lâu không quay lại mua	144, 143, 142, 141, 124, 123, 122, 121, 114, 112, 111

PHÂN TÍCH CHI TIẾT



Trong tổng số 114,081 khách hàng

- Phân khúc At Risk có nhiều số lượng khách hàng nhất (chiếm tới 36.66% khách hàng). Điều này gây ra rủi ro rất lớn cho doanh nghiệp. Cần phải xác định xem điều gì khiến họ không hài lòng và không quay lại mua hàng. Đó có thể là giá của sản phẩm, chất lượng sản phẩm chưa đảm bảo, sản phẩm chưa đáp ứng được nhu cầu của khách hay thị trường có nhiều sản phẩm thay thế...
- Phân khúc VIP đứng thứ 2 với tổng số lượng khách hàng chiếm 34.18%. Doanh nghiệp bằng mọi giá phải giữ chân được những khách hàng này vì đây là phân khúc đã có sự tin tưởng cao vào thương hiệu.
- Phân khúc Potential Loyalist và Loyal Customers lần lượt chiếm 19.63% và 4.93%. Đây là 2 nhóm khách hàng mua khá thường xuyên, nhưng mức chi tiêu ở mức không cao. Doanh nghiệp có thể đưa ra các ưu đãi gắn liền với các ngưỡng chi tiêu, nhằm nâng giá trị cho mỗi đơn hàng.



- Phân khúc At Risk với nhiều số lượng khách hàng nhất và cũng là nhóm mang lại doanh thu nhiều nhất với 42.38%. Doanh nghiệp cần kích thích để những khách hàng này giao dịch lại sớm nhất có thể. Nếu không, thương hiệu có thể sẽ bị khách hàng lãng quên.
- Những khách hàng VIP chiếm 36% tổng doanh thu. Doanh nghiệp có thể đề xuất những chương trình mang giá trị khác biệt và được cá nhân hóa cho nhóm khách hàng này.
- Để đạt được hiệu quả tốt nhất, doanh nghiệp cần tăng trưởng nhóm khách hàng VIP và hạn chế số lượng các khách hàng At Risk. Điều này giúp doanh nghiệp phát triển bền vững và tối ưu ngân sách marketing.

FILE CODE

```
with CTE1 as (
select CustomerID,
       datediff('2022-09-01', max(Purchase_Date)) as Recency,
       count(distinct(Purchase_Date)) / (datediff('2022-09-01', created_date) / 365) as
Frequency,
       sum(GMV) / (datediff('2022-09-01', created_date) / 365) as Monetary,
       row_number () over (order by datediff('2022-09-01', max(Purchase_Date))) as
rn_recency,
       row_number () over (order by count(distinct(Purchase_Date)) / (datediff('2022-
09-01', created_date) / 365)) as rn_frequency,
       row_number () over (order by sum(GMV) / (datediff('2022-09-01',
created_date) / 365)) as rn_monetary
from hw7.customer_registered cr
join hw7.customer_transaction ct on cr.ID=ct.CustomerID
group by CustomerID),
CTE2 as (
select CustomerID, Recency, Frequency, Monetary,
       case
           when rn_recency >= (select min(rn_recency) from CTE1) and rn_recency <
(select count(rn_recency)*0.25 from CTE1) then 4
           when rn_recency >= (select count(rn_recency)*0.25 from CTE1) and
rn_recency < (select count(rn_recency)*0.5 from CTE1) then 3
           when rn_recency >= (select count(rn_recency)*0.5 from CTE1) and
rn_recency < (select count(rn_recency)*0.75 from CTE1) then 2
           else 1
       end as R,
       case
           when rn_frequency >= (select min(rn_frequency) from CTE1) and
rn_frequency < (select count(rn_frequency)*0.25 from CTE1) then 1
           when rn_frequency >= (select count(rn_frequency)*0.25 from CTE1) and
rn_frequency < (select count(rn_frequency)*0.5 from CTE1) then 2
           when rn_frequency >= (select count(rn_frequency)*0.5 from CTE1) and
rn_frequency < (select count(rn_frequency)*0.75 from CTE1) then 3
           else 4
       end as F,
       case
           when rn_monetary >= (select min(rn_monetary) from CTE1) and rn_monetary <
(select count(rn_monetary)*0.25 from CTE1) then 1
           when rn_monetary >= (select count(rn_monetary)*0.25 from CTE1) and
rn_monetary < (select count(rn_monetary)*0.5 from CTE1) then 2
           when rn_monetary >= (select count(rn_monetary)*0.5 from CTE1) and
rn_monetary < (select count(rn_monetary)*0.75 from CTE1) then 3
           else 4
       end as M
from CTE1
group by CustomerID)
select concat(R, F, M) as RFM,
       count(*) as total_customers,
       case
           when concat(R, F, M) in (444, 443, 344, 343) then 'VIP'
           when concat(R, F, M) in (442, 441, 424, 423, 342, 341) then 'Loyal
Customers'
           when concat(R, F, M) in (422, 421, 414, 413, 412, 411, 324, 323, 322,
321, 314, 313) then 'Potential Loyalist'
           when concat(R, F, M) in (312, 311) then 'Recent Customers'
           when concat(R, F, M) in (244, 243, 242, 241, 224, 223, 222, 221, 214,
213, 212, 211) then 'At Risk'
           else 'Lost'
       end as customer_type
from CTE2
group by RFM
order by total_customers desc
```