На основе выбранных продуктов из первого вебинара, выберите 2 A/B-теста

и составьте план эксперимента, чтобы избежать Peeking Problem

Распишите, какие метрики и как вы будете анализировать

План экспиримента:

1. Определение цели

Нужно добиться роста количества пользователей. Компания стремится к росту количества активных пользователей (метрика DAU), определяемых как среднее количество зарегистрированных пользователей сайта в день за последние 30 дней. Предположим, что этого можно добиться либо путем улучшения показателей удержания (процент пользователей, возвращающихся для повторного использования продукта), либо путем увеличения числа новых регистрирующихся пользователей.

В процессе исследования воронки мы заметили, что 60% пользователей уходят до завершения регистрации. Это означает, что можно повысить количество регистраций, изменив страницу регистрации, что, в свою очередь, должно помочь увеличить количество активных пользователей.

2. Определение метрики

В качестве метрики выбираем долю зарегистрированных пользователей (registration rate), определяемую как количество новых пользователей, которые регистрируются, поделенное на общее количество новых посетителей сайта.

3. Разработать гипотезу

На текущей странице регистрации есть баннер и форма регистрации. Есть несколько пунктов, которые мы можем протестировать: поля формы, позиционирование, размер текста, но баннер на главной странице визуально наиболее заметен, поэтому сначала надо узнать, увеличится ли доля регистраций, если изменить изображение на нём.

Общая гипотеза заключается в следующем: «Если изменить главную страницу регистрации, то больше новых пользователей будут регистрироваться внутри продукта, потому что новое изображение лучше передает его ценности».

Нам необходимо определить две гипотезы, которые помогут понять, является ли наблюдаемая разница между версией A (изначальной) и версией B (новой, которую мы хотим проверить) случайностью или результатом изменений, которые мы произвели.

— Нулевая гипотеза предполагает, что результаты, А и В на самом деле не отличаются и что наблюдаемые различия случайны. Мы надеемся опровергнуть эту гипотезу.

— Альтернативная гипотеза — это гипотеза о том, что B отличается от A, и мы хотим сделать вывод об её истинности.

4. Подготовить эксперимент

Для того, чтобы тест выдавал корректные результаты сделаем следующее:

— Создадим новую версию (B), отражающую изменения, которые мы хотите протестировать.

— Определим контрольную и экспериментальную группы. Каких пользователей мы хотитм протестировать: всех пользователей на всех платформах или только пользователей из одной страны? Определим группу испытуемых, отобрав их по типам пользователей, платформе, географическим показателям и т. п. Затем определим, какой процент исследуемой группы составляет контрольная группа (группа, видящая версию A), а какой процент — экспериментальная группа (группа, видящая версию B).

— Убедимся, что пользователи будут видеть версии A и B в случайном порядке. Это значит, у каждого пользователя будет равный шанс получить ту или иную версию.

— Определим уровень статистической значимости (α). Это уровень риска, который мы принимаем при ошибках первого рода (отклонение нулевой гипотезы, если она верна), обычно α = 0.05. Это означает, что в 5% случаев мы будем обнаруживать разницу между A и B, которая на самом деле обусловлена случайностью. Чем ниже выбранный нами уровень значимости, тем ниже риск того, что мы обнаружим разницу, вызванную случайностью.

— Определим минимальный размер выборки. Калькуляторы есть здесь и здесь, они рассчитывают размер выборки, необходимый для каждой версии. На размер выборки влияют разные параметры и наши предпочтения. Наличие достаточно большого размера выборки важно для обеспечения статистически значимых результатов.

— Определим временные рамки. Возьмем общий размер выборки, необходимый нам для тестирования каждой версии, и разделим его на наш ежедневный трафик, так мы получим количество дней, необходимое для проведения теста. Как правило, это одна или две недели.

На существующем сайте в разделе регистрации мы изменим главную страницу — это и будет нашей версией B. Мы решаем, что в эксперименте будут участвовать только новые пользователи, заходящие на страницу регистрации. Мы также обеспечиваем случайную выборку, то есть каждый пользователь будет иметь равные шансы получить A или B, распределенные случайным образом.

Важно определить временные рамки. Допустим, ежедневно на нашу страницу регистрации в среднем приходит трафик от 10 000 новых пользователей, это означает, что только 5000 пользователей могут увидеть каждую версию. Тогда минимальный размер выборки составляет около 100 000 просмотров каждой версии. 100 000/ 5000 = 20 дней — столько должен продлиться эксперимент.

5. Проведите эксперимент

— Обсудим параметры эксперимента с исполнителями.

— Выполним запрос на тестовой закрытой площадке. Это поможет проверить данные.

— В самом начале проведения тестирования проверим, действительно ли оно работает.

— Преждевременный просмотр результатов может испортить статистическую значимость.

6. Анализ результатов.

Нам нужно получить данные и рассчитать значения выбранной ранее метрики успеха для обеих версий (A и B) и разницу между этими значениями. Если не было никакой разницы в целом, Мы также можем сегментировать выборку по платформам, типам источников, географическим параметрам и т. п., если это применимо. Мы можете обнаружить, что версия B работает лучше или хуже для определенных сегментов.

Проверим статистическую значимость. Статистическая теория, лежащая в основе этого подхода, но основная идея в том, чтобы выяснить, была ли разница в результатах между A и B связана с изменениями или это результат случайности или естественных изменений. Это определяется путем сравнения тестовых статистических данных (и полученного p-значения) с нашим уровнем значимости.

Если p-значение меньше уровня значимости, то можно отвергнуть нулевую гипотезу, имея доказательства для альтернативы.

Если p-значение больше или равно уровню значимости, мы не можем отвергнуть нулевую гипотезу о том, что A и B не отличаются друг от друга.

A/B-тестирование может дать следующие результаты:

— Контрольная версия, А выигрывает или между версиями нет разницы. Если исключить причины, которые могут привести к недействительному тестированию, то проигрыш новой версии может быть вызван, например, плохим сообщением и брендингом конкурентного предложения или плохим клиентским опытом.

— Версия B выигрывает. A/B-тест подтвердил вашу гипотезу о лучшей производительности версии B по сравнению с версией A. Отлично! Опубликовав результаты, мы можем провести эксперимент на всей аудитории и получить новые результаты.

Заключение

По результатам теста сделаем заключение.

Если вдруг у нас будет нестандартная ситуация, так как на носу новый год и возможность нагрузки на сайт велик из-за большого количества посещения сайта. И это может повлиять на результат тестирования в таком случае нам придется перевести период теста после праздников.