

Подготовка и анализ А/В-теста

Автор: Плохотников Виталий

Дата: 05.03.2024

[ссылка на дашборд](#)

Постановка задачи

- Проверить гипотезы о влиянии ввода новой подкатегории «Игровые ноутбуки» на исследуемые метрики:
 - Конверсия в покупку в категории «Компьютеры» увеличится на 100%.
 - Средний чек в категории не изменится.
- Рассчитать параметры теста, оценить корректность его проведения и проанализировать результаты эксперимента.
- Построить дашборды для каждого из этапов работы.
- По результатам теста сделать выводы и дать рекомендации.

Краткие выводы

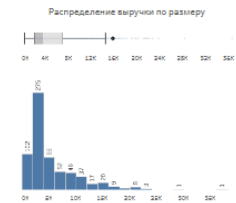
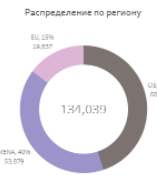
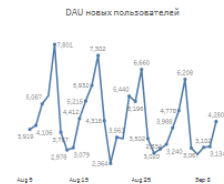
- Проанализированы исторические данные, на основе которых подготовлены параметры A/B-теста. Разработано ТЗ для проведения теста.
- Исследованы данные за 1-й день тестирования. Проведена оценка корректности проведения A/B-теста
- Проведен анализ данных по окончании A/B-теста. Проведена проверка полученных результатов на статистическую значимость.
- Сделаны выводы и рекомендации.

Общий вид дашбордов

Подготовка и проведение А/В-теста

< 1 2 3 >

Анализ исторических данных



5,421 средняя выручка

4,324 средняя стоимость пользователя

4,697 средняя стоимость привлечения

0.50% конверсия в покупку

Подготовка и проведение А/В-теста

< 1 2 3 >

Анализ данных во время проведения теста



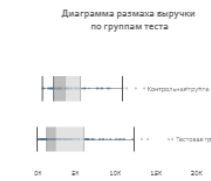
Сводная таблица участников тестирования по группам, региону, типу устройства

test_name	test_group	first_device	CIS	EU	MEA
daily_water_test	Контрольная группа	Android	309	97	243
		iPhone	110	32	136
		Mac	60	17	81
		PC	100	70	145
	Тестовая группа	Android	747	301	271
		iPhone	130	49	154
		Mac	61	20	83
		PC	163	81	142

Подготовка и проведение А/В-теста

< 1 2 3 >

Анализ результатов А/В-теста



Анализ исторических данных

Основные наблюдения

- За рассматриваемый период средняя посещаемость ресурса составляет 4 324 новых пользователя в день. Посещаемость имеет явный сезонный характер: со вторника по субботу посещаемость растет, затем резко падает в воскресенье и понедельник. Разница между минимальным и максимальным значениями лежит в диапазоне от 3 374 до 4 324 за период. Также на графике заметен тренд на понижение, причем, уменьшение количества новых пользователей в пиковые дни (пятница, суббота) более ярко выражено чем в дни минимальной посещаемости.



4 324

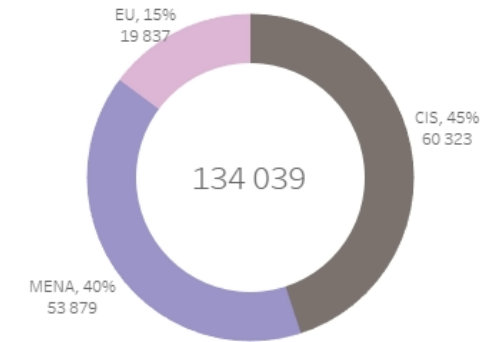
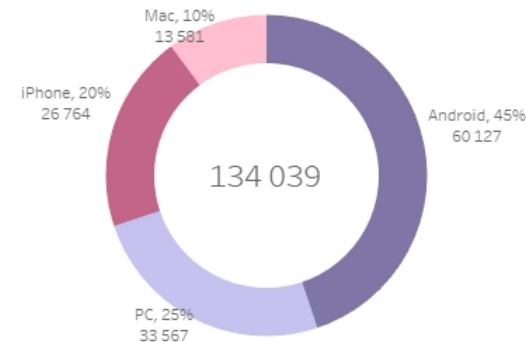
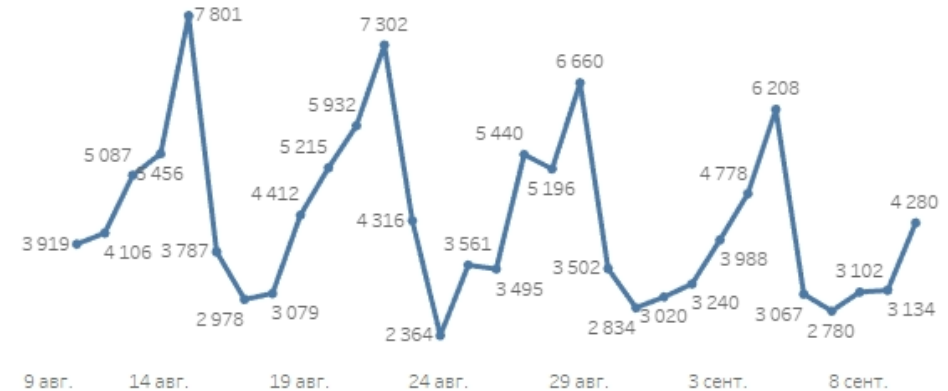
среднее количество
пользователей

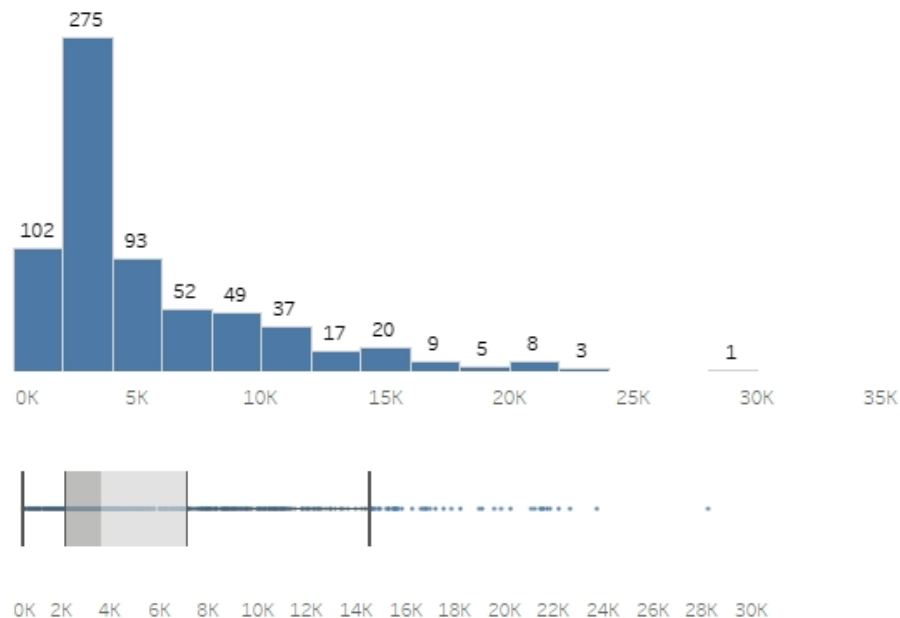


0,50%

конверсия в покупку

- Конверсия в покупку составляет 0,5%, что является вполне нормальным результатом для столь узкой категории товаров.
- Из 134 039 новых пользователей с 11.08 по 10.09 45% пользователей (60 127) зашло в маркетплейс с устройств на Android, с персональных компьютеров – 25%, с айфонов – 20% и еще 10% с макбуков. Распределение пользователей между мобильными и десктопными устройствами примерно соответствует существующим тенденциям.
- Приблизительно равные доли имеют пользователи из региона CIS (45%) и MENA (40%). Еще 10% пользователей приходят из стран EU.





- Медианное значение выручки (3 649) смещено в левую сторону относительно среднего значения (5 421), что говорит о наличии большого «правого хвоста» - покупок на значительно большие суммы. 25% покупок укладывается в диапазон 2 259 – 3 649. На гистограмме корзина с покупками от 2 000 до 4 000 имеет наибольшее количество заказов (275). Еще 25% покупок лежит в диапазоне 3 649 – 7 182. По графикам видно большое количество выбросов – покупок за гранью 14 556.
- СКО = 4 697 также показывает наличие большого разброса размеров покупок относительно среднего значения.



5 421
средняя выручка



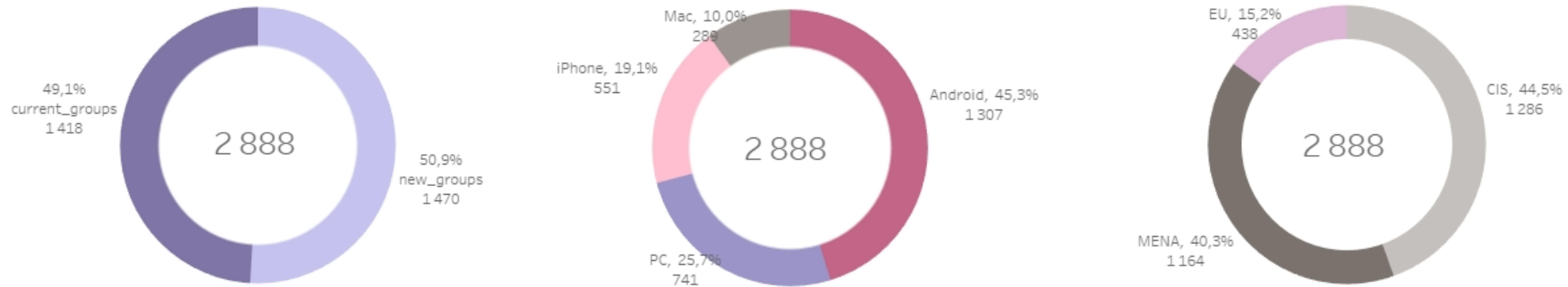
4 697
среднеквадратичное
отклонение

Техническое задание на проведение A/B-теста

- **Название:** Выделение игровых ноутбуков в отдельную категорию товаров на маркетплейсе «Всё.Техника»
- **Цель:** Проверить гипотезу, что выделение игровых ноутбуков в отдельную категорию будет способствовать увеличению спроса на товары из данной категории.
- **Метрики:**
 - Конверсия в покупку в разделе «Компьютеры». Прогноз: рост – 100%.
 - Средний чек в категории «Компьютеры» (computer equipments). Прогноз: не изменится.
- **Аудитория:** новые пользователи, все регионы.
- **Тестовые группы**
 - Количество групп – 2.
 - Названия групп: тестовая группа (new_groups), контрольная группа (current_groups).
 - Доля пользователей для каждой группы – 50%.
- **Длительность теста** – 7 дней.
- **Размер выборки** – 9 432.
- **Дата запуска теста** – 14.10.2020 (включительно).
- **Дата остановки теста** – 20.10.2020 (включительно).

Анализ данных во время проведения теста

Основные наблюдения



- За рассматриваемый период (14 октября) распределение пользователей между группами произошло в соотношении 50,9% / 49,1% (new_groups / current_groups). Отклонение от планируемых долей около 1%. Общее количество пользователей, принявших участие в тесте – 2 888, что вполне возможно для дня недели «среда» с учетом нисходящего тренда, наблюдавшегося в августе-сентябре, который мог продолжиться и в октябре.
- Отличие соотношения долей пользователей в разрезе типа устройства в день 14 октября и по историческим лежит в пределах 1%, что также подтверждает корректность сбора данных. Соотношение долей пользователей по регионам за 14 октября также аналогично соотношению по историческим данным. Какие-то другие регионы в выборку не попали, т.е. здесь данные тоже собираются корректно.

- В сводной таблице за 14 октября явных аномалий, характерных при сбое в сборе данных также не наблюдается.

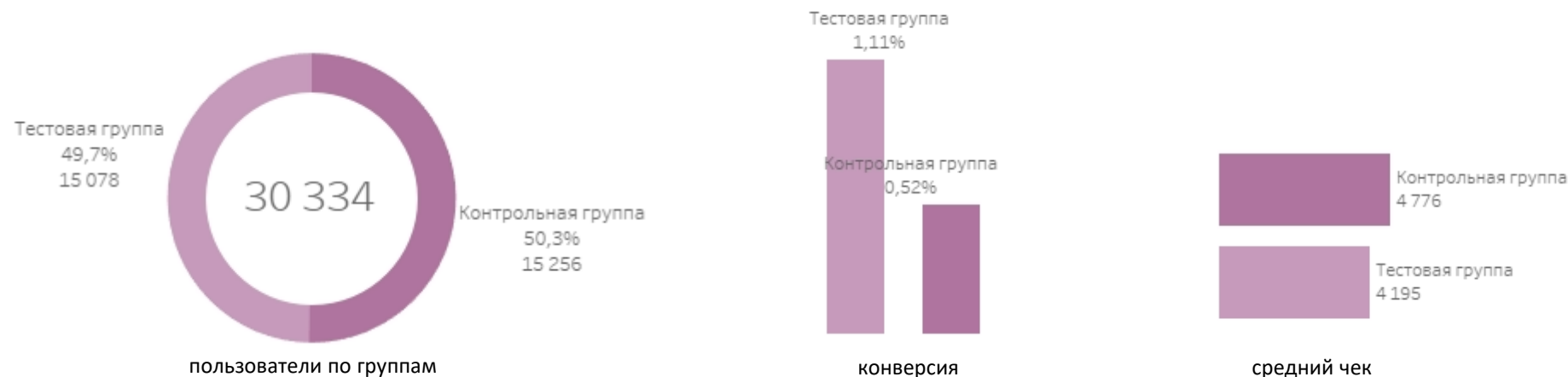
test_name	test_group	first_device	CIS	EU	MENA
gaming_laptops_test	current_groups	Android	309	97	245
		iPhone	110	32	116
		Mac	66	17	51
		PC	160	70	145
	new_groups	Android	297	86	273
		iPhone	120	49	124
		Mac	61	26	68
		PC	163	61	142

Оценка правильности проведения теста

- Применение одновыборочного z-теста для распределения пользователей по группам показало, что при выбранном уровне значимости (0.05) $p\text{-value} = 0,3338$, что не позволяет нам отвергнуть нулевую гипотезу, т.е. пользователи распределены по группам равномерно. Тест продолжаем.
- Размер накопленной выборки (2 888) пользователей также не позволяет остановить тест, т.к. необходимый объем (30 268) не набран. Тест продолжаем.

Анализ данных после завершения A/B-теста

Основные наблюдения

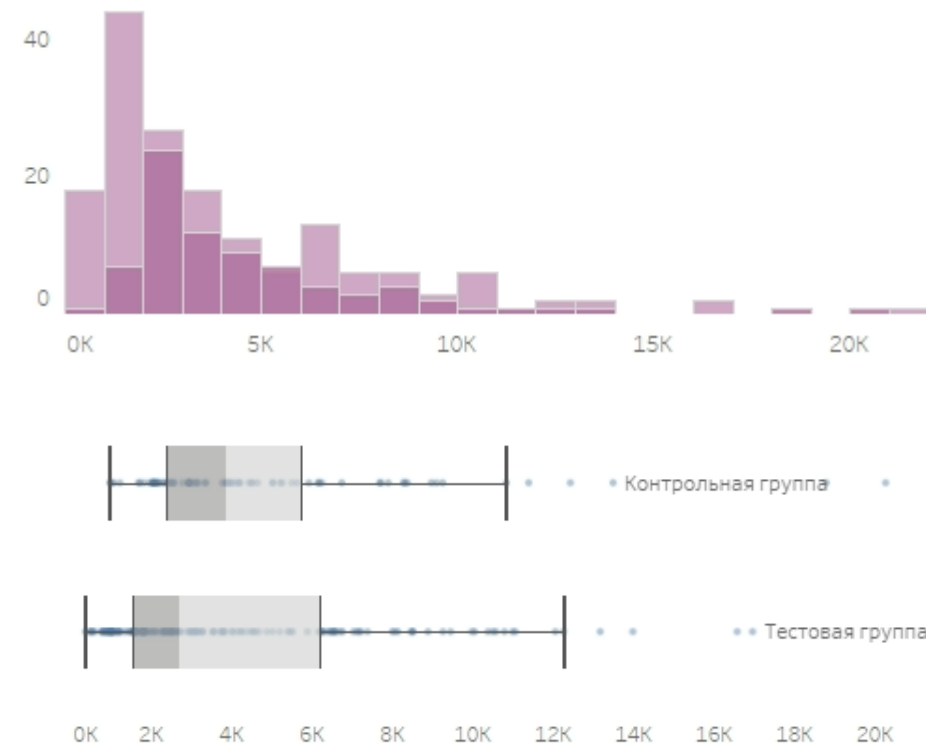


- Всего в тесте приняло участие 30 334 пользователя, которые примерно в равных пропорциях распределены между тестовой (49,7%) и контрольной (50,3%) группами. Количество участников теста значительно превосходит минимальное требуемое значение объема выборки (9 432). Так количество пользователей в тестовой (15 078) и контрольной (15 256) группах отличаются, необходимо будет проверить с помощью одновыборочного z-теста, является ли это статистически значимым.
- В тестовой группе конверсия в покупку (1,11%) в 2 раза превышает аналогичный показатель в контрольной группе (0,52%). Предварительно можно сказать, что гипотеза об увеличении конверсии в 2 раза при выделении игровых ноутбуков в отдельную категорию подтвердилась.
- Однако, нужно отметить, что в тестовой группе снизилась средняя выручка. Средний чек в тестовой группе – 4 195, а в контрольной – 4 776. Предположение о том, что при выделении ноутбуков в отдельную категорию средний чек не изменится предварительно не подтверждается.

- Выбросы присутствуют как в тестовой так и контрольной группе.
- Стоит отметить, что выделение игровых ноутбуков в отдельную категорию привело к смещению распределения выручки в левую сторону. Медиана в тестовой группе – 2 682, в контрольной – 3 864. Это произошло за счет увеличения количества покупок в тестовой группе в низком ценовом сегменте.
- В целом, можно сказать, что в тестовой группе произошло увеличение межквартильного размаха. В большей степени это заметно на нижнем квартиле. 50% покупок в тестовой группе делается в пределах сумм 1 597 – 6 210, тогда как в контрольной группе этот диапазон уже 2 404 – 5 781.

Проверка результатов на статистическую значимость

- Применение одновыборочного z-теста для распределения пользователей по группам показало, что при выбранном уровне значимости (0,05) $p\text{-value} = 0,29601$, что не позволяет нам отвергнуть нулевую гипотезу, т.е. пользователи распределены по группам равномерно.
- Двухвыборочный z-тест для пропорций при проверке полученной конверсии выдал значение $p\text{-value} = 0,0000$, что меньше уровня значимости 0,05. Это позволяет нам отвергнуть нулевую гипотезу о равенстве конверсий в сторону первой гипотезы. Конверсии действительно отличаются, и это не случайный результат.



Распределение выручки и по группам

- Применение t-теста для среднего чека при уровне значимости $\alpha=0,05$ показал результат $p\text{-value} = 0,26938$. Результат не позволяет отвергнуть нулевую гипотезу о равенстве средних чеков.
- Однако тест Манна-Уитни для среднего чека при том же уровне значимости выдал другой результат $p\text{-value} = 0,01016$. Это означает, что выборки получены из разных генеральных совокупностей, т.е. средние чеки могут отличаться. Результат, полученный t-тестом (см. выше) может быть объяснен наличием выбросов в наших выборках.
- С учетом небольшого количества выбросов и сопоставимого их количества в тестовой и контрольной группах примем решение о равенстве средних значений согласно результатам t-теста.

Выводы и рекомендации

- Выделение ноутбуков в отдельную категорию способствовало увеличению конверсии в покупку ноутбука до 1,11% (0,52% в контрольной группе).
- Однако, это также привело к небольшому снижению среднего чека покупки: 4 195 – в тестовой группе и 4 776 – в контрольной группе. Медианные значения – 2 682 и 3 864 соответственно.
- Не смотря на некоторое уменьшение значения среднего чека в тестовой группе по сравнению с контрольной группой, будем рекомендовать введение подраздела «Игровые ноутбуки» для всей аудитории. Решение основывается на результаты проверки гипотезы о равенстве средних t-тестом. Кроме того, основной метрикой A/B-теста является конверсия, а средний чек всего лишь дополнительной метрикой. Третий довод в пользу такого решения основан на понимании бизнес-задач. Изменение в каталоге товаров позволило привлечь новых покупателей, которые в дальнейшем могут принести маркетплейсу дополнительные деньги, т.к. «дотянуться» до них маркетинговыми активностями будет проще чем до тех, кто покупки в магазине не совершал. И самое главное, нововведение принесло выручку 700 570, почти в 2 раза большую чем в контрольной группе (377 279), а следовательно и доход маркетплейсу как % от продаж продавцов на платформе.
- Как отмечалось ранее, причиной уменьшения среднего чека является значительное увеличение количества покупок в низком ценовом сегменте (в среднем и высоком ценовых сегментах рост не столь существенный). Гипотеза может быть следующей. Игровые ноутбуки как более мощные и следовательно более дорогие были перенесены в отдельную категорию. Большинство покупателей ноутбуков вряд ли является игроманами. Они просто приходят в маркетплейс и покупают ноутбук. В новой ситуации они видят отдельную категорию «Игровые ноутбуки» и, игнорируя ее, идут в общий раздел. Таким образом, покупатели игнорируют категорию товаров с наиболее высокими ценами и, выбирая другие ноутбуки с более низкой ценой, снижают средний чек в категории.