

“Алло, ну как там с Тинькофф Премиум?”

ТИНЬКОФФ

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



Введение для Центрального университета



Наша команда проанализировала данные о клиентах Тинькофф Премиум. После участия я поработал над ошибками и немного улучшил проект.
Приятного...просмотра?

ТИНЬКОФФ

Структура данных



В датасете представлены данные о заявках клиентов, которые хотели подключить и подписку и карту в марте 2023 года.

1

Финансовая “история” клиента:

2

Заявки

3

Дальнейшие действия



Финансовое “прошлое” клиента



Данные о пути заявки

Дата подачи заявки



Флаг встречи



Флаг получения карты



Флаг отказа по
заявке

Тинькофф Премиум
одобрен

Какие данные есть?

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



Тинькофф

Дальнейшие действия

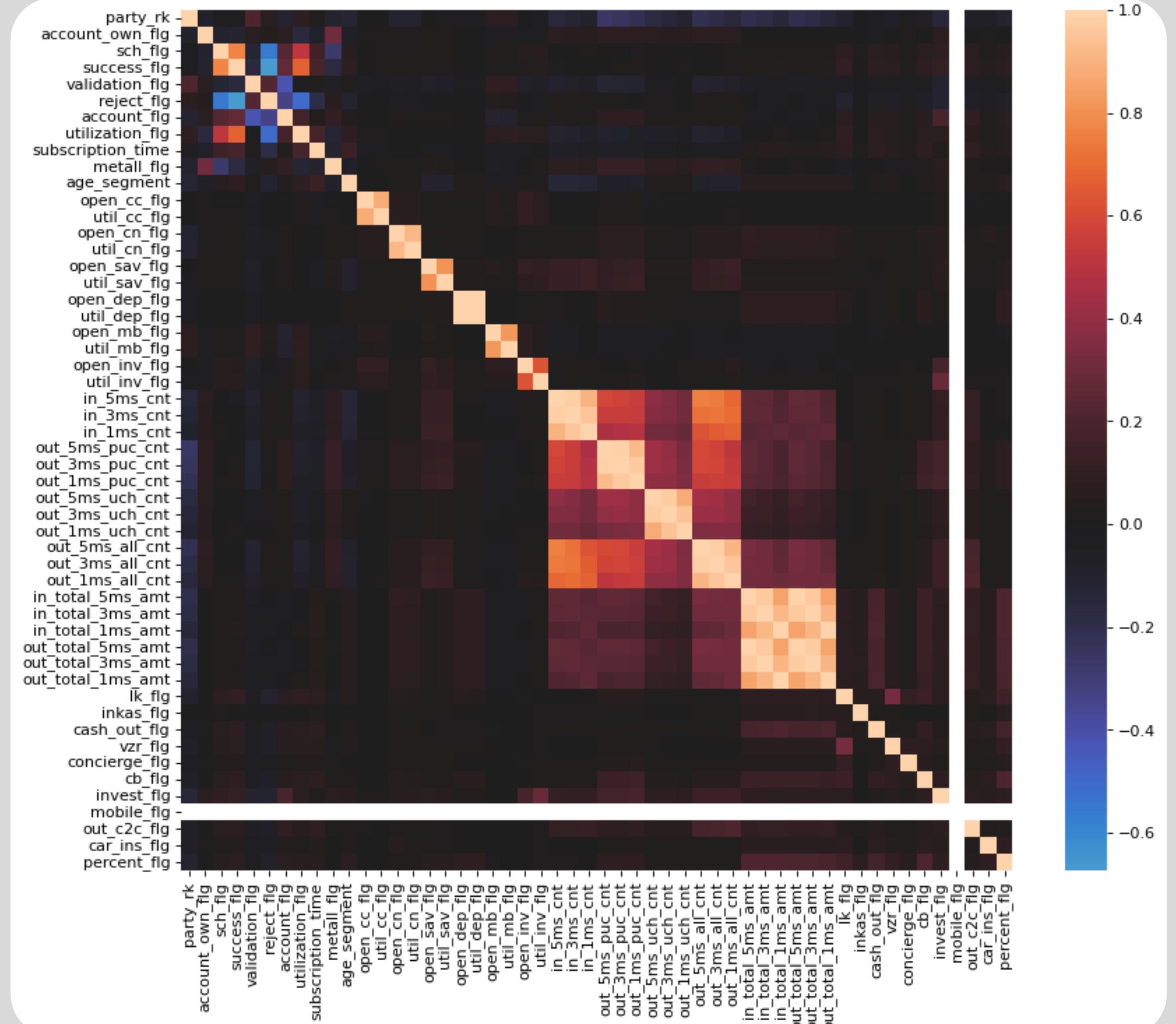


ТИНЬКОФФ

Разведанализ

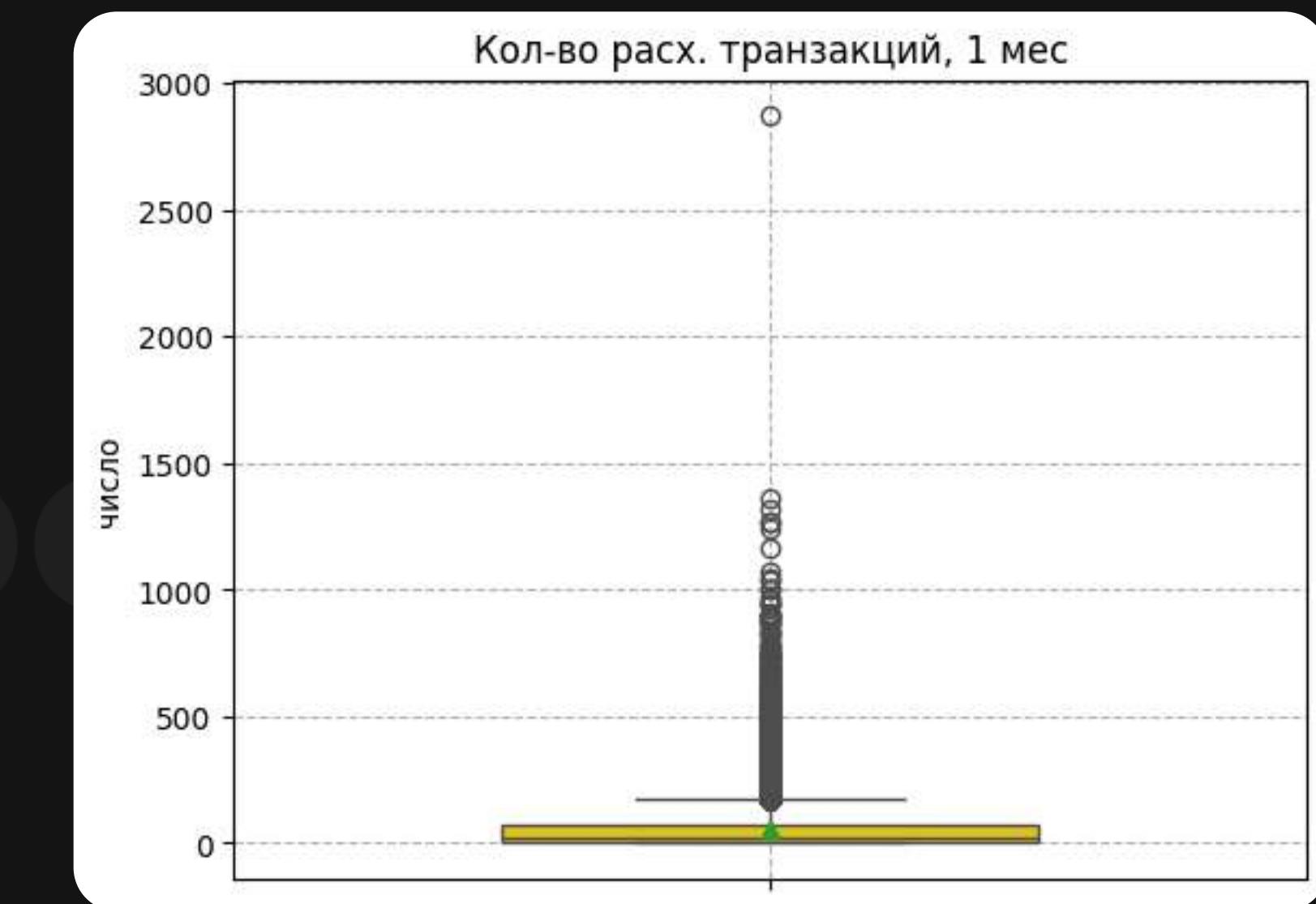
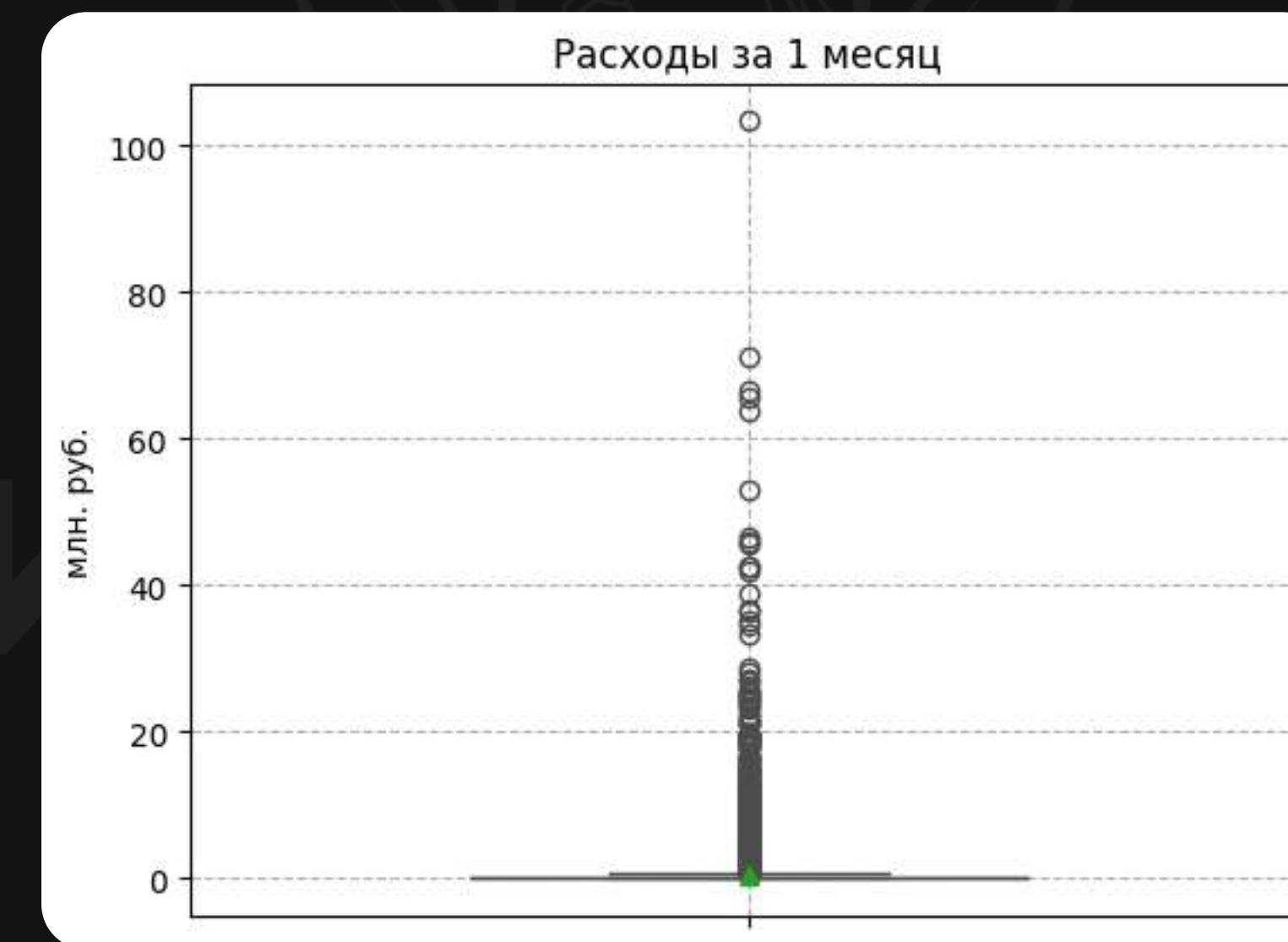
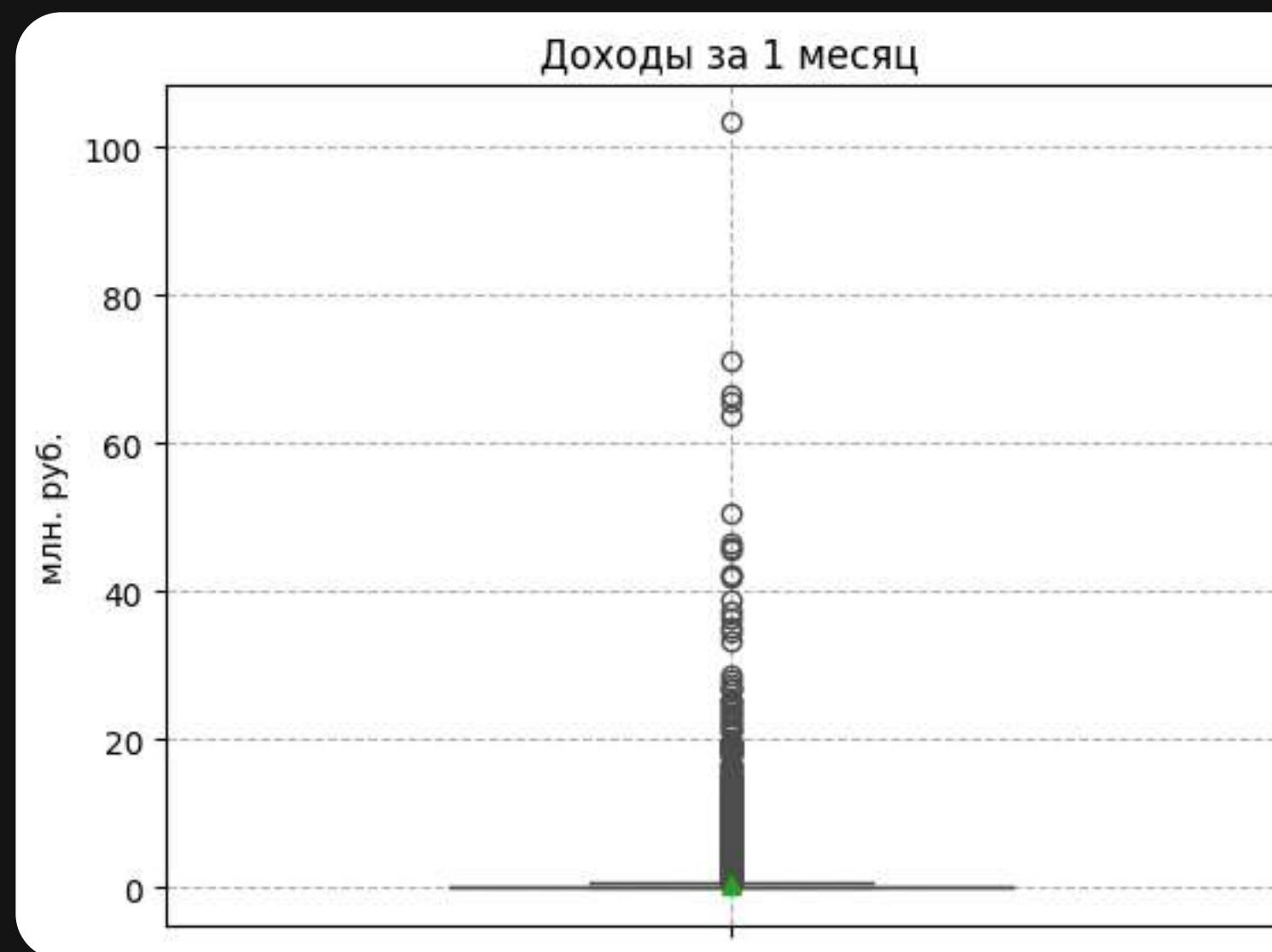
Корреляций в данных
довольно мало.
Поэтому попробуем
обработать данные.

Корреляционная матрица



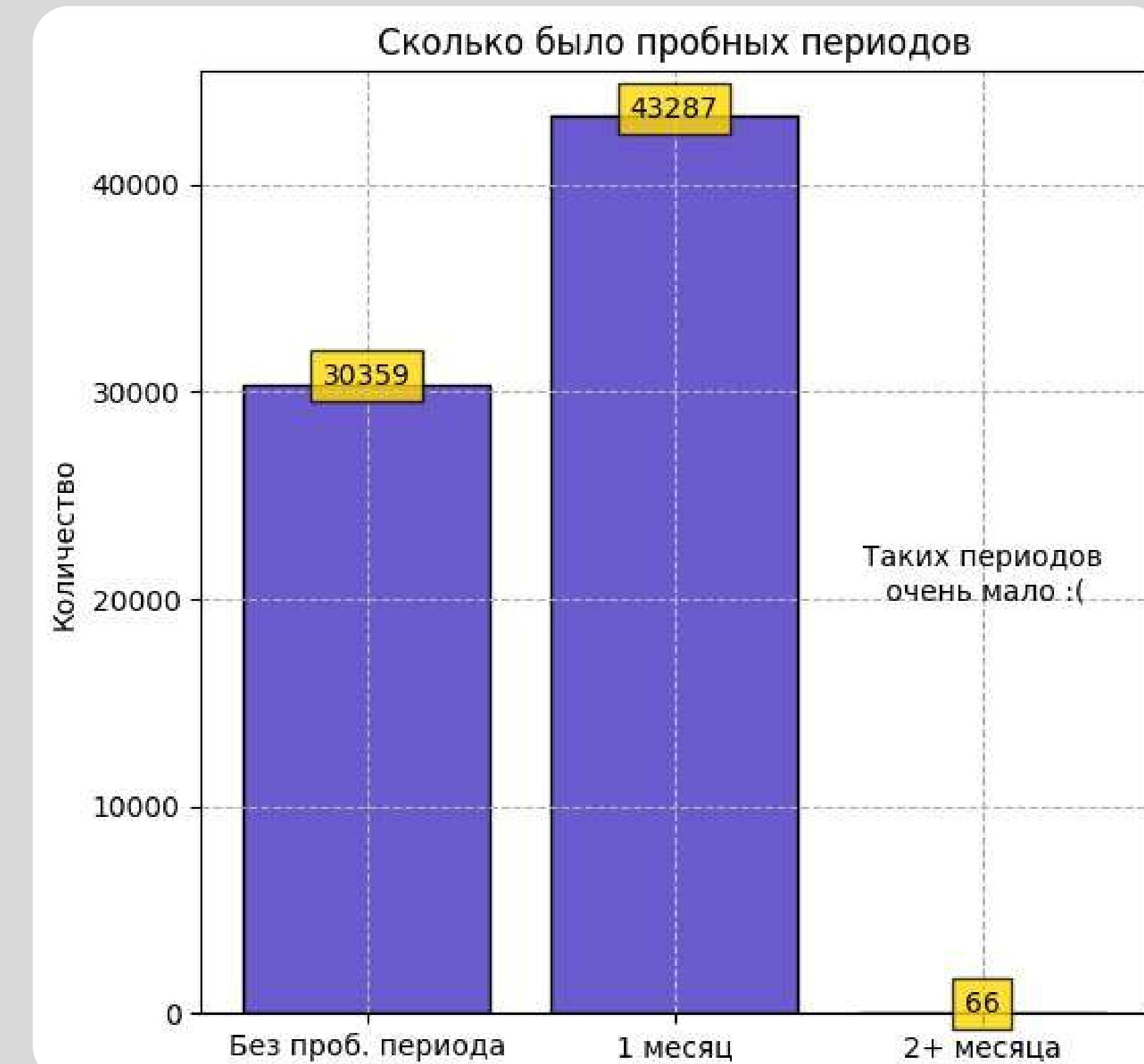
Разведанализ

Очень много выбросов, причем во всех параметрах, связанных с транзакциями.
Причем выбросы значительные (присутствует доход больше 100 млн. руб за 1 месяц!)



Разведанализ

Пробные
периоды
больше, чем на
два месяца
давали только в
исключительных
случаях. Поэтому
их крайне мало.



Что влияет на длительность подписки Тинькофф Премиум?

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

Гипотеза 1

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

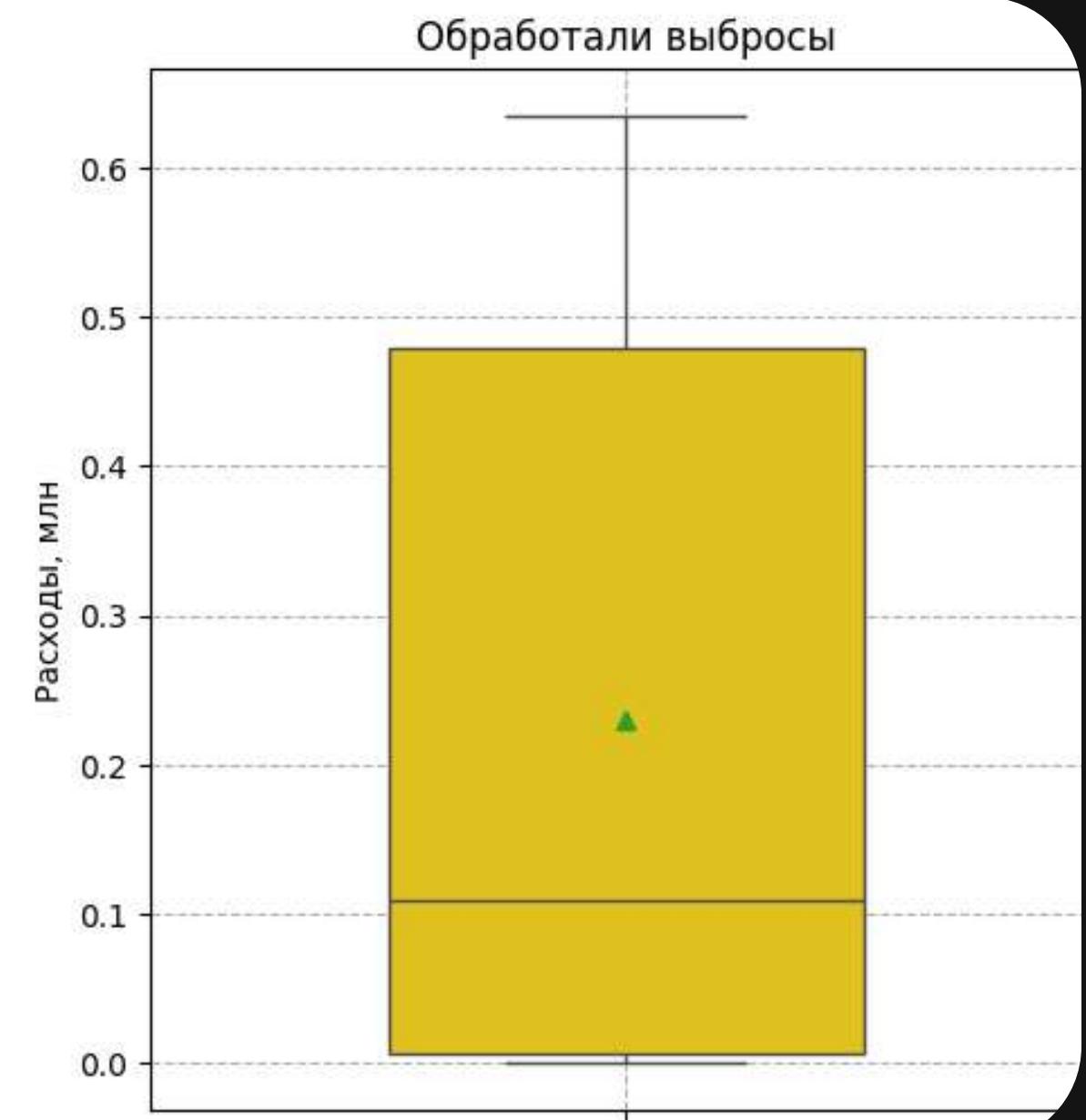
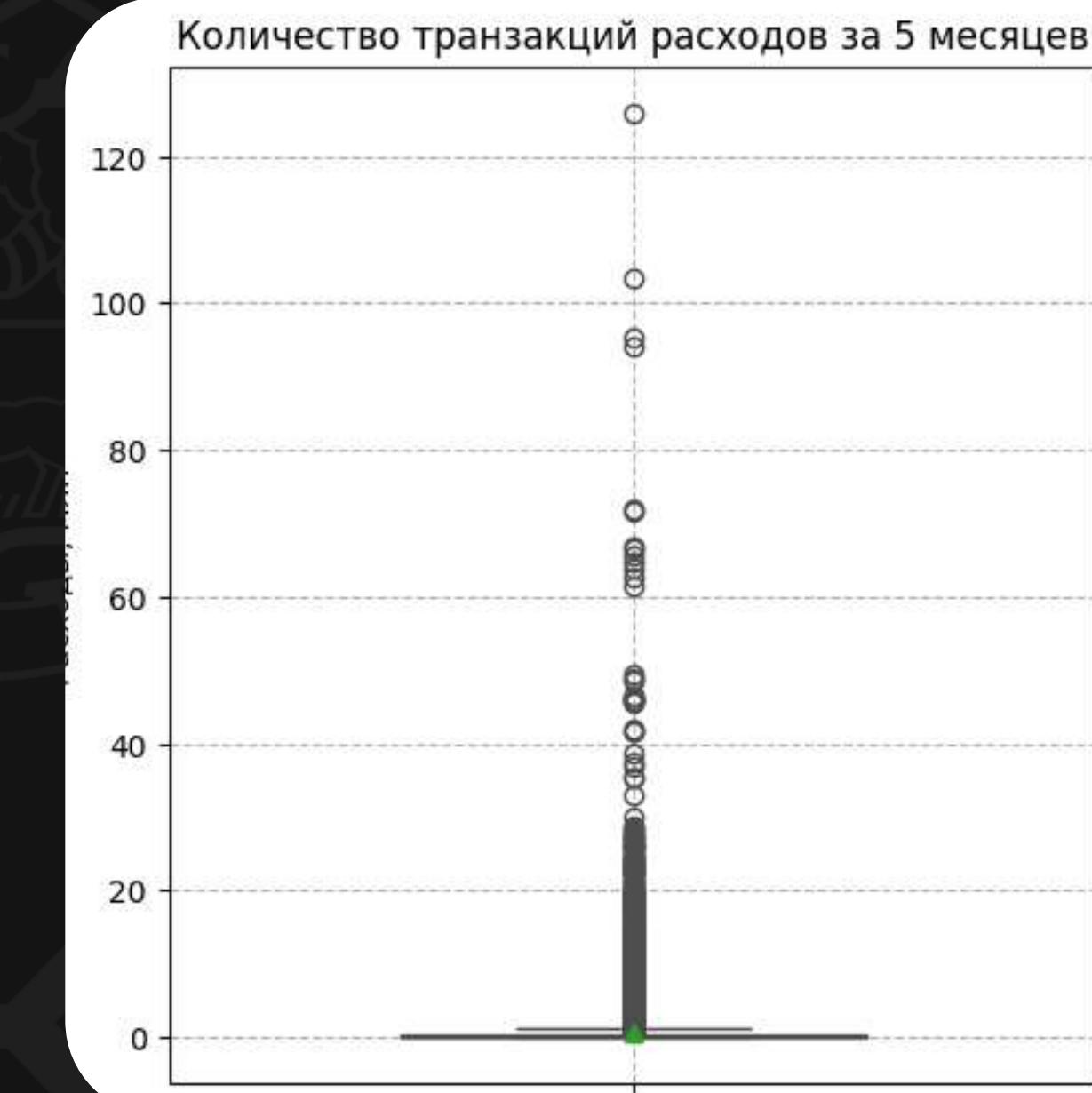
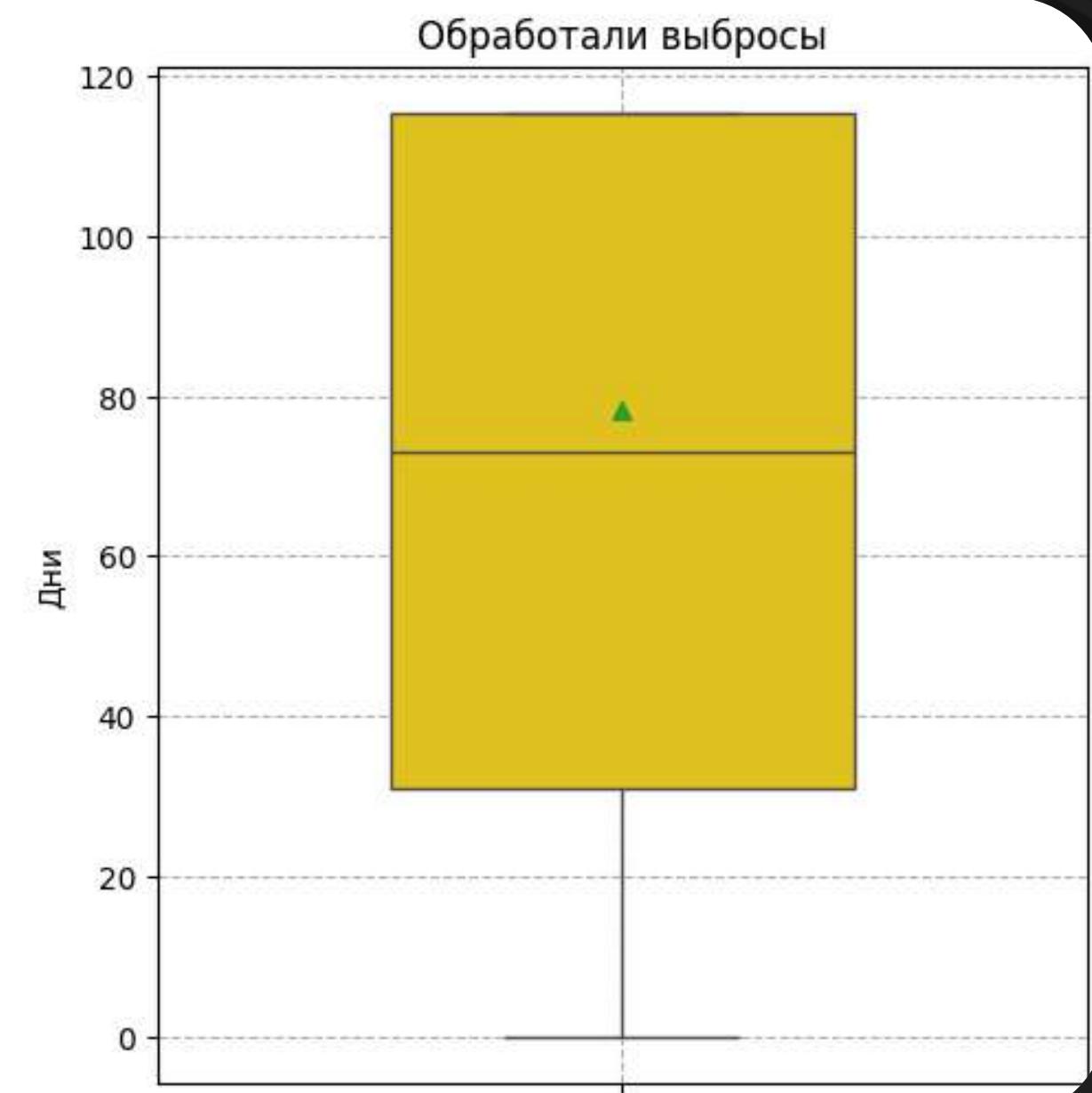
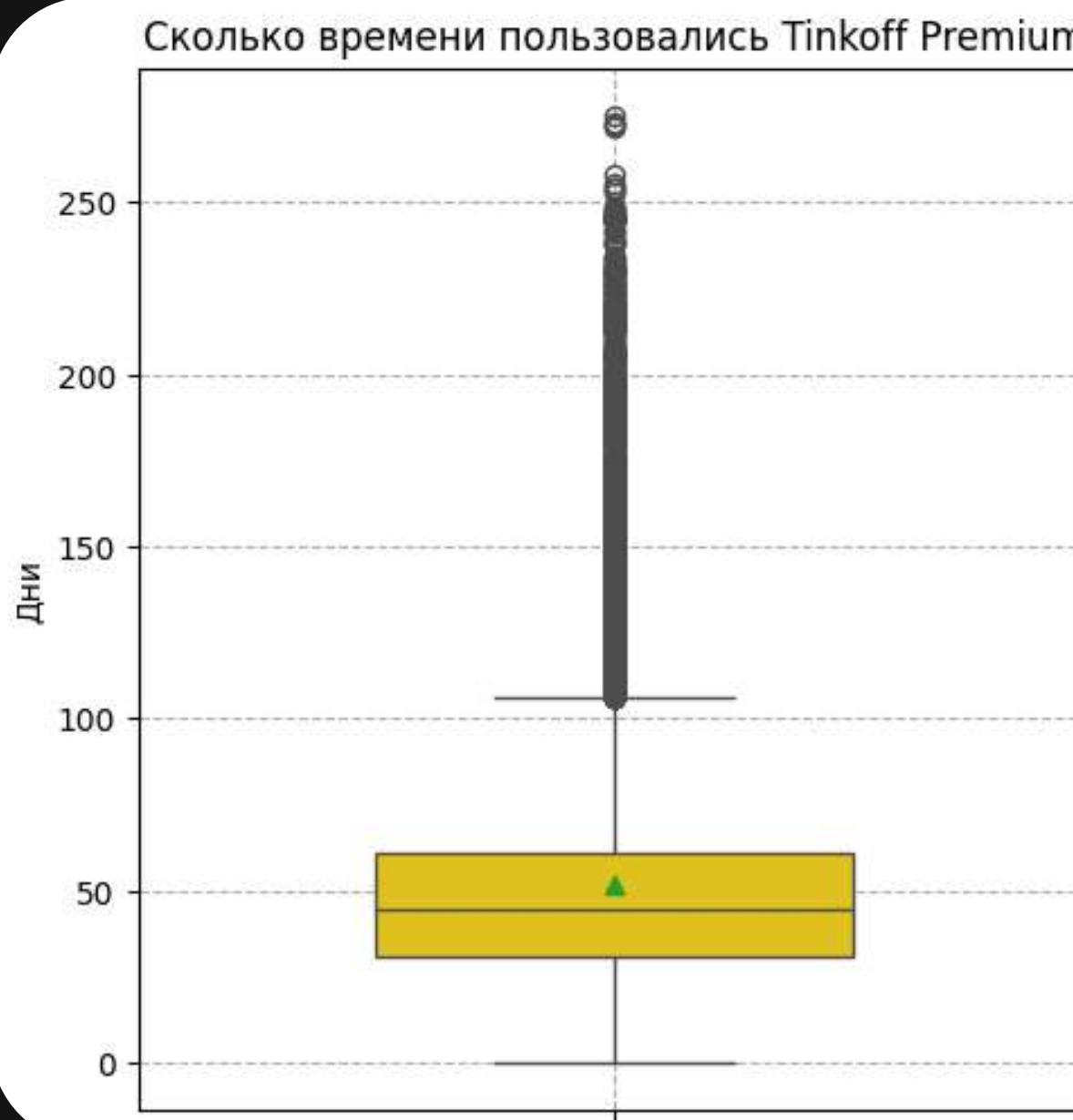
Люди, которые тратят много денег и много получают, скорее всего захотят получить какую-то выгоду со своих расходов в виде кэшбека. Более того, людям с огромным доходом с большей вероятностью позволят себе премиум.

Гипотеза 2

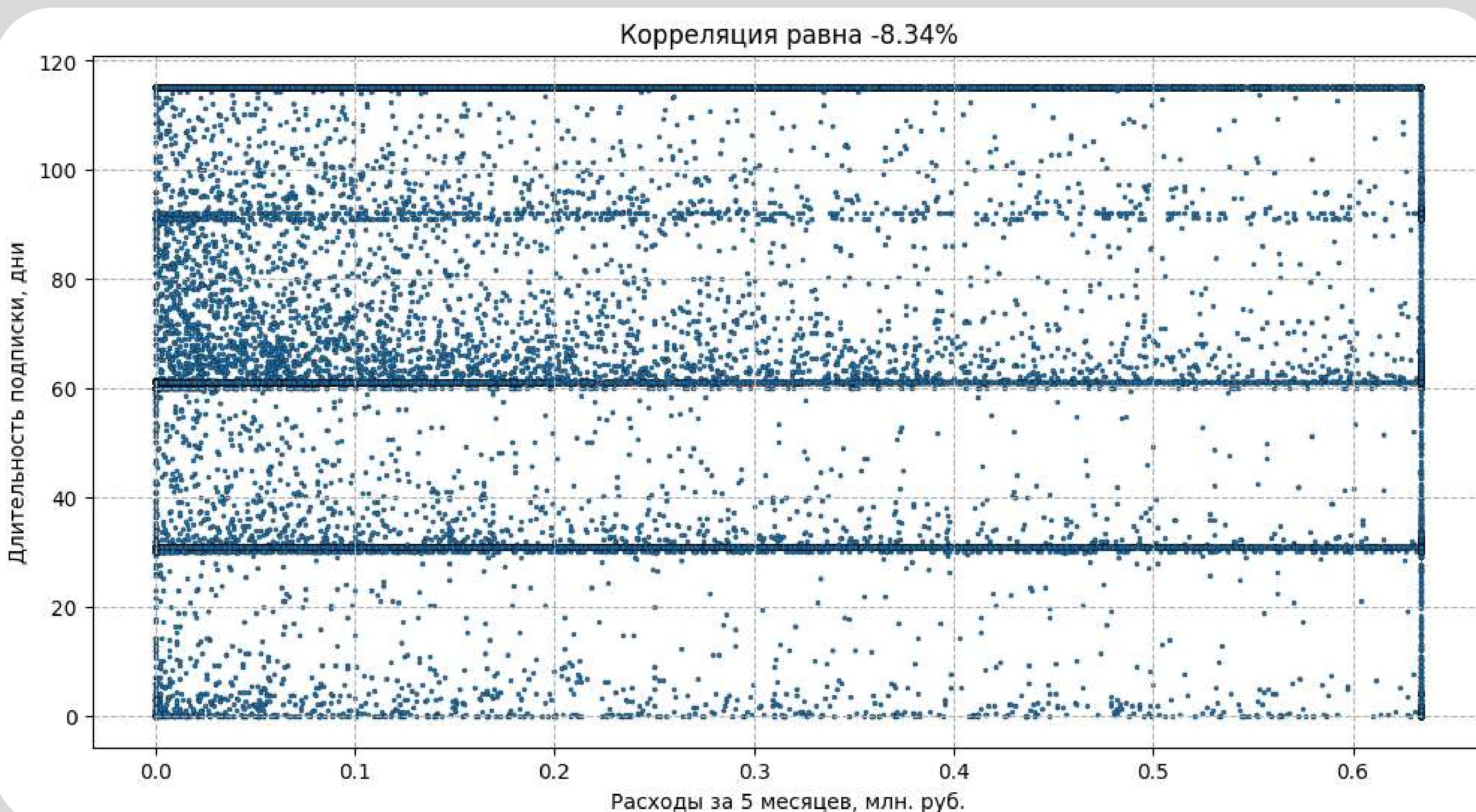
Длительность подписки зависит от того, насколько активно клиент пользуется её преимуществами.

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

Обрабатаем выбросы



Проверка гипотезы 1



Линейная
зависимость
отрицательна и
крайне
незначительна. Стоит
обратить внимание и
на “полосы” на
определенных днях, к
ним ещё вернёмся...

Мат. модель

Разобьём датасет на группы [0; 100 тыс. руб), [100 тыс. руб; 200 тыс. руб.), ...
[400 тыс. руб.; 500 тыс. руб)

H_null

Различий в группах нет

H_alt

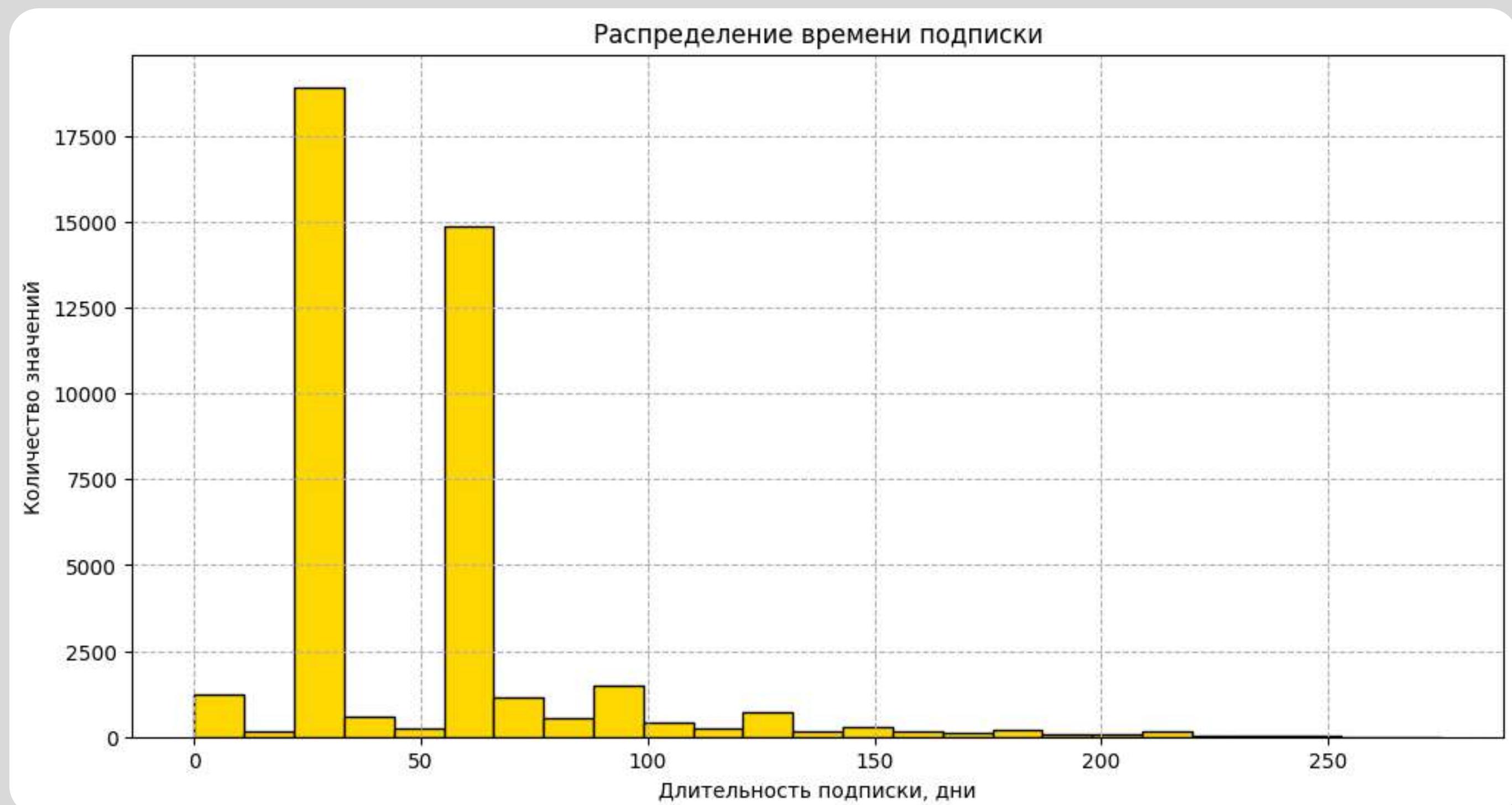
Различия в группах есть

```
F_onewayResult(statistic=68.523165, pvalue=1.39e-57)
```

Линейная независимость отрицательна и незначительна. При этом нельзя сказать, что разницы между группами нет, ведь pvalue < 0.05.

Остается только сделать вывод, что количество расходов мало чем влияет на время подписки, и это не то, на что стоит обращать много внимания.

А на что стоит обратить внимание?



Двумя слайдами ранее я рассказывал о неких полосах на определенных днях. Стоит обратить внимание, что самые толстые линии приходятся на 30й и 60й день - 1й и 2й месяца. Именно столько пробных периодов раздавали больше всего.

Вывод: много клиентов по окончании пробного периода уходят. Скорее всего потому, что не готовы платить.

Не забываем про вторую гипотезу...

Вторая гипотеза заключается в том, что частота использования привилегий может показать, насколько долго клиент будет подписчиком премиума.

Однако, у нас только бинарные значения (да/нет). Тут на помощь приходит feature engineering.

Индекс использования привилегий

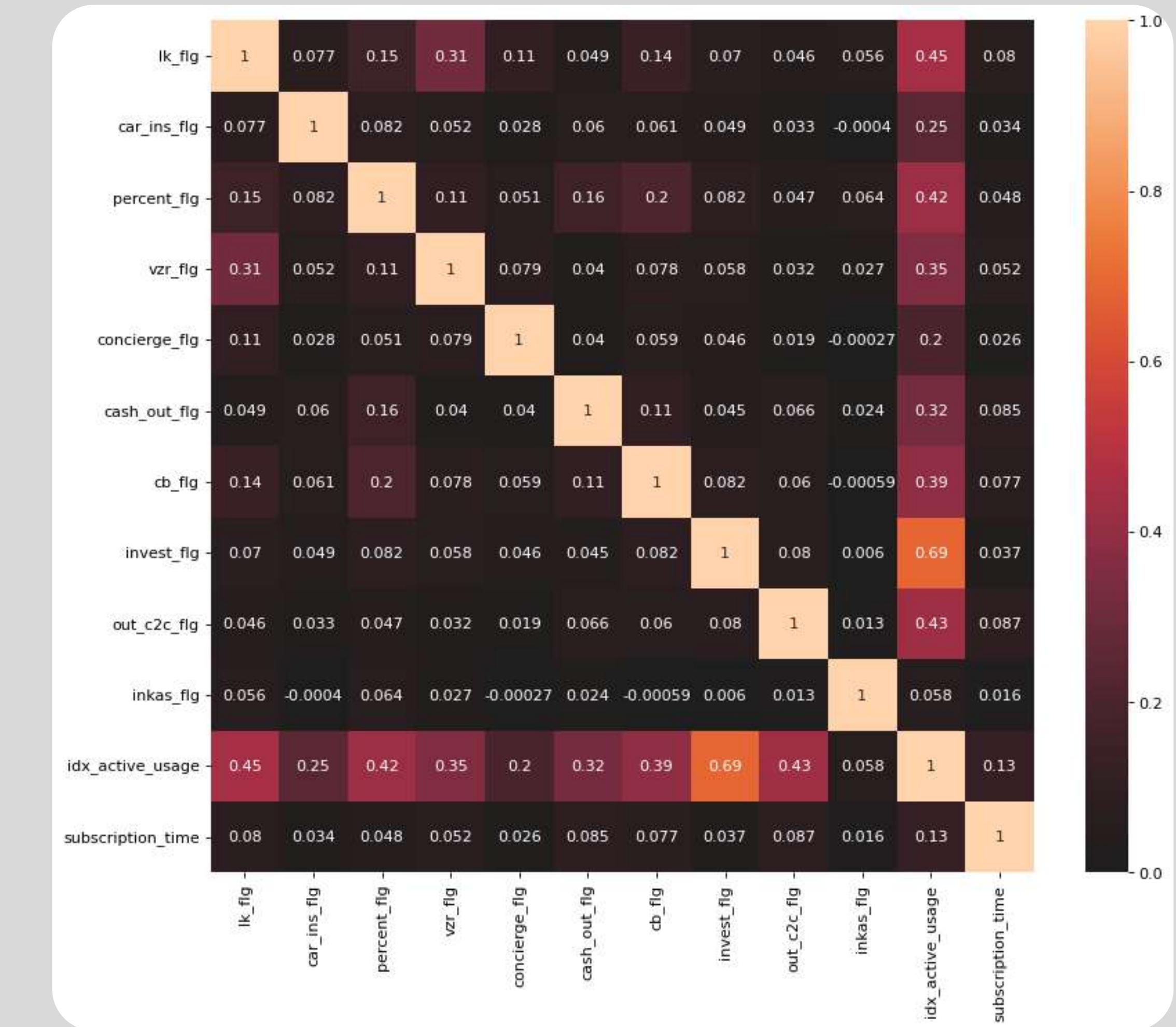
$$\text{Idx} = \text{lk_flg} + \text{inkas_flg} + \text{cashout_flg} + \dots$$

Суть заключается в том, чтобы просуммировать все колонки привилегий.

ТИНЬКОФФ

Проверка гипотезы 2

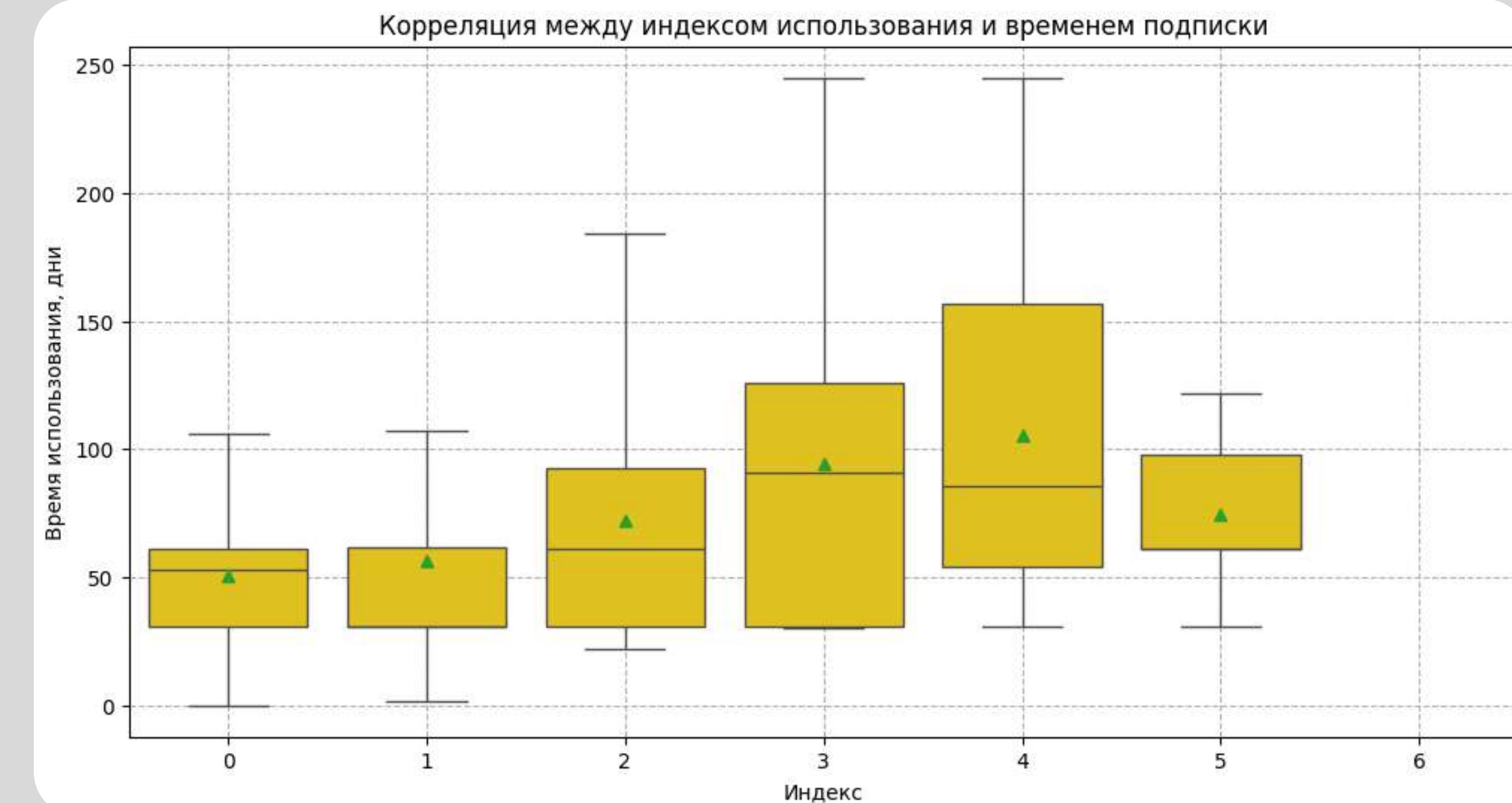
У отдельных значений корреляция с целевой метрикой была не больше 9%. Но у нового значения корреляция с `subscription_time` целых 13% в датасете с выбросами - незначительная, но это уже результат.



Проверка гипотезы 2

Посмотрим на другой график. Видно, что нижняя планка повышается с повышением индекса. Также почти везде повышается медиана. Верхняя планка тоже повышается. Но вот начиная с $idx = 5$ показатели падают. Но это можно объясняется малым количеством записей.

Результаты стат.теста не противоречат: pvalue < 0.05, а значит разница между группами действительно есть.



F_onewayResult(statistic=402.8, pvalue=1.403e-87)

Выводы

- Гипотеза №1 ошибочна: Клиенты с большими расходами не пользуются подпиской больше других (а возможно даже меньше)
- Гипотеза №2 верна: активные клиенты действительно в среднем сидят дольше других, но таких меньшинство.
- Множество клиентов отписываются под конец пробного периода.

Policy implication

- Стоит разными способами привлекать внимание клиентов к возможностям подписки.
- Можно разбить пробную подписку на две фазы: первая фаза открывает доступ только к части всех возможностей и длится примерно 25 дней. Но вторая фаза длится последние 5 дней и предоставляет все возможности подписки Тинькофф Премиум. Вторую фазу можно начать с активной рекламы всех привилегий подписки, что также решает проблему из первого пункта.

Таким образом мы мотивируем клиентов оставаться на больший срок и чаще пользоваться услугами Тинькофф Премиум.

Спасибо за внимание

**Михаил
Боков**

Математик-статист

**Павел
Маляренко**

Математик-статист

**Роман
Локманов**

Программист

**Дмитрий
Бусыгин**

Программист

**Александр
Агафонов**

Ментор

**Мария
Франк**

Ментор



Ссылка на Github