

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ТАРАСА ШЕВЧЕНКА  
Факультет інформаційних технологій  
Кафедра інтелектуальних технологій**

**КВАЛІФІКАЦІЙНА РОБОТА  
на здобуття освітнього ступеня «магістр»**

НА ТЕМУ:

**Інструментальні засоби бізнес-аналізу  
діяльності комерційної організації**

Галузь знань: 12 «Інформаційні технології»

Спеціальність: 122 «Комп'ютерні науки»

Освітньо-наукова програма «Технології штучного інтелекту»

Виконав:

студент 2 курсу магістратури групи ТШП-21

Красновідов Сергій Володимирович

Науковий керівник:

Сорока Петро Миколайович

кандидат фізико-математичних наук, доцент

(науковий ступінь, вчене звання)

Засвідчую, що в цій кваліфікаційній роботі  
немає запозичень з праць інших авторів без  
відповідних посилань

Студент \_\_\_\_\_

Кваліфікаційна робота допущена до захисту  
рішенням кафедри *інтелектуальних технологій*

Протокол № \_\_\_\_ від «\_\_\_\_» травня 2021 р.

Зав. кафедри \_\_\_\_\_ доц. Іларіонов О.Є.

**Київ 2021**

## РЕФЕРАТ

Кваліфікаційна робота складається зі вступу, 3 розділів, висновків, списку використаної літератури із 59 джерел. Загальний обсяг роботи 100 сторінок. Робота містить 13 таблиць та 31 рисунок.

**Актуальність теми.** Підприємства можуть не працювати на повну потужність за різних обставин та економічної ситуації, тому для кожної комерційної організації завжди актуально підтримувати конкурентоспроможність, аналізувати поточну діяльність та прогнозувати майбутню. Досягнути цього можна завдяки інформаційним технологіям бізнес-аналітики, які дозволяють керівникам приймати рішення за суттєвої економії часу на пошук і аналіз необхідної інформації на основі вже існуючих даних про організацію.

**Метою кваліфікаційної роботи** є аналітичний огляд та консолідація поширених та ефективних інформаційних технологій бізнес-аналітики та їх використання в аналізі діяльності комерційної організації, розробка сценаріїв для проведення бізнес-аналізу в цілому.

**Об'єктом дослідження** є комерційна організація Будмаркет.

**Предметом дослідження** є оцінювання та аналіз діяльності комерційної організації.

**Результати роботи.** У ході роботи було розглянуто аналітично-інформаційні системи бізнес-аналізу, їхні переваги та недоліки, а також описано технології та методи, які вони використовують. На основі однієї з платформ був проведений аналіз діяльності магазину будівельних та побутових матеріалів у формі сегментування клієнтів, виділення шаблонів їхньої поведінки та наборів популярних товарів, аналізу продаж у розрізі днів, тижнів та місяців. Також було побудовано прогнозну модель щоденних продаж. На основі результатів аналізу було сформовано рекомендації організації та зроблено прогноз продажів на наступні декілька днів.

**Апробація результатів кваліфікаційної роботи.** Основні положення кваліфікаційної роботи були представлені на VII Міжнародній науково-практичній конференції «Інформаційні технології та взаємодії» (м. Київ, 4 грудня 2020 року). За матеріалами кваліфікаційної роботи було опубліковано 1 роботу в збірнику матеріалів цієї конференції.

**Ключові слова:** комерційна організація, бізнес-аналіз, аналітична платформа, прогнозування продаж, аналіз ринкового кошику, data mining, ARIMA.

## ABSTRACT

A qualification work consists of an introduction, 3 chapters, conclusions, a list of references (59 sources). The total volume of work is 100 pages. The work contains 13 tables and 31 figures.

**Actuality of the theme.** Businesses may not operate at full capacity under different circumstances and economic situations, therefore it is always important for every commercial organization to maintain competitiveness, analyze its current and forecast its future activities. This can be achieved through business intelligence information technologies, which allows managers to make significant time-saving decisions to search and analyze the necessary information based on existing data about the organization.

**The purpose of the qualification work** is to analytically review and consolidate common and effective information technologies of business analytics and their use in the analysis of the commercial organization, and to develop scenarios for business analysis in general.

**The object of the research** is commercial organization Budmarket.

**The subject of the research** is the evaluation and analysis of the activities of a commercial organization.

**Results of the work.** The analytical-information systems of business analysis, their advantages and disadvantages, as well as the technologies and methods they use were considered and described. Using one of the systems, an analysis of the store of construction and household materials was done in the form of customer segmentation, selection patterns of their behavior and sets of popular products, sales analysis in terms of days, weeks and months. Also, a forecast model of daily sales was designed. Based on the results of the analysis, the organization's recommendations were formed and its sales were forecast for the next few days.

**Publications.** Some of the results of the qualification work were presented as thesis at VII International conference «Information Technology and Interactions» (Kyiv, December 04, 2020). Based on the materials of the qualification work, the scientific paper was published in the materials of this conference.

**Keywords:** commercial organization, business analysis, business analytics, analytical platform, sales forecasting, market basket analysis, data mining, ARIMA.

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	7
ВСТУП.....	8
РОЗДІЛ 1. БІЗНЕС-АНАЛІТИКА ТА КОМЕРЦІЙНА ДІЯЛЬНІСТЬ .....	10
1.1. Сутність бізнес-аналітики .....	10
1.2. Технології бізнес-аналітики.....	12
1.3. Підходи до аналізу діяльності комерційної організації .....	22
РОЗДІЛ 2. АНАЛІТИЧНИЙ ОГЛЯД ІСНУЮЧИХ РІШЕНЬ БІЗНЕС-АНАЛІТИКИ .....	26
2.1. Аналіз існуючих рішень бізнес-аналітики .....	26
2.2. Основні характеристики розглянутих рішень бізнес-аналітики.....	39
2.3. Математичний апарат інформаційних технологій бізнес-аналітики..	45
РОЗДІЛ 3. АНАЛІЗ ДІЯЛЬНОСТІ КОМЕРЦІЙНОЇ ОРГАНІЗАЦІЇ.....	72
3.1. Підготовка даних та аналітична звітність .....	72
3.2. Аналіз та прогнозування продаж магазину будівельних матеріалів ..	74
3.3. Аналіз ринкового кошика організації.....	89
ВИСНОВКИ .....	94
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	96

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

BI	Business Intelligence
БА	бізнес-аналітика
ІС	інформаційна(-і) система(-и)
БД	база даних
ІІ	штучний інтелект
KD	Knowledge Discovery, пошук знань
KDD	Knowledge Discovery in Databases, пошук знань у базах даних
ETL	Data Extracting, Transforming and Loading, видобуток, перетворення та завантаження даних
EIS (KIC)	Enterprise Information Systems (корпоративна інформаційна система)
CRM	Customer Relationship Management, система управління відносинами з клієнтами
ERP	Enterprise Resource Planning, система планування ресурсів підприємства
SCM	Supply Chain Management, системи управління ланцюгом поставок
OLAP	Online Analytical Processing, аналітична онлайн обробка
ACF	AutoCorrelation Function, функція автокореляції
SES	просте експоненціальне згладжування
AR	AutoRegressive model, модель авторегресії
MA	Moving Average model, модель ковзного середнього
ARIMA	AutoRegressive Integrated Moving Average, модель авторегресійного інтегрованого ковзного середнього

## ВСТУП

Сьогоднішня економічна ситуація, що склалася, зважаючи на світову кризу, не дозволяє великим та малим підприємствам функціонувати на повну потужність. У таких умовах для кожної комерційної організації є актуальним підтримання рентабельності та конкурентоспроможності. З цією задачею допомагають впоратися системи бізнес-аналітики, які дозволяють керівникам приймати обґрунтовані управлінські рішення за суттєвої економії часу на пошук і аналіз необхідної інформації [29].

Аналітика охоплює нові галузеві практики, такі як видобуток даних, аналіз великих даних, машинне навчання, ШІ та прогнозована аналітика. Це термін, який використовується для опису технічних аспектів аналітики, які мають передбачувальні можливості і можуть бути використані для вирішення бізнес-проблем.

Аналітика може перетворити необроблені дані у велику колекцію інформації, яка класифікує дані для виявлення та аналізу поведінкових даних та моделей. Організації можуть використовувати цю інформацію для аналізу поточного стану своєї діяльності та прогнозування майбутньої поведінки та тенденцій, задаючи питання "що-якщо". Крім того, аналітика може допомогти у виявленні шахрайства та часто замінює громіздкі ручні процеси.

Основною метою аналітичних інформаційних систем є забезпечення швидкого доступу до даних, виконання аналізу даних та інформаційна підтримка процесу прийняття рішень [1].

Призначення бізнес-аналітики (Business Intelligence, BI) – видобути знання про бізнес з даних, використовуючи різні апаратно-програмні технології. Такі технології дають можливість організаціям перетворювати дані в інформацію, а потім інформацію в знання [30].

*Метою цієї кваліфікаційної роботи є аналітичний огляд та консолідація поширених та ефективних інформаційних технологій бізнес-аналітики та їх*

використання в аналізі діяльності комерційної організації, розробка сценаріїв для проведення бізнес-аналізу в цілому.

Для досягнення мети необхідно розв'язати *наступні задачі*:

- розглянути та опанувати основні концепції бізнес-аналізу;
- проаналізувати сучасний ринок систем та технологій бізнес-аналізу;
- вибрати технології та, за можливості, аналітичну платформу чи систему, яка враховує специфіку галузі комерційної організації, діяльність якої аналізується;
- розробити сценарії для проведення бізнес-аналізу комерційної організації.

*Об'єкт дослідження* роботи – комерційна організація Будмаркет.

*Предметом дослідження* є оцінювання та аналіз її діяльності.

## РОЗДІЛ 1. БІЗНЕС-АНАЛІТИКА ТА КОМЕРЦІЙНА ДІЯЛЬНІСТЬ

### 1.1. Сутність бізнес-аналітики

Сучасні умови ведення бізнесу, що характеризуються зростаючою жорсткою конкуренцією і нестабільністю економічних умов, пред'являють підвищені вимоги до оперативності і якості рішень, що приймаються, на всіх рівнях управління підприємством. Підтримка ухвалення рішень припускає володіння актуальною всеосяжною інформацією про стан і тенденції розвитку бізнесу методами і засобами Business Intelligence. При цьому обсяг інформації, яку необхідно враховувати для формування оптимальних обґрунтованих рішень, неухильно росте. Це призводить до ситуації, коли стає неможливо ефективно управляти підприємством без використання сучасних засобів інформаційного забезпечення, а саме методів і засобів бізнес-аналітики.

На сьогодні, економічній діяльності рівня підприємств притаманні такі ознаки, як: критичний рівень конкуренції на ринку і підвищена чутливість підприємств до впливу економічних умов. У кінцевому результаті – нестабільність стану на ринку, яка вимагає оперативних і високоякісних рішень менеджерів на всіх рівнях управління підприємством [2]. Основою для прийняття управлінських рішень є повна інформація про стан бізнесу і тенденцій розвитку ринку, а управління такою інформацією – складний процес, що вимагає комплексного підходу.

Для того, щоб розглядати бізнес-аналітику як комплексне поняття, потрібно спочатку визначити сам термін «аналіз». «Аналіз (грец. – розпад, розчленування, розбір) – логічний прийом, метод дослідження, який полягає в тому, що досліджуваний предмет подумки розділяється на складові елементи, які потім досліджуються окремо як частина розподіленого цілого, для того, щоб виділені в ході аналізу елементи з'єдналися за допомогою іншого логічного прийому – синтезу – в ціле, збагачене новими знаннями» [3].

«Розглядаючи аналітику як комплексну наукову дисципліну, можна використати принципово нові, урізноманітненні підходи при аналізі інформації

в контексті підготовки управлінських рішень. Зауважимо, що процес аналізу називають аналізуванням, а спосіб проведення аналізу – аналітичним методом. Розглядаючи аналітику в контексті бізнес-аналізу, наведені поняття характеризуватимуть спосіб ведення бізнесу, прийняття управлінських рішень, налагодження системи підтримки прийняття рішень в предметній галузі, шляхом детального аналізу всіх ключових елементів об'єкта аналізу та їх когнітивних зв'язків» [3]. Виділяють наступні процеси, які описують бізнес-аналітику як особливий вид діяльності в комерційній організації:

- обґрунтування інформаційно-аналітичних завдань і виділення цілей управління;
- збір інформації, формування знань та управління ними для розв'язання управлінських задач в умовах невизначеності;
- оцінка та аналіз зібраної інформації та знань в контексті завдань управління;
- побудова моделі предметної області та перевірка її адекватності, координація;
- проведення експериментів за участю побудованої моделі;
- аналіз результатів роботи моделі, синтез нових знань для розв'язання управлінських задач.

Діяльність підприємства, організації або системи відображається в інформаційний простір. Підготовка прийняття рішень і, відповідно, аналіз відбуваються в цьому просторі як середовищі – з одного боку і проводяться операції з елементами структури цього простору і сполучених з ним інформаційних просторів . Аналітик та інші особи, пов'язані з інформаційними процесами використовують у своїй роботі поняття, що визначають сутність, структуру, елементи інформаційного простору, особливо при використанні сучасних інформаційних технологій [4]. Бізнес-аналітика наголошує на тому, що процес організації і аналізу вже існуючих даних для прийняття найінформативніших бізнес-рішень є одним з найважчих і основних завдань

менеджменту. Управлінці, які усвідомлюють, що стратегічні аспекти управління сучасним автоматизованим, комп'ютеризованим, інформатизованим підприємством, інтегрованим із зовнішнім середовищем, потребує іншої теоретичної бази порівняно, наприклад, з принципами управлінського раціоналізму, знайдуть додаткові можливості вдосконалення власної діяльності в концепції всезагальної організаційної науки.

Бізнес-аналітика – це розширена форма аналізу даних або різновид застосування даних аналітиком, включаючи аналіз великих даних у бізнесі. Аналіз даних також часто визначається як етап вивчення, узагальнення та отримання висновків з даних [5]. Це визначення також підкреслює три функції аналізу даних: вивчення даних, узагальнення даних та висновок з даних. Наприклад, менеджер з маркетингу може використовувати аналіз даних для вивчення маркетингової ситуації на основі інформації про продажі за останній квартал та прогнозувати тенденцію збуту в найближчі півроку або рік. Лікар зазвичай використовує аналіз даних для діагностики захворювання пацієнта, щоб призначити для нього ліки [31].

## 1.2. Технології бізнес-аналітики

Бізнес-аналітика та аналіз даних в цілому сприяють розвитку корпоративних інформаційних систем (KIC), або Enterprise information systems (EIS). Це, перш за все, зумовлено делегацією управління цінними даними, інформацією та знаннями, а також підтримки прийняття бізнес-рішень такої системі [32].

Як в EIS, так і в аналітиці, дані розглядають як стратегічний ресурс для розвитку комерційних організацій. EIS зазвичай передбачає інтерактивну візуалізацію для дослідження та виявлення підмножин даних та закономірностей, які можуть розглядатися як частина аналізу великих даних. EIS включає аналітичні інструменти для використання даних, записаних системою, для оцінки ефективності бізнесу. Вони є частиною будь-якої системи бізнес-

аналітики. Це означає, що EIS та бізнес-аналітика базуються на однакових підходах та технологіях для підтримки прийняття бізнес-рішень та вдосконалення результатів діяльності підприємств [6]. Такі системи привертають все більшу увагу в наукових колах, організаціях та підприємствах останні кілька десятиліть. Існує багато різних визначень EIS. Розглянемо три найпоширеніших визначення EIS:

1. Системи, які допомагають менеджерам та компаніям покращувати свою ефективність і завдяки яким можна легко обмінюватися даними та інформацією між департаментами та зовнішніми структурами. Ці системи інтегрують такі функціональні системи, як бухгалтерський облік, фінанси та маркетинг, а також інші операції.
2. Комерційне програмне забезпечення на основі набору інтегрованих програмних модулів та централізованого сховища даних.
3. Інформаційні системи, що підтримують діяльність у декількох підрозділів підприємства.

Перше визначення є самостійним для EIS і наголошує на обміні даними та інформацією. Друге наголошує на корпоративному програмному забезпеченні із централізованим сховищем даних. Третє – найбільш загальне визначення. Поєднуючи ці три визначення, ми визначаємо EIS як інформаційну систему, що має спільний централізований доступ до сховища даних організації та підтримує діяльність кількох підрозділів, завдяки інтеграції функціональних інформаційних підсистем (IC), таких як бухгалтерський облік, фінанси та маркетинг, а також доступ до Інтернет-ресурсів. Вони допомагають організаціям покращити ефективність ведення бізнесу та прийняття рішень [7], надаючи можливість безперебійного обміну даними та інформацією між підрозділами та клієнтами.

EIS в основному складаються з ERP (планування ресурсів підприємства), SCM (ланцюг поставок) систем управління, CRM (управління взаємовідносинами із клієнтами) системи та системи управління знаннями (KM).

Основні постачальники великих корпоративних EIS, наприклад, SAP Business Suite, Oracle e-Business Suite та Microsoft Dynamics Suite надають такі рішення, що дозволяють інтегрувати ці EIS між собою і пов'язувати їх з клієнтськими системами на основі веб-сервісів, такі як Amazon Web Services, хмарних сервісів, наприклад, Microsoft Azure, та орієнтованих на сервіс архітектур [33].

Як було зазначено вище, EIS засновані на підходах і технологіях бізнес-аналітики та аналізу даних. Розглянемо детальніше їх нижче.

### 1.2.1. KDD

Knowledge Discovery in Databases (KDD) – це нетривіальний процес ідентифікації дійсних, нових, потенційно корисних і зрозумілих закономірностей у даних у вигляді знань. Цими знаннями можуть бути класифікації (нейронні мережі) та кластери даних (мапи Кохонена), шаблони або моделі даних (асоціативні правила) та зв'язки між атрибутами даних (дерева рішень).

KDD можна розглядати як інтерактивний та ітеративний процес, включаючи кроки з багатьма рішеннями, прийнятими безпосередньо користувачем [34]:

1. Розуміння предметної області діяльності, збір попередніх знань та визначення мети процесу KDD.
2. Створення цільового набору даних.  
Вибір набору даних або зосередження уваги на підмножині змінних чи зразках даних (sampling).
3. Очищення та попередня обробка даних.  
Основні операції включають видалення шуму, якщо необхідно, збираючи необхідну інформацію для моделювання або врахування шуму, приймаючи рішення обробки відсутніх полів даних, та облік інформації про часові послідовності та відомі зміни.
4. Проектування даних.

Пошук корисних функцій для представлення даних залежно від мети завдання. За допомогою методів зменшення або перетворення розмірностей даних, ефективна кількість розглянутих змінних може бути зменшена або може бути інваріантним поданням знайдених даних.

5. Відповідність мети KDD (крок 1) конкретному методу Data Mining. Наприклад, узагальнення, класифікація, регресія, кластеризація тощо.

6. Дослідницький аналіз, вибір моделі та гіпотез.

Вибір алгоритму(-ів) обробки даних та вибір методу(-ів) для пошуку шаблонів даних. Цей крок включає вибір моделей і параметрів, що можуть бути доречними (наприклад, моделі категоріальних даних відрізняються від моделей векторних даних) і узгодження методу видобутку даних із загальними критеріями процесу KDD (наприклад, кінцевий користувач може бути більше зацікавлений у розумінні самої моделі, ніж у її прогнозувальних можливостях).

7. Data Mining.

Пошук моделей у конкретній формі або наборі таких форм, включаючи правила класифікації або дерева, регресію та кластеризацію. Користувач може значно допомогти цьому методу, правильно виконавши попередні кроки.

8. Інтерпретація видобутих шаблонів.

Можливе повернення до будь-якого з кроків 1-7 для наступних ітерацій. Цей крок також може включати візуалізацію отриманих шаблонів та моделей або візуалізацію даних, наданих цими моделями.

9. Безпосереднє використання отриманих знань, включаючи їх у іншу систему або просто звітування.

Цей процес також включає перевірку та вирішення потенційних конфліктів із раніше видобутими знаннями.

Скорочений процес KDD зображений на рис. 1.1.

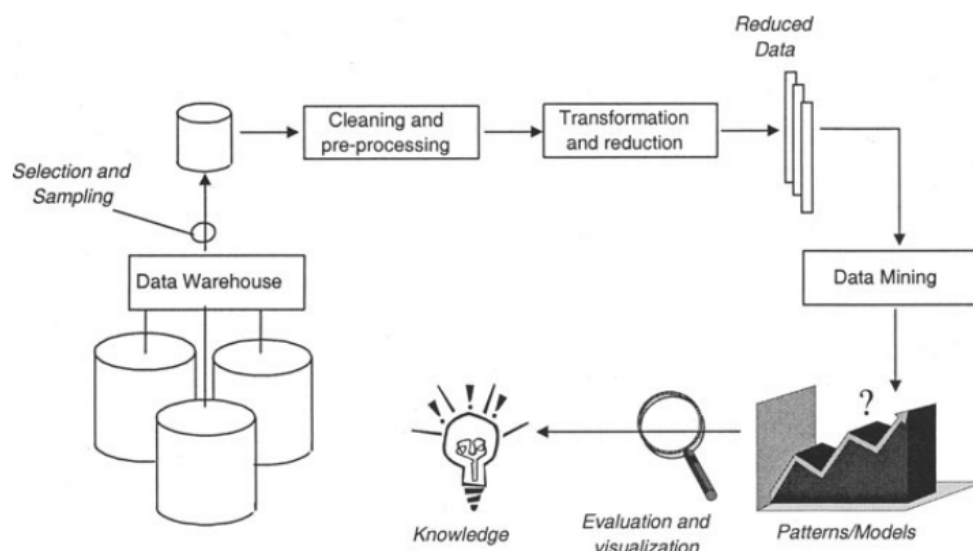


Рисунок 1.1 Схема-візуалізація скороченого процесу KDD [8]

Процес KDD може складатися із багатьох ітерацій і може містити цикли між будь-якими кроками. Більшість літературних джерел про KDD [1;8] зосереджені на кроці 7, тобто Data Mining. Однак інші кроки є не менш важливими для успішного застосування KDD на практиці.

### 1.2.2. Data Mining

Data Mining є міждисциплінарною підгалуззю інформатики. Це обчислювальний процес виявлення закономірностей у великих наборах даних, що включають методи на стику штучного інтелекту, машинного навчання, статистики та систем баз даних. Загальна мета аналізу даних полягає у видобуванні інформації з набору даних та перетворенні її в зрозумілу структуру для подальшого використання [9]. Окрім етапу первинного, «сирого» аналізу, він включає аспекти управління даними, їхню попередню обробку, розгляд моделі даних, метрики цікавості (interestingness), складності, після обробки виявлених структур, візуалізацію та оновлення даних. Data Mining – це аналітичний етап процесу «виявлення знань у базах даних», або KDD [34].

Цей термін часто неправильно розуміють, оскільки метою насправді є вилучення зразків та знань із великого обсягу даних, а не видобуток безпосередньо самих даних. Його також часто застосовують до будь-якої форми

обробки великих наборів даних або інформації (збір, вилучення, збереження, аналіз та статистика), а також до будь-якого застосування комп'ютерної системи підтримки прийняття рішень, включаючи штучний інтелект, машинне навчання та Business Intelligence [35].

Фактичним завданням Data Mining є автоматичний або напівавтоматичний аналіз великої кількості даних для виявлення раніше невідомих, цікавих записів, таких як групи деяких пов'язаних записів (кластерний аналіз), незвичні записи (виявлення аномалії) та залежні записи (видобуток правил асоціації). Це зазвичай передбачає використання таких методів баз даних, як просторові індекси. Потім ці закономірності можна розглядати як проміжний підсумок вхідних даних, що може бути використаний для подальшого аналізу або, наприклад, для машинного навчання та прогнозу аналітики. Ні збір даних, ні підготовка даних, ні інтерпретація результатів та звітування не є частиною етапу Data Mining, але належать до загального процесу KDD як додаткові етапи [36].

Видобуток даних складається з шести загальних класів завдань:

1. Виявлення аномалій/змін/відхилень – ідентифікація незвичних записів даних, які можуть бути цікавими, або помилки даних, які вимагають подальшого дослідження.
2. Правила асоціацій (моделювання залежностей) – пошук взаємозв'язків між змінними. Наприклад, супермаркет може збирати дані про звички покупців. За допомогою вивчення правил асоціацій супермаркет може визначити, які товари часто купують разом, і використовувати цю інформацію для маркетингових цілей. Це іноді називають аналізом ринкового кошика.
3. Кластеризація – це завдання виявлення груп і структур в даних, які тим чи іншим способом подібні, без використання відомих структур.
4. Класифікація – це завдання узагальнення відомої структури для застосування до нових даних. Наприклад, програма електронної пошти

може спробувати класифікувати повідомлення електронної пошти як «спам» або «реклама».

5. Регресія – спроба знайти функцію, яка моделює дані з найменшою помилкою.
6. Узагальнення забезпечує більш компактне представлення набору даних, включаючи візуалізацію та генерацію звітів.

### 1.2.3. OLAP-системи

OLAP (Online Analytical Processing) – це система аналітичної обробки даних. Вона призначена для підготовки звітів, побудови прогнозів і виконання статистичних розрахунків на базі великих інформаційних масивів, що мають складну структуру [10; 37].

OLAP системи складаються з:

1. бази даних (БД) – джерело, з якого береться інформаційний матеріал для обробки. Тип БД визначається різновидом OLAP системи і порядком виконання дій OLAP сервера. Найчастіше використовуються реляційні та багатовимірні БД і сховища даних;
2. OLAP сервер – ядро системи, за допомогою якого проводиться обробка багатовимірних структур даних, і забезпечується зв'язок між БД і користувачами систем;
3. Програмні додатки для роботи користувачів, у яких формуються запити і візуалізуються отримані відповіді.

Три основні операції в OLAP – це агрегація, деталізація та зрізи і кубікування.

1. Агрегація (roll-up), також відома як консолідація, або згортка, полягає у зменшенні вимірності набору даних та «підняття» рівня конкретних вимірів даних, наприклад, квартальний звіт усіх областей України перетворюється на річний звіт 10 економічних зон.

2. Деталізація (drill-down) – фрагментування великих наборів даних на менші набори, тобто «зменшення» рівня виміру даних та збільшення вимірності всього набору.
3. Зрізи і кубування (slice&dice) – представлення OLAP-куба у вигляді таблиці (slice) або під-кубу (dice).

Специфіка обробки даних OLAP системами полягає в побудові багатовимірних масивів інформації. Для формування таких масивів OLAP система збирає дані з різних джерел, наприклад, зі сховищ даних, з інформаційних систем управління підприємством (ERP) або з системи взаємодії з клієнтами (CRM). Після цього інформація обробляється на OLAP сервері і передається у програми користувача [11].

Зберігання та обробка даних із застосуванням OLAP систем можуть здійснюватися як безпосередньо на робочих місцях користувачів, у формі реляційних баз даних - при спільній роботі OLAP систем з ERP або CRM-системами, так і у формі багатовимірних баз даних на віддалених серверах у вигляді зіркових схем (star schemas).

Перевагами таких систем є багатовимірність моделі даних з великою кількістю проєкцій, «пакетність» видобування даних, що дозволяє ефективно підтримувати паралельний багатокористувацький доступ, а також обробка надлишкових даних з різних джерел.

OLAP також характеризується відносно низькою кількістю операцій. Запити часто дуже складні та включають агрегації. Для систем OLAP, час відгуку – це прямий показник ефективності: допустима затримка зазвичай становить від декількох секунд до декількох годин, залежачи майже цілком від обсягу даних [10].

Тому OLAP системи в основному застосовуються в сфері прийняття стратегічних управлінських рішень і використовуються для проектування бюджетів, побудови прогнозування, підготовки фінансової звітності, зберігання результатів.

#### 1.2.4. Нереляційні бази даних

Бази даних SQL, що базуються на реляційній моделі даних, не легко масштабуються для обробки зростаючих обсягів структурованих даних і мають обмежені можливості для обробки напівструктурованих та неструктурованих даних. Отже, існує потреба в альтернативних моделях зберігання даних.

Сукупності баз даних, побудовані навколо цих альтернативних моделей зберігання, стали називатися базами даних NoSQL, що може означати 'NotOnlySQL' або 'No, NeverSQL' залежно від моделі зберігання, що розглядається. Зараз доступно понад 150 різних баз даних NoSQL. Усі вони високопродуктивні і позбавлені традиційних обмежень реляційних баз даних, що позначається на їхній масштабованості та розподіленій обробці [14].

Основними категоріями баз даних NoSQL є сховища ключових значень, сховища документів, сховища розширюваних записів (або ширококолонкові записи) та бази даних графіків, хоча існує багато інших типів баз даних NoSQL. Зберігання ключ-значення - це місце, де дані можуть зберігатися без схеми, з відношенням «key-value», що складається з ключа, зазвичай рядка, і значення, яке є фактичними даними, що цікавлять. Саме значення може зберігатися за допомогою типу даних мови програмування або як об'єкт [38].

Сховище документів – це сховище ключ-значення, де значеннями є конкретно власні документи, такі як Microsoft Office (MS Word, MS Excel тощо), PDF, XML або подібні документи. Хоча кожен рядок таблиці в базі даних SQL матиме однакову послідовність стовпців, кожен документ може мати елементи даних, які абсолютно різні.

Як і бази даних SQL, розширювані сховища записів або широкі сховища колонок мають «таблиці» (так звані «супер сімейства колонок»), які містять колонки (так звані «супеколонки»). Однак кожна з них містить поєднання "атрибутів", подібних до сховищ ключ-значення. Найпоширенішими базами даних NoSQL, такими як Hadoop, є розширювані сховища записів.

Бази даних на графах складаються із взаємопов'язаних елементів з невизначеною кількістю взаємозв'язків і використовуються для зберігання даних, що представляють такі поняття, як соціальні відносини, зв'язки громадського транспорту, дорожні карти або топології мережі.

#### 1.2.5. Неструктуровані дані

Для того, щоб дані могли бути використані, їх потрібно проаналізувати, і для цього потрібен цілий новий комплекс складних методів, включаючи машинне навчання, обробку природної мови, прогнозне моделювання, нейронні мережі та картографування соціальних мереж. Поряд із цими методами знаходиться додатковий набір засобів візуалізації даних.

Неструктуровані дані або неструктурована інформація стосується інформації, яка не зберігається в електронній таблиці чи базі даних або не вміщується в розпізнаний шаблон чи модель. Він може бути текстовим або нетекстовим, що включає дати, номери, повідомлення електронної пошти, миттєві повідомлення тощо. Зазвичай він зберігається на паперових та електронних файлах, таких як документи Word, PowerPoint, електронна пошта, зображення, аудіо- та відео-файли, а дедалі частіше – у стрічках соціальних мереж. Неструктуровані дані важко зрозуміти за допомогою традиційних комп'ютерних програм, оскільки вони порівнюються з даними, що зберігаються у визначених полях таблиць баз даних або в тегах документів.

На неструктуровану інформацію припадає понад 70-80% усіх даних в організаціях, особливо з активним використанням соціальних медіа. Якщо не управляти такою інформацією, величезний обсяг неструктурованих даних, що генеруються щороку на підприємстві, може коштувати дорого з точки зору зберігання, потенційної відповідальності, доступу та неефективності, оскільки дані не можуть бути проаналізовані або просто не можуть передаватися між користувачами та системами.

Неструктуровані дані, що зберігаються в електронних файлах, можуть мати певну нав'язану структуру, принаймні для цілей подання, наприклад, імена файлів, імена папок та підпапок тощо. Призначені структури подання забезпечують певний ступінь управління та контролю над колекціями документів так само, як теги в HTML служать для обробки інформації в браузері, але не передають безпосередньо семантичного значення вмісту, позначеного тегами.

### 1.3. Підходи до аналізу діяльності комерційної організації

Аналіз діяльності комерційної організації можна поділити на логічні блоки, що описані нижче. Ці блоки використовують вище описані технології бізнес-аналітики, а також принципи фінансового менеджменту [12; 15]. Розглянемо їх детальніше.

*Аналіз цінності клієнта* дозволяє прогнозувати поведінку клієнта. Завдяки успіхам у сфері зберігання даних, цей напрям став збирати та використовувати ще більше додаткової інформації про клієнтів, отриманої, наприклад, від аналітичної платформи веб-сторінки комерційної організації за допомогою Google Analytics, соціальних мереж тощо. Водночас даний підхід спирається на класичну комбінацію зберігання даних, інтелектуального аналізу даних та прогнозу аналітики для створення моделей, які дозволяють тестувати комерційні теорії, пов'язані з поведінкою клієнта та впливом на неї.

*Моделювання ризиків.* Найчастіше кредитори, що використовують цей підхід для класифікації клієнтів, вибирають кредитні стратегії на основі оцінювання ризиків видачі та повернення кредиту тим чи іншим клієнтом. Моделювання кредитного ризику є частиною кредитної оцінки для таких продуктів, як кредитні картки та овердрафти, де модель використовується для встановлення кредитних лімітів. Моделі ризиків також можуть бути використані для розрахунку суми застави, необхідної для того, щоб гарантувати, що позика забезпечена активами [39].

*Дослідження ринку.* Дослідження ринку є одним із ключових факторів, що використовуються для підтримки конкурентоспроможності і надає важливу інформацію для виявлення та аналізу потреби ринку, його розміру та конкуренції. Методи дослідження ринку охоплюють такі якісні методи як фокус-групи, поглиблені інтерв'ю та етнографію, а також кількісні методи, такі як опитування клієнтів та аналіз вторинних даних [16].

Дослідження ринку, яке включає дослідження соціальних та громадських думок, є систематичним збором і інтерпретацією інформації про осіб чи організації для підтримки прийняття рішень. За допомогою інформації про ринок можна дізнатись про ситуацію попиту та пропозиції. Дослідники ринку відіграють більш широку роль, ніж раніше, допомагаючи своїм клієнтам зрозуміти соціальні, технічні та навіть юридичні аспекти ринків.

*Сегментація ринку* – це поділ ринку або населення на підгрупи з подібними мотиваціями. Він широко використовується для сегментування за географічними відмінностями, особистими відмінностями, демографічними відмінностями, технологічними відмінностями, використанням товарних відмінностей, психографічними відмінностями та гендерними відмінностями. Для сегментації B2B зазвичай використовується фірмографія. Ринкові тенденції - це рух ринку вниз або вгору вниз протягом певного періоду часу. Визначити розмір ринку може бути складніше, якщо починати з нової інновації. В цьому випадку, доводиться отримувати цифри за кількістю потенційних клієнтів або сегментів клієнтів.

*Візуалізація даних* у бізнес-аналітиці може використовуватися для візуального виявлення чітких закономірностей, кореляцій та тенденцій, які можуть бути не такими очевидними для програм автоматизованого аналізу даних та розпізнавання образів [11].

*Моделювання поведінки схильності клієнтів* дозволяє створити гіпотезу щодо ймовірності того, що потенційний клієнт придбає певний товар або послугу, тобто вони «схильні до покупки». Наприклад, що «вам сподобалось А,

тож вам може сподобатися В». Це досягається застосуванням узагальнень, отриманих на основі прогнозів тенденцій, Data Mining та грубо- (coarse-) і тонко-зернистого (fine-grained) аналізу.

Це стало класичною формою прогнозної аналітики. Моделювання схильності допомагає орієнтуватися на потрібних клієнтів, оптимізувати ресурси та значно збільшує ймовірність позитивної реакції рекламних кампаній, позбавляючи потреби загального загального підходу.

Цей тип моделювання схильності допомагає таким компаніям, як Netflix, рекомендувати телевізійні шоу та фільми на основі інтерпретації нашої схильності і минулої поведінки та аналізу подібної поведінки інших клієнтів. На цьому рівні це подібне на нечіткий аналіз. Він також може бути використаний для успішного перехресного продажу та перепродажу. Хоча без достатнього надходження адекватної, відповідної та своєчасної інформації запропоновані варіанти не завжди можуть бути релевантними [40].

Аналітика виявлення шахрайства існує вже давно і, як правило, спирається на традиційні засоби аналізу даних. Першими напрямками діяльності, які запровадили та застосовували виявлення шахрайства, були банки, телекомунікаційні та страхові компанії. Управління шахрайством – це діяльність, що потребує значних обсягів даних, процесів та знань. Методи аналітики виявлення шахрайства спираються на статистичний аналіз, та штучний інтелект.

Статистичні підходи до виявлення шахрайства включають попередню обробку даних (за допомогою зберігання даних); розрахунок статистичних обмежень та категорій; моделі розподілу ймовірностей; профілювання; аналіз часових рядів даних, що залежать від часу; кластеризація, класифікація, асоціація та відповідність.

III в аналітиці виявлення шахрайства включає аналіз даних; експертні системи; розпізнавання шаблонів (також є частиною видобутку даних); додаткові методи машинного навчання, включаючи навчання під наглядом та

без нагляду; та штучні нейронні мережі (наприклад, для виявлення відповідних, але поверхнево інтуїтивних моделей, які в іншому випадку не виявлялися б). Інші методи, що використовуються в аналітиці виявлення шахрайства, включають Байєса (для прогнозування ймовірності події на основі умов, які можуть бути пов'язані з цією подією); теорію рішень; узгодження послідовностей; інтерпретацію сигналу.

SWOT – це письмовий аналіз сильних і слабких сторін, можливостей та загроз бізнесу. SWOT повинен бути використаний не тільки на етапі створення компанії, але також може бути використаний протягом усього життєвого циклу компанії. SWOT також може бути записаний для участі в конкурсі зрозуміти, як розробляти маркетингові стратегії.

Доступність досліджень за допомогою Інтернету вплинула на величезну кількість споживачів використання цього носія інформації; для отримання знань, що стосуються практично кожного типу доступного товару та послуги. Зручність та доступність Інтернету створили глобальний дослідницький центр електронної комерції B2C для розгалуженої мережі Інтернет-магазинів, яка мотивує роздрібні ринки країн. У 2010 році цей засіб створив дохід від 400 до 600 мільярдів доларів, передбачалося, що в 2015 році цей онлайн-ринок принесе дохід від 700 до 950 мільярдів доларів [17]. Вплив маркетингових досліджень, незалежно від того, які форми він приймає, є надзвичайно потужний стимул для будь-якого типу споживачів та їх постачальників.

## РОЗДІЛ 2. АНАЛІТИЧНИЙ ОГЛЯД ІСНУЮЧИХ РІШЕНЬ БІЗНЕС-АНАЛІТИКИ

### 2.1. Аналіз існуючих рішень бізнес-аналітики

Інформаційні технології та системи бізнес-аналітики – це набір інструментів, що використовуються компаніями для отримання, аналізу та перетворення даних у корисну інформацію про їхній бізнес, або знання. Приклади інструментів бізнес-аналітики включають візуалізацію даних, зберігання даних, інформаційні панелі та звітність [17]. На відміну від конкурентної розвідки (competitive intelligence), технології бізнес-аналітики використовують внутрішні дані, які виробляє організація, а не зовнішні джерела.

По мірі того, як технологія та підхід Big Data набувала популярності, зросла популярність і технологій бізнес-аналітики в цілому. Компанії генерують, відстежують та компілюють дані у дійсно великих масштабах [41]. Але всі ці дані залишаються просто даними доти, доки ми не зможемо зрозуміти їх і використовувати для покращення ведення та результатів підприємницької діяльності.

Щоби зробити обґрунтований вибір того чи іншого рішення, підприємства повинні базувати їх на доказах або закономірностях. Великий об'єм даних, які виробляють підприємства та їхні клієнти, містять докази структури закупівель та ринкових тенденцій. Об'єднуючи, стандартизуючи та аналізуючи ці дані, підприємства можуть краще розуміти своїх клієнтів, краще прогнозувати зростання доходів та краще захищати себе від бізнес-пасток, невдач.

Бізнес-аналітика традиційно приймає форму щоквартальних або щорічних звітів, але сучасні технології бізнес-аналітики працюють постійно (online, real-time) та набагато швидше. Таким чином, компанії приймають рішення за лічені хвилини.

Програмне забезпечення бізнес-аналітики інтерпретує достатню кількісно вимірюваних дій підприємств і їхніх клієнтів та повертає запити на основі шаблонів у даних. Бізнес-аналітика має різні види і охоплює значну кількість

технологій, тому нижче наведено порівняння популярного програмного забезпечення для бізнес-аналітики та розглянуто три основні етапи, через які повинні пройти дані для надання такої аналітики.

### 2.1.1. Microsoft Power BI

Power BI – це загальна назва асортименту хмарних програм та служб, які допомагають організаціям збирати, управляти та аналізувати дані з різних джерел за допомогою зручного інтерфейсу.

Інструменти бізнес-аналітики, такі як Power BI, можна використовувати для багатьох цілей. У першу чергу Power BI збирає дані та обробляє їх, перетворюючи на зрозумілі представлення (view), часто використовуючи візуально переконливі та прості в обробці діаграми та графіки. Це дозволяє користувачам створювати та ділитися чіткими та корисними знімками того, що відбувається у їхньому бізнесі.

Power BI підключається до ряду джерел даних, починаючи від базових електронних таблиць Excel і закінчуючи базами даних, а також хмарними та локальними програмами.

Power BI – це загальний термін, який може стосуватися як настільної програми Windows під назвою Power BI Desktop, онлайн-сервісу SaaS (Програмне забезпечення як послуга) під назвою Power BI Service, так і мобільних програм Power BI, доступних на телефонах і планшетах Windows, а також для пристроїв iOS та Android. Power BI побудований на основі Microsoft Excel, і, таким чином, крива навчання від Excel до Power BI не така крута; кожен, хто може використовувати Excel, може використовувати Power BI, але останній є набагато потужнішим, ніж його електронна таблиця.

Переваги використання Power BI:

- Підприємства можуть вводити величезні обсяги даних у Power BI, які багато інших платформ намагалися б обробити.

- Вбудовані функції машинного навчання можуть аналізувати дані та допомагати користувачам виявляти цінні тенденції та робити прогнози з високою достовірністю.
- Інформацію можна візуалізувати за допомогою потужних шаблонів, щоб підприємства могли краще зрозуміти свої дані.
- Power BI базується на хмарі, тому користувачі отримують сучасні інтелектуальні технології та потужні алгоритми, які регулярно оновлюються.
- Потужні можливості персоналізації дозволяють користувачам створювати інформаційні панелі, щоб вони могли швидко отримати доступ до необхідних даних.
- На кожен показник ефективності користувача можна встановити сповіщення у разі будь-яких їхніх оновлень.
- Power BI має інтуїтивно зрозумілий інтерфейс, що робить його набагато привабливішим та зручнішим для навігації, ніж складні електронні таблиці
- Платформа інтегрується з іншими популярними інструментами управління бізнесом, такими як SharePoint, Office 365 та Dynamics 365, а також з іншими продуктами, що не належать Microsoft, такими як Spark, Hadoop, Google Analytics, SAP, Salesforce та MailChimp.
- Power BI забезпечує безпеку даних, пропонуючи детальний контроль над доступністю як зсередини, так і зовні.

### 2.1.2. SAP BusinessObjects

Програмний продукт SAP BusinessObjects Business Intelligence (BI) Suite від компанії SAP – це гнучка, масштабована платформа бізнес-аналітики, яка здійснює пошук та обмін інформацією для прийняття якісних бізнес-рішень. Інтегрована, уніфікована інфраструктура програмного продукту Забезпечує масштабованість від одного до багатьох інструментів та інтерфейсів на локальному рівні, у хмарах або в якості гібридного підходу.

Ключові можливості системи SAP BusinessObjects Business Intelligence (BI):

- *Звітність і аналіз.* Усі бізнес-користувачі можуть розуміти та оцінювати тенденції даних за допомогою інструментів бізнес-аналітики для нерегламентованих запитів, досліджень та аналізу веб-середовищ.
- *Візуалізація та аналітика даних.* Створення ефективної візуалізації BI, панелей моніторингу та аналітичних додатків для оцінки ризиків, підвищення ефективності, виділення унікальних даних, а також використання функцій загального доступу для обміну бізнес-інформацією з колегами та підтримкою прийняття рішень організації.
- *Інтеграція.* Підтримка Microsoft Office дозволяє використовувати інтуїтивно зрозумілі інструменти для фільтрації та обробки даних, визначення тенденцій та аномалій з Microsoft Excel та обміну результатами з онлайн презентаціями Microsoft PowerPoint. Можливість підключення SAP Business Warehouse і SAP HANA онлайн.

SAP BusinessObjects – найбільш зріла і функціональна система роботи зі звітами на підприємстві. Рішення BusinessObjects Business Intelligence оптимізують пошук і накопичення корпоративних відомостей, інтегрують інформацію з різних корпоративних джерел. Джерела інформації – різноманітні системи управління, бази даних і навіть електронні листи. Зібраний матеріал інформативно оформляється, що дозволяє оперативно аналізувати його і приймати стратегічні рішення.

Результати застосування SAP BusinessObjects впливають не тільки на технічні, а й на творчі процеси. Скорочуються фінансові та репутаційні ризики, адже людський фактор не впливає на процес збору корпоративної інформації. Професіоналам не доводиться витрачати час на отримання та перевірку необхідних даних і аналітики. Вони можуть повністю зосередитися на виконанні безпосередньої задачі. Керівникам зручніше оперативно оцінити ефективність

команди і прийняти стратегічне рішення, спираючись на інформативно представлені дані.

Всі продукти SAP BusinessObjects Business Intelligence мають ряд загальних функціональних характеристик, але кожен з них має свою специфіку. Наведемо деякі можливості модуля для корпоративних клієнтів:

- Crystal Reports створює звітні форми з точністю до «окремого пікселя». Ідеально підходить для друкованої продукції – типової звітності для контролюючих органів або внутрішньо-корпоративного використання.
- WebIntelligence, навпаки, створює корпоративну аналітику для перегляду на моніторі. Його особливість – легка інтеграція з MS Office.
- Design Studio допоможе в проведенні стратегічного аналізу. Додаток допомагає формувати інформаційні панелі, відомі як дашборди, з можливістю аналізу за принципом «що - якщо».
- Унікальне рішення Data Services інтегрується з усіма поширеними системами управління підприємствами. Додаток витягує і аналізує як завгодно великі масиви даних, незалежно від їх місця розташування і формату.
- Information Design Tool – потужний засіб для створення корпоративних інформаційних просторів, так званих Юніверсів. У середині них приховано те, що користувачі зазвичай не повинні бачити в стандартних ERP-системах: механізми виявлення, структуризації, аналізу і трансформації даних. Після створення Юніверсом корпоративний користувач працює з ними самостійно, у нього відсутня необхідність в допомозі з боку ІТ-відділу.

Більшість продуктів модуля мають інтуїтивно зрозумілий і візуально привабливий інтерфейс. Працювати з корпоративними даними можуть користувачі з будь-яким рівнем технічної грамотності. Необхідні дані для аналізу збираються і візуалізуються в зручній і інформативній формі: у вигляді таблиць, графіків і діаграм. Таким чином, рішення SAP BO BI оптимізує

корпоративну роботу всіх відділів і допомагають знизити навантаження на IT-фахівців.

Рішення SAP BusinessObjects BI легко взаємодіє з більшістю ERP-систем. Особливо ефективно використовувати модуль спільно з SAP High-Performance Analytic Appliance – високопродуктивної платформою для зберігання і аналізу корпоративної інформації.

### 2.1.3. Tableau

Як провідний інструмент візуалізації даних, Tableau має велику кількість унікальних функцій. Його потужна програма виявлення та дослідження даних дозволяє вам відповісти на важливі питання за лічені секунди. Ви можете використовувати інтерфейс перетягування Tableau, щоб візуалізувати будь-які дані, досліджувати різні подання та навіть легко поєднувати декілька баз даних. Це не вимагає складних сценаріїв. Той, хто розуміється на бізнес-проблемах, може вирішити це питання шляхом візуалізації відповідних даних. Після аналізу ділитися з іншими настільки ж просто, як і публікувати на сервері Tableau.

Для чіткого розуміння аналітичні інструменти в Tableau можна класифікувати на два розділи [59]:

1. Інструменти розробника використовуються для розробки інформаційних панелей, діаграм, звітів, візуалізації. Продукти Tableau, що належать до цієї категорії – це Tableau Desktop та Tableau Public.
2. Метою інструментів спільного використання є спільний доступ до візуалізацій, звітів, інформаційних панелей, створених за допомогою інструментів розробника. До цієї категорії належать продукти Tableau Online, Server та Reader.

Tableau з'єднує та витягує дані, що зберігаються в різних місцях. Він може отримувати дані з будь-якої платформи, яку тільки можна собі уявити. Просту базу даних, таку як Excel, до такої складної бази даних, як Oracle, бази даних у

хмарі, такі як веб-сервіси Amazon, база даних Microsoft Azure SQL, Google Cloud SQL та різні інші джерела даних, можуть бути отримані за допомогою Tableau.

Після запуску Tableau доступні готові роз'єми для передачі даних, що дозволяє підключатися до будь-якої бази даних. Залежно від версії Tableau кількість з'єднувачів даних, що підтримуються Tableau, буде відрізнятись.

Добуті дані можна підключити в режимі реального часу або додати до механізму обробки даних Tableau Desktop. Тут аналітик даних та інженер обробки даних розроблюють візуалізації. Створені інформаційні панелі потім надаються користувачам. Користувачі, які отримують інформаційні панелі, переглядають файл за допомогою програми читання таблиць.

Дані з Tableau Desktop можна опублікувати на сервері Tableau. Це корпоративна платформа, де підтримуються одночасна робота декількох аналітиків, управління, модуль безпеки та функції автоматизації. Завдяки серверу Tableau кінцеві користувачі мають кращий досвід доступу до файлів з будь-якого місця, будь то робочий стіл, мобільний телефон або електронна пошта.

Tableau пропонує рішення для всіх видів галузей, відділів та середовищ даних. Нижче наведено кілька унікальних функцій, які дозволяють Tableau обробляти різні сценарії.

- Швидкість аналізу. Оскільки він не вимагає високого рівня програмування, будь-який користувач, який має доступ до даних, може почати використовувати його для отримання цінної інформації.
- Самостійний. Tableau не потребує складної настройки програмного забезпечення. Настільна версія, яка використовується більшістю користувачів, легко встановлюється і містить усі функції, необхідні для запуску та завершення аналізу даних.
- Visual Discovery. Користувач досліджує та аналізує дані за допомогою візуальних інструментів, таких як кольори, лінії трендів, діаграми та

графіки. Сценарію для написання майже не потрібно, оскільки майже все робиться перетягуванням.

- Змішати різноманітні набори даних. Tableau дозволяє поєднувати різні реляційні, напівструктуровані та необроблені джерела даних у режимі реального часу, без значних витрат на інтеграцію. Користувачам не потрібно знати деталі збереження даних.
- Робота в реальному час. Tableau може фільтрувати, сортувати та обговорювати дані на льоту та вбудовувати поточну інформаційну панель на портали, такі як сайт SharePoint або Salesforce. Ви можете зберегти свій перегляд даних і дозволити колегам підписатися на ваші інтерактивні інформаційні панелі, щоб вони бачили найсвіжіші дані, просто оновивши веб-браузер.
- Централізовані дані. Сервер Tableau забезпечує централізоване розташування для управління всіма опублікованими джерелами даних організації. Ви можете видаляти, змінювати дозволи, додавати теги та управляти розкладами в одному зручному місці. На сервері даних легко запланувати оновлення та управляти ними. Адміністратори можуть централізовано визначати графік витягів на сервері як для покрокового, так і для повного оновлення.

#### 2.1.4. QlickView

QlickView – це програмний інструмент для виявлення даних та спостереження за замовниками від Qlik, лідера у сфері аналізу та розвідки. Він побудований для надання даних самообслуговування, які можуть допомогти приймати рішення та генерувати значну рентабельність інвестицій для користувачів будь-якого рівня технічної кваліфікації. Він пропонує кілька функцій, включаючи "попередньо консервовані" інформаційні панелі - панелі інструментів, які попередньо налаштовані для аналізу та інтерпретації даних.

Переваги QlickView:

- Підключається безпосередньо до джерела даних: підключається майже до будь-якого джерела даних, включаючи хмари, великі дані, файлові та локальні дані. Видобуває інформацію з ряду служб (Salesforce, HIVE, Teradata).
- Підприємницька співпраця: користувачі мають спільний доступ до однієї інформаційної панелі, з можливістю переглядати одне і те ж представлення даних або навіть відстежування один одного під час навігації в додатку.
- Сильні асоціації: QlikView містить потужну платформу побудови асоціацій для проведення прямих та непрямих пошуків у даних або в межах одного поля.
- Додаток In-Memory: розміщення програмного забезпечення в оперативній пам'яті дозволяє оброблювати запити швидше та ефективніше.
- Швидке розгортання та конфігурація: дані не потрібно зберігати в таблицях чи кубах.
- Надійний контроль даних за допомогою унікальних інформаційних панелей та комбінації фільтрів.

Недоліки:

- Не підтримує аналітичну обробку онлайн (OLAP)
- Користувачі не можуть передбачити прибутковість проектів або підрозділів за допомогою QlikView.
- Рішення не допомагає користувачам визначати закономірності в даних.
- Не має спеціальних функцій звітування.
- Не дозволяє користувачам планувати отримання звітів про ВІ в певний час у певному форматі.

#### 2.1.5. Deductor + Loginom

Deductor є аналітичною платформою, базою для «створення прикладних рішень в області аналізу даних. Реалізовані в Deductor технології дозволяють на базі єдиної архітектури пройти всі етапи побудови аналітичної системи від

створення сховища даних до автоматичного підбору моделей і візуалізації отриманих результатів» [26].

Deductor містить все необхідне для вирішення завдань Data Mining. Застосування методів Data Mining – фактично єдина можливість отримати користь з накопиченої інформації, в іншому випадку зібрані дані будуть лежати «мертвим вантажем». Data Mining дозволяє витягувати з даних знання і перетворити в конкурентні переваги: якісно прогнозувати, точніше, виявляти цільові аудиторії, прогнозувати розвиток подій, управляти ризиками тощо.

Deductor складається з 2-х основних частин:

1. Deductor Warehouse – це «багатовимірне сховище даних, що акумулює всю необхідну для аналізу предметної області інформацію. Використання єдиного сховища дозволяє забезпечити несуперечність даних і централізоване зберігання, а також автоматично забезпечує всю необхідну підтримку процесу аналізу даних. Deductor Warehouse оптимізований для вирішення саме аналітичних задач, що позитивно позначається на швидкості доступу до даних» [26]. У ряді випадків має сенс відмовитися від традиційного сховища даних і скористатися альтернативою – віртуальним сховищем Virtual Warehouse.
2. Deductor Studio – це «програма, яка реалізує функції імпорту, обробки, візуалізації і експорту даних. Deductor Studio може функціонувати і без сховища даних, отримуючи інформацію з будь-яких інших джерел, але найбільш оптимальним є їх спільне використання. В Deductor Studio включений повний набір механізмів, що дозволяє отримати інформацію з довільного джерела даних, провести весь цикл обробки (очищення, трансформацію даних, побудову моделей), відобразити отримані результати найбільш зручним чином (OLAP, таблиці, діаграми, дерева тощо) і експортувати їх в найбільш поширені формати» [26].

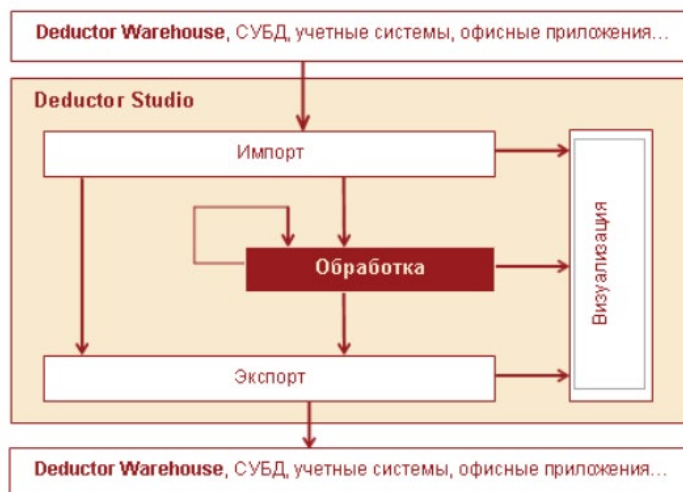


Рисунок 2.1 Схема рівня додатків Deductor Studio [26]

У 2018 року розвиток Deductor припинилося у зв'язку із закінченням його життєвого циклу. На зміну йому прийшла нова платформа – Loginom. Loginom здатний обробляти масиви даних в десятки разів більше за Deductor, підтримує значну кількість джерел даних, має зручний і сучасний інтерфейс [27].

Loginom – інтегрована Low-code платформа для реалізації всіх аналітичних процесів: від консолідації і підготовки даних до моделювання, розгортання і візуалізації. Вона підтримує візуальне налаштування логіки обробки і повторного використання компонентів, а також надає просунуту аналітику: від найпростіших формул до машинного навчання. В ній, на відміну від Deductor, підтримується багато інших ІС та їхніх форматів даних для імпорту/експорту, що полегшує інтеграцію. Loginom забезпечує підтримку JavaScript для повторного використання готових рішень популярних браузерних CRM-систем.

#### 2.1.6. Project Expert

Програмне забезпечення Project Expert є розробкою фірми «Проінвест» і являє собою систему підтримки прийняття рішень (СППР) для розробки, аналізу і вибору оптимального плану розвитку бізнесу, створення та аналізу фінансових планів та інвестиційних проектів.

Дана система дозволяє розробити імітаційну модель нового чи вже існуючого підприємства незалежно від його розмірів. Project Expert дає

можливість моделювати діяльність підприємств будь-якої галузевої належності та різних масштабів – від невеликих приватних підприємств до холдингових структур [13].

За допомогою вибору різних параметрів зовнішнього середовища та параметрів діяльності підприємства, що вводяться безпосередньо користувачем (фінансовим менеджером), у діалоговому режимі можна створити досить достовірну імітаційну модель конкретного інвестиційного проекту та оцінити отримані економічні показники і показники ефективності інвестицій. Регулюючи встановлені параметри імітаційної моделі, можна спостерігати, до яких наслідків призведуть ухвалені рішення або зміни зовнішніх факторів.

Як аналітична платформа, Project Expert здійснює дві основні функції:

1. перетворює опис діяльності підприємства з мови користувача у формалізований запис грошових потоків;
2. обчислює комплекс показників, відповідно до яких фінансовий менеджер може судити про (оцінювати) результативність ухвалених рішень.

Імітаційна модель фінансової діяльності підприємства Project Expert дозволяє розв'язувати такі задачі:

- розробити детальний фінансовий план та визначити потреби в грошових коштах на перспективу;
- визначити схему фінансування підприємства, оцінити можливість і ефективність залучення коштів з різних джерел;
- розробити план розвитку підприємства або реалізації інвестиційного проекту, визначивши найбільш ефективну стратегію маркетингу, а також стратегію виробництва, яка буде забезпечувати раціональне використання матеріальних, людських та фінансових ресурсів;
- розрахувати різні варіанти розвитку проекту, змінюючи значення факторів, що здатні впливати на його фінансові результати та порівняти їх між собою;

- сформуувати стандартні фінансові документи, розрахувати найпоширеніші фінансові показники, провести всебічний аналіз ефективності поточної та перспективної діяльності підприємства;
- провести аналіз проекту із урахуванням неточності початкових даних із застосуванням методу статистичного аналізу Монте-Карло;
- підготувати бізнес-план інвестиційного проекту, який відповідає міжнародним стандартам;
- здійснювати контроль за виконанням проектів, порівнюючи проектні та реальні грошові потоки.
- побудувати інтегровані фінансові потоки для групи проектів; оцінити ефективність діяльності компанії, яка впроваджує низку різних проектів; скласти бюджет для групи проектів.

Програмний продукт Project Expert реалізує функції обміну даними (проектами, стартовими балансами, планами збуту, підсумковими таблицями, текстовими файлами) з різними системами планування та управління проектами (MS Project, Primavera Project Planner), імпорту даних із програми фінансового аналізу Audit Expert та програми розроблення стратегічного плану маркетингу Marketing Expert, експорту даних у MS Word, MS Excel, MS Access, MS FoxPro тощо.

#### 2.1.7. Результати аналізу та порівняння ІС БА

Вище проведений огляд та аналіз існуючих інформаційних систем бізнес-аналітики, широко представлених на сьогоднішньому ринку, дозволяє порівняти розглянуті ІС за критеріями, наведеними у таблиці 2.1.

Таблиця 2.1 Порівняння ІС БА за загальними критеріями

Критерій \ ІС БА	Microsoft BI	SAP Business Online	QlikView	Tableau	Loginom	Project Expert
Підтримка скриптів R і Python	+	-	+	+	Python	-

Критерій \ ІС БА	Microsoft BI	SAP Business Online	QlikView	Tableau	Loginom	Project Expert
Динамічна крос-фільтрація	+	-	+	+	+	+
OLAP	+	+	+	+	+	-
Пошукова аналітика NLP	+	+	+	+	-	-
Підготовка даних	+	+	+	+	+	-
Модель даних	Star	Star	Snowflake	Flat	Snowflake	Flat
Незалежна від БД	+	+	+	+	+	-
Вбудована row-level безпека	+	+	+	+	+	+
Інтеграція з іншими ІС	+	-	-	-	+	+
Прогнозна аналітика	+	+	-	+	+	+

## 2.2. Основні характеристики розглянутих рішень бізнес-аналітики

### 2.2.1. Зберігання даних для бізнес-аналітики

Дані існують у ряді систем по всій організації. Для найбільш точного аналізу компанії повинні забезпечити стандартизоване форматування для всіх типів даних з кожної з цих систем. Наприклад, великі підприємства можуть тримати інформацію про своїх клієнтів у своїй програмі управління взаємовідносинами з клієнтами (CRM), а фінансові дані - у своїй програмі планування ресурсів підприємства (ERP). Ці окремі програми можуть по-різному маркувати та класифікувати дані, та їм обом доведеться стандартизувати дані перед аналізом [58].

Деякі програми програмного забезпечення для бізнес-аналітики отримують дані для аналізу безпосередньо з вихідних програм через власне підключення API або обробників веб-подій. Інші системи бізнес-аналітики вимагають використання сховищ даних для агрегування різноманітних наборів

даних в одному місці. Малий бізнес, окремі підрозділи або окремі користувачі можуть виявити, що пряме підключення працює добре, але великим корпораціям, корпоративним компаніям та компаніям, які генерують великі масиви даних, знадобиться більш гнучкі інструменти бізнес-аналітики.

Якщо вони вибирають рішення для централізованого зберігання, підприємства можуть використовувати сховище даних або інформаційний портал для зберігання інформації про свою компанію та придбання програмного забезпечення для виявлення, перетворення та завантаження (ETL) для полегшення зберігання даних. Крім того, вони можуть використовувати структуру зберігання даних, таку як Hadoop, для управління своїми даними.

### 2.2.2. Сховища даних

Бізнес-аналітика поєднує різні джерела даних в одну базу даних, будуючи сховище даних. Сховища даних виконують функцію центрального сховища даних, які запитуються та аналізуються іншими програмами BI. За допомогою методу виявлення, перетворення та завантаження сховища даних узагальнюють дані з усієї організації та полегшують іншим програмам швидкий доступ до них.

Інструменти аналітики та звітності все ще можуть функціонувати без сховищ даних, але запуск звітів через програмне забезпечення CRM або навіть програмне забезпечення торгових точок (POS) не тільки обмежує фокус розвідки, але й негативно впливає на продуктивність цих додатків. Крім того, дані в цих системах існують у різних форматах, що робить надзвичайно важким робити висновки та визначати закономірності без реструктуризації даних у загальний формат та розміщення їх у спільній зоні.

Дані, що зберігаються у сховищі даних, мають форму вимірів або фактів, які витягуються із систем, що виробляють дані. Факти представляють цифри для певної дії, як-от продаж віджета. Виміри дають контекст фактам, додаючи дати та місцезнаходження, а також їх називають метаданими. Наприклад, розміри

можуть розбити продажі віджетів на місяці або роки, що полегшує виконання запитів.

### 2.2.3. Data Marts

Фактично простіші, спеціальні версії сховищ даних, ринки даних фокусуються на конкретній підмножині даних, замість того, щоб зберігати дані з усієї компанії. Вони можуть зберігати частіше використовувані дані або дані, які використовує лише один відділ. Компаніям буде дешевше впроваджувати ринки даних, ніж цілі сховища даних, і вони можуть забезпечити не ІТ-персонал кращим досвідом роботи, обмежуючи складність бази даних [42].

### 2.2.4. Видобуток, перетворення та завантаження (ETL)

Призначені процесом, за допомогою якого дані передаються в сховище даних, програми ETL нормалізують дані в центральному місці. Компанії можуть придбати програмне забезпечення ETL із програмним забезпеченням для зберігання даних або як додаткову програму. Давайте розглянемо кожен частину процесу ETL:

Видобуток даних – це процес отримання даних із системи походження. Часто найскладніший аспект процесу, ступінь успішності видобування даних із вихідних систем, ERP або CRM-систем, впливає на успіх решти процесу. Неструктуровані дані погано відформатовані для розміщення в рядки та колонки, що ускладнює аналіз після зберігання у сховищі даних. Позначення неструктурованих даних такими метаданими, як інформація про автора, тип вмісту та інші фактори, що ідентифікують, може допомогти командам знайти потрібні дані, коли вони зберігаються в сховищі даних і з часом завантажуються в програмне забезпечення BI [43].

Трансформувати: після видобування даних із програми, з якої вони походять, ці дані повинні бути нормалізовані, перш ніж зберігатись у сховищі даних для подальшого використання. Щоб аналізи в системі бізнес-аналітики

працювали належним чином, дані з різних програм походження повинні існувати в одному форматі, інакше запити будуть неточними.

Завантажити: видобуті та нормалізовані дані готові до завантаження в центральну базу даних, найчастіше сховище даних. Частоти навантаження будуть різнитися залежно від організації. Деякі компанії можуть вводити нові дані щотижня, а інші – щодня.

#### 2.2.5. Видобуток даних

Видобуток даних включає автоматизований та напівавтоматизований аналіз наборів даних для виявлення закономірностей та невідповідностей. Загальні кореляції, отримані в результаті інтелектуального аналізу даних, включають групування конкретних наборів даних, знаходження аномалій у даних та створення зв'язків або залежностей з різних наборів даних [44].

Видобуток даних часто виявляє закономірності, що використовуються в більш складних аналізах, таких як прогнозне моделювання, що робить його важливою частиною процесу ВІ.

Зі стандартних процесів, що виконуються за допомогою видобування даних, вивчення правил асоціацій представляє найбільшу користь. Вивчаючи дані для визначення залежностей та побудови співвідношень, правило асоціації може допомогти компаніям краще зрозуміти спосіб взаємодії клієнтів із їх веб-сайтом або навіть те, які фактори впливають на їхню купівельну поведінку.

#### 2.2.6. Прогнозна аналітика

Мабуть, одним із найцікавіших аспектів бізнес-аналітики, програми прогносної аналітики функціонують як вдосконалена підмножина обробки даних. Як підказує назва, прогнозована аналітика прогнозує майбутні події на основі поточних та історичних даних. Складаючи зв'язки між наборами даних, ці програмні програми передбачають ймовірність майбутніх подій, що може призвести до величезної конкурентної переваги для бізнесу.

Прогностичний аналіз передбачає детальне моделювання і навіть заходить у сферу машинного навчання, де програмне забезпечення насправді вчиться на минулих (попередніх) подіях, щоб передбачити майбутні наслідки. Для наших цілей зупинимось на двох основних формах прогнозного аналізу [45]:

1. *Прогнозне моделювання.* Цей тип програмного забезпечення, найвідоміший сегмент прогнозової аналітики, робить те, що впливає з його назви: він передбачає, особливо стосовно одного елемента. Прогнозні моделі шукають кореляцію між певною одиницею виміру та принаймні однією або кількома ознаками, що належать до цієї одиниці. Мета полягає в тому, щоб знайти однакову кореляцію між різними наборами даних.
2. *Описове моделювання.* Тоді як прогнозне моделювання шукає єдину кореляцію між підрозділом та його ознаками, то для того, щоб, наприклад, передбачити ймовірність зміни клієнтом постачальників послуг, описове моделювання прагне зменшити дані до розмірів та групувань, що піддаються управлінню. Описова аналітика добре працює для узагальнення такої інформації, як унікальні перегляди сторінок або згадування в соціальних мережах.

#### 2.2.7. Обробка природної мови

Дані мають три основні форми: структуровану, напівструктуровану та неструктуровану. Неструктуровані дані є найпоширенішими і включають текстові документи та інші типи файлів, що існують у форматах, які комп'ютери не можуть легко прочитати.

Неструктуровані дані не можна зберігати в рядках або колонках, що робить неможливим аналіз традиційного програмного забезпечення для аналізу даних. Однак ці дані часто мають вирішальне значення для розуміння результатів бізнесу. Маючи стільки даних у неструктурованому вигляді, аналіз тексту повинен стати ключовим фактором при спробі знайти найкраще програмне забезпечення для бізнес-аналітики.

Програмне забезпечення для обробки природної мови (NLP), також відоме як програмне забезпечення для аналізу тексту, поєднує великі набори неструктурованих даних, щоб знайти приховані закономірності. NLP особливо цікавий для підприємств, які працюють із соціальними мережами. Використовуючи відповідне програмне забезпечення, компанія може встановити правило для відстеження ключових слів або фраз - наприклад, назви компанії - для пошуку закономірностей використання клієнтами цієї мови. Інструменти обробки природними мовами також вимірюють настрої споживачів, дають уявлення про цінність клієнта протягом усього життя та вивчають тенденції споживачів, які можуть бути інформацією про майбутні асортименти [46].

#### 2.2.8. Оперативна база даних

Обробка бази даних в пам'яті використовує оперативну пам'ять замість обробки диска або жорсткого диска для зчитування інформації. Доступ до інформації таким чином збільшує продуктивність програми в геометричній прогресії. Зростаюча потужність оперативної пам'яті в наших обчислювальних середовищах у поєднанні з попитом на більш гнучкі системи означає, що це програмне забезпечення має великий вклад у майбутнє ВІ. Різке падіння цін на пам'ять робить його більш популярним варіантом проведення аналізу за допомогою багатовимірних баз даних та кубів.

Все більше і більше користувачів ВІ не є ІТ-персоналом. Це співробітники зі стандартним обсягом технологічної кмітливості, які хочуть використати потужність ВІ для отримання конкурентних переваг.

Отже, дизайн механізмів звітності та простота використання аналітичних функцій рухаються до нижчого бар'єру доступу. Більше не достатньо чудових функцій аналізу чи зберігання даних. Вони повинні використовуватися як ІТ-експертами, так і бізнес-користувачами, які не мають аналітичного досвіду.

### 2.2.9. Вбудована аналітика

Програмне забезпечення бізнес-аналітики обіцяє пояснити бізнес-аналітику для найбільш нетехнічних співробітників, що зумовило попит на вбудовані засоби аналітики. Ці інструменти дозволяють компаніям створювати візуалізацію даних у своєму програмному забезпеченні ВІ та динамічно обслуговувати ці візуалізації внутрішнім та зовнішнім клієнтам у програмах компанії [18].

Вбудована аналітика заощаджує компаніям тисячі годин і сотні тисяч доларів, які вони в іншому випадку використали б для створення інформаційних панелей звітності та аналітики з нуля. Тепер ці інструменти дають діловим користувачам доступ до власних візуалізацій plug-and-play, що значно прискорює час виходу на ринок.

## 2.3. Математичний апарат інформаційних технологій бізнес-аналітики

### 2.3.1. Аналіз ринкового кошика

Аналіз ринкового кошика (Market Basket Analysis) – це методика, яка визначає міцність асоціації між парами товарів, що купуються разом, та визначає закономірності їх спільних появ. Спільна поява – це коли дві або більше подій відбуваються разом [19].

Аналіз ринкового кошика створює правила сценарію «Якщо-тоді», тобто логічної імплікації. Наприклад, якщо елемент *A* придбано, то елемент *B*, ймовірно, буде придбаний. Правила носять імовірнісний характер, або, іншими словами, вони походять від частоти спільних появ у спостереженнях. Частота – це частка кошиків, що містять предмети, що нас цікавлять.

*Ринковий кошик* визначається як набір предметів, куплений клієнтом під час одного відвідування магазину. Під час відвідування супермаркету ми, як правило, купуємо багато товарів різних категорій і ставимо їх усіх разом в

одному єдиному кошику, що розглядається як одна операція. Аналіз ринкового кошика – це аналіз таких кошиків разом.

Аналіз ринкового кошика охоплює широкий набір аналітичних методів, спрямованих на виявлення асоціацій та зв'язків між конкретними об'єктами, виявлення поведінки споживачів та взаємозв'язку між предметами. У роздрібній торгівлі він використовується, виходячи з наступної ідеї: якщо клієнт купує певну групу предметів, з деякою ймовірністю він купує й іншу групу товарів [20]. Наприклад, відомо, що коли клієнт купує пиво, в більшості випадків він купує також і чіпси.

Різні типи поведінки, що виникають під час покупок, цікавлять компанії, що продають свою продукцію. Вони зацікавлені в аналізі товарів, які купуються разом, щоби створити нові стратегії маркетингу/збуту, які можуть бути корисними для збільшення переваг компанії, а також кращого досвіду здійснення покупок клієнтами.

### 2.3.2. Асоціативні правила

Асоціативні правила дозволяють знаходити закономірності між пов'язаними подіями. Прикладом такої закономірності служить правило, яке вказує, що з події  $X$  слідує подія  $Y$  з деякою ймовірністю. Встановлення таких залежностей дає можливість знаходити дуже прості й інтуїтивно зрозумілі правила.

Вперше завдання пошуку асоціативних правил було запропоновано для «знаходження типових шаблонів покупок, що здійснюються в супермаркетах, тому іноді її ще називають аналізом ринкової корзини» [19].

Нехай маємо базу даних, що складається транзакцій клієнтів супермаркету, де кожна транзакція – це набір товарів, придбаних клієнтом за один візит. Таку транзакцію ще називають ринковим кошиком.

Нехай  $I=(i_1, i_2, \dots, i_n)$  – множина (набір) товарів, що називаються елементами.

Нехай  $D$  – множина транзакцій, де кожна транзакція  $T$  – це набір елементів  $I$ ,  $T \in I$ . Кожна транзакція представляє собою бінарний вектор, де  $T(k) = 1$ , якщо  $i_k$  присутній у транзакції, інакше  $T(k)=0$ . Говорять, що транзакція  $T$  містить  $X$ , деякий набір елементів з  $I$ , якщо  $X \in T$ . Асоціативним правилом називається імплікація  $X \Rightarrow Y$ , де  $X \in I$ ,  $Y \in I$  та  $X$  перетин  $Y$  є порожньою множиною. Правило  $X \Rightarrow Y$  має підтримку  $s$  (*support*), якщо  $s$  відсотків транзакцій з  $D$  містять  $X \cup Y$ , тобто  $support(X \Rightarrow Y) = support(X \cup Y)$ . Достовірність правила показує яка ймовірність того, що з  $X$  витікає  $Y$ . Правило  $X \Rightarrow Y$  справедливе з достовірністю (*confidence*)  $c$ , якщо  $c$  відсотків транзакцій з  $D$ , що містять  $X$ , також містять  $Y$ ,  $confidence(X \Rightarrow Y) = support(X \cup Y) / support(X)$ .

Наприклад, 46% транзакцій, що містять «Чай», також містять «Печиво». 10% усіх транзакцій містять обидва товари. 46% – це достовірність (*confidence*) правила, іншими словами, «Чай» купується з «Печивом» з імовірністю 46%, а 10% - це підтримка (*support*), або те, що умова «Чай» і наслідок «Печиво» зустрічається у 10% розглянутих транзакцій.

Мета цього аналізу – встановити залежності виду якщо в транзакції зустрічається набір елементів  $X$ , то це може означати, що інший набір елементів  $Y$  повинен зустрічатися в цій або подібних транзакціях теж. Визначення цих залежностей допомагає знаходити дуже прості й інтуїтивно зрозумілі правила [47].

Алгоритми пошуку асоціативних правил [47; 48; 49] призначаються для знаходження всіх правил  $X \Rightarrow Y$ , підтримка і достовірність яких повинні бути вищими за деякі попередньо визначені пороги, які називаються мінімальною підтримкою (*support*) і мінімальною достовірністю (*confidence*) відповідно.

Пошук асоціативних правил зазвичай розбивається на 2 підзавдання:

- 1) Знайти всі набори елементів, які задовольняють поріг мінімальної підтримки. Такі набори елементів називаються таким, що часто зустрічаються.

- 2) Згенерувати правила з наборів елементів, знайдених на попередньому кроці, з достовірністю, яка задовольняє мінімальну достовірність.

Одним з перших представлених алгоритмів, який ефективно розв'язує подібні класи завдань, був алгоритм Apriori. Крім нього були розроблені й інші алгоритми: DHP, Partition, DIC тощо [49].

Значення порогів мінімальної підтримки та мінімальної достовірності вибираються так, щоб обмежити кількість знайдених правил. Адже якщо підтримка правила велика, то це означає, що алгоритм буде знаходити такі правила, які вже відомі аналітикам або настільки очевидні, що немає жодного сенсу проводити такий аналіз.

Однак, низька підтримка може призвести до генерації набагато більшої кількості правил, що, звісно, вимагає суттєвої обчислювальної потужності. Проте аналітиками прийнято вважати, що більшість цікавих правил знаходиться саме за низького порогу мінімальної підтримки, хоча і надто низьке значення генерує статистично необґрунтовані правила.

При цьому пошук асоціативних правил – це не тривіальна задача, як може здатися на перший погляд. Одна з проблем – алгоритмічна складність [49] при знаходженні наборів популярних елементів, тому що зі зростанням кількості елементів в  $I$  ( $|I|$ ) експоненціально зростає число потенційних наборів елементів.

### 2.3.2.1. Узагальнені асоціативні правила

Підрозділом вище було припущено, що всі елементи однорідні. Якщо повернутися до аналізу ринкового кошика, то з'ясується, що це товари однакових атрибутів, окрім їх назви. Проте ми завжди можемо доповнити транзакцію інформацією про те, до якої групи належить конкретний товар, і побудувати їхню ієрархію.

Нехай дано базу транзакції, при цьому відомо, що її елементи входять у ту чи іншу групу. Тоді з цих даних можна добувати правила, які поєднують одну групу з іншими, одні елементи з декількома групами тощо.

Наприклад, якщо клієнт придбав товар з групи «Напої», то він придбає і товар групи «Молочні продукти» або «Соки». Ці правила називаються узагальнено асоціативними.

Узагальненим асоціативним правилом називається імплікація  $X \Rightarrow Y$ , де  $X \in I$ ,  $Y \in I$  й  $X \cap Y = \emptyset$  (порожня множина) і де жоден з елементів, що входять у набір  $Y$ , не є батьком жодного з елементів, що входять у  $X$ . Підтримка та достовірність обчислюються аналогічно до простих асоціативних правил.

Якщо ввести додаткову інформацію про групи елементів у вигляді ієрархії, це дає наступні переваги:

- 1) Асоціативні правила встановлюються не лише серед окремих елементів, а й між рівнями ієрархії (групами).
- 2) Навіть якщо деякі елементи мають недостатню підтримку, група в цілому може задовольняти поріг мінімальної підтримки.

Щоби знайти такі правила, достатньо використовувати будь-який з попередньо розглянутих алгоритмів. З цією метою, кожна транзакція доповнюється всіма предками (батьками) кожного елемента, який присутній у транзакції. Проте якщо застосовувати ці алгоритми без попередньої обробки, це може призвести до наступних проблем:

- 1) Елементи нижніх рівнів ієрархії прагнуть до нижчих значень підтримки, порівнюючи з елементами верхніх рівнів.
- 2) При додаванні груп у транзакції, також збільшується кількість атрибутів і розмірність простору входів відповідно. Таким чином ускладнюється завдання пошуку, генеруючи набагато більшу кількість правил.
- 3) З'являються надлишкові правила, які можуть суперечити узагальненому асоціативному правилу, наприклад, «Соки» – «Прохолодні напої». Видно, що цінність такого правила на практиці – нульова при достовірності 100%. Тому також доцільно розглядати спеціальні оператори, що видаляють подібні надлишкові правила.

Шукаючи узагальнені асоціативні правила, рекомендують використовувати спеціалізований алгоритм [48], який усуває згадані проблеми, а також у 2-5 разів швидший за стандартний Apriori.

Отже, елементи групуються не тільки за входженням у конкретну товарну категорію, а й за іншими характеристиками, такими як ціна (дешево, дорого), бренд тощо.

#### 2.3.2.2. Чисельні асоціативні правила

До цього було розглянуто спрощене завдання пошуку асоціативних правил, оскільки все призводило до отримання такого факту, який свідчив би чи присутній цей елемент у транзакції чи ні. Розглядаючи випадок ринкової корзини, було розглянуто два стани: товар було придбано чи ні, ігноруючи інформацію, наприклад, про кількість товару, хто саме його купив, характеристики покупця тощо. Тож можна стверджувати, що розглядалися «булеві» асоціативні правила. Переглянувши будь-яку базу даних, помітно, що кожна транзакція складається з числових, категоріальних та інших типів даних. Для обробки цих записів і пошуку чисельних асоціативних правил було запропоновано спеціальний алгоритм пошуку [47].

#### 2.3.3. Класифікація

Класифікація в аналізі даних – це розбиття множини об'єктів або спостережень на попередньо задані групи, або класи, всередині кожної з яких вони передбачаються подібними один на одного, мають приблизно однакові властивості й ознаки. При цьому рішення виходить на основі аналізу значень атрибутів (ознак).

Формалізуючи, нехай  $X$  – множина опису об'єктів,  $Y$  – скінчена множина нумерацій класів. Існує одна невідома цільова залежність – відображення  $y^*$ :  $X \rightarrow Y$ , значення якої відомі тільки на об'єктах скінченної навчальної вибірки

$X_m = (x_{1m}, y_{1m}), \dots, (x_{mm}, y_{mm})$ . Необхідно побудувати такий алгоритм  $a: X \rightarrow Y$ , що здатен класифікувати довільний об'єкт  $x$  з множини  $X$ .

Класифікація є однією з найважливіших задач Data Mining, адже якщо відомо властивості об'єктів кожного класу, то коли нове спостереження відноситься до певного класу, ці властивості автоматично поширюються і на нього [21].

Якщо число класів обмежена двома, то має місце бінарна класифікація, до якої можуть бути зведені набагато складніші завдання. Наприклад, замість визначення таких ступенів кредитного ризику, як «Високий», «Середній» або «Низький», можна використовувати лише дві – «Видати» або «Відмовити».

Для класифікації в Data Mining використовується багато різних моделей: нейронні мережі, дерева рішень, машини опорних векторів, метод  $k$ -найближчих сусідів, алгоритми покриття тощо.

#### 2.3.4. Методи прогнозування

##### 2.3.4.1. Якісні методи прогнозування

Методи прогнозування значною мірою залежать від наявних даних: якщо дані відсутні або наявні, але не зовсім підходять до прогнозування, тоді слід використовувати якісні методи прогнозування. Ці методи не є суто здогадками – існують добре розроблені структуровані підходи отримання хороших прогнозів без використання історичних даних.

Суб'єктивні судження (judgmental forecasts) є суб'єктивними, і тому можуть містити різні види упереджень та обмежень.

Такі прогнози можуть бути суперечливими. На відміну від статистичних прогнозів, які кожного разу можуть формуватися за допомогою одних і тих же математичних формул, суб'єктивні судження значною мірою залежать від пізнання людини і є вразливими до його обмежень. Наприклад, обмеження у вигляді пам'яті може зробити останні події важливішими, ніж вони є насправді,

і може ігнорувати важливі події з більш далекого минулого; або обмежена тривалість уваги може призвести до втрати важливої інформації; або нерозуміння причинно-наслідкових зв'язків може призвести до помилкових висновків.

Дослідження в цій галузі [50; 51] показали, що точність суб'єктивних суджень покращується, коли прогнозуючий має:

- 1) важливі знання в області,
- 2) більш своєчасну, актуальну інформацію.

Суб'єктивні судження можуть швидко пристосуватися до таких змін, інформації чи подій.

Існує три загальні ситуації, в яких використовується суб'єктивні судження:

- 1) дані відсутні, тому статистичні методи не застосовуються, а суб'єктивні судження є єдиним можливим підходом;
- 2) дані доступні, формуються статистичні прогнози, які потім коригуються з використанням суб'єктивних суджень;
- 3) дані доступні, статистичні та суб'єктивні судження формуються незалежно, а потім об'єднуються.

Варто зазначити, що коли дані доступні, рекомендується застосовувати тільки статистичні методи.

#### 2.3.4.1.1. Метод Дельфі (представник методів суб'єктивних суджень)

Метод Дельфі був винайдений Олафом Хелмером та Норманом Далкі з корпорації «Ренд» у 1950-х роках для вирішення конкретної військової проблеми. Метод спирається на ключове припущення, що прогнози групи зазвичай є більш точними, ніж прогнози окремих людей. Метою методу Дельфі є побудова консенсусних прогнозів групи експертів у структурованому ітеративному порядку. Для впровадження та управління процесом призначається фасилітатор. Метод Дельфі зазвичай включає наступні етапи:

- 1) Зібрана група експертів.
- 2) Завдання прогнозування встановлюються та розподіляються серед експертів.
- 3) Експерти повертають початкові прогнози та обґрунтування. Вони складаються та узагальнюються з метою надання зворотного зв'язку.
- 4) Відгуки надаються експертам, які зараз переглядають свої прогнози з урахуванням відгуків. Цей крок може повторюватися до досягнення задовільного рівня консенсусу.
- 5) Остаточні прогнози будуються шляхом узагальнення прогнозів експертів.

#### 2.3.4.2. Кількісні методи прогнозування

Кількісне прогнозування може застосовуватися, коли виконуються дві умови [45]:

- 1) доступна числова інформація про минуле;
- 2) припускають, що деякі аспекти минулого продовжуватимуться і в майбутньому.

Існує широкий спектр методів кількісного прогнозування, які часто розробляються в рамках конкретних дисциплін для конкретних цілей. Кожен метод має свої властивості, точність та витрати, які необхідно враховувати при виборі конкретного методу.

Більшість проблем кількісного прогнозування використовують дані часових рядів (зібрані через рівні проміжки часу) або дані поперечного перерізу (зібрані в один момент часу).

Все, що спостерігається послідовно з часом, є часовим рядом. Мета прогнозування часових рядів – оцінка того, як послідовність продовжиться у майбутньому.

Найпростіші методи прогнозування часових рядів використовують лише інформацію про змінну, яку потрібно прогнозувати, і не роблять спроб виявити фактори, що впливають на її поведінку. Тому вони будуть екстраполювати

тенденції та сезонні структури, ігноруючи всю іншу інформацію, таку як маркетингові ініціативи, діяльність конкурентів, зміни економічних умов тощо.

Моделі часових рядів, що використовуються для прогнозування, включають моделі декомпозиції, моделі експоненціального згладжування та моделі ARIMA.

Змінні предикторів часто корисні при прогнозуванні часових рядів. Наприклад, припустимо, ми хочемо прогнозувати погодинну потребу в електроенергії (ПЕ) спекотного регіону протягом літнього періоду. Модель із предикторними змінними може мати такий вигляд:

$$ПЕ = f(\text{температура, економ. стан, населення, час, день, похибка}) \quad (2.1)$$

Співвідношення не є точним – завжди будуть відбуватися зміни у попиті на електроенергію, які неможливо врахувати за допомогою змінних прогнозу. Похибка дозволяє випадкові зміни та наслідки відповідних змінних, які не включені у модель. Це називається пояснювальною моделлю, оскільки вона допомагає пояснити, що спричиняє зміни у потребі в електроенергії [45].

Оскільки дані про попит на електроенергію утворюють часовий ряд, ми також можемо використовувати модель часових рядів для прогнозування. У цьому випадку відповідне рівняння прогнозування часових рядів має вигляд

$$ПЕ = f(ПЕ_{\tau}, ПЕ_{\tau-1}, \dots, \text{похибка}) \quad (2.2)$$

де  $\tau$  – це поточна година,  $\tau + 1$  – наступна година,  $\tau - 1$  – попередня година тощо. Тут прогнозування майбутнього базується на минулих значеннях змінної, але не на зовнішніх змінних, які можуть вплинути на систему. Знову ж таки, похибка дозволяє випадкові зміни та наслідки відповідних змінних, які не включені в модель.

Існує також третій тип моделі, який поєднує в собі ознаки двох вищезазначених моделей. Наприклад, це може бути подано як

$$ПЕ_{\tau+1} = f(ПЕ_{\tau}, \text{температура, час, день, похибка}) \quad (2.3)$$

Цей тип «змішаних моделей» отримав різні назви в різних дисциплінах. Вони відомі як моделі динамічної регресії, моделі даних панелей, моделі функцій передачі та моделі лінійних систем (припускаючи, що функція є лінійною).

Розглядаючи часові ряди, варто розрізняти наступні терміни:

- Тренд, або тенденція, існує тоді, коли спостерігається довгострокове збільшення або зменшення даних. Він не повинен бути лінійним. Іноді тренд позначається як «зміна напрямку», коли відбувається перехід від зростаючої до спадної тенденції.
- Сезонна закономірність виникає, коли на часовий ряд впливають сезонні фактори, такі як пора року або день тижня. Сезонність завжди має фіксовану і відому частоту.
- Цикл виникає, коли зростання падіння даних не мають фіксованої частоти. Ці коливання, як правило, зумовлені економічними умовами і часто пов'язані з «діловим циклом». Тривалість цих коливань зазвичай становить не менше 2 років.

Коли ми розкладаємо часовий ряд на компоненти, ми зазвичай поєднуємо тренд і цикл в один компонент – цикл тренду (іноді його називають просто трендом). Таким чином, часовий ряд складається з трьох компонентів: компонента трендового циклу, сезонної складової та залишкової складової (що містить все інше у часових рядах).

#### 2.3.4.2.1. Примітивні методи прогнозування

##### 2.3.4.2.1.1. Метод середніх

Прогнози всіх майбутніх значень дорівнюють середньому історичних даних. Якщо ми позначимо ці значення як  $y_1, \dots, y_T$ , то прогнозовані значення можна отримати наступним чином

$$\hat{y}_{T+h|T} = \bar{y} = \frac{(y_1 + \dots + y_T)}{T}, \quad (2.4)$$

де  $\hat{y}_{T+h|T}$  – це оцінка  $\hat{y}_{T+h}$  на основі  $y_1, \dots, y_T$ .

#### 2.3.4.2.1.2. Наївний метод

У наївних прогнозах говорять, що всі прогнозовані значення мають значення останнього спостереження, тобто

$$\hat{y}_{T+h|T} = y_T \quad (2.5)$$

Один із представників групи таких методів – це метод випадкового блукання (random walk method).

У випадку сезонності даних, кожне прогнозоване значення відповідає останньому спостережуваному значенню сезону з того самого року (наприклад, того ж місяця попереднього року). Описуючи формально, прогнозоване значення для часу  $T + h$  обчислюється за наступною формулою

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)} \quad (2.6)$$

де  $m$  – це сезон, а  $k$  – ціла частина  $\frac{h-1}{m}$  (тобто кількість повних років у прогнозі до моменту  $T+h$ ).

#### 2.3.4.2.1.3. Метод дрейфування

Частковий випадок наївного методу, який полягає у тому, щоб прогнози могли збільшуватися чи зменшуватися з часом, де величина змін у часі (дрейф) встановлюється як середня зміна, яка спостерігається в історичних даних. Таким чином прогноз для часу  $T + h$  задається наступною формулою

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1} \sum_{t=2}^T (y_t - y_{t-1}) = y_T + h \left( \frac{y_T - y_1}{T-1} \right) \quad (2.7)$$

Це еквівалентно проведенню прямої між першим та останнім спостереженнями та екстраполяції цієї прямої у майбутнє.

### 2.3.5. Кореляція

Загальноприйнятим є обчислення коефіцієнтів кореляції для вимірювання сили взаємозв'язку між двома змінними. Кореляція між змінними  $x$  та  $y$  визначається як (2.8)

$$r = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum(x_t - \bar{x})^2} \sqrt{\sum(y_t - \bar{y})^2}}. \quad (2.8)$$

Значення  $r$  завжди знаходиться у проміжку від  $-1$  до  $1$ , при цьому негативні значення вказують на негативний зв'язок (збільшення  $x$  зменшує  $y$ ), а позитивні – на позитивний (збільшення  $x$  збільшує  $y$ ). Коефіцієнт кореляції вимірює лише силу лінійного співвідношення, тому іноді може вводити в оману при аналізі нелінійних даних. Візуалізація різних рівнів кореляції наведена на рис. 2.3.

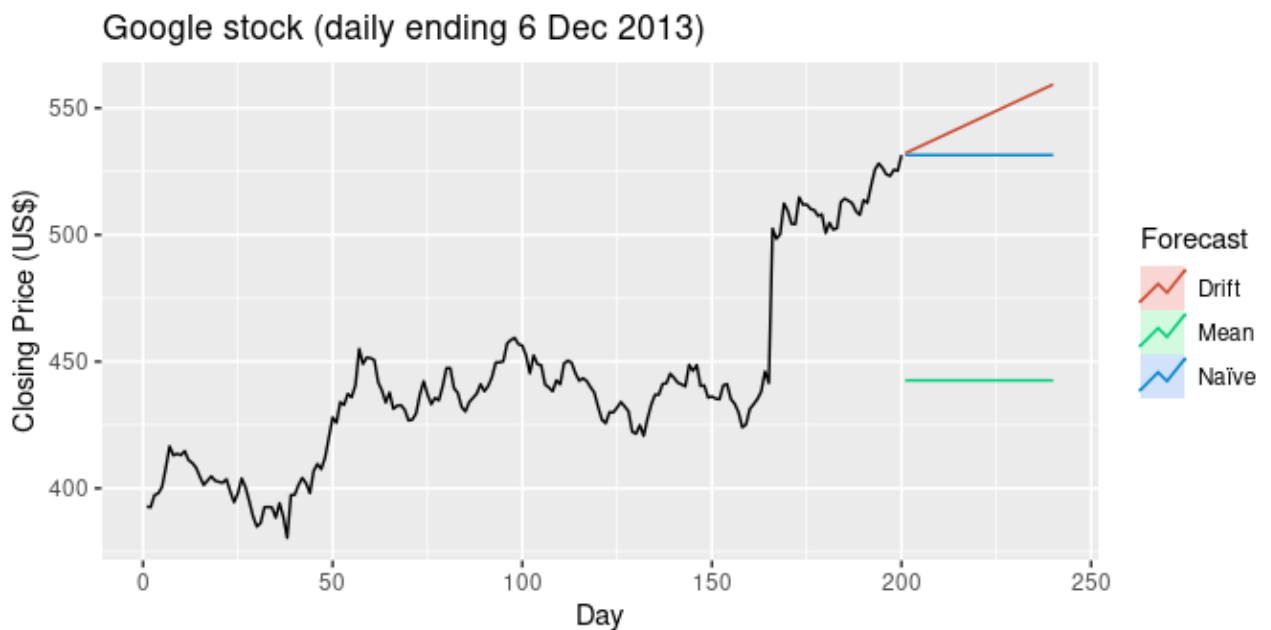


Рисунок 2.2 Порівняння примітивних методів прогнозування [45]

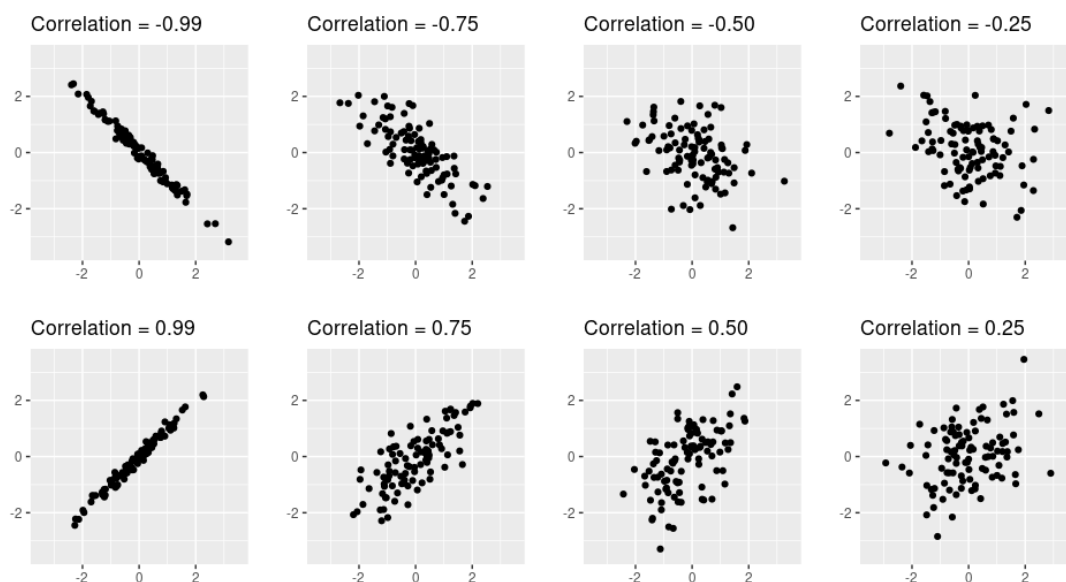


Рисунок 2.3 Приклади різної кореляції [45]

### 2.3.5.1. Автокореляція

Подібно до того, як кореляція вимірює ступінь лінійної залежності між двома змінними, автокореляція вимірює лінійну залежність між минулими значеннями часового ряду [45; 22].

Існує кілька коефіцієнтів автокореляції, що відповідають кожній панелі в графіку відставання (lag plot). Наприклад,  $r_1$  вимірює взаємозв'язок між  $y_t$  та  $y_{t-1}$ ,  $r_2$  вимірює взаємозв'язок між  $y_t$  та  $y_{t-2}$  тощо.

Значення  $r_k$  можна записати як

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (2.9)$$

де  $t$  – тривалість часового ряду.

Наприклад, перші 9 коефіцієнтів автокореляції даних виробництва пива наведені у табл. 2.2.

Таблиця 2.2 Перші 9 коефіцієнтів автокореляції показників виробництва пива

$r_1$	$r_2$	$r_3$	$r_4$	$r_5$	$r_6$	$r_7$	$r_8$	$r_9$
-0.102	-0.657	-0.060	0.869	-0.089	-0.635	-0.054	0.832	-0.108

Ці коефіцієнти відповідають графіку функції автокореляції (ACF), який наведено на рис. 2.4. Такий графік також називають корелограмою.

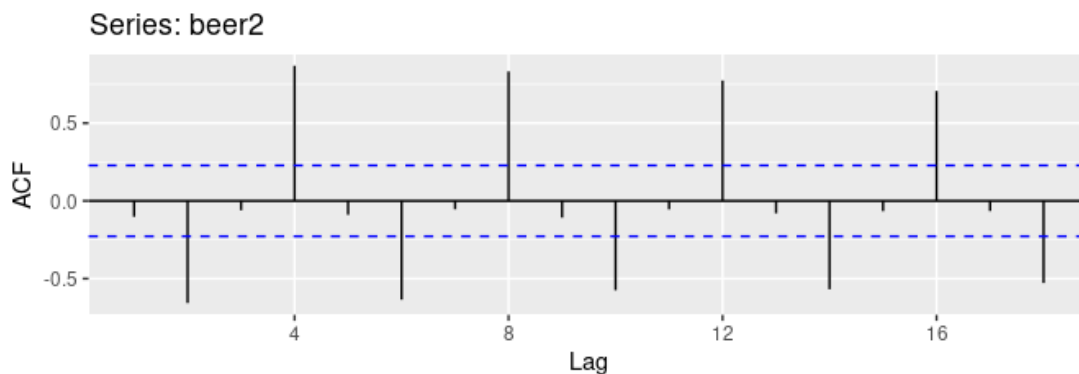


Рисунок 2.4 Корелограма часового ряду показників виробництва пива [45]

На рис. 2.4.:

- $r_4$  вище, ніж для інших відставань (lags). Це пов'язано із сезонною структурою даних: вершини, як правило, розташовані на відстані чотири чверті, а западини, як правило, на відстані дві чверті.
- $r_2$  є більш негативним, ніж для інших відставань, оскільки западини, як правило, відстають на дві чверті від вершин.
- пунктирні сині лінії вказують, чи кореляція взагалі існує (суттєво відрізняється від нуля).

Коли дані мають тенденцію (тренд), автокореляції для невеликих відстаней, як правило, є великими додатними значеннями, оскільки спостереження у певний період часу також мають приблизно однаковий розмір. Таким чином, ACF тенденційних часових рядів, як правило, має позитивні значення, які повільно зменшуються із збільшенням відстані. Якщо дані сезонні, автокореляції будуть більшими для сезонних відстаней (із кратним значенням сезонної частоти), ніж для інших відстаней [45]. У випадку, якщо дані трендові та сезонні одночасно, ці «ефекти» можуть поєднуватися.

### 2.3.5.2. Шуми

Часові ряди, які не показують автокореляції, називаються білим шумом,

або просто шумом. Для рядів білого шуму очікується, що кожна автокореляція буде близькою до нуля. Варто зазначити, що вони можуть і не дорівнювати нулю, оскільки завжди можуть бути деякі випадкові відхилення. Для білого шуму природньо, що 95% стрибків функції автокореляції лежать у межах  $\pm 2\sqrt{T}$ , де  $T$  – це довжина часового ряду.

Загальноприйнято наносити ці межі на графік AFC: якщо 5% «викидів» функції перебувають поза цими межами, тоді ряд, швидше за все, не є білим шумом.

На рис. 2.5 зображено приклад часового ряду, що є білим шумом. Оскільки його довжина  $T=50$ , тоді межі «білого шуму» становлять  $\pm 2\sqrt{50} \approx 0,28$ . Усі коефіцієнти автокореляції лежать у цих межах (рис. 2.6), підтверджуючи, що дані є білим шумом.

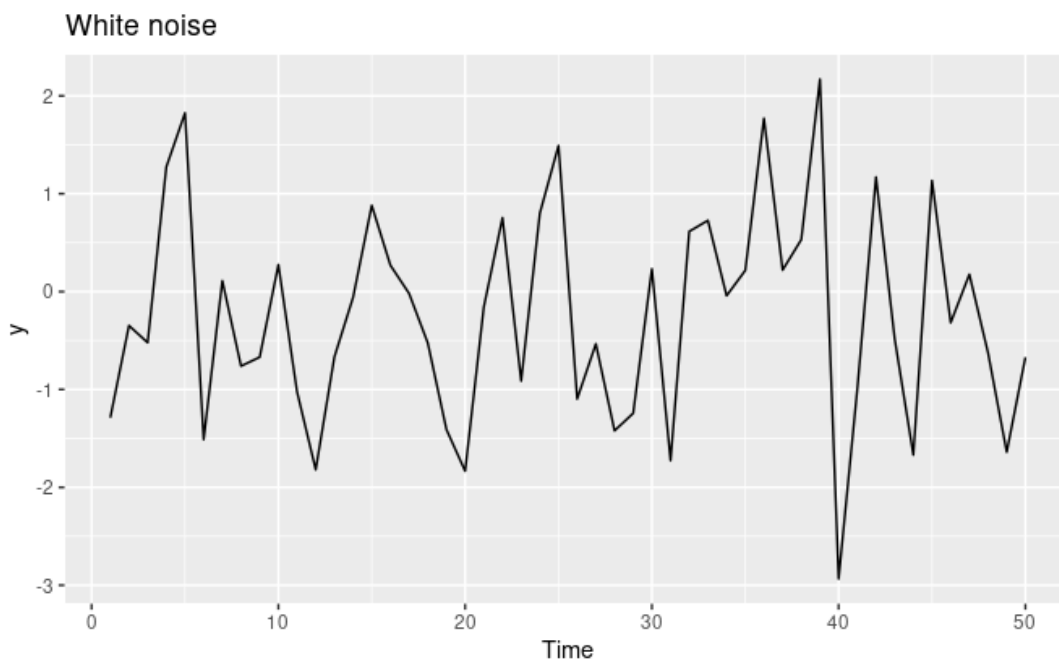


Рисунок 2.5 Графік часового ряду, що є білим шумом [45]

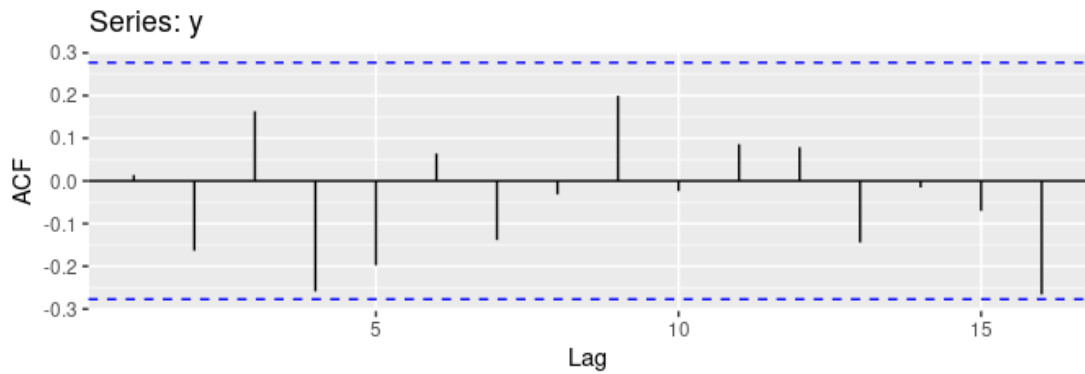


Рисунок 2.6 Корелограма білого шуму [45]

### 2.3.6. Регресія

У найпростішому випадку регресійна модель дозволяє отримати лінійну залежність між прогнозованою змінною  $y$  та єдиною змінною-предиктором  $x$  [51].

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (2.10)$$

Наприклад, така штучна модель наведена на рис 2.7. Коефіцієнти  $\beta_0$  і  $\beta_1$  позначають перетин і нахил прямої відповідно. Перетин  $\beta_0$  представляє передбачуване значення  $y$ , коли  $x = 0$ . Нахил  $\beta_1$  являє собою середню прогнозовану зміну  $y$  в результаті збільшення  $x$  на одну одиницю.

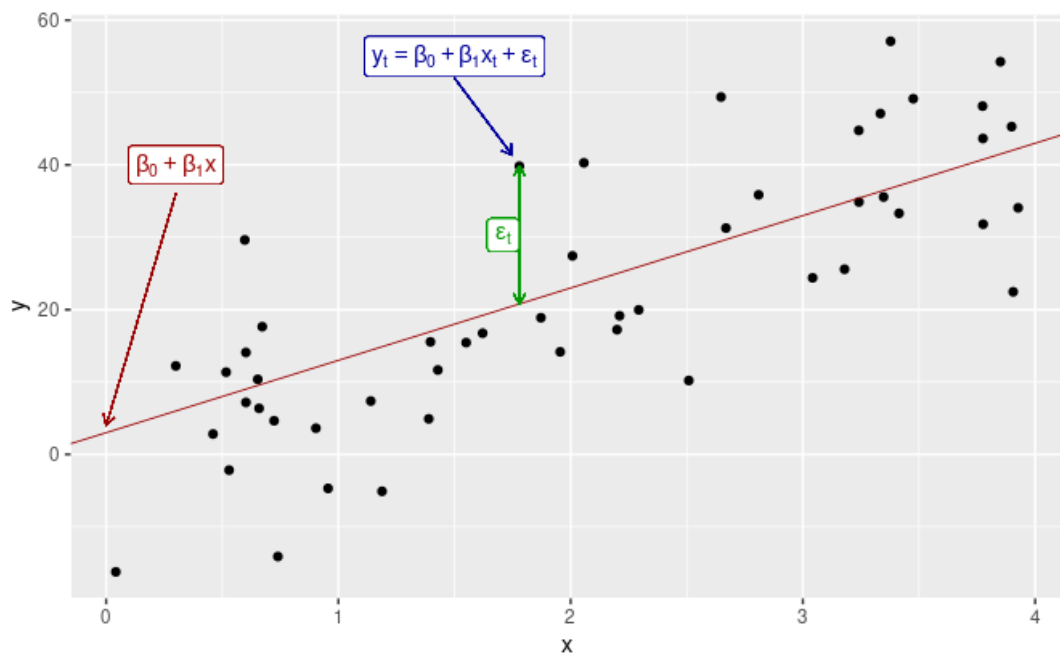


Рисунок 2.7 Графік деякої функції для демонстрації параметрів регресії [45]

Спостереження не лежать на прямій, а розкидані навколо неї. Ми можемо розглядати кожне спостереження  $y_t$  як таке, що складається із систематизованої або пояснюваної частини моделі  $\beta_0 + \beta_1 x_1$  і похибки  $\varepsilon_t$ . Термін похибка означає не помилку, а відхилення від базової прямолінійної моделі. Він фіксує все, що може вплинути на  $y_t$ , крім  $x_t$ .

Коли є дві або більше змінних-предикторів, модель називається моделлю множинної регресії. Загальною формою моделі множинної регресії є

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t \quad (2.11)$$

де  $y$  – змінна, яку слід прогнозувати,  $x_1, \dots, x_k$  –  $k$  змінних-предикторів. Кожна змінна-предиктор має бути числовою. Коефіцієнти  $\beta_0, \dots, \beta_t$  вимірюють ефект кожного предиктора після врахування ефектів усіх інших предикторів у моделі. Таким чином, коефіцієнти вимірюють граничні ефекти змінних-предикторів [45].

Коли ми використовуємо модель лінійної регресії, ми неявно робимо деякі припущення щодо змінних у рівнянні. По-перше, ми припускаємо, що модель є розумним наближенням до реальності; тобто зв'язок між прогнозованою змінною та змінними-предикторами задовольняє це лінійне рівняння. По-друге, ми робимо такі припущення щодо похибок:

- Їхнє середнє дорівнює нулю; інакше прогнози будуть систематично упередженими.
- Вони не є автокорельованими; інакше прогнози будуть неефективними, оскільки в даних є більше інформації, яку можна використати.
- Вони не пов'язані зі змінними-предикторами; інакше було б більше інформації, яка повинна бути включена в систематичну частину моделі.

Також корисно те, що якщо помилки розподіляються з постійною дисперсією  $\sigma^2$ , то легко отримати інтервали прогнозування.

Іншим важливим припущенням у моделі лінійної регресії є те, що кожен предиктор  $x$  не є випадковою величиною. Якби ми проводили контрольований

експеримент у лабораторії, ми могли б контролювати значення кожного  $x$  (щоб вони не були випадковими) і спостерігати отримані значення  $y$ . Отримавши дані спостережень (включаючи більшість даних у бізнесі та економіці), неможливо контролювати значення  $x$ , тому ми просто спостерігаємо і тільки припускаємо.

### 2.3.7. Експоненціальне згладжування

Експоненціальне згладжування було запропоновано наприкінці 1950-х [45], результуючи у найбільш успішні методи прогнозування. Прогнози, створені з використанням експоненціальних методів згладжування, є зваженими середніми показниками минулих спостережень, при цьому ваги експоненціально зменшуються у міру того, як спостереження старіють. Іншими словами, чим спостереження нещодавніше, тим вища його вага. Цей підхід генерує надійні прогнози швидко та на широкий діапазон часових рядів, що є великою перевагою для застосувань у промисловості.

Найпростіший з експоненціальних методів згладжування природно називається простим експоненціальним згладжуванням (SES). Цей метод підходить для прогнозування даних без чітких тенденцій або сезонної закономірності.

Використовуючи наївний метод, усі прогнози на майбутнє дорівнюють останньому спостережуваному значенню серії (2.5) для  $h=1,2,\dots$ . Цей метод передбачає, що останнє спостереження – єдине важливе, а всі попередні спостереження не дають інформації про майбутнє. Це можна сприймати як середньозважене, де вся вага дається останньому спостереженню.

Використовуючи метод середніх, усі майбутні прогнози дорівнюють простому середньому серед спостережуваних даних, тобто

$$\hat{y}_{T+h|T} = \frac{1}{T} \sum_{t=1}^T y_t \quad (2.12)$$

для  $h=1,2,\dots$ . Цей метод передбачає, що всі спостереження мають однакову важливість, і надає їм однакові ваги при формуванні прогнозів.

Насправді, істина десь посередині. Наприклад, доцільніше прикріпити більші ваги до останніх спостережень, ніж до спостережень з далекого минулого. Саме така концепція стоїть за простим експоненціальним згладжуванням. Прогнози обчислюються із зваженими середніми, де ваги експоненціально зменшуються, оскільки спостереження походять з минулого. Найменші ваги пов'язані з найдавнішими спостереженнями:

$$\hat{y}_{T+h|T} = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \dots, \quad (2.13)$$

де  $0 \leq \alpha \leq 1$  параметри згладжування. Прогноз на один крок уперед  $T + 1$  – це середньозважене значення всіх спостережень у серії  $y_1, \dots, y_T$ . Швидкість зниження ваги контролюється параметром  $\alpha$ .

### 2.3.8. Моделі ARIMA

Моделі ARIMA (AutoRegressive Integrated Moving Average) забезпечують інший підхід до прогнозування часових рядів. Експоненціальне згладжування та моделі ARIMA є двома найбільш широко використовуваними підходами до прогнозування часових рядів та забезпечують додаткові підходи до проблеми. Експоненціальні моделі згладжування базуються на описі тенденції та сезонності даних, у той час як моделі ARIMA описують автокореляції в даних.

#### 2.3.8.1. Стаціонарність

Стаціонарний часовий ряд – це такий ряд, властивості якого не залежать від часу, протягом якого він спостерігається. Таким чином, часові ряди з тенденціями чи сезонністю не є стаціонарними. Вони впливатимуть на значення часового ряду в різний час. З іншого боку, часовий ряд білих шумів стаціонарний – час його спостереження неважливий, оскільки ряд повинен виглядати приблизно однаково у будь-який момент часу.

Проте часовий ряд із циклічною поведінкою (але без тенденції та сезонності) є стаціонарним. Це пов'язано з тим, що цикли не мають фіксованої довжини, тому перед тим, як ми спостерігатимемо ряд, ми не можемо бути впевнені, де будуть локальні екстремуми циклів.

Загалом, стаціонарні часові ряди не матимуть передбачуваних закономірностей у довгостроковій перспективі. Графіки часу покажуть, що ряд є приблизно горизонтальним (хоча можливі певні циклічні поведінки), з постійною дисперсією.

Наприклад, на рис. 2.8 наведені графіки 9 різних часових рядів: (a) ціна акцій Google протягом 200 днів поспіль; (b) Щоденна зміна ціни акцій Google протягом 200 днів поспіль; (c) Щорічна кількість страйків у США; (d) щомісячні продажі нових будинків на одну сім'ю, що продаються в США; (e) річна ціна десятка яєць у США (у доларах США); (f) щомісячна кількість свиней, забитих у штаті Вікторія, Австралія; (g) річна кількість рисі, що потрапила в район річки Маккензі на північному заході Канади; (h) щомісячне виробництво австралійського пива; (i) Щомісячне виробництво електроенергії в Австралії. Необхідно визначити стаціонарні часові ряди.

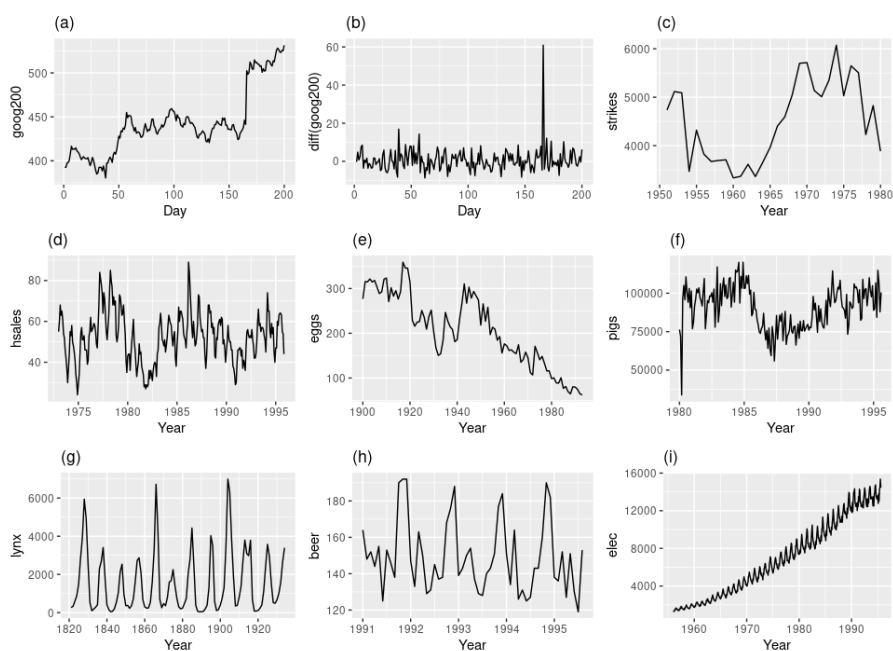


Рисунок 2.8 Приклади 9 часових рядів [45]

Очевидна сезонність виключає серії (d), (h) та (i). Тенденції та зміна рівнів виключають ряди (a), (c), (e), (f) та (i). Збільшення дисперсії також виключає ряд (i). Тому залишаються лише (b) та (g) як стаціонарні часові ряди.

Помітні цикли ряду (g) можуть здатися нестаціонарними. Але ці цикли є періодичними. Вони спричинені великою популяцією рисі для наявного корму, що призводить до припинення розмноження, так що популяція падає, тоді відновлення джерел їжі дозволяє популяції зростати знову і так далі. У довгостроковій перспективі терміни цих циклів непередбачувані. Отже, ряд є стаціонарним.

### 2.3.8.2. Різниця (Differencing)

На рис. 2.8. секція (a) показує нестаціонарний ряд, у той час як секція (b) – стаціонарний (щоденні зміни акції Google). Це показує один із способів зробити нестаціонарний часовий ряд стаціонарним. Достатньо обчислити різницю між послідовними спостереженнями. Це відомо як операція різниці (differencing).

Такі перетворення, як логарифмування, можуть допомогти стабілізувати дисперсію часового ряду. Різниця може допомогти стабілізувати середнє значення часового ряду, усуваючи зміни рівня часового ряду, а отже, усуваючи (або зменшуючи) тенденцію та сезонність.

Окрім аналізу часового графіку даних, корелограма також корисна для виявлення нестаціонарних часових рядів. Для стаціонарних часових рядів графік автокореляції збігається до нуля відносно швидко, тоді як для нестаціонарних даних він повільно зменшується. Також для нестаціонарних даних значення першого коефіцієнту автокореляції  $r_1$  часто є великим додатнім значенням.

Корелограма різниць цін акцій Google виглядає точно так само, як і серії білого шуму: там відсутні автокореляції, які б виходили за межі 5% (рис. 2.9), а статистика Лjung-Бокса  $Q^*$  має  $p=0,355$  (для  $h=10$ ). Це свідчить про те, що щоденна зміна ціни акцій Google, по суті, є випадковою і не співвідноситься (корелює) з попередніми днями.

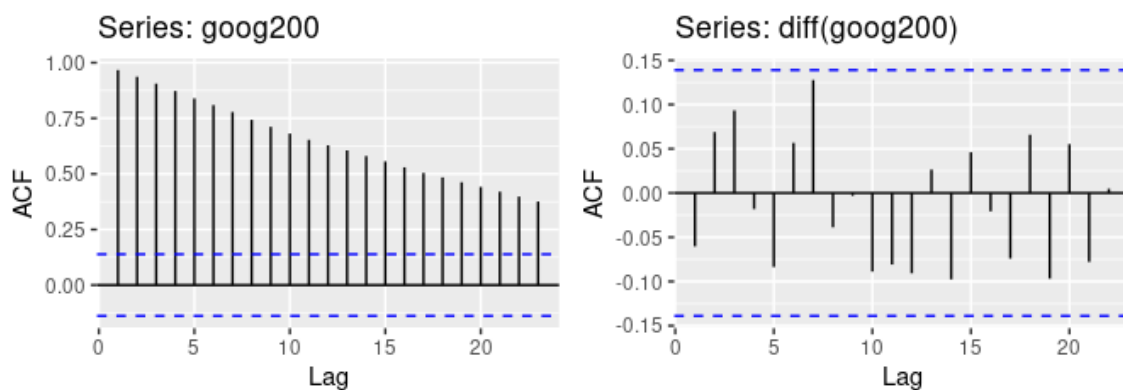


Рисунок 2.9 Порівняння корелограм різниць нестационарного (ліворуч) та стаціонарного (праворуч) рядів [45]

### 2.3.8.3. Авторегресійні моделі

У моделі множинної регресії значення прогнозується, використовуючи лінійну комбінацію предикторів. У моделі авторегресії значення залежної змінної прогнозується, використовуючи лінійну комбінацію попередніх значень цієї змінної. Термін авторегресія вказує на те, що це регресія змінної щодо себе.

Модель авторегресії порядку  $p$  можна записати як 2.14

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t, \quad (2.14)$$

де  $\varepsilon_t$  – це білий шум. Це як множинна регресія, але з попередніми (lagged) значеннями  $y_t$  як предикторами. Вона називається моделлю AR(p), тобто авторегресійною моделлю порядку  $p$ .

Авторегресивні моделі надзвичайно гнучкі при роботі з широким спектром різних моделей часових рядів [28]. Два часових ряди на рис. 2.10 показують ряди з моделі AR(1) та AR(2). Зміна параметрів  $\phi_1, \dots, \phi_p$  призводить до різних моделей часових рядів. Дисперсія помилки  $\varepsilon_t$  зміниться лише при масштабуванні ряду, а не закономірностей.

Для моделі AR(1):

- коли  $\phi_1 = 0$ ,  $y_t$  еквівалентно білому шуму;
- коли  $\phi_1 = 1$  та  $c = 0$ ,  $y_t$  еквівалентно випадковій ходьбі;
- коли  $\phi_1 = 1$  і  $c \neq 0$ ,  $y_t$  еквівалентно випадковому блуканню з дрейфом;

- коли  $\phi_1 < 0$ ,  $y_t$  має тенденцію коливатися навколо середнього.

Авторегресивні моделі зазвичай обмежуються стаціонарними даними, і в цьому випадку потрібні певні обмеження на значення параметрів:

- Для моделі AR(1):  $-1 < \phi_1 < 1$ ;
- Для моделі AR(2):  $-1 < \phi_2 < 1, \phi_1 + \phi_2 < 1, \phi_2 - \phi_1 < 1$ .

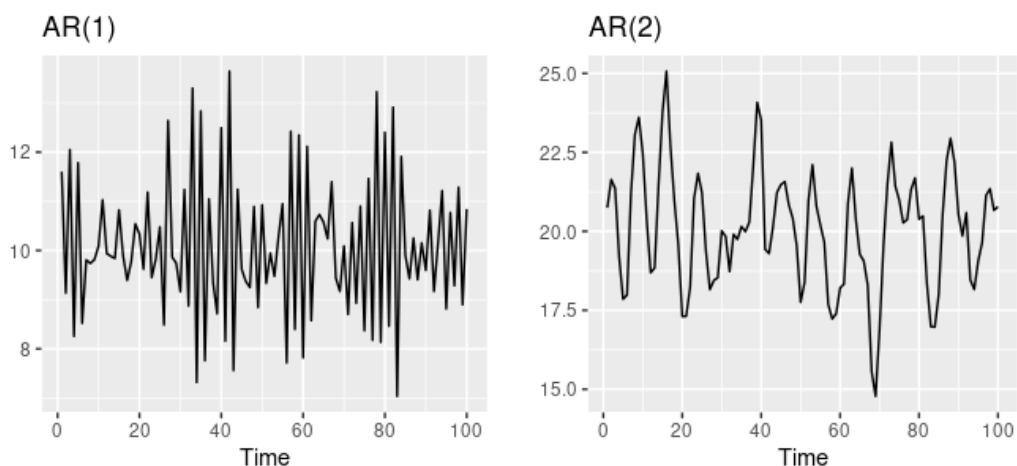


Рисунок 2.10 Приклад авторегресійних моделей. Ліва  $AR(1) = y_t = 18 + 0,8y_{t-1} + \varepsilon_t$ . Права  $AR(2) = y_t = 8 + 1,3y_{t-1} + 0,7y_{t-2} + \varepsilon_t$ . В обох моделях  $\varepsilon_t$  нормально розподілений білий шум із середнім 0 та дисперсією 1 [45]

#### 2.3.8.4. Моделі ковзного середнього

Замість того, щоб використовувати попередні значення прогнозованої змінної в регресії, модель ковзного середнього використовує попередні помилки прогнозу в регресійній моделі.

$$y_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_p\varepsilon_{t-p}, \quad (2.15)$$

де  $\varepsilon_t$  – білий шум. Це модель ковзного середнього  $q$ -порядку MA( $q$ ). Тут не спостерігаються значення  $\varepsilon_t$ , тож насправді це не зовсім регресія у звичному розумінні.

Кожне значення  $y_t$  можна сприймати як зважене ковзне середнє серед кількох останніх похибок прогнозу. Однак моделі ковзного середнього не слід плутати зі згладжуванням ковзного середнього: модель ковзного середнього

використовується для прогнозування майбутніх значень, тоді як згладжування ковзного середнього використовується для оцінки цикл-тренду минулих значень.

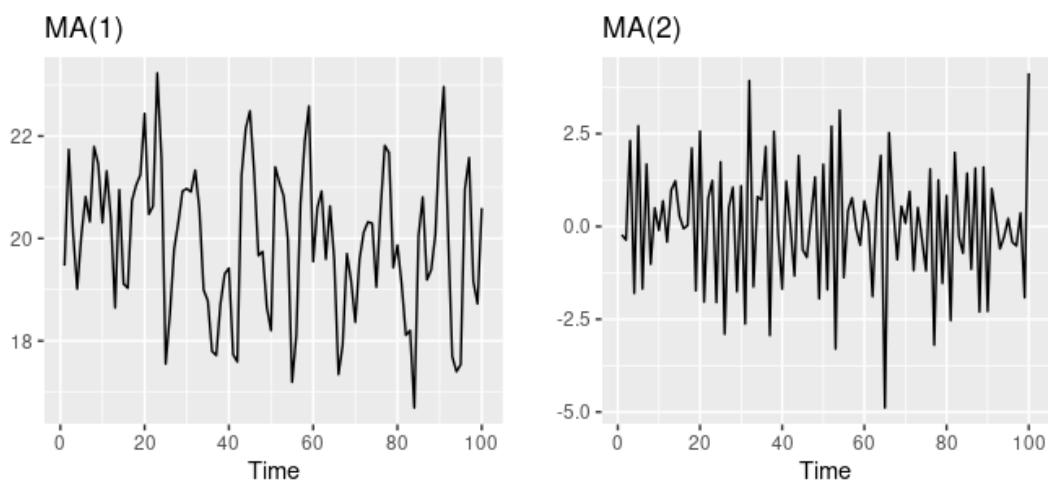


Рисунок 2.11 Приклад MA моделей з різними параметрами. Ліва  $MA(1) = y_t = 20 + \varepsilon_t + 0,8\varepsilon_{t-1}$ . Права  $MA(2) = y_t = \varepsilon_t - \varepsilon_{t-1} + 0,8\varepsilon_{t-2}$ . В обох моделях  $\varepsilon_t$  нормально розподілений білий шум із середнім 0 та дисперсією 1 [45]

Як і в авторегресивних моделях, дисперсія похибки  $\varepsilon_t$  буде змінюватися лише при масштабуванні часового ряду, а не його закономірностей.

Можна записати будь-яку стаціонарну модель  $AR(p)$  як модель  $MA(\infty)$ . Наприклад, використовуючи повторну заміну, ми можемо продемонструвати це для моделі  $AR(1)$ : (2.16)

$$y_t = \phi_1 y_{t-1} + \varepsilon_t = \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t = \dots \quad (2.16)$$

За умови  $-1 < \phi_1 < 1$ , значення  $\phi_k$  буде меншим, оскільки  $k$  стає більшим. Тож в решті-решт ми отримуємо

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \dots, \quad (2.17)$$

тобто процес  $MA(\infty)$ .

Зворотний результат має місце, якщо ми накладаємо деякі обмеження на параметри MA. Тоді модель MA називається оберненою. Тобто ми можемо записати будь-який зворотний процес  $MA(q)$  як процес  $AR(\infty)$ . Обернені моделі вводяться не просто, щоб дозволити нам перейти з моделей MA на моделі AR. Вони також мають деякі інші математичні властивості [45].

### 2.3.8.5. Несезонні моделі ARIMA

Якщо поєднати операцію різниці з авторегресією та моделлю ковзного середнього, ми отримуємо несезонну модель ARIMA. ARIMA є аббревіатурою від AutoRegressive Integrated Moving Average (у цьому контексті «інтеграція» протилежне до різниці). Повну модель можна записати як

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_q \varepsilon_{t-q} + \dots + \varepsilon_t \quad (2.18)$$

де  $y'_t$  – диференційований (можливо, і не раз) ряд. До «предикторів» з правого боку належать як попередні значення  $y_t$ , так і попередні помилки. Це і є моделлю ARIMA( $p, d, q$ ), де  $p$  – порядок авторегресивної частини,  $d$  – ступінь різниці;  $q$  – порядок частини ковзного середнього.

Ті самі умови стаціонарності та оберненості, що використовуються для моделей авторегресії та ковзного середнього, також застосовуються до моделі ARIMA.

Деякі з вище описаних моделей є особливими випадками моделі ARIMA:

- Білий шум – ARIMA(0,0,0)
- Випадкове блукання – ARIMA(0,1,0) без константи
- Випадкове блукання з дрейфом – ARIMA(0,1,0) з константою
- Авторегресія – ARIMA( $p,0,0$ )
- Ковзне середнє – ARIMA(0,0, $q$ )

Константа  $c$  має важливий вплив на довгострокові прогнози, отримані за допомогою цих моделей:

- Якщо  $c = 0$  і  $d = 0$ , довгострокові прогнози будуть нульовими.
- Якщо  $c = 0$  і  $d = 1$ , довгострокові прогнози будуть мати ненульову константу.
- Якщо  $c = 0$  і  $d = 2$ , довгострокові прогнози будуть прямими лініями.
- Якщо  $c \neq 0$  і  $d = 0$ , довгострокові прогнози будуть переходити до середнього значення даних.
- Якщо  $c \neq 0$  і  $d = 1$ , довгострокові прогнози будуть прямими.

- Якщо  $c \neq 0$  і  $d = 2$ , довгострокові прогнози будуть дотримуватися квадратичної тенденції.

Значення  $d$  також впливає на інтервали прогнозування. Чим вище значення  $d$ , тим швидше збільшуються розміри інтервалів прогнозування. Для  $d = 0$  стандартне відхилення довгострокового прогнозу перейде до стандартного відхилення історичних даних, тому інтервали прогнозування по суті будуть однаковими.

#### 2.3.8.6. Сезонні моделі ARIMA

Сезонна модель ARIMA формується шляхом включення додаткових сезонних термів до несезонних моделей ARIMA, тобто  $ARIMA(p, d, q)(P, D, Q)_m$ , де  $(p, d, q)$  – несезонна частина моделі,  $(P, D, Q)$  – сезонна частина моделі,  $m$  – кількість спостережень на рік.

Сезонна частина моделі складається з термів, подібних на несезонні компоненти моделі, але передбачає зворотний зсув сезонного періоду. Наприклад, модель  $ARIMA(1,1,1)(1,1,1)_4$  (без константи) призначена для щоквартальних даних ( $m = 4$ ) і може бути записана як

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)y_t = (1 + \phi_1 B)(1 + \Theta_1 B^4)\varepsilon_t \quad (2.19)$$

де  $B$  – оператор зворотного зсуву, тобто  $B y_t = y_{t-1}$ ,  $B^4 y_t = y_{t-4}$ , а додаткові сезонні умови просто множаться на несезонні.

## РОЗДІЛ 3. АНАЛІЗ ДІЯЛЬНОСТІ КОМЕРЦІЙНОЇ ОРГАНІЗАЦІЇ

### 3.1. Підготовка даних та аналітична звітність

Для аналізу та прогнозування діяльності комерційної організації, у нашому розпорядженні є база даних продажів магазину будівельних товарів Будмаркет. Ця база даних (БД) складається з 3 таблиць – Клієнти, Продажі, Приход.

Аналіз та прогнозування діяльності магазину буде проводитися у програмному забезпеченні Loginom, що є наступником аналітичної платформи Deductor Studio.

Для того, щоб почати аналізувати продажі магазину, необхідно спочатку створити новий аналітичний сценарій та підключити цю БД у Loginom. Оскільки база даних, що розглядається, експортована з Deductor у спеціальному сховищі даних Deductor Warehouse і доступна у форматі Firebird, налаштуємо підключення до цього сховища за допомогою вбудованих засобів Loginom (рис. 3.1).

Рисунок 3.1 Вікно підключення до сховища даних

Після успішного підключення до БД, розглянемо таблицю Продаж. Вона містить інформацію про клієнтів і їхні покупки лише за 2019-ий рік, а саме за період березень-грудень 2019 року. Всього ж у базі даних магазину нараховується 8509 продаж (табл. 3.1, рис. 3.2), що цілком достатньо для аналізу та прогнозування діяльності невеликої організації.

Таблиця 3.1 Фрагмент записів (фактів) таблиці «Продажі»

№ клієнта	Група клієнта	Місто	Сума продажу (грн)	Дата продажу
00000010	Постійний клієнт	Чернігів	16935,67	01.03.2019
00000015	Постійний клієнт	Чернігів	275582,41	01.03.2019
00000016	Клієнт	Чернігів	41323,80	01.03.2019
00000081	VIP клієнт	Ірпінь	201650,41	01.03.2019
00001158	Клієнт	Бориспіль	13708,80	01.03.2019
00001160	VIP клієнт	Київ	1278,09	01.03.2019

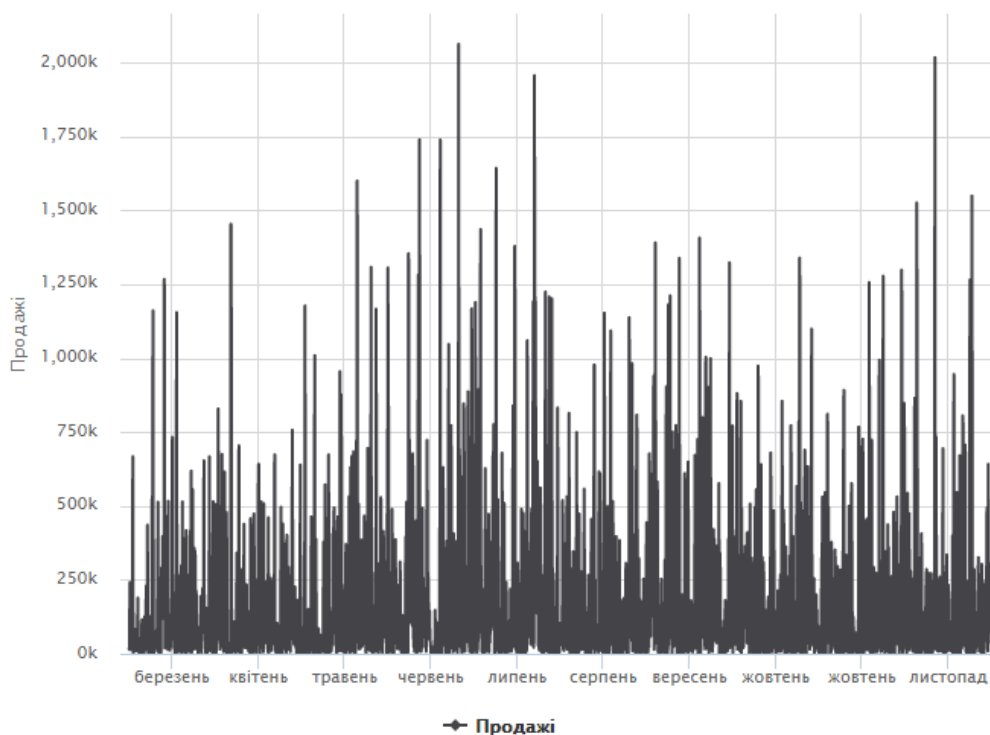


Рисунок 3.2 Графік продаж за березень-листопад 2019 року

Відповідно до технології Knowledge Discovery in Databases (KDD), одним із перших кроків у видобуванні знань має бути очистка та перетворення даних до потрібного формату. У нашому випадку графік продаж не дає зрозуміти повної ситуації щодо діяльності магазину, тому звернемося до інструментів бізнес-аналітичної звітності – розрізу даних, OLAP-кубу, крос-таблиць, статистичних показників, тощо – для визначення коректного формату даних.

Виберемо ключових клієнтів магазину. Зазвичай, це 50% клієнтів, що принесли дохід за певний період [23]. Це зручно зробити, використовуючи віджет-візуалізатор OLAP-куба.

Для цього створимо новий вузол сценарію і задаймо поля «Група клієнта», «Номер клієнта» вимірами, а «Сума зі знижкою» фактами куба. Для того щоб у звіті були дані тільки по клієнтам, які приносять 50% доходу, налаштуємо фільтр фактів (записів), у якому зазначена умова відбору клієнтів – частка від загальної суми повинна становити 50% продаж за увесь розглянутий період продаж. Результат наведено на рис. 3.3.

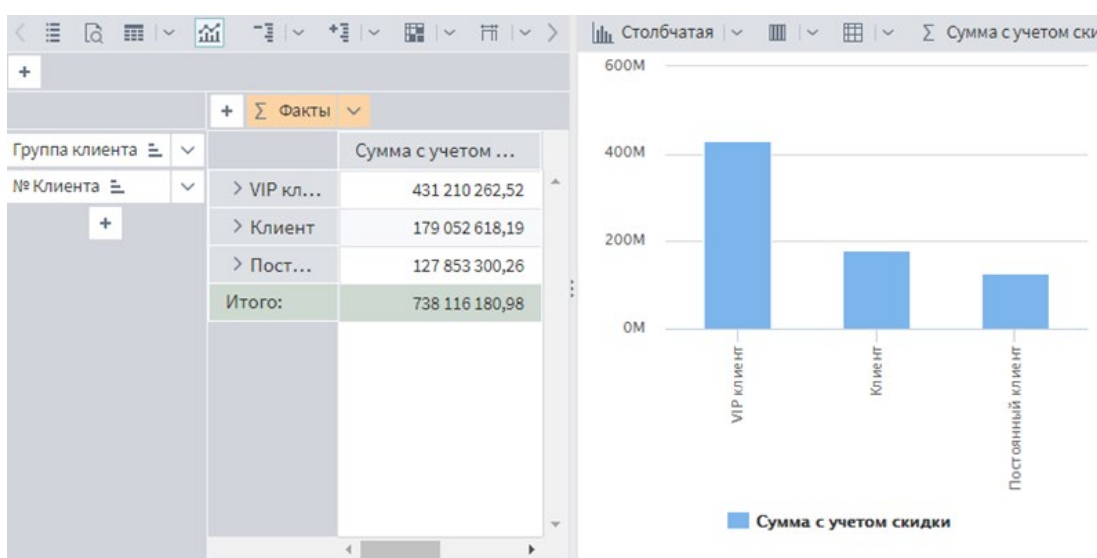


Рисунок 3.3 Звіт по клієнтам, що забезпечують 50% доходу магазину з продаж

Із рисунка видно, що приблизно половина всіх важливих клієнтів – це VIP-клієнти, тому магазину рекомендовано звернути більше уваги саме на них для детальної сегментації, з метою їх стимулювання та залучення до повторних покупок.

### 3.2. Аналіз та прогнозування продаж магазину будівельних матеріалів

*Аналіз сезонності та тренду продаж.* «Як і більшість інших видів аналізу, аналіз часових рядів передбачає, що дані містять систематичну складову (зазвичай включає кілька компонент) і випадковий шум (помилку, похибку), який ускладнює виявлення регулярних компонент. Більшість методів

дослідження часових рядів включає різні способи фільтрації шуму, що дозволяють побачити регулярну складову більш чітко» [28].

У нашому випадку, таким часовим рядом є факти продаж магазину [24]. Цей ряд доступний у вигляді записів таблиці «Продажі» бази даних магазину будівельних матеріалів. Для ефективної обробки такого ряду, відкинемо всі атрибути таблиці «Продажі», окрім двох – «Дата продажу» та «Сума продажу», та додаймо вимірності даних у вигляді деталізації атрибуту «Дата продажу» на день, тиждень та місяць продажу. Це може допомогти відшукати закономірності продаж на різних рівнях [15]. Графіки щоденних, щотижневих та щомісячних продаж наведені на рис. 3.4-3.6 відповідно.

Графік щоденних продаж порідшав і тепер відрізняється від графіку всіх продаж магазину. Це зумовлено групуванням і сумуванням всіх продаж за конкретний день. По осі абсцис розташовані порядкові номери дня продаж у календарному 2019 році. При цьому графік не виділяється жодними закономірностями візуально, окрім декількох вершин (піків) у його другій половині з кроком у 50 днів. Це може свідчити про деяку сезонність.

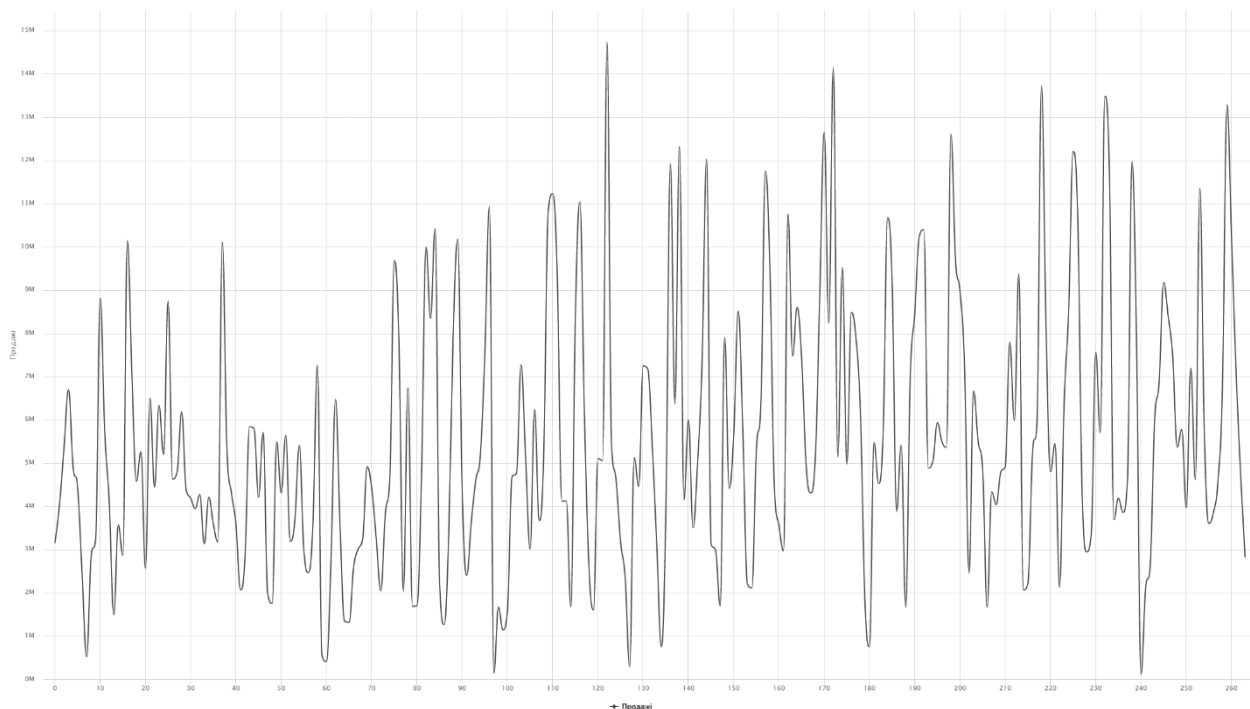


Рисунок 3.4 Щоденні продажі магазину будівельних матеріалів за березень-грудень 2019 року

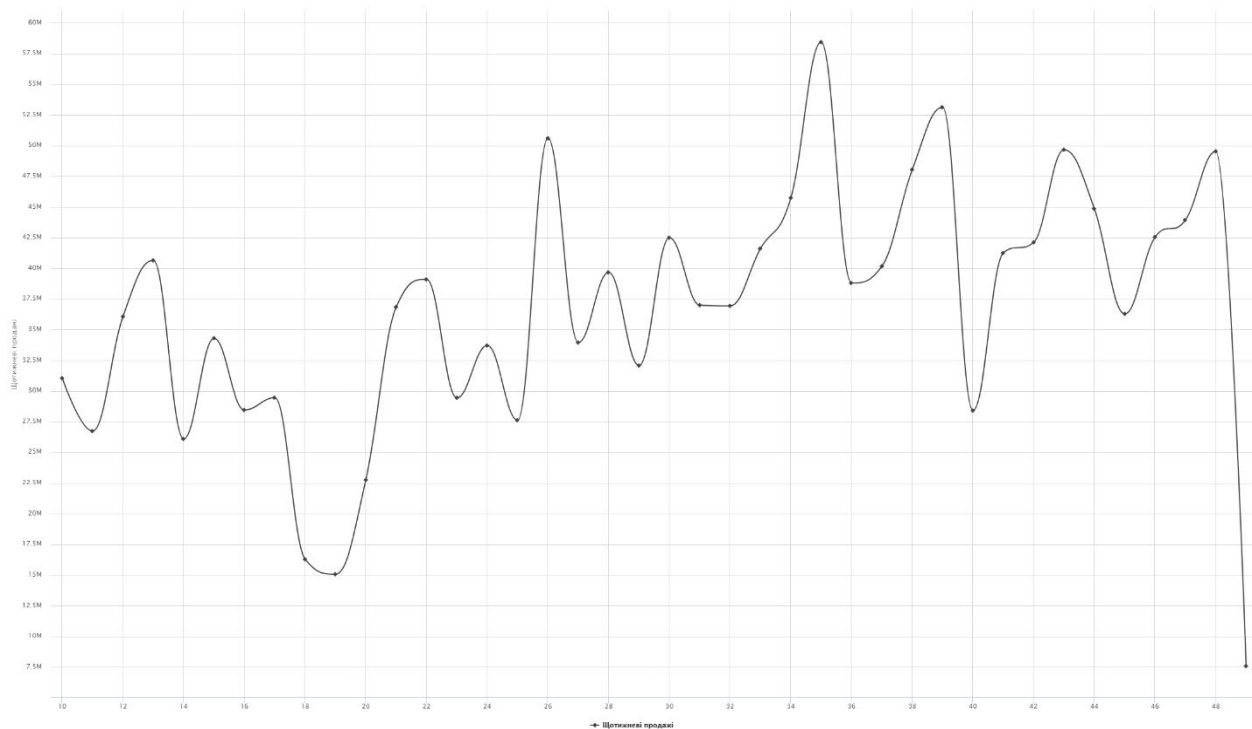


Рисунок 3.5 Щотижневі продажі магазину за березень-грудень 2019 року

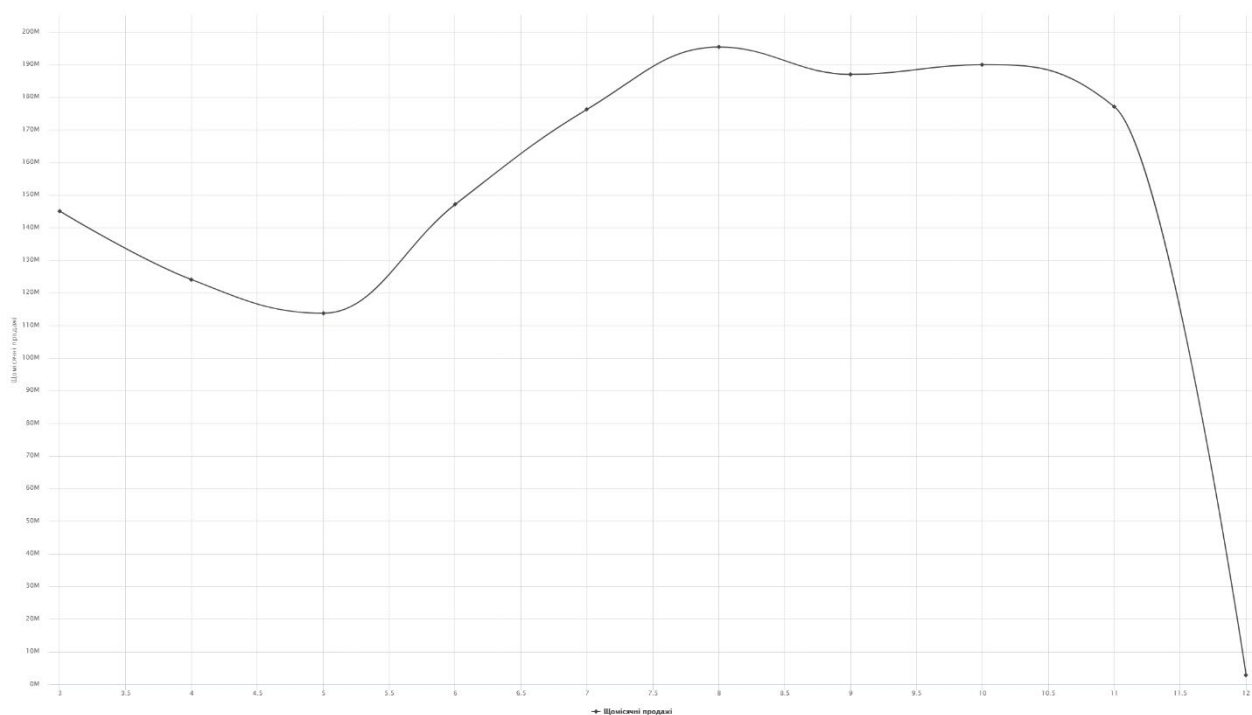


Рисунок 3.6 Щомісячні продажі магазину за березень-грудень 2019 року

Графіки щотижневих та щомісячних продаж здаються, на перший погляд, лаконічними, проте мають спільну рису – продажі різко падають у грудні місяці. Переглянувши факти продаж магазину за грудень 2019 року, з'ясувалося, що в БД магазину за грудень записано лише 17 продаж і тільки за його перший день.

У порівнянні зі 1000 записів (у середньому) для інших місяців, це означає, що дані за грудень можуть зсунути негативну оцінку аналізу продаж у сторону останнього кварталу 2019 року, тому не повинні бути враховані при аналізі взагалі.

З цією метою скористаємося ще однією трансформацією даних для фільтрації і відокремлення продаж магазину будівельних матеріалів за грудень. У результаті отримуємо 8492 записи тільки в період березня-листопада 2019 року. Для подальшого аналізу будемо використовувати дані за цей період.

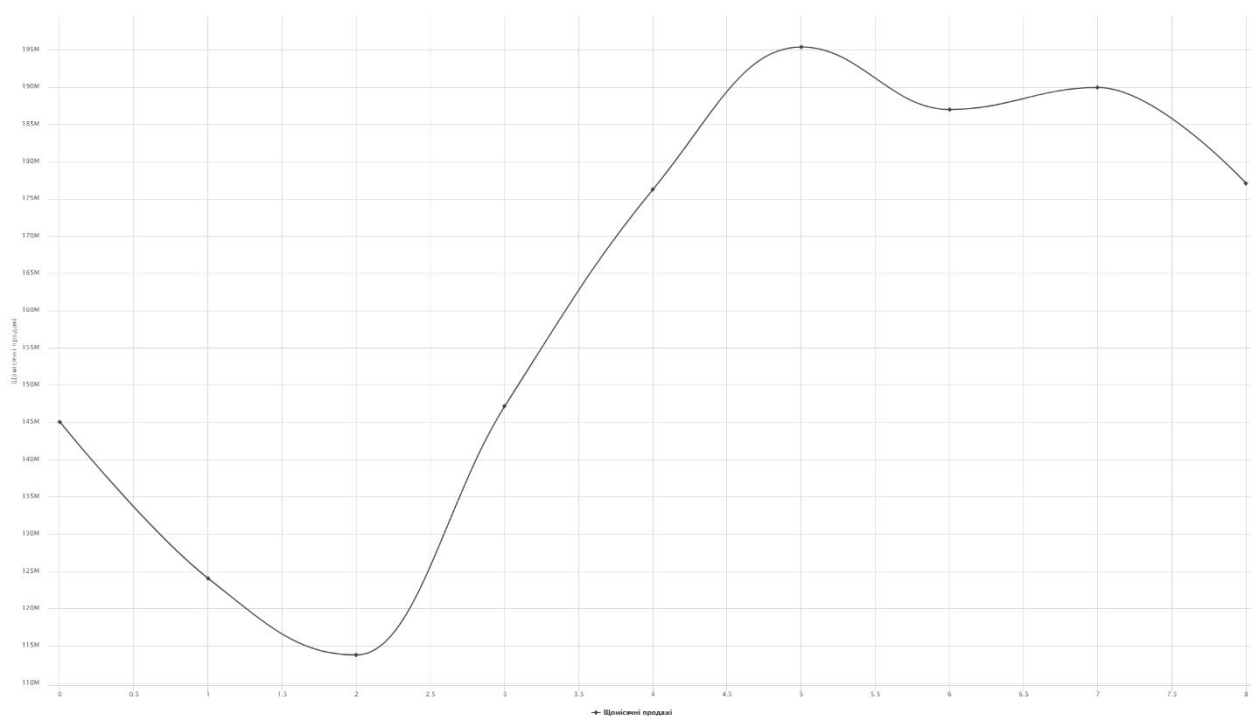


Рисунок 3.7 Графік щомісячних продаж магазину за період березень-грудень 2019 року

Оновлений графік щомісячних продаж магазину, що наведений на рис 3.7, нагадує за формою дзвін або ковбойську шляпу (рис. 3.8). Можна припустити, що продажі магазину за період березень 2019 року – березень 2020 року матимуть нормальний розподіл.

Для щоденних (263) та щотижневих (40) продаж ряд містять достатню кількість (значення у дужках) записів, спостережень для аналізу та прогнозування. У випадку щомісячних продаж, для аналізу залишається тільки 9 спостережень за березень-листопад 2019 року відповідно, що очевидно замало

для ґрунтового аналізу. Тому далі для аналізу продаж як часового ряду будуть використовуватися записи лише щоденних та щотижневих продаж.

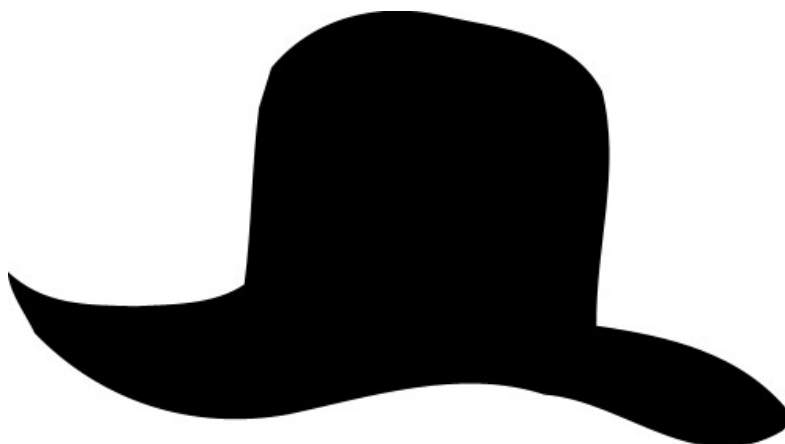


Рисунок 3.8 Силует ковбойської шляпи як візуалізація можливої траєкторії щомісячних продаж магазину

Більшість регулярних складових часових рядів належить до двох класів: вони є або трендом, або сезонною складовою. Тренд представляє загальну систематичну лінійну або нелінійну компоненту, яка може змінюватися в часі. Сезонністю, або сезоном, називається складова часового ряду, яка описує регулярні зміни його значень в межах деякого періоду і представляє собою послідовність майже повторюваних циклів. Обидва ці види регулярних компонент часто присутні в ряді одночасно.

Не існує «автоматичного» способу виявлення тренду в часовому ряді. Однак якщо тренд є монотонним (строго зростає або спадає), то аналізувати такий ряд зазвичай неважко. Якщо часові ряди містять значну похибку, то першим кроком виділення тренду є згладжування.

Для того щоб зрозуміти чи містить потік продаж магазину значну похибку, оцінимо модель за допомогою автокореляційного аналізу. Для цього побудуємо автокореляцію перших 60 днів та 9 тижнів (2 місяці в середньому) продаж (рис. 3.9-3.10).

Сезонний компонент може бути прив'язаний до певного календарного часового інтервалу: дня, тижня, кварталу, місяця або року, а може бути

пов'язаний з якою-небудь подією, яка безпосередньо не має відношення до календарного інтервалу. В економіці багато явищ характеризуються періодично повторюваними сезонними ефектами. Відповідно, часові ряди, що відображають ці сезонні ефекти, містять періодичні коливання.

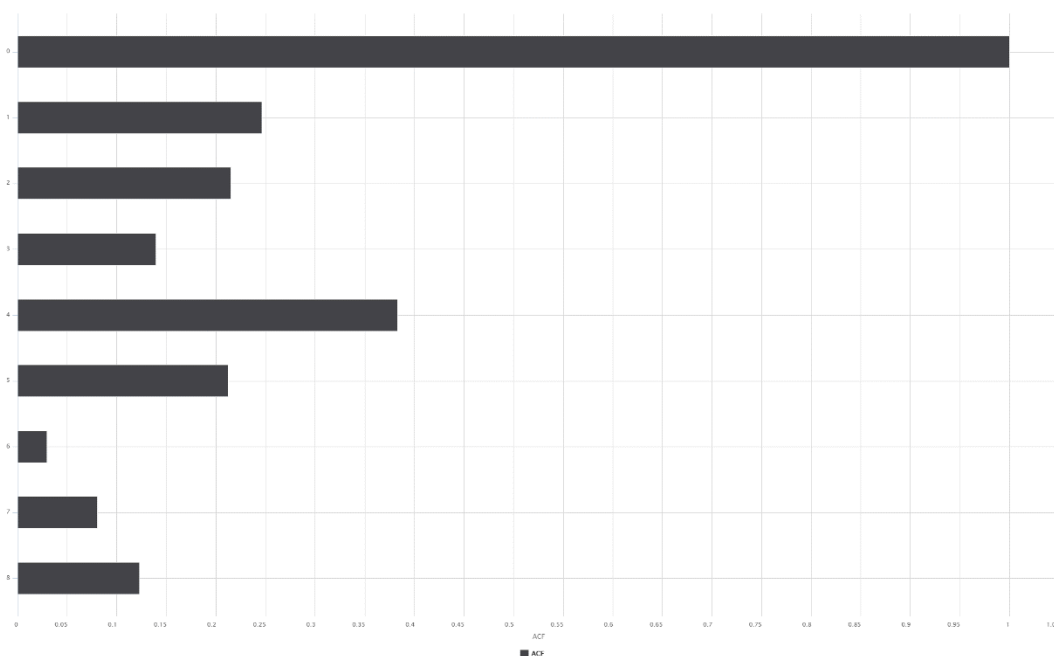


Рисунок 3.9 Графік функції автокореляції продаж магазину за перших 9 тижнів

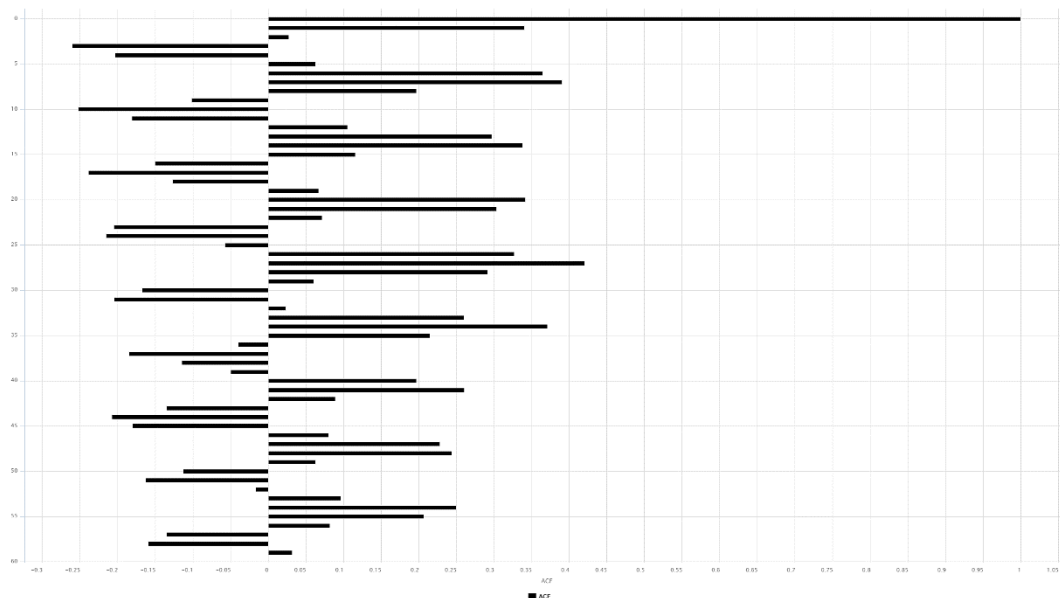


Рисунок 3.10 Графік функції автокореляції продаж магазину за перших 60 днів

Корелограма щоденних продаж описує деякий поки невідомий цикл, у той час як графік автокореляційної функції щотижневих продаж магазину свідчить про наявність позитивно-спадного тренду.

На перший погляд, продажі повторюються кожні 3-4 дні. Проте якщо зіставити значення автокореляційної функції та календар 2019 року, починаючи з 1-го березня (п'ятниця), сезонність (сезонний цикл) щоденних продаж, швидше за все, становитиме 7 і пояснюється семиденним календарним (або п'ятиденним робочим) тижнем, тобто продажі під кінець тижня (п'ятниця- понеділок) будуть зростати, а посеред тижня – спадати.

Корелограму продаж за перші 9 тижнів зручніше передати у вигляді таблиці коефіцієнтів автокореляційної функції для кожного лагу (табл. 3.2).

Таблиця 3.2 Коефіцієнти автокореляції продаж за перші 9 тижнів

r1	r2	r3	r4	r5	r6	r7	r8	r9
1,00	0,25	0,21	0,14	0,38	0,21	0,03	0,08	0,12

З таблиці видно, що тижневі продажі не дуже корелюють між собою, принаймні на цьому проміжку, і щонайбільше є частиною більшого тренду, тому варто побудувати ще одну корелограму розміром усієї вибірки щотижневих продаж. Її наведено на рис. 3.11.

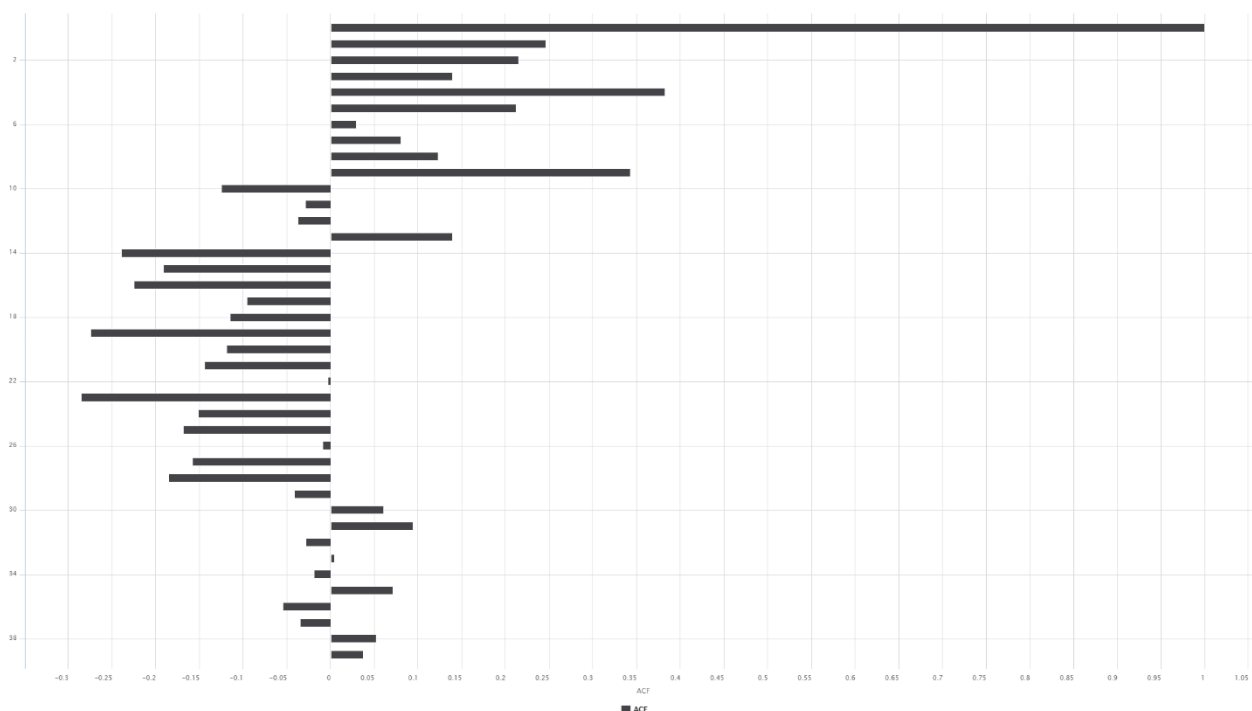


Рисунок 3.11 Корелограма всіх тижневих продаж магазину

За корелограмою всіх тижневих продаж можна припускати про наявність щоквартального тренду, тобто в середньому 14 тижнів, але насправді таких короткотривалих трендів майже не існує, тому це може бути щоквартальною сезонністю (циклом). Значення автокореляційної функції тенденційних часових рядів, як правило, має позитивні значення, які повільно зменшуються із збільшенням часу, оскільки спостереження у певний період часу також мають приблизно однаковий розмір.

Для перевірки цих припущень необхідно мати більше даних про продажі магазину, розподілених у часі принаймні на рік уперед. Можна також скористатися згладжуванням існуючих даних. Logiном надає інструмент «Згладжування» у вигляді фільтру Ходрика-Прескотта [52]. Для цього спочатку визначимо значення лямбда за наступною формулою:

$$\lambda = \left(\frac{f}{4}\right)^2 * 1600 \quad (3.1)$$

де  $f$  – це частота ряду за рік. Хоч і пропонуються інші підходи до обчислення  $\lambda$  [53; 54], скористаємося опцією Logiном, безпосередньо вказуючи частоту ряду, що дорівнює 52,17 для щотижневих записів. Графік згладженого ряду щотижневих продаж магазину наведено на рис. 3.12.

Як бачимо, що графік фільтру близький до графіку лінійної функції, а тому жодного тренду не існує у конкретному цьому ряді, що представляє продажі у вигляді 40 щотижневих записів.

Отже, підтверджено припущення про сезонну циклічність ряду щоденних продаж вікном у 7 днів, а також спростована гіпотеза про будь-який тренд ряду щотижневих продаж на вибірці 40 тижнів. Далі ці 2 ряди можуть бути розглянуті під час побудови прогнозних моделей.

*Побудова моделі прогнозування продаж.* Процедури оцінки параметрів і прогнозування за допомогою усунення шумів й аналізу сезонності та трендів, припускають, що математична модель процесу відома. У реальних даних часто немає яскраво виражених регулярних складових. Окремі спостереження містять

значну похибку, тоді як під час аналізу бажано виділити не тільки регулярні компоненти, а також і побудувати прогноз.

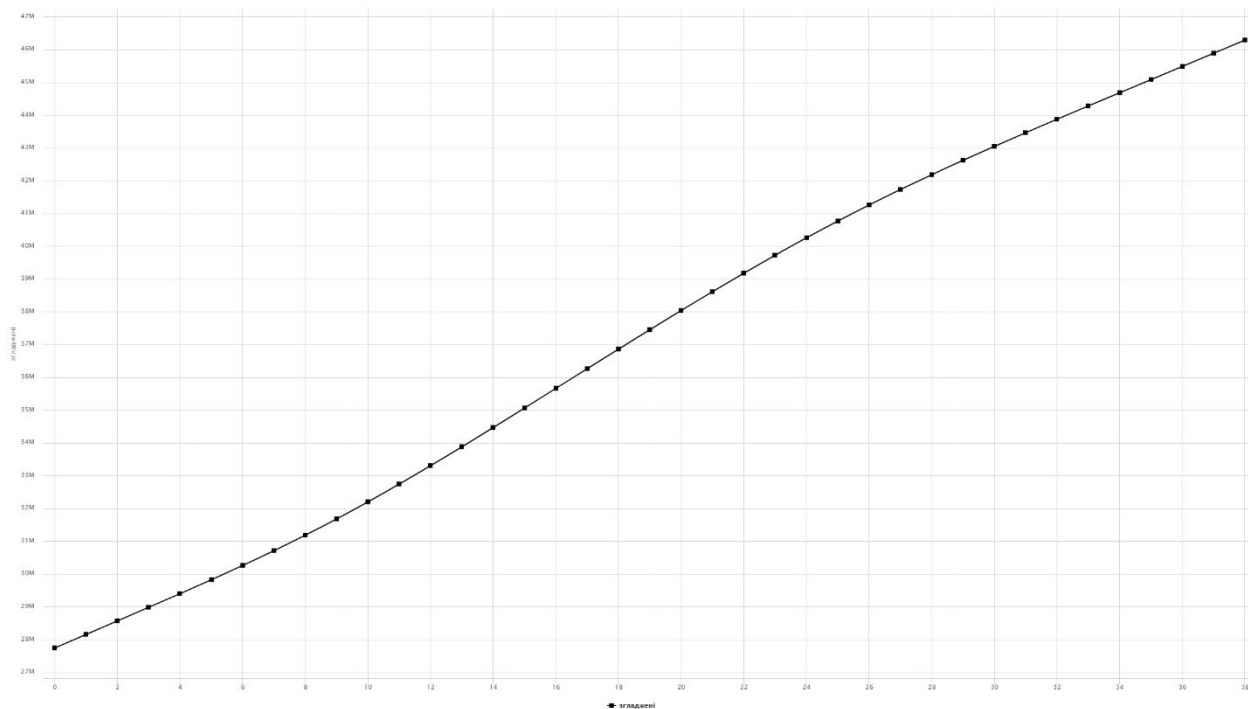


Рисунок 3.12 Графік фільтра Ходріка-Прескотта для щотижневих продаж магазину,  $\lambda = 4764,46$

Методологія ARIMA, розроблена Боксом і Дженкінсом [55], дозволяє це зробити. Цей метод надзвичайно популярний у багатьох додатках, Deductor Studio та Logiном зокрема, і також підтверджувалася його потужність [56].

Водночас моделі ARIMA – непростий метод, адже для його використання потрібен великий практичний досвід, і хоча він часто дає задовільні результати, вони також залежать від кваліфікації користувача [57].

Як було зазначено й детально описано у попередньому розділі, ці моделі складаються з 2 основних процесів – стаціонарної авторегресії та ковзного середнього.

Для моделі ARIMA(p,d,q) необхідно, щоб ряд був стаціонарним. Це означає, що його середнє постійне, а вибірккові дисперсія та автокореляція не змінюються протягом часу. Тому зазвичай необхідно брати різницю ряду доти, доки він не стане стаціонарним (також застосовують логарифмічні перетворення для стабілізації дисперсії).

Число різниць, які були взяті, щоб досягти стаціонарності, визначаються параметром  $d$ . Для того щоб визначити необхідний ступінь (порядок) різниці, потрібно досліджувати графік ряду і автокорелограму:

- сильні зміни рівня зазвичай вимагають взяття несезонної різниці першого порядку;
- сильні зміни нахилу вимагають взяття різниці другого порядку.

Сезонна складова вимагає взяття відповідної сезонної різниці. Якщо є повільне зменшення вибірових коефіцієнтів автокореляції залежно від лага, зазвичай беруть різницю першого порядку. Однак слід пам'ятати, що для деяких часових рядів потрібно брати різниці невеликого порядку або зовсім не брати їх. Також варто зауважити, що надмірна кількість взятих різниць призводить до менш стабільних оцінок коефіцієнтів.

Таким чином модель ARIMA дозволяє з високою якістю прогнозувати як стаціонарні часові ряди, так і нестаціонарні часові ряди з трендом і навіть ряди зі сезонністю (тоді модель називається Seasonal ARIMA, або SARIMA).

Відповідно до корелограми часового ряду щоденних продаж магазину та прикладів стаціонарних рядів у попередньому розділі, можна стверджувати, що цей ряд нестаціонарний.

Застосуємо операцію різниці першого порядку на ряді щоденних продаж, послідовно використовуючи наступні інструменти Loginom – «Редагування викидів» (для усунення викидів даних), «Ковзне вікно» (для отримання значень ряду, зсунутого на 1 тиждень), «Калькулятор» (для створення нового, обчислювального факту у таблиці «Продажі») – і перевіримо, чи достатньо цього для перетворення ряду на стаціонарний. Корелограма всього ряду різниць першого порядку наведено на рис. 3.13.

Якщо скористатися операцією згладжування (фільтр Ходріка-Прескотта з  $\lambda = 1$ ) для ряду щоденних продаж і повторити процес отримання корелограми, ряд яскраво показує цикли та певний позитивно-спадний тренд (рис. 3.14).

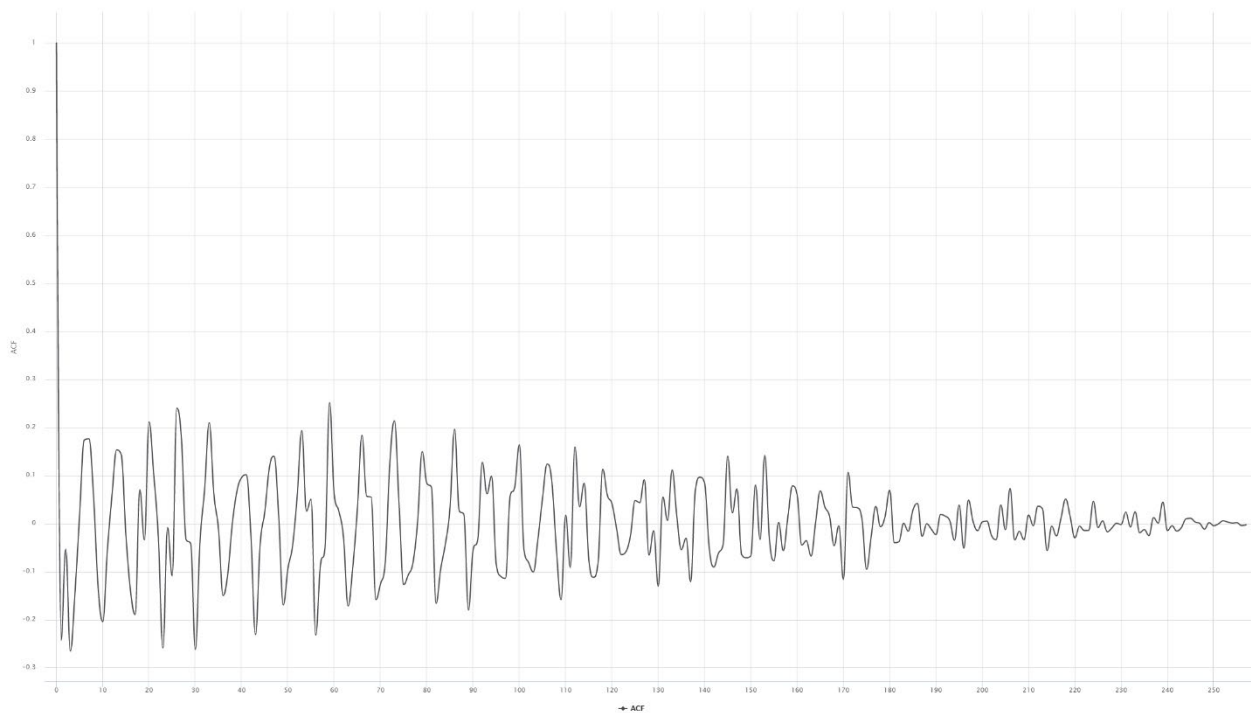


Рисунок 3.13 Корелограма різниці 1-го порядку ряду щоденних продаж

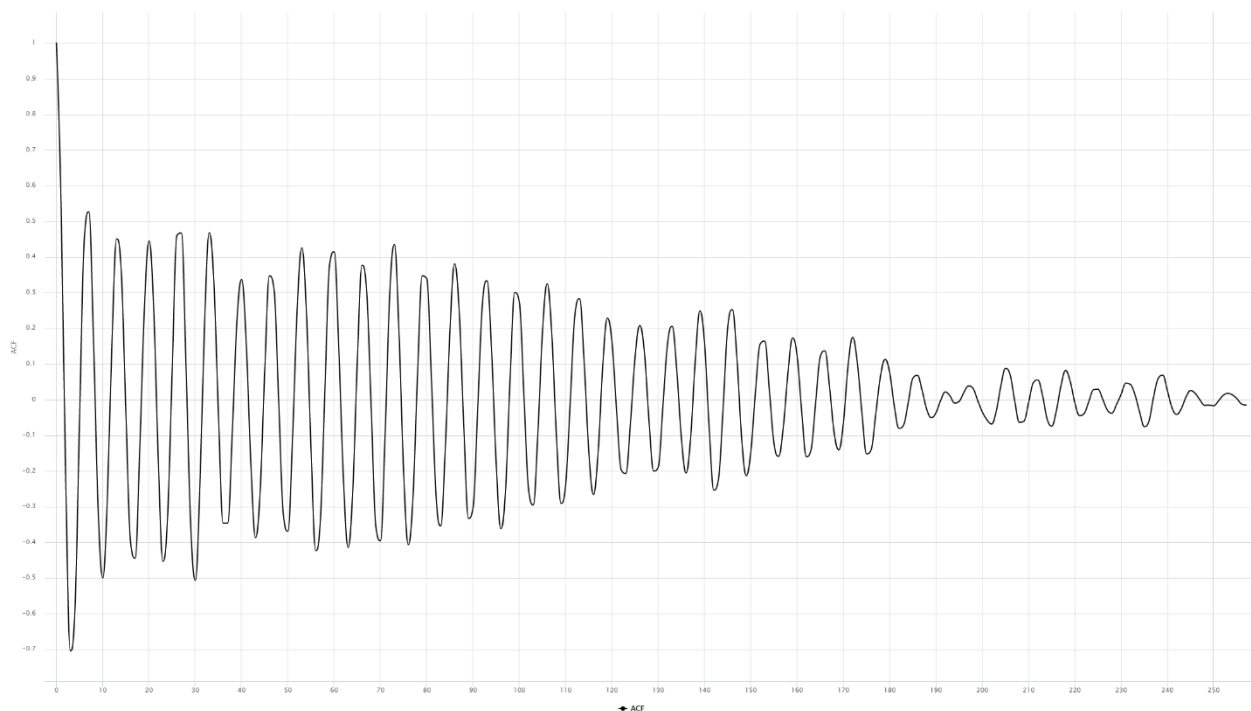


Рисунок 3.14 Корелограма згладженого ряду різниці першого порядку щоденних продаж

Різниця першого порядку усунула нестационарність щоденних продаж. Використаємо цей новий ряд різниць для побудови моделі ARIMA. Для цього скористаємося інструментом Logiном «ARIMAX», що також дозволяє розширити модель ARIMA іншими вхідними параметрами. Підбір параметрів

моделі залишимо Loginom'у, крім сезонної складової – ми виявили, що вона становить 7 днів.

Конфігурація моделі у Loginom здійснюється за інформаційним критерієм Акаїке (Akaike information criterion, AIC). Автоматично розраховані коефіцієнти моделі ARIMA наведені в табл. 3.3, графік фактичних продаж та трьох наступних прогнозованих різниць щоденних продаж зображено на рис. 3.15.

Згідно з табл. 3.3, отриману модель ARIMA можна записати в наступній нотації –  $ARIMA(5,0,5)(2,0,2)$ . Пригадаємо, що ці параметри і нотація припустимі лише для прогнозованої моделі ARIMA *стаціонарного ряду різниць першого* порядку щоденних продаж. Для того щоб отримати прогнозовані значення безпосередньо нестационарного ряду щоденних продаж необхідно включити порядок операції різниці у параметри моделі ARIMA, тобто маємо наступну еквівалентну модель –  $ARIMA(5,1,5)(2,0,2)$ . Графіки ряду фактичних щоденних продаж та значень моделі  $ARIMA(5,1,5)(2,0,2)$  на наступні 3 дні наведено на рис. 3.16.

Таблиця 3.3 Автоматично сконфігурована модель  $ARIMA(5,0,5)(2,0,2)_7$

Параметр	Коефіцієнт
AR[1]	1,036658328
AR[2]	-0,741741972
AR[3]	-0,083749922
AR[4]	0,018536296
AR[5]	0,016357354
MA[1]	0,081657173
MA[2]	0,064779591
MA[3]	0,030440825
MA[4]	-0,005799253
MA[5]	-0,025048554

Параметр	Коефіцієнт
Сезонна AR[7]	0,003027996
Сезонна AR[14]	-0,00361616
Сезонна MA[7]	-0,050087531
Сезонна MA[14]	0,033696559
Константа	9872,51007

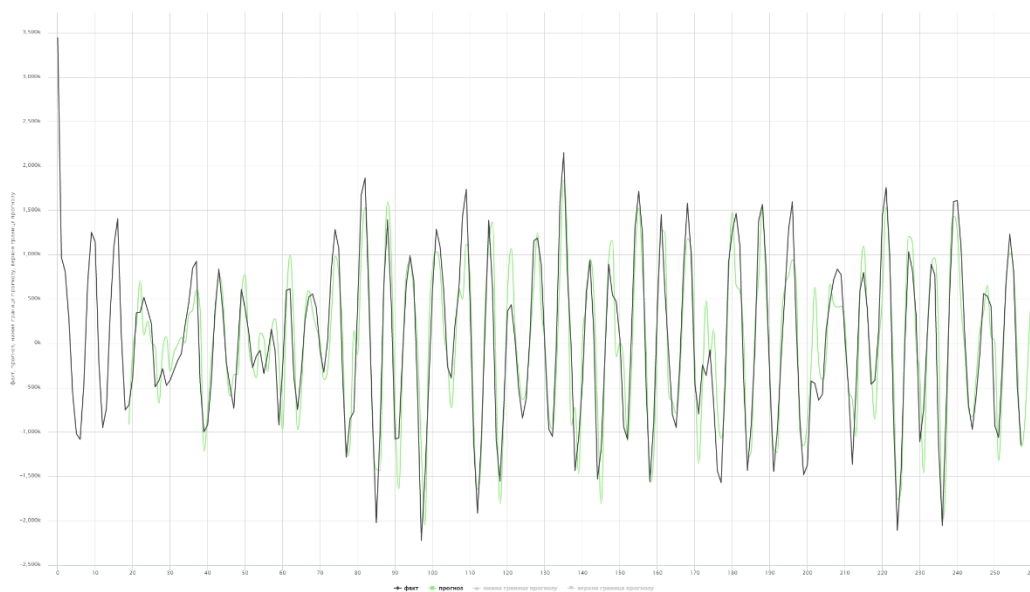


Рисунок 3.15 Графіки фактичних та прогнозованих моделлю  $ARIMA(5,0,5)(2,0,2)_7$  різниць першого порядку щоденних продаж

Таблиця 3.4 Фактичні та прогнозовані моделлю  $ARIMA$  продажі за останній тиждень по дням

Фактичні продажі (грн)	Прогнозовані продажі (грн)
6738120,93	6506607,77
7557790,67	7616492,54
7075928,32	7392554,34
5929824,98	5921945,26
-	5005407,93
-	4937084,52
-	5685446,67

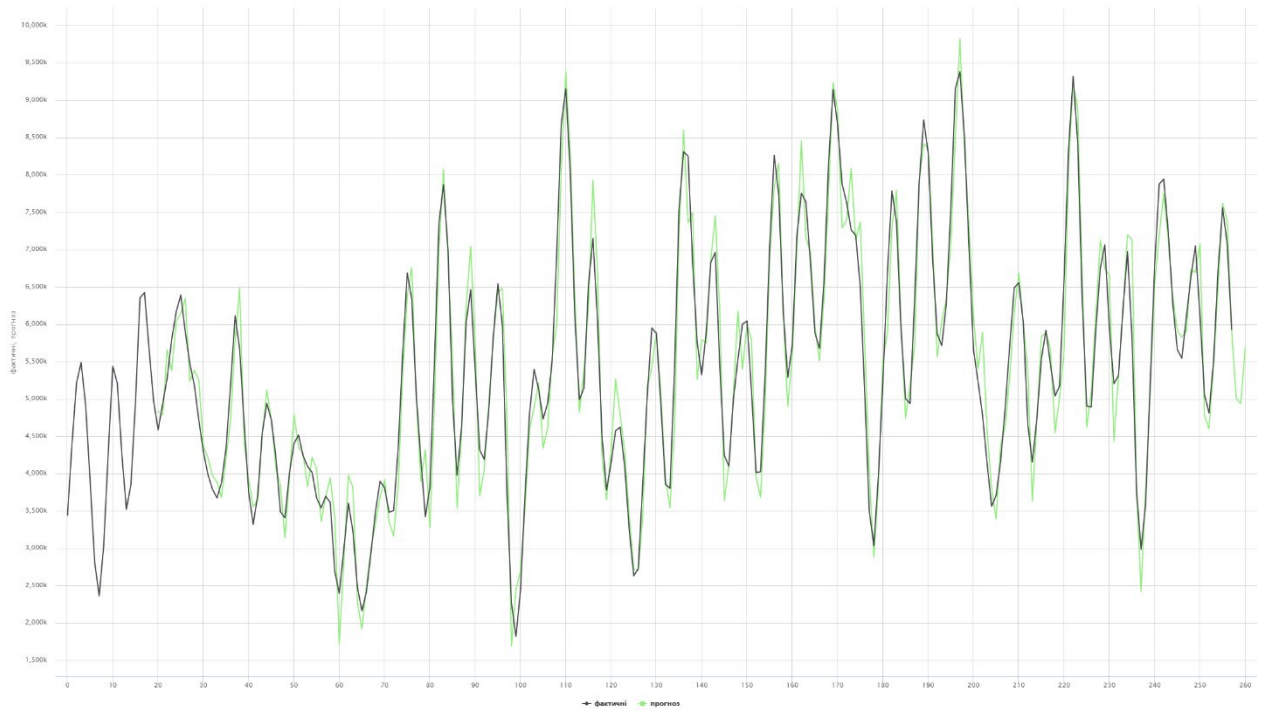


Рисунок 3.16 Графіки ряду фактичних щоденних продаж та значень моделі  $ARIMA(5,1,5)(2,0,2)$  на наступні 3 дні

Якісна модель ARIMA повинна давати не тільки досить точний прогноз, але бути економною і мати незалежні похибки, які в ідеалі містять лише шум без систематичних компонент (зокрема, АКФ залишків не повинна мати жодної періодичності). Якщо залишки систематично розподілені (наприклад, від'ємні в першій частині ряду і приблизно рівні нулю в другій) або включають деяку періодичну компоненту, то це свідчить про неадекватність моделі.

Перевіркою моделі ARIMA є графік похибок (рис. 3.17), вивчення їх трендів та перевірка АКФ залишків (рис. 3.18; на графіку АКФ зазвичай чітко видно періодичність).

З цих графіків видно, що похибки моделі  $ARIMA(5,0,5)(2,0,2)$  не корелюють між собою, а натомість є фактично білим шумом, тому модель можна вважати адекватною.

Таким чином, було проведено аналіз продаж магазину будівельних матеріалів у різних розрізах – по місяцям, тижням та дням. Також було створено адекватну прогнозу модель ARIMA (коефіцієнт детермінації  $r^2$  якої становить 0,8) та спрогнозовано продажі на наступні 3 дні.

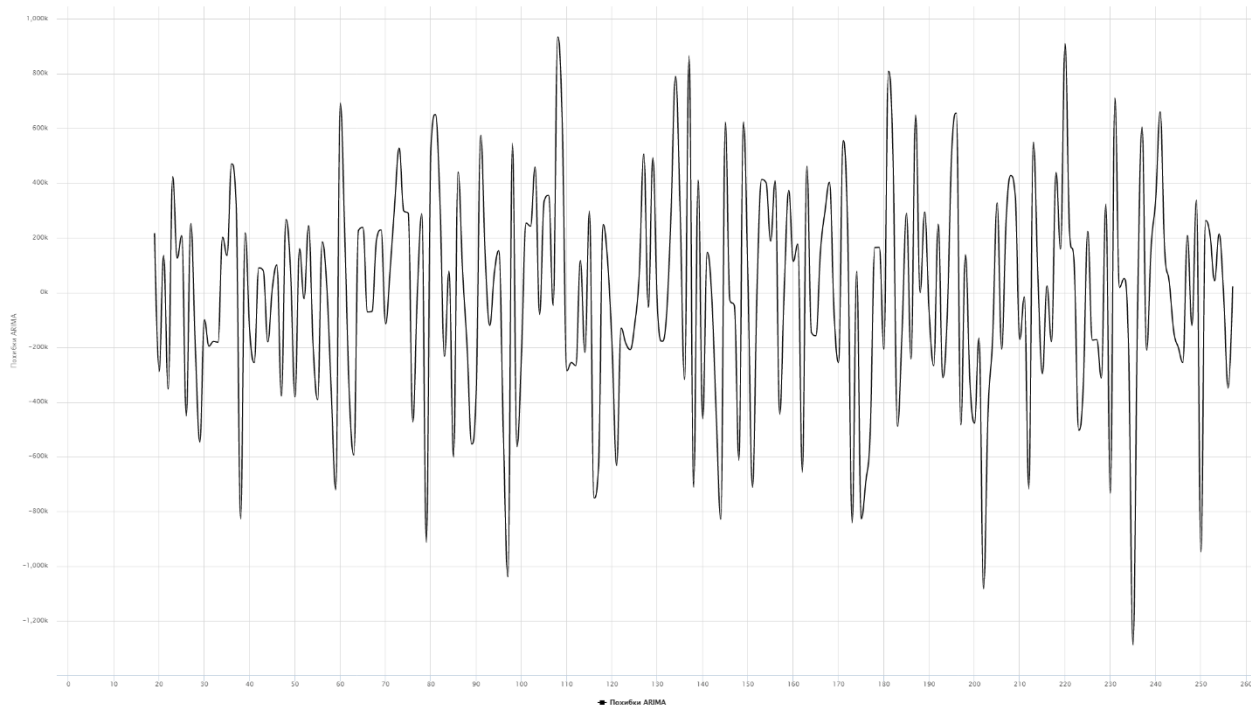


Рисунок 3.17 Графік похибок моделі ARIMA прогнозування різниць щоденних продаж першого порядку

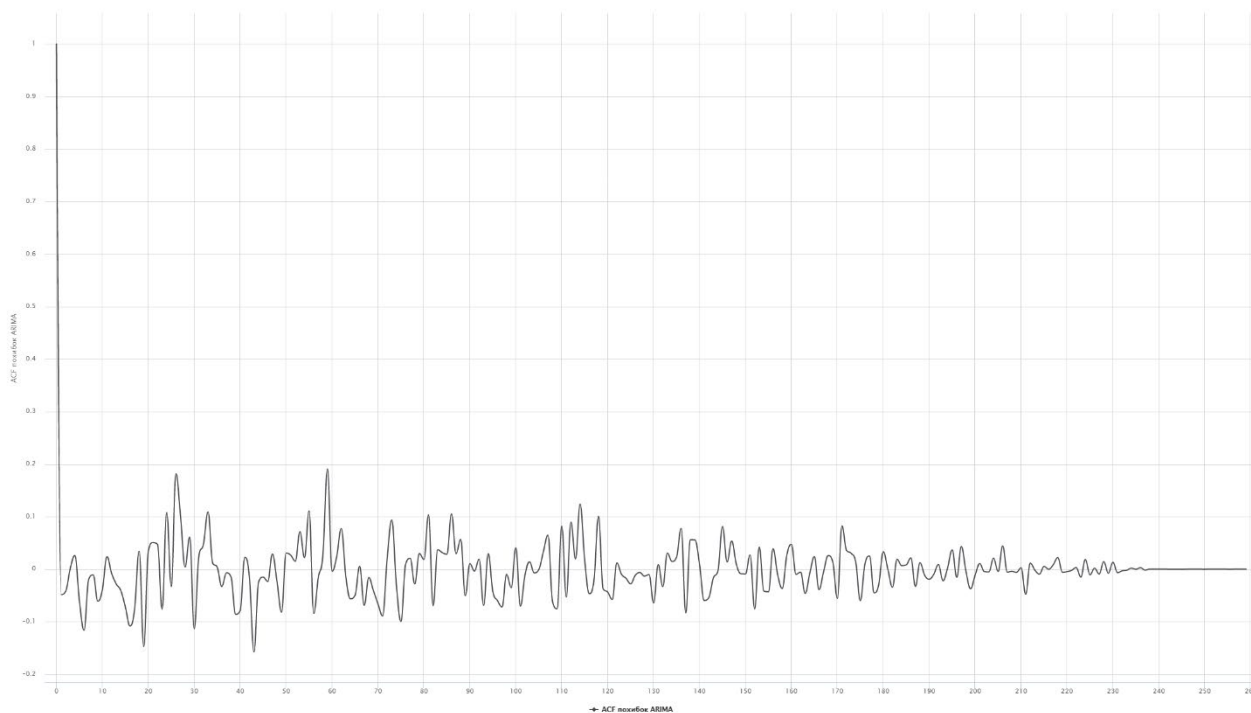


Рисунок 3.18 Корелограма похибок моделі ARIMA прогнозування різниць щоденних продаж першого порядку

### 3.3. Аналіз ринкового кошика організації

Для виявлення спільно придбаних товарів в магазинах проводиться аналіз ринкового кошика. У його основі лежить пошук асоціативних правил, що дозволяє знаходити закономірності між пов'язаними подіями. Алгоритм може виявляти типові шаблони покупок і спільно придбаних товарів.

Отримані результати дозволяють оптимізувати асортимент і його розміщення в торгових залах, поліпшити управління запасами, збільшити обсяги продажів за рахунок пропозиції клієнтам супутніх товарів [25].

Нижче, на рис. 3.19, наведено фрагмент набору транзакцій (чеків) продаж відділу побутової хімії магазину будівельних матеріалів. Увесь набір налічує близько 5000 товарів. Створимо новий сценарій у програмі LogiNot та налаштуємо імпорт текстового файлу, фрагмент даних якого описано нижче.

```

S056560 Перчатки резиновые
S056626 Средство для прочистки труб
S056626 Стиральный порошок ручной
S056624 Чистящий порошок универсальный
S056624 Микроспрей
S056624 Средство для чистки металлических изделий
S056622 Средство по уходу за зеркалами и стеклами
S056608 Средство для мытья посуды
S056608 Мыло кусковое
S056608 Зубная паста
S056601 Средство от накипи
S056601 Мыло кусковое

```

Рисунок 3.19 Фрагмент набору для аналізу ринкового кошику

Для пошуку асоціативних правил часто використовуються 2 найвідоміших алгоритми – Apriori [47] та FPG [48]. Головна перевага останнього перед іншими – обчислювальна ефективність для великої кількості транзакцій. Програмне забезпечення LogiNot реалізує алгоритм FPG, тому пошук асоціативних правил відбуватиметься за цим алгоритмом з наступними параметрами (табл. 3.5):

Таблиця 3.5 Параметри процесу пошуку асоціативних правил

Параметр	Значення
Мінімальна підтримка	5%
Максимальна підтримка	14%

Параметр	Значення
Максимальна кількість наслідків	3
Мінімальна достовірність правила	60%
Мінімальний ліфт правила	1

Не було знайдено жодного асоціативного правила, яке задовольняло б вказані параметри табл. 3.5. Це нормально, оскільки пошук асоціативних правил також базується на попередньому відлагодженні його параметрів аналітиком-експертом, яким у цьому випадку може бути досвідчений керівник цього відділу. Нові параметри алгоритму наведені в табл. 3.6.

Таблиця 3.6 Оновлені параметри пошуку асоціативних правил

Параметр	Значення
Мінімальна підтримка	3%
Максимальна підтримка	автоматично
Максимальна кількість наслідків	3
Мінімальна достовірність правила	50%
Мінімальний ліфт правила	1

У результаті 2-го пошуку було знайдено 6 асоціативних правил, які наведені у табл. 3.7.

Таблиця 3.7 Результати 2-го пошуку асоціативних правил

№	Підтримка	Достовірність	Ліфт	Умова	Наслідок
1	0,039	0,55	12,1	Пральний порошок-автомат	Кондиціонер для білизни
2	0,039	0,85	12,1	Кондиціонер для білизни	Пральний порошок-автомат
3	0,047	0,72	5,1	Сода	Порошок універсальний
4	0,037	0,70	4,9	Засіб від накипу	Порошок універсальний
5	0,064	0,86	4,5	Засіб для миття посуду	Мило
6	0,082	0,84	4,4	Мило рідке	Мило

Було отримано тривіальні правила та набори, про які, швидше за все, магазин вже і так знає, наприклад, пральний порошок та кондиціонер для білизни, які майже в усіх магазинах розміщені на стендах поруч.

При цьому Logipom також надає результати пошуку популярних наборів (12 штук), що відрізняються від асоціативних правил фактичною присутністю в наборі аналізованих даних. Такі набори наведені в табл. 3.8.

Таблиця 3.8 Популярні набори 2-го пошуку

№	Потужність	Достовірність	Товар	Товар
1	2	0,082	Мило	Мило рідке
2	2	0,064	Мікроспрей	Порошок універсальний
3	2	0,064	Мило	Засіб для миття посуду
4	2	0,047	Порошок універсальний	Сода
5	2	0,044	Мікроспрей	Мило
6	2	0,039	Пральний порошок-автомат	Кондиціонер для білизни
7	2	0,038	Порошок універсальний	Зубна паста
8	2	0,037	Порошок універсальний	Засіб від накипу
9	2	0,032	Мікроспрей	Освіжувач повітря
10	2	0,032	Гель антимікробний	Мило рідке
11	2	0,032	Гель антимікробний	Пральний порошок ручний
12	2	0,030	Мікроспрей	Гель антимікробний

Всі ці набори виглядають тривіальними, доволі очевидними та давно втіленими маркетологами та мерчендайзерами магазину, тому проведемо третій, останній пошук асоціативних правил з параметрами, що дозволять знайти приховані набори товарів, які добре продаються разом. Конфігурація та результати 3-го пошуку наведені в табл. 3.9-3.11 відповідно.

Таблиця 3.9 Параметри 3-го пошуку асоціативних правил

Параметр	Значення
Мінімальна підтримка	0,5%
Максимальна підтримка	15%
Максимальна кількість наслідків	3
Мінімальна достовірність правила	25%
Мінімальний ліфт правила	2,5

Таблиця 3.10 Фрагмент асоціативних правил після 3-го пошуку зі значенням ліфту в проміжку [6; 8]

№	Підтримка	Достовірність	Ліфт	Умова	Наслідок
14	0,007	0,35	8,0	Хустинка	Засіб для догляду за склом
15	0,008	0,33	7,5	Зубна паста	Балон освіжувача повітря
16	0,005	0,73	7,5	Гель антимікробний	Мило рідке
17	0,014	0,31	7,3	Балон освіжувача повітря	Піна/сіль для ванни
18	0,014	0,32	7,3	Піна/сіль для ванни	Балон освіжувача повітря
19	0,006	0,43	6,7	Зубна паста	Сода

Таблиця 3.11 Фрагмент результатів аналізу «Що-Якщо» 3-го пошуку

№	Підтримка	Достовірність	Ліфт	Чек	Умова	Наслідок 1	Наслідок 2
48	0,005	0,85	19,69	SO51488	Зубна паста	Піна/сіль для ванни	Балон освіжувача повітря
60	0,006	0,35	8,04	SO52014	Хустинка		Засіб для догляду за склом
28	0,006	0,92	17,91	SO52923	Зубна паста	Засіб для чищення кахелю	Серветки паперові
37	0,008	0,94	13,41	SO57321	Зубна паста	Кондиціонер для білизни	Пральний порошок-автомат
45	0,013	0,31	7,31	SO56808	Балон освіжувача повітря		Піна/сіль для ванни

На перший погляд набір «хустинка – засіб для догляду за склом» має низьку підтримку (35 таких наборів у 5000 чеках) та достовірність і не вартий розгляду. Проте достатньо велике значення його ліфта може свідчити про початок тренду на цей набір товарів: магазину буде достатньо перенести стенд з хустинками ближче до стенду зі засобами для догляду за скляними поверхнями.

Правило «піна для ванни – запасний балон освіжувача повітря» зафіксовано більше, ніж у 100 чеках, що може говорити про ситуації, коли під час прийому ванни з піною також потрібен і приємний запах у всій кімнаті, наприклад, для привчання купання дітей, які не дуже цього любляють.

У цілому, шукані правила з такими параметрами доволі необґрунтовані для негайного втілення магазином, але слугують основою для детальних аналізів, наприклад, на сезонність продаж та сегментації клієнтів (за віком, сімейним статусом тощо).

## ВИСНОВКИ

У результаті проведеного аналізу діяльності комерційної організації на прикладі магазину будівельних матеріалів Будмаркет можна сформулювати наступні висновки.

Розробка бізнес-плану дозволяє планувати не тільки новий проект чи створення організації, а й знаходити невідповідності у поточній діяльності вже існуючого підприємства. Недоліком і водночас перевагою бізнес-планування є величезна кількість вхідної інформації та параметрів, які на виході мають дати достовірну імітаційну модель інвестиційного проекту.

Було підготовано шаблони сценаріїв, якими компанія в майбутньому може користуватися у своїй діяльності. З їх допомогою організація має можливість консолідувати дані та робити зрізи, інтерактивні аналітичні дашборди про динаміку продажів, завантаженості магазинів, про найбільш куповані товари, на основі спроектованого або вже існуючого сховища даних, не переживаючи за його формат. Реалізована також можливість створювати необхідні для бізнесу звіти.

Використання моделей ARIMA, з одного боку, дозволило побудувати точну та надійну прогнозну модель, а з іншого боку, процес побудови таких моделей нечіткий і складний та вимагає експертних знань, тому деякі інформаційні системи бізнес-аналітики відмовляються від їхньої підтримки, надаючи комерційним організаціям інші засоби аналізу.

За допомогою асоціативних правил виявлено часто обрані набори товарів та найбільш вигідні товарні позиції. При виконанні кваліфікаційної роботи були отримані результати, які показують, що при правильному виборі та впровадженні системи інтелектуального аналізу, бізнес отримує суттєві конкурентні переваги та можливості оптимізації бізнес-процесів.

Механізм управління діяльністю комерційної організації повинен відображати взаємозв'язок між її підсистемами та із зовнішнім середовищем. Це дає змогу вибудувати ефективну систему управління діяльністю підприємства з

організаційним та економічним забезпеченням. Створені моделі та сценарії дають змогу менеджерам управляти процесом досягнення прогнозованих результатів діяльності організації, а також його контролювати.

Показники діяльності досліджуваної комерційної організації свідчать про наявність добротного планування, що забезпечує її рентабельність та конкурентоспроможність на ринку відповідних послуг. Тому для такого роду підприємств впровадження будь-якої аналітичної платформи бізнес-аналітики дозволить вийти на новий рівень не лише прибутковості, а й оптимізації бізнес-процесів та прийняття управлінсько-економічних рішень взагалі. Отже, розглянуті аналітичні платформи бізнес-аналітики і Logiном зокрема, містять всі необхідні функції, щоби зробити бізнес-аналіз першим правильним рішенням управлінців.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Лукіяничук С. Застосування засобів data mining з метою оптимізації діяльності підприємства / Лукіяничук С. М., Ковалевська В. В // Вісник Волинського інституту економіки та менеджменту. – 2018. – No 20. – С. 178-185.
2. Крисько Ж. Сутність та види комерційної діяльності підприємства / Крисько Ж // Українська наука: минуле, сучасне, майбутнє. – 2013. – No 8. – С. 75-83.
3. Штефан Б.М. Моделі та інформаційні технології бізнес-аналітики в маркетинговій діяльності / Б.М. Штефан. – Тернопіль, 2015. – 182 с.
4. Помазун О. Моделі та інформаційні технології підтримки прийняття рішень з управління бізнес-процесами підприємства. / Помазун О. // КНЕУ – Київ, 2016. – 203 с.
5. Сисоєнко І. А. Теоретичні узагальнення щодо розуміння сутності поняття «бізнес-аналітика» / І. А. Сисоєнко // Таврійський науковий вісник. – 2013. – Вип. 83. – С. 382-386.
6. Мінц О.Ю. Методологія моделювання інноваційних інтелектуальних систем прийняття рішень в економіці. Маріуполь : Приазовський державний технічний університет, 2017. – 214 с.
7. Пшенянік Є. Збір та аналіз вимог для оптимізації бізнес процесів підприємства. / Пшенянік Є. В., Самченко Н. К. // Інноваційна Україна: креативні ідеї та проекти. – Київ, КНЕУ, 2020. – С. 352-353.
8. Роскладка А. Моделювання процесу консолідації даних агропромислового підприємства / Роскладка А.А. // Вісник ОНУ ім. І.І.Мечникова. – 2015. – No 2/1. – С. 191-194.
9. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). – К.: КНУ ім. Тараса Шевченка, 2018. – 152 с.
10. Сорока П. М. Інформаційний менеджмент: Навч. посіб. для дистанційного навчання / П. М. Сорока, Б. П. Сорока; за наук. ред. д-а екон. наук, професора О. Д. Гудзинського. – К.: Університет «Україна», 2007. – 535 с. – ISBN 978-966-388-204-8.

11. Гороховатський В. Візуалізація та інтелектуальний аналіз бізнес-даних із використанням засобів аналітичної платформи Deductor / Гороховатський В., Вагіна М. // Вісник Університету банківської справи – 2015. – № 2. – С.163-170.
12. Попович П. Я. Економічний аналіз діяльності суб'єктів господарювання. Підручник. – Тернопіль: Економічна думка, 2001. – 365 с.
13. Бізнес-планування підприємницької діяльності : навч. посіб. / З. С. Варналій, Т. Г. Васильців, Р. Л. Лупак, Р. Р. Білик. Чернівці: Технодрук, 2019. – 264 с. ISBN 978-617-7611-59-1
14. Инмон Б. Производительность систем хранилищ данных. // Data Warehouse Environment – 2000. – № 4. – С. 41-48.
15. Марченко И. С. Системный подход к управлению коммерческой деятельностью организаций / И. С. Марченко // Вестник МГТУ. – 2010. – № 1, т. 13. – С. 27–30.
16. Подпорина А. Разработка бизнес-плана по открытию магазина автохимии и автокосметики // НИУ «БелГУ». – Белгород, 2018. – 54 с.
17. Лебедев В.В. Информационные технологии бизнес-аналитики: Учебно-методическое пособие. – Пермь: НИУ ВШЭ ПФ, 2011.
18. Паклин Н. Бизнес-аналитика: от данных к знаниям. 2-е изд., испр / Н.Б. Паклин, В.И. Орешков // СПб.: Питер. – 2013. – 704 с. – ISBN 9785459007176
19. Бабин Д. Генетический алгоритм решения задачи анализа рыночной корзины / Д.В. Бабин // Журнал «Штучний інтелект», Інститут проблем штучного інтелекту – Донецьк, 2006 – № 4. - С.129-132.
20. В.В. Сычев. Поиск ассоциативных правил с помощью аналитической платформы deductor на примере комплекса услуг автотранспортного предприятия / В.В. Сычев, И.Э. Игумнова // Вестник Современных Исследований. – Омск, 2019. – С. 219-222.
21. Френкс Б. Укрощение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики – М.: «Манн, Иванов и Фербер». – 2014. – 352 с.



34. Shmueli G. Data Mining for Business Intelligence. 2<sup>nd</sup> ed / Shmueli G., Patel N., Bruce P. // John Wiley & Sons. – 2010. – P. 428. - ISBN 9780470526828
35. Berry M. Data Mining Techniques for Marketing, Sales and Customer Relationship management. / Berry M., Linoff G. // John Wiley & Sons. – 2011. – P. 888. – ISBN 9780470650936
36. Larose D. Discovering Knowledge in Data // John Wiley & Sons. – 2014. – P. 316. – ISBN 9780470908747
37. Perner P. Advances in Data Mining: Applications and Theoretical Aspects // Springer – 2010. – 662 p. – ISBN 978-3-319-08976-8
38. Kimball R. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses. // John Wiley – 2002. – 436 p.
39. Collier P. Accounting for managers. Third edition. / Paul M. Collier // Wiley. –, 2009. – 532 p. – ISBN 978-0-470-7764-0
40. Podeswa H. The Business Analyst's Handbook. // Course Technology. – 2008. – P. 411. – ISBN 1598635654
41. Helfert E. Techniques of Financial Analysis // McGraw-Hill – 1996. – 600 p. – ISBN 0786311207
42. Morton J. Big Data: Opportunities and Challenges / Morton J., Runciman B., Gordon K // BCS Learning & Development Limited. – Swindon, 2014.
43. Liebowitz J. Big Data and Business Analytics // Auerbach Publications. – 2013. – P. 304. – ISBN 9781466565784
44. Voulgaris Z. AI for Data Science / Voulgaris Z., Bulut. Y // Technics Publications. – 2018. – P. 290. - ISBN 9781634624121
45. Hyndman R. Forecasting: principles and practice, 2nd edition / Hyndman R.J., Athanasopoulos, G. // OTexts - Australia, Melbourne, 2018. – 384 p.
46. Mueller J. Artificial Intelligence / Mueller J., Massaron L. // John Wiley & Sons. – 2018. – P. 336. - ISBN 978-1-119-46765-6
47. Srikant R. Mining quantitative association rules in large relational tables / Srikant R., Agrawal R. // IBM Almaden Research Center. – 1996. – P. 12.

48. Brin S. Dynamic Itemset Counting and Implication Rules for Market Basket Data / Brin S., Motwani R., Ullman J., Tsur S. // ACM Press – New York, 1997. – P. 255-264.
49. Han J. Mining frequent patterns without candidate generation. / Han J., Jian P., Yiwen Y // School of Computing Science. Simon Fraser University – 2000. – vol. 29, no. 2, pp. 1-12.
50. Rouaud M. Probability, Statistics and Estimation (Short Edition). [Электронный ресурс] – 2013. // Режим доступа: <http://www.incertitudes.fr/book.pdf>
51. Neter J. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs / John Neter, William Wassermann, Michael H. Kutner. – Irwin, Illinois, 1990.
52. Hodrick R. Postwar US business cycles: an empirical investigation / Hodrick R. J., Prescott E. C // Journal of Money, credit, and Banking. – 1997. – P. 1-16.
53. Ravn M. On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations / Ravn, Morten & Uhlig, Harald // The Review of Economics and Statistics. – 2002.
54. Pedersen, Torben Mark. The Hodrick-Prescott filter, the Slutsky effect, and the distortionary effect of filters // Journal of Economic Dynamics and Control, 25, issue 8. – P. 1081-1101.
55. Box G. Time series analysis: forecasting and control. / Box G. E. P., Jenkins G. M. // Holden-Day. – San Francisco, 1976. – 575 p.
56. Pankratz A. Forecasting with Univariate Box-Jenkins Models. // Wiley & Sons, Inc. - New York, 1983.
57. Business fluctuations; forecasting techniques and applications. / Bails, Dale G., Peppers, Larry C. // Prentice Hall Inc. – New Jersey, 1982. – p.482.
58. A guide to the Business Analysis Body of Knowledge (BABOK guide), version 3.0 / International Institute of Business Analysis // International Institute of Business Analysis – Toronto, 2015. – ISBN 9781927584026
59. What is Tableau? [Электронный ресурс] – 2018. // Режим доступа: <https://www.guru99.com/what-is-tableau.html>