

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**Факультет інформатики та обчислювальної техніки
Кафедра інформаційних систем та технологій**

До захисту допущено:

Завідувач кафедри

_____ Олександр РОЛІК

«__» _____ 20__ р.

**Дипломний проєкт
на здобуття ступеня бакалавра
за освітньо-професійною програмою «Інформаційні управляючі системи та
технології»
спеціальності 126 «Інформаційні системи та технології»
на тему: «Інформаційна система аналізу та обробки статистичних даних
інтернет магазину»**

Виконав:

студент IV курсу, групи ІС-91
Шерстюк Ігор Олександрович

Керівник:

доцент, д.т.н., доцент
Дорогий Я.Ю.

Рецензент:

доцент кафедри ПЗКС ФПМ
к.т.н., доцент Цуркан В. В.

Засвідчую, що у цьому дипломному проєкті
немає запозичень з праць інших авторів без
відповідних посилань. Студент

Київ – 2023 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет інформатики та обчислювальної техніки
Кафедра інформаційних систем та технологій

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 126 «Інформаційні системи та технології»

Освітньо-професійна програма «Інформаційні управляючі системи та технології»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Олександр РОЛІК

«___» _____ 20__ р.

ЗАВДАННЯ

на дипломний проєкт студенту

Шерстюку Ігорю Олександровичу

1. Тема проєкту «Інформаційна система аналізу та обробки статистичних даних інтернет магазину», керівник проєкту Дорогий Ярослав Юрійович., доцент, к.т.н., доцент, затверджені наказом по університету від «31» травня 2023 р. № 2101-с

2. Термін подання студентом проєкту: 12 червня 2023 р.

3. Вихідні дані до проєкту: технічна документація

4. Зміст пояснювальної записки:

Теоретичні дослідження аналізу діяльності інтернет-магазину, розробка моделі для аналізу статистики інтернет-магазину, розробка інформаційної системи статистичного аналізу діяльності інтернет-магазину

5. Перелік графічного матеріалу

1. Діаграма прецедентів, 2. Схеми алгоритмів, 3. Діаграми бізнес-процесів BPMN 1, 4. Діаграми бізнес-процесів BPMN 2

6. Дата видачі завдання 06.03.2023

Календарний план

№ з/п	Назва етапів виконання дипломного проєкту	Термін виконання етапів проєкту	Примітка
1.	Затвердження теми роботи	06.03.2023	
2.	Теоритичне дослідження	09.03.2023-20.03.2023	
3.	Аналіз існуючих рішень	20.03.2023-01.04.2023	
4.	Вивчення та аналіз завдання	01.04.2023-22.04.2023	
5.	Розробка моделі	20.04.2023-03.05.2023	
6.	Реалізація системи	03.05.2023-26.05.2023	
7.	Тестування системи	26.05.2023-28.05.2023	
8.	Оформлення пояснювальної записки	28.05.2023-04.06.2023	
9.	Предзахист	05.06.2023	
10.	Захист		

Студент
Керівник

Ігор ШЕРСТЮК
Ярослав ДОРОГИЙ

АНОТАЦІЯ

Шерстюк І. О. Інформаційна система аналізу та обробки статистичних даних Інтернет магазину. КПП ім. Ігоря Сікорського, Київ, 2023.

Проект містить 63 с. тексту, 27 рисунків, 2 таблиці, посилання на 34 літературні джерела, додатки.

Ключові слова: аналіз діяльності, Інтернет-магазин, інформаційна система, статистичні дані, процес проектування.

Об'єктом дослідження є аналіз статистичних даних інтернет-магазину.

Предметом дослідження є інформаційні системи для аналізу статистичних даних інтернет-магазину.

Метою дослідження є проектування та розробка інформаційної системи аналізу статистичних даних інтернет-магазину.

У дипломному проєкті були розглянуті алгоритми, які реалізують окремі блоки аналізу статистичних даних інтернет-магазину. Дані алгоритми були реалізовані в системі. Система реалізована на мові Python у вигляді веб-додатку, що знімає питання установки. Проведене тестування показало, що система працює у відповідності з вимогами.

Практичне значення отриманих результатів полягає в побудові системи, яка дозволяє проаналізувати основні складові елементи діяльності підприємства, а саме, дохід, клієнтів та товари.

Запропоноване рішення використовує внутрішню статистичну інформацію інтернет-магазину, на відміну, від існуючих сервісів, орієнтовано на внутрішні бізнес-процеси магазину, реалізовано у вигляді простого в управлінні веб-додатку.

SUMMARY

Sherstiuk I. O. Information system of analysis and processing of statistical data of an Internet store. National Technical University of Ukraine 'Igor Sikorsky Kyiv Polytechnic Institute', Kyiv, 2023.

The project contains 63 pages. text, 27 figures, 2 tables, references to 34 literary sources, appendices.

Keywords: activity analysis, Internet store, information system, statistical data, design process.

The object of the study is the analysis of statistical data of the online store.

The subject of research is information systems for the analysis of statistical data of an online store.

The purpose of the research is the design and development of an information system for the analysis of statistical data of an online store.

Algorithms that implement separate blocks of statistical data analysis of an online store were considered in the diploma project. These algorithms were implemented in the system. The system is implemented in the Python language in the form of a web application, which eliminates installation issues. The conducted testing showed that the system works in accordance with the requirements.

The practical significance of the obtained results lies in the construction of a system that allows analyzing the main components of the enterprise's activity, namely, income, customers and goods.

The solution uses the internal statistical information of the online store, in contrast to existing services, is focused on the internal business processes of the store, implemented in the form of an easy-to-manage web application.

Номер рядка	Формат	Позначення	Найменування	Кільк. аркушів	Номер екзем.	Примітка
1			<u>Документація загальна</u>			
2						
3			Знову розроблена			
4						
5	A4	IC91.290БАК.004 ПЗ	Пояснювальна записка	63		
6	A3	IC91.290БАК.004 Д1	Інформаційна система аналізу	1		
7			та обробки статистичних.			
8			Діаграма прецедентів			
9	A3	IC91.290БАК.004 Д2	Інформаційна система аналізу	1		
10			та обробки статистичних.			
11			Схема алгоритму			
12	A3	IC91.290БАК.004 Д3	Інформаційна система аналізу	1		
13			та обробки статистичних.			
14			Діаграма бізнес-процесів			
15			ВРМН 1			
16	A3	IC91.290БАК.004 Д4	Інформаційна система аналізу	1		
17			та обробки статистичних.			
18			Діаграма бізнес-процесів			
19			ВРМН 2			
20	A4	IC91.290БАК.004 Д5	Інформаційна система аналізу	7		
21			та обробки статистичних.			
22			Лістинг програми			
23						
24						
25						
26						
27						
28						
Зм.	Аркуш	№ докум.	Підпис	Дата	IC91.290БАК.004 ТП Інформаційна система аналізу та обробки статистичних даних інтернет магазину Відомість проекту	
Розроб.		Шерстюк І. О.				
Керівн.		Дорогий Я. Ю.				
Затв.		Дорогий Я. Ю.				
					Літ.	Аркуш
					Т	Аркушів
						1
					КПІ ім. Ігоря Сікорського	
					Група IC-91	

**Пояснювальна записка
до дипломного проєкту
на тему: «Інформаційна система аналізу та обробки
статистичних даних інтернет магазину»**

Київ – 2023 року

ЗМІСТ

ВСТУП.....	3
1 ТЕОРЕТИЧНІ ДОСЛІДЖЕННЯ АНАЛІЗУ ДІЯЛЬНОСТІ ІНТЕРНЕТ-МАГАЗИНУ	5
1.1 Особливості функціонування інтернет-магазину	5
1.2 Аналітика інтернет-магазину: метрики та способи вимірювання	16
1.3 Огляд існуючих сервісів для аналізу аналітики інтернет-магазину	29
Висновки до розділу	27
2 РОЗРОБКА МОДЕЛІ ДЛЯ АНАЛІЗУ СТАТИСТИКИ ІНТЕРНЕТ-МАГАЗИНУ	29
2.1 Методи статистичного аналізу показників діяльності інтернет-магазину	29
2.2 Модель системи аналізу діяльності інтернет-магазину	46
2.3 Розробка інфологічного забезпечення інформаційної системи	46
Висновки до розділу 2	47
3 РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ СТАТИСТИЧНОГО АНАЛІЗУ ДІЯЛЬНОСТІ ІНТЕРНЕТ-МАГАЗИНУ	48
3.1. Розробка алгоритму роботи системи	48
3.2 Обґрунтування засобів розробки.....	56
3.3 Тестування роботи системи	58
Висновки до розділу 3	67
ВИСНОВКИ.....	68
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	64

					ІС91.290БАК.004 ПЗ						
Змн.	Арк.	№ докум.	Підпис	Дата							
Розроб.		Шерстюк І. О.			Інформаційна система аналізу та обробки статистичних даних інтернет магазину Пояснювальна записка			Літ.	Арк.	Аркушів	
Перевір.		Дорогий Я. Ю.								2	68
								КПІ ім. Ігоря Сікорського Група ІС-91			
Н. Контр.											
Затверд.		Дорогий Я. Ю.									

ВСТУП

Зі зміною середовища та розвитком технологій онлайн-ринок є актуальною сферою, яка потребує постійного вивчення. Кількість користувачів Інтернету зростає, оскільки онлайн-технології вдосконалюються, а компанії пропонують продукти та послуги онлайн, але тільки 47% населення України купує ці продукти та послуги.

Розвиток Інтернет-комунікацій означає великі можливості та зміни у способах ведення бізнесу. Це має економічні та політичні, а також технологічні наслідки в країнах, що розвиваються. Внаслідок цього відбуваються зміни в структурі ринків, ринкові взаємодії підприємств, їх стратегії та мобільності окремих осіб. Впровадження нових технологій дозволяє робити бізнес швидше, ніж раніше.

Сучасні організації перебувають під зростаючим тиском з боку зацікавлених сторін, щоб знайти нові способи ефективної конкуренції на динамічних ринках. Цей динамізм є результатом великих змін у структурі ринків, мобільності людей і значного зростання глобальних подорожей.

Посилення конкуренції призводить до того, що будь-яке підприємство повинно постійно досліджувати ефективність своєї роботи. Так, для будь-якого інтернет-магазину постає питання проведення аналізу щодо ефективності своєї діяльності. Існує досить багато різних метрик, за допомогою яких можна оцінити сайт інтернет-магазину. Вони дають можливість оцінити якість створення сайту, ефективність проведення рекламної компанії, відвідуваність інтернет-магазину тощо. Всі ці показники ґрунтуються, в основному, на відвідуваності сайту. Для аналізу цих показників можна використовувати різні онлайн інструменти, наприклад, GoogleAd тощо.

Однак, будь-який інтернет-магазин має внутрішні дані щодо своєї роботи, а саме, інформацію про транзакції, про клієнтів та товари, які купуються. Дана інформація також може виступати джерелом даних для прийняття рішень щодо стратегії роботи магазину. Використання такої інформації може суттєво вплинути

					ІС91.290БАК.004 ПЗ	Арк.
						3
Змн.	Арк.	№ докум.	Підпис	Дата		

на подальші плани магазину. В той же час відмітимо, що така інформація, зазвичай, не може бути оброблена вручну, оскільки вона може бути великою за обсягом (понад десяток мільйонів транзакцій за рік) і містити приховані зв'язки. Тому для вирішення таких задач необхідно застосовувати автоматизовані системи, алгоритми роботи яких використовують спеціальні методи обробки великих даних та виявлення прихованих закономірностей. Отже тема даного дослідження є актуальною.

Об'єктом дослідження є аналіз статистичних даних інтернет-магазину.

Предметом дослідження є інформаційні системи для аналізу статистичних даних інтернет-магазину.

Метою дослідження є проектування та розробка інформаційної системи аналізу статистичних даних інтернет-магазину.

Для досягнення мети потрібно вирішити такі завдання:

- дослідити основні показники, що характеризують діяльність інтернет-магазину;
- проаналізувати статистичні методи, які використовуються для оцінки діяльності магазину;
- розробити модель інформаційної системи для аналізу статистичних даних інтернет-магазину;
- розробити інформаційну систему на основі побудованої моделі;
- провести тестування системи.

Практичне значення отриманих результатів полягає в побудові системи, яка дозволяє проаналізувати основні складові елементи діяльності підприємства, а саме, дохід, клієнтів та товари.

Запропоноване рішення використовує внутрішню статистичну інформацію інтернет-магазину, на відміну, від існуючих сервісів, орієнтовано на внутрішні бізнес-процеси магазину, реалізовано у вигляді простого в управлінні веб-додатку. Робота складається з вступу, 3 розділів, висновків, списку використаних джерел, додатків.

					ІС91.290БАК.004 ПЗ	Арк.
						4
Змн.	Арк.	№ докум.	Підпис	Дата		

1 ТЕОРЕТИЧНІ ДОСЛІДЖЕННЯ АНАЛІЗУ ДІЯЛЬНОСТІ ІНТЕРНЕТ-МАГАЗИНУ

1.1 Особливості функціонування інтернет-магазину

Ринок Інтернет-торгівлі на сьогоднішній день є невід'ємною частиною економіки будь-якої держави. Це один з найдинамічніших технологічних ринків у світі. Швидке зростання цього ринку обумовлено, в першу чергу, досить швидким глобальним поширенням широкосмугового (фіксованого та мобільного) доступу до мережі Інтернет.

Дані з семи країн, які разом складають близько половини світового ВВП (включаючи Сполучені Штати та Китай), свідчать про значне зростання роздрібних онлайн-продажів у цих країнах. У 2019 році, перед пандемією, сума цих продажів склала близько 2 трлн доларів США. Але вже у 2020 році вона зросла до майже 2,5 трлн доларів США, а в 2021 році досягла 2,9 трлн доларів США. Варто зазначити, що Китай забезпечує більше половини роздрібних онлайн-продажів у цих країнах, тоді як Сполучені Штати відповідають за близько 30% цієї суми [1].

У 2019 році продажі 13 найбільших компаній електронної комерції, орієнтованих на споживачів, становили 2,4 трильйона доларів. Після початку пандемії COVID-19 у 2020 році цей показник різко зріс до 2,9 трильйона доларів США, а у 2021 році відбулося ще одне збільшення до 3,9 трильйонів доларів США (рисунок 1.1).

У 2019 році лише 17,8% продажів припадало на онлайн-покупки, тоді як у 2022 році частка он-лайн комерції в загальних продажах становитиме 21%. Щодо прогнозу, то в 2025 році очікується зростання до 24,5% [1].

За даними прогнозів, до 2023 року роздрібні онлайн-продажі досягнуть 6,17 трильйона доларів США, а веб-сайти електронної комерції будуть забезпечувати до 22,3% від загального обсягу роздрібних продажів [1].

					ІС91.290БАК.004 ПЗ	Арк.
						5
Змн.	Арк.	№ докум.	Підпис	Дата		

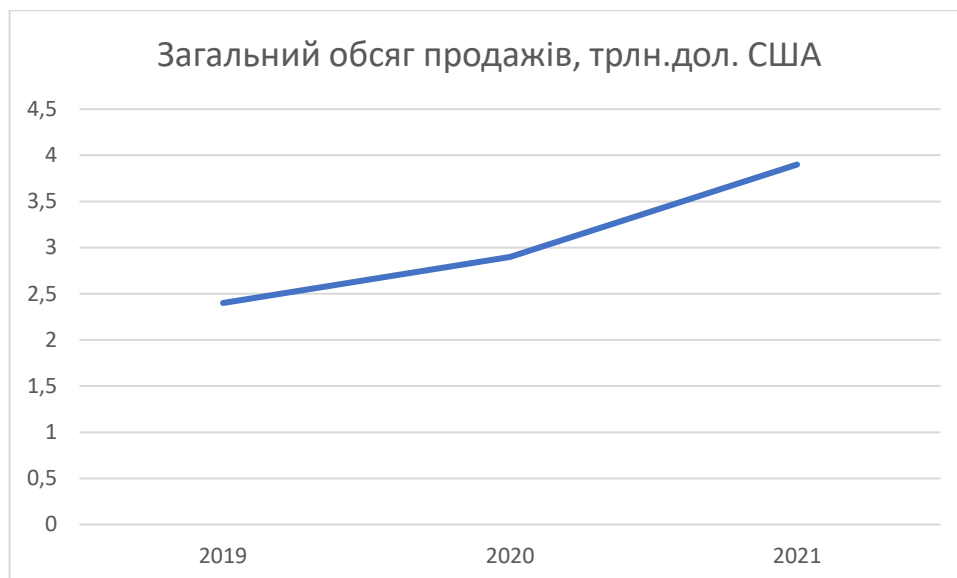


Рисунок 1.1 – Динаміка онлайн продажів в світі

У відповідності з статистичними дослідженнями в 2021 році Україна посіла 65-е місце за доходом в 1,1 млрд доларів США на ринку електронної комерції (об'єм ринку – понад 4 млрд доларів США, що становить 2,6% ВВП країни). Український ринок демонструє пришвидшений темп зростання цього ринку у 27% порівняно зі світовим темпом 15% [2].

Згідно прогнозних оцінок у 2025 році обсягу ринку електронної комерції мав становити понад 7 млрд доларів з річним темпом зростання 18,76%. Це був найбільші показники серед країн Центральної та Східної Європи.

На жаль, на розвиток електронної комерції в Україні впливають об'єктивні чинники. Так, Україна має найнижчий рівень проникнення Інтернету в Європі – 71,8% на початок 2022 року. І серед них лише 44% інтернет-користувачів здійснюють покупки онлайн.

Зараз війна також суттєво вплинула на розвиток ринку електронної комерції. Наприкінці лютого, у березні та квітні 2022 року електронна комерція майже зникла в Україні. Доходи від неї знизилися на 87 %. Але вже в травні обсяги почали наближатися до 80 %, а у червні-липні становили 90 % порівняно з довоєнними показниками. Так що, ринок електронної комерції залишається суттєвим рушієм економіки країни [1].

Інтернет-магазин – це web-сайт, що забезпечує продаж через Інтернет з використанням електронного каталогу або іншого способу представлення продукції. Особливості функціонування Інтернет-магазину можна описати наступним чином [3]:

– відкритість 24/7: Інтернет-магазини працюють цілодобово, що дозволяє клієнтам зробити покупку в зручний для них час. Крім того, це дозволяє власникам магазинів збільшити свій потенційний охоплюваний аудиторію, оскільки люди можуть зробити покупку з будь-якої точки світу, де є доступ до Інтернету.

– швидкість та зручність пошуку: Інтернет-магазини зазвичай мають вбудовані інструменти пошуку, які дозволяють клієнтам знайти потрібний товар за декількома параметрами, такими як назва товару, виробник, ціна та інші.

– широта асортименту: Інтернет-магазини можуть запропонувати більший вибір товарів, ніж традиційні магазини, оскільки вони не обмежені розміром свого простору і можуть зберігати товари на складах. Крім того, вони можуть пропонувати товари різних виробників та поставників з різних країн світу.

– онлайн оплата: Інтернет-магазини дозволяють клієнтам здійснювати оплату онлайн, що робить процес покупки швидшим та зручнішим. Клієнти можуть скористатися різними способами оплати, такими як кредитні картки, PayPal, банківські перекази тощо.

– доставка товарів: Інтернет-магазини зазвичай пропонують різні способи доставки товарів, в залежності від потреб клієнтів. Це може бути доставка кур'єром, поштою або самовивіз зі складу магазину. Крім того, деякі магазини пропонують безкоштовну доставку при певних умовах, наприклад, якщо замовлення перевищує певну суму.

– зручність та економія часу: Оскільки процес покупки здійснюється онлайн, клієнти можуть зробити покупку без потреби виходити з дому або офісу. Це робить процес купівлі більш зручним та економить час.

					ІС91.290БАК.004 ПЗ	Арк.
						7
Змн.	Арк.	№ докум.	Підпис	Дата		

– Можливість порівняння товарів та цін: В Інтернет-магазинах клієнти можуть порівняти товари та їх ціни з іншими магазинами, що дозволяє зробити більш обдуманий вибір та отримати кращу пропозицію.

– Контроль якості: В Інтернет-магазинах клієнти можуть давати відгуки про товари та послуги, що дозволяє іншим клієнтам зробити більш обдуманий вибір та відчувати більший контроль над якістю товарів.

Торгові системи електронних магазинів можуть бути різними. За ознакою ступеня автоматизації їх можна класифікувати як веб-вітрини, інтернет-магазини та торгові інтернет-системи (TIC) [3].

Веб-вітрина представляє собою сукупність каталогу, системи навігації та оформлення замовлення з подальшою передачею менеджера для подальшої обробки. Це відносно прості і недорогі сайти, що представляють товари торгової компанії у вигляді стандартного каталогу. Вони можуть виконувати операції оформлення замовлення, а іноді і виставлення рахунку [3].

Можливості інтернет-магазину значно вищі, але разом із цим вищі і вартість його реалізації. Системи інтернет-магазину виконують більшу частину завдань, що не вирішуються в рамках веб-вітрини, наприклад, завдяки динамічній обробці інформації і роботі з базами даних в ньому є можливість працювати індивідуально з кожним покупцем, що зареєструвався.

Інтернет-магазини та TIC можуть здійснювати повний торговельний цикл. Відмінною рисою TIC при цьому є те, що вона додатково інтегрована в систему автоматизації всієї діяльності компанії.

Потенційний клієнт може залишити інтернет-магазин набагато легше, ніж реальний магазин. Ця обставина визначає необхідність високого рівня сервісу, запропонованого у віртуальних магазинах. У той самий час особливості контакту з покупцем призводять до важливої відмінності наданих ними послуг [3].

У власника віртуального магазину є можливість отримувати досить повну інформацію про відвідувачів веб-сайту та будувати відповідно до неї систему просування свого інтернет-магазину.

					ІС91.290БАК.004 ПЗ	Арк.
						8
Змн.	Арк.	№ докум.	Підпис	Дата		

Крім того, за рахунок використання комп'ютерних технологій можлива персоналізація підходу до кожного з клієнтів, виходячи з історії його відвідувань магазину та зроблених раніше покупок.

Опишемо схему взаємодії покупця з віртуальним магазином. За допомогою браузера покупець заходить на веб-сайт інтернет-магазину. Сайт містить електронну вітрину, на якій представлений каталог товарів та необхідні елементи інтерфейсу для введення реєстраційної інформації, формування замовлення, проведення через Інтернет платежів, оформлення доставки, отримання інформації про компанію-продавця та інтерактивну допомогу.

Реєстрація покупця здійснюється або при оформленні замовлення або при вході в магазин. Після вибору товару йому потрібно заповнити форму, в якій вказується, яким чином буде здійснено оплату та доставку. Для захисту персональної інформації взаємодія повинна здійснюватися за захищеним каналом, наприклад, за протоколом SSL. По закінченні формування замовлення та реєстрації вся зібрана інформація про покупця з електронної вітрини надходить у торговельну систему інтернет-магазину. Там здійснюється перевірка наявності затребуваного товару на складі, ініціюється запит до платіжної системи. За відсутності товару на складі надсилається запит постачальнику, а покупцю повідомляється про час затримки.

У разі, якщо оплата здійснюється під час передачі товару покупцю, необхідно підтвердження факту замовлення. Найчастіше це відбувається за допомогою електронної пошти або телефоном.

Якщо магазин має можливість оплати покупки по мережі, тоді на етапі оплати підключається платіжна система. Після повідомлення про проведення віртуального платежу торговою системою формується замовлення служби доставки.

					ІС91.290БАК.004 ПЗ	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дата		



Рисунок 1.2 – Схема взаємодії покупця з віртуальним магазином

1.2 Аналітика інтернет-магазину: метрики та способи вимірювання

Аналітика електронної комерції є основою для прийняття багатьох бізнес-рішень. Показники є однією з найважливіших, якщо не найважливішою частиною всієї концепції аналітики. Можна регулярно отримувати величезну кількість даних, але якщо не перетворюєте ці дані на корисну інформацію, це не допоможе розвитку бізнесу.

Аналітика електронної комерції – це процес збору, аналізу та використання даних для вимірювання впливу інтернет-магазину на бізнес. Це допомагає компаніям зрозуміти й оцінити поведінку клієнтів, тенденції онлайн-покупок, показники продажів і рентабельність інвестицій [5].

Аналітика електронної комерції має вирішальне значення для виявлення та інтерпретації даних, що допомагає приймати кращі рішення, які можуть скоротити витрати, збільшити продажі та покращити загальну роботу.

Основною перевагою аналітики електронної комерції є можливість збирати, аналізувати та стратегічно діяти на основі даних клієнтів.

Така аналітика може відстежувати, вимірювати та допомагати вивчати кожну взаємодію клієнта з магазином і кожну дію, яку клієнти виконують в онлайн-магазині.

Метрики мають узгоджувати ефективність із цілями. Якщо суб'єкт господарювання встановив цілі та знає, як їх вимірювати, це свідчить про те, що він знає, як пов'язати свою діяльність із бізнес-цілями. Відповідно, метрики, встановлені таким чином, є частиною дорожньої карти для досягнення стратегічних цілей; вони визначають дії, необхідні для досягнення цілей [5].

Існує кілька напрямків вимірювань для оцінки ефективності підрозділів електронної комерції [5]:

- показники трафіку та використання сайту;
- маркетингові показники;
- фінансові показники;
- інші показники ефективності.

Показники трафіку Найперша форма вимірювання онлайн-бізнесу, пов'язана з вимірюванням трафіку. Хоча такі показники бувають дуже грубими, вони можуть надати кількість корисної інформації про обсяги трафіку, використання сайту, джерела трафіку та атрибути відвідувачів.

Аналіз відвідуваності сайту може допомогти визначити обсяг трафіку на веб-сайті, а також визначити, які розділи веб-сайту відвідуються, коли вони відвідуються, ким вони відвідуються та звідки прийшли відвідувачі. Обсяг трафіку можна виміряти різними способами: відвідуваннями, переглядами сторінок, переглядами реклами чи показами банерів, відвідуваннями та унікальними відвідувачами [6].

Найдавніша метрика трафіку передбачає вимірювання звернень. Звертання зараховується щоразу, коли щось запитується від веб-сервера. Це може бути сторінка, зображення або просто посилання на інше місце. Щоразу, коли ви натискаєте сторінку, це буде називатися хітом. Хіт – це дуже широка інформація. Коли хтось створює хіт, ви не знаєте, на що він дивився і що бачив. Крім того, одна

					ІС91.290БАК.004 ПЗ	Арк.
						11
Змн.	Арк.	№ докум.	Підпис	Дата		

веб-сторінка може генерувати дюжину звернень, якщо на ній багато фотографій, у той час як текстова сторінка може створити лише одне звернення. Як наслідок, актуальність відвідувань для вимірювання трафіку на веб-сайті була поставлена під сумнів, і почали використовувати альтернативні способи вимірювання обсягу трафіку [7].

Наступним типом інформації, яка використовується, є перегляд сторінки, який відбувається, коли завантажується ціла сторінка інформації. Це буде вважатися більш цінним, оскільки всю інформацію, яка відображається на цій сторінці, користувач переглядатиме або, принаймні, вважається, що вона переглядається.

Перегляди реклами, покази реклами або покази банерів – це більш точне вимірювання кількості переглядів сторінки з рекламним банером або іншою рекламою протягом певного періоду часу.

Відвідування – це не просто перегляд однієї сторінки, це ціла колекція сторінок, які переглядає один користувач протягом певного періоду часу – зазвичай 30 хвилин. Збираючи всі ці сторінки в один блок, який називається відвідуванням, який ідентифікується з конкретним користувачем, суб'єкт може отримати кращу інформацію, ніж переглядаючи тисячі потенційно непов'язаних звернень [5].

Кількість неповторюваних відвідувачів за певний часовий проміжок вимірює обсяг активності з точки зору передбачуваних осіб, хоча насправді в цьому типі аналізу враховуються комп'ютери (тобто IP-адреси), а не особи.

Інструменти аналізу трафіку зазвичай покладаються на IP-адресу для групування різних звернень і приписування їх одному користувачеві. Однак проксі та "плаваючі" IP-адреси роблять цей метод неточним. Оскільки вміст файлу cookie можна зберігати у файлі журналу, деякі інструменти дозволяють точніше відстежувати, призначаючи унікальний ідентифікатор у файлі cookie, а потім скидаючи вміст файлу cookie до файлу журналу при кожному зверненні, завдяки чому можна точніше відстежувати сеанс користувача [5].

					ІС91.290БАК.004 ПЗ	Арк.
						12
Змн.	Арк.	№ докум.	Підпис	Дата		

Трафік є загальним показником потенціалу сайту; однак, щоб отримати значущу інформацію, аналіз має бути глибшим. Показники використання сайту допомагають електронному бізнесу оцінити ефективність веб-сайту електронної комерції. Аналіз використання сайту може допомогти відстежити втрати в різних точках шляху через веб-сайт від точки входу до виходу та допомогти усунути діри, які допускають "витоки", замість того, щоб перетворювати відвідувачів на покупців.

Використання сайту можна аналізувати просторово або часово. Аналіз просторового використання може визначити поведінку відвідувачів відповідно до [7]:

- сторінок найпопулярніших записів – сторінок, які найчастіше вводяться;
- головних сторінок виходу – сторінок, з яких найчастіше виходять (сторінки виходу можуть означати витік із веб-сайту, вказуючи на передчасний вихід відвідувачів);
- сторінок, які найчастіше відвідуються;
- найменш відвідуваних сторінок;
- сторінок з одним відвідуванням – сторінок, які відвідуються лише один раз (вони можуть бути проблемними точками на веб-сайті);
- шляхів в межах сайту – найпоширеніших шляхів, якими користуються відвідувачі (вони можуть свідчити про проблеми з дизайном, якщо найпоширеніший шлях не веде до успішного завершення транзакції електронної комерції чи іншої мети сайту).

Часовий аналіз використання сайту може визначити відвідуваність за місяць, тиждень, день тижня та годину дня. Такі розбивки можуть бути корисними для співвіднесення з часом рекламних акцій або змін на веб-сайті для визначення їх ефективності. Насправді більшість вимірювань трафіку слід повідомляти за певний і послідовний період часу, щоб забезпечити порівнянність показників.

Мета маркетингового вимірювання полягає в тому, щоб допомогти краще просувати продукти та/або послуги суб'єкта господарювання. З появою систем

					ІС91.290БАК.004 ПЗ	Арк.
						13
Змн.	Арк.	№ докум.	Підпис	Дата		

CRM, інструментів спільної фільтрації та персоналізації були розроблені складні методи, орієнтовані на клієнта, для збору важливої маркетингової інформації [8].

Файли журналу веб-сервера Referrer Analysis містять інформацію про попередню URL-адресу відвідувача. Інформацію про реферера також можна придбати в інформаційних посередників, таких як DoubleClick. За цією інформацією можна визначити, звідки надходить трафік, і таким чином виміряти ефективність рекламних кампаній. Якщо суб'єкт господарювання розміщує рекламу на веб-сайті, але не отримує з нього трафік, тоді витрати на рекламу можуть бути неефективними. І навпаки, якщо трафік надходить із місць, де відсутня реклама суб'єкта, тоді йому слід розглянути, які фактори можуть відповідати за залучення такого трафіку [6, 7].

Аналіз місцезнаходження передбачає розбивку поведінки клієнтів за географічним походженням. Аналіз місця розташування має велике значення для маркетингу, оскільки деякі акції/реклама мають регіональне охоплення. Показники розташування також можуть допомогти оцінити ефективність менеджерів різних географічних регіонів і ефективність місцевих маркетингових зусиль у створенні трафіку та доходів. За відсутності чітких даних приблизне місцезнаходження можна визначити за IP-адресою; однак проксі є серйозним обмеженням для цього методу. Все частіше опитування користувачів або реєстрація відвідувачів використовуються для отримання такої інформації, як місцезнаходження, і, звичайно, покупці вказують своє місцезнаходження як частину інформації про доставку.

Аналіз профілю клієнта передбачає розбивку поведінки відвідувача за атрибутами профілю. Інтернет дозволяє здійснювати вибірку цільової аудиторії в режимі реального або квазіреального часу. Нові відвідувачі заходять на веб-сайт електронного бізнесу як "анонімні", але їх дії одразу визначають у профілі. Кожна дія записується у файл журналу веб-сервера. Взаємодіючи з сайтом електронної комерції, відвідувачі надають інформацію, яку можна використовувати для

					ІС91.290БАК.004 ПЗ	Арк.
						14
Змн.	Арк.	№ докум.	Підпис	Дата		

формування "профілів користувачів", які, у свою чергу, можуть використовуватися для маркетингових цілей, наприклад для вибору продуктів для показу тощо [7].

Підвищити ефективність веб-сайту можна шляхом створення профілю типового користувача або класів користувачів шляхом відстеження таких атрибутів, як вік, стать, дохід домогосподарства, середня покупка, загальна кількість покупок тощо. Дані профілю клієнта можна отримати з наданих даних клієнтами явно або виведеними з неявної інформації, притаманної атрибутам або поведінці клієнта. Явна інформація передається користувачем під час підписки на послугу, здійснення покупки, заповнення анкети тощо, наприклад ім'я, адреса та сімейний стан. Неявну інформацію можна отримати з дій користувача та/або історії покупок, наприклад улюблений колір, вікова група або бажані теми. Для таких висновків може знадобитися використання інструментів інтелектуального аналізу даних для виявлення ключових моделей, що представляють інтерес для організації. Аналіз профілю користувача може допомогти суб'єкту оцінити, чи охоплює він цільову аудиторію, і, таким чином, ефективність його маркетингових кампаній.

Аналіз кошика для покупок передбачає розбивку товарів, придбаних у зразковій транзакції. Аналіз кошика для покупок може допомогти розвинути розуміння купівельної поведінки клієнтів і ціннісних міркувань клієнтів. Набори елементів – це сукупність елементів, які з'являються разом у багатьох транзакціях. Знання про набори товарів і їх частоту можна використовувати для покращення позиціонування продукту або для перехресних продажів і продажів з більшою ціною. Наприклад, використовуючи аналіз кошика для покупок, компанія може запропонувати клієнту такі рекомендації, як "Клієнти, які придбали X, також придбали Y", які можна бачити на сайтах великих інтернет-магазинів [7].

Аналіз кошика для покупок може допомогти суб'єкту господарювання оцінити та спрогнозувати попит на продукт.

Фінансові показники відстежують доходи, витрати, повернення інвестицій і створення вартості для власників бізнесу.

					ІС91.290БАК.004 ПЗ	Арк.
						15
Змн.	Арк.	№ докум.	Підпис	Дата		

Потоки доходів у типовому електронному бізнесі включають дохід від продажів продуктів і послуг, дохід від реклами. Аналіз продажів передбачає розбивку потоків доходу за періодами, категоріями продуктів, джерелом переходу, часом доби тощо. Він також може включати розбивку продажів за профілем покупця, демографічними показниками, географічним розташуванням тощо. Розширена аналітика продажів може включати аналіз кластерів продуктів, моделей покупок, аналіз кошика для покупок і прогнозування майбутніх продажів на основі спостережуваних моделей [7].

Показники витрат включають постійні витрати, операційні витрати та витрати на маркетинг. Постійні витрати включають витрати на інфраструктуру та експлуатаційні витрати. Витрати на інфраструктуру включають розвиток магазину, розробку каталогу, послуги розміщення магазину. Операційні витрати включають обслуговування сайту, рекламу/просування, маркетинг і операції. Змінні витрати включають витрати на продукти/послуги, витрати на ліцензування на одиницю, комісійні від продажів за кліками тощо, а також витрати на розповсюдження неелектронних товарів або послуг.

Рентабельність інвестицій (ROI) розраховується як $(\text{Доходи} - \text{Витрати}) / \text{Активи}$. Сильна сторона цього показника полягає в тому, що він поєднує три ключові фінансові параметри в єдиний зведений показник. Слабкість полягає в тому, що він пропускає нефінансові показники, такі як обслуговування клієнтів, якість роботи та інновації [6].

Для оцінки ефективності електронного бізнесу можна використовувати інші показники ефективності. Наприклад, показники продуктивності можуть стосуватися доступності мережі, часу відповіді системи, точності обробки транзакцій, своєчасності обробки транзакцій, реагування на запити про зміни системи, ефективності служби підтримки, обробки інцидентів безпеки, задоволеності клієнтів, ступеня інновацій тощо. Як правило, такі показники вбудовані в угоди про рівень обслуговування, які визначають критерії ефективності, адаптовані до послуг, що надаються, а також штрафи за невиконання.

					ІС91.290БАК.004 ПЗ	Арк.
						16
Змн.	Арк.	№ докум.	Підпис	Дата		

Багатовимірні системи показників об'єднують кілька показників, щоб створити багатовимірний показник ефективності. Часто використовуваним методом для створення внутрішньої системи показників є збалансована система показників, розроблена Робертом Капланом і Девідом Нортоном [7].

Каплан і Нортон створили "картку показників" із чотирма основними розділами: клієнт, внутрішній процес, навчання та зростання та фінансова ефективність. Ідея полягає в тому, що в кожній із цих чотирьох областей необхідно визначити невелику кількість показників, зосереджуваних на довгострокових рушіях успіху бізнесу.

Показники, орієнтовані на клієнта, включають частку ринку, аналіз трафіку, аналіз покупок, вартість придбання, обізнаність і задоволеність/лояльність. Показники внутрішнього процесу включають інноваційність, ефективність, результативність та економічність операцій електронного бізнесу та надійність системи. Показники навчання та зростання стосуються можливостей і мотивації співробітників. Фінансові показники зосереджені на доходах, витратах і рентабельності інвестицій (ROI).

Отже, в будь-якому разі організація повинна зосереджуватись на тих показниках, які характеризують її роботу.

Визначимо, які джерела метрик може використовувати інтернет-магазин для аналізу своєї роботи.

Користувачі, які переглядають веб-сайт, користувачі, які натискають рекламні банери, користувачі, які взаємодіють з опитуваннями та опитуваннями, покупці, які обирають продукт, покупці, які роблять покупки, покупці, які звертаються до служби підтримки клієнтів, – усе це діяльність клієнтів можна відстежувати та аналізувати, щоб краще зрозуміти поведінку клієнтів і узгодити дизайн електронного бізнесу з поведінкою клієнтів [8].

Найбільш очевидним джерелом є дані, які зберігаються користувачами в файлах cookie та гаманцях. Наступним найбільш використовуваним джерелом є дані, знайдені в файлах журналу веб-сервера. Третім ключовим джерелом є бази

					ІС91.290БАК.004 ПЗ	Арк.
						17
Змн.	Арк.	№ докум.	Підпис	Дата		

даних транзакцій, які записують інформацію про продажі. Четвертим джерелом є бази даних профілів користувачів, створені інформаційною системою магазину [7].

Аналіз кліків – це метод, за допомогою якого можна відстежувати та аналізувати шлях користувача на веб-сайті. Ключовим тут є аналіз даних для виявлення важливих тенденцій, таких як труднощі з пошуком конкретних сайтів чи інформації, пошуки, які не дали результатів, або постійне «зворотне» переміщення сайтом. Аналізуючи дані потоку кліків, електронний бізнес може виявити помилки дизайну веб-сайту, які можна виправити до того, як на них зіткнуться інші клієнти. За інших обставин аналіз потоку кліків може бути пов'язаний з іншими факторами клієнта, такими як демографічні показники чи галузь бізнесу, що може призвести до виявлення тенденцій, пов'язаних із конкретними цільовими ринками. Наприклад, кореляція реєстраційних даних користувача, як-от типу галузі чи іншої офлайн-інформації, з аналізом потоку кліків може привести організацію до висновку, що її веб-сайт не організований у спосіб, який добре підходить для певних галузей. Потім сайт можна переробити, щоб краще відповідати конкретним галузям, або розробити спеціалізований дизайн для вирішення цієї проблеми.

Дані кліків підтримуються кількома технологіями. Файли cookie – це невеликі текстові файли, які зберігаються веб-сайтом на окремих комп'ютерах і дозволяють сайту відстежувати переміщення відвідувача. Файли cookie часто додаються до комп'ютера особи під час першого відвідування сайту. Через них можна відстежувати велику кількість інформації, як-от час, проведений на сайті, паролі, інформацію про кошик для покупок, уподобання та останній відвіданий сайт. Використання файлів cookie може персоналізувати веб-сайт, запам'ятовуючи профіль користувача та організовуючи презентацію веб-сайтів. Основне занепокоєння щодо файлів cookie полягає в тому, що вони зберігаються на комп'ютері користувача та потенційно можуть надавати інформацію про інші відвідувані сайти, уподобання та звички покупців. Однак інформацію в файлі cookie зазвичай не читає ніхто, окрім сайту, який її туди розмістив; зазвичай це цифри та літери, які ні для кого нічого не означатимуть, за винятком поєднання з

					ІС91.290БАК.004 ПЗ	Арк.
						18
Змн.	Арк.	№ докум.	Підпис	Дата		

інформацією, яка вже зберігається на стороні веб-сервера. Використання файлів cookie як онлайн-технології стало стандартною практикою [8].

Важливо зазначити, що дані, збережені в файлі cookie, можуть бути легко стерті користувачем або користувач може прийняти рішення взагалі не приймати файли cookie. Тому файли cookie часто використовуються для зберігання некритичних даних, таких як персоналізація для анонімних користувачів, інформація для входу (лише ім'я для входу, а не пароль), дата останнього входу користувача тощо.

Електронні гаманці є більш складною версією файлів cookie, оскільки файли cookie можуть містити лише два або три елементи інформації, тоді як гаманець може містити більше інформації. Електронні гаманці спочатку забезпечували більш безпечний спосіб передачі інформації про кредитні картки через Інтернет, але згодом вони також стали більш загальним засобом передачі інформації. Електронний гаманець може містити номер кредитної картки користувача, ім'я, адресу, адресу доставки тощо. Наявність цієї інформації в електронному гаманці була б особливо корисною для он-лайн покупців, оскільки це позбавило б необхідності вводити цю інформацію щоразу, коли користувач хоче щось купити. Крім того, інформація гаманця може зберігатися в надійному місці в спеціально відведеному місці зберігання, яке не знаходиться ні в онлайн-магазині, ні на власному комп'ютері користувача (де її може прочитати хтось, хто має доступ до комп'ютера). Використання авторизації для зберігання цифрових гаманців може забезпечити як безпеку, так і зручність під час покупок.

Файли журналів містять детальну інформацію про кожну дію, яка сталася на веб-сайті, і є одним із найпоширеніших джерел даних для аналізу трафіку. Файли журналу веб-сервера – це текстові файли, що зберігаються на веб-сервері. Кожен рядок представляє «хіт», що містить таку інформацію: HTTP-запит, дата, час, ОС, тип браузера та URL-адреса джерела переходу. URL-адреса переходу – це джерело, з якого надійшов запит. Ця інформація дуже корисна. Маючи лише цю невелику інформацію, система може визначити, звідки прийшов відвідувач, чи було це

					ІС91.290БАК.004 ПЗ	Арк.
						19
Змн.	Арк.	№ докум.	Підпис	Дата		

перенаправлення від рекламного банера чи від іншого постачальника реклами. Інша інформація, яку можна отримати з журналу, включає, наприклад, які браузері люди використовують (наприклад, Google Chrome, Opera, Microsoft Edge тощо), які операційні системи використовуються (наприклад, MacOS, Windows, Unix, Linux), які обчислювальні платформи тощо [7].

Аналіз файлу журналу – це процес аналізу інформації щодо переміщень користувачів сайтом на основі даних, зібраних у файлах журналу сервера, і може здійснюватися за допомогою низки засобів, включаючи базові методи баз даних або розширені інтернет-продукти.

Файли журналу часто використовуються для інших цілей, а не лише для аналізу продуктивності. Вони використовуються для відновлення системи після збою. Якщо сервер виходить з ладу під час обробки транзакції, файл журналу можна використовувати для повторної обробки транзакції або принаймні для повернення системи до стану, в якому вона була до збою.

Веб-баги (також відомі як чисті GIF-файли) – це файли зображень, вбудовані у веб-сторінку, які можуть відстежувати рухи користувача без відома користувача. По суті, веб-баги – це чіткі зображення або зображення, які зливаються з фоном сторінки, яку ви переглядаєте. Веб-баг надсилає інформацію на сервер (належить самій компанії або рекламній фірмі) кожного разу, коли ви запитуєте нову сторінку, і таким чином відстежує ваші переміщення. Це дозволяє аналітичній фірмі зрозуміти ваші переміщення сайтом і є одним із методів збору даних потоку кліків.

База даних транзакцій – це файл облікових даних, який веде облік проданих товарів, кількості, засобів розрахунку, умов, дати тощо. Цей тип інформації може бути корисним для розуміння поведінки покупців; однак багатство інформації, що міститься в такій базі даних, залежить від реалізації. Як і у випадку з даними профілю користувача, будуть явні дані, що відображають транзакцію, і неявні дані, отримані в результаті аналізу поведінки користувача. Такі умовиводи можуть вимагати використання інструментів бізнес-аналітики або інтелектуального

					ІС91.290БАК.004 ПЗ	Арк.
						20
Змн.	Арк.	№ докум.	Підпис	Дата		

аналізу даних, щоб ідентифікувати ключові моделі, що представляють інтерес для організації [7].

База даних профілів користувачів містить інформацію про відвідувачів і покупців. База даних може містити явні дані, надані користувачами, наприклад, інформацію про їхню адресу, ім'я чи вік, якщо вони відповідали на запити цієї інформації у формах або надавали інформацію в їхній гаманець. База даних також може містити неявні дані, дані, отримані шляхом аналізу характеристик користувача. Для таких висновків може знадобитися використання інструментів інтелектуального аналізу даних для виявлення ключових моделей, що представляють інтерес для організації.

Таким чином, для організації аналізу ефективності діяльності магазину можна використовувати різні джерела інформації. Це залежить від можливостей компанії, її цілей та інструментів аналізу, які є в наявності.

Для сучасного інтернет-магазину особливо важливим є взаємовідношення з клієнтами, оскільки завдяки клієнтам, можливості їх утримання, залучення до нових покупок формується основний дохід магазину. Тому аналіз клієнтів для інтернет-магазинів висувається на першу позицію. Сучасні компанії стурбовані підвищенням цінності споживача шляхом аналізу життєвого циклу клієнта. Інструменти та технології сховищ даних, інтелектуального аналізу даних та інших методів управління взаємовідносинами з клієнтами (CRM) надають підприємствам нові можливості діяти на основі концепцій маркетингу відносин [9].

Показник ефективності змінюється з частки ринку на так звану "частку гаманця". Підприємства не просто мають справу з клієнтами, щоб здійснювати операції; вони перетворюють можливість продавати продукти на досвід обслуговування та намагаються встановити довгострокові відносини з кожним клієнтом [9].

Поява Інтернету, безсумнівно, сприяла зміні фокусу маркетингу. Оскільки он-лайн інформація стає доступнішою та багатшою, споживачі стають більш поінформованими та досвідченими. Одним із ефективних способів виділити себе є

					ІС91.290БАК.004 ПЗ	Арк.
						21
Змн.	Арк.	№ докум.	Підпис	Дата		

системи, які можуть точно й узгоджено взаємодіяти з клієнтами. Збір демографічних даних і даних про поведінку клієнтів робить можливим точне націлювання. Таке націлювання також допомагає при розробці ефективного плану просування, щоб протистояти жорсткій конкуренції, або при виявленні потенційних клієнтів, коли з'являються нові продукти. Послідовна взаємодія з клієнтами означає, що компанії повинні зберігати записи про транзакції та відповіді в онлайн-системі, доступній для досвідчених співробітників, які знають, як з нею взаємодіяти. Важливість налагодження тісних стосунків із клієнтами вимагає використання CRM. Індивідуальні каталоги, персоналізовані сторінки та цільові пропозиції продуктів можуть спростити процес закупівлі та підвищити ефективність для компанії. Сповіщення електронною поштою та нові інформація про продукт, адаптована до різних покупців, може допомогти підвищити ефективність реклами. Усе це можна віднести до переваг CRM [9].

Завдяки використанню кредитних карток інтернет-магазин може вести детальний облік кожної транзакції покупки. Це дозволяє їм краще розуміти різні сегменти клієнтів. Деякі роздрібні програми включають [10]:

Виконання аналізу кошика. Також відомий як аналіз спорідненості, аналіз кошика показує, які товари клієнти зазвичай купують разом. Ці знання можуть покращити запаси, стратегії розташування магазинів і рекламні акції.

Прогнозування продажів. Вивчення часових закономірностей допомагає роздрібним торговцям приймати рішення щодо запасів. Якщо клієнт придбає товар сьогодні, коли він, швидше за все, придбає додатковий товар?

Маркетинг на базі даних. Інтернет-продавці можуть розробляти профілі клієнтів із певною поведінкою, наприклад, тих, хто купує одяг від дизайнерських лейблів або тих, хто відвідує розпродажі. Цю інформацію можна використати для фокусування економічно ефективних рекламних кампаній.

					IC91.290БАК.004 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		22

1.3 Огляд існуючих сервісів для аналізу аналітики інтернет-магазину

Зараз існує досить багато різних інструментів для отримання аналітики веб-сайтів.

Найбільш поширеним сервісом є Google Analytics. Google Analytics – це інструмент аналітики електронної комерції для будь-якого типу веб-сайту (рисунок 1.3). Він став стандартним інструментом веб-аналітики завдяки тому, що він безкоштовний, простий у використанні та містить інформативні звіти.

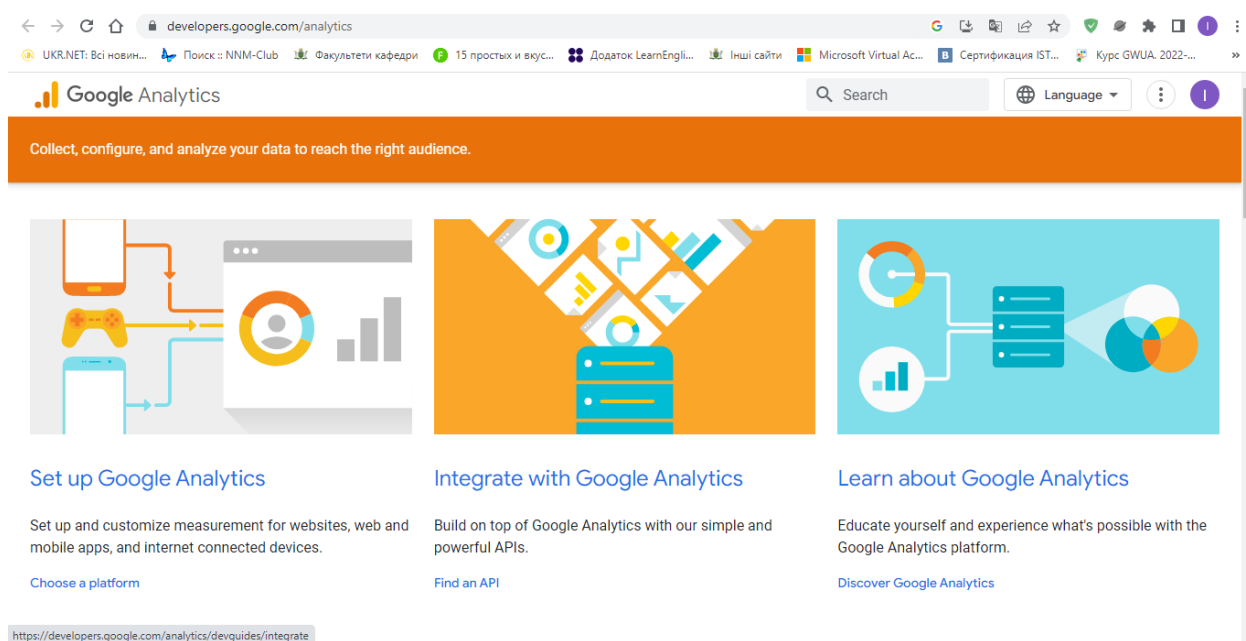


Рисунок 1.3 – Google Analytics

Як правило, Google Analytics складається з 5 основних звітів: звіти реального часу, звіти про аудиторію, про придбання, поведінку, перетворення. В свою чергу, вони складаються з інших звітів та підзвітів.

Власник встановлює ресурс від Google, який автоматично відстежує всі переходи. Система починає записувати дії та переміщення, як тільки людина відвідує будь-яку сторінку сайту. Причому вона враховує не лише прямі переходи. Враховуються відвідування з пошукової видачі, соцмереж, рекламних сервісів та інших джерел.

					ІС91.290БАК.004 ПЗ	Арк.
						23
Змн.	Арк.	№ докум.	Підпис	Дата		

Вся інформація надсилається на сервери Google, де обробляється та систематизується у загальну базу даних. Через деякий час дані з'являються на панелі керування, де можна скористатися різними фільтрами та віджетами для перегляду конкретної інформації.

Завдяки збору статистики власники сайтів можуть виконувати аналіз наступних аспектів:

- аналітичних інструментів. Користувальницькі змінні, API-інструменти, зведення відвідуваності, візуалізація трафіку, підтримка доступу для співробітників тощо;

- змісту. Статистика сторінок ресурсу, відстеження подій, аудит швидкості завантаження. Різноманітність інструментів дозволяє отримати дані про найбільш популярні сторінки сайту;

- мобільних даних. Аналіз за програмами, статистика переходів з гаджетів, оцінка ефективності оголошень на мобільних пристроях;

- соціальної активності Детальна інформація про репости матеріалів, вплив соцмереж та переходи, комерційна вигода;

- конверсії. Детальна статистика для конверсій та цілей, оптимізація продажів, контроль ефективності РК, візуалізація переходів на ресурсі;

- реклами. Інструмент, що дозволяє аналізувати та порівнювати різні джерела реклами, підвищувати ефективність рекламних кампаній та інтеграції з AdWords.

Зі звітів можна дізнатися:

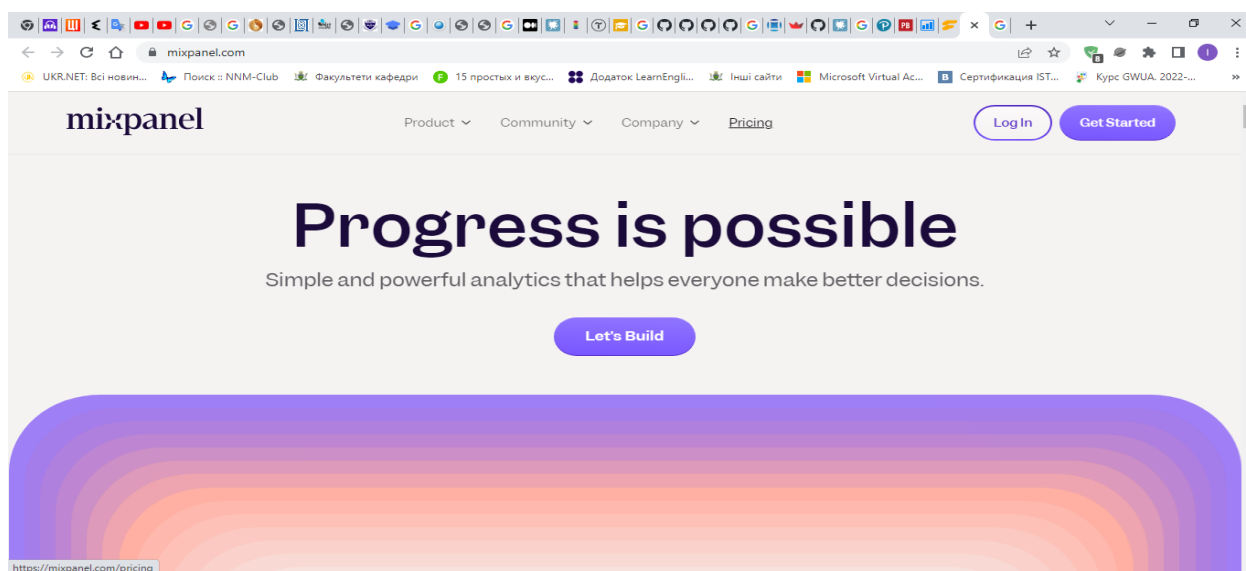
- про швидкість завантаження сторінок ресурсу;
- про події. Надається повна інформація про натискання на різні кнопки, перехід між сторінками, заповнення різних форм;
- про розподіл платформ;
- про продаж та конверсію;
- про рекламні місця;
- про активність у соцмережах.

					ІС91.290БАК.004 ПЗ	Арк.
						24
Змн.	Арк.	№ докум.	Підпис	Дата		

MixPanel – це одне з найдосконаліших аналітичних рішень (рисуюнок 1.4). Як і Google Analytics, його можна використовувати як для мобільних пристроїв, так і для Інтернету.

MixPanel використовує інший підхід до веб-аналізу, вимірюючи події. Його основними функціями є:

- Закладки.
- А/В тестування.
- Візуальне кодування.
- Мобільні опитування.
- Особливості утримання.
- Сумісність з iOS і Android.
- Залучення клієнтів.



Рисуюнок 1.4 – Головна сторінка MixPanel

MixPanel відстежує події, створює воронки та виявляє тенденції, що є більш цілеспрямованим порівняно з іншими аналітичними платформами.

Наступним сервісом є CleverTap (рисуюнок 1.5). Цей інструмент пропонує інформацію про клієнтів у реальному часі, допомагаючи створювати маркетингові

					ІС91.290БАК.004 ПЗ	Арк.
						25
Змн.	Арк.	№ докум.	Підпис	Дата		

кампанії. Він має можливість створювати багатоканальні кампанії на основі інформації про дії користувачів.

CleverTap є прикладом універсального маркетингового інструменту для мобільних додатків завдяки функціям маркетингової аналітики, автоматизації маркетингу, мобільної аналітики, мобільного маркетингу та push-повідомлень.

Він оснащений простим інтерфейсом перетягування, що робить закріплення, клонування або експорт будь-якого звіту на спеціальну інформаційну панель дуже легким.

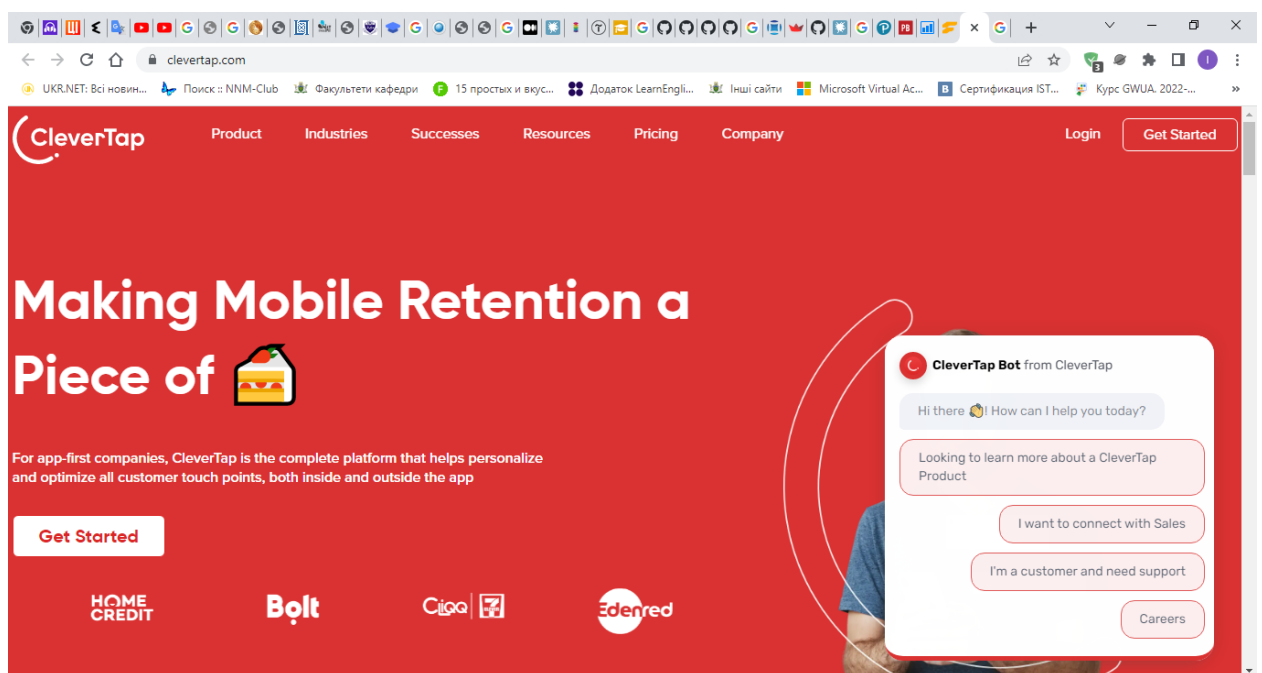


Рисунок 1.5 – Головна сторінка CleverTap

CleverTap пропонує автоматичне сегментування клієнтів для створення високоперсоналізованих кампаній, що є ключем до перетворення їх на лояльних клієнтів. CleverTap наразі є єдиною платформою такого типу, яка використовує ШІ та машинне навчання для персоналізації взаємодії з клієнтами за допомогою даних у реальному часі.

Наступний сервіс LeanPlum (рисунок 1.6). Функціональність LeanPlum зосереджена на побудові міцних відносин із клієнтами. Завдяки міжканальним кампаніям LeanPlum об'єднує повідомлення та взаємодію в програмі.

					ІС91.290БАК.004 ПЗ	Арк.
						26
Змн.	Арк.	№ докум.	Підпис	Дата		

LeanPlum підтримує канали зв'язку як поза програмою, так і всередині програми. У такому випадку клієнт може отримувати сповіщення про нові товари та послуги через push-сповіщення, електронну пошту, вебхук, повідомлення в додатку або певну функцію додатка.

Він дозволяє створювати варіанти маркетингових кампаній і проводити A/B-тестування, щоб створити виграшну кампанію.

LeanPlum пропонує вам огляд усіх послідовностей, кампаній і звітів для отримання інформації щодо співвідношення витрат та прибутків.

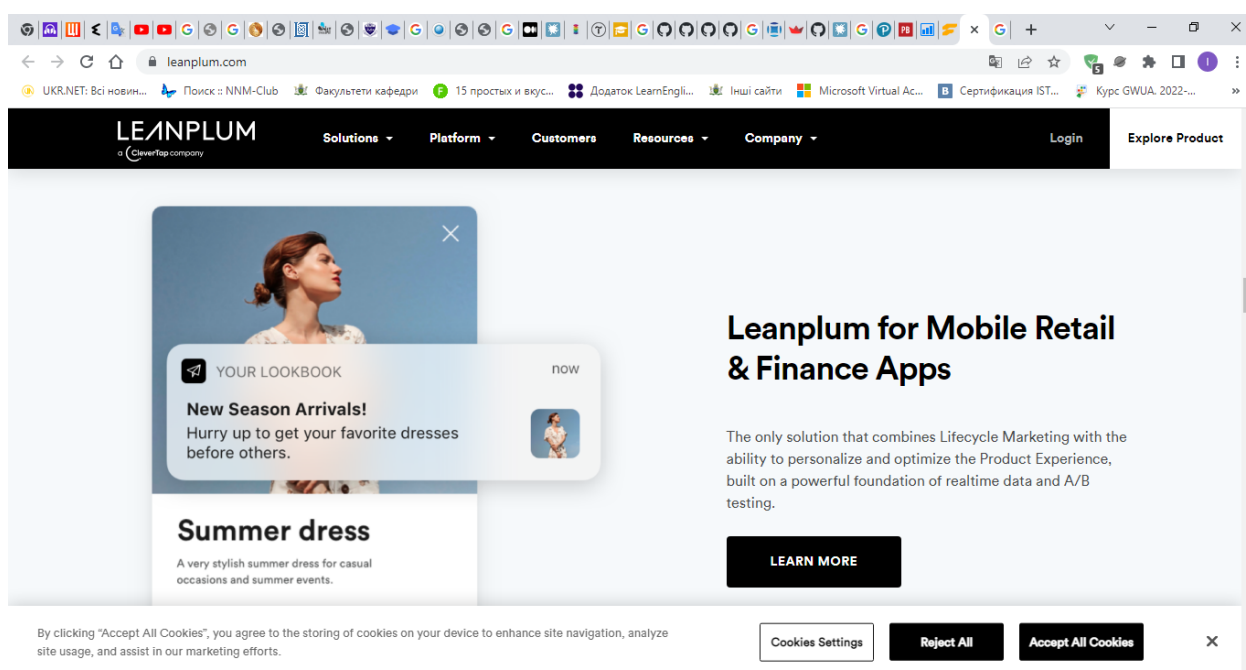


Рисунок 1.6 – Головне вікно LeanPlum

Висновки до розділу

Електронна комерція, а разом з нею, інтернет-магазини, стала однією з основних тенденцій ведення бізнесу. Це обумовлено перевагами в здійсненні такого виду діяльності, серед яких швидкість, простота, можливість постійного доступу до товарів.

Але підвищення ефективності використання інтернет-магазинів залишається на часі. Ефективність електронного бізнесу можна виміряти декількома способами, починаючи від показників трафіку на сайті електронної комерції до показників

					ІС91.290БАК.004 ПЗ	Арк.
						27
Змн.	Арк.	№ докум.	Підпис	Дата		

поведінки клієнтів і, зрештою, до показників ефективності електронного бізнесу як ROI.

Існує велика кількість показників, які описують електронну комерцію, але, в будь-якому разі, їх використання повинно бути пов'язані з конкретними цілями. Такі показники можуть допомогти керівництву досягти своїх стратегічних цілей. Їх можна використовувати як основу для прийняття рішень.

					ІС91.290БАК.004 ПЗ	Арк.
						28
Змн.	Арк.	№ докум.	Підпис	Дата		

2 РОЗРОБКА МОДЕЛІ ДЛЯ АНАЛІЗУ СТАТИСТИКИ ІНТЕРНЕТ-МАГАЗИНУ

2.1 Методи статистичного аналізу показників діяльності інтернет-магазину

Бізнес-аналітика – це процес, за допомогою якого підприємства використовують статистичні методи та технології для аналізу даних, щоб отримати розуміння та покращити прийняття стратегічних рішень.

Існує три типи аналітики, які компанії використовують для прийняття рішень; описова аналітика, яка розповідає нам про те, що вже сталося; прогнозна аналітика, яка показує нам, що може статися в недалекому майбутньому, і, нарешті, прескриптивна аналітика, яка повідомляє, що має статися в майбутньому [11].

Описова аналітика – це аналіз історичних даних за допомогою двох ключових методів – агрегації даних і інтелектуального аналізу даних – які використовуються для виявлення тенденцій і закономірностей. Описова аналітика не використовується для того, щоб робити висновки або прогнозувати майбутнє на основі її результатів, даний метод пов'язаний з представленням того, що сталося в минулому.

Описова аналітика часто відображається за допомогою візуальних представлень даних, таких як лінійні, стовпчасті та секторні діаграми, і, хоча вони самі по собі дають корисну інформацію, часто виступають основою для майбутнього аналізу. Оскільки описова аналітика використовує досить прості методи аналізу, будь-які висновки мають бути легкими для розуміння широкої бізнес-аудиторії.

З цієї причини описова аналітика є основою повсякденної звітності в багатьох компаніях. Річні звіти про доходи є класичним прикладом описової аналітики, а також інші звіти, такі як інвентаризація, складські дані та дані про продажі, які можна легко агрегувати та надати чіткий знімок операцій компанії. Іншим широко використовуваним прикладом є соціальні медіа та інструменти Google Analytics, які узагальнюють певні групи на основі простого підрахунку подій, таких як кліки та лайки [11].

					ІС91.290БАК.004 ПЗ	Арк.
						29
Змн.	Арк.	№ докум.	Підпис	Дата		

Хоча описові дані можуть бути корисними для швидкого виявлення тенденцій і закономірностей, аналіз має свої обмеження. Як окремий інструмент описова аналітика може не дати повної картини. Щоб отримати більше розуміння, вам потрібно заглибитися глибше.

Прогностична аналітика – це вдосконалений метод аналізу даних, який використовує ймовірності для оцінки того, що може статися в майбутньому. Як і описова аналітика, прогностична аналітика використовує інтелектуальний аналіз даних, однак вона також використовує методи статистичного моделювання та машинного навчання для визначення ймовірності майбутніх результатів на основі історичних даних. Щоб зробити прогнози, алгоритми машинного навчання беруть наявні дані та намагаються заповнити відсутні дані якомога найкращими припущеннями [11].

Потім ці прогнози можна використовувати для вирішення проблем і визначення можливостей для зростання. Наприклад, організації використовують прогнозну аналітику, щоб запобігти шахрайству, шукаючи закономірності злочинної поведінки, оптимізуючи свої маркетингові кампанії, виявляючи можливості для перехресних продажів, і зменшуючи ризик, використовуючи минулу поведінку, щоб передбачити, які клієнти, найімовірніше, не сплатять платежі.

Інша галузь прогновної аналітики – глибоке навчання, яке імітує процеси прийняття рішень людиною, щоб робити ще більш складні прогнози. Наприклад, завдяки використанню кількох рівнів соціального та екологічного аналізу глибоке навчання використовується для точнішого прогнозування кредитних балів, а в галузі медицини воно використовується для сортування цифрових медичних зображень, таких як МРТ та рентгенівські знімки, щоб забезпечити автоматизоване прогнозування, яке лікарі використовують для діагностики пацієнтів.

У той час як прогнозна аналітика показує компаніям необроблені результати їхніх потенційних дій, преспективна аналітика показує компаніям, який варіант є найкращим [11].

					ІС91.290БАК.004 ПЗ	Арк.
						30
Змн.	Арк.	№ докум.	Підпис	Дата		

Прескриптивної використовує різноманітні статистичні методи.

Хоча прескриптивна аналітика тісно пов'язана як з описовою, так і з прогнозною аналітикою, вона наголошує на практичних ідеях замість моніторингу даних. Це досягається шляхом збору даних із ряду описових і прогностичних джерел і застосування їх у процесі прийняття рішень. Потім алгоритми створюють і відтворюють можливі моделі прийняття рішень, які можуть по-різному вплинути на організацію.

Підприємства, які можуть використовувати потужність прескриптивної аналітики, використовують її різними способами. Наприклад, вона оптимізувати бізнес-результати, рекомендуючи найкращий курс дій для клієнтів і постачальників. Вони також дають змогу фінансовим компаніям знати, наскільки знизити вартість продукту, щоб залучити нових клієнтів і зберегти високий прибуток

Методи описової статистики є відносно простими, оскільки вони включають базові математичні розрахунки та агрегацію даних для отримання важливих значень для оцінки історичної ділової практики та її ефективності. Деякі поширені методи описового статистичного аналізу включають [10, 11]:

1. Міри частоти включають математичні розрахунки, зокрема підрахунок, обчислення відсотків і частоту повторень. Міра частоти використовуються для підрахунку кількості разів, коли конкретна змінна, подія або число з'являється в наборі даних, тобто

2. Міри центру розподілу описують характеристики центру розподілу для даного набору даних. Вони використовуються для відображення середніх відповідей шляхом аналізу частоти точок вибірки даних і вираження її за допомогою середнього значення, медіани та моди.

Центральна міра тенденції визначає найбільш загальні тенденції або спільні характеристики у вибіркових даних.

3. Міра варіації описує центральні позиції розподілу для даного набору даних. Вона використовується для відображення розподілу середніх варіантів

					ІС91.290БАК.004 ПЗ	Арк.
						31
Змн.	Арк.	№ докум.	Підпис	Дата		

шляхом аналізу частоти точок вибірки даних і вираження її за допомогою середнього значення, медіани та моди. Центральна міра тенденції визначає найбільш загальні тенденції або спільні характеристики у вибіркових даних.

Для прогнозної статистики можуть використовувати більш складні методи.

Факторний аналіз використовується для встановлення базової структури більшого набору корельованих змінних. Мета полягає в тому, щоб ущільнити інформацію, що міститься в кількох оригінальних цінностях, у менший набір зі складеними розмірами, забезпечуючи при цьому мінімальну втрату інформації. Простіше кажучи, ми зменшуємо дані, створюючи одну змінну, якою легко керувати, представляючи набір спостережуваних змінних (зазвичай семантичних диференціальних шкал) у термінах загальних факторів, які можуть пояснити кореляцію, яку можна застосувати до більшої сукупності.

Сумісний аналіз (conjoint analysis – конджойнт-аналіз) використовується, щоб розрізнити, як респонденти дослідження ринку приймають складні рішення про покупку, які включають сприйняття й оцінку різних змінних, пов'язаних із продуктом або послугою. Конджойнт-аналіз вимагає від респондентів оцінити компроміси, пов'язані з різними факторами, такими як ціни, брендинг тощо, і визначити їх вплив на обмірковування покупки, щоб оцінити критерії прийняття рішень клієнтами.

Перехресна таблиця визначає взаємозв'язки, які можуть бути неочевидними, розміщуючи змінні поруч одна з одною у двовимірній таблиці, яка забезпечує перспективу розуміння таких співвідношень. Ця техніка використовується для оцінки закономірностей, тенденцій, зв'язків і ймовірностей шляхом групування змінних для розуміння кореляцій між різними змінними, задіяними у вибірці даних.

Аналіз TURF (повністю недубльований аналіз охоплення та частоти) використовується для ранжування й оптимізації комбінацій продуктів, а також для точного налаштування комунікаційних стратегій шляхом аналізу охоплення джерел зв'язку та частоти. Аналіз TURF дозволяє оцінити медіа та ринковий потенціал для розробки оптимальних стратегій комунікації та розміщення. Він

					IC91.290БАК.004 ПЗ	Арк.
						32
Змн.	Арк.	№ докум.	Підпис	Дата		

визначає кількість користувачів, охоплених кожним методом зв'язку, і частоту їх охоплення, щоб можна було краще контролювати настрої ринку.

Кореляційний аналіз забезпечує поглиблене дослідження взаємозв'язку між двома чи більше змінними з набору даних та їх застосування до загального пулу сукупності. Це може допомогти підприємствам робити прогнози щодо майбутньої поведінки шляхом встановлення причинно-наслідкових зв'язків або залежностей, які можуть бути позитивними чи негативними.

Регресійний аналіз широко використовується для прогнозування щільності зв'язку між двома чи більше змінними. Під час аналізу оцінюється вплив незалежних змінних на залежну змінну, щоб зрозуміти, які змінні мають більший вплив.

Управління взаємовідносинами з клієнтами визначається чотирма елементами простої структури: знання, ціль, продаж, обслуговування [9]. CRM вимагає від фірми знання та розуміння своїх ринків і клієнтів. Це передбачає детальну розвідку клієнтів з метою вибору найприбутковіших клієнтів і визначення тих, на кого більше не варто орієнтуватися. CRM також передбачає розробку пропозиції: які продукти продавати яким клієнтам.

Компанії можуть дізнаватися про своїх клієнтів за допомогою внутрішніх даних клієнтів або вони можуть дані про покупки із зовнішніх джерел. Є кілька джерел внутрішніх даних: зведені таблиці, що описують клієнтів (наприклад, записи про рахунки) опитування підгрупи клієнтів, які відповідають на детальні запитання, поведінкові дані, що містяться в системах транзакцій (веб-журнали, записи кредитних карток, тощо) [9].

Далі система CRM повинна проаналізувати дані за допомогою статистичних інструментів, OLAP і Data Mining.

Одним з видів аналізу, який дає змогу поєднати інформацію про клієнтів та результати продажів є RFM-аналіз, який дозволяє сегментувати клієнтів за частотою та сумою покупок та виявляти тих, які приносять більше грошей [12].

Абревіатура RFM розшифровується [13]:

					ІС91.290БАК.004 ПЗ	Арк.
						33
Змн.	Арк.	№ докум.	Підпис	Дата		

Recency – давність (як давно клієнти робили покупки в Інтернет-магазині). Високий показник давності означає, що у клієнта вже склалося досить добре враження про магазин, тому він нещодавно відвідував вас..

Frequency – частота (як часто клієнти здійснюють покупки). Високий показник частоти говорить про те, що клієнту подобається здійснювати покупки, тому він часто повертається до магазину.

Monetary – гроші (загальна сума витрат). Високий рівень цього показника означає, що клієнту подобається витратити гроші саме в цьому інтернет-магазині.

В результаті аналізу утворюється множина груп. Кожна група – певна кількість клієнтів, які по-різному купують в інтернет-магазині і реагують на його контент.

На основі цих груп компанія може вибрати індивідуальний підхід до кожної групи.

В результаті сегментації утворюються такі групи [14]:

1. Втрачені – це клієнти, котрі колись купували у компанії, але згодом перестали.

2. Сплячі - це клієнти, які купували в магазині, але, можливо, прямо зараз у них немає потреби в товарах або вони збираються купити, але весь час забувають оформити замовлення.

3. Лояльні клієнти, які перестали купувати – це тип клієнтів, які раніше купували регулярно, але з останнього часу перестали.

4. Новачки – це клієнти, які нещодавно познайомилися з товарами бренду та щось із них придбали.

5. Клієнти, що розвиваються – ці клієнти купують часто, але невеликі суми. Згодом вони можуть перейти до групи постійних клієнтів.

6. Постійні клієнти часто купують і на великі суми.

Поділ клієнтів на групи може відбуватися за допомогою таких методів.

На основі простих фіксованих діапазонів, які визначаються дослідником для кожної оцінки. Порогові значення діапазону залежить від характеру бізнесу. Цей

					ІС91.290БАК.004 ПЗ	Арк.
						34
Змн.	Арк.	№ докум.	Підпис	Дата		

метод оцінки залежить від типу підприємства, для якого проводиться аналіз. Основним недоліком є те, що у міру зростання бізнесу діапазони оцінок можуть потребувати частого коригування. Також при регулярних платежах з різними умовами оплати розрахунки оцінок будуть неправильними.

Другий спосіб базується на розрахунку квантилів, тобто 5 рівних 20% інтервалів на основі доступних значень. Даний спосіб вирішує проблеми попереднього методу. Квантилі працюють з будь-якою галуззю, оскільки діапазони вибираються із самих даних, вони розподіляють клієнтів рівномірно і не мають перехресних значень.

Квантилі - це найчастіше рекомендований метод розрахунку RFM-оцінки. Він використовується для створення RFM-сегментації, яка, в свою чергу, є рішенням для бізнес-аналітики та маркетингового аналізу для онлайн-продавців.

Сучасні засоби машинного навчання також дозволяють проводити складний аналіз статистики онлайн магазину. Вони дають змогу виявити приховані залежності, які можуть бути неочевидними при використанні описової статистики. Так, важливим елементом будь-якої онлайн торгівлі є кошик покупця. Аналіз ринкового кошика можна здійснювати шляхом застосування пошуку асоціативних правил (Association Rule Mining). Це заснований на правилах метод машинного навчання, який допомагає виявити значущі кореляції між різними продуктами відповідно до їхньої появи в наборі даних [11].

Одним з основних недоліків цього методу є те, що він складається з різних формул і параметрів, і це може ускладнити його застосування.

У навчанні асоціативним правилам використовуються три основні показники, а саме: Підтримка (Support), Довіра (Confidence) та Підйом (Lift) [16].

Підтримка – це проста базова ймовірність того, що подія відбудеться. Вона вимірюється часткою транзакцій, у яких з'являється набір елементів. Інакше кажучи, $Support(A)$ – це кількість транзакцій, що включають A, поділена на загальну кількість транзакцій (2.1) [17].

$$Support(A) = \frac{n_A}{n} \quad (2.1)$$

де n_A – кількість появ товару А,

n – загальна кількість транзакцій

Підтримку можна розраховувати для декількох товарів:

$$Support(A, B) = \frac{n_{AB}}{n} \quad (2.2)$$

де n_{AB} – кількість появ в транзакції товарів А та В (разом)

Довіру до наступної події за наявності попередньої події можна описати з допомогою умовної ймовірності. Простіше кажучи, це ймовірність настання події А за умови, що подія В вже відбулася.

Цей принцип можна використовувати для опису ймовірності придбання товару, коли інший товар вже знаходиться у кошику. Показник довіри вимірюється шляхом розподілу частки транзакцій з товарами А і В на частку транзакцій з товаром В [17].

$$Confidence(A, B) = \frac{n_{AB}}{n_B} \quad (2.3)$$

Підйом – це відношення спостережуваного значення до очікуваного. Показником підйому вимірюють ймовірність того, що товар буде куплено при купівлі іншого товару, при цьому враховується популярність обох товарів. Його можна розрахувати, розділивши ймовірність того, що обидва товари виявляться разом, на добуток ймовірностей того, що обидва товари виявляться разом, якби між ними не було зв'язку [17].

					ІС91.290БАК.004 ПЗ	Арк.
						36
Змн.	Арк.	№ докум.	Підпис	Дата		

$$Lift(A, B) = \frac{Support(A, B)}{Support(A) \cdot Support(B)} \quad (2.4)$$

Значення Lift, що дорівнює 1, говорить про те, що обидва елементи фактично незалежні і ніяк не пов'язані між собою. Будь-яке значення Lift, що перевищує 1, показує, що існує зв'язок. Чим вище значення, тим вищий рівень зв'язку.

Алгоритм Apriori – один із найпопулярніших алгоритмів, які використовуються в навчанні правил асоціацій на реляційних базах даних. Він визначає елементи в наборі даних і далі розширює їх на все більші та більші набори елементів. Однак алгоритм Apriori розширюється тільки в тому випадку, якщо набори елементів є частинами, тобто якщо існує ймовірність того, що набір елементів перевищує заздалегідь визначений поріг [18].

Алгоритм Apriori передбачає наступне.

Ймовірність того, що набір елементів не є частим, можлива у двох випадках:

$P(I) < \text{мінімального порогу підтримки}$, де I – будь-який непустий набір елементів.

Будь-яка підмножина в наборі елементів має значення менше мінімальної підтримки.

Друга характеристика окреслює властивість антимонотонності.

Кроки алгоритму Apriori

Алгоритм починає з генерації набору елементів за допомогою кроку Join Step, тобто генерації набору елементів $(K+1)$ із наборів елементів K .

Відразу після цього алгоритм переходить до кроку обрізки, тобто видаляє будь-який набір елементів-кандидатів, який не відповідає вимогам мінімальної підтримки. Наприклад, алгоритм видалить елемент A , якщо $Support(A)$ буде нижчим від заданого мінімального значення $Support$.

Алгоритм повторює обидва кроки до того часу, поки не залишиться жодної можливості розширення.

Наведемо формальне визначення. Для цього введемо такі поняття.

					IC91.290БАК.004 ПЗ	Арк.
						37
Змн.	Арк.	№ докум.	Підпис	Дата		

Множина об'єктів

$$X \subseteq I = \{x_1, x_2, \dots, x_n\} \quad (2.5)$$

Множина ідентифікаторів транзакцій

$$T = \{t_1, t_2, \dots, t_m\} \quad (2.5)$$

Множина транзакцій

$$\{(t, x): t \in T, x \in X\} \quad (2.6)$$

Розглянемо дерево префіксів (prefix tree), де 2 елементи X і Y з'єднані, якщо X є прямою підмножиною Y . Таким чином ми можемо пронумерувати всі підмножини множини I (рисунок 2.1).

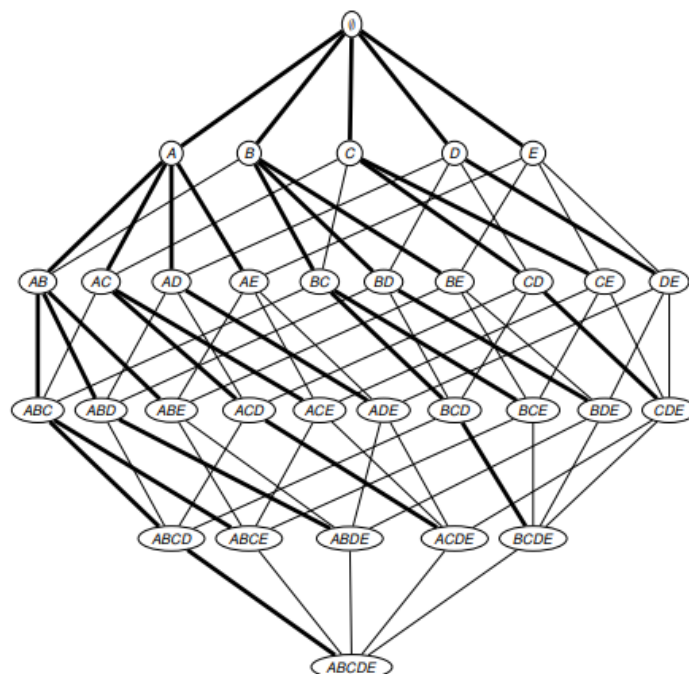


Рисунок 2.1 – Дерево префіксів

Сформульоване твердження щодо антимонотонності має вигляд

Якщо $X \subseteq Y$, то $Support(X) \geq Support(Y)$

Звідси випливає дві властивості

а) якщо Y зустрічається часто, то будь-яка підмножина $X: X \subseteq Y$ теж зустрічається часто;

б) якщо X зустрічається рідко, то будь-яка супермножина $Y: Y \supseteq X$ теж зустрічається рідко.

Апріорі алгоритм по-рівнево проходить по префіксному дереву і розраховує частоту появи підмножин X в D . Таким чином, відповідно до алгоритму:

виключаються рідкісні підмножини і всі їх супермножини

розраховується $Support(X)$ для кожного відповідного кандидата X розміру k на рівні k (рисунок 2.2).

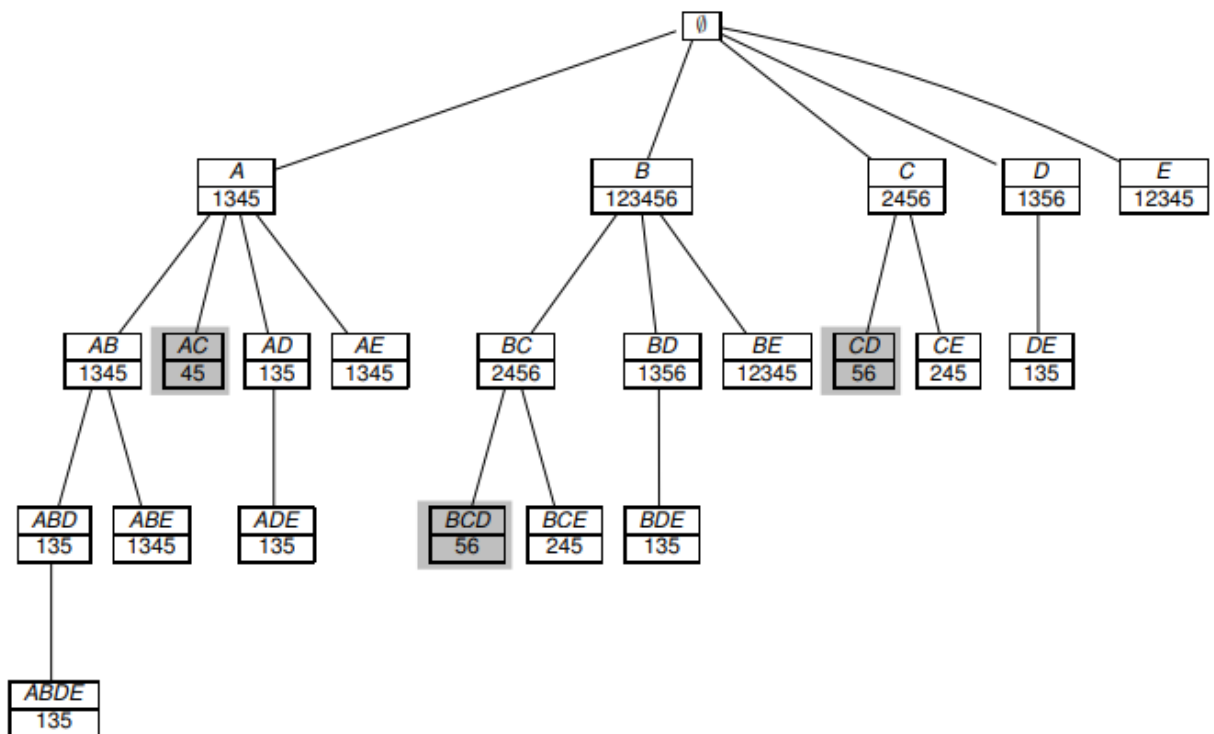


Рисунок 2.2 – Проходження алгоритму Апріорі

Алгоритм є доволі швидким, однак, оскільки він повинен проходити по всьому датасету, то він може потребувати багато оперативної пам'яті, що дещо обмежує його застосування на великих масивах історичних даних.

2.2 Модель системи аналізу діяльності інтернет-магазину

Для розробки проекту системи аналізу статистики онлайн магазину використовується об'єктно-орієнтована методологія проектування інформаційних систем, основу якої становить об'єктно-орієнтована концепція представлення моделей предметної галузі, що мають структурні властивості та поведінку [19].

Модель варіантів використання представляє описи функціонального призначення системи, а також надає можливі варіанти взаємодії різних типів користувача із системою [20]. Мета розробки моделі варіантів використання – досягти взаємного розуміння роботи системи між розробником та замовником. Дана модель призначена виявити можливості, технології, нефункціональні вимоги та зовнішнє оточення, що взаємодіють із системою.

У рамках об'єктно-орієнтованого підходу виділяють такі основні діаграми [19]:

Діаграма варіантів використання;

Діаграма послідовності;

Діаграма діяльностей;

Діаграма станів;

Діаграма класів.

Діаграма варіантів використання (сценаріїв поведінки, прецедентів) є вихідним концептуальним уявленням системи у процесі її проектування та розробки. Ця діаграма складається з акторів, варіантів використання та відносин між ними [23].

Суть цієї діаграми полягає в наступному: проєктована система представляється у вигляді множини акторів, що взаємодіють із системою за допомогою варіантів використання. При цьому актором називається будь-який об'єкт, суб'єкт чи система, що взаємодіє з модельованою системою ззовні.

Варіант використання – це специфікація сервісів (функцій), які система надає актору. Іншими словами, кожен варіант використання визначає деякий набір дій,

					ІС91.290БАК.004 ПЗ	Арк.
						40
Змн.	Арк.	№ докум.	Підпис	Дата		

що здійснюються системою при взаємодії з актором. При цьому на моделі ніяк не відображається те, яким чином буде реалізований цей набір дій.

Реалізація системи аналізу статистики онлайн-магазину передбачає її використання менеджерами компанії, які за її допомогою зможуть отримати зручне подання інформації про істотні характеристики роботи магазину за певний період. В своїй роботі система буде використовувати файл, сформований за зразком, з інформаційної системи онлайн-магазину (варіант CRM-системи). Даний файл має містити інформацію про транзакції, які здійснювались за певний період часу.

Отже, можна визначити таких акторів: менеджер – співробітник, який обробляє замовлення та отримує інформацію щодо діяльності магазину, система онлайн продажів.

Виділимо такі варіанти використання даної інформаційної системи: перегляд основної інформації щодо роботи магазину за певний період, аналіз клієнтів (сегментація), аналіз товарних послідовностей (побудова асоціативних правил), завантаження інформації про транзакції.

На основі вихідних даних було побудовано діаграму прецедентів, яка описує загальну схему взаємодії акторів у межах інформаційної системи (кресленик ІС91.290БАК.004 Д1).

Основний потік варіанта використання "Завантажити дані" починається після запуску системи.

Система після запуску звертається до вивантаженого файлу з БД онлайн-продажів та завантажує його.

Користувач бачить на екрані повідомлення про початок роботи та можливі варіанти подальшого використання системи

Альтернативний потік. В разі відсутності файлу видається повідомлення про помилку. Користувач може скопіювати файл в каталог та перезавантажити сесію роботи з системою.

Основний потік варіанта використання "Переглянути основну статистику".

Користувач обирає з запропонованого системою період для аналізу.

					ІС91.290БАК.004 ПЗ	Арк.
						41
Змн.	Арк.	№ докум.	Підпис	Дата		

Система з завантаженого файла формує вибірку згідно заданої умови.

Користувач обирає варіант "Основна статистика".

Система на основі відібраних даних розраховує основні показники та виводить їх користувачу.

Альтернативний потік. Користувач може змінити період аналізу.

Основний потік варіанта використання "Побудувати профілі клієнтів".

Користувач обирає з запропонованого системою період для аналізу.

Система з завантаженого файла формує вибірку згідно заданої умови.

Користувач обирає варіант "Побудувати профілі клієнтів".

Система на основі відібраних даних здійснює RFM-аналіз, видає користувачу інформацію.

Альтернативний потік. Користувач може використати період, який був обраний для попереднього аналізу.

Основний потік варіанта використання "Побудувати асоціативні правила для товарів".

Користувач обирає з запропонованого системою період для аналізу.

Система з завантаженого файла формує вибірку згідно заданої умови.

Користувач обирає варіант "Побудувати асоціативні правила для товарів".

Система на основі відібраних даних здійснює побудову правил, видає користувачу інформацію.

Альтернативний потік. Користувач може використати період, який був обраний для попереднього аналізу.

Для подальшого моделювання системи використаємо діаграми послідовності (англ. sequence diagram), де показані взаємодії об'єктів, упорядковані за часом їх прояви. Основними елементами діаграми послідовності є позначення об'єктів (прямокутники), вертикальні лінії (англ. lifeline), що відображають протягом часу при діяльності об'єкта, та стрілки, що показують виконання дій об'єктами. На цій діаграмі об'єкти розташовуються зліва направо. Її недоліком є те, що вона займає багато місця.

					IC91.290БАК.004 ПЗ	Арк.
						42
Змн.	Арк.	№ докум.	Підпис	Дата		

Відобразимо на діаграмах послідовності варіанти використання "Показати основну статистику", "Побудувати профілі клієнтів", "Побудувати асоціативні правила" (рисунки 2.3-2.5).

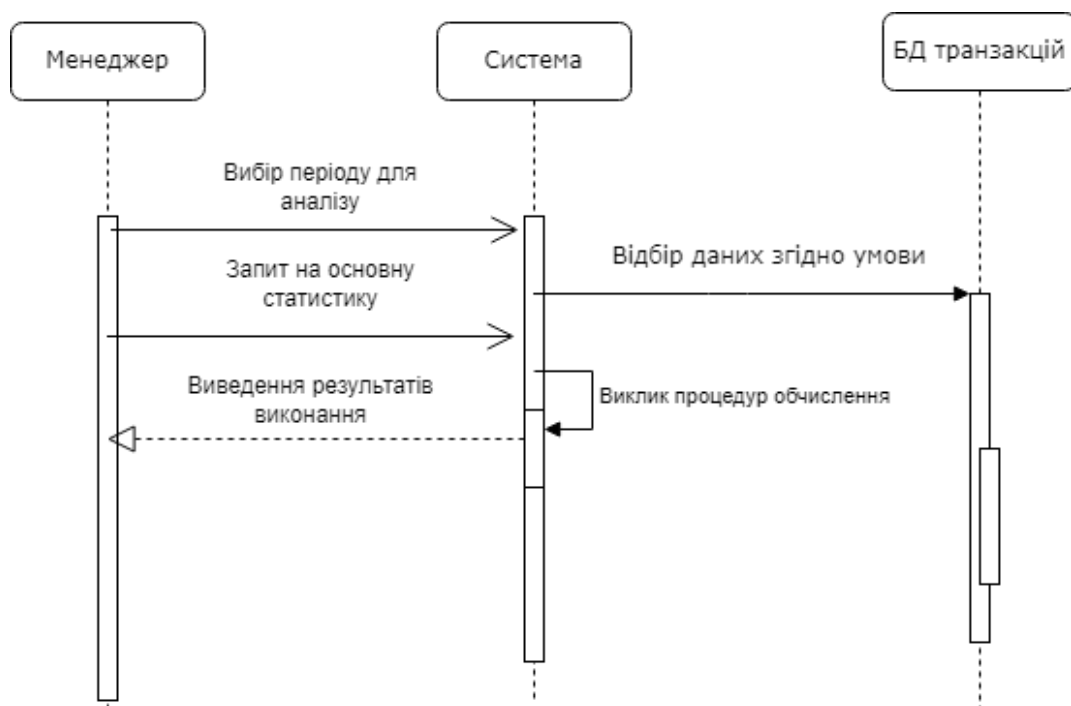


Рисунок 2.3 – Діаграма послідовності для варіанта використання «Переглянути основну статистику»

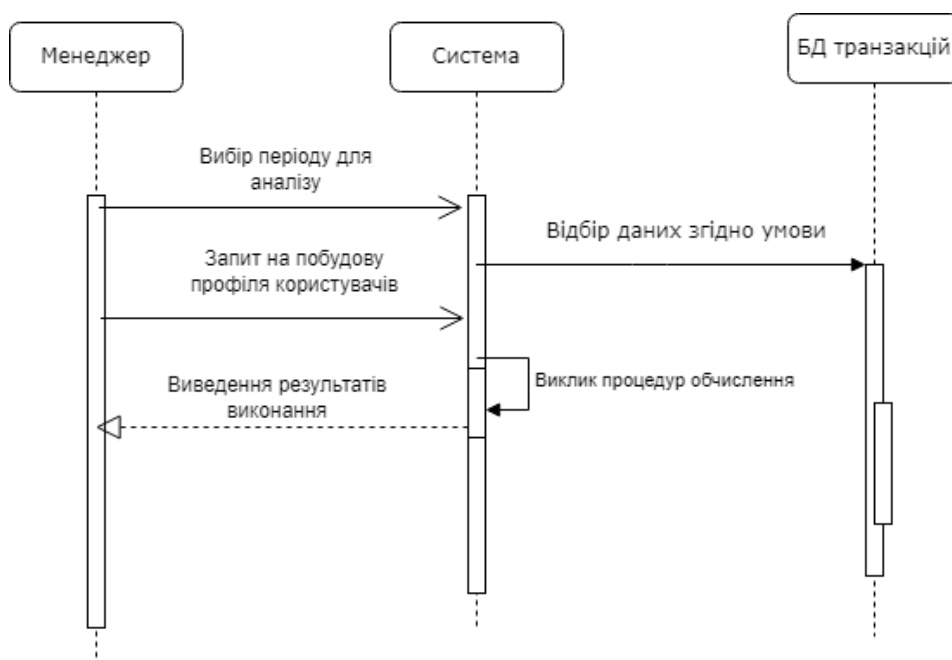


Рисунок 2.4 – Діаграма послідовності для варіанта використання «Переглянути основну статистику»

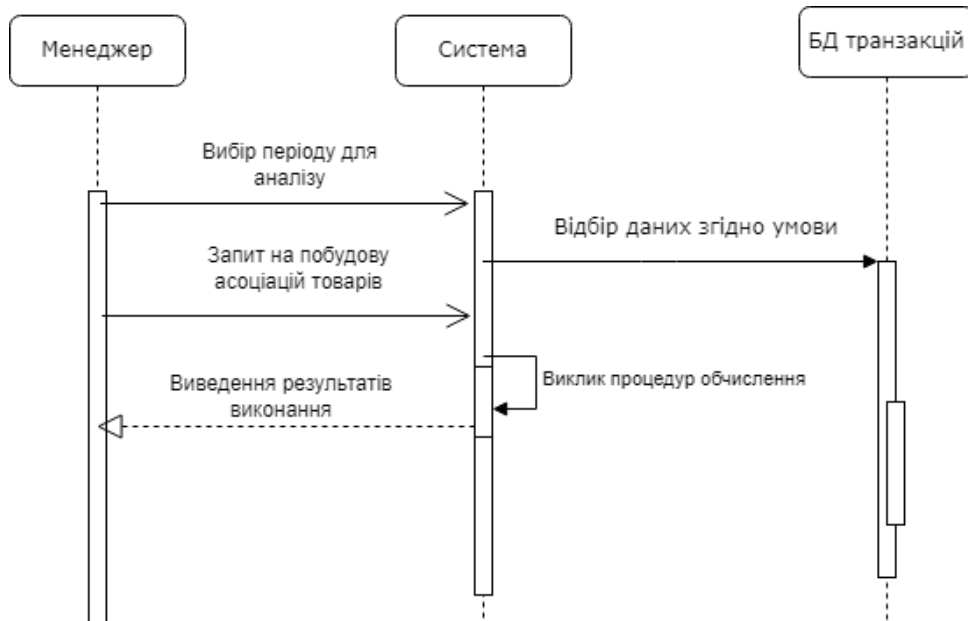


Рисунок 2.5 – Діаграма послідовності для варіанта використання
«Переглянути основну статистику»

На всіх цих діаграмах відображена послідовність дій користувача при виконанні того чи іншого варіанту використання.

Для варіанту використання "Завантажити дані" побудуємо діаграму активності, яка дозволяє більш детально візуалізувати варіант використання. Вона використовується для побудови потоку діяльності через систему. В нашому випадку вона буде відображати послідовність запуску системи (рисунок 2.6).

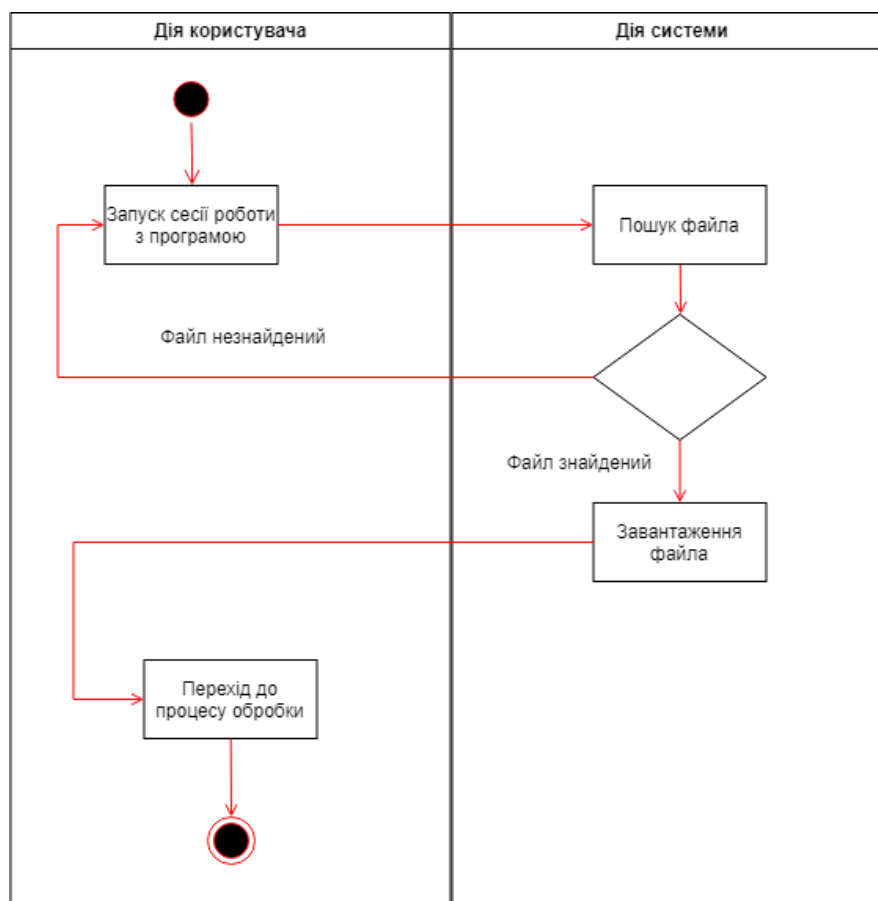


Рисунок 2.6 – Діаграма активності для варіанта "Завантажити дані"

Розроблювальну програму передбачається створювати у вигляді веб-додатку. За відображення даних, що формуються додатком відповідає компонент View, який представляється пакетом streamlit. Даний компонент відповідає за представлення інформації та результатів обробки даних. Дані, що поступають для обробки, надходять від системи транзакцій онлайн-магазину. Отже, діаграму компонентів (component diagram) можна показати таким чином, як зображено на рисунку 2.7.

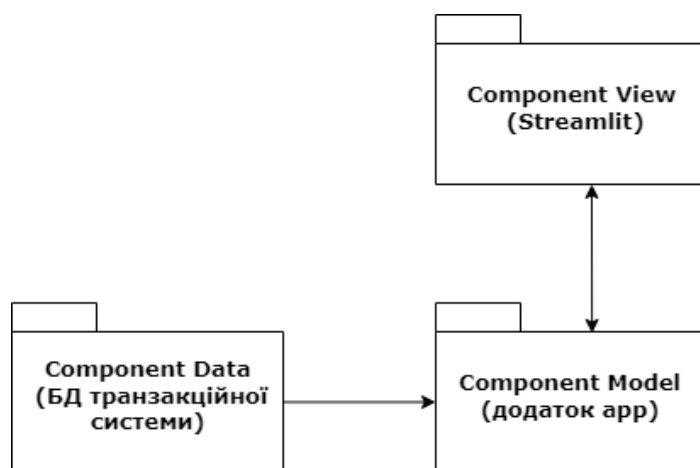


Рисунок 2.7 – Діаграма компонентів розроблювальної системи

2.3 Розробка інфологічного забезпечення інформаційної системи

Перейдемо до процесу проектування моделі даних. Оскільки система для своєї роботи використовує дані від транзакційної системи, то необхідно визначити вимоги до формату даних, що будуть використовуватись в системі.

Зазначимо, що формат видачі будь-якої транзакційної системи є таблиці, які формуються у відповідності з зовнішніми запитами. Тому проектувана система буде отримувати інформацію для обробки у табличному вигляді. Оскільки завжди існують певні проблеми з представленням різних даних, то такі файли представляють у форматі CSV, текстовий формат, призначений для представлення табличних даних. В цьому форматі рядок таблиці відповідає рядку тексту, який містить одне або кілька полів, розділених комами. Зазвичай, більшість мов програмування мають засоби для обробки таких файлів. Крім того, такі файли завжди менші за розміром, на відміну від файлів, що представляються в звичайному табличному форматі.

Для роботи системи кожний рядок даного файла повинен представляти собою одну транзакція, яка була здійснена в онлайн магазині.

Кожна транзакція містить таку інформацію (в дужках зазначені назви полів):

1. Номер чеку – ціле число, InvoiceNo.
2. Код товару – буквено-цифрова послідовність, StockCode.

3. Опис товару – текст, Description.
4. Кількість товару – ціле число, Quantity.
5. Дата покупки – дата формату PPPP-ММ-ДД ГГ:ХХ:СС, InvoiceDate.
6. Ціна одиниці товару – число, UnitPrice.
7. Код покупця – цифрова послідовність формату ЦЦЦЦЦ.Ц CustomerID,
8. Країна – текст, Country,
9. Рік – ціле число, Year.
10. Місяць – ціле число в межах від 1 до 12, Month.
11. Дата – ціле число в межах від 1 до 31, Day.
12. Квартал – ціле число в межах від 1 до 4, Quarter,
13. Загальна вартість покупки – число. TotalPrice.

Висновки до розділу 2

В результаті проведеного дослідження було проведено аналіз методів, які можуть використовуватись при аналізі статистики онлайн магазинів. Їх можна представити у вигляді методів, які описують результати роботи підприємства, та методів, що дозволяють спрогнозувати та прийняти рішення щодо подальшої діяльності.

Були розглянуті методи, які використовуються для аналізу покупців магазину та для аналізу товарів, а саме, RFM-аналіз та аналіз на основі асоціативних правил.

Було проведено проектування системи аналізу статистики інтернет-магазину на основі об'єктно-орієнтованого підходу, в результаті чого була отримана низка моделей, які описують діяльність системи.

					ІС91.290БАК.004 ПЗ	Арк.
						47
Змн.	Арк.	№ докум.	Підпис	Дата		

3 РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ СТАТИСТИЧНОГО АНАЛІЗУ ДІЯЛЬНОСТІ ІНТЕРНЕТ-МАГАЗИНУ

3.1. Розробка алгоритму роботи системи

Аналіз статистики роботи інтернет-магазину повинен охоплювати всі складові його ефективності, а саме, продажі, клієнтів та товари. Тому розроблювальна система повинна забезпечувати можливість проаналізувати загальні показники діяльності, профілі клієнтів, та товарів. Для цього пропонується обчислювати такі показники, як:

- загальна виручка за період, середня вартість чеку, кількість продажів, сума щоденної виручки для характеристики загальної діяльності підприємства;
- сегментацію клієнтів (активні, сплячі тощо) для характеристики профілів клієнтів;
- популярність груп товарів (асоціативні правила) для характеристики товарів.

Для користувача повинна існувати можливість вибору як періоду, за який він хоче отримати аналіз даних, так і вид аналізу, який він хоче здійснити. Це забезпечить гнучкість системи та можливість її налаштування для окремого менеджера.

Загальний алгоритм роботи програми з точки зору користувача наведений на кресленику ІС91.290БАК.004 Д2.

Опишемо більш детально алгоритми аналізу, що використовуються в даній системі.

Алгоритм розрахунку основних показників діяльності підприємства є лінійним алгоритмом.

Будемо вважати, що відбір даних за період був виконаний, і кількість транзакцій становить N .

Алгоритм. Розрахунок основних статистик

ВХІД. Датасет, що містить транзакції.

					ІС91.290БАК.004 ПЗ	Арк.
						48
Змн.	Арк.	№ докум.	Підпис	Дата		

Алгоритм

1. Розрахунок загальної суми продажів $\sum_{i=1}^N TotalPrice_i$
2. Розрахунок кількості покупок $count(InvoiceNo)$
3. Розрахунок середнього значення чеку $\frac{\sum_{i=1}^N TotalPrice_i}{count(InvoiceNo)}$
4. Розрахунок середньої кількості покупців за день $\frac{\sum_{i=1}^N CustomerID_i}{T_{кін} - T_{поч}}$
5. Розрахунок найпопулярнішого товару $\max_i count(StockCode_i)$
6. Розрахунок найвигіднішого товару $\max_i \sum_{StockCode_i} TotalPrice$

Наведемо алгоритм RFM-аналізу [14]

ВХІД. Датасет, що містить транзакції.

1. Розрахунок дати останньої покупки для всіх $CustomerID_i$
2. Розрахунок кількості покупок для кожного $CustomerID_i$
3. Розрахунок загальної вартості покупок для кожного $CustomerID_i$
4. Визначення рангів значення R_rank для дат останньої покупки
5. Визначення рангів значення F_rank для частоти покупок
6. Визначення рангів значення M_rank для суми покупок
7. Розрахунок нормалізованих рангів
8. Розрахунок порогового значення

$$Score_i = 0.15 \cdot R_{rank_i} + 0.28 \cdot F_{rank_i} + 0.57 \cdot M_{rank_i}$$

9. Визначення сегменту користувача за правилом

$$Cust_Segment_i = \begin{cases} Top\ Customers, & \text{if } Score_i > 4.5 \\ High\ value\ Customers, & \text{if } 4 < Score_i < 4.5 \\ High\ value\ Customers, & \text{if } 4 < Score_i < 4.5 \\ Medium\ value\ Customers, & \text{if } 3 < Score_i < 4 \\ Low\ value\ Customers, & \text{if } 1.6 < Score_i < 3 \\ Lost, & \text{if } Score_i < 1.6 \end{cases}$$

Наведемо алгоритм побудови асоціативних правил [18]

ВХІД. Датасет, що містить транзакції, δ – порогове значення

1. $C_1 \leftarrow [\{i\}: i \text{ in } J]$

2. $k \leftarrow 1$

3. Поки $C_k \neq 1$ виконувати

4. Підраховуються всі *support* для всіх кандидатів

Для всіх транзакцій $(tid, I) \in D$ виконати

Для всіх кандидатів $X \in C_k$ виконати

Якщо $X \in I$ $support(X)++$

5. Видобуваємо всі часті елементи $F_k = \{X | support(X) > \delta\}$

6. Генеруються нові кандидати

$\forall X, Y \in F_k, X[i] = Y[i] \text{ for } 1 \leq i \leq k-1 \text{ and } X[k] \leq Y[k] \text{ do}$

$I = X \cup \{Y[k]\}$

$\text{if } \forall J \subseteq I, |J| = k, J \in F_k \text{ then } C_{k+1} \leftarrow C_{k+1} \cup I, k++$

3.2 Обґрунтування засобів розробки

Вибір мови програмування та середовища розробки обумовлювався кількома вимогами:

- розроблена цією мовою програма повинна працювати під більшістю кількістю ОС;
- середовище розробки має бути загальнодоступним і недорогим або безкоштовним;
- програма повинна працювати на різних платформах. Це поширює область застосування цієї системи;
- бажана можливість реалізації системи у вигляді веб-додатку, що в деяких випадках допомагає уникати процес установки на комп'ютер, оптимізувати процес оновлення баз, керувати та змінювати інтерфейс програми в реальному часі.

Одним із поширених засобів розробки є високорівнева мова програмування Python. Python має підтримку різних парадигм програмування (структурного, об'єктно-орієнтованого функціонального та інших); у ньому використовується динамічна типізація змінних, забезпечується періодичне звільнення пам'яті від

					ІС91.290БАК.004 ПЗ	Арк.
						50
Змн.	Арк.	№ докум.	Підпис	Дата		

об'єктів, що не використовуються. Python володіє досить лаконічним синтаксисом, з більш високим рівнем абстракції, порівняно з PHP, та дозволяє передати таку кількість інформації в меншому обсязі вихідного коду [27; 28].

Python є мовою загального призначення, тому може застосовуватися практично у будь-якій галузі розробки ПЗ (standalone, клієнт-сервер, Web-додатки) та у будь-якій предметній області. Крім того, Python легко інтегрується з вже існуючими компонентами, що дозволяє впроваджувати Python вже написані програми [29].

Також Python має розвинену систему модулів. Це можуть бути як стандартні модулі, також написаних на Python, які мають такі ж переваги, як кросплатформність, і дозволяють реалізовувати найбільш загальні завдання. Крім того існує багато бібліотек для реалізації різних спеціальних задач, таких, як машинне навчання, обробка зображень, рішення задач оптимізації тощо.

До середовища виконання Python входить лише інтерпретатор, який одночасно є і компілятором, однак компілює вихідний код Python безпосередньо в цільовий машинний код платформи.

Через широку низку функціональних можливостей Python може використовуватися в самих різних сферах; за його допомогою можна здійснювати розробку веб-додатків, автоматизованих інформаційних систем, наукових обчислювальних комплексів, графічних пакетів тощо [29].

Все це обумовило вибір даної мови програмування для реалізації інформаційної системи.

В Python існує доволі багато різних бібліотек, які дозволяють створювати веб-додатки. Найбільш звичною зв'язкою є використання Python у зв'язці з Django. Серед популярних фреймворків є Flask. Але проблема з використанням цих фреймворків полягає в тому, що розробник повинен мати певні знання про HTML, CSS і JavaScript.

Реалізація системи аналізу статистики сайту не потребує створення повноцінного сайту, оскільки вона більш орієнтована на обробку даних і скоріше

					IC91.290БАК.004 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		51

представляє собою дашборд, що забезпечує користувачу зручним інструментом для наочного подання даних. Внаслідок цього був обраний фреймворк Streamlit, який забезпечує створення веб-додатків з використанням лише мови Python [32].

Streamlit – це бібліотека Python з відкритим вихідним кодом, яка дозволяє легко створювати та ділитися спеціальними веб-програмами для машинного навчання та обробки даних. Вона дозволяє доволі просто створювати та розгортати потужні програми для обробки даних.

Виконання програми здійснюється через командний рядок Python.

Даний фреймворк успішно інтегрується з бібліотекою Pandas, яка використовувалась в даному додатку для обробки даних.

Pandas – це високорівнева бібліотека Python для аналізу даних. Вона побудована поверх низькорівневої бібліотеки NumPy (написана на C), що є великим плюсом у продуктивності. В екосистемі Python, Pandas є найбільш просунутою бібліотекою, що швидко розвивається, для обробки та аналізу даних. Основою даної бібліотеки виступають такі структури даних, як DataFrame та Series [33; 34].

3.3 Тестування роботи системи

Реалізація системи здійснювалась на основі модульного підходу. Всі розрахунки здійснюються в окремих модулях, які імпортуються в основну програму. Структура проекту подана на рисунку 3.1.

					ІС91.290БАК.004 ПЗ	Арк.
						52
Змн.	Арк.	№ докум.	Підпис	Дата		

User > AppData > Local > Programs > Python > Python310 > data_analysis >					Поиск в: data_analysis
Имя	Дата изменения	Тип	Размер		
__pycache__	24.04.2023 20:52	Папка с файлами			
analiz_app	24.04.2023 20:27	Python File	5 КБ		
associate_rule	24.04.2023 20:52	Python File	2 КБ		
customer_retention_rate	31.01.2023 19:29	Python File	1 КБ		
data	31.01.2023 19:29	Файл Microsoft E...	42 729 КБ		
filter_by_year_and_quarter	31.01.2023 19:29	Python File	1 КБ		
mean_customers_per_day	31.01.2023 19:29	Python File	1 КБ		
median_order_value	31.01.2023 19:29	Python File	1 КБ		
most_popular_product	31.01.2023 19:29	Python File	1 КБ		
most_profitable_product	31.01.2023 19:29	Python File	2 КБ		
order_value_variability	31.01.2023 19:29	Python File	1 КБ		
rfm	24.04.2023 0:40	Python File	3 КБ		
total_sales	31.01.2023 19:29	Python File	1 КБ		

Рисунок 3.1 – Структура проекту

Як зазначалось, файл з даними повинен відповідати заданому формату. Для тестування використовувався файл, сформований умовною CRM. Структура файлу наведена на рисунку 3.2. Файл містить інформацію про 3973925 транзакцій, здійснених в період з грудня 2021 року по грудень 2022 року.

data - Excel																		Поиск	Вход	—	□	×
Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Справка Нагрузочный тест Acrobat Рабочая группа Поделиться Вставить Буфер обмена Шрифт Выравнивание Число Условное форматирование Стили Ячейки Общий Форматировать как таблицу Стили Сортировка и фильтр Найти и выделить																						
ВОЗМОЖНА ПОТЕРЯ ДАННЫХ При сохранении книги в формате с разделителями-запятыми (CSV-файл) некоторые возможности могут быть потеряны. Если вы не хотите терять эти возможности, сохраните ее в формате файла Excel. Больше не показывать Сохранить как...																						
A1 InvoiceNo,StockCode,Description,Quantity,InvoiceDate,UnitPrice,CustomerID,Country,Year,Month,Day,Quarter,TotalPrice																						
1	InvoiceNo,StockCode,Description,Quantity,InvoiceDate,UnitPrice,CustomerID,Country,Year,Month,Day,Quarter,TotalPrice																					
2	536365,85123A,WHITE HANGING HEART T-LIGHT HOLDER,6,2021-12-01 08:26:00,2.55,17850.0,United Kingdom,2021,12,1,4,15.299999999999999																					
3	536365,71053,WHITE METAL LANTERN,6,2021-12-01 08:26:00,3.39,17850.0,United Kingdom,2021,12,1,4,20.34																					
4	536365,844068,CREAM CUPID HEARTS COAT HANGER,8,2021-12-01 08:26:00,2.75,17850.0,United Kingdom,2021,12,1,4,22.0																					
5	536365,840296,KNITTED UNION FLAG HOT WATER BOTTLE,6,2021-12-01 08:26:00,3.39,17850.0,United Kingdom,2021,12,1,4,20.34																					
6	536365,84029E,RED WOOLLY HOTTIE WHITE HEART.,6,2021-12-01 08:26:00,3.39,17850.0,United Kingdom,2021,12,1,4,20.34																					
7	536365,22752,SET 7 BABUSHKA NESTING BOXES,2,2021-12-01 08:26:00,7.65,17850.0,United Kingdom,2021,12,1,4,15.3																					
8	536365,21730,GLASS STAR FROSTED T-LIGHT HOLDER,6,2021-12-01 08:26:00,4.25,17850.0,United Kingdom,2021,12,1,4,25.5																					
9	536366,22633,HAND WARMER UNION JACK,6,2021-12-01 08:28:00,1.85,17850.0,United Kingdom,2021,12,1,4,11.100000000000001																					
10	536366,22632,HAND WARMER RED POLKA DOT,6,2021-12-01 08:28:00,1.85,17850.0,United Kingdom,2021,12,1,4,11.100000000000001																					
11	536367,84879,ASSORTED COLOUR BIRD ORNAMENT,32,2021-12-01 08:34:00,1.69,13047.0,United Kingdom,2021,12,1,4,54.08																					
12	536367,22745,POPPY'S PLAYHOUSE BEDROOM,6,2021-12-01 08:34:00,2.1,13047.0,United Kingdom,2021,12,1,4,12.600000000000001																					
13	536367,22748,POPPY'S PLAYHOUSE KITCHEN,6,2021-12-01 08:34:00,2.1,13047.0,United Kingdom,2021,12,1,4,12.600000000000001																					
14	536367,22749,FELTCRAFT PRINCESS CHARLOTTE DOLL,8,2021-12-01 08:34:00,3.75,13047.0,United Kingdom,2021,12,1,4,30.0																					
15	536367,22310,IVORY KNITTED MUG COSY,6,2021-12-01 08:34:00,1.65,13047.0,United Kingdom,2021,12,1,4,9.899999999999999																					
16	536367,84969,BOX OF 6 ASSORTED COLOUR TEASPOONS,6,2021-12-01 08:34:00,4.25,13047.0,United Kingdom,2021,12,1,4,25.5																					
17	536367,22623,BOX OF VINTAGE JIGSAW BLOCKS,3,2021-12-01 08:34:00,4.95,13047.0,United Kingdom,2021,12,1,4,14.850000000000001																					
18	536367,22622,BOX OF VINTAGE ALPHABET BLOCKS,2,2021-12-01 08:34:00,9.95,13047.0,United Kingdom,2021,12,1,4,19.9																					

Рисунок 3.2 – Файл з умовними даними щодо роботи онлайн-магазину

Фреймворк Streamlit надає можливість розмістити додатки в мережі. Для цього використовується Streamlit Community Cloud, де можна зареєструватися за допомогою акаунта Google, GitHub або електронної пошти. Через Streamlit Community Cloud можна розгортати за кілька хвилин. Для додатка виділяється унікальний піддоменний URL, що забезпечує можливість перегляду його іншими

					ІС91.290БАК.004 ПЗ		Арк.
Змн.	Арк.	№ докум.	Підпис	Дата			53

користувачами. Таким чином, додаток може бути доступний для роботи декільком працівникам підприємства.

Для тестування локальний комп'ютер за адресою <http://localhost:8501>.

Після запуску додатку через фреймворк в браузері з'являється вкладка з запущеним додатком (рисунок 3.3). Якщо файл вдалося завантажити, то користувачу пропонується найперший період за замовчуванням (в нашому випадку, це 4 квартал 2021 року).

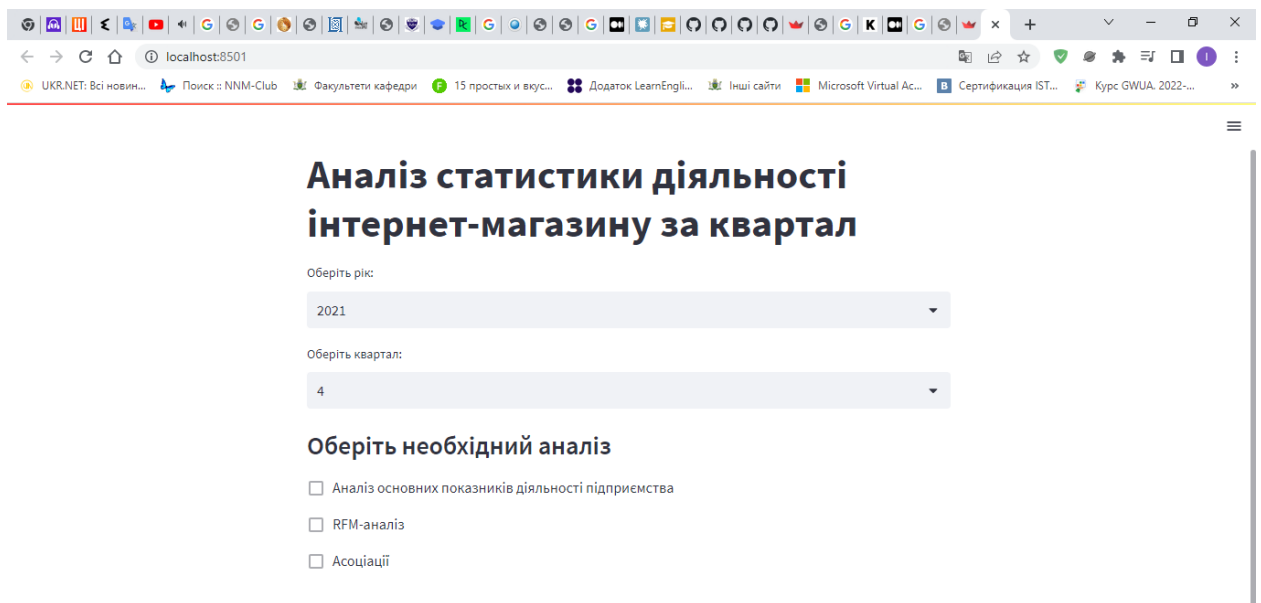


Рисунок 3.3 – Початкова сторінка додатку

Користувач може залишити період без змін або обрати інший період (рік та квартал відповідно). Оскільки дані вибираються зі списку, який формується на основі завантажених даних, то у користувача немає можливості ввести неправильне значення.

					ІС91.290БАК.004 ПЗ	Арк.
						54
Змн.	Арк.	№ докум.	Підпис	Дата		

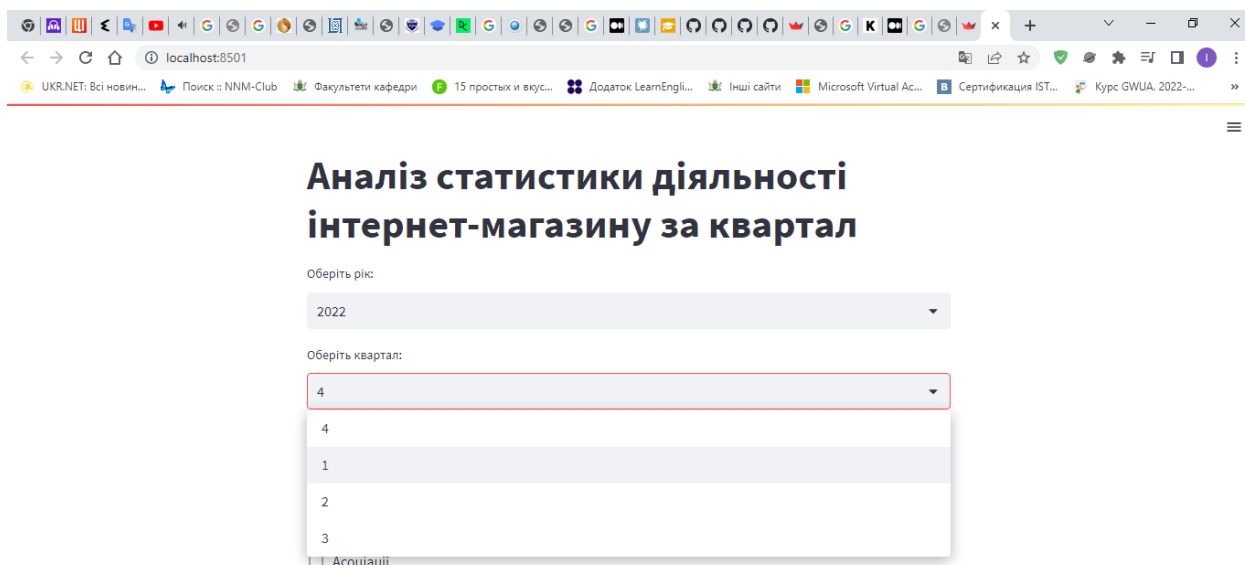


Рисунок 3.4 – Вибір періоду для аналізу

Далі користувачу пропонується три варіанти аналізу. Оскільки додаток відображає результати аналізу, то інформація щодо декількох видів аналізу може залишатися одночасно у вікні браузера. Для того, щоб запустити аналіз потрібно встановити відповідний прапорець.

На рисунках 3.5-3.8 відображений основний аналіз результатів роботи онлайн-магазину.

На рисунку 3.5 показані розрахунки загальної суми, середньої кількості покупців в день, середня вартість замовлення та стандартне відхилення замовлення (для визначення на скільки обчислена статистика відображає сукупність), найбільш популярний товар, найбільш вигадний товар та % утриманих покупців.

					ІС91.290БАК.004 ПЗ	Арк.
						55
Змн.	Арк.	№ докум.	Підпис	Дата		

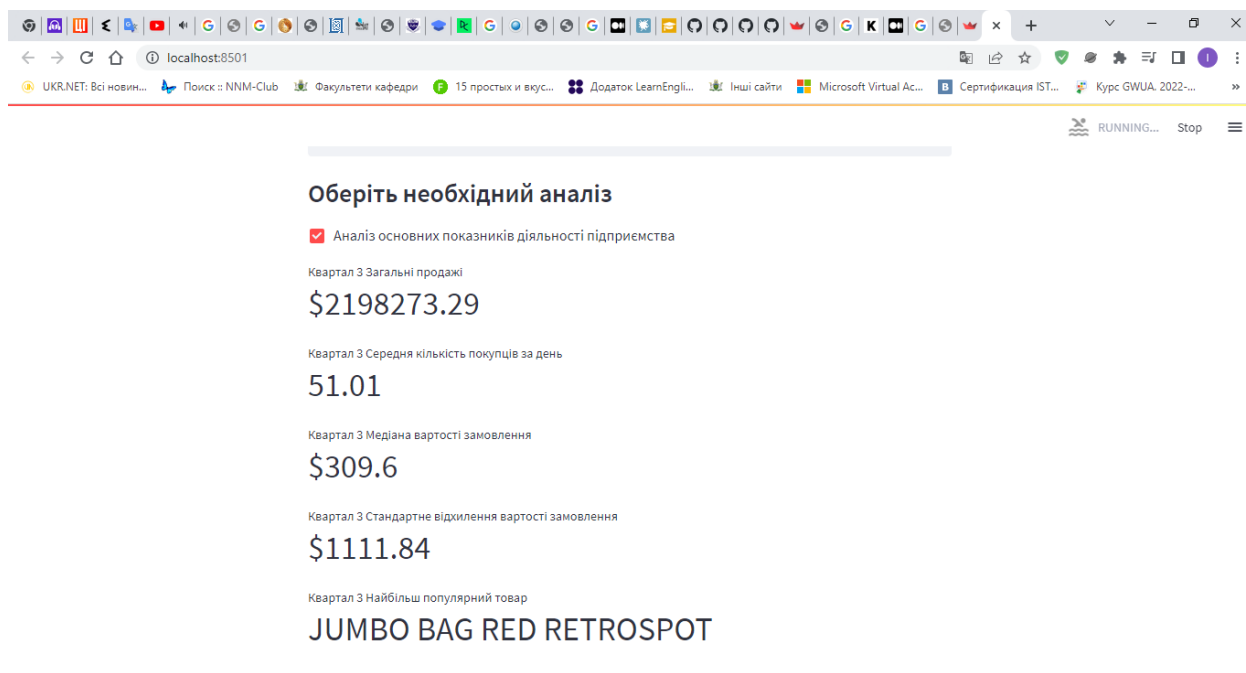


Рисунок 3.5 – Розрахунок основних показників онлайн-магазину

На рисунку 3.6 відображається графік розподілу цін на товари, які продавалися за цей період, що дає змогу визначити дорогі та дешеві товари та побачити ціновий діапазон проданих товарів.

На рисунку 3.7 відображається динаміка щоденних продажів за обраний період, що дає можливість виявити пікові дні продажів, визначити періодичність і, таким чином, планувати подальшу діяльність магазину.

Відмітимо, що використання відповідної бібліотеки Python `plotpy` дає змогу будувати інтерактивні графіки, для яких можна змінювати масштаб, переглядати значення точок даних. Бібліотека також надає інструменти для зберігання графіків у форматі `png` для подальшого використання, наприклад, у звітних документах (рисунки 3.7).

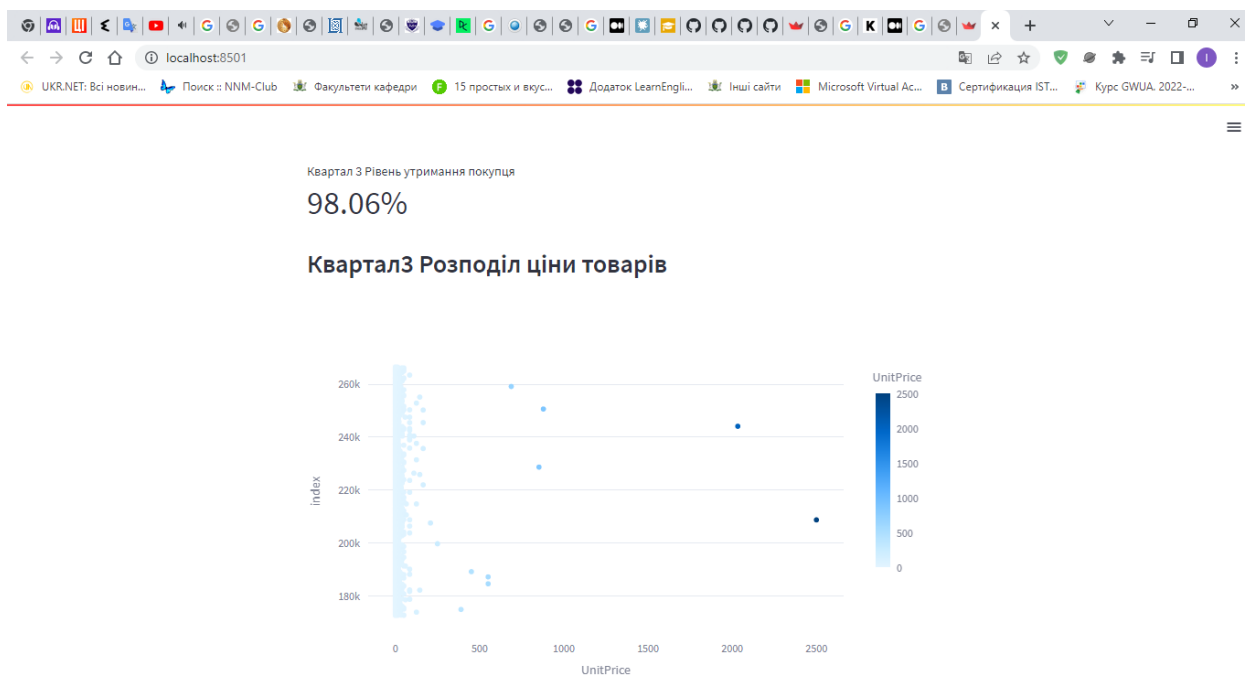
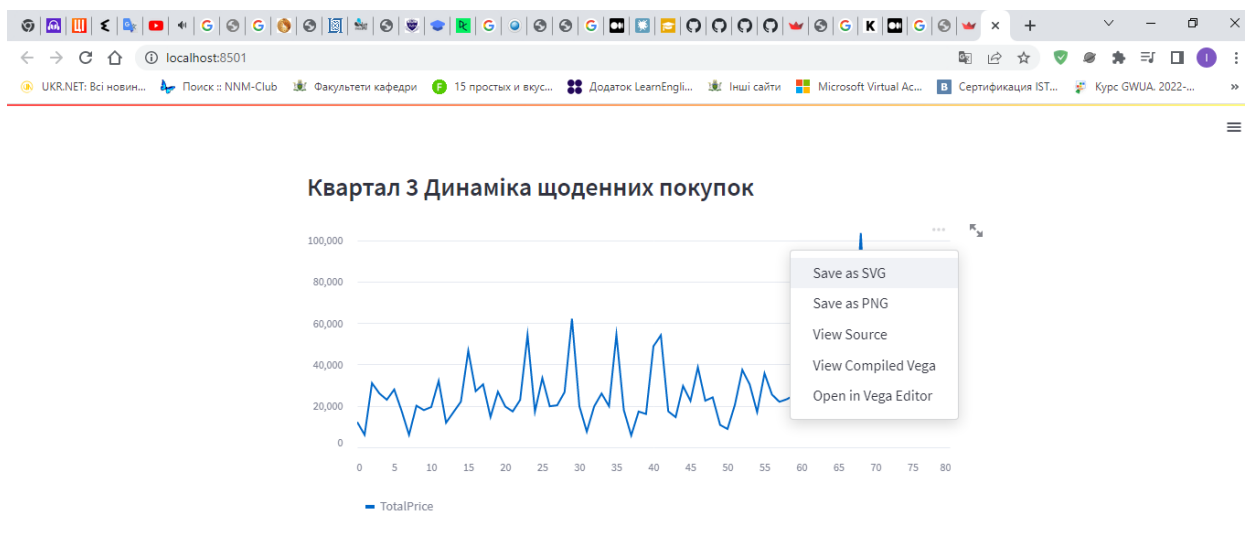


Рисунок 3.6 – Графік розподілу цін на товари в магазині



Квартал 3 Розподіл продажів за країнами

localhost:8501/#

Рисунок 3.7 – Динаміка щодених продажів в онлайн-магазині

Останнім графіком в цьому блоці є графік розподілу покупців за країнами (рисунок 3.8), що дає можливість більш точно вибудувати стратегію магазину.



Рисунок 3.8 – Розподіл покупців за країнами

Для відміни виведення даних аналізу користувач може зняти прапорець, що призводить до вилучення відповідної інформації зі сторінки додатку.

Наступний вид аналізу, що доступний користувачу, – це RFM-аналіз. Даний аналіз проводиться за обраний період.

Результати аналізу виводяться у вигляді інтерактивної таблиці (рисунок 3.9), де зазначаються ідентифікаційні номери користувачів, їх рейтинговий бал та належність до певного сегменту. Користувач може задавати порядок сортування, виокремити таблицю, виділити потрібні значення.

RFM-аналіз

	CustomerID	RFM_Score	Customer_segment
0	12,346	0.06	Lost Customers
1	12,347	4.48	High value Customer
2	12,348	2.09	Low Value Customers
3	12,349	3.41	Medium Value Customer
4	12,350	1.1	Lost Customers
5	12,352	3.46	Medium Value Customer
6	12,353	0.33	Lost Customers
7	12,354	2.67	Low Value Customers
8	12,355	0.93	Lost Customers
9	12,356	3.11	Medium Value Customer

Рисунок 3.9 – Таблиця з результатами RFM-аналізу

Також за результатами аналізу виводиться кругова діаграма, яка показує відсоток клієнтів, що належать до цієї чи іншої групи (рисунок 3.10).

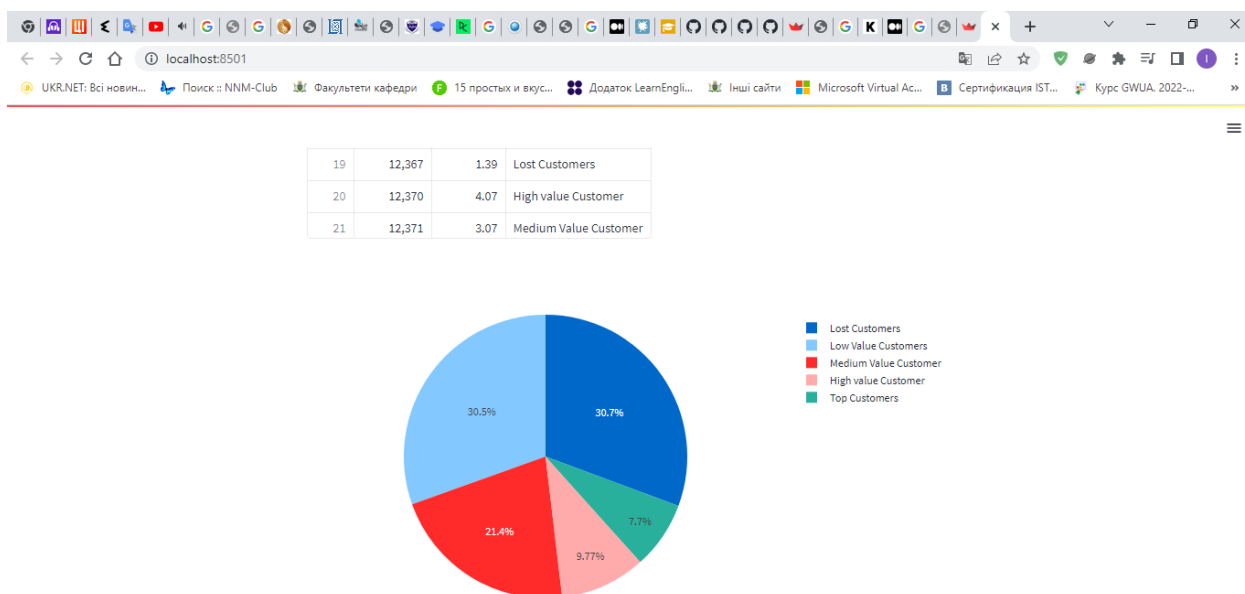


Рисунок 3.10 – Діаграма розподілу клієнтів за групами

Наступний аналіз, що буде виконуватись – це побудова асоціацій між товарами. Даний аналіз є доволі важким як за розміром даних, що опрацьовуються,

так і за машинними часом. Тому було вирішено на локальному комп'ютері проводити його для місячних даних, що дозволить знизити навантаження.

При перенесенні додатка в хмару можна буде скористатися потужностями сервера і перейти до побудови квартальних асоціацій.

На рисунку 3.11 наведені визначені правила покупки відповідних товарів у вигляді інтерактивної таблиці.

Асоціації

Оберіть місяць:

2

	antecedents	consequents	support	confidence	↓ lift
1,667	frozenset(["21464"])	frozenset(["21463"])	0.006	1	124.875
1,666	frozenset(["21463"])	frozenset(["21464"])	0.006	0.75	124.875
4,677	frozenset(["22685"])	frozenset(["22683"])	0.006	0.8571	122.3265
4,676	frozenset(["22683"])	frozenset(["22685"])	0.006	0.8571	122.3265
4,673	frozenset(["22685"])	frozenset(["22682"])	0.006	0.8571	122.3265
4,672	frozenset(["22682"])	frozenset(["22685"])	0.006	0.8571	122.3265
4,671	frozenset(["22682"])	frozenset(["22683"])	0.006	0.8571	122.3265
4,670	frozenset(["22683"])	frozenset(["22682"])	0.006	0.8571	122.3265
4,669	frozenset(["22681"])	frozenset(["22686"])	0.008	0.8889	111
4,668	frozenset(["22686"])	frozenset(["22681"])	0.008	1	111

Рисунок 3.11 – Побудовані асоціативні правила

На рисунку 3.12 побудований тривимірний графік розподілу правил.

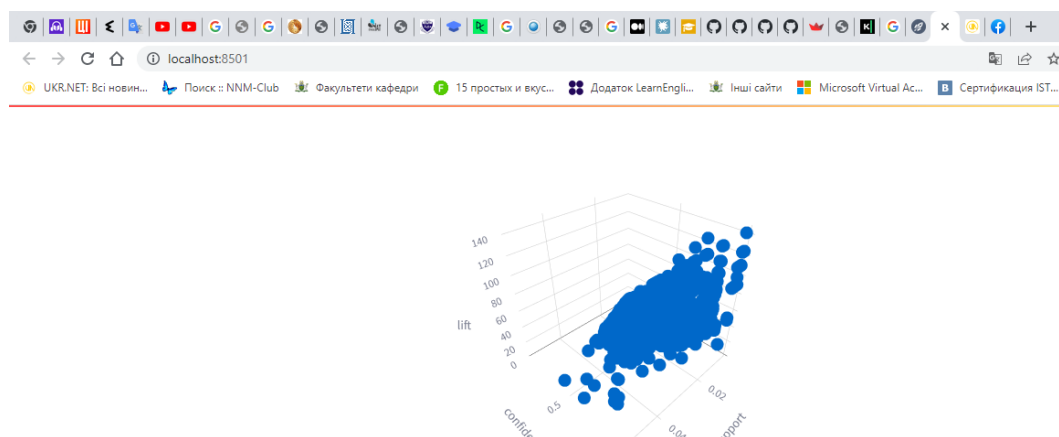


Рисунок 3.12 – Графік розподілу правил відносно їх характеристик

Отже, розроблена система дає можливість досить чітко проаналізувати дані статистики інтернет-магазину. На основі такого аналізу можна приймати рішення щодо ефективності роботи магазину, підвищення ступеню лояльності клієнта, співвідношенні в акціях тих чи інших товарів.

Висновки до розділу 3

В даному розділі були описані загальний алгоритм роботи системи та алгоритми роботи окремих модулів системи.

Було обґрунтовано вибір мови Python як мови реалізації проекту. Для розробки веб-додатку був обраний фреймворк Streamlit, що дозволило використовувати лише засоби мови Python для розробки.

Проведене тестування показало відповідність розробленої програми поставленим завданням.

					ІС91.290БАК.004 ПЗ	Арк.
						61
Змн.	Арк.	№ докум.	Підпис	Дата		

ВИСНОВКИ

На основі проведеного дослідження зробимо такі висновки.

Роль і сприйняття Інтернету в різних контекстах його використання швидко змінюється – від раннього фокусування на поданні різноманітної інформації для клієнтів та іншими користувачами до того, коли Інтернет стає одним із центральних компонентів багатоканальної інформації та комунікаційної стратегії. Інтернет дає компаніям можливість збирати, аналізувати та використовувати все більшу кількість цифрової інформації про споживачів. Аналіз і використання онлайн-даних приносять переваги компаніям при розробці різних стратегій діяльності.

В ході даної роботи були отримані такі результати:

1. Аналіз даних щодо електронної комерції, як у світі, так і в Україні за останні роки показує, що, незважаючи на проблеми пов'язані з COVID-19 та війною, ринок онлайн торгівлі продовжує зростати. В Україні він, завмерши у перші місяці вторгнення, почав відновлюватись і доганяє передвоєнні темпи росту.

2. Існує досить багато різних метрик, на основі яких можна оцінити роботу Інтернет-магазину. Вони охоплюють як фінансову сторону діяльності підприємства, визначаючи його дохід від онлайн діяльності, так і взаємодію з клієнтами, яка зараз виходить на перший план. Сучасні споживачі вимагають постійного підвищення рівня своєї задоволеності. Тому в умовах жорсткої конкуренції підприємства повинні зосередити свої зусилля на утриманні клієнтів. Для цього підприємству краще орієнтуватися на свої внутрішні джерела інформації, які допоможуть виявити та утримати лояльних клієнтів.

3. Для обробки статистичних даних інтернет-магазинів використовуються широкий набір методів, починаючи від описової статистики і закінчуючи методами машинного навчання та штучного інтелекту. Ці методи дають можливість отримати з даних різнопланову інформацію. Так, RFM-аналіз дає змогу сегментувати базу клієнтів та виявити найбільш лояльних. Метод побудови

					ІС91.290БАК.004 ПЗ	Арк.
						62
Змн.	Арк.	№ докум.	Підпис	Дата		

асоціативних правил дозволяє визначити пари товарів, які можна пропонувати користувачу з великою ймовірністю покупки

4. Об'єктно-орієнтований підхід до процесу моделювання інформаційних систем дає змогу повно описати діяльність користувачів з розроблювальною системою. На основі даного підходу був розроблений проект інформаційної системи аналізу статистичних даних. Проект складається з набору UML діаграм: діаграми варіантів використання, діаграм послідовності, діаграми активності.

5. Були розглянуті алгоритми, які реалізують окремі блоки аналізу статистичних даних інтернет-магазину. Дані алгоритми були реалізовані в системі. Система реалізована на мові Python у вигляді веб-додатку, що знімає питання установки. Проведене тестування показало, що система працює у відповідності з вимогами.

Отже, всі завдання даного дослідження виконані.

					ІС91.290БАК.004 ПЗ	Арк.
						63
Змн.	Арк.	№ докум.	Підпис	Дата		

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Щербак О. Електронна торгівля. Українська та світова практика правового регулювання. *Юрист&Закон*, №43. URL: https://uz.ligazakon.ua/ua/magazine_article/EA016163 (дата звернення 20.04.2023)
2. Козицька Г. В. Розвиток міжнародної електронної торгівлі в умовах діджиталізації економіки. *Економічний вісник*, 2021, №2. С.101-114.
3. Краус К.М., Краус Н.М., Манжура О.В. Електронна комерція та Інтернет-торгівля: навчально-методичний посібник. Київ : Аграр Медіа Груп, 2021. 454 с.
4. Куклинова Т.В. Сучасні тенденції та фактори Інтернет-торгівлі в Україні. *Вісник соціально-економічних досліджень*, 2018, № 1(65). С.95-102
5. Metrics for Performance Measurement in E-Commerce URL: https://www.pearsoned.ca/highered/divisions/virtual_tours/trites/data/Trites_EBus_Ch1_2.pdf (дата звернення 20.04.2023)
6. Ahmed H, Tahseen D, Haider W, Asad M, Nand S, Kamran S. Establishing Standard Rules for Choosing Best KPIs for an E-Commerce Business based on Google Analytics and Machine Learning Technique. *International Journal of Advanced Computer Science and Applications*, 2017. Jan 1;8.
7. eCommerce analytics: What should you measure and why? URL: <https://factory.dev/blog/ecommerce-analytics> (дата звернення 20.04.2023)
8. Miroglia B, Zeber D, Kaye J, Weiss R. The Effect of Ad Blocking on User Engagement with the Web. *Proceedings of the 2018 World Wide Web Conference. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee*, 2018. P. 813–821.
9. Hill L. CRM: easier said than done. *Intelligent Enterprise*, 1999, 2(18):53.
10. Thearling K. Data mining and CRM: zeroing in on your best customers. *DM Direct*. December, 1999. URL: <http://www.dmreview.com /editorial/dmreview/print-action.cfm?EdID=1744> (дата звернення 20.04.2023)

					IC91.290БАК.004 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		64

11. Decision Support Solutions: Compaq. Object relational data mining technology for a competitive advantage. URL: <http://www.tandem.com/brfs—wps?odmadvwp/odmadvwp.htm> (дата звернення 20.04.2023)

12. Wong E., Wei Y. Customer online shopping experience data analytics: Integrated customer segmentation and customised services prediction model, *International Journal of Retail & Distribution Management*, 2018, Vol. 46 No. 4. P. 406-420.

13. Mittal M., Chauhan S., Gupta S., Singh H. A Statistical Analysis on Website Quality For Purchase Intention during Online Shopping, *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, Bhopal, India, 2021. P. 125-129.

14. Christy A. Joy, Umamakeswari A., Priyatharsini L., Neyaa A. RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*, 2021, Volume 33, Issue 10, P.1251-1257.

15. Ernawati E., Baharin S.S.K., Kasmin F. A review of data mining methods in RFM-based customer segmentation. *Journal of Physics: Conference Series*, 2021, 1869:012085.

16. Bhargava M., Selwal A. Association rule mining using Apriori algorithm: a review. *International Journal of Advanced Research in Computer Science*, 2013, Vol 4, No 2.

17. Haoyu X. Research and Case Analysis of Apriori Algorithm Based on Mining Frequent Item-Sets. *Open Journal of Social Sciences*. DOI: 09.458-468.10.4236/jss.2021.94034.

18. Agarwal P., Yadav M.L., Anand N. Study on Apriori Algorithm and its Application in Grocery Store. *International Journal of Computer Application*, 2013, July 74(14):1-8

19. Rainer R.K., Prince B. Introduction to Information Systems. Wiley, 2019. 560 p.

					IC91.290БАК.004 ПЗ	Арк.
						65
Змн.	Арк.	№ докум.	Підпис	Дата		

20. Fundamentals of Information Systems. URL: https://www.radford.edu/mhtay/ITEC110/Fundamental_Info_Sys/Lecture/ch01_5e.pdf (дата звернення 20.04.2023)

21. Dennis N.H. and Shirley D.G. Information Systems Foundations Theory, Representation and Reality. URL: <http://library.oapen.org/bitstream/id/a16645a2-7778-4c30-a227-47504caba732/459291.pdf>. (дата звернення 20.04.2023)

22. Порохня В.М. Експертні системи в економіці. Запоріжжя : ВПК «Запоріжжя», 2005. 272 с.

23. Проектування інформаційних систем: Загальні питання теорії проектування ІС (конспект лекцій): навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки» / КПІ ім. Ігоря Сікорського; уклад.: О. С. Коваленко, Л. М. Добровська. Електронні текстові дані (1 файл: 2,02 Мбайт). Київ : КПІ ім. Ігоря Сікорського, 2020. 192с.

24. Авраменко В.С., Авраменко А.С. Проектування інформаційних систем: навчальний посібник / В.С. Авраменко, А.С. Авраменко. Черкаси: Черкаський національний університет ім. Б. Хмельницького, 2017. 434 с.

25. Томашевський О. М., Цегелик Г. Г., Вітер М. Б., Дудук В. І. Інформаційні технології та моделювання бізнес-процесів : навчальний посібник. Київ : «Видавництво «Центр учбової літератури»», 2012. 296 с.

26. Литвин В.В., Пасічник В.В., Шаховська Н.Б. Проектування інформаційних систем. Навчальний посібник. Київ : Магнолія 2006, 2021. 380 с.

27. Joshi P. Artifical Intelligence with Python. Birmingham-Mumbai: Packt Publishing, 2017. 437 p.

28. Яковенко А.В. Основи програмування. Python. Частина 1. Підручник для студ. спеціальності 122 "Комп'ютерні науки". Київ : КПІ ім. Ігоря Сікорського, 2018. 195 с. URL: <https://ela.kpi.ua/bitstream/123456789/25111/1/Python.pdf>. (дата звернення 20.04.2023)

29. Путівник мовою програмування Python. URL: <https://pythonguide.rozh2sch.org.ua/> (дата звернення 20.04.2023)

					ІС91.290БАК.004 ПЗ	Арк.
						66
Змн.	Арк.	№ докум.	Підпис	Дата		

30. Goodrich M.T., Tamassia R., Goldwasser M.H. Data Structures and Algorithms in Python. Hoboken, NJ: Wiley, 2013. 748 p.

31. Лубко Д.В., Шаров С.В. Методи та системи штучного інтелекту: навч. посіб. Мелітополь : ФОП Однорог Т.В., 2019. 264 с.

32. Фреймворк Streamlit. URL: <https://streamlit.io/> (дата звернення 20.04.2023)

33. Бібліотека Pandas. URL: <https://pandas.pydata.org/> (дата звернення 20.04.2023)

34. Вивчаємо pandas. Урок 2. Структури даних Series і DataFrame. URL: <https://ruqui.tiosa.pp.ua/mismi/jak-zrobiti-tablicju-v-pandas.html> (дата звернення 20.04.2023)

					ІС91.290БАК.004 ПЗ	Арк.
						67
Змн.	Арк.	№ докум.	Підпис	Дата		

ДОДАТКИ

Додаток А

Інформаційна система аналізу та обробки статистичних даних інтернет магазину

Лістинги програми

Аркушів 7

Київ – 2023 року

Модуль analiz_app.py

```
import pandas as pd
import plotly.express as px
import plotly.graph_objs as go
import streamlit as st
from filter_by_year_and_quarter import *
from total_sales import *
from mean_customers_per_day import *
from median_order_value import *
from most_popular_product import *
from most_profitable_product import *
from customer_retention_rate import *
from order_value_variability import *
from rfm import *
from associate_rule import *

# Display title and text
st.title("Аналіз статистики діяльності інтернет-магазину за квартал")

# Read dataframe
df = pd.read_csv('./data.csv')

# Create a selection widget to choose the quarter
year = st.selectbox("Оберіть рік:", df["Year"].unique())
quarter = st.selectbox("Оберіть квартал:", df["Quarter"].unique())

# Filter the data for the selected quarter
filtered_df = filter_by_year_and_quarter(df, year, quarter)

st.subheader('Оберіть необхідний аналіз')
if st.checkbox('Аналіз основних показників діяльності підприємства'):

    # SUMMARIZE THE DATA
    st.metric(f"Квартал {quarter} Загальні продажі", f"${total_sales(filtered_df, year, quarter)}",
    delta=None, delta_color="normal", help=None, label_visibility="visible")

    # MEAN
    st.metric(f"Квартал {quarter} Середня кількість покупців за день",
    f"{mean_customers_per_day(filtered_df, year, quarter)}", delta=None, delta_color="normal",
    help=None, label_visibility="visible")
```

					ІС91.290БАК.004 Д5		
Змн.	Арк.	№ докум.	Підпис	Дата	Інформаційна система аналізу та обробки статистичних даних інтернет магазину Лістинг програми		
Розроб.		Шерстюк І. О.					
Перевір.		Дорогий Я. Ю.					
Н. Контр.							
Затверд.		Дорогий Я. Ю.					
					Лім.		
					Арк.		
					Акрушів		
					1		
					7		
					КПІ ім. Ігоря Сікорського		
					Група ІС-91		

```

# MEDIAN
st.metric(f"Квартал {quarter} Медіана вартості замовлення ", f"${median_order_value(filtered_df,
year, quarter)}", delta=None, delta_color="normal", help=None, label_visibility="visible")

# STANDARD DEVIATION
st.metric(f"Квартал {quarter} Стандартне відхилення вартості замовлення",
f"${order_value_variability(filtered_df, year, quarter)}", delta=None, delta_color="normal", help=None,
label_visibility="visible")

# the most popular product(mode) in the selected quarter
st.metric(f"Квартал {quarter} Найбільш популярний товар", f"${most_popular_product(filtered_df,
year, quarter)}", delta=None, delta_color="normal", help=None, label_visibility="visible")

# Calculate the most profitable product(mode) in the selected quarter
st.metric(f"Квартал {quarter} Найбільш вигідний товар", f"${most_profitable_product(filtered_df,
year, quarter)}", delta=None, delta_color="normal", help=None, label_visibility="visible")

# Display the customer retention rate for the selected quarter
st.metric(f"Квартал {quarter} Рівень утримання покупця", f"${customer_retention_rate(filtered_df,
year, quarter)}%", delta=None, delta_color="normal", help=None, label_visibility="visible")

# VISUALIZE THE DATA
# Create a scatterplot of unit prices for the selected quarter
st.subheader(f"Квартал {quarter} Розподіл ціни товарів")
fig = px.scatter(filtered_df, x="UnitPrice", color="UnitPrice")
st.plotly_chart(fig)

#Create a plot for day total price
st.subheader(f"Квартал {quarter} Динаміка щоденних покупок ")
ch1=filtered_df.groupby(['Month','Day']).agg({'TotalPrice': 'sum'})
ch1.reset_index(drop= True , inplace= True )
chart=ch1['TotalPrice']
fig1 = px.scatter(ch1)
st.line_chart(chart)

#Create a plot for day total price
st.subheader(f"Квартал {quarter} Розподіл продажів за країнами ")
filtered_df.groupby(["Country"]).count().reset_index()
fig2 = px.bar(filtered_df, x="Country",y='Quantity')
st.plotly_chart(fig2)

if st.checkbox('RFM-аналіз'):
    st.subheader('RFM-аналіз')
    table=rfm(df)
    st.write(table[['CustomerID', 'RFM_Score', 'Customer_segment']])
    fig3=px.pie(table,values=table.Customer_segment.value_counts(),

```

					IC91.290БАК.004 Д5	Арк.
						2
Змн.	Арк.	№ докум.	Підпис	Дата		

```

names=table.Customer_segment.value_counts().index)
st.plotly_chart(fig3)

if st.checkbox('Асоціації'):
    st.subheader('Асоціації')
    month = st.selectbox("Оберіть місяць:", filtered_df["Month"].unique())
    rules=associate_rule(filtered_df,month)
    st.write(rules[['antecedents', 'consequents', 'support', 'confidence', 'lift']])
    st.write("Rules identified: ", len(rules))
    fig4 = px.scatter_3d(rules, x='support', y='confidence', z='lift')
    st.plotly_chart(fig4)

```

Модуль total_sales.py

```

def total_sales(data: pd.DataFrame, year: int, quarter: int) -> float:
    """
    Calculate the total sales for a given year and quarter
    :param data: The dataframe containing the data
    :param year: The year to calculate the total sales for
    :param quarter: The quarter to calculate the total sales for
    :return: The total sales for the given year and quarter
    """

    # Filter the data by the given year and quarter
    data = filter_by_year_and_quarter(data, year, quarter)

    # Calculate the total sales
    return round(data['TotalPrice'].sum(),2)

```

Модуль customer_retention_rate.py

```

import pandas as pd
from filter_by_year_and_quarter import filter_by_year_and_quarter

def customer_retention_rate(data: pd.DataFrame, year: int, quarter: int) -> float:
    """
    Calculate the customer retention rate for a given year and quarter
    :param data: The dataframe containing the data
    :param year: The year to calculate the customer retention rate for
    :param quarter: The quarter to calculate the customer retention rate for
    :return: The customer retention rate for the given year and quarter
    """

    # Filter the data by the given year and quarter
    data = filter_by_year_and_quarter(data, year, quarter)

    # Calculate the number of unique customers
    num_customers = data['CustomerID'].nunique()

    # Calculate the number of repeat customers

```

					IC91.290БАК.004 Д5	Арк.
						3
Змн.	Арк.	№ докум.	Підпис	Дата		

```

num_repeat_customers = sum(data.groupby(['CustomerID']).count()['Day'] > 1)

# Calculate the customer retention rate
return round(num_repeat_customers/num_customers *100,2)

```

Модуль mean_customers_per_day.py

```

import pandas as pd
from filter_by_year_and_quarter import filter_by_year_and_quarter

# Write a function to calculate mean number of customers per day
def mean_customers_per_day(data: pd.DataFrame, year: int, quarter: int) -> float:
    """
    Calculate the mean number of customers per day for a given year and quarter
    :param data: The dataframe containing the data
    :param year: The year to calculate the mean number of customers per day for
    :param quarter: The quarter to calculate the mean number of customers per day for
    :return: The mean number of customers per day for the given year and quarter
    """

    # Filter the data by the given year and quarter
    data = filter_by_year_and_quarter(data, year, quarter)

    # Calculate the mean number of customers per day
    return round(data.groupby(['Month', 'Day']).nunique()['CustomerID'].mean(), 2)

```

Модуль most_popular_product.py

```

import pandas as pd
from filter_by_year_and_quarter import filter_by_year_and_quarter

# Write a function to find the most popular product for a given year and quarter
def most_popular_product(data: pd.DataFrame, year: int, quarter: int) -> str:
    """
    Find the most popular product for a given year and quarter
    :param data: The dataframe containing the data
    :param year: The year to find the most popular product for
    :param quarter: The quarter to find the most popular product for
    :return: The most popular product for the given year and quarter
    """

    # Filter the data by the given year and quarter
    data = filter_by_year_and_quarter(data, year, quarter)

    # Find the most popular product
    most_popular_stock_code = data['StockCode'].mode()[0]
    return data.loc[data['StockCode'] == most_popular_stock_code, 'Description'].values[0]

```

Модуль most_profitable_product.py

					IC91.290БАК.004 Д5	Арк.
						4
Змн.	Арк.	№ докум.	Підпис	Дата		

```

import pandas as pd
from filter_by_year_and_quarter import filter_by_year_and_quarter

# Write a function to find the most profitable product for a given year and quarter
def most_profitable_product(data: pd.DataFrame, year: int, quarter: int) -> str:
    """
    Find the most profitable product for a given year and quarter
    :param data: The dataframe containing the data
    :param year: The year to find the most profitable product for
    :param quarter: The quarter to find the most profitable product for
    :return: The most profitable product for the given year and quarter
    """

    # Filter the data by the given year and quarter
    data = filter_by_year_and_quarter(data, year, quarter)

    # Find the most profitable product
    total_price_by_stock_code = pd.DataFrame(data.groupby('StockCode')['TotalPrice'].sum())
    #print(total_price_by_stock_code.head())
    sorted_by_price_stock = total_price_by_stock_code.sort_values(by='TotalPrice', ascending=False)
    #print(sorted_by_price_stock.head())
    return data.loc[data['StockCode'] == sorted_by_price_stock.index[0], 'Description'].values[0]

```

Модуль order_value_variability.py

```

# Write a function to calculate the order value variability for a given year and quarter
def order_value_variability(data: pd.DataFrame, year: int, quarter: int) -> float:
    """
    Calculate the order value variability for a given year and quarter
    :param data: The dataframe containing the data
    :param year: The year to calculate the order value variability for
    :param quarter: The quarter to calculate the order value variability for
    :return: The order value variability for the given year and quarter
    """

    # Filter the data by the given year and quarter
    data = filter_by_year_and_quarter(data, year, quarter)

    # Group data by order number and calculate the order total
    total_order = data.groupby('InvoiceNo')['TotalPrice'].sum()

    # Calculate the order value variability
    return round(np.std(total_order), 2)

```

Модуль rfm.py

```

import pandas as pd
import datetime as dt
import numpy as np

```

					IC91.290БАК.004 Д5	Арк.
						5
Змн.	Арк.	№ докум.	Підпис	Дата		

```

def rfm(df: pd.DataFrame):
    """
    Calculate RFM-analysis
    param df: The dataframe containing the data
    :return: The dataframe with segmentation of customers

    """

    df['InvoiceDate'] = pd.to_datetime(df['InvoiceDate'])
    df_recency = df.groupby(by='CustomerID', as_index=False)['InvoiceDate'].max()
    df_recency.columns = ['CustomerID', 'LastPurchaseDate']
    recent_date = df_recency['LastPurchaseDate'].max()
    df_recency['Recency'] = df_recency['LastPurchaseDate'].apply(lambda x: (recent_date - x).days)

    frequency_df = df.drop_duplicates().groupby(
        by=['CustomerID'], as_index=False)['InvoiceNo'].count()
    frequency_df.columns = ['CustomerID', 'Frequency']

    df['Total'] = df['TotalPrice']
    monetary_df = df.groupby(by='CustomerID', as_index=False)['Total'].sum()
    monetary_df.columns = ['CustomerID', 'Monetary']

    rf_df = df_recency.merge(frequency_df, on='CustomerID')
    rfm_df = rf_df.merge(monetary_df, on='CustomerID').drop(columns='LastPurchaseDate')

    rfm_df['R_rank'] = rfm_df['Recency'].rank(ascending=False)
    rfm_df['F_rank'] = rfm_df['Frequency'].rank(ascending=True)
    rfm_df['M_rank'] = rfm_df['Monetary'].rank(ascending=True)

    # normalizing the rank of the customers
    rfm_df['R_rank_norm'] = (rfm_df['R_rank']/rfm_df['R_rank'].max())*100
    rfm_df['F_rank_norm'] = (rfm_df['F_rank']/rfm_df['F_rank'].max())*100
    rfm_df['M_rank_norm'] = (rfm_df['M_rank']/rfm_df['M_rank'].max())*100
    rfm_df.drop(columns=['R_rank', 'F_rank', 'M_rank'], inplace=True)

    rfm_df['RFM_Score'] = 0.15*rfm_df['R_rank_norm']+0.28 * \
        rfm_df['F_rank_norm']+0.57*rfm_df['M_rank_norm']
    rfm_df['RFM_Score'] *= 0.05
    rfm_df = rfm_df.round(2)
    rfm_df['Customer_segment'] = np.where(rfm_df['RFM_Score'] >
        4.5, "Top Customers",
        (np.where(
            rfm_df['RFM_Score'] > 4,
            "High value Customer",
            (np.where(
                rfm_df['RFM_Score'] > 3,
                "Medium Value Customer",
                np.where(rfm_df['RFM_Score'] > 1.6,

```

					IC91.290БАК.004 Д5	Арк.
						6
Змн.	Арк.	№ докум.	Підпис	Дата		

```

        'Low Value Customers', 'Lost Customers')))))))

return(rfm_df)

```

Модуль associate_rule.py

```

import pandas as pd
import numpy as np
import seaborn as sns
import re
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
from mlxtend.preprocessing import TransactionEncoder

def associate_rule(filtered_df: pd.DataFrame, month: int):
    """
    Calculate RFM-analysis
    param df: The dataframe containing the data
    :return: The dataframe with segmentation of customers

    """
    basket1 = filtered_df[(filtered_df['Month'] == month)]

    basket = basket1[['CustomerID', 'StockCode', 'InvoiceDate']]
    basket.StockCode = basket.StockCode.transform(lambda x: [x])
    basket = basket.groupby(['CustomerID', 'InvoiceDate']).sum()['StockCode'].reset_index(drop=True)

    encoder = TransactionEncoder()
    transactions = pd.DataFrame(encoder.fit(basket).transform(basket), columns=encoder.columns_)
    frequent_itemsets = apriori(transactions, min_support= 6/len(basket), use_colnames=True, max_len =
2)
    rules = association_rules(frequent_itemsets, metric="lift", min_threshold = 1.5)

    return rules

```

					IC91.290БАК.004 Д5	Арк.
						7
Змн.	Арк.	№ докум.	Підпис	Дата		