A/B test landing page

experiment design

Description

On the current version of the landing page, the user first selects a retailer, goes to the catalog, searches for the goods he needs, and only then can he find out that there may not be a selected store at his address. We want to test the hypothesis that choosing an address before the store selection stage will help to avoid “wrong” scenarios without increasing bounces from the landing itself.

Issues

The experiment is running on all users, not just new ones, which leads to a number of problems:

- old users have already formed their attitude towards us as a product, and this can influence whether they fall off more than a new flow
- we know the addresses and stores where old users shopped, and if they fall into the experimental group and click on the familiar store, we will send them on a UX-meaningless journey through the screens and require extra data

Issues

- The experiment was launched on all types of devices (phones, desktop) and operating systems, although it would be reasonable to assume that users of different devices may be more comfortable following different scenarios.
- Experimental group and control group of different sizes; it would be necessary to take 20% of users, divide them into experiment and control by 10%, follow up on A / A that they do not differ, and then make changes to the experimental group.

Methodology of experiment

We need to understand how the new flow will affect the bounce rate of the landing page and the conversion to actual add to cart. For this experiment, we decided to count the conversion as adding to cart from the session.

For one session, we counted all actions of a user with one anonymous_id within one hour (for example, from 12:00 to 13:00). We assume that in this way we lost a certain number of conversions (if a person entered at 12:30 pm and performed the target action at 13:01 pm, we will not count it), and we assume that this will affect the results of both groups equally and will not affect experiment results.

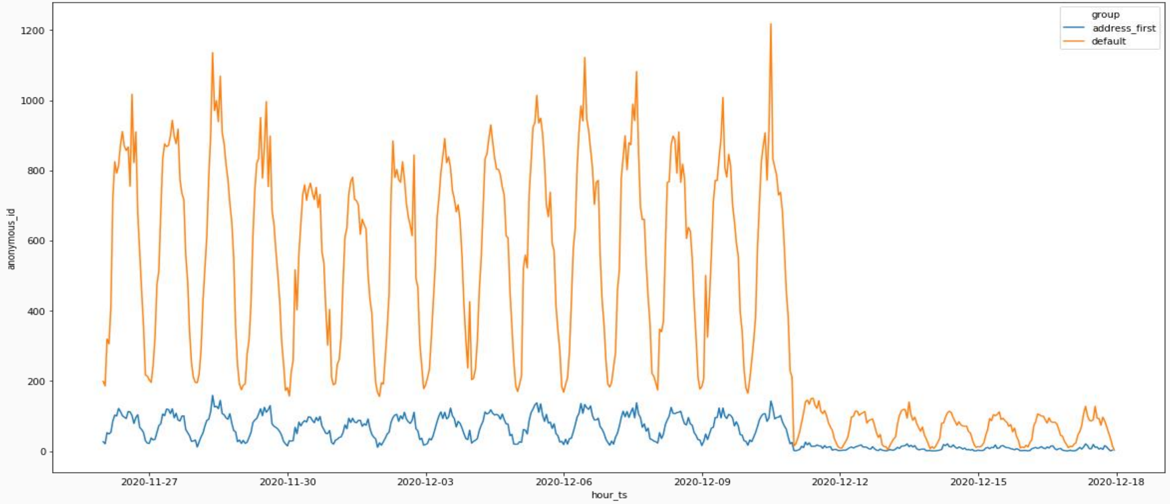
Both metrics were calculated using the Mann-Whitney test.

Anomalies

- 528 users fell into both groups of the experiment at once. We have removed them from the data.
- <1% of users were enrolled in the experiment several times in the same group, but with different browsers and device types. We left some one entry for such users and believe that this did not affect the results.

Anomalies

Since December 11, 2020, the number of users has decrease.



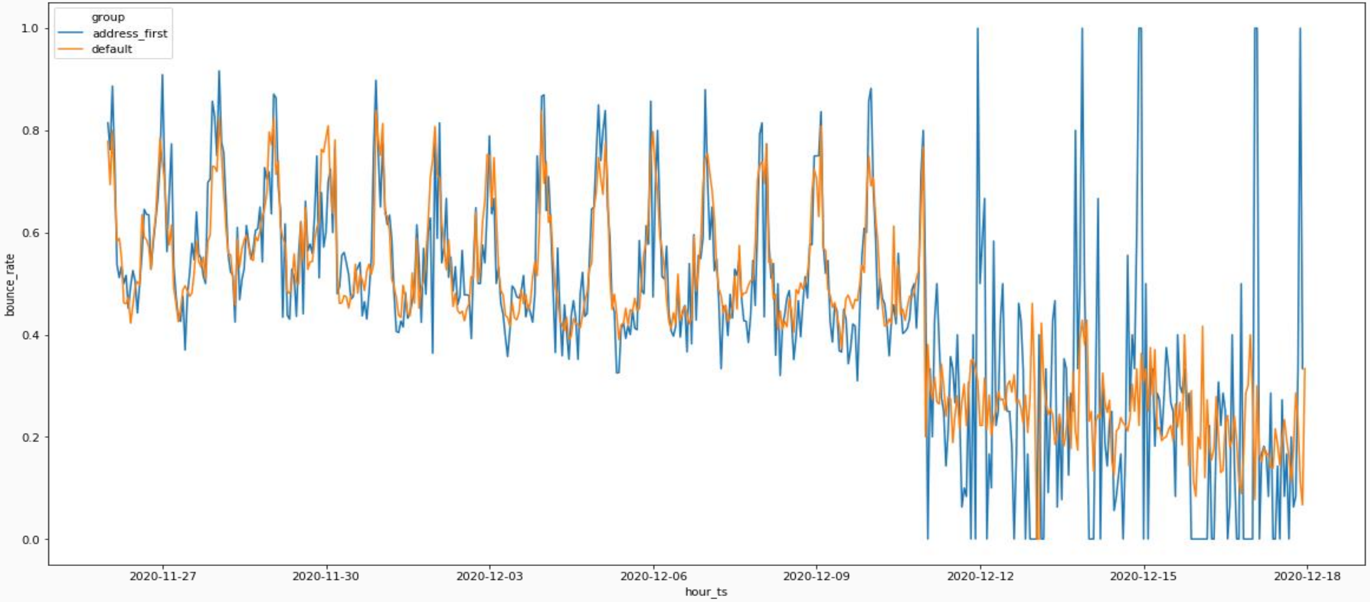
after December 11 in both groups of people becomes much smaller. It looks as if not all data began to come to us and we need to check where the problem with filling them is.

Anomalies

Since December 11, 2020, the number of users has decrease.

When calculating, we relied on data up to December 10 inclusive; in the case of conversion, this limitation resulted in the conversion rate difference being statistically significant (unlike full data).

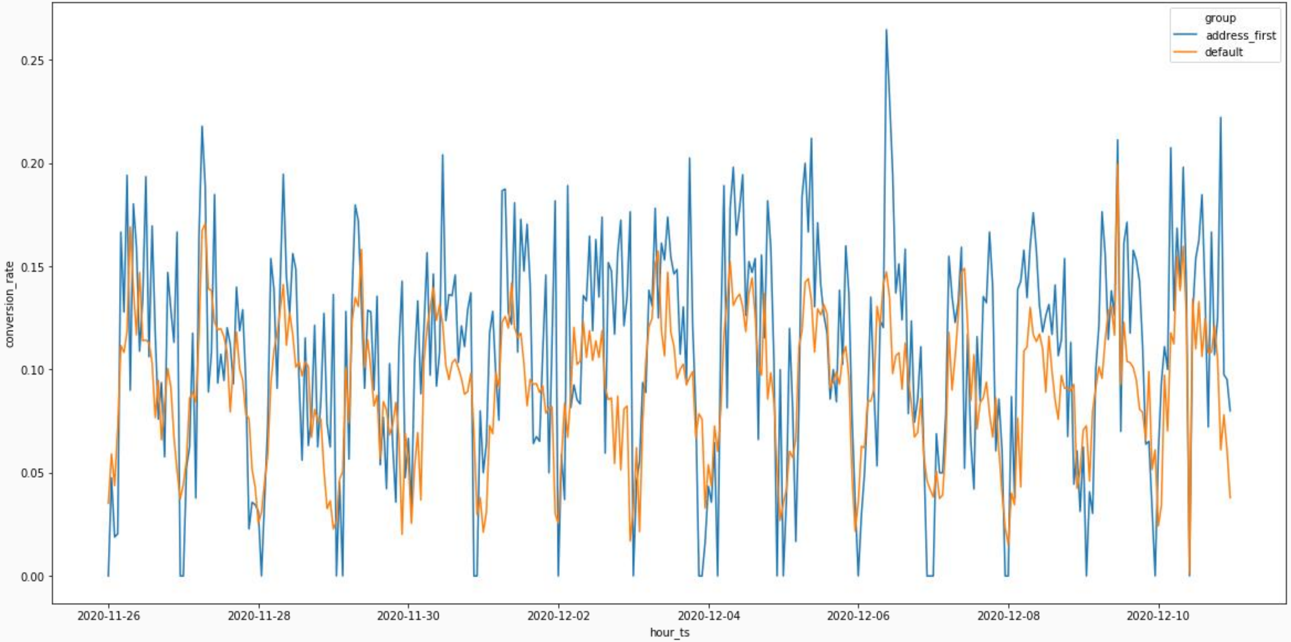
Bounce rate



MannwhitneyuResult(statistic=61718.5, pvalue=0.13477210463377265)

At first glance, the bounce rate graph does not show a significant difference between the groups.
The Mann-Whitney test also does not see significant differences between them, even if we take only data up to December 11 – p-value=0.13.

Conversion rate



MannwhitneyuResult(statistic=49275.0, pvalue=1.3221586226759172e-08)

The graph shows that the conversion of the experimental group is always higher than the conversion of the default group. TThe Mann-Whitney test confirms this. p-value=0.00000001 (statistically significant).

conclusions

New feature:

- does not spoil the bounce rate
- statistically significantly increases conversion

We should use it!*

*in perfect world we should retake test avoid all issues, but probably don’t have time\money for that in this case