

Завдання 1

Завдання №1

Ви провели експеримент в якому зменшили кількість точок входу в чат з іншими користувачами для чоловічої аудиторії

1. Користувачі тестової групи мають доступ до чатів в розділі "Чати", в якому відображається список контактів:
 - які відповіли юзерам взаємною реакцією
 - яких підкинули логіки підбору продукту
2. І не мають доступ до переходу в чати з інших місць продукту (профіль контакту, розділи пошуку)
3. Обмеження стосуються неплатників

Між групами розподілились як "нові" користувачі (які зареєструвались після дати старту експерименту), так і "старі" (які зареєструвались до дати старту експерименту і вже мали певний досвід користування продуктом раніше)

Мотивація експерименту - переосмислення концепції активності на продукті, створюємо умови для кращого керування активностями/підбором в подальших експериментах.

Порахуємо спочатку метрики для груп і створимо таблицю результатів:

a. Конверсії в оплату (CR_payment)

```
#####  
###  
/* Конверсія в оплату*/  
  
SELECT t1.group_id,  
       CONCAT(ROUND(t2.conversion_count / t1.count * 100, 2),  
             '%') AS conversion_rate  
FROM   (SELECT tr.group_id,  
              COUNT(tr.user_id) AS count  
        FROM   user_split_groups tr  
        GROUP BY tr.group_id) t1  
LEFT JOIN (SELECT COUNT(DISTINCT usg.user_id) AS  
conversion_count,  
           group_id  
        FROM   payments  
        LEFT JOIN knu_quarks_workshops.user_split_groups  
usg  
           ON payments.user_id = usg.user_id  
        GROUP BY usg.group_id) t2  
ON t1.group_id = t2.group_id  
GROUP BY t1.group_id;
```

Результат запиту:

	group_id	conversion_rate
1	1	3.82%
2	0	3.71%

b. ARPU / ARPPU

1) ARPU

```
#####
/* ARPU */

SELECT t1.group_id,
       amount / count_user AS ARPU
FROM (SELECT COUNT(DISTINCT user_id) AS count_user,
            group_id
FROM user_split_groups
GROUP BY group_id) t1
JOIN (SELECT group_id,
            SUM(amount_usd) AS amount
FROM user_split_groups usg
JOIN payments p
ON usg.user_id = p.user_id
GROUP BY group_id) t2 ON t1.group_id = t2.group_id;
```

Результат запису:

	group_id	ARPU
1	0	0.43102482033298073
2	1	0.44353964917736666

2) ARPPU

```
#####
/* ARPPU */

SELECT t1.group_id,
       amount / count_payuser AS ARPPU
FROM (SELECT COUNT(DISTINCT p.user_id) AS count_payuser,
            group_id
FROM user_split_groups
RIGHT JOIN payments p
ON user_split_groups.user_id = p.user_id
GROUP BY group_id) t1
JOIN (SELECT group_id,
            SUM(amount_usd) AS amount
```

```
FROM user_split_groups usg
JOIN payments p
ON usg.user_id = p.user_id
GROUP BY group_id) t2 ON t1.group_id = t2.group_id;
```

Результат запиту:

	group_id		ARPPU
1		0	11.630170858049148
2		1	11.611549451074547

c. retention N day

N візьмемо 3, 7, 14, 21, 28

SQL query буде аналогічне для всіх N тому наведемо тільки для 3 днів, а для всіх інших буде аналогічним тільки буде зміна для на відповідне значення, тому будемо наводити відразу результати

```
WITH registration_day_users AS (
    SELECT
        usg.group_id,
        COUNT(DISTINCT usg.user_id) AS users_on_0_day
    FROM
        knu_quarks_workshops.user_split_groups usg
    JOIN
        registrations r ON usg.user_id = r.user_id
    JOIN
        activities_aggregated aa ON usg.user_id = aa.user_id
    GROUP BY
        usg.group_id
)

SELECT
    t1.group_id,
    COUNT(DISTINCT aa.user_id) AS
users_retention_count_on_day_3,
    COALESCE(rdu.users_on_0_day, 0) AS users_on_0_day,
    COUNT(DISTINCT aa.user_id) / COALESCE(rdu.users_on_0_day, 0)
AS retention_rate
FROM
    activities_aggregated aa
LEFT JOIN (
    SELECT
```

```

        usg.user_id,
        DATE(r.reg_dt) AS reg_dt,
        DATE_ADD(DATE(r.reg_dt), INTERVAL 3 DAY) AS
registration_plus_n_days,
        usg.group_id
    FROM
        registrations r
    RIGHT JOIN
        knu_quarks_workshops.user_split_groups usg
    ON
        r.user_id = usg.user_id
) t1 ON aa.user_id = t1.user_id
LEFT JOIN
    registration_day_users rdu ON t1.group_id = rdu.group_id
WHERE
    DATE(aa.dt) = t1.registration_plus_n_days
GROUP BY
    t1.group_id;

```

Результат запросу:

N = 3

	group_id	users_retention_count_on_day_3	users_on_0_day	retention_rate
1	0	4174	76096	0.0549
2	1	4207	76206	0.0552

N = 7

	group_id	users_retention_count_on_day_7	users_on_0_day	retention_rate
1	0	2564	76096	0.0337
2	1	2545	76206	0.0334

N = 14

	group_id	users_retention_count_on_day_14	users_on_0_day	retention_rate
1	0	1562	76096	0.0205
2	1	1631	76206	0.0214

N = 21

	group_id	users_retention_count_on_day_21	users_on_0_day	retention_rate
1	0	1302	76096	0.0171
2	1	1277	76206	0.0168

N = 28

	group_id	users_retention_count_on_day_28	users_on_0_day	retention_rate
1	0	1077	76096	0.0142
2	1	1038	76206	0.0136

N day	Group id	count_retention users on N day	count users on 0 day	retention rate
-------	----------	-----------------------------------	-------------------------	----------------

3	0	4174	76096	0.0549
	1	4207	76206	0.0552
7	0	2564	76096	0.0337
	1	2545	76206	0.0334
14	0	1562	76096	0.0205
	1	1631	76206	0.0214
21	0	1302	76096	0.0171
	1	1277	76206	0.0168
28	0	1077	76096	0.0142
	1	1038	76206	0.0136

d.

Порахуємо

1) Конверсію в активність (CR_act)

```
#####
/* Конверсія в активність */

SELECT t1.group_id,
       activity_cnt / count_user AS ARPU
FROM (SELECT COUNT(DISTINCT user_id) AS count_user,
           group_id
FROM user_split_groups
GROUP BY group_id) t1
JOIN (SELECT group_id,
            SUM(activity_cnt) AS activity_cnt
FROM activities_aggregated
LEFT JOIN user_split_groups
ON activities_aggregated.user_id = user_split_groups.user_id
JOIN knu_quarks_workshops.activity_type_dict atd on
activities_aggregated.activity_type = atd.activity_type
GROUP BY group_id) t2 ON t1.group_id = t2.group_id;
```

Результат запиту:

	group_id	÷	ARPU	÷
1		0	331.4552	
2		1	311.2510	

2) Середня кількість виконаних активностей (average_act)

```
#####
/* Середня кількість виконаних активностей */

SELECT group_id,
       AVG(activity_cnt) AS avg_activity_cnt
FROM activities_aggregated
LEFT JOIN user_split_groups
ON activities_aggregated.user_id = user_split_groups.user_id
JOIN knu_quarks_workshops.activity_type_dict atd on
activities_aggregated.activity_type = atd.activity_type
GROUP BY group_id
```

Результат запиту:

	group_id		avg_activity_cnt	
1	1		47.5796	
2	0		49.8432	

3) Середній чек юзера з кожної групи

```
#####
/* Середній чек юзера */

SELECT group_id,
       AVG(amount_usd)
FROM user_split_groups usg
INNER JOIN knu_quarks_workshops.payments p
ON usg.user_id = p.user_id
GROUP BY usg.group_id;
```

Результат запиту:

	group_id		`AVG(amount_usd)`	
1	0		8.116118437165667	
2	1		8.168248933713537	

4) Порахуємо кількість конкретних конверсій (тобто тільки like, dislike, message і visit)

a. CR_like

```
#####
/* CR_like */
```

```

SELECT t1.group_id,
       activity_cnt / count_user AS ARPU
FROM (SELECT COUNT(DISTINCT user_id) AS count_user,
           group_id
FROM user_split_groups
GROUP BY group_id) t1
JOIN (SELECT group_id,
             SUM(activity_cnt) AS activity_cnt
FROM activities_aggregated
LEFT JOIN user_split_groups
ON activities_aggregated.user_id = user_split_groups.user_id
JOIN knu_quarks_workshops.activity_type_dict atd on
activities_aggregated.activity_type = atd.activity_type
WHERE activity_name = 'like'
GROUP BY group_id) t2 ON t1.group_id = t2.group_id;

```

Результат запису:

	group_id	ARPU
1	0	51.1870
2	1	49.3466

b. CR_dislike

```

#####
/* CR_dislike */

SELECT t1.group_id,
       activity_cnt / count_user AS ARPU
FROM (SELECT COUNT(DISTINCT user_id) AS count_user,
           group_id
FROM user_split_groups
GROUP BY group_id) t1
JOIN (SELECT group_id,
             SUM(activity_cnt) AS activity_cnt
FROM activities_aggregated
LEFT JOIN user_split_groups
ON activities_aggregated.user_id = user_split_groups.user_id
JOIN knu_quarks_workshops.activity_type_dict atd on
activities_aggregated.activity_type = atd.activity_type
WHERE activity_name = 'dislike'
GROUP BY group_id) t2 ON t1.group_id = t2.group_id;

```

Результат запису:

	group_id	÷	ARPU	÷
1		0		233.6621
2		1		218.8009

c. CR_message

```
#####
/* CR_message */

SELECT t1.group_id,
       activity_cnt / count_user AS ARPU
FROM (SELECT COUNT(DISTINCT user_id) AS count_user,
            group_id
FROM user_split_groups
GROUP BY group_id) t1
JOIN (SELECT group_id,
            SUM(activity_cnt) AS activity_cnt
FROM activities_aggregated
LEFT JOIN user_split_groups
ON activities_aggregated.user_id = user_split_groups.user_id
JOIN knu_quarks_workshops.activity_type_dict atd on
activities_aggregated.activity_type = atd.activity_type
WHERE activity_name = 'message'
GROUP BY group_id) t2 ON t1.group_id = t2.group_id;
```

Результат запиту:

	group_id	÷	ARPU	÷
1		0		4.1735
2		1		3.8856

d. CR_visit

```
#####
/* CR_visit */

SELECT t1.group_id,
       activity_cnt / count_user AS ARPU
FROM (SELECT COUNT(DISTINCT user_id) AS count_user,
            group_id
FROM user_split_groups
GROUP BY group_id) t1
JOIN (SELECT group_id,
            SUM(activity_cnt) AS activity_cnt
FROM activities_aggregated
LEFT JOIN user_split_groups
ON activities_aggregated.user_id = user_split_groups.user_id
JOIN knu_quarks_workshops.activity_type_dict atd on
activities_aggregated.activity_type = atd.activity_type
```



```
WHERE activity_name = 'visit'
GROUP BY group_id) t2 ON t1.group_id = t2.group_id;
```

Результат запиту:

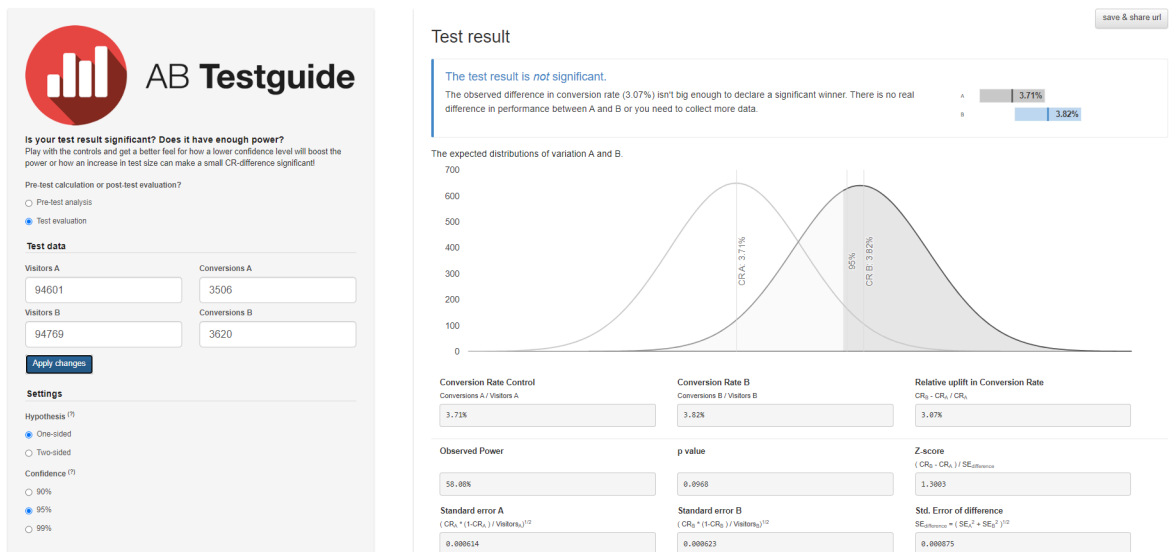
	group_id	÷	ARPU	÷
1	0		42.4326	
2	1		39.2179	

Побудуємо тепер таблицю зі всіма метриками разом

GROUP_ID	0	1
Total users in Group	94601	94769
CR_payment	0.0371	0.0382
CR_payment(nums)	3506	3620
ARPU	0.4310	0.4436
ARPPU	11.6302	11.6115
CR_act	331.4552	311.2510
Average_act	76.4779	73.2272
Середній чек юзера	8.1161	8.1683
CR_like	51.1870	49.3466
CR_dislike	233.6621	218.8009
CR_message	4.1735	3.8856
CR_visit	42.4326	39.2179

Перевіримо через калькулятори чи можемо довіряти статистиці

Результат з першого калькулятора

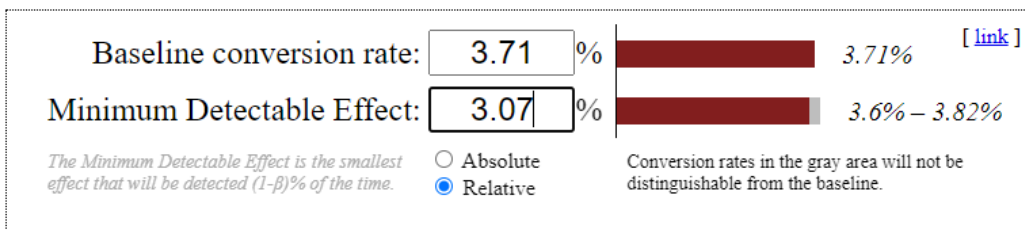


Перевіримо чи достатній розмір вибірки за допомогою іншого калькулятора

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

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

Question: How many subjects are needed for an A/B test?



Sample size:

434,192

per variation

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

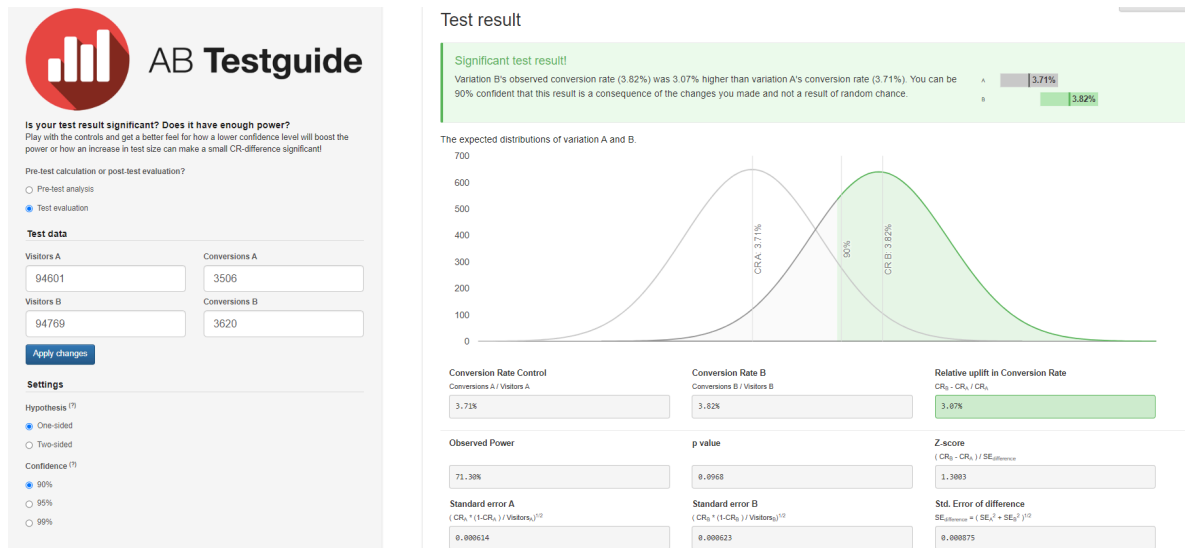
Significance level α: 5% *Percent of the time a difference will be detected, assuming one does NOT exist*

See also: [How Not To Run an A/B Test](#)

Need A/B sample sizes on your iPhone or iPad? Download [A/B Buddy](#) today.

Тобто для **Confidence level** $\alpha=0.05$ в нас недостатньо даних і потрібно значно більше даних

Тому можемо зменшити **Confidence level** $\alpha = 0.1$ і перевірити гіпотезу знову.



Тобто ми можемо прийняти нашу B гіпотезу на **Confidence level** $= 0.1$

В результаті експерименту ми отримали наступні зміни

- Вплив на метрику **CR_payment**, **ARPU**, **Середній чек юзера**, **retention_3_day**, **retention_14_day** є позитивним
- Вплив на метрику **CR_payment**, **ARPPU**, **CR_act**, **Average_act**, **CR_like**, **CR_dislike**, **CR_message**, **CR_visit**, **retention_7_day**, **retention_21_day**, **retention_28_day** є негативним, але незначними (окрім **CR_act**, **CR_dislike**)

Враховуючи зміни в комплексі пропонуємо:

- Якщо є допустимим **Confidence level** $\alpha = 0.1$ пропонуємо закрити тест та прийняти гіпотезу B, різниця в метриках незначна, метрики активності на продукті зменшилися (**CR_act** на ~20) в основному зменшилась кількість дизлайків (**CR_dislike** на ~15)
- Якщо **Confidence level** $\alpha = 0.1$ не є допустимим пропонуємо продовжити збирати дані

Завдання 2

Завдання №2

Ви провели експеримент на воронці реєстрації, метою якого є збільшення конверсії у тріал після проходження онбордингу за рахунок зміни дизайну, який користувач бачить при першому запуску додатку після його встановлення.

У вас було 2 версії: контрольна (Var A) та тестова (Var B).

Тест був активним 10 днів, у ньому прийняло участь 9000 користувачів, які рівномірно розподілялись між обома версіями.

За результатами тесту конверсія у тріал* у Var A = 12,29%, у Var B = 10,82%.

Кроки воронки та їх конверсії вказано нижче 📌

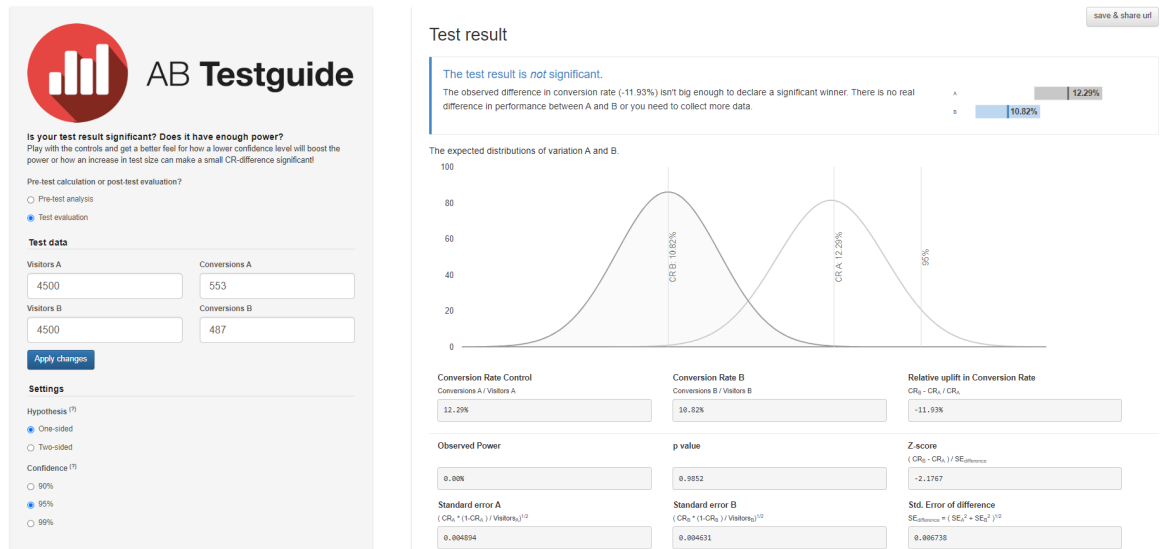
* Тріал - тестовий період користування можливостями платної підписки

Порахуємо кількість юзерів які зробили конверсії

$$\begin{aligned} \text{Var A} &= 12.29\% & CR &= \frac{\text{Total conv}}{\text{Total visit.}} \cdot 100\% (*) \\ \text{Var B} &= 10.82\% \\ \text{Total conv} &= X \\ \text{Всього юзерів в тесті} &= 9000 \\ \text{Розподіл рівномірний між групами A і B} \\ \Rightarrow \{ \text{Total visit} \}_{A,B} &= \frac{9000}{2} = 4500 \\ (*) \Rightarrow \frac{x_A}{4500} &= 0.1229 \Rightarrow x_A = 553.05 \approx 553 \\ \frac{x_B}{4500} &= 0.1082 \Rightarrow x_B = 486.9 \approx 487 \end{aligned}$$

	Var A	Var B
Total visitors	4500	4500
Conversion rate	12.29%	10.82%
Total conversion	553	487

Результат першого калькулятора:

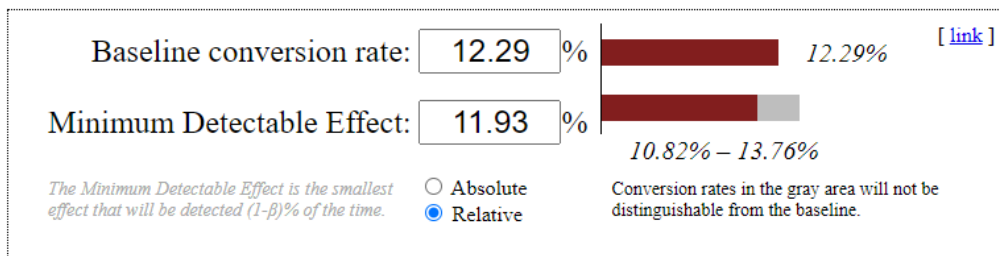


Результати другого калькулятора:

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

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

Question: How many subjects are needed for an A/B test?



Sample size:

7,989

per variation

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

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

See also: [How Not To Run an A/B Test](#)

Need A/B sample sizes on your iPhone or iPad? Download [A/B Buddy](#) today.

Калькулятори нам говорять про те що не хватає юзерів у вибірці, а саме 3489 юзерів. Якщо припустити що розподіл кількості юзерів рівномірний тобто кожного дня ~450 юзерів приходить на кожний варіант тесту то нам щоб набрати повну вибірку приблизно потрібно ~8 днів.

Якщо розглянути дані з кроків воронки та їх конверсій то можемо замітити що на другому ж кроці є різниця між Var A(86.5%) проти Var B(84.3%) (різниця з попереднім, тобто 1 кроком де 100% в кожному), далі розподіл конверсій приблизно однаковий до 6 кроку коли в Var A(80.7%) і Var B(76.3%) (на 5 кроці було Var A(81.7%) та Var B(80.1%)) і до кінця розподіли конверсій також є приблизно однаковими.

Враховуючи зміни в комплексі пропоную:

- Змінити 1 та 5 кроки воронки (різниця між падінням конверсій у Var A та Var B 2.2% та 2.8% відповідно до кроку воронки) та перезапустити тест враховуючи нові зміни