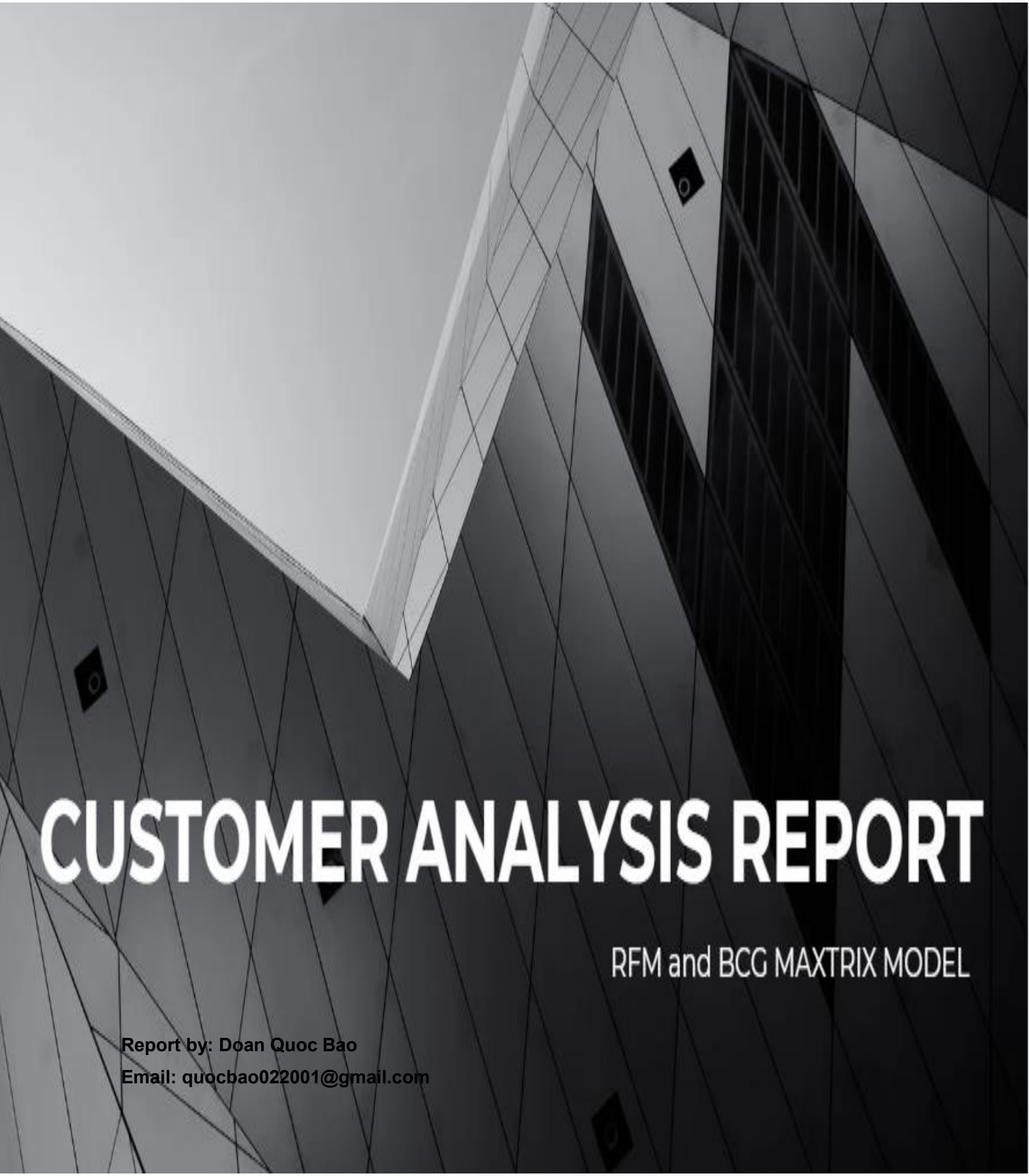


Report by: Doan Quoc Bao
Email: quocbao022001@gmail.com

Contents

1. Introduction

1.1 Customer 360

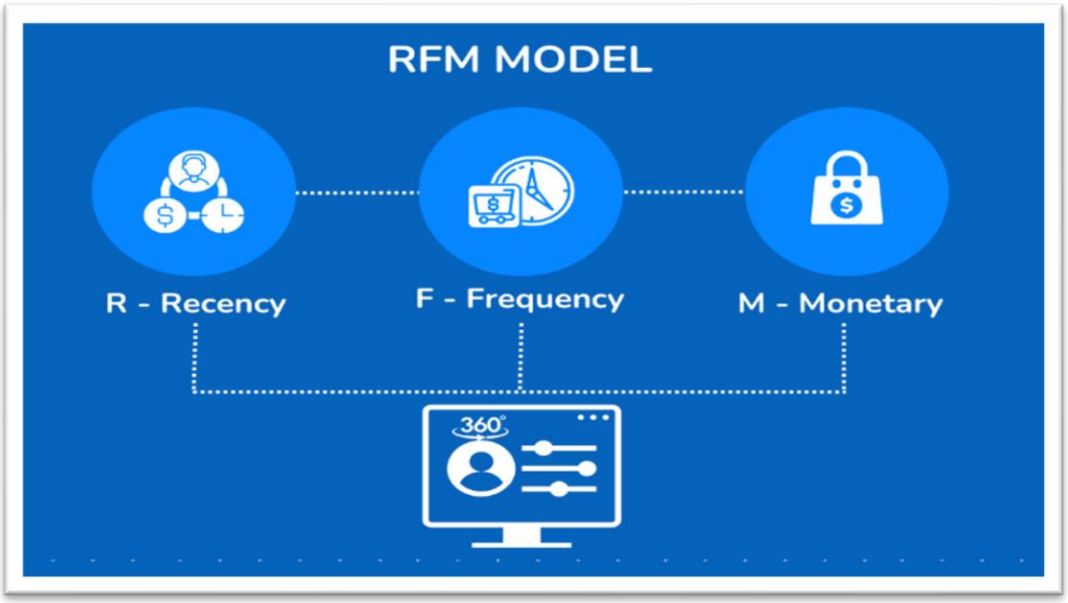
Definition: Customer 360 is a holistic view of all customer interactions and data across all channels, systems, and stages in the customer journey. Its goal is to provide businesses with a unified, complete view of their customers, regardless of the channel they interact with (website, social media, store, customer service, etc.).

Application: Customer360 helps businesses gain deeper insights into their customers, personalize experiences, enhance marketing strategies, and improve customer satisfaction and retention.

Picture 1. Customer 360

1.2 RFM Model

RFM (Recency, Frequency, Monetary) is a marketing and customer analytics technique used to evaluate and categorize customers based on their purchasing behavior. The RFM model helps businesses identify the most valuable customers, segment their customer base, and tailor marketing strategies. Here's what each component of RFM stands for:



Picture 2. RFM Model

RFM analysis ranks each customer on the following factors:

Recency (R): Measures how recently a customer made a purchase. Customers who made a purchase more recently are more likely to engage with the business again. The more recent the purchase, the higher the score for recency.

Frequency (F): Measures how often a customer makes a purchase within a specific period. Customers who buy more frequently are usually more loyal and valuable to the business. More frequent customers receive higher scores for frequency.

Monetary (M): Measures how much a customer spends. Higher spending customers are more valuable, so they receive higher scores for monetary value.

- **Benefits of RFM Analysis**
  - Customer Segmentation: RFM helps identify groups like loyal customers, at-risk customers, and big spenders, enabling targeted marketing.
  - Predicting Customer Behavior: Businesses can predict the likelihood of future purchases based on past behavior.
  - Improving Marketing ROI: By focusing on high RFM score customers, businesses can increase their return on marketing investments.
- **RFM Score**

In this report, we utilize the RFM (Recency, Frequency, Monetary) model to segment customers based on their purchasing behavior. The RFM model assigns scores to customers based on

three key dimensions: Recency, Frequency, and Monetary value. Each of these dimensions is divided into four levels, with scores ranging from 1 to 4. The score will be segmented as below:

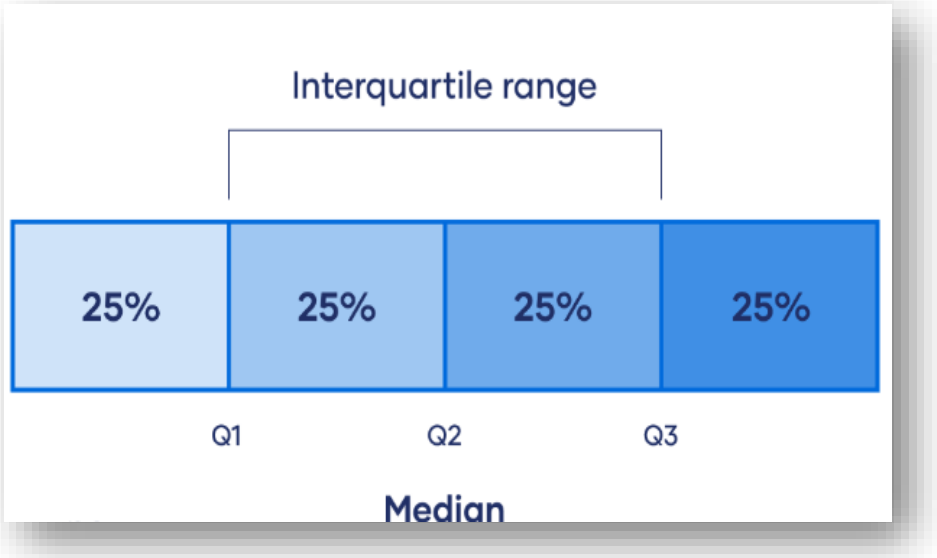
Table 1: Definition of RFM Model

Score	Recency (R)	Frequency (F)	Monetary (M)
1	Latest order (most recent)	Lowest frequency (least frequent buyer)	Lowest spending (least money spent)
2	The second most recent order	Low frequency	Low spending
3	Older order	High frequency	High spending
4	First order (least recent)	Highest frequency (most frequent buyer)	Highest spending (most money spent)

1.3 IQR (Interquartile range)

The Interquartile Range (IQR) is a statistical measure used to assess the spread of a dataset by dividing it into four quartiles. After sorting the data in ascending order, the dataset is split into four equal parts. The first quartile (Q1) represents the 25th percentile, meaning 25% of the data falls below this value. The second quartile (Q2) is the median, representing the 50th percentile. The third quartile (Q3) corresponds to the 75th percentile, indicating that 75% of the data lies below this value. Finally, the fourth quartile (Q4) covers the highest 25% of the data, from Q3 to the maximum value, capturing the top-performing data points.

Picture 3. Interquartile range



1.4 BCG Maxtrix

The Boston Consulting group's product portfolio matrix (BCG matrix) is designed to help with long-term strategic planning, to help a business consider growth opportunities by reviewing its portfolio of products to decide where to invest, to discontinue, or develop products. It's also known as the Growth/Share Matrix.

The Matrix is divided into 4 quadrants based on an analysis of market growth and relative market share, as shown in the diagram below.

Picture 4. BCG Matrix Model

BCG Matrix in RFM is a strategic tool that classifies customers into four distinct categories—Stars, Cash Cows, Big Questions, and Dogs—based on their RFM scores. Each category reflects the customers' purchasing behavior and their potential value to the business.

- Stars: Customers with high scores in Recency, Frequency, and Monetary value. They are recent and frequent buyers who contribute significantly to revenue. This segment represents the most valuable customers and should be prioritized for retention efforts.
- Cash Cows: Customers who demonstrate moderate Recency but high Frequency and Monetary scores. These customers consistently generate revenue but may not engage as recently. Maintaining their loyalty is crucial for ongoing profitability.
- Big Questions: Customers with high Recency but low Frequency and moderate Monetary scores. They are recent purchasers but do not buy often or spend much. This

segment presents an opportunity for growth through targeted marketing strategies aimed at increasing their engagement and spending.

- Dogs: Customers with low scores across Recency, Frequency, and Monetary value. They have not purchased recently and contribute minimally to revenue. Strategies for this segment may include re-engagement efforts or reassessing the viability of marketing investments directed at them.

2. Analysis

2.1 Descriptive Data

The dataset has 2 tables: customer_transaction and customer_registered

- **Customer_transaction Table (Fact table)**

This table captures information related to customer purchases, tracking each individual transaction along with associated details such as customer ID, transaction date, and the revenue generated. It can be used to analyze customer behavior, transaction frequency, and the monetary value of each transaction.

Table 2. Columns of Customer_transaction Table

Column name	Data type	Meaning
transaction_id	int	Transaction code of customers
customerid	int	Customer code
purchase_date	varchar(50)	On the day of a transaction by customer
gmv	int	Revenue of transaction

- **Customer_registered (DIM table)**

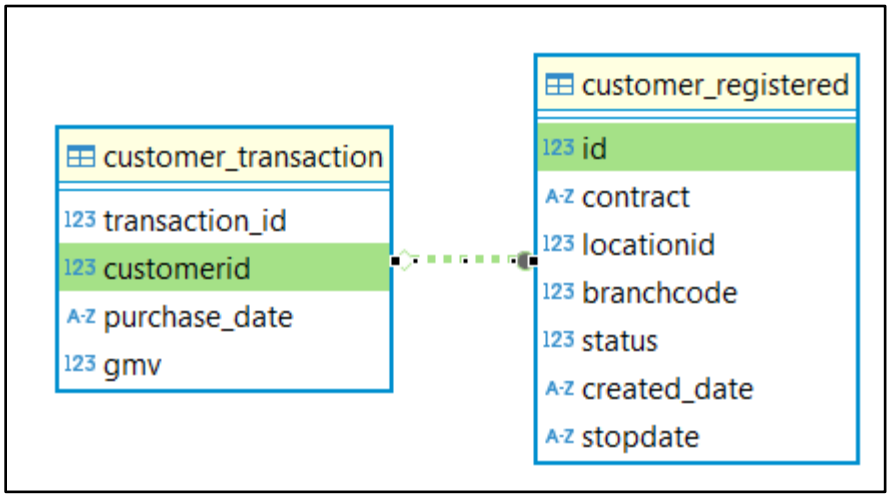
This table records information related to customer contracts, including contract details, location, branch information, and contract status. It is useful for tracking when customers initiate and terminate contracts and their association with different locations and branches.

Table 3. Columns of Customer_registered Table

Column name	Data type	Meaning
Id	Int	Customer code
contract	varchar(50)	The contract code of customers

Locationid	Int	Location code
Branchcode	Int	Branch Code
Status	Int	Status of the contract
Created_date	varchar(50)	On the day customers created a contract
stopdate	varchar(50)	On the day customers ended a contract

- **ERD Model**



Picture 5. ERD Model

The relationship of 2 tables is many to one relationship, which is between customerid and id. It means the table customer_transaction has many customer codes and it is recorded in the customer_registered table.

2.1.1 Cleaning Data

- Remove transactions that didn't have customer_id in the dataset.
- Total transaction after cleaning: 114081.

2.1.2 Segment customers based on IQR Range

- Recency: calculate by the latest transaction until 1/9/2022
- Frequency: calculate the total transaction by one customer from they created transaction, and it will be divided by the age contract of each customer.
- Monetary: total revenue of each customer.

As we mentioned, we will use IQR to calculate and categorize customers to classify each customer as belonging to which group. The table below shows how we classified them:

Table 4. Definition of calculation in RFM Model

Quarter	1	2	3	4
Recency	<= 31 days	<= 62 days	<= 92 days	> 92 days
Frequency	<= 0.19	<= 0.23	<= 0.26	> 0.26
Monetary	<=15625	<=19173	<= 23832	>23832

Described:

- Recency in Quarter 1 was the customers who had bought under 32 days from the latest transaction until 01/09/2022. So, Quarters 2,3 and 4 have the same meanings, but it has the days purchased in ascending compared to Quarter 1, respectively.
- Frequency in Quarter 1 was the customers who had bought with the lowest frequency relative to their average contract age at the platform. And Quarters 2,3,4 had a higher frequency compared to Quarter 1 in Ascending, respectively.
- Monetary in Quarter 1 was the customers who had bought with the lowest revenue relative to their average contract age at the platform; and Quarters 2,3,4 had a higher revenue compared to Quarter 1 in Ascending, respectively.
    - As a result, we can take the score for this customer, and collect the Combinations by joining these numbers of each customer into RFM scores. After that, we can organize them into the BCG Matrix model based on these combinations by their properties of BCG Matrix groups.

2.2 RFM Score and BCG Matrix

With BCG Matrix, we can identify these customers by their behavior on platforms and have matched these RFM Scores with corresponding cells in BCG Maxtrix. The table below shows how these customers will be label.

Table 5. Group Scores of BCG Matrix Model

BCG Matrix	RFM Scores
Star	444, 434, 344, 343, 433, 443, 344, 434
Cash Cow	324, 314, 324, 334
Big Question	322, 312, 232, 223, 233, 322, 323, 232
Dog	111, 122, 121, 112
Other	123, 132, 134, 142, 143, 211, 213, 221, 224, 234, 241, 242, 243, 244, 311, 313, 321, 331, 332, 333,341, 342, 411, 421, 431

After segmentation and by RFM Score and BCG Maxtrix, I named '4+1' BCG Matrix ingredients by other names for easy analysis, as shown in the picture below:

- Star as Champion
- Cash cow as Loyal Customer
- Big question as Potential Loyalty
- Dog as At risk
- Other as the remaining group and is Need Attraction.

CustomerID	recency	frequency	monetory	age_contract	R	F	M	RFM_Score	BCG_Matrix	Customer_Group
71739	62	0.14	14430	7	2	1	1	211	Other	Need Attraction
72014	62	0.27	34931	7	2	4	4	244	Star	Champion
72052	31	0.14	19934	7	1	1	3	113	Big Question	Potential Loyalty
72657	62	0.14	27506	7	2	1	4	214	Other	Need Attraction
74549	62	0.14	17224	7	2	1	2	212	Big Question	Potential Loyalty
74570	62	0.14	17224	7	2	1	2	212	Big Question	Potential Loyalty
74604	31	0.28	21357	7	1	4	3	143	Star	Champion
74662	31	0.14	14468	7	1	1	1	111	Other	Need Attraction
74797	62	0.14	14473	7	2	1	1	211	Other	Need Attraction
74819	31	0.28	26190	7	1	4	4	144	Star	Champion
74977	92	0.14	15852	7	3	1	2	312	Dog	At Risk
74979	31	0.14	21929	7	1	1	3	113	Big Question	Potential Loyalty
75205	62	0.28	35037	7	2	4	4	244	Star	Champion
75284	31	0.14	0	7	1	1	1	111	Other	Need Attraction
75428	92	0.14	14479	7	3	1	1	311	Dog	At Risk
75457	31	0.14	6895	7	1	1	1	111	Other	Need Attraction
75524	31	0.14	9652	7	1	1	1	111	Other	Need Attraction
75751	31	0.14	15864	7	1	1	2	112	Big Question	Potential Loyalty
75796	92	0.14	14484	7	3	1	1	311	Dog	At Risk
75816	92	0.14	18622	7	3	1	2	312	Dog	At Risk
75821	62	0.14	14484	7	2	1	1	211	Other	Need Attraction
75879	62	0.14	13105	7	2	1	1	211	Other	Need Attraction
75918	31	0.14	13105	7	1	1	1	111	Other	Need Attraction
75983	92	0.14	17243	7	3	1	2	312	Dog	At Risk

Picture 6. Example Data After Querying

Above is the final data after I cleaned, normalized, and queried data to collect exactly what we need to prepare for visualization and give insights behind the data to help the company see the behavior of their customers; moreover, they can find out their situations and solutions.

Picture 7. The distribution of RFM Score.

R	Total Customers
1	42049
2	30995
3	41037
Total	114081

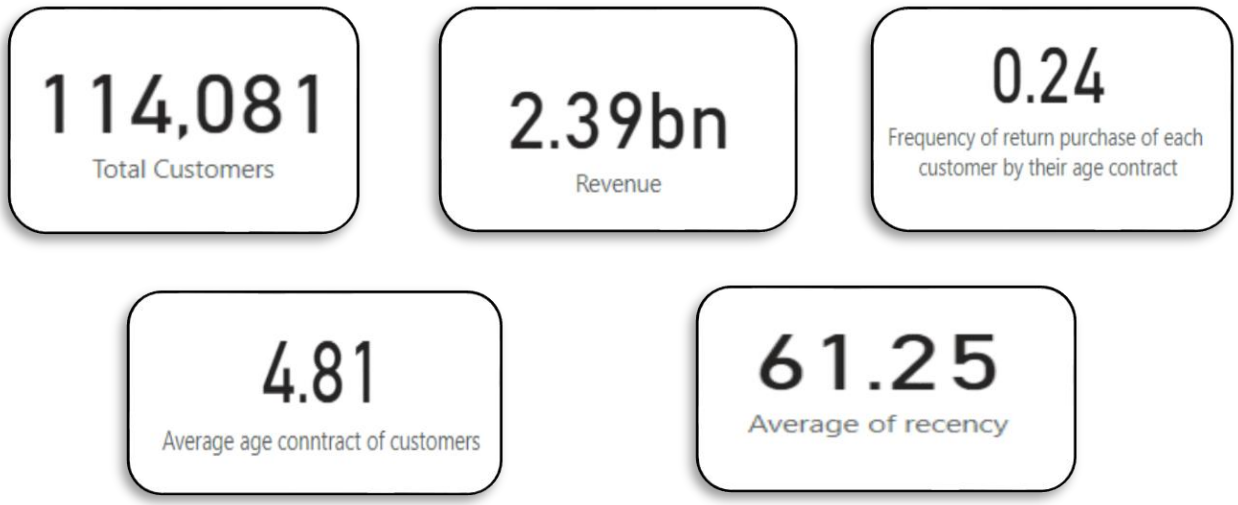
M	Total Customers
1	28569
2	28471
3	28542
4	28499
Total	114081

F	Total Customers
1	31532
2	30957
3	29428
4	22164
Total	114081

After evaluating the scores for each customer according to the RFM score from 1-4 according to 3 indicators; it was found that there was not much difference in the above score groups in terms of quantity in the groups. Especially in the Monetary group, when this group had an almost even distribution in terms of quantity.

2.3 Analysis Report

❖ ***Overview***

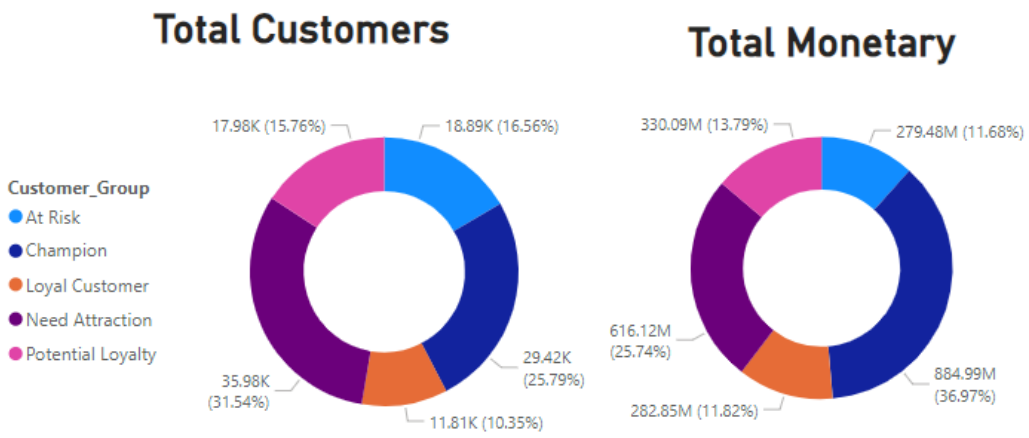


Picture 8. Overview of Customer Landscape

The data chart above shows that the company had 114,081 customers, generating 2.39 billion in revenue. With each customer on average, the percentage of making purchasing products is 24% over 100% in one year. The average time contracts for customers is 4.81 years and the average latest transaction until 01/09/2022 is 61.25 days.

Champion group has the highest contribution revenue and second-largest of customers

Picture 9. The market share and quantities of customers in each group

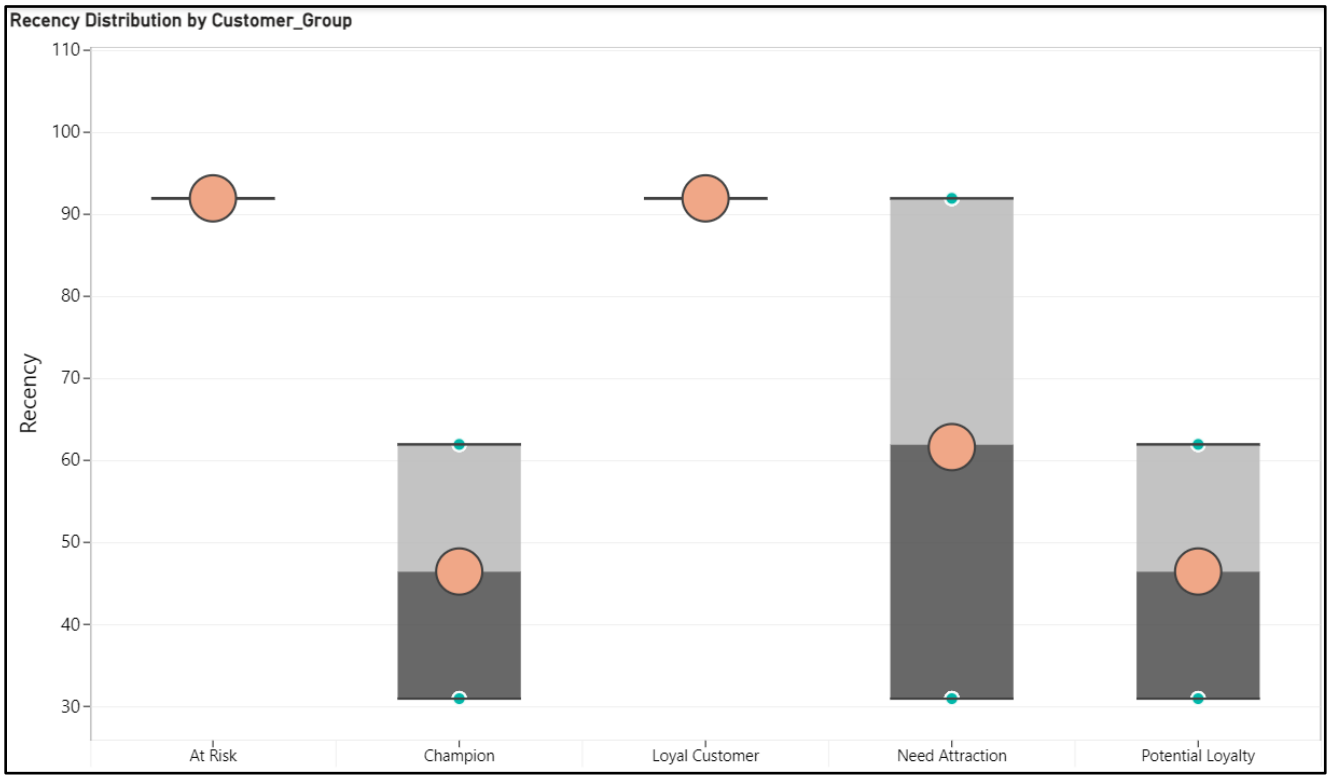


Champions serve as both the largest customer segment and the highest revenue contributor, with over 29,420 customers generating 884.99 million**.** It's surprising that Need Attraction generates over $616.12 million as the second largest, with a total of nearly 35,980 customers. Loyal customers accounted for 11.82% of revenue (282.85 million) and only 10.35% of the total number of customers. Potential loyalty is the third portion of revenue and accounted for 15.76% of customers. Finally, there is the At Risk Group customer, which accounted for 11.68% of revenue.

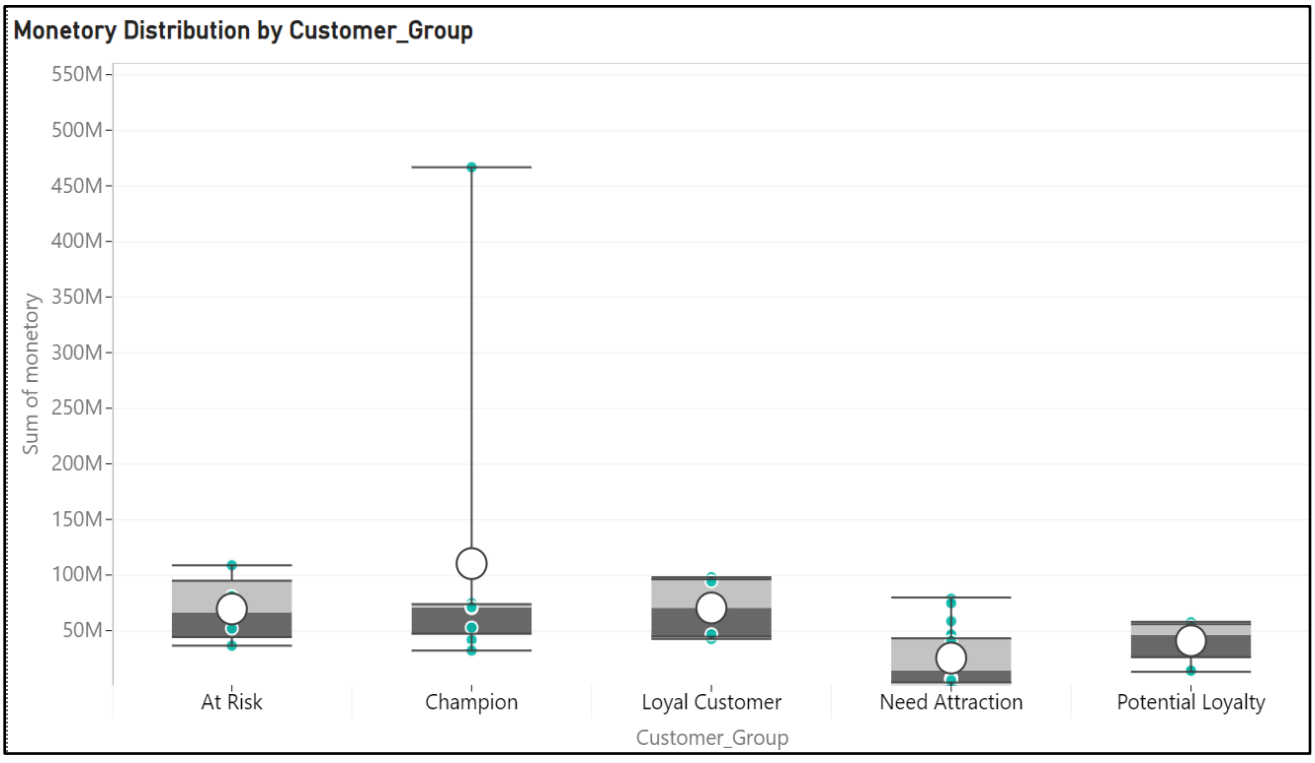
⇨ We need to take action in the Need Attraction group due to the high number of customers, as they have the potential to become loyal customers. Additionally, we still focus on 3 group: potential, loyal and champion customers because they are mainly primary customers.

The Champion Group leads in value, with the highest RFM score

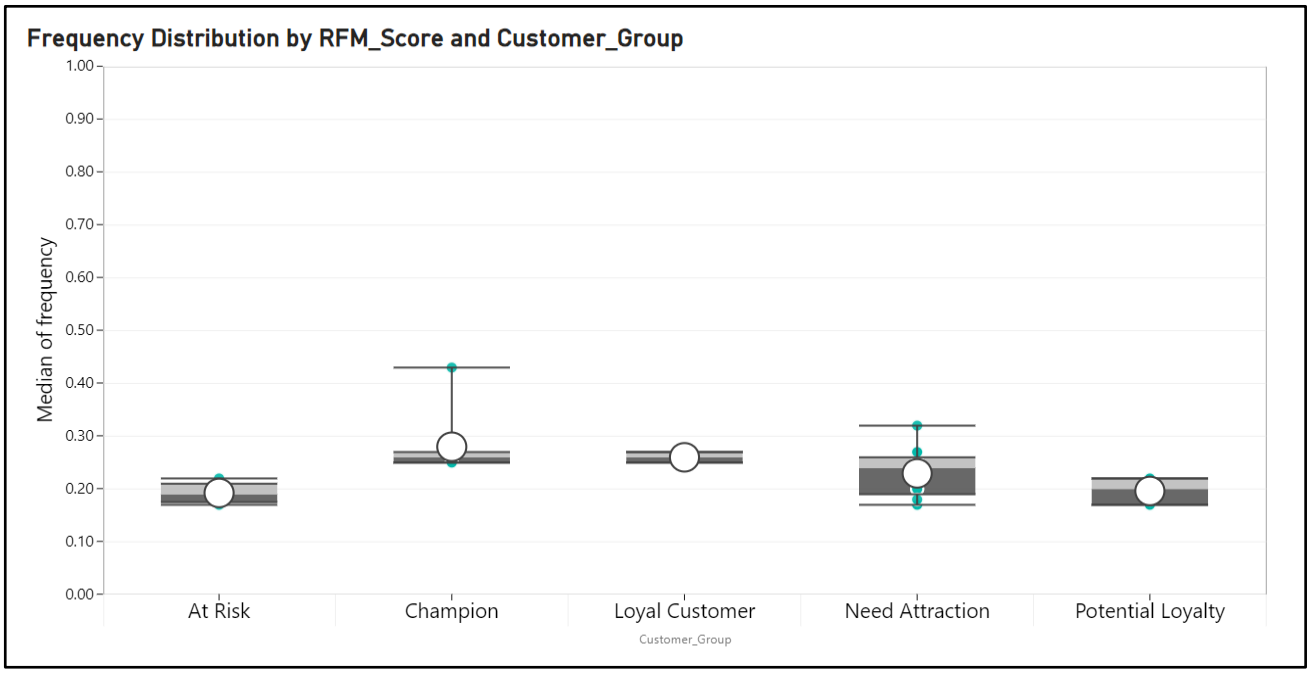
Picture 10. Boxplot chart of Recency distribution by Customer Group



Picture 11. Boxplot chart of Monetary distribution by Customer Group



Picture 12. Boxplot chart of Frequency distribution by Customer Group



Looking at the RFM score distribution, we collect insights as below:

- **Champion Group:** This group has **the highest mean purchase frequency**, approximately 0.28. This indicates that one customer in this group **makes purchases around 0.28 times per year**. Additionally, there is one outlier in terms of monetary contribution, which amounts to 467,071,432 million and accounts for 20% of total revenue, and the mean is 110 million per customer. This outlier corresponds to a group score of 144. The mean recency is 46.37 days, meaning that customers typically return to make a purchase about 46 days after their last purchase. => **Recommendation**: Maintain engagement and loyalty programs to sustain high-value contributions.
- **"Loyal customer"**: has a frequency as 0.26 which means customer in this group also buy 3 times per year and generate 70 million. But the distance day they will purchase again is also high with nearly 92 days. => **Recommendation**: Implement marketing campaigns to encourage more frequent purchases and engagement.
- **"Potential Loyalty"** customer, has a frequency score is 0.2, which tends to buy 0.2 times per year. One customer brought in around 40 million in revenue. But the duration of time they will come back is as same as Champion customers (about 46 days). =>

Recommendation: Focus on increasing spending and purchase frequency through targeted promotions and rewards.

- **“Need Attraction”**: The purchase frequency for this customer group is 0.23 times per year, indicating that they make purchases somewhat infrequently. On average, each customer spends only 25 million, and the duration between purchases is approximately 60 days. These figures suggest that this group is currently not a strong market potential. Therefore, the platform could offer additional benefits to attract them, particularly because they represent the largest segment, accounting for 31% of the total customer base. => **Recommendation**: Offer benefits or promotions to encourage more frequent purchases and improve engagement.
- **“At Risk** has a frequency score of 0.19, which is the lowest among our customer segments. Though they generate 69 million per year, the average time between their purchases is 91 days. This indicates that they are not the primary customers of this platform. => This group can be existing customers, so it needs attractive strategies to activate them. => **Recommendation**: Design reactivation strategies to bring these customers back into regular purchasing behavior.

2. Coding Process

```sql
CREATE DATABASE DRM_Project;
USE drm_project;
create table rfm_calculation as
with RFM_Data as(
SELECT
    cr.id,
    DATEDIFF('2022-09-01', MAX(STR_TO_DATE(ct.purchase_date, '%m/%d/%Y'))) AS recency,
    ROUND(1.00 * COUNT(ct.purchase_date) / (DATEDIFF('2022-09-01', STR_TO_DATE(cr.created_date, '%m/%d/%Y'))/365), 2) AS frequency,
    ROUND(1.00 * SUM(ct.GMV) / (DATEDIFF('2022-09-01', STR_TO_DATE(cr.created_date, '%m/%d/%Y'))/365), 0) AS monetory,
    round( DATEDIFF('2022-09-01', STR_TO_DATE(cr.created_date, '%m/%d/%Y'))/365,0)  AS age_contract
FROM drm_project.customer_transaction ct
LEFT JOIN drm_project.customer_registered cr ON ct.customerid = cr.id
where cr.id is not null
GROUP BY cr.id, cr.created_date)
 select *,
        row_number () over ( order by recency) as rn_recency,
        row_number () over ( order by frequency) as rn_frequency,
        row_number () over ( order by monetory) as rn_monetory
 from RFM_Data;

with RFM_Data as(
SELECT
    cr.id,
    DATEDIFF('2022-09-01', MAX(STR_TO_DATE(ct.purchase_date, '%m/%d/%Y'))) AS recency,
    ROUND(1.00 * COUNT(ct.purchase_date) / (DATEDIFF('2022-09-01', STR_TO_DATE(cr.created_date, '%m/%d/%Y'))/365), 2) AS frequency,
    ROUND(1.00 * SUM(ct.GMV) / (DATEDIFF('2022-09-01', STR_TO_DATE(cr.created_date, '%m/%d/%Y'))/365), 0) AS monetory,
    round( DATEDIFF('2022-09-01', STR_TO_DATE(cr.created_date, '%m/%d/%Y'))/365,0)  AS age_contract
FROM drm_project.customer_transaction ct
LEFT JOIN drm_project.customer_registered cr ON ct.customerid = cr.id
where cr.id is not null
GROUP BY cr.id, cr.created_date)
 select *,
        row_number () over ( order by recency) as rn_recency,
        row_number () over ( order by frequency) as rn_frequency,
        row_number () over ( order by monetory) as rn_monetory
 from RFM_Data;
```

```sql
use drm_project;
create table rfm_final as
with RFM_Segment as (SELECT *, CASE
    WHEN recency >= min(recency)
    and recency <= (select recency from drm_project.rfm_calculation WHERE rn_recency = (select round(count(CustomerID) * 0.25, 0) from drm_project.rfm_calculation)) then 1
    WHEN rn_recency > (select recency from drm_project.rfm_calculation WHERE rn_recency = (select round(count(CustomerID) * 0.25, 0) from drm_project.rfm_calculation))
    and recency <= (select recency from drm_project.rfm_calculation WHERE rn_recency = (select round(count(CustomerID) * 0.5, 0) from drm_project.rfm_calculation)) then 2
    WHEN recency > (select recency from drm_project.rfm_calculation WHERE rn_recency = (select round(count(CustomerID) * 0.5, 0) from drm_project.rfm_calculation))
    and recency <= (select recency from drm_project.rfm_calculation WHERE rn_recency = (select round(count(CustomerID) * 0.75, 0) from drm_project.rfm_calculation)) then 3
    ELSE 4 END as R,
    CASE
    WHEN frequency >= min(frequency)
    and frequency <= (select frequency from drm_project.rfm_calculation WHERE rn_frequency = (select round(count(CustomerID) * 0.25, 0) from drm_project.rfm_calculation)) then 1
    WHEN frequency > (select frequency from drm_project.rfm_calculation WHERE rn_frequency = (select round(count(CustomerID) * 0.25, 0) from drm_project.rfm_calculation))
    and frequency <= (select frequency from drm_project.rfm_calculation WHERE rn_frequency = (select round(count(CustomerID) * 0.5, 0) from drm_project.rfm_calculation)) then 2
    WHEN frequency > (select frequency from drm_project.rfm_calculation WHERE rn_frequency = (select round(count(CustomerID) * 0.5, 0) from drm_project.rfm_calculation))
    and frequency <= (select frequency from drm_project.rfm_calculation WHERE rn_frequency = (select round(count(CustomerID) * 0.75, 0) from drm_project.rfm_calculation)) then 3
    ELSE 4 END as F,
    CASE
    WHEN monetory >= min(monetory)
    and monetory <= (select monetory from drm_project.rfm_calculation WHERE rn_monetory = (select round(count(CustomerID) * 0.25, 0) from drm_project.rfm_calculation)) then 1
    WHEN monetory > (select monetory from drm_project.rfm_calculation WHERE rn_monetory = (select round(count(CustomerID) * 0.25, 0) from drm_project.rfm_calculation))
    and monetory <= (select monetory from drm_project.rfm_calculation WHERE rn_monetory = (select round(count(CustomerID) * 0.5, 0) from drm_project.rfm_calculation)) then 2
    WHEN monetory > (select monetory from drm_project.rfm_calculation WHERE rn_monetory = (select round(count(CustomerID) * 0.5, 0) from drm_project.rfm_calculation))
    and monetory <= (select monetory from drm_project.rfm_calculation WHERE rn_monetory = (select round(count(CustomerID) * 0.75, 0) from drm_project.rfm_calculation)) then 3
    ELSE 4 END as M
 from drm_project.rfm_calculation
group by CustomerID, rn_recency,rn_frequency,rn_monetory, recency,frequency,monetory)
select  *, concat(R,F,M) as RFM_Score,
        CASE
        WHEN R IN (1,2) AND F IN (3, 4) AND M IN (3, 4) THEN 'Star'
        WHEN R IN (2,3) AND F IN (3,4) AND M IN (3, 4) THEN 'Cash Cow'
        WHEN R IN (1,2) AND F IN (1,2) AND M IN (2,3) THEN 'Big Question'
        WHEN R IN (3,4) AND F IN (1, 2) AND M IN (1, 2) THEN 'Dog'
        ELSE 'Other' end as BCG_Matrix
from RFM_Segment
where customerID is not null;
```