

Организация массового A/B-тестирования

# Сплиттинг и стратификация

# Цель урока

Давайте теперь поговорим, как разбивать пользователей на группы для проведения тестирования.

# Задачи урока

- ✓ Основные характеристики группы пользователей (сегменты клиентов по разным признакам)
- ✓ Требования к контрольной и экспериментальной группам пользователей
- ✓ Однородность выборки (пример)
- ✓ Выбор нужного количества пользователей из контрольной и экспериментальной групп по ID
- ✓ Разделение группы пользователей на контрольную и экспериментальную группы (процентное соотношение)

# Разделение на группы: кейс

Необходимо разбить пользователей на две группы.  
В одну группу распределяем тех, кто делает  
в среднем более 5 покупок в месяц, в другую —  
оставшихся. Что мы делаем не так?

# Разделение на группы: кейс

Необходимо разбить пользователей на две группы. В одну группу распределяем тех, кто делает в среднем более 5 покупок в месяц, в другую — оставшихся. Что мы делаем не так?

Ответ: группы получились неоднородными, так как, возможно, не функционал будет влиять на лояльность, а неравномерность в поведении.

# Варианты разделения: случайно

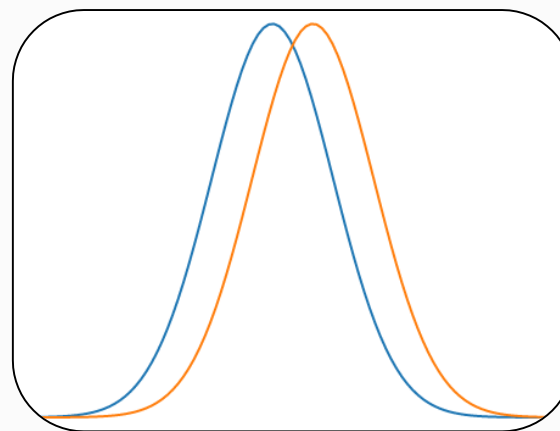


# Варианты разделения: случайно



Необходимо убедиться, что после случайного разделения группы стали однородными. Для этого можно сравнить средние значения метрик: должны примерно совпадать.

Рандом иногда выдаёт неоднородные группы. Чаще это происходит, когда метрик сравнения более трёх.



Пример, где средние двух групп не совпадают — повторяем разбиение.

А что если после разбиения  
среднее двух групп совпало,  
но в одной группе оказались  
только мужчины? Есть  
ли тут однородность?

A



B





# Варианты разделения: стратифицировано случайно

- 1 Делим пользователей на группы по какому-то признаку — страты. **Набираем** пользователей из каждой страты рандомно и пропорционально размеру страты
- 2 Ранее выбранных пользователей убираем из каждой страты. **Набираем** нужное количество из оставшихся пользователей для другой группы

# Варианты разделения: стратифицировано случайно

## Кейс

Проводим тестирование на  $\frac{1}{2}$  пользователей.

Имеем в каждой группе по  $\frac{1}{4}$  всех пользователей.

Страта	< 1 мес.	1–6 мес.	6–12 мес.	> 12 мес.
Размер	500	1 000	1 500	2 000
Размер	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
Сколько берём из каждой страты	125	250	375	500

# Стратифицировано случайно: кейс

Происходит размытие эффекта функционала, так как не учитываются все характеристики пользователей, проверка на однородность не помогает.

- **Генеральная совокупность:**  
покупатели в аптеке  
из Москвы
- **Разделение на группы:**  
случайно
- **Изменение:** новый дизайн
- **Первичные метрики:**  
среднее количество покупок
- **Нюанс:** при тестировании  
в Москве были  
магнитные бури
- **Проблема:** покупатели  
в какой-то группе более  
восприимчивые  
к погоде, соответственно,  
совершили больше  
покупок в аптеке

# Сегментация

## Зачем?

- 1 Разные сегменты могут реагировать на изменения по-разному
- 2 Помогает обнаружить баги для каких-то сегментов



Когда выбирать сегменты  
**до** или **после** проведения теста?

- До — если предполагается, где возникнет разница
- После — если не предполагается разницы

# Плюсы выбора сегментов заранее

- ✓ Минимизация вероятности ошибки первого рода

Плохо искать различия по большому числу сегментов после тестирования без поправок на множественное тестирование

А поправки уменьшают мощность

- ✓ Расчёт необходимого размера выборок

В сегментах может быть недостаточно пользователей, чтобы увидеть MDE

**Применяйте сегментирование, если это можно объяснить.**

Организация массового A/B-тестирования

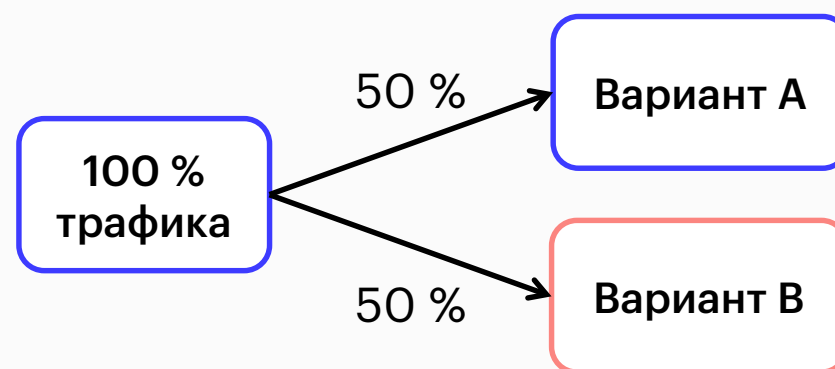
# СПЛИТИНГ

# Разделение на группы: кейс

Вы правильно разделили пользователей на однородные группы. С точки зрения тестирования всё вообще сделали правильно.

Но тестируемая гипотеза оказалась негативной и все пользователи, попавшие в тестовые группы, перестали пользоваться вашим сервисом. Как итог, доход компании упал в два раза.

## А/В-тестирование



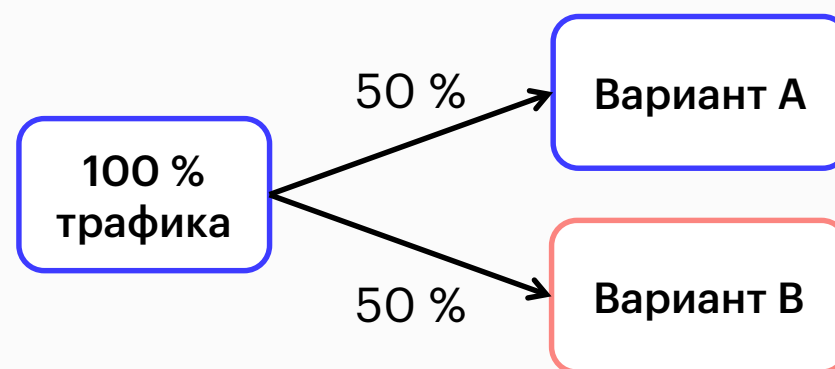
# Разделение на группы: кейс

Вы правильно разделили пользователей на однородные группы. С точки зрения тестирования всё вообще сделали правильно.

Но тестируемая гипотеза оказалась негативной и все пользователи, попавшие в тестовые группы, перестали пользоваться вашим сервисом. Как итог, доход компании упал в два раза.

**Как такого можно было избежать?**

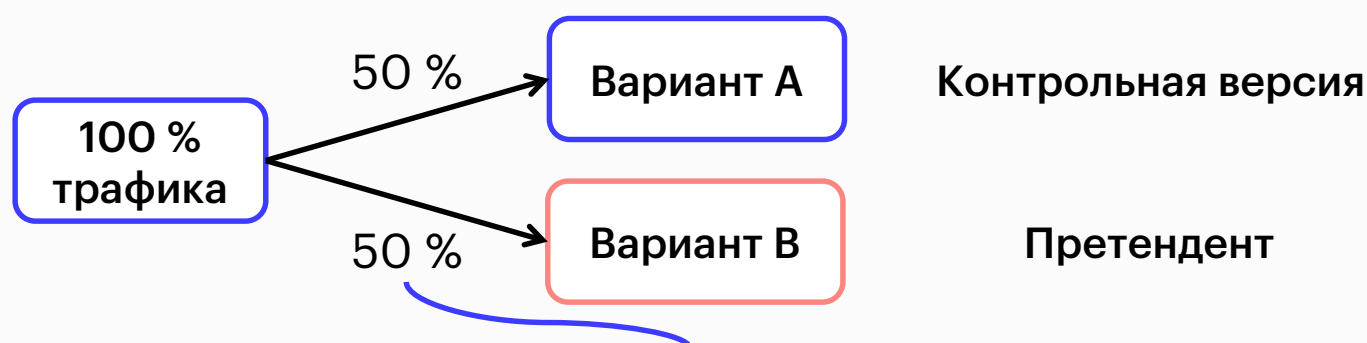
## А/В-тестирование





# Разделение на группы: кейс

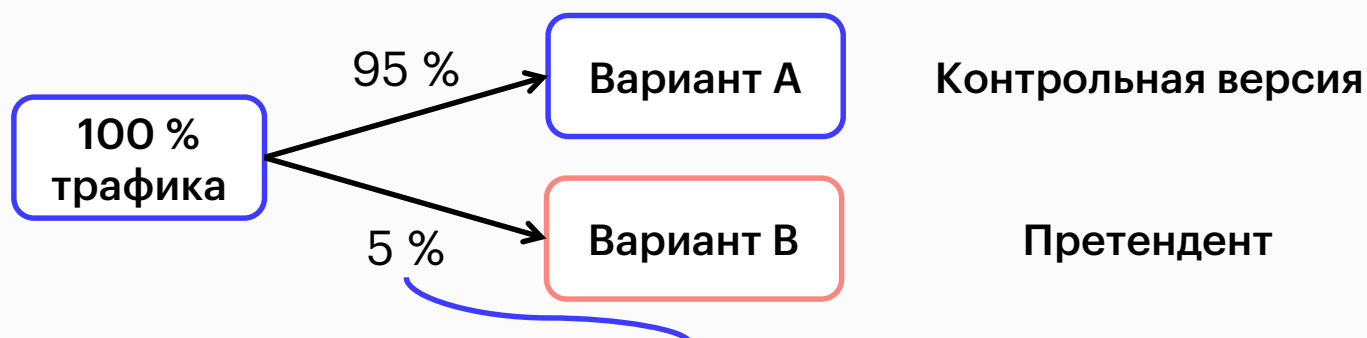
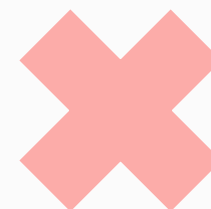
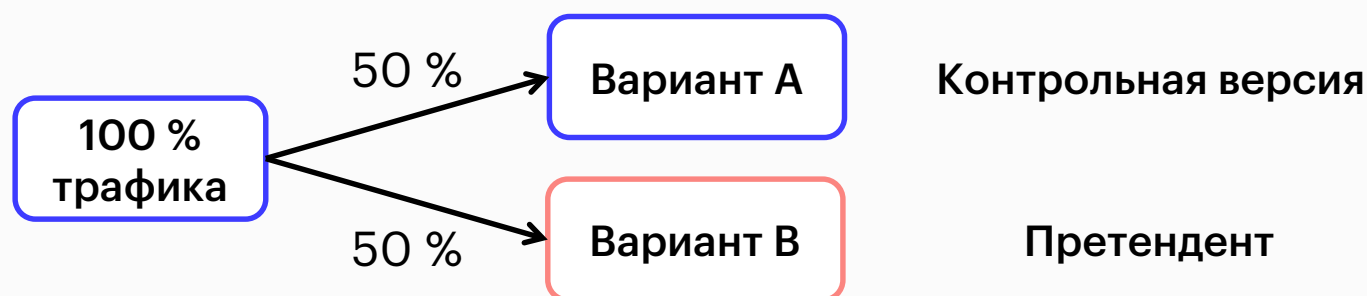
## A/B-тестирование



Так выглядит обычное разбиение 50/50. В случае провала проверяемой гипотезы затрагивается большая часть пользователей.

# Разделение на группы: кейс

## А/В-тестирование



Чтобы минимизировать риски, отводите для теста небольшой процент пользователей, например, 5 %. Так вы и гипотезу проверите, и риски будут не так велики.

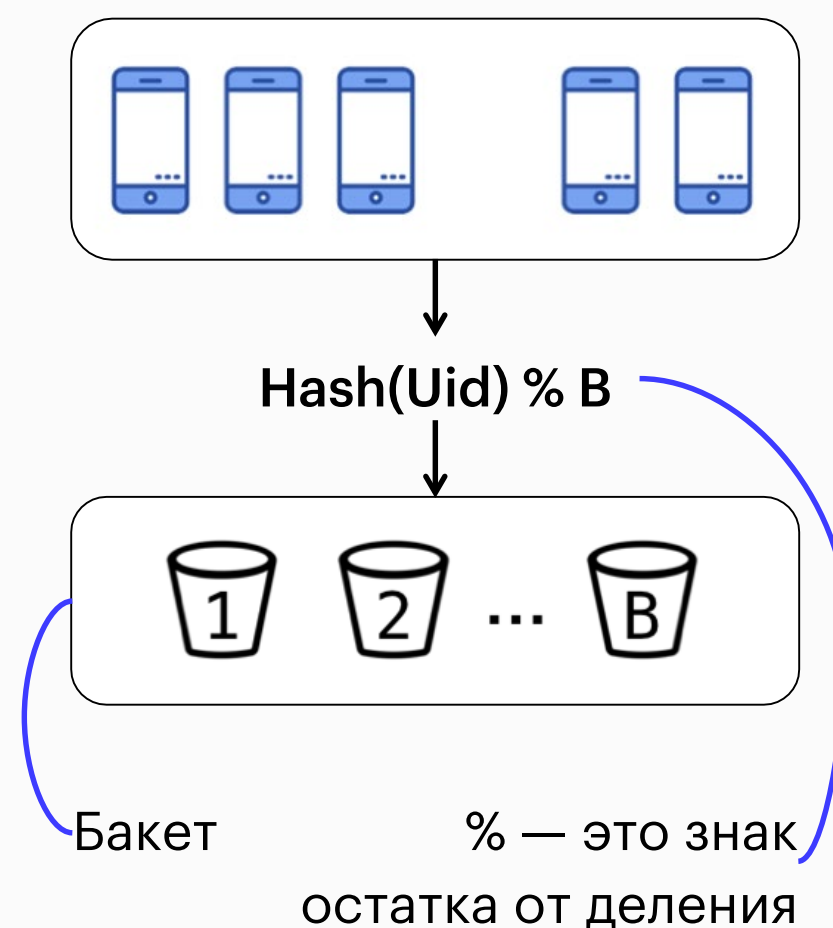
Организация массового A/B-тестирования

# Сплиттинг на много групп

## Сплиттинг на много групп

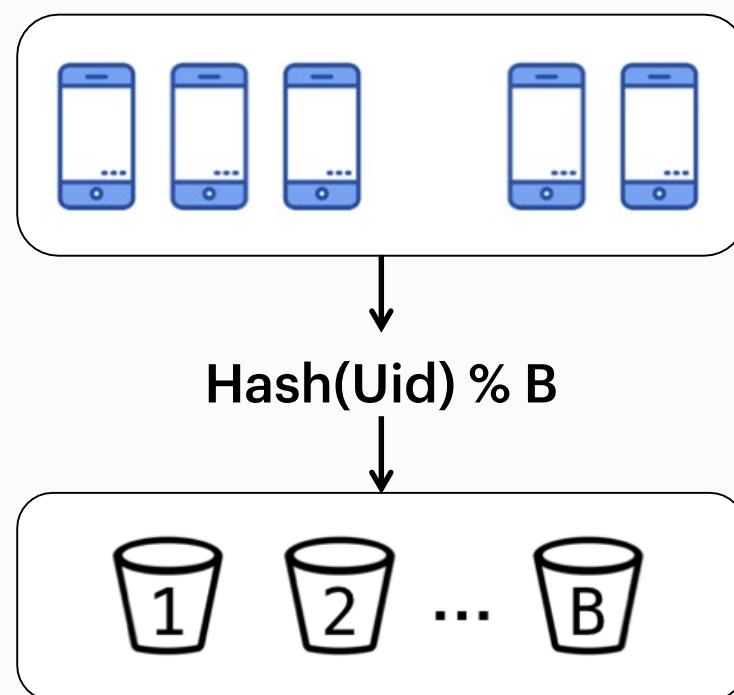
В случае когда данных у вас много, а делить на группы данные не хочется, можно вспомнить знакомый нам метод бакетинга. Обратим внимание на детали этого разбиения:

Идея проста: хешируем Uid'ы пользователей и по остатку от деления «разбрасываем» их на некоторое количество бакетов (обозначим их число за **B**):



### Преимущества:

- бакеты получаются быстро
- размер данных может быть сжат, если использовать среднее значение метрики по всему бакету
- если бакеты формировать при занесении нового пользователя в базу, то время разделения будет ещё меньше — надо будет выбрать только свободные бакеты
- пересечение пользователей между экспериментами контролируется



# Итоги урока

- ✓ Правильная сегментация пользователей даст вам больше уверенности в том, что различия в группах вызваны именно изменениями, а не из-за неравномерного разделения на группы
- ✓ Разное процентное соотношение между группами А и В снижает риски последствий от неудачных гипотез
- ✓ Разбиение на группы (бакеты) даёт выигрыш по времени, а также возможность переиспользовать одни и те же бакеты в разных экспериментах