

Fefelov_Final Project-1

Тест-план для А/В тестування нового підходу онлайн-реклами

1. Загальна інформація про експеримент

1.1. Контекст та бізнес-проблема

Команда вирішила випробувати новий підхід до розміщення онлайн-реклами продукту з метою покращення ефективності маркетингових витрат та збільшення конверсії користувачів у покупців.

1.2. Об'єкт тестування

- **Що тестуємо:** Новий підхід до розміщення онлайн-реклами
- **Де тестуємо:** Онлайн реклами платформи (імовірно Facebook/Instagram Ads або аналогічні)
- **На ком тестуємо:** Користувачі, які бачать нашу рекламу та відвідують вебсайт

2. Гіпотеза

2.1. Формульовання гіпотези

Нульова гіпотеза (H_0): Новий підхід до розміщення реклами не впливає на ключові метрики ефективності кампанії (CTR, конверсія у покупку, вартість придбання покупця).

Альтернативна гіпотеза (H_1): Новий підхід до розміщення реклами покращує ключові метрики ефективності кампанії порівняно з поточним підходом.

2.2. Передбачуваний ефект

Очікується, що новий підхід може:

- Підвищити Click-Through Rate (CTR) на 5-15%
- Покращити конверсію з кліку у покупку на 3-10%

- Знизити вартість залучення покупця (Cost per Purchase)
- Підвищити якість трафіку (більш цільова аудиторія)

3. Дизайн експерименту

3.1. Тип тесту

A/B тест з двома групами:

- Контрольна група (A): Використовує поточний підхід до розміщення реклами
- Тестова група (B): Використовує новий експериментальний підхід

3.2. Метод рандомізації

- Рівень рандомізації: На рівні рекламних кампаній/ad sets
- Розподіл трафіку: 50% / 50% між контрольною та тестовою групами
- Спосіб розподілу: Незалежні рекламні кампанії з однаковими бюджетними обмеженнями

3.3. Одиниця аналізу

- Primary unit: День (daily metrics)
- Secondary unit: Окремий користувач (для глибинного аналізу)

4. Первинна та вторинні метрики

4.1. Первинна (Primary) метрика

Click-Through Rate (CTR):

- Визначення: Відсоток користувачів, які клікнули на рекламу відносно кількості показів
- Формула: $CTR = (\text{Website Clicks} / \text{Impressions}) \times 100\%$
- Чому первинна: Перший індикатор ефективності реклами; швидко реагує на зміни; достатньо обсягів для статистичної значущості

Базовий рівень (Control): 4.86%

Очікуване покращення: +5-15% відносно (до ~5.5% абсолютно)

4.2. Ключові вторинні метрики

1. Конверсія Add to Cart → Purchase (Cart-to-Purchase Rate)

- **Визначення:** Відсоток користувачів, які зробили покупку після додавання товару в кошик
- **Формула:** $(\# \text{ of Purchase} / \# \text{ of Add to Cart}) \times 100\%$
- **Базовий рівень:** 40.21%
- **Чому важлива:** Показує якість залученого трафіку та його намір придбати продукт

2. Загальна конверсія (Overall Conversion Rate)

- **Визначення:** Відсоток користувачів, які зробили покупку після кліку на рекламу
- **Формула:** $(\# \text{ of Purchase} / \# \text{ of Website Clicks}) \times 100\%$
- **Базовий рівень:** 9.83%
- **Чому важлива:** Комплексний показник ефективності всієї воронки

3. Cost per Purchase (Вартість залучення покупця)

- **Визначення:** Середня вартість залучення одного покупця
- **Формула:** Spend / # of Purchase
- **Базовий рівень:** \$4.41
- **Чому важлива:** Ключова фінансова метрика для оцінки ROI

5. Guardrail метрики (метрики захисту)

Метрики, які не повинні погіршитися в результаті експерименту:

5.1. Фінансові guardrails

1. Total Spend (контроль бюджету)

- Тестова група не повинна перевищувати бюджет контрольної більш ніж на 20%
- Моніторинг щоденних витрат

2. ROAS proxy (Return on Ad Spend)

- Кількість покупок на долар витрат не повинна суттєво погіршитись
- Мінімально допустиме погіршення: -15%

5.2. Якісні guardrails

3. Reach / Impressions ratio

- Унікальність охоплення повинна залишатись на прийнятному рівні
- Не допускається зниження нижче 70%

4. Bounce rate / Engagement proxy

- Користувачі повинні продовжувати взаємодіяти з сайтом (Searches, View Content)
- Моніторинг співвідношення Clicks → Searches

6. Розрахунок розміру вибірки та тривалості

6.1. Параметри для розрахунку

Для первинної метрики (CTR):

- Базовий рівень (r_1): 4.857%
- Minimum Detectable Effect (MDE): 10% відносного покращення → 5.343%
- Significance level (α): 0.05
- Statistical power (1- β): 0.80
- Тип тесту: Двосторонній (two-tailed)

6.2. Обсяг даних

Наявні дані з експерименту:

- Контрольна група: 29 днів (через відсутність даних 5 серпня)
- Тестова група: 30 днів
- Загальний період: 1-30 серпня 2019

6.3. Достатність вибірки

Агреговані обсяги:

- Control Impressions: 3,177,233
- Test Impressions: 2,237,544
- Control Clicks: 154,303
- Test Clicks: 180,970

Висновок: Обсяг даних **БІЛЬШ НІЖ ДОСТАТНІЙ** для виявлення статистично значущих відмінностей навіть при малих ефектах ($MDE < 5\%$).

6.4. Тривалість експерименту

Рекомендована тривалість: 28-30 днів

Обґрунтування:

- Мінімум 4 повні тижні для врахування тижневих патернів (weekday vs weekend)
- Достатньо для накопичення статистично значущої кількості конверсій
- Зменшує вплив випадкових флюктуацій та зовнішніх подій
- Дозволяє виявити novelty effect (ефект новизни)

Фактична тривалість проведеного експерименту: 30 днів (1-30 серпня 2019)

7. Критерії успіху та прийняття рішення

7.1. Умови для визнання тесту успішним

Тест вважається успішним, якщо виконуються ВСІ наступні умови:

Обов'язкові умови:

1. Primary метрика (CTR): Статистично значуще покращення ($p\text{-value} < 0.05$)
2. Ключова вторинна метрика (Cart → Purchase): Статистично значуще покращення або без погіршення
3. Guardrail метрики: Жодна не погіршилась за встановлені межі
4. Мінімальна тривалість: Експеримент тривав мінімум 28 днів

Додаткові бажані умови: 5. Покращення Cost per Purchase (бажано, але не критично) 6. Загальна конверсія не погіршилась 7. Жодних технічних проблем або аномалій в даних

7.2. Рішення за результатами

Сценарій	Рішення
Primary метрика покращилася, guardrails OK	<input checked="" type="checkbox"/> Запустити новий підхід на 100%

Сценарій	Рішення
Primary метрика покращилася, але guardrail порушенено	🔄 Доопрацювати та повторити тест
Primary метрика без змін ($p > 0.05$)	✗ Залишити поточний підхід
Primary метрика погіршилась	✗ Відхилити новий підхід
Неоднозначні результати	🔄 Провести додатковий тест або збільшити вибірку

8. Ризики та план мітігації

8.1. Ідентифіковані ризики

Ризик	Імовірність	Вплив	Мітігація
Сезонність (серпень)	Висока	Середній	Повторити тест в інший сезон для валідації
Зовнішні події	Середня	Високий	Моніторинг новин, аномалій в даних
Технічні збої трекінгу	Низька	Високий	Щоденна перевірка повноти даних, алерти
Novelty effect	Середня	Середній	Подовжений період тесту (30 днів), пост-ланч моніторинг
Бюджетні обмеження	Низька	Середній	Встановлення daily budget caps
Simpson's paradox	Низька	Високий	Сегментований аналіз (device, geo, demographics)

8.2. План реагування на проблеми

Якщо виявлено аномалії в даних:

- Зупинити тест для технічної перевірки
- Виключити аномальні дні з аналізу (як день 5 серпня)
- За необхідності - перезапустити тест

Якщо бюджет перевищено:

- Призупинити кампанію з вищими витратами
- Перерахувати cost metrics та оцінити ROAS

Якщо guardrail метрика порушена:

- Негайно зупинити тест
- Провести root cause analysis
- Доопрацювати підхід перед повторним тестуванням

9. План моніторингу під час експерименту

9.1. Частота перевірок

- Щодня (перші 3 дні): Інтенсивний моніторинг для виявлення критичних проблем
- Кожні 3 дні (дні 4-14): Регулярні перевірки метрик та guardrails
- Щотижня (дні 15-30): Проміжний аналіз прогресу до статистичної значущості
- Фінальний аналіз (день 30+): Повний статистичний аналіз та прийняття рішення

9.2. Дашборд та інструменти

Ключові метрики для моніторингу:

- Daily Spend по групах
- Cumulative CTR
- Running p-values (обережно, з поправкою на multiple testing)
- Guardrail metrics trends

Інструменти:

- Автоматизований дашборд з щоденним оновленням
- Alerts при значних відхиленнях

- Python/Excel для статистичного аналізу

10. Пост-експериментальний аналіз

10.1. Додаткові аналізи після завершення

1. Сегментований аналіз:

- По днях тижня (weekday vs weekend)
- По періодах (перша vs друга половина місяця)
- За наявності даних: по device type, географії, демографії

2. Аналіз воронки:

- Детальне порівняння проходження кожного етапу воронки
- Ідентифікація вузьких місць
- Drop-off rates на кожному кроці

3. Cohort аналіз (якщо можливо):

- LTV користувачів, залучених через різні підходи
- Retention rates
- Repeat purchase rates

10.2. Документація рішення

Після завершення аналізу підготувати:

- Підсумковий звіт з результатами
- Візуалізації порівняння метрик
- Рекомендації щодо roll-out або відхилення
- Lessons learned для майбутніх експериментів

11. Етична та юридична складова

11.1. Згоди користувачів

- Користувачі не повинні знати, що вони є частиною експерименту (стандартна практика для A/B тестів реклами)

- Жодних змін на сайті або в UX - тільки спосіб показу реклами

11.2. Конфіденційність даних

- Всі дані агреговані та знеособлені
- Дотримання GDPR та інших регуляцій щодо персональних даних
- Безпечно зберігання та обробка даних

12. Критерії валідності тесту

Тест вважається валідним, якщо:

- Рандомізація: Групи сформовані випадково без systematic bias
- Відсутність contamination: Контрольна та тестова групи не перетинаються
- Стабільність умов: Відсутні значні зовнішні зміни під час тесту
- Повнота даних: Мінімальна кількість missing data (виключено лише 1 день з 30)
- Відсутність fraud: Жодних аномалій, що вказують на bot traffic або fraud
- Sample Ratio Mismatch: Співвідношення розмірів груп відповідає очікуваному

13. Висновок та готовність до експерименту

13.1. Чеклист готовності

- Гіпотеза чітко сформульована
- Primary та secondary метрики визначені
- Guardrail метрики встановлені
- Розрахований розмір вибірки та тривалість
- План моніторингу підготовлений
- Ризики ідентифіковані з планом мітігації
- Критерії успіху та прийняття рішення затверджені
- Технічна інфраструктура для трекінгу готова
- Stakeholders погодили план

13.2. Наступні кроки

1. Затвердження тест-плану - Завершено
2. Налаштування трекінгу - Готово (tracking plan в Task 2)
3. Запуск експерименту - Проведено (1-30 серпня 2019)

4. ➤ Аналіз результатів - В процесі (Task 3-5)
5. ➤ Прийняття рішення - Після аналізу