

Міністерство освіти і науки України
Державний університет «Житомирська політехніка»
Факультет інформаційно-комп'ютерних технологій
Кафедра інженерії програмного забезпечення

КУРСОВА РОБОТА

(ПОЯСНЮВАЛЬНА ЗАПИСКА)

з дисципліни: «Моделювання інформаційних процесів та системи обробки
великих даних»

на тему:

**«Моделі інформаційних процесів для системи прогнозу попиту на товари з
урахуванням відгуків і рейтингів покупців»**

Виконав магістр 1-го курсу, групи ІПЗм-24-1
спеціальності 121 «Інженерія програмного
забезпечення»

Холодніцький Владислав Ігорович

(прізвище, ім'я та по-батькові)

Керівник: к.т.н., доцент кафедри КН Сугоняк І.І.

Дата захисту: " ____ " _____ 2024 р.

Національна шкала _____

Кількість балів: _____

Оцінка: ECTS _____

Члени комісії

(підпис)

Інна СУГОНЯК

(прізвище та ініціали)

(підпис)

Марина ГРАФ

(прізвище та ініціали)

(підпис)

(прізвище та ініціали)

Житомир – 2024

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ 1. МОДЕЛІ ІНФОРМАЦІЙНИХ ПРОЦЕСІВ ДЛЯ СИСТЕМИ ПРОГНОЗУ ПОПИТУ НА ТОВАРИ З УРАХУВАННЯМ ВІДГУКІВ І РЕЙТИНГІВ ПОКУПЦІВ.....	4
1.1 Постановка задачі	4
1.2 Інформаційні процеси, обґрунтування вибору засобів моделювання та програмного середовища.....	5
1.3 Побудова моделей бізнес-процесів та потоків даних	6
Висновки до розділу 1	14
РОЗДІЛ 2. ОРГАНІЗАЦІЯ СХОВИЩА ДАНИХ ДЛЯ АНАЛІТИКИ ТА ETL ПРОЦЕСУ.....	15
2.1 Визначення даних для аналізу та розробка сховища даних	15
2.2 ETL процес та використання Azure Synapse Analytics для копіювання та трансформації даних.....	21
Висновки до розділу 2	29
РОЗДІЛ 3. ОРГАНІЗАЦІЯ ЗАСОБІВ ВІЗУАЛІЗАЦІЇ ТА АНАЛІЗУ ДАНИХ.....	30
3.1 Технологія Power BI та її основні можливості	30
3.2 Публікація звіту на прикладній дошці Power BI.....	31
Висновки до розділу 3	36
ВИСНОВКИ	37
ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	38

					«Житомирська політехніка» 24.121.33.000			
Змн.	Арк.	№ докум.	Підпис	Дата				
Розроб.		Холодницький В. І.			Пояснювальна записка до курсової роботи		Літ.	Арк.
Керівник		Сугоняк І.І.						Аркушів
								2
								47
Реценз.					ФІКТ, гр. ПЗМ-24-1			
Затверд.		Граф М. С.						

ВСТУП

У сучасному світі для забезпечення ефективного функціонування бізнес-процесів необхідно обробляти великий обсяг інформації. У сфері торгівлі ключовим завданням є прогнозування попиту на товари, адже вірний прогноз дозволить оптимізувати запаси, забезпечити задоволення потреб усіх споживачів та мінімізувати витрати. Враховуючи стрімкий розвиток електронної комерції та поширення онлайн-платформ, відгуки та рейтинги покупців стали важливим джерелом інформації, яка впливає на прийняття рішень як споживачами, так і продавцями.

Актуальність теми полягає в необхідності розробки моделей, що зможуть ефективно використовувати рейтингові дані та відгуки для прогнозування попиту на визначені товари. Даний метод однозначно покращить управління запасами, зміцнить позиції компанії на ринку та підвищить задоволення споживачів. Враховуючи усі зазначені фактори в процесі прогнозування, можна вірно передбачити зміни попиту, що дасть змогу оптимізувати маркетингові та продажні стратегії. Це є важливою складовою для підтримки конкурентоспроможності компанії в динамічному ринковому середовищі.

Метою курсової роботи є дослідження та розробка моделей інформаційних процесів для прогнозування попиту на товари, що враховують відгуки та рейтинги покупців.

Об'єктом дослідження є інформаційні процеси в системах прогнозування попиту.

Предметом дослідження є моделі інформаційних процесів, що враховують відгуки та рейтинги споживачів для покращення прогнозів попиту на товари.

		Холодніцький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		3

РОЗДІЛ 1. МОДЕЛІ ІНФОРМАЦІЙНИХ ПРОЦЕСІВ ДЛЯ СИСТЕМИ ПРОГНОЗУ ПОПИТУ НА ТОВАРИ З УРАХУВАННЯМ ВІДГУКІВ І РЕЙТИНГІВ ПОКУПЦІВ

1.1 Постановка задачі

Основною метою моделювання інформаційних процесів у системі прогнозування попиту на товари є створення діючого інструменту для передбачення можливого попиту з урахуванням відгуків та рейтингів споживачів. Ця система дозволить не лише покращити точність прогнозів, а й забезпечить кращу взаємодію з клієнтами завдяки аналізу їхніх думок та оцінок щодо певних товарів.

Для досягнення мети курсової роботи необхідно вирішити низку завдань, спрямованих на створення ефективної моделі, яка буде використовувати різні алгоритми машинного навчання та статистичні моделі для прогнозування попиту на основі аналізу відгуків та рейтингів покупців.

Основні етапи виконання роботи:

1. Здійснення моделювання інформаційних процесів;
 - 1.1. Розробка IDEF0 діаграми для відображення функцій та процесів системи;
 - 1.2. Проектування BPMN діаграми для ілюстрації основних етапів та взаємодій у процесі прогнозування попиту;
2. Створення сховища даних для аналітичних потреб і підтримки ETL процесів;
 - 2.1. Проектування схем баз даних OLAP та OLTP;
 - 2.2. Генерація даних для заповнення таблиць;
 - 2.3. Завантаження даних в Azure Synapse Analytics;
 - 2.4. Розробка аналітичних запитів до сховища даних;
3. Проведення аналізу та створення візуалізацій за допомогою Power BI;

		Холодніцький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				4
Змн.	Арк.	№ докум.	Підпис	Дата		

1.2 Інформаційні процеси, обґрунтування вибору засобів моделювання та програмного середовища

У контексті системи прогнозу попиту на товари з урахуванням відгуків та рейтингів покупців важливим аспектом є правильно підібрані методи моделювання інформаційних процесів і програмні засоби, що забезпечать точний опис усіх функцій системи, їх взаємодії між собою, потоків даних та бізнес-процесів. Це сприятиме аналізу, вдосконаленню та впровадженню ефективної системи підтримки прийняття рішень.

Для моделювання інформаційних процесів було використано нотації IDEF0 та BPMN.

IDEF0 - є методом функціонального моделювання, який дозволяє описати основні функції системи, їх взаємозв'язки, ресурси, управління та виходи. Його основною перевагою є ієрархічна структура, що полегшує розуміння складних систем і дозволяє відобразити ключові аспекти взаємодії інформаційних потоків [1]. Цей підхід допоможе ідентифікувати логічні зв'язки між компонентами системи прогнозування попиту.

BPMN - використовується для моделювання бізнес-процесів у зрозумілому та уніфікованому вигляді. Ця нотація дозволяє створювати моделі, які легко інтерпретуються як технічними фахівцями, так і кінцевими користувачами. BPMN є особливо корисною для опису послідовності дій, взаємодії між підсистемами та прийняття рішень на основі аналітичних даних [2].

Для реалізації сховищ баз даних будуть використані такі архітектурні підходи як: OLTP та OLAP.

OLTP - використовується для оперативної обробки транзакцій, таких як оновлення баз даних про замовлення, залишки на складах чи зміну клієнтських профілів. Ця архітектура забезпечує швидкість і надійність обробки даних у реальному часі [3].

OLAP - застосовується для аналітичної обробки даних, що дозволяє виявляти тенденції та здійснювати багатовимірний аналіз. Цей підхід ефективний для формування звітів і візуалізації прогнозів попиту [4].

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				5
Змн.	Арк.	№ докум.	Підпис	Дата		

Для зберігання та аналізу великих обсягів даних використано Azure Synapse Analytics, яка інтегрує потужні засоби обробки і аналітики в хмарному середовищі. Ця платформа забезпечує інтеграцію з різними джерелами даних, гнучке управління ними та високу продуктивність при виконанні складних запитів [5].

Для візуалізації та створення звітів застосовано Power BI, яка дозволяє інтерактивно аналізувати дані, створювати динамічні дашборди та отримувати корисні аналітичні висновки. Інтеграція Power BI із хмарними сервісами та базами даних значно полегшує створення комплексної картини прогнозів попиту, враховуючи зворотний зв'язок та оцінки споживачів [6].

Поєднання зазначених методів і інструментів забезпечує всебічне моделювання, аналіз та впровадження ефективної системи прогнозування попиту на товари [7].

1.3 Побудова моделей бізнес-процесів та потоків даних

В середовищі Ramus було створено IDEF0 діаграму (рис. 1.1), яка відображає всі можливі рівні декомпозицій для процесу прогнозування попиту на товари. На контекстному рівні наведено загальний опис системи, а на наступних рівнях деталізовано основні функції та їх взаємозв'язки. Діаграма чітко ілюструє вхідні та вихідні дані, механізми, ресурси й обмеження, що забезпечують виконання процесів. Основними елементами діаграми є:

- Вхідні дані: історичні дані продажів, відгуки покупців, рейтинги покупців, зовнішні фактори.
- Вихідні дані: прогноз попиту на товари, звіти, графіки, рекомендації.
- Механізми: Алгоритми машинного навчання, інформаційна система з аналізу даних.
- Обмеження: точність даних, обмеження бюджету на аналіз.

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		6

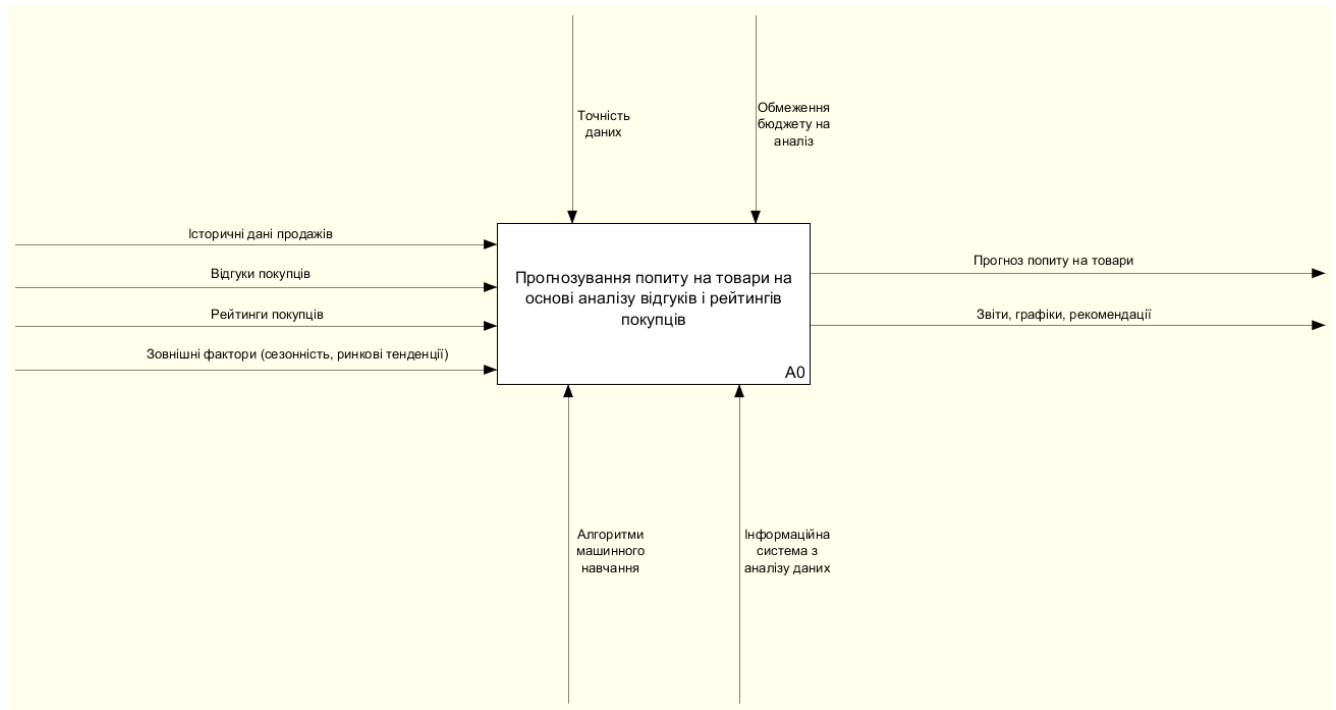


Рисунок 1.1 – Контекстна діаграма IDEF0

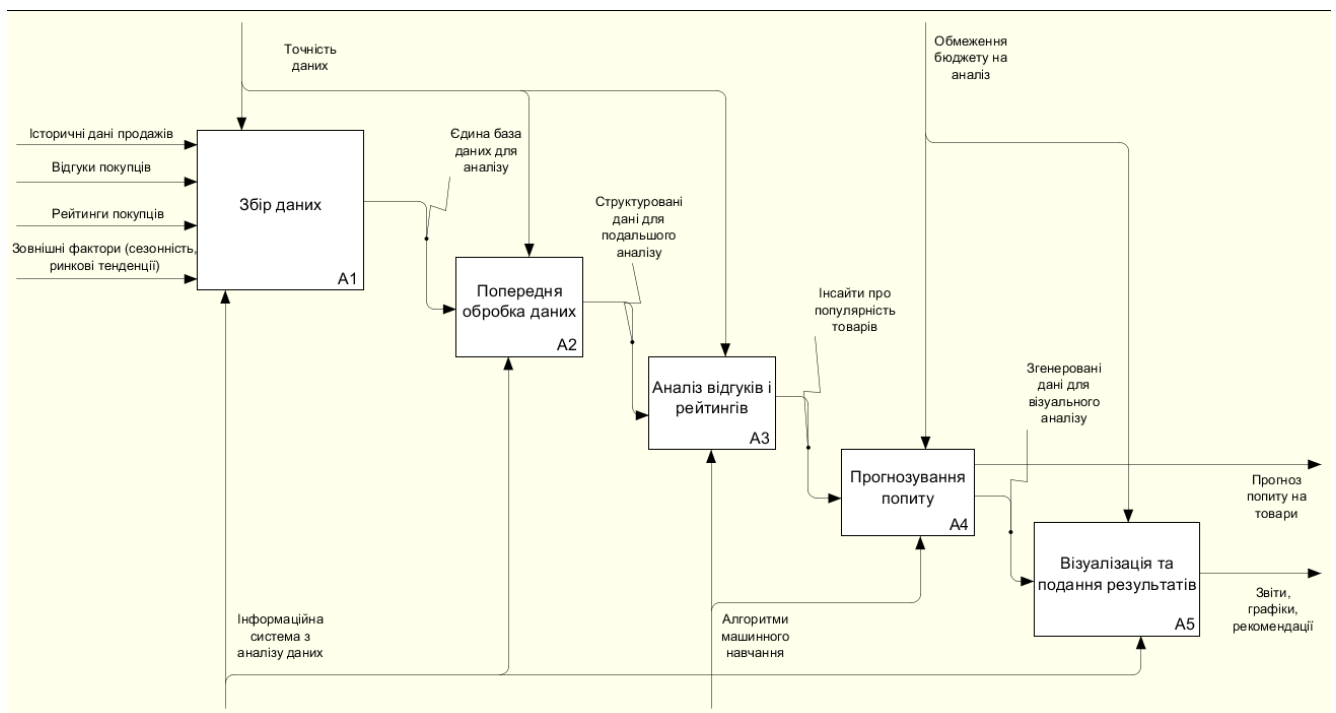


Рисунок 1.2 – Декомпозиція контекстної діаграми

Декомпозиція контекстної діаграми (рис. 1.2) складається з п'яти процесів.

Розглянемо детально кожен процес:

- Збір даних:
 - Вхідні дані: історичні дані продажів, відгуки покупців, рейтинги покупців, зовнішні фактори.
 - Вихідні дані: єдина база даних для аналізу.

- Механізми: інформаційна система з аналізу даних.
- Обмеження: точність даних.
- Попередня обробка даних:
 - Вхідні дані: єдина база даних для аналізу.
 - Вихідні дані: структуровані дані для подальшого аналізу.
 - Механізми: інформаційна система з аналізу даних.
 - Обмеження: точність даних.
- Аналіз відгуків і рейтингів:
 - Вхідні дані: структуровані дані для подальшого аналізу.
 - Вихідні дані: інсайти про популярність товарів.
 - Механізми: алгоритми машинного навчання.
 - Обмеження: точність даних.
- Прогнозування попиту:
 - Вхідні дані: інсайти про популярність товарів.
 - Вихідні дані: прогноз попиту на товари, згенеровані дані для візуального аналізу.
 - Механізми: алгоритми машинного навчання.
 - Обмеження: обмеження бюджету на аналіз.
- Візуалізація та подання результатів:
 - Вхідні дані: згенеровані дані для візуального аналізу.
 - Вихідні дані: звіти, графіки, рекомендації.
 - Механізми: інформаційна система з аналізу даних.
 - Обмеження: обмеження бюджету на аналіз.

Далі було проведено декомпозицію кожного процесу (рис. 1.3 – 1.7):

Збір даних охоплює історичну інформацію про продажі, відгуки та рейтинги покупців, які інтегруються в єдину базу даних. Процес також враховує зовнішні фактори, такі як сезонність і ринкові тенденції, для створення основи точного аналізу (рис. 1.3).

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				8
Змн.	Арк.	№ докум.	Підпис	Дата		

Попередня обробка даних включає очищення бази від некоректної інформації, заповнення пропусків і структурування для підготовки до подальшого аналізу (рис. 1.4).

Аналіз відгуків і рейтингів включає оцінку тональності текстів відгуків, виділення ключових факторів із текстів та розрахунок популярності товарів на основі рейтингів для отримання інсайтів (рис. 1.5).

Прогнозування попиту включає аналіз вхідних даних для отримання інсайтів про популярність товарів, побудову моделі прогнозування з використанням підготовленого набору факторів та алгоритмів машинного навчання, а також оцінку точності прогнозу й його оптимізацію для створення точного прогнозу попиту та візуального аналізу (рис. 1.6).

Візуалізація та подання результатів включає форматування даних для забезпечення їхньої готовності до побудови графіків і таблиць, створення структурованих візуалізацій у вигляді графіків і таблиць, а також генерування звітів із рекомендаціями на основі візуалізованих даних для підтримки прийняття рішень (рис. 1.7).

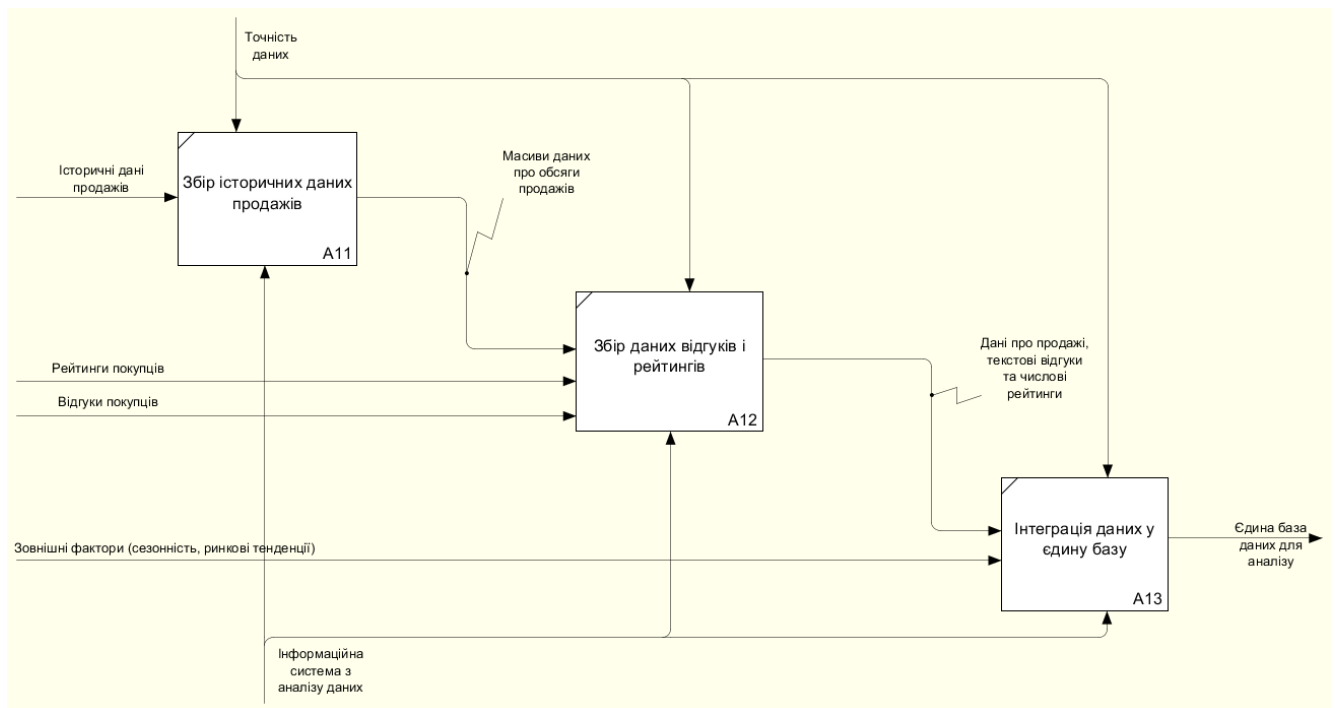


Рисунок 1.3 – Декомпозиція збору даних

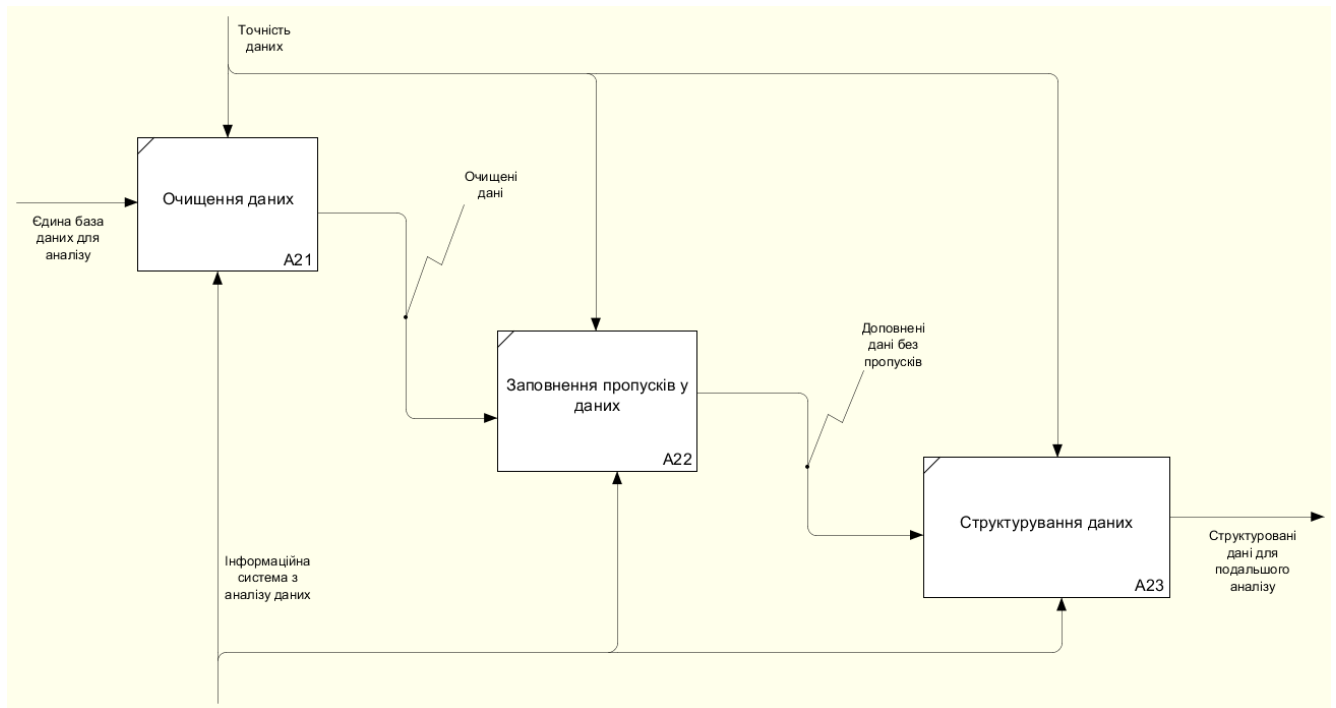


Рисунок 1.4 – Декомпозиція попередньої обробки даних

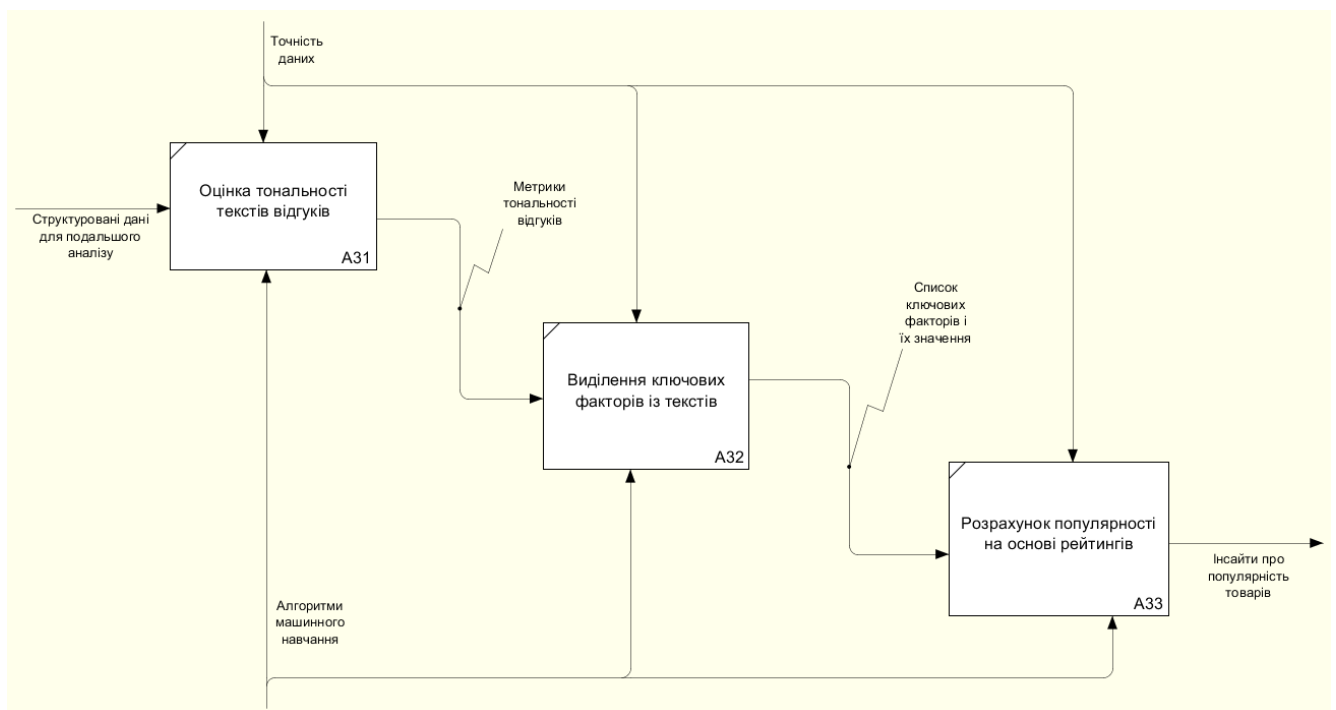


Рисунок 1.5 – Декомпозиція аналізу відгуків та рейтингів

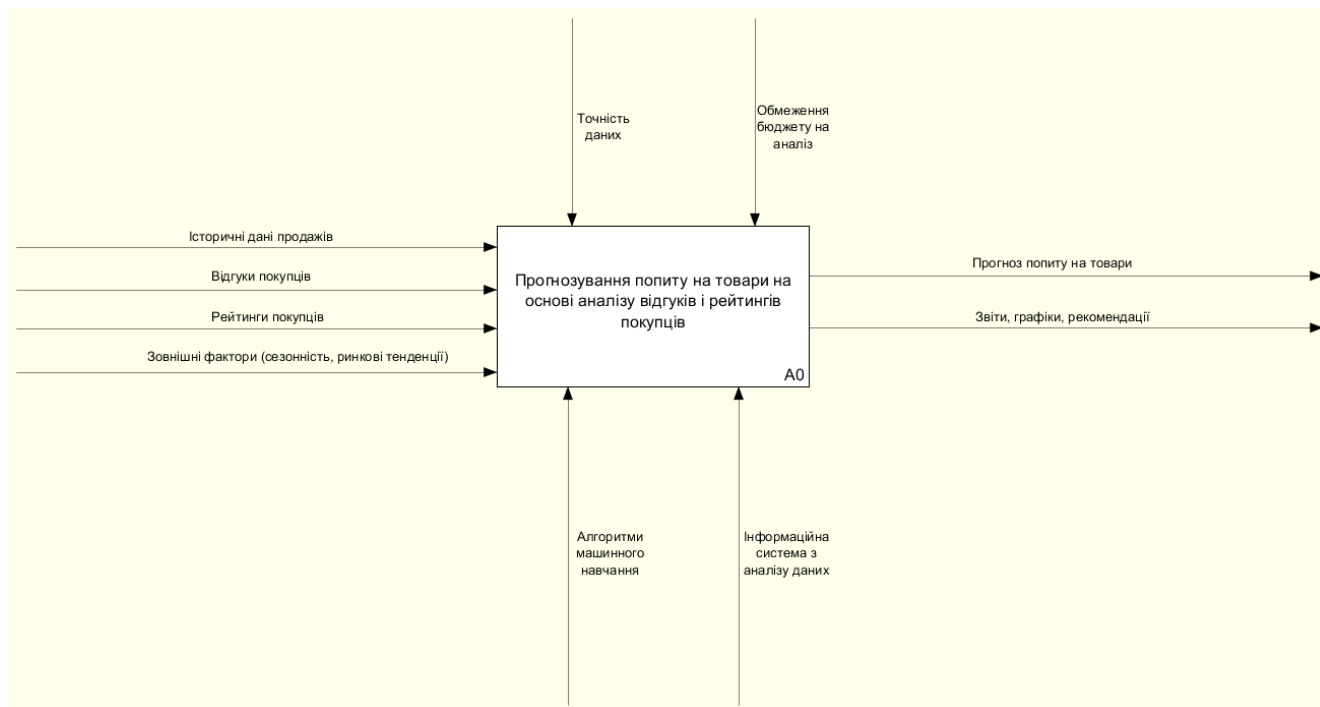


Рисунок 1.6 – Декомпозиція прогнозування попиту

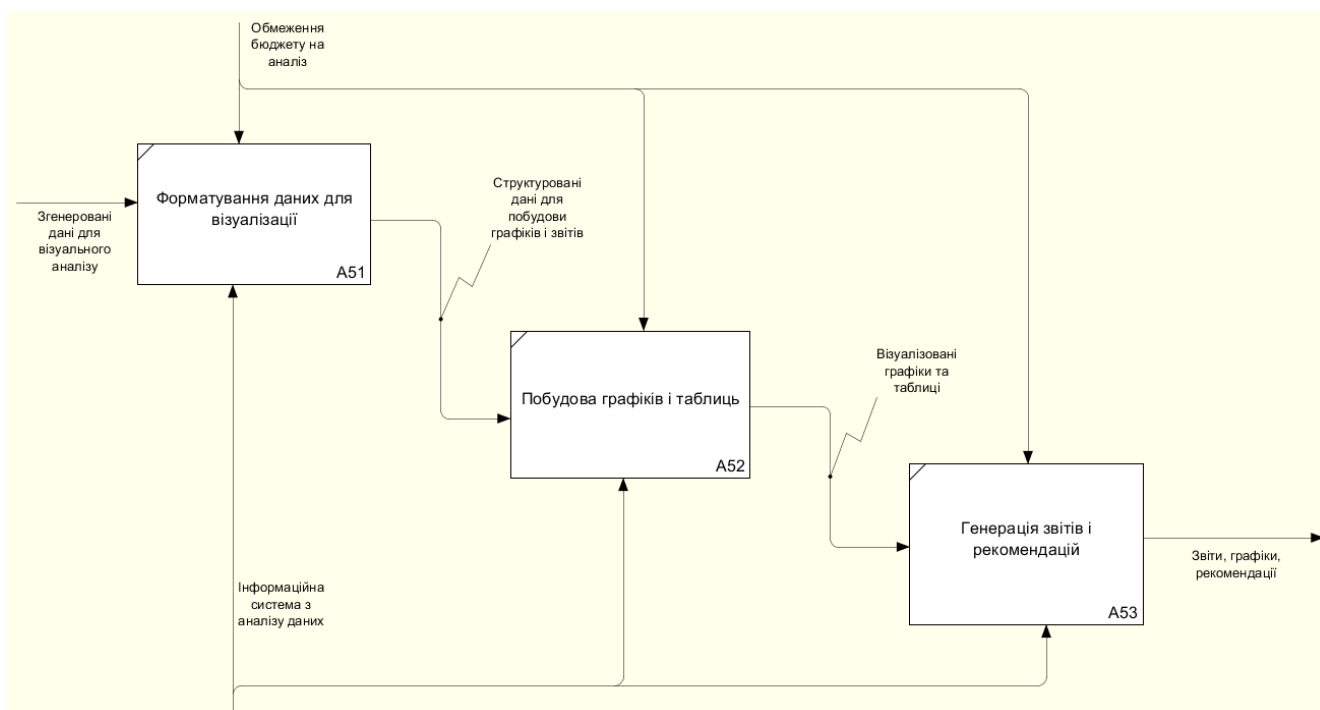


Рисунок 1.7 – Декомпозиція візуалізації та подання результатів

Також було створено BPMN модель бізнес процесів системи прогнозування попиту на товари (рис. 1.8) за допомогою онлайн сервісу bpmn.io.

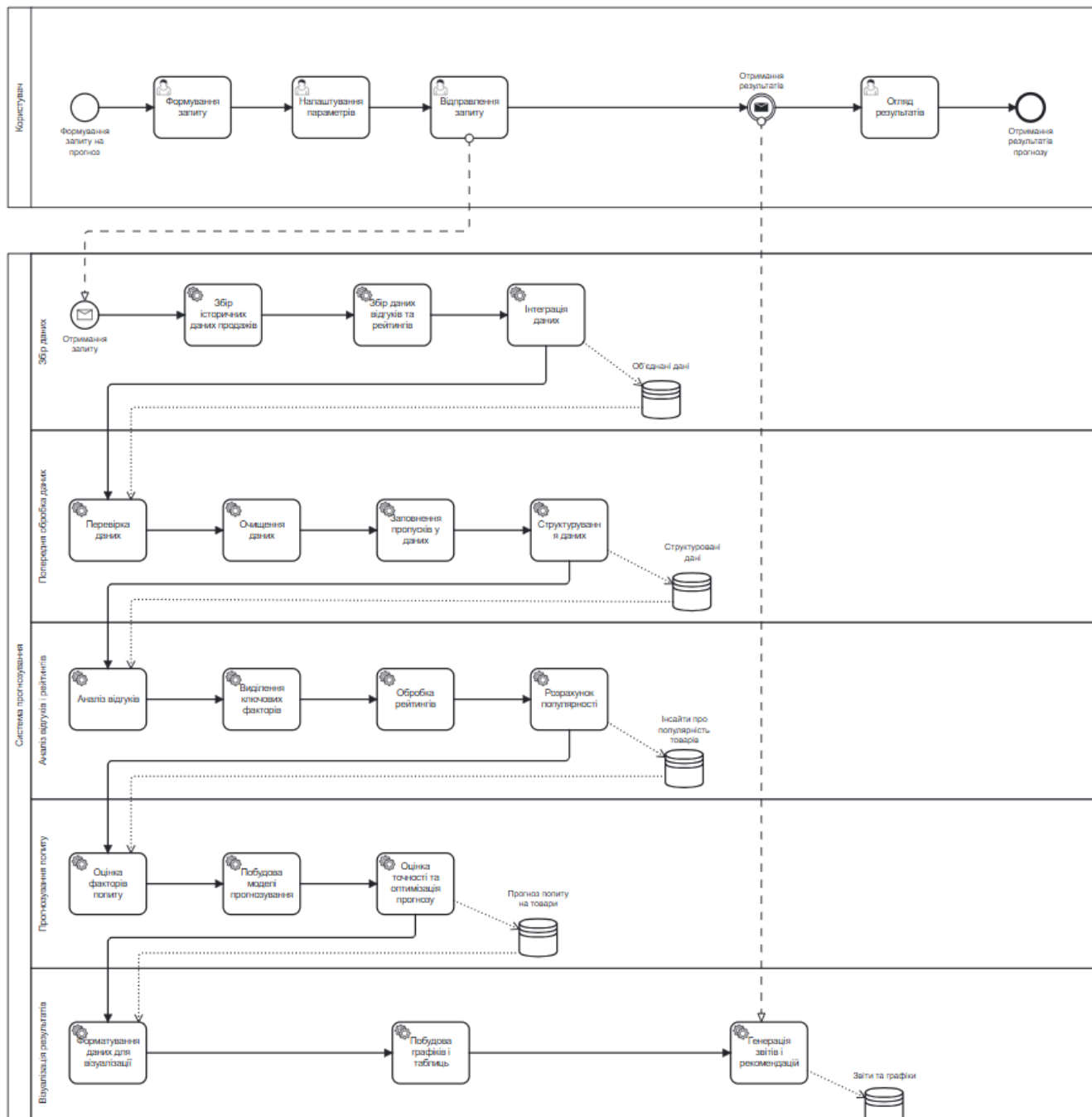


Рисунок 1.8 – BPMN діаграма

BPMN модель починається з події "Формування запити", яка ініціює початок процесу. На першому етапі уточнюються параметри запити, проводиться його формування, після чого готовий запит надсилається для обробки. Після виконання всіх необхідних операцій результат запити відправляється клієнту, завершуючи цей етап процесу.

Далі процес переходить до обробки даних. Спершу проводиться перевірка якості отриманих даних, що дозволяє виявити будь-які можливі помилки або невідповідності. Після перевірки здійснюється очікування або завантаження

додаткових даних із зовнішніх джерел. Отримані дані інтегруються, проходять етап структуризації і зберігаються у вигляді готових до аналізу структурованих даних.

На етапі аналітики відбувається аналіз бізнес-показників, визначення ключових факторів, що впливають на результати, і розрахунок релевантних рейтингів. Ці дії спрямовані на підготовку звіту, який містить детальну інформацію про попередні результати аналізу.

Після цього модель переходить до прогнозування попиту. На цьому рівні здійснюється оцінка чинників попиту, аналіз точності попередніх прогнозів і генерація оновлених моделей прогнозування. Результати цього етапу готуються для подальшого використання у візуалізації даних.

Фінальний етап передбачає візуалізацію даних, яка включає створення таблиць, графіків і рекомендацій. Ці елементи дозволяють наочно представити отримані результати та сприяють прийняттю обґрунтованих рішень. Модель демонструє послідовність і взаємозв'язок процесів, які забезпечують ефективну обробку, аналіз і презентацію даних.

		Холодніцький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		13

Висновки до розділу 1

Було проведено дослідження інформаційних процесів, які формують основу роботи досліджуваної системи. Для цього застосовано методи моделювання за допомогою діаграм IDEF0 та BPMN. Розроблені моделі забезпечили структуроване уявлення функцій системи, потоків даних і взаємодій між її основними елементами, що дозволило деталізувати всі ключові етапи процесу.

Моделювання інформаційних процесів дало змогу описати основні функції системи, включаючи визначення вхідних даних, керуючих факторів і механізмів виконання операцій. Завдяки цьому вдалося уточнити послідовність і взаємозалежність етапів, що включають обробку даних, виконання операцій і отримання кінцевого результату. Також було виявлено вузькі місця в процесах, які потенційно можна оптимізувати для покращення загальної продуктивності.

Особливу увагу було приділено деталізації залежностей між процесами, які забезпечують логічну зв'язаність і ефективність системи. Використання діаграм BPMN дозволило більш наочно представити динаміку роботи системи, тоді як IDEF0 сприяв кращому розумінню функціональної структури та ролі кожного з елементів у досягненні загальних цілей.

Отримані результати створюють необхідну основу для подальших досліджень і вдосконалення процесів у системі. Використання побудованих моделей дає змогу приймати обґрунтовані рішення щодо оптимізації ресурсів, підвищення ефективності виконання операцій та покращення загальної продуктивності роботи системи.

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				14
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 2. ОРГАНІЗАЦІЯ СХОВИЩА ДАНИХ ДЛЯ АНАЛІТИКИ ТА ETL ПРОЦЕСУ

2.1 Визначення даних для аналізу та розробка сховища даних

Перед створенням аналітичного сховища даних, яке буде використовуватись для системи прогнозування попиту на товари, важливо спочатку розробити транзакційне сховище. Воно виконує роль основи, яка забезпечує зберігання і обробку первинних даних, необхідних для подальшого аналізу. Це сховище дозволить акумулювати інформацію про відгуки, рейтинги покупців і характеристики товарів, що є критично важливими для точної оцінки тенденцій та потреб споживачів.

Проектуючи OLTP схему, варто приділити особливу увагу таким аспектам, як раціональна структура, яка сприятиме мінімізації дублювання даних і підвищенню ефективності роботи системи. Не менш важливо забезпечити ізоляцію транзакцій, що є ключовим фактором для підтримання цілісності та достовірності інформації в умовах паралельної обробки даних. Висока продуктивність системи, яка забезпечить швидке виконання запитів, є ще одним важливим елементом, який дозволить системі оперативно реагувати на зміни в даних.

Транзакційне сховище даних включатиме в свою структуру такі таблиці.

Таблиця 2.1

Структура таблиці “Products”

Поле	Ключ	Тип	Призначення
id	PK	int	Унікальний ідентифікатор продукту
name	-	varchar(255)	Назва продукту
price	-	decimal(10, 2)	Ціна продукту
stock	-	int	Кількість товару на складі
description	-	varchar(max)	Детальний опис продукту
image_url	-	varchar(255)	Посилання на зображення продукту
category_id	FK	int	Зв'язок із категорією продукту

Таблиця 2.2

Структура таблиці “Customers”

Поле	Ключ	Тип	Призначення
id	PK	int	Унікальний ідентифікатор клієнта
name	-	varchar(255)	Ім'я клієнта
email	-	varchar(255)	Електронна пошта клієнта
phone	-	varchar(20)	Телефон клієнта
loyalty_points	-	int	Бонусні бали клієнта
registration_date	-	datetime	Дата реєстрації клієнта
address_id	FK	int	Зв'язок із адресою клієнта

Таблиця 2.3

Структура таблиці “Product_categories”

Поле	Ключ	Тип	Призначення
id	PK	int	Унікальний ідентифікатор категорії
name	-	varchar(255)	Назва категорії

Таблиця 2.4

Структура таблиці “Addresses”

Поле	Ключ	Тип	Призначення
id	PK	int	Унікальний ідентифікатор адреси
country	-	varchar(100)	Країна
city	-	varchar(100)	Місто
street	-	varchar(255)	Вулиця
zip_code	-	varchar(20)	Поштовий індекс
region	-	varchar(100)	Регіон або область

Таблиця 2.5

Структура таблиці “Reviews”

Поле	Ключ	Тип	Призначення
id	PK	int	Унікальний ідентифікатор відгуку
rating	-	tinyint	Рейтинг (1–5)
comment	-	varchar(max)	Текст відгуку
date	-	datetime	Дата залишення відгуку
sentiment	-	varchar(20)	Настрій відгуку (позитивний, нейтральний, негативний)
product_id	FK	int	Зв’язок із продуктом
customer_id	FK	int	Зв’язок із клієнтом

Таблиця 2.6

Структура таблиці “Transactions”

Поле	Ключ	Тип	Призначення
id	PK	int	Унікальний ідентифікатор транзакції
transaction_date	-	datetime	Дата транзакції
total_amount	-	decimal(10, 2)	Загальна сума транзакції
payment_method	-	varchar(50)	Спосіб оплати
customer_id	FK	int	Зв’язок із клієнтом

Таблиця 2.7

Структура таблиці “Promotions”

Поле	Ключ	Тип	Призначення
id	PK	int	Унікальний ідентифікатор акції
name	-	varchar(255)	Назва акції
start_date	-	datetime	Початок акції
end_date	-	datetime	Кінець акції
discount_percentage	-	decimal(5, 2)	Відсоток знижки

Таблиця 2.8

Структура таблиці “Promotion_products”

Поле	Ключ	Тип	Призначення
promotion_id	FK	int	Зв'язок із акцією
product_id	FK	int	Зв'язок із продуктом

Таблиця 2.9

Структура таблиці “Transaction_details”

Поле	Ключ	Тип	Призначення
id	PK	int	Унікальний ідентифікатор деталізації транзакції
quantity	-	int	Кількість проданих одиниць товару
price_per_unit	-	decimal(10, 2)	Ціна за одиницю
discount	-	decimal(5, 2)	Знижка на товар у цій транзакції
transaction_id	FK	int	Зв'язок із транзакцією
product_id	FK	int	Зв'язок із продуктом

За допомогою наведених транзакційних таблиць сховищ даних можна розробити OLTP діаграму, яка відображатиме зв'язки між первинними та зовнішніми ключами та забезпечить інтеграцію даних.



Рисунок 2.1 - OLTP діаграма

Основаючись на транзакційних таблицях та OLTP-діаграмі, можна розробити OLAP-діаграму для системи прогнозування попиту на товари з урахуванням відгуків і рейтингів покупців. Така діаграма дозволить здійснювати багатовимірний аналіз даних, агрегувати інформацію для точнішого прогнозування попиту та формувати звіти для прийняття рішень щодо асортименту і цінової політики.

Для побудови використовувалася зіркова схема, яка є однією з найбільш ефективних моделей організації даних у сховищах. У цій моделі використано дві таблиці фактів: перша містить історичні дані про продажі, а друга – результати прогнозування попиту. Обидві таблиці пов'язані з кількома таблицями вимірів, що включають інформацію про товари, відгуки, рейтинги покупців та інші аспекти. Такий підхід забезпечує гнучкість моделі, легкість у структурі та дозволяє швидко виконувати як аналітичні запити, так і прогнозування.

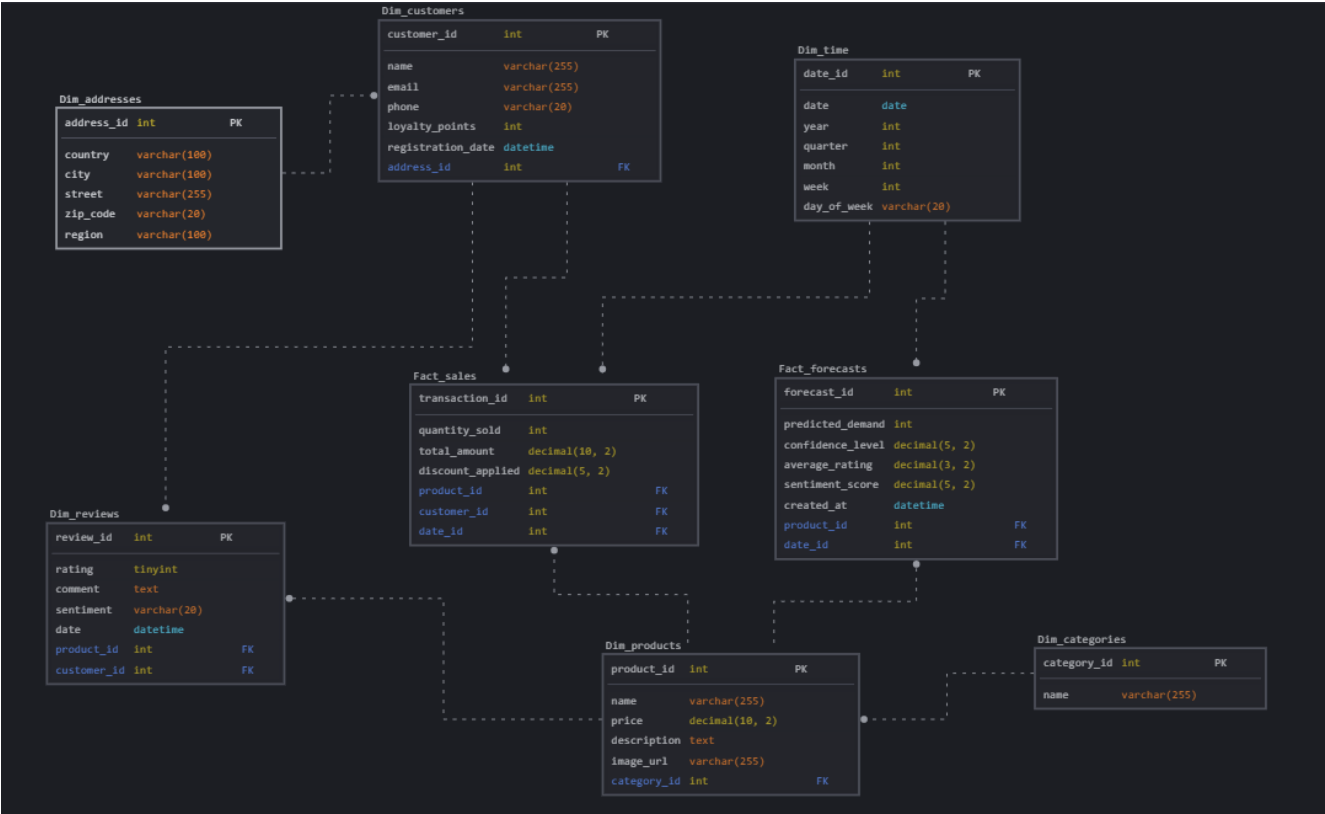


Рисунок 2.2 - OLAP діаграма

Оскільки реальне транзакційне сховище (OLTP) відсутнє, для всіх таблиць в моделі OLAP будуть випадково згенеровані дані. Це дозволить провести тестування та аналіз моделі без необхідності мати доступ до реальних даних. Генерація даних буде включати як історичні транзакційні записи, так і прогнозовані дані для попиту, що дозволить перевірити ефективність та точність побудованої системи прогнозування. Для кожної таблиці будуть створені дані відповідно до їхньої структури, що дозволить здійснити необхідні запити та перевірити працездатність зіркової схеми на основі випадкових значень.

	dim_reviews.csv	25.12.2024 12:33	Excel.CSV	36 KБ
	fact_forecasts.csv	25.12.2024 12:27	Excel.CSV	25 KБ
	fact_sales.csv	25.12.2024 12:27	Excel.CSV	15 KБ
	dim_time.csv	25.12.2024 12:27	Excel.CSV	18 KБ
	dim_customers.csv	25.12.2024 12:27	Excel.CSV	41 KБ
	dim_addresses.csv	25.12.2024 12:27	Excel.CSV	23 KБ
	dim_products.csv	25.12.2024 12:27	Excel.CSV	42 KБ
	dim_categories.csv	25.12.2024 12:27	Excel.CSV	1 KБ

Рисунок 2.3 - Згенеровані csv файли

2.2 ETL процес та використання Azure Synapse Analytics для копіювання та трансформації даних

Azure Synapse Analytics є основним інструментом для організації та оптимізації процесів обробки великих обсягів даних у цій роботі. Ця платформа дозволяє ефективно виконувати завдання з копіювання, трансформації та зберігання даних, забезпечуючи масштабоване середовище для їх обробки. Використання механізму ETL (Extract, Transform, Load) дозволяє інтегрувати різноманітні джерела даних, здійснювати їх необхідну обробку та підготовку до подальшого аналізу, що є критично важливим для прийняття обґрунтованих рішень на основі цих даних.

На першому етапі реалізації проекту необхідно створити надійне сховище для файлів, які містять дані, що підлягають обробці. Для цього буде використано сервіс Storage accounts від Azure, що дозволяє зберігати великі обсяги інформації в хмарному середовищі з високою доступністю та масштабованістю (рис. 2.4). Цей підхід забезпечує ефективне керування даними та їх безпечне зберігання, що є основою для подальшої роботи з ними в рамках платформи Azure Synapse Analytics.

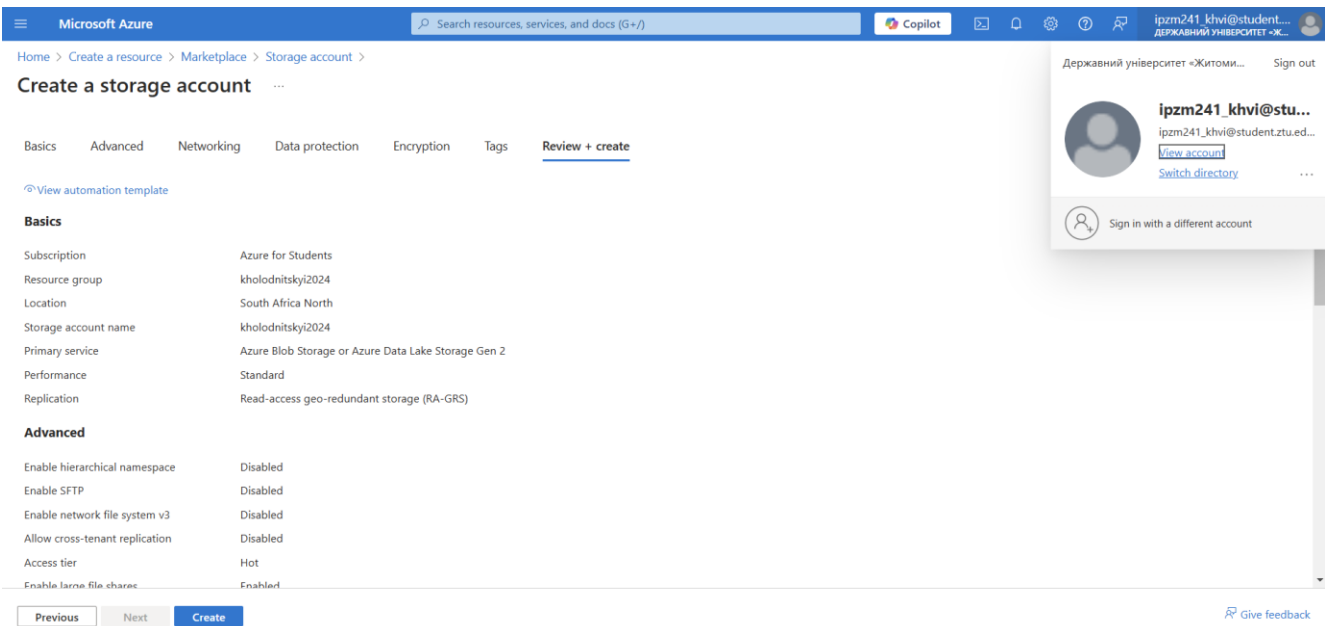


Рисунок 2.4 - Створення сховища файлів

Наступним етапом створимо новий робочий простір у сервісі Azure Synapse Analytics і налаштуємо його інтеграцію зі сховищем файлів для забезпечення зручного доступу та обробки даних (рис. 2.5).

		Холодніцький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		21

Microsoft Azure | Search resources, services, and docs (G+)

Home > Create a resource > Marketplace > Azure Synapse Analytics >

Create Synapse workspace

* Basics * Security Networking Tags Review + create

Create a Synapse workspace to develop an enterprise analytics solution in just a few clicks.

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all of your resources.

Subscription *

Resource group * [Create new](#)

Managed resource group

Workpace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name *

Region *

Select Data Lake Storage Gen2 * ☒ From subscription ☐ Manually via URL

[Review + create](#) [< Previous](#) [Next: Security >](#)

Рисунок 2.5 - Створення робочого простору в Azure Synapse Analytics

На наступному етапі здійснимо завантаження згенерованих CSV файлів до нашого сховища даних, що дозволить ефективно зберігати та організувати інформацію для подальшої обробки, аналізу та використання в процесах прийняття рішень (рис. 2.6).

Microsoft Azure | Synapse Analytics | synapsekholodnitskiy

Synapse live | Validate all | Publish all

Workspace | Linked

Filter resources by name

Azure Data Lake Storage Gen2 | 2

synapsekholodnitskiy (Primary - st...)

kholodnitskiy (Primary)

(Attached Containers)

Upload | New folder | Select all | Refresh

Destination folder: /

File Upload

File name | Size | Action

dim_reviews.csv	35.6 KB	Remove
fact_forecasts.csv	24.9 KB	Remove
fact_sales.csv	14.8 KB	Remove
dim_time.csv	17.1 KB	Remove
dim_customers.csv	40.0 KB	Remove
dim_addresses.csv	22.6 KB	Remove
dim_products.csv	42.0 KB	Remove
dim_categories.csv	149 B	Remove

[Upload](#) [Cancel](#)

Рисунок 2.6 – Завантаження CSV файлів до сховища даних

Створюємо виділений SQL пул (рис. 2.7), що дозволяє ефективно організувати обробку великих обсягів даних. Після цього підготовлюємо скрипт для створення таблиць, які будуть використовуватися в аналітичному сховищі

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		22

даних, що є важливим етапом для подальшої роботи з великими масивами інформації. Процес налаштування сховища є ключовим для забезпечення швидкого та ефективного доступу до необхідних даних (рис. 2.8).

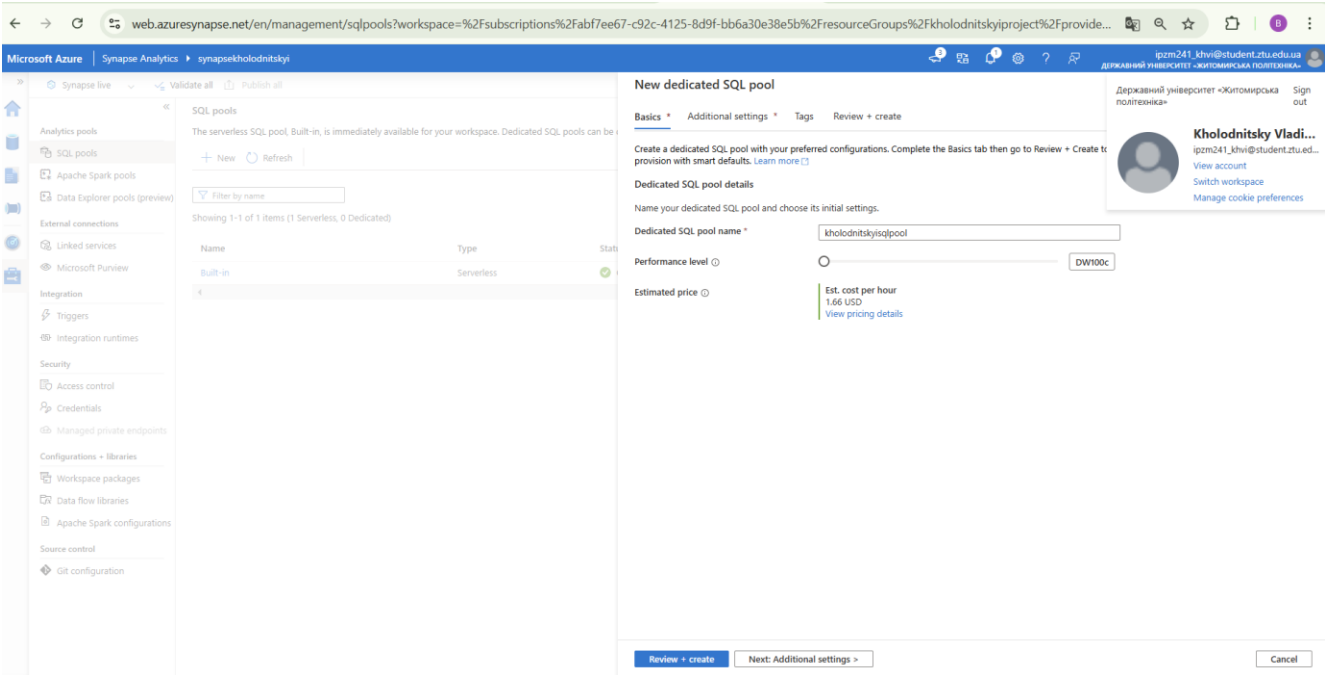


Рисунок 2.7 – Створення виділенного SQL пулу

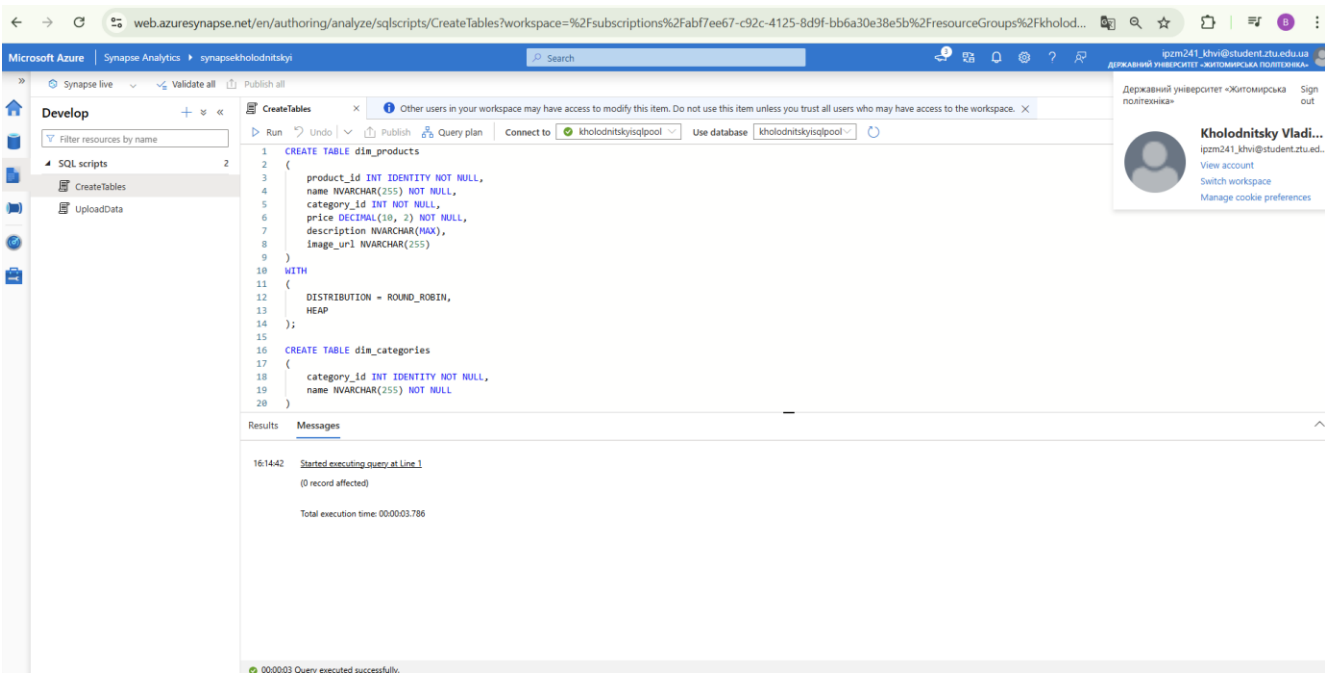


Рисунок 2.8 – Виконання скрипта CreateTables

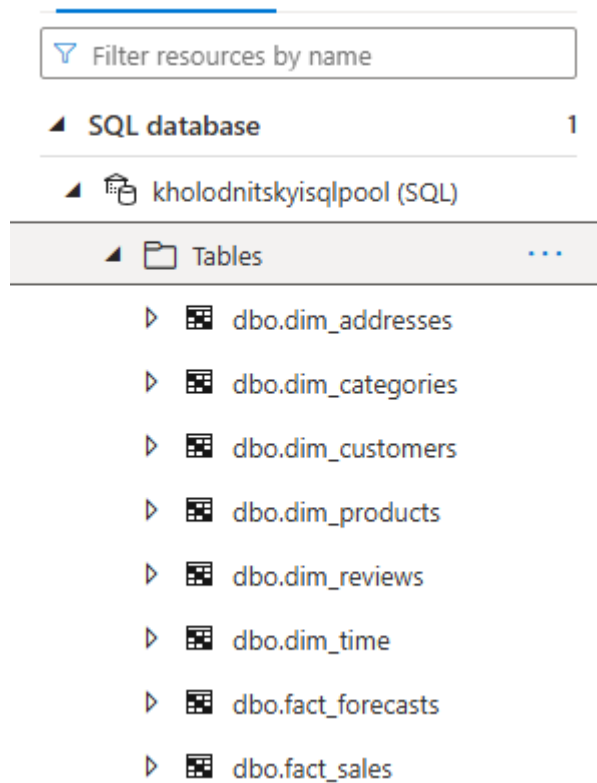


Рисунок 2.9 – Створені таблиці

Розробимо та виконаємо скрипт для автоматичного завантаження даних із файлового сховища до таблиці в сховище даних (рис. 2.10).

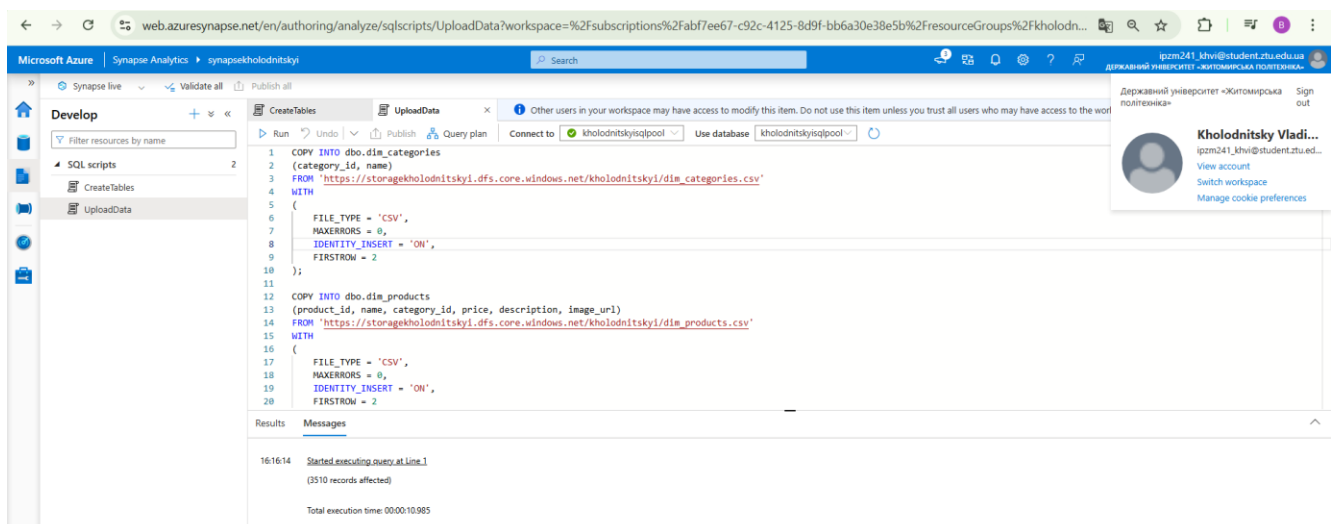


Рисунок 2.10 – Завантаження даних до таблиць

Аналітичне сховище для системи прогнозування попиту на товари з урахуванням відгуків і рейтингів покупців вже створене і містить всі необхідні дані для проведення подальшого аналізу. На наступному етапі буде розроблено набір SQL-запитів (рис. 2.11 – 2.20), які дозволять здійснити глибокий аналіз взаємозв'язку між рейтингами, відгуками покупців і попитом на товари. Результати

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				24
Змн.	Арк.	№ докум.	Підпис	Дата		

виконання цих запитів будуть експортовані у формат CSV, що забезпечить можливість їх подальшої обробки і візуалізації для отримання точних прогнозів попиту на товари на основі аналізу відгуків та оцінок споживачів.

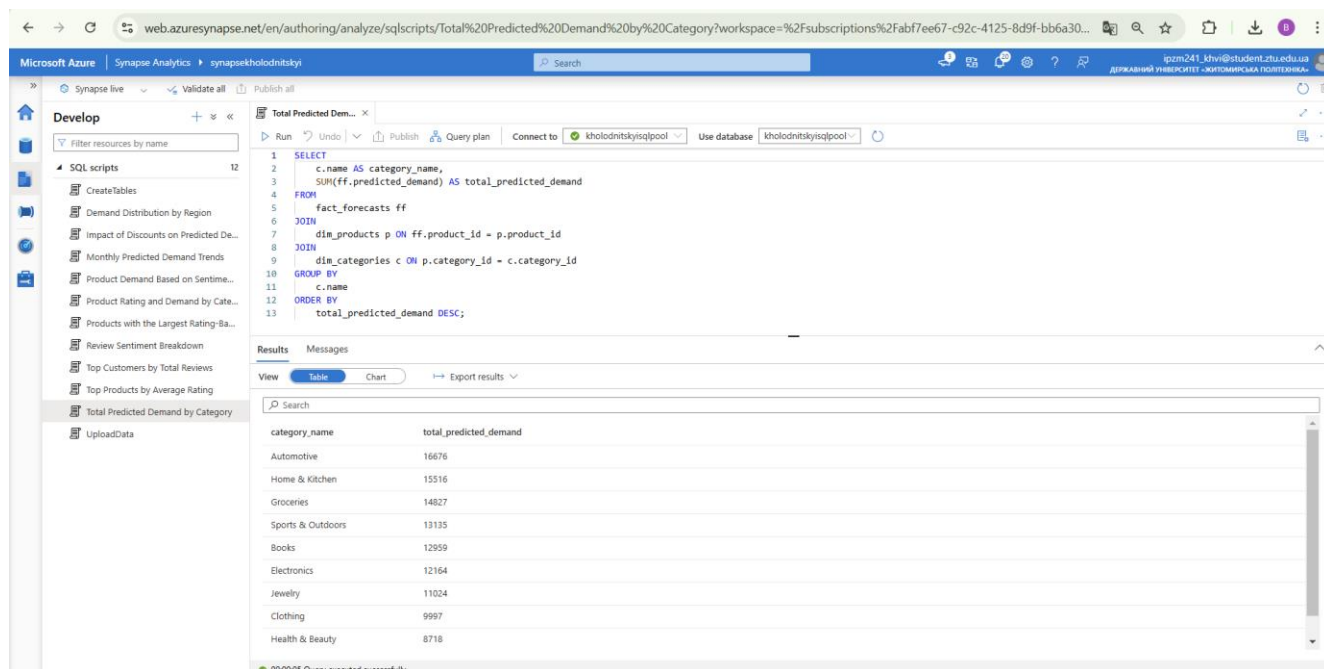


Рисунок 2.11 – Загальний прогнозований попит за категоріями

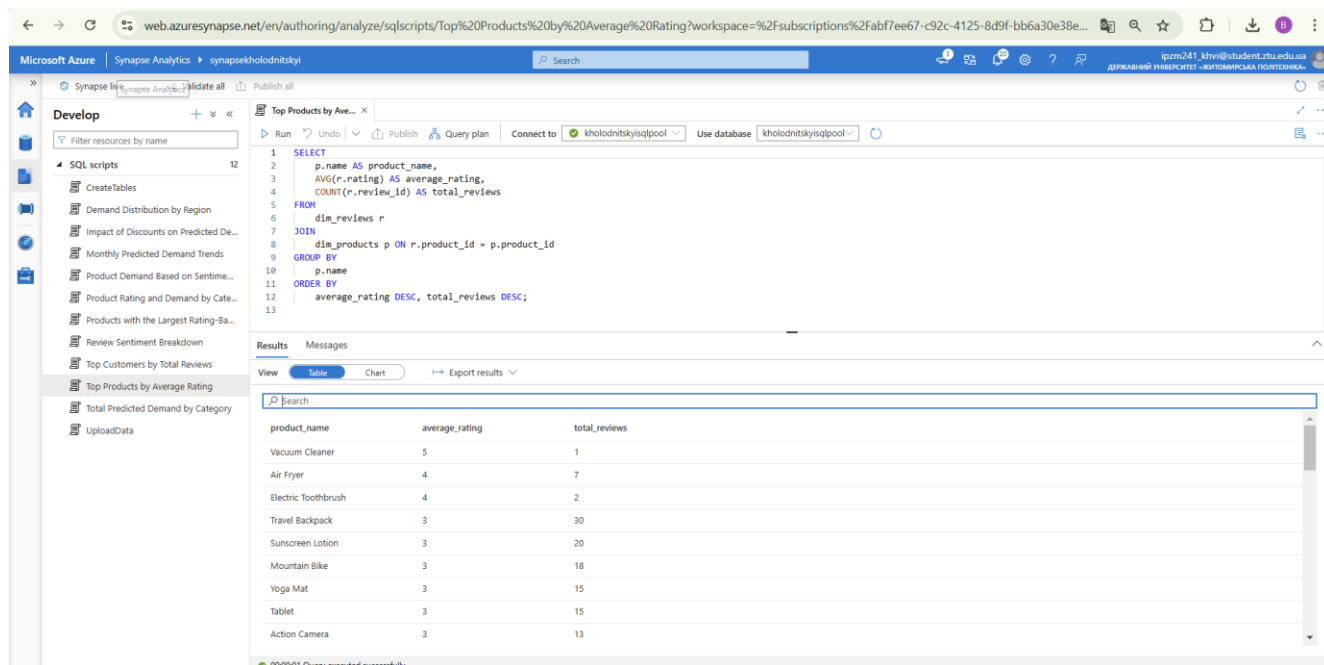


Рисунок 2.12 – Топ товарів за середнім рейтингом

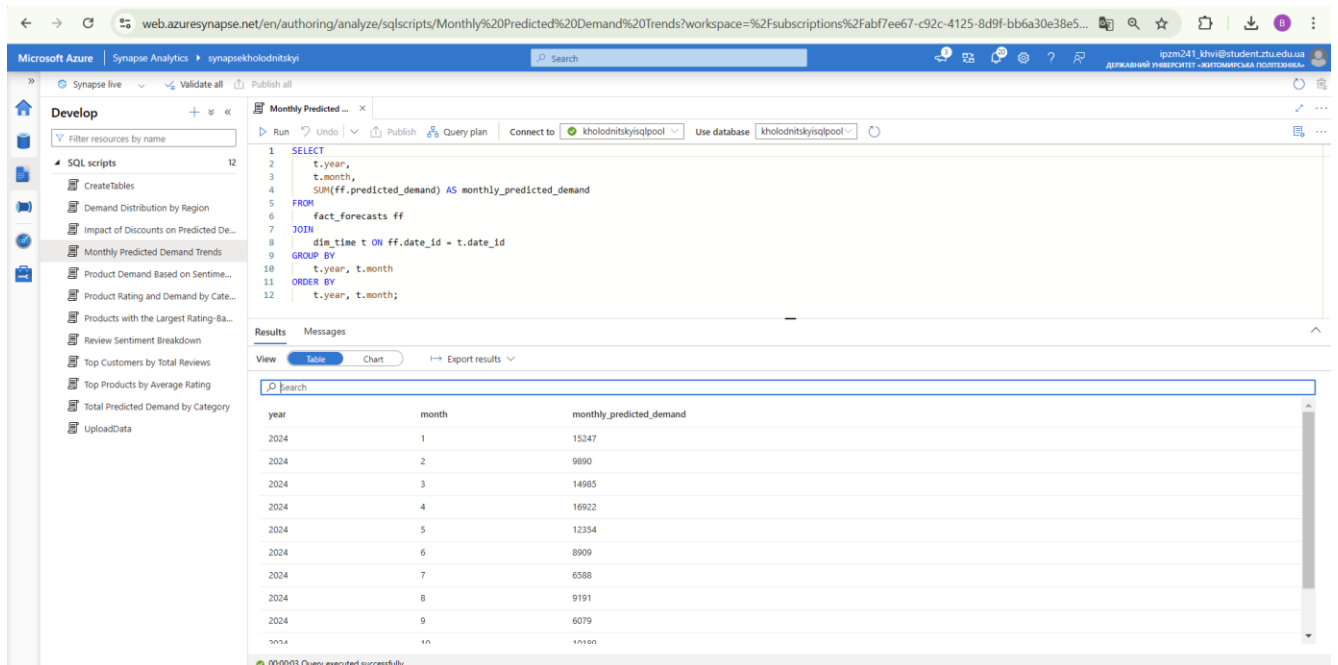


Рисунок 2.13 – Щомісячні тенденції прогнозованого попиту

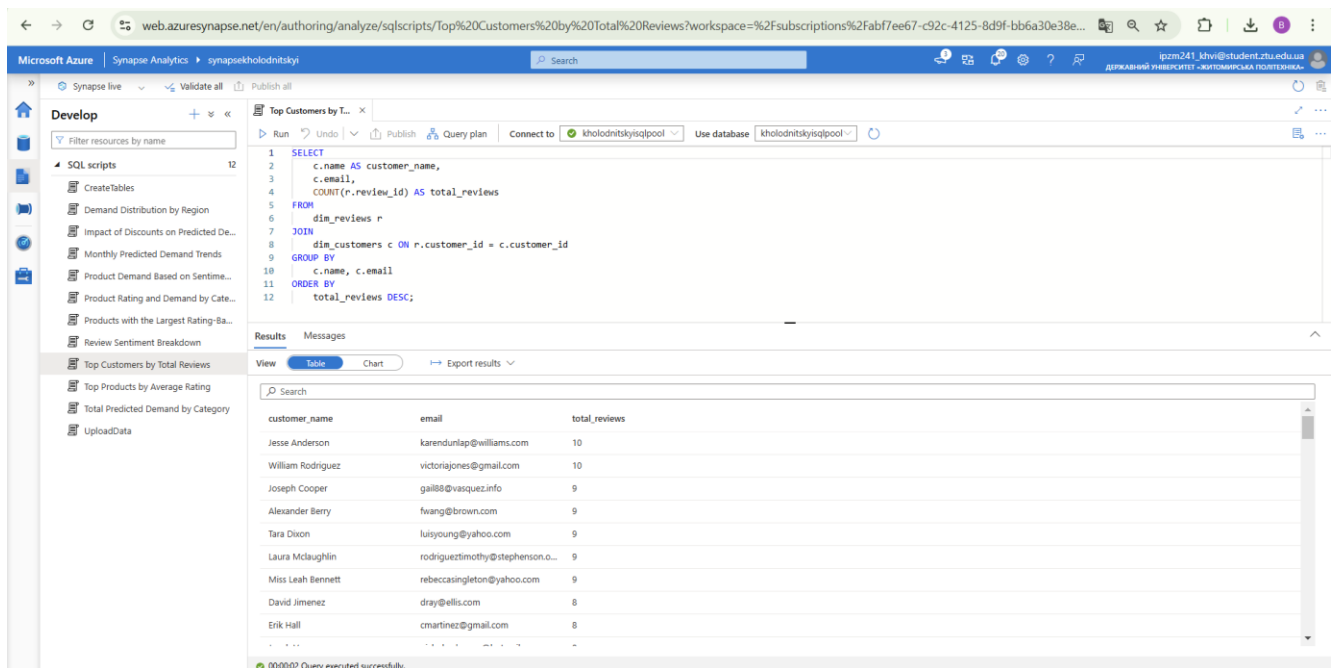


Рисунок 2.14 – Топ клієнтів за кількістю відгуків

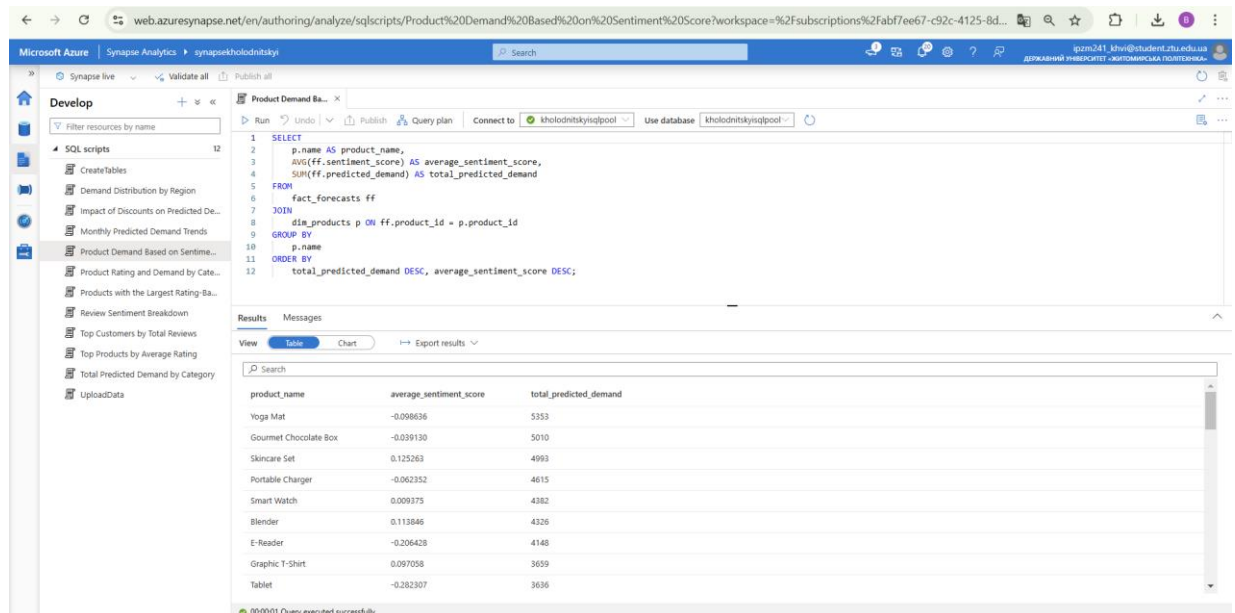


Рисунок 2.15 – Попит на товари на основі настрою відгуків

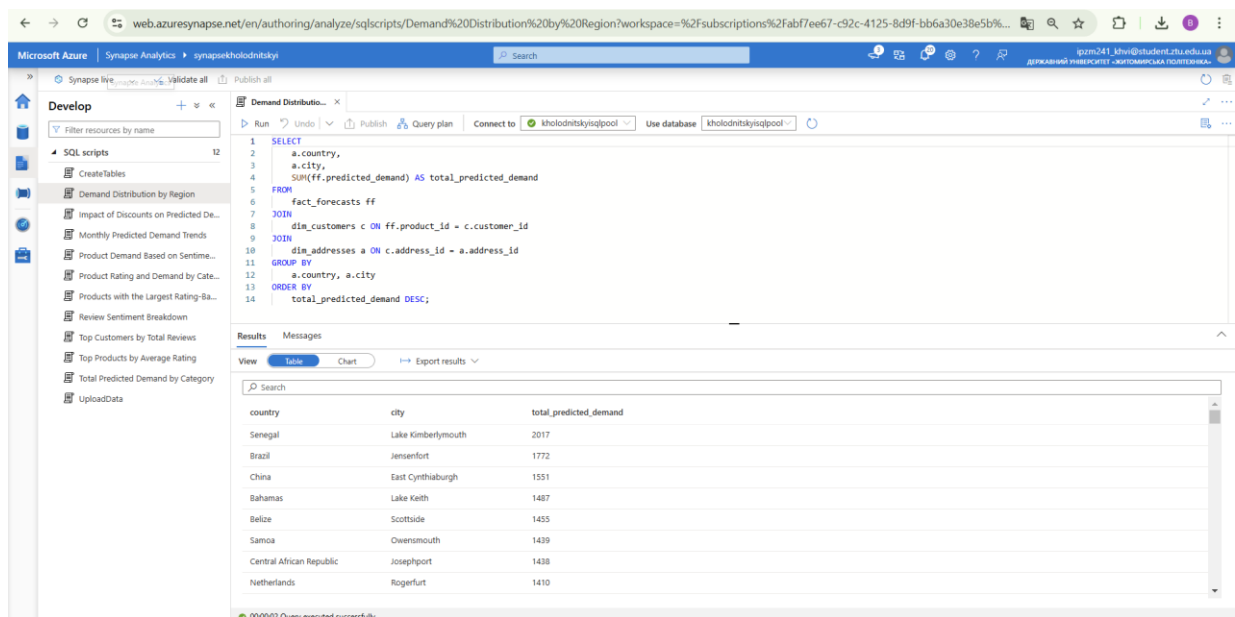


Рисунок 2.16 – Розподіл попиту за регіонами

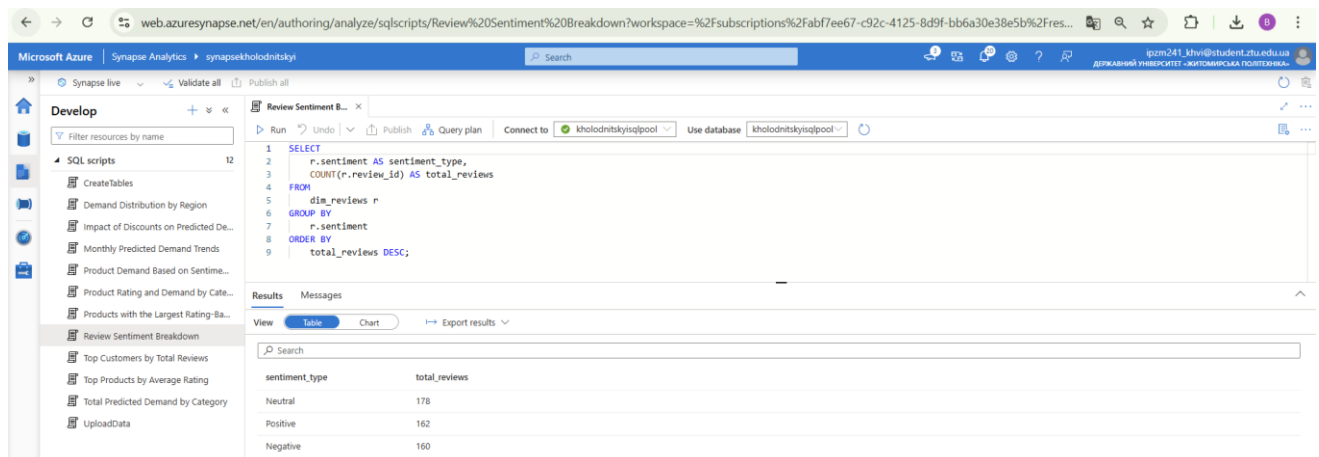


Рисунок 2.17 – Розподіл відгуків за настроєм

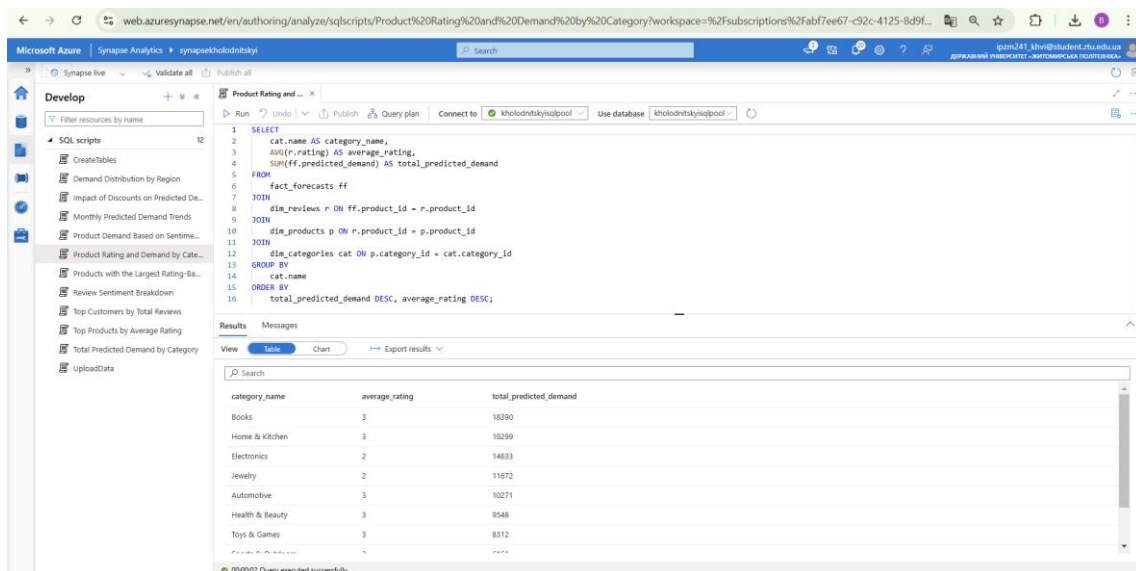


Рисунок 2.18 – Рейтинг і попит на товари за категоріями

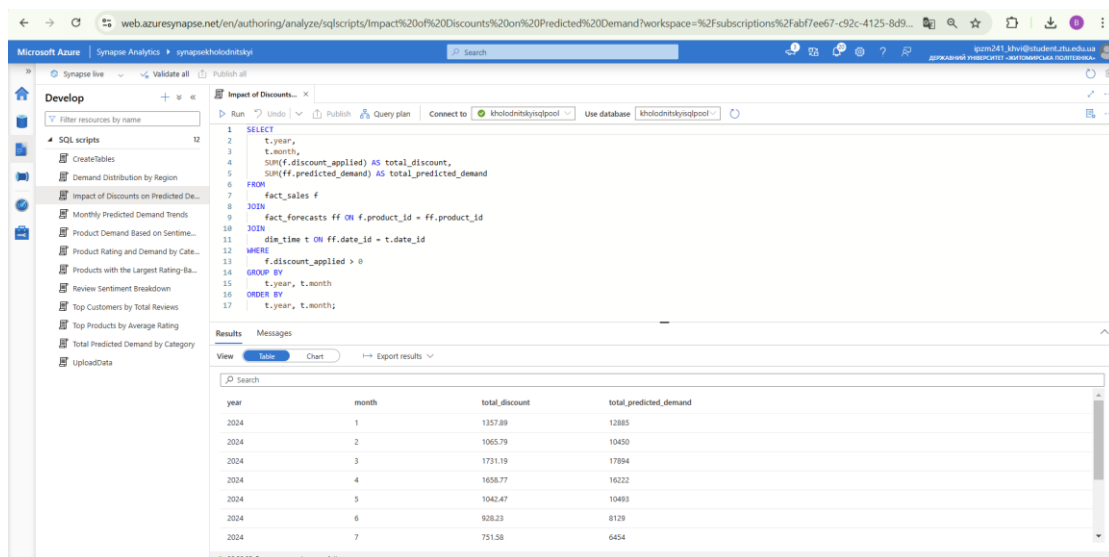


Рисунок 2.19 – Вплив знижок на прогнозований попит

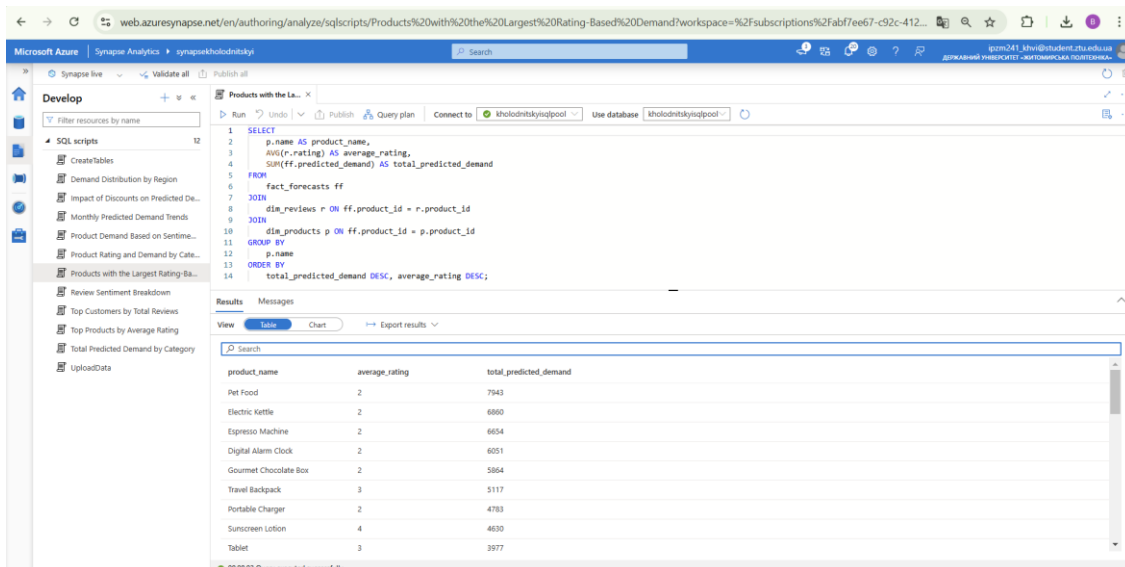


Рисунок 2.20 – Товари з найбільшим попитом на основі рейтингів

Висновки до розділу 2

У цьому розділі було розроблено структуру даних для прогнозування попиту на товари, зокрема транзакційні таблиці, які забезпечують нормалізацію інформації та ефективну обробку запитів. На основі отриманих даних було створено модель для обробки та зберігання даних, а також побудовано структуру для подальшого багатовимірного аналізу.

Використовуючи інструменти для обробки великих обсягів інформації, було сформовано сховище даних, що забезпечило швидкий доступ до необхідних даних для подальшого аналізу. У результаті розроблено кілька аналітичних запитів, які дозволяють здійснювати глибокий аналіз прогнозованого попиту, включаючи вплив різних факторів, таких як знижки та відгуки клієнтів.

Отримані результати є основою для подальшої візуалізації даних і створення звітів, що дозволяють отримати більш детальне уявлення про ефективність процесів та підтримувати прийняття обґрунтованих рішень у бізнесі.

Отримані дані також створюють основу для подальшого удосконалення системи прогнозування попиту, що дозволяє адаптувати стратегії до змін на ринку. Розроблені моделі та аналітичні запити забезпечують високу точність прогнозів, що сприяє оптимізації бізнес-процесів та покращенню ефективності прийняття рішень на основі реальних даних.

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				29
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 3. ОРГАНІЗАЦІЯ ЗАСОБІВ ВІЗУАЛІЗАЦІЇ ТА АНАЛІЗУ ДАНИХ

3.1 Технологія Power BI та її основні можливості

Power BI — це сучасна платформа бізнес-аналітики, створена компанією Microsoft, яка призначена для інтеграції, аналізу та візуалізації даних. Вона є універсальним інструментом, що дозволяє ефективно працювати з даними різного формату, консолідувати їх з різних джерел та створювати інтерактивні звіти. Power BI допомагає бізнесам усіх розмірів отримувати цінну інформацію, необхідну для прийняття обґрунтованих управлінських рішень [7].

Платформа забезпечує повний цикл роботи з даними: від збору до представлення результатів. Її функціонал дозволяє підключатися до різноманітних джерел, таких як локальні бази даних, хмарні сервіси, файли Excel, веб-ресурси та інші API. Це забезпечує інтеграцію розрізнених даних у єдине середовище, що сприяє гнучкості й зручності роботи. Power BI також надає інструменти для очищення та трансформації даних за допомогою вбудованого редактора Power Query. Ця функція дозволяє автоматизувати процеси обробки даних (ETL-процеси), що особливо важливо при роботі з великими обсягами інформації [7].

Одним із головних переваг Power BI є її можливість створювати інтуїтивно зрозумілі візуалізації. Панелі моніторингу, діаграми, графіки, карти та таблиці дозволяють представити навіть складну аналітичну інформацію у зрозумілому та доступному форматі. Візуалізації є інтерактивними: користувачі можуть змінювати параметри фільтрації, детально аналізувати окремі сегменти даних та отримувати оновлені результати у реальному часі. Це робить аналіз більш гнучким і адаптивним до потреб бізнесу [7].

Power BI також підтримує хмарні технології. Використання Power BI Service дає змогу отримувати доступ до звітів із будь-якого пристрою з підключенням до Інтернету. Це сприяє зручному використанню платформи для віддаленої роботи та спільної аналітики в командах. Завдяки можливості публікації звітів у хмарі всі учасники процесу прийняття рішень отримують актуальну інформацію одночасно [7].

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		30

Окрім цього, Power BI інтегрується з інструментами машинного навчання, такими як Python і Azure Machine Learning. Це відкриває нові можливості для прогнозування майбутніх трендів, аналізу поведінки споживачів і виявлення прихованих закономірностей у великих наборах даних. Завдяки таким функціям платформа стає не лише аналітичним інструментом, але й потужним засобом для стратегічного планування [7].

Універсальність Power BI полягає також у тому, що він доступний для користувачів із різними рівнями технічної підготовки. Інтуїтивний інтерфейс дозволяє швидко освоїти основні функції платформи навіть тим, хто не має спеціалізованих навичок у сфері роботи з даними [7].

Таким чином, Power BI є потужним інструментом, який об'єднує гнучкість, функціональність і простоту у використанні. Це робить його надзвичайно ефективним у вирішенні завдань бізнес-аналітики, аналізу великих обсягів даних та прийняття рішень на основі чітких, структурованих і візуалізованих показників [7].

3.2 Публікація звіту на прикладній дошці Power BI

Для виконання візуалізації та аналізу в Power BI необхідно завантажити дані, підготовлені у форматі CSV. Ці файли були сформовані після виконання SQL-запитів, описаних у розділі 2.2 (рис. 3.1). CSV-файли містять структуровану інформацію, яка підходить для подальшої обробки та аналізу в Power BI.

Формат CSV забезпечує сумісність із платформою, що спрощує процес імпорту даних. Після завантаження інформація може бути трансформована за допомогою вбудованих інструментів Power BI для підготовки до створення інтерактивних звітів і візуалізацій.

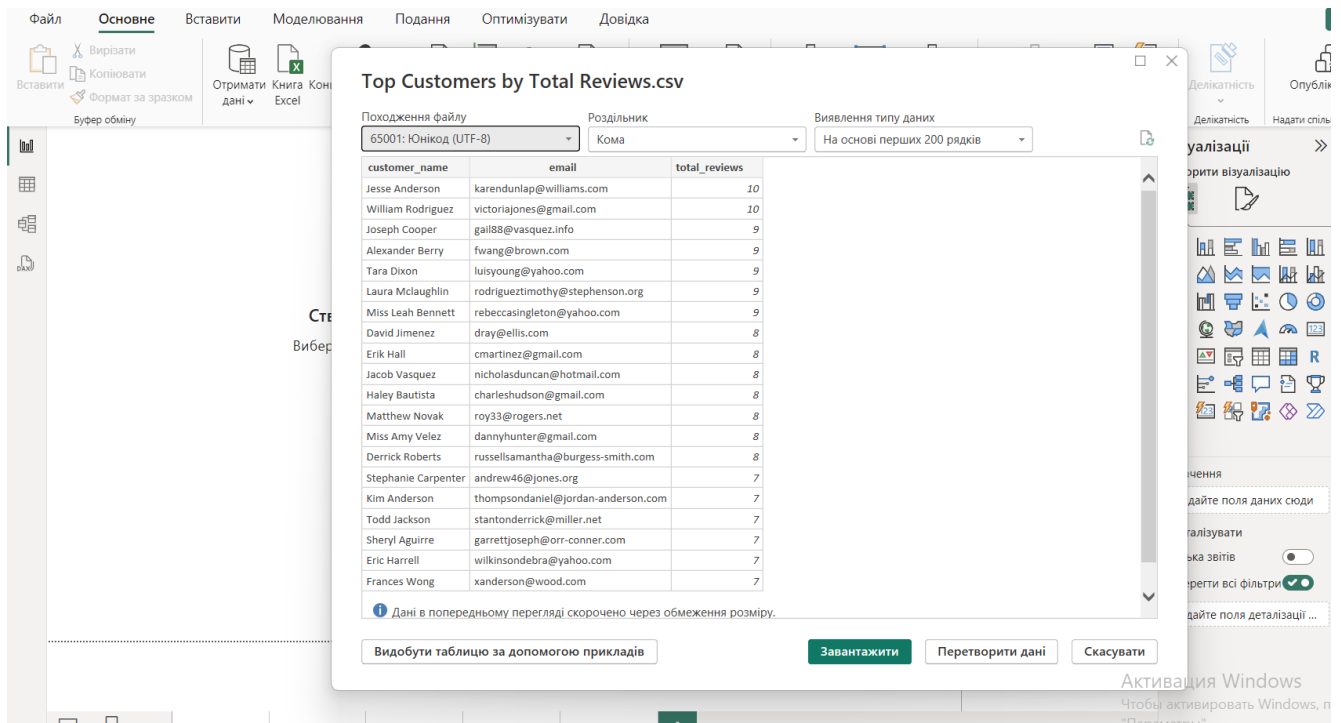


Рисунок 3.1 – Завантаження даних у Power BI

Після завантаження даних у Power BI була сформована структура, що відображає зв'язки між таблицями та атрибутами (рис 3.2).

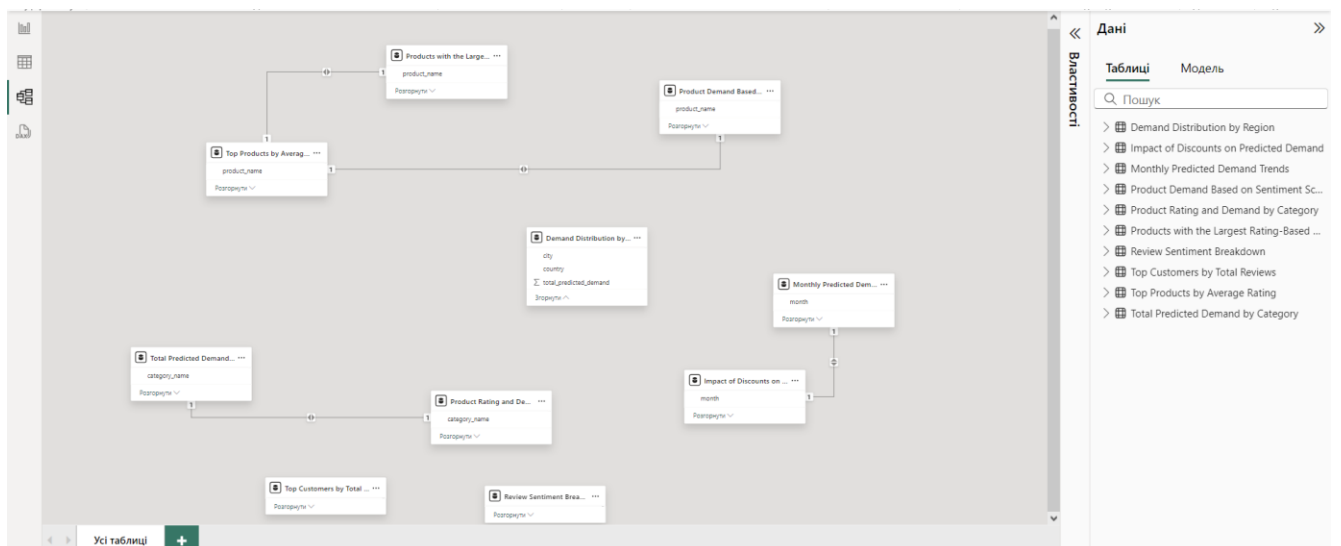


Рисунок 3.2 – Структура даних у Power BI

Наступним кроком буде створення першого дашборду, на якому будуть представлені щомісячні тенденції прогнозованого попиту, загальний прогнозований попит за категоріями, а також вплив знижок на прогнозований попит (рис. 3.3). Цей дашборд дозволить візуалізувати ключові показники та зручним способом відстежувати зміни попиту, що забезпечить більш ефективне прийняття рішень щодо планування запасів та цінової політики.

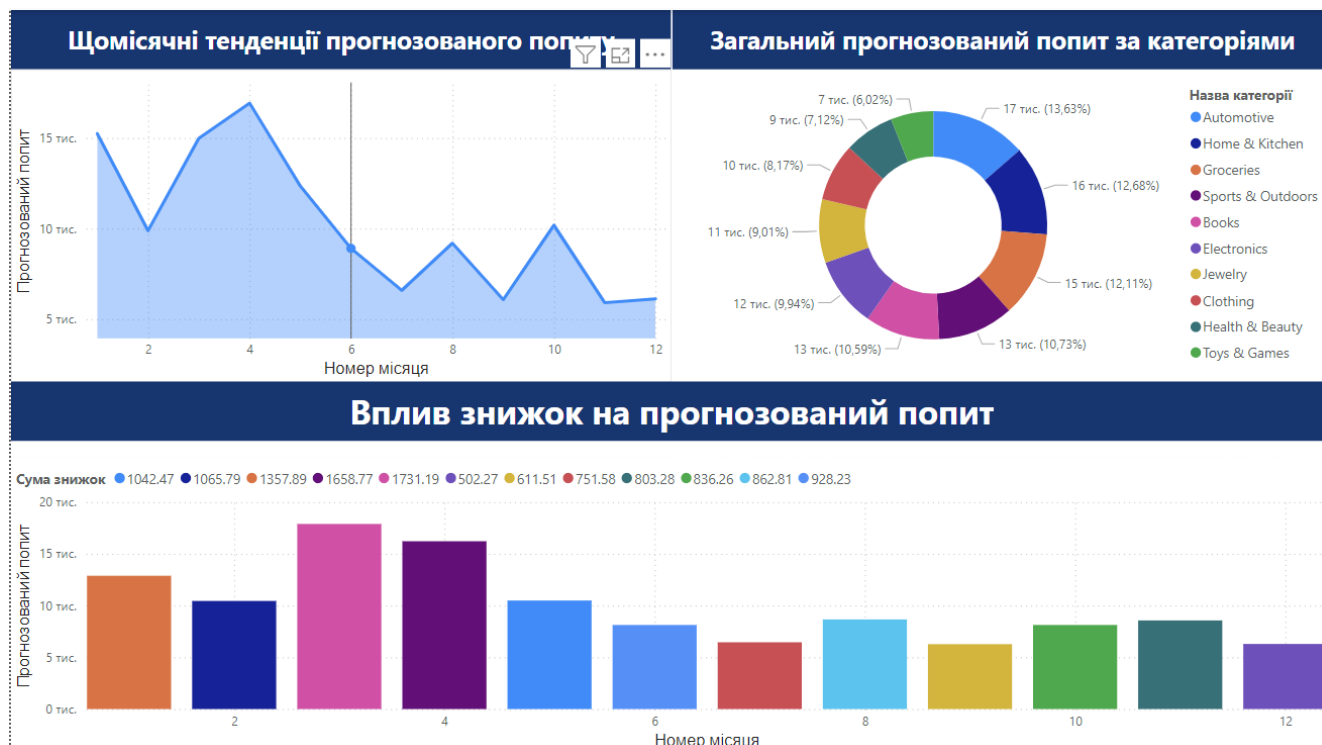


Рисунок 3.3 – Прогнозований попит з аналізом тенденцій та впливу знижок

Далі було створено дашборд з однією візуалізацією — розподіл попиту за регіонами (рис. 3.4). Для цього була використана карта, яка дозволяє візуалізувати точки попиту, що дає змогу наочно побачити географічне розташування попиту в різних регіонах. Така візуалізація сприяє кращому розумінню просторових тенденцій і дозволяє приймати більш обґрунтовані рішення щодо розподілу ресурсів та маркетингових стратегій.

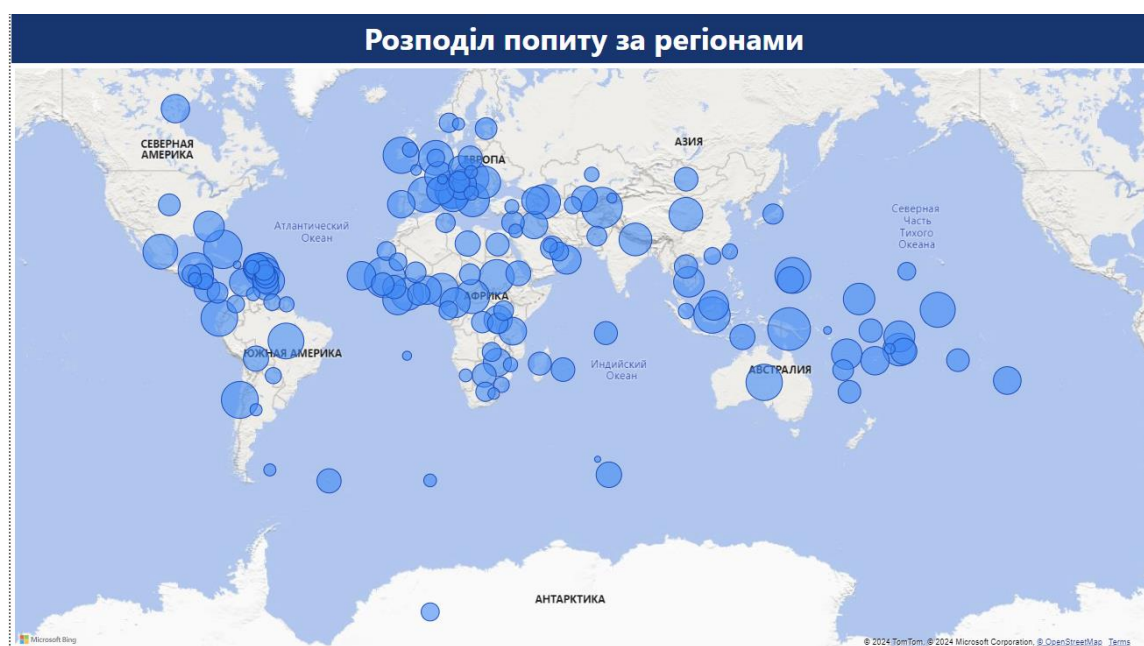


Рисунок 3.4 – Розподіл попиту за регіонами

		Холодніцький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				33
Змн.	Арк.	№ докум.	Підпис	Дата		

Далі був створений дешборд для візуалізації середнього рейтингу та кількості відгуків по товарам, а також прогнозованого попиту на ці товари (рис. 3.5). Для зручності вибору конкретного товару було використано роздільник у форматі плиток, що дозволяє легко переглядати дані для кожного товару окремо. Такий підхід дозволяє користувачам швидко отримувати необхідну інформацію щодо рейтингу, відгуків і попиту, що допомагає в ухваленні більш обґрунтованих рішень.

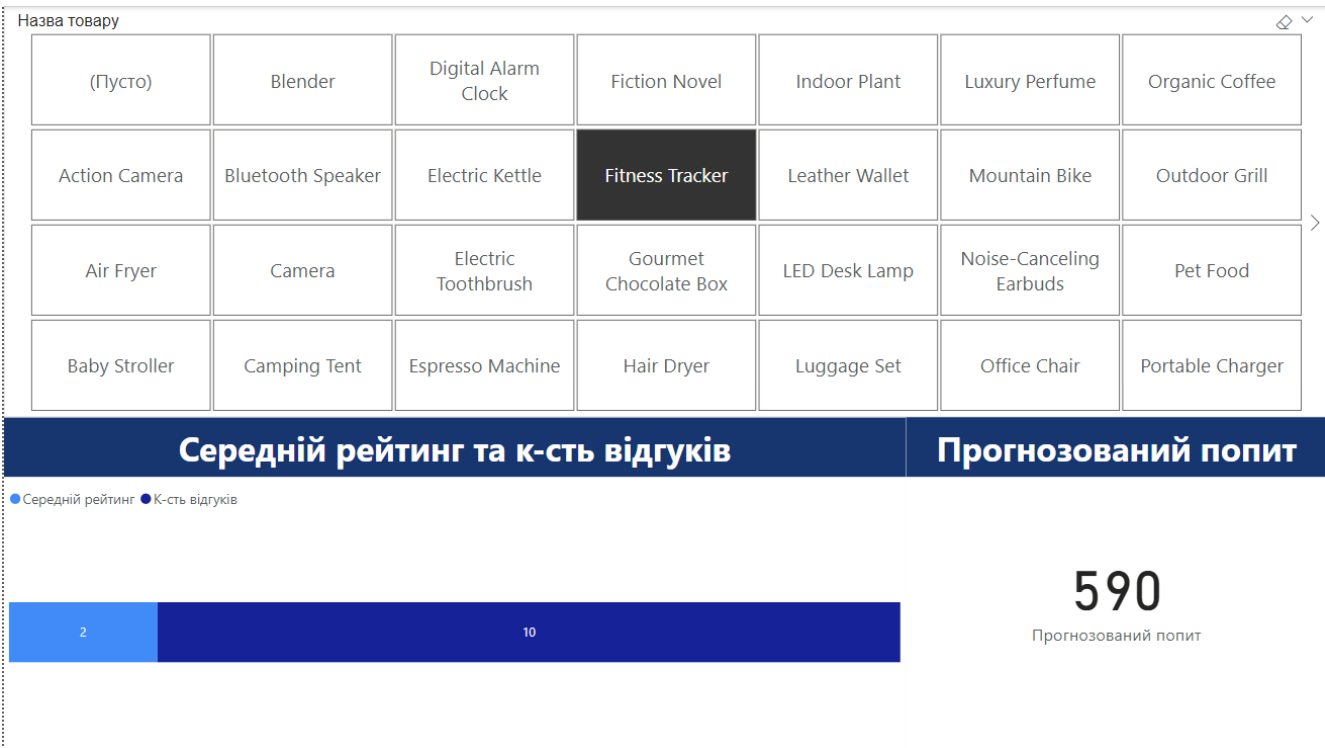


Рисунок 3.5 – Візуалізація рейтингу та прогнозованого попиту по товарам

Останнім етапом стала візуалізація топ клієнтів за кількістю відгуків, розподілу відгуків за настроєм та прогнозованого попиту на товари на основі настрою відгуків (рис. 3.6). Така візуалізація дозволяє не тільки визначити найактивніших клієнтів, але й аналізувати, як позитивні та негативні відгуки впливають на попит на товари. Цей підхід дає можливість виявити важливі зв’язки між настроєм відгуків та поведінкою споживачів, що є корисним для стратегічного планування.



Рисунок 3.6 – Візуалізація розподілу відгуків за настроєм та попиту на товари

Висновки до розділу 3

У даному розділі було детально розглянуто процеси візуалізації та аналізу даних, що стосуються прогнозу попиту на товари з урахуванням відгуків і рейтингів покупців. За допомогою платформи Power BI було створено інтерактивні панелі, які забезпечують наочне представлення ключових показників, таких як тенденції попиту, вплив рейтингів на продажі та динаміка споживчих уподобань.

Реалізовані візуалізації дозволяють користувачам оперативно отримувати динамічну інформацію про основні аспекти роботи системи, включаючи аналіз популярності товарів, виявлення сезонних коливань і визначення товарів із високим потенціалом продажу.

Окрім того, візуалізаційні панелі сприяють більш глибокому розумінню взаємозв'язків між оцінками покупців і реальними показниками продажів, що відкриває можливості для коригування маркетингових стратегій та планування асортименту. Це допомагає бізнесу реагувати на потреби ринку більш гнучко й ефективно.

У підсумку, використання Power BI значно спростило аналіз ключових даних, зробивши його більш інтуїтивно зрозумілим для кінцевих користувачів. Завдяки інтерактивним звітам та візуалізаціям, система стала ефективним інструментом для прийняття обґрунтованих управлінських рішень і своєчасної адаптації бізнес-процесів до змін ринкових умов.

		Холодніцький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		36

ВИСНОВКИ

У результаті виконання курсової роботи було проведено комплексне дослідження моделей інформаційних процесів у системі прогнозу попиту на товари з урахуванням відгуків і рейтингів покупців. Аналіз існуючих підходів до моделювання та впровадження сучасних інформаційних технологій дозволив створити ефективну систему, здатну забезпечити високу точність прогнозування та підтримку прийняття рішень.

У рамках роботи було використано методології моделювання, такі як IDEF0 та BPMN, які дозволили структурувати ключові етапи обробки даних, необхідних для прогнозування попиту. Цей підхід сприяв виявленню слабких місць у системі та визначенню шляхів для її оптимізації.

На основі побудованих моделей було розроблено транзакційне та аналітичне сховища даних, що забезпечило ефективну обробку великих обсягів інформації, включаючи відгуки та рейтинги покупців. Використання сучасних технологій обробки даних дозволило автоматизувати ETL-процеси, інтегрувати дані з різних джерел для їх подальшого аналізу, а також створити аналітичні звіти, які враховують вплив суб'єктивних оцінок на попит.

Застосування візуалізацій у Power BI надало змогу створити інформативні звіти, які допомагають аналізувати тенденції попиту та забезпечувати ефективне прийняття управлінських рішень.

Результати роботи підтвердили, що використання сучасних моделей і підходів до моделювання інформаційних процесів у прогнозуванні попиту дозволяє не лише підвищити точність прогнозів, а й гнучко адаптувати систему до змін у поведінці покупців. Це робить систему ефективною, конкурентоспроможною та здатною до подальшого розвитку.

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		37

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. IDEF0 - Wikipedia. *Wikipedia, the free encyclopedia*. URL: <https://en.wikipedia.org/wiki/IDEF0> (дата звернення: 26.12.2024).
2. Business Process Model and Notation - Wikipedia. *Wikipedia, the free encyclopedia*. URL: https://en.wikipedia.org/wiki/Business_Process_Model_and_Notation (дата звернення: 26.12.2024).
3. What Is Online Transactional Processing (OLTP)? | IBM. *IBM - United States*. URL: <https://www.ibm.com/think/topics/oltp> (дата звернення: 26.12.2024).
4. What is OLAP? - Online Analytical Processing Explained - AWS. *Amazon Web Services, Inc.* URL: https://aws.amazon.com/what-is/olap/?nc1=h_ls (дата звернення: 26.12.2024).
5. OLTP vs OLAP - Difference Between Data Processing Systems - AWS. *Amazon Web Services, Inc.* URL: <https://aws.amazon.com/compare/the-difference-between-olap-and-oltp/> (дата звернення: 26.12.2024).
6. Azure Synapse Analytics - Azure Synapse Analytics. *Microsoft Learn: Build skills that open doors in your career*. URL: <https://learn.microsoft.com/en-us/azure/synapse-analytics/> (дата звернення: 26.12.2024).
7. What is Power BI? - Power BI. *Microsoft Learn: Build skills that open doors in your career*. URL: <https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview> (дата звернення: 26.12.2024).

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		38

ДОДАТКИ

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		39

Скрипт для генерації таблиць в аналітичному сховищі даних.

```
CREATE TABLE dim_products
(
    product_id INT IDENTITY NOT NULL,
    name NVARCHAR(255) NOT NULL,
    category_id INT NOT NULL,
    price DECIMAL(10, 2) NOT NULL,
    description NVARCHAR(MAX),
    image_url NVARCHAR(255)
)
WITH
(
    DISTRIBUTION = ROUND_ROBIN,
    HEAP
);

CREATE TABLE dim_categories
(
    category_id INT IDENTITY NOT NULL,
    name NVARCHAR(255) NOT NULL
)
WITH
(
    DISTRIBUTION = ROUND_ROBIN,
    HEAP
);

CREATE TABLE dim_customers
(
    customer_id INT IDENTITY NOT NULL,
    name NVARCHAR(255) NOT NULL,
    email NVARCHAR(255) NOT NULL,
    phone NVARCHAR(20),
    address_id INT NOT NULL,
    loyalty_points INT DEFAULT 0,
    registration_date DATETIME NOT NULL
)
WITH
(
    DISTRIBUTION = ROUND_ROBIN,
    HEAP
);

CREATE TABLE dim_addresses
(
    address_id INT IDENTITY NOT NULL,
    country NVARCHAR(100) NOT NULL,
    city NVARCHAR(100) NOT NULL,
    street NVARCHAR(255),
```



```

        zip_code NVARCHAR(20),
        region NVARCHAR(100)
    )
WITH
(
    DISTRIBUTION = ROUND_ROBIN,
    HEAP
);

CREATE TABLE dim_time
(
    date_id INT IDENTITY NOT NULL,
    date DATE NOT NULL,
    year INT NOT NULL,
    quarter INT NOT NULL,
    month INT NOT NULL,
    week INT NOT NULL,
    day_of_week NVARCHAR(20) NOT NULL
)
WITH
(
    DISTRIBUTION = ROUND_ROBIN,
    HEAP
);

CREATE TABLE dim_reviews
(
    review_id INT IDENTITY NOT NULL,
    product_id INT NOT NULL,
    customer_id INT NOT NULL,
    rating TINYINT NOT NULL,
    sentiment NVARCHAR(20),
    comment NVARCHAR(MAX),
    date DATETIME NOT NULL
)
WITH
(
    DISTRIBUTION = ROUND_ROBIN,
    HEAP
);

CREATE TABLE fact_sales
(
    transaction_id INT IDENTITY NOT NULL,
    date_id INT NOT NULL,
    customer_id INT NOT NULL,
    product_id INT NOT NULL,
    quantity_sold INT NOT NULL,
    total_amount DECIMAL(10, 2) NOT NULL,
    discount_applied DECIMAL(5, 2) DEFAULT 0
)

```

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		41

```

WITH
(
    DISTRIBUTION = ROUND_ROBIN,
    CLUSTERED COLUMNSTORE INDEX
);

CREATE TABLE fact_forecasts
(
    forecast_id INT IDENTITY NOT NULL,
    product_id INT NOT NULL,
    date_id INT NOT NULL,
    predicted_demand INT NOT NULL,
    confidence_level DECIMAL(5, 2),
    average_rating DECIMAL(3, 2),
    sentiment_score DECIMAL(5, 2),
    created_at DATETIME NOT NULL
)
WITH
(
    DISTRIBUTION = ROUND_ROBIN,
    CLUSTERED COLUMNSTORE INDEX
);

```

Скрипт для заповнення таблиць згенерованими даними.

```

COPY INTO dbo.dim_categories
(category_id, name)
FROM
'https://storagekholodnitskyi.dfs.core.windows.net/kholodnitskyi/dim_categories.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO dbo.dim_products
(product_id, name, category_id, price, description, image_url)
FROM
'https://storagekholodnitskyi.dfs.core.windows.net/kholodnitskyi/dim_products.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO dbo.dim_addresses
(address_id, country, city, street, zip_code, region)

```

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		42

```

FROM
'https://storagekholodnitskyi.dfs.core.windows.net/kholodnitskyi/dim_addresses.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO dbo.dim_customers
(customer_id, name, email, phone, address_id, loyalty_points, registration_date)
FROM
'https://storagekholodnitskyi.dfs.core.windows.net/kholodnitskyi/dim_customers.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO dbo.dim_time
(date_id, date, year, quarter, month, week, day_of_week)
FROM 'https://storagekholodnitskyi.dfs.core.windows.net/kholodnitskyi/dim_time.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO dbo.dim_reviews
(review_id, product_id, customer_id, rating, sentiment, comment, date)
FROM
'https://storagekholodnitskyi.dfs.core.windows.net/kholodnitskyi/dim_reviews.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO dbo.fact_sales
(transaction_id, date_id, customer_id, product_id, quantity_sold, total_amount,
discount_applied)
FROM 'https://storagekholodnitskyi.dfs.core.windows.net/kholodnitskyi/fact_sales.csv'
WITH
(

```

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		43

```

FILE_TYPE = 'CSV',
MAXERRORS = 0,
IDENTITY_INSERT = 'ON',
FIRSTROW = 2
);

COPY INTO dbo.fact_forecasts
(forecast_id, product_id, date_id, predicted_demand, confidence_level,
average_rating, sentiment_score, created_at)
FROM
'https://storagekholodnitskyi.dfs.core.windows.net/kholodnitskyi/fact_forecasts.csv'
WITH
(
FILE_TYPE = 'CSV',
MAXERRORS = 0,
IDENTITY_INSERT = 'ON',
FIRSTROW = 2
);

```

Скрипт Demand Distribution by Region.

```

SELECT
a.country,
a.city,
SUM(ff.predicted_demand) AS total_predicted_demand
FROM
fact_forecasts ff
JOIN
dim_customers c ON ff.product_id = c.customer_id
JOIN
dim_addresses a ON c.address_id = a.address_id
GROUP BY
a.country, a.city
ORDER BY
total_predicted_demand DESC;

```

Скрипт Impact of Discounts on Predicted Demand.

```

SELECT
t.year,
t.month,
SUM(f.discount_applied) AS total_discount,
SUM(ff.predicted_demand) AS total_predicted_demand
FROM
fact_sales f
JOIN
fact_forecasts ff ON f.product_id = ff.product_id
JOIN
dim_time t ON ff.date_id = t.date_id
WHERE
f.discount_applied > 0
GROUP BY

```

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		44

```

        t.year, t.month
ORDER BY
        t.year, t.month;

```

Скрипт Monthly Predicted Demand Trends.

```

SELECT
    t.year,
    t.month,
    SUM(ff.predicted_demand) AS monthly_predicted_demand
FROM
    fact_forecasts ff
JOIN
    dim_time t ON ff.date_id = t.date_id
GROUP BY
    t.year, t.month
ORDER BY
    t.year, t.month;

```

Скрипт Product Demand Based on Sentiment Score.

```

SELECT
    p.name AS product_name,
    AVG(ff.sentiment_score) AS average_sentiment_score,
    SUM(ff.predicted_demand) AS total_predicted_demand
FROM
    fact_forecasts ff
JOIN
    dim_products p ON ff.product_id = p.product_id
GROUP BY
    p.name
ORDER BY
    total_predicted_demand DESC, average_sentiment_score DESC;

```

Скрипт Product Rating and Demand by Category.

```

SELECT
    cat.name AS category_name,
    AVG(r.rating) AS average_rating,
    SUM(ff.predicted_demand) AS total_predicted_demand
FROM
    fact_forecasts ff
JOIN
    dim_reviews r ON ff.product_id = r.product_id
JOIN
    dim_products p ON r.product_id = p.product_id
JOIN
    dim_categories cat ON p.category_id = cat.category_id
GROUP BY
    cat.name
ORDER BY
    total_predicted_demand DESC, average_rating DESC;

```

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				45
Змн.	Арк.	№ докум.	Підпис	Дата		

Скрипт Products with the Largest Rating-Based Demand.

```
SELECT
    p.name AS product_name,
    AVG(r.rating) AS average_rating,
    SUM(ff.predicted_demand) AS total_predicted_demand
FROM
    fact_forecasts ff
JOIN
    dim_reviews r ON ff.product_id = r.product_id
JOIN
    dim_products p ON ff.product_id = p.product_id
GROUP BY
    p.name
ORDER BY
    total_predicted_demand DESC, average_rating DESC;
```

Скрипт Review Sentiment Breakdown.

```
SELECT
    r.sentiment AS sentiment_type,
    COUNT(r.review_id) AS total_reviews
FROM
    dim_reviews r
GROUP BY
    r.sentiment
ORDER BY
    total_reviews DESC;
```

Скрипт Top Customers by Total Reviews.

```
SELECT
    c.name AS customer_name,
    c.email,
    COUNT(r.review_id) AS total_reviews
FROM
    dim_reviews r
JOIN
    dim_customers c ON r.customer_id = c.customer_id
GROUP BY
    c.name, c.email
ORDER BY
    total_reviews DESC;
```

Скрипт Top Products by Average Rating.

```
SELECT
    p.name AS product_name,
    AVG(r.rating) AS average_rating,
    COUNT(r.review_id) AS total_reviews
FROM
    dim_reviews r
```

		Холодницький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		46

```

JOIN
    dim_products p ON r.product_id = p.product_id
GROUP BY
    p.name
ORDER BY
    average_rating DESC, total_reviews DESC;

```

Скрипт Total Predicted Demand by Category.

```

SELECT
    c.name AS category_name,
    SUM(ff.predicted_demand) AS total_predicted_demand
FROM
    fact_forecasts ff
JOIN
    dim_products p ON ff.product_id = p.product_id
JOIN
    dim_categories c ON p.category_id = c.category_id
GROUP BY
    c.name
ORDER BY
    total_predicted_demand DESC;

```

		Холодніцький В. І.			«Житомирська політехніка» 24.121.33.000	Арк.
		Сугоняк І. І.				
Змн.	Арк.	№ докум.	Підпис	Дата		47