

RFM Analysis for Customer 360 Project

Customer 360 là gì?



Trước đây, sự tương tác giữa công ty và khách hàng còn hạn chế nên yêu cầu quản lý dữ liệu đơn giản hơn. Tất cả những gì cần thiết cho công việc kinh doanh hoàn chỉnh là các yếu tố dữ liệu khách hàng cốt lõi.

Trong thời buổi hiện nay, tiến bộ công nghệ đã dẫn đến vô số cách mà các công ty kinh doanh với khách hàng. Các loại dữ liệu khác nhau, chẳng hạn như dữ liệu giao dịch (transaction data), hành vi (behavioral data), tương tác (interaction data) và nhân khẩu học (demographics data) đóng vai trò rất quan trọng để nắm bắt góc nhìn rộng hơn, sâu hơn về khách hàng.

RFM Analysis là gì?

RFM là viết tắt của Recency, Frequency, Monetary.

RFM là một phương pháp phân tích và phân đoạn khách hàng dựa trên ba yếu tố quan trọng: Recency (Tần suất gần nhất), Frequency (Tần suất mua hàng) và Monetary (Giá trị tiêu dùng). **RFM** được sử dụng để xác định giá trị của từng khách hàng và nhóm khách hàng lại với nhau dựa trên hành vi mua hàng của họ.

Mục tiêu chính của **RFM** là phân đoạn khách hàng và hiểu rõ hơn về hành vi mua hàng của họ. Bằng cách phân tích Recency, Frequency và Monetary, giúp xác định giá trị của từng khách hàng và nhóm khách hàng lại với nhau. Phân tích RFM giúp trả lời những câu hỏi:

Đâu là những khách hàng chi tiêu nhiều và trung thành nhất?

Đâu là những khách hàng có khả năng rời bỏ công ty?

Đâu là những khách hàng đã từ bỏ thương hiệu?, ...

RFM định lượng giá trị khách hàng dựa trên 3 thông tin chính:

Recency (R): khoảng thời gian tính từ lần cuối giao dịch đến thời điểm xác định để làm báo cáo. Chỉ số này cho biết liệu gần đây khách hàng có mua hàng hay đã lâu rồi chưa mua hàng.

Frecency (F): tần suất giao dịch, được tính bằng cách lấy tổng số lần mua hàng. Chỉ số này cho biết khách hàng có mua hàng thường xuyên không.

Monetary (M): tổng số tiền khách hàng đã chi trả trong suốt vòng đời mua hàng của họ. Chỉ số này cho biết khách hàng đã chi bao nhiêu cho công ty

Sau khi có 3 thông tin trên, sử dụng phương pháp chia khoảng Interquatile range. Chia điểm 3 yếu tố trên thành 4 điểm theo các khoảng:

1 điểm: min (0%) → lower quartile (25%)

2 điểm: lower quartile (25%) → median (50%)

3 điểm: median (50%) → upper quartile (75%)

4 điểm: upper quartile (75%) → max (100%)

Phân khúc khách hàng từ phân tích RFM

R, F, M có thang điểm 1-4, do đó có khoảng 64 tổ hợp RFM khác nhau.

Phân khúc khách hàng	RFM	Đặc điểm
Loyal	144, 244	Là những khách hàng mới giao dịch, mua hàng thường xuyên và chi tiêu nhiều nhất. Những khách hàng này trung thành, sẵn hàng chi tiêu và sẽ thường xuyên mua hàng trong tương lai.
Potential Loyalist	142, 141 131, 133, 132, 242, 241, 233, 223	Là những khách hàng có giao dịch gần đây, chi tiêu khá, đã mua hơn một lần và mua thường xuyên
New Customers	122, 121, 112, 111, 211	Đã mua gần đây, chi tiêu thấp, không mua hàng thường xuyên
Promising	124, 113, 114, 214, 213	Đã mua hàng gần đây, chi tiêu cao nhưng chưa thường xuyên
Need Attention	143, 134, 243, 234, 224	Chưa mua hàng gần đây, chi tiêu khá, đã từng mua hàng thường xuyên
About To Sleep	231, 221, 212, 321, 313	Đã lâu chưa mua hàng, chi tiêu thấp, không mua hàng thường xuyên
At Risk	344, 343, 342, 334, 324, 442, 434, 433, 424	Rất lâu không quay lại mua hàng, chi tiêu khá, đã từng mua hàng thường xuyên

Cannot Lose Them	444, 314, 414, 313, 413	Rất lâu không quay lại, chi tiêu cao, đã từng mua hàng rất thường xuyên
Hibernating customers	232, 222, 331, 341, 333, 332, 323, 322, 432, 423, 422, 312, 311	Khá lâu chưa mua lại, chi tiêu thấp, mua hàng không thường xuyên
Lost customers	411, 412, 421, 431, 441	Rất lâu không quay lại, chi tiêu thấp, mua hàng chỉ 1 vài lần duy nhất

Dataset

Customer_Registered

Column name	Data type	Length	Note
ID	bigint	8	Example: 1, 2, 3, ..., 123321123
Contract	varchar	100	Mã hợp đồng
LocationID	int	4	Mã vị trí
BranchCode	tinyint	1	Mã chi nhánh
Status	tinyint	1	0 / 1 / 2 / 3
created_date	datetime	8	Ngày đăng ký
stop_date	datetime	8	Ngày hủy

Customer_Transaction

Column name	Data type	Length	Note
ID	bigint	8	TransactionID
CustomerID	varchar	200	Example: 1, 2, 3, ..., 123321123
Purchase_Date	datetime	8	Ngày giao dịch
GMV	bigint	8	Gross Monetary Value

Code

Bước 1: Tính giá trị Recency, Frecency, Monetary

```
select max(purchase_date) from Customer_Transaction
Output: 2022-08-31 00:00:00.000
```

-> Chọn ngày 09/01/2022 làm báo cáo

```
SELECT CustomerID,
DATEDIFF(DAY , MAX(Purchase_Date), '2022/09/01') as Recency,
round((CAST (COUNT(ID) as decimal (4,2)) / cast(DATEDIFF(DAY
,MIN(Purchase_Date), '2022/09/01') as decimal(4,2))),2) AS Frequency,
SUM(GMV) AS Monetary
INTO #rfmdata
FROM Customer_Transaction
GROUP BY CustomerID
```

CustomerID	Recency	Frequency	Monetary
-----+	-----+	-----+	-----+
0		1 0.9200000	400000
1000000		31 0.0300000	75000
1000004		92 0.0100000	95000
1000009		31 0.0300000	75000
1000036		31 0.0300000	150000
1000047		62 0.0200000	115000
1000049		62 0.0200000	95000
1000051		62 0.0200000	75000
1000056		92 0.0100000	95000
1000057		62 0.0200000	85000

942339 records

Bước 2: Chấm điểm các chỉ số Recency, Frequency, Monetary dựa trên IQR method

```

SELECT CustomerID,
Recency,
Frequency,
Monetary,
NTILE(4) OVER (ORDER BY Recency) AS R_Score,
NTILE(4) OVER (ORDER BY Frequency) AS F_Score,
NTILE(4) OVER (ORDER BY Monetary) AS M_Score
INTO #rfmscore
FROM #rfmdata

```

CustomerID	Recency	Frequency	Monetary	R_Score	F_Score	M_Score
-----+	-----+	-----+	-----+	-----+	-----+	-----+
3176185		67 0.0100000	0	3	1	1
3906061		67 0.0100000	0	3	1	1
4151960		67 0.0100000	0	3	1	1
525549		67 0.0100000	0	3	1	1

```

3628834 | 68|0.0100000| 0| 3| 1| 1|
3874110 | 68|0.0100000| 0| 3| 1| 1|
3519142 | 68|0.0100000| 0| 3| 1| 1|
4072628 | 70|0.0100000| 0| 3| 1| 1|
178884 | 67|0.0100000| 0| 3| 1| 1|
2381022 | 67|0.0100000| 0| 3| 1| 1|
942339 records

```

Bước 3: Gộp 3 chỉ số R, F, M thành RFM, sau đó kiểm tra tổ hợp nào chiếm đa số

```

SELECT CustomerID,
concat(R_Score, F_Score, M_Score) as RFM_Score
into #segmentation
from #rfmscore

```

```

CustomerID|RFM_Score|

```

```

-----+-----+

```

```

3176185 |311 |

```

```

3906061 |311 |

```

```

4151960 |311 |

```

```

525549 |311 |

```

```

3628834 |311 |

```

```

3874110 |311 |

```

```

3519142 |311 |

```

```

4072628 |311 |

```

```

178884 |311 |

```

```

2381022 |311 |

```

```

942339 records

```

Đó là các bước để thực hiện, ta có thể đúc kết trong 1 câu lệnh:

```

SELECT *,
concat(R_Score, F_Score, M_Score) as RFM_Score
from (SELECT CustomerID,

```

```

Recency,
Frequency,
Monetary,
NTILE(4) OVER (ORDER BY Recency) AS R_Score,
NTILE(4) OVER (ORDER BY Frequency) AS F_Score,
NTILE(4) OVER (ORDER BY Monetary) AS M_Score
FROM (SELECT CustomerID,
DATEDIFF(DAY , MAX(Purchase_Date), '2022/09/01') as Recency,
round((CAST (COUNT(ID) as decimal (4,2)) / cast(DATEDIFF(DAY
, MIN(Purchase_Date), '2022/09/01') as decimal(4,2))), 2) AS Frequency,
SUM(GMV) AS Monetary
FROM Customer_Transaction
GROUP BY CustomerID ) a)a1

```

Kết quả: [rfmoutput](#)

Bước 4: RFM Analysis:

Import Segmentation vào database

```

select CustomerID,
Recency,
Frequency ,
Monetary ,
R_Score ,
F_Score ,
M_Score ,
RFM_Score,
segment
from segmentationname s right join rfmoutput r
on s.Scores = r.RFM_Score
order by CustomerID

```

CustomerID	Recency	Frequency	Monetary	R_Score	F_Score	M_Score	RFM_Score	segmentation
0	1	0.92	400000	1	4	4	144	Loyal
71739	62	0.02	105000	2	2	4	224	Need Attention
72014	62	0.02	254091	2	2	4	224	Need Attention

Output: final1

Kiểm tra xem tổ hợp nào chiếm đa số:

```

select segment,
count(CustomerID) as Member_Amount,

```

```

count(CustomerID)/(select count(customerID)
                    from (select CustomerID,
                                Recency,
                                Frequency ,
                                Monetary ,
                                R_Score ,
                                F_Score ,
                                M_Score ,
                                RFM_Score,
                                segment
                        from segmentationname s right join rfmoutput
r
                        on s.Scores = r.RFM_Score
                        order by CustomerID ) hmm )*100 as
Percentage
from
    (select CustomerID,
    Recency,
    Frequency ,
    Monetary ,
    R_Score ,
    F_Score ,
    M_Score ,
    RFM_Score,
    segment
    from segmentationname s right join rfmoutput r
    on s.Scores = r.RFM_Score
    order by CustomerID ) hmm
group by segment
order by Member_Amount desc

```

segmentation	Member_Amount	Percentage
--------------	---------------	------------

-----+-----+-----+

Hibernating Customers	267718	28.0497
-----------------------	--------	---------

Potential Loyalist	141178	14.7917
--------------------	--------	---------

About To Sleep	118186	12.3827
----------------	--------	---------

Cannot Lose Them	115906	12.1439
------------------	--------	---------

Need Attention	102530	10.7424
----------------	--------	---------

Loyal	76138	7.9772
-------	-------	--------

Lost Customers	68871	7.2158
----------------	-------	--------

At Risk	63914	6.6965
---------	-------	--------

Output: final2

Mô tả dữ liệu Recency, Frequency, Monetary

index	Recency	Frequency	Monetary
count	942339.0	942339.0	942339.0
mean	60.36328115465878	0.021835326777306263	87826.7821282999
std	26.244773943660924	0.029325968690441455	37153.64356435095
min	1.0	0.01	0.0
25%	31.0	0.01	70000.0
50%	62.0	0.02	75000.0
75%	92.0	0.03	95000.0
max	92.0	2.0	8400000.0

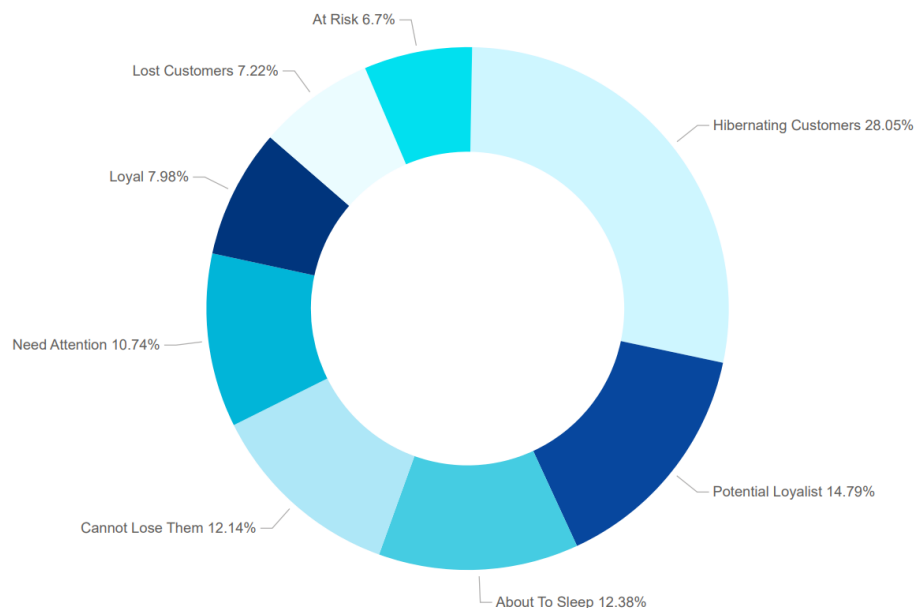
Visualization

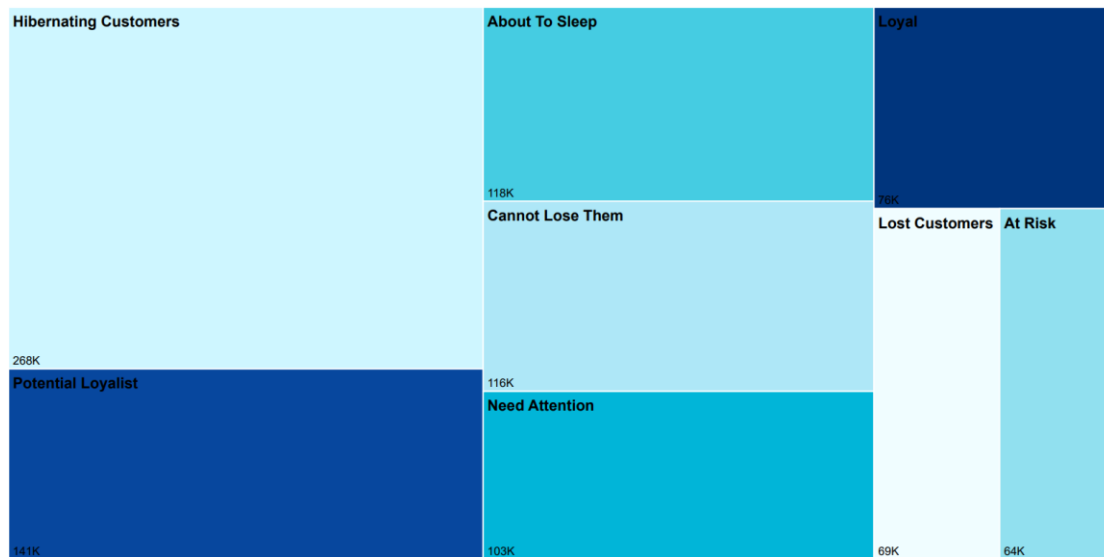
Biểu đồ Treemap & Donut thể hiện tỷ trọng các phân khúc khách hàng.

Mức độ quan trọng tăng dần từ nhạt đến đậm

Trong đó nhóm Hibernating chiếm cao nhất (28,05%). Do đó cần có những chiến dịch để kích thích nhóm này mua hàng và chi tiêu nhiều hơn.

Hai tập khách hàng quan trọng nhất của doanh nghiệp chiếm hơn 22%, đây cũng là một chỉ số tốt. Trong đó tập KH Loyal có 76,138 khách hàng, và Potential Loyalist có 141,176 khách hàng.





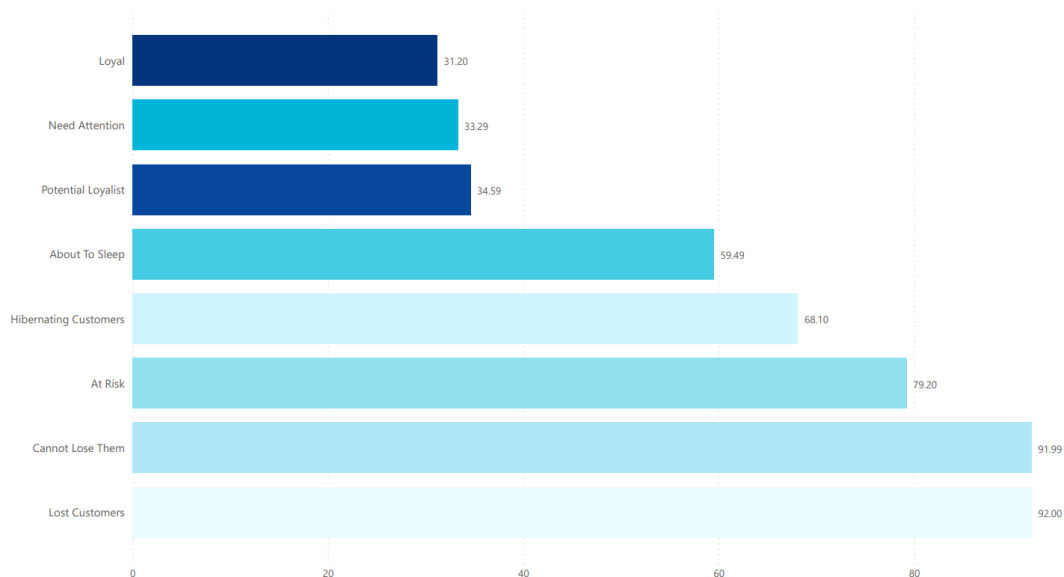
Phân tích cụ thể theo các yếu tố R, F, M

Mặc dù những phân khúc khách hàng có đặc điểm riêng, nhưng mỗi doanh nghiệp lại có số lượng khách hàng mỗi phân khúc khác nhau cho nên những yếu tố về Recency, Frequency, Monetary của tổng thể doanh nghiệp sẽ khác nhau. Do đó cần đánh giá, sự tương quan giữa 2 yếu tố cơ cấu của số khách hàng trong mỗi phân khúc và giá trị mà mỗi nhóm đem lại cho doanh nghiệp.

Màu sắc về mức độ của phân khúc được đậm dần từ Hibernating -> Loyal

Số ngày mua hàng gần nhất theo phân khúc khách hàng

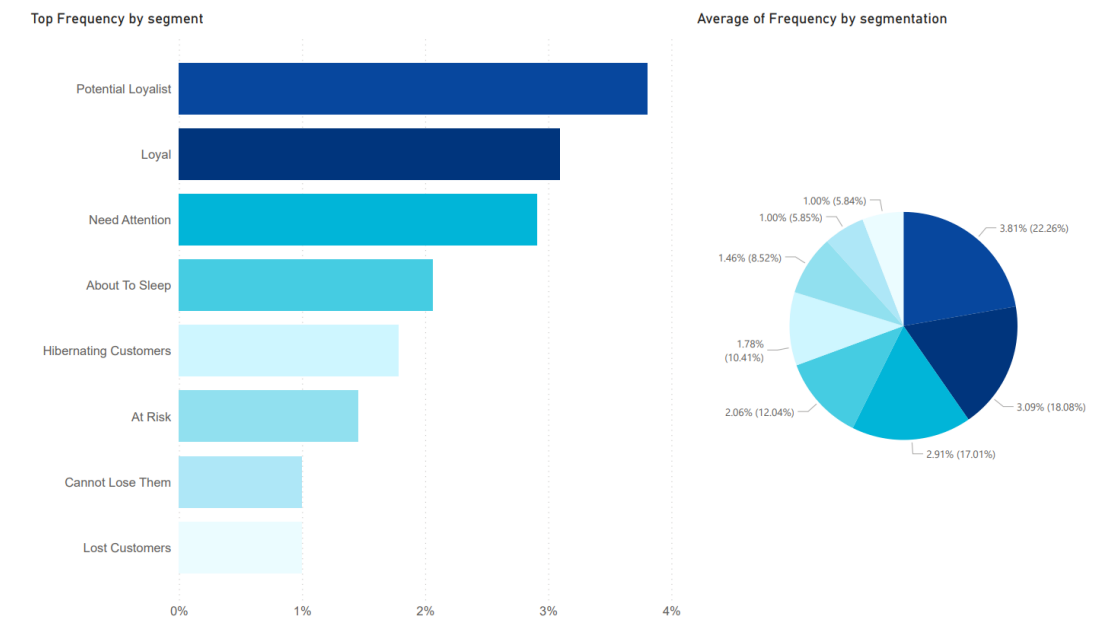
Top Recency by segmentation



Phân khúc mua hàng gần đây nhiều nhất là Loyal, New Attention và Potential Loyalist. Những KH ở các phân khúc này có đơn hàng gần nhất trung bình khoảng 30 ngày.

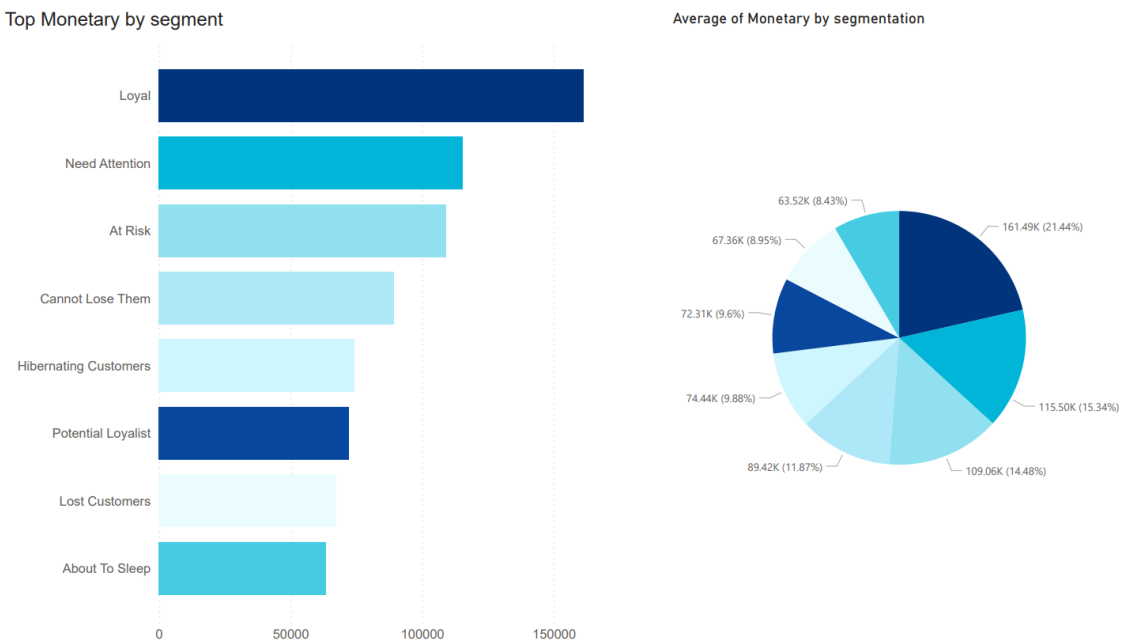
Nhóm Cannot Lose Them có số ngày mua hàng gần nhất ~ Lost Customer. Nên có những ưu đãi, cải tiến SP để kích thích nhóm này mua lại vì họ sắp có nguy cơ thành nhóm Lost Customer.

Tần suất mua hàng theo phân khúc khách hàng



Nhóm Potential Loyalist đứng vị trí thứ nhất, họ là những người mua hàng rất thường xuyên. Kết hợp với Monetary, có thể nhận định nhóm Loyal và Need Attention là những khách hàng đáng giá , có thể là những ưu đãi hoàn tiền, tặng voucher để kích thích họ mua thường xuyên hơn vì GMV họ đem lại cho doanh nghiệp rất cao

Doanh thu theo phân khúc khách hàng

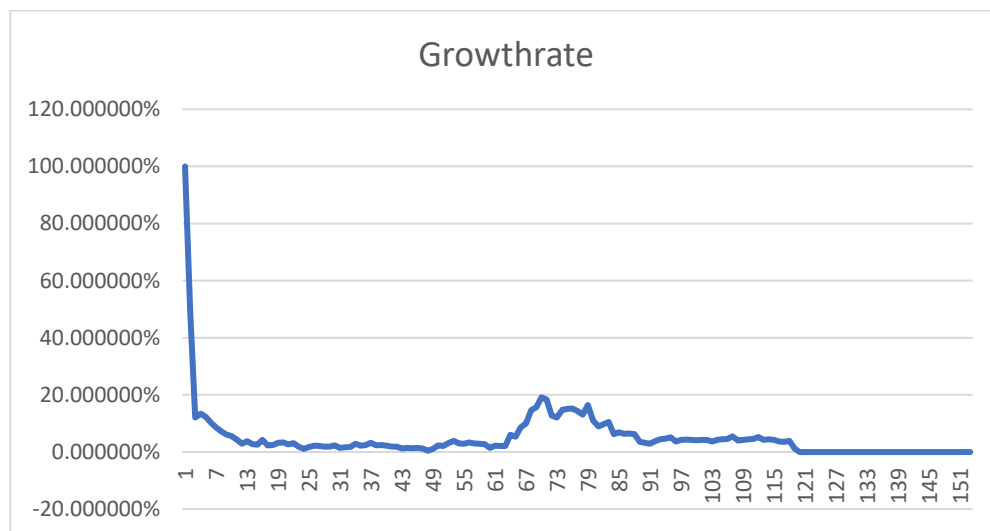


Customer Growth Rate for Business Intelligence Reporting

Sử dụng dữ liệu truy vấn từ Customer_Registered

```
SELECT  
FORMAT (created_date, 'yyyy/MM') AS Date,  
LocationID,  
BranchCode,  
COUNT(DISTINCT CASE WHEN stopdate IS NULL THEN ID END) AS  
new_customers,  
COUNT(DISTINCT CASE WHEN stopdate IS NOT NULL THEN ID END) AS  
lost_customers  
FROM Customer_Registered cr  
GROUP BY FORMAT (created_date, 'yyyy/MM'), LocationID, BranchCode
```

Output: <https://drive.google.com/file/d/1fkSxTWuCI7ztslMGnYnRziskRqk4plZ/view?usp=sharing>



Customer Retention Rate

$$\frac{\begin{array}{c} \text{\# of Customers} \\ \text{at End of} \\ \text{Period} \end{array} - \begin{array}{c} \text{\# of New} \\ \text{Customers Added} \\ \text{During Period} \end{array}}{\begin{array}{c} \text{\# of Customers at Beginning of Period} \end{array}} \times 100 = \text{Retention Rate}$$

$$R = (\text{End-New})/\text{Begin}$$

Dựa trên tập dữ liệu, xác định period để tính tỉ lệ này là 5 năm (2016-2021).

Trong data:

Date: 2015/12 – 2021/08

End (2016/01–2021/08) = End (2009/03-2015/12) + New (2016/01–2021/08) – Lose (2016/01–2021/08)

New (2016/01–2021/08)

Begin (2016/01–2021/08) = End (2009/03-2015/12)

Kết quả:

	Nem Customer	Customer Lose	Number of Customers	Retentionrate
2009/03 - 2015/12	184070	11	184059	-
2016/01-2021/8	864448	42	1048465	99.9772%

Trong 5 năm được thống kê có 864,448 khách hàng mới, đạt mốc khách hàng 1tr

Retention rate của doanh nghiệp trong 5 năm là 99,9772%

Hic em tính làm thêm mà bận quá