

Элементы

статистики для построения
гипотез A/B-тестирования

Артур Сапрыкин

Data Scientist
Freelancer

Цель модуля

Понять, как создавать гипотезы A/B-тестирования и применять базовые элементы статистики к ним.

Что вы будете знать по итогам изучения модуля?

- ✓ Как выглядят данные о продукте
- ✓ Разницу между количественными и качественными данными
- ✓ Меры центральной тенденции для первичной оценки данных
- ✓ Поймёте, какими способами можно визуализировать и сравнивать распределение данных
- ✓ Как формулировать гипотезу, которая будет проверяться в A/B-тесте
- ✓ Что такое p-value и доверительный интервал и как их можно применять

Элементы статистики для построения гипотез A/B-тестирования

**Данные для проведения
A/B-тестирования
и определение выборки**

Данные для проведения A/B-тестирования и определение выборки

Цель урока

Понять, как собирать данные
о продукте и правильно
создавать выборку.

Данные для проведения A/B-тестирования и определение выборки

Задачи урока

- ✓ Узнать, какие данные о продукте нужны для его оценки
- ✓ Познакомиться с определением качественных и количественных данных
- ✓ Узнать, что такое генеральная совокупность и выборка

Данные для проведения А/В-тестирования и определение выборки

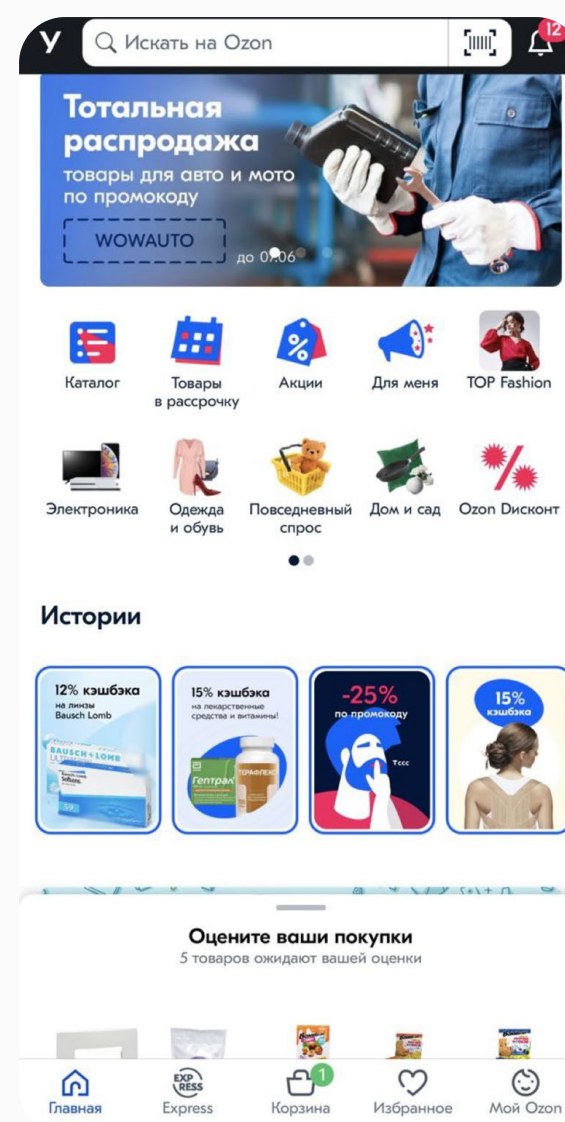
Что такое данные о продукте и где их взять?

Что из себя представляют данные о продукте?

Это информация, которую мы можем получить, когда с продуктом взаимодействует пользователь.

Например, рассмотрим приложение OZON.

OZON — это продукт, а все действия, которые совершает пользователь с этим продуктом, и есть данные.

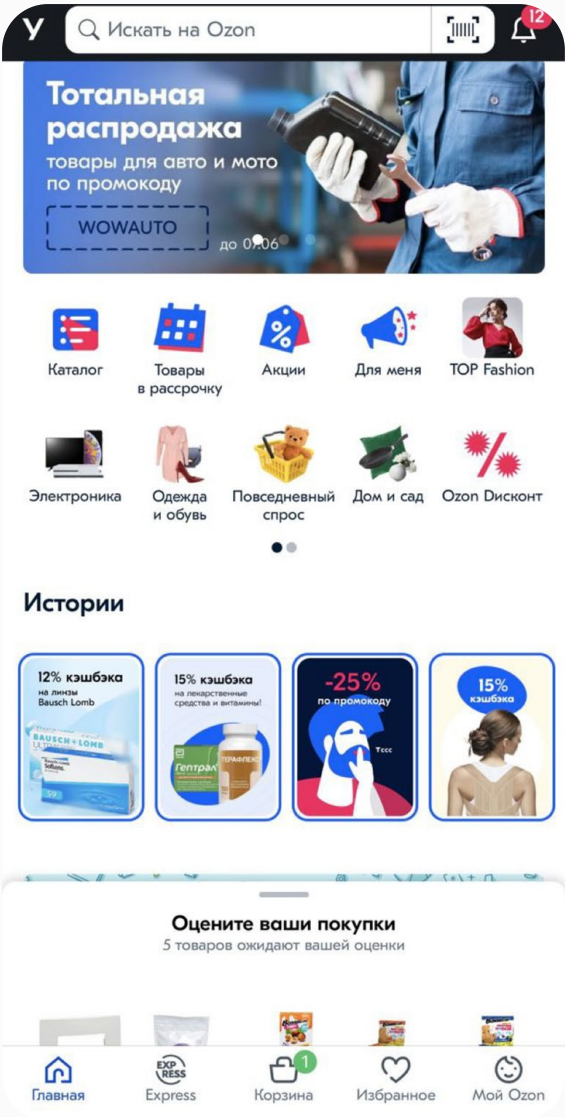


Данные для проведения А/В-тестирования и определение выборки

Что такое данные о продукте и где их взять?

Выглядят они примерно так:

Вход
Иванов зашёл в каталог в 8:00
Петров открыл истории в 9:00
Сидоров открыл акции в 9:30
Смирнов оценил покупку в 11:00
Иванов вышел из каталога в 12:05
Попов закрыл приложение в 13:07
Соколов выбрал уютг в 16:14
Волков зашёл в избранное в 17:40



Что такое данные о продукте и где их взять?

Выглядят они примерно так:

Вход

Иванов зашёл в каталог в 8:00

Петров открыл истории в 9:00

Сидоров открыл акции в 9:30

Смирнов оценил покупку в 11:00

Иванов вышел из каталога в 12:05

Попов закрыл приложение в 13:07

Соколов выбрал уют в 16:14

Волков зашёл в избранное в 17:40

То есть мы можем посчитать:

- сколько раз пользователь зашёл в «Каталог»
- или в «Истории»
- сколько товаров добавил в «Корзину»
- добавление товара в «Избранное»
- и любые другие действия пользователя в приложении

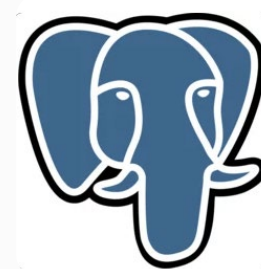
Где взять данные?

Как правило, **все данные о действиях пользователя логируются и находятся в базе данных (БД).**

При этом есть минимум информации о пользователях, необходимый для описания продукта: id пользователя | дата действия | название действия (клик по кнопке, добавление и т. д.) | в какой части продукта происходит действие («Каталог», «Истории» и т. д.).

Как реализован сбор данных о действиях пользователей в БД, зависит от компании, в которой вы работаете.

Это может быть БД, развёрнутая на Postgres, Oracle, MySQL или других сервисах.



ORACLE



Данные для проведения А/В-тестирования и определение выборки

Иные источники данных

webflow



Google Analytics



Яндекс метрика



Данные для проведения А/В-тестирования и определение выборки

Как собрать все данные вместе?

Сбор данных состоит из трёх стадий:

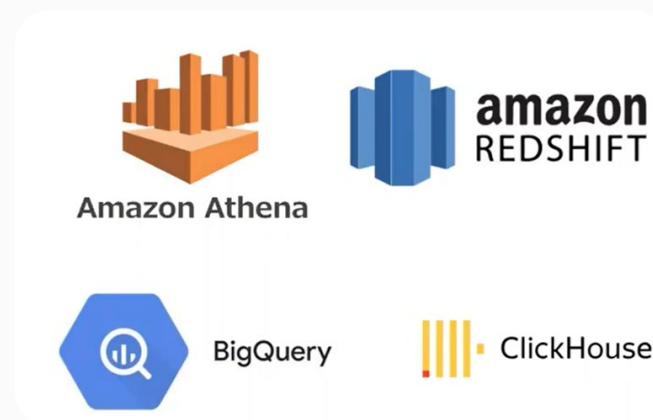
Export → Transform → Load

Выгрузка → Преобразование → Загрузка

Существует уже достаточно много сервисов для ETL-цикла.

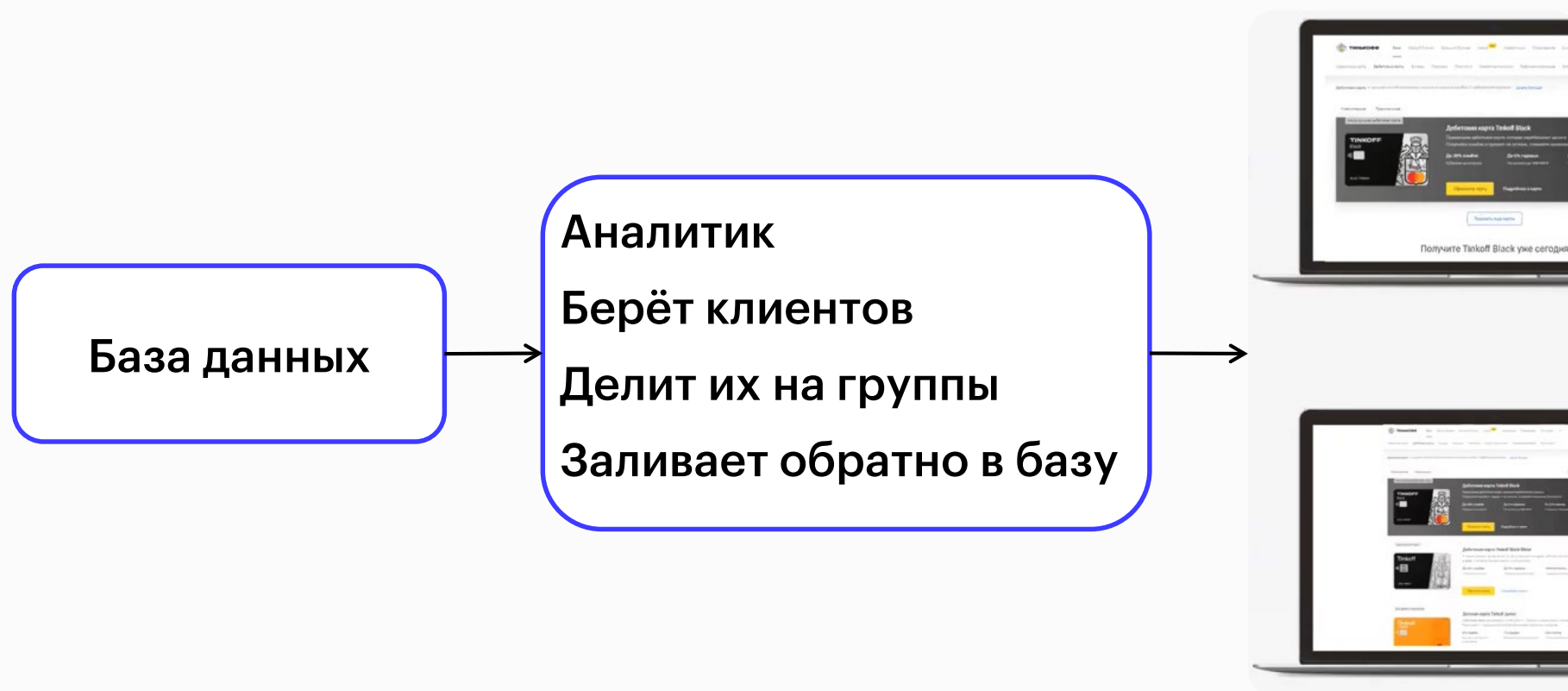
В разных компаниях от аналитика требуется разная степень погружённости в data engineering. Поэтому в штате обычно присутствует специальный человек — **Data Engineer**.

На практике сложно достичь идеального ETL-цикла, постоянно что-то ломается.



Данные для проведения А/В-тестирования и определение выборки

Где делят данные?



Количественные и качественные данные. Чем отличаются?

- ✓ **Количественные данные** — всё, что можно посчитать или пронаблюдать

Например, количество действий пользователя:
клики по кнопке, входы в приложение,
количество просмотров страниц сайта и т. д.

Количественные данные собираются путём
взаимодействия пользователя с продуктом

Количественные и качественные данные. Чем отличаются?

- ✓ **Качественные данные** — это субъективная информация, которую нельзя измерить цифрами

Например: насколько какая-либо функция удобна в использовании. Вкус, цвет, архитектурный стиль, семейное положение — всё это типы качественных данных

Качественные данные собираются из интервью с пользователями, звонков, почтовой переписки

Какие данные нужно использовать для того, чтобы провести А/В-тестирование?

Количественные данные считаются более точными. Они измеримы и однозначны, поэтому мы будем их использовать для статистической оценки А/В-тестирования.

Качественные данные, напротив, недостоверны, то есть субъективны, и отражают личную точку зрения отдельных людей.

Но при этом качественные данные важны для нас после проведения исследования.

Задача и проблема A/B-тестирования

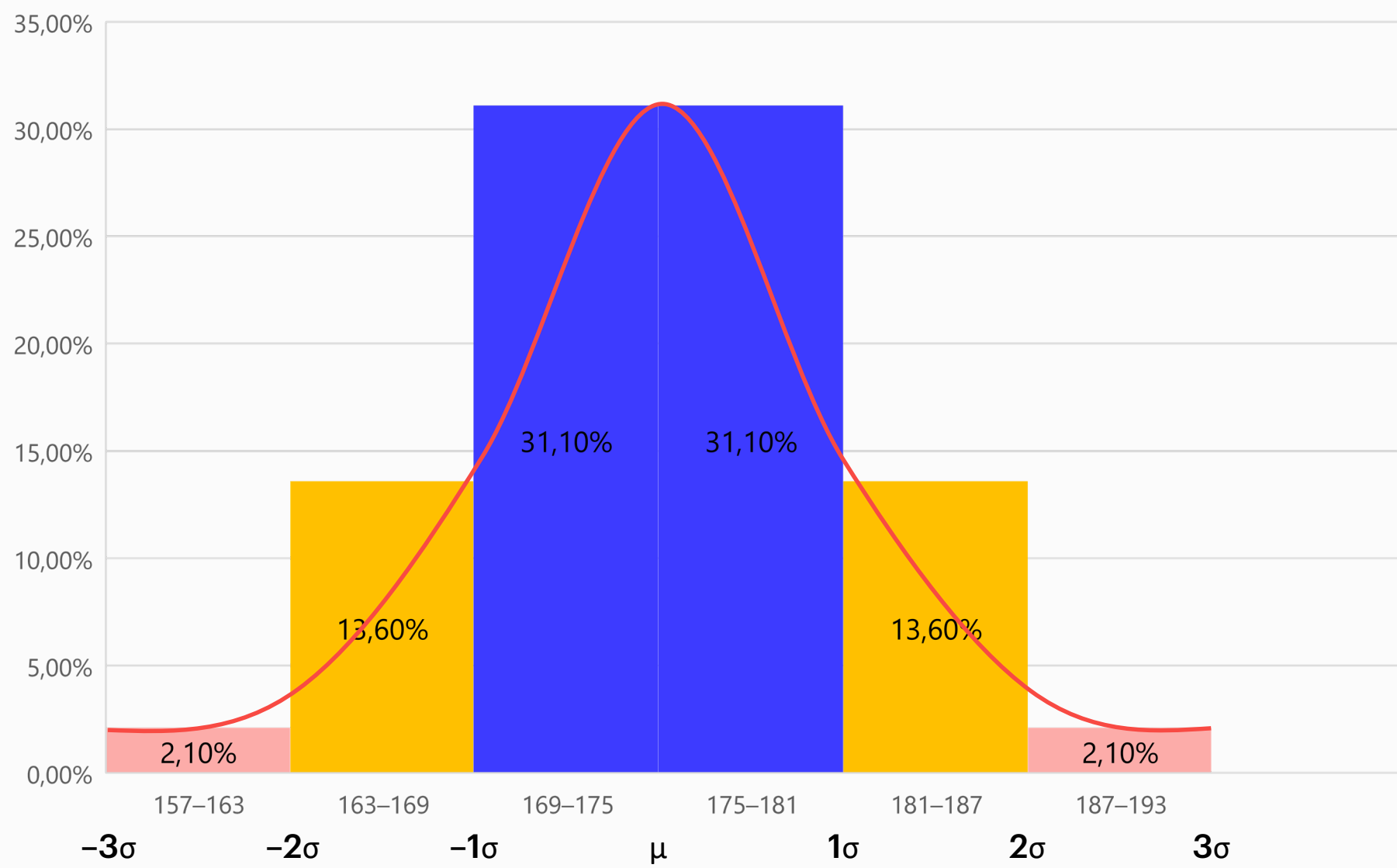
- ✓ **Задача:** сделать вывод о большой совокупности клиентов
- ✓ **Проблема:** невозможно собрать всех существующих в мире клиентов и поделить пополам, чтобы протестировать новую фичу. Это потребует больших финансовых затрат и т. п.

Генеральная совокупность и выборка

- ✓ **Генеральная совокупность** — множество объектов, относительно которых мы хотели бы делать выводы в рамках исследования некоторой проблемы
- ✓ **Выборка** — часть генеральной совокупности
- ✓ **Задача:** сделать вывод о большой совокупности клиентов
- ✓ **Проблема:** невозможно собрать всех существующих в мире клиентов и поделить пополам, чтобы протестировать новую фичу. Это потребует больших финансовых затрат и т. п.

Данные для проведения А/В-тестирования и определение выборки

Нормальное распределение по Гауссу



Генеральная совокупность и выборка

- ✓ **Решение:** от общего числа клиентов, которое называется **генеральной совокупностью**, отсекаем часть, которая называется **выборкой**



Выбор генеральной совокупности

Группы набираем из того сегмента, на который предположительно повлияет изменение, иначе будет размытие эффекта.

Данные для проведения A/B-тестирования и определение выборки

Плохой выбор генеральной совокупности

Хотим оценить эффект от новой фичи
для пользователей, которые пользуются
приложением менее 1 года.

Берём всех пользователей и получаем вывод,
что эффекта не наблюдается. Почему
такое может быть?

Данные для проведения A/B-тестирования и определение выборки

Плохой выбор генеральной совокупности

Хотим оценить эффект от новой фичи для пользователей, которые пользуются приложением менее 1 года.

Берём всех пользователей и получаем вывод, что эффекта не наблюдается. Почему такое может быть?

Ответ: выбрали неподходящую генеральную совокупность. Если фича влияет не на всех пользователей, то не надо смотреть метрику по всем.

Требования к выборке: репрезентативность

- ✓ Выборка должна отражать свойства генеральной совокупности
- ✓ При любой сегментации выборки её распределение по сегментам должно быть похоже на распределение генеральной совокупности



Итоги урока



Данные о продукте берутся из взаимодействия пользователя с продуктом



Данные для проведения А/В-тестирования можно взять из БД, в которой имеется основная информация по действиям пользователей



Для А/В-тестирования в первую очередь используются количественные данные



Залогом успеха в тестировании является правильный выбор генеральной совокупности и выборки