

Міністерство освіти і науки України
Тернопільський національний педагогічний університет
імені Володимира Гнатюка

Фізико-математичний факультет
Кафедра інформатики та методики її навчання

Кваліфікаційна робота
МЕТОДИ ВІЗУАЛІЗАЦІЇ ВЕЛИКИХ ДАНИХ У БІЗНЕС АНАЛІТИЦІ

Спеціальність 122 Комп'ютерні науки
Освітня програма «Комп'ютерні науки»

Здобувача вищої освіти освітньо-
кваліфікаційного рівня «магістр»
Шовага Івана Михайловича

НАУКОВИЙ КЕРІВНИК:
кандидат педагогічних наук,
доцент Карабін Оксана Йосифівна

РЕЦЕНЗЕНТ:
доктор технічних наук, професор
кафедри комп'ютерних наук
Тернопільського національного
технічного університету
ім. Івана Пулюя
Литвиненко Ярослав
Володимирович

Тернопіль – 2025

АНОТАЦІЯ

Шоваг І. М. Методи візуалізації великих даних у бізнес аналітиці. Кваліфікаційна робота на здобуття освітнього ступеня «магістр» зі спеціальності 122 Комп'ютерні науки. ТНПУ ім. В. Гнатюка. Тернопіль, 2025. 61 с.

У кваліфікаційній роботі здійснено теоретичне обґрунтування основ застосування технологій візуалізації великих даних у системах підтримки прийняття рішень, розглянуто концепцію Big Data та специфіку її застосування в бізнес-аналітиці; охарактеризовано сучасні методи та інструментальні засоби візуального аналізу, проведено порівняння архітектурних підходів до обробки даних (Import vs DirectQuery); розроблено оптимізовану семантичну модель даних типу «Галактика» та реалізовано ETL-процеси для підготовки масивів інформації; описано процес створення динамічних метрик мовою DAX та налаштування інтерактивної взаємодії візуальних елементів; розроблено аналітичний дашборд у середовищі Power BI з використанням алгоритмів штучного інтелекту для аналізу ключових бізнес-показників.

Ключові слова: великі дані, бізнес-аналітика, візуалізація даних, Power BI, дашборд, модель даних.

ABSTRACT

Shovah I. M. Methods of Big Data visualization in business analytics. The qualification work for obtaining a master's degree in the specialty 122 Computer Science. Ternopil Volodymyr Hnatiuk National Pedagogical University. Ternopil, 2025. 61 p.

In the qualification work a theoretical justification of the basics of applying Big Data visualization technologies in decision support systems was carried out, the concept of Big Data and the specifics of its application in business analytics were analyzed; modern methods and tools of visual analysis were characterized, a comparison of architectural approaches to data processing (Import vs DirectQuery) was conducted; an optimized "Galaxy" type semantic data model was developed and ETL processes for information arrays preparation were implemented; the process of creating dynamic metrics using DAX language and setting up interactive interaction of visual elements was described; an analytical dashboard in the Power BI environment using artificial intelligence algorithms for analyzing key business indicators was developed.

Keywords: big data, business analytics, data visualization, Power BI, dashboard, data model.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	4
ВСТУП.....	5
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ЗАСТОСУВАННЯ	
ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ВІЗУАЛІЗАЦІЇ ВЕЛИКИХ	
ДАНИХ У БІЗНЕС-АНАЛІТИЦІ.....	8
1.1 Теоретичні аспекти застосування інформаційних технологій та візуалізації великих даних у бізнес-аналітиці.....	8
1.2 Бізнес-аналітика, її структурні етапи візуалізації великих даних	10
1.3 Особливості аналітичних інструментів для роботи з великими даними у бізнес-аналітиці	13
Висновки до першого розділу.....	16
РОЗДІЛ 2. МЕТОДИ ВІЗУАЛІЗАЦІЇ ВЕЛИКИХ ДАНИХ У БІЗНЕС-АНАЛІТИЦІ	18
2.1 Організація та застосування технологій візуалізації великих даних у бізнес-аналітиці	18
2.2 Методи візуалізації великих даних у бізнес-аналітиці	28
2.3 Типи візуалізацій та побудова інтерактивних звітів	35
Висновки до другого розділу	39
РОЗДІЛ 3. РОЗРОБКА МОДЕЛІ ТА ВІЗУАЛІЗАЦІЇ ВЕЛИКИХ ДАНИХ У БІЗНЕС-АНАЛІТИЦІ ЗАСОБАМИ POWER BI.....	41
3.1 Розробка моделі великих даних у бізнес-аналітиці.....	41
3.2 Етапи розробки візуалізації великих даних у бізнес-аналітиці засобами Power BI.....	45
3.3 Реалізація методів інтелектуального аналізу та розрахунок ключових показників ефективності.....	49
3.4 Розробка візуалізації великих даних у бізнес-аналітиці засобами Power BI.....	52
Висновки до третього розділу.....	57
ЗАГАЛЬНІ ВИСНОВКИ.....	60
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	62
ДОДАТКИ	66

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ЦТ – цифрові технології

BI – Бізнес-інтелект (Business Intelligence)

Big Data – Великі дані

CRM – Управління взаємовідносинами з клієнтами (Customer Relationship Management)

ERP – Планування ресурсів підприємства (Enterprise Resource Planning)

ETL – Вилучення, трансформація, завантаження (Extract, Transform, Load)

GCP – Google Cloud Platform

HDFS – Розподілена файлова система Hadoop (Hadoop Distributed File System)

IoT – Інтернет речей (Internet of Things)

KPI – Ключові показники ефективності (Key Performance Indicators)

NoSQL – Нереляційні бази даних (Not only SQL)

YARN – Ще один ресурсний менеджер (Yet Another Resource Negotiator)

AWS – Платформа хмарних обчислень (Amazon Web Services)

SSMS – середовище для керування, конфігурації та адміністрування всіх компонентів SQL Server (SQL Server Management Studio)

ВСТУП

Актуальність теми дослідження. У сучасному етапі розвитку світова економіка супроводжується експоненціальним зростанням обсягів даних, які генеруються у бізнес-середовищі. Інформаційні потоки охоплюють транзакції, поведінку клієнтів та ринкові показники, вимагаючи від менеджменту нових підходів до їх обробки. Здатність ефективно сприймати та аналізувати цю інформацію залишається обмеженою людським фактором. Компанії накопичують масиви даних, відомі як Big Data, але їхній потенціал часто залишається нереалізованим через відсутність дієвих інструментів інтерпретації.

Нормативно-правовим підґрунтям впровадження сучасних аналітичних систем є Закон України «Про хмарні послуги», який визначає засади обробки та зберігання великих обсягів даних із використанням хмарних обчислень. Важливу роль відіграє Розпорядження Кабінету Міністрів України «Про схвалення Концепції розвитку штучного інтелекту в Україні», що актуалізує використання інтелектуальних алгоритмів для автоматизації бізнес-процесів. Правові аспекти роботи з інформаційними масивами також регулюються Законами України «Про захист інформації в інформаційно-комунікаційних системах» та «Про основні засади забезпечення кібербезпеки України», які створюють фундамент для безпечної цифровізації економіки.

Проблематика обробки великих даних та бізнес-аналітики перебуває в центрі уваги наукової спільноти. Вагомий внесок у розвиток теоретичних та прикладних аспектів цієї галузі зробили вітчизняні та закордонні науковці. Зокрема, у працях В. Григорківа висвітлено основи бізнес-аналітики та моделювання економічних процесів. Дослідженню систем штучного інтелекту та методів прийняття рішень присвячені роботи В. Литвина та Н. Шаховської. Економічні аспекти впровадження технологій Big Data та інструментарій їх аналізу розкрито у наукових напрацюваннях І. Шевчука. Фундаментальні принципи побудови сховищ даних та багатовимірного

моделювання, що є основою сучасних BI-систем, обґрунтовано у класичних працях W. Inmon та R. Kimball.

Традиційні табличні звіти не відображають повною мірою прихованих закономірностей у масивах даних, що зумовлює необхідність застосування новітніх методів візуалізації. Розробка інтерактивних дашбордів та систем візуалізації є пріоритетним напрямком розвитку інформаційних технологій, спрямованих на підтримку прийняття управлінських рішень. Ефективне управління сучасним бізнесом потребує впровадження інструментів Business Intelligence, які забезпечують миттєвий доступ до аналітичних інсайтів. В інформаційній сфері бізнесу наявна суперечність між існуванням колосальних обсягів даних та недостатньою ефективністю існуючих підходів до їх подання. Вирішення цього протиріччя можливе шляхом системного використання спеціалізованих BI-систем, зокрема Power BI, що поєднують класичну звітність з елементами штучного інтелекту. Необхідність створення адаптивних моделей візуалізації, здатних працювати з великими даними, зумовлює вибір теми дослідження **«Методи візуалізації великих даних у бізнес аналітиці»**.

Мета дослідження – теоретично обґрунтувати основи візуалізації великих даних та розробити візуалізацію великих даних у бізнес-аналітиці.

Відповідно до мети визначено такі **завдання дослідження**:

1. Теоретично обґрунтувати основи застосування інформаційних технологій та візуалізації великих даних у бізнес-аналітиці.
2. Охарактеризувати та конкретизувати методи візуалізації великих даних у бізнес-аналітиці.
3. Розробити модель великих даних в бізнес-процесах.
4. Розробити візуалізацію великих даних у бізнес-аналітиці засобами Power BI.

Об’єкт дослідження – візуалізація великих даних у бізнес-аналітиці.

Предмет дослідження – технології розробки візуалізації великих даних у бізнес-аналітиці.

Для досягнення мети та реалізації поставлених завдань використано комплекс взаємопов'язаних методів дослідження: теоретичні (аналіз наукової літератури та документації, порівняння функціональних можливостей BI-платформ, системний аналіз); емпіричні (моделювання схеми даних, прототипування дашборду, експериментальна перевірка ефективності візуалізацій).

Наукова новизна одержаних результатів полягає у теоретичному обґрунтуванні та практичній реалізації підходів до візуалізації великих даних, а саме: удосконалено методику побудови семантичних моделей даних для бізнес-аналітики шляхом застосування схеми «Галактика» та оптимізації запитів, що забезпечує масштабованість системи; дістало подальшого розвитку застосування інструментів інтелектуального аналізу даних у Business Intelligence через інтеграцію вбудованих AI-алгоритмів у візуальні звіти; набули подальшого розвитку методи візуалізації великих даних за рахунок систематизації принципів побудови інтерактивних дашбордів, орієнтованих на кінцевого користувача.

Практичне значення одержаних результатів полягає в систематизації теоретичного матеріалу щодо методів роботи з Big Data та у розробці повнофункціональної аналітичної системи на платформі Power BI. Розроблені візуалізації та модель даних можуть бути впроваджені в практику роботи підприємств для аналізу продажів, а також використані як навчально-методичний матеріал в освітньому процесі закладів освіти.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ОСНОВИ ЗАСТОСУВАННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ВІЗУАЛІЗАЦІЇ ВЕЛИКИХ ДАНИХ У БІЗНЕС- АНАЛІТИЦІ

1.1 Теоретичні аспекти застосування інформаційних технологій та візуалізації великих даних у бізнес-аналітиці

Фундаментальні трансформації у світовій економіці останнього десятиліття тісно пов'язані з переходом до парадигми Індустрії 4.0, де дані стають ключовим активом підприємств, порівнюваним за значущістю з капіталом та людськими ресурсами. Стрімка цифровізація бізнес-процесів, повсюдне впровадження мобільних технологій та інтернету речей призвели до виникнення феномену великих даних. Згаданий термін вийшов за межі суто технологічного визначення і став об'єктом глибокого дослідження в економіці, менеджменті та комп'ютерних науках, оскільки змінює саму природу конкуренції на ринках. Компанії, здатні ефективно акумулювати та інтерпретувати інформаційні потоки, отримують стратегічні переваги, тоді як ігнорування наявного ресурсу призводить до втрати ринкових позицій та зниження ефективності операційної діяльності [3; 9].

У сучасних науково-технічних джерелах науковці під поняттям великі дані розуміють не лише масиви інформації екстремальних обсягів, які неможливо обробити вручну, але й специфічні технології їх зберігання, обробки та аналізу [2; 5; 6; 13; 19; 21]. Застосування новітніх технологій дозволяє отримувати якісно нові результати, недосяжні при використанні традиційних систем керування базами даних. Аналіз джерел свідчить, що еволюція поняття пройшла складний шлях від класичної моделі «3V», запропонованої аналітиками на початку 2000-х років, до розширених концепцій «5V», «7V» та навіть «10V». Розширення переліку характеристик

підкреслює багатогранність явища та його вплив на всі сфери діяльності сучасного підприємства [15].

Критичний аналіз наукових підходів дозволяє виділити та деталізувати характеристики, що є визначальними для побудови ефективних систем бізнес-аналітики:

– *обсяг* – накопичення та експоненціальне зростання терабайтів, петабайтів і навіть зеттабайтів даних. Джерелами масивів виступають транзакційні системи підприємств, логи веб-серверів, дані з сенсорів та активність у соціальних мережах. Традиційні реляційні сховища даних часто виявляються економічно неефективними та технічно неспроможними для зберігання та індексації наявних масштабів інформації, що стимулює перехід до розподілених файлових систем та хмарних рішень [24];

– *швидкість* – відображає інтенсивність надходження даних та необхідність їх оперативної обробки. Вимога до аналізу в режимі реального часу або близькому до нього стає критичною для галузей ритейлу, логістики та фінансового сектору. Затримка в аналізі навіть на хвилини може призвести до втрати актуальності інформації та, як наслідок, до прямих фінансових збитків через несвоєчасне прийняття рішень [19];

– *різноманітність* – необхідність обробки різноманітних даних. Окрім структурованих табличних даних, сучасні аналітичні системи повинні обробляти напівструктуровані формати та неструктурований контент, такий як текст електронних листів, зображення та відеопотоки. Наявна різноманітність вимагає застосування нових підходів до моделювання даних, відмінних від жорстких схем класичних реляційних баз, та використання гнучких механізмів інтеграції [21].

У контексті дослідження найважливішими є додаткові характеристики, які трансформують «сирі» дані в реальний бізнес-інструмент:

– *достовірність* – вказує на проблему надійності та якості даних. В умовах роботи з великими масивами інформація часто містить шуми, пропуски, дублікати або є неповною. Робота з неякісними даними може

призвести до помилкових управлінських висновків, тому процеси автоматизованого очищення та валідації стають критично важливими етапами побудови аналітичної системи;

– *цінність* – визначає економічну доцільність впровадження аналітичних систем. Накопичені дані самі по собі є пасивом, що потребує значних витрат на інфраструктуру зберігання та захист. Активом інформація стає лише після вилучення з неї цінності через інструменти глибокої аналітики та візуалізації (рис. 1.1), що дозволяє оптимізувати бізнес-процеси та знаходити нові точки росту [5].

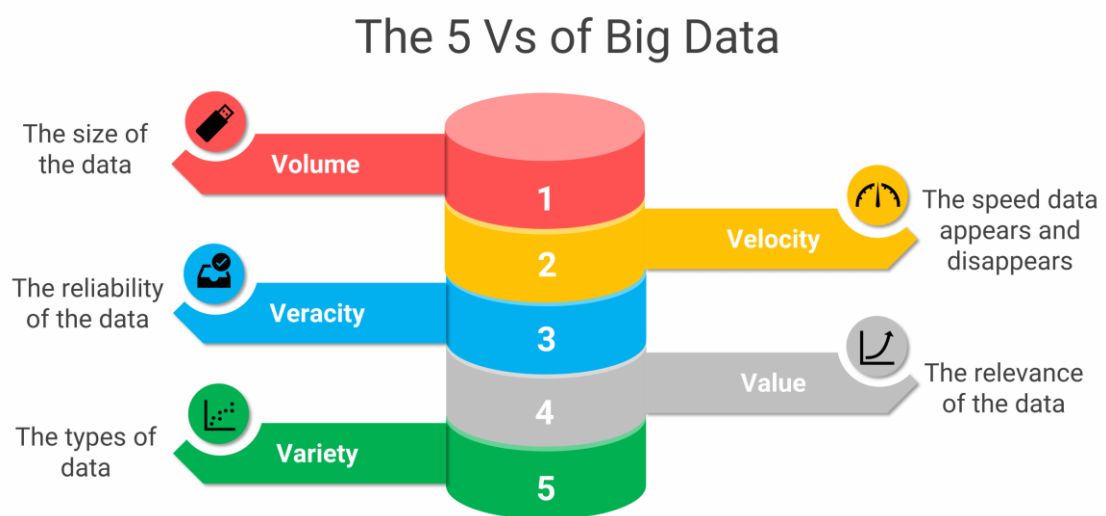


Рис. 1.1. Характеристика моделі «5V» великих даних

Таким чином, значення великих даних у сучасному бізнесі полягає не в простому накопиченні інформації «про запас», а у створенні надійного технологічного підґрунтя для переходу від інтуїтивного менеджменту до управління, що базується на фактах. Описаний підхід дозволяє мінімізувати ризики та підвищити точність стратегічного планування.

1.2 Бізнес-аналітика, її структурні етапи візуалізації великих даних

Ефективне використання потенціалу великих даних неможливе без застосування методології бізнес-аналітики. У науково-технічних працях поняття визначається як сукупність технологій, практик, методологій та навичок для ітеративного дослідження минулих бізнес-показників з метою

отримання інсайтів та стратегічного планування майбутньої діяльності. Зазначений процес є комплексним перетворенням даних на знання, а знань – на конкретні управлінські дії, що підвищують прибутковість компанії [5; 11; 13; 14]. Варто чітко розмежовувати поняття Business Intelligence та Business Analytics, хоча на практиці вони часто використовуються як синоніми. Згідно з дослідженнями провідних консалтингових агенцій, Business Intelligence традиційно фокусується на описовому аналізі, відповідаючи на питання «що сталося?» і «чому подія відбулася?», та надаючи звітність про поточний стан справ. Натомість просунута аналітика орієнтована на майбутнє: прогнозування та оптимізацію. Однак у сучасних конвергентних платформах, таких як Power BI, межі між напрямками поступово стираються, утворюючи єдину екосистему, що покриває весь спектр аналітичних завдань від моніторингу до прогнозування [25].

Процес бізнес-аналітики структурується за моделлю ієрархії цінності, де кожен наступний етап додає більше користі для бізнесу, але вимагає застосування складніших технологій та алгоритмів [8]:

– *описова аналітика* – базовий рівень, що відповідає за агрегацію та візуалізацію історичних даних. Інструментами тут виступають стандартні звіти, панелі ключових показників ефективності та дашборди, які дають відповідь на питання про минулі події. Описаний рівень виступає фундаментом, на якому будуються всі подальші висновки, оскільки без якісного розуміння історії неможливо будувати адекватні прогнози;

– *діагностична аналітика* – передбачає поглиблений пошук причинно-наслідкових зв'язків. На даному етапі використовуються інструменти заглиблення в дані, аналіз кореляцій та пошук аномалій, щоб зрозуміти фундаментальні причини успіхів або невдач. Аналіз дозволяє виявити фактори, що найбільше впливають на результативність бізнесу, та відокремити системні тренди від випадкових відхилень;

– *предиктивна аналітика* – використовує статистичні моделі, регресійний аналіз та алгоритми машинного навчання для виявлення

прихованих патернів і прогнозування майбутніх трендів з певною ймовірністю. Вона відповідає на питання про ймовірність настання тих чи інших подій, що дозволяє бізнесу завчасно підготуватися до змін ринкової кон'юнктури [11];

– *прескриптивна аналітика* – найвищий рівень, що генерує рекомендації щодо оптимальних дій. Вона використовує методи імітаційного моделювання та складні оптимізаційні алгоритми для вибору найкращого сценарію розвитку подій серед багатьох можливих альтернатив. Методика дозволяє не просто передбачити майбутнє, а й активно впливати на нього, обираючи оптимальну стратегію (рис. 1.2).

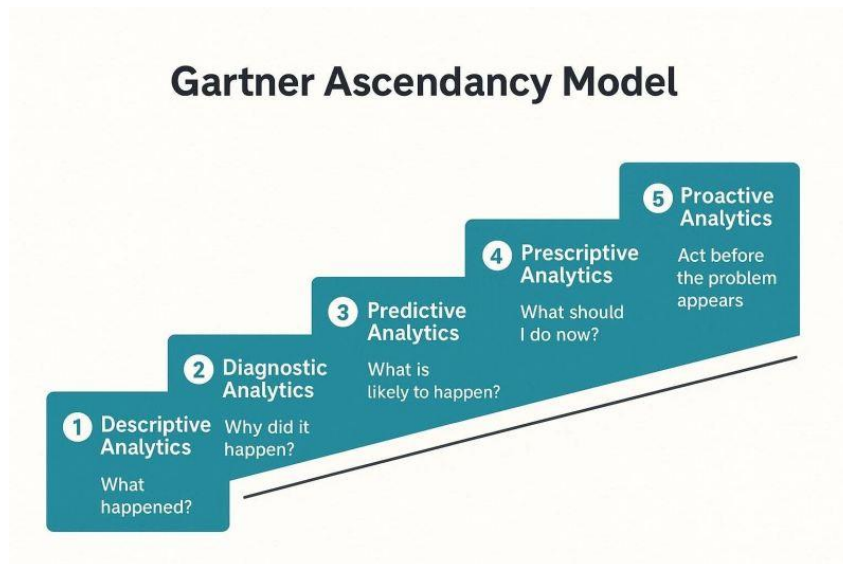


Рис. 1.2. Модель зрілості бізнес-аналітики

Аналіз ринку аналітичних платформ демонструє стійке лідерство таких інструментів як Microsoft Power BI, Tableau та Qlik. Їхньою спільною рисою є глобальний перехід від моделі, орієнтованої на IT-фахівців, до моделі самообслуговування. Зазначена парадигма дозволяє бізнес-користувачам та аналітикам без глибоких знань програмування самостійно підключатися до даних, моделювати їх та будувати візуалізації, що значно прискорює цикл прийняття рішень та зменшує залежність бізнесу від технічних департаментів [13; 21]. Упровадження бізнес-аналітики неминуче стикається з фундаментальною проблемою якості та інтеграції даних. Приклад програмної реалізації процесів трансформації даних наведено у додатку А.

Таким чином, за статистикою, значна частина часу будь-якого аналітичного проєкту витрачається не на побудову графіків чи моделей, а на процеси вилучення, трансформації та завантаження даних. Вважаємо, що належної попередньої обробки навіть найдосконаліші алгоритми візуалізації будуть відображати викривлену картину реальності.

1.3 Особливості аналітичних інструментів для роботи з великими даними у бізнес-аналітиці

В умовах інформаційного перевантаження, коли менеджери змушені оперувати терабайтами інформації, візуалізація даних трансформується з допоміжного засобу ілюстрації звітів у ключовий когнітивний інструмент аналізу. Згідно з дослідженнями у сфері когнітивної психології, візуалізація використовує особливості людського зорового сприйняття для миттєвого розпізнавання патернів, аномалій, кластерів та трендів, які практично неможливо виявити при перегляді рядків у табличних звітах [6; 30]. Графічне представлення дозволяє «стиснути» величезні масиви даних до зрозумілих образів, значно прискорюючи процес сприйняття інформації.

Теоретичні основи ефективної візуалізації базуються на принципах графічної досконалості, ясності та мінімізації «візуального шуму». Головна мета візуалізації в бізнесі – не естетична привабливість, а швидкість та точність передачі інформації. Графік повинен відповідати на конкретне бізнес-питання без необхідності додаткових пояснень чи коментарів. Кожен елемент на екрані має нести смислове навантаження, а декоративні елементи, що не додають інформаційної цінності, повинні бути виключені [28].

У контексті великих даних виникає специфічна проблема: спроба відобразити мільйони точок даних на одному екрані призводить до ефекту перенасичення, коли інформація зливається в нечитабельний шум. Для вирішення завдання в сучасній аналітиці застосовують методи інтелектуальної агрегації, семантичного зумування та інтерактивної деталізації.

Можна виділити категорії візуалізацій, що є найбільш релевантними для завдань бізнес-аналітики [33]:

- *порівняльні* (стовпчасті діаграми, гістограми, лінійчаті діаграми) – використовуються для ранжування об'єктів та аналізу категорій, дозволяючи швидко виявити лідерів та аутсайдерів;

- *часові* (лінійні графіки, діаграми областей, каскадні діаграми) – необхідні для виявлення сезонності, циклічності та довгострокових трендів у часових рядах;

- *кореляційні* (діаграми розсіювання, бульбашкові діаграми) – застосовуються для пошуку статистичних залежностей між двома або більше змінними, що допомагає перевіряти гіпотези;

- *геопросторові* (карти, хороплети) – критично важливі для аналізу регіональних показників та логістичних потоків, дозволяючи бачити просторовий розподіл даних;

- *композиційні* (дерезовидні карти, кругові діаграми) – використовуються для аналізу структури цілого та визначення внеску окремих частин у загальний результат.

Особливого значення набувають комплексні інтерактивні панелі – дашборди. На відміну від статичних звітів, дашборди забезпечують моніторинг ключових показників ефективності в реальному часі та дозволяють користувачеві взаємодіяти з даними через систему фільтрів, зрізів та перехресного виділення [7]. Успіх дашборду залежить не лише від вдалого вибору типів діаграм, але й від логіки побудови семантичної моделі даних, що лежить в його основі та забезпечує швидкий відгук інтерфейсу [10]. Технологічний стек для роботи з великими даними та їх візуалізації являє собою складну екосистему, яку можна умовно розділити на три взаємопов'язані рівні: зберігання, обробка та представлення.

На рівні зберігання відбувся парадигмальний зсув від монолітних реляційних баз даних до розподілених файлових систем та гнучких нереляційних рішень (NoSQL). Технології екосистеми Hadoop та концепція

«Озер даних» на хмарних платформах дозволяють економічно ефективно зберігати величезні масиви неструктурованих даних у їх первісному форматі [3; 17]. Проте для завдань швидкої аналітичної звітності структуровані сховища та сучасні колоночні бази даних залишаються стандартом завдяки високій швидкості виконання запитів на агрегацію [26].

На рівні обробки ключову роль відіграють технології обробки в оперативній пам'яті. Зокрема, фреймворк Apache Spark став стандартом індустрії, оскільки дозволяє виконувати складні ітеративні обчислення та алгоритми машинного навчання безпосередньо в оперативній пам'яті кластера. Технологія забезпечує значне прискорення аналізу порівняно з традиційним дисковим підходом, що є критичним для забезпечення інтерактивності аналітичних систем [19].

На рівні візуалізації сучасні платформи, насамперед Microsoft Power BI, виступають конвергентною точкою, що об'єднує дані з різномірних джерел. Важливою технологічною особливістю платформи є вбудований аналітичний двигун VertiPac, який використовує стовпчиковий метод зберігання даних, словникове кодування та агресивні алгоритми стиснення в оперативній пам'яті. Зазначений підхід дозволяє досягати високої продуктивності відгуку візуалізацій навіть на масивах у сотні мільйонів рядків [23; 26].

Окрім того, для роботи з надвеликими даними, які фізично не вміщуються в пам'ять, реалізовано підтримку технології DirectQuery та композитних моделей. Механізм дозволяє візуалізувати дані, залишаючи їх у джерелі, або комбінувати «гарячі» кешовані дані з «холодними» архівними даними в одному звіті [20; 18].

Таким чином, аналіз технологічного ландшафту дозволяє стверджувати, що сучасні засоби візуалізації є невід'ємною частиною екосистеми великих даних, а їх ефективність безпосередньо залежить від тісної інтеграції з технологіями розподіленого зберігання та попередньої обробки даних. Тільки комплексне поєднання потужних алгоритмів обробки, оптимізованих моделей даних та ергономічних інтерактивних інтерфейсів

дозволяє подолати виклики, пов'язані з обсягом та різноманітністю інформації. Синергетичний підхід забезпечує трансформацію сирих даних у реальну бізнес-цінність, створюючи надійну основу для прийняття обґрунтованих управлінських рішень та подальшої практичної реалізації аналітичної системи.

Висновки до першого розділу

У першому розділі кваліфікаційної роботи здійснено комплексний та системний аналіз теоретико-методологічних засад застосування сучасних інформаційних технологій та методів візуалізації великих даних у сфері бізнес-аналітики. На основі опрацювання науково-технічних джерел та аналітичних звітів сформульовано наступні узагальнення:

1. Встановлено, що феномен великих даних трансформувався у ключовий фактор стратегічної конкурентоспроможності підприємств. Критичний аналіз характеристик великих даних дозволив розширити класичну модель критично важливими для бізнесу вимірами – Достовірністю та Цінністю. З'ясовано, що традиційні підходи до управління даними втрачають ефективність, що зумовлює необхідність переходу до гібридних архітектур, які поєднують гнучкість «Озер даних» та швидкість корпоративних сховищ.

2. Проаналізовано сутність бізнес-аналітики як ітеративного процесу дослідження даних. Визначено, що сучасна аналітика еволюціонує від описової до діагностичної, предиктивної та прескриптивної. Обґрунтовано критичну важливість інструментів ETL у складі сучасних ВІ-платформ для трансформації розрізнених масивів інформації у узгоджені семантичні моделі, без яких неможлива якісна візуалізація.

3. Досліджено психофізіологічні основи сприйняття візуальної інформації та доведено, що графічна візуалізація є ефективним способом подолання когнітивного бар'єру при роботі з великими даними. Систематизовано методи візуалізації та визначено специфіку їх застосування

для різних типів аналітичних завдань. Відзначено трансформацію ролі звітності: перехід від статичних звітів до інтерактивних дашбордів, які забезпечують можливість сценарного аналізу.

4. Виявлено, що платформи класу Self-Service BI, зокрема Microsoft Power BI, стають стандартом для сучасного бізнесу. Їх ключовою перевагою у контексті Big Data є використання технологій In-Memory та підтримка режимів DirectQuery, що дозволяє візуалізувати дані безпосередньо з джерела без їх повного завантаження в оперативну пам'ять.

Узагальнюючи результати теоретичного дослідження, можна стверджувати, що на сьогодні існує об'єктивна потреба у розробці та впровадженні комплексних методик візуалізації, які б органічно поєднували потужність сучасних технологій обробки великих даних з принципами ефективного візуального дизайну та сторітелінгу. Теоретичний аналіз підтвердив, що обраний технологічний стек та методологічні підходи є релевантними та достатніми для вирішення поставлених завдань, що створює необхідне наукове підґрунтя для переходу до наступних етапів роботи, присвячених практичній реалізації візуалізації великих даних у бізнес-аналітиці засобами Power BI.

РОЗДІЛ 2

МЕТОДИ ВІЗУАЛІЗАЦІЇ ВЕЛИКИХ ДАНИХ У БІЗНЕС-АНАЛІТИЦІ

2.1 Організація та застосування технологій візуалізації великих даних у бізнес-аналітиці

Ефективність функціонування сучасних систем бізнес-аналітики визначається їхньою архітектурною здатністю забезпечувати безперебійний процес обробки інформації на всіх етапах її життєвого циклу: від підключення до різномірних джерел даних та їх трансформації у придатний для аналізу формат до створення оптимізованих логічних моделей, які слугують фундаментом для генерації точних та оперативних аналітичних висновків. Платформа Power BI демонструє виняткові можливості на кожному з етапів, пропонуючи гнучкий та масштабований підхід до побудови аналітичних систем корпоративного рівня [7; 25].

Одним із визначальних компонентів архітектури аналітичної платформи є широкий спектр підтримуваних конекторів та механізмів інтеграції. Універсальність інструментарію дозволяє користувачам імпортувати дані з локальних файлових систем, хмарних онлайн-сховищ або встановлювати прямі тунельні підключення до корпоративних баз даних та веб-сервісів.

Середовище розробки пропонує різноманітний набір нативних конекторів, що значно спрощує процес інтеграції даних з різних технологічних платформ, нівелюючи необхідність розробки складних проміжних шлюзів. Система здатна підключатися до сотень різномірних джерел, забезпечуючи формування уніфікованого погляду на бізнес-операції підприємства [19].

В умовах роботи з великими даними, які за своєю природою часто зберігаються в розрізнених джерелах та мають різну структуру, здатність

системи до безперешкодної інтеграції інформації набуває критичного значення.

Зазначений підхід дозволяє організаціям долати проблему фрагментації даних, коли інформація про різні аспекти діяльності компанії ізольована в окремих системах, та будувати концепцію «єдиного джерела істини».

Інтеграція є необхідним фундаментом для проведення комплексного аналізу, що охоплює всі сфери діяльності підприємства та дозволяє виявляти приховані залежності між різними бізнес-процесами.

Спектр підтримуваних джерел даних охоплює практично всі сучасні технології зберігання інформації. Серед файлових сховищ підтримуються як традиційні формати електронних таблиць та текстові файли з розділювачами, так і формати, оптимізовані для зберігання великих даних, зокрема JSON та Parquet, а також підключення до корпоративних файлових серверів.

У сегменті баз даних забезпечується інтеграція як з класичними реляційними системами, так і з високопродуктивними аналітичними платформами, орієнтованими на стовпчикове зберігання. Окрему групу становлять конектори до хмарних сервісів та платформ, що дозволяє отримувати дані безпосередньо з CRM та ERP систем.

Критично важливим архітектурним рішенням, яке безпосередньо впливає на продуктивність візуалізації та швидкість обробки запитів при роботі з великими даними, є вибір стратегії зберігання та доступу до даних.

Платформа пропонує кілька базових режимів роботи, кожен з яких базується на різних принципах взаємодії з даними та має специфічні сценарії застосування залежно від обсягу інформації, вимог до актуальності даних та доступних обчислювальних ресурсів.

Для кращого розуміння цих компромісів, у таблиці 2.1 представлено порівняння режимів підключення Power BI для великих даних.

Таблиця 2.1

Порівняння режимів підключення Power BI для великих даних

Режим підключення	Принцип роботи	Переваги для великих даних	Недоліки/Обмеження	Ідеальні сценарії використання
Import Mode	Дані повністю завантажуються в пам'ять Power BI (VertiPaq)	Найвища швидкість запитів, повний функціонал Power Query та DAX, ефективне стиснення даних	Споживання пам'яті, дані не в реальному часі, обмеження розміру моделі (до 12 ГБ для Premium)	Для даних, що не потребують миттєвого оновлення, або для агрегованих даних, що поміщаються в пам'ять
DirectQuery Mode	Запити надсилаються безпосередньо до вихідного джерела даних у реальному часі. Модель зберігає лише метадані	Дані в реальному часі, немає обмежень на розмір моделі Power BI	Продуктивність залежить від швидкості джерела, обмеження функціоналу Power Query та DAX, високе навантаження на джерело	Для дуже великих наборів даних, які неможливо імпортувати, або коли актуальність даних є критичною
Dual Mode	Гібридний режим: таблиця кешується в пам'яті, але може поширювати фільтри на таблиці DirectQuery	Поєднання швидкості імпорту та гнучкості DirectQuery, покращення продуктивності композитних моделей	Потребує ретельного налаштування, потенційні проблеми з коректністю даних, якщо використовується необережно	Для таблиць вимірів у змішаних моделях (наприклад, "Дата", "Продукт"), де факт-таблиця знаходиться в DirectQuery
Direct Lake Storage Mode	Завантаження великих обсягів даних з Delta-таблиць у	Продуктивність запитів, порівнянна з Import, швидкі	Автоматичний перехід у DirectQuery для певних функцій	Для семантичних моделей, що підключаються до великих Fabric

	OneLake пам'ять потребою (транскодування)	в за Fabric lakehouses/wareh ouses	оновлення (копіювання лише метаданих), ідеально для лакехаусів/складів	(наприклад, RLS), обмеження на композитне моделювання та обчислювані стовпці	lakehouses/warehous es, де потрібна висока продуктивність без повного імпорту
--	--	--	---	--	---

Режим імпорту базується на концепції завантаження повної копії даних у внутрішню пам'ять аналітичної системи. При використанні підходу дані витягуються з джерела, проходять етап трансформації та завантажуються в оперативну пам'ять, де обробляються високопродуктивним колоночним рушієм. Рушій використовує передові алгоритми стиснення даних, такі як кодування значень, словникове кодування та кодування довжин серій, що дозволяє суттєво зменшити обсяг займаної пам'яті порівняно з вихідним розміром даних [21]. Завдяки зберіганню даних в оперативній пам'яті, режим забезпечує найвищу швидкість виконання аналітичних запитів та миттєвий відгук візуальних елементів звіту. Додатково він надає повний доступ до всього функціоналу мови формул, включаючи складні функції аналізу часу та роботи з ієрархіями. Використання режиму імпорту є оптимальним сценарієм для більшості аналітичних завдань, де обсяг даних дозволяє розмістити їх у доступній оперативній пам'яті, а вимоги до актуальності інформації допускають періодичне оновлення за розкладом. Однак при роботі з надвеликими масивами даних підхід може зіткнутися з апаратними обмеженнями, оскільки розмір моделі лімітований обсягом виділеної оперативної пам'яті. Також слід враховувати час, необхідний для повного оновлення даних, який може бути значним при великих обсягах інформації.

Альтернативним архітектурним підходом є режим прямого запиту, який реалізує концепцію роботи з даними без їх фізичного переміщення в аналітичну систему. У зазначеному режимі дані залишаються у вихідному джерелі, а аналітична система зберігає лише метадані про структуру таблиць

та зв'язки між ними. При кожній взаємодії користувача зі звітом (фільтрація, деталізація, зміна вимірів) система генерує запит, який автоматично транслюється у відповідний нативний запит до джерела даних [29]. Метод дозволяє працювати з масивами даних практично необмеженого обсягу, оскільки ліміти визначаються лише потужністю та обсягом сховища самого джерела даних, а не ресурсами аналітичної станції. Крім того, режим прямого запиту забезпечує доступ до актуальної інформації в реальному часі, оскільки кожен запит відображає поточний стан бази даних. Можливість є критично важливою для завдань операційного моніторингу та ситуаційних центрів. Водночас слід враховувати, що продуктивність візуалізацій у цьому режимі безпосередньо залежить від швидкодії джерела даних та пропускну здатності мережі. Деякі складні функції та трансформації можуть бути недоступними або мати обмежену функціональність через неможливість їх трансляції у запит до джерела.

Для вирішення компромісу між швидкістю роботи та обсягом даних застосовується гібридний режим, який дозволяє комбінувати переваги імпорту та прямого запиту в межах однієї семантичної моделі. Архітектура композитних моделей дозволяє частину таблиць (наприклад, довідники та агреговані дані) зберігати в оперативній пам'яті для швидкого доступу, тоді як таблиці фактів з деталізованими транзакціями залишаються у джерелі та опитуються через прямі запити. Гібридний підхід дозволяє досягти балансу між високою продуктивністю для загальних звітів та можливістю глибокого аналізу детальних даних без перевантаження пам'яті системи [20; 18].

Невід'ємною та критично важливою складовою архітектури будь-якої аналітичної системи є підсистема трансформації та підготовки даних. У середовищі Power BI функцію виконує компонент Power Query, який виступає потужним рушієм для вилучення, очищення та трансформації даних перед їх завантаженням у модель. Згідно з галузевою статистикою, аналітики даних витрачають значну частину робочого часу саме на процеси підготовки інформації, що робить наявність ефективного інструментарію ключовим

фактором успіху аналітичного проєкту [2]. Функціональне ядро підсистеми трансформації базується на спеціалізованій мові формул, яка є функціональною мовою програмування, чутливою до регістру, призначеною для маніпулювання даними. Кожна дія користувача у графічному інтерфейсі редактора запитів автоматично транслюється у відповідний скрипт. Фрагмент лістингу коду попередньої обробки даних мовою М представлено у додатку А. Механізм забезпечує повну відтворюваність процесів обробки даних: визначена послідовність кроків (підключення, фільтрація, заміна значень, об'єднання) зберігається як запит і автоматично виконується при кожному оновленні даних. Описаний підхід дозволяє автоматизувати рутинні операції з підготовки звітності та мінімізувати ризик людських помилок. Важливою технологічною особливістю при роботі з великими даними є механізм згортання запитів. Технологія дозволяє транслювати кроки трансформації даних, визначені у редакторі запитів, у єдиний нативний запит, який виконується на стороні джерела даних. Функція дозволяє перенести обчислювальне навантаження з локальної машини аналітика на потужний сервер бази даних, що є критично важливим для продуктивності при обробці мільйонів рядків. Використання згортання запитів є однією з найкращих практик оптимізації процесів підготовки даних у корпоративних середовищах [30].

Інструментарій трансформації дозволяє ефективно вирішувати проблеми, пов'язані з якістю та структурою вхідних даних. Вбудовані засоби профілювання даних надають можливість аналізувати розподіл значень, виявляти помилки, порожні значення та аномалії на ранніх етапах розробки. Функції очищення дозволяють стандартизувати формати, видаляти недруковані символи, заповнювати пропуски та виправляти помилки типізації. Окрім базових операцій, система підтримує складні сценарії трансформації, необхідні для формування аналітичних моделей. Операції злиття дозволяють збагачувати дані шляхом приєднання інформації з інших таблиць за ключовими полями, реалізуючи логіку реляційного об'єднання.

Операції додавання використовуються для консолідації даних з однотипних джерел, наприклад, об'єднання файлів продажів за різні місяці. Функції зведення та розведення дозволяють змінювати структуру таблиць, перетворюючи їх з широкого формату у довгий і навпаки, що часто є необхідним для коректної візуалізації даних.

Наступним фундаментальним етапом після фізичної підготовки даних є процес моделювання. Семантичне моделювання перетворює набір розрізнених таблиць на цілісну логічну структуру, яка описує бізнес-предметну область та слугує основою для побудови звітів. Якісно спроектована модель даних є запорукою високої швидкодії звітів, коректності розрахунків та зручності для кінцевих користувачів. Процес моделювання в аналітичних системах суттєво відрізняється від проектування транзакційних баз даних. Якщо в транзакційних системах пріоритетом є нормалізація даних для забезпечення швидкості запису та економії дискового простору, то в аналітичних системах ключовим є оптимізація швидкості читання та агрегації великих обсягів даних. Оптимізація досягається шляхом використання денормалізованих схем, серед яких найбільш поширеними є схеми «Зірка» та «Сніжинка» [26].

Схема «Зірка» є золотим стандартом моделювання для аналітичних систем. Структура передбачає чіткий поділ таблиць на два типи: таблиці фактів та таблиці вимірів. Таблиці фактів містять числові показники (метрики) бізнес-процесів (наприклад, сума продажу, кількість одиниць, собівартість) та зовнішні ключі для зв'язку з вимірами. Вони зазвичай є вузькими, але дуже довгими, містять мільйони або мільярди рядків транзакцій. Таблиці вимірів містять описові атрибути (наприклад, назва товару, категорія, ім'я клієнта, адреса магазину), за якими здійснюється фільтрація та групування даних. Використання схеми «Зірка» забезпечує максимальну продуктивність аналітичного рушія, оскільки дозволяє виконувати фільтрацію та агрегацію за мінімальну кількість операцій з'єднання. Крім того, структура є інтуїтивно зрозумілою для бізнес-

користувачів, що спрощує створення звітів у режимі самообслуговування. Денормалізація вимірів (рис. 2.1) дозволяє уникнути зайвих з'єднань під час виконання запиту, що особливо важливо при роботі з великими обсягами даних, де кожна додаткова операція може суттєво вплинути на час відгуку [26].

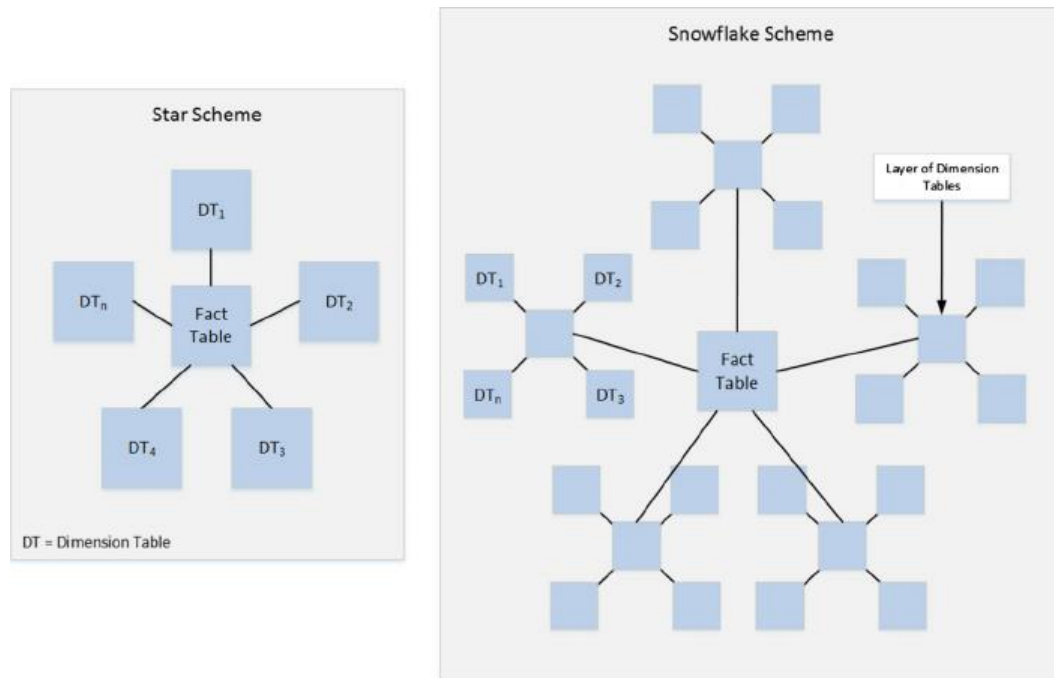


Рис. 2.1. Структурні відмінності схем моделювання даних «Зірка» та «Сніжинка»

У більш складних сценаріях, коли аналітична система охоплює кілька бізнес-процесів, застосовується схема «Галактика», або схема сузір'я фактів. Модель передбачає наявність кількох таблиць фактів (наприклад, факти продажів та факти закупівель), які використовують спільні (конформні) виміри (наприклад, вимір часу та вимір продукту). Приклад практичної реалізації схеми моделі даних типу «Галактика» наведено у додатку В. Зазначений підхід (рис. 2.2) дозволяє проводити перехресний аналіз різних аспектів діяльності компанії в межах єдиного інформаційного простору, забезпечуючи цілісність аналітики та можливість порівняння показників різних департаментів [21].

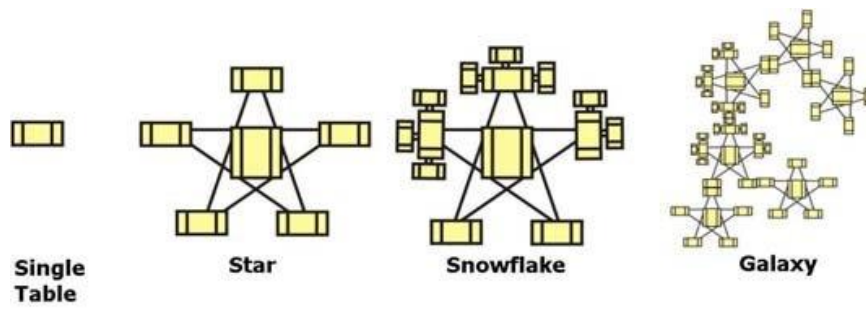


Рис. 2.2. Структурні відмінності схем моделювання даних «Сузір'я» та інших схем

Критично важливим елементом моделювання є налаштування зв'язків між таблицями. Зв'язки визначають, як фільтри, застосовані до однієї таблиці, поширюються на інші таблиці моделі. Основні типи зв'язків включають «один-до-багатьох», «один-до-одного» та «багато-до-багатьох». Найбільш ефективним та рекомендованим є зв'язок «один-до-багатьох», коли один запис у таблиці вимірів відповідає багатьом записам у таблиці фактів. Зазначений тип зв'язку забезпечує однозначність фільтрації та високу продуктивність. Напрямок крос-фільтрації визначає вектор поширення фільтрів через зв'язок. Однонаправлена фільтрація є базовим та найбільш продуктивним варіантом, коли фільтри поширюються від вимірів до фактів. Двонаправлена фільтрація дозволяє фільтрам поширюватися в обох напрямках, що забезпечує гнучкість у складних сценаріях аналізу, але може негативно впливати на продуктивність та призводити до неоднозначностей у схемі даних. При роботі з великими даними рекомендується мінімізувати використання двонаправленої фільтрації для забезпечення стабільності та швидкодії моделі.

Функціональним інструментом реалізації бізнес-логіки та аналітичних розрахунків у моделі виступає мова DAX. Інструмент являє собою бібліотеку функцій та операторів, які можуть бути використані у формулах для створення виразів аналізу даних. На відміну від формул електронних таблиць, які оперують комірками та діапазонами, мова оперує цілими таблицями та стовпцями, що робить її ідеальним інструментом для роботи з реляційними моделями. DAX дозволяє створювати три типи обчислень:

обчислювані стовпці, обчислювані таблиці та міри [22]. Обчислювані стовпці розраховуються під час оновлення даних та зберігаються у моделі, займаючи оперативну пам'ять. Елементи корисні для створення нових атрибутів фільтрації, але їх надмірне використання може призвести до роздування розміру моделі. Міри є динамічними формулами, які обчислюються «на льоту» у момент виконання запиту візуалізації. Вони не займають місця в пам'яті для зберігання значень і реагують на контекст звіту. Для завдань аналізу великих даних використання мір є кращою практикою, оскільки забезпечує оптимізацію використання ресурсів та гнучкість аналізу. Приклади ключових мір аналізу мовою DAX подано у додатку А.

Розуміння концепції контекстів є ключем до ефективного використання мови формул. Контекст фільтра – набір фільтрів, які застосовуються до моделі даних у момент обчислення формули. Зазначені фільтри можуть надходити з візуалізацій, зрізів, фільтрів сторінки або бути програмно модифіковані всередині самої формули. Контекст рядка – концепція, яка дозволяє формулі «бачити» поточний рядок під час ітерації по таблиці. Вміння керувати контекстами дозволяє реалізовувати складну аналітичну логіку, таку як порівняння періодів, розрахунок накопичувальних підсумків, аналіз частки від цілого та ранжування елементів.

Важливим аспектом моделювання для великих даних є оптимізація продуктивності запитів. Процес включає використання змінних для уникнення повторних обчислень, заміну дорогих функцій фільтрації на більш ефективні аналоги, мінімізацію використання ітераторів та правильне управління гранулярністю даних. Оптимізація моделі дозволяє забезпечити швидкий відгук навіть при роботі з мільярдами рядків даних, що є критичним для забезпечення якісного користувацького досвіду [22; 26].

Процес моделювання даних доцільно розглядати як ієрархічну структуру, що складається з трьох послідовних рівнів абстракції:

– *концептуальне моделювання* – визначає загальну структуру бізнесу та інформаційних потоків на високому рівні, ідентифікуючи ключові бізнес-

сутності та їхні семантичні взаємозв'язки без прив'язки до технічної реалізації. Етап дозволяє узгодити розуміння даних між бізнес-замовниками та розробниками;

– *логічне моделювання* – деталізує атрибутивний склад сутностей, типи даних, бізнес-правила та обмеження цілісності, формуючи технічну специфікацію структури даних;

– *фізичне моделювання* – є конкретною технічною реалізацією логічної моделі у середовищі аналітичної платформи, що включає налаштування типів даних, індексацію, розділення таблиць та конфігурацію параметрів зберігання для забезпечення максимальної продуктивності.

Таким чином, зазначений системний, багаторівневий підхід до організації та застосування технологій візуалізації гарантує, що побудована аналітична система буде не лише функціональною, але й надійною, масштабованою та оптимізованою для обробки великих обсягів інформації.

2.2 Методи візуалізації великих даних у бізнес-аналітиці

У науковому дискурсі візуалізація великих даних розглядається не просто як процес побудови графіків, а як комплексна дисципліна, що базується на специфічних методах перетворення абстрактних багатовимірних даних у сприйнятні візуальні образи. Методи візуалізації являють собою фундаментальні підходи до мапінгу даних, які визначають стратегію подання інформації залежно від її структури, обсягу та аналітичної мети. На відміну від типів візуалізації, які є конкретними інструментами реалізації (наприклад, стовпчикова діаграма), методи описують загальну логіку проєкції даних у візуальний простір.

Класифікація методів візуалізації великих даних, прийнята в сучасній аналітиці, включає кілька основних категорій, кожна з яких вирішує специфічні завдання обробки масивів інформації:

– *геометричні методи* – базуються на відображенні даних у декартовій або інших системах координат. Підхід включає методи проєкції

багатовимірних даних на двовимірну площину, такі як діаграми розсіювання (Scatter Plots) та паралельні координати. Геометричні методи є найбільш ефективними для виявлення кореляцій, кластерів та викидів у числових масивах даних;

– *ієрархічні методи* – застосовуються для візуалізації даних, що мають складну вкладену структуру. Методика використовує концепцію вкладеності просторових областей для відображення батьківських та дочірніх елементів. Типовими представниками групи є деревоподібні карти та сонячні промені, які дозволяють компактно відобразити ієрархію великого обсягу на одному екрані [29];

– *піксельно-орієнтовані методи* – є критично важливими саме для аналізу великих даних, оскільки ставлять за мету відобразити максимальну кількість одиниць інформації на екрані. Ідея полягає у відображенні кожного значення даних як окремого кольорового пікселя. Групування пікселів за певними алгоритмами дозволяє аналітику миттєво оцінювати патерни та щільність даних у масивах, що містять мільйони записів, без необхідності їх попередньої агрегації;

– *графові методи* – фокусуються на візуалізації зв'язків та відношень між об'єктами. Методи є незамінними при аналізі соціальних мереж, логістичних ланцюжків та транзакційних зв'язків, дозволяючи виявляти центральні вузли та спільноти в неструктурованих даних [15];

– *методи інтерактивної взаємодії* – доповнюють статичні методи механізмами динамічного дослідження даних. До них відносяться методи деталізації, масштабування, фокусування на контексті та динамічної фільтрації. Інтерактивність дозволяє подолати обмеження екрану, надаючи користувачеві можливість самостійно керувати рівнем деталізації інформації.

Практична реалізація розглянутих теоретичних методів здійснюється через конкретні типи візуалізацій, вибір яких залежить від специфіки бізнес-задачі. Серед поширених типів візуалізацій та їх використання можна виділити наступні категорії:

– *лінійні діаграми* – використовуються для візуалізації даних протягом безперервного періоду часу. Графіки складаються з серії точок даних, з'єднаних лініями, що ілюструють зміну певної змінної з часом. Часто застосовуються для відображення тенденцій, таких як зміни цін акцій, температури або розміру населення, а також для порівняння різних наборів даних, як-от показники продажів для різних продуктів або регіонів. У контексті великих даних візуалізації дозволяють агрегувати великі часові ряди, показуючи загальні тенденції без перевантаження деталями [33];

– *гістограми та стовпчикові діаграми* – призначені для порівняння різних категорій даних. Гістограми складаються з прямокутних стовпців, де кожен стовпець представляє окрему категорію даних, а його висота або довжина вказує на величину даних. Інструменти є ефективними для відображення як загальних тенденцій, так і конкретних відмінностей між категоріями. Графічні елементи дозволяють швидко порівнювати агреговані показники великих категорій, виділяючи ключові відмінності;

– *кругові та кільцеві діаграми* – ілюструють відносні розміри різних категорій даних як частки цілого. Кругова діаграма ділиться на сектори, де кожен сектор представляє окрему категорію, а його розмір пропорційний величині даних. Кільцева діаграма має аналогічну структуру, але з отвором посередині, який може використовуватися для відображення додаткової інформації. Зазначені типи найкраще підходять для високорівневого огляду структури, наприклад, частки ринку або розподілу бюджету, але при роботі з великими даними їх слід використовувати обережно, щоб уникнути перевантаження деталями [28];

– *картки* – використовуються для відображення одного, але важливого числа, такого як загальний обсяг продажів, частка ринку або загальна кількість можливостей. Тип візуалізації часто є найкращим способом відстеження ключової метрики, отриманої з величезних наборів даних, забезпечуючи миттєвий фокус на найважливішому;

– *комбіновані діаграми* – поєднують два або більше типів діаграм в одній візуалізації, дозволяючи відображати кілька типів даних одночасно. Поєднання корисне для порівняння різних типів даних або виділення конкретних закономірностей та тенденцій, наприклад, візуалізації взаємозв’язків між обсягом продажів та маржею прибутку [32];

– *матриці* – відображають дані у форматі сітки, схожому на таблицю, але з додатковими параметрами форматування та можливостями агрегації. Таблиці дозволяють відображати дані з кількох джерел та порівнювати різні набори даних, забезпечуючи ефективне дослідження великих обсягів інформації з можливістю деталізації;

– *дерево декомпозиції* – спеціалізована діаграма, що використовується для розбиття складного процесу або проблеми на менші, більш керовані компоненти. Вона дозволяє проводити аналіз першопричин, автоматично визначаючи фактори, що найбільше впливають на результат. Інструмент допомагає візуально досліджувати ієрархічні структури великих даних, виявляючи приховані рушійні фактори;

– *воронкові діаграми* – візуалізують потік даних або кроки процесу від однієї стадії до наступної. Діаграми є корисними для виявлення вузьких місць або областей для покращення в процесах, що обробляють великі обсяги даних, наприклад, у воронці продажів;

– *візуальні елементи ключових показників ефективності* – відображають статус одного або кількох показників, які є вимірними значеннями, що відстежують продуктивність організації щодо цілей. Елементи надають швидкий огляд стану важливих метрик для моніторингу ефективності бізнесу.

У таблиці 2.2 представлено поширені типи візуалізацій Power BI та їх застосування для бізнес-аналітики.

Таблиця 2.2

Поширені типи візуалізацій Power BI та їх застосування для бізнес-аналітики

Тип візуалізації	Основне призначення	Приклади використання в бізнес-аналітиці	Переваги для великих даних
Лінійна діаграма	Відображення тенденцій з часом	Зміна продажів за місяцями, динаміка цін на акції, зростання кількості клієнтів	Ефективно агрегує великі часові ряди, дозволяючи бачити загальні тенденції без перевантаження деталями
Гістограма/Стовпчикова діаграма	Порівняння категорій даних.	Продажі за регіонами, порівняння продуктивності відділів, розподіл клієнтів за віком	Дозволяє швидко порівнювати агреговані показники великих категорій, виділяючи ключові відмінності
Кругова/Кільцева діаграма	Відображення пропорцій частин до цілого	Частка ринку різних продуктів, розподіл бюджету за статтями витрат	Швидко показує відносний внесок кожної категорії у загальний обсяг, що важливо для високорівневих оглядів великих даних
Карта	Відображення одного ключового показника	Загальний дохід, кількість нових клієнтів, поточна температура	Ідеально для відображення найважливіших КРІ, отриманих з величезних наборів даних, забезпечуючи миттєвий огляд
Комбінована діаграма	Поєднання двох або більше типів діаграм.	Продажі (стовпці) та маржа (лінія) за місяцями, кількість замовлень та середня вартість замовлення	Дозволяє візуалізувати взаємозв'язки між різними метриками з великих даних, виявляючи кореляції
Матриця	Відображення даних у сітці з можливістю деталізації	Таблиця продажів за продуктами та регіонами, фінансовий звіт з деталізацією за рахунками	Дозволяє досліджувати великі обсяги даних у табличному форматі з можливістю агрегації та деталізації

Дерево декомпозиції	Розбиття складної проблеми на компоненти	Аналіз причин зниження продажів (від загального до конкретного продукту/регіону/клієнта)	Допомагає візуально досліджувати ієрархічні структури великих даних, виявляючи рушійні фактори
Воронкова діаграма	Візуалізація потoku даних або кроків процесу	Етапи воронки продажів (від потенційних клієнтів до закритих угод), прогрес у маркетинговій кампанії	Допомагає виявляти вузькі місця у процесах, що обробляють великі обсяги даних, та оптимізувати їх
KPI	Відображення статусу ключових показників ефективності	Поточний CSAT, 30-денний рівень повторних госпіталізацій, рівень задоволеності співробітників	Надає швидкий огляд стану важливих метрик, отриманих з великих даних, для моніторингу ефективності

Наступним рівнем візуалізації є створення інтерактивних дашбордів та звітів, які є потужними інструментами для перетворення даних на візуально захоплюючі та дієві інсайти. Дашборд – коротко описуючи є одною сторінкою, яка показує різні діаграми та візуалізації, призначені для легкого дослідження та взаємодії з даними [32]. Інструмент надає високорівневий, консолідований огляд ключових показників бізнесу. Звіти, на відміну від дашбордів, є більш деталізованими та можуть містити кілька сторінок, призначених для глибокого аналізу та розкриття найдрібніших деталей з одного набору даних.

Основні компоненти дизайну дашбордів включають візуалізації та плитку, що передають інформацію; фільтри та інтерактивні елементи, такі як слайсери, які дозволяють користувачам динамічно взаємодіяти з даними; ключові показники ефективності, що пропонують негайні висновки про стан бізнесу; а також механізми навігації та деталізації, які покращують користувацький досвід. Професійний дашборд підтримується візуально привабливим дизайном з послідовними елементами брендування [13].

Принципи створення ефективних звітів базуються на чіткому визначенні цілей та цільової аудиторії, пріоритезації ключових метрик та уникненні інформаційного перевантаження. Використання послідовних тем, кольорів та шрифтів створює цілісний вигляд, а вибір відповідних візуалізацій покращує розуміння інформації. Забезпечення інтерактивності через слайсери та фільтри дозволяє користувачам динамічно досліджувати дані, що призводить до більш персоналізованого досвіду.

Перехід від статичних звітів до динамічних інструментів прийняття рішень є однією з найважливіших трансформацій у бізнес-аналітиці. Традиційні звіти часто були статичними документами, що надавали знімок даних на певний момент часу. Сучасні платформи дозволяють створювати інтерактивні дашборди з оновленнями в реальному часі, що змінює парадигму від пасивного споживання інформації до активного дослідження [25]. Інтерактивність є першочерговою для великих даних, оскільки вона дозволяє користувачам аналізувати масивні набори даних «на льоту», виявляючи приховані закономірності.

У випадках, коли стандартних візуалізацій недостатньо (рис. 2.3) для задоволення специфічних аналітичних потреб, платформа надає можливість використовувати користувацькі візуалізації.

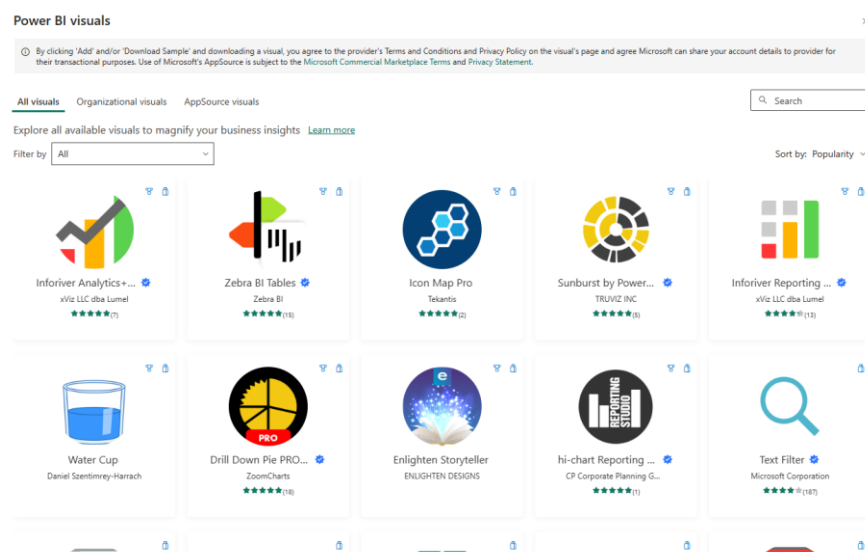


Рис. 2.3. Інтерфес внутрішньої платформи користувацьких та корпоративних візуалізацій в Power BI

Розширення дозволяють візуалізувати дані у персоналізований спосіб, використовуючи бібліотеки з відкритим вихідним кодом.

Компоненти можуть базуватися на відомих бібліотеках JavaScript, таких як D3 або jQuery, а також на скриптах мови R, що відкриває широкі можливості для інтеграції передових методів аналізу.

Процес створення включає використання інструментів розробника, тестування, налагодження та, за необхідності, сертифікацію.

Таким чином, використання користувацьких візуалізацій дозволяє подолати обмеження стандартних інструментів для нішевих або складних сценаріїв великих даних, забезпечуючи необхідну гнучкість для точного представлення нюансів.

2.3 Типи візуалізацій та побудова інтерактивних звітів

Ефективність візуального аналізу безпосередньо залежить від коректного вибору типу графічного представлення даних. В сучасних системах бізнес-аналітики доступний широкий спектр візуалізацій, кожна з яких призначена для вирішення специфічних аналітичних завдань.

При роботі з великими даними вибір ускладнюється необхідністю врахування продуктивності рендерингу та когнітивного навантаження на користувача. Класифікація візуалізацій здійснюється за їх функціональним призначенням [6; 29].

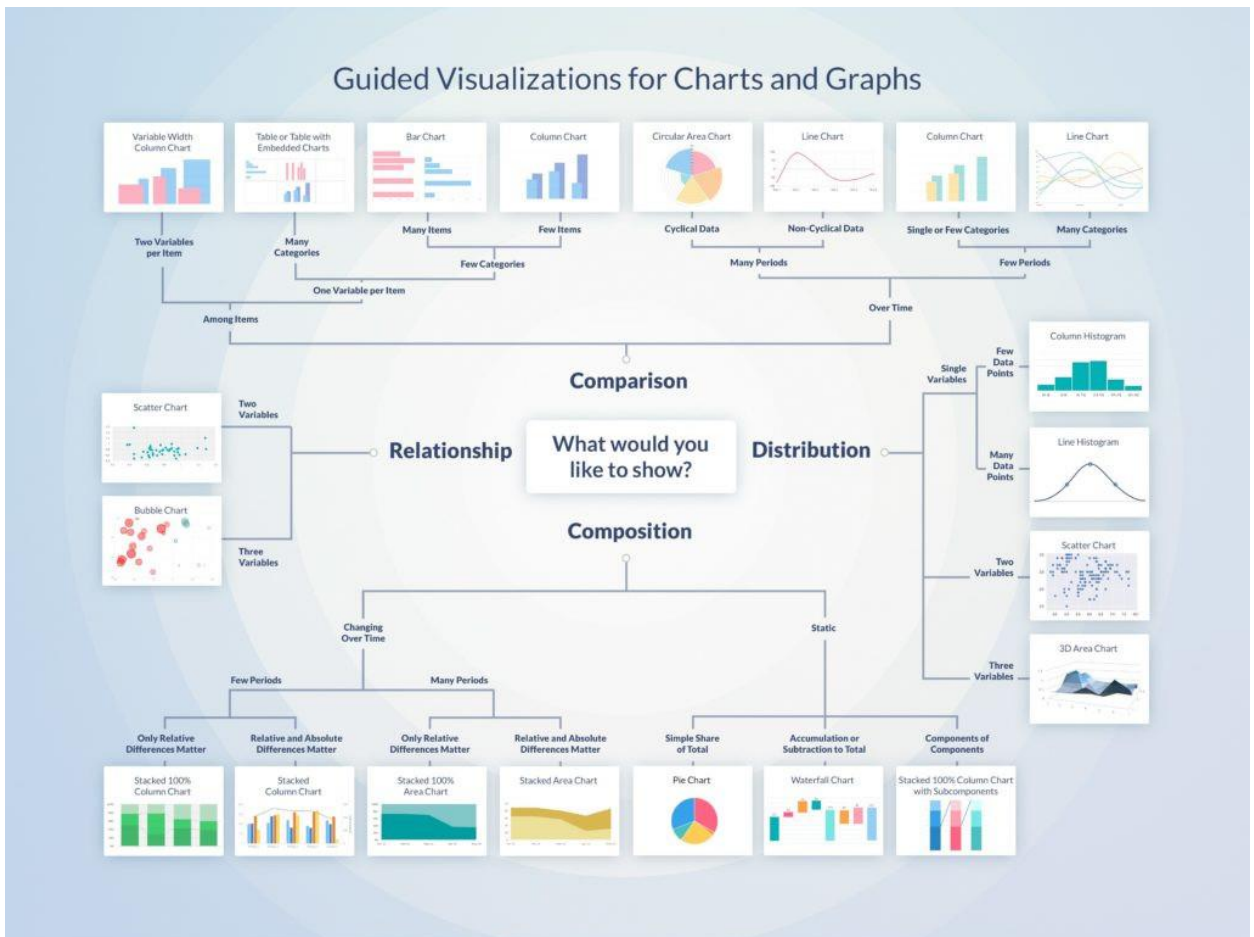


Рис. 2.4. Класифікація методів візуалізації даних

– *лінійні діаграми* – є основним інструментом для аналізу часових рядів. Графіки складаються з послідовності точок даних, з'єднаних лініями, що дозволяє чітко відслідковувати динаміку змін показника в часі. У контексті великих даних тип візуалізації є незамінним для виявлення довгострокових трендів, сезонності та циклічності. Завдяки здатності агрегувати мільйони транзакцій у просту лінію, діаграми дозволяють аналізувати величезні часові проміжки без втрати інформативності. Важливим аспектом є можливість накладання кількох ліній для порівняльного аналізу динаміки різних категорій або періодів;

– *гістограми та стовпчикові діаграми* – призначені для категоричного порівняння. Елементи використовують висоту або довжину прямокутників для відображення величини показника. Тип є найбільш ефективним для ранжування об'єктів, порівняння результатів діяльності різних підрозділів або аналізу розподілу частот. При роботі з великими даними важливо

використовувати функції сортування та обмеження кількості відображуваних елементів (наприклад, «Топ-10»), щоб уникнути перевантаження графіка [33];

– *секторні та кільцеві діаграми* – використовуються для відображення структури цілого, ілюструючи частку кожної категорії у загальному обсязі. Незважаючи на їх популярність, у професійній аналітиці великих даних використання рекомендується обмежувати ситуаціями з невеликою кількістю категорій (до 5–6), оскільки велика кількість секторів ускладнює візуальне порівняння площ. Більш ефективною альтернативою для складних ієрархічних структур є деревоподібні карти, які використовують вкладені прямокутники для відображення ієрархії та величини показників;

– *картки та індикатори* – є ключовими елементами для відображення агрегованих показників ефективності. Елементи фокусують увагу користувача на найважливіших цифрах (загальний дохід, прибуток, кількість клієнтів), отриманих шляхом обробки всього масиву даних. Використання карток дозволяє миттєво оцінити статус виконання цілей без необхідності аналізу детальних графіків;

– *комбіновані діаграми* – поєднують переваги стовпчастих та лінійних діаграм, дозволяючи відображати різноманітні метрики на одному полі (наприклад, абсолютний обсяг продажів у стовпцях та відносну маржу у лінії). Поєднання є потужним інструментом для пошуку кореляцій та залежностей між показниками в масивах великих даних;

– *матриці та таблиці* – забезпечують можливість детального перегляду даних у структурованому вигляді. На відміну від простих списків, матриці підтримують ієрархічне групування рядків і стовпців, а також функції згортання та розгортання рівнів. Функціонал дозволяє аналізувати дані з різним ступенем деталізації – від підсумків по роках до окремих днів – в межах одного візуального елемента.

– *інтелектуальні візуалізації* – становлять окрему групу, яка використовує алгоритми машинного навчання для автоматичного пошуку

інсайтів. Дерево декомпозиції дозволяє проводити глибокий аналіз першопричин, автоматично розбиваючи загальний показник на фактори впливу. Інструмент аналізу ключових факторів автоматично визначає атрибути, що найбільше впливають на цільову метрику, використовуючи регресійний аналіз. Інструменти є критично важливими для роботи з великими даними, оскільки дозволяють виявити приховані закономірності, які неможливо знайти вручну [19].

Наступним рівнем організації візуального простору є створення інтерактивних дашбордів. Дашборд – єдиний інтерфейс, що консолідує найважливішу інформацію, необхідну для досягнення конкретних цілей моніторингу та управління. На відміну від класичних багатосторінкових звітів, дашборд проєктується за принципом «одного екрану», надаючи миттєвий зріз ситуації [32]. Приклади інтерфейсів розроблених інтерактивних звітів наведено у додатку Б. Принципи побудови ефективних дашбордів для великих даних базуються на ергономіці та психології сприйняття. Ключовим правилом є пріоритезація інформації: найбільш важливі показники розміщуються у верхньому лівому куті, а ступінь деталізації зростає при русі вниз та вправо. Важливим аспектом є мінімізація неінформативних елементів дизайну на користь чистоти даних.

Інтерактивність є визначальною рисою сучасних аналітичних звітів. Використання наскрізних фільтрів, часових шкал та механізмів деталізації (drill-through) трансформує дашборд зі статичної картинки в інструмент активного дослідження даних. Механізм дозволяє користувачеві самостійно формувати аналітичні запити, переходячи від загального огляду до аналізу конкретних транзакцій, що є основою концепції самообслуговування в аналітиці [25]. Підхід забезпечує гнучкість прийняття рішень та дозволяє оперативно реагувати на зміни в бізнес-середовищі, спираючись на достовірні дані.

Таким чином, сучасний арсенал візуальних засобів у поєднанні з принципами інтерактивного дизайну дозволяє вирішити головне протиріччя

бізнес-аналітики: як надати користувачу повний доступ до масивів великих даних, зберігши при цьому ясність та простоту сприйняття інформації. Систематизація розглянутих типів візуалізацій та методів їх компонування створює необхідну методичну базу для переходу до наступного етапу роботи – практичного проектування та розробки аналітичної системи, яка буде здатна ефективно функціонувати в умовах реальних бізнес-навантажень.

Висновки до другого розділу

У другому розділі кваліфікаційної роботи проведено комплексне дослідження методологічних та технологічних аспектів побудови систем візуалізації великих даних. На основі системного аналізу архітектури платформи Microsoft Power BI та супутніх інструментів обробки інформації сформульовано наступні узагальнення:

1. Визначено, що ефективність аналітичної системи в умовах роботи з Big Data першочергово залежить від правильно обраної стратегії інтеграції даних. Критичний аналіз режимів підключення показав, що для обробки масивів інформації, обсяг яких перевищує доступну оперативну пам'ять, безальтернативним є використання режиму DirectQuery або гібридних композитних моделей. Такий архітектурний підхід дозволяє перенести обчислювальне навантаження на сторону джерела даних, забезпечуючи доступ до актуальної інформації в реальному часі без необхідності її фізичного дублювання в моделі.

2. Обґрунтовано ключову роль процесів підготовки та трансформації даних у забезпеченні достовірності аналітичних висновків. Встановлено, що використання інструментарію Power Query та мови M дозволяє автоматизувати етапи очищення, типізації та нормалізації даних, вирішуючи проблеми їх різноманітності та низької якості ще до моменту завантаження в модель. Доведено, що застосування технології згортання запитів є критично

важливим фактором оптимізації продуктивності, оскільки мінімізує обсяг даних, що передаються мережею.

3. Доведено, що фундаментом швидкодії аналітичної системи є якісно спроектована семантична модель даних. Порівняльний аналіз схем моделювання підтвердив переваги використання денормалізованих структур типу «Галактика» для комплексних аналітичних завдань, що охоплюють кілька бізнес-процесів. Така організація даних, у поєднанні з використанням мови DAX для створення динамічних метрик, забезпечує високу швидкість агрегації мільярдів рядків транзакцій та гнучкість у розрахунку показників ефективності.

4. Систематизовано методи візуалізації та розроблено рекомендації щодо їх застосування залежно від типу аналітичного завдання. Визначено, що для виявлення прихованих закономірностей у великих масивах даних доцільно використовувати не лише класичні графіки, а й інтелектуальні візуалізації на базі алгоритмів машинного навчання. Аргументовано необхідність переходу від статичних звітів до інтерактивних дашбордів, які завдяки механізмам наскрізної фільтрації та деталізації трансформують процес роботи з даними в активне дослідження, дозволяючи користувачеві самостійно знаходити відповіді на непередбачені питання.

5. Встановлено, що сучасні ВІ-платформи надають замкнений екосистемний цикл роботи з даними, який покриває всі етапи від підключення до джерела до фінальної візуалізації. Можливість розширення базового функціоналу через користувацькі візуалізації та інтеграцію скриптів дозволяє адаптувати систему до специфічних потреб будь-якої предметної області. Визначені архітектурні принципи, методи моделювання та підходи до візуалізації створюють надійне підґрунтя для переходу до наступного етапу роботи – практичної розробки моделі та візуалізації великих даних у бізнес-аналітиці засобами Power BI та імплементації системи інтелектуального аналізу даних, що буде реалізовано у третьому розділі.

РОЗДІЛ 3

РОЗРОБКА МОДЕЛІ ТА ВІЗУАЛІЗАЦІЇ ВЕЛИКИХ ДАНИХ У БІЗНЕС-АНАЛІТИЦІ ЗАСОБАМИ POWER BI

У третьому розділі магістерської роботи детально висвітлено процес практичної реалізації системи інтелектуального аналізу даних, спроектованої для обробки великих масивів інформації у сфері роздрібної торгівлі. Основна увага приділяється технічним аспектам побудови архітектури, яка поєднує класичні підходи Business Intelligence (BI) з передовими алгоритмами машинного навчання (Machine Learning). Розділ охоплює повний життєвий цикл роботи з даними: від обґрунтування вибору інструментарію та розгортання експериментального середовища до реалізації ETL-процесів, побудови семантичної моделі, імплементації AI-алгоритмів та фінальної візуалізації результатів у вигляді інтерактивного дашборду. Окремий акцент зроблено на методології масштабування розробленого прототипу для роботи з Big Data.

3.1 Розробка моделі великих даних у бізнес-аналітиці

В умовах стрімкого зростання обсягів неструктурованої та напівструктурованої інформації проєктування ефективної системи бізнес-аналітики вимагає ретельного та виваженого підходу до вибору архітектурних рішень. Сучасні вимоги до аналітичних систем виходять далеко за межі простої візуалізації історичних даних; вони включають потребу в інтеграції прогнозних моделей, забезпеченні мінімальної затримки при обробці запитів та можливості горизонтального масштабування.

У дослідження розроблено прототип аналітичної системи, головною вимогою до якої була масштабованість, тобто здатність до безшовної міграції від роботи з локальними наборами даних до обробки великих даних у хмарному середовищі без необхідності зміни основної бізнес-логіки та алгоритмів розрахунку показників. Для реалізації практичної частини

дослідження обрано технологічний стек компанії Microsoft. Рішення базується на аналізі ринкових позицій та функціональних можливостей платформи. Згідно з аналітичними звітами *Gartner*, Microsoft Power BI стабільно займає лідерські позиції як за повнотою візії, так і за здатністю до реалізації [21; 20].

Варто зазначити, що вибір технологічного стеку та архітектурних рішень для розробки системи значною мірою ґрунтується на практичному досвіді, отриманому під час роботи в компанії «iTransition». Окремі підходи до моделювання даних та побудови інтерактивних дашбордів, досліджені та систематизовані в рамках цієї кваліфікаційної роботи, були успішно адаптовані та використані при виконанні реальних проєктних завдань. Це дозволило перевірити ефективність запропонованих методів на практиці та підтвердити їх відповідність сучасним вимогам ІТ-індустрії до аналітичних систем корпоративного рівня.

Розроблена архітектура рішення є трирівневою і включає наступні компоненти: рівень даних реалізовано на базі системи управління базами даних Microsoft SQL Server. У контексті роботи SQL Server виконує роль корпоративного сховища даних. Вибір реляційної СКБД зумовлений необхідністю забезпечення цілісності транзакційних даних та наявністю потужних інструментів індексації, що критично важливо для швидкодії при виконанні складних SQL-запитів. Для структурованих фінансових даних (рис. 3.1) та даних про продажі реляційна модель залишається стандартом, незважаючи на поширення NoSQL рішень.

Figure 3.1 shows the SSMS interface displaying a query result for the *FactInternetSales* table. The query is a `SELECT TOP (1000)` statement listing various columns including `ProductKey`, `OrderDateKey`, `ShipDateKey`, `CustomerKey`, `PromotionKey`, `CurrencyKey`, `SalesTerritoryKey`, `SalesOrderNumber`, `SalesOrderLineNumber`, `RevisionNumber`, `OrderQuantity`, `UnitPrice`, `ExtendedAmount`, `UnitPriceDiscountPct`, `DiscountAmount`, and `ProductStandard`. The results are displayed in a grid with 1000 rows and 15 columns. The status bar at the bottom indicates "Query executed successfully".

Рис. 3.1. Інтерфейс SSMS з вибіркою даних із таблиці *FactInternetSales*

Рівень логіки та трансформації відповідає за вилучення даних з джерела, їх очищення, трансформацію та завантаження у модель. Для реалізації завдань використано інструментарій Power Query та мову формул М. Інструмент дозволяє створювати відтворювані скрипти обробки даних, які автоматично застосовуються при оновленні інформації. Для реалізації бізнес-логіки, зокрема розрахунку ключових показників ефективності та аналізу відхилень, використано мову DAX. Мова дозволяє створювати динамічні міри, які перераховуються «на льоту» залежно від контексту фільтрації, обраного користувачем.

Рівень візуалізації та штучного інтелекту представлено клієнтською частиною Microsoft Power BI Desktop. Засіб обрано через його потужні можливості візуалізації та наявність вбудованих функцій штучного інтелекту, які дозволяють застосовувати складні алгоритми машинного навчання, такі як кластеризація та регресійний аналіз, у режимі low-code.

Однією з ключових методологічних проблем при проведенні досліджень у сфері великих даних на етапі прототипування є обмежений доступ до реальних корпоративних кластерів та сувора конфіденційність реальних бізнес-даних. Крім того, технічні характеристики стандартного

робочого середовища фізично не дозволяють розгорнути повноцінний кластер на базі Apache Hadoop або Apache Spark для обробки терабайтів інформації в прийнятний час. Виходячи з наявних обмежень, у роботі застосовано методологію «масштабованого прототипування».

В якості джерела даних для розробки та тестування системи використано базу даних Adventure Works DW 2022, офіційно розроблену та підтримувану компанією Microsoft. Вибір набору даних є науково та технічно обґрунтованим з кількох причин:

- по-перше, схема бази даних спроектована за принципами, ідентичними до реальних сховищ великих даних у сфері глобального ритейлу. Вона представляє собою повноцінну схему «Сніжинка», що містить нормалізовані таблиці фактів, такі як FactInternetSales, FactResellerSales, та розгалужені таблиці вимірів, наприклад DimCustomer, DimProduct. Структура дозволяє на етапі розробки відпрацювати складні алгоритми побудови семантичної моделі та оптимізації швидкодії, які будуть ідентичними при роботі з реальними петабайтами даних;

- по-друге, набір даних містить повний спектр типів змінних, необхідних для тренування та валідації ML-моделей. Безперервні числові змінні, наприклад SalesAmount, Freight, TaxAmt, використовуються для задач регресії та прогнозування. Категоріальні змінні, такі як EnglishProductCategoryName, MaritalStatus, застосовуються для задач класифікації та кластеризації. Часові ряди OrderDate, ShipDate з високою гранулярністю дозволяють проводити аналіз сезонності та трендів. Географічні дані City, StateProvinceCode, CountryRegionCode необхідні для просторового аналізу. Наявна різноманітність дозволяє повноцінно протестувати роботу візуальних елементів штучного інтелекту;

- по-третє, аналітична модель, розроблена на базі SQL Server, є архітектурно агностичною до обсягу даних. Факт означає, що при переході від прототипу до продуктивного середовища джерело даних може бути безшовно замінене на хмарне сховище великих даних, таке як Azure Synapse

Analytics або Azure Data Lake Storage. При міграції вся розроблена бізнес-логіка, структура звіту та налаштовані алгоритми збережуться повністю. Зміниться лише режим підключення: з режиму імпорту, який використовується для прототипування, на режим прямого запиту, який генерує запити до хмарного кластера при кожній взаємодії користувача з дашбордом. Відтак, використання Adventure Works дозволяє створити повнофункціональний «цифровий двійник» реальної системи великих даних, уникнувши апаратних обмежень локального середовища, але зберігши повну методологічну чистоту експерименту та готовність рішення до масштабування.

Для забезпечення відтворюваності результатів експерименту наводимо характеристики середовища, в якому проводилася розробка та тестування системи: операційна система Windows 11 Pro, система управління базами даних Microsoft SQL Server 2025 Developer Edition, середовище розробки Microsoft Power BI Desktop (версія за листопад 2025 року) та середовище виконання скриптів Python 3.9 з бібліотеками pandas, scikit-learn, matplotlib для розширеної аналітики. Використання локального сервера дозволило емулювати мережеві затримки та наблизити умови тестування до реальних.

3.2 Етапи розробки візуалізації великих даних у бізнес-аналітиці засобами Power BI

У проєктах пов'язаних з аналізом великих даних та машинним навчанням, етап підготовки даних є фундаментом, від якого залежить коректність роботи всіх подальших алгоритмів. Згідно з методологією CRISP-DM, етап займає від 60 % до 80 % часу всього дослідження. У розробленій системі реалізація процесів вилучення, трансформації та завантаження даних здійснювалася засобами вбудованого в Power BI інструменту Power Query Editor, що базується на функціональній мові програмування M.

Першим кроком практичної реалізації стало налаштування підключення до локального сховища даних. Для виконання завдання використано нативний конектор SQL Server Database. Процес підключення включав наступні параметри: сервер – localhost, база даних – AdventureWorksDW2022, режим підключення – Import. Застосування режиму імпорту для прототипу є стратегічним рішенням. У зазначеному режимі рушій VertiPaaS завантажує копію даних в оперативну пам'ять, виконує агресивне стиснення та індексацію стовпців. Механізм забезпечує мілісекундну швидкість відгуку при фільтрації візуалізацій, що є критичним для користувацького досвіду. Однак, архітектура системи спроектована таким чином, що при переході до реальних обсягів великих даних, модель може бути переключена в режим DirectQuery або Composite Model, де агреговані дані зберігаються в пам'яті, а детальні транзакції залишаються в хмарному сховищі.

З бази даних імпортовано набір таблиць, що формують ядро аналітичної моделі: FactInternetSales (основна таблиця фактів, що містить деталізовану інформацію про кожну транзакцію продажу), DimCustomer (таблиця виміру з даними про клієнтів), DimProduct, DimProductCategory, DimProductSubcategory (нормалізовані таблиці продуктового каталогу), DimSalesTerritory (дані про географічну приналежність продажів) та DimDate (спеціалізована таблиця календаря). Детальний вигляд інтерфейсу джерела даних у середовищі SSMS наведено у додатку Г.

Дані з корпоративних сховищ, часто потребують додаткової обробки для коректної роботи алгоритмів машинного навчання. У Power Query Editor реалізовано послідовний ланцюжок трансформацій. Перед початком трансформацій проведено аудит якості даних за допомогою інструментів профілювання, що дозволило виявити стовпці з великою кількістю порожніх значень, аномальні викиди у числових полях та помилки у типах даних.

Критично важливим етапом для ML-алгоритмів є коректна типізація. Для фінансових показників (SalesAmount, TaxAmt, Freight, UnitCost)

примусово встановлено тип даних Fixed Decimal Number (Currency). Формат використовує чотири знаки після коми і забезпечує точність фінансових розрахунків, уникаючи помилок плаваючої коми. Для полів, що містять дати (OrderDate, ShipDate), встановлено тип Date, що дозволяє системі автоматично створювати ієрархії дат і використовувати їх у DAX-функціях.

У контексті великих даних кожен зайвий байт інформації, помножений на мільярди рядків, створює суттєве навантаження на інфраструктуру. Тому застосовано принцип мінімалізму: завантажувати лише те, що необхідно для аналізу. З даних FactInternetSales видалено технічні стовпці: RevisionNumber, CarrierTrackingNumber, CustomerPONumber. Оптимізація не лише зменшила розмір файлу, а й прискорила час оновлення даних.

У базі Adventure Works дані про продукти нормалізовані, тобто розбиті на три таблиці: Товари, Підкатегорії, Категорії. Для оптимізації моделі Power BI та спрощення написання DAX-формул прийнято рішення денормалізувати ці виміри в одну широку таблицю DimProduct. Завдання реалізовано за допомогою операції об'єднання запитів. До таблиці DimProduct приєднано DimProductSubcategory за ключем ProductSubcategoryKey, розгорнуто стовпець EnglishProductSubcategoryName, а до отриманого результату приєднано DimProductCategory за ключем ProductCategoryKey. В результаті отримано єдину таблицю DimProduct (рис. 3.2), що містить повну ієрархію товару. Метод зменшує кількість дорогих операцій з'єднання під час виконання запитів візуалізацій.

```
let
    Source = Sql.Database("localhost", "AdventureWorksDW2020"),
    DimProduct = Source[[Schema="dbo",Item="DimProduct"]][Data],
    #"Merged Subcategory" = Table.NestedJoin(DimProduct, {"ProductSubcategoryKey"},
        DimProductSubcategory, {"ProductSubcategoryKey"}, "Subcategory", JoinKind.LeftOuter),
    #"Expanded Subcategory" = Table.ExpandTableColumn(#"Merged Subcategory", "Subcategory",
        {"EnglishProductSubcategoryName", "ProductCategoryKey"}, {"SubcategoryName", "ProductCategoryKey"}),
    #"Merged Category" = Table.NestedJoin(#"Expanded Subcategory", {"ProductCategoryKey"},
        DimProductCategory, {"ProductCategoryKey"}, "Category", JoinKind.LeftOuter),
    #"Expanded Category" = Table.ExpandTableColumn(#"Merged Category", "Category",
        {"EnglishProductCategoryName", {"CategoryName"}}, {"CategoryName"}),
    #"Removed Columns" = Table.RemoveColumns(#"Expanded Category",
        {"ProductSubcategoryKey", "ProductCategoryKey"})
in
    #"Removed Columns"
```

Рис. 3.2. Лістинг коду мовою M, що демонструє операцію підготовки даних

Для підвищення якості роботи ML-алгоритмів, зокрема візуалізації Key Influencers, створено похідні стовпці ще на етапі ETL. Числове поле YearlyIncome дискретизовано на категорії «High», «Medium», «Low» за допомогою умовної логіки. Дискретизація дозволяє алгоритмам шукати патерни не в хаотичних числах, а в чітких соціальних групах. Також об'єднано поля FirstName та LastName у поле FullName для зручного відображення у звітах.

Після завершення трансформацій та завантаження даних розпочався етап моделювання. Якість семантичної моделі визначає не лише коректність цифр, а й швидкість роботи звітів. Враховуючи складну структуру джерела даних, спроектовано схему «Галактика». Схема є архітектурним патерном, що дозволяє об'єднати різні бізнес-процеси в єдину модель. У центрі моделі знаходяться таблиці фактів, які використовують спільні виміри DimDate, DimProduct та DimCustomer. Структура дозволяє проводити перехресний аналіз (рис. 3.3), наприклад, порівнювати продажі одного й того ж товару через різні канали збуту.

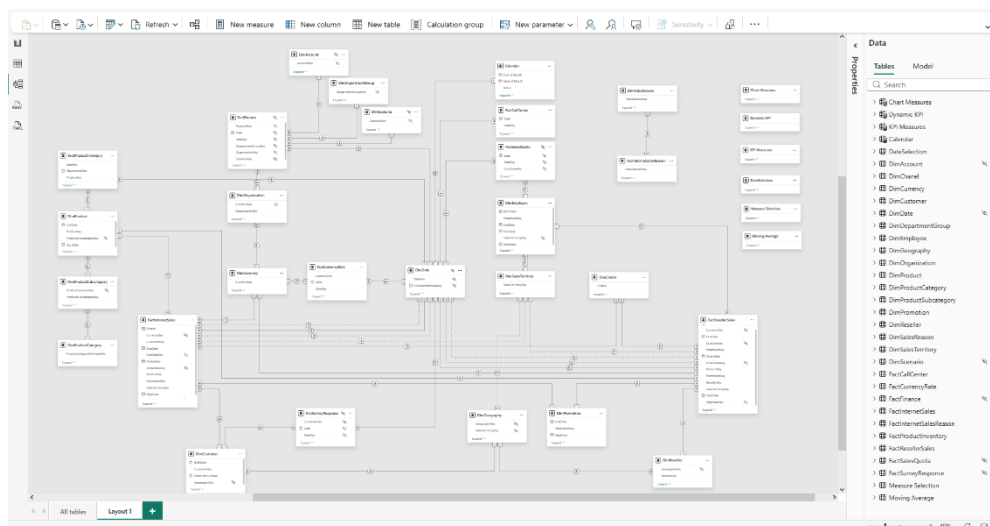


Рис. 3.3. Схема моделі даних типу «Галактика» у Power BI

Між таблицями налаштовано зв'язки типу «один-до-багатьох», де виміри фільтрують таблиці фактів. Напрямок крос-фільтрації встановлено як однонаправлений, що запобігає виникненню неоднозначностей у складній схемі та забезпечує коректність обчислень. Для коректної роботи функцій аналізу часу таблицю DimDate позначено як офіційну таблицю дат, що

вимикає вбудовані автоматичні ієрархії та дозволяє використовувати єдиний календар для всіх фактів. Усі технічні зовнішні ключі, такі як CustomerKey та ProductKey, приховані у поданні звіту (рис. 3.4), залишаючи доступними лише бізнес-поля.

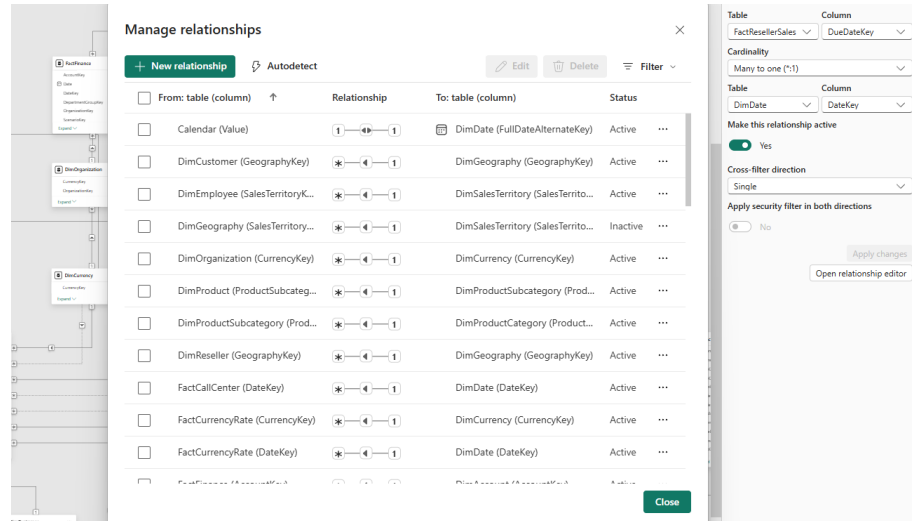


Рис. 3.4. Інтерфейс внутрішнього менеджера зв'язків таблиць

Таким чином, у результаті виконання етапу попередньої обробки створено оптимізовану, очищену та структуровану модель даних, яка повністю готова до застосування складних аналітичних обчислень та алгоритмів машинного навчання.

3.3 Реалізація методів інтелектуального аналізу та розрахунок ключових показників ефективності

Після завершення етапу моделювання даних та побудови схеми «Галактика» реалізовано аналітичний шар системи. Процес включає дві складові: детермінований аналіз (розрахунок чітких фінансових показників та ключових метрик ефективності) та імовірнісний аналіз (застосування алгоритмів машинного навчання для пошуку прихованих закономірностей).

Для перетворення транзакційних даних у бізнес-інсайти використано DAX міри. На відміну від обчислюваних стовпців, міри обчислюються динамічно в момент виконання запиту залежно від контексту фільтрації, який задає користувач через зрізи або взаємодію з графіками. Створено окрему

таблицю `_Measures` для централізованого зберігання всіх обчислень, що відповідає кращим практикам розробки [26].

Першим рівнем ієрархії метрик є базові розрахунки, що виконують пряму агрегацію стовпців таблиць фактів. Основний показник доходу компанії `Total Sales` розраховується як сума стовпця `SalesAmount` у таблиці `FactInternetSales`: $\text{Total Sales} = \text{SUM}(\text{FactInternetSales}[\text{SalesAmount}])$

Метрика є адитивною, тобто коректно підсумовується по будь-яких вимірах. Для розрахунку прибутковості створено міру `Total Cost` (собівартість реалізованої продукції), яка базується на стовпці `TotalProductCost`. Валовий прибуток `Total Profit` розраховується як різниця між продажами та собівартістю. Застосовано механізм галуження мір, коли одна міра базується на інших, що дозволяє змінювати логіку в одному місці, автоматично оновлюючи всі залежні метрики: $\text{Total Profit} = [\text{Total Sales}] - [\text{Total Cost}]$

Для порівняння ефективності різних сегментів бізнесу розраховано відносні метрики. Показник `Profit Margin %` (рентабельність продажів) показує прибутковість на одиницю продажу. Використано функцію `DIVIDE` для безпечного ділення з автоматичною обробкою помилки ділення на нуль: $\text{Profit Margin \%} = \text{DIVIDE}([\text{Total Profit}], [\text{Total Sales}], 0)$

Для аналізу поведінки користувачів розроблено метрики підрахунку унікальних сутностей, зокрема `Unique Customers` (кількість активних клієнтів) та `Average Ticket` (середній чек), який є ключовим показником для ритейлу та показує купівельну спроможність аудиторії: $\text{Avg Ticket} = \text{DIVIDE}([\text{Total Sales}], \text{DISTINCTCOUNT}(\text{FactInternetSales}[\text{SalesOrderNumber}]))$

Ключовою відмінністю розробленої системи від традиційних звітів є інтеграція штучного інтелекту для автоматизованого пошуку інсайтів. У `Power BI` завдання реалізовано через групу інструментів `AI Visuals`, які використовують потужності сервісів `Azure ML` [30]. Для вирішення завдання

пояснювальної аналітики використано візуальний елемент Key Influencers (Аналіз ключових факторів впливу).

Бізнес-завдання полягало у визначенні характеристик клієнта або товару, що найбільше впливають на збільшення суми чека. Візуалізація автоматично виконує сканування даних і, залежно від типу цільової змінної, обирає метод аналізу. Оскільки цільовою змінною є SalesAmount (число), алгоритм будує лінійну регресійну модель, знаходячи коефіцієнти кореляції для кожного фактора. У поле аналізу поміщено міру Total Sales. Для пояснення впливу додано набір атрибутів з таблиці DimCustomer: EnglishCountryRegionName, MaritalStatus, TotalChildren, EnglishOccupation, HouseOwnerFlag. Після обробки даних система згенерувала список факторів впливу. Наприклад, виявлено, що статус «Professional» збільшує середній чек (рис. 3.5.), а відсутність дітей корелює зі зростанням суми покупок.

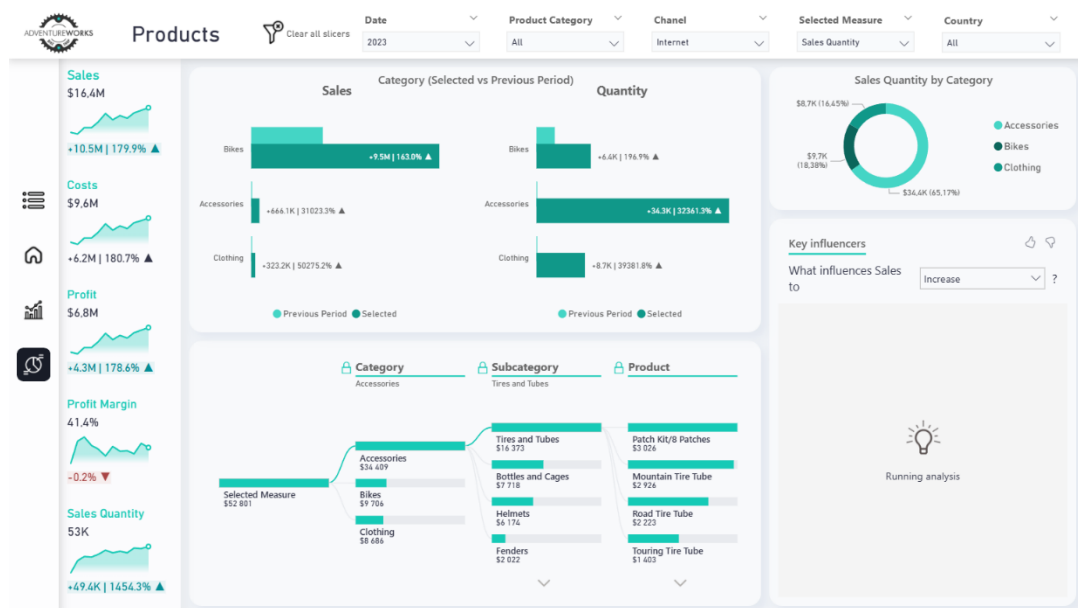


Рис. 3.5. Використання AI-візуалізацій Key Influencers та Decomposition Tree для аналізу факторів впливу на продажі

Крім того, інструмент автоматично згенерував сегменти за допомогою алгоритму дерев рішень, що дозволило сформувати портрет ідеального клієнта.

Для реалізації глибокого діагностичного аналізу застосовано інструмент Decomposition Tree. Компонент дозволяє візуалізувати ієрархічну

структуру даних та зрозуміти внесок кожного елемента в загальний результат. Інструмент налаштовано для декомпозиції загальної суми продажів за вимірами: Category, Subcategory, Product, Country. Головною перевагою є функція інтелектуального вибору шляху. Користувачеві надається вибір: High Value або Low Value. Алгоритм сканує всі доступні виміри і автоматично обирає той, який містить елемент з екстремальним значенням метрики. Функціонал дозволяє реалізувати сценарій експлоративного аналізу, коли менеджер може швидко знайти причину падіння прибутку, не перебираючи сотні комбінацій фільтрів вручну [29].

Для забезпечення цілісності аналітичного досвіду налаштовано взаємодію між усіма елементами звіту. Реалізовано перехресну фільтрацію, коли вибір стовпчика на графіку продажів автоматично фільтрує дерево декомпозиції та фактори впливу. Також налаштовано наскрізну деталізацію, що дає можливість переходу з загального огляду на детальну сторінку аналізу клієнтів із збереженням контексту фільтрації.

Таким чином, програмна реалізація аналітичного блоку поєднала в собі точність математичних розрахунків DAX та евристичну потужність алгоритмів машинного навчання, забезпечуючи користувача інструментами як для контролю поточного стану, так і для стратегічного планування.

3.4 Розробка візуалізації великих даних у бізнес-аналітиці засобами Power BI

Фінальним та інтеграційним етапом програмної реалізації стало створення візуального інтерфейсу системи – інтерактивного звіту, який слугує єдиною точкою входу користувачів до аналітичних даних. При проєктуванні архітектури звіту суворо дотримано принципів «розповіді даних» та використано сучасний підхід до UX/UI дизайну, що нагадує навігацію у повноцінних веб-додатках. Головною метою було створення ергономічного середовища, яке мінімізує когнітивне навантаження на

користувача та дозволяє отримувати відповіді на бізнес-питання за мінімальну кількість кліків.

Натомість традиційного використання стандартних вкладок внизу сторінки, реалізовано вертикальне навігаційне меню зліва. Елемент містить інтуїтивно зрозумілі ікони для переходу між основними сторінками звіту Home, Sales, Products та є статичним елементом, який залишається на екрані під час навігації, забезпечуючи контекстну безперервність роботи. Ліва частина дашборду виділена під наскрізну панель ключових показників ефективності (Global KPI Sidebar). Незалежно від обраної сторінки та глибини деталізації, користувач завжди бачить 5 критичних метрик, що дозволяють тримати руку на пульсі бізнесу: Sales, Costs, Profit, Profit Margin та Sales Quantity.

Для кожного показника реалізовано міні-графік (Sparkline), що візуалізує тренд за останній звітний період, дозволяючи миттєво оцінити стабільність показника без необхідності переходу на детальні сторінки. Під графіком динамічно відображається індикатор росту (YoY Growth) у абсолютному грошовому вимірі та у відсотках порівняно з аналогічним періодом минулого року. Застосовано умовне форматування: зелений колір індикатора вказує на позитивну динаміку, червоний – сигналізує про падіння, що дозволяє миттєво ідентифікувати проблемні зони.

Головна сторінка звіту виконує роль стратегічного центру управління для топ-менеджменту. Основна мета сторінки – надати інструмент для гнучкого, багаторівневого аналізу динаміки ключових показників у різних часових розрізах. У верхній частині сторінки розміщено панель фільтрів з випадаючими списками, що дозволяють фільтрувати дані за звітним періодом, категорією продукту, каналом продажів, країною походження замовлення та типом бізнесу (B2B/B2C). Для зручності користувача реалізовано кнопку повного очищення фільтрів (рис. 3.6), яка дозволяє скинути всі налаштування одним кліком та повернутися до огляду всієї компанії.

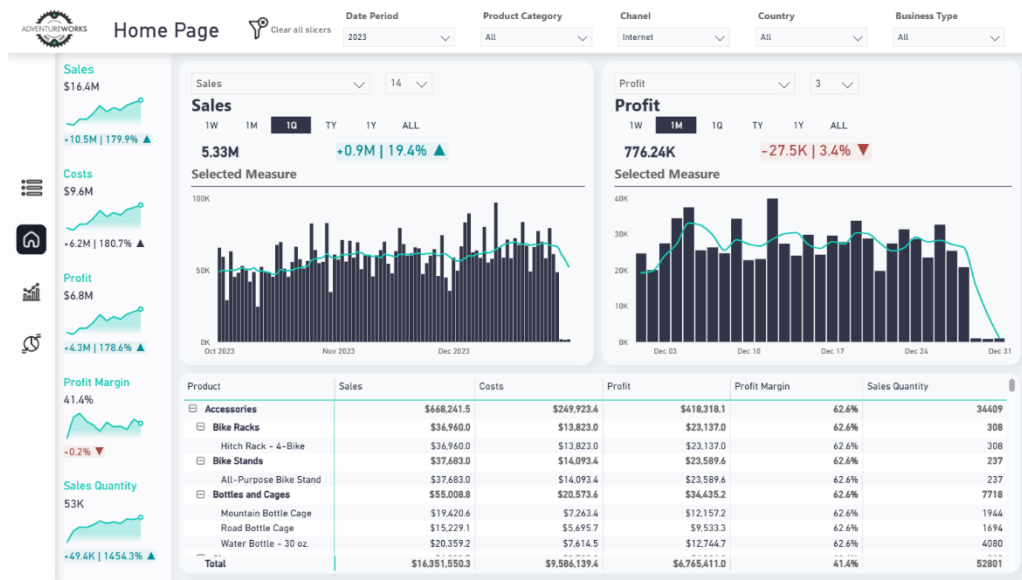


Рис. 3.6. Головна панель аналітичного звіту

Основну частину екрану займає модуль динамічного аналізу часу. Для забезпечення гнучкості аналізу реалізовано кастомні кнопки перемикавання гранулярності часу (1W – тиждень, 1M – місяць, 1Q – квартал, TY – поточний рік, 1Y – рік, ALL – весь період), які за допомогою DAX-логіки та Bookmarks дозволяють динамічно змінювати масштаб осі X на графіку. Інтегровано випадаючий список вибору міри, що дозволяє змінити метрику, яка відображається на графіку (наприклад, переключитися з аналізу доходу на аналіз маржинальності), без необхідності переходу на іншу сторінку звіту. Комбінована діаграма використовує стовпці для відображення фактичних значень за період та лінію для показу згладженого тренду, що допомагає візуально відокремити системні зміни від випадкових коливань.

Сторінка аналізу продажів сфокусована на поглибленому вивченні структури прибутку, сезонності попиту та географічних факторів впливу. Комбінована діаграма у першому квадранті візуалізує фінансові результати по кварталах, одночасно показуючи абсолютну суму прибутку у вигляді стовпців та відносну рентабельність у вигляді лінії. Підхід дозволяє виявити небезпечні ситуації, коли, наприклад, валовий прибуток зростає за рахунок збільшення обсягів, але рентабельність падає через надмірні знижки. Кільцева діаграма поруч відображає структуру прибутку за товарними

категоріями (Bikes, Accessories, Clothing), дозволяючи швидко оцінити внесок кожного напрямку (рис. 3.7) в загальний результат.

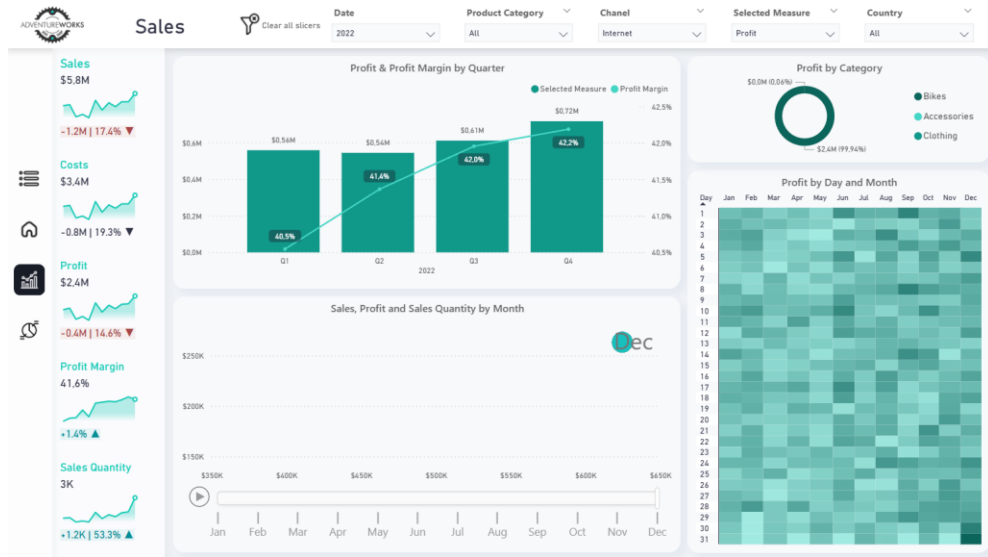


Рис. 3.7. Сторінка детального аналізу продажів

У нижній частині сторінки розміщено динамічну візуалізацію з віссю відтворення. При активації функції графік анімує зміни показників місяць за місяцем, що дозволяє побачити еволюцію продажів у динаміці та зрозуміти, як змінювалася структура попиту протягом року. Календарна теплова карта дозволяє миттєво виявляти дні з аномально високими або низькими продажами, які на звичайному лінійному графіку часто зливаються в загальний шум. Інтенсивність кольору кожної комірки календаря відповідає обсягу продажів, що дозволяє візуально ідентифікувати сезонні піки (наприклад, передсвяткові розпродажі) або дні «просадки».

Сторінка аналізу продуктів поєднує класичний товарний аналіз із просунутими алгоритмами машинного навчання. Діаграма типу «Торнадо» порівнює поточні показники з попереднім періодом для основних категорій, дозволяючи чітко бачити приріст або падіння як у грошовому, так і в кількісному вимірах. Дерево декомпозиції дозволяє користувачеві інтерактивно розгортати загальну суму продажів по ієрархічному ланцюжку категорія-підкатегорія-продукт, виявляючи конкретні товарні позиції, які є драйверами росту (рис. 3.8) або, навпаки, тягнуть показники вниз.

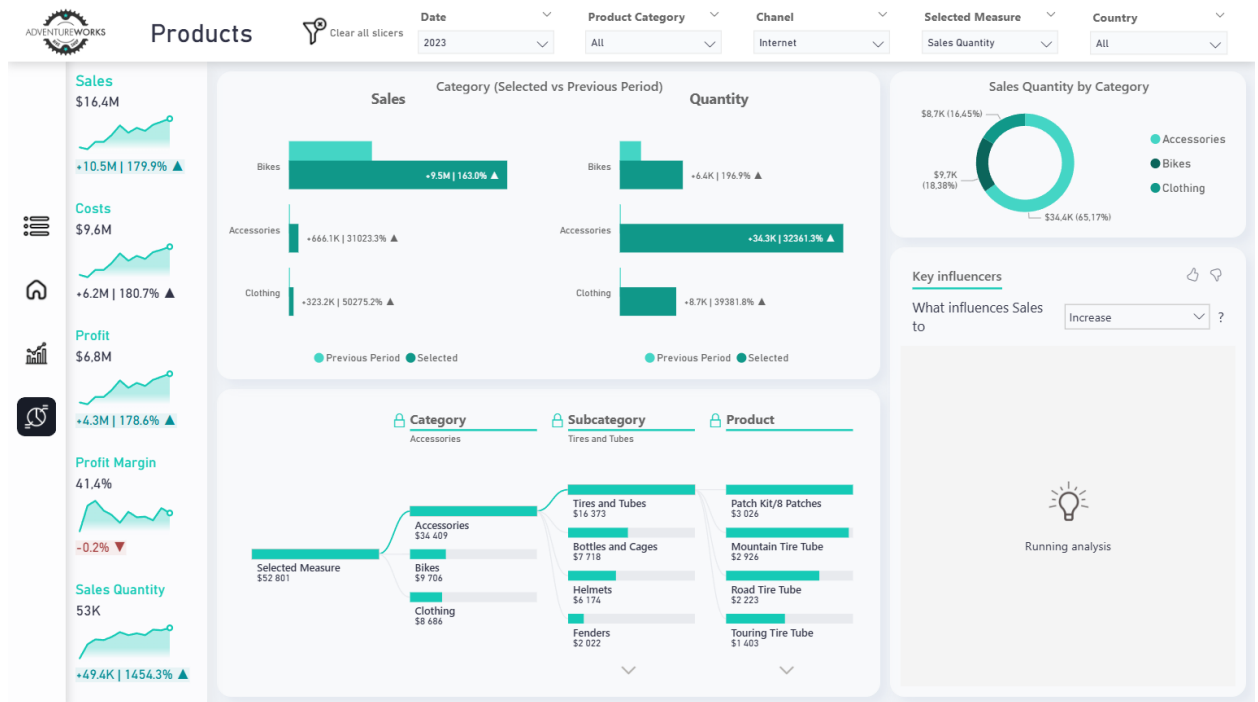


Рис. 3.8. Сторінка AI-інсайти та продуктового аналізу

Візуальний елемент Key Influencers автоматично аналізує фактори, що впливають на ріст продажів, будуючи регресійну модель «на льоту» та пояснюючи виявлені математичні залежності зрозумілою бізнес-мовою (наприклад, «Країна замовлення – США збільшує ймовірність високого чека на 25 %»).

Важливим архітектурним аспектом розробки є забезпечення масштабованості рішення для завдань великих даних. Архітектура системи підтримує технології, що дозволяють обробляти терабайти даних без зміни логічної структури. Хоча у поточному прототипі використано режим імпорту для забезпечення максимальної швидкодії інтерфейсу, система спроектована з можливістю безшовного перемикання в режим DirectQuery. Опція дозволяє візуалізувати дані безпосередньо з хмарного сховища Azure Synapse Analytics, використовуючи обчислювальну потужність кластера бази даних, а не локального комп'ютера користувача.

Найбільш передовим підходом до масштабування, закладеним в архітектуру рішення, є використання агрегацій. Технологія дозволяє створити гібридну таблицю, де агрегована копія даних (наприклад, суми продажів по днях) завантажується в пам'ять для миттєвої швидкодії

візуалізацій верхнього рівня, а детальна таблиця з мільярдами транзакцій залишається у джерелі. Система автоматично визначає тип запиту і непомітно для користувача перемикається між швидким кешем і повним джерелом даних. Для забезпечення актуальності даних передбачено використання політик інкрементального оновлення, що дозволяє розділити масив даних на архівні (статичні) та оперативні (динамічні) партиції, оновлюючи лише останні зміни, що суттєво зменшує навантаження на систему оновлення.

Таким чином, розроблений інтерактивний звіт є не лише засобом візуалізації, а й повноцінним аналітичним додатком, який інтегрує в собі потужність математичних моделей, гнучкість ETL-процесів та передові UX-практики. Рішення створює надійне підґрунтя для переходу до фінального етапу дослідження – узагальнення отриманих результатів та формулювання висновків щодо ефективності запропонованого рішення.

Висновки до третього розділу

У третьому розділі кваліфікаційної роботи представлено результати програмної реалізації системи інтелектуального аналізу даних. Практична частина дослідження підтвердила валідність обраних теоретичних підходів та дозволила створити повнофункціональне аналітичне рішення на базі платформи Microsoft Power BI.

За результатами проведеної роботи сформульовано наступні висновки:

1. Реалізовано інженерно-технологічну складову системи. У ході дослідження успішно апробовано методологію «масштабованого прототипування». На базі технологічного стеку SQL Server та Power BI розгорнуто архітектуру, яка емулює роботу реального корпоративного сховища даних. Ключовим результатом на цьому етапі стала розробка автоматизованого конвеєра обробки даних (ETL-pipeline) мовою M, що дозволило вирішити проблему якості та різномірності вхідної інформації ще до етапу візуалізації: реалізовано алгоритми очищення шумів, нормалізації

довідників та типізації даних, що забезпечило скорочення обсягу моделі та гарантувало цілісність аналітичної бази.

2. Побудовано семантичне ядро та розрахункову логіку. Розроблено та оптимізовано семантичну модель даних за схемою «Галактика», яка стала фундаментом для високошвидкісних аналітичних запитів. На відміну від класичних нормалізованих схем, запропонована структура дозволила об'єднати різноманітні бізнес-процеси в єдиний інформаційний простір через конформні виміри. Функціональність системи забезпечено шляхом написання комплексу динамічних метрик мовою DAX. Застосування патернів аналізу часу та змінних дозволило автоматизувати складні фінансові розрахунки, такі як річний приріст (YoY Growth) та накопичувальний підсумок, усунувши необхідність ручної обробки звітів.

3. Впроваджено інтелектуальну візуалізацію та ергономічний дизайн. Створено інтерактивний користувацький інтерфейс, який трансформує підхід до роботи з даними від пасивного перегляду до активного дослідження. Інтеграція алгоритмів штучного інтелекту надала системі здатності до діагностичної аналітики: алгоритми автоматично виявляють приховані кореляції та фактори впливу на бізнес-показники. Візуальна частина реалізована у форматі навігаційного ' зі спеціалізованими сторінками для різних рівнів менеджменту, що в поєднанні з кастомними елементами управління значно підвищило ергономіку та знизило когнітивне навантаження на користувача.

4. Верифіковано масштабованість рішення для великих даних. Архітектурний аналіз розробленого прототипу підтвердив його готовність до роботи з надвеликими обсягами даних. Обґрунтовано, що закладені механізми – зокрема технологія DirectQuery для прямих запитів до хмарних сховищ та гібридні моделі агрегацій – дозволяють системі зберігати високу продуктивність при експоненціальному зростанні обсягу транзакцій. Доводячи, що запропоноване рішення є не просто навчальним макетом, а

універсальним шаблоном, придатним для розгортання в реальних умовах високонавантажених корпоративних систем.

Узагальнюючи практичні результати, можна стверджувати, що розроблена система є завершеним програмним продуктом, який вирішує завдання візуалізації великих даних шляхом синергії інженерних методів обробки інформації, математичного апарату бізнес-логіки та сучасних принципів візуального проєктування. Результати тестування на еталонному наборі даних підтвердили коректність роботи алгоритмів та високу ефективність запропонованих візуальних рішень.

ЗАГАЛЬНІ ВИСНОВКИ

У кваліфікаційній роботі здійснено комплексне вирішення дослідження щодо методів візуалізації великих даних у бізнес аналітиці та підвищення ефективності аналітичних процесів підприємства шляхом розробки та впровадження системи візуалізації великих даних. На основі проведеного дослідження отримано наступні результати:

1. Теоретично обґрунтовано основи застосування інформаційних технологій та візуалізації великих даних у бізнес-аналітиці. Встановлено, що цифрова трансформація економіки та експоненціальне зростання обсягів даних вимагають переходу від традиційних методів звітності до інтерактивних аналітичних систем. Визначено, що еволюція концепції Big Data до моделі «5V» (обсяг, швидкість, різноманітність, достовірність, цінність) робить використання спеціалізованих інструментів візуалізації безальтернативним способом подолання когнітивного розриву при прийнятті управлінських рішень. Доведено, що впровадження платформ самообслуговування дозволяє демократизувати доступ до даних та трансформувати їх у стратегічний актив підприємства. З'ясовано, що сучасні технології дозволяють не просто накопичувати дані, а й видобувати з них цінність, необхідну для стратегічного планування та оптимізації бізнес-процесів.

2. Охарактеризовано та конкретизовано методи візуалізації великих даних у бізнес-аналітиці. Систематизовано підходи до графічного представлення інформації залежно від типу аналітичного завдання: від порівняльного аналізу за допомогою гістограм до виявлення кореляцій через діаграми розсіювання. Визначено, що для роботи з надвеликими масивами даних найбільш ефективними є методи агрегації, семантичного зумування та інтелектуальної фільтрації. Обґрунтовано доцільність використання інтерактивних дашбордів, які, на відміну від статичних звітів, забезпечують можливість сценарного аналізу та глибокої деталізації до рівня окремих

транзакцій без втрати контексту. Інтеграція методів візуалізації з алгоритмами машинного навчання дозволяє автоматизувати процес пошуку інсайтів, роблячи аналітику доступною для широкого кола користувачів.

3. Розроблено модель великих даних в бізнес-процесах. Спроектовано та реалізовано оптимізовану семантичну модель даних за схемою «Галактика», що дозволило інтегрувати різноманітні бізнес-процеси (продажі, товарні запаси, клієнтська активність) в єдиний інформаційний простір. Для забезпечення цілісності та якості даних розроблено автоматизований ETL-конвеєр мовою M, який виконує очищення, нормалізацію та типізацію вхідних потоків інформації. Розроблена модель підтримує роботу в режимах імпорту та прямого запиту, що забезпечує її масштабованість та здатність обробляти терабайти транзакційних даних зі збереженням високої швидкості відгуку. Використання спільних вимірів дозволило проводити перехресний аналіз різних аспектів діяльності компанії, забезпечуючи цілісність аналітики.

4. Розроблено візуалізацію великих даних у бізнес-аналітиці засобами Power BI. Практично реалізовано комплексну аналітичну систему, що включає інтерактивні панелі для стратегічного моніторингу та операційного аналізу. В систему інтегровано алгоритми штучного інтелекту, які автоматизують процес діагностичної аналітики та виявлення прихованих закономірностей. Функціональність візуалізацій розширено за допомогою динамічних метрик мовою DAX, що дозволило реалізувати складні розрахунки часових рядів та порівняльний аналіз ефективності (YoY Growth). Результати тестування підтвердили, що розроблена система значно скорочує час на підготовку звітності та підвищує точність управлінських рішень. Розроблений інтерфейс відповідає сучасним стандартам UX/UI, забезпечуючи інтуїтивно зрозумілу навігацію та ефективну взаємодію з даними.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Про хмарні послуги: Закон України від 17.02.2022 р. № 2075-IX. Відомості Верховної Ради України. 2022. № 24. Ст. 132.
2. Бабич С. В. Інтелектуальний аналіз даних: навч. посіб. Київ: КПІ ім. Ігоря Сікорського, 2021. 286 с.
3. Веретенникова Г. Б. Методи та моделі аналізу великих даних: конспект лекцій. Харків: ХНЕУ ім. С. Кузнеця, 2020. 145 с.
4. Глибовець М. М., Олецький О. В. Штучний інтелект: підручник. Київ: КМ-Т, 2020. 436 с.
5. Григорків В. С., Ярошенко О. І. Бізнес-аналітика: навч. посіб. Чернівці: Чернівецький нац. ун-т, 2022. 352 с.
6. Данильчук Г. Б. Візуалізація даних у бізнес-аналітиці: принципи та інструментарій. *Економіка та суспільство*. 2023. № 48. DOI: <https://doi.org/10.32782/2524-0072/2023-48-52>.
7. Єрємін Н. В., Ситник Н. В. Використання Power BI для аналізу та візуалізації даних у бізнесі. *Цифрова економіка та економічна безпека*. 2022. № 1. С. 45–51.
8. Ковальчук О. Я. Системи підтримки прийняття рішень: навч. посіб. Тернопіль: ТНПУ ім. В. Гнатюка, 2021. 210 с.
9. Криволапов С. О. Технології Big Data в управлінні сучасним підприємством. *Бізнес Інформ*. 2021. № 5. С. 338–344.
10. Кузьмінський Ю. Б. Хмарні технології в бізнес-аналітиці: переваги та ризики. Вісник Хмельницького національного університету. *Економічні науки*. 2023. № 2(316). С. 120–125.
11. Литвин В. В. Системи та методи прийняття рішень: навч. посіб. Львів: Видавництво Львівської політехніки, 2020. 316 с.
12. Різник В. В., Яцишин А. В. Технології Business Intelligence: аналіз ринку та перспективи розвитку. *Інформаційні технології і засоби навчання*. 2020. Т. 77, № 3. С. 312–325.

13. Шаховська Н. Б. Системи штучного інтелекту: навч. посіб. Львів: Видавництво Львівської політехніки, 2021. 392 с.
14. Шевчук І. Б. Big Data в економіці: основи та інструментарій: монографія. Львів: ЛНУ імені Івана Франка, 2021. 360 с.
15. Шоваг І. М., Карабін О. Й. Принципи проєктування та розробки комп'ютерної гри в Unreal Engine. *Сучасні цифрові технології та інноваційні методики навчання: досвід, тенденції, перспективи* : матеріали XIII Міжнародної науково-практичної інтернет-конференції (м. Тернопіль, 5 квітня, 2024 р.). Тернопіль : ТНПУ ім. В. Гнатюка, 2024. С. 253–255.
16. Шоваг І. М., Карабін О. Й. Інноваційні підходи до організації виробничої практики у сфері інформаційних технологій. *Шлях в науку: перші кроки* : матеріали III Всеукраїнської науково-практичної конференції молодих учених та здобувачів освіти (м. Тернопіль, 9 квітня 2025 р.). ТНПУ. хім.-біо., 2025. С. 230–233.
17. Шоваг І. М., Карабін О. Й. Інструменти та методи візуалізації Big Data у бізнес-аналітиці. *Сучасні цифрові технології та інноваційні методики навчання: досвід, тенденції, перспективи* : матеріали XV Міжнародної науково-практичної інтернет-конференції (м. Тернопіль, 10 квітня, 2025 р.). Тернопіль : ТНПУ ім. В. Гнатюка, 2025. С. 323–325.
18. Шоваг І. М., Карабін О. Й. Технологічні інновації у візуалізації візуалізації Big Data в бізнес-аналітиці. *Сучасні цифрові технології та інноваційні методики навчання: досвід, тенденції, перспективи* : матеріали XVI Міжнародної науково-практичної інтернет-конференції (м. Тернопіль, 6-7 листопада, 2025 р.). Тернопіль : ТНПУ ім. В. Гнатюка, 2025. С. 343–345
19. Clark D. Beginning Power BI: A Practical Guide to Self-Service Data Analytics. 3rd ed. New York: Apress, 2023. 368 p.
20. Coles M., Roff C. The Big Data Ecosystem: Managing Volume, Velocity, and Variety in the Cloud. London: Kogan Page, 2021. 280 p.

21. Deckler G. Learn Power BI: A comprehensive guide to building data models and creating visualization reports. *2nd ed. Birmingham: Packt Publishing, 2022. 530 p.*
22. Ferrari A., Russo M. The Definitive Guide to DAX: Business Intelligence for Microsoft Power BI, SQL Server Analysis Services, and Excel. 2nd ed. Redmond: Microsoft Press, 2020. 768 p.
23. Franks B. Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics. Hoboken: Wiley, 2020. 336 p.
24. Gartner Magic Quadrant for Analytics and Business Intelligence Platforms 2024. Gartner. URL: <https://www.gartner.com/en/documents/5519595> (accessed: 10.12.2024).
25. Inmon W. H. Building the Data Warehouse. 4th ed. Indianapolis: Wiley, 2005. 576 p.
26. Kimball R., Ross M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3rd ed. Indianapolis: Wiley, 2013. 600 p.
27. Knaflitz C. N. Storytelling with Data: A Data Visualization Guide for Business Professionals. Hoboken: Wiley, 2015. 288 p.
28. Microsoft Learn. AdventureWorks sample databases. URL: <https://learn.microsoft.com/en-us/sql/samples/adventureworks-install-configure> (accessed: 05.12.2024).
29. Microsoft Learn. Power BI Documentation: Data Visualization and Modeling. URL: <https://learn.microsoft.com/en-us/power-bi/> (accessed: 08.12.2024).
30. Powell B. Microsoft Power BI Cookbook: Gain expertise in Power BI to enhance your business intelligence analysis. 2nd ed. Birmingham: Packt Publishing, 2021. 624 p.
31. Provost F., Fawcett T. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. Sebastopol: O'Reilly Media, 2013. 414 p.

32. Wexler S. The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios. Hoboken: Wiley, 2017. 448 p.

33. Wilke C. O. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. Sebastopol: O'Reilly Media, 2019. 390 p.

ДОДАТКИ

Додаток А

Лістинг програмного коду системи

А.1. Скрипт попередньої обробки даних (мова M)

```
let
    Source = Sql.Database("localhost", "AdventureWorksDW2020"),
    DimProduct = Source[Schema="dbo",Item="DimProduct"][Data],
    #"Merged Subcategory" = Table.NestedJoin(DimProduct, {"ProductSubcategoryKey"},
        DimProductSubcategory, {"ProductSubcategoryKey"}, "Subcategory", JoinKind.LeftOuter),
    #"Expanded Subcategory" = Table.ExpandTableColumn(#"Merged Subcategory", "Subcategory",
        {"EnglishProductSubcategoryName", "ProductCategoryKey"}, {"SubcategoryName", "ProductCategoryKey"}),
    #"Merged Category" = Table.NestedJoin(#"Expanded Subcategory", {"ProductCategoryKey"},
        DimProductCategory, {"ProductCategoryKey"}, "Category", JoinKind.LeftOuter),
    #"Expanded Category" = Table.ExpandTableColumn(#"Merged Category", "Category",
        {"EnglishProductCategoryName"}, {"CategoryName"}),
    #"Removed Columns" = Table.RemoveColumns(#"Expanded Category",
        {"ProductSubcategoryKey", "ProductCategoryKey"})
in
    #"Removed Columns"
```

Рис. 1. Лістинг коду мовою M, що демонструє операцію підготовки даних

А.2. Ключові міри аналізу (мова DAX)

```
Sales Previous Period =
VAR SelectedYear = SELECTEDVALUE('Calendar'[Year])
VAR SelectedQuarter = SELECTEDVALUE('Calendar'[QuarterNumber])
VAR SelectedMonth = SELECTEDVALUE('Calendar'[Month])
VAR SelectedDay = SELECTEDVALUE('Calendar'[Day])
RETURN
SWITCH(
    TRUE(),
    NOT(ISBLANK(SelectedDay)),
    IF(
        SelectedMonth = 1 && SelectedDay = 1,
        CALCULATE([Sales],
            FILTER(
                ALL('Calendar'),
                'Calendar'[Year] = SelectedYear - 1 &&
                'Calendar'[Month] = 12 &&
                'Calendar'[Day] = 31
            )
        ),
        CALCULATE([Sales], PREVIOUSDAY('Calendar'[Value])
    ),
    NOT(ISBLANK(SelectedMonth)),
    IF(
        SelectedMonth = 1,
        CALCULATE([Sales],
            FILTER(
                ALL('Calendar'),
                'Calendar'[Year] = SelectedYear - 1 &&
                'Calendar'[Month] = 12
            )
        ),
        CALCULATE([Sales], PREVIOUSMONTH('Calendar'[Value])
    )
)
```

```

    )
  ),
  NOT(ISBLANK(SelectedQuarter)),
  IF(
    SelectedQuarter = 1,
    CALCULATE([Sales],
      FILTER(
        ALL('Calendar'),
        'Calendar'[Year] = SelectedYear - 1 &&
        'Calendar'[QuarterNumber] = 4
      )
    ),
    IF(
      SelectedYear = 2024 && SelectedQuarter = 1,
      CALCULATE([Sales],
        FILTER(
          ALL('Calendar'),
          'Calendar'[Year] = 2023 &&
          'Calendar'[QuarterNumber] = 4
        )
      ),
      CALCULATE([Sales], PREVIOUSQUARTER('Calendar'[Value]))
    )
  ),
  NOT(ISBLANK(SelectedYear)),
  IF(
    SelectedYear = 2024,
    CALCULATE([Sales],
      FILTER(
        ALL('Calendar'),
        'Calendar'[Year] = 2023
      )
    ),
    CALCULATE([Sales], PREVIOUSYEAR('Calendar'[Value]))
  )
)
)

```

Рис. 2. Динамічне порівняння продажів з минулим періодом

```

Sales Reference Label =
VAR Current_ = [Sales]
VAR Past_ = [Sales Previous Period]
VAR Growth_ = DIVIDE(Current_ - Past_, Past_)
VAR NumOfDigits =
  IFERROR(
    LEN(
      CONVERT(
        INT(
          [Sales]
        ), STRING ))
    ,0)
VAR Suffix =
  ".0" &
  SWITCH(
    TRUE( ),
    NumOfDigits >= 13, "T",
    NumOfDigits >= 10, "B",
    NumOfDigits >= 7, "M",
    NumOfDigits >= 4, "K"
  )
VAR Commas =

```

```

REPT(",",
  SWITCH(
    TRUE(),
    NumOfDigits >= 13, "4",
    NumOfDigits >= 10, "3",
    NumOfDigits >= 7, "2",
    NumOfDigits >= 4, "1"
  )
)
)
VAR Format_ = "+#,##0" & Commas & Suffix & ";-#,##0" & Commas & Suffix
return
IF (Growth_ <> 0,
  "" & FORMAT(Current_ - Past_, Format_)
  &
  FORMAT(Growth_, " | 0.0 % ▲; | 0.0 % ▼;") & "", " ")

```

Рис. 3. Динамічне форматування продажів цифрових значень як текстові

```

Sparkline Sales =
VAR LineColour = "%2315CAB6"
VAR PointColour = "%23ffffff"
VAR Defs = "<defs>
  <linearGradient id='grad' x1='0' y1='25' x2='0' y2='50' gradientUnits='userSpaceOnUse'>
    <stop stop-color=' %2315CAB6' offset='0' />
    <stop stop-color=' %2315CAB6' offset='0.3' />
    <stop stop-color=' %23fbfbfd' offset='1' />
  </linearGradient>
</defs>"
VAR XMinDate = MIN('Calendar'[Start of Month])
VAR XMaxDate = MAX('Calendar'[Start of Month])
// Obtain overall min and overall max measure values when evaluated for each date
VAR YMinValue = MINX(Values('Calendar'[Start of Month]),CALCULATE([Sales]))
VAR YMaxValue = MAXX(Values('Calendar'[Start of Month]),CALCULATE([Sales]))
// Build table of X & Y coordinates
VAR SparklineTable = ADDCOLUMNS(
  SUMMARIZE('Calendar', 'Calendar'[Start of Month]),
  "X", INT(150 * DIVIDE('Calendar'[Start of Month] - XMinDate, XMaxDate - XMinDate)),
  "Y", INT(50 * DIVIDE([Sales] - YMinValue, YMaxValue - YMinValue))
)
// Concatenate X & Y coordinates to build the sparkline
VAR Lines = CONCATENATEX(SparklineTable, [X] & ", " & 50 - [Y], " ", 'Calendar'[Start of Month])
// Last data points on the line
VAR LastSparkYValue = MAXX( FILTER(SparklineTable, 'Calendar'[Start of Month] = XMaxDate), [Y])
VAR LastSparkXValue = MAXX( FILTER(SparklineTable, 'Calendar'[Start of Month] = XMaxDate), [X])
// Add to SVG, and verify Data Category is set to Image URL for this measure
VAR SVGImageURL =
  "data:image/svg+xml;utf8," &
  --- gradient---
  "<svg xmlns='http://www.w3.org/2000/svg' x='0px' y='0px' viewBox='-7 -7 164 64'>" & Defs &
  "<polyline fill='url(#grad)' fill-opacity='0.3' stroke='transparent'
  stroke-width='0' points=' 0 50 " & Lines &
  " 150 150 Z />" &
  --- Lines---
  "<polyline
  fill='transparent' stroke='" & LineColour & "'
  stroke-linecap='round' stroke-linejoin='round'
  stroke-width='3' points=' " & Lines &
  " />" &
  --- Last Point---
  "<circle cx='" & LastSparkXValue & "' cy='" & 50 - LastSparkYValue & "' r='4' stroke='" & LineColour & "
  " stroke-width='3' fill='" & PointColour & "' />" &
  "</svg>"
RETURN SVGImageURL

```

Рис. 4. Динамічна векторна візуалізація лінійного графіку продажів

Інтерфейс розробленого звіту

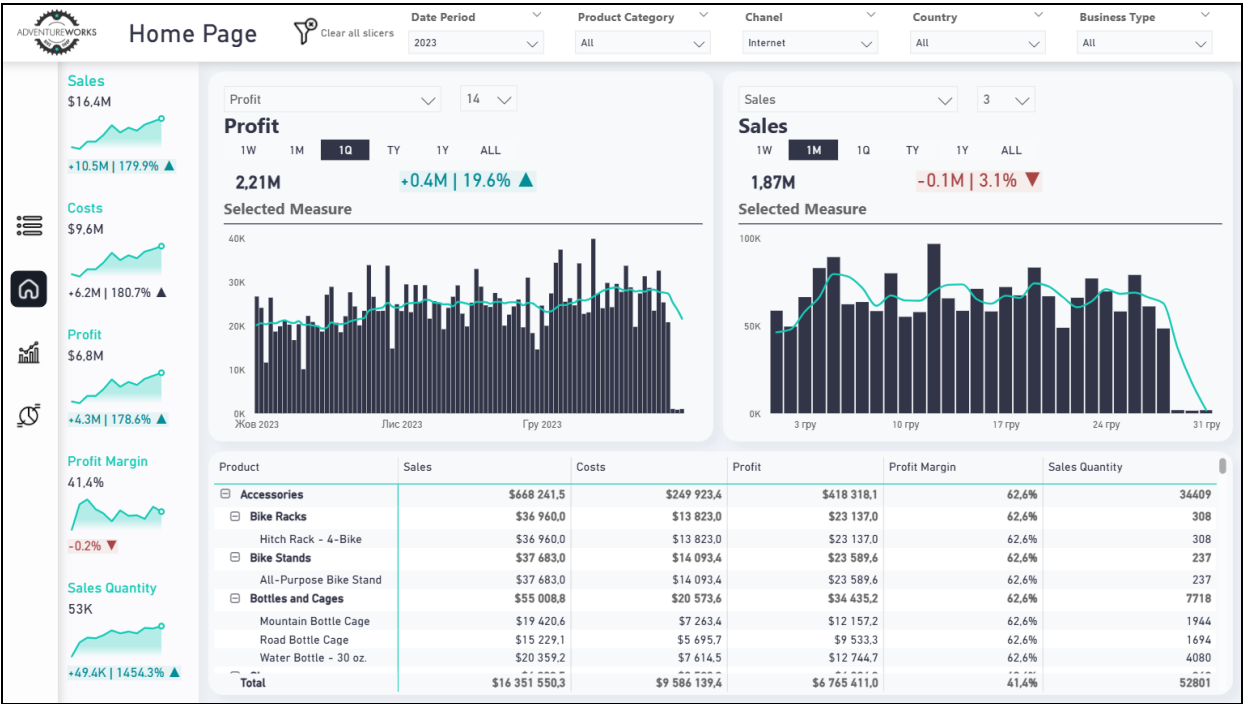


Рис. 1. Головна панель (Home Page) аналітичного звіту

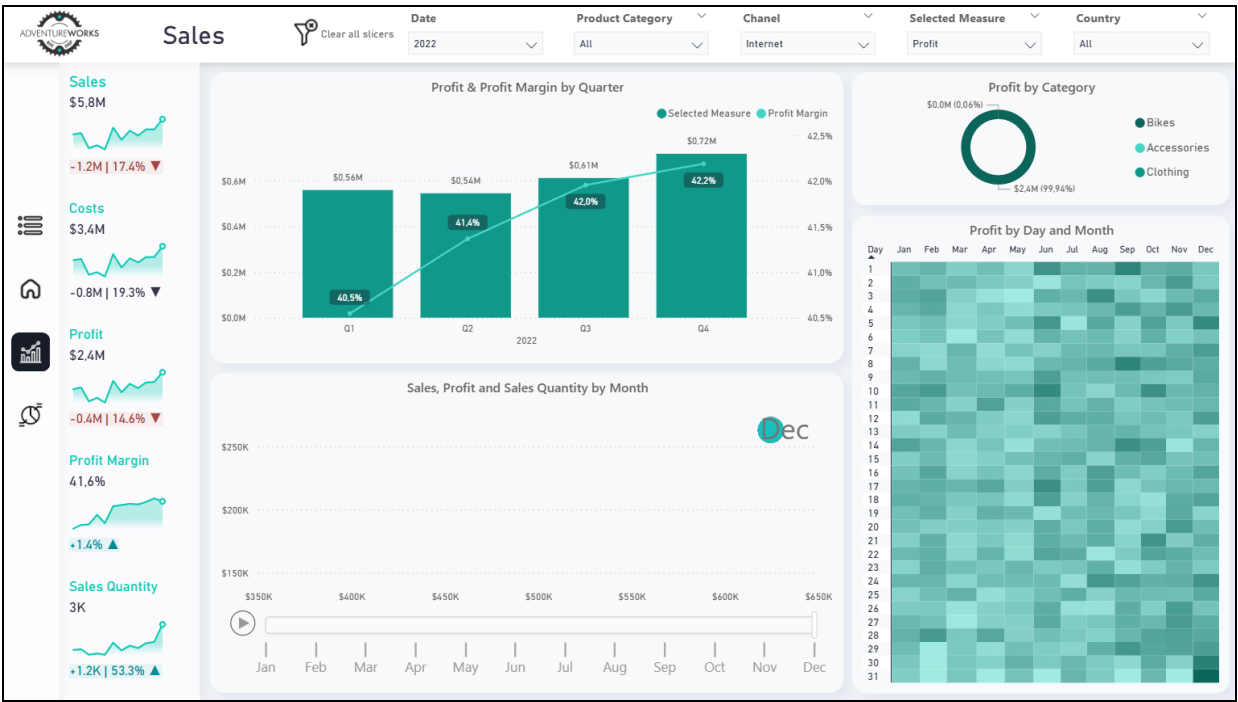


Рис. 2. Сторінка детального аналізу продажів (Sales Analysis)

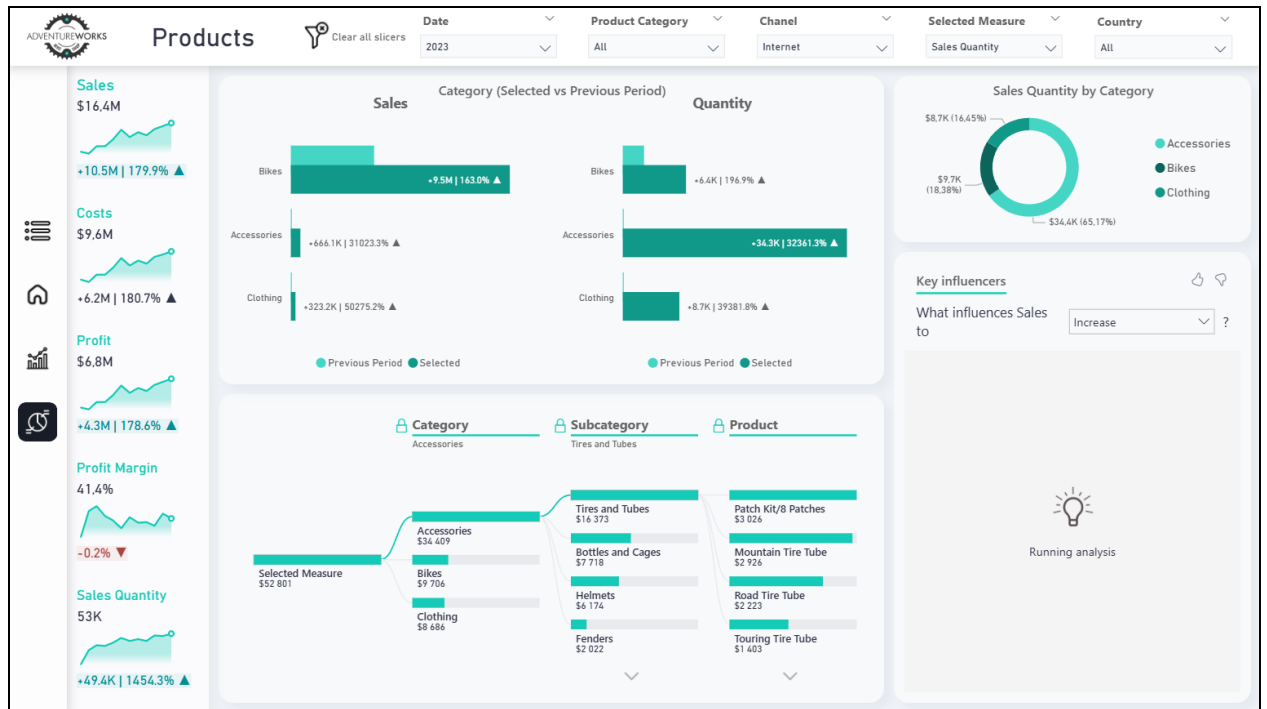


Рис. 3. Сторінка AI-інсайтів та продуктового аналізу

Рис. 1. Реалізація схеми даних типу «Галактика» у Power BI

Інтерфейс SSMS з базою даних

FileEditViewQueryGitProjectToolsExtensionsWindowHelp

FileEditViewQueryGitProjectToolsExtensionsWindowHelp

FileEditViewQueryGitProjectToolsExtensionsWindowHelp

AdventureWorksDW

AdventureWorksDW

AdventureWorksDW

Object Explorer

Connect

DESKTOP-FSQQA65 (SQL Server 17.0.10007 - DESKTOP-...)

Databases

System Databases

Database Snapshots

AdventureWorks2005

AdventureWorksDW2008

Database Diagrams

Tables

System Tables

FileTables

External Tables

Graph Tables

dboAdventureWorksDWBuildVersion

dboAdventureLog

dboAccount

dboCurrency

dboCustomer

dboDate

dboDepartmentGroup

dboEmployee

dboGeography

dboOrganization

dboProduct

dboProductCategory

dboProductSubcategory

dboPromotion

dboReseller

dboSalesReason

dboSalesTerritory

dboScenario

dboFactAdditionalInformationProduct

dboFactCallCenter

dboFactCurrencyRate

dboFactFinance

dboFactInternetSales

dboFactInternetSalesReason

dboFactProductInventory

dboFactResellerSales

dboFactSalesQuota

dboFactSalesResponse

dboNewFactCurrencyRate

dboProspectiveBuyer

Dropped Ledger Tables

Views

External Resources

Synonyms

Programmability

Service Broker

Spinecase

SQLQuery1.sql (730)

SELECT TOP (1000) ([ProductKey], [OrderDateKey], [DueDateKey], [ShipDateKey], [CustomerKey], [PromotionKey], [CurrencyKey], [SalesTerritoryKey], [SalesOrderNumber], [SalesOrderLineNumber], [RevisionNumber], [OrderQuantity], [UnitPrice], [ExtendedAmount], [UnitPriceDiscountPct], [DiscountAmount], [ProductStandardKey]) FROM [AdventureWorksDW].[FactSalesReason]

100%No insert found

ResultsMessages

ProductKeyOrderDateKeyDueDateKeyShipDateKeyCustomerKeyPromotionKeyCurrencyKeySalesTerritoryKeySalesOrderNumberSalesOrderLineNumberRevisionNumberOrderQuantityUnitPriceExtendedAmountUnitPriceDiscountPctDiscountAmountProductStandardKey

1	310	20211229	20220110	20220105	21768	1	19	6	SO43687	1	1	1	3578.27	3578.27	0	0	2171.2942
2	346	20211229	20220110	20220105	20389	1	39	7	SO43688	1	1	1	3399.99	3399.99	0	0	1912.1544
3	346	20211229	20220110	20220105	20383	1	100	1	SO43689	1	1	1	3399.99	3399.99	0	0	1912.1544
4	336	20211229	20220110	14501	1	100	4	SO43700	1	1	1	699.0982	699.0982	0	0	413.1463	
5	346	20211229	20220110	11003	1	6	9	SO43701	1	1	1	3399.99	3399.99	0	0	1912.1544	
6	311	20211230	20220111	27645	1	100	4	SO43702	1	1	1	3578.27	3578.27	0	0	2171.2942	
7	310	20211230	20220111	16424	1	6	9	SO43703	1	1	1	3578.27	3578.27	0	0	2171.2942	
8	351	20211230	20220111	11005	1	6	9	SO43704	1	1	1	3374.99	3374.99	0	0	1886.0944	
9	344	20211230	20220111	11011	1	6	9	SO43705	1	1	1	3399.99	3399.99	0	0	1912.1544	
10	312	20211231	20220112	27621	1	100	4	SO43706	1	1	1	3578.27	3578.27	0	0	2171.2942	
11	312	20211231	20220112	27616	1	100	4	SO43707	1	1	1	3578.27	3578.27	0	0	2171.2942	
12	330	20211231	20220107	20042	1	98	10	SO43708	1	1	1	699.0982	699.0982	0	0	413.1463	
13	313	20211231	20220112	16351	1	6	9	SO43709	1	1	1	3578.27	3578.27	0	0	2171.2942	
14	314	20211231	20220107	16517	1	6	9	SO43710	1	1	1	3578.27	3578.27	0	0	2171.2942	
15	314	20220101	20220113	27606	1	100	1	SO43711	1	1	1	3578.27	3578.27	0	0	2171.2942	
16	311	20220101	20220113	13513	1	29	8	SO43712	1	1	1	3578.27	3578.27	0	0	2171.2942	
17	310	20220102	20220114	20220109	27601	1	100	4	SO43713	1	1	1	3578.27	3578.27	0	0	2171.2942
18	311	20220102	20220114	13551	1	98	10	SO43714	1	1	1	3578.27	3578.27	0	0	2171.2942	
19	314	20220102	20220114	16403	1	6	9	SO43715	1	1	1	3578.27	3578.27	0	0	2171.2942	
20	311	20220102	20220114	16529	1	6	9	SO43716	1	1	1	3578.27	3578.27	0	0	2171.2942	
21	336	20220102	20220114	20249	1	6	9	SO43717	1	1	1	699.0982	699.0982	0	0	413.1463	
22	311	20220103	20220115	20220110	27604	1	100	1	SO43718	1	1	1	3578.27	3578.27	0	0	2171.2942
23	312	20220103	20220115	20220110	27612	1	100	4	SO43719	1	1	1	3578.27	3578.27	0	0	2171.2942
24	311	20220103	20220115	13264	1	29	8	SO43720	1	1	1	3578.27	3578.27	0	0	2171.2942	
25	310	20220103	20220115	20220110	13590	1	98	10	SO43721	1	1	1	3578.27	3578.27	0	0	2171.2942
26	312	20220104	20220116	13561	1	98	10	SO43722	1	1	1	3578.27	3578.27	0	0	2171.2942	
27	332	20220104	20220116	14530	1	100	1	SO43723	1	1	1	699.0982	699.0982	0	0	413.1463	
28	311	20220104	20220116	16520	1	6	9	SO43724	1	1	1	3578.27	3578.27	0	0	2171.2942	
29	311	20220105	20220117	13208	1	29	8	SO43725	1	1	1	3578.27	3578.27	0	0	2171.2942	
30	332	20220105	20220117	14560	1	100	4	SO43726	1	1	1	699.0982	699.0982	0	0	413.1463	
31	311	20220105	20220117	16507	1	6	9	SO43727	1	1	1	3578.27	3578.27	0	0	2171.2942	

Query executed successfully.

DESKTOP-FSQQA65 (17.0 RTM)

DESKTOP-FSQQA65 (AdventureWorksDW2008)

00:00:00

Row: 1, Col: 1

1000 rows

Рис. 1. Інтерфейс SSMS з даними з таблиці FactInternetSales