

RFM анализ клиентов аптечной сети

RFM-анализ — метод анализа, позволяющий сегментировать клиентов по частоте и сумме покупок и выявлять тех клиентов, которые приносят больше денег.

Аббревиатура RFM расшифровывается:

- Recency — давность (как давно ваши пользователи что-то у вас покупали);
- Frequency — частота (как часто они у вас покупают);
- Monetary — деньги (общая сумма покупок).

По этим признакам можно разделить всех клиентов на группы, понять, кто из клиентов покупает часто и много, кто — часто, но мало, а кто вообще давно ничего не покупал. Такая сегментация используется для построения схем коммуникации с клиентами в зависимости от группы, к которой они принадлежат.

В данной работе проведен RFM-анализ с целью сегментации клиентов аптечной сети. Расчеты выполнены на основе таблицы **bonuscheques** базы данных **Аптека** (таблица транзакций участников бонусной системы аптеки).

Описание полей таблицы:

- datetime - дата и время совершения транзакции;
- shop - название аптеки;
- card - номер бонусной карты;
- bonus_earned - количество начисленных бонусов;
- bonus_spent - количество потраченных бонусов;
- summ - сумма чека;
- summ_with_disc - сумма чека с учетом скидок и списаний бонусов;
- doc_id - номер документа.

По каждому из признаков Recency, Frequency, Monetary были выделены по три равные группы. Каждой группе было присвоено числовое обозначение от 1 до 3:

Номер группы	По давности заказа (recency)	По частоте покупок (frequency)	По сумме покупок (monetary)
1	давние клиенты	покупает очень редко	маленькая сумма
2	относительно недавние клиенты	покупает нечасто	средняя сумма
3	недавние клиенты	покупает часто	большая сумма

Визуализация сделана в BI-системе Yandex DataLens:

<https://datalens.yandex.ru/mbchddpt0bawa-rfm-analiz>

1. Для начала посчитаем общее количество клиентов аптечной сети. В нашем случае идентификатором клиента будет номер бонусной карты:

```
select count (distinct card) from bonuscheques  
where card like '2000%' - - исключаем фейковые карты по условию задачи
```

Результат запроса составит 5926. Это общее количество клиентов, которых необходимо сегментировать по параметрам recency, frequency, monetary.

2. Проанализируем распределение клиентов по параметрам recency, frequency, monetary.

2.1. Параметр Recency (давность покупок)

Для начала обратим внимание, что мы имеем дело с историческими данными. Запрос *select max(datetime) from bonuscheques* выведет в качестве самой последней дату покупки 9 июня 2022 года.

Поскольку мы имеем дело с историческими данными, то сравнивать даты покупки с сегодняшним днем будет не вполне корректно и введет в заблуждение. Поэтому давность покупок будем сравнивать в самой свежей датой в имеющихся данных - 09.06.2022.

```
with tmp as (  
  select  
    card,  
    '2022-06-09'::date - max(datetime)::date as recency  
  from bonuscheques  
  where card like '2000%'  
  group by card)  
select  
  count(card) as cnt,  
  recency  
from tmp  
group by recency
```

Результат запроса в виде графика:

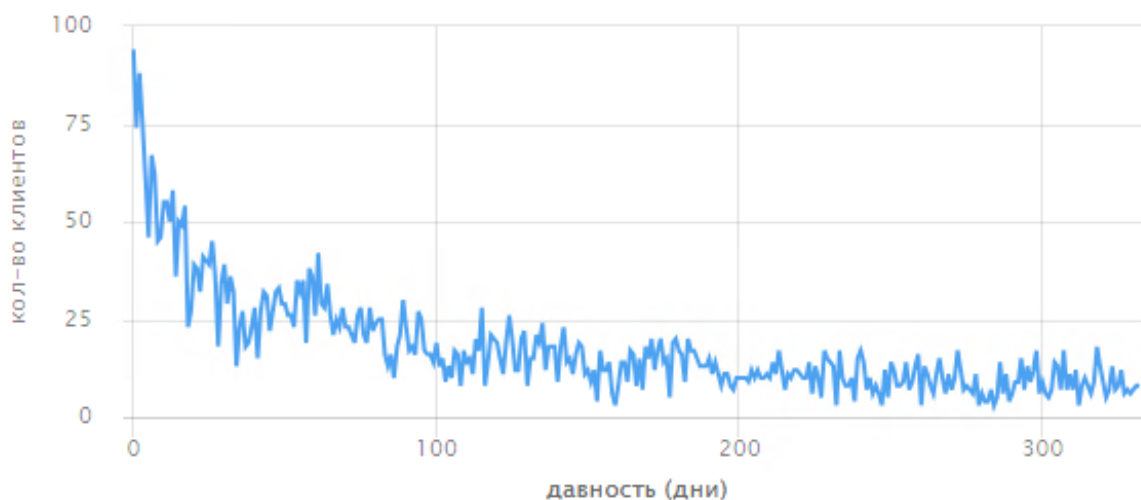
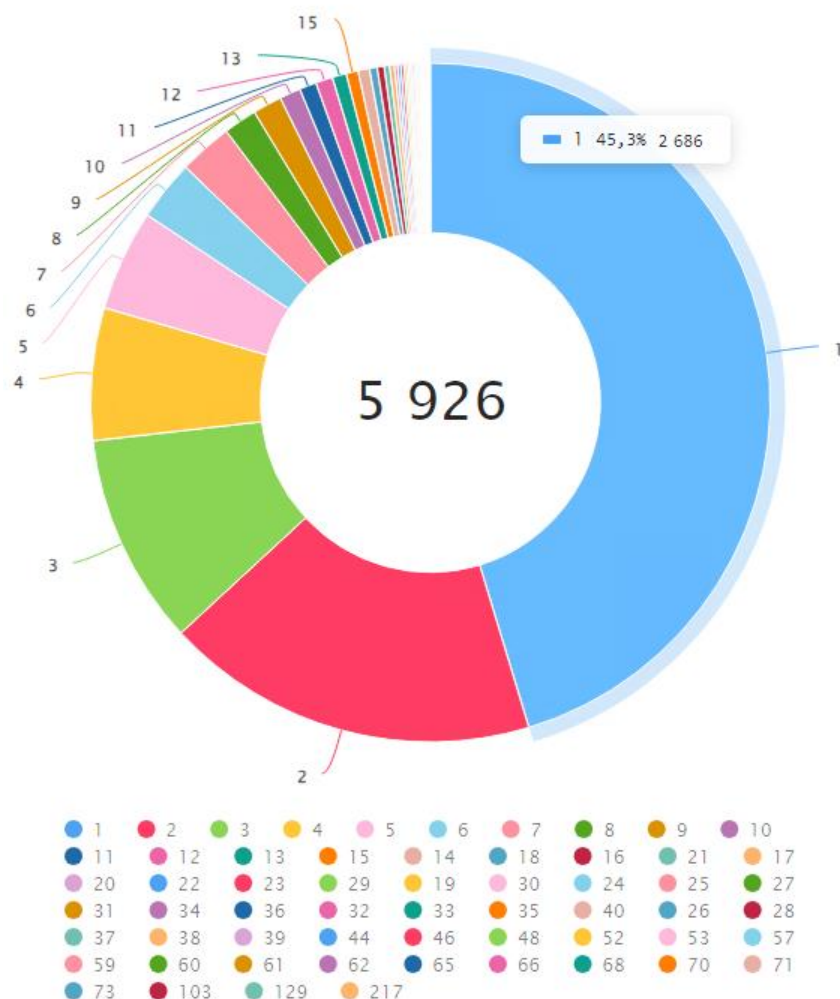


График имеет сильный уклон на промежутке от 0 до 33 дней, потом постепенное равномерное снижение показателя. Минимальное значение параметра Resensу составляет 0 дней, максимальное - 332 дня.

2.2. Параметр Frequency (частота покупок):

```
with tmp as(
  select
    card,
    count(datetime) as frequency
  from bonuscheques
  where card like '2000%'
  group by card
)
select
  frequency,
  count(card) as cnt
from tmp
group by frequency
```

Результат запроса в виде круговой диаграммы распределения количества покупок на клиента:



2686 из 5926 клиентов (45%) сделали в аптеке всего одну покупку. 1-3 раза покупали 73% клиентов. Из этого распределения интуитивно обозначаются границы сегментирования клиентов по параметру frequency - 1 и 3 покупки.

2.3. Параметр Monetary (общая сумма покупок):

```
with tmp as (  
  select  
    card,  
    sum(summ_with_disc) as monetary  
  from bonuscheques  
  where card like '2000%'  
  group by card)  
select  
  monetary,  
  count(card) as cnt  
from tmp  
group by monetary
```

Результат запроса в виде столбчатой диаграммы:

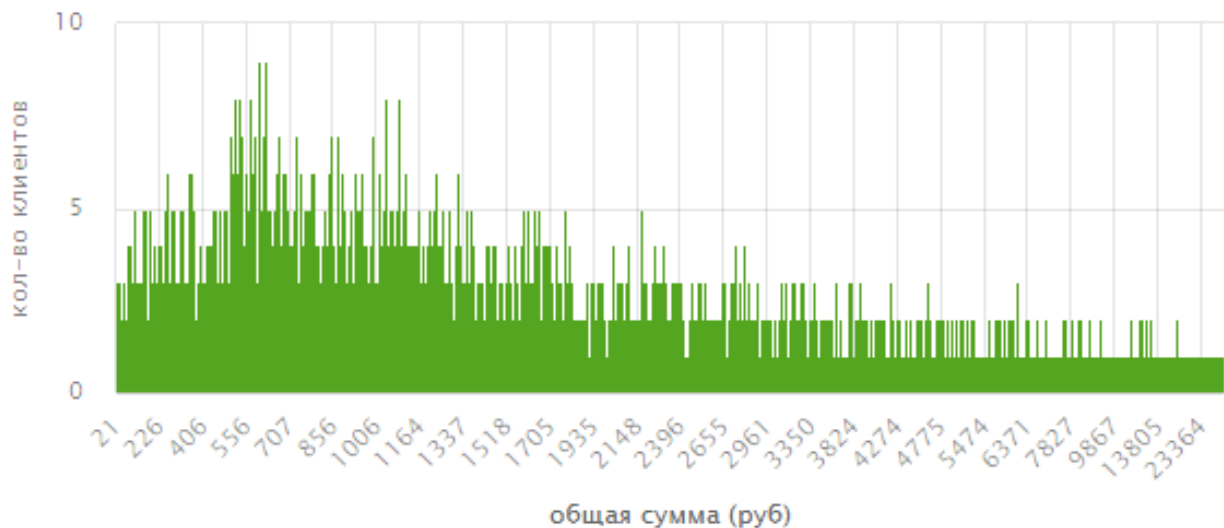


График по параметру monetary сильно скошен вправо, крупные суммы денег в аптеке клиенты оставляли редко.

3. Считаем пороговые значения для определения границ сегментов по параметрам Recency, Frequency, Monetary по 33-ему и 66-му перцентилям:

```
with tmp as (  
  select  
    card,  
    '2022-06-09'::date - max(datetime)::date as recency,  
    count(datetime) as frequency,  
    sum(summ_with_disc) as monetary  
  from bonuscheques  
  where card like '2000%'  
  group by card  
)  
select  
  min('recency') as показатель,  
  percentile_disc(0.33) within group(order by recency) as perc_33,  
  percentile_disc(0.66) within group(order by recency) as perc_66  
from tmp  
union  
select  
  min('frequency'),  
  percentile_disc(0.33) within group(order by frequency),  
  percentile_disc(0.66) within group(order by frequency)  
from tmp  
union  
select  
  min('monetary'),  
  percentile_disc(0.33) within group (order by monetary) as perc_33,  
  percentile_disc(0.66) within group (order by monetary) as perc_66  
from tmp
```

Результат запроса в виде таблицы:

показатель	значение 33-его процентиля	значение 66-го процентиля
recency	48	143
monetary	920	2394
frequency	1	3

Отметим, что пороговые значения по параметру frequency совпадают с предположениями, выдвинутыми на основе анализа круговой диаграммы.

4. Выделяем группы клиентов на основе выполненного гfm анализа:

4.1. По параметру recency:

1 группа - покупали давно, более 143 дней назад от взятой за точку отсчета даты (или более, чем 4 месяца назад);

2 группа - покупки делались 48 - 143 дня назад от взятой за точку отсчета даты (более месяца назад, но не более 5 месяцев назад);

3 группа - покупали недавно, менее 48 дней назад относительно взятой за точку отсчета даты (примерно в течение последнего месяца - полутора)

4.2. По параметру frequency:

1 группа - клиенты, которые совершили всего одну покупку в аптеке;

2 группа - клиенты, которые совершили 2-3 покупки в аптеке;

3 группа - клиенты, которые совершили более 3 покупок в аптеке.

4.3. По параметру frequency

По данному параметру округлим пороговые значения показателя до сотен рублей, чтобы пороговые значения не смотрелись странно в классификации:

3 группа - клиенты, которые совершили покупки на общую сумму до 900 р.;

2 группа - клиенты, которые совершили покупки на общую сумму 900 - 2400 р.;

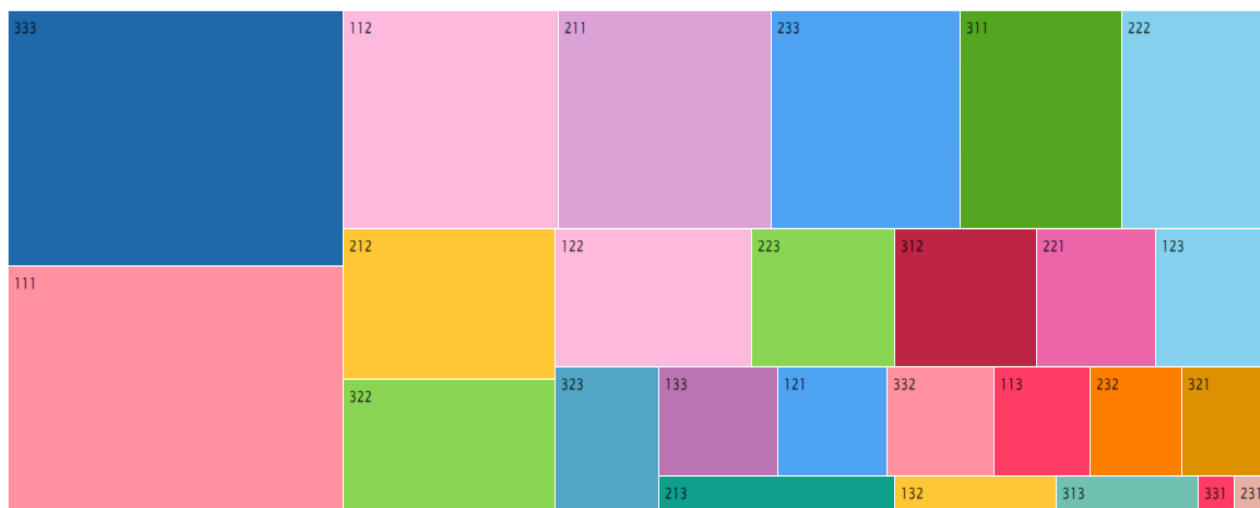
1 группа - клиенты, которые совершили покупки на общую сумму более 2400 р..

5. Сегментируем клиентов по выделенным группам:

```
with tmp as (  
  select  
    card,  
    '2022-06-09'::date - max(datetime)::date as recency,  
    count(datetime) as frequency,  
    sum(summ_with_disc) as monetary  
  from bonuscheques  
  where card like '2000%'  
  group by card  
) , tmp2 as (  
  select  
    card,  
    case  
      when recency <= 48 then '3'  
      when recency > 143 then '1'  
      else '2' end as recency_group,  
    case  
      when frequency > 3 then '3'  
      when frequency = 1 then '1'  
      else '2' end as frequency_group,  
    case  
      when monetary > 2400 then '3'  
      when monetary <=900 then '1'  
      else '2' end as monetary_group  
  from tmp  
)
```

```
select
  count(card),
  concat(recency_group, frequency_group, monetary_group) as rfm_group
from tmp2
group by rfm_group
order by count(card) desc
```

Результат запроса в виде древовидной диаграммы:



И в виде таблицы:

RFM группа	Кол-во клиентов, чел.	Доля в общем кол-ве, %
333	799	13,48
111	775	13,08
112	437	7,37
211	432	7,29
233	382	6,45
311	327	5,52
222	302	5,10
212	297	5,01

322	268	4,52
122	253	4,27
223	185	3,12
312	184	3,10
221	153	2,58
123	148	2,50
323	142	2,40
133	121	2,04
121	111	1,87
332	109	1,84
113	98	1,65
232	93	1,57
321	90	1,52
213	85	1,43
132	58	0,98
313	51	0,86
331	13	0,22
231	11	0,19

131	2	0,03
Всего	5 926	100,00

6. Краткие выводы по результатам RFM-анализа:

1. Примерно одинаковые доли по 13% в распределении занимают самые лояльные и “потерянные экономные клиенты”. Эти две группы составляют более четверти клиентов, остальные группы представлены долями от 7% и ниже и распределены более равномерно.

2. Клиенты распределены примерно одинаково по показателю Recency:

- 33% постоянных клиентов (группы, начинающиеся с цифры 3) - это группы, с которыми нужно активно и разнообразно работать;
- 33% спящих клиентов (группы, начинающиеся с цифры 2) - клиенты, которых нужно постараться перевести в группу постоянных;
- 34% уходящих клиентов (группы, начинающиеся с цифры 1) - клиенты, часть из которых, вероятно, безвозвратно потеряна, а часть можно постараться вернуть.

3. Из уходящих клиентов группы 111, 112, 113 - скорее всего безвозвратно потерянные клиенты, на которых не стоит тратить много усилий, они составляют в сумме 1310 клиентов или 22%.

4. В целом наибольшие проблемы видятся по показателю Frequency, 45% клиентов совершали покупки не просто редко, а всего один раз.

Ниже представлен вариант маркетинговых действий с разбивкой по каждой группе клиентов:

Группа	Характеристика	Что отправляем
Уходящие		
111	Потерянные экономные	Скорее всего, эти клиенты уже ушли от нас, и не стоит тратить слишком больших усилий на этот сегмент. Отправляем автоматическую цепочку реактивации.
112	Одноразовые	
113	Одноразовые	
121	Уходящие редкие	Эти клиенты сделали у нас несколько покупок, а потом перестали покупать. Их можно попытаться вернуть и обязательно получить обратную связь, почему они от нас ушли. Отправляем реактивационную рассылку с опросом.
122		
123		

131	Уходящие постоянные	Этих клиентов обязательно нужно попытаться вернуть. Отправляем им реактивационные письма, запрашиваем обратную связь, предлагаем бонусы, программы лояльности.
132	Уходящие хорошие постоянные	
133	Уходящие VIP	
Спящие		
211	Спящие разовые с маленьким чеком	Эти клиенты ещё помнят о нас. Нужно попробовать расшевелить их стимулирующими акциями. Отправляем им рассылки с интересными акциями и выгодными предложениями.
212		
213		
221	Спящие редкие с маленьким чеком	
222	Спящие редкие со средним чеком	
223	Спящие редкие с высоким чеком	
231	Спящие постоянные с маленьким чеком	Ещё не так давно они часто покупали у вас, а сейчас перестали. Нужно обязательно узнать, в чём дело, и попытаться вернуть их. Отправляем им рассылку с интересными акциями и рассылку с опросом.
232	Спящие постоянные со средним чеком	
233	Спящие постоянные с высоким чеком	
Постоянные		
311	Новички с маленьким чеком	Отправляем welcome-цепочку с описанием преимуществ, ответами на частые вопросы, обучающие материалы
312	Новички со средним чеком	
313	Новички с высоким чеком. Потенциальные VIP	В дополнение к welcome-цепочке отправляем им интересное предложение, чтобы удержать интерес
321	Постоянные с маленьким чеком	Нужно попытаться увеличить средний чек допродажей. Отправляем им

322	Постоянные со средним чеком	рассылки с сопутствующими товарами
323	Постоянные с высоким чеком	Эти клиенты и так хорошо покупают, не стоит утомлять их лишними рассылками. Отправляем только обычные рассылки
331	Постоянные с маленьким чеком	Можно попытаться увеличить сумму чека допродажей. Отправляем им рассылки с сопутствующими товарами
332	Постоянные со средним чеком	Наши самые хорошие клиенты, можно попробовать увеличить сумму чека. Отправляем им специальные предложения, сопутствующие товары
333	VIP	Отправляем им особые предложения, например, участие в программе лояльности — приглашение в клуб любимых клиентов