

ТПОЭ - 23/24

Лекция 4

Нерсес Багиян

В предыдущих сериях

1. Мы разобрались, какие метрики бывают
2. Разобрались, что такое сплит система и за какими ошибками следим в рамках А/Б тестов
3. Давайте вернемся к вопросу, а как работать с ретеншеном и LTV?

Главные метрики в продукте

1. Retention - показываем как много пользователей будут возвращаться
2. LTV - показывает, сколько прибыли принесет пользователь на протяжении своей жизни

Вспомним: какие тут проблемы?

Вопрос аудитории

С какими метриками можно проводить эксперимент?

Чувствительность

Насколько долго нужно ждать, чтобы увидеть изменения в метрике

Достоверность

Можно ли получить точное подтверждение из данных?



Интерпретируемость

Можем ли четко ответить, что улучшили?

Способ подсчета

Среднее, отношение или квантиль

Пример: площадка объявлений

Представим, что у нас есть сайт с объявлениями — классифайд, — который умеет только публиковать карточки товаров и показывать номера телефонов продавцов. Потенциальный покупатель должен кликнуть на кнопку «позвонить», чтобы увидеть номер.

Вопрос: что для такого продукта будет метрикой роста? А NSM?

Продам кота-ублюдка

Размещено 10 авг. в 20:08. [Редактировать](#), [удалить](#), [поднять объявление](#)



AVITO.ru

Цена 9 999 руб.

Продавец [alexander](#) (компания) < написать письмо

Телефон 8 999 XXX-XX-XX < показать номер

Город Алтайский край, Бийск < показать на карте

Вы владелец объявления? [Откройте магазин на AVITO.ru](#) и увеличьте продажи!

Порода: Британская
НАСТОЯЩИЙ ЗАСРАНЕЦ!!!
ОБМЕН на вашу резину или диски ЛЮБЫХ размеров в любом состоянии!!!!
ОТПРАВКА В РЕГИОНЫ!!!!

Пример: площадка объявлений

Представим, что у нас есть сайт с объявлениями – классифайд, – который умеет только публиковать карточки товаров и показывать номера телефонов продавцов. Потенциальный покупатель должен кликнуть на кнопку «позвонить», чтобы увидеть номер.

Вопрос спорный, но среди таких метрик могут быть:

- Сумма проданных товаров на площадке
- Количество сделок
- Количество посетителей ежедневно

В чем проблема GMV?

Представим, что у нас есть сайт с объявлениями – классифайд, – который умеет только публиковать карточки товаров и показывать номера телефонов продавцов. Потенциальный покупатель должен кликнуть на кнопку «позвонить», чтобы увидеть номер.

В идеале было бы замерять GMV, но тут возникает проблема с достоверностью такой метрики. Ее нельзя точно узнать из данных

Достоверность

Можно ли получить точное подтверждение из данных?

Прокси метрики

Прокси метрика - косвенная метрика целевой метрики, с которой она сильно коррелирует и с которой есть причинно-следственная связь.

Корреляция

Высоко коррелирована с основной метрикой

Казуальная связь

Имеет причинно-следственную связь с целевой метрикой

Интерпретируемая

Метрику может объяснить каждый сотрудник

Чувствительная

Насколько долго нужно ждать, чтобы увидеть изменения в метрике

Достоверная

Можно ли получить точное подтверждение из данных?

Какая прокси метрика может быть в GMV классифайда?

Вопрос аудитории

Пример: площадка объявлений

В нашем кейсе можно было выбрать метрику **количество контактов**

Корреляция

Чем больше контактов, тем больше продаж

Казуальная связь

Изменение в количестве контактов ведет к изменению продаж

Интерпретируемая

Понятная

Чувствительная

Можно детектировать изменения

Достоверная

Можем считать из данных

Как искать прокси метрику?

Алгоритм:

1. Нужно собрать исторический пул в метрике, к которой ищем прокси
2. Подбираем пул кандидатов для исследований
3. Проверяем корелляцию
4. Проверяем причинно-следственную связь
5. Проводим проверку ошибок 1-ого и 2-ого рода

Как искать прокси метрику?

Пример с площадки объявлений

1. Нужно собрать исторический пул в метрике, к которой ищем прокси

Обзваниваем продавцов, которые убрали свои объявления с публикации и собираем информацию была продажа или нет

2. Подбираем пул кандидатов для исследований

Собираем пул кандидатов: количество контактов, конверсия в контакт, отзыв на исполнителя

3. Проверяем корреляцию

Считаем корреляцию между двумя метриками

4. Проверяем причинно-следственную связь

Используем методы causal inference для проверки причинно-следственных взаимосвязей

5. Проводим проверку ошибок 1-ого и 2-ого рода

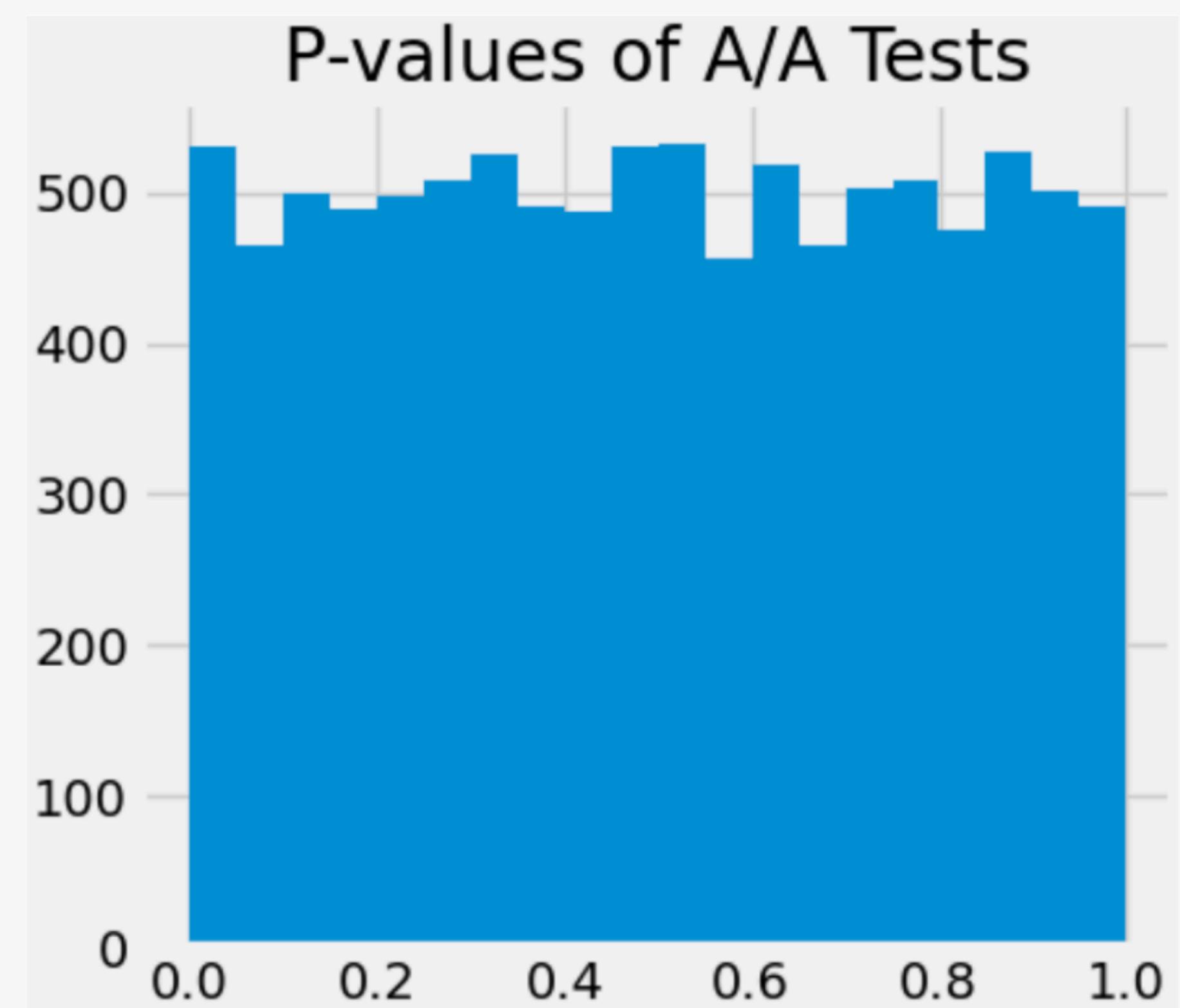
Проводим А/А тесты и синтетический А/В тест

Reminder: A/A тест

Способ проверить, что прокси метрика ничего не ломает

Пусть сплит система на основе исторических данных выдала две группы, которые не участвовали в эксперименте. Если мы сравним их между собой - мы проведем A/A тест. Тест, где между двумя группами нет изменений (нулевая гипотеза верна).

Если провести много-много таких тестов, то распределение будет равномерным и это значит, что сплит система работает правильно с точки зрения ошибки 1-ого рода

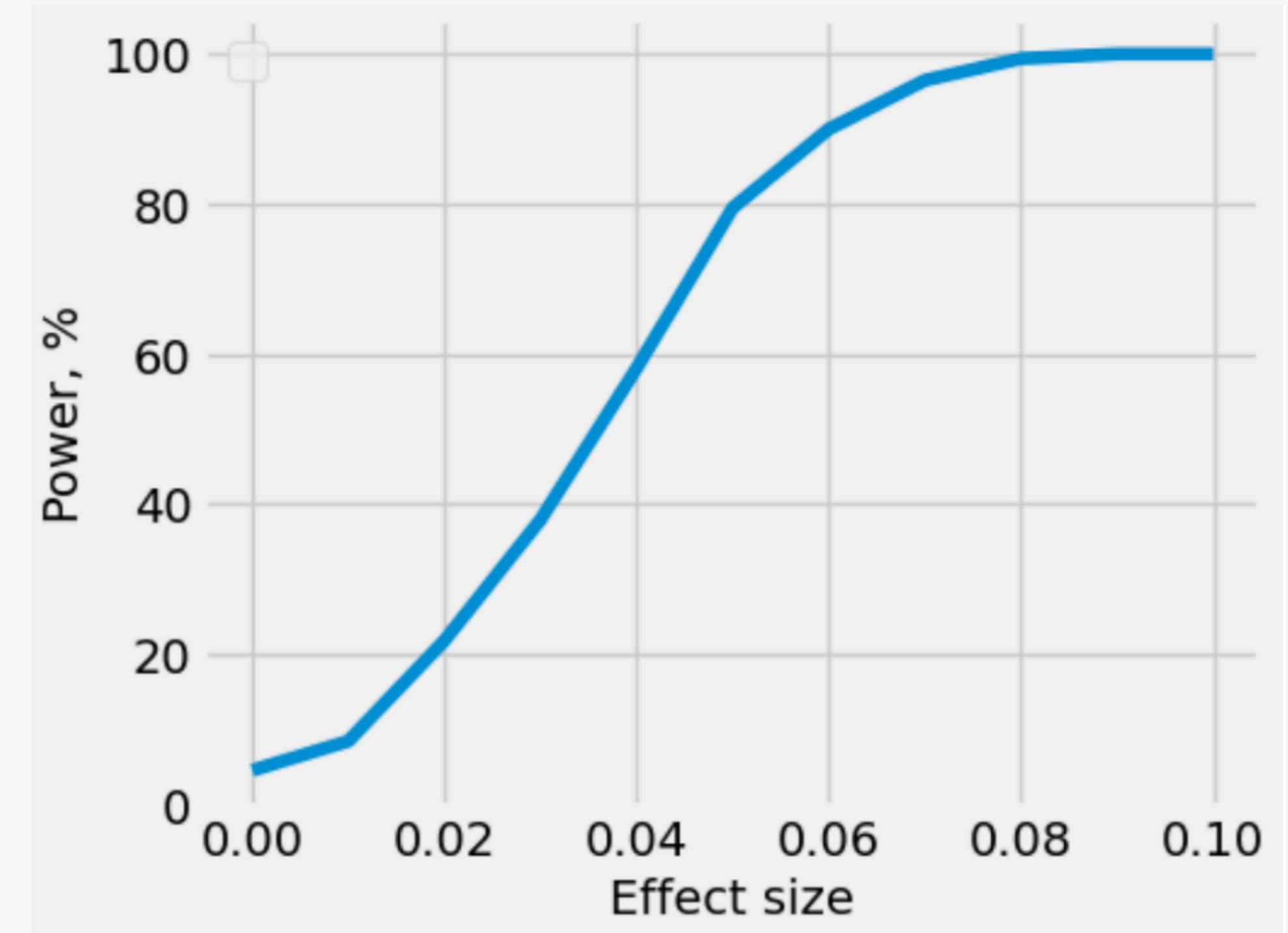


Reminder: Синтетический А/В тест

Способ смоделировать реальный эксперимент

Алгоритм:

1. Берем две группы из сплит системы
2. Генерируем шум, чтобы добавить одной из групп
3. Считаем метрики, натравливаем тест
4. Повторяюм много-много раз с разными значениями среднего у шума (это будет размер эффекта)
5. Получаем картинку как справа



Какая прокси метрика может быть для retention?

Вопрос аудитории

Прокси для retention

Пример с площадки объявлений

Алгоритм:

1. Нужно собрать исторический пул в метрике, к которой ищем прокси
2. Подбираем пул кандидатов для исследований
3. Проверяем корелляцию
4. Проверяем причинно-следственную связь
5. Проводим проверку ошибок 1-ого и 2-ого рода

Прокси для retention

Пример с площадки объявлений

1. Нужно собрать исторический пул в метрике, к которой ищем прокси

Можно собрать исторические данные по поведению клиентов

2. Подбираем пул кандидатов для исследований

Что здесь собираем?

3. Проверяем корреляцию

Считаем корреляцию между двумя метриками

4. Проверяем причинно-следственную связь

Используем методы causal inference для проверки причинно-следственных взаимосвязей. Но можно и забить. Вопрос: почему?

5. Проводим проверку ошибок 1-ого и 2-ого рода

Проводим A/A тесты и синтетический A/B тест

Reminder: что отображает retention?

- Частоту возникновения задачи, для решения которой пользователь рассматривает ваш продукт (характеристика задачи);
- Добавочную ценность вашего продукта относительно альтернатив (характеристика эффективности решения по сравнению с другими способами).

Какие еще метрики могут отображать добавочную ценность?

Вопрос аудитории

Прокси для retention

Пример с площадки объявлений

1. Нужно собрать исторический пул в метрике, к которой ищем прокси

Можно собрать исторические данные по поведению клиентов

2. Подбираем пул кандидатов для исследований

Конверсия в контакт, скорость от начала поиска товаров до контакта, отзыв

3. Проверяем корреляцию

Считаем корреляцию между двумя метриками

4. Проверяем причинно-следственную связь

Используем методы causal inference для проверки причинно-следственных взаимосвязей. Но можно и забить. Вопрос: почему?

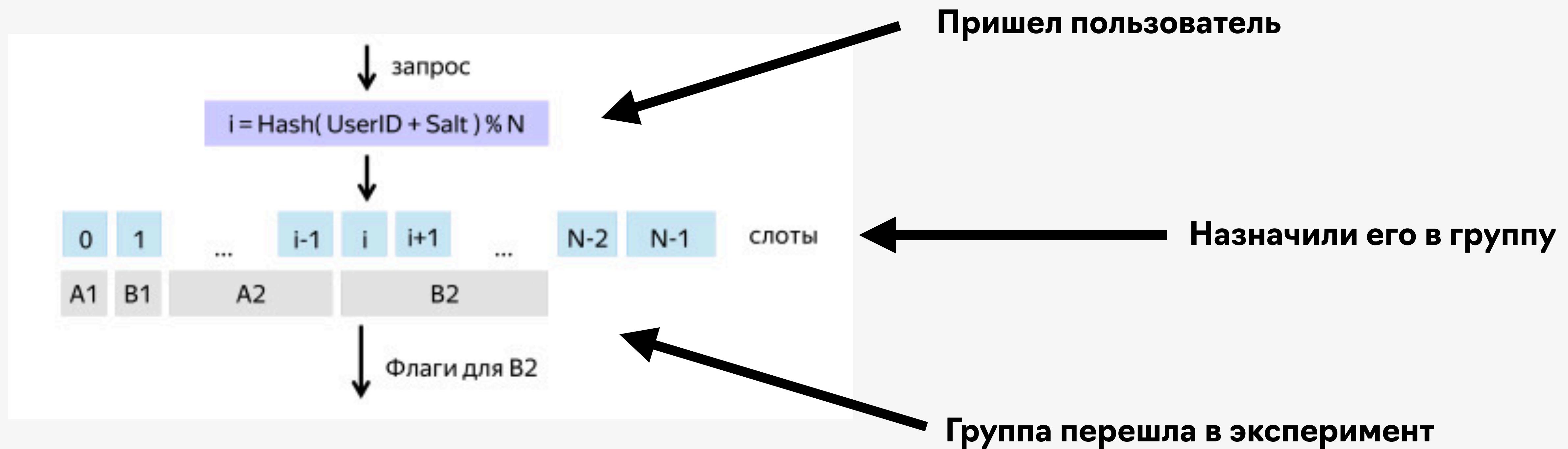
5. Проводим проверку ошибок 1-ого и 2-ого рода

Проводим A/A тесты и синтетический A/B тест

Как делать сплит-системы?

Хеширование с солью

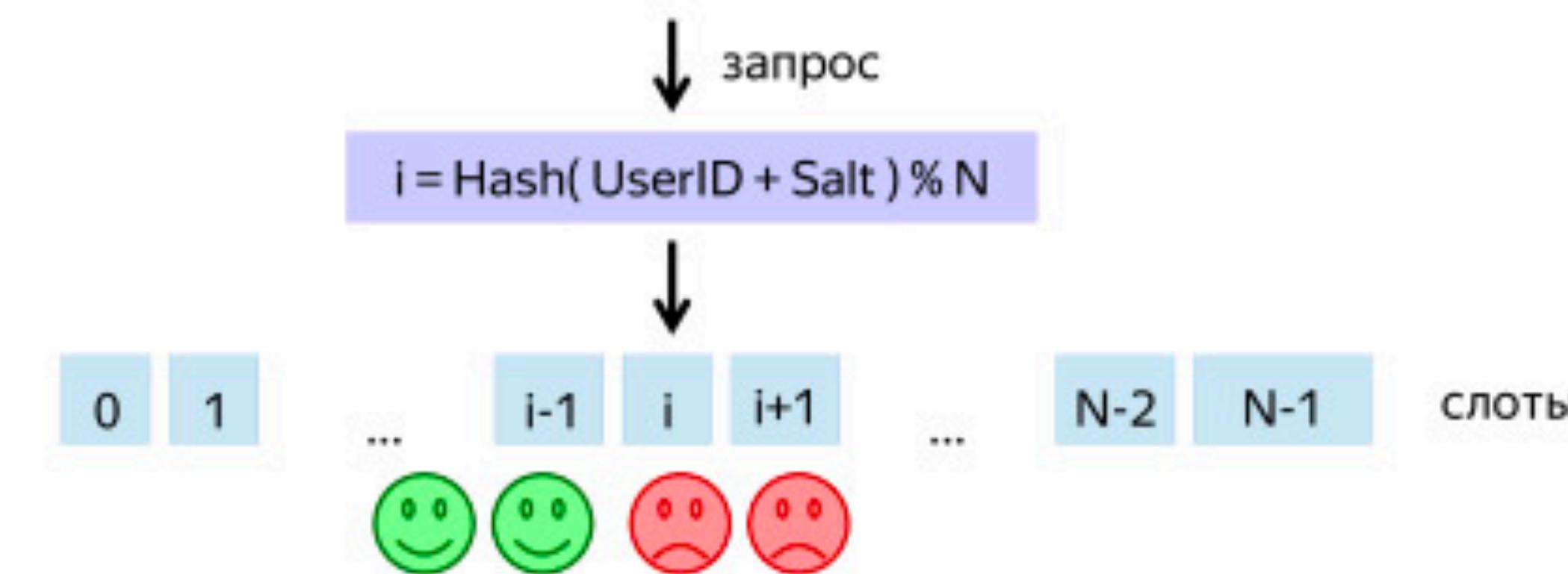
Использовать, когда данных много



Зачем нужна соль?

Вопрос аудитории

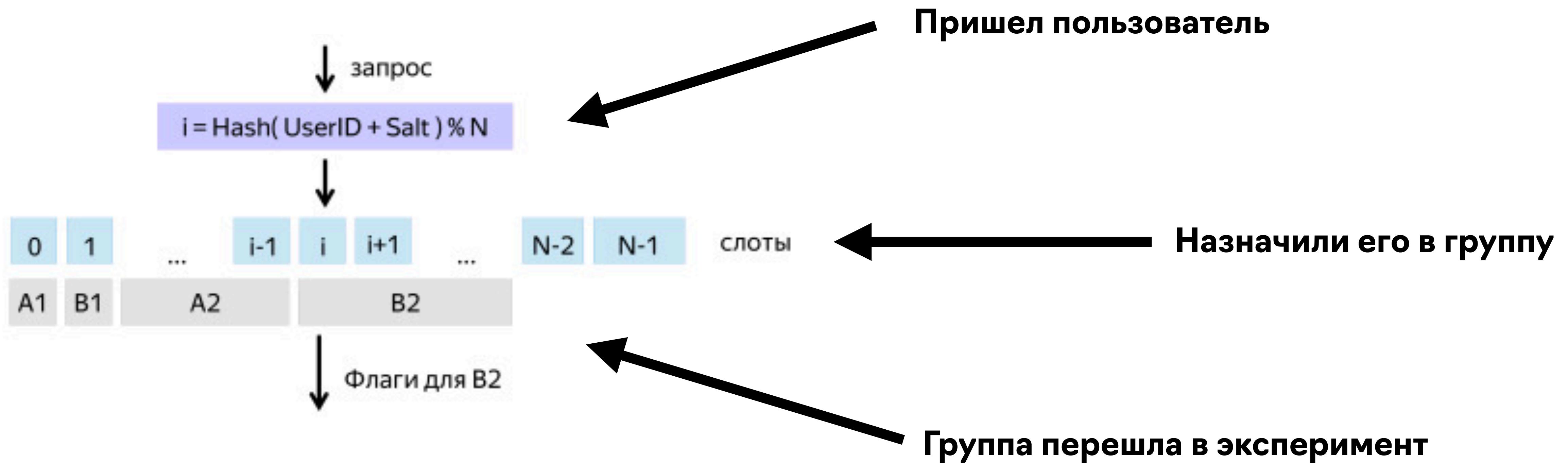
Проблема памяти пользователей



Даже после завершения эксперимента пользователи ведут себя по-разному

Хеширование с солью

Использовать, когда данных много



Как думаете на малых данных такое заработает?

Вопрос аудитории

Жадный отбор

Жадный отбор - противоположность случайному отбору, поочереди выбираем из оставшихся юнитов для добавления в группу А или В, оптимизируя функцию потерь

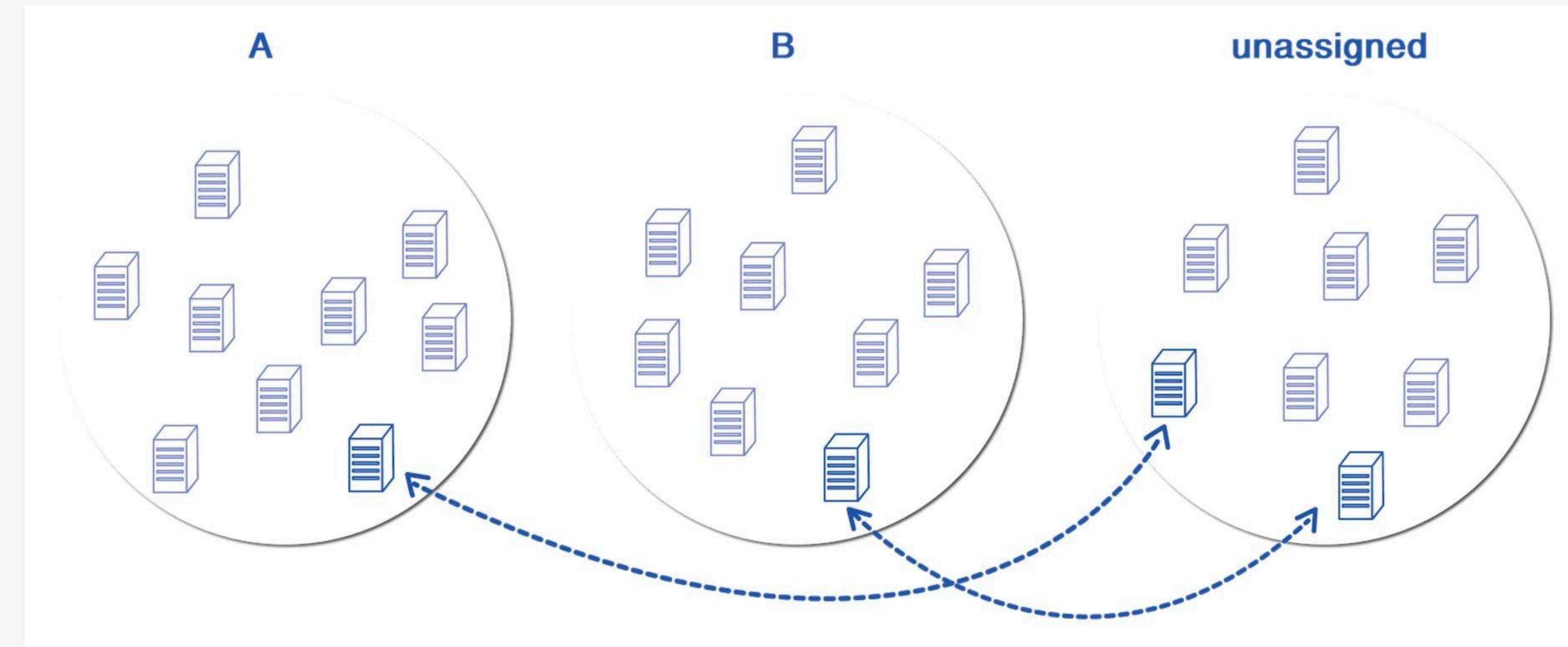
Как эта функция потерь выглядит?

$$\text{cost}(A, B) = \sum_{m \in M, d \in D} \lambda_m (X_{m,d}^A - X_{m,d}^B)$$

- М - набор метрик
- D - набор исторических дат
- $X_{m,d}$ - значение какой-то метрики m в конкретный день d
- λ_m - вес метрики m малое

Жадный отбор

Жадный отбор - противоположность случайному отбору, поочереди выбираем из оставшихся юнитов для добавления в группу A или B, оптимизируя функцию потерь



Пример: площадка объявлений

Мы работаем все в той же площадке объявлений. Теперь мы хотим провести А/Б тест, который влияет на наши алгоритмы монетизации. Ограничение этого теста в том, что люди в одном городе общаются и мы не можем делать разбиение внутри одного города, поэтому единица рандомизации у нас это город



Губозакаточная
машинка

999 ₽



...

Пример: площадка объявлений

Мы работаем все в той же площадке объявлений. Теперь мы хотим провести А/Б тест, который влияет на наши алгоритмы монетизации. Ограничение этого теста в том, что люди в одном городе общаются и мы не можем делать разбиение внутри одного города, поэтому единица рандомизации у нас это город

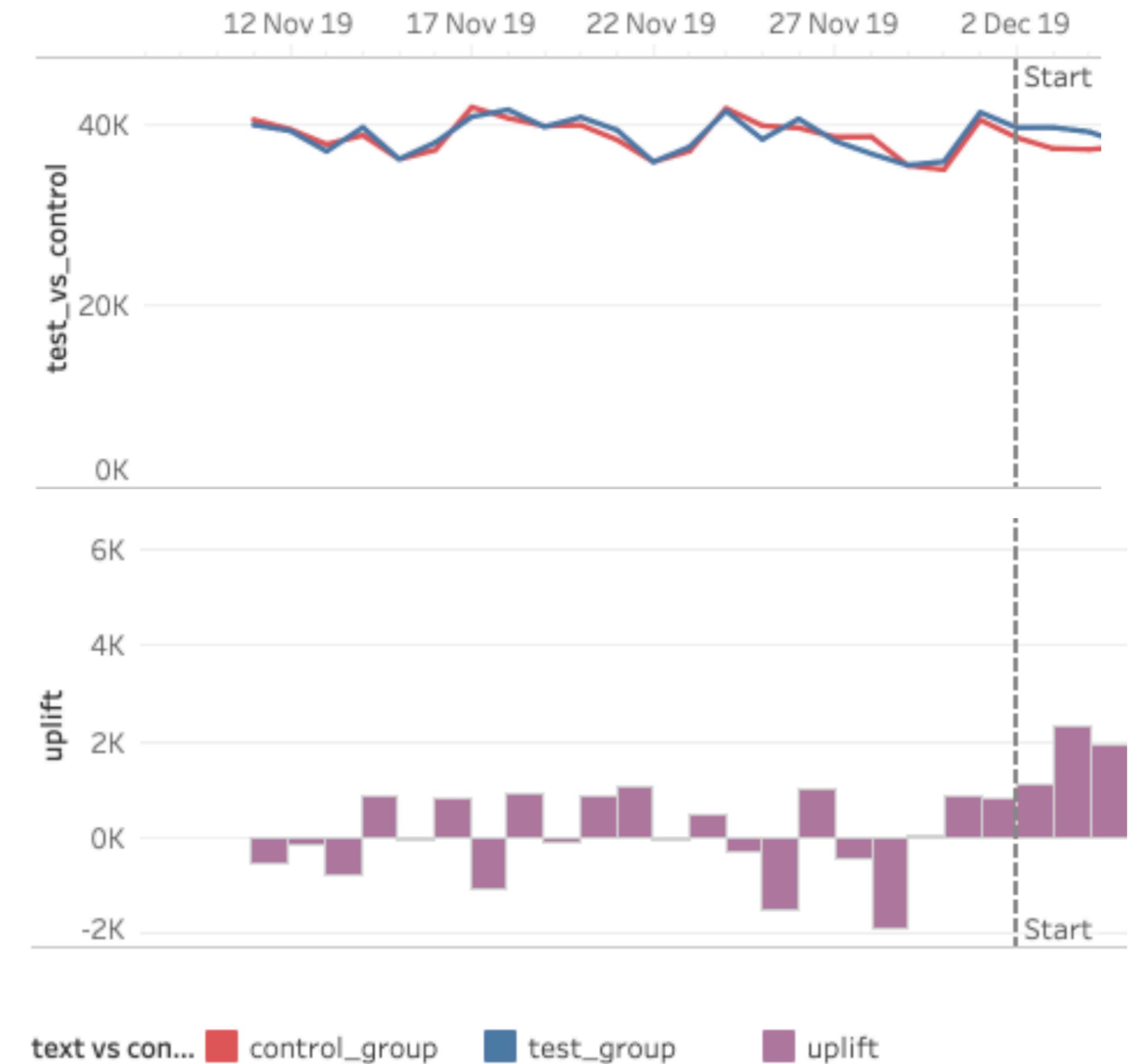
Для простоты выберем одну метрику за которой будем следить - выручка



Губозакаточная
машинка
999 ₽

Что значит, что выручка похожа?

1. Схожая динамика тестовой и контрольной групп
2. Маленькое日内ное отклонение тест vs
контроль
3. Постоянное отношение значений метрики тест
vs контроль



Какой критерий сюда подойдет?

Вопрос аудитории

Жадный отбор на примере площадки с объявлениями

$$\text{cost}(A, B) = \sum_{m \in M, d \in D} \lambda_m (X_{m,d}^A - X_{m,d}^B)$$

- M - набор метрик, в нашем случае средняя выручка и дисперсия выручки
- D - набор исторических дат, в нашем случае
- $X_{m,d}$ - значение какой-то метрики m в конкретный день d
- λ_m - вес метрики m малое

Алгоритм жадного отбора

Изначально группы A и B
случайно
инициализированы

Возвращаем группы

Находим swap (A, unassigned)
или (B, unassigned), которая
максимально снижает cost
функцию

Есть такой swap?

Да

Нет

Как бы улучшили жадный алгоритм?

Вопрос аудитории

Улучшенный алгоритм жадного отбора

Возможные варианты:

1

Делаем swap с
несколькими
объектами

2

Меняем объекты
между группами

3

Другие алгоритмы
оптимизации

Изначально группы A и B
случайно
инициализированы

Возвращаем группы

Да

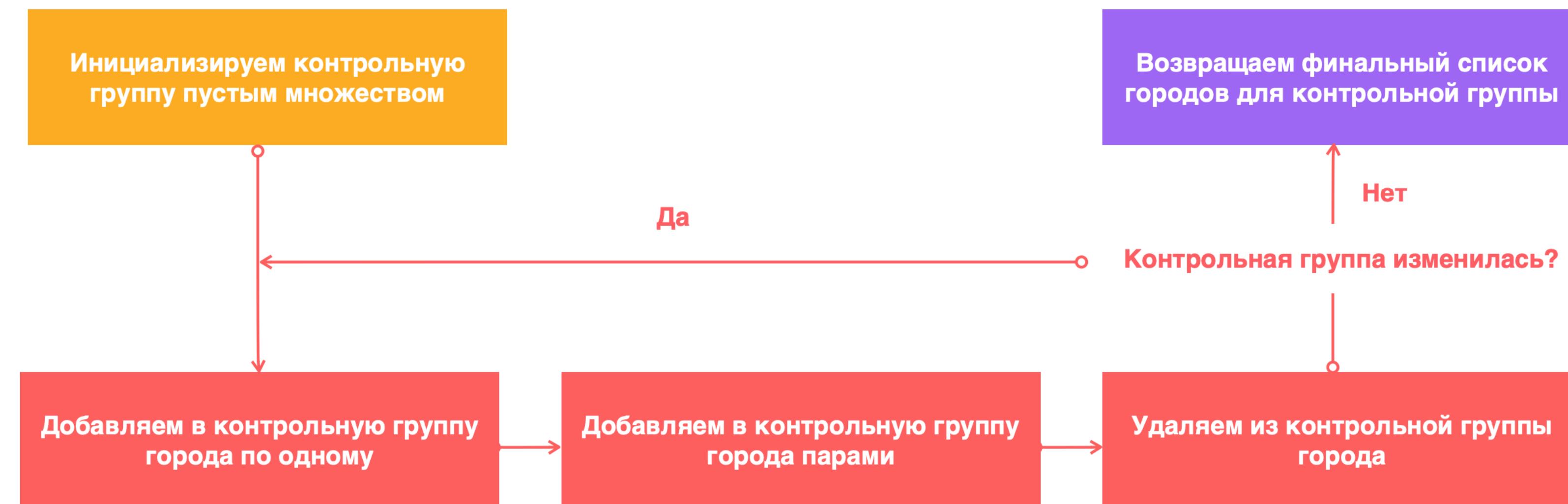
Нет

Находим swap (A, unassigned)
или (B, unassigned), которая
максимально снижает cost
функцию

Есть такой swap?

Пример: площадка объявлений

Схема работы алгоритма формирования контрольной группы



На каждом шаге алгоритма контрольная группа последовательно “приближается” к тестовой.



Пример: площадка объявлений

Мы работаем все в той же площадке объявлений. Теперь мы хотим провести А/Б тест, который влияет на наши алгоритмы монетизации. Ограничение этого теста в том, что люди в одном городе общаются и мы не можем делать разбиение внутри одного города, поэтому единица randomизации у нас это город.

Однако у нас появилось ограничение - мы хотим, чтобы наше изменение повлияло на всех одновременно, что не создавать преимущества большим продавцам

Когда закончил институт и выложил свое первое резюме:



Собака просит о помощи

Бесплатно

- Площадь Маркса ⚑ 21–30 мин.

Пример: площадка объявлений

Мы работаем все в той же площадке объявлений. Теперь мы хотим провести А/Б тест, который влияет на наши алгоритмы монетизации. Ограничение этого теста в том, что люди в одном городе общаются и мы не можем делать разбиение внутри одного города, поэтому единица рандомизации у нас это город.

Однако у нас появилось ограничение - мы хотим, чтобы наше изменение повлияло на всех одновременно, что не создавать преимущества большим продавцам

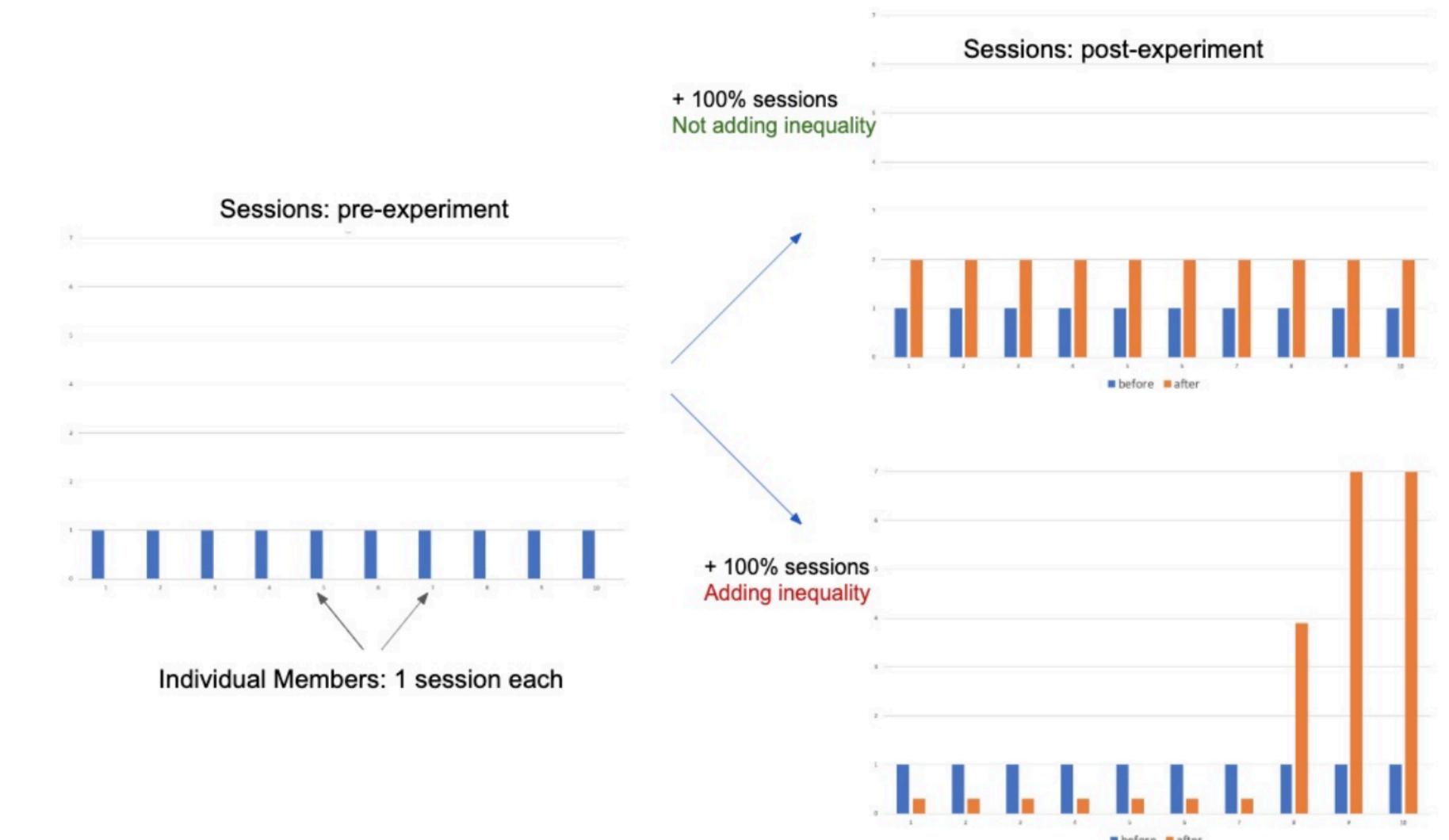


Figure 1: Two products with the same average impact, but different individual effect patterns.

Что сделать с метрикой, чтобы такое учесть?

Вопрос аудитории

Индекс Аткинсона

Индекс Аткинсона - экономический индекс, используемый для сравнения неравенства в доходах между странами

$$A_\epsilon^C(X^C) = 1 - \frac{\left(\frac{1}{n} \sum_i^n x_i^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}}}{\sum_i^n x_i}$$

Индекс Аткинсон
контрольная
группа

$$A_\epsilon^T(X^T) = 1 - \frac{\left(\frac{1}{n} \sum_i^n x_i^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}}}{\sum_i^n x_i}$$

Индекс Аткинсон
группа
эксперимента

Что про него надо знать: если зафиксировать значение ϵ , то индекс Аткинсона маленький, если в выборке значения метрик похожи, а ϵ – это уровень неприязни к неравенству.

Жадный отбор на примере площадки с объявлениями

$$\text{cost}(A, B) = \sum_{m \in M, d \in D} \lambda_m (X_{m,d}^A - X_{m,d}^B)$$

- M - набор метрик, в нашем случае средняя выручка и дисперсия выручки + **добавляем разницу между индексами Аткинсона**
- D - набор исторических дат, в нашем случае
- $X_{m,d}$ - значение какой-то метрики m в конкретный день d
- λ_m - вес метрики m малое

Итого

1. Разобрались, что делать с метриками, которые долго мерять - подбирать прокси метрики
2. Разобрались как эти прокси метрики искать и как проверять, что они норм
3. Разбрали несколько способов разбиения на группы