

CUSTOMER ANALYSIS REPORT



Võ Xuân Đạt

0822.635.618

Datvx150102@gmail.com

MỤC LỤC

I.	MỤC TIÊU.....	1
II.	PHƯƠNG PHÁP THỰC HIỆN.....	2
1.	<i>Customer 360</i>	2
2.	<i>Mô hình phân tích RFM</i>	2
3.	<i>BCG Matrix</i>	3
4.	<i>IQR (interquartile range)</i>	5
III.	BIẾN ĐỔI DỮ LIỆU.....	6
1.	<i>Nhập dữ liệu đầu vào</i>	6
2.	<i>Xây dựng thang điểm RFM và phân khúc khách hàng</i>	6
IV.	TRỰC QUAN HOÁ VÀ PHÂN TÍCH DỮ LIỆU	10
V.	CODE REFERENCE.....	14

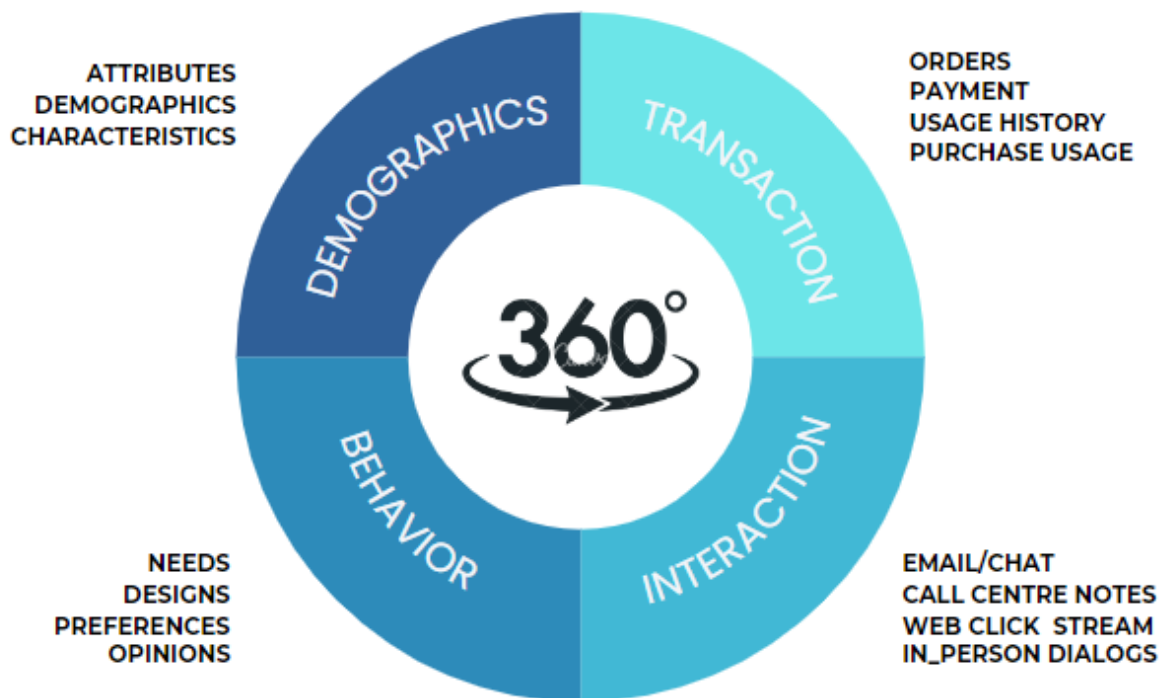
I. MỤC TIÊU

Mục tiêu chính của bài báo cáo này là dựa trên nền tảng customer 360, sử dụng mô hình phân tích RFM (Recency, Frequency, Monetary) và BCG Matrix để **phân loại khách hàng** dựa trên bộ dữ liệu của họ từ đó hỗ trợ đội Chăm Sóc Khách hàng (CSKH) trong việc đạt được mục tiêu là **tối ưu hóa chiến dịch tiếp thị cuối năm**.

II. PHƯƠNG PHÁP THỰC HIỆN

1. Customer 360

Customer 360 là một chiến lược phân tích xoay quanh khách hàng dựa trên dữ liệu về giao dịch (*Transaction Data*); dữ liệu tương tác với website hoặc ứng dụng (*Interaction Data*); dữ liệu về hành vi tiêu dùng của khách hàng (*Behavioral Data*) và dữ liệu về nhân khẩu học của khách hàng (*Demographics Data*)

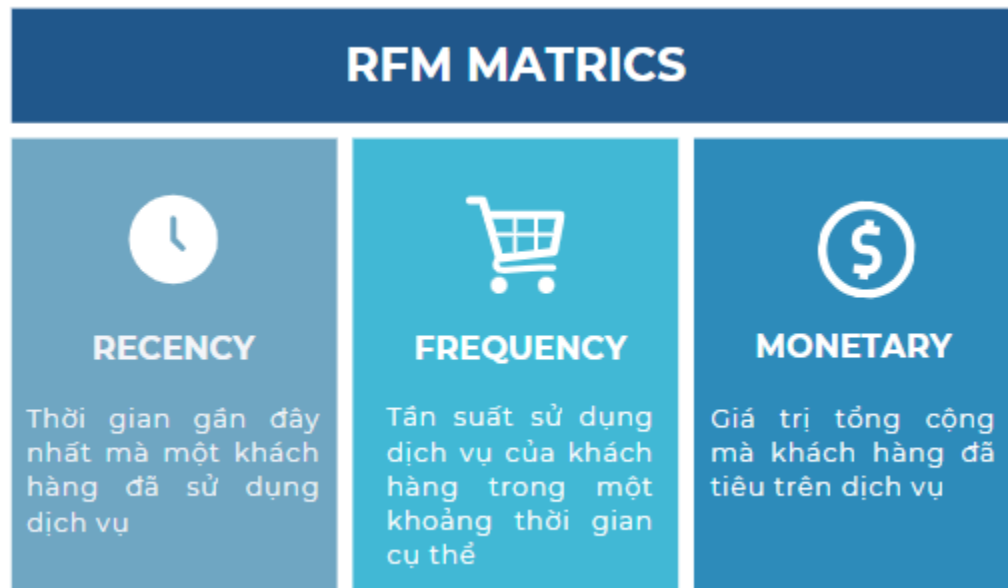


Customer 360 tập trung các thông tin và dữ liệu về khách hàng về một mối giúp cho doanh nghiệp có cái nhìn tổng quan, đa diện hơn về tệp khách hàng của mình. Từ đó doanh nghiệp có thể xây dựng các chiến lược cụ thể nhắm đến từng phân loại khách hàng

2. Mô hình phân tích RFM

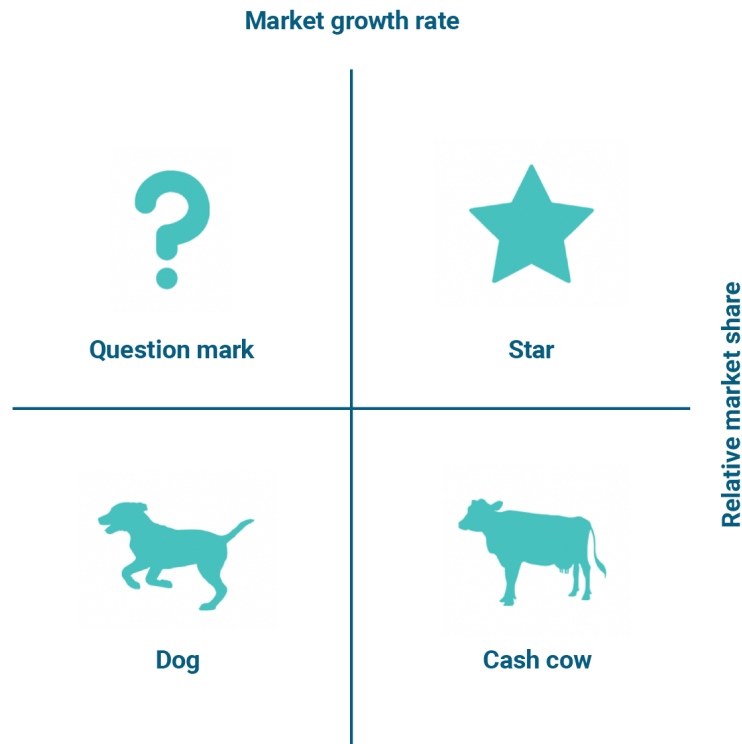
RFM là viết tắt của ba yếu tố chính: **Recency** (R), **Frequency** (F), **Monetary** (M)

Những con số R, F, và M này cung cấp cho chúng ta cái nhìn chi tiết từng khách hàng, cho phép chúng ta phân loại họ vào các nhóm dựa trên độ tương tác, tần suất mua sắm, và giá trị tiêu dùng.



3. BCG Matrix

BCG Matrix _ còn được gọi là ma trận danh mục sản phẩm, là một công cụ kế hoạch kinh doanh được sử dụng để **đánh giá vị trí chiến lược của danh mục thương hiệu** của một công ty.



- **Con hỏi (Question Marks):**

- Sản phẩm có tốc độ tăng trưởng thị trường cao nhưng thị phần thấp.
- Đây là những sản phẩm đòi hỏi nhiều quản lý và đầu tư để tăng thị phần.
- Thường được tài trợ bằng dòng tiền từ phân khúc “Bò sữa”.
- Mục tiêu tốt nhất là biến chúng thành “Ngôi sao”.

- **Bò sữa (Cash Cows):**

- Sản phẩm có thị phần cao nhưng tốc độ tăng trưởng thấp.
- Tạo ra dòng tiền và tự duy trì.
- Mục tiêu là duy trì và tối ưu hóa lợi nhuận từ chúng.

- **Ngôi sao (Stars):**

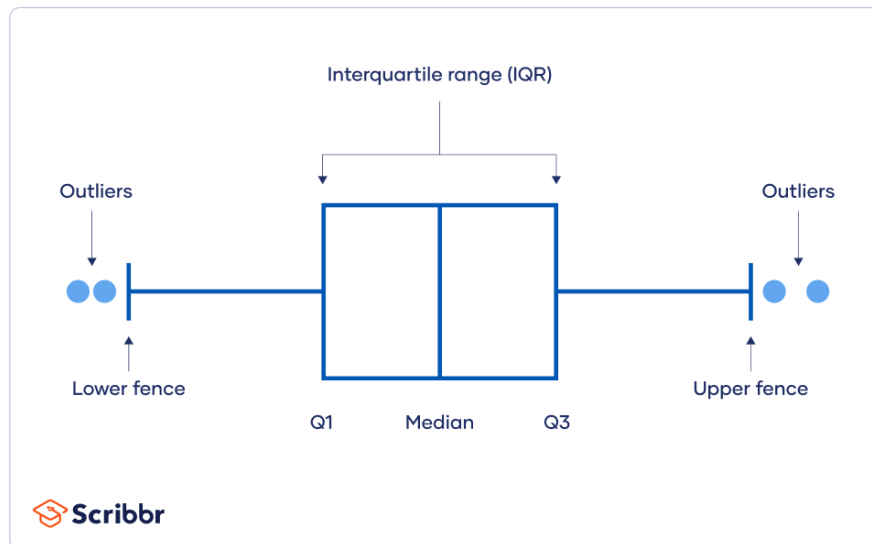
- Sản phẩm có thị phần cao và tốc độ tăng trưởng cao.
- Cạnh tranh mạnh mẽ trong thị trường.
- Đòi hỏi đầu tư để duy trì vị thế dẫn đầu.

- **Con chó (Dogs):**

- Sản phẩm có thị phần thấp và tốc độ tăng trưởng thấp.
- Tạo ra dòng tiền nhưng không bao giờ đạt đến phân khúc “Ngôi sao”.

4. IQR (interquartile range)

IQR là độ trải giữa hay còn gọi là khoảng tứ phân vị của tập dữ liệu. Khái niệm này được sử dụng trong phân tích thống kê để giúp đưa ra kết luận về một tập số. IQR thường được sử dụng cho khoảng biến thiên vì nó loại trừ hầu hết giá trị bất thường của dữ liệu.



III. BIẾN ĐỔI DỮ LIỆU

1. Nhập dữ liệu đầu vào

- Bảng `customer_registered`:

Column Name	#	Data type	Identity	Collation	Not Null
123 id	1	int4			[]
ABC contract	2	varchar(16)		default	[]
123 location_id	3	int4			[]
123 branch_code	4	int4			[]
123 status	5	int4			[]
🕒 created_date	6	timestamp			[]
🕒 stop_date	7	timestamp			[]

`customer_registered` mô tả thông tin **942.338** mã khách hàng (*id*) đã đăng ký thẻ thành viên, với mã hợp đồng (*contract*), mã vị trí (*location_id*), mã chi nhánh (*branch_code*), tình trạng (*code*), ngày đăng ký (*created_date*), ngày hủy (*stop_date*)

- Bảng `customer_transaction`:

Column Name	#	Data type	Identity	Collation	Not Null
123 transaction_id	1	int4			[]
123 customer_id	2	int4			[]
🕒 purchase_date	3	date			[]
123 gmv	4	int4			[]

`customer_transaction` mô tả giao dịch khách hàng đã thực hiện **từ tháng 6/2022 đến tháng 8/2022**, mỗi giao dịch gồm mã giao dịch (*transaction_id*), mã khách hàng thực hiện giao dịch (*customer_id*), ngày giao dịch (*purchase_date*), số tiền giao dịch (*gmv*)

2. Xây dựng thang điểm RFM và phân khúc khách hàng

➤ STEP 1: Tính số liệu Recency, Frequency, Monetary

- Sử dụng **JOIN** kết hợp 2 bảng `customer_registered` và `customer_transaction`
- **Recency**: khoảng thời gian giao dịch gần nhất cho tới thời điểm thực hiện báo cáo
- **Frequency**: tần suất giao dịch từ tính lần đầu khách hàng giao dịch

- **Monetary:** trung bình tổng số tiền khách hàng giao dịch kể từ lần đầu khách hàng đăng ký
- Kết quả:

123 customer_id ▼	123 recency ▼	123 frequency ▼	123 monetary ▼
3,729,649	92	1	60,000
3,650,678	92	1	60,000
3,888,678	92	1	60,000
3,539,570	92	1	60,000
2,960,928	92	1	60,000
4,020,864	92	1	60,000
3,895,624	92	1	60,000
3,768,378	92	1	60,000
3,386,915	92	1	60,000
3,548,972	92	1	60,000
3,803,485	92	1	60,000
4,081,666	92	1	60,000
4,080,790	92	1	60,000
3,559,274	92	1	60,000
3,993,660	92	1	60,000
3,239,098	92	1	60,000
3,644,661	92	1	60,000
4,126,692	92	1	60,000

➤ STEP 2: Ghi điểm RFM theo phương pháp IQR

- Dựa trên **IQR**, mỗi chỉ số R,F,M được chia làm 4 cấp độ 1-2-3-4 tương ứng với mức độ hành vi tích cực tăng dần

IQR	1	2	3	4
recency	92	62-92	31-62	1-31
frequency	1	1	1	1-6
monetary	0-70.000	70.000-75.000	75.000-95.000	95.000-8.400.000

- Kết quả:

123 customer_id ▼	123 recency ▼	123 frequency ▼	123 monetary ▼	123 r ▼	123 f ▼	123 m ▼
3,729,649	92	1	60,000	1	1	1
3,650,678	92	1	60,000	1	1	1
3,888,678	92	1	60,000	1	1	1
3,539,570	92	1	60,000	1	1	1
2,960,928	92	1	60,000	1	1	1
4,020,864	92	1	60,000	1	1	1
3,895,624	92	1	60,000	1	1	1
3,768,378	92	1	60,000	1	1	1
3,386,915	92	1	60,000	1	1	1
3,548,972	92	1	60,000	1	1	1
3,803,485	92	1	60,000	1	1	1
4,081,666	92	1	60,000	1	1	1
4,080,790	92	1	60,000	1	1	1
3,559,274	92	1	60,000	1	1	1
3,993,660	92	1	60,000	1	1	1
3,239,098	92	1	60,000	1	1	1
3,644,661	92	1	60,000	1	1	1
4,126,692	92	1	60,000	1	1	1

➤ STEP 3: Phân khúc khách hàng dựa trên RFM

- Từ 64 tổ hợp RFM, chia thành 10 nhóm khách hàng như sau:

RFM	SEGMENTATION	DESCRIPTION
Champions	444, 443, 434	KH có tầm quan trọng cao nhất, chi tiêu nhiều, sử dụng dịch vụ thường xuyên, gần đây
Loyal Customers	244, 243, 234, 344, 343, 334	KH trung thành với thương hiệu, thường xuyên sử dụng dịch vụ và mức chi tiêu khá cao
Potential Loyalist	313, 331, 314, 341, 324, 342, 323, 332, 322, 333, 413, 431, 414, 441, 424, 442, 423, 432, 422, 433	KH có tiềm năng trở thành KH trung thành, sử dụng dịch vụ khá thường xuyên và gần đây
New Customer	411, 412, 421	KH mới sử dụng dịch vụ gần đây, chưa chi tiêu chưa nhiều và chưa thường xuyên
Promising	311, 312, 321	KH có sử dụng dịch vụ gần đây, có tiềm năng trở thành New Customer
Can not lose them	144, 143, 134	KH chi tiêu nhiều và thường xuyên sử dụng dịch vụ nhưng từ lâu rồi

Needs Attention	213, 231, 214, 241, 224, 242, 223, 232, 233	KH chi tiêu nhiều hoặc sử dụng dịch vụ thường xuyên, cần tiếp cận lại
About to sleep	211, 212, 222	KH từng sử dụng dịch vụ, tuy nhiên không chi tiêu nhiều, không thường xuyên, và không gần đây, tuy chưa quá lâu
At risk Customer	113, 131, 114, 141, 124, 142, 123, 132, 122, 133	KH từng chi tiêu nhiều hoặc sử dụng dịch vụ thường xuyên nhưng đã rất lâu rồi, dễ bị mất
Lost Customer	111, 121, 112	KH đã sử dụng dịch vụ tuy nhiên không còn quay lại nữa

- Kết quả:

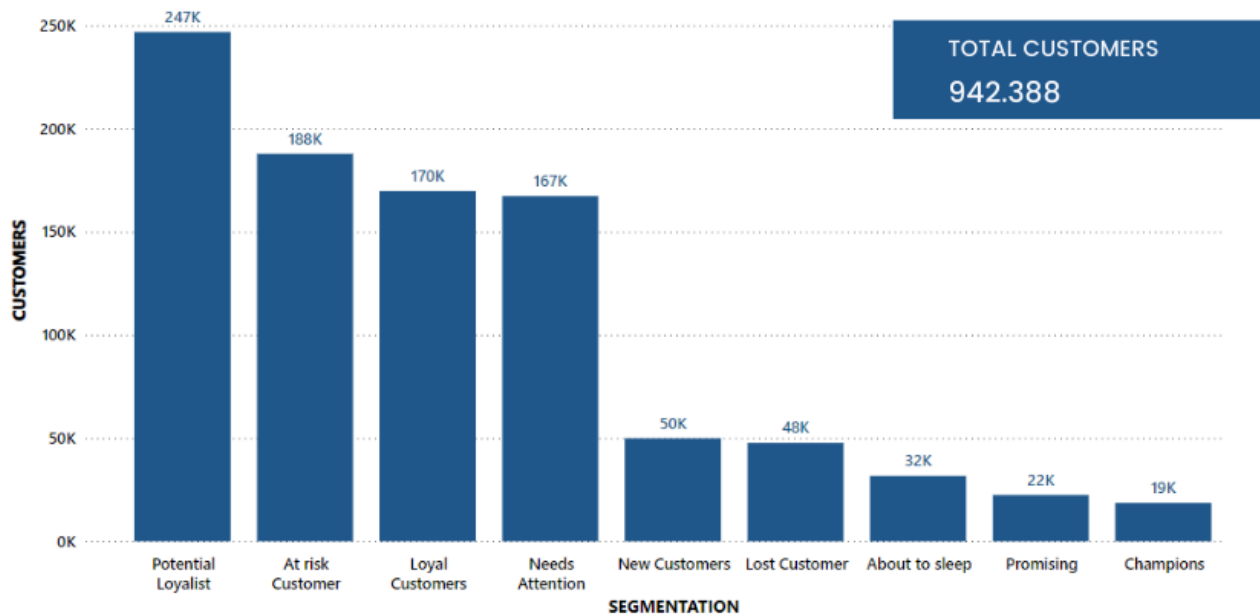
123 customer_id ↑ ▼	123 rfm ▼	ABC rfm_segmentation ▼
75,284	411	New Customers
75,428	244	Loyal Customers
75,457	411	New Customers
75,524	432	Potential Loyalist
75,751	344	Loyal Customers
75,796	244	Loyal Customers
75,816	244	Loyal Customers
75,821	344	Loyal Customers
75,879	333	Potential Loyalist
75,918	444	Champions
75,983	244	Loyal Customers
76,087	344	Loyal Customers
76,133	344	Loyal Customers
76,187	244	Loyal Customers
76,217	344	Loyal Customers
76,249	244	Loyal Customers
76,303	344	Loyal Customers
76,326	344	Loyal Customers
76,332	344	Loyal Customers

IV. TRỰC QUAN HOÁ VÀ PHÂN TÍCH DỮ LIỆU

Sử dụng Power BI kết nối với cơ sở dữ liệu

CUSTOMER BY SEGMENTATION

- Tổng cộng **942.338 khách hàng** được chia vào **9 phân khúc**.



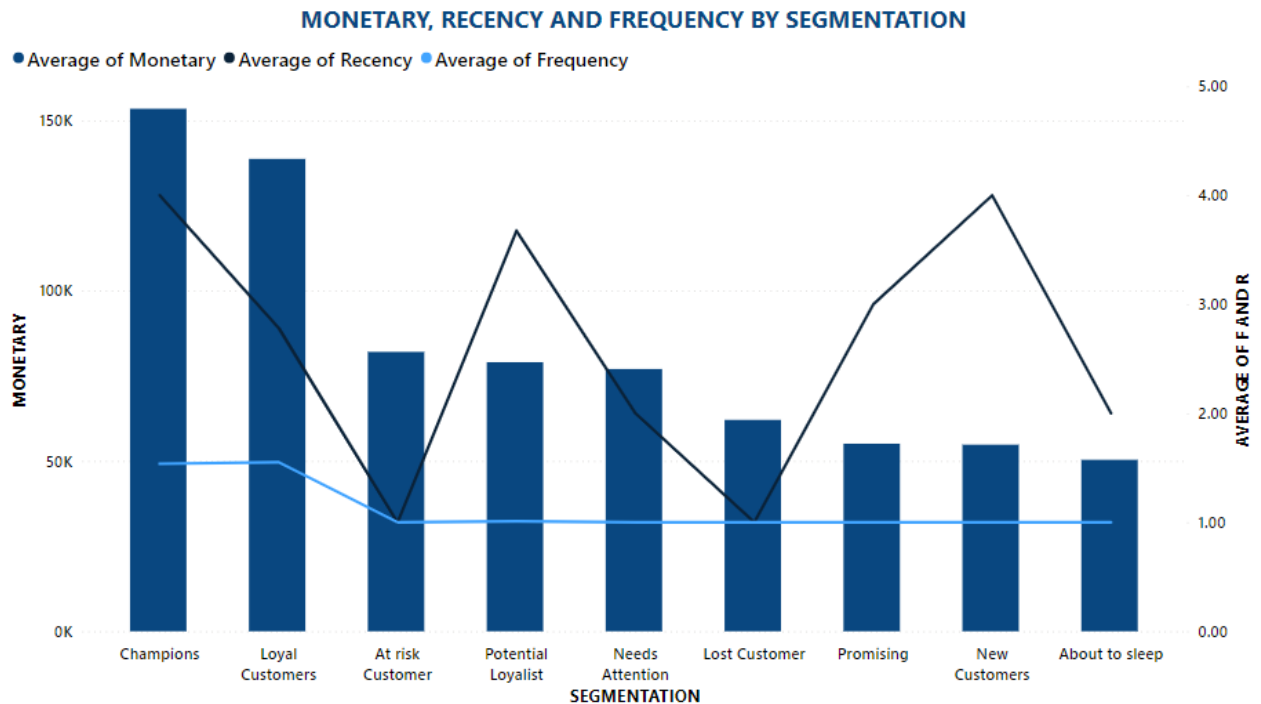
- Phân khúc Potential Loyalist **sử dụng dịch vụ khá thường xuyên** và gần đây có **số lượng KH nhiều nhất** với gần 250k.
- Theo sau đó là 3 phân khúc tương đối bằng nhau là At risk Customer, Loyal Customers và Needs Attention có chưa từ gần 170k đến gần 190k khách hàng.
⇒ Đây là những phân khúc khá nhạy cảm khi cần có những chiến lược để **thúc đẩy thêm khả năng chi tiêu** của Loyal Customers và **tăng tiếp cận** đối với At risk Customer và Needs Attention
- Tập khách hàng quan trọng nhất là Champions xếp sau cùng với số lượng khách hàng ít nhất với chỉ gần 20k

FREQUENCY AND RECENCY BY SEGMENTATION

- Không ngạc nhiên khi tỉ lệ doanh thu / người của phân khúc Champions và Loyal Customers là **lớn nhất, chi tiêu thường xuyên và gần đây nhất**

- Phân khúc At risk Customer có tỉ lệ doanh thu / người cao thứ 3 nhưng họ **không đến thường xuyên và mức 'làm mới' cũng thấp**

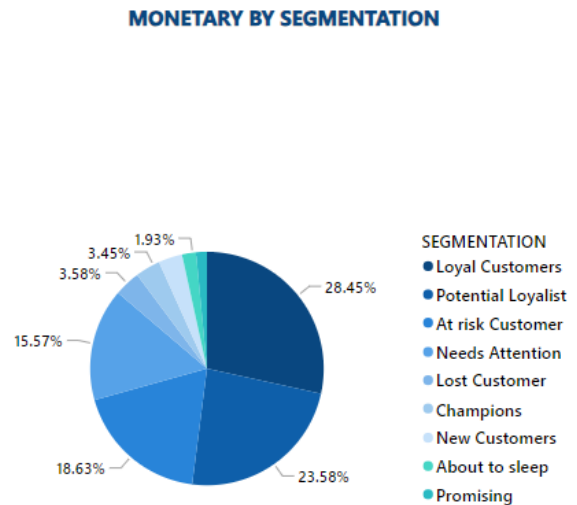
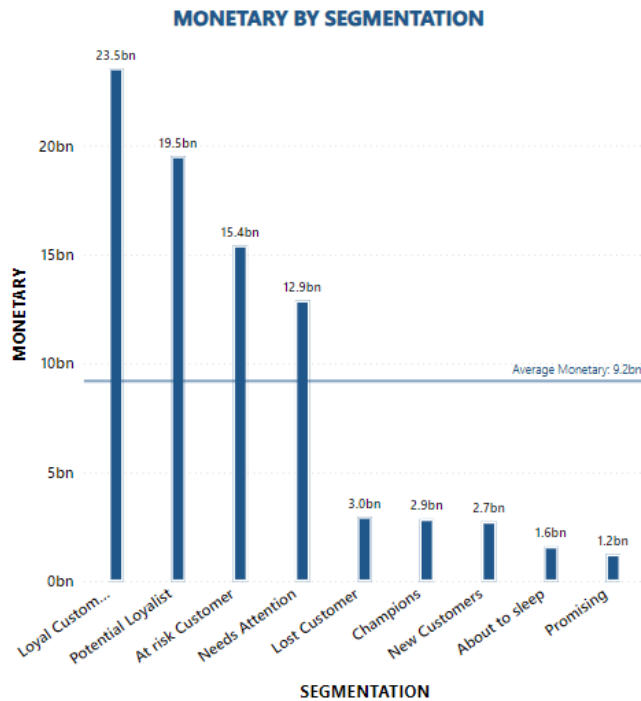
⇒ Cần có chiến lược để phân khúc KH này **sử dụng dịch vụ nhiều hơn và thương xuyên hơn**



- Phân khúc Potential Loyalist , Promising, New Customers sử dụng gần đây nhất nhưng chưa thường xuyên

⇒ Cần có chiến lược thúc đẩy phân khúc KH này **đến thường xuyên và chi tiêu nhiều hơn**

MONETARY BY SEGMENTATION



- 4 phân khúc với **số lượng KH lớn nhất cũng đem về doanh thu cao nhất**: Loyal Customers, Potential Loyalist, At risk Customer và Needs Attention
- Phân khúc quan trọng nhất Champions với số lượng KH ít nhất chỉ đem lại được **3,45% doanh thu**
- Phân khúc New Customer với số lượng KH đứng top 5 (hơn 50k, gấp 3 số lượng KH của phân khúc Champions) đem về tổng doanh thu gần bằng phân khúc Champions
- Phân khúc Promising chiếm tỉ lệ doanh thu **ít nhất với 1.93%**

Recommendations:

- Tạo ra các **ưu đãi đặc biệt, giảm giá và dịch vụ độc quyền với quyền lợi riêng biệt** chỉ dành cho khách hàng Champions để giữ chân họ.
- ⇒ Điều này cũng **thu hút thêm nhóm khách hàng Loyal Customers** để họ bỏ tiền ra chi tiêu nhiều hơn

- Đối với nhóm khách hàng At risk Customers và Needs Attention **xây dựng lại mối quan hệ với khách hàng** thông qua email, gọi điện, gửi thư cảm ơn hoặc cung cấp ưu đãi/ khuyến mãi dành riêng cho họ, để **thúc đẩy động lực quay lại mua hàng**.
- ⇒ Điều này cũng giúp cho nhóm Potential Loyalist quay lại chi tiêu thương xuyên hơn.
- Ngoài ra cần **tạo trải nghiệm khách hàng mạnh mẽ** và cung cấp **trải nghiệm cá nhân hoá** cho nhóm Promising để thúc đẩy họ trở thành các New Customers.

V. CODE REFERENCE

```
WITH rfm_table as
(
    SELECT
        customer_id,
        '2022-09-01'::date - max(purchase_date) recency,
        count(transaction_id) frequency,
        sum(gmv) monetary
    FROM customer_transaction ct
    JOIN customer_registered cr ON ct.customer_id = cr.id
    WHERE customer_id <> 0
    GROUP BY customer_id
),
rfm_calculation AS
(
    SELECT
        *,
        NTILE(4) OVER (ORDER BY recency DESC) r,
        NTILE(4) OVER (ORDER BY frequency) f,
        NTILE(4) OVER (ORDER BY monetary) m
    FROM rfm_table
),
rfm_chart AS
(
    SELECT
        *,
        CAST(concat(r,f,m) AS int) AS rfm
    FROM rfm_calculation
),
rfm_segmentation AS
(
    SELECT
        *,
        CASE WHEN rfm IN (444,443,434) THEN 'Champions'
        WHEN rfm IN (244,243,234,344,343,334) THEN 'Loyal Customers'
        WHEN rfm IN
        (313,331,314,341,324,342,323,332,322,333,413,431,414,
        441,424,442,423,432,422,433) THEN
        'Potential Loyalist'
        WHEN rfm IN (311,312,321) THEN 'Promising'
        WHEN rfm IN (411,412,421) THEN 'New Customers'
        WHEN rfm IN (144,143,134) THEN 'Can not lose them'
        WHEN rfm IN (213,231,214,241,224,242,223,232,233) THEN
        'Needs Attention'
        WHEN rfm IN (211,212,221,222) THEN 'About to sleep'
        WHEN rfm IN (113,131,114,141,124,142,123,132,122,133) THEN
        'At risk Customer'
        WHEN rfm IN (111,121,112) THEN 'Lost Customer'
        END AS rfm_segmentation
    FROM rfm_chart
)
SELECT *
FROM rfm_segmentation;
```