Tanzim Amin
San Jose State University
ISDA 111 - Info/Data Sci
Professor Gupta

YouTube Sentimental Data Analysis

3rd December 2022

OVERVIEW

With 2 billion monthly active users, YouTube is the second most widely used social media network and, in terms of total users and page views, the second largest website in the world, following behind Google. YouTube videos offer a plethora of information and useful insights for any company, and YouTube viewers are increasing daily. Playing video games alone and then witnessing others show off their gaming skills naturally complement one another. Consequently, it is not surprising to discover a large number of gaming influencers on YouTube. When you have the appropriate tools in place, you can easily take advantage of all those unreserved thoughts and unfiltered client feedback. This report will use the "YouTube Data API request" dataset from YouTube API for Developers Reference (https://developers.google.com/youtube/v3). Utilizing the YouTube Data API, I will be able to add features that are typically performed on the YouTube website to my own website or application. There are many resources that I can obtain using the API that are included on the reference sheet. In this project, I will be building a Python Project to Scrape YouTube data using YouTube Data API. Using that particular YouTube API, I then will be extracting the data and then loading this data into a Python Pandas DataFrame and then will proceed to analyze this data. Finally, from all that I will build a simple visualization from this data using the Python Seaborn library.

BRAINSTORM

In order to access YouTube data for my project, I must first create a YouTube API Key. To retrieve the API key, I need to go to the Google developers page, where a reference sheet lies with the information needed. I will figure out how to utilize the created API key to access various YouTube data after it has been generated. Then I can use the Google documentation for using the YouTube API as an example. To gather the different data I need to develop the system, I will look at the various portions of the documentation. Lastly, I'll also take a look at the sample Python code provided by Google to call various resources and strategies to get data from YouTube.

```
# Rough draft planning stage – created by Tanzim Amin
* i'll be using Jupyter Notebook to write my python code.
* new virtual environment Anaconda
* first create a YouTube API Key which will be our credential
to access youtube data.
* first part: extract channel details
* second part: extract video details from a particular
channel.
* analyze all the data by loading it into a pandas data frame.


* then create some simple visualization using the Seaborn
python library.
```

Figure 1 - Planning Phase (Source: Tanzim Amin)

Now, I can begin developing the Python code necessary to construct this project. To write the Python code, I will use the Jupyter Notebook. By using Anaconda, I will construct a fresh virtual environment for this project. After creating the virtual environment, I will install all necessary Python packages. So next, I will install "google-api-python-client" (which is the google python package required to access youtube api data), I already have pandas and seaborn installed.

```
(base) C:\Users\tanzi>conda install -c conda-forge google-api-python-client
Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.

Retrieving notices: ...working... done
```

Figure 2 - installed Google API client through Anaconda terminal (Source: Tanzim Amin)

For the first section, I will take information about YouTube channels. In other words, I will gather information on each channel's name, total number of subscribers, views, and

videos posted. I will collect these facts for a select few channels that cater to a particular niche which can be related to gaming or family friendly content, and then I can compare the channel data to one another. I will check out which channels have the most subscribers, the most views, and the number of videos they upload. All of this data will be loaded into a pandas dataframe before being analyzed. Using this data, I will also create some simple visualizations that will allow me to compare these various channels with ease.

In the second section, I will construct a logic to extract video information from a particular channel. I'll pull information like the video title, the total number of views each video has received, the total number of likes, dislikes, and comments each video has had. I will gather this information for each video that a certain channel has released. The data will subsequently be loaded into a pandas dataframe for analysis. In the end, I'll use the Python Seaborn package to do some straightforward visualization.

	Channel_name	Subscribers	Views	Total_videos	playlist_id
0	elrubiusOMG	40400000	7435947593	704	UUXazgXDIYyWH-yXLAkcrFxw
1	VEGETTA777	33000000	14757045945	7031	UUam8T03EOFBsNdR0thrFHdQ
2	Fernanfloo	44900000	9864867755	541	UUV4xOVpbcV8SdueDCOxLXtQ
3	PewDiePie	111000000	28292462097	4476	UU-lHJZR3Gqxm24_Vd_AJ5Yw
4	JuegaGerman	45600000	13459125807	1929	UUYiGq8XF7YQD00x7wAd62Zg
5	Dream	29100000	2573131049	109	UUTkXRDQl0luXxVQrRQvWS6w

Figure 3

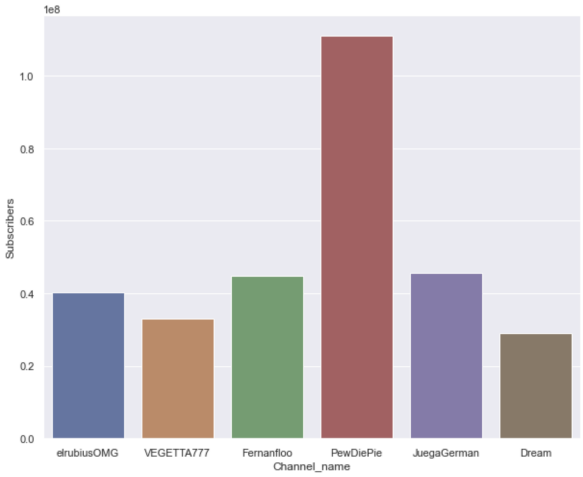

Figure 4

Figure 3 & 4 - Top gaming channels on YouTube (Source: Tanzim Amin)

On YouTube, a substantial genre is gaming. Playing video games yourself and then watching others show off their gaming skills naturally complement one another. Therefore, it is not surprising to discover a large number of gaming influencers on YouTube. Gaming channel content producers are working hard in order to create worlds, connections, and friendships to support their goals, with many entering into long-term commitments. In Figure 3, you can see who the top 6 gaming channels are on the video-sharing platform, with the number of subscribers and views they receive in systematic order. Figure 4 translates to a histogram, showing which of the six channels has the highest number of subscribers. Furthermore, I could have extended the results to

capture more than six channels but wanted to keep the data population low since the next section will go more in-depth providing comprehensive analysis for one of the channels.

Figure 5

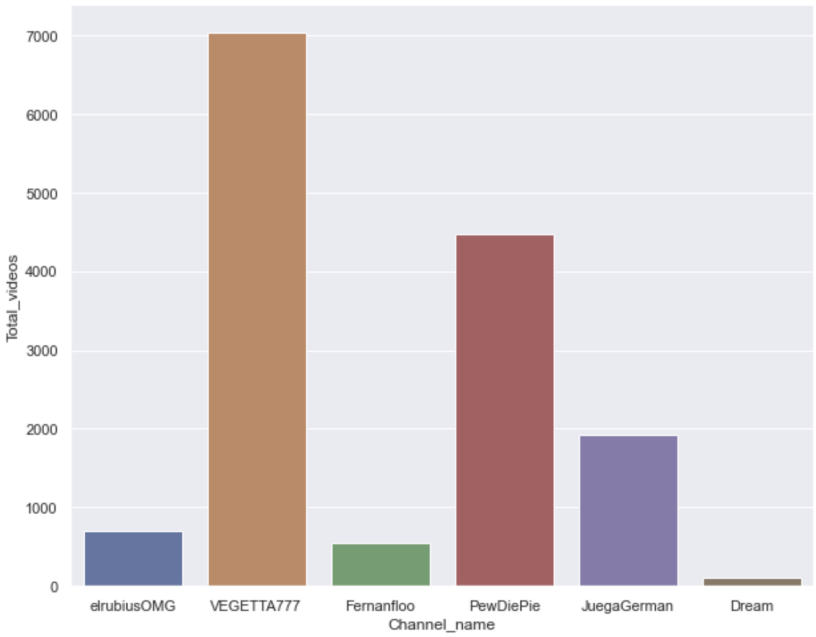

Figure 5 - Channel with the most videos posted (Source: Tanzim Amin)

It is not easy to upload a high quantity of videos in a set timeframe, while maintaining exceptional video quality. The video-sharing platform is flooded with Youtubers, someone who is an influencer producing videos on the YouTube platform, typically posting on a consistent basis to attract a vast audience. However, one specific Youtuber, VEGETTA777 seems to dominate the other Youtubers in terms of how many videos he uploads. As shown in Figure 5, VEGETTA777 has a staggering amount of uploads in total. This signifies his passion and dedication for pushing out videos consistently.

SCOPE

Ideally, a user can earn money on YouTube by applying for and being accepted to the YouTube Partner Program. In addition to subscription and view count limitations, each element has its own specific eligibility requirements. Specific functionalities might not be accessible if our reviewers decide that their channel or video is not qualified. There are two basic justifications for these higher criteria. The most significant one is that in every

location where the function is offered, YouTube must adhere to legal restrictions. Then, YouTube ensures that they have enough context on your channel in order to reward good creators. With VEGETTA777's astounding success on the platform, cashing in a large amount of revenue would come to no surprise.

For a considerably relevant channel, like him, it would be useful for him to have an analytical reference chart to measure the immense success/growth of his channel. I can gather details like the title of each video, how many times it has been seen overall, as well as how many likes and comments each video has gotten. In order to perform the analysis, the data will next be put into a pandas dataframe. Then using the Python Seaborn data visualization library, based on matplotlib, I will create this all into a stunning visualization which can be used as a reference to measure the level of success for his channel.

EXECUTION

Extracting video details from the channel: ***VEGETTA777***

Refer to YT-envProject .ipynb file and .csv file

Note: Youtube Removed the Dislike Count in late-2021:
https://www.youtube.com/watch?v=kxOuG8jMIgI&ab_channel=YouTubeCreators

Figure 6

	Title	Published_date	Views	Likes	Comments
0	Marvel's Midnight Suns: Dia de Super Héroes!	2022-12-02	100169	10543	75
1	KARMALAND 5: Mundial en el Coliseo!	2022-12-01	369233	37478	385
2	Vegetta777 vuelve a BATTLEFIELD 2042 *GRATIS* 😱	2022-11-30	118163	14775	290
3	God of War Ragnarök: Rescatando al VERDADERO T...	2022-11-29	275328	23275	147
4	KARMALAND 5: Mi refugio de gatos ¿FUTURO FARO?	2022-11-28	318872	29761	197
...	...	...	...	...	...
7247	ZONA DEMO: SSX	2012-02-28	103480	3309	305
7248	UNCHARTED 3: LA DIOSA DE CHUECA CON PACOSTRATO...	2012-02-27	134119	4316	287
7249	MW3: NADA ESTA PERDIDO HASTA EL FINAL! CON DEF...	2012-02-26	113769	4648	397
7250	MW3: LIVE RAPIDITO DE MW3 21 - 0	2012-02-23	276796	12979	1540
7251	UNCHARTED 3: PRESENTACION DEL CANAL!	2012-02-21	3366899	215375	29880

7252 rows × 5 columns

Figure 6 - All VEGETTA777'S videos and statistical information (Source: Tanzim Amin)

From the data set in Figure 6, it consists of 6 columns and more than 7,000 rows. The first column represents the index, followed by the Video Title, then the date of when that video was published. Then it shows the total number of views that particular video received, and shows how many people used the like feature, a button signifying how many people like the video and lastly, then number of comments the video has received typically revolving around positive feedback to some brutal criticism.

Figure 7

	Title	Published_date	Views	Likes	Comments
5218	PLANETA VEGETTA: UNA GRAN AVENTURA NOS ESPERA #1	2015-01-24	72363205	500016	50886
5030	GTA V ONLINE: CARRERAS ENTRE AMIGOS-ENEMIGOS	2015-04-29	50725354	653051	17128
7052	ESPECIAL 7777 - MINECRAFT: EL GRAN COMBATE	2012-09-21	39007551	447914	36752
6503	ESPECIAL 777.777 EN PLANETA VEGETTA: LA GRAN V...	2013-07-23	31907717	440907	35129
5962	ESCAPA DE LA BESTIA #2 | OLIMPOCRAFT	2014-03-28	29087675	262011	11128
4577	GTA V ONLINE: YATE WIGETTA Y NUEVO APARTAMENTO	2015-12-15	27748559	358017	16656
3105	SOY UN NIÑO - HELLO NEIGHBOR (JUEGO FINAL) #1	2017-12-10	25877317	588528	14718
2548	SPIDERMAN - COMIENZA LA AVENTURA #1	2018-09-06	21839080	589307	18153
5894	PLANETA VEGETTA: EL DIA DE MI BODA #70	2014-05-01	21817206	308503	24743
3497	GTA V - LADRONES EN PRIMERA PERSONA!	2017-05-27	21337217	356588	14184

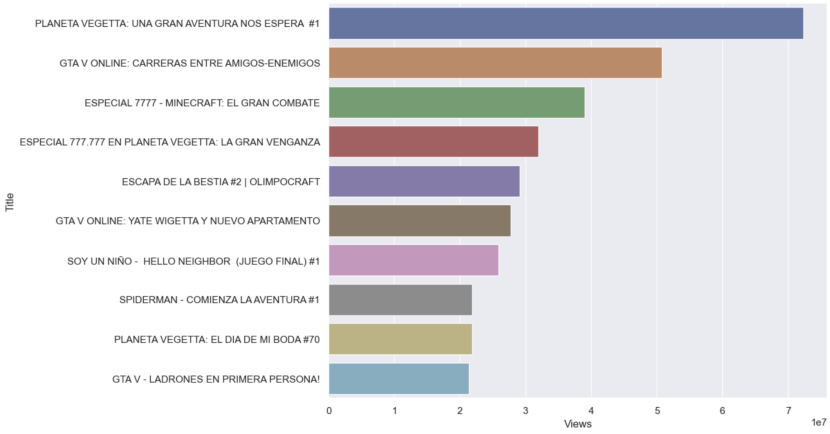

Figure 8

Figure 7 & 8 - VEGETTA777'S top 10 most relevant videos (Source: Tanzim Amin)

Here, I organized the data set to show the top 10 most popular or relevant videos from his channel. As you can see in Figure 7, which contains a horizontal bar chart, nicely color-coded to differentiate the video titles and give someone looking at a title, a glimpse of what video is trending.

Additionally, YouTube already has a method for a user to look at what the creator's most popular videos are, but that does not really provide an in-depth breakdown of the statistics. For example, in Figure 9, if someone goes to that channel, they can filter the videos to "Popular" and see this on their screen. This render lacks a lot of

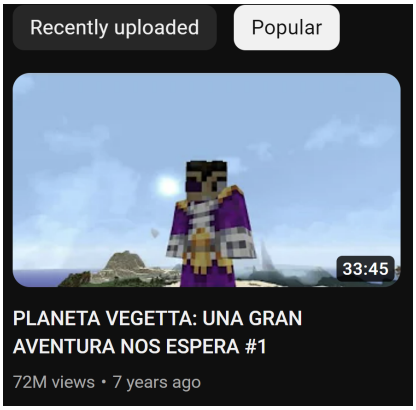

Figure 9 - VEGETTA777'S most popular video (Source: youtube.com/vegetta777)

critical information regarding the total number of views, published date, likes, and comments.

Now if the video creator wants to see what month in the previous year he uploaded the most videos, he can do that as well, along with figuring out which month got the most views. With this information, the creator can make a hypothesis on which month to push out the more videos in the following year. This can lead to more ad sense revenue for the creator as well as more exposure for his channel.

Figure 10

	Month	size
0	Apr	585
1	Aug	630
2	Dec	625
3	Feb	543
4	Jan	607
5	Jul	616
6	Jun	556
7	Mar	617
8	May	577
9	Nov	625
10	Oct	644
11	Sep	627

Figure 11

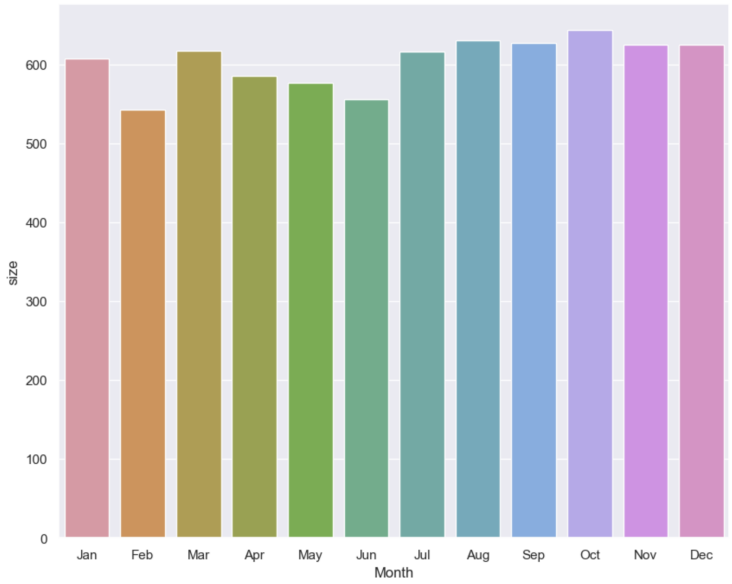

Figure 10 & 11 - VEGETTA777'S most videos uploaded/month

(Source: Tanzim Amin)

As you can see in Figure 11, VEGETTA777 produced the most videos in October and the least videos in February. Staying consistent with uploading videos is one factor to being successful on the platform. It involves performing the same tasks again and again, receiving positive and negative feedback from the viewers, and making the appropriate adjustments to the video content in order to help make it thrive.

ANALYSIS

Ultimately, with the help of important indicators and reports in this data visualization, you can utilize analytics to better understand the success of your videos and channels. Also, the channel is doing good in terms of views, likes and comments this year so I think they can manage the channel with their gaming content, rather than shifting to a different

niche. Using all the factors and filters, YouTube shows trending videos and updates this list every 15 minutes.

There was not much raw data in this dataset to begin with, so not much cleaning was needed. Though, this made it a little more convenient to read and analyze. After exploring the dataset and visually analyzing the different relationships among time, subscribers, views, likes, and comments. Thankfully, this can be done for any channel and one can download any channel's whole data by using their API in a nonmalicious manner, and using it for the purpose of analyzing the data. Essentially, YouTube has various methods to analyze all the channels on their platform using their internal mechanism. This is used for the algorithm, to push certain videos out to more people and increase engagement so people are hooked.

QUESTION AND ANSWER

1. **How can we utilize this data to produce videos that will reach a bigger audience?**

   I believe that it all comes down to using the metrics to analyze why those uploaded videos performed so well, as opposed to the user's other videos. For example, it can be based on the time of day to the YouTube "adpocalypse", which is a site-wide term coined to describe the mass advertiser boycotts and withdrawals on YouTube. You can also get in-depth information about certain trends and what your audience may want to see and how they connect with your channel by using a data-driven video marketing plan.

2. **How can a user integrate Youtube Analytics into the overall marketing strategy?**

   I suggest trying to treat YouTube more like a traffic source than a growth platform if you want to use YouTube videos as promotional content. Keep in mind that YouTube has a large number of videos, so anything you select to start with probably already has competitors. You can attract viewers and keep them on your YouTube channel by using key data in YouTube Analytics. By using YouTube Analytics data properly, you may significantly increase the amount of people who watch your YouTube videos.

3. **By using the data, can advertising a YouTube video be worth it?**

   YouTube is not always the preferred marketing platform for companies. It is thought of being a costly platform that needs investment and is better suited for large businesses. Choosing an advertising firm with YouTube expertise is advisable if you have made the decision to use YouTube as a marketing channel. Try requesting your

YouTube Analytics on a frequent basis or to switch from YouTube Studio to other, more illuminating data solutions.

4. **Would it be more practical for a YouTuber to resort to Whatagraph as opposed to using the Python dataset?**

   Whatagraph is a cross-channel reporting platform that assists companies in monitoring, evaluating, and discussing marketing effectiveness with teams and clients. Whatagraph offers automated integrations with YouTube and YouTube Ads. All essential data flows into Whatagraph's user-friendly analytics tool and lets you create detailed reports. Now this might seem as a more viable choice to go with, however Whatagraph application paid service with premium rates ranging up to $300 per month, which might not be very striking to those content creators struggling financially.

5. **Does it take a certain amount of hours of watch-time before a video is recommended by YouTube's algorithm?**

   I believe most people don't choose what they want to watch depending on when a video was uploaded or watch videos in the most current sequence. Because it's normal for people to exhibit interest in older videos, channels may realize that some of their videos are gaining popularity months after they were first uploaded. This can be the result of a certain topic becoming more popular or the fact that new channel subscribers are revisiting older recordings. As a result, films that were posted weeks, months, or even years earlier are frequently seen on a user's home page.

SUMMARY

To summarize everything, for this project we've done two things that can be quite beneficial for the YouTube user to know. For the first section, we took information about YouTube channels such as gathering information on each channel's name, total number of subscribers, views, and videos posted. In the second section, we constructed a logic to extract video information from a particular channel, in the example I used the YouTuber: VEGETTA777, ideally the information retrieved ranged from the video title, the total number of views each video has received, the total number of likes, dislikes, and comments each video has had. Finally, we executed a few simple visualization graphs using the Python Seaborn module.

CONCLUSION

The rise of social media platforms has also drawn users to video-sharing websites like YouTube. On these websites, viewers of videos can post their thoughts or feelings in the comments section. Everyone is welcome to share their thoughts in relation to the current YouTube discussions. As a result, everyone is free to share what they think about the performance. From this analysis, I learned that there are many metrics that go behind a YouTube video's success and those metrics can be calculated using data visualization in Python. There is a necessity to do an analysis on opinion mining due to the rise of numerous critiques that surface in a short period of time. This could be further implemented into a sentiment analysis through machine learning, which is the technique of looking for patterns or societal moods in natural language in relation to certain goods. But I am not at that level of competence yet, so until then I feel like this project is a great foundation for the start of something new.

LINKS

I posted this project on my Github, check it out here:

https://github.com/Tanzaik/YouTube-Sentimental-Data-Analysis-2