

# Аналіз результатів А/В тесту мобільного застосунку

Виконав: **ВЛАДИСЛАВ Гончарук**

Мета: Визначити, чи впливає вказівка про «50% знижку» на конверсію у покупку підписки.

Інструменти: Python (pandas, scipy, seaborn, matplotlib)

Проект виконано в рамках курсу Data Analytics  
від **GoIT School**

pandas

seaborn

# Опис даних

## Дані:

Один рядок = один користувач

Поля: user\_id, test\_group (a/b), conversion (0/1), event\_time

Мета: Порівняти конверсію (install → purchase) між групами

Група	Кількість Користувачів	Конверсія
A	10 013	6.10%
B	9 985	8.90%

## Перевірка якості даних:

- ✓ Відсутні пропущені значення
- Унікальні користувачі: **19 998**
- Дата початку тесту: **03.07.2023**
- Дата завершення: **25.07.2023**
- Тривалість: **23 дні**

# 2. Огляд і базова статистика

```
print(data.groupby('test_group')['user_id'].nunique())  
print(data.groupby('test_group')['conversion'].mean())
```

```
test_group  
a      10013  
b       9985  
Name: user_id, dtype: int64  
test_group  
a      0.061021  
b      0.089034  
Name: conversion, dtype: float64
```

pandas

seaborn

# Динаміка користувачів

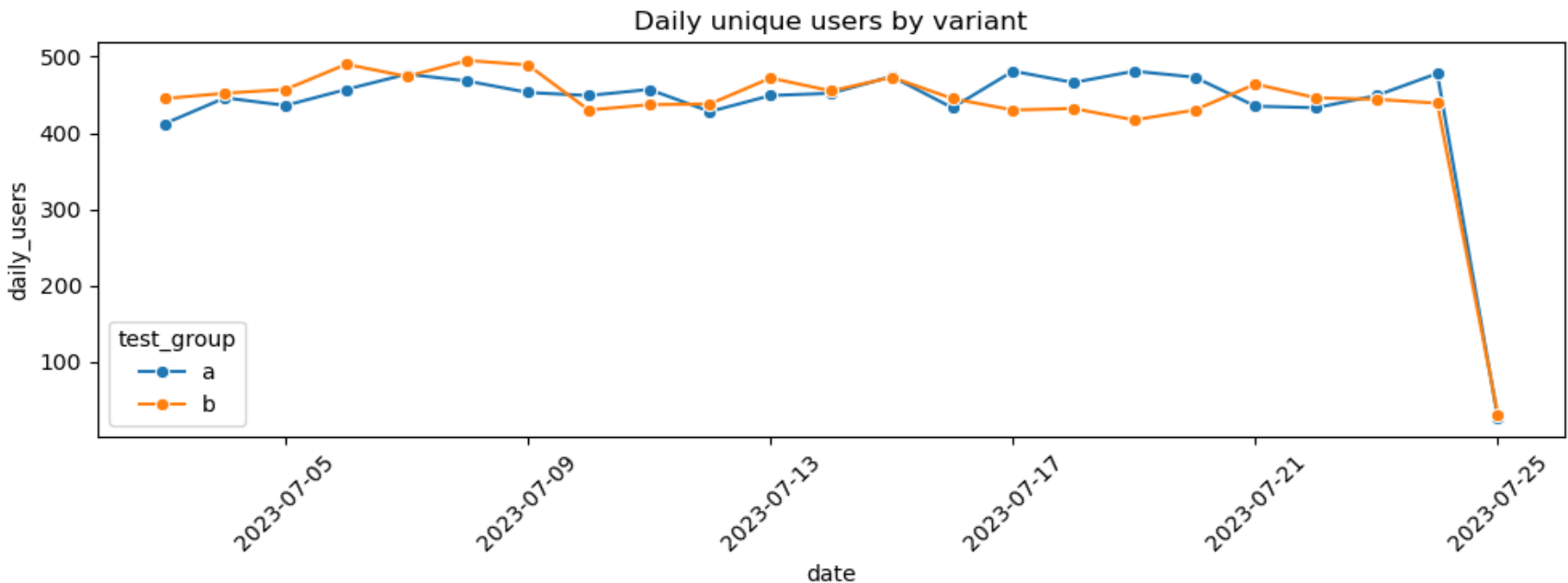
## Висновок:

Трафік стабільний, аномалій не спостерігається.

На графіку: ось X — дати,  
ось Y — кількість користувачів

```
# 3. Статистика за день
daily = data.groupby([data['date'], 'test_group'])['user_id'].nunique().reset_index()
daily.columns = ['date', 'test_group', 'daily_users']

# Побудова графіку
plt.figure(figsize=(10,4))
sns.lineplot(data=daily, x='date', y='daily_users', hue='test_group', marker='o')
plt.title('Daily unique users by variant')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



Лінійний графік показує кількість унікальних користувачів по днях для груп А та В.

# Статистичний аналіз

**Метод:** Оскільки метрика — бінарна (conversion), застосовано **z-test для двох пропорцій**,  $\chi^2$ -test і пермутаційний тест.

$\alpha$  - 5%

## Результати:

Конверсія у групі **B**

**статистично значуще вища,**

ніж у групі **A**.

Тест	Статистика	p-value	Висновок
Z-test	-7.52	0.0000000000000549	Відхиляємо $H_0$
$\chi^2$ -test	56.14	0.0000000000000674	Відхиляємо $H_0$
Permutation	-0.028	0.0002	Відхиляємо $H_0$

```
# proportions z-test (для перевірки різниці пропорцій)
z_stat, p_z = proportions_ztest(summary['conversions'], s

# chi2 test (для незалежності груп і конверсії)
chi2, p_chi2, _, _ = chi2_contingency(pd.crosstab(data['t

# permutation test (робастний тест для різниці середніх)
res = permutation_test(
    (data.loc[data['test_group']=='a', 'conversion'],
     data.loc[data['test_group']=='b', 'conversion']),
    lambda a,b: a.mean()-b.mean(),
    n_resamples=10000, alternative='two-sided'
)
```

=== Summary ===

	conversions	n	cr
test_group			
a	611	10013	0.0610
b	889	9985	0.0890

=== Statistical Tests ===

Z-test: z= -7.520, p=0.00000000000005491218 → Reject  $H_0$  - option B wins.

Chi2-test:  $\chi^2$ = 56.142, p=0.00000000000006740441

Permutation test: stat= -0.028, p=0.000200

Пояснення: p-value < 0.05 → відхиляємо  $H_0$

# Порівняння Conversion Rate

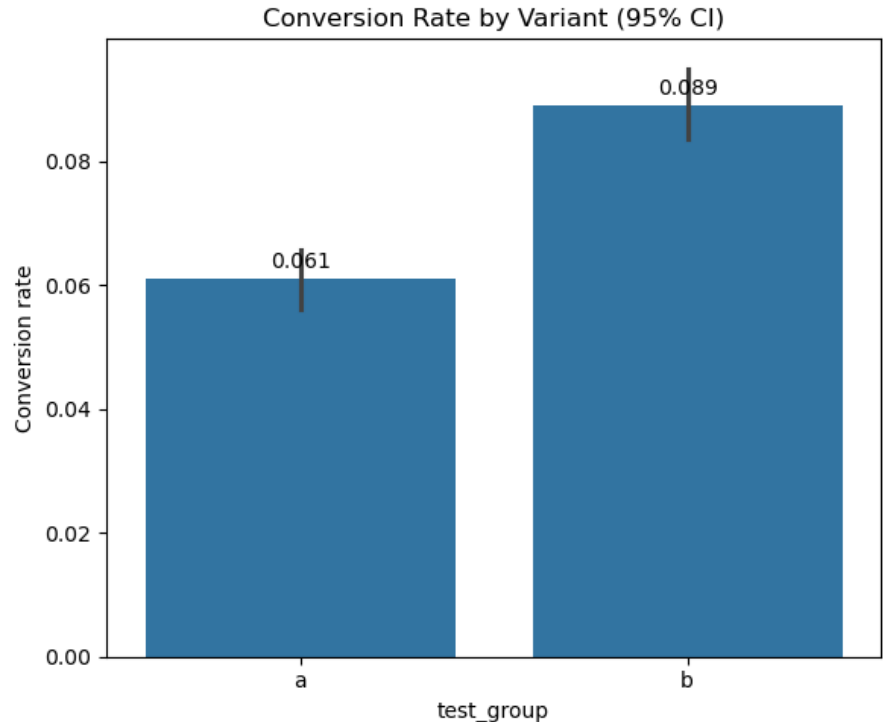
## Результати:

**Абсолютна різниця  $\Delta_{abs}$  (B-A) = +2.8 відсот.п.**

**Відносне зростання (Uplift) = +45.9%**

```
plt.figure(figsize=(6,5))
ax = sns.barplot(data=data, x='test_group', y='conversion', errorbar=sns.barplot_errdata)
plt.ylabel('Conversion rate')
plt.title('Conversion Rate by Variant (95% CI)')
plt.tight_layout()

# Підписи CR
for c in ax.containers:
    ax.bar_label(c, fmt='%.3f', label_type='edge', padding=3)
plt.show()
```



**Стовпчиковий графік із 95% довірчими інтервалами, що показує різницю CR між A та B.**

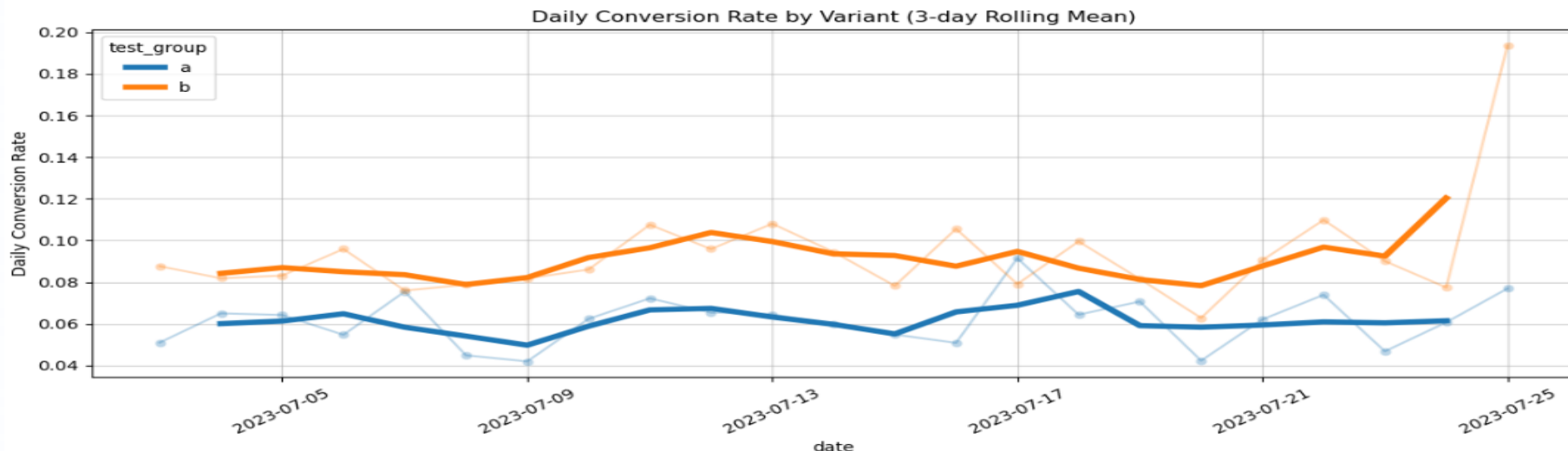
```
# Числові значення
print(summary[['conversions', 'n', 'cr']].round(4))
# обчислення абсолютної (Δ) та відносної (%) різниці
delta = summary.loc['b', 'cr'] - summary.loc['a', 'cr']
uplift = delta / summary.loc['a', 'cr']
print(f'Δabs(B-A)={delta:.4f}, uplift={uplift:.2%}')
```

	conversions	n	cr
test_group			
a	611	10013	0.061
b	889	9985	0.089

Δabs(B-A)=0.0280, uplift=45.91%

# Динаміка конверсії в часі

**Результати:** Ефект не є випадковим, тренд послідовний.



Графік щоденного CR для кожної групи демонструє **стабільну не випадкову перевагу варіанту В протягом усього періоду тесту**. Також коказано лінії трендів, що згладжують розкид даних

```
# Daily Conversion Rate
daily_conv = (
    data.groupby(['date', 'test_group'])
        .agg(users=('user_id', 'nunique'),
             conversions=('conversion', 'sum'))
        .reset_index()
)
daily_conv['cr'] = daily_conv['conversions'] / daily_conv['users']

plt.figure(figsize=(12,5))
sns.lineplot(data=daily_conv, x='date', y='cr', hue='test_group', marker='o')
```



# Висновок, рекомендація

**Результати:** Ефект не є випадковим, тренд послідовний.

**Резюме:**

Показник	A	B
Користувачів	10 013	9 985
Conversion Rate	6.10%	8.90%
$\Delta$ abs	+2.8 п.п.	
Uplift	+45.9%	

N\_A: 10013

CR\_A: 0.061

N\_B: 9985

CR\_B: 0.089

$\Delta$ abs(B-A): 0.0280

Uplift: 45.91%

p\_z: 0.000000000000005491218

p\_chi2: 0.000000000000006740441

p\_perm: 0.000200

**Висновок:** Різниця статистично значуща ( $p < 0.05$ ).

- ✓ **Рекомендація:** Впровадити новий дизайн із позначкою «50% знижка».
- ✓ Додатково — перевірити вплив на **LTV** та **відтік користувачів**.



# Дякую за увагу!

## ГОТОВИЙ до співпраці

*Контакти*



[Email](#)



[LinkedIn](#)



[GitHub](#)



pandas



seaborn