

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВАСИЛЯ СТУСА

МАРТИНЕНКО ОЛЕКСАНДР ОЛЕКСІЙОВИЧ

Допускається до захисту:

в.о. завідувача кафедри
інформаційних технологій
канд. техн. наук, доцент

_____ О.В. Зелінська

«_____» _____ 2024 р.

**ОЦІНКА ЕФЕКТИВНОСТІ РЕКЛАМИ НА ІНТЕРНЕТ-РЕСУРСАХ НА
ОСНОВІ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ**

Спеціальність 122 «Комп'ютерні науки»

Кваліфікаційна (магістерська) робота

Науковий керівник:

Зелінська О.В., доцент кафедри
інформаційних технологій,
канд. техн. наук, доцент

(підпис)

Оцінка: _____ / _____ / _____
(бали за шкалою ECTS /за національною
шкалою)

Голова ЕК: _____
(підпис)

Вінниця – 2024

АНОТАЦІЯ

Мартиненко О.О. Оцінка ефективності реклами на інтернет-ресурсах на основі методів інтелектуального аналізу. Спеціальність 122 «Комп'ютерні науки», Освітня програма «Комп'ютерні технології обробки даних (Data Science)». Донецький національний університет імені Василя Стуса, Вінниця, 2024.

У кваліфікаційній (магістерській) роботі досліджено аналіз ефективності реклами на інтернет-ресурсах із застосуванням методів інтелектуального аналізу. Розраховано ключові показники ефективності рекламної компанії, когортний аналіз для оцінки утримання користувачів, а також прогноз денної і тижневої активності. Застосовано логістичну регресію для моделювання ймовірності утримання користувачів.

86 с., 33 рис., 6 табл., 47 джерел.

Ключові слова: інтелектуальний аналіз даних; оцінка ефективності реклами; логістична регресія; когортний аналіз.

ABSTRACT

Martynenko O.O. Evaluation of the effectiveness of advertising on Internet resources based on methods of intellectual analysis. Specialty 122 "Computer science", Educational program "Computer technologies of data processing (Data Science)". Vasyl Stus Donetsk National University, Vinnytsia, 2024.

In the qualification (master's) thesis, it was investigated the analysis of the effectiveness of advertising on Internet resources using the methods of intellectual analysis. The object of the study is the system of evaluating the effectiveness of advertising on Internet resources. The subject of research is models, methods and technologies for evaluating the effectiveness of advertising. Key performance indicators of the advertising company, cohort analysis to assess user retention, as well as forecast of daily and weekly activity were calculated. Logistic regression was applied to model the probability of user retention. The obtained results make it possible to automate data analysis in order to optimize the budget, personalize advertising and increase user engagement in Internet resources.

86 p., 33 figures., 6 tab., 47 ref.

Keywords: intelligent data analysis; assessment of advertising effectiveness; logistic regression; cohort analysis.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ	
ВСТУП	5
РОЗДІЛ 1 ОСНОВНІ ПОНЯТТЯ ТА ПІДХОДИ ДО ОЦІНКИ ЕФЕКТИВНОСТІ РЕКЛАМИ НА ІНТЕРНЕТ-РЕСУРСАХ	7
1.1. Основні поняття і тенденції рекламної діяльності на інтернет-ресурсах	7
1.2. Сучасні підходи до оцінки ефективності інтернет-реклами	15
1.3. Застосування інтелектуального аналізу та штучного інтелекту в маркетинговій діяльності та постановка завдання	21
1.4. Висновки за розділом 1	22
РОЗДІЛ 2 МЕТОДИ ТА ТЕХНОЛОГІЇ ЗБОРУ ДАНИХ І ОЦІНКИ ЕФЕКТИВНОСТІ РЕКЛАМИ	28
2.1. Методи збору даних для аналізу ефективності реклами	28
2.2. Інтелектуальні методи аналізу і оцінки ефективності реклами	30
2.3. Технології оцінки ефективності реклами	35
2.4. Висновки за розділом 2	45
РОЗДІЛ 3 ПРОГРАМНА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ АНАЛІЗУ ЕФЕКТИВНОСТІ РЕКЛАМИ	46
3.1 Збір і попередня обробка даних рекламних кампаній (на прикладі комп'ютерних ігор)	46
3.2. Програмна реалізація інтелектуальної системи аналізу ефективності реклами	60
3.3. Висновки за розділом 3	70
ВИСНОВКИ	71
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ	72
ДОДАТКИ	78

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ

CR - Conversion Rate

CPC - Cost Per Click

CPL - Cost Per Lead

CPO - Cost Per Order

CPM - Cost Per Mille

CPA - Cost Per Action

CAC - Customer Acquisition Cost

ІІІ – штучний інтелект

ВСТУП

Актуальність. Високий рівень конкуренції та розширення можливостей рекламної діяльності на інтернет-ресурсах та медіа, що в свою чергу призводить для підвищення цього рівня спонукає рекламодавців до збільшення та/або їх оптимізації з метою підвищення ефективності.

Постановка проблеми в загальному вигляді. Ефективне використання інтернет-маркетингу сприяє: збільшенню обсягів продажів, покращенню сприйняття бренду, залученню нових клієнтів, збільшенню рентабельності підприємств. Результативність реклами – її основна характеристика.

Таким чином існує проблема оцінки ефективності промоакцій на етапі запуску та після, та її оптимізації для поліпшення ефективності. Ефективність залежить від багатьох факторів: сегмент, ринку, аудиторії, характеристик розміщення реклами (вид, тип розміщення, вид інтернет-ресурсу та інше).

Об'єкт дослідження – система оцінки ефективності реклами на інтернет-ресурсах.

Предмет дослідження – науково-теоретичні основи, методичні підходи та практичні аспекти інформаційної технології оцінки ефективності реклами на інтернет-ресурсах.

Мета дослідження – теоретичне обґрунтування та розробка практичних рекомендацій застосування методів інтелектуального аналізу для оцінки ефективності реклами на інтернет-ресурсах.

Задачі дослідження:

- визначити сучасний стан технологій і тенденції рекламної діяльності на інтернет-ресурсах;
- дослідити інтелектуальні технології аналізу ефективності реклами;
- виконати постановку завдання, збір і попередню обробку даних рекламних кампаній, очищення та організацію даних;
- виконати аналіз і прогноз динаміки активності користувачів на основі інтелектуальних методів і моделей;
- розробити модель оцінки ймовірності утримання користувачів в

залежності від їх характеристик;

- виконати програмну реалізацію інтелектуальної системи аналізу ефективності реклами.



РОЗДІЛ 1

ОСНОВНІ ПОНЯТТЯ ТА ПІДХОДИ ДО ОЦІНКИ ЕФЕКТИВНОСТІ РЕКЛАМИ НА ІНТЕРНЕТ-РЕСУРСАХ

1.1 Основні поняття і тенденції рекламної діяльності на інтернет-ресурсах

Інтернет-реклама --- реклама, що розміщується в інтернеті, переважно популярних веб-сайтах, що добре себе зарекомендували (наприклад, Google.com — найвідвідуваніший сайт у світі (поряд із facebook.com) розроблений компанією Google, яка займається розробкою, розвитком і дизайном найпопулярнішого в Інтернеті пошукового сервісу. Google підтримує та розробляє низку інтернет-сервісів і продукції, отримуючи дохід передусім від реклами, завдяки своїй програмі Ads.); представлення товарів, послуг або підприємства на інтернет-ресурсах. Інтернет-реклама адресована масовому клієнту і має характер переконання [1].

Маркетингове дослідження --- це функція, яка зв'язує споживача, замовника та громадськість із маркетологом через інформацію, яка використовується для:

- ідентифікації та визначення можливостей і проблем;
- генерації, уточнення та оцінки дій;
- контролю продуктивності;
- покращення розуміння процесу дії інтернет-реклами.

Маркетолог визначає інформацію, необхідну для вирішення цих проблем, розробляє метод збору інформації, керує та реалізує процес збору даних, аналізує результати та повідомляє результати та їхні наслідки [2].

Вид, можливості поширення і вартість реклами залежать від типу інтернет-ресурсу, що в кінцевому випадку впливає на її ефективність. Розглянемо класифікацію за типами інтернет-ресурсів, які в більшості є платформами для розміщення реклами [3].

За доступністю сервісів:

- відкриті (всі сервіси доступні для будь-яких відвідувачів і користувачів);

– напіввідкриті (для доступу необхідно зареєструватися (зазвичай безкоштовно));

– закриті (повністю закриті службові сайти організацій (у тому числі корпоративні), особисті сайти приватних осіб. Такі сайти доступні для вузького кола користувачів. Доступ новим користувачам зазвичай надається через запрошення).

З фізичним розташуванням:

– загальнодоступні сайти (мережі Інтернет);

– локальні сайти (доступні тільки в межах локальної мережі).

За технологічними особливостями:

– статичні (складаються з статистичних html (htm, dhtml) сторінок, що становлять єдине ціле. Користувачеві видаються в тому вигляді в якому вони зберігаються на сервері);

– динамічні (складаються з динамічних html (htm, dhtml) сторінок-шаблонів, інформації, скриптів у вигляді окремих файлів. Вміст генерується за запитом спеціальними скриптами (програмами) на основі інших даних з будь-якого джерела);

– сайти створені із застосуванням Flash-технологій, коли весь сайт розташовується на одній веб-сторінці, призначеної виключно для завантаження Flash-файлу, а вся навігація і контент реалізовані в самому Flash-ролику.

За типами макетів: фіксованої ширини, гумовий макет, динамічно-еластичний.

За схемою подання інформації і категорії вирішуваних задач:

– Інтернет-портал - багатокомпонентна розгалужена структура, сформована з функціонально самодостатніх сайтів самостійних організацій або підрозділів корпоративних структури;

– Інформаційні ресурси:

• тематичний сайт (надає специфічну вузькотематичну інформацію з будь-якої теми);

• тематичний портал (великий веб-ресурс, який надає інформацію з

певної тематики, але додатково містять засоби взаємодії з користувачами і дозволяють користувачам спілкуватися в рамках порталу (форуми, чати));

- Інтернет-представництва власників бізнесу:

- Сайт-візитка (містить загальну інформацію про власника сайту);
- Представницький сайт (сайт-візитка з розширеною функціональністю);
- Корпоративний сайт (містить повну інформацію компанії-власника, послуги/продукцію, події в житті компанії. Часто містить різні функціональні інструменти для роботи с контентом (пошук і фільтри, корпоративні блоги, форуми). Може бути інтегрований з внутрішніми системами власника (КІС, бухгалтерськими системами). Може містити закриті для певних груп користувачів (співробітників дилерів, контрагентів));

- Каталог продукції (містить докладний опис товарів/послуг, сертифікати, технічні і споживчі дані, відгуки);

- Інтернет-магазин -- сайт з каталогом продукції, за допомогою якого клієнт може замовити продукції. Використовує різні системи розрахунків;

- Промосайт – сайт про конкретну торгову марку або продукт, якій містить інформацію про бренд, різних рекламних акціях.

Веб-сервіс-сайт (сайт створений для виконання будь-яких завдань або надання послуг в рамках мережі WWW).

- Дошка оголошень – ресурс, на якому є можливість розміщення публічного оголошення про продаж або купівлю товарі і послуг або короткої інформації

- Каталог сайтів – це ресурс, на якому розміщується сайти блоги (наприклад, Open Directory Project).

- Пошукові сервіси (наприклад, Google, Bing);

- Поштовий сервіс (наприклад, Gmail, Outlook);

- Файлообмінний пірінговий (peer-to-peer) сервіс (наприклад, BitTorrent);

- Хмарне середовище даних (Google Dick, OneDrive);

- Сервіс редагування даних (наприклад, Google Docs);
- Фотохостінг (наприклад, Picnik, ImageShack);
- Відео-хостінг (наприклад, YouTube);
- Соціальні медіа;
- Комбіновані веб-сервіси (соціальні мережі (наприклад, Facebook, Twitter); спеціалізовані соціальні мережі (наприклад, MySpace, LinkedIn)).

Основною тенденцією сучасного ринку реклами є інтенсивний розвиток інтернет-реклами як у світі, так і в Україні.

Найпопулярніших видів реклами, як зазначають дослідники [6-10] є наступні види: контекстна; таргетингова; медійна; банерна; SEO; вірусна, в соціальних мережах, тизерної, E-mail- розсилок та інші.

Для класифікації інтернет-реклами використовуються наступні ознаки: спосіб розміщення; спосіб поширення; канал поширення, спосіб створення, форма подання рекламної інформації [11].

За способом розміщення реклами виокремлюють наступні види:

- таргетингова (контекстна, локальна, геоконтекстна тощо) – реклама, що надсилається певним категоріям споживачів, обраним за певною ознакою (інтереси, географічне розміщення, тематика тощо);
- медійна – розміщення текстово-графічних рекламних матеріалів на рекламних майданчиках.

За способом поширення виокремлюють наступні види:

- звичайна – реклама поширюється через всі можливі канали на договірних засадах з рекламодавцями та клієнтами;
- вірусна – характеризується прогресуючою передачею рекламного матеріалу від користувача до користувача з ініціативи самих користувачів;
- спонсорська – певний тип присутності бренду в програмі, який підбирається індивідуально з урахуванням завдань рекламодавця (поліпшення іміджу бренду, формування попиту та ін.);
- партизанська – малобюджетні чи безоплатні способи просування товару, що ґрунтуються на винахідливості та креативності;

- оплачувана – користувач отримує певну оплату за перегляд реклами або підписку на рекламу.

За каналом поширення виокремлюють наступні види:

- поштова – рекламні матеріали, що надсилаються електронною поштою;
- пошукова – реклама в результатах пошуку пошукових систем;
- соціально-мережева – рекламна інформація, що поширюється через соціальні мережі (купони, конкурси, групи в соціальних мережах).
- Youtube-реклама.

За формою подання рекламної інформації виокремлюють наступні види:

- банерна – реклама, яка для поширення потрібної інформації в інтернеті використовує графічні носії банери;
- блочна – має форму текстових блоків, котрі, на відміну від банерів, містять лише текстову інформацію;
- відео-реклама – форма реклами, що розміщується в інтернеті, і направлена на створення іміджу компанії, просування послуг чи товарів, представлення інформації з ціллю підвищення продажу;
- віконна (спливаючі вікна та інш.);
- текстова – в якості рекламного оголошення використовується текст з гіперпосиланням;
- тизерна – реклама, що складається з тексту та картинки, мета якої зацікавити споживача та спонукати до дії (натиснути, зателефонувати, замовити тощо).

За способом створення виокремлюють наступні види:

- нативна – природна реклама, що розробляється відповідно до тематики, стилю оформлення ресурсу, на якому розміщується;
- новинна – створення реклами на основі поточних новин;
- продакт-плейсмент – реклама, впроваджена в тканину ресурсу (відео, аудіо, книги);
- SEO-наповнення.

В роботах зазначається [12], що носій інформації якнайбільше впливає на сприйняття інформації, оскільки задіює різні органи чуття, застосовує різні способи привернення та утримання уваги споживача.

Таблиця 1.1 -Характеристика та особливості Інтернет-реклами

Назва	Характеристика	Особливості
Search	Вибір демонстраційних рекламних повідомлень визначається з урахуванням пошукових запитів користувачів	Має переваги в залученні відвідувачів та при роботі з брендами, ціна реклами залежить від основних слів, значний потенціал трафіку
Email	Розсилання електронних рекламних листів цільовій аудиторії, при цьому використовуючи уже створену базу накопичених адрес	Дає можливість встановити моментальний контакт із користувачами, персоналізація усіх повідомлень. Має складність роботи зі значною кількістю даних.
Display	Графічно-текстові рекламні матеріали (гіперпосилання) на сайті рекламодавця.	Гнучкість у цінах та форматах. Недоліки: після завершення рекламної кампанії є велика ймовірність скорочення відвідувачів
Affiliate	Просування продукту чи бізнесу за допомогою партнерів.	Оплата за фактом дії чи продажу, має швидкий запуск, нескладний контроль. Недоліки полягають у складності прогнозування, наявності репутаційних ризиків
Social	Залучення уваги до бренду, продукту через соціальні мережі чи платформи.	Наявність великої кількості бази користувачів, доступність для різного сегментування. Можливість працювати із повідомленнями, лідерами думок.
SEO	Оптимізація видачі сайту у пошукових, соціальних мережах.	Є відносно недорогим і достатньо довговічним джерелом трафіку на сайт. Та потребує грамотну роботу. Відмінна робота у системі з іншими каналами залучення.

Особливості функціонування ринку товарів і послуг України у фізичному і цифровому просторі у 2023 році.

1. Війна з росією. Головною проблемою українських підприємців, залишається війна та усі наслідки, які вона з собою несе: пошкодження складів з

готовою продукцією, що негативно вплинуло на терміни поставки серійної продукції, ризик вкладання грошей в розширення, зниження продажів у періоди масованих обстрілів тощо.

2. Конкуренція. Українським підприємцям у 2023 році доводилось активно конкурувати за клієнтів. Опитувані помітили недобросовісну конкуренцію, спроби переманити клієнта усіма можливими методами та демпінг цін.

3. Зниження попиту. Власники інтернет-магазинів розповіли, що у 2023 році помітили зниження попиту та купівельної спроможності. Також, на їх погляд, покупки стали більш вимогливими. Деякі підприємці відмічають, що кількість замовлень стала меншою у порівнянні з попередніми роками.

4. Міжнародна експансія Нової Пошти. У 2023 році Нова Пошта активно відкривала міжнародні відділення. Наразі у компанії більше 80 відділень у 11 країнах Європи. Також логістичний оператор заявив, що за результатами 2023 року вони доставили на 30% більше посилок, ніж у 2022. Загалом компанія опрацювала 412 млн відправлень. Протягом 2023 року в Україні було відкрито 2242 нових відділення та встановлено 1853 поштомати.

5. Штучний інтелект. Стрімке поширення штучного інтелекту на початку 2023 року. Chat GPT, Gemini, Jasper, Midjourney. Багато підприємців намагались впровадити штучний інтелект у роботу на постійній основі.

6. Міжнародні інструменти. У вересні 2023 року Google знову дозволив українським підприємцям реєструвати свій бізнес на Google Maps і повноцінно користуватися сервісом Google Business. Ці корисні інструменти з міркувань безпеки не працювали з початку повномасштабного вторгнення. Із запуском Etsy Payments українці можуть офіційно продавати на платформі клієнтам з Америки, Європи, Азії тощо.

7. Rozetka збільшила власну мережу відділень. З січня по жовтень 2023 маркетплейс Rozetka відкрив 141 точку видачі. За даними компанії, за рік їхня кількість зросла у більш ніж півтора рази. Станом на жовтень 2023 року, 416 відділень працювали у 122 населених пунктах України.

8. Розвиток експорту. Все більше представників різного українського бізнесу виходять на міжнародний ринок. Моделі роботи відрізняються в залежності від розмірів компаній, асортименту та цілей бізнесу. Зросла і популярність іноземних маркетплейсів, підприємці активніше виходять на Amazon, Ebay, Allegro та інші майданчики.

Таким чином, у 2023 році українські інтернет-магазини зіткнулися з різними проблемами та мали багато викликів. Але e-commerce продовжує розвиватися, опановувати різні канали просування та забезпечувати українців необхідними для життя товарами:

Споживчі тенденції. За дослідженням найпопулярніший канал для просування інтернет-магазинів у 2023 році – це Google Shopping. На другому місці – продажі через органічний трафік на сайт, на третьому – продаж на маркетплейсах.

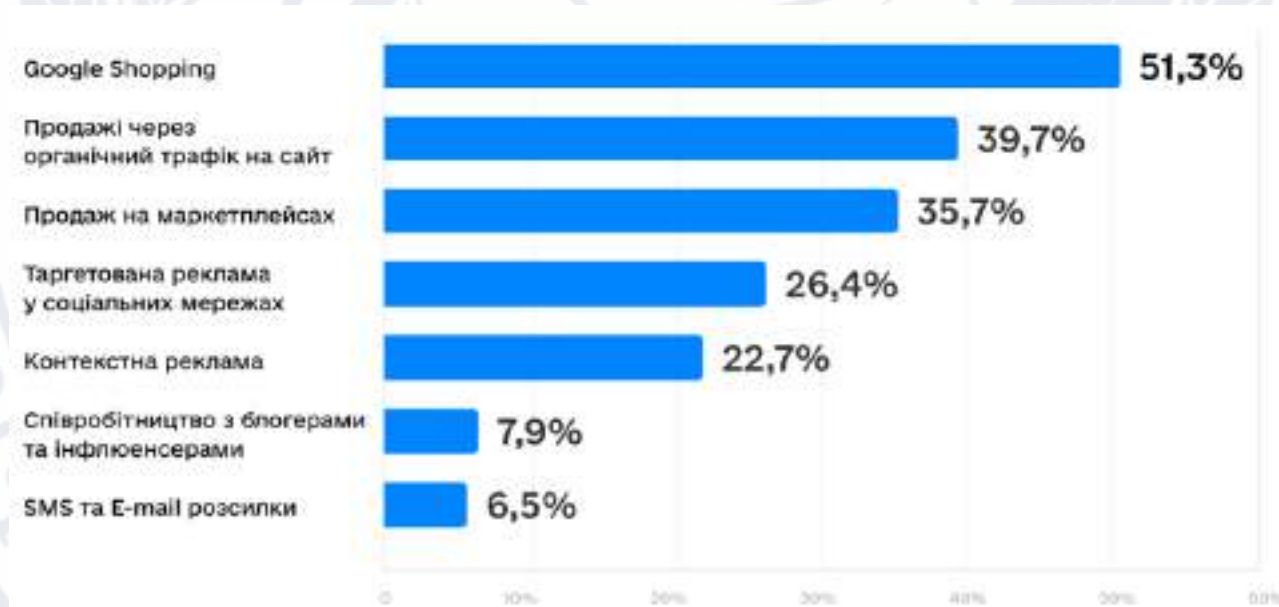


Рисунок 1.1. Доля найбільш успішних каналів продажу для інтернет-магазинів у 2023 році

<https://ua-retIII.com/2024/03/yak-pracyuvav-ukra%1%97nskiy-e-commerce-u-2023-roci-doslidzhennya/>

Повномасштабна війна принесла зміни в життя всіх українців. І люди незалежно від віку, статі чи регіону проживання переглянули свої пріоритети у споживанні, дозвіллі та плануванні майбутнього. Компанія Gradus Research спеціально для 17 Українського маркетинг-форуму провела дослідження настроїв трьох поколінь українських споживачів на третьому році повномасштабної війни та порівняла їх між собою. Досліджувалось, що купують, як розважаються та які плани будують три різних покоління українців (покоління Х, міленіалів та зумерів) під впливом війни.

Щодо пріоритетів споживання, з поміж категорій, на яких економлять українські споживачі: люди старші за 45 років та міленіали найчастіше обирають відмову від розваг та відпочинку (67% та 60% відповідно); на другому місці – їжа поза домом, а на третьому – нова побутова техніка та електроніка.

Щодо культурних запитів, то більшу частину вільного часу зети і міленіали проводять за прослуховуванням музики та переглядом фільмів чи серіалів онлайн. Ікси теж надають перевагу фільмам і серіалам, а також читанню новин. Близько третини міленіалів та зетів слухають українську музику, ще половина і тих, й інших слухає і українських, і зарубіжних виконавців.

Світовий ринок інтернет-реклами характеризується високою динамікою розвитку, яка проявляється в зростанні його обсягів та структурної перебудови. Найбільшими темпами зростання серед різних видів інтернет-реклами вирізняються SMM та пошукова реклама. Аналіз локального ринку інтернет-реклами підтвердив цю тенденцію [13].

1.2 Сучасні підходи до оцінки ефективності інтернет-реклами

Теоретичні аспекти, зокрема щодо економічної сутності ефективності рекламної діяльності, містяться в працях Даніеля Старча, Пола Черінгтона, Клауда Хопкінса, Харлоу Гейла, Волтера Скотта. Цими вченими було розглянуто питання загальної оцінки ефективності реклами.

Дослідженням проблем пошукової реклами та оцінки її ефективності займалися вітчизняні науковці, такі як: І. Бойчук, В. Голик, Є. Домашева, Л.

Капінус, В. Ковшик, М. Макарова, І. Савельєва, О. Сологуб.

Виокремлюють наступні способи оцінки якісних і кількісних показників ефективності інтернет-реклами.

За напрямками:

- комунікативна (інформаційна) ефективність реклами – дозволяє встановити, наскільки ефективне конкретне рекламне звернення передає цільовій аудиторії необхідні відомості, або формує бажану точку зору;
- економічна ефективність реклами – співвідношення між результатом, отриманим від реклами та розмірами витрат (матеріальних фінансових) за певний період (доба, тиждень, місяць, період рекламної кампанії і т.д.).

Основний критерій комунікативної ефективності реклами є індекс поінформованості AW.

Індекс поінформованості AW чисельно дорівнює відношенню числа користувачів, які знають про вміст реклами, до загального обсягу цільової аудиторії. В межах Інтернету підрахунок AW досить проблематичний, однак є ряд параметрів, які на відміну від AW піддаються більш точним підрахункам.

На вибір ключових метрик впливають завдання рекламної кампанії, проте кожна з них здатна показати, наскільки добре працюють оголошення. Найпоширенішими KPI вважаються:

CR - коефіцієнт конверсій показує, який відсоток відвідувачі сайту зробили певну цільову дію (покупку, телефонний дзвінок, залишили заявку чи зареєструвалися на сайті). Для розрахунку CR кількість цільових дій ділять на кількість відвідувачів з множенням на 100%.

CTR - визначає, яка частина користувачів з тих які побачили оголошення перейшло по ньому на сайт.

CPC – гроші, які сплачує замовник оголошення за кожен перехід по посиланню. Для виконання розрахунків необхідно розділити суму, витрачену на одну промоакцію, на кількість людей, які вчинили переходи.

CPL – ціна одного ліда, визначається кількістю заявок, що потрібно для отримання клієнта і яка його цінність.

CPO – ціна одного замовлення. У даній метриці враховуються тільки ті користувачі, які залишили заявку на товар. При цьому факт оплати самого замовлення не має значення.

CPM – вартість 1 тисячі показів. При запуску реклами для підвищення охоплення ця метрика дозволить визначити ефективність оголошення. Дозволяє оцінити ефективність підвищення впізнаваності бренду та довіри користувачів та охопленість аудиторії.

CPA – ціна за кожну реєстрацію, завантаження файлу, а також перегляд контенту і певної кількості сторінок. У цей показник може бути включена будь-яка дія, виконана користувачем на вашому сайті.

CAC – це сума, яку рекламодавець витрачає на залучення кожного нового клієнта. Для розрахунку CAC потрібно загальні витрати поділити на кількість нових клієнтів.

Ефективність реклами визначається такими показниками, як надійність, доступність, інноваційність, користь та інші. [12]

Ефективне використання інтернет-маркетингу сприяє:

- збільшенню обсягів продажів,
- покращенню сприйняття бренду,
- залученню нових клієнтів,
- збільшенню рентабельності підприємств.

Узагальнену інформацію щодо системи показників оцінки ефективності пошукової реклами представлено в таблиці 1.2.

Також розглянемо ще два підходи до оцінювання ефективності інтернет-маркетингу, такі як Post-click та Post-view аналіз.

Концепція Post-click аналізу дозволяє підприємствам глибше зрозуміти поведінку своїх користувачів після того, як вони взаємодіяли з рекламою або потрапили на веб-сайт.

Таблиця 1.2 - Система показників оцінки ефективності пошукової реклами

Група показників	
Маркетингові показники	Фінансові показники
Джерела інформації для відповідних груп показників	
Сервіси Google Ads, Ringostat, CallTracking, Roistat, RankMoniror, Serpstat, Allpositions, Seo Gadget	Інформація корпоративної CRM-системи або бухгалтерської програми
Перелік показників, що застосовуються	
CTR (клікабельність) CR (коефіцієнт конверсії) CPC (вартість кліка) CPA (вартість цільової дії) CPL (ціна за лід)	ROAS (коефіцієнт рентабельності інвестицій в рекламу) ROI (коефіцієнт рентабельності інвестицій) ROMI (коефіцієнт рентабельності інвестицій в маркетинг) CAC (вартість залучення нового клієнта) LTV (прибуток за час співпраці з клієнтом)

Post-click аналіз — це процес вивчення поведінки користувачів на веб-сайті після того, як вони відкрили рекламний банер або посилання та здійснили перехід на сторінку підприємства. Цей аналіз дозволяє визначити, наскільки ефективно реклама перетворюється на конкретні дії користувачів, такі як реєстрація, покупка, заповнення форми або інші дії, які є цільовими для маркетингової стратегії. Post-click зосереджений на вивченні дій користувачів на веб-сайті або в додатку, таких як перегляд сторінок, реєстрація, здійснення покупок або інші форми взаємодії. Він допомагає компаніям зрозуміти, як реклама впливає на потенційних клієнтів, а також дозволяє оцінити повернення вкладень від рекламних кампаній.

Таблиця 1.3 - Переваги застосування веб-аналітики в маркетинговій діяльності

Покращення веб-сайту	Веб-аналітика надає відомості про те, як користувачі взаємодіють із веб-сайтом, що дозволяє ідентифікувати слабкі місця та можливості для поліпшення. Підприємства можуть оптимізувати інтерфейс, структуру сторінок і контент для забезпечення кращого користувацького досвіду.
Сегментація аудиторії	За допомогою веб-аналітики підприємства можуть розуміти свою аудиторію, визначити її інтереси та потреби. Це допомагає створювати більш спрямовану та ефективну маркетингову стратегію.
Вимірювання конверсій	Веб-аналітика дозволяє відстежувати конверсії – тобто дії, які ведуть до бажаного результату, наприклад, такого як покупка продукту або заповнення форми зворотного зв'язку. Це дозволяє визначити, які маркетингові заходи та ресурси приносять найкращі результати
Оптимізація рекламних кампаній	За допомогою веб-аналітики можна визначити, які рекламні кампанії привертають більше трафіку та генерують більше конверсій. Це дозволяє підприємствам зменшити витрати на неефективну рекламу та зосередитися на більш успішних стратегіях.
Розробка змісту	Аналіз даних веб-аналітики допомагає розуміти, який контент привертає найбільше уваги користувачів. Це дозволяє створювати більш привабливий та цікавий контент для аудиторії.

Однією з основних ролей Post-click аналізу в маркетингу для підприємств є вимірювання конверсій. Конверсія — це дія, яку очікується від відвідувача після переходу з реклами. Наприклад, це може бути покупка продукції, заповнення контактної форми для отримання інформації або відправка запиту на консультацію. Post-click аналіз дозволяє вимірювати, скільки користувачів виконують ці дії після переходу з реклами. Це важливо для визначення ефективності рекламних кампаній та розуміння, які з них приносять найкращі результати. Також Post-click аналіз дозволяє визначити, які елементи веб-сайту привертають найбільше уваги користувачів і які можуть бути оптимізовані для підвищення конверсій. Наприклад, ви можете виявити, що певний тип контенту або дизайн сторінки привертає більше уваги і генерує більше конверсій, і в

такому разі вам буде корисно зосередити увагу на розвитку цих елементів.

Принципово іншим підходом до оцінювання ефективності маркетингу в Інтернеті є Post-view-аналіз, який дає змогу відстежувати дії користувачів на сайті після контакту з рекламним оголошенням. За допомогою конверсій Post-view, як впливає з назви, не лише клік, але й перегляди, які користувач має у вказаному вікні огляду, отримують частину кредиту за конверсію. Аналіз Post-view дозволяє оцінити ефективність рекламної кампанії після того, як користувачі перейшли по рекламі. Цей вид аналізу допомагає визначити, наскільки успішно реклама впливає на конверсії, залучення клієнтів та інші ключові метрики.

Інструменти інтернет-маркетингу, які повинні відслідковуватися за допомогою Post-click-і Post-view-аналізу, представлені в табл. 2

Таблиця 1.4 - Розподіл інструментів інтернет-маркетингу за видами аналізу

Post-click	Post-view
Контекстна реклама	Медійна реклама
Пошуковий і зворотній ретаргетинг	Ведення спільнот у соціальних мережах і в блогах
Пошукова оптимізація	Репутаційний моніторинг
Партнерська програма	Просування через статті
Продаючі email-розсилки	Інформаційні email-розсилки

Таким чином, аналізи Post-click та Postview є важливими інструментами, оскільки вони надають цінну інформацію про ефективність рекламних кампаній та вплив на аудиторію. Post-click аналіз дозволяє виміряти конкретні дії користувачів після переходу за посиланням, такі як покупки, реєстрації або інші цільові дії, що допомагає зрозуміти ефективність рекламних кампаній у стимулюванні конверсій. З іншого боку, аналіз Post-view дозволяє визначити кількість відображень рекламного контенту перед аудиторією після її взаємодії зі сторінкою в інтернеті або в соціальних медіа. Ці дані допомагають зрозуміти, наскільки ефективно реклама привертає увагу аудиторії після перегляду. Обидва

аналізи допоможуть бізнесам оптимізувати свої рекламні зусилля, розуміти потреби своєї аудиторії та планувати майбутні кампанії з більшою ефективністю.

1.3. Застосування інтелектуального аналізу та штучного інтелекту в маркетинговій діяльності та постановка завдання

У пошуках альтернативних джерел прибутку ритейлери шукають способи використання наявних активів. Одним з таких активів є дані про клієнтів. У поєднанні з рекламою в місцях покупок – як онлайн, так і в магазині – це дає величезну можливість розвивати свій медіа-бізнес.

На сьогодні більшість інтернет-браузерів поступово припиняють використання сторонніх файлів cookie, тому можливість ритейлу використовувати власні дані, щоб допомогти рекламодавцям контактувати зі споживачами стає все більш актуальною. Це дозволяє або створити новий потік доходу або скоротити власні витрати на рекламу.

Завдяки текстам, зображенням, маркетинговим кампаніям, товарним пропозиціям, створеним за допомогою штучного інтелекту, ритейлери зможуть краще комунікувати з цільовою аудиторією. Крім того, генеративний ШІ може поліпшити та пришвидшити аналіз великих масивів даних, а також надавати прогнози, виокремлювати тенденції тощо.

Використання штучного інтелекту в маркетингу є одним із актуальніших трендів сучасного бізнесу. ШІ дозволяє підприємствам покращити ефективність своєї рекламної діяльності, аналізуючи дані про споживачів та пропонуючи персоналізовані рекламні кампанії. Завдяки алгоритмам машинного навчання штучний інтелект здатний оптимізувати бюджети на рекламу, прогнозувати попит на товари та послуги, а також покращувати взаємодію з клієнтами. Крім того, ШІ дозволяє автоматизувати процеси в маркетингу, такі як сегментація аудиторії, націлення та аналіз ефективності рекламних кампаній. Це допомагає підприємствам заощадити час та ресурси, а також підвищити точність прогнозування результатів своєї діяльності. Використання штучного інтелекту в маркетингу дозволяє підприємствам бути більш конкурентоспроможними на

ринку та ефективно взаємодіяти зі споживачами (рис. 1.2).

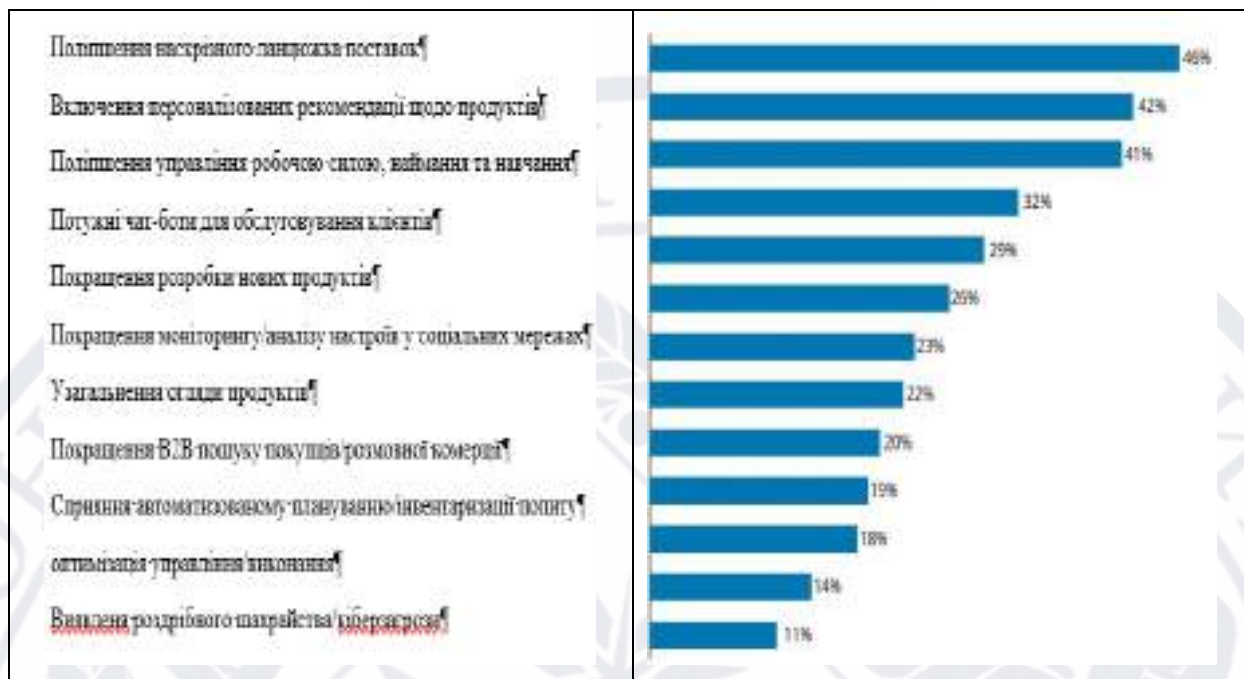


Рисунок 1.2 - Пріоритети використання ІІІ в Інтернет-маркетингу

Автоматизація рекламної діяльності за допомогою штучного інтелекту — це процес використання технологій машинного навчання, аналізу даних та інших методів ІІІ для оптимізації рекламних кампаній і підвищення ефективності маркетингових стратегій.

Основною метою використання штучного інтелекту в рекламній і маркетинговій діяльності є скорочення витрат. Найчастіше за все саме виробництво контенту зумовлює його застосування. Використання баз даних і аналіз аналогічних матеріалів, визначення точно заданої тематики дозволяє ІІІ писати як чернетки для копірайтерів так і якісні рекламні тексти. І це не єдина його властивість. ІІІ може формувати звіти, результати, аналізу даних, що робить його відмінним помічником спеціаліста з реклами.

У сфері цифрового маркетингу поява індивідуальних рекламних стратегій зробили революцію у способах зв'язку бізнесу зі своєю аудиторією. Цей перехід від широких демографічних кампаній до індивідуального обміну повідомленнями є свідченням важливості аналізу даних та досліджень поведінки споживачів.

Використовуючи величезні обсяги даних, що генеруються в результаті онлайн-взаємодій, підприємства тепер можуть створювати повідомлення, які знаходять відгук на особистому рівні, значно підвищуючи ймовірність взаємодії та конверсії.

В основі персоналізованої реклами лежить можливість аналізувати великі набори даних для виявлення закономірностей та переваг, унікальних для кожного користувача. Наприклад, платформа електронної комерції може аналізувати історію відвідувань користувача, записи про покупки та пошукові запити, щоб рекомендувати продукти, що відповідають його минулій поведінці.

Інтеграція складних алгоритмів та методів машинного навчання дозволяє рекламодавцям прогнозувати майбутні потреби споживачів. Яскравим прикладом є використання прогнозової аналітики потоковими сервісами, такими як Netflix, які пропонують шоу та фільми на основі історії переглядів та рейтингів.

Персоналізація переносить фокус із продукту на споживача. Такі бренди, як Spotify, створюють індивідуальні плейлисти, такі як Discover Weekly відображають музичні уподобання користувачів, сприяючи глибшому зв'язку із сервісом.

Ефективність персоналізованої реклами можна безпосередньо виміряти за допомогою таких показників, як рейтинг кліків (CTR) та коефіцієнт конверсії, що дає реальні докази їхнього впливу на успіх бізнесу.

У сфері персоналізованої реклами поворот до стратегій, заснованих на даних, здійснив прорив у способах взаємодії підприємства зі своїми споживачами. Ретельний аналіз даних про споживачів який виявляє не тільки закономірності купівельної поведінки, а й основні мотиви, що керують цими діями.

Підлаштовуючи контент та пропозиції, що відповідають індивідуальним уподобанням та потребам, підприємства можуть створити більш якісний та потрібний продукт для своїх клієнтів. Способи реалізації персоналізації для покращення досвіду клієнтів:

1. Індивідуальні рекомендації щодо продукту: Один ефективний спосіб персоналізації досвіду клієнтів – це надання індивідуальних рекомендацій щодо

продукту, заснованих на індивідуальних уподобаннях та історії покупок. Використовуючи алгоритми, що працюють на ШІ, підприємства можуть аналізувати дані клієнтів і робити точні прогнози, які продукти чи послуги можуть їх зацікавити. Збільшення ймовірності купівлі та покращення загальної якості обслуговування клієнтів.

2. Оптимізована доставка контенту: персоналізація також може застосовуватися до доставки контенту, що забезпечує клієнтам отримання контенту, який є актуальним та цінним для них. Сегментуючи клієнтів на основі їхньої демографії, інтересів чи попередніх взаємодій, підприємства надають цільовий контент, який резонує з кожним клієнтом. Наприклад, Netflix рекомендує фільми та телешоу, засновані на історії та уподобаннях користувача, що робить досвід потокової передачі більш приємним та персоналізованим.

3. Адаптивний досвід веб-сайту: ще один спосіб використання персоналізації – це створення адаптивного досвіду веб-сайтів, які динамічно коригуються на основі поведінки та переваг користувача. Аналізуючи взаємодію з користувачами, підприємства оптимізують макети веб-сайтів, розміщення контенту та навігацію, щоб забезпечити персоналізований досвід перегляду. Наприклад, Spotify коригує свою домашню сторінку на основі прослуховувань користувача, демонструючи персональні плейлисти та рекомендації для покращення досвіду пошуку музики користувачем.

4. Персоналізовані електронні кампанії: маркетинг електронною поштою може бути зроблений ефективнішим завдяки персоналізації. Використовуючи дані клієнтів, такі як історія покупок, поведінка перегляду або демографічна інформація, підприємства створюють цільові та персоналізовані електронні кампанії електронною поштою.

Однією з основних функцій штучного інтелекту у рекламі є створення персоналізованого контенту. Використовуючи алгоритми ШІ, підприємства можуть створити адаптований контент для своєї цільової аудиторії:

1. Генерація динамічного змісту. Інструменти з ШІ аналізують великі обсяги даних клієнтів, включаючи поведінку перегляду, історію покупок та

демографічну інформацію, щоб генерувати динамічний контент, який обслуговує переваги кожної людини. Наприклад, веб-сайт електронної комерції може використовувати ШІ для рекомендації продуктів на основі попередніх покупок клієнта або моделей перегляду, збільшуючи шанси на конверсію.

2. Інформація і використання. Не залежно від типу контенту (формальний, чи неформальний, чи гумористичний) алгоритми ШІ генерують текст, який відповідає унікальному стилю спілкування користувача. Цей рівень персоналізації допомагає встановити глибші зв'язки з клієнтами та підвищити їхній спільний досвід.

3. Індивідуальні електронні кампанії: електронний маркетинг залишається потужним інструментом для залучення клієнтів, але ШІ може підняти його на наступний рівень. Аналізуючи дані клієнтів, ШІ сегментує користувачів на конкретні групи на основі їхніх інтересів, переваг чи демографії. Ця сегментація дозволяє підприємствам відправляти цілеспрямований та відповідний контент, що призводить до вищих показників відкритого та зробленого кліку. Наприклад, інструмент електронної пошти з ШІ може автоматично надсилати персональні рекомендації щодо продуктів клієнтам на основі їх попередніх покупок або перегляду.

4. Оптимізація контенту в соціальних мережах: ШІ допомагає в оптимізації контенту в соціальних мережах, аналізуючи взаємодію з користувачами, настрої та теми тренду. Алгоритми ШІ пропонують найкращий час для публікації, найбільш привабливих форматів контенту та навіть генерувати ідеї контенту на основі переваг користувачів. Цей рівень персоналізації гарантує, що підприємства доставляють правильний контент потрібній аудиторії у потрібний час, максимізуючи охоплення та залучення.

Завдання:

- визначити сучасний стан технологій і тенденції рекламної діяльності на інтернет-ресурсах;
- дослідити інтелектуальні технології аналізу ефективності реклами;
- виконати постановку завдання, збір і попередню обробку даних

рекламних кампаній, очищення та організацію даних;

- виконати аналіз і прогноз динаміки активності користувачів на основі інтелектуальних методів і моделей;
- розробити модель оцінки ймовірності утримання користувачів в залежності від їх характеристик;
- виконати програмну реалізацію інтелектуальної системи аналізу ефективності реклами.

1.4. Висновки за розділом 1

1. Досліджено основні поняття і тенденції рекламної діяльності на інтернет-ресурсах.
2. Проведений аналіз сучасних підходів до оцінки ефективності інтернет-реклами.
3. Визначені особливості застосування інтелектуального аналізу та штучного інтелекту в маркетинговій діяльності та виконано постановку завдань дослідження.

РОЗДІЛ 2 МЕТОДИ ТА ТЕХНОЛОГІЇ ЗБОРУ ДАНИХ І ОЦІНКИ ЕФЕКТИВНОСТІ РЕКЛАМИ

2.1 Методи збору даних для аналізу ефективності реклами

У сучасну цифрову епоху підприємства мають доступ до великих обсягів даних клієнтів. Від перегляду поведінки до історії покупки, ці дані мають цінну інформацію, яка може допомогти стимулювати індивідуальний досвід клієнтів. Однак аналіз великих обсягів даних вручну – це складне та трудомістке завдання. Саме тут допомагає методи інтелектуального аналізу та штучний інтелект. Використовуючи ці методи аналізу даних клієнтів, підприємства можуть оптимізувати процес, отримати більш глибоке розуміння клієнта і підійти до вирішення його завдань і потреб більш персоналізовано.

Перший крок включає збір даних за допомогою різних точок взаємодії, таких як історія онлайн-переглядів, транзакції покупок та активність у соціальних мережах. Наприклад, ритейлер модного одягу може відстежувати товари, які покупець переглядає в Інтернеті, щоб визначити його стильові уподобання.

Інструменти, засновані на ШІ, можуть автоматично збирати та систематизувати дані клієнтів із різних джерел, таких як веб- сайти, платформи соціальних мереж та CRM- системи. Ці інструменти використовують алгоритми для отримання відповідної інформації, усуваючи необхідність ручного введення даних та зниження шансів на людську помилку. Наприклад, інструменти контенту штучного інтелекту можуть зібрати дані з відгуків клієнтів та згадок про соціальні мережі, надаючи бізнесу зворотний зв'язок у реальному часі та аналіз настроїв.

Отже основні технології збору даних на інтернет-ресурсах можна виокремити наступні:

- збір статистичних даних технологічним моніторингом (логи-файли серверу, лічильник відвідувань);
- отримання даних від Web-видавництв (графік та вид розміщення реклами, показники AD Exposure, AD Reach, AD Frequency, AD Impression, CTR);

- соціологічні опитування серед користувачів Інтернету (наприклад, дані професійних дослідницьких агентств);
- отримання додаткових даних про відвідувачів (реєстраційні дані, заповнені анкети, ідентифікація за допомогою файлів cookies з прив'язкою до БД по географії і т.д.);
- отримання внутрішніх даних компанії рекламодавця (показники, що характеризують маркетингову діяльність, продажі і т.д.).

Аналітичні інструменти надають ряд метрик поведінки користувача, які відображають поточну інформацію про ефективність, релевантність трафіку і привабливості оголошення. Сервісом розраховуються окремі метрики, такі як:

- кліки (кількість разів відвідувачі сайту що кликали на банер, щоб потрапити на сайт);
- покази (кількість показів конкретної реклами); сеанси (число відвідувань, в ході яких гості взаємодіють з веб-сайтом); середня тривалість сеансу (середня кількість часу, що провів відвідувач на сторінці);
- сторінки за сеанс (скільки сторінок було переглянуто під час сеансу); показник відмов (відсоток відвідувань, під час яких гості покинули веб-сторінку або не виконали жодних дій протягом 15 секунд) [5, с 14].

2.2 Інтелектуальні методи аналізу і оцінки ефективності реклами

Для інтерпретації даних інтернет-маркетингу використовуються класичні методи аналізу, статистичні методи і методи інтелектуального аналізу. Використовуючи методи інтелектуального аналізу, зокрема машинного навчання, підприємства можуть ідентифікувати закономірності, кореляції та тенденції даних.

Щодо методів інтелектуального аналізу, це вони можуть бути використанні для прогнозування майбутніх покупок на основі минулої поведінки споживача. Це дозволяє прогнозувати поведінку клієнтів, яка використовується для прогнозної оцінки ефективності реклами.

Також для оцінки ефективності реклами спочатку споживачі можуть

сегментуватися на окремі групи за поведінкою при здійсненні покупок.

Однією з ключових переваг використання штучного інтелекту для аналізу даних клієнтів є здатність надавати високоперсоналізовані рекомендації та цільові маркетингові кампанії. Така персоналізація виконується на основі моделей прогнозу ймовірності здійснення цільової дії потенційного споживача з певним набором характеристик споживача та його дій на інтернет-ресурсах.

Реакція споживачів на зворотній зв'язок відстежується для постійного вдосконалення підходу.

На основі перелічених вище показників ефективності реклами на різних типах інтернет-ресурсах в залежності від цілей рекламної компанії для оцінки (прогнозу) ефективності реклами в якості відгука можуть бути обрані наступні показники:

Y_1 - кількість користувачів, які перейшли по оголошенню на сайт;

Y_2 - кількість заявок, що потрібно для отримання клієнта;

Y_3 - кількість користувачів, які залишили заявку на товар;

Y_4 - кількість користувачів, які виконали реєстрацію;

Y_5 - кількість користувачів, які виконали завантаження файлу;

Y_6 - кількість користувачів, які виконали перегляд контенту і певної кількості сторінок;

Y_7 - кількість залучених нових клієнтів;

Y_8 - CR (коефіцієнт конверсій);

Y_9 - CTR (частина користувачів з тих які побачили оголошення перейшло по ньому на сайт);

Y_{10} - CPL (ціна одного ліда, визначається кількістю заявок, що потрібно для отримання клієнта і яка його цінність);

Y_{11} - CAC (сума, яку рекламодавець витрачає на залучення кожного нового клієнта);

Y_{12} - ROAS (коефіцієнт рентабельності інвестицій в рекламу);

Y_{13} - ROMI (коефіцієнт рентабельності інвестицій в маркетинг);

Y_{14} - САС (вартість залучення новогоклієнта);

Y_{15} - LTV (прибуток за час співпраці зклієнтом);

В якості показників, що впливають на ефективність реклами можуть бути обрані наступні показники:

X_1 - вік користувача (у випадку, введення персональних даних при реєстрації);

X_2 - стать, користувача (у випадку, введення персональних даних при реєстрації);

X_3 - кількість відвідувачів сайту;

X_4 - час перегляду контенту,

X_5 - кількість переходів на сайті,

$\overline{X_6}$ - набір характеристик товарів, що переглянуті,

$\overline{X_7}$ - набір характеристик сайтів, що переглянуті,

$\overline{X_8}$ - набір характеристик з логі-файлів,

$\overline{X_9}$ - графік розміщення реклами,

$\overline{X_{10}}$ - вид розміщення реклами,

$\overline{X_{11}}$ - набір характеристик географії користувача,

$\overline{X_{12}}$ - набір характеристик цільових дій користувача.

Персептрон для бінарної класифікації визначено у вигляді

$$y = f(\mathbf{x}, \mathbf{w}) == f\left(\sum_{j=1}^J w_j x_j + w_0\right), \quad f(s) = \begin{cases} 0, & s \leq 0, \\ 1, & s > 0, \end{cases} \quad (2.1)$$

де

w_j – ваговий коефіцієнт для j -го ознаки,

w_0 – зміщення.

Середньоквадратична помилка, отримана за i -м об'єктом, визначена у вигляді

$$e_i = \frac{1}{2} (d_i - f(\mathbf{x}_i, \mathbf{w}))^2,$$

де

$\mathbf{x}_i$ – i -й вектор значень ознак,

d_i – i -е значення навчального відгуку, $d_i \in \{0,1\}$.

Функція втрат у випадку бінарної класифікації:

$$F(\mathbf{w}) = R(\mathbf{w}) + E(\mathbf{w}),$$

$$E(\mathbf{w}) = \sum_{i=1}^I e_i,$$

де

$E(\mathbf{w})$ – втрати на навчальних даних (сума середньоквадратичних помилок по всій множини об'єктів),

$R(\mathbf{w})$ – втрати регуляризації.

Завдання ідентифікації вектору параметрів представлено у вигляді

$$F(\mathbf{w}) \rightarrow \min_{\mathbf{w}}.$$

Моделі прогнозу ефективності.

Лінійний регресійний поліном для апроксимації представлений у вигляді:

$$y = f(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^J w_j x_j + w_0, \quad (2.2)$$

де

w_j – ваговий коефіцієнт j -ої ознаки,

w_0 – зміщення.

Функція втрат у разі апроксимації представлена у вигляді

$$F(\mathbf{w}) = R(\mathbf{w}) + RSS(\mathbf{w}),$$

$$RSS(\mathbf{w}) = \sum_{i=1}^I (d_i - f(\mathbf{x}_i, \mathbf{w}))^2,$$

де

$RSS(\mathbf{w})$ – втрати на навчальних даних (сума квадратів залишків),

$R(\mathbf{w})$ – втрати регуляризації,

$\mathbf{x}_i$ – i -й вектор значень ознак,

d_i – i -е значення навчального відгуку.

Втрати регуляризації в випадку:

– відсутність регуляризації (звичайна лінійна регресія)

$$R(\mathbf{w}) = 0,$$

– регуляризації l_1 (регресія LASSO (Least Absolute Shrinkage and Selection Operator))

$$R(\mathbf{w}) = \lambda \|\mathbf{w}\|_{l_1},$$

– регуляризації l_2 (регресія Ridge)

$$R(\mathbf{w}) = \lambda \frac{1}{2} \|\mathbf{w}\|_{l_2}^2,$$

– комбінації регуляризації l_1 і l_2 (еластична мережа)

$$R(\mathbf{w}) = \lambda \left(\frac{1-\rho}{2} \|\mathbf{w}\|_{l_2}^2 + \rho \|\mathbf{w}\|_{l_1} \right),$$

де

λ – параметр регуляризації, $\lambda > 0$,

ρ – параметр змішування, $0 < \rho < 1$.

Завдання ідентифікації вектору параметрів представлено у вигляді

$$F(\mathbf{w}) \rightarrow \min_{\mathbf{w}}.$$

Для класифікації користувачів щодо здійснення цільової дії використаємо наступні моделі класифікації клієнтів.

В логістичній регресії ми застосовуємо логістичну функцію:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (2.3)$$

Для підгонки моделі (2.3) ми застосовуємо метод максимальної правдоподібності.

Після невеликих перетворень (2.3) виявимо, що

$$\frac{p(X)}{1-p(X)} = e^{\beta_0 + \beta_1 X} \quad (2.4)$$

Величина $\frac{p(X)}{1-p(X)}$ називається ризик події та може набувати будь-яких значень від 0 до ∞ .

Якщо ми прологарифмуємо обидві частини (2.4), то отримаємо

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X \quad (2.5)$$

Ліва частина у цьому рівнянні називається логарифмом ризику, або логітом. Як ми бачимо логіт регресійної моделі (2.5) є лінійним по X .

В лінійній регресійній моделі β_1 відображає середню зміну Y , викликану збільшенням X на 1. У той же час в логістичній регресійній моделі збільшення X на одну одиницю змінює на β_1 логарифм ризику (2.5), або, що те ж саме, примножує ризик на e^{β_1} (2.4). Незалежно від значення X при додатньому β_1 зростання X буде супроводжуватися зростанням $p(X)$, а при від'ємному β_1 зростання X призведе до зниження $p(X)$.

Оцінка коефіцієнтів моделі. Коефіцієнти β_0 та β_1 у (4.1) невідомі і підлягають оцінюванню за доступними навчальними даними. Застосуємо метод максимальної правдоподібності. Функція правдоподібності має вигляд:

$$l(\beta_0, \beta_1) = \prod_{i: y_i=1} p(x_i) \prod_{i': y_{i'}=0} (1 - p(x_{i'})). \quad (2.6)$$

Оцінки $\hat{\beta}_0$ та $\hat{\beta}_1$ обирають таким чином, щоб максимізувати функцію правдоподібності.

Метод максимальної правдоподібності є загальним і використовується для підгонки багатьох нелінійних моделей.

2.3 Технології оцінки ефективності реклами

Ефективність реклами визначається не лише тим, наскільки вона привертає увагу цільової аудиторії, а й тим, чи досягає вона поставлених бізнесом цілей. Сучасні технології оцінки ефективності реклами дозволяють комплексно виміряти її вплив, аналізувати реакції споживачів та оптимізувати рекламні кампанії. До найпоширеніших технологій оцінки ефективності реклами відносяться системи аналітики, трекінгові технології, A/B тестування, моделі атрибуції, біг дата та інструменти штучного інтелекту.

1. Системи аналітики

Основою оцінки ефективності реклами є системи аналітики, такі як Google Analytics, Facebook Insights, Adobe Analytics та інші платформи. Вони дозволяють відслідковувати зокрема наступні ключові показники:

- кількість кліків,
- час перебування на сторінці,
- кількість конверсій,
- коефіцієнт відмов,
- вартість за клієнта (CPA).

Такі дані допомагають зрозуміти, наскільки успішною була реклама в досягненні конкретних цілей.

Google Analytics – інструмент, що дозволяє аналізувати поведінку користувачів, які переходять за рекламними оголошеннями, і відстежувати, які канали генерують найбільшу кількість конверсій.

Розглянемо, наприклад, задачу прогнозування відтоку для розробників ігор, які використовують Google Analytics 4 (GA4) і BigQuery ML.

Утримання користувачів може бути серйозною проблемою для розробників мобільних ігор. Щоб утримати більший відсоток користувачів після першого використання програми, розробники можуть вжити заходів, щоб мотивувати та стимулювати певних користувачів повернутися. Але для цього розробникам потрібно визначити схильність кожного конкретного користувача повертатися після перших 24 годин.

Ресурси Google Analytics 4 (GA4) об'єднують вимірювання додатків і веб-сайтів на одній платформі та тепер є стандартними в Google Analytics. Будь-яка компанія, яка хоче виміряти свій веб-сайт, програму або обидва, може використовувати GA4 для повнішого уявлення про те, як клієнти взаємодіють із їхнім бізнесом. Із запуском Google Analytics 4 експорт даних Google Analytics у BigQuery тепер доступний для всіх користувачів.

Налаштувавши експорт BigQuery, можна досліджувати дані в BigQuery. Google Analytics 4 використовує модель вимірювання на основі подій. Кожен рядок у даних є подією з додатковими параметрами та властивостями. Схема для BigQuery Export може допомогти зрозуміти структуру даних.

Використаємо загальнодоступні зразки даних експорту з фактичної мобільної ігрової програми під назвою "Flood It!" (Android, iOS), щоб створити модель прогнозування відтоку.

Ось як виглядають дані. Кожен рядок у наборі даних є унікальною подією, яка може містити вкладені поля для параметрів події.

```
SELECT *  
FROM `firebase-public-project.analytics_153293282.events_*`  
TABLESAMPLE SYSTEM (1 PERCENT)
```

Результат запиту представлений на рисунку 2.1.

```
SELECT  
  COUNT(DISTINCT user_pseudo_id) as count_distinct_users,  
  COUNT(event_timestamp) as count_events  
FROM  
  `firebase-public-project.analytics_153293282.events_*`
```

Результат запиту представлений на рисунку 2.2.

Row	event_date	event_timestamp	event_name	event_params.key	event_params.value.string_value	event_params.value.int_value	event_params.value.float_value	event_params.value.double_value	event_previous_timestamp	event
1	20180826	1533621577502129	screen_view	feedback_screen_id	null	-8217158045436942191	null	null	15356214489831029	
				feedback_screen_class	FeedbackActivity	null	null	null		
				feedback_event_origin	auto	null	null	null		
				feedback_screen_id	null	-8217158045436942191	null	null		
				feedback_screen_class	game_board	null	null	null		
2	20180829	1537755613657019	user_engagement	feedback_screen_class	FeedbackViewController	null	null	null	15377550327037919	
				feedback_event_origin	auto	null	null	null		
				feedback_screen_id	null	1770655604720481032	null	null		
				engagement_time_msec	null	14719	null	null		
3	20180829	153775561143035	user_engagement	feedback_screen_class	game_board	null	null	null	15377550866414916	
				feedback_event_origin	auto	null	null	null		
				feedback_screen_id	null	1770655604720481034	null	null		
				engagement_time_msec	null	1548	null	null		
4	20180829	153775561000009	user_engagement	feedback_screen_class	FeedbackViewController	null	null	null	1537755086703039	
				feedback_event_origin	auto	null	null	null		
				feedback_screen_id	null	1770655604720481037	null	null		

Рисунок 2.1 - Структура набору даних про користувачів мобільної ігрової програми

Цей набір даних містить 5,7 млн. подій від понад 15 тисяч користувачів.

	count_distinct_users	count_events
0	15175	5700000

Рисунок 2.2 - Характеристики набору даних

Мета аналізу даних полягає в тому, щоб використовувати BigQuery ML на прикладі набору даних додатка, щоб передбачити схильність до відтоку чи відтоку користувачів на основі демографічних показників і активності користувачів протягом перших 24 годин після встановлення додатка.

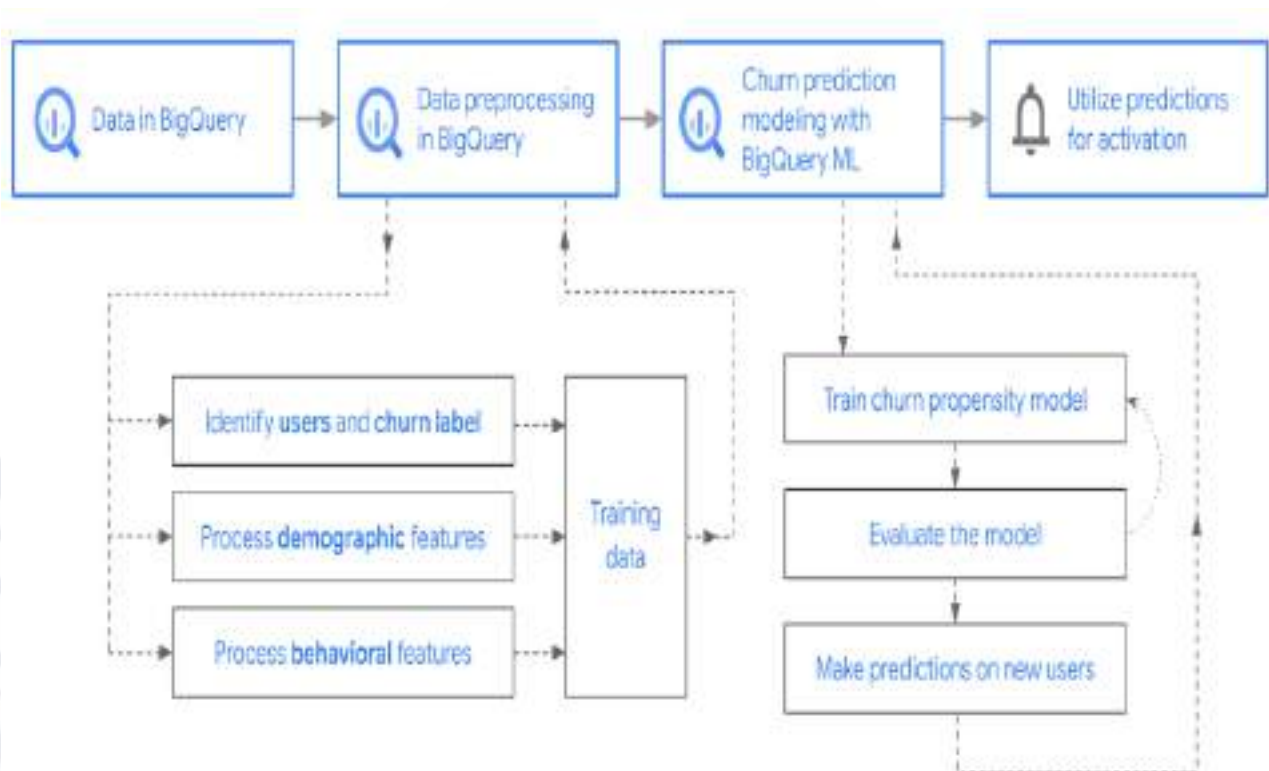


Рисунок 2.3 - Схема інформаційної технології BigQuery ML

Facebook Insights – технологія, що надає дані про ефективність реклами в соціальній мережі Facebook. Вона показує:

- кількість користувачів, які побачило рекламу,
- кількість користувачів, які взаємодіяли з рекламою,
- демографічні дані цільової аудиторії.

Adobe Analytics – дозволяє збирати детальні дані про взаємодії користувачів з рекламою, аналізувати багатоканальні кампанії та створювати глибокі аналітичні звіти.

2. Трекінгові технології

Технології трекінгу є невід'ємною частиною процесу оцінки ефективності реклами. Вони дозволяють відслідковувати шлях користувача від моменту його знайомства з рекламою до кінцевої дії (наприклад, покупки чи заповнення форми). До основних трекінгових інструментів відносяться:

- пікселі відстеження,
- куки-файли;

- інструменти для збору UTM-міток.

Пікселі відстеження – це маленькі зображення, які вставляються в рекламу або на вебсайт для збору інформації про дії користувачів. Наприклад, Facebook Pixel відслідковує поведінку користувачів після кліку на рекламу в Facebook.

UTM-мітки – спеціальні теги, які додаються до URL і дозволяють розрізняти трафік з різних рекламних кампаній. Вони дають змогу відслідковувати ефективність кожного оголошення, кожної платформи та конкретного каналу.

Наприклад, посилання з UTM-мітками:

https://example.com/motorcycles/cruiser/2021/7/?utm_source=google&utm_medium=cpc&utm_campaign=sales_2021

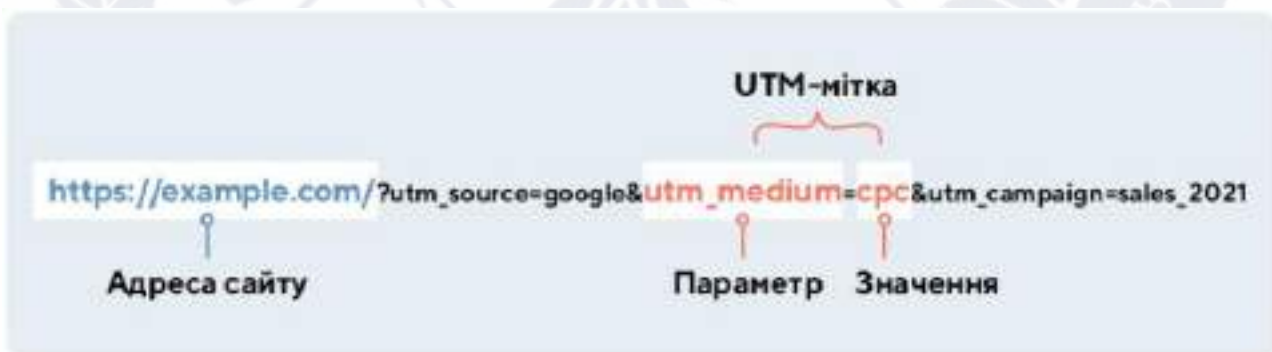


Рисунок 2.4 - Структура посилання з UTM-мітками

Отже, кожна UTM-мітка містить такі елементи:

- фіксований параметр, який потрібно відстежити, наприклад, `utm_source`;
- значення параметра, де можна вказати будь-яке слово;
- знак рівності (=) — з'єднує параметр і його значення;
- амперсанд (&) — розділяє параметри UTM між собою;
- знак питання (?) — відокремлює першу UTM-мітку від оригінального посилання.

URL-адреса може містити лише 5 параметрів:

- 3 обов'язкових — `utm_source`, `utm_medium`, `utm_campaign`;
- 2 опціональних — `utm_term` та `utm_content`.

Статистику та переходи за UTM-мітками можна відстежувати в аналітичних системах, таких як Google Analytics або Open Web Analytics. Кожна система унікальна, тому дії можуть відрізнятися.

3. A/B тестування

A/B тестування — це метод, що дозволяє порівняти ефективність двох або більше варіантів реклами. Його суть полягає у тому, щоб створити різні версії оголошення з незначними відмінностями (наприклад, іншим текстом, зображенням або заклик до дії) і визначити, який варіант приносить кращі результати.

A/B тестування використовується для оптимізації оголошень у Google Ads, Facebook та інших рекламних платформах. Воно дозволяє дізнатися, який текст, зображення або інші елементи найкраще працюють для досягнення цілей.

Мультиваріантне тестування – більш просунутий тип A/B тестування, який дозволяє одночасно тестувати декілька елементів оголошення. Це забезпечує більш комплексний підхід до визначення найефективніших компонентів реклами.

4. Моделі атрибуції

Атрибуція – це процес визначення ролі кожного каналу у прийнятті кінцевого рішення користувача. Існує кілька моделей атрибуції, які дозволяють оцінити ефективність реклами в контексті багатоканальних взаємодій.

Моделі на основі останньої взаємодії – приписують всю цінність конверсії останньому каналу, через який пройшов користувач. Це проста модель, але вона не враховує всі попередні взаємодії, що могли вплинути на рішення користувача.

Моделі на основі першої взаємодії – приписують всю цінність конверсії першому каналу, через який користувач дізнався про продукт. Вона допомагає зрозуміти, які канали генерують початковий інтерес.

Лінійна модель – рівномірно розподіляє цінність конверсії між усіма взаємодіями, що враховує вплив кожного каналу на користувача.

Модель часової декадності – приписує більшу цінність каналам, з якими користувач взаємодіяв ближче до моменту конверсії.

5. Big Data та штучний інтелект.

Зростання обсягів даних (Big Data) та розвиток технологій штучного інтелекту (ШІ) дозволяють здійснювати глибоку та масштабну оцінку ефективності реклами. Аналітичні алгоритми ШІ використовуються для обробки великих обсягів інформації, виявлення прихованих закономірностей та прогнозування поведінки споживачів.

Розглянемо, наприклад, формат і схему даних ресурсу Google Analytics 4 та Google Analytics для Firebase, які експортуються в BigQuery. Записи:

- event (ЗАПИС event_params, ЗАПИС item_params)
- user (Поля privacy_info, ЗАПИС user_properties, ЗАПИС user_ltv)
- device
- geo
- app_info
- items (ЗАПИС item_params)
- collected_traffic_source
- session_traffic_source_last_click
- traffic_source
- stream і platform
- ecommerce
- publisher

У проєкт BigQuery для кожного зв'язаного ресурсу Google Analytics 4 і проєкту Firebase додається єдиний набір даних під назвою analytics_<property_id>. Property_id позначає ідентифікатор ресурсу Analytics, який указано в налаштуваннях ресурсу Google Analytics 4 і в налаштуваннях аналітики додатків у Firebase. Дані з кожного ресурсу Google Analytics 4 і кожного додатка, для яких увімкнено експорт BigQuery, експортуватимуться в цей єдиний набір даних.

У випадки якщо увімкнений щоденний експорт даних, для кожного набору даних щодня створюватиметься таблиця events_PPPRММДД (табл. 2.1).

Таблиця 2.1 - Поля event схеми даних ресурсу Google Analytics 4

Назва поля	Тип даних	Опис
batch_event_index	INTEGER	Число, яке вказує на послідовність кожної події в групі відповідно до порядку їх появи на пристрої
batch_ordering_id	INTEGER	Число, яке монотонно зростає щоразу, коли з певної сторінки надсилається запит мережі
batch_page_id	INTEGER	Послідовне число, яке призначається сторінці й збільшується для кожної наступної сторінки під час взаємодії
event_date	STRING	Дата реєстрації події в журналі (формат PPPPMMDD в зареєстрованому часовому поясі вашого додатка)
event_timestamp	INTEGER	Час у мікросекундах (UTC), коли подію було зареєстровано в журналі клієнта
event_previous_timestamp	INTEGER	Час у мікросекундах (UTC), коли подію було раніше зареєстровано в журналі клієнта
event_name	STRING	Назва події
event_value_in_usd	FLOAT	Значення параметра value події, виражене у валютному еквіваленті (у доларах США)
event_bundle_sequence_id	INTEGER	Послідовний ідентифікатор пакета, у якому завантажено ці події
event_server_timestamp_offset	INTEGER	Зміщення позначки часу отримання й завантаження в мікросекундах

Для потокового експорту даних створюється таблиця events_intraday_PPPPMDD. Дані в ній з'являються автоматично протягом дня в міру того як реєструються нові події. Ця таблиця видаляється наприкінці кожного дня після заповнення таблиці events_PPPPMDD.

Дані з деяких пристроїв надсилаються в Analytics не в той самий день, коли активуються події. Через цю затримку дані про події в щоденних таблицях Analytics (events_PPPPMDD) оновлюються протягом трьох днів після таких подій. Події отримують правильні позначки часу незалежно від того, коли було відправлено дані. Якщо дані надходять пізніше, ніж через три дні, події не реєструються.

Кожен стовпець у таблиці events_PRRPMMDD представляє параметр події. При цьому деякі параметри включені в ЗАПИСИ, а деякі ЗАПИСИ, зокрема items і event_params, повторюються.

Алгоритми машинного навчання – здатні аналізувати великі обсяги даних і будувати прогнози на основі минулих поведінкових патернів користувачів. Це дає можливість не лише оцінювати ефективність поточних кампаній, а й прогнозувати результати майбутніх.

Аналіз настроїв – використовується для аналізу коментарів, відгуків, соціальних мереж і показує, яке враження справляє реклама на цільову аудиторію.

Прогнозна аналітика – дозволяє передбачити, як поведінка користувачів зміниться у відповідь на певну рекламну кампанію.

6. Інструменти CRM та персоналізація

Інтеграція даних реклами з CRM-системами дозволяє отримувати глибший аналіз ефективності рекламних кампаній, особливо у випадку з B2B маркетингом. Це дозволяє побудувати індивідуальний підхід до кожного клієнта.

CRM-системи (наприклад, Salesforce, HubSpot) дозволяють зберігати інформацію про взаємодію з клієнтом на всіх етапах, що дозволяє точніше оцінювати, як саме реклама сприяла отриманню нових клієнтів.

Персоналізована реклама – завдяки збору даних про поведінку користувачів, компанії можуть створювати індивідуальні рекламні пропозиції, що підвищує їх ефективність.

7. Соціальні мережі та впливові особи

Соціальні мережі, як Facebook, Instagram, Twitter, стали потужними платформами для оцінки ефективності реклами. Оцінка взаємодій (лайки, коментарі, репости) допомагає зрозуміти, наскільки реклама привертає увагу аудиторії. Використання впливових осіб також дозволяє ефективно збільшувати охоплення реклами та оцінювати її за рахунок кількості згадувань, переглядів і переходів.

Ефективна оцінка рекламних кампаній залежить від використання різноманітних технологій і методів, що забезпечують всебічний аналіз та оптимізацію. Застосування сучасних технологій, таких як аналітичні платформи, інструменти трекінгу, моделі атрибуції, штучний інтелект та CRM-системи, дозволяє маркетологам краще розуміти свою аудиторію, адаптувати контент і досягати більш високих показників ефективності.

2.4. Висновки за розділом 2

1. Систематизовано методи збору даних для аналізу ефективності реклами.
2. Досліджено інтелектуальні методи аналізу і оцінки ефективності реклами.
3. Виконано системний аналіз технологій оцінки ефективності реклами.

РОЗДІЛ 3 ПРОГРАМНА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ АНАЛІЗУ ЕФЕКТИВНОСТІ РЕКЛАМИ

3.1 Збір і попередня обробка даних рекламних кампаній (на прикладі комп'ютерних ігор)

Найчастіше при дослідженні реклами ігор завдання формуються наступним чином. Необхідно дослідити ефективність реклами ігор взагалі та окремих функцій з метою визначення дій щодо утримання і розширення та збільшення залученості користувачів.

Для аналізу даних були використані три джерела даних:

- статистичні дані отримані технологічним моніторингом (логи-файли серверу, лічильник відвідувань);
- даних про відвідувачів (реєстраційні дані, заповнені анкети, ідентифікація за допомогою файлів cookies);
- отримання внутрішніх даних компанії рекламодавця (показники, що характеризують маркетингову діяльність, продажі).

Вихідні дані про активність користувачів комп'ютерних ігор були представлені файлами даних щодо здійснення плати за гру, характеристик користувачів, характеристик ігор і імпортовані у листи Google Sheets, визначені типи даних в кожному стовпці, в результаті чого мають наступний вигляд (рис. 3.1 - 3.3). Розраховані узагальнені значення показників витрат по іграм і кількості користувачів (рис. 3.4).

	A	B	C	D	E
1	user_id	activity_date	game_activity_name	total_seconds	week_start_date
2	5Xx7q8xw22ftCybZ1XZVQ==	2022-09-25	game 3: Main campaign	266	2022-09-19
3	5Xx7q8xw22ftCybZ1XZVQ==	2022-09-25	game 3: Special events	43	2022-09-19
4	C/7YatYfa4qFbB1omZ7Jjw==	2022-09-05	game 2: Battle pass challenges	51	2022-09-05
5	C/7YatYfa4qFbB1omZ7Jjw==	2022-09-28	game 2: Battle pass challenges	846	2022-09-26
6	C/7YatYfa4qFbB1omZ7Jjw==	2022-10-13	game 2: Battle pass challenges	656	2022-10-10
7	C/7YatYfa4qFbB1omZ7Jjw==	2022-10-26	game 2: Battle pass challenges	559	2022-10-24
8	C/7YatYfa4qFbB1omZ7Jjw==	2022-11-07	game 2: Battle pass challenges	1248	2022-11-07
9	C/7YatYfa4qFbB1omZ7Jjw==	2022-11-23	game 2: Battle pass challenges	802	2022-11-21
10	C/7YatYfa4qFbB1omZ7Jjw==	2022-11-28	game 2: Battle pass challenges	129	2022-11-28
11	C/7YatYfa4qFbB1omZ7Jjw==	2022-12-03	game 2: Battle pass challenges	1881	2022-11-28
12	C/7YatYfa4qFbB1omZ7Jjw==	2022-12-16	game 2: Battle pass challenges	197	2022-12-12
13	C/7YatYfa4qFbB1omZ7Jjw==	2022-12-25	game 2: Battle pass challenges	17332	2022-12-19
14	C/7YatYfa4qFbB1omZ7Jjw==	2022-08-08	game 2: Campaign	17	2022-08-08
15	C/7YatYfa4qFbB1omZ7Jjw==	2022-08-14	game 2: Campaign	3	2022-08-08
16	C/7YatYfa4qFbB1omZ7Jjw==	2022-09-12	game 2: Campaign	20	2022-09-12
17	C/7YatYfa4qFbB1omZ7Jjw==	2022-09-29	game 2: Campaign	845	2022-09-26
18	C/7YatYfa4qFbB1omZ7Jjw==	2022-10-25	game 2: Campaign	306	2022-10-24
19	C/7YatYfa4qFbB1omZ7Jjw==	2022-10-26	game 2: Campaign	4	2022-10-24
20	C/7YatYfa4qFbB1omZ7Jjw==	2022-11-25	game 2: Campaign	157	2022-11-21
21	C/7YatYfa4qFbB1omZ7Jjw==	2022-11-26	game 2: Campaign	557	2022-11-21
22	C/7YatYfa4qFbB1omZ7Jjw==	2022-12-04	game 2: Campaign	18	2022-11-28
23	C/7YatYfa4qFbB1omZ7Jjw==	2022-12-30	game 2: Campaign	5	2022-12-26
24	C/7YatYfa4qFbB1omZ7Jjw==	2022-07-26	game 2: Campaign	109	2022-07-25
25	C/7YatYfa4qFbB1omZ7Jjw==	2022-09-25	game 2: Special events and chall	1	2022-09-19
26	C/7YatYfa4qFbB1omZ7Jjw==	2022-10-13	game 2: Special events and chall	13	2022-10-10
27	C/7YatYfa4qFbB1omZ7Jjw==	2022-10-19	game 2: Special events and chall	2659	2022-10-17
28	C/7YatYfa4qFbB1omZ7Jjw==	2022-10-20	game 2: Special events and chall	76	2022-10-17
29	C/7YatYfa4qFbB1omZ7Jjw==	2022-10-26	game 2: Special events and chall	375	2022-10-24
30	C/7YatYfa4qFbB1omZ7Jjw==	2022-10-29	game 2: Special events and chall	549	2022-10-24
31	C/7YatYfa4qFbB1omZ7Jjw==	2022-11-15	game 2: Special events and chall	1	2022-11-07

Рисунок 3.1 - Вихідні дані про оплату за ігри. Початок.

На рисунку 3.1 наведені вихідні дані про оплату за ігри (перші 30 спостережень).

	A	B	C	D	E
43986	SXF4nkTSldHSgQVJ65JcQ==	2022-12-19	game 2: Campaign	1	2022-12-19
43987	SXF4nkTSldHSgQVJ65JcQ==	2022-12-23	game 2: Campaign	1	2022-12-19
43988	SXF4nkTSldHSgQVJ65JcQ==	2022-12-26	game 2: Campaign	18	2022-12-26
43989	SXF4nkTSldHSgQVJ65JcQ==	2022-12-31	game 2: Mini-games	7	2022-12-26
43990	SXF4nkTSldHSgQVJ65JcQ==	2022-12-18	game 2: Clan activities	1	2022-12-12
43991	SXF4nkTSldHSgQVJ65JcQ==	2022-12-17	game 2: Campaign	2	2022-12-12
43992	SXF4nkTSldHSgQVJ65JcQ==	2022-12-22	game 2: Campaign	1	2022-12-19
43993	SXF4nkTSldHSgQVJ65JcQ==	2022-12-28	game 2: Campaign	1	2022-12-26
43994	SXF4nkTSldHSgQVJ65JcQ==	2022-12-29	game 2: Campaign	41	2022-12-26
43995	SXF4nkTSldHSgQVJ65JcQ==	2022-12-31	game 2: Campaign	1	2022-12-26
43996	SXF4nkTSldHSgQVJ65JcQ==	2022-12-25	game 2: Clan activities	1	2022-12-26
43997	SXF4nkTSldHSgQVJ65JcQ==	2022-12-20	game 2: Campaign	13	2022-12-19
43998	SXF4nkTSldHSgQVJ65JcQ==	2022-12-16	game 2: Mini-games	33	2022-12-12
43999	SXF4nkTSldHSgQVJ65JcQ==	2022-12-24	game 2: Mini-games	62	2022-12-19
44000	SXF4nkTSldHSgQVJ65JcQ==	2022-12-25	game 2: Mini-games	1	2022-12-19
44001	SXF4nkTSldHSgQVJ65JcQ==	2022-12-19	game 2: Clan activities	3	2022-12-19
44002	SXF4nkTSldHSgQVJ65JcQ==	2022-12-15	game 2: Campaign	7	2022-12-12
44003	SXF4nkTSldHSgQVJ65JcQ==	2022-12-24	game 2: Campaign	5	2022-12-19
44004	SXF4nkTSldHSgQVJ65JcQ==	2022-12-14	game 2: Mini-games	4	2022-12-12
44005	SXF4nkTSldHSgQVJ65JcQ==	2022-12-15	game 2: Mini-games	104	2022-12-12
44006	SXF4nkTSldHSgQVJ65JcQ==	2022-12-26	game 2: Mini-games	1	2022-12-26
44007	SXF4nkTSldHSgQVJ65JcQ==	2022-12-30	game 2: Mini-games	9	2022-12-26
44008	SXF4nkTSldHSgQVJ65JcQ==	2022-12-17	game 2: Clan activities	1	2022-12-12
44009	SXF4nkTSldHSgQVJ65JcQ==	2022-12-21	game 2: Campaign	5	2022-12-19
44010	SXF4nkTSldHSgQVJ65JcQ==	2022-12-22	game 2: Mini-games	3	2022-12-19
44011	SXF4nkTSldHSgQVJ65JcQ==	2022-12-23	game 2: Mini-games	7	2022-12-19
44012	SXF4nkTSldHSgQVJ65JcQ==	2022-12-29	game 2: Mini-games	84	2022-12-26
44013	SXF4nkTSldHSgQVJ65JcQ==	2022-12-05	game 2: Clan activities	1	2022-12-05

Рисунок 3.2 - Вихідні дані про оплату за ігри. Продовження.

На рисунку 3.2 наведені вихідні дані про оплату за ігри (останні 30 спостережень).

A1	A	B	C	D	E
1	user_id	game_name	language	has_older_device_mor age	
2	LuDHi0aoGmgpUITGreh0tQ==	game 3	uk	TRUE	22
3	G9053rs6jBKitKUhrUdLew==	game 3	uk	FALSE	27
4	54ouP7IBhVCeV0tS1Oj4Q==	game 3	ru	FALSE	21
5	lIdaqkizLaFesmkjCvIzmQ==	game 3	uk	FALSE	24
6	yINOQh7e6gNGQNC0T3RBxw==	game 3	ru	TRUE	26
7	4EiWC2ewIX2A7i9NuBAefQ==	game 3	uk	TRUE	18
8	VGEh7S+cxW9PT7H6KNgEGQ==	game 3	uk	FALSE	24
9	pz7LU0s9I7xgncmHVxQvOw==	game 3	ru	FALSE	24
10	QUmnu9bjwIT1/vYgzDEDQ==	game 3	uk	TRUE	16
11	BjIqmsfEshc/QjsD7VgEjg==	game 3	ru	FALSE	18
12	rw8yxS9+t5PvgiNa7EhMvQ==	game 3	uk	TRUE	18
13	7Em69vKH6zh6HgArCzG0+w==	game 3	ru	FALSE	17
14	72brUr/DKItStTenSpHYg==	game 3	uk	FALSE	24
15	i8yTxTIERu5jtlOFWB0dFO==	game 3	ru	FALSE	33
16	PWvj9YoQTX9FGLGg1CWJQ==	game 3	uk	FALSE	16
17	Ix3SsUV7N0rv93iCexVp5A==	game 3	uk	FALSE	16
18	A6QTeGOKjaGkGkRCFqNybw==	game 3	ru	FALSE	33
19	2onLvYXEEj2aH+D4N9ldYQ==	game 3	uk	FALSE	16

Рисунок 3.3 - Вихідні дані про користувачів ігор.

На рисунку 3.3 наведені вихідні дані про користувачів ігор. (перші 19 спостережень).

	A	B	C
1	game_name	Ad spend	Total users count
2	game 1	0	600
3	game 2	944.33	1266
4	game 3	17322.71	15333

Рисунок 3.4 - Вихідні дані про витрати на рекламу ігор.

Для подальшої роботи проведемо очищення даних, видалення дублікатів та агрегацію даних за користувачами.

Для попереднього аналізу ефективності рекламних кампаній розрахуємо наступні показники, що характеризують загальні фінансові показники ігор і характеристики користувачів (рис. 3.5 -3.7).

- 1) Total revenue -- сума по стовпцю revenue для кожної гри;
- 2) Paid users count -- загальна кількість користувачів, що платять;
- 3) CR to Paid -- конверсія з користувачів в платних користувачів у відсотках;
- 4) ARPPU -- середній revenue на користувача, який платить;
- 5) Average Age (paid users) -- середній вік користувачів, які платять;
- 6) Median age (paid users) -- медіанний вік користувачів, які платять (розрахувати для всіх ігор загалом, а не для кожної гри окремо);
- 7) Minimum age (paid users) -- мінімальний вік користувачів, які платять;
- 8) Maximum age (paid users) -- максимальний вік користувачів, які платять.

	A	B	C	D
1	user_id	game_name	payment_date	revenue_amount_usd
2	fDgyPW0yy0+9letMvl0	game 3	2022-08-18	5,29
3	MSxDpyVU8A+9Wgmng	game 3	2022-08-09	6,37
4	b6IW5fOrJiJUE7dwhEp	game 3	2022-05-04	11,94
5	nkIUUKirU5GYU5ZtNO	game 3	2022-10-25	6,19
6	06RPvisFPCmM9ag+iS	game 3	2022-08-14	6,30
7	f9cWBjiN4c8IXAkBhdql	game 3	2022-10-25	12,56
8	7L4SKPGMd33lnSOq6	game 3	2022-12-04	5,43
9	ngOdQ8blXsraLxTHS8	game 3	2022-07-02	4,64
10	IHTZiMgVj6HJZc31yCC	game 3	2022-11-11	4,91
11	FbbhrtuL39vfmF2Ob\	game 3	2022-11-28	6,32
12	GetHmCsisRTAnwhuBl	game 3	2022-12-03	6,50
13	YC2fZQTQXV2UR1/Wl	game 3	2022-11-06	6,80
14	e/6IajqZ922YY1rh2ErM	game 3	2022-12-17	4,50
15	dpfu6t8G4blqMn55KMf	game 3	2022-05-23	5,27
16	BOcGIFXXESDUbttxDz	game 3	2022-08-29	5,83
17	QZBIEFECZAgVI5ikN9	game 3	2022-05-14	4,88
18	orbgZNBGuHZZx63HT/	game 3	2022-09-23	4,90
19	4oH5e29koOH9zNYWf	game 3	2022-07-06	6,02
20	pLOCMIX+imLizoDehd	game 3	2022-05-29	7,03

Рисунок 3.5 - Розрахунок показника total revenue для ігри 3. Початок.

На рисунку 3.5 наведені результати розрахунку показника total revenue для ігри 3 (перші 20 спостережень).

3014	qqYyR2vRI+7pgluyTnV	game 3	2022-09-25	10,76
3015	fvYhEtD56gROaXqCi9	game 3	2022-07-27	12,62
3016	Veep00i4vvZbTAVD6G	game 3	2022-11-01	4,95
3017	GetHmCsisRTAnwhuBl	game 3	2022-12-06	8,34
3018	UdsILd3VPsla6C2r+wy	game 3	2022-12-26	4,57
3019	1KzIDBp5hORL7Y5PB;	game 3	2022-11-18	15,45
3020	UIYPls8zzm4uX6+FZv	game 3	2022-10-21	4,80
3021	8ncYOMYXeKDelzX40	game 3	2022-09-15	4,90
3022	Vn7Z6/B2x4f/w8R6xVf	game 3	2022-12-03	5,55
3023	q/Tz1C31XWaBUDrRn	game 3	2022-06-12	8,78
3024	zmZPan/KZoyrIR2r/Svl	game 3	2022-10-26	10,56
3025	fCdElzmUtnjNYQLbx81	game 3	2022-04-14	6,04
3026	xt/PY4N2zkZQgtlZ9Cgl	game 3	2022-11-04	8,53
3027	BJlqmsfEshc/QjsD7Vgl	game 3	2022-06-19	5,21

Рисунок 3.6 - Розрахунок показника total revenue для ігри 3. Продовження.

На рисунку 3.6 наведені результати розрахунку показника total revenue для ігри 3 (останні спостереження).

	A	B	C	D	E	F	G	H	I	J	K
1	game name	Ad spend	Total users count	Total revenue	Paid users count	CR to Paid	ARPPU	Average Age (paid users)	Median age (paid users)	Minimum age (paid users)	Maximum age (paid users)
2	game 1	0.00	600	23.13	3	0.50%	7.71	22.33		20	26
3	game 2	944.33	1266	73.61	7	0.55%	10.52	25.00		18	35
4	game 3	17 322.71	15333	20 950.15	373	2.43%	56.17	23.25		14	44
5	Total	18 267.04	17 199.00	21 046.89	383	2.23%	54.95	23.53	22	17.33	38
6								23.27		14	44

Рисунок 3.7 - Узагальнені характеристик рекламних кампаній ігор.

На основі розрахунків можемо зробити наступні висновки:

1. ігри під номером 1 і 2 можуть мають нижчу CR to Paid конверсію в оплату;
2. значна різниця ARPPU між іграми.

Можна висунути наступні гіпотези щодо зазначених проблем:

1. Менший діапазон користувачів по віку в 1 і 2 гри.
2. Можливо 1 і 2 гра мають складний тарифний план, мало рекламуються.

Таким чином, щоб збільшити ефективність реклами потрібно проаналізувати гру, рекламу, зробити відповідні зміни і після цього рекламувати гру.

Для аналізу віку користувачів визначимо наступні характеристики:

1. середній вік;
2. стандартне відхилення (standard deviation);
3. медіану;
4. інтерквартильний розмах (interquartile range);
5. 10 перцентиль віку;
6. 90 перцентиль віку.

Після видалення дублікатів і застосування відповідних функцій в отримаємо значення характеристик віку користувачів ігор (рис. 3.8).

	A	B	C	D	E	F	G	H
1	user_id	game_name	language	has_older_devs	age	Назва показника	Пояснення	
2	5xx7q8w228TcYbZ1xZVQ==	game 3	uk	FALSE	26	COUNT	n=	216
3	C7Yao7YqFqFb1onZ7Jje==	game 2	uk	FALSE	52			
4	9nB4N3m62ypCK7nU2tHlg==	game 3	uk	FALSE	18	AVERAGE	Хочед=	28,01385889
5	GUNmu8bjwT1VYgzDEDQ==	game 3	uk	TRUE	16	Standard deviation	SD	11,26096043
6	8Jqmr5EshcQjsD7VgEjg==	game 3	ru	FALSE	18	MEDIAN	MED=	25
7	72brUnOK8SiTten5pHYg==	game 3	uk	FALSE	24	Quartile	Q1=	19
8	lc3SsUv7NOn50CesVp6A==	game 3	uk	FALSE	16	Quartile	Q3=	34
9	0pBh7SPmBBGuxQJUY5Q==	game 3	uk	FALSE	14	Interquartile range	IQR=	15
10	+Xa5t2+cFw0h4nMGdKg==	game 1	uk	FALSE	28	Percentile	Per(10%)=	17
11	dANFmH1miX5nKDMtbow==	game 3	uk	FALSE	19	Percentile	Per(90%)=	42
12	q8TV+qAgGOnF1B3Hux2A==	game 2	uk	FALSE	32			
13	08T00LQC16Lgl6vAeww==	game 3	uk	FALSE	17			
14	Yh8mYycPmc6CH8Rbjhu3ZQ==	game 3	uk	FALSE	28	Вилучити копії		
15	eadWR0qGk5SjWfG5Kq==	game 3	uk	TRUE	24	Потрапити на DAU не вийшло		
16	TndhXWlnewzKgS2ERKqHg==	game 1	uk	TRUE	51	Застосовано 213 унікальних ігор		
17	zEYHPUpq7lc4PH1p8RVg==	game 3	uk	FALSE	28			
18	6mN0Jsk2ACKHZA86dnpw==	game 3	uk	FALSE	27			
19	0eReSP1mJiaCfNarA17cg==	game 3	uk	FALSE	30			
20	lvNww0deBWVQHR74LODA==	game 1	uk	TRUE	47			

Рисунок 3.8 - Узагальнені характеристик віку користувачів ігор.

Для подальшого аналізу визначимо метрики залученості користувачів ігор. Коли клієнти взаємодіють із продуктом, вебсайтом, послугою або застосунком, це показує, що вони вважають їх цінними; та підтверджує: те, що ти пропонуєш, допомагає клієнтам задовольняти їхні потреби. Якщо залученість користувачів — здорова, це є показником того, що бізнес надає клієнтам або користувачам цінність. Висока залученість допоможе запровадити утримання клієнтів, що призведе до збільшення рекомендацій та прибутку для твоєї організації.

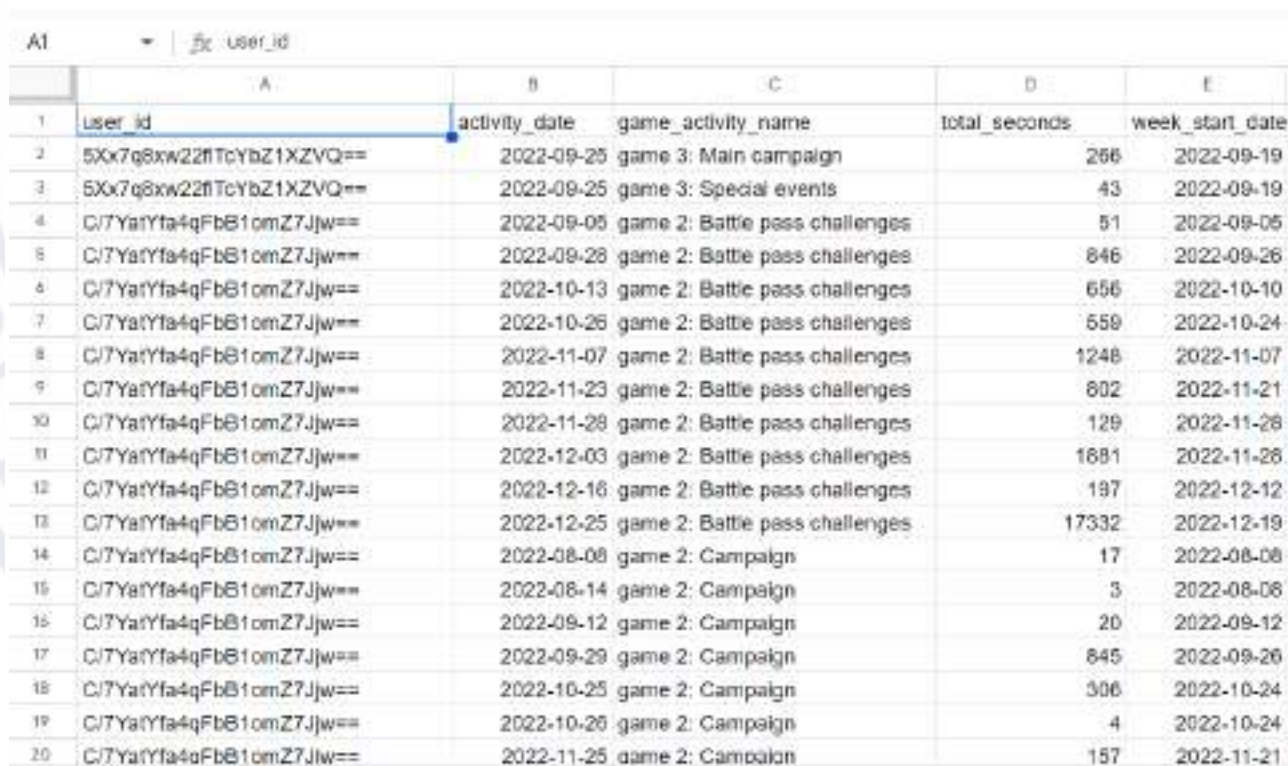
Дуже поширеною практикою є відстеження кількості щоденних активних користувачів (DAU), щотижневих активних користувачів (WAU) та щомісячних активних користувачів (MAU). Ці метрики показують кількість користувачів, які взаємодіють з твоїм продуктом чи послугою протягом вказаного періоду часу. Метрики визначаються наступним чином:

- DAU (daily active users) — це кількість унікальних користувачів, які були активними у вказаний день.

- WAU (weekly active users) вимірює кількість унікальних користувачів, які були активними у вказаний тиждень.
- MAU (monthly active users) — це кількість унікальних користувачів, які були активними у вказаний місяць.

Важливо визначити активність, яку користувач повинен здійснити у бізнесі, щоб бути визначеним як "активний" користувач. Така активність або дія не є однаковими для кожного бізнесу. І завдання аналітика — допомогти визначитися з дією або набором дій, які будуть вважатися за активність. Чи є активним користувачем той, хто відкриває мобільний застосунок, заходить на сайт? Або це той, хто виконує більш конкретну дію з твоїм продуктом: читання статті, подання запиту або написання коментаря, завантаження документа тощо.

Розрахуємо DAU, WAU та stickiness користувачів. Для цього на листі “activity” додамо колонку week_start_date. Використовуємо формули для виведення дати початку тижня для кожної дати у колонці activity_date (рис. 3.9).



	A	B	C	D	E
	user_id	activity_date	game_activity_name	total_seconds	week_start_date
2	5Xx7q8xw22fTcYbZ1XZVQ==	2022-09-25	game 3: Main campaign	266	2022-09-19
3	5Xx7q8xw22fTcYbZ1XZVQ==	2022-09-25	game 3: Special events	43	2022-09-19
4	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-09-05	game 2: Battle pass challenges	51	2022-09-05
5	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-09-28	game 2: Battle pass challenges	846	2022-09-26
6	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-10-13	game 2: Battle pass challenges	656	2022-10-10
7	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-10-26	game 2: Battle pass challenges	559	2022-10-24
8	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-11-07	game 2: Battle pass challenges	1248	2022-11-07
9	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-11-23	game 2: Battle pass challenges	802	2022-11-21
10	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-11-28	game 2: Battle pass challenges	129	2022-11-28
11	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-12-03	game 2: Battle pass challenges	1881	2022-11-28
12	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-12-16	game 2: Battle pass challenges	197	2022-12-12
13	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-12-25	game 2: Battle pass challenges	17332	2022-12-19
14	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-08-08	game 2: Campaign	17	2022-08-08
15	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-08-14	game 2: Campaign	3	2022-08-08
16	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-09-12	game 2: Campaign	20	2022-09-12
17	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-09-29	game 2: Campaign	845	2022-09-26
18	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-10-25	game 2: Campaign	306	2022-10-24
19	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-10-26	game 2: Campaign	4	2022-10-24
20	Cj7YatYfa4qFbS1cmZ7Jjw==	2022-11-25	game 2: Campaign	157	2022-11-21

Рисунок 3.9 - Визначення тижнів активності користувачів.

Для подальших розрахунків створимо лист “DAU” з колонкою activity_date з унікальні дати активності (activity_date) (рис. 3.10).

	A	B	C	D	E	F	G	H	I	J
	activity_date	DAU	week_start_date	Week_day						
1	2022-09-25	76	2022-09-19	неділя						
2	2022-09-26	68	2022-09-26	понеділок						
3	2022-09-28	84	2022-09-26	середа						
4	2022-10-13	98	2022-10-10	четвер						
5	2022-10-26	97	2022-10-24	середа						
6	2022-11-07	96	2022-11-07	понеділок						
7	2022-11-23	71	2022-11-21	середа						
8	2022-11-28	82	2022-11-28	понеділок						
9	2022-12-03	82	2022-11-28	субота						
10	2022-12-16	78	2022-12-12	п'ятниця						
11	2022-12-25	78	2022-12-19	неділя						
12	2022-08-08	71	2022-08-08	понеділок						
13	2022-08-14	69	2022-08-08	неділя						
14	2022-09-12	79	2022-09-12	понеділок						
15	2022-09-29	83	2022-09-26	четвер						
16	2022-10-25	98	2022-10-24	неділя						
17	2022-11-25	76	2022-11-21	п'ятниця						
18	2022-11-26	83	2022-11-21	субота						
19	2022-12-04	84	2022-11-28	неділя						

Рисунок 3.10 - Визначення унікальних дат активності користувачів

На наступному кроці за допомогою функції COUNTUNIQUEIFS визначимо унікальну кількість користувачів, які проводили час на продукті у відповідну дату (рис.3.10).

	A	B	C	D	E	F	G	H	I	J
	activity_date	DAU	week_start_date	Week_day						
1	2022-09-25	76	2022-09-19	неділя						
2	2022-09-26	68	2022-09-26	понеділок						
3	2022-09-28	84	2022-09-26	середа						
4	2022-10-13	98	2022-10-10	четвер						
5	2022-10-26	97	2022-10-24	середа						
6	2022-11-07	96	2022-11-07	понеділок						
7	2022-11-23	71	2022-11-21	середа						
8	2022-11-28	82	2022-11-28	понеділок						
9	2022-12-03	82	2022-11-28	субота						
10	2022-12-16	78	2022-12-12	п'ятниця						
11	2022-12-25	78	2022-12-19	неділя						
12	2022-08-08	71	2022-08-08	понеділок						
13	2022-08-14	69	2022-08-08	неділя						
14	2022-09-12	79	2022-09-12	понеділок						
15	2022-09-29	83	2022-09-26	четвер						
16	2022-10-25	98	2022-10-24	неділя						
17	2022-11-25	76	2022-11-21	п'ятниця						
18	2022-11-26	83	2022-11-21	субота						

Рисунок 3.11 - Визначення кількості унікальних користувачів за день

week_start_date WAU		Average DAU DAU/WAU		
1	2022-09-19	114	551	7
2	2022-09-26	99	494	7
3	2022-09-26	120	585	7
4	2022-10-03	127	621	7
5	2022-10-24	131	638	7
6	2022-11-07	139	668	7
7	2022-11-21	130	571	7
8	2022-11-28	127	590	7
9	2022-11-28	127	580	7
10	2022-12-05	132	614	7
11	2022-12-19	129	574	7
12	2022-09-08	91	483	7
13	2022-09-08	91	483	7
14	2022-09-12	187	521	7
15	2022-09-26	120	585	7
16	2022-10-24	131	638	7
17	2022-11-21	130	571	7
18	2022-11-21	130	571	7

Вилучити копії

Вибрано рядки: 44013, стовпці: 6

☐ Дати мають рядок для заголовка

Стовпці для аналізу

☐ Вибрати все:

☐ Стовпець A

☐ Стовпець B

☐ Стовпець C

☐ Стовпець D

☒ Стовпець E

☐ Стовпець F

Скасувати Вилучити копії

Рисунок 3.12 - Визначення кількості унікальних користувачів за тиждень

На рисунку 3.12 наведені результати визначення кількості унікальних користувачів за тиждень (перші 20 спостережень).

week_start_date WAU		Average DAU DAU/WAU		
2	2022-03-07	28	89	7
3	2022-03-14	32	130	7
4	2022-03-21	36	168	7
5	2022-03-28	40	209	7
6	2022-04-04	43	205	7
7	2022-04-11	43	229	7
8	2022-04-18	44	227	7
9	2022-04-25	43	214	7
10	2022-05-02	47	245	7
11	2022-05-09	52	269	7
12	2022-05-16	56	292	7
13	2022-05-23	61	321	7
14	2022-05-30	71	385	7
15	2022-06-06	70	392	7
16	2022-06-13	68	383	7
17	2022-06-20	73	389	7
18	2022-06-27	93	437	7
19	2022-07-04	96	451	7
20	2022-07-11	81	473	7

Рисунок 3.13 - Визначення середньої активності за кожен день тижня і співвідношення Average DAU до WAU.

На рисунку 3.13 наведені результати визначення середньої активності за кожен день тижня і співвідношення Average DAU до WAU (перші 20 спостережень).

Для виявлення особливостей користувачів за характеристиками наявними в наборі даних виконаємо аналіз і візуалізація структури користувачів ігор (рис.3.14, 3.15).



Рисунок 3.14 - Розподіл користувачів за мовою гри.



Рисунок 3.15 - Розподіл користувачів за ознакою *has_older_device_model*.

Для більш детального аналізу залученості користувачів проведемо аналіз за типами активності в іграх (рис. 3.16).

Розіб'ємо колонку *game_activity_name* на дві частини: назва гри та назву активності. Всього в назв активностей об'єднаємо в 5 типів активності (рис. 3.17).

	A	B	C	D	E	F	G	H	I	J	K	L
	Activity											
1	user_id	activity_date	game_activity_name	game_num	game_name	activity_type	language	total_sec	week_start	Activity	First_active	Activity_month_number
2	561796w22f1v521XZ	2022-09-28	game 3: Main campaign	game 3	Main campaign	Campaign	en	366	2022-09-19	9	9	0
3	561796w22f1v521XZ	2022-09-28	game 3: Special events	game 3	Special events	Events	en	48	2022-09-19	9	9	0
4	Q7VnYf54qF8f1omZL	2022-09-06	game 2: Battle pass challenges	game 2	Battle pass challenges	Challenges	uk	51	2022-09-05	9	7	2
5	Q7VnYf54qF8f1omZL	2022-09-28	game 2: Battle pass challenges	game 2	Battle pass challenges	Challenges	uk	848	2022-09-25	9	7	2
6	Q7VnYf54qF8f1omZL	2022-10-10	game 2: Battle pass challenges	game 2	Battle pass challenges	Challenges	uk	656	2022-10-10	10	7	3
7	Q7VnYf54qF8f1omZL	2022-10-26	game 2: Battle pass challenges	game 2	Battle pass challenges	Challenges	uk	558	2022-10-24	10	7	3
8	Q7VnYf54qF8f1omZL	2022-11-07	game 2: Battle pass challenges	game 2	Battle pass challenges	Challenges	uk	1248	2022-11-07	11	7	4
9	Q7VnYf54qF8f1omZL	2022-11-23	game 2: Battle pass challenges	game 2	Battle pass challenges	Challenges	uk	802	2022-11-21	11	7	4
10	Q7VnYf54qF8f1omZL	2022-11-28	game 2: Battle pass challenges	game 2	Battle pass challenges	Challenges	uk	128	2022-11-28	11	7	4
11	Q7VnYf54qF8f1omZL	2022-12-03	game 2: Battle pass challenges	game 2	Battle pass challenges	Challenges	uk	1081	2022-11-28	11	7	4
12	Q7VnYf54qF8f1omZL	2022-12-16	game 2: Battle pass challenges	game 2	Battle pass challenges	Challenges	uk	190	2022-12-12	12	7	5
13	Q7VnYf54qF8f1omZL	2022-12-26	game 2: Battle pass challenges	game 2	Battle pass challenges	Challenges	uk	1712	2022-12-19	12	7	5
14	Q7VnYf54qF8f1omZL	2022-09-08	game 2: Campaign	game 2	Campaign	Campaign	uk	17	2022-09-08	8	7	1
15	Q7VnYf54qF8f1omZL	2022-09-14	game 2: Campaign	game 2	Campaign	Campaign	uk	3	2022-09-08	8	7	1
16	Q7VnYf54qF8f1omZL	2022-09-12	game 2: Campaign	game 2	Campaign	Campaign	uk	28	2022-09-12	9	7	2
17	Q7VnYf54qF8f1omZL	2022-09-29	game 2: Campaign	game 2	Campaign	Campaign	uk	846	2022-09-25	9	7	2
18	Q7VnYf54qF8f1omZL	2022-10-28	game 2: Campaign	game 2	Campaign	Campaign	uk	308	2022-10-24	10	7	3
19	Q7VnYf54qF8f1omZL	2022-10-28	game 2: Campaign	game 2	Campaign	Campaign	uk	4	2022-10-24	10	7	3
20	Q7VnYf54qF8f1omZL	2022-11-25	game 2: Campaign	game 2	Campaign	Campaign	uk	157	2022-11-21	11	7	4
21	Q7VnYf54qF8f1omZL	2022-11-26	game 2: Campaign	game 2	Campaign	Campaign	uk	523	2022-11-21	11	7	4

Рисунок 3.16. - Виокремлення типу активності користувачів

	A	B
	activity_name	activity_type
2	Main campaign	Campaign
3	Special events	Events
4	Battle pass challenges	Challenges
5	Campaign	Campaign
6	Special events and challenges	Events
7	Mini-games	Mini-games
8	Clan activities	Clan
9	Clan multiplayer	Clan

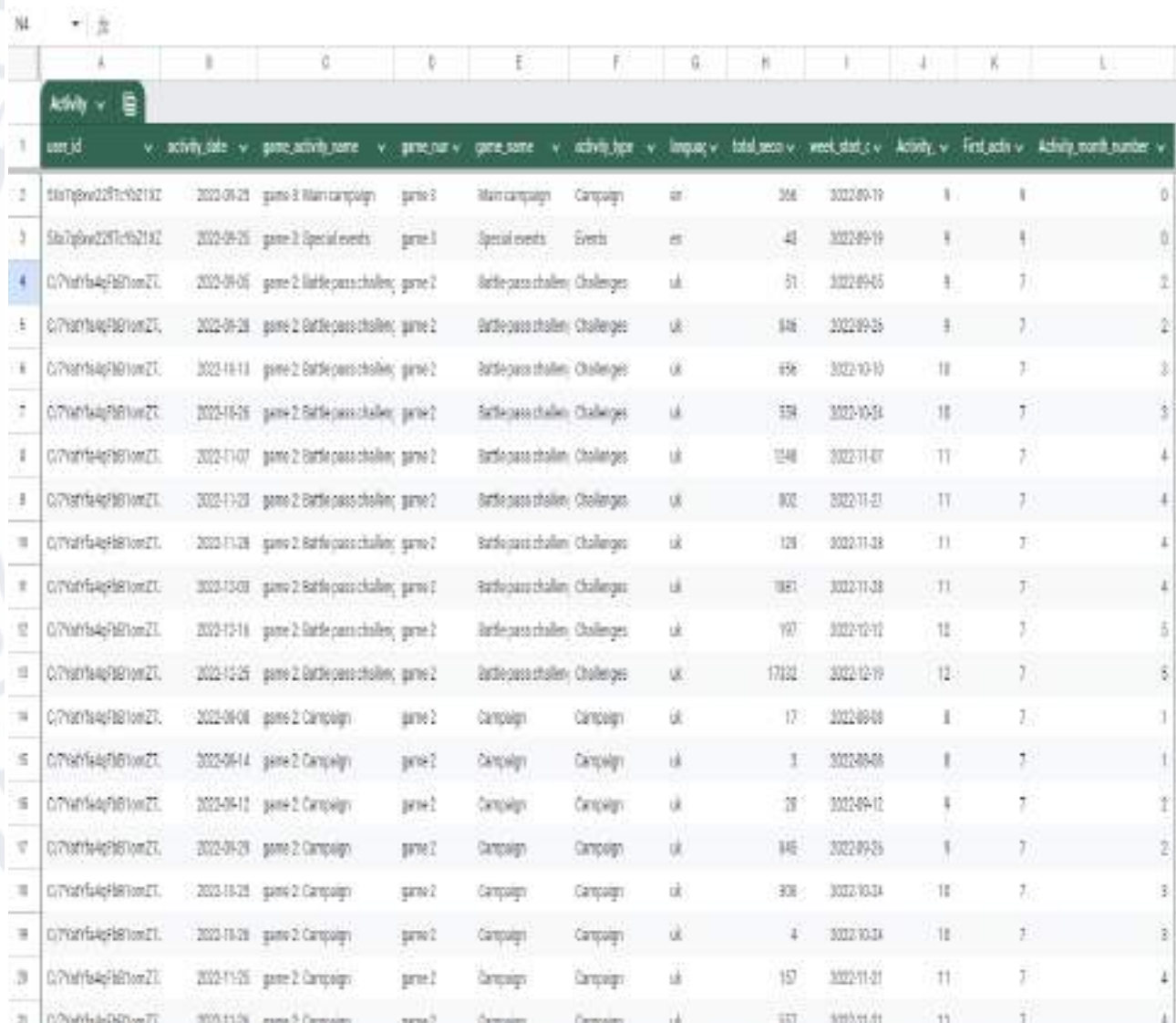
Рисунок 3.17- Визначення типу активності користувачів

Зробимо наступні перетворення даних: створимо три колонки, що є похідними від колонки activity_date (рис. 3.18):

a) Activity month — місяць активності, тобто місяць, до якого входить activity_date.

b) First activity month— перший місяць активності для кожного користувача.

c) Activity month number — номер місяця активності. Тобто, скільки місяців пройшло від First activity month до Activity month.



	A	B	C	D	E	F	G	H	I	J	K	L
	Activity											
1	user_id	activity_date	game_activity_name	game_name	game_type	activity_type	impact	total_score	week_start	Activity	First activity	Activity month number
2	56a79b0e227c1b2142	2022-09-29	game 3 Main campaign	game 3	Main campaign	Campaign	en	366	2022-09-19	9	9	0
3	56a79b0e227c1b2142	2022-09-26	game 3 Special events	game 3	Special events	Events	en	40	2022-09-19	9	9	0
4	079a7fa49f8f10e2f	2022-09-05	game 2 Battle pass challenge	game 2	Battle pass challenge	Challenges	uk	51	2022-09-05	9	7	2
5	079a7fa49f8f10e2f	2022-09-28	game 2 Battle pass challenge	game 2	Battle pass challenge	Challenges	uk	846	2022-09-26	9	7	2
6	079a7fa49f8f10e2f	2022-10-13	game 2 Battle pass challenge	game 2	Battle pass challenge	Challenges	uk	656	2022-10-10	10	7	3
7	079a7fa49f8f10e2f	2022-10-26	game 2 Battle pass challenge	game 2	Battle pass challenge	Challenges	uk	559	2022-10-24	10	7	3
8	079a7fa49f8f10e2f	2022-11-07	game 2 Battle pass challenge	game 2	Battle pass challenge	Challenges	uk	1248	2022-11-07	11	7	4
9	079a7fa49f8f10e2f	2022-11-29	game 2 Battle pass challenge	game 2	Battle pass challenge	Challenges	uk	802	2022-11-27	11	7	4
10	079a7fa49f8f10e2f	2022-11-28	game 2 Battle pass challenge	game 2	Battle pass challenge	Challenges	uk	128	2022-11-28	11	7	4
11	079a7fa49f8f10e2f	2022-12-03	game 2 Battle pass challenge	game 2	Battle pass challenge	Challenges	uk	1881	2022-11-28	11	7	4
12	079a7fa49f8f10e2f	2022-12-16	game 2 Battle pass challenge	game 2	Battle pass challenge	Challenges	uk	707	2022-12-12	12	7	5
13	079a7fa49f8f10e2f	2022-12-26	game 2 Battle pass challenge	game 2	Battle pass challenge	Challenges	uk	17132	2022-12-19	12	7	5
14	079a7fa49f8f10e2f	2022-09-08	game 2 Campaign	game 2	Campaign	Campaign	uk	17	2022-09-08	8	7	1
15	079a7fa49f8f10e2f	2022-09-14	game 2 Campaign	game 2	Campaign	Campaign	uk	3	2022-09-08	8	7	1
16	079a7fa49f8f10e2f	2022-09-12	game 2 Campaign	game 2	Campaign	Campaign	uk	28	2022-09-12	9	7	2
17	079a7fa49f8f10e2f	2022-09-29	game 2 Campaign	game 2	Campaign	Campaign	uk	846	2022-09-26	9	7	2
18	079a7fa49f8f10e2f	2022-10-29	game 2 Campaign	game 2	Campaign	Campaign	uk	308	2022-10-24	10	7	3
19	079a7fa49f8f10e2f	2022-10-28	game 2 Campaign	game 2	Campaign	Campaign	uk	4	2022-10-24	10	7	3
20	079a7fa49f8f10e2f	2022-11-26	game 2 Campaign	game 2	Campaign	Campaign	uk	157	2022-11-21	11	7	4
21	079a7fa49f8f10e2f	2022-11-06	game 2 Campaign	game 2	Campaign	Campaign	uk	557	2022-11-01	11	7	4

Рисунок 3.18 - Перетворення даних активності користувачів

Узагальнимо дані про активність користувачів. Виведемо Activity month number та як значення — кількість унікальних користувачів (рис. 3.18) та візуалізуємо (рис. 3.19) і виконаємо візуалізацію (рис. 3.20).

	A	B
1	Activity_month_number	COUNTUNIQUE з user_id
2	0	216
3	1	184
4	2	165
5	3	147
6	4	115
7	5	84
8	6	66
9	7	58
10	8	40
11	9	26
12	10	14

Рисунок 3.19 - Узагальнені дані за кількістю місяців активності користувачів

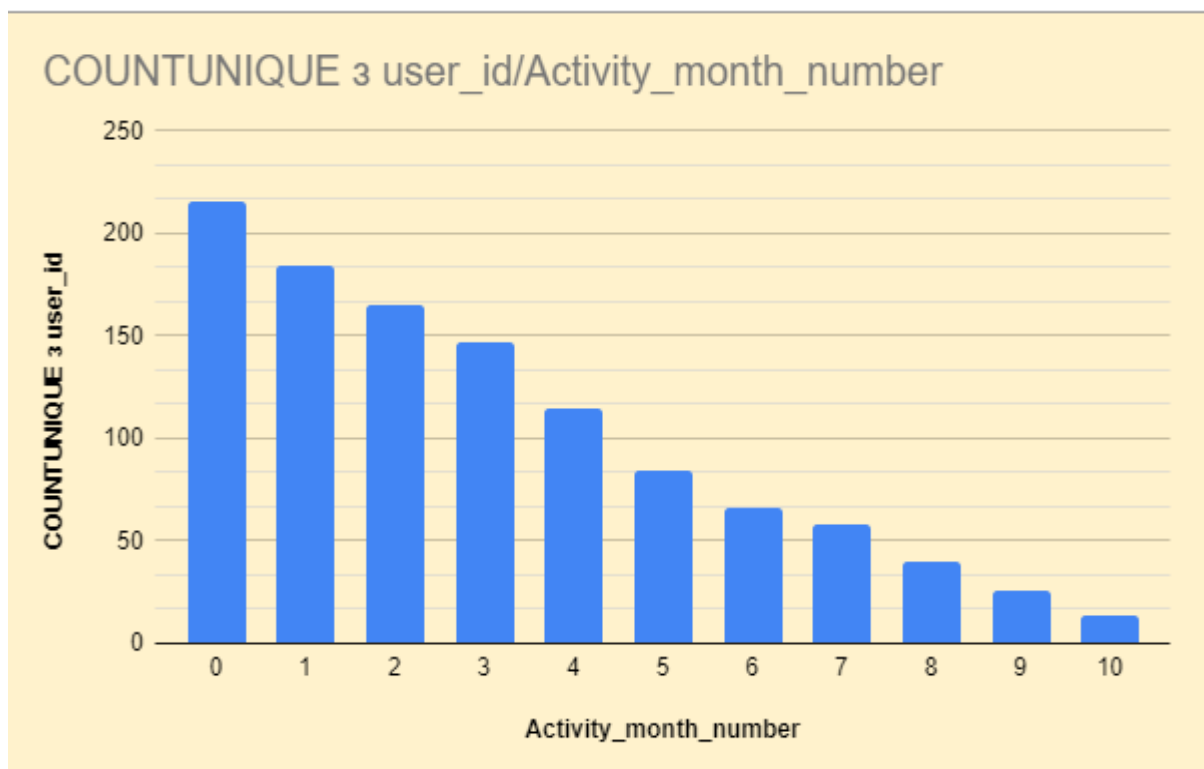


Рисунок 3.20 - Діаграма кількості місяців активності користувачів

3.2. Програмна реалізація інтелектуальної системи аналізу ефективності реклами.

На наступному кроці виконаємо прогноз показників Average DAU та WAU на наступні 20 днів і тижнів (рис. 3.21) і візуалізуємо данні (рис. 3.22, 3.23):

C45 fx =ROUND(FORECAST(A45;C\$2:C\$44;A\$2:A\$44);0)

	A	B	C	D	E
	Таблиця2				
1	с	Average DAU	WAU	DAU/WAU	Стовпець 1
40	2022-11-28	84	127	66,37%	листопада
41	2022-12-05	88	129	68,22%	грудня
42	2022-12-12	88	132	66,45%	грудня
43	2022-12-19	82	119	68,91%	грудня
44	2022-12-26	69	107	64,64%	грудня
45	2023-01-02	100	147	68,03%	січня
46	2023-01-09	102	149	68,46%	січня
47	2023-01-16	104	152	68,42%	січня
48	2023-01-23	106	155	68,39%	січня
49	2023-01-30	107	157	68,15%	січня
50	2023-02-06	109	160	68,13%	лютого
51	2023-02-13	111	162	68,52%	лютого
52	2023-02-20	113	165	68,48%	лютого
53	2023-02-27	114	168	67,86%	лютого
54	2023-03-06	116	170	68,24%	березня
55	2023-03-13	118	173	68,21%	березня
56	2023-03-20	120	176	68,18%	березня
57	2023-03-27	121	178	67,98%	березня
58	2023-04-03	123	181	67,96%	квітня
59	2023-04-10	125	183	68,31%	квітня
60	2023-04-17	126	186	67,74%	квітня
61	2023-04-24	128	189	67,72%	квітня
62	2023-05-01	130	191	68,06%	травня
63	2023-05-08	132	194	68,04%	травня
64	2023-05-15	133	197	67,51%	травня

Рисунок 3.21 - Прогнозування показників Average DAU до WAU.



Рисунок 3.22 - Активності користувачів



Рисунок 3.23 - Прогноз тижневої активності користувачів

Для аналізу динаміки активності користувачів виконаємо когортний аналіз. Когорти у цьому побудуємо за найбільш розповсюдженим шаблоном — використовуючи дату першої активності (точніше, місяць першої активності): В результаті отримаємо дані про кількість унікальних користувачів в кожній когорті протягом періоду активності (рис. 3.24).

	A	B	C	D	E	F	G	H	I	J	K	L	M
13	COUNTUNIQUE(рис. Activity_month_number)												
14	First activity month	1	1	2	3	4	5	6	7	8	9	10	11
15	2	19	17	18	18	18	17	15	15	14	13	14	19
16	3	28	27	24	27	26	20	19	19	21	13		28
17	4	8	7	8	8	7	6	5	4	5			9
18	5	25	25	25	22	19	21	20	20				25
19	6	15	16	8	8	8	8	7					16
20	7	21	15	14	15	15	12						21
21	8	38	27	25	27	22							38
22	9	25	21	22	21								25
23	10	18	18	17									18
24	11	9	9										9
25	12	8											8
26	Загальний результат	216	184	165	147	115	84	66	58	40	26	14	216

Рисунок 3.24 - Когортний аналіз динаміки активності користувачів

Для аналізу динаміки залученості користувачів у відносних показниках створимо три зрізи (рис. 3.25 – 3.28): назва гри, тип активності та мова користувача для показника retention rate — це кількість унікальних користувачів у відповідний місяць поділена на кількість унікальних користувачів у місяць залучення користувачів.

Результати фільтрації за зрізами представлені на рисунках 3.26- 3.28.

		0	1	2	3	4	5	6	7	8	9	10
2	100.00%	88.47%	94.74%	94.74%	94.74%	89.47%	78.85%	78.85%	77.88%	68.42%	73.68%	
3	100.00%	96.43%	85.71%	96.43%	92.86%	71.43%	67.86%	67.86%	75.89%	46.43%		
4	100.00%	77.78%	88.89%	88.89%	77.78%	66.67%	65.56%	44.44%	55.56%			
5	100.00%	100.00%	100.00%	88.89%	76.67%	84.00%	80.00%	88.00%				
6	100.00%	100.00%	50.00%	50.00%	50.00%	50.00%	43.75%					
7	100.00%	65.22%	68.87%	68.57%	65.22%	52.17%						
8	100.00%	75.90%	88.56%	75.89%	61.11%							
9	100.00%	92.00%	88.00%	84.89%								
10	100.00%	100.00%	94.44%									
11	100.00%	100.00%										
12	100.00%											

Рисунок 3.25. - Зрізи показника retention rate активності користувачів

		1	2	3	4	5	6	7	8	9	10
4%	88.66%	94.44%	94.44%	83.33%	83.33%	83.33%	77.78%	88.89%	72.22%		
6%	86.71%	82.14%	75.00%	71.43%	67.86%	64.29%	75.89%	46.43%			
8%	87.68%	87.50%	87.50%	75.00%	62.50%	50.00%	62.50%				
10%	88.00%	84.00%	76.00%	84.00%	80.00%	80.00%					
12%	88.68%	80.00%	80.00%	88.68%	70.00%						
14%	68.68%	70.00%	70.00%	68.68%							
16%	77.42%	87.74%	50.00%								
18%	111.76%	105.89%									
20%	83.33%										

Рисунок 3.26 - Зрізи показника retention rate активності користувачів
відфільтровані за двома іграми

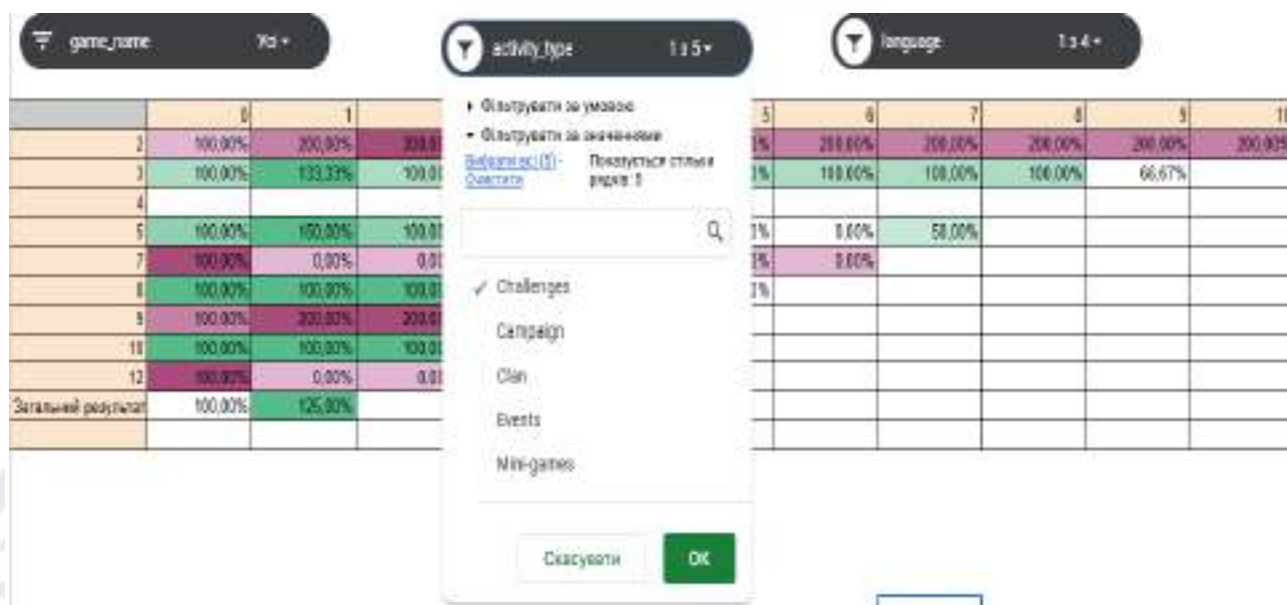


Рисунок 3.27 - Зрізи показника retention rate активності користувачів відфільтровані за типом активності



Рисунок 3.28 - Зрізи показника retention rate активності користувачів відфільтровані за мовою гри

Отримані результати аналізу в залежності від бажаних показників, термінів залученості і аудиторії дозволяють приймати рішення щодо рекламної і тарифної політики кампанії.

Постановка задачі моделювання і оцінювання. Розглянемо набір даних *Active_useres*, в якому введемо залежну змінну Y - *Retention*, яка характеризує ймовірність утримання користувача більше певного часу і представлена двома категоріями - *yes* і *no*.

$$Retention = \begin{cases} 1, total_sekonds \geq T_0 \\ 0, total_sekonds < T_0 \end{cases}$$

Логістична регресія моделює ймовірність приналежності Y до тієї чи іншої категорії. Для набору даних *Active_useres* логістична регресія моделює ймовірність утримання користувача більше певного часу. Ймовірність утримання для деякого значення *age* можна записати як

$$\Pr(Retention = yes|age).$$

Значення $\Pr(Retention = yes|age)$, які ми скорочено позначаємо як $p(age)$, будуть змінюватися від 0 до 1. Далі для заданого значення *age* можна передбачити ймовірний статус утримання користувача.

Наприклад, для клієнта з $p(age) > 0,5$, можна було б передбачити $Retention = yes$. Якщо компанія бажає бути консервативною в прогнозі того, які користувачі повертаються до користування програмним продуктом (грою), то в якості альтернативи вона може використовувати більш низьке порогове значення ймовірності, наприклад $p(age) > 0,1$.

Отримаємо описову статистику показників набору даних, які можуть бути використані як фактори моделі (табл.3.2).

Побудова моделі логістичної регресії. Для моделювання зв'язку між $p(X) = \Pr(Y = 1|X)$ та X використаємо кодування відгуку у вигляді 0 та 1.

$$Retention = \begin{cases} 1, total_sekonds \geq T_0 \\ 0, total_sekonds < T_0 \end{cases}$$

де $T_0 = Median(Total_seconds) = 310055$

Табл.3.2 - Описова статистика показників набору даних *Active_useres*

	age	First_activity_ month	Total_ seconds	Total_ minutes	Total_ hours	Total_ day
Mean	28,00	6,5945	1122794,889	18713,25	311,89	13,56
Standard Error	0,76	0,1928	158650,5137	2644,18	44,07	1,92
Median	25	7	310055	5167,58	86,13	3,74
Mode	17	8	2736757	45612,62	760,21	33,05
Standard Deviation	11,26	2,8401	2337068,003	38951,13	649,19	28,23
Sample Variance	126,75	8,0663	5461886852393	91031447539,89	1517190792,33	65964817,06
Kurtosis	3,38	-1,0304	59,42858256	0,99	0,02	0,00
Skewness	1,43	-0,0526	6,344595767	0,11	0,00	0,00
Range	69	10	25906146	431769,10	7196,15	312,88
Minimum	14	2	1	0,02	0,00	0,00
Maximum	83	12	25906147	431769,12	7196,15	312,88
Sum	6075	1431	243646491	4060774,85	67679,58	2942,59
Count	217	217	217	3,62	0,06	0,00
Largest(1)	83	12	25906147	431769,12	7196,15	312,88
Smallest(1)	14	2	1	0,02	0,00	0,00
Confidence Level(95%)	1,5063	0,38001	312701,3401	5211,69	86,86	3,78

Програмний код побудови і оцінки логістичної регресії реалізований на мові Python (Додаток А).

У табл. 3.3 наведені оцінки коефіцієнтів і супутня їм інформація, отримана в результаті підгонки логістичної регресійної моделі за даними *Active_useres* з метою передбачення ймовірності $\Pr(Retention = yes|age)$ по *age*.

Таблиця 3.3 – Регресійна статистика логістичної моделі ймовірності $\Pr(Retention = yes|age)$ по *age* при $T_0 = Median(Total_seconds)$

	Коефіцієнт	Ст. помилка	р- значення	відношення шансів	Нижня границя 95%	Верхня границя 95%
вільний член	2,2975	0,4655	0,0000008	9,9494	3,9958	24,7739
<i>Age</i>	-0.0843	0,0166	0,0000004	0,9192	0,8897	0,9496

Побудована логістична модель, яка прогнозує ймовірність, що кількість часу (total_seconds), яку проведе новий користувач гри буде більше медіанного значення, поточних користувачів.

Виходячи з результату регресійного аналізу, можна зробити наступні висновки.

1. Перевірка значущості моделі

- P-value для змінної Age: дуже мале значення (0.0000004) свідчить про те, що цей коефіцієнт є статистично значущим на рівні значущості 0.05, тобто змінна Age дійсно впливає на залежну змінну.

2. Інтерпретація коефіцієнтів

- Коефіцієнт вільного члена становить 2.2975, що свідчить про базове значення залежної змінної, коли незалежна змінна Age дорівнює нулю.

- Відношення шансів для вільного члена дорівнює 9.9494. Це значення вказує на те, що в базовому стані (коли Age = 0) шанси на настання події зростають приблизно в 10 разів.

- Коефіцієнт змінної Age дорівнює -0.0843, що означає негативний вплив цієї змінної на залежну змінну.

- Відношення шансів для Age дорівнює 0.9192, що менше одиниці. Це свідчить про те, що з кожним збільшенням змінної Age шанси на настання події зменшуються на приблизно 8% ($1 - 0.9192 \approx 0.081$ або 8%).

3. Інтервали довіри

- 95% довірчий інтервал для коефіцієнта Age: від 0.8897 до 0.9496. Оскільки обидва значення нижчі за 1, це додатково підтверджує, що Age зменшує шанси на настання події.

4. Висновки щодо змінної Age

- Оскільки коефіцієнт для Age є негативним і статистично значущим, можна зробити висновок, що зі зростанням значення Age зменшуються шанси на настання події.

- Ймовірно, Age має зворотний зв'язок з імовірністю цільової події в моделі.

Таким чином, змінна Age має статистично значущий негативний вплив на залежну змінну. Шанси на настання події зменшуються приблизно на 8% з кожним збільшенням значення Age.

Прогнозування. Маючи оцінки коефіцієнтів, обчислюємо вірогідність утримання користувачів певного віку. Наприклад, використовуючи оцінки коефіцієнтів з табл. 3.1, ми передбачаємо, що ймовірність утримання користувача віком в 20 років становить

$$\hat{p}(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} = \frac{e^{2,2975 - 0,0843 \cdot 20}}{1 + e^{2,2975 - 0,0843 \cdot 20}} = 0,65,$$

Тобто ймовірність утримання користувача віком в 20 років більше медіанного значення для поточних користувачів $T_0 = \text{Median}(\text{Total_seconds}) = 310055$ дорівнює 65%.

Таблиця 3.4 - Ймовірність утримання користувача більше медіанного значення для поточних користувачів

age		Ймовірність утримання користувача
Median	25	0,5469
Mode	17	0,7032
Minimum	14	0,7532
Maximum	83	0,0090

Отримані результати є основою для розробки маркетингової політики наступних рекламних компаній.

Алгоритм інформаційної технології оцінки ефективності реклами

Вхідні дані: набори даних про активність користувачів ігор за результатами рекламних кампаній.

Вихідні дані: прогноз денної, тижневої активності, динаміка показників активності по когортам і зрізам, прогноз ймовірності утримання користувачів в залежності від віку і часу гри.

Алгоритм інформаційної технології оцінки ефективності реклами.

- Крок 1. Завантаження набору даних для аналізу. Визначення типів даних.
- Крок 2. Розрахунок узагальнених значень показників витрат по іграм і кількості користувачів.
- Крок 3. Аналіз віку користувачів за характеристиками:
- Крок 4. Визначення середньої активності за кожен день тижня і співвідношення Average DAU до WAU та stickiness користувачів.
- Крок 5. Прогноз показників Average DAU та WAU.
- Крок 6. Аналіз динаміки активності користувачів на основі показника retention rate — це кількість унікальних користувачів у відповідний місяць поділена на кількість унікальних користувачів у місяць залучення користувачів.
- Крок 7. Когортний аналіз динаміки активності користувачів
- Крок 8. Зрізи показника retention rate активності користувачів
- Крок 9. Побудова логістичної моделі, яка прогнозує ймовірність, що кількість часу (total_seconds), яку проведе новий користувач гри буде більше заданого (зокрема, медіанного) значення, поточних користувачів.
- Крок 10. Отримання прогнозу утримання користувачів для різних значень віку і часу утримання.

3.3. Висновки за розділом 3

1. Виконано збір і попередня обробка даних рекламних кампаній (на прикладі комп'ютерних ігор).
2. Створена програмна реалізація інтелектуальної системи аналізу ефективності реклами.

ВИСНОВКИ

1. Визначений сучасний стан технологій і тенденції рекламної діяльності на інтернет-ресурсах.
2. Проведений аналіз існуючих інтелектуальних технологій аналізу ефективності реклами.
3. Виконана постановка завдання, збір і попередня обробку даних рекламних кампаній, очищення та організацію даних.
4. Виконаний аналіз і прогноз динаміки активності користувачів на основі інтелектуальних методів і моделей.
5. Розроблена модель оцінки ймовірності утримання користувачів в залежності від їх характеристик;
6. Виконана програмна реалізацію інтелектуальної технології аналізу ефективності реклами.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ

1. Романенко Л. Ф. Інтернет-реклама: її види та інструменти. Соціально-економічні проблеми сучасного періоду України. 2017. Вип. 1. С. 71-73. URL: http://nbuv.gov.ua/UJRN/sepspu_2017_1_21. (дата звернення: 14.05.2024).
2. Definitions of Marketing. *American Marketing Association*: веб-сайт. URL: <https://www.ama.org/the-definition-of-marketing-what-is-marketing/> (дата звернення: 14.05.2024).
3. Concept_of_advertising. URL: https://wikieducator.org/CONCEPT_OF_ADVERTISING (дата звернення: 14.05.2024).
4. Hardy J. The History of Marketing: From Trade to Tech. *History Cooperative*. 2016. URL: <https://historycooperative.org/the-evolution-of-marketing-from-trade-to-tech/> (дата звернення: 14.05.2024).
5. Глобальні рекламні тренди від WARC (<https://vrk.org.ua/news-events/2024/global-ad-trends-warc.html>)
6. Об'єм рекламно-комунікаційного ринку України 2023 і прогноз об'ємів ринку 2024. URL: <https://vrk.org.ua/news-events/2023/ad-volume-2024.html> (дата звернення: 14.05.2024).
7. Липчук В.В., Дудяк Р.П., Бугіль С.Я., Янишин Я.С. Маркетинг: навч. посіб. за заг. ред. В. В. Липчука. 2-ге вид. стереотипне, Львів: «Магнолія-плюс», 2016. 456 с.
8. Цифровий маркетинг – модель маркетингу ХХІ сторіччя: колективна монографія. Одеський національний політехнічний університет. Одеса: Астропринт, 2017. 292 с.
9. Окландер М.А., Романенко О.О. Специфічні відмінності цифрового маркетингу від інтернет-маркетингу. *Економічний вісник Національного технічного університету України "Київський політехнічний інститут"*. 2015 № 12 URL: <http://surl.li/cqevdi>
10. Соколова Л. В., Мазура А. Ю. Можливості Інтернет-маркетингу в сучасному інформаційному полі. *Сучасний рух науки: тези доп. VIII міжнар. наук.-практ. інтернет-конференції, 3-4 жовтня 2019 р. Дніпро, 2019. Т. 2.С.*

451-455. URL: <http://www.wayscience.com/wp-content/uploads/2019/10/TOM-2>

11. Vincent Aguirre. Digital Marketing 101: The Target Audience. *Become Distinct*. 2020. URL: <https://becomedistinct.com/digital-marketing-101-the-target-audience/> (дата звернення: 14.05.2024)

12. Joshi S., Bhatia S., RIIIkar K., Pall H. Customer experience and associated customer behavior in end user devices and technologies (smartphones, mobile internet, mobile financial services). *International Journal of High Performance Computing and Networking*. 2017 Vol. 10, No.1-2. pp. 118-126. URL: <https://www.inderscienceonline.com/doi/abs/10.1504/IJHPCN.2017.083209> (дата звернення: 14.05.2024)

13. Wolny J., CharoensuksIII N. Mapping customer journeys in multichannel decision-making. *Journal of Direct, Data and Digital Marketing Practice*. 2014. Vol. 15, No. 4. pp. 317-326. URL: <https://link.springer.com/article/10.1057/dddmp.2014.24> (дата звернення: 14.05.2024)

14. Ashman R., Solomon M. R., Wolny J. An old model for a new age: Consumer decision making in participatory digital culture. *Journal of Customer Behaviour*. 2015. Vol 14, No. 2, pp.127-146. URL: <https://www.ingentaconnect.com/content/westburn/jcb/2015/00000014/00000002/art00004> (дата звернення: 14.05.2024)

15. Solomon M., Bamossy G., Askegaard S., & Hogg, M. (2014). Consumer behaviour: A European perspective, 5th edition. 2014. Harlow: Pearson Higher Education

16. Bettman J.R., Park C.W. Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of Consumer Research*, 1980. Vol. 7, No. 3, pp. 234-238. URL: <https://academic.oup.com/jcr/article-abstract/7/3/234/1820883> (дата звернення: 14.05.2024)

17. Solomon M.R. Consumer behaviour: buying, having, and being, 11th edition. 2015. Upper Saddle River: Pearson Education.

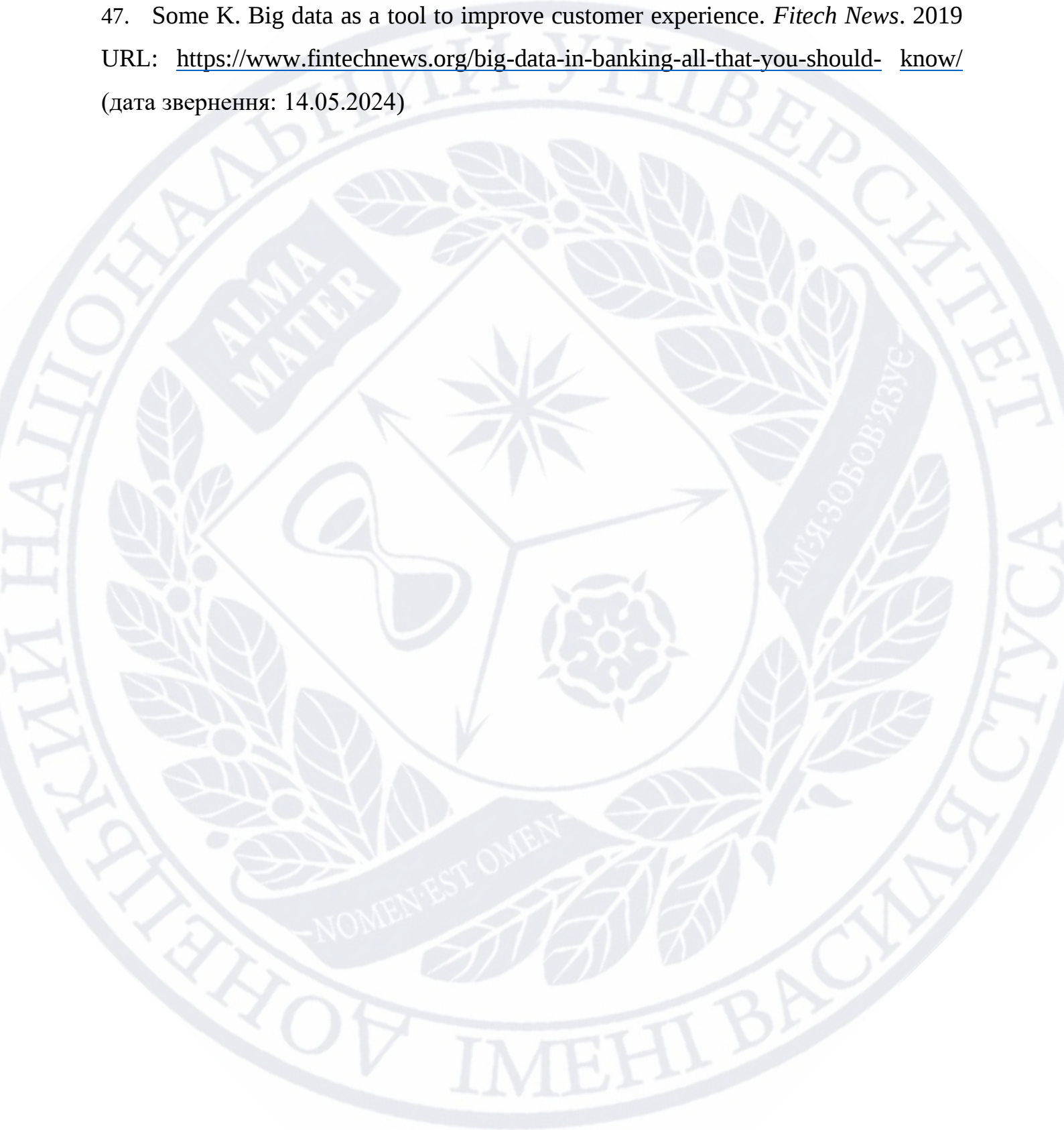
20. Simonson I., Rosen E. Absolute Value: What Really Influences Customers in the Age of (Nearly) Perfect Information. 2014. New York: HarperCollins.
21. Reddy N. S. Particle Swarm Optimized Neural Network for Predicting Customer Behaviour in Digital Marketing. *International Conference on Inventive Research in Computing Applications*. 2020. Vol. 2. pp. 70-77. URL: https://www.aconf.org/conf_174919.html (дата звернення: 18.11.2020)
22. Cui Y., Tobossi R., Vigouroux O. Modelling customer online behaviours with neural networks: applications to conversion prediction and advertising retargeting, *arXivLabs*, 2018. Vol.1. URL: <https://arxiv.org/pdf/1804.07669.pdf> (дата звернення: 14.05.2024)
23. Miklosik A., Kuchta M., Evans N., Zak S. Towards the Adoption of Machine Learning-Based Analytical Tools in Digital Marketing. *IEEE Access*, 2019. Vol. 7, pp. 85705-85718. URL: <https://ieeexplore.ieee.org/abstract/document/8746184> (дата звернення: 14.05.2024)
24. Mackey T., Kalyanam J., Klugman J., Kuzmenko E. Solution to Detect, Classify, and Report Illicit Online Marketing and Sales of Controlled Substances via Twitter: Using Machine Learning and Web Forensics to Combat Digital Opioid Access, *Journal of Medical Internet Research* 2018, Vol. 20, No. 4. URL: <https://www.jmir.org/2018/4/e10029/> (дата звернення: 14.05.2024)
25. Kim J., Kang S., Hoon Lee K., Evolution of digital marketing communication: Bibliometric analysis and network visualization from key articles, *Journal of Business Research*, 2019. Vol. 1. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0148296319305727> (дата звернення: 14.05.2024)
26. Narkhede S. Understanding AUC-ROC Curve. *Towards Data Science*. 2018 URL: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5> (дата звернення: 14.05.2024)
27. Khodabandehlou S., Rahman M. Z. Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. *Journal of Systems and Information Technology*, 2017. Vol 1. URL:

- <https://www.emerald.com/insight/content/doi/10.1108/JSIT-10-2016-0061/full/html>
(дата звернення: 14.05.2024)
28. Old Second Bank Uses Machine Learning to Predict Customer Behavior & Improve Marketing Effectiveness. *LBDO Digital* URL: <https://www.bdo.com/digital/client-success/old-second-bank-uses-machine-learning-to-predict-c> (дата звернення: 14.05.2024)
29. Сайт “Ритейл в Україні” [Ua-RetIII.com] <https://ua-retIII.com/2021/09/sposobi-ocinki-efektivnosti-internet-reklami/> (дата звернення: 14.05.2024)
30. Яловега Н. (2023). Аналіз ефективності інтернет-реклами: нові методики вимірювання та оцінки результатів. *Економіка та Суспільство*, 57. <https://doi.org/10.32782/2524-0072/2023-57-52>
31. Резнік Н., Луцій О. (2024). Аналіз ефективності застосування інструментів інтернет маркетингу в маркетингову діяльність підприємств аграрного сектору. *Економіка та Суспільство*, 60. <https://doi.org/10.32782/2524-0072/2024-60-145>
32. Кобернюк С., Струнгар А., Завгородня Л. (2024). Аналіз ролі та ефективності використання штучного інтелекту у вдосконаленні персоналізованої реклами та взаємодії з аудиторією. *Економіка та Суспільство*, 61. <https://doi.org/10.32782/2524-0072/2024-61-112>
33. Булах Т.Д. (2021). До проблеми класифікації інтернет-реклами. *The european development trends in journalism, pr, media and communication*. <https://doi.org/10.30525/978-9934-26-042-1-28>
34. Юзов А.М. (2022). Оцінка ефективності пошукової реклами. *Efektivna Ekonomika*, 8. <https://doi.org/10.32702/2307-2105.2022.8.38>
35. Мазур І. І., Стаднік Г. В. (2023). Ефективність застосування інструментів інтернет-реклами в електронній комерції. *Efektivna Ekonomika*, 2. <https://doi.org/10.32702/2307-2105.2023.2.18>
36. Шикіна О., Нечева Н. (2023). Аналіз інструментів інтернет-маркетингу національних готельних мереж України. *Економіка Та Суспільство*, 47. <https://doi.org/10.32782/2524-0072/2023-47-82>

37. Гадецька З. (2022). Оцінка ефективності застосування сучасних інтернет-комунікацій та технологій цифрового маркетингу для оптимізації роботи підприємства. Економіка та суспільство, 40. <https://doi.org/10.32782/2524-0072/2022-40-44>
38. Гарбар Ж. В., Гонтарук Я. В. (2022). Розвиток реклами в мережі інтернет для просування послуг в сфері екотуризму. Індустрія Туризму і Гостинності в Центральній Та Східній Європі, 6, 5–11. <https://doi.org/10.32782/tourismhospcee-6-11>
39. Жуковський Д., Лозовська Л. (2023). Аналіз використання методів машинного навчання в аналітиці показників інтернет ресурсів. Сталий Розвиток Економіки, 2(47), 65–69. <https://doi.org/10.32782/2308-1988/2023-47-9>
40. Нестеренко В., Сідельнікова В. (2021). Аналіз маркетингової діяльності підприємства: актуальні аспекти оцінювання ефективності. Проблеми і Перспективи Розвитку Підприємництва, 26, 89. <https://doi.org/10.30977/ppb.2226-8820.2021.26.89>
41. Мальчик М. В., Адасюк І. П. Реклама в інтернеті: теоретичний аналіз та особливості URL: <https://science.lpnu.ua/sites/default/files/journal-paper/2021/may/23590/210488verstka-77-87.pdf> (дата звернення: 24.05.2022).
42. IT-індустрія України. URL: <http://www.itukrIIIine.org.ua/ru/it> (дата звернення: 21.11.2019).
43. Сайт холдингу KM Core. URL: <http://www.kmcore.com> (дата звернення: 21.11.2019).
44. Сайт компанії KM Ware. URL: www.km-ware.com (дата звернення: 21.11.2019).
45. Методи оцінки ефективності реклами. URL: <https://memosales.ru/reklama/ocenka-effektivnosti-kampanii> (дата звернення: 14.05.2024).
46. Vieira V. A. Inácio Severo de Almeida M., Agnihotri R., Arunachalam S. In pursuit of an effective B2B digital marketing strategy in an emerging market. *Journal of the Academy of Marketing Science*, 2019. Vol. 47, pp. 1085-1108. URL:

<https://link.springer.com/article/10.1007/s11747-019-00687-1> (дата звернення: 14.05.2024)

47. Some K. Big data as a tool to improve customer experience. *Fitech News*. 2019
URL: <https://www.fitechnews.org/big-data-in-banking-all-that-you-should-know/>
(дата звернення: 14.05.2024)





ДОДАТКИ

Програмний код

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Підключення даних
activity = pd.read_csv('activity.csv')
active_users = pd.read_csv('active_users.csv')

# Попередній аналіз
# Визначаємо загальні фінансові показники
# Total revenue
total_revenue = activity.groupby('game_id')['revenue'].sum()

# Paid users count
paid_users_count = activity[activity['paid'] ==
True].groupby('game_id')['user_id'].nunique()
# CR to Paid (Conversion Rate)
total_users = activity.groupby('game_id')['user_id'].nunique()
cr_to_paid = (paid_users_count / total_users) * 100

# ARPPU (Average Revenue Per Paying User)
arppu = total_revenue / paid_users_count

# Average Age, Median Age, Min Age, Max Age (paid users)
paid_users = active_users[active_users['user_id'].isin(activity[activity['paid'] ==
True]['user_id'])]
average_age_paid = paid_users['age'].mean()
median_age_paid = paid_users['age'].median()
min_age_paid = paid_users['age'].min()
max_age_paid = paid_users['age'].max()

# Характеристики віку
average_age = active_users['age'].mean()
```

```

std_age = active_users['age'].std()
median_age = active_users['age'].median()
iqr_age = active_users['age'].quantile(0.75) - active_users['age'].quantile(0.25)
percentile_10 = active_users['age'].quantile(0.1)
percentile_90 = active_users['age'].quantile(0.9)
# Видаляємо дублікатів користувачів
active_users = active_users.drop_duplicates()
# Розрахунок DAU, WAU, Stickiness
activity['week_start_date'] = pd.to_datetime(activity['activity_date']) -
pd.to_timedelta(pd.to_datetime(activity['activity_date']).dt.weekday, unit='d')
# DAU (Daily Active Users)
dau = activity.groupby('activity_date')['user_id'].nunique()
# WAU (Weekly Active Users)
wau = activity.groupby('week_start_date')['user_id'].nunique()
# Stickiness (DAU/WAU)
stickiness = (dau.mean() / wau.mean()) * 100
# Графіки
# Розподіл користувачів за мовою гри
language_distribution = active_users['language'].value_counts()
language_distribution.plot(kind='bar', title='Розподіл користувачів за мовою гри')
plt.show()
# Розподіл користувачів за ознакою has_older_device_model
device_model_distribution = active_users['has_older_device_model'].value_counts()
device_model_distribution.plot(kind='bar', title='Розподіл користувачів за ознакою
has_older_device_model')
plt.show()
# Перетворення даних
activity['activity_date'] = pd.to_datetime(activity['activity_date'])
activity['activity_month'] = activity['activity_date'].dt.month

```

```

activity['first_activity_month'] =
activity.groupby('user_id')['activity_date'].transform('min').dt.month
activity['activity_month_number'] = activity['activity_month'] -
activity['first_activity_month']
# Діаграма кількості місяців активності користувачів
activity_months_distribution =
activity.groupby('activity_month_number')['user_id'].nunique()
activity_months_distribution.plot(kind='line', title='Кількість місяців активності
користувачів')
plt.show()
# Виведення розрахованих значень
print("Total Revenue by Game:\n", total_revenue)
print("Paid Users Count by Game:\n", paid_users_count)
print("CR to Paid by Game (%):\n", cr_to_paid)
print("ARPPU by Game:\n", arppu)
print("Average Age (Paid Users):", average_age_paid)
print("Median Age (Paid Users):", median_age_paid)
print("Min Age (Paid Users):", min_age_paid)
print("Max Age (Paid Users):", max_age_paid)
print("Average Age (All Users):", average_age)
print("Standard Deviation of Age:", std_age)
print("Median Age (All Users):", median_age)
print("Interquartile Range of Age:", iqr_age)
print("10th Percentile of Age:", percentile_10)
print("90th Percentile of Age:", percentile_90)
print("DAU/WAU Stickiness Ratio (%)", stickiness)
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
import datetime
# Прогноз Average DAU на 20 днів

```

```

X = dau_reset[['days_since_start']]
y = dau_reset['user_id']
model = LinearRegression()
model.fit(X, y)
# Прогноз на наступні 20 днів
future_days = pd.DataFrame({'days_since_start':
                             range(dau_reset['days_since_start'].max() + 1,
                                     dau_reset['days_since_start'].max() + 21)})
future_dau = model.predict(future_days)
# Графік прогнозу Average DAU
plt.figure(figsize=(10, 5))
plt.plot(dau_reset['activity_date'], dau_reset['user_id'], label='Actual DAU')
plt.plot(dau_reset['activity_date'].max() +
          pd.to_timedelta(future_days['days_since_start'] - dau_reset['days_since_start'].max(),
                           unit='d'),
          future_dau, label='Predicted DAU', linestyle='--', color='red')
plt.title('Прогноз DAU на 20 днів')
plt.xlabel('Дата')
plt.ylabel('DAU')
plt.legend()
plt.show()
# Прогноз WAU
wau_reset = wau.reset_index()
wau_reset['week_start_date'] = pd.to_datetime(wau_reset['week_start_date'])
wau_reset['weeks_since_start'] = (wau_reset['week_start_date'] -
wau_reset['week_start_date'].min()).dt.days // 7
X_wau = wau_reset[['weeks_since_start']]
y_wau = wau_reset['user_id']
model_wau = LinearRegression()
model_wau.fit(X_wau, y_wau)

```

```

# Прогноз на наступні 20 тижнів
future_weeks = pd.DataFrame({'weeks_since_start':
range(wau_reset['weeks_since_start'].max() + 1,
wau_reset['weeks_since_start'].max() + 21)})

future_wau = model_wau.predict(future_weeks)
# Графік прогнозу WAU
plt.figure(figsize=(10, 5))
plt.plot(wau_reset['week_start_date'], wau_reset['user_id'], label='Actual WAU')
plt.plot(wau_reset['week_start_date'].max()
pd.to_timedelta(future_weeks['weeks_since_start']
wau_reset['weeks_since_start'].max(), unit='W'),
future_wau, label='Predicted WAU', linestyle='--', color='red')
plt.title('Прогноз WAU на 20 тижнів')
plt.xlabel('Дата')
plt.ylabel('WAU')
plt.legend()
plt.show()
# Розрахунок Retention Rate
activity['activity_month'] = activity['activity_date'].dt.to_period('M')
activity['cohort_month'] =
activity.groupby('user_id')['activity_date'].transform('min').dt.to_period('M')
retention = activity.groupby(['cohort_month',
'activity_month'])['user_id'].nunique().reset_index()
cohort_sizes = retention[retention['cohort_month'] ==
retention['activity_month']].set_index('cohort_month')['user_id']
retention['retention_rate'] = retention.apply(lambda x: x['user_id'] /
cohort_sizes[x['cohort_month']], axis=1)
# Візуалізація когортного аналізу
retention_pivot = retention.pivot(index='cohort_month', columns='activity_month',
values='retention_rate')

```

```

plt.figure(figsize=(12, 8))
plt.title('Когортний аналіз Retention Rate')
plt.xlabel('Місяць активності')
plt.ylabel('Місяць залучення')
sns.heatmap(retention_pivot, annot=True, fmt='.0%', cmap='Blues')
plt.show()

# Назва гри
game_retention = activity.groupby(['game_id', 'cohort_month',
'activity_month'])['user_id'].nunique().reset_index()
game_cohort_sizes = game_retention[game_retention['cohort_month'] ==
game_retention['activity_month']].set_index(['game_id', 'cohort_month'])['user_id']
game_retention['retention_rate'] = game_retention.apply(lambda x: x['user_id'] /
game_cohort_sizes[(x['game_id'], x['cohort_month'])], axis=1)

# Тип активності
activity_type_retention = activity.groupby(['activity_type', 'cohort_month',
'activity_month'])['user_id'].nunique().reset_index()
activity_type_cohort_sizes =
activity_type_retention[activity_type_retention['cohort_month'] ==
activity_type_retention['activity_month']].set_index(['activity_type',
'cohort_month'])['user_id']
activity_type_retention['retention_rate'] = activity_type_retention.apply(lambda x:
x['user_id'] / activity_type_cohort_sizes[(x['activity_type'], x['cohort_month'])],
axis=1)

# Мова користувача
language_retention = activity.merge(active_users, on='user_id').groupby(['language',
'cohort_month', 'activity_month'])['user_id'].nunique().reset_index()
language_cohort_sizes = language_retention[language_retention['cohort_month'] ==
language_retention['activity_month']].set_index(['language',
'cohort_month'])['user_id']

```

```

language_retention['retention_rate'] = language_retention.apply(lambda x:
x['user_id'] / language_cohort_sizes[(x['language'], x['cohort_month'])], axis=1)
# Виведення зрізів
print("Retention Rate by Game:\n", game_retention)
print("Retention Rate by Activity Type:\n", activity_type_retention)
print("Retention Rate by Language:\n", language_retention)
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, roc_curve, auc
# Завантаження даних Active_users
active_users = pd.read_csv('active_users.csv')
# Додавання залежної змінної Retention
T_0 = 3600 * 24 * 7 # Наприклад, T_0 = 1 тиждень у секундах
active_users['Retention'] = np.where(active_users['total_seconds'] >= T_0, 1, 0)
# Перевірка кількості категорій
print(active_users['Retention'].value_counts())
# Підготовка змінних для моделі
X = active_users[['age']] # Незалежна змінна (вік користувачів)
y = active_users['Retention'] # Залежна змінна (утримання)
# Розділення на навчальну та тестову вибірки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)
# Побудова моделі логістичної регресії
log_model = LogisticRegression()
log_model.fit(X_train, y_train)
# Оцінка коефіцієнтів моделі
beta_0 = log_model.intercept_[0] #  $\beta_0$ 
beta_1 = log_model.coef_[0][0] #  $\beta_1$ 
print(f"Коефіцієнти моделі:  $\beta_0 = \{beta_0\}$ ,  $\beta_1 = \{beta_1\}$ ")
# Прогноз для тестових даних

```

```
y_pred_prob = log_model.predict_proba(X_test)[: , 1] # Ймовірність Retention=yes
y_pred = (y_pred_prob > 0.5).astype(int) # Класифікація з порогом 0.5
# Оцінка якості моделі
print("Класифікаційний звіт:")
print(classification_report(y_test, y_pred))
# Визначення ймовірності утримання для різних значень віку
age_range = np.linspace(X['age'].min(), X['age'].max(), 100).reshape(-1, 1)
p_age = log_model.predict_proba(age_range)[: , 1]
plt.figure(figsize=(10, 6))
plt.plot(age_range, p_age, color='green', label="Ймовірність утримання p(age)")
plt.axhline(y=0.5, color='red', linestyle='--', label="Попіг = 0.5")
plt.axhline(y=0.1, color='orange', linestyle='--', label="Попіг = 0.1")
plt.title("Ймовірність утримання користувачів залежно від віку")
plt.xlabel("Вік користувачів")
plt.ylabel("Pr(Retention = yes | age)")
plt.legend()
plt.show()
```