



Исследование поведения пользователей
при активации повышенного кэшбэка

Кэшбэк – всегда плюс?

sqlite9

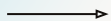
Дуванов Лука

Мелехов Даниил

Передёра Арина

Подъяпольский Ярослав

Седельников Данила



01

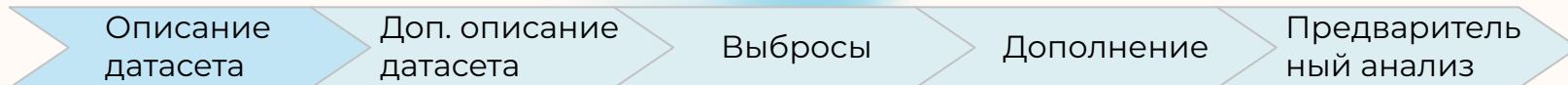
Обзор датасета

Повышенный кэшбэк

— датасет с категориями покупок пользователей, тратами и кэшбэком по ним

Используемые данные:

- *Ключ клиента*
- *Возраст клиента*
- *Оборот по категориям*
- *Активация категорий*
- *Полученный кэшбэк*



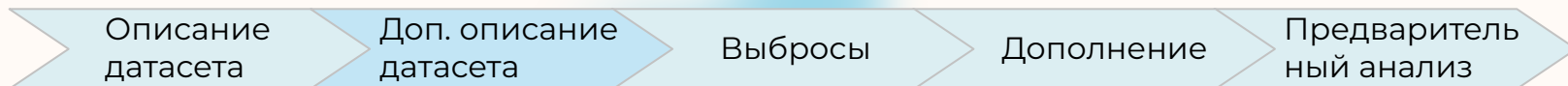
Повышенный кэшбэк

— датасет с категориями покупок пользователей, тратами и кэшбэком по ним

Категории кэшбэка:

аптеки	жд_билеты
рестораны	образование
одежда_и_обувь	дом_и_ремонт
автоуслуги	спорттовары
супермаркеты	животные
такси	цветы
красота	фастфуд
развлечения	каршеринг
	аренда_авто

- Оборот
- Активация категории
- Полученный кэшбэк



Обработка выбросов



Пользователи с полом None



Пользователи с кэшбэком больше оборота в любой категории

Пользователи с оборотом или кэшбэком в любой из категорий, у которых в отсортированном по убыванию столбце выполняется неравенство:

$$\frac{|a_{i+1} - M_a|}{2} < a_i - a_{i+1}$$

M_a – медиана столбца
 a_i – значение столбца

3942 (6,6%) выброса
найдено

Описание датасета

Доп. описание датасета

Выбросы

Дополнение

Предварительный анализ

Дополнение датасета

is_blowout

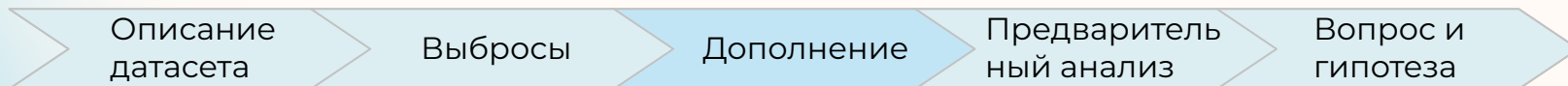
Добавлен столбец для определения выбросов

age_cat

Столбец для определения возрастной категории (в строчном виде, "14-24", "76+")

Замена None на 0

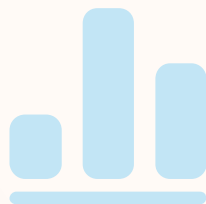
В категориях с оборотом/кэшбэком, для большей читаемости датасета



Предварительный анализ



Проведено сравнение значений до
и после очистки выбросов



Построены графики распределения
для оборота, графики распределения
пользователей по возрасту

Выбросы

Дополнение

Предварительный анализ

Вопрос и гипотеза

Механизм

Средний оборот по категориям

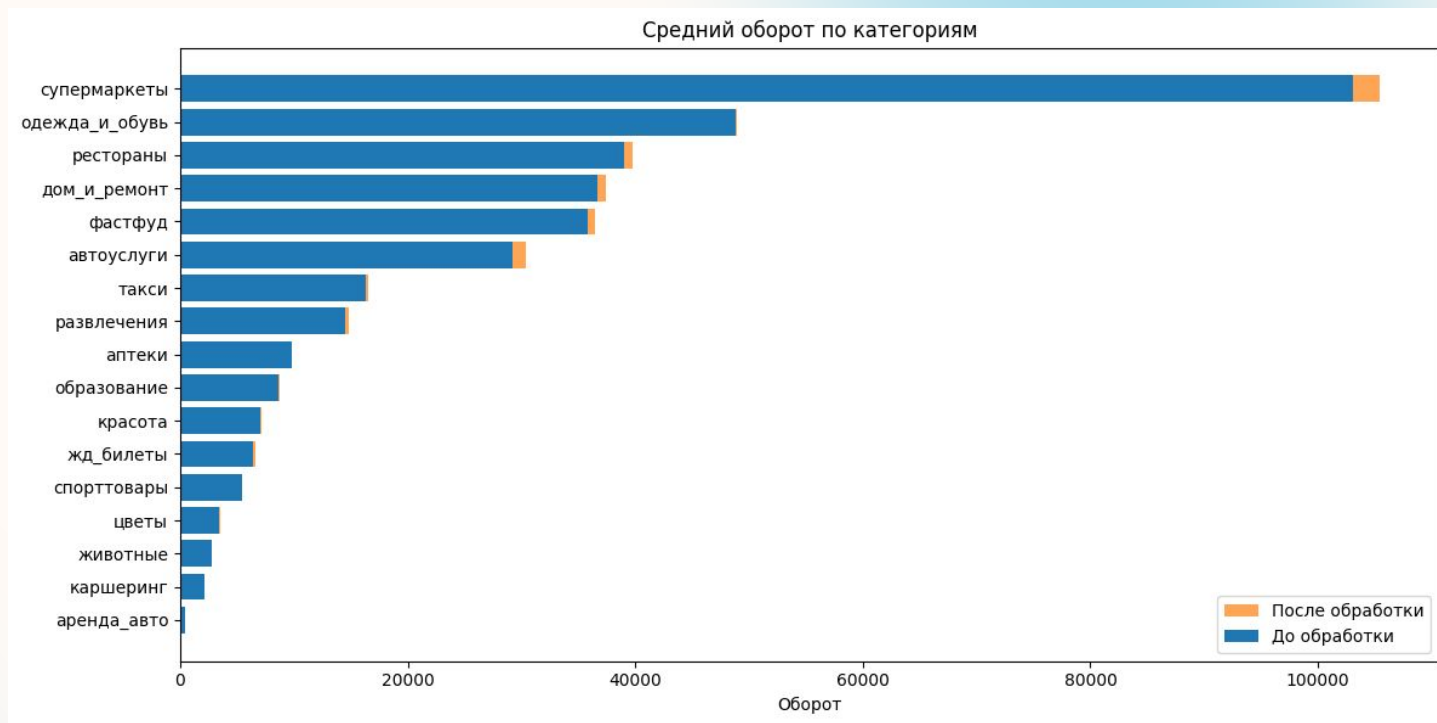
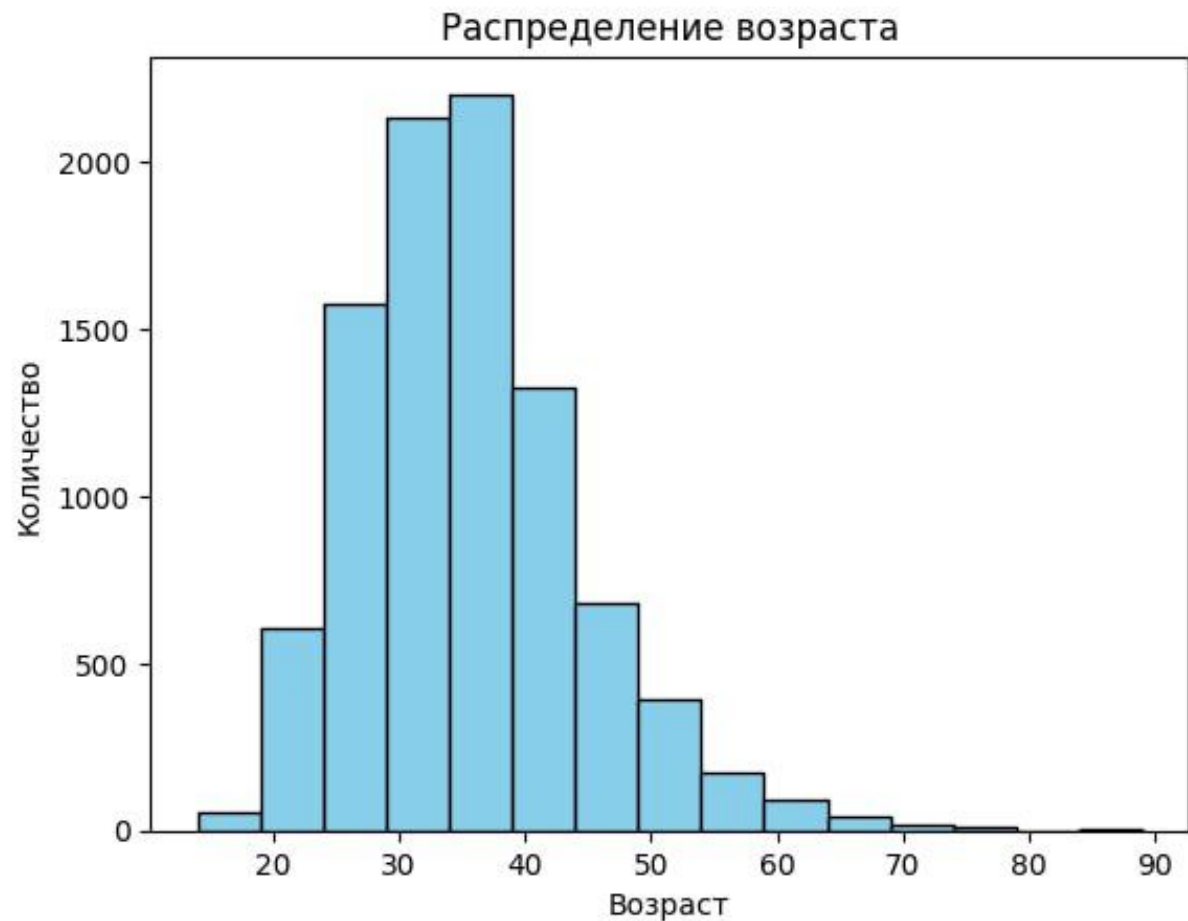


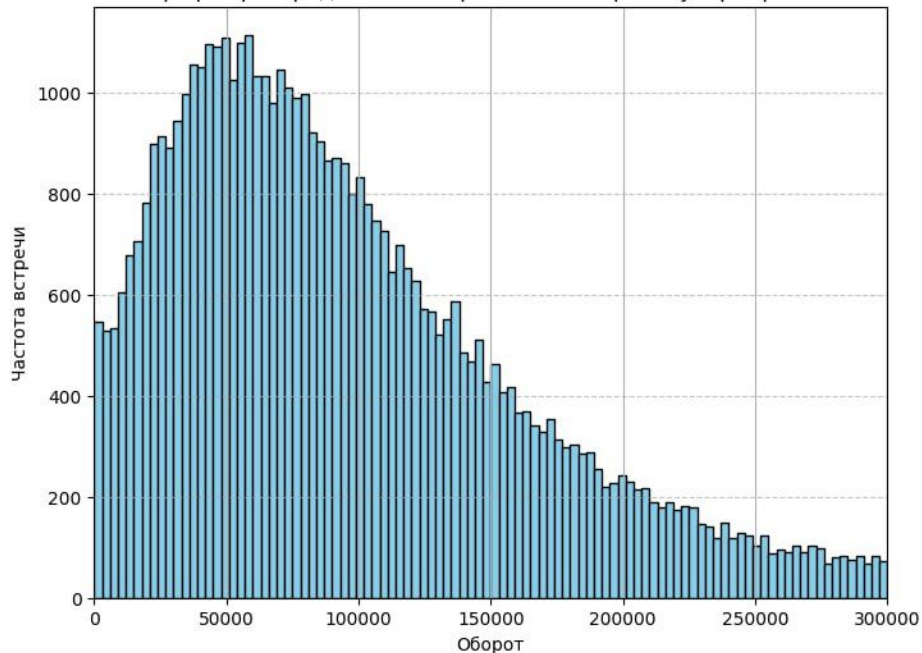
График сравнения средних оборотов по всем категориям до и после обработки от выбросов

Распределение пользователей по возрасту



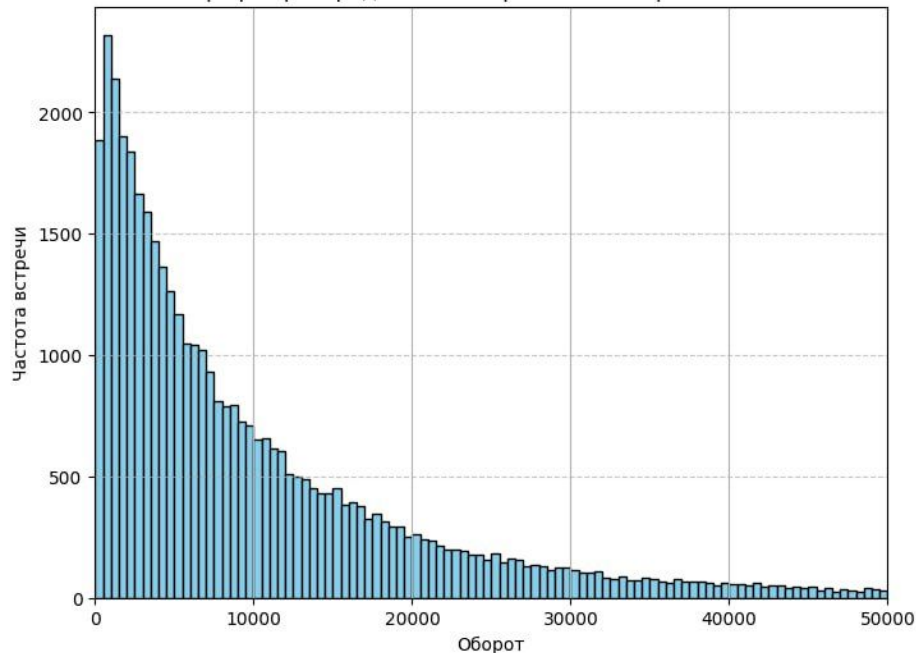
Сравнение распределения оборотов

График распределения оборота в категории: супермаркеты



Супермаркеты

График распределения оборота в категории: аптеки



Все остальные категорий
на примере категории “аптеки”



02

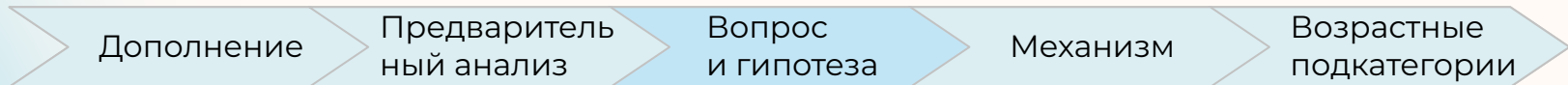
Постановка задачи

Исследовательский вопрос и гипотеза

Как активация категории повышенного кэшбэка влияет на оборот пользователя в этой категории?

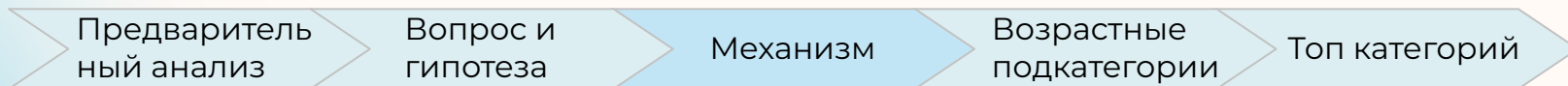


Активация пользователем категории повышенного кэшбэка положительно влияет на увеличение оборота пользователей в этой категории



Механизм гипотезы

1. Распределение пользователей на подкатегории по возрасту
2. Определить топ категорий кэшбэка по кол-ву предложений активации для возрастных категорий
3. Посчитать средний оборот по выбранным категориям с активированным, неактивированным и не предложенным кэшбэком
4. Найти зависимость оборота в категории от активации повышенного кэшбэка с помощью матмодели





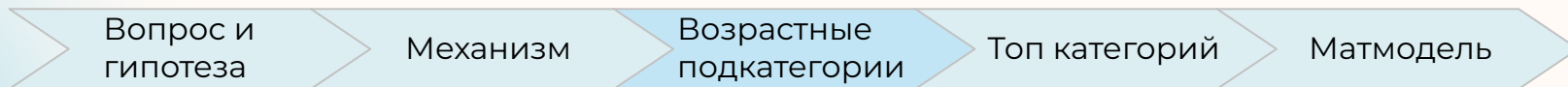
03

Аналитический блок

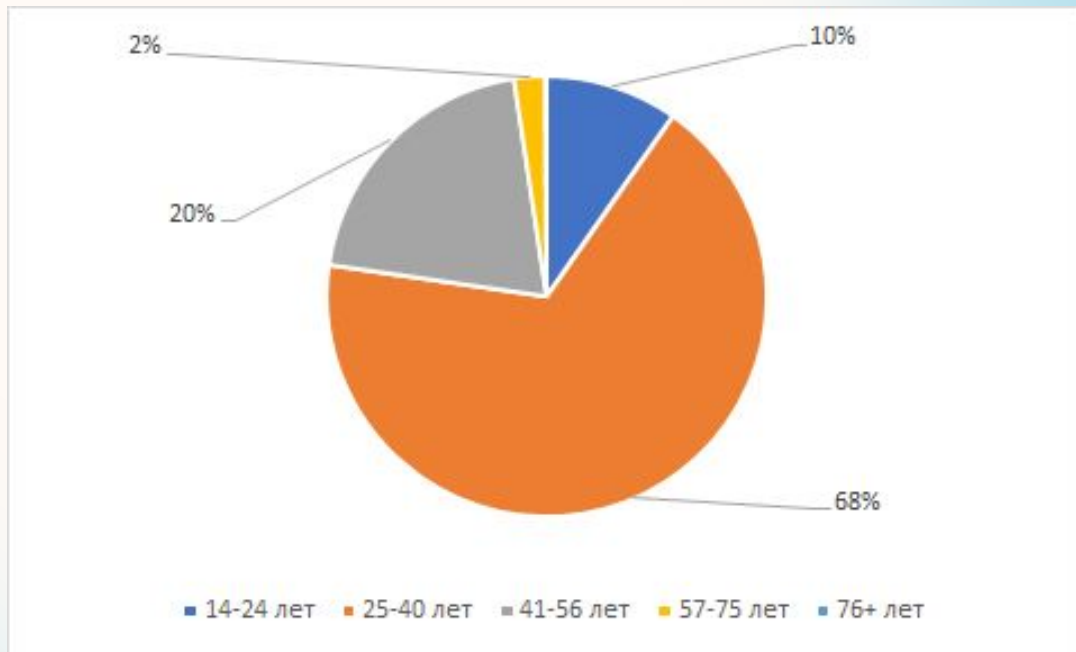
Распределение пользователей на группы по возрасту ¹



1 – [Статья financejar](#), “Average Credit Score by Age: Statistics and Trends” Victoria Scanlon



Полученное распределение пользователей на подкатегории по возрасту



14-24 лет	5 436 чел.
25-40 лет	37 700 чел.
41-56 лет	11 346 чел.
57-75 лет	1 284 чел.
76+ лет	48 чел.

Вопрос и гипотеза

Механизм

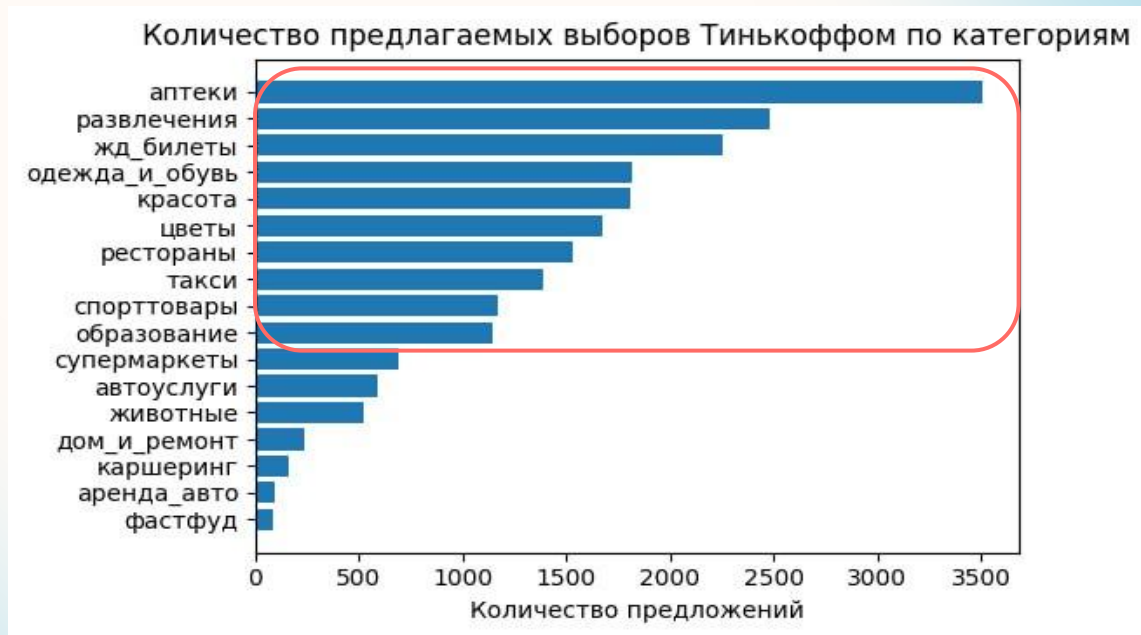
Возрастные подкатегории

Топ категорий

Матмодель

Критерий определения категорий повышенного кэшбэка для исследования.

Количество предложений категорий повышенного кэшбэка на примере 1-ой (14-24 лет) возрастной группы



Определен топ-10 для каждой возрастной группы

Механизм

Возрастные подкатегории

Топ категорий

Матмодель

Результаты матмодели

Топовые категории кэшбэка по возрастным группам.

14-24 лет

25-40 лет

41-56 лет

57-75 лет

76+ лет

Аптеки, Развлечения, Ж/Д билеты,
Красота, Одежда и обувь, Рестораны,
Цветы

- | | | | | |
|---------------|---------------|---------------|---------------|---------------------|
| → Спорттовары | → Спорттовары | → Спорттовары | → Спорттовары | Животные |
| → Образование | → Образование | → Образование | → Образование | Супермаркеты |
| Такси | → Автоуслуги | → Автоуслуги | → Автоуслуги | → Автоуслуги |

Механизм

Возрастные
подкатегории

Топ категорий

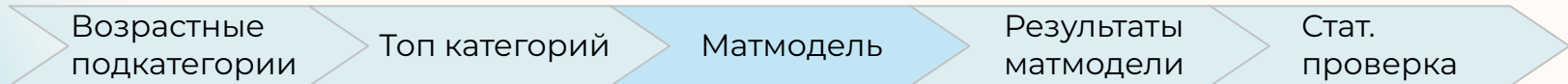
Матмодель

Результаты
матмодели

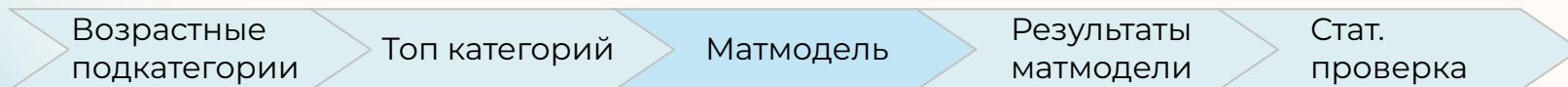
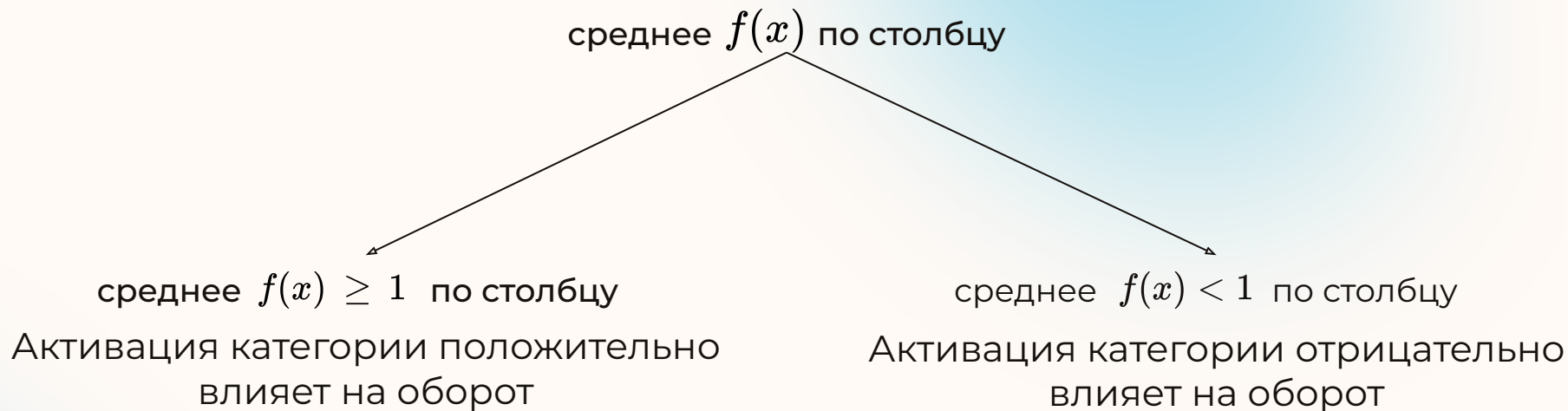
Математическая модель определения зависимости оборота в категории от ее активации

$$f(x) = \frac{\frac{1}{\text{sum}(\text{активация}_x)} \sum_{i=1}^6 \text{оборот}_{x_i} * \text{активация}_{x_i}}{\frac{1}{\text{sum}(\text{not}(\text{активация}_x) + (\text{активация}_x == \text{None}))} \sum_{i=1}^6 (\text{оборот}_{x_i} * \text{not}(\text{активация}_{x_i}) + \text{оборот}_{x_i} * (\text{активация}_{x_i} = \text{None}))}$$

- $f(x)$ – функция определения зависимости оборота в категории от активации кэшбэка по ней для отдельного пользователя.
- x – id пользователя
- активация – матрица массивов из 0, 1 или None, показатель активации категории.
- оборот – матрица массивов оборотов в категории за 6 месяцев.



Интерпретация математической модели определения зависимости оборота в категории от ее активации



Результаты матмодели

Для групп 14-24, 25-40, 41-56, 57-75 для всех пользователей среднее $f(x) \geq 1$ в топ 10 категорий. Это показывает положительную зависимость.

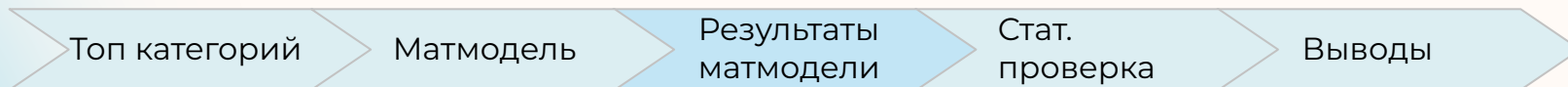
Для категории 76+:

5 категорий где среднее $f(x) \geq 1$:

- Рестораны
- Жд билеты
- Автоуслуги
- Развлечения
- супермаркеты

5 категории где среднее $f(x) < 1$:

- Одежда и обувь
- Аптеки
- Цветы
- Красота
- животные



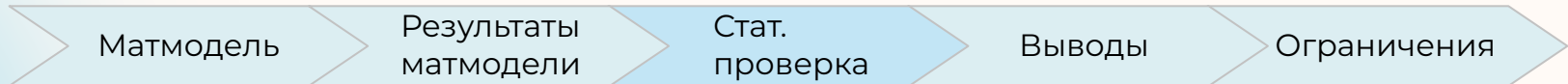
Статистическая проверка

Нулевая гипотеза (H_0):

Активация пользователем категории повышенного кэшбэка НЕ ВЛИЯЕТ на увеличение оборота пользователей в этой категории.

Альтернативная гипотеза (H_1):

Активация пользователем категории повышенного кэшбэка ВЛИЯЕТ на увеличение оборота пользователей в этой категории.



Результаты T-test-ов для каждой возрастной группы по обороту с активированным кэшбеком и без него.

14-24 лет: H0: 0/10 категорий, H1: подтверждено для 10/10 топовых категорий

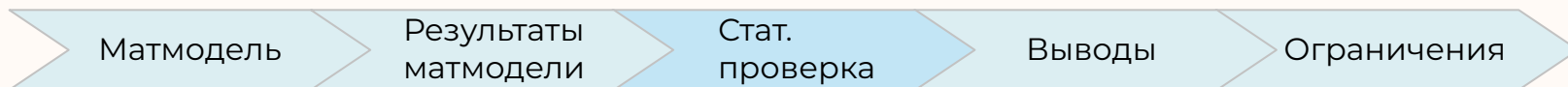
25-40 лет: H0: 0/10 категорий, H1: подтверждено для 10/10 топовых категорий

41-56 лет : H0: 0/10 категорий, H1: подтверждено для 10/10 топовых категорий

57-75 лет: H0: 0/10 категорий, H1: подтверждено для 10/10 топовых категорий

76+ лет: H0: 5/10 категорий, H1: подтверждено для 5/10 топовых категорий

Это означает, что у нас есть достаточно доказательств, чтобы подтвердить альтернативную гипотезу.



Средний оборот по каждой из категорий повышенного кэшбэка

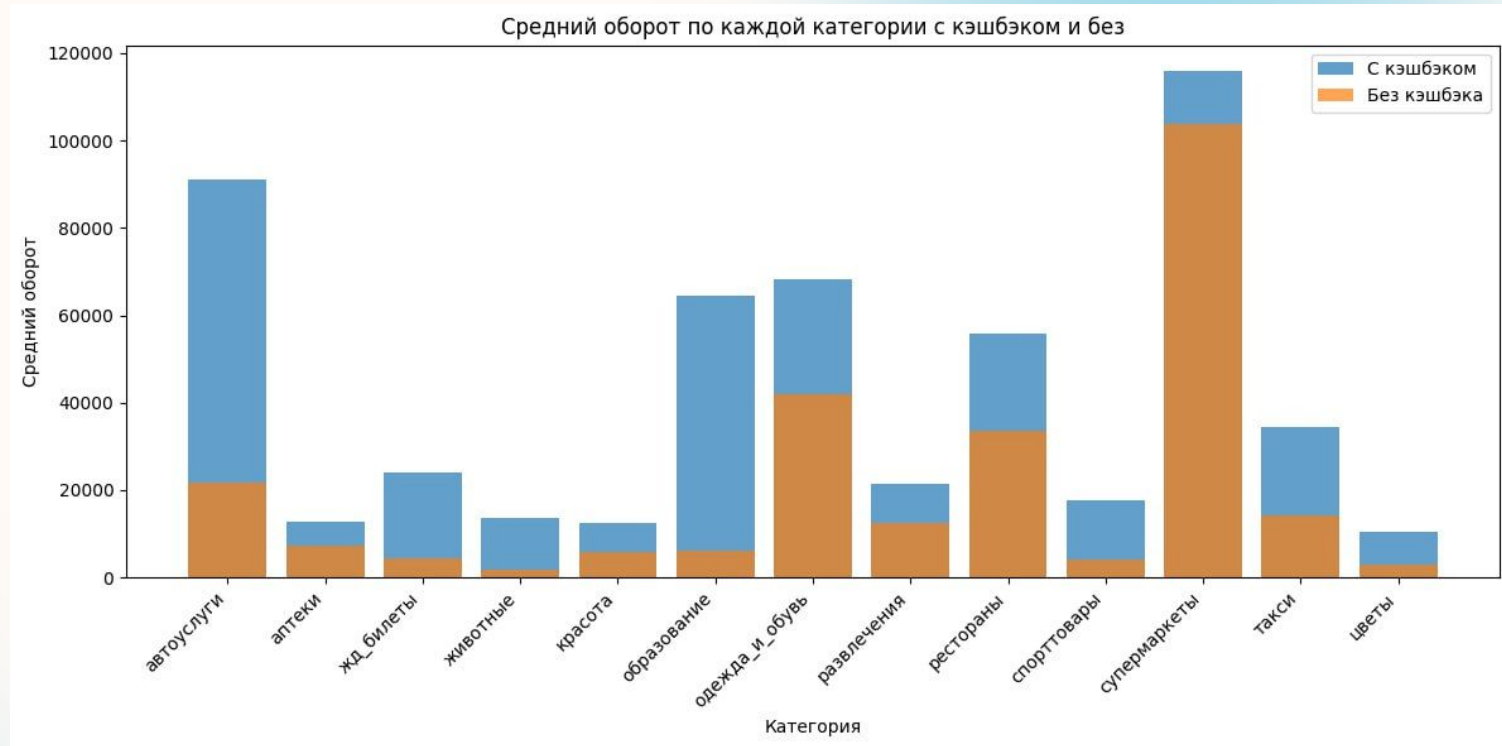


График сравнения оборотов пользователей с активированным кэшбэком (синий) и неактивированным кэшбэком в этой категории (оранжевый).

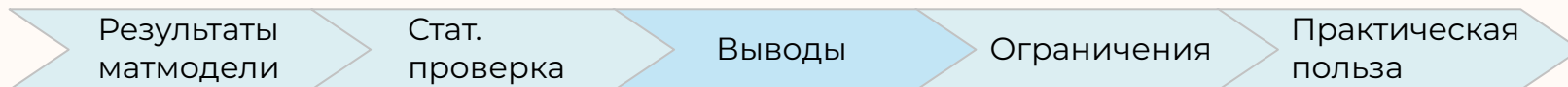


04

Заключение

Вывод по гипотезе

С помощью приведенной математической модели и статистической проверки **гипотеза подтвердилась** для всех возрастных групп пользователей в топ-10 категориях повышенного кэшбэка – активация повышенного кэшбэка положительной влияет на увеличение оборота пользователя в этой категории.



Ограничения



TINKOFF

Результаты исследования можно обобщить только на пользователей Тинькофф



Данные не обо всех регионах России (нет 8 регионов)



Нельзя обобщить исследование для пользователей 76+ из-за слишком маленького количества пользователей

Стат.
проверка

Выводы

Ограничения

Практическая
польза

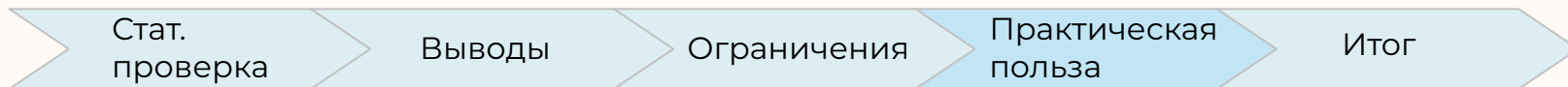
Итог

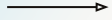
Практическая польза

Улучшение алгоритма рекомендации повышенного кэшбэка пользователям с целью повышения их оборотов в данных категориях.

Policy implication

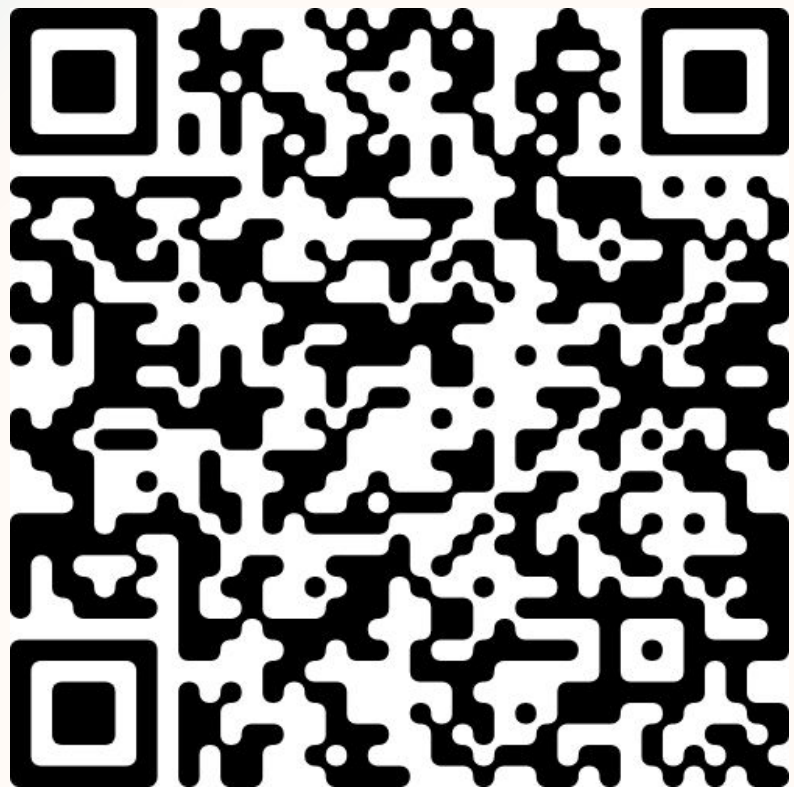
При сотрудничестве с партнерами Банк может повышать кэшбэк в определенных категориях для привлечения к ним внимания, чтобы пользователи заинтересовались в совершении трат по партнерской программе.





**ГОТОВЫ ОТВЕТИТЬ
на Ваши вопросы**

Наш Colab с расчетом T-test

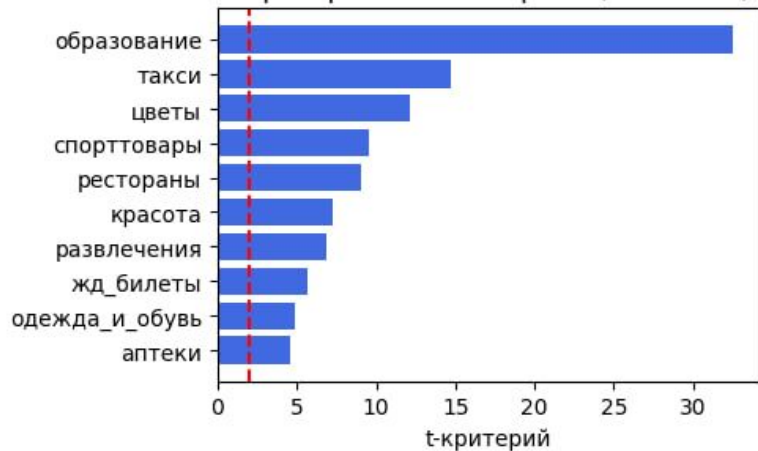


<https://clck.ru/37BtLm>

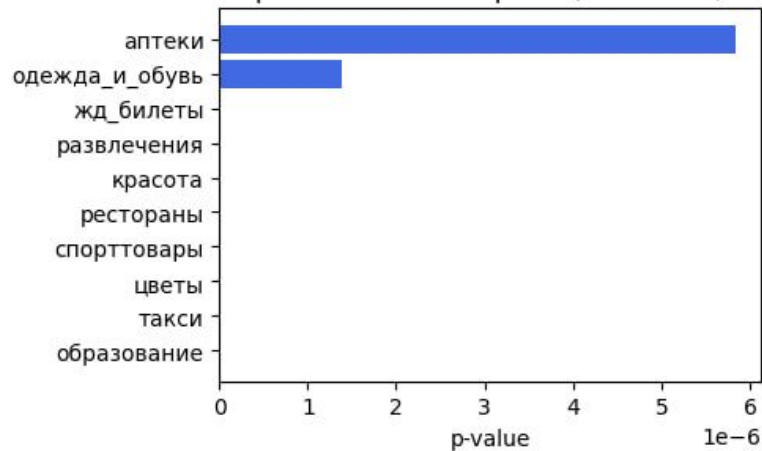


Приложение

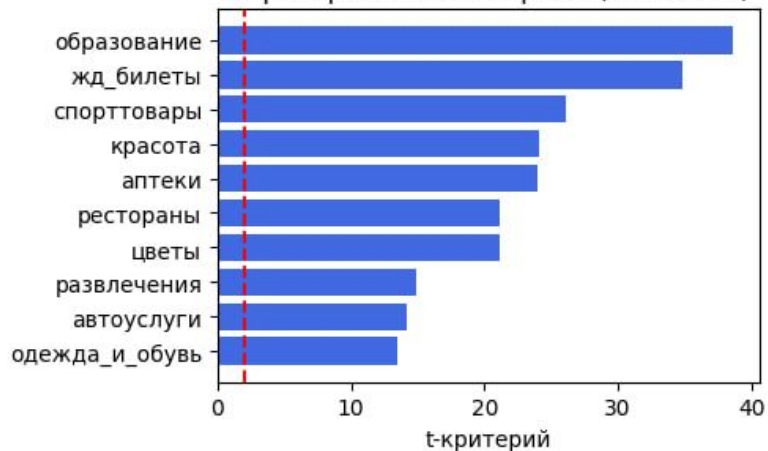
t-критерий по категориям (14-24 лет)



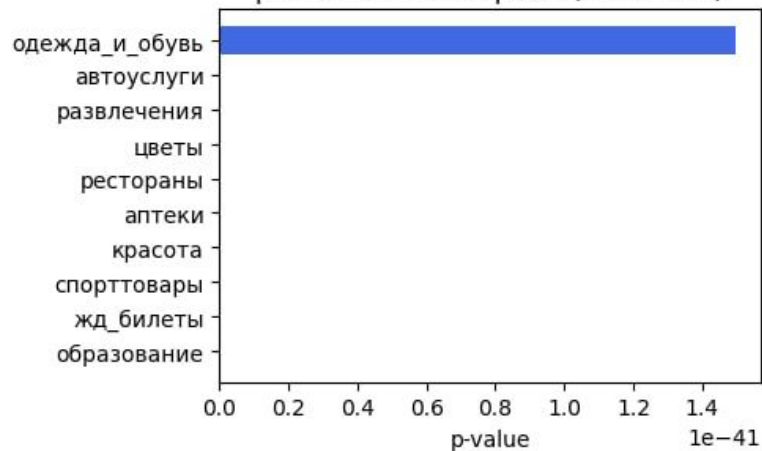
p-value по категориям (14-24 лет)



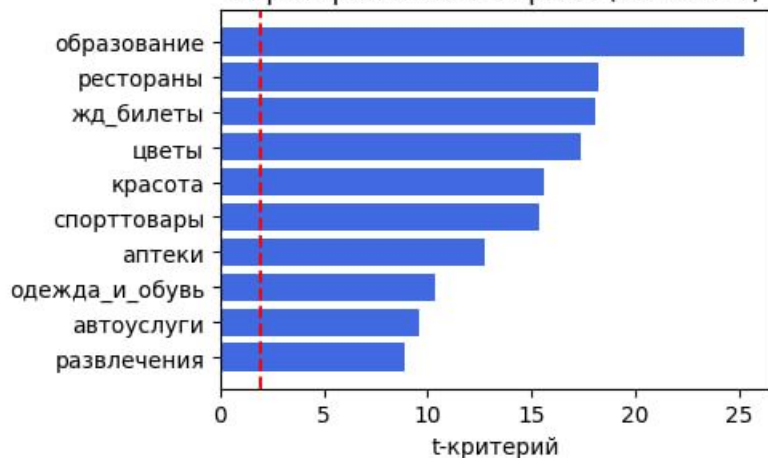
t-критерий по категориям (25-40 лет)



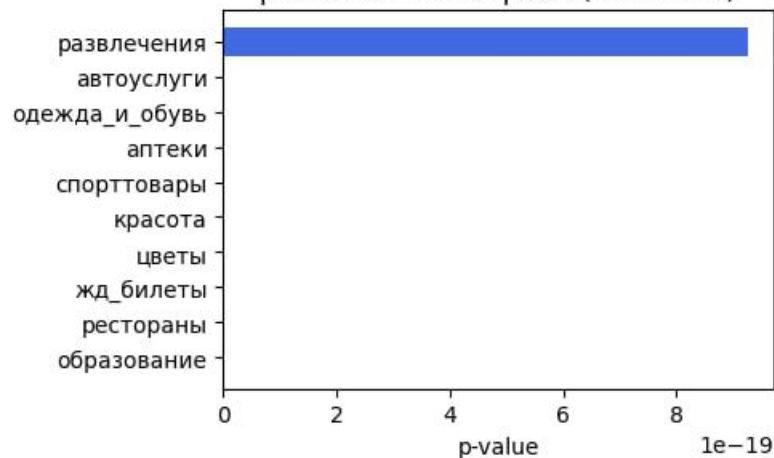
p-value по категориям (25-40 лет)



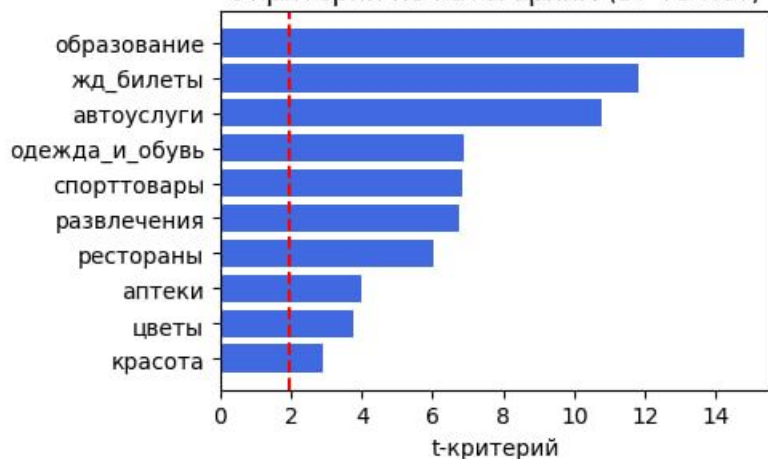
t-критерий по категориям (41-56 лет)



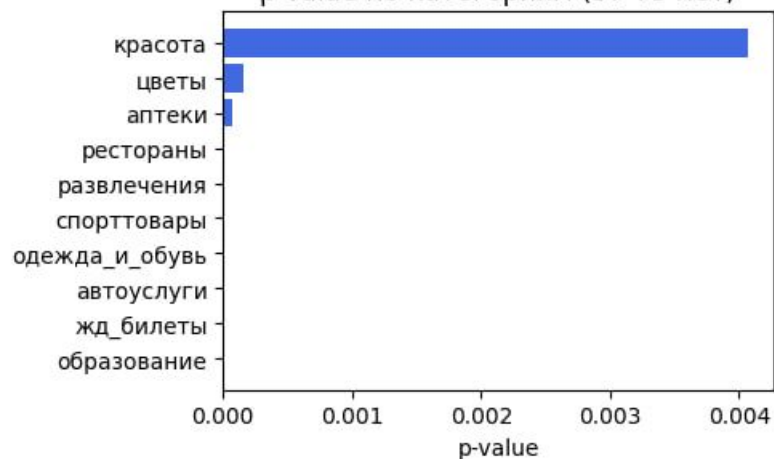
p-value по категориям (41-56 лет)



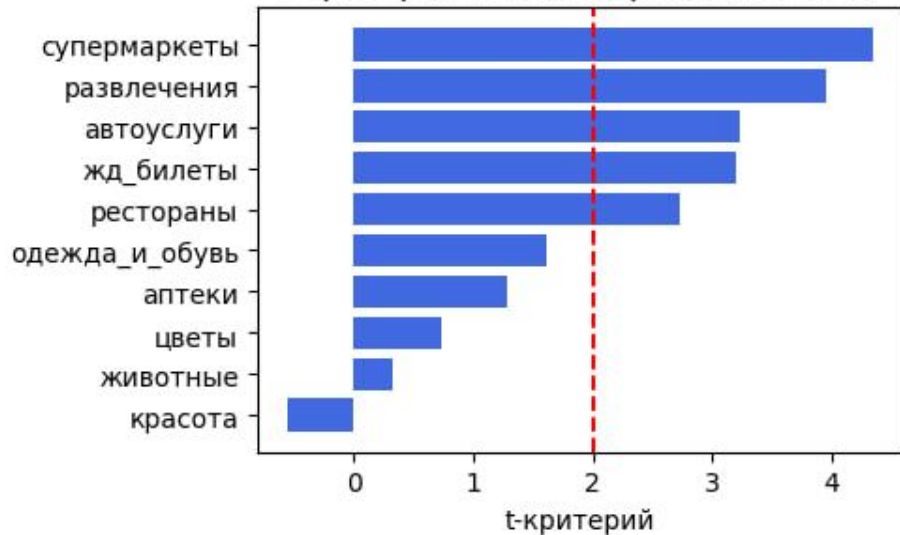
t-критерий по категориям (57-75 лет)



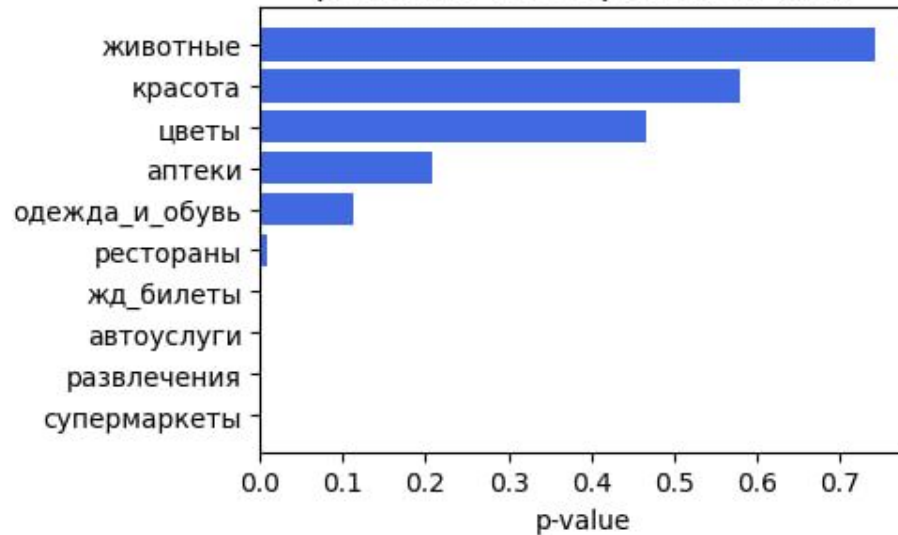
p-value по категориям (57-75 лет)



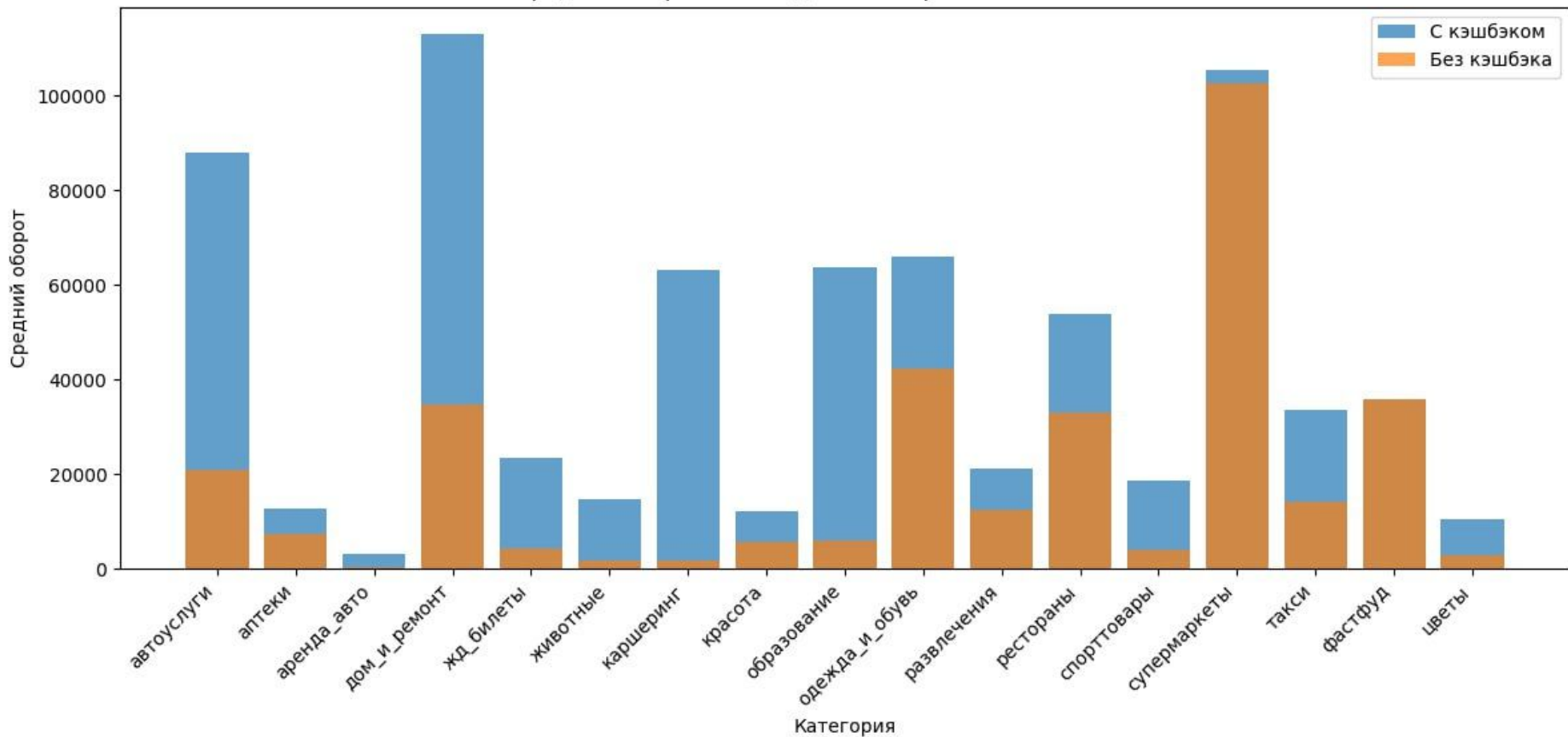
t-критерий по категориям (76+ лет)



p-value по категориям (76+ лет)



Средний оборот по каждой категории с кэшбэком и без



Распределение категорий по влиянию активации повышенного кэшбэка на оборот

	14-24	25-40	41-56	57-75	76+
подтвердилось	аптеки развлечения жд билеты одежда и обувь красота цветы рестораны такси спорттовары образование	аптеки развлечения красота одежда и обувь рестораны жд билеты цветы спорттовары автоуслуги образование	аптеки развлечения красота рестораны одежда и обувь жд билеты цветы спорттовары автоуслуги образование	аптеки развлечения красота цветы жд билеты рестораны одежда и обувь спорттовары автоуслуги образование	рестораны развлечения автоуслуги супермаркеты красота жд билеты
нет					одежда и обувь Цветы Животные Красота аптеки

Сколько раз банк предлагал каждую категорию

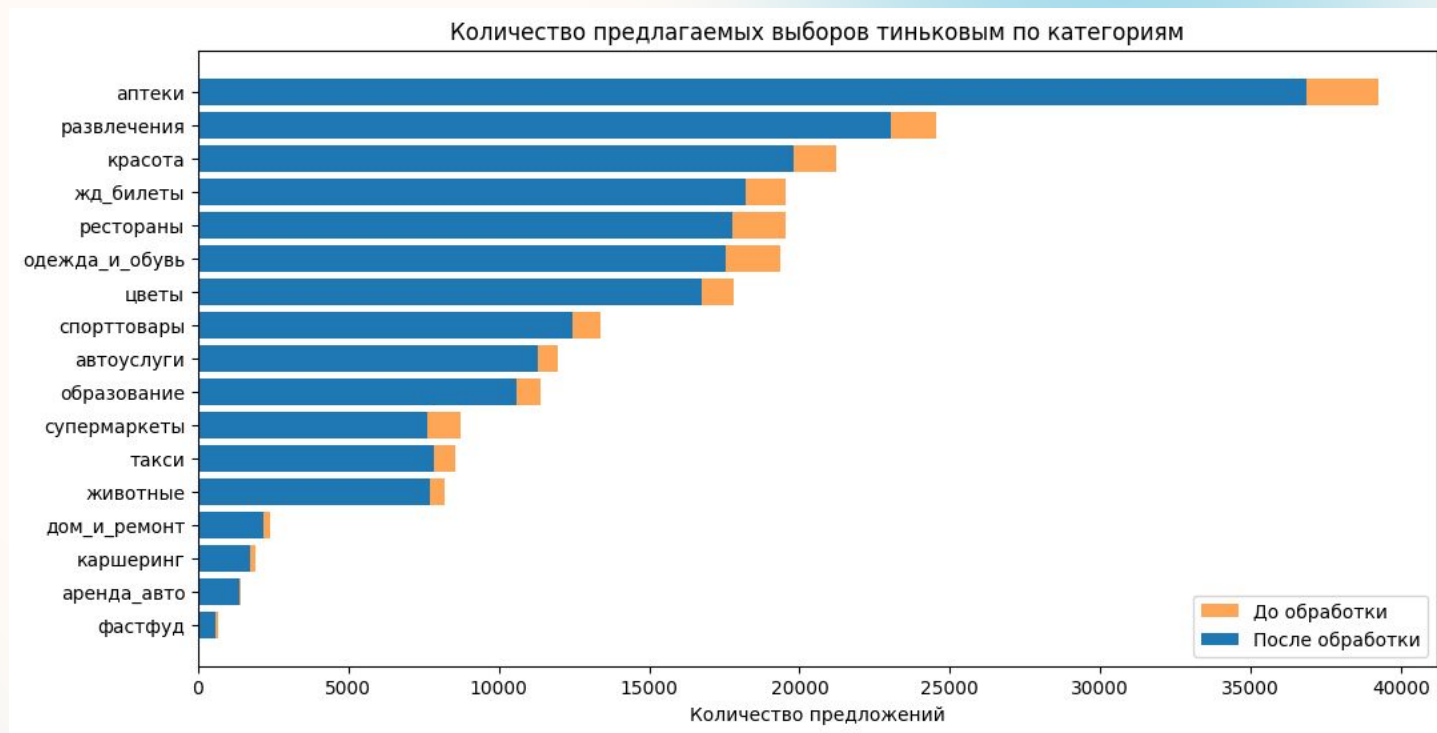


График сравнения предложений категорий повышенного кэшбэка до и после обработки от выбросов

Сколько раз пользователь активировал каждую категорию

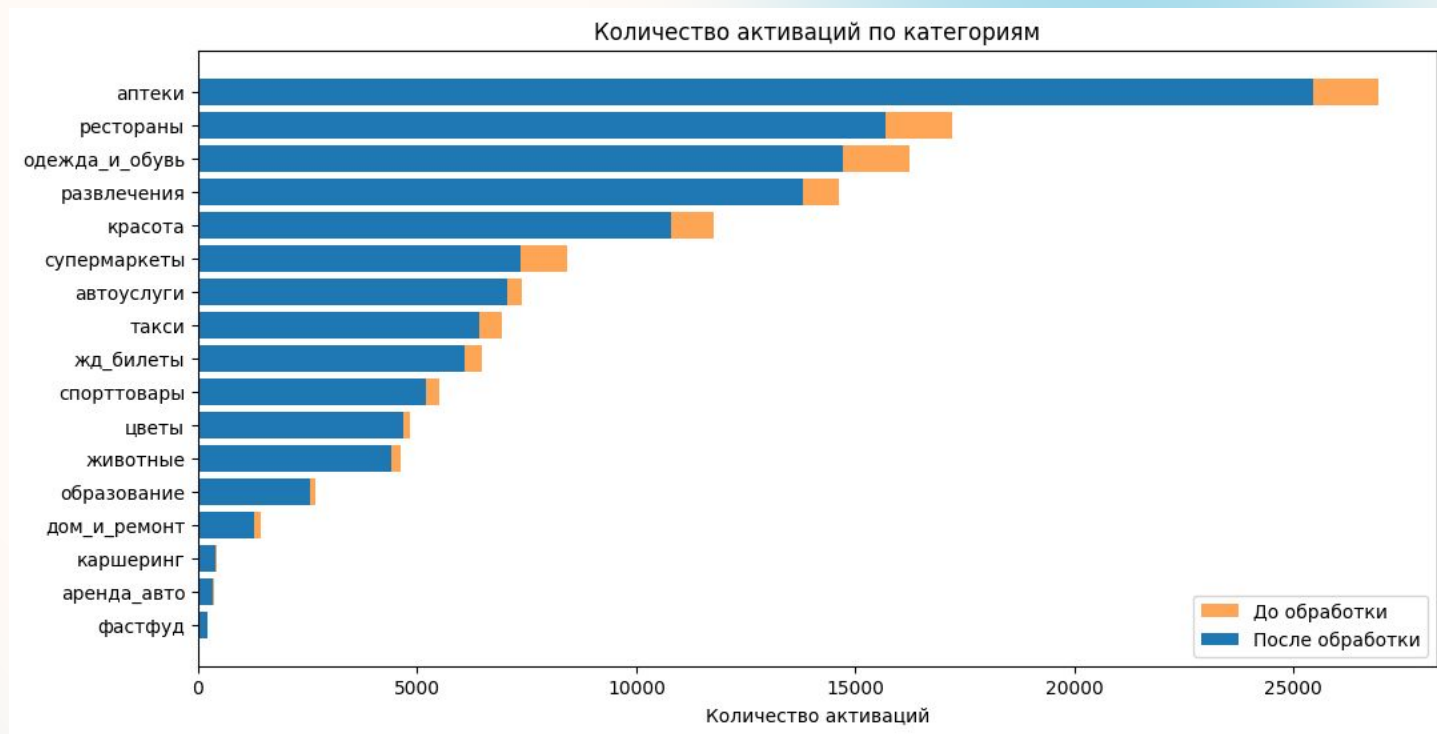
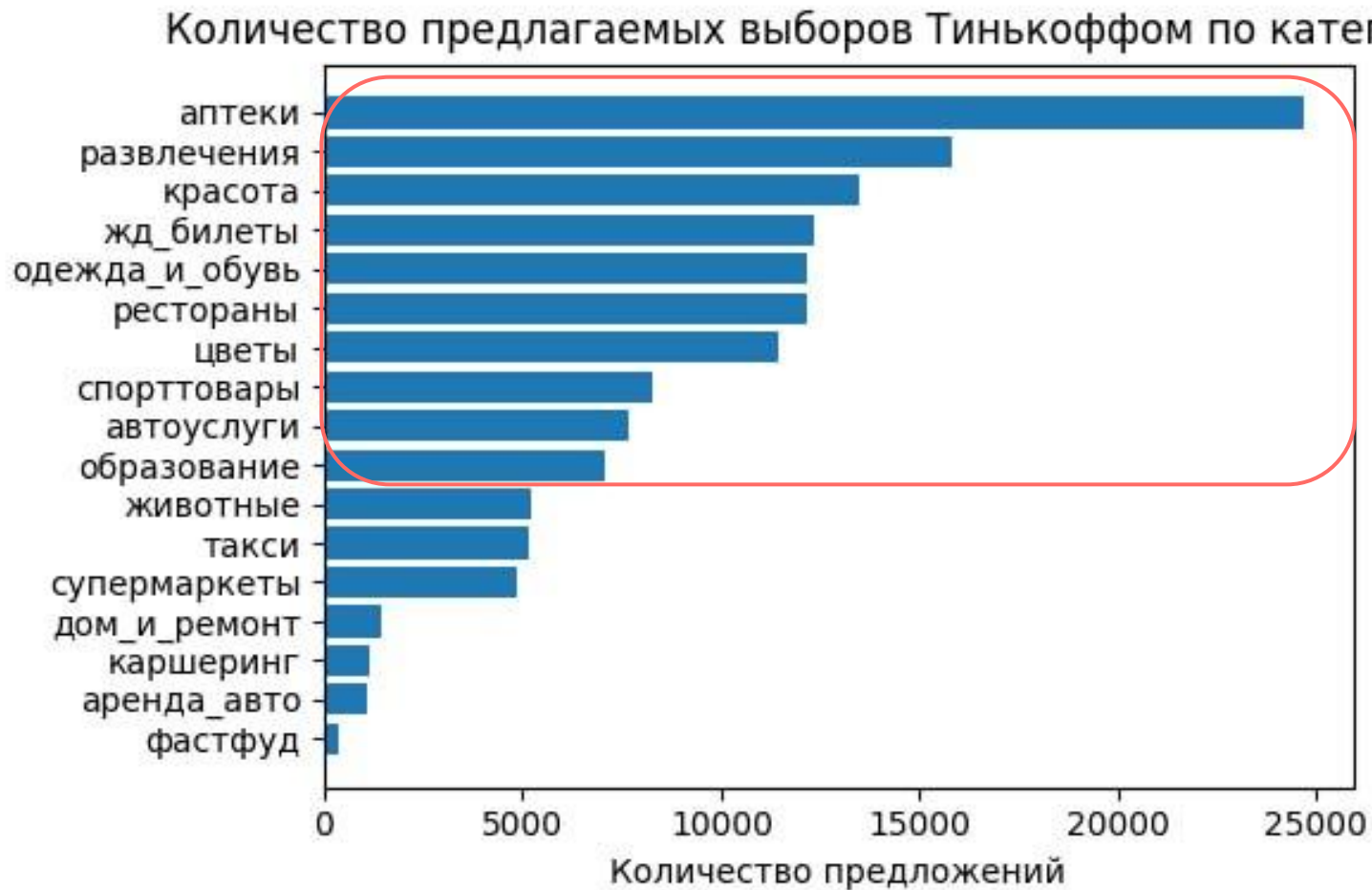


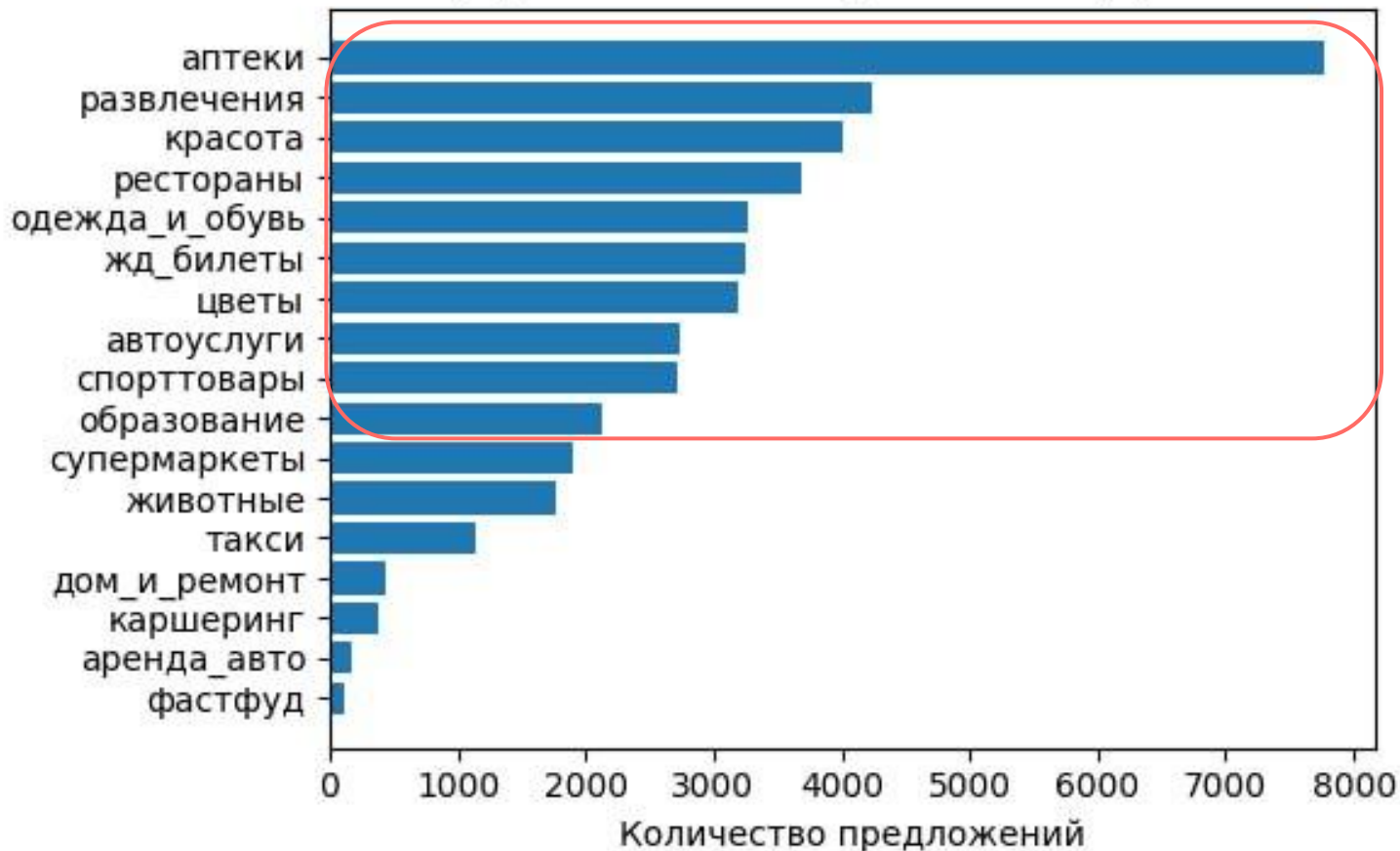
График сравнения количества активаций категорий повышенного кэшбэка до и после обработки от выбросов

Количество предложений категорий повышенного кэшбэка. 25-40 лет



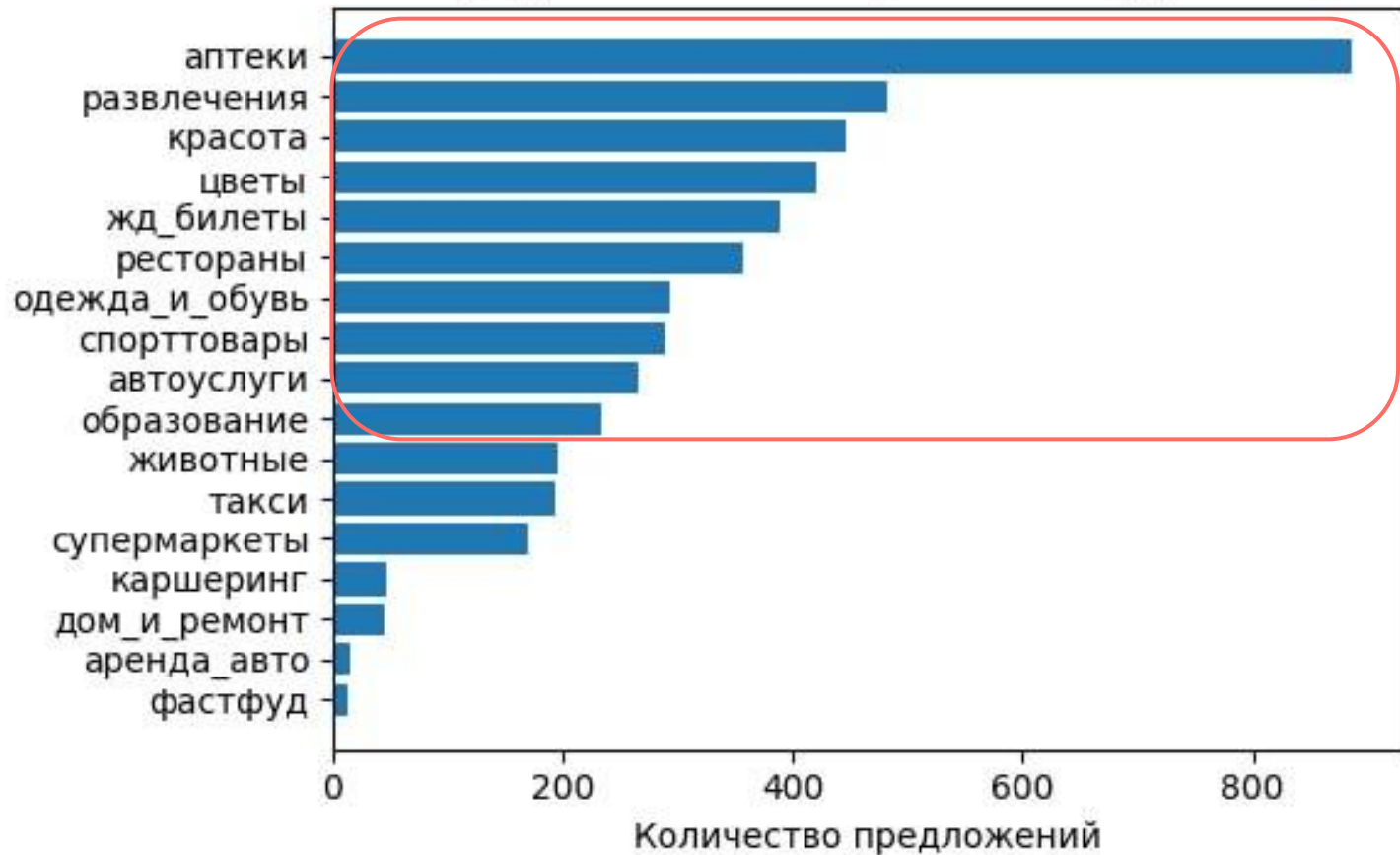
Количество предложений категорий повышенного кэшбэка. 41-56 лет

Количество предлагаемых выборов Тинькоффом по категориям



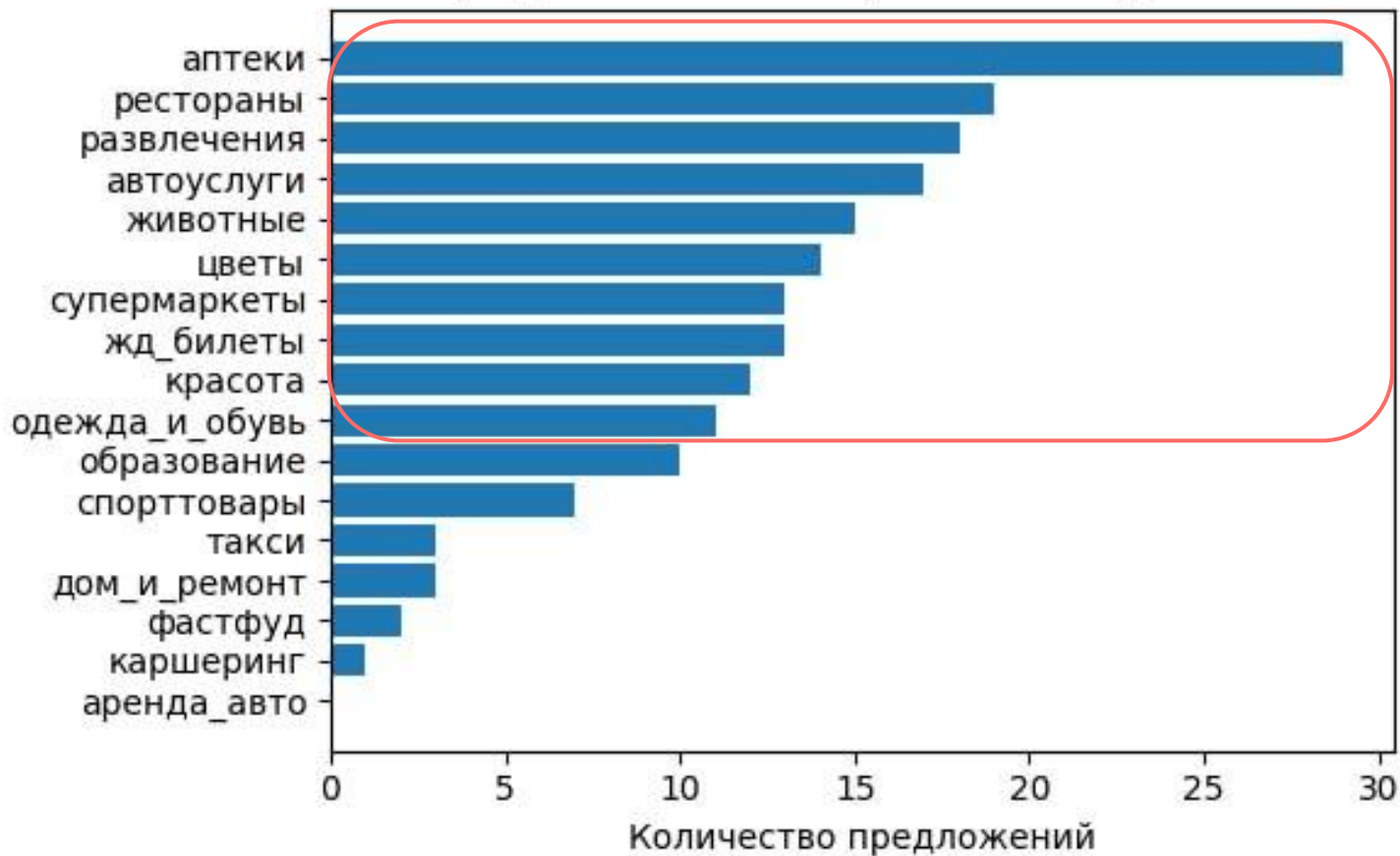
Количество предложений категорий повышенного кэшбэка. 57-75 лет

Количество предлагаемых выборов Тинькоффом по категориям



Количество предложений категорий повышенного кэшбэка.76+ лет

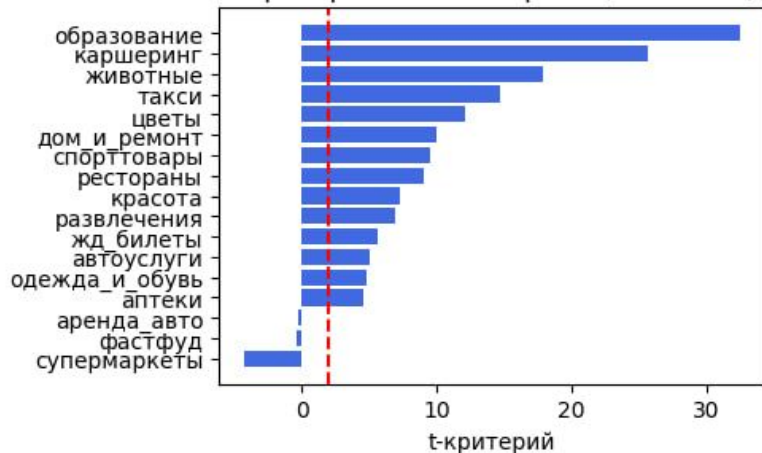
Количество предлагаемых выборов Тинькоффом по категориям



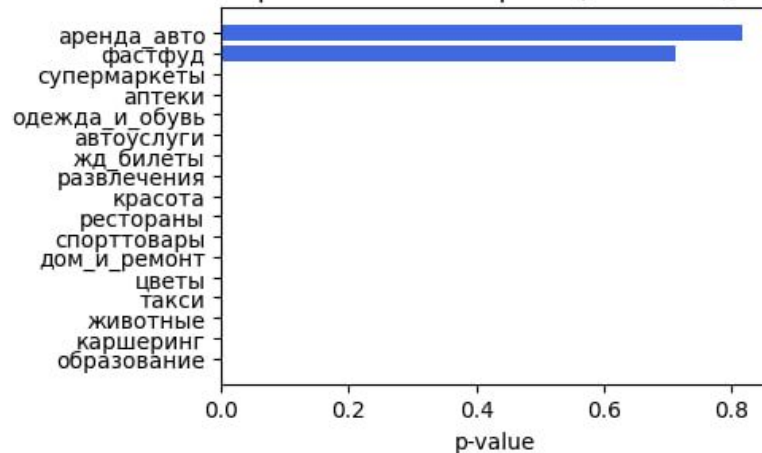
Отсутствующие в датасете регионы

- Донецкая Народная Республика
- Запорожская область
- Луганская Народная Республика
- Республика Алтай
- Республика Ингушетия
- Херсонская область
- Чеченская Республика
- Чукотский автономный округ

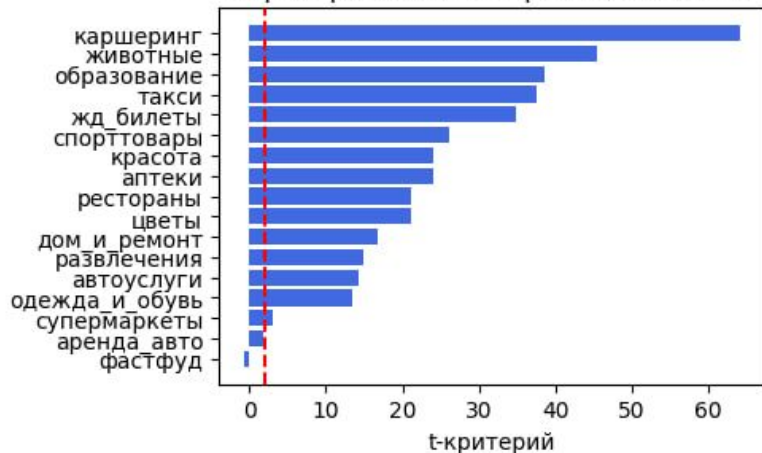
t-критерий по категориям (14-24 лет)



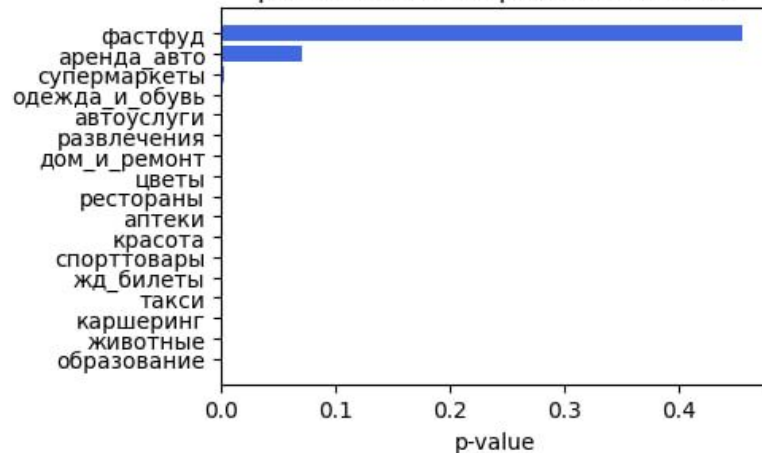
p-value по категориям (14-24 лет)



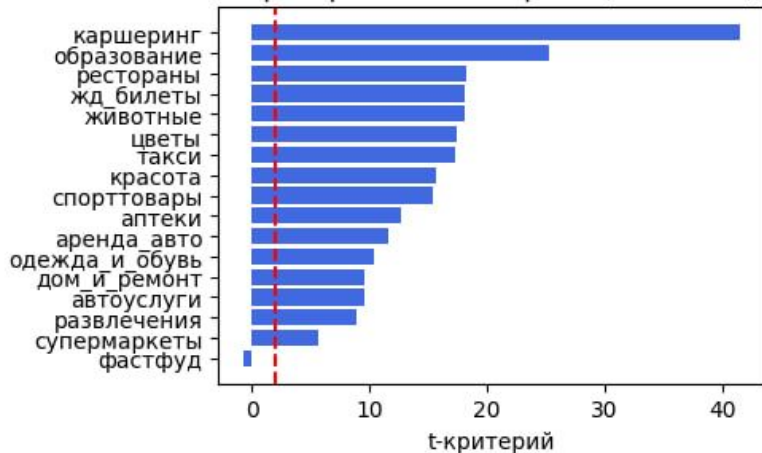
t-критерий по категориям (25-40 лет)



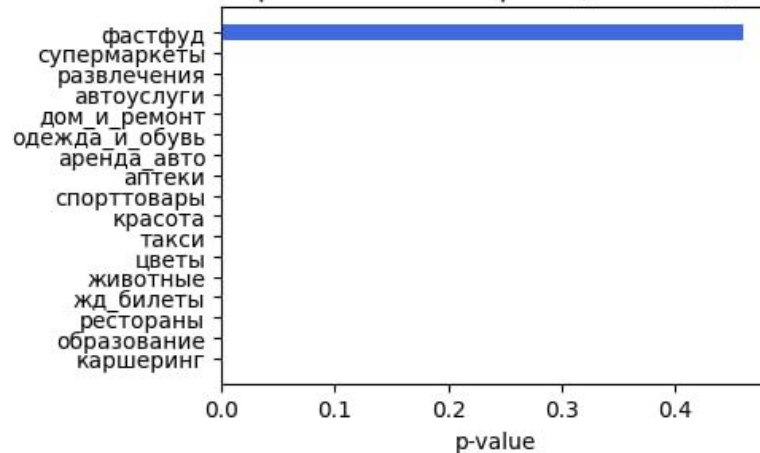
p-value по категориям (25-40 лет)



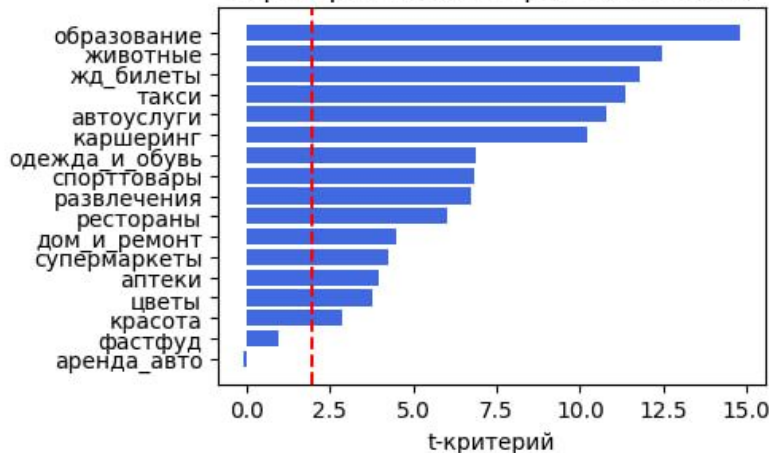
t-критерий по категориям (41-56 лет)



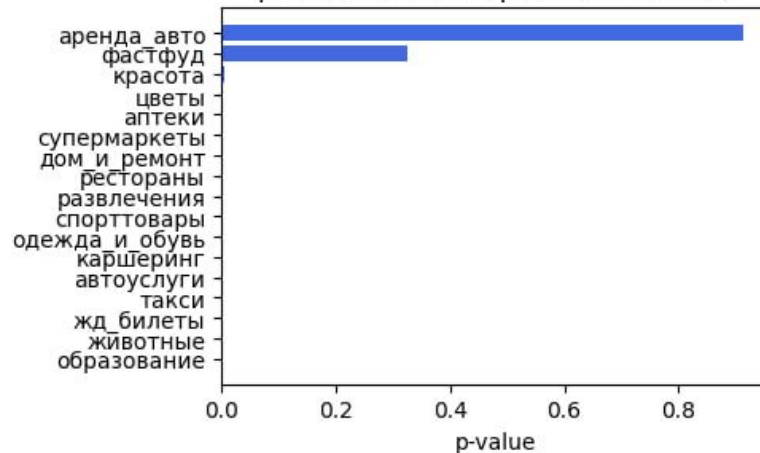
p-value по категориям (41-56 лет)



t-критерий по категориям (57-75 лет)



p-value по категориям (57-75 лет)



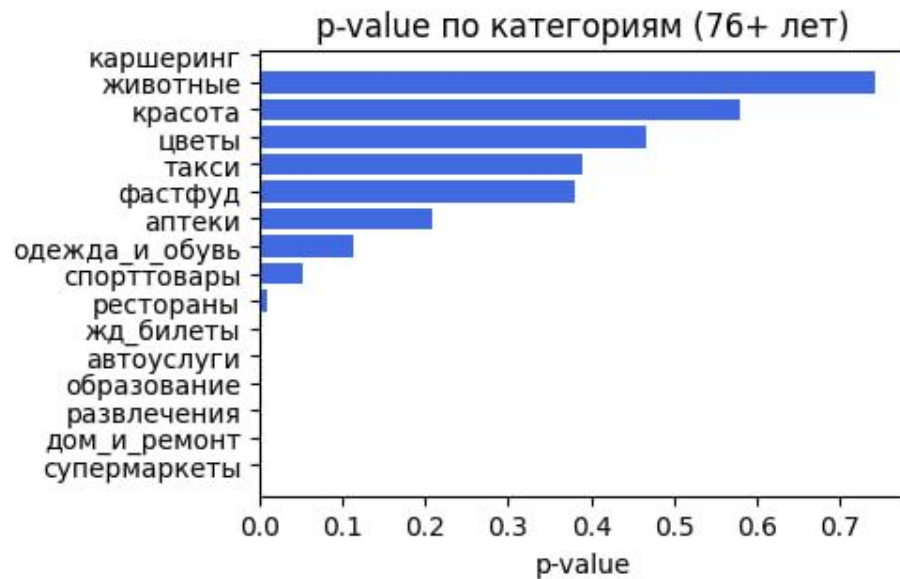
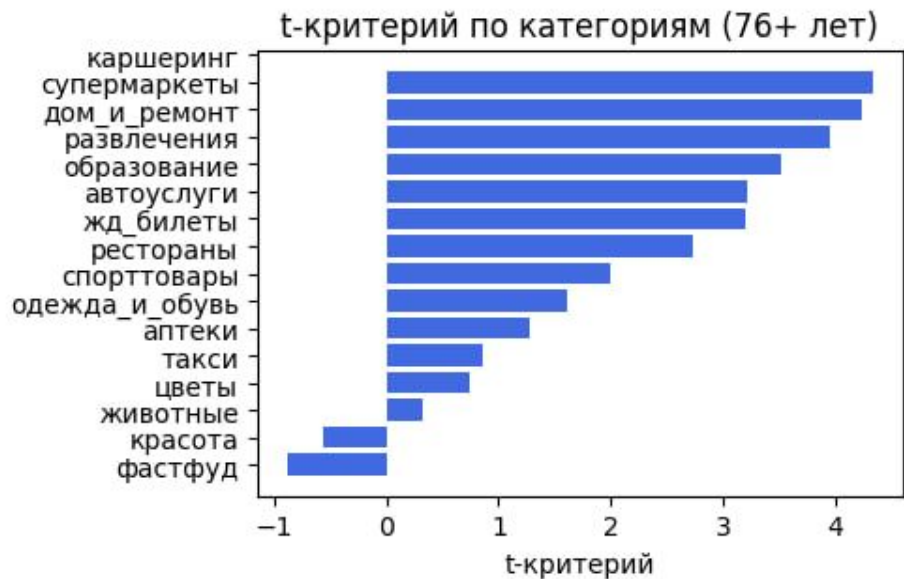


График распределения оборота в категории: рестораны

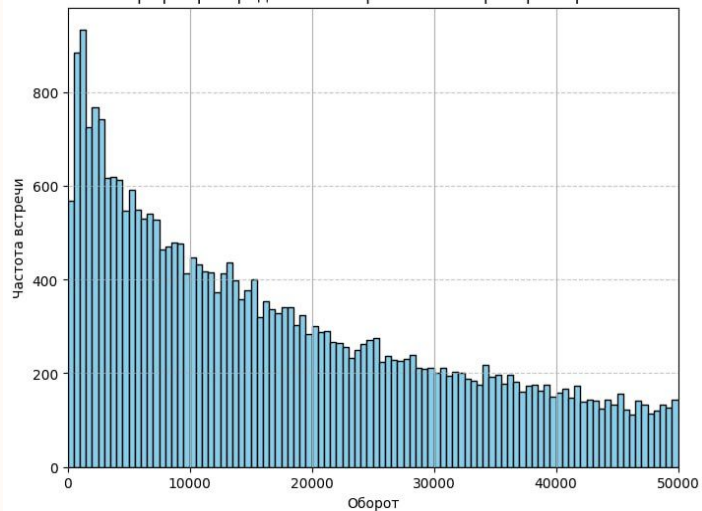


График распределения оборота в категории: одежда_и_обувь

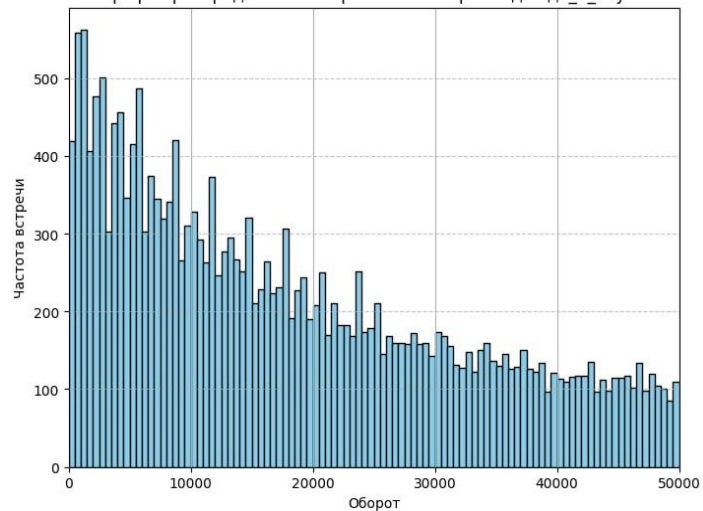


График распределения оборота в категории: автоуслуги

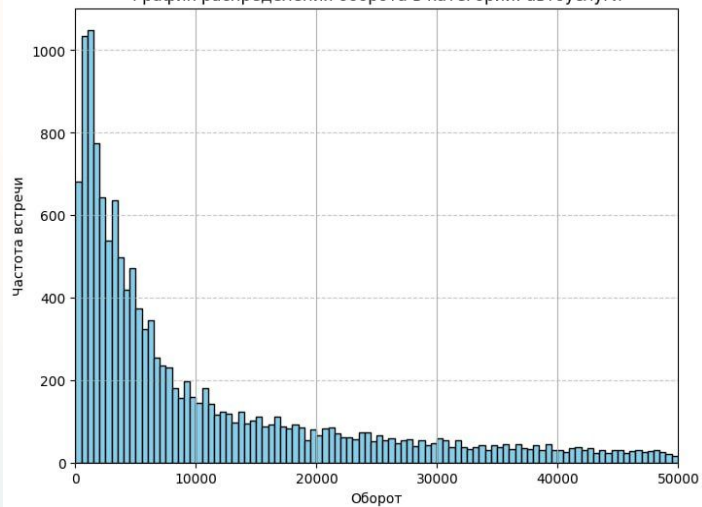


График распределения оборота в категории: такси

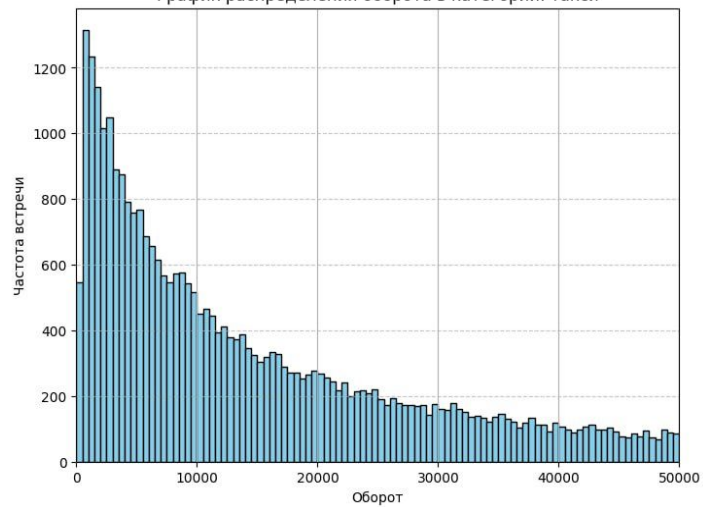


График распределения оборота в категории: красота

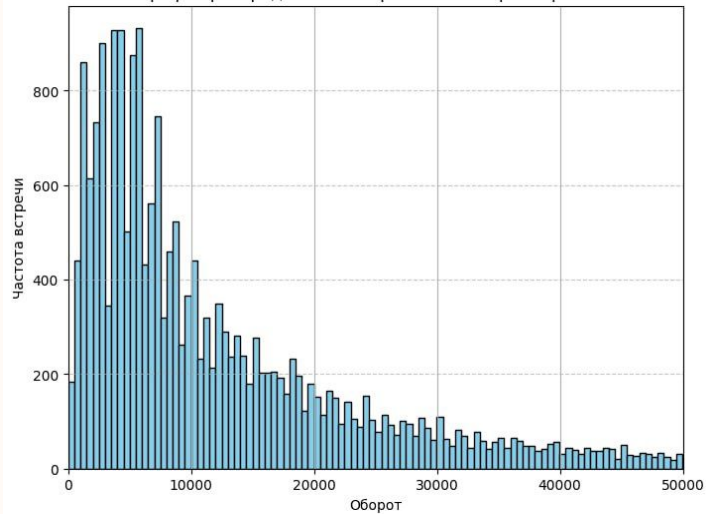


График распределения оборота в категории: развлечения

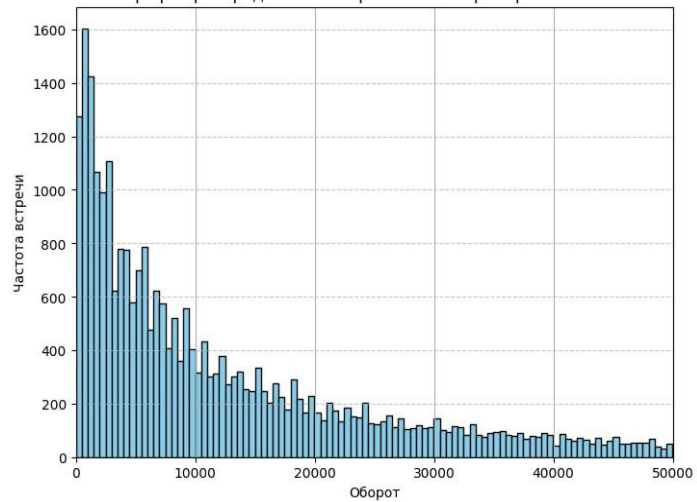


График распределения оборота в категории: образование

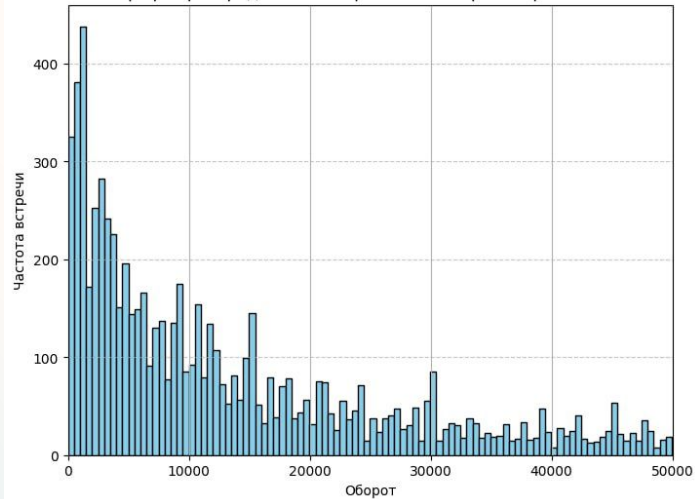


График распределения оборота в категории: жд_билеты

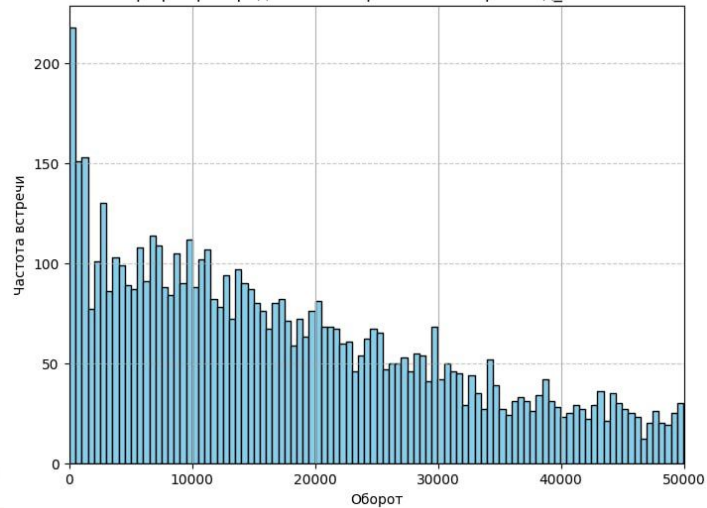


График распределения оборота в категории: дом и ремонт

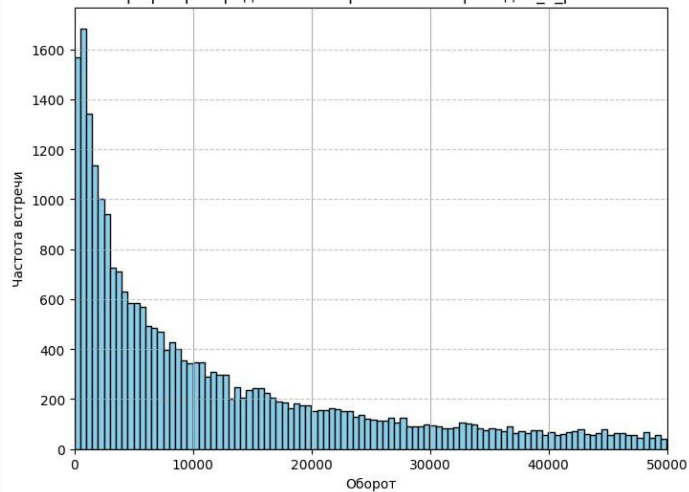


График распределения оборота в категории: животные

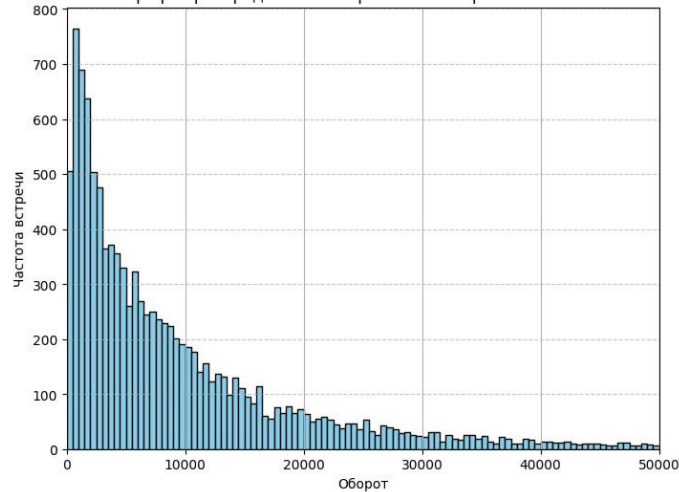


График распределения оборота в категории: спорттовары

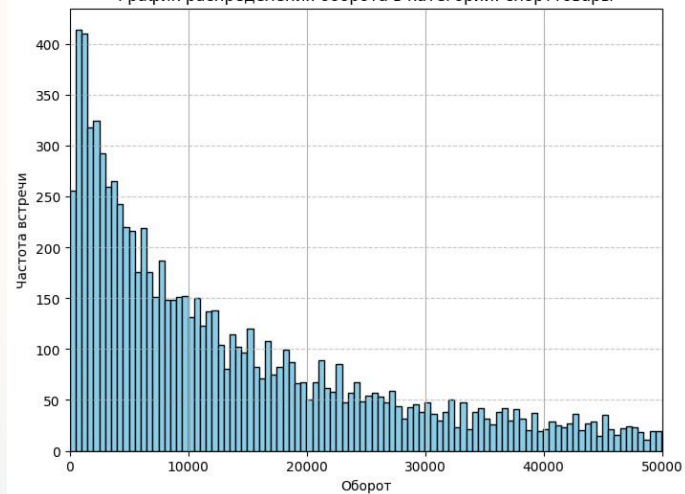


График распределения оборота в категории: цветы

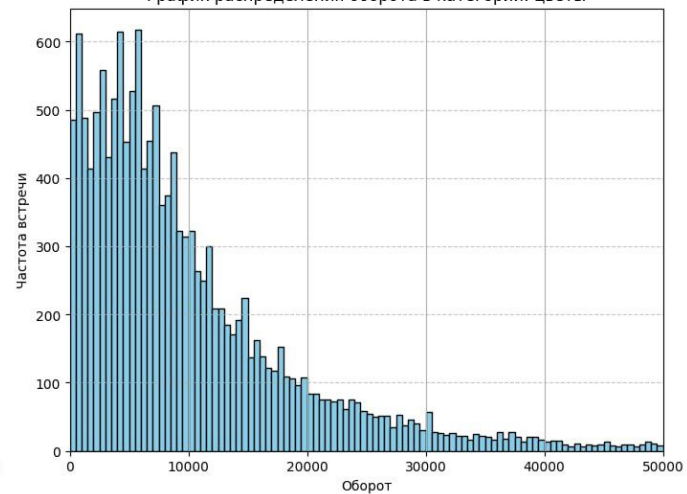


График распределения оборота в категории: каршеринг

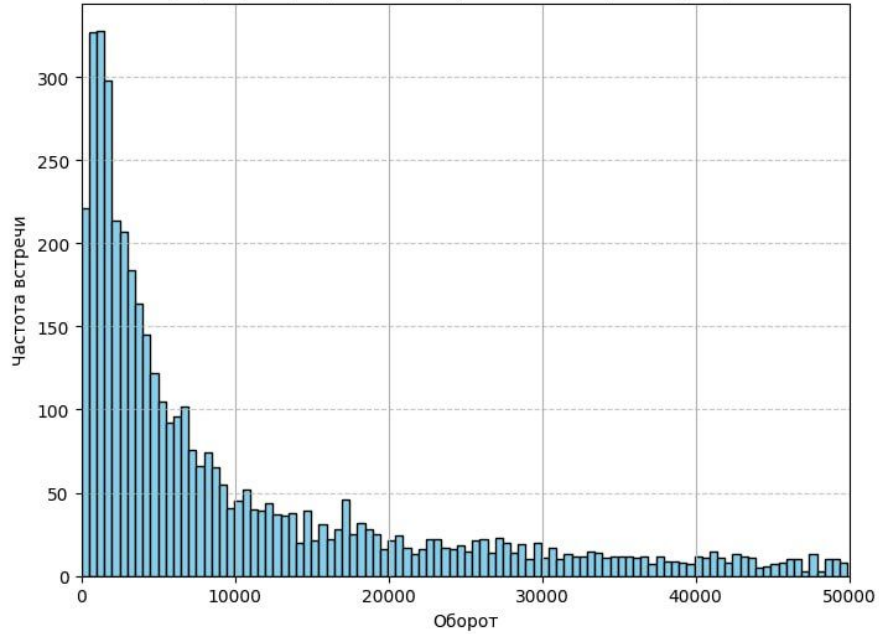
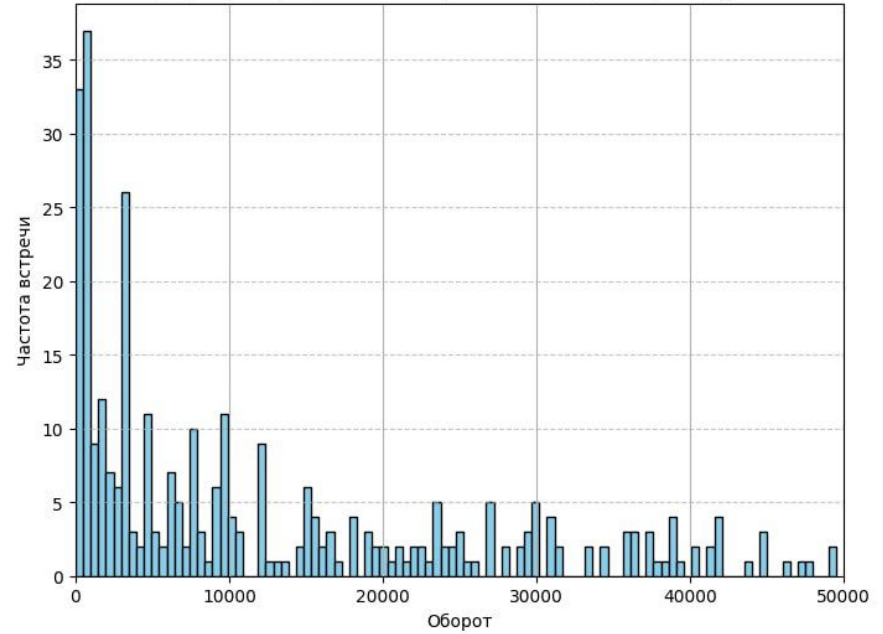


График распределения оборота в категории: аренда_авто



U