INTRODUCTION

Aims and objectives of this project
Within this project, I would like to explore the following:

Getting to know Youtube API and how to obtain video data.
Analyzing video data and verifying different common "myths" about what makes a video do well on Youtube, for example:
Does the number of likes and comments matter for a video to get more views?
Does the video duration matter for views and interaction (likes/ comments)?

From the creator I take into consideration, how often they upload new videos?. On which days in the week? What are the trends in viewership? And so on.

JAMES SCHOLZ has 3 youtube channels and we can analyze each channel.
The analysis consists of two parts;

- How does each channel fare compared to the other channels
- Analysis within the channel i.e Trends and video statistics

THE ANALYSIS PROCESS

- ❖ Data Retrieval from the youtube API
- ❖ Data Frame Creation
- ❖ Data Normalization
- ❖ Analysis

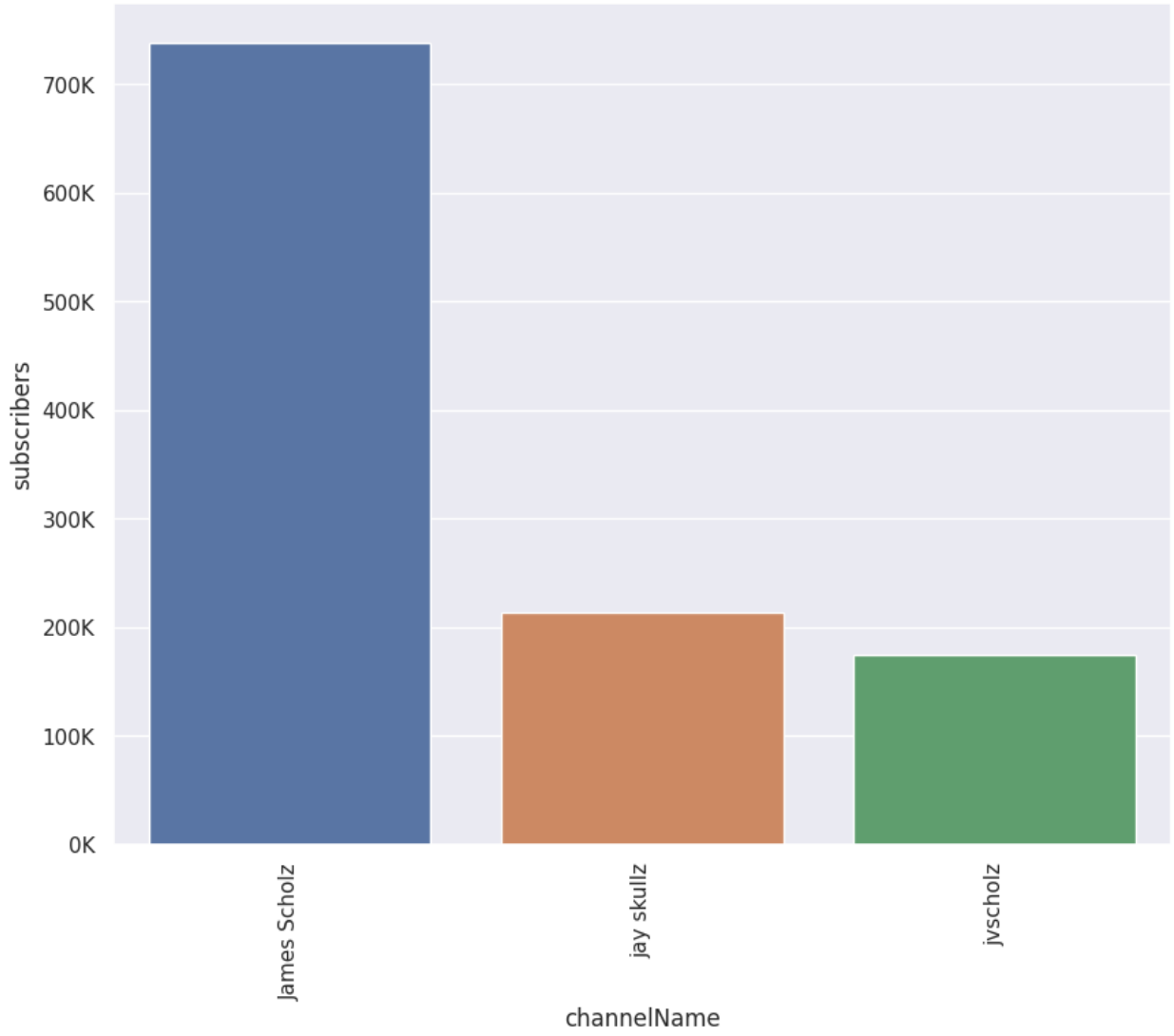
You can find the full analysis process including the analysis of the 3 youtube channels on my GitHub;
https://github.com/CYBERIC-EA/JAMES-S-YOUTUBE-CHANNEL-ANALYSIS

CHANNEL COMPARISON ANALYSIS

1. SUBSCRIBERS PER CHANNEL

We can compare the suscribers of different channels by updating the channel id list.

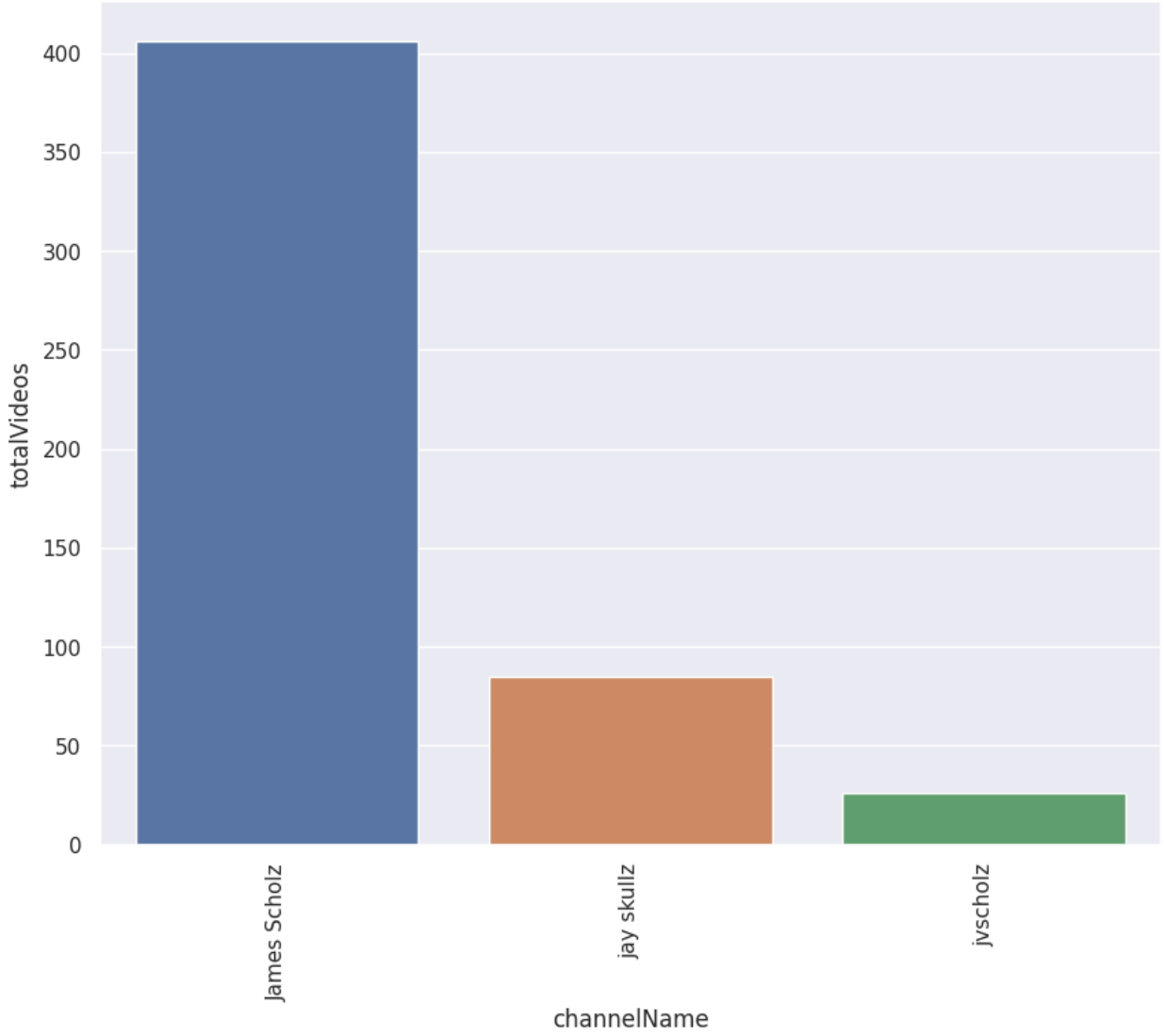
Let's take a look at the number of subscribers per channel to have a view of how popular the channels are when compared with one another.



2. VIDEOS PER CHANNEL

We can compare the total number of uploaded videos of different channels by updating the channel id list.

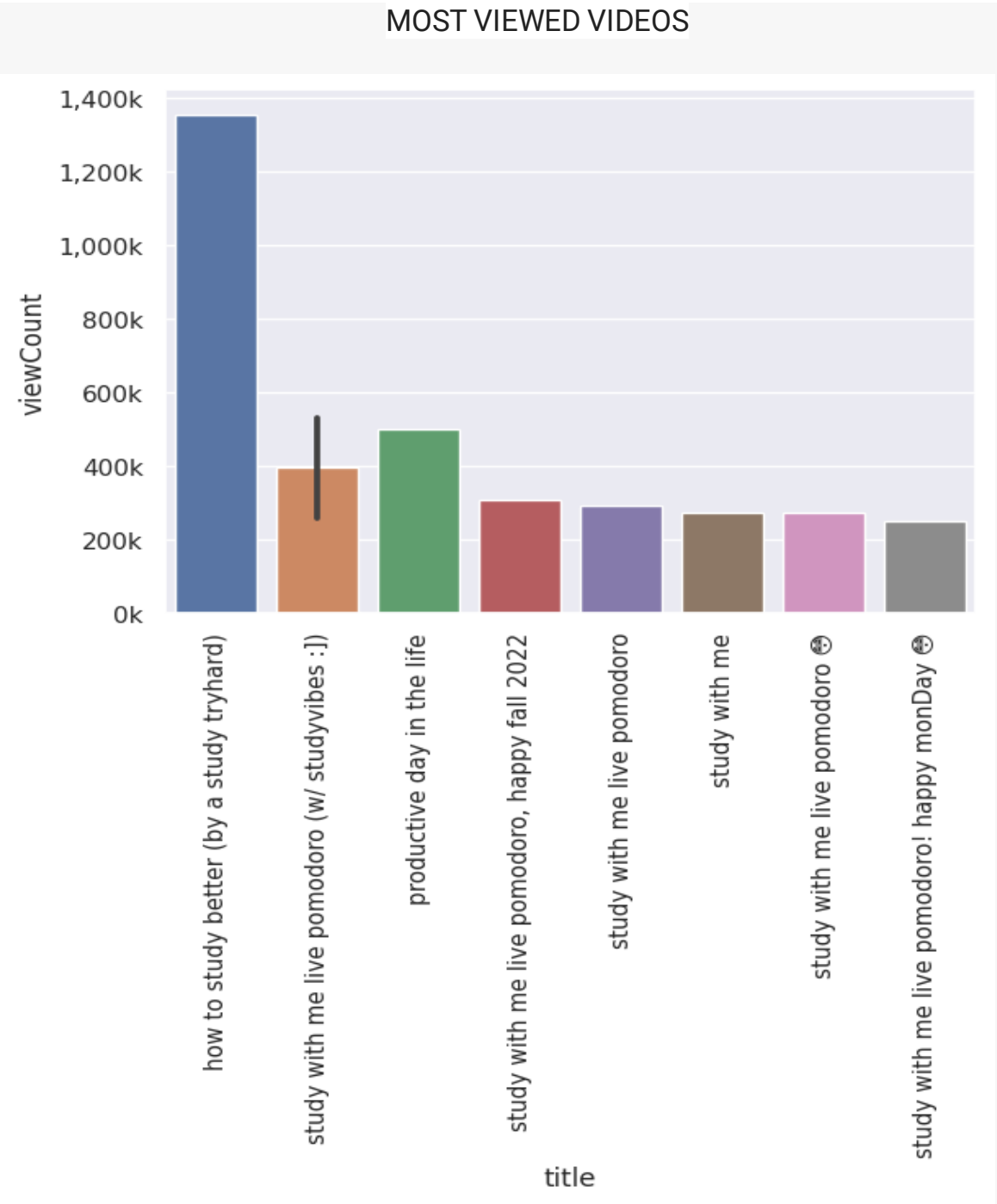
Let's take a look at the number of subscribers per channel to have a view of how popular the channels are when compared with one another.

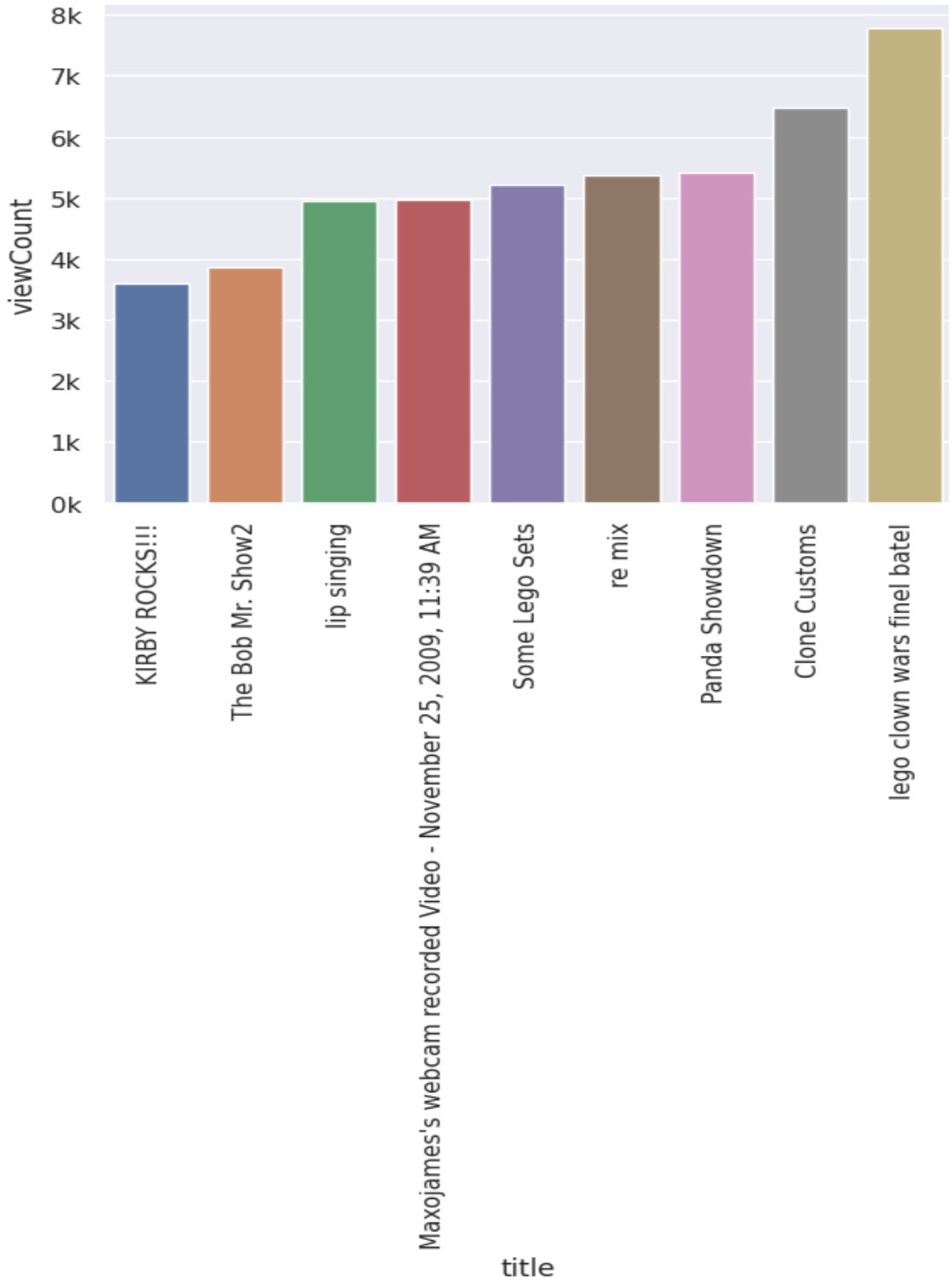


Based on the two visualizations above you can deduce that there is a correlation between the amount of videos uploaded and the number of suscribers the channel has.

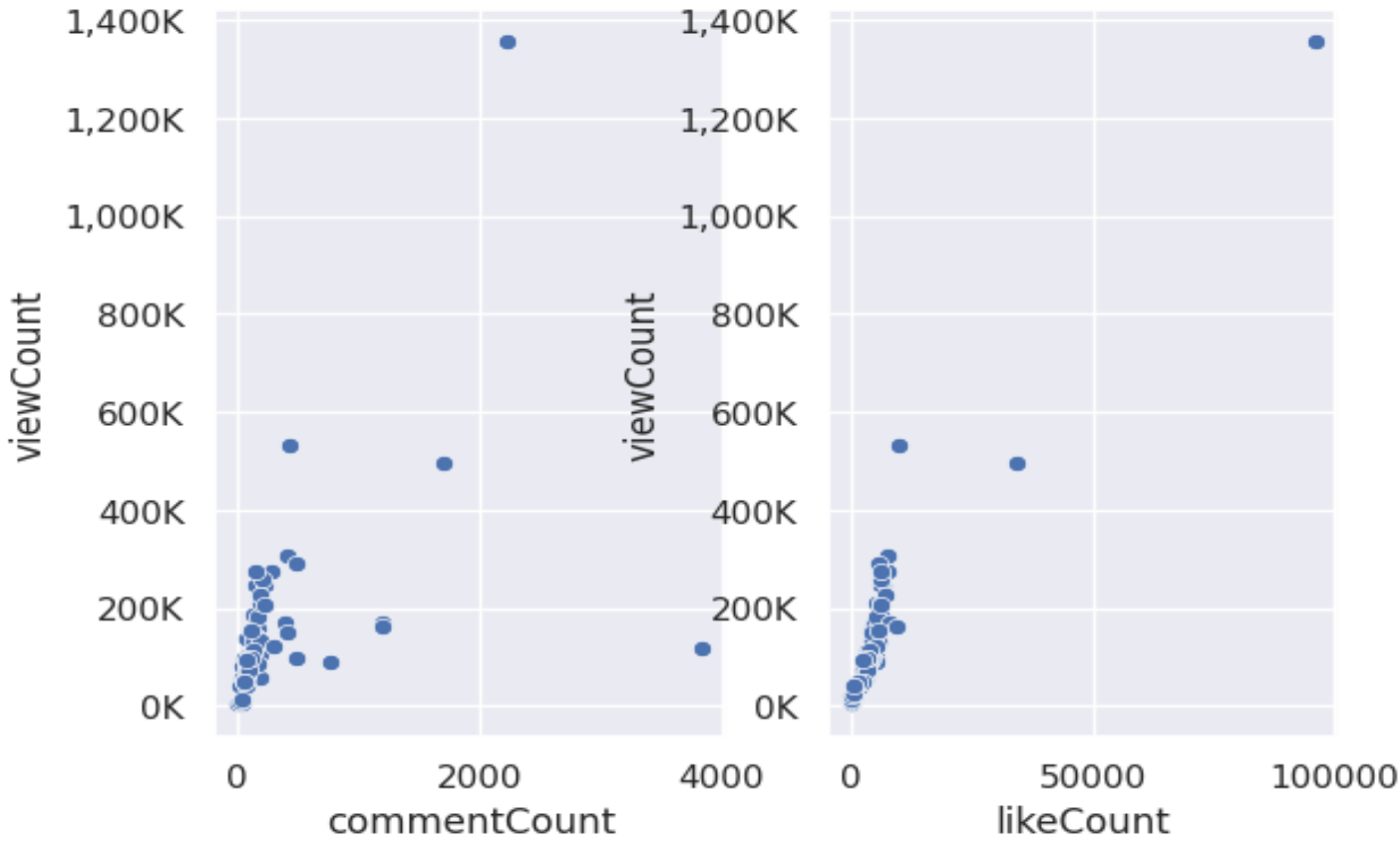
jay skullz channel analysis

1. VIEWS PER VIDEO (Most and Least Viewed Videos)




LEAST VIEWED VIDEOS
viewCount
8k
7k
6k
5k
4k
3k
2k
1k
0k
KIRBY ROCKS!!!
The Bob Mr. Show2
lip singing
Maxojames's webcam recorded Video - November 25, 2009, 11:39 AM
Some Lego Sets
re mix
Panda Showdown
Clone Customs
lego clown wars finel batel
title

2. Does the number of likes and comments matter for a video to get more views? Firstly, I would like to check if comments and likes do correlate with how many views a video would get. In the plots below, it can be observed that the number of views and number of comments/ likes strongly correlated with each other. The number of likes seems to suggest stronger correlation than the number of comments. However, this is expected as the more people watching a video, the more likely this video will get comments and likes. To correct for this factor, we will plot these relationships again using the comments per 1000 view and likes per 1000 view ratios.

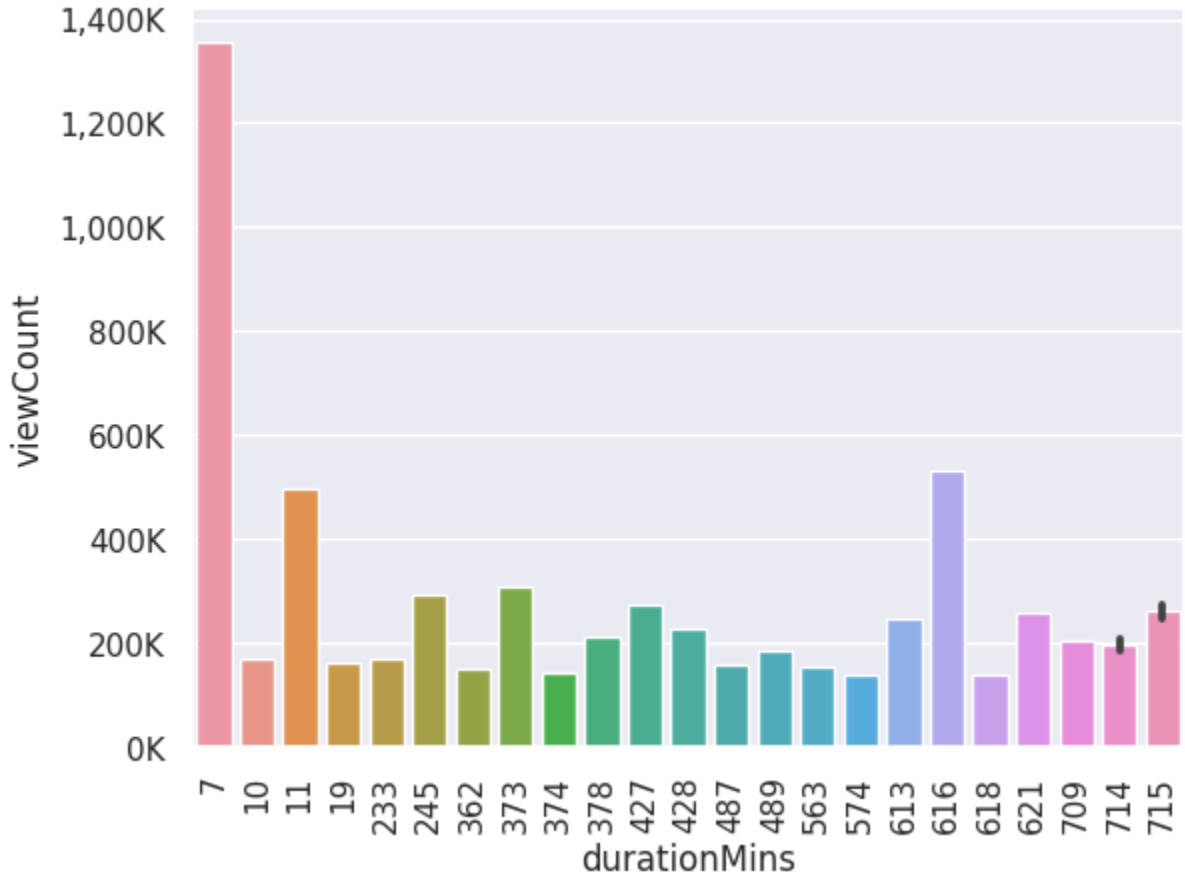


After correcting for the absolute number of views, it turns out that the correlation is much less clear. The comment-view relationship seems to completely disappear: a lot of videos have thousands of views and very few comments, while some videos have very few views have better interaction. However, it is understandable that comments take more effort than views and likes, and normally comments would die off when the video gets older.

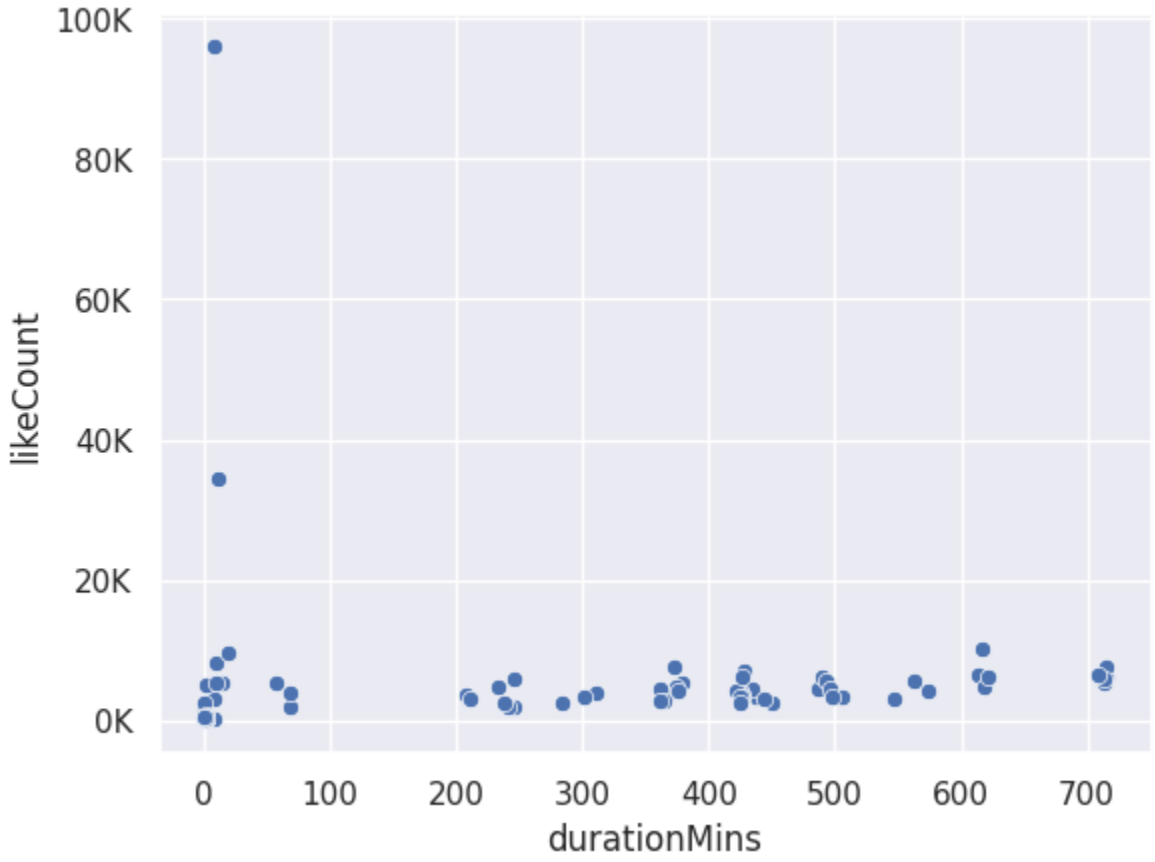
As for the like-view relationship, we can still see some positive correlation between views and like ratio (though very subtle), which means that the more views a video has, the more people would hit the like button! This seems to support the idea of social proof,

which means that people tend to like better the products that are already liked by many other people.

3. Does the video duration matter for views and interaction (views/likes/ comments)? As can be seen in the histogram below, most videos are between 7 to 20 minutes. In this case the videos that are above 200 minutes should be considered as streamed videos.



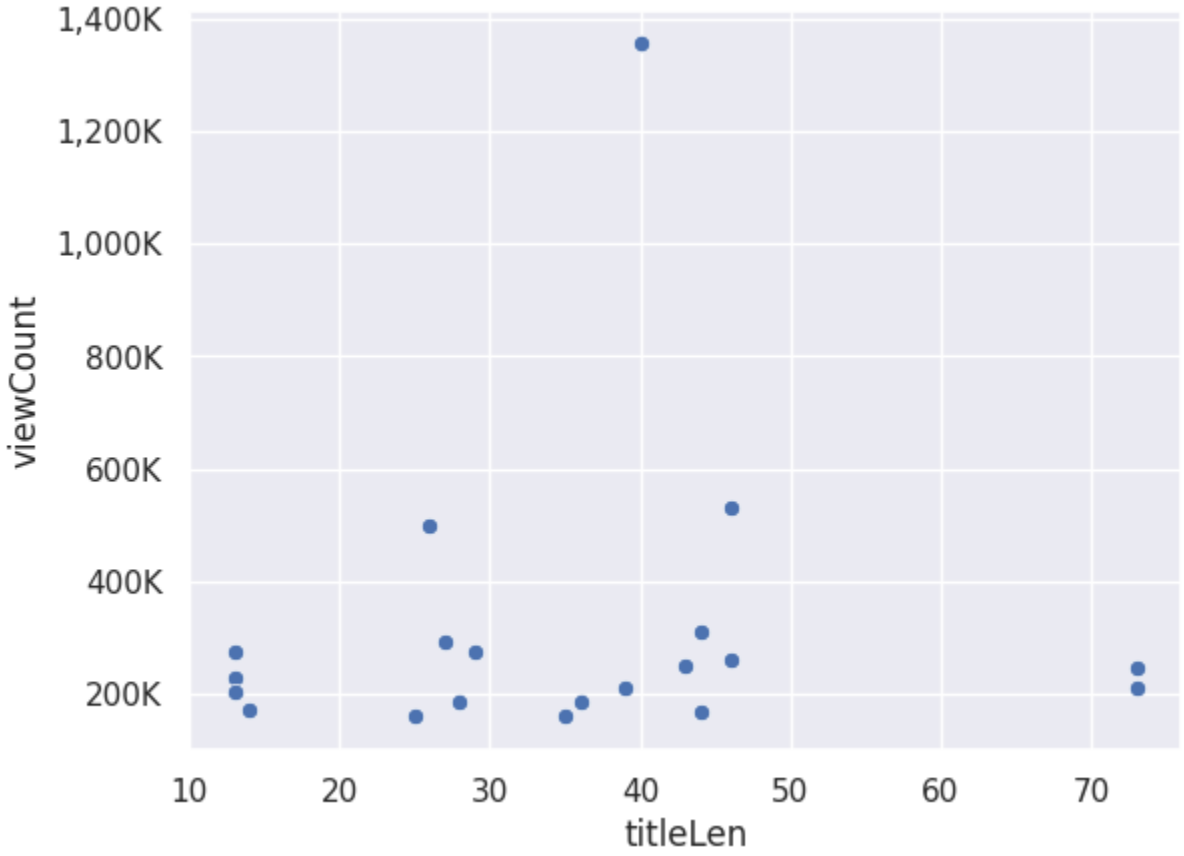
4. VIDEO DURATION AGAINST LIKE COUNT



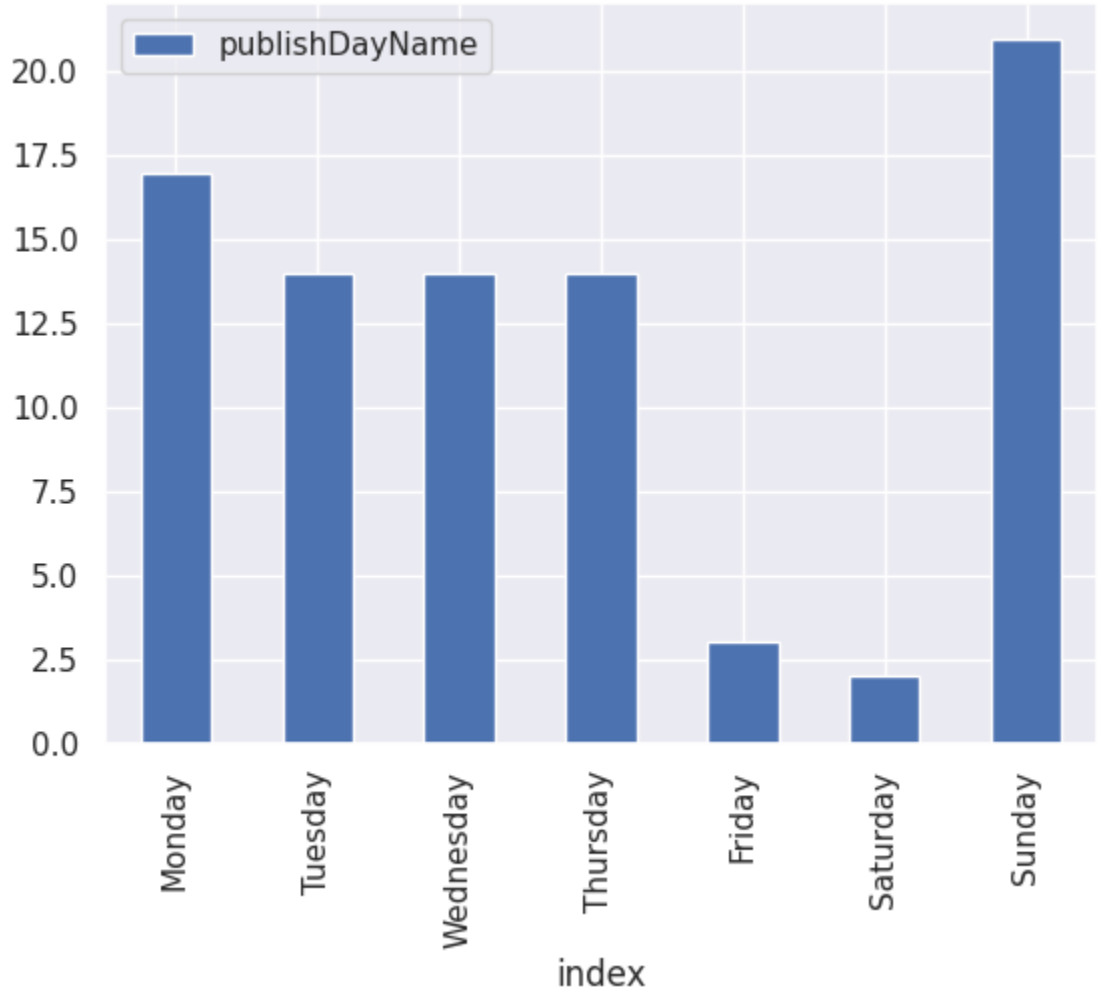
The duration of the video actually impacts the number of views an comments on the videos. This may be due to the fact that the videos which are long i.e streams are;

1. Not rewatched often as compared to shorter videos
2. Not commented on as the people may be reading along
3. Missed by the viewers

5. How does the length of the title affect the amount of views the person gets?

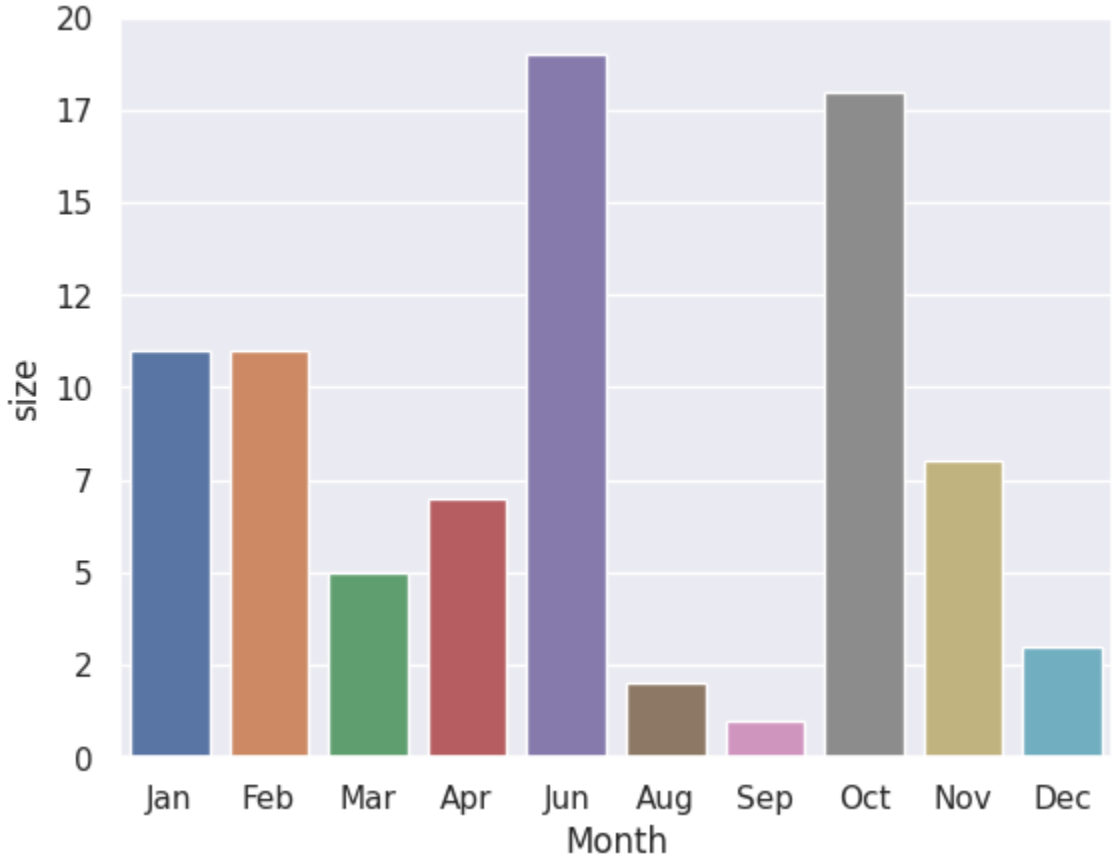


6. UPLOAD PATTERNS AND HOW THEY AFFECT THE CHANNEL



It's interesting to see that more videos are uploaded on Sundays. More videos are uploaded on Sunday compared to the other days. This could be because of the nature of the niche, which heads towards productivity and studying. Although the videos can be consumed at any time, It could mean that Sunday is when the creator and potentially the viewers have the most free time. It could also just means that he begins working on his videos during the weekend(i.e Friday or Saturday) with the hope of uploading it later.

7. We can also take a look at Monthly Upload Schedule.

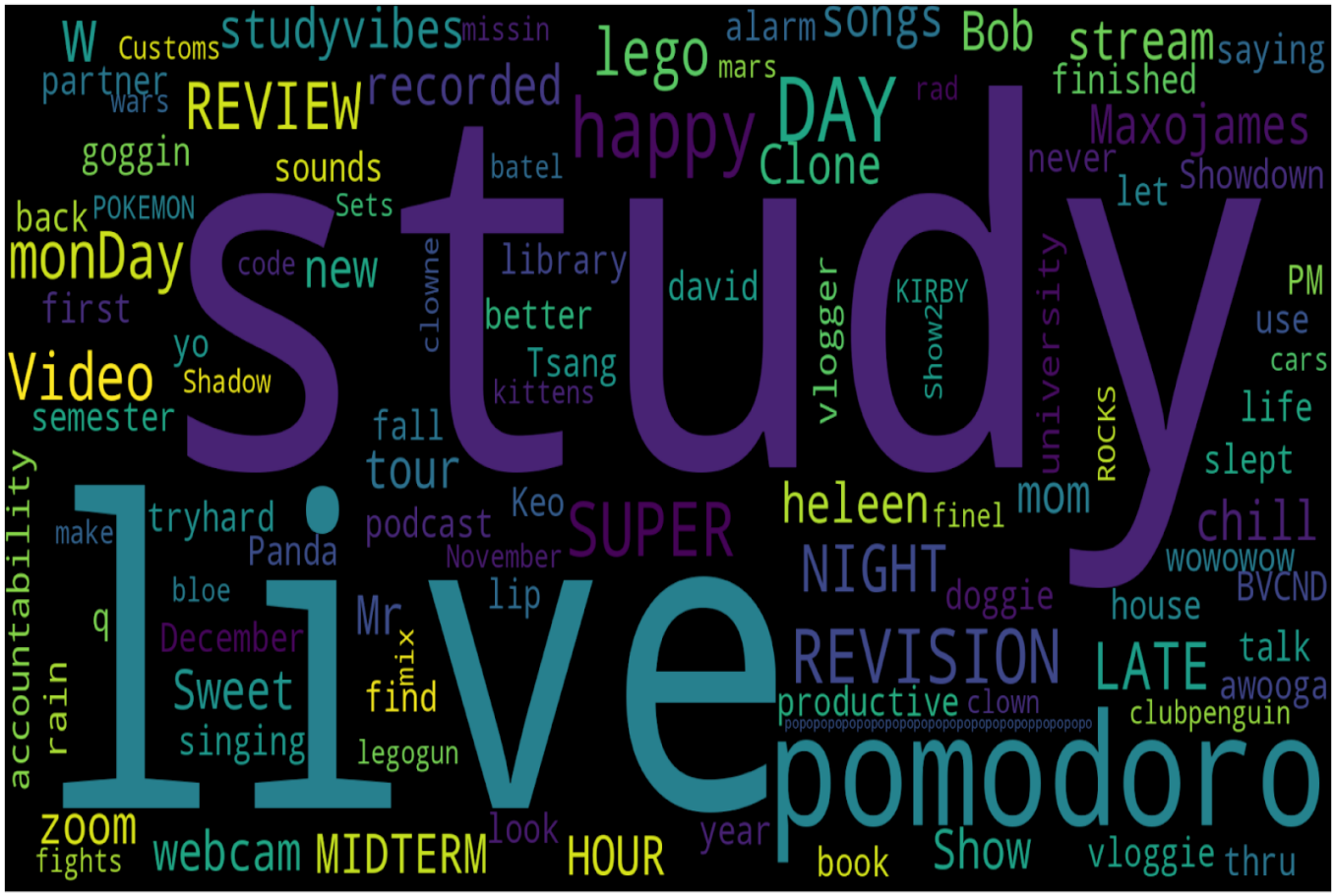


Although videos are uploaded every month, there seems to be a spike every two months or so. This could be because the creator is a student and thus is able to put up more videos during semester or mid-semester breaks.

8. Wordcloud for video titles

A word cloud is a visualization technique that represents textual data, typically a collection of words or text documents, in which the size of each word indicates its frequency or importance within the text. The words are usually displayed in a visual arrangement, with more frequent or important words appearing larger and bolder.

Word clouds are commonly used to gain insights from text data and highlight key themes or patterns.



The word cloud shows Each word in all of the video titles an represents them based on frequency. With this, we can determine that most of the videos contain the word study, live, and Pomodoro.

Conclusions and future research ideas

Conclusions

In this project, I explored the video data of james scholz and revealed many interesting findings:

The more likes and comments a video has, the more views the video gets(it is not guaranteed that this is a causal relationship, it is simply a correlation and can work both ways).

Likes seem to be a better indicator of interaction than comments and the number of likes seems to follow the "social proof",

which means the more views the video has, the more people will like it.

Most videos have between 5 and 30 tags.

Most-viewed videos tend to have an average title length of 20-70 characters. Too short or too long titles seem to harm viewership.

Videos are mostly uploaded on Sunday and other weekdays and Friday and Saturday aren't popular upload times

Project limitation:

The findings should also be taken with a grain of salt for a number of reasons:

The number of videos is quite small(the dataset has less than 500 videos)

Many other factors haven't been taken into analysis,

including the marketing strategy of the creator and many random effects that would affect how successful a video is.

Ideas for future research:

To expand and build on this research project, one can:

Expand the dataset to also include bigger and smaller channels within the productivity and study niche

Perform sentiment analysis on the comments and find out which videos get more positive comments and which videos get less positive comments

Do market research by analyzing questions in the comment threads and identifying common questions/market gaps which could potentially filled

Conduct this research for other niches(e.g. vlogs or beauty channels), to compare different niches with each other to see the different patterns in viewership and video characteristics.

References / Resources used:

[1] Youtube API. Avaiable at https: // developers.google.com/youtube/v3

[2] Converting video durations to time function. https: // stackoverflow.com/questions/15596753/how-do-i-get-video-durations-with-youtube-api-version-3

[3] P. Covington, J. Adams, E. Sargin. The youtube video recommendation system. In Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '16, pages 191-198, New York, NY, USA, 2016. ACM.