YouTube Trends Analysis

Ricardo Leal
rickyleal5@gmail.com
GitHub Repo: https://github.com/rickyleal5/youtube-trends
November 8th 2023

Abstract

Analysis of daily top-trending YouTube videos from a dataset on Kaggle.com. As an active user of the platform, I was interested in calculating some statistics about the top-trending videos and curious about what makes them trendy. For these reasons, I made a program to extract, transform, and load the data into a visualization platform.

Introduction

YouTube is one of the most popular video-sharing platforms for content creators. According to Ceci (2023), in "2022, YouTube was among the most downloaded mobile entertainment apps worldwide, with 154 million downloads from global users." Worldwide, the popularity of YouTube as a social media platform where videos are published seems to increase every year. Therefore, an analysis of this dataset was ideal to gain insight into what the trendy videos have in common.

Method

The dataset was acquired from Kaggle, a data science competition platform. A Node.js program streams the data in the dataset files to a TimescaleDB database. The database structure was created using Sequelize models that represent the tables in the database. The data visualizations were displayed on interactive dashboards made on Grafana. Both Grafana and TimescaleDB ran in Docker containers. Unit testing was done using Mocha and Chai and can be run locally or on a CI/CD workflow using GitHub Actions.

Analysis

1. Exploratory Data Analysis

The dataset contains multiple JSON and CSV files, one of each file type per region. The JSON files contain information about the eleven regions and thirty-two categories available per region. However, the categories are repeated in every region except for the United States region which includes an additional "Nonprofits & Activism" category. About each category, only the IDs and titles of them were used. The file names are prefixed with a two-letter region code which is the only information provided about the regions. The CSV files contain the data about the two hundred daily trendy videos for multiple days. Each row of the file has five numeric, two date, two boolean, and seven alphanumeric attributes which are described below.

Attributes of the CSV files in the format (attribute: description. data category):
   • video_id: unique identifier of a video. Alphanumeric
   • title: title of the video. Alphanumeric(includes symbols and emojis)
   • publishedAt: the date when the video was published and available to be viewed. Date
   • channelId: unique identifier of a channel. Alphanumeric
   • channelTitle: title of the video. Alphanumeric
   • categoryId: unique identifier of a category. Number
   • trending_date: the date when the video was trendy. Date
   • tags: tags of a video. Alphanumeric
   • view_count: number of views. Number
   • likes: number of likes. Number
   • dislikes: number of dislikes. Number
   • comment_count: number of comments posted. Number
   • thumbnail_link: URL for the thumbnail image of the video. Alphanumeric
   • comments_disabled: Whether or not viewers can post a comment. Boolean
   • ratings_disabled: Whether or not viewers can like or dislike the video. Boolean
   • description: description of the video. Alphanumeric(includes symbols and emojis)

The program does a basic cleaning before loading the data into the database. It removes the trends where the number of views is negative or zero. In addition, it removes the trends where the number of likes, dislikes, or comments is negative.

   2.   Basic statistics

The dataset only contains trends from August 11th, 2020 through November 6th,2023. All dates are in the Coordinated Universal Time (UTC) standard.

Total number of trends, regions, categories, channels, and videos.

Counts

Trends	Regions	Categories	Channels	Videos
2,569,418	11	32	44,414	444,142

The following statistics are made from the data available when the videos were trendy on a certain date. However, this does not represent the values the videos could have today. A video could be trendy with a relatively low number of views compared to other trendy videos. The so-called "YouTube Algorithm" decides what a trendy video is and based on this data, a trendy video could have low numbers and still be trendy somehow.

Video Averages

Views	Likes	Dislikes	Comments
1,210,847	56,203	757	3,155

Video Maximum

Views	Likes	Dislikes	Comments
296,314,174	16,464,253	12,341,474	6,939,302

Video Minimum

Views	Likes	Dislikes	Comments
7,841	0	0	0

This table shows that it's possible to have a trendy video with a low number of views, likes, and comments, or a high number of dislikes. It's also possible to have a trendy video that does not allow a user to rate or comment on it.

Note: Sheesh! A video indeed has that many dislikes. I looked it up and the video is not showing anymore the number of dislikes but the comments talking about them do not lie.

3. What categories have the most videos?

Top Categories With Most Videos		
Entertainment 25.6%	**People & Blogs** 13.4%	**Gaming** 12.6%

These are estimated values calculated with TimescaleDB hyperfunction mcv_agg() which gets the most common elements of a set and their relative frequencies.

4. What are the most viewed trendy channels and categories worldwide?

Most Viewed Categories

Category	Category ID	Views
Music	10	108,302,363,645
People & Blogs	22	67,141,978,997
Entertainment	24	164,662,853,021

Most Viewed Channels

Channel	Channel ID	Views
T-Series	UCq-Fj5jknLsUf-MWSy4_brA	5,106,791,121
A4	UC2tsySbe9TNrI-xh2IximHA	4,813,891,220
MrBeast	UCX6OQ3DkcsbYNE6H8uQQuVA	4,776,502,511

The most watched category is Music. This makes sense since plenty of artists post their official music videos on the platform. YouTube Records, shows that most records come from music videos. The most viewed channel, T-Series, is India's largest music label and movie studio in the country.

5. What are the most viewed, liked, disliked, and commented videos?

Most Viewed Videos

Video Title	Channel	Channel ID	Category	Category ID	Views
BTS (방탄소년단) 'Butter' Official MV	Big Hit Labels	UC3IZKseVpdzPS...	Music	10	296,314,174
BLACKPINK - 'Pink Venom' M/V	BLACKPINK	UCOmHUn--16B9...	Music	10	277,791,741
BTS (방탄소년단) 'Dynamite' Official MV	Big Hit Labels	UC3IZKseVpdzPS...	Music	10	262,319,276

Most Liked Videos

Video Title	Channel	Channel ID	Category	Category ID	Likes
BTS (방탄소년단) 'Butter' Official MV	Big Hit Labels	UC3IZKseVpdzPS...	Music	10	16,464,253
BTS (방탄소년단) 'Dynamite' Official MV	Big Hit Labels	UC3IZKseVpdzPS...	Music	10	16,254,791
BLACKPINK - 'Pink Venom' M/V	BLACKPINK	UCOmHUn--16B9...	Music	10	12,993,894

Most Disliked Videos					
Video Title	Channel	Channel ID	Category	Category ID	Dislikes
Sadak 2 | Official Trailer | Sanjay | Pooja ...	Star Studios	UCGqvJPRcv7aVF...	Entertainment	24	12,341,474
Khaali Peeli | Teaser | Ishaan | Ananya P...	Zee Studios	UC3jMepkLKF8y4...	Film & Animation	1	1,733,752
Prime Minister Narendra Modi's Mann Ki ...	Bharatiya Janata Party	UCrwE8kVqtIUVU...	News & Politics	25	1,229,004

| Most Commented Videos					
Video Title	Channel	Channel ID	Category	Category ID	Comments
BTS (방탄소년단) 'Butter' Official MV	Big Hit Labels	UC3IZKseVpdzPS...	Music	10	6,939,302
BTS (방탄소년단) 'Dynamite' Official MV	Big Hit Labels	UC3IZKseVpdzPS...	Music	10	6,303,708
BTS (방탄소년단) 'Life Goes On' Official MV	Big Hit Labels	UC3IZKseVpdzPS...	Music	10	4,225,989

The Music category(ID #10) leads again. The most viewed, liked, and commented videos are all K-pop music videos. On the other hand, the most disliked video is the trailer of Sadak 2, a movie produced in India.

6. What channels were the most trendy in 2022 and how 2023 is going so far?

Trendy channels in 2022
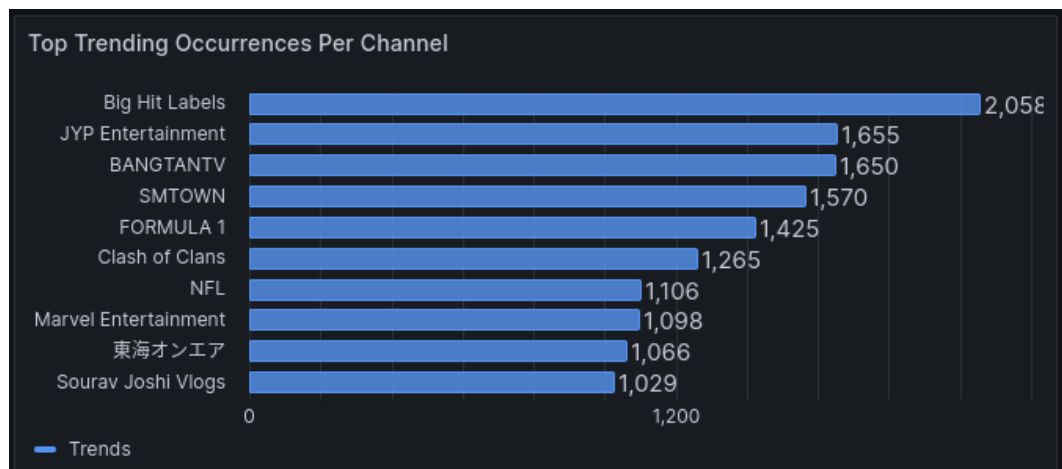

Trendy channels as of November 6th,2023:
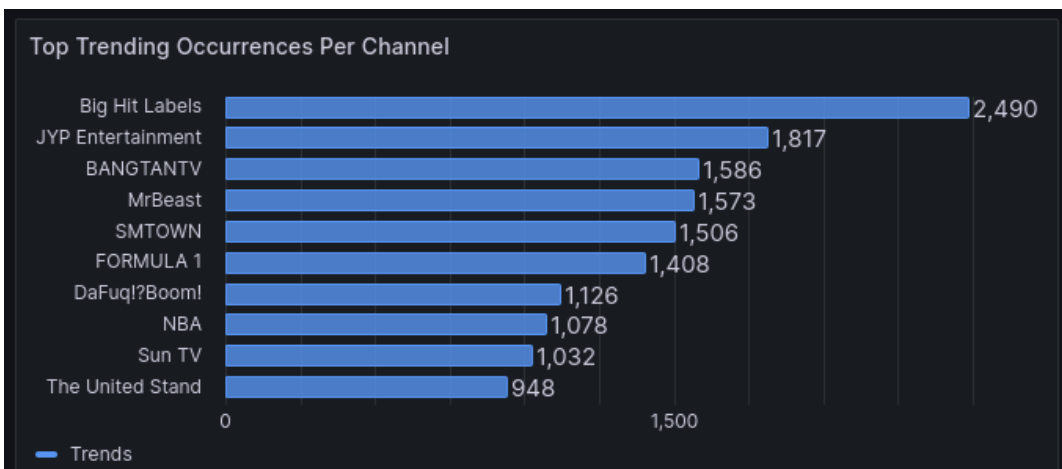
The top three channels so far have stayed in their places. These three mostly post videos in the Music category. Some other channels post videos of Music as well but others about Sports, Entertainment, Gaming, and Video blogging.

7. What is the most popular day trendy videos with the most views get published?
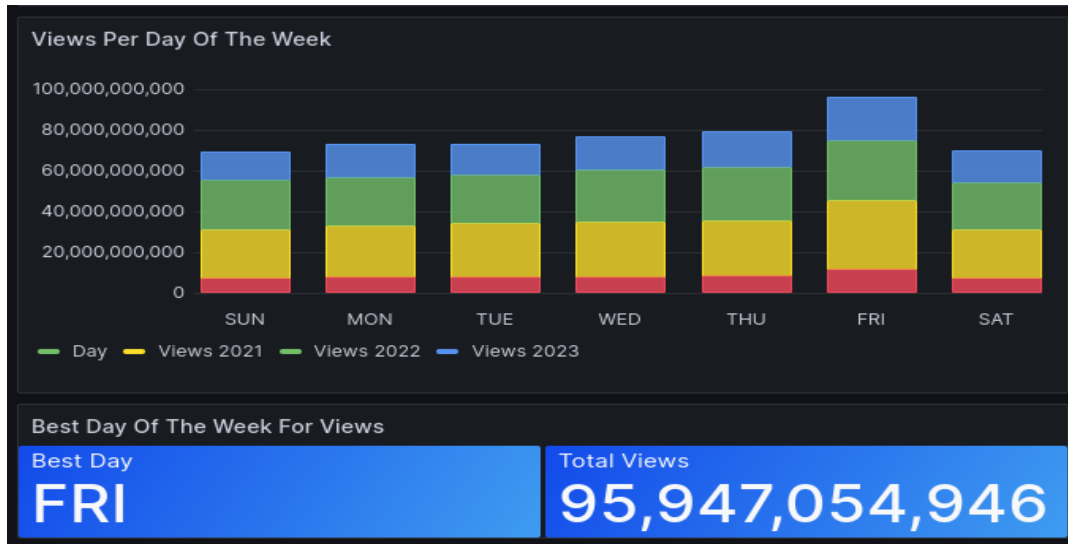
The most popular day is Friday, the day before the weekend starts in most regions.

8. How many views on average and in total does a channel get depending on the number of trendy videos published?
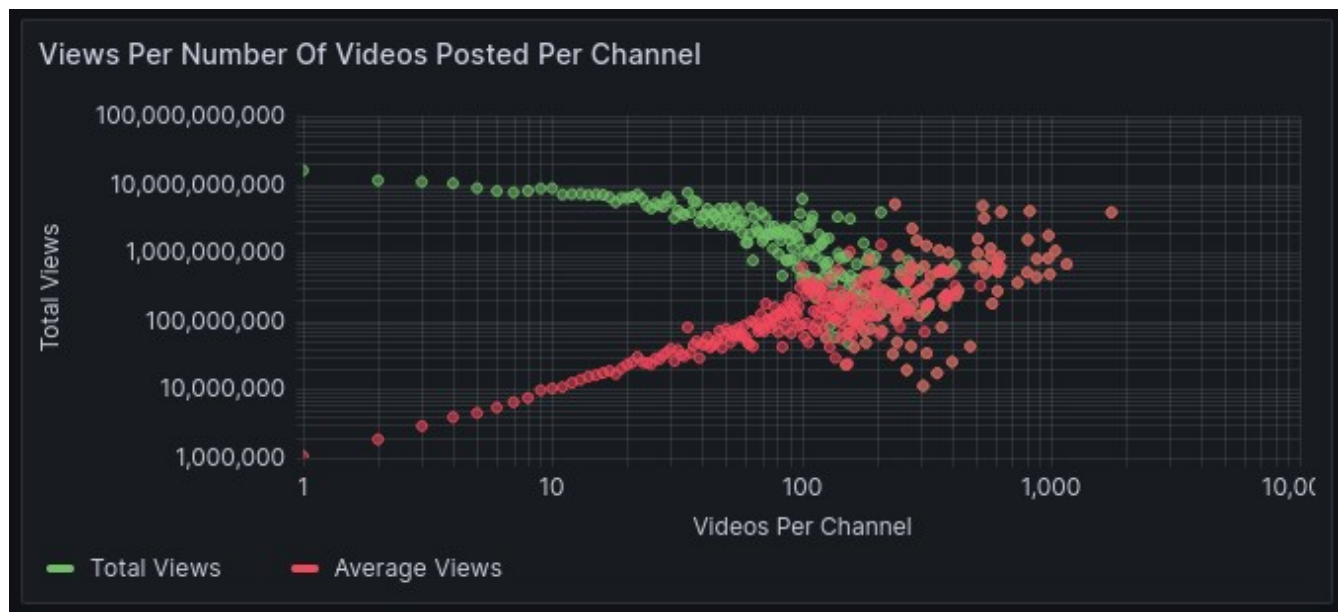
On average, the more trendy videos a channel publishes, the more views the channel gets. However, the number of views added by channels with lower number of trendy videos is higher than the number of views added by channels with more trendy videos.

Conclusion

Many factors influence how the so-called "YouTube Algorithm" decides what a trendy video is and it is not only a matter of getting the most views. It is possible to have a trendy video even without a high number of views or likes which is likely what most content creators are looking for. The Music and Entertainment categories are the most trendy among all the categories. Based on the data, the best day to publish trendy videos is Friday or at least before the weekend starts. Most channels have few trendy videos but this should not stop them from continuing to publish more videos on the platform.

The dataset is good for getting an insight into trendy videos. However, if it had the number of followers, the number of videos published per channel, and the data of the videos when they were not trending on a date, a better insight into the videos could be gained.

References

Sharma, R. (2023, November). YouTube Trending Video Dataset (updated daily), Version 1186. Retrieved November 7, 2023 from https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset/versions/1186.

Ceci, L. (2023, October). Countries with the most YouTube users 2023. Retrieved November 7, 2023 from https://www.statista.com/statistics/280685/number-of-monthly-unique-youtube-users.