

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Факультет інформаційних технологій

УДК 004.4:[004:658.114.25]

«ПОГОДЖЕНО»

«ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ»

Декан факультету
інформаційних технологій

Завідувач кафедри комп'ютерних наук

Болбот І.М., д.т.н., професор

Голуб Б.Л., к.т.н., доцент

_____ 2024 р.

_____ 2024р.

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему Аналітична система ведення електронної комерції

Спеціальність 121 Інженерія програмного забезпечення

(код і назва)

Освітня програма Програмне забезпечення інформаційних систем

(назва)

Орієнтація освітньої програми

(освітньо-професійна або освітньо-наукова)

Гарант освітньої програми

к.т.н., професор

Голуб Б. Л.

(науковий ступінь та вчене звання)

(підпис)

(ПІБ)

Керівник магістерської кваліфікаційної роботи

д.т.н., професор

Бушма О. В.

(науковий ступінь та вчене звання)

(підпис)

(ПІБ)

Виконала

Трофимчук С.О.

(підпис)

(ПІБ студента)

КИЇВ-2024

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	3
ВСТУП	4
1. АНАЛІЗ ТА ПРОЄКТУВАННЯ.....	10
1.1. Аналіз предметної області	10
1.2. Аналіз існуючих рішень та проблем.....	14
1.3. Постановка завдання.....	18
1.4. Проєктування системи.....	25
2. МЕТОДИ ТА ТЕХНОЛОГІЇ СТВОРЕННЯ АНАЛІТИЧНОЇ СИСТЕМИ	31
2.1 Основні концепції розробки аналітичних систем	31
2.2 Використані технології.....	38
2.3 Моделювання сховища даних.....	42
3. РОЗРОБКА АНАЛІТИЧНОЇ СИСТЕМИ	44
3.1 ELT процесинг	44
3.2 Моделювання даних.....	46
3.3 Побудова звітності в середовищі Power BI	50
3.4 Аналіз даних шляхом 1Rule алгоритму	54
3.5 Кластерний аналіз.....	61
3.6 Результати дослідження та рекомендації.....	66
ВИСНОВКИ	69
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	71

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

OLAP – online analytical processing, аналітична обробка в реальному часі.

СД – сховище даних.

БД – база даних.

ІС – інформаційна система.

SQL – structured query language, мова структурованих запитів.

OLE DB – Object Linking and Embedding, Database.

ETL – Extract, Transform, Load

ELT – Extract, Load, Transform.

ВСТУП

На сьогоднішній день бізнес у сфері електронної комерції (eCommerce) дуже стрімко змінюється і зараз одним із ключових рушіїв цих змін є цифровізація та активний розвиток цієї сфери. Через те, що зростає популярність купувати онлайн та переходом більшої частини користувачів у покупки в Інтернеті, компанії змушені адаптовувати свої стратегії розвитку та цифровізуватися для того, аби зберігати свою конкурентоспроможність. У таких сучасних умовах недостатньо лише пропонувати якісний продукт чи певну послугу, а також необхідно ефективно використовувати дані та аналітику в цілому. Дані, у сьогоднішній час, стали новим двигуном бізнесу, а аналітика у свою чергу, ключовим інструментом для управління цим двигуном.

Аналітика дозволяє бізнесу приймати обґрунтовані рішення, спираючись на реальні факти, а не на припущення. Для кращого розуміння її ролі можна розглянути приклад аналізу рекламних кампаній та продажів. Давайте уявимо бізнес, який процвітає: він має широкий асортимент продукції, сучасний веб-сайт і постійно зростаючу кількість клієнтів. Але що далі? Як підтримувати темпи зростання та адаптуватися до змін у поведінці споживачів? Згідно з дослідженням Forbes, понад половина онлайн-покупців (53%) віддають перевагу компаніям, які персоналізують їхній досвід[1]. Це підкреслює важливість використання аналітики для збору даних про клієнтів і створення релевантних стратегій[1].

Сам процес аналітики стосується процесу збору, аналізу та інтерпретації даних, пов'язаних з діяльністю в веб-мережі, шляхом відстеження показників і ключових показників ефективності (KPI), щоб отримати розуміння поведінки клієнтів, продажів, маркетингу та загальних бізнес-операцій на цифровому ринку. Аналізуючи ці дані за допомогою таких інструментів, як статистичний аналіз, візуалізація даних[2] і машинне навчання, компанії можуть отримати цінну інформацію.

Аналітика дає шанс відповісти на питання «Як покращити бізнес?» і може використовуватися для прогнозування майбутніх показників бізнесу. Власники бізнесу використовують аналітику для виявлення тенденцій і взаємозв'язків, щоб отримати значущу інформацію з даних для здійснення стратегічних дій - наприклад, аналізу загальних настроїв відгуків клієнтів або розрахунку рентабельності інвестицій (ROI) маркетингової кампанії.

Зі 100% сучасних компаній лише 20 планують і реалізують маркетингові активності на основі великих даних, вигравши тим самим у 80%, що залишилися. Маркетингова стратегія, розроблена на основі даних, завжди більш точна та результативна в порівнянні з тією, яка була побудована на здогадах та особистих припущеннях. Всі ті, хто ігнорує Big Data, в результаті нераціонально витрачають бюджети, не розвиваються і добре, якщо стоять на одному місці, а не залишають ринок назавжди [3].

Data Driven підхід дозволяє відстежити поведінку цільової аудиторії, передбачити її потреби та вплинути на прийняття рішень. Пропонуємо разом з нами розібратися у тому, що таке Data Driven маркетинг, як його застосовувати та чого з його допомогою можна досягти [3].

Data Driven marketing - це маркетинг, заснований на аналізі великих даних про людей та дії, які вони здійснюють в інтернеті та офлайн. Data Driven маркетинг покликаний спрогнозувати поведінку користувачів у мережі та оптимізувати рекламні меседжі таким чином, щоб вони були максимально персоналізованими та ефективними.[3]

Згідно з даними дослідження агенції Adweek, більша половина маркетологів активно використовують Big Data в своїй роботі:



Рис. 1.1 Використання Big Data у маркетингу

- **77%** з них переконані в тому, що дані добре допомагають при роботі з аудиторією;
- **69%** за їх допомогою налаштовують таргетований контент;
- **55%** без Big Data не зможуть розробити маркетингову стратегію просування бренду;
- **49%** вважають, що саме великі дані допомагають їм налагоджувати комунікації з клієнтами;
- **44%** будують на даних усі свої аналітичні звіти[4].

Саме тому аналітика є одним із найпотужніших інструментів, який сучасні компанії можуть використовувати для досягнення своїх цілей. Вона дозволяє отримати цінну інформацію, яка допомагає приймати обґрунтовані рішення, підвищувати ефективність і досягати максимального впливу. Однією з ключових сфер, де аналітика відіграє вирішальну роль, є аналіз рекламних кампаній та продажів. Використання даних у цих процесах дозволяє підвищити рентабельність інвестицій (ROI) та дає можливість глибше зрозуміти споживача, адаптувати маркетингові стратегії та сприяти зростанню бізнесу.

Тут можна виділити кілька пунктів:.

1. Оцінка ефективності рекламних кампаній.

Рекламні кампанії є основою для залучення нових клієнтів, підвищення впізнаваності бренду та стимулювання продажів. Проте без аналітики компанії часто витрачають значні бюджети без розуміння, що саме працює, а що ні.

Аналітика дозволяє відповісти на такі запитання:

- які маркетингові канали (соціальні мережі, email-маркетинг, контекстна реклама) приносять найбільше конверсій?
- які рекламні повідомлення або формати оголошень резонують із цільовою аудиторією?
- який сегмент аудиторії є найціннішим для кампанії?

Завдяки аналітиці можна відстежувати ключові показники ефективності (KPI), такі як рейтинг кліків (CTR), вартість залучення клієнта (CAC), коефіцієнт конверсії, а також загальну рентабельність інвестицій у рекламу (ROAS). Наприклад, якщо кампанія у соціальних мережах має високий CTR, але низький коефіцієнт конверсії, це може свідчити про те, що користувачі зацікавлені оголошенням, але не завершують покупку через недосконалий процес оформлення замовлення на сайті.

2. Оптимізація бюджету.

Один із найбільших викликів у рекламних кампаніях - ефективний розподіл бюджету. Аналітика дозволяє зрозуміти, які канали приносять найвищу віддачу на вкладені кошти. Наприклад, компанія може побачити, що реклама на пошукових платформах має вищий коефіцієнт конверсії порівняно з банерною рекламою. Використовуючи ці дані, бюджет можна перерозподілити на більш результативні канали, мінімізуючи витрати на менш ефективні формати.

3. Персоналізація реклами.

Сучасні споживачі очікують персоналізованого підходу. Аналітика дає змогу сегментувати аудиторію за демографічними показниками, поведінковими моделями або вподобаннями. Наприклад, завдяки аналізу даних можна

створювати таргетовані рекламні кампанії, які відповідають інтересам кожного сегмента.

Наприклад, дані показують, що молодша аудиторія (18–25 років) частіше купує через мобільний додаток, тоді як старша аудиторія (45–55 років) надає перевагу покупкам через сайт. Знаючи це, компанія може створити окремі рекламні кампанії для кожної групи: для молодшої аудиторії — акцент на рекламі мобільного додатка, а для старшої — на зручності використання сайту.

Саме через це дане дослідження є актуальним у наш час, адже лише за допомогою аналітики компанії можуть зростати.

Мета дослідження - розробити аналітичну систему для виявлення найефективніших каналів трафіку та покращити процес прийняття рішень на конкретному підприємстві.

Об'єктом дослідження: обсяг продажів підприємства у сфері електронної комерції

Предметом дослідження: ключові метрики та фактори, що визначають успішність продажів в електронній комерції

З цього випливає, що **завданням дослідження** є: розробка аналітичної система, яка надасть змогу визначати ефективність показників компанії та приймати правильні рішення в подальшому.

Наукова новизна даного дослідження полягає у комплексному підході до аналізу трафіку та її взаємодія з продажами у сфері електронної комерції. Це дозволить автоматизувати аналітику ефективності рекламних джерел трафіку і оцінювати їхній внесок у обсяг продажів.

У цьому аспекті комплексність пояснює системність та узгодженість використання сучасних інструментів, методів та аналітичних підходів для дослідження впливу каналів трафіку на продажі на конкретному підприємству. Цей підхід базується, у свою чергу, на інтеграції різних етапів обробки даних,

починаючи від збору цих даних та підготовки до візуалізації та інтерпретації результатів.

Апробація результатів дослідження

Апробація постановки проблеми та аналізу рішень дослідження відбулася в рамках V Всеукраїнської науково-практичної інтернет-конференції студентів та аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем 2023» 26 квітня 2023 (додаток А).

Результати дослідження було апробовано в рамках XIV Міжнародної науково-практичної конференції молодих вчених «Інформаційні технології: економіка, техніка, освіта 2023» 26-27 жовтня 2023 року м. Київ, Україна (додаток Б).

Також було створено постер, де показані основні моменти проведеного дослідження.

Магістерська робота містить вступ, три розділи, висновки, використані джерела та додатки.

Нижче наведено короткий опис до кожного розділу.

У першому розділі було розглянуто задачу, яку необхідно було дослідити, було надано детальний опис предметної області, розкрито постановку завдання, створено діаграми прецедентів та опис архітектури системи.

Другий розділ був присвячений

1. АНАЛІЗ ТА ПРОЄКТУВАННЯ

1.1. Аналіз предметної області

Аналітика у сфері eCommerce – це така галузь, яка зосереджується на зборі, обробці та аналізі даних, пов'язаних з роботою компаній, які займаються продажами у мережі Інтернет. Вона дає змогу компаніям краще розуміти своїх клієнтів, оцінювати ефективність маркетингових кампаній, управляти продажами, а також оптимізувати всі бізнес-процеси.

Збір та правильна обробка даних є однаково важливими як для офлайн так і для онлайн бізнесу. На відміну від офлайн бізнесу, який тільки починає масово впроваджувати BI (Business Intelligence) – системи, схожі на Microsoft PowerBI, онлайн-бізнес уже активно використовує інструменти для збору та консолідації даних[5].

Однією з ключових задач аналітики у маркетингу є розуміння джерел трафіку. Це включає аналіз того, звідки приходять відвідувачі – через пошукові системи, рекламні кампанії, соціальні мережі чи інші канали. Завдяки таким інструментам, як Google Analytics чи інших спеціалізованих аналітичних систем, компанії можуть визначити, які канали приносять найбільший обсяг відвідувачів, а які – конверсії. Важливо не лише кількість, але й якість трафіку, тобто наскільки ймовірно, що користувач здійснить покупку або іншу цільову дію.

Ще одним важливим напрямом є аналіз поведінки клієнтів у продукті. Це включає вивчення того, як довго користувачі перебувають на платформі, які сторінки відвідують, які товари додають до кошика і на якому етапі вони можуть залишити процес покупки. Наприклад, аналіз покинутих кошиків дає змогу зрозуміти, що заважає користувачам завершувати покупки, і відповідно вдосконалити процес замовлення.

Не менш важливою складовою є аналіз продажів. Він дозволяє визначити, які товари користуються найбільшою популярністю, які категорії товарів

забезпечують найбільші доходи, а також як змінюються продажі залежно від сезонності чи акцій. Це дозволяє не лише краще прогнозувати попит, але й ефективно управляти запасами, зменшуючи надлишкові витрати.

Ключовим аспектом аналітики є сегментація клієнтів. Поділ клієнтів на групи за певними характеристиками, такими як вік, географія, частота покупок чи середній чек, дає змогу створювати персоналізовані пропозиції. Наприклад, для клієнтів, які часто купують товари певної категорії, можна розробити спеціальні акції чи рекомендації, що підвищують ймовірність повторної покупки.

Маркетингові кампанії також є важливим об'єктом аналізу. Завдяки аналітиці можна оцінити, наскільки ефективними були витрати на рекламу, які кампанії принесли найбільший дохід, а які – не виправдали очікувань. Використання метрик, таких як ROI (рентабельність інвестицій) чи CPA (вартість залучення клієнта), допомагає компаніям оптимізувати рекламні бюджети.

Окрім цього, аналітика електронної комерції все більше використовує прогнозування. Завдяки сучасним алгоритмам машинного навчання компанії можуть передбачати поведінку клієнтів, прогнозувати попит на товари або виявляти ризик втрати клієнтів. Також активно впроваджуються рекомендаційні системи, які пропонують клієнтам товари, що можуть їх зацікавити, на основі їхньої історії покупок.

Аналітика електронної комерції також охоплює управління логістикою. Наприклад, аналіз часу доставки чи оптимізація маршрутів дозволяє зменшити витрати і підвищити задоволеність клієнтів. Оцінка ефективності роботи складів і ланцюгів постачання є невід'ємною частиною цього процесу.

Сучасні технології, такі як ETL/ELT, хмарні бази даних, BI-платформи (наприклад, Power BI або Tableau), забезпечують швидку обробку великих обсягів даних та надають компаніям потужні інструменти для аналізу. Інтеграція

даних з різних джерел дозволяє створювати цілісну картину бізнесу та приймати обґрунтовані рішення в реальному часі.

Таким чином, аналітика eCommerce бізнесу є складним, але надзвичайно потужним інструментом, який дозволяє компаніям не лише аналізувати минулі результати, але й прогнозувати майбутні тенденції та адаптуватися до змін на ринку. Це ключ до побудови успішного бізнесу в умовах жорсткої конкуренції та швидкої діджиталізації.

Діаграма прецедентів використовується для відображення взаємозв'язків між об'єктами, що взаємодіють, та способами їх взаємодії у визначеній області. Таким чином, на цій діаграмі виокремлюються основні компоненти, а саме:

Актор - це представник користувача, особа або інший фізичний об'єкт, який існує у даній предметній області.

Прецедент - це конкретна дія або операція, яку виконує конкретний актор. Ця операція може виконуватися безпосередньо актором або спільно з іншим прецедентом.

Було визначено наступних акторів:

- 1) користувач – людина, яка взаємодіє із покупкою товарів;
- 2) менеджер з продажу - займається обробкою замовлень клієнтів та контролем за доставкою замовлення;
- 3) маркетолог - фокусується на створенні маркетингових стратегій;
- 4) product owner - відповідає за формування стратегії розвитку продукту та A/B тестування;
- 5) аналітик – відповідає за обробку і організацію аналітичних даних.

Актори взаємодіють через різні процеси або завдання. Процеси включають дії, такі як "Пошук товару", "Обробка замовлень клієнтів", "Контроль за доставкою замовлення", "Створення маркетингових стратегій", "Формування

стратегії розвитку продукту", "A/B тестування", "Підготовка аудиторії", "Обробка та організація аналітичних даних", і "Збір аналітичних даних".

Стрілки показують напрямок взаємодії між ролями та процесами, вказуючи на потік інформації або завдань, які необхідно виконати. Наприклад, менеджер з продажу займається обробкою замовлень клієнтів та контролем за доставкою замовлення, тоді як маркетолог фокусується на створенні маркетингових стратегій. Product Owner може відповідати за формування стратегії розвитку продукту та A/B тестування, а аналітик - за обробку і організацію аналітичних даних.

За допомогою визначених прецедентів і операцій, які виконують актори, була розроблена діаграма прецедентів. Ця діаграма представлена на рисунку 2.

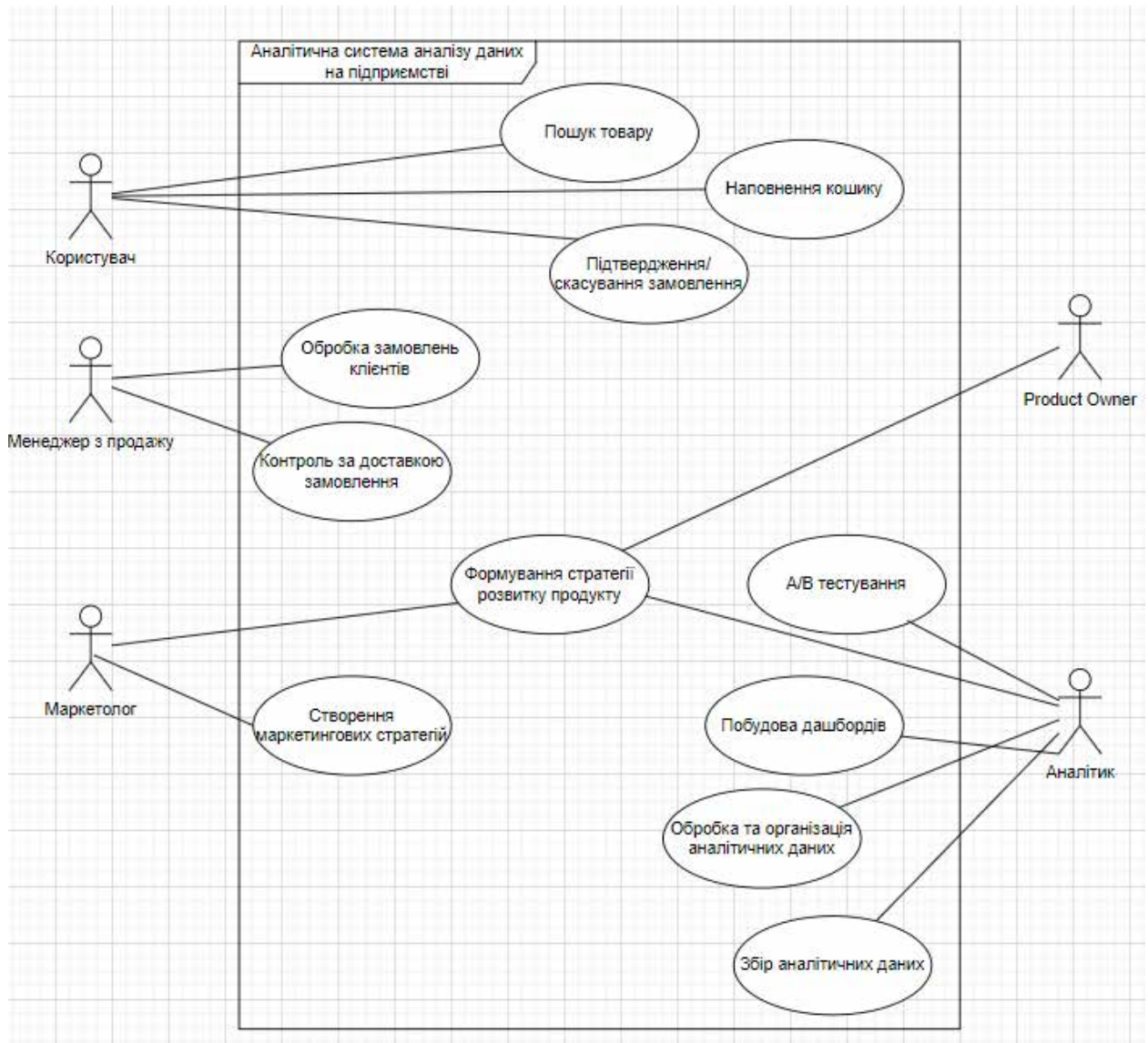


Рис.2. Діаграма прецедентів предметної області

1.2. Аналіз існуючих рішень та проблем

На сьогоднішній момент стан справ у сфері аналізу даних у електронній комерції вказує на те, що компанії деталі частіше та активніше почали використовувати аналітику для підвищення ефективності власних бізнес-процесів. Стосується це також і оптимізації маркетингових стратегій та покращення клієнтського досвіду користування продуктами.

Галузь електронної комерції стикається з великим викликом обробки величезних обсягів даних і отримання прибутку від аналізу даних. Це викликано інформаційною революцією. Big Data Analytics (BDA) - це методи, які використовуються для керування та вивчення цієї купи даних, яка складається з текстових повідомлень, коментарів у соціальних мережах тощо. На додаток до цього, Big Data Analytics можна використовувати в секторі електронної комерції, щоб отримати збільшити доходи та залучити більше клієнтів[6].

У цій галузі аналіз даних охоплює дуже широкий спектр завдань: від відстеження поведінки користувачів у продукті до створення персоналізації пропозицій користувачам, прогнозування продажів, витрат, оптимізація логістики точно.

Цю позицію підтверджує аналітична стаття **Analytical evaluation of big data applications in E-commerce: a mixed method approach**, у якій описано дослідження, як аналітика великих даних трансформує сектор електронної комерції. У свою чергу, можемо стверджувати, що роль великих даних у електронній комерції дуже значна - аналітика дозволяє компаніям обробляти величезні обсяги структурованих та неструктурованих даних. Ці дані використовуються для отримання корисної інформації, оптимізації бізнес-процесів та прийняття рішень[7].

Також дане дослідження акцентує увагу на створенні аналітичної інфраструктури та підкреслює, що для розробки системи обробки великих даних найчастіше застосовують інструменти Apache Hadoop та Spark і суміш хмарних рішень (AWS кластери) для обробки та зберігання великих наборів даних.

З точки зору потокової обробки найчастіше використовуються інструменти Apache Kafka та Spark Streaming. Вони дозволяють аналізувати великі набори даних у реальному часі.

Але з точки зору слабких місць у аналізі даних у цій сфері, дослідження визнає проблеми у конфіденційності даних, високі витрати на впровадження та недостатню кількість кваліфікованого персоналу для керування складними аналітичними системами

У іншому дослідження, **Development of E-Commerce Website with Analytics** описано, яким чином компанії у сфері електронної комерції збирають та аналізують дані, які технології і інструменти використовують при цьому.

Можна виділити три етапи, у яких компанії збирають дані.

1. **Збір та інтеграція даних.** Компанії зачасту збирають дані з різних джерел, таких як веб-аналітика, власні CRM системи, платіжні системи тощо. Деякі компанії у свою чергу використовують Google Analytics 4 для відстеження поведінки користувачів на сайті.

2. **Маркетингова аналітика.** У цьому напрямі фокус більше спрямовано на оптимізацію рекламних кампаній, аналіз ефективності рекламного трафіку та розробки персоналізованих стратегій по комунікації з клієнтом. У статті наведений приклад про сегментацію клієнтів на основі їхньої поведінки або попередніх покупок, де компанії адаптують свої маркетингові повідомлення майже для кожного користувача.

3. Також важливим етапом є **прогнозування та автоматизація.** Це здійснюється шляхом використання машинного навчання для прогнозування продажів, визначення відтоку клієнтів тощо[8].

Щодо технологій, то для великих даних використовують Hadoop та Spark, для хмарної інфраструктури використовують AWS, Google Cloud, Microsoft Azure. Інструментами для аналізу даних є PowerBi, Tableau, Looker – вони допомагають візуалізувати дані та приймати бізнес-рішення. Для трансформації даних використовують dbt [9][12][13].

У дослідженні **Implementing Big Data Analytics in Business Management** описано дві основні методології для аналізу даних: K-means Clustering та Sentiment Analyst.

Компанії використовують метод середніх для когортного аналізу користувачів за поведінковими атрибутами. Виділені кластери допомагають розробляти цільові маркетингові стратегії для кожної когорти користувачів.

Також дослідження описує методи аналізу клієнтів за допомогою методів обробки природньої мови (NLP), аналізує текстові дані для визначення емоційних тонів[10].

З інструментів аналізу у сфері eCommerce дослідження описує мову програмування Python та бібліотеки Pandas, NumPy, Scikit-learn для кластеризації[11].

Отже, можемо підсумувати наявні дослідження та виявити, що для технологій обробки великих даних застосовуються інструменти Hadoop та Spark. Python є основною мовою для аналізу даних та проведенню глибокого аналізу даних. Компанії можуть мати безліч джерел даних, які необхідно об'єднувати між собою. Найчастіше використовують саме ETL підхід, хоча вибір підходу інтеграції даних залежить від специфіки компанії. Великі компанії більше використовують хмарні рішення, що для малих підприємств деколи є дуже дорогим. Основними проблемами у сфері електронної комерції вбачається конфіденційність даних та великі витрати на обслуговування аналітичних інфраструктур. Через свою «неповороткість» компанії платять занадто великі кошти для пришвидшення аналізу даних, хоча проблема вбачається саме у гомісткій інфраструктурі.

Також спостерігається проблема з масштабуванням, інтеграцією нових інструментів для аналітики та недоліками з якістю даних, які виникають у процесі аналізу.

1.3. Постановка завдання

Етап постановки завдання необхідний для визначення основної мети дослідження та формулювання шляхів досягнення цієї мети. Було визначено, що головною метою даної роботи є **розробка аналітичної системи для управління електронною комерцією**, а саме створення аналітики та аналітичної інфраструктури для вимірювання ефективності продажів підприємства у розрізі рекламних кампаній та джерел трафіку.

Робота проводилася з дотриманням законодавства України про захист персональних даних та конфіденційність інформації. При роботі з персональними даними використовувалися лише анонімізовані набори, що відповідають вимогам GDPR.

Першим етапом є аналіз вхідної інформації, яке накопичує та зберігає підприємство під час ведення свого бізнесу. Для цього необхідно проаналізувати існуючу архітектуру системи компанії щоб визначити, які дані ми уже отримуємо.

Нижче, на Рис.1.3 представлено основні модулі архітектури системи компанії. На основі неї зробимо висновок щодо типів наявних даних у компанії.

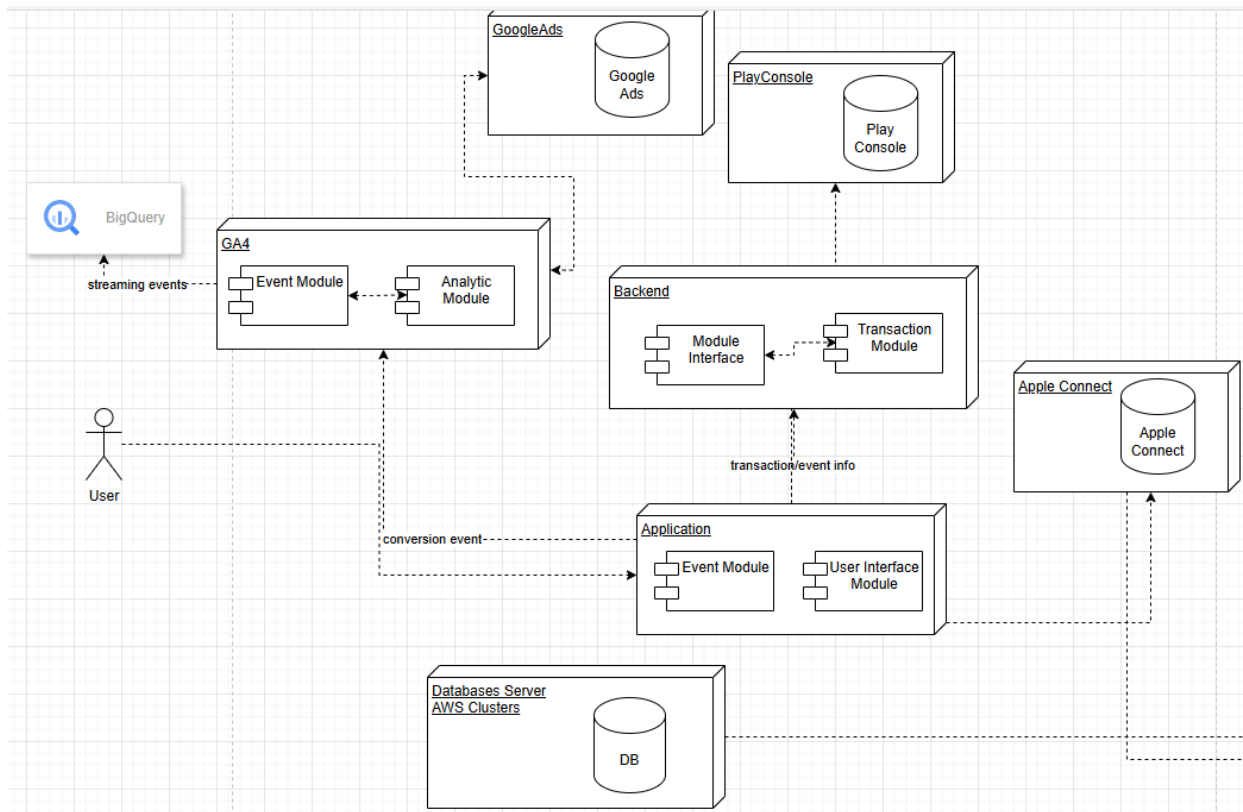


Рис. 1.3 Основні складові архітектури системи компанії

Виходячи з Рис.1.3 бачимо такі основні компоненти архітектури.

1. **Користувач (User)** – це кінцевий користувач, який взаємодіє з додатком, генеруючи події (event) та виконуючи дії, які можуть конвертуватися у транзакції або інші аналітичні події.
2. **Додаток (Application)** – це основний інтерфейс для користувачів, який включає два модулі:
 - а. **Event Module** – модуль, який фіксує події користувача, такі як натискання кнопок, перегляд сторінок, конверсії тощо.
 - б. **User Interface Module** – інтерфейс користувача, через який він взаємодіє з додатком, а також надсилає події у систему.
3. **Backend** – серверна частина, яка обробляє інформацію з додатка та передає її в інші системи.

- a. **Module Interface** – інтерфейс для обміну даними між модулями.
 - b. **Transaction Module** – обробляє інформацію про транзакції та події, які генеруються користувачем.
4. **Google Analytics 4 (GA4)** – платформа для аналітики, яка використовується для збору подій користувача та їх аналізу.
 - a. **Event Module** – модуль збору подій, який отримує інформацію про події від Backend та додатка.
 - b. **Analytic Module** – модуль аналітики, який обробляє події та надає аналітичні звіти
 5. **GoogleAds** – інструмент, який відстежує дані про рекламні кампанії та взаємодію користувачів з рекламою. Інформація з Google Ads потрапляє у GA4 і може використовуватися для оцінки ефективності реклами.
 6. **Play Console та Apple Connect** – ці компоненти використовуються для моніторингу мобільних додатків на платформах Google Play і Apple App Store відповідно. Вони надають дані про використання додатка, завантаження, помилки тощо.
 7. **BigQuery** – платформа для зберігання та обробки великих обсягів даних. В даній архітектурі BigQuery отримує потоки подій (streaming events) з GA4 для збереження і подальшого аналізу.

Ці компоненти взаємодіють таким чином.

1. **Користувач** взаємодіє з **додатком**, генеруючи події, які обробляються через Event Module.
2. Події передаються до **GA4** для збору даних і аналізу користувацьких взаємодій.
3. **Backend** отримує інформацію про події та транзакції з додатка і передає ці дані до аналітичних платформ, таких як GA4 та **Amplitude**.

4. Події з **GA4** надходять у **BigQuery** через потоки подій для подальшого аналізу великих даних.

5. **Google Ads** передає інформацію про рекламні кампанії в **GA4** для зв'язку взаємодії користувача з рекламою та подальшого аналізу.

6. **Amplitude** отримує події з **Backend**, що дозволяє детально аналізувати поведінку користувачів.

7. **Play Console** та **Apple Connect** забезпечують додаткові дані про взаємодію з мобільними додатками. Ці дані імпортуються у **BigQuery** датасети.

Можемо бачити, що підприємство уже збирає дані про дії користувачів, але у рамках поточного дослідження не передбачено використання цих даних, а лише тих, які стосуються продажів та каналів трафіку, адже розробка такою системи є дуже трудомісткою задачею.

Найголовнішими нашими джерелами даних у даному випадку виступають **GoogleAds**, **Google Analytics 4** та база даних компанії. У цих джерелах уже є необхідна інформація щодо каналів трафіку та продажів, необхідно лише її коректно зібрати, змоделювати та проаналізувати.

Google Analytics 4 – це платформа для аналітики від **Google**, яка дозволяє компаніям різної величини безкоштовно відстежувати взаємодію їх користувачів на веб-сайтах або мобільних додатках[14].

Цей сервіс популярний через свою специфіку збору даних про взаємодію користувачів, яка реалізована на подієво-орієнтованій моделі (**event-based**). Ця модель дозволяє досліджувати взаємодії користувача у вигляді певних подій (клік, перегляд сторінки, перехід за посиланням тощо).

Також **GA4** можна напряму інтегрувати з **GoogleAds**, що дозволяє обмінюватися інформацією та точніше налаштовувати рекламні кампанії (персоніфікувати їх за когортами).

Звісно, існують і інші сервіси, за допомогою яких можна збирати статистику по користувачах, але важливо зауважити, що не всі сервіси є безкоштовними і не всі дають можливість напряму безкоштовно вивантажувати дані.

Серед сервісів, які можна було використовувати під час дослідження є Amplitude. Цей сервіс також компанія почала інтегрувати до своєї системи, але поки безуспішно. Цей сервіс є інструментом для глибокого аналізу поведінки користувачів, що також у свою чергу підтримує подієво-орієнтовану модель [15]. Але, на жаль, Amplitude заточена більше під продуктову аналітику (дослідження воронки (шляхів) користувача у системі і тд), що не дозволяє використовувати цей інструмент у рамках цієї задачі. Також немає безкоштовних інструментів для атрибуції трафіку (що доступно у Google Analytics 4 безкоштовно).

Можемо зробити висновок, що не всі джерела даних, які присутні у компанії необхідні для реалізації задачі аналізу продажів за каналами трафіку. Наступним же етапом виконання мети є збір цієї інформації у централізованій базі даних.

Розробляти базу даних не довелося, адже вона була реалізована підприємством самостійно. У базі даних підприємство зберігає різноманітну інформацію, але розглянемо основні сутності, які необхідні нам у рамках дослідження:

1. Користувач – збереження інформації про користувачів, які взаємодіяли із системою. Доступна інформація про:

- a. Логін користувача
- b. Нік у системі
- c. Прізвище користувача

- d. Ім'я
- e. Дата реєстрації
- f. Статус користувача

2. Заовлення – сутність, яка містить інформацію про заовлення користувачів. Містить інформацію про:

- a. Ідентифікатор заовлення
- b. Ідентифікатор користувача
- c. Дата заовлення
- d. Статус заовлення
- e. Сума заовлення
- f. Товар

3. Кошик заовлення – сутність, яка зберігає інформацію про товари, які були додані користувачем до кошика перед оформленням заовлення. Вона включає інформацію про:

- a. Ідентифікатор кошика
- b. Ідентифікатор користувача
- c. Ідентифікатор товару
- d. Кількість товару
- e. Дата створення кошика
- f. Статус кошика – вказує на те, чи кошик є активний чи перетворений на заовлення

4. Джерело користувача – сутність, яка описує, яким чином користувач потрапив у систему (через який канал трафіку). Поля сутності включають такі атрибути:

- a. Ідентифікатор джерела трафіку
- b. Назва джерела трафіку

- c. Дата реєстрації користувача
- d. Ідентифікатор користувача

База даних підприємства створена у СУБД PostgreSQL. Саме тому сховище даних також було вирішено розробляти саме там. На це було кілька причин:

1. Підприємство уже використовує PostgreSQL для поточної бази даних, тому у межах компанії уже є компетенції щодо підтримки майбутнього сховища
2. Легша сумісність даних із базою даних та майбутнім сховищем.
3. Оптимальна кількість даних та розмір сховища, що дозволяє використовувати PostgreSQL для аналітичних розрахунків, хоча вона не є повністю OLAP системою.

Сховище даних було побудовано на основі бази даних, даних з Google Analytics 4 та GoogleAds, які містять основну інформацію щодо продажів, каналів трафіку користувачів та додаткової інформації щодо рекламних кампаній, які необхідні підприємству для аналізу.

Саме тому першою задачею, яка стояла, була допомога у створенні вітрини даних для аналізу продажів. Необхідно було забезпечити зберігання таких даних:

1. Інформація про користувача та його рекламні мітки, а саме:
 - a. Параметри джерела кампанії (utm_source);
 - b. Параметри каналу кампанії (utm_medium);
 - c. Параметри назви кампанії (utm_campaign).
 - d. Ідентифікатор користувача
2. Інформацію про товар, а саме:
 - a. Назва товару

а також:

- a. Інформація про підкатегорію товару
 - b. Категорію товару
 - c. Розділу товару
3. Інформація про суму замовлень та к-сть замовлень
 4. Інформацію про статуси замовлень
 5. Інформація за кількістю проданих товарів

Для виконання цієї задачі необхідно розробити структуру сховища даних, щоб в подальшому завантажити туди потрібну інформацію з оперативного БД.

Саме тому для досягнення поставленої мети було визначено такі основні завдання.

1. Аналіз вихідних даних компанії та структуризація цих даних для подальшої побудови DataMart
2. Вибір СУБД для реалізації сховища даних
3. Підбір інструментів для інтеграції даних у сховище
4. Реалізація інтеграції даних у сховищі
5. Моделювання даних
6. Реалізація сховища даних
7. Аналіз даних у середовищі PowerBi
 - a. Оцінка продажів за джерелами трафіку
 - b. Виявлення тенденцій продажів за останні 2 роки та аналіз найбільш популярних товарів у розрізі категорій
 - c. Візуалізація отриманих результатів у середовищі PowerBi.

1.4. Проєктування системи

Проєктування системи відбувалося у 2 етапах та покривало процеси не лише пов'язаних із темою цього дослідження. Опишу кожен з цих етапів окремо.

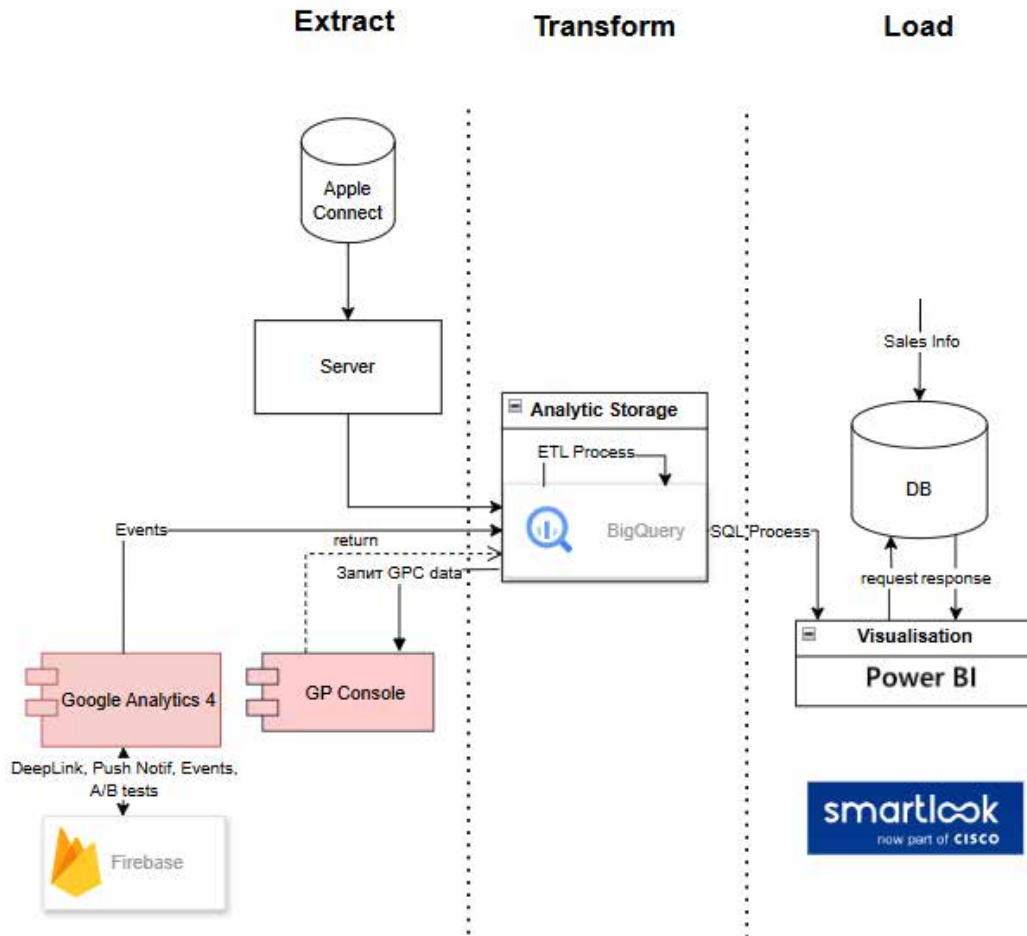


Рис. 1.4 Спроектвана базава аналітика

На Рис. 1.4 представлена перша версія спроектованої системи аналітики для підприємства. У першому варіанті структура системи була поділена на 3 складових ETL процесу: Extract, Transform, Load.

Extract, Transform, Load (ETL) - процес, який використовується в базах даних, у сховищах даних та у засобах Business Intelligence для забезпечення їх роботи для підтримки прийняття рішень [16] [17]. ETL використовують для витягнення даних з усіх джерел у спільне центральне сховище даних, яке зветься сховищем даних або Data Warehouse. У нашому випадку це буде Data Mart.

ETL використовує набір бізнес-правил для очищення, впорядкування необроблених даних та підготовки їх до зберігання, аналізу тощо. Детальніше про концепцію ETL процесів у Розділі 2.

Отже, схема представляє такі блоки:

1. Extract:

a. Apple Connect, Google Analytics 4 та Firebase служать джерелами даних для збору подій користувачів та взаємодії їх із глибокими посиланнями (deep links, push сповіщеннями тощо)

b. Із цих джерел дані надходять у центральний сервер, а потім у аналітичне сховище BigQuery

2. Transfort:

a. Дані проходять **процес ETL** у BigQuery і саме тут ці дані готуються до подальшого аналізу.

3. Load:

a. Перетворені дані із минулого кроку завантажуються в **БД** (базу даних), де уже зберігається інформація про продажі та інші дані, необхідні для звітності.

b. Інструментами візуалізації буде **Power BI** та використання сервісу **Smartlook** для аналізу поведінки користувачів.

З рисунку видно, що у першому варіанті реалізації основний фронт робіт по обробці даних мав бути на боці BigQuery, а уже потім оброблені дані мали завантажувати у PostgreSQL сховище. Але на цьому етапі ми з командою розробників стикнулися з такими проблемами та викликами.

1. Проблема із непотрібним зберіганням (дублюванням) даних на BigQuery. Ми дійшли висновку, що оброблені дані необхідно буде продовжувати зберігати на боці BigQuery, що породжує розміщення одних і тих самих даних у різних місцях, що не є коректним.

2. Збільшення бюджету на обробку даних на боці BigQuery. Хоч і оплата за оброблені байти у BigQuery доволі гнучка, компанія не захотіла платити

додаткові кошти для витрати на обробку даних на сторонньому сервісі, дешевше це було б зробити уже на стороні PostgreSQL

3. Процес переміщення даних спочатку у BigQuery, а потім у PostgreSQL виглядає не зовсім коректним та занадто важким, що ускладнюватиме процес інтеграції даних.

Саме через вищенаведені причини було обрано PostgreSQL як основу для сховища даних. Більш детально про причини буде описано у Розділі 2.

У фінальній версії було прибрано BigQuery як проміжну систему для обробки даних та використано інструмент для інтеграції даних Airbyte та засіб для моделювання даних у сховищі dbt.

Фінальним етапом та затверджену архітектуру системи можна побачити на Рис.1.5.

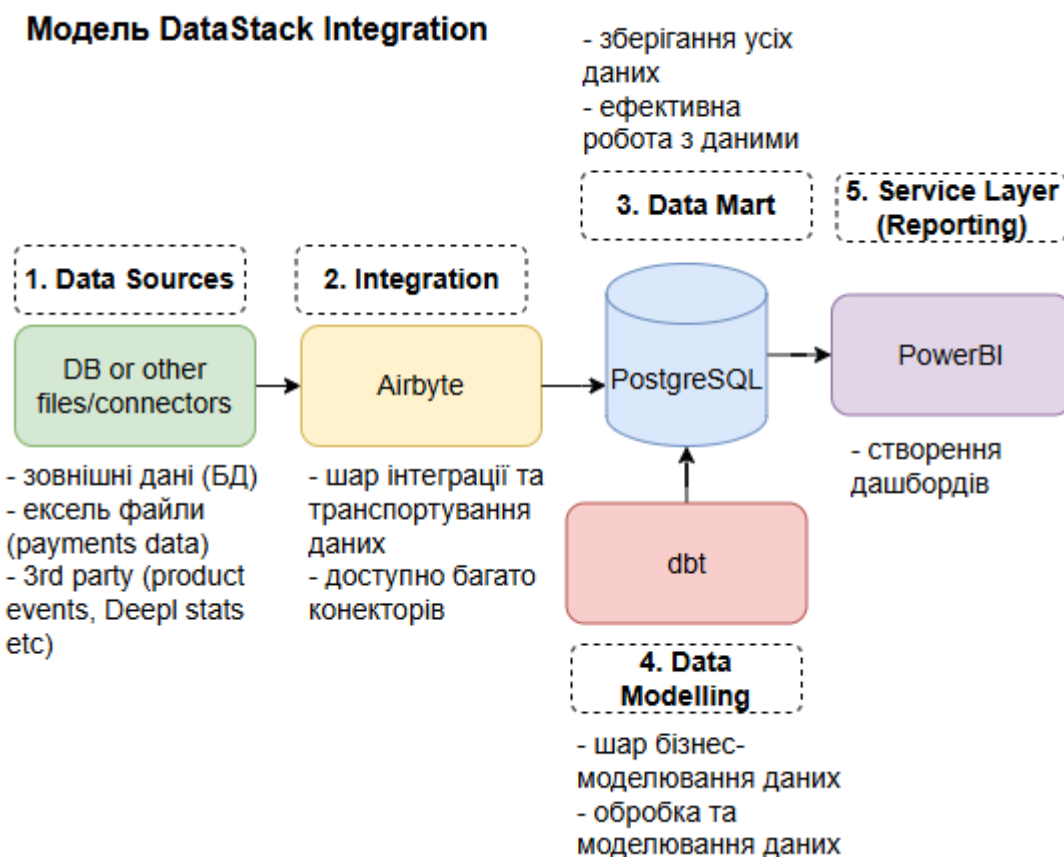


Рис. 1.5 Затверджена версія проекту системи аналітики

Дана архітектура побудована на основі кількох ключових компонентів, кожен з яких відповідає за певний етап обробки даних. Це задумано для того, щоб забезпечити якісне аналітичне середовище згідно Modern Data Stack підходу. Нижче було наведено докладний опис кожного з етапів.

1. **Джерела даних (Data Sources).** На цьому етапі було зібрані різні зовнішні джерела даних. У нашому випадку, це були дані з бази даних (БД), дані з Google Analytics 4 з каналами трафіку тощо.

2. **Інтеграція (Integration).** Для інтеграції даних було використано сервіс **Airbyte**. Цей інструмент забезпечує транспортування даних від різних джерел у сховище даних PostgreSQL.

3. **Сховище даних (Data Mart).** PostgreSQL тут виступає у ролі сховища даних, де зберігаються всі зібрані дані.

4. **Моделювання даних (Data Modeling).** Було використано сервіс **dbt** (data build tool) для перетворення даних у відповідну бізнесову структуру. Цей інструмент у першу чергу допомагає здійснювати моделювання даних, трансформуючи дані всередині PostgreSQL у формат та структуру, яка буде зручною для аналізу.

5. **Сервісний рівень (Service Layer - Reporting).** Даний рівень забезпечує сервіс побудови дашбордів **Power BI**, який є інструментом для візуалізації та аналітики, який легко можна інтегрувати з PostgreSQL для створення дашбордів.

Результатом проєктування системи стала фінальна версія топології системи підприємства включно з аналітичним модулем.

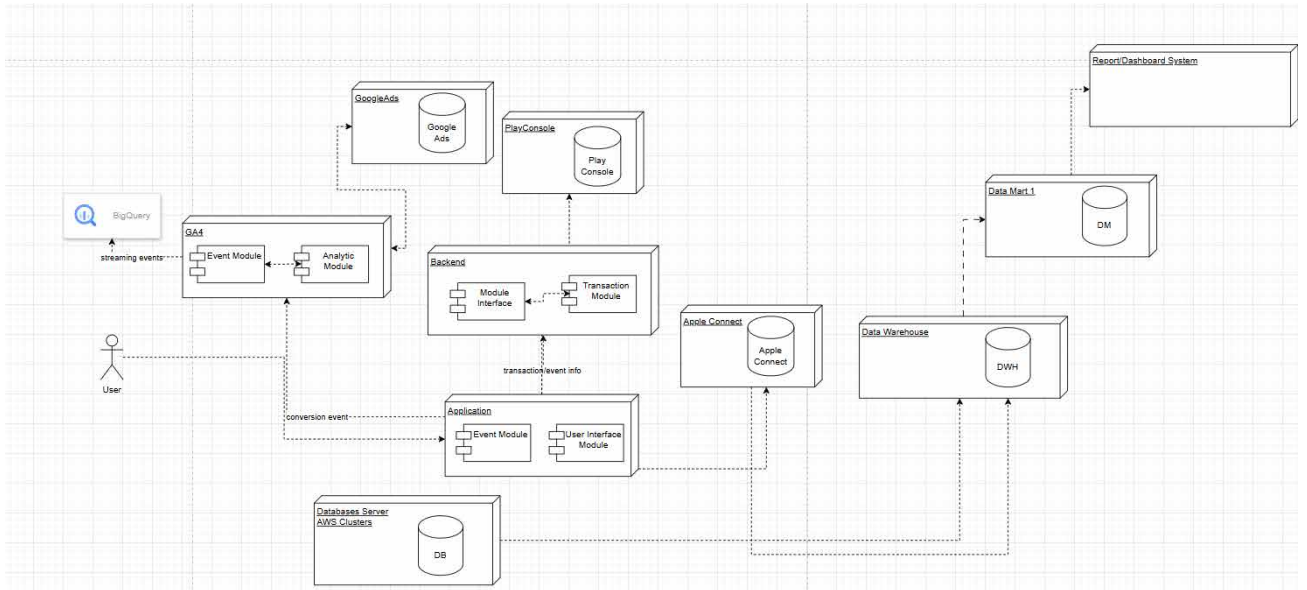


Рис. 1.6. Фінальна версія топологія системи включно з аналітичним модулем

Висновки до Розділу 1. Отже, результатом Розділу 1 є проаналізована предметна область, визначено основні методи та технології розробки та аналізу у сфері eCommerce, сформовано завдання на розробку та спроектована система. У Розділі 2 буде опис методів та технологій створення аналітичної системи, які були використані у даному дослідженні. У Розділі 3 буде опис самого процесу розробки системи

2. МЕТОДИ ТА ТЕХНОЛОГІЇ СТВОРЕННЯ АНАЛІТИЧНОЇ СИСТЕМИ

2.1 Основні концепції розробки аналітичних систем

Першою концепцією, яку необхідно розглянути, є концепція сховищ даних.

Мета концепції сховищ даних - прояснити відмінності в характеристиках даних в операційних і аналітичних системах, визначити вимоги даним, що поміщаються в сховище, визначити загальні принципи і етапи його побудови, основні джерела даних, дати рекомендації по рішенню потенційних проблем, що виникають при їх вивантаженні, очищенні, узгодженні, транспортуванні і завантаженні до цільової бази даних сховища.

Предметом концепції сховищ даних є не аналіз даних, а власне дані, тобто концепція їх підготовки для подальшого аналізу. В той же час концепція сховища даних визначає не просто єдиний логічний погляд на корпоративні дані, а реалізацію єдиного інтегрованого джерела даних [18].

Типовою проблемою, яку вирішує сховище даних, є аналіз продажів. Сам по собі аналіз продажів допомагає виявляти тенденції, допомагає планувати продажі по товарам, клієнтам і тд. Завдяки сховищам даних можна отримати більш інтерактивне та ширше уявлення про результати продажів.

Якщо описувати технології побудових сховищ даних, то варто висвітлити основну її ідею. Ця ідея заключається у тому, що проводити оперативний аналіз даних безпосередньо у оперативній базі нефективно. Це через те, що такі бази більше заточені під швидкодію роботи системи для клієнта, а виконання трудомістких аналітичних запитів зменшить швидкодію усієї бази.

Натомість, всі необхідні для аналізу дані витягуються з декількох традиційних баз даних (в основному, реляційних), перетворюються і потім поміщаються в одне джерело даних – сховище даних.

В процесі занурення дані:

- очищаються – усунення непотрібної інформації;

- агрегуються – обчислення сум, середніх;
- трансформуються – перетворення типів даних, реорганізація структур зберігання;
- об'єднуються із зовнішніх і внутрішніх джерел – приведення до єдиних форматів;
- синхронізуються – відповідність одному моменту часу [18].

Наступним важливим концептом розробки аналітичної системи є Data Warehouse (корпоративні сховища даних) та DataMart (вітрини даних).

Корпоративні сховища даних містять інформацію, що відноситься до всієї корпорації і зібрану з безлічі оперативних джерел для консолідованого аналізу. Зазвичай такі сховища охоплюють цілий ряд аспектів діяльності корпорації і використовуються для ухвалення як тактичних, так і стратегічних рішень. Корпоративне сховище містить детальну і узагальнену інформацію, його об'єм може досягати від 50 Гбайт до одного або декількох терабайт. Вартість створення і підтримки корпоративних сховищ може бути дуже високою. Зазвичай їх створенням займаються централізовані відділи інформаційних технологій, причому створюються вони зверху вниз, тобто спочатку проектується загальна схема, і тільки тоді починається заповнення даними. Такий процес може займати декілька років [18].

Вітрини даних містять підмножину корпоративних даних і будуються для відділів або підрозділів усередині організації. Вітрини даних часто будуються силами самого відділу і охоплюють конкретний аспект, що цікавить співробітників даного відділу. Кіоск даних може отримувати дані з корпоративного сховища (залежний кіоск) або, що поширеніше, дані можуть поступати безпосередньо з оперативних джерел (незалежний кіоск) [18] [19].

Найчастіше, при проектуванні вітрин даних використовуються стандартні моделі такі як схема «зірки» та «сніжинки».

Схема даних «Зірка» є найпопулярнішим типом моделі даних. Вона характеризується наявністю таблиць фактів, які поєднані з нею таблиці розмірностей, які потім поєднуються між собою через центральну таблицю. Вона характеризується високою продуктивністю запитів.

Схема «сніжинка» використовується для нормалізації «зірки». Вона була створена та започаткована як Snowflake схема. Характеристикою цієї схеми є те, що ця схема заточена для можливості більшої деталізації даних (наприклад, за періодом дати, тижня, місяця)

OLAP (Online Analytical Processing) - це технологія обробки даних, яка використовується для аналітичної роботи з великими обсягами даних. Основна мета OLAP - забезпечити швидкий доступ до багатовимірного аналізу даних для прийняття управлінських рішень. Вона є фундаментальною складовою сучасних систем бізнес-аналітики (BI) і використовується для роботи з історичними даними, виявлення тенденцій, а також для прогнозування та моделювання.

За допомогою OLAP аналітики даних можуть швидко аналізувати складні запити, щоб перетворювати величезні набори даних у ефективні висновки, які часто вимагають наші керівники. За допомогою OLAP аналітики можуть виконувати багатовимірний аналіз, тобто вони розглядають такі параметри, як час, географія та ієрархічні категорії, щоб побачити закономірності та тенденції, які, ймовірно, було б важко виявити, якби ви працювали з таблицями та знаходили зв'язки без OLAP.

Наприклад, OLAP допомагає нам отримувати дані про продажі за регіонами, деталізувати деталі, щоб проаналізувати ефективність продажів для кожного окремого магазину, а потім узагальнювати всю інформацію для визначення загальних тенденцій продажів.

Будь-яке обговорення OLAP є упуцтеним, не говорячи про куби OLAP. Куб OLAP, також відомий як гіперкуб, — це свого роду структура даних, яка дозволяє

швидко та багатовимірно аналізувати велику кількість даних, що перевищує те, що можна досягти за допомогою реляційних баз даних. Серед його ключових особливостей:

1. **Розміри** : вони представляють різні категорії для аналізу.
2. **Міри** : це числові значення, які потрібно проаналізувати.
3. **Попереднє агрегування** : дані попередньо обчислюються та зберігаються тут для швидкого запиту.
4. **Ієрархії** : вони дозволяють деталізувати виміри.
5. **Швидка продуктивність запитів** : вони дозволяють швидко аналізувати великі набори даних.

Простіше кажучи, куби OLAP беруть наші таблиці, які є плоскими та двовимірними (думають як електронні таблиці), і підносять їх у щось, що має три (або багато більше) вимірів. Куби OLAP додають додаткові шари та асоціації. Вони належним чином організовують дані для підтримки складних аналітичних запитів і звітності з набагато вищою швидкістю.

OLAP зосереджується на аналізі даних для отримання бізнес-аналізу, тоді як обробка онлайн-транзакцій (OLTP) зосереджена на обробці онлайн-транзакцій у реальному часі. OLTP використовується для виконання онлайн-транзакцій бази даних, які генерують такі працівники, як касири та банківські касири. Програми самообслуговування клієнтів, такі як онлайн-банкінг, подорожі та електронна комерція, також створюють транзакції бази даних і підключаються до систем OLTP. OLTP може бути джерелом даних для систем OLAP.

Ось візуальне зображення, щоб ви могли побачити місце OLAP у цьому:

The OLAP process

How data is prepared for online analytical processing (OLAP)

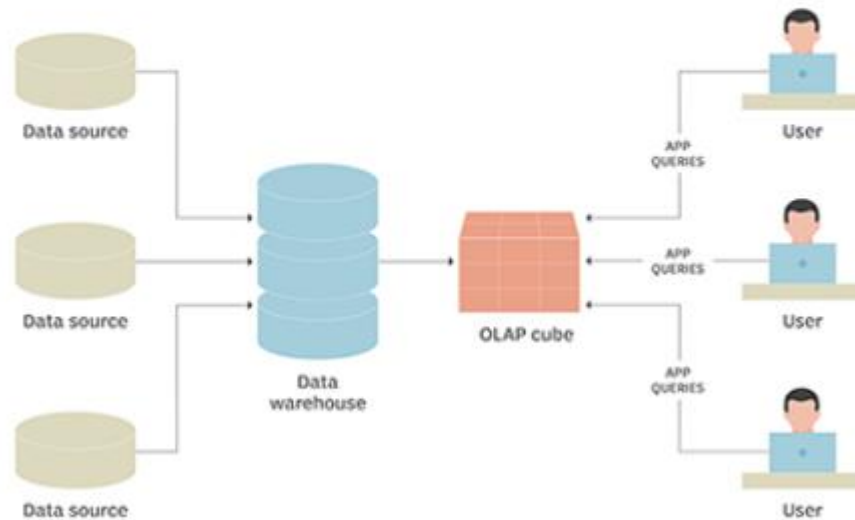


Рис. 2.1. OLAP процес

Наступним концептом розробки архітектури системи є процес інтеграції даних у сховище.

ETL (Extract, Transform, Load) та ELT (Extract, Load, Transform) є двома ключовими підходами до інтеграції даних, що застосовуються в системах бізнес-аналітики та обробки великих обсягів інформації. Обидва методи мають спільну мету - перенесення даних із джерел до сховища даних (Data Warehouse), для подальшого аналізу. Головна відмінність між підходами полягає в тому, коли і де відбувається трансформація даних: ETL виконує цю операцію до завантаження у сховище, тоді як ELT — після завантаження. Можемо розглянути кожен із цих підходів детальніше.

Якщо детальніше, **ETL** здійснює підготовку даних для подальшої аналітики одразу після її вивантаження. Для цього система отримує інформацію

з різних баз, очищає від помилок, призводить до єдиного формату та рівня деталізації. Таким чином, збагачені дані, що одержуються з різних джерел, наводяться до єдиного формату, що дозволяє ефективно з ними взаємодіяти. Тільки після всіх цих маніпуляцій інформація надходить до цільової репозиторії і стає доступною для вивчення з використанням технологій BI та data science [19].

Але також у цього підходу є видимі недоліки.

1. Значні часові витрати. Для консолідації даних необхідно розробити єдині вимоги, які мають бути застосовні до різних джерел інформації. Їх використання може розтягнутися на тижні і навіть місяці. Дуже багато часу може знадобитися на оновлення цільового репозиторію.

2. Відсутність гнучкості. Це є проблемою, бо конвєси даних спочатку проектуються відповідно до поточної специфіки джерел інформації та цілей користувача, але при збільшенні кількості джерел даних або появі нових варіантів їх використання, необхідно заново модернізувати систему.

3. Залежність передачі від коректності процесів їх перетворення. Якщо будуть помилки на одному з підготовчих етапів вивантаження інформації зупиняється [19].

ELT є новим підходом до підготовки даних. У наш час швидкість аналізу стає важливою перевагою, тому впровадження ELT процесів стає більш актуальним.

Прискорення процесу передачі досягається з допомогою оптимізації підходу внаслідок зміни послідовності операцій. Спочатку дані виймаються та завантажуються, а лише після обробляються. У процесі їх трансформації задіяні хмарні технології, що дозволяє оптимізувати швидко і непомітно для користувача [19].

Особливістю **ELT** є також поступове опрацювання інформації, що здійснюється в міру формування запитів. При цьому користувач може

використовувати єдиний репозиторій для різних програм. На відміну від процесу підготовки інформації в конвеєрі, вихідні дані не будуть втрачені – для вирішення конкретного завдання створюється репліка. Наявність подібної проміжної бази даних дозволяє повертатися до вивчення вихідної інформації неодноразово, не вносячи технічних змін до **ELT**-системи [19].

Головна відмінність між ETL та ELT полягає у послідовності операцій та розподілі навантаження між системами. У ETL трансформації виконуються до завантаження у сховище, що забезпечує підготовлені дані, але потребує більше часу для обробки. У ELT сирі дані одразу завантажуються у сховище, а трансформації виконуються за потреби, використовуючи ресурси сховища.

ETL:

- Дані трансформуються перед завантаженням, що робить їх готовими до використання одразу після імпорту у сховище.
- Підходить для традиційних сховищ даних, які мають обмежені обчислювальні можливості.
- Процес є більш контрольованим, але менш гнучким для динамічних бізнес-запитів.

ELT:

- Сирі дані зберігаються у сховищі, що дає більше можливостей для роботи з ними, але вимагає потужного сховища.
- Використовується у хмарних рішеннях, які можуть масштабувати обчислювальні ресурси.
- Забезпечує гнучкість і швидкість при роботі з великими обсягами даних.

Саме через переваги ELT процесу його і було обрано для інтеграції даних у рамках даної розробки та дослідження.

І остання важлива концепція зветься Modern Data Stack (MDS) підхід. Вона загально описує побудову аналітичної інфраструктури, яка націлена на швидкодію прийняття рішень.

Modern Data Stack – це концепція, особливістю якої є використання аналітичних інструментів та технологій з відкритим кодом, які допомагають створити сучасну, мінливу та набагато більш незалежну структуру даних.

Основними складовими такого підходу є:

1. Збір даних у поєднанні з використанням сучасних інструментів для ETL/ELT процесу щоб інтегрувати дані у сховище.
2. Трансформування даних у потрібну структуру. Наприклад, інструментом dbt.
3. Зберігання даних. Для цього використовують найчастіше хмарні рішення.
4. Аналіз та візуалізація. Для цього найчастіше використовують Tableau або PowerBi.
5. Автоматизація процесів.

Тобто, можемо підсумувати, що головною перевагою цього підходу є швидку інтеграцію даних та їх аналіз. Це дозволяє компаніям ефективно масштабувати свої рішення без великих витрат.

2.2 Використані технології

Для реалізації сховища даних було обрано PostgreSQL.

PostgreSQL — це потужна об'єктно-реляційна база даних з відкритим вихідним кодом, активна розробка якої триває понад 35 років, завдяки чому вона заслужила міцну репутацію надійності, надійності функцій і продуктивності[21].

Перевагами PostgreSQL для аналітики виступають його основні аналітичні модулі, які допомагають розподілити навантаження на сховище і виконувати складні аналітичні запити швидше [22].

Також з переваг можна назвати легке масштабування через реплікації, шардінг тощо для розподілення навантаження на різні сервера.

Також варто зауважити, що для підприємства було важливо не витратити багато коштів на аналітичну інфраструктуру, що неможливо із використанням хмарних технологій.

У контексті аналітичного сховища даних PostgreSQL був вибраний кількома факторами. Одним із них було бажання підприємства зберегти свою інфраструктуру в межах однієї технології. А оскільки Postgres пропонує також і модулі для аналітики, які на поточних вимірах даних працюють адекватно, було вирішено зупинитися саме на цьому варіанті [23].

Наступним використаним інструментом став Airbyte.

Airbyte – це проект із відкритим вихідним кодом, який стрімко набирає популярності. Проект доступний на GitHub (3.800+ stars), а спільнота у Slack налічує 2.500+ осіб. По суті, це сучасний стандарт для вибудовування потоків інтеграції даних з різних додатків, баз даних і API в аналітичні сховища даних, озера даних. Нижче я коротко розгляну ключові переваги інструменту [24].

Великий набір конекторів, доступних для підключення за лічені хвилини. У списку всі найпопулярніші СУБД, а також безліч популярних сьогодні додатків: Intercom, Zendesk, Stripe, Salesforce, Jira. Зусиллями спільноти користувачів список конекторів постійно зростає. Додавання нових конекторів зведено до простого конфігурування – оркестрацією та способами реплікації займається Airbyte [26].

Однією з ключових переваг Airbyte є його гнучкість. Сервіс підтримує роботу із багатьма конекторами, які допомагають легко підключитися до будь-якої бази даних, CRM системи або навіть Google Analytics. Також є можливість імпорту навіть файлів.

Сам же процес роботи Airbyte базується на принципі ETL або ELT, залежно від потреб користувача або підприємства. На етапі Extract дані отримуються з джерел через відповідні конектори. Load процес може бути реалізований у сирому вигляді для подальшої трансформації у сховищі (ELT) або з попередньою обробкою [26] [27].

Ще важливою перевагою платформи є налаштування типів синхронізації. Доступні варіанти інкрементального оновлення за ключем, повного оновлення таблиці тощо. Також можна виставити cron для запуску процесів.

Найголовніше є те, що це повністю open-source проект, а це означає, що сервісом можна користуватися безкоштовно.

Для моделювання даних у сховищі вибрано dbt.

DBT (data build tool) — це інструмент, покликаний додати елементи програмної інженерії в роботу з даними (SQL) [27]. Основна мета dbt полягає у впровадженні методів CI/CD процесів в роботу з даними, таких як контроль версій, модульність, повторне використання коду, автоматизація.

Даний інструмент вирішує такі проблеми:

1. **Система контролю версій.** DBT зберігає проект у репозиторії, де ви описуєте моделі даних (скажімо, таблиць). Те, що проект перебуває в репозиторії, дає можливість контролювати зміни, проводити код-рев'ю.

2. **Генерація SQL.** Моделі, описані SQL-кодом, в якому також може бути використана Jinja. Якщо коротко, то в нас з'являються змінні, цикли, галуження та функції.

3. **Data Lineage (граф залежності).** У моделях, залежних від інших, ви вказуєте назву моделі, а не «хардкодите» таблицьку/представлення. Таким чином DBT знає граф залежностей і може збирати таблицьки послідовно.

4. **Інкрементальне оновлення.** DBT вміє оновлювати таблицьки повністю і має два типи інкрементального оновлення: «видалити та замінити новим»

(наприклад, партицію за певну дату) та «оновити за ключем» (наприклад, рядок для певного користувача).

5. Тести на якість даних. У DBT є вбудовані тести для певних колонок. Зокрема, `unique`, `not_null`, допустимі значення, `referential integrity` тощо. Крім того, можна написати SQL-тест, який буде перевіряти специфічні сценарії.

6. Документація. На базі моделей, їхніх залежностей, опису, тестів і структури DWH генерується гарна HTML-сторінка. А якість заповнення документації можна контролювати на етапі підтвердження `merge request` [27].

Dbt працює з даними, які вже завантажені в сховище (наприклад, Snowflake, BigQuery чи Redshift), і трансформує їх за допомогою SQL-запитів. Це означає, що він реалізує принципи підходу ELT, коли всі трансформації виконуються безпосередньо в сховищі. Dbt не переносить дані між системами, а фокусується на забезпеченні трансформації всередині сховища.

Таким чином, dbt є фаворитом у світі інструментів для моделювання даних. Саме тому було обрано саме цей інструмент.

Для візуалізації даних обрано Microsoft PowerBi.

PowerBi – це хмарна уніфікована платформа для бізнес-аналітики. Вона допомагає перетворити інформацію з різних джерел даних у візуальні інтерактивні дашборди, що дозволяє відстежувати ключові показники бізнесу в одному інтерфейсі [29].

Важливою, на мій погляд, перевагою Power BI є вбудований інструмент Power Query, який дає змогу виконувати всі необхідні перетворення даних. В аналітиці дані часто є "брудними" - вони можуть містити пропуски, дублікати або різні невідповідності форматів. Завдяки інструменту Power Query можна швидко очистити інформацію, трансформувати її в потрібний формат і створити готовий набір даних для подальшого аналізу. Наприклад, у рамках дослідження було

використано Power Query для витягнення та очищення непотрібних даних для звіту.

Окремо варто згадати, що у Power BI можна гарно моделювати дані. Тут важливо враховувати складність кількості зв'язків між таблицями даних. За допомогою моделювання у Power BI можна встановити зв'язки між таблицями і при цьому створити логічну модель, яка дозволить проводити складні аналізи.

Power BI також має потужні можливості для візуалізації даних та їх динамічності.. У контексті цього дослідження інтерактивні дашборди, які були розроблені для підприємства, містили динамічні фільтри, які дають можливість досліджувати дані під різними кутами.

Крім того, Power BI ідеально інтегрувався з іншими інструментами, які було обрано для розробки. Наприклад, з dbt та Airbyte. Дані, які трансформуються в dbt по схемі передаються до Power BI для створення фінальних звітів і дашбордів.

2.3 Моделювання сховища даних

Сховище даних або корпоративне сховище даних (EDW) — це система, яка об'єднує дані з різних джерел в єдине центральне узгоджене сховище даних для підтримки аналізу даних, інтелектуального аналізу даних, штучного інтелекту (AI) і машинного навчання. Система сховища даних дозволяє організації запускати потужну аналітику величезних обсягів історичних даних у спосіб, який не може зробити стандартна база даних [3].

Схему вітрини даних представлено на Рис 3.

Для сховища даних було виділено такі виміри:

1) dim_product – містить інформацію про товар, його підкатегорію, категорію, розділ

2) dim_date – містить інформацію про дати, рік, квартал, номер тижня, місяць, день тижня

3) dim_user – містить інформацію про користувача та його рекламні мітки

4) dim_status_order – містить інформацію щодо статусу замовлень

А також створено таблицю фактів fact_order. Вона містить ключі із таблиць product, dim_date, user_dim та значення статусу, к-сть, пропоновану ціну та суму замовлення. На рисунку 3 зображено схему сховища.

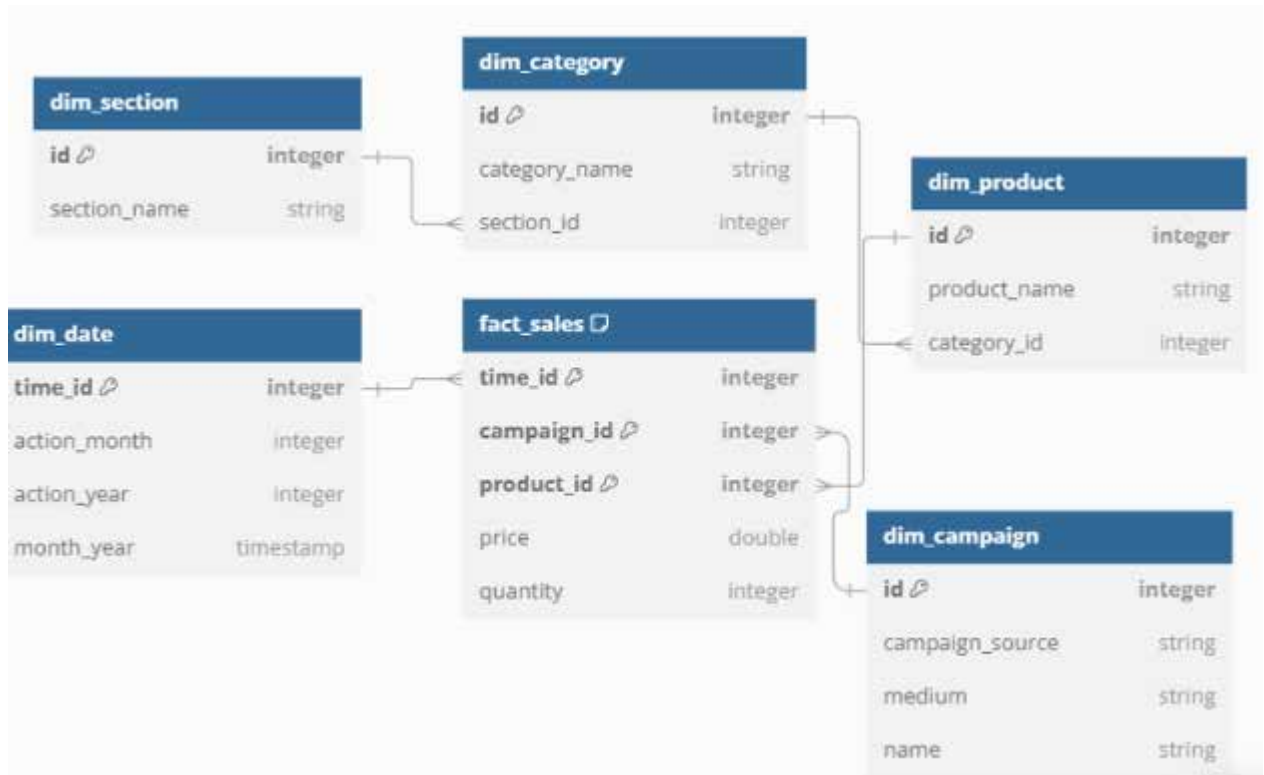


Рис. 3 Вітрина даних

Висновки до Розділу 2. Результатом Розділу 2 є розгляд основних концепцій розробки аналітичних систем, визначено такі поняття як сховище даних, OLAP, Modern Data Stack, ETL/ELT. Також описані використані технології з детальним поясненням причин їх вибору. І на завершення, була змодельована схема сховища даних. У Розділі 3 буде опис фактчної розробки аналітичної системи за допомогою технологій і підходів, які описані у Розділі 2.

3. РОЗРОБКА АНАЛІТИЧНОЇ СИСТЕМИ

3.1 ELT процесинг

Для забезпечення ефективного зберігання та обробки великих обсягів даних, ми вирішили використовувати колонкову базу даних. Основною причиною цього вибору стала висока швидкість виконання аналітичних запитів та можливість ефективного стиснення даних, що забезпечує колонкова архітектура.

ELT процесинг складався з наступних етапів:

1. **Витягнення даних (Extract)** - дані витягувалися з репліки бази даних. Основні таблиці, які підлягали переносу, включали дані про користувачів, товари, підкатегорії, категорії, розділи, статуси замовлень тощо. Використання репліки, у свою чергу, забезпечувало відсутність навантаження на основну базу даних, що було критично важливим для забезпечення стабільності роботи системи.

2. **Завантаження даних (Load)** - дані, які витягували з оперативної бази даних, завантажували у сховище даних. На цьому етапі було використано сервіс Airbyte, що забезпечило автоматизації та автоматичну нормалізацію даних.

3. **Трансформація даних (Transform)** - передбачає трансформацію даних у бізнесову структуру. Так як дані з оперативної Бд завантажуються у сирому вигляді, необхідно їх представити у формі сховища даних. Також на цьому етапі усі дані перетворюються у потрібний вигляд.

Основними таблицями, які були перенесені з оперативної бази даних у сховище даних, були:

- **Таблиця користувачів** – вона містить інформацію про зареєстрованих користувачів, їхні профілі покупок, їхня активність та історію покупок.

– **Таблиця товарів** включає дані про всі доступні товари, їхні характеристики, ціни та доступність на складі.

– **Таблиця підкатегорій** містить інформацію про підкатегорії товарів та ключі з їхніми категоріями.

– **Таблиця категорій** включає дані про основні категорії товарів.

– **Таблиця розділів** містить дані про різні розділи товарів

– **Таблиця статусів замовлень** включає у себе інформацію про статуси замовлень

На Рис можемо побачити запущені джоби для трансферу даних.

Job History		Cancel Sync	102 jobs
All jobs	All jobs to date		
Sync Running		10:33AM 07/12/2024	
Sync Succeeded	2.44 MB 12,182 records extracted 12,182 records loaded 1m 49s	4:45PM 07/05/2024	
Sync Succeeded	25.65 KB 97 records extracted 97 records loaded 1m 40s	3:45PM 07/05/2024	
Sync Succeeded	8.38 KB 26 records extracted 26 records loaded 1m 34s	2:45PM 07/05/2024	
Sync Succeeded	11.48 KB 37 records extracted 37 records loaded 1m 57s	1:45PM 07/05/2024	
Sync Succeeded	180.91 KB 795 records extracted 795 records loaded 1m 50s	12:29PM 07/05/2024	
Sync Succeeded	1.46 MB 4,907 records extracted 4,907 records loaded 1m 24s	11:29AM 07/05/2024	

Рис. 3.1 Джоби для трансферу даних

Подальшим кроком став вибір типу синхронізації даних.

Саме тому і варто зазначити, що існують такі методи синхронізації даних:

1. Повне оновлення (Full Refresh). У цьому режимі дані з джерела повністю перезаписують існуючі дані у сховищі. Кожен раз, коли запускається синхронізація, Airbyte витягує усі існуючі дані з таблиці та перезаписує її.

2. Інкрементальне оновлення (Incremental Sync). В цьому режимі синхронізуються лише ті зміни, які відбулися за час останнього завантаження. Цей підхід ефективніший за попередній.

3. Додавання (Append). При цьому підході завантаження відбувається усіх нових записів без заміни або оновлення старих даних. Це є корисним для потокових даних, де нам не потрібно змінювати усі наявні записи.

На рис 3.2 представлено частину таблиць, які були перенесені. Можемо також бачити, що у Airbyte можна налаштовувати тип синхронізації. Для більшості таблиць було налаштування інкрементальне оновлення, що дозволяє не оновлювати усі дані відразу, а за хеш-ключем здійснювати пошук останніх змін та дозаписувати інші дані.

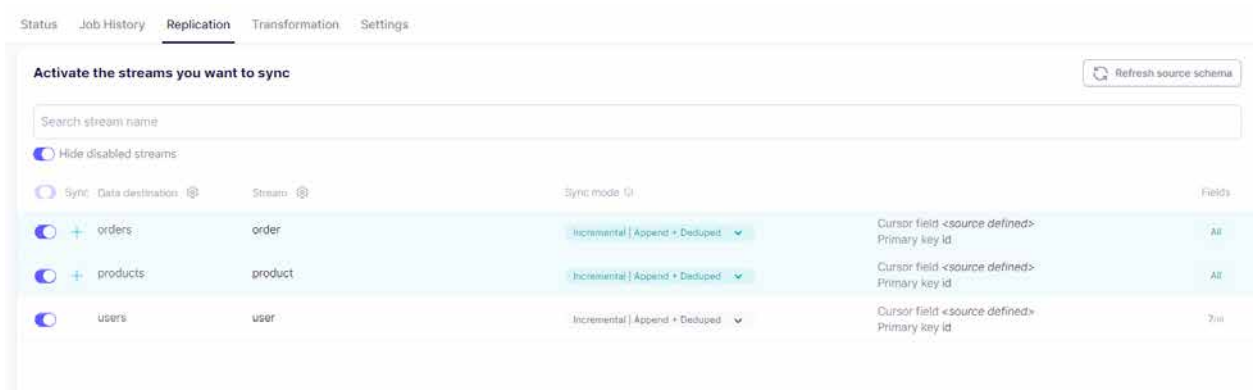


Рис. 3.2 Таблиці, які були перенесені

Можна із впевненістю сказати, що впровадження ELT процесів за допомогою Airbyte дозволило значно автоматизувати перенесення даних та забезпечити їхню актуальність у сховищі.

3.2 Моделювання даних

Наступним етапом було створення моделей даних та загальне моделювання даних у вітрині даних. На цьому етапі однією з ключових задач було

моделювання даних за допомогою інструменту dbt (Data Build Tool) для того, аби забезпечити високої якості структурованості даних.

Dbt – це відкритий інструмент для трансформації даних, який допомагає нам створювати, тестувати та автоматизувати процеси побудови моделей даних у сховищах даних.

Вибір цього інструменту обґрунтований його здатністю спрощувати у певній мірі автоматизацію та обробку великих обсягів даних, а також цей інструмент надає легко впроваджувати зміни у бізнес-логіку без потреби переписувати багато запитів, особливо якщо вони мають залежності.

Сам процес моделювання даних за допомогою dbt складався з наступних етапів.

1. **Імпорт даних** - спочатку було імпортовано дані з оперативного сховища, яка була заповнена в результаті ETL процесів. Основні таблиці, що підлягали моделюванню, включали дані про користувачів, товари, підкатегорії, категорії, розділи, статуси замовлень і тд.

2. **Створення моделей** - було створено моделі для кожної з основних таблиць. Ці моделі дозволяли структурувати дані та підготувати їх для подальшого аналізу. В процесі створення моделей було використано SQL-запити для трансформації даних.

3. **Побудова шарів трансформацій** - для створення правильної структури треба було створити кілька шарів трансформацій, щоб забезпечити поступове вдосконалення даних. Перший шар включав базові трансформації для підготовки сирих даних. Другий шар включав більш складні трансформації, такі як об'єднання даних з різних джерел та обчислення агрегованих показників. Третій шар включав кінцеві моделі, що використовувалися для аналітики та звітності.

4. Після побудови моделей dbt надає можливості для автоматичного документування моделей. Це дозволило створити детальну документацію для кожної таблиці та трансформації. Це значно спростило розуміння структури даних та процесу їхньої обробки.

5. Останнім та важливим етапом процесу моделювання було тестування даних у моделях. Було створено тести для перевірки цілісності та коректності даних, що дозволило виявити та виправити можливі помилки на ранніх етапах.

Основні таблиці, які були змодельовані за допомогою dbt, включали:

- 1) таблиця користувачів;
- 2) таблиця товарів;
- 3) таблиця підкатегорій;
- 4) таблиця розділів.

У методології dbt (Data Build Tool) існує концепція різних шарів або ступенів, які допомагають організувати та керувати процесом обробки даних.

Основні шари включають:

Base Models - тут відбувається очищення та підготовка сирих даних, що надходять з джерел. Ці моделі часто використовуються для переіменування колонок, видалення дублікатів, перетворення даних у відповідний формат тощо.

У base шарі було створено базові моделі кошика, логів замовлення, таблиці замовлення, історії цін та даних про користувача.

Stage Models - використовується для завантаження даних з базових моделей в проміжні структури, які дозволяють виконувати складніші операції трансформації. Вони можуть включати об'єднання даних з різних джерел або агрегацію даних для подальшого аналізу.

На цьому етапі створено моделі обробки замовлення, пропонованої ціни, моделі об'єднання товарів, підкатегорій, категорій, розділів та даних про користувачів.

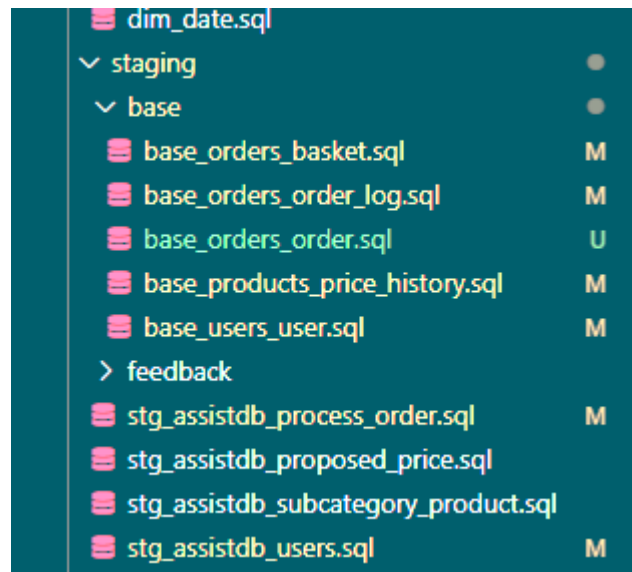


Рис. 3.3 Моделі рівня Base та Stage

Intermediate Models – тут відбувається виконання складніших операцій трансформації, таких як обчислення показників або створення агрегованих таблиць. Проміжні моделі можуть бути базою для побудови кінцевих аналітичних моделей.

На цьому етапі створено 3 моделі.

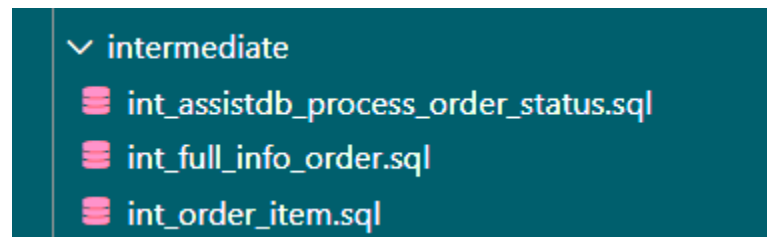


Рис. 3.4 Моделі рівня Intermediate

Marts Цей шар включає кінцеві моделі, які використовуються для зберігання готових до аналізу даних. Marts даних містять оптимізовані структури для швидкого доступу до інформації та виконання аналітичних запитів.

Було створено основні моделі, які будуть використовуватися у вітрині даних для візуалізації.

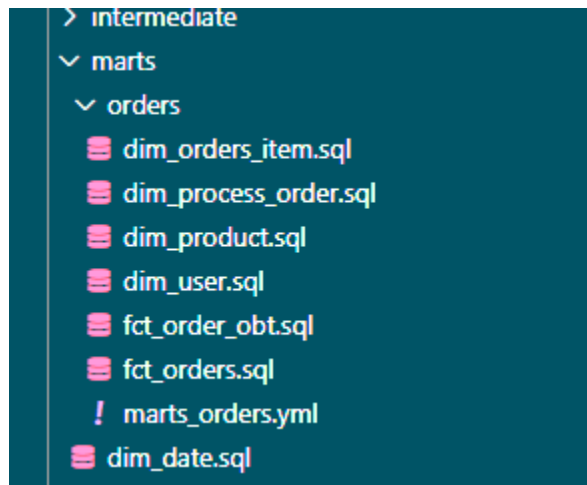


Рис.3.5 Моделі рівня Marts

Моделювання даних за допомогою dbt дозволило створити структуровану та якісну вітрину даних, що значно полегшило доступ до інформації та її аналіз. Використання dbt для трансформації та моделювання даних забезпечило високу ефективність та точність обробки даних, що позитивно вплинуло на прийняття рішень та покращення бізнес-процесів.

3.3 Побудова звітності в середовищі Power BI

Microsoft PowerBi – це платформа для бізнес-аналізу від компанії Microsoft. Вона дозволяє її користувачам збирати, обробляти, аналізувати та візуалізувати дані з різних джерел. Це допомагає у свою чергу приймати обґрунтовані бізнес-рішення на основі аналітики.

У свою чергу, звітність у PowerBi дає змогу створювати різні типи візуалізацій (графіки, діаграми, таблиці тощо). Вони допомагають краще візуалізувати та сприймати інформацію та бізнес-інсайти.

Для створення звітності у Power BI спочатку потрібно було імпортувати уже готові дані із сховища даних. Було підключено нову вітрину даних до PowerBi. На етапі завантаження PowerBi дає можливість виконати очищення та обробку даних, видалення дублюючих записів і заповнення відсутніх значень. Також на цьому етапі створюються нові колонки та міри для подальшого аналізу.

Наступним кроком є створення візуалізацій. Для цього використовуються різні типи діаграм та графіків, зокрема стовпчасті діаграми, кругові діаграми, лінійні графіки та індикатори. Налаштовуються фільтри для відображення даних за категоріями продуктів і створюються індикатори темпу росту продажів.

Для побудови графіку суми продажів за періодом необхідно спочатку створити міру, відносно якої будемо переглядати інформацію. У нашому випадку це поле «Сума продажів».

Міра – це динамічно обчислювальне значення, яке написано за допомогою мови DAX. Міра надає можливість виконувати розрахунки на основі даних, які зберігаються у моделях даних. Також основна перевага мір у тому, що їх відразу можна агрегувати в залежності від контексту – тобто значення міри може змінюватися від фільтра, який на неї впливає.

Нижче зображено формулу DAX міри для обрахунку суми продажів.

$$\text{SumSales} = \text{SUM}(\text{fact_sales}[\text{sum_price}]), \quad (3.1)$$

де fact_sales – таблиця фактів,

Sum_sales – поле із сумою продажів

Наступним кроком є саме створення графіку суми продажів за періодами. Для цього у панелі PowerBi треба вибрати тип діаграми «Графік» та перенести туди період та суму продажів, як це показано на Рис 3.6.

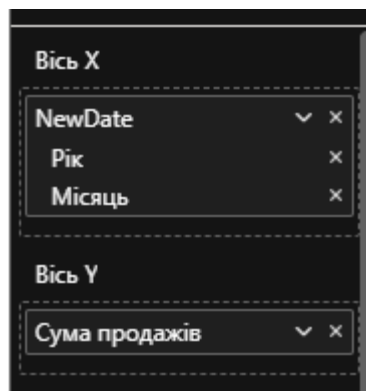


Рис. 3.6 Налаштування графіку суми продажів

І в результаті, на Рис 3.7 представлено побудований графік, який відображає суму продажів за місяцями.

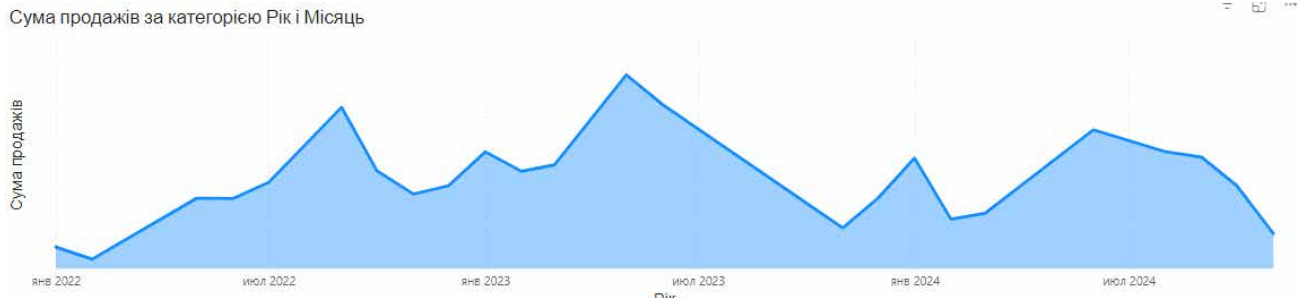


Рис. 3.7. Виручка за кожним місяцем

Бачимо, що з початку 2022 до середини 2024 року сума продажів має зростаючий тренд із кількома піками, але після цього можемо побачити невеликий спад. Це можна пояснити нижчою рекламною активністю та сезонністю.

Це також вказує на успішність рекламних кампаній в окремі періоди, але й також на проблеми з утриманням клієнтів: залучаємо нових, але не повертаємо старих, що заважає стабільному зростанню продажів. Виявлені піки у 2023 та 2024 роках свідчать про сезонність і це в свою чергу вказує на потенційні точки росту.

На наступному Рис 3.8 представлено помісячний розподіл суми продажів у розрізі рекламних кампаній. Тут варто зауважити, що аналізується 2023-2024 роки. Бачимо, що спостерігається певна сезонність по каналу трафіку Facebook за березень-червень. Але загалом, потрібно також враховувати і те, що запуск рекламних кампаній також не є стабільним і не постійним.

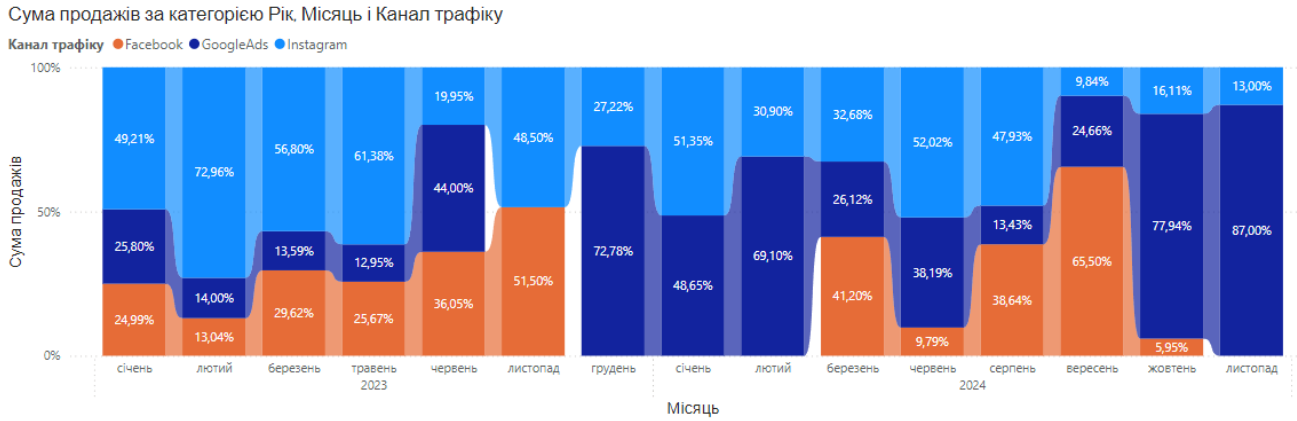


Рис. 3.8 Розподіл виручки за каналами трафіку за місяцями

Наступний графік на Рис 3.9 уже вказує на фактичний розподіл продажів за каналами трафіку. З графіку бачимо, що майже половину виручки приходить із продажів клієнтам, які прийшли через Instagram (44%), далі йде GoogleAds 33% та Facebook 24%.

Розподіл продажів за каналом трафіку

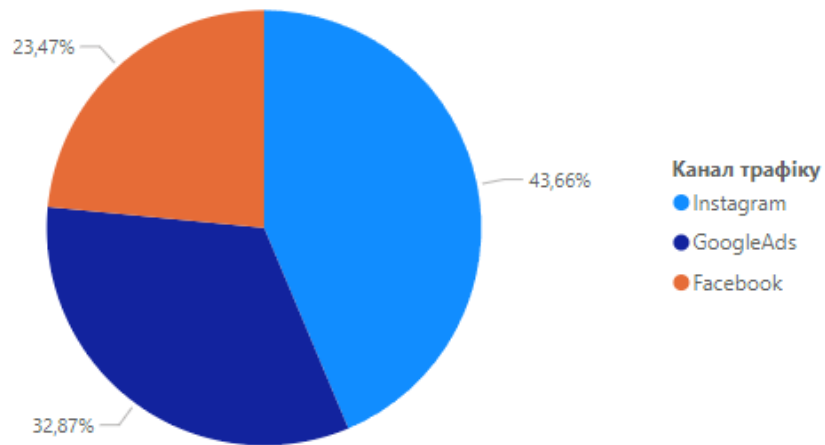


Рис. 3.9 Розподіл продажів за каналом трафіку

Наступний графіку (Рис 3.10) показує розподіл продажів за каналами трафіку та категорією товарів. Можемо бачитт, що користувачі з каналу трафіку

Instagram купують найбільше з категорій Gaming, Ноутбуки та Монітори, з GoogleAd розподіл вказує на більшу зацікавленість у категорії Gaming (це пов'язано із націленістю рекламних кампаній). Найбільші % продажів у категорії Ноутбуки з трафіку Facebook .

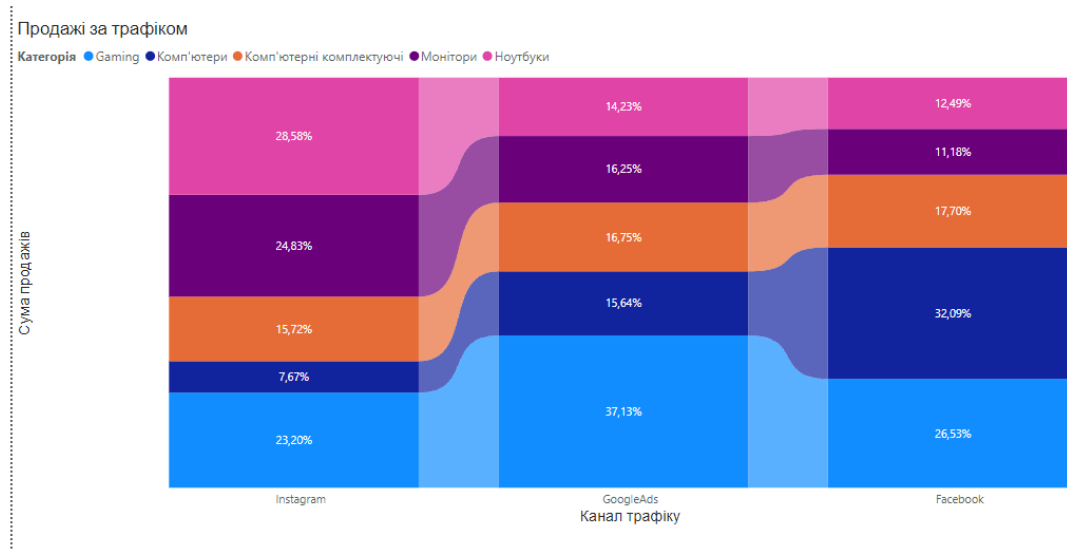


Рис. 3.10 Розподіл продажів за каналами трафіку та категорією товарів

3.4 Аналіз даних шляхом 1Rule алгоритму

Алгоритм **1R (One Rule)** є простим та інтерпретованим методом машинного навчання, який використовується для задач класифікації. Цей алгоритм отримав свою назву через те, що він генерує лише одне правило для класифікації даних, обираючи для цього найкращу ознаку. Завдяки своїй простоті, 1R часто використовують як базову модель для порівняння з іншими складнішими алгоритмами класифікації.

Алгоритм 1R починає з вибору однієї ознаки (фічі) з набору даних і перевіряє, як вона може допомогти класифікувати цільову змінну (мітку класу). Якщо ознака є числовою, алгоритм може здійснити її дискретизацію, перетворивши значення на категорії. Далі для кожного унікального значення цієї ознаки створюється правило, яке асоціює це значення з найбільш поширеним класом серед усіх об'єктів, що мають таке значення цієї ознаки.

Алгоритм використовує ці правила для оцінки того, як добре кожна ознака класифікує дані, і підраховує, скільки разів правило правильно класифікує екземпляри. В кінцевому підсумку вибирається правило, яке дає найменший рівень помилок, тобто правило з найменшою кількістю неправильних класифікацій. Тому 1R вибирає ту ознаку, яка має найкращу здатність до класифікації серед усіх ознак.

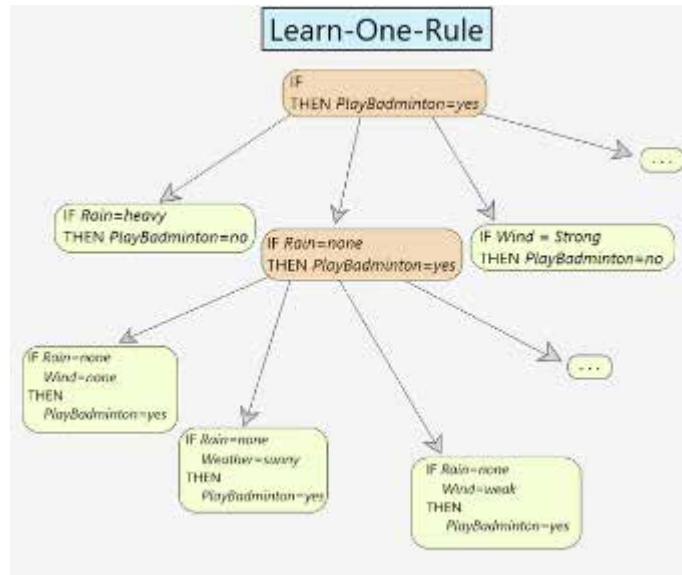


Рис. 3.11. Алгоритм виконання 1Rule

Після створення правила для вибраної ознаки алгоритм використовує це правило для класифікації нових даних. В нових екземплярах алгоритм просто перевіряє значення вибраної ознаки і прогнозує клас, асоційований з цим значенням.

Для аналізу даних за допомогою 1Rule алгоритму було використано у середовищі PowerVi.

Для реалізації цього алгоритму спочатку потрібно було створити розподіл даних по продажах за категоріями та кампаніями. Для цього було створено розподіл продажів за категоріями у DAX, що відображено формулою нижче.

$$\text{TotalSales} = \text{SUM}(\text{fact_sales}[\text{sum_price}]), \quad (3.2)$$

де fact_sales – таблиця фактів,

Sum_sales – поле із сумою продажів

Далі було створено міри для розподілу. Вона необхідна для визначення розподілу продажів за категоріями товарів:

$$\text{Sales Distribution by Category} = \text{DIVIDE}(\text{SUM}(\text{fact_sales}[\text{sum_price}]), \quad (3.3) \\ \text{CALCULATE}(\text{SUM}(\text{fact_sales}[\text{sum_price}]), \text{ALL}(\text{dim_category})), 0)$$

де fact_sales – таблиця фактів,

sum_sales – поле із сумою продажів,

dim_category – таблиця категорія товарів

На Рис. 3.12 можемо бачити розподіл продажів за категорією товарів. Виходить, що 29% продажів приходять з категорії Gaming, 20% з категорії Ноутбуки, 19% з категорії Монітори, 16% з категорії Комп'ютерні комплектуючі.

Розподіл продажів за Категорією товарів

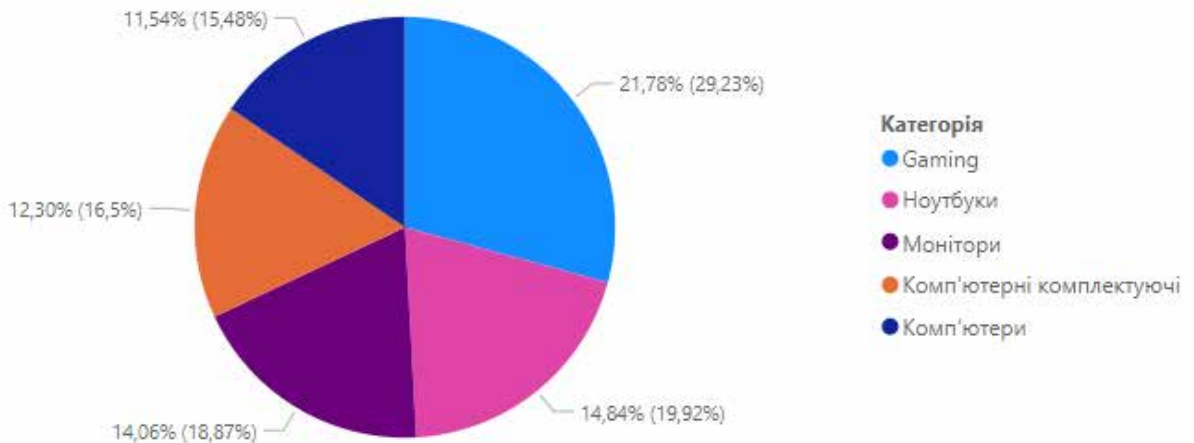


Рис. 3.12 Розподіл продажів за категорії товарів

Наступним кроком за правилом 1Rule було визначено клас ціни «Високі продажі» та «Низькі продажі».

Категорію ціни було визначено відносно медіани суми продажів за параметрами, обрахунок цього параметру можна побачити на формулі нижче.

$$\text{SalesCategory} = \text{VAR MedianPrice} = \text{MEDIAN}(\text{Sales}[\text{sum_price}]) \quad (3.4)$$

$\text{RETURN IF}(\text{Sales}[\text{sum_price}] > \text{MedianPrice}, \text{"High Sales"}, \text{"Low Sales"})$

де fact_sales – таблиця фактів,

MedianPrice – медіанна ціна продажів

sum_sales – поле із сумою продажів

Якщо сума продажів у розрізі категорії більша медіанного значення від усіх продажів, ми визначаємо цей клас як «Високі продажі», якщо менше – «Низькі продажі».

На рис. 3.13 бачимо результат обрахунків, таблиця представляє собою розподіл категорії товарів, каналів трафіку, суми продажів та визначені категорії продажі.

Категорія	Канал трафіку	Категорія продажів	Сума продажів
Gaming	Facebook	Low Sales	71 471,61
Комп'ютери	Facebook	High Sales	91 382,71
Комп'ютерні комплектуючі	Facebook	High Sales	96 813,75
Монітори	Facebook	Low Sales	87 920,76
Ноутбуки	Facebook	Low Sales	87 978,56
Gaming	GoogleAds	High Sales	92 913,35
Комп'ютери	GoogleAds	Low Sales	83 994,76
Комп'ютерні комплектуючі	GoogleAds	Low Sales	85 015,58
Монітори	GoogleAds	High Sales	89 660,10
Ноутбуки	GoogleAds	High Sales	101 550,48
Gaming	Instagram	High Sales	95 185,38
Комп'ютери	Instagram	Low Sales	88 797,58
Комп'ютерні комплектуючі	Instagram	Low Sales	77 789,09
Монітори	Instagram	High Sales	91 492,76
Ноутбуки	Instagram	High Sales	92 568,88
Усього			1 334 535,35

Рис. 3.13 Розподіл категорії продажів за категорією товарів та каналом трафіку

Далі було обраховано дві міри, які рахують кількість категорій, які мають високі та низькі продажі. Це необхідно, щоб визначити подальшу ймовірність в розрізі окремо каналів трафіку та категорій товарів.

$$\begin{aligned} \text{High Sales} = & \text{CALCULATE(COUNTROWS('Sales'),} & (3.5) \\ & \text{'Sales'[SalesCategory] = "High Sales")} \end{aligned}$$

де Sales – таблиця продажів,

Salescategory – категорія продажів (висока, низька),

Нижче також обрахунок категорій товарів з позначкою «Низькі продажі»:

$$\begin{aligned} \text{Low Sales} = & \text{CALCULATE(COUNTROWS('Sales'),} & (3.6) \\ & \text{'Sales'[SalesCategory] = "Low Sales")} \end{aligned}$$

де Sales – таблиця продажів,

Salescategory – категорія продажів (висока, низька),

Формула описує обрахунок класу для параметра (категорія товару або каналу трафіку). Цим ми визначаємо найпопулярніший клас у розрізі параметра, якого ми аналізуємо.

$$\text{Class} = \text{IF}([\text{High Sales}] \geq [\text{Low Sales}], \text{"High Sales"}, \text{"Low Sales"}) \quad (3.7)$$

де High Sales – обрахована міра у формулі (3.5),

Low Sales – обрахована міра у формулі (3.6)

Після визначення класу для параметра необхідно порахувати ймовірність, з якою вибраний клас описує наші дані. Реалізація DAX формули представлена формулою нижче.

$$\text{Probability} = \text{IF}([\text{Class}] = \text{“High Sales”}, [\text{High Sales}] / ([\text{High Sales}] + [\text{Low Sales}]), [\text{Low Sales}] / ([\text{High Sales}] + [\text{Low Sales}])) \quad (3.8)$$

де [Class] – міра визначення класу, обрахована у формулі (3.7)

High Sales – обрахована міра у формулі (3.5),

Low Sales – обрахована міра у формулі (3.6)

Для оцінки якості ймовірності було обраховано помилку, DAX формули якої представлено формулою нижче.

$$\text{Mistake} = \text{IF}([\text{Class}] = \text{“High Sales”}, \text{FORMAT}([\text{Low Sales}] / \text{FORMAT}([\text{High Sales}] + [\text{Low Sales}]), \text{FORMAT}([\text{High Sales}] / \text{FORMAT}([\text{High Sales}] + [\text{Low Sales}])) \quad (3.9)$$

де Class – міра визначення класу, обрахована у формулі (3.7)

High Sales – обрахована міра у формулі (3.5),

Low Sales – обрахована міра у формулі (3.6)

На Рис. 3.14 представлена таблиця, яка відображає категорії товарів, визначено клас категорії продажів, сума продажів, ймовірність та помилка ймовірності. Можемо бачити, що товари з категорії Gaming у нашому розподілі даних більшість класів розподілу мали High Sales. За такою статистикою, ймовірність високої ціни у даній категорії складає 67% з помилкою 1/3.

У категорії «Комп'ютери» та «Комп'ютерні комплектуючі» з ймовірністю 67% клас продажів буде High Sales із помилкою 1/3.

По категорія «Монітори» та «Ноутбуки» статистика показує, що із ймовірністю 100% клас продажів буде High Sales. Це вказує нам на те, що із цих категорій користувачі купують більш дорогі товари.

Категорія	Class	Сума продажів	Probability	Mistake
Gaming	High Sales	259 570,34	66,67%	1/3
Комп'ютери	High Sales	264 175,05	66,67%	1/3
Комп'ютерні комплектуючі	Low Sales	259 618,42	66,67%	1/3
Монітори	High Sales	269 073,62	100,00%	/3
Ноутбуки	High Sales	282 097,92	100,00%	/3
Усього	High Sales	1 334 535,35	73,33%	4/15

Рис. 3.14 Ймовірність класу за категорією «Категорія товарів»

За статистикою на Рис. 3.15 , по розподілу за каналом трафіку можна побачити наступне:

1. По каналу трафіку з Facebook найчастіше категорія продажів була Low Price, ймовірність низьких продажів складає 80% з помилкою 1/5.
2. З GoogleAds клас продажів складає High Sales з ймовірністю 60% з помилкою 2/5.
3. Трафік з Instagram має High Sales з ймовірністю 80% з помилкою 1/5.

Канал трафіку	Class	Сума продажів	Probability	Mistake
Facebook	High Sales	435 567,39	80,00%	1/5
GoogleAds	High Sales	453 134,27	60,00%	2/5
Instagram	High Sales	445 833,69	80,00%	1/5
Усього	High Sales	1 334 535,35	73,33%	4/15

Рис. 3.15 Ймовірність класу за категорією «Канал трафіку»

На основі аналізу інформації за правилом 1Rule можна зробити такі висновки.

1. Ймовірність того, що продажі з категорії «Комп'ютери» та «Комп'ютерні комплектуючі» будуть **High Sales** (високий продаж), складає 67%, з помилкою 1/3. Це означає, що для цих них є значний шанс на отримання статусу високий продаж, але також є і досить велика ймовірність того, що продажі можуть бути низькими. Ці товари мають різні варіанти ціни, що впливає на частоту продажів високого та низького рівня

2. За статистикою, ймовірність того, що з категорії «Монітори» та «Ноутбуки» мають **High Sales**, складає 100%. Це вказує на те, що користувачі, які купують монітори та ноутбуки, зазвичай обирають більш дорогі товари. Можемо зробити висновок, що ці категорії товарів користуються попитом серед покупців, які готові витратити більшу суму грошей на техніку.

3. У випадку трафіку з Facebook, найбільш поширеним класом продажів є **Low Price** (низька ціна), з ймовірністю 80% і помилкою 1/5. Це каже нам про те, що користувачі, які приходять з Facebook, частіше купують дешевші товари і може свідчити про те, що аудиторія з Facebook обирає товари, які мають більш доступну вартість.

4. У випадку GoogleAds, ймовірність високих продажів **High Sales** складає лише 60%, з помилкою 2/5. Це означає, що кампанії з GoogleAds не завжди приносять очікувані високі продажі, навіть попри те, що GoogleAds генерує більше трафіку, але цей трафік менш цільовим або менш схильним до покупки дорогих товарів.

5. Трафік з Instagram має **High Sales** з ймовірністю 80%, з помилкою 1/5. Це означає, що Instagram є ефективним каналом для продажу дорогих товарів. Користувачі з цього каналу мають високу ймовірність купувати більш дорогі продукти і це свідчить про якість аудиторії.

3.5 Кластерний аналіз

Кластеризація, або кластерний аналіз - це статистична процедура, задача якої полягає у розбитті вибірки даних на певні підмножини, які не перетинаються. Ці підмножини називаються кластерами.

Кожен кластер складається зі схожих за своїми характеристиками даних, а об'єкти різних кластерів мають істотно відрізнятися один від одного. Задача кластеризації відноситься насамперед до статистичної обробки, а також до

широкого класу задач навчання без вчителя. Ще її можна описати через задачу класифікації.

Задача кластеризації - це по факту задача класифікації, бо в обох випадках ми ділимо об'єкти на основі їх подібності між собою, але у випадку кластеризації приналежність навчальних об'єктів будь-яким класам не задається.

Така задача є загальною, тому для її розв'язання використовуються різні підходи. Алгоритми побудови кластерів дуже відрізняються у підходах до того, що відносити в один кластер і як їх взагалі ефективніше шукати. Кластери можна утворювати ґрунтуючись на відстані між ними, на щільності ділянок у просторі даних, інтервалах або на конкретних статистичних розподілах. Це все залежить від конкретного набору даних та мети використання результатів. Кластерний аналіз не є автоматизованим, це скоріше ітераційний процес, тому що часто доводиться змінювати метод опрацювання даних та параметри моделі, поки не буде отримано з результат з заданими властивостями [5].

Задачу кластерного аналізу було реалізовано у PowerBi.

Для створення кластерів було обрано кампанію продажу та виділено 6 кластерів. У процесі кластеризації лежить алгоритм k-means для групування даних на основі вибраних полів.

Метод k-means для групування даних полягає у мінімізації відстані між точками для кожного кластера. У свою чергу, кожен кластер має свій центр. Точки даних на графіку відображаються як частини кластера залежно від того, до якого центру вони найближчі.

На Рис. 3.16 представлено налаштування кластерів у PowerBi.

Рис. 3.16 Налаштування кількості кластерів

Опишемо кожен кластер детальніше:

1. Кластер 1 складається з рекламних кампаній через Instagram. Це можна побачити і через більшу к-сть проданих товарів користувачам, які приєдналися через цей канал трафіку. Також цей кластер описує продажі з GoogleAds, які генерують більш менше продажів, ніж через Instagram. Також можемо сказати, що даний кластер має найвищі показник по загальній сумі продажів та по кількості проданих товарів. Це говорить про те більша частина користувачів купують дешевші, але більше товарів, за рахунок чого і така велика сума продажів

2. Кластер 2 найкраще описує продажі з GoogleAds. Тут загальні продажі значно домінують над іншими каналами трафіку, зокрема Instagram.

3. Кластер 3 представляє користувачів, які витрачають найбільше коштів та купують більше товарів. Тут найнижчі показники по загальним продажам, що вказує на те, що основний сегмент користувачів знаходиться у Кластері 1. Основну частку доходу у цьому кластері генерує Instagram, хоча і присутня доля Facebook

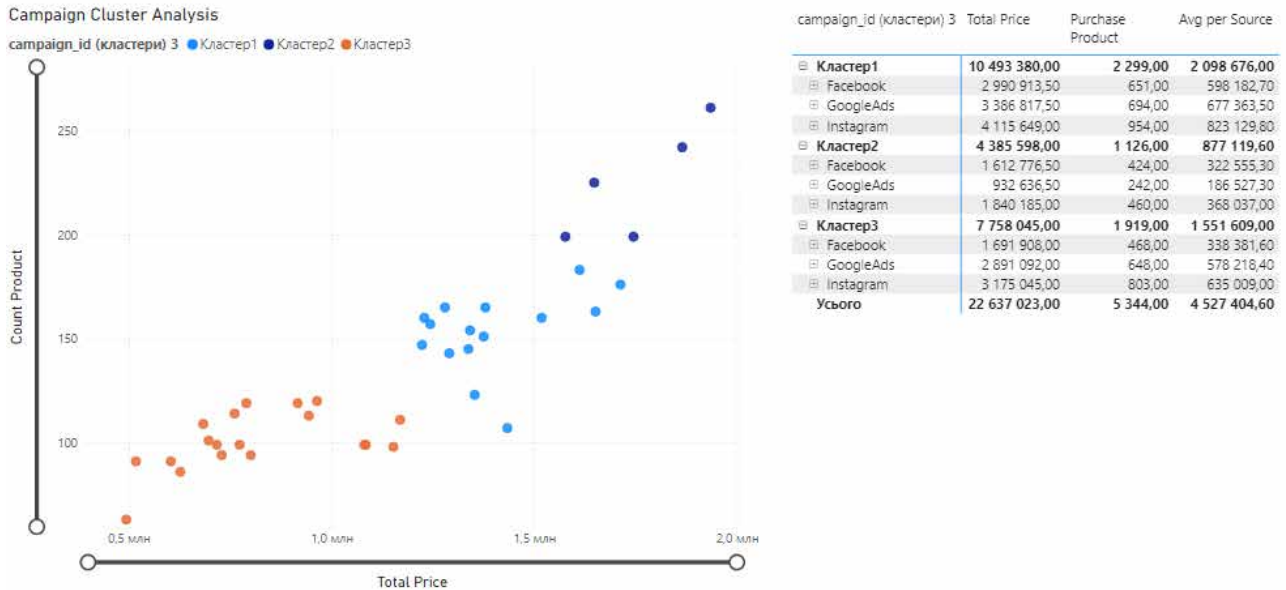


Рис. 3.17 Кластерний аналіз

На Рис. 3.18 можна побачити розподіл за категоріями товарів у Кластері 1.

Cluster	Total Price	Purchase Product	Avg per Source
Кластер1	10 493 380,00	2 299,00	2 098 676,00
Facebook	2 990 913,50	651,00	598 182,70
Gaming	750 808,50	190,00	150 161,70
Комп'ютери	843 544,00	191,00	168 708,80
Комп'ютерні комплектуючі	408 073,00	63,00	81 614,60
Монітори	440 919,50	120,00	88 183,90
Ноутбуки	547 568,50	87,00	109 513,70
GoogleAds	3 386 817,50	694,00	677 363,50
Gaming	1 005 432,50	180,00	201 086,50
Комп'ютери	631 444,00	165,00	126 288,80
Комп'ютерні комплектуючі	768 331,00	162,00	153 666,20
Монітори	429 130,00	84,00	85 826,00
Ноутбуки	552 480,00	103,00	110 496,00
Instagram	4 115 649,00	954,00	823 129,80
Gaming	1 665 720,50	365,00	333 144,10
Комп'ютери	827 989,00	192,00	165 597,80
Комп'ютерні комплектуючі	229 618,00	70,00	45 923,60
Монітори	806 151,50	201,00	161 230,30
Ноутбуки	586 170,00	126,00	117 234,00

Рис. 3.18 Розподіл продажів у Кластері 1 за категоріями товарів

За таблицею бачимо, що найбільший дохід з Instagram та GoogleAds принесла категорія товарів Gaming, а ось у Facebook ситуація інша – тут переважає категорія Комп'ютери.

Також бачимо і найменш прибуткові категорії товарів у цьому Кластері – для Instagram та Facebook це Комп'ютерні комплектуючі, для GoogleAds це Монітори.

Ноутбуки	508 170,00	126,00	117 234,00
Кластер2	4 385 598,00	1 126,00	877 119,60
Facebook	1 612 776,50	424,00	322 555,30
Gaming	434 719,50	164,00	86 943,90
Комп'ютери	313 186,00	97,00	62 637,20
Комп'ютерні комплектуючі	227 988,00	49,00	45 597,60
Монітори	299 918,00	51,00	59 983,60
Ноутбуки	336 965,00	63,00	67 393,00
GoogleAds	932 636,50	242,00	186 527,30
Gaming	368 343,00	94,00	73 668,60
Комп'ютери	161 823,00	51,00	32 364,60
Комп'ютерні комплектуючі	246 054,50	63,00	49 210,90
Монітори	127 024,50	31,00	25 404,90
Ноутбуки	29 391,50	3,00	5 878,30
Instagram	1 840 185,00	460,00	368 037,00
Gaming	624 033,50	158,00	124 806,70
Комп'ютери	321 924,50	72,00	64 384,90
Комп'ютерні комплектуючі	382 505,00	105,00	76 501,00
Монітори	267 720,00	89,00	53 544,00
Ноутбуки	244 002,00	36,00	48 800,40

Рис. 3.19 Розподіл продажів у Кластері 2 за категоріями товарів

На рис. 3.19 зображено розподіл продажів у кластері 2 за категоріями товарів. Бачимо, що найбільш прибуткові категорії товарів у всіх трьох каналах трафіку це Gaming. З найменш прибуткових можемо виділити Ноутбуки.

На рис. 3.20 представлено розподіл за кластером 3. Цей кластер є найменшим із всіх, але представляє товари, середній чек який найдорожчий. Бачимо, що з Instagram у цьому кластері це комп'ютерні комплектуючі. Більше всього це пов'язано із продажами у пік продажів відеокарт Nvidia. У інший каналах трафіку найприбутковішою категорією є Gaming.

Кластер3	7 758 045,00	1 919,00	1 551 609,00
Facebook	1 691 908,00	468,00	338 381,60
Gaming	565 526,50	178,00	113 105,30
Комп'ютери	396 776,00	86,00	79 355,20
Комп'ютерні комплектуючі	332 076,50	100,00	66 415,30
Монітори	182 330,50	38,00	36 466,10
Ноутбуки	215 198,50	66,00	43 039,70
GoogleAds	2 891 092,00	648,00	578 218,40
Gaming	856 145,00	211,00	171 229,00
Комп'ютери	592 646,50	112,00	118 529,30
Комп'ютерні комплектуючі	515 895,00	130,00	103 179,00
Монітори	340 446,50	69,00	68 089,30
Ноутбуки	585 959,00	126,00	117 191,80
Instagram	3 175 045,00	803,00	635 009,00
Gaming	565 807,50	143,00	113 161,50
Комп'ютери	537 240,50	157,00	107 448,10
Комп'ютерні комплектуючі	710 291,50	193,00	142 058,30
Монітори	664 491,00	190,00	132 898,20
Ноутбуки	697 214,50	120,00	139 442,90

Рис. 3.20 Розподіл продажів у Кластері 3 за категоріями товарів

За результатами кластерного аналізу можна надати такі висновки:

1. Необхідно сфокусуватися на кластері 1, адже тут знаходиться основний сегмент користувачів. Необхідно масштабувати кампанії, особливо через Instagram та GoogleAds, які уже показують непогані результати
2. Із другого кластеру можна посилити роботу з GoogleAds, адже цей канал приносить найбільше трафіку, хоча не є в топ 1 прибутковості.
3. Оскільки кластер 3 є найменшим по сумі продажів, слід посилити маркетингові стратегії через Instagram. Таким чином ми можемо спробувати залучити більше користувачів, які матимуть більший середній чек.

3.6 Результати дослідження та рекомендації

У результаті проведеного аналізу каналів трафіку та їх вплив на продажі можна зробити декілька важливих висновків. По-перше, за період з початку 2022 року до середини 2024 обсяги продажів показували зростаючий тренд. Це свідчить про ефективність рекламних кампаній у ті періоди. Це є хорошим і поганим фактором одночасно. Ця ситуація демонструє що сезонні коливання а

зменшення рекламної активності сильно впливають на обсяги продажів підприємства. Також, з точки зору залученості клієнтів також бачимо, що нових клієнтів вдається залучати, але старих утримувати стає все важче. Саме тому важливо у цьому аспекті приділити увагу стратегії утримання клієнтів, зокрема використовуючи ретаргетинг аудиторії.

Наступними результатами є аналіз каналів трафіку і там виявився цікавий тренд. Канал трафіку Instagram приносить найбільшу частку продажів – 43,66%, Google Ads у свою чергу забезпечує 32,87% продажів, хоча тут є з чим працювати, адже загальна конверсія низька. Facebook приносить 23,47%.

Аналіз категорій товарів показав, що користувачі з Instagram найчастіше купують товари з категорії Gaming та Ноутбуки та монітори. Це пов'язано із таргетингом реклами саме на цю тематику та акцент на молодшу аудиторію. GoogleAds також найбільше залучає у категорію Gaming і це теж пов'язано із специфікою рекламних кампаній для цієї аудиторії.

Аналіз за 1Rule показав.

1. Ймовірність високих продажів з категорії «Комп'ютери та комплектуючі» складає 67%.
2. 100% ймовірність високих продажів у категорії «Монітори та Ноутбуки» вказує на стабільний попит на ці товари, але це залежить і від рекламної кампанії.
3. Користувачі з Facebook має високий шанс на покупку дешевших товарів (80%).
4. Ймовірність високих продажів з GoogleAds складає лише 60%, що вказує на менш цільовий трафік, який не завжди призводить до покупки дорогих товарів.

5. Трафік з Instagram має 80% ймовірність високих продажів, що свідчить про високу ймовірність покупки дорогих товарів завдяки якісній аудиторії.

Загалом, можна надати такі рекомендації.

1. Варто зменшити витрати на рекламу у Facebook, адже ефективність цього каналу є більш сезонною.
2. Необхідно посилити роботу з контент-маркетингом, що дозволить залучати нових клієнтів дешевше.
3. Збільшення бюджету на ретаргетинг аудиторії.
4. Спостерігати за ростом продажів у категорії Gaming та акцентувати увагу на категоріях «Монітори та ноутбуки», адже за правилом 1Rule вони мають 100% ймовірність продажів, що свідчить про стабільний попит на даний момент.

Висновки до Розділу 3. Результатом Розділу 3 є розробка та опис процесу ELT процесу, моделювання даних та побудови звітності можливостями PowerBi. Також було проведено аналіз даних шляхом використання правила 1Rule та проведено кластерний аналіз, який допоміг визначити когорти користувачів.

ВИСНОВКИ

Результатом даного дослідження стала розробка системи аналізу продажів та аналіз рекламних кампаній і як вони впливають на продажі на конкретному підприємстві. У першому розділі було досліджено предметну область та визначено, які технології та підходи використовують інші компанії у сфері eCommerce для аналізу даних.

На початковому етапі було проведено аналіз предметної області за допомогою діаграми прецедентів, що допомогло виявити ключові процеси та взаємозв'язки в системі продажів, ставши основою для подальшого моделювання та розробки.

Для проектування системи були проаналізовані дані та їх джерела, сформовано схему побудови аналітичної системи. Результатом Розділу 1 стала побудована архітектура системи та топологія системи підприємства з імплементованим аналітичним модулем.

У Розділі 2 було детально описано методи та технології створення аналітичної системи. Було виділено основні концепції, такі як Modern Data Stack, Data Mart, описано підходи у ETL ELT процесах. Також було описано використані технології та представлено аргументи, чому було вибрано саме їх. PostgreSQL було вибрано через бажання компанії мати аналітичне сховище в одній екосистемі, для інтеграції даних вибрано сервіс Airbyte, який допоміг швидко синхронізувати дані між сховищем та оперативним джерелом даних, а також налаштувати інкрементальне оновлення даних. Для моделювання даних обрано dbt як основний інструмент для моделювання даних. Для візуалізації обрано PowerBi. Результатом Розділу 2 стало побудоване сховище даних.

У Розділі 3 детально описано процес розробки аналітичної системи. Наведено процес створення ELT процесу для завантаження даних у сховище, описано процес моделювання даних за допомогою dbt та описано концепцію розподілу даних за шарами. Була побудована звітність у середовищі PowerBi, яка дозволила проаналізувати основну інформацію щодо рекламних кампаній та їх вплив на продажі. За допомогою технологій Data Mining та правила 1Rule було створено класифікацію високої та низької продажів за джерелом реклами та категорією товару. Також був проведений кластерний аналіз, який дав уявлення про розподіл продажів та допоміг виділити когорти користувачів. Результатом цього розділу стали представлені результати дослідження та рекомендації для підприємства.

У результаті проведеного аналізу каналів трафіку було зроблено декілька важливих висновків. Аналіз показав кілька піків, які вказували на ефективність продажів у ті періоди, але і на проблеми із утриманням клієнтів. Аналіз каналів трафіку показав найефективніші канали та визначено найпопулярніші категорії товарів. Визначено, що трафік з каналу GoogleAds здебільшого спрямований на категорію товарів Gaming, хоча при цьому загальна конверсія з цього каналу низька.

За аналізом з правила 1Rule визначено, що категорія «Монітори та ноутбуки» показала 100% високих продажів, що вказує на їх стабільний попит. Трафік з Facebook частіше генерує продажі дешевших товарів з ймовірністю 80%.

У результаті було надано рекомендації щодо покращення становища: зменшення витрат на рекламу у Facebook, посилити контент-маркетинг, збільшити бюджети на ретаргетинг та звернути увагу на категорії Gaming, «Монітори та ноутбуки», які складають стабільний попит.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. <https://www.thoughtspot.com/data-trends/analytics/ecommerce-analytics>
2. <https://www.forbes.com/advisor/business/customer-experience-trends/>
3. <https://www.adindex.ua/uk/data-driven-marketynh/>
4. <https://www.adweek.com/performance-marketing/infographic-what-data-driven-marketing-looks-2015-163607/>
5. <https://forbes.ua/business/analitika-v-ecommerce-dopomaga-ne-vtrachati-groshi-3-kroki-dlya-ii-nalashtuvannya-15022021-1016>
6. <https://ijsrcseit.com/home/issue/view/article.php?id=CSEIT228324>
7. <https://www.researchgate.net/publication/369506184> Analytical evaluation of big data applications in E-commerce A mixed method approach
8. <https://www.warse.org/IJATCSE/static/pdf/file/ijatcse301042021.pdf>
9. <https://www.researchgate.net/publication/363039657> E-Commerce Analytics And Predictions
10. <https://www.researchgate.net/publication/374247258> Big data analytics and its application in E-commerce
11. <https://www.semanticscholar.org/paper/Implementing-Big-Data-Analytics-in-Business-Lahoti-Panse/59d03b0d9f811f8f6381d9c58bcd1320329c9e9>
12. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9367128>
13. <https://www.researchgate.net/publication/365950503> Web Analytics Tools for e-Commerce An Overview and Comparative Analysis
14. <https://support.google.com/analytics/answer/10089681?hl=uk>
15. <https://amplitude.com/>
16. <https://uk.wikipedia.org/wiki/ETL>

17. <https://aws.amazon.com/what-is/etl/>
18. https://moodle.znu.edu.ua/pluginfile.php/767343/mod_resource/content/1/%D0%A2%D0%B5%D0%BC%D0%B0%202.pdf
19. <https://www.rbcgrp.com/ua/etl-ili-elt-kakoj-process-raboty-s-dannymi-daet-optimalnyj-rezultat/>
20. <https://www.indicium.tech/blog/modern-data-stack-the-guide-to-business-success>
21. <https://www.postgresql.org/>
22. <https://habr.com/ru/companies/yoomoney/articles/748598/>
23. <https://www.hydra.so/blog-posts/2022-12-13-how-we-built-fastest-postgres-db-for-analytics>
24. <https://airbyte.com/>
25. <https://habr.com/ru/companies/otus/articles/574704/>
26. <https://www.theseattledataguy.com/what-is-airbyte-and-why-you-should-use-it/#page-content>
27. <https://dou.ua/forums/topic/42886/>
28. <https://data-life-ua.com/modelling/what-does-dbt-actually-do/>
29. <https://www.getdbt.com/product/what-is-dbt>
30. <https://hub.kyivstar.ua/articles/shho-take-power-bi-mozhlyvosti-dlya-biznes-analityky>
31. <https://www.span.eu/ua/%D1%96%D0%BD%D1%81%D0%B0%D0%B9%D1%82%D0%B8/power-bi-what-is-it-about/>
32. <https://learn.microsoft.com/en-us/power-bi/>
33. Що таке архітектура програмного забезпечення [Електронний ресурс]. – 2017. – Режим доступу до ресурсу: <https://soandso.biz/blog/software-engineering/arhitektura-programnogo-zabezpechennya.html>.

34. OLAP (online analytical processing) [Электронный ресурс] – Режим доступа до ресурсу: <https://www.techtarget.com/searchdatamanagement/definition/OLAP>.
35. What is a data warehouse? [Электронный ресурс] – Режим доступа до ресурсу: <https://www.ibm.com/topics/data-warehouse>.
36. What is SSAS (SQL Server Analysis Services)? [Электронный ресурс] – Режим доступа до ресурсу: <https://www.solarwinds.com/resources/it-glossary/ssas>.
37. What Are SQL Server Integration Services? [Электронный ресурс] – Режим доступа до ресурсу: <https://www.solarwinds.com/resources/it-glossary/ssi-sql-server-integration-services>.
38. SSRS Tutorial: What is SQL Server Reporting Services? [Электронный ресурс]. – 2023. – Режим доступа до ресурсу: <https://www.guru99.com/ssrs-tutorial.html>.
39. What is a KPI? [Электронный ресурс] – Режим доступа до ресурсу: <https://www.qlik.com/us/kpi>.