2024

CUSTOMER SEGMENTATION REPORT

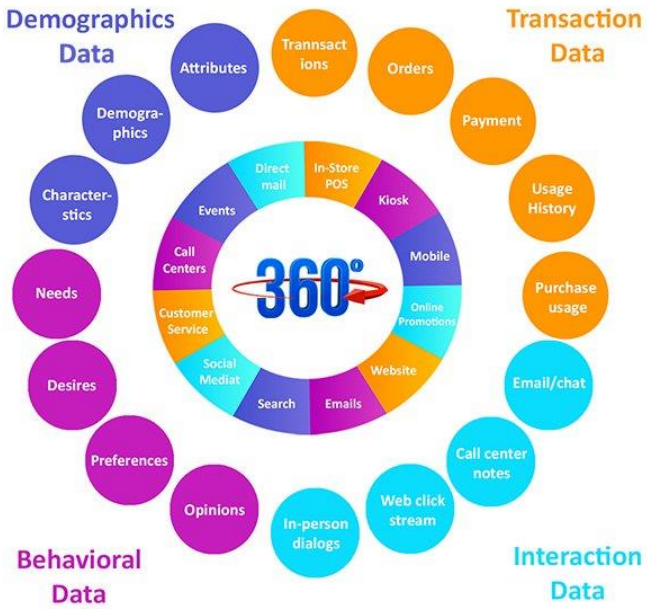
Presented by:
Tran Thien Ha My

+84931599459
MYITRAN97@GMAIL.COM

TABLE OF CONTENTS

I. Overview

1.1. Customer360 strategy



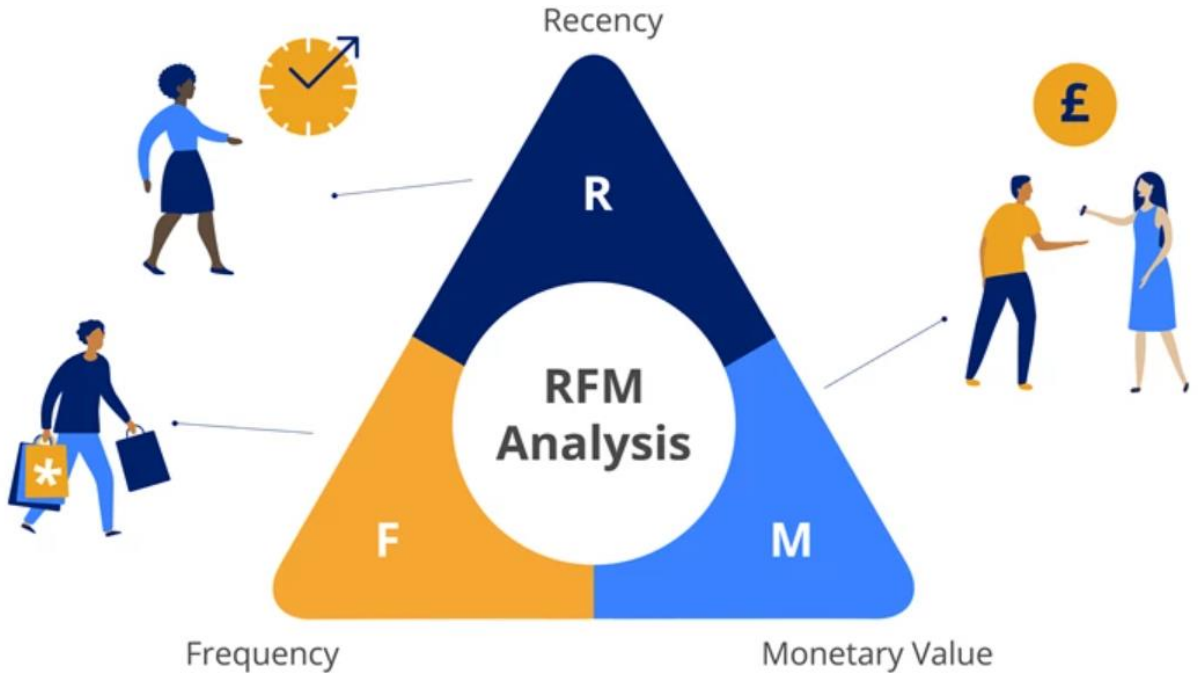
Customer360 is a strategy or approach used by businesses to gain a comprehensive understanding of their customers. This involves analysing data from various touchpoints and sources across the customer's journey to create a complete, unified profile of each customer. Customer360 data includes**:**

- **Transaction data:** customer’s transaction history like payments, orders, usage history, purchase history
- **Demographics data:** customer’s basic information such demographics data
- **Behavioral data:** customer’s needs, desires, preferences, opinions
- **Interaction data:** interaction history of customers such as call logs, emails, web click

Benefits of using Customer360 analysis:

- Personalized Customer Experiences
- Improved Customer Service
- Increased Sales and Revenue
- Enhanced Customer Retention
- Better Decision-Making
- Cross-Selling and Upselling Opportunities

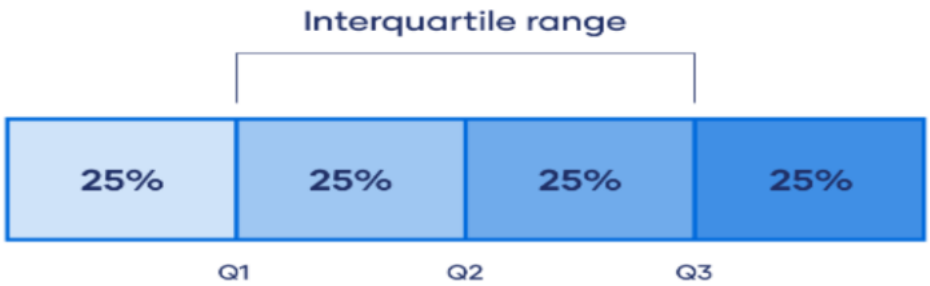
1.2. The RFM Model



RFM analysis is a marketing technique used to rank and segment customers based on their purchasing behavior. This method helps understand and target business's most valuable customers by evaluating three key factors:

- **Recency**: How recent a customer has made a purchase. Customers who recently made a purchase will be more likely to come again
- **Frequency**: How often a customer makes a purchase. Customers who buys often is more valuable
- **Monetary**: How much money does a customer spend. Customers who spends more money is more valuable

1.3. Interquartile range



The interquartile range (IQR) is a measure of statistical dispersion, which is the spread of the middle 50% of a data set. It is calculated by subtracting the first quartile (Q1) from the third quartile (Q3).

Steps to Calculate the Interquartile Range (IQR):

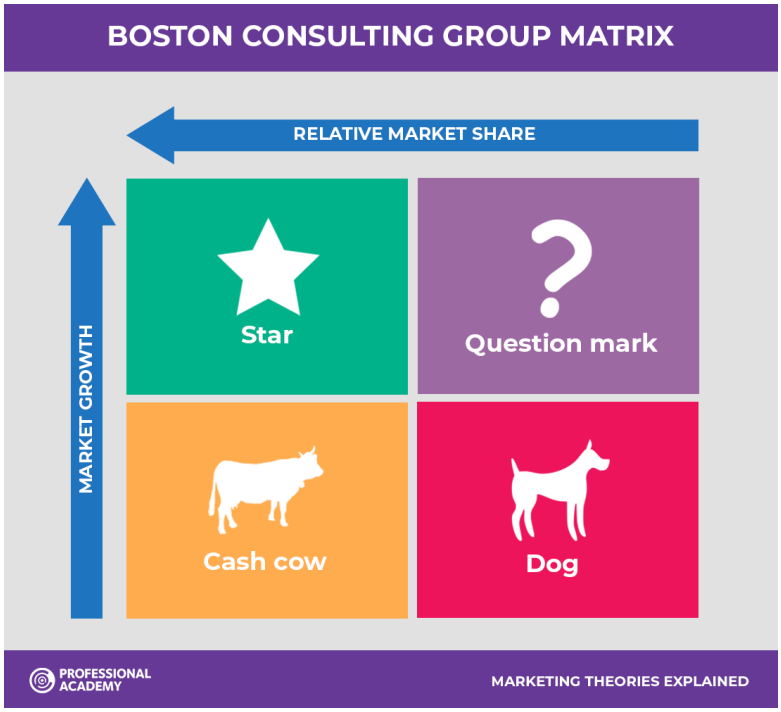
1. Arrange the Data: Organize the data points in ascending order.
2. Determine the Quartiles:

❖ First Quartile (Q1): This is the median of the lower half of the data

❖ Second Quartile (Q2): This is the median of the whole data set.

❖ Third Quartile (Q3): This is the median of the upper half of the data

1.4. The BCG matrix



The Boston Consulting Group Matrix, or BCG matrix, is a strategic business tool used to help organizations analyze their product lines or business units for strategic planning purposes.

According to the BCG matrix, customers are classified into 4 groups. They are:

1. **Stars:** These are high-value customers with strong potential for growth. They are often frequent buyers and are highly engaged with the website.
2. **Question Marks**: These customers have potential but currently do not bring much value, possibly because they are new or not fully engaged.
3. **Cash Cows**: These are reliable customers who provide steady revenue without the need for much investment.
4. **Dogs**: These customers contribute minimally to profitability and have low growth prospects.

II. Method

2.1. Getting available customers for analysis

Customers' data is stored inside two tables which is Customer_Registered and Customer_Transaction

Table 1: Customer_Registered

	ID	Contract	LocationID	BranchCode	Status	created_date	stopdate
1	0	SGDN00215	8	1	0	2011-11-25 00:00:00.000	2012-01-05 00:00:00.000
2	1	SGDN00214	8	1	0	2012-06-14 00:00:00.000	[NULL]
3	2	SGD374348	8	1	0	2012-11-01 00:00:00.000	[NULL]
4	3	SGD022064	8	1	2	2011-06-22 00:00:00.000	2013-05-29 00:00:00.000
5	4	SGD041015	8	5	2	2011-12-17 00:00:00.000	2014-11-11 00:00:00.000
6	5	SGDN00211	[NULL]	[NULL]	2	2015-06-09 00:00:00.000	2015-09-09 00:00:00.000
7	6	SGD374348	8	1	3	2012-11-26 00:00:00.000	2012-12-13 00:00:00.000

Table 2: Customer_Transaction

	ID	CustomerID	Purchase_Date	GMV
1	0	1327813	2022-06-01 00:00:00.000	95,000
2	1	1157830	2022-06-01 00:00:00.000	75,000
3	2	873915	2022-07-01 00:00:00.000	95,000
4	3	3505071	2022-07-01 00:00:00.000	90,000
5	4	2930918	2022-07-01 00:00:00.000	109,091
6	5	899882	2022-06-01 00:00:00.000	105,000

The set of available customers for analysis will be selected by joining the Customer_Registered and Customer_Transaction table, but choosing customers where stopdate = NULL (means the customer's contract has not been terminated)

2.2. Calculate RFM values

We will calculate each customer's Recency, Frequency and Monetary value by:

- **Recency**: the number of days between today and the customer's last visit
- **Frequency**:

$$\frac{2022/09/01 - \text{the day of the customer's most recent purchase}}{2022/09/01 - \text{the contract's creation date}}$$

- **Monetary**:

$$\frac{total\ money\ spent}{number\ of\ times\ of\ purchase}$$

2.3. Calculate Inter Quartile Range

	Q1	Q2	Q3
Recency	31	62	92
Frequency	0.0005	0.0006	0.0007
Monetary	75,000	86,000	95,000

2.4. Customer segmentation

Each customer is evaluated based on three key metrics: Recency, Frequency, and Monetary value. Here's how each metric is scored:

Recency: A score of 4 is assigned to customers who have visited the website very recently, indicating high engagement. Conversely, a score of 1 is given to customers who have low engagement.

Frequency: Customers who visit the website very frequently receive a high score of 4, reflecting their regular engagement. Those who rarely visit the website are scored as 1.

Monetary: A score of 4 is assigned to customers who spend large amounts per visit. A score of 1 is given to customers who spend minimal amounts

Using the interquartile range, each customer is assigned a RFM score.

	1	2	3	4
Recency	>92	62 < R <= 92	31 < R =< 62	R <= 31
Frequency	F < 0.0005	0.0006 >=F > 0.0005	0.0007 > F >= 0.0006	F > 0.0007
Monetary	M < 75,000	75,000 < M < 85,000	85,000 < M < 95,000	M>95,000

III. RESULTS

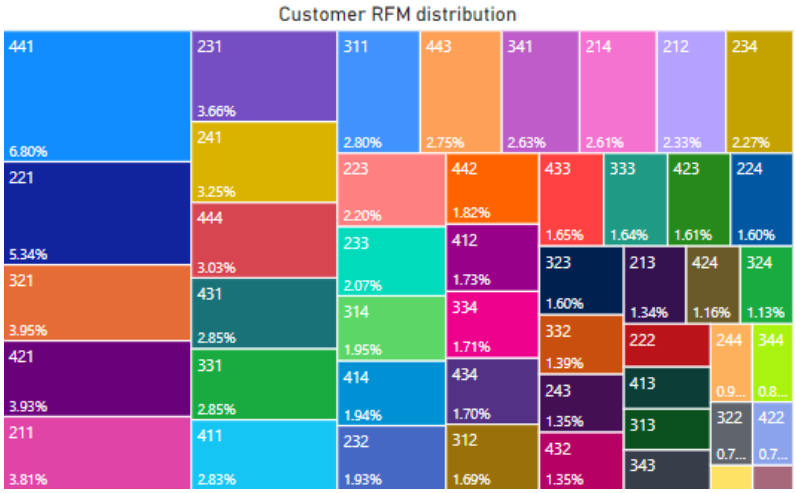
3.1. OLAP output

After applying the above steps, we have this table for analysis

	customerid	Recency	Frequency	Amount spent per transaction	R_Q1	R_Q2	R_Q3	F_Q1	F_Q2
1	1018163	31	0.0007390983	70,000	38.75	62	84.5	0.0005784515	0.0006849636
2	141697	92	0.000396668	143,182	38.75	62	84.5	0.0005784515	0.0006849636
3	569661	62	0.0005714286	30,000	38.75	62	84.5	0.0005784515	0.0006849636
4	653077	62	0.0005995204	75,000	38.75	62	84.5	0.0005784515	0.0006849636

F_Q3	M_Q1	M_Q2	M_Q3	R	F	M	RFM_result
0.0006974945	75,000	80,000	102,500	4	4	1	441
0.0006974945	75,000	80,000	102,500	1	1	4	114
0.0006974945	75,000	80,000	102,500	3	1	1	311
0.0006974945	75,000	80,000	102,500	3	2	1	321

3.2. Customer RFM distribution

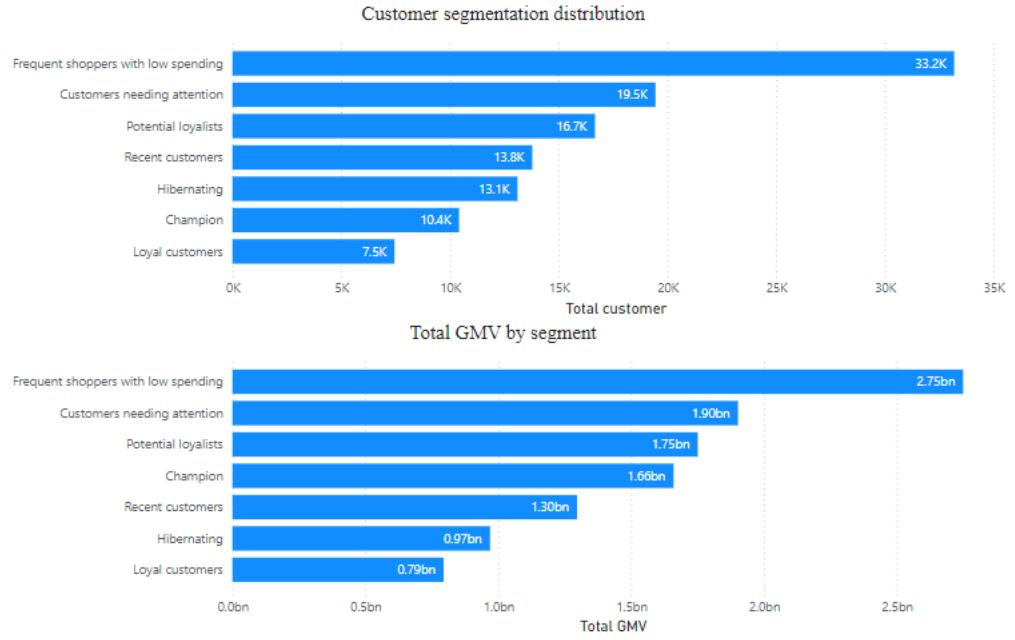


3.3. Customer segmentation criteria

Segment	Characteristics	RFM scores
Champions	Top customers. Purchase frequently, have recently made a purchase and spend a lot	444, 443, 434, 433
Loyal Customers	Loyal customers who frequently visit, engage with us recently and spend well (not as much as champions)	344, 343, 432, 334, 333

Potential loyalists	Customers who have made recent purchase, buy frequently and spend moderately	243, 342, 442, 424, 423, 324, 244, 422, 322, 323, 332, 323, 232
Frequent shoppers with low spending	Customers who visits a lot but do not spend much	441, 341, 241, 411, 321, 421, 431, 331
Recent customers	Customers who made recent purchase but do not buy often	414, 314, 413, 412, 312, 313, 311
Customers needing attention	These customers have not purchased recently but used to buy frequently and spend moderately. Efforts should be made to re-engage them	214, 242, 213, 222, 224, 223, 234, 231, 233
Hibernating	Customers spends little and do not visit often	211, 212, 221
Lost	Customers who have not purchased for a long time, do not spend much and do not buy frequently	114, 113, 131, 132

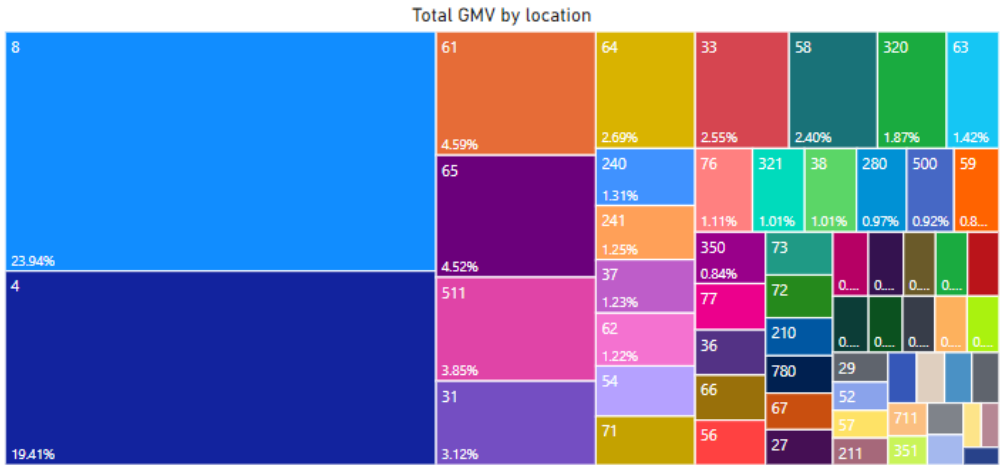
3.4. Results visualization and Analysis



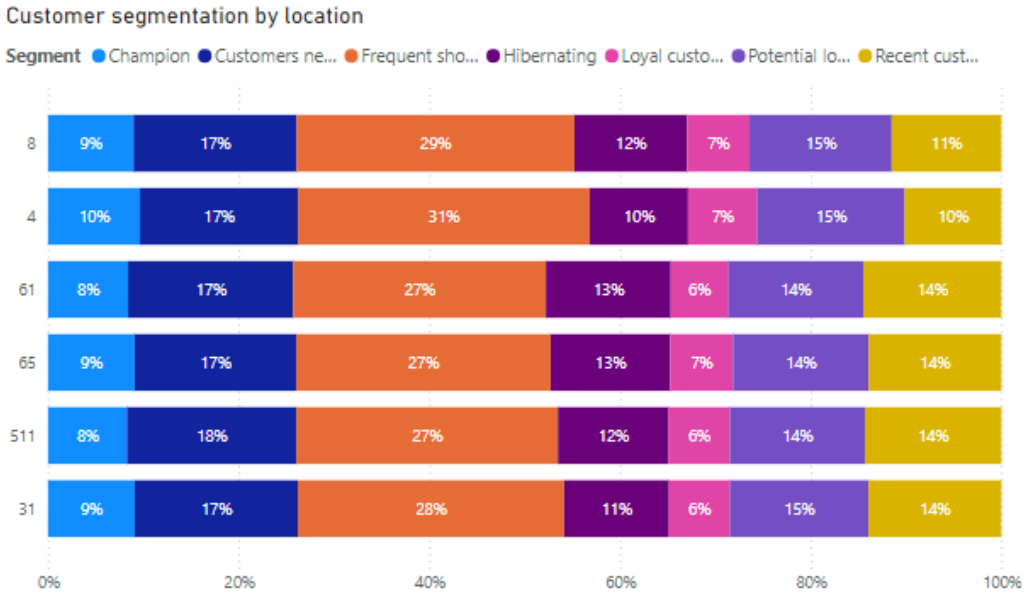
Most of our customers are **frequent shoppers with low spending** (33.2K customers). Customers needing attention have the **second-highest customer**, followed by Potential loyalists and Recent customers. Champions generate a high GMV (1.66bn) despite having fewer customers (10.4K). The number and GMV of loyal customers is **not very impressive** as they have the lowest customer count (7.5K and 0.79bn).

The high number of frequent shoppers with low spending may indicate **problems in product mix**, as evidenced by customers **making small purchase frequently** but **show little interest in additional items**. The low number of loyal customers indicate **declining customer satisfaction** and **ineffective retention strategies**. We should identify the customers who have converted from loyal to "customers needing attention" and try to understand patterns among those who have stopped being loyal. For Champions customers,

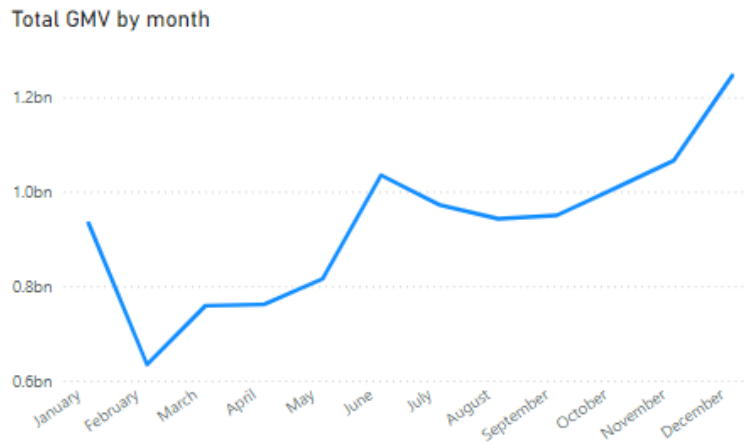
as this is a very important segment we should try to retain and maximize value from this segment by giving them exclusive rewards & benefits or conducting personalized engagement.



Most of our GMV was generated in location 8 (23%) and 4 (19.41%), followed by 61 (4.59%), 65 (4.52%), 511 (3.85%), 31 (3.12%). These are the cities we should focus on as this indicates these markets have high demands, better customer engagement or more effective marketing and sales strategies. For the cities that is not generating good revenue, this indicates they might not be profitabe or they may require different marketing strategy.



Looking at the distribution of each segmentation of our customers by location, we see that the distribution is pretty similar across locations. This indicates that there is little to no difference in customer behavior across location. This means we can cut cost by standardizing marketing strategy for different locations. Insights from one location can be applied to others which will make it easier to develop new products or services.



Overall trend for the shop's revenue in a year can be seen: sales usually drops from Jan to Feb (Lunar New Year might be a reason why) and go back up again and reach a local peak near the beginning of Summer, and highest by the end of the year (Nov – Dec). This means we should focus marketing efforts and promotions during these time of year.

The shop experienced rapid increase in GMV in Q2 2015 and continued to rise well into Q3 2019, but GMV plummeted in Q4 2019 when revenue supposed to be the highest of the year. This is a peculiar pattern and need a thorough analysis into the reason why the shop experienced this big drop in revenue.