

RFM Analysis of a Pharmacy Retail Chain's Customer Base

Task

🎯 The company has an SMS-mailing system, but it sees no point in wasting the budget and instead wants to launch a personalized campaign.

📊 The purpose of the research:

- *to classify customers* – since it makes no sense to offer personalized discounts aimed at increasing the average purchase amount to a non-target audience (e.g., those who have only made a single low-value purchase).
- *to develop recommendations* – specifying what kind of offers should be made to each customer group.

This task can be addressed by conducting an **RFM analysis**.

RFM analysis is a method of evaluating customers based on their purchases.

We look at how often people buy, how recently they made their last purchase, and how much money they spend.

RFM - definition

Recency: how recently the customer made their last purchase

Frequency: how often the customer makes purchases

Monetary: how much money the customer spends

The essence of RFM analysis is to divide customers into groups depending on how recently they made a purchase, how frequently they buy, and how much they spend. Then we assign each group a category marker from 1 to 3 (1 – low, 2 – medium, 3 – high).

Recency	Frequency	Monetary
1 - a long time ago	1 - rarely	1 - low spend
2 - an average time ago	2 - infrequently	2 - medium spend
3 - recently	3 - frequently	3 - high spend

The Structure of the Database

The Bonuscheques table from a retail chain database was used in this research. The table contains information about transactions made using bonuscards (i.e. regular customers who give their personal information/phone numbers, emails etc.)

If the cash register was offline at the moment of purchase then an encrypted sequence of characters was recorded instead of the bonus card number. In such cases, the card number

cannot be recovered using the available database. Therefore, only purchases that could be identified were included in the analysis.

Customers with only one purchase were excluded from the analysis, since offering a personalized discount wouldn't help increase their average spend.

The Choice of Metrics for Analysis

To determine the markers boundaries, the distribution of the values was plotted and the percentiles, mean and median values were calculated.

Recency:

The value distribution indicates that the range boundaries can be set at 28 and 86. Additionally, the percentiles distribution shows that 30% of customers made purchases within 25 days. This group is classified as “Fresh” and is assigned a marker 3. The boundary for the next segment is set at the 60th percentile - value 85. The average (85) and median (61) values fall into this middle segment, which is assigned a marker 2. Customers who made a purchase more than 85 days ago are classified as “long inactive” and are assigned a marker 1.



Recency_avg_median



Frequency:

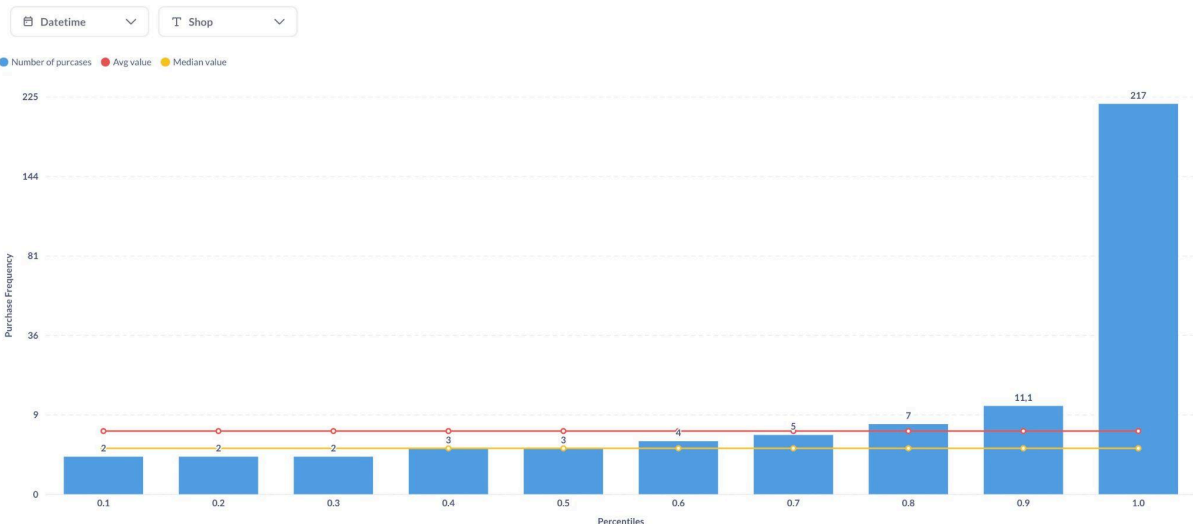
Quite a lot of customers made only 2 purchases (since those who made just 1 purchase were excluded). This corresponds to the 30th percentile - we assign them marker 1.

This can be observed on the purchase frequency distribution graph.

Visually, we can define the second group as those who made between 3 and 6 purchases - marker 2. The median (3) and the average (5.68, which we round to 6) fall within this range. Customers who made more than 7 purchases fall into the 80th percentile - we assign them marker 3.



Frequency_avg_median



Monetary (Amounts shown in units):

We segmented customers based on the spread of purchase values, using percentiles to determine the thresholds.

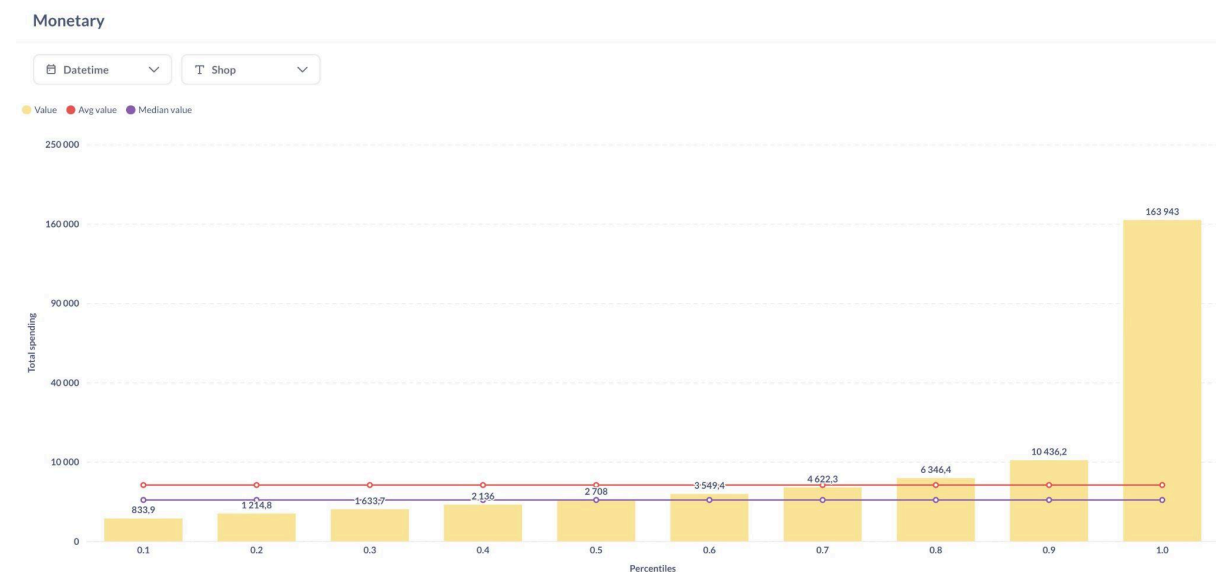
Customers who spent less than 2,136 units are assigned to marker 1 (40th percentile - the smallest purchase amounts).

Those who spent between 2,136 and 6,346 units fall into group 2.

20% of customers spent 6,345 monetary units or more (80th percentile) - these are considered “VIP” customers and we assign them marker 3.

The average purchase amount (5,081) and the median (2,708) both fall into group 2.

The large gap between the average and the median indicates the presence of outliers.



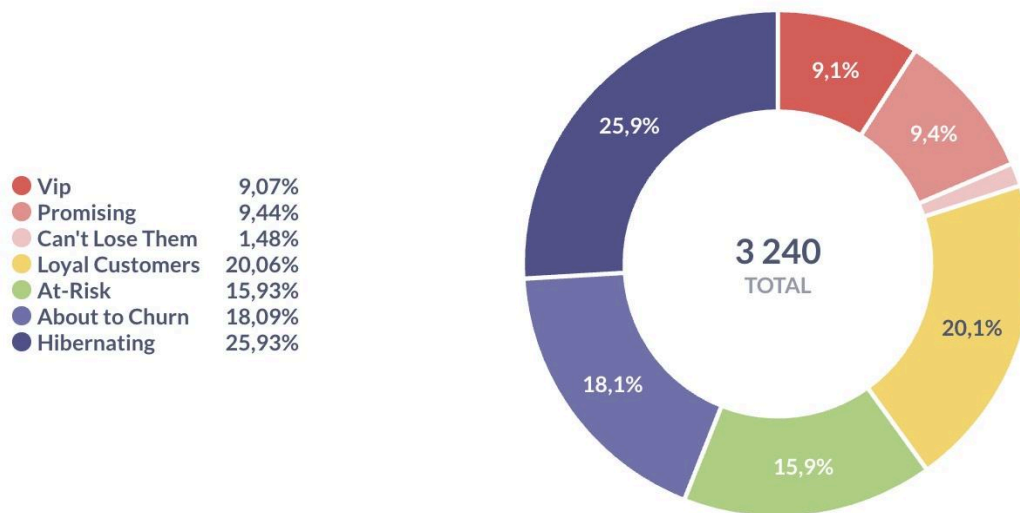
Marker	Recency (in days)	Frequency (number of purchases)	Monetary (Units)
1	over 84	2 purchases	up to 2136
2	up to 83	from 3 to 6	from 2137 to 6345
3	up to 25	7 and more purchases	6346 and more

All customers were divided into 27 segments, with the number of customers per segment ranging widely - from a maximum of 428 to a minimum of 3.

	A-Z rfm	123 Number of customers	A-Z Description (Recency, Frequency, Monetary)
1	111	428	a long time ago, rarely, low spend
2	112	157	a long time ago, rarely, medium spend
3	113	7	a long time ago, rarely, high spend
4	121	223	a long time ago, infrequently, low spend
5	122	316	a long time ago, infrequently, medium spend
6	123	56	a long time ago, infrequently, high spend
7	131	3	a long time ago, frequently, low spend
8	132	50	a long time ago, frequently, medium spend
9	133	48	a long time ago, frequently, high spend
10	211	186	a while ago, rarely, low spend
11	212	66	a while ago, rarely, medium spend
12	213	5	a while ago, rarely, high spend
13	221	181	a while ago, infrequently, low spend
14	222	244	a while ago, infrequently, medium spend
15	223	50	a while ago, infrequently, high spend
16	231	9	a while ago, frequently, low spend
17	232	83	a while ago, frequently, medium spend
18	233	146	a while ago, frequently, high spend
19	311	148	recently, rarely, low spend
20	312	56	recently, rarely, medium spend
21	313	3	recently, rarely, high spend
22	321	114	recently, infrequently, low spend
23	322	202	recently, infrequently, medium spend
24	323	39	recently, infrequently, high spend
25	331	5	recently, frequently, low spend
26	332	121	recently, frequently, medium spend
27	333	294	recently, frequently, high spend

Since our task is to optimize the mechanism of SMS mailings, it does not make sense for the marketing department to work with a group of 3 people, so groups with a small number of customers are combined into enlarged groups on a similar principle of work.

After distribution 7 groups were formed.



To identify customer value we examine the percentage distribution by groups using the Pareto Principle (80/20 rule) - an empirical rule that in its most general form, states that "20% of efforts generate 80% of the results, while the remaining 80% of efforts produce only 20% of the results."

We will use this Principle as a baseline approach for analyzing the factors influencing our performance and optimizing its outcomes: by correctly identifying a minimal set of the most impactful actions, we can quickly achieve a significant portion of the intended overall result, while further improvements may be inefficient or unjustified.

The analysis allowed us to segment customers into groups and identify the most effective approach for each group.

Our goal is to retain the best customers, convert loyal ones into potential VIPs and encourage less active customers to move toward loyalty.

In other words, we aim to improve at least one marker parameter of the group: from level 1 → 2, from level 2 → 3.

Final RFM table with assignment of category and group to each client

card	Days without purchase	R	Number of Purchase	F	Purchase amount	M	RFM	Customer_group
2000200150053	84	1	2	1	713	1	111	Hibernative
2000200150091	4	3	5	2	3 549	2	322	Loyal Customers
2000200150107	7	3	3	2	1 735	2	322	Loyal Customers
2000200150145	21	3	14	3	12 426	3	333	VIP
2000200150237	45	2	3	2	2 772	2	222	Loyal Customers
2000200150367	30	2	16	3	11 064	3	233	Promising
2000200150381	39	2	8	3	3 803	3	233	Promising
2000200150398	7	3	5	2	4 118	3	323	Promising
2000200150411	6	3	29	3	90 386	3	333	VIP
2000200150428	36	2	7	2	13 419	3	223	Loyal Customers
2000200150442	65	2	2	1	667	1	211	Hibernative
2000200150459	318	1	2	1	2 152	2	112	About to Churn
2000200150466	46	2	2	1	937	1	211	Hibernative
2000200150480	7	3	2	1	3 639	3	313	At Risk
2000200150497	15	3	7	2	17 183	3	323	Promising
2000200150589	95	1	5	2	5 056	3	123	About to Churn

Recommendations

Let's take a closer look at each group and provide recommendations.

In our case we assume that 20% of the customers generate 80% of the profit, which means our key groups should consist of approximately 648 clients (20% of 3,240).

It turns out that the main revenue is generated by VIPs (333) - 294 customers and Promising with the potential to become VIPs - 306 customers, which makes a total of 600 customers (about 18.5%), already close to 20%.

RFM	Number of customers	Description (Recency, Friquency, Monitary)	Customer group
333	294	recently, frequently, high spend	VIP

VIP group

🎯 The main goal for the VIP group (9,07%) is to ensure they remain in this group. The goal for the Promising group is to offer targeted promotions aimed at improving their 'weaker' metrics - increasing average spend and purchase frequency.

✅ It is important for these groups to show their value, so we can offer personal discounts or make exclusive offers in gratitude for loyalty, but it is important not to be annoying so that the offers are not perceived as intrusive.

Can't Lose Them group (1,48%) - These are customers who haven't made a purchase in a long time, but were previously active, they used to buy frequently with medium or high spend.

RFM	Number of customers	Description (Recency, Frequency, Monetary)	Customer group
133	48	a long time ago, frequently, high spend	Can't Lose Them

🎯 The goal for this group is to bring the customer back to making purchases.

✅ A good motivator for returning customers can be personalized bonuses - such as a discount on their next purchase or exclusive 'just for you' offers, ideally with a limited validity period to encourage a quicker return.

The Loyal Customers group (20,06%) - this is quite a large group consisting of customers who made purchases with an average recency, either frequently or infrequently and with a medium or large check.

RFM	Number of customers	Description (Recency, Frequency, Monetary)	Customer group
232	83	a while ago, frequently, medium spend	Loyal Customers
223	50	a while ago, infrequently, high spend	Loyal Customers
222	244	a while ago, infrequently, medium spend	Loyal Customers
213	5	a while ago, rarely, high spend	Loyal Customers
322	202	recently, infrequently, medium spend	Loyal Customers
212	66	a while ago, rarely, medium spend	Loyal Customers
	650		

🎯 The goal for this group is customer retention and increase their average transaction value.

✅ Key purchase drivers here can be Loyalty Programs (the more purchases, the higher the discounts) or personalized offers based on purchase history.

The At Risk group (15.93%) have purchased recently, but they buy infrequently with a middle or low spend.

RFM	Number of customers	Description (Recency, Frequency, Monetary)	Customer group
312	56	recently, rarely, medium spend	At Risk
331	5	recently, frequently, low spend	At Risk
311	148	recently, rarely, low spend	At Risk
321	114	recently, infrequently, low spend	At Risk
221	181	a while ago, infrequently, low spend	At Risk
231	9	a while ago, frequently, low spend	At Risk
313	3	recently, rarely, high spend	At Risk
	516		

🎯 The goal is to increase average spend and purchase frequency.

✅ This group can be encouraged to purchase by offering more attractive deals ('buy more - get a gift') or by sending targeted mailings with promotional items.

The Hibernating group (18,09%) has not purchased for a long time, visits infrequently, but has a medium or high spend.

RFM	Number of customers	Description (Recency, Friquency, Monitary)	Customer group
112	157	a long time ago, rarely, medium spend	Hibernating
113	7	a long time ago, rarely, high spend	Hibernating
132	50	a long time ago, frequently, medium spend	Hibernating
122	316	a long time ago, infrequently, medium spend	Hibernating
123	56	a long time ago, infrequently, high spend	Hibernating
	586		

🎯 We need to prevent the loss of customers.

✅ It would be great to get feedback on why customers visit rarely - for this you can send information on promotions along with a short questionnaire with a bonus for participation, offer a discount on the next purchase or make a personal offer.

The About to Churn group (25,93%) purchased a long time ago, rarely buys with a low spend.

RFM	Number of customers	Description (Recency, Friquency, Monitary)	Customer group
121	223	a long time ago, infrequently, low spend	About to Churn
111	428	a long time ago, rarely, low spend	About to Churn
211	186	a while ago, rarely, low spend	About to Churn
131	3	a long time ago, frequently, low spend	About to Churn
	840		

This is a large group of mostly one-time customers. On the other hand, these customers have left their contact details (they are in the system), but they do not consider this pharmacy as their regular place of purchase.

🎯 Perhaps the reason for this is the low awareness of the customers about the range of products, so it is necessary to provide more information about the range of products and promotions.

✅ This is the most challenging group, and attempts to win them back can be costly. You can try to re-engage customers with large discounts, promotions, or gifts, using various methods. Then, they can be segmented into 'truly lost' (have not purchased for a very long time) and 'potentially recoverable'. For the former, test SMS campaigns with significant discounts, but only if they respond.

Conclusion

In conclusion, I would like to note that from time to time you need to review the group and do the RFM analysis again. How often depends on how dynamic your base is.

If you are a large online store with a huge number of visitors and frequent purchases, then you can update the analytical data once a month or two. If customers rarely make orders, updating the RFM analysis once a quarter or even once every six months will be enough.

The point of the new RFM analysis group is to see which of your customers have changed theirs in order to effectively apply activation methods.

In general, RFM analysis is suitable for everyone, the main thing is to have information about purchases and a sufficient Customer Base.