

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ТАРАСА ШЕВЧЕНКА  
Економічний факультет  
Кафедра економічної кібернетики**

**КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА  
«Розробка системи оптимізації рівня конверсії веб сайту»**

студента 4 курсу  
спеціальності 051 «Економіка»  
ОПП «Економічна кібернетика»  
денної форми навчання  
Буліка Романа Вікторовича

**Науковий керівник:**

Кандидат фізико-математичних  
наук, доцент  
Кравець Тетяна Вікторівна

Засвідчую, що у цій дипломній  
роботі немає запозичень із  
праць інших авторів без  
відповідних посилань  
Студент \_\_\_\_\_

(підпис)

Роботу допущено до захисту перед ЕК  
рішенням кафедри економічної кібернетики  
від 12 червня 2023 р., протокол № 17  
Завідувач кафедри:  
доктор економічних наук, професор  
Ляшенко Олена Ігорівна

\_\_\_\_\_  
(підпис)

КИЇВ – 2023

## РЕФЕРАТ

**Кваліфікаційна робота бакалавра містить:** 55 ст., 8 рис., 1 табл., 22 джерел, 1 додаток

**Ключові слова:** А/В тести, веб-сайт, цифровий маркетинг, оптимізація рівня конверсії, ВІ інструменти.

**Об'єкт дослідження:** веб-сайт компанії, до має потенціал для оптимізації рівня конверсії статистичними методами.

**Мета дослідження:** отримання автоматизованої системи прийняття рішень що дозволяє тестувати гіпотези про оптимізацію рівня конверсії веб-сайту.

**Методи дослідження:** А/В тестування частотним підходом, аналіз даних з допомогою ВІ інструментів.

**Наукова новизна, теоретична значимість дослідження:** систематизація досліджень та напрацювань по темі інших науковців та поєднанні знань зі сфери маркетингу зі статистичними методами для вирішення реальних задач бізнесу.

**Практична цінність:** отримання робочого інструменту, що виконує свої функції, допомагаючи оптимізувати рівень конверсії веб-сайту компанії.

## RESUME

Taras Shevchenko National University of Kyiv,

Faculty of Economics, Department of Economic Cybernetics

Key words: A/B testing, website, digital marketing, conversion rate optimization, BI tools.

The graduation research of student Roman Bulik deals with development of CRO statistics based systems.

The work is interesting for digital marketers, applied scientists in business sphere.

Pages 55, tables 1, bibliog. 22, append. 1

## ЗМІСТ

ВСТУП	4
РОЗДІЛ 1. ОПТИМІЗАЦІЯ ЧАСТИН МАРКЕТИНГУ ДЛЯ ПОБУДОВИ УСПІШНОГО БІЗНЕСУ	6
1.1. Роль маркетингу в сучасній компанії	6
1.2. Потенціал частин маркетингу до оптимізації статистичними методами	13
1.3. ВІ інструменти як основа для розробки системи оптимізації рівня конверсії веб сайту	17
РОЗДІЛ 2. МЕТОДИ ОПТИМІЗАЦІЇ РІВНЯ КОНВЕРСІЇ ВЕБ-САЙТУ	22
2.1. Основні підходи до оптимізації рівня конверсії веб-сайту	22
2.2. Частотний підхід до А/В тестування	26
РОЗДІЛ 3. ПЕРЕВІРКА ГІПОТЕЗИ ПРО ОПТИМІЗАЦІЮ РІВНЯ КОНВЕРСІЇ ВЕБ-САЙТУ	30
3.1. Опис гіпотези про оптимізацію	30
3.2. Опис даних та процесу проведення тесту	33
3.3. Застосування частотного підходу до оцінки тестів	39
ВИСНОВКИ	47
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	49
ДОДАТКИ	52

## ВСТУП

**Актуальність** теми дослідження зумовлена популярністю веб-сайтів та їх роллю в успіху сучасного бізнесу. Також підприємства мають за мету максимізацію прибутку. У цьому їм може допомогти оптимізація рівня конверсії веб-сайту.

За оцінкою Forbes станом на 2023 рік веб-сайти мають 71% підприємств. Така популярність власних веб-сайтів говорить про їх затребуваність та увагу до того як вони функціонують та наскільки ефективно виконують свої функції. Так як веб-сайт є одним із інструментів продажів, то компанії готові вкладатися у його розвиток та покращення.

Сучасний веб-сайт має багато функцій починаючи від відображення інформації про діяльність підприємства, його цілі, продукти та контактні дані закінчуючи спеціальними пропозиціями та просто корисною інформацією про сферу діяльності підприємства.

Підприємства хочуть мати ефективніші веб-сайти та готові платити за це. Проте, оптимізація рівня конверсії веб-сайту це постійний процес, що приносить свої плоди тільки на дистанції. Під рівнем конверсії веб-сайту розуміється частка візитерів, що зробили цільову дію, наприклад реєстрацію на сайті. Цільові дії називаються конверсіями. У ІТ сфері вважається за норму отримувати 1 успішний тест з 5 запущених. Тому варто підходити до цього питання системно і використовувати data driven підхід у прийнятті рішень. Розроблена автоматизована система оптимізації рівня конверсії веб-сайту може стати інструментом для ґрунтовної роботи з CRO будь-якої компанії.

**Об'єктом** дослідження є веб-сайт компанії, що має потенціал для оптимізації рівня конверсії статистичними методами.

**Предметом** дослідження є розробка автоматизованої системи оптимізації рівня конверсії веб-сайту компанії.

**Метою** дослідження є отримання автоматизованої системи прийняття рішень що дозволяє тестувати гіпотези про оптимізацію рівня конверсії веб-сайту.

**Завдання роботи:** застосувати статистичні методи оптимізації рівня конверсії веб-сайту, розробити систему оптимізації рівня конверсії веб-сайту.

**Практична значущість** полягає у отриманні робочого інструменту, що виконує свої функції, допомагаючи оптимізувати рівень конверсії веб-сайту компанії.

## РОЗДІЛ 1. ОПТИМІЗАЦІЯ ЧАСТИН МАРКЕТИНГУ ДЛЯ ПОБУДОВИ УСПІШНОГО БІЗНЕСУ

### 1.1. Роль маркетингу в сучасній компанії

Маркетинг - це комплексна система організації виробництва і збуту продукції, спрямована на задоволення потреб потенційних споживачів і досягнення прибутку шляхом проведення досліджень та прогнозування ринку, аналізу внутрішнього і зовнішнього середовища підприємства, розробки стратегії просування на ринку з використанням сучасних маркетингових технологій. Важливим етапом маркетингової діяльності є вивчення ринку та його динаміки, що включає аналіз попиту та визначення потреб споживачів, їхніх уподобань і поведінки на ринку. Дослідження ринку також дозволяє виявити конкурентну обстановку, визначити ключових конкурентів та знайти ніші, де компанія може зайняти сильну позицію [19].

Розробка стратегії маркетингу включає в себе визначення мети підприємства на ринку, вибір цільової аудиторії, розробку позиціонування продукції та формування конкурентної переваги. На основі стратегії розробляються тактики маркетингу, які включають в себе вибір маркетингових інструментів (таких як продукт, ціна, збут) і розробку маркетингових програм. Маркетингові програми включають в себе детальний план дій, спрямованих на досягнення маркетингових цілей підприємства. Ці програми орієнтовані на впровадження маркетингових стратегій, встановлення цінової політики, розробку каналів збуту, створення рекламних кампаній та взаємодію зі споживачами. Тому маркетинг є необхідною складовою сучасного підприємства, оскільки він дозволяє досягти успіху на ринку шляхом адаптації до потреб і вимог споживачів, ефективного використання ресурсів та створення конкурентних переваг [19].

Важливість маркетингу для бізнесу важко переоцінити. Маркетинг є стратегічною функцією, яка допомагає компанії привернути нових клієнтів, підвищити свою репутацію, зростати і займати конкурентну позицію на ринку.

Ось декілька ключових аспектів, що підкреслюють важливість сучасного маркетингу для бізнесу:

- Стратегічний аналіз ринку.
- Брендинг і позиціонування.
- Залучення нових клієнтів.
- Утримання існуючої клієнтської бази.
- Розширення ринку та інтернаціоналізація.
- Прийняття рішень на основі даних.

Стратегічний аналіз ринку є одним із ключових аспектів, що підкреслюють важливість маркетингу. Такого роду аналіз починається з комплексного дослідження ринкового середовища, включаючи конкурентів, потреб потенційних клієнтів і галузевих тенденцій. Проводячи дослідження ринку та збираючи інформацію, компанії можуть виявити невикористані можливості, зрозуміти уподобання клієнтів і розробити ефективні маркетингові стратегії. Завдяки ухваленню рішень на основі отриманих даних, маркетинг допомагає компаніям порівняти свої продукти чи послуги з ринковим попитом. Також цей аспект допомагає компанії розуміти конкурентне оточення, в якому вона працює. Це включає аналіз конкурентів, їхніх стратегій, сильних та слабких сторін, а також потенційних загроз для компанії. Ця інформація дозволяє розробити стратегію, що дозволяє підприємству виділитись на ринку та мати конкурентну перевагу над іншими операторами ринку [20].

Брендинг та позиціонування є важливими аспектами сучасного маркетингу для компанії. Вони сприяють побудові сильного і впізнаваного бренду та встановленню унікальної позиції на ринку. Брендинг - це процес створення та управління іміджем компанії, продукту або послуги. Він включає в себе визначення ідентичності бренду, його цінностей, обіцянок та унікальності, що дозволяє відрізнити його від конкурентів. Бренд створюється через елементи, такі як логотип, найменування, слоган, дизайн упаковки, комунікаційні матеріали та інші аспекти, що сприяють його впізнаваності та асоціаціям у свідомості споживачів. Ефективний бренд створює сильну сприйняття

споживачами і викликає емоційну зв'язок з брендом. Він допомагає побудувати довіру, лояльність і вірність споживачів, створюючи позитивний імідж компанії. Крім того, брендинг дозволяє відрізнятись від конкурентів, створюючи унікальну пропозицію на ринку і формуючи впізнавану марку.

Позиціонування, з іншого боку, відображає спосіб, яким компанія вирізняється на ринку та сприймається споживачами. Це стратегічний процес, в результаті якого компанія займає чітку та унікальну позицію у свідомості споживачів в порівнянні з конкурентами. Вона визначає свої сильні сторони, цільову аудиторію та значення, яке вона пропонує споживачам. Позиціонування вимагає глибокого розуміння цільової аудиторії та конкурентного середовища. Воно включає в себе вибір унікальних характеристик, які відрізняють компанію від інших гравців на ринку, і акцентує увагу на них у комунікації зі споживачами. Позиціонування також визначає спосіб, яким компанія сприймається споживачами в контексті цінностей, переваг та пропозицій. Успішне позиціонування дозволяє компанії отримати конкурентну перевагу, покращити свої продажі та розширити свою базу клієнтів. Це допомагає встановити унікальну ідентичність на ринку та зайняти певне місце в свідомості споживачів.

Залучення нових клієнтів є одним з найголовніших аспектів маркетингу. За допомогою цільових маркетингових кампаній, підприємства можуть охопити велику кількість потенційних клієнтів, залучити нових користувачів і перетворити їх на клієнтів, які будуть готові щось купити. Серед найпопулярніших підходів сучасного маркетингу можна виділити три:

- PPC (pay per click) модель закупки.
- CPA (cost per action) модель закупки трафіку.
- Органічне просування.

PPC модель закупки трафіку - це стратегія маркетингу, в якій рекламодавець платить лише за отримані результати або дії, які виконує споживач у відповідь на рекламний матеріал. Як можна дізнатися з назви моделі, оплата відбувається за кліки по рекламних оголошеннях, які ведуть на

веб-сайт компанії. За допомогою PPC моделі рекламодавець може ефективно використовувати свій рекламний бюджет, оскільки він платить лише за конкретні результати. Замість того, щоб платити за покази рекламних матеріалів, він сплачує тільки за кліки, тобто знімає з себе відповідальність за розміщення оголошень та доведення їх до відома потенційних клієнтів [7].

Ця модель закупки трафіку може бути реалізована через різні канали реклами, такі як контекстна реклама, електронна пошта, соціальні медіа або партнерські мережі. Однією з ключових переваг PPC моделі закупки трафіку є зниження ризику для рекламодавця. Він платить лише за результати, що означає, що витрати на рекламу пов'язані безпосередньо з отриманими цільових користувачів. Це дозволяє більш точно контролювати ефективність рекламної кампанії та вкладати бюджет у найбільш вигідні канали. Великою перевагою даної моделі є наявність великої кількості даних про поведінку користувачів, їх взаємодію з рекламними матеріалами. Саме дані дозволяють гнучко управляти таргетингом на самі релевантні аудиторії та обирати більш ефективні канали залучення користувачів [8].

CPA (cost per action) модель закупки трафіку - це стратегія маркетингу, в якій рекламодавець платить тільки за конкретні дії, виконані споживачами на його веб-сайті, наприклад, за реєстрацію, покупку, заповнення форми або завантаження файлу. У цій моделі рекламодавець встановлює конкретні дії, які він вважає цінними для своєї бізнес-моделі, і платить партнерам, рекламним мережам тільки за кожну успішну дію, яка відбувається завдяки їх рекламній діяльності. Ця модель закупки трафіку є вигідною для рекламодавців, оскільки вони платять лише за конкретні результати, які більш ймовірно ведуть до отримання виручки, ніж просто клік по рекламному оголошенню, як у PPC моделі. Вона зменшує ризик фінансових витрат на непродуктивну рекламу і дозволяє точно контролювати витрати та ефективність рекламних кампаній. CPA модель також сприяє стимулюванню рекламних партнерів до більш активних зусиль, оскільки вони отримують оплату лише за конкретні дії, а не за

простий показ або клік. Це стимулює їх працювати над ефективним привертанням цільового трафіку [10].

Головним недоліком даної моделі є ціна та якість приведеної аудиторії. Закупівля за даною моделлю значно дорожча за аналогічні об'єми за PPC моделлю. Також відсутня інформація про те, що відбувається з користувачем до потрапляння на наш веб-сайт, що може ускладнювати оцінку якості приведених користувачів на ранніх етапах користування веб-сайтом. Крім цього якість трафіку значною мірою залежить від домовленостей та якості партнерських зв'язків із підрядником.

Також існує напрямок органічного просування, що тісно пов'язаний із брендингом та репутацією. Органічне просування - це підхід у перформанс маркетингу, який спрямований на залучення цільової аудиторії та покращення результатів без прямого витрачання грошей на рекламу. Він базується на створенні і підтримці якісного та цінного контенту, оптимізації веб-сайту та активному залученні аудиторії через органічні канали комунікації, наприклад соціальні мережі або оффлайн заходи.

Органічне просування використовує методи, які засновані на рості та розвитку природним шляхом. Це може включати пошукову оптимізацію (SEO), створення цікавого та корисного контенту, активну присутність у соціальних медіа, взаємодію зі спільнотами та впливовими особами, а також використання реферальних програм та партнерських відносин. Одним з головних переваг органічного просування є його вартість. Органічні канали комунікації не вимагають прямого фінансового вкладення, оскільки їх використання передбачає власні зусилля та креативний підхід. Він може бути особливо корисним для маленьких підприємств або стартапів з обмеженим бюджетом, які не можуть конкурувати з великими рекламними кампаніями [17].

Маркетинг виходить за рамки залучення клієнтів, він також зосереджений на зміцненні довгострокових відносин із клієнтами. Завдяки ініціативам із збереження клієнтів, таким як персоналізоване спілкування, програми лояльності та підтримка клієнтів, маркетинг допомагає компаніям виходити на

новий рівень взаємодії з клієнтами. Розуміючи їхні потреби, уподобання та відгуки, компанії можуть адаптувати свої пропозиції, підвищити рівень задоволеності клієнтів і стимулювати повторні покупки.

Утримання існуючої клієнтської бази великою мірою описується терміном “ремаркетинг”. Ремаркетинг - це стратегія маркетингу, яка спрямована на залучення та утримання вже існуючих користувачів на веб-сайті. Цей підхід використовується для спонукання відвідувачів, які вже мали контакт з брендом, повертатися на сайт та здійснювати бажані дії, наприклад покупки.

Ремаркетинг може включати показ рекламних банерів, текстових оголошень або відео на інших веб-сайтах або в соціальних медіа, що відвідуються цими користувачами. Це дозволяє бренду знову звернутися до цільової аудиторії та нагадати про свої продукти або послуги. Ремаркетинг використовує куки (записи, які зберігаються на пристрої користувача) для відстеження взаємодії відвідувача з веб-сайтом. За допомогою цієї інформації рекламні кампанії можуть бути специфічно налаштовані для цих відвідувачів під час їх подальшої інтернет-активності. Загалом, цей напрям використовує всі ті самі інструменти, що і для залучення нових користувачів, але доносить інформацію, яка пов’язана із закликком повторити користувацький досвід, а не стати новим користувачем. Одна з важливих переваг ремаркетингу полягає в тому, що він дозволяє персоналізувати повідомлення для конкретних користувачів. Це може бути важливо для залучення їх уваги та стимулювання повторних дій. Наприклад, якщо користувач покинув сайт під час процесу оформлення замовлення, ремаркетингова реклама може нагадати йому про незавершену покупку та пропонувати спеціальні знижки або промокоди для завершення операції [1].

Також для утримання часто використовують листи на електронну пошту, яку користувач залишив під час реєстрації. Такі листи є персоналізованими і закликають повернутися на веб-сайт, пропонують особливі умови чи пропозиції або просто нагадують про бренд, дають додаткову інформацію чи інформують про щось.

Також для утримання часто використовують листи на електронну пошту, яку користувач залишив під час реєстрації. Такі листи є персоналізованими і закликають повернутися на веб-сайт, пропонують особливі умови чи пропозиції або просто нагадують про бренд, дають додаткову інформацію чи інформують про щось.

Розширення ринку та інтернаціоналізація теж виокремлює маркетинг з-поміж інших функцій підприємства. Маркетинг відіграє важливу роль у сприянні розширенню ринку за межі країни. Аналізуючи ринковий потенціал, реалізуючи стратегії виходу на ринок і адаптуючи маркетингові зусилля до різних культурних контекстів, компанії можуть успішно виходити на нові ринки. Маркетинг відіграє вирішальну роль у визначенні цільових сегментів, розумінні поведінки місцевих споживачів і розробці ефективних маркетингових кампаній, адаптованих до конкретних ринків. Завдяки розширенню ринку компанії можуть отримати нові джерела доходу та диверсифікувати свою клієнтську базу [20].

У епоху цифрових технологій прийняття рішень на основі даних це обов'язковий атрибут успішної компанії. Цей напрямок притаманний не тільки маркетингу, але сучасний маркетинг без data driven підходу важко уявити. Data driven підхід у сучасному маркетингу - це стратегія, яка базується на зборі, аналізі та використанні даних для прийняття обґрунтованих та ефективних рішень при управлінні. В цьому підході дані стають основою для прийняття рішень, а не суб'єктивні оцінки чи припущення. Data driven маркетинг використовує різні джерела даних, включаючи дані про споживачів, ринок, продажі, поведінку користувачів на веб-сайтах, соціальних медіа та інші. Ці дані збираються, систематизуються та аналізуються з метою отримання інсайтів та розуміння поведінки цільової аудиторії. На основі аналізу даних, маркетологи можуть визначити ефективність певних маркетингових каналів, рекламних кампаній, стратегій спільнот, а також ідентифікувати потенційні можливості для покращення результатів. Вони можуть визначити, які маркетингові заходи призводять до кращих результатів, та зосередити зусилля

на них. Data driven маркетинг також дозволяє персоналізувати комунікацію зі споживачами. На основі аналізу даних про інтереси, поведінку та покупки користувачів, маркетологи можуть надавати індивідуально налаштовані пропозиції та повідомлення, що підвищує ефективність комунікації та можливість залучення клієнтів.

Data driven підхід також дозволяє маркетологам здійснювати прогнозування та передбачення на основі наявних даних. Вони можуть використовувати аналітичні моделі та алгоритми для виявлення тенденцій, прогнозування попиту, оцінки ефективності рекламних кампаній та багато іншого.

Таким чином роль маркетингу в діяльності сучасної компанії є важливою. Процеси у цій зоні компанії тісно пов'язані з діяльністю інших частин підприємства. Існує багато інструментів сучасного маркетингу, що можуть опиратися на аналіз та обробку даних. Data driven підхід у прийнятті рішень є невід'ємною характеристикою сучасних компаній.

## **1.2. Потенціал частин маркетингу до оптимізації статистичними методами**

Сучасний маркетинг має достатньо розгалужену структуру і розвивається з кожним роком. Так як ІТ технології розвиваються швидко, то деякі ролі перестають бути актуальними з часом, нові ролі з'являються, а деякі просто видозмінюються. Загалом, на даний момент можна виділити такі частини маркетингу:

- Залучення нових користувачів за PPC моделлю.
- Залучення нових користувачів за CPA моделлю.
- Органічний напрям залучення користувачів.
- Напрямок оптимізації конверсії сторінок веб-сайту (CRO).
- Ремаркетинг та утримання користувачів.
- Прямі продажі.

- R&D (research and development).

Про залучення нових користувачів за PPC моделлю уже багато описано у пункті 1.1 роботи. Але варто звернути увагу на те, що чому залучення користувачів за PPC та CPA моделями варто розділяти в контексті оптимізації процесів та підвищення ефективності. Головними відмінностями цих напрямів є асиметрія інформації та ступінь залученості партнерів до отримання результатів.

Асиметрія інформації полягає у тому, що за PPC моделлю маркетинг компанії отримує максимально можливий об'єм даних про користувача та його поведінку до того як потенційний клієнт потрапить на веб-сайт і залишить свої персональні дані. Ці дані можуть включати інформацію про те з якої рекламної мережі прийшов користувач, який рекламний матеріал зацікавив людину, а також потенційні інтереси користувача. Вся ця інформація не доступна при CPA моделі закупки через те, що трафік залучає сторонній партнер, який не розголошує спосіб в який він залучив юзера. Відповідно повну інформацію про людину починаємо отримувати тільки з моменту візиту на наш веб-сайт [7]. Залученість партнерів до отримання бажаного результату полягає у тому, що за PPC моделлю самі купуємо рекламу напряму в рекламних мережах таких як Google Ads, Meta, TikTok Ads та ін. Процес покупки уніфікований для всіх користувачів цих платформ та дає рівні умови по контролю якості закупки реклами. При роботі за CPA моделлю компанія домовляється із підрядниками про те якої якості мають виходити користувачі та покладаємося на добропорядність партнера. Багато питань вирішуються на рівні відносин конкретних людей і ці відносини не стандартизовані для всіх покупців та продавців на ринку, що робить використання цього методу непрозорим для ринку.

PPC напрям залучення користувачів має великий потенціал для оптимізації статистичними методами так як отримує багато інформації про користувачів під час розміщення рекламних матеріалів. Проте, існують обмеження та особливості рекламних платформ, які ускладнюють проведення повноцінних

тестів та досліджень аудиторії та того як потенційні клієнти взаємодіють з рекламою. Головною проблемою цього напрямку є неможливість створити достатньо рівні умови для того, щоб порівнювати між собою різні підходи залучення користувачів. Занадто багато важливих факторів залишають непідконтрольними маркетологу при проведенні тестів, тому будь-який результат тестування можна списати на випадковість [8].

CPA напрям має низький потенціал для оптимізації статистичними методами через брак інформації та обмежені важелі впливу на рекламну діяльність. Цей напрям значно краще розвивається методами бізнес девелопменту та побудовою відносин із партнерами.

Органічний напрям залучення користувачів має обмежені можливості для оптимізації статистичними методами. Обмеження пов'язані з проблемами, описаними для PPC напрямку. Основна діяльність у цьому напрямі пов'язана зі спробами зрозуміти алгоритми роботи новинних стрічок у соціальних мережах. Там краще працюють експертні методи.

Напрямок оптимізації конверсії сторінок веб-сайту має високий потенціал для оптимізації статистичними методами. Цей напрям полягає у постійній розробці та покращенні сторінок веб-сайту таким чином, щоб більше користувачів робили цільові дії, наприклад залишили свою електронну пошту або купили товар чи послуги. Конверсією називають будь-яку цільову дію користувача. Рівнем конверсії веб сайту називають частку візитерів на веб-сайт, що зробили конверсію, тобто цільову дію.

Важливою перевагою цього напрямку є наявність великої кількості точних даних про характеристики користувача та його поведінку. Серед даних, що можуть мати інформацію про користувача можуть бути куки його браузера, юзер-агент, Ір адреса. Дані про поведінку користувача містять інформацію про те як людина взаємодіє з веб-сайтом. Такі дані, зазвичай, збираються за допомогою власного трекінгу, який записує логи про натискання кнопок, введення даних та переходи між сторінками веб-сайту. Також із цією задачею можуть допомогти сторонні сервіси, які дадуть можливість трекінгу без

необхідності розробляти власний трекінг чи взагалі писати будь-який код. Самим популярним з таких рішень є Google Tag Manager. Покращення користувацького досвіду є важливим аспектом CRO, що включає врахування швидкості завантаження сторінок, вирішення проблем навігації, забезпечення простоти та зручності використання, адаптивного дизайну для різних пристроїв тощо.

Саме ця зона маркетингу є вигідною частиною бізнесу для тестування гіпотез. Головною передумовою для цього є повний контроль над усім, що відбувається на веб-сайті та коректний трекінг, який дозволяє оцінювати ситуацію та зважено приймати рішення за data driven підходом. Контроль над досвідом користувача дозволяє створити достатньо рівні умови для різних груп A/B тесту. Такі передумови дозволяють довіряти результатам тестів та впроваджувати зміни відповідно до зроблених висновків.

Незважаючи на те, що більшість бізнесів мають власний веб-сайт, лише близько 200 млн веб-сайтів є активними, підтримуються власниками та мають відвідувачів, що все одно дуже великий об'єм та підтверджує актуальність напряму оптимізації [12].

Ремаркетинг та утримання користувачів має хороший потенціал для оптимізації статистичними методами. Для цього обмеження схожу з PPC через частково спільні інструменти. Але цей напрям стратегічно важливий і навіть прості дії у цьому напрямі можуть значно впливати на фінансовий успіх компанії [1].

Прямі продажі схожі до CPA напрямку через те, що відносини між людьми грають основну роль в успіху напрямку. Для досягнення кращих результатів у цьому напрямку ефективніше займатися покращенням самого процесу продажів, який важко оцінити цифрами та застосувати до нього статистику. Потенціал цього напрямку в контексті цієї роботи низький.

R&D напрям має високий потенціал для покращень ефективності маркетингу статистичними методами. R&D (Research and Development) в маркетингу - це напрямок, який спрямований на дослідження та розробку нових

маркетингових стратегій, ідей, технологій і інновацій з метою покращення конкурентоспроможності компанії та задоволення потреб споживачів. Напрямок сприяє інноваціям у маркетингових стратегіях та інструментах. Це може включати впровадження нових технологій, використання цифрових маркетингових каналів, персоналізацію комунікації зі споживачами, розробку креативних рекламних кампаній тощо. R&D в маркетингу передбачає проведення тестування нових ідей, стратегій та інструментів з метою виявлення їх ефективності. Це дозволяє компанії визначати успішні стратегії та вдосконалювати свої підходи на основі отриманих даних. Проте, цей напрям має не великий вплив на фінансові показники компанії в моменті, відповідно варто обирати напрями, які при оптимізації забезпечать більший приріст у фінансових показниках.

Серед всіх розглянутих частин маркетингу є ті, де статистичні методи справді застосовні та можуть дати значний приріст у ефективності та ті частини, де краще використовувати інші методи, наприклад експертні. Серед тих частин, де статистика застосовна, можна виділити напрям оптимізації конверсії сторінок веб-сайту (CRO). Цей напрям впливає на ефективність всього залучення користувачів на веб-сайт, тобто PPC, CPA, органічного залучення та прямих продажів. Робота у цьому напрямі дозволяє отримувати більше клієнтів при тих же самих витратах на рекламу, що дуже вигідно на великих масштабах, коли витрати на утримання команди CRO незмірно нижчі, ніж маркетингові бюджети на залучення нових користувачів. Крім важливості цей напрям ще й зручний для проведення тестів та перевірки гіпотез. Ці фактори роблять напрям CRO пріоритетним для застосування статистичних методів в цілях оптимізації рівня конверсії веб-сайту [8].

### **1.3. BI інструменти як основа для розробки системи оптимізації рівня конверсії веб сайту**

Business Intelligence (BI) — це основні інструменти, які використовуються бізнесом для збору, аналізу та представлення даних і інформації для прийняття

обґрунтованих рішень. Ці інструменти допомагають організаціям отримати уявлення про свою діяльність, ринкові тенденції, поведінку клієнтів та інші ключові фактори прийняття стратегічних і тактичних рішень. Цей напрям допомагає інтерпретувати велику кількість даних, звертаючи увагу лише на ключові фактори, що впливають на ефективність, моделювати результат різних варіантів розвитку подій, відстежувати результати прийняття рішень. ВІ мають 4 основні функції:

- Збереження даних
- Обробка даних
- Аналіз даних
- Представлення даних

Збереження даних полягає у отриманні даних та їх зберіганні у доступності для користувача. Отримання може відбуватися з інших систем, наприклад від користувача, із внутрішніх та зовнішніх трекінгових систем, з інших сховищ даних. Зберігання даних відбувається у базах даних. Системи управління даними забезпечують безпечне зберігання та управління даними. Вони забезпечують механізми для створення баз даних, таблиць, схем, полів та інших об'єктів для організації та структурування даних. Компанії можуть використовувати реляційні чи нереляційні бази даних для зберігання користувальницької інформації. Системи управління даними дозволяють створювати таблиці або колекції, де кожен запис містить дані про окремого користувача, такі як ім'я, адреса, контактна інформація та інші деталі. Сьогодні часто використовують хмарні сервіси для зберігання даних. Компанії можуть використовувати такі сервіси як Amazon Web Services (AWS), Google Cloud або Microsoft Azure, для зберігання даних про користувачів. Ці сервіси надають масштабовані рішення для зберігання та управління великим обсягом даних [2].

Обробка даних передбачає інтеграцію даних, яка виражається у об'єднанні даних з різних джерел в єдине сховище, що дозволяє аналізувати дані в комплексі. Також під час обробки відбувається очищення даних, яке полягає у виявленні та виправлення помилок, дублікатів та інших неправильностей в

даних для забезпечення їх якості та достовірності. Цей процес допомагає зробити дані більш надійними та придатними для подальшого використання в цілях аналізу. Також на етапі обробки можуть з'являтися нові дані на основі старих. Прикладом можуть бути розрахунки нових метрик із вже існуючих (наприклад, з даних про витрачену на рекламу суму та кількість залучених користувачів можна порахувати ціну за користувача) або отримання додаткових параметрів користувача, які можна отримати із вже існуючих параметрів, наприклад назви рекламної кампанії.

Можливості для аналізу даних є однією із найважливіших частин ВІ інструментів. Аналіз даних - це процес виявлення, інтерпретації та розуміння закономірностей, трендів та інсайтів, які містяться в наборі даних. Цей процес включає в себе застосування різноманітних методів, моделей та статистичних підходів для виявлення корисної інформації та висновків з даних. Аналіз даних може здійснюватися за допомогою візуалізаційних інструментів, статистичного аналізу, машинного навчання та інших технік, що дозволяють розкрити приховані залежності та зробити прогнози. Цей процес допомагає компаніям зрозуміти свій бізнес, виявити можливості для покращення та прийняти обґрунтовані рішення. Отримані знання та інсайти використовуються для прийняття рішень в компанії. Це може включати вдосконалення стратегії, зміни в продуктах або послугах, оптимізацію маркетингових кампаній та реагування на зміни в ринковому середовищі. Аналіз даних має позитивний вплив на компанію в разі, якщо він проводиться вірно та базується на правильних даних, що ще раз підкреслює важливість попередніх функцій ВІ інструментів. Проте варто слідкувати за якістю цього процесу, бо поганий аналіз або аналіз на основі поганих даних приводить до невірно прийнятих рішень та втрачених або недоотриманих прибутків компанії.

Представлення та візуалізація даних є важливою функцією ВІ інструментів та наступним етапом після етапу аналізу даних, оскільки вони допомагають зрозуміти та сприйняти інформацію легше і швидше, що дуже важливо для

бізнесу. Графічне зображення даних дозволяє легше сприймати і розуміти великі обсяги інформації та швидше виявити взаємозв'язки, тренди та патерни.

Візуалізація даних полягає у використанні графічних зображень, діаграм, графіків та інших візуальних елементів для подання даних у зрозумілій для користувача формі. Часто візуалізація є компактним відображенням результатів аналізу в зручній формі. Візуалізація даних дозволяє ефективно комунікувати результати аналізу з зацікавленими сторонами. Графіки та діаграми мають бути зрозумілими для різних аудиторій, навіть для тих, хто не має глибоких знань у сфері аналізу даних. Хороша візуалізація знижує когнітивне навантаження зацікавлених осіб та зберігає значну кількість часу при прийнятті рішень. Також важливим аспектом є надійність інформації та однозначна інтерпретація візуалізації [16].

Одним із найпопулярніших у сфері бізнесу та маркетингу Ві інструментів є Tableau. Популярність пов'язані з тим, що інструмент дозволяє легко візуалізувати та аналізувати дані, зробити їх зрозумілими та доступними для прийняття обґрунтованих рішень. Він може бути використаний для моніторингу результатів маркетингових кампаній, аналізу продажів, виявлення трендів та інших аналітичних завдань, спрямованих на вдосконалення бізнес-процесів та стратегій. Дані цей продукт отримує зі сторонніх баз даних, тому не покриває перший пункт функцій Ві систем, але добре справляється з усіма іншими функціями [15].

Головною перевагою цього інструменту є можливість зберегти виконану роботу з візуалізації та поширити її між іншими користувачами, зберігши інтерактивний формат. Також важливою особливістю Tableau є висока гнучкість та великий набір функцій. Tableau дозволяє розробити систему, що буде автоматично забирати потрібну інформацію з бази даних із заданою періодичністю, наприклад раз на день о конкретній годині, обробляти ці дані за наперед заданими формулами. Також є можливість автоматизувати прийняття рішень і перетворити кінцевий продукт із системи підтримки прийняття рішень на повноцінну систему прийняття рішень.

Також цей інструмент достатньо швидко працює, щоб обробляти великі об'єми даних та проводити складні статистичні обчислення на базі цих даних. Ця особливість є важливою так як зручний інструмент для візуалізації не буде задовольняти потреби користувачів, якщо працюватиме повільно чи нестабільно.

Таким чином, можливо розробити повноцінну систему, яка буде допомагати при оптимізації конверсії веб-сайту статистичними методами на базі ВІ інструментів, наприклад розглянутого Tableau. Такі інструменти допоможуть опрацювати, аналізувати та візуалізувати дані. Для забезпечення роботи з даними можна використати різноманітні бази даних та операції над ними. Самі дані потрібно брати з уже готових інших систем по збору та передачі даних.

## РОЗДІЛ 2. МЕТОДИ ОПТИМІЗАЦІЇ РІВНЯ КОНВЕРСІЇ ВЕБ-САЙТУ

### 2.1. Основні підходи до оптимізації рівня конверсії веб-сайту

Існує два основні напрями методів, що можуть лягти в основу системи оптимізації рівня конверсії веб-сайту:

- **Експертні методи.** Ця група методів включає використання знань та експертизи фахівців для оцінки, аналізу та тестування гіпотез щодо оптимізації конверсії веб-сайту. Цей підхід може бути особливо корисним на ранніх стадіях оптимізації, коли відсутні достатні обсяги даних або коли потрібна швидка оцінка певних аспектів веб-сайту.

- **Статистичні методи.** Ця група методів полягає у перевірці статистичних гіпотез. Статистичні методи грають важливу роль у покращенні конверсії веб-сайту, оскільки вони надають об'єктивні дані та висновки на основі зібраних даних. За допомогою такого підходу можна отримувати цінну інформацію про об'єктивну реальність і мати високий рівень довіри до отриманих висновків.

Експертні методи мають велику кількість інструментів та часто застосовуються у класичному маркетингу. Можна виділити популярні підходи цього напрямку:

- **Експертні огляди (Expert Reviews):** фахівці проводять детальний огляд веб-сайту з точки зору дизайну, навігації, вмісту та інших факторів, які можуть впливати на конверсію. Вони шукають потенційні проблеми та пропонують покращення на основі своєї експертизи та знань про користувачів.

- **Експертні інтерв'ю (Expert Interviews):** фахівці спілкуються з ключовими зацікавленими сторонами, включаючи маркетологів, дизайнерів, розробників та інших фахівців, для збору інсайтів та думок щодо потенційних покращень. Це дозволяє врахувати різні перспективи та експертні знання при формулюванні гіпотез.

- **Карти персонажів (User Personas):** фахівці розробляють детальні профілі та описи типових користувачів веб-сайту. Це допомагає уявити, які цільові аудиторії використовують сайт, їх потреби та перешкоди. Інформація

про персонажів може використовуватись для ідентифікації можливих покращень, спрямованих на конкретні групи користувачів.

- **Фокус групи (Focus Groups):** фахівці можуть бути запрошені для участі в групових дискусіях або мозкових штурмах щодо підвищення конверсії. Вони обговорюють ідеї, аналізують різні можливості та надають свої рекомендації на основі своїх знань та досвіду.

Серед статистичних методів популярними самими популярними є А/В тестування. Цей метод полягає у тому, що випадковим чином розподіляються відвідувачі між двома групами: контрольною групою (А), де застосовується поточний варіант, і тестовою групою (В), де впроваджується новий варіант. Новий варіант веб-сайту створюється на основі раніше згенерованої гіпотези про оптимізацію. В основі цього методу лежать статистичні підходи до тестування гіпотез. Нульова гіпотеза ( $H_0$ ) та альтернативна гіпотеза ( $H_1$  або  $H_a$ ) є ключовими поняттями у цьому методі. Ці гіпотези є конкуруючими твердженнями, які висуваються для перевірки статистичних даних, які збираються на реальних версіях веб-сайту [18].

Нульова гіпотеза ( $H_0$ ) виражає припущення про відсутність будь-якої статистично значущої різниці, ефекту. Вона стверджує, що будь-яка спостережувана різниця між групами або змінними є випадковою або статистично незначущою. Нульова гіпотеза зазвичай формулюється таким чином, щоб її можна було перевірити за допомогою статистичного тесту. Альтернативна гіпотеза ( $H_1$  або  $H_a$ ) протилежна нульовій гіпотезі і виражає припущення про наявність статистично значущої різниці, ефекту або зв'язку між змінними. Вона стверджує, що спостережувана різниця або ефект мають статистичну значущість і не пояснюються випадковістю [21].

Процес тестування гіпотез полягає у зборі даних, обчисленні статистики тесту і прийнятті рішення щодо прийняття або відхилення нульової гіпотези на підставі отриманих результатів. Якщо дані надають достатні докази на користь альтернативної гіпотези і проти нульової гіпотези, то нульова гіпотеза

відхиляється. В іншому випадку, якщо докази недостатні, нульова гіпотеза приймається.

Існує два основні підходи до проведення А/В тестів:

1. А/В тестування за частотним підходом. Частотний підхід до ймовірності визначає ймовірність події по відношенню до того, як часто певна подія відбувається у великій кількості випробувань. цей підхід використовує ймовірність лише для моделювання певних процесів, які широко описуються як *sampling* - формування вибірки. Тобто за цим підходом дані це багато випадкових вибірок, звідси і назва [14].

2. А/В тестування за Байєсівським підходом. Байєсовська статистика розраховує ймовірність того, що подія є істинною, враховуючи отримані результати (дані) і знання про цю ймовірність до початку дослідження (апріорна ймовірність). Цей підхід базується на теоремі Байєса, сформульованій Томасом Байєсом, англійським статистиком 18 століття. Байєсівська система має справу з умовною ймовірністю, яка є ймовірністю події за умови, що відбулася інша подія. Основною характеристикою байєсівського методу є невизначеність щодо значення параметра сукупності, на відміну від частотного аналізу, у якому параметри вважаються фіксованими. Байєсівський підхід використовує ймовірність для моделювання як вибірки, так і інших видів невизначеності при перевірці гіпотез [5].

Головним недоліком експертних методів є те, що вони покладаються на суб'єктивну оцінку спеціалістів та користувачів. Їх думка може бути помилковою у випадку саме нашого веб-сайту, але працювати на іншому. Так як будується система, що буде стабільно та послідовно вносити зміни у веб сайт, то варто обрати більш надійні статистичні методи, які дозволяють добре контролювати прийняття рішень щодо змін. Отже, основою нашої системи стануть статистичні методи [22].

Проте, у обраному напрямку є важливий вибір щодо підходу до реалізації А/В тестів. Жоден із двох основних підходів не є однозначно кращим за інший.

Вони використовують у різних цілях через свої особливості у інтерпретації ймовірностей.

Частотний підхід має такі переваги над Байєсівським:

- **Об'єктивність:** частотний підхід ґрунтується на принципах частотної ймовірності та математичних статистик, що надає об'єктивну основу для аналізу даних та прийняття рішень.

- **Відносно простий у реалізації:** частотний підхід має широкий набір статистичних методів, які легко використовувати для аналізу даних. Це включає такі методи, як t-тест, ANOVA, регресійний аналіз та інші.

- **Добре підходить для великих вибірок:** частотний підхід зазвичай працює краще з великими вибірками, оскільки він заснований на законах великих чисел та центральній граничній теоремі.

- **Простота у інтерпретації.**

Недоліки частотного підходу:

- **Залежність від припущень:** частотний підхід часто ґрунтується на припущеннях про розподіл даних, незалежність спостережень тощо. Якщо ці припущення не виконуються, можуть виникнути проблеми з точністю та надійністю отриманих результатів.

- **Фіксований підхід до невизначеності:** частотний підхід не забезпечує прямого розгляду невизначеності або попередньої інформації про параметри системи. Він розглядає параметри як фіксовані значення, що може призводити до обмеженої роботи з невизначеністю.

Попри все у іншого підходу також є своє переваги та недоліки, які варто врахувати при виборі. Переваги Байєсівського підходу:

- **Урахування невизначеності:** Байєсівський підхід дозволяє явно враховувати невизначеність шляхом використання апіорних розподілів та оновлення їх з урахуванням даних. Це дозволяє отримати розподіли ймовірностей для параметрів моделі.

- Гнучкість: Байєсівський підхід дозволяє використовувати різноманітні моделі та пристосовувати їх до конкретних випадків. Він також дає змогу враховувати експертні знання та попередню інформацію при моделюванні та прийнятті рішень.

Недоліки Байєсівського підходу:

- Вимоги до попередньої інформації: Байєсівський підхід вимагає наявності попередньої інформації або апріорного розподілу для параметрів моделі. Якщо така інформація відсутня або недостатня, можуть виникати проблеми з точністю та надійністю результатів.

- Вища складність обчислень: Байєсівський підхід вимагає проведення обчислень, таких як оновлення апріорного розподілу з урахуванням даних. Це може бути більш складним та витратним обчислювальним процесом порівняно з частотним підходом.

- Інтерпретація ймовірностей: Байєсівський підхід надає не пряму інтерпретацію ймовірностей. Це призводить до складнощів у формулюванні гіпотез та складнішої інтерпретації результатів тестування [6].

Враховуючи всі вище перелічені переваги та недоліки підходів будемо використовувати частотний підхід до проведення А/В тестів. Головною перевагою з-поміж всіх є зрозуміліша інтерпретація результатів, що має позитивно вплинути на досвід користування системою оптимізації рівня конверсії веб-сайту. Саме швидкість та простота рішень в пріоритеті бізнесу, коли це стосується пошуку можливостей для росту та покращень.

## **2.2. Частотний підхід до А/В тестування**

Частотний підхід до А/В тестування це підхід, що за основу бере гіпотезу про відсутність різниці між досліджуваними величинами. Ця гіпотеза перевіряється з використанням таких понять як помилки першого та другого роду, p-value та power.

Базовий план використання цього підходу складається із наступних пунктів:

1. Формулювання нульової та альтернативної гіпотези. Нульова гіпотеза полягає у тому, що різниці між тестовими групами немає, а видно різницю, то це випадковість. Альтернативна гіпотеза протилежна до нульової та полягає у тому, що існує статистично значуща різниця між групами тесту, яка не пояснюється випадковістю.

2. Розрахунок вибірки для проведення тесту. Вибірka має бути достатньою, щоб результати тесту виявилися статистично значущими із заданим рівнем надійності.

3. Розрахунок статистичних показників після завершення тесту. Рахуються такі показники як  $p$ -value та power. Обраховані значення порівнюються із цільовими попередньо заданими для цього тесту.

4. Прийняття рішення про відхилення або прийом нульової гіпотези. На основі порахованих статистичних показників відбувається прийняття рішення про результати тесту. Інтерпретація гіпотез дає змогу запровадити отримані результати в життя.

Ідентифікація гіпотези для перевірки, як правило, відбувається через знання предметної області проблеми, що вирішується. Нульова гіпотеза – це, як правило, твердження щодо популяції, яке вважається істинним. Альтернативна гіпотеза - це твердження, яке суперечить нульовій гіпотезі. Простий приклад можна окреслити в наступному сценарії: компанія хоче підвищити коефіцієнт конверсії користувачів, які відвідують наш веб-сайт, додавши окрему функцію. Нульова гіпотеза полягала б у тому, що додавання цієї відмінної функції на веб-сайт не матиме впливу на коефіцієнт конверсії. Альтернативна гіпотеза полягає в тому, що додавання цієї нової функції вплине на коефіцієнт конверсії.

Нульова гіпотеза є початковим припущенням або твердженням, яке стверджує відсутність будь-якого ефекту, різниці чи зв'язку між порівнюваними змінними. Зазвичай нульова гіпотеза формулюється таким чином, щоб вона була протилежна до того, що дослідник сподівається показати або виявити в аналізованій вибірці.

Альтернативна гіпотеза є протилежною нульовій гіпотезі. Вона формулюється для вираження конкретних ефектів, різниць чи зв'язків, які дослідник хоче підтвердити на основі даних. Альтернативна гіпотеза може бути односторонньою, коли дослідник очікує лише позитивний або лише негативний ефект, або двосторонньою, коли дослідник очікує будь-яку відмінність в будь-якому напрямку. У нашому випадку буде сформульована саме двостороння гіпотеза через те, що не відомо як саме внесені зміни впливають на користувачів [5].

Для розрахунку розміру вибірки потрібно визначити очікувану точність. Задані параметри точності відображають ймовірності зробити помилку приймаючи рішення. Помилка першого роду (Type I Error) відбувається, коли нульова гіпотеза відхиляється, хоча вона насправді є правильною. Це означає, що приймається статистично значущий ефект там, де його немає. Рівень значущості ( $\alpha$ ) визначає ймовірність зробити помилку першого роду. Помилка другого роду (Type II Error) відбувається, коли нульова гіпотеза приймається, хоча насправді альтернативна гіпотеза є правильною. Це означає, що не був виявлений статистично значущого ефекту, коли він існує. Ймовірність помилки другого роду позначається як  $\beta$  (бета). Зазвичай, стандартним рівнем надійності обирають  $\alpha = 0.05$ , а  $\beta = 0.8$ .

Розмір вибірки для кожної групи тесту рахується за наступною формулою:

$$n = \frac{(Z_{\alpha/2} \sqrt{2p_1(1-p_1)} + Z_{\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)})^2}{|p_2 - p_1|^2}$$

, де

$p_1$  - конверсія контрольної групи,  $p_2$  – сума конверсії контрольної групи та мінімального очікуваного ефекту зміни, тобто очікувана конверсія тестової групи,  $Z(\alpha/2) = 1.96$ ,  $Z(\beta) = 0.84$ .

Статистичні показники такі як p-value і power, а також  $Z(\alpha/2)$  і  $Z(\beta)$  розраховуються по таблицях-довідниках Z-Score для двосторонніх тестів [9].

Рішення приймається на основі порівняння розрахованих значень  $p$ -value і  $power$  із попередньо заданими при плануванні тесту. Можливі наступні варіанти рішень:

- $p\text{-value} > 0.5$ ,  $power > 0.8$  - приймаємо нульову гіпотезу, бо ймовірність помилки першого роду вища, ніж планували;
- $p\text{-value} > 0.5$ ,  $power < 0.8$  - приймаємо нульову гіпотезу, бо помилки ймовірності помилок першого та другого роду вищі, ніж очікувано;
- $p\text{-value} < 0.5$ ,  $power < 0.8$  - приймаємо нульову гіпотезу, бо ймовірність помилки другого роду вище, ніж запланована;
- $p\text{-value} < 0.5$ ,  $power > 0.8$  - відхиляємо нульову гіпотезу через те, що зміни між тестовими групами виявилися достатньо великими, щоб ймовірності помилок першого та другого роду були на рівні не вище заданого нами перед тестом;

## **РОЗДІЛ 3. ПЕРЕВІРКА ГІПОТЕЗИ ПРО ОПТИМІЗАЦІЮ РІВНЯ КОНВЕРСІЇ ВЕБ-САЙТУ**

### **3.1. Опис гіпотези про оптимізацію**

Для того, щоб почати покращувати власний веб-сайт за допомогою контрольованих тестів потрібно знайти точку на яку можна максимально вплинути, щоб отримати найефективніші абсолютні результати. Під абсолютними результатами маються на увазі натуральні показники приросту проходження юзерів через сторінку веб-сайту, що оптимізується. Так як головною ціллю компанії є максимізація прибутку, то найефективніше впливати на групи користувачів, які приносять більше виручки [4].

Процес генерації гіпотез про оптимізацію рівня конверсії веб-сайту шляхом А/В тесту зазвичай включає наступні кроки:

1. Збір даних та пошук можливостей: першим етапом є аналіз даних про веб-сайт, таких як метрики відвідуваності, конверсії, час на сайті та інші показники взаємодії користувачів. Цей етап може включати використання веб-аналітики, інструментів відстеження та інших джерел даних. На основі аналізу даних можна виявити області, які потребують поліпшень або оптимізації. Це можуть бути, наприклад, сторінки зі слабкою конверсією, проблеми зі швидкістю завантаження або незадовільні елементи дизайну.

2. Формулювання гіпотез: на основі виявлених можливостей формулюються конкретні гіпотези про те, які зміни або експерименти можуть призвести до поліпшення веб-сайту. Гіпотези повинні бути конкретними та реальними для перевірки.

3. Планування А/В-тесту: розробляється план для проведення А/В-тесту, включаючи визначення варіантів (А і В), розподіл відвідувачів між групами, тривалість тесту та метрики, які будуть вимірюватися.

4. Проведення тесту: А/В-тест включає розподіл відвідувачів між варіантами А і В, запуск змін (наприклад, нового дизайну, тексту або функцій) для варіанту В і збір даних про поведінку користувачів.

5. Аналіз результатів: після закінчення тесту проводиться аналіз результатів, включаючи порівняння метрик та оцінку статистичної значущості різниці між варіантами А і В за допомогою статистичних методів. Із доступних підходів до оцінки результатів тестування будемо користуватися частотним підходом.

6. Висновки і впровадження: на основі результатів тесту робляться висновки щодо ефективності та впливу змін. Якщо варіант В показує статистично значущі покращення, то він може бути впроваджений на веб-сайт. Також, виходячи із поставлених гіпотез, можуть бути винесені додаткові знання про ефективність тих чи інших підходів.

Експертним методом була сформульована гіпотеза про оптимізацію. **Гіпотеза** звучить наступним чином: можливість поміняти розмір інтерфейсу на веб-сайті може допомогти реєструвати більше цільових користувачів. Під цільовими користувачами маються на увазі ті, хто готові платити за продукт.

Для того, щоб підтвердити коректність формулювання гіпотези, та мати можливість приблизно оцінити потенційний вплив майбутніх змін на загальну ефективність веб-сайту, проведемо попередній аналіз. Також попередній аналіз дозволяє нам бути впевненими в тому, що інформація на базі якої будується гіпотеза вірна.

Подивимося на ефективність різних сегментів користувачів в розрізі вікових груп. Такий погляд на дані може допомогти краще зрозуміти потреби та стиль користування веб-сайтом тих груп користувачів, що більше впливають на прибуток, ніж інші.

У Табл 3.1 показаний розподіл користувачів, що зробили хоч одну оплату за весь період користування сервісом по вікових групах. Бачимо, що більшість із таких користувачів належать до старших вікових груп, а саме біля 30% користувачів мають вік між 46 та 55, а також біля 41% користувачів мають вік більше 56 років.

Таблиця 3.1

Розподіл зареєстрованих користувачів по віку серед тих, хто заплатив  
принаймні 1 раз

Age	Users share
65+	13,92%
56-65	27,17%
46-55	30,30%
36-45	17,61%
18-35	11,00%

*Джерело: складено автором на основі даних компанії*

Люди старшого віку надають перевагу простим та зрозумілим інтерфейсам, що не мають лишніх деталей та не перевантажені контентом. Крім цього у значної частини цих користувачів можуть бути проблеми із зором, що може затрудняти отримання важливої для подальшого користування інформації, а також критично важливих для реєстрації даних, таких як ім'я, вік, електронна пошта, стать та пароль.

Потенційно можемо допомогти деякій частині старших юзерів простіше сприймати інформацію та реєструватися без складнощів. Можливо, деякі з юзерів мали критичні проблеми при реєстрації саме через складнощі зі сприйняттям інформації. Допомогти цим юзерам можна, наприклад, надавши можливість самостійно обирати розмір шрифту на веб-сайті. Таким чином було підтверджено те, що вплив на вікових юзерів має бізнесовий сенс та сформульована гіпотеза не протирічить наявним у нас даним.

В тесті буде приймати участь 2 групи:

- Базова - веб-сайт без змін, тобто без налаштувань розміру шрифту.
- Тестова - веб-сайт із функціоналом, що дозволяє змінити розмір шрифту в тому числі збільшити його до комфортного для юзера розміру.

Для оцінки термінів проведення тесту можна подивитися на середню кількість візитів на веб-сайт у тому сегменті трафіку, на якому збираємося проводити тест. За тиждень перед тестом веб-сайт отримував в середньому 53

тисячі візитів, тобто біля 26,5 тисяч візитів на кожну з груп тесту. Відповідно до частотного підходу проведення тестів нам потрібно отримати принаймні по 54 тисячі візитів на кожну з груп тесту. Очікувана тривалість тестування складає 3 дні.

### **3.2. Опис даних та процесу проведення тесту**

Для проведення тесту була обрана частина платного трафіку, що йде на веб сайт із рекламної мережі Google Display Network та закупляється внутрішніми PPC спеціалістами компанії [13].

Такий вибір зумовлений наступними фактами:

- Даний сегмент трафіку займає велику частку у структурі платного трафіку. Відповідно оптимізація такого значного сегменту дасть більший ефект на ефективність всього платного трафіку загалом, ніж оптимізація менших сегментів.
- Цей сегмент приводить стабільно велику кількість візитів на веб сайт щодня. Завдяки цьому можна достатньо точно оцінити терміни проведення A/B тесту, що важливо для планування роботи команди.
- GDN трафік має достатньо однорідну структуру та стабільну якість аудиторії. Цей фактор дозволяє отримувати достатньо однорідну вибірку для тестових груп навіть при порівняно невеликих об'ємах трафіку.

Для розподілу візитерів між групами, тобто для процесу сплітування трафіку, на ринку існує кілька готових рішень. Найбільші серед них це Optimizely, VWO та Google Optimize. Проте, кожне з цих рішень має непрозору механіку сплітування трафіку, що ускладнює прийняття рішень по тесту, призводить до нерівних по розміру вибірки груп та, відповідно, сумніві щодо однорідності трафіку саме в кожному конкретному A/B тесті.

Також сучасні сайти можуть бути побудовані як на основі застарілих технологій, так і на основі самих нових та передових фреймворків, мов програмування та технологій. Тому ринкові інструменти побудовані таким чином, щоб працювати з веб-сайтами різного рівня сучасності. Така особливість

призводить до ситуації, коли сторонні сервіси для сплітування трафіку працюють менш швидко та надійно, ніж аналогічний функціонал, розроблений спеціально для конкретного веб сайту з урахуванням особливостей всіх технологій, які лежать в основі веб сайту.

Також ринкові інструменти коштують значно дорожче, ніж власноруч створений інструмент із аналогічним функціоналом. Це є наслідком об'ємів трафіку та кількості тестів, що проводяться через інструмент для сплітування трафіку.

Зважаючи на всі вище описані фактори був створений власний сплітер, який дешевше розробляти і підтримувати і який дає високу гнучкість у проведенні тестів. Головною перевагою такого рішення є повний контроль над алгоритмами сплітування та можливість перевірити кожне прийняте рішення.

У цьому інструменті для сплітування трафіку був створений для двох груп, одна з яких базова, а інша тестова. З інтерфейсом та внесеними параметрами можна ознайомитися на рис 3.1. Для того щоб тест пройшов максимально швидко результат був отриманий якнайшвидше, було вирішено 50% трафіку відправляти на базову групу і 50% трафіку на тестову. З моменту натискання на кнопку “Start Split” тест починає проводитися, кожен новий візит починає потрапляти в одну з груп.

SPLIT ID: 640 STATUS: inactive

Mirror
  Container ID
  Partner
  4 Landings

Group 2 Landing page % ADD

GROUP	LANDING PAGE	WEIGHT
0	/25/11	50%
1	/96	50%

▶ START SPLIT

Рис. 3.1. Інтерфейс інструменту для сплітування візитів із налаштуваннями А/В тесту

*Джерело: складено автором*

Після створення тесту у внутрішньому інструменті для сплітування трафіку, нові візити за посиланням, по якому проводиться тест, отримують свою групу: базову або тестову. Обрання групи відбувається випадковим чином, тобто має працювати як підкидання монети і у випадку нашого тест із розподілом трафіку 50/50. При цьому не враховуються такі фактори як стать, вік чи рівень доходу, бо не маємо такої інформації на етапі візиту на наш сайт, поки юзер сам не надасть ці дані. Також не враховуються такі фактори як девайс, браузер, операційна система, хоча їх можна отримати на етапі візиту юзера. Завдяки великій вибірці тесту ці фактори мають рівномірно розподілитися між групами, що буде пізніше перевірено.

Під час кожного з візитів юзера на досліджуваний сайт проводиться трекінг таких важливих для тесту даних як ідентифікатор кожного візиту, група яка присвоєна цьому візиту, результат візиту (закінчена реєстрація чи відсутність закінченої реєстрації) та дані про візит (маркетингові параметри з посилання, через яке прийшов юзер, юзерагент, IP адреса). Трекінг побудований

таким чином, щоб не проводити ніяких операцій із даними, що зберігаються в браузері користувача. Такий підхід не вимагає додаткових проблем із законодавством країн світу щодо збору даних. Таким чином всі ці дані не потребують згоди людини для їх збору, тому ці дані коректно збираються майже у всіх випадках.

Всі дані по трекінгу записуються у MySQL базу даних. Такий вибір зумовлений особливостями цього типу баз даних. Головною перевагою такої бази даних є можливість працювати з великими об'ємами даних. Також важливою перевагою є мова запитів SQL, яка має високу популярність, відповідно багато спеціалістів можуть ефективно працювати з такою системою. Проте, така база має проблеми зі швидкістю роботи та доступністю. Через ці особливості база погано підходить у ролі аналітичної бази даних, до якої будуть звертатися із різнотипними, не оптимізованими запитами великого об'єму [11].

Через вище описані особливості зберігання даних було обрано іншу базу даних, яка виконує роль аналітичної бази даних. Вибір впав на Google BigQuery. До головних переваг такого рішення можна віднести:

**Масштабованість:** Google BigQuery дозволяє обробляти великі обсяги даних за дуже короткий час завдяки своїм можливостям масштабування та паралельної обробки.

**Швидкість:** сервіс дозволяє робити запити до даних значно швидше, ніж аналогічні рішення, а також має багато особливостей пов'язаних із кешуванням даних, до яких частіше робляться запити.

**Вартість:** Google пропонує гнучку систему тарифів, що дозволяє користувачам платити лише за те, що вони використовують. При цьому немає оплати за завантаження та зберігання даних. Головною статтею витрат є безпосередньо оплата за об'єм даних, що використані для виконання запиту.

**Різноманітність джерел даних:** Google BigQuery підтримує інтеграцію з багатьма різноманітними джерелами даних, такими як Google Analytics, Google Cloud Storage, та багато інших.

**Простота використання:** сервіс має простий інтерфейс користувача, що дозволяє швидко і ефективно створювати запити до даних та візуалізувати результати аналізу.

**Безпека:** обране рішення є хмарним, тому фізичні ризики щодо втрати накопичувачів із даними малоймовірними. Також є гнучка система доступів до частин даних тільки конкретним групам користувачів [3].



Рис. 3.2. Схема шляху візитера на веб-сайті

*Джерело: складено автором*

Схему шляху візитера на веб-сайті можна побачити на рис. 3.2. За результатом попереднього етапу отримуємо сирі дані в аналітичній базі даних Google BigQuery. Для того, щоб оцінювати результату тесту чекаємо, поки збереться достатній розмір вибірки.

Серед даних, що є у аналітичній базі даних будуть використовуватися наступні:

- `date_visit` - дата та час візиту юзера. Формат даних: `datetime`.

- split\_id - унікальний ідентифікатор спліт тесту. Формат даних: string
- split\_group\_id - унікальний ідентифікатор спліт групи, в яку потрапив користувач під час тесту. Формат даних: string.
- platform - розділення користувачів на групи за їх платформою, а саме на мобільний девайс чи десктоп. Формат даних: string.
- device - деталізація девайсів всередині платформи користувача. Формат даних: string.
- browser - назва браузера, в якому відбувся візит. Формат даних: string.
- visits\_cnt - кількість унікальних візитів на лендінгову сторінку сайту, який хочемо оптимізувати за допомогою тесту. Формат даних: string.
- regs\_cnt - кількість унікальних реєстрацій користувачів. Формат даних: string.

Унікальність візитів для поля visit\_cnt забезпечується за допомогою куків (cookies) користувача та юзер агента (User-agent).

Куки - це невеликі файли, що зберігаються на комп'ютері або пристрої користувача через веб-браузер і використовуються для збереження даних про користувача та його взаємодію з веб-сайтом. Ці файли дозволяють веб-сайтам розпізнавати та ідентифікувати користувачів, зберігати вибрані настройки, забезпечувати персоналізований досвід використання та збирати аналітичну інформацію для поліпшення функціональності та взаємодії з веб-сайтом.

User-Agent (юзер-агент) - це інформація, що передається веб-браузером або іншим клієнтським програмним забезпеченням на веб-сервер при виконанні запиту. Він містить інформацію про програмне забезпечення та пристрій, з якого здійснюється запит, і використовується сервером для визначення, які дані надіслати відповідно до можливостей та вимог клієнта. User-Agent може включати наступну інформацію: назва браузера або програми, наприклад, "Mozilla Firefox" або "Google Chrome", версія браузера або програми, наприклад, "Version 92.0.4515.107" операційна система, на якій запущений браузер, наприклад, "Windows 10" або "Mac OS X" та іншу додаткову

інформацію, таку як тип пристрою (наприклад, смартфон, планшет або персональний комп'ютер) або мовні налаштування.

Для зручності аналізу та візуалізації результатів всі дані були агреговані та підключені у Ві інструмент Tableau.

### **3.3. Застосування частотного підходу до оцінки тестів**

Після отримання необхідної кількості користувачів у кожній з груп тесту починаємо аналізувати результати та підходити до прийняття рішення. Під час проведення тесту важливо не дивитися на те як себе показують групи. Приймати рішення до отримання необхідної кількості не можна навіть за умов отримання статистично значимого результату.

Проблема підглядання (англ. "peeking problem") виникає під час проведення А/В тестування, коли люди, які відповідають за аналіз результатів тесту, мають доступ до даних тесту в режимі реального часу і впливають на прийняття рішень на підставі цих даних до завершення експерименту. Це може спотворити результати тесту та призвести до неправильних висновків. Основна причина проблеми підглядання полягає в тому, що люди, які мають доступ до даних тесту, можуть впливати на тест, переглядаючи його результати та змінюючи або припиняючи експеримент залежно від того, що вони бачать. Наприклад, якщо вони помічають, що одна версія (А або В) показує кращі результати, вони можуть надати перевагу цій версії і зупинити тест раніше або внести зміни на льоту, не дочекавшись достатньої статистичної значимості.

Це може призвести до некоректного висновку про ефективність тестованих змін. Існує ризик, що експеримент не буде проведений на достатньо великій вибірці або протягом достатньо тривалого часу для отримання надійних результатів. Саме тому приступаємо до аналізу лише після отримання всього необхідного об'єму візитів на дві версії сайту, що приймають участь у тестуванні.

За результатами тесту було отримано 68519 унікальних візитів у базову групу (позначена як група 0) та 67839 унікальних візитів у тестову групу

(позначена як група 1). Такого об'єму візитів достатньо, щоб набрати розрахований раніше необхідний об'єм вибірки для тесту.

Для того щоб прийняти рішення на правильних даних важливо перевірити розподіл візитів між групами по доступних критеріях. Візити мають розподілятися між групами однорідно. Загалом, однорідна вибірка дозволяє отримати надійні та об'єктивні результати A/B тестів і робити обґрунтовані висновки про ефективність тестованих змін. При плануванні та проведенні A/B тестування слід приділяти достатню увагу формуванню однорідної вибірки.

Для зручного контролю рівномірності вибірки був розроблений дашборд, що візуалізує розподіл візитів між групами у всіх доступних розрізах. з його інтерфейсом можна ознайомитися на рис. 3.3.

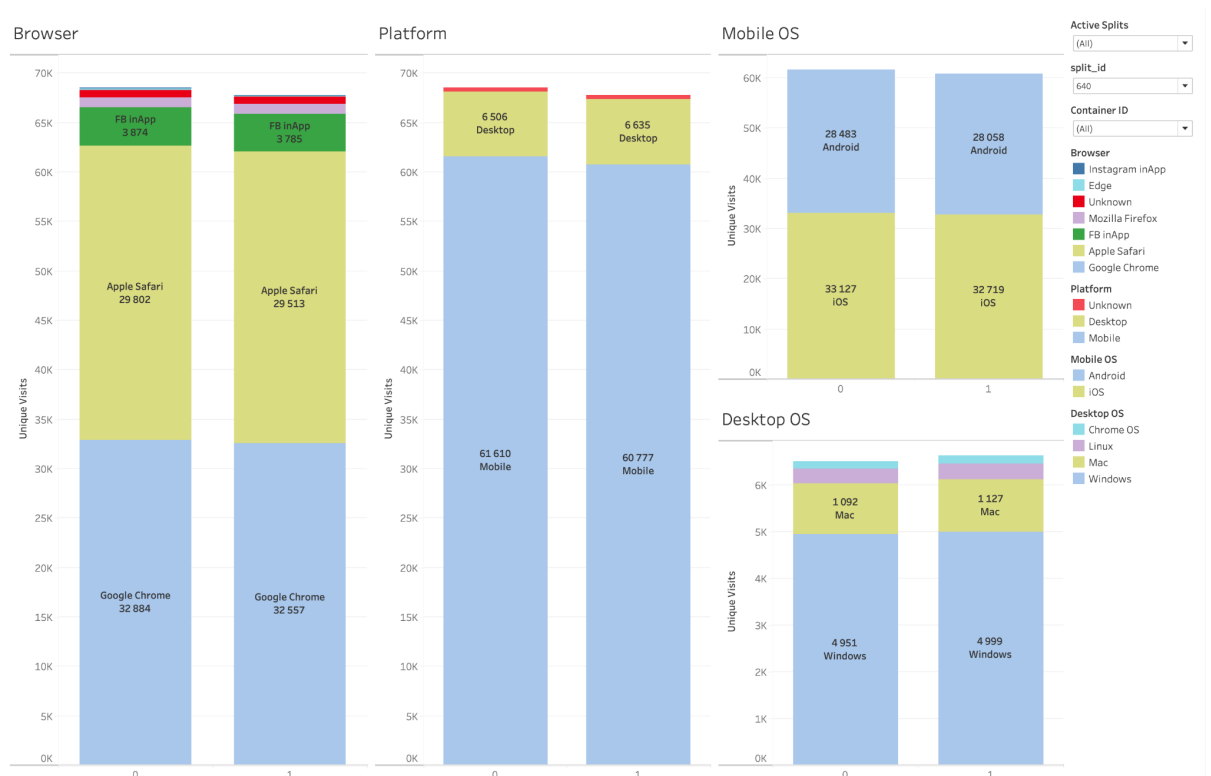


Рис. 3.3. Повний інтерфейс сторінки із аудиторними факторами для контролю рівномірності розподілення користувачів по групах тесту

*Джерело: складено автором на основі власних розрахунків*

В контексті платформи, як видно на рис. 3.4, користувачі рівномірно розподілені між групами. Біля 90% візитів відбулися з мобільних девайсів у обох групах. Більше 9% візитів відбулися із настільних комп'ютерів та

ноутбуків. Інші візити відбулися з невідомих девайсів або відбувся збій у зчитуванні девайсу користувача.

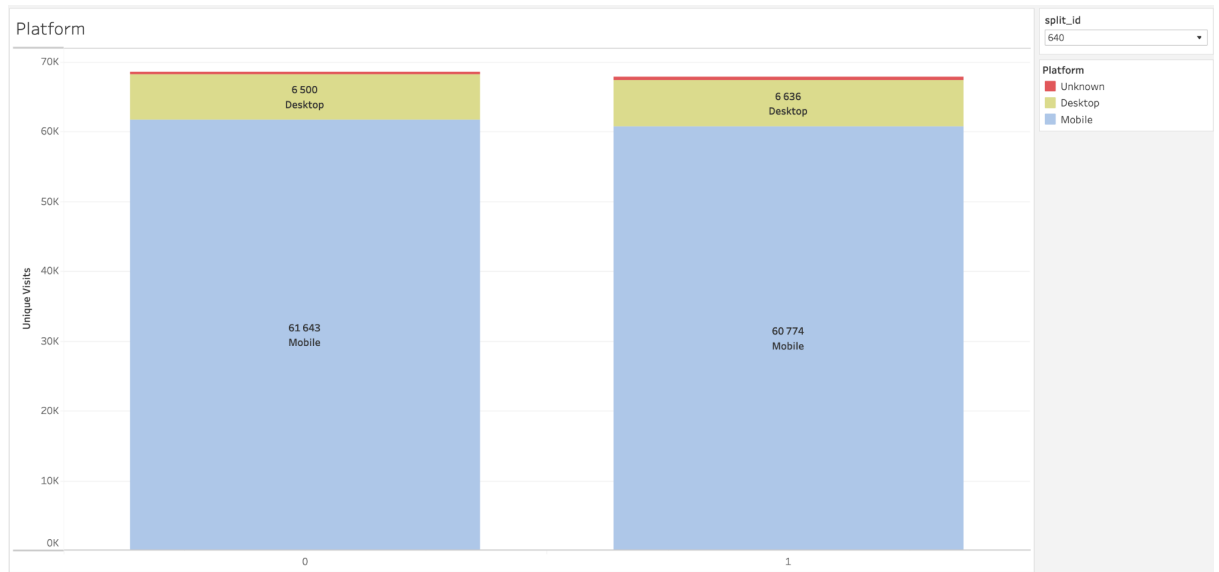


Рис. 3.4. Розподіл користувачів по групах тесту в розрізі платформи

*Джерело: складено автором на основі власних розрахунків*

Серед мобільних девайсів найбільшу частку займають IOS девайси, тобто мобільні девайси виробництва компанії Apple. У базовій групі їх 54%, а у тестовій 54%, як видно на рис. 3.5. Всі інші візити у групах відбулися із техніки з операційною системою Android. По цьому критерію вибірка також рівномірно розподіляється між групами.

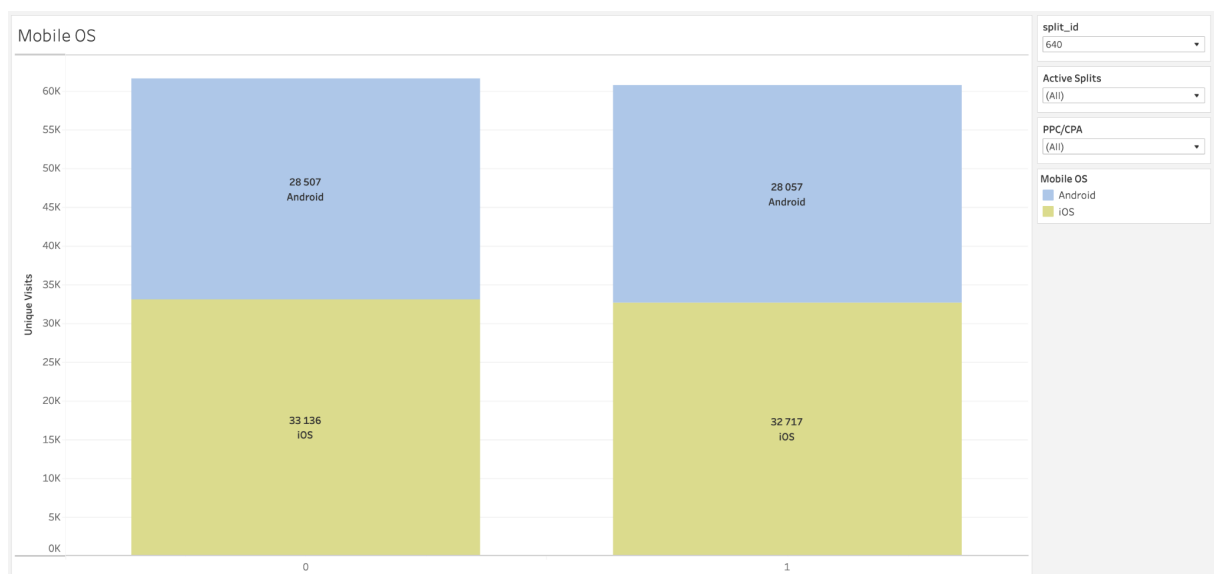


Рис. 3.5. Розподіл користувачів по групах тесту в розрізі операційної системи мобільної платформи

*Джерело: складено автором на основі власних розрахунків*

Серед комп'ютерів більшість візитів відбулася із техніки, що працює на операційній системі Windows. У базовій групі таких візитів було 76%, а у тестовій 75%. Тобто трафік із настільних комп'ютерів та ноутбуків розподіляється між групами рівномірно, що візуалізовано на рис 3.6.

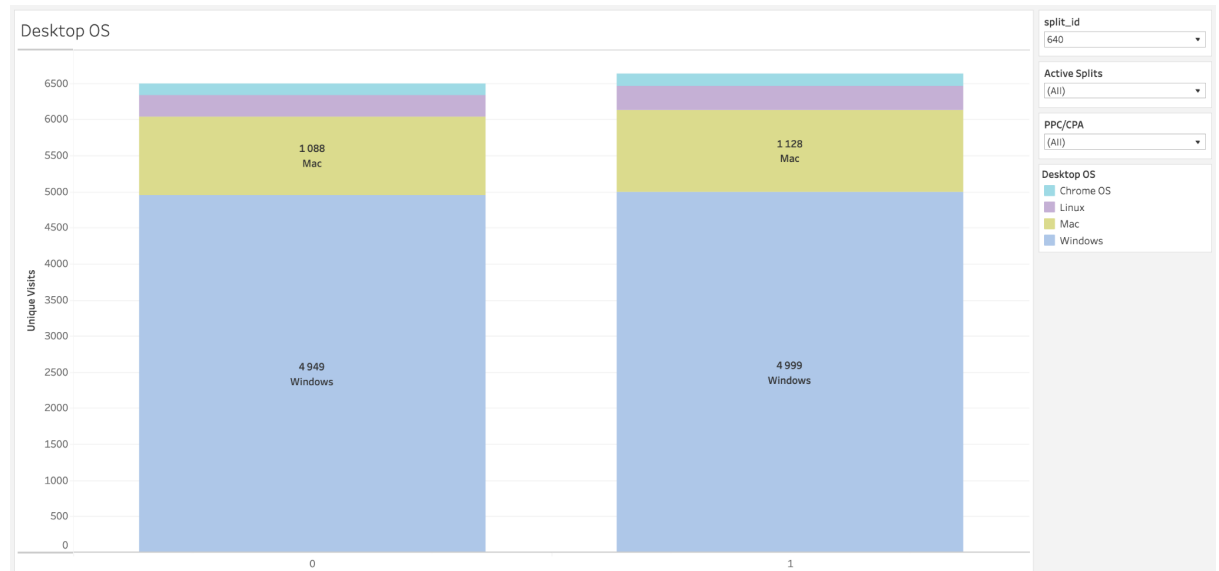


Рис. 3.6. Розподіл користувачів по групах тесту в розрізі операційної системи на настільних комп'ютерах

*Джерело: складено автором на основі власних розрахунків*

Також варто перевірити чи рівномірно користувачі розподілені між браузерами. Цей критерій може значно вплинути на перебіг тесту та на результати кожної з груп у тесті, бо різні браузери по-різному працюють із сайтами і впливають на швидкість роботи сайту.

Бачимо на рис. 3.7 що браузер Google Chrome був у 48% візитів двох груп тесту. По 43.5% візитів у кожній з груп мали браузер Apple Safari. Невеликі групи користувачів рівномірно розподілилися між групами серед тих, хто користувався веб сайтом із Facebook InApp Browser, Instagram InApp Browser, Edge та Mozilla Firefox. Бачимо, що по критерію браузера люди розділилися між групами рівномірно.

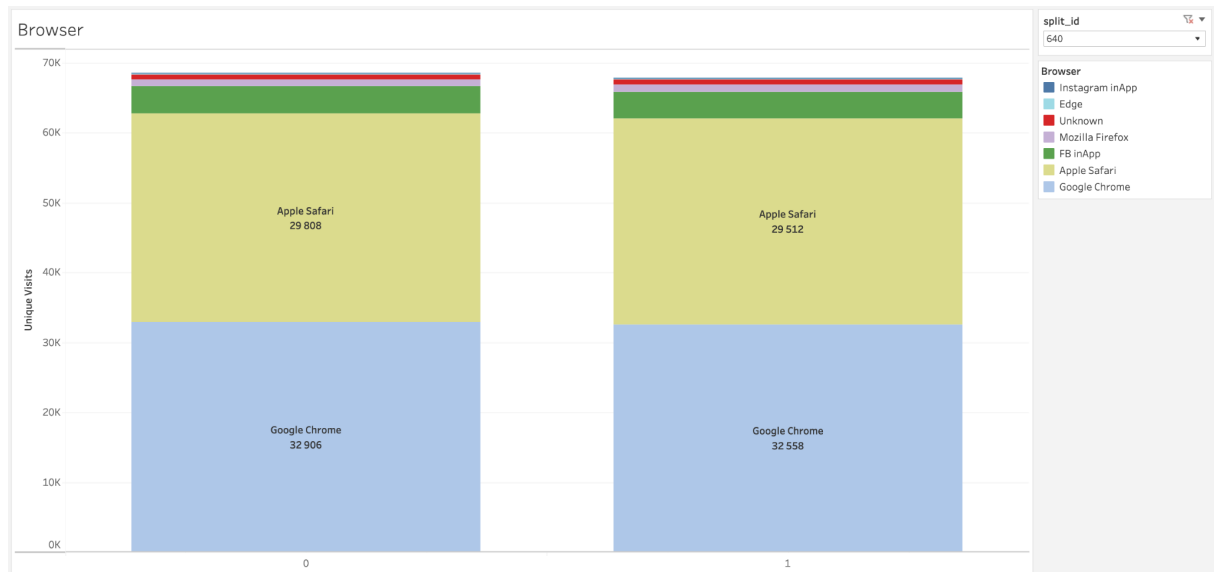


Рис. 3.7. Розподіл користувачів по групах тесту в розрізі браузерів

*Джерело: складено автором на основі власних розрахунків*

Для візуалізації результатів усіх тестів була розроблена сторінка із переліком тестів. З інтерфейсом сторінки можна ознайомитися на рис. 3.8. У таблиці маємо всі необхідні для прийняття рішення поля, а саме:

- `split_id` - унікальний ідентифікатор спліт тесту.
- `is_active` - поле, що вказує на те активний тест чи трафік перестав потрапляти у групи тесту і вибірка не буде збільшуватися.
- `container_id` - унікальний ідентифікатор деякої частини трафіку, що потрапляє на наш веб сайт і бере участь у тесті (поле пuste через фільтр `Container Level = False`).
- `Sample Size` - кількість візитів, що потрібна для того, щоб прийняти рішення про успішність тесту при очікуваному ефекті зміни розміром 10%.
- `split_group` - номер спліт групи, де 0 це базова, а від 1 до нескінченності це тестові групи.
- `Result` - рішення про успішність тесту. Може приймати значення `Not enough visits`, коли розмір вибірки недостатній, `No difference`, коли зібрана достатня вибірка, але різниця між конверсіями нижча за очікуваний ефект, `Won`, коли зібрана достатня вибірка і тестова група показала статистично значиму

зміну не меншого розміру, ніж очікуваний ефект, Lost, коли зібрана достатня вибірка, але тестова група показала статистично значиму негативну зміну.

- visits - кількість унікальних візитів на кожну із груп тесту.
- regs - кількість реєстрацій у кожній з груп. Реєстрації є конверсіями у нашому тесті.
- CR\_reg - показник конверсії із візиту на сайт у реєстрацію
- CR\_diff - приріст конверсії у тестовій групі в порівнянні з базовою групою тесту.
- p-value - ймовірність отримання таких або більш екстремальних результатів, якщо нульова гіпотеза (припущення про відсутність ефекту або різниці) є правильною.
- power - ймовірність виявлення статистично значущих різниць або ефектів, якщо вони насправді існують.

split...	is_active...	Container ID	Sample Size	split_group...	Result	visits	regs	CR_reg	CR_diff	p-value	Power
640	False	54 099	0	1	No difference	68 519	1 957	2,86%			
						67 839	1 905	2,81%	-1,7%	0,5961	11,51%

date\_visit: 01.02.2023 - 31.12.2023

mirror\_name: (All)

partner: (All)

split\_id: 640

Container ID: (All)

Minimum Detectable Effect: 10%

Active Splits: (All)

PPC/CPA: (All)

Container Level: False

?

Рис. 3.8. Інтерфейс дашборду для оцінки результатів тестів

*Джерело: складено автором на основі власних розрахунків*

Для зручності користування та пришвидшення пошуку потрібних тестів у правій частині дашборда є блок фільтрів, до яких належать:

- date\_visit - фільтр меж по часу.
- mirror\_name - фільтр по назві дзеркала нашого веб сайту.
- partner - фільтр по маркетинговому параметру, яким позначена частина трафіку.
- split\_id - фільтр по ідентифікатору тесту.
- Container ID - фільтр по частинах трафіку, які об'єднують однорідні джерела трафіку.
- Minimum Detectable Effect - параметр, що закладає у розрахунки Sample Size очікування менеджера по приросту рівня конверсії.
- Active Splits - фільтр, що дозволяє показувати активні тести, неактивні тести та всі тести.
- PPC/CPA - фільтр, що дозволяє відображати тести, що проводяться на PPC трафіку або на CPA трафіку.
- Container Level - параметр, що додає у візуалізацію деталізацію по container\_id або прибирає цю деталізацію.

Таблична частина може містити багато тестів одночасно, тому менеджер може приймати рішення по групі тестів без переключення фільтрів та переходів на інші дашборди.

Аналізуючи конкретно цей тест із ідентифікатором 640, було вже перевірено однорідність розподілу вибірки між групами. Умови тесту виявились достатніми, щоб оцінювати його результати за статистичними показниками частотного підходу та бути впевненими з високою ймовірністю про те, що прийняті рішення будуть правильними.

За результатами тесту отримали  $p\text{-value} = 0,5961$ . Отримане значення показника показує ймовірність отримати таке або сильніше відхилення рівня конверсії за умови, що різниця рівнів конверсії між групами насправді відсутня у генеральній сукупності. Тобто прийнявши рішення про те, що різниця існує, ми з високою ймовірністю приймемо хибне рішення. Даний показник мав бути менше 0,05 для того, щоб ми були достатньо впевнені при прийнятті рішень.

Power = 11,51% говорить про низьку ймовірність виявити відхилення за умови, що це відхилення насправді існує. Цільовий показник становить 80% і більше. Тільки при рівні power  $\geq 80\%$  будемо достатньо впевнені, щоб приймати рішення про закриття тесту на користь тестової групи.

Так як статистичні показники p-value та power говорять про те, що приріст у тестовій групі на рівні 10% і більше мало ймовірний, то приймаємо нульову гіпотезу про те, що тестова група не відрізняється від базової. Відповідно до результатів у візуалізації відображається рішення “No difference”, яке показує прийняте рішення та сигналізує про необхідність завершення тесту.

Було отримано результати з точки зору статистики, які потрібно інтерпретувати в контексті висунутої гіпотези про оптимізацію рівня конверсії веб-сайту. Гіпотеза була сформульована наступним чином: можливість поміняти розмір інтерфейсу на веб-сайті може допомогти зареєструвати більше цільових користувачів, що з більшою ймовірністю, ніж інші готові платити за сервіс.

Враховуючи результати тесту зі сторони статистики можна зробити висновок, що покращення зручності користуванням шрифтом на веб-сайті не дало значного приросту в конверсії. Не було отримано значно більше цільових користувачів протягом тесту, тому не будемо застосовувати протестований функціонал до всього веб-сайту. Також опосередковано можна зробити висновок про те, що переважна більшість користувачів має достатньо зручний розмір інтерфейсу на базовій версії веб-сайту та не потребує змін у цій сфері. Ці висновки можна буде використати при висуванні майбутніх гіпотез, зменшивши кількість невдалих тестів, пов'язаних зі шрифтом на веб-сайті у майбутньому.

## ВИСНОВКИ

Проведене дослідження досягли своєї мети та показали як можна впливати на успішність компанії статистичними методами. Було визначено, що маркетинг є комплексною системою, яка організує виробництво і збут продукції з метою задоволення потреб споживачів і отримання прибутку. Він включає аналіз ринку, розробку стратегій і спрямований на залучення нових клієнтів, утримання існуючих, розширення ринку і прийняття обґрунтованих рішень. Серед популярних підходів сучасного маркетингу виділяються PPC, CPA і органічне просування, які дозволяють ефективно використовувати рекламний бюджет і залучати цільову аудиторію. Однак, маркетинг виходить за рамки лише залучення нових клієнтів і концентрується на зміцненні довгострокових відносин з користувачами, використовуючи стратегії збереження, розширення ринку і міжнародну експансію, а також data driven підхід. Маркетинг є важливим фактором успіху бізнесу, який допомагає будувати сильний бренд і досягати конкурентної переваги на ринку.

Серед частин маркетингу було обрано CRO як пріоритетну частину для оптимізації. Було визначено, що для оптимізації конверсії веб-сайту зручно використовувати статистичні методи. Основним методом було визначено A/B тестування. Такий вибір зумовлений тим, що статистичні методи забезпечують об'єктивні дані та стабільність внесення змін. Частотний підхід до A/B тестування рекомендується як більш надійний та простий у реалізації вибір для проведення контрольованих тестів. Використання статистичних методів вимагає знань про предметну область для коректного формулювання гіпотез та обчислення статистичних показників, таких як p-value і power.

Була розроблена система підтримки прийняття рішень, в основі якої лежить частотний підхід до A/B тестування. Базою для розробки такого рішення стали база даних Google Big Query та BI інструмент Tableau. Результатом став дашборд у Tableau, який виконує роль автоматизованої системи прийняття рішень. Він показує розподіл користувачів між групами, розмір необхідної

вибірки на кожен з груп, статистичні показники та прийняте рішення про результати тесту.

Вибір трафіку з Google Display Network для тесту зумовлений його перевагами над іншими джерелами трафіку в контексті впливу на загальну успішність маркетингу. Використання внутрішнього інструменту для сплітування трафіку дозволяє контролювати процес та перевірити прийняті рішення.

Після отримання необхідної вибірки на базову та тестову групи був проведений аналіз тесту. Розподіл візитів між групами був однорідним за різними критеріями (операційна система, платформа, девайс), тому результати тесту можна приймати з високою довірою. За такими статистичними показниками як  $p$ -value та power було прийнято нульову гіпотезу про відсутність статистично значущої різниці між групами тесту. Роботу системи було продемонстровано на одному прикладі перевірки гіпотези. Хоча продемонстрований тест не допоміг підвищити рівень конверсії веб-сайту, проте розроблена система продемонструвала свою працездатність. Варто продовжувати тестувати гіпотези про оптимізацію рівня конверсії веб сайту.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Arya V., Sethi D., Paul J. Does digital footprint act as a digital asset? – Enhancing brand experience through remarketing. *International Journal of Information Management*. 2019. Vol. 49. P. 142–156. URL: <https://doi.org/10.1016/j.ijinfomgt.2019.03.013> (date of access: 10.06.2023).
2. Chen M., Mao S., Liu Y. Big Data: A Survey. *Mobile Networks and Applications*. 2014. Vol. 19, no. 2. P. 171–209. URL: <https://doi.org/10.1007/s11036-013-0489-0> (date of access: 10.06.2023).
3. Cloud data warehouse to power your data-driven [Електронний ресурс] / 2023. – Режим доступу до ресурсу: <https://cloud.google.com/bigquery>.
4. Controlled experiments on the web: survey and practical guide / R. Kohavi et al. *Data Mining and Knowledge Discovery*. 2008. Vol. 18, no. 1. P. 140–181. URL: <https://doi.org/10.1007/s10618-008-0114-1> (date of access: 10.06.2023).
5. Frequentist and Bayesian Hypothesis Testing: An Intuitive Guide for Urologists and Clinicians / J. Gaona et al. *Revista Urología Colombiana / Colombian Urology Journal*. 2022. Vol. 31, no. 03. P. e130-e140. URL: <https://doi.org/10.1055/s-0042-1756171> (date of access: 10.06.2023).
6. Kruschke J. K., Liddell T. M. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*. 2017. Vol. 25, no. 1. P. 178–206. URL: <https://doi.org/10.3758/s13423-016-1221-4> (date of access: 10.06.2023).
7. Kumar V., Mittal S. Mobile marketing campaigns: practices, challenges and opportunities. *International Journal of Business Innovation and Research*. 2020. Vol. 21, no. 4. P. 523. URL: <https://doi.org/10.1504/ijbir.2020.105996> (date of access: 10.06.2023).
8. Setting the future of digital and social media marketing research: Perspectives and research propositions / Y. K. Dwivedi et al. *International Journal of Information Management*. 2020. P. 102168. URL: <https://doi.org/10.1016/j.ijinfomgt.2020.102168> (date of access: 10.06.2023).

9. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations / S. Greenland et al. *European Journal of Epidemiology*. 2016. Vol. 31, no. 4. P. 337–350. URL: <https://doi.org/10.1007/s10654-016-0149-3> (date of access: 10.06.2023).
10. Szymanski G., Lininski P. Model of the Effectiveness of Google Adwords Advertising Activities. 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, 11–14 September 2018. 2018. URL: <https://doi.org/10.1109/stc-csit.2018.8526633> (date of access: 10.06.2023).
11. The rise of “big data” on cloud computing: Review and open research issues / I. A. T. Hashem et al. *Information Systems*. 2015. Vol. 47. P. 98–115.
12. Top Website Statistics For 2023 [Електронний ресурс] / 2023. – Режим доступу до ресурсу: <https://www.forbes.com/advisor/business/software/website-statistics/>.
13. Tucker C. E. Social Networks, Personalized Advertising, and Privacy Controls. *Journal of Marketing Research*. 2013. Vol. 51, no. 5. P. 546–562. URL: [<https://sci-hub.se/10.1509/jmr.10.0355>]
14. Unified frequentist and Bayesian testing of a precise hypothesis Comment by Dennis V. Lindley Comment by Thomas A. Louis Comment by David Hinkley Rejoinder by J. O. Berger, B. Boukai and Y. Wang / J. O. Berger et al. *Statistical Science*. 1997. Vol. 12, no. 3. P. 133–160. URL: <https://doi.org/10.1214/ss/1030037904> (date of access: 10.06.2023).
15. What Do We Talk About When We Talk About Dashboards? / A. Sarikaya et al. *IEEE Transactions on Visualization and Computer Graphics*. 2019. Vol. 25, no. 1. P. 682–692. URL: <https://doi.org/10.1109/tvcg.2018.2864903> (date of access: 10.06.2023).
16. Wilke C. O. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media, 2019. 390 p.

17. Yalçın N., Köse U. What is search engine optimization: SEO?. *Procedia - Social and Behavioral Sciences*. 2010. Vol. 9. P. 487–493. URL: <https://doi.org/10.1016/j.sbspro.2010.12.185> (date of access: 10.06.2023).
18. Zaslavsky B. G. Bayesian Versus Frequentist Hypotheses Testing in Clinical Trials with Dichotomous and Countable Outcomes. *Journal of Biopharmaceutical Statistics*. 2010. Vol. 20, no. 5. P. 985–997. URL: <https://doi.org/10.1080/10543401003619023> (date of access: 10.06.2023).
19. Амеліна І.В, Лукашенко К.С. РОЛЬ КОМПЛЕКСУ МАРКЕТИНГУ В УПРАВЛІННІ ДІЯЛЬНІСТЮ ПІДПРИЄМСТВА В УМОВАХ РИНКОВОЇ КОНКУРЕНЦІЇ. *Економіка та управління підприємствами*. 2022.
20. Мосійчук І. В. ОСОБЛИВОСТІ УПРАВЛІННЯ МАРКЕТИНГОВОЮ ДІЯЛЬНІСТЮ ПІДПРИЄМСТВ В УКРАЇНІ. 2017.
21. Черняк О. І., Обушна О.М., Ставицький А.В. Теорія ймовірностей та математична статистика. Збірник задач: навчальний посібник. 2-ге видання, виправлене. К.: Т-во «Знання», КОО, 2002. 199 с.
22. Шапочка М., Макарюк О. Застосування експертних оцінок при прийнятті рішень за умов невизначеності. *Механізм регулювання економіки*. 2006. № 4.

## ДОДАТКИ

## Додаток А

```

1 with partners as (
2   select distinct id_partner
3   from product.users
4 ),
5 visit_partners as (
6   select id_visit, (select value from lv.query_params where key = 'partner') partner
7   from marketing.landing_visit lv
8   where date(date_created) >= '2022-12-09'
9
10  -- union all
11
12  -- select id, any_value(value) partner
13  -- from marketing.landing_visitors lv
14  -- left join unnest(params) on key = 'partner'
15  -- where date(date_created) < '2022-12-09'
16  -- group by 1
17 ),
18
19 splitter_data as (
20  select distinct
21    cast(lsd.id_split as string) as id_split,
22    cast(lsd.id_split_container as string) as id_split_container,
23    cast(sg.id as string) as id_split_group,
24    sg.number as split_group_number,
25    sg.landing,
26    s.date_start,
27    s.date_end,
28    s.is_active,
29    comment
30  from (select distinct id_split_container, id_split, id_split_group from
31 `marketing.landing_splitter_decision`) lsd
32 left join marketing.splits s on s.id = lsd.id_split
33 left join marketing.splitter_containers sc on sc.id = lsd.id_split_container
34 left join marketing.split_groups sg on sg.id = lsd.id_split_group
35 where true
36   and concat(sc.id, sc.date_updated) in (
37     select concat(id,
38       max(date_updated) over (partition by id) last_update
39     from marketing.splitter_containers
40     order by 1 desc
41   )
42   and concat(s.id, s.date_updated) in (
43     select concat(id,
44       max(date_updated) over (partition by id) last_update
45     from marketing.splits
46     order by 1 desc
47   )
48 order by cast(id_split as int64) desc
49 ),
50
51 events_product_pageview AS(
52  SELECT e.userId id_user
53    , MIN(e.date_created) date_seen_nf
54  FROM product.events e
55  JOIN product.users u ON u.id = e.userId
56  WHERE true
57  AND date(e.date_created) > date_sub(CURRENT_DATE(), interval 33 day) --дані лише за
58 останні 33 днів

```

```

59     AND e.event.subtype = 'page_view'
60     AND e.source.pageType != 'wizard_page'
61     AND e.source.pageType is not null
62     AND u.date_created >= '2023-01-01'
63 GROUP BY 1
64 ),
65
66 email_confirm AS (
67     SELECT id_user
68         , MIN(um.date_created) date_confirmed
69     FROM retention.user_meta um
70         JOIN product.users u ON u.id = um.id_user
71     WHERE true
72         AND date(um.date_created) > date_sub(CURRENT_DATE(), interval 33 day) --дані лише за
73 останні 33 днів
74         AND id_key = 13
75         AND u.date_created >= '2023-01-01'
76     GROUP BY 1
77 ),
78
79 spent_15cr AS (
80     SELECT DISTINCT cs.id_regular_user
81         , FIRST_VALUE(cs.date_created) OVER (PARTITION BY cs.id_regular_user ORDER BY
82 cs.date_created) spent_15_credits
83     FROM product.credits_spends cs
84         JOIN product.users u ON u.id = cs.id_regular_user
85     WHERE cs.is_regular_user_spent
86         AND cs.date_created >= '2023-01-01'
87         AND u.date_created >= '2023-01-01'
88     QUALIFY SUM(spent) OVER (PARTITION BY cs.id_regular_user ORDER BY cs.date_created) >= 15
89 )
90
91 , for_audience_check as (
92     select distinct id_visit,
93         user_agent
94         , split_group_id
95         , CASE
96             WHEN lower(user_agent) LIKE '%iphone%' THEN 'iPhone'
97             WHEN lower(user_agent) LIKE '%macintosh%' THEN 'Mac'
98             WHEN lower(user_agent) LIKE '%ipad%' THEN 'iPad'
99             WHEN lower(user_agent) LIKE '%android%' THEN 'Android'
100            WHEN lower(user_agent) LIKE '%win%' THEN 'Windows'
101            WHEN lower(user_agent) LIKE '%linux%' THEN 'Linux'
102            WHEN lower(user_agent) LIKE '%cros%' THEN 'Chrome OS'
103            ELSE 'Unknown'
104        END AS Device
105
106     , CASE
107         WHEN user_agent LIKE '%edge%' THEN 'Edge'
108         WHEN user_agent LIKE '%MSIE%' THEN 'Internet Explorer'
109         WHEN user_agent LIKE '%Firefox%' THEN 'Mozilla Firefox'
110         WHEN user_agent LIKE '%Chrome%' THEN 'Google Chrome'
111         WHEN user_agent LIKE '%Safari%' THEN 'Apple Safari'
112         WHEN user_agent LIKE '%Opera%' THEN 'Opera'
113         WHEN user_agent LIKE '%Outlook%' THEN 'Outlook'
114         WHEN user_agent LIKE '%FBAN%' THEN 'FB inApp'
115         WHEN user_agent LIKE '%Instagram%' THEN 'Instagram inApp'
116         ELSE 'Unknown'
117     END AS Browser
118     from marketing.visit_view vv
119     left join marketing.landing_visit lv using(id_visit)
120     where date_visit > '2023-01-01'
121 )

```

```

122
123 , audience_factors as (
124 select *
125 , case
126     when (Device like 'iPhone' or Device like 'Android') then 'Mobile'
127     when (Device like 'Mac' or Device like 'Windows' or Device like 'Linux' or Device like
128 'Chrome OS') then 'Desktop'
129     else 'Unknown'
130 end as Platform
131 from for_audience_check
132 )
133
134 , costs_for_visits as (
135
136 select
137     id_visit,
138     v.id_external,
139     cost / count(distinct id_visit) over (partition by v.id_external) as
140 visit_acquisition_cost
141 from (select distinct id_visit, id_external from marketing.visit_view where date_visit >=
142 '2023-01-01') v
143 left join (select id_external, sum(cost) as cost from marketing.ssot_regs where date_action
144 >= '2023-01-01' group by 1) s on v.id_external = s.id_external
145
146 )
147
148 select
149     date(v.date_visit) date_visit,
150     v.container_id,
151     sd.id_split as split_id,
152     sd.split_group_number split_group_id,
153     sd.date_start,
154     sd.date_end,
155     sd.is_active,
156     v.landing,
157     af.Device,
158     af.Browser,
159     af.Platform,
160     if(lv.partner in (select * from partners), lv.partner, null) partner,
161     v.mirror_name,
162     if(id_split is not null
163         , if(concat(sd.landing, '/') = v.landing, 1, 0)
164         , 1
165     ) as not_fantoms,
166     count(distinct v.id_visit) visits_cnt,
167     count(distinct v.id_user) regs_cnt,
168     sum(visit_acquisition_cost) as costs,
169     sum(rcd_cnt) as rcd_cnt,
170     count( distinct if (s.is_eu is true, s.id_user, null)) eu_cnt,
171     count( distinct if (s.revenue_7d > 0, s.id_user, null)) cl_cnt_7d,
172     count( distinct if (ec.date_confirmed is not null, ec.id_user, null)) confirmed_email,
173     count( distinct if (nf.date_seen_nf is not null, nf.id_user, null)) seen_mainpage,
174     count( distinct if (sp.spent_15_credits is not null, sp.id_regular_user, null)) crl5_spent,
175
176 from marketing.visit_view v
177 left join visit_partners lv on v.id_visit = lv.id_visit
178 left join splitter_data sd on v.container_id = cast(sd.id_split_container as string)
179     and v.split_id = cast (sd.id_split as string)
180     and v.split_group_id = cast(sd.id_split_group as string)
181 left join marketing.ssot s on s.id_action = v.id_user
182 join costs_for_visits c on c.id_visit = v.id_visit
183 left join audience_factors af on af.id_visit = v.id_visit
184 left join email_confirm ec on ec.id_user = v.id_user

```

```
185 left join events_product_pageview nf on nf.id_user = v.id_user
186 left join spent_15cr sp on sp.id_regular_user = v.id_user

where date(v.date_visit) >= '2022-12-09'
and path = 'splitter'
and concat(sd.landing, '/') = v.landing
and v.id_visit not in (select distinct id_visit from marketing.landing_visit_map where type
like "autologin") --тут виключаються автологіни
group by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14
having not_fantoms = 1
order by 1, 2, 3, 4
```