

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ
ЛЬВІВСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ ВНУТРІШНІХ СПРАВ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ УПРАВЛІННЯ, ПСИХОЛОГІЇ
ТА БЕЗПЕКИ**

Кафедра інформаційних технологій

**АНАЛІЗ ВІДКРИТИХ ІНФОРМАЦІЙНИХ РЕСУРСІВ ЗАСОБАМИ
ХМАРНИХ ТЕХНОЛОГІЙ**

Кваліфікаційна робота
здобувача вищої освіти
4 курсу денної форми навчання
Станіслава ВОЗНЮКА

Науковий керівник:
доцент, канд. фіз.-мат. наук_
Тетяна МАГЕРОВСЬКА

Рецензент:

Кваліфікаційна робота допущена до захисту
«___» _____ 2026 р., протокол № _____

Завідувач кафедри інформаційних технологій
_____ Олег ЗАЧЕК
(підпис)

Львів
2026

СПИСОК УМОВНИХ СКОРОЧЕНЬ

API	Application Programming Interface
DBaaS	Database-as-a-service
PaaS	Platform as a service
IaaS	Infrastructure as a service
SaaS	Software as a service
MPP	Massively Parallel Processing
SQL	Structured query language (мова структурованих запитів)
БД	База даних
ПЗ	Програмне забезпечення
СУБД	Система управління базами даних

АНОТАЦІЯ

Вознюк С. Аналіз відкритих інформаційних ресурсів засобами хмарних технологій. – Рукопис.

Дослідження на здобуття освітнього ступеня «бакалавр» за спеціальністю F6 – «інформаційні системи та технології». – Львівський державний університет внутрішніх справ, МВС України. Львів. 2026.

У кваліфікаційній роботі спроектовано та розроблено технологію аналізу відкритих інформаційних ресурсів із використанням хмарних технологій. З цією метою у роботі: проаналізовано сучасний стан розвитку відкритих інформаційних ресурсів та визначено їх основні характеристики; досліджено технологічні можливості хмарних обчислень у контексті опрацювання й аналізу інформаційних ресурсів; досліджено перспективні напрями еволюції відкритих інформаційних ресурсів і технологій їх аналізу; оцінено функціональні можливості хмарних платформ для зберігання, опрацювання, аналітики та візуалізації відкритих даних; обрано інструментарій для дослідження; розроблено технологічну модель організації процесу аналізу відкритих інформаційних ресурсів у хмарному середовищі; продемонстровано ефективність моделі організації процесу аналізу відкритих інформаційних ресурсів з використанням хмарних сервісів.

За результатами дослідження систематизовано основні проблеми використання хмарних сервісів для аналізу відкритих інформаційних ресурсів з урахуванням якості даних, масштабованості та безпеки; запропоновано узагальнений підхід до організації процесу аналізу відкритих інформаційних ресурсів у хмарному середовищі, що враховує специфіку різномірних джерел даних та вимоги до їх інтеграції; обґрунтовано доцільність використання хмарних технологій як базової інфраструктури для аналітичного опрацювання відкритих інформаційних ресурсів.

Ключові слова: відкриті інформаційні ресурси, аналіз даних, хмарні технології, Python, BigQuery, Looker Studio

ABSTRACT

Vozniuk S. Analysis of Open Information Resources Using Cloud Technologies. – Manuscript.

Thesis for obtaining the bachelor's degree in specialty F6 – Information Systems and Technologies. – Lviv State University of Internal Affairs, Ministry of Internal Affairs of Ukraine. Lviv, 2026.

The qualification work designs and develops a technology for analyzing open information resources using cloud technologies. For this purpose, the study: analyzes the current state of development of open information resources and identifies their main characteristics; investigates the technological capabilities of cloud computing in the context of processing and analyzing information resources; examines promising directions for the evolution of open information resources and technologies for their analysis; evaluates the functional capabilities of cloud platforms for storage, processing, analytics, and visualization of open data; selects appropriate tools for the research; develops a technological model for organizing the process of analyzing open information resources in a cloud environment; and demonstrates the effectiveness of the proposed model for organizing the analysis process of open information resources using cloud services.

Based on the research results, the main challenges of using cloud services for analyzing open information resources are systematized, taking into account data quality, scalability, and security. A generalized approach to organizing the process of analyzing open information resources in a cloud environment is proposed, considering the specifics of heterogeneous data sources and the requirements for their integration. The feasibility of using cloud technologies as a basic infrastructure for the analytical processing of open information resources is substantiated.

Keywords: open information resources, data analysis, cloud technologies, Python, BigQuery, Looker Studio.

ЗМІСТ

ВСТУП.....	8
Розділ 1.ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ ДОСЛІДЖЕННЯ.....	13
1.1. Відкриті дані: поняття та їх характеристики.....	13
1.2. Аналіз і візуалізація відкритих даних хмарними сервісами.....	16
1.3. Дашборт як засіб аналізу даних.....	17
1.4. Хмарні сховища даних.....	23
1.5. Перспективні напрями еволюції відкритих даних і технологій їх аналізу...28	
Розділ 2.ОБРАННЯ ТЕХНОЛОГІЙ ТА ІНСТРУМЕНТАРІЮ ДЛЯ АНАЛІЗУ ВІДКРИТИХ ІНФОРМАЦІЙНИХ РЕСУРСІВ.....	32
2.1. GoogleLookerStudio.....	32
2.2. GoogleBigQuery.....	38
2.3. Python.....	42
2.3. Визначення технології аналізу відкритих інформаційних ресурсів.....	46
Розділ 3.ПРАКТИЧНЕ ЗАСТОСУВАННЯ ТЕХНОЛОГІЇ АНАЛІЗУ ВІДКРИТИХ ІНФОРМАЦІЙНИХ РЕСУРСІВ ЗА ДОПОМОГОЮ ХМАРНИХ СЕРВІСІВ.....	49
3.1. Підготовка середовищ для аналізу інформаційних ресурсів.....	49
3.2. Проведення процедури отримання відкритих інформаційних ресурсів для подальшого аналізу.....	50
3.3. Проектування дашборду для аналізу відкритих інформаційних ресурсів...56	
ВИСНОВКИ.....	59
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	61

ВСТУП

Стрімке зростання обсягів відкритих інформаційних ресурсів є однією з характерних ознак сучасного цифрового суспільства. Державні установи, органи місцевого самоврядування, міжнародні організації, наукові інституції та громадські ініціативи активно публікують дані у відкритому доступі. Такі ресурси охоплюють соціально-економічні показники, статистичну інформацію, результати досліджень, освітні матеріали, геопросторові дані тощо. Водночас сам факт відкритості даних ще не гарантує їх ефективного використання.

На практиці виникає низка проблем, пов'язаних із пошуком, опрацюванням, аналізом і візуалізацією відкритих інформаційних ресурсів. Дані часто подаються у різних форматах, мають неоднорідну структуру, відрізняються за якістю, повнотою та актуальністю. Значні обсяги інформації ускладнюють її опрацювання традиційними засобами, особливо в умовах обмежених технічних ресурсів окремих установ чи дослідників. У результаті потенціал відкритих даних залишається реалізованим лише частково.

У цьому контексті особливого значення набувають хмарні технології, які забезпечують масштабованість обчислювальних ресурсів, доступ до інструментів аналітики, зберігання великих масивів даних та спільну роботу користувачів незалежно від їхнього місця перебування. Хмарні сервіси дозволяють автоматизувати процеси опрацювання інформації, здійснювати складні аналітичні операції та створювати інтерактивні візуалізації без потреби у розгортанні власної інфраструктури. Проте, питання ефективного та доцільного застосування таких технологій для аналізу відкритих інформаційних ресурсів потребує ґрунтовного дослідження.

Незважаючи на активний розвиток хмарних платформ та зростання кількості відкритих даних, залишається невирішеною проблема

комплексного підходу до їх інтеграції: від вибору інструментів до забезпечення безпеки, надійності, економічної доцільності та відповідності нормативним вимогам. Також потребують уточнення методичні засади організації процесу аналізу відкритих ресурсів у хмарному середовищі з урахуванням специфіки галузі застосування.

Викладеним обумовлюється **актуальність дослідження**: необхідність розроблення та обґрунтування підходів до ефективного аналізу відкритих інформаційних ресурсів із використанням можливостей хмарних технологій. Розв'язання цієї проблеми сприятиме підвищенню якості управлінських рішень, розвитку наукових досліджень, забезпеченню прозорості діяльності установ та розширенню практичного використання відкритих даних у різних сферах суспільного життя.

Аналіз останніх досліджень і публікацій. У науковій літературі хмарні технології розглядаються як ключовий інструмент для ефективного опрацювання великих обсягів інформаційних ресурсів (big data). Так, у статті [1] автори проводять огляд можливостей хмарних платформ у контексті аналітики великих даних, підкреслюючи, що традиційні інструменти опрацювання вже не справляються з обсягами та різноманітністю інформаційних ресурсів, тоді як хмарні сервіси забезпечують масштабованість і зниження технічних бар'єрів для аналізу (big data analytics) у режимі реального часу.

Цей висновок переформується з роботою Ф. Бериша та ін. [2], у якій хмарні обчислення розглядаються як ефективний механізм для опрацювання, зберігання й аналізу великих обсягів інформаційних ресурсів із різноманітних джерел, що є важливою умовою для аналізу відкритих даних.

Українські дослідники відзначають, що відкриті дані (open data) є складовою частиною відкритої науки й інформаційної відкритості. Так, аналіз теоретичних аспектів розвитку інформаційних комунікацій демонструє, що поняття відкритих даних охоплює не лише доступ до відомостей, а й

безперешкодну можливість їх аналізу та повторного використання в інших дослідженнях та системах [3].

Це важливо для розуміння суті відкритих інформаційних ресурсів як предмета аналізу, адже саме доступність і стандартизованість даних визначають можливості їх автоматизованої обробки у хмарних середовищах.

Ряд українських статей розглядають аспекти безпеки, методології та практичного застосування хмарних технологій. Наприклад, А. Жилін, А. Дівіцький та А. Козачок у роботі [4] аналізують проблеми захисту інформаційних ресурсів при використанні хмарних обчислень, що є важливою частиною процесу аналізу відкритих даних, особливо коли мова йде про обробку чутливої інформації чи інтеграцію даних з різних джерел.

В. Івкова та Р. Банах [5] у своїй роботі пропонують методологію дослідження захищеності хмарних середовищ за допомогою OSINT-інструментів, що показує практичні підходи до оцінки ризиків та аналізу вразливостей у відкритих джерелах даних.

У міжнародному контексті існують підтримувані наукові журнали та праці, які охоплюють широке коло тем, пов'язаних із хмарними обчисленнями та обробкою даних. Журнал *Journal of Big Data* [6], наприклад, включає дослідження з таких напрямів, як пошук, зберігання, аналіз, архітектури обробки даних, інструменти машинного навчання, а також хмарні платформи для масштабної аналітики, що створює важливу базу для розуміння сучасних практик у цій сфері. Хоча це джерело не присвячене виключно відкритим даним, воно демонструє важливість хмарних обчислень як фундаментальної складової інфраструктури для великих даних та аналітичних платформ – включно з відкритими типами даних.

Метою роботи є розроблення практичних підходів до ефективного аналізу відкритих інформаційних ресурсів із використанням хмарних технологій.

Для досягнення поставленої мети слід виконати наступні **завдання**:

- проаналізувати сучасний стан розвитку відкритих інформаційних ресурсів та визначити їх основні характеристики;
- дослідити технологічні можливості хмарних обчислень у контексті опрацювання й аналізу інформаційних ресурсів;
- дослідити перспективні напрями еволюції відкритих даних і технологій їх аналізу;
- оцінити функціональні можливості хмарних платформ для зберігання, опрацювання, аналітики та візуалізації відкритих даних й обрати інструментарій для дослідження;
- розробити технологічну модель організації процесу аналізу відкритих інформаційних ресурсів у хмарному середовищі;
- продемонструвати ефективність моделі організації процесу аналізу відкритих інформаційних ресурсів з використанням хмарних сервісів.

Об'єктом дослідження у роботі є процеси збирання, зберігання, опрацювання й аналізу відкритих інформаційних ресурсів у цифровому середовищі.

Предметом дослідження є методи, моделі та інструменти аналізу відкритих інформаційних ресурсів, що реалізуються із застосуванням хмарних технологій.

Методи досліджень. У дослідженні використано комплекс взаємопов'язаних теоретичних і прикладних методів. Зокрема, застосовано аналіз, синтез і узагальнення наукових джерел для визначення сучасного стану розвитку відкритих інформаційних ресурсів та хмарних технологій; порівняльний аналіз – для оцінювання можливостей різних хмарних платформ (додатків); моделювання – для розроблення технологічної схеми організації процесу аналізу відкритих інформаційних ресурсів у хмарному середовищі; а також методи опрацювання, статистичного аналізу та візуалізації даних для перевірки практичної ефективності запропонованих підходів.

Структура роботи. Кваліфікаційна робота складається із вступу, трьох розділів, висновків, списку використаних джерел, додатків. Обсяг основного тексту роботи складає 53 сторінки, 14 рисунків, 2 лістингів і 23 бібліографічних джерел. Загальний обсяг роботи – 61 сторінка.

Розділ 1. ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ ДОСЛІДЖЕННЯ

1.1. Відкриті дані: поняття та їх характеристики

В сучасних умовах розвитку суспільства, стану його інформатизації, цифровізації державного управління важливого значення набуває питання відкритості та доступності інформаційних ресурсів. Одним із важливих аспектів для оцінки діяльності органів державної та виконавчої влади, розвитку інноваційної економіки, ефективності управлінських процесів є можливість аналізу відкритих даних.

Концептуально поняття «відкриті дані» формувалися внаслідок розвитку інформаційно-комунікаційних технологій, їх активного використання у електронному урядуванні, а, головне, необхідності забезпечення вільного доступу до публічної інформації.

Нині відкриті дані розглядаються як невід'ємний аспект сучасного суспільного життя, який сприяє розвитку науково-технічної діяльності, економіки, громадянського суспільства та демократизації суспільних відносин. Зрештою, відкриті дані забезпечують можливість створення нових інформаційних (аналітичних) продуктів, сприяють стимулюванню створення прозорості державного управління та формуванню ефективних підходів до взаємодії між державою і громадянами.

У загальному випадку під поняттям «відкриті дані» розумують інформацію, яка може вільно використовуватися, повторно використовуватись і поширюватись будь-якою особою без певних обмежень, за винятком вимоги щодо зазначення джерела чи дотримання умов поширення.

Зважаючи на важливість тлумачення цього поняття законодавством України закріплено дану дефініцію нормативно. Так, згідно із Законом

України «Про доступ до публічної інформації» (ст. 10) [7], відкриті дані – це «публічна інформація у форматі, що забезпечує її автоматизовану обробку електронними засобами, вільний та безоплатний доступ до неї, а також можливість її подальшого використання».

У цій же статті зазначається: «Будь-яка особа може вільно копіювати, публікувати, поширювати, використовувати, у тому числі в комерційних цілях, у поєднанні з іншою інформацією або шляхом включення до складу власного продукту, публічну інформацію у формі відкритих даних з обов'язковим посиланням на джерело отримання такої інформації» [7].

На підставі цього визначення можемо означити наступні характеристики (ознаки) відкритих даних:

- вільний доступ до інформаційних ресурсів;
- можливість автоматизованого їх опрацювання;
- можливість повторного застосування;
- відсутність дискримінаційних обмежень;
- відкритий формат подання інформації.

Відкриті дані можуть містити різнотипну інформацію, наприклад, статистичні відомості, геопросторові дані, результати наукових досліджень, економічні показники, державні реєстри, інформацію про транспортні засоби й інші інформаційні ресурси.

Відкриті дані володіють специфічними характеристиками, що впливають на їх функціональність та різняться від інших видів інформаційних ресурсів. До таких характеристик належать:

- доступність – одна з ключових характеристик. Дані мають надаватися за допомогою інтернет-мережі без будь-яких дискримінаційних обумовлень та, як правило, безплатно або за мінімальні кошти. Дана характеристика надає можливість одержання, копіювання й використання інформаційних ресурсів без потреби отримання будь-якого дозволу;

- машиночитаність – представлення інформації у форматах, які придатні для процесів автоматизованого опрацювання (наприклад, .CSV, .XML, .XLSX). Це ключова характеристика, яка забезпечує можливість аналізування значних обсягів інформаційних даних, проєктування інформаційних систем та інших сервісів. Дана характеристика суттєво впливає на ефективність використання відкритих даних у створенні аналітичних продуктів;
- придатність для повторного використання – відкриті дані мають бути придатними для повторного використання й розповсюдження. Це забезпечує створення нових аналітичних продуктів (знань), розвитку інноваційних технологій та цифрового менеджменту;
- повнота та достовірність – дані мають оприлюднюватись у початковому вигляді без будь-яких спотворень та вміщувати максимально повну інформацію, чим досягається їх практична цінність. Достовірність інформаційних ресурсів забезпечує рівень довіри до них та ефективність їх опрацювання;
- актуальність – відкриті дані мають постійно оновлюватись згідно до змін, що відбуваються у певній галузі. Актуальність є необхідною вимогою ефективного використання інформації під час прийняття управлінських рішень;
- інтеперабельність – визначає здатність до сумісного використання різних інформаційних ресурсів й їх інтеграції з різними інформаційними системами. Це забезпечується завдяки використанню стандартизації форматів і метаданих;
- наявність метаданих – забезпечують структурування, форматування та умови використання інформаційних ресурсів. За їх допомогою досягається правильне інтерпретування інформації, підвищується ефективність використання даних

1.2. Аналіз і візуалізація відкритих даних хмарними сервісами

Хмарні технології (cloud computing) надають доступ до обчислювальних ресурсів, ПЗ і сховищ даних за допомогою інтернет-мережі без потреби інсталяції локальних програм. У галузі опрацювання відкритих даних вони забезпечують: збереження великих масивів інформаційних ресурсів; автоматизоване опрацювання інформації; проведення аналізу в режимі онлайн; створення інтерактивної візуалізації (дашбордів); колективну діяльність користувачів; інтеграцію різних джерел відкритої інформації. До основних переваг цих технологій слід віднести масштабованість, доступність, економічну ефективність, а, головне, можливість швидкого оновлення даних під час процесів аналізу.

Нині на ринку існує значна кількість хмарних програмних продуктів (сервісів), придатних для проведення аналізу та візуалізації відкритих даних. Враховуючи обмеженість роботи коротко розглянемо лише найпопулярніші рішення: Microsoft Power BI Service (<https://learn.microsoft.com/uk-ua/power-bi/>) [8]; Tableau Software Tableau Cloud (<https://data.softico.ua/>) [9]; Amazon Web Services QuickSight (<https://www.amazonaws.cn/en/quicksight/>) [10]; Google Looker Studio (<https://lookerstudio.google.com/>) [11].

Microsoft Power BI Service – це хмарна платформа бізнес-аналітики, яка дозволяє: під'єднання до відкритих наборів даних; постійне автоматизоване оновлення інформаційних ресурсів; підготовку та візуалізацію інтерактивних дашбордів; опрацювання значних обсягів інформаційних ресурсів; інтегрування із різними джерелами інформації.

Даний сервіс активно використовується органами державної та виконавчої влади, комерційними структурами з метою аналізування статистичних і соціально-економічних даних.

Tableau Software Tableau Cloud – сервіс, що надає потужні можливості інтерактивної візуалізації інформаційних ресурсів у хмарному середовищі. Даний інструмент забезпечує: аналізування великих наборів

відкритих інформаційних ресурсів; підготовку складних інтерактивних візуалізацій; здійснення прогнозування; формування аналітичних панелей; доступ до результатів аналізування з використанням вебінтерфейсу.

Amazon Web Services QuickSight – хмарний інструмент із системою бізнес-аналітики, що застосовує технології машинного навчання для аналізу інформаційних ресурсів. До основних функцій цього сервісу слід віднести: швидке опрацювання великих обсягів інформації; застосування інтерактивної аналітики; автоматизацію процесів виявлення закономірностей; створення візуалізацій; масштабованість процесів аналітики.

Google Looker Studio – хмарний інструмент, призначений для створення інтерактивних звітів і дашбордів на основі аналізу даних. До основних можливостей даного сервісу слід віднести: візуалізація відкритих інформаційних ресурсів у режимі онлайн; інтегрування з джерелами даних, що мають відкритий характер; підготовка інтерактивних звітів; можливість спільної роботи користувачів; автоматизація оновлення даних.

Зазвичай, даний інструмент активно застосовується під час аналізування вебресурсів, статистичних даних та інших соціально-економічних чи суспільно-політичних факторів.

Основною перевагою Google Looker Studio є безкоштовність, що зумовило його популярність під час проведення аналізу даних і вплинуло у подальшому на вибір під час нашого дослідження.

Порівнюючи характеристики розглянутих сервісів можемо знайти схожості і відмінності, однак кожному із інструментів властива можливість візуалізації аналізованих процесів. У кожному випадку це досягається застосуванням дашбордів.

1.3. Дашборд як засіб аналізу даних

Дашборд являє собою інформаційну панель. По суті – це інтерактивний засіб візуалізації, на якому відображаються головні показники

(результат аналізу інформації) у графічному вигляді. На ньому об'єднуються різні види інформації в одному вікні для швидкого аналізу даних.

Інформація, яка відображається на дашборді, легше сприймається, ніж, наприклад, текстова аналітична записка. При цьому, як правило, наявна інформація відповідно структурована, а усі параметри, що піддавалися аналізу, виділені різними кольорами й мають різну конфігурацію.

Дашборди забезпечують відображення ключових показників ефективності, відстеження змін даних у режимі онлайн, виявлення тенденції й закономірностей; можливість проведення порівняльного аналізу; підтримання прийняття управлінських рішень. Приклад дашборду, створеного за допомогою Google Looker Studio приведено на рис. 1.1.



Рис. 1.1. Приклад дашборду

У практиці аналізу відкритих даних розрізняють наступні види дашбордів [12]:

- операційні – відображають поточні процеси в режимі онлайн;

- аналітичні – застосовуються для поглибленого аналізування інформаційних ресурсів і встановлення закономірностей;
- стратегічні – забезпечують оцінювання перспективних показників розвитку.

До проектування дашбордів застосовують певні вимоги. Під час його створення слід подбати, щоб інформація на ньому відображалась у зручному для сприйняття вигляді. Тобто, під час першого погляду було зрозумілим, про що саме на ньому йдеться.

Під час створення діаграм і графіків, які будуються на основі різних даних, їх слід оформляти різними кольорами. Різні дані не можна відображати одним кольором, інакше здійснювати аналіз буде важко і в результаті через плутанину можуть бути сформульовані хибні висновки.

Дашборди мають також бути адаптивними за датою. Це дозволяє за допомогою зміни дат проаналізувати інформацію за різні часові періоди і встановити тенденції.

Не слід забувати про інтерактивність дашборту. Це дозволить проводити сортування за різними показниками (часом, регіонами, засобами тощо).

Під час створення дашборду слід забезпечити утримування функції відстежування за часом. Це питання є особливо актуальним під час контролю за динамікою певних процесів і, у такому випадку, необхідно забезпечити оновлення даних за визначені часові проміжки. Для прикладу, через кожну годину, день, тиждень тощо.

Не слід забувати на утриманні функціоналу порівняння, щоб можна було порівнювати обрані показники із визначеним періодом. А також забезпечити візуалізацію точок зростання/спадання, щоб зрозуміти максимальні і мінімальні показники проєкту

Не слід перевантажувати дашборт інформацією. Якщо на ньому більше ніж три графіки (діаграми), його буде важко сприймати й аналізувати.

Зрештою, слід пам'ятати про мету дашборду (як правило відображається) у його назві. Без неї немає сенсу в аналізі даних [13].

Дашборд займає важливе місце під час проведення аналізу тих чи інших інформаційних ресурсів. Він дозволяє швидко отримувати відповіді на визначенні питання. Наведемо приклад. Університетом запущено процедуру проведення профорієнтаційних заходів. Для цього задіяно ряд інформаційних каналів: email-розсилки, рекламу через успішних випускників, контекстну рекламу, використано ряд соціальних мереж. Через тиждень потрібно проаналізувати, який із задіяних каналів приніс найбільше трафіку на сайт університету, переглядів. Мета такого аналізу – залишити у використанні лише ті канали, які приносять найбільше переглядів (користі) та відмовитись від не робочих, особливо, якщо реклама потребує вкладання постійних коштів.

Використовуючи засоби аналітики Google доцільно створити дашборд і налаштувати передавання інформації з усіх задіяних до профорієнтації каналів з метою постійного відстежування результатів ефективності їх використання. При цьому може бути встановлено, що певний канал зовсім не приносить результатів або його використання є малоефективним. У такому випадку слід відмовитись від нього або прийняти заходи щодо його покращення. Якщо ж метою є омніканальний підхід, за якого усі канали працюють спільно без конкуренції, то ліпше обрати шлях удосконалення роботи кожного окремого каналу.

Також дашборд сприяє передбаченню та визначенню проблем. У нашому випадку, якщо дашборд вказує, наприклад, що перегляди рекламних профорієнтаційних роликів падає у порівнянні з цим же періодом минулого року, значить ролики застаріли або існують ще якісь певні причини, які, зрештою, призведуть до погіршення ситуації і в подальшому. Однак, у такому випадку менеджмент університету матиме можливість прийняти заходи, щоб уникнути такої ситуації.

Таким чином вимальовуються основні переваги використання дашбордів. До таких переваг слід віднести наступні:

- кожна мета, завдання підприємства, установи, комерційної чи державної структури відображається на єдиному екрані. Немає потреби відкривати різні документи, вкладки у браузері під час нарад, засідань чи інших зустрічей;
- усі необхідні інформаційні ресурси на дашборді відображаються в автоматичному режимі. Відсутня потреба у їх постійному ручному зведенні;
- дані легко порівнювати, наприклад, використовуючи часові періоди (місяці, квартали, роки тощо);
- результати порівняльного аналізу відображаються швидко та у зручному вигляді для прийняття управлінських рішень;
- простота й інтуїтивна зрозумілість під час створення та побудови графіків і діаграм (хоча потрібні деякі знання із галузі аналітичної діяльності);
- на дашборді не можна підмінити (сфальсифікувати) певні параметри, завищити чи знизити показники, тому візуалізація інформації є прозорою та максимально вірогідною;
- дашборд сприяє простоті плануванню маркетингових чи прийняттю інших управлінських рішень, виявляти напрями у діяльності, які «просідають»;
- дашборди дозволяють економити час на проведенні управлінських нарад, засідань чи інших зустрічей.

Звичайно, під час використання дашбордів можна знайти і певні недоліки, хоча їх не так багато і вони зовсім не перекривають отримувану користь від їх використання. До основних таких недоліків можемо віднести: потреба у базових знаннях з аналітичної діяльності; для побудови дашборду потрібно витратити певний час (спроєктувати його у кілька кліків мишки,

скоріш за все, не вийде); під кожен проєкт потрібно розробляти окремий дашборд.

Порівнюючи використання дашборду, наприклад, із звітом, підготовленого у форматі табличного процесора (рис. 1.2), можемо констатувати, що звіт – статичний інструмент, він не забезпечує можливість відслідковувати дані у динаміці, а це не забезпечує можливості проводити моніторинг різних показників у режимі реального часу [14] і відповідно сприяє прийняттю своєчасних управлінських рішень. Звіт можна підготувати разово і демонструвати у презентаціях, або можна готувати регулярно, додаючи до таблиці нову інформацію. Дашборд – динамічний інструмент, у якому відомості можуть оновлюватися як в автоматичному, так і у ручному режимі. Дашборд сприяє відстежуванню різних показників в онлайн-режимі.

The image shows a screenshot of Microsoft Excel with a PivotTable and its PivotTable Field List. The PivotTable is located in the range A1:C15 and has the following data:

Номер	Місто	Дата
1	Київ	Позначки рядків
2	Львів	Січ
3	Харків	Тра
4	Київ	Чер
5	Донецьк	Лип
6	Суми	Сер
7	Полтава	Вер
8	Броди	Жов
9	Стрий	Лис
10	Самбір	Загальний підсумок
11	Суми	
12	Львів	
13	Одеса	

The PivotTable Field List on the right shows the following settings:

- Field: **Дата** (checked)
- Filters: **Місяці**, **Дата**, **Місто**
- Rows: **Рядки**
- Columns: **Значення**
- Options: Відкласти оновлення макета, **Оновити**

Рис. 1.2. Приклад створення звіту (зведеної таблиці) [14]

Звичайно, дашборд слід вважати звітом, оскільки він сконцентровує відомості на одній сторінці та забезпечує візуалізацію зібраної інформації. Однак, не будь-яке зібрання інформації та її візуалізація це дашборд. Якщо у створеному звіті не передбачено можливості автоматизації оновлення інформації у онлайн-режимі, то такий звіт не вважатиметься дашбордом. На щоденному звіті можна підготувати зріз відомостей упродовж дня. Зазвичай, такий звіт вважається періодичним. За допомогою дашборду можна прослідкувати зміну показників на поточний час (кількість переглядів рекламних роликів, рівень прибутку на поточний момент чи інші показники).

Нині на ринку присутня значна кількість інструментів, використовуючи які можна швидко та ефективно спроектувати дашборд. Для цього лише потрібно зібрати необхідні інформаційні ресурси. Google Analytics забезпечує автоматичне їх збирання та об'єднання у єдине сховище даних.

1.4. Хмарні сховища даних

Останнє десятиліття характеризується стрімким розвитком хмарних технологій в усьому світі. Нині важко уявити діяльність державних чи комерційних структур, які б не використовували можливості таких технологій у своїй діяльності. Доказом цього може слугувати прийняття Верховною Радою України Закону «Про хмарні послуги» [15], який визначає правовідносини, що виникають під час надання хмарних послуг, та «... встановлює особливості використання хмарних послуг органами державної влади, органами місцевого самоврядування, військовими формуваннями, державними підприємствами, установами та організаціями, суб'єктами владних повноважень та іншими суб'єктами, яким делеговані такі повноваження».

Хмарні сервіси забезпечують доступ до різноманітних інформаційних ресурсів і програмного забезпечення лише із використанням будь-якого

браузера. При цьому, програмне забезпечення та інформаційні ресурси зберігаються у хмарі.

У даному підрозділі ми розглянемо деякі можливості доступних (безкоштовних) хмарних сховищ даних (баз даних), акцентуючи увагу на можливості їх застосування в аналізі даних, а також використання окремих сервісів під час підготовки дашбордів.

Хмарні послуги для служб сховищ даних нині є привабливою альтернативою через наявність таких атрибутів, як масштабованість і доступність. Однак, зберігання важливих інформаційних ресурсів у ненадійного постачальника хмарних послуг створює велику небезпеку, яка полягає у надійності захисту даних. Тому постачальники таких послуг мають гарантувати клієнтам конфіденційність даних та надійність зберігання відповідних інформаційних ресурсів.

З цією метою розробляються хмарні підходи до збереження та управління структурованими інформаційними ресурсами. Такі підходи є однією із найпопулярніших методик у галузі керування даними. Зокрема, концепція DBaaS концентрується на тому, що користувачу немає потреби встановлювати та здійснювати підтримку сховища даних. Йому лише потрібно відправити запит для отримання з її допомогою БД. Щоб досягти цього слід застосувати ресурси корпоративних, публічних або гібридних хмар [16].

Для прикладу можемо навести сервіс AmazonWebServices. Розробники цієї послуги пропонують наступні БД у хмарних середовищах: SimpleDB–NoSQL, яка представляє концепцію, у якій інформаційні ресурси зберігаються у «ключ-значення»; AmazonRelationalDatabaseService, що є уже SQL БД із інтерфейсом MySQL; DynamoDB.

Користувач одержує потрібні інформаційні ресурси із сервісу, які підібрані за визначеними у запиті параметрами. При цьому користувач має можливість здійснювати операції у такій БД, як із звичайною: модифікувати дані, розсилати визначеним користувачам, зберігати, обмінюватися із іншими

БД тощо. Послуга DBaaS надає функціонал подібний до того, що забезпечують реляційні СУБД, наприклад, такі як: Oracle Database, MySQL і SQLServer. З іншої погляду, хмарний сервіс DBaaS надає досить гнучкі, масштабовані можливості за вимогою, які дозволяють здійснювати самообслуговування й просте управління, особливо з огляду проектування свого середовища [17].

Пропозиції сервісу DBaaS надають достатньо можливостей для контролю продуктивності та використання, а також формують попередження для споживачів послуги про наявність можливих проблеми. DBaaS здатен здійснювати найпростіший аналіз даних.

DBaaS є прикладом хмарних сервісів: PaaS – платформа як послуга; IaaS – інфраструктура як послуга, SaaS – програмне забезпечення як послуга. Бази даних, які управляються з використанням описаної моделі, часто називають хмарними сховищами.

У моделі DBaaS (база даних як послуга) не є за можливе гарантувати дотримання конфіденційності за допомогою застосування нескладного шифрування даних користувачів, так як застосовувані методи шифрування перешкоджають виконанню SQL-запитів.

Є моделі, що забезпечують виконання SQL-запитів із зашифрованими відомостями. Такі архітектури ґрунтуються на використанні надійного проміжного проксі-сервера, що одержує доступ до сховища даних від імені своїх користувачів. Цей вибір моделі притаманний для вебкористувачів, яким надається доступ до системи управління базою даних за допомогою інших проміжних серверів. Однак, тут є один недолік: застосовуючи довірений проксі-сервер обмежується доступність й масштабованість шифрованої БД.

Таким чином, наявні архітектури, що базуються на використанні проксі-серверів, не відповідають змісту хмарного сховища даних, у якому є можливість розподіленням користувачам отримувати відомості з віддаленої СУБД.

У 2011 році компанія Google представила новий сервіс – GoogleCloudSQL, в основі якого закладено БД MySQL. У результаті вебсервіси отримали можливість опрацьовувати, створювати й використовувати реляційні БД, що по-факту знаходяться у хмарному середовищі компанії Google. Така послуга забезпечила загальне обслуговування БД для клієнтів сервісу, звільнивши їх від певних додаткових фінансових витрат, заощадивши час та відповідні технічні ресурси. Послуга викликала цікавість у напрямі застосування можливостей GoogleCloudSQL для проектування віртуальних БД, які можна застосовувати під час ознайомлення з базами даних у освітньому процесі. GoogleCloudSQL – це послуга, яка надає можливість створення, опрацювання та використання реляційних БД із програмами користувачів на базі ПЗ AppEngine, підготовлених на мові програмування Python.

GoogleCloudSQL – це повноцінний керований сервіс, який забезпечує підтримку БД і управляє нею, надаючи можливість проектувальникам сконцентруватися на реалізуванні ПЗ та необхідних послуг. Ключовими особливостями послуги є наступне:

- простота у використанні – наявний графічний інтерфейс забезпечує користувачу зручні процедури проектування, опрацювання, керування та безпосередній контроль за БД;
- простота в управлінні сервісом – не слід турбуватися про реплікацію БД, її конфігурацію чи про інші подібні аспекти;
- надійність у наданні послуги – забезпечується високим ступенем доступності ПЗ і служб. За умов розриву зв'язку із сервісом послуга буде і надалі доступною. Проектування нової віртуальної СУБД, створення резервної копії та можливість її подальшого відновлення не потребує значних зусиль.

Корпорацією Microsoft запропоновано публічне хмарне середовище – сервіс Microsoft Azure, який призначений для проектування та реалізації

хмарних сервісів. До базових компонентів сервісу Microsoft Azure належать компонент SQLAzure – реляційна БД, яка надається як сервіс. SQLAzure являє собою копію звичної системи управління базою даних MSSQLServer, яка забезпечує можливість роботи з БД з використанням вебсервісів.

Технологія забезпечує можливості зберігання неструктурованих і структурованих інформаційних ресурсів, виконання реляційних запитів та надає можливість проведення пошуку, об'єднання, синхронізування відомостей, а також підготовку аналітичних звітів. Використання даного сервісу забезпечує можливість створення проєкту реляційної БД з урахуванням усіх переваг, які забезпечуються будь-якою хмарною технологією.

Варто відзначити, що SQLAzure забезпечує достатній рівень безпеки через вбудовані системи захисту інформації, резервного копіювання й самовідновлення.

Користування MicrosoftSQLAzure будується на таких аспектах: обліковий запис, сервер і БД. Обліковий запис може керувати одним чи декількома серверами. Сервер являє собою логічну концепцію, схожу на основну БД, у якій міститься ще одна чи декілька БД з метаданими та відомостями про використання БД. Усі БД на сервері зберігають звичайні SQL-об'єкти: користувачів, таблиці, сутності тощо.

До основних переваг використання хмарних сховищ даних слід віднести:

- можливість створення резервної копії;
- різні способи зберігання інформаційних ресурсів;
- постійний доступ до інформаційних ресурсів, розміщених у хмарі;
- можливість колективної роботи з даними;
- значна економія фінансових, програмно-технічних ресурсів та часу.
- користувачі зберігають свої завдання в хмарі.

Із врахуванням описаних вище можливостей та характеристик хмарних сервісів MicrosoftSQLAzure і GoogleCloudSQL, зокрема,

доступності, простоти налагодження та управління хмарними інформаційними ресурсами, дані середовища широко використовуються під час інформаційно-аналітичної діяльності.

Разом з цим хочемо звернути увагу на платформу Google BigQuery. Google BigQuery є керованим хмарним сховищем даних (data warehouse). Даний додаток призначений для зберігання, опрацювання й аналізування значних обсягів як структурованих, так і напівструктурованих даних за допомогою SQL-запитів.

Важливою особливістю BigQuery є те, що ця платформа належить до класу аналітичних БД (так званих OLAP-систем), які оптимізовані для швидкого виконання складних запитів над великими масивами інформаційних ресурсів. Важливим аспектом цього вебдодатку є його серверлес-архітектура (serverless), що забезпечує відсутність необхідності у налаштуванні й адмініструванні серверної інфраструктури. Аналітики опрацьовують дані без необхідності керування ресурсами, масштабуванням чи оновленням. Усі ці процеси автоматичні.

Описані особливості зумовили широке застосування цього додатку у різних галузях: бізнес-аналітиці, фінансах, маркетингу, електронній комерції, логістиці та інших сферах, де обсяг інформаційних ресурсів постійно зростає.

Враховуючи це, вважаємо, що даний інструмент може бути використаний під час проведення нашого подальшого дослідження. Через те, детальніше платформа Google BigQuery розглядатиметься у розділі 2.

1.5. Перспективні напрями еволюції відкритих даних і технологій їх аналізу

Відкриті дані суттєво впливають на процеси реорганізації державного управління та на інші процеси практично в усіх галузях суспільної діяльності. Сучасні ініціативи у сфері Open Data спрямовуються на створення

національних та місцевих порталів, підвищення якості інформаційних ресурсів і практичних компетентностей користувачів, а також на гармонізацію стандартів опрацювання інформаційних ресурсів відповідно до міжнародних стандартів. У цьому напрямку Міністерством цифрової трансформації України презентовано нову візію розвитку відкритих даних, що включає модернізацію порталів, підвищення якості наборів даних і стандартизацію метаданих для забезпечення вищого рівня їхньої доступності та використання громадянськістю та бізнесом [18].

Відповідно органами місцевого самоврядування на місцевому рівні організовуються навчання для представників громад стосовно запровадження політики відкритих даних, що забезпечує прозорість й ефективність муніципального управління [19].

Процеси аналізу відкритих даних стають окремою галуззю практичної діяльності, що забезпечує створення платформних рішень для моніторингу і контролю як державних, так і комерційних процесів, наприклад, у галузі державних закупівель, бюджетних витрат, проведення банківських операцій тощо. Створення та запровадження нових аналітичних підходів, що ґрунтуються на відкритих даних, забезпечують суспільству розуміння діяльності державних та комерційних інституцій, роботу публічних політик, роблячи її більш підзвітною та ефективнішою.

Досягти цього дозволяє використання сучасних, високопродуктивних обчислювальних засобів для аналізу великих масивів інформаційних ресурсів, що дозволяє спрогнозувати тенденції у галузі соціально-економічного розвитку, наукових досліджень та політичних процесів.

На нинішній день можна виокремити три основні напрями розвитку відкритих даних і технологій їх аналізу.

1. Підвищення якості інформаційних ресурсів й їх інтеграції. Нині однією із головних проблем є досягнення високої якості інформаційних ресурсів та їх сумісності з різними джерелами. Сутність цієї проблематики

полягає у потребі узгодження метаданих, форматів представлення даних та впровадження єдиних стандартів обміну інформаційними ресурсами, що забезпечило б підвищення цінності даних як первинного джерела аналітичних відомостей. З метою вирішення названої проблеми проводиться розроблення автоматизованих підходів до ідентифікації «високоякісних» інформаційних ресурсів у базах даних, що забезпечить ефективніше спрямування аналітичних ресурсів на релевантні джерела для досліджень та прийняття управлінських рішень.

2. Розвиток аналітичних технологій. Нинішні тенденції у розвитку аналітики відкритих даних тісно пов'язуються з інтеграцією методик машинного навчання, штучного інтелекту та автоматизованого опрацювання природного мовлення, що забезпечує створення гнучкіших та функціонально насичченіших інформаційно-аналітичних систем аналізу інформаційних ресурсів. Майбутні наукові і практичні дослідження, наприклад у галузі Open Learning Analytics, можуть продемонструвати зростання кількості програмно-технічних та методологічних інструментів для аналізу відкритих даних у відповідних середовищах, що має потенціал бути адаптованим як для комерційних, так і для урядових даних. Це може відкрити нові можливості у прогнозуванні поведінкових тенденцій, оптимізації процесів та розроблення інноваційних рішень у різних галузях соціального життя.

3. Підтримання сталого розвитку й соціальних інновацій. Відкриті інформаційні ресурси також є важливою базою під час реалізування цілей сталого розвитку, економіки, фінансів та держави в цілому. За допомогою відкритих інформаційних ресурсів органи державного управління, місцевого самоврядування й громадяни отримують можливість ухвалювати обґрунтовані рішення стосовно національних, соціальних, воєнних, економічних, фінансових, екологічних та інших пріоритетів розвитку суспільних відносин. Доступ до таких інформаційних ресурсів забезпечує підвищення соціальної згуртованості й прозорості у прийнятті управлінських

рішень, що, своєю чергою, забезпечує підсилення довіри до інституцій влади й сприяє прогресивному розвитку територіальних громад.

Підсумовуючи слід зазначити, що еволюція відкритих даних та технологій їх аналізу є доволі складним, багатофакторним процесом, що включає розвиток як нормативно-правових засад, так і технічної інфраструктури для збирання, опрацювання й використання інформаційних ресурсів. Перспективи розвитку відкритих даних тісно пов'язуються із підвищенням якості, стандартів і доступності інформаційних ресурсів, впровадженням сучасних аналітичних технологій, а також зі сприянням сталому розвитку суспільства.

Названі вище тенденції відкривають нові можливості для інтеграції відкритих інформаційних ресурсів у процеси суспільного розвитку – від ефективного державного управління й прозорості до підтримки науково-практичних досліджень та інновацій. У даному контексті подальші дослідження і використання відкритих інформаційних ресурсів є не лише технологічним викликом, але й соціально важливим завданням XXI століття.

Розділ 2.

ОБРАННЯ ТЕХНОЛОГІЙ ТА ІНСТРУМЕНТАРІЮ ДЛЯ АНАЛІЗУ ВІДКРИТИХ ІНФОРМАЦІЙНИХ РЕСУРСІВ

Дослідивши у розділі 1 предметну область нашого дослідження слід визначитись із інструментарієм, який будемо у подальшому використовувати для аналізу відкритих інформаційних ресурсів. Враховуючи результати аналізу існуючих хмарних сервісів, що використовуються для дослідження відкритих даних (викладеного у підрозділі 1.2.), а також хмарних сховищ даних (викладеного у підрозділі 1.4.), оберемо для проведення подальших розробок інструменти: GoogleLookerStudio, GoogleBigQuery та одну із найпопулярніших на сьогодні мов програмування Python.

2.1. GoogleLookerStudio

LookerStudio – представляє собою безплатну хмарну платформу від компанії Google призначену для аналізування інформаційних ресурсів і візуалізації результатів опрацювання даних. Основний зміст її використання – підготовка інтерактивних аналітичних документів і дашбордів, за допомогою яких неопрацьовані відомості перетворюються на зрозумілі, структуровані та наочні дані, що є підставою для прийняття необхідних управлінських рішень.

У загальному випадку додаток орієнтований на маркетологів, фахівців з фінансів, логістики, аналітиків, менеджерів з різних галузей державного управління та господарської діяльності тощо. Володіючи інтуїтивним графічним інтерфейсом (рис. 2.1) користувачам надається можливість опрацьовувати дані без необхідності володіння глибокими знаннями у галузі інформаційних технологій чи програмування.

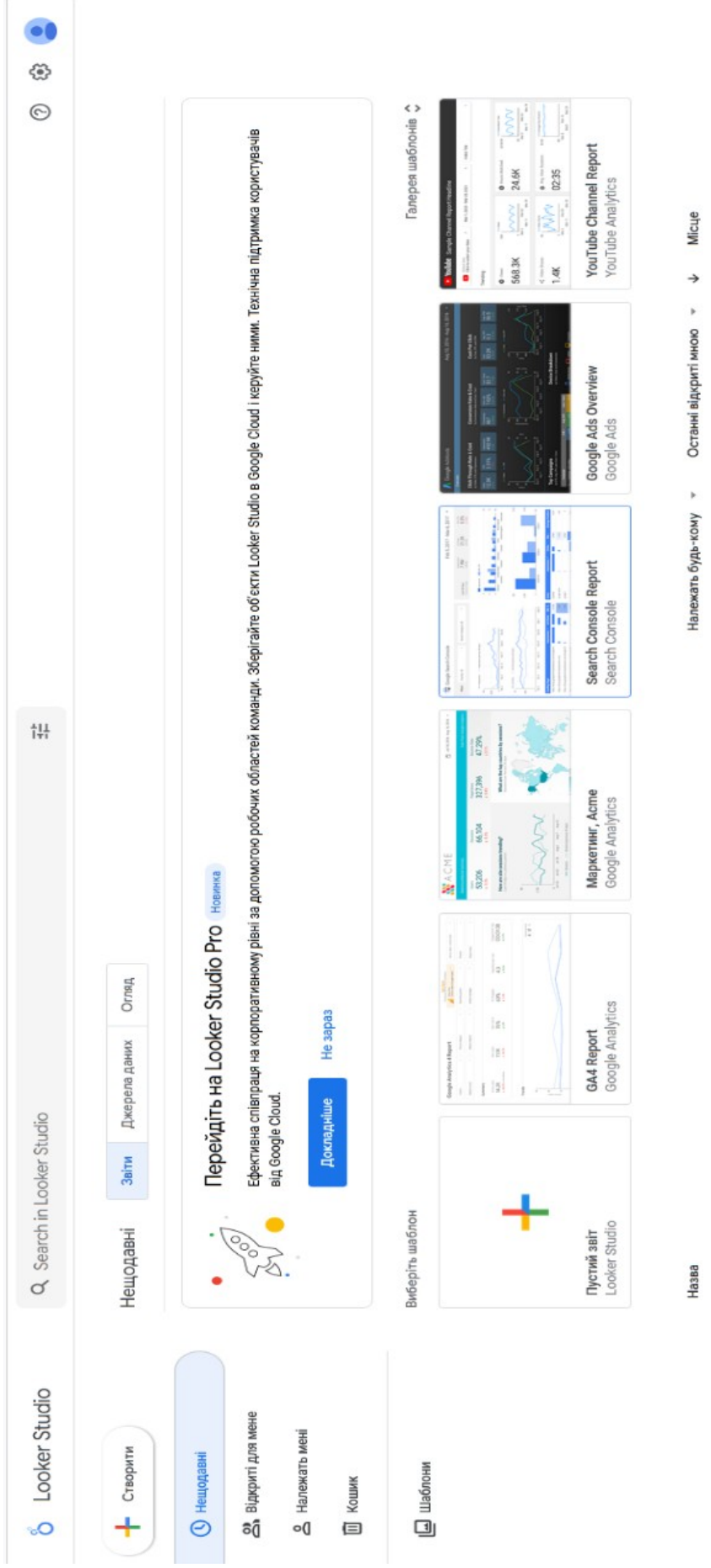


Рис. 2.1. Графічний інтерфейс Looker Studio

Перетворити неопрацьовані інформаційні ресурси на ефектні графіки (діаграми), застосувати відповідні ефекти та налаштування, швидко поділитися результатами аналітичної роботи з необхідними користувачами, надати до них спільний доступ – усе це дозволяє вивести прийняття управлінських рішень на основі даних на інший (вищий) щабель лише за допомогою кількох клацань мишки.

2.1.1. Характеристики LookerStudio

Характеризуючи розглядуваний додаток варто сконцентруватися на деяких можливостях цієї платформи. До основних таких характеристик LookerStudio слід віднести наступне:

1. *Здатність об'єднувати необхідні відомості в одному блоці.* Користувач може об'єднати інформаційні ресурси, одержані з кількох джерел, в єдиному звіті, що дозволяє здійснити комплексний аналіз бізнес-процесу. Інтегрування даних може здійснюватися з наступних джерел: Google Analytics, Google Ads, Google Sheets, BigQuery, YouTube Analytics, CSV-файли, SQL-бази даних.

Окрім цього можна просто здійснювати імпорт та візуалізацію наявної інформації з табличних процесорів, під'єднувати сторонні сервіси за допомогою конекторів (FacebookAds, CRM тощо) та інших інформаційних додатків чи БД, проводити аналіз актуальних звітів з декількох проєктів і джерел інформації в єдиному зручному інтерфейсі.

2. *Вивчення інформаційних ресурсів.* За допомогою LookerStudio просто здійснювати перетворення різних типів даних у прості для розуміння аналітичні матеріали (звіти) та інформаційні панелі, що відповідають вимогам бізнес-аналітики. За їх допомогою зручно відстежувати бізнес-процеси, які цікавлять користувача. Наприклад: фінансові витрати, трафік на сайті тощо.

3. *Ведення історії даних.* Здатність візуалізувати дані за різні часові періоди у вигляді естетично створених графіків і діаграм.

4. *Налаштування доступу та спільна робота.* Як і більшість продуктів Google, LookerStudio підтримує можливості спільного редагування аналітичних звітів, їх публікацію за посиланням чи розміщенням на вебсайті. Це робить додаток зручним для командної діяльності та презентації результатів менеджменту.

5. *Створення інтерактивних дашбордів.* Кожен елемент на панелі є інтерактивним. При цьому є можливість використовувати різне фільтрування даних, задавати періоди для аналізу, сегментувати відомості й переглядати детальну інформацію.

6. *Об'єднання та трансформація даних.* Застосування LookerStudio дозволяє: об'єднувати інформаційні ресурси, отримані з різних джерел; створювати розрахункові поля; застосовувати формули чи проводити математичні розрахунки, здійснювати групування, фільтрування даних та їх сегментацію. Усе це забезпечує можливості для проектування складних інформаційно-аналітичних моделей без застосування окремих ВІ-систем чи програмування.

7. *Автоматизація оновлення даних.* Дані у дашбордах оновлюються автоматично згідно з налаштуваннями джерела. Це забезпечує можливість отримання аналітичної інформації в онлайн-режимі, знижує витрати часу на підготовку звітності у ручному режимі, мінімізувати вплив людського фактору і помилок.

2.1.2. Переваги й обмеження використання LookerStudio

Безперечно, як уже вказувалось вище, однією з найважливіших переваг розглядуваної платформи те, що базова версія безкоштовна. Для малого та середнього бізнесу це забезпечує суттєву економію фінансових ресурсів у порівнянні із використанням платних ВІ-рішень.

Також, оскільки LookerStudio є хмарною технологією, то його робота забезпечується наявністю браузера. У цьому випадку не має необхідності встановлювати додаткове ПЗ – достатньо мати Google-акаунт та доступ до інтернет-мережі.

Окрім цього слід відзначити простоту у використанні сервісу. Це досягається завдяки інтерфейсу, спроектованому за принципом drag-and-drop (перетягування елементів). А це значить, що немає необхідності розробляти програмний код. Основні налаштування проводяться за допомогою візуального редактора. Нові користувачі здатні швидко освоїти додаток, що дозволяє навіть без досвіду аналітичної роботи створити професійний дашборд упродовж декількох годин.

До переваг сервісу також слід віднести гнучкість у налаштуванні дизайну. Зокрема, LookerStudio забезпечує налаштування кольорової палітри, використання корпоративного стилю, зміну структури сторінки, створення багатосторінкових звітів.

Слід також виокремити можливість підтримки масштабування діяльності. Із розширенням діяльності зростає й обсяг інформаційних ресурсів, які необхідно опрацьовувати. LookerStudio забезпечує підтримку великих обсягів даних за допомогою BigQuery, що своєю чергою забезпечує можливість створення складних аналітичних структур, дозволяє працювати з багатьма інформаційними ресурсами одночасно. Така перевага робить платформу LookerStudio привабливою не лише для невеликих організацій, а й для великих установ.

І, зрештою, як позитивний аспект слід відзначити можливість централізації звітності. Це означає, що замість великої кількості файлів, створених за допомогою табличних процесорів і розрізнених звітів, організація може впровадити єдину аналітичну систему. Це дозволяє стандартизувати підходи до звітності та підвищити прозорість аналітичних процесів.

Реалізацію окремих можливостей і переваг, перелічених вище, відображено на рис. 2.2.



Рис. 2.2. Можливості LookerStudio

Не дивлячись на розглянуті позитивні аспекти щодо використання сервісу LookerStudio, слід зазначити і про певні обмеження, які йому притаманні. Такими обмеженнями є:

- обмежена продуктивність під час роботи з надвеликими масивами інформаційних ресурсів без BigQuery;
- нижча глибина аналітичних можливостей у порівнянні з професійними BI-системами;
- залежність від стабільності роботи інтернет-мережі.

Проте, для більшості інформаційно-аналітичних потреб перелічених вище характеристик цілком достатньо.

Усе це вказує на те, що LookerStudio є потужним, гнучким і доступним інструментом для підготовки інтерактивної аналітики й звітності.

2.2. GoogleBigQuery

BigQuery – це хмарне середовище, яке є складовою GoogleCloud Platform і здатне опрацьовувати надвеликі обсяги даних (Big Data). Цей додаток є хмарним сховищем інформаційних ресурсів, що забезпечує можливість проведення інформаційно-аналітичних операцій із застосуванням стандарту SQL.

GoogleBigQuery є PaaS-сервісом («платформа як послуга»), що здатний підтримувати значку кількість функціоналу СУБД. Архітектурно це сервіс організовано на технології MPP, що забезпечує масово-паралельне опрацювання складних інформаційних запитів навіть з петабайтами інформації за секунду. Такий підхід забезпечує надвисоку продуктивність сервісу, надійність і простоту масштабування.

Взаємодія із додатком відбувається за допомогою вебінтерфейсу, API, CLI-засобів чи користувацьких бібліотек за допомогою використання різних мов програмування.

Йому притаманний різносторонній функціонал, використовуючи який клієнти мають можливість швидко вантажити надвеликі обсяги інформаційних ресурсів, формувати їх у формі електронних таблиць, створювати запити з використанням мови SQL, здійснювати зберігання та вивантаження їх результатів.

Детально особливості роботи з розглядуваним сховищем даних описано в [14].

2.1.1. Характеристики GoogleBigQuery

Як вже було названо GoogleBigQuery притаманний широкий різноманітний функціонал, який пропонуємо розглянути детальніше.

1. Збереження й опрацювання великих обсягів інформаційних ресурсів.

BigQuery забезпечує збереження фактично необмежену інформацію у табличній формі з можливістю її підготовки до аналітичних запитів. Сервіс

зданен підтримуються наступні типи даних: структуровані (табличні із фіксованою структурою); напівструктуровані (.JSON); потокові (streaming inserts); файлові (.CSV, .ORC тощо). Усі ці дані можна завантажувати як із хмарних середовищ, так і з локальних джерел з використанням інструментів інтеграції.

2. *Проведення SQL-аналітики й опрацювання складних запитів.* BigQuery зданен підтримувати стандартний SQL із додатковими розширеннями, які використовуються для опрацювання значних обсягів інформаційних ресурсів. Для цього призначені вкладені й аналітичні функції, передбачена агрегація даних, можливість опрацювання масивів і структур, застосування геопросторових і часових функції.

3. *Інтегрування із різними аналітичними та BI-додатками.* Як уже зазначалось, розглядуваний сервіс легко інтегрується із різними сервісами, що виконують функції візуалізації чи аналітики. Як правило, даний сервіс використовують разом із такими додатками, як LookerStudio, Tableau, MicrosoftPowerBI. Таке інтегрування забезпечує можливості для проєктування інтерактивних дашбордів на основі великих обсягів інформаційних ресурсів.

4. *BigQueryML (машинне навчання).* Хмарне сховище даних має закладені функціональні можливості, які забезпечують проєктування та навчання моделей машинного навчання на основі SQL-запитів. При цьому здійснюється підтримка лінійної й логістичної регресій, класифікація та кластеризація; моделі часових рядів; інтегрування з TensorFlow. Таким чином забезпечується можливість для фахівців-аналітиків проводити прогностичне моделювання за відсутності спеціалізованих знань з інформаційних технологій та програмування.

5. *Дотримання безпеки та контролю доступу.* З цією метою вебдодаток використовує шифрування інформаційних ресурсів як у статичному стані спокою, так і під час їх передавання, забезпечує перевірку

прав доступу на рівні проекту, формування даних, таблиць, а також інтегрування із системами управління автентифікацією.

6. *Масштабованість і продуктивність.* Хмарне середовище здатне масштабуватися в автоматичному режимі в залежності від навантажень. Клієнтам сервісу не має потреби управляти окремо певними кластерами або серверами. BigQuery здатний одночасно опрацьовувати десятки тисяч запитів, не втрачаючи при цьому продуктивність.

2.1.2. Переваги й обмеження використання GoogleBigQuery

Розглянутий нами вище функціонал дозволяє сформувати і окреслити основні позитивні аспекти використання розглядуваного хмарного сховища даних.

Насамперед слід акцентувати увагу на такій перевазі, як відсутність потреби у адмініструванні інфраструктури. Це дає змогу установі, організації, фірмі зосередитись на інформаційно-аналітичній діяльності, а не на потребах технічного супроводження інформаційної системи.

Наступною особливістю є те, що, завдяки можливості зберігання інформації у табличній (колонковій) формі та паралельному її опрацюванню, цей сервіс забезпечує практично миттєве створення складних аналітичних документів, використовуючи при цьому великий обсяг інформаційних ресурсів.

Важливим позитивним аспектом є також те, що вебдодаток BigQuery просто інтегрувати з різними додатками GoogleCloud, існуючими системами потокового опрацювання інформації й засобами для аналітики. Усе це забезпечує можливість роботи у створеному єдиному аналітичному середовищі, а здатність до потокового завантаження сприяє реалізації near real-time аналітики.

І, зрештою, описані вище засоби машинного навчання (BigQueryML) дозволяють максимально спростити процеси аналізу інформаційних ресурсів

із застосуванням алгоритмів машинного навчання без необхідності розроблення окремої інфраструктури.

Незважаючи на перелічені вище переваги сервісу GoogleBigQuery слід зазначити, що йому властиві і окремі недоліки (обмеження) у використанні. Серед них:

- безкоштовна версія обмежена, а вартість оплати за опрацьований значний обсяг інформаційних ресурсів може призвести до відчутних фінансових вкладень, особливо у випадку неефективного використання через часто повторювані запити без оптимізації;
- сервіс BigQuery призначений для роботи виключно у хмарному середовищі. Установи, організації, фірми та компанії через наявність, наприклад, регуляторних обмежень можуть відчувати проблеми під час використання ресурсу;
- даний вебдодаток оптимізований під аналітичні, а не транзакційні навантаження, тому його не можна використовувати як заміну операційним БД;
- ефективність використання ресурсу вимагає необхідного рівня знань щодо створення і оптимізації SQL-запитів. Без таких знань продуктивність цього додатку може бути неоптимальною;
- хоча розробники сервісу заявляють про підтримку streaming, в окремих випадках спостерігається певна затримка доступності інформаційних ресурсів під час складних запитів.

Незважаючи на перелічені вище недоліки, GoogleBigQuery на нинішній день є прогресивним високопродуктивним хмарним сховищем інформаційних ресурсів, спрямованим на опрацювання надвеликих масивів даних із використанням режиму аналітики.

Із врахуванням сказаного вважаємо, що даний вебсервіс є ефективним рішенням для установ, фірм, компаній, інших організацій, що опрацьовують великі обсяги інформаційних ресурсів, потребуючи при цьому отримання

швидких аналітичних результатів за мінімальних фінансових затрат на підтримання програмно-апаратної інфраструктури.

Таким чином, GoogleBigQuery слід сприймати як один із провідних вебсервісів хмарної аналітики, якому притаманна масштабованість, висока продуктивність, гнучкий функціонал у рамках сучасної інформаційно-аналітичної сфери діяльності.

2.3. Python

Мову програмування Python включено до складу нашого інструментарію з наступних міркувань. На сьогоднішній день ця мова є високорівневою інтерпретованою мовою програмування широкого призначення. Основне її орієнтування направлене на читабельність коду, ефективність і простоту його написання та, зрештою, універсальність під час практичного застосування.

Нині дана мова програмування знайшла широке застосування під час проектування вебдодатків, аналізі інформаційних ресурсів, штучному інтелекті, процесів автоматизації під час використання інформаційно-комунікаційних систем, а також під час наукових досліджень й підготовці ПЗ різного рівня складності.

Основні аспекти, які сприяють широкому застосуванню цієї мови програмування, полягають у тому, що у ній приховано деталізацію реалізування. Це забезпечує користувачам можливість зосереджуватися на логіці завдань. Також у ній виконання коду відбувається через інтерпретатор, без потреби завчасної компіляції, що сприяє спрощенню тестування й налагодження. Слід згадати і синтаксис, який є надзвичайно простим. Застосування відступів несе синтаксичне значення, що структурує код.

Такі підходи вказують на те, що ця мова програмування базується на об'єктно-орієнтованій парадигмі і включає як процедурне, так і функціональне програмування.

Найпопулярнішою платформою, яка дозволяє використовувати дану мову для програмування є Anaconda. Нижче розглянемо деякі особливості використання даного сервісу.

2.3.1. Anaconda. Загальна характеристика.

Дистрибутив Anaconda можна завантажити з офіційного сайту [21] (рис. 2.3). AnacondaPython – це безплатний додаток з відкритим кодом, який надає можливості створювати і виконувати програми мовою Python. Цей додаток набув широкої популярності під час вивчення мови програмування Python, а також для проведення різноманітних наукових досліджень, розрахунків, машинного навчання тощо. Однією із причин набутої популярності стало те, що у цьому додатку значно спрощено процедури розгортання пакетів й управління ними. Також важливим аспектом популярності стало те, що він постачається зі значною кількістю бібліотек, які просто використовувати під час підготовки власних проєктів.

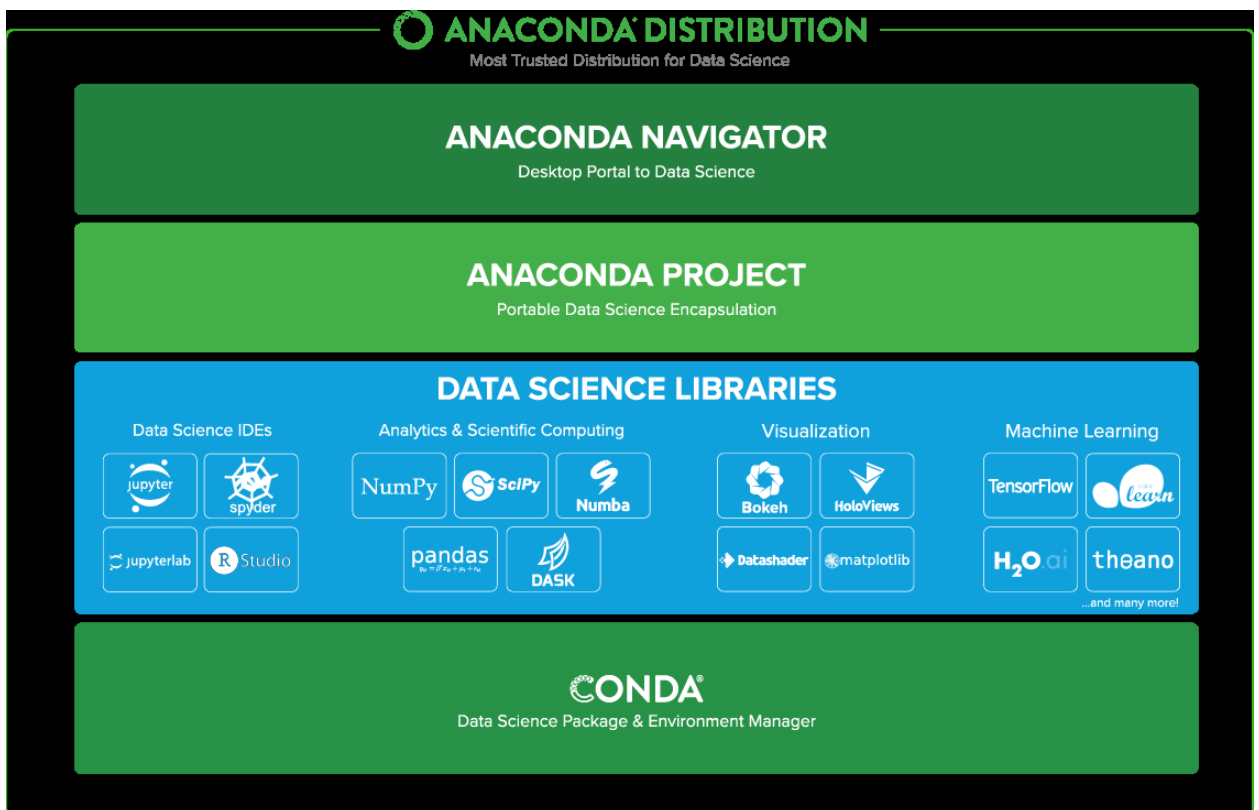


Рис. 2.3. Вікно дистрибутиву Anaconda

ПЗ Anaconda надає можливості підготовки різних середовищ для обраних версій Python і, відповідно, версій пакетів. Цю платформу можна також застосувати під час інсталяції, вилучення й оновлення додатків у середовищі розроблювального проєкту. Окрім цього, клієнт може застосувати Anaconda під час розгортання будь-якого потрібного проєкту за допомогою мишки. Такі можливості сприяють популяризації даного середовища і саме він є ідеальним під час вивчення мови програмування Python [22].

2.4.1 Anaconda. Інструментарій.

Під час інсталяції Anaconda клієнт одночасно встановлює ряд інших додатків (Conda, AnacondaNavigator, Miniconda тощо). У цьому підпункті опишемо окремі із цих додатків та означимо їх корисність.

Anaconda Navigator – по-факту це стільниця чи графічний інтерфейс клієнта, з якого зручно здійснювати запуск програм на виконання й ефективно управляти середовищем, каналами, додатками без наявності рядка команд. Графічний інтерфейс додатку Anaconda Navigator представлено на рис. 2.4.

Conda – додаток, який забезпечує керування пакетами у середовищі Anaconda. За його допомогою виконуються завдання інсталяції, оновлення й видалення додатків з інтерфейсу. Клієнт має можливість використати даний додаток для підготовки та проєктування власних пакетів підпрограм й забезпечити підтримку їх версій.

Miniconda – цей додаток є спрощеною версією Anaconda, який керує пакетами Python і Conda.

Spyder – являє собою середовище розроблення Python з різноманітним функціоналом для опрацювання коду Python, такими, наприклад, як текстовий процесор, відлагоджувач, інтерактивна консоль тощо.

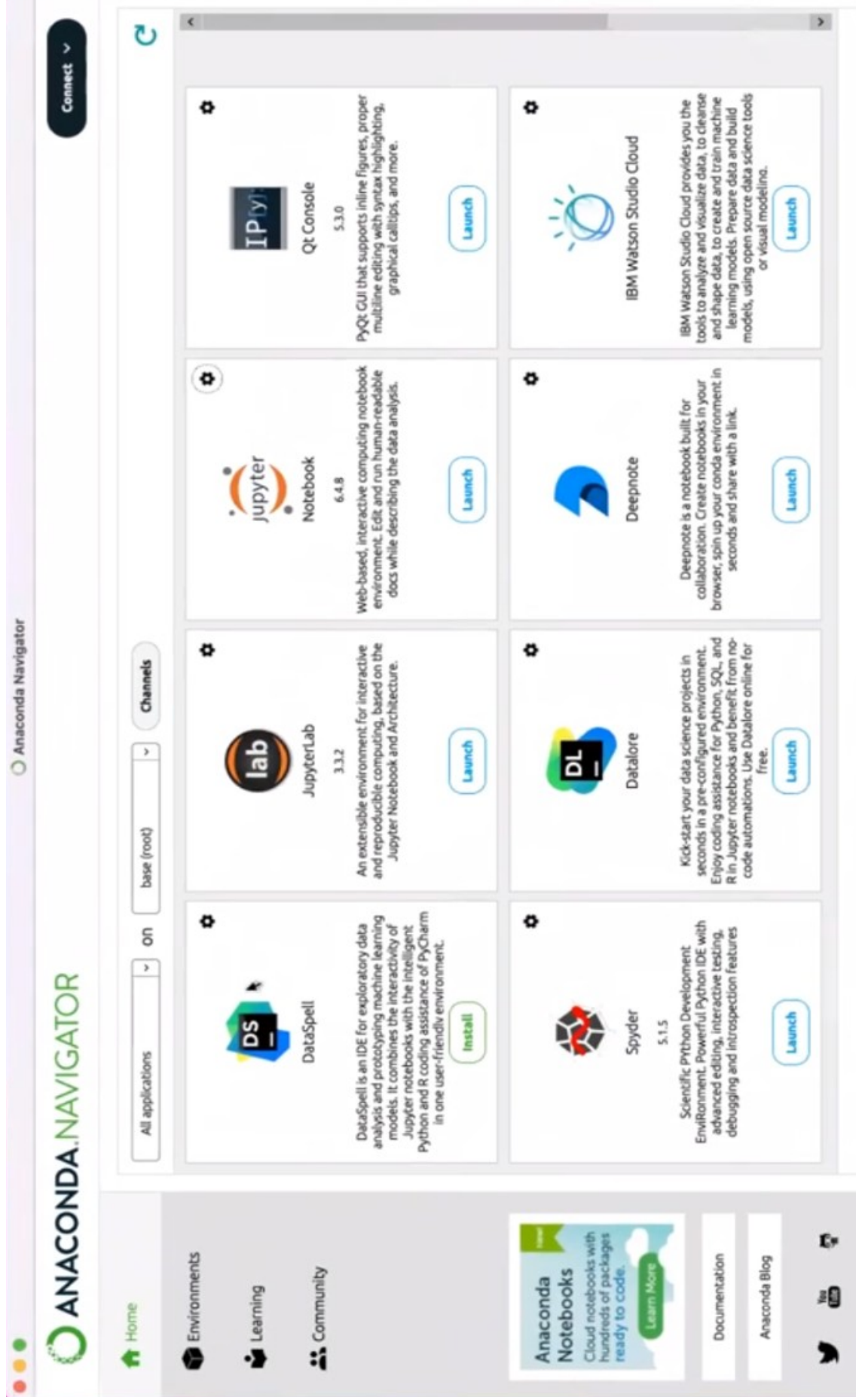


Рис. 2.4. Графічний інтерфейс Анаconda Navigator

Jupyter Notebook – вебдодаток, зміст якого полягає у можливості підготовки документів, що включають активний код, текст, візуалізацію з метою їх подальшого поширення.

Anaconda Prompt – інструмент взаємодії з командним рядком AnacondaDistribution.

Пакет – набір модулів, інстальованих з використанням додатку Conda. Тут під модулем розуміється файл, який створений мовою програмування Python, і має розширення .py.

Environment – папка, у якій розміщуються всі файли, що забезпечують роботу окремої програми. Для прикладу, інтерпретатор Python, пакети й файли конфігурування програм.

2.3. Визначення технології аналізу відкритих інформаційних ресурсів

Загальний вигляд спроектованої нами технології аналізу відкритих інформаційних ресурсів показано на рис 2.5.

Звичайно, що першопочатковим кроком будь-якого аналізу є наявність джерела (джерел) відкритих інформаційних ресурсів. Такими джерелами інформації можуть слугувати як певні посилання, що ведуть до вебсторінки із даними, так і власне сама БД. Тобто, мають бути деякі інформаційні ресурси, які підлягатимуть опрацюванню.

Відповідно до схеми (див. рис. 2.5.) застосовуємо додаток Anaconda, у який завчасно додаємо необхідні бібліотеки та засоби для опрацювання даних. Використовуємо додаток JupyterNotebook для підготовки створення потрібного програмного коду. Використовуючи JupyterNotebook і мову програмування Python опрацьовуємо відкриті інформаційні ресурси, перетворивши їх у необхідні типи та формати даних.

Опрацьовані інформаційні ресурси відправляємо через GoogleCloud Platform у БД BigQuery. Тут уже можна опрацьовувати інформаційні ресурси, які знаходяться у базі даних. Одночасно BigQueryWorkspace дає змогу

підготувати до вибірки дані за допомогою SQL-запиту, здійснити перетворення даних чи здійснити інші операції, передбачені додатком.



Рис. 2.5. Технологія аналізу відкритих інформаційних ресурсів

На наступному етапі переходимо до проєктування дашборду із гістограмами (діаграмами, графіками), які забезпечують краще розуміння

аналізованої ситуації. Як нами уже зазначалось, для цього завдання використаємо LookerStudio.

Завершальним етапом є аналіз відкритих інформаційних ресурсів аналітиком чи іншою особою, яка здатна їх проаналізувати та надати менеджменту структури (організації, фірми, підприємства тощо) рекомендації для прийняття відповідних управлінських рішень.

Розділ 3.

ПРАКТИЧНЕ ЗАСТОСУВАННЯ ТЕХНОЛОГІЇ АНАЛІЗУ ВІДКРИТИХ ІНФОРМАЦІЙНИХ РЕСУРСІВ ЗА ДОПОМОГОЮ ХМАРНИХ СЕРВІСІВ

3.1. Підготовка середовищ для аналізу інформаційних ресурсів

Перед проведенням аналізу інформаційних ресурсів хмарними сервісами необхідно здійснити їх попередню підготовку. Для цього необхідно, щоб на комп'ютері була встановлена Anaconda та мати облікові записи Google для подальшої роботи в GoogleBigQuery та GoogleLooker Studio (у даних додатках слід авторизуватися).

Інсталювати Anaconda можна з офіційного сайту (див. підрозділ 2.3.1). Після інсталяції необхідно запустити JupiterNotebook (див. рис. 2.4). У результаті відкриється вікно браузера, у якому можна починати проєкт. Далі потрібно обрати проєкт Python, що дозволяє розпочати процедуру підготовки коду.

JupiterNotebook є зручним, тому що дозволяє інсталювати необхідні бібліотеки відразу у вікні, не вдаючись до вказування команди на встановлення (оновлення) бібліотек в додатку.

Для продовження нашого дослідження можемо використати бази відкритих даних із порталу Єдиний державний вебпортал відкритих даних (data.gov.ua) [23], наприклад, дані про публічні закупівлі із системи Prozorro. Метою при цьому може бути аналіз ефективності державних закупівель, зокрема: визначення найбільших замовників і постачальників; аналіз економії бюджетних коштів; виявлення аномалій у тендерах; оцінка конкурентності закупівель тощо.

У процесі роботи дані з API або файлів Prozorro завантажуються до хмарного сховища та імпортуються в аналітичну базу даних Google BigQuery

для опрацювання і зберігання. У BigQuery можна здійснити агрегацію закупівель по регіонах, аналіз динаміки витрат за роками, пошук аномалій (наприклад, закупівлі з одним учасником), рейтинг постачальників.

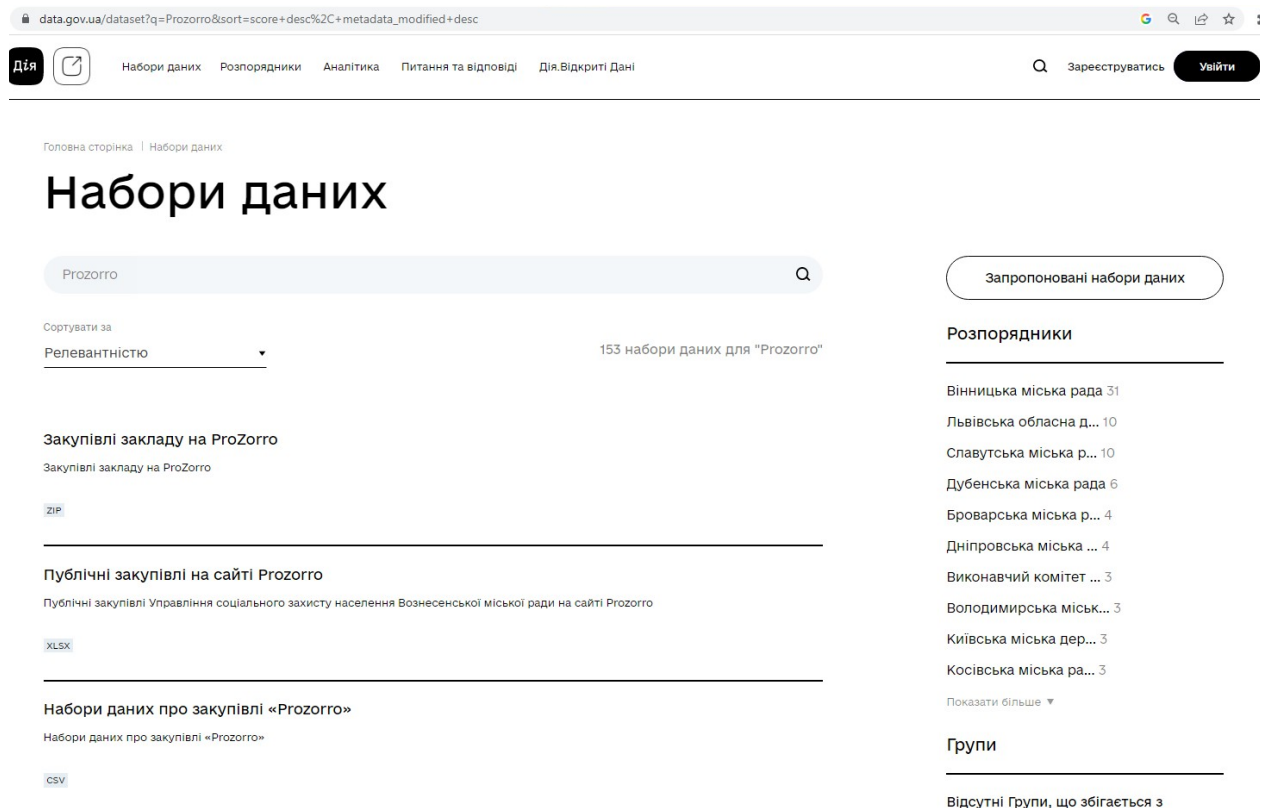


Рис. 3.1. Набори відкритих інформаційних ресурсів

3.2. Проведення процедури отримання відкритих інформаційних ресурсів для подальшого аналізу

Проведення процедури аналізу відкритих інформаційних ресурсів здійснюємо у наступній послідовності:

1. У Google BigQuery створюємо новий проєкт. У нашому прикладі він має назву «prozorro1dataset» (рис. 3.2.).
2. Встановлюємо бібліотеки (одноразово), за умови, що їх ще немає:

```
pip install requests google-cloud-bigquery python-dateutil
```

```
export GOOGLE_APPLICATION_CREDENTIALS="/path/to/service-account.json"
```

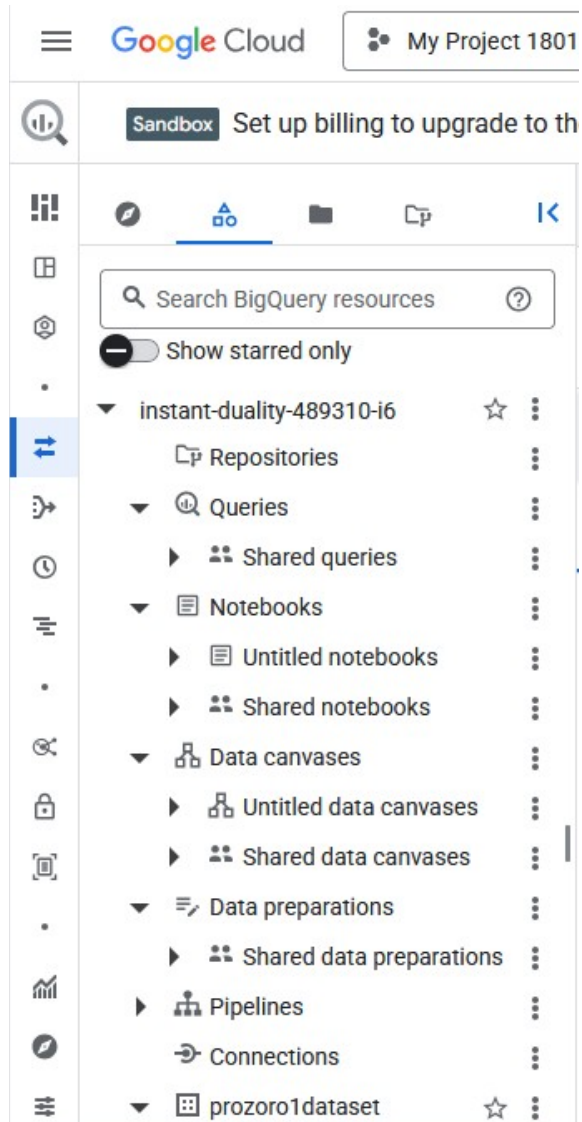


Рис. 3.2. Створення проєкту у BigQuery

3. Запускаємо у BigQuery (в консолі або через BigQuery) лістинг 3.1 для створення порожніх таблиць. Для цього готуєм SQL-запит:

Лістинг 3.1.

```
CREATE TABLE IF NOT EXISTS `PROJECT_ID.DATASET.prozorro_contracts_raw` (
  contract_id STRING,
  tender_id STRING,
  date_signed TIMESTAMP,
  amount NUMERIC,
  currency STRING,
  status STRING,
  contract_number STRING,
  date_modified TIMESTAMP,
  source STRING
)
PARTITION BY DATE(date_signed);
```

```

CREATE TABLE IF NOT EXISTS `PROJECT_ID.DATASET.etl_state` (
  pipeline STRING,
  last_offset STRING,
  updated_at TIMESTAMP
);

INSERT INTO `PROJECT_ID.DATASET.etl_state` (pipeline, last_offset, updated_at)
SELECT 'prozorro_contracts', NULL, CURRENT_TIMESTAMP()
WHERE NOT EXISTS (
  SELECT 1 FROM `PROJECT_ID.DATASET.etl_state` WHERE pipeline='prozorro_contracts'
);

```

У результаті у BigQuery створюються порожні таблиці (рис. 3.3).

The screenshot shows the Google Cloud BigQuery interface. The left sidebar displays a navigation menu with categories like Repositories, Queries, Notebooks, Data canvases, Data preparations, Pipelines, and Connections. The 'prozorro1dataset' is selected, and the 'etl_state' table is highlighted. The main panel shows the 'Schema' view for the 'etl_state' table. Below the schema, there are buttons for 'Edit schema' and 'View row access policies'.

Field name	Type	Mode	Description	Key
pipeline	STRING	NULLABLE	-	-
last_offset	STRING	NULLABLE	-	-
updated_at	TIMESTAMP	NULLABLE	-	-

Рис. 3.3. Створення таблиць у BigQuery

4. Переносимо відкриті інформаційні ресурси із баз даних системи державних закупівель «Prozorro» до таблиць створених у Google BigQuery за допомогою коду, підготовленого на основі мови програмування Python (лістинг 3.2).

Лістинг 3.2.

```

import time
import requests
from datetime import datetime, timezone
from dateutil import parser as dtparser

from google.cloud import bigquery

API_BASE = "https://public.api.openprocurement.org/api/2.5"
PIPELINE_NAME = "prozorro_contracts"

# -----
# Helpers: BigQuery state
# -----
def get_last_offset(client: bigquery.Client, state_table: str) -> str | None:
    q = f"""
    SELECT last_offset
    FROM `{state_table}`
    WHERE pipeline = @pipeline
    LIMIT 1
    """
    job = client.query(
        q,
        job_config=bigquery.QueryJobConfig(
            query_parameters=[bigquery.ScalarQueryParameter("pipeline", "STRING", PIPELINE_NAME)]
        ),
    )
    rows = list(job.result())
    return rows[0]["last_offset"] if rows else None

def set_last_offset(client: bigquery.Client, state_table: str, offset: str | None) -> None:
    q = f"""
    UPDATE `{state_table}`
    SET last_offset = @offset,
        updated_at = CURRENT_TIMESTAMP()
    WHERE pipeline = @pipeline
    """
    client.query(
        q,
        job_config=bigquery.QueryJobConfig(
            query_parameters=[
                bigquery.ScalarQueryParameter("offset", "STRING", offset),
                bigquery.ScalarQueryParameter("pipeline", "STRING", PIPELINE_NAME),
            ]
        ),
    ).result()

# -----
# Helpers: API calls
# -----
def api_get(url: str, params: dict | None = None, timeout: int = 30) -> dict:
    # Простий retry на тимчасові збої
    for attempt in range(5):
        r = requests.get(url, params=params, timeout=timeout)
        if r.status_code == 200:
            return r.json()
        if r.status_code in (429, 500, 502, 503, 504):
            time.sleep(2 * (attempt + 1))
            continue
        raise RuntimeError(f"API error {r.status_code}: {r.text[:500]}")
    raise RuntimeError(f"API retry exceeded for {url}")

def parse_ts(value: str | None):
    if not value:
        return None

    dt = dtparser.isoparse(value)

    if not dt.tzinfo:
        dt = dt.replace(tzinfo=timezone.utc)

    # BigQuery JSON ingest очікує ISO string
    return dt.astimezone(timezone.utc).isoformat()

```

```

# -----
# Extract + Transform: contracts pages
# -----
def fetch_contracts_page(offset: str | None, limit: int = 100) -> tuple[list[dict], str | None]:
    url = f"{API_BASE}/contracts"
    params = {"limit": limit}
    if offset:
        params["offset"] = offset

    payload = api_get(url, params=params)
    data = payload.get("data", []) or []
    next_offset = None
    next_page = payload.get("next_page")
    if isinstance(next_page, dict):
        next_offset = next_page.get("offset")
    return data, next_offset

def fetch_contract_detail(contract_id: str) -> dict:
    url = f"{API_BASE}/contracts/{contract_id}"
    payload = api_get(url)
    return payload.get("data") or {}

def to_bq_row(contract: dict) -> dict:
    # Мінімальна модель під "динаміку витрат за роками"
    value = contract.get("value") or {}
    return {
        "contract_id": contract.get("id"),
        "tender_id": contract.get("tender_id") or contract.get("tenderID") or contract.get("tender_id"),
        "date_signed": parse_ts(contract.get("dateSigned")),
        "amount": value.get("amount"),
        "currency": value.get("currency"),
        "status": contract.get("status"),
        "contract_number": contract.get("contractNumber"),
        "date_modified": parse_ts(contract.get("dateModified")),
        "source": "prozorro_api_contracts",
    }

# -----
# Load: BigQuery insert (JSON rows)
# -----
def load_rows(client: bigquery.Client, table_id: str, rows: list[dict]) -> None:
    # insert_rows_json добре підходить для невеликих/середніх батчів
    errors = client.insert_rows_json(table_id, rows)
    if errors:
        # Повертає список помилок по рядках
        raise RuntimeError(f"BigQuery insert errors: {errors[:3]}")

def run(project_id: str, dataset: str, limit_per_page: int = 100, max_pages: int = 50):
    client = bigquery.Client(project=project_id)

    raw_table = f"{project_id}.{dataset}.prozorro_contracts_raw"
    state_table = f"{project_id}.{dataset}.etl_state"

    offset = get_last_offset(client, state_table)
    print(f"[ETL] start offset = {offset}")

    pages_done = 0
    total_loaded = 0

    while pages_done < max_pages:
        page, next_offset = fetch_contracts_page(offset=offset, limit=limit_per_page)
        if not page:
            print("[ETL] no more data")
            break

        # Деталі контрактів (щоб отримати value/dateSigned стабільно)
        batch_rows: list[dict] = []
        for item in page:
            cid = item.get("id")
            if not cid:
                continue
            detail = fetch_contract_detail(cid)
            row = to_bq_row(detail)

            # Фільтр: пропускаємо рядки без суми або дати підписання
            if row["amount"] is None or row["date_signed"] is None:
                continue

```

```

        batch_rows.append(row)

    if batch_rows:
        load_rows(client, raw_table, batch_rows)
        total_loaded += len(batch_rows)
        print(f"[ETL] loaded batch rows = {len(batch_rows)} (total {total_loaded})")

    # Зберігаємо offset і рухаємося далі
    offset = next_offset
    set_last_offset(client, state_table, offset)

    pages_done += 1
    print(f"[ETL] next offset = {offset}, pages_done={pages_done}")

    if not offset:
        break

print(f"[ETL] done. total_loaded={total_loaded}")

if __name__ == "__main__":
    PROJECT_ID = "instant-duality-489310-i6" #os.environ.get("instant-duality-489310-i6")
    DATASET = "prozoro1dataset" # os.environ.get("prozoro1dataset")

    if not PROJECT_ID or not DATASET:
        raise SystemExit("Set env vars: BQ_PROJECT_ID and BQ_DATASET")

    run(PROJECT_ID, DATASET, limit_per_page=100, max_pages=50)

```

У результаті відкриті дані з «Prozoro» переносяться до створених таблиць у BigQuery (рис 3.4).

The screenshot shows the Google Cloud BigQuery interface. The top navigation bar includes the Google Cloud logo, the project name 'My Project 18015', and a search bar. The main content area displays the table 'prozorro_contracts_raw' within the 'prozoro1dataset' dataset. A notification indicates that this is a partitioned table. Below the notification, there are tabs for 'Schema', 'Details', 'Preview', 'Table Explorer', 'Insights', 'Lineage', and 'Data Profile'. The 'Preview' tab is active, showing a table with 5 columns: 'contract_id', 'tender_id', 'date_signed', and 'amount'. The table contains 12 rows of data, with the first row having a contract_id of 'a177544261f0423fb271c5fbc...' and an amount of 819.

Row	contract_id	tender_id	date_signed	amount
83	a177544261f0423fb271c5fbc...	40b04bf54ae14a209460154fde...	2016-08-29 07:50:02.770000 UTC	819
84	ed75404ea75d45e498e0dbc32...	7885b173356d4c9c90fd9d2fc1...	2016-08-29 07:40:31.204000 UTC	2924
85	0708f30962b44ba9b73bf7e958...	9511659c94984c32afc310d2f8...	2016-08-29 07:28:26.288000 UTC	4030.75
86	681abbb5dcd4a7c89cb7c33d...	014c2cdf7a93437b8d6a1e59c2...	2016-08-25 14:10:13.960000 UTC	190176
87	59c1aba3e87c4d59902861963...	4d1028f4967f439b9e35c9a668...	2016-08-26 13:07:50.689000 UTC	45710
88	7078d625a02547309f89ee07d...	041332b85e494e94915d477ff9...	2016-08-01 05:55:14.053000 UTC	47354
89	2c01ff35af394b6ebdff40c8589...	d12259f9773a4818a0d4a3ea7a...	2016-08-01 06:14:57.510000 UTC	66197.18
90	e6ba0e0ee3cae4ed58eacae42ec...	3aa0321700fd48349d0e0606c6...	2016-05-24 06:54:31.381000 UTC	981718.8
91	9ae13274d38d4c15aed26d737...	3607c62602444f929bd7413ed...	2016-08-30 09:35:00 UTC	493740
92	51b5398c1f4e41eba01885ab1c...	aacdef2137c1489e9ab8902f90...	2016-07-22 09:43:27.295000 UTC	11400
93	7a01a51e593a40d09be0cb7d6...	cd4ef15052d4492ab33061ef15...	2016-06-19 12:55:33.139000 UTC	50000
94	392f7844e77b4aea978db0d815...	4dbf31a32140441588c245577...	2016-08-31 13:43:07.503588 UTC	1499995

Рис. 3.4. Таблиця у BigQuery із перенесеними з «Prozoro» даними

3.3. Проектування дашборду для аналізу відкритих інформаційних ресурсів

Для створення дашборду заходимо у LookerStudio та розпочинаємо проектування звіту на основі даних з BigQuery.

На початковому етапі у вікні LookerStudio слід активувати іконку «Пустий звіт» (рис. 3.5).

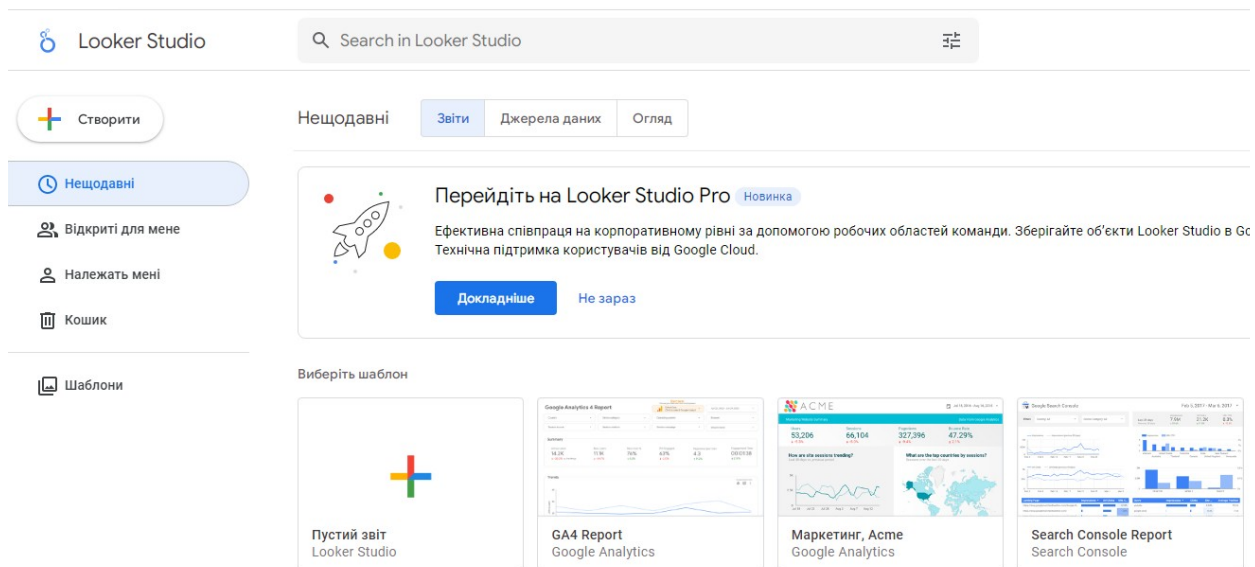


Рис. 3.5. Початкове вікно Looker Studio

Для проектування дашборду слід означити джерела інформаційних ресурсів, на основі яких його буде створено (у нашому проєкті це таблиці з BigQuery (рис. 3.6.)) та додати чарти (рис. 3.7).

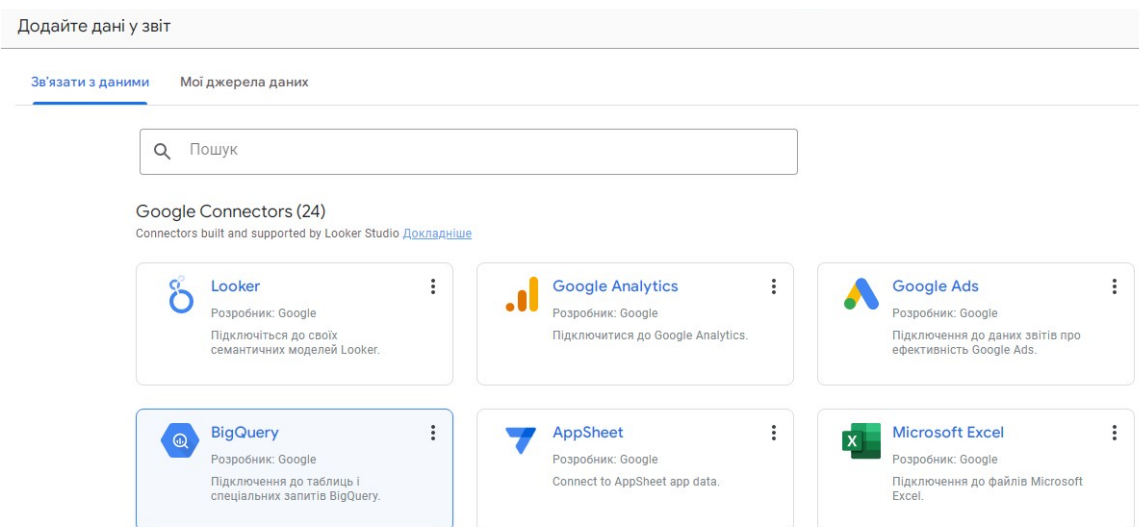


Рис. 3.6. Вибір інформаційних ресурсів для дашборду

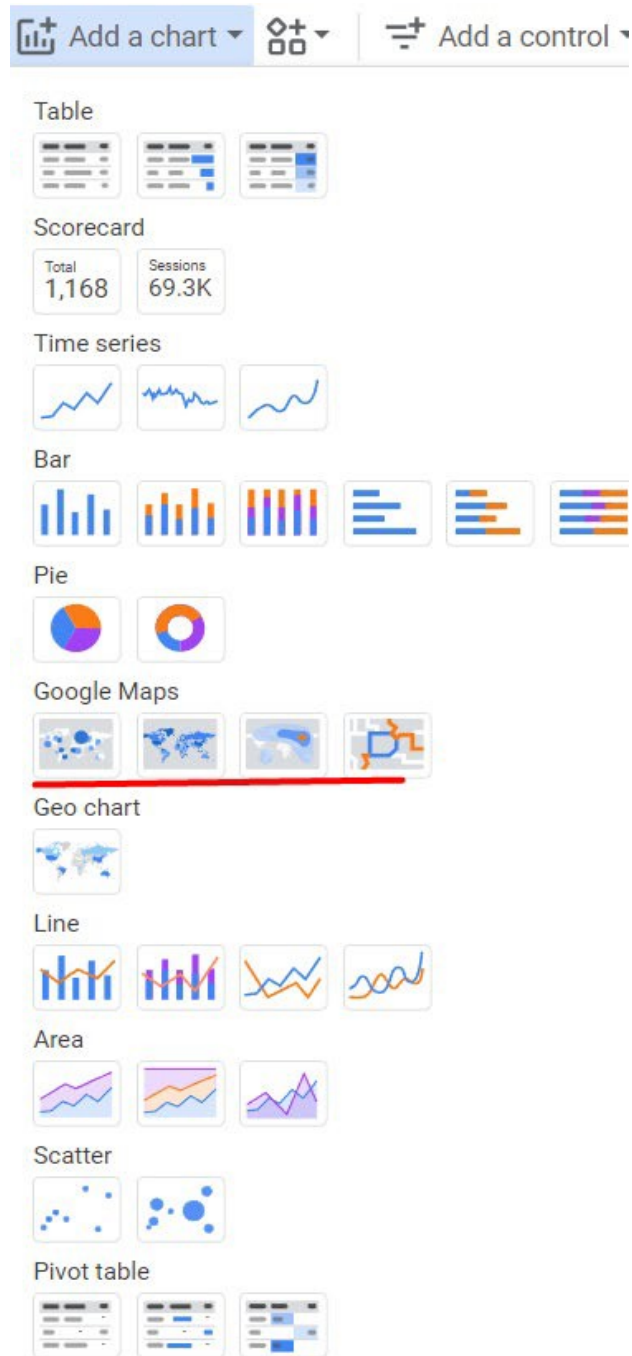


Рис. 3.7. Доступні чарти для проектування дашборду у Looker Studio

Після вибору джерел інформації та чартів з'явиться дашборд (рис. 3.8).

LookerStudio й BigQuery дуже швидко взаємодіють під час зміни чи оновлення інформації. Усі ці процеси відразу відображаються у дашборді, що дозволяє швидко аналізувати інформацію й приймати правильні управлінські рішення.

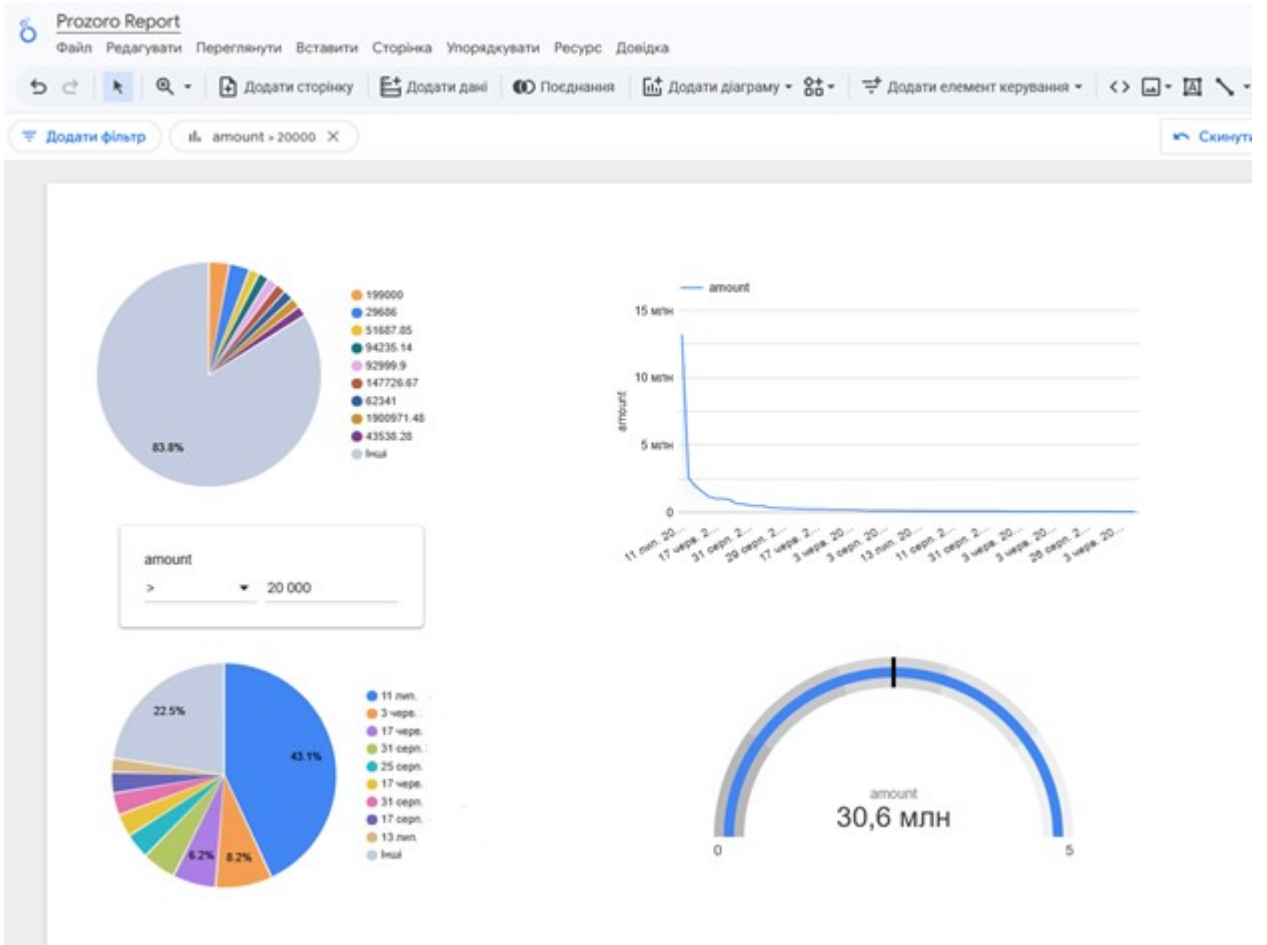


Рис. 3.7. Приклад спроектованого у LookerStudio дашборду для аналізу інформаційних ресурсів з «Prozoro»

ВИСНОВКИ

У кваліфікаційній роботі спроектовано та розроблено технологію аналізу відкритих інформаційних ресурсів із використанням хмарних технологій.

За результатами роботи реалізовано наступні завдання:

- проаналізовано сучасний стан розвитку відкритих інформаційних ресурсів та визначено їх основні характеристики;
- досліджено технологічні можливості хмарних обчислень у контексті опрацювання й аналізу інформаційних ресурсів;
- досліджено перспективні напрями еволюції відкритих інформаційних ресурсів і технологій їх аналізу;
- оцінено функціональні можливості хмарних платформ для зберігання, опрацювання, аналітики та візуалізації відкритих даних та обрано інструментарій для дослідження;
- розроблено технологічну модель організації процесу аналізу відкритих інформаційних ресурсів у хмарному середовищі;
- продемонстровано ефективність моделі організації процесу аналізу відкритих інформаційних ресурсів з використанням хмарних сервісів.

Для реалізації вказаних завдань обрано наступний інструментарій:

- Python – мова програмування (додаток Anaconda);
- Google BigQuery – хмарне сховище даних для зберігання інформаційних ресурсів з метою їх аналізу;
- Google Looker Studio – хмарний інструмент, призначений для створення інтерактивних звітів і дашбордів на основі аналізу даних.

За результатами дослідження систематизовано основні проблеми використання хмарних сервісів для аналізу відкритих інформаційних

ресурсів з урахуванням якості даних, масштабованості та безпеки; запропоновано узагальнений підхід до організації процесу аналізу відкритих інформаційних ресурсів у хмарному середовищі, що враховує специфіку різномірних джерел даних та вимоги до їх інтеграції, обґрунтовано доцільність використання хмарних технологій як базової інфраструктури для аналітичного опрацювання відкритих інформаційних ресурсів.

Практичне значення дослідження полягає в можливості використання отриманих результатів: у науково-дослідній роботі та освітньому процесі під час вивчення дисциплін, пов'язаних із аналізом даних, інформаційними системами та хмарними технологіями; для розроблення аналітичних інформаційних систем і дашбордів на основі відкритих інформаційних ресурсів; як рекомендацій для фахівців, які працюють з відкритими даними та хмарними сервісами.

В результаті виконання проєкту отримано нові знання і практичні навички з аналізу відкритих даних засобами хмарних технологій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Nuno Miguel Carvalho Galego, Domingos Santos Martinho, & Nelson Martins Duarte (2024). Cloud Computing for Big Data Analytics How Cloud Computing Can Handle Processing Large Amounts of Data And Improve Real-Time Data Analytics, *Procedia Computer Science*, 237, 297-304. <https://doi.org/10.1016/j.procs.2024.05.108>
2. Berisha, F., Mëziu, K., & Shala, B. (2022). Cloud computing: An overview and its impact in big data analytics. *Cluster Computing*, 25, 2821–2837. <https://doi.org/10.1007/s10586-022-03500-8>
3. Nosenko , Y. (2021). Еволюція засобів і технологій відкритої науки. *Науковий вісник Ужгородського університету. Серія: «Педагогіка. Соціальна робота»*, 1(48), 293–298. <https://doi.org/10.24144/2524-0609.2021.48.293-298>
4. Zhylin, A., Divitskyi, A., & Kozachok, A. (2019). Problems of protection of informational resources when using cloud technologies. *Collection "Information Technology and Security"*, 7(2), 171–182. <https://doi.org/10.20535/2411-1031.2019.7.2.190565>
5. Івкова, В. С., & Банах Р. І. (2023). Методика дослідження захищеності хмарних середовищ засобами OSINT. *Захист інформації*, 25(1), 34–42. <https://doi.org/10.31673/2409-7292.2025.026087>
6. Journal of Big Data. Взято з https://en.wikipedia.org/wiki/Journal_of_Big_Data?
7. Про доступ до публічної інформації: Закон України від 13 січня 2011 року № 2939-VI. Взято з <https://zakon.rada.gov.ua/laws/show/2939-17#Text>
8. Power BI documentation. Взято з <https://learn.microsoft.com/uk-ua/power-bi/>
9. Перетворюйте Ваші дані на миттєві результати разом із ВІ-системою Tableau. Взято з <https://data.softico.ua/>

10. Amazon QuickSight. Unified business intelligence at large-scale. Взято з <https://www.amazonaws.cn/en/quicksight/>
11. Google Looker Studio. Взято з <https://lookerstudio.google.com/u/0/navigation/reporting>
12. Що таке дашборд, як працює і як зробити самому. Взято з <https://goit.global/ua/articles/shcho-take-dashbord-iak-pratsiuie-i-iak-zrobyty-samomu/>
13. Що таке дашборд. *mc.today*. Взято з <https://mc.today/chto-takoe-dashbord-i-zachem-on-vam-nuzhen/#a>
14. Зачек, О. І., Сенік, В. В., Магеровська, Т. В., Пелех, Я. М., Рудий, Т. В., Д'яков, А. В., & Кулешник Я. Ф. (2022). *Інформаційні технології: навчальний посібник*. Львів: Львівський державний університет внутрішніх справ, 432 с.
15. Про хмарні послуги: Закон України від 17 лютого 2022 року № 2075-ІХ. Взято з <https://zakon.rada.gov.ua/laws/show/2075-20#Text>
16. Паралельні бази даних. Взято з <https://studfile.net/preview/7757346/page:75/>
17. Yarova, T. V., & Miakshylo, O. M. The Innovation Technology of Data Bases is SQL Azure Database. Взято з <https://dspace.nuft.edu.ua/server/api/core/bitstreams/853fd0c5-8829-4585-b4b3-d31e17bfb0f9/content>
18. Мінцифра презентувала візію розвитку відкритих даних в Україні. Взято з <https://thedigital.gov.ua/news/technologies/mintsifra-prezentovala-viziyu-rozvitku-vidkritikh-danikh-v-ukraini>
19. Прозорість Та Підзвітність: Розвиток відкритих даних у місцевому самоврядуванні. Тренінг Open Data Academy у Житомирі. Взято з https://oda.zht.gov.ua/news/prozoristtapidzvitnist-rozvytok-vidkrytyh-danyh-u-mistsevomu-samovryaduvanni-trening-open-data-academy-u-zhytomyri/?utm_source=chatgpt.com

- 20.Огляд BigQuery: інтерфейс сервісу, основні плюси та мінуси системи.
Взято з <https://profitstore.ua/uk/blog/expert/oglyad-bigquery-interfejs-servisu>
- 21.Download Anaconda Distribution. Взято з <https://www.anaconda.com/download>
- 22.Start coding immediately. Взято з <https://www.anaconda.com/products/distribution/start-coding-immediately>
- 23.Набори даних. Взято з https://data.gov.ua/dataset?q=Prozorro&sort=score+desc%2C+metadata_modified+desc