

RFM-анализ клиентской базы аптечной сети

Метрики

Для анализа проведения персонализированной рассылки в нашем случае лучше всего подходит RFM-анализ. Он дает комплексное представление о поведении клиента.

Для проведения RFM- анализа были выбраны 3 основные метрики, свойственные RFM - это:

- Recency (Давность): как давно клиент совершил последнюю покупку. Ключевой показатель вовлеченности покупателя. Клиент, который недавно покупал лекарства с большей вероятностью отреагирует на письмо о скидках, на сегмент товаров связанных с его покупкой, чем человек, который покупал 300 дней назад.
- Frequency (Частота): как часто клиент совершает покупки. Служит показателем привычки у покупателя. Она позволяет точно отсеять клиентов случайно забежавших в аптеку за одним лекарством, от постоянных клиентов, на которых стоит сделать акцент. Помогает не тратить бюджет на разовых клиентов.
- Monetary (Деньги): общая сумма покупок клиента. Является главным показателем экономической ценности. Она помогает правильно определить какому клиенту дать большую скидку и необходима для принятия экономически обоснованных решений. Предлагать клиенту купившему лекарства на небольшую сумму высокую скидку - экономически невыгодно.

Также для разделения на группы клиентов был проведен ABC анализ по сумме покупок, который позволил выделить сегмент «A» — клиентов, генерирующих 80% всей выручки. Были найдены медианы, среднее, перцентили. Это позволило нам определить естественные, обоснованные границы между группами, опираясь на реальное поведение клиентской базы.

Гипотезы

- Значительная часть клиентов - это случайные покупатели, купившие один раз "аскорбинку".
- Есть группа лояльных, постоянных клиентов, которые дают большую часть выручки, которые являются ценными для компании. По принципу Парето это 20% от всей клиентской базы.
- Подавляющая часть - это средние клиенты, которые не ходят каждую неделю, но и не забывают на годы.
- Существует сегмент клиентов, которые раньше были лояльными, но сейчас на грани ухода.

Что нужно сделать?

Для проведения эффективных рассылок и оптимизации бюджета нам нужно отличать ценных, лояльных от случайных, неактивных. Это классическая задача поведенческой сегментации для решения , которой лучше всего подходит RFM анализ.

Изучаем данные и выбираем столбцы по которым будем проводиться RFM анализ.

Проверяем данные на повторяющиеся строки, неправильные записи номера бонусной карты, затем для каждого уникального клиента рассчитываем RFM, выбираем по каким столбцам будем проводить. Делаем анализ всех данных для определения сегментации клиентов. Дальше стоит вопрос как правильно классифицировать на самых ценных, лояльных, требующих внимания, спящих. Далее определяем какие клиенты в какой группе и какое их количество для проверки гипотез. Делаем выводы какие сообщения отправлять каждой группе клиентов.

Замечание.

Есть аномалия в данных при таком запросе 21073 строк

```
select distinct doc_id  
from bonuscheques  
where card ~ '^[0-9]+$'
```

А при таком 21075

```
select distinct doc_id  
from bonuscheques  
where card ~ '^[0-9]+$'
```

2022-01-06 12:57:28.000	Аптека 10	200020020821166	0	2207	2207	
15#17000057#66115#17_17						
2022-01-06 12:57:45.000	Аптека 10	200020020821166	0	2207	2207	
15#17000057#66115#17_17						
2022-01-18 17:21:49.000	Аптека 10	2000200189985	19	21	771	771
15#17000300#66127#17_51						
2022-01-18 17:26:16.000	Аптека 10	2000200188902	11	0	371	371
15#17000300#66127#17_51						

Также если в момент покупки касса была в оффлайн-режиме, то вместо номера карты записывается зашифрованная последовательность символов.

Анализ метрик

Подготавливаем данные, изучаем их для корректной сегментации клиентов.

ABC - анализ по сумме покупок

Основная цель этого запроса - сегментировать клиентскую базу по их экономической ценности, чтобы определить самых прибыльных клиентов.

Сначала мы подготавливаем данные отбираем нужные чеки с корректными номерами карт. Затем считаем общую сумму покупок для каждого уникального клиента и с помощью оконных функций рассчитываем долю вклада в общую выручку, сортируя клиентов по убыванию трат и присваиваем сегмент. И далее находим максимальную и минимальную сумму для каждой группы.

```
with sortbonuscheques as (
select
    distinct on
        (doc_id)
    card,
    datetime,
    summ
from
    bonuscheques
where
    card ~ '^[0-9]+\$',1
abc as(select card, sum(summ) as monetary
from sortbonuscheques
group by card),
a as (select
    card,
    monetary,
```

```

case
when sum(monetary) over (order by monetary desc)/
sum(monetary) over() <= 0.7 then 'A'
when sum(monetary) over (order by monetary desc)/
sum(monetary) over() <= 0.9 then 'B'
else 'C'
end m
from abc)
select
m as segment,
max(monetary) as max_monetary,
min(monetary) as min_monetary
from
a
group by
m
order by
m

```

Запрос вернул таблицу со стоками, определяющими денежные границы для каждого сегмента.

	A-Z segment	123 max_mon	123 min_mon
1	A	163 943	3 494
2	B	3 492	1 322
3	C	1 321	21

Получены конкретные пороги для RFM monetary. Сегмент C четко очерчивает группу клиентов с наименьшей экономической ценностью. Мы видим, что сегмент A, приносящий 70% дохода, имеет достаточно высокий порог входа. В данном анализе мы используем порог 0,7, создавая более эксклюзивную группу, а порог 0,9 для большей экономии бюджета, исключая большее количество случайных покупателей из дорогих рассылок.

Средние, медианы и перцентили 25-й и 75-й

Хотим рассчитать показатели разброса и тенденции. Эти данные позволяют оценить, что является низким, стандартным и высоким показателем для метрики.

Вначале подготавливаем данные, находим recency, frequency, monetary для уникального клиента и вычисляем четыре ключевые статистические метрики для каждой метрики.

1. Monetary

```

with sortbonuscheques as (
select

```

```

distinct on
(doc_id)
card,
datetime,
summ
from
bonuscheques
where
card ~ '^[0-9]+\$',  

rfm as(select card, count(*) as frequency, ('2022-06-10':: date - max(datetime) :: date) as
recency, sum(summ) as monetary
-- последняя дата 2022-06-09
from sortbonuscheques
group by card),
rfm_stats as (
select
    AVG(monetary) as avg_monetary,
    PERCENTILE_CONT(0.25) within group (
order by
    monetary) as percentile_25,
    PERCENTILE_CONT(0.50) within group (
order by
    monetary) as median_monetary,
    PERCENTILE_CONT(0.75) within group (
order by
    monetary) as percentile_75
from
rfm
)
select
stats.avg_monetary as avg,
stats.percentile_25 as percentile_25,
stats.median_monetary as median,
stats.percentile_75 as percentile_75
from
rfm_stats stats

```

Также для проверки правильности сегментации по сумме покупок были найдены эти метрики, которые дали такие показатели, что подтверждает выбор критериев в ABC-анализе.

Мы видим, что есть большая разница между средним и медианным значением. Это говорит нам о том, что есть покупатели, совершающие покупки на очень большие суммы, которые утягивают среднее значение вверх. Медиана показывает какое количество денег тратит стандартный клиент.



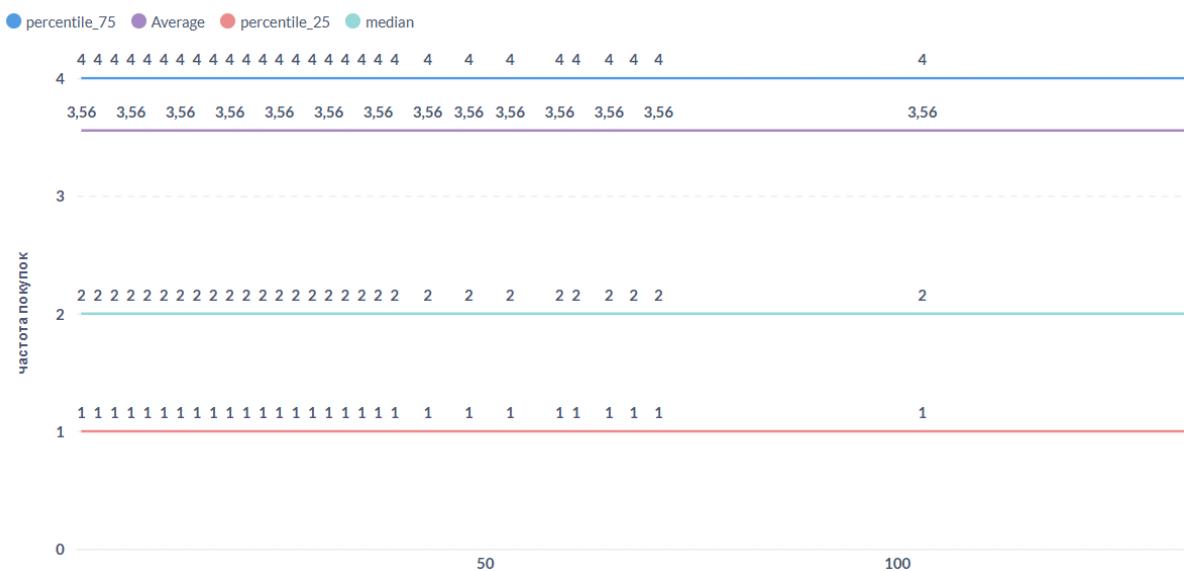
2. Frequency

```

with sortbonuscheques as (
select
    distinct on
        (doc_id)
    card,
    datetime,
    sum
from
    bonuscheques
where
    card ~ '^[0-9]+\$'),
rfm as(select card, count(*) as frequency, ('2022-06-10)::date - max(datetime) :: date) as recency,
sum(sum) as monetary
from sortbonuscheques
group by card),
rfm_stats as (
select
    AVG(frequency) as avg_frequency,
    PERCENTILE_CONT(0.25) within group (
order by
    frequency) as percentile_25,
    PERCENTILE_CONT(0.50) within group (
order by
    frequency) as median_frequency,
    PERCENTILE_CONT(0.75) within group (
order by
    frequency) as percentile_75,
    PERCENTILE_CONT(0.90) within group (
order by
    frequency) as median_recency,
    PERCENTILE_CONT(0.95) within group (
order by
    frequency) as percentile_recency
)
)
```

```
PERCENTILE_CONT(0.75) within group (0.75)
order by
    frequency) as percentile_75
from
    rfm
)
select
    stats.avg_frequency as avg,
    stats.percentile_25 as percentile_25,
    stats.median_frequency as median,
    stats.percentile_75 as percentile_75
from
    rfm_stats as stats
```

На основе получившихся данных, мы видим, что гипотеза о том, что значительная часть нашей клиентской базы - это одноразовые покупатели. Среднее значение значительно выше, чем медианное, что указывает на наличие часто покупающих клиентов, которые утягивают среднее значение вверх. Чтобы войти в топ 25% самых частых покупателей нужно совершить более 4 покупок.



3. Recency

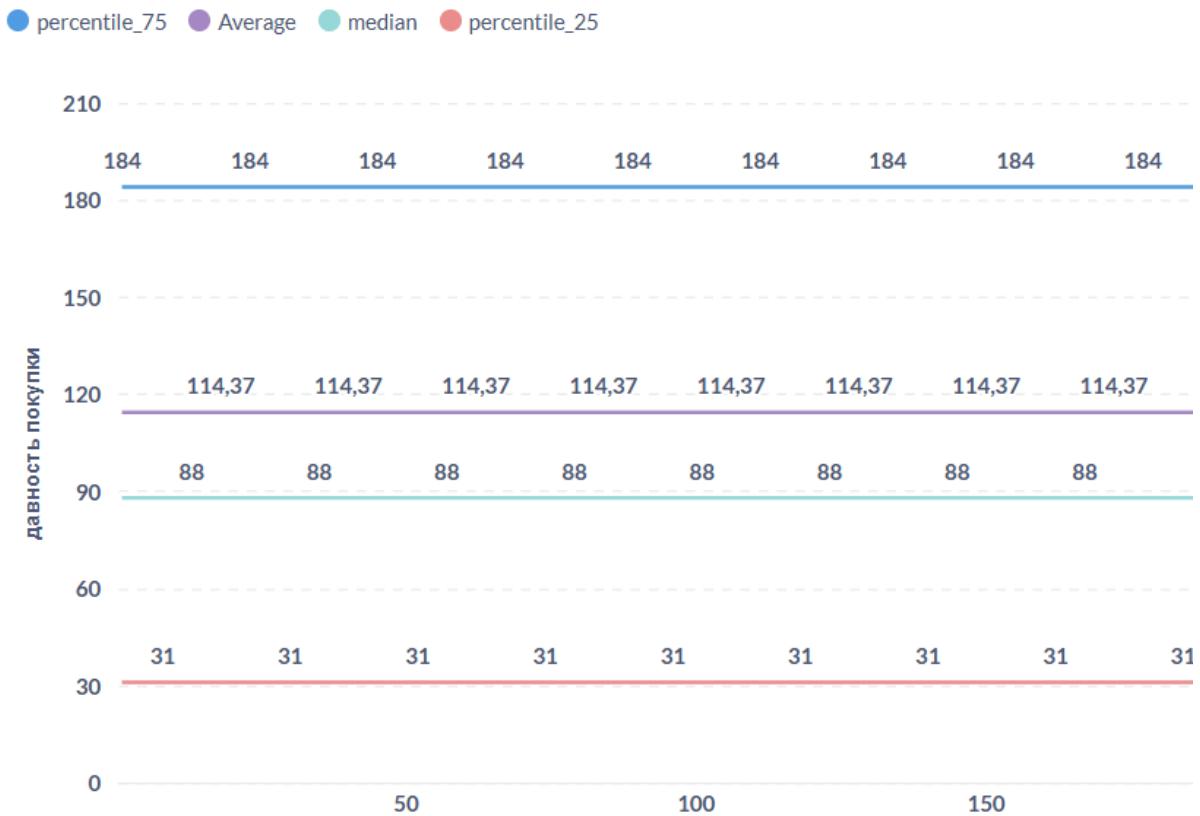
```
with sortbonuscheques as (
select
    distinct on
        (doc_id)
card,
        datetime,
        summ
```

```

from
    bonuscheques
where
    card ~ '^[0-9]+\$'),
rfm as(select card, count(*) as frequency, ('2022-06-10':: date - max(datetime) :: date) as recency,
sum(summ) as monetary
from sortbonuscheques
group by card),
rfm_stats as (
select
    AVG(recency) as avg_recency,
    PERCENTILE_CONT(0.25) within group (
order by
    recency) as percentile_25,
    PERCENTILE_CONT(0.50) within group (
order by
    recency) as median_recency,
    PERCENTILE_CONT(0.75) within group (
order by
    recency) as percentile_75
from
    rfm
)
select
    stats.avg_recency as avg,
    stats.percentile_25 as percentile_25,
    stats.median_recency as median,
    stats.percentile_75 as percentile_75
from
    rfm_stats as stats

```

25% клиентов совершили покупку в течение последнего месяца. Стандартный клиент совершил покупку около трех месяцев назад. Спящий клиент не совершал покупку за последние 184 дня. Среднее значение выше медианного это указывает на наличие клиентов, которые очень давно не покупали в аптеке.



RFM - анализ

Это итоговый запрос всего исследования. Его цель — классифицировать каждого клиента в один из семи сегментов на основе его покупательского поведения и найти клиентов относящихся к каждому сегменту для корректной рассылки.

- **Recency:** метрика была рассчитана путем вычитания даты последней транзакции каждого клиента из общей даты окончания анализируемого периода.
- **Frequency:** метрика была получена путем подсчета количества всех уникальных транзакций для каждого клиента после предварительной очистки данных от дубликатов.
- **Monetary :** рассчитана как сумма всех его покупок.

Делаем подготовку данных и считаем recency, frequency, monetary для уникального клиента. Далее присваиваем каждому клиенту баллы от 1 до 3, основываясь на пороговых значениях, которые мы определили ранее. И далее применяем систему правил case when, которая присваивает каждому клиенту финальное название сегмента.

- Чемпионы:

Элита клиентской базы. Покупали недавно, часто и на самые большие суммы.

- В зоне риска:

Раньше покупали много и часто, но давно не возвращались.

- Лояльные клиенты:

Покупают очень часто и на значительные суммы, формируя стабильный доход.

- Растущие:

Перспективные. Совершили недавнюю покупку и уже имеют опыт повторных визитов.

- Новички:

Недавно совершили первую покупку. Ключевая задача — мотивировать их на второй визит.

- Спящие:

Давно совершили одну покупку. Наименее активная группа, неинтересная для бизнеса.

- Требуют внимания:

Группа со средними показателями. Самая большая с огромным, но нераскрытым потенциалом.

И в самом конце считаем количество покупателей в каждой группе, чтобы понять распределение клиентов.

```
with sortbonuscheques as (
```

```
select
```

```
    distinct on
```

```
        (doc_id)
```

```
    card,
```

```
        datetime,
```

```
        summ
```

```
from
```

```
    bonuscheques
```

```
where
```

```
    card ~ '^[0-9]+\$',
```

```

rfm as(select card, count(*) as frequency, ('2022-06-10':: date - max(datetime) :: date) as recency,
sum(summ) as monetary
-- последняя дата 2022-06-09
from sortbonuscheques
group by card),
rfm_scored as(select
    *,
    case
        when recency <= 31 then 3
        when recency <= 184 then 2
        else 1
    end as r_score,
    case
        when frequency > 4 then 3
        when frequency >= 2 then 2
        else 1
    end as f_score,
    case
        when monetary >= 2400 then 3
        when monetary >= 900 then 2
        else 1
    end as m_score
from
    rfm
),
segmention as(select
    card,
    recency,
    frequency,
    monetary,
    concat(r_score, f_score, m_score) as rfm,
    case
        when r_score = 3 and f_score = 3 and m_score = 3 then 'Чемпионы'
        when r_score = 1 and f_score >= 2 and m_score >= 2 then 'В зоне риска'
        when f_score = 3 and m_score >= 2 then 'Лояльные клиенты'
        when r_score = 3 and f_score = 2 then 'Растущие'
        when r_score = 3 and f_score = 1 then 'Новички'
        when r_score = 1 and f_score = 1 then 'Спящие'
    else 'Требуют внимания'
    end as segment
from
    rfm_scored
order by rfm desc)
select
    segment,
    count(*)
from
    segmention
group by
    segment

```

Анализ разделил всю клиентскую базу на осмысленные, измеримые группы. Размеры сегментов подтверждают наши гипотезы. Есть небольшое, но значимое ядро лояльных клиентов и чемпионов, которые составляют 20% от всех клиентов. Есть большая группа "Спящих", которые пришли один раз и то очень давно, чуть меньше 20%. Самым большим ся является сегмент «Требуют внимания». Их показатели не являются ни выдающимися, ни провальными - это среднестатистический клиент. Даже небольшое улучшение поведения этой группы даст хороший эффект.



Выводы

Был проведен комплексный RFM-анализ. В ходе работы была получена четкая структура клиентской базы, основанная не на стандартных порогах, а на основе наших данных, то есть обоснованных порогах, которые были получены на основе таких метрик как ABC-анализ, среднее, медианное значения, перцентили.

- Анализ показал, что лояльные клиенты и чемпионы составляют около 20% от всей базы, то есть работает принцип Парето. Именно они формируют 70% всей выручки, поэтому являются очень ценными для нашей компании.
- Гипотеза о наличии "одноразовых", неактивных клиентов подтвердилась их в районе 1000 и 25-й перцентиль равный 1 показал, что минимум 25% - это клиенты, купившие один раз.
- Также подтвердилась гипотеза о что существует сегмент раньше лояльных клиентов, которые сейчас на грани ухода. Анализ показал, что таких около 400 клиентов.
- Гипотеза о большом количестве стабильных клиентов подтвердилась. Их около 45% от всей базы. Это значит, что даже если хоть немного улучшить показатели этой группы то бизнес будет приносить больший доход.

Универсальная рассылка для каждого сегмента не будет работать. Эффективность будет достигнута при дифференцированной стратегии. Это повысит рентабельность маркетинговой компании.

1. Самый ценный сегмент - чемпионы и лояльные клиенты. Для них нужно разработать и внедрить VIP программу. Они уже лояльны и готовы тратить большие деньги, поэтому стандартная рассылка со скидками им не так интересна. Им нужно ощущение эксклюзивности. Поэтому предлагается сделать к примеру следующие рассылки:

- Поздравление с праздником в виде материального подарка, учитывая прошлые покупки клиента. Таким образом покупатель придет в аптеку и возможно еще что нибудь приобретет и соответственно он почувствует персонализированный подход и будет еще более лояльным.

"Имя" поздравляем с праздником! Ваш небольшой подарок уже ждет вас в нашей аптеке.

- После покупки отправить сообщение о том, что у клиента есть возможность проконсультироваться с нами по поводу препарата.

Ваш персональный фармацевт на связи. Если у вас есть вопросы по препарату, готовы проконсультировать.

- Ранний доступ к препаратам и возможность забронировать, также возможность получить в подарок пробник

Новые витамины уже в наличии. Закажите до старта продаж и получите в подарок [пробник].

2. Новички. Их нужно сподвигнуть на повторную покупку, тем самым создавать привычку покупать именно у нас. Это даст возможность со временем перевести их в группу постоянных покупателей.

- Рассылка сразу после покупки.

Спасибо за вашу первую покупку! Надеемся, вам все понравилось. Можете оставить отзыв по ссылке. Ждем вас у нас за новыми покупками.

- Напоминание о себе в несколько этапов.
 1. Ссылка на статью связанную с купленными товарами, в которой описано как правильно принимать, возможно как улучшить эффект.
 2. Персонализированные скидки на вторую покупку с ограниченным сроком действия.

- Напоминание и последний шанс

Ваша персональная скидка на вторую покупку сгорит через 2 дня! Успейте воспользоваться.

3. Спящие. Вероятность их отклика стремится к нулю. Поэтому чтобы не тратить бюджет впустую можно отправить всего одно "агрессивное" сообщение напоминание о нас.

Скидка 30% на всё, только 3 дня!

Всех, кто не отреагирует, перевести в "холодную базу" и прекратить платные коммуникации.

4. В зоне риска. Вернуть этих клиентов дешевле, чем привлекать новых, которые будут в группе лояльных покупателей

- Персональные предложения

Мы заметили, что вы давно не заходили. Специально для вас скидка на все товары или к примеру скидка на товары, которые покупал клиент ранее.

- Бонус

Мы соскучились! Дарим 500 бонусов на покупку от 2000 рублей. Срок действия 7 дней.

5. Требуют внимания. Это наши стандартные клиенты, которые покупаю ни так часто и ни на такие большие суммы. Самая большая группа со средними показателями. Без правильного взаимодействия они скатятся в группу спящих, а при хорошем дорастут до лояльных клиентов.

- Общие акции

Только на этой неделе — скидка 15% на все сезонные товары

- Стимуляция регулярности

1. Если клиент купил препарат который необходимо принимать каждый месяц, можно предложить бесплатное напоминание о том что пришло время следующей покупки. Тем самым привлечь к покупке именно у нас.
2. Вы не покупали витамины уже 2 месяца. Возможно, пора пополнить запасы? Дарим скидку 10% на ваш любимый бренд

6. Растущие клиенты. Они не так давно пришли, но уже возвращались, они уже не случайные прохожие. Нужно превратить визиты в привычку и есть высокая вероятность отклика, так как они были недавно

- Увеличение чека

К вашим покупкам из категории идеально подойдут наши средства. Познакомьтесь с линейкой со скидкой 20%

- Знакомство с товарами

Знаете ли вы, что у нас есть не только лекарства, но и большой выбор ортопедических товаров? Получите бесплатную консультацию.

- Программа лояльности

После второй покупки вы автоматически стали участником нашей бонусной программы. С каждой следующей покупки вам начисляется 5% кэшбэка.