CUSTOMER RFM ANALYTICAL REPORT

Viet Tuan Dinh

1.Project Overview

The project aims to analyze data utilizing MySQL Server to transform data from Online Transactional Processing (OLTP) to Online Analytical Processing (OLAP). Subsequently, it focuses on constructing an RFM (Recency, Frequency, Monetary) model to categorize customers into groups using the Customer 360 approach and RFM model.

1.1 What is Customer 360?

Customer 360 is a business management and customer analysis strategy in which all information about each customer is gathered into a common platform. This platform contains comprehensive data about various aspects of the customer, from demographic details to communication methods to the customer's relationships with the business. Consolidating customer information and data into a single source helps companies gain a holistic, multidimensional view of their customer base. By understanding customers better, companies can develop specific strategies tailored to each customer segment, enhance customer experience, and improve the competitiveness of the business in the market.

BENEFITS OF CUSTOMER360 PLATFORM

1. **Demographic:** This entails personal information about customers such as age, gender, address, occupation, education, family, etc. Demographic data helps businesses better understand their customer base and categorize them into different target groups.
2. **Behavioural:** This includes information about customer behaviour such as purchasing habits, frequency of purchases, preferred products or services, and time and manner of interaction with the business. Behavioural data helps companies gain a deeper understanding of customer needs and preferences, enabling the creation of effective interaction and marketing strategies.
3. **Transaction:** This refers to information about customer purchase transactions or financial transactions, such as purchase history, quantity of purchases, transaction value, payment methods, etc. Transaction data helps businesses assess performance, predict purchasing trends, and optimize sales strategies.
4. **Interaction:** This encompasses information about interactions between customers and the business through channels such as phone, email, website, social media, etc. Interaction data helps companies to understand how customers interact and respond to marketing strategies and services, thereby improving the customer experience when using the business's products or services.
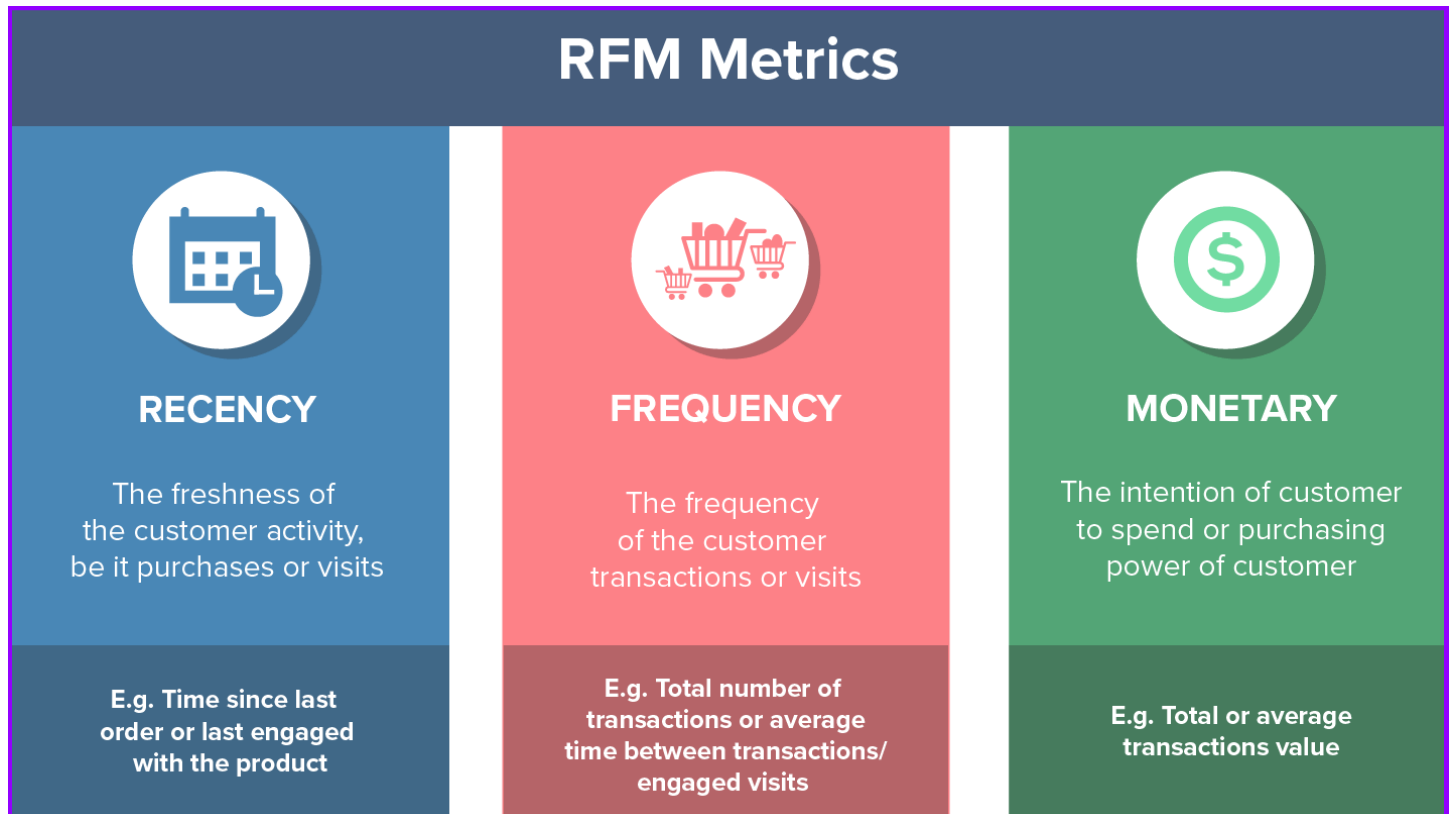
1.2 RFM Model

RFM is a customer segmentation model based on the Customer360 philosophy. Data on the transaction history between customers and businesses is collected and analyzed based on three

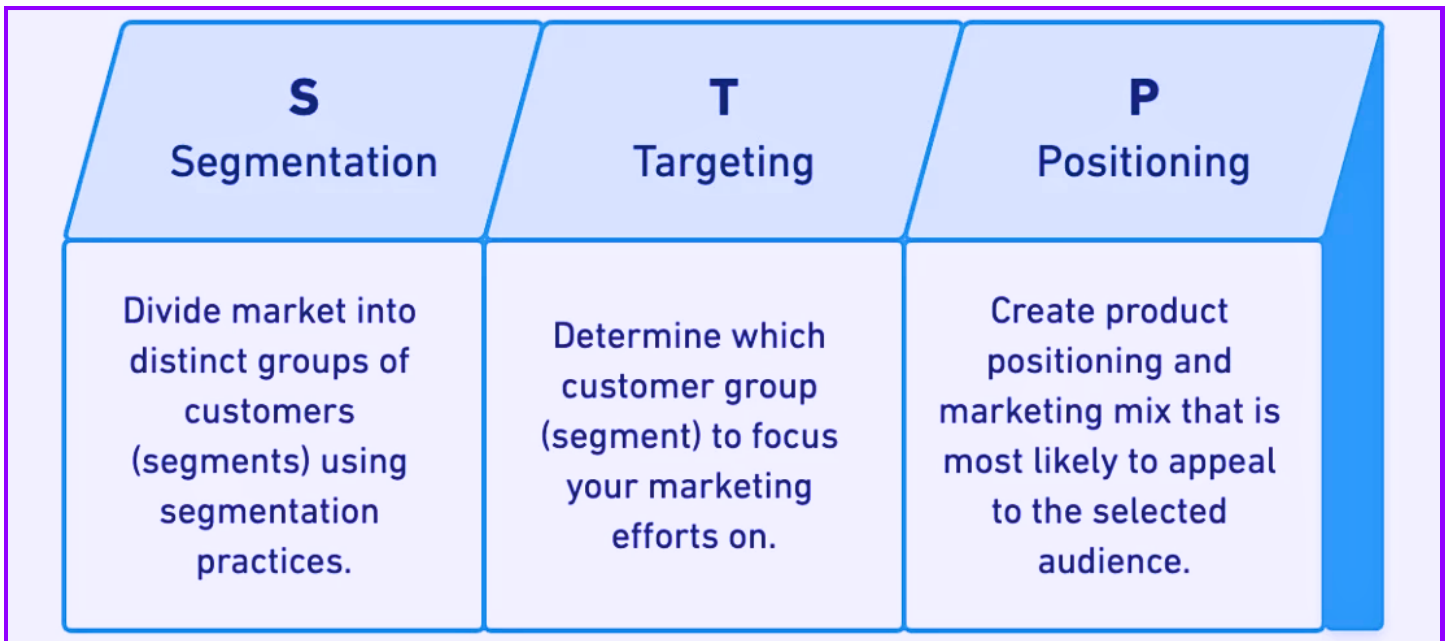key factors: Recency, Frequency, and Monetary.

RFM Metrics

- Recency: The most recent time a customer purchased or used a service.
- Frequency: How often a customer makes purchases or uses services.
- Monetary: The amount of money a customer has spent on purchases or services.

Recency, Frequency, and Monetary are three crucial factors in quantifying customer behaviour and interactions with the business. Among these, Frequency and Monetary are determinants of the Customer Lifetime Value, while Recency influences Customer Retention.

Benefits of RFM Metris

The RFM model helps businesses enhance customer retention by aiding them in understanding fundamental marketing principles:

S Segmentation	T Targeting	P Positioning
Divide market into distinct groups of customers (segments) using segmentation practices.	Determine which customer group (segment) to focus your marketing efforts on.	Create product positioning and marketing mix that is most likely to appeal to the selected audience.

The RFM model allows businesses to gain crucial insights into customers that they can act upon, thereby shaping business strategies around these insights. This model enables companies to understand the significance of their brand to current customers, assisting them in managing customer perceptions and converting positive emotions into purchasing actions.

RFM scoring utilizes quartiles in statistical probability.

Quartiles in statistical probability involve dividing data into four equal parts based on their order. Specifically, quartiles divide the data into four sections with the same number of observations, meaning each section contains 25% of the total observations.

Quartiles consist of three values: the first quartile (Q1), the second quartile (Q2), and the third quartile (Q3). These three values divide a dataset (sorted in ascending order) into four parts with equal observations.

Assuming the dataset has n observations:

- The first quartile is calculated using $Q1 = 25 * (n+1) / 100$.
- The second quartile, the median, is $Q2 = (n+1)/2$.
- The third quartile is calculated using the formula $Q3 = 75 * (n+1) / 100$.
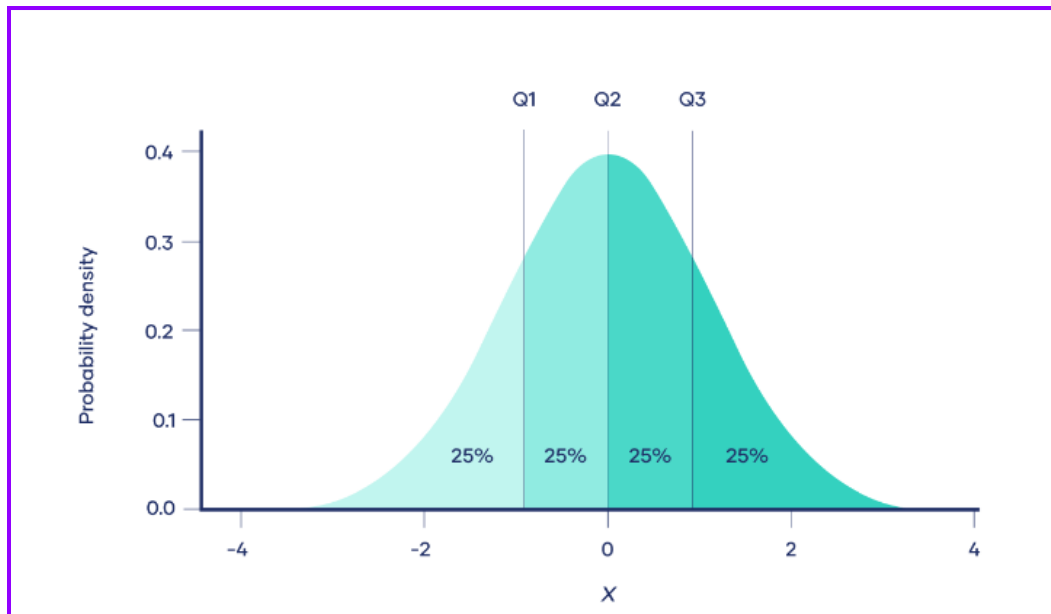
For each variable (R, F, M), we categorize into groups from 1 to 4, with group 4 representing the highest values and group 1 representing the lowest values. Specifically for R, a score of 4 indicates recently purchased, while 1 signifies a long absence of transactions with the business.
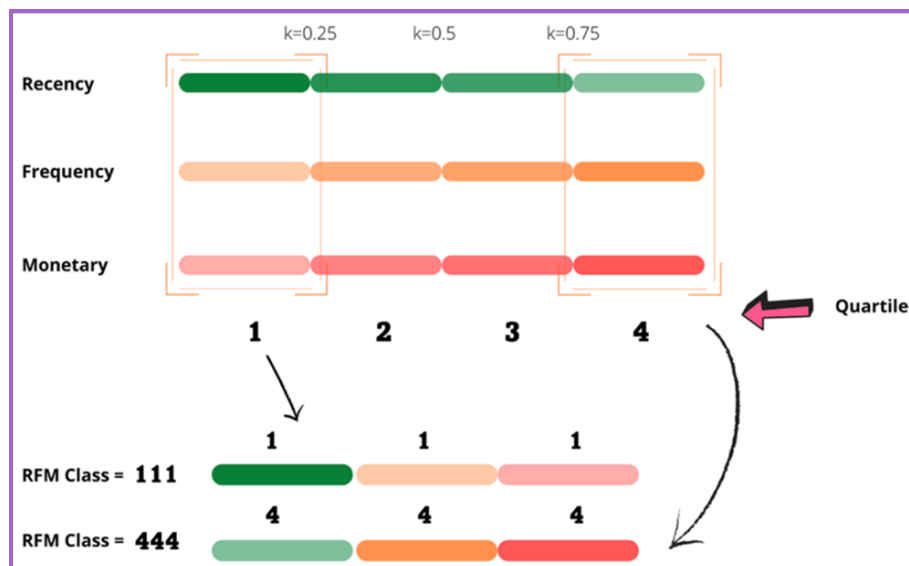
Customer Segmentation Based on RFM Score

Each customer will be assessed based on three key factors: Recency (R), Frequency (F), and Monetary Value (M). Each factor is evaluated compared to the entire customer base and is assigned a score ranging from 1 to 4. A score of 1 indicates the lowest level, while a score of 4 represents the highest level relative to the overall customer base. These individual scores are subsequently amalgamated into a unified RFM index, which serves as the basis for customer classification.

Customer Segment	Activity	RFM group
Champions	Bought recently, buy often and spend the most!	444, 443, 434, 344
Loyal Customers	Spend good money with us often. Responsive to promotions.	442, 441, 432, 431, 433, 343, 342, 341
Potential Loyalist	Recent customers, but spent a good amount and bought more than once.	424, 423, 324, 323, 413, 414, 343, 334
Recent Customers	Bought most recently, but not often.	333, 332, 331, 313
Promising Customers	Recent shoppers, but haven't spent much.	422, 421, 412, 411, 311, 321, 312, 322
Customers Needing Attention	Above average recency, frequency and monetary values. May not have bought very recently though.	333, 332, 331, 313
About To Sleep / New Customer	Below average recency, frequency and monetary values. Will lose them if not reactivated.	422, 421, 412, 411, 311, 321, 312, 322
At Risk / Price Sensitive	Spent big money and purchased often. But long time ago.	131, 132, 141, 142, 231, 232, 241, 242
Can't Lose Them	Made biggest purchases, and often. But haven't returned for a long time.	244, 234, 243, 233, 224, 214, 213, 134, 144, 143, 133
Hibernating	Last purchase was long back. Low spenders and low number of orders.	223, 221, 222, 211, 212, 124
Lost	Lowest recency, frequency and monetary scores.	111, 112, 113, 114, 121, 122, 123

1.3 How to do RFM Analysis?

Steps to Implement RFM Analysis and Segmentation

01 Gather and Analyze Customer Data

02 Calculate RFM Scores

03 Segment Customers Based on RFM Scores

04 Tailor Marketing Strategies for Each Segment

2.RFM ANALYSIS

2.1 Data Overview:

The data is compiled from 113773 customers, with 1040158 transactions, purchase data spanning from June 1, 2022, to August 31, 2022, stored in 2 tables:

- **Customer_Registered Table:**

Column name	Data type	Meaning
ID	Bigint	Index of customer
Contract	Varchar	ID of Contract
LocationID	Int	ID of Location
BranchCode	Tinyint	ID of Branch
Status	Tinyint	Example: 0 - 1 - 2
Created_date	Datetime	Register Date
Stop_date	Datetime	Cancel Date

- **Customer_Transaction table:**

Column name	Data type	Meaning
ID	Bigint	Index of customer
CustomerID	Varchar	ID of customer
Purcharse_Date	Datetime	Date of Transaction
GMV	Bigint	Revenue of Transaction

2.2 Data Cleansing

In this phase, we conduct data cleansing to eliminate potential noise:

• Omit customers with ID=0, indicating test accounts in the system.

• Exclude records lacking information on created_date or with empty created_date fields.

• Remove customers with stop_date != NULL, indicating discontinued service usage.

• Filter out transactions without revenue generation, i.e., GMV equals 0.

2.3 RFM Segmentation

To segment RFM effectively, we follow these steps:

1. Step 1: Calculate each customer's R, F, and M values.
   - R: Determine the recency of the last purchase, measured as the days between the current date and the customer's most recent purchase.
   - F: Assess the frequency of purchases, calculated as the total number of orders the customer has made over the contract age. Contract age is computed from the registration date to the current date, ensuring an unbiased assessment of usage time across different customers.
   - M: Evaluate the monetary value spent by the customer on purchasing products or utilizing business services. This metric is also computed based on the customer's contract age.
2. Step 2: Assign scores to R, F, M using quartiles.
3. Step 3: Determine RFM segments by amalgamating R, F, M scores.

After using IQR(Interquartile Range) Method for Calculation RFM score, each customer's RFM score is categorized as follows:
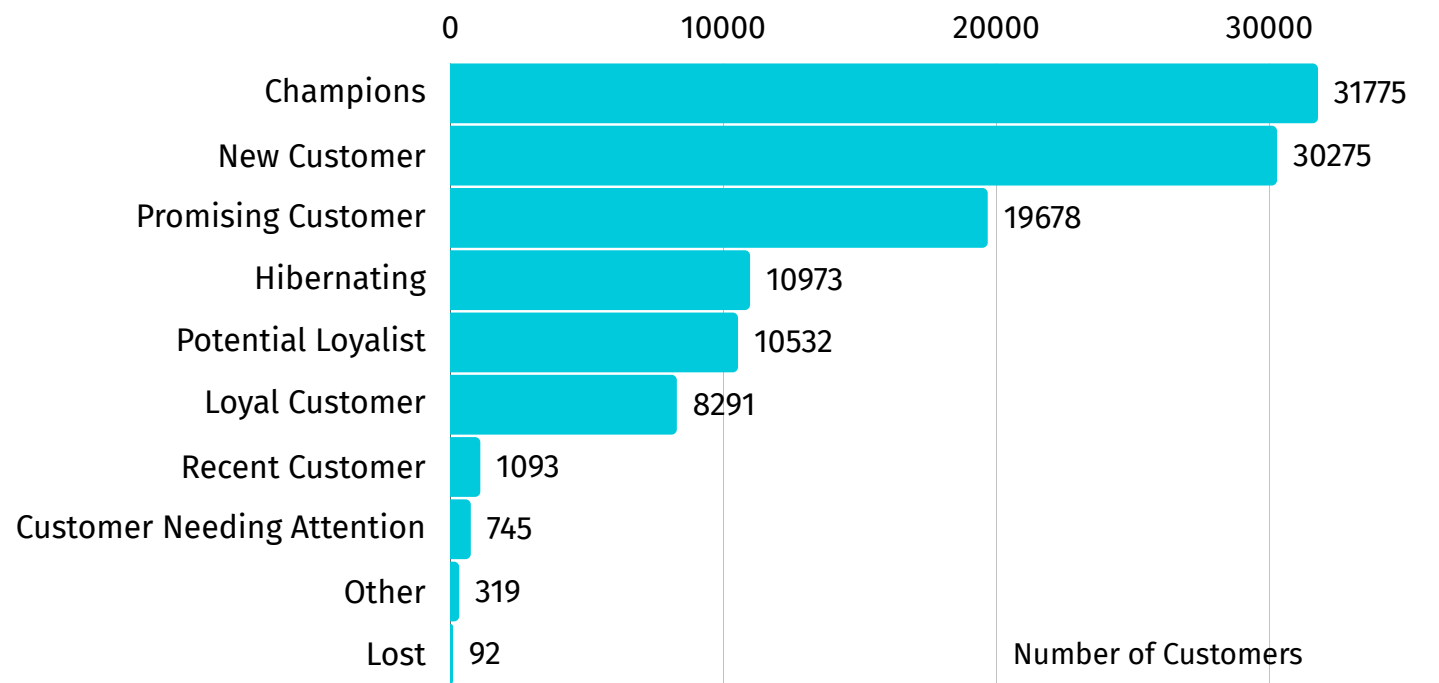
Score (IQR)	1	2	3	4
Recency	>= 92 days	62 to 91 days	32 to 61 days	<= 31 days
Frequency(per year) (For only GMV >0)	0.00 to 0.13	0.14 to 0.19	0.20 to 0.24	0.25 to 2.00
Monetary(per year)	=< 15833 VND	15833 to 18750 VND	18750 to 23750VND	>= 23750 VND

2.4 VISUALIZATION AND ANALYTICS

Utilize **PowerBI** to present dashboards and analyze the obtained results. Details regarding the analysis outcomes are elaborated in the subsequent section.

Segments Distribution

Segment	Number of Customers
Champions	31775
New Customer	30275
Promising Customer	19678
Hibernating	10973
Potential Loyalist	10532
Loyal Customer	8291
Recent Customer	1093
Customer Needing Attention	745
Other	319
Lost	92

2.5 Actionable Tips:

Customer Segment	Actionable Tip
Champions	Reward them. They can be early adopters for new products. They will promote your brand.
Loyal Customer	Upsell them higher-value products. Ask them for reviews. Increase engagement with them.
Potentail Loyalist	Offer loyalty/membership programs, introduce other products.
Recent Customer	Provide onboarding support, create early success for them, start building a relationship.
Promising Customer	Create brand awareness, offer free trials.
Customers Needing Attention	Create time-limited offers, recommend based on previous purchases. Re-engage them.
Hibernating	Share valuable resources, introduce popular products/renewals at a discounted price, reconnect with them.
At Risk	Send personal emails to reconnect, offer renewal, provide helpful resources.
Can't lose them	Win them back with renewal or newer products, don't let them become churned, have a conversation with them.
Hibernating	Offer related products and special discounts. Re-establish brand value.
Lost	Rekindle interest with outreach campaigns, ignore if not.

2.6 Comments:

Customer Distribution Analysis:

1. The largest customer segment is "Champions", with 31,775 individuals, indicating a robust base of brand advocates.
2. Conversely, the "Lost" segment comprises only 92 individuals, suggesting a relatively minor loss in terms of customer base.
3. "Hibernating" customers, totalling 10,973, represent a significant portion of the customer base, necessitating targeted re-engagement efforts.

Customer Engagement and Potential Value:

1. The "Champions" segment presents an opportunity to leverage their influence for new product launches and brand promotion.
2. "Loyal Customers" (8,291) and "Potential Loyalists" (10,532) are valuable segments likely to respond positively to upselling and loyalty programs, respectively.

Risk Mitigation Strategies:

1. For "Customers Needing Attention" (745), time-limited offers and personalized recommendations based on previous purchases could help re-engage them.
2. "At Risk" customers require urgent attention, necessitating personalized outreach and renewal offers to prevent churn.

Resource Allocation and Prioritization:

1. Allocating resources towards onboarding support for "Recent Customers" (1,093) can enhance early success and foster long-term relationships.
2. "Promising Customers" (19,678) represent a substantial opportunity for brand awareness initiatives and free trials to convert them into loyal patrons.

Re-engagement Strategies:

1. "Hibernating" customers may respond well to discounts and the reintroduction of popular products to reignite their interest.
2. Although small, the "Lost" segment still presents an opportunity for rekindling interest through targeted outreach campaigns.

Segment Specific Campaigns:

1. Tailoring marketing campaigns and offers based on customer segments can lead to higher conversion rates and improved customer satisfaction.
2. Implementing customer feedback loops and monitoring engagement metrics can help refine strategies over time for each segment.

In summary, a nuanced approach that considers each customer segment's unique characteristics and behaviors is essential for maximizing retention, driving growth, and fostering lasting customer relationships.

Thank you!

Appendix:

-- CLEAN DATA

-- 1. Omit customers with ID=0

```sql
DELETE FROM [dbo].[Customer_Transaction]

WHERE CustomerID = 0;

DELETE FROM [dbo].[Customer_Registered]

WHERE ID = 0;


-- 2. Exclude records lacking information on created_date or with empty created_date fields.

DELETE FROM [dbo].[Customer_Registered]

WHERE created_date IS NULL OR created_date = '';


-- 3. Remove customers with stop_date != NULL, indicating discontinued service usage.

DELETE FROM [dbo].[Customer_Registered]

WHERE stopdate IS NOT NULL;


-- 4. Filter out transactions without revenue generation, i.e., GMV equals 0.

DELETE FROM [dbo].[Customer_Transaction]

WHERE gmv = 0;



-- ANALYTICS

-- OVERVIEW

-- 1. CUSTOMER AND TRANSACTION:

-- Number of customers

SELECT COUNT(DISTINCT CustomerID) AS NumberOfCustomers

FROM [dbo].[Customer_Transaction];


-- Number of transactions

SELECT COUNT(*) AS NumberOfTransactions
```

```sql
FROM [dbo].[Customer_Transaction];


-- Average number of transactions per customer
SELECT
    MIN(NumberOfTransactions) AS MinTransactions,
    MAX(NumberOfTransactions) AS MaxTransactions,
    AVG(NumberOfTransactions) AS AvgTransactions
FROM (
    SELECT
        CustomerID,
        COUNT(*) AS NumberOfTransactions
    FROM [dbo].[Customer_Transaction]
    GROUP BY CustomerID
) AS TransactionCounts;


-- TRANSACTION MONTHLY
SELECT
    YEAR(Purchase_Date) AS Year,
    MONTH(Purchase_Date) AS Month,
    COUNT(Transaction_ID) AS NumberOfTransactions
FROM
    [dbo].[Customer_Transaction]
GROUP BY
    YEAR(Purchase_Date),
    MONTH(Purchase_Date)
ORDER BY
    YEAR(Purchase_Date),
```

```sql
    MONTH(Purchase_Date);


-- GMV

-- Calculate the average GMV (Gross Merchandise Volume)

SELECT AVG(CAST(GMV AS BIGINT)) AS AverageGMV

FROM [dbo].[Customer_Transaction];


-- TIME

-- Select first date and last date

SELECT

    MIN(Purchase_Date) AS min_Date,

    MAX(Purchase_Date) AS max_date

FROM

    [dbo].[Customer_Transaction];



                        -- RFM MODELS


-- Drop temporary tables if they exist to avoid conflicts

IF OBJECT_ID('tempdb..#calculation') IS NOT NULL

    DROP TABLE #calculation;

IF OBJECT_ID('tempdb..#result') IS NOT NULL

    DROP TABLE #result;

IF OBJECT_ID('tempdb..#customer_segmentation') IS NOT NULL

    DROP TABLE #customer_segmentation;


-- Perform calculations and store intermediate results
```

```sql
SELECT
    CustomerID,
    DATEDIFF(DAY, MAX(CAST(Purchase_Date AS DATE)), '2022-09-01') AS recency,
    ROUND(
        CAST(COUNT(DISTINCT CAST(Purchase_Date AS DATE)) AS FLOAT) /
        CAST(DATEDIFF(YEAR, CAST(created_date AS DATE), '2022-09-01') AS FLOAT), 2
    ) AS frequency,
    SUM(gmv) / DATEDIFF(YEAR, CAST(created_date AS DATE), '2022-09-01') AS monetary,
    ROW_NUMBER() OVER (ORDER BY DATEDIFF(DAY, MAX(CAST(Purchase_Date AS DATE)), '2022-09-01')) AS rn_recency,
    ROW_NUMBER() OVER (ORDER BY ROUND(
                            CAST(COUNT(DISTINCT CAST(Purchase_Date AS DATE)) AS FLOAT) /
                            CAST(DATEDIFF(YEAR, CAST(created_date AS DATE), '2022-09-01') AS FLOAT), 2
                        )) AS rn_frequency,
    ROW_NUMBER() OVER (ORDER BY SUM(gmv)) AS rn_monetary
INTO #calculation
FROM
    [dbo].[Customer_Transaction] T
JOIN
    [dbo].[Customer_Registered] R ON T.CustomerID = R.ID
GROUP BY
    CustomerID, created_date;


-- Check count of records in the calculation table
SELECT
    COUNT(*)
FROM
```

```sql
    #calculation;

-- Perform RFM grouping and mapping

SELECT

    *,

    CASE

        WHEN recency < (SELECT recency FROM #calculation WHERE rn_recency = (SELECT
CAST(COUNT(DISTINCT customerid) * 0.25 AS INT) FROM #calculation) AND recency >= (SELECT
recency FROM #calculation WHERE rn_recency = 1)) THEN '1'

        WHEN recency >= (SELECT recency FROM #calculation WHERE rn_recency = (SELECT
CAST(COUNT(DISTINCT customerid) * 0.25 AS INT) FROM #calculation)) AND recency < (SELECT
recency FROM #calculation WHERE rn_recency = (SELECT CAST(COUNT(DISTINCT customerid) * 0.5 AS
INT) FROM #calculation)) THEN '2'

        WHEN recency >= (SELECT recency FROM #calculation WHERE rn_recency = (SELECT
CAST(COUNT(DISTINCT customerid) * 0.5 AS INT) FROM #calculation)) AND recency < (SELECT recency
FROM #calculation WHERE rn_recency = (SELECT CAST(COUNT(DISTINCT customerid) * 0.75 AS INT)
FROM #calculation)) THEN '3'

        ELSE '4'

    END AS R,

    CASE

        WHEN frequency < (SELECT frequency FROM #calculation WHERE rn_frequency = (SELECT
CAST(COUNT(DISTINCT customerid) * 0.25 AS INT) FROM #calculation)) AND frequency >= (SELECT
frequency FROM #calculation WHERE rn_frequency = 1) THEN '1'

        WHEN frequency >= (SELECT frequency FROM #calculation WHERE rn_frequency = (SELECT
CAST(COUNT(DISTINCT customerid) * 0.25 AS INT) FROM #calculation)) AND frequency < (SELECT
frequency FROM #calculation WHERE rn_frequency = (SELECT CAST(COUNT(DISTINCT customerid) *
0.5 AS INT) FROM #calculation)) THEN '2'

        WHEN frequency >= (SELECT frequency FROM #calculation WHERE rn_frequency = (SELECT
CAST(COUNT(DISTINCT customerid) * 0.5 AS INT) FROM #calculation)) AND frequency < (SELECT
frequency FROM #calculation WHERE rn_frequency = (SELECT CAST(COUNT(DISTINCT customerid) *
0.75 AS INT) FROM #calculation)) THEN '3'

        ELSE '4'

    END AS F,

    CASE
```

```sql
        WHEN monetary < (SELECT monetary FROM #calculation WHERE rn_monetary = (SELECT
CAST(COUNT(DISTINCT customerid) * 0.25 AS INT) FROM #calculation)) AND monetary >= (SELECT
monetary FROM #calculation WHERE rn_monetary = 1) THEN '1'

        WHEN monetary >= (SELECT monetary FROM #calculation WHERE rn_monetary = (SELECT
CAST(COUNT(DISTINCT customerid) * 0.25 AS INT) FROM #calculation)) AND monetary < (SELECT
monetary FROM #calculation WHERE rn_monetary = (SELECT CAST(COUNT(DISTINCT customerid) *
0.5 AS INT) FROM #calculation)) THEN '2'

        WHEN monetary >= (SELECT monetary FROM #calculation WHERE rn_monetary = (SELECT
CAST(COUNT(DISTINCT customerid) * 0.5 AS INT) FROM #calculation)) AND monetary < (SELECT
monetary FROM #calculation WHERE rn_monetary = (SELECT CAST(COUNT(DISTINCT customerid) *
0.75 AS INT) FROM #calculation)) THEN '3'

        ELSE '4'

    END AS M

INTO #result

FROM

    #calculation;


-- Display the results with concatenated RFM group

SELECT

    *,

    CONCAT(R, F, M) AS [group]

INTO

    #customer_segmentation

FROM

    #result;


-- Mapping data to customer segments

SELECT

    CONCAT(R, F, M) AS [group],

    COUNT(*) AS total_client,
```

```sql
    CASE

        WHEN CONCAT(R, F, M) IN ('444', '443', '434', '344') THEN 'champions'

        WHEN CONCAT(R, F, M) IN ('442', '441', '432', '431', '433', '343', '342', '341') THEN 'loyal customer'

        WHEN CONCAT(R, F, M) IN ('424', '423', '324', '323', '413', '414', '343', '334') THEN 'potential loyalist'

        WHEN CONCAT(R, F, M) IN ('333', '332', '331', '313') THEN 'recent customer'

        WHEN CONCAT(R, F, M) IN ('422', '421', '412', '411', '311', '321', '312', '322', '314') THEN 'promising
customers'

        WHEN CONCAT(R, F, M) IN ('142', '231', '232', '241') THEN 'customer needing attention'

        WHEN CONCAT(R, F, M) IN ('244', '234', '243', '233', '224', '144', '143', '133') THEN 'new customer'

        WHEN CONCAT(R, F, M) IN ('131', '132', '141', '242') THEN 'at risk customer'

        WHEN CONCAT(R, F, M) IN ('214', '213', '134') THEN 'cant lose them'

        WHEN CONCAT(R, F, M) IN ('223', '221', '222', '211', '212', '124') THEN 'hibernating'

        WHEN CONCAT(R, F, M) IN ('111', '112', '113', '114', '121', '122', '123', '222') THEN 'lost'

        ELSE 'other'

    END AS customer_type

INTO

    #customer_segmentation

FROM

    #result

GROUP BY

    CONCAT(R, F, M)

ORDER BY

    total_client;


-- Display final customer segmentation

SELECT

    *
```

```sql
FROM
    #customer_segmentation;


-- CREATE TABLE CUSTOMER_SEGMENTATION

SELECT * INTO Customer_Segmentation FROM #customer_segmentation;

-- Calculate total customers per customer type

SELECT
    SUM(total_client) AS TOTAL_CUSTOMER,
    customer_type
FROM
    Customer_Segmentation
GROUP BY
    customer_type
ORDER BY
    SUM(total_client) DESC;


-- Calculate Quartile R, F, và M

WITH CTE AS (
    SELECT
        [recency],
        ROUND([frequency] * 12, 4) AS frequency_scaled, -- Multiply Frequency by 12
        [monetary],
        ROW_NUMBER() OVER (ORDER BY [recency]) AS R_RowNum,
        ROW_NUMBER() OVER (ORDER BY [frequency]) AS F_RowNum,
        ROW_NUMBER() OVER (ORDER BY [monetary]) AS M_RowNum,
        COUNT(*) OVER () AS TotalRows
    FROM
```

```sql
    #calculation
)
SELECT
    (SELECT [recency] FROM CTE WHERE R_RowNum = CEILING(TotalRows * 0.25)) AS
R_25th_Percentile,

    (SELECT [recency] FROM CTE WHERE R_RowNum = CEILING(TotalRows * 0.50)) AS
R_50th_Percentile,

    (SELECT [recency] FROM CTE WHERE R_RowNum = CEILING(TotalRows * 0.75)) AS
R_75th_Percentile,

    (SELECT [frequency_scaled] FROM CTE WHERE F_RowNum = CEILING(TotalRows * 0.25)) AS
F_25th_Percentile, -- Use frequency_scaled instead of frequency

    (SELECT [frequency_scaled] FROM CTE WHERE F_RowNum = CEILING(TotalRows * 0.50)) AS
F_50th_Percentile,

    (SELECT [frequency_scaled] FROM CTE WHERE F_RowNum = CEILING(TotalRows * 0.75)) AS
F_75th_Percentile,

    (SELECT [monetary] FROM CTE WHERE M_RowNum = CEILING(TotalRows * 0.25)) AS
M_25th_Percentile,

    (SELECT [monetary] FROM CTE WHERE M_RowNum = CEILING(TotalRows * 0.50)) AS
M_50th_Percentile,

    (SELECT [monetary] FROM CTE WHERE M_RowNum = CEILING(TotalRows * 0.75)) AS
M_75th_Percentile;
```