ПобеDano

НАКРУТКА ОТЗЫВОВ

Тинькофф.
Юридические лица

Ключ к успеху – это создание у клиентов реалистичных ожиданий

Выполнили:
Архипов Владимир
Филинов Дмитрий
Утенков Дмитрий
Стрекалова Анастасия
Садчикова Арина

Оглавление

База данных

Записи в телефонных книгах клиентов Tinkoff

- Общее количество записей номеров
- Количество матерных/ фродовых слов в записях компании/руководителей компании

Взаимодействия с банками и судами

- количество заявок на кредит
- количество жалоб на компанию
- количество и процент дел в суде в качестве ответчика
- сумма исков в качестве ответчика

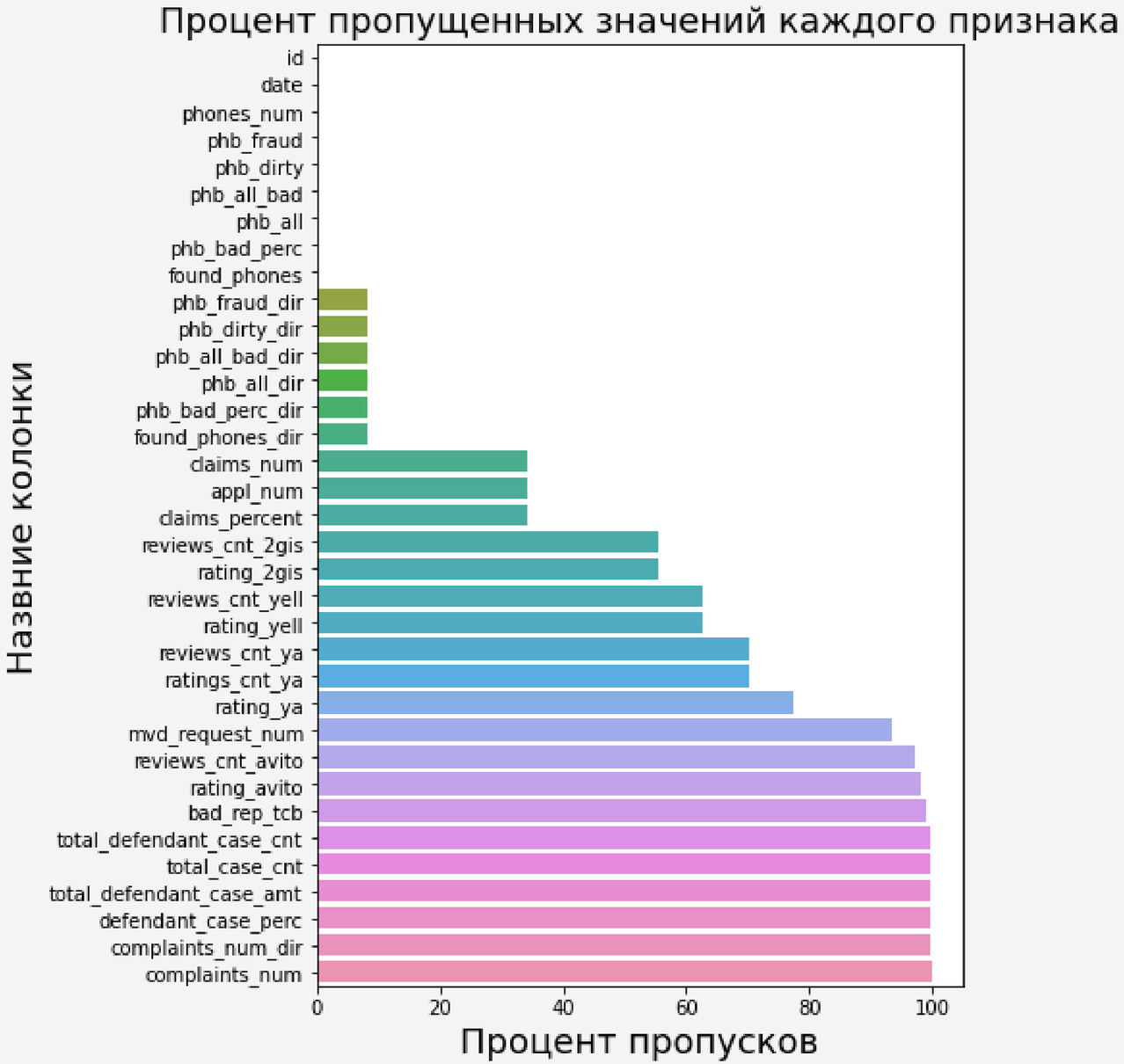
Количество отзывов на платформах

- Яндекс карты
- 2gis
- Avito
- Yell

Рейтинг на платформах

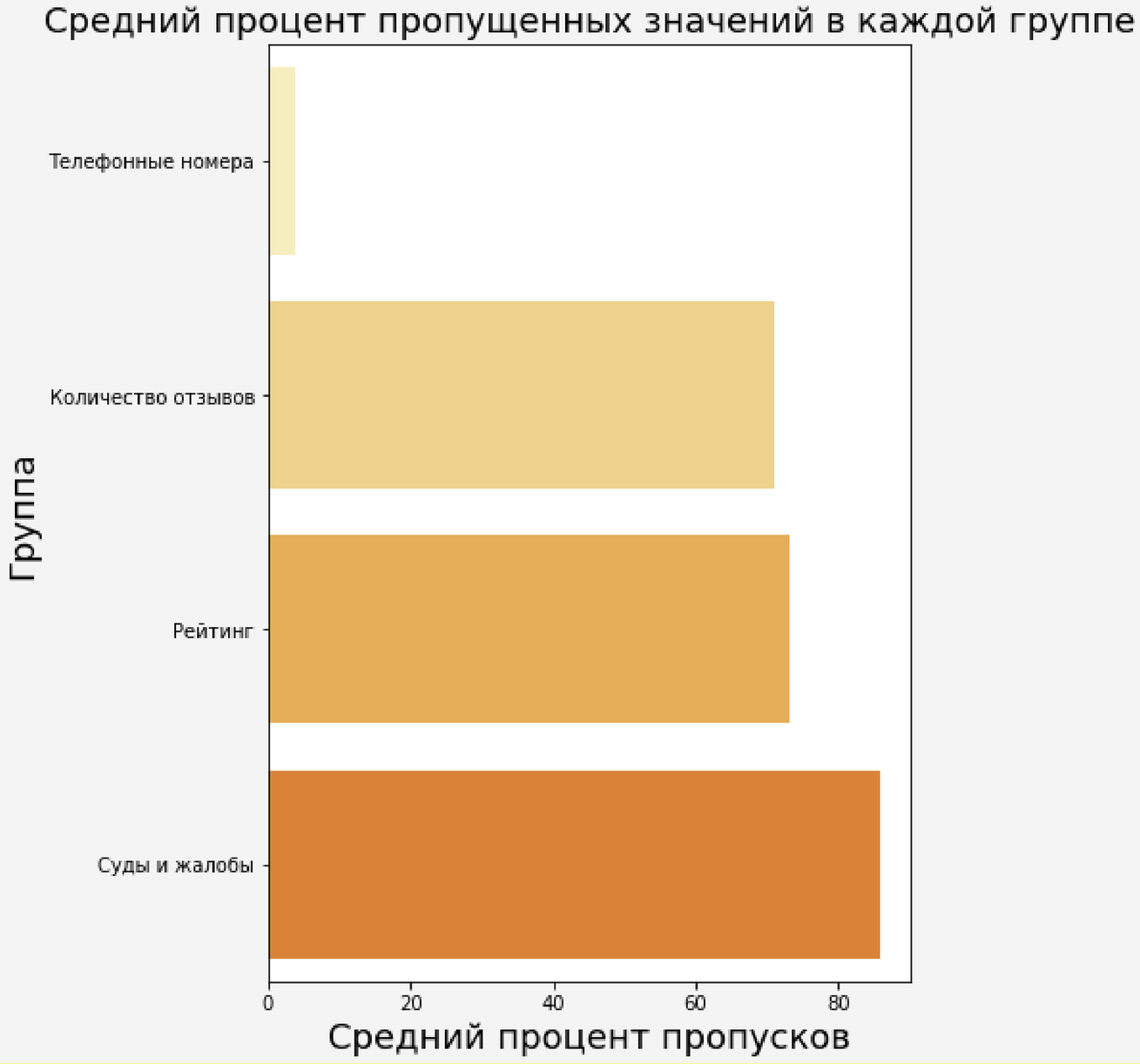
- Яндекс карты
- 2gis
- Avito
- Yell

Пропущенные значения в базе данных



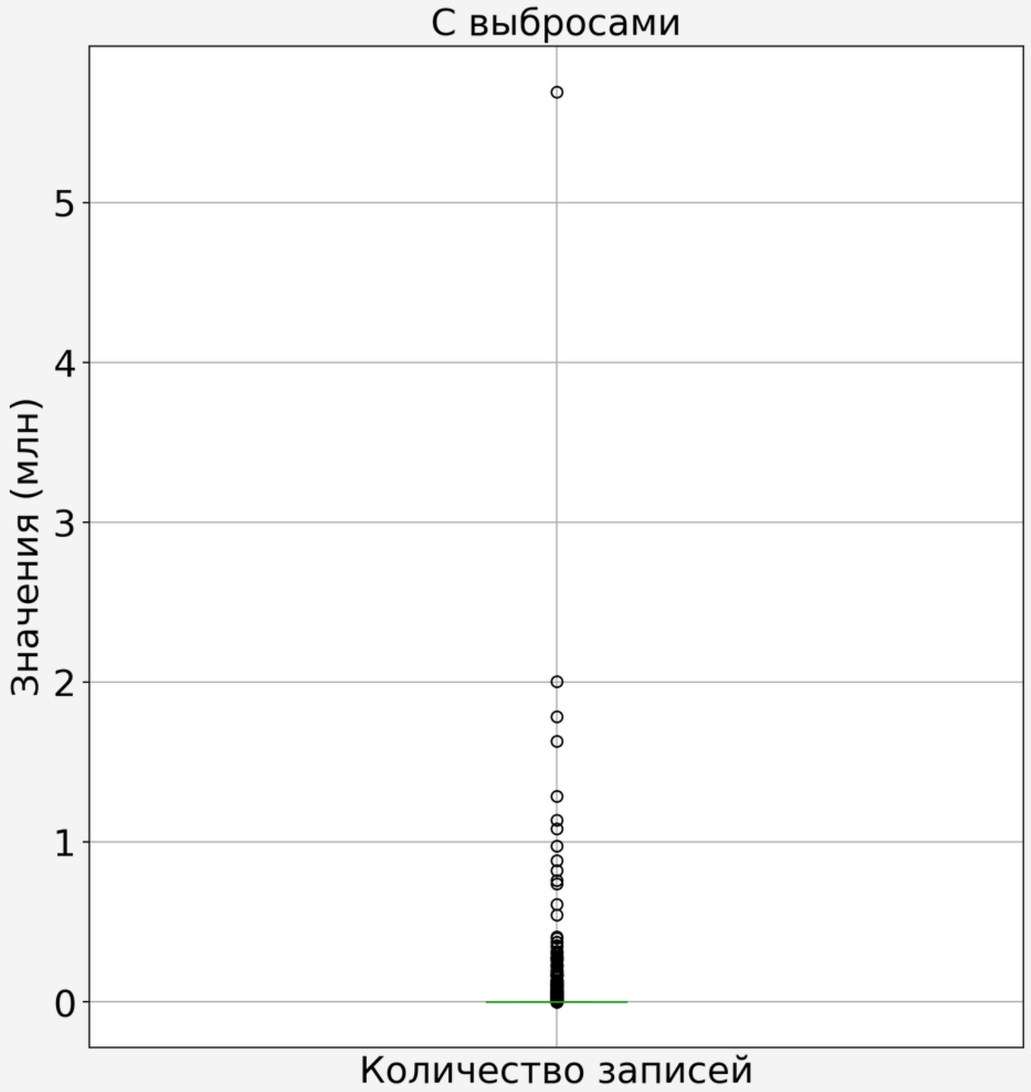
Данные из группы взаимодействий с банками и судами - около 97 % пропусков

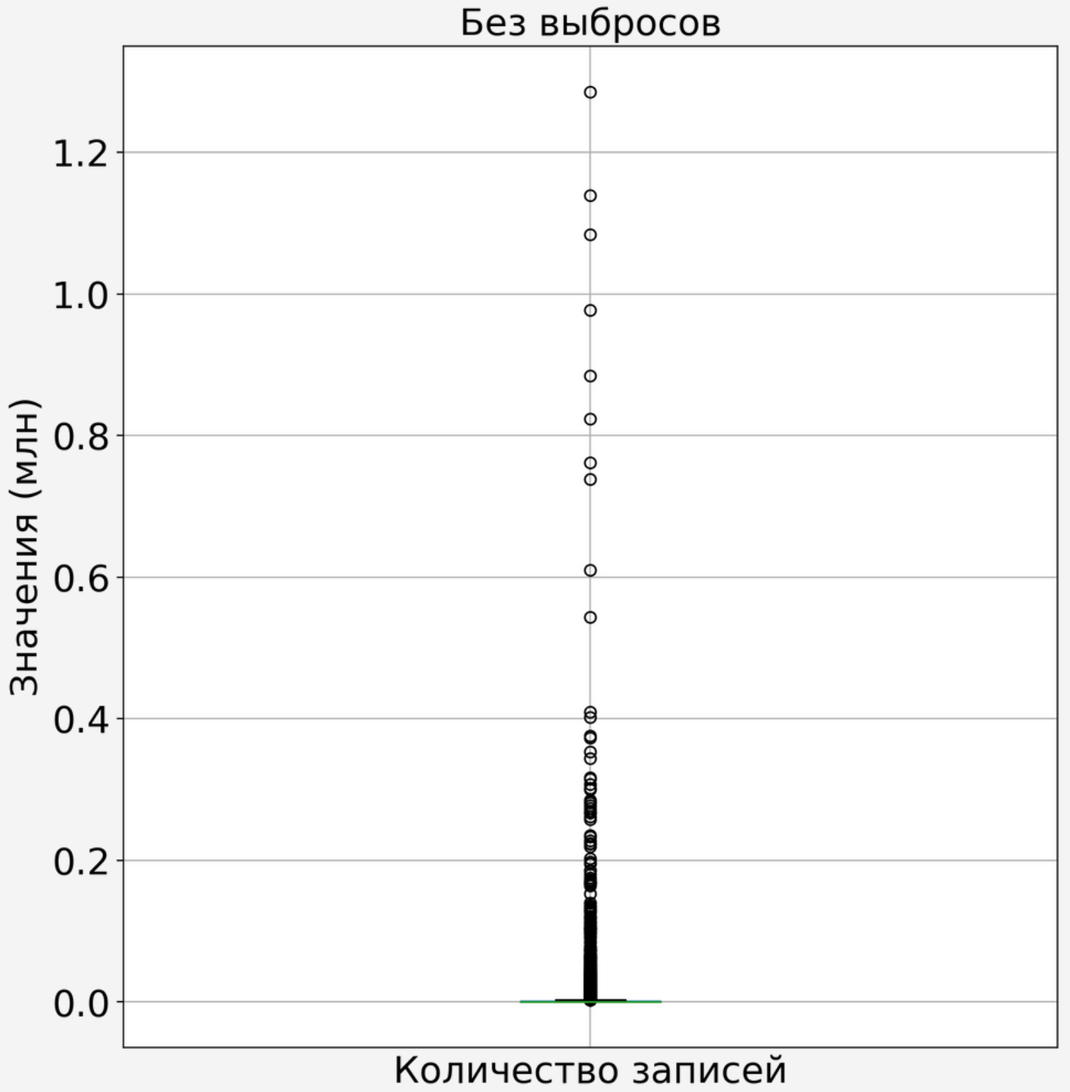
Пропущенные значения в базе данных



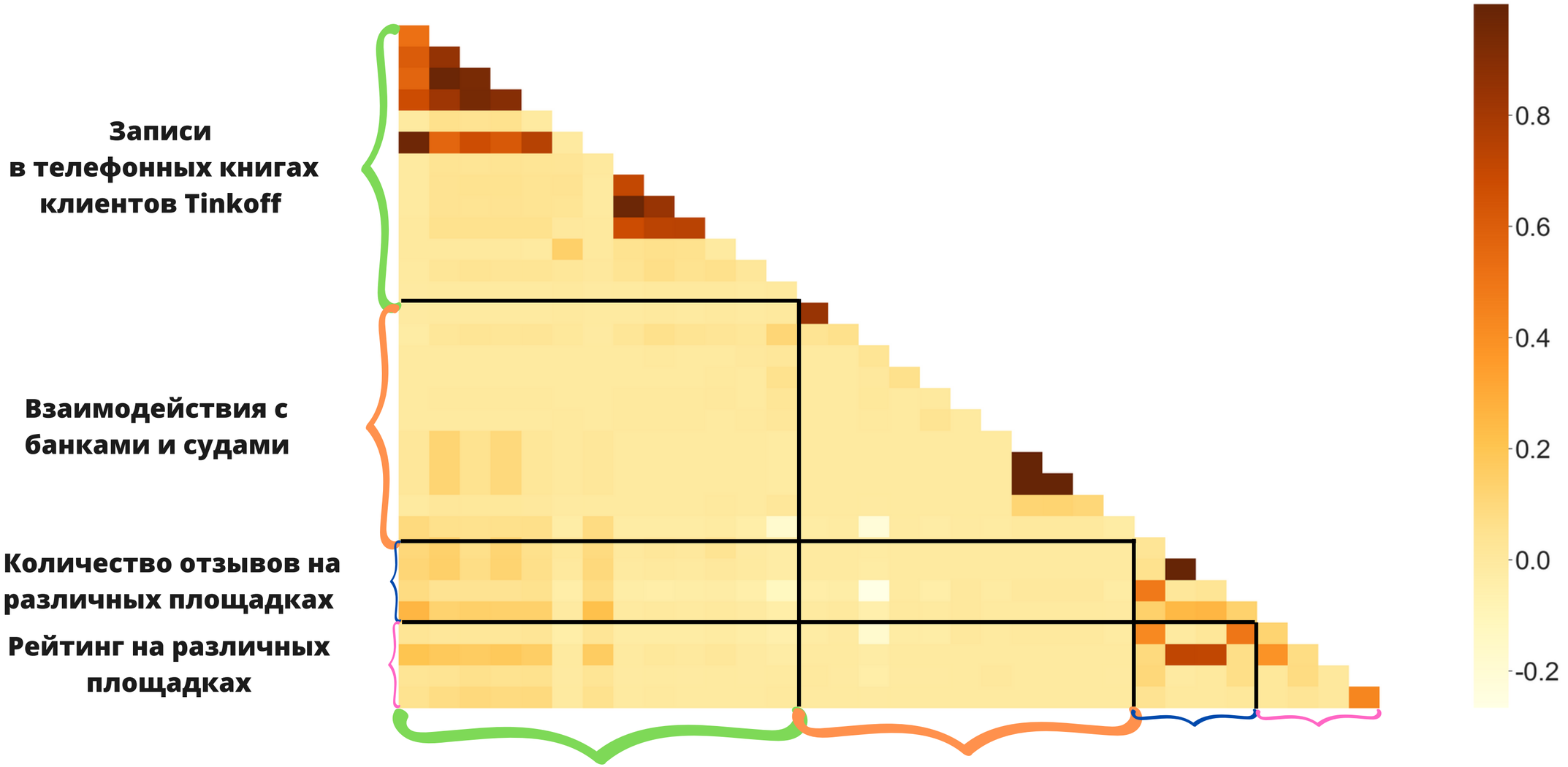
Обработка выбросов

Количество записей номеров компаний в ТК клиентов





Матрица корреляции



Исследовательский вопрос

Какие компании больше беспокоятся за свою репутацию?

***Репутация** - это **совокупность факторов**, таких как **отзывы и рейтинг** на платформах, количество **жалоб** на компании, **исков в суд**, **процент фродовых слов** из ТК клиентов*

Гипотеза

Непопулярные компании чаще пользуются инструментами накрутки отзывов

***Непопулярные** - это компании, чьи номера мало записывали в ТК клиентов*

Новые колонки

Количество всех отзывов

– сумма всех отзывов на разных площадках

Средний рейтинг

$$\frac{\sum \text{рейтинг} * \text{кол-во отзывов на платформе}}{\sum \text{кол-во отзывов на платформе}}$$

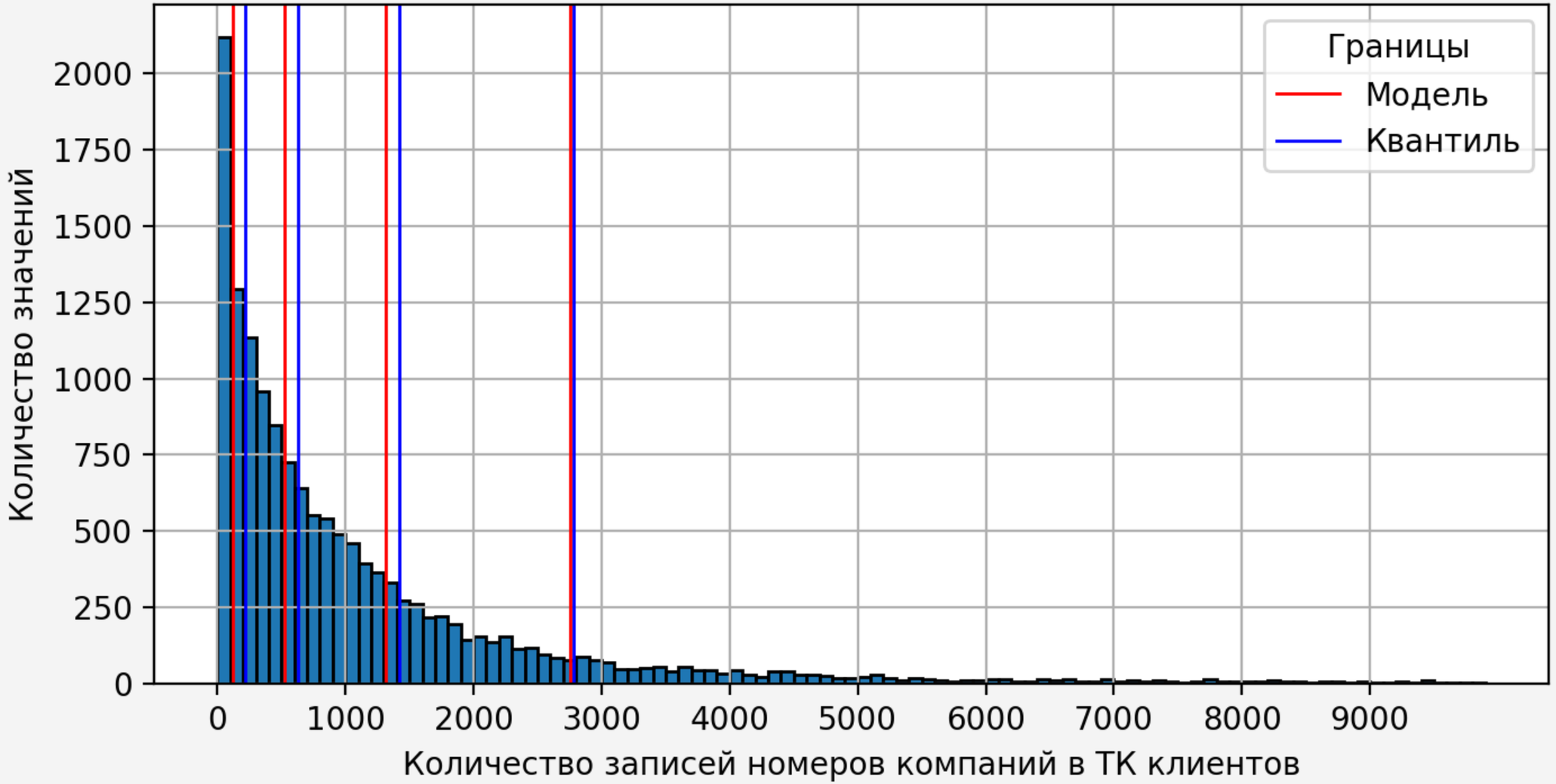
Популярность

1 – (< 122 записей в ТК клиентов)
2 – (123-520 записей в ТК клиентов)
3 – (521-1318 записей в ТК клиентов)
4 – (1319-2757 записей в ТК клиентов)
5 – (> 2757 записей в ТК клиентов)

cheating

0 - Нет подозрений в накрутке
1 - Компания подозревается в накрутке

Гистограмма распределения кол-ва записей в ТК клиентов



Механизм


Непопулярная компания получает негативный отзыв
Рейтинг компании сильно падает
Прибыль компании снижается
Компании приходится прибегать к накрутке отзывов

Критерии определения накрутки отзывов

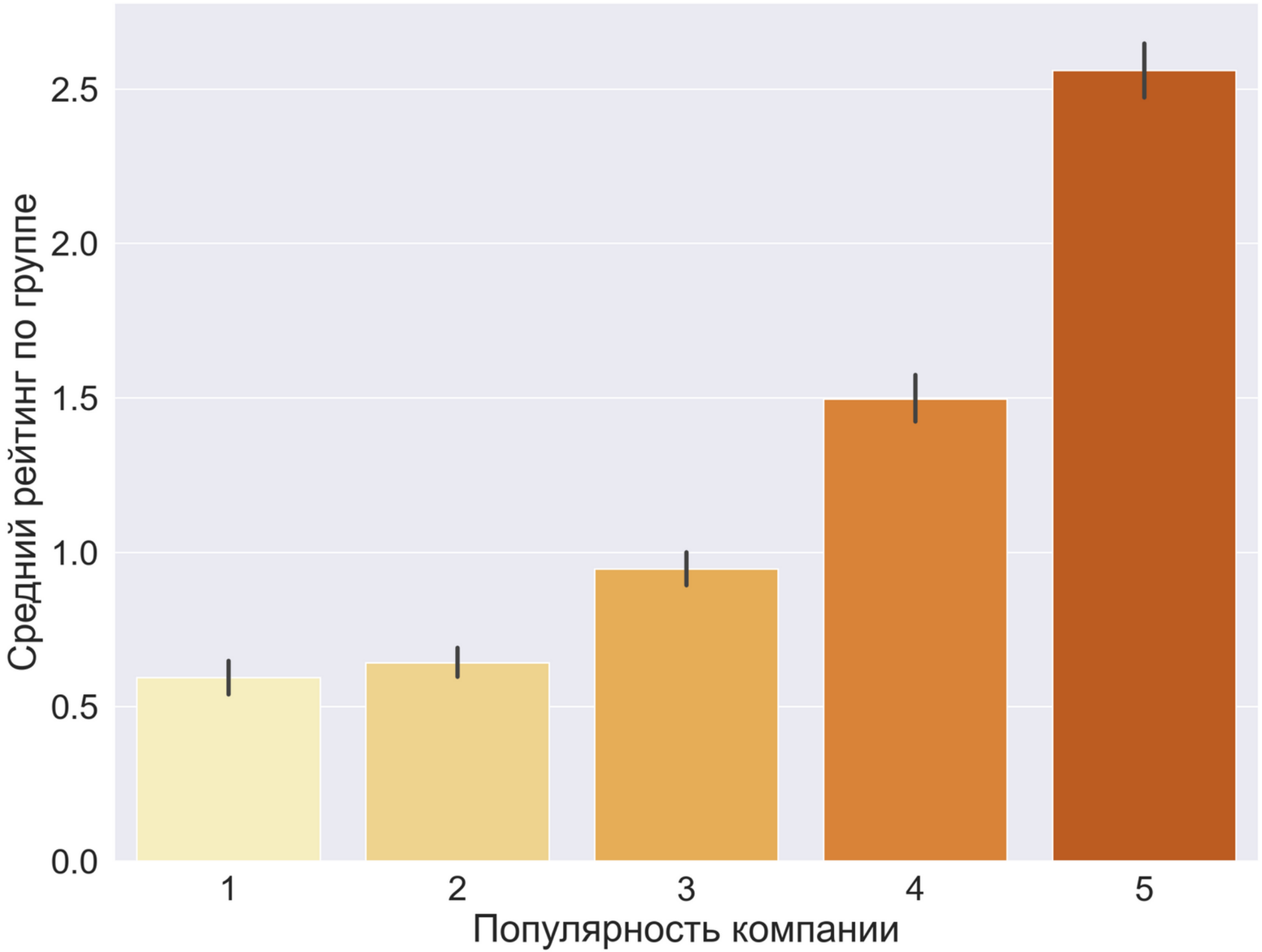
Компании у которых

Q(count_reviews) - Q(phb_all) > 0.5

Q - персентиль
count_reviews - количество отзывов
phb_all - количество записей
номеров компании в ТК клиентов

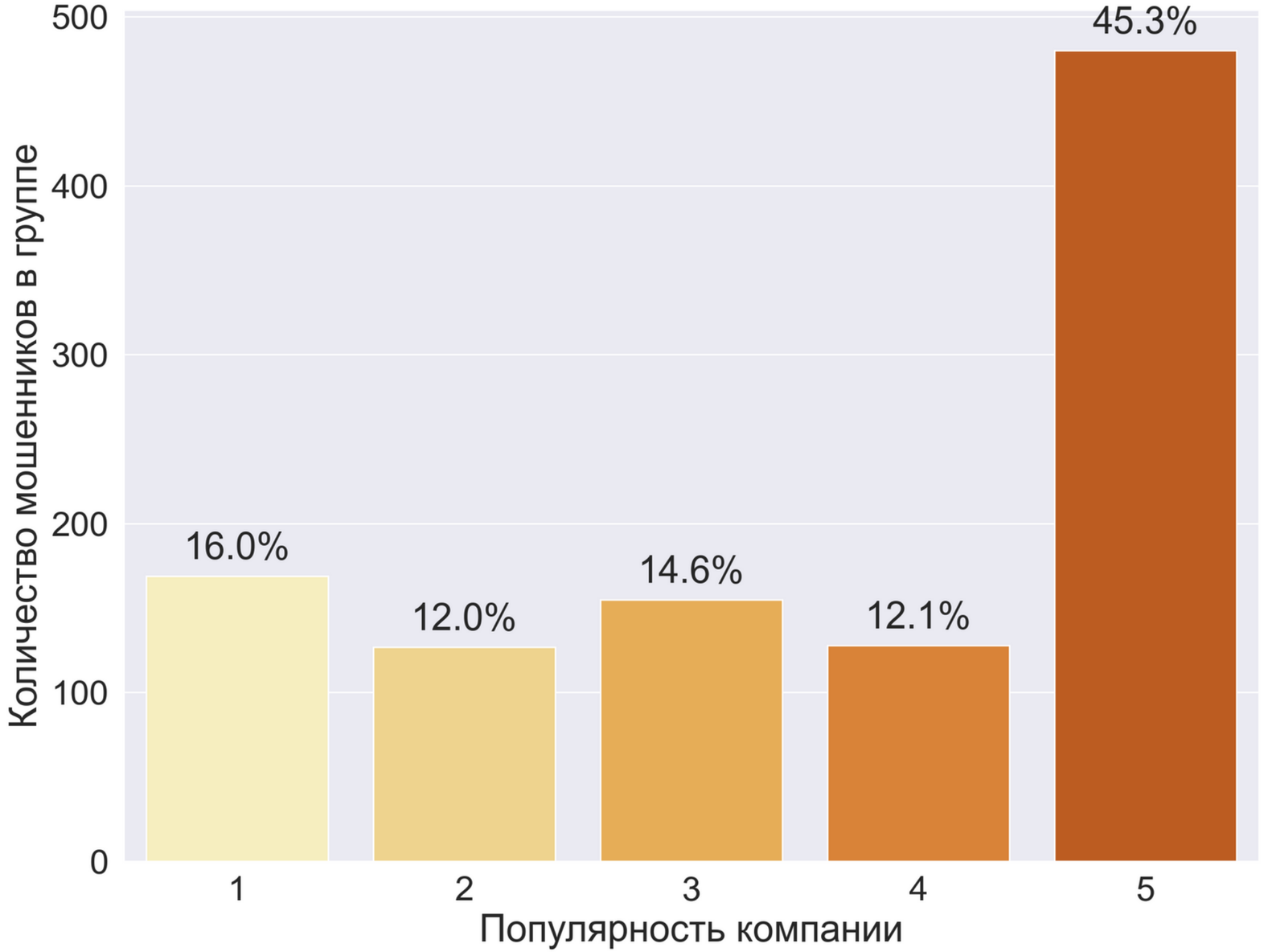
рейтинг компании на двух платформах отличается > 2.5

График зависимости среднего рейтинга от популярности



1 – практически неизвестная компания
5 - крайне популярная компания

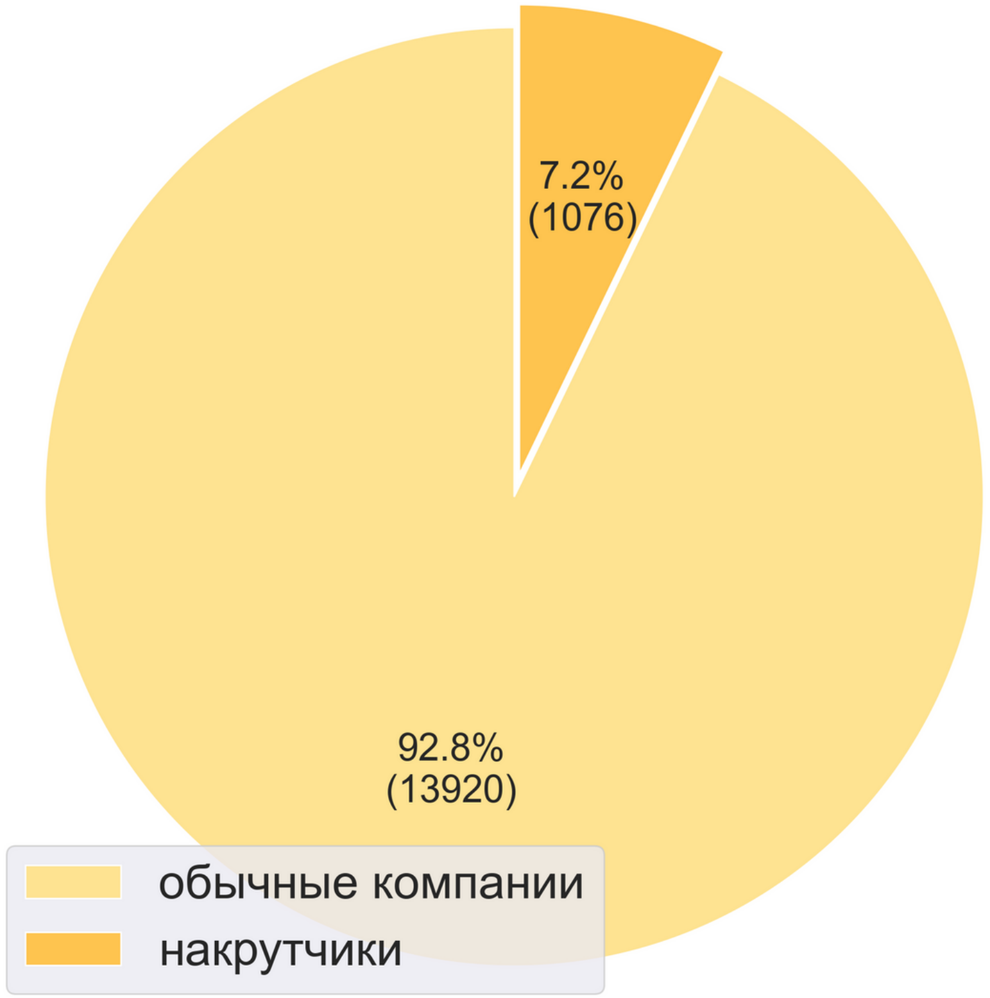
Количество мошенников по группам



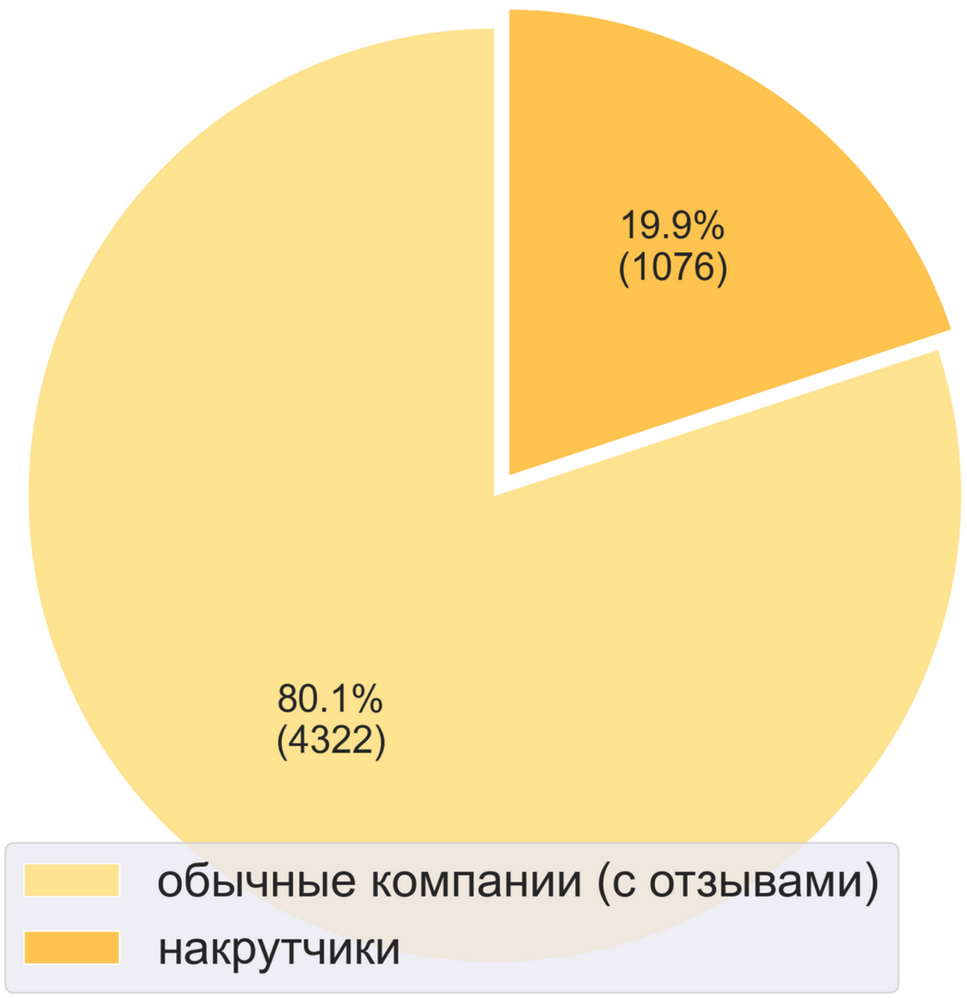
1 – практически неизвестная компания
5 - крайне популярная компания

Анализ результатов

С учетом компаний без рейтинга



Без учета компаний без рейтинга



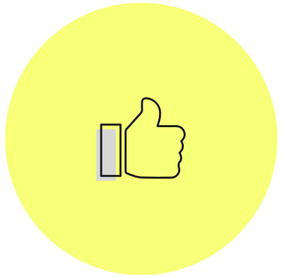
Выводы

Популярные компании чаще накручивают отзывы

Так как они:

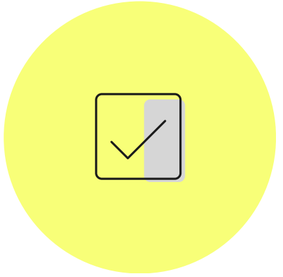
- заботятся о персональном имени
- имеют бóльшие бюджеты на маркетинг и пиар
- меньше подозреваются в накрутке отзывов

Развитие проекта



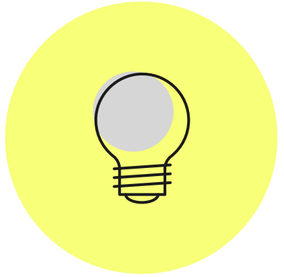
Улучшения

- более точная система определения накрутки



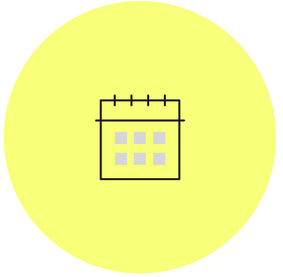
Необходимые данные

- расположение компаний
- mcc коды.



Чем будет полезно

- первичные отбор юридических лиц
- повышение доверия к платформам с отзывами

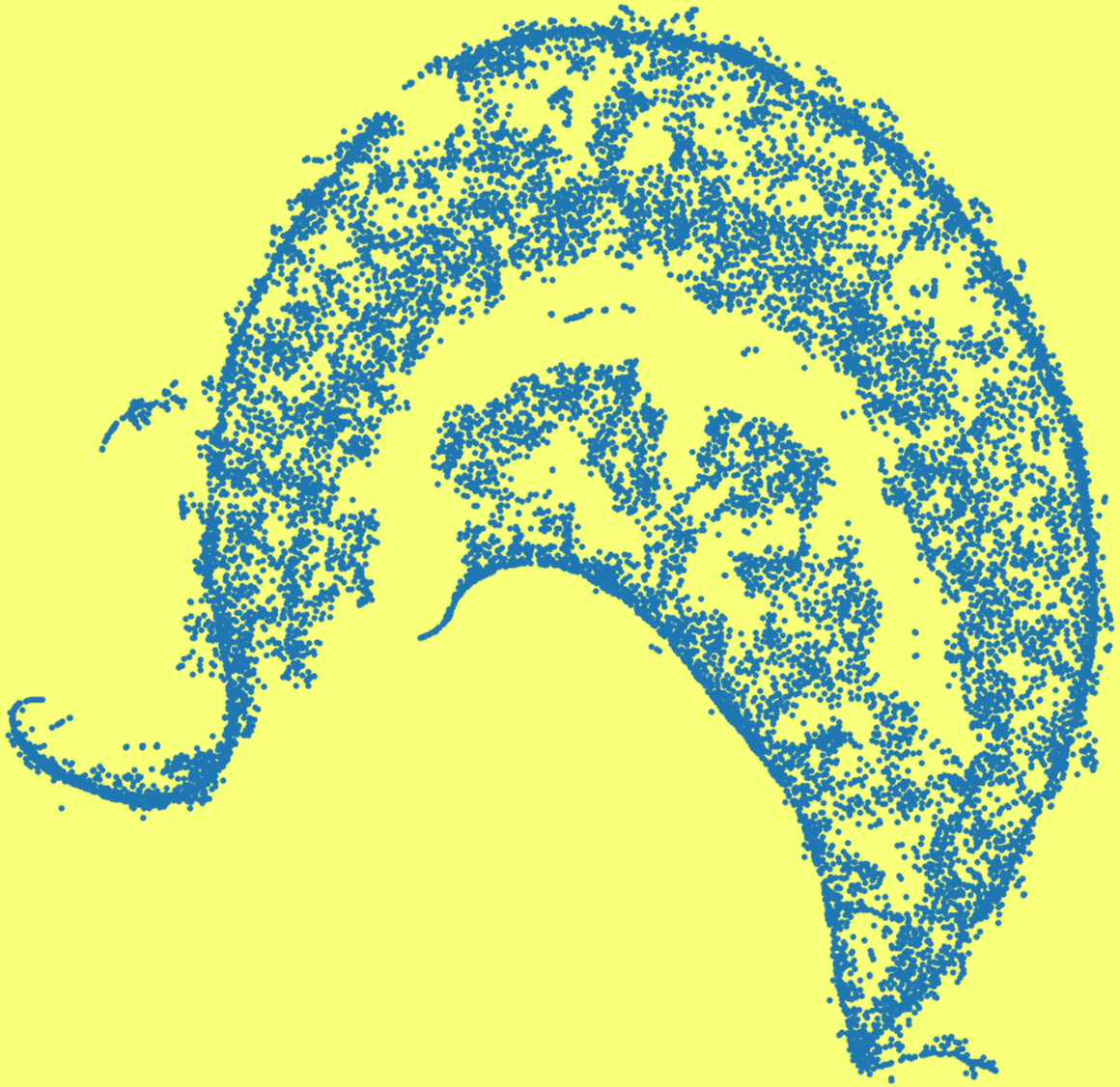


Кому будет полезно

- Компании Тинькофф
- площадкам, где располагаются отзывы

Реализация модели

Т-распределенное стохастическое встраивание соседей (TSNE)
+
Кластеризация методом К-ближайших соседей (KMeans)



Наша команда

Анастасия Стрекалова

Аналитик

Дмитрий Филинов

Программист

Владимир Архипов

Тимлид

Дмитрий Утенков

Программист

Арина Садчикова

Дизайнер

GitHub