

Анализ A/B теста для сервиса доставки продуктов Сбермаркет

Ноутбук с решением: <https://drive.google.com/file/d/1CRgbIJi4-ocSsj-aUEbyZeFzECIS1Vgf/view?usp=sharing>

Замечания по дизайну эксперимента

Основная претензия — это нечетко сформулирована гипотеза.

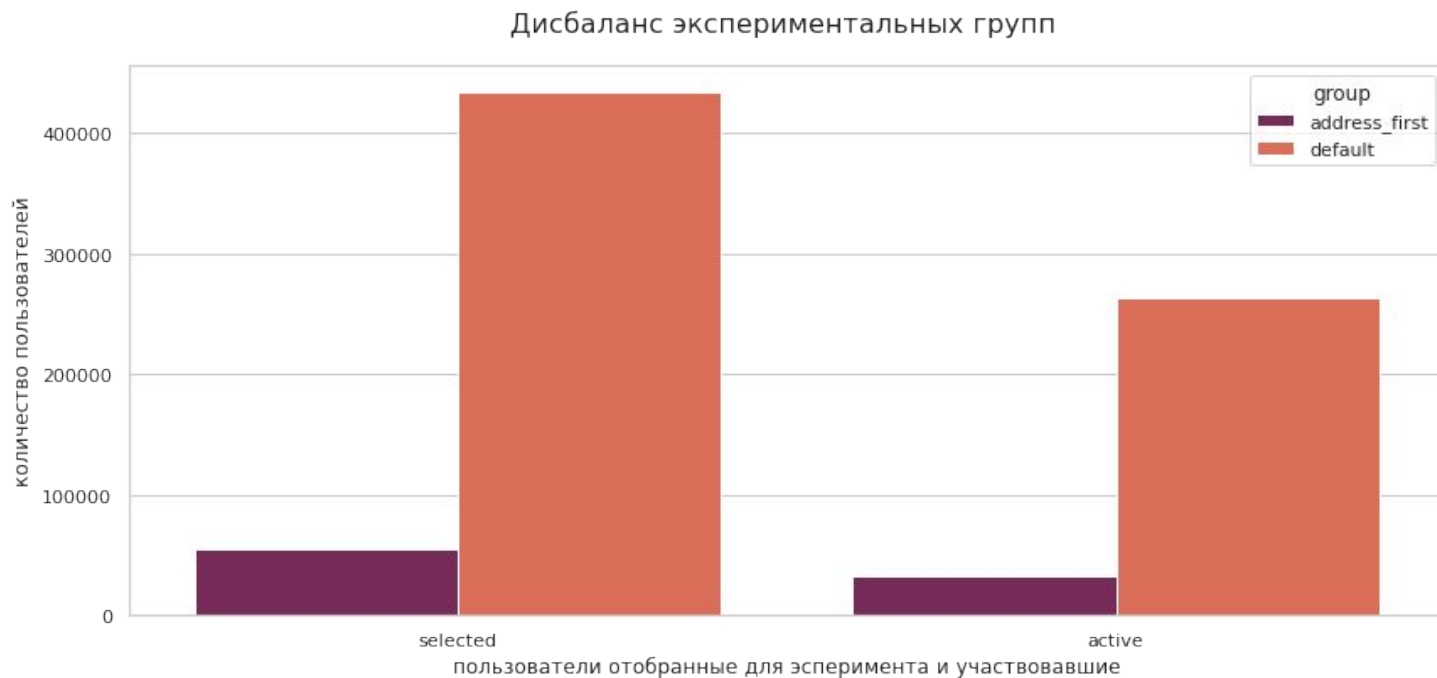
Не указаны критерии "кривизны" сценариев и метрики по которой ее нужно оценивать. Возможно подразумевалось, что это достигается автоматически.

По сути, мы меняем сценарий. И правильнее было бы указать на какую метрику это окажет влияние и в каких пределах.

Пример формулировки: "Выбор адреса до этапа выбора магазина увеличит конверсию в добавление товара в пределах 5%, без статистически значимого роста отказов с лендинга."

Тогда основная метрика будет конверсией в добавление в корзину (то что должно улучшиться, ради чего производим изменения). Отказ с лендинга будет контрольной метрикой (то что не должно ухудшиться).

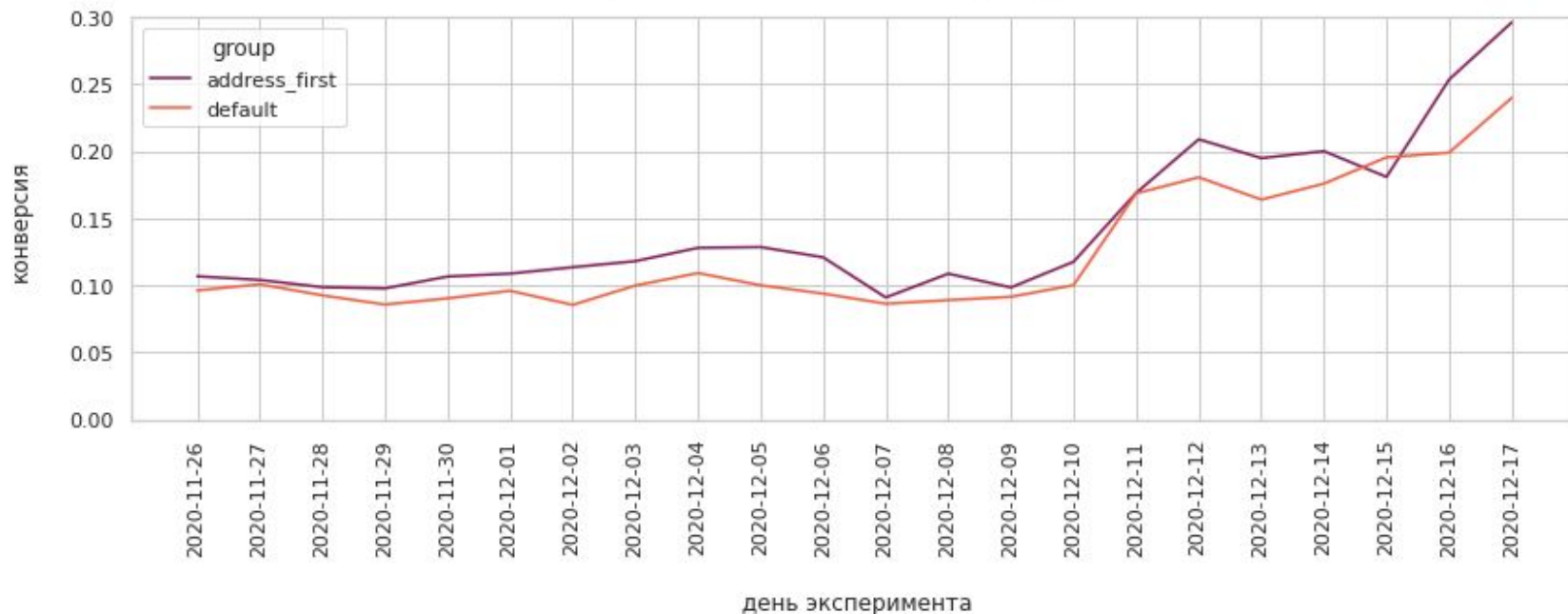
По группам пользователи распределены со значительным дисбалансом: тестовая группа в восемь раз меньше контрольной.



В имеющихся у нас данных о событиях нет событий значительной части пользователей из числа распределенных по группам.

Эксперимент проводился с 26 ноября по 17 декабря.
Набор пользователей в экспериментальные группы прекратился за неделю до завершения теста.

Конверсия в добавление в корзину по дням



На протяжении всего времени эксперимента, конверсия тестовой группы выше контрольной. В последнюю неделю эксперимента, когда новых пользователей не добавлялось, посетители реже заходят на сайт с лендинга, по-этому растет конверсия и снижается доля отказов.

Нас интересует не количество товаров добавленных в корзину, а сам факт того, что пользователь дошел до этого этапа. По-этому конверсия в нашем случае — это отношение количества сеансов пользователей, имеющих хотя бы одно успешное добавление в корзину, к количеству сеансов пользователей начавшихся с посещения лендинга.

Конверсия в добавление в корзину по очищенным данным

Тестовая группа	Посещений лендинга	Добавлений в корзину	Конверсия
address_first	25570	2583	0.101
default	207424	18026	0.087

В тестовой группе конверсия в добавление товаров в корзину выше на 1.4%

Отказ в нашем случае – это отношение количества уникальных пользователей посетивших только лэндинг, к количеству уникальных пользователей, тех кто прошел на страницу ритейлера, но при этом был на лендинге.



Доля отказов тестовой группы значительно выше на протяжении всего времени эксперимента.

Изменения вызвали значительное увеличение и без того очень высокой доли отказов с лендинга.

Доля отказов по очищенным данным

Тестовая группа	Посещений лендинга	Количество отказов	Доля отказов
address_first	25570	18088	0.707
default	207424	119223	0.575

Доля отказов с лендинга в тестовой группе по очищенным данным на 14% выше контрольной.

Итак, у нас значительно увеличилась доля отказов, и немного выросла конверсия. Если исходить из условий эксперимента, то значительно выросшая доля отказов не дает нам оснований говорить об успешности теста.

Однако, на конверсию в покупки изменения сценария статистически значимо не повлияли:

Конверсия в завершённые заказы

Посещений лендинга	Завершённых заказов	Конверсия в завершённые заказы
26172.0	458.0	0.017
211696.0	3309.0	0.016

Поведение пользователей с намерением завершить заказ — не изменилось.

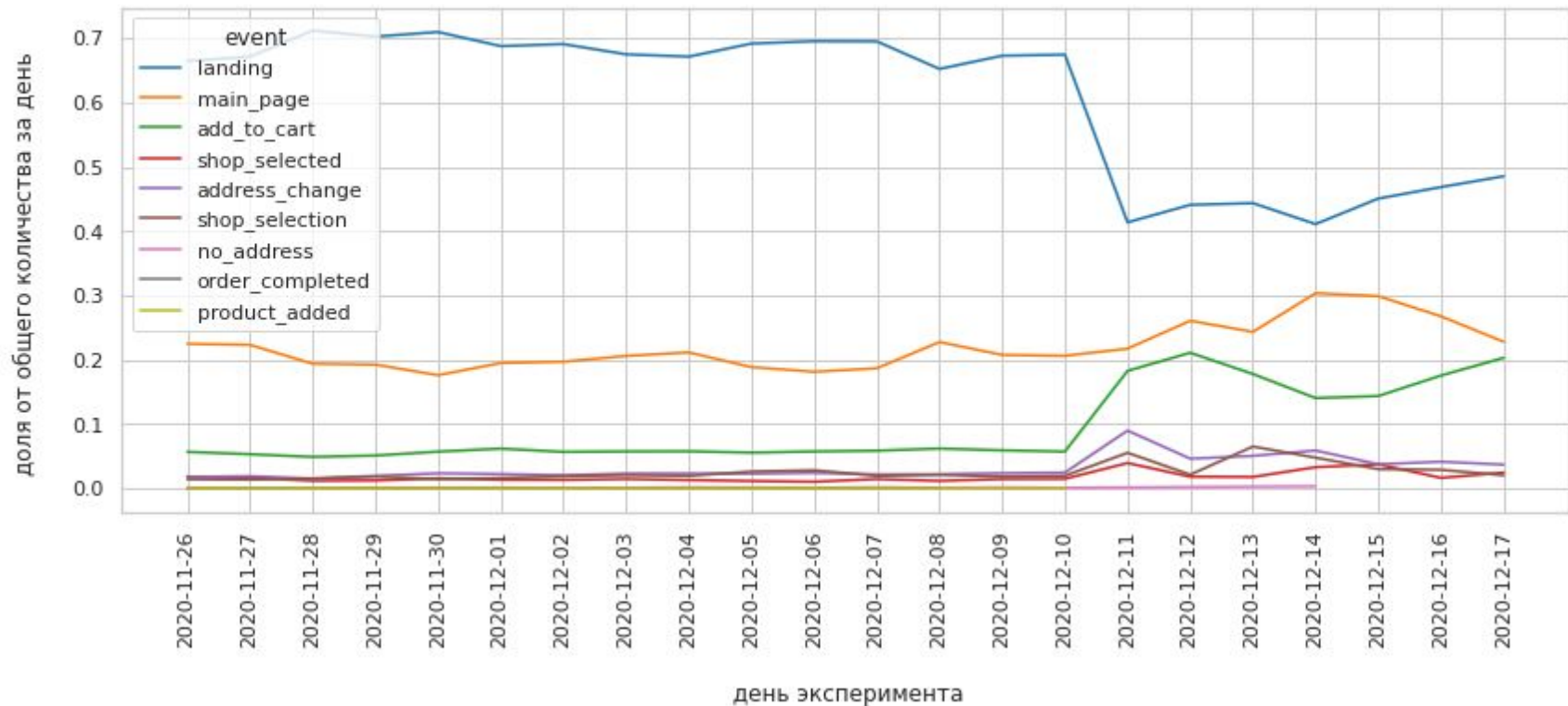
Можно предположить, что поменялось поведение новых пользователей зашедших с целью "посмотреть". Те кого отпугнула необходимость вводить персональные данные (адрес доставки) ушли в отказ, те кто остался продемонстрировали увеличение конверсии в добавление в корзину. Значит интерфейс стал удобнее.

Поэтому наше предложение, изменения принимать, а проблему отказа на стадии введения персональных данных решать с помощью повышения доверия к сервису. Например с помощью рекламной кампании с лидером мнений целевой аудитории.

Аномалии в данных

30% посетителей начали сессии не с просмотра лендинга, а с последующих событий.

Первые события пользователей на сайте



Либо эксперимент проводился не среди новых пользователей, либо у нас отсутствует часть данных по активности пользователей.

Аномалии в данных

- В последовательностях событий пользователей встречаются цепочки из одного и того же события. 11.59% от общего количества событий приходится на такие повторы. Их возникновение не зависит от времени или от техники пользователя. Скорее всего повторы генерируются движком сайта в момент бездействия. Возможно из-за автоматического обновления страниц.
- По имеющимся у нас данным, не все пользователи из числа распределенных по группам, совершали действия в период эксперимента. Очень вероятно что у нас не все данные. Либо не все дни эксперимента, либо не все пользователи.
- События "Shop Selection Started" и "Shop Selected" последовательны, второе не может произойти без первого. По этому странно, что событий "Shop Selected" больше чем "Shop Selection Started". В качестве источника обоих событий есть 'header'. Либо часть пользователей вызвала окно выбора магазина оттуда, либо поле с магазином в шапке заполняется автоматически при повторном посещении сайта.

Недочеты эксперимента

Не определено, на каком сегменте пользователей проведен эксперимент. Подобные изменения правильно тестировать на новых пользователях. Вернувшиеся пользователи возможно помнят, что где-то нужно указывать адрес и корректируют свое поведение. Судя по тому, что 30% пользователей первым событием имеют не просмотр лендинга, пользователи не новые.

Отсутствуют данные о пересечении с одновременно проходящими тестами. Упоминание о слове "test" есть в схеме эксперимента, но этого столбца нет в таблице распределения по группам эксперимента.