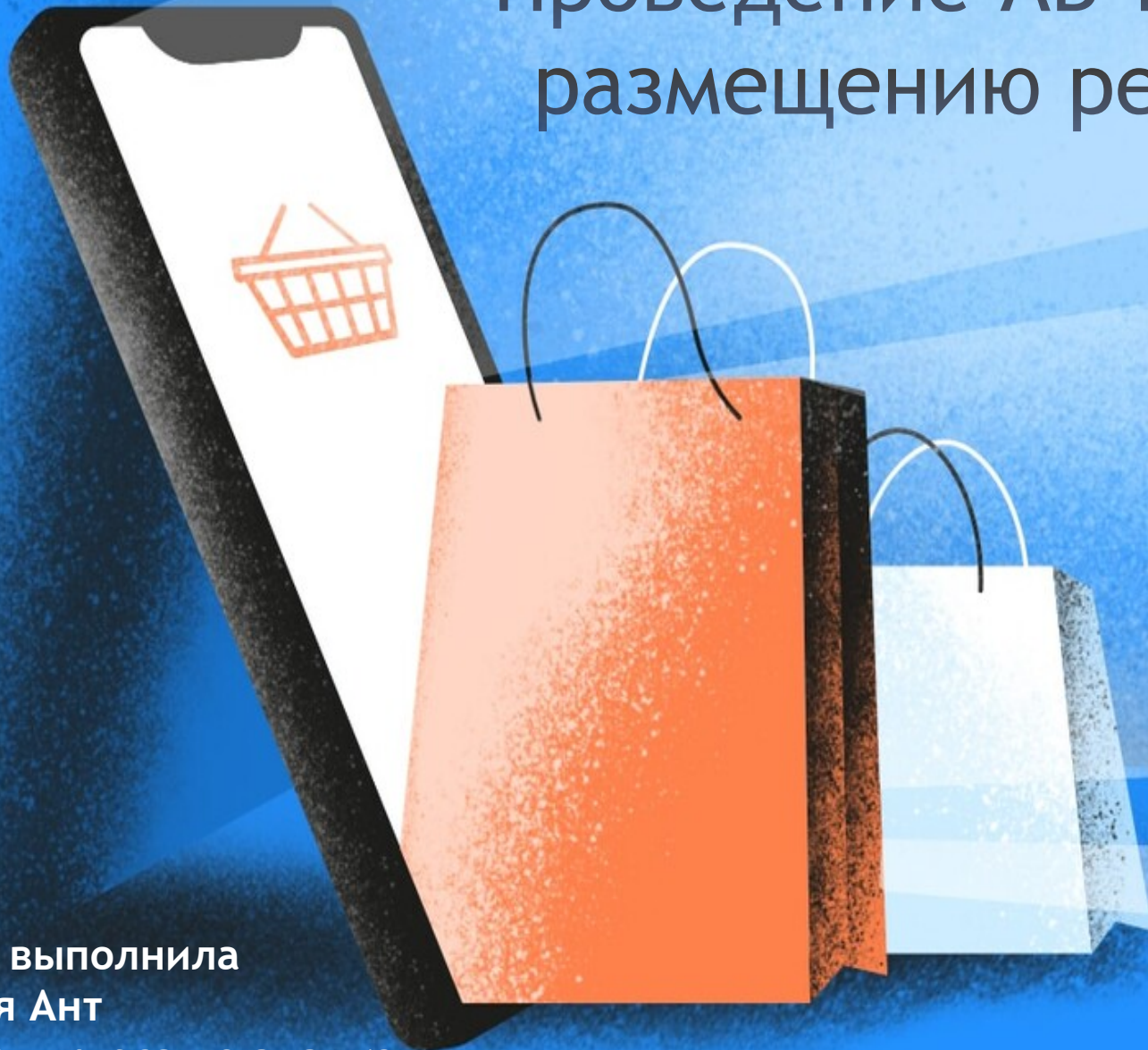


Проведение АБ теста по размещению рекламы



Анализ выполнила
Евгения Ант
в рамках курса по анализу
данных

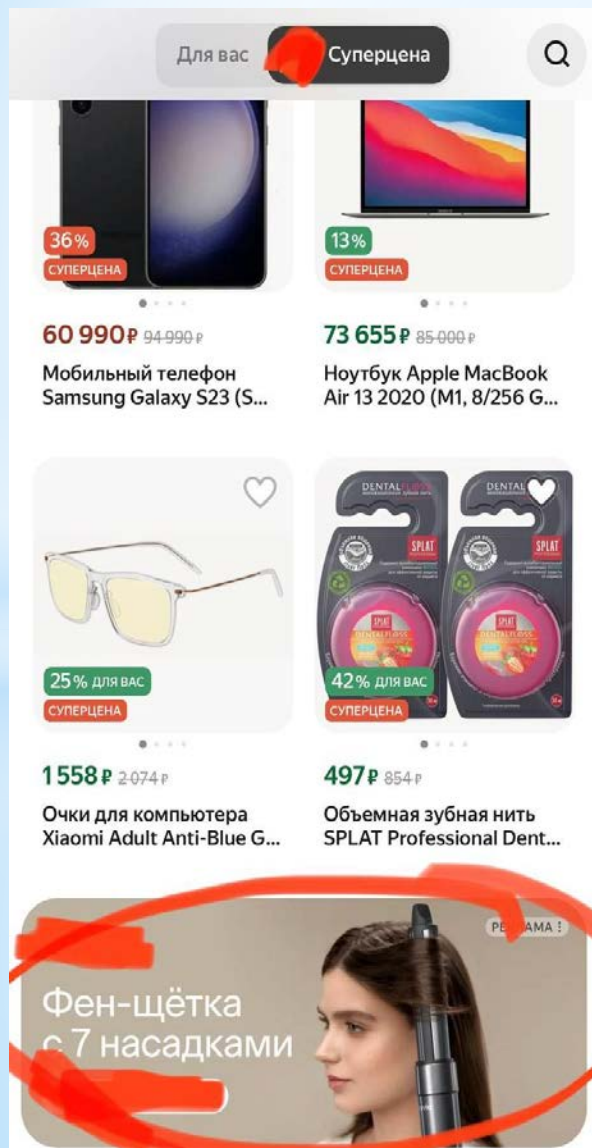
Исходные данные

Наша компания является маркетплейсом товаров. У нас в приложении есть рекламные блоки, с помощью которых продавцы могут продвигать свои товары. Также через них мы можем информировать об акциях и скидочных днях, например, о Черной пятнице.

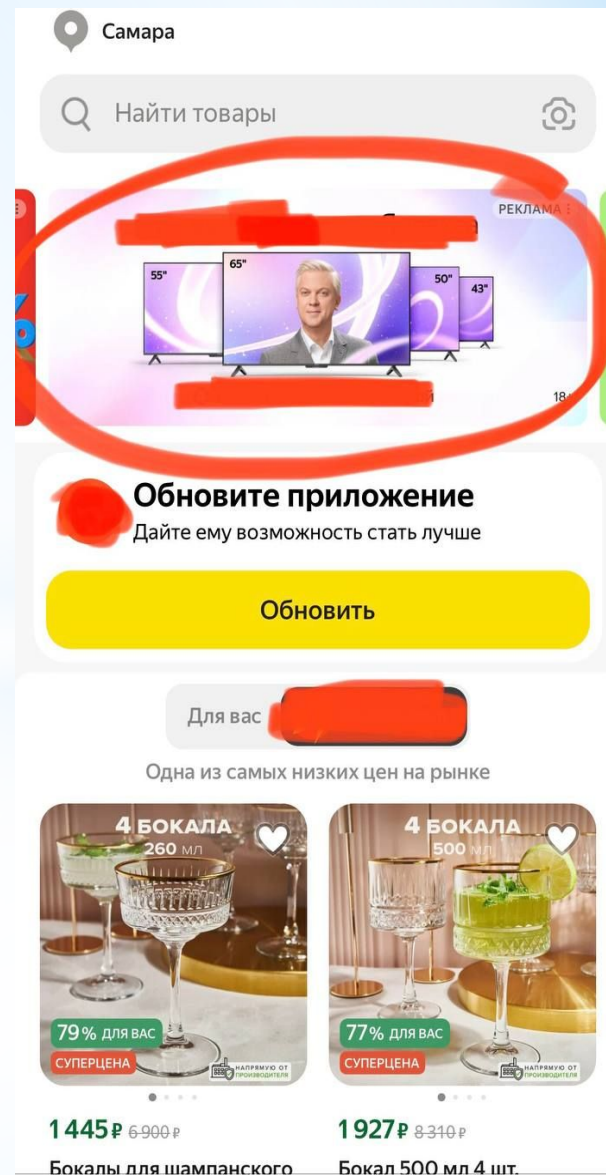
Сейчас первый рекламный блок расположен после 3 линий товаров бесконечной ленты (ленту товаров можно скроллить и рекламные блоки будут появляться среди товаров с какой-то периодичностью). В данном эксперименте мы будем рассматривать только ПЕРВЫЙ рекламный блок. К нам пришел менеджер и просит провести АБ тест по размещению рекламы и сравнить места размещения.

Сейчас первый рекламный блок размещается в ленте товаров, менеджер хочет подвинуть рекламный блок наверх и сравнить результаты. Примеры размещения рекламы ниже. При клике на рекламный блок открывается подборка товаров.

Контроль



Тест



Задача: Необходимо протестировать изменение расположения рекламы: для тест-группы секция рекламы должна будет располагаться над основной секцией.

Цель: проверить гипотезу, что изменение расположения влияет на взаимодействие пользователей с рекламой

▣ **Гипотеза:** такое расположение рекламы положительно влияет на метрики (основные - кликабельность и конверсия).

Основные метрики, которые хотим отслеживать:

- Конверсия в добавление товаров в корзину от захода в рекламу
- Кликабельность секции рекламы
- Конверсия в добавление товаров в корзину из рекламы от просмотра главной
- Конверсия в заказ от захода в рекламу
- Конверсия в заказ от просмотра главной
- Выручка на пользователя (ARPU) по товарам, добавленным из секции
- Выручка на покупателя (ARPPU) по товарам, добавленным из секции
- Успешным тестированием считаем, если ключевая метрика AB теста секции рекламы значимо увеличиваются.

▣ **Ограничения:**

Мы уже проводили подобный эксперимент и получили реальный эффект увеличения ключевой метрики (которую вы выбрали) на 2,5%. Эксперимент проводился 1 неделю. Такой же период необходимо взять и в этот раз, потому что потом в каталоге магазина планируются изменения.

Дизайн проекта.

1. Изучить пользовательское поведение до эксперимента: вывести с помощью графиков ключевые метрики пользователей (средний чек корзины, частота кликов, конверсия в добавления).
2. Сделать выводы о поведении пользователей
3. Вывести таблицу соотношения MDE и размера выборки.
4. Выбрать оптимальный MDE с учетом наших ограничений.

У нас есть файл `data_raw.csv` с полями: `client_id`, `dt`, `event_type`:

- `client_id` - идентификатор клиента

- `dt` - дата в формате YYYY-MM-DD

- `event_type` - действия пользователей:

 - `view` - это значит, что пользователь смотрел главный экран нашего магазина (вошел в приложение)

 - `click` - это значит, что пользователь кликнул на рекламу

 - `add` - это значит, что пользователь добавил товары из рекламы.

Преобразуем файл к формату воронки:

Если хотя бы раз на конкретную дату пользователь заходил на главный экран магазина, то присвоено 1, иначе - 0.

Получили файл `processed_data.csv`, следующего формата:

event_type	client_id	dt	adds	clicks	views
------------	-----------	----	------	--------	-------

0	1	2024-06-17	0	0	1
---	---	------------	---	---	---

1	1	2024-06-19	0	0	1
---	---	------------	---	---	---

2	3	2024-06-17	0	1	1
---	---	------------	---	---	---

Результат обработки файла processed_data.csv:

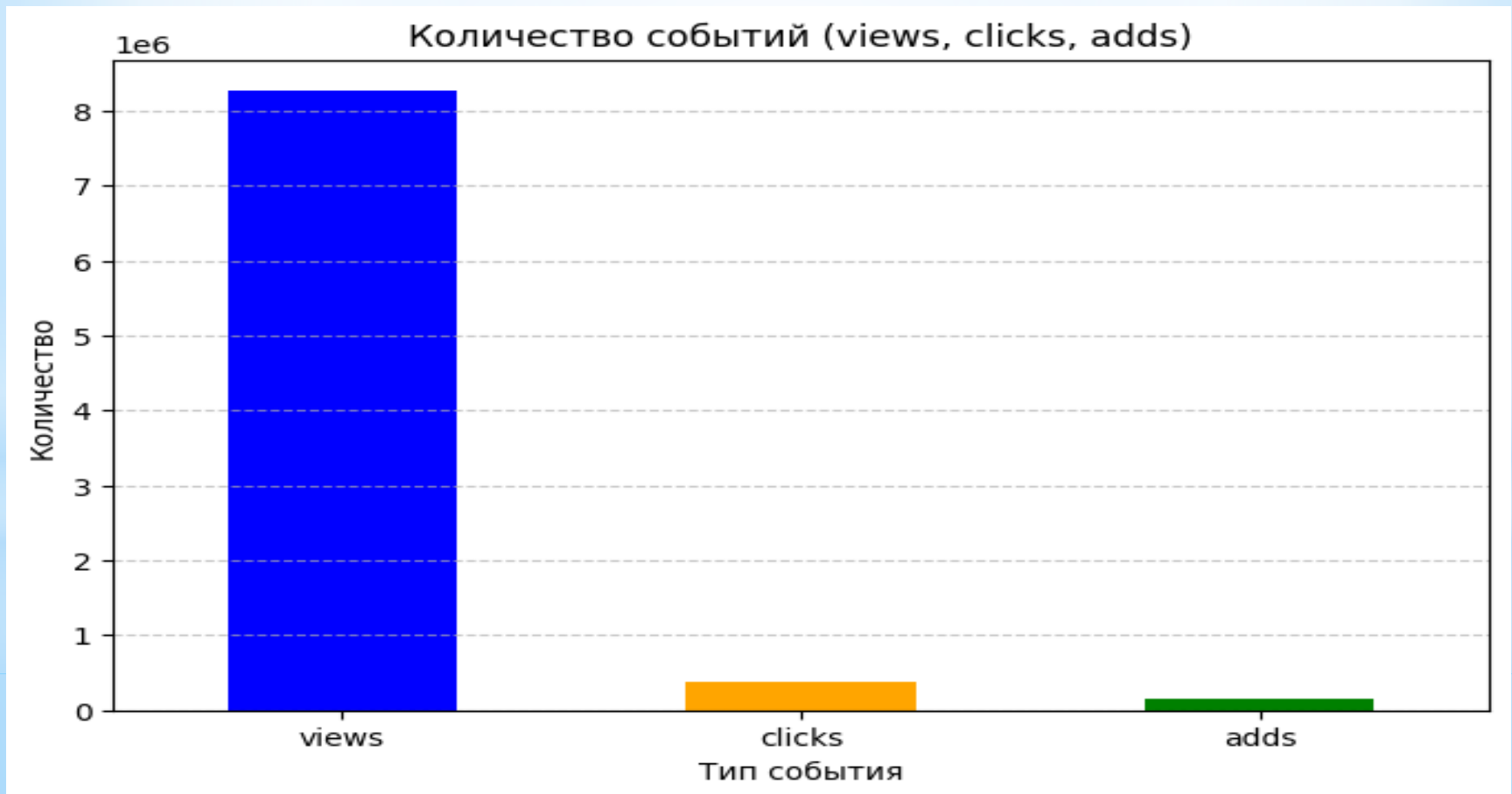
views 8 267 094

clicks 370 039

adds 148 700

CTR (конверсия в просмотр): 4.48%

CR (конверсия в добавление): 40.18%



Рассчитаем **MDE** (Минимальный наблюдаемый эффект) для CTR и CR.

Формулы расчета

Дисперсия для пропорции (бинарного показателя):

$$\sigma^2 = p(1 - p)$$

где p — текущее значение метрики (CTR или CR).

Расчет **MDE** (по формуле для двух выборок):

$$MDE = \sqrt{\frac{2 \cdot \sigma^2}{n}} \cdot (Z_{\alpha/2} + Z_{\beta})$$

где:

σ^2 — дисперсия метрики.

n — размер выборки.

$Z_{\alpha/2}$ — квантиль нормального распределения для уровня значимости (например, для 95% доверия $\alpha=0.05 \Rightarrow Z_{0.025}=1.96$).

Z_{β} — квантиль нормального распределения для мощности теста (обычно 80%, значит $\beta=0.2 \Rightarrow Z_{0.2}=0.84$).

▣ Результаты расчета **MDE**

MDE для CTR: 0.03%

MDE для CR: 0.07%

Выводы:

Для CTR (конверсия в просмотр)

Если разница в CTR между группами меньше 0.03%, тест не сможет ее достоверно зафиксировать.

То есть **тест выявит только значимые изменения** в поведении пользователей.

Для CR (конверсия в добавление)

Тест обнаружит разницу в CR только от 0.07% и выше.

Меньшие изменения останутся статистически незначимыми.

Обоснование выбора MDE

Если мы ожидаем небольшие изменения в CTR и CR (например, за счет UX-изменений), **нам подойдет такой MDE.**

Если нужно выявить даже минимальные колебания, придется **увеличивать выборку** для снижения MDE.

Построим таблицу зависимости MDE ↔ размер выборки и выберем оптимальный вариант с учетом наших ограничений (не более 20% пользователей).

Формирование таблицы MDE и размера выборки

	Выборка	MDE_CTR	MDE_CR
0	10000	0.005795	0.013735
1	20000	0.004098	0.009712
2	30000	0.003346	0.007930
3	40000	0.002898	0.006868
4	50000	0.002592	0.006143
5	60000	0.002366	0.005607
6	70000	0.002190	0.005191
7	80000	0.002049	0.004856
8	90000	0.001932	0.004578
9	100000	0.001833	0.004343
10	110000	0.001747	0.004141
11	120000	0.001673	0.003965
12	130000	0.001607	0.003809
13	140000	0.001549	0.003671
14	150000	0.001496	0.003546
15	160000	0.001449	0.003434
16	170000	0.001406	0.003331
17	180000	0.001366	0.003237
18	190000	0.001330	0.003151
19	200000	0.001296	0.003071

Анализ таблицы MDE и размеров выборки

Тренд изменения MDE

По мере увеличения размера выборки MDE (минимальный детектируемый эффект) уменьшается.

Это ожидаемо, так как при большем объеме данных дисперсия снижается, и можно выявить более мелкие эффекты.

MDE_CTR (Минимальный детектируемый эффект для CTR)

При выборке 10,000 пользователей $MDE_CTR \approx 0.0058$ (0.58%).

При выборке 100,000 пользователей $MDE_CTR \approx 0.00183$ (0.183%).

При 200,000 пользователей $MDE_CTR \approx 0.001296$ (0.13%).

Это значит, что если мы хотим обнаружить разницу в CTR на уровне 0.13%, нам понадобится 200,000 пользователей.

MDE_CR (Минимальный детектируемый эффект для CR)

При выборке 10,000 пользователей $MDE_CR \approx 0.0137$ (1.37%).

При выборке 100,000 пользователей $MDE_CR \approx 0.00434$ (0.43%).

При 200,000 пользователей $MDE_CR \approx 0.00307$ (0.31%).

Это значит, что если мы хотим обнаружить разницу в CR на уровне 0.31%, нам нужно 200,000 пользователей.

Выбор оптимального MDE с учетом ограничений

Мы хотим затронуть не более 20% пользователей.

Если всего в системе, например, 1,000,000 пользователей, то доступно для эксперимента 200,000 пользователей.

При 200,000 пользователей мы сможем определять разницу в CTR от 0.13% и в CR от 0.31%.

Если мы готовы анализировать только 100,000 пользователей, то $MDE_CTR \approx 0.18\%$, а $MDE_CR \approx 0.43\%$.

Выводы

Чем больше выборка, тем меньшие изменения в CTR и CR мы можем зафиксировать.

При ограничении в 200,000 пользователей мы можем зафиксировать разницу в CTR от 0.13% и в CR от 0.31%.

Оптимальный размер выборки зависит от ожидаемого эффекта.

Если ожидаемый эффект в CTR более 0.2%, то можно взять выборку 100,000.

Если ожидаемый эффект в CTR менее 0.15%, то потребуется выборка 200,000.

Если разница в конверсии ожидается в пределах 0.3-0.4%, достаточно выборки в 100,000.

Мы согласовали эксперимент и запустили его.
Прошла неделя и пришло время подвести результаты.

У нас есть файл с результатами [final_results_to_analyze.xlsx](#)

Столбцы:

date дата

client_id клиент

ab_group группа эксперимента

is_view_ads флаг, были ли просмотры рекламы

cnt_view_ads количество просмотров рекламы

is_adds_ads флаг наличия добавлений из рекламы

cnt_adds_ads количество добавлений из рекламы

sum_adds_ads сумма добавлений из рекламы (стоимость
добавленных товаров в рублях)

is_orders_ads флаг наличия заказа из рекламы

cnt_orders_ads количество заказов из рекламы

sum_orders_ads сумма заказов из рекламы (стоимость
заказанных товаров в рублях)

- Рассчитали основные метрики (CTR, CR_Adds, CR_Orders, ARPU, ARPPU).
- Разделили данные на тестовую и контрольную группы.
- Провели t-тесты для оценки значимости различий в CTR, CR_Adds и выручке.
- Определили, являются ли различия статистически значимыми ($p\text{-value} < 0.05$).

Метрики по группам:

	ab_group	total_users	total_views	total_clicks	total_adds	total_orders
0	control	200786	25389	43469	15799	5579
1	test	200474	25852	43805	16337	5881

	total_revenue	CTR	CR_Adds	CR_Orders	ARPU	ARPPU
0	1589899	1.712119	0.363454	0.128344	7.918376	284.979208
1	1687895	1.694453	0.372948	0.134254	8.419521	287.008162

Результаты статистического теста:

```
'CTR p-value': 0.7079439461730692
'CR_Adds p-value': 0.16736601313485483
'Revenue p-value': 0.042774140139963536
'CTR Significant': False
'CR_Adds Significant': False
'Revenue Significant': True
```

Вывод по результатам A/B-теста:

CTR (Кликабельность рекламы):

Контрольная группа: 1.71

Тестовая группа: 1.69

p-value = 0.7079 (значительно больше 0.05)

Вывод: Различие в CTR между группами не является статистически значимым. Внедренные изменения не повлияли на кликабельность рекламы.

CR_Adds (Конверсия в добавление товара в корзину):

Контрольная группа: 36.35%

Тестовая группа: 37.29%

p-value = 0.1674 (больше 0.05)

Вывод: Различие в конверсии добавления товара в корзину не является статистически значимым. Это означает, что пользователи в тестовой группе добавляют товары чуть чаще, но разница могла возникнуть случайно.

Выручка (Total Revenue):

Контрольная группа: 1,589,899 руб.

Тестовая группа: 1,687,895 руб.

p-value = 0.0428 (меньше 0.05)

Вывод: Различие в выручке является статистически значимым, что означает, что изменения в тестовой группе действительно привели к увеличению дохода.

Общий итог:

Изменения не повлияли на CTR (кликабельность рекламы) и конверсию добавлений в корзину.

Однако выручка в тестовой группе увеличилась статистически значимо. Это может быть связано с более дорогими товарами в заказах, изменением пользовательского поведения или другими факторами.

Для принятия решения стоит изучить поведение пользователей в тестовой группе подробнее (например, средний чек, категории товаров и т. д.).

Рекомендация:

Так как выручка увеличилась значимо, можно рассмотреть внедрение тестового варианта в продакшн, но параллельно провести дополнительный анализ, чтобы понять, почему произошел рост выручки.

Проведем дополнительный анализ по группам:

Метрики по группам:

	ab_group	total_users	total_views	total_clicks	total_adds	total_orders
0	control	200786	25389	43469	15799	5579
1	test	200474	25852	43805	16337	5881

	total_revenue	avg_order_value	avg_orders_per_user	avg_cart_value
0	1589899	3.245632	0.011389	5.026804
1	1687895	3.433109	0.011962	5.138580

	avg_adds_per_user	CTR	CR_Adds	CR_Orders	ARPU	ARPPU
0	0.032252	1.712119	0.363454	0.128344	7.918376	284.979208
1	0.033229	1.694453	0.372948	0.134254	8.419521	287.008162

Результаты статистического теста:

```
'cnt_view_ads p-value': 0.7079439461730692,  
'cnt_adds_ads p-value': 0.16736601313485483,  
'cnt_orders_ads p-value': 0.04059213036963724,  
'sum_orders_ads p-value': 0.042774140139963536
```

Значимые различия:

```
'cnt_view_ads p-value': False  
'cnt_adds_ads p-value': False  
'cnt_orders_ads p-value': True  
'sum_orders_ads p-value': True
```

Выводы по результатам A/B-теста:

Основные метрики:

Кликабельность рекламы (CTR): Практически не отличается между группами (1.71% в контрольной, 1.69% в тестовой), $p\text{-value} = 0.707 \rightarrow$ статистически незначимо.

Конверсия в добавление (CR_Adds): Немного выше в тестовой группе (0.373 против 0.363), но $p\text{-value} = 0.167 \rightarrow$ разница незначима.

Конверсия в заказ (CR_Orders): Немного выше в тестовой группе (0.134 против 0.128), но разница небольшая.

Выручка и средние показатели:

Общая выручка (Total Revenue): В тестовой группе выше (1,687,895 против 1,589,899).

Средний чек (Avg Order Value): Немного выше в тестовой группе (3.43 против 3.24).

Средняя сумма добавленных товаров в корзину (Avg Cart Value): Практически одинакова (5.13 против 5.03).

Средняя частота добавления товаров (Avg Adds per User): Незначительно выше в тестовой группе.

Статистическая значимость различий:

Значимые различия обнаружены в количестве заказов ($p = 0.040$) и общей выручке ($p = 0.042$), что говорит о реальном эффекте тестовой группы на заказы и продажи.

Просмотры рекламы ($p = 0.707$) и добавления в корзину ($p = 0.167$) статистически не различаются, что означает, что тестовая группа не повлияла на вовлечение в начальных этапах воронки.

Заключение

- ▣ Тестовая группа показывает значимое увеличение количества заказов и общей выручки.
- ▣ Однако влияние на клики и добавления в корзину незначительно.
- ▣ **Рекомендация:** можно развернуть тестовое изменение на всех пользователей, если цель — увеличение заказов и выручки, но при этом проанализировать, почему не изменилось вовлечение на ранних этапах.

После запуска изменений на всех пользователей продолжить мониторинг ключевых метрик, особенно:

Выручку и средний чек - подтвердить, что рост сохраняется.

Конверсию на разных этапах (добавления, заказы) - проверить, не меняется ли поведение пользователей после масштабирования.

Влияние на LTV и повторные покупки - возможно, тестовая группа дает не только краткосрочный эффект.

Можно запустить дополнительный анализ через неделю-две после полного развертывания.

Спасибо за внимание!



Евгения Ант
email:
telegram:
tel.: