

CUSTOMER ANALYSIS REPORT

Prepared by: Huynh Cong Hieu
Email: hch.b.1710@gmail.com

Table of content

I. Project Customer 360 overview

II. Model analysis

1. Customer 360

2. RFM metrics

3. IQR method

4. BGC matrix

III. Data analysis process

1. Data preparation

2. Data model

3. RFM calculation

4. Customer segmentation

5. Visualization and insights

I. Project customer 360 overview

- Customer 360 is an analytics platform centered around customers, encompassing all data to categorize customers into distinct groups based on various attributes for analysis.
- The goal of the Customer 360 project is that, after analysis, we can obtain an overview of customer information. This information is then distributed to the company's departments to develop strategies for marketing, sales, and promotional programs that are effective and cost-efficient, ultimately generating the highest possible revenue.

II. Model analysis

1. Customer 360

Definition: The Customer 360 analysis model is a comprehensive method that helps businesses gain a deeper understanding of their customers by aggregating and analyzing data from various sources. It is a powerful tool in customer relationship management (CRM) strategy, providing businesses with a more holistic and detailed view of each customer segment.



Figure 1: Customer 360

The Customer 360 model is divided into four main groups in data analysis:

- 1. Demographics data:** This includes basic information about customers such as name, address, email, identification number, phone number, education level, and income. This data is used to gain an overview of customer information, understand who they are, and categorize customers by age, income, and education level.
- 2. Transaction data:** This data related to customer transactions, including preferred payment methods (credit card, cash, etc.), transaction history, and spending levels in transactions.
- 3. Interaction data:** This data related to customer interactions, often involving large volumes of big data. For example, customer interactions on TikTok within a day, including shares, comments, and the duration of viewing posts.
- 4. Behavioral data:** This data on customer behavior, derived from comments and feedback on services. This helps in identifying customer preferences, needs, and desires. Notably, this is an amalgamation of transaction and interaction data.

2. RFM metrics

RFM metrics is a method for analyzing and segmenting customers based on three main criteria: Recency (R), Frequency (F), and Monetary (M). This is a popular technique in customer relationship management (CRM) and marketing, helping businesses identify and assess the value of each customer across different segments.

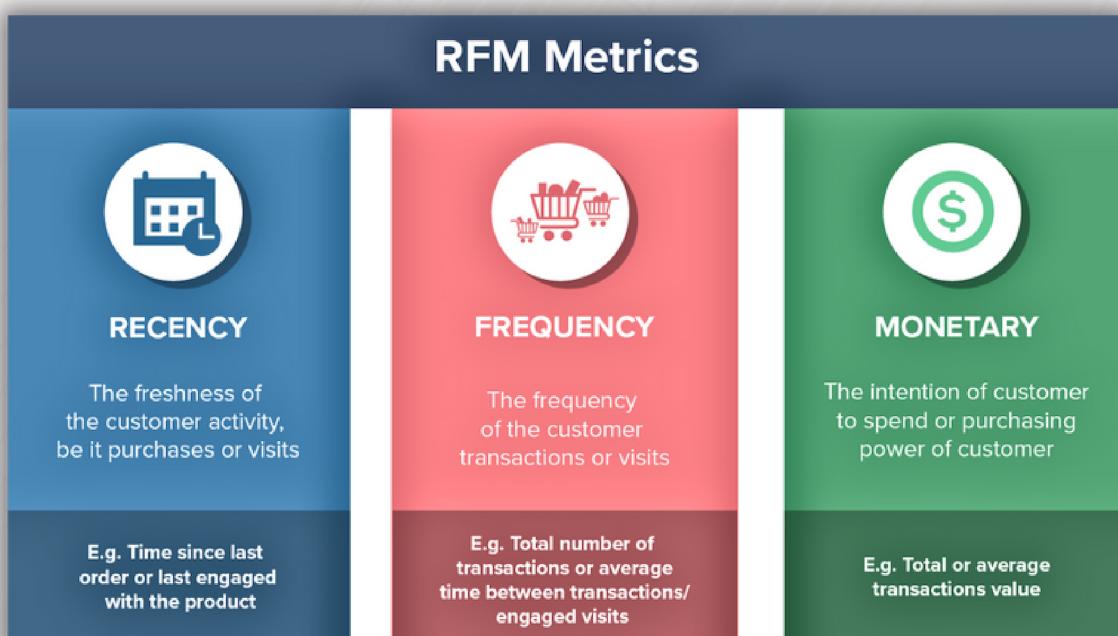


Figure 2: RFM metrics

Definition of the three RFM metrics:

- 1. Recency:** This metric measures the time elapsed since a customer's last purchase up to the present moment. A lower recency score indicates that the customer has a high purchasing need and a low churn rate.
- 2. Frequency:** This metric measures the number of transactions or purchases a customer has made within a specific period. A higher frequency score suggests that the customer has a high transaction rate, is likely to become a loyal customer, and will contribute significant value to the business in the future.
- 3. Monetary:** This metric measures the total amount of money each customer has spent across all transactions. A higher monetary score indicates a potential high-value customer who has spent a significant amount and is likely to return for more purchases if their buying frequency is high."

3. IQR method

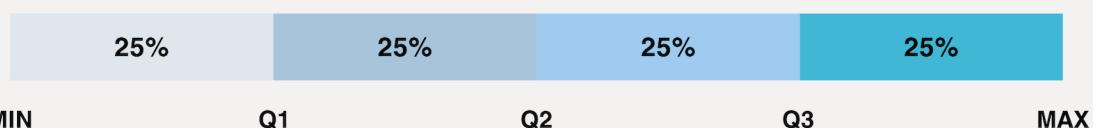


Figure 3: IQR method

IQR abbreviated as 'interquartile range,' is the distance between the 25th percentile and the 75th percentile of a data set. We apply the IQR method in this case to divide the data into four equal parts for scoring customers based on their RFM scores, as follows:

- MIN to Q1: 1 point
- Q1 to Q2: 2 points
- Q2 to Q3: 3 points
- Q3 to MAX: 4 points

Method for determining IQR:

1. Sort the data in ascending order.
2. Determine the positions of the 25th percentile (Q1) and the 75th percentile (Q3):
 - Q1 (First Quartile): The value at the 25th percentile of the data set.
 - Q3 (Third Quartile): The value at the 75th percentile of the data set.

4. BGC matrix

The BCG matrix is a business model analysis tool that helps companies evaluate and categorize their strategic business units (SBUs) or products based on two main factors: market growth rate and relative market share.

Additionally, in data analysis, we can apply the BCG matrix along with domain knowledge to analyze and segment customers. This allows us to classify each customer according to their RFM score, thereby clearly identifying the characteristics and behaviors of each customer type to develop optimal marketing strategies and customer care methods



Figure 3: BGC matrix

The BCG matrix is divided into four main areas (Stars, Question Marks, Cash Cows, Pets), each corresponding to the characteristics and growth potential of different customer groups for a business

BGC metrix table

BGC metrix	Description	Mapping group	RFM score
Stars	Customers with high recency, frequency, and monetary scores generate significant value and revenue for the business.	Champions	343, 333, 433, 443, 434, 444, 344, 334
Cash cows	Customers with high frequency, who are still using services and products but do not have the same monetary value as Stars	Loyalty customer	422, 332, 423, 323, 313, 413, 233, 424, 324, 224, 244, 234, 314, 414
Question marks	Recently emerged customers generate moderate revenue for the business but may leave at any time.	Potential customer	222, 322, 442, 242, 342, 432, 232, 223, 243, 341, 441
Dogs (Pets)	Customers with poor recency, frequency, and monetary metrics, indicating low engagement and interest in the business's products and services	Bad customer	The rest

II. Data analysis process

1. Data preparation

The following customer data is collected from the internal CRM system, with the data timeframe from June 1, 2022, to August 1, 2022, comprising two tables: Customer_register and Customer_Transaction.

Table name	Description
Customer_register	This table contains information related to customer contract registration.
Customer_Transaction	This table contains information related to customer contract transactions.

Table Customer_Registered

Tên cột	data_type	định nghĩa
ID	bigint	Customer ID
Contract	varchar(50)	Contract ID
LocationID	int	Position ID
BranchCode	tinyint	Branch ID
Status	tinyint	Status
created_date	datetime	Registration date
stop_date	datetime	Contract termination date

123 ID	abc Contract	123 LocationID	123 BranchCode	123 Status	⌚ created_date	⌚ stopdate
16	SGDN00017		0	0	2017-12-05 00:00:00	2017-12-06 00:00:00
17	SGH068462		0	0	2017-12-06 00:00:00	2017-12-06 00:00:00
18	SGH068460		0	0	2017-12-06 00:00:00	2017-12-06 00:00:00
19	SGDN00017		0	0	2017-12-06 00:00:00	2018-01-10 00:00:00
20	SGH068462		0	0	2017-12-06 00:00:00	2017-12-07 00:00:00
21	SGH068462		0	0	2017-12-07 00:00:00	2017-12-07 00:00:00
22	SGH068462		0	0	2017-12-07 00:00:00	2017-12-07 00:00:00
23	HNFDN0025		0	0	2017-12-07 00:00:00	2018-05-29 00:00:00
24	SGH068460		0	0	2017-12-07 00:00:00	2017-12-07 00:00:00
25	SGH068460		0	0	2017-12-07 00:00:00	2017-12-07 00:00:00
26	SGH068460		0	0	2017-12-07 00:00:00	2018-01-03 00:00:00

Table Customer_Transaction

Column names	data_type	Definition
ID	bigint	Transaction ID
CustomerID	varchar(50)	Customer ID
Purchase_Date	datetime	Transaction date
GMV	bigint	Gross Monetary Value

123 Transaction_ID	123 CustomerID	⌚ Purchase_Date	123 GMV
0	1,327,813	2022-06-01 00:00:00	95,000
1	1,157,830	2022-06-01 00:00:00	75,000
2	873,915	2022-07-01 00:00:00	95,000
3	3,505,071	2022-07-01 00:00:00	90,000
4	2,930,918	2022-07-01 00:00:00	109,091
5	899,882	2022-06-01 00:00:00	105,000

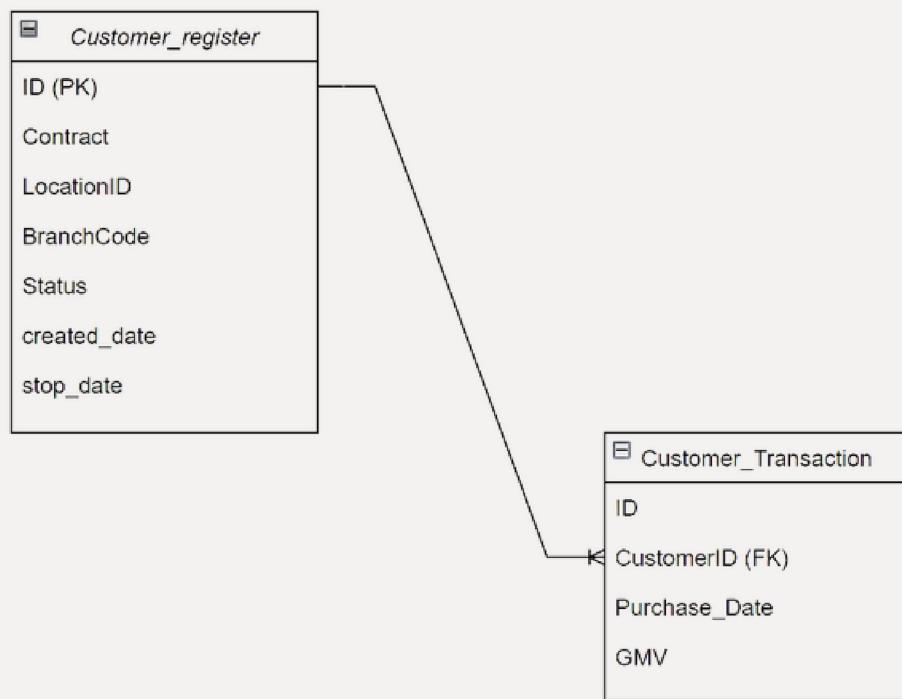
Data processing

- Delete rows where ID = 0.
- Delete rows where GMV = 0.
- Delete rows where Created_date = 0 or is empty

Total customers	Total transactions	Total GMV
113.773	126,849	22,248,522,396

After filtering the data, we have **113,773 customers** and **126,849 transactions** with a **total expenditure of 22,248,522,396 VND**.

2. Data model



The Customer register table has a primary key 'ID' linked to the Customer transaction table with a foreign key 'CustomerID'. Therefore, we can delve deeper into customer transaction information through the Customer transaction table, including transaction ID, purchase date, and total amount spent by customers.

3. RFM caculation

Based on the IQR method mentioned above, we calculate the quartiles: Q1, Q2, Q3, Q4, and assign scores to each group as follows:

To determine Recency, Frequency, and Monetary for each customer in the Customer 360 dataset

- **Recency** = Recency = datediff(day, MAX(Purchase_date), '2022-09-01'), the number of days between '2022-09-01' and the most recent Purchase_date for each customer.
- **Frequency** = count(Purchase_date), total number of transactions or service usages by a customer
- **Monetary** = sum(GMV), total amount spent by each customer on services
- However, to calculate R, F, M more accurately and avoid bias, we need to consider the customer's tenure (customer age) using their contract duration.
- Calculation of customer contract age: DATEDIFF(year, created_date, '2022-09-01'): the number of years from the current date '2022-09-01' minus the year the contract was created for each customer.
- Then, use Frequency and Monetary divided by the year the contract = customer age.

customer_id	recency	frequency	monetary	RFM_score	segmentation
76658	62	0.14	13571	344	VIP
76675	62	0.14	17857	343	VIP
76692	31	0.14	15000	444	VIP
76755	62	0.14	22727	342	Potential customer
76770	31	0.14	14155	444	VIP
76782	92	0.14	15000	244	Loyalty customer
76812	31	0.29	30000	411	Bad customer
76845	62	0.14	15000	344	VIP
76900	31	0.14	20714	442	Potential customer
76910	31	0.14	15000	444	VIP
77015	92	0.14	17857	243	Potential customer
77067	31	0.14	15000	444	VIP
77089	92	0.14	15000	244	Loyalty customer
77096	31	0.14	13571	444	VIP
77127	62	0.14	17857	343	VIP
77132	62	0.14	25714	342	Potential customer
77135	31	0.29	25829	412	Bad customer
77560	92	0.14	20714	242	Potential customer
77609	92	0.29	38311	211	Bad customer
77639	62	0.29	34155	311	Bad customer

4. Customer segmentation

After calculating the R, F, M scores, we will categorize customers based on their scores and calculate the total number of customers and total expenditure within each segment.

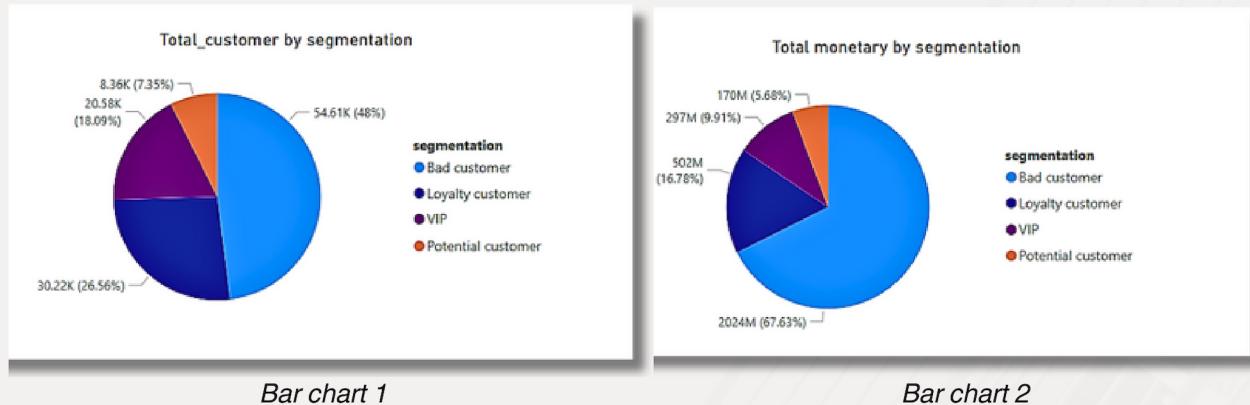
Table customer segmentation

Mapping group	RFM score	Total customer	Total GMV
VIP	343, 333, 433, 443, 434, 444, 344, 334	20,577	296,634,207
Loyalty customer	422, 332, 423, 323, 313, 413, 233, 424, 324, 224, 244, 234, 314, 414	8,362	169,948,087
Potential customer	222, 322, 442, 242, 342, 432, 232, 223, 243, 341, 441	30,221	502,204,681
Bad customer	The rest	54,613	2,024,462,647

ABC segmentation	123 total_customer	123 total_monetary
VIP	20,577	296,634,207
Potential customer	8,362	169,948,087
Loyalty customer	30,221	502,204,681
Bad customer	54,613	2,024,462,647

5. Visualization and insights

Data overview



Bar chart 1

Bar chart 2

Chart 1 illustrates that currently, bad customers account for nearly half (48%) of the total across the four customer segments. This reflects a relatively high churn rate, possibly due to perceived inadequacies in the products or services they are using, or poor customer care services.

Additionally, **Chart 2** shows that the highest total expenditure comes from **bad customers (67.63%)**, despite their low purchase frequency or prolonged absence. This suggests that although they make fewer transactions, each transaction involves a significant amount of spending. The reason could stem from inadequate customer care services, potentially leading them from being potential or loyal customers to becoming transient customers.

Following bad customers are **VIP customers (9.91%)**, **Loyalty customers (16.78%)**, and **Potential customers (5.68%)**. These three groups currently contribute nearly equal expenditure, indicating the need for sales departments to focus on providing better care to reduce churn rates among these groups.

General Solution:

- Improvements are needed in services and products to enhance customer experience and reduce churn rates among transient customers, who spend significantly but may not return due to various reasons.
- Launch promotional campaigns and incentives targeting VIP, Loyalty, and Potential customers, as they exhibit higher purchase frequency despite lower spending per transaction.

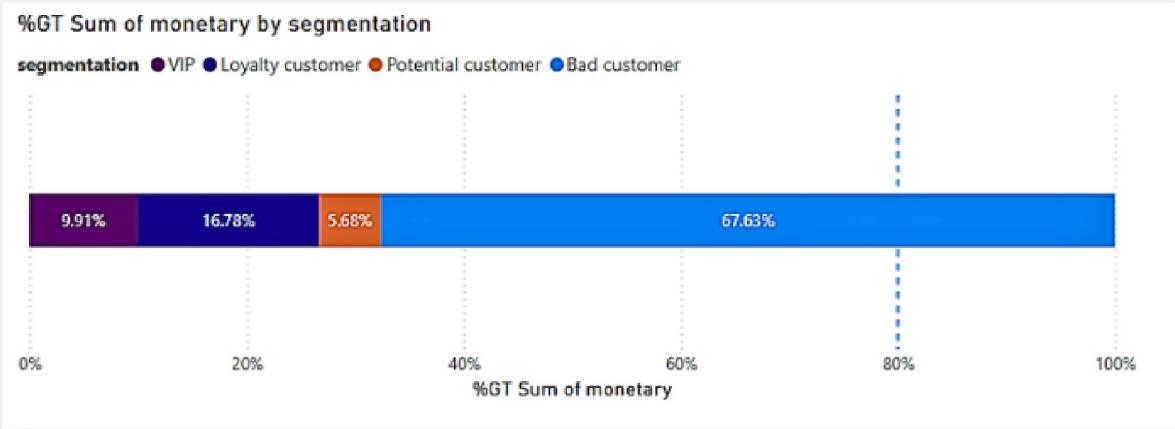


Chart 3

According to Pareto's principle as shown in the chart above, we can observe that the top 20% of customers currently do not generate 80% of the revenue for the business. This indicates that the business has not fully tapped into the potential of its best customers.

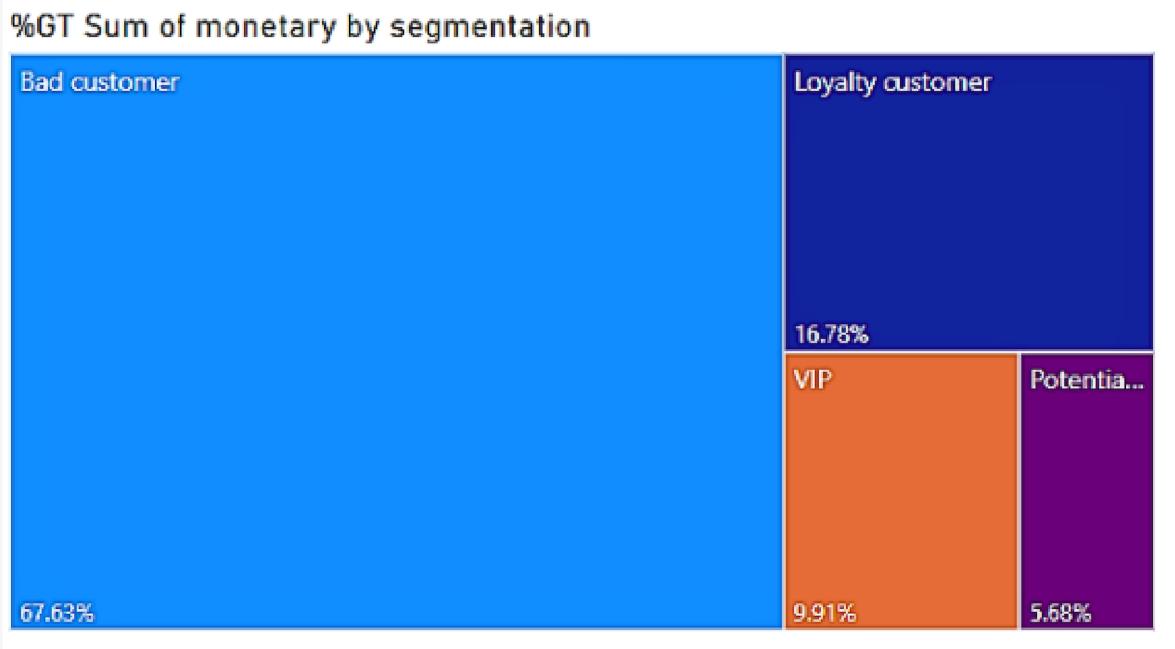


Chart 4

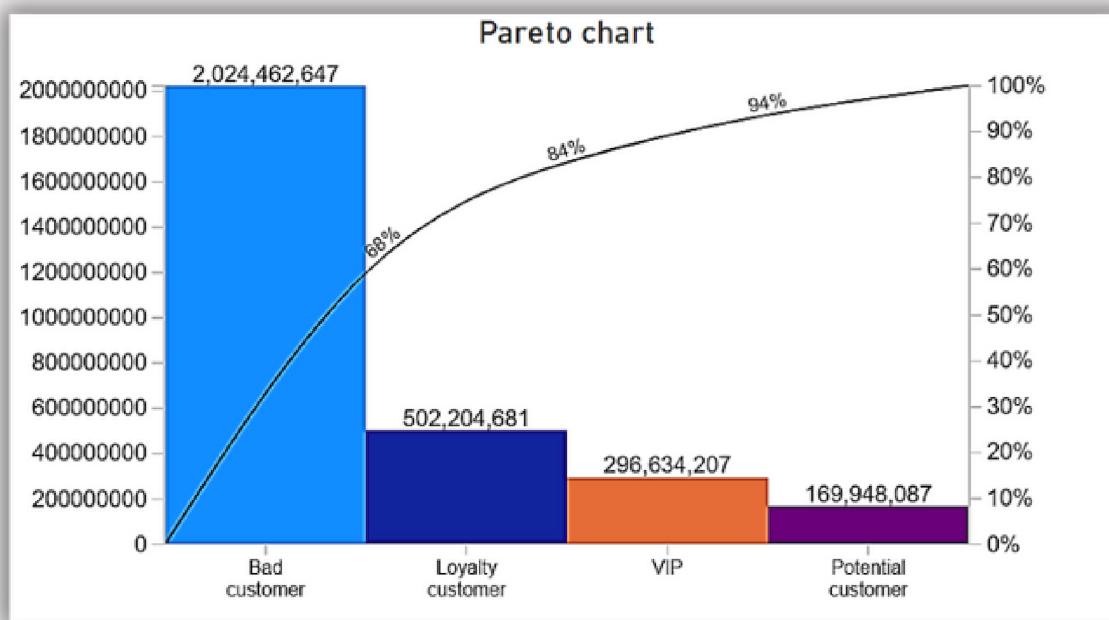


Chart 5

From charts 4 and 5, we can observe the distribution of revenue across each customer segment, specifically noting that 80% of the revenue comes from **Bad customers (67.63%)** and **Loyalty customers (16.78%)**, with the majority falling under the Bad customer category. This result indicates that the main revenue stream comes from transient customers, which is a concerning sign as these customers only provide temporary revenue without sustainability.

This high churn rate suggests issues such as misaligned customer targeting and ineffective previous promotional and marketing campaigns.

Solution:

- The sales department should allocate resources and provide better care for **Loyalty, VIP, and Potential customers**, focusing on understanding why their spending is not higher. This understanding will help in devising solutions and improving products to boost sales.
- Reevaluate the outcomes of previous marketing campaigns and promotions to draw conclusions and make improvements going forward.
- Improving customer retention rates is crucial for ensuring long-term business sustainability, maintaining stable revenue streams, and supporting operational and developmental activities.