**RFM-анализ клиентской базы аптечной сети**

**Цели исследования:**

1. Провести классификацию пользователей.
2. Сформировать рекомендации для полученных групп пользователей.

В качестве исходных данных у нас есть база данных “Apteka”, которая состоит из 4 таблиц:

***bonuscheques*** - таблица с информацией о транзакциях по бонусной системе;

***employee*** - таблица с информацией о сотрудниках;

***shops*** - таблица с информацией об аптеках;

***sales*** - таблица с информацией о продажах.

Для решения этой задачи воспользуемся таким инструментом, как RFM-анализ — метод анализа, позволяющий сегментировать клиентов по частоте и сумме покупок и выявлять тех, которые приносят больше денег.

Также помогает понять, кто из клиентов лучше всего реагирует на текущие рекламные кампании и на будущие активности. Как правило, небольшой процент пользователей реагирует на общие рекламные предложения. RFM-анализ является отличным методом прогнозирования реакции клиента и улучшения взаимодействия. RFM использует поведение покупателей, чтобы определить, как работать с каждой группой клиентов.

в RFM входят:

* ***Recency*** — давность (как давно клиенты были у нас). Давность в срезе клиентской базы можно посмотреть, если отсортировать клиентов по дате последней покупки.
* ***Frequency*** — частота (как часто они у нас покупают). Для расчета частоты посещения нужно количество покупок за период времени.
* ***Monetary*** — деньги (общая сумма трат).

Для расчетов воспользуемся таблицей ***bonuscheques***, из которой возьмем в качестве клиентов номера их бонусных карт. Дату последнего посещения возьмем из столбца ***datetime*** - дата и время совершения транзакции. Источником денежных трат выступит столбец ***summ\_with\_disc*** - сумма чека с учетом скидок и списаний бонусов.

Определим за какой период времени представлены данные в таблице:

**SELECT** **min**(datetime::**date**) **AS** *min\_date*,

**max**(datetime::**date**) **AS** *max\_date*

**FROM** bonuscheques

В таблице собраны данные почти за 11 месяцев: с 12.07.2021 г. по 09.06.2022 г. Жаль, что не за весь год, но в целом хороший набор, так как, учитывая, что это аптеки, то в данный период как раз попадают сезонные заболевания, во время которых людям чаще требуются медицинские препараты.

Учитывая тот факт, что если в момент покупки касса была в оффлайн-режиме, то вместо номера карты записывается зашифрованная последовательность символов, то необходимо отфильтровать наши данные, чтобы в них остались только верифицированные клиенты, для которых мы и будем разрабатывать рекомендации.

**with** *filtered* **as** (

**select** datetime::**date** **as** dt,

card,

summ\_with\_disc **as** *summ*,

doc\_id

**from** bonuscheques *b*

**where** **length**(card) < 14

),

Далее, сгруппировав отфильтрованную таблицу по нашим клиентам, вычислим необходимые нам данные: дату последнего визита, количество покупок и на какую сумму эти покупки были сделаны.

*grouped* **as** (

**select** card **as** *clients*,

**max**(***dt***) **as** *last\_visit*,

**count**(**distinct** doc\_id) **as** *cnt\_purchases*,

**sum**(*summ*) **as** *sum\_purchases*

**from** *filtered*

**group** **by** card

),

За точку отсчета для определения количества дней с момента последней покупки возьмем 13 июля 2022 года.

*agg* **as** (

**select** *clients*,

*last\_visit*,

**date**(**'2022-07-13'**) - *last\_visit* **as** *diff\_days*,

*cnt\_purchases*,

*sum\_purchases*

**from** *grouped*

),

Определим границы сегментов для трех показателей. Учитывая, что мы работаем с данными клиентов аптеки, то границы между лучшими и худшими клиентами выглядят размыто, так как сильно влияет фактор сезонности, к тому же сезоны заболеваний склонны сдвигаться по времени. Также есть фактор инфляции потребительских цен, из-за чего одного и того же клиента, купившего на определенную сумму, мы можем отнести сперва к одной группе, а спустя условно полгода он уже окажется в другой. Поэтому в данном случае считаю более рационально разделить соответствующие массивы с помощью перцентилей на три группы: до 33%, до 66% и более 66%.

*perc* **as** (

**select** **'perc\_of\_day'** **as** *perc\_name*,

**percentile\_disc**(0.33) **within** **group** (**order** **by** *diff\_days*) **as** *"33%"*,

**percentile\_disc**(0.66) **within** **group** (**order** **by** *diff\_days*) **as** *"66%"*

**from** *agg*

**union**

**select** **'perc\_of\_cnt'** **as** *perc\_name*,

**percentile\_disc**(0.33) **within** **group** (**order** **by** *cnt\_purchases*),

**percentile\_disc**(0.66) **within** **group** (**order** **by** *cnt\_purchases*)

**from** *agg*

**union**

**select** **'perc\_of\_sum'** **as** *perc\_name*,

**percentile\_disc**(0.33) **within** **group** (**order** **by** *sum\_purchases*),

**percentile\_disc**(0.66) **within** **group** (**order** **by** *sum\_purchases*)

**from** *agg*

),

На выходе получаем такие результаты:

| **perc\_name** | **33%** | **66%** |
| --- | --- | --- |
| **perc\_of\_sum** | *920* | *2394* |
| **perc\_of\_cnt** | *1* | *3* |
| **perc\_of\_day** | *82* | *177* |

Из чего формируем следующие сегменты:

По количеству дней прошедших с последней покупки:

* от 1 до 82 дней - 1 группа (недавний клиент),
* от 83 дней до 177 дней - 2 группа (ожидающий клиент),
* более 177 дней - 3 группа (давний клиент)

По количеству совершенных покупок:

* более 3 покупок - 1 группа (частый клиент),
* 2 - 3 покупки - 2 группа (редкий клиент),
* 1 покупка - 3 группа (разовый клиент)

По количеству общих денежных трат:

* более 2394 рублей - 1 группа (высокий чек),
* от 920 рублей до 2394 рублей - 2 группа (средний чек),
* до 920 рублей - 3 группа (низкий чек)

Формируем таблицу для RFM-анализа:

*rfm* **as** (

**select** *clients*,

*diff\_days*,

**case**

**when** *diff\_days* < 83 **then** **'1'**

**when** *diff\_days* < 178 **then** **'2'**

**else** **'3'**

**end** **as** *"r\_(recency)"*,

*cnt\_purchases*,

**case**

**when** *cnt\_purchases* = 1 **then** **'3'**

**when** *cnt\_purchases* <= 3 **then** **'2'**

**else** **'1'**

**end** **as** *"f\_(frequency)"*,

*sum\_purchases*,

**case**

**when** *sum\_purchases* < 921 **then** **'3'**

**when** *sum\_purchases* < 2395 **then** **'2'**

**else** **'1'**

**end** **as** *"m\_(monetary)"*

**from** *agg*

)

На основе полученной таблицы выделим категории и определим, сколько клиентов находится в каждой из них.

**select** *category*,

**count**(*clients*)

**from** (**select** *clients*,

**concat**(*"r\_(recency)"* || *"f\_(frequency)"* || *"m\_(monetary)"*) **as** *category*

**from** *rfm*

) *t*

**group** **by** *category*

**order** **by** *category*

| **category** | **111** | **112** | **113** | **121** | **122** | **123** | **131** | **132** | **133** | **211** | **212** | **213** | **221** | **222** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **count** | *799* | *109* | *13* | *143* | *264* | *93* | *51* | *181* | *330* | *383* | *92* | *11* | *185* | *301* |

| **category** | **223** | **231** | **232** | **233** | **311** | **312** | **313** | **321** | **322** | **323** | **331** | **332** | **333** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **count** | *154* | *86* | *286* | *442* | *121* | *57* | *3* | *148* | *248* | *116* | *98* | *418* | *794* |

Получилось 27 сегментов, начиная от супер лояльных с категорией 111 - недавно сделали много покупок на большую сумму, до 333 - были давно, сделали одну покупку на маленькую сумму. Так как все клиенты из этой таблицы являются обладателями бонусных карт, то необходимо и для всех разработать рекомендации для улучшения взаимодействия с ними и, как следствие, потенциально увеличить прибыль от них.

Как видно из таблицы, сегменты получились неравномерными, поэтому целесообразно группы объединять. Поэтому выделим следующие объединенные группы и пропишем рекомендации для них:

1. VIP - клиенты из сегмента 111. Они очень лояльны к нашей сети аптек, поэтому для них можно сделать привилегированную программу лояльности, которая позволит, к примеру, участвовать в распродажах раньше остальных.
2. Лояльные - к ним отнесем клиентов из сегментов 112, 113, 121. Эти клиенты также довольны нашим продуктом, так как покупали недавно и в большом количестве, но у них необходимо повысить сумму чека. Поэтому им необходимо отправлять предложения с более дорогими товарами или с сопутствующими товарами.
3. Растущие - клиенты из сегментов 122, 123. Видно, что они заинтересованы в наших товарах, но пока тратятся не сильно, поэтому их надо “подтолкнуть” и постараться перевести в разряд лояльных клиентов. Для чего можно отправлять рассылки с сопутствующими товарами или более частыми периодами скидок.
4. Новые - клиенты из сегментов 131, 132, 133. Новых клиентов нужно познакомить с нашим брендом и рассказать о всех преимуществах пользования им. Можно предоставить скидку на крупную покупку.
5. Умеренные - клиенты из сегментов 211, 212, 213, 221, 222. Можно сказать это наш средний класс. У них уже есть определенное поведение по взаимодействию с нами, поэтому не стоит утомлять их излишней рекламой. Лучше отправлять рассылки во время сезонных скидок.
6. В зоне риска - клиенты из сегментов 223, 231, 232, 233. У данных клиентов необходимо подогреть интерес к нашим продуктам. Рассылаем им сообщения с интересными акциями и выгодными предложениями.
7. Уходящие лояльные - клиенты из сегментов 311, 312, 313. Для данных клиентов лучше провести опрос, и выяснить причины изменения отношения к нашей продукции. Отправить им новую полезную информацию. Напомнить о бонусах и программах лояльности.
8. Уходящие малоактивные - клиенты из сегментов 321, 322, 323. От этих клиентов также желательно получить обратную связь, все таки они тоже обладатели бонусных карт. Отправить им персонализированную рассылку. Предложить попробовать ранее понравившиеся продукты еще раз. Отправить им новую полезную информацию.
9. Разовые - клиенты из сегментов 331, 332, 333. Скорее всего они воспользовались когда-то нашей аптекой не целенаправленно, а просто она им подвернулась по пути домой или на работу. В связи с чем не рационально прикладывать значительные усилия по рекламе, но так как этот сегмент клиентов достаточно большой, то стоит отправить им реактивационную рассылку с напоминанием о себе и о своих преимуществах. В результате произойдет какая-то конверсия и некоторые клиенты вернутся.

Как видно из проведенного исследования - RFM-анализ помогает во многих аспектах: уменьшить количество потерянных клиентов, предлагать допродавать товары сегментам, которые лучше всего идут навстречу, увеличивать лояльность клиентов, продавать дорогие товары. Но надо не забывать, и время от времени пересматривать сегменты и снова делать RFM-анализ, чтобы посмотреть, какие из наших клиентов изменили свой статус, что позволит нам понять эффективность наших действий.