

Тестовое задание для вакансии Аналитик данных LOYMAX AI

Эльвира Шараева 19.06.2025

Анализ NWOE и NIV для ATE

Цель: Оценить средний эффект воздействия по признакам из A/B тестов рекламной платформы Criteo

Исходные данные

Для исследования были предоставлены данные рекламной платформы Criteo в виде таблицы порядка 13,9 млн записей. Наименования 16 колонок: 12 анонимных признаков(f0-f11), treatment- воздействие, conversion - конверсия, visit - посещение, exposure - показ. Пропущенных значений в таблице не обнаружено, очистка данных не потребовалась

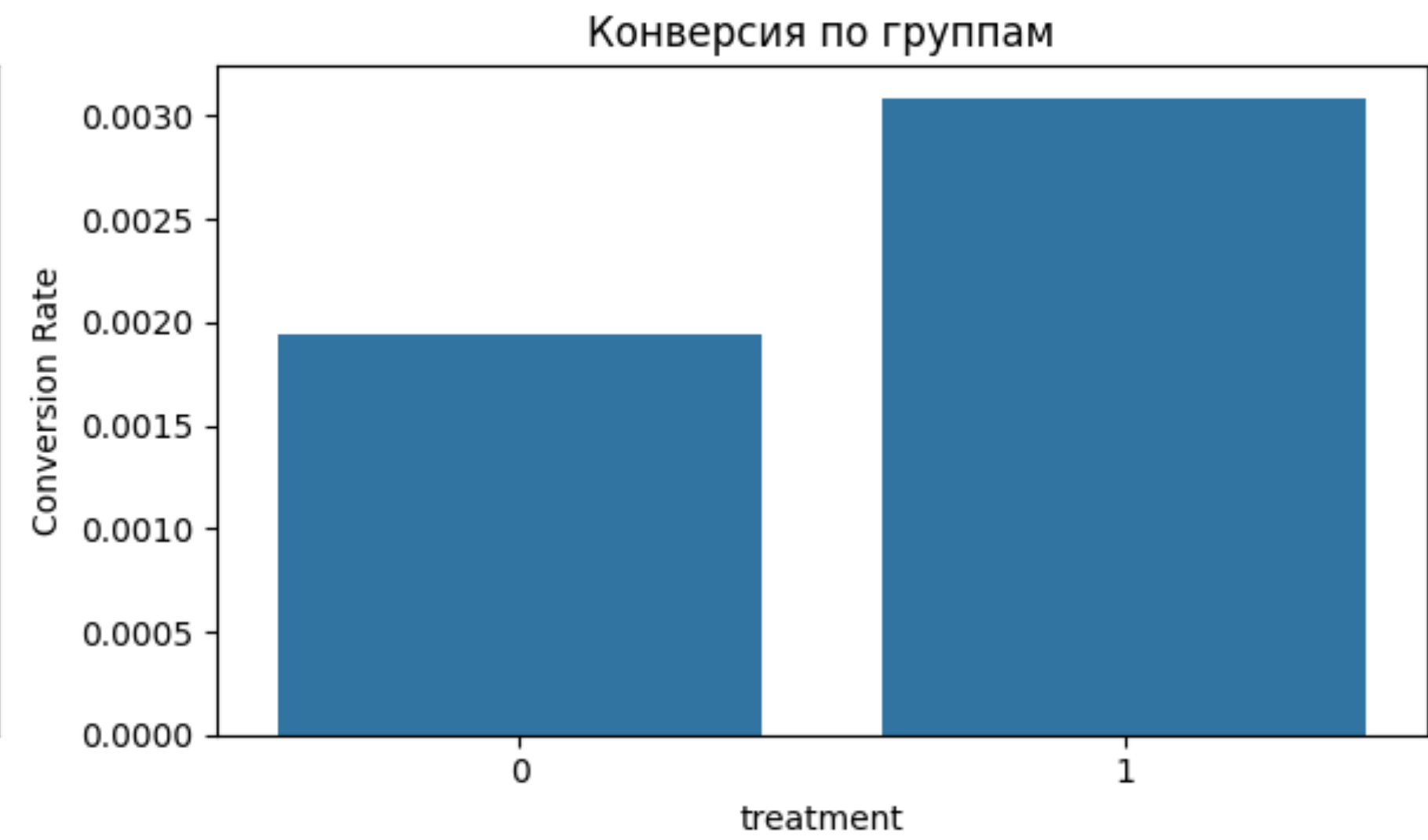
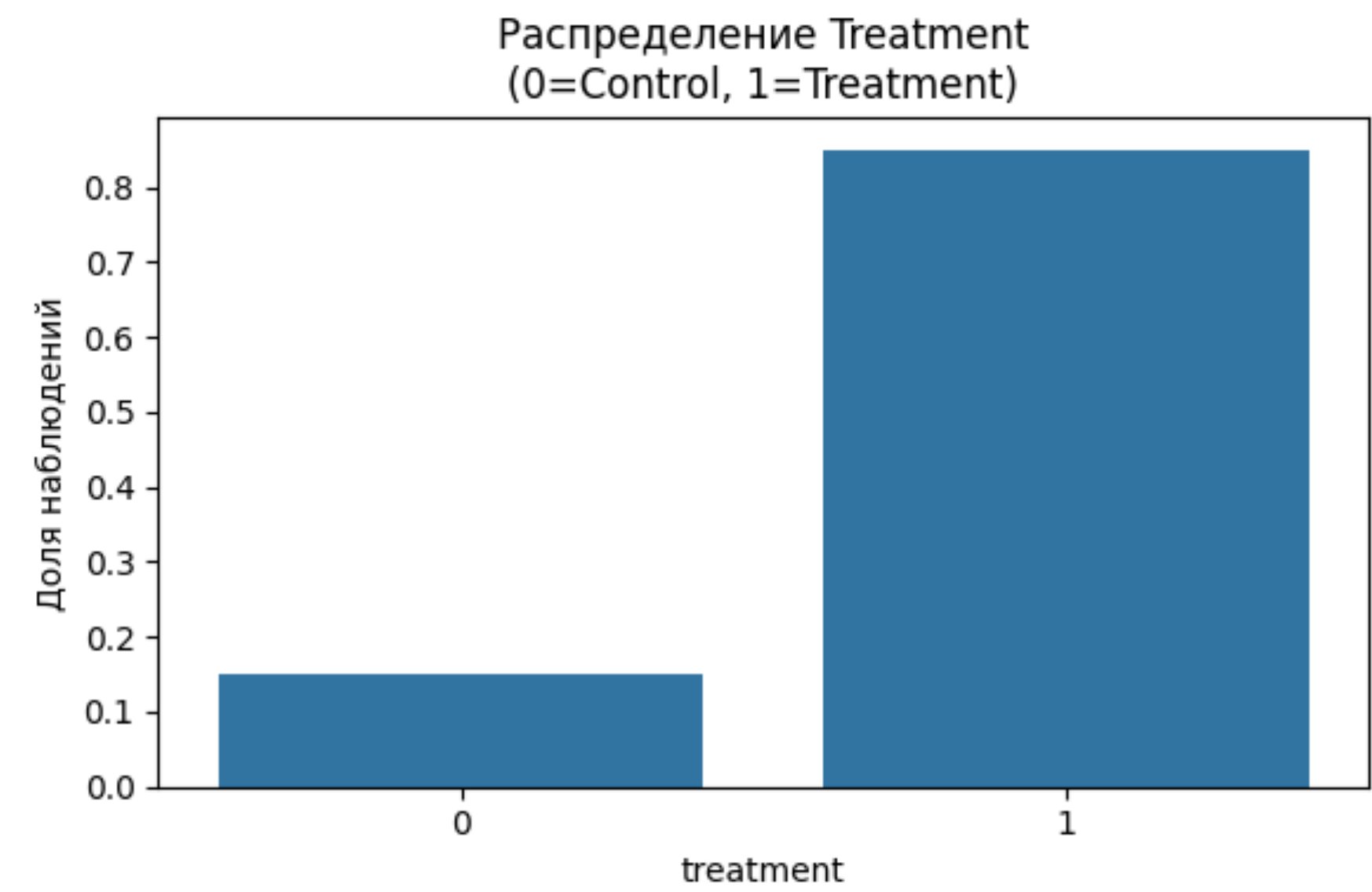
Сравнение конверсии в посещение для тестовой и контрольной группы

Распределение по группам:

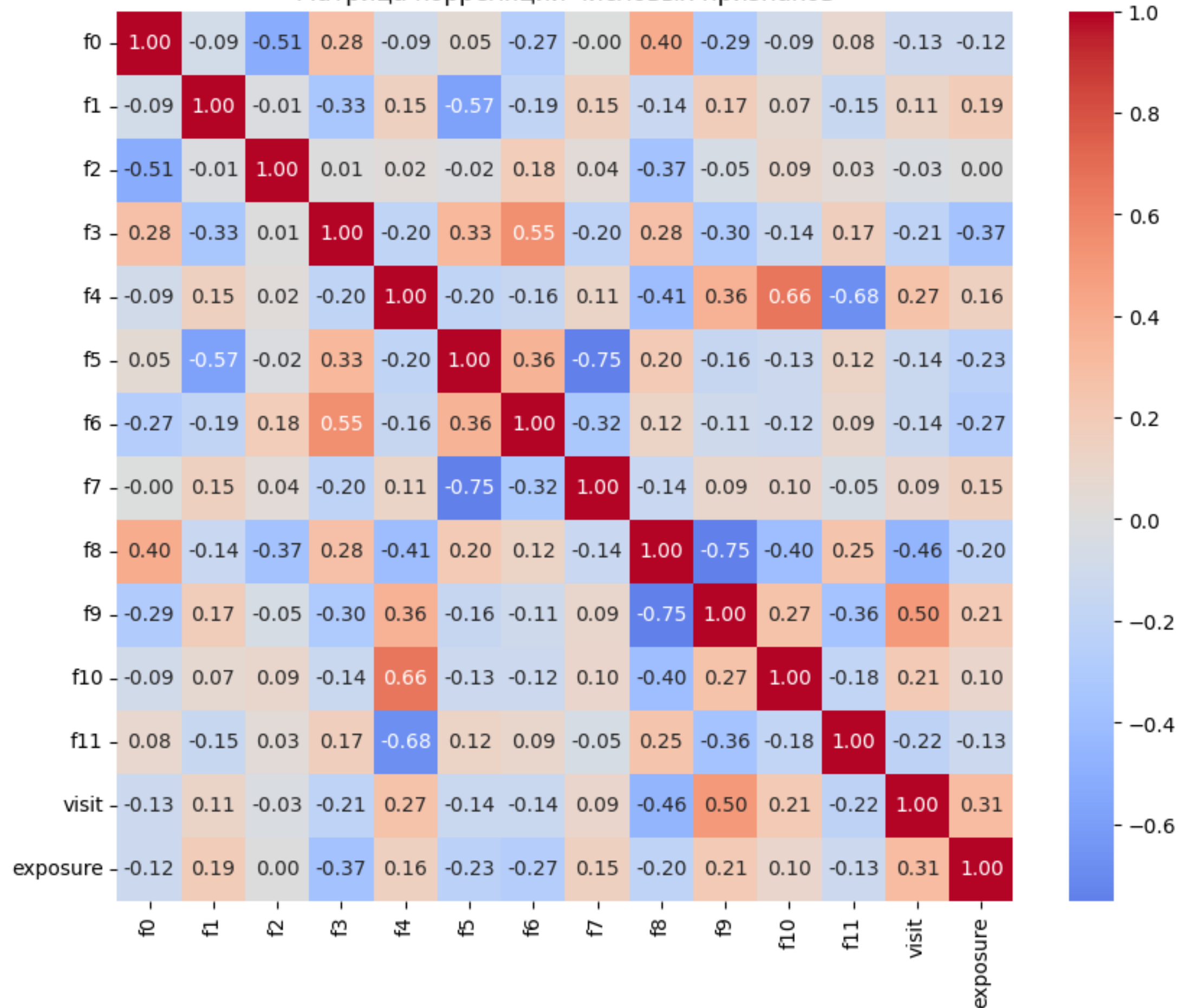
proportion	
treatment	
1	0.85
0	0.15

Конверсия по группам:

conversion	
treatment	
0	0.001938
1	0.003089

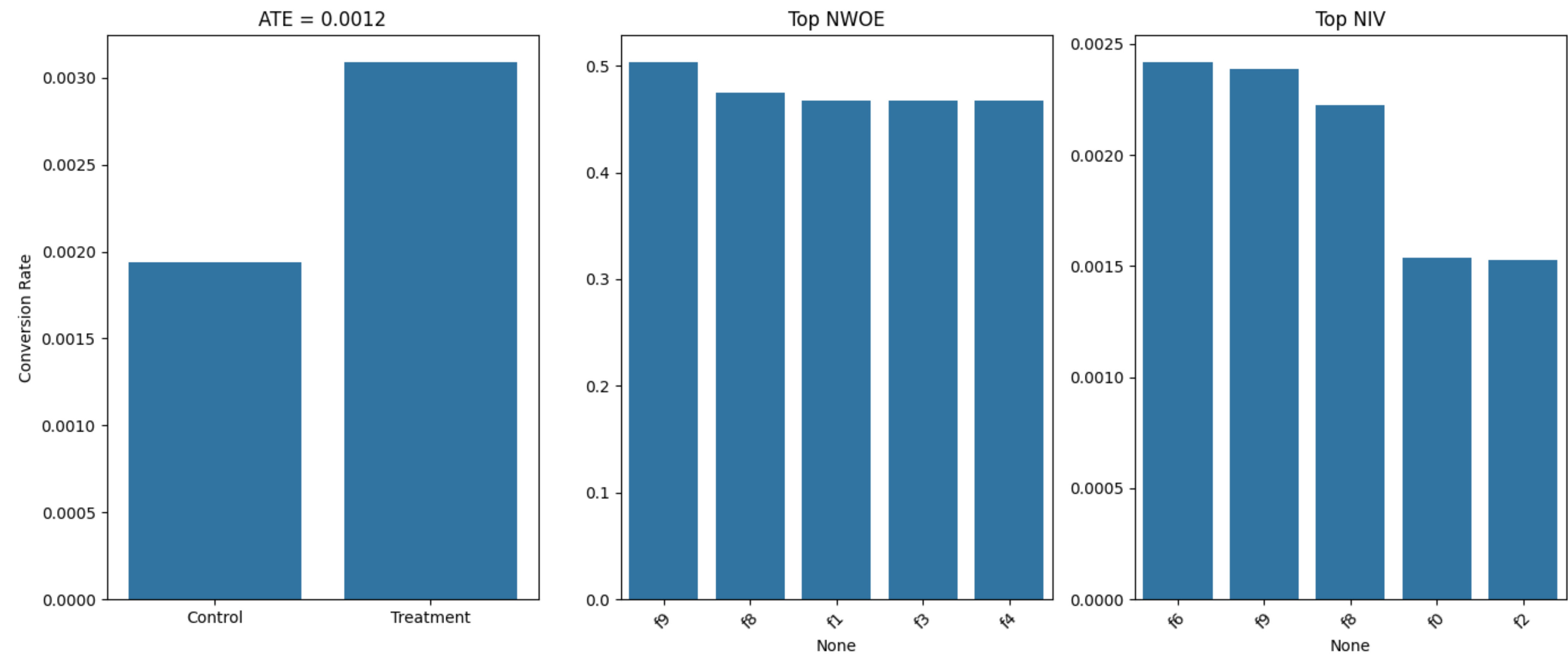


Матрица корреляций числовых признаков



- Высокую линейную корреляцию показали признаки f4 и f10 (КК 0,66) обратную корреляцию f11 и f4(КК -0,68), f8 и f9(КК - 0.75) , f5 и f7 (КК - 0.75)

ATE для тестовой и контрольной групп, NWOE и NIV для анонимных характеристик(топ 5)



Выводы

1. $ATE = 0.0012$: Общий средний эффект воздействия крайне слабый (почти нулевой).
В среднем воздействие практически не повлияло на конверсию.

2. Топ-3 признаков по $NWOE$: $f9: 0.50$, $f8: 0.47$, $f1: 0.47$

Все три признака показывают сильный положительный эффект ($NWOE > 0.3$) для определенных подгрупп.

Несмотря на сильные эффекты в подгруппах, общий ATE близок к нулю. Это означает, что:
Положительные эффекты в одних подгруппах компенсируются отрицательными в других (которые не попали в топ).

Эти подгруппы слишком малы, чтобы повлиять на общий ATE .

3. Топ-3 признаков по NIV^{**} (информативность для сегментации эффекта): $f6: 0.0025$, $f9: 0.0024$, $f8: 0.0022$

Все три признака не информативны ($NIV \approx 0$), включая те, что показали высокий $NWOE$.

Вывод

1. Сильные эффекты ($NWOE$) наблюдаются в очень маленьких подгруппах (не влияют на общую информативность).

2. Возможно, биннинг (разбиение на группы) был неудачным — например, бины содержат нерепрезентативные или шумовые данные.

Анализ изменения поведения клиентов

Цель: Проанализировать влияние маркетинговой кампании на состав покупок, расширение товарной корзины, покупку новых товаров и выявить чувствительность клиентов до, во время и после кампании.

Исходные данные

Данные:

Person.csv – характеристики клиентов

Goods.csv – каталог товаров с категориями

List_Purchases.parquet – строки покупок

List_Cheques.parquet – чеки

Action.csv – история кампаний

Поля включают: Person_BKEY, Purchase_ID, Operation_datetime, баллы (toCard_, fromCard_), Amount_Cheque, Merchant_id, Goods_BKEY, Qnt, Amount, Amount_Original

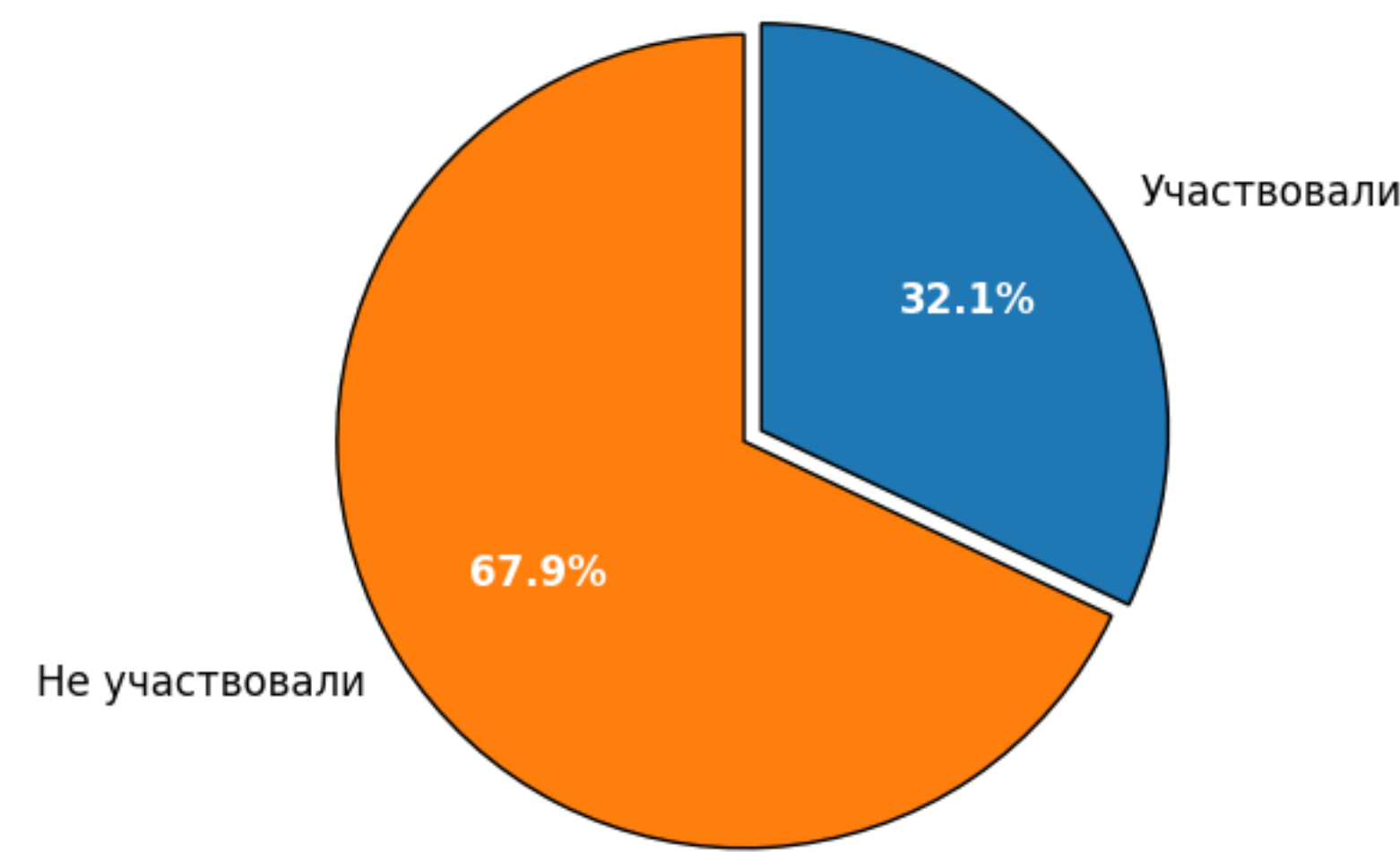
Этапы анализа

1. Сегментация по периодам: до/во время/после акции
2. Создание признаков доли новых товаров и категорий в покупках до и во время акции
3. Сравнительный анализ: чеки, сумма, товарные доли и доли категорий каталога
4. Определение чувствительных клиентов: создание и выделение признаков и бинов с максимальным влиянием на изменение покупательского поведения в разрезе количества и состава товаров и категорий, покупки новых товаров и категорий
5. Визуализация: временные графики, диаграммы категорий
6. Выводы + бизнес-рекомендации

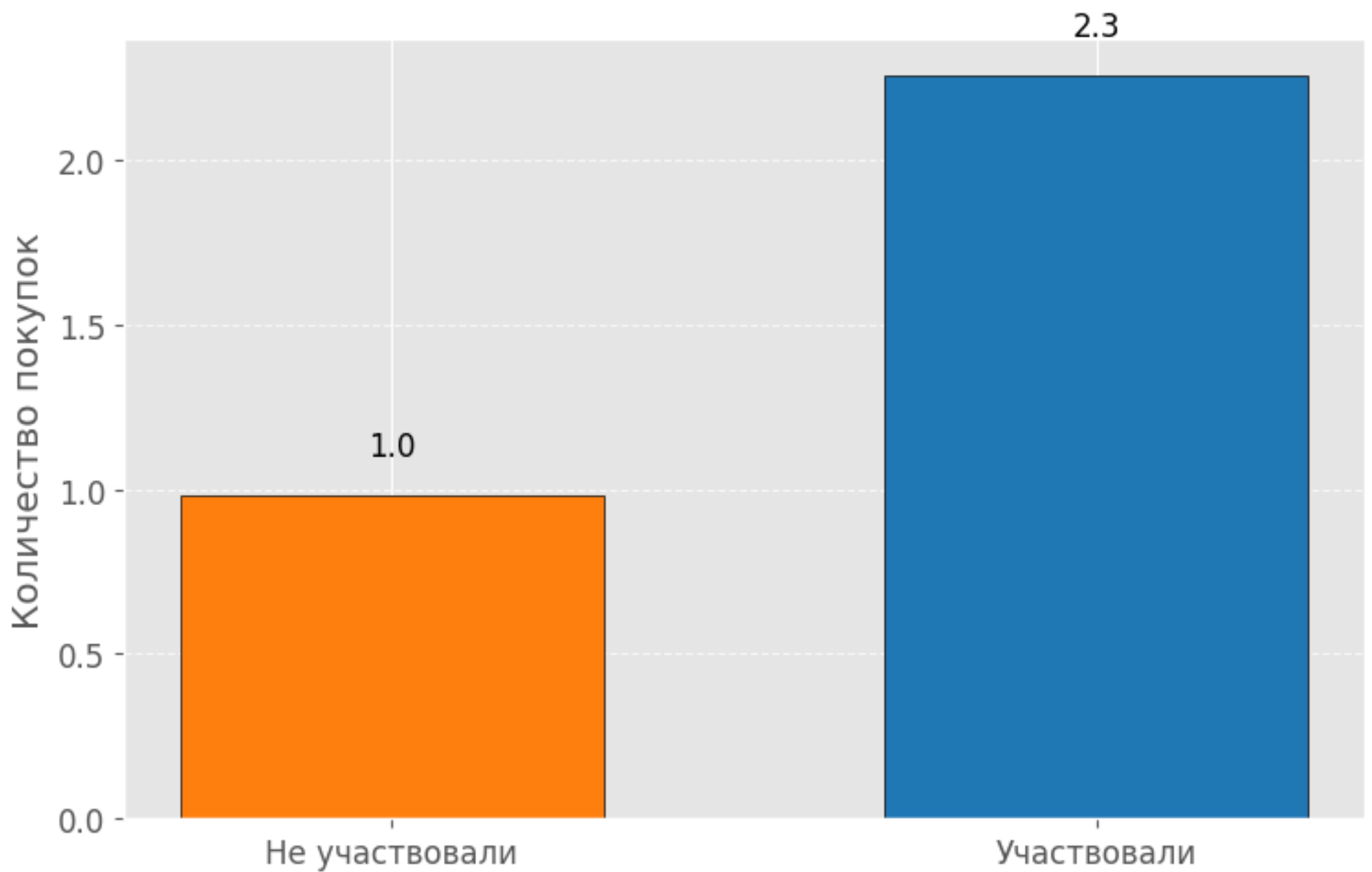
Покупатели, принимавшие участие в акции 21.03.2019* г.

*дата проведения акции определена по данным массива Actions.csv

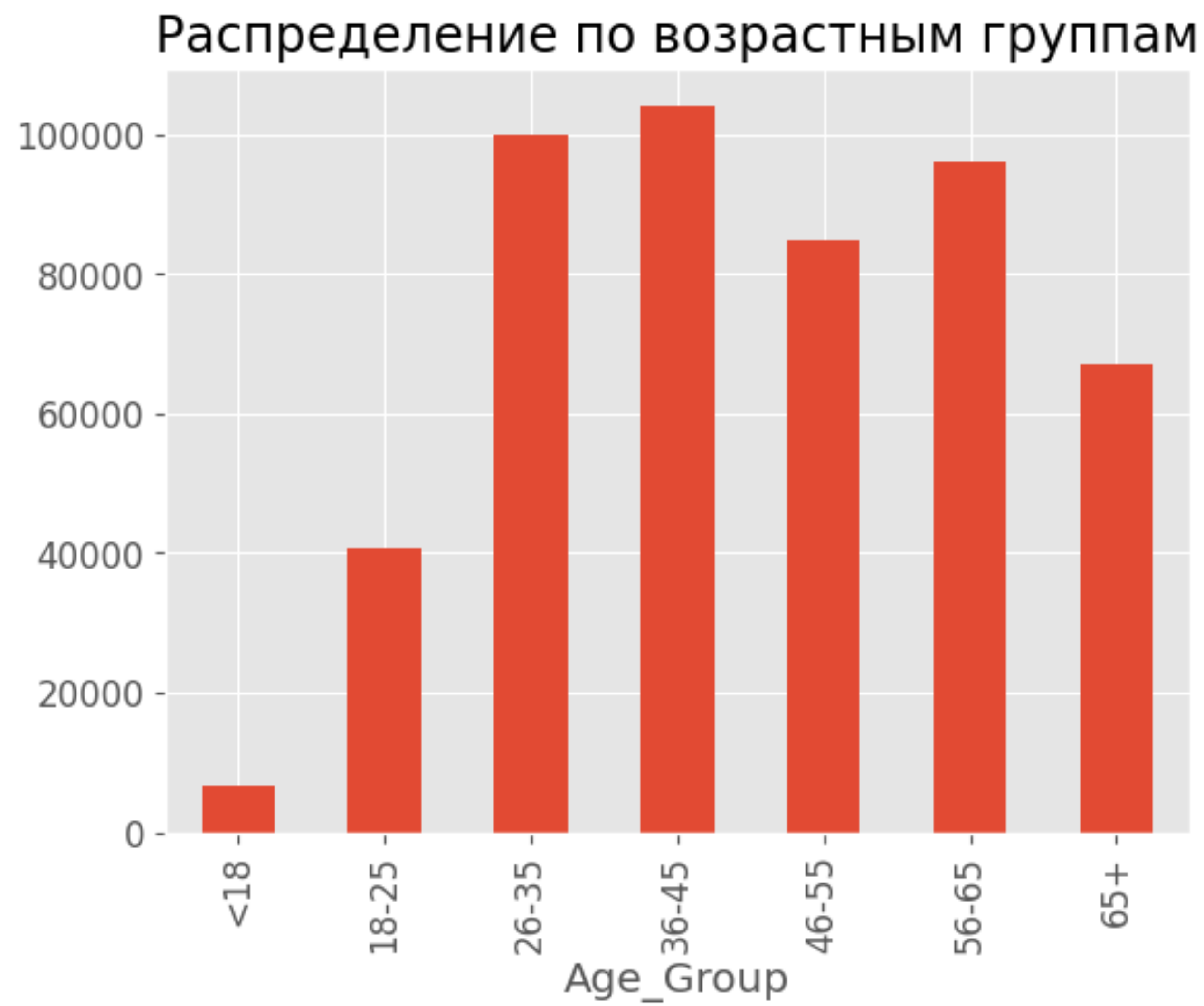
Распределение покупателей по участию в акции



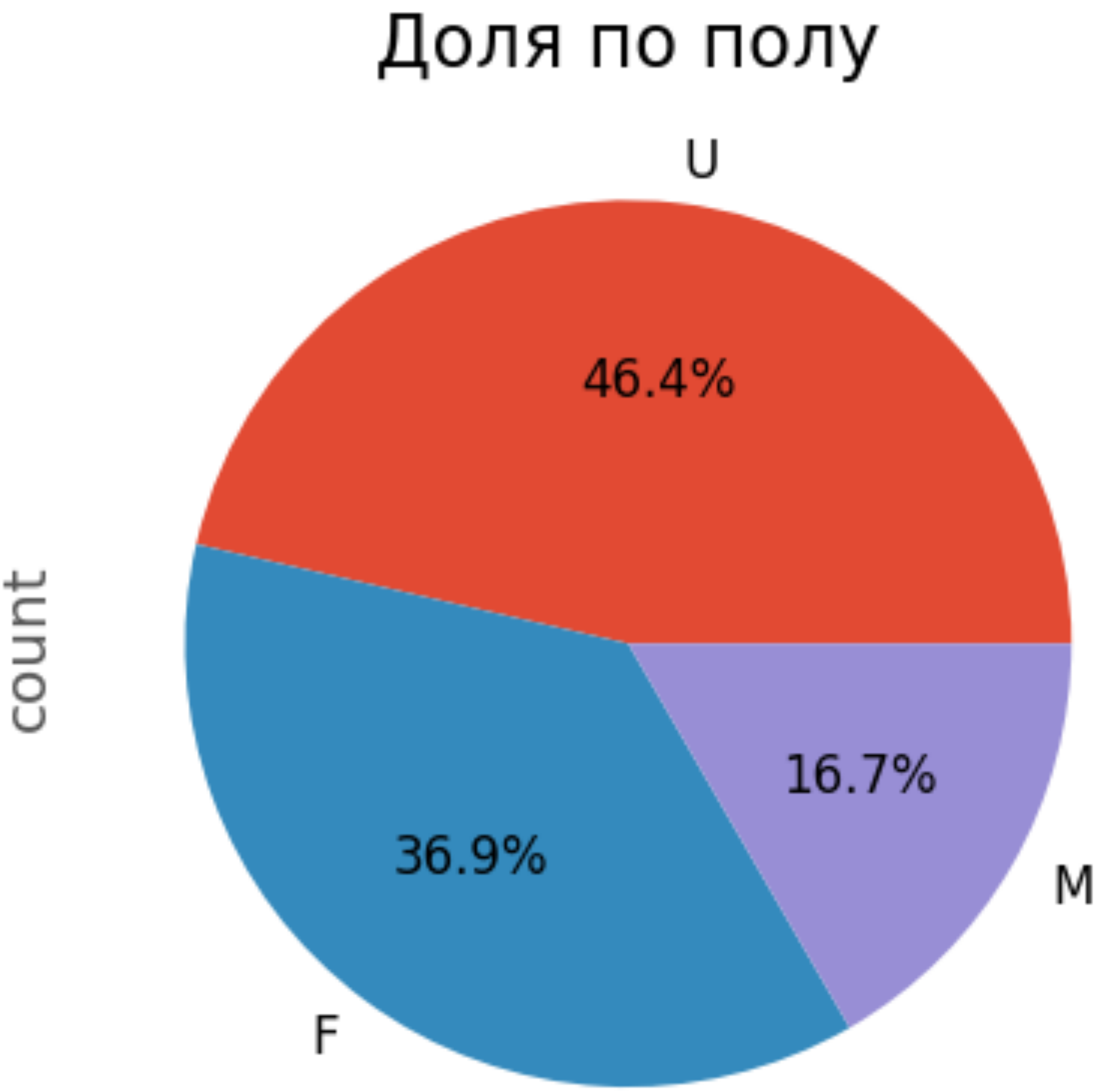
Среднее количество покупок



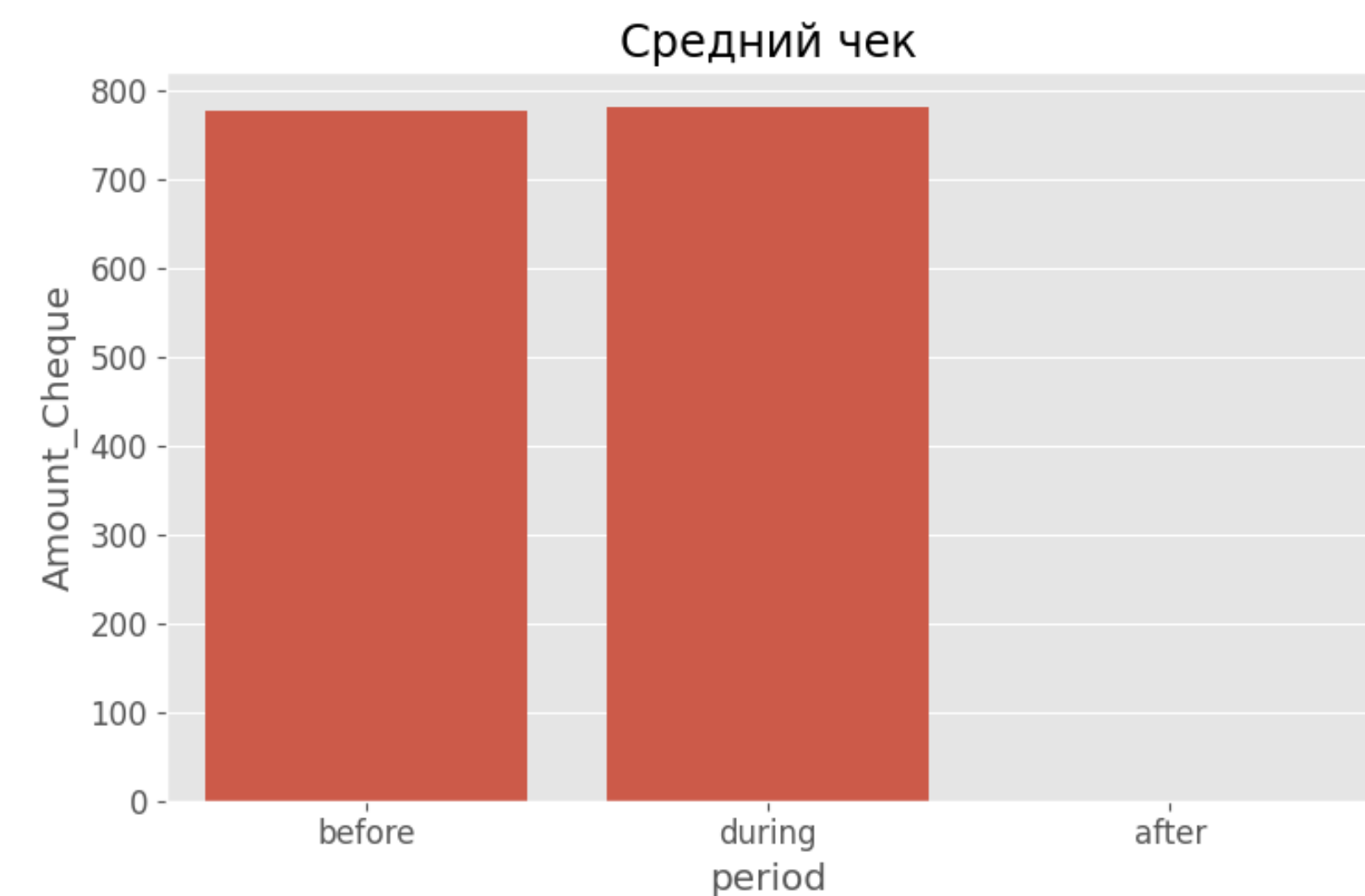
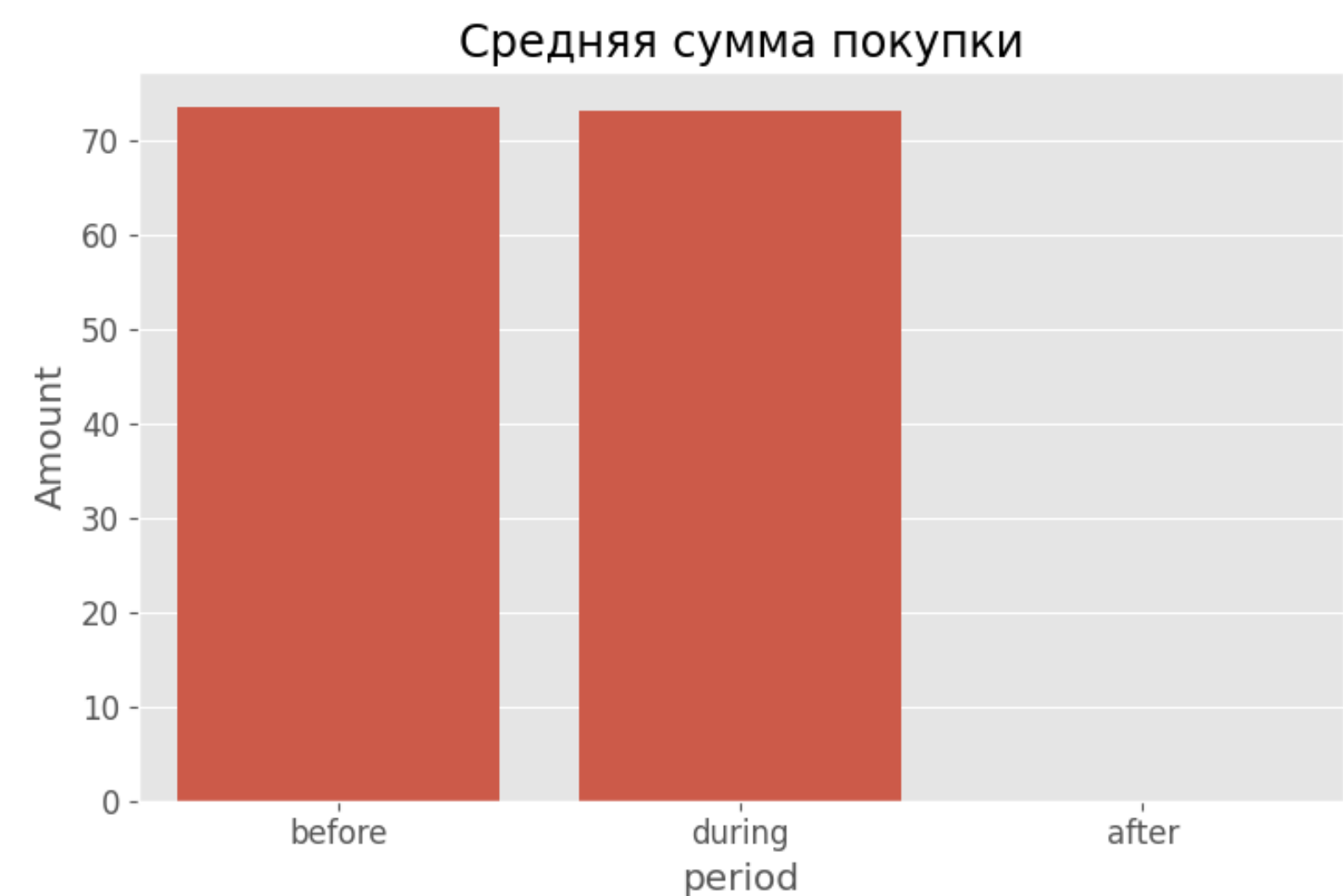
Основные возрастные категории участников акции



Гендерный состав участников акции

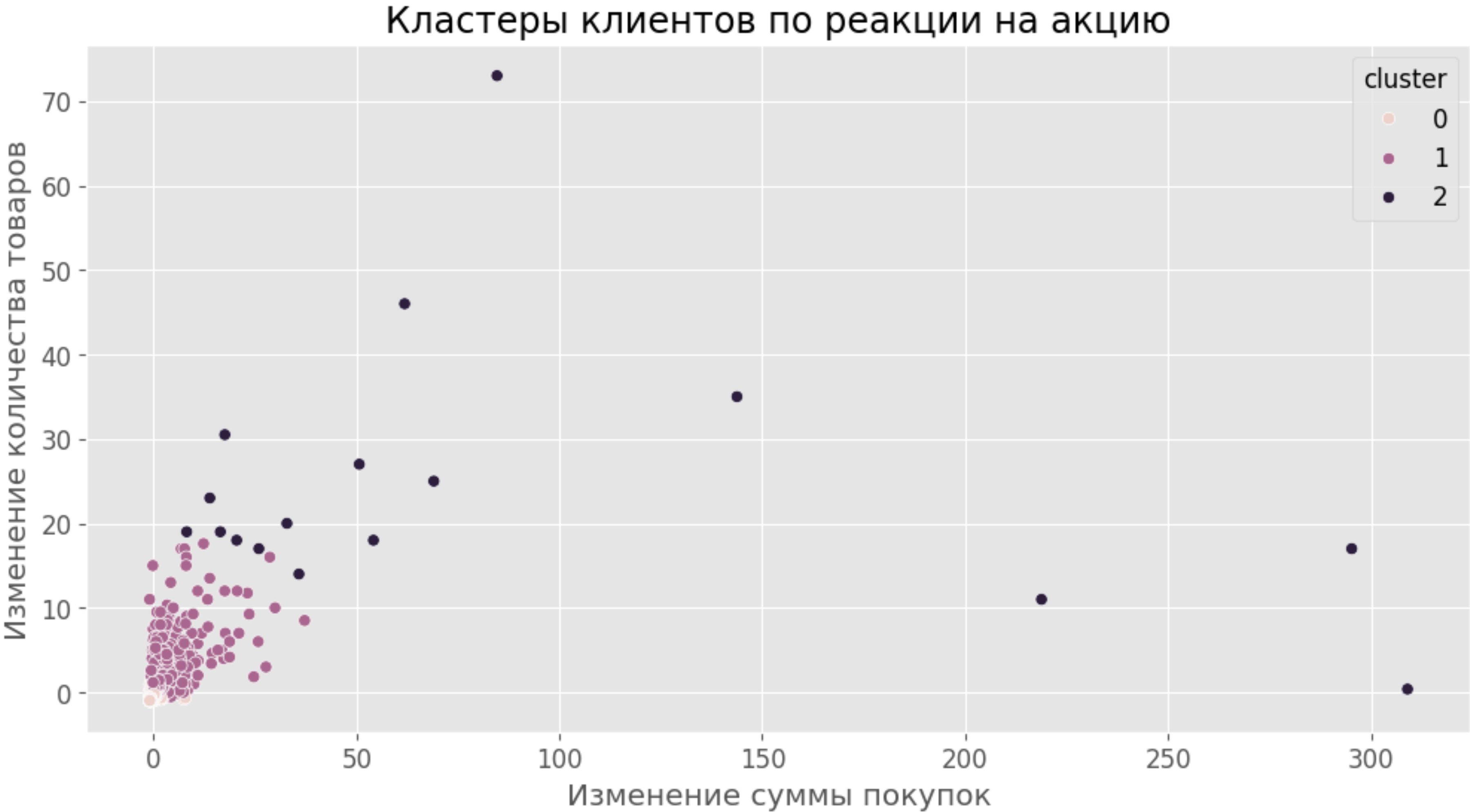


Сравнительный анализ



Изменение количества и стоимости покупок покупателей разной чувствительности

0- покупатели, снизившие покупки 1- не изменившие покупки покупатели 2- покупатели, количество и суммы покупок которых значительно увеличились



Изменение продаж категории товаров 1- го уровня до/во время/после акции



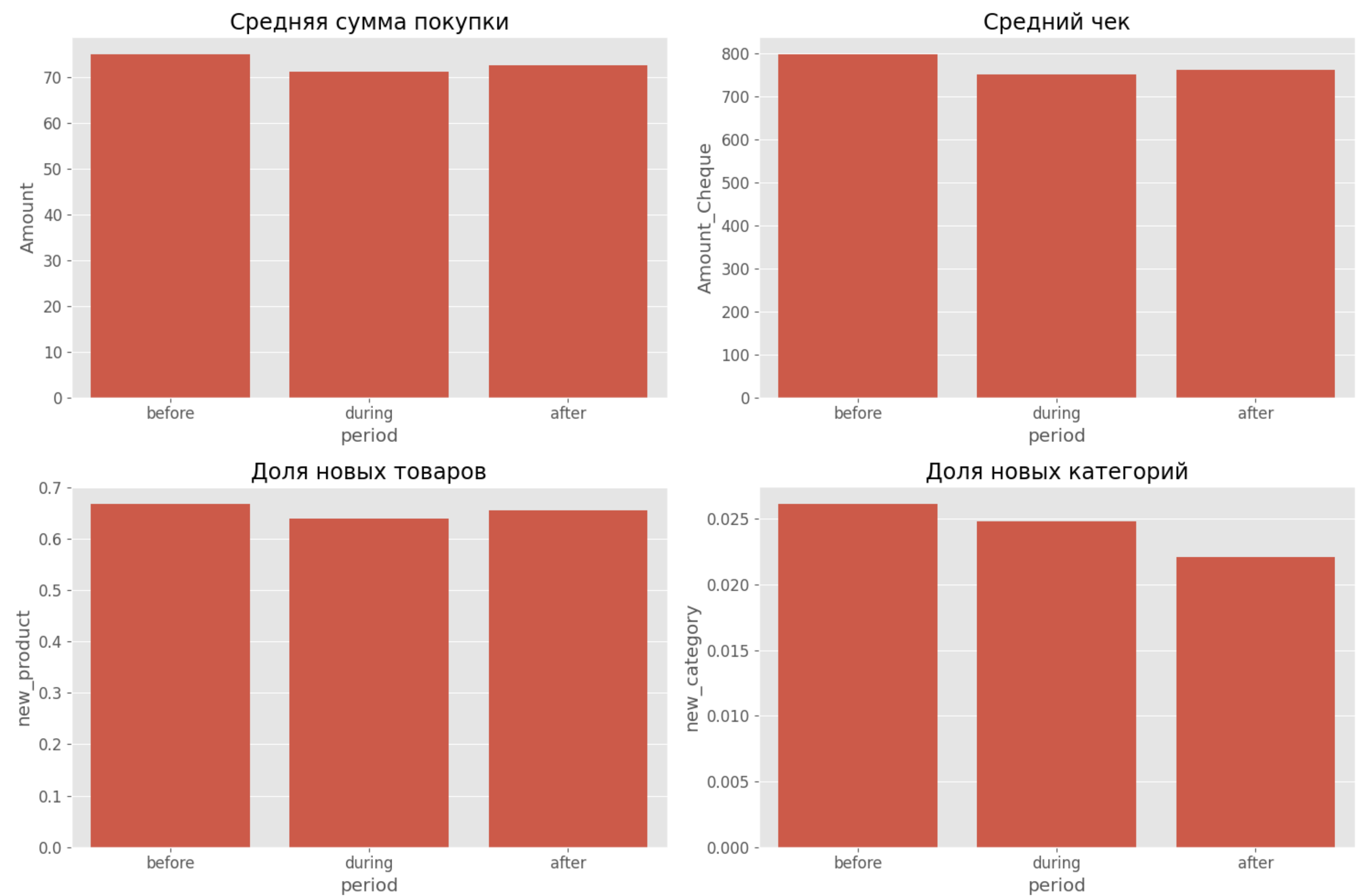
Изменение количества продаж



Выводы

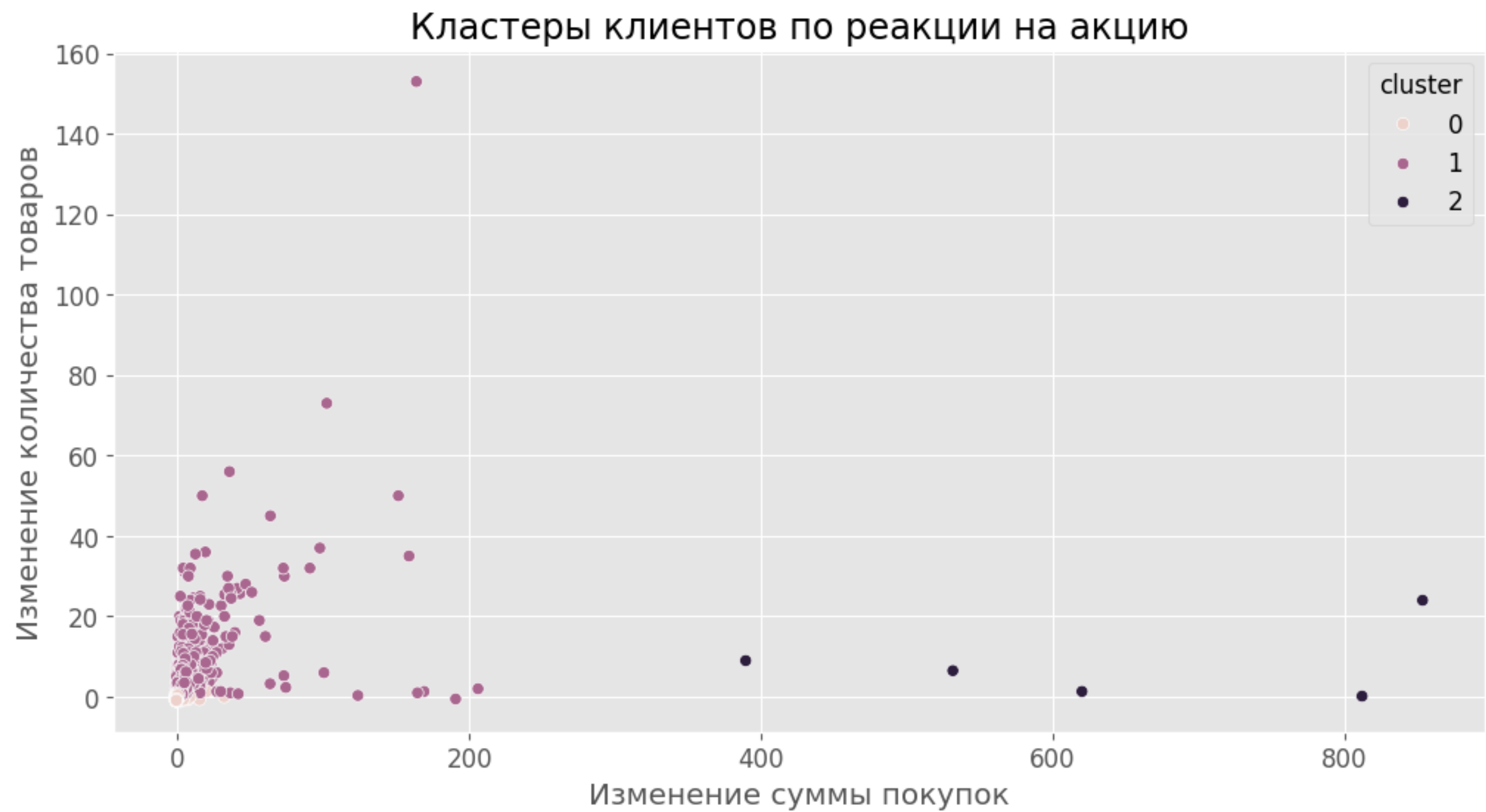
1. Анализ массива List_Purchases(покупки) показал что все покупки совершены до 18.03.2019 г., в то время как акция проходила 21.03.2021 г. Период после акции не определяется.
2. Всплеск покупок приходится на периоды- конец декабря 2018г., вторая половина февраля 2019 г., первая декада марта 2019 г., что скорее всего связано с праздничными днями календаря(Новый год, День защитника Отечества и Международный Женский День)

Для демонстрации поэтапного сравнительного анализа была принята другая дата проведения акции - 21.01.2019 г.

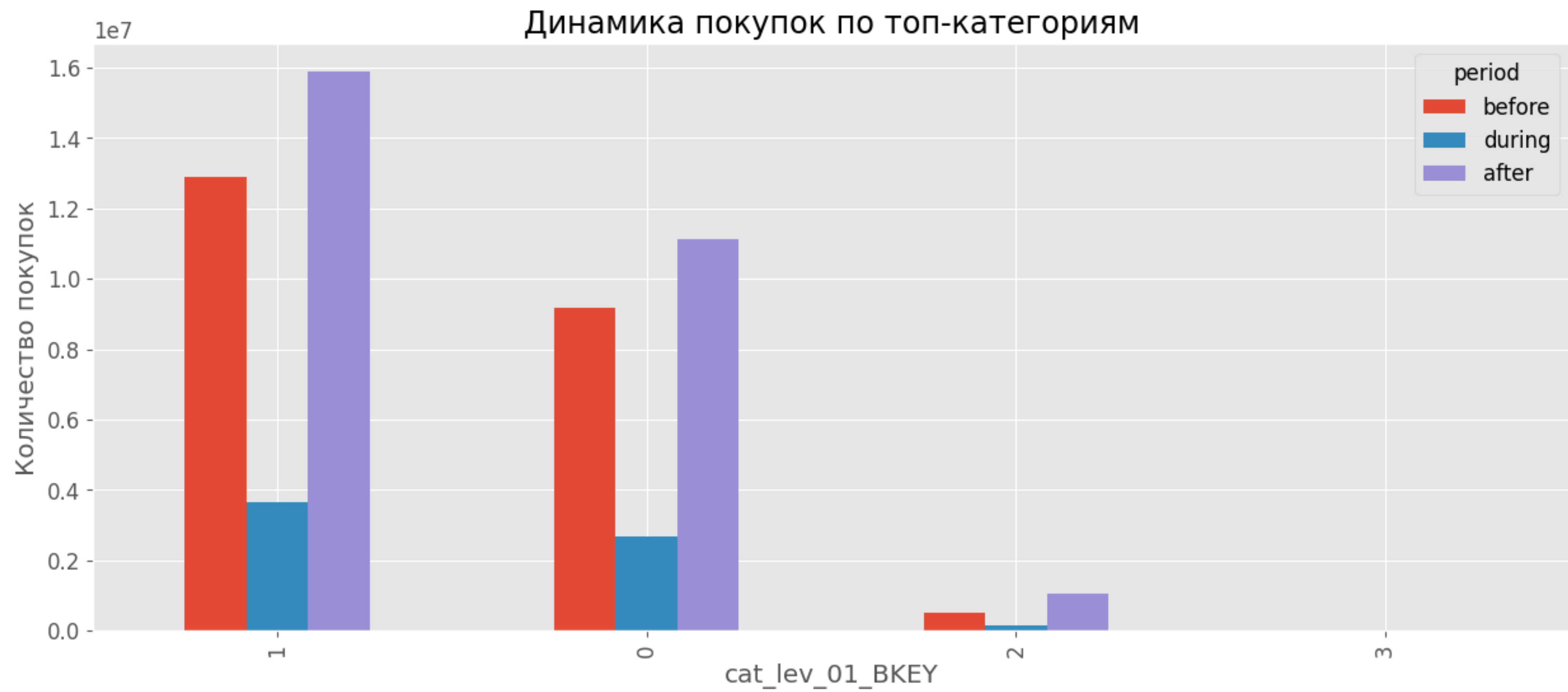


Изменение количества и стоимости покупок покупателей разной чувствительности

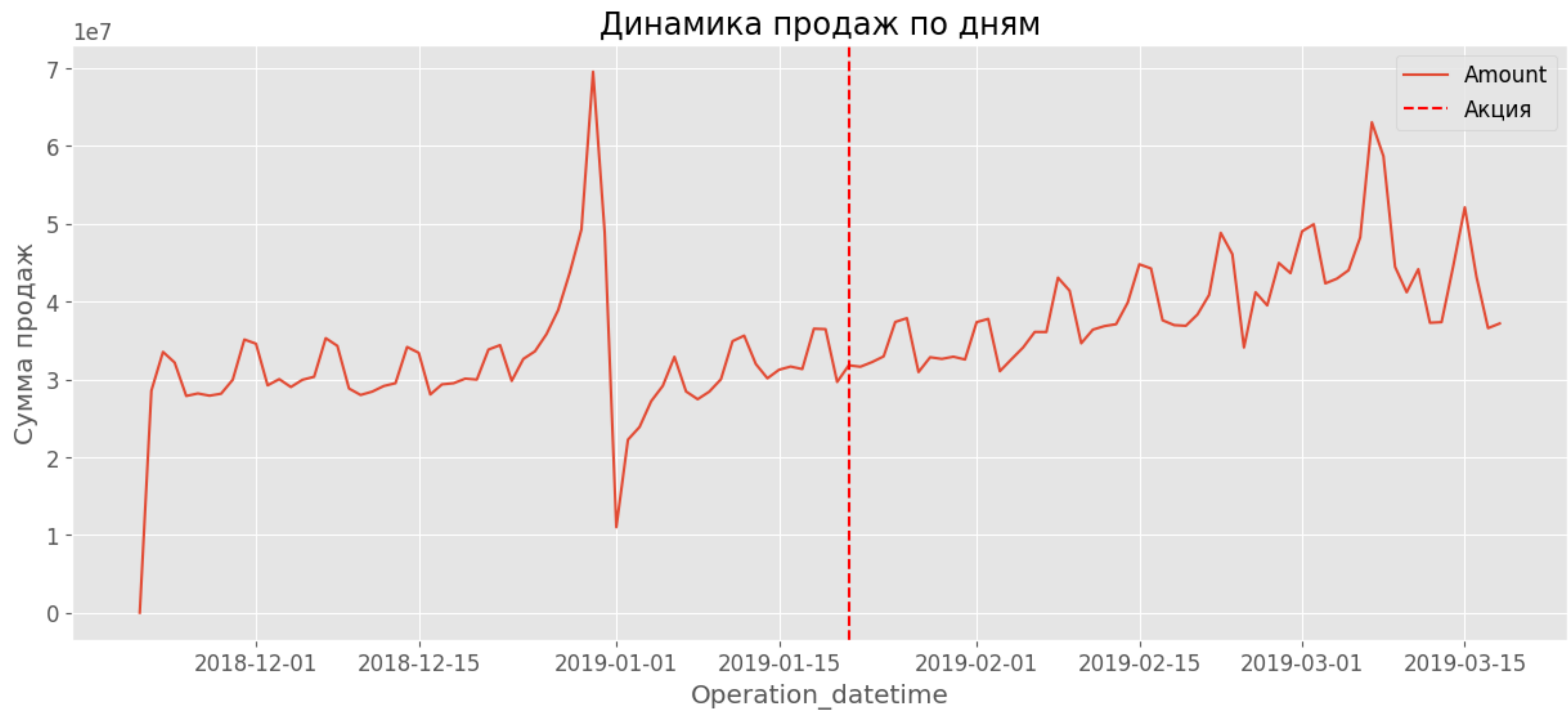
0- покупатели, снизившие покупки 1- не изменившие покупки покупатели 2- покупатели, количество и суммы покупок которых значительно увеличились



Изменение продаж товаров категории 1-го уровня до/во время/после акции



Изменение количества продаж



Выводы

1. Необходимо учитывать что дата проведения акции была выбрана случайно, и все выводы случайны, исследование проведено для демонстрации корректного определения периодов до/во время/после акции.
2. Средний чек снизился на 5.8% во время акции
3. Доля новых товаров составила 63.9% во время акции (-2.9% изменение)

Дополнительно можно провести исследования для других уровней категорий товаров.