

1. Dataset Overview

У проєкті використано датасет із Kaggle, який містить інформацію про відео, що потрапили у розділ YouTube Trending.

Дані включають кількість переглядів, лайків, дизлайків, коментарів, дати публікації та дати появи у трендах, а також іншу метаінформацію.

Джерело: Kaggle (YouTube Trending Videos Analysis)

Структура сирого датасету:

- 40 950 рядків
- 16 колонок
- 16 категорій
- 6 351 унікальних відео

Сирі дані розміщено в аркуші "USvideos_raw".

2. Data Cleaning & Preparation

Для обробки була створена копія сирої таблиці – "USvideos_clean".

У ній виконано такі кроки:

1. Форматування дат:
 - створено `trending_date_clean` та `publish_time_clean`
 - дати переведені у стандартний формат (через формули DATE / DATEVALUE).
2. Додавання назви категорій:
 - створено колонку `category_name`
 - значення додані через `ARRAYFORMULA + VLOOKUP` на основі `category_id`.
3. Перевірка форматів колонок:

- числові колонки приведено до Number
 - дати – до Date/DateTime
 - текстові – до Plain Text
 - пропусків у ключових метриках (views, likes, dislikes, comment_count, category_id) не виявлено
 - пропуски були лише у `description`, але ця колонка не використовується в аналізі.
-

3. Deduplication & Feature Engineering

Оскільки одне відео може потрапляти у тренди багато разів, у сирій таблиці воно повторюється по днях.

Для даного аналізу створено окремий аркуш – "USvideos_per_video", який зберігає кожне відео один раз.

У цій таблиці сформовані такі колонки:

- фінальні значення views, likes, dislikes, comments (останній день перебування у трендах)
- `first_trending_date`
- `last_trending_date`
- `days_in_trending`
- `ratio_likes_dislike` – likes / dislikes
- `high_view` – бінарна ознака для графіків (наприклад, views > X).

Ці дані стали основою для побудови всіх графіків.

4. Metrics Used

Для аналізу основними метриками були:

- views
- likes
- dislikes
- comments
- likes-to-dislikes ratio
- share of high-view videos

Для порівняння категорій обрано медіану, а не середнє, оскільки в багатьох категоріях присутні аутлайєри, які сильно спотворюють середні значення.

Процес імпорту JSON

JSON-файл категорій було завантажено у Google Sheets та конвертовано у формат таблиці через вбудовану функцію імпорту (File → Import → Upload). З цієї таблиці були взяті відповідності `category_id` → `category_name`.

5. Insights & Visualizations

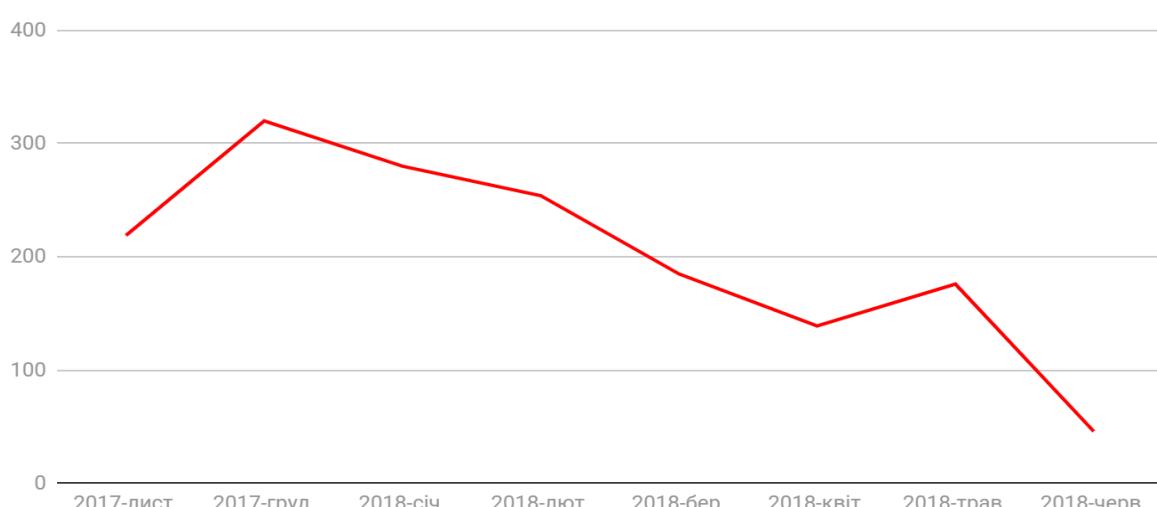
У цьому розділі представлені ключові візуалізації та інсайти, отримані в результаті аналізу трендових відео YouTube. Кожна діаграма відповідає на конкретне аналітичне питання та супроводжується коротким поясненням.

5.1 Time Trends

Що показує діаграма:

Динаміку кількості унікальних відео, що потрапляли у тренди, з розбивкою по місяцях для обраної категорії.

Trending Video Count Over Time



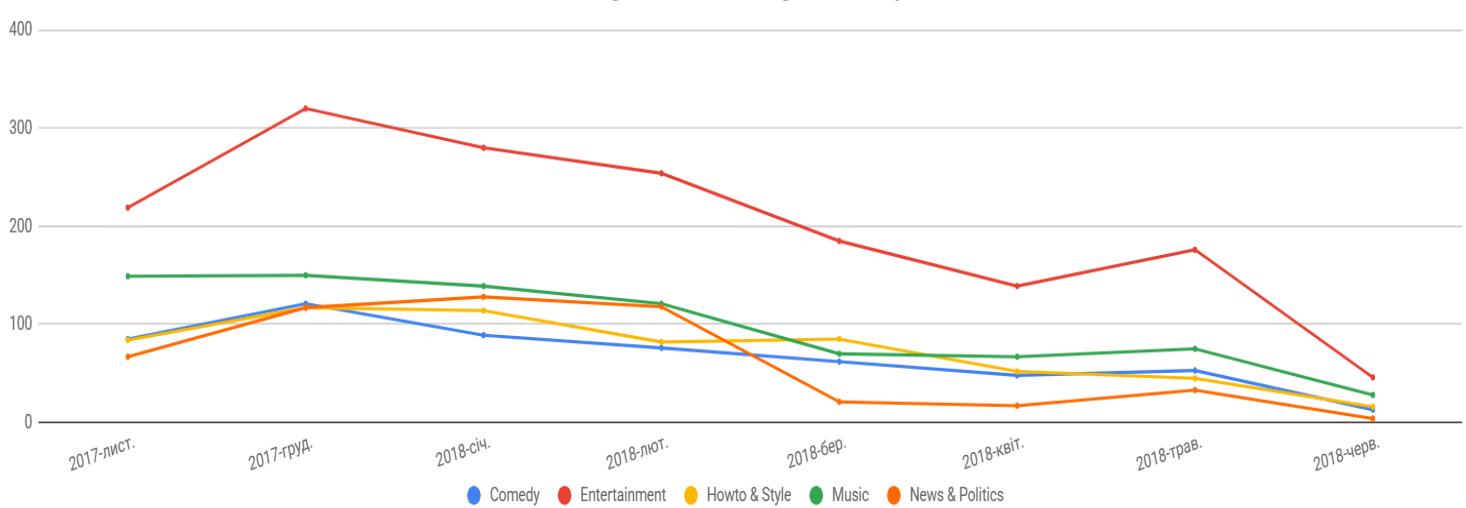
*Діаграма оновлюється динамічно залежно від обраної категорії. На зображені показано приклад для категорії Entertainment.

Інсайт:

Кількість трендових відео має чітку сезонність: для більшості категорій пікові значення припадають на грудень–січень.

Що показує діаграма: Динаміку кількості унікальних відео, що потрапляли у тренди, для топ-5 категорій у розрізі місяців.

TOP 5 Categories – Trending Volume by Month



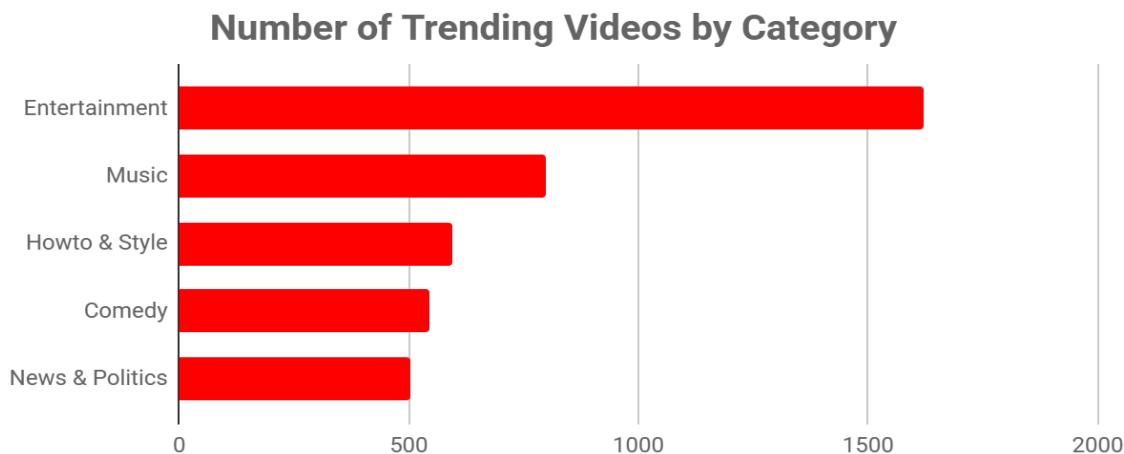
Інсайт:

Категорія *Entertainment* демонструє найбільшу волатильність із різким зростанням у зимові місяці, тоді як *Comedy* та *Howto & Style* мають більш стабільну динаміку протягом року.

5.2 Category Performance

Що показує діаграма:

Топ-5 категорій за кількістю унікальних відео, які потрапили у тренди.

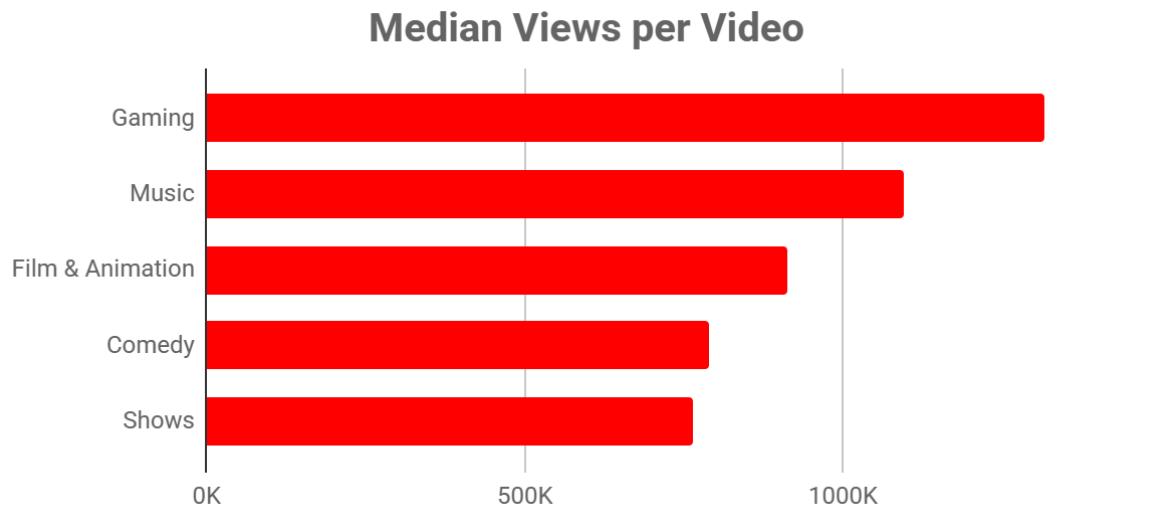


Інсайт:

Категорія *Entertainment* домінує за кількістю трендових відео ($\approx 25\%$), що вказує на її широку представленість у трендах, але не обов'язково на найвищу залученість.

Що показує діаграма:

Медіанну кількість переглядів на одне відео для топ-5 категорій.

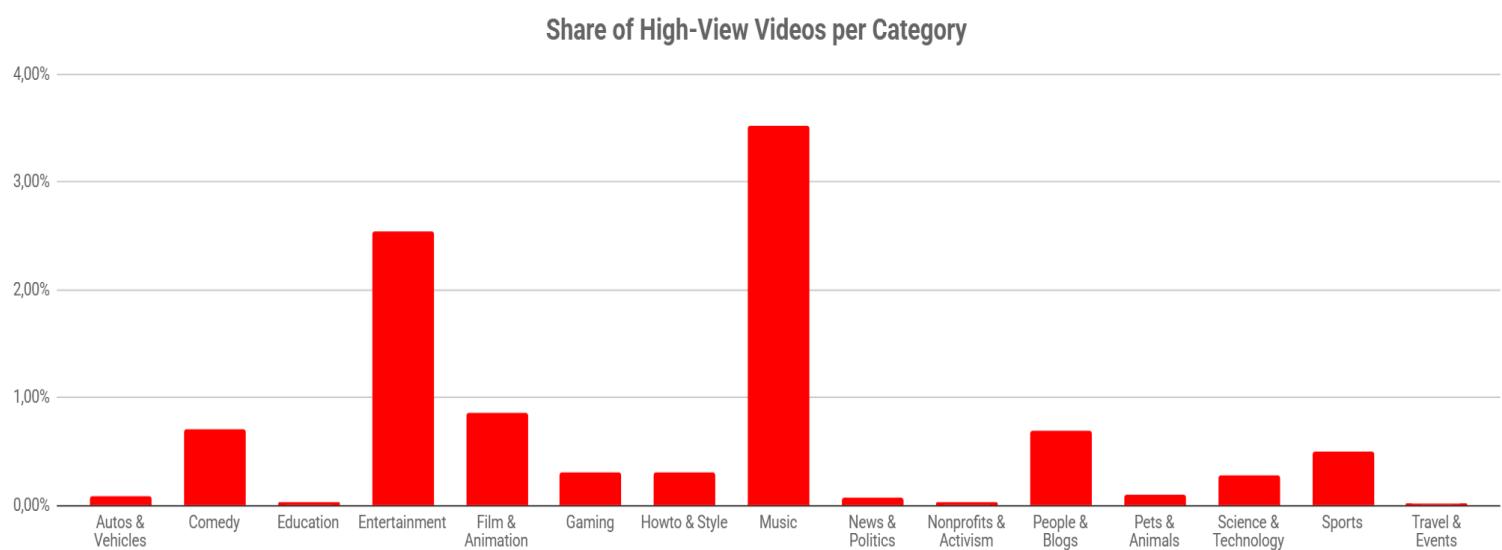


Інсайт:

Music та *Gaming* мають вищу медіану переглядів, ніж *Entertainment*, що свідчить про вищий середній інтерес до відео цих категорій.

Що показує діаграма:

Частку відео з дуже високою кількістю переглядів (вище 90-го перцентиля) в межах кожної категорії.



Інсайт:

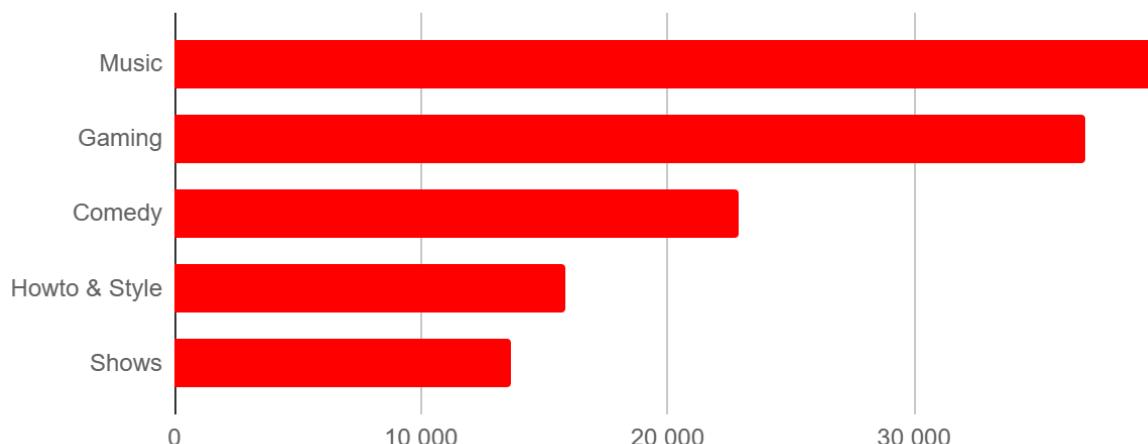
Категорії *Music* та *Entertainment* мають найбільшу частку відео з дуже високими переглядами, що підтверджує їхню здатність регулярно генерувати вірусний контент.

5.3 Engagement Metrics

Що показує діаграма:

Медіанну кількість вподобайок на одне відео для топ-5 категорій.

Median Likes per Video

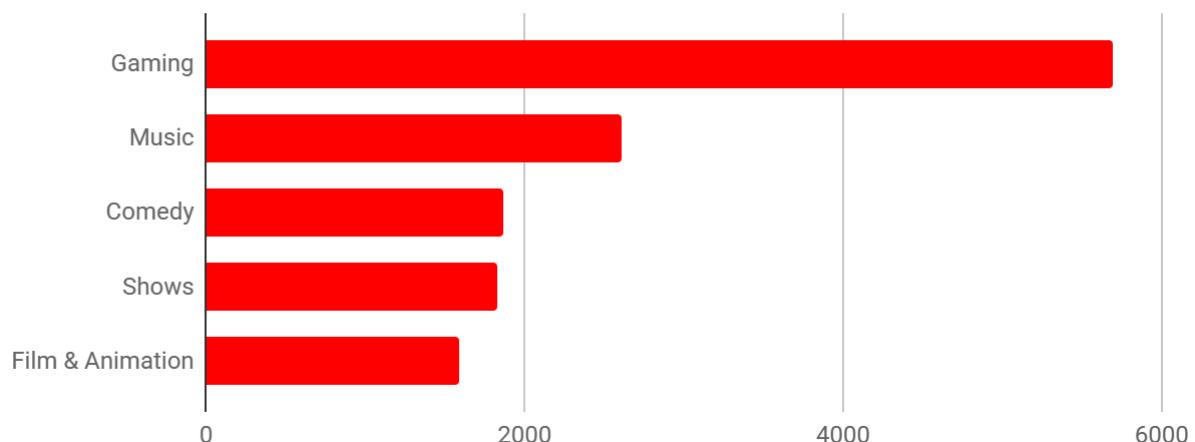


Відео з категорій *Music* та *Gaming* отримують більше вподобайок у розрахунку на одне відео, що свідчить провищий рівень емоційної залученості аудиторії порівняно з іншими категоріями.

Що показує діаграма:

Медіанну кількість коментарів на одне трендове відео для топ-5 категорій.

Median Comments per Video



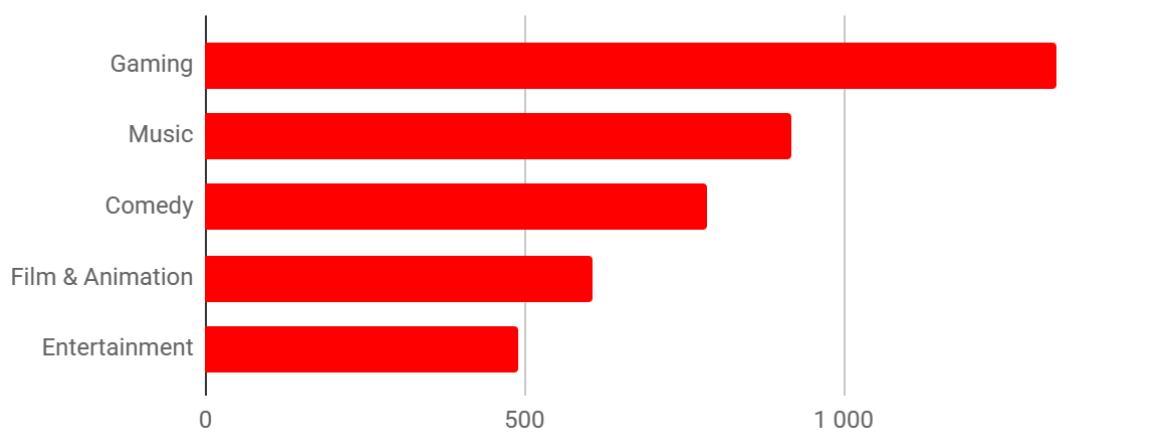
Інсайт:

Категорія *Gaming* демонструє найвищий рівень коментування, що може свідчити про активну взаємодію спільноти та високий рівень обговорення контенту серед глядачів.

Що показує діаграма:

Медіанну кількість дизлайків на одне трендове відео для топ-5 категорій.

Median Dislikes per Video

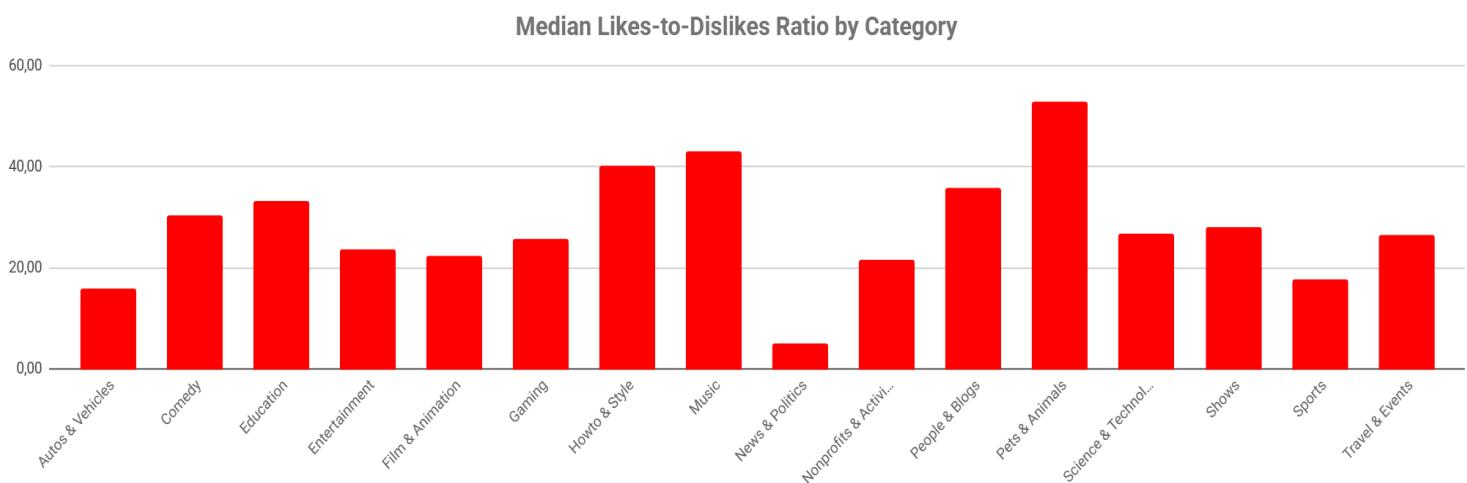


Інсайт:

Підвищені значення медіани дизлайків у деяких категоріях можуть вказувати не лише на негативне сприйняття контенту, а й на поляризацію аудиторії та загалом високий рівень залученості.

Що показує діаграма:

Співвідношення вподобайок до дизлайків (медіанне значення) для кожної категорії.



Інсайт:

Найвищий показник *likes-to-dislikes ratio* спостерігається у категорії *Pets & Animals*, що свідчить про стабільно позитивне сприйняття контенту. Найнижчі значення характерні для категорії *News & Politics*, що може вказувати на контроверсійність тем.

6. Key Findings & Conclusions

Даний аналіз дозволив дослідити поведінку трендових відео на YouTube та виявити ключові патерни, пов'язані з категоріями контенту, рівнем залученості аудиторії та динамікою трендів у часі.

Основні висновки:

- Категорії **Music** та **Entertainment** домінують за кількістю відео в трендах та часткою відео з дуже високими переглядами, що підтверджує їхню здатність

стабільно створювати вірусний контент.

- Категорії **Gaming** та **Comedy** демонструють високий рівень залученості аудиторії, зокрема за кількістю лайків і коментарів у розрахунку на одне відео.
- Контент у категоріях **Pets & Animals** має найвищий likes-to-dislikes ratio, що свідчить про переважно позитивне сприйняття аудиторією.
- Категорія **News & Politics** характеризується нижчим співвідношенням лайків до дизлайків, що вказує на поляризацію думок та активні дискусії навколо контенту.
- Аналіз у часовому розрізі показав наявність сезонних коливань, зокрема зростання активності в окремих категоріях у зимові місяці.

Загалом, результати аналізу підтверджують, що потрапляння відео у тренди залежить не лише від кількості переглядів, але й від рівня взаємодії аудиторії, характеру контенту та його здатності викликати емоційну реакцію.

Обмеження аналізу:

Даний аналіз базується на даних лише для трендових відео та охоплює обмежений часовий період, тому результати не можуть бути повністю узагальнені на весь контент YouTube.