

Заголовок: "A/B-тест нової кнопки оплати"

Дата початку: //_____

Дата завершення: //_____

Відповідальні: Сніжана Заярна

Опис

Мета тесту: підвищити `begin_checkout` на 5% і не погіршити `session with orders`.

Гіпотеза: ми віримо, що якщо зробимо кнопку в оплату більшою, то збільшимо конверсію початку проходження оплати та не зменшимо кількість сесій з покупками.

Варіанти

Варіант A (Контрольний): без змін. Кнопка про оплату звичайного розміру.

Варіант B (Експериментальний): зі змінами. Кнопку про оплату зробили більшою.

Вибірка

Цільова аудиторія: усі відвідувачі сайту

Розподіл трафіку: 50% на варіант А, 50% на варіант В.

Метрики

Основна метрика: `begin_checkout / session`, ріст на 5%.

Додаткові метрики: `session with orders / session`, падіння не спостерігається.

Результати

Event Name	Value		Value (%)
	Test Group	1	
add_payment_info	1,988	2,229	12.12%
add_shipping_info	3,034	3,221	6.16%
add_to_cart	1,395	1,366	-2.08%
begin_checkout	3,784	4,021	6.26%
first_visit	30,596	30,512	-0.27%
new_account	3,823	3,681	-3.71%
page_view	191,543	198,050	3.40%
scroll	73,244	73,376	0.18%
session	45,362	45,193	-0.37%
session_start	45,905	45,649	-0.56%
session_with_orders	4,514	4,526	0.27%
user_engagement	171,788	179,081	4.25%
view_item	62,335	65,337	4.82%
view_promotion	29,188	29,117	-0.24%

Аналіз

Перевірка розподілення 50% на 50%:



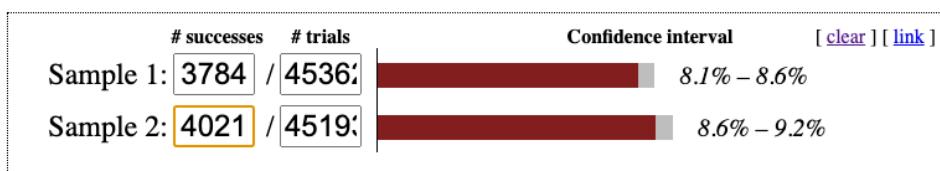
Зміна в основних метриках: у тоталі метрика **begin_checkout / session** статистично значуще виросла з 8.34% до 8.90%, що складає 6.66%:

	Test Group		
Event Name	1	2	
begin_checkout	3784	4021	
session	45362	45193	
	8,34%	8,90%	6,66%

Evan's Awesome A/B Tools ([home](#)):

[Sample Size Calculator](#) | [Chi-Squared Test](#) | [Sequential Sampling](#) | [2 Sample T-Test](#) | [Survival Times](#) | [Count Data](#)

Question: Does the rate of success differ across two groups?



Verdict:

Sample 2 is more successful

(p = 0.00289)

Зміна в додаткових метриках: метрика **session_with_orders / session** виросла з 9.95% до 10.01%, що складає 0.64%, результат не є статистично значущим. Метою було не просадити метрику, що в результаті і маємо:

	Test Group		
Event Name	1	2	
session_with_orders	4514	4526	
session	45362	45193	
	9,95%	10,01%	0,64%

Evan's Awesome A/B Tools ([home](#)):

[Sample Size Calculator](#) | [Chi-Squared Test](#) | [Sequential Sampling](#) | [2 Sample T-Test](#) | [Survival Times](#) | [Count Data](#)

Question: Does the rate of success differ across two groups?



Verdict:

No significant difference

(p = 0.75)

Зміни у розрізі каналів: статистично значущі позитивні результати дали канали Direct та Social Search.

Для каналу Undefined просіла додаткова метрика, проте результати не є статистично значущими:

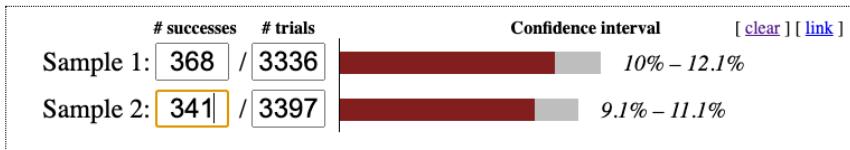
F	G	H	I
	Test Group		
Event Name	1	2	
begin_checkout	353	569	
session	3336	3397	
	10,58%	16,75%	58,30%

	Test Group		
Event Name	1	2	
session_with_orders	368	341	
session	3336	3397	
	11,03%	10,04%	-9,00%

Evan's Awesome A/B Tools ([home](#)):

[Sample Size Calculator](#) | [Chi-Squared Test](#) | [Sequential Sampling](#) | [2 Sample T-Test](#) | [Survival Times](#) | [Count Data](#)

Question: Does the rate of success differ across two groups?



Verdict:

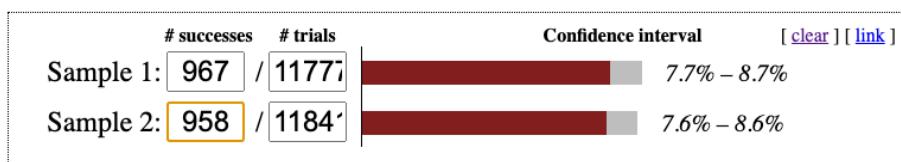
No significant difference

(p = 0.18)

Для каналу Paid Search просіла основна та додаткова метрики, проте результати не є статистично значущими:

Evan's Awesome A/B Tools ([home](#)):
[Sample Size Calculator](#) | [Chi-Squared Test](#) | [Sequential Sampling](#) | [2 Sample T-Test](#) | [Survival Times](#) | [Count Data](#)

Question: Does the rate of success differ across two groups?

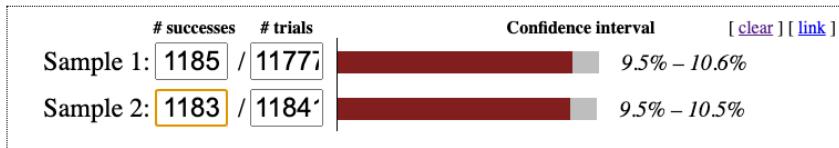


Verdict:

No significant difference

(p = 0.74)

Question: Does the rate of success differ across two groups?



Verdict:

No significant difference

($p = 0.86$)

	Test Group		
Event Name	1	2	
begin_checkout	967	958	
session	11777	11841	
	8,21%	8,09%	-1,47%

	Test Group		
Event Name	1	2	
session_with_order	1185	1183	
session	11777	11841	
	10,06%	9,99%	-0,71%

Канал Organic Search дав негативні статистично значущі результати:

	Test Group		
Event Name	1	2	
begin_checkout	1249	1144	
session	15675	15631	
	7,97%	7,32%	-8,15%

	Test Group		
Event Name	1	2	
session_with_orders	1502	1534	
session	15675	15631	
	9,58%	9,81%	2,42%

Question: Does the rate of success differ across two groups?



Verdict:

Sample 1 is more successful

($p = 0.0306$)

Зміни у розрізі девайсів: статистично значущі позитивні результати для пристроїв desktop.

Для мобільних пристроїв просіла основна метрика, проте результати не є статистично значущими:

	Test Group		
Event Name	1	2	
begin_checkout	1593	1561	
session	17896	17767	
	8,90%	8,79%	-1,30%

	Test Group		
Event Name	1	2	
session_with_order	1738	1771	
session	17896	17767	
	9,71%	9,97%	2,64%

Є негативний тренд по планшетах (*tablet*), даних дуже мало, але негативний тренд спостерігається по переважній кількості метрик:

Event Name	Value		Value (%)	
	Test Group		Test Group	
	1	2		2
add_payment_info	48	31	-35.42%	
add_shipping_info	66	49	-25.76%	
add_to_cart	37	31	-16.22%	
begin_checkout	83	56	-32.53%	
first_visit	699	695	-0.57%	
new_account	90	93	3.33%	
page_view	4,829	3,814	-21.02%	
scroll	1,751	1,440	-17.76%	
session	999	1,009	1.00%	
session_start	1,032	1,021	-1.07%	
session_with_orders	98	78	-20.41%	
user_engagement	4,010	3,438	-14.26%	
view_item	1,692	1,125	-33.51%	
view_promotion	660	651	-1.36%	

Зміни у розрізі континентів: статистично значущі позитивні результати спостерігались для континентів Asia, Europe та Oceania.

Для континенту America основна метрика виросла лише на 3.22% і просіла додаткова метрика. Результати не є статистично значущими:

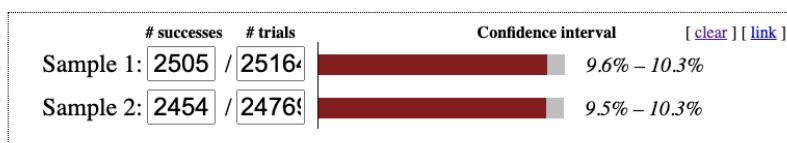
Event Name	Test Group		
	1	2	
begin_checkout	2130	2164	
session	25164	24769	
	8,46%	8,74%	3,22%

Event Name	Test Group		
	1	2	
session_with_orders	2505	2454	
session	25164	24769	
	9,95%	9,91%	-0,47%

Evan's Awesome A/B Tools ([home](#)):

[Sample Size Calculator](#) | [Chi-Squared Test](#) | [Sequential Sampling](#) | [2 Sample T-Test](#) | [Survival Times](#) | [Count Data](#)

Question: Does the rate of success differ across two groups?



Verdict:

No significant difference

(p = 0.86)

Для континенту Africa даних дуже мало. Тест дав негативні статистично значущі результати по основній метриці:

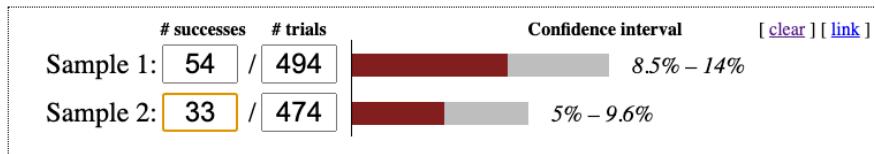
	Test Group		
Event Name	1	2	
begin_checkout	54	33	
session	494	474	
	10,93%	6,96%	-36,31%

	Test Group		
Event Name	1	2	
session_with_order	46	45	
session	494	474	
	9,31%	9,49%	1,95%

Evan's Awesome A/B Tools ([home](#)):

[Sample Size Calculator](#) | [Chi-Squared Test](#) | [Sequential Sampling](#) | [2 Sample T-Test](#) | [Survival Times](#) | [Count Data](#)

Question: Does the rate of success differ across two groups?



Verdict:

Sample 1 is more successful

(p = 0.0309)

Висновки:

Тотал: Основна метрика (Begin Checkout) статистично значуще зросла на 6,6%, що перевищує поставлену ціль у 5%. Додаткова метрика (Session with Orders) не зазнала статистично значущого погіршення, що відповідає вимогам тесту.

Канали: Позитивний ефект: Direct та Social Search (статистично значущі). Негативний ефект: Organic Search (статистично значуще). Просідання без значущості: Undefined та Paid Search.

Девайси: Позитивний ефект: Desktop (статистично значущий). Негативний тренд: Tablets (низька вибірка, але переважно негативні результати). Мобільні пристрої: Негативний ефект, але без значущості.

Континенти: Позитивний ефект: Азія, Європа, Океанія (статистично значущі). Негативний ефект: Африка (статистично значущий спад). Америка: Нестатистично значущий ріст основної метрики, але спад додаткової.

Рекомендації:

Оскільки позитивний ефект статистично значуще зафікований лише на десктопних пристроях, рекомендується впроваджувати зміни лише для цього сегмента. Для інших каналів та пристройів варто продовжити додаткове тестування або провести глибший аналіз.

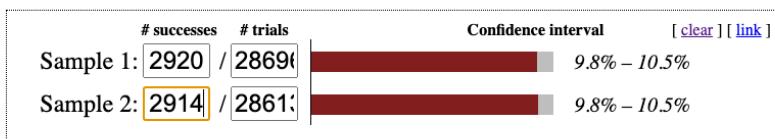
Можна спробувати виключити з аналізу Organic channel, пристройі tablet та континент Africa. В такому випадку ми отримаємо більший ніж в тоталі ріст основної метрики, однак додаткова метрика зазнає незначного статистично не значущого погрішення:

Event Name	Value		Value (%)	
	Test Group		Test Group	
	1	2	1	2
add_payment_info	1,298	1,680	29.43%	
add_shipping_info	1,934	2,247	16.18%	
add_to_cart	854	955	11.83%	
begin_checkout	2,432	2,820	15.95%	
first_visit	17,534	17,610	0.43%	
new_account	2,427	2,303	-5.11%	
page_view	118,010	125,071	5.98%	
scroll	45,999	48,129	4.63%	
session	28,696	28,613	-0.29%	
session_start	28,932	28,789	-0.49%	
session_with_orders	2,920	2,914	-0.21%	
user_engagement	106,366	113,716	6.91%	
view_item	38,333	41,470	8.18%	
view_promotion	18,278	18,215	-0.34%	

Evan's Awesome A/B Tools ([home](#)):

[Sample Size Calculator](#) | [Chi-Squared Test](#) | [Sequential Sampling](#) | [2 Sample T-Test](#) | [Survival Times](#) | [Count Data](#)

Question: Does the rate of success differ across two groups?



Verdict:

No significant difference

(p = 0.97)

Додаткові нотатки: Потрібно запустити окремий аналіз для мобільних пристройів та континенту Америка, щоб зрозуміти причини слабкої динаміки. Перевірити глибше канал Paid Search, оскільки обидві метрики показали негативну тенденцію, навіть без статистичної значущості. Загалом, тест можна вважати успішним, але потрібна додаткова оптимізація для окремих сегментів.