# A/B Testing Project

## Task

You've received an analytical task from an international online store. Your predecessor failed to complete it: they launched an A/B test and then quit (to start a watermelon farm in Brazil). They left only the technical specifications and the test results.

### Technical description

* Test name: recommender\_system\_test
* Groups: А (control), B (new payment funnel)
* Launch date: 2020-12-07
* Date when they stopped taking up new users: 2020-12-21
* End date: 2021-01-01
* Audience: 15% of the new users from the EU region
* Purpose of the test: testing changes related to the introduction of an improved recommendation system
* Expected result: within 14 days of signing up, users will show better conversion into product page views (the product\_page event), instances of adding items to the shopping cart (product\_cart), and purchases (purchase). At each stage of the funnel product\_page → product\_cart → purchase, there will be at least a 10% increase.
* Expected number of test participants: 6000

Download the test data, see whether it was carried out correctly, and analyze the results.

### Description of the data

Download the datasets from [Notion](https://www.notion.so/A-B-Testing-e7e9c322b8d0433e988589103a57521d#c79206a4335f4f07979c2f8ba0f00b23).

To access the datasets on the platform, add /datasets/ to the beginning of the file path (for instance, /datasets/ab\_project\_marketing\_events\_us.csv).

* ab\_project\_marketing\_events\_us.csv — the calendar of marketing events for 2020
* final\_ab\_new\_users\_upd\_us.csv — all users who signed up in the online store from December 7 to 21, 2020
* final\_ab\_events\_upd\_us.csv — all events of the new users within the period from December 7, 2020 through January 1, 2021
* final\_ab\_participants\_upd\_us.csv — table containing test participants

Structure of ab\_project\_\_marketing\_events\_us.csv:

* name — the name of the marketing event
* regions — regions where the ad campaign will be held
* start\_dt — campaign start date
* finish\_dt — campaign end date

Structure of final\_ab\_new\_users\_upd\_us.csv:

* user\_id
* first\_date — sign-up date
* region
* device — device used to sign up

Structure of final\_ab\_events\_upd\_us.csv:

* user\_id
* event\_dt — event date and time
* event\_name — event type name
* details — additional data on the event (for instance, the order total in USD for purchase events)

Structure of final\_ab\_participants\_upd\_us.csv:

* user\_id
* ab\_test — test name
* group — the test group the user belonged to

### Instructions for completing the task

* Describe the goals of the study.
* Explore the data:
  + Do types need to be converted?
  + Are there any missing or duplicate values? If so, how would you characterize them?
* Carry out exploratory data analysis:
  + Study conversion at different stages of the funnel.
  + Is the number of events per user distributed equally among the samples?
  + Are there users who are present in both samples?
  + How is the number of events distributed among days?
  + Are there any peculiarities in the data that you have to take into account before starting the A/B test?
* Evaluate the A/B test results:
  + What can you say about the A/B test results?
  + Use a z-test to check the statistical difference between the proportions.
* Describe your conclusions regarding the EDA stage and the A/B test results.