



ТИНЬКОФФ



Что влияет на активность пользователей в экосистеме Тинькофф?



**Почему этот
исследовательский
вопрос?**



Повышение доходов



**Анализ поведения
потребителей**



Наша гипотеза - ...

Существует «идеальный» портрет пользователя «Тинькофф Города», к которому можно приблизить любого другого пользователя



Механизм работы гипотезы

Неоднородность потребительской аудитории

Потребители различаются по:

- Региону проживания
- Полу
- Возрасту
- Категории трат

Возможность перехода пользователя к другому блоку

Пользователя можно мотивировать на траты в определённой категории для приближения его к портрету «идеального» клиента



Переменные

user_id	Уникальный идентификатор клиента
region	Актуальный регион
bundle	Наличие подписки (No bundle = нет подписки, Tinkoff Pro, Tinkoff Premium - соответствующие подписки)
num_sessions	Количество сессий* у пользователя за 2022 год
avg_session	Средняя длина сессии* (в секундах)
fuel, cinema, food, travel, restaurants, concerts, games, books, theater, sports, beauty, flowers, technology	Бинарные переменные, описывающие категорию заказа Например, Топливо, Красота и т.д.
price	Стоимость заказа

* Сессия - это время от открытия приложения до его закрытия, включает в себя все действия совершенные в мобильном банке



Предобработка

1

- Поиск и удаление дубликатов в `user_id`
- Поиск и удаление строк с пустыми значениями

2

- Выбор необходимых переменных в соответствии с гипотезой

3

- Составление матрицы корреляций (без `user_id`)



Работа над гипотезой

- Подсчёт количества повторений каждого `user_id` в оригинальной data base
- Вывод в **отдельный** столбец `count` в таблицу с уникальными `user_id`
- Подсчёт количества категорий у каждого `user_id`
- Вывод в **отдельный** столбец в таблицу с уникальными `user_id`
- Удаление значений в `count` (столбец числа заказов), равных **1** (явно не активные пользователи)
- Построение точечного графика на основе количества заказов
- Выделение диапазона перелома с наименьшим значением, равным **10**
- Разделение пользователей на группы по количеству заказов (**больше или равно 10** и **меньше 10**)
- Вычисление медианы по числу заказов (**4,2**)
- Отбрасывание всех значений в столбце числа заказов, меньших округлённой вниз медианы



Работа над гипотезой

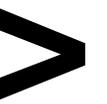
- Распределение пользователей по количеству различных категорий от **1 до 5**
- Определение количества пользователей в каждой категории (**1** категория – больше половины пользователей)
- Определение границы между блоком большого количества категорий и малого на **2** (так как при **3** остаётся слишком маленькая выборка)
- Выделение основных **4 блоков** (портретов клиентов)



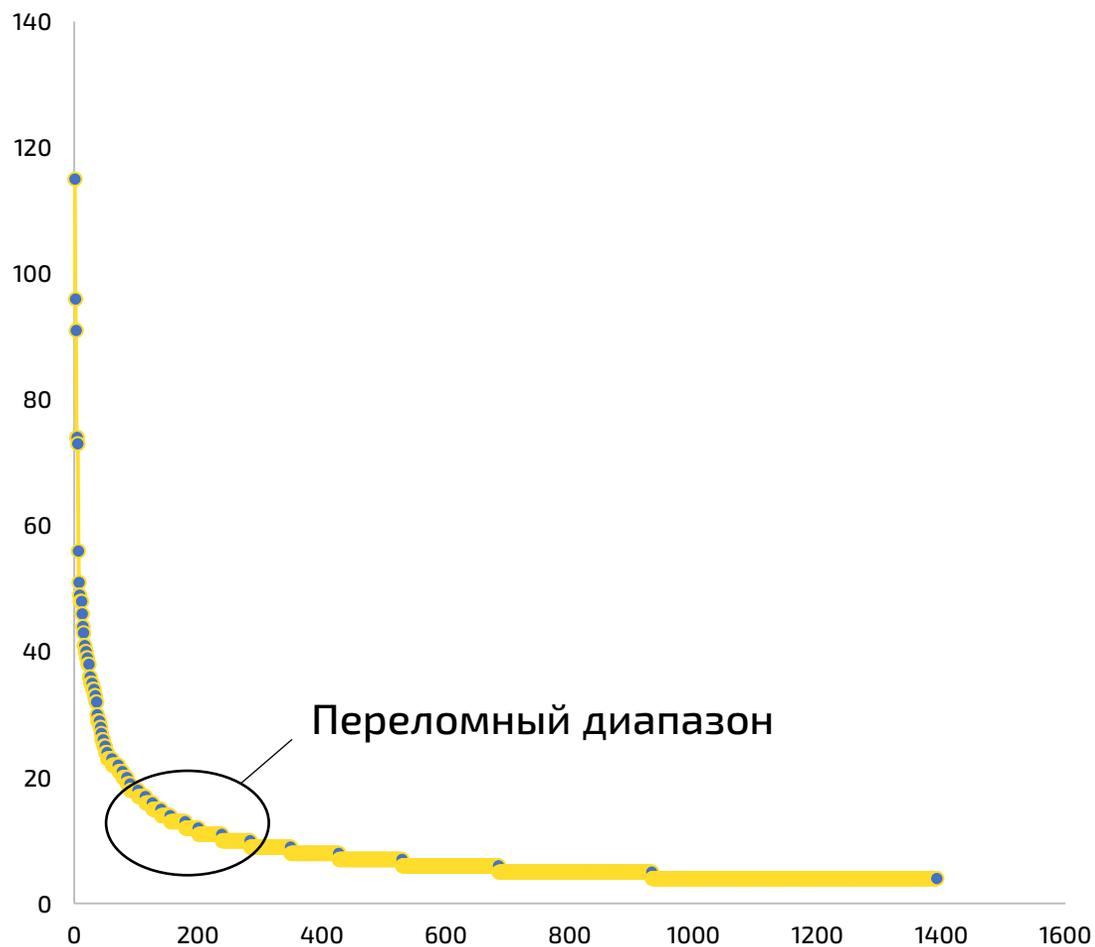
Визуальные данные

Количество категорий

		Б	М	
Количество заказов	Б	Эталон		115
	М			≥ 4
				< 4



Визуальные данные



Распределение количества операций

Duplicate rows

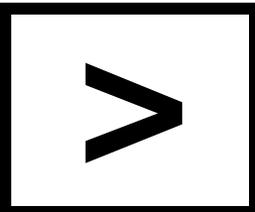
Most frequently occurring

	user_id	sex	age	region	bundle	num_sessions	avg_session	service	order_date	pric
79	19221	M	16	Санкт-Петербург	No bundle	266.0	78.0	Игры	28.04.2022	4
481	82385	M	17	Челябинская область	Tinkoff Premium	1531.0	184.0	Путешествия	26.05.2022	384
552	93619	M	15	Челябинская область	No bundle	350.0	164.0	Путешествия	22.10.2022	388
223	45050	F	17	Санкт-Петербург	No bundle	269.0	68.0	Путешествия	30.10.2022	184
341	74120	F	16	Новосибирская область	Tinkoff Premium	79.0	334.0	Путешествия	16.08.2022	140
453	82385	M	17	Челябинская область	Tinkoff Premium	1531.0	184.0	Путешествия	18.04.2022	480
484	82385	M	17	Челябинская область	Tinkoff Premium	1531.0	184.0	Путешествия	26.05.2022	404
500	82385	M	17	Челябинская область	Tinkoff Premium	1531.0	184.0	Путешествия	28.07.2022	621
504	82385	M	17	Челябинская область	Tinkoff Premium	1531.0	184.0	Путешествия	29.07.2022	641
21	4232	F	15	Московская область	No bundle	472.0	116.0	Кино	13.07.2022	38

Комплект дубликатов по id пользователя



Выводы



Устойчивость результатов

Данное исследование актуально исключительно для регионов Российской Федерации



Какие перспективы?



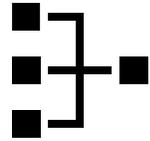
Повышение количества заказов и используемых категорий



Подбор индивидуальных предложений каждому пользователю



Ограничения

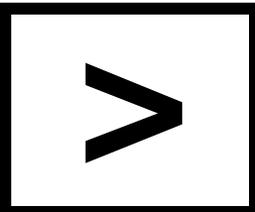


Неполнота данных

Например, отсутствие информации об акциях в тот период времени



Ограниченность возрастной категории



**Спасибо за
внимание**

!

Наша команда

Рудольф

Что сделал

Арина

Что сделала

Даня

Что сделал

Паша

Что сделал

Родион

Что сделал