**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**

**“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**

**ІМЕНІ ІГОРЯ СІКОРСЬКОГО”**

Факультет прикладної математики

Кафедра програмного забезпечення комп’ютерних систем

**КУРСОВА РОБОТА**

з дисципліни “Бази даних”

спеціальність 121 – Програмна інженерія

на тему: Система аналізу цін споживчих товарів

|  |  |  |
| --- | --- | --- |
| **Студентка групи** КП-93 | **Коротюк Марія Ігорівна** | \_\_\_\_\_\_\_\_\_  (підпис) |
| **Викладач**  **к.т.н, доцент кафедри СПіСКС** | **Петрашенко А.В.** | \_\_\_\_\_\_\_\_\_  (підпис) |

Захищено з оцінкою \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Київ – 2020

**АНОТАЦІЯ**

Виконуючи курсову роботу ми створили програму для зберігання та аналізу поточних цін на товари у магазинах косметики та парфумерії за певною категорією. Дана система призначена для використання у сфері торгівлі та маркетингу.

Результатом роботи є можливість визначення цінових трендів, через порівняння середніх вартостей товарів однієї категорії у різних магазинах, також можна отримати перелік товарів, які є популярними серед клієнтів і у майбутньому користуватимуться ще більшим попитом, а також перелік товарів, які потребують додаткової реклами, адже мають гарні відгуки, але малу кількість продажів.

**ЗМІСТ**

Вступ………………………………………………………………………………4

1. Аналіз інструментарію для виконання курсової роботи…………………...5
2. Структура бази даних………………………………………………………...6
3. Опис програмного забезпечення…………………………………………….7

3.1 Загальна структура програмного забезпечення………………………...7

3.2 Опис модулів програмного забезпечення………………………………7

3.3 Опис основних алгоритмів роботи……………………………………...7

1. Аналіз функціонування засобів реплікації………………………………...13
2. Аналіз функціонування засобів резервування/відновлення……………...13
3. Аналіз результатів підвищення швидкодії запитів……………………….13
4. Опис результатів аналізу предметної галузі……………………………....16

Висновки………………………………………………………………………...17

Література……………………………………………………………………….18

Додаток 1………………………………………………………………………..19

Додаток 2………………………………………………………………………..20

Додаток 3………………………………………………………………………..21

Додаток 4………………………………………………………………………..22

Додаток 5………………………………………………………………………..23

Додаток 6………………………………………………………………………..24

**ВСТУП**

Метою розробки є здобуття навичок використання СУБД та створення програмного забезпечення, яке буде взаємодіяти з реляційною базою даних, яка містить велику кількість даних, а також подальший аналіз цих даних і оформлення результатів у вигляді діаграм, графіків та таблиць.

Зокрема база даних містить актуальну інформацію про товари у магазинах косметики та парфумерії, таких як Prostor, Watsons, Brocard та Kosmo, дані про покупки та відгуки клієнтів, що дозволяє проаналізувати популярність товарів, та різницю середніх цін за категоріями у цих магазинах.

Кожен товар у базі даних належить до певної категорії, зокрема косметика для обличчя, очей та брів, губ та парфумерія. Оскільки більшість товарів у магазинах схожа та належить до однакових виробників, то ми аналізуємо середні ціни в межах певної категорії і визначаємо різницю у відсотках.

Дана робота є актуальною, адже дозволяє визначити цінові тренди, порівнюючи ціни у різних магазинах, і тому може використовуватись у сферах торгівлі та маркетингу, які зараз активно розвиваються.

Для кращого представлення більшість результатів аналізу також додатково зображуються як графіки та таблиці excel.

**1. Аналіз інструментарію для виконання курсової роботи**

Для виконання курсової роботи було обрано СУБД PostgreSQL, тому що ця СУБД підтримує бази даних необмеженого розміру, має потужні і надійні механізми транзакцій і реплікації, може бути встановлена на будь-якій операційній системі [1].

Також система є розширювана, оскільки можна створювати власні типи даних та індекси, а також розширювати поведінку за допомогою мов програмування [2].

Курсова робота написана мовою програмування python. Для генерації частини даних було використано фреймворк Scrapy, для аналізу даних бібліотека Data Science – pandas, для побудови графіків – бібліотека Matplotlib. Для взаємодії із базою даних за допомогою мови програмування python ми використали модуль psycopg2 та бібліотеку SQLAlchemy.

Scrapy – фреймворк для обходу сайтів та вилучення з них структурованих даних, які можуть використовуватися для широкого кола корисних застосувань: видобуток та обробка інформації чи історичне архівування [3].

pandas – програмна бібліотека, написана для мови програмування Python для маніпулювання даними та їхнього аналізу. Вона, зокрема, пропонує структури даних та операції для маніпулювання чисельними таблицями та часовими рядами [4].

Matplotlib – бібліотека на мові програмування Python для візуалізації даних двовимірною 2D графікою (3D графіка також підтримується). Отримувані зображення можуть бути використані як ілюстрації в публікаціях [5].

psycopg2 – найпопулярніший адаптер бази даних PostgreSQL, який є легким та ефективним у використанні [6].

SQLAlchemy - інструментарій SQL та об'єктно-реляційне відображення для мови програмування Python випущене під ліцензією MIT [7].

**2. Структура бази даних**

База даних містить наступні сутності (див. додаток 1):

1. store – сутність, яка містить дані про магазини, які будуть аналізуватись у роботі.
2. category – сутність, яка містить дані про категорії магазинів.
3. product – сутність, яка містить дані про товари, що належать до певного магазину та певної категорії.
4. purchase – сутність, яка містить дані про дату та час здійснення кожної покупки.
5. customer – сутність, яка містить дані користувача, який може створювати відгуки на товари.
6. review – сутність, яка містить дані про відгуки, створені певними користувачами на певні товари.

Між усіма таблицями зв’язки – це один до багатьох.

**3. Опис програмного забезпечення**

**3.1 Загальна структура програмного забезпечення**

Розроблене програмне забезпечення складається з двох частин: перша – програма для отримування даних з сайтів магазинів, друга – програма з консольним інтерфейсом для взаємодії з базою даних, проведення аналізу даних, здійснення пошуку, а також додатково було створено CRUD операції, що дозволяють додавати, видаляти і оновлювати інформацію у більшості таблиць.

**3.2 Опис модулів програмного забезпечення**

Програмне забезпечення розроблене відповідно до структури MVC. У проекті є консольне меню, яке дозволяє обрати користувачу необхідну операцію. Програмне забезпечення містить один контролер, одне подання, окремі модулі для кожного класу, який відповідає певній сутності бази даних, зокрема store, category, product, purchase, customer, review. Також програма містить моделі, які відповідають сутностям (store\_model, category\_model, product\_model, purchase\_model, customer\_model, review\_model) та моделі для аналізу та пошуку даних (comparison\_model i search\_model). Моделі безпосередньо взаємодіють із базою даних і дозволяють отримувати програмі доступ до потрібної інформації.

**3.3 Опис основних алгоритмів роботи.**

Дані, що не генеруються за допомогою фреймворка Scrapy, генеруються спеціальними SQL запитами за допомогою моделей сутностей. Імена для таблиці customer створюються з використанням бібліотеки names.

CRUD операції виконуються з використанням бібліотеки SQLAlchemy, є валідація даних, які вводить користувач. Також перед додаванням даних у таблицю product здійснюється перевірка чи точно такого ж рядка не існує ще у таблиці. Для цього був написаний тригер (Рис. 1):

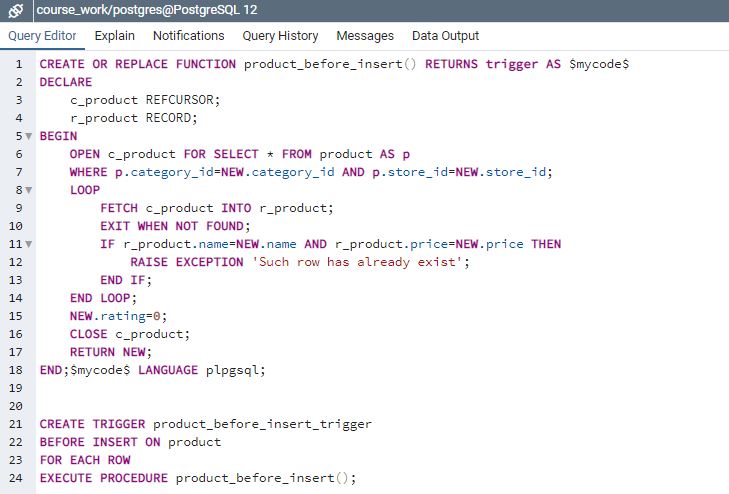


Рис. 1 product\_before\_insert\_trigger

Також були написані тригери, які перевіряють чи дати покупок, які мають бути додані у таблицю purchase при додаванні чи редагуванні, є не ранішими, ніж 01.01.2019 та не пізнішими, ніж поточна дата (Рис. 2):

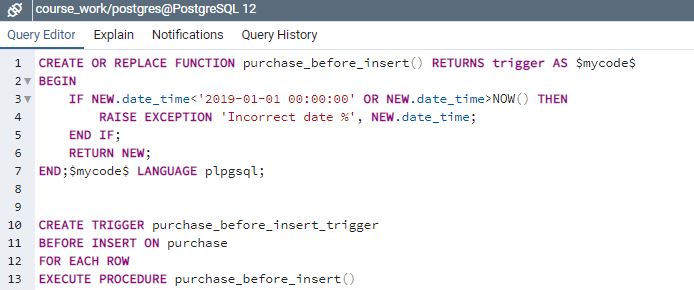


Рис. 2 а) purchase\_before\_insert\_trigger

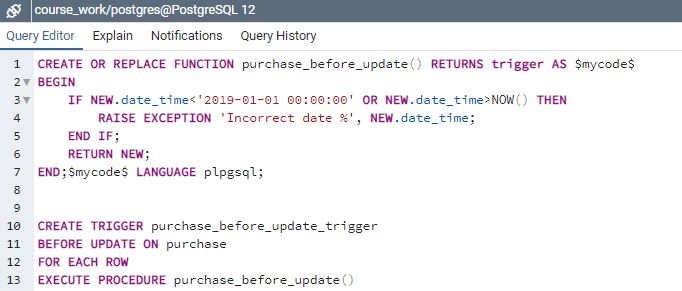


Рис. 2 б) purchase\_after\_insert\_trigger

Було створено тригер, який після додавання, редагування чи видалення запису у таблиці review обраховує і перезаписує середнє значення у полі rating таблиці product та перед додавнням, редагуванням у таблиці review здійснюється перевірка чи значення rating знаходиться в межах від 1 до 5 (Рис. 3):

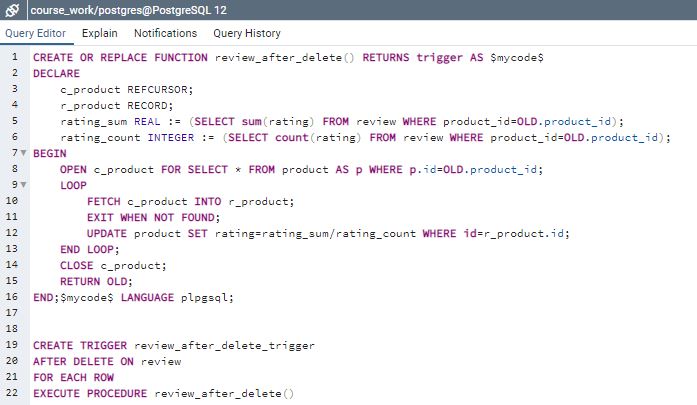


Рис. 3 а) review\_after\_delete\_trigger

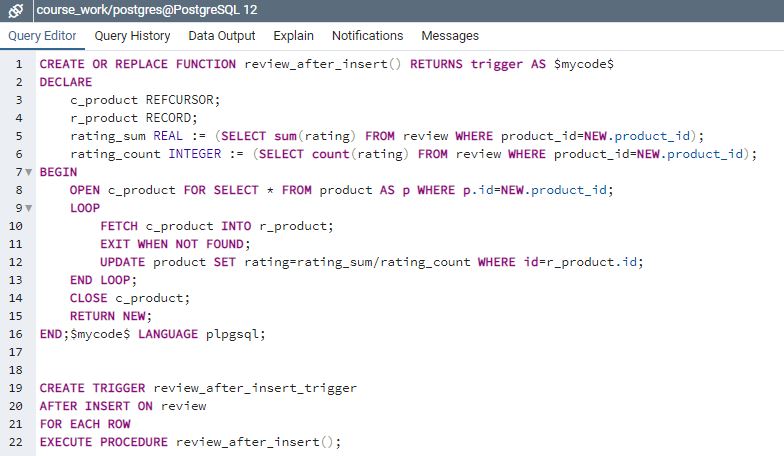
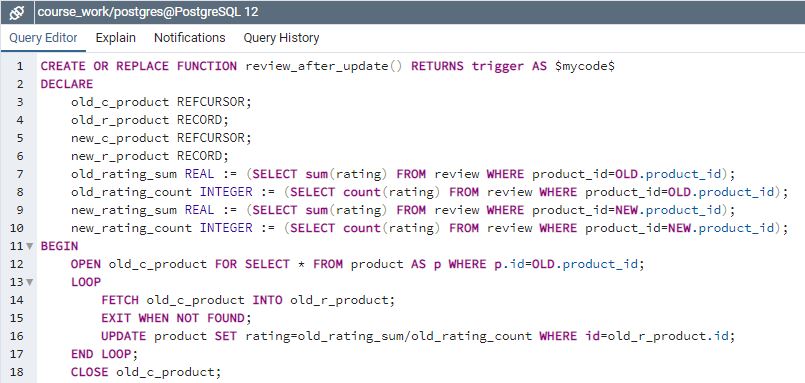


Рис. 3 б) review\_after\_insert\_trigger



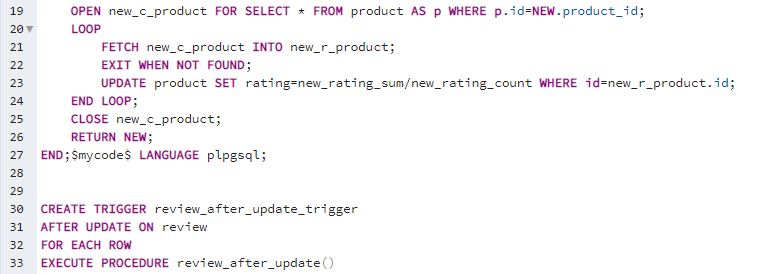


Рис. 3 в) review\_after\_update\_trigger

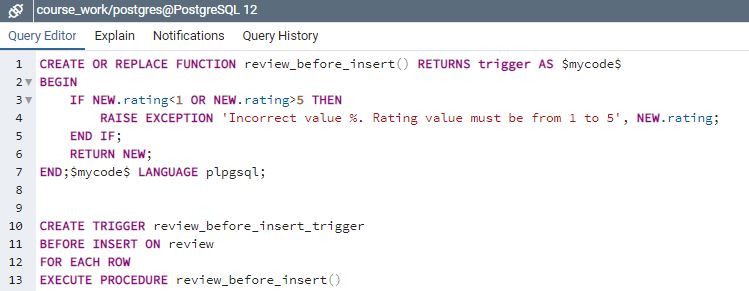


Рис. 3 г) review\_before\_insert\_trigger

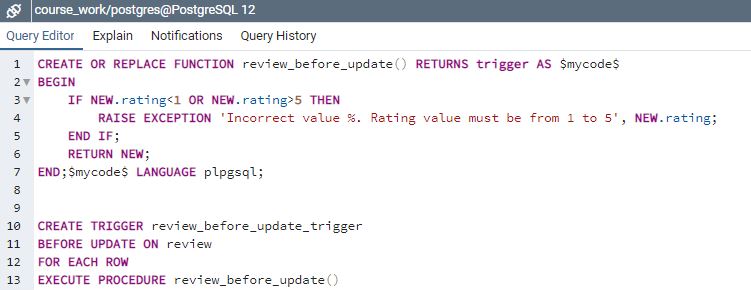


Рис. 3 д) review\_before\_update\_trigger

Порівняння проданих продуктів відбувається наступним чином: визначається кількість проданих товарів у кожній категорії для кожного магазину і ці результати записуються у таблицю за допомогою бібліотеки pandas. Далі створюються графіки чи гістограми та записуються отримані дані у Excel фай .Для визначення різниці цін спочатку визначаються 2 магазини, для яких буде проводитись аналіз, далі знаходяться середні ціни за категорією у відповідному магазині та записуються у таблицю, створюються графіки та здійснюється запис у Excel файл. Потім вираховується різниця у відсотках і дана інформація виводиться для користувача у консолі.

Для прогнозування популярності товарів потрібно визначити які товари купують частіше, які мають кількість відгуків більше середнього та їх рейтинг є не нижчим 4. Саме це і товари є найпопулярнішими і клієнти задоволені їх якістю, отже попит на них буде продовжувати рости.

Для визначення переліку товарів, які потребують додаткової реклами, потрібно знайти товари які купують рідко і вони мають мало відгуків проте їх рейтинг є не нижчим 4.2, що свідчить про те, що клієнти, які їх купували ними задоволені. Отже, такі товари потрібно рекламувати, щоб збільшити аудиторію їх споживачів і зробити їх популярнішими.

У програмі також є можливість пошуку записів у таблиці product за рейтингом у визначеному користувачем діапазоні, та пошук у таблиці purchase за датою – також у визначеному користувачем діапазоні,

**4. Аналіз функціонування засобів реплікації**

При виконанні курсової роботи реплікація бази даних не була реалізована.

**5. Аналіз функціонування засобів резервування/відновлення**

При виконанні курсової роботи резервування/відновлення бази даних не були реалізовані.

**6. Аналіз результатів підвищення швидкодії запитів**

Для запитів, які використовуються для пошуку, були використані індекси для підвищення їх швидкодії. В обох випадках були застосовані індекси BTree, адже пошук здійснюється у заданому діапазоні з використанням знаків < > в умові (Рис. 4):

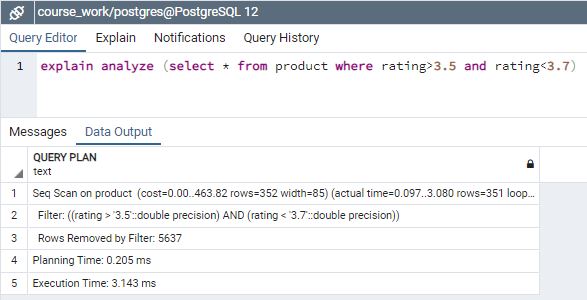


Рис. 4 а) Приклад запиту до створення індекса для таблиці product

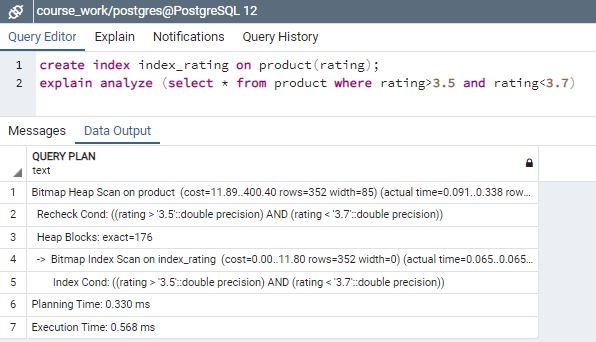


Рис. 4 б) Приклад запиту після створення індекса для таблиці product

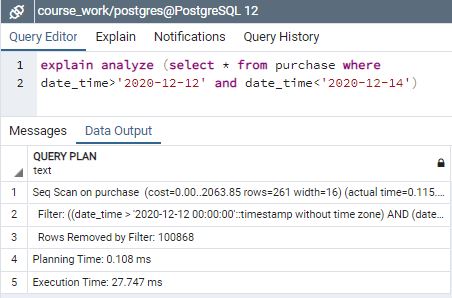


Рис. 4 в) Приклад запиту до створення індекса для таблиці purchase

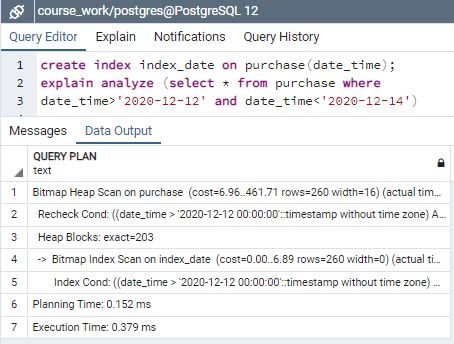


Рис. 4 г) Приклад запиту після створення індекса для таблиці purchase

Зокрема швидкість пошуку у таблиці product збільшилась у 3 рази з використанням індексів, для цього ми порівняли швидкість виконання до створення і після створення 10 разів (Рис 5 (а)), а у таблиці purchase відбулось пришвидшення приблизно у 50 разів, порівняння також проводилось 10 разів (Рис 5 (б)). Гістограми порівняння швидкодії запитів наведені у Додатку 2.

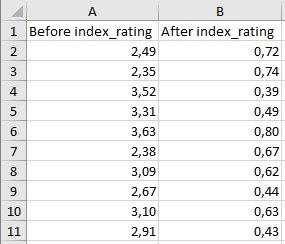


Рис. 5 а) Порівняння швидкодії запиту для пошуку у таблиці product

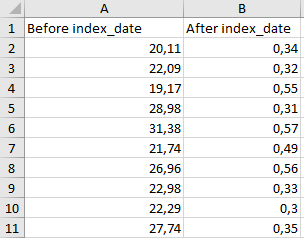


Рис. 5 б) Порівняння швидкодії запиту для пошуку у таблиці purchase

**7. Опис результатів аналізу предметної галузі**

При порівнянні кількості проданих товарів за категоріями у різних магазинах за 12.12.2020 отримуємо, що у трьох з чотирьох категорій найбільша кількість проданих товарів належить магазину Brocard. Найбільша кількість проданих товарів у цей день – 18, а найменша – 2. Також можемо порівняти кількості продажів за місяць і отримати їх графік (Додаток 3).

При порівнянні середніх цін на товари у магазинах Brocard і Watsons отримуємо, що в Brocard ціни в рази вищі, особливо різняться ціни на парфумерію (Додаток 4).

У Додатках 5 та 6 наведені гістограми, які характеризують рейтинги популярних та не популярних товарів та їх кількість продажів та відгуків за категоріями.

**Висновки**

Отже, у курсовій роботі ми створили програмне забезпечення, яке дозволяє отримувати актуальні дані з інтернет-магазинів Prostor, Watsons, Brocard та Kosmo, записувати ці дані у реляційну базу даних PostgreSQL, проводити аналіз цих даних за категоріями із створенням таблиць, графіків та гістограм, та отримувати перелік популярних товарів і непопулярних, які потребують додаткової реклами. Тому дане програмне забезпечення може використовуватись у сфері торгівлі та маркетингу.

Для написання програми ми використали фреймворк Scrapy, модуль psycopg2, бібліотеки pandas, matplotlib та SQLAlchemy.

Також програма дозволяє генерувати дані за допомогою SQL запитів і здійснювати пошук, який виконується швидше завдяки використанню індексів.

**Література**

1. What is PostgreSQL? [Електронний ресурс] – Режим доступу до ресурсу: <https://www.educba.com/what-is-postgresql/>.
2. Chapter 11. Indexes [Електронний ресурс] – Режим доступу до ресурсу: <https://www.postgresql.org/docs/9.1/indexes-intro.html>.
3. Створення веб павука з Python та Scrapy [Електронний ресурс] – Режим доступу до ресурсу: <https://codeguida.com/post/254/>.
4. pandas [Електронний ресурс] – Режим доступу до ресурсу: <https://uk.wikipedia.org/wiki/Pandas>.
5. matplotlib [Електронний ресурс] – Режим доступу до ресурсу: <https://uk.wikipedia.org/wiki/Matplotlib>.
6. Доступ к базе данных PostgreSQL с помощью psycopg2 [Електронний ресурс] – Режим доступу до ресурсу: <https://riptutorial.com/ru/python/example/18257/доступ-к-базе-данных-postgresql-с-помощью-psycopg2>.
7. SQLAlchemy [Електронний ресурс] – Режим доступу до ресурсу: <https://uk.wikipedia.org/wiki/SQLAlchemy>.

ДОДАТОК 1

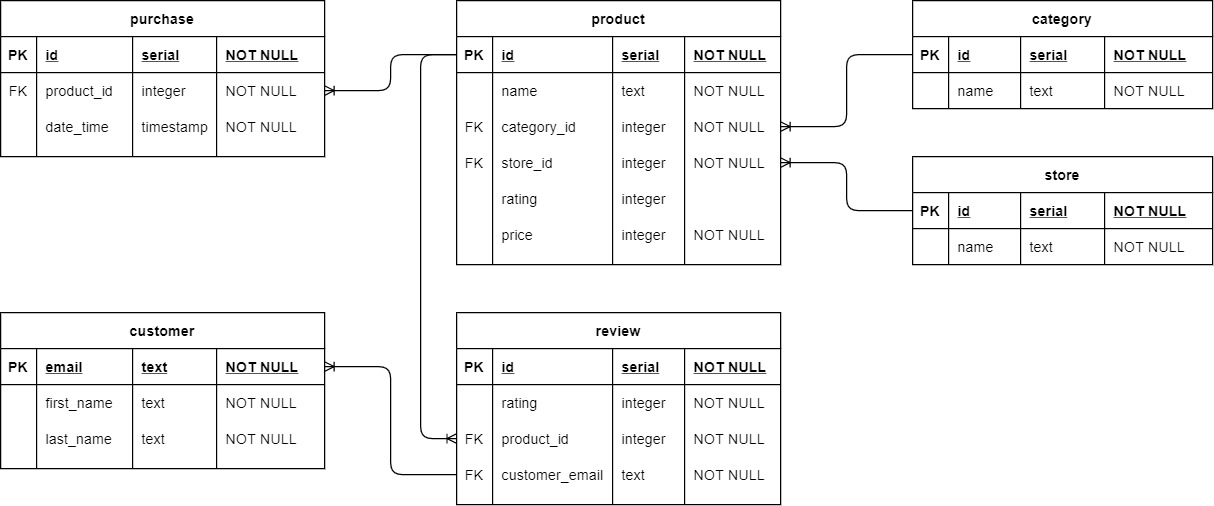
****

Рис. Структура бази даних

ДОДАТОК 2

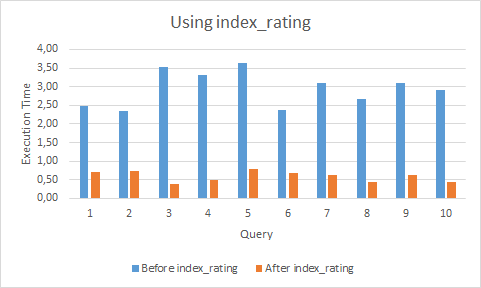


Рис. Порівняння швидкодії запитів для таблиці product

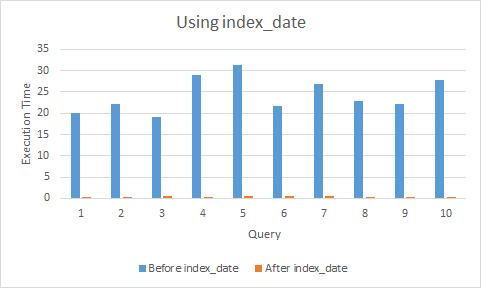


Рис. Порівняння швидкодії запитів для таблиці purchase

ДОДАТОК 3

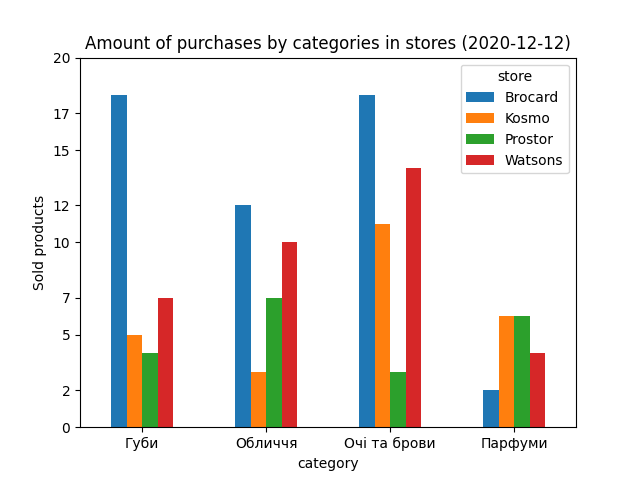


Рис. Приклад порівняння кількості покупок за категоріями у різних магазинах за певний день

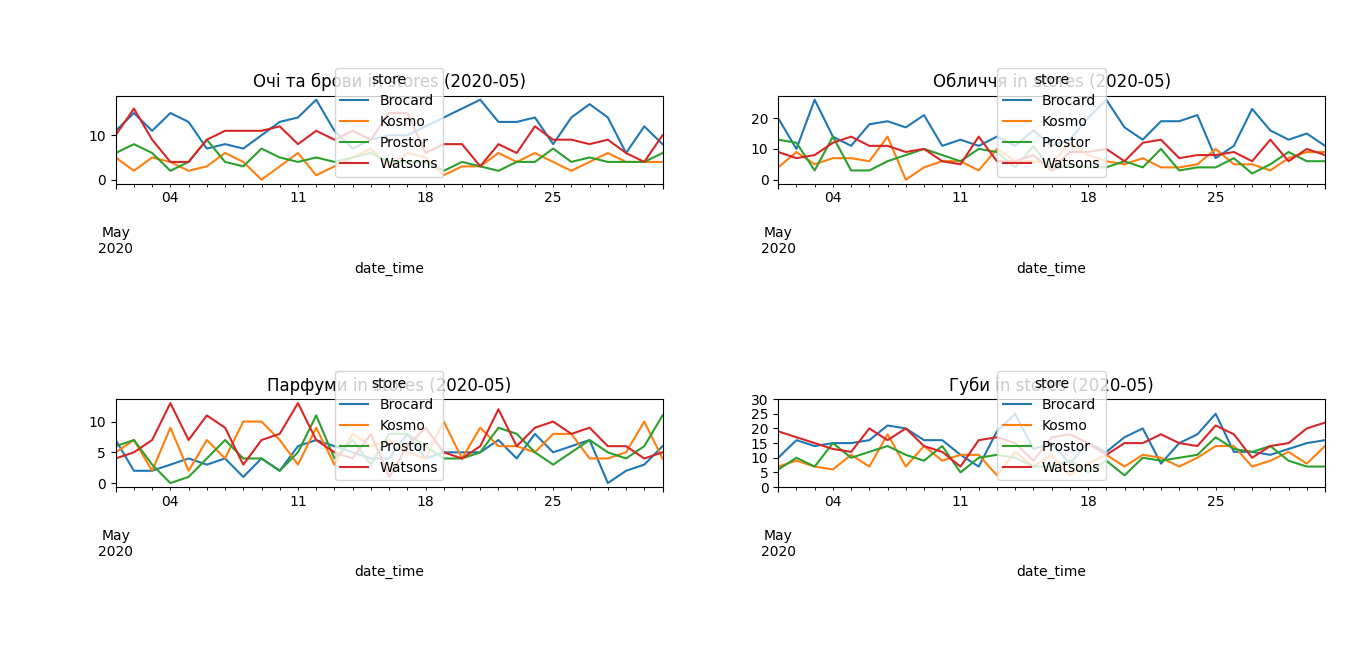


Рис. Приклад порівняння кількості покупок за категоріями у різних магазинах за певний місяць

ДОДАТОК 4

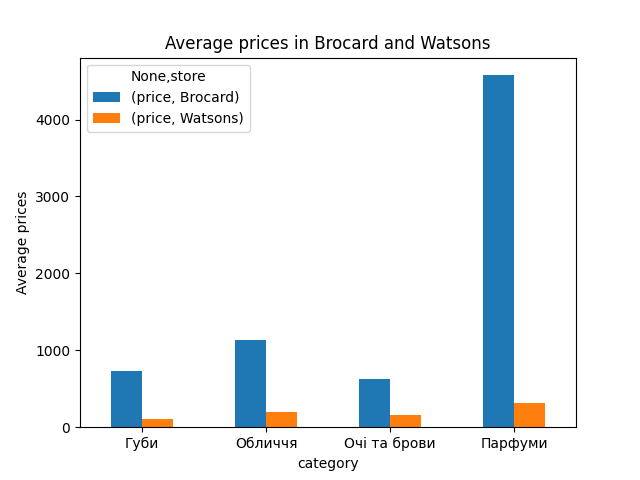


Рис. Приклад порівняння середніх цін за категоріями у двох магазинах

ДОДАТОК 5

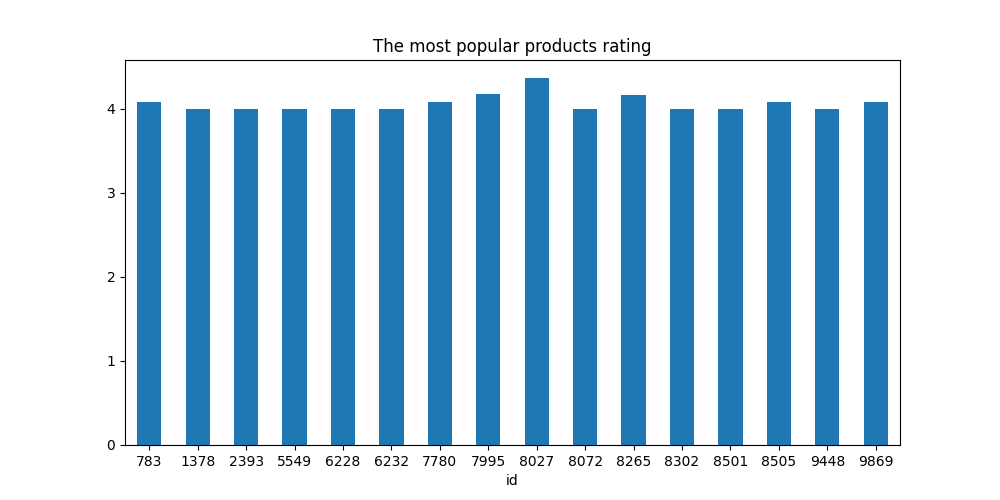


Рис. Приклад порівняння популярних товарів за рейтингом

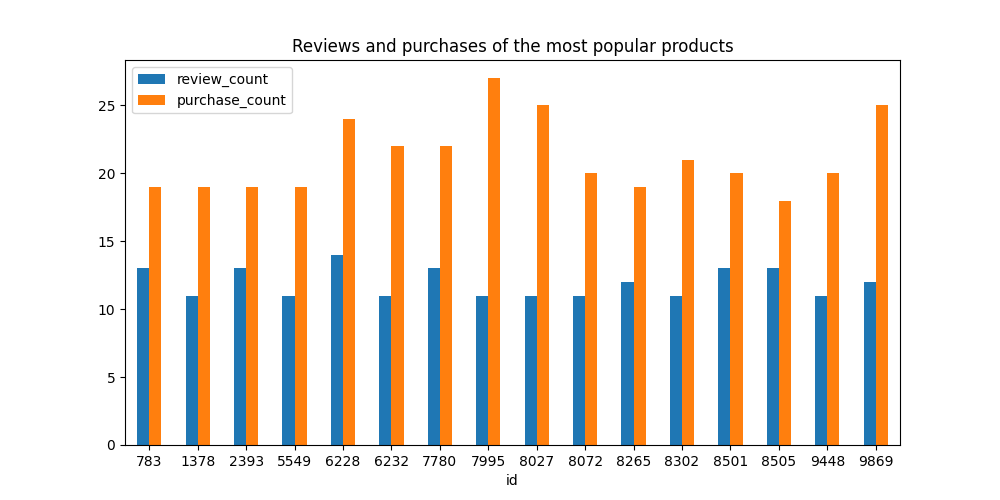


Рис. Приклад порівняння популярних товарів за кількістю покупок та відгуків

ДОДАТОК 6

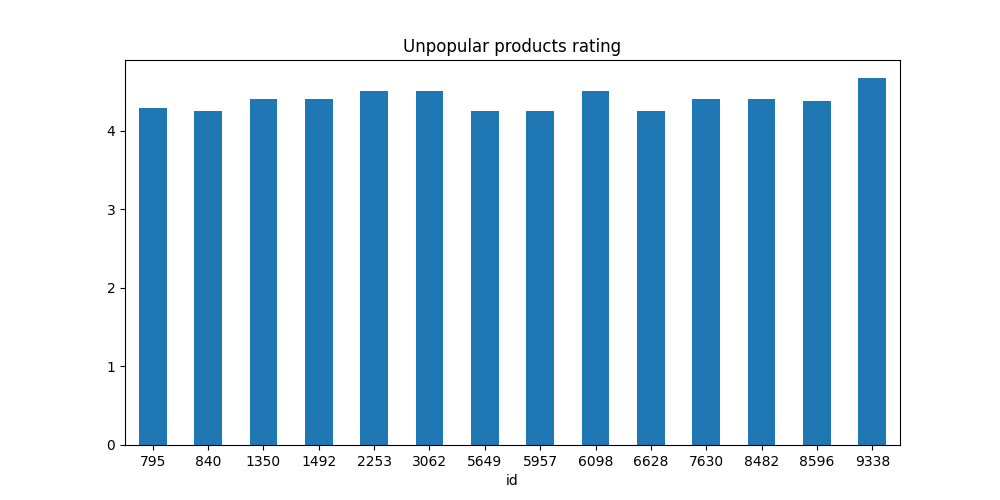


Рис. Приклад порівняння непопулярних товарів за рейтингом

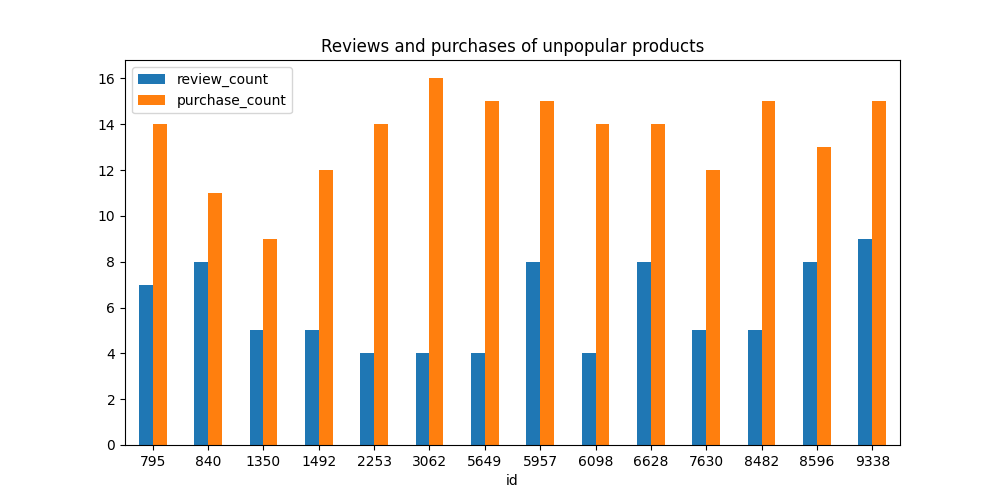


Рис. Приклад порівняння непопулярних товарів за кількістю покупок та відгуків