RFM-анализ клиентской базы аптечной сети

## Выбор метрик

Для целей проведения RFM-анализа используем таблицу bonuscheques, так как она содержит информацию о номере бонусной карты, с помощью которой мы можем идентифицировать клиента.

Допущение: не будем учитывать зашифрованные карты, поскольку в нашем случае нет возможности их расшифровать и они будут искажать картину (т.к. зашифрованный код уникален и нет возможности осуществить разбивку по клиентам).

Также цель нашего исследования заключается в формировании рекомендаций по работе с имеющейся клиентской базой, а возможность взаимодействовать с “зашифрованными” клиентами у нас отсутствует.

Recency: Посчитаем количество дней с момента покупки. За дату отсчета возьмем максимальную дату из базы + 1 день.

Frequency: Будем считать количество уникальных doc\_id.

Monetary: Нет возможности проанализировать количество позицией в чеке, поэтому будем анализировать общую сумму чека.

Для Основных метрик посчитаем 33 и 66 перцентили и скользящую медиану:

|  |  |  |
| --- | --- | --- |
| RFM | Показатель | Значение |
| Recency | 33-percentile | 49 |
| sliding median | 88 |
| 66-percentile | 144 |
| Frequency | 33-percentile | 1 |
| sliding median | 2 |
| 66-percentile | 3 |
| Monetary | 33-percentile | 548 |
| sliding median | 739 |
| 66-percentile | 979 |

На основании полученных границ осуществим сегментацию:

|  |  |  |
| --- | --- | --- |
| RFM | Сегмент | Описание |
| 111 | Ключевые | Недавно /Покупают часто /Большая сумма |
| 112 | Лояльные | Недавно /Покупают часто /Средняя сумма |
| 113 | Недавно /Покупают часто /Маленькая сумма |
| 122 | Недавно /Средняя частота /Средняя сумма |
| 121 | Крупные | Недавно /Средняя частота /Большая сумма |
| 211 | Давно /Покупают часто /Большая сумма |
| 221 | Давно /Средняя частота /Большая сумма |
| 212 | Неактивные | Давно /Покупают часто /Средняя сумма |
| 213 | Давно /Покупают часто /Маленькая сумма |
| 222 | Давно /Средняя частота /Средняя сумма |
| 223 | Давно /Средняя частота /Маленькая сумма |
| 232 | Давно /Покупают редко /Средняя сумма |
| 233 | Давно /Покупают редко /Маленькая сумма |
| 231 | Неактивные крупные | Давно /Покупают редко /Большая сумма |
| 311 | Очень давно /Покупают часто /Большая сумма |
| 321 | Очень давно /Средняя частота /Большая сумма |
| 331 | Очень давно /Покупают редко /Большая сумма |
| 123 | Новые | Недавно /Средняя частота /Маленькая сумма |
| 131 | Недавно /Покупают редко /Большая сумма |
| 132 | Недавно /Покупают редко /Средняя сумма |
| 133 | Недавно /Покупают редко /Маленькая сумма |
| 312 | Почти потерянные | Очень давно /Покупают часто /Средняя сумма |
| 313 | Очень давно /Покупают часто /Маленькая сумма |
| 322 | Очень давно /Средняя частота /Средняя сумма |
| 323 | Очень давно /Средняя частота /Маленькая сумма |
| 332 | Очень давно /Покупают редко /Средняя сумма |
| 333 | Ушедшие | Очень давно /Покупают редко /Маленькая сумма |

## Аналитика

### Запрос 1. Основные метрики

Посчитаем основные метрики по всем клиентам: общая выручка (Revenue), количество уникальных клиентов, количество заказов, средняя стоимость заказов (AOV), средний доход на пользователя (ARPU), Lifetime.

Код запроса:

**with** *a* **as** (

**select**

card,

**min**(datetime::**date**) **as** *mn\_dt*,

**max**(datetime::**date**) **as** *mx\_dt*

**from**

bonuscheques

**where**

**length**(card) <= 13

**group** **by**

card

)

**select**

**sum**(*b*.summ) **as** *revenue*,

**count**(**distinct** *b*.doc\_id) **as** *purchases*,

**count**(**distinct** *b*.card) **as** *clients*,

round(**sum**(*b*.summ)\* 1.0 / **count**(**distinct** *b*.doc\_id), 1) **as** *aov*,

round(**sum**(*b*.summ)\* 1.0 / **count**(**distinct** *b*.card), 1) **as** *arpu*,

(

**select**

round(**avg**(a.mx\_dt - a.mn\_dt), 1)

**from**

a) **as** *lifetime*

**from**

bonuscheques *b*

### Запрос 2. Доля зашифрованных клиентов

Посчитаем основные метрики в разрезе идентифицированных и неидентифицированных клиентов: выведем общую выручку (Revenue), количество уникальных клиентов, количество заказов. А также посмотрим какую долю от средней стоимость заказов (AOV) составляет скидка.

Код запроса:

**select**

**sum**(summ) **as** *revenue*,

**count**(**distinct** doc\_id) **as** *purchases*,

**count**(**distinct** card) **as** *clients*,

**sum**(summ)\* 1.0 / **count**(**distinct** doc\_id) **as** *aov*,

**sum**(summ)\* 100.0 / **cast**(**sum**(summ\_with\_disc) **as** **numeric**) - 100 **as** *disc\_percent*,

**case**

**when** **length**(card) <= 13 **then** 'ident'

**else** 'unident'

**end** *client\_type*

**from**

bonuscheques

**group** **by**

*client\_type*

Согласно полученным данным зашифрованные записи составляют существенную долю от имеющихся у нас данных. Всего за рассматриваемый период совершено 38480 покупок, из них невозможно идентифицировать 17407 (45%). Также зашифрованные записи составляют 40% от общей выручки. Необходимо установить причину столь частого отсутствия связи в кассовых аппаратах.

Также можно заметить, что размер скидки составляет всего 0,3%, что, на первый взгляд, кажется крайне незначительным.

### Запрос 3. RFM- анализ

Распределим имеющуюся базу клиентов по сегментам RFM и посмотрим на количество клиентов, среднюю частоту и среднее количество покупок по каждому сегменту.

Код запроса:

-- Соберем данные по идентифицированным клиентам

**with** *data* **as** (

**select**

card,

**sum**(summ) **as** *sm*,

**count**(\*) **as** *cnt*,

**max**(datetime::**date**) **as** *mx\_dt*,

**make\_date**(2022, 6, 10) - **max**(datetime::**date**) **as** *delta\_dt*

**from**

bonuscheques

**where**

**char\_length**(card) <= 13

**group** **by**

card

),

-- В отдельную группу вынесем клиентов, которых невозможно идентифицировать

*unident* **as** (

**select**

card,

**sum**(summ) **as** *sm*,

**count**(\*) **as** *cnt*,

**max**(datetime::**date**) **as** *mx\_dt*,

**make\_date**(2022, 6, 10) - **max**(datetime::**date**) **as** *delta\_dt*

**from**

bonuscheques

**where**

**char\_length**(card) > 13

**group** **by**

card

),

-- Рассчитаем перцентили

*percentile* **as** (

**select**

**percentile\_cont**(0.33) **within** **group** (

**order** **by** *sm*) **as** *monetary\_per33*,

**percentile\_cont**(0.66) **within** **group** (

**order** **by** *sm*) **as** *monetary\_per66*,

**percentile\_cont**(0.33) **within** **group** (

**order** **by** *cnt*) **as** *frequency\_per33*,

**percentile\_cont**(0.66) **within** **group** (

**order** **by** *cnt*) **as** *frequency\_per66*,

**percentile\_cont**(0.33) **within** **group** (

**order** **by** *delta\_dt*) **as** *recency\_per33*,

**percentile\_cont**(0.66) **within** **group** (

**order** **by** *delta\_dt*) **as** *recency\_per66*

**from**

*data*

),

-- Каждому клиенту присвоим соответствующий рейтинг в 3 категориях

*rfm\_long* **as** (

**select**

*d*.card,

**case**

**when** *d*.*delta\_dt* < *p*.*recency\_per33* **then** 1

**when** *d*.*delta\_dt* < *p*.*recency\_per66* **then** 2

**else** 3

**end** *recency*,

**case**

**when** *d*.*cnt* <= *p*.*frequency\_per33* **then** 3

**when** *d*.*cnt* <= *p*.*frequency\_per66* **then** 2

**else** 1

**end** *frequency*,

**case**

**when** *d*.*sm* < *p*.*monetary\_per33* **then** 3

**when** *d*.*sm* < *p*.*monetary\_per66* **then** 2

**else** 1

**end** *monetary*

**from**

*data* *d*

**cross** **join** *percentile* *p*

),

-- Сформируем признак rfm в виде кода и словесного описания

*rfm\_short* **as** (

**select**

*rfml*.card,

*d*.*delta\_dt*,

*d*.*cnt*,

*d*.*sm*,

**concat**(*recency*, *frequency*, *monetary*) **as** *rfm*,

**case**

**when** *recency* = 1 **then** 'Недавно'

**when** *recency* = 2 **then** 'Давно'

**when** *recency* = 3 **then** 'Очень давно'

**end** *r*,

**case**

**when** *frequency* = 1 **then** 'Покупают часто'

**when** *frequency* = 2 **then** 'Средняя частота'

**when** *frequency* = 3 **then** 'Покупают редко'

**end** *f*,

**case**

**when** *monetary* = 1 **then** 'Большая сумма'

**when** *monetary* = 2 **then** 'Средняя сумма'

**when** *monetary* = 3 **then** 'Маленькая сумма'

**end** *m*

**from**

*rfm\_long* *rfml*

**left** **join** *data* *d* **on**

*d*.card = *rfml*.card

**order** **by**

*rfm*

)

-- Сгруппируем клиентов по сегментам, присвоим рейтинг (num) для сортировки

(

**select**

**count**(**distinct** card) **as** *number\_of\_clients*,

**avg**(*cnt*) **as** *avg\_cnt*,

**avg**(*sm*) **as** *avg\_sm*,

**avg**(*delta\_dt*) **as** *avg\_dt*,

*rfm*,

**concat**(*r*, ' /', *f*, ' /', *m*) **as** *"Сегмент\_описание"*,

**case**

**when** *rfm* = '111' **then** 'Ключевые'

**when** *rfm* **in** ('112', '113', '122') **then** 'Лояльные'

**when** *rfm* **in** ('121', '211', '221') **then** 'Крупные'

**when** *rfm* **in** ('212', '213', '222', '223', '232', '233') **then** 'Неактивные'

**when** *rfm* **in** ('231', '311', '321', '331') **then** 'Крупные неактивные'

**when** *rfm* **in** ('123', '131', '132', '133') **then** 'Новые'

**when** *rfm* **in** ('312', '313', '322', '323', '332') **then** 'Почти потерянные'

**when** *rfm* **like** '333' **then** 'Ушедшие'

**end** *segment*,

**case**

**when** *rfm* = '111' **then** 1

**when** *rfm* **in** ('112', '113', '122') **then** 2

**when** *rfm* **in** ('121', '211', '221') **then** 3

**when** *rfm* **in** ('212', '213', '222', '223', '232', '233') **then** 4

**when** *rfm* **in** ('231', '311', '321', '331') **then** 5

**when** *rfm* **in** ('123', '131', '132', '133') **then** 6

**when** *rfm* **in** ('312', '313', '322', '323', '332') **then** 7

**when** *rfm* **like** '333' **then** 8

**end** *num*

**from**

*rfm\_short*

**group** **by**

*rfm*,

*segment*,

*num*,

*"Сегмент\_описание"*

**order** **by**

*num*,

*rfm*

)

-- Объединим с не идентифицированными клиентами

**union** **all**

(

**select**

**count**(**distinct** card),

**avg**(*cnt*),

**avg**(*sm*),

**avg**(*delta\_dt*),

'-',

'Без карты',

'Без карты',

9

**from**

*unident*

)

### Запрос 4. Объем выручки в разрезе rfm-сегментов

Посмотрим объем выручки в разрезе полученных сегментов.

Код запроса:

-- Соберем данные по идентифицированным клиентам

**with** *data* **as** (

**select**

card,

**sum**(summ) **as** *sm*,

**count**(\*) **as** *cnt*,

**max**(datetime::**date**) **as** *mx\_dt*,

**make\_date**(2022, 6, 10) - **max**(datetime::**date**) **as** *delta\_dt*

**from**

bonuscheques

**where**

**char\_length**(card) <= 13

**group** **by**

card

),

-- Рассчитаем перцентили

*percentile* **as** (

**select**

**percentile\_cont**(0.33) **within** **group** (

**order** **by** *sm*) **as** *monetary\_per33*,

**percentile\_cont**(0.66) **within** **group** (

**order** **by** *sm*) **as** *monetary\_per66*,

**percentile\_cont**(0.33) **within** **group** (

**order** **by** *cnt*) **as** *frequency\_per33*,

**percentile\_cont**(0.66) **within** **group** (

**order** **by** *cnt*) **as** *frequency\_per66*,

**percentile\_cont**(0.33) **within** **group** (

**order** **by** *delta\_dt*) **as** *recency\_per33*,

**percentile\_cont**(0.66) **within** **group** (

**order** **by** *delta\_dt*) **as** *recency\_per66*

**from**

*data*

),

-- Каждому клиенту присвоим соответствующий рейтинг в 3 категориях

*rfm\_long* **as** (

**select**

*d*.card,

**case**

**when** *d*.*delta\_dt* < *p*.*recency\_per33* **then** 1

**when** *d*.*delta\_dt* < *p*.*recency\_per66* **then** 2

**else** 3

**end** *recency*,

**case**

**when** *d*.*cnt* <= *p*.*frequency\_per33* **then** 3

**when** *d*.*cnt* <= *p*.*frequency\_per66* **then** 2

**else** 1

**end** *frequency*,

**case**

**when** *d*.*sm* < *p*.*monetary\_per33* **then** 3

**when** *d*.*sm* < *p*.*monetary\_per66* **then** 2

**else** 1

**end** *monetary*

**from**

*data* *d*

**cross** **join** *percentile* *p*

),

*rfm\_short* **as** (

**select**

*rfml*.card,

*d*.*delta\_dt*,

*d*.*cnt*,

*d*.*sm*,

**concat**(*recency*, *frequency*, *monetary*) **as** *rfm*

**from**

*rfm\_long* *rfml*

**left** **join** *data* *d* **on**

*d*.card = *rfml*.card

**order** **by**

*rfm*

)

-- Рассчитаем выручку и сгруппируем клиентов по сегментам, присвоим рейтинг (num) для сортировки

**select**

**round**(**sum**(*sm*)/ 1000) **as** *total\_sm*,

*rfm*,

**case**

**when** *rfm* = '111' **then** 'Ключевые'

**when** *rfm* **in** ('112', '113', '122') **then** 'Лояльные'

**when** *rfm* **in** ('121', '211', '221') **then** 'Крупные'

**when** *rfm* **in** ('212', '213', '222', '223', '232', '233') **then** 'Неактивные'

**when** *rfm* **in** ('231', '311', '321', '331') **then** 'Крупные неактивные'

**when** *rfm* **in** ('123', '131', '132', '133') **then** 'Новые'

**when** *rfm* **in** ('312', '313', '322', '323', '332') **then** 'Почти потерянные'

**when** *rfm* **like** '333' **then** 'Ушедшие'

**end** *"Сегмент"*,

**case**

**when** *rfm* = '111' **then** 1

**when** *rfm* **in** ('112', '113', '122') **then** 2

**when** *rfm* **in** ('121', '211', '221') **then** 3

**when** *rfm* **in** ('212', '213', '222', '223', '232', '233') **then** 4

**when** *rfm* **in** ('231', '311', '321', '331') **then** 5

**when** *rfm* **in** ('123', '131', '132', '133') **then** 6

**when** *rfm* **in** ('312', '313', '322', '323', '332') **then** 7

**when** *rfm* **like** '333' **then** 8

**end** *num*

**from**

*rfm\_short*

**group** **by**

*rfm*,

*"Сегмент"*,

*num*

**order** **by**

*num*,

*rfm*

### Запрос 5. Сводная таблица по основным показателям в разрезе сегментов

Выведем основные показатели в разрезе rfm-сегментов: выручка, количество уникальных клиентов, количество заказов, средняя стоимость заказа (AOV), средний доход на пользователя (ARPU).

Код запроса:

-- Соберем данные по идентифицированным клиентам

**with** *data* **as** (

**select**

card,

**sum**(summ) **as** *sm*,

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

**max**(datetime::**date**) **as** *mx\_dt*,

**make\_date**(2022, 6, 10) - **max**(datetime::**date**) **as** *delta\_dt*

**from**

bonuscheques

**where**

**length**(card) <= 13

**group** **by**

card

),

-- В отдельную группу вынесем клиентов, которых невозможно идентифицировать

*unident* **as** (

**select**

card,

**sum**(summ) **as** *sm*,

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

**max**(datetime::**date**) **as** *mx\_dt*,

**make\_date**(2022, 6, 10) - **max**(datetime::**date**) **as** *delta\_dt*

**from**

bonuscheques

**where**

**length**(card) > 13

**group** **by**

card

),

-- Рассчитаем перцентили

*percentile* **as** (

**select**

**percentile\_cont**(0.33) **within** **group** (

**order** **by** *sm*) **as** *monetary\_per33*,

**percentile\_cont**(0.66) **within** **group** (

**order** **by** *sm*) **as** *monetary\_per66*,

**percentile\_cont**(0.33) **within** **group** (

**order** **by** *cnt*) **as** *frequency\_per33*,

**percentile\_cont**(0.66) **within** **group** (

**order** **by** *cnt*) **as** *frequency\_per66*,

**percentile\_cont**(0.33) **within** **group** (

**order** **by** *delta\_dt*) **as** *recency\_per33*,

**percentile\_cont**(0.66) **within** **group** (

**order** **by** *delta\_dt*) **as** *recency\_per66*

**from**

*data*

),

-- Каждому клиенту присвоим соответствующий рейтинг в 3 категориях

*rfm\_long* **as** (

**select**

*d*.card,

**case**

**when** *d*.*delta\_dt* < *p*.*recency\_per33* **then** 1

**when** *d*.*delta\_dt* < *p*.*recency\_per66* **then** 2

**else** 3

**end** *recency*,

**case**

**when** *d*.*cnt* <= *p*.*frequency\_per33* **then** 3

**when** *d*.*cnt* <= *p*.*frequency\_per66* **then** 2

**else** 1

**end** *frequency*,

**case**

**when** *d*.*sm* < *p*.*monetary\_per33* **then** 3

**when** *d*.*sm* < *p*.*monetary\_per66* **then** 2

**else** 1

**end** *monetary*

**from**

*data* *d*

**cross** **join** *percentile* *p*

),

-- Сформируем признак rfm в виде кода

*rfm\_short* **as** (

**select**

*rfml*.card,

*d*.*delta\_dt*,

*d*.*cnt*,

*d*.*sm*,

**concat**(*recency*, *frequency*, *monetary*) **as** *rfm*

**from**

*rfm\_long* *rfml*

**left** **join** *data* *d* **on**

*d*.card = *rfml*.card

**order** **by**

*rfm*

)

-- Рассчитаем необходимые метрики и сгруппируем клиентов по сегментам, присвоим рейтинг (num) для сортировки

**select**

**sum**(*sm*)\* 1.0 / **sum**(*cnt*) **as** *aov*,

**sum**(*sm*)\* 1.0 / **count**(**distinct** card) **as** *ARPU*,

**sum**(*sm*) **as** *"Продажи"*,

**count**(card) **as** *"Клиенты"*,

**sum**(*cnt*) **as** *"Заказы"*,

*rfm*,

**case**

**when** *rfm* = '111' **then** 'Ключевые'

**when** *rfm* **in** ('112', '113', '122') **then** 'Лояльные'

**when** *rfm* **in** ('121', '211', '221') **then** 'Крупные'

**when** *rfm* **in** ('212', '213', '222', '223', '232', '233') **then** 'Неактивные'

**when** *rfm* **in** ('231', '311', '321', '331') **then** 'Крупные неактивные'

**when** *rfm* **in** ('123', '131', '132', '133') **then** 'Новые'

**when** *rfm* **in** ('312', '313', '322', '323', '332') **then** 'Почти потерянные'

**when** *rfm* **like** '333' **then** 'Ушедшие'

**end** *segment*

**from**

*rfm\_short*

**group** **by**

*segment*,

*rfm*

**union** **all**

**select**

**sum**(*sm*)\* 1.0 / **sum**(*cnt*),

**sum**(*sm*)\* 1.0 / **count**(**distinct** card),

**sum**(*sm*),

**count**(**distinct** card),

**sum**(*cnt*),

'-',

'Без карты'

**from**

*unident*

### Запрос 6. Количество клиентов в разрезе сегментов по месяцам

Посчитаем количество клиентов в каждом сегменте, проведя помесячный RFM-анализ. Количество клиентов будем учитывать накопительным итогом, чтобы была возможность отследить “почти потерянных” и “ушедших” клиентов, а также “новых” клиентов. В качестве допущения все клиенты, совершившие 1 покупку в первом месяце рассматриваемого периода (июль 2021) отнесены в категорию “новые”. В качестве “почти потерянных” будем учитывать клиентов, у которых с момента последней покупки прошло не более 2 месяцев, “ушедшими” будем считать тех, у кого прошло 3 и более месяца с даты последней покупки. “Новые” клиенты - клиенты совершившие первую и единственную покупку в рассматриваемом месяце.

Для Frequency зададим следующие границы: R<1 - покупают мало, R<2 - покупают средне, R>=2 - покупают много. Границы определены исходя из того, что в среднем 69% от всех покупок это покупки 1 раз в месяц.

Код запроса:

-- Формируем список месяцев за рассматриваемый период

**with** *months\_list* **as** (

**select**

\*

**from**

**generate\_series**(

**date\_trunc**('month', (**select** **min**(datetime) **from** bonuscheques)),

**date\_trunc**('month', (**select** **max**(datetime) **from** bonuscheques)),

'1 month') **as** mnth

),

-- Для каждого клиента определим дату первой покупки за весь период

min\_dt\_client **as** (

**select**

card,

**min**(datetime) **as** mn\_dt

**from**

bonuscheques

**where**

**char\_length**(card) <= 13

**group** **by**

card

),

-- Сформируем список клиентов по всем месяцам с месяца первой покупки для дальнейшего анализа

client\_months **as** (

**select**

**distinct** c.card,

m.mnth

**from**

(

**select**

**distinct** card

**from**

bonuscheques

**where**

**char\_length**(card) <= 13) c

**cross** **join** months\_list m

**left** **join** min\_dt\_client mndt **on**

mndt.card = c.card

**where**

m.mnth >= **date\_trunc**('month', mndt.mn\_dt)

),

-- Собираем данные по идентифицированным клиентам

ident **as** (

**select**

card,

**date\_trunc**('month', datetime) **as** mnth,

**sum**(summ) **as** sm,

**count**(\*) **as** cnt,

**max**(datetime::**date**) **as** mx\_dt,

**extract**('day' **from** (**date\_trunc**('month', **max**(datetime::**date**)) + **interval** '1 month' - **interval** '1 day' - **max**(datetime::**date**))) **as** delta\_dt

**from**

bonuscheques

**where**

**true**

**and** **char\_length**(card) <= 13

**group** **by**

mnth,

card

),

-- В отдельную группу вынесем клиентов, которых невозможно идентифицировать

unident **as** (

**select**

card,

**date\_trunc**('month', datetime) **as** mnth,

**sum**(summ) **as** sm,

**count**(\*) **as** cnt,

**max**(datetime::**date**) **as** mx\_dt,

**extract**('day' **from** **date\_trunc**('month', **max**(datetime::**date**)) + **interval** '1 month' - **max**(datetime::**date**)) **as** delta\_dt

**from**

bonuscheques

**where**

**true**

**and** **char\_length**(card) > 13

**group** **by**

mnth,

card

),

-- Рассчитаем перцентили (frequency рассмотрели отдельно)

percentile **as** (

**select**

mnth,

**percentile\_cont**(0.33) **within** **group** (

**order** **by**

sm) **as** monetary\_per33,

**percentile\_cont**(0.66) **within** **group** (

**order** **by**

sm) **as** monetary\_per66,

**percentile\_cont**(0.33) **within** **group** (

**order** **by**

delta\_dt) **as** recency\_per33,

**percentile\_cont**(0.66) **within** **group** (

**order** **by**

delta\_dt) **as** recency\_per66

**from**

ident

**group** **by**

mnth

),

-- К списку клиентов по месяцам добавляем в каждый месяц дату последней покупки и считаем кол-во месяцев с последней покупки

**data** **as** (

**select**

c.card,

c.mnth,

**max**(mx\_dt) **over**(**partition** **by** c.card **order** **by** c.mnth, c.card **rows** **between** **unbounded** **preceding** **and** **current** **row**) **as** last\_dt,

**extract**('month' **from** **age**(c.mnth, **date\_trunc**('month', **max**(mx\_dt) **over**(**partition** **by** c.card **order** **by** c.mnth, c.card **rows** **between** **unbounded** **preceding** **and** **current** **row**)))) **as** months\_since\_last\_purchase,

i.sm,

i.cnt

**from**

client\_months c

**left** **join** ident i **on**

i.card = c.card

**and** i.mnth = c.mnth

**order** **by**

card

),

-- Каждому клиенту присвоим соответствующий рейтинг в 3 категориях

rfm\_long **as** (

**select**

d.card,

d.mnth,

last\_dt,

mndt.mn\_dt **as** first\_dt,

months\_since\_last\_purchase,

d.sm,

d.cnt,

i.delta\_dt,

**case**

**when** i.delta\_dt < p.recency\_per33 **then** 1

**when** i.delta\_dt < p.recency\_per66 **then** 2

**else** 3

**end** recency,

**case**

**when** d.cnt < 1 **then** 3

**when** d.cnt < 2 **then** 2

**else** 1

**end** frequency,

**case**

**when** d.sm < p.monetary\_per33 **then** 3

**when** d.sm < p.monetary\_per66 **then** 2

**else** 1

**end** monetary

**from**

**data** d

**left** **join** ident i **on**

i.card = d.card

**and** i.mnth = d.mnth

**left** **join** percentile p **on**

p.mnth = d.mnth

**left** **join** min\_dt\_client mndt **on**

d.card = mndt.card

),

-- Распределяем по сегментам с учетом накопительного итога

final\_data **as** (

**select**

\*,

**concat**(recency, frequency, monetary) **as** rfm,

**case**

**when** last\_dt = first\_dt

**and** cnt = 1 **then** 'Новые'

**when** months\_since\_last\_purchase >= 3 **then** 'Ушедшие'

**when** months\_since\_last\_purchase **in** (1, 2) **then** 'Почти потерянные'

**end** segment,

**case**

**when** **concat**(recency, frequency, monetary) = '111' **then** 'Ключевые'

**when** **concat**(recency, frequency, monetary) **in** ('112', '113', '122') **then** 'Лояльные'

**when** **concat**(recency, frequency, monetary) **in** ('121', '211', '221') **then** 'Крупные'

**when** **concat**(recency, frequency, monetary) **in** ('212', '213', '222', '223', '232', '233') **then** 'Неактивные'

**when** **concat**(recency, frequency, monetary) **in** ('231', '311', '321', '331') **then** 'Крупные неактивные'

**when** **concat**(recency, frequency, monetary) **in** ('123', '131', '132', '133') **then** 'Новые'

**when** **concat**(recency, frequency, monetary) **in** ('312', '313', '322', '323', '332') **then** 'Почти потерянные'

**when** **concat**(recency, frequency, monetary) **like** '333' **then** 'Ушедшие'

**end** segment1

**from**

rfm\_long rfml

**order** **by**

card,

mnth

),

-- Формируем единый признак сегмента и объединяем с не идентифицированными клиентами

**final** **as** (

(

**select**

mnth,

**case**

**when** segment **is** **null** **then** segment1

**else** segment

**end** rfm\_segment,

rfm,

**count**(**distinct** card) **as** number\_of\_clients

**from**

final\_data

**group** **by**

mnth,

rfm\_segment,

rfm

**order** **by**

mnth,

rfm\_segment,

rfm)

**union** **all**

(

**select**

u.mnth,

'Зашифрованные',

'-',

(

**select**

**count**(**distinct** u2.card)

**from**

unident u2

**where**

u2.mnth <= u.mnth)

**from**

unident u

**group** **by**

mnth)

)

-- Выводим кол-во клиентов в разрезе сегментов

**select**

mnth,

rfm,

rfm\_segment,

**sum**(number\_of\_clients)

**from**

**final**

**group** **by**

mnth,

rfm,

rfm\_segment

**order** **by**

mnth,

rfm\_segment,

rfm

## Выводы

1. Выявлено большое количество неидентифицированных операций. Одна из первоочередных задач: выяснить, с чем связано такое количество “зашифрованных” записей. Возможно это связано с проблемами с кассовыми аппаратами или с проблемами со связью, или с какими-то иными проблемами. В течение рассматриваемого периода наблюдается положительная динамика и с 53% в июле 2021 доля “зашифрованных” клиентов сократилась до 35%. Однако влияние данного фактора все равно остается очень существенно.
2. Сумма скидок с учетом списаний бонусов составляет всего 0,3%. Согласно открытым данным рекомендуемый уровень скидок составляет около 3-5%. Необходимо более тщательно проанализировать существующую бонусную программу. Увеличение скидок для определенных категорий клиентов позволит эффективно влиять на лояльность клиентов и увеличение продаж. Например, это может позитивно сказаться на увеличении продаж у “лояльных” клиентов, число которых составляет 4%, а доля выручки всего 2%.

**Рекомендации по сегментам:**

Сгруппируем клиентов по сегментам, для каждого сегмента выберем стратегию взаимодействия.

|  |  |  |
| --- | --- | --- |
| Сегмент | Тип клиентов | Стратегия и инструменты |
| 111 | **Ключевые/ «Ядро»:**  Доля клиентов: 8%  Доля выручки: 27%   * заказывали недавно; * часто обращаются; * много тратят | Продемонстрировать заинтересованность со стороны компании:   * VIP-карта; * подарки ко дню рождения; * выявить ключевые ценности и поддерживать их в продукте; * скидки и бонусы не требуются. |
| 112, 113, 122 | **Лояльные:**  Доля клиентов: 4%  Доля выручки: 2%   * заказывали недавно; * покупают часто или средне; * тратят не очень много | Стимулировать покупки:   * накопительные скидки; * кешбэк; * подборки товаров на основе истории заказов * в момент покупки предложить сопутствующие товары |
| 121, 211, 221 | **Крупные/ «Киты»:**  Доля клиентов: 8%  Доля выручки: 13%   * заказывают нерегулярно; * тратят много | Необходимо изучить потребности клиентов:   * получить обратную связь; * кастомизировать предложения; * скидки и бонусы не требуются. |
| 212, 213, 222, 223, 232, 233 | **Неактивные/ «Спящие»:**  Доля клиентов: 14%  Доля выручки: 4%   * давно не обращались в компанию | Возобновить сотрудничество:   * провести информационную рассылку; * разработать стимулирующие акции; * начислить бонусы на карту. |
| 231, 311, 321, 331 | **Крупные неактивные:**  Доля клиентов: 5%  Доля выручки: 6%   * заказывали давно; * тратят много | Возобновить сотрудничество:   * провести информационную рассылку; * разработать стимулирующие акции; * начислить бонусы на карту; * направить небольшой опросный лист, чтобы получить обратную связь и узнать причину прекращения взаимодействия. |
| 123, 131, 132, 133 | **Новые/ «новички»:**  Доля клиентов: 7%  Доля выручки: 2%   * недавно заказали впервые | Наладить крепкие взаимовыгодные отношения:   * знакомить с брендом, рассказывать о продукте; * собирать фидбэк; * оказывать активную помощь в выборе товара, давать полезную информацию; * предлагать акции. |
| 312, 313, 322, 323, 332 | **Почти потерянные/ «уходящие»:**  Доля клиентов: 9%  Доля выручки: 4%     * в прошлом много заказывали, но давно не обращались | Вернуть и сделать постоянными клиентами:   * делать рассылки; * предложить скидку на часто покупаемые товары; * напомнить о покупке; * предложить другие товары на основе рекомендаций; * начислить бонусы. |
| 333 | **Ушедшие:**  ​​Доля клиентов: 8%  Доля выручки: 1%   * заказывали немного; * давно не обращались | * Настроить автоматическую email-рассылку и не вкладываться в привлечение |