

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ**

Факультет інформаційних технологій

ПОГОДЖЕНО

Декан факультету (Директор ННІ)

Інформаційних технологій
(назва факультету (ННІ))

(підпис)

Ігор БОЛБОТ

(ім'я ПРІЗВИЩЕ)

“ ___ ” _____ 20__ р.

ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ

Завідувач кафедри

комп'ютерних наук
(назва кафедри)

(підпис)

Белла ГОЛУБ

(ім'я ПРІЗВИЩЕ)

“ ___ ” _____ 20__ р.

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему Програмне забезпечення системи аналізу реалізації високоякісного насіннєвого матеріалу

Спеціальність 121 Інженерія програмного забезпечення
(код і найменування)

Освітня програма “Програмне забезпечення інформаційних систем”
(назва)

Орієнтація освітньої програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Гарант освітньої програми

к.ф.-м.н., доцент
(науковий ступінь та вчене звання)

(підпис)

Віктор КИРИЧЕНКО
(ім'я ПРІЗВИЩЕ)

Керівник магістерської кваліфікаційної роботи

к. ф. — м. н., доцент
(науковий ступінь та вчене звання)

(підпис)

Яніна КРИВОРУЧКО
(ім'я ПРІЗВИЩЕ)

Виконав

(підпис)

Микита ЗАВАЦЬКИЙ
(ім'я ПРІЗВИЩЕ здобувача)

КИЇВ – 2025

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Факультет інформаційних технологій

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук

доцент, к.т.н. _____ Голуб Б.Л.
(науковий ступінь, вчене звання) (підпис)
“ 10 ” листопада 2024 року

ЗАВДАННЯ

ДО ВИКОНАННЯ МАГІСТЕРСЬКОЇ КВАЛІФІКАЦІЙНОЇ РОБОТИ СТУДЕНТУ

_____ Завацький Микита Сергійович

(прізвище, ім'я, по батькові)

Спеціальність 121 “Інженерія програмного забезпечення”

(код і найменування)

Освітня програма Програмне забезпечення інформаційних систем

(назва)

Орієнтація освітньої програми _____ освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Тема магістерської кваліфікаційної роботи Програмне забезпечення системи аналізу реалізації високоякісного насінневого матеріалу

затверджена наказом проректора НУБіП України від “ 1 ” листопада 2024 р. № 1963 “С”

Термін подання завершеної роботи на кафедру _____ 14 листопада 2025 р.

(рік, місяць, число)

Вихідні дані до магістерської кваліфікаційної роботи набори даних Держстату, Єдиного державного реєстру суб'єктів насінництва та Мінагрополітики України

Перелік питань, що підлягають дослідженню:

1. Аналіз ключових проблем системи реалізації високоякісного насінневого матеріалу на сільськогосподарських підприємствах.

2. Дослідження можливостей застосування OLAP та Data Mining методів для підвищення ефективності аналізу реалізації насінневого матеріалу.

3. Визначення факторів, що впливають на якість та обсяги реалізації насінневого матеріалу.

4. Виявлення закономірностей та трендів у процесах постачання і збуту високоякісного насінневого матеріалу.

Перелік графічного матеріалу (за потреби)

Дата видачі завдання “ 7 ” листопада 2024 р.

Керівник магістерської кваліфікаційної роботи _____

(підпис)

Криворучко Я.С.
(прізвище та ініціали)

Завдання прийняв до виконання _____ Завацький М.С.

ЗМІСТ

ВСТУП.....	6
РОЗДІЛ 1. АНАЛІЗ ПОСТАНОВКИ ЗАВДАННЯ ТА ПРЕДМЕТНОЇ ОБЛАСТІ.....	9
1.1 Загальна характеристика предметної області.....	9
1.2 Огляд існуючих рішень.....	11
1.3 Постановка завдання.....	15
РОЗДІЛ 2. МОДЕЛЮВАННЯ СИСТЕМИ.....	17
2.1 Діаграма прецедентів.....	17
2.2 Архітектура системи.....	19
2.3 Діаграма послідовності.....	20
2.4 Діаграма класів.....	21
РОЗДІЛ 3. РОЗРОБКА СИСТЕМИ.....	22
3.1 Проектування реляційної бази даних.....	22
3.2 Проектування сховища даних.....	25
3.3 Розгортання OLAP-кубу.....	29
3.4 Отримання даних за допомогою Data Flow.....	33
РОЗДІЛ 4. РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ.....	38
4.1 Вимоги до апаратного та програмного забезпечення.....	38
4.2 Побудова звітності.....	39
4.3 Розрахунок КРІ.....	43
4.4 Інтелектуальний аналіз даних за допомогою Data Mining...	44
4.5 Побудова програмного забезпечення.....	52
4.5.1 Вступ та вибір технологічного стеку.....	52
4.5.2 Архітектура програмного забезпечення.....	53

4.5.3 Опис функціональних модулів системи.....	53
ВИСНОВКИ.....	58
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	60

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

БД – база даних.

СД – сховище даних.

СППР – система підтримки прийняття рішень.

СУБД – система управління базами даних.

SQL – SQL Server Analysis Services (SSAS).

SSAS – SQL Server Integration Services.

SSIS – SQL Server Integration Services.

SSRS – SQL Server Reporting Services.

OLAP – On-Line Analytical Processing

DSV - Data Source View.

DF – Data Flow.

DSS - Decision Support System

ВСТУП

Актуальність теми. В аграрному секторі України ринок насіння є ключовим, але водночас одним із найменш прозорих. На даний момент, основний відсоток продажів іде через посередників, що створює "інформаційний вакуум" для кінцевого споживача, який не завжди може зробити економічно вигідний вибір. Проблематика полягає у відсутності єдиних інструментів для аналізу ефективності постачання та якості насіння.

Сьогодні конкурентоспроможність агробізнесу напряду залежить від здатності працювати з даними. Впровадження аналітичних платформ - це не просто технологічний тренд, а інструмент для глибокого розуміння ринку. Такі системи допомагають визначити найефективніші канали збуту, оптимізувати логістику та, найголовніше, забезпечити аграріїв якісним насінням за найкращою ціною.

Об'єкт дослідження: процес реалізації високоякісного насінневого матеріалу.

Предмет дослідження: система аналізу реалізації високоякісного насінневого матеріалу.

Мета дослідження: підвищити рівень реалізації високоякісного насінневого матеріалу.

Завдання:

- Провести системний аналіз предметної області ринку реалізації насінневого матеріалу.
- Спроекувати архітектуру програмного комплексу, що включає оперативну базу даних (БД) та багатовимірне аналітичне сховище даних (СД).
- Розробити структуру інформаційного забезпечення та ETL-процес (на базі SSIS) для перенесення даних з БД у СД.
- Побудувати багатовимірний OLAP-куб (на базі SSAS) для забезпечення швидкої агрегації та аналізу даних.
- Застосувати методи інтелектуального аналізу даних (Data Mining)

- Розробити програмне забезпечення для візуалізації результатів та підтримки прийняття управлінських рішень.
- Представити та проаналізувати отримані результати, сформулювавши висновки щодо ефективності запропонованого рішення.

Методи дослідження. Теоретичні: системний аналіз; аналіз науково-технічної літератури методи теорії ймовірностей та математичної статистики.

Практичні: методи проектування баз даних та сховищ даних; технології OLAP для побудови багатовимірних кубів; методи інтелектуального аналізу даних; методи програмної інженерії для розробки прототипу програмного забезпечення.

Наукова новизна. Удосконалено підхід до аналізу ринку насінництва шляхом інтеграції класичних OLAP-технологій (для швидкої агрегації даних) з сучасними методами Data Mining (для глибокого аналізу та прогнозування), що реалізовані у єдиному програмному комплексі.

Запропоновано комбіноване використання алгоритмів Prophet (для прогнозування сезонного попиту), K-Means (для сегментації регіональних ринків за профілем попиту) та Apriori (для виявлення патернів закупівель) саме для вирішення задачі аналізу реалізації насіння в умовах українського ринку.

Отримало подальший розвиток проектування систем підтримки прийняття рішень для аграрного сектору, що враховує специфічні фактори ринку, такі як висока сезонність та непрозора структура збуту.

Апробація результатів дослідження. В процесі розробки кваліфікаційної роботи на тему “Програмне забезпечення системи аналізу реалізації високоякісного насіннєвого матеріалу” відбувся виступ XVI Міжнародній науково-практичній конференції студентів, аспірантів та молодих вчених «ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ: ЕКОНОМІКА, ТЕХНІКА, ОСВІТА» 28-29 жовтня 2025 року, НУБіП України, Київ, 2025.

Публікації. Опубліковані тези “Програмне забезпечення системи

аналізу реалізації високоякісного насінневого матеріалу”, що ввійшли до Збірника матеріалів Міжнародної науково-практичної конференції молодих вчених “ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ: ЕКОНОМІКА, ТЕХНІКА, ОСВІТА 2025”, (28-29 жовтня 2025 року, НУБіП України, Київ, 2025).

Структура роботи. Магістерська кваліфікаційна робота складається із 61 сторінки. У роботі використано 34 джерела. Кількість розділів - 4. У першому розділі виконується аналіз постановки завдання та предметної області. У другому розділі описується моделювання системи. Третій розділ описує розробку системи. У четвертому розділі представлені результати дослідження.

РОЗДІЛ 1. АНАЛІЗ ПОСТАНОВКИ ЗАВДАННЯ ТА ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Загальна характеристика предметної області

Предметною областю даного дослідження є процес реалізації та аналізу ринку високоякісного насіннєвого матеріалу в Україні. Цей ринок є однією з ключових та найбільш технологічно ємних складових усього агропромислового комплексу країни.

Ринок насіння в Україні характеризується складною структурою, яка поєднує потужне внутрішнє виробництво та значну залежність від імпорту високопродуктивних гібридів.

1. Внутрішнє виробництво: В Україні офіційно зареєстровано понад 400 суб'єктів насінництва та розсадництва. У 2024 році ними було задекларовано до реалізації близько 500 тис. тонн насіння, де ключові позиції займають пшениця, кукурудза та ячмінь.

2. Імпорт: Незважаючи на власне виробництво, Україна залишається великим імпортером, особливо у сегментах насіння соняшнику, кукурудзи та ріпаку. Загальний обсяг імпорту у 2024 році склав близько 388 мільйонів доларів. Ключовими постачальниками є країни ЄС (Франція, Німеччина, Румунія, Угорщина, Австрія) та США.

3. Ключові гравці: На ринку домінують великі міжнародні компанії-виробники, такі як Bayer (Monsanto), Corteva (Pioneer), Syngenta, KWS, Lidea та Maisadour. Багато з них мають в Україні власні виробничі потужності та насіннєві заводи.

Ключова проблематика ринку полягає не стільки у виробництві, скільки в складній та непрозорій системі дистрибуції та високій частці фальсифікату.

Канали збуту чітко діляться на B2B та B2C сегменти:

- B2B (Business-to-Business): Основний обсяг продажів проходить через мережу офіційних дистриб'юторів та посередників, які реалізують

продукцію агрохолдингам та фермерським господарствам.

- B2C (Business-to-Consumer): Цей сегмент, хоча і менший, активно розвивається через інтернет-магазини та онлайн-майданчики (Prom.ua, Rozetka), але він орієнтований переважно на дрібних фермерів та приватних осіб.

Ця багатоканальна структура створює "інформаційний вакуум", де аграрію складно відстежити ланцюг постачання та бути впевненим у якості продукції.

Ключові проблеми, що стоять перед аграріями:

1. Ризик фальсифікату: Це головна загроза. За експертними оцінками, до 20% ринку насіння в Україні може мати сумнівне походження. Фальсифікат часто мімікрує під брендову продукцію, маючи майже ідентичну ціну, але не дає прогнозованої врожайності. Це прямі фінансові втрати.
2. Економічний тиск: Аграрії змушені шукати шляхи здешевлення виробництва. Однак, якщо вони економлять на засобах захисту рослин (купуючи генерики), то на якісному насінні вони економити не намагаються, розуміючи, що це основний фактор ризику. Це створює високий попит на перевірену, але дорогу продукцію.
3. Кліматичні виклики: Посилення примх погоди змушує фермерів відмовлятися від "простих" сортів на користь дорожчих, але більш стійких та адаптованих гібридів.

Саме ці проблеми - високий ризик фальсифікату, економічний тиск та необхідність вибору складних гібридів - формують гостру потребу в програмному забезпеченні для аналізу.

Існуючі ERP-системи чи облікові програми на підприємствах здебільшого фіксують лише факт продажу. Вони не дають інструментів для глибокого аналізу.

Розроблювана система покликана вирішити ці завдання шляхом:

1. Аналізу постачальників: Дозволяючи відстежувати надійність партнерів та перевіряти їх наявність у Державному реєстрі суб'єктів насінництва.
2. Прогнозування попиту: Допомагаючи дистриб'юторам та виробникам планувати запаси сезонних товарів.
3. Сегментації ринку: Дозволяючи аналізувати, які регіони купують які типи насіння, та оптимізувати логістику і маркетинг.
4. Виявлення закономірностей: Знаходячи приховані зв'язки між категоріями, регіонами та обсягами замовлень для виявлення ринкових ніш.

1.2 Огляд існуючих рішень

Аналіз існуючих рішень було проведено за декількома напрямками: дослідження існуючих інформаційних систем з реалізації насіння та аналіз комерційних аналітичних систем.

1 напрям: Для аналізу було обрано одного з провідних гравців ринку e-commerce насіння в Україні — інтернет-магазин «Semena.in.ua»(рис. 1.1). Даний сайт є класичною інформаційною системою, орієнтованою на клієнта (B2C), та виконує функції операційної обробки транзакцій (OLTP).

Його головна сторінка та структура каталогу демонструють типову архітектуру для онлайн-ритейлу:

- **Каталог продукції(представлений на рис. 1.2):** Має глибоку ієрархічну структуру, що є основою для майбутнього аналізу. Категорії включають "Насіння овочів", "Насіння квітів", "Засоби захисту рослин", "Добрива", "Агроволокно" тощо.
- **Модулі взаємодії:** Сайт містить стандартні сторінки, такі як "Оплата і доставка", "Акції", "Блог" (статті та поради), "Про компанію" та "Контакти".

- **Функціонал:** Система ефективно вирішує завдання з пошуку товару, фільтрації (за виробником, ціною), оформлення замовлення та збору відгуків.

Такі системи є джерелом первинних даних (чеків, замовлень, клієнтської поведінки), але вони не надають вбудованих інструментів для глибокого аналізу. Вони не можуть автоматично сегментувати ринок за областями або прогнозувати попит. Це підкреслює необхідність створення окремої аналітичної системи (OLAP), яка б використовувала ці оперативні дані для стратегічного аналізу.

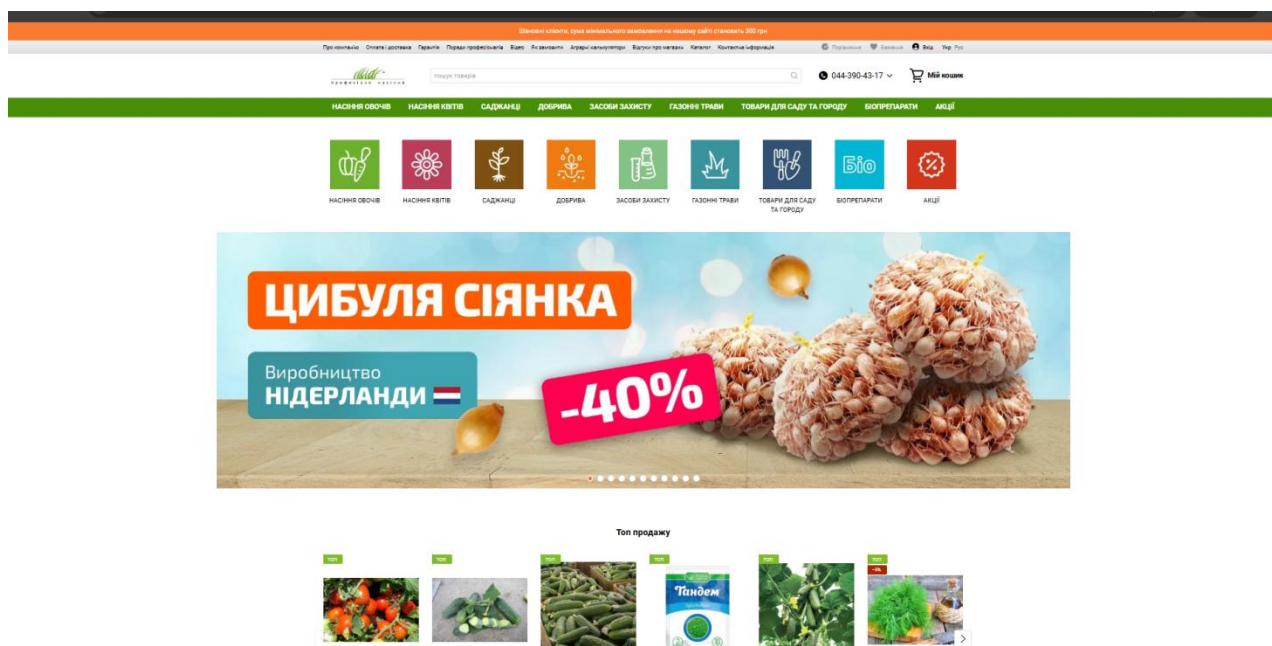


Рис. 1.1 Головна сторінка сайту “Semena.in.ua”

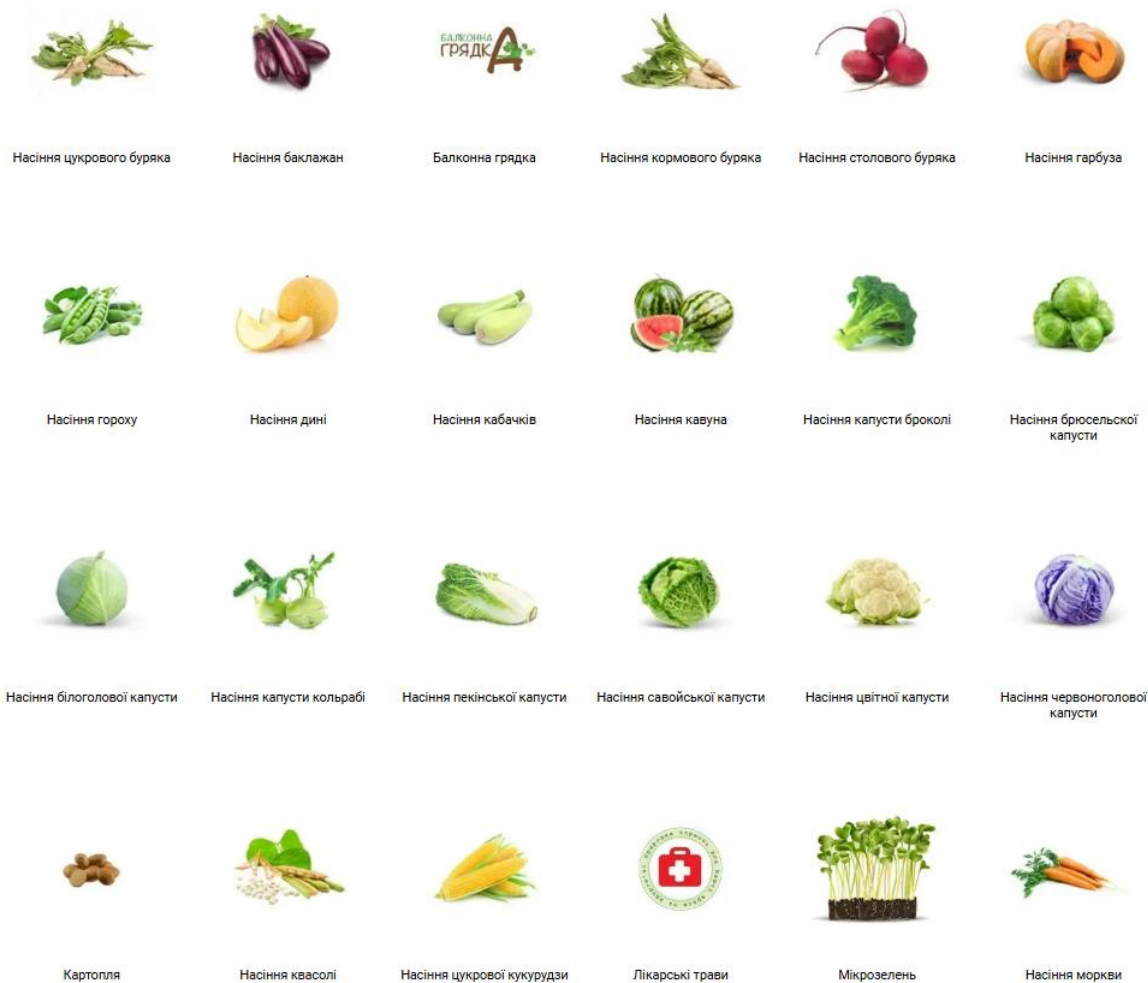


Рис. 1.2 Каталог сайту

2 напрям: Аналіз існуючих систем бізнес-аналітики.

Розглянемо існуючу систему Odoo - це популярна ERP-система з відкритим кодом, яка здобула визнання завдяки своїй модульності. Вона пропонує широкий набір інтегрованих бізнес-додатків, що охоплюють CRM (управління відносинами з клієнтами), управління складом, бухгалтерію, виробництво та управління продажами. Можливості цієї платформи представлені на рис. 1.3.

Аналітичні можливості (Переваги): Сильною стороною Odoo є глибока інтеграція аналітичних інструментів з операційними процесами. Аналітичні модулі Odoo є потужними для операційної звітності (OLTP). Вони дозволяють:

- Управління продажами: Відстежувати виконання планів, аналізувати воронки продажів та ефективність менеджерів.

- Управління бюджетами: Вести план-фактний аналіз витрат та доходів у реальному часі.
- Обробка даних: Генерувати стандартні звіти (наприклад, "залишки на складі", "рух товарів", "обсяг продажів за місяць").
- Кількісний та якісний аналіз: Вбудовані дашборди надають керівництву загальну картину стану бізнесу.

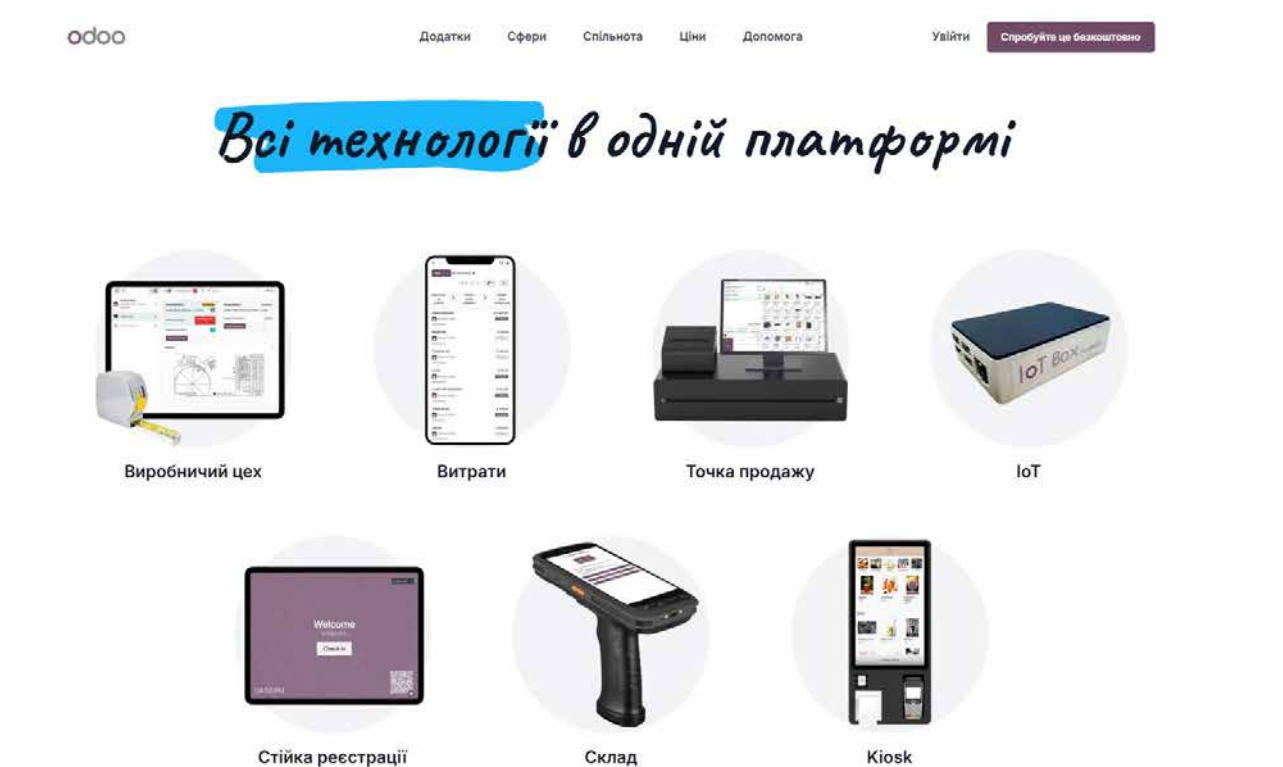


Рис. 1.3 Платформа Odoo

Недоліки: незважаючи на потужний операційний функціонал, аналітичні можливості Odoo обмежені описовою аналітикою. Вони чудово відповідають на питання "Що сталося?", але не мають вбудованих інструментів для прогностичного чи приписового аналізу.

Іншим рішенням є Microsoft Dynamics 365 (рис. 1.4) - комплексна флагманська платформа від Microsoft, що тісно поєднує функціонал ERP та CRM. Її ключова перевага - нативна інтеграція з екосистемою Microsoft,

зокрема з Power BI.

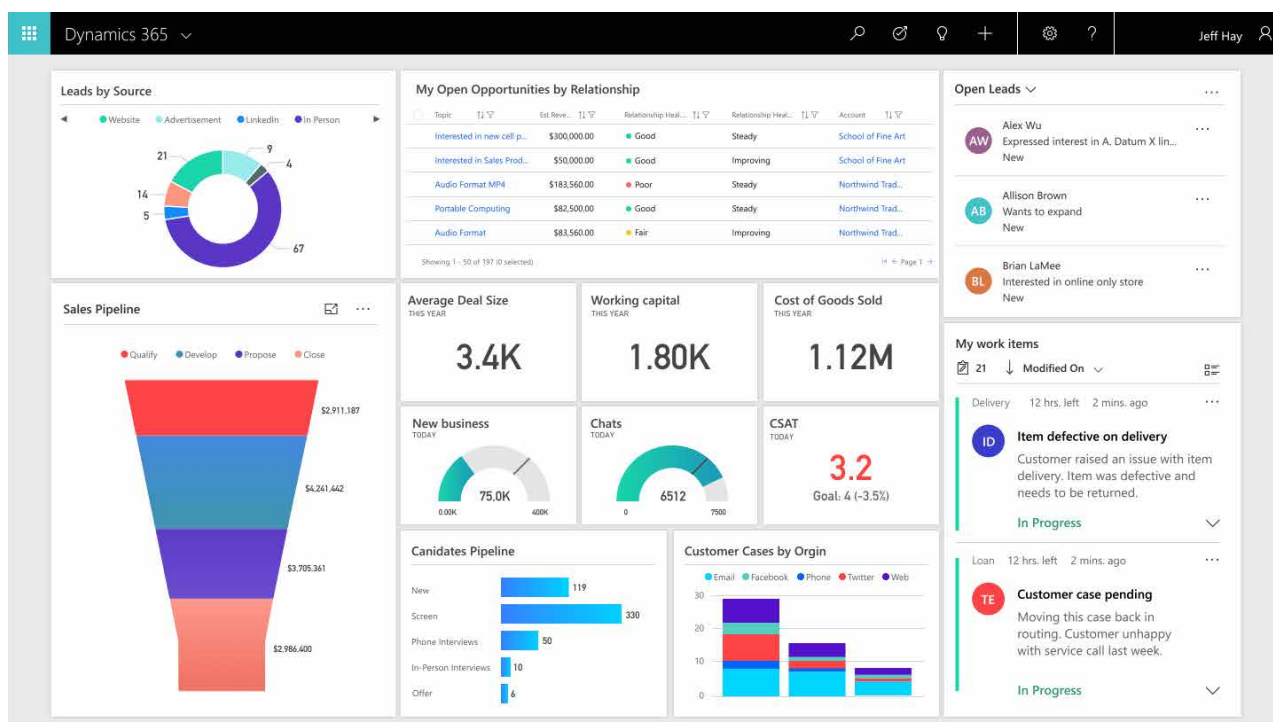


Рис. 1.4 Інтерфейс Microsoft Dynamics 365

Аналітичні можливості (Переваги): Аналітичний розділ Dynamics 365, посилений Power BI, надає значно ширші можливості, ніж стандартні ERP:

- Візуалізація даних: Побудова складних, інтерактивних дашбордів.
- Розрахунок KPI: Інструменти для відстеження ключових показників ефективності в реальному часі.
- Інтерактивний аналіз: Можливість "провалюватися" (drill-down) у даних, від загальних показників до окремих транзакцій, що схоже на функціонал OLAP-кубів.

Недоліки: проте, як і всі системи загального призначення, Dynamics 365 та Power BI є платформами, а не готовими рішеннями. Вони надають потужний інструментарій, але не спеціалізовані моделі для ринку насіння.

1.3 Постановка завдання

Проведений аналіз предметної області та ознайомлення з існуючими рішеннями дозволили сформулювати ключові завдання для розробки системи

аналізу.

1. Формування моделей предметної області.
2. Побудова архітектури системи аналізу. Основними вузлами системи є:
 - Оперативні джерела.
 - Багатовимірне сховище даних (де агрегована інформація зберігається для аналізу).
 - Аналіз даних (OLAP): Здійснюється за допомогою технологій OLAP, що дозволяє візуалізувати дані та розраховувати KPI.
 - Інтелектуальний аналіз даних (Data Mining): Здійснюється на основі технологій Data Mining для виявлення прихованих закономірностей.
3. Дослідження роботи системи та випробування. Це завдання полягає у практичній реалізації аналітичних моделей.
4. Генерування висновків та рекомендацій. На основі отриманих результатів (графіків прогнозу та звітів) формулюються висновки щодо ефективності системи та надаються рекомендації для бізнесу щодо реалізації насінневого матеріалу.

Для аналізу даних є перелік питань, на які має відповісти система:

- У яких виробників насіння більше всього замовляють?
- З яких областей було здійснено найбільше покупок насіння?
- Яка категорія насіння має найбільшу кількість продажів?
- Коли найбільше замовляють насіння?

РОЗДІЛ 2. МОДЕЛЮВАННЯ СИСТЕМИ

2.1 Діаграма прецедентів

Діаграма такого виду показує зв'язок між акторами та прецедентами системи, яка встановлює обмеження. Основними елементами даної діаграми є актори та прецеденти.

Актор у діаграмі прецедентів - це будь-який користувач чи об'єкт, який взаємодіє з системою.

Прецедент - це певна дія, яка виконується конкретним актором у системі. Прецедент може бути напряму пов'язаний з актором, з яким взаємодіє, або з іншим прецедентом, від виконання якого залежить виконання даного прецеденту.

У діаграмі прецедентів можуть бути такі варіанти взаємодії:

- використання;
- включення;
- розширення;
- вимога;
- схожість;
- рівнозначність.

Актор може бути пов'язаний з прецедентом напряму або через інший прецедент. Якщо немає прямого зв'язку між актором та прецедентом, то такий прецедент називається абстрактним.

Між прецедентами можуть бути такі зв'язки:

- Включення, або `include` - це взаємозв'язок, коли залежний прецедент обов'язково виконується на основі базового.
- Розширення, або `extend` - це різновид взаємозв'язку, коли базовий прецедент не обов'язково використовує дочірній [2, 4].

Для побудови діаграми прецедентів розроблюваної системи перш за все необхідно визначити акторів, які взаємодіють з нею.

Основні актори діаграми прецедентів:

- **Клієнт** - будь-яка особа, що може:
 - переглянути асортимент товарів у магазині;
 - замовити товар;
 - отримати консультацію.
- **Працівник** - співробітник магазину:
 - додавання інформації про товари на сайті магазину;
 - оформлення замовлень;
 - ведення звітності;
 - регулярне поповнення товарів на складі магазину.
- **Постачальник** - виступає посередником між магазином та виробником товарів, і має такі обов'язки:
 - постачання товарів;
 - надання звітності.
- **Аналітик** - співробітник, що займається однією зі складових діяльності магазину, зокрема аналізом інформації, а також відповідає за всю діяльність магазину, і виконує такі функції:
 - аналіз попиту споживачів, заснований на інформації про замовлення;
 - аналіз попиту на насіння виробників;
 - ведення звітності.

На рис. 2.1 представлена діаграма прецедентів розроблюваної системи.

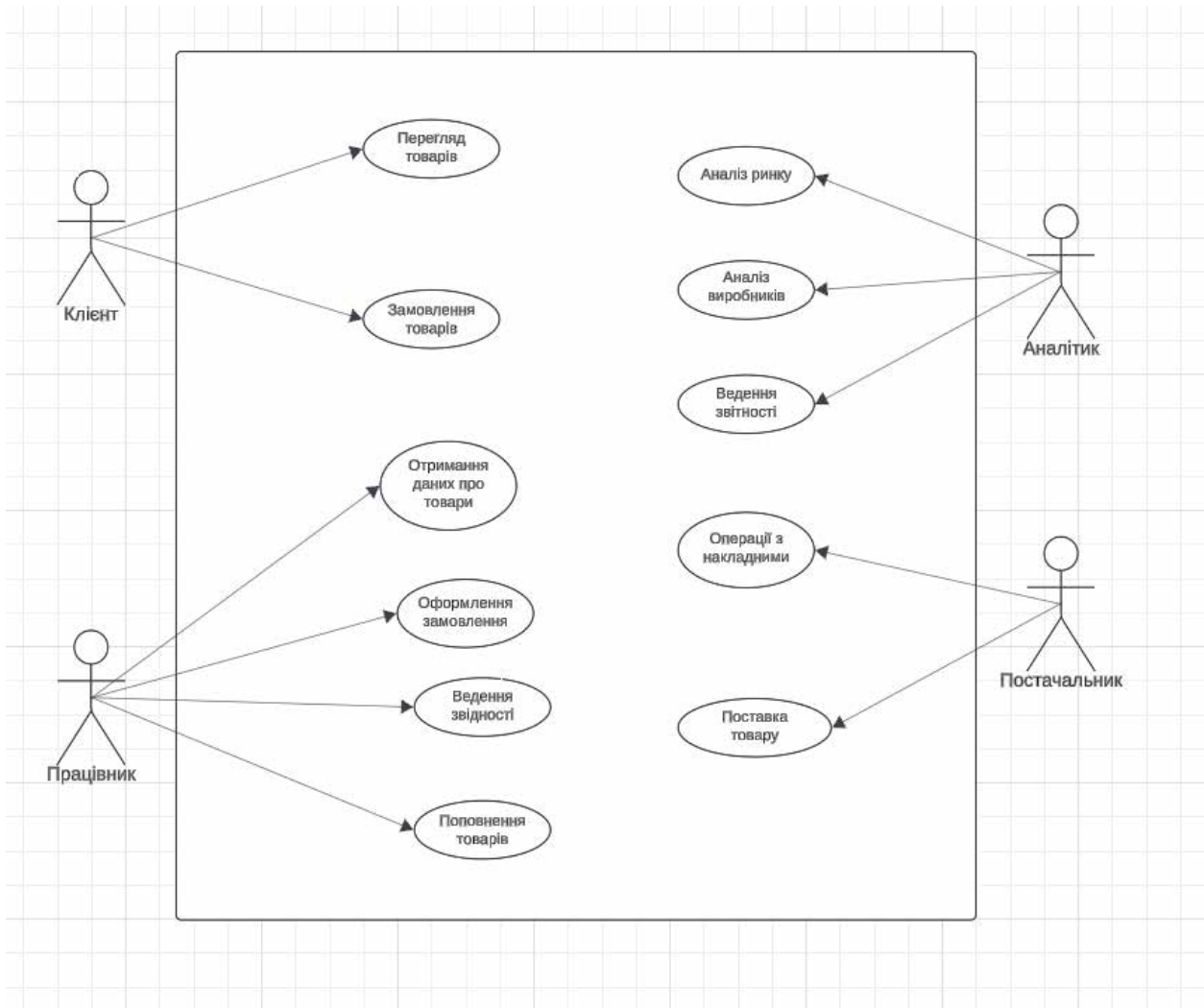


Рис. 2.1 Діаграма прецедентів

2.2 Архітектура системи

Архітектура системи аналізу для управління високоякісним насіннєвим матеріалом на рис. 2.2 складається з кількох ключових компонентів:

1. Device PC — це робоча станція, яка забезпечує доступ до системи через браузер із будь-якого комп'ютера, що підключений до інтернету.
2. Device Web Server — серверний вузол, який розміщує основні компоненти системи, включаючи:
3. Shop View — модуль, що дозволяє керувати операціями магазину та взаємодіяти з базою даних (Data Base).
4. Analysis View — компонент для виконання аналізу даних і формування звітів. Він працює із закритими даними, забезпечуючи їхню обробку.
5. Data Base — централізоване сховище даних, яке використовується для

управління інформацією, необхідною для оперативної роботи системи.

6. Data Warehouse — спеціалізоване сховище даних для тривалого зберігання архівованих даних і для подальшого аналізу.

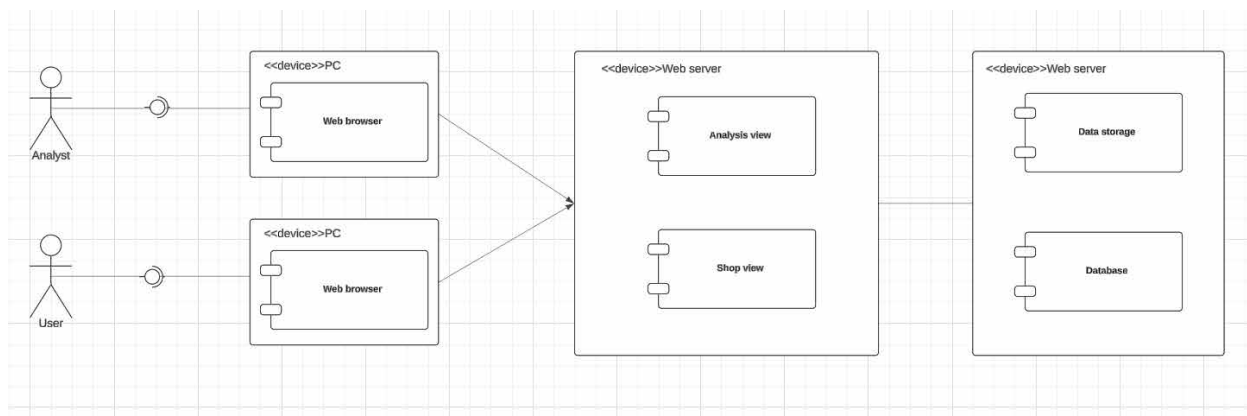


Рис. 2.2 Архітектура системи

2.3 Діаграма послідовності

Діаграма послідовності використовується для візуалізації взаємодії між об'єктами системи у часі. Вона показує, які повідомлення та в якій послідовності надсилаються між учасниками для виконання конкретного сценарію (прецеденту). Нижче на рис. 2.3 представлена діаграма послідовності для створення звіту аналітиком.

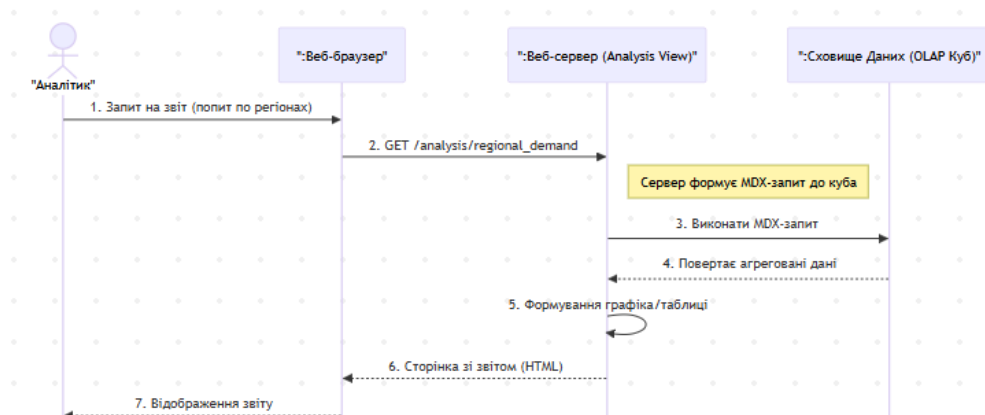


Рис. 2.3 Діаграма послідовності

2.4 Діаграма класів

Після створення діаграми послідовності можна створити діаграму класів.

Вона показує статичну структуру системи, тобто з яких "блоків" (класів) вона складається, які в них є дані (атрибути) та які між ними зв'язки. На відміну від діаграми послідовності, яка показує поведінку в часі, діаграма класів показує архітектуру даних.

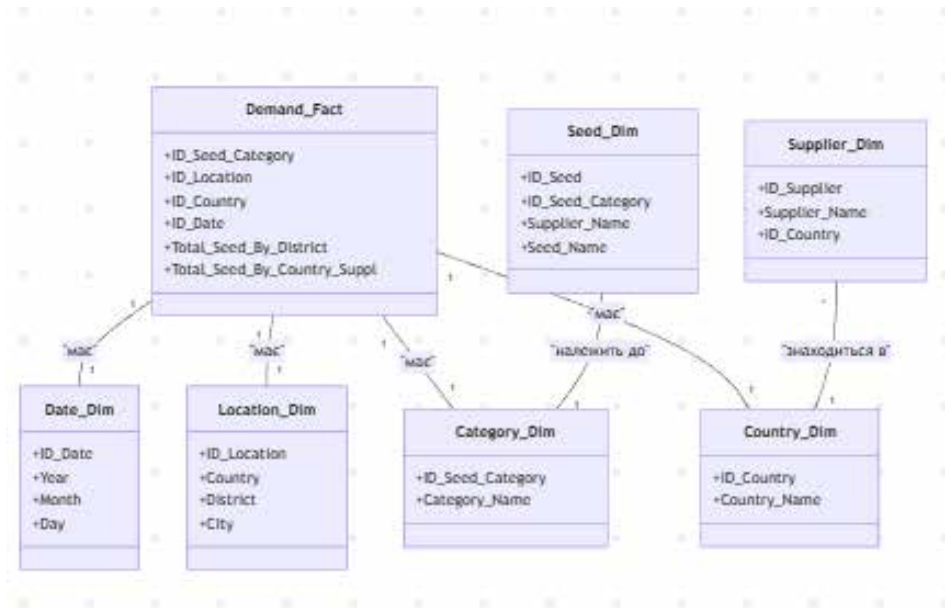


Рис. 2.4 Діаграма класів

Діаграма класів на рис. 2.4 показує структуру сховища даних (СД):

- **Центральний клас:** Demand_Fact (Таблиця фактів). У ньому зберігаються числові показники (Total_Seed_By_District) та зовнішні ключі (ID), які пов'язують його з вимірами.
- **Класи вимірів:** Date_Dim, Location_Dim, Category_Dim, Country_Dim та ін. Вони описують, що, де і коли відбувалося.

РОЗДІЛ 3. РОЗРОБКА СИСТЕМИ

3.1 Проектування реляційної бази даних

Джерелом даних системи виступає оперативна база даних представлена на рис. 3.1.

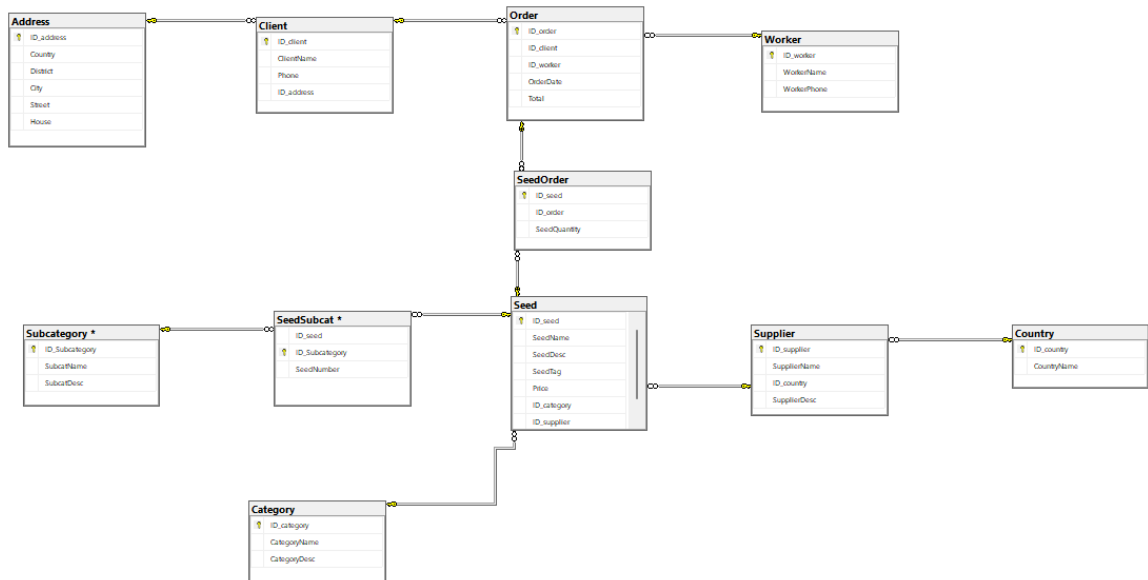


Рис. 3.1 Оперативна база даних

Оперативна БД містить такі таблиці:

- **Client**: інформація про клієнтів .
- **Address**: адреси клієнтів.
- **Order**: таблиця замовлень .
- **Worker**: інформація про співробітників .
- **Seed**: інформація про насіння .
- **Category**: довідник категорій насіння.
- **Subcategory**: підкатегорія насіння.
- **Supplier**: інформація про постачальників.
- **Country** : інформація про країну виробника.
- **SeedOrder**: таблиця-посередник між Seed та Order.

- SeedSubcat: таблиця-посередник між Subcategory та Seed.

Опис таблиць:

1. **Address:** Містить інформацію про місця доставки клієнтів. Поля:
 - ID_address – первинний ключ, унікальний ідентифікатор адреси.
 - Country – країна доставки.
 - District – область.
 - City – місто.
 - Street – вулиця.
 - House – номер будинку.
2. **Client:** Інформація про клієнтів магазину. Поля:
 - ID_client – унікальний ідентифікатор клієнта.
 - ClientName – ім'я клієнта.
 - Phone – номер телефону клієнта.
 - ID_address – зв'язок із адресою доставки.
3. **Order:** Дані про замовлення. Поля:
 - ID_order – ідентифікатор замовлення.
 - ID_client – зв'язок із клієнтом.
 - ID_worker – зв'язок із працівником.
 - OrderDate – дата замовлення.
 - Total – загальна сума замовлення.
4. **Worker:** Інформація про працівників магазину. Поля:
 - ID_worker – ідентифікатор працівника.
 - WorkerName – ім'я працівника.
 - WorkerPhone – телефон працівника.
5. **SeedOrder:** Посередня таблиця, що містить інформацію про замовлене насіння. Поля:
 - ID_seed – ідентифікатор насіння.
 - ID_order – ідентифікатор замовлення.

- SeedQuantity – кількість насіння в замовленні.
6. **Seed:** Інформація про насіння. Поля:
- ID_seed – унікальний ідентифікатор насіння.
 - SeedName – назва насіння.
 - SeedDesc – опис насіння.
 - SeedTag – тег насіння.
 - Price – ціна насіння.
 - ID_category – зв’язок із категорією.
 - ID_supplier – зв’язок із постачальником.
7. **SeedSubcat:** Посередня таблиця між насінням та підкатегоріями.

Поля:

- ID_seed – ідентифікатор насіння.
 - ID_subcategory – ідентифікатор підкатегорії.
 - SeedNumber – кількість насіння в підкатегорії.
8. **Subcategory:** Інформація про підкатегорії насіння. Поля:
- ID_subcategory – унікальний ідентифікатор підкатегорії.
 - SubcatName – назва підкатегорії.
 - SubcatDesc – опис підкатегорії.
9. **Category:** Інформація про категорії насіння. Поля:
- ID_category – унікальний ідентифікатор категорії.
 - CategoryName – назва категорії.
 - CategoryDesc – опис категорії.
10. **Supplier:** Інформація про постачальників насіння. Поля:
- ID_supplier – унікальний ідентифікатор постачальника.
 - SupplierName – назва постачальника.
 - SupplierDesc – опис постачальника.
 - ID_country – зв’язок із країною.
11. **Country:** Інформація про країни постачальників. Поля:
- ID_country – унікальний ідентифікатор країни.

- CountryName – назва країни.

3.2 Проектування сховища даних

Сховище даних (data warehouse) — це тип системи керування даними, яка призначена для забезпечення та підтримки діяльності бізнес-аналітики (BI). Сховища даних призначені виключно для виконання запитів та аналізу і часто містять великі обсяги історичних даних. Дані в сховищі даних зазвичай отримують з широкого діапазону джерел, таких як файли журналів програм і програми транзакцій.

Сховище даних централізує та консолідує великі обсяги даних із кількох джерел. Його аналітичні можливості дозволяють організаціям отримувати цінну бізнес-ідею зі своїх даних для покращення процесу прийняття рішень .

Під час розробки системи було створено сховище даних, яке дозволить проводити аналіз у різних розрізах.

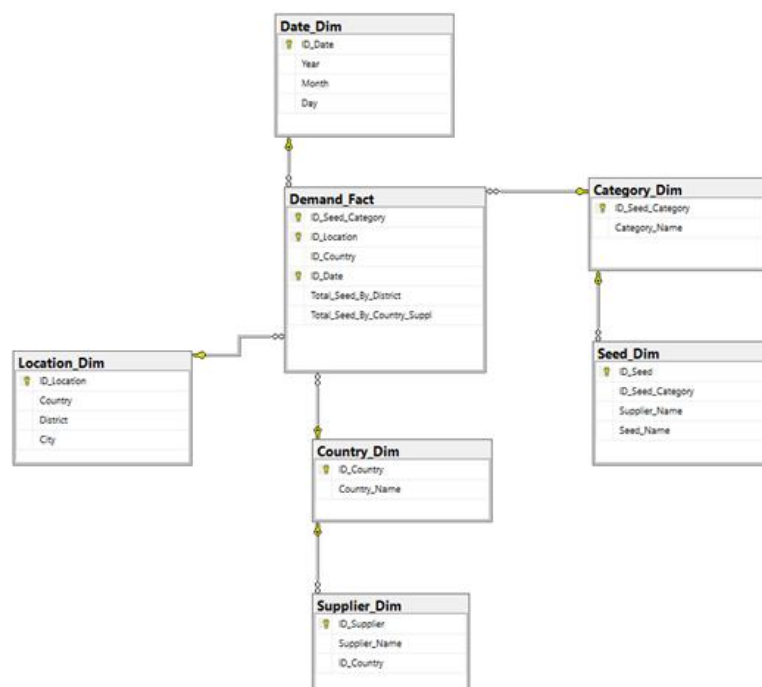


Рис. 3.2 Структура сховища даних

Опис таблиць у сховищі даних:

1. **Location_Dim**(рис. 3.3) - ця таблиця зберігає інформацію про місцезнаходження клієнтів. Вона містить такі поля:

- ID_Location (int) – первинний ключ, який однозначно ідентифікує місцезнаходження.
- Country (varchar(50)) – назва країни.
- District (varchar(50)) – назва області.
- City (varchar(50)) – назва міста.

2. **Category_Dim** (рис 3.4) - ця таблиця містить інформацію про категорії насіння. Вона включає:

- ID_Seed_Category (int) – первинний ключ, який ідентифікує категорію насіння.
- Category_Name (varchar(50)) – назва категорії.

3. **Date_Dim** (рис 3.5) - ця таблиця зберігає інформацію про дати. Вона включає такі поля:

- ID_Date (int) – первинний ключ, який ідентифікує дату.
- Year (char(4)) – рік.
- Month (char(2)) – місяць.
- Day (char(2)) – день.

4. **Supplier_Dim** (рис 3.6) - ця таблиця зберігає дані про виробників насіння. Вона містить такі поля:

- ID_Supplier (int) – первинний ключ, який ідентифікує виробника.
- Supplier_Name (varchar(50)) – назва виробника.
- ID_Country (int) – посилання на країну, де розташований виробник.

5. **Demand_Fact** - ця таблиця містить агреговані дані про попит. Вона включає такі поля:

- ID_Seed_Category (int) – посилання на категорію насіння.
- ID_Location (int) – посилання на місцезнаходження клієнта.
- ID_Country (int) – посилання на країну виробника.

- ID_Date (int) – посилання на дату.
 - Total_Seed_By_District (int) – кількість замовленого насіння у певному районі.
 - Total_Seed_By_Country_Suppl (int) – кількість насіння, замовленого у постачальників з конкретної країни.
6. **Seed_Dim**(рис. 3.7) – таблиця з інформацією про насіння:
- ID_Seed (int) – первинний ключ, який ідентифікує насіння.
 - Seed_Name (varchar(50)) – назва насіння.
 - ID_Seed_Category (int) – посилання на категорію насіння.
 - Supplier_Name (int) – назва постачальника.
7. **Country_Dim**(рис.3.8) – таблиця з інформацією про країни:
- ID_Country (int) – первинний ключ, який ідентифікує країну.
 - Country_Name (varchar(50)) – назва країни.

ID_Location	Country	District	City
1	Україна	Вінницька обл...	Вінниця
2	Україна	Волинська обл...	Луцьк
3	Україна	Дніпропетровс...	Дніпро
4	Україна	Донецька обла...	Донецьк
5	Україна	Житомирська ...	Житомир
6	Україна	Закарпатська ...	Ужгород
7	Україна	Запорізька об...	Запоріжжя
8	Україна	Івано-Франків...	Івано-Франків...
9	Україна	Київська облас...	Київ
10	Україна	Кіровоградськ...	Кропивницький
11	Україна	Луганська обл...	Луганськ
12	Україна	Львівська обла...	Львів
13	Україна	Миколаївська ...	Миколаїв
14	Україна	Одеська область	Одеса
15	Україна	Полтавська об...	Полтава
16	Україна	Рівненська об...	Рівне
17	Україна	Сумська облас...	Суми
18	Україна	Тернопільська ...	Тернопіль
19	Україна	Харківська обл...	Харків
20	Україна	Херсонська об...	Херсон
21	Україна	Хмельницька ...	Хмельницький
22	Україна	Черкаська обл...	Черкаси
23	Україна	Чернівецька о...	Чернівці
24	Україна	Чернігівська о...	Чернігів

Рис 3.3

ID_Seed_Cateq...	Name_Category
1	Зернові
2	Овочеві
3	Фрукти
4	Олійні культури
5	Бобові
6	Кормові культ...
7	Технічні культ...
8	Квіти
* NULL	NULL

Рис.3.4

ID_Date	Year	Month	Day
1	2024	1	1
2	2024	2	15
3	2024	3	3
4	2024	4	12
5	2024	5	22
6	2024	6	18
7	2024	7	27
8	2024	8	5
9	2024	9	14
10	2024	10	10
* NULL	NULL	NULL	NULL

Рис.3.5

ID_Supplier	Name_Supplier	ID_Country
1	АгроТех Постачання	1
2	Зелене Поле Корпорація	2
3	Майстри Насіння Індії	3
4	Рости Більше Китай	4
5	ЕкоНасіння Бразилія	5
6	ВрожайПро Україна	1
7	АгроНасіння США	2
8	ПриродаРіст Індія	3
9	ФермаПрямо Китай	4
10	РішенняДляРослин Бразилія	5

Рис.3.6

ID_Seed	ID_Seed_Cateq...	Name_Supplier	Seed_Name
1	1	АгроТрейд	Пшениця
2	1	ГрінСідс	Кукурудза
3	2	Українські Нас...	Картопля
4	2	Садко	Морква
5	3	ЕкоСад	Яблука
6	3	ЕкоСад	Груші
7	4	СонцеСід	Соняшник
8	4	АгроТрейд	Ріпак
9	5	ГрінСідс	Соя
10	5	Садко	Горох
11	6	ФлораУкраїна	Люцерна
12	6	ФлораУкраїна	Клевер
13	7	АгроТрейд	Цукровий буряк
14	7	ТабакСід	Тютюн
15	8	Квітка	Троянди
16	8	Квітка	Тюльпани

Рис.3.7

ID_Country	Name_Country
1	Україна
2	США
3	Індія
4	Китай
5	Бразилія
*	NULL

Рис.3.8

3.3 Розгортання OLAP-кубу

Служби аналізу SQL Server (SSAS) — це багатовимірний OLAP-сервер, а також механізм аналітики, який дозволяє розділяти великі обсяги даних. Він є частиною Microsoft SQL Server і допомагає виконувати аналіз за допомогою різних вимірів. Він має два варіанти багатовимірний і табличний.

Для розробки кубу було використано середовище Visual Studio з розширенням SSAS. На першому етапі необхідно визначити джерело даних — база даних OLAP або сховище даних. На основі визначеного джерела

даних будуть імпортуватись необхідні дані.

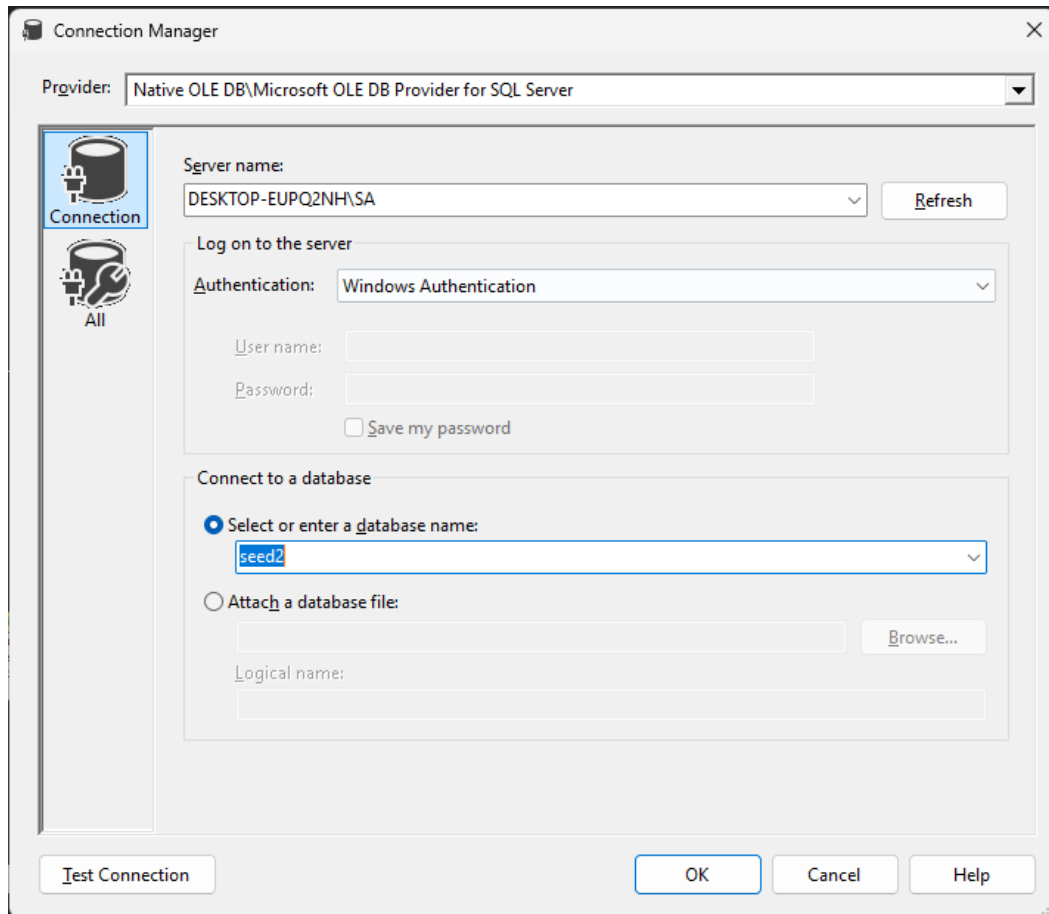


Рис. 3.9 Підключення сховища даних

Наступним етапом у розробці куба є створення джерела даних, яке являє собою абстракцію реляційного джерела даних, яка стає основою кубів і вимірів, які ви створюєте в багатовимірному проекті. Метою DSV є надання контролю над структурами даних, які використовуються у проекті, і працювати незалежно від базових джерел даних (наприклад, можливість перейменувати або об'єднати стовпці без безпосередньої зміни вихідного джерела даних). У процесі розробки можна створити кілька представлень джерел даних у проекті або базі даних Analysis Services на одному або кількох джерелах даних і створити кожне з них, щоб задовольнити вимоги до іншого рішення.

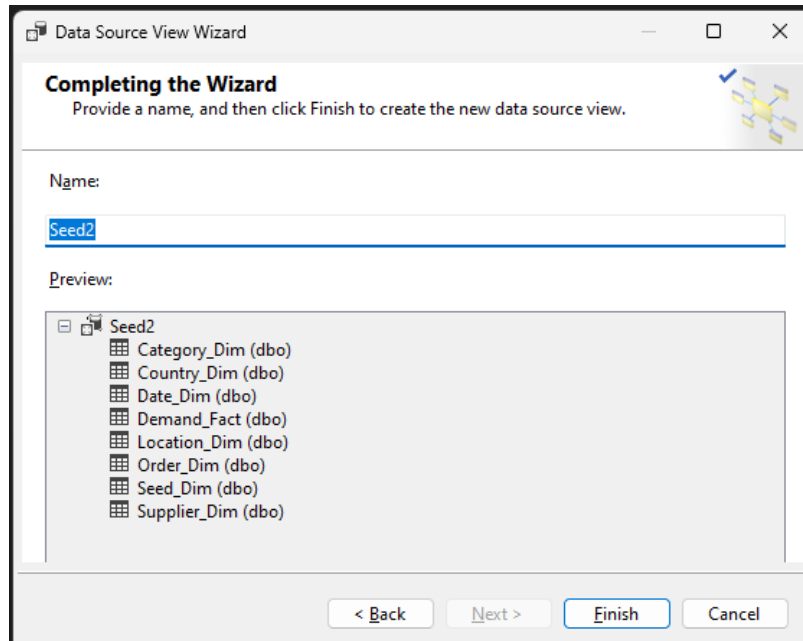


Рис. 3.10 Процес створення куба

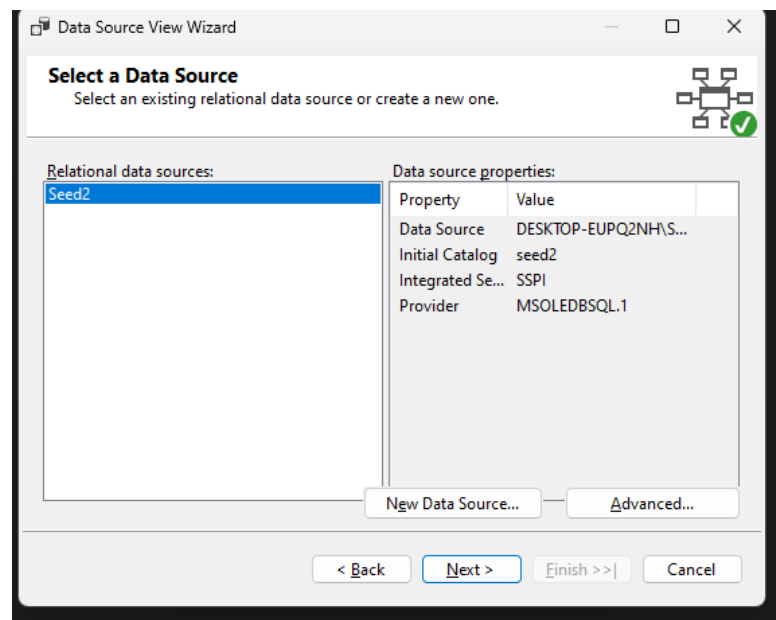


Рис. 3.11 Процес створення куба

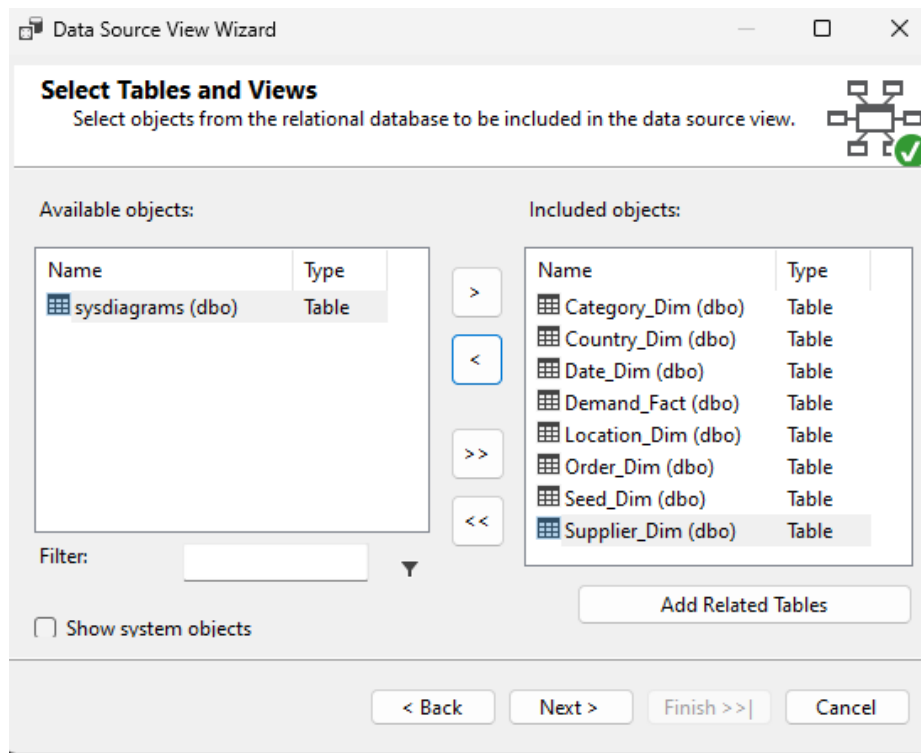


Рис. 3.12 Процес створення куба

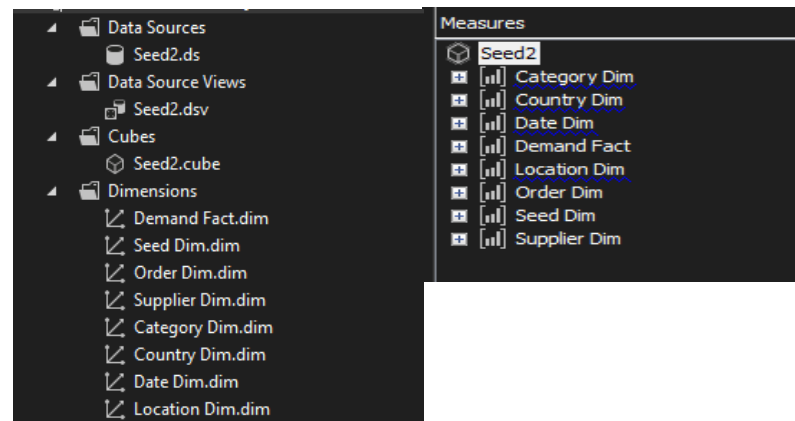


Рис.3.13 Структура куба

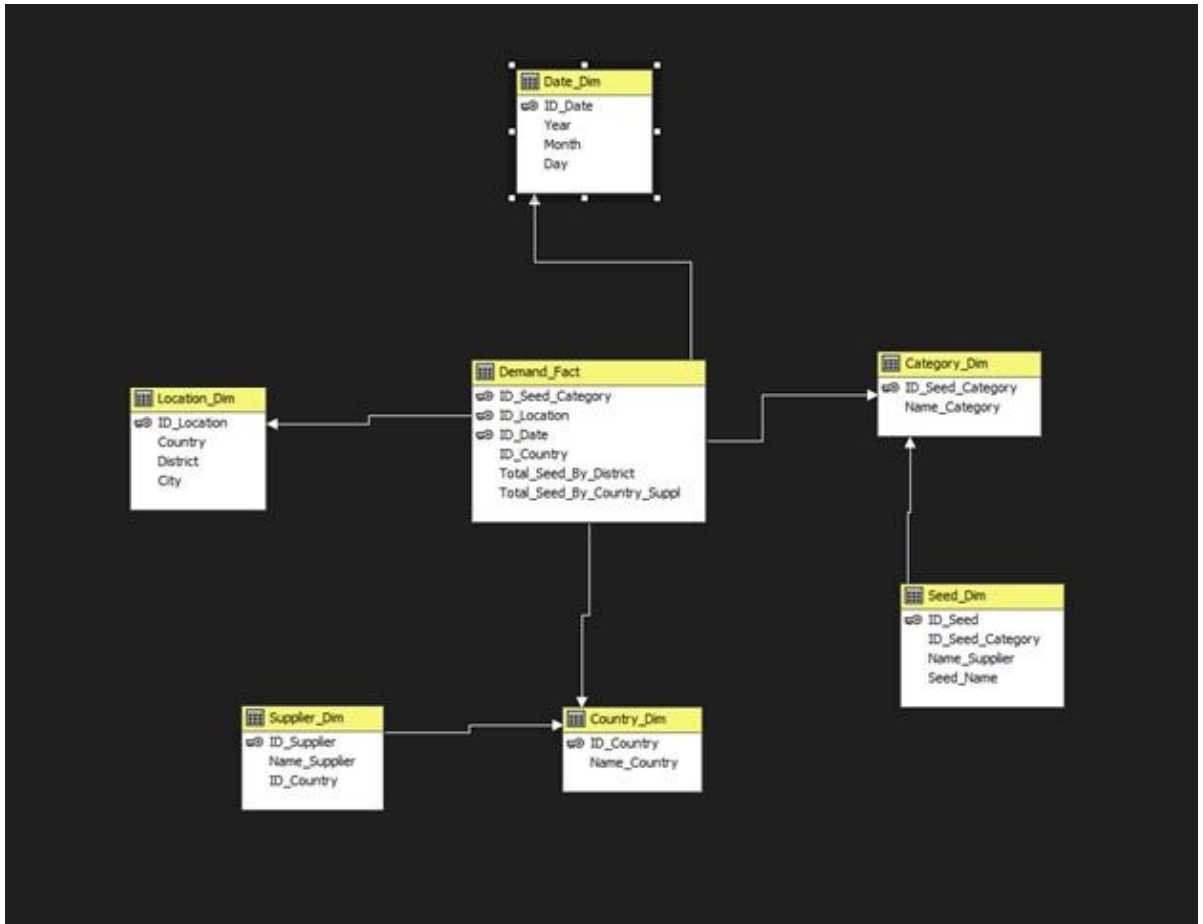


Рис. 3.14 Розгорнутий куб

Процес передачі даних було реалізовано за допомогою служби SQL Server Integration Services. SSIS – це інструмент, що дозволяє у зручному вигляді реалізувати інтеграцію, тобто. реалізувати процес перенесення даних з одного джерела до іншого. Цей процес іноді називають ETL (від англ. Extract, Transform, Load – дослівно «вилучення, перетворення, завантаження»).

В інструменті SSIS є служба Data Flow, за допомогою якої було проведено заповнення таблиць вимірів та фактів.

3.4 Отримання даних за допомогою Data Flow

Для забезпечення аналітичної системи актуальними даними необхідно реалізувати механізм їх перенесення з оперативної бази даних до сховища даних (СД). Цей процес, відомий як ETL (Extract, Transform, Load), у даній

роботі реалізовано за допомогою SQL Server Integration Services (SSIS) .

SSIS — це інструмент, що дозволяє у зручному візуальному вигляді проєктувати процеси інтеграції та перетворення даних. Ключовим компонентом для цього є служба Data Flow.

Було створено кілька потоків даних (Data Flow Tasks) для послідовного наповнення таблиць вимірів та таблиці фактів.

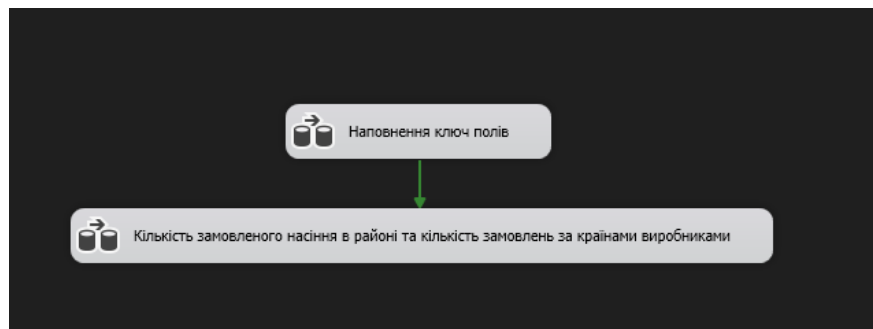


Рис. 3.15 Потоки даних для наповнення СД

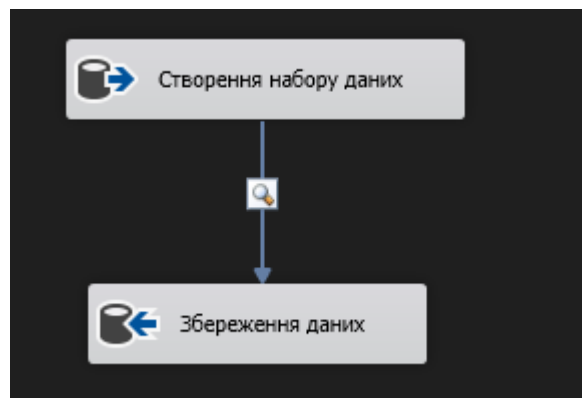


Рис. 3.16 Перший крок наповнення(Наповнення ключ полів)

Блок “Створення набору даних” – блок для сформування набору даних, над якими будуть проведені операції.

Блок “Збереження даних” – блок призначений для збереження даних у визначені таблиці.

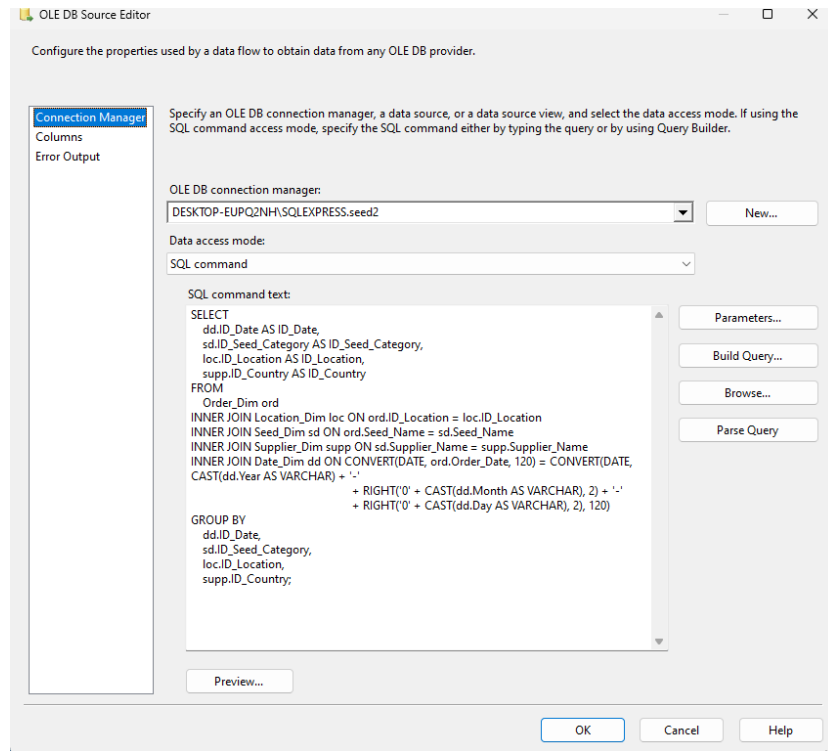


Рис.3.17 Запит у блоці “Створення набору даних”

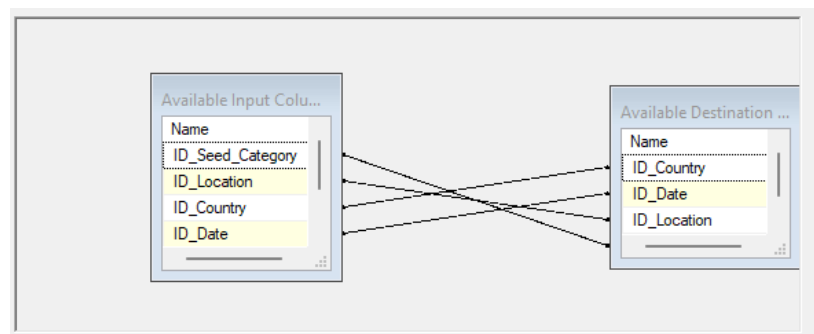


Рис. 3.18 Блок “Збереження даних”

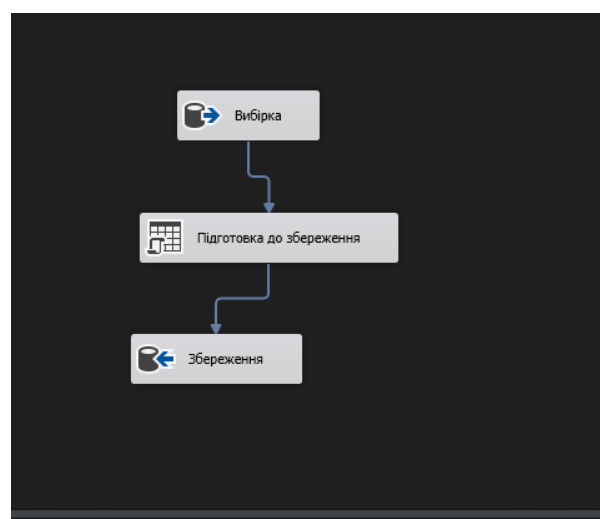


Рис.3.19 Блок ”Кількість замовленого насіння в районі та кількість замовлень за країнами виробниками”

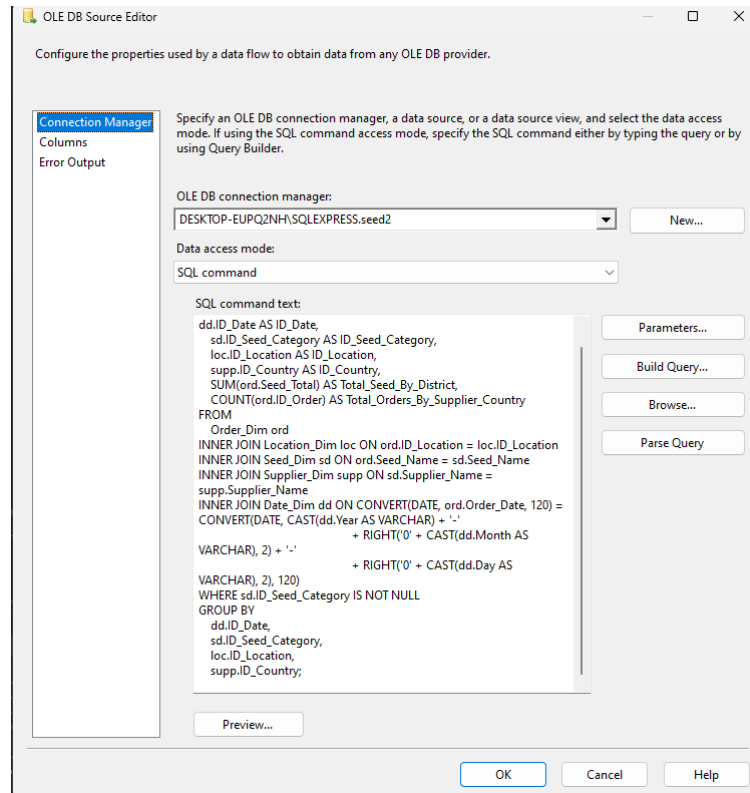


Рис. 3.20 Запит у блоці “Вибірка”

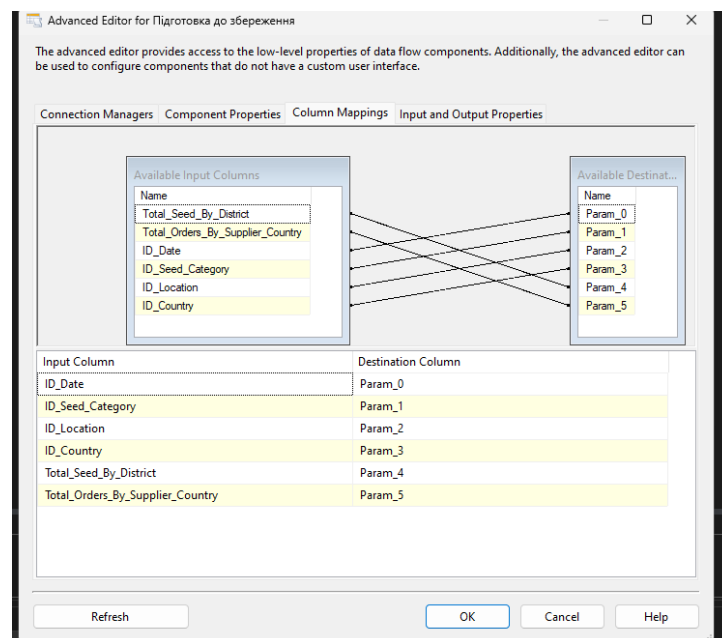


Рис. 3.21 Блок ”Підготовка до збереження”

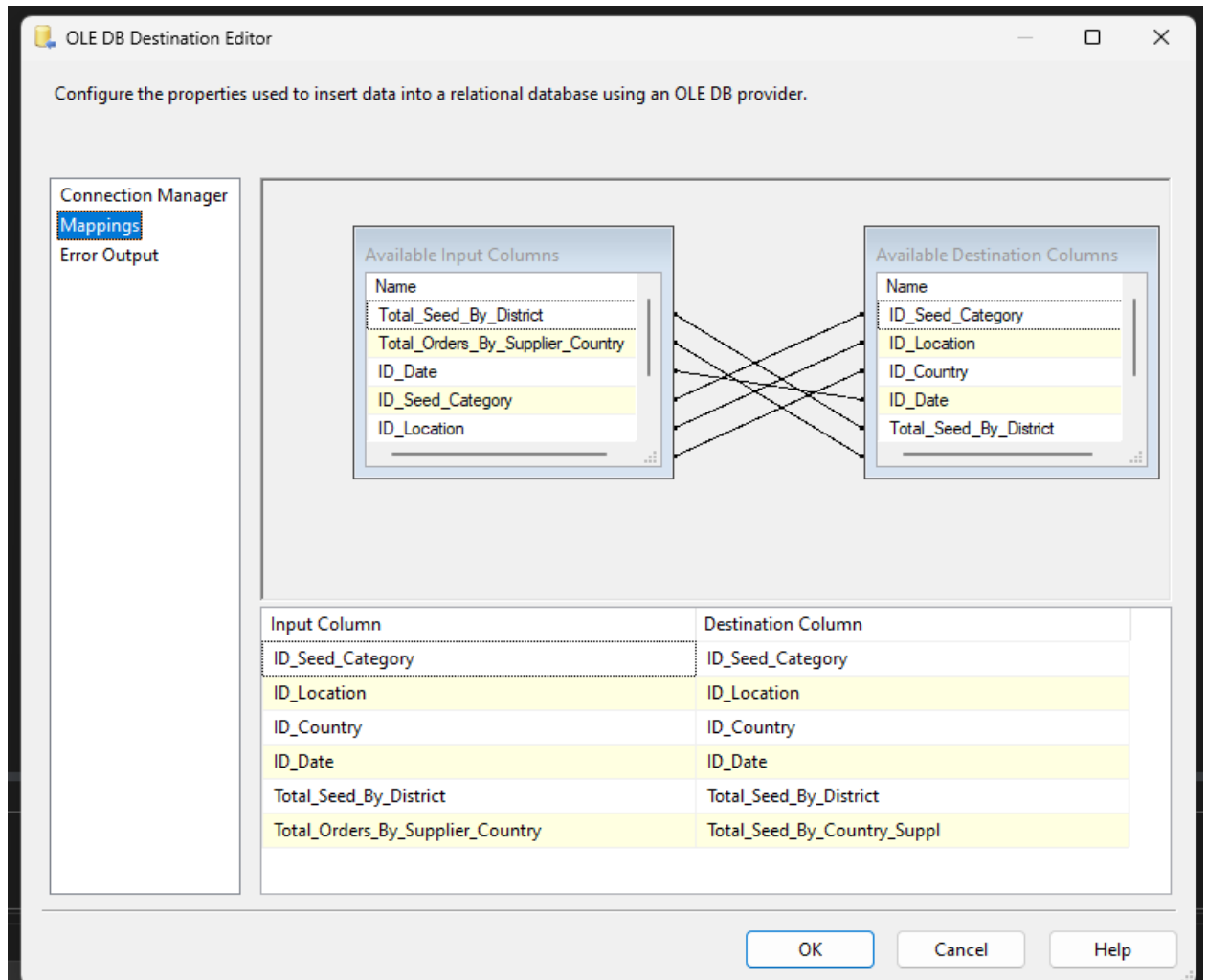


Рис. 3.22 Блок "Збереження даних"

	ID_Seed_Cateq...	ID_Location	ID_Country	ID_Date	Total_Seed_By_District	Total_Seed_By_Country_Suppl
▶	1	1	1	1	500	1
	1	2	2	2	300	1
	2	3	3	3	200	1
	2	7	3	7	350	1
	4	4	1	4	150	1
	5	5	1	5	400	1
	6	10	2	10	220	1
	7	8	4	8	180	1
	8	6	5	6	50	1
	8	9	5	9	70	1
•	NULL	NULL	NULL	NULL	NULL	NULL

Рис. 3.23 Дані, занесені у СД після виконаних потоків

РОЗДІЛ 4. РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

4.1 Вимоги до апаратного та програмного забезпечення

Для забезпечення належного функціонування та розробки програмного забезпечення системи аналізу реалізації насінневого матеріалу, сформульовано вимоги до апаратних та програмних засобів. Вимоги розділено на дві категорії: конфігурація для розробника системи та конфігурація для кінцевого користувача (аналітика).

Вимоги до середовища розробника

Це конфігурація, необхідна для створення, налаштування та адміністрування сховища даних, OLAP-кубів, ETL-процесів та аналітичних моделей.

Апаратне забезпечення (Рекомендовано):

- Процесор: 64-розрядний процесор з тактовою частотою 2.0 ГГц або вище (наприклад, Intel Core i5/i7 або AMD Ryzen 5/7).
- Оперативна пам'ять (RAM): Мінімум 16 ГБ (рекомендовано 32 ГБ для комфортної роботи з Visual Studio, SQL Server та аналітичними моделями).
- Вільне місце на диску: Мінімум 40 ГБ для встановлення Visual Studio, SQL Server та компонентів.

Програмне забезпечення:

- Операційна система: Windows 10 (64-bit) або Windows 11 (64-bit).
- Середовище розробки (IDE): Microsoft Visual Studio 2022 (наприклад, Enterprise Edition), з встановленими робочими навантаженнями:
- "Data storage and processing" (Зберігання та обробка даних).
- "Data science and analytical applications" (для інтеграції з Python).
- Сервер Баз Даних: Microsoft SQL Server 2022 (наприклад, Developer або Evaluation Edition).
- Ключові компоненти SQL Server:

- Database Engine Services (для оперативної БД та СД).
- Analysis Services (SSAS), встановлений у багатовимірному режимі (Multidimensional Mode) (для OLAP-кубів та Data Mining).
- Integration Services (SSIS) (для ETL-процесів).
- Інструменти керування: SQL Server Management Studio (SSMS) версії 21 або новішої.
- Розширення Visual Studio:
- SQL Server Integration Services Projects.
- Microsoft Analysis Services Projects (для роботи з кубами).
- Microsoft Reporting Services Projects (для звітів SSRS).
- Середовище Python (для Data Mining): Anaconda або стандартний Python 3.9+ з бібліотеками: pandas, prophet, scikit-learn, mlxtend, streamlit.

Вимоги до середовища кінцевого користувача (Аналітика)

Це конфігурація, необхідна аналітику або менеджеру для доступу та використання розробленого програмного забезпечення (інтерактивного дашборду).

Апаратне забезпечення (Мінімальні):

- Будь-який сучасний персональний комп'ютер (ПК) або мобільний пристрій, здатний запустити веб-браузер.
- Стабільне підключення до мережі (Інтернет або внутрішньої мережі компанії).
- Операційна система: Будь-яка (Windows, macOS, Linux, Android, iOS).
- Веб-браузер: Сучасний браузер, такий як Google Chrome, Mozilla Firefox, Safari або Microsoft Edge.

4.2 Побудова звітності

Під час проведення аналізу даних одним з найважливіших етапів є

формування звітності. Саме за допомогою звітів аналітик та підприємство можуть побачити цілісну картину розвитку підприємства.

Для створення звітної інформації та візуалізації результатів у даній роботі, замість традиційних BI-систем, було використано гнучкий та потужний стек технологій з екосистеми Python. Це забезпечує повний контроль над обробкою даних та широкі можливості для кастомізації візуалізацій.

Основними інструментами для формування звітів стали:

- **Pandas:** Провідна бібліотека Python для маніпуляції та аналізу даних. Вона була використана для завантаження даних з CSV-файлу, а також для їх агрегації, групування (`.groupby()`) та сортування (`.sort_values()`).
- **Matplotlib та Seaborn:** Бібліотеки для візуалізації даних. Вони дозволили створити чіткі та інформативні графіки (стовпчасті діаграми, секторні діаграми) для представлення знайдених закономірностей.

Процес створення звітів у Python починається із завантаження джерела даних у `DataFrame`. Далі, замість використання "майстра" чи "конструктора" звітів, застосовуються програмні методи для формування вибірки. Дані агрегуються за допомогою `.groupby()` та сортуються, а потім візуалізуються за допомогою виклику відповідних функцій `sns.barplot()` або `plt.pie()`. Це дозволяє наочно продемонструвати результати вибірки.

Для відображення інформації було створено звіт, який дає відповіді на питання:

1. З яких областей було здійснено найбільше покупок насіння?
2. Яка категорія насіння має найбільшу кількість продажів?

Створений код для формування вибірки:

```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns
```

```

df = pd.read_csv('analytics_data_large.csv')

# --- 1. Аналіз: З яких областей було найбільше покупок? ---

district_demand =
df.groupby('District')['Total_Seed_By_District'].sum().sort_values(ascending=False)

print("Таблиця попиту за областями (Топ-10):")

print(district_demand.head(10).to_string()) # .to_string()

plt.figure(figsize=(12, 8))

sns.barplot(x=district_demand.head(10).values, y=district_demand.head(10).index,
palette='viridis')

plt.title('Топ-10 областей за попитом на насіння', fontsize=16)

plt.xlabel('Загальний попит (одиниць)', fontsize=12)

plt.ylabel('Область', fontsize=12)

plt.show()

```

Результатом виконання коду є графік який показує топ-10 областей за попитом на насіння(Рис. 4.1).

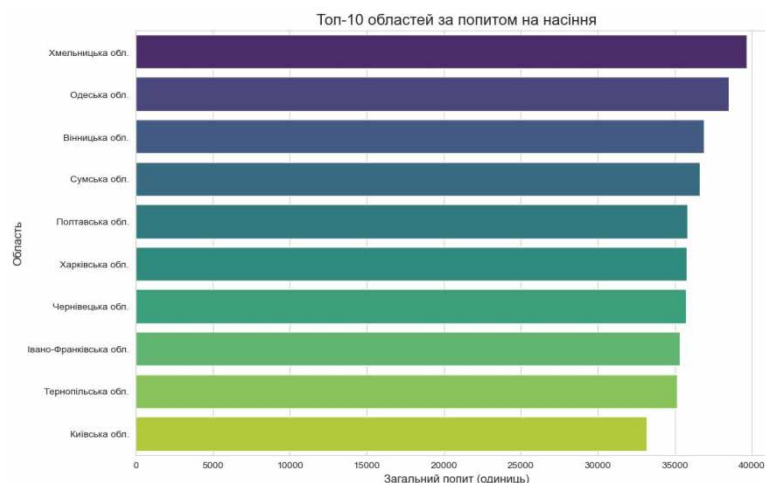


Рис. 4.1 Графік областей за попитом на насіння

```

# --- 2. Аналіз: Яка категорія насіння має найбільшу кількість продажів? ---

```

```

category_demand =
df.groupby('Category_Name')['Total_Seed_By_District'].sum().sort_values(ascending=False)

print("Таблиця попиту за категоріями:")

print(category_demand.to_string())

plt.figure(figsize=(10, 10))

colors = sns.color_palette('viridis', len(category_demand))

plt.pie(category_demand, labels=category_demand.index, autopct='%1.1f%%',
startangle=140, colors=colors)

plt.title('Частка категорій насіння у загальному попиті', fontsize=16)

plt.axis('equal')

plt.show()

```

Результатом виконання коду є діаграма яка показує частки категорій насіння у загальному попиті(рис. 4.2).

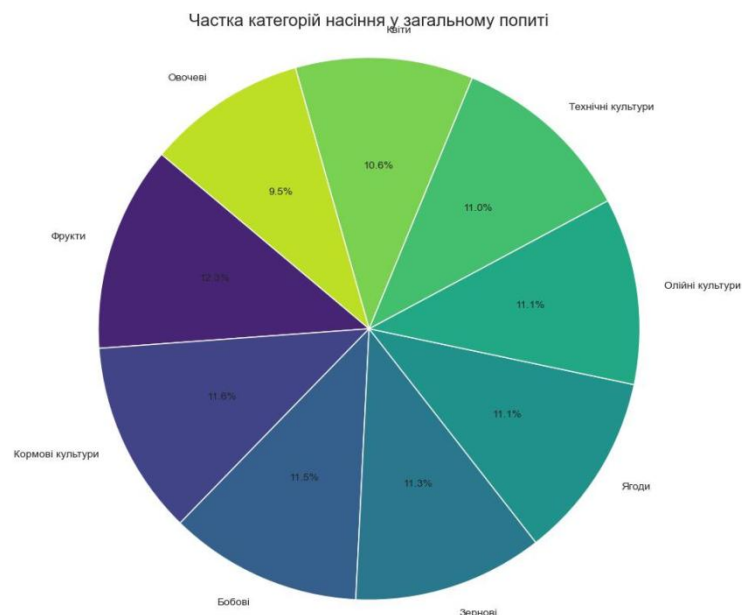


Рис.4.2 Попит на категорії насіння

4.3 Розрахунок KPI

Для визначення успішності роботи підприємства аналітики обчислюють коефіцієнт ефективності.

KPI - це набір обчислень, пов'язаних з групою показників у кубі, які використовуються для оцінки успіху бізнесу. Зазвичай ці обчислення є комбінацією виразів багатовимірних виразів (MDX) або обчислюваних елементів. KPI також мають додаткові метадані, які надають інформацію про те, як клієнтські програми повинні відображати результати обчислень KPI. KPI обробляє інформацію про поставлену ціль, фактичну формулу продуктивності, записану в кубі, і вимірювання, щоб показати тенденцію та статус ефективності.

Компоненти для розрахунку KPI:

- Ціль - числовий вираз MDX або обчислення, яке повертає цільове значення KPI.
- Значення - числовий вираз MDX, який повертає фактичне значення KPI.
- Статус - вираз MDX, який представляє стан KPI на певний момент часу. Вираз MDX стану має повертати нормалізоване значення від -1 до 1. Значення, рівні або менші за -1, будуть інтерпретуватися як "погані" або "низькі". Значення нуль (0) інтерпретується як «прийнятне» або «середнє». Значення, що дорівнюють або перевищують 1, будуть інтерпретуватися як «добре» або «високе».

Для аналізу фактів, занесених у куб, було визначено таке значення KPI:

Seed_Sell - визначення ефективності реалізації насіння по всій країні за 2024 рік.



Рис.4.3 Результат обрахунку KPI

Ціль магазину полягає в реалізації 300000 одиниць насіння в рік. План був виконаний на 93 відсотки. Магазин не досяг цілі зовсім трохи, але і не зазнав збитку.

4.4 Інтелектуальний аналіз даних за допомогою Data Mining

Після побудови та наповнення сховища даних, яке агрегує інформацію, наступним кроком є застосування методів інтелектуального аналізу даних (Data Mining) для виявлення прихованих закономірностей. На відміну від вбудованих інструментів MS SQL Server, для даного дослідження було обрано більш гнучкий та потужний підхід з використанням мови Python та її аналітичних бібліотек (pandas, prophet, scikit-learn, mlxtend).

Цей підхід дозволяє відповісти на ключові бізнес-питання, такі як сегментація ринку, прогнозування попиту та аналіз купівельної поведінки.

1. Прогнозування попиту (Метод Prophet)

Завдання: Спрогнозувати загальний попит на насіння на 2025 рік, щоб оптимізувати планування закупівель та уникнути дефіциту товару.

Реалізація: Було використано бібліотеку Prophet від Facebook, яка спеціально розроблена для прогнозування часових рядів з сильною сезонністю. Модель була навчена на даних про попит за 2023-2024 роки.

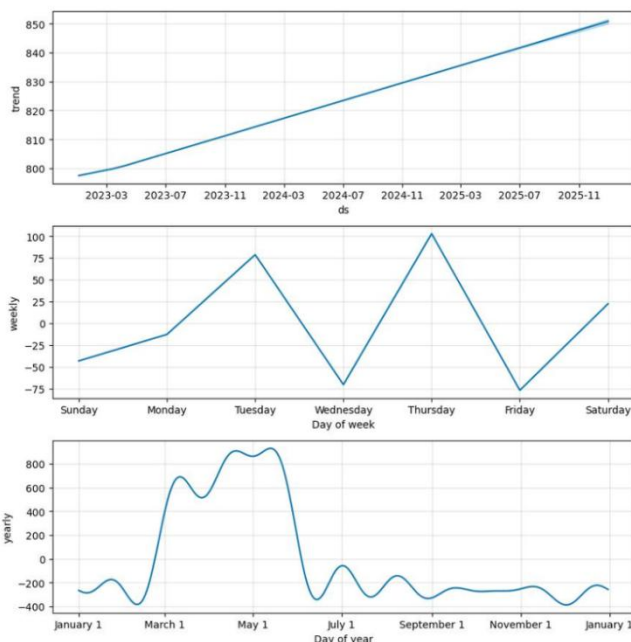
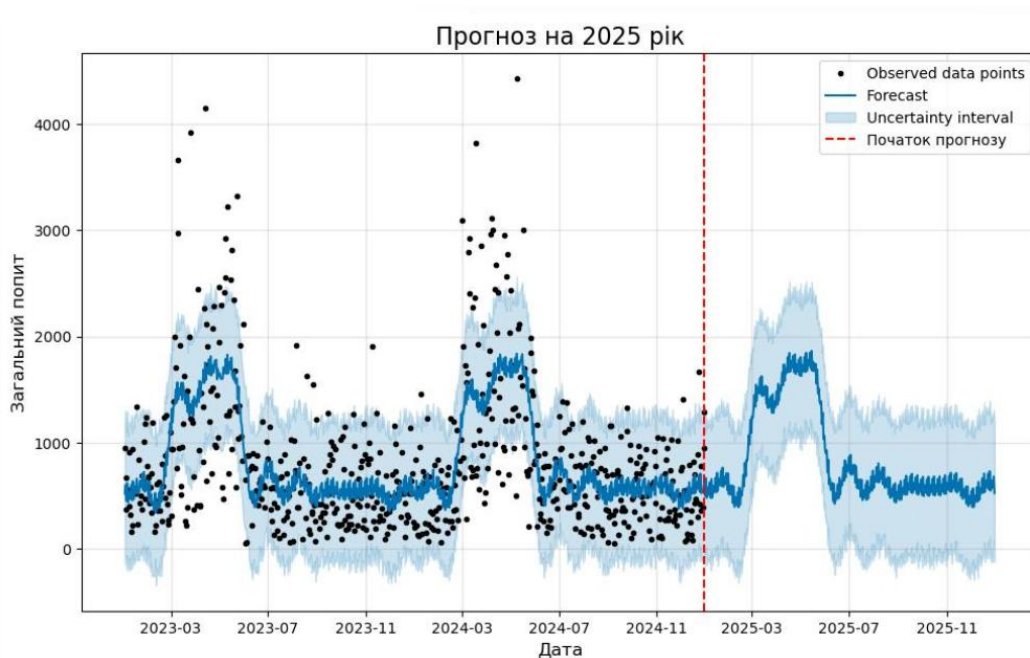


Рис.4.4 Результат прогнозування

Результати прогнозування попиту(рис. 4.4):

- **Графік прогнозу:** Модель успішно ідентифікувала ключові патерни в даних. На графіку видно, як прогноз (синя лінія) коректно відтворює річну сезонність (очікуваний пік попиту навесні 2025 року), спираючись на історичні дані (чорні точки).
- **Компоненти:** Аналіз компонентів моделі підтвердив наявність зростаючого тренду та сильної тижневої (спад у середу) і річної (пік у квітні-травні) сезонності.
- **Оцінка моделі:** Розрахований R-квадрат (R^2) = 0.3813. Це значення вказує на те, що щоденні дані є дуже (волатильними). Проте, модель коректно вловила загальний тренд та сезонні патерни, що є головним для стратегічного планування.

2. Кластеризація ринку (Метод K-Means)

Завдання: Сегментувати ринок, відповівши на питання: "З яких регіонів надходять замовлення?" та "Чи існують групи областей зі схожою купівельною поведінкою?".

Реалізація: Було використано алгоритм K-Means з бібліотеки scikit-learn. Алгоритм проаналізував області України за профілями їхнього попиту на різні категорії насіння. Для візуалізації багатовимірних результатів було використано метод головних компонент (PCA).

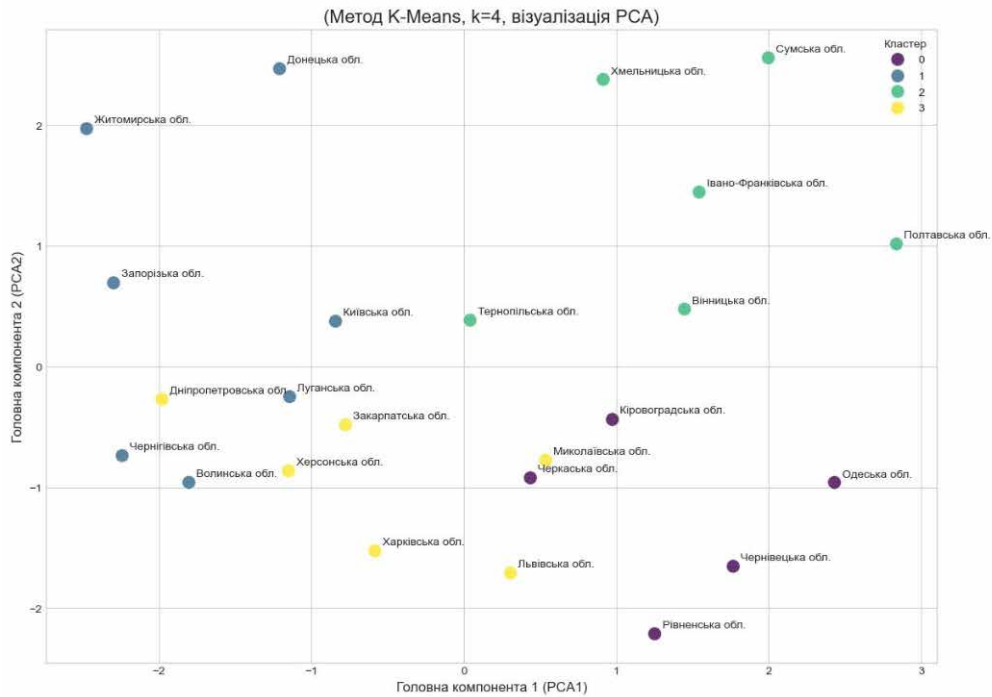


Рис.4.5 Результат кластеризації

Результати кластеризації(рис. 4.5):

Сегментація: Алгоритм успішно розділив ринок на 4 унікальні кластери (групи) регіонів з яких виділив кластери з унікальним профілем попиту:

- Кластер 0: Фокус на Технічних культурах та Ягодах.
- Кластер 1: Переважають Фрукти та Олійні культури.
- Кластер 2: Основний попит на Зернові та Олійні культури.
- Кластер 3: Високий попит на Квіти та Кормові культури.

2D-графік PCA чітко показує, як області з одного кластера (одного кольору) групуються разом, підтверджуючи схожість їхніх моделей попиту.

Це дає змогу розробляти цільові маркетингові стратегії для кожної групи регіонів.

3. Пошук асоціативних правил (Алгоритм Apriori)

Мета полягала у виявленні не випадкових зв'язків між трьома ключовими атрибутами замовлення: Category_Name, Country_Name (Країна-постачальник) та District. Ми шукали відповідь на питання: "Які комбінації регіонів, категорій та країн-постачальників з'являються разом частіше, ніж можна було б очікувати випадково?"

Виявлення правил: Алгоритм виявив стійкі зв'язки між характеристиками замовлень.

Результати пошуку асоціативних правил(рис 4.6):

	Якщо (замовлення)	То (Наслідок)	Lift (Невипадковість)	Confidence (Впевненість)	Support (Підтримка)
89	Country_Name_Іспанія	District_Запорізька обл.	3.008021	0.132353	0.018
88	District_Запорізька обл.	Country_Name_Іспанія	3.008021	0.409091	0.018
149	District_Львівська обл.	Country_Name_Україна	2.700617	0.388889	0.014
64	District_Одеська обл.	Category_Name_Технічні культури	2.631579	0.263158	0.010
48	District_Житомирська обл.	Category_Name_Олійні культури	2.539683	0.320000	0.016
49	Category_Name_Олійні культури	District_Житомирська обл.	2.539683	0.126984	0.016
29	District_Херсонська обл.	Category_Name_Квіти	2.532928	0.238095	0.010
28	Category_Name_Квіти	District_Херсонська обл.	2.532928	0.106383	0.010
97	District_Івано-Франківська обл.	Country_Name_Китай	2.307692	0.300000	0.018
96	Country_Name_Китай	District_Івано-Франківська обл.	2.307692	0.138462	0.018
91	District_Луганська обл.	Country_Name_Іспанія	2.297794	0.312500	0.010
101	District_Дніпропетровська обл.	Country_Name_Китай	2.262443	0.294118	0.010
72	Category_Name_Фрукти	District_Донецька обл.	2.248876	0.103448	0.012
73	District_Донецька обл.	Category_Name_Фрукти	2.248876	0.260870	0.012
95	District_Черкаська обл.	Country_Name_Іспанія	2.237852	0.304348	0.014

Рис.4.6 Результат пошуку асоціативних правил

1. Якщо (замовлення) (Antecedent): Це умова або посилка. Це те, що ми вже бачимо у "кошику" (тобто в одному рядку-замовленні).

2. То (Наслідок) (Consequent): Це те, що, як передбачає правило, також буде в цьому "кошику".

3. Lift (Невипадковість): Це показник "цікавості" правила. Він показує, наскільки зв'язок між "Якщо" і "То" сильніший за чисту випадковість.

- Lift = 1: Зв'язку немає, це чиста випадковість.

- Lift > 1: Зв'язок позитивний.

- Lift < 1: Зв'язок негативний (тобто, якщо є "Якщо", то "То", ймовірно, не буде).

4. Confidence (Впевненість): Це % ймовірності. Він відповідає на питання: "Якщо ми вже бачимо в замовленні Якщо (замовлення), то з якою

ймовірністю там також буде То (Наслідок)?" Confidence = 0.409 означає 41% ймовірність.

5. **Support (Підтримка):** Це показник того, наскільки часто це правило взагалі трапляється у даних.

Аналіз ключових правил:

Правило №88: District_Запорізька обл. -> Country_Name_Іспанія

- Значення: Якщо замовлення йде до Запорізької області, то з 41% впевненістю воно буде з Іспанії.
- Важливість: Показник Lift (3.008) є надзвичайно високим. Це означає, що зв'язок між цим регіоном та Іспанією в 3 рази сильніший за будь-який випадковий збіг. Це вказує на чітку бізнес-закономірність (можливо, ексклюзивний постачальник для регіону або регіональна спеціалізація).

Правило №149: District_Львівська обл. -> Country_Name_Україна

- Значення: Якщо замовлення йде до Львівської області, то з 39% впевненістю це буде насіння від українського постачальника.
- Важливість: Lift (2.7) також дуже високий, що вказує на сильний пріоритет внутрішнього ринку для цього регіону.

Правило №64: District_Одеська обл. -> Category_Name_Технічні культури

- Значення: Якщо замовлення йде до Одеської області, то з 26% впевненістю це будуть Технічні культури.
- Важливість: Lift (2.6) показує сильний зв'язок між регіоном та специфічною категорією товару, що свідчить про аграрну спеціалізацію регіону.

Ці правила допомагають в оптимізації асортименту (крос-продажі) та логістики.

4. Спрощений алгоритм Байєса

Мета: "Чи можемо ми передбачити обсяг замовлення (велике воно чи мале), базуючись на тому, що замовляють, звідки і коли?"

```

Ознаки (X): ['Category_Name', 'Country_Name', 'Month']
Медіанний обсяг замовлення: 272.0
Цільова змінна (Y): 'High_Volume' / 'Low_Volume' (створено успішно)
Розмір вибірки: 2500 (Train: 1750, Test: 750)

Точність моделі (Accuracy): 0.6907 (або 69.1%)

--- Звіт по класифікації (High/Low Volume) ---
              precision    recall  f1-score   support

High_Volume    0.85     0.46     0.59     373
Low_Volume     0.63     0.92     0.75     377

   accuracy                   0.69     750
  macro avg    0.74     0.69     0.67     750
weighted avg    0.74     0.69     0.67     750

```

Рис.4.7 Спрощений алгоритм Байєса

1. Створення цільової змінної (Y): Було взято стовпець Total_Seed_By_District (Обсяг).

- Знайдено медіанне (середнє) значення: 272.0.

- Створено новий бінарний стовпець: якщо замовлення > 272, воно отримувало клас High_Volume (Велике), інакше - Low_Volume (Мале).

2. Вибір ознак (X): Для прогнозу було обрано три логічні категоріальні ознаки: Category_Name, Country_Name та Month.

3. Модель: Використано класифікатор CategoricalNB (Категоріальний Наївний Баєс), який ідеально підходить для роботи з категоріальними ознаками.

4. Навчання: Дані були розділені у співвідношенні 70% (1750 записів) на навчання та 30% (750 записів) на тестування.

Інтерпретація результатів

Головний результат: Точність моделі (Accuracy): 0.6907 (або 69.1%)

Це означає що модель правильно вгадала клас (Велике/Мале) у 69.1% випадків на тестових даних, яких вона раніше не бачила.

Важливість: Для бінарної класифікації випадкове вгадування дало б 50%. Цей результат на 19.1% кращий за випадковість, що вказує на те, що обрані нами ознаки (Категорія, Країна, Місяць) дійсно мають статистично значущий вплив на обсяг замовлення.

1. Детальний аналіз (Звіт по класифікації):

- precision (Точність): Наскільки можна довіряти прогнозу моделі.
- recall (Повнота): Наскільки добре модель знаходить усі реальні випадки.

2. Аналіз класу High_Volume (Великі замовлення):

- precision: 0.85: Це найсильніший показник моделі. Якщо модель каже "Це велике замовлення", вона права у 85% випадків. Це робить її надійним інструментом для підтвердження великих замовлень.

- recall: 0.46: Модель знаходить лише 46% (менше половини) усіх реальних великих замовлень. Вона пропускає багато великих замовлень, класифікуючи їх як малі.

3. Аналіз класу Low_Volume (Малі замовлення):

- precision: 0.63: Якщо модель каже "Це мале замовлення", вона права у 63% випадків.

- recall: 0.92: Модель чудово знаходить малі замовлення. Вона успішно ідентифікувала 92% усіх реальних малих замовлень, що робить її ефективною для "відсіювання" дрібних транзакцій.

4.5 Побудова програмного забезпечення

4.5.1 Вступ та вибір технологічного стеку

Для вирішення завдання аналізу реалізації насінневого матеріалу було розроблено спеціалізоване програмне забезпечення у вигляді інтерактивного веб-додатку. Цей підхід дозволяє не просто обробляти дані, але й надавати кінцевому користувачу (наприклад, аналітику чи менеджеру) зручний графічний інтерфейс для взаємодії з даними та моделями в реальному часі.

Основою програмного забезпечення є мова програмування Python 3.x та фреймворк Streamlit, який дозволяє швидко перетворювати скрипти аналізу даних на інтерактивні веб-сторінки.

Ключові бібліотеки, що склали технологічний стек системи:

- Streamlit: Ядро програмного забезпечення, що відповідає за побудову всього графічного інтерфейсу користувача (GUI), включаючи вкладки, кнопки, вивід графіків та таблиць.
- Pandas.
- Prophet (від Facebook): Використовується для реалізації модуля прогнозування часових рядів.
- Scikit-learn (sklearn): Фундаментальна бібліотека машинного навчання, що використовується для:
 - Кластеризації (KMeans).
 - Масштабування даних (StandardScaler) та зниження розмірності (PCA).
 - Побудови класифікатора Наївного Баєса (CategoricalNB).
 - Підготовки даних (OrdinalEncoder, LabelEncoder) та оцінки моделей (classification_report).
- Mlxtend: Використовується для реалізації модуля пошуку асоціативних правил (алгоритм Apriori).

- Matplotlib та Seaborn.

4.5.2 Архітектура програмного забезпечення

Архітектура системи побудована на базі Streamlit і є односторінковим веб-додатком. Для зручності користувача, інтерфейс розділено на п'ять логічних модулів, кожен з яких вирішує окрему аналітичну задачу:

1. Головний огляд (BI-звіти)
2. Прогнозування (Prophet)
3. Кластеризація (K-Means)
4. Асоціативні правила (Apriori)
5. Наївний Баєс (ML)

4.5.3 Опис функціональних модулів системи

1. Модуль "Головний огляд (BI-звіти)" (рис. 4.8) Цей модуль виконує базовий бізнес-аналіз (Business Intelligence). Він фільтрує дані за поточний рік (2024), агрегує попит (Total_Seed_By_District) за країною-виробником (Country_Name) та візуалізує результат у вигляді стовпчастої діаграми.

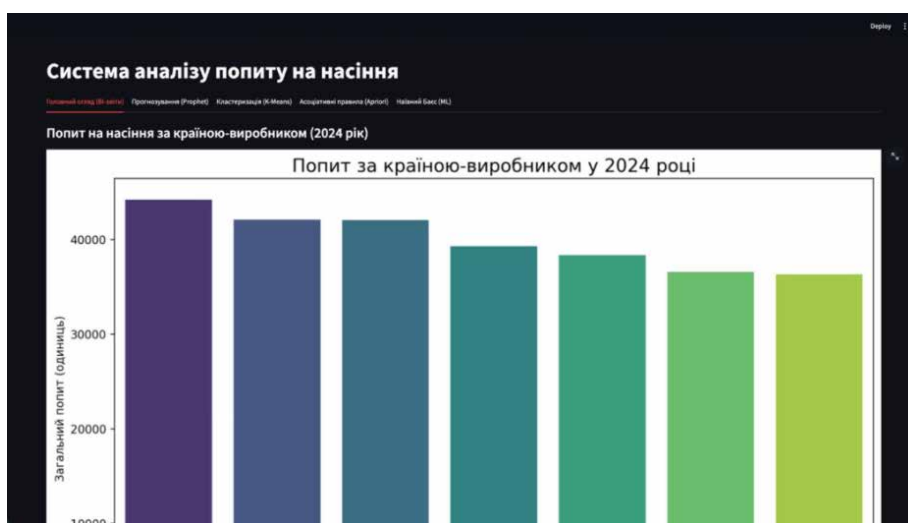


Рис.4.8 Головний огляд

2. Модуль "Прогнозування (Prophet)" (рис. 4.9) Реалізує функцію `run_prophet_analysis()`. Цей модуль вирішує задачу прогнозування попиту.

Навчання: Модель Prophet навчається на агрегованих даних за 2023-2024 роки.

Прогнозування: Модель буде прогнозувати на 365 днів (2025 рік).

Результат: У інтерфейс виводяться два графіки: загальний прогноз (`model_prophet.plot()`) та аналіз компонентів (тренди, сезонність) (`model_prophet.plot_components()`).

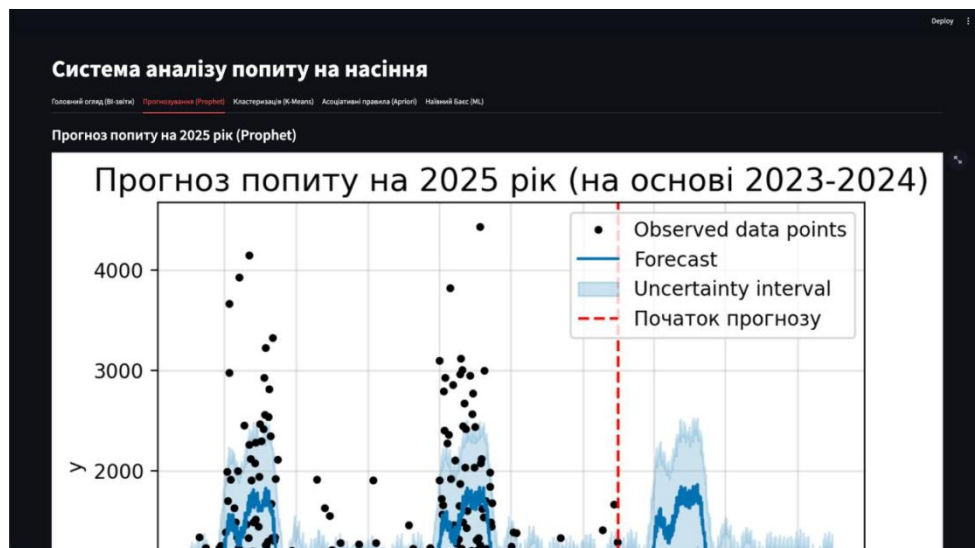


Рис.4.9 Вкладка Прогнозування

3. Модуль "Кластеризація (K-Means)" (рис. 4.10) Реалізує функцію `run_kmeans_analysis()` для сегментації ринку.

Підготовка: Дані зводяться у матрицю (`pivot_table`), де рядки — це District (райони), а стовпці - Category_Name (категорії).

Обробка: Дані масштабуються за допомогою `StandardScaler`, щоб різні обсяги попиту були порівнянними.

Кластеризація: Використовується алгоритм `KMeans` з фіксованою кількістю кластерів ($k = 4$) для групування схожих районів.

Візуалізація: Для відображення багатовимірних кластерів у 2D-просторі застосовується метод головних компонент (PCA). Результати (профілі кластерів та їх склад) виводяться в інтерфейс.

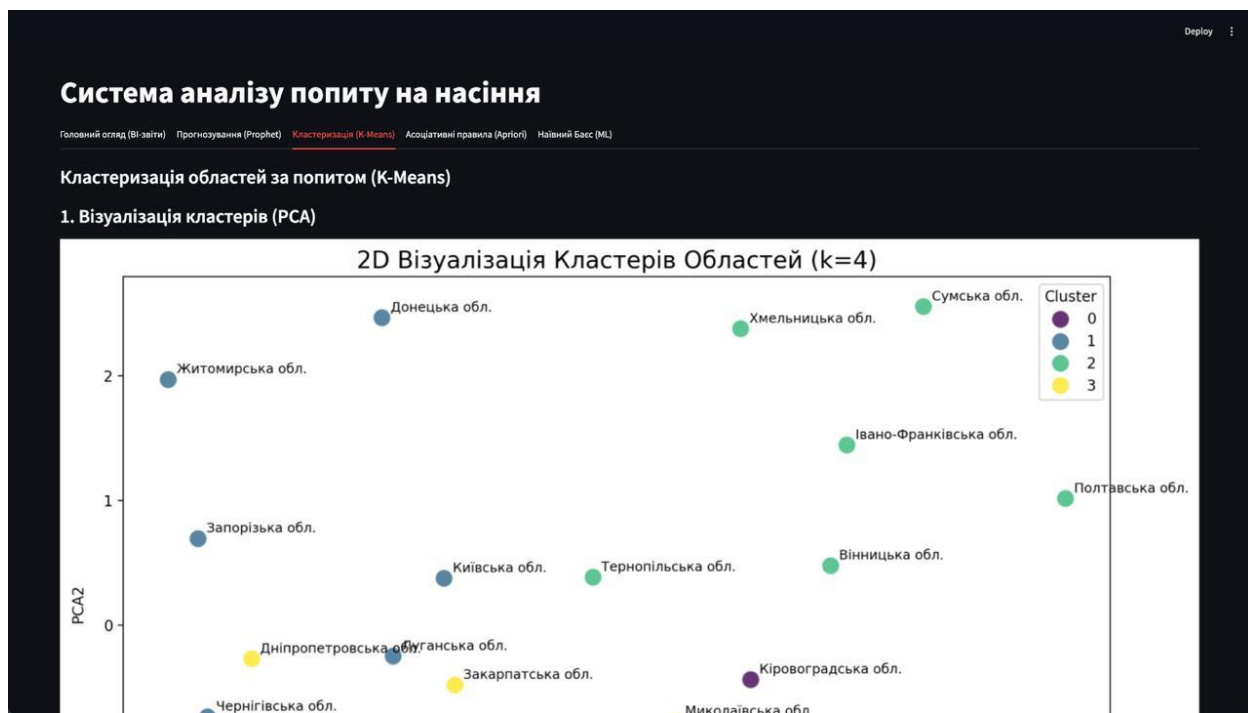


Рис.4.10 Вкладка Кластеризація

4. Модуль "Асоціативні правила (Apriori)" (рис.4.11) Реалізує функцію `run_apriori_analysis()` для аналізу ринкового кошика.

Підготовка: Дані (`Category_Name`, `Country_Name`, `District`) перетворюються на транзакційну матрицю (`one-hot encoding`) за допомогою `pd.get_dummies()`.

Аналіз: Застосовуються функції `apriori` та `association_rules` з `mlxtend` для пошуку правил з `lift > 1.1` та `confidence > 0.1`.

Результат: Знайдені правила з перекладеними назвами стовпців виводяться у `st.dataframe`.

Система аналізу попиту на насіння

Головний огляд (BI-звіт) Прогнозування (Prophet) Кластеризація (K-Means) **Асоціативні правила (Apriori)** Наївний Баєс (ML)

Аналіз ринкового кошика (Apriori) <=>

Знайдені асоціативні правила

	Якщо (Посилка)	То (Наслідок)	Lift (Невипадковість)	Confidence (Впевненість)	Support (Підтримка)
89	Country_Name_Іспанія	District_Запорізька обл.	3.008	0.1324	0.018
88	District_Запорізька обл.	Country_Name_Іспанія	3.008	0.4091	0.018
148	District_Львівська обл.	Country_Name_Україна	2.7006	0.3889	0.014
65	District_Одеська обл.	Category_Name_Технічні культури	2.6316	0.2632	0.01
48	Category_Name_Олійні культури	District_Житомирська обл.	2.5397	0.127	0.016
49	District_Житомирська обл.	Category_Name_Олійні культури	2.5397	0.32	0.016
29	District_Херсонська обл.	Category_Name_Квіти	2.5329	0.2381	0.01
28	Category_Name_Квіти	District_Херсонська обл.	2.5329	0.1064	0.01
97	District_Івано-Франківська обл.	Country_Name_Китай	2.3077	0.3	0.018
96	Country_Name_Китай	District_Івано-Франківська обл.	2.3077	0.1385	0.018
--	---	---	---	---	---

Рис.4.11 Вкладка Асоціативні правила

5. Модуль "Наївний Баєс (ML)" (рис. 4.11) Реалізує функцію `run_naive_bayes_analysis()` для побудови прогностичної моделі.

Постановка задачі: Модуль прогнозує не категорію, а обсяг замовлення.

Інжиніринг ознак: Обчислюється медіана обсягу (272.0), і всі замовлення діляться на два класи: `High_Volume` та `Low_Volume`.

Навчання: Модель `CategoricalNB` навчається передбачати цей клас на основі трьох ознак: `Category_Name`, `Country_Name` та `Month`.

Результат: В інтерфейс виводиться метрика `st.metric` із загальною точністю моделі (69.1%) та повний звіт про класифікацію (`classification_report`), що дозволяє оцінити її ефективність.

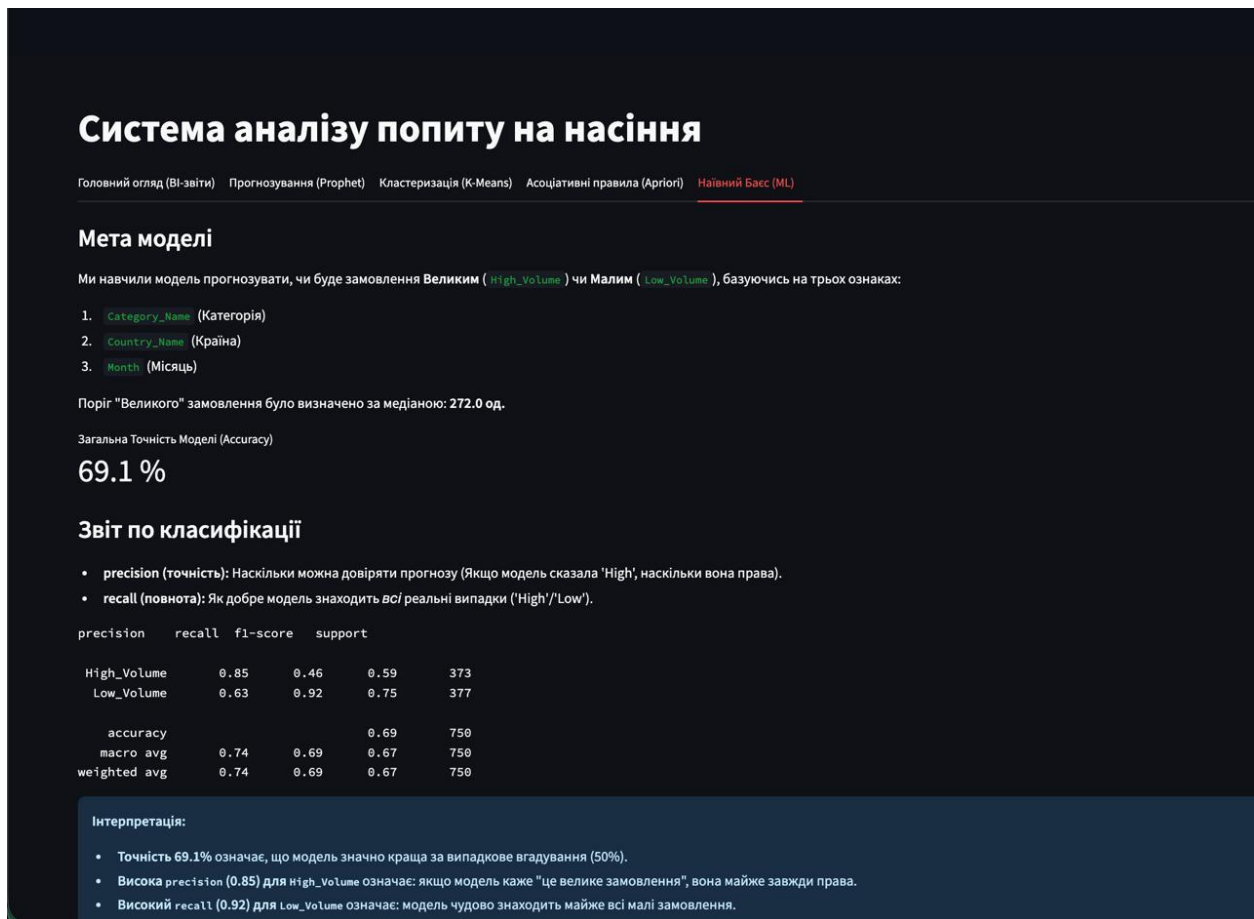


Рис.4.12 Вкладка Наївний Байєс

ВИСНОВКИ

У магістерській кваліфікаційній роботі було вирішено науково-прикладне завдання щодо розробки програмного забезпечення системи аналізу реалізації високоякісного насінневого матеріалу. У процесі виконання роботи було отримано наступні результати:

- Проаналізовано предметну область аналізу реалізації насінневого матеріалу та вивчено сучасні методи інтелектуального аналізу даних. Визначено ключові бізнес-задачі, що потребують вирішення, зокрема: прогнозування попиту в часі, сегментація ринку (районів), виявлення прихованих асоціативних зв'язків та класифікація обсягів замовлень.

- Визначено основні вимоги до програмного забезпечення, здатного надавати комплексний інструментарій для підтримки прийняття рішень. Ключовою вимогою стала розробка єдиного інтерактивного інтерфейсу, який би об'єднував різні аналітичні моделі.

- Розроблено архітектуру програмної у вигляді інтерактивного веб-додатку. Запропонована архітектура базується на фреймворку Streamlit та включає п'ять функціональних аналітичних модулів: модуль ВІ-звітів, модуль прогнозування (Prophet), модуль кластеризації (K-Means), модуль асоціативного аналізу (Apriori) та модуль класифікації (Naive Bayes).

- Реалізовано та протестовано програмний продукт, який довів свою ефективність у вирішенні поставлених завдань. Модуль Prophet успішно побудував прогноз попиту на 2025 рік, виявивши сезонні компоненти. Модуль K-Means провів валідну сегментацію ринку, згрупувавши райони у 4 кластери зі схожими профілями попиту. Модуль Apriori виявив сильні не випадкові зв'язки. Модуль Naive Bayes підтвердив можливість прогнозування обсягу замовлення з точністю 69.1%, що значно перевищує випадкове вгадування та доводить наявність закономірностей у даних.

Таким чином, розроблене програмне забезпечення є повноцінною системою підтримки прийняття рішень (DSS), яка перетворює необроблені дані на готові аналітичні висновки, прогнози та сегментацію ринку. Робота підтверджує, що поєднання сучасних бібліотек (Streamlit, Prophet, Scikit-learn, Mlxtend) дозволяє створити інструмент для аналізу реалізації насінневого матеріалу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. What is a Decision Support System (DSS)? – [Електронний ресурс]. – Режим доступу: <https://www.techtarget.com/searchbusinessanalytics/definition/decision-support-system-DSS>
2. Use Cases. – [Електронний ресурс]. – Режим доступу: <https://www.usability.gov/how-to-and-tools/methods/use-cases.html>
3. Overview of Online Analytical Processing (OLAP). – [Електронний ресурс]. – Режим доступу: <https://support.microsoft.com/en-us/office/overview-of-online-analytical-processing-olap-15d2cdde-f70b-4277-b009-ed732b75fdd6>
4. Fowler, M. (2003). UML Distilled: A Brief Guide to the Standard Object Modeling Language. – [Електронний ресурс]. – Режим доступу: https://ptgmedia.pearsoncmg.com/images/9780321193681/samplepages/9780321193681_Sample.pdf
5. Python 3.12.0 documentation. – [Електронний ресурс]. – Режим доступу: <https://docs.python.org/3/>
6. McKinney, W. (2017). Python for Data Analysis, 2nd Edition. – [Електронний ресурс]. – Режим доступу: <https://www.lkhibra.ma/books/Python-for-Data-Analysis.pdf>
7. pandas 2.1.3 documentation. – [Електронний ресурс]. – Режим доступу: <https://pandas.pydata.org/docs/>
8. NumPy v1.26 User Guide. – [Електронний ресурс]. – Режим доступу: <https://numpy.org/doc/stable/>
9. Harris, C. R., et al. (2020). Array programming with NumPy. – [Електронний ресурс]. – Режим доступу: <https://www.nature.com/articles/s41586-020-2649-2.pdf>

10. Streamlit Documentation: Main concepts. – [Електронний ресурс]. – Режим доступу: <https://docs.streamlit.io/library/get-started/main-concepts>
11. Prophet: Automatic Forecasting Procedure (Official Documentation). – [Електронний ресурс]. – Режим доступу: <https://facebook.github.io/prophet/>
12. Taylor, S. J., & Letham, B. (2018). Forecasting at scale. – [Електронний ресурс]. – Режим доступу: https://www.researchgate.net/publication/344989540_Forecasting_at_scale
13. Прогнозування часових рядів за допомогою Prophet і Python. – [Електронний ресурс]. – Режим доступу: <https://www.analyticsvidhya.com/blog/2023/04/time-series-forecasting-using-prophet-and-python/>
14. sklearn.cluster.KMeans User Guide. – [Електронний ресурс]. – Режим доступу: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
15. Кластеризація методом к–середніх – [Електронний ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/Кластеризація_методом_к–середніх
16. sklearn.decomposition.PCA User Guide. – [Електронний ресурс]. – Режим доступу: <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>
17. Метод головних компонент (PCA): інтуїтивне пояснення. – [Електронний ресурс]. – Режим доступу: <https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c>
18. mlxtend.frequent_patterns.apriori Documentation. – [Електронний ресурс]. – Режим доступу: http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/

19. mlxtend.frequent_patterns.association_rules Documentation. – [Електронний ресурс]. – Режим доступу: http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/
20. Алгоритм Apriori та асоціативні правила: пояснення. – [Електронний ресурс]. – Режим доступу: <https://www.datacamp.com/tutorial/association-rule-mining-via-apriori-algorithm-in-python>
21. sklearn.naive_bayes.CategoricalNB User Guide. – [Електронний ресурс]. – Режим доступу: https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.CategoricalNB.html
22. sklearn.preprocessing.OrdinalEncoder User Guide. – [Електронний ресурс]. – Режим доступу: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OrdinalEncoder.html>
23. sklearn.metrics.classification_report User Guide. – [Електронний ресурс]. – Режим доступу: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html
24. Matplotlib 3.8.2 documentation. – [Електронний ресурс]. – Режим доступу: <https://matplotlib.org/stable/index.html>
25. Seaborn 0.13.1 documentation. – [Електронний ресурс]. – Режим доступу: <https://seaborn.pydata.org/>
26. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. – [Електронний ресурс]. – Режим доступу: <https://www.scribd.com/document/714843620/J-D-Hunter-2007-Matplotlib-A-2D-Graphics-En>
27. Магазин насіння. – [Електронний ресурс]. – <https://semena.in.ua/>

28. Поляков, М. В., Ханін, І. Г., Шевченко, Г. Я., Білозубенко, В. С., & Марченко, О. А. (2025). Data Mining для дослідницьких цілей. Science and Science of Science, (2(128), 27–47.
29. Use Case. – [Електронний ресурс]. – Режим доступу: <https://searchsoftwarequality.techtarget.com/definition/use-case>
30. Що таке OLAP? – [Електронний ресурс]. – Режим доступу: <https://uk.education-wiki.com/5528785-what-is-olap>
31. What is MOLAP (Multidimensional OLAP) in Data Warehouse? – [Електронний ресурс]. – Режим доступу: <https://www.guru99.com/multidimensional-online-analytical-processing.html>
32. What Is a Data Warehouse? – [Електронний ресурс]. – Режим доступу: <https://www.oracle.com/database/what-is-a-data-warehouse/>
33. SSAS Tutorial: What is SSAS Cube, Architecture & Types . – [Електронний ресурс]. – Режим доступу: <https://www.guru99.com/ssas-tutorial.html>
34. Data Source Views in Multidimensional Models. – [Електронний ресурс]. – Режим доступу: <https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/data-source-views-in-multidimensional-models?view=asallproducts-allversions>