

Дизайн и реализация А/В-теста по изменению позиции блока с премьерами в стриминговом сервисе



Выполнил проект
Уляшев Виталий

Введение

Цель: оценить, влияет ли изменение позиции блока «Премьеры недели» на пользовательское поведение.

Гипотеза: если поднять блок «Премьеры недели» в самый верх главной страницы, то это увеличит его кликабельность и вовлечённость, а также позитивно скажется на метриках потребления контента.

Условия и ограничения:

- Аналогичный эксперимент проводился ранее на других блоках — тогда мы получили реальный прирост ключевой метрики на 2.5%, поэтому ожидаемое минимальное детектируемое изменение (MDE) = 2.5%.
- Продолжительность теста: 7 дней (важно уложиться до запуска большого промо и смены блока на новые премьеры).
- Эксперимент охватывает всех пользователей, заходящих на домашнюю страницу в указанный период.
- Тест считается успешным, если тестовая группа показывает значимое улучшение по кликабельности блока и конверсии в начало просмотра.

Первичный анализ

- У нас есть две группы: тестовая и контрольная. В тестовой группе наблюдается больше событий.
- Есть 6 видов событий:
 - *view_home* - посещение домашней страницы,
 - *view_block* - просмотр блока,
 - *click_block* - нажатие на карточку из блока,
 - *start_watch* - начало просмотра,
 - *finish_watch* - досмотрел до конца,
 - *revenue* - оплата за просмотр, если есть.
- Явных выбросов в распределении оплаты за контент не обнаружено. Минимальная оплата - 0.63 у.е., максимальная - 2.81 у.е. Чаще покупали контент за 2.34 у.е.
- Продолжительность эксперимента: 5 дней 23:19:10, с "2025-04-01 00:03:02" по "2025-04-06 23:22:12".
- Количество уникальных клиентов в эксперименте: 120 человек.
- Пересечений в группах между пользователями не обнаружено.

Ключевые и дополнительные метрики

Ключевые метрики:

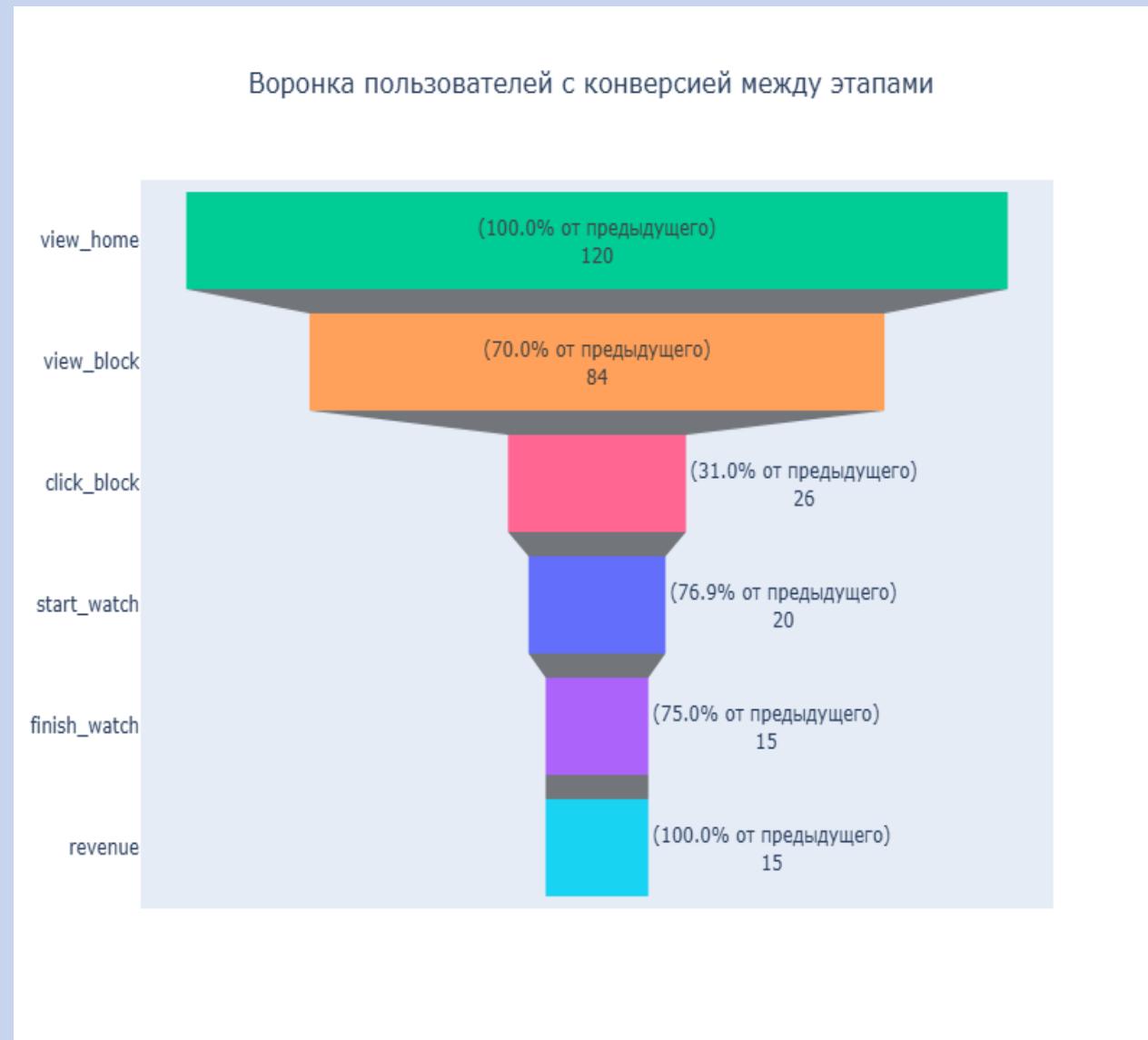
- CTR_block - кликабельность блока «Премьеры недели» (кол-во кликов на карточки в блоке / кол-во просмотров блока),
- CR_start_watch - конверсия в начало просмотра (от просмотра главной страницы --> до старта любого контента из «Премьер недели»).

Дополнительные метрики:

- CR_finish_watch - конверсия в полное просмотренное видео (начал просмотр контента --> досмотрел до конца),
- ARPU - выручка на пользователя за контент из блока «Премьеры недели» (сумма взносов / кол-во уникальных пользователей).

Поведение пользователей до эксперимента

- Узкое место образовалось между просмотром блока и нажатием на него. Просматривать видео начинает каждый 6 пользователь. Из тех кто начал смотреть контент 75% просматривают его до конца.
- Общее поведение пользователей по заданным метрикам:
 - общая кликабельность блока: 30.95%,
 - общая конверсия в начало просмотра: 16.67%,
 - общая конверсия в полное просмотренное видео: 75.00%,
 - среднюю продолжительность просмотра определить не получилось из-за ошибки логирования,
 - ARPU: 0.22 у.е.



Дизайн эксперимента

Параметры для теста:

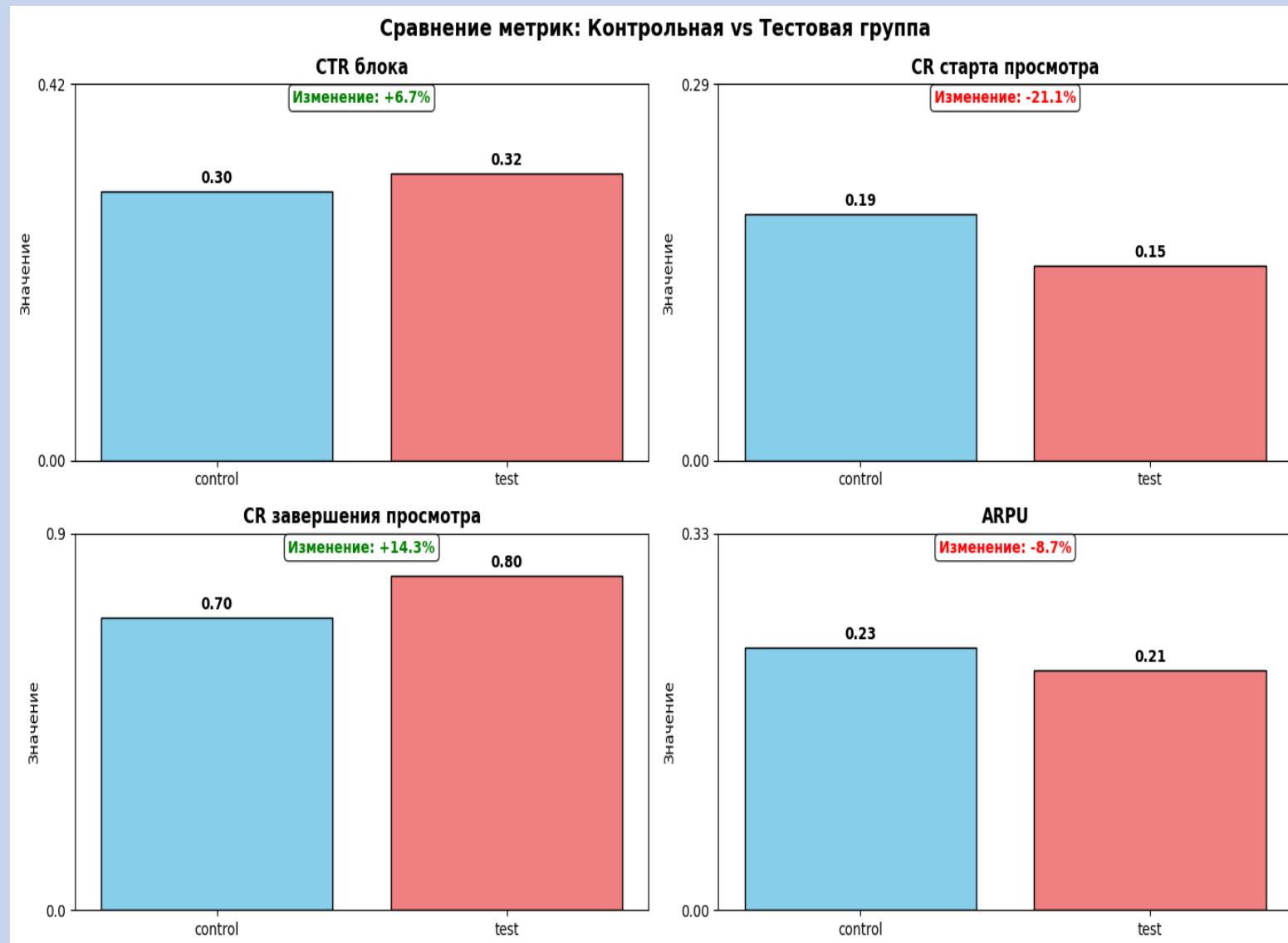
- уровень значимости: 0.05,
- мощность: 0.8,
- MDE: 0.025 (2.5%).

Внимание!

- Расчетный размер выборки значительно превышает нашу экспериментальную выборку. На нашей выборке мы сможем обнаружить MDE, который будет превышать 24-25%. Для обнаружения эффекта в 2.5% потребуется выборка из 8638 пользователей.
- Можно применить для статистического теста метод бутстрэп, но стоит быть осторожным с результатами. Мы хотим обнаружить маленький эффект на маленькой выборке, что крайне сложно и результат может быть недостоверным.

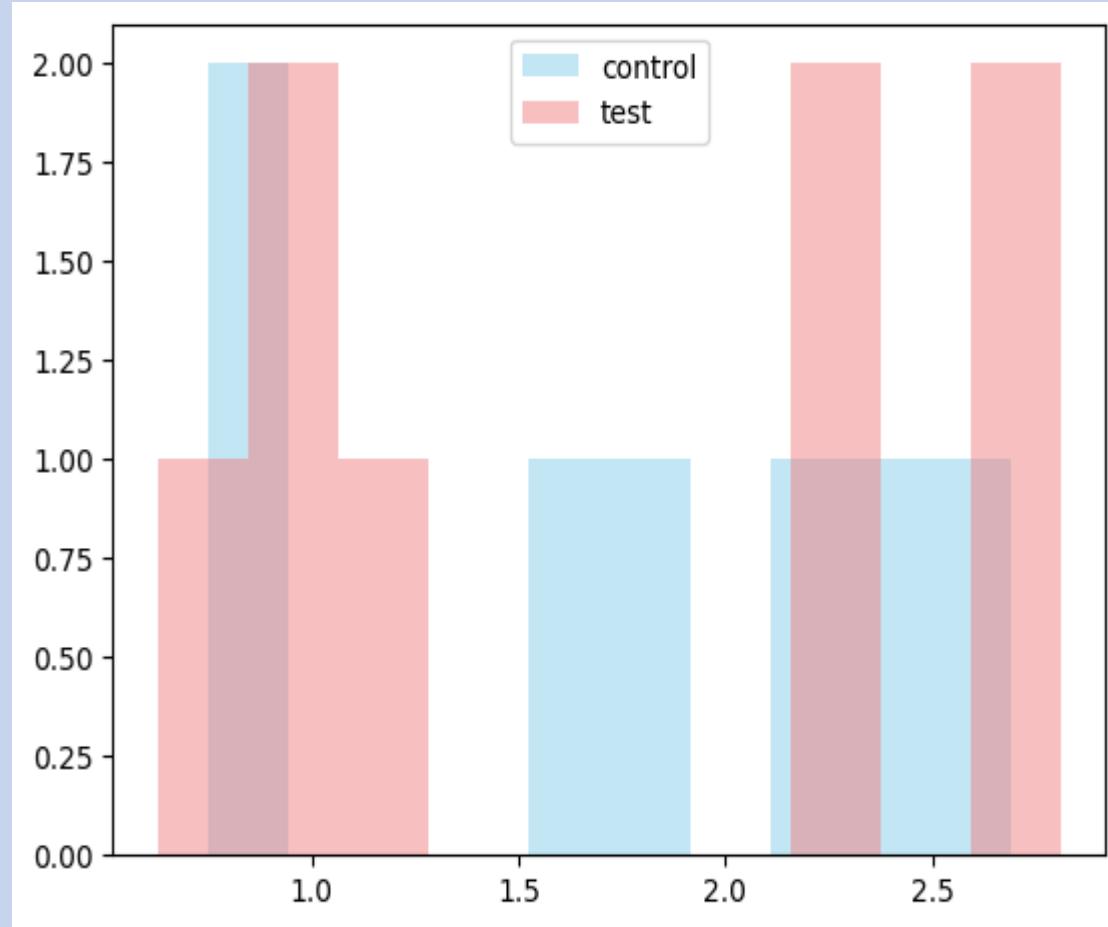
Сравнение метрик в контрольной и тестовой группах

Нельзя однозначно выделить одну из групп, как лучшую.
Тестовая группа превосходит контрольную в метриках CTR блока на (+6.7%) и CR завершения просмотра на (+14.3%).
Но уступает контрольной группе в метриках CR старта просмотра (-21.1%) и ARPU (-8.7%).



Сравнение метрик в контрольной и тестовой группах

Распределение показывает, что тестовая группа больше тратит на оплату контента.
Но стоит учесть, что в тестовой группе больше пользователей, а по метрике ARPU (оплата на пользователя) она уступает контрольной.

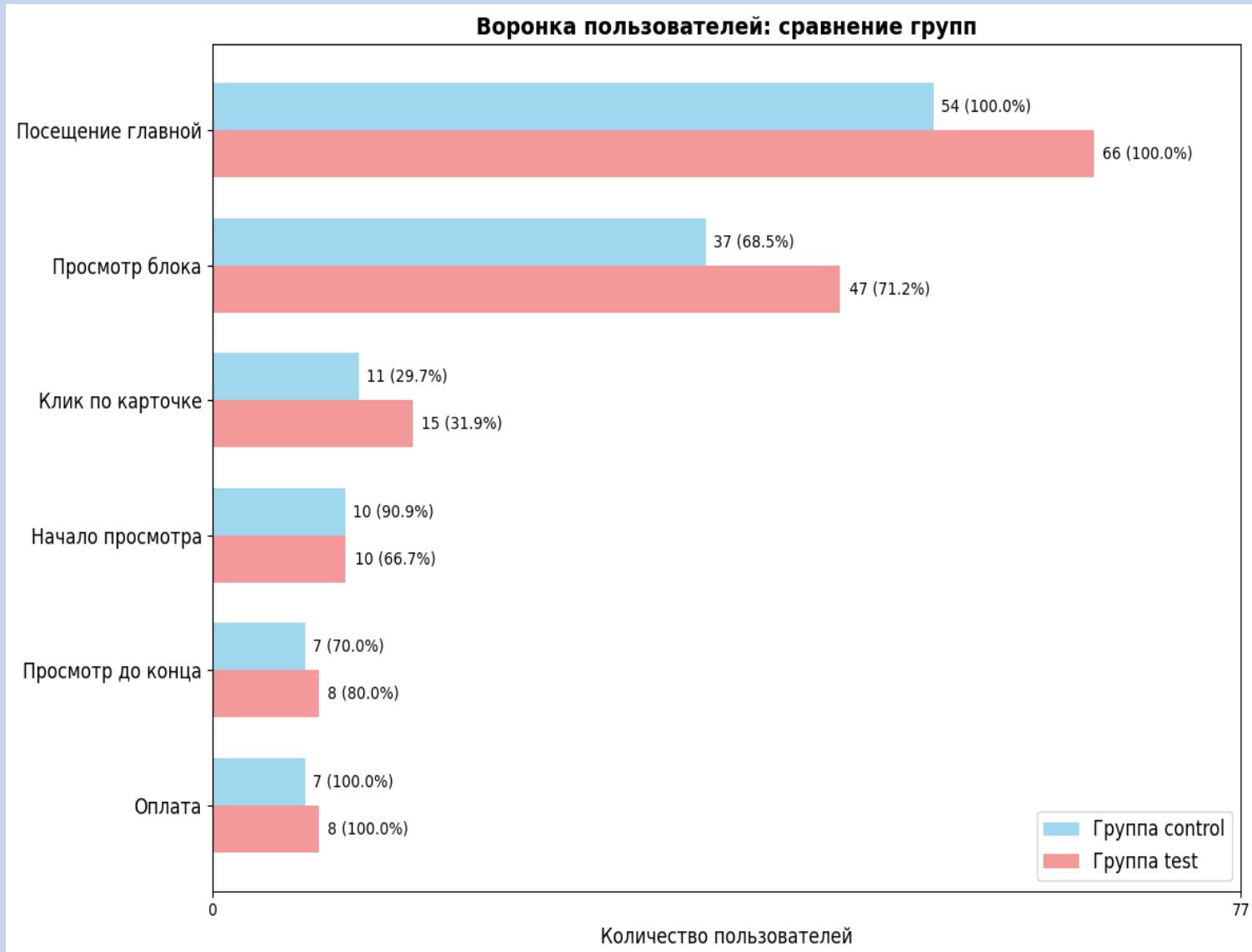


Сравнение метрик в контрольной и тестовой группах

На пользовательской воронке тестовая группа однозначно смотрится лучше. На это может влиять большее количество пользователей в teste.

В начало просмотра контента приходит одинаковое количество пользователей, не смотря на изначальную их разницу в группах. В контрольной группе в начало просмотра после клика по карточке переходят охотнее.

Стоит обратить внимание, что в тестовой группе 28.8% пользователей посетившие домашнюю страницу не просматривают блок с премьерами, хотя он находится в самом верху. В относительном сравнении с контрольной группой разница не велика, там не просматривают блок 31.5% пользователей.



Бутстрэп тест для CTR блока

Нулевая гипотеза (H_0):

Различие CTR блока между тестовой и контрольной группами статистически незначимо и не превышает 2.5%.

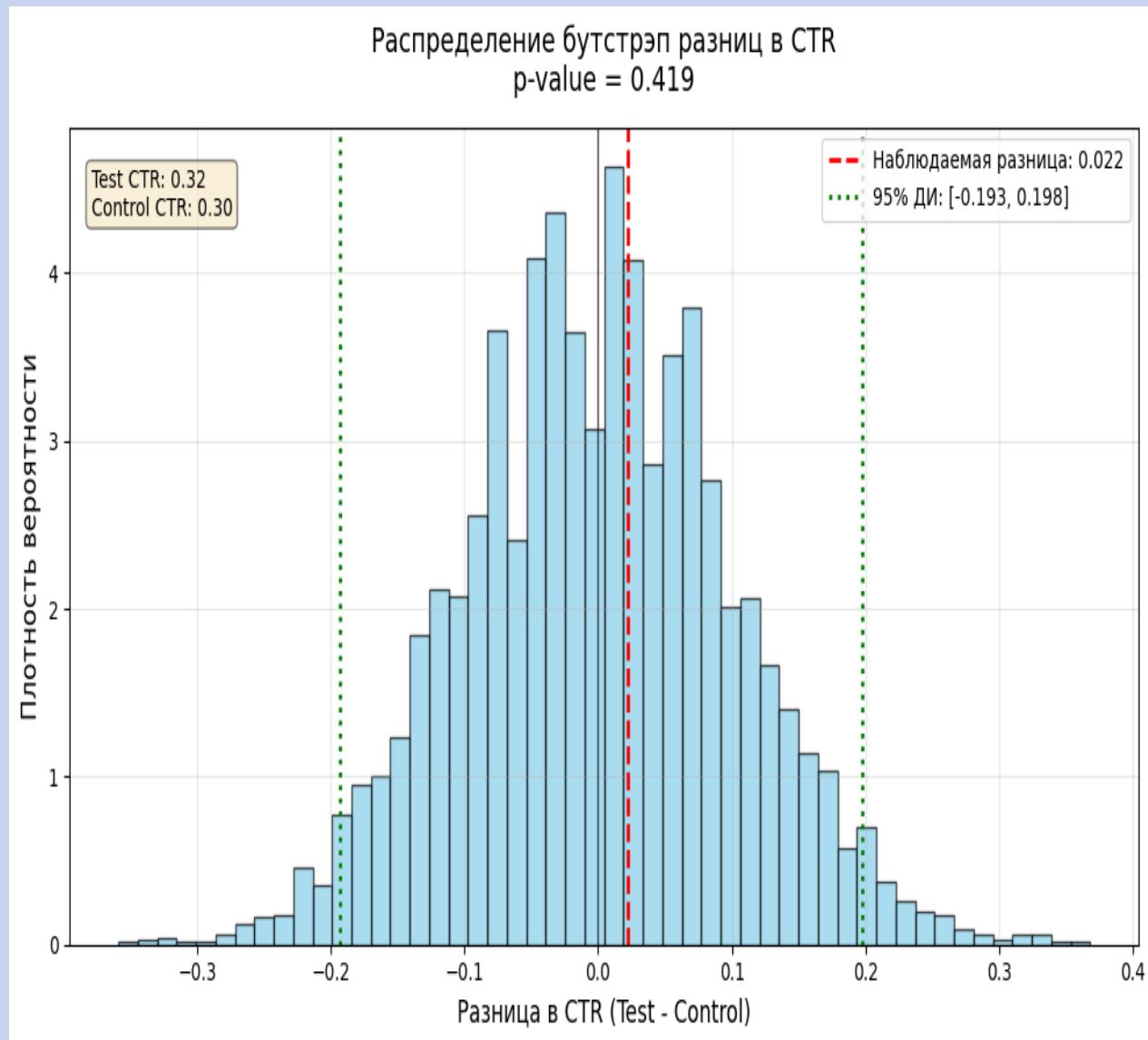
Альтернативная гипотеза (H_1):

CTR блока в тестовой группе превышает CTR блока в контрольной группе более чем на 2.5%.

Результаты теста:

- Разница средних значений в группах: 0.022
- P-value: 0.419
- 95% Доверительный интервал: [-0.193, 0.198]

Проведенный А/В-тест для CTR блока не выявил статистически значимой разницы между тестовой и контрольной группами. Наблюданное различие в 2.2 процентных пункта может быть объяснено случайной вариацией данных, об этом говорит слишком широкий доверительный интервал, включающий ноль.



Бутстрэп тест для CR в просмотр

Нулевая гипотеза (H_0):

Различие CR в просмотр между тестовой и контрольной группами статистически незначимо и не превышает 2.5%.

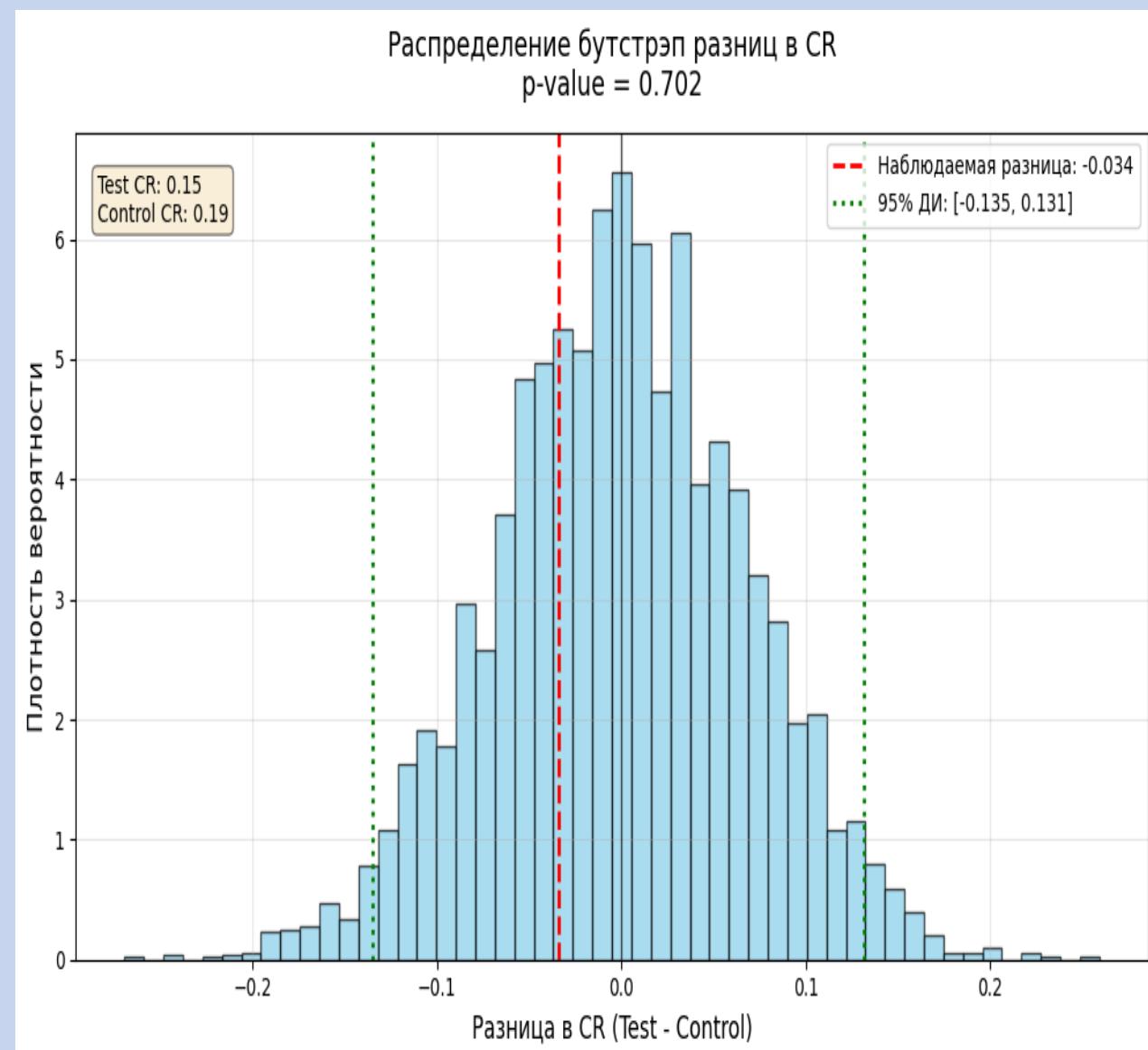
Альтернативная гипотеза (H_1):

CR в просмотр в тестовой группе превышает CR в просмотр в контрольной группе более чем на 2.5%.

Результаты теста:

- Разница средних значений в группах: -0.034
- P-value: 0.702
- 95% Доверительный интервал: [-0.135, 0.131]

Проведенный А/В-тест для CR в начало просмотра не выявил статистически значимой разницы между тестовой и контрольной группами. Конверсия в тестовой группе на 3.4 процентных пункта ниже, чем в контрольной. Но эта разница не статистически значима, вероятность получить такую разницу случайно: 70.4%. Эффект слишком неопределенный для выводов.



Общий вывод

- Собрано только 120 пользователей, расчетный размер выборки значительно превышает нашу экспериментальную группу. На нашей выборке мы сможем обнаружить MDE, который будет превышать 24-25%. Для обнаружения эффекта в 2.5% потребуется выборка из более чем 8000 пользователей.
- Продолжительность теста была 6 дней вместо 7 дней.
- Исходя из того, что выборка маленькая, был проведен бутстрэп тест для двух метрик: CTR блока и CR в начало просмотра. Оба теста не показали статистически значимой разницы между тестовой и контрольной группой.
- Бутстрэп тесты показали широкие доверительные интервалы, включающие ноль, поэтому сложно оценить влияние изменения на наши метрики. Также широкие интервалы показывают, что у нас мало данных для выявления стабильного и значимого результата.
- Результаты статистически недёжны и не могут быть основанием для принятия решения. Требуется большее количество пользователей в teste, чтобы сделать качественные заключения.
- Сравнение групп по метрикам показывает, что тестовая группа уступает в конверсии в начало просмотра, 19% - контроль и 15% - тест; в средних тратах на пользователя (ARPU), 0.23 у.е. - контроль, 0.21 у.е. - тест.
- После посещения домашней страницы относительная разница просмотра блока с премьерами в группах не особо велика, 28.8% в тестовой группе и 31.5% в контрольной. Хотя в тестовой группе этот блок находится в самом верху.

Рекомендации

1. Пока не стоит раскатывать фичу, есть вероятность, что изменения негативно повлияют на метрики. Рекомендуется сохранить текущую версию интерфейса.
2. Рекомендуется перезапустить тест на более ёмком канале, для качественных выводов нужно увеличить выборку и продлить эксперимент.
3. Стоит доработать блок с премьерами в тестовой группе. Из эксперимента видно, что относительная разница просмотра блока с премьерами между группами не особо велика, хотя в тестовой группе этот блок находится в самом верху.
4. Использовать этот эксперимент как пилотное исследование для формирования новых гипотез и проанализировать отзывы клиентов из тестовой группы.

Спасибо за внимание!

