

A/B-тест лендинга

Дизайн эксперимента

Описание

На текущем варианте лендинга пользователь сначала выбирает ритейлера, переходит в каталог, ищет нужные ему товары и только потом может узнать, что по его адресу выбранного магазина может не оказаться. Хотим проверить гипотезу, что выбор адреса до этапа выбора магазина поможет избежать “кривых” сценариев без повышения отказов с самого лендинга.

Проблемы дизайна

Эксперимент запущен на всех пользователях, а не только на новых, это приводит к ряду проблем:

- старые пользователи уже сформировали своё отношение к нам как к продукту, и это может влиять на то, будут ли они отваливаться, сильнее, чем новый флоу
- мы знаем адреса и магазины, где старые пользователи делали покупки, и если они попадут в экспериментальную группу и кликнут на привычный магазин, мы отправим их в бессмысленное с точки зрения UX путешествие по экранам и потребуем лишние данные

Проблемы дизайна

- Эксперимент запущен на всех типах устройств (телефоны, десктоп) и операционных системах, хотя разумно было бы предположить, что пользователям разных девайсов может быть удобнее идти по разным сценариям.
- Экспериментальная группа и контрольная разного размера; следовало бы взять, скажем, 20% пользователей, разделить на эксперимент и контроль по 10%, проследить на A/A, что они не отличаются, и тогда уже катить на экспериментальную группу изменения.

Методология исследования

Нам нужно понять, как новый флоу повлияет на bounce rate лендинга (метрика не должна статистически значимо упасть) и на конверсию в фактическое добавление в корзину. Для этого эксперимента мы решили считать конверсию в добавление в корзину из сессии.

За одну сессию мы считали все действия пользователя с одним `anonymous_id` в течение одного астрономического часа (напр. с 12:00 до 13:00). Мы предполагаем, что таким образом мы теряли некоторое количество конверсий (если человек зашёл в 12:30, а в 13:01 совершил целевое действие, мы его не засчитаем), и предполагаем, что это отразится одинаково на результатах обеих групп и не повлияет на результаты эксперимента.

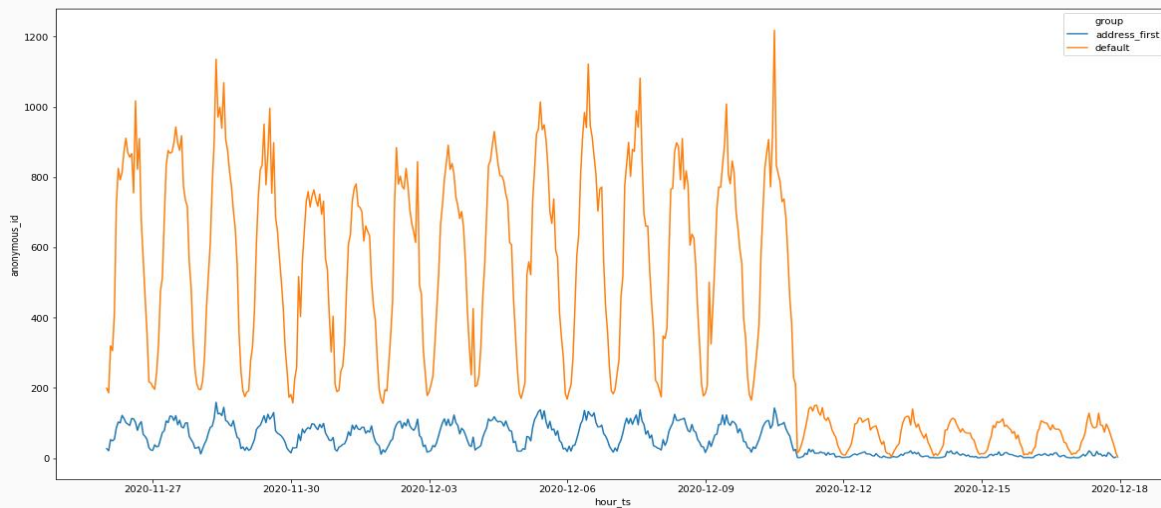
Для расчёта обеих метрик использовался тест Манна-Уитни.

Аномалии в данных

- 528 пользователей попали сразу в обе группы эксперимента. Мы удалили их из данных.
- <1% пользователей несколько раз были записаны в эксперимент в одну и ту же группу, но с разными браузерами, ОС и типами устройств. Мы оставляли для таких пользователей какую-то одну запись и считаем, что это не повлияло на результаты.

Аномалии в данных

С 11 декабря 2020 года количество пользователей резко падает.



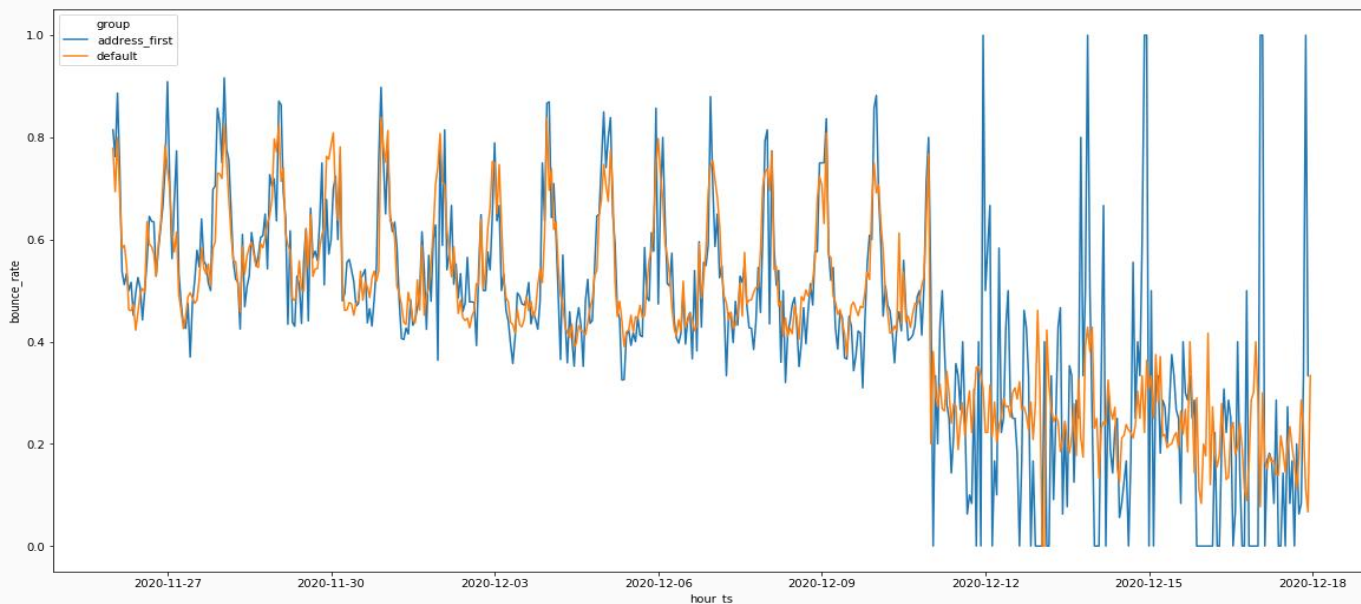
На графике изображено количество уникальных посетителей за каждый час с группировкой по группам эксперимента. Мы видим, что после 11 декабря в обеих группах людей становится сильно меньше. Выглядит как будто к нам стали поступать не все данные и нужно проверить, где проблема с их заливкой.

Аномалии в данных

С 11 декабря 2020 года количество пользователей резко падает.

При расчётах мы опирались на данные до 10 декабря включительно; в случае с конверсией это ограничение привело к тому, что разница conversion rate стала статистически значимой (в отличие от полных данных).

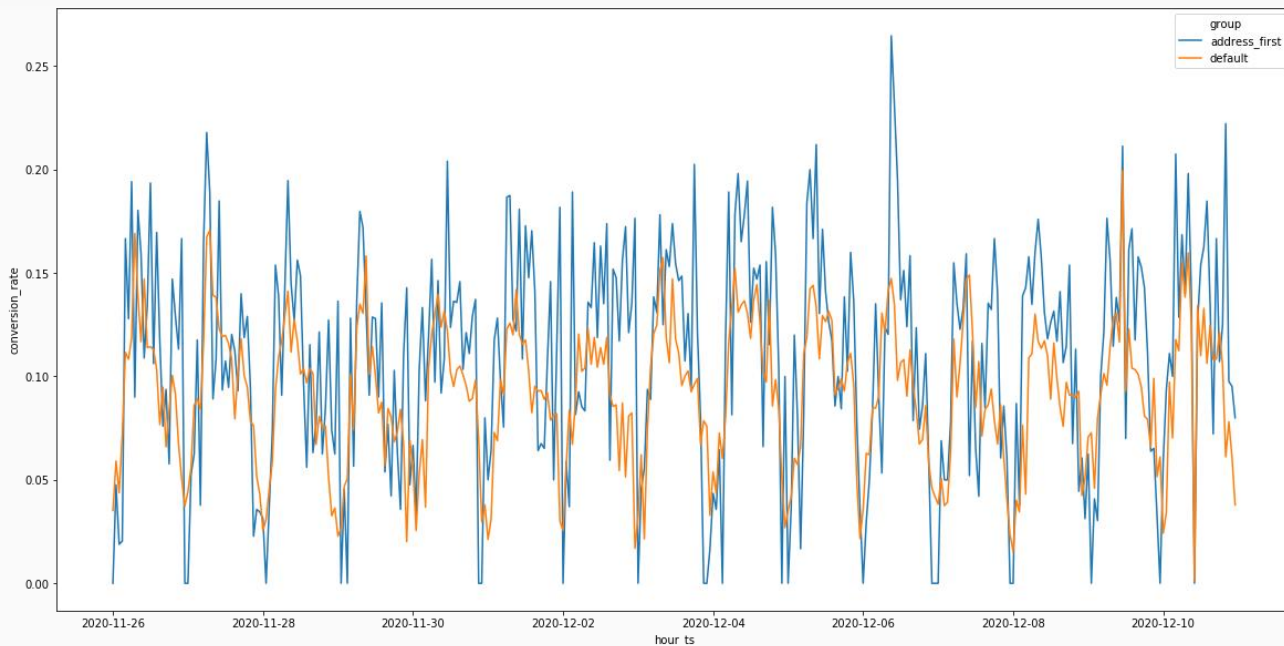
Bounce rate



На первый взгляд, на графике bounce rate не видно значимого различия между группами. Тест Манна-Уитни тоже не видит между ними значимых различий, даже если брать только данные до 11 декабря – $p\text{-value}=0.13$.

MannwhitneyResult(statistic=61718.5, pvalue=0.13477210463377265)

Conversion rate



На графике видно, что конверсия экспериментальной группы всё время оказывается выше, чем конверсия дефолтной группы. Тест Манна-Уитни это подтверждает с $p\text{-value}=0.00000001$ (то есть статистически значимо).

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

Выводы

Новая фишка:

- не портит bounce rate
- статистически значительно повышает конверсию

Значит, надо катить!*

*вообще лучше бы учесть ошибки в дизайне эксперимента и запустить ещё один только на новых пользователях и только на десктопе и проследить, чтобы посреди эксперимента данные не начали никуда утекать, но если мы будем дизайнить идеальный эксперимент, мы так никогда ни одну фишку не запустим