

Проведение и анализ А/В-теста для сервиса «Всё.техника»

Предмет исследования: выделение игровых ноутбуков в отдельную категорию

Аналитик: Анастасия Домнина

Предварительные гипотезы

Есть гипотеза, что спрос на игровые ноутбуки мог бы быть больше, но пользователи не могут найти их среди других ноутбуков.

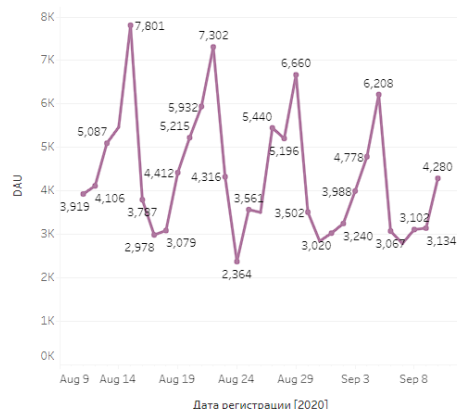
Требуется провести A/B-тест и вынести игровые ноутбуки в отдельную категорию товаров.

Ожидается, что в результате изменений, конверсия в покупку вырастет на 100%, а средний чек в категории «Компьютеры» не изменится. Тест планируется провести только для новых пользователей.

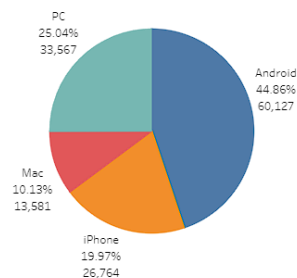
1 этап. Подготовка к тесту

Данные для подготовки к А/В тесту

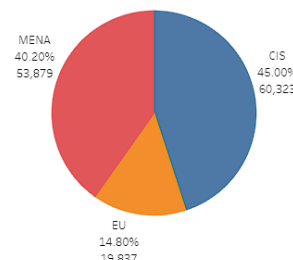
DAU новых пользователей



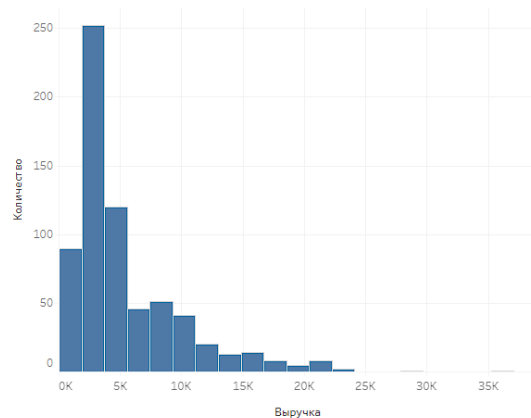
Распределение новых пользователей по типам устройств



Распределение новых пользователей по регионам



Гистограмма выручки



Средняя выручка

5,421

Дисперсия выручки

4,697

Средний DAU новичков

4,324

Конверсия в покупку

0.5013 %

Для начала соберём исторические данные для определения параметров А/В-теста.

Данные были отобраны только по новым пользователям, которые зарегистрировались в период с 11.08.2020 по 10.09.2020.

Показатели выручки и конверсии в покупку были рассчитаны для категории «Компьютерная техника».

Рассмотрим графики подробнее.

Общий вид [дашборда](#) с данными для подготовки к А/В-тесту.

DAU новых пользователей

DAU новых пользователей

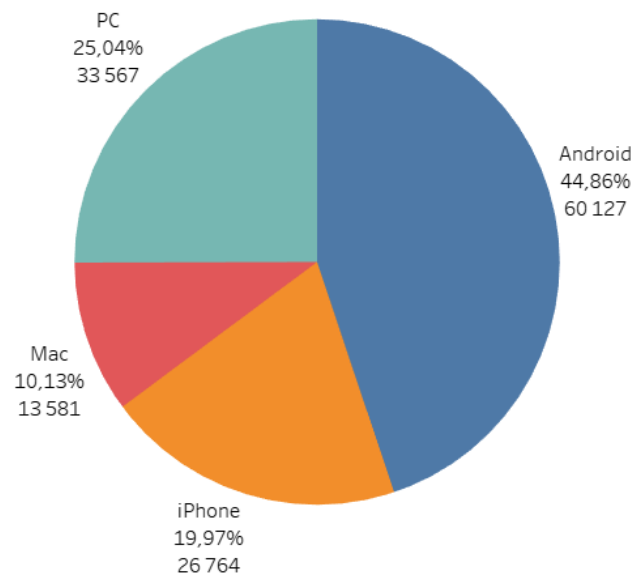


Средний DAU для сервиса за исследуемый период – 4324. При этом максимальное значение – 7801, а минимальное – 2364.

График имеет отчетливую сезонность привлечения новых пользователей: пики минимума приходятся на понедельник, далее мы наблюдаем стабильный рост показателя до пика максимума в субботу, сильное падение в воскресенье и следующий пик минимума в понедельник.

Распределение новых пользователей по типам устройств

Распределение новых пользователей по типам устройств

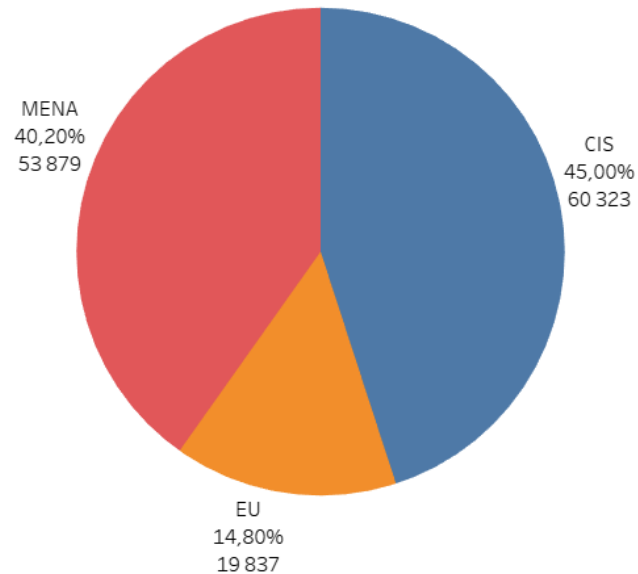


По типам устройств пользователи имеют следующее распределение:

- ~ 45% пользователи Android
- ~ 25% пользователи PC
- ~ 20% пользователи iPhone
- ~ 10% пользователи Mac

Распределение новых пользователей по регионам

Распределение новых пользователей по регионам



По регионам пользователи имеют следующее распределение:

- ~ 45% пользователи региона CIS
- ~ 40% пользователи региона MENA
- ~ 15% пользователи региона EU

Конверсия в покупку

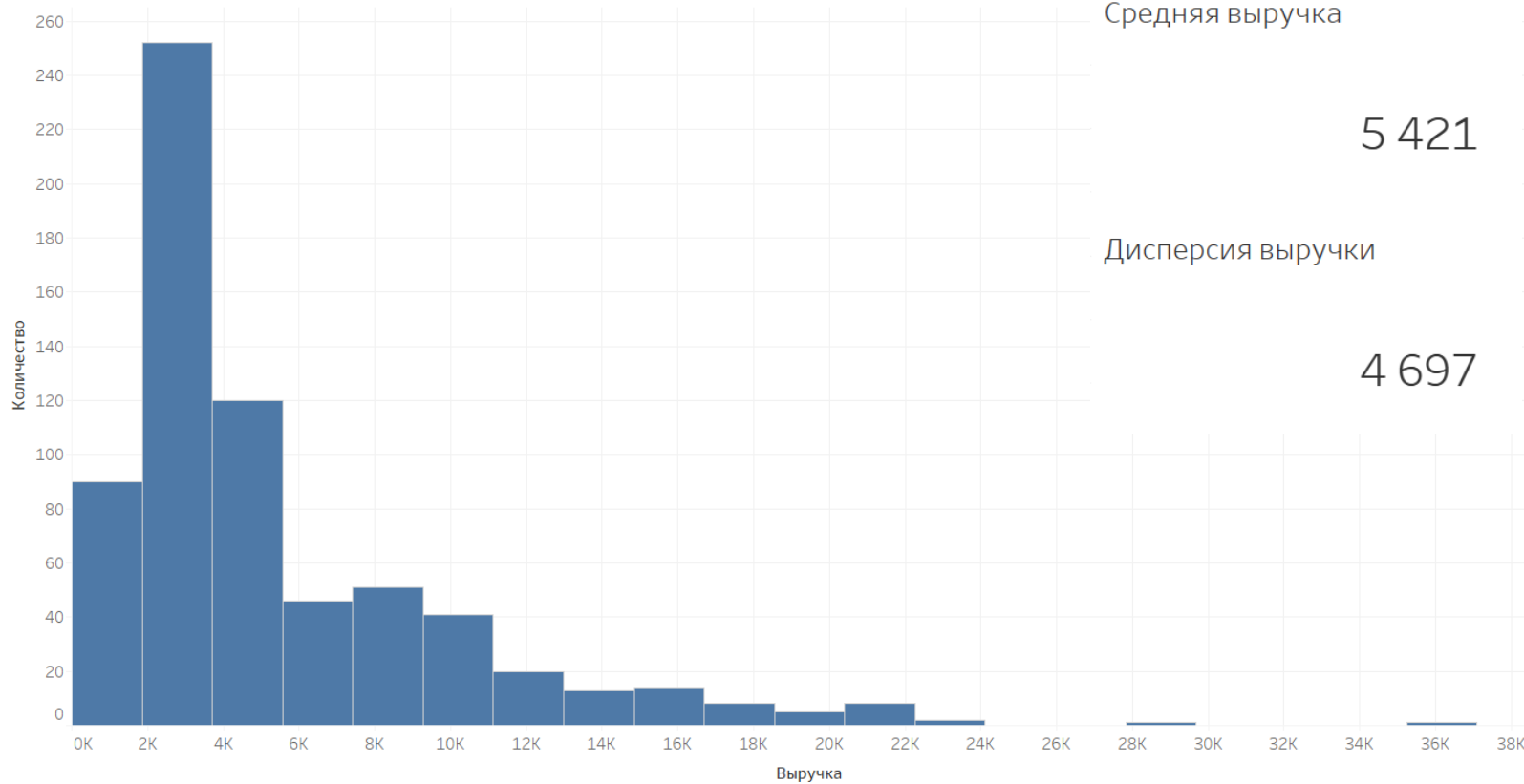
Для периода с 11.08.2020 по 10.09.2020 конверсия
в покупку компьютерной техники составила ~ 0.5%

Конверсия в покупку

0,5013 %

Выручка от покупки

Гистограмма выручки



По гистограмме распределения выручки можно заметить пик для значений от 2000 до 4000, а также длинный «хвост» справа со значениями до 36000.

Средняя выручка составляет 5421, а СКО – 4697.

Параметры для проведения теста

На основе полученных данных определим параметры для проведения теста: длительность и время запуска

Найдем размер выборки, который нужен, чтобы найти изменения в конверсии в покупку компьютера.

Исторический уровень конверсии – 0.5%

Ожидаемый эффект – 100%

Среднее число новых пользователей за неделю – $4324 * 7 = 30268$

Уровень значимости – 0.05

Мощность – 80%

Тест двусторонний.

Получаем: 4673 пользователя в каждой тестовой группе.

Далее найдем размер выборки, который нужен, чтобы найти изменения в среднем чеке в категории компьютеров.

Историческое значение среднего чека – 5421

Ожидаемый эффект – $5421 + 5\% = 5692$

СКО – 4697

Уровень значимости – 0.05

Мощность – 80%

Получаем: 4716 пользователя в каждой тестовой группе.

Для расчёта длительности теста берём наибольший размер выборки (4716), делим его на среднее число новых пользователей (4324), умножаем на число групп. Учитывая сезонность DAU нам необходима минимум полная неделя для проведения теста, поэтому округляем вверх до ближайшей недели и получаем длительность теста в 7 дней.

Параметры для проведения теста

Далее необходимо выбрать дату начала теста.

Слева представлен календарь значимых праздников, на которые запланированы маркетинговые события в 2020 году.

Согласно календарю – самым удачным временем является запуск сразу после Chinese Moon Festival. Например, 14 октября.

С одной стороны, разработчики успеют всё сделать и протестировать, с другой — пользователи в регионе APAC успеют немного отдохнуть от праздничной активности, при этом сам тест не откладывается далеко.

Название события	Регион	Дата начала	Дата окончания
New year	CIS, MENA, EU	2020-01-01	2020-01-01
May Day	CIS, EU	2020-05-01	2020-05-04
Chinese New Year	APAC	2020-01-25	2020-01-25
Valentines day	CIS, MENA, EU	2020-02-14	2020-02-14
St. Patric's Day	EU	2020-03-17	2020-03-17
Victory Day (May 9th)	CIS	2020-05-09	2020-05-09
Orthodox Christmas	CIS	2020-01-07	2020-01-07
International Women's Day	CIS, EU	2020-03-08	2020-03-08
Chinese Moon Festival	APAC	2020-10-01	2020-10-07
Catholic Christmas	MENA, EU	2020-12-25	2020-12-25

Техническое задание на проведение теста

Обобщим собранные данные.

Название теста: gaming_laptops_test

Назначение теста: проверка гипотезы - спрос на игровые ноутбуки будет больше, если вынести их в отдельную категорию. Т.е. ожидается увеличение конверсии в покупку, при этом средний чек должен остаться неизменным

Тестируемые метрики: конверсия в покупку (ожидание роста на 100%), средний чек (ожидание неизменности)

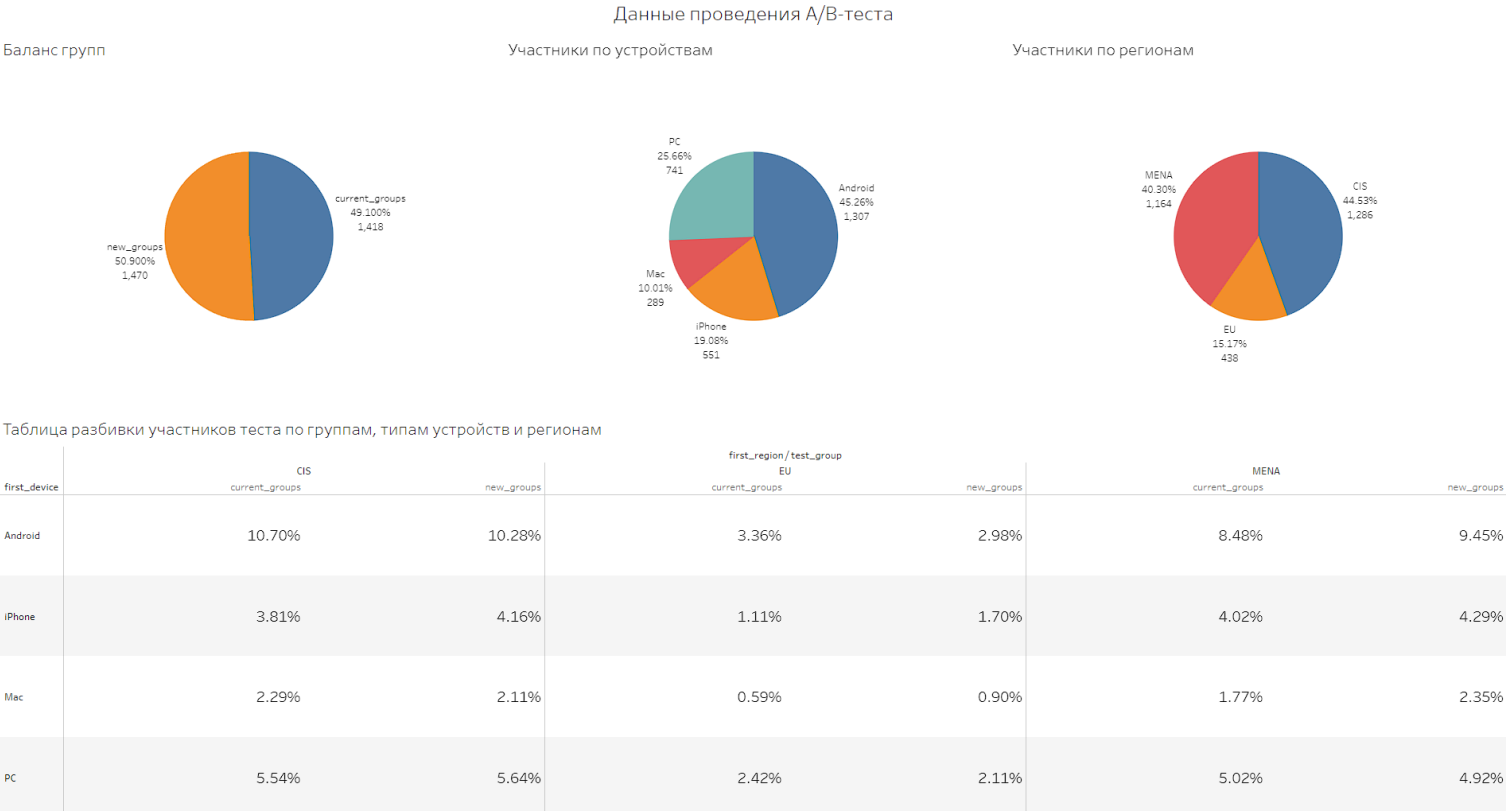
Число тестовых групп, их названия, доля пользователей для каждой группы: 2 группы, тестовая и контрольная с равномерным распределением (50/50)

Длительность теста и нужное число пользователей: 7 дней, не менее 9432 пользователей (по 4716 в каждой группе)

Дата запуска теста: 14.10.2020

Дата остановки теста: 20.10.2020 (включительно)

2 этап. Проведение теста



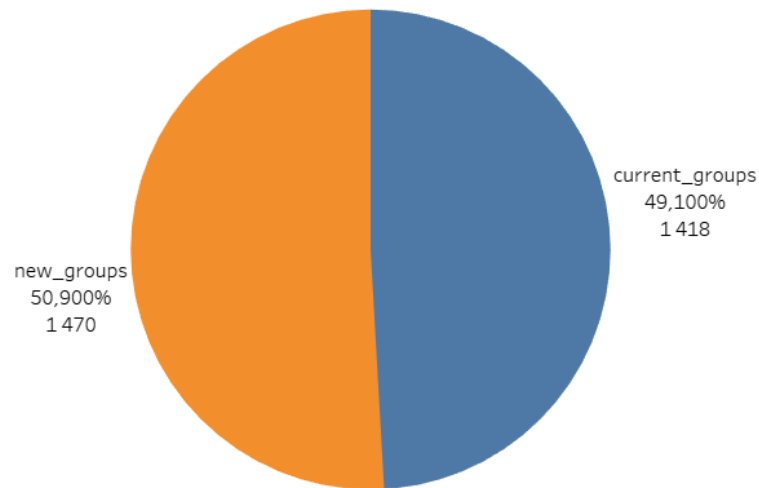
Следующий этап – проведение теста и проверка корректности собираемых данных. Данные были отобраны только по новым пользователям, которые зарегистрировались 14.10.2020

Рассмотрим графики подробнее.

Общий вид [дашборда](#) с данными проведения А/В-теста

Баланс групп

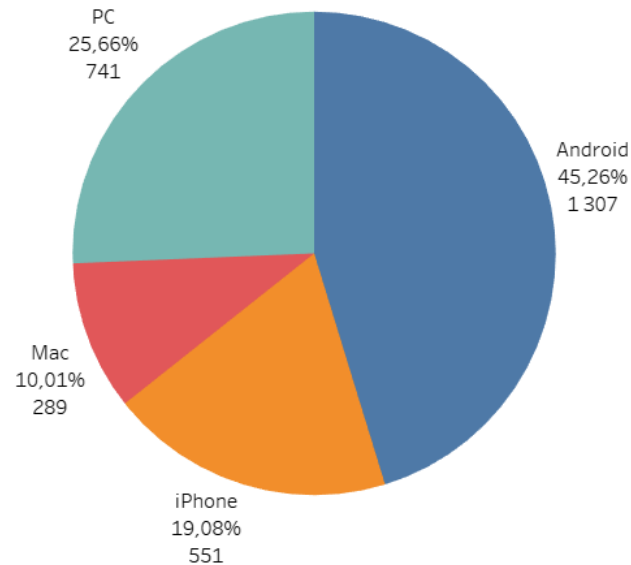
Баланс групп



Всего в тесте приняло участие 2888 пользователей.
Из них 50.9% попали в тестовую группу, 49.1% - в контрольную.
Далее будут произведены расчеты для вычисления равномерности распределения.

Участники по устройствам

Участники по устройствам



По типам устройств пользователи
распределились следующим образом:

~ 45% пользователи Android

~ 26% пользователи PC (~ 25% в исторических значениях)

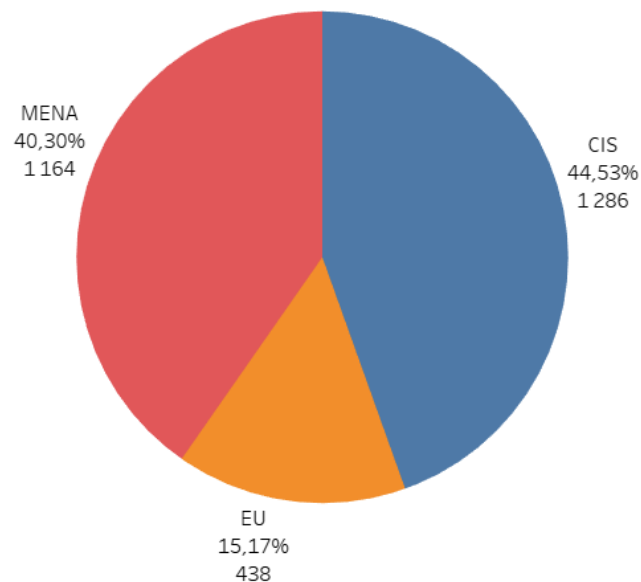
~ 19% пользователи iPhone (~ 20% в исторических значениях)

~ 10% пользователи Mac

Небольшие расхождения с историческими значениями есть, в пределах 1%.

Участники по регионам

Участники по регионам



По регионам пользователи распределились следующим образом :

- ~ 45% пользователи региона CIS
- ~ 40% пользователи региона MENA
- ~ 15% пользователи региона EU

Расхождения с историческими значениями минимальны.

Общая таблица распределения пользователей по группам, устройствам и регионам

Таблица разбивки участников теста по группам, типам устройств и регионам

first_dev..	CIS		first_region / test_group EU		MENA	
	current_groups	new_groups	current_groups	new_groups	current_groups	new_groups
Android	10,70%	10,28%	3,36%	2,98%	8,48%	9,45%
iPhone	3,81%	4,16%	1,11%	1,70%	4,02%	4,29%
Mac	2,29%	2,11%	0,59%	0,90%	1,77%	2,35%
PC	5,54%	5,64%	2,42%	2,11%	5,02%	4,92%

Оценка правильности проведения теста

Оценим корректность распределения участников теста на группы.

Доля пользователей в тестовой группе – 50.9%

Ожидаемая доля – 50%

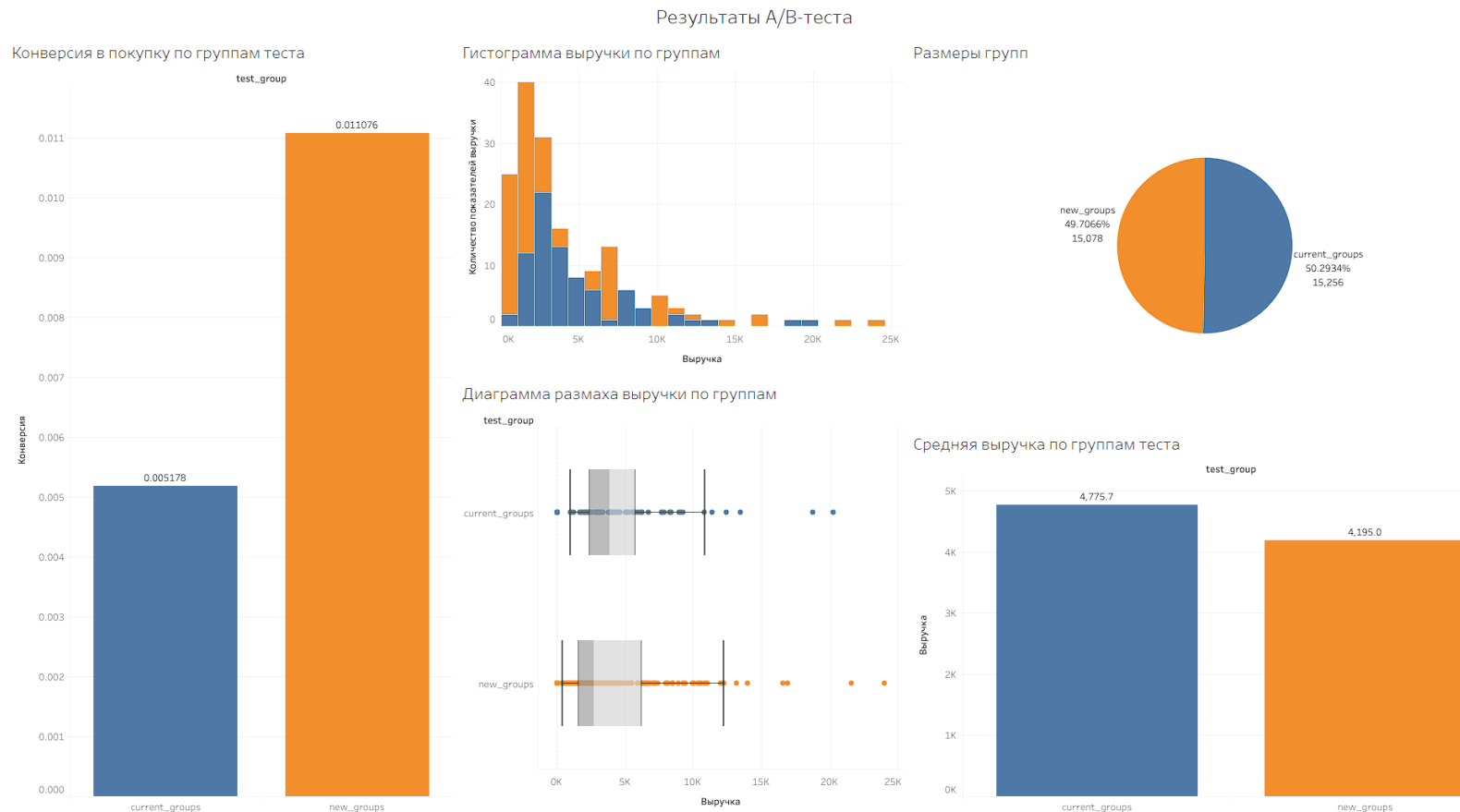
Размер выборки – 2888

Уровень значимости – 0.05

P-value равен 0.33338, что выше заданного уровня значимости, и значит, что наблюдаемый несущественный перекося доли пользователей в сторону группы new_groups — результат случайности, а не систематическое явление.

Тест проводится корректно, его необходимо продолжать для достижения целей по объему выборок и продолжительности в связи с сезонностью.

3 этап. Результаты теста



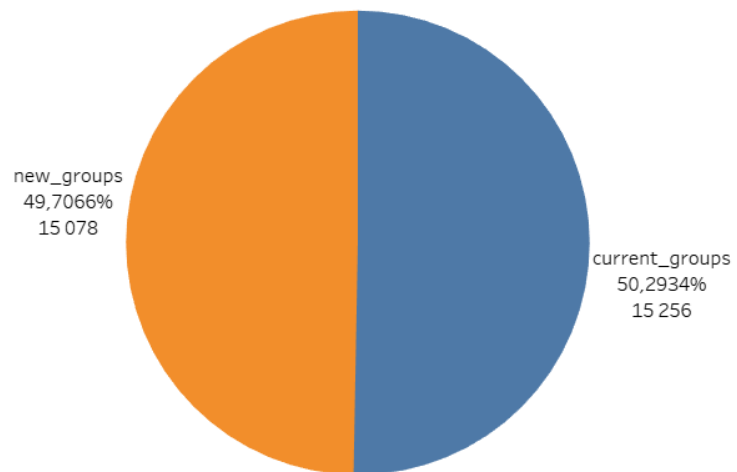
Последний этап – исследование результатов проведенного теста. Тест проводился с 14 по 20 октября 2020 года.

Рассмотрим графики подробнее.

Общий вид [дашборда](#) с результатами A/B-теста

Баланс групп

Размеры групп



Всего в тесте приняло участие 30334 пользователя.

Из них 49.7% попали в тестовую группу, 50.3% - в контрольную.

Распределение равномерно.

Конверсия в покупку по группам

Конверсия в покупку по группам
теста



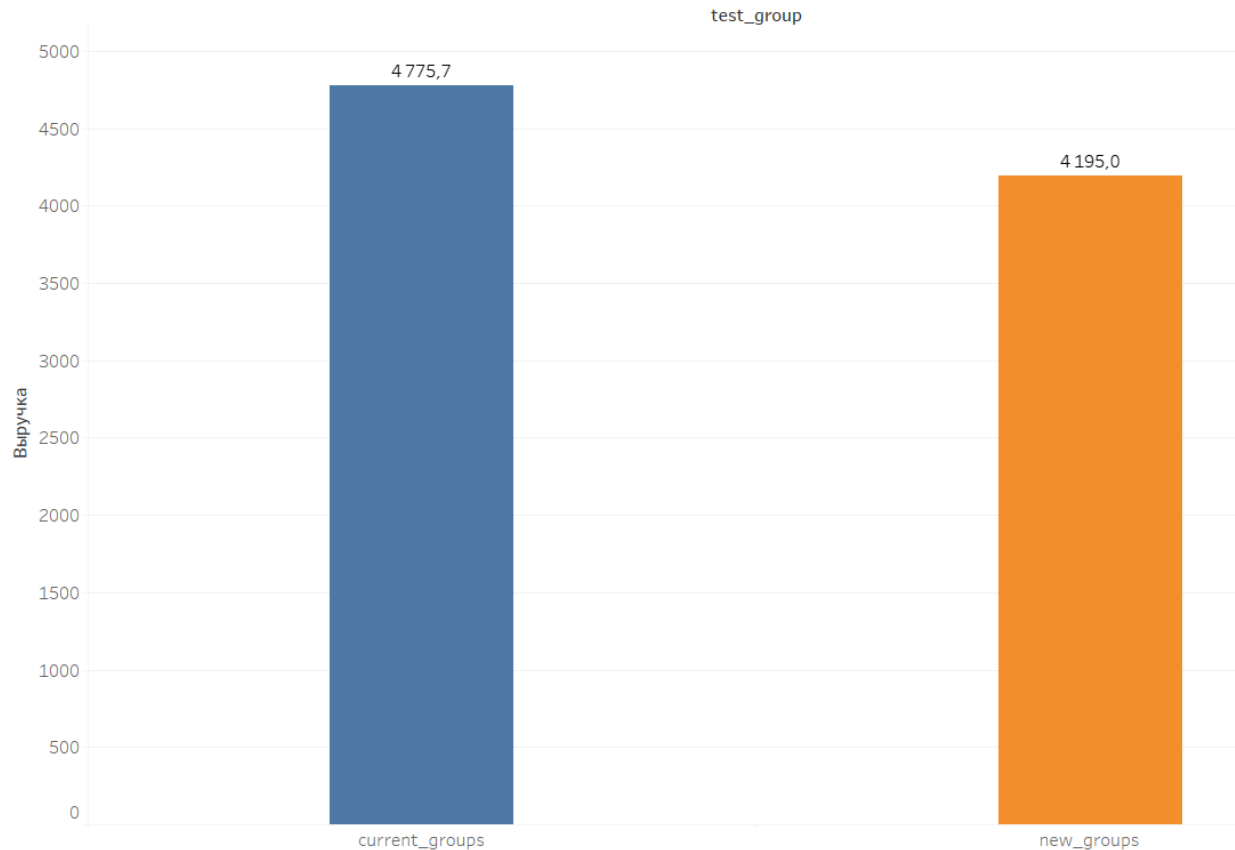
Рассмотрим конверсию наших групп. По графику можно заметить значительное увеличение конверсии в покупку для пользователей из тестовой группы – ~ в 2 раза, или, на 100%

Значение для контрольной группы – 0.52%

Значение для тестовой группы – 1.1%

Средний чек по группам

Средняя выручка по группам теста



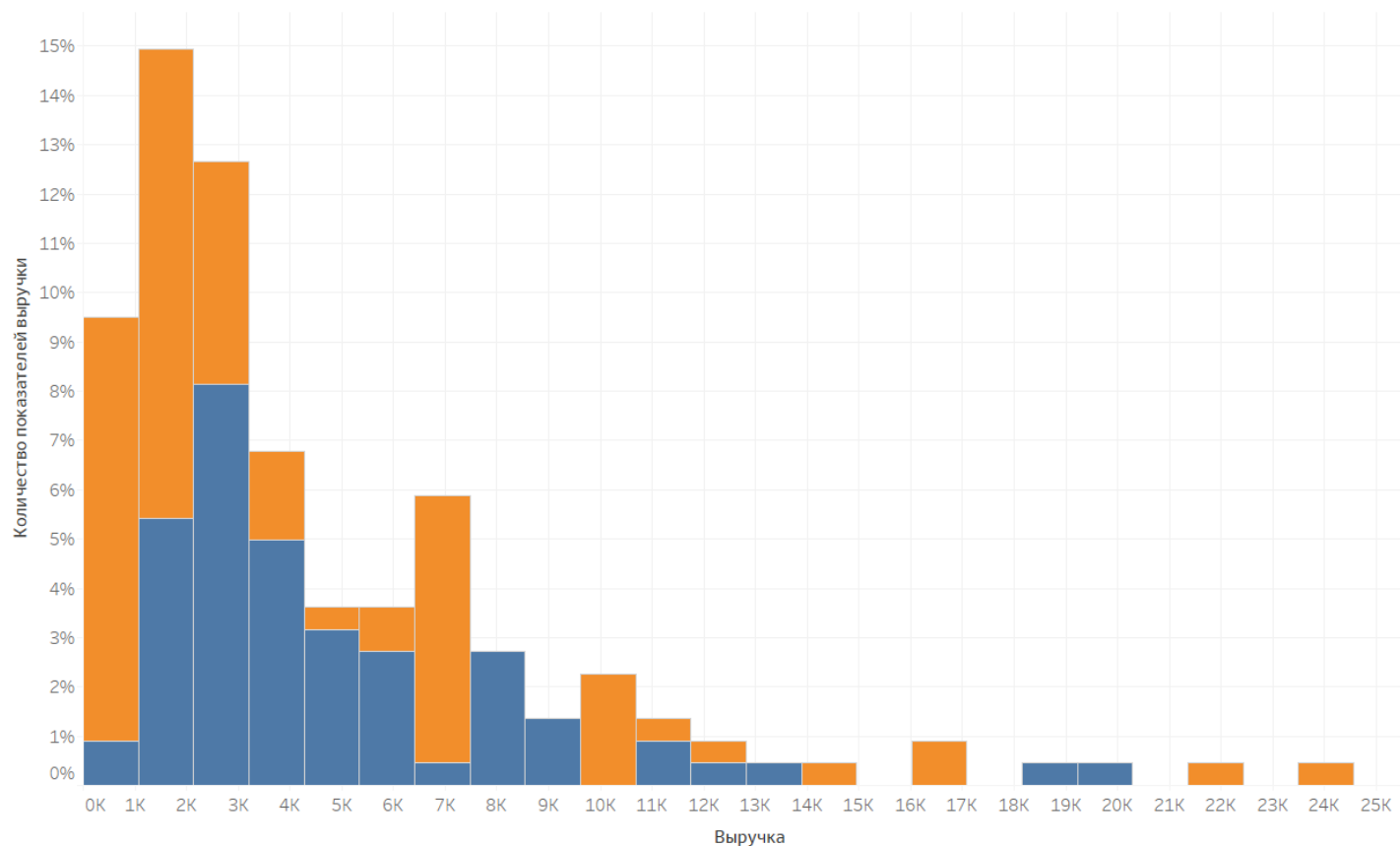
Далее рассмотрим средний чек (выручку). По графику можно заметить, что средний чек тестовой группы значительно ниже, чем у контрольной.

Значение для контрольной группы – 4775.7

Значение для тестовой группы – 4195

Выручка по группам

Гистограмма выручки по группам



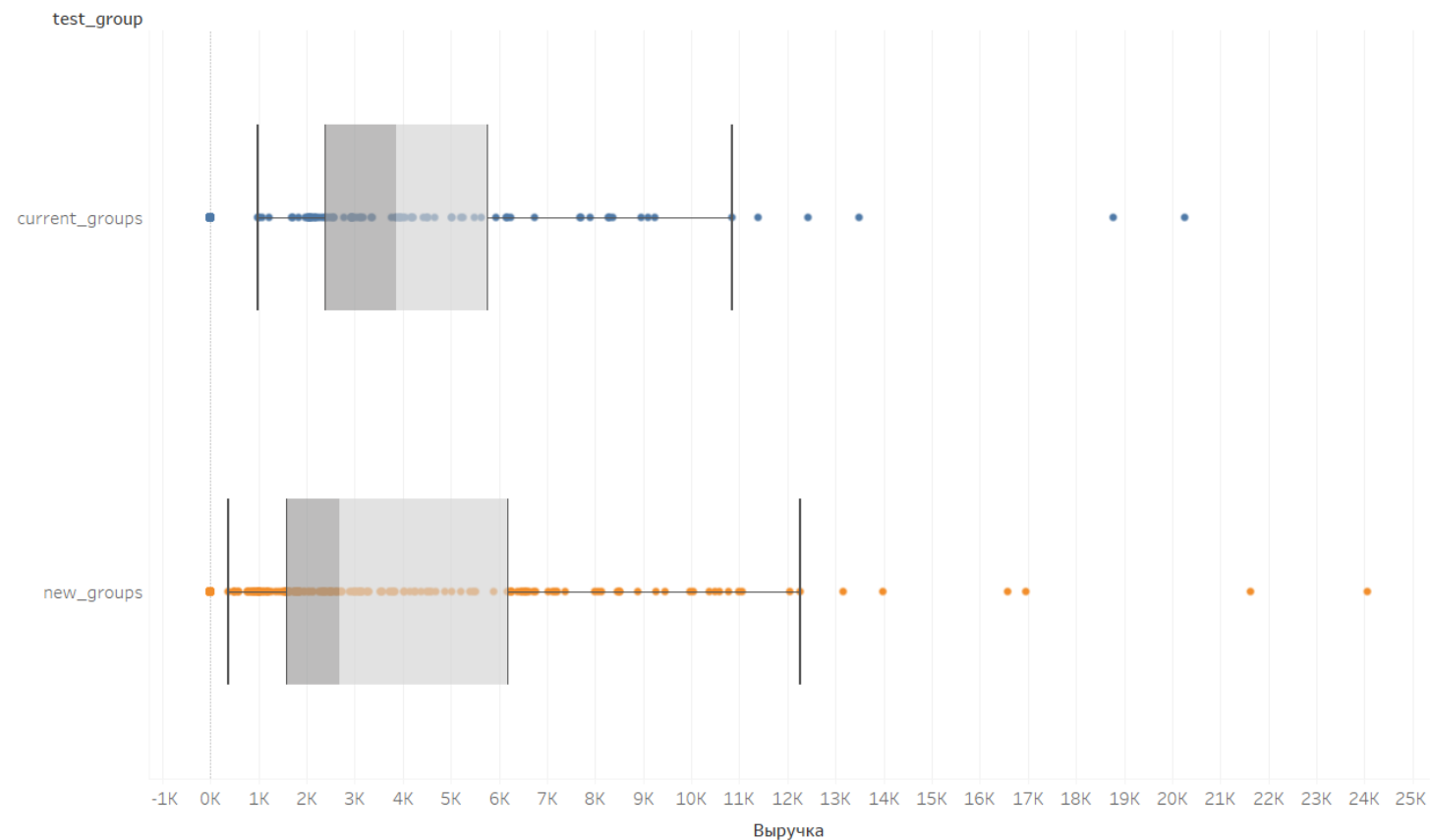
Сравним гистограммы распределения выручки по группам.

По графику можно заметить, что у тестовой группы значительно больше покупок, что коррелирует с повышением показателя конверсии в покупку. При этом также видно, что значительная часть покупок в этой группе на чек до 3000.

Посмотрим распределение выручки на диаграмме размаха.

Выручка по группам

Диаграмма размаха выручки по группам



На графике можно заметить, что в обеих группах есть 5-6 существенных выброса справа, которые могут сильно повлиять на среднее и по одному выбросу слева.

Рассмотрим численные показатели.

Для тестовой группы:

Медиана – 2682

Нижний «ус» – 366

Нижний квантиль – 1597

Верхний «ус» – 12255

Верхний квантиль – 6210

Для контрольной группы:

Медиана – 3864

Нижний «ус» – 975

Нижний квантиль – 2404

Верхний «ус» – 10836

Верхний квантиль – 5781

Тестовая выборка имеет больший размах и при этом меньшую медиану.

Анализ результатов теста

Проведем двухвыборочный z-тест для пропорций и сравним конверсию в покупку в тестовых группах

Конверсия в контрольной группе – 0.52%

Конверсия в тестовой группе – 1.1%

Размер контрольной группы – 15256

Размер тестовой группы – 15078

Уровень значимости – 0.05

Тест двусторонний

P-value равен 0.00000, поэтому можно утверждать, что эффект неслучайный— между тестовой и контрольной группами действительно есть различия.

Перейдём к оценке статистической значимости для средних чеков.

С помощью t-теста рассчитаем p-value для выручки в категории компьютеров. Для полученных значений (без неплатящих пользователей) в двустороннем тесте и уровне значимости 0.05 получаем p-value — 0.26938. Не отвергаем нулевую гипотезу, средние выручки от продаж компьютеров в обеих группах не отличаются.

Далее с помощью теста Манна-Уитни рассчитаем p-value для выручки в категории компьютеров. Для полученных значений (без неплатящих пользователей) в двустороннем тесте и уровне значимости 0.05 получаем p-value — 0.01016. Отвергаем нулевую гипотезу. Данные в выборках групп получены из разных генеральных совокупностей, их параметры отличаются.

Результаты последних двух тестов значительно отличаются. Так как в значениях мы обнаружили значительные выбросы, то тест Манна-Уитни будет более достоверным.

Выводы по результатам проведения теста

На основе проведенного анализа можно сделать следующие выводы:

1. Введение отдельной категории игровых ноутбуков положительно сказалось на конверсии в покупку компьютеров вообще (0.5% -> 1.1%).
2. Введение отдельной категории игровых ноутбуков отрицательно сказалось на среднем чеке в категории компьютеров (4775.7 -> 4195)
3. И исходя из предыдущих двух пунктов - нельзя рекомендовать выделение игровых ноутбуков к внедрению для всей аудитории, т.к. наблюдается очень серьезное падение среднего чека.

Необходимо исследовать причину падения среднего чека при повышении конверсии.

Вероятно, когда все товары находятся в одной общей категории для пользователя проще «докупать» что-то дополнительное к основной покупке. В то время как пользователи, которые видят только игровые ноутбуки и покупки совершают только по этой категории.