

ДЗ №3

А/В-тести

Завдання 1

Нам задано наступний набір даних:

user id	event_time	event name	group	country	source	
433952333075	6/13/2023	click	Rounded rectangle	Germany		6
434445569649	44991	click	Squircle	Germany		1
434445569649	44991	6842824074	purchase	Squircle	Germany	1
434461599357	45175	click	Rounded rectangle	Germany		1
434941603456	45205	click	Squircle	Turkey		1
434941603456	6/30/2023	click	Squircle	Turkey		1
434941603456	7/21/2023	click	Squircle	Turkey		1
435158812174	45022	click	Rounded rectangle	Burkina Faso		6
435268765331	45113	click	Squircle	El Salvador		1
435692140999	6/13/2023	click	Rounded rectangle	Spain		1
435795539828	45022	click	Rounded rectangle	Canada		1
435949621718	45113	click	Rounded rectangle	Spain		6
436221819732	6/13/2023	click	Squircle	United States		2
436221819732	6/13/2023	click	Squircle	United States		4

Як можна помітити, датасет має деякі проблеми зі зміщенням даних та кодуванням. За допомогою Excel та Python (код буде прикріплено при здачі) усунемо ці недоліки. Також було виявлено, що для рядків з «purchase» є дублікати з точністю до секунд в часі, тому їх було видалено. Було видалено записи, які потрапили в кожну групу. Інші записи були залишені з власних міркувань. Загалом отримано ось такий датасет:

Id	Date	Event	Group	Country	Source
433952333075	2023-06-13 0:00:00	click	Rounded rectangle	Germany	6
434445569649	2023-06-03 0:00:00	click	Squircle	Germany	1
434445569649	2023-06-03 0:00:00	purchase	Squircle	Germany	1
434461599357	2023-06-09 0:00:00	click	Rounded rectangle	Germany	1
434941603456	2023-06-10 0:00:00	click	Squircle	Turkey	1
434941603456	2023-06-30 0:00:00	click	Squircle	Turkey	1
434941603456	2023-07-21 0:00:00	click	Squircle	Turkey	1
435158812174	2023-06-04 0:00:00	click	Rounded rectangle	Burkina Faso	6
435268765331	2023-06-07 0:00:00	click	Squircle	El Salvador	1
435692140999	2023-06-13 0:00:00	click	Rounded rectangle	Spain	1
435795539828	2023-06-04 0:00:00	click	Rounded rectangle	Canada	1
435949621718	2023-06-07 0:00:00	click	Rounded rectangle	Spain	6
436221819732	2023-06-13 0:00:00	click	Squircle	United States	2
436221819732	2023-06-13 0:00:00	click	Squircle	United States	4
436221819732	2023-06-13 0:00:00	click	Squircle	United States	4
436340213058	2023-06-03 0:00:00	click	Squircle	Sweden	6
436340213058	2023-06-03 0:00:00	click	Squircle	Sweden	6
437153889617	2023-06-08 0:00:00	click	Squircle	Netherlands	1
437232605120	2023-06-08 0:00:00	click	Rounded rectangle	Norway	6
437355112928	2023-06-08 0:00:00	click	Squircle	France	1

Зрозуміло, що конверсію в даній задачі будемо визначати для купівлі асистенту, тобто $\frac{\text{count}(\text{purchase})}{\text{count}(\text{click})}$.

Подальший огляд даних проведемо за допомогою Python. Бачимо, що датасет розбитий майже порівну по контрольній та тестовій групі:

		Id	Date	Country	Source
Group	Event				
Rounded rectangle	click	123366	123366	123366	116352
	purchase	3565	3565	3565	3398
Squircle	click	124883	124883	124883	117273
	purchase	3725	3725	3725	3561

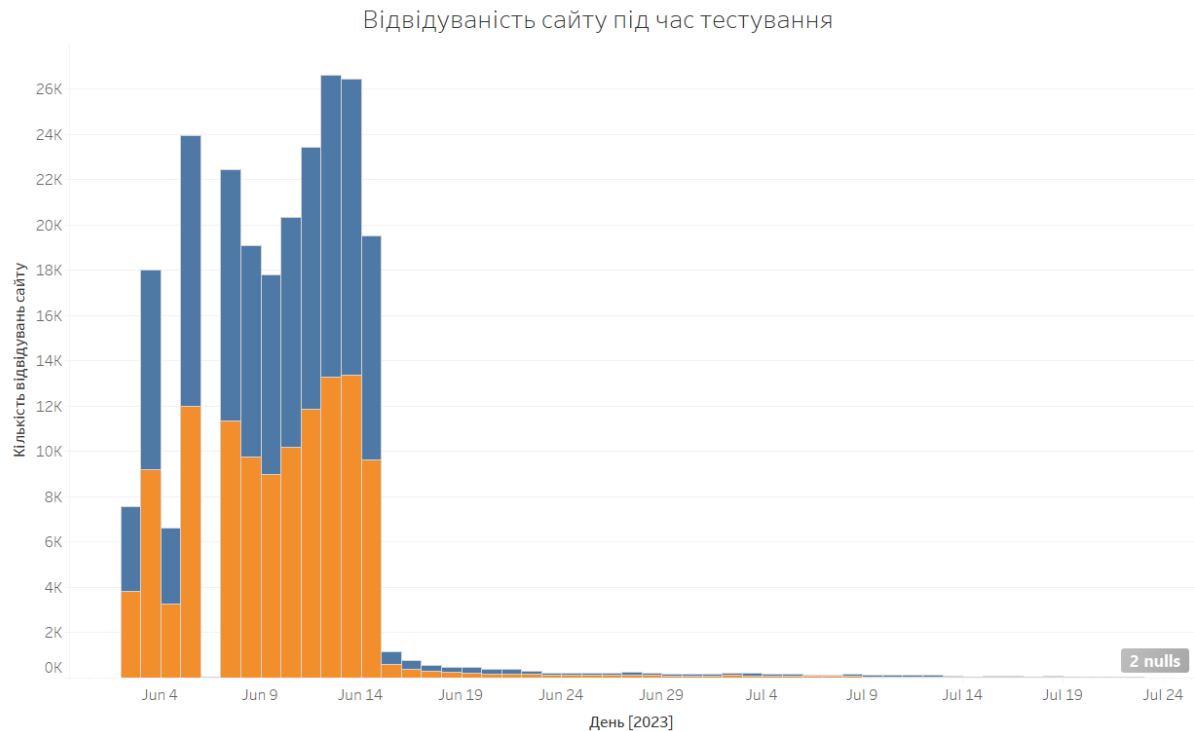
Також, бачимо, що загалом кількість юзерів також майже однакова. Розрізи по країнах і джерелах також показують, що в кожній групі маємо представників всіх джерел і однакової кількості країн. Визначати відсоткові відношення країн та джерел для А і Б групи вважаю недоречним:

```
df1.groupby(['Group', 'Event']).nunique()
```

✓ 0.1s

		Id	Date	Country	Source
Group	Event				
Rounded rectangle	click	105771	51	212	10
	purchase	3558	672	108	10
Squircle	click	106631	51	213	10
	purchase	3714	702	105	10

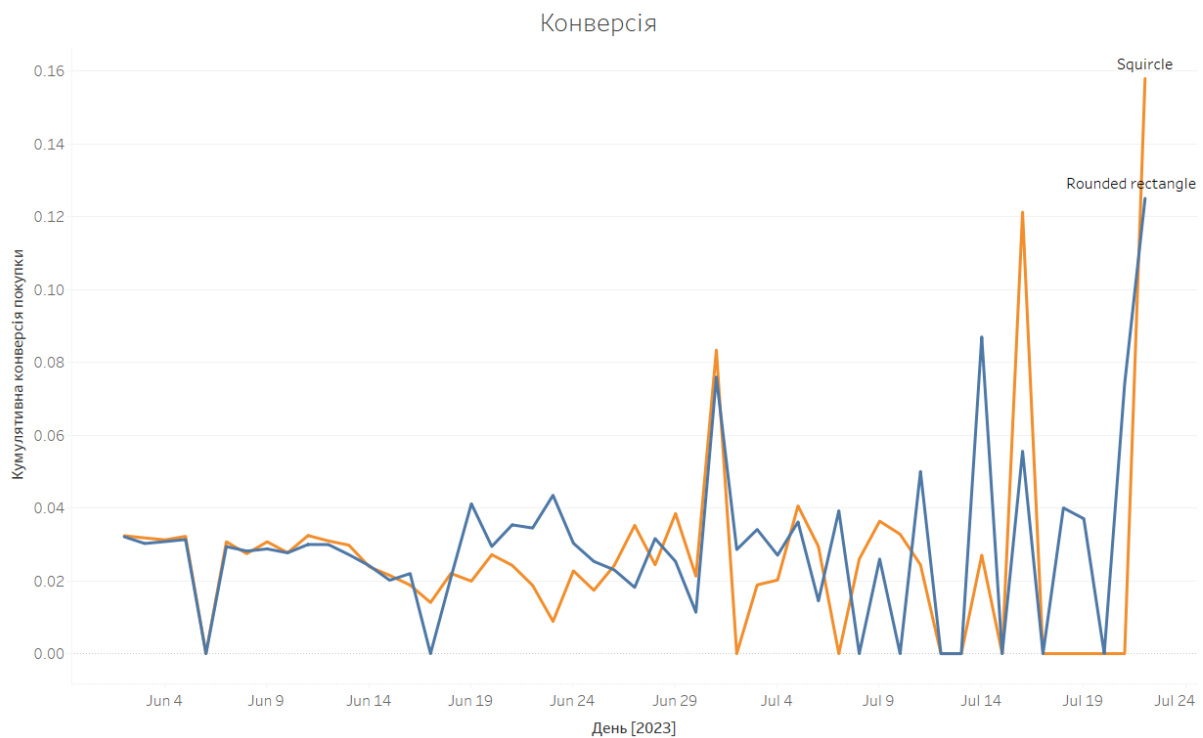
Якщо подивитися на відвідування сайту під час тесту, то бачимо, що нерівномірно по днях були зроблені записи до датасету:



Але в цілому з тестової та контрольної групи людей в день приходило однаково.

Можна було б зробити припущення, що не варто проводити тест після 14 червня, оскільки маємо невелику кількість відвідувань і мабуть вони не сильно вплинуть на результат. Проте слід глянути на конверсію по кожному з днів:

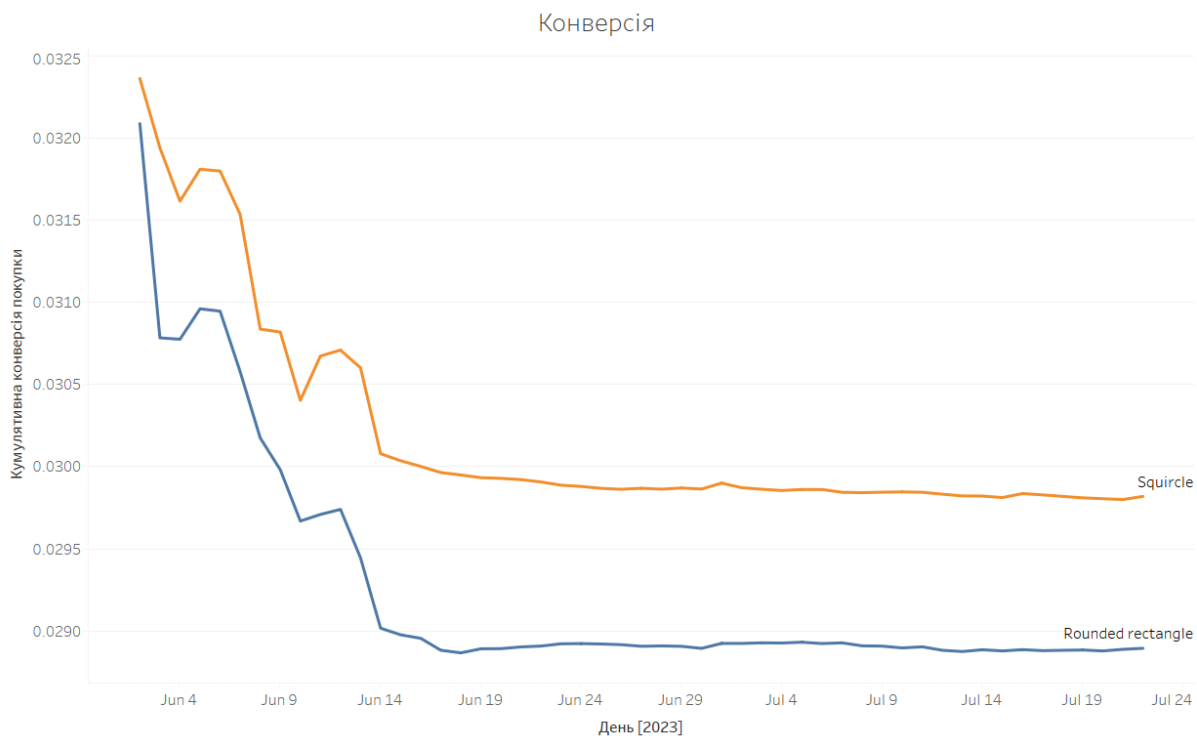
Event	Date	Group	Click Count	Purchase Count	Conversion
0	2023-06-02	Rounded rectangle	3613.0	116.0	0.032106
1	2023-06-02	Squirrel	3676.0	119.0	0.032372
2	2023-06-03	Rounded rectangle	8532.0	258.0	0.030239
3	2023-06-03	Squirrel	8846.0	281.0	0.031766
4	2023-06-04	Rounded rectangle	10787.0	332.0	0.030778
...
97	2023-07-20	Squirrel	22.0	0.0	0.000000
98	2023-07-21	Rounded rectangle	27.0	2.0	0.074074
99	2023-07-21	Squirrel	22.0	0.0	0.000000
100	2023-07-22	Rounded rectangle	8.0	1.0	0.125000
101	2023-07-22	Squirrel	19.0	3.0	0.157895



Єдине, що можна сказати з цього графіку, те що в останні дні конверсія значно виросла. Можливо, що ті покупки, які були зроблені в ці дні, якось вплинуть на результат.

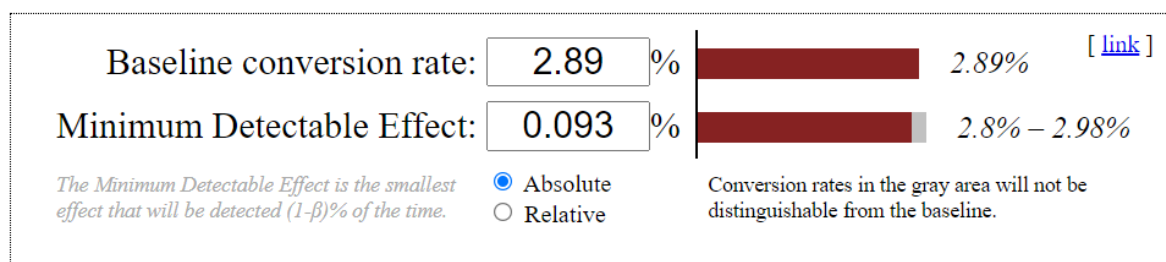
Подивимося як себе поводить кумулятивна конверсія з плином експерименту:

Event	Date	Group	Click Count	Purchase Count	Conversion
0	2023-06-02	Rounded rectangle	3615.0	116.0	0.032089
1	2023-06-02	Squirrel	3677.0	119.0	0.032363
2	2023-06-03	Rounded rectangle	12149.0	374.0	0.030784
3	2023-06-03	Squirrel	12523.0	400.0	0.031941
4	2023-06-04	Rounded rectangle	22940.0	706.0	0.030776
...
97	2023-07-20	Squirrel	124911.0	3723.0	0.029805
98	2023-07-21	Rounded rectangle	123406.0	3565.0	0.028888
99	2023-07-21	Squirrel	124933.0	3723.0	0.029800
100	2023-07-22	Rounded rectangle	123414.0	3566.0	0.028895
101	2023-07-22	Squirrel	124952.0	3726.0	0.029819



Здається, що тестова група показує конверсію на 0,1% відсотковий пункт більшу, але ось це плато, яке ми отримали для тестової та контрольної групи, якраз пов'язане з тим, що маємо мало записів після 14 червня.

Отже, визначимо розмір вибірки для кожної групи, яка була б необхідна для визначення значущості результату експерименту, якщо взяти за MDE зовсім трохи менше число ніж ми маємо різницю насправді. Бачимо, що нам не вистачає для цього даних:



Sample size:
511,750
per variation

Statistical power 1-β: 80% *Percent of the time the minimum effect size will be detected, assuming it exists*

Significance level α: 5% *Percent of the time a difference will be detected, assuming one does NOT exist*

В такому випадку ми не можемо відштовхуватися від нашого MDE. Тому просто сформулюємо наші нульову і альтернативну гіпотезу наступним чином (зробимо наш тест одностороннім):

H_0 :

«Конверсія сайту з кнопками «Squirrel» насправді є некращою за конверсію з сайту з кнопками «Rounded rectangle» з рівним статистичної значимості – 0.05 і потужністю критерію – 0.8»

H_1 :

«Конверсія сайту з кнопками «Squirrel» краща за конверсію з сайту з кнопками «Rounded rectangle» з рівним статистичної значимості – 0.05 і потужністю критерію – 0.8»

Маючи такі дані:

Event	Group	Click Count	Purchase Count	Conversion
0	Rounded rectangle	123366	3565	0.028898
1	Squirrel	124883	3725	0.029828

Введемо їх в онлайн калькулятор і отримаємо наступний результат:

Calculate your statistical significance

	Visitors	Conversions		Conversion rate
A	123414	3566	→	2.89%
B	124952	3726	→	2.98%

Hypothesis ⓘ

☒ One-sided ☐ Two-sided

Confidence ⓘ

☐ 90% ☒ 95% ☐ 99%

Calculate

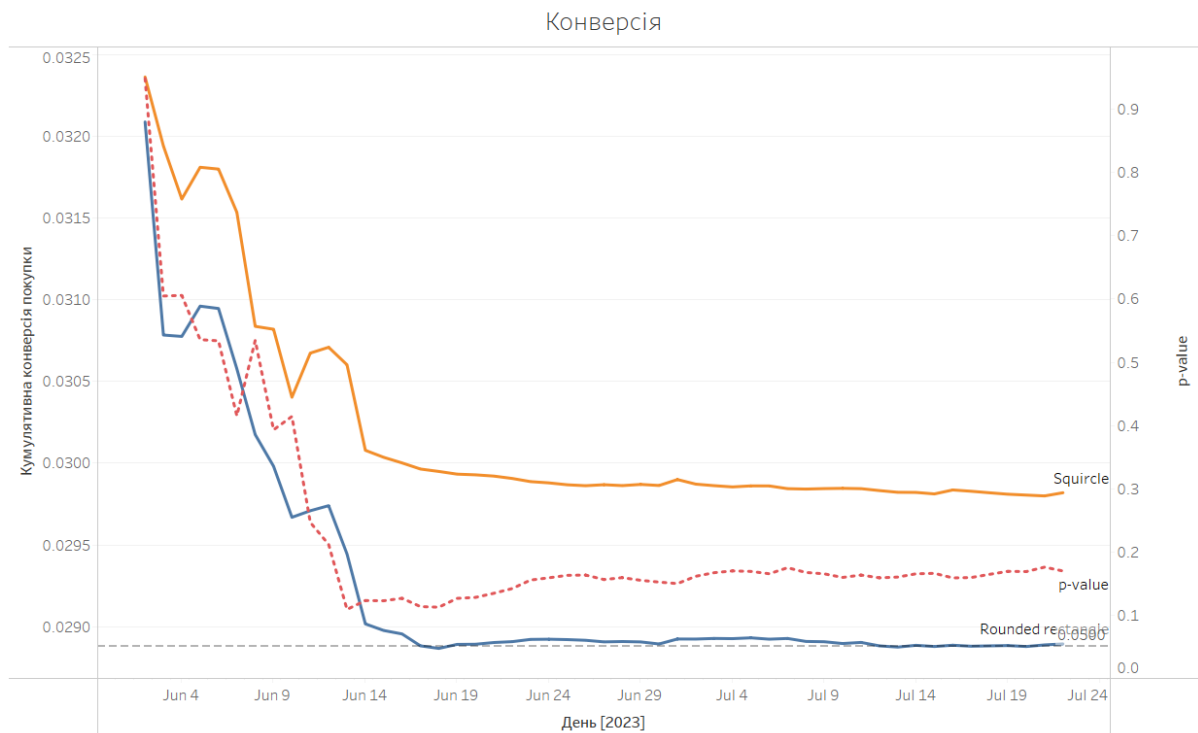
Result not significant!

Variant B's conversion rate (2.98%) was 3.2% higher than variant A's conversion rate (2.89%), but you cannot say, with 95% confidence, that variant B will perform better than variant A.

Power	p value
61.49%	0.0861

Бачимо, що даний результат не є статистично значимим при значеннях ймовірності помилки першого і другого роду 0.05 і 0.2 відповідно. Відповідно слід прийняти основну гіпотезу H_0 , що тестова група не є кращою за контрольну.

Глянемо загалом на значення p-value:



Тут добре видно, що значення p -value не досягає потрібного нам значення, більш того, навіть при ймовірності похибки першого роду – 0.1, ми не можемо бути впевненими, що результат є статистично значимим.

Загальний висновок:

В даному завданні важко однозначно надати відповідь на поставлене питання, оскільки не було надано допустимі ймовірності помилок першого і другого роду. З останнього графіку видно, що p -value доволі високий, аби бути впевненим, що кнопки типу «Squirrel» підвищують нашу конверсію. Проте основуючись на отриманих результатах, я б оновив всі кнопки до «Squirrel», адже навряд ми можемо сказати, що вони гірші за контрольний варіант «Rounded rectangle». Щодо тривалості експерименту, то видно, що немає було сенсу його проводити після 19 червня.

Завдання 2

Маємо такий датасет:

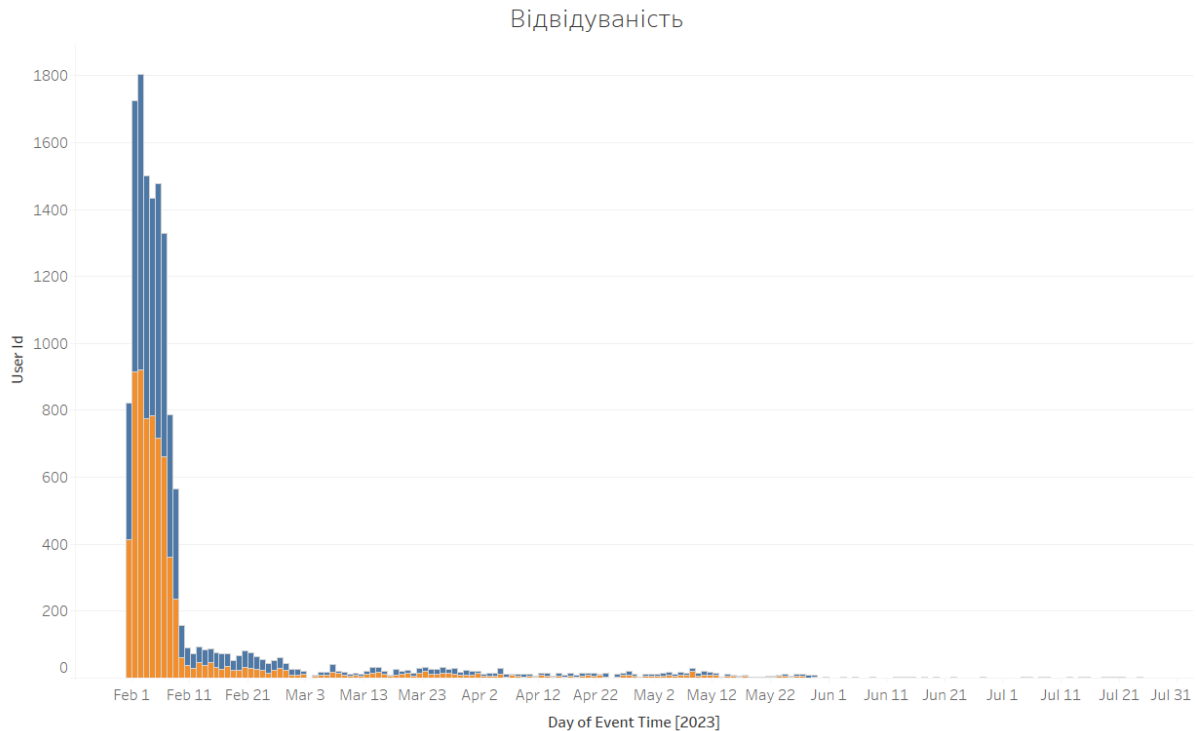
user id	event time	event name	group	country	source
5.62E+16	1/2/23	click	test(6.99)	Canada	
5.62E+16	1/2/23	click	test(6.99)	Chile	
5.62E+16	1/2/23	click	test(6.99)	Colombia	
5.61E+16	1/2/23	click	test(6.99)	Italy	
5.62E+15	1/2/23	click	test(6.99)	Mauritius	
5.61E+16	1/2/23	click	test(6.99)	Netherlands	
5.61E+16	1/2/23	click	test(6.99)	United States	
5.62E+15	2/2/23	click	test(6.99)	Austria	
5.63E+16	2/2/23	click	test(6.99)	France	
5.62E+16	2/2/23	click	test(6.99)	Germany	
5.62E+16	2/2/23	click	test(6.99)	Italy	
5.63E+16	2/2/23	click	test(6.99)	Romania	
5.63E+16	2/2/23	click	test(6.99)	Singapore	
5.63E+15	2/2/23	click	test(6.99)	Singapore	
5.63E+16	2/2/23	click	test(6.99)	Singapore	
5.63E+16	2/2/23	click	test(6.99)	Singapore	
5.63E+16	2/2/23	click	test(6.99)	Singapore	

Оскільки в даному завданні нам треба визначити чи зросла кількість покупок асистенту, то будемо рахувати конверсію по людях. Тобто який відсоток людей натискає кнопку купити. Відповідно видалимо з датасету дублікати дій для однакового ID.

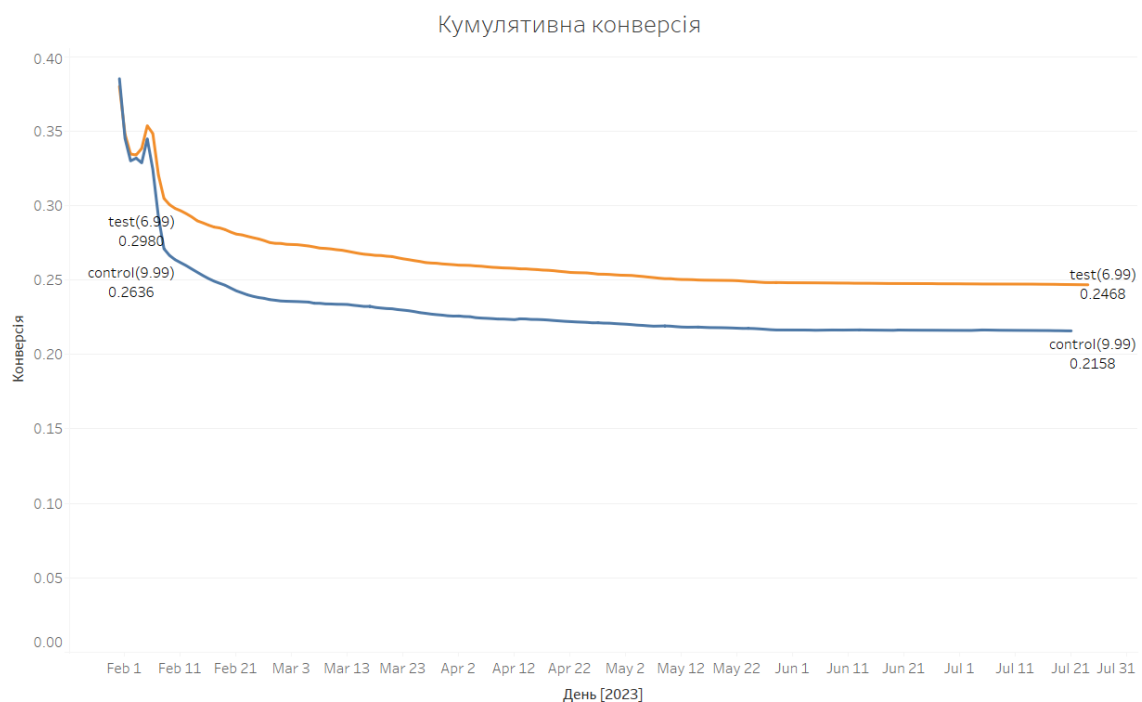
Перевіримо як розподілені дані по групах А і Б:

<pre>df.groupby('group').count()</pre>					
✓ 0.0s					
	user id	event time	event name	country	source
group					
control(9.99)	7217	7217	7217	7217	3425
test(6.99)	7129	7129	7129	7129	3494
<pre>df.groupby('group').nunique()</pre>					
✓ 0.0s					
	user id	event time	event name	country	source
group					
control(9.99)	6619	149	2	153	11
test(6.99)	6446	144	2	150	13

Як і в попередньому завданні будемо порівнювати А і Б групи по значенню конверсії покупки бота. Оскільки в файлі з даними немає інформації про порядковий номер кожної дії, то визначимо як і коли бралися дані для експерименту по днях:

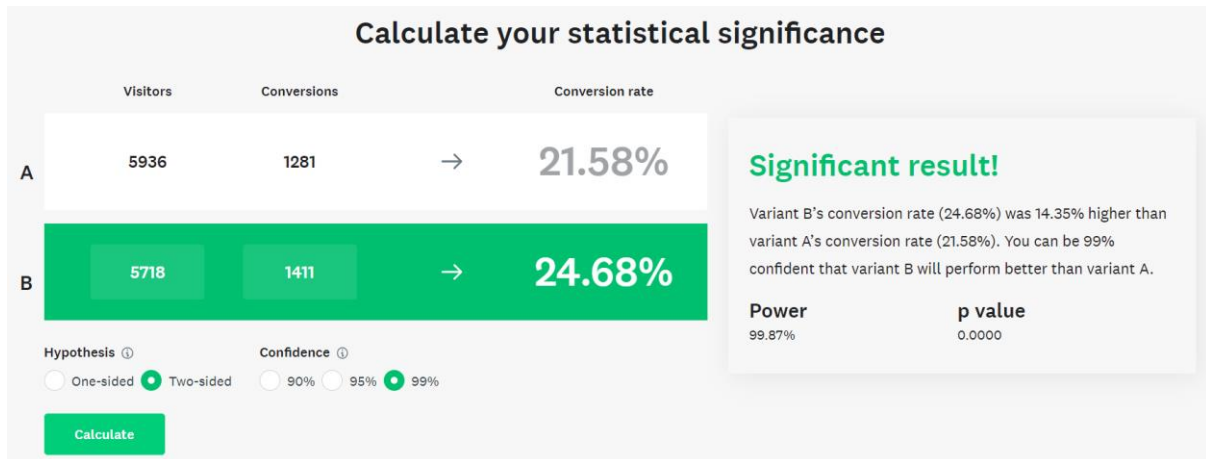


Загалом, кількість записів в день по кожній групі майже однакова. Видно, що після лютого дані в датасет майже не записувалися, менше 30 на день. Подивимось на конверсію покупок з плином часу експерименту:



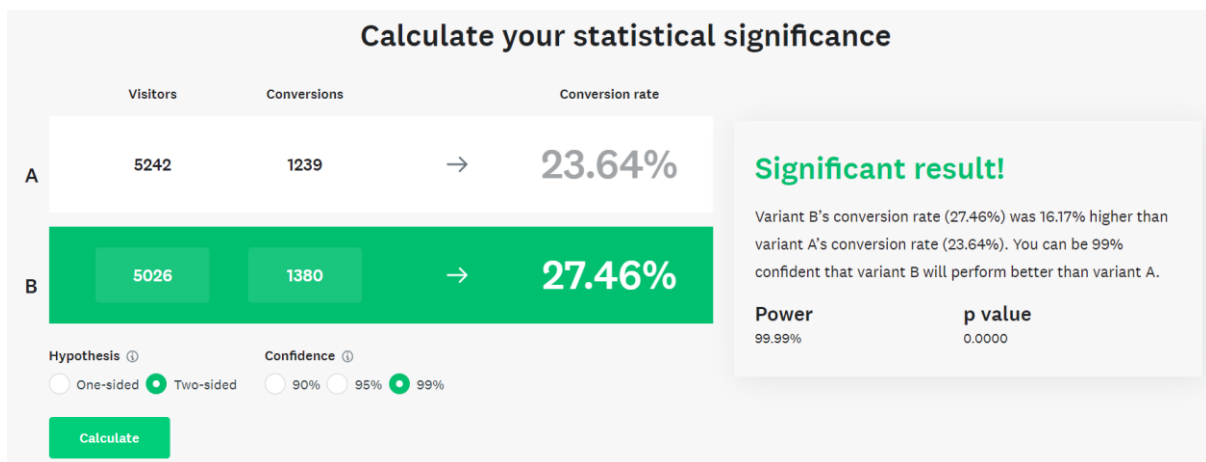
За рахунок того, що майже всі записи були зроблені (кількість записів на графіку) в перший тиждень експерименту, конверсія після лютого сильно не змінювалася. Видно, що конверсія між тестовою і контрольною групою дійсно помітна. Перевіримо результат на стат. значимість.

Перевіримо спочатку для всього експерименту:



Дійсно, така помітна різниця на доволі великій вибірці виявилася стат. значимою. Ми можемо з 99% впевненістю сказати, що результат є не випадковим. Авжеж ми не можемо отримати такі ідеальні значення для p-value, скоріше за все отримані значення α і β були близькі до нуля і даний онлайн калькулятор округлив їх.

А тепер перевіримо результат на статистичну значимість по тих значеннях, які ми отримали за перший місяць експерименту (2023-01-31 – 2023-02-28):



Результат також виявився статистично значимим.

Даний результат наочно продемонстрований на наступному графіку:



З графіку видно, що нам досить навіть першого тижня, аби стверджувати, що результат є статистично значимий. Якщо б ми проводили даний тест в реальному часі, то однозначно не варто було б проводити його довше ніж 1 тиждень.

Можна стверджувати майже напевно, що зменшення ціни за покупку збільшує кількість покупок. При зменшенні ціни з 10\$ до 7\$ даний факт був доволі зрозумілий ще до самого початку експерименту.

Проте дослідження на цьому не можна завершити, оскільки варто зрозуміти чи відіб'ємо ми втрати в ціні на нових клієнтах. Для цього були дані значення LTV користувача за 6 місяців:

Control group – 29.27 USD

Test group – 23.08 USD

В даному завданні краще використовувати значення конверсій по всіх наявних даних, аби точніше визначити у скільки разів з'явиться більше клієнтів на продукті.

Тепер асистента будуть купувати у стільки разів більше:

$$\frac{0.246765}{0.215802} \approx 1.143$$

Але в середньому дохід з одного клієнта за 6 місяців впав в:

$$\frac{23.08}{29.27} \approx 0.789$$

А тому дохід загалом впав в стільки разів:

$$1.143 \times 0.789 \approx 0.9$$

Або ж на 10%.

Загальний висновок:

В даному завданні ми можемо стверджувати, що тестова ціна асистента залучає більше клієнтів ніж контрольна з confidence level = 0.99. Проте, відштовхуючись від того, що в даному завданні основною метрикою була середній дохід з клієнта за пів року, то насправді ця метрика вже гірша за контрольну на 10%. Тому, на мою думку, варто залишити ціну на рівні 9,99\$.