**RFM-анализ клиентской базы аптечной сети**

*Анализ исходного сегмента*

1. Для RFM-анализа, собираем информацию по кол-ву дней с ‘2022-06-09 21:49:45.000’ (последняя дата покупки по данным), кол-ву покупок на одного клиента и сумме покупки каждого клиента.

**select**

*b*.card,

**count**(*b*.doc\_id) **over**(**partition** **by** b.card) **as** cnt\_purchuases, --кол-во покупок

**extract**(**day** **from** **max**(*b*.datetime) **over**() - **max**(*b*.datetime) **over**(**partition** **by** *b*.card)) + 1 **as** *days\_diff*, --кол-во дней с момента последней покупки клиента

**sum**(*b*.summ\_with\_disc) **over**(**partition** **by** b.card) **as** **sum** --сумма покупки без учета бонусов (взяла это поле для выявления фактических денег поступивших за счет продажи)

**from** bonuscheques *b*

**where** **substring**(b.card **from** **'^[\d]+$'**) **notnull** --отбираем только валидные номера карт

**order** **by** card

1. Далее я начала смотреть как в целом распределены кол-во клиентов по каждому из параметров.

Для параметра Recency:

**with** *agg\_info* **as** (

**select** **distinct**

*b*.card,

**count**(*b*.doc\_id) **over**(**partition** **by** b.card) **as** cnt\_purchuases, --кол-во покупок

**extract**(**day** **from** **max**(*b*.datetime) **over**() - **max**(*b*.datetime) **over**(**partition** **by** *b*.card)) + 1 **as** *days\_diff*, --кол-во дней с момента последней покупки клиента

**sum**(*b*.summ\_with\_disc) **over**(**partition** **by** b.card) **as** **sum**, --сумма покупки без учета бонусов (взяла это поле для выявления фактических денег поступивших за счет продажи),

**round**(**sum**(*b*.summ\_with\_disc) **over**(**partition** **by** *b*.card) \* 1.0 / **count**(*b*.doc\_id) **over**(**partition** **by** *b*.card), 2) **as** *avg\_sum*

**from** bonuscheques *b*

**where** **substring**(b.card **from** **'^[\d]+$'**) **notnull** --отбираем только валидные номера карт

**order** **by** card

)

**select**

days\_diff,

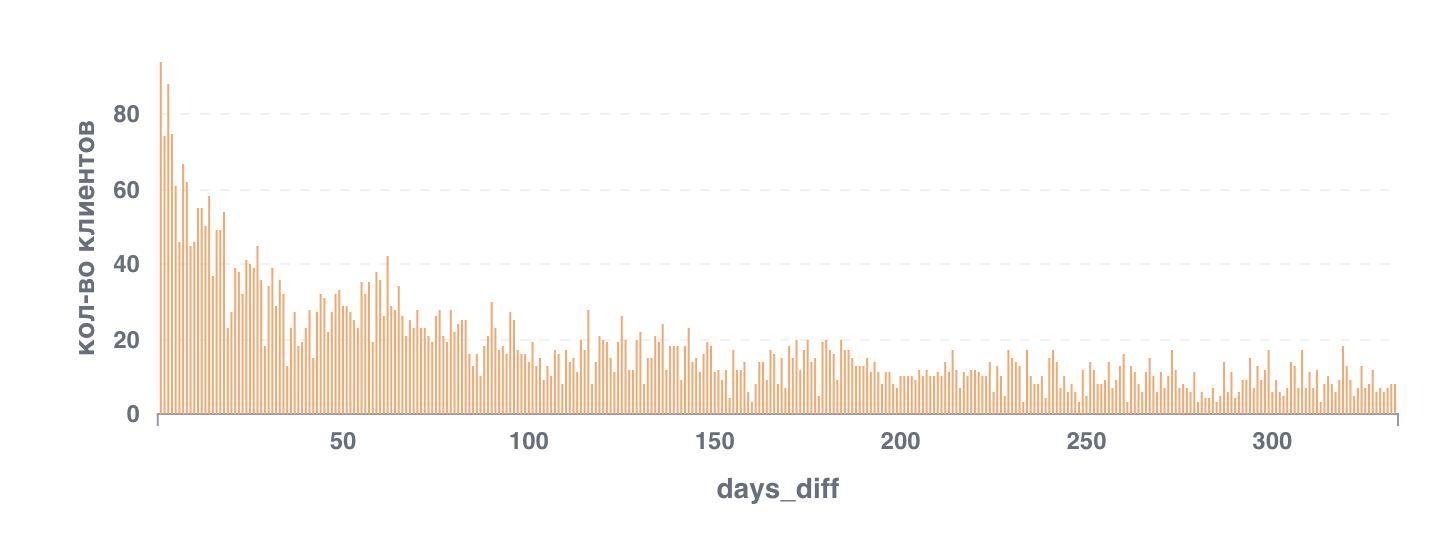
**count**(card) **as** **"среднее значение"**

**from** agg\_info

**group** **by** days\_diff

**order** **by** days\_diff

Вывод - данные не распределены нормально, у больше половины клиентов прошло более 70 дней с момента последней покупки:



**with** *agg\_info* **as** (

**select** **distinct**

*b*.card,

**count**(*b*.doc\_id) **over**(**partition** **by** b.card) **as** cnt\_purchuases, --кол-во покупок

**extract**(**day** **from** **max**(*b*.datetime) **over**() - **max**(*b*.datetime) **over**(**partition** **by** *b*.card)) + 1 **as** *days\_diff*, --кол-во дней с момента последней покупки клиента

**sum**(*b*.summ\_with\_disc) **over**(**partition** **by** b.card) **as** **sum**, --сумма покупки без учета бонусов (взяла это поле для выявления фактических денег поступивших за счет продажи),

**round**(**sum**(*b*.summ\_with\_disc) **over**(**partition** **by** *b*.card) \* 1.0 / **count**(*b*.doc\_id) **over**(**partition** **by** *b*.card), 2) **as** *avg\_sum*

**from** bonuscheques *b*

**where** **substring**(b.card **from** **'^[\d]+$'**) **notnull** --отбираем только валидные номера карт

**order** **by** card

)

**select**

**case**

**when** days\_diff::**integer** / 10 = 0 **then** **'< 10'**

**else** **to\_char**(days\_diff::**integer** / 10 \* 10, **'999'**)

**end** **as** dd,

**count**(**distinct** card) **as** cnt

**from** agg\_info

**group** **by**

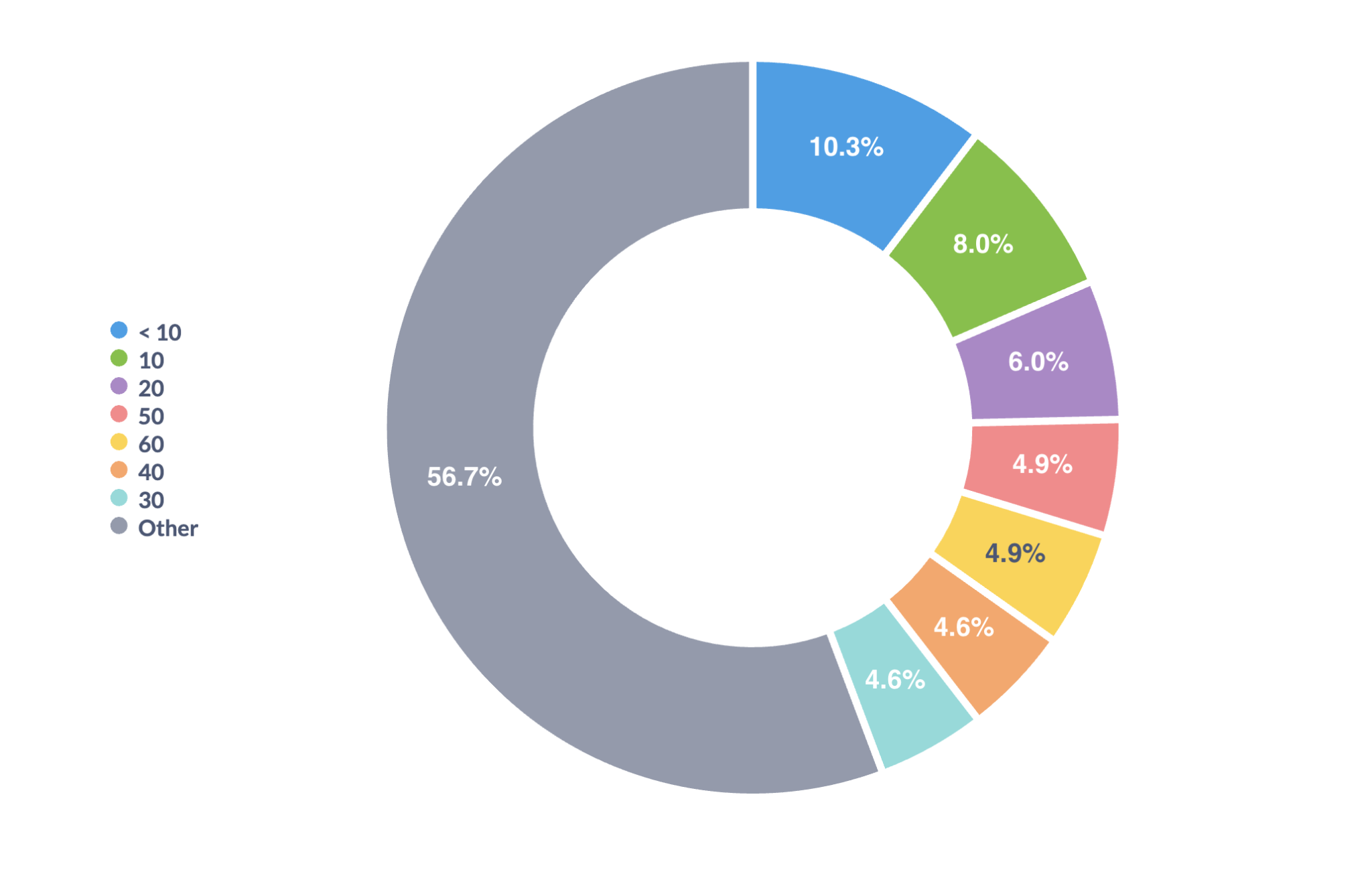
**case**

**when** days\_diff::**integer** / 10 = 0 **then** **'< 10'**

**else** **to\_char**(days\_diff::**integer** / 10 \* 10, **'999'**)

**end**

**order** **by** cnt **desc**



Для параметра Frequency:

**with** *agg\_info* **as** (

**select** **distinct**

*b*.card,

**count**(*b*.doc\_id) **over**(**partition** **by** b.card) **as** cnt\_purchuases, --кол-во покупок

**extract**(**day** **from** **max**(*b*.datetime) **over**() - **max**(*b*.datetime) **over**(**partition** **by** *b*.card)) + 1 **as** *days\_diff*, --кол-во дней с момента последней покупки клиента

**sum**(*b*.summ\_with\_disc) **over**(**partition** **by** b.card) **as** **sum**, --сумма покупки без учета бонусов (взяла это поле для выявления фактических денег поступивших за счет продажи),

**round**(**sum**(*b*.summ\_with\_disc) **over**(**partition** **by** *b*.card) \* 1.0 / **count**(*b*.doc\_id) **over**(**partition** **by** *b*.card), 2) **as** *avg\_sum*

**from** bonuscheques *b*

**where** **substring**(b.card **from** **'^[\d]+$'**) **notnull** --отбираем только валидные номера карт

**order** **by** card

)

**select**

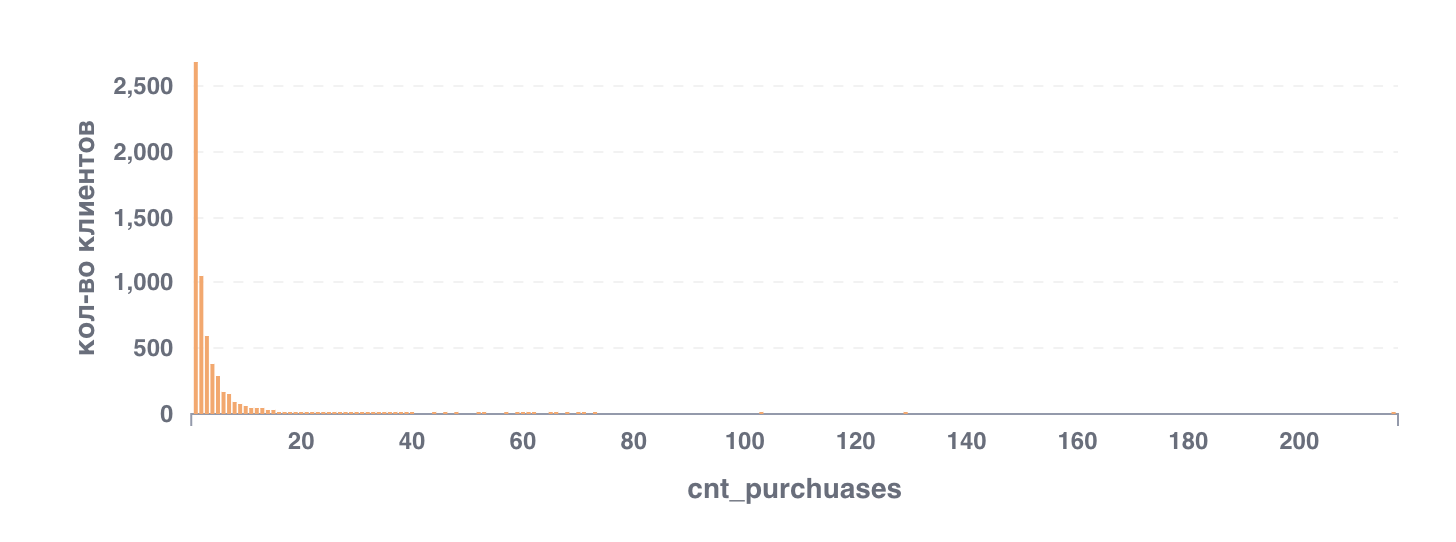
cnt\_purchuases,

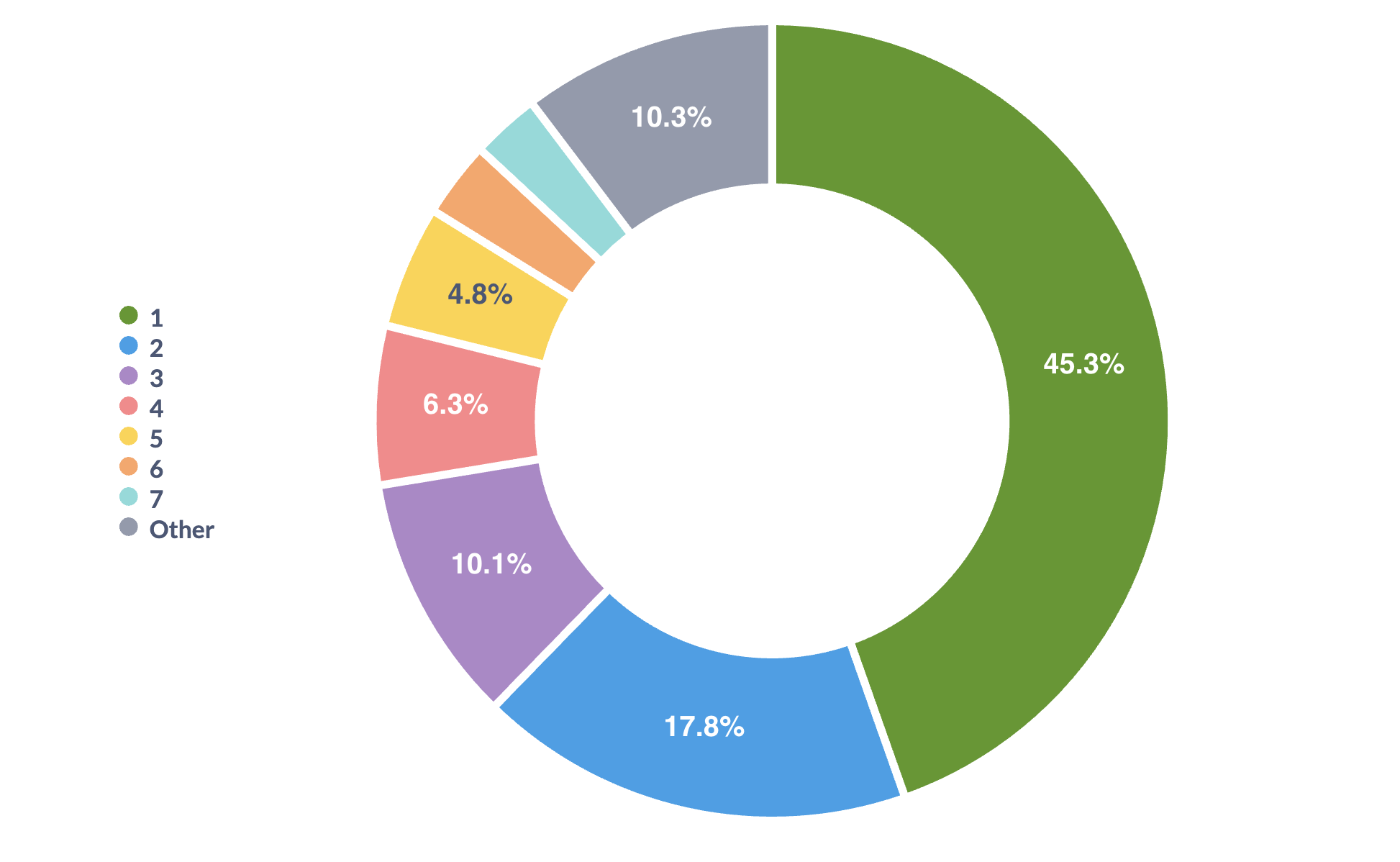
**count**(card)

**from** agg\_info

**group** **by** cnt\_purchuases

Вывод - явно есть большие выбросы в сегменте, большинство клиентов покупали 1-2 раза за все время, но при этом есть клиенты, кто делал до 217 покупок.





По параметру Monetary примерно такая же картина (подробнее в дашборде).

*RFM-анализ и классификация клиентов*

Сложнее всего было определять пороги для каждого из параметров.

Так как есть существенно большая разница между покупателями. Поэтому я решила просегментировать клиентов по методу ABC-анализа и далее оценивать клиентов из каждой группы по RFM-анализу:

**with** *agg\_info* **as** (

**select** **distinct**

*b*.card,

**count**(*b*.doc\_id) **over**(**partition** **by** b.card) **as** cnt\_purchuases, --кол-во покупок

**extract**(**day** **from** **max**(*b*.datetime) **over**() - **max**(*b*.datetime) **over**(**partition** **by** *b*.card)) + 1 **as** *days\_diff*, --кол-во дней с момента последней покупки клиента

**sum**(*b*.summ\_with\_disc) **over**(**partition** **by** b.card) **as** **sum**, --сумма покупки без учета бонусов (взяла это поле для выявления фактических денег поступивших за счет продажи),

**round**(**sum**(*b*.summ\_with\_disc) **over**(**partition** **by** *b*.card) \* 1.0 / **count**(*b*.doc\_id) **over**(**partition** **by** *b*.card), 2) **as** *avg\_sum*

**from** bonuscheques *b*

**where** **substring**(b.card **from** **'^[\d]+$'**) **notnull** --отбираем только валидные номера карт

**order** **by** card

),

abc\_clients **as** (

**select**

a.\*,

(**case**

**when** **sum**(a.**sum**) **over**(**order** **by** a.**sum** **desc**) / **sum**(a.**sum**) **over**() <= 0.8 **then** **'A'**

**when** **sum**(a.**sum**) **over**(**order** **by** a.**sum** **desc**) / **sum**(a.**sum**) **over**() <= 0.95 **then** **'B'**

**else** **'C'**

**end**

) **as** abc

**from** agg\_info a

),

check\_percentile **as** (

**select**

abc,

percentile,

**percentile\_disc**(percentile) **within** **group** (**order** **by** **sum**) perc\_value\_sum,

**percentile\_disc**(percentile) **within** **group** (**order** **by** cnt\_purchuases) perc\_value\_cnt,

**percentile\_disc**(percentile) **within** **group** (**order** **by** cnt\_purchuases) perc\_value\_days

**from** abc\_clients, **generate\_series**(0.0, 1.0, 0.105) **AS** percentile

**group** **by** abc, percentile

),

thresholds **as** (

**select**

abc,

**percentile\_disc**(0.38) **within** **group** (**order** **by** days\_diff) **as** r1,

**percentile\_disc**(0.75) **within** **group** (**order** **by** days\_diff) **as** r2,

**percentile\_disc**(1) **within** **group** (**order** **by** days\_diff) **as** r3,

**percentile\_disc**(1) **within** **group** (**order** **by** cnt\_purchuases) **as** f1,

**percentile\_disc**(0.68) **within** **group** (**order** **by** cnt\_purchuases) **as** f2,

**percentile\_disc**(0.45) **within** **group** (**order** **by** cnt\_purchuases) **as** f3,

**percentile\_disc**(1) **within** **group** (**order** **by** **sum**) **as** m1,

**percentile\_disc**(0.66) **within** **group** (**order** **by** **sum**) **as** m2,

**percentile\_disc**(0.33) **within** **group** (**order** **by** **sum**) **as** m3

**from** abc\_clients

**group** **by** abc

),

rfm\_result **as** (

**select**

a.\*,

t.f3,

t.f2,

(

**case**

**when** a.days\_diff <= t.r1 **then** 1

**when** a.days\_diff <= t.r2 **then** 2

**else** 3

**end**

) **as** recency,

(

**case**

**when** a.cnt\_purchuases <= t.f3 **then** 3

**when** a.cnt\_purchuases <= t.f2 **then** 2

**else** 1

**end**

) **as** frequency,

(

**case**

**when** a.**sum** <= t.m3 **then** 3

**when** a.**sum** <= t.m2 **then** 2

**else** 1

**end**

) **as** monetary,

(**case** **when** a.days\_diff <= t.r1 **then** 1 **when** a.days\_diff <= t.r2 **then** 2 **else** 3 **end**)\*100 +

(**case** **when** a.cnt\_purchuases <= t.f3 **then** 3 **when** a.cnt\_purchuases <= t.f2 **then** 2 **else** 1 **end**)\*10 +

(**case** **when** a.**sum** <= t.m3 **then** 3 **when** a.**sum** <= t.m2 **then** 2 **else** 1 **end**) **as** rfm

**from** abc\_clients a

**join** thresholds t

**on** a.abc = t.abc

)

**select**

rfm,

**count**(card) **as** cnt

**from** rfm\_result

**group** **by** rfm

**order** **by** rfm

*Результат и выводы*

После сравнения результатов, пороги группы A значительно выделяются по сравнению с B и C (возможно, я бы объединила 2 последние группы в один), но мне кажется, чем детальнее разделены сегменты, тем точнее можно будет настроить коммуникации на клиентов.

Если говорить про краткосрочные перспективы:

**Кейс:** не выполняется ежемесячный план по сумме продаж.

**Какие сегменты вовлекаются:** все, у кого в показатель Monetory - 1 и два остальных параметра - 2 или 3 и из группы A; 111/ 121/ 131/ 211/ 221/ 231 из групп B и 111 - из группы C

**Почему:** Это те клиенты, которые в целом оставляют неплохие деньги, но делают покупки не так часто и уже давно.

**Как запустить коммуникации:** Реклама сезонных витаминов, запустить персональную акцию для таких партнеров с ограниченным сроком.

**Кейс:** увеличить рейтинг аптеки (например, в 2ГИС )

**Какие сегменты вовлекаются:**111, клиентов с оценкой 1 по параметру Frequency или клиентов, кто не так давно заходил (оценка 1 по Recency) и сделал покупку на большую сумму (оценка 1 по Monetary) из каждого сегмента

**Почему:** Это клиенты, кто часто или не так давно делал покупку на большую сумму, скорее всего это лояльные клиенты и есть причина по которой они часто пользуются услугами аптеки. Для частых клиентов можно запустить рассылку со ссылкой в 2гис и просьбой поставить оценку в приложении; для “недавних” покупателей можно запустить ту же рассылку с дальнейшей мотивацией вернуться - предоставить определенную скидку на следующую покупку от конкретной суммы (все зависит от бюджета)

Если говорить про долгосрочные перспективы:

**Кейс:** увеличить кол-во покупателей

**Какие сегменты вовлекаются:** здесь можно брать клиентов с сегментами 2 или 3 по каждому из параметров из групп B/C

**Почему:** Грубо говоря, это “лиды”. Я бы детально проанализировала почему они не покупают часто, чтобы понять как с ними работать в дальнейшем и увеличить сегмент лояльных покупателей в аптеке.