

События пользовательского сценария

выброс котов балбесов:

Коровина Ульяна

Кожевников Сергей

Мазитова Нелли

Федорова Мария

Илюхин Семён

Executive Summary



Исследовали влияние пола, возраста и уровня развития связи в регионе на конверсии на этапах покупки



Обнаружили, что у пожилых мужчин из регионов на одном из этапов конверсия ниже на 8%



Разработали решение, которое, как мы предполагаем, должно увеличить конверсию, а значит и количество покупок

Описание данных

Данные о рекламной кампании и трафике: название рекламной кампании, источник и канал трафика

Данные о событии и сессии: тип устройства, id клиента, id сессии, точное время события, название события

Данные о пользователе: id клиента, тариф клиента, является ли участником программы лояльности, авторизован ли, тип покупателя (b2b/b2c), название региона, возраст, пол, rfm*, подключен ли домашний интернет

Все данные не являются количественными



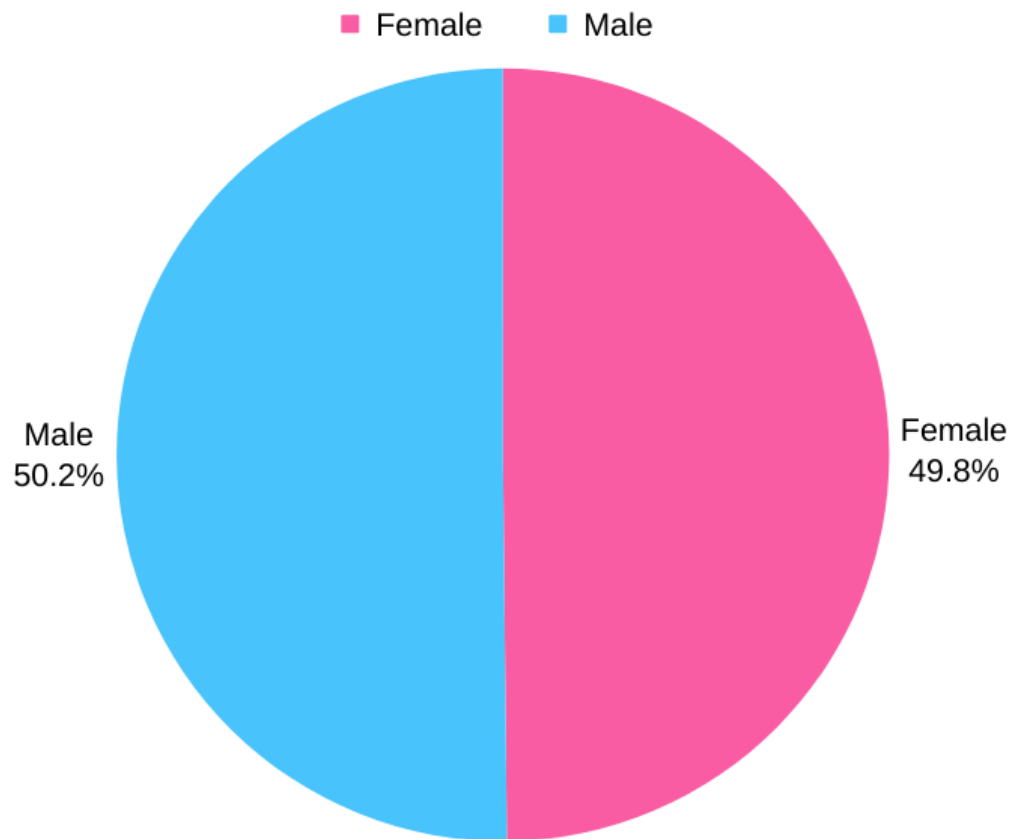
Данные о тарифе, статусе авторизации, типе пользователя, rfm и наличии домашнего интернета есть только у авторизованных в рамках текущего месяца пользователей

rfm - аббревиатура (Recency, Frequency, Monetary — давность, частота, деньги) — сегментация клиентов по лояльности

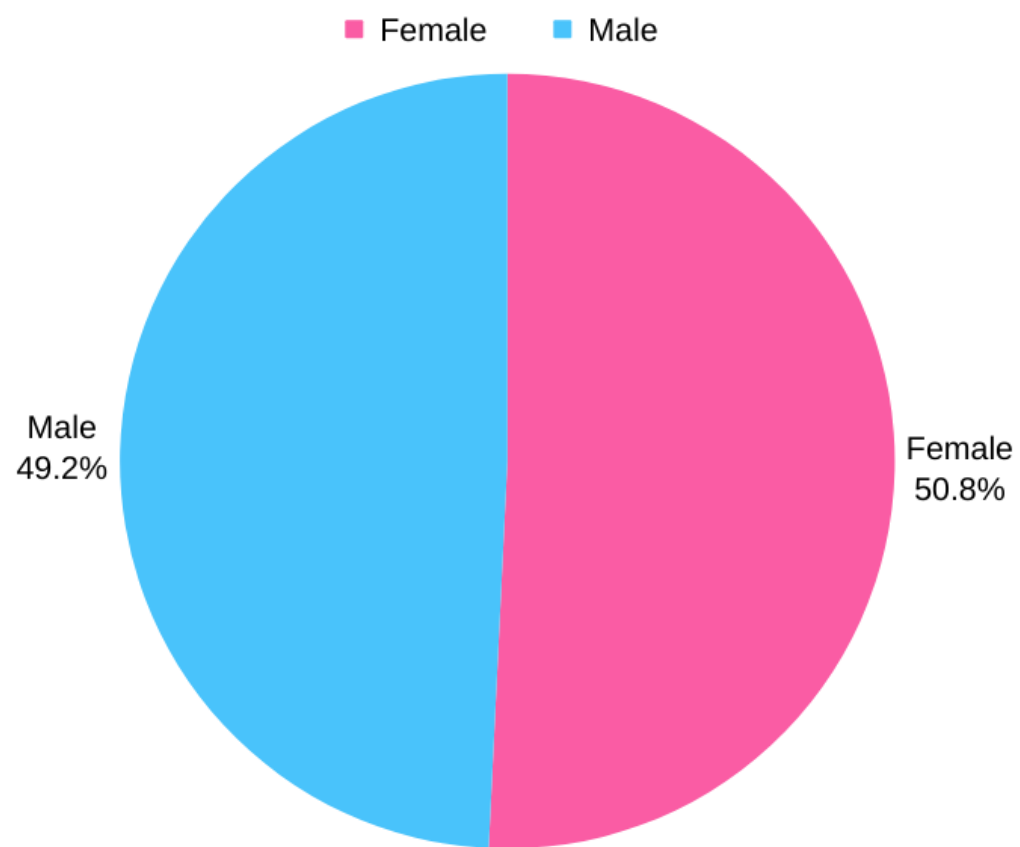
Преданализ

Исследование соотношения полов

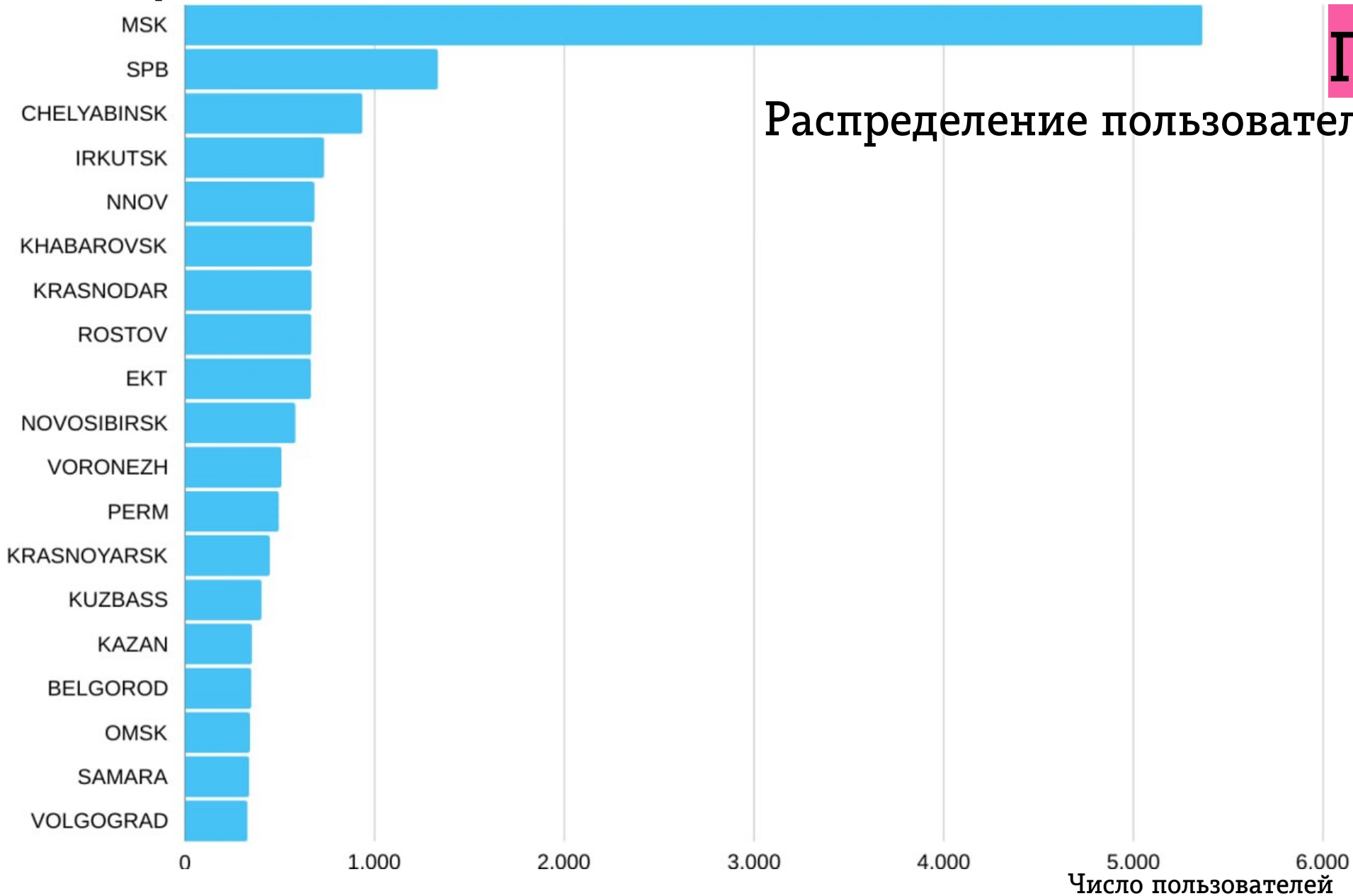
Распределение клиентов по полу



Распределение клиентов, совершивших покупку, по полу



Название региона

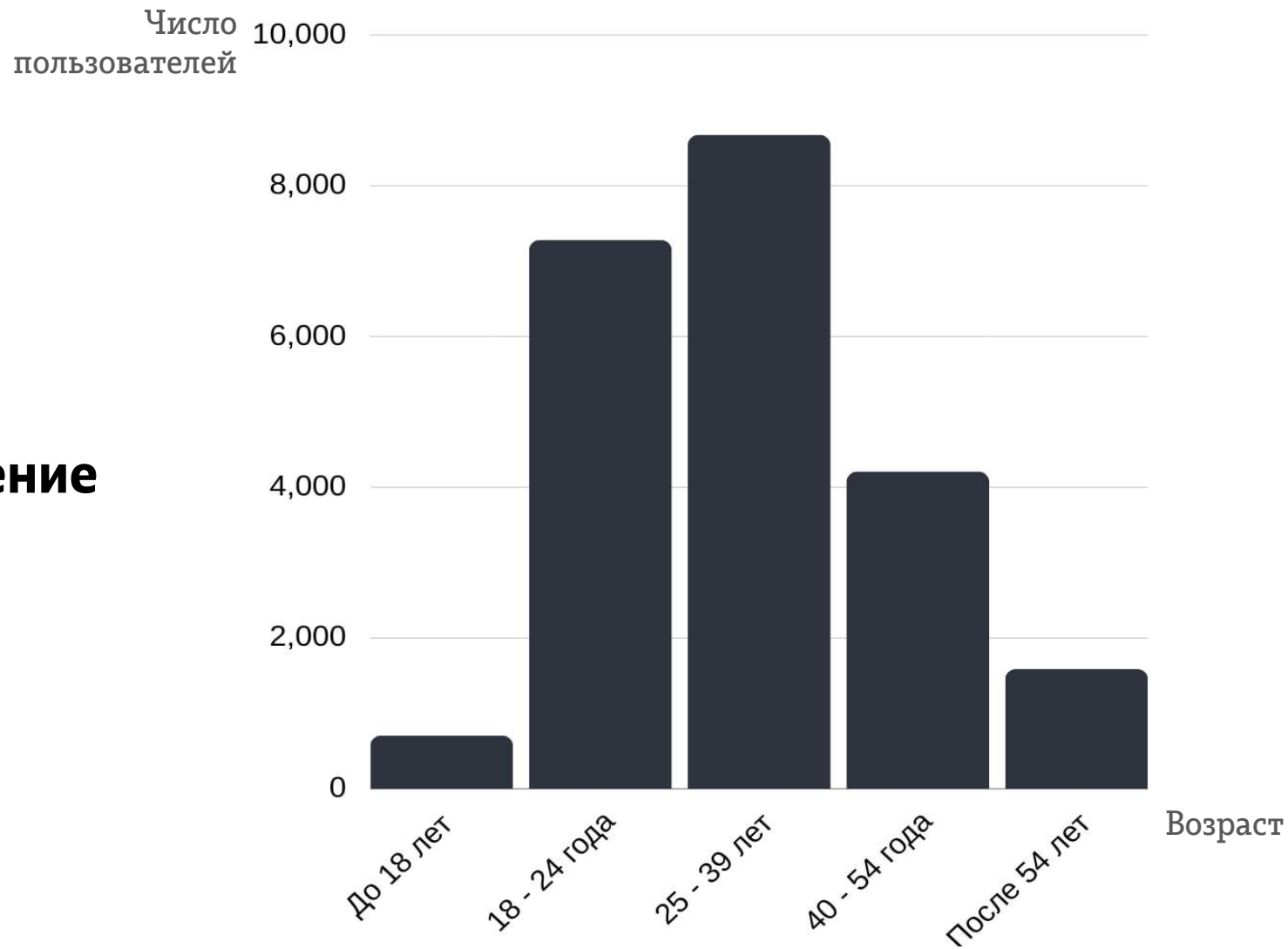


Преданализ

Остальные регионы
имеют число
пользователей меньше
300
(49 регионов)

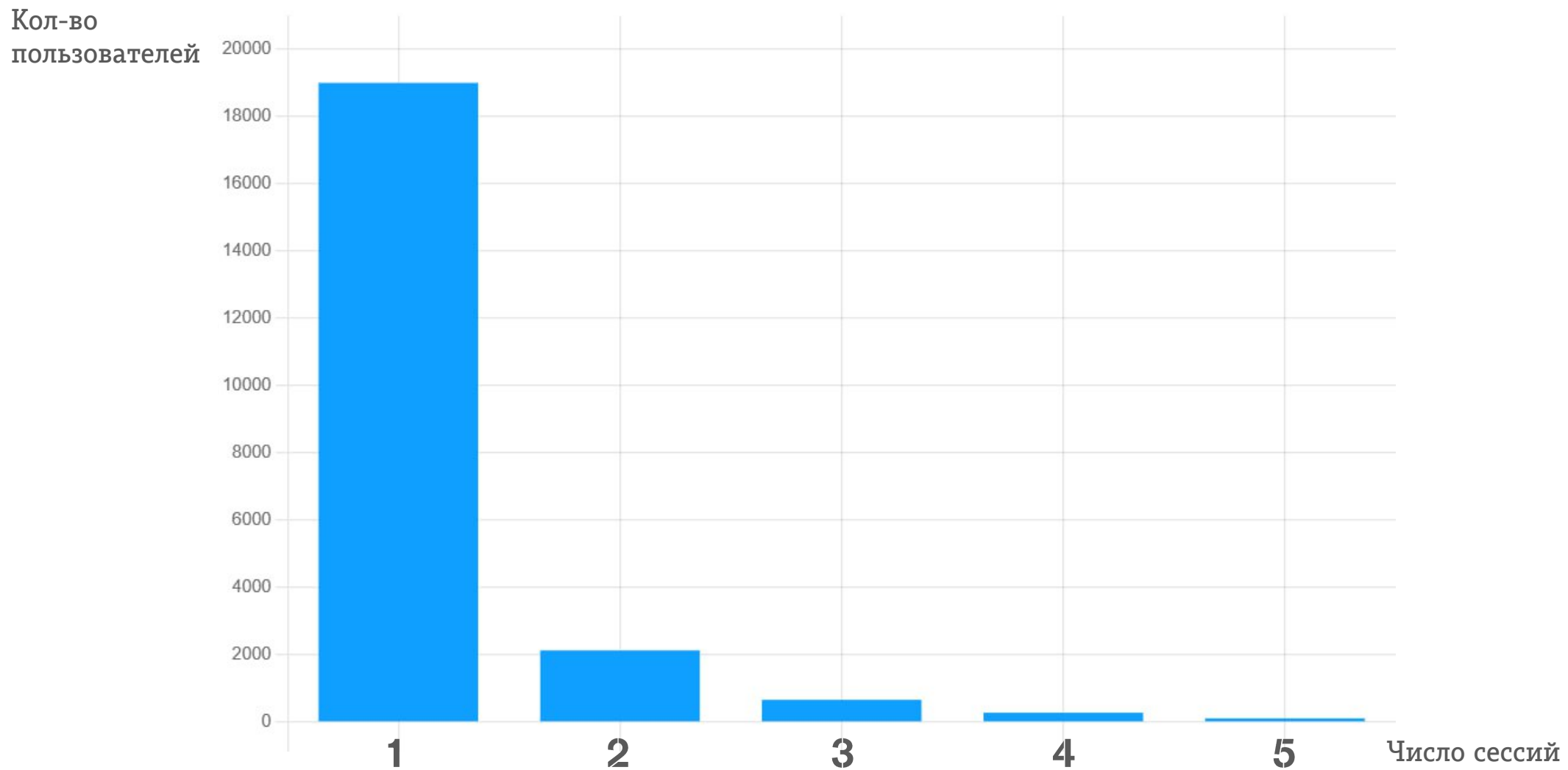
Преданализ

Возрастное распределение



Преданализ

Кол-во пользователей совершивших определенное число сессий



Исследовательский вопрос и гипотеза

Исследовательский вопрос:

Отличается ли конверсия на некоторых этапах покупки в зависимости от возраста, пола и уровня развития связи в регионе?

Гипотеза:

Если пользователь из региона с низким уровнем развития связи старше 40 лет (независимо от пола), то конверсия на некоторых этапах оформления покупки меньше.

Механизм

Люди старше сорока лет¹
из регионов с низким
уровнем развития связи²
имеют более низкий
уровень цифровой
грамотности.

Они путаются в
интерфейсе.

Сессии чаще
заканчиваются по
причине проблем
во взаимодействии
с интерфейсом.

Конверсия на этапах,
ведущих к покупке,
меньше.

¹ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7889415/>

² <https://issek.hse.ru/news/767681612.html/>

Подготовка данных

Обработка и чистка сессий от выбросов

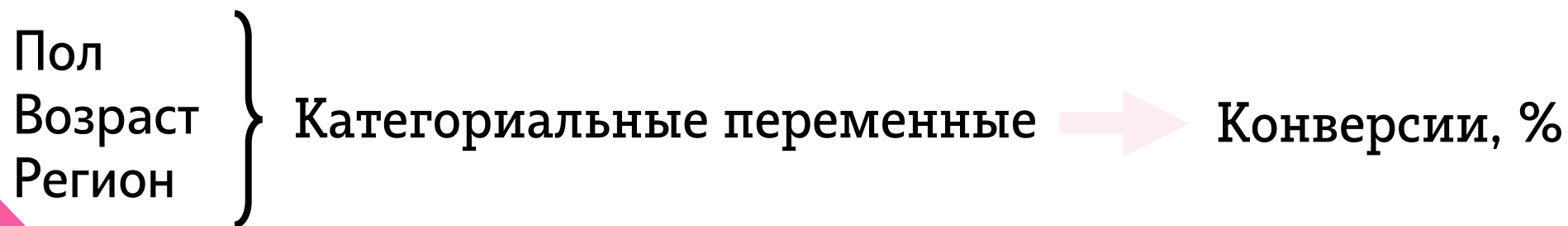
	Session_id	Client_id
✓	Session_1	Client_1
✗	Session_2	Client_2, Client_3

Исследуемые показатели

Объясняемая переменная:

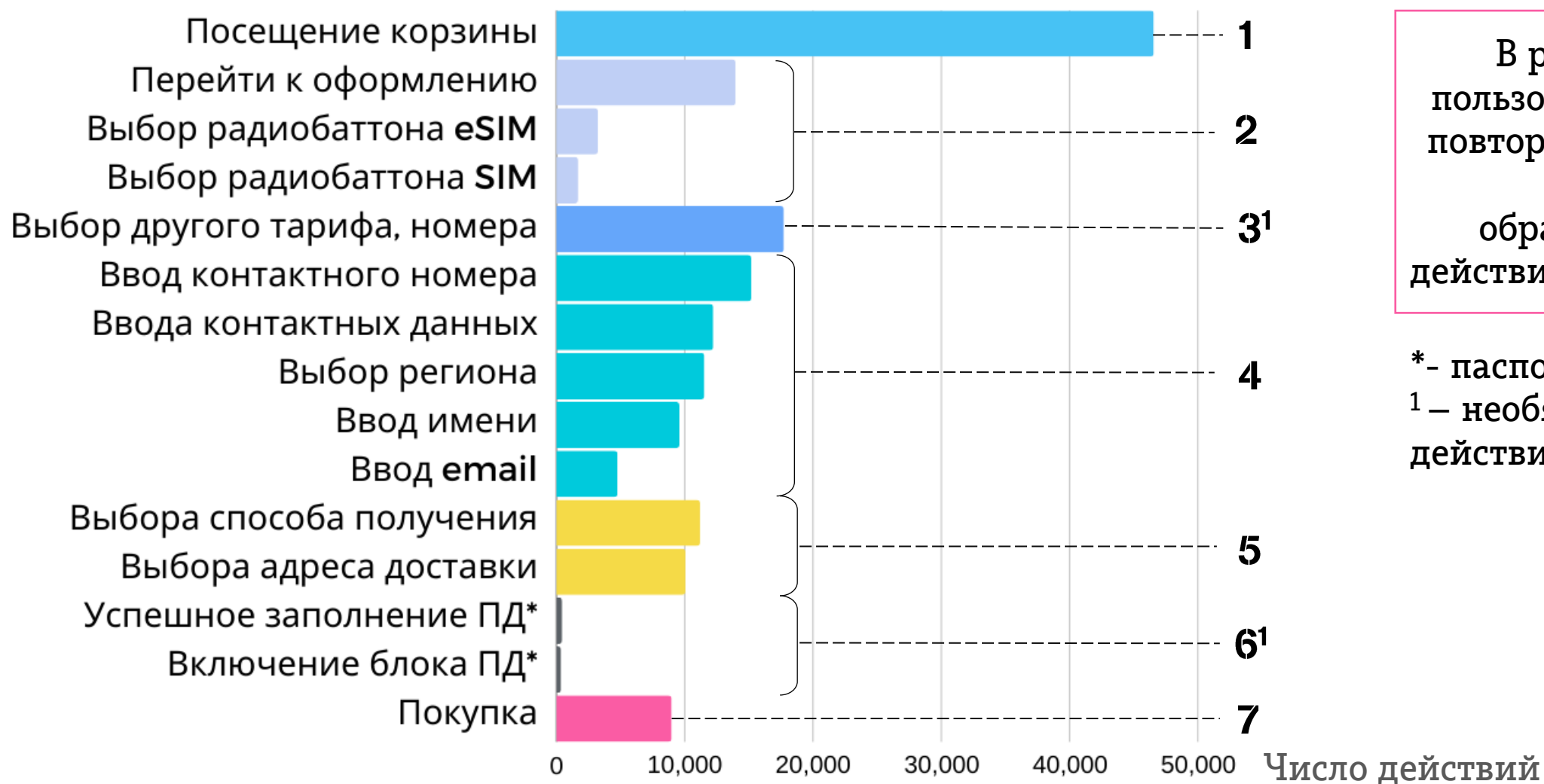
Конверсия - отношение числа посетителей сайта, выполнивших на нём текущее действие к числу пользователей, выполнивших предыдущее.

Объясняющие переменные:



Подготовка данных

Распределение частоты действий



В рамках сессии пользователь может повторять действия, из-за чего до обработки кол-во действий не линейно

*- паспортные данные
1 – необязательные действия

Подготовка данных

Группировка регионов

График телефонной плотности¹ по регионам за 2022 год **без учета**

Москвы и

Санкт-Петербурга

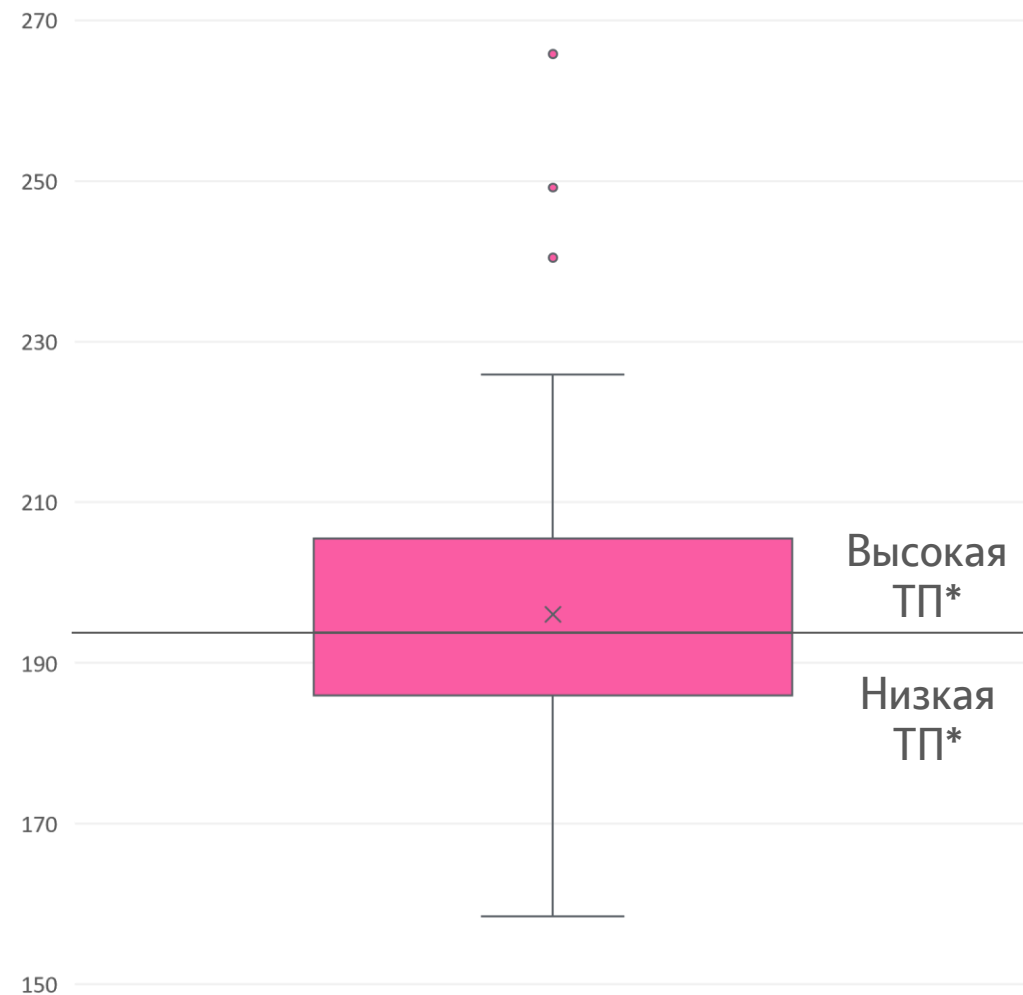
ТП - количество телефонных подключений активных устройств деленное на население

✔ **Москва и Санкт-Петербург выделены отдельно**

*ТП - аббревиатура (Телефонная плотность)

¹Источник Сведения о телефонной плотности (статистические данные)

ТП



Выборка

По итогу обработк данных мы имеем следующие выборки людей

Пол	Регион	Количество	
		<40	>40
Мужчины	СПб	515	168
	Москва	2035	645
	Регионы высокой ТП	4875	1778
	Регионы низкой ТП	934	321
Женщины	СПб	445	206
	Москва	1968	708
	Регионы высокой ТП	4876	1751
	Регионы низкой ТП	905	335
Генеральная		21774	

Математическое подтверждение

H_0 – конверсия группы по переходам с одного ивента на другой не отличается от аналогичной конверсии генеральной совокупности

H_1 – между конверсией генеральной совокупности и конверсией группы есть различия

5% принимаемый уровень
значимости

Математическое подтверждение

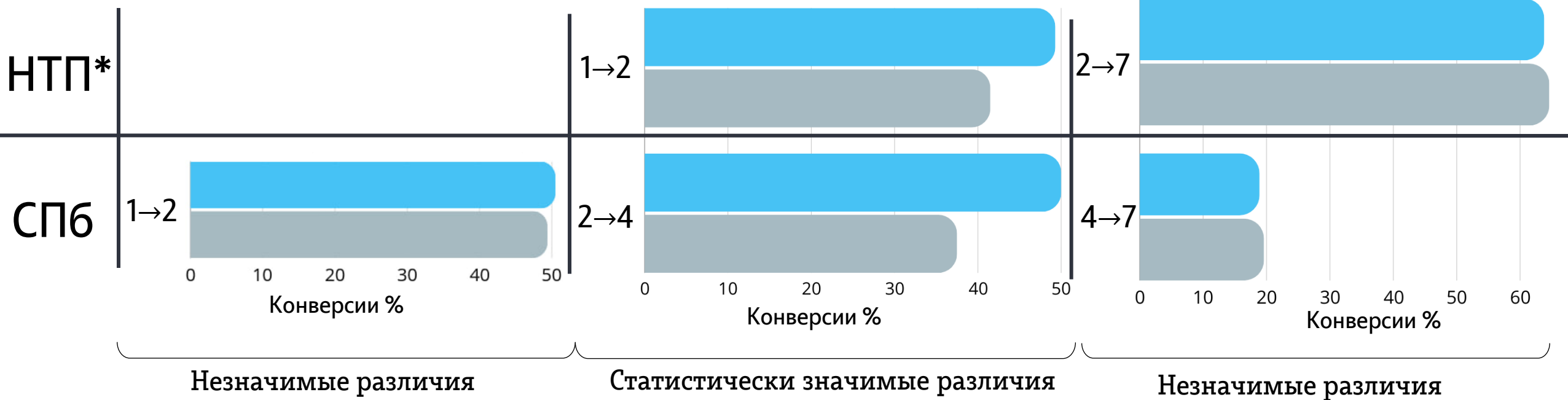
Для проверки $H_0: p = p_0$ воспользуемся $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$,

где \hat{p} – наблюдаемая конверсия группы, а p_0 – наблюдаемая конверсия в генеральной совокупности

Математическое подтверждение

Статически значимые различия подтвердились только у **мужчин**

● Генеральная выборка соответствующего действия ● Выборка старше 40



Интерпретация



Гипотеза частично подтверждена



Мужчины старше 40 лет из регионов с низким уровнем развития связи на данном этапе переходят реже, чем пользователи в целом

1→2



Мужчины старше 40 лет из Санкт-Петербурга на данном этапе переходят реже, чем пользователи в целом

2→4



Для женщин старше 40 таких различий нет



Конверсии на всех остальных этапах покупки у обеих групп **не имеют** статистических различий

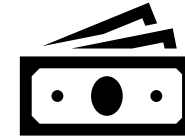
Мы можем предполагать, что при увеличении конверсий на указанных этапах увеличится конверсия в покупку

Policy implication

1. Помощь людям зрелого и пожилого возрастов, испытывающим проблемы с интерфейсом.



2. Увеличение конверсии на этапах, с которыми возникают сложности, увеличит конверсию в покупку.



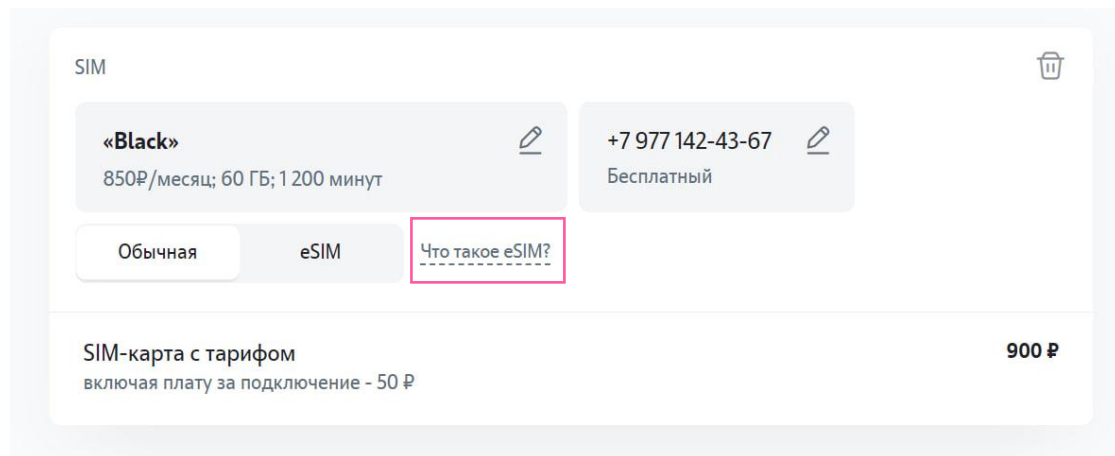
MVP расширения в chrome: подсказки для Tele-2



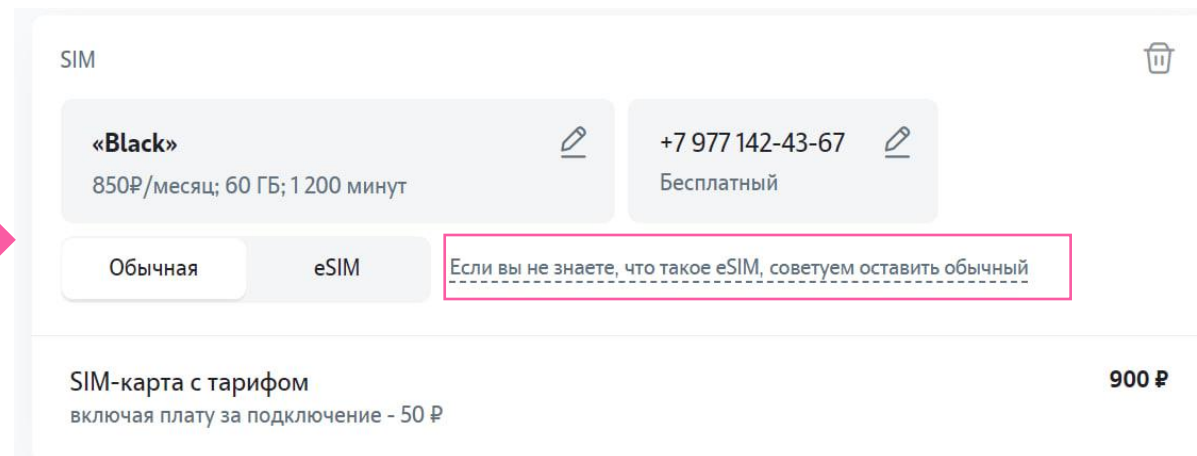
GitHub

MVP расширения в chrome: подсказки для Tele-2

Обычный вид:



Для выявленных групп:



Для выявленных групп с низкой конверсией расширение будет менять детали интерфейса на более простые

Ограничения

Отсутствуют данные о том, как пользователи попали в корзину

Не дана информация о том, как точно интерпретируется метрика RFM*

Присутствуют данные только за один месяц

Нельзя использовать информацию о тарифах

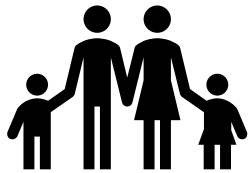
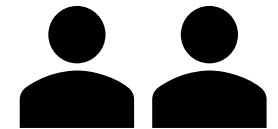
rfm - аббревиатура (Recency, Frequency, Monetary — давность, частота, деньги) — сегментация клиентов по лояльности

Перспективы



Применить к исследованию значения
device

Провести A/B тест расширения



Получить сведения о детях
пользователей

Кожевников Сергей
Аналитик, статистик



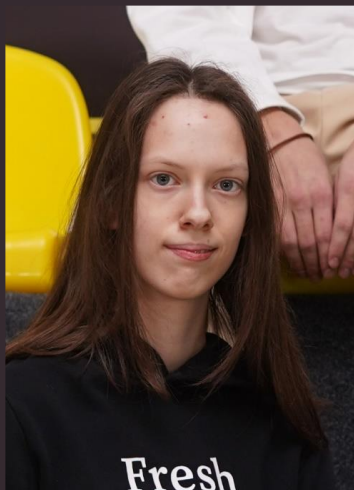
Илюхин Семён
Программист



Программист
Мазитова Нелли



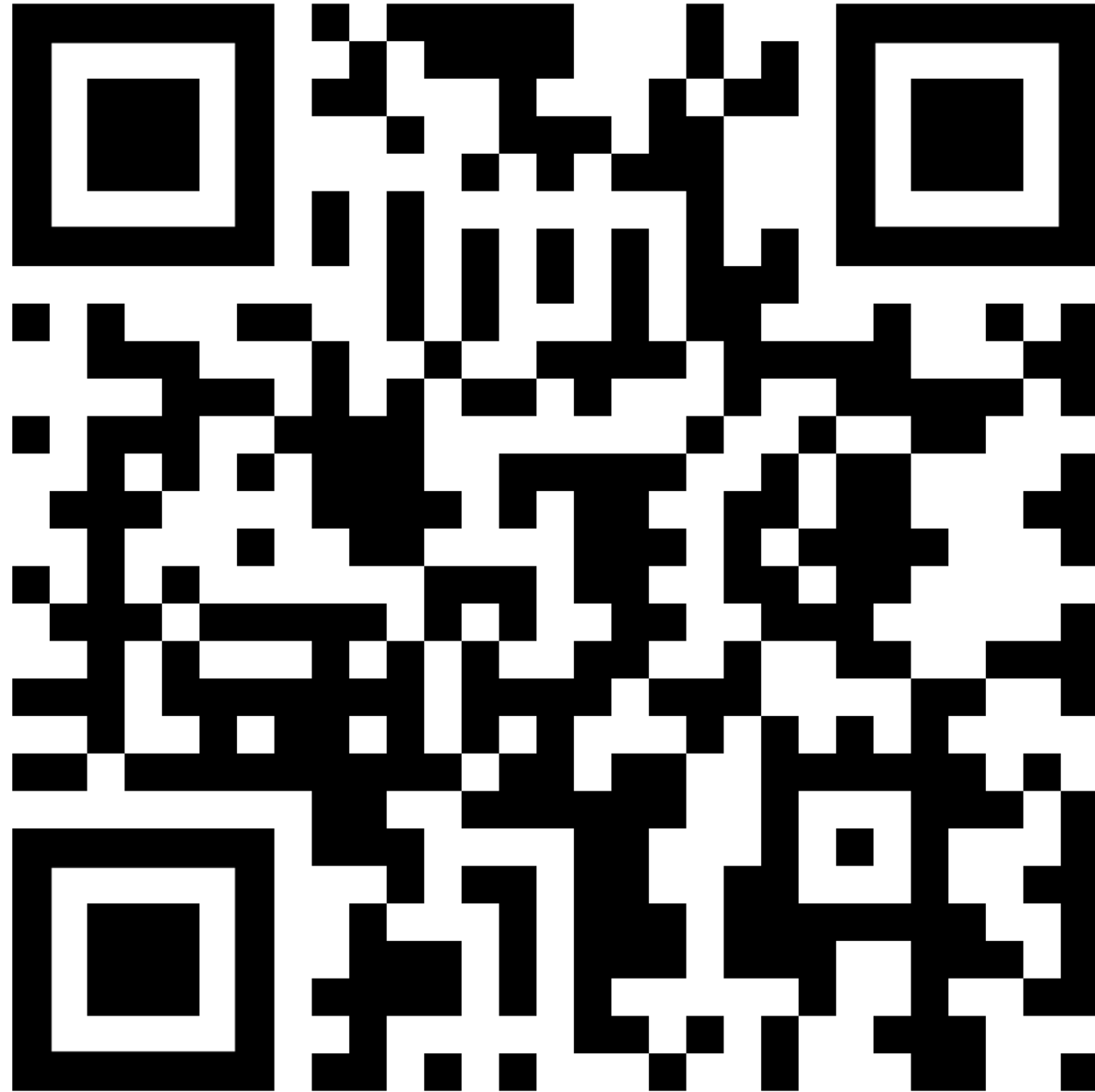
Фёдорова Мария
Программист



Коровина Ульян
Тимлид, статистик



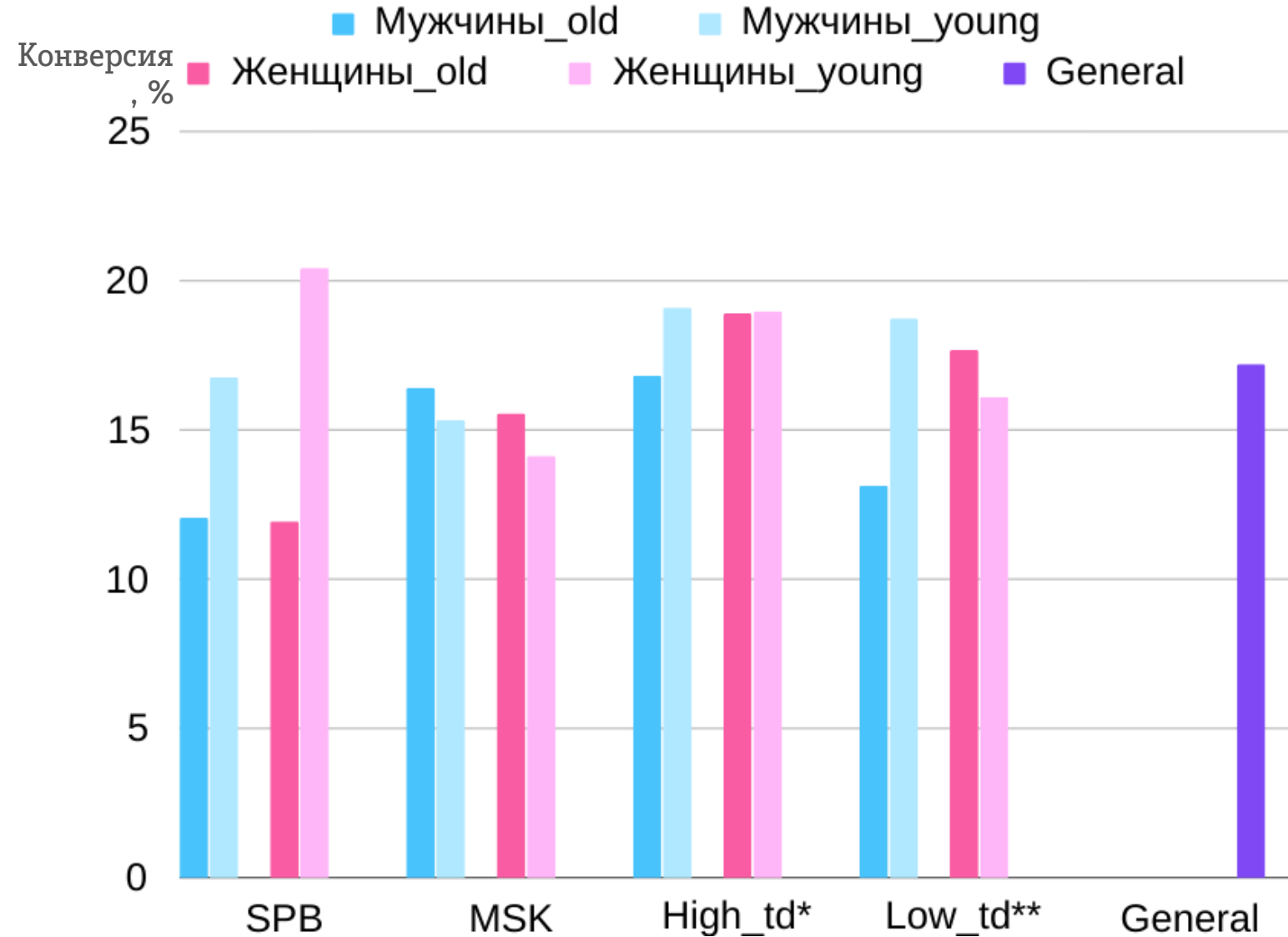
**Ссылка на
гитхаб**



ОТВЕТЫ НА ВОПРОСЫ

Приложение

Распределение процента конверсии покупки в зависимости от группы



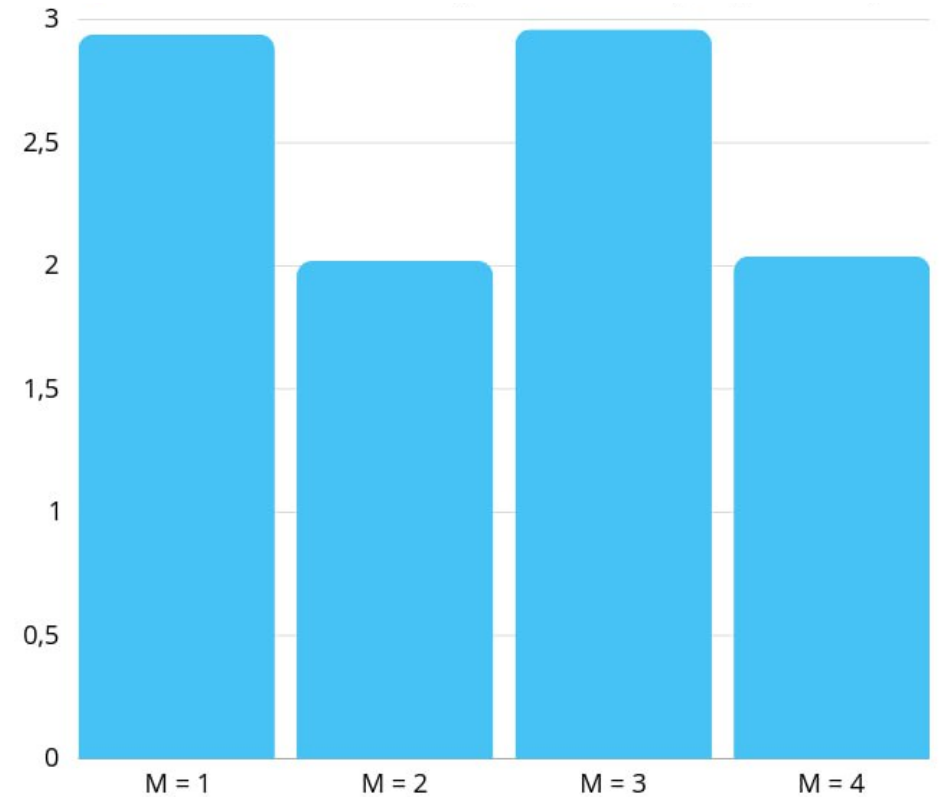
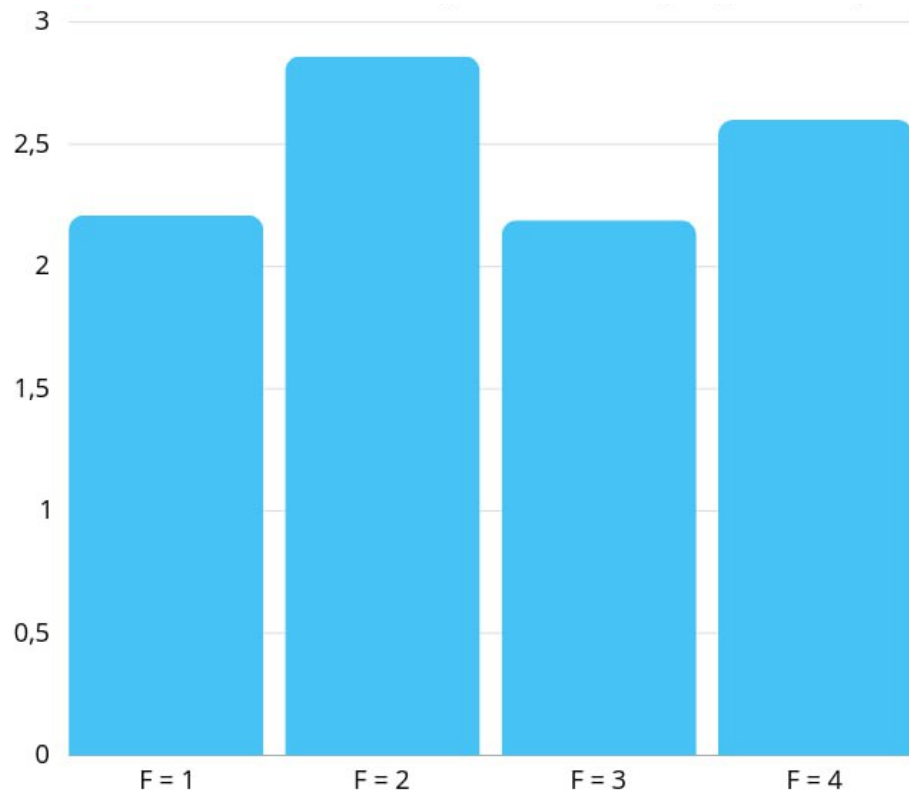
*High_td – Регионы с высокой телефонной плотностью

**Low_td – Регионы с низкой телефонной плотностью

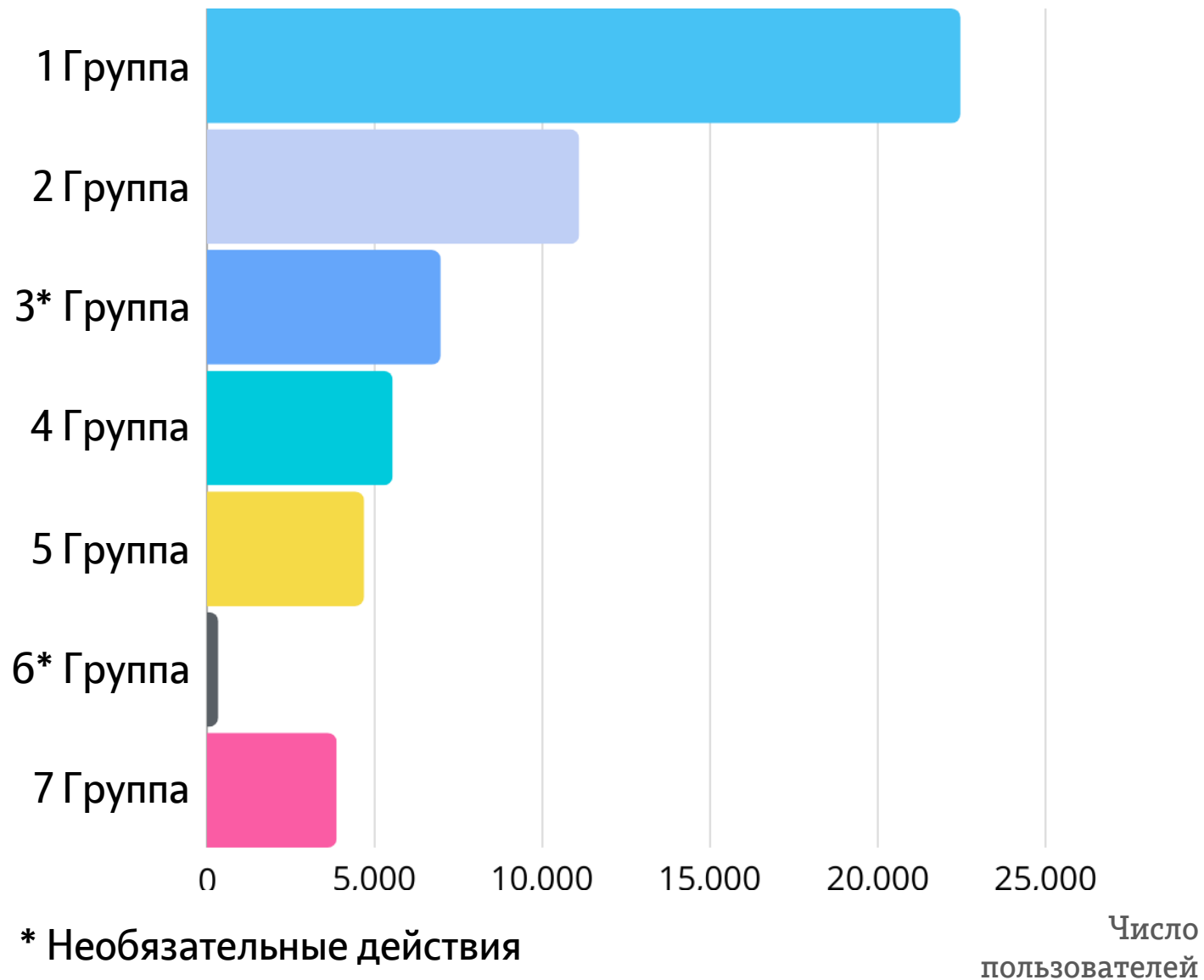
Old > 40, Young < 40

Вывод зависимости параметров RFM

Отношение количества покупок к количеству людей в выборке



Приложение



Распределение количества действий по группам после их чистки и группировки для каждого пользователя в рамках всех сессий

Переименовать сессии

Удалить повторы действий

Соединить сессии каждого пользователя

Подготовка данных
Группировка действий

Поделить сессий на блоки до покупки

Заполнить обязательные действия

Вычисление конверсий