

Міністерство освіти і науки України

Луцький національний технічний університет

(повне найменування закладу вищої освіти)

Факультет комп'ютерних та інформаційних технологій

(повне найменування факультету)

Кафедра комп'ютерної інженерії та кібербезпеки

(повне найменування кафедри)

**КВАЛІФІКАЦІЙНА РОБОТА
ЗА СТУПЕНЕМ ВИЩОЇ ОСВІТИ «БАКАЛАВР»**

МОНІТОРИНГ ПЛАТФОРМИ YOUTUBE ЗАСОБАМИ SQL

YOUTUBE PLATFORM MONITORING USING SQL

спеціальність 123 Комп'ютерна інженерія

(шифр і назва спеціальності)

освітня програма Комп'ютерна інженерія

(назва освітньої програми)

Виконав: здобувач вищої освіти

групи КІс-21

Савчук Ольга Володимирівна

(підпис)

Керівник:

к.т.н., доцент

Лавренчук Світлана Василівна

(підпис)

Кваліфікаційну роботу

допущено до захисту

« 07 » червня 2024 р.

Гарант освітньої програми:

к.т.н., доцент

Лавренчук Світлана Василівна

(підпис)

Луцьк – 2024 року

ЛУЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

Факультет комп'ютерних та інформаційних технологій

Кафедра комп'ютерної інженерії та кібербезпеки

Ступінь вищої освіти: бакалавр

Галузь знань: 12 Інформаційні технології

Спеціальність: 123 Комп'ютерна інженерія

Освітня програма: «Комп'ютерна інженерія»

ЗАТВЕРДЖУЮ

Завідувач кафедри

проф. Н.Черняшук

« 10 » 01 2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧУ ВИЩОЇ ОСВІТИ

Савчук Ользі Володимирівні

(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи Моніторинг платформи YouTube засобами SQL

Керівник роботи к.т.н., доцент Лавренчук Світлана Василівна

затверджені наказом закладу вищої освіти від «30» грудня 2023 року № 459/01-02

2. Строк подання здобувачем вищої освіти кваліфікаційної роботи 11.06.2024р.

3. Вихідні дані до роботи Джерелом розробки є науково-технічна література та публікації в періодичних виданнях з даного питання, опубліковані зарубіжні та вітчизняні роботи в даній області та різні інтернет-ресурси технічного спрямування

4. Зміст пояснювальної записки (перелік питань, які потрібно розробити):

Вступ

Аналіз існуючих підходів до моніторингу платформи YouTube

Особливості моніторингу платформи YouTube

Запити мовою SQL для отримання потрібних даних

Розробка веб-інтерфейсу моніторингової системи

Висновки

5. Перелік графічного (ілюстративного) матеріалу:

Структура таблиці з вхідними даними моніторингу

Приклади використання мови SQL для отримання додатку

Архітектура системи, структура додатку

Інтерфейс системи

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис	
		завдання видав	завдання прийняв
<i>Аналіз існуючих підходів до моніторингу платформи YouTube</i>	<i>Лавренчук С.В., доцент</i>		
<i>Теоретичне дослідження особливостей моніторингу платформи YouTube</i>	<i>Лавренчук С.В., доцент</i>		
<i>Практична реалізація моніторингової системи</i>	<i>Лавренчук С.В., доцент</i>		
<i>Нормоконтроль</i>	<i>Багнюк Н.В., доцент</i>		
<i>Гарант ОП</i>	<i>Лавренчук С.В., доцент</i>		
<i>Показник запозичень тексту</i>	____%		
<i>Академічна доброчесність</i>	<i>Міскевич О.І., асистент</i>		

7. Дата видачі завдання 10.01.2024 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1.	<i>Розділ 1. Аналіз існуючих підходів до моніторингу платформи YouTube</i>	до 15.02.2024 р.	Виконано
2.	<i>Розділ 2. Особливості моніторингу платформи YouTube</i>	до 15.03.2024 р.	Виконано
3.	<i>Розділ 3. Реалізація моніторингової системи</i>	до 04.05.2024 р.	Виконано
4.	<i>Висновки та пропозиції</i>	до 07.05.2025 р.	Виконано
5.	<i>Формування списку використаних джерел</i>	до 10.05.2024 р.	Виконано
6.	<i>Формування додатків</i>	до 15.05.2024 р.	Виконано
7.	<i>Оформлення ілюстративного матеріалу</i>	до 20.05.2024 р.	Виконано
8.	<i>Нормоконтроль</i>	до 01.06.2024 р.	Виконано
9.	<i>Інструментальна перевірка на академічний плагіат</i>	до 04.06.2024 р.	Виконано
10.	<i>Представлення кваліфікаційної роботи бакалавра до захисту</i>	до 11.06.2024 р.	Виконано

Здобувач вищої освіти

(підпис)

Савчук О.В.

(прізвище, ініціали)

Керівник кваліфікаційної роботи

(підпис)

Лавренчук С.В.

(прізвище, ініціали)

АНОТАЦІЯ

Савчук Ольга Володимирівна. Моніторинг платформи YouTube засобами SQL.

Кваліфікаційна робота бакалавра ОП «Комп'ютерна інженерія» спеціальності 123 Комп'ютерна інженерія. Луцький національний технічний університет. Луцьк, 2024. 57 с.

Кваліфікаційна робота складається з вступу, трьох розділів, висновків, списку використаних джерел, додатків.

Перший розділ присвячено аналізу існуючих способів моніторингу платформи YouTube. Також в цьому розділі здійснено огляд сервісів, за допомогою яких проводять аналітику платформи (Unbox Social та Quintly).

В другому розділі розглянуто засоби та методи моніторингу даних за допомогою YouTube Analytics та SQL.

Третій розділ присвячено опису розробленої програми для моніторингу даних платформи YouTube, візуалізації роботи запитів. Також показано які функції можна робити з фреймворком Django, для створення веб-застосунку.

Ключові слова: платформа YouTube, аналітика, база даних, Python, статистика.

ANNOTATION

Savchuk Olha Volodymyrivna. YouTube platform monitoring using SQL.

Bachelor's qualifying thesis of the OP «Computer Engineering» specialty 123 Computer Engineering. Lutsk National Technical University. Lutsk, 2024. 57 p.

The qualification work consists of an introduction, three sections, conclusions, a list of used sources, and appendices.

The first section is devoted to the analysis of existing ways of monitoring the YouTube platform. Also, in this section, an overview of the services used to conduct platform analytics (Unbox Social and Quintly) is carried out.

In the second section, the means and methods of monitoring data using YouTube Analytics and SQL are considered.

The third section is devoted to the description of the developed program for data monitoring of the YouTube platform, visualization of the work of requests. It also shows what features you can do with the Django framework to create a web application.

Keywords: YouTube platform, analytics, database, Python, statistics.

ЗМІСТ

ВСТУП	7
РОЗДІЛ 1 АНАЛІЗ ІСНУЮЧИХ ПІДХОДІВ ДО МОНІТОРИНГУ	
ПЛАТФОРМИ YOUTUBE	9
1.1 Огляд існуючих інструментів та програмних рішень	9
1.1.1 Unbox Social.....	9
1.1.2 Quintly	10
1.2 Оцінка переваг та недоліків існуючих підходів	12
РОЗДІЛ 2 ОСОБЛИВОСТІ МОНІТОРИНГУ ПЛАТФОРМИ YOUTUBE	14
2.1 Особливості моніторингу відео контенту	14
2.2 Засоби та методи моніторингу даних	16
2.3 Можливості SQL для аналізу даних на платформі YouTube	18
РОЗДІЛ 3 РЕАЛІЗАЦІЯ МОНІТОРИНГОВОЇ СИСТЕМИ.....	21
3.1 Вхідні дані.....	21
3.2 Запити мовою SQL для отримання потрібних даних.....	24
3.3 Розробка веб-інтерфейсу моніторингової системи	31
3.4 Приклади використання моніторингової системи.....	36
ВИСНОВКИ.....	38
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	39
ДОДАТКИ.....	41

ВСТУП

Актуальність теми. YouTube є однією з найпопулярніших платформ для завантаження та перегляду відео. В останні роки спостерігається надзвичайно значне зростання кількості відеоконтенту, що потребує ефективного моніторингу для виявлення та розуміння ключових тенденцій. Творцям відео, маркетологам і компаніям, які використовують YouTube для просування свого вмісту, потрібні інструменти аналізу даних.

Використання SQL дозволяє проводити глибокий аналіз і отримувати цінну інформацію для вирішення стратегічних завдань. SQL є потужним інструментом для роботи з базами даних, і його використання в контексті моніторингу YouTube дозволяє ефективно обробляти великі обсяги даних і отримувати точні результати.

Моніторинг та аналіз даних з YouTube може стати ключовим елементом стратегії розвитку бізнесу, дозволяючи приймати обґрунтовані рішення та підвищувати ефективність використання платформи для досягнення поставлених цілей.

Мета роботи – створення програми для постійного моніторингу та аналізу даних на платформі YouTube з метою виявлення ключових тенденцій та отримання звіту.

Об'єктом дослідження є платформа YouTube як основна соціальна мережа для обміну та перегляду відеоконтенту, зокрема, дослідження спрямоване на аналіз і моніторинг різних аспектів, пов'язаних із вмістом, користувачами та їхньою взаємодією на платформі.

Предмет дослідження – використання мови запитів SQL для отримання різноманітної статистики та інформації з бази даних YouTube.

Завдання дослідження:

- здійснити огляд існуючих систем моніторингу web-платформ;
- проаналізувати вхідні дані статистики YouTube та підготувати до використання;

- створити SQL-запити для визначення співвідношення підписок та переглядів відео, використовуючи базу статистичних даних YouTube;
- розробити систему моніторингу з web-інтерфейсом, який автоматично виконуватиме SQL-запити;
- провести аналіз отриманих результатів, виявити ключові тенденції переглядів платформи YouTube.

Апробація. За результатами дослідження опубліковано статтю у фаховому журналі, яка подана в додатку А [1].

РОЗДІЛ 1

АНАЛІЗ ІСНУЮЧИХ ПІДХОДІВ ДО МОНІТОРИНГУ ПЛАТФОРМИ YOUTUBE

1.1 Огляд існуючих інструментів та програмних рішень

Платформа Youtube, стала надзвичайно популярна серед користувачів не тільки, як засіб для розваг, а також як засіб для розвитку власного бізнесу чи стартапу. Також це є великий ресурс для поширення контенту різного жанру, від освітнього до розважального. Для ефективного використання Youtube, як і будь-якого інструменту для бізнесу, важливо передбачити можливість моніторингу для подальшого необхідного аналізу.

YouTube надає власний інструмент для аналізу відеостатистики – YouTube Analytics. Цей інструмент дозволяє користувачам отримувати дані про кількість переглядів, підписників, взаємодію з відео та багато іншого. Він має зручний інтерфейс і базовий функціонал для аналізу, але може бути обмежений у великих обсягах даних або специфічних запитах. Недоліком є те, що використання базового функціоналу обмежує нас у масштабному вивченні ринку та потреб аудиторії, що ускладнює можливість оперативно слідувати за трендами, та визначати закономірності.

Відповідно до потреб аналізу, з'явилося безліч готових рішень для проведення моніторингу та аналітики. Розглянемо декілька із них.

1.1.1 Unbox Social

Unbox Social дає повне уявлення про ефективність відео на YouTube. Можна отримувати інформацію про ефективність після публікації, щоб побачити, які відео приваблюють потенційну аудиторію, а які ні. Це дозволяє відстежувати кількість залучених і втрачених підписників, щоб краще зрозуміти зростання каналу. Він також надає глибокий аналіз демографічних даних безпосередньо аудиторії, щоб можна було оптимізувати свою стратегію. За допомогою цього інструменту аналітики YouTube можна бачити як з часом

змінюються перегляди відео. Таким чином, це дає змогу точно визначити, коли кількість переглядів різко зросла, і чи вплинуло на це конкретне відео [2].

Використання цього сервісу для дослідження аудиторії допоможуть зрозуміти поведінку аудиторії бренду, отримати глибоке розуміння того, хто є потенційною аудиторією та наскільки вона активна. Це допоможе оцінити, чи споживається контент і чи націлений він на потрібну аудиторію.

Також завдяки аналізу настроїв аудиторії цей інструмент допоможе зрозуміти аудиторію. Функція аналізу настроїв дозволить дізнатися, що люди говорять про наш бренд у соціальних мережах. Ця інформація допоможе зрозуміти, як наш бренд сприймається аудиторією, а також масштаби маркетингових стратегій.

Наявні інструменти для звітування в соціальних мережах нададуть дані за різними показниками, які допоможуть масштабувати існуючі стратегії. Є можливість отримувати звіти про соціальні мережі за будь-яку дату, місяць або рік. Звіти можна отримувати у потрібних нам форматах, включаючи PDF, PPT та XLS (Excel). Також передбачена кастомна можливість, для створення звітів власного формату [3].

1.1.2 Quintly

Quintly – один з найнадійніших інструментів аналітики YouTube на ринку. Він пропонує сотні показників, які можна відстежувати, щоб отримати повне уявлення про ефективність. Сюди входить статус підписки на канал і джерело трафіку. Він також показує місце відтворення та метрики карток для каналу. Інструмент дозволяє заглибитися в поверхневі показники і дає розбивку переглядів, що надходять від підписників і приватних відео. Він також дає краще розуміння того, як канал працював з самого початку. Для цього можна переглянути історичну кількість підписників і метрики за весь час існування.

Основними функціями Quintly є збір, аналіз і візуалізація даних з різних соціальних мереж, що дозволяє користувачам отримувати важливу інформацію для прийняття рішень та вдосконалення своїх стратегій в соціальних медіа.

Розглянемо основні можливості Quintly:

- дозволяє збирати дані з YouTube та інших соціальних медіа, такі як кількість підписників, перегляди відео, взаємодія з контентом, поширення публікації тощо;
- платформа надає інструменти для детального аналізу відеометрики, включаючи дані про перегляди, час перегляду, вподобання та коментарі;
- дозволяє порівнювати показники каналу на YouTube з конкурентами, щоб зрозуміти їхню активність та ефективність;
- платформа дозволяє користувачам створювати різноманітні звіти та дашборди для візуалізації даних та моніторингу результатів;
- має API, що дозволяє інтегрувати дані з платформи у власні програмні рішення чи інші сервіси;
- добре підходить для бізнесу, маркетологів та агентств, які шукають інструмент для аналізу та моніторингу своєї присутності в соціальних медіа, включаючи YouTube. Платформа надає широкий набір функцій та можливостей для глибокого аналізу даних та ефективного управління соціальними медіа.

Отже, Youtube став невід'ємним інструментом користувачів, які використовують його для розваг, а також і для розвитку власного бізнесу або стартапу. Для ефективного використання цього сервісу у бізнесі необхідно проводити моніторинг діяльності. Для цього є Youtube Analytics, що є вбудованим інструментом для аналізу відеостатистики. Проте його базовий функціонал може бути обмеженим у великих обсягах даних або для специфічних запитів. Нові готові рішення для моніторингу та аналітики, такі як Unbox Social та Quintly, дозволяють отримати більше інформації про ефективність відео. Вони надають глибокий аналіз демографічних даних аудиторії, допомагають відстежувати тренди та розуміти, як наш контент сприймається аудиторією. У виборі між цими інструментами важливо враховувати свої потреби, масштаб діяльності та бюджет. Кожен з них має свої переваги та недоліки, але загалом допомагає зробити моніторинг та аналіз відеоконтенту на YouTube більш ефективним та продуктивним.

1.2 Оцінка переваг та недоліків існуючих підходів

YouTube став однією з найвпливовіших платформ сучасності. З роками вона почала сильно розвиватись, і на сьогоднішій день використовується не лише для висвітлювання своїх думок та життя, але й для розвитку бізнесу та стартапів. Однак, для ефективного використання YouTube у бізнесі необхідно проводити моніторинг діяльності та аналіз відеоконтенту. Нижче розглянемо переваги та недоліки існуючих підходів до моніторингу на YouTube.

YouTube Analytics є вбудованим інструментом платформи, що дозволяє легко отримувати доступ до відеостатистики. Інтерфейс YouTube Analytics досить зручний та інтуїтивно зрозумілий, що спрощує процес аналізу даних. YouTube Analytics надає базовий набір функцій для аналізу, таких як кількість переглядів, підписників, взаємодія з відео тощо.

Базовий функціонал YouTube Analytics може бути обмеженим у великих обсягах даних або для специфічних запитів. Використання базового функціоналу обмежує можливість масштабного вивчення ринку та потреб аудиторії [4].

Якщо говорити про Unbox Social, то він дозволяє отримати глибокий аналіз демографічних даних аудиторії, що допоможе оптимізувати стратегію контенту. За допомогою Unbox Social можна відстежувати зростання кількості підписників та взаємодію з контентом. Недоліком цієї системи можуть стати фінансові витрати на використання цієї системи.

Щодо Quintly, вона пропонує широкий набір функцій для збору, аналізу та візуалізації даних з різних соціальних мереж. Особливістю цієї системи є доступний функціонал для порівняння з конкурентами. Це дозволяє порівнювати показники нашого каналу на YouTube з конкурентами для кращого розуміння ефективності та подальшого аналізу дій. Хоч і ця функція дуже корисна, проте інтерфейс Quintly може бути занадто складним для новачків та вимагати додаткового часу для освоєння.

Також варто розглянути варіант із написанням власної системи для моніторингу. Така система дозволяє отримати більше переваг, ніж недоліків,

тому що її можна максимально адаптувати та пристосувати для власних потреб враховуючи усю необхідну специфіку використання. Викликом для такого рішення може стати необхідність знань в сфері аналітики даних, та розробки програмного забезпечення. Усе це знадобиться у ході розробки додатку та для вирахування необхідних статистичних даних. Але варто пам'ятати про гнучкість такого рішення, та максимальну ефективність оскільки це рішення буде не універсальним, а націленим на конкретні завдання та потреби.

Отже, існують різноманітні підходи моніторингу, від використання вбудованих інструментів, таких як YouTube Analytics, до використання спеціалізованих платних сервісів, наприклад, Unbox Social або Quintly. Кожен з цих підходів має свої переваги та недоліки, і вибір залежить від конкретних потреб та можливостей користувача. YouTube Analytics забезпечує базовий набір функцій без додаткових витрат, але може бути обмеженим у великих обсягах даних. У свою чергу, платні сервіси, такі як Unbox Social та Quintly, надають більше функціоналу для аналізу та порівняння з конкурентами, але вимагають фінансових витрат та часу на освоєння інтерфейсу. Також варто розглянути можливість розробки власної системи для моніторингу, яка забезпечить максимальну гнучкість та адаптивність до конкретних потреб.

РОЗДІЛ 2

ОСОБЛИВОСТІ МОНІТОРИНГУ ПЛАТФОРМИ YOUTUBE

2.1 Особливості моніторингу відео контенту

Як уже згадували, Youtube потужний інструмент для поширення відео контенту, та використання його у різних цілях. І для якісного моніторингу потрібно проаналізувати провідні показники та метрики, які є важливими для аналізу.

Перерахуємо доступні метрики які можна отримати використовуючи різні інструменти:

- перегляди;
- залучення;
- місця відтворення;
- тип пристрою;
- ключові слова;
- джерела трафіку.

Це лише декілька метрик, їх список можна доповнювати, наприклад, середня тривалість перегляду, загальний час перегляду, утримання аудиторії та інші [5].

Перегляди відображають кількість разів, коли відео було переглянута. Важливо відслідковувати динаміку цього показника, щоб зрозуміти, як ефективно контент привертає увагу аудиторії [6]. Крім того, взаємодія з відео, така як коментарі, вподобання та підписки, є ключовими метриками, які допомагають зрозуміти рівень зацікавленості та залучення аудиторії.

Перегляди – це кількість разів, коли відео відтворюється, і вони також включають повторні перегляди. Це може здатися простою метрикою, але це складний показник. Для того, щоб було зараховано «перегляд», відео має бути переглянута більше 30 секунд. З точки зору того, хто намагається утримати увагу аудиторії протягом перших кількох секунд свого відео, 30 секунд – це дуже багато часу [7].

Залученість аудиторії відображається завдяки поширенням, лайкам, дизлайкам. Коментарі можуть вказувати на емоційну складову глядача, та ставлення його до переглянутого контенту. Поширення мають здатність викликати довіру до контенту і водночас зумовлюють переходи із інших ресурсів до перегляду відео. У сукупності ці показники відображатимуть рівень залученості контенту.

Глядачі знаходять відео через різні джерела, і YouTube Analytics дозволяє визначити, що це за джерела. Вони можуть знайти контент через пошукові запити, плейлисти, запропоновані відео тощо, на які, звісно, впливає алгоритм YouTube. Але на додаток до алгоритму YouTube, який впливає на охоплення відео, глядачі також переглядають відео, переходячи за прямими або зовнішніми посиланнями. Аналіз того, як глядачі знайшли контент, допоможе визначити джерела, які приносять найбільше трафіку. Потім можна скористатися цими джерелами і більше працювати над тими, які не отримують стільки трафіку, скільки повинні [8].

Поза базовими метриками, які було згадано раніше, існують інші важливі показники, які можуть бути корисні для аналізу відеоконтенту на YouTube.

Середня тривалість перегляду – цей показник вказує на середню кількість часу, яку глядачі витрачають на перегляд нашого відео. Чим вищий цей показник, тим більше зацікавленість аудиторії нашим контентом. Але важливо враховувати, що не завжди довші відео означають кращу взаємодію з аудиторією. Якщо звертатись до короткотривалих тем або швидкого споживання контенту, коротші відео можуть бути ефективнішими.

Загальний час перегляду. Ця метрика вказує на загальну кількість часу, який глядачі витрачають на перегляд всіх наших відео. Вона дає загальний огляд того, як контент узгоджується зі споживачами та чи зберігається їх інтерес протягом тривалого періоду часу.

Утримання аудиторії. Цей показник показує, яка частина аудиторії залишалася на відео протягом тривалості відео. Високий рівень утримання аудиторії свідчить про те, що контент цікавий та захоплюючий для глядачів. Але

низький рівень утримання може вказувати на те, що слід переглянути та покращити якість або привабливість контенту.

Отже, ці метрики, разом із базовими показниками, надають комплексний огляд ефективності відеоконтенту на YouTube. Важливо регулярно аналізувати ці дані для того, щоб оптимізувати свою стратегію контенту та забезпечити максимальну взаємодію з аудиторією.

2.2 Засоби та методи моніторингу даних

До початку розробки системи моніторингу, варто розібратись в особливостях, засобах та правилах моніторингу.

На процес моніторингу та аналізу безпосередньо впливає якість даних. Моніторинг якості даних – це оцінка, вимірювання та управління даними організації з точки зору точності, узгодженості та надійності. Він використовує різні методи для виявлення та вирішення проблем з якістю даних, гарантуючи, що високоякісні дані використовуються для бізнес-процесів та прийняття рішень. Важливість якості даних важко переоцінити, оскільки неякісні дані можуть призвести до неправильних висновків, неефективних операцій та відсутності довіри до інформації, що надається системами компанії. Моніторинг може гарантувати, що проблеми з якістю даних будуть виявлені на ранній стадії, до того, як вони зможуть вплинути на бізнес-операції організації або на її клієнтів [9].

Розглянемо ключові виміри якості даних, на які зазвичай звертають увагу під час моніторингу якості даних:

- точність вимірює ступінь коректності при порівнянні значень з їхнім справжнім представленням;
- повнота оцінює ступінь присутності та доступності всіх необхідних даних;
- узгодженість стосується однорідності даних у різних джерелах або системах;

- своєчасність оцінює, наскільки актуальною є інформація щодо її цільового використання;
- достовірність стосується дотримання заздалегідь визначених форматів, правил або стандартів для кожного атрибуту в наборі даних;
- унікальність гарантує, що в наборі даних не існує дублікатів записів;
- цілісність допомагає підтримувати референтні зв'язки між наборами даних без будь-яких розривів.

Використання систем моніторингу дозволяє не лише виявляти потенційні проблеми, а й своєчасно реагувати на них, що забезпечує надійність та вірогідність даних, необхідних для прийняття стратегічних рішень та забезпечення успіху бізнесу [10].

Розглянемо основні підходи та методи, які допомагають забезпечити високу якість та надійність даних.

- Профілювання даних. Основною метою профілювання даних є вивчення та аналіз структури, вмісту та взаємозв'язків даних для виявлення аномалій і недоліків. Це сприяє отриманню інформації про якість даних, включаючи типи, довжини, шаблони та унікальні значення. Існують три основних типи профілювання: за стовпцями, за залежностями та за надмірністю. Кожен з цих типів дозволяє отримати всебічне розуміння даних та виявити потенційні проблеми з якістю.

- Аудит даних полягає в оцінці точності та повноти даних шляхом порівняння їх з певними правилами або стандартами. Цей процес допомагає виявити неправильні, відсутні або суперечливі дані. Аудит може проводитися як вручну, так і за допомогою автоматизованих інструментів, що допомагають виявити розбіжності в даних.

- Правила якості даних визначають критерії, яким повинні відповідати дані для забезпечення їхньої точності, повноти та надійності. Вони можуть включати перевірку на дублікати, відповідність еталонним даним та відповідність форматам.

– Очищення даних – це процес виявлення та виправлення помилок, невідповідностей і неточностей у даних. Цей процес передбачає використання різних методів, таких як перевірка, перетворення та дедуплікація, для забезпечення точності та повноти даних.

– Моніторинг даних у режимі реального часу полягає в безперервному відстеженні та аналізі даних з моменту їхнього створення. Це дозволяє виявляти та вирішувати проблеми з якістю даних негайно.

– Тестування продуктивності даних – це процес оцінки ефективності та масштабованості систем обробки даних та інфраструктури.

– Управління метаданими – це організація, підтримка та використання метаданих для покращення якості, узгодженості та зручності використання даних.

Отже, моніторинг якості даних є важливим етапом для будь-якого проведення аналізу чи майбутнього використання. Цей процес передбачає систематичне відстеження, аналіз та оновлення даних з метою забезпечення їхньої точності, повноти, актуальності та достовірності. Моніторинг якості даних допомагає виявляти та усувати проблеми з якістю, такі як дублікати, помилки введення даних, відсутність або неправильність інформації. Це забезпечує високу якість даних, що є ключовим для прийняття обґрунтованих рішень та успішного функціонування бізнесу.

2.3 Можливості SQL для аналізу даних на платформі YouTube

Найпопулярніше використання SQL сьогодні – це мова запитів, яка дозволяє користувачам створювати інформаційні панелі та інструменти звітності. Складні інструкції можна легко передавати до баз даних, а маніпулювати даними можна за лічені секунди. Простота використання дозволяє користувачам відображати дані за допомогою інтуїтивно зрозумілих інформаційних панелей, які здатні ілюструвати дані різними способами. Коли SQL використовується для аналізу даних, мова запитів бази даних одночасно

взаємодіє з декількома базами даних, а також використовує реляційні бази даних. Ця гнучка мова є доступною для користувачів і водночас забезпечує необхідну глибину, що дозволяє створювати розширені інформаційні панелі та інструменти для аналізу даних. Хоча мова SQL проста, вона здатна виконувати складний аналіз даних [11].

Вона існує з 1970-х років і сьогодні є найпоширенішим методом доступу до даних у базах даних. SQL має безліч функцій, які дозволяють користувачам читати, маніпулювати та змінювати дані. Хоча SQL зазвичай використовується інженерами при розробці програмного забезпечення, вона також популярна серед аналітиків даних з кількох причин:

- проста семантика, легка для розуміння та вивчення;
- можливість використання не використовуючи сторонні додатки;
- порівняно з інструментами електронних таблиць, аналіз даних, виконаний на SQL, легко піддається аудиту та реплікації [12].

Розглянемо деякі можливості SQL для аналізу даних на платформі YouTube:

1. Аналіз переглядів відео. SQL дозволяє витягувати дані про кількість переглядів відео, які можна далі аналізувати за різними критеріями. Наприклад, можна визначити найпопулярніші відео на каналі, тенденції у зміні кількості переглядів з часом, а також вивчити ефективність різних типів контенту.

2. Взаємодія з контентом. Використовуючи SQL можна проаналізувати взаємодію аудиторії з контентом, включаючи кількість лайків, дизлайків, коментарів та підписок під відео. Це допомагає зрозуміти, як аудиторія реагує на вміст та які типи контенту викликають найбільше зацікавлення.

3. Демографічні дані аудиторії. Отримавши демографічні дані аудиторії, такі як вік, стать, місце проживання та інші можна розуміти склад та інтереси аудиторії каналу, що може бути корисним для розробки персоналізованих стратегій залучення аудиторії.

4. Тренди та прогнозування. Можна аналізувати тренди у зміні популярності відео та прогнозувати майбутні тенденції. Це дозволяє бізнесу та

маркетологам бути впевненими у своїх стратегіях та адаптувати їх відповідно до змін у споживчому попиті.

Використання SQL для аналізу даних на платформі YouTube відкриває широкі можливості для отримання цінної інформації та прийняття обґрунтованих рішень. Поєднуючи дані з YouTube з іншими джерелами, такими як дані веб-аналітики або соціальних мереж, бізнес може отримати повний образ ефективності своїх маркетингових стратегій та покращити результати своєї діяльності.

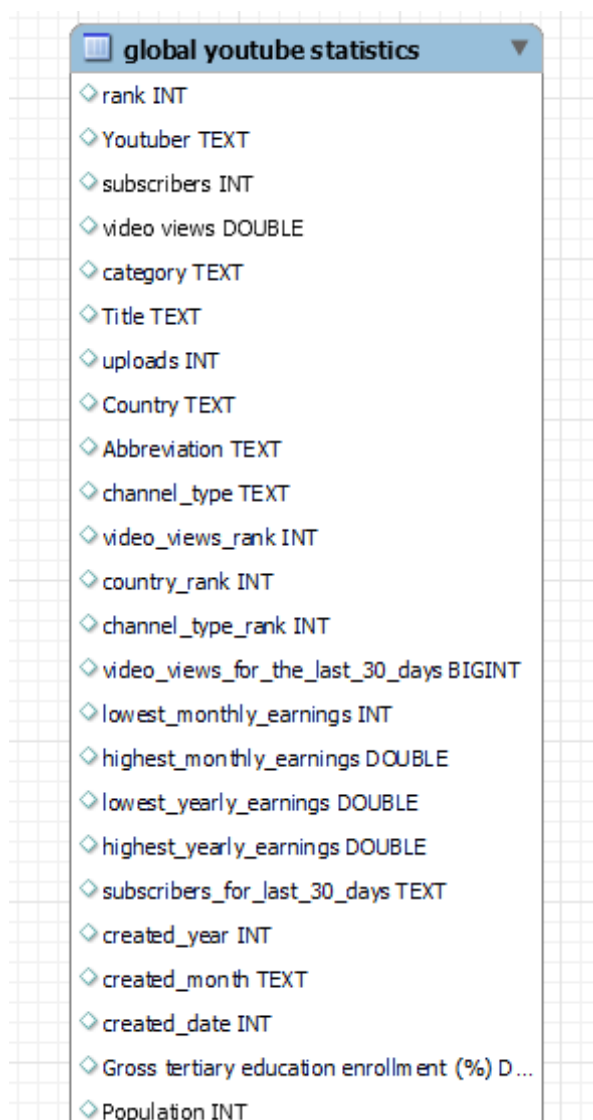
SQL сьогодні використовується як мова запитів для створення інформаційних панелей та інструментів звітності, особливо у сфері аналізу даних. Вона є найпоширенішим методом доступу до даних у базах даних і володіє безліччю функцій для їх читання, маніпулювання та зміни. Хоча SQL зазвичай використовується розробниками програмного забезпечення, вона стає все більш популярною серед аналітиків даних завдяки своїй простоті та можливості виконувати складний аналіз даних.

РОЗДІЛ 3

РЕАЛІЗАЦІЯ МОНІТОРИНГОВОЇ СИСТЕМИ

3.1 Вхідні дані

Створено базу даних в СУБД MySQL, яка містить статистику каналів YouTube з найбільшою кількістю підписок та створено таблицю (рисунок 3.1). Дані було взято з платформи YouTube за 2023 рік у форматі csv та завантажено в MySQL Workbench – інтегроване середовище розробки баз даних, яке надає зручні інструменти для роботи з різними системами управління базами даних (СУБД) за допомогою мови запитів SQL.



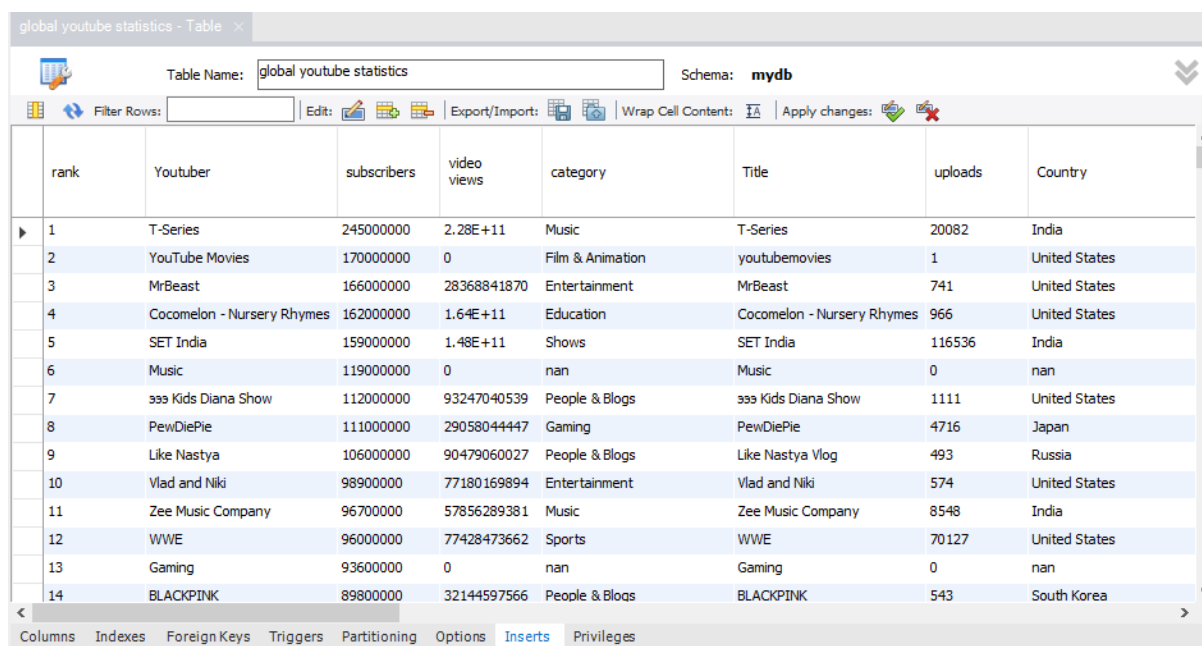
The image shows a screenshot of the MySQL Workbench interface. At the top, there is a tab labeled 'global youtube statistics'. Below the tab, a list of columns is displayed, each with a diamond icon to its left. The columns and their data types are as follows:

Column Name	Data Type
rank	INT
Youtuber	TEXT
subscribers	INT
video views	DOUBLE
category	TEXT
Title	TEXT
uploads	INT
Country	TEXT
Abbreviation	TEXT
channel_type	TEXT
video_views_rank	INT
country_rank	INT
channel_type_rank	INT
video_views_for_the_last_30_days	BIGINT
lowest_monthly_earnings	INT
highest_monthly_earnings	DOUBLE
lowest_yearly_earnings	DOUBLE
highest_yearly_earnings	DOUBLE
subscribers_for_last_30_days	TEXT
created_year	INT
created_month	TEXT
created_date	INT
Gross tertiary education enrollment (%) D...	
Population	INT

Рисунок 3.1 – Структура таблиці з даними для дослідження

SQL Workbench є потужним інструментом для розробки та адміністрування баз даних, який відповідає потребам розробників і адміністраторів даних будь-якого рівня. Перевагами є інсталювання в різних операційних системах, включаючи Windows, macOS і Linux, що робить його універсальним інструментом для роботи з базами даних, незалежним від платформи. Підтримка різноманітних СУБД, включаючи MySQL, PostgreSQL, Oracle, Microsoft SQL Server, зручний інтерфейс, має гнучку архітектуру, яка дозволяє встановлювати різні розширення та плагіни для розширення його функціональності відповідно до певних потреб – це все сприяє його популярності.

Для зберігання даних в базі даних було створено таблицю «global youtube statistics» (рисунок 3.2), що містить інформацію про статистику YouTube-каналів.



rank	Youtuber	subscribers	video views	category	Title	uploads	Country
1	T-Series	245000000	2.28E+11	Music	T-Series	20082	India
2	YouTube Movies	170000000	0	Film & Animation	youtubemovies	1	United States
3	MrBeast	166000000	28368841870	Entertainment	MrBeast	741	United States
4	Cocomelon - Nursery Rhymes	162000000	1.64E+11	Education	Cocomelon - Nursery Rhymes	966	United States
5	SET India	159000000	1.48E+11	Shows	SET India	116536	India
6	Music	119000000	0	nan	Music	0	nan
7	эээ Kids Diana Show	112000000	93247040539	People & Blogs	эээ Kids Diana Show	1111	United States
8	PewDiePie	111000000	29058044447	Gaming	PewDiePie	4716	Japan
9	Like Nastya	106000000	90479060027	People & Blogs	Like Nastya Vlog	493	Russia
10	Vlad and Niki	98900000	77180169894	Entertainment	Vlad and Niki	574	United States
11	Zee Music Company	96700000	57856289381	Music	Zee Music Company	8548	India
12	WWE	96000000	77428473662	Sports	WWE	70127	United States
13	Gaming	93600000	0	nan	Gaming	0	nan
14	BLACKPINK	89800000	32144597566	People & Blogs	BLACKPINK	543	South Korea

Рисунок 3.2 – Дані з таблиці «global youtube statistics»

Таблиця складається з рядків (записів) та стовпців (полів), де кожен стовпець визначає тип даних, які можуть бути збережені у відповідному рядку, а саме:

- «rank» (int) – ранг каналу;

- «Youtuber» (text) – ім'я або назва власника каналу;
- «subscribers» (int) – кількість підписників на каналі;
- «video views» (double) – кількість переглядів відео на каналі;
- «category» (text) – категорія каналу;
- «Title» (text) – назва каналу або заголовок відео;
- «uploads» (int) – кількість відео, які завантажені на канал;
- «Country» (text) – країна, до якої відноситься канал;
- «Abbreviation» (text) – аббревіатура країни;
- «channel_type» (text) – тип каналу;
- «video_views_rank» (int) – ранг кількості переглядів відео;
- «country_rank» (int) – ранг країни;
- «channel_type_rank» (int) – ранг типу каналу;
- «video_views_for_the_last_30_days» (bigint) – кількість переглядів відео за останні 30 днів;
- «lowest_monthly_earnings» (int) – найнижчі щомісячні доходи;
- «highest_monthly_earnings» (double) – найвищі щомісячні доходи;
- «lowest_yearly_earnings» (double) – найнижчі щорічні доходи;
- «highest_yearly_earnings» (double) – найвищі щорічні доходи;
- «subscribers_for_last_30_days» (text) – інформація про кількість підписників за останні 30 днів;
- «created_year» (int) – рік створення каналу;
- «created_month» (text) – місяць створення каналу;
- «created_date» (int) – дата створення каналу;
- «Gross tertiary education enrollment (%)» (double) – відсоток брутної вищої освіти;
- «Population» (int) – населення країни;
- «Unemployment rate» (double) – рівень безробіття;
- «Urban_population» (int) – кількість міського населення;
- «Latitude» (double) – широта місця розташування каналу;
- «Longitude» (double) – довгота місця розташування каналу.

Для створення таблиці були використані такі типи даних:

INTEGER або INT – використовується для зберігання цілих чисел, таких як ідентифікатори та кількість елементів;

DOUBLE – використовується для зберігання чисел з плаваючою комою, таких як десяткові дробки;

TEXT – використовується для зберігання текстових рядків.

3.2 Запити мовою SQL для отримання потрібних даних

Структура SQL-запитів може бути різною залежно від типу запиту та його складності, що подано на рисунку 3.3.

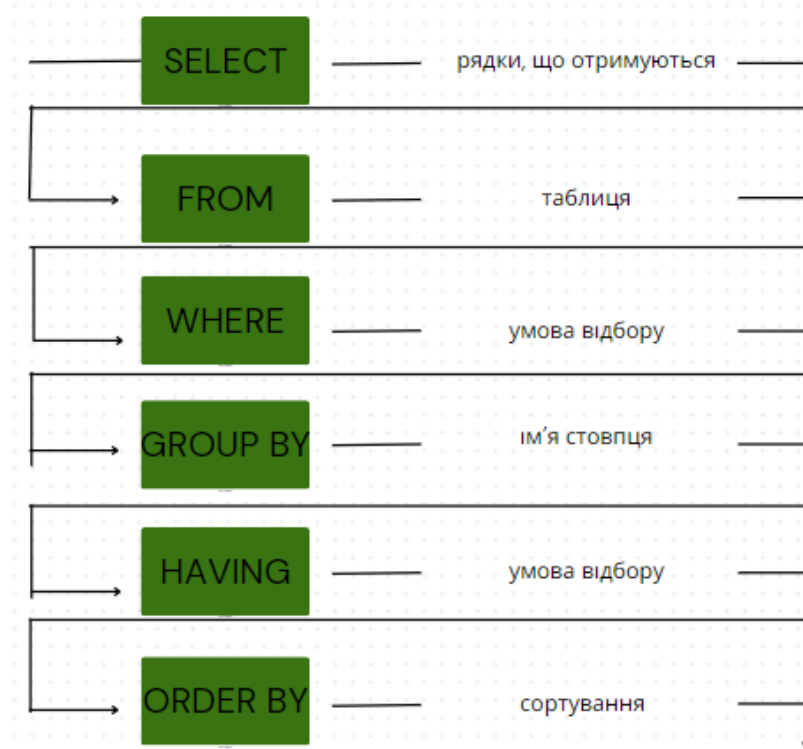


Рисунок 3.3 – Структура SQL-запиту

Розглянемо детально. Частина SELECT визначає список стовпців, які будуть отримані в результаті запиту. Стовпці можуть містити значення, отримані зі стовпців у таблицях бази даних, або значення, обчислені під час виконання запиту [13].

Частина FROM визначає список таблиць, що містять елементи даних, до яких потрібно зробити запит.

Частина WHERE містить умови, які вибирають рядки для включення в результати запиту.

Блоки GROUP BY використовуються для створення підсумкових запитів. Типовий запит включає в себе запис для кожного рядка в таблиці, яка є результатом запиту. Потім підсумковий запит спочатку групує рядки бази даних за вказаними атрибутами, а потім включає підсумковий рядок для кожної групи в результати запиту.

Блок HAVING визначає, що лише деякі групи, створені за допомогою GROUP BY, включені до результатів запиту. У цій частині ми виберемо групи за критеріями відбору.

Блок ORDER BY упорядковує результати запиту на основі даних, що містяться в одному або кількох стовпцях.

Розглянемо на прикладі SQL-запиту на рисунку 3.4, який вибирає назви Youtuber-ів та типи їхніх каналів із таблиці «global youtube statistics», де тип каналу дорівнює «Music». Результати сортуються за зменшенням алфавіту назв Youtuber-ів та обмежуються 25 рядками.

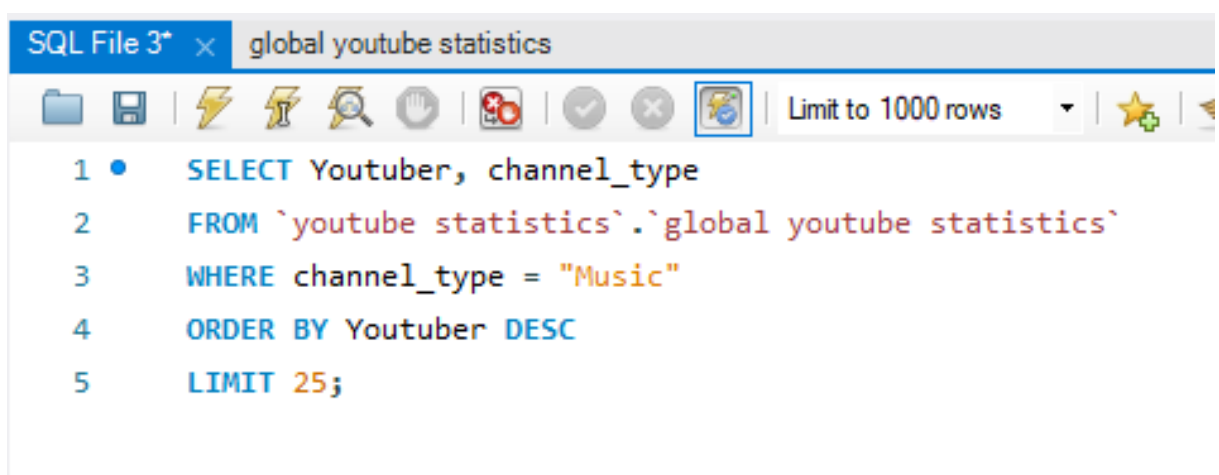
The image shows a screenshot of a SQL editor window. The title bar reads "SQL File 3* x global youtube statistics". The editor contains a SQL query with five lines, numbered 1 to 5. The query is: 1 SELECT Youtuber, channel_type; 2 FROM `youtube statistics`.`global youtube statistics`; 3 WHERE channel_type = "Music"; 4 ORDER BY Youtuber DESC; 5 LIMIT 25; The editor interface includes a toolbar with icons for file operations, a search icon, and a "Limit to 1000 rows" dropdown menu.

Рисунок 3.4 – Приклад SQL-запиту

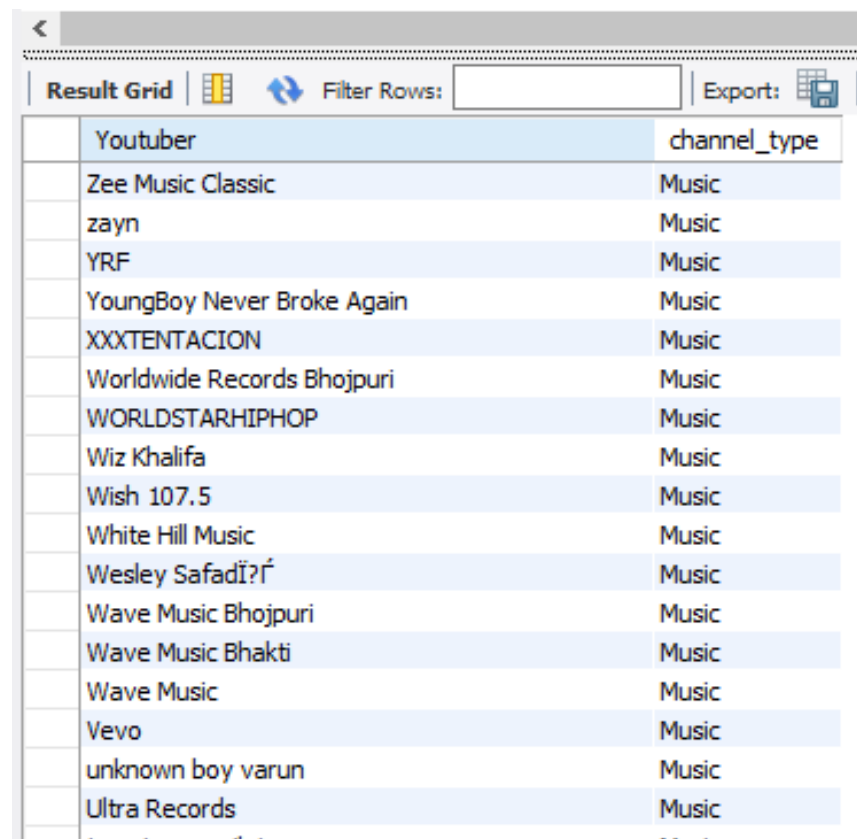
SELECT Youtuber, channel_type – вибір стовпців, які містять назву «Youtuber» та «channel_type».

FROM `youtube statistics`.`global youtube statistics` – вибір даних з таблиці «global youtube statistics».

WHERE channel_type = «Music» – вибір рядків, де тип каналу дорівнює «Music»

ORDER BY Youtuber DESC – сортування результатів за назвою «Youtuber» у спадному порядку.

LIMIT 25 – обмеження кількості результатів. Тобто буде подано лише 25 результатів, як на рисунку 3.5.



Youtuber	channel_type
Zee Music Classic	Music
zayn	Music
YRF	Music
YoungBoy Never Broke Again	Music
XXXTENTACION	Music
Worldwide Records Bhojpuri	Music
WORLDSTARHIPHOP	Music
Wiz Khalifa	Music
Wish 107.5	Music
White Hill Music	Music
Wesley Safadī?ġ	Music
Wave Music Bhojpuri	Music
Wave Music Bhakti	Music
Wave Music	Music
Vevo	Music
unknown boy varun	Music
Ultra Records	Music

Рисунок 3.5 – Результат запиту

Ще декілька прикладів SQL-запитів подано нижче.

1. Запит на виведення даних з таблиці «global youtube statistics» та його результат подано на рисунку 3.6.

SQL File 3*
Limit to 1000 rows

```
1 SELECT * FROM `youtube statistics`.`global youtube statistics`
```

rank	Youtuber	subscribers	video views	category	Title	uploads	Country	Abbreviation	channel_type	video_views_rank	country
1	T-Series	245000000	228000000000	Music	T-Series	20082	India	IN	Music	1	1
2	YouTube Movies	170000000	0	Film & Animation	youtubemovies	1	United States	US	Games	4055159	7670
3	MrBeast	166000000	28368841870	Entertainment	MrBeast	741	United States	US	Entertainment	48	1
4	Cocomelon - Nursery Rhymes	162000000	164000000000	Education	Cocomelon - Nursery Rhymes	966	United States	US	Education	2	2
5	SET India	159000000	148000000000	Shows	SET India	116536	India	IN	Entertainment	3	2
7	ннн Kids Diana Show	112000000	93247040539	People & Blogs	ннн Kids Diana Show	1111	United States	US	Entertainment	5	3
8	PewDiePie	111000000	29058044447	Gaming	PewDiePie	4716	Japan	JP	Entertainment	44	1
9	Like Nastya	106000000	90479060027	People & Blogs	Like Nastya Vlog	493	Russia	RU	People	630	5
10	Vlad and Niki	98900000	77180169894	Entertainment	Vlad and Niki	574	United States	US	Entertainment	8	5
11	Zee Music Company	96700000	57856289381	Music	Zee Music Company	8548	India	IN	Music	12	3
12	WWE	96000000	77428473662	Sports	WWE	70127	United States	US	Sports	7	6
14	BLACKPINK	89800000	32144597566	People & Blogs	BLACKPINK	543	South Korea	KR	Music	32	1
16	Sony SAB	83000000	101000000000	Shows	Sony SAB	71270	India	IN	Entertainment	4	5
17	5-Minute Crafts	80100000	26236790209	Howto & Style	5-Minute Crafts 2.0	1	United Kingdom	GB	Entertainment	4057901	4797
18	BANGTANTV	75600000	20826993957	Music	BANGTANTV	2281	South Korea	KR	Music	112	2
19	Sports	75000000	0	nan	sports	3	United States	US	Entertainment	3898122	6266

Рисунок 3.6 – Результат виведення даних з таблиці «global youtube statistics»

2. Запит на рисунку 3.7 показує канал, його тип, країну, та кількість переглядів, за умови, що кількість переглядів перевищує ліміт 200000, а також тип каналу «Music» або «Comedy». Дані відсортовано за спадаючим списком з лімітом 8, що подано на рисунку 3.8.

```
1 SELECT Youtuber, channel_type, Country, subscribers
2 FROM `youtube statistics`.`global youtube statistics`
3 WHERE channel_type = "Music" OR channel_type = "Comedy"
4 ORDER BY
5 CASE
6   WHEN subscribers > 200000 THEN subscribers
7   ELSE Youtuber
8 END ASC
9 LIMIT 8;
```

Рисунок 3.7 – SQL-запит

	Youtuber	channel_type	Country	subscribers
▶	Make Joke Of	Comedy	India	123000000
	Avril Lavigne	Music	United States	124000000
	vector	Comedy	United States	124000000
	CKM	Comedy	India	125000000
	Wave Music Bhakti	Music	India	125000000
	Mundo Bitá	Music	Brazil	125000000
	NickiMinajAtVEVO	Music	United States	126000000
	Ellie Goulding	Music	United Kingdom	126000000

Рисунок 3.8 – Результат

rank	Youtuber	subscribers	video views	category	Title	uploads	Country	Abbreviation	channel_type	video_views_rank	country
2	YouTube Movies	17000000	0	Film & Animation	youtubemovies	1	United States	US	Games	4055159	7670
3	MrBeast	166000000	28368841870	Entertainment	MrBeast	741	United States	US	Entertainment	48	1
4	Cocomelon - Nursery Rhymes	162000000	16400000000	Education	Cocomelon - Nursery Rhymes	966	United States	US	Education	2	2
5	SET India	159000000	14800000000	Shows	SET India	116536	India	IN	Entertainment	3	2
7	ннн Kids Diana Show	112000000	93247040539	People & Blogs	ннн Kids Diana Show	1111	United States	US	Entertainment	5	3
8	PewDiePie	111000000	29058044447	Gaming	PewDiePie	4716	Japan	JP	Entertainment	44	1
9	Like Nastya	106000000	90479060027	People & Blogs	Like Nastya Vlog	493	Russia	RU	People	630	5
10	Vlad and Niki	98900000	77180169894	Entertainment	Vlad and Niki	574	United States	US	Entertainment	8	5
12	WWE	96000000	77428473662	Sports	WWE	70127	United States	US	Sports	7	6
16	Sony SAB	83000000	10100000000	Shows	Sony SAB	71270	India	IN	Entertainment	4	5
17	5-Minute Crafts	80100000	26236790209	Howto & Style	5-Minute Crafts 2.0	1	United Kingdom	GB	Entertainment	4057901	4797
19	Sports	75000000	0	nan	sports	3	United States	US	Entertainment	3898122	6266
22	Zee TV	70500000	73139054467	Entertainment	Zee TV	129204	India	IN	Entertainment	9	6
23	Pinkfong Baby Shark - Kids' S...	68200000	38843229963	Education	Pinkfong Baby Shark - Kids' S...	2865	United States	US	Education	23	8
25	ChuChu TV Nursery Rhymes ...	65900000	45757850229	Education	ChuChu TV Nursery Rhymes ...	633	India	IN	Education	18	7
27	Colors TV	64600000	61510906457	Shows	Colors TV	112915	India	IN	Entertainment	10	9

rank	Youtuber	subscribers	video views	category	Title	uploads	Country	Abbreviation	channel_type	video_views_rank	country
30	Movieclips	59500000	59316472754	Film & Animation	Movieclips	39113	United States	US	Film	11	9
34	Aaj Tak	57600000	25307753534	News & Politics	Aaj Tak	283775	India	IN	News	71	13
41	Infobells - Hindi	53300000	30516172739	Education	Infobells - Hindi	577	India	IN	Education	37	16
44	BillionSurpriseToys - Nurser...	52200000	9877365274	Education	BillionSurpriseToys - Nurser...	847	United States	US	Education	450	15
45	JuegaGerman	48100000	14631710289	Gaming	JuegaGerman	2052	Chile	CL	Games	237	1
47	Get Movies	47500000	30788679536	Entertainment	Get Movies	3322	Russia	RU	Film	35	1
48	Shemaroo	47400000	22519705183	Entertainment	Shemaroo	11451	India	IN	Entertainment	97	17
51	A4	46300000	22936630813	Gaming	aefour	1	Cuba	CU	People	4053056	99
53	Fernanflo	46100000	10323391593	Gaming	Fernanflo	543	El Salvador	SV	Games	419	1
54	Voot Kids	45500000	21388725229	Entertainment	Voot Kids	4660	India	IN	Entertainment	105	18
55	Felipe Neto	45200000	16602198273	Entertainment	Felipe Neto	4331	Brazil	BR	Entertainment	174	2
56	Vocі?і?і?і?і?	44700000	7828610828	Entertainment	Vocі?і?і?і?і?	1558	Brazil	BR	Entertainment	681	3
57	HAR PAL GEO	44600000	41139050371	Entertainment	HAR PAL GEO	100755	Pakistan	PK	Entertainment	20	1
60	whinderssonunes	44200000	4274709210	Comedy	whinderssonunes	558	Brazil	BR	Comedy	1741	4
61	ABS-CBN Entertainment	44200000	50292540392	People & Blogs	ABS-CBN Entertainment	193890	Philippines	PH	Entertainment	16	1
62	Alan Walker	43700000	12884264778	Music	alanwalker	8	Thailand	TH	Film	4056116	4026

Рисунок 3.14 – Результат

6. Запит на рисунку 3.15 вибирає всі дані з таблиці «global youtube statistics», де тип каналу не є музичним (тобто `channel_type != «Music»`) і кількість підписників більше 25,600,000 (`subscribers > 25600000`). Результати на рисунку 3.16 сортуються за кількістю підписників у зростаючому порядку (`ORDER BY subscribers ASC`) і обмежуються першими 15 рядками (`LIMIT 15`).

```

1 • SELECT *
2 FROM `youtube statistics`.`global youtube statistics`
3 WHERE channel_type != "Music" and subscribers > 25600000
4 ORDER BY subscribers asc
5 LIMIT 15

```

Рисунок 3.15 – SQL-запит

rank	Youtuber	subscribers	video views	category	Title	uploads	Country	Abbreviation	channel_type	video_views_rank	country_rank	channel_type
225	LosPolinesios	25700000	7466926260	People & Blogs	LosPolinesios	982	Mexico	MX	People	746	8	7
223	VanossGaming	25800000	15541421838	Gaming	VanossGaming	1762	Canada	CA	Games	206	5	17
222	Indosiar	25900000	11372071889	Music	Indosiar	65286	Indonesia	ID	Entertainment	358	5	60
217	Smosh	26100000	10435474336	Comedy	Smosh	1619	United States	US	Comedy	410	65	13
216	Junya.いづみ?いづみ?	26200000	16097531087	Entertainment	Junya.いづみ?いづみ?	5985	Japan	JP	Entertainment	184	2	56
215	D Billions	26200000	31977463002	nan	D Billions	775	United States	US	Entertainment	33	64	58
214	BB Ki Vines	26300000	4749833967	Entertainment	BB Ki Vines	190	India	IN	Entertainment	1462	51	57
213	PANDA BOI	26400000	17211600007	Entertainment	PANDA BOI	967	Italy	IT	Entertainment	158	1	55
212	Alan Chikin Chow	26400000	27006526665	Comedy	Alan Chikin Chow	865	United States	US	Comedy	61	63	12
211	Goldmines Bollywood	26400000	8595760553	Film & Animation	Goldmines Bollywood	3622	India	IN	Film	588	50	15
209	Raffy Tulfo in Action	26500000	15065753455	News & Politics	Raffy Tulfo in Action	10022	Philippines	PH	News	219	3	5
206	Beast Reacts	26700000	4388047013	People & Blogs	Beast Reacts	241	United States	US	Entertainment	1632	61	54
208	Netflix	26700000	7173668905	Entertainment	Netflix	6471	United States	US	Entertainment	797	62	55
207	enchufetv	26700000	10317306313	Comedy	enchufetv	975	Colombia	CO	Comedy	420	5	11
205	Renato Garcia YT	26900000	7938616641	Entertainment	Renato Garcia YT	3956	Brazil	BR	Entertainment	664	11	54

Рисунок 3.16 – Результат

3.3 Розробка веб-інтерфейсу моніторингової системи

Після опрацювання таблиці за допомогою запитів потрібно розробити інтерфейс, за допомогою якого, можна отримати різні статистичні дані.

Для розробки веб-інтерфейсу за основу взято фреймворк Django. Фреймворк обрано через простоту розробки базових проєктів, та за наявністю у мові програмування Python інших бібліотек які допоможуть легко візуалізувати наші дані.

Django надає набір зареєстрованих та готових до використання компонентів, що дозволяє розробникам швидко створювати веб-додатки без необхідності писати все з нуля. Він має гнучку архітектуру, яка дозволяє легко масштабувати додаток з ростом обсягу даних та трафіку.

Фреймворк вбудовує ряд заходів безпеки, таких як захист від SQL-ін'єкцій, Cross-Site Scripting (XSS) та Cross-Site Request Forgery (CSRF), що допомагає уникнути багатьох типових загроз безпеки веб-додатків.

Базовий проєкт на Django надає готовий адміністративний інтерфейс, який можна легко налаштувати та використовувати для управління даними в базі даних.

Django надає ORM, що дозволяє розробникам працювати з базою даних за допомогою об'єктів Python, що спрощує роботу з даними та зменшує ризик виникнення помилок [14].

Фреймворк має потужний механізм роботи з шаблонами, що дозволяє розробникам розділяти логіку та представлення, що полегшує управління кодом та робить його більш читабельним.

Django має високоякісну та докладну документацію, яка допомагає розробникам швидко вивчити фреймворк та знайти відповіді на питання [15].

Розглянемо послідовність розгортання проєкту на Django:

1. Необхідно підготувати віртуальне середовище та довантажити бібліотеку фреймворку. Використання віртуального середовища такого як `pipenv` допоможе уникнути проблем із не співпадінням версій та зробить розгортання проєкту на різних платформах однаковим. Для завантаження віртуального середовища та бібліотеки Django використовуємо `pip` (рисунок 3.17 та рисунок 3.18).

```
PS E:\TEST DJANGO> pip install pipenv
you should consider upgrading via the C:\Users\NICK\A
PS E:\TEST DJANGO> pipenv install
Creating a virtualenv for this project...
```

Рисунок 3.17 – Завантаження віртуального середовища

```
Alternatively, run a command inside the v
PS E:\TEST DJANGO> pip install django
```

Рисунок 3.18 – Завантаження бібліотеки Django, використовуючи `pip`

2. Розгортання базової структури веб-додатку на базі Django демонструє простоту та швидкість розгортання проєкту (рисунок 3.19).

```
PS E:\TEST DJANGO> django-admin startproject youtube_app
```

Рисунок 3.19 – Розгортання базової структури

Після цього ми можемо побачити кістяк проєкту (рисунок 3.20). Варто переконатись чи проєкт вдало розгорнувся, для цього можна локально запустити його за допомогою команди, наведеної в додатку Б, спробувати провзаємодіяти із файлом `manage.py` (рисунок 3.20), та відкрити за посиланням у браузері.

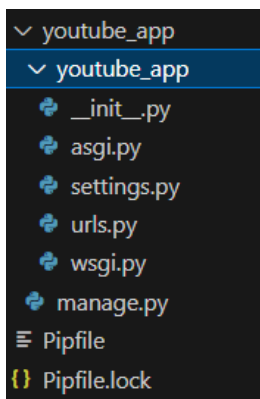


Рисунок 3.20 – Складові проєкту

3. Створення додатку. Кожен Django-проєкт має один або кілька додатків. Кожен додаток відповідає за конкретну функціональну частину проєкту. Наприклад, ми можемо мати додаток для аутентифікації користувачів, додаток для роботи з продуктами тощо. У нашому випадку створюємо додаток `main`, (рисунок 3.21) який і буде включати у себе весь функціонал (рисунок 3.22).

```
PS E:\TEST DJANGO> django-admin startapp main
```

Рисунок 3.21 – Створення додатку `main`

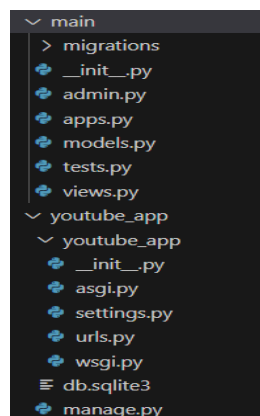


Рисунок 3.22 – Складові додатку `main`

Попередні пункти показують розгортання шаблонного проєкту на Django на далі розглядаємо необхідні дії для того щоб веб-додаток став інструментом для моніторингу.

Для подальшої роботи нам потрібно забезпечити логічну частину обчислень статистики та додати теки. Для зручності розробки розглянемо оновлену структуру нашого додатку та його файлів (додаток В). Отже призначення цих тек та файлів:

`static` – ця папка призначена для зберігання статичних файлів, таких як CSS, JavaScript, зображення тощо. Ці файли використовуються для відображення веб-сторінок.

– `statistics`: це власна назва папки, яку ми використовуємо для збереження csv-файлів які будуть використовуватися для обрахування статистики у додатку;

– `templates` – ця папка містить файли шаблонів HTML, які використовуються для відображення веб-сторінок. Для зручності виконаний базовий шаблон, на основі якого будуть надбудовуватись інші сторінки (додаток Г). Ми створили блок, вміст якого буде мінятись у рядках 29-31 (додаток Г). У різних файлах прописаний вміст сторінки, який буде відображатись, приклад такого вмісту наведено в додатку Д;

– `admin.py` – цей файл містить конфігурацію адміністративного інтерфейсу Django для моделей даного додатку. Використовується для налаштування відображення та поведінки моделей у веб-інтерфейсі адміністратора;

– `models.py` – містить опис моделей даних для даного додатка. Моделі визначають структуру та взаємозв'язки даних у базі даних;

– `urls.py` – цей файл містить визначення маршрутів URL для даного додатка. Він вказує, які веб-адреси пов'язані з кожною сторінкою у додатку (додаток Е);

– `views.py` – цей файл містить визначення сторінок (`views`) для даного додатка. Відображення виконують логіку обробки запитів за допомогою функцій та відображення відповідних шаблонів які зберігаються у папці `templates`.

Приклад вигляду функції у файлі views наведено в додатку Ж. Можна простежити частину коду який відповідає за вирахування даних та побудову таблиці.

Розглянувши усі особливості варто окреслити логіку за якою працює додаток та послідовність його опрацювання запиту (рисунок 3.23).

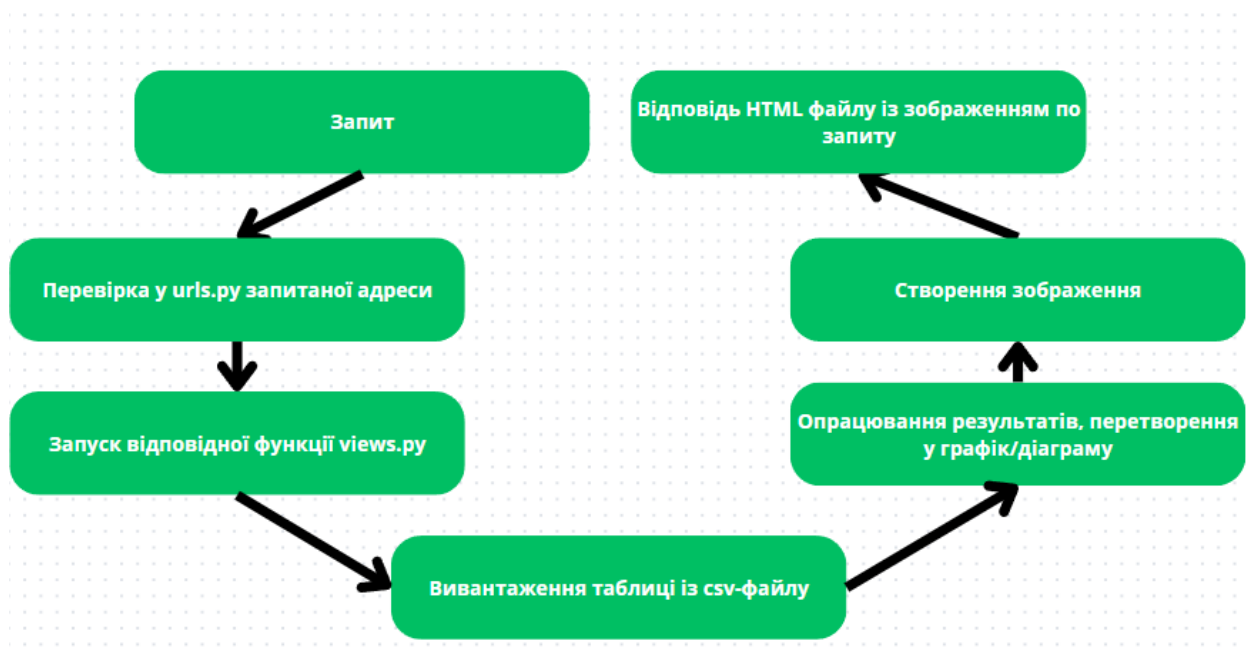


Рисунок 3.23 – Схема роботи додатку

1. Отримання запиту за адресою сторінки. Наш додаток опрацьовує запит в адресі, наприклад 127.0.0.1:8000/last25, де адреса після домену буде вказувати на адресу запиту «last25».

2. У файлі urls.py шукається відповідний зв'язок за запитом. У нашому випадку «last25». Відповідний 8 рядок (додаток Е).

3. Запускається функція яка була прописана як відповідь у файлі urls.py 8 рядок (додаток Е), внаслідок чого, відбувається опрацювання запиту.

4. Кожне обчислення статистики передбачає постійне перечитування csv-файлу та вивантаження даних із нього.

5. Отримана таблиця опрацьовується згідно запиту, проводиться вибірка даних та створення із цих даних графіку/діаграми.

6. Експорт графіку/діаграми у зображення. Це спрощений елемент додатку для легкості подальшого використання отриманих даних та імпортування їх у HTML файл

7. Формування відповіді у вигляді HTML-файлу із зображенням графіку/діаграми.

Отже, розглянувши послідовність виконання та структуру проекту можна зробити висновок щодо легкого використання фреймворку для написання власних застосунків. При цьому варто відзначити модульність проекту, кожна логічна частина знаходиться в окремому файлі та ці файли зв'язуються між собою.

3.4 Приклади використання моніторингової системи

Розглянемо приклад використання власної розробленої моніторингової системи YouTube, побудованої на фреймворку Django. Проаналізуємо можливості та доцільність її використання.

Для початку розглянемо інтерфейс (додаток І). Інтерфейс додатку простий та надає змогу вибрати уже заздалегідь підготовлену статистику. Для цього використовуємо активну кнопку «Обери статистику». Після чого у нас з'являється викидний список із переліком доступних графіків/діаграм. Коли ми обираємо одну із статистик, наприклад, «25 найпопулярніших каналів», тоді нас переадресує на нову сторінку. Далі відбудеться обчислення даних із таблиці, яка у нас збережена в директорії statistic.

Спробуємо проаналізувати цю статистику (додаток К). Отже, у топі знаходиться канал «T-Series». Виконавши пошук цього каналу, можна побачити що канал викладає індійські пісні та їх підбірки. Така велика кількість переглядів зумовлена великою кількістю відео. В цілому, якщо проаналізувати цей топ-25, то бачимо, що усі канали спрямовані на розважальний контент та контент для дітей.

Проаналізуємо «останні 25 переглянутих каналів за їх типом» (додаток Л). Бачимо зовсім протилежні показники. У списку останніх переглянутих каналів лідером є канали категорії «Спорт», наступні «Як зробити?», треті у графіку «Навчальні». Ігри та розваги у цьому топі зайняли останні позиції.

Після детального огляду інтерфейсу та аналізу різноманітності доступних статистичних даних, можемо зробити кілька висновків щодо можливостей та доцільності використання такої моніторингової системи.

Інтерфейс системи (додаток И) приваблює своєю простотою та зручністю використання. Вибір доступної статистики за допомогою викидного списку дозволяє користувачам швидко та легко отримувати необхідну інформацію.

Широкий спектр доступних статистичних даних дозволяє користувачам аналізувати різні характеристики каналів та їх вмісту.

Система надає можливість порівняння різних каналів за різними показниками, що допомагає користувачам отримати уявлення про конкурентну ситуацію та тренди на платформі YouTube.

Отже, можна зробити висновок, що використання моніторингової системи YouTube на Django є доцільним для власників каналів та маркетологів, оскільки вона надає широкі можливості для аналізу та оптимізації їх контенту та стратегій просування. Система є гнучкою для покращень та внесення змін. Зручний інтерфейс та різноманітність доступних статистичних даних роблять цю систему ефективним інструментом для успішного управління каналом на YouTube та маркетологам для оцінювання стану справ в індустрії. Недоліком цієї системи є ж її гнучкість, оскільки покращення або внесення нових можливостей вимагатиме додаткових знань та обов'язкових навичок роботи із python для створення нових статистичних даних.

ВИСНОВКИ

В кваліфікаційній роботі було проведено аналіз та здійснено розробку моніторингової системи для платформи YouTube з використанням мови SQL. В процесі дослідження було виявлено, що можливості SQL дозволяють ефективно здійснювати аналіз великих обсягів даних, що відображають активність користувачів та контенту на платформі YouTube.

Основні поставлені завдання були успішно виконані, а саме:

- Здійснено огляд існуючих систем моніторингу web-платформ таких як Unbox Social та Quintly.
- Проаналізовано вхідні дані статистики YouTube та підготовано до використання, зокрема конвертовано з формату csv до sql.
- Створено SQL-запити для визначення співвідношення підписок та переглядів відео, використовуючи базу статистичних даних YouTube.
- Розроблено систему моніторингу з web-інтерфейсом, який автоматично виконуватиме SQL-запити.
- Проведено аналіз отриманих результатів, виявлено ключові тенденції переглядів платформи YouTube.

Результати дослідження дозволили створити ефективну систему, яка забезпечує зручний доступ до важливої інформації, необхідної для аналізу та прийняття рішень, стосовно створення контенту на платформі YouTube.

Використання мови SQL в контексті моніторингу платформи YouTube виявилось дуже перспективним, дозволяючи ретельно аналізувати різні параметри, такі як популярність відео, залучення глядачів, охоплення переглядів.

Таким чином, отримані дані демонструють успішність використання SQL для моніторингу платформи YouTube і можуть бути корисними для компаній, які займаються відеомаркетингом, аналітикою та розробкою веб-платформи.

Це дослідження дозволяє нам зрозуміти динаміку змін на платформі та сприяє ефективному управлінню контентом і ресурсами.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Лавренчук, С., Христинець, Н., & Савчук, О. Моніторинг платформи YouTube засобами SQL. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво*. 2024. № 54, С. 130-136. URL: <https://doi.org/10.36910/6775-2524-0560-2024-54-15> (дата звернення: 18.04.2024).
2. YouTube analytics tools to improve your marketing in 2024. Sprout social. URL: <https://sproutsocial.com/insights/youtube-analytics-tools/> (дата звернення: 19.04.2024).
3. Social media analytics social media tools unbox social.unboxsocial. URL: <https://www.unboxsocial.com/features/social-media-analytics> (дата звернення: 19.04.2024).
4. Ashwini, T., Sahana, L. M., Mahalakshmi, E., Shweta, S. P. Youtube data analysis using hadoop framework URL: https://www.researchgate.net/figure/Analysis-of-YouTube-data-using-Hadoop-MapReduce-framework_fig3_355342772 (дата звернення: 20.04.2024).
5. Top websites ranking. most visited websites in Ukraine. URL: <https://www.similarweb.com/top-websites/ukraine/> (дата звернення: 19.02.2024)
6. Аналітика YouTube: 14 показників для відстеження ефективності вашого відеоконтенту. URL: <https://wave.video/ua/blog/youtube-analytics-metrics/> (дата звернення: 20.02.2024)
7. YouTube Analytics. URL: <https://vidiq.com/ru/blog/post/youtube-shorts-analytics/> (дата звернення: 20.04.2024).
8. Manko B. A. Video advertising: using youtube analytics for the target audience. *Journal of Information Technology Teaching Cases*. 2023 13(1), 77-81.
9. Data Quality monitoring techniques metrics to Watch.IBM Blog. URL: <https://www.ibm.com/blog/8-data-quality-monitoring-techniques/> (дата звернення: 22.04.2024).
10. Top 50 youtube channels. boost your views and subscribers on youtube - vidiq. URL: <https://vidiq.com/youtube-stats/top/50/> (дата звернення: 10.03.2024).

11. Ginsberg C. Why every Data Analyst should know SQL.noble desktop design coding classes and certificate programs. URL: <https://www.nobledesktop.com/classes-near-me/blog/why-data-analysts-need-sql> (дата звернення: 22.04.2024).
12. SQL for Data Analysis: Tutorial introduction — mode. Mode resources. URL: <https://mode.com/sql-tutorial/introduction-to-sql> (дата звернення: 22.04.2024).
13. Огляд основних SQL запитів. URL: <https://itvdn.com/ua/blog/article/m-sql#nvukckd7gfvs> (дата звернення: 20.04.2024).
14. Django REST framework. URL: <https://www.django-rest-framework.org/> (дата звернення: 23.04.2024).
15. Django The web framework for perfectionists with deadlines. URL: <https://www.djangoproject.com/> (дата звернення: 23.04.2024).