



CUSTOMER ANALYTICS REPORT

By Nguyễn Doãn Hoàng
March, 2024

Mục lục

- 01** **Customer360**
- 02** **RFM model**
- 03** **BCG matrix và
Thang điểm RFM**
- 04** **Phân nhóm khách hàng**
- 05** **Phân tích chi tiết**
- 06** **Kết luận**
- 07** **Phụ lục**

Đặt vấn đề

Thế giới kinh doanh ngày nay đang chứng kiến sự bùng nổ của dữ liệu. Lượng thông tin khổng lồ được tạo ra mỗi ngày từ các hoạt động kinh doanh, tương tác khách hàng và các nguồn khác mang đến tiềm năng to lớn cho các doanh nghiệp trong việc **thấu hiểu thị trường, tối ưu hóa chiến lược và gia tăng lợi nhuận**.

Tuy nhiên, **biến dữ liệu thành những thông tin hữu ích và khả dụng** lại là một thách thức lớn. Các doanh nghiệp, đặc biệt là các doanh nghiệp lớn, cần có **công cụ mạnh mẽ để thu thập, phân tích và tổng hợp dữ liệu** từ nhiều nguồn khác nhau, từ đó **xây dựng chiến lược hiệu quả và nhắm đúng đối tượng mục tiêu**.

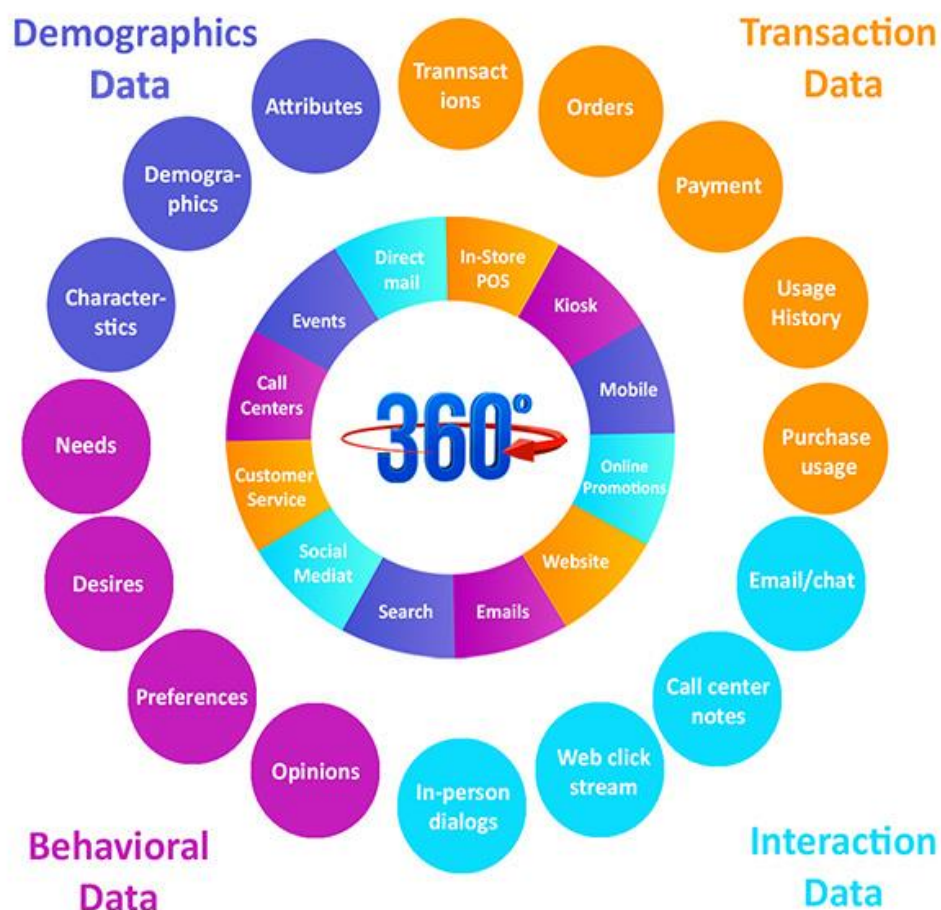
Customer360 chính là **chìa khóa vàng** giúp doanh nghiệp giải quyết bài toán này. Là một nền tảng tiên tiến, Customer360 cung cấp cho doanh nghiệp **giải pháp toàn diện để quản lý thông tin khách hàng một cách hiệu quả**.

1. CUSTOMER360

Vẽ chân dung khách hàng từ từng mảnh ghép dữ liệu

Customer360 không chỉ là một công cụ, mà còn là một phương pháp tiên tiến để phân tích thông tin khách hàng, tập trung vào việc xây dựng hồ sơ khách hàng toàn diện bằng cách gắn kết dữ liệu từ nhiều nền tảng khác nhau, thông qua các mảnh ghép dữ liệu đa dạng, bao gồm:

- **Dữ liệu giao dịch (Transaction Data):** Thông tin về lịch sử mua hàng, bao gồm sản phẩm đã mua, số lượng, giá cả và tần suất mua hàng.
- **Dữ liệu tương tác (Interaction Data):** Cách thức khách hàng tương tác với website, ứng dụng di động, email hoặc các kênh khác của doanh nghiệp.
- **Dữ liệu hành vi (Behavioral Data):** Thông tin về sở thích, thói quen và quan điểm
- **Dữ liệu nhân khẩu học (Demographics Data):** Thông tin cơ bản về khách hàng như tuổi tác, giới tính, nghề nghiệp, thu nhập và địa điểm.



1. CUSTOMER360

Vẽ chân dung khách hàng từ từng mảnh ghép dữ liệu

Bằng cách kết hợp tất cả các nguồn dữ liệu trên, Customer360 giúp doanh nghiệp **hiểu rõ từng khách hàng ở một mức độ sâu sắc**. Doanh nghiệp có thể nhận diện các mẫu hình hành vi, **dự đoán nhu cầu** và **cá nhân hóa trải nghiệm** cho từng khách hàng, từ đó **nâng cao mức độ hài lòng và lòng trung thành**.

Báo cáo này sẽ tập trung phân tích **Transaction data**, từ đó tiến hành phân thành các nhóm khách hàng.

Mục đích: doanh nghiệp có thể xây dựng các chiến lược cụ thể nhắm đến từng phân loại khách hàng, nâng cao trải nghiệm của khách hàng và năng lực cạnh tranh của doanh nghiệp trên thị trường.



2. RFM MODEL

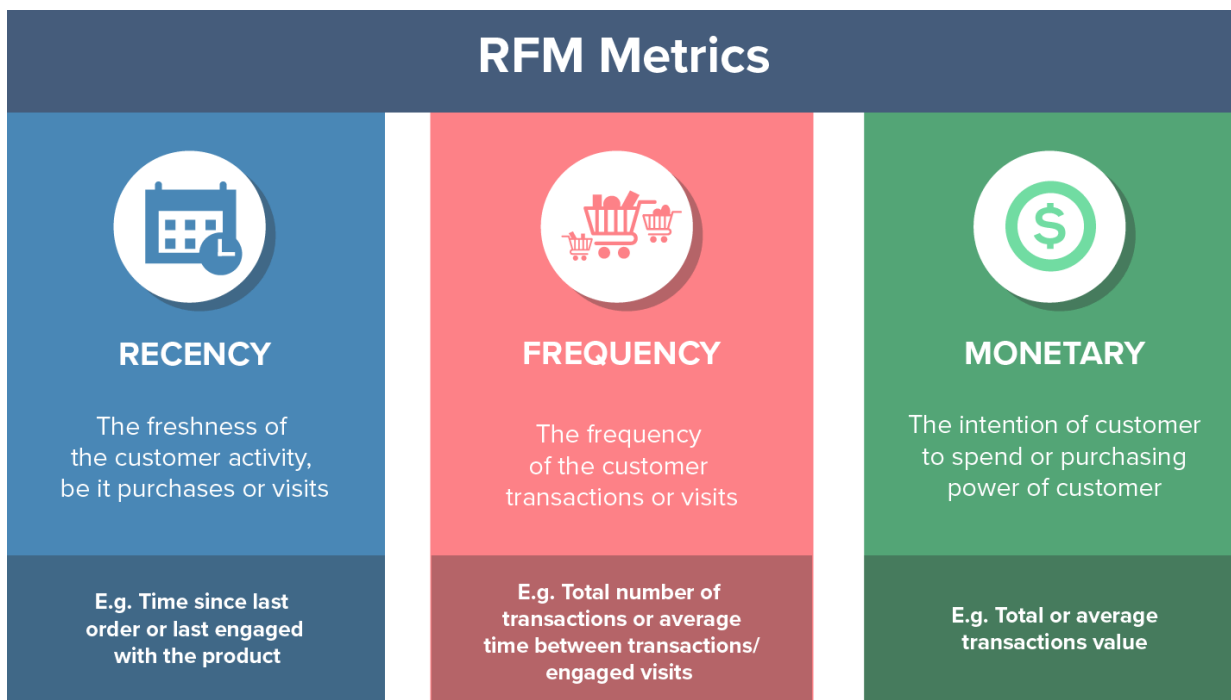
Giải mã hành vi khách hàng qua 3 yếu tố cốt lõi

RFM là một mô hình phân tích và phân khúc khách hàng dựa trên hành vi tiêu dùng của họ. Mô hình này tận dụng dữ liệu giao dịch từ lịch sử mua hàng để đánh giá khách hàng trên ba yếu tố cốt lõi:

R (Recency): Tần suất mua gần đây: Yếu tố này cho biết khoảng thời gian kể từ lần mua hàng cuối cùng của khách hàng. Khách hàng có giao dịch gần đây được đánh giá cao hơn vì thể hiện sự quan tâm duy trì với doanh nghiệp.

F (Frequency): Mức độ mua hàng thường xuyên: Yếu tố này cho biết tần suất mua hàng của khách hàng trong một khoảng thời gian nhất định. Khách hàng mua hàng thường xuyên được coi là khách hàng trung thành và có giá trị cao.

M (Monetary): Giá trị đơn hàng và tổng giá trị mua hàng: Yếu tố này đánh giá mức độ chi tiêu của khách hàng. Khách hàng chi tiêu nhiều hơn được coi là khách hàng có giá trị cao.



DỮ LIỆU

Dữ liệu được lấy từ 2 bảng:

- customer_registered: Bảng chứa thông tin khi đăng ký của user
- customer_transaction: Bảng chứa thông tin giao dịch của user từ tháng 6/2022 đến tháng 8/2022. Mỗi hàng trên bảng là mỗi giao dịch.

customer_registered

Column Name	#	Data Type	Comment
123 ID	1	int	Mã khách hàng
ABC Contract	2	varchar(50)	Mã hợp đồng
123 LocationID	3	int	Mã vùng
123 BranchCode	4	int	Mã chi nhánh
123 Status	5	int	Trạng thái
ABC created_date	6	varchar(50)	Ngày đăng kí
ABC stopdate	7	varchar(50)	Ngày hủy

customer_transaction

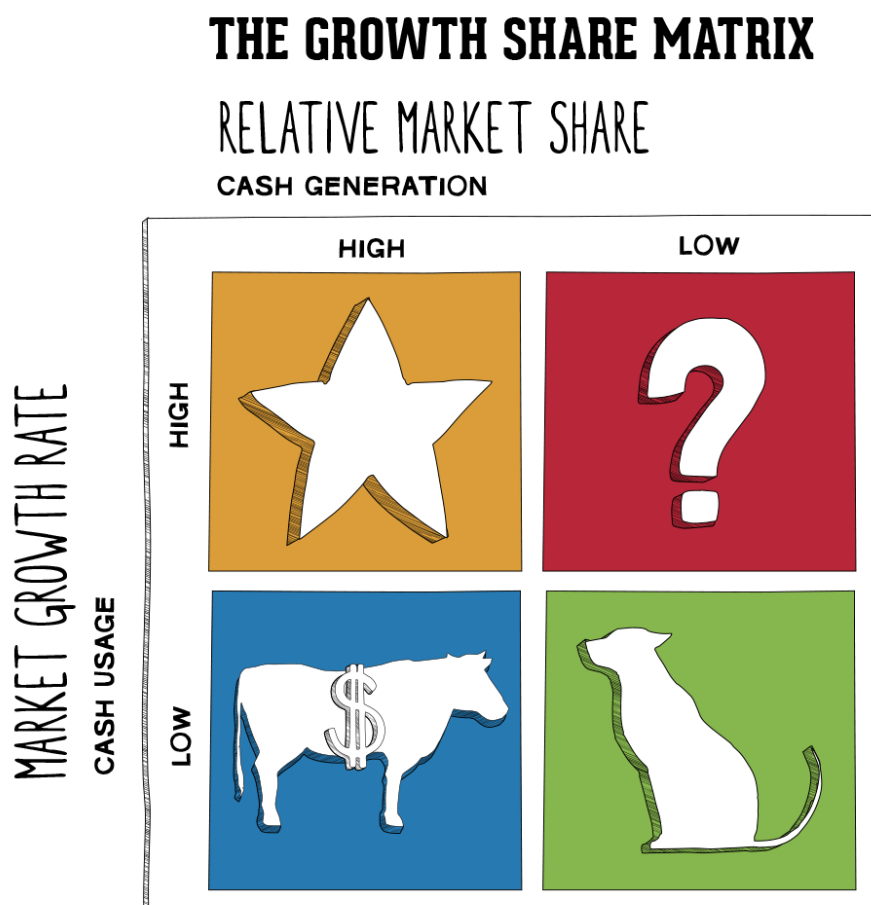
Column Name	#	Data Type	Comment
123 Transaction_ID	1	int	Mã giao dịch
123 CustomerID	2	int	Mã khách hàng
ABC Purchase_Date	3	varchar(50)	Ngày thực hiện giao dịch
123 GMV	4	int	Giá trị giao dịch

3. BCG MATRIX

Báo cáo này ứng dụng lí thuyết của ma trận BCG để phân khách hàng thành các nhóm. BCG là viết tắt của Boston Consulting Group, nghĩa là chiến lược kinh doanh được xây dựng bởi Boston BCG nhằm định hướng xây dựng chiến lược tăng trưởng thị phần cho một loại sản phẩm hay dịch vụ bất kỳ

Ma trận Boston phân tích và định hướng giúp doanh nghiệp đánh giá được thị phần, cơ hội tiềm năng đầu tư hay nguy cơ rủi ro để quyết định ra mắt sản phẩm/dịch vụ hay loại bỏ để tránh thua lỗ. Cụ thể, ma trận BCG sẽ phân tích 2 yếu tố quan trọng:

- Thị phần (Market share): Thị phần của sản phẩm thấp hoặc cao trên thị trường
- Triển vọng phát triển (Market growth): Triển vọng của khách hàng tiềm năng trên thị trường



3. THANG ĐIỂM RFM

Bằng cách **phân tích và kết hợp cả ba yếu tố RFM**, doanh nghiệp có thể phân loại khách hàng thành các nhóm khác nhau. Điều này giúp doanh nghiệp:

- **Hiểu rõ hơn về hành vi và giá trị** của từng nhóm khách hàng.
- **Xác định nhóm khách hàng tiềm năng** và nhóm cần gia tăng tương tác.
- **Xây dựng các chiến lược marketing và chăm sóc khách hàng** phù hợp cho từng nhóm.

Mỗi khách hàng sẽ được tính điểm về 3 yếu tố R, F, M. Mỗi yếu tố được tính điểm tương đối với toàn bộ khách hàng, và được chấm trên **thang điểm từ 1 tới 4**, trong đó 1 là điểm thấp nhất và 4 là điểm cao nhất so với toàn bộ tệp khách hàng. Giá trị của từng tham số thống kê như sau:

	Min	25%	50%	75%	Max
Recency	1	31	62	91	92
Frequency	0.1375	0.1904	0.2286	0.2579	1.4494
Monetary	0	75000	85000	105000	568182

Trong đó:

- Recency (ngày) được tính bằng cách lấy ngày 09-08-2022 trừ đi ngày giao dịch gần nhất
- Frequency (giao dịch/năm) được tính bằng cách lấy số lượng giao dịch chia cho tuổi của khách hàng tính theo năm.
- Monetary được tính bằng cách tính tổng giá trị các giao dịch của khách hàng trong khoảng thời gian phân tích.

Điểm của từng yếu tố được gán thành các số từ 1 đến 4

Dựa vào phân bố dữ liệu, mức điểm RFM của mỗi khách hàng được chia như sau:

RECENCY	
Mức 1	Lần sử dụng dịch vụ gần nhất cách đây ≥ 92 ngày
Mức 2	Lần sử dụng dịch vụ gần nhất cách đây từ 62 đến 91 ngày
Mức 3	Lần sử dụng dịch vụ gần nhất cách đây từ 31 đến 61 ngày
Mức 4	Lần sử dụng dịch vụ gần nhất cách đây từ 1 đến 30 ngày
FREQUENCY trong thời gian từ 01/06/22 đến 31/08/22	
Mức 1	Sử dụng dịch vụ trung bình ≤ 0.1904 lần trên năm
Mức 2	Sử dụng dịch vụ trung bình > 0.1904 và ≤ 0.2286 lần trên năm
Mức 3	Sử dụng dịch vụ trung bình > 0.2286 và ≤ 0.2579 lần trên năm
Mức 4	Sử dụng dịch vụ trung bình > 0.2579 lần trên năm
MONETARY	
Mức 1	Tổng số tiền chi tiêu ≤ 75000 VND
Mức 2	Tổng số tiền chi tiêu > 75000 và ≤ 85000 VND
Mức 3	Tổng số tiền chi tiêu > 85000 và ≤ 105000 VND
Mức 4	Tổng số tiền chi tiêu > 105000 VND

4. PHÂN NHÓM KHÁCH HÀNG

Theo thang điểm RFM

Sau khi tính điểm 3 yếu tố cho từng khách hàng, 3 số điểm này sẽ được gộp lại thành một chỉ số RFM duy nhất dùng để phân loại khách hàng vào từng nhóm:

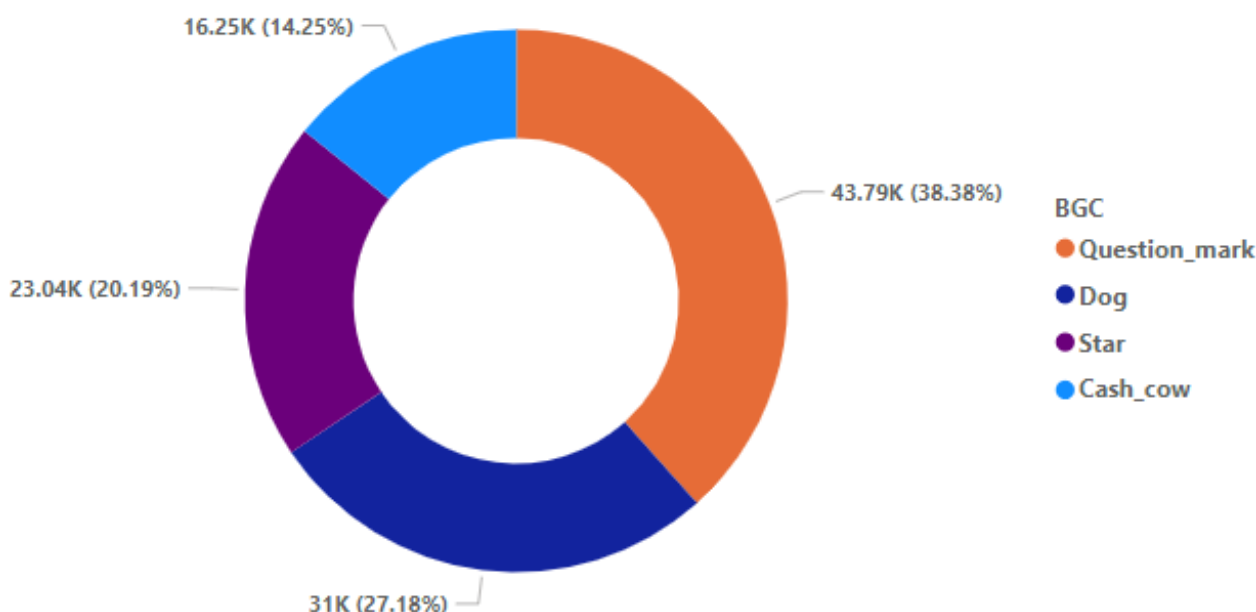
Nhóm	Mô tả	Tổ hợp điểm
Nhóm khách hàng VIP (Star)	Khách hàng có tầm quan trọng cao nhất; họ chi tiêu nhiều, sử dụng dịch vụ thường xuyên, và sử dụng dịch vụ gần đây	444, 443, 434, 433, 344, 343, 334, 333
Nhóm khách hàng thân thiết (Cash cow)	Khách hàng trung thành với thương hiệu; họ chi tiêu nhiều tương đương với các khách hàng VIP, có thể họ thường xuyên sử dụng dịch vụ nhưng lại sử dụng dịch vụ trong thời gian trước đây	424, 423, 413, 414, 313, 314, 323, 324, 233, 234, 244, 243,
	Khách hàng trung thành với thương hiệu; họ chi tiêu nhiều tương đương với các khách hàng VIP, có thể họ ít thường xuyên sử dụng dịch vụ nhưng lại sử dụng dịch vụ trong thời gian gần đây.	133, 134, 143, 144
Nhóm khách hàng tiềm năng (Question mark)	Khách hàng có tiềm năng trở thành các khách hàng thân thiết; họ sử dụng dịch vụ ít thường xuyên hơn, nhưng có mức chi tiêu tương đối.	441, 442, 431, 432, 341, 342, 331, 332, 131,
	Khách hàng chi tiêu ít hơn khách hàng trung thành, nhưng sử dụng dịch vụ khá thường xuyên và khá gần đây.	132, 141, 142, 241, 242, 231,
	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.	232, 113, 114, 123, 124, 213, 214, 223, 224
	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.	
Nhóm khách hàng vắng lai (Dog)	Khách hàng đã từng sử dụng dịch vụ, tuy nhiên không chi tiêu nhiều, không thường xuyên, và không gần đây.	Các tổ hợp còn lại
	Khách hàng đã sử dụng dịch vụ tuy nhiên không còn quay lại nữa	

5. PHÂN TÍCH CHI TIẾT

Số lượng khách hàng từng nhóm

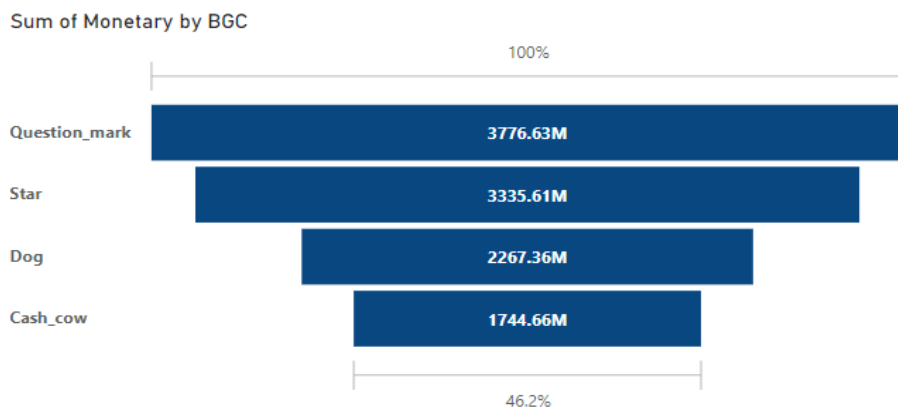
Dữ liệu được tổng hợp từ 114,081 khách hàng, với dữ liệu mua hàng từ 01-06-2022 đến 09-08-2022.

Count of CustomerID by BGC

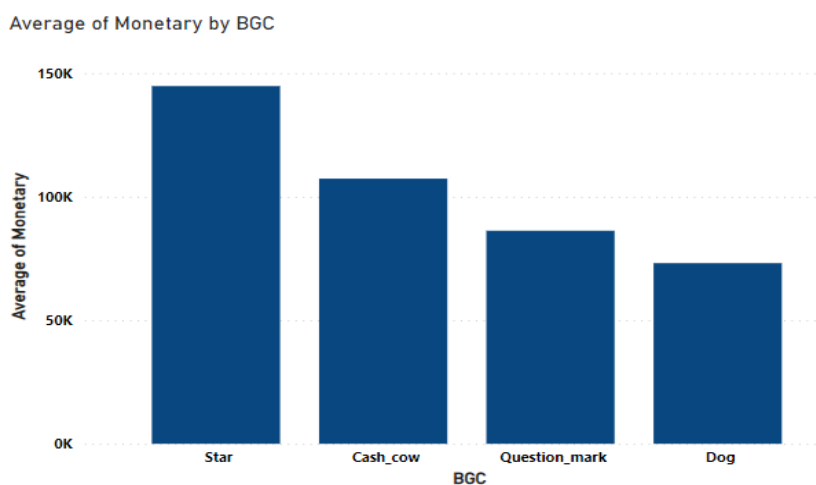


Trong tổng số **hơn 110K khách hàng**, có đến **38.38%** khách hàng **tiềm năng**, tương đương với **43.79K** người. Xếp **thứ 2** là nhóm khách **vãng lai**, có **31K** người tương ứng với **27.18%**. Tập khách hàng quan trọng nhất và có nhiều sức ảnh hưởng nhất là nhóm **khách VIP**, chiếm tỉ trọng **20.19%** và xếp **thứ 3**. Chiếm số lượng ít nhất là nhóm khách **than thiết** với **16.2K** người, tương đương **14.25%**.

Doanh thu theo từng nhóm khách hàng

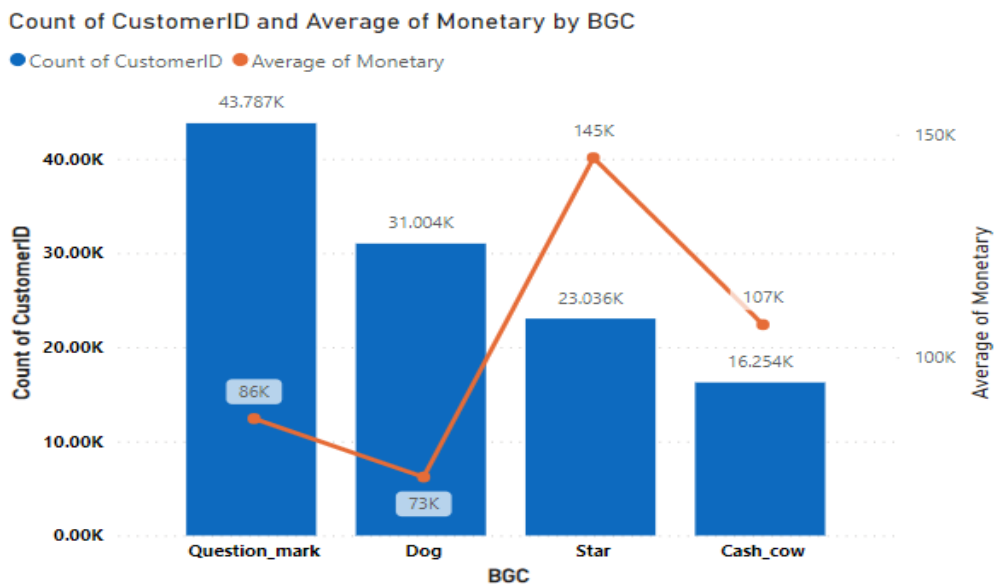


Do số lượng khách hàng tiềm năng là rất cao, lớn **gần gấp đôi** khách hàng nhóm VIP nhưng doanh thu chỉ **gấp 1.13 lần**. Mỗi nhóm này đều đóng góp **hơn 3 tỉ** vào doanh thu và đây cũng là 2 **nhóm khách hàng chủ chốt** trong việc tạo ra doanh thu cho doanh nghiệp. Mức độ đóng góp cao thứ 3 là nhóm khách vãng lai. Mặc dù điểm M của nhóm này ở mức **thấp nhất** và điểm R, F cũng **không cao**, nhưng **số lượng lại rất lớn** nên nhóm này đã tạo ra **hơn 2.2 tỉ doanh thu**, cao hơn nhóm khách than thiết – chỉ tạo ra **1.7 tỷ đồng**.

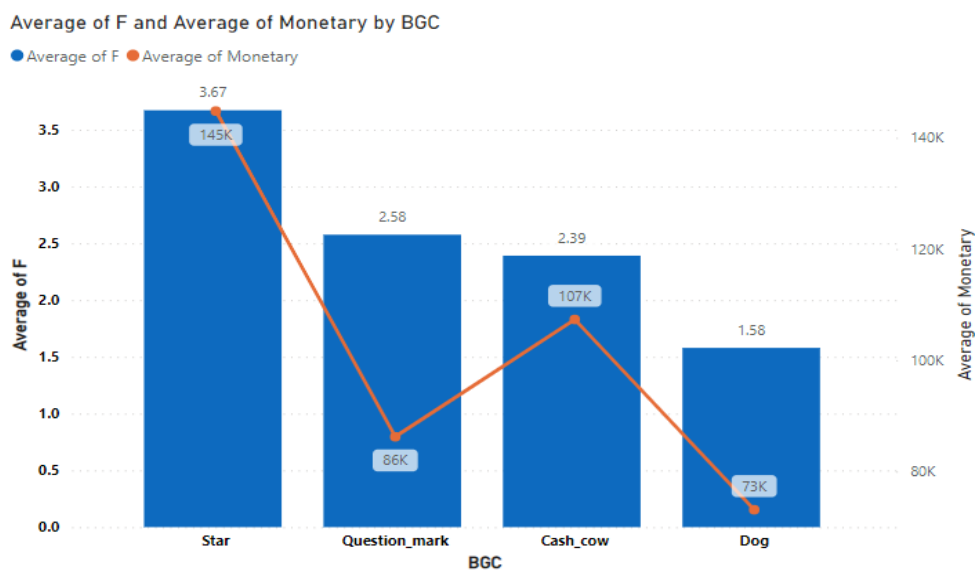


Xét về mức độ đóng góp trung bình vào tổng doanh thu, nhóm khách **VIP thường chi trả ở mức cao nhất**, trung bình mỗi khách hàng của nhóm này đóng góp gần **150K**. Tiếp đến là nhóm **khách hàng thân thiết với trung bình là hơn 100K**. Mặc dù tổng doanh thu được tạo ra của nhóm này ở mức thấp nhất nhưng do số lượng ít nhất nên mức đóng góp trung bình của nhóm này tương đối cao. Nhóm này bao gồm những khách hàng có số điểm M khá cao **và có tiềm năng rất lớn để trở thành các khách hàng VIP** và khách hàng VIP nếu được tác động một cách hợp lý.

Giá trị đóng góp trung bình cao thứ 3 là nhóm **khách tiềm năng**, khoảng **90K/người**. Cuối cùng là nhóm khách **vãng lai** với giá trị khoảng **75K**.



Biểu đồ này thể hiện **số lượng khách hàng và giá trị trung bình doanh thu** mỗi khách đem lại. Từ biểu đồ có thể thấy nhóm **khách VIP và nhóm khách thân thiết** có số lượng ít nhất nhưng **doanh thu trung bình lại cao nhất**. Các chương trình chăm sóc khách hàng đặc biệt và các chương trình tặng thưởng tri ân khách hàng nên được tập trung vào nhóm này.



Xét về cả tần suất sử dụng trung bình và doanh thu đóng góp trung bình, nhóm **khách VIP và khách thân thiết** đều có các giá trị rất tốt, vừa hoạt động sôi nổi, vừa có **giá trị các đơn hàng rất cao**. Nhóm **khách hàng tiềm năng** là nhóm khách hoạt động khá **tích cực** nhưng giá trị trung bình của doanh thu mỗi người đem lại thì **tương đối thấp**, có thể nhóm khách này thường giao dịch với các đơn hàng có giá trị thấp. Điều này cho thấy đây chính xác là nhóm khách trong tương lai sẽ giúp **công ty tăng trưởng đột phá**. Xếp cuối là nhóm khách **vãng lai**, tần suất mua hàng **thấp** cũng như giá trị các đơn hàng ở mức vô cùng **thấp**.

6. KẾT LUẬN

- **Nhóm khách hàng VIP:** là nhóm khách hàng “toàn năng” với các điểm R,F,M đều cao, doanh nghiệp cần có chính sách riêng để giữ chân những người trong nhóm này, khiến họ gắn bó lâu dài hơn.
- **Nhóm khách hàng thân thiết:** tạo ra **dòng tiền ổn định và khá cao** nhưng số lượng rất **ít**, mặc dù vậy vẫn nên có các ưu đãi nhằm cố gắng phát triển nhóm này trở thành nhóm VIP.
- **Nhóm khách hàng tiềm năng:** chiếm số lượng đông nhất, là “mỏ” **chưa khai thác tối ưu** của doanh nghiệp dù tần suất sử dụng dịch vụ tương đối nhiều. Đối mặt với sự khó tính của nhóm khách hàng này, doanh nghiệp nên đưa ra những danh mục sản phẩm mới, phù hợp nhằm **giải quyết tính “tiết kiệm”** của nhóm này.

7. PHỤ LỤC

Code SQL

```
with data1 as (
    select CustomerID ,
    datediff('2022-09-01', STR_TO_DATE(max(Purchase_Date), '%m/%d/%Y')) as
    Recency,
    count(Purchase_date) as trans_num,
    sum(GMV) as Monetary
    from customer_transaction ct
    where CustomerID != '0'
    group by CustomerID
    order by trans_num desc),
data2 as (
    select ID, case when stopdate = '' then datediff('2022-09-01',
    STR_TO_DATE(created_date, '%m/%d/%Y'))/365.25
    else datediff(STR_TO_DATE(stopdate, '%m/%d/%Y'), STR_TO_DATE(created_date,
    '%m/%d/%Y'))/365.25
    end as Contract_term
    from customer_registered cr ),
RFM_cal as (select CustomerID,
    Recency,
    trans_num/Contract_term as Frequency,
    Monetary
    from data1
    join data2
    on data1.CustomerID = data2.ID),
RFM_score as (
    select *,
    ntile(4) over(order by Recency DESC) as R,
    ntile(4) over(order by Frequency) as F,
    ntile(4) over(order by Monetary) as M
    from RFM_cal)
select *,concat(R,F,M) as RFM,
case
    when concat(R,F,M) in ('444','443','434','433','344','343','334','333') then
    'Star'
    when concat(R,F,M) in ('424','423','413','414','313','314','323','324',
    '233','234','244','243','133','134','143','144') then 'Cash_cow'
    when concat(R,F,M) in
    ('441','442','431','432','341','342','331','332','131','132','141','142',
    '241','242','231','232','113','114','123','124','213','214','223','224') then
    'Question_mark'
    else 'Dog'
    end as BGC
from RFM_score
```


7. PHỤ LỤC

Code SQL

```
-- Tính các quartiles cho từng yếu tố RFM
with a as
  (select r1.customerID, recency,
    row_number() over (order by recency desc) as rn_recency
  from rfm_score r1) ,
b as
  (select r2.customerID, Frequency,
    row_number() over (order by Frequency desc) as rn_Frequency
  from rfm_score r2) ,
c as
  (select r3.customerID, Monetary,
    row_number() over (order by Monetary desc) as rn_Monetary
  from rfm_score r3),
d as (
  select a.customerID, a.recency, a.rn_recency , b.Frequency,
    b.rn_Frequency, c.Monetary, c.rn_Monetary
  from a
  left join b on a.customerID=b.customerID
  left join c on a.customerID=c.customerID)
select min(recency) as min,
  (select recency from d where rn_recency = (select
    FLOOR(count(CustomerID)*0.75) from rfm_score)) as Q1,
  (select recency from d where rn_recency = (select
    FLOOR(count(CustomerID)*0.5) from rfm_score)) as Q2,
  (select recency from d where rn_recency = (select
    FLOOR(count(CustomerID)*0.25) from rfm_score)) as Q3,
  max(recency) as max
from d
union
select min(Frequency) as min,
  (select Frequency from d where rn_Frequency = (select
    FLOOR(count(CustomerID)*0.75) from rfm_score)) as Q1,
  (select Frequency from d where rn_Frequency = (select
    FLOOR(count(CustomerID)*0.5) from rfm_score)) as Q2,
  (select Frequency from d where rn_Frequency = (select
    FLOOR(count(CustomerID)*0.25) from rfm_score)) as Q3,
  max(Frequency) as max
from d
union
select min(Monetary) as min,
  (select Monetary from d where rn_Monetary = (select
    FLOOR(count(CustomerID)*0.75) from rfm_score)) as Q1,
  (select Monetary from d where rn_Monetary = (select
    FLOOR(count(CustomerID)*0.5) from rfm_score)) as Q2,
  (select Monetary from d where rn_Monetary = (select
    FLOOR(count(CustomerID)*0.25) from rfm_score)) as Q3,
  max(Monetary) as max
from d
```