

Державний торговельно-економічний університет

Кафедра цифрової економіки та системного аналізу

ВИПУСКНА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

«Інтерактивна візуалізація діяльності компанії Apple»

Студента 4 курсу, 14 групи,
першого (бакалаврського)
рівня вищої освіти
спеціальності

124 «Системний аналіз»

освітньої програми

«Інформаційні технології та
бізнес-аналітика (Data Science)»

підпис студента

Антоновського Данііла
Олеговича

Науковий керівник
кандидат економічних наук,
старший викладач

підпис керівника

Лазоренко Віталій
Валерійович

Гарант освітньої програми
кандидат економічних наук,
доцент

підпис гаранта

Кулаженко Володимир
Валерійович

Київ 2023

Факультет інформаційних технологій

Кафедра цифрової економки та системного аналізу

Освітній ступінь бакалавр

Спеціальність 124 «Системний аналіз»

Освітня програма «Інформаційні технології та бізнес-аналітика (Data Science)»

Затверджую

Зав. кафедри _____ Роскладка А.А.
«15» грудня 2022 р.

**Завдання
на випускню кваліфікаційну роботу студента**

Антоновському Даніілу Олеговичу

(*прізвище, ім'я, по батькові*)

1. Тема випускної кваліфікаційної роботи (проекту)

«Інтерактивна візуалізація діяльності компанії Apple»

Затверджена наказом ДТЕУ від «09» грудня 2022 р. № 3333

2. Строк здачі студентом закінченої роботи 09 червня 2023 року

3. Цільова установка та вихідні дані до роботи

Мета роботи полягає у створенні інтерактивного аналітичного звіту діяльності компанії Apple.

Об'єкт дослідження є процеси діяльності компанії Apple.

Предметом дослідження слугує звітність компанії Apple.

4. Зміст випускної кваліфікаційної роботи (перелік питань за кожним розділом)

ВСТУП

РОЗДІЛ 1 ОСНОВНІ ПОЛОЖЕННЯ АНАЛІЗУ ДАНИХ

1.1. Основи Data Science

1.2. Класифікація методів Data Science

1.3. Огляд існуючих аналітичних платформ, які можуть бути використані для аналізу даних і побудови аналітичних звітів

Висновки до розділу 1

РОЗДІЛ 2 ПОБУДОВА АНАЛІТИЧНОГО ЗВІТУ ЗАСОБАМИ POWER BI

2.1. Створення аналітичного звіту на платформі Microsoft Power BI

2.2. Публікація аналітичного звіту

Висновки до розділу 2

РОЗДІЛ 3 АНАЛІЗ ДАНИХ В СЕРЕДОВИЩІ R STUDIO

3.1. Використання середовища R Studio для аналізу даних

3.2. Моделювання і аналіз цін на акції Apple

Висновки до 3 розділу

ВИСНОВКИ

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

5. Календарний план виконання роботи

№ пор.	Назва етапів випускної кваліфікаційної роботи	Строк виконання етапів роботи	
		за планом	фактично
1	2	3	4
1	Вибір теми випускної кваліфікаційної роботи	01.12.2022	01.12.2022
2	Розробка та затвердження завдання на випускну кваліфікаційну роботу	15.12.2022	15.12.2022
3	Вступ	01.02.2023	
4	Розділ 1. Основні положення аналізу даних	13.03.2023	
5	Розділ 2. Побудова аналітичного звіту засобами Power BI	24.04.2023	
6	Розділ 3. Аналіз даних в середовищі R Studio	01.05.2023	
7	Висновки та пропозиції	08.05.2023	
8	Здача випускної кваліфікаційної роботи на кафедру науковому керівнику	22.05.2023	
9	Попередній захист випускної кваліфікаційної роботи	30.05.2023	
10	Виправлення зауважень, зовнішнє рецензування випускної кваліфікаційної роботи	06.06.2023	
11	Представлення готової зшитої випускної кваліфікаційної роботи на кафедру	09.06.2023	
12	Публічний захист випускної кваліфікаційної роботи	За розкладом роботи ЕК	

6. Дата видачі завдання «15» грудня 2022 р.

7. Науковий керівник випускної кваліфікаційної роботи

Лазоренко В. В.
(прізвище, ініціали)

8. Гарант освітньої програми

Кулаженко В. В.
(прізвище, ініціали)

9. Завдання прийняв до виконання студент

Антоновський Д.О.
(прізвище, ініціали)

10. Відгук наукового керівника випускної кваліфікаційної роботи

Науковий керівник випускної кваліфікаційної роботи _____

(підпис)

(дата)

Відмітка про попередній захист _____

(підписи членів комісії попереднього захисту)

30.05.2023р

(дата)

11. Висновок про випускну кваліфікаційну роботу

Випускна кваліфікаційна робота студента АНТОНОВСЬКОГО Д.О.

(прізвище, ініціали)

може бути допущена до захисту екзаменаційній комісії.

Гарант освітньої програми _____

(підпис)

Кулаженко В. В.

(прізвище, ініціали)

Завідувач кафедри _____

(підпис)

Роскладка А. А.

(прізвище, ініціали)

« _____ » _____ 2023 р.

АНОТАЦІЯ

В даній випускній кваліфікаційній роботі було розглянуто основні задачі та етапи проведення аналізу даних, основні задачі консолідації даних, типи та структури сховищ даних, методи та алгоритми переробки даних, основи програмування мовою DAX, методи публікації поширення та захисту аналітичних звітів, характеристика аналітичної платформи Microsoft Power BI.

А також було проведено аналіз інтерактивної діяльності компанії Apple засобами Microsoft Power BI та RStudio.

Ключові слова: аналіз даних, передобробка, трансформація, візуалізація, ключові показники, аналітичний звіт.

Annotation

In this final qualifying work, I solved the main tasks and stages of data analysis, the main tasks of data consolidation, types and structures of data stores, and data processing algorithms, the basics of programming in the DAX language, methods of publishing and protecting analytical reports, characteristics of the Microsoft Power analytical platform. BI.

An analysis of Apple's interactive activities using Microsoft Power BI and RStudio was also conducted.

Keywords: data science, pre-processing, transformation, visualization, key indication, analytical report.

ЗМІСТ

ВСТУП	2
РОЗДІЛ 1 ОСНОВНІ ПОЛОЖЕННЯ АНАЛІЗУ ДАНИХ.....	3
1.1 Основи Data Science	3
1.2 Класифікація методів Data Science	9
1.3 Огляд існуючих аналітичних платформ, які можуть бути використані для аналізу даних і побудови аналітичних звітів	12
Висновки до розділу 1	14
РОЗДІЛ 2 ПОБУДОВА АНАЛІТИЧНОГО ЗВІТУ ЗАСОБАМИ POWER BI	15
2.1 Створення аналітичного звіту на платформі Microsoft Power BI.....	15
2.2 Публікація аналітичного звіту	25
Висновки до розділу 2	26
РОЗДІЛ 3 АНАЛІЗ ДАНИХ В СЕРЕДОВИЩІ R STUDIO.....	28
3.1 Використання середовища R Studio для аналізу даних.....	28
3.2 Моделювання і аналіз цін на акції Apple	38
Висновки до розділу 3	47
ВИСНОВКИ.....	48
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ	50
ДОДАТКИ.....	52

ВСТУП

Apple Inc – є американською технологічною компанією. Офіс розташовується у Купертино в Каліфорнії. Їхня продукція відома усьому світові, бо вона найбільша компанія в області інформаційних наук і технологій за виручкою, найбільший у нашому світі виробник мобільних гаджетів. Саме приставка “I” (“ай”) та емблема яблука стала невід’ємною частиною бренду.

Apple розробляє різні гаджети, а саме такі як:

- телефони (Iphone);
- планшети (Ipad);
- комп’ютери (IMac);
- ноутбуки (MacBook);
- смарт-годинники (Apple Watch);
- навушники (AirPods, Earpods);
- Apple TV та інше...

Мета випускної кваліфікаційної роботи полягає у створенні інтерактивного аналітичного звіту діяльності компанії.

Об’єктом даного дослідження є процеси діяльності компанії Apple.

Предметом слугує звітність компанії Apple.

Завдання випускної кваліфікаційної роботи було розглянути основи основи Data Science; дослідити класифікації методів Data Science; проаналізувати існуючі аналітичні платформи, які можуть бути використані для аналізу даних та побудови аналітичних звітів; провести аналіз на платформі MS Power BI; розробити код, що показує ключові показники компанії.

РОЗДІЛ 1 ОСНОВНІ ПОЛОЖЕННЯ АНАЛІЗУ ДАНИХ

1.1 Основи Data Science

Етапи аналізу даних можна виділити такі:

- отримання даних;
- обробка;
- аналіз;
- інтерпретація результату обробки.

Дослідження або отримання даних є первинним процесом пізнання об'єкту, предметної області. В процесі вивчення предметної області має бути створена модель. Складається вона з певних об'єктів, котрі саме мають різноманітний параметр властивостей й перебувають у певних відносинах чи взаємодії будь-яким чином.

Завдання *обробки* даних – це фаза розбору необробленої кількості даних, які мають бути в подальшому організовані для аналізу.

Аналіз даних включає в себе алгоритмічне виконання логічних дій з зібраних даних, подальше перетворення на статичну форму.

Методом *інтерпретації результатів* цих процедур є способи планування збірки даних, з метою простішого, точного аналізу в разі разом з механізмами статистики.

Загальна класифікація задач інтелектуального аналізу даних:

- класифікація,
- кластеризація,
- прогнозування,
- асоціація,
- візуалізація.

Завдання *класифікації* зводиться до визначення класу об'єкта за його ознаками. Є процесом сортування даних за різними типами, формами або будь-якими різними категоріями чи класами.

Завдання *кластеризації* належить до статистичної обробки та великого класу навчальних завдань («завдання без вчителя»). Тобто не є єдиним

алгоритмом, а загальною проблемою для вирішення котрої використовується різноманітні підходи.

Прогнозування – це процес передбачення майбутнього стану об’єкта або явища, котрий заснований на аналізі, зокрема і минулого, і сьогодення. Задача полягає у систематичній оцінці інформації про якісні та кількісні характерні ознаки подальшого розвитку обраного об’єкта чи явища.

Спосіб побудови *асоціативних правил* застосовані до пошуку комбінацій значень атрибутів та спрогнозувати другу множину («наслідок») на основі вже першої множини («умови»).

Задача візуалізації даних надає можливість компактно охопити та узагальнити великий обсяг інформації, нібито стискаючи її саму, результатом є графічне зображення проаналізованих даних. [1]

Консолідація даних – є одним із зручних варіацій, щоб об’єднати дані з різних джерел в один звіт. При процесі консолідації, дані упорядковуються таким чином, щоб їх було зручно узагальнювати і оновлювати.[2]

Основи задачі консолідації даних:

- вибірка конкретних джерел даних;
- розробка стратегії консолідації;
- збагачення та очищення;
- оцінювання якості даних;
- перенесення даних у сховище. [3]

Саме поняття сховище даних (або *data warehouse*) – це особлива форма організації бази даних, призначена для зберігання в узгодженій формі агрегованої інформації, отриманої на основі баз даних різних систем і зовнішніх джерел.

Характеризуються сховища даних за такими ознаками:

- *предметна орієнтованість* – це коли в репозиторій сховища має в собі дані, що всебічно описують конкретну предметну область;
- *інтегрованість* – збір даних з різноманітних джерел, їх узагальнення та зберігання в єдиному загальному репозиторії;
- *забезпечення несуперечності даних* – це дані з джерел, що можуть

містити продубльовану та суперечливу інформацію, тому ми виконуємо такі кроки, як: перевірка, звіряння, узагальнення та доповнення перед завантаженням в репозиторій сховища;

- *незмінюваність* – дані репозиторію використовуються тільки в режимі читання та ніяк не коригуються;
- *підтримка хронології* – збереження даних за датою і максимального часового періоду, котрий можливий;
- *оптимізація під виконання складних аналітичних запитів* – проектування самого сховища за для мінімізації часу щодо формулювання аналітичного звіту, який необхідний для підтримки рішень.

Корпоративні сховища даних містять інформацію про всю компанію або підприємство, зібрану з оперативних джерел для консолідованого аналізу. Як правило, такі репозиторії сховища охоплюють багато аспектів діяльності компанії та використовуються для прийняття тактичних й стратегічних рішень. Архів підприємства містить детально зведену інформацію. За вартістю підтримка і створення корпоративного сховища є високою. Більшість із них створюються централізованими відділами інформаційних технологій, і створюються вони за нисхідним методом – це спочатку проектується загальна схема, а вже потім починається наповнення даних. Даний процес триває декілька років.

Основною складовою структури сховищ даних є *таблиця фактів* та *таблиця вимірів*.

Таблиця фактів – це основна таблиця у сховищі даних. Зазвичай містить в собі дані про об'єкти чи події, сукупність котрих аналізуватиметься надалі. В більшості випадків зустрічаються чотири види випадків такі як:

- *транзакційні факти* (базуються на певних подіях);
- *факти «миттєвих знімків»* (основані станом об'єкту в певний період часу (кінець дня/місяця));
- *факти зв'язані з елементами документа* (складається на основі того чи іншого документа і містить детальну інформацію про елементи цього документа);

- факти події чи стану об'єкта (ті що уявляють подію без її подробиць).

Таблиці вимірів містять інформацію, які є постійними або рідко змінюються. У більшості випадків ці дані одиничним записом для всіх елементів нижнього рівня ієрархії у вимірі.

Існують три типи сховищ даних:

1. *Корпоративне сховище* – база даних, котра поєднує різноманітний функціонал в області організації та об'єднує їх у єдиний порядок.
2. *Оперативне сховище* – використовується в якості альтернативного застосування програми для підтримання прийняття операційного рішення.
3. *Data Mart* (або *вітрине сховище*) – тримає увагу на зберіганні інформації для конкретної функціональної області, в якій міститься підмножина даних, що зберігаються у репозиторії сховища.[4]

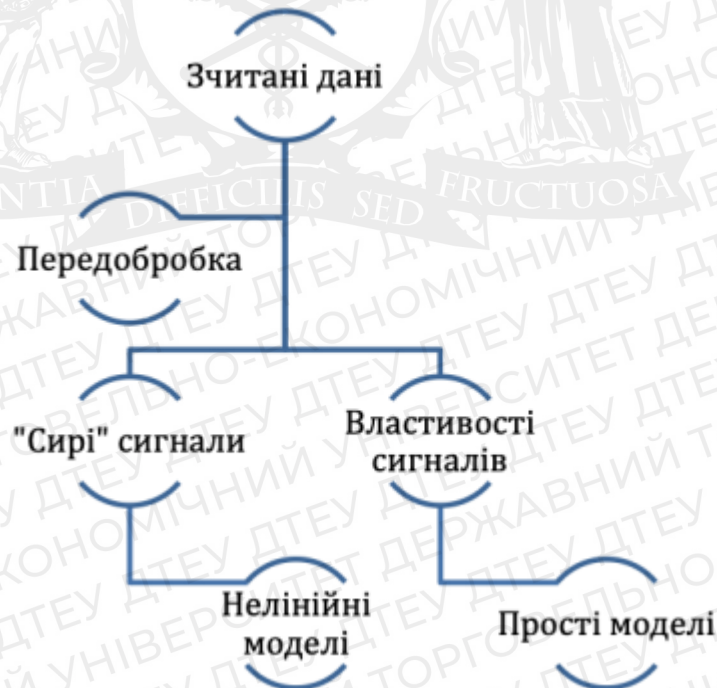


Рисунок 1. 1 Алгоритм передобробки даних

Задля підвищення якості даних йде ужиток комплекс методів і алгоритмів під назвою «очищення даних». Тобто перед використанням методу очищення треба зробити оцінку якостей даних. Постає питання вибору де проводити дану оцінку якості. Ці місця можуть бути:

- в джерелах даних (вишукування помилок орфографії, пропущень, дублювання, протирічних значень на рівні запису і таблиці);
- процес Extract transform load (визначення проблеми даних та відповідно використовувати метод очищення, щоб переконатися, що вже очищені дані завантажуються в пам'ять);
- аналітична система (виконується перед використанням різних методів інтелектуального аналізу інформації в процесі передобробки самих даних).

Найпоширеніший метод перевірки якості даних є *профайлінг*. Він виконується, основуючись на аналізі meta-даних, які описуються саме структурою даних. В цьому процесі відбувається наступний аналіз даних:

- тип поля (атрибуту) – викривлення типу можуть бути викликані неправильним використанням символів-розподільників для цілих і дробових чисел, коли замість крапок використовуються коми та й навпаки, а також неузгодженими роздільниками для груп розряду у числах пробілу, коми, крапки або ж їх відсутність;
- довжина значень – визначає кількість максимальної дозволених символів в значенні атрибута. Якщо йде перевищення встановленого числа, тоді потрібно обмежити довжину значення.
- дискретні значення – перевірка унікальності і частоти появи;
- діапазон допустимих значень – мінімальне/максимальне значення, котрий приймає атрибут;
- аналіз рядкових шаблонів – рядки обробляються згідно з деякими задалегідь визначеними правилами.

Призначення профайлінгу це виявлення найпоширеніших «регулярних» технічних проблем.[5]

Створення і використання моделей інтелектуального аналізу (Data Mining) є одним із ключових пунктів на початку розуміння, осмислювання і подальшого прогнозу тенденцій в проаналізованих об'єктах.

Побудова моделі даних здійснюється задля дослідження чи вивчення об'єктів які моделюються, процес чи явище отримання набутих знань, котрі є необхідними для прийняття рішень.

Використання даних моделей надає можливість у визначенні найліпшого рішення для певної ситуації.

Аналітики створюють моделі, як подібність досліджувальних об'єктів. Описані моделі можуть бути показані у різних зображеннях, формулах та схемах. [6]

1.2 Класифікація методів Data Science

Візуалізація даних – це графічне представлення інформації, яке дозволяє стисло представити на зображенні те, що потребує кількох абзаців у еквівалентному тексті.

Основна ідея візуального аналізу даних полягає в представленні даних у певній візуальній формі, яка дозволяє використовувати візуальні представлення даних, розуміти їх природу, робити висновки та безпосередньо взаємодіяти з даними. Візуальний аналіз проводиться в три етапи:

- 1) *Побіжний аналіз* - дозволяє виявити корисні закономірності та зосередитися на одному або кількох з них;
- 2) *Посилення та фільтрація* - шаблони, виявлені на попередньому етапі, фільтруються та розглядаються в більшому масштабі;
- 3) *Деталізування за потреби* - якщо потрібна додаткова інформація, виконайте детальну візуалізацію даних.

Одні з найпоширеніших способів візуалізації даних:

- графік;
- діаграма;
- інфографіка;
- схема;
- бізнес-аналітика;

- карта або картограма.

Існують критерії оцінювання, за якими можна відібрати оптимальні методи візуалізації:

- 1) *простір відображення* (залежить від визначеного методу зображення буває двовимірним/тривимірним);
- 2) *поведінка в часі* (визначає «спосіб поведінки» відтвореного малюнку);
- 3) *обхід перепон* (зазвичай визначає чи метод візуалізації в час створення елементів зображення враховує можливість їхнього накладання один на одного та усунення накладання);
- 4) *спосіб формування зображення* (визначає тип створеного зображення);
- 5) *сітковість малювання* (вказує на лінії, які можуть бути намальовані лише під певним кутом).[7]

DAX (Data Analysis eXpressions) – це мова формул, або ж формули для зведених таблиць.

Більшість функцій DAX мають імена та характеристики, подібні до звичайних функцій Excel, але доступні лише в PowerPivot. Як відомо, звичайна зведена таблиця може містити обчислювані поля, над якими можна виконувати прості арифметичні операції (додавання, віднімання, множення, ділення). DAX було створено, щоб додати функціональність обчислюваного поля. Загалом DAX не є складною мовою. На перший погляд, достатньо зрозуміти, як працює DAX. Різниця між формулами Excel і формулами DAX полягає в тому, що Excel працює з окремими клітинками та діапазонами, тоді як DAX дозволяє отримати доступ лише до цілих таблиць та їхніх стовпців.

Які бувають основні функції DAX:

- функція фільтру (фільтрування);
- математична функція;
- статистична функція;
- логічна функція;
- функції дати і часу;
- текстова функція;
- інформаційна функція.

Існують два типи створення розрахункових полів:

- розрахунковий стовпець;
- заходи.

Розрахунковий стовпець – це такий додатковий стовпець, створений користувачем в якій-небудь таблиці, що вже завантажена в PowerPivot.

Міра – це розрахункове поле в «Зведеній таблиці», яка використовується в області значення зведеної таблиці. Задля розміщення результату в іншій області зведеної таблиці використовується розрахунковий стовпець.

В DAX існують три типи контексту:

- *контекст фільтра* (набір дозволених значень для кожного стовпця, залежно від обмеження фільтра або виразу, які застосовуються до рядка або певних критеріїв фільтра у формулі);
- *контекст рядку* (відповідає концепції поточного рядка. До прикладу, якщо був стверний обчислюваний стовпець, тоді контекст рядка – це значенням в кожному окремому рядку та значення в стовпці, пов'язаному з поточним рядком);
- *контекст запиту* (відноситься до підмножини даних, яке неявно створюється для кожного осередку комірки зведеної таблиці на основі заголовків рядків і стовпців зведених таблиці).[8]

Одна з функцій Power BI є *публікація звіту*, котра доступна для всіх зареєстрованих користувачів програми. Щоб опублікувати потрібно на вкладці «Основне» натиснути «Опублікувати». За потреби залогінитися у власному акаунті Microsoft. Далі вибрати розташування звіту, натиснувши кнопку *Вибрати*. До рекомендацій опублікування щоб спростити доступ до SharePoint звіту, опублікувати треба до робочої області групи через «*Керування проектами*». Після служба Power BI самостійно дозавантажить ваш звіт до браузеру.

Якщо у вас є звіт в Power BI і ви хочете надати доступ комусь з ваших колег чи друзів, одним із способів поширення – це функція *надати спільний доступ*. Нажаль надання спільного доступу потрібна Pro версія Power BI. Якщо ця версія у вас є, тоді ви можете ділитися звітом за допомогою посилань, що надають доступ:

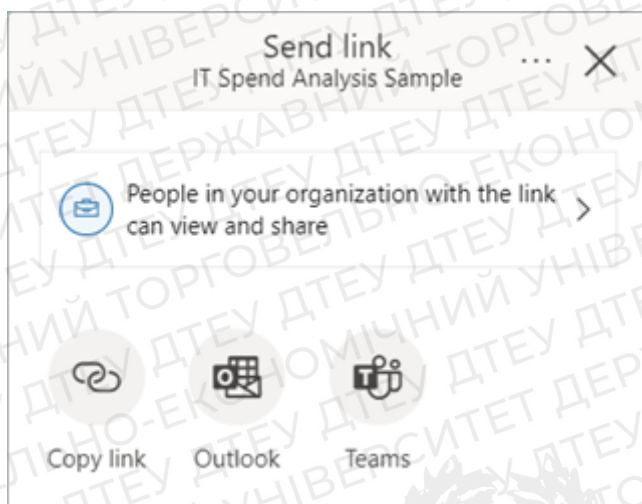


Рисунок 1. 2 Посилання на звіт

Так само можна поділитися звітом, надавши користувачу на пряму:

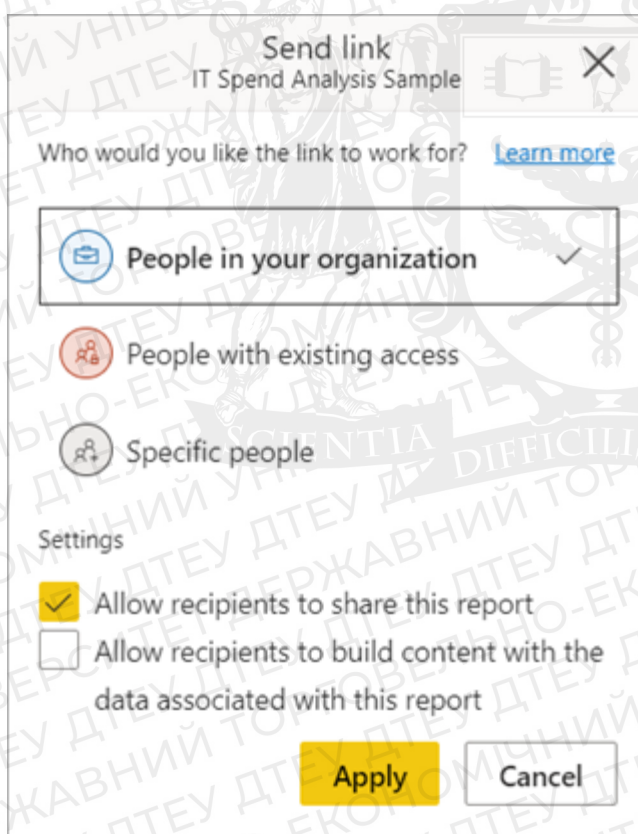


Рисунок 1. 3 Посилання на звіт прямим доступом

1.3 Огляд існуючих аналітичних платформ, які можуть бути використані для аналізу даних і побудови аналітичних звітів

Існує багато різних аналітичних платформ, які слугують для аналізу даних та побудови аналітичних звітів. Особисто наша група працювала в таких як:

1. *Power BI* – це програмне забезпечення від Microsoft для вивчення

бізнес-аналітики (BI). Націлена на візуалізацію та можливості бізнес-аналітики з доволі простим інтерфейсом, в якому створюються власні звіти і інформаційні панелі. [9]

2. *Deductor Studio* – це програмне забезпечення для аналізу даних, ціль якого формулювати процес прийняття рішень та становлення його «у потік». Схема роботи *Deductor Studio* доволі простий: спочатку йде імпорт інформації; подалі його обробка та візуалізація та експорт, що призводить до прийому оброблених даних. [10]
3. *Loginom* – це аналітична платформа, котра забезпечує інтеграцію, очищення та аналіз інформації для прийняття більш ефективних управлінських рішень. [11]

Microsoft Power BI – одна з хмарних сервісів бізнес-аналітики для створення і перетворення необроблених даних у значну інформацію, інтуїтивно зрозумілими таблицями та візуалізаціями.

Серед можливих візуалізацій: гистограма з накопиченням, лінійна діаграма з накопичення, графік, діаграма з областями, діаграма кругова, точкова діаграма, карта, таблиця, діаграма дерева, матриця, воронка, картка, багатостороння картка, ключовий показник ефективності.

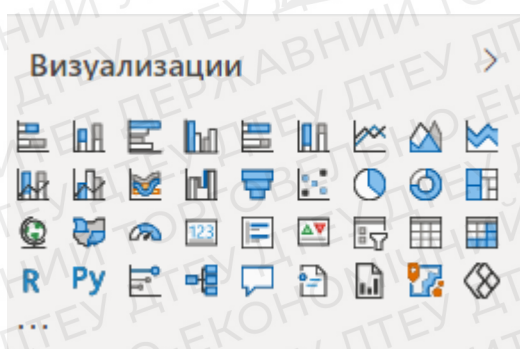


Рисунок 1. 4 Види візуалізацій

Серед можливих завантажень даних:

1. Група «Файл»;
2. Група «База даних»;
3. Група «Power Platform»;
4. Група «Azure»;
5. Група «Веб служби»;

6. Група «Інше».

Висновки до розділу 1

Спеціаліст з Data Science вирішує проблему отримання корисної інформації з надмірної кількості даних, використовуючи різні методи і етапи аналізу інформаційних даних.

Щоб досягнути успішного аналізу даних для отримання найкращого результату – використовують найбільш поширені методи, такі як: кореляційний і регресійний аналіз.

Оскільки аналіз даних є однією з важливих частин сучасного бізнесу, то на даний момент доступно багато різноманітних платформ Data Science з різними можливостями та підтримкою великого об'єму. Більшість платформ підтримують візуалізацію даних і готові звіти для користувачів.

РОЗДІЛ 2 ПОБУДОВА АНАЛІТИЧНОГО ЗВІТУ ЗАСОБАМИ POWER BI

2.1 Створення аналітичного звіту засобами Microsoft Power BI

Задачею є проаналізувати дані компанії Apple за певний проміжок часу та зробити їх візуалізацію на аналітичній платформі Power BI. Обов'язковими елементами візуалізації є таблиця (або матриця), стовпчикова діаграма, кругова або кільцева діаграма, графік та карта. Результатом виконання задачі є аналітичний звіт.

Вхідної інформацією для розв'язування задачі є інформація із сайту <https://www.kaggle.com/>, а саме:

1. Звіт компанії щодо прайс-листу в 26 країнах світу;
2. Історичні котирування акцій компанії;
3. Магазини Apple по країнам (або регіонам);
4. Загальна продукція Apple.

Отож, сформовано містку базу даних, що містить основні показники фінансів; продажі продукції та прибуток від них та іншої основної діяльності; зріст/спадання, середнє значення акцій; моделі продукції, ціни на них, кількість переглядів на сайті, початок їхнього продажу, властивості гаджету; місце знаходження та кількість магазинів в конкретній країні або регіону; додано карту країн та до якої групи доходів вона відноситься (високий прибуток, більше середнього або менше середнього).

Кінцевою або ж вихідною інформацією є складений аналітичний звіт, де за допомогою вбудованих можливостей Microsoft Power BI було візуалізовано основні показники і досягнення та діяльність компанії Apple за певні роки.

Перший етап, який передбачає практична частина, є збір даних. Отож, джерелами є певні чотири Excel-таблиці з даними, які були описані трохи вище.

Другий етап – це імпорт цих даних саме до системи Power BI. А саме «Файл → Отримати дані → Вибрати тип даних Excel → Імпортувати дані». Далі нам на допомогу йде Power Query, саме через нього ми і здійснюємо трансформацію даних, змінюючи тип, формат даних та інше. В цій частині

було здійснено трансформацію типу даних щодо цілих чисел та видозміна формату часу. Грошовий формат та грошову одиницю встановлено згідно тих даних, указаних в таблиці.

Одним з наступних етапів є побудова *моделі даних*, тобто структури зв'язаних між собою таблиць даних. Було побудовано схему даних типу «сузір'я» (рис. 2.1). Використовується коли в сховищі присутні декілька таблиць-фактів, що застосовують загальні таблички вимірювань і консольні таблиці, тоді сховище будується за схемою «сузір'я», тому дане розташування розмірності виглядає, неначе сукупність зірок у моделі схеми «Галактика».

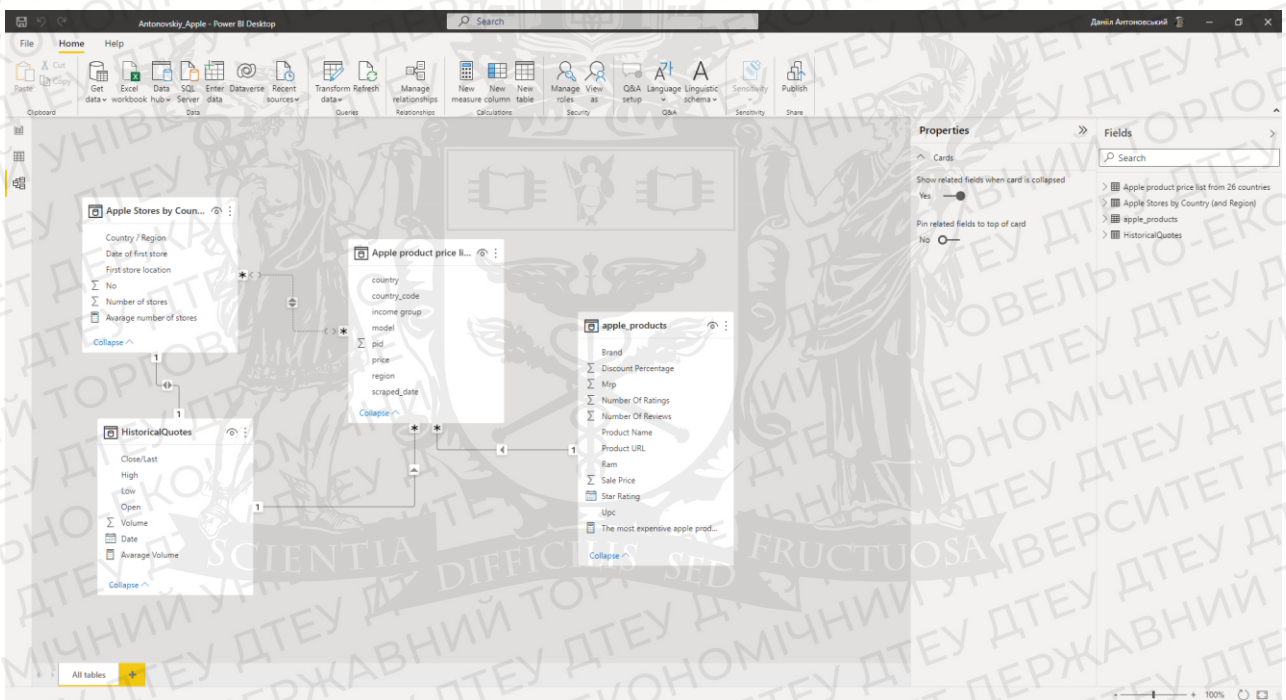


Рисунок 2. 1 Схема даних типу «сузір'я»

Як можна зауважити: центральною таблицею вийшла «Apple products price list». Поєднані між собою за датами, окрім таблиця «Apple products», в зв'язок між назвами і брендом продукту.

Задля отримання додаткової кількості інформації було додано певну кількість обчислюваних мір, а саме:

- 1) *Average number of stores* – таблиця до якої вставлена міра «Apple Stores by Country (and Region)». Отож, цей стовпець має функцію обчислення середнього значення кількості магазинів у кожному з регіонів або країн. Даний показник вираховується через *AVERAGE*. Кінцева формула використання (рис. 2.2): $Average\ number\ of\ stores = AVERAGE('Apple\ Stores\ by\ Country\ (and$

Region)=[Number of stores]);

No	Country / Region	Date of first store	First store location	Number of stores
1	United States	15.05.2001	Tysons Corner Center, Tysons Corner, Virginia	272
2	Japan	30.12.2003	Okinaka, Tokyo	10
3	United Kingdom	20.12.2004	Regent Street, London	38
4	Canada	21.05.2005	Yorkdale Shopping Centre, Toronto	28
5	Italy	31.03.2007	Centro Commerciale Romita Est, Rome	17
6	Australia	19.06.2008	George Street, Sydney	22
7	China mainland	19.07.2008	Sanjium, Beijing	43
8	Switzerland	25.09.2008	Rue de Riv, Geneva	4
9	Germany	08.10.2008	Alte Fassade, Munich, Bavaria	4
10	France	07.12.2009	Carroussel du Louvre, Paris	20
11	Spain	04.09.2010	La Maquillista, Barcelona	11
12	Hong Kong	24.09.2011	KK Mall, Central and Western District	6
13	Netherlands	03.02.2012	Hirschlopplein, Amsterdam	3
14	Sweden	15.09.2012	Tibby Centrum, Stockholm	3
15	Brazil	15.02.2014	VillageMall, Barra da Tijuca, Rio de Janeiro	2
16	Turkey	05.04.2014	Zorlu Center, Istanbul	2
17	Belgium	18.06.2015	Avenue de la Triperie, Brussels	1
18	United Arab Emirates	29.10.2015	Mall of the Emirates, Dubai	1
19	United Arab Emirates	29.10.2016	Yas Mall, Abu Dhabi	4
20	Macau	25.08.2016	Galaxy Macau	2
21	Mexico	24.09.2016	Centro Santa Fe, Santa Fe, Mexico City	2
22	Singapore	27.05.2017	Orchard Road	3
23	Taiwan	01.07.2017	Taipei 101, Taipei	2
24	South Korea	27.01.2018	Gyeongju, Seoul	1
25	Austria	24.02.2018	Kalmanstrasse, Vienna	3
	Thailand	10.11.2018	Konnam, Bangkok	2

Рисунок 2. 2 Розрахунок середньої кількості магазинів

2) The most expensive apple products – таблиця до якої вставлена міра «apple_products». Дана міра створена для вимірювання найбільш дорогого та продаваного продукту компанії Apple. Даний показник вираховується через MAX. Кінцева формула використання (рис. 2.3): The most expensive apple products = MAX(apple_products[Sale Price]);

Product Name	Product URL	Brand	Sale Price	Qty	Discount Percentage	Number of Ratings	Number of Reviews
APPLE iPhone 8 Plus (64 GB)	https://www.rikart.com/apple-iphone-8-plus-64-gb/...	Apple	49900	49900	0	3411	356
APPLE iPhone 8 Plus (Black, 256 GB)	https://www.rikart.com/apple-iphone-8-plus-256-gb/...	Apple	49900	49900	0	3411	356
APPLE iPhone 8 Plus (Silver, 256 GB)	https://www.rikart.com/apple-iphone-8-plus-256-gb/...	Apple	49900	49900	0	3411	356
APPLE iPhone 8 (64 GB)	https://www.rikart.com/apple-iphone-8-64-gb/...	Apple	77000	77000	0	21202	774
APPLE iPhone 8 (128 GB)	https://www.rikart.com/apple-iphone-8-128-gb/...	Apple	82900	82900	0	3411	356
APPLE iPhone 8 Plus (Space Grey, 64 GB)	https://www.rikart.com/apple-iphone-8-plus-space-grey-64-gb/...	Apple	49900	49900	0	3411	356
APPLE iPhone 8 Plus (Space Grey, 256 GB)	https://www.rikart.com/apple-iphone-8-plus-space-grey-256-gb/...	Apple	77000	77000	0	21202	774
APPLE iPhone X Max (Silver, 64 GB)	https://www.rikart.com/apple-iphone-x-max-silver-64-gb/...	Apple	89900	89900	0	1424	249
APPLE iPhone X Max (Silver, 256 GB)	https://www.rikart.com/apple-iphone-x-max-silver-256-gb/...	Apple	42990	42990	20	79512	6798
APPLE iPhone X (Black, 64 GB) (includes EarPods, Power Adapter)	https://www.rikart.com/apple-iphone-x-black-64-gb/...	Apple	76990	42990	16	79512	6798
APPLE iPhone X (Space Grey, 64 GB)	https://www.rikart.com/apple-iphone-x-space-grey-64-gb/...	Apple	42990	42990	20	79512	6798
APPLE iPhone X (White, 128 GB) (includes EarPods, Power Adapter)	https://www.rikart.com/apple-iphone-x-white-128-gb/...	Apple	42990	42990	20	79512	6798
APPLE iPhone 11 Pro Max (Midnight Green, 256 GB)	https://www.rikart.com/apple-iphone-11-pro-max-midnight-green-256-gb/...	Apple	212000	212000	0	2035	308
APPLE iPhone 11 Pro Max (Midnight Green, 64 GB)	https://www.rikart.com/apple-iphone-11-pro-max-midnight-green-64-gb/...	Apple	212000	212000	0	16178	101
APPLE iPhone 11 Pro Max (Midnight Green, 512 GB)	https://www.rikart.com/apple-iphone-11-pro-max-midnight-green-512-gb/...	Apple	212000	212000	0	1019	103
APPLE iPhone 11 Pro Max (Midnight Green, 64 GB)	https://www.rikart.com/apple-iphone-11-pro-max-midnight-green-64-gb/...	Apple	149000	149000	29	7088	521
APPLE iPhone 11 Pro Max (Midnight Green, 512 GB)	https://www.rikart.com/apple-iphone-11-pro-max-midnight-green-512-gb/...	Apple	212000	149000	15	7088	521
APPLE iPhone 12 Pro Max (Midnight Green, 128 GB)	https://www.rikart.com/apple-iphone-12-pro-max-midnight-green-128-gb/...	Apple	127000	127000	0	1019	103
APPLE iPhone 12 Pro Max (Midnight Green, 512 GB)	https://www.rikart.com/apple-iphone-12-pro-max-midnight-green-512-gb/...	Apple	209000	127000	18	2001	180
APPLE iPhone 12 Pro Max (Midnight Green, 256 GB)	https://www.rikart.com/apple-iphone-12-pro-max-midnight-green-256-gb/...	Apple	139000	127000	6	180	43
APPLE iPhone 12 Pro Max (Midnight Green, 128 GB)	https://www.rikart.com/apple-iphone-12-pro-max-midnight-green-128-gb/...	Apple	127000	127000	6	180	43
APPLE iPhone 12 Pro Max (Midnight Green, 256 GB)	https://www.rikart.com/apple-iphone-12-pro-max-midnight-green-256-gb/...	Apple	139000	127000	6	180	43
APPLE iPhone 12 Pro Max (Midnight Green, 512 GB)	https://www.rikart.com/apple-iphone-12-pro-max-midnight-green-512-gb/...	Apple	209000	127000	6	180	43
APPLE iPhone 12 Pro Max (Midnight Green, 128 GB)	https://www.rikart.com/apple-iphone-12-pro-max-midnight-green-128-gb/...	Apple	127000	127000	6	180	43
APPLE iPhone 12 Pro Max (Midnight Green, 256 GB)	https://www.rikart.com/apple-iphone-12-pro-max-midnight-green-256-gb/...	Apple	139000	127000	6	180	43
APPLE iPhone 12 Pro Max (Midnight Green, 512 GB)	https://www.rikart.com/apple-iphone-12-pro-max-midnight-green-512-gb/...	Apple	209000	127000	6	180	43

Рисунок 2. 3 Розрахунок топ продажу

3) Average Volume – таблиця до якої вставлена міра «HistoricalQuotes». Наступна міра обчислює середній об'єм кількості продажу акцій компанії по

конкретним рокам. Даний показник вираховується через *Avarage*. Кінцева формула використання (рис. 2.4): $Avarage\ Volume = AVERAGE(HistoricalQuotes[Volume])$.

The screenshot shows the Tableau Desktop interface. At the top, the calculated field editor displays the formula: `Average Volume = AVERAGE(HistoricalQuotes[Volume])`. Below this, a data table is visible with columns for Date, Close/Last, Volume, Open, High, and Low. The data spans from 28.02.2020 to 01.06.2020. A watermark of the Ukrainian coat of arms is overlaid on the table.

Date	Close/Last	Volume	Open	High	Low
28.02.2020	\$273.36	106721200	\$257.26	\$278.41	\$256.37
27.02.2020	\$273.52	80215180	\$281.3	\$286	\$272.96
26.02.2020	\$292.65	49678420	\$286.53	\$297.88	\$286.5
25.02.2020	\$288.08	57668800	\$300.95	\$302.13	\$286.13
24.02.2020	\$296.18	55548850	\$297.26	\$304.18	\$289.23
21.02.2020	\$313.05	82426420	\$318.62	\$320.45	\$310.5
20.02.2020	\$320.9	25141490	\$322.63	\$324.65	\$318.21
19.02.2020	\$323.82	23495990	\$320	\$324.57	\$320
18.02.2020	\$319	38210550	\$315.36	\$319.75	\$314.61
14.02.2020	\$324.95	20028450	\$324.74	\$329.98	\$322.85
13.02.2020	\$324.87	33688800	\$324.19	\$326.22	\$320.35
02.12.2020	\$972.3	28423570	\$211.37	\$322.23	\$321.47
02.11.2020	\$319.41	23580780	\$323.6	\$323.9	\$318.71
02.10.2020	\$321.55	27337220	\$314.18	\$321.55	\$313.85
02.07.2020	\$323.03	28421020	\$322.37	\$323.4	\$318
02.06.2020	\$325.21	26356890	\$322.57	\$325.22	\$320.2648
02.05.2020	\$323.45	29708720	\$323.52	\$324.76	\$318.95
02.04.2020	\$318.85	34154120	\$315.31	\$319.64	\$315.9345
02.03.2020	\$308.66	24868800	\$304.1	\$313.49	\$302.21
21.01.2020	\$309.51	48827190	\$320.93	\$322.68	\$308.28
30.01.2020	\$323.87	31685820	\$320.544	\$324.09	\$318.75
29.01.2020	\$324.94	51419990	\$324.45	\$327.85	\$321.38
28.01.2020	\$317.69	40584890	\$312.6	\$318.4	\$312.19
27.01.2020	\$308.95	48485020	\$310.06	\$311.77	\$304.88
24.01.2020	\$316.31	36638890	\$320.25	\$323.33	\$317.538
23.01.2020	\$319.23	26127990	\$317.92	\$319.36	\$315.65
22.01.2020	\$317.7	24618200	\$318.18	\$319.97	\$317.31
21.01.2020	\$316.57	27721820	\$317.19	\$319.02	\$316
17.01.2020	\$318.73	34454120	\$316.37	\$318.74	\$315
16.01.2020	\$312.24	27287250	\$313.59	\$315.7	\$312.09
15.01.2020	\$311.34	30480880	\$311.85	\$315.5	\$309.55
14.01.2020	\$312.68	4863480	\$316.7	\$317.57	\$312.17
13.01.2020	\$316.96	30527220	\$311.64	\$317.07	\$311.15
01.10.2020	\$310.33	35217070	\$310.6	\$312.67	\$308.25
01.09.2020	\$309.65	42621540	\$307.219	\$310.49	\$306.27
01.08.2020	\$309.19	33090950	\$297.16	\$304.439	\$297.156
01.07.2020	\$298.39	27877660	\$299.84	\$300.9	\$297.48
01.06.2020	\$299.8	25644640	\$299.79	\$299.96	\$292.75

Рисунок 2. 4 Розрахунок об'єму продажу акції

Для створення міри потрібно знайти вкладку «Основне» в розділі «Обчислення» натиснути «Нова міра». Наступний крок це введення виразу DAX, який надасть змогу обчислення певного значення, основуючись вже отриманими даними. Відбувається створення міри у відповідній таблиці, результат якої ми побачимо лише на візуалізації.

Згідно виконаними вище перелічених етапів, я перейшов до візуалізації моєї роботи, в результаті якої було створено 2 інформаційних сторінки, що містять в собі:

Сторінка №1 (рис. 2.9): має кластерну діаграму (рис. 2.5), матричну таблицю (мал 2.6), інтерактивну карту (мал 2.7) та кругову діаграму (рис. 2.8).

Отож, кластерна діаграма показує статистику, що сумує кількість *Last/Close*. *The Last Close price* – це найостанніша ціна закриття акції компанії Apple на відповідній фондовій біржі. З боку графіка показується сума кількості, а вже синім *Close/Last* в грошовій одиниці долару.[14]

Поняття *кластерна діаграма* означає, що це стовпчаста діаграма, що представляє собою дані у вертикальних стовпцях послідовно, хоч діаграми

прості і виготовлені, але ці діаграми є також доволі складними для візуального перегляду, якщо існують конкретна категорія із кількома порівняльними серіями, то це легко передивитись дякуючи цій діаграмі.[13]

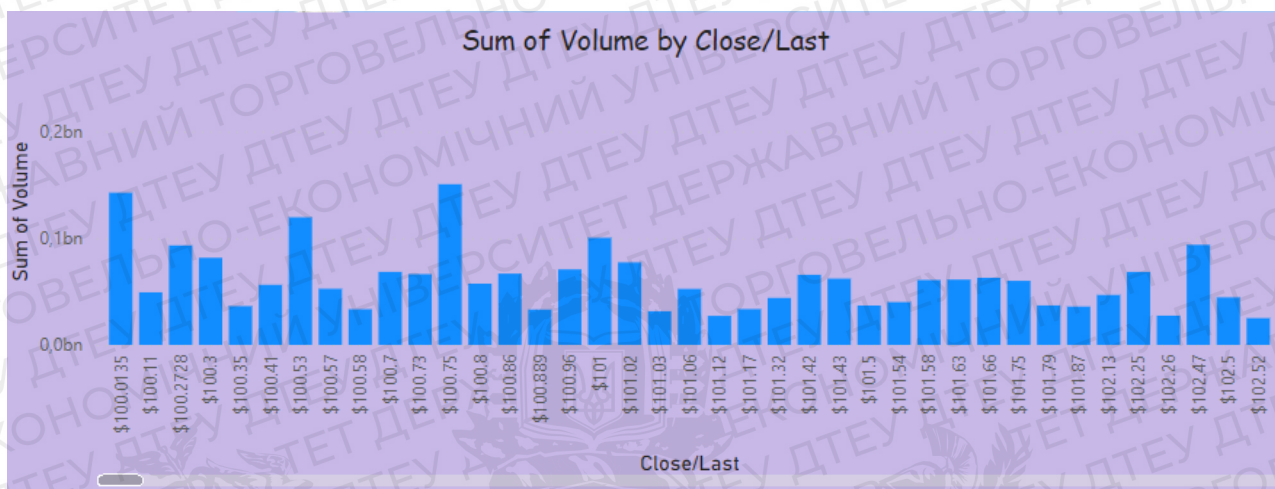


Рисунок 2. 5 Кластерна діаграма

В матричній таблиці зображено кількість магазинів, країни або регіони, в яких вони були відкриті та першу локацію цих пунктів прийому покупців. Внизу таблиці сформовано підсумовану найбільшу кількість цих магазинів та прив'язану до неї країну. Найбільшу кількість має, звичайно ж США, адже вони є творцями своєї продукції та й загалом створювали продукцію саме для споживання у власній країні, але згодом почали поширювати і на країни Європи і Азії.

Матрична таблиця – це таблиця, яка розширюється і вертикально, і горизонтально. Має певну кількість строк та стовбців згідно з унікальних значень вказаних полях. Як правило слугують таблицями для пошуку даних.[15]

No	First Country / Region	Sum of Number of stores
1	United States	272
2	Japan	10
3	United Kingdom	38
4	Canada	28
5	Italy	17
6	Australia	22
7	China mainland	43
8	Switzerland	4
9	Germany	16
10	France	20
11	Spain	11
12	Hong Kong	6
13	Netherlands	3
Total	Australia	522

Рисунок 2. 6 Матрична таблиця

Наступне, що я маю на першій сторінці – це інтерактивну карту, що відображає країни та середню кількість магазинів. Як ми можемо побачити, саме США є країною, яка має найбільшу кількість магазинів. Причину якої, було названо вище.

Інтерактивна карта – це веб-карта, яка дозволяє користувачам натискати, панаромувати та масштабувати для вилучання додаткової інформації. Вони можуть представляти собою географічні карти для відображення даних по конкретному місцерозташуванню та використовувати шари демонстрування кількох наборів даних.[16]



Рисунок 2. 7 Інтерактивна карта

Останньою візуалізацією є кругова діаграма, яка відтворює в відсотковому співвідношенні середню кількість продажів акцій по рокам. Розписано з 2010 до 2020 років. Найбільший показник саме за 2010 рік, 18,67%.

Круговою діаграмою використовують, як поліпшену версію всіх кругових діаграм, що усуває складність зчитування кругових діаграм. Як відомо, зображають співвідношення частин до цілих, де всі частини представляють 100% при зборі разом.

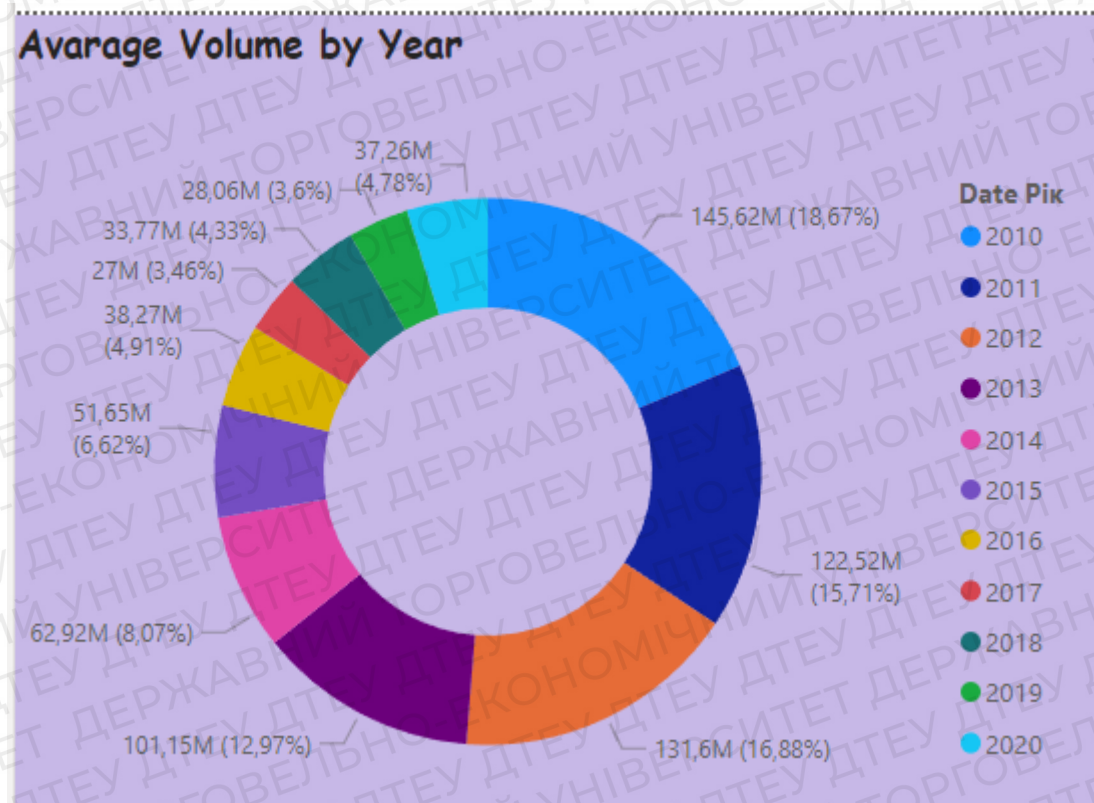


Рисунок 2. 8 Кругова діаграма

Отож, загальний вигляд ця сторінка має такий:

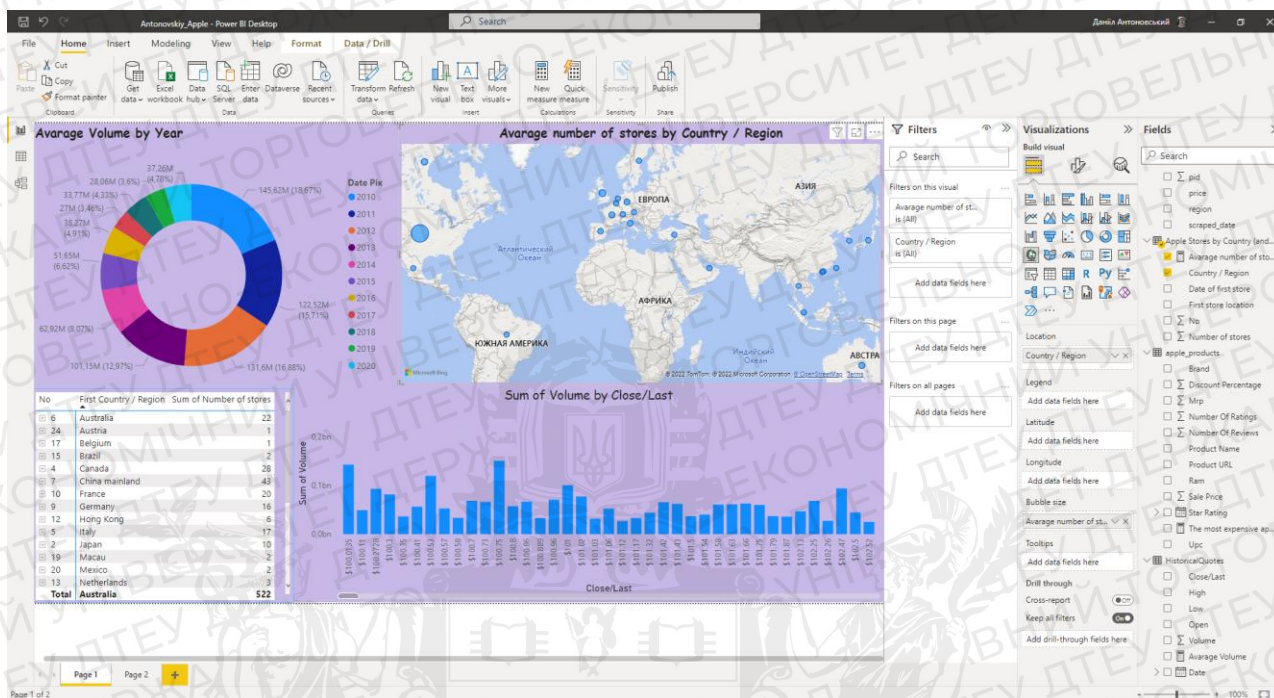


Рисунок 2.9. Вигляд першої сторінки

Сторінка №2 (рис. 2.12): включає діаграму з областями або площинну діаграму (рис. 2.11) та картку (рис. 2.10).

Перше, що було додано, це міра, яка вираховує найдорожчий продукт Apple, а саме iPhone 12 Pro, за допомогою візуалізації «картка».

Одне з чисел, наприклад обсяг продажів, узагальнена кількість певного підрахунку, частка ринку, подеколи є найважливішою частиною, що вам потрібно відстежити. Тому за допомогою типу візуалізації в Power BI, яка називається *картка*, може бути найліпшим шляхом перегляду конкретного числа.[17]

Як і у всіх власних візуалізаціях Power BI, карту можна створити за допомогою редакторів запитів чи запитань і відповідей.

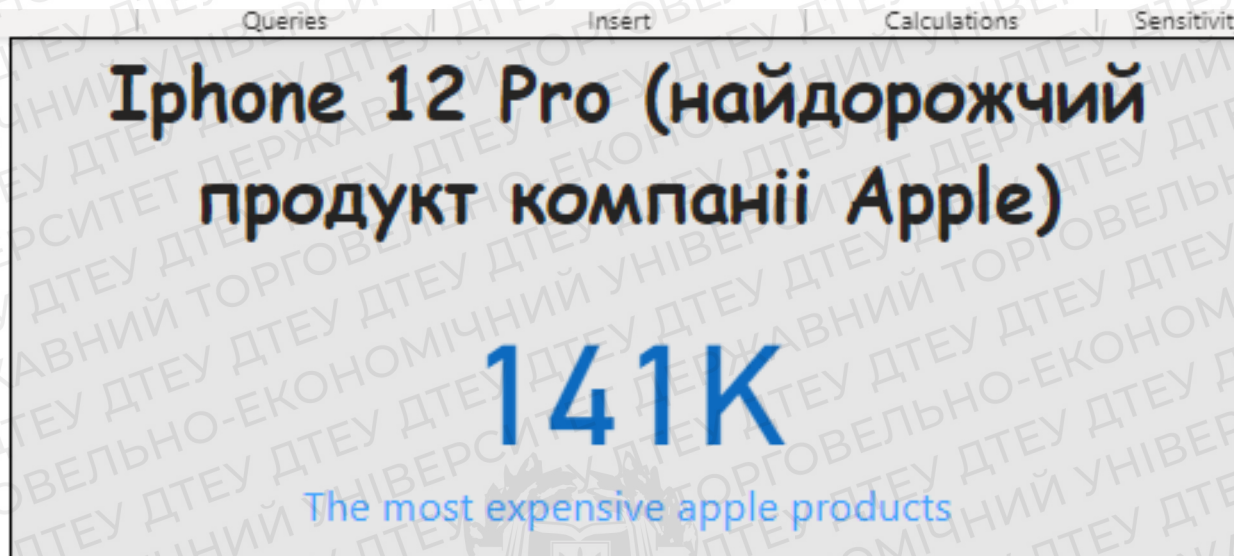


Рисунок 2. 10 Картка

Наступний етап, це діаграма з областями або ж площинна діаграма підкреслює величину змін у часі та використовуються, щоб привернути увагу до загального значення певного тренду.

Наприклад, в мене вийшла візуалізація, яка підкріплена слайдером для більш детальної візуалізації суми цін продажу на конкретний продукт компанії Apple.

Основні площинні діаграми є чудовим вибором:

- задля того, щоб побачити і порівняти тенденцію конкретного обсягу за часовий ряд;
- для окремого ряду, який і представляє собою фізичну лічильну множину.[18]

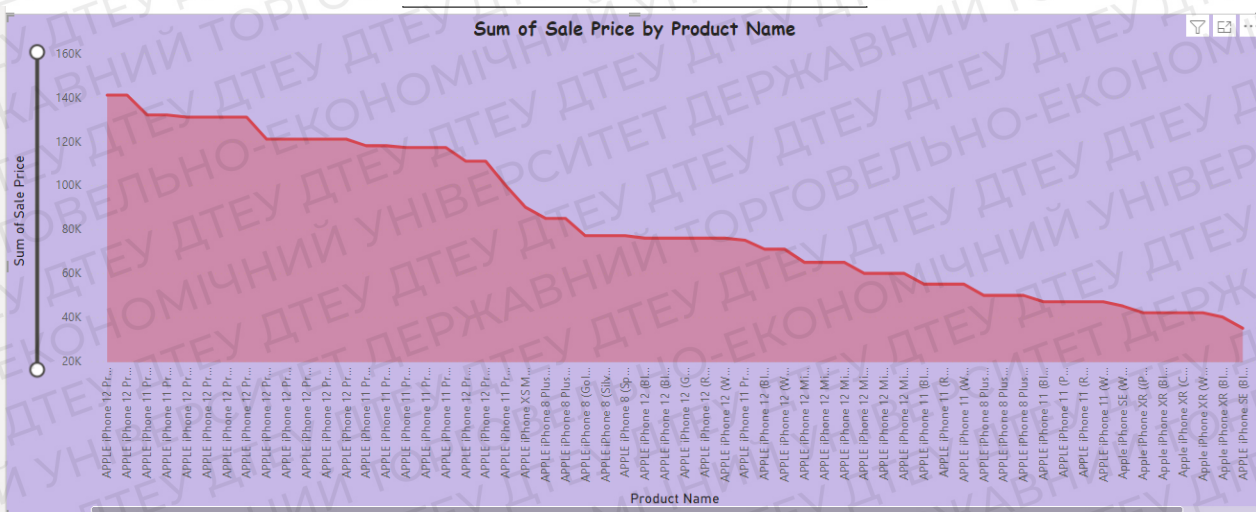


Рисунок 2. 11 Діаграма з областями

Отож, загальний вигляд інша сторінка має вже такий:

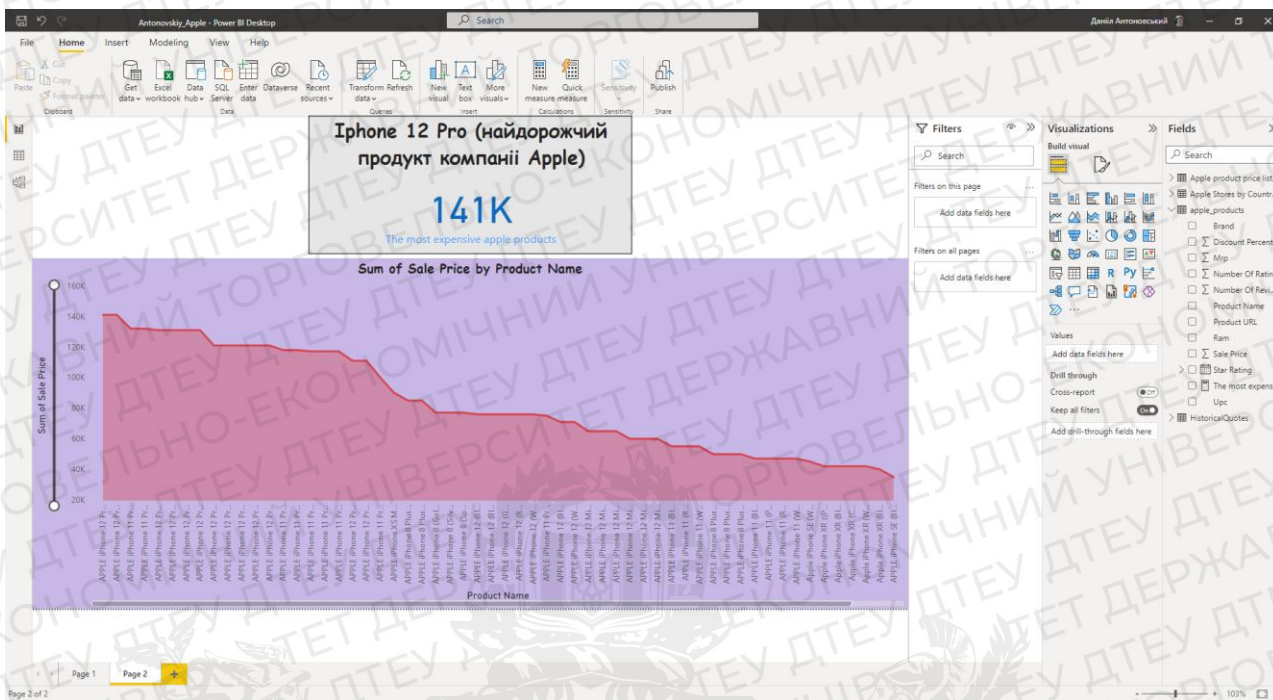


Рисунок 2. 12 Видяг другої сторінки

Обов'язково хочеться відмітити, що кожна візуалізація є інтерактивною, при натисканні конкретної ланки, ряду таблиці, частки кругової діаграми та інших елементів буде конкретизовано у всіх вище перерахованих візуальних об'єктів.

Тобто підсвічено на графіках, діаграмах, таблицях іншим виділяючим кольором. На карті відображено конкретне місцезрешташування.

2.2 Публікація аналітичного звіту

При завершенні формуванні аналітичного звіту, він був опублікований на порталі Power BI Service.

Отож, для початку публікації звіту необхідно перейти у вкладку «Основне», вибрати категорію «Надати спільний доступ» та натиснути «Опублікувати» (рис. 2.13).

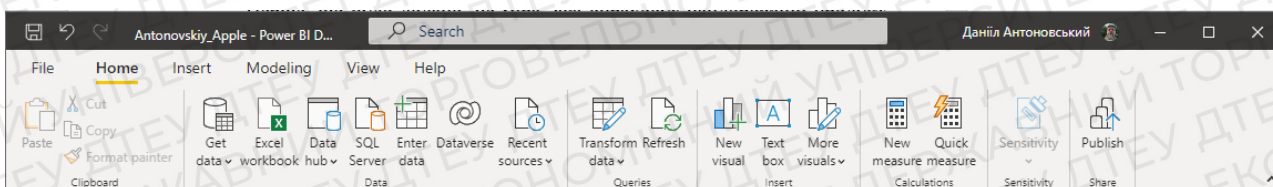


Рисунок 2. 13 Процес публікування звіту

Наступне, що ми робимо це вибираємо місце призначення публікованого звіту. В моєму ж випадку обирається власну робочу область Power BI (рис. 2.14).

Publish to Power BI

Select a destination

Search

My workspace

Select

Cancel

Рисунок 2. 14 Місце призначення звіту

Останнім етапом, при успішній публікації, з'явиться готовий звіт на порталі Power BI Service (рис. 2.15).

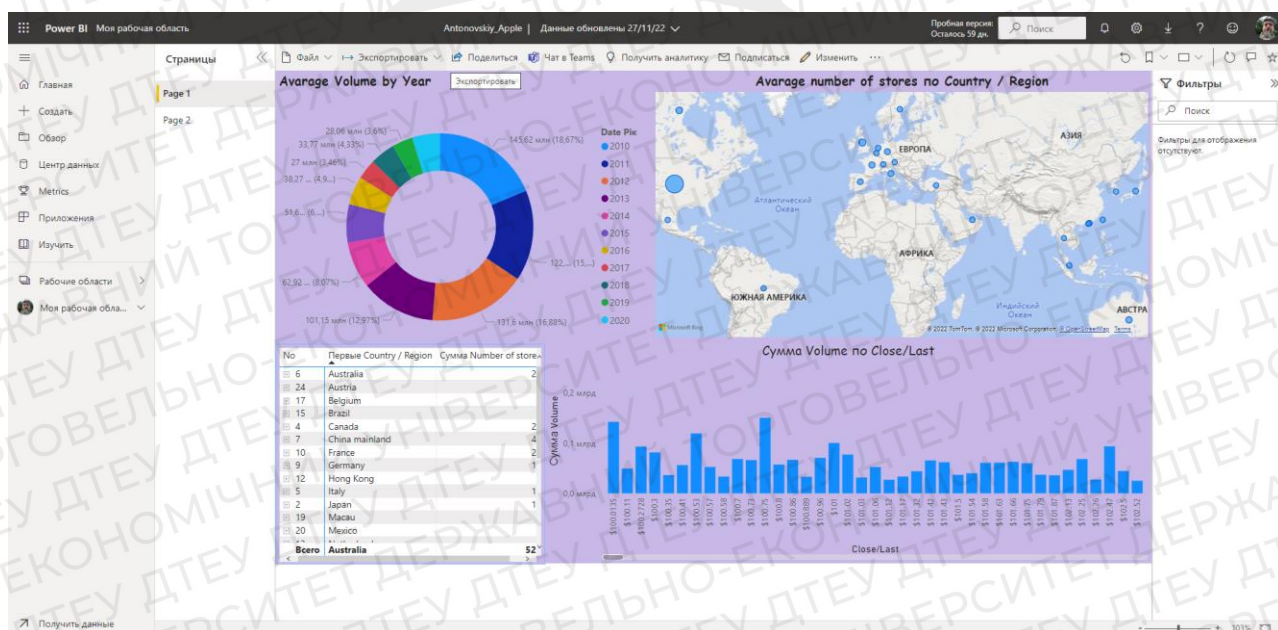


Рисунок 2. 15 Місце призначення звіту

Функції доступу для опублікованого звіту:

- редагування сторінок аналітичного звіту;

- надання спільного доступу;
- збереження копії;
- друк сторінки;
- завантаження РВІХ-файлу;
- експортування файлу в Power Point, Excel, PDF.

Щоб оновити дані у звітах Power BI на сервері звітів Power BI, потрібно створити розклад оновлення. Щоб створити даний план, треба перейти до налаштувань робочої області Power BI, далі вибрати частоту оновлення, часовий пояс і час.

Також є можливість вказати адресу електронної пошти, якщо щось піде не так під час оновлення.

Висновки до розділу 2

Імпортувавши дані та обробивши їх, здійснюючи розрахунки за допомогою вбудованих функцій, як «міра» та в решті-решт розробивши аналітичний звіт в платформі Power BI, було проведено аналіз даних

та було визначено середню кількість магазинів компанії Apple у світі, розрахунок топ продажу певних продуктів компанії та найдорожчого гаджету, обчислення середнього об'єму кількості продажу акцій компанії по конкретним рокам.

Дані були виведені в аналітичному звіті за допомогою візуалізаційних інструментів: кластерної діаграми, матречної таблиці, інтерактивної карти, кругової діаграми, картки, діаграми з областями.

Країна, яка мала найбільші продажі у 2010 році – США. А топ продуктом став за останні 10 років став iPhone 12 Pro.



РОЗДІЛ 3 АНАЛІЗ ДАНИХ В СЕРЕДОВИЩІ R STUDIO

3.1 Використання середовища R Studio для аналізу даних

Обробка даних – це основний інструмент щодо пізнання нашого світу, який людина отримує з різноманітних джерел. Сама суть системного аналізу є інтерактивний процес, включаючи дослідження інформаційного потоку зовні, які можна представити за допомогою візуалізації.

В минулому високоефективні комп'ютери та доступне ПЗ (програмне забезпечення) дозволили впроваджувати повну інформацію та технічні цикли процесів. Зазвичай цей процес включає в собі наступні кроки:

- анотування даних;
- отримання загальних відомостей структури даних;
- редагування завантажених показників, заміна або видалення пропущених значень;
- доступ до обробки даних та їх завантаження з джерел і об'єднання взаємопов'язаних вхідних таблиць;
- графічне представлення даних та результатів обчислень інформативною формою;
- оформлення цих результатів, підготовка таблиць, діаграм до публікаційної якості.

Системний аналітик – це спеціаліст, який має глибокі знання у своїй професійній діяльності і невеликий досвід у цілому спектрі аналітичних навиків.

Основними інструментами аналітиків є:

- Excel;
- SQL;
- Python;
- SAS;
- Oracle та інші.

Але є ще один із найважливіших інструментів аналітика, що має назву “R”, тобто це мова програмування та середовище для статистичних обчислень та графічного аналізу.

Дана мова аналізує дані відкритого коду, які підтримуються великими та активними науково-дослідницькими спільнотами світу.

З появою дешевого і легкого доступу аналізу даних парадигма змінилася. Замість попереднього налаштування всіх параметрів аналізу, цей процес став інтерактивним у значній мірі. У той же час результатом кожного етапу аналізу слугує вхідними даними для подальшого введення.

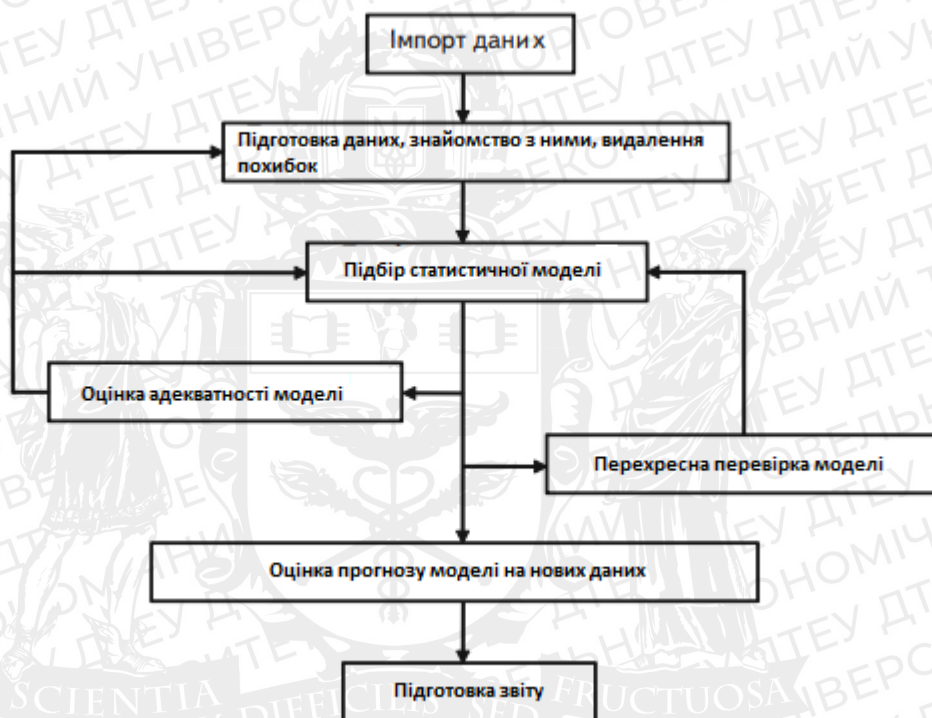


Рисунок 3.1 Схема типового аналізу даних

Сьогодні R беззаперечний лідер серед вільно розповсюджених систем статистичного аналізу, а про це свідчать численні переможці категорій щорічного відкритого конкурсу програмного забезпечення Bossie Awards для системи R в декількох номінаціях. Провідні університети світу, великі корпоративні аналітики та дослідницькі центри постійно використовують R під час виконання наукових та інженерних розрахунків і створення масштабних інформаційних проектів.

Завдяки широкій статистичній знань на основі пакетів цього середовища та широкій підтримці наукової спільноти відтворення скриптів R поступово стає загальноприйнятим «стандартом», як у журнальних публікацій, так і для неформального спілкування вчених у всьому світі.

Система статистичного аналізу та візуалізації даних R складається з наступних частин:

- мова програмування R високого рівня, яка дозволяє виконувати різноманітні операції над об'єктами, векторами, матрицями, списками і т.д.;
- великий набір функцій обробки даних, що були зібрані в окремі пакети (package);
- розвинена система підтримки, що включає оновлення компонентів даного середовища, його інтерактивну допомогу та різні освітні ресурси, що призначені для початкового вивчення R.

Двоє молодих новозеландських вчених Росс Ихак і Роберт Джентльмен, які афішували розробку R, шлях якого належить до 1993 року. За основу було взято мову програмування розвинутої комерційної системи статистичної обробки даних S-Plus, після чого було створено її безкоштовну і вільну реалізацію, яка різниця від свого попередника легко розширеною модульною архітектурою.

Зараз мова R має велику численність користувачів, що повідомляють про нові помилки авторам додаткових бібліотек, включно і самої системи R, котрі в майбутньому швидко вирішуються.

Програма, в якій виконується аналіз даних, за допомогою мови R називається R Studio (рис. 3.2). [19]



Рисунок 3.2 Логотип R Studio

Приклад вигляду самого інтерфейсу в операційній системі (рис. 3.3):

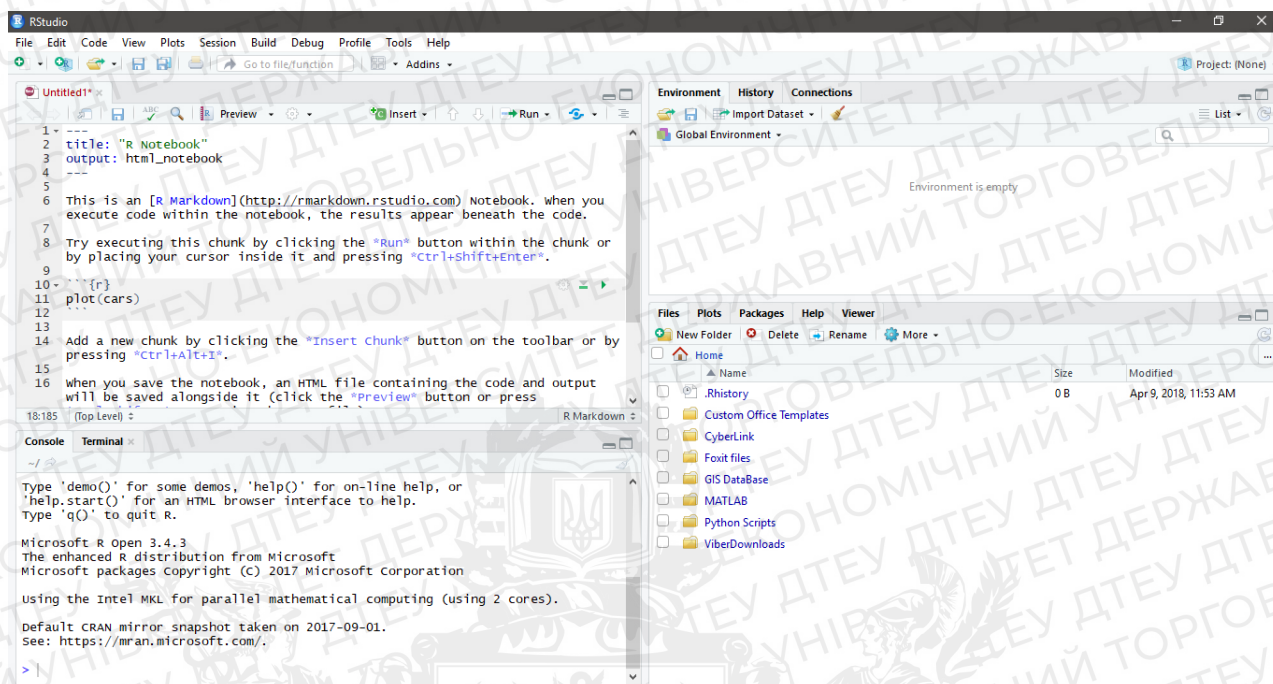


Рисунок 3.3 Приклад інтерфейсу R

В R вбудовано великий обсяг довідникового матеріалу, які допоможуть користувачу у разі роботи в програмі. Вбудованій системі допомоги міститься:

- роз'яснення;
- посилання на літературу;
- приклади для функції із встановлених пакетів.

Викликання даного довідника з використання функцій можна викликати таким чином (рис. 3.4):

Вік (місяці)	Вага (кг)
<code>help.start()</code>	Загальна довідка
<code>help(func)</code> або <code>??func</code>	Довідка щодо функції <code>func</code>
<code>help.search(func)</code> або <code>??func</code>	Пошук в довідці записів, що містять <code>func</code>
<code>example(func)</code>	Приклади використання функції <code>func</code>
<code>RSiteSearch(func)</code>	Пошук записів, що містять <code>func</code> в онлайн-довідниках і заархівованих розсилках
<code>apropos(func, mode='function')</code>	Список всіх доступних функцій, в назві яких є <code>func</code>
<code>data()</code>	Список всіх демонстраційних даних, що містяться в завантажених пакетах
<code>vignette()</code>	Список всіх доступних інструкцій щодо завантажених пакетів
<code>vignette(func)</code>	Список інструкцій за темою <code>func</code>

Рисунок 3.4 Функції виклику довідки R

Робочий простір – це поточна область R у пам'яті комп'ютера, яке містить створені користувачем об'єкти (тобто вектори, матриці, функції, списки даних і т.д.). Після завершення кожного сеансу ви маєте змогу зберегти свою робочий простір і він буде автоматично завантажатися у наступному запуску програми. Використовуйте стрілки «вгору» та «вниз» для переходу між вже введеними командами. Це дозволить вам викликати попередню команду, замінити її і одразу виконати її, натискаючи «Enter».

Поточна робоча директорія – це директорія, яка містить файли даних, куди зберігаються результати роботи. За допомогою функції `getwd()` можна дізнатися робочу директорію на даний момент та за допомогою `setwd()` призначити іншу. У випадку необхідності імпорту файлу, що перебуває не в робочій директорії, то тоді необхідно написати повний шлях до нього, завжди укладаючи назви файлів та директорій у лапки (рис. 3.5).

```
setwd("C:/myprojects/project1")
options()
options(digits=3)
x <- runif(20)
summary(x)
hist(x)
savehistory()
save.image()
q()
```

Рисунок 3.5 Приклад коду у робочому просторі

Запуск R розпочинає інтерактивну сесію за замовчуванням, де введення даних здійснюється за допомогою клавіатури, а результат виводиться на екран. Проте можна і запустити також команди з скрипта, а виведення можливе безпосередньо в різних пристроях.

Щодо введення: за допомогою функції `source("filename")` відбувається запуск скрипту в тому разі, якщо не було прописано шляху до файлу, тобто він міститься у поточній робочій директорії. Розширення має бути `.R`, але це не обов'язкова умова.

Щодо текстового висновку: за допомогою функція `sink("ім'я_файлу")` відбувається вивід всіх результатів із назвою «ім'я_файлу» та за замовчуванням

даний файл вже наявний і нова версія записується поверх старої. Параметр `append=TRUE` дозволяє додавати новий текст до файлу замість того, щоб писати його в старий текст. Параметр `split=TRUE` дозволяє вивід результатів на екран і до текстового файлу включно. Виконання команди `sink()` без аргументів просто відновить вихід на екран.

Щодо графічного висновку: команда `sink()` хоча й управляє висновком текстової частини, але не робить жодного впливу на виведення графіки.

Наприклад, маємо три файли програмного коду (`Script1.R`, `Script2.R` тощо), які команда `source("Script1")` запускає і виведе результати на екран. Якщо наберем дані команди (рис. 3.6), то виконається код із файлу `Script2.R`, після чого результати з'являться на екрані. Окрім, до файлу `myoutput` буде додано текстове виведення, а графічне – збережено у `mygraphs.pdf`.

```
sink("myoutput", append=TRUE, split=TRUE)
pdf("mygraphs.pdf")
source("script2.R")
```

Рисунок 3.6 Код перший

Зрештою, якщо введемо вже такі команди (мал 3.7), то виконаним буде код файлу `Script3.R`, після чого також виведуться результати на екрані. Отож, ані текстове, ані графічне збереженням у файлі не будуть.

```
sink()
dev.off()
source("script3.R")
```

Рисунок 3.7 Код другий

Графічно послідовність можна представити таким чином (рис. 3.8):

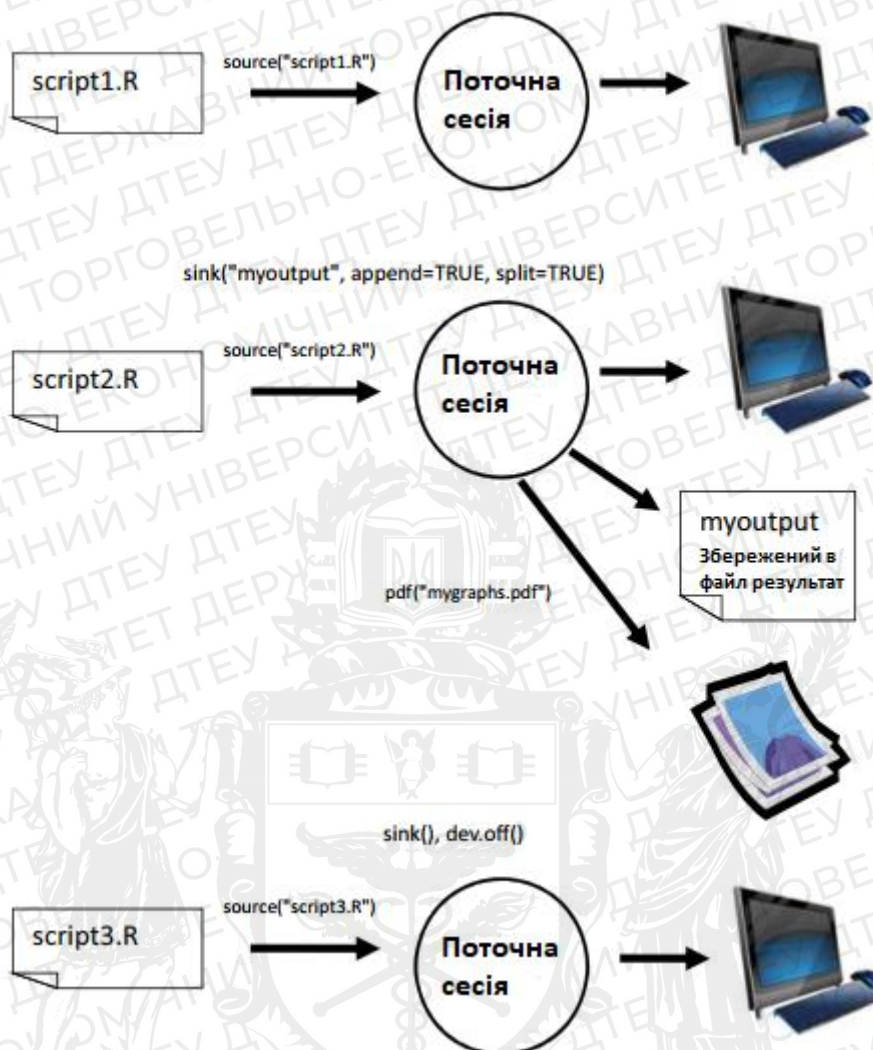


Рисунок 3.8 Графічне зображення проробленого коду

В базовій установці мови R є багато можливостей, проте деякі опції програми були реалізовані в додаткових модулях, які власне можна встановити. Їх є більше ніж 2 500 створених різними користувачами модулів, яке називаються *packages* пакети.

Тобто *пакети* – це зібрання певного функціоналу R, включаючи дані і скомпільованого програмного коду в певному форматі. *Бібліотекою* називають ту директорію, в якій пакети містяться на вашому ПК. Розташування бібліотеки показує функція `.libPath()`, а за вивід на екран назв всіх наявних в конкретній бібліотеці пакетів відповідна функція `library()`.

R містить велику кількість функцій, призначення яких є управління пакетами. Завантаження певного пакету виконується через команду `install.packages()` – введена без аргументів, використовується для виклику списку дзеркал з CRAN сайту. Вже після вибору певного дзеркала, то буде доступним список всіх відкритих пакетів. Вибираємо один із наведених, який

буде завантаженим. У випадку, якщо ви знаєте пакетну назву, що встановлюватиметься, тоді потрібно повідомити функції цю назву у вигляді аргументу. Наприклад *install.packages("base")*.

Пакет необхідно встановити один раз. Зауважу, що як і всі програми, ці пакети можуть оновлятися їхнім розробником. Використання команди *update.package()* допоможе у вирішенні даної проблеми.

Для отримання докладної інформації з приводу встановленого пакету, використовуйте функцію *installed.packages()*, яка виводить всі наявні з номерами версій, назв та іншої інформації на екран.

При встановленні пакету, він завантажується із сайту CRAN у вашу бібліотеку. Щоб ним скористатися в поточній сесії, необхідно за допомогою команди *library()* якщо завантажити.

Після завантаження пакету стають доступними нові функції, набори даних тощо. Невеликий набір даних надається разом із демонстраційним програмним кодом, який дає можливість тестування нових можливостей. Система довідника містить опис функцій разом з їх застосуванням та інформаційним блоком про кожен вбудований набір. Вивід короткого опису пакету і алфавітний покажчик всіх функцій і наборів даних можна за допомогою функції *help(package="назва_пакету")*. Загалом за допомогою команди *help()* можна з'ясувати нові деталі наборів даних чи функцій.

Існує багато помилок, які схильні робити як новачки, так і досвідчені R програмісти. Якщо було отримано повідомлення про помилку, переконаймося, що ви виконали наведені нижче дії:

- використовуйте неправильний регістр: *help()*;
- ви забули поставити дужки де потрібно: *install.packages("base")* працює, а *install.packages(base)* – видає помилку;
- забули поставити дужки при виклику функції. Наприклад, слід ввести *help()* замість *help*. Дужки обов'язкові, навіть якщо немає аргументів;
- в ОС Windows використовуйте \ у шляху до файлу. R розглядає зворотні косі риски, як контрольні символи. *setwd("c:/mycomputer")* створює помилку, а *setwd("c:\mycomputer")* навпаки;

- було введено функцію з пакету, що ще не є завантаженим. Функція `order.clusters()` є частиною пакета `glus`. Спроба використати цей пакет до його завантаження призведе до програмної помилки.

Першочерговий етап для аналізу даних – є створення набору даних, який містить інформацію для вивчення в необхідному форматі. R розподіляє ці задачі на такі підзадачі, як:

- вибірка типу даних;
- введення даних у протилежному форматі. [20]

Набір даних – зазвичай прямокутний масив даних, рядки в якому відповідають спостереженням, стовпці їхнім ознакам.

R обробляє широкий спектр структур даних, таких як скаляри, вектори, матриці, масиви даних, списки і таблиці. Вони різняться за типом даних, способом створення, складністю пристрою та способом розмітки вилучення окремих елементів. Дані структури представлені схематично на рис. 3.9:

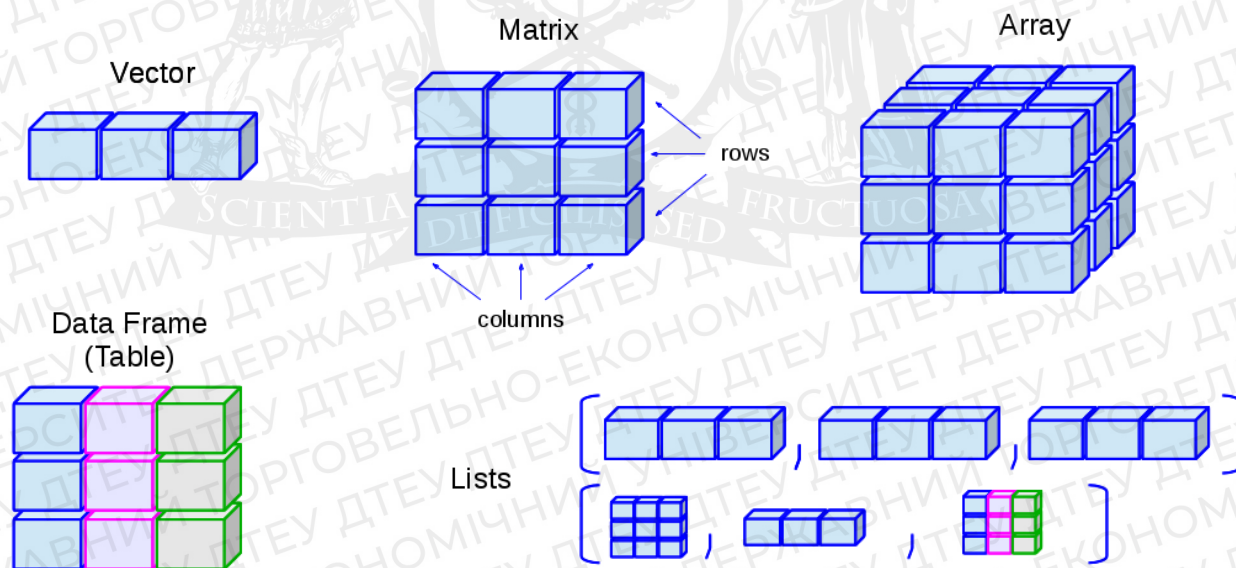


Рисунок 3.9 Тип структури даних

Вектор (vector) – це одномірний масив даних, який може містити як числові, так і текстові і логічні значення. Для його створення використовується функція `c()` (рис. 3.10).

```
a <- c(1, 2, 5, 3, 6, -2, 4)
b <- c("one", "two", "three")
c <- c(TRUE, TRUE, TRUE, FALSE, TRUE, FALSE)
```

Рисунок 3.10 Приклад вектору

Матриця (matrix) – це двохвимірний масив даних, кожен елемент якого має однаковий тип (числовий/текстовий/логічний). *Vector* містить елементи матриці, команди *nrow* та *ncol* визначають числа рядків і стовпців. Має формат: `myMatrix <- matrix(vector, nrow=число_рядків, ncol=число_стовпців)`.

Масиви даних (array) мають схожість з матрицями, але можуть мати понад два виміри. Тут *vector* містить лише дані, *dimensions* є числовим вектором, який визначає розмірність кожного вимірювання, *dimnames* – не обов'язковий список імен розмірів (мал 3.11).



```

dim1 <- c("A1", "A2")
dim2 <- c("B1", "B2", "B3")
dim3 <- c("C1", "C2", "C3", "C4")
z <- array(1:24, c(2, 3, 4), dimnames=list(dim1, dim2, dim3))
z

```

Output structure:

```

, , C1
  B1 B2 B3
A1  1  3  5
A2  2  4  6
, , C2
  B1 B2 B3
A1  7  9 11
A2  8 10 12
, , C3
  B1 B2 B3
A1 13 15 17
A2 14 16 18
, , C4
  B1 B2 B3
A1 19 21 23
A2 20 22 24

```

Рисунок 3.11 Приклад масиву даних

Таблиця даних (data frame) – є більш поширеним об'єктом порівняно з матрицею, так як різні стовпці можуть містити різні типи даних (числа, текст тощо). Функції `col1`, `col2`, `col3`, ... – вектори типу текстового/числового/логічного, які будуть стовпцями таблиці. Оголошення назв кожного стовпця виконується за допомогою `names()` (рис. 3.12). [21]

```
mydata <- data.frame(col1, col2, col3, ...)
```

patientID	age	diabetes	status
<dbl>	<dbl>	<fctr>	<fctr>
1	25	Type1	Poor
2	34	Type2	Improved
3	28	Type1	Excellent
4	52	Type1	Poor

4 rows

Рисунок 3.12 Приклад таблиці даних

3.2 Моделювання і аналіз цін на акції Apple

Задачею є ситуаційне моделювання та аналіз цін на акції компанії Apple за певний проміжок часу та відобразити результат візуально на платформі RStudio. Обов'язковим елементом візуалізації є таблиці з графіками, діаграмами, на яких зображено результати роботи.

За основу, взято бібліотеки, а саме:

1. `quantmod`;
2. `xts`;
3. `PerformanceAnalytics`;
4. `rugarch`.

Отже, в даних бібліотеках міститься дані про продажі Open (ціна відкриття першої угоди), Close (ціна закриття першої угоди), High (максимальна ціна), Low (мінімальна ціна), Adjusted метод, загальна кількість цих продажів у певний період часу.

Кінцевою інформацією є графіки з даними, в яких було проведено аналіз денного продажу, щоденного дохід, прогнозування на приклад створення ситуаційного випадку.

Отже, перше ми зробимо після завантаження бібліотек, це визначимо коливання денної ціни на продукцію компанії Apple. Візьмемо за приклад проміжок часу в одинадцять років, а саме від 01.01.2008 до 31.12.2019. Для того, щоб працювати з даними, використовуємо функцію `getSymbols()` з пакету `quantmod`, який забезпечить послідовний інтерфейс саме для імпорту даних з різних джерел в наш робочій простір. За замовчуванням імпортує дані як `xts`-об'єкт. Символ-артикуль для Apple є «AAPL». Задля перевірки точності даних та дат можемо викликати таблицю через функцію `View()`, або також можна перевірити за допомогою `head()` та `tail()`, вививши із таблиці пару перших та

останніх стрічок нашого датасету. Наступний інструмент, який буде і в подальшому фігурувати це `chartSeries()`, який використовується для створення стандартних фінансових діаграм таким з тимчасовими рядом, як об'єкт і служить базовою функцією майбутньому доповненню технічного аналізу. Якщо ми впишемо дату, наприклад грудень 2019 року (`chartSeries(AAPL["2019-12"])`), тоді маємо наступне (рис. 3.13):

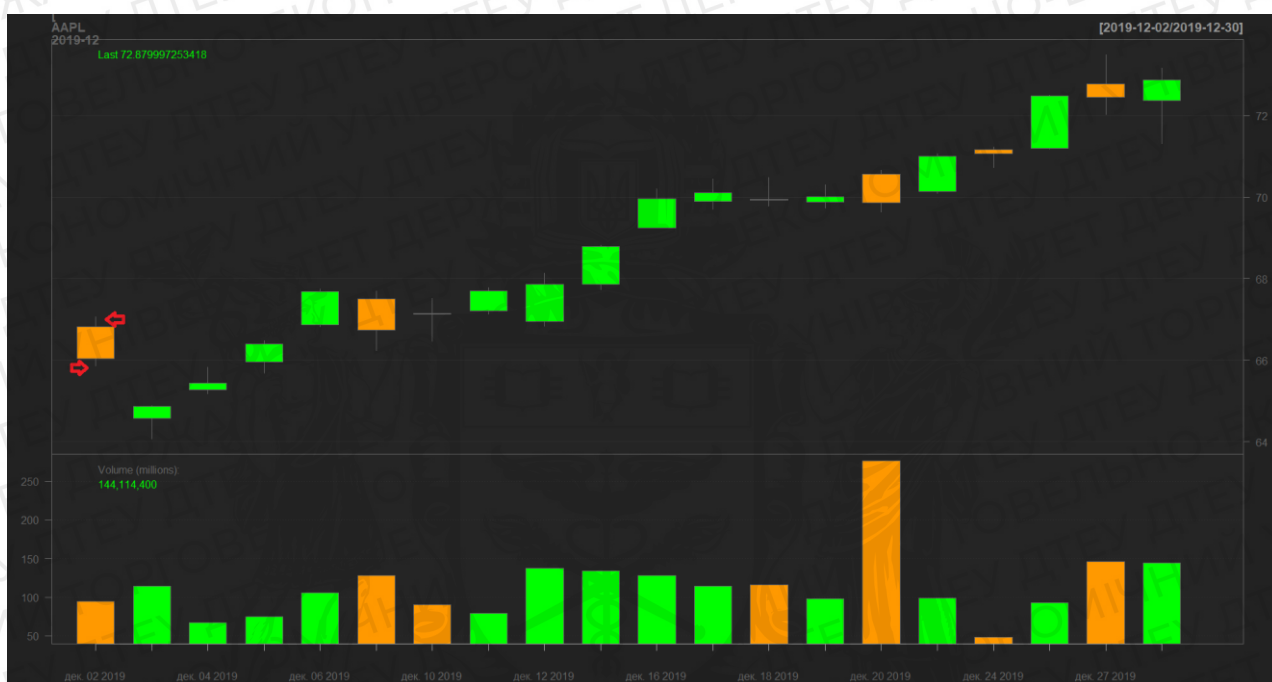


Рисунок 3.13 Продажі компанії за грудень 2019 року

Отже, бачимо, що на графіку зображено зелені та помаранчеві кубики. На кожному з них я рисочки (позначено на першому кубі), верхня свідчить про максимальну ціну дня (High of the day), інша ж – мінімальну (Low of the day). Помаранчевий колір свідчить, що ціна закриття першої угоди (Close price) була менше ніж її відкриття (Open price). З зеленим кольором навпаки, Open price був менше ніж Close price.

Якщо ж ми приберемо дату в (`chartSeries(AAPL)`), можемо бачити всі зміни за весь період від 01.01.2008 до 31.12.2019 (рис. 3.14). Лише зауважмо, що ми не можемо використовувати прямо такі дані з графіку, краще конвертувати в днівні продажі компанії Apple.



Рисунок 3.14 Продажі компанії за 11 років

Далі переходимо до самих денних продажів. Щоб розрахувати простий чи складний дохід цін використовуємо функцію *CalculateReturns*. Операцію проводимо через *return*, в дужках залишаємо *APPL.Close* та відкриваємо його через *View()*. Бачимо, що перший рядок *NA*, інші вже є пораховані. Переконалися в розрахунках не складно, достатньо відкрити таблицю і відняти другий від першого в стовпці *Close* та розділити знову перший (тобто $(B-A)/A$). Рядок *NA* в принципі не потрібен, тому залишаємо як *return[-1]*.

Вививши гістограму, бачимо, як дані періодичності денного прибутку. Середнє число близько нулю. Деякі дні мають високий показник прибутку (*High return*), так і низький показник (*Low return*).

Створимо гістограму прибутку з додатковою кривою (рис. 3.15), що підходить для методів *density* (відображає графік щільності) та *normal* (для відображення встановленої звичайно лінії над середньою). На графіку *density* зображений зеленим, *normal* червоним, а саму гістограму – синім кольорами.

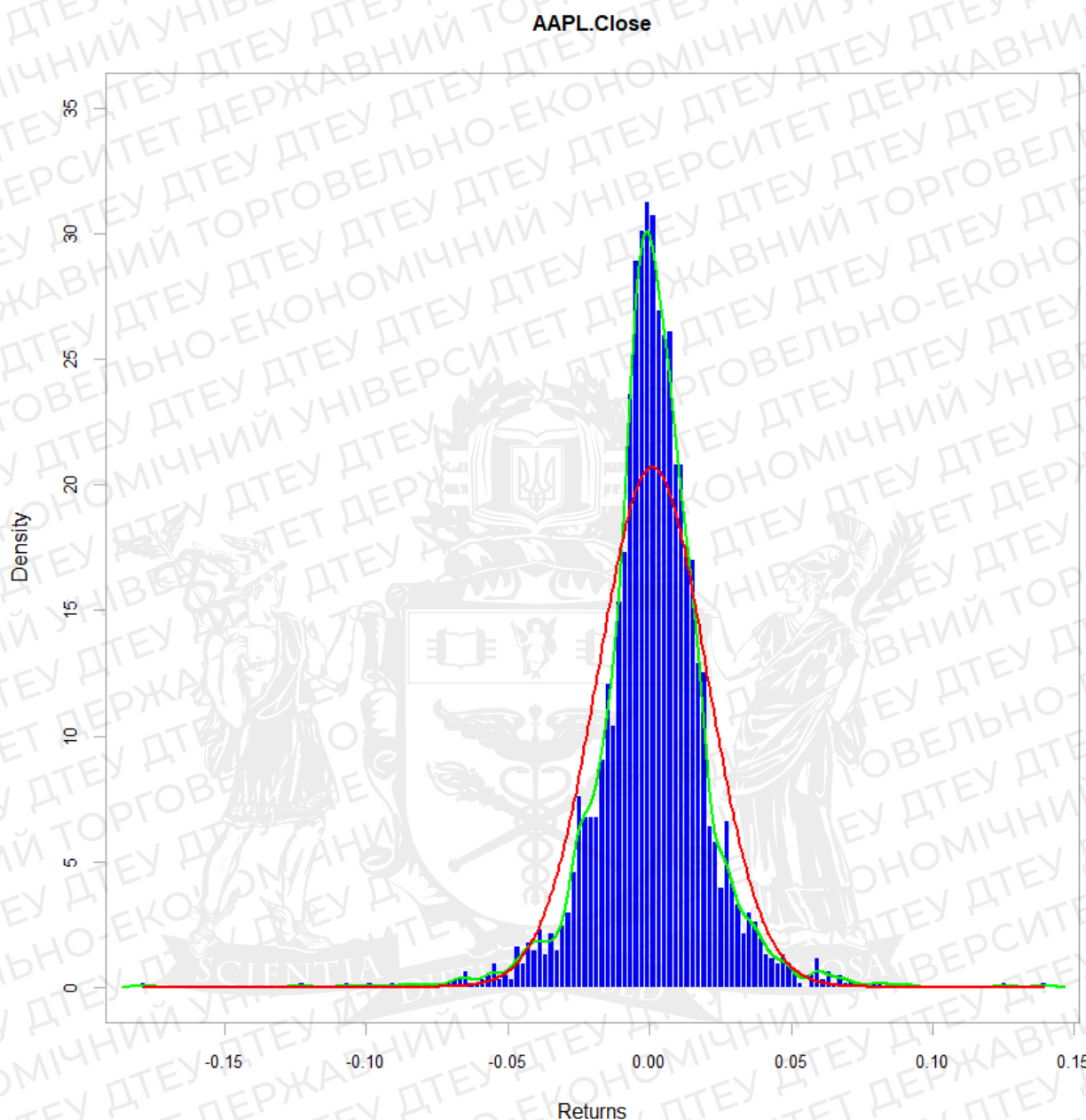


Рисунок 3.15 Гістограма прибутку з кривою

Знову створимо `chartSeries` для нашого `return`'ну і бачимо, що графік зеленого кольору, що є коливанням мінливості прибутку. Даний плагін говорить про оцінку за певний часовий період, на якому ми не бачимо ніяких тенденцій та сигналів. Отже є періоди, де ми маємо багато різних коливань на ринку, зокрема на лівому крайньому періоді, де ми маємо фінансовий кризу в 2008 році, коли ринок був найбільш хитким. Та й також можемо бачити далі, меншу мінливість, що означає, що ринок більш стабільний і спокійний (рис. 3.16).

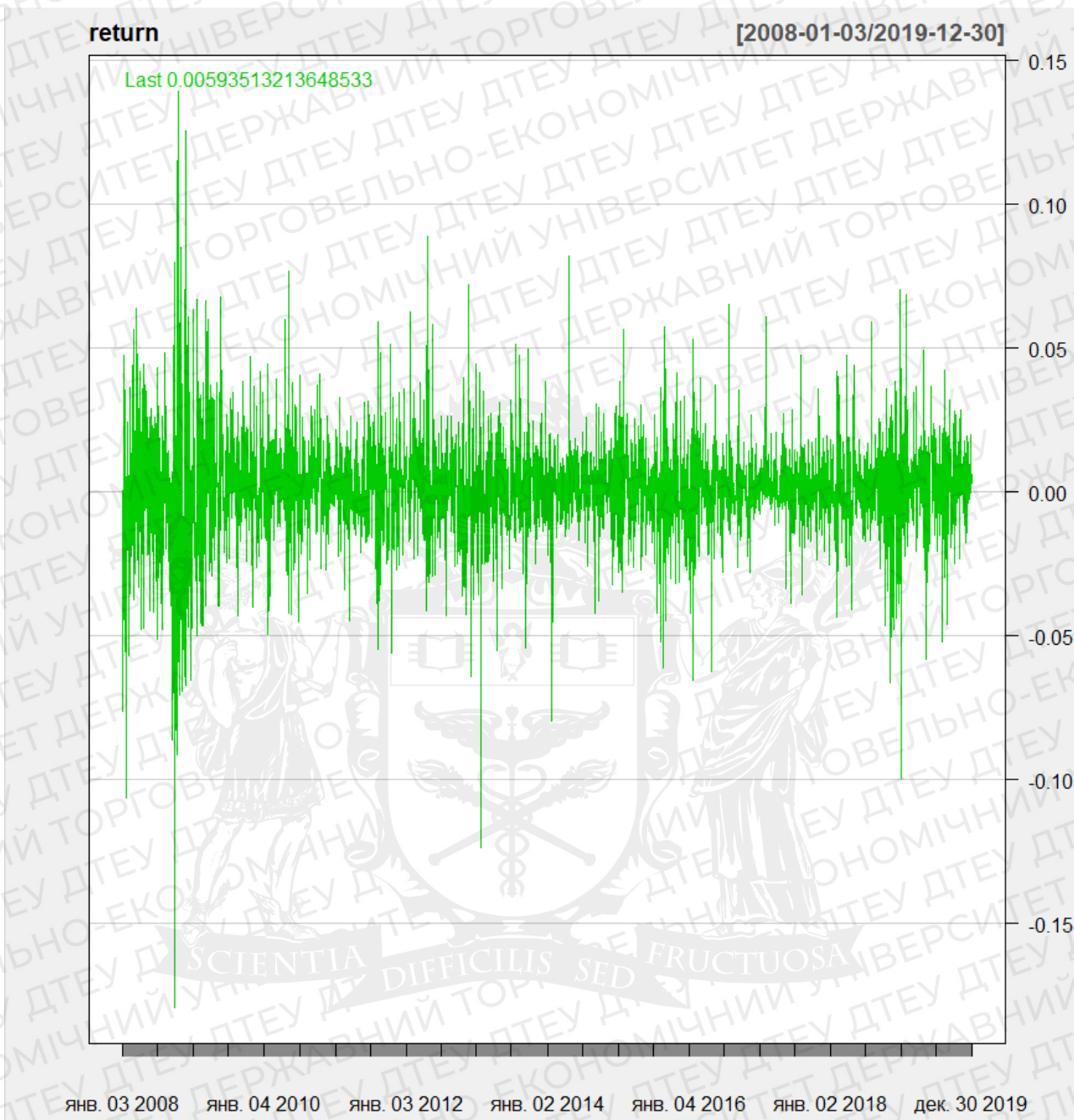


Рисунок 3.16 Гістограма річного мінливості на ринку

Наступним кроком є побудова пакувальника для створення діаграм показників продуктивності в лінійному графіку. Їй відповідає команда `chart.RollingPerformance`. Дані беремо з тих же одинадцять років (2008-2019 рр.), далі вираховуємо місячну мінливість за допомогою команди `width`. В середньому в одному місяці ми маємо 22 днів продажу. Функцію (`FUN`), яку ми використовуємо, це `sd.annualized`. Вказуємо в `scale` 252, що є кількістю днів продажу в рік. Ну а на останнє в цій функції ми можемо вказати назву таблиці, а саме «Місячна волатильність (мінливість) компанії Apple» (рис. 3.17). Даний графік буде значно краще відображати дані з графіку вище (рис. 3.16). На ньому зображено високу мінливість зліва та і всі інші показники, де була висока та

низька волатильність компанії.

Місячна волатильність (мінливість) Apple

2008-01-03 / 2019-12-30

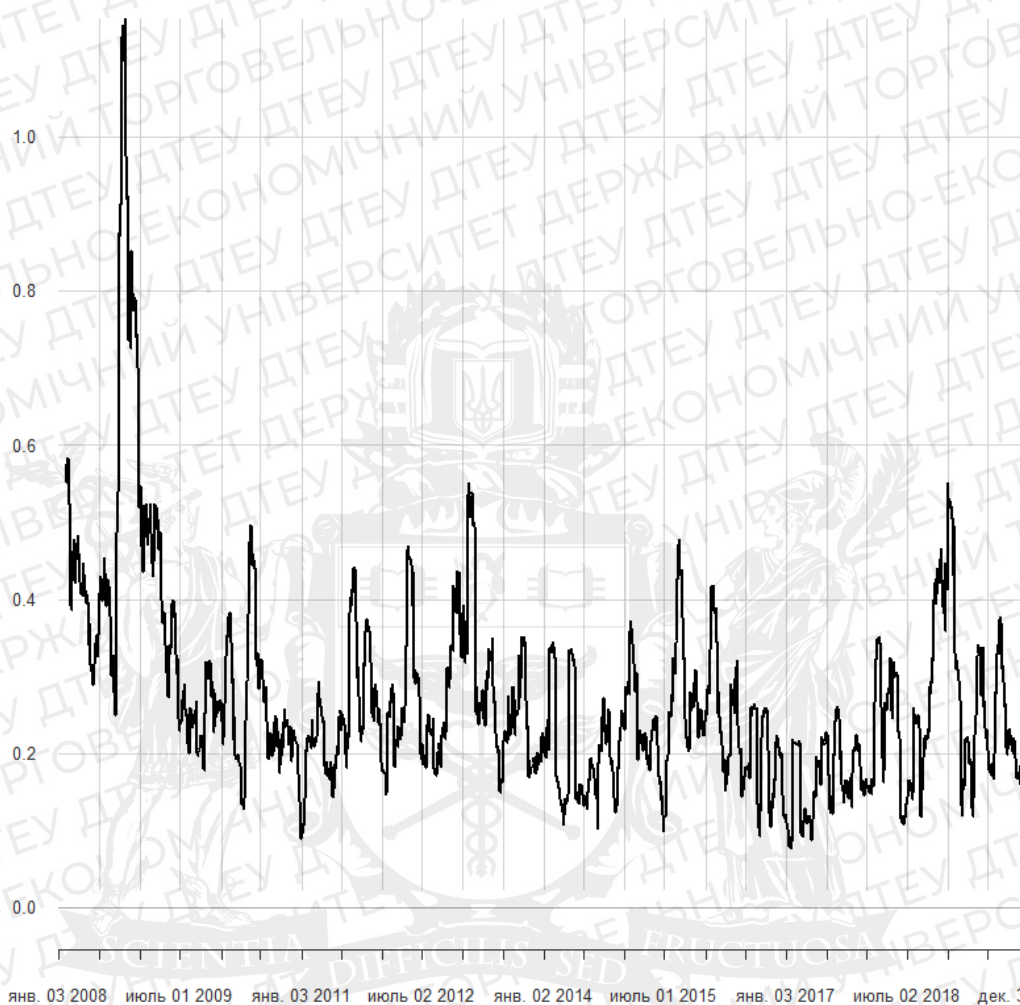


Рисунок 3.17 Графік місячна волатильності (мінливісті) компанії Apple

В випадку, якщо ми поставимо, там де ми вираховували помісячно 22 дні на 252, матимемо більш гладкий для ока графік (рис. 3.18).

Річна волатильність (мінливість) Apple

2008-01-03 / 2019-12-30



Рисунок 3.18 Графік річної волатильності (мінливісті) компанії Apple

Побудуємо код для ситуаційного випадку над яким ми будемо працювати з прогнозами на майбутні роки. Після приписання коду, скористаємося методом прогнозування на основі різних одновимірних моделей GARCH, а саме *ugarchforecast*, де дані рівні прибутку з 2008 року. Спрогнозуємо на один рік, тобто 252 днів. Після копіюємо строки та змінюємо для 2019 року. Зобразимо графік в дві колони, для того, щоб одразу бачити видозміни (рис. 3.19).

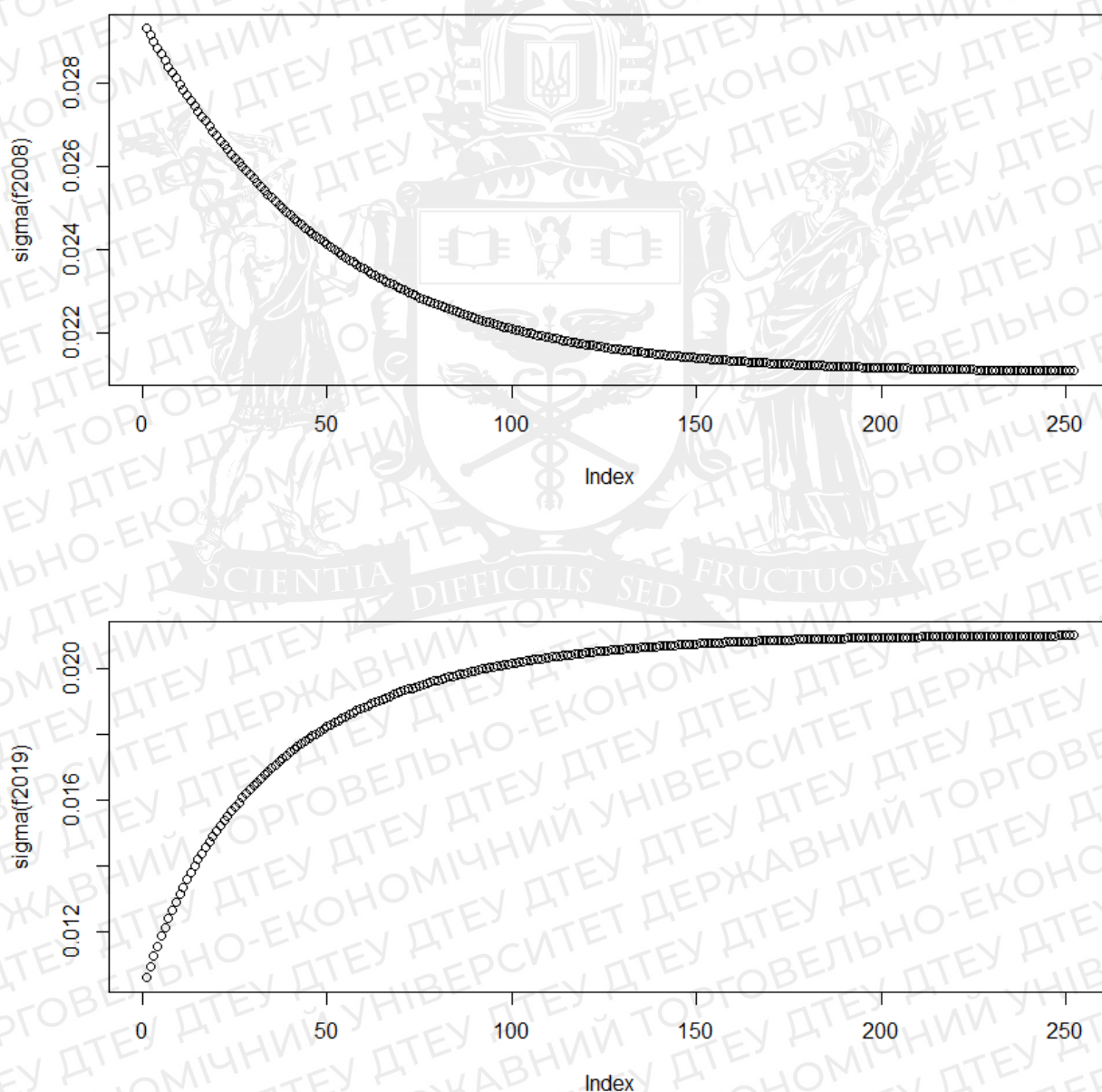


Рисунок 3.19 Приклади прогнозу на рік певних періодів часу

Відзначимо, що за 2008 рік ми маємо високу мінливість, яка трохи впала. В 2019 році вона була мала, але спрогнозувавши, бачимо що мінливість зросте у 2020 році.

На останок зробимо випадок, де ми введемо через функцію *ugarchpath*

кількість часових рядів прогнозованого прибутку, який ви хочете. В нашому випадку ми будемо робити три різних графік с прогнозам на рік продажів. Матимемо такі графіки, базовані на моделі, отримавши із встановленого значення об'єктів, що повертаються за допомогою функції моделювання *fitted* (мал 3.20).

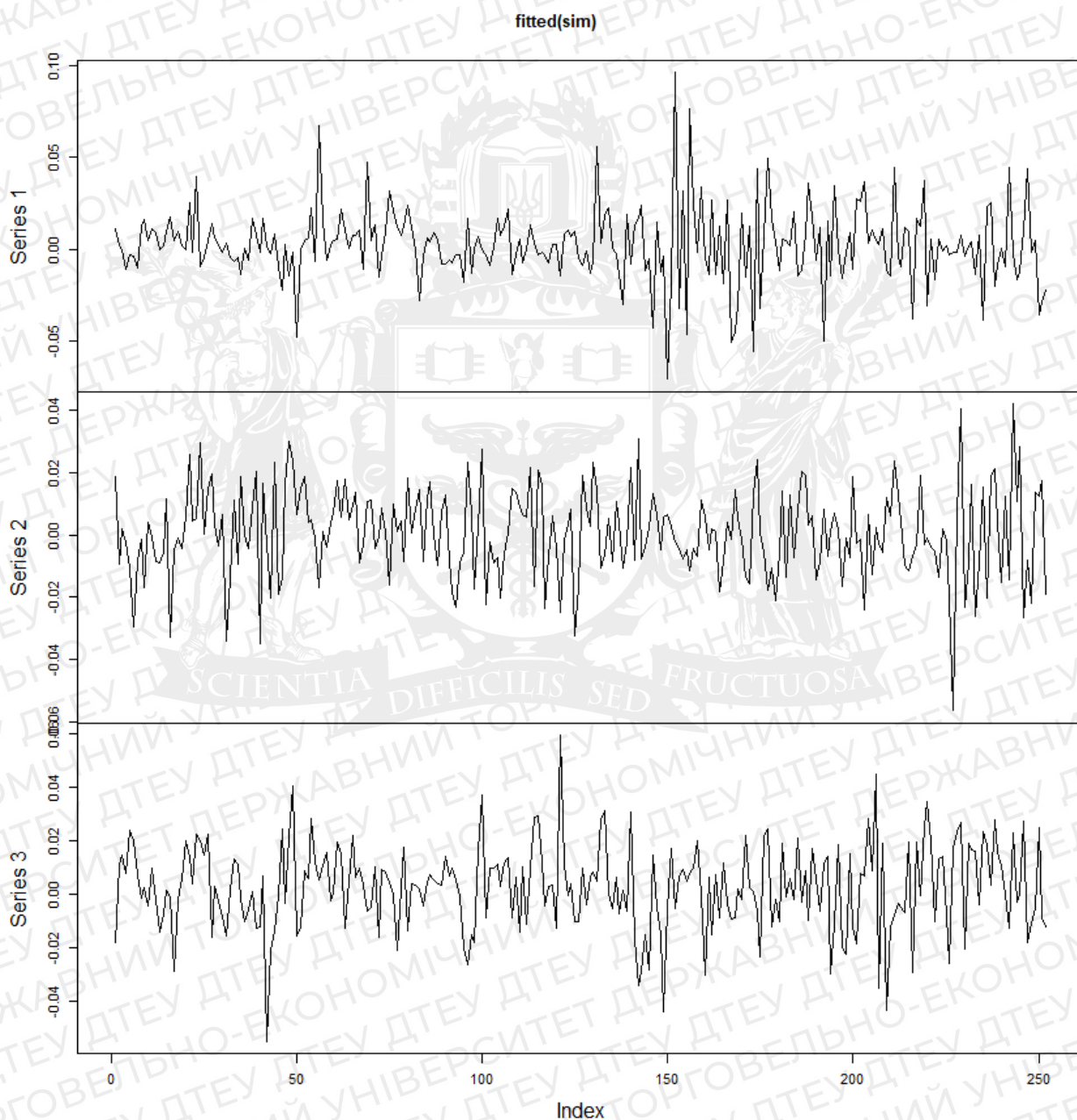


Рисунок 3.20 Часові ряди мінливості створенні за допомогою *fitted*

Використовуючи *sigma*, отримаємо стандартне відхилення похибок (рис. 3.21).

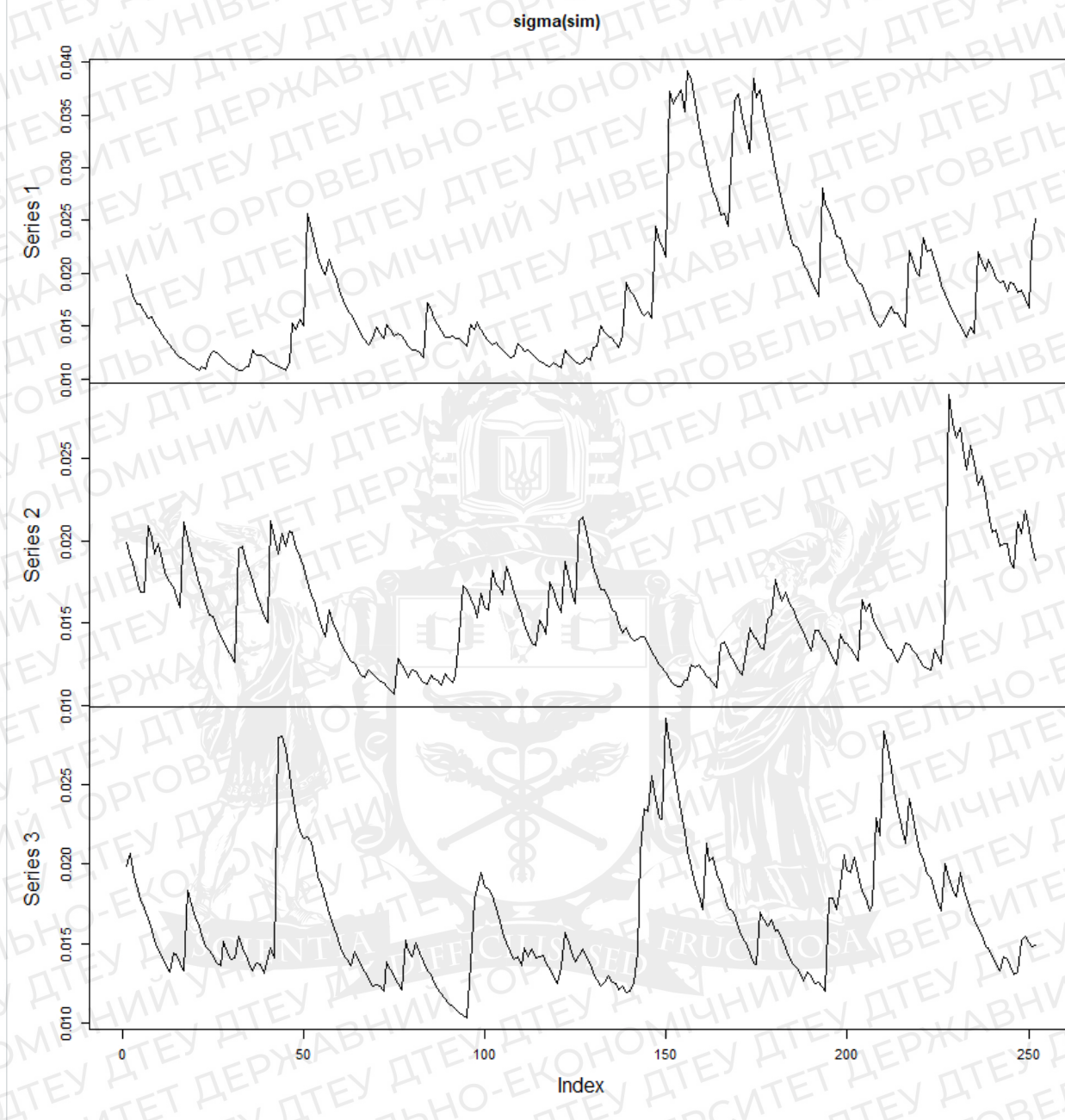


Рисунок 3.21 Часові ряди мінливості створенні за допомогою *sigma*

Щоб дізнатися, що ми будемо мати в кінці 2019 року, ми використовуємо значення ціни закриття угоди (Close). В нашому випадку це 78.88 на 30.12.2019. Отож, використовуємо це число як початкове значення, використовуючи *apply* функцію разом із *cumsum*, тобто функцію, що повертає вектор елементами якого є сумарні суми, мінімуми чи максимуми елементів аргументу. Чорна лінія показує найкращий результат продажів за рік, за стокову ціну відповідає червона лінія, а зелена – значення їх в середньому (рис. 3.22).

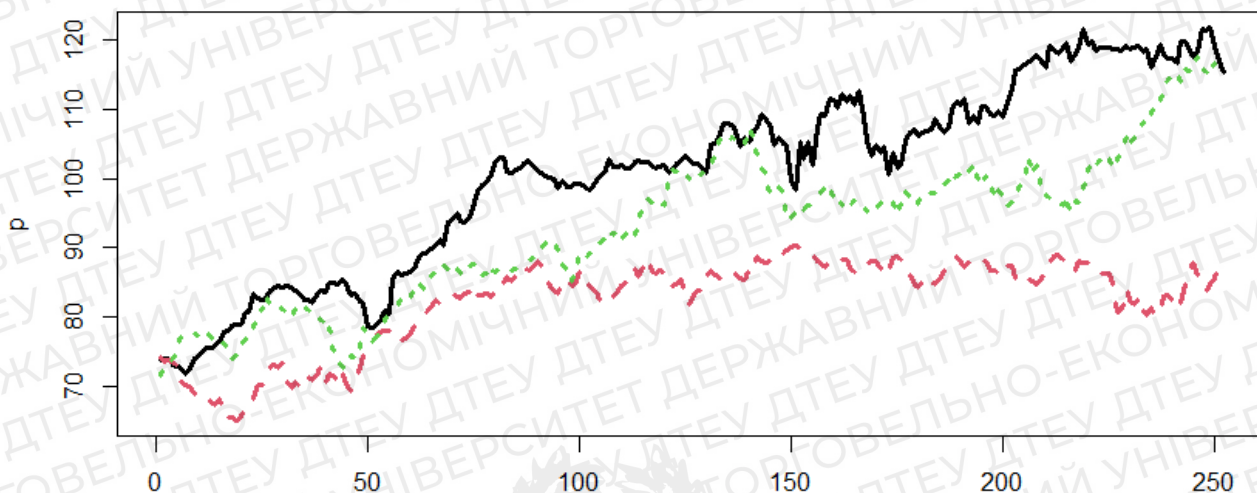


Рисунок 3.22 Завершений результат симуляції прогнозу

Висновки до розділу 3

В цьому розділі, було описано одну з мов для аналізу даних, а саме R, в програмі RStudio. Проаналізовано використання даної мови як з теоретичної, так і з практичної частини.

В теоретичній частині проговорено на основі чого було створено мову R, суть та методи її використання, історію становлення та популяризації, можливості та актуальність.

В практичній відповідно створено наглядний код, за допомогою якого можна було проаналізувати характеристику мінливості на ринку продажів, підвести статистику прибутку, симулювати прогноз на майбутній рік.

ВИСНОВКИ

Основуючись вище перерахованої інформації, успішно вдалося провести аналіз основної діяльності компанії Apple та інтерактивно відобразити загальні ключові показники в аналітичному звіті.

У теоретичній частині випускної кваліфікаційної роботи викладено основні завдання та етапи аналізу даних, задачі консолідації даних, їхні типи, методи і структури сховищ (репозиторіїв) даних та алгоритми передобробки даних.

Відповідно до поставлених завдань, розглянуто програмні засоби аналізування даних, побудови аналітичного звіту, в особистості виділено 3 найбільш поширених платформ для аналізу даних та опис їхніх можливостей у цілому.

Аналіз даних компанії Apple показав чіткий ріст фінансових і стратегічних показників з певні періоди досліджуваного часу, починаючи з 2010 року, закінчуючи 2020 роком. Маємо зазначити, що популярність бренду є на високому рівні. Про це свідчить нам збільшення економічних показників в сфері продажу та попит на продукцію компанії.

В процесі побудови аналітичного звіту, основууючись даними компанії, було опрацьовано основні можливості використання даних з аналітичною платформи Microsoft Power BI, а саме: трансформація та візуалізація; побудова моделі даних; оновлення даних. Між елементами візуалізації було використано для зображення показників інструменти візуалізації такі елементи як: кругові та стовпчасті діаграми, картки, матричні таблиці. Проаналізовано їх користь, простоту і комфортність у використанні.

Останнім пунктом в процесі створення аналітичного звіту є його публікація. Місце його публікації було обрано власну робочу область Power BI.

Окрім всього цього, було на практиці було для більш детального аналізу мову R. Задачею якого було ситуаційне моделювання та аналіз цін на акції компанії Apple за певний проміжок часу. Відображення результату візуально на платформі RStudio. Елементами візуалізації є таблиці з графіками, діаграмами, на яких зображено результати роботи.

Кінцевою інформацією стали графіки з даними, в яких було проведено

аналіз денного продажу, щоденного дохід, прогнозування на приклад створення ситуаційного випадку.



СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ

1. Задачі Data Mining та їх класифікація. Інформація та знання
https://moodle.znu.edu.ua/pluginfile.php?file=/486125/mod_resource/content/1/%d0%9b%d0%b5%d0%ba%d1%86%d1%96%d1%8f%205.pdf [Електронний ресурс]
2. Консолідація даних. Зведені таблиці.
<https://sites.google.com/site/lutskschool1yasenchuk/materiali-do-urokiv/10-klas/urok-11-1> [Електронний ресурс]
3. Сховище даних:
https://stud.com.ua/121123/informatika/shovischa_daniv [Електронний ресурс]
4. Процес побудови та використання моделі
<https://studfile.net/preview/5554364/page:78/> [Електронний ресурс]
5. Візуалізація даних: що це таке і для чого вона потрібна
<https://gurt.org.ua/articles/37609/> [Електронний ресурс]
6. Методи візуалізації даних у розподілених системах:
<https://ena.lpnu.ua:8443/server/api/core/bitstreams/4f41f803-40ed-4071-9cda-2bb33b3ae54d/content> [Електронний ресурс]
7. DAX-Урок 1. Що таке DAX. Основні поняття.
http://moonexcel.com.ua/uroki-dax1_ua [Електронний ресурс]
8. Power BI: https://uk.wikipedia.org/wiki/Power_BI [Електронний ресурс]
9. Deductor: https://kursoviks.com.ua/bd_kompyuternyye/article_post/868-laboratorna-robot-a-no-1-bazovi-printsipi-roboti-v-analitichniy-platformi-deductor#:~:text=Deductor%20E2%80%93%20D1%86%D0%B5%20D0%B0%D0%BD%D0%B0%BB%D1%96%D1%82%D0%B8%D1%87%D0%BD%0%B0%20D0%BF%D0%BB%D0%B0%D1%82%D1%84%D0%BE%D1%80%D0%BC%D0%B0%2C%20D0%BF%D1%80%D0%B8%D0%B7%D0%BD%D0%B0%D1%87%D0%B5%D0%BD%D0%B0,%D1%80%D1%96%D1%88%D0%B5%D0%BD%D1%8C%20D0%B2%20D0%BE%D0%B1%D0%BB%D0%B0%D1%81%D1%82%D1%96%20D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D0%B7%D1%83%20D0%B4%D0%B0%D0%BD%D0%B8%D1%85
 [Електронний ресурс]
10. Loginum: <https://loginom.com/platform/extract-transform-load>

[Електронний ресурс]

11. Типи схем у моделюванні сховища даних - схема Star & Snowflake:

<https://uk.myservername.com/schema-types-data-warehouse-modeling-star-snowflake-schema> [Електронний ресурс]

12. Опис кластерної діаграми:

<https://uk.mcfairbanks.com/1033-clustered-column-chart-in-excel>

[Електронний ресурс]

13. Last close price: [https://www.stockopedia.com/ratios/last-close-price-quotes-](https://www.stockopedia.com/ratios/last-close-price-quotes-5025/)

[5025/](https://www.stockopedia.com/ratios/last-close-price-quotes-5025/) [Електронний ресурс]

14. Матрична таблиця: [https://www.65bit.com/docs/what-is-a-matrix-](https://www.65bit.com/docs/what-is-a-matrix-table/#:~:text=A%20matrix%20table%20is%20a,generally%20'look%20up'%20table)

[table/#:~:text=A%20matrix%20table%20is%20a,generally%20'look%20up'%20table](https://www.65bit.com/docs/what-is-a-matrix-table/#:~:text=A%20matrix%20table%20is%20a,generally%20'look%20up'%20table)
[s.](https://www.65bit.com/docs/what-is-a-matrix-table/#:~:text=A%20matrix%20table%20is%20a,generally%20'look%20up'%20table) [Електронний ресурс]

15. Інтерактивна карта: [https://www.versantus.co.uk/blog/how-interactive-](https://www.versantus.co.uk/blog/how-interactive-maps-can-improve-user-experience-your-site#:~:text=Interactive%20maps%20are%20web%2Dbased,to%20showcase%20multiple%20data%20sets.)

[maps-can-improve-user-experience-your-site#:~:text=Interactive%20maps%20are%20web%2Dbased,to%20showcase%20multiple%20data%20sets.](https://www.versantus.co.uk/blog/how-interactive-maps-can-improve-user-experience-your-site#:~:text=Interactive%20maps%20are%20web%2Dbased,to%20showcase%20multiple%20data%20sets.) [Електронний ресурс]

16. Картка: [https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-](https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-card?tabs=powerbi-service)

[visualization-card?tabs=powerbi-service](https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-card?tabs=powerbi-service) [Електронний ресурс]

17. Діаграма з областями: [https://learn.microsoft.com/en-us/power-](https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-basic-area-chart)

[bi/visuals/power-bi-visualization-basic-area-chart](https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-basic-area-chart) [Електронний ресурс]

18. Лекція 1. Знайомство з мовою програмування R:

[https://rstudio-pubs-](https://rstudio-pubs-static.s3.amazonaws.com/378130_5600736fb2734e01bf109c83e6d83676.html)
[static.s3.amazonaws.com/378130_5600736fb2734e01bf109c83e6d83676.html](https://rstudio-pubs-static.s3.amazonaws.com/378130_5600736fb2734e01bf109c83e6d83676.html)

[Електронний ресурс]

19. Лекція 2. Робота з мовою R:

[https://rstudio-pubs-](https://rstudio-pubs-static.s3.amazonaws.com/378105_599bcd2892bf46498a6371290149267d.html)
[static.s3.amazonaws.com/378105_599bcd2892bf46498a6371290149267d.html](https://rstudio-pubs-static.s3.amazonaws.com/378105_599bcd2892bf46498a6371290149267d.html)

[Електронний ресурс]

20. Лекція 3. Типи та структури даних

[https://rstudio-pubs-](https://rstudio-pubs-static.s3.amazonaws.com/378106_3b96b5f50ef54c2bb90c9186ffae1dc0.html)
[static.s3.amazonaws.com/378106_3b96b5f50ef54c2bb90c9186ffae1dc0.html](https://rstudio-pubs-static.s3.amazonaws.com/378106_3b96b5f50ef54c2bb90c9186ffae1dc0.html)

[Електронний ресурс]

Додаток А

Код програмування для RStudio

```
# Бібліотеки
library(quantmod)
library(xts)
library(PerformanceAnalytics)
library(rugarch)

# Днівні продажі Apple
getSymbols("AAPL",
           from = "2008-01-01",
           to = "2019-12-31")
View(AAPL)
chartSeries(AAPL)

# Щоденний прибуток
return <- CalculateReturns(AAPL$AAPL.Close)
View(return)
return <- return[-1]
hist(return)
chart.Histogram(return,
                 methods = c('add.density', 'add.normal'),
                 colorset = c('blue', 'green', 'red'))
chartSeries(return, theme = 'white')

# Аналіз мінливості показників
chart.RollingPerformance(R = return["2008::2019"],
                         width = 252,
                         FUN = "sd.annualized",
                         scale = 252,
                         main = "Річна волатильність (мінливість) Apple")
```

```
# Симуляція ситуації, прогноз
```

```
s <- ugarchspec(mean.model = list(armaOrder = c(0,0)),  
                variance.model = list(model = "gjrGARCH"),  
                distribution.model = 'sstd')
```

```
m <- ugarchfit(data = return, spec = s)
```

```
sfinal <- s
```

```
setfixed(sfinal) <- as.list(coef(m))
```

```
f2008 <- ugarchforecast(data = return["/2008-12"],
```

```
                    fitORspec = sfinal,
```

```
                    n.ahead = 252)
```

```
f2019 <- ugarchforecast(data = return["/2019-12"],
```

```
                    fitORspec = sfinal,
```

```
                    n.ahead = 252)
```

```
par(mfrow = c(2,1))
```

```
plot(sigma(f2008))
```

```
plot(sigma(f2019))
```

```
sim <- ugarchpath(spec = sfinal,
```

```
                m.sim = 3,
```

```
                n.sim = 1*252,
```

```
                rseed = 123)
```

```
plot.zoo(fitted(sim))
```

```
plot.zoo(sigma(sim))
```

```
tail(AAPL)
```

```
p <- 72.8800*apply(fitted(sim), 2, 'cumsum') + 72.8800
```

```
matplot(p, type = "l", lwd = 3)
```