

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

Навчально-науковий інститут атомної та теплової енергетики

Кафедра цифрових технологій в енергетиці

«До захисту допущено»

Завідувач кафедри

_____ Наталія АУШЕВА

“ ” _____ 2025 р.

**Дипломна робота
на здобуття ступеня бакалавр**

За освітньою програмою “Цифрові технології в енергетиці”
Спеціальності 122 “Комп’ютерні науки”
на тему: “Веб-додаток аналізу транзакцій в закладі обслуговування”

Виконав: студентка 4 курсу, групи ТР-15

Чубукова Анастасія Олександрівна
(прізвище, ім’я, по батькові)

(підпис)

Керівник ст.викладач Ольга БЕСПАЛА
(посада, науковий ступінь, вчене звання, ім’я, ПРІЗВИЩЕ)

(підпис)

Рецензент зав.кафедри ТАЕ, д.т.н., проф. Ольга Чорноусенко
(посада, науковий ступінь, вчене звання, ім’я, ПРІЗВИЩЕ)

(підпис)

Н.контроль асистент Антон ПАСІЧНЮК
(посада, ім’я, ПРІЗВИЩЕ)

(підпис)

Засвідчую, що у цій дипломній роботі немає
запозичень з праць інших авторів без
відповідних посилань.

Студент _____

Київ - 2025

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО”**

**НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ АТОМНОЇ ТА ТЕПЛОВОЇ
ЕНЕРГЕТИКИ**

Кафедра ЦИФРОВИХ ТЕХНОЛОГІЙ В ЕНЕРГЕТИЦІ

Рівень вищої освіти – перший (бакалаврський)

спеціальність 122 “Комп’ютерні науки”

Освітньо-професійна програма “Цифрові технології в енергетиці”

ЗАТВЕРДЖУЮ

Завідувач кафедри ЦТЕ

Наталія АУШЕВА

(підпис)

“ ” 2025 р.

ЗАВДАННЯ

на дипломну роботу студенту

Чубуковій Анастасії Олександрівні

(прізвище, ім’я, по батькові)

1. Тема роботи “ Веб-додаток аналізу транзакцій в закладі обслуговування ”

Науковий керівник Беспала Ольга Миколаївна, ст.викладач

(прізвище, ім’я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від “ 02 ” червня 2025 року № 1875-с

2. Термін подання студентом роботи 09.06.2025

3. Вихідні дані до роботи мова програмування R, веб-фреймворк Shiny, середовище розробки RStudio, СУБД SQLite

4. Перелік питань, які потрібно розробити _____

1) провести аналіз існуючих аналогів

2) визначити та аргументувати методи сегментації, кластеризації та прогнозування

3) спроєктувати архітектуру програмного забезпечення

4) реалізувати інтерфейс користувача для різних ролей з функціональним розмежуванням доступу;

- 5) описати процес використання веб-застосунку, включно з прикладами роботи
5. Орієнтований перелік ілюстративного матеріалу архітектурна схема веб-застосунку, модель представлення даних, діаграма прецедентів, скріншоти інтерфейсу для кожного модуля.
6. Дата видачі завдання 13.09.2024

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1.	Вибір теми роботи	16.09.2024	Виконано
2.	Аналіз методів та засобів розв'язання задачі	14-20.04.2025	Виконано
3.	Розробка архітектури та загальної структури системи	21-27.04.2025	Виконано
4.	Розробка окремих підсистем	28.04-01.05.2025	Виконано
5.	Програмна реалізація системи	02-13.05.2025	Виконано
6.	Оформлення пояснювальної записки	14-27.05.2025	Виконано
7.	Захист програмного забезпечення	14.05.2025	Виконано
8.	Передзахист	27.05.2025	Виконано
9.	Захист	16-20.06.2025	Виконано

Студент

(підпис)

Анастасія ЧУБУКОВА
(прізвище та ініціали)

Керівник

(підпис)

Ольга БЕСПАЛА
(прізвище та ініціали)

АНОТАЦІЯ

Дипломна робота виконана на 58 сторінках, містить 29 ілюстрацій, 2 таблиці, 2 додатки, 25 джерел у переліку посилань.

Мета роботи – розробити веб-додаток для аналізу транзакцій клієнтів у книгарні-кав'ярні з використанням методів сегментації, кластеризації, оцінки ефективності акцій та прогнозування ймовірності відтоку.

Методи та засоби: мова програмування R, фреймворк Shiny, бібліотеки ggplot2, plotly, randomForest, cluster, dplyr, caret, shinyWidgets; база даних SQLite; середовище розробки RStudio.

Результат – програмна система, яка дозволяє проводити аналіз клієнтської поведінки та візуалізувати ключові бізнес-показники.

Ключові слова: АНАЛІТИКА КЛІЄНТІВ, ВІДТІК КЛІЄНТІВ, RFM, КЛАСТЕРИЗАЦІЯ, ТРАНЗАКЦІЇ, R SHINY.

ABSTRACT

The thesis consists of 58 pages, includes 29 illustrations, 2 tables, 2 appendices, and 25 references in the bibliography.

The aim of the work is to develop an web application for analyzing customer transactions in a bookstore café using methods of segmentation, clustering, campaign effectiveness evaluation, and churn prediction.

Methods and tools: R programming language, Shiny framework, libraries ggplot2, plotly, randomForest, cluster, dplyr, caret, shinyWidgets; SQLite database; RStudio development environment.

Result – a software system that enables multilevel analysis of customer behavior, visualization of key business indicators.

Keywords: CUSTOMER ANALYTICS, CUSTOMER CHURN, RFM, CLUSTERING, TRANSACTIONS, R SHINY.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ.....	8
ВСТУП.....	9
1 ПОСТАНОВКА ЗАДАЧІ АНАЛІЗУ ТРАНЗАКЦІЙ ЗАКЛАДУ ОБСЛУГОВУВАННЯ.....	11
1.1 Задача аналізу транзакцій у закладах обслуговування.....	11
1.1.1 Постановка задачі.....	12
1.1.2 Користувачі системи.....	13
1.2 Огляд програмних засобів для аналізу транзакцій.....	14
1.2.1 ВІ-платформа для бізнес-аналітики.....	14
1.2.2 Комерційні сервіси.....	15
1.2.3 Open-source рішення для поведінкової аналітики.....	16
1.2.4 Вибір підходу до реалізації системи.....	17
1.3 Технології для реалізації веб-додатку.....	18
2 МЕТОДИ АНАЛІЗУ ТРАНЗАКЦІЙНИХ ДАНИХ.....	20
2.1 Метод RFM-сегментації клієнтів.....	20
2.2 Метод кластеризації K-means.....	21
2.3 Метод дерева рішень.....	24
2.4 Прогнозування відтоку клієнтів методом Random Forest.....	26
2.5 Оцінка ефективності акцій за допомогою ключових показників.....	30
3 ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ.....	32
3.1 Архітектура програмної системи.....	32
3.2 Використані технології та середовище розробки.....	34
3.2.1 Мова програмування та середовище розробки.....	34

	7
3.2.2 Веб-фреймворк для інтерфейсу	35
3.2.3 Пакети для аналізу та візуалізації.....	35
3.3 База даних та структура таблиць	36
3.4 Діаграма прецедентів	39
4 РОБОТА КОРИСТУВАЧА З ПРОГРАМНОЮ СИСТЕМОЮ.....	41
4.1 Системні вимоги	41
4.2 Вхід у систему та опис головної сторінки	42
4.3 Інтерфейс адміністратора системи	43
4.4 Інтерфейс аналітика	44
4.4.1 Вкладка «Аналіз акцій»	45
4.4.2 Кластеризація клієнтів	47
4.4.3 Кластеризація товарів	50
4.4.4 Прогнозування відтоку	53
4.5 Інтерфейс власника	54
ВИСНОВКИ.....	57
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	59
ДОДАТОК А.....	62
ДОДАТОК Б	66

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

RFM – модель сегментації клієнтів, що базується на трьох показниках: Recency (давність останньої покупки), Frequency (частота покупок), Monetary (грошова вартість покупок).

K-means – алгоритм кластеризації, який поділяє об'єкти на задану кількість груп (кластерів) на основі схожості.

Random Forest – ансамблевий алгоритм машинного навчання, який використовує множину дерев рішень для класифікації або регресії.

R Shiny – фреймворк для створення інтерактивних веб-додатків на мові програмування R.

UI (User Interface) – інтерфейс користувача, графічне або текстове середовище для взаємодії з програмою.

CRUD – скорочення від Create, Read, Update, Delete; базові операції над даними у базі.

SQL (Structured Query Language) – мова структурованих запитів, що використовується для взаємодії з реляційними базами даних.

SQLite – легка вбудована реляційна база даних, яка не потребує окремого серверного розгортання.

CRM (Customer Relationship Management) – система управління взаємовідносинами з клієнтами.

ROC-крива (Receiver Operating Characteristic) – графічне представлення діагностичної здатності бінарного класифікатора.

KPI (Key Performance Indicators) – ключові показники ефективності, що використовуються для оцінки результативності бізнес-процесів.

PDP (Partial Dependence Plot) – це візуалізація, яка показує середнє значення передбачення моделі при фіксованому значенні однієї або двох ознак, а всі інші ознаки змінюються в межах своїх значень у навчальній вибірці.

ВСТУП

Із розвитком інформаційних технологій дедалі більше бізнесів використовують інструменти аналітики даних для покращення операційної ефективності та ухвалення рішень. У сфері роздрібної торгівлі та обслуговування транзакційна аналітика вже стала стандартною практикою – зокрема в супермаркетах, торгових мережах і цифрових платформах. Проте специфіка у закладах зі змішаним сервісним форматом, як-от книгарні-кав'ярні, залишається малодослідженою в контексті глибинного аналізу поведінки клієнтів.

Протягом останніх кількох років в Україні спостерігається стрімкий розвиток галузі книговидання та книготоргівлі, що спричинило відкриття великої кількості нових книгарень [1]. Це свідчить про зростаючий інтерес українців до літератури та культурних подій, який стимулює розвиток мережі закладів, що продають книжки. Традиційні книгарні все частіше перетворюються на сучасні книгарні-кав'ярні, де покупці можуть не лише придбати книги, а й насолодитися чашкою якісної кави в затишній атмосфері, поспілкуватися чи провести час у приємній атмосфері [2]. Такий формат значно розширює аудиторію закладів, залучаючи не лише постійних читачів, а й студентів, молодь, туристів і випадкових відвідувачів. Це зумовлює появу нових сценаріїв поведінки споживачів, які не завжди можна оцінити стандартними методами. Важливою складовою бізнес-моделі стають саме кавові напої, які нерідко забезпечують вагомую частку виручки.

Акції у книгарнях-кав'ярнях можуть бути як класичними знижками, так і тематичними пропозиціями, бонусами або подарунковими акціями. Вивчення їх впливу на виручку, кількість покупців, середній чек і поведінку окремих клієнтських сегментів відкриває можливості для кращого планування та управління такими заходами. При цьому, як і в багатьох невеликих форматах роздрібної торгівлі, ці ефекти часто не аналізуються системно, а рішення щодо акцій приймаються інтуїтивно.

Використання адаптованих систем для аналізу транзакцій дозволяє перейти від інтуїтивного підходу до обґрунтованих даними рішень. Інтерактивна система

аналізу даних допомагає закладам краще розуміти своїх клієнтів, коригувати маркетингові стратегії та підвищувати ефективність управління на основі реальних показників.

Програмний продукт, створений у межах цієї роботи, дозволяє на практиці реалізувати такі підходи. Його архітектура поєднує інструменти поведінкової аналітики, сегментації клієнтів та оцінки ефективності акцій у зручному для користувача інтерфейсі. Система спеціально адаптована під потреби невеликих закладів, яким важливо ухвалювати рішення не лише інтуїтивно, а й на основі достовірних даних і аналітичних висновків.

Дипломна записка містить чотири розділи. У першому розділі наведено постановку прикладної задачі, обґрунтовано актуальність теми, визначено мету, задачі та цільову аудиторію розробленого програмного продукту, а також здійснено огляд існуючих рішень і програмних засобів, що можуть бути використані для аналізу транзакцій. У другому розділі описано методи, застосовані в системі: RFM-сегментацію, кластеризацію, прогнозування відтоку клієнтів і аналіз ефективності акцій на основі KPI. Третій розділ присвячено архітектурі програмної системи, моделі представлення даних, інтерфейсу, реалізації ключових модулів, а також діаграмам, що відображають логіку роботи. У четвертому розділі продемонстровано сценарії використання веб-застосунку та представлено результати його роботи на практичних даних.

Під час написання дипломної роботи було прийнято участь у конференції «Сучасні проблеми наукового забезпечення енергетики» з публікацією тез на тему «Аналіз транзакцій для оцінки впливу поведінкових факторів на продажі» [3] (Додаток Б).

1 ПОСТАНОВКА ЗАДАЧІ АНАЛІЗУ ТРАНЗАКЦІЙ ЗАКЛАДУ ОБСЛУГОВУВАННЯ

У даному розділі описано прикладну задачу, яка стала основою для реалізації веб-додатку з аналізу транзакцій у форматі книгарні-кав'ярні. Зважаючи на актуальність використання даних про клієнтську поведінку та потребу в оцінці ефективності промоційних заходів, розробка програмного інструменту для аналітики у сфері ритейлу має практичне значення для малого бізнесу. Перший підрозділ містить постановку задачі, в якій викладено передумови, основні вимоги до функціональності системи, а також описані ключові задачі, що розв'язуються в рамках проєкту. Також увагу присвячено огляду аналогів систем для аналізу транзакцій. Описано програмні засоби, які були обрані для реалізації веб-додатку: мова програмування, бібліотеки аналізу даних і візуалізації, платформа розробки та обґрунтування їх вибору.

1.1 Задача аналізу транзакцій у закладах обслуговування

В умовах зростаючої конкуренції у сфері закладів обслуговування [4] все більш важливого значення набуває ефективне використання даних про клієнтську поведінку. Особливо це стосується гібридних форматів закладів, на зразок книгарень-кав'ярень, які не лише надають покупцям змогу придбати книгу, а й створюють приємну атмосферу для проведення часу, підвищуючи ймовірність додаткових покупок. Саме тому для такого бізнесу стає все важливішим глибше розуміти своїх клієнтів – не просто знати, скільки вони витрачають, а як часто повертаються, що поєднують у покупках, як реагують на акції, які товари «працюють» разом.

У невеликих закладах аналітична підтримка часто потребує інструментів, які поєднують просте використання з можливістю глибокого аналізу даних – без необхідності у складних CRM або BI-системах. Вони мають багатий масив даних

про транзакції - чеки, час покупок, категорії товарів, дані клієнтів за програмою лояльності. Це створює потенціал для побудови доступних і зручних аналітичних інструментів, які допоможуть ухвалювати обґрунтовані рішення – наприклад, коли і як варто проводити кавові акції, які знижки стимулюють не разові покупки, а повернення, як змінюється структура чека після запуску промо.

У цьому контексті виникає задача побудови веб-застосунку, який дозволить власникам книгарні-кав'ярні проводити аналіз даних у зручному форматі. Такий застосунок має надавати інтерфейс для огляду основних бізнес-показників, сегментування клієнтів за частотою та обсягами покупок, моделювання впливу кавових акцій на фінансові результати. Він має допомагати відповідати на прості, але стратегічно важливі питання: «чи варто запускати цю акцію ще раз?», «як змінився склад клієнтів після акції?», «чи вигідна знижка на каву, якщо її пов'язати з покупкою книги?»

1.1.1 Постановка задачі

Система повинна охоплювати можливість фільтрації та агрегації даних за будь-який обраний період, розрахунок ключових бізнес-показників – таких як загальна виручка, середній чек, частка кави у виручці, кількість унікальних клієнтів. Крім того, доцільно реалізувати модулі моделювання ефекту впроваджених акцій за допомогою методів статистичного аналізу, зокрема лінійної регресії, а також побудову сегментаційних моделей на основі підходу RFM.

Отже, головна задача дипломної роботи полягає у створенні веб-системи, яка дає змогу здійснювати аналіз транзакцій у книгарні-кав'ярні, оцінювати вплив маркетингових дій на фінансові результати, а також формувати практичні рекомендації для підвищення загальної виручки закладу. Це включає в себе такі задачі:

- підготовка та попереднє структурування транзакційних даних, включаючи об'єднання інформації про клієнтів, чеки, товари та категорії, а також очищення масиву від дублікативних або некоректних записів;
- розрахунок ключових бізнес-показників: загальної виручки, середнього чека, кількості унікальних клієнтів, частки окремих категорій у структурі продажів;
- реалізація RFM-сегментації клієнтів (Recency, Frequency, Monetary) для виявлення стабільних, неактивних, нових або потенційно цінних клієнтів;
- побудова моделі прогнозування ймовірності відтоку клієнтів за допомогою алгоритму random forest, що дозволяє виявити ризикові групи;
- моделювання впливу умовних акцій на фінансові показники та поведінку клієнтів з використанням статистичних методів, зокрема регресійного аналізу;
- створення зручного веб-інтерфейсу для перегляду аналітичних даних, у якому реалізовано інтерактивні фільтри, графіки, табличні виводи та можливість налаштування параметрів користувачем.

Ці задачі було реалізовано у вигляді веб-застосунку на платформі R Shiny, що поєднує в собі функціональність збору, обробки, візуалізації та інтерпретації даних. Система орієнтована на потреби малого сервісного бізнесу з багатокомпонентною клієнтською взаємодією та дозволяє підвищити якість управлінських рішень на основі даних.

1.1.2 Користувачі системи

Розроблений веб-застосунок орієнтований на потреби кількох типів користувачів, кожен з яких взаємодіє із системою відповідно до своєї ролі. Власник закладу отримує доступ до зведених показників і узагальненої аналітики, зокрема до дашборду ключових метрик, огляду динаміки виручки, кількості клієнтів та структури продажів, а також можливість експорту звітів. Адміністратор системи відповідає за технічну підтримку даних: завантаження та оновлення транзакцій з бази даних, перегляд журналу змін (аудиту), а також CRUD-операції з довідниками

– клієнтів, товарів і категорій. Аналітик є ключовим користувачем розширених модулів системи: він аналізує ефективність акцій, досліджує зміни поведінки клієнтів, проводить кластеризацію товарів і споживачів, а також використовує модель прогнозування відтоку для виявлення ризикових груп.

1.2 Огляд програмних засобів для аналізу транзакцій

Оскільки аналітика транзакцій є важливим інструментом для вивчення поведінки клієнтів, оцінки ефективності маркетингових заходів і формування управлінських рішень у сфері роздрібної торгівлі, протягом останніх років з'явилося багато потужних комерційних та open-source рішень, які дозволяють компаніям аналізувати великі обсяги даних у режимі реального часу, будувати прогнозні моделі та візуалізувати результати в інтерактивному форматі. У цьому розділі розглянемо приклади сучасних аналітичних систем, що частково або повністю покривають задачі, подібні до тих, що реалізовані в межах даної дипломної роботи.

1.2.1 BI-платформа для бізнес-аналітики

Одними з найпоширеніших рішень для візуалізації й аналізу бізнес-даних є системи класу BI (Business Intelligence), зокрема Tableau та Power BI [5]. Tableau це інструмент для створення дашбордів і візуалізацій, який надає широкі можливості інтеграції даних з різних джерел, зокрема баз даних SQL, електронних таблиць, CRM-систем.

Power BI (рисунок 1.1) від Microsoft має подібну функціональність, але тісніше інтегрується з іншими продуктами Microsoft, такими як Excel, Azure чи SharePoint. Обидві платформи підтримують фільтри, сегментацію, візуальну агрегацію показників і дозволяють формувати персоналізовані аналітичні панелі.

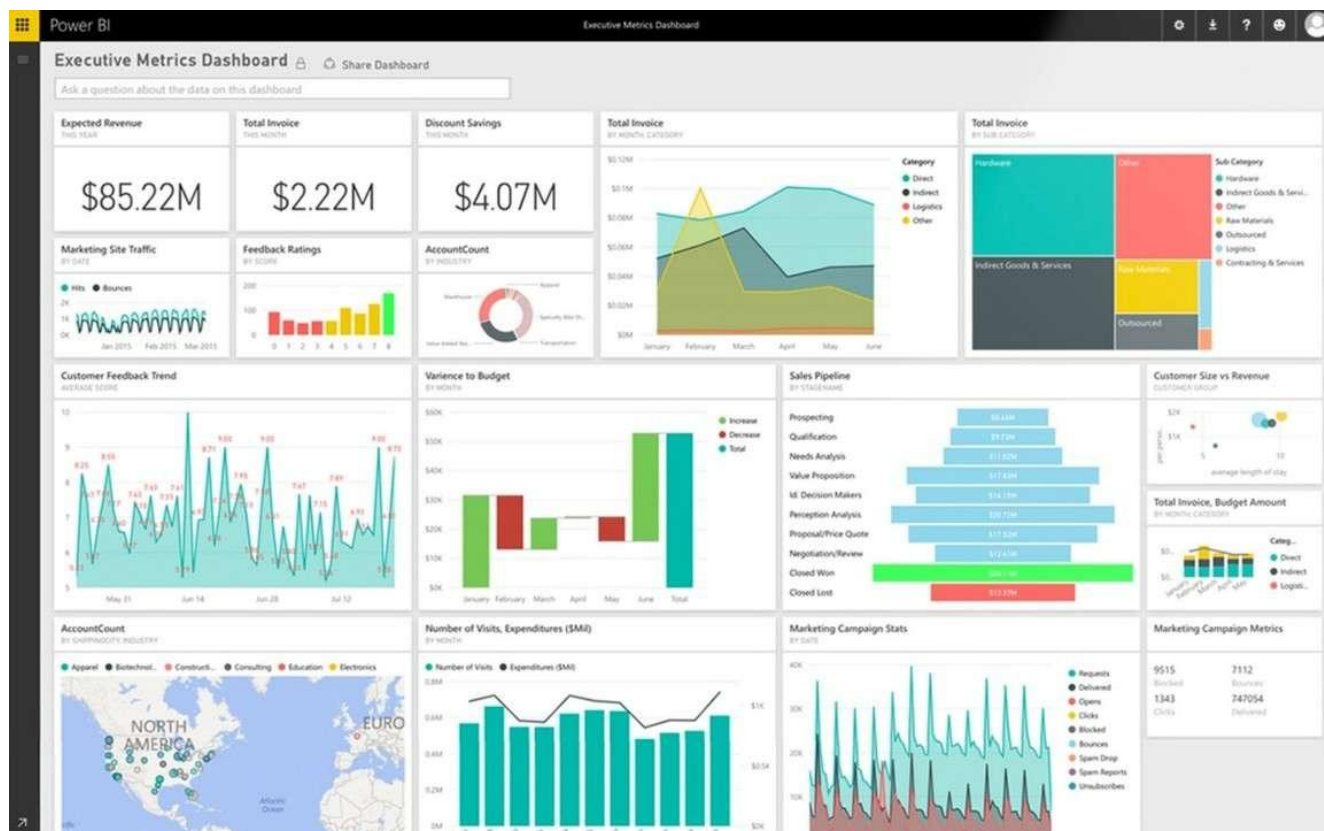


Рисунок 1.1 – Система Power BI

Ці системи активно використовуються у великих компаніях і дозволяють працювати з великими обсягами даних у хмарному середовищі. Такі системи мають дуже широкий спектр функцій, але орієнтовані передусім на корпоративних користувачів і потребують значних ресурсів для налаштування [6].

Для малих закладів, які не мають окремих аналітичних відділів або складної інфраструктури, впровадження подібних рішень може бути дуже затратним як фінансово, так і організаційно.

1.2.2 Комерційні сервіси

Іншим прикладом є сервіси, що спеціалізуються на поведінковій аналітиці користувачів, зокрема Kissmetrics та Mixpanel (рисунок 1.2). Вони надають можливість відстеження поведінки клієнтів у цифровому середовищі – на сайтах, у

застосунках, під час підписки чи купівлі [7]. Дані агрегуються в реальному часі, що дозволяє формувати сегменти, будувати воронки та проводити А/В-тести.

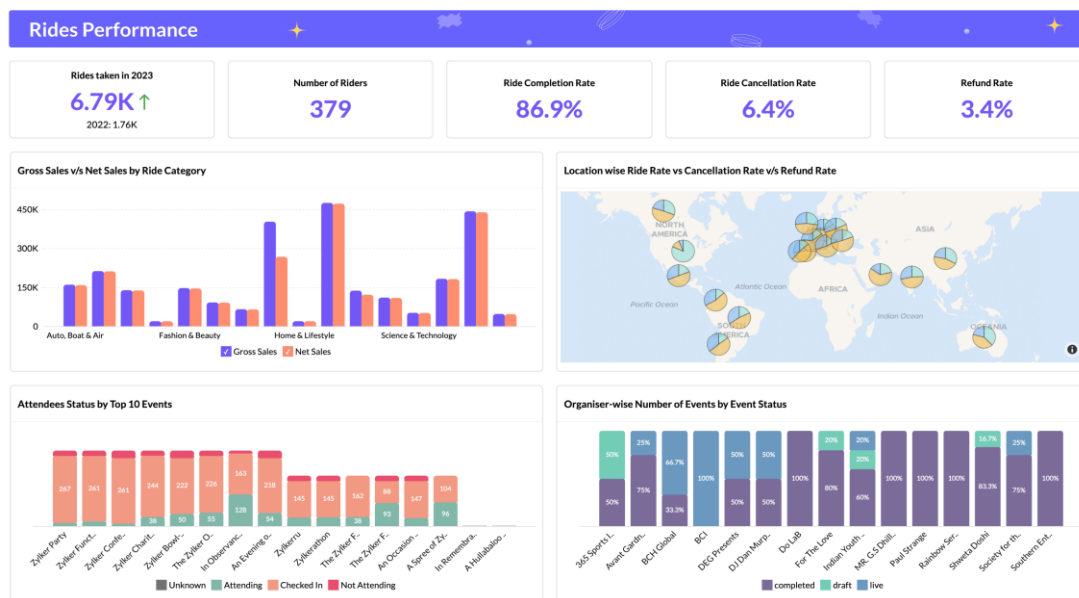


Рисунок 1.2 – Сервіс Mixpanel

Проте функціональність таких платформ переважно розрахована на онлайн-бізнеси та SaaS-сервіси. Їх застосування в офлайн-торгівлі, зокрема у книгарнях або змішаних форматах, потребує складної інтеграції через зовнішні сенсори, API, або залучення стороннього ПЗ.

1.2.3 Open-source рішення для поведінкової аналітики

Серед відкритих рішень особливу увагу варто приділити проєктам, створеним за допомогою мов програмування R та Python. У репозиторіях GitHub доступні багато прикладів дашбордів на базі Shiny (для R) або Streamlit (для Python), які реалізують RFM-аналіз, кластеризацію, прогнозування відтоку та інші елементи поведінкової аналітики. Вони мають відкритий код, який можна адаптувати під власні задачі, та не вимагають ліцензійної оплати [8].

Найчастіше ці системи сфокусовані на якійсь вузькій темі: аналіз електронної комерції, оцінка клієнтів банку, вивчення активності користувачів веб-сервісів. Через це їхнім недоліком є відсутність підходу до роздрібного бізнесу з невеликою кількістю клієнтів і поєднанням товарів і послуг. Крім того, більшість готових рішень потребують адаптації – перекладу, зміни структури даних, інтерфейсу.

У таблиці 1.1 узагальнено ключові характеристики трьох основних типів рішень, що використовуються для поведінкової аналітики в ритейлі. Порівняння охоплює призначення, цільовий бізнес, адаптивність до змішаних форматів, вартість, потребу в адаптації та рівень контролю.

Таблиця 1.1 – Порівняння існуючих аналогічних систем

Ознака	ВІ-платформи	Комерційні сервіси	Open-source рішення
Призначення	Бізнес-аналітика, візуалізація	Онлайн-аналітика дій користувачів	Гнучка поведінкова аналітика
Цільовий бізнес	Великий / корпоративний	SaaS, e-commerce	Малий / середній бізнес
Придатність для книгарні	Складні та затратні для локального використання	Не адаптовані до офлайн-торгівлі	Можна адаптувати до змішаного формату
Ліцензії та вартість	Висока вартість ліцензії, обмежені безкоштовні версії	Платна підписка, залежить від обсягу даних і функцій	Безкоштовно (open source), адаптація залежить від ресурсів
Потреба в адаптації	Висока	Висока	Середня
Гнучкість і контроль	Обмежена (залежність від платформи)	Обмежена (SaaS-модель)	Висока (повний контроль над кодом)

1.2.4 Вибір підходу до реалізації системи

Попри різноманіття доступних рішень, жодне з них не повною мірою відповідає потребам малих закладів культурно-комерційного формату. Стандартні

ВІ-системи є надто громіздкими, поведінкова аналітика орієнтована на онлайн, а open-source приклади потребують значних доробок. У контексті дипломної роботи постала задача створити просту у використанні, інтерактивну систему, яка працює з наявною локальною базою даних, не вимагає розгортання серверів, має відкриту архітектуру та включає спеціалізовані модулі саме для аналізу чеків, клієнтів і акцій у книгарні-кав'ярні.

Розроблена система дозволяє здійснювати сегментацію клієнтів, аналіз ефективності акцій і прогнозування відтоку без потреби в зовнішніх сервісах. Її перевагою є адаптація до реалій українського малого бізнесу та можливість подальшого масштабування функціоналу. У наступних розділах буде детально описано структуру системи, принципи її роботи та практичні результати, отримані в процесі тестування.

1.3 Технології для реалізації веб-додатку

Для реалізації веб-застосунку, призначеного для обробки й візуалізації транзакційних даних книгарні-кав'ярні, було обрано низку програмних засобів, які забезпечують як ефективну роботу з даними, так і створення зручного користувацького інтерфейсу. Вибір інструментів був зумовлений необхідністю гнучко працювати з таблицями транзакцій, виконувати сегментаційний аналіз, моделювати поведінкові показники, а також інтерактивно представляти результати для кінцевих користувачів – власника бізнесу та аналітика.

Основним середовищем реалізації системи стала мова програмування R, яка забезпечує потужний набір інструментів для статистичного аналізу, роботи з даними, візуалізації та побудови інтерактивних застосунків. Однією з ключових переваг мови R є її відкритість, велика кількість спеціалізованих бібліотек, орієнтованих на обробку даних, а також активна спільнота користувачів, що полегшує підтримку та розвиток проєкту.

Для створення веб-інтерфейсу було використано фреймворк Shiny, який забезпечує двосторонній зв'язок між користувачем і серверною логікою, що дає змогу в режимі реального часу змінювати параметри аналізу, фільтрувати дані, переглядати графіки та таблиці залежно від введених умов.

Для візуалізації результатів було використано пакет ggplot2, який забезпечує гнучке й естетично привабливе створення графіків. Його інтеграція зі Shiny дає змогу будувати графіки, які оновлюються динамічно під час взаємодії з користувачем. Для відображення числових показників, таблиць і дашбордів застосовувалися пакети DT (таблиці з фільтрами та сортуванням), shinydashboard (компонування сторінки у вигляді панелей із KPI) та shinyWidgets (розширені елементи управління, зокрема перемикачі, фільтри, вибір дат тощо).

Для моделювання сегментації клієнтів було використано реалізацію RFM-аналізу, виконану за допомогою пакетів dplyr і lubridate для обробки дат і угруповань. Для побудови моделей прогнозування ймовірності відтоку клієнтів застосовувався алгоритм Random Forest, реалізований через пакет randomForest.

2 МЕТОДИ АНАЛІЗУ ТРАНЗАКЦІЙНИХ ДАНИХ

У даному розділі проаналізовано методи, які використовувались у програмному продукті. Зокрема, розглянуто методи RFM-сегментації, кластеризація методом K-means, метод дерева рішень та метод Random Forest.

2.1 Метод RFM-сегментації клієнтів

У межах аналізу активності клієнтів спираючись на дані транзакцій, важливо не лише фіксувати кількість покупок чи суму витрат, а й розуміти поведінкові особливості кожного покупця. Одним із класичних, і водночас ефективних підходів до такого аналізу є RFM-сегментація – метод, який дозволяє здійснити поведінкову класифікацію клієнтів на основі трьох ключових параметрів: Recency (давність останньої покупки), Frequency (частота покупок) та Monetary (загальна сума витрат).

Цей підхід широко застосовується у маркетингу, CRM та аналітиці клієнтської бази. Його перевага полягає в простоті обчислень, прозорості інтерпретації та високій практичній цінності для прийняття управлінських рішень. На відміну від складніших алгоритмів машинного навчання, RFM не потребує попереднього навчання моделі, має зрозумілу бізнес-логіку та забезпечує швидкий результат навіть на невеликому масиві даних [9].

Суть методу полягає в тому, що кожен клієнт отримує три числові оцінки:

- 1) параметр Recency – час (у днях, тижнях або місяцях) від останньої покупки до дати аналізу;
- 2) параметр Frequency – кількість покупок за певний період;
- 3) параметр Monetary – загальна сума, витрачена клієнтом у торговій точці.

Ці значення можуть бути нормалізовані або віднесені до певних категорій (рангів). Найпоширенішим є поділ кожної змінної на п'ять квантилів (рангів від 1 до 5), де, наприклад, клієнти з найменшою давністю покупки (тобто найбільш

«свіжі») отримують $Recency = 5$, а ті, хто давно не купував — $Recency = 1$. Для $Frequency$ та $Monetary$ ранги зростають зі збільшенням кількості покупок та витрат відповідно.

Таким чином, для кожного клієнта формується вектор категорій, що дозволяє побудувати багатовимірну матрицю клієнтів і групувати їх у сегменти. Наприклад, комбінація 5-5-5 відповідає найбільш цінним клієнтам, які нещодавно здійснили багато дорогих покупок, тоді як сегмент 1-1-1 охоплює неактивних і малоприбуткових клієнтів.

У межах цієї роботи було обрано саме кватильний підхід до шкалювання показників, що дозволило здійснити адаптивну сегментацію відповідно до фактичного розподілу значень у вибірці. На відміну від жорстких порогових значень, такий підхід гарантує рівномірний розподіл клієнтів по сегментах навіть за умов змінної структури даних.

RFM-сегментація демонструє низку переваг: інтерпретованість, простоту реалізації, гнучкість у налаштуванні та можливість використання як в автономному режимі, так і як попередній етап для подальших аналітичних алгоритмів, таких як кластеризація або прогнозування відтоку. Саме ці якості роблять метод особливо корисним у малому та середньому бізнесі, де відсутні складні CRM-системи [10].

Отже, RFM-сегментація є оптимальним підходом для початкового поділу клієнтів і побудови поведінкових профілів. Її результати можна надалі доповнювати глибшими моделями або використовувати як основу для автоматизованих рішень (наприклад, запуску тригерних кампаній, оцінки життєвого циклу клієнта, виявлення груп із підвищеним ризиком відтоку).

2.2 Метод кластеризації K-means

Після базової сегментації клієнтів за допомогою RFM-моделі доцільно застосувати більш гнучкі методи виявлення закономірностей у поведінці, які не передбачені наперед визначеними порогами чи фіксованими правилами. Одним із

найбільш поширених та ефективних алгоритмів для цього завдання є кластеризація методом K-means.

Кластеризація – це метод неконтрольованого навчання, який дозволяє згрупувати об’єкти (у даному випадку клієнтів або товари) за схожістю їх характеристик, без попереднього знання міток або класів. Метод K-means ґрунтується на принципі мінімізації внутрішньокластерної дисперсії: кожен об’єкт належить до того кластера, центр якого є для нього найближчим у багатовимірному просторі ознак [11]. Математично це можна виразити так:

$$\min_{C_1, \dots, C_k} \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2,$$

де C_j – множина точок, що належать кластеру j

μ_j – центр цього кластера (середнє значення ознак усіх елементів у кластері),

$\|x_i - \mu_j\|^2$ – квадрат евклідової відстані між об’єктом і центром.

У рамках даного проєкту кластеризація використовувалася у двох напрямках: для групування клієнтів за поведінковими параметрами (RFM-метрики, середній чек, кількість візитів) та для кластеризації товарів за показниками продажів, популярності, середньої ціни та впливу на структуру чека.

Алгоритм K-means був обраний з-поміж інших методів кластеризації завдяки простоті реалізації, високій швидкодії та можливості масштабування на великі вибірки. У порівнянні з ієрархічною кластеризацією, яка є обчислювально затратною і менш гнучкою у випадках із численними спостереженнями, K-means дозволяє задавати бажану кількість кластерів і оперативно отримувати результати навіть на великих наборах даних [12].

Процедура реалізації кластеризації складалась з кількох етапів:

– вибір змінних для кластеризації – включено такі показники, як Recency, Frequency, Monetary (із попереднього RFM-аналізу), а також похідні: середній чек, кількість товарів у покупці, кількість унікальних категорій.

– попередня обробка даних – нормалізація значень (z-score або мін-макс трансформація) для забезпечення однакового масштабу ознак, що важливо для алгоритму K-means.

– вибір кількості кластерів (k) – здійснювався за допомогою методу “лікоть” (elbow method), при якому будується графік зміни середньої внутрішньокластерної варіації в залежності від кількості кластерів, а точка згину графіка обирається як оптимальне значення k .

– запуск алгоритму – початкові центроїди обирались випадково (із подальшою ініціалізацією методом k -means++ для покращення стабільності результатів), після чого ітераційно перераховувались центри до стабілізації кластерів.

У результаті кожен клієнт або товар отримував мітку відповідного кластера. Кількість кластерів зазвичай варіювалася від 3 до 5 – цього достатньо для виявлення ключових груп з різною поведінкою.

Переваги застосування K-means:

– гнучкість у побудові різних варіантів кластеризації без зміни структури системи;

– простота інтерпретації – результати виводяться у вигляді таблиць із кластерами, графіків і дашбордів, зрозумілих навіть нефаховому користувачу;

– інтеграція з іншими модулями – кластери використовуються у звітах, при побудові моделей відтоку, а також у плануванні промоцій.

Варто зазначити, що попри загальну ефективність, метод K-means має й певні обмеження:

– потребує попереднього визначення кількості кластерів;

– чутливий до початкових центрів (хоча це частково вирішується методом k -means++);

– не гарантує виявлення кластерів у разі складної геометрії розподілу.

Тим не менш, з урахуванням задач і обсягів даних у даному проєкті, кластеризація методом K-means виявилася оптимальним і практично ефективним рішенням, що дозволило реалізувати додатковий рівень аналітики без ускладнення системи.

2.3 Метод дерева рішень

У процесі розробки системи аналізу транзакцій важливою задачею стало не лише виявлення сегментів клієнтів, але й пояснення факторів, які впливають на їхню поведінку, зокрема активність або ймовірність відтоку. Для цього був використаний інтерпретований метод машинного навчання – дерево рішень, що дозволяє будувати прозорі та логічно зрозумілі моделі на основі реальних ознак клієнтів.

Дерева рішень належать до алгоритмів контролюваного навчання й використовуються для класифікації або регресії, залежно від характеру вихідної змінної [13]. З математичної точки зору, дерево рішень – це модель, що рекурсивно розбиває простір ознак на підмножини таким чином, щоб максимізувати певний критерій чистоти у кожному вузлі. Для задач класифікації найчастіше використовують ентропію або індекс Джині.

Формула ентропії наступна:

$$E(S) = \sum_{i=1}^C -p_i \times \log_2 p_i,$$

де S – поточна вибірка прикладів,

C – кількість класів,

p_i – частка об'єктів класу.

Індекс Джині математично можна виразити так:

$$Gini = 1 - \sum_{i=1}^C (p_i)^2,$$

де C – кількість класів,

p_i – частка об'єктів класу.

Кожне розбиття вибирається так, щоб зменшити невизначеність між вузлом-батьком і його дочірніми вузлами. Алгоритм формує структуру у вигляді дерева, де кожна внутрішня вершина – це перевірка умови на певну ознаку, а кожен лист – це прогноз.

У цій роботі метод застосовувався для класифікації клієнтів за різними критеріями: активний/пасивний, схильний до повернення/втрати, реактивний/байдужий до акцій. Основна ідея алгоритму полягає в послідовному

розбитті вибірки на підгрупи за найбільш значущими предикторами, що дозволяє сформувати гілки логіки поведінки, які легко інтерпретуються як умовні правила (рисунок 2.1).

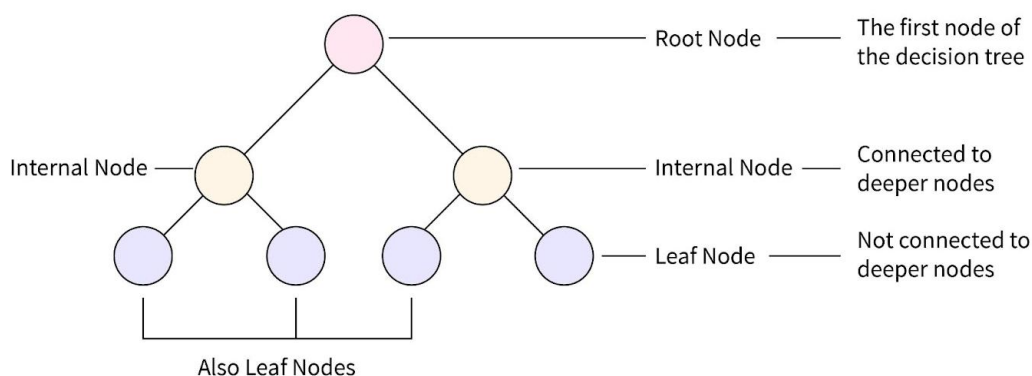


Рисунок 2.1 – Схема роботи алгоритму Дерево рішень

У межах реалізованої системи дерево рішень будувалося на основі попередньо зібраних клієнтських ознак: давність останньої покупки, частота відвідувань, середній чек, частка покупок у вихідні, участь в акціях, кількість категорій товарів, що купувалися. Також враховувалися змінні з попередньої кластеризації та RFM-сегментації. Цільова змінна при цьому залежала від конкретного сценарію: в одному випадку – це була ймовірність того, що клієнт зменшить частоту візитів, в іншому – ймовірність реакції на акцію.

На відміну від “чорних скриньок”, таких як нейронні мережі чи ансамблеві моделі, дерево рішень дозволяє побачити повну структуру ухвалення рішень – тобто, які ознаки є ключовими у прогнозі, яка їхня послідовність і як змінюється результат за різних значень параметрів [14]. Це має особливе значення для малого бізнесу, де управлінські рішення часто базуються не лише на результатах прогнозу, а й на довірі до логіки системи.

Завдяки своїй наочності, дерево рішень може використовуватись не лише для внутрішньої аналітики, а й як інструмент для побудови стратегій управління клієнтськими сегментами. Наприклад, якщо з моделі видно, що клієнти, які рідко купують у вихідні й мають середній чек до 200 грн, найчастіше не повертаються

після першого візиту, – це дає підстави для запуску спеціальних пропозицій саме для цієї групи.

У порівнянні з іншими пояснювальними моделями, такими як логістична регресія, дерево рішень не вимагає лінійності залежностей і має нижчий бар'єр входу для користувачів без математичної підготовки. Логістична регресія, хоча й є точним інструментом у випадку простої бінарної класифікації, не забезпечує тієї гнучкості в описі складних умов, які легко моделюються за допомогою дерева. Натомість дерево дозволяє відразу зрозуміти, які фактори впливають на рішення системи, а також як вони поєднуються у реальному житті.

Саме тому дерево рішень було інтегроване у функціонал системи як окрема частина модуля аналітики клієнтів. В результаті користувач може не лише переглянути аналітичні показники, а й візуально побачити логіку прогнозу, що особливо корисно для власника або маркетолога, який приймає рішення на основі отриманих сегментів і прогнозів. Такий підхід забезпечує інтерпретованість, практичну цінність та довіру до системи, що є ключовими факторами для малого сервісного бізнесу.

2.4 Прогнозування відтоку клієнтів методом Random Forest

В умовах конкурентного ринку ефективне управління клієнтською базою передбачає не лише ретроспективний аналіз, а й можливість прогнозувати майбутню поведінку споживачів, зокрема – ймовірність втрати клієнтів. У сфері роздрібного бізнесу, де залучення нових покупців зазвичай значно дорожче за утримання існуючих, виявлення ознак потенційного відтоку стає важливим елементом маркетингової стратегії. Для вирішення цього завдання в даній роботі було застосовано метод Random Forest – сучасний алгоритм ансамблевого навчання, що є ефективним у численних практичних застосуваннях, зокрема в задачах класифікації поведінки споживачів.

Random Forest поєднує в собі велику кількість окремих дерев рішень (рисунок 2.2), кожне з яких навчається на випадковій підвибірці даних з випадковим набором ознак. Кінцевий результат визначається шляхом голосування дерев у випадку класифікації або усереднення прогнозів у задачах регресії. Така архітектура дозволяє досягти високої точності за рахунок зниження дисперсії, властивій окремим деревам, і суттєво зменшує ризик переобучення [15].

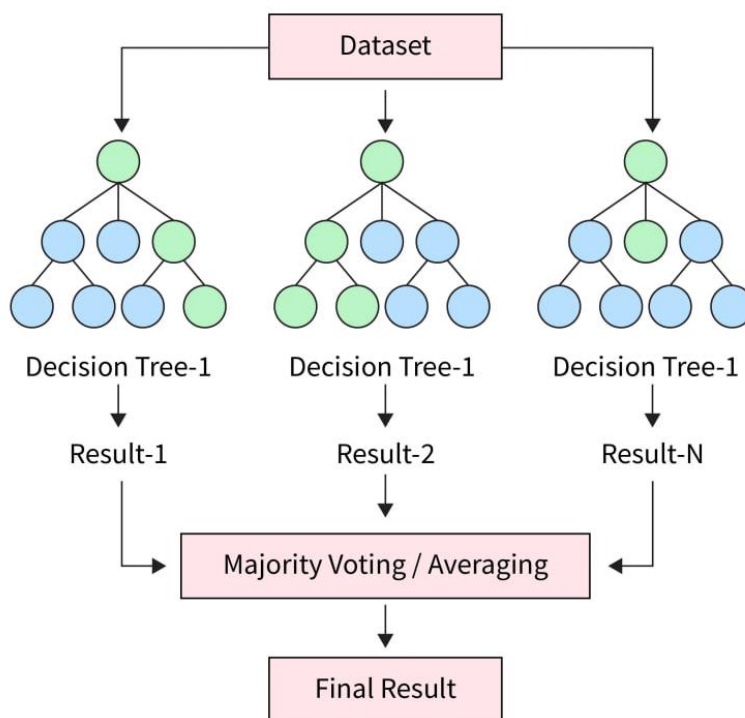


Рисунок 2.2 – Схема роботи алгоритму методу Random Forest

Перевагою методу Random Forest є його стійкість до шумів у даних, здатність обробляти як числові, так і категоріальні змінні без складної попередньої обробки, а також можливість працювати з великим числом ознак. Крім того, модель автоматично надає оцінки важливості ознак, що дає змогу зрозуміти, які фактори мають найбільший вплив на прогноз – тобто не лише зробити передбачення, а й пояснити його. Це особливо цінно в практичному контексті, де бізнес-користувачам важливо мати не просто «чорну скриньку», а інтерпретовану логіку ухвалення рішень.

У системі, що розробляється, метод Random Forest було обрано через його здатність виявляти складні, нелінійні взаємозв'язки між поведінковими показниками клієнтів і фактом їхньої втрати. Модель дозволяє аналізувати комбінований вплив таких факторів, як частота відвідувань, тривалість перерв між покупками, зміни структури чека, участь в акціях та багато інших. Завдяки цьому, система отримує змогу не лише ідентифікувати клієнтів із підвищеним ризиком відтоку, а й з'ясувати, що саме вплинуло на цей ризик.

У теоретичному контексті Random Forest належить до класу багатокомпонентних методів, які поєднують прогнози кількох слабких моделей для створення однієї сильної. Він виник як розвиток і вдосконалення класичних дерев рішень, які, хоч і легко інтерпретуються, мають обмежену стабільність і часто демонструють нестійку поведінку при незначній зміні вхідних даних. Random Forest усуває ці недоліки завдяки механізму бутстрапінгу і випадковій вибірці ознак, що дозволяє кожному дереву «бачити» лише частину інформації, роблячи загальний ансамбль менш схильним до переобучення.

У порівнянні з іншими методами прогнозування, наприклад, логістичною регресією, яка також широко застосовується для прогнозування відтоку, Random Forest не потребує припущення про лінійну форму залежностей між предикторами та цільовою змінною. Логістична модель має обмежену здатність враховувати складні взаємодії між ознаками, які в реальній поведінці клієнтів часто мають нелінійний характер.

Ще однією альтернативою є градієнтний бустинг (наприклад, XGBoost або LightGBM), який часто досягає ще вищої точності, ніж Random Forest. Проте бустинг є більш чутливим до налаштувань, має більший ризик переобучення й складніший для інтерпретації, що ускладнює його використання у системах, орієнтованих на широку аудиторію користувачів. У цьому сенсі Random Forest забезпечує оптимальний баланс між точністю, стійкістю та зрозумілістю – саме ті характеристики, які є критично важливими для систем, створених для малого бізнесу [16].

На додаток до технічних переваг, метод також добре масштабується – тобто дозволяє розширювати обсяг даних або додавати нові змінні без суттєвих змін в архітектурі моделі. Це дає змогу у майбутньому доповнити систему додатковими джерелами даних, наприклад, про взаємодію в соціальних мережах, вплив сезонності, або результати опитувань.

Для побудови моделі були використані ті самі дані, що й у попередніх модулях: транзакції, агреговані в розрізі клієнтів, з додаванням змінних, отриманих у результаті сегментації та кластеризації. До моделі було включено ознаки, які потенційно можуть мати зв'язок із ймовірністю відтоку: частота відвідувань за останні місяці, остання дата покупки, середній чек, участь у промо-акціях, кількість категорій товарів, динаміка зміни активності, день тижня останнього візиту тощо.

Цільовою змінною була бінарна ознака: 1 – якщо клієнт віднесений до категорії «ризик відтоку» (відсутність покупок упродовж останніх 90 днів), 0 – якщо клієнт залишається активним. Для навчання моделі було сформовано навчальну вибірку з історичних даних, розділену на тренувальний і тестовий набори у пропорції 80/20.

Модель була реалізована з використанням пакету `randomForest` у середовищі R. При побудові моделі оптимізувалися такі параметри, як кількість дерев (`ntree`) і кількість змінних, які враховуються при кожному поділі (`mtry`). Для оцінки якості моделі використовувалися метрики точності, повноти, специфічності, а також побудова матриці помилок і ROC-кривої [17]. ROC-крива (Receiver Operating Characteristic) відображає співвідношення між чутливістю (True Positive Rate) та хибнопозитивною частотою (False Positive Rate) за різних порогів класифікації, що дає змогу оцінити здатність моделі розрізняти між клієнтами, які залишаються активними, та тими, що мають ризик відтоку. В результаті вдалося досягти достатньо високого рівня точності при прогнозуванні відтоку, а найважливішими предикторами були визначені: `Recency`, `Frequency`, участь в акціях, середній чек та кількість унікальних днів візиту.

Результати моделі були інтегровані в окремий модуль веб-застосунку. Користувач отримує не лише список клієнтів із підрахованою ймовірністю відтоку,

а й візуалізацію найважливіших факторів, які впливають на це значення. Крім того, реалізована можливість фільтрації клієнтів за ризиком, сегментом та поведінковими ознаками для подальших цільових дій – наприклад, запуску персоналізованих пропозицій для збереження лояльності.

Отже, застосування Random Forest у проєкті дозволяє не лише ідентифікувати клієнтів з високим ризиком відтоку, а й надати аналітично обґрунтовані підказки для подальших дій – наприклад, запуску персоналізованих акцій, зміни комунікації або пріоритизації сегментів для глибшої роботи. Це переводить систему з рівня описової аналітики на рівень предиктивної та рекомендаційної, що відповідає сучасним трендам у сфері управління клієнтським досвідом.

2.5 Оцінка ефективності акцій за допомогою ключових показників

У сфері роздрібного обслуговування маркетингові акції є одним із ключових інструментів стимулювання попиту, формування лояльності та підвищення середнього чека. Проте без кількісної оцінки їхнього впливу рішення щодо повторного запуску тієї чи іншої акції часто ухвалюються інтуїтивно. Тому у рамках даної дипломної роботи було передбачено модуль, який дозволяє аналізувати ефективність промоційних заходів на основі ключових бізнес-показників – KPI (Key Performance Indicators) [18].

Оцінка ефективності акцій здійснюється шляхом порівняння динаміки основних показників до та після проведення акції. Зокрема, аналізуються такі метрики:

- загальна виручка – дозволяє оцінити, чи призвела акція до зростання обсягу продажів;
- середній чек – демонструє зміну структури споживання внаслідок застосування знижок або бонусів;

- кількість унікальних клієнтів – свідчить про здатність акції залучити нових відвідувачів або активізувати неактивних клієнтів;

- частка акційних товарів у загальному обсязі продажів – дозволяє визначити, чи не призвела акція до канібалізації інших товарів;

- поведінкові індикатори – наприклад, частота повернення після акції, комбінації товарів у чеках, реакція різних сегментів клієнтів.

Для побудови такого аналізу в системі реалізовано порівняння показників за обраними періодами: до акції, під час акції та після її завершення. Це дозволяє виявити короткостроковий та середньостроковий вплив заходу на продажі. Користувач може отримати візуалізацію результатів у вигляді графіків, таблиць та інтерактивних підсумкових панелей.

Отже, розглянутий підхід до аналізу маркетингових заходів дозволяє отримати кількісно обґрунтовані висновки щодо доцільності впровадження, повторного запуску або модифікації акцій.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ

У цьому розділі описано технічну основу розробки аналітичної системи для аналізу транзакційних даних у книгарні-кав'ярні. Реалізація системи здійснювалася з урахуванням обмежених ресурсів малого бізнесу, тому ключовими критеріями були простота розгортання, відсутність ліцензійних обмежень, швидкодія на середньостатистичних ПК та можливість локального використання.

Середовище реалізації обрано так, щоб об'єднати інструменти імпорту, обробки, аналізу, візуалізації й інтеграції результатів у єдину структуру. У підрозділах нижче наведено опис архітектури програмного продукту, мови програмування, використаних бібліотек і фреймворків, бази даних, середовища розробки, а також пояснено вибір технологій.

3.1 Архітектура програмної системи

Архітектура створеного програмного забезпечення побудована за принципами модульності, прозорості взаємодії компонентів та мінімальної залежності від зовнішнього середовища. Система реалізована у вигляді локального веб-додатку з використанням мови програмування R та фреймворку Shiny, що забезпечує двосторонню взаємодію між користувачем та серверною логікою через графічний інтерфейс у браузері [19].

Функціонально програмний продукт поділяється на три основні рівні:

1) рівень користувача (User Interface Layer) – реалізований за допомогою бібліотек shiny, shinydashboard та shinyWidgets. Користувач взаємодіє з системою через вкладки, фільтри, перемикачі та кнопки дій, обираючи періоди аналізу, типи сегментацій, параметри моделей. Інтерфейс адаптований до трьох основних ролей користувачів: власника, адміністратора та аналітика;

2) серверна логіка (Server Layer) – відповідає за обробку введених параметрів, виконання запитів до бази даних, формування аналітичних таблиць,

побудову графіків, запуск моделей сегментації, кластеризації та прогнозування. Реактивна модель Shiny забезпечує миттєве оновлення виводу у відповідь на зміни параметрів без потреби перезавантаження сторінки;

3) рівень даних (Data Layer) – представлений у вигляді локальної бази даних SQLite. У базі зберігається інформація про транзакції, клієнтів, товари, категорії. З'єднання з БД виконується через бібліотеку RSQLite, а доступ до таблиць – через SQL-запити або за допомогою пакетів dplyr та DBI.

Логіка взаємодії в системі передбачає наступний такий сценарій: користувач у графічному інтерфейсі задає параметри аналізу (наприклад, період або вибір сегмента клієнтів), після чого ці параметри надсилаються на сервер, де запускається відповідний блок обробки даних (рисунок 3.1). Результати – у вигляді графіків, таблиць, дашбордів або PDF-звітів – повертаються до інтерфейсу й відображаються у відповідному модулі.



Рисунок 3.1 – Діаграма архітектури програми

Кожен функціональний блок системи – аналіз KPI, сегментація клієнтів (RFM), кластеризація товарів і покупців, аналіз акцій, прогнозування відтоку – реалізований як окремий модуль, що підключається до основної архітектури через

вкладки й реактивні елементи. Це дозволяє масштабувати систему: додавати нові модулі або змінювати наявні без впливу на роботу інших компонентів.

Також у межах архітектури реалізовано базову систему керування доступом на основі ролей користувачів. Залежно від ролі, користувач отримує доступ лише до відповідного функціоналу: наприклад, адміністратор працює з формами завантаження та оновлення даних, аналітик – із модулями кластеризації та прогнозування, а власник – зі зведеними звітами та дашбордами.

Таким чином, архітектура програмного продукту поєднує в собі простоту реалізації, гнучкість у налаштуванні, низьку залежність від зовнішніх ресурсів і можливість подальшого розвитку. Обраний підхід дозволяє ефективно використовувати наявні транзакційні дані для прийняття обґрунтованих рішень у середовищі малого сервісного бізнесу.

3.2 Використані технології та середовище розробки

Інструменти, що використовуються у системі, мають забезпечити як точність аналітики, так і зручність інтерпретації результатів для користувача. Вибір технологій був орієнтований на можливість реалізації складних обчислювальних алгоритмів, побудови зручного інтерфейсу й забезпечення підтримки локальної бази даних без потреби в складній інфраструктурі.

3.2.1 Мова програмування та середовище розробки

Основою для розробки обрано мову програмування R, яка використовується для статистичного аналізу та побудови моделей машинного навчання. R має велику кількість пакетів, що дозволяють ефективно працювати з табличними даними, візуалізувати інформацію, будувати моделі кластеризації, регресії, прогнозування та оцінювати ефективність маркетингових заходів [20].

Розробка здійснювалася в інтегрованому середовищі RStudio, яке забезпечує зручний інтерфейс для написання коду, тестування та інтеграції з іншими елементами системи. Використання R дозволило реалізувати весь цикл роботи з даними: від імпорту до виведення фінального звіту у вигляді PDF-документа чи дашборду [21].

3.2.2 Веб-фреймворк для інтерфейсу

Для створення інтерактивного інтерфейсу системи використано фреймворк Shiny, який дозволяє зручно створювати веб-додатки. Shiny забезпечує двосторонню взаємодію з користувачем: дані вводяться через елементи керування (фільтри, селектори, дати), а результати автоматично оновлюються у вигляді графіків, таблиць чи значень показників [22].

Інтерфейс реалізовано у вигляді структурованої панелі з декількома вкладками: огляд основних метрик, RFM-сегментація, аналіз акцій, прогноз відтоку, інструменти для адміністратора. Це дозволяє користувачам з різними правами доступу працювати з системою відповідно до своїх потреб.

3.2.3 Пакети для аналізу та візуалізації

Для реалізації аналітичної частини системи було використано низку спеціалізованих пакетів мови R, які охоплюють візуалізацію, кластеризацію, машинне навчання та інтерфейсну інтеграцію. Зокрема, пакети ggplot2 і plotly забезпечують побудову якісних графіків – як статичних, так і інтерактивних, що особливо зручно для дашбордів. Для виконання кластерного аналізу клієнтів застосовуються пакети cluster, factoextra та вбудована функціональність kmeans, яка дозволяє групувати клієнтів за подібністю транзакційної поведінки та візуалізувати ці групи для подальшого аналізу.

Алгоритми прогнозування ймовірного відтоку реалізовано через пакети `randomForest` і `rpart`, які забезпечують побудову дерев рішень та моделей випадкових лісів, а також `caret`, що дозволяє стандартизувати підхід до побудови моделей і оцінювання їх якості. Інтерфейсна частина програми збагачена можливостями `shinyWidgets`, `shinydashboard` та `DT`, які додають сучасні елементи взаємодії з користувачем, як-от інтерактивні таблиці, динамічні панелі керування та стилізовані віджети.

Експорт результатів аналізу у форматі PDF реалізується за допомогою `rmarkdown` і `knitr`. Це дозволяє швидко формувати підсумкові звіти з візуалізаціями та ключовими показниками, що зручно для внутрішньої аналітики.

3.3 База даних та структура таблиць

У системі використовується база даних SQLite [23], яка містить чотири основні таблиці: `transactions`, `clients`, `products`, `promotions` (рисунок 3.2). Таблиця `clients` містить дані про клієнтів, які здійснюють покупки. Поле `customerID` є первинним ключем і унікальним ідентифікатором кожного клієнта. Додаткові поля включають ім'я (`customerName`), прізвище (`customerSurname`), дату народження (`customerBD`), телефон (`customerPhone`) та адресу електронної пошти (`customerEmail`). Ця інформація використовується як для ідентифікації клієнта, так і для побудови сегментів за демографічними або поведінковими ознаками. У таблиці `products` зберігається перелік товарів, що доступні у книгарні-кав'ярні. Основним ключем є `productID`. Крім назви товару (`productName`), таблиця включає кілька рівнів категорій (`category`, `category2`, `category3`), що дозволяє гнучко класифікувати продукцію. Поле `price` (ціна товару) використовується для аналізу структури виторгу. Така структура дає змогу аналізувати продажі за категоріями, виявляти лідерів продажів і розраховувати рентабельність.

Інформація про маркетингові акції знаходиться в таблиці `promotions`, де для кожної акції зазначено її тип, розмір знижки, цільову категорію, порогові умови, а

також період дії. Це дозволяє проводити оцінку ефективності кампаній як у цілому, так і для окремих сегментів клієнтів чи товарів.

Центральна таблиця для всього аналізу – transactions (таблиця 3.1). Вона відображає кожну окрему покупку, здійснену клієнтами. Первинним ключем є purchaseID, що унікально ідентифікує кожну транзакцію. Таблиця також містить дату (purchaseDate) та час (purchaseTime) покупки, а також зовнішні ключі до таблиць clients, products та promotions. Поле amount вказує кількість одиниць кожного придбаного товару. Ця таблиця використовується для побудови усіх основних аналітичних метрик: виручки, середнього чека, частоти покупок, поведінкових кластерів тощо.

Таблиця 3.1 – Приклад заповненої таблиці транзакцій

check	date	customer	promo	product	category	category2	qty	total
20250 4- 29035	2025- 04-29	2556	1	Убивство Роджера Екройда	Книги	Детектив	3	690
20250 4- 29073	2025- 04-29	4652		Зіткнення	Книги	Романтич на проза	1	660
20250 4- 29065	2025- 04-29	3931	3	Хірург. Книга 1	Книги	Трилери та жахи	3	900
20250 4- 29076	2025- 04-29	3931		Гарячий шоколад	Меню	Чай та інші напої	2	150
20250 4- 29061	2025- 04-29	2228		Американ о	Меню	Кавові напої	1	60

Взаємозв'язки реалізовано через зовнішні ключі: клієнти пов'язані з транзакціями через customerID, товари – через productID, акції – через promoID. При об'єднанні таблиць створюється зведена вітрина з додатковими обчисленими змінними: сума знижки, ознака участі в акції, категорія товару тощо.

Дані зчитуються за допомогою SQL-запитів (dbGetQuery) і обробляються через dplyr. Для визначення ідентифікаторів чеків використовується генерація унікальних newCheckID. Ідентифікатор створюється на основі дати транзакції та її унікального номеру. Обробка дат і знижок реалізована із застосуванням lubridate та case_when.

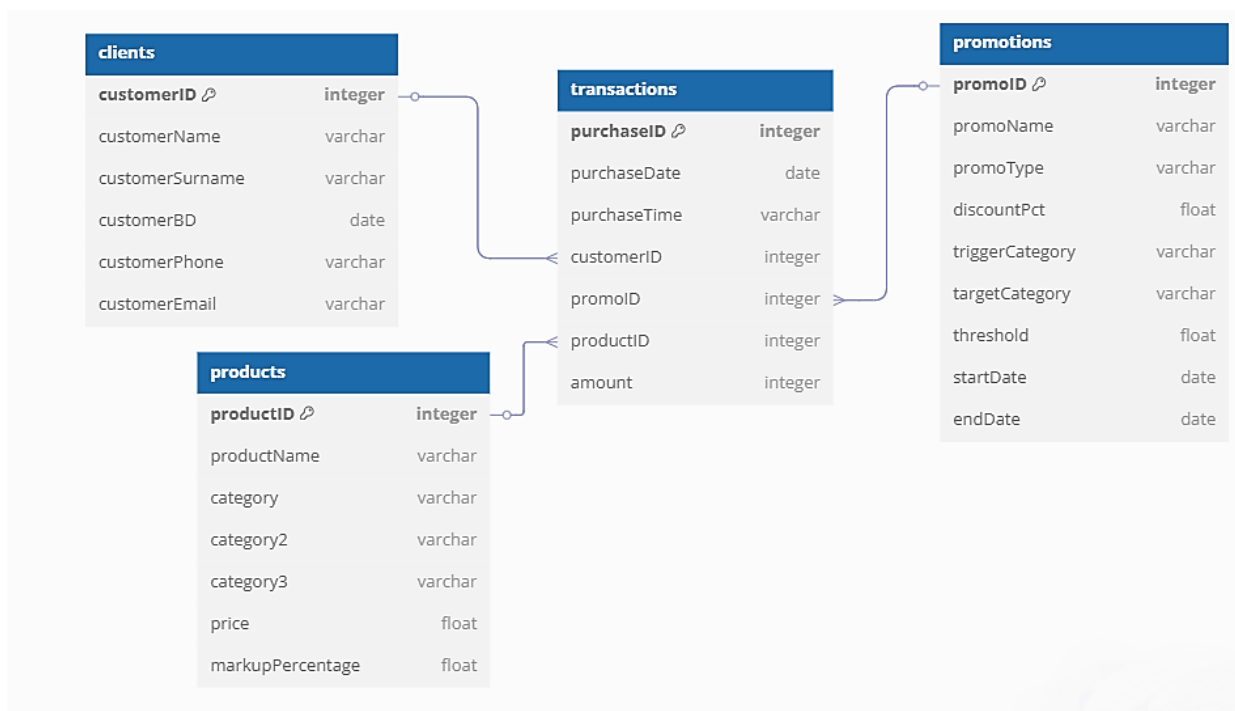


Рисунок 3.2 – ER-діаграма

При запуску додатку дані з таблиць об'єднуються в «плоску аналітичну таблицю», яка служить основою для візуалізацій, кластеризації, сегментації клієнтів, прогнозування відтоку та оцінки ефективності маркетингових дій. Завдяки нормалізованій структурі та чітким зв'язкам система зберігає високу гнучкість і масштабованість навіть при зростанні обсягу транзакцій. У процесі агрегації автоматично обчислюються ключові показники клієнтів, такі як частота візитів, середній чек, давність останньої покупки, частка покупок акційних товарів

3.4 Діаграма прецедентів

Для візуалізації логіки роботи користувача з системою створено діаграму прецедентів [24], що показує взаємодію трьох основних ролей: власника, аналітика і адміністратора. На рисунку 3.3 представлена діаграма прецедентів, яка ілюструє взаємодію трьох основних ролей користувачів системи: адміністратора, аналітика та власника. Кожна роль має доступ до відповідного набору функцій, що реалізовані у вигляді вкладок у системі.



Рисунок 3.3 – Діаграма прецедентів

Адміністратор має доступ до перевірки якості та коректності даних, завантаження транзакцій, перегляду таблиць, виявлення аномалій і завантаження CSV-файлів із помилками. Ці дії забезпечують підтримку актуальності та коректності інформації в системі.

Аналітик працює з модулями, що відповідають за сегментацію клієнтів, прогнозування відтоку, аналіз ефективності акцій, кластеризацію товарів і

побудову дерева рішень. Ці функції дозволяють проводити комплексну оцінку динаміки продажів, виявляти поведінкові патерни та формувати висновки щодо цільових груп товарів або клієнтів.

Власник має спрощений інтерфейс із доступом до перегляду зведеного звіту та завантаження PDF-файлу з результатами. Це дозволяє отримати ключові бізнес-показники без заглиблення в технічні деталі.

З діаграми можна побачити, що деякі прецеденти включають або розширюють інші дії. Наприклад, прогнозування відтоку включає попередню кластеризацію клієнтів; перегляд дерева рішень є розширенням функції кластеризації. Це дозволяє забезпечити логічну послідовність взаємодії та гнучке управління доступом.

4 РОБОТА КОРИСТУВАЧА З ПРОГРАМНОЮ СИСТЕМОЮ

Цей розділ описує процес взаємодії користувача з розробленим веб-додатком, побудованим на платформі R Shiny. У ньому розглянуто вимоги до комп'ютера та програмного забезпечення, інструкції з інсталяції, а також сценарії роботи для різних ролей користувачів – адміністратора, аналітика та власника.

4.1 Системні вимоги

Розроблена система аналізу транзакцій реалізована як веб-застосунок на платформі R Shiny та розгорнута на хмарному сервісі shinyapps.io [25]. Такий підхід забезпечує максимальну зручність для кінцевого користувача, оскільки не потребує встановлення жодного додаткового програмного забезпечення – достатньо мати доступ до Інтернету та сучасний браузер.

Для роботи з системою користувачу необхідно перейти за наданим посиланням, яке веде до опублікованої версії застосунку. Інтерфейс запускається автоматично у веб-браузері, і всі обчислення, фільтрація, аналіз та візуалізація даних виконуються на хмарному сервері. Це дозволяє використовувати систему на будь-якому пристрої – ноутбучі, ПК або планшеті – без налаштувань і технічних складностей.

Для роботи з системою достатньо мати пристрій із доступом до Інтернету та встановленим сучасним веб-браузером в актуальній версії. Додаткове програмне забезпечення або встановлення локальних компонентів не потрібні.

Інтерфейс застосунку адаптований до трьох ролей користувачів: власника закладу, адміністратора й аналітика. Усі функціональні блоки згруповані у вкладки: перегляд основних показників, сегментація клієнтів, аналіз акцій, прогнозування відтоку, робота з довідниками тощо. Візуалізація результатів, таблиці та аналітичні звіти доступні безпосередньо у браузері.

Таким чином, система є повністю готовим до використання онлайн-рішенням, що не потребує інсталяції, та може бути ефективно застосована в малому бізнесі.

4.2 Вхід у систему та опис головної сторінки

Після запуску системи користувач бачить форму аутентифікації (рисунок 4.1).

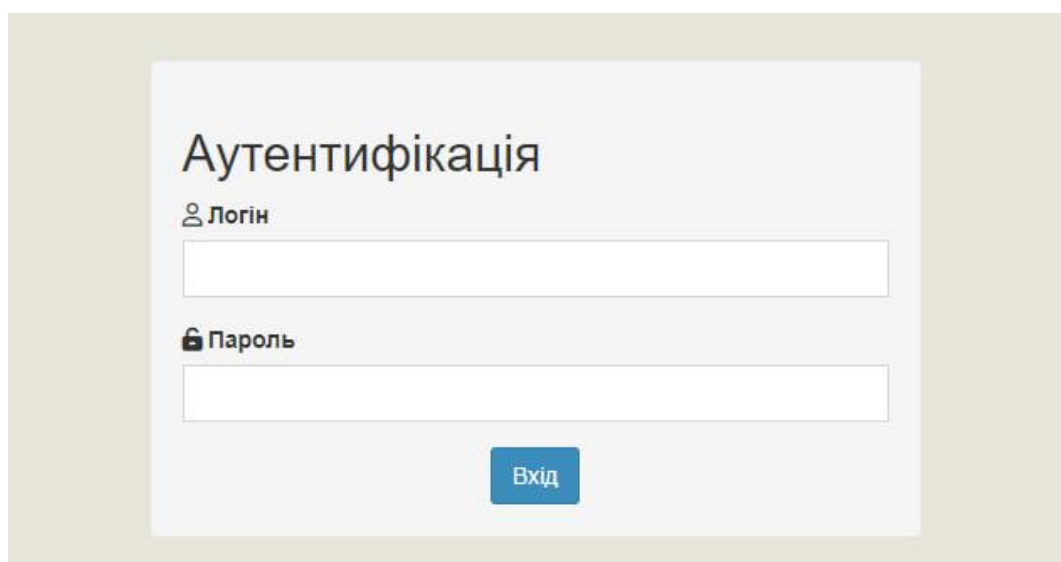
The image shows a web form for authentication. At the top, the title 'Аутентифікація' is displayed in a large, dark font. Below the title, there are two input fields. The first field is labeled 'Логін' (Login) and has a small user icon to its left. The second field is labeled 'Пароль' (Password) and has a small lock icon to its left. Below the password field, there is a blue button with the text 'Вхід' (Login) in white. The entire form is set against a light gray background, which is itself centered on a larger, light beige background.

Рисунок 4.1 – Форма авторизації

Це вікно авторизації побудоване для обмеження доступу до функціоналу програми відповідно до ролі користувача (адміністратор, аналітик або власник). Інтерфейс складається з двох обов'язкових полів: логін та пароль, а також кнопки «Вхід». Після введення облікових даних та натискання кнопки, система автоматично визначає роль користувача й відкриває лише ті вкладки, які йому дозволені.

4.3 Інтерфейс адміністратора системи

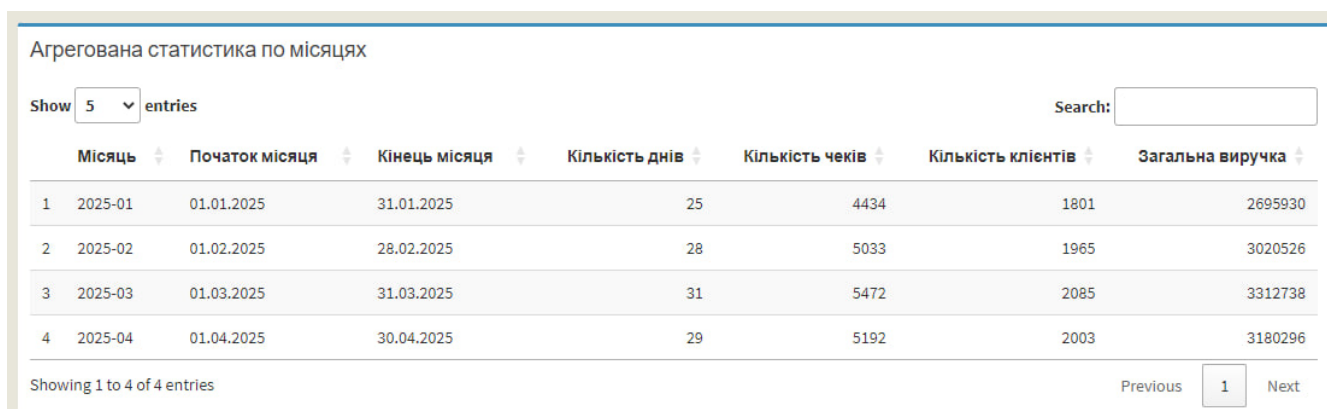
Розпочнемо з авторизації під обліковим записом адміністратора. На головній вкладці «Дані» адміністратор має доступ до інтерактивного перегляду основних таблиць, зокрема транзакцій, клієнтів, товарів та акцій. Інтерфейс реалізований у вигляді зручної фільтрованої таблиці з можливістю пошуку, сортування та фільтрації по кожному стовпцю – це зображено на рисунку 4.2. Це дозволяє швидко знайти необхідну інформацію, перевірити її коректність або підготувати до експорту. Наприклад, у таблиці транзакцій відображаються поля CheckID, дата, ID клієнта, назви товарів, категорії, кількість, ціна за одиницю та загальна сума. Усі дані витягуються з бази SQLite в режимі реального часу.

Рисунок 4.2 – Вкладка «Дані»

На вкладці «Аудит» адміністратор бачить статистику щодо повноти даних. Система автоматично перевіряє наявність пропущених значень, нульових або негативних сум, невалідних дат або ID. Якщо помилки виявлено, вони з’являються у таблиці нижче з деталізацією по кожному полю. За потреби адміністратор може натиснути кнопку «Завантажити CSV з проблемними рядками» – система згенерує

окремий файл для подальшого виправлення. Якщо ж жодних критичних помилок не виявлено, таблиця залишається порожньою.

На рисунку 4.3 представлено агреговану статистику по місяцях, що включає кількість днів, чеків, клієнтів і загальну виручку. Такий огляд дозволяє оцінити обсяг даних у кожному періоді та своєчасно виявити аномалії – наприклад, дублювання інформації за окремі місяці або пропущені дні. Це допомагає уникнути використання неповних або спотворених даних у подальшому аналізі.



Агрегована статистика по місяцях

Show entries Search:

	Місяць	Початок місяця	Кінець місяця	Кількість днів	Кількість чеків	Кількість клієнтів	Загальна виручка
1	2025-01	01.01.2025	31.01.2025	25	4434	1801	2695930
2	2025-02	01.02.2025	28.02.2025	28	5033	1965	3020526
3	2025-03	01.03.2025	31.03.2025	31	5472	2085	3312738
4	2025-04	01.04.2025	30.04.2025	29	5192	2003	3180296

Showing 1 to 4 of 4 entries Previous Next

Рисунок 4.3 – Агрегована статистика аудиту даних по місяцях

Третя вкладка – «Завантаження даних» – дозволяє імпортувати нові Excel-файли з транзакціями. Після вибору файлу користувач натискає кнопку «Обробити завантаження», і система автоматично імпортує нові дані у відповідну таблицю бази даних.

4.4 Інтерфейс аналітика

Розглянемо наступного користувача системи – аналітика. Йому доступно три вкладки: аналізу акцій, кластеризації клієнтів та товарів, і прогнозування відтоку. Далі буде детально розказано про кожен з цих сторінок.

4.4.1 Вкладка «Аналіз акцій»

Перший з доступних аналітику модулів – це «Аналіз акцій», який дозволяє оцінити ефективність конкретних промоцій. У верхній частині вкладки відображаються три ключові показники: загальна виручка під час дії акції, кількість унікальних клієнтів та середній чек. Для кожного з них зазначено також відсоткову зміну відносно контрольного періоду до акції, що дозволяє оперативно оцінити вплив промоції на основні метрики. Колір картки автоматично змінюється залежно від характеру змін: зелений позначає позитивну динаміку (зростання виручки, кількості клієнтів або середнього чека), а червоний – про негативну (зменшення цих показників).

У полі «Оберіть акцію» користувач може обрати зі списку потрібну маркетингову кампанію. Таким чином можна порівняти ефективність акцій за ключовими метриками. А саме як змінилася загальна виручка, кількість унікальних клієнтів та середній чек до та під час кожної акції. Наприклад, на рисунках 4.4 та 4.5 порівняно результати двох акцій, і видно, що друга з них значно ефективніша, адже під час її проведення усі три показники зросли, тоді як під час першої акції спостерігається зменшення виручки та кількості унікальних клієнтів. Такий формат дозволяє не лише швидко оцінити ефективність акції, а й порівнювати між собою кілька промоцій на основі зрозумілих та візуально підкріплених індикаторів.



Рисунок 4.4 – Ключові показники менш ефективної акції «5% знижка при покупці від 1000 грн»

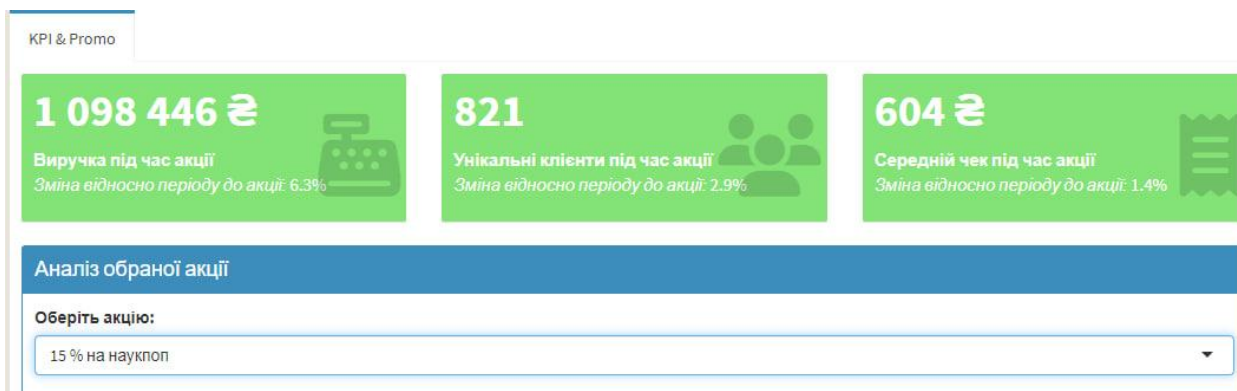


Рисунок 4.5 – Ключові показники більш ефективної акції «15% знижка на науково-популярну літературу»

Також система формує динамічні графіки, що відображають зміну виручки у розрізі трьох часових періодів: до початку акції, безпосередньо під час її проведення та після завершення (рисунок 4.6). Це дає змогу користувачеві оцінити, як саме промоційна активність вплинула на поведінку покупців. Кожен із періодів позначений окремим кольором: синій – для доакційного етапу, жовтий – для періоду дії акції, рожевий – для післяакційного. Під графіком подається таблиця з агрегованими значеннями для кожного з періодів: сума виручки, кількість клієнтів, середній чек та тривалість періоду в днях. Це дозволяє зіставити кількісні дані з візуальною картиною та отримати чітке уявлення про результати акції.



Рисунок 4.6 – Аналіз ефективності акції за трьома періодами

Завершальним елементом є автоматично сформований висновок по акції, в якому наводиться коротка інтерпретація результатів – це показано на рисунку 4.7.



Рисунок 4.7 – Автоматизовані висновки по акції

Якщо результати є негативними, це може свідчити про неефективність знижки та потребу в перегляді стратегії. У разі позитивного впливу система вказує на потенціал акції для повторного використання.

4.4.2 Кластеризація клієнтів

Вкладка «Кластеризація» дозволяє аналітику сегментувати клієнтську базу на основі RFM-аналізу (Recency, Frequency, Monetary). Інтерфейс модуля складається з інструментів налаштування кількості кластерів, автоматичного запуску кластеризації та візуалізації отриманих результатів. Спочатку користувачу необхідно обрати бажану кількість кластерів на повзунку (наприклад, 4 кластери, як показано на рисунку 4.8) і після цього натиснути кнопку запуску RFM-кластеризації.

Система автоматично розраховує метрики для кожного клієнта, масштабує їх і застосовує алгоритм K-means. Результати відображаються у вигляді кольорової діаграми розсіювання з групуванням клієнтів за сегментами. Додатково відображається показник середнього силуету, який допомагає оцінити якість розділення кластерів.

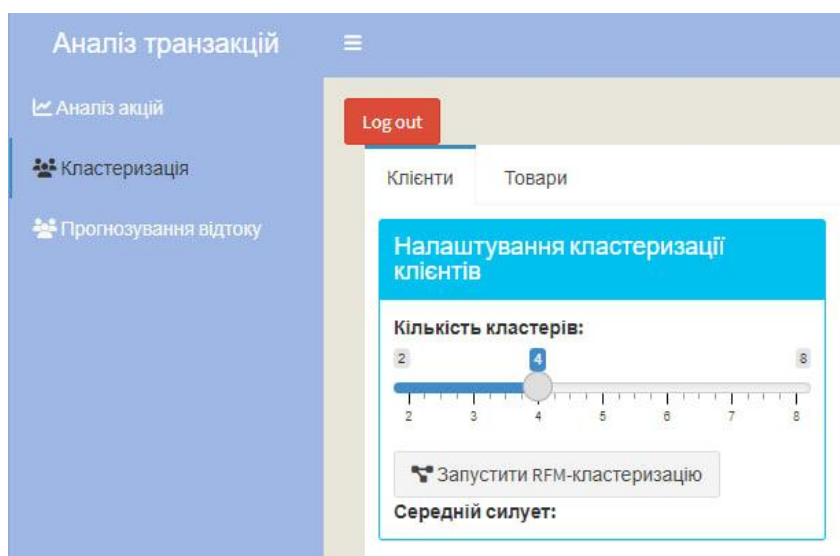


Рисунок 4.8 – Вибір кількості кластерів та запуск кластеризації

На рисунку 4.9 показано графік візуалізації результатів кластеризації, де кожен клієнт позначений точкою, а колір відповідає кластеру. По осі X відкладається давність останньої покупки (Recency), по осі Y – загальні витрати клієнта (Monetary). Такий підхід дозволяє швидко виявити ключові патерни – наприклад, виділити активних покупців, разових відвідувачів або клієнтів із високими витратами.

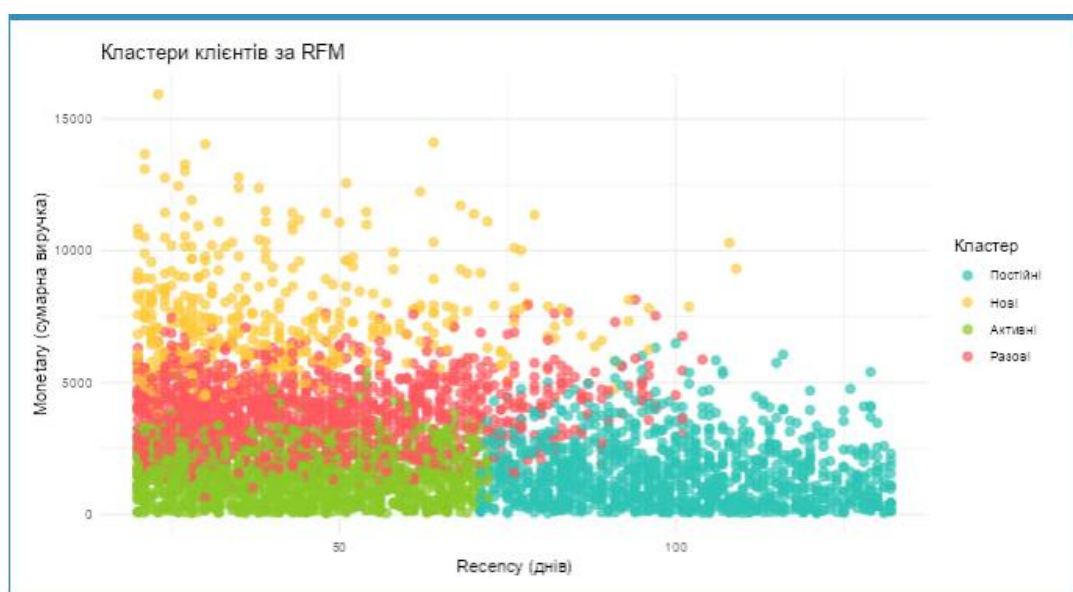


Рисунок 5.9 – Візуалізація кластеризації клієнтів

Під графіком подається таблиця з узагальненою статистикою по кожному кластеру: середні значення всіх трьох RFM-метрик, загальна виручка, яку приносить кластер, та його частка в загальному обсязі продажів (рисунк 4.10). Це дозволяє оцінити значущість кожної групи для бізнесу.

Підсумок кластерів клієнтів					
cluster_label	Середній_Recency	Середній_Frequency	Середній_Monetary	Загальна_виручка	Частка_виручки
Разові	45	3.1	3799.3	5125310	42
Нові	40	4.5	7405.7	3236278	26.5
Активні	44	1.6	1492.9	2093094	17.1
Постійні	99	1.4	1528.6	1754810	14.4

Рисунок 4.10 – Таблиця статистики по кластерах

Також користувач бачить візуалізацію розподілу клієнтів між кластерами у вигляді діаграми з кількістю і відсотковим співвідношенням, що допомагає оцінити розмір кожного сегмента. Праворуч відображено таблицю участі клієнтів у акціях за кожним кластером – скільки осіб брали участь і який відсоток від загальної кількості це становить. Це показано на рисунку 4.11.

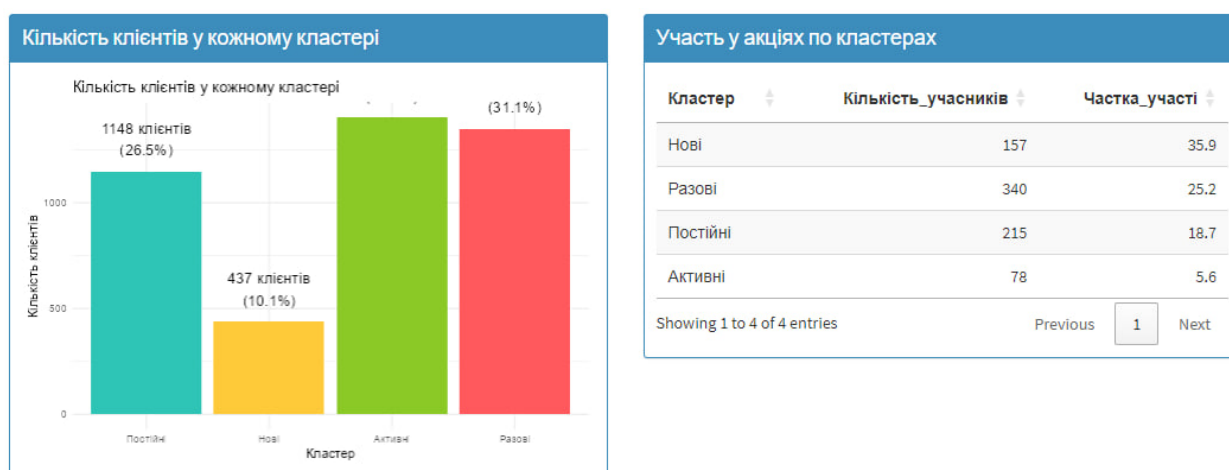


Рисунок 4.11 – Додаткова статистика по кластерах

У нижній частині вкладки розміщене дерево рішень, яке наочно інтерпретує межі між кластерами. Воно будується за алгоритмом gpart і дозволяє побачити, за якими правилами клієнти потрапляють до певної групи. Поруч подано таблицю, що дублює ці правила в зручному для аналізу форматі (рисунок 4.12).

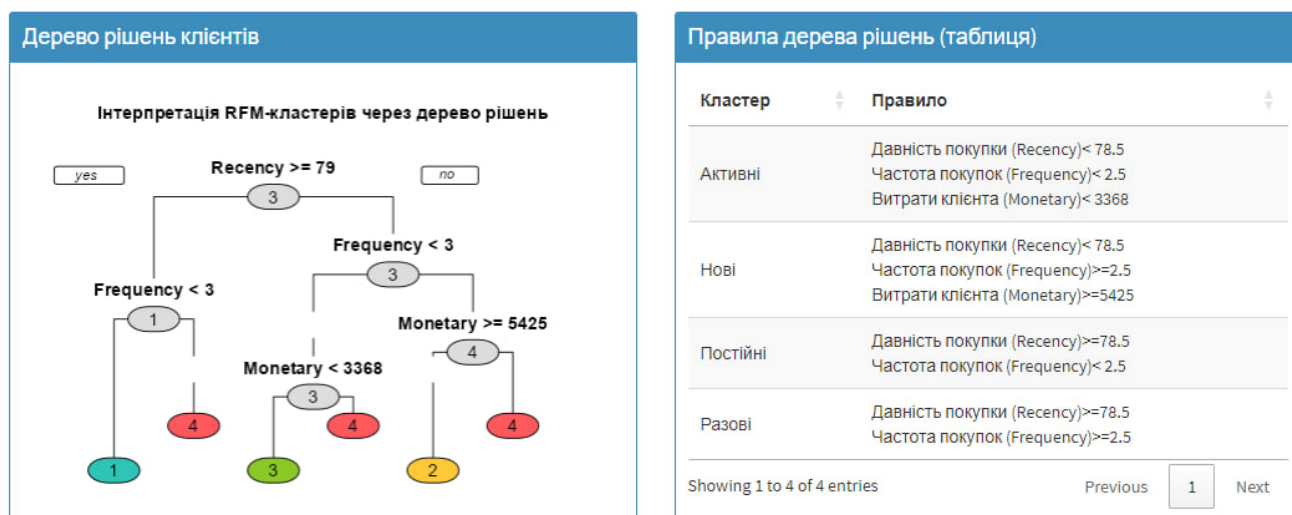


Рисунок 4.12 – Дерево рішень

4.4.3 Кластеризація товарів

На вкладці «Товари» користувач аналітик отримує можливість дослідити характеристики різних товарів на основі історії продажів. У верхній частині інтерфейсу розміщено текстові підказки з поясненням логіки кластеризації: система формує групи товарів за схожістю ключових показників – кількості продажів, кількості днів у продажу, загальної виручки та кількості клієнтів. В поясненні також зазначено, що ефективність кластеризації можна оцінити за допомогою середнього силуету, а за допомогою z-індексу можна дослідити аномальні товари. Крім цього, пояснення містять типові тлумачення результатів кластеризації: товари можуть бути класифіковані як хіти продажу, стабільні позиції, нішеві продукти, сезонні, або інші, що не вписались у чіткий профіль і вимагають додаткового аналізу. Панель підказок можна побачити на рисунку 4.13.

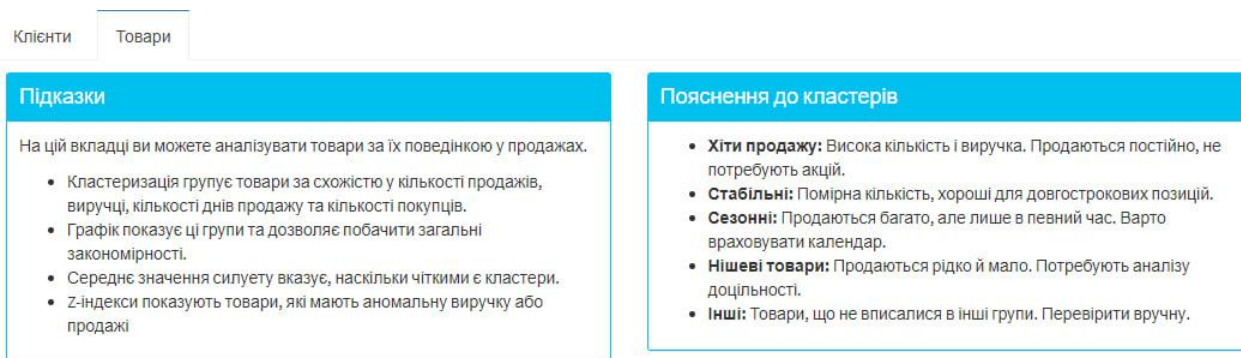


Рисунок 4.13 – Текстові підказки

Аналітик може самостійно обрати кількість кластерів (від 2 до 5) і задати осі для побудови двовимірної проєкції: наприклад, кількість продажів по осі X та виручка – по осі Y. Після натискання кнопки запуску кластеризації відбувається обчислення, і на правій панелі з'являється візуалізація розподілу товарів по кластерах. Кожна група позначена своїм кольором, що дозволяє легко виявити закономірності: наприклад, зелена група може відповідати стабільним товарам із помірною виручкою, а фіолетова – нішевим товарам із коротким циклом продажів (рисунок 4.14).

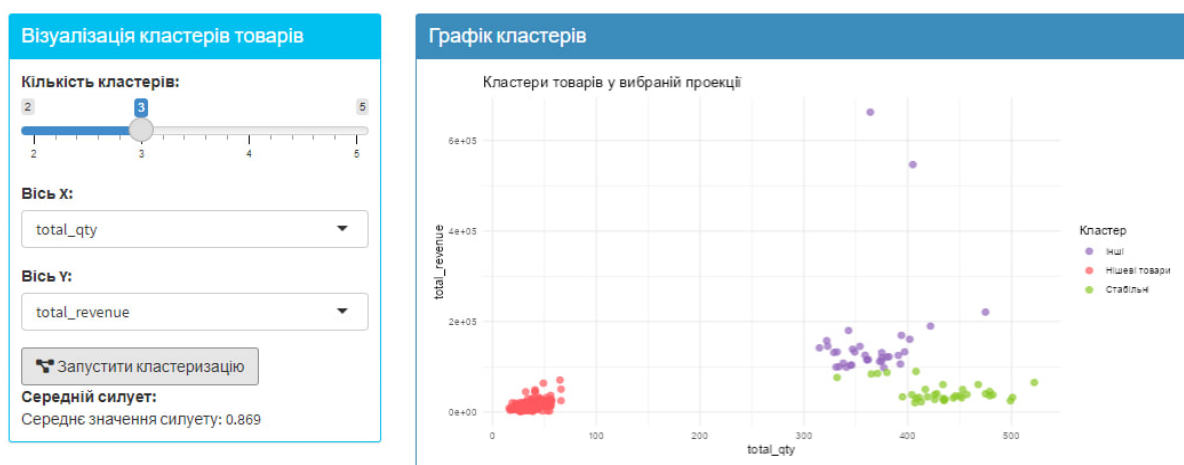


Рисунок 4.14 – Візуалізація кластеризації товарів

Під графіком подається таблиця, де для кожного кластеру наведено зведену статистику: кількість товарів, середня кількість продажів, середня виручка, середня

тривалість продажу та кількість клієнтів (рисунок 4.15). Це дозволяє оцінити внесок кожного кластеру в загальну структуру продажів.

Характеристики кластерів					
Show	5	▼	entries	Search:	
cluster_label	Кількість_товарів	Середня_кількість	Середня_виручка	Днів_продавалось	Кількість_клієнтів
Інші	35	364.9	157835.9	90.6	178.6
Нішеві товари	335	37.1	15397.6	17.1	18.4
Стабільні	35	434	43630.1	96.4	212

Рисунок 4.15 – Таблиця характеристики кластерів

Також користувач може переглянути таблицю, що демонструє типові товари для кожного кластеру, тобто найбільш характерні приклади позицій, які були віднесені до певної групи, а ще таблицю, яка виводить по кожному кластеру топ-5 найкращих і найгірших товарів за кількістю продажів.

Крім цього, система дозволяє виявляти аномальні товари – ті, що мають надмірно високу або низьку виручку чи кількість продажів відносно середнього значення по кластеру. Аномальність визначається за допомогою Z-індексу, і товари з надто високими чи низькими значеннями виводяться у відповідній таблиці. Їх також можна переглянути на графіку розподілу Z-значень, де легко виявити винятки в загальній структурі продажів – рисунок 4.16.

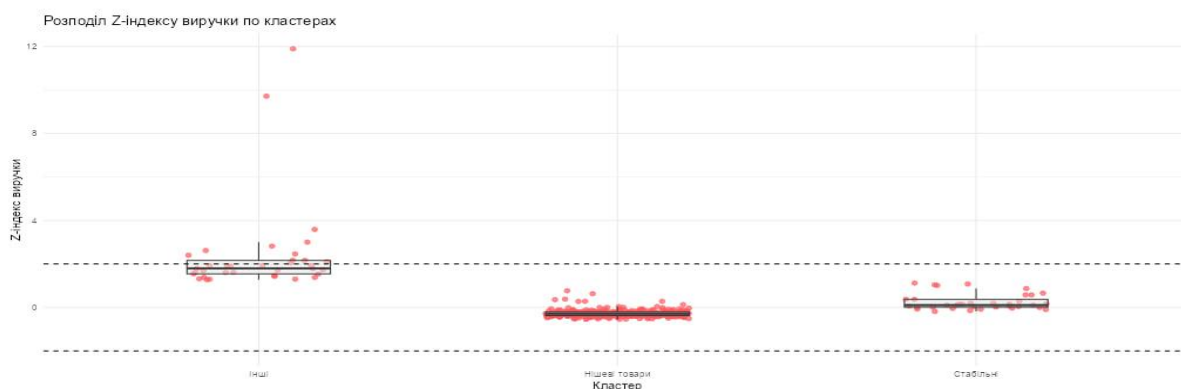


Рисунок 4.16 – Графік розподілу Z-значень по кластерах

4.4.4 Прогнозування відтоку

Вкладка прогнозування ймовірності відтоку клієнтів дає змогу оцінити ризики втрати клієнтів на основі алгоритму Random Forest. Інтерфейс починається з кнопки «Запустити прогноз», яка активує побудову моделі на основі результатів RFM-сегментації. На рисунку 4.17, представлена діаграма важливості змінних, де показано, які з RFM-параметрів найбільше впливають на модель. У прикладі найбільш вагомим фактором є давність останньої покупки.

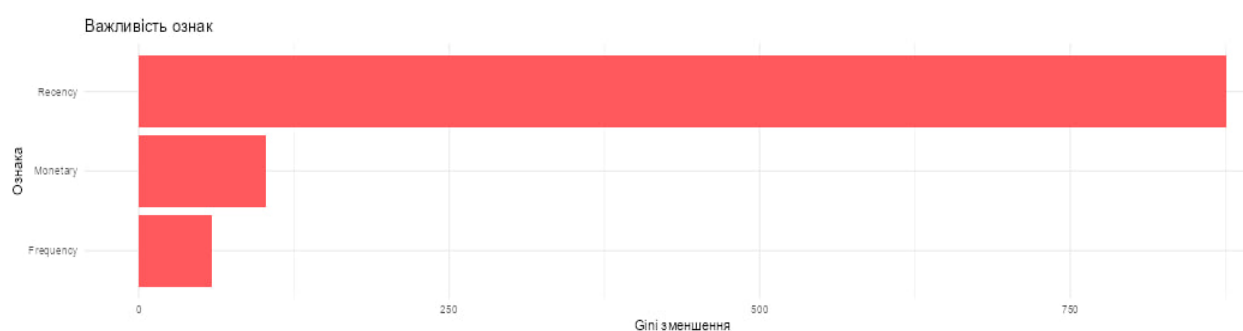


Рисунок 4.17 – Діаграма важливості змінних

Далі користувач бачить гістограму розподілу ймовірностей відтоку серед клієнтів, що допомагає виявити концентрацію ризикових груп (рисунок 4.18). Таблиця на рисунку 4.19 містить список клієнтів з ймовірністю відтоку кожного – від 0 до 1, разом із відповідними значеннями RFM-параметрів. Це дозволяє аналітику сформувати пріоритетний список клієнтів для подальшої роботи.

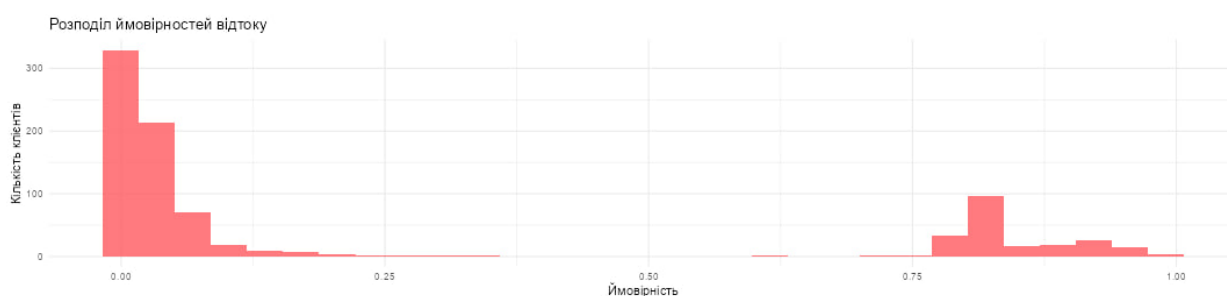


Рисунок 4.18 – Гістограма розподілу ймовірностей відтоку

Show entries Search:

customerID	Recency	Frequency	Monetary	churn_prob
2114	93	4	4880	0.99
215	97	3	3069	0.98
1754	99	3	805	0.98
4634	96	4	6250	0.98
1526	87	3	1685	0.97
2564	99	3	1650	0.97
2786	102	3	3425	0.97
274	101	4	3055	0.96
2200	105	3	3265	0.96

Рисунок 4.19 – Список клієнтів з високою ймовірністю стати відтоком

Окремим кольоровим блоком система надає практичні рекомендації щодо дій у відповідь на ризики: наприклад, клієнтам із ризиком понад 0.7 рекомендовано запропонувати персональну пропозицію, тоді як для середньої групи ризику варто застосовувати легкі нагадування через SMS чи e-mail.

В останньому блоці користувач може проаналізувати вплив кожної окремої змінної на ймовірність відтоку через графік PDP (Partial Dependence Plot). Наприклад, зміна частоти покупок дозволяє оцінити, як зростання чи спад цього показника впливає на ризик втрати клієнта.

4.5 Інтерфейс власника

Вкладка, доступна для власника, слугує основним зведеним дашбордом, що дозволяє швидко оцінити ключові показники діяльності закладу. У верхній частині інтерфейсу відображаються три основні метрики: загальна виручка за весь період, кількість унікальних клієнтів та середній чек.

Під показниками розміщено графік динаміки виручки по місяцях, що дає змогу оцінити зростання або спад продажів у короткостроковій перспективі. Поряд розташована кругова діаграма, яка ілюструє частку виручки по категоріях товарів.

Це дозволяє швидко визначити, які категорії є домінуючими, а які – другорядними. Це все зображено на рисунку 4.20.

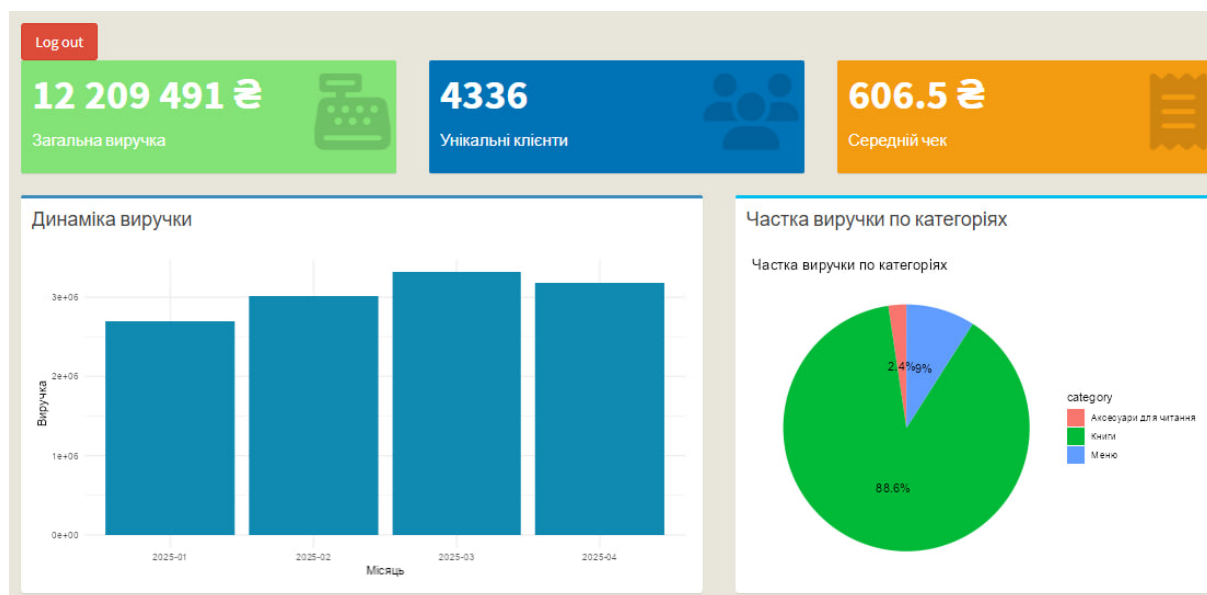


Рисунок 4.20 – Верхня частина дашборду на вкладці власника

Перший з графіків на рисунку 4.21 демонструє зміну середнього чека по місяцях, а другий – кількість клієнтів у ті самі періоди. Разом вони формують загальне уявлення про купівельну активність клієнтів та її коливання протягом аналізованого періоду.

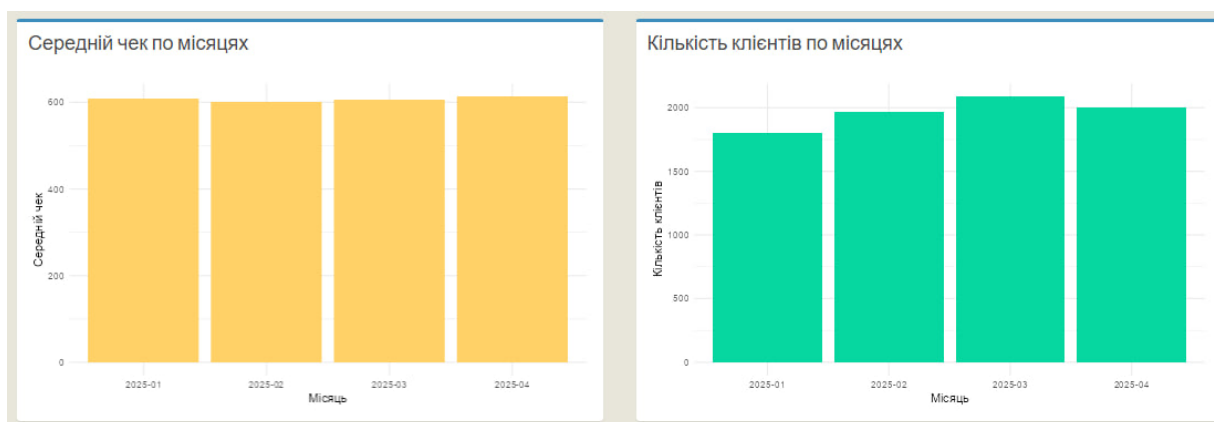


Рисунок 4.21 – Графіки динаміки показників по місяцях

У нижньому блоці розміщено таблицю з результатами проведених акцій. Вона містить дані про кількість транзакцій, виручку, середній чек та частку чеків, що включають комбінацію “книга + кава”. Цей показник важливий для оцінки перехресного попиту на основні позиції кафе та книгарні.

Поряд із таблицею акцій знаходиться перелік топ-5 товарів за кількістю покупок та виручкою. Цей список дозволяє визначити найприбутковіші позиції в асортименті. Також у нижній частині сторінки розміщено графік змін виручки під час акцій, що дозволяє наочно побачити, які промо-кампанії були ефективними, а які – мали негативний або нейтральний вплив. Під графіком формується автоматичний текстовий висновок на основі даних за останній місяць. У ньому підсумовується динаміка виручки, визначається провідна товарна категорія та виводиться частка чеків із комбінацією “книга + кава” (рисунок 4.22).

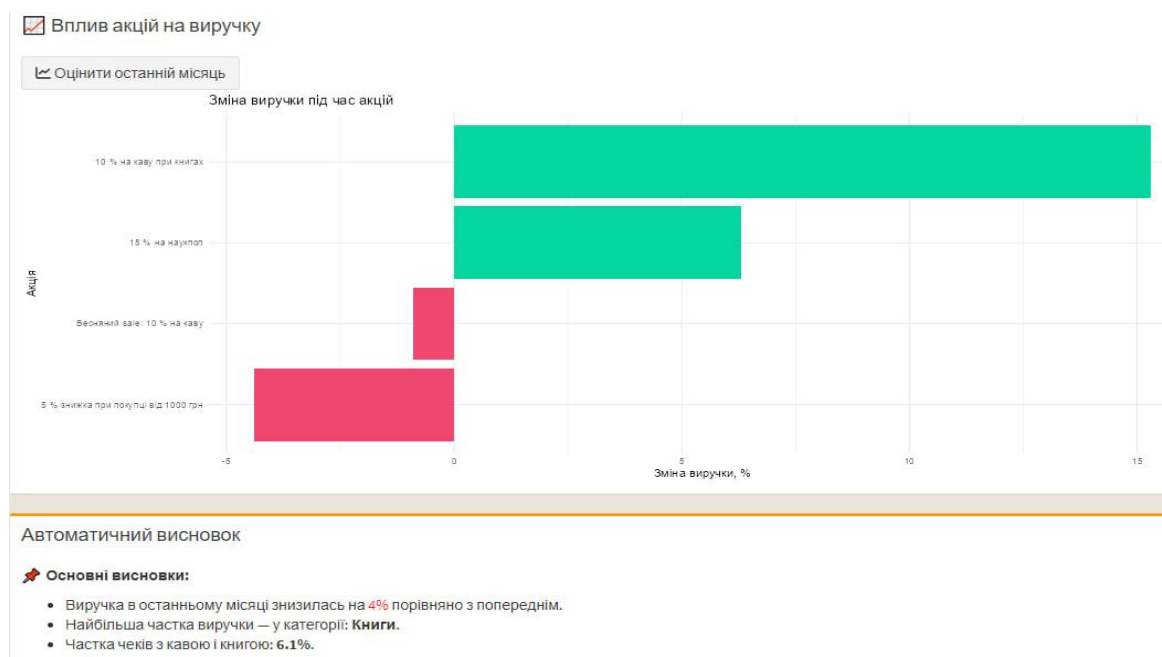


Рисунок 4.22 – Нижня частина дашборду на вкладці власника

Крім цього, користувач може завантажити переглянуті дані у вигляді PDF-звіту. Зведена вкладка для власника забезпечує комплексний огляд ключових показників, дозволяючи швидко оцінити ефективність діяльності, виявити найсильніші сторони бізнесу та вчасно зреагувати на негативні зміни.

ВИСНОВКИ

У процесі виконання дипломної роботи було проведено аналіз підходів до вирішення задачі аналізу транзакцій клієнтів у сфері роздрібного обслуговування. Розглядалися як класичні статистичні методи, так і сучасні алгоритми машинного навчання, які можуть бути ефективно застосовані в умовах малого бізнесу з обмеженими ресурсами. На основі вивчення теоретичних та практичних джерел обґрунтовано доцільність використання таких методів: RFM-аналіз для сегментації клієнтів за частотою, обсягами та давністю покупок; кластеризація методом K-means для поділу клієнтів на групи зі схожими характеристиками; алгоритм Random Forest – для прогнозування ймовірності відтоку; а також оцінка ефективності маркетингових заходів на основі ключових KPI-метрик, таких як виручка, середній чек, кількість унікальних клієнтів та поведінкові індикатори до, під час та після акції. Застосування саме цих методів дозволяє реалізувати як описову, так і прогностичну аналітику на прикладному рівні.

У ході дослідження програмних рішень, доступних для реалізації систем поведінкової аналітики, було обґрунтовано використання відкритого середовища R та фреймворку Shiny. Такий вибір забезпечив оптимальний баланс між гнучкістю, відсутністю ліцензійних обмежень, зручністю інтерактивної візуалізації та широким вибором бібліотек для роботи з даними. Також використано допоміжні пакети, зокрема ggplot2, plotly, DT, randomForest, cluster, shinyWidgets, які дозволили реалізувати повний аналітичний цикл — від обробки сировинних даних до виведення підсумкових звітів.

У рамках дипломної роботи було створено інтерактивну програмну систему, яка дозволяє здійснювати аналіз транзакцій в закладі обслуговування гібридного формату - книгарні-кав'ярні. Система включає інтерфейс для огляду ключових бізнес-показників, сегментацію клієнтів, оцінку ефективності промоцій, кластеризацію, прогнозування відтоку, а також модулі адміністрування. Система реалізована у вигляді веб-застосунку, доступного через браузер, з розподілом прав

доступу для різних типів користувачів. Користувачами системи є власник, аналітик, адміністратор).

Розроблена система поєднує класичні підходи до аналізу клієнтської поведінки з сучасними інструментами візуалізації та побудови прогнозних моделей. Завдяки гнучкій архітектурі, відкритому коду та простому веб-інтерфейсу, її можна адаптувати до потреб різних сервісних закладів малого бізнесу. Система може бути використана як основа для подальших прикладних досліджень або впровадження в реальних умовах, де аналітика клієнтських транзакцій є ключовим елементом прийняття рішень. Перспективним напрямом удосконалення є розширення функціоналу, автоматизація генерації інсайтів, а також інтеграція з CRM-системами або платіжними модулями.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. MC.TODAY. Чому в Києві виник книжковий бум: все про книгарні, кав'ярні та з чим їх їдять. URL: <https://mc.today/uk/chomu-v-kiyevi-vinik-knizhkovij-bum-vse-pro-knigarni-kav-yarni-ta-z-chim-yih-yidyat/> (дата звернення 27.02.2025).
2. Андрійчук М. Т. Особливості діяльності київських книгарень у 2022–2024 роках: науковий журнал. Київ : КПП. 2024. С. 175-187.
3. Беспала О.М., Чубукова А.О. Аналіз транзакцій для оцінки впливу поведінкових факторів на продажі. *Сучасні проблеми наукового забезпечення енергетики*: матеріали XXII міжнародної науково-практичної конференції молодих вчених та студентів. Київ, 22-25 квітня 2025 року. Київ, 2025. С. 268-270.
4. Анастасія Федор. Ресторанний ринок: тренди 2025 року та очікування експертів. URL: <https://www.komersant.info/restorannyy-rynok-trendy-2025-roku-ta-ochikuvannia-ekspertiv/> (дата звернення: 20.04.2025).
5. Travis Cuzick. Power BI vs. Tableau: Comprehensive Comparison Guide. URL: <https://zerotomastery.io/blog/power-bi-vs-tableau-comparison-guide/> (дата звернення: 17.05.2025).
6. Thinklytics. Understanding the Cost Differences: Tableau vs. PowerBI in 2023. URL: <https://thinklytics.com/understanding-the-cost-differences-between-tableau-and-powerbi-2023/> (дата звернення: 17.05.2025).
7. Dom. Kissmetrics vs. Mixpanel: A Comprehensive Comparison for 2024. URL: <https://www.kissmetrics.io/blog/kissmetrics-vs-mixpanel-a-comprehensive-comparison-for-2024/> (дата звернення: 17.05.2025).
8. Streamlit Showcase Gallery. URL: <https://streamlit.io/gallery> (дата звернення: 18.03.2025).
9. Andriy Boychuk. RFM Analysis as an Effective Customer Segmentation in 2025. URL: <https://flowium.com/blog/rfm-analysis/> (дата звернення: 18.03.2025).
10. Rsquared Academy Blog. RFM Analysis in R. URL: <https://www.r-bloggers.com/2019/02/rfm-analysis-in-r/> (дата звернення: 17.05.2025).

11. Sorumlu Yazar. Customer Segmentation Using K-Means Clustering Algorithm and RFM Model. İzmir: DEÜ. 2023. p. 491-503.
12. Andrew Tate. When To Choose Density-Based Methods. URL: <https://hex.tech/blog/comparing-density-based-methods/> (дата звернення: 17.05.2025).
13. GeeksforGeeks. Decision Tree in R Programming. URL: <https://www.geeksforgeeks.org/decision-tree-in-r-programming/> (дата звернення: 17.05.2025).
14. Max Kuhn , Kjell Johnson. Applied Predictive Modeling. USA: Springer. 2013. p. 287-376.
15. GeeksforGeeks: Random Forest Algorithm – Explanation & Intuition. URL: <https://www.geeksforgeeks.org/random-forest-classifier-using-scikit-learn/> (дата звернення: 17.05.2025).
16. Nattapoj Apichardsilkij. Basic Comparison Between RandomForest, SVM, and XGBoost. URL: https://medium.com/@ap.nattapoj_st/basic-comparison-between-randomforest-svm-and-xgboost-0e5862871175 (дата звернення: 17.05.2025).
17. Zach Bobbitt. How to Build Random Forests in R (Step-by-Step). URL: <https://www.statology.org/random-forest-in-r/> (дата звернення: 18.05.2025)
18. Alexandra Twin. KPIs: What Are Key Performance Indicators? Types and Examples. URL: <https://www.investopedia.com/terms/k/kpi.asp> (дата звернення: 16.04.2025).
19. Vaibhav Tiwari. Understanding System Architecture: A Beginner's Guide. URL: <https://medium.com/@vaibhavtiwari.945/understanding-system-architecture-a-beginners-guide-d2aad48f0007> (дата звернення: 17.05.2025).
20. Brown D. S. Statistics and Data Visualization Using R: The Art and Practice of Data Analysis. USA: SAGE Publications. 2021. 616 p.
21. Wickham H., Cetinkaya-Rundel M., Grolemund G. R for Data Science: Import, Tidy, Transform, Visualize and Model Data. USA: O'Reilly Media. 2023. 576 p.
22. Appsilon. R Shiny: The Definitive Implementation Guide. URL: <https://www.appsilon.com/guides/r-shiny> (дата звернення: 17.05.2025).

23. Tutorialspoint. SQLite Tutorial. URL: <https://www.tutorialspoint.com/sqlite/index.htm> (дата звернення: 17.05.2025).

24. Visual Paradigm: What is Use Case Diagram? URL: <https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-use-case-diagram/> (дата звернення: 17.05.2025).

25. RStudio Support. How to deploy an app to shinyapps.io. URL: https://docs.posit.co/shinyapps.io/guide/getting_started/index.html (дата звернення: 16.04.2025).

ДОДАТОК А

Фрагменти коду основних складових застосунку

«Веб-додаток аналізу транзакцій в закладі обслуговування»

УКР.НТУУ «КПІ ім. Ігоря Сікорського»_ІАТЕ_ЦТЕ_ТР_15

Аркушів 3

Програмні засоби, використані у застосунку:

- мова програмування – R;
- фреймворк для створення веб-додатків – Shiny;
- побудова дашбордів – shinydashboard;
- мова запитів до бази даних – SQL (через пакет DBI та драйвер RSQLite);
- робота з таблицями – dplyr, tidyr, lubridate;
- візуалізація даних – ggplot2, corrplot, plotly, rpart.plot;
- побудова дерев рішень – rpart;
- кластеризація – kmeans, cluster;
- прогнозування відтоку клієнтів – randomForest, pdp;
- інтерфейс автентифікації – shinyauthr;
- робота з таблицями в інтерфейсі – DT;
- створення PDF-звітів – rmarkdown;

Реалізація підключення до бази даних та отримання транзакцій

```
library(readxl)
library(DBI)
library(RSQLite)
library(dplyr)

dir.create("database", showWarnings = FALSE)
con <- dbConnect(SQLite(), "database/bookstore_cafe.sqlite")

clients <- read_excel("C:/Users/Я/Documents/4 курс/диплом/довідник_клієнтів.xlsx")
products_raw <- read_excel("C:/Users/Я/Documents/4 курс/диплом/довідник_товарів.xlsx")
promotions <- read_excel("C:/Users/Я/Documents/4 курс/диплом/довідник_акцій.xlsx")
transactions <- read_excel("C:/Users/Я/Documents/4 курс/диплом/транзакції.xlsx")
products <- products_raw %>%
distinct(productID, .keep_all = TRUE)

con <- dbConnect(RSQLite::SQLite(), "database/bookstore_app.sqlite")

dbWriteTable(con, "clients", clients, overwrite = TRUE)
dbWriteTable(con, "products", products, overwrite = TRUE)
dbWriteTable(con, "promotions", promotions, overwrite = TRUE)
dbWriteTable(con, "transactions", transactions, overwrite = TRUE)

print(dbListTables(con))
dbDisconnect(con)
```

Реалізація RFM-сегментації та кластеризації клієнтів

```
observeEvent(input$run_clust, {
  rfm <- transactions %>%
    group_by(customerID) %>%
    summarise(
      Recency = as.numeric(Sys.Date() - max(purchaseDate)),
      Frequency = n_distinct(newCheckID),
      Monetary = sum(totalPrice, na.rm = TRUE),
```

```

    .groups = "drop"
  )

rfm_scaled <- scale(rfm[, c("Recency", "Frequency", "Monetary")])

set.seed(123)
km <- kmeans(rfm_scaled, centers = input$х_clusters)
rfm$х_cluster <- factor(km$х_cluster)

labels <- c("Постійні", "Нові", "Лояльні", "Разові", "Сплячі", "Цінні", "Інші")
rfm$х_cluster_label <- factor(
  rfm$х_cluster,
  levels = sort(unique(rfm$х_cluster)),
  labels = labels[1:length(unique(rfm$х_cluster))]
)

client_clusters(rfm)
sil <- silhouette(km$х_cluster, dist(rfm_scaled))
client_silhouette(list(avg = mean(sil[, "sil_width"]), obj = sil))

fit <- rpart(
  cluster ~ Recency + Frequency + Monetary,
  data = rfm,
  control = rpart.control(cp = 0.01, maxdepth = 3),
  model = TRUE
)
client_tree_model(fit)
})

```

Реалізація прогнозування відтоку клієнтів

```

observeEvent(input$х_run_churn_model, {
  req(client_clusters())

  df <- client_clusters() %>%
    filter(!is.na(Recency), !is.na(Frequency), !is.na(Monetary))

  set.seed(123)
  df$х_churn <- ifelse(df$Recency > quantile(df$Recency, 0.75), 1, 0)
  df$х_churn <- ifelse(runif(nrow(df)) < 0.05, 1 - df$х_churn, df$х_churn)

  set.seed(42)
  idx <- sample(1:nrow(df), size = 0.8 * nrow(df))
  train_df <- df[idx, ]
  test_df <- df[-idx, ]

  fit <- randomForest(
    as.factor(churn) ~ Recency + Frequency + Monetary,
    data = train_df,
    ntree = 100,
    importance = TRUE
  )
  rf_model(fit)

  probs <- predict(fit, test_df, type = "prob")[, 2]
  pred_labels <- ifelse(probs > 0.5, 1, 0)

  test_df$х_churn_prob <- round(probs, 3)
  test_df$х_churn_pred <- pred_labels
  rf_predictions(test_df)

  cm <- table(Факт = test_df$х_churn, Прогноз = test_df$х_churn_pred)
  acc <- mean(test_df$х_churn == test_df$х_churn_pred)

  output$х_churn_model_summary <- renderPrint({
    cat(" Confusion Matrix (Test Set):\n")
    print(cm)
    cat("\n Accuracy:", round(acc * 100, 2), "%")
    cat("\n\n ✨ Побудовано на", nrow(train_df), "тренувальних клієнтах")
  })

  showNotification("Прогноз завершено", type = "message")
})

```

Реалізація KPI-аналізу ефективності акцій

```

promo_metrics_all <- reactive({
  req(input$selected_promo)
  p <- promotions %>% filter(promoName == input$selected_promo)
  start <- as.Date(as.POSIXct(as.numeric(p$startDate), origin = "1970-01-01"))
  end <- as.Date(as.POSIXct(as.numeric(p$endDate), origin = "1970-01-01"))
  days <- as.integer(end - start) + 1

  get_m <- function(from, to) {
    df <- transactions %>% filter(purchaseDate >= from, purchaseDate <= to)
    data.frame(
      Period = "_",
      Revenue = sum(df$totalPrice, na.rm = TRUE),
      Customers = n_distinct(df$customerID),
      Avg_Check = if (nrow(df) > 0) round(mean(df$totalPrice), 2) else 0,
      Days = length(unique(df$purchaseDate)),
      stringsAsFactors = FALSE
    )
  }

  before <- get_m(start - days, start - 1); before$Period <- "До акції"
  during <- get_m(start, end); during$Period <- "Під час акції"
  after <- get_m(end + 1, end + days); after$Period <- "Після акції"

  do.call(rbind, list(before, during, after))
})

output$promo_plot <- renderPlot({
  p <- promotions %>% filter(promoName == input$selected_promo)
  req(nrow(p) == 1)
  start <- as.Date(as.POSIXct(as.numeric(p$startDate), origin = "1970-01-01"))
  end <- as.Date(as.POSIXct(as.numeric(p$endDate), origin = "1970-01-01"))
  days <- as.integer(end - start) + 1

  df_time <- transactions %>%
    filter(purchaseDate >= (start - days), purchaseDate <= (end + days)) %>%
    group_by(purchaseDate) %>%
    summarise(dailyRevenue = sum(totalPrice, na.rm = TRUE)) %>%
    mutate(period = case_when(
      purchaseDate < start ~ "До акції",
      purchaseDate >= start & purchaseDate <= end ~ "Під час акції",
      purchaseDate > end ~ "Після акції"
    ))

  ggplot(df_time, aes(purchaseDate, dailyRevenue, color = period)) +
    geom_line(size = 1.2) +
    geom_vline(xintercept = as.numeric(start), color = "#118ab2", linetype = "dashed") +
    geom_vline(xintercept = as.numeric(end), color = "#118ab2", linetype = "dashed") +
    scale_color_manual(values = c(
      "До акції" = "#118ab2",
      "Під час акції" = "#ffca3a",
      "Після акції" = "#f15bb5"
    )) +
    labs(x = "Дата", y = "Виручка", color = "Період") +
    theme_minimal()
})

output$period_summary <- renderDT({
  df <- promo_metrics_all()
  datatable(df, rownames = FALSE, options = list(pageLength = 3, dom = "t"))
})

output$promo_conclusion <- renderUI({
  m <- promo_metrics_all()
  delta_revenue <- round(100 * (m$Revenue[m$Period == "Під час акції"] / m$Revenue[m$Period == "До акції"] - 1), 1)
  delta_customers <- round(100 * (m$Customers[m$Period == "Під час акції"] / m$Customers[m$Period == "До акції"] - 1), 1)
})

```

ДОДАТОК Б

Апробація

Матеріали XXII Міжнародної науково-практичної конференції молодих вчених і студентів «Сучасні проблеми наукового забезпечення енергетики»

«Аналіз транзакцій для оцінки впливу поведінкових факторів на продажі»

УКР.НТУУ «КПІ ім. Ігоря Сікорського»_ІАТЕ_ЦТЕ_ТР_15

Аркушів 5

Київ — 2025

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

СУЧАСНІ ПРОБЛЕМИ НАУКОВОГО ЗАБЕЗПЕЧЕННЯ ЕНЕРГЕТИКИ

Матеріали XXII Міжнародної
науково-практичної конференції
молодих вчених і студентів
м. Київ, 22–25 квітня 2025 року

ТОМ 2



Київ- 2025

контролю.	243
<i>КАРКУШЕВСЬКИЙ В.Л., бакалаврант</i>	
<i>Керівник - доц., к.ф.-м.н. Тарнавський Ю.А.</i>	
Порівняння технологій захоплення мережевого трафіку для розробки системи глибокого аналізу пакетів.	245
<i>КОВАЛЬОВ О.О., бакалаврант</i>	
<i>Керівник - асист. Кардашов О.В.</i>	
Градентні методи для сегментації зображень з метою виявлення пожеж.	248
<i>КУЛИКОВСЬКИЙ В.Г., бакалаврант</i>	
<i>Керівник - доц., к.т.н. Кузьменко І.М.</i>	
Веб-застосунок для планування харчування та обліку калорій групи осіб.	250
<i>МАРТИНЧУК А.В., бакалаврант</i>	
<i>Керівник - доц., к.е.н. Селеда І.В.</i>	
Система ідентифікації музичної композиції за аудіофрагментом.	252
<i>ПЕСТЕНКОВ Д.О. студент 4 курс, гр. ТР-14</i>	
<i>Доцент, к.іст.н. н. Ошійко А.І.</i>	
Система моніторингу відеопотоків для виявлення потенційних загроз.	254
<i>РЕЗЕНКОВ Д.А. студент 4 курс, гр. ТР-14</i>	
<i>Доцент, к.е.н. Селеда І.В.</i>	
Дистанційне керування пристроями IoT на базі wifi технології.	256
<i>РУДЕНКО В.І., бакалаврант; ІМАСВА Т.А., бакалаврант</i>	
<i>Керівник - проф., д.т.н. Отрох С.І.</i>	
Автоматизоване робоче місце фахівця АЕС з дозиметрії та захисту від іонізуючого випромінювання.	258
<i>САПОЖНИКОВ О.Ю., бакалаврант</i>	
<i>Керівник - доц., к.ф.-м.н. Тарнавський Ю.А.</i>	
Інтерактивна веб-платформа пропозицій спортивних товарів.	260
<i>СКОПЦОВА В.О. студентка 4 курс, гр. ТР-14</i>	
<i>Доцент, к.е.н. Селеда І.В.</i>	
Тренування нейронної мережі для гри "камінь-ножиці-папір".	262
<i>СТЕЦЬ А.О., бакалаврант</i>	
<i>Керівник - доц., к.т.н. Кублій Л.І.</i>	
Графічний калькулятор побудови логічних схем.	265
<i>ТУРЛАК С.С., бакалаврант</i>	
<i>Керівник - доц., к.т.н. Кублій Л.І.</i>	
Аналіз транзакцій для оцінки впливу поведінкових факторів на продажі.	268
<i>ЧУБУКОВА А.О., бакалаврант</i>	
<i>Керівник - ст.викл. Беспіла О.М.</i>	
Fuzzy Logic and Computer Vision-Based Quadcopter Landing Control Algorithm.	271
<i>HLAVATSKYI O.S., MS's programme</i>	
<i>Academic leader - prof., doc.eng.sc. Shushura O.M.</i>	

УДК 004.6

¹ Бакалаврант 4 курсу Чубукова А.О.² Ст. викл. Беспала О.М.<https://scholar.google.com.ua/citations?user=gLUrLHUAAAAAJ&hl=uk>³ КПІ ім. Ігоря Сікорського

АНАЛІЗ ТРАНЗАКЦІЙ ДЛЯ ОЦІНКИ ВПЛИВУ ПОВЕДІНКОВИХ ФАКТОРІВ НА ПРОДАЖІ

Постановка проблеми та її актуальність. У сучасних умовах високої конкуренції підприємства сфери обслуговування стикаються з необхідністю аналізу факторів, що впливають на обсяг продажів [1]. Тому існує потреба у створенні системи аналізу, яка б дозволила оцінити поведінкові чинники та їх вплив на фінансові показники підприємства.

На практиці система може бути використана в різних предметних областях. Протягом останніх кількох років в Україні активно розвивається галузь книговидавництва та книготоргівлі, і як наслідок – відкриття великої кількості нових книгарень [2]. Також спостерігається тенденція перетворення книгарень на книгарні-кав'ярні та мультиформатні культурні простори [3]. В свою чергу, кáva часто складає доволі високий відсоток покупок [4]. Однак на даний момент підхід до моделювання виручки книгарні-кав'ярні з урахуванням параметрів, пов'язаних із продажами кави, все ще вимагає подальшого вдосконалення. Створення системи моделювання виручки буде дуже корисним, оскільки вона дозволить точно оцінювати вплив продажів кавових напоїв на виручку закладу.

Аналіз останніх досліджень. Сьогодні цифрові технології є одним з найважливіших чинників економічного розвитку, оскільки дозволяють швидко збирати та аналізувати великі обсяги інформації. Останні дослідження показують, що застосування штучного інтелекту, Інтернету речей і хмарних сервісів допомагає зменшити витрати, підвищити якість послуг та створити нові, інноваційні бізнес-моделі. Інтеграція сучасних технологій дозволяє підприємствам не тільки краще аналізувати фінансові показники, а й прогнозувати їх зміни [5].

Формулювання мети. Метою цієї роботи є створення системи аналізу транзакційних даних для оцінки поведінкових факторів та їх впливу на обсяг продажів в закладах обслуговування.

Основна частина. Основна ідея розробленої системи полягає у створенні інтерактивного веб-додатку, який дозволяє моделювати збільшення виручки як залежну змінну. Система реалізована за допомогою R Shiny, що забезпечує зручний графічний інтерфейс та доступність через веб-браузер.

У даній роботі розглядається система, що містить набір параметрів N_r , що представлені у вигляді набору згенерованих даних. Ці параметри охоплюють такі характеристики операцій, як дата та час покупки, ідентифікатор клієнта, інформацію про заклад, категорії товарів та фінансові показники (ціна, кількість, загальна сума чеку). Приклад набору даних представлений в Таблиці 1.

Задача полягає в тому, щоб визначити зв'язок між цими параметрами та загальними фінансовими результатами, що дозволить оцінити вплив поведінкових факторів (частота покупок, тощо) на обсяг продажів.

Для обробки транзакційних даних використано бібліотеки R `readxl` і `dplyr`, які дозволяють імпортувати дані з Excel-файлів, проводити очищення, групування та узагальнення даних за релевантними параметрами, такими як: дата покупки та назва закладу [6].

Таблиця 1 - Приклад набору даних

ID	date	customer ID	bookstore ID	category2	price	amount	total Price
9	2025-01-05	466	1	Класична література	310	2	620
10	2025-01-06	95	2	Комікси та манги	345	2	690
11	2025-01-06	444	1	Кавові напої	70	1	70
12	2025-01-06	794	1	Класична література	240	1	240
12	2025-01-06	794	1	Детективи	380	1	380
12	2025-01-06	794	1	Трилери та жаки	750	1	750
13	2025-01-06	580	2	Історична та пригодницька проза	525	3	1575
14	2025-01-06	110	2	Історія, факти та біографії	570	1	570
15	2025-02-04	113	2	Кавові напої	65	1	65

На основі агрегованих даних застосовується економіко-статистичний підхід із використанням лінійної регресії для оцінки впливу продажів кави на фінансові результати закладів [7]. Сценарне моделювання дає можливість змоделювати різні умови – наприклад, підвищення продажів певної категорії товарів – і прогнозувати відповідні зміни загальних фінансових показників. Для візуалізації результатів використовується бібліотека `ggplot2`, що дозволяє наочно представити виявлені залежності через графіки та діаграми [8]. Система використовує MySQL для зберігання та швидкого доступу до даних.

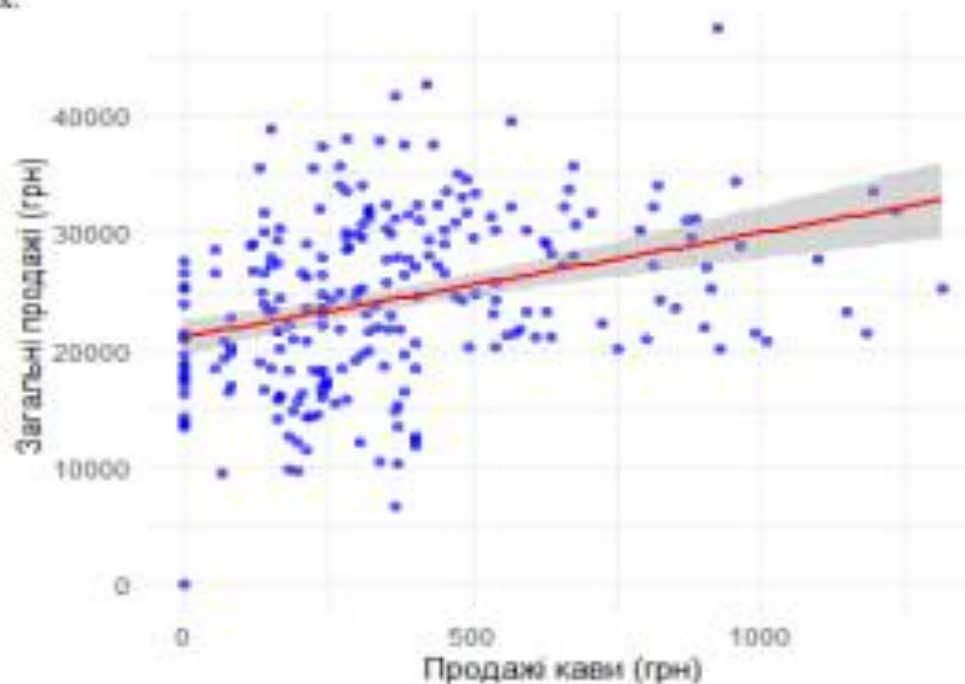


Рисунок 2 - Вплив продажу кави на загальні продажі закладу

Модель регресії, задана функцією:

$$Total_Sales = 21083,5 + 8,928 \times Coffee_Sales$$

демонструє статистично значущий вплив продажів кави на загальні продажі. Інтерсепт, що дорівнює 21083,5, відображає базовий рівень виручки, а коефіцієнт при змінній *Coffee_Sales* (8,928) свідчить, що кожна додаткова одиниця продажів кави збільшує загальні продажі приблизно на 9 гривень. Висока t-статистика (5,377) та дуже малий p-value (1,81e-07) підтверджують надійність оцінок. Хоча Multiple R-squared (0,1087) показує, що продажі кави пояснюють лише близько 11% варіації загальних продажів, це відкриває можливості для подальшого розширення моделі шляхом включення додаткових факторів, що впливають на виручку.

Результати. Отримані показники свідчать про наявність статистично значущого позитивного зв'язку між продажами кави та загальними продажами. Коефіцієнти моделей показують, що збільшення обсягів продажів кавових напоїв веде до зростання загальної виручки. Графіки підтвердили наявність чіткої тенденції до зростання фінансових показників при збільшенні продажів кави.

Висновки. Таким чином, в результаті аналізу було встановлено, що продаж кавових напоїв є важливим чинником, який суттєво впливає на загальний обсяг продажів кав'ярні-кав'ярні. Лінійна регресія дозволила оцінити цей вплив, що обернулося впровадження спеціальних маркетингових заходів – наприклад, проведення акцій та оптимізацію асортименту. Практичні рекомендації, розроблені на основі отриманих результатів, можуть сприяти підвищенню фінансової ефективності підприємства, а подальше покращення моделі дозволить врахувати додаткові фактори для ще точнішої оцінки.

Перелік посилань

1. Smith J., Doe A. Digital Technologies and Innovation: A New Paradigm // Technological Forecasting and Social Change. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0278431917300646> (дата звернення 27.02.2025).
2. MC.TODAY. Чому в Києві виник книжковий бум: все про книгарні, кав'ярні та з чим їх їдять. URL: <https://mc.today/uk/chomu-v-kyievi-vinik-knizhkovij-bum-vse-pro-knigarni-kav-yarni-ta-z-chim-ih-vidyat/> (дата звернення 27.02.2025).
3. Аядрішчук, М. Т. Особливості діяльності київських книгарень у 2022–2024 роках: науковий журнал. Київ : КПІ. 2024. С. 175-187.
4. Феліксон, Д. Скільки треба грошей, щоб відкрити книгарню-кав'ярню: досвід Дмитра Феліксона, засновника мережі REadeAT. URL: <https://mc.today/uk/blog/skylki-treba-groshej-shhob-vidkriti-knigarnyu-kav-yarnyu-dosvid-dmitra-feliksova-zasnovnika-merezhi-readeat/> (дата звернення 27.02.2025).
5. Ткаченко С. М. Цифровізація економіки як стратегічна платформа розвитку держави : тези доповідей. Запоріжжя : Запорізький національний університет. 2024. С. 153 - 160.
6. Wickham H., Cetinkaya-Rundel M., Grolemund G. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. USA: O'Reilly Media. 2023. 576 p.
7. Hoffmann J. P. Linear Regression Models: Applications in R. UK: Chapman and Hall/CRC. 2021. 420 p.
8. Brown D. S. Statistics and Data Visualization Using R: The Art and Practice of Data Analysis. USA: SAGE Publications. 2021. 616 p.