# **A/B-тестирование**

### **Постановка задачи**

Ваша задача — провести оценку результатов A/B-теста. В вашем распоряжении есть датасет с действиями пользователей, техническое задание и несколько вспомогательных датасетов.

* Оцените корректность проведения теста. Проверьте:
  + Соответствие данных требованиям технического задания. В случае нарушения требований оцените, насколько существенно негативное влияние на результаты теста. При проверке условий технического задания и фильтрации данных, контролируйте количество уникальных пользователей, принимающих участие в тесте – это поможет детальнее изучить влияние каждого условия на аудиторию теста.
  + Время проведения теста. Убедитесь, что оно не совпадает с маркетинговыми и другими активностями.
  + Аудиторию теста. Удостоверьтесь, что нет пересечений с конкурирующим тестом и нет пользователей, участвующих в двух группах теста одновременно. Проверьте равномерность распределения пользователей по группам и правильность их формирования.
* Проанализируйте результаты теста

### **Техническое задание**

* Название теста: recommender\_system\_test;
* группы: А — контрольная, B — новая платёжная воронка;
* дата запуска: 2020-12-07;
* дата остановки набора новых пользователей: 2020-12-21;
* дата остановки: 2021-01-04;
* аудитория: 15% новых пользователей из региона EU;
* назначение теста: тестирование изменений, связанных с внедрением улучшенной рекомендательной системы;
* ожидаемое количество участников теста: 6000.
* ожидаемый эффект: за 14 дней с момента регистрации пользователи покажут улучшение каждой метрики не менее, чем на 10%:
  + конверсии в просмотр карточек товаров — событие product\_page,
  + просмотры корзины — product\_cart,
  + покупки — purchase.

### **Данные**

[final\_ab\_events.csv](https://disk.yandex.ru/d/pgKgZ9lRp0Enwg)

[ab\_project\_marketing\_events.csv](https://disk.yandex.ru/d/ZVC7GUYYckKnGQ)

[final\_ab\_new\_users.csv](https://disk.yandex.ru/d/Qrto8Gerpu424g)

[final\_ab\_participants.csv](https://disk.yandex.ru/d/nhGUR3ZkNIG4VA)

### **Описание данных**

ab\_project\_marketing\_events.csv — календарь маркетинговых событий на 2020 год.

Структура файла:

* name — название маркетингового события;
* regions — регионы, в которых будет проводиться рекламная кампания;
* start\_dt — дата начала кампании;
* finish\_dt — дата завершения кампании.

final\_ab\_new\_users.csv — пользователи, зарегистрировавшиеся с 7 по 21 декабря 2020 года.

Структура файла:

* user\_id — идентификатор пользователя;
* first\_date — дата регистрации;
* region — регион пользователя;
* device — устройство, с которого происходила регистрация.

final\_ab\_events.csv — действия новых пользователей в период с 7 декабря 2020 по 4 января 2021 года.

Структура файла:

* user\_id — идентификатор пользователя;
* event\_dt — дата и время события;
* event\_name — тип события;
* details — дополнительные данные о событии. Например, для покупок, purchase, в этом поле хранится стоимость покупки в долларах.

final\_ab\_participants.csv — таблица участников тестов.

Структура файла:

* user\_id — идентификатор пользователя;
* ab\_test — название теста;
* group — группа пользователя.

### **Как сделать задание?**

* Опишите цели исследования
* Исследуйте данные:
  + Требуется ли преобразование типов?
  + Опишите природу пропущенных значений и дубликатов, если их обнаружите.
* Оцените корректность проведения теста. Обратите внимание на:
  + Соответствие данных требованиям технического задания. Проверьте корректность всех пунктов технического задания.
  + Время проведения теста. Убедитесь, что оно не совпадает с маркетинговыми и другими активностями.
  + Аудиторию теста. Удостоверьтесь, что нет пересечений с конкурирующим тестом и нет пользователей, участвующих в двух группах теста одновременно. Проверьте равномерность распределения по тестовым группам и правильность их формирования.
* Проведите исследовательский анализ данных:
  + Количество событий на пользователя одинаково распределены в выборках?
  + Как число событий в выборках распределено по дням?
  + Как меняется конверсия в воронке в выборках на разных этапах?
  + Какие особенности данных нужно учесть, прежде чем приступать к A/B-тестированию?
* Оцените результаты A/B-тестирования
  + Что можно сказать про результаты A/В-тестирования?
  + Проверьте статистическую разницу долей z-критерием.
* Опишите выводы по этапу исследовательского анализа данных и по проведённой оценке результатов A/B-тестирования. Сделайте общее заключение о корректности проведения теста.