Task statement

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: A (control), B (new payment funnel)
- Launch date: 2020-12-07
- The 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), product card views (product_card) and purchases (purchase). At each of the stage of the funnel product_page → product_card → 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. @

Data

final_ab_events_us.csv.csv

final_ab_new_users_upd.csv

final_ab_participants_upd.csv

ab_project_marketing_events.csv

- ab_project_marketing_events_us.csv — the calendar of marketing events for 2020
- final_ab_new_users_upd.csv — all users who signed up in the online store from December 7 to 21, 2020
- final_ab_events_upd.csv — all events of the new users within the period from December 7, 2020 to January 1, 2021
- final_ab_participants_upd.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.csv:

- user_id
- first_date — sign-up date
- region
- device — device used to sign up

Structure of final_ab_events_upd.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.csv:

- user_id
- ab_test — test name
- group — the test group the user belonged to

Instructions on completing the task

- Describe the goals of the research

- Explore the data
    - Does it need converting types?
    - Are there any missing or duplicate values? If so, what's their nature?
- Carry out exploratory data analysis
    - Study conversion at different funnel stages
    - Is the number of events per user distributed equally in the samples?
    - Are there users who enter both samples?
    - How is the number of events distributed by days?
    - Think of the possible details 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 tell about the A/A test results?
    - Use the z-criterion to check the statistical difference between the proportions

- Describe the conclusions on the EDA stage, as well as on the evaluation of the A/B test results