



Что влияет на длительность подписки на Tinkoff Premium?

This is fine: Дмитрий Бусыгин, Павел Маляренко,
Роман Локманов, Михаил Боков, а также
менторы - Александр Агафонов и Мария Франк

ТИНЬКОФФ



Структура данных



Финансовая “история” клиента:

- Даты открытия различных счётов: накопительные, вклады, кредиты, инвестиционные и.т.д.
- Количество транзакций расходов и пополнения, а также их суммарные размеры.
- Личные данные: возраст, гражданство, место проживания.

Заявка:

- Дата подачи заявки
- Канал подачи заявки
- Флаг назначения встречи
- Флаг успешного получения карты
- Флаг валидации
- Флаг отказа по заявке
- Флаг заказа металлической карты

Дальнейшие действия:

- Пользовался ли клиент привилегиями (повышенный кэшбек, инкассация, бизнес-зал, Тинькофф Мобайл)
- Открыл ли счета (накопительные, вклады, кредиты, ...)

Что влияет на длительность подписки на Tinkoff Premium?

В интересах компании удержать клиентов на как можно дольший срок. Но для этого нужно понять, когда и почему клиенты отписываются.

Гипотезы

1

Чем больше количество транзакций и их размеры, тем дольше люди будут пользоваться премиумом.

Люди, которые тратят много денег и много получают, скорее всего захотят получить какую-то выгоду со своих расходов в виде кэшбека. Более того, людям с огромным доходом с большей вероятностью позволят себе премиум.

2

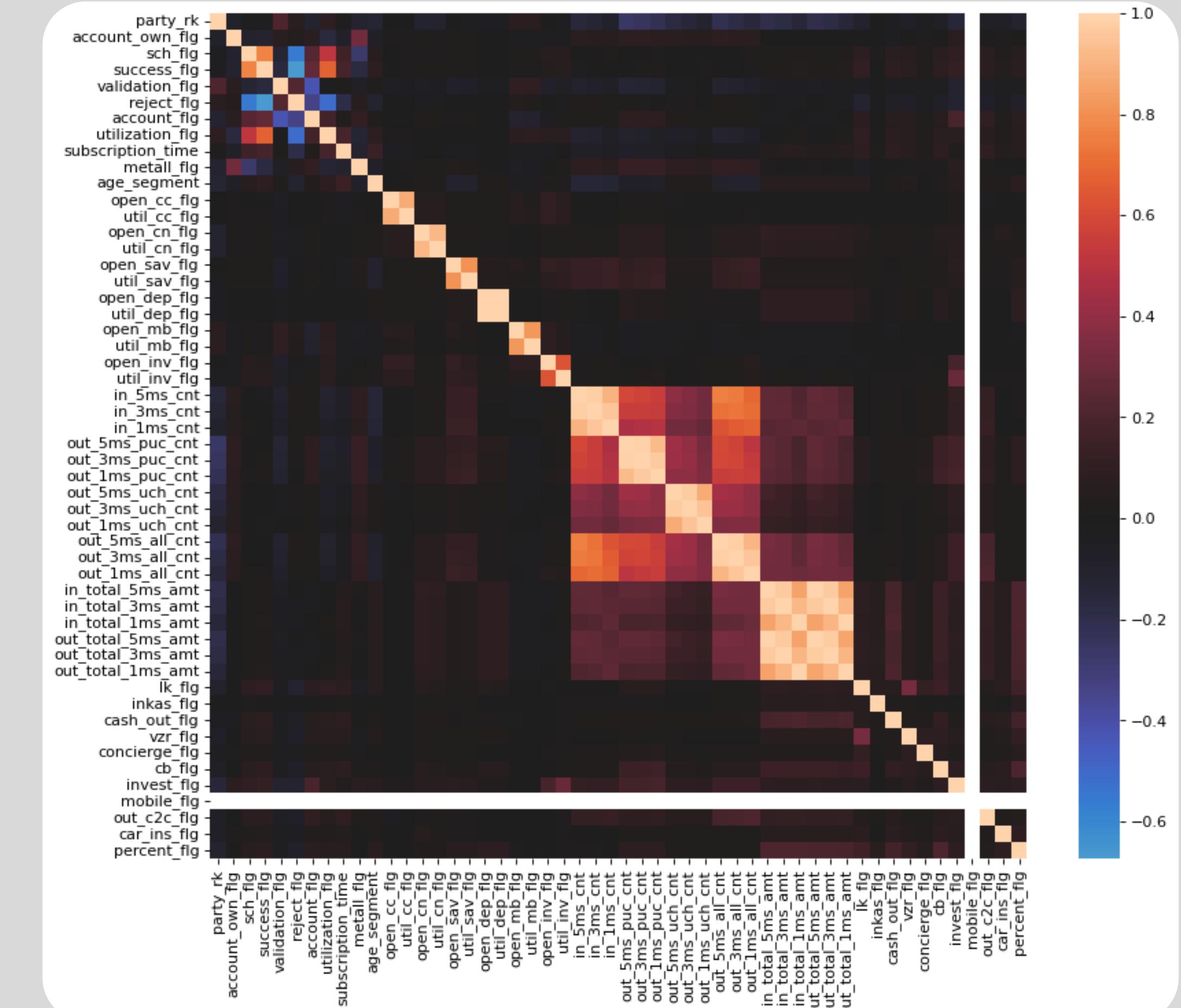
Длительность подписки зависит от того, насколько активно клиент пользуется её преимуществами.

Клиенты, не пользующиеся подпиской, попросту отключат её, так как она им не нужна (трата 1990 руб в никуда). Активные пользователи, наоборот, получат максимальную выгоду от предложения и, возможно, подписка даже окупится.

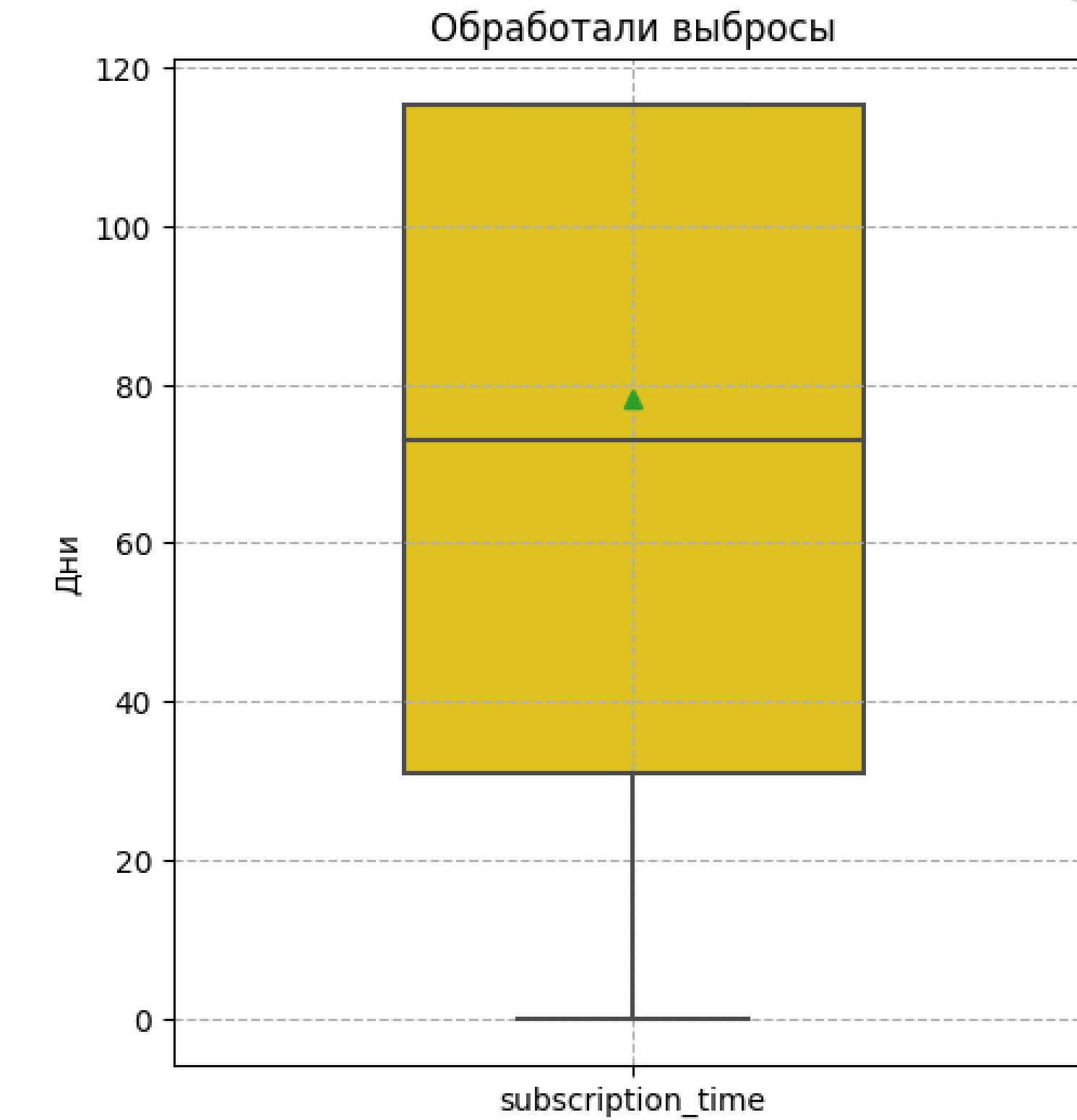
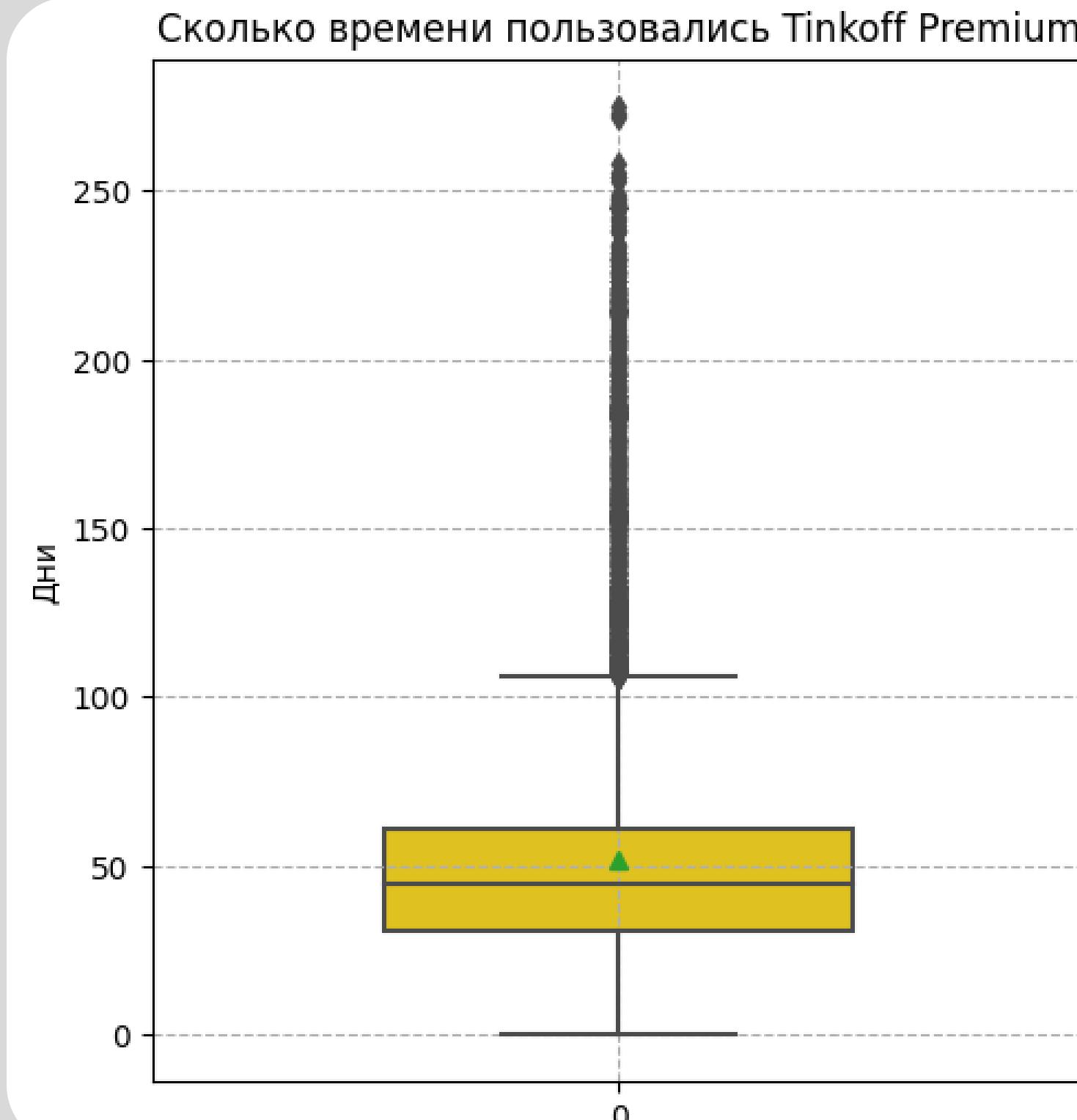
Разведанализ

Корреляций в данных довольно мало. С длительностью подписки вообще ничего не коррелирует, поэтому попробуем обработать данные.

Корреляционная матрица

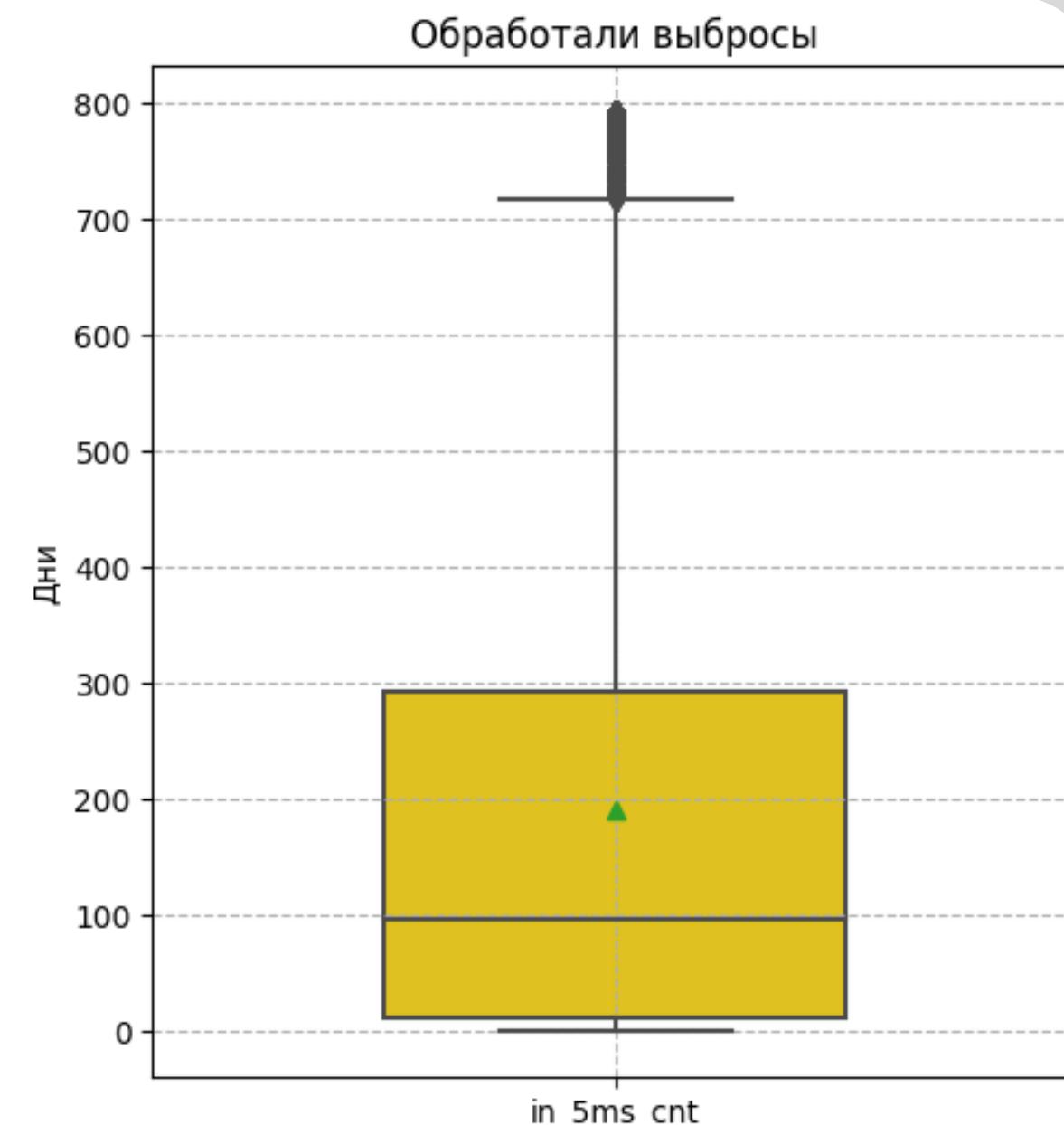
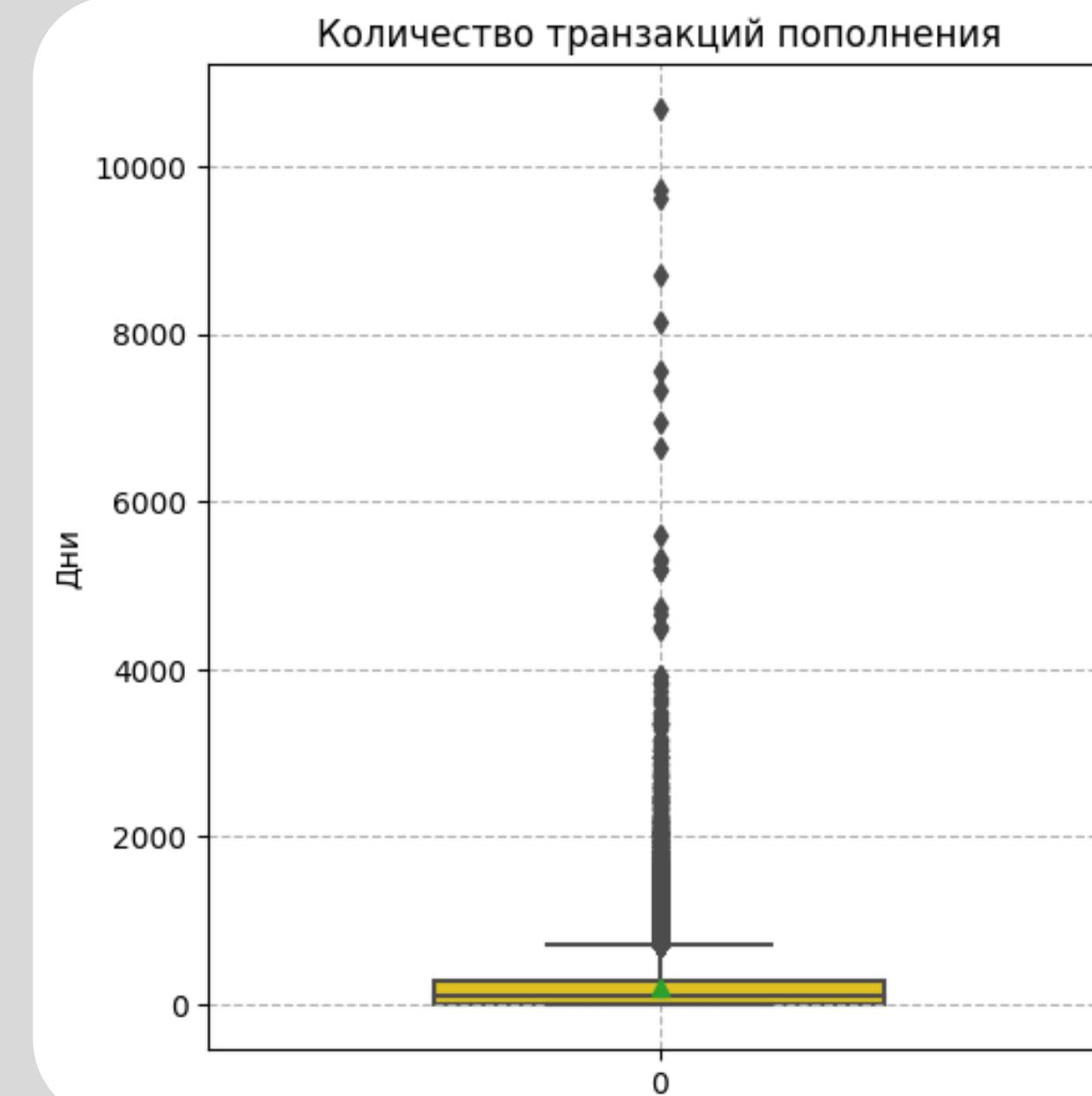
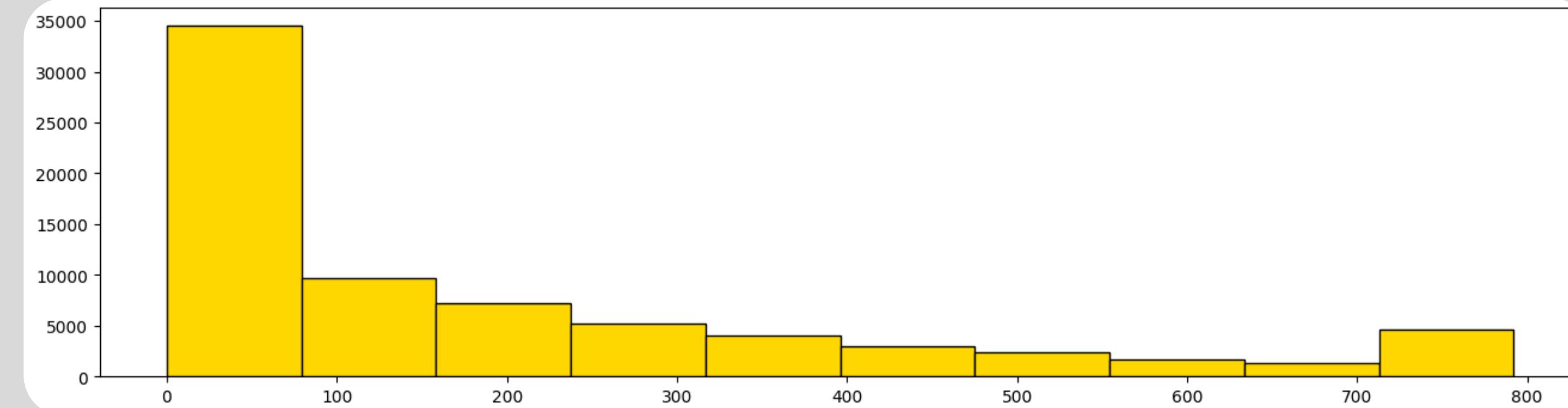


Наша целевая метрика - длительность подписки. Выбросы начинаются после 100 дней. Поэтому мы обрабатывали все, что было больше 95% всех значений - меняли на 95 процентиль.



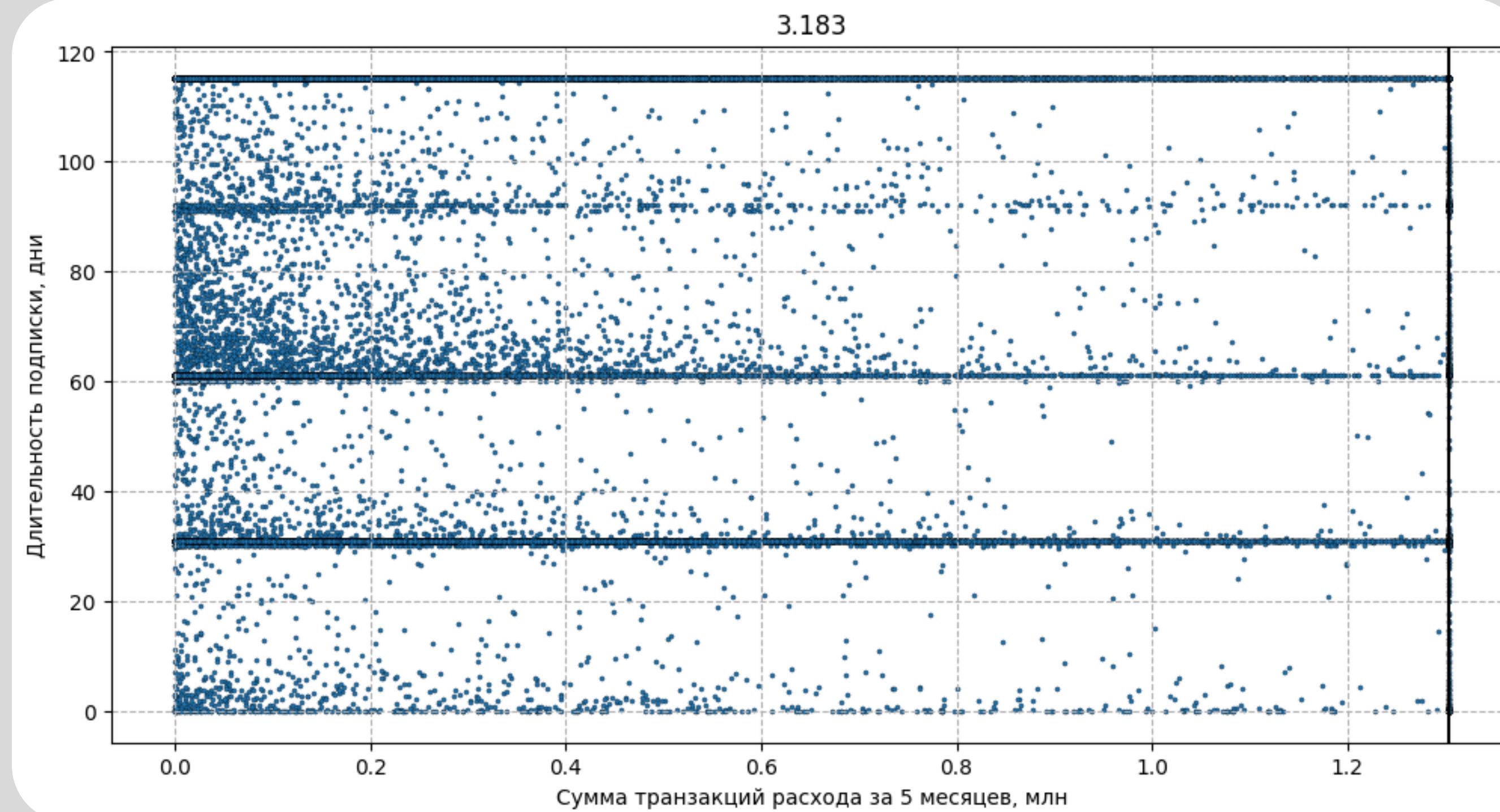
Разведанализ

В колонках с транзакциями тоже оказалось полно аномально высоких значений. У большинства клиентов их кол-во около 100 за последние 5 месяцев, но при этом макс. значение >10000.



Проверка гипотезы 1

Даже после обработки данных, корреляции между длительностью подписки и суммой всех транзакций нет. Поэтому можно сделать микровывод, что гипотеза ошибочна.



Проверка гипотезы 2

Вторая гипотеза заключается в том, что частота использования привилегий может показать, насколько долго он будет подписчиком премиума. Однако, у нас только бинарные значения (да/нет). Тут на помощь приходит feature engineering.

Индекс использования привилегий

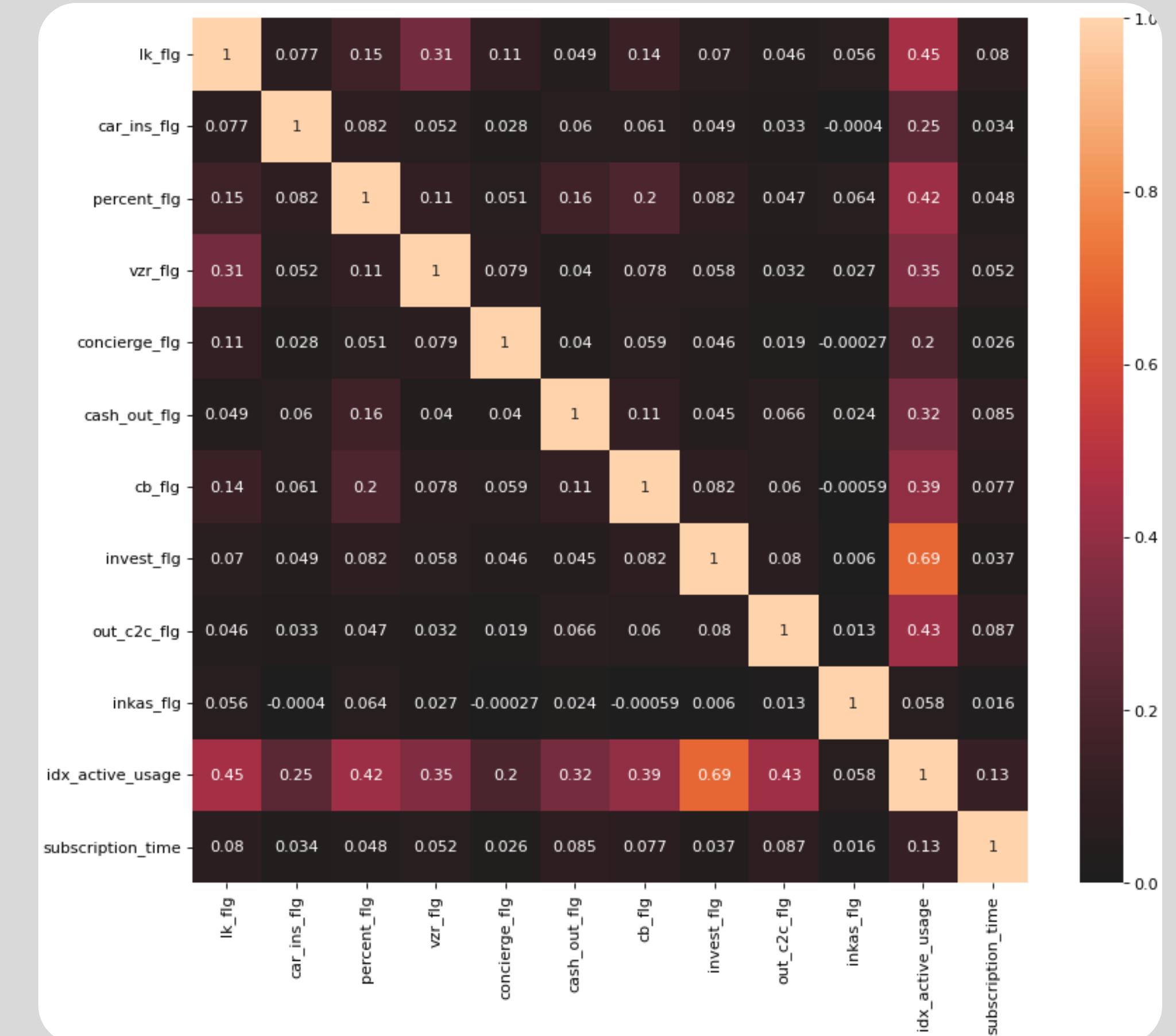
$$\text{Idx} = \text{lk_flg} + \text{inkas_flg} + \text{cashout_flg} + \dots$$

Суть заключается в том, чтобы просуммировать все колонки привилегий.

ТИНЬКОФФ

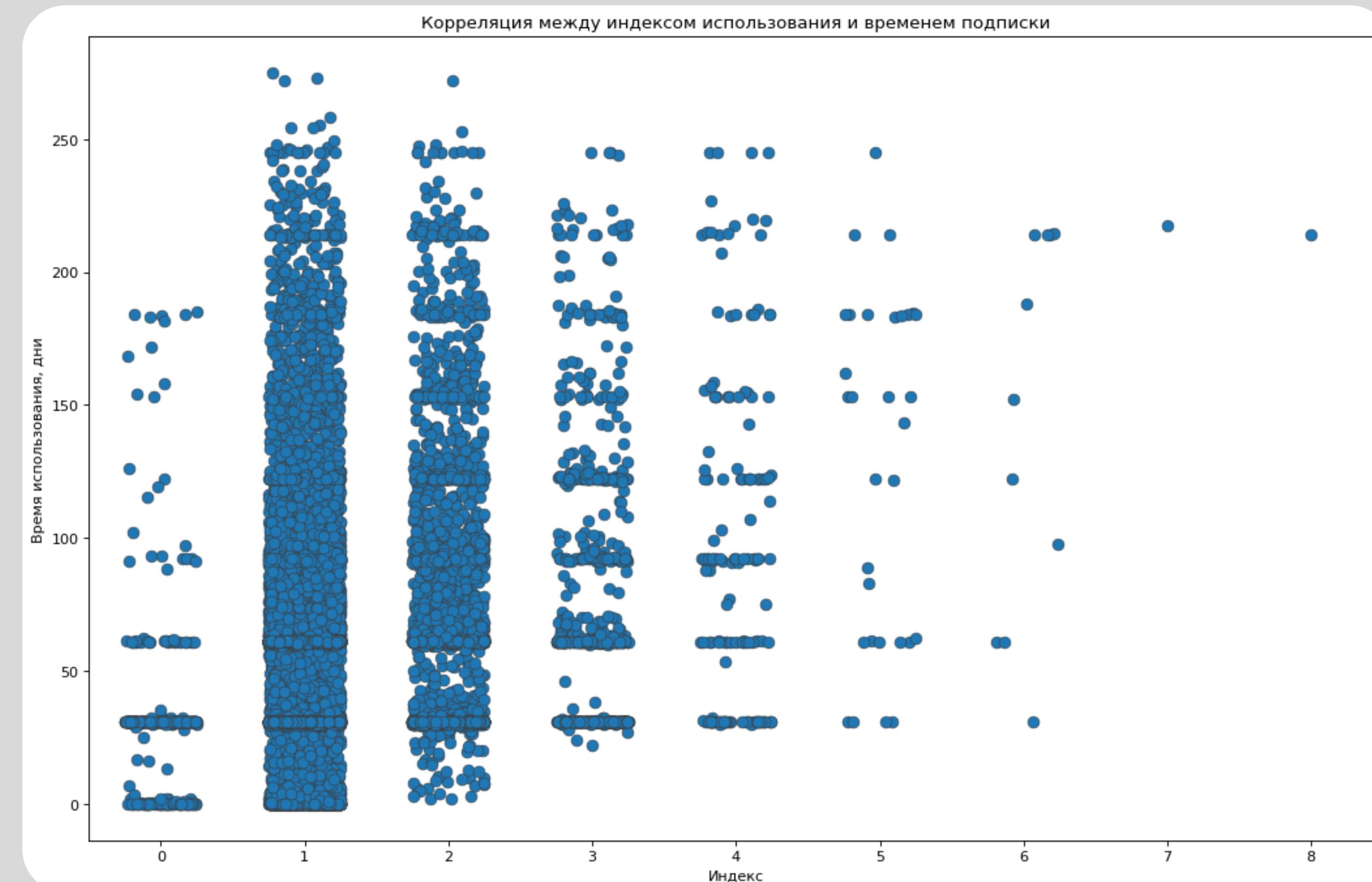
Проверка гипотезы 2

У отдельных значений корреляция с целевой метрикой была не больше 9%. Но у нового значения корреляция с subscription_time целых 13% в датасете с выбросами - незначительная, но это уже результат.



Проверка гипотезы 2

Посмотрим на точечный график. Видно, что нижняя планка повышается с повышением индекса, однако верхняя планка высокая у клиентов с $\text{Idx} = 1$. Можно сделать вывод, что чем активнее человек использует привелегии, тем меньше вероятность, что он рано отключит подписку.

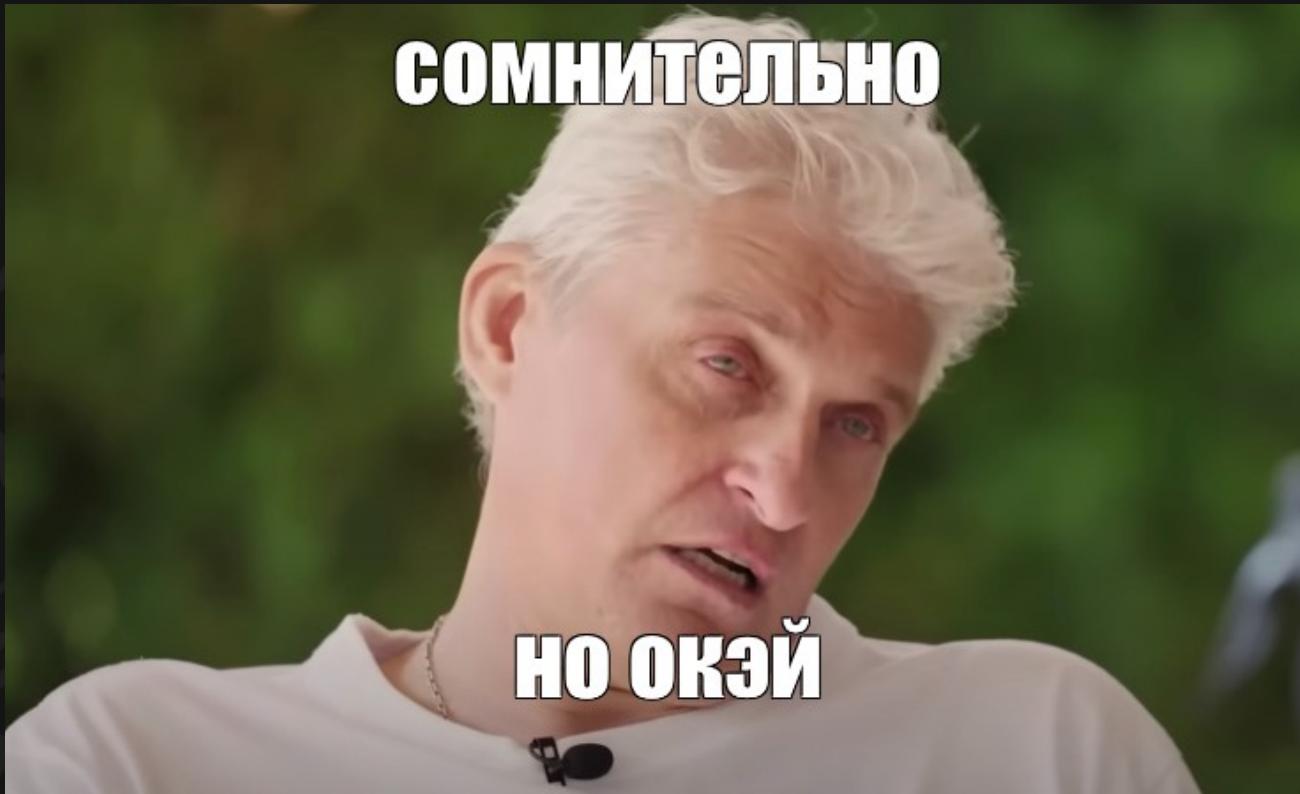


Какие выводы можно сделать



1

Первая гипотеза
опровергнута: количества
транзакций никак не влияет
на длительность подписки



2

Вторая гипотеза имеет место
быть. По активному
пользованию нельзя точно
сказать, сколько времени он
будет подписчиком, однако
можно дать примерную
оценку снизу

Спасибо за внимание

**Михаил
Боков**

Математик-статист

**Павел
Маляренко**

Математик-статист

**Роман
Локманов**

Программист

**Дмитрий
Бусыгин**

Программист (и капитан, но
это не так важно)

**Александр
Агафонов**

Ментор

**Мария
Франк**

Ментор



Ссылка на Colab