YouTube Trends

STA14B Final Project By Miners

Haolin Li | Haochen Zhan | Yiting Kuang

Instructor: Nick Ulle

We choose to analyze the trending videos on YouTube for our final project. We broke down the report into 6 parts with our analysis and observation within each part, for the sake of it will provide an overview in detail:

Project Content:

Introduction and Motivation:
- Why is our research important? Who is our target audience?

Part 1: Summary about YouTube Trend
- What is YouTube Trend? How does YouTube determine trending videos?

Part 2: Explore the Dataset

Part 3: YouTube's determination analysis
- To prove YouTube's measurement of how they determine trending videos

Part 4: Discovery of other components
- To find new factors that might make a video trending, that YouTube did not stated.

Conclusion and Considerations:
- What did we find from our analysis? What can audiences take out from this research?

These parts will be explained in detail in the body of the report. Each part will contain deep and precise analytical results and information regarding the topic of each part, to help readers better understand our dataset, perspective, approaches and decisions that take us to our conclusions. Additionally, our code is commented to allow readers to follow along and understand our methodology with our process.

Collaborators:

This project is a collaborative project between all three of us. Any repetitive code is mostly a result of our working on separate parts of the project. Most of the code should be efficient and iterable, but anything seems repetitive is simply because of a disjoint collaboration.

Introduction and Motivation:

For our project, we want to see if there are patterns within the trending videos and specific components of a video that can make an ordinary video to become trending on YouTube. There are many trending videos on YouTube daily and they can be very different or similar to each other. Our goal is to see if they share certain similarities that influence them to be trending. More about the YouTube trend will be discussed in Part 1 and referenced throughout our report. Our ultimate goal is to become like the consultants for Youtuber content creators, to show and suggest how videos can be trending.

We believe that trending videos share certain similarities that make them to be trending, we want to test this hypothesis as well as others during our research. Do trending videos relate to what was happening in the world at that time? We believe that there is a connection between the trending videos and what was happening in the world, videos talking about the heated issues will be more likely to be recommended in trending.

Additionally, we will follow what YouTube has stated about how they determine trending videos, such as avoid using misleading titles, more details with YouTube's statement will be shown in Part 1 and Part 3. We want to prove YouTube's measurement of trending videos, as well as guidelines they stated about videos.

Also, we want to explore deeper into the dataset, to discover new factors that are possible to make videos to be trending, that YouTube did not stated or included in their determination. We want to approach from a different perspective to this project and show our audiences of how trending videos work on YouTube.

There are many components that add up in order to bring a video to the trending section, we believe that our report will be covered some of the important reasons of how videos can be trending, as well as minor reasons that others might not think of.

Part 1: Summary about YouTube Trend

On YouTube, there is a designated section for anyone who wants to know what is happening on YouTube and in the world right now.

More details about what YouTube Trends is defined by YouTube here:

http://youtube-trends.blogspot.com/p/about-youtube-trends.html

Below is from YouTube Help that shows YouTube's statement of how they determine if a video is ranked on Trending.

https://support.google.com/youtube/answer/7239739?hl=en

As mentioned above, despite YouTube published about their determination of trending videos, we want to show if what they said is true, as well as to explore more unstated guidelines to determine a trending video.

Part 2: Explore the Dataset

To begin our analysis, we first explore any unusual columns and cells within the dataset.

There are 40949 videos within the dataset and there are 16 different columns that represent there are 16 different components of the videos.

We noticed that there are many duplicated videos within the dataset, because one video can be trending for multiple days, and being recorded in the dataset, so we clean up the dataset by creating a new variable within the dataset, 'frequency', to represent how many times a video has being recorded as trending within the time period. Then we use drop_duplicates function to clean up the dataset, and now we have only unique videos within the dataset.

We also noticed that the category id variable for each video is represented by numbers. To make future analysis to be more sufficient, we added their category name according to their category id to the dataset.

Details of exploring and cleaning data can be found within our coding files.

Now we are ready to explore deeper into this dataset and begin our research.

Part 3: YouTube's Determination Analysis

Within this part, we will show how YouTube claimed the method they used to determine trending videos. We will prove their measurements, as well as to provide our observations in detail. The following is extracted from the YouTube Help Center:

Trending aims to surface videos that:

- Are appealing to a wide range of viewers
- Are not misleading, clickbaity or sensational
    - **Loud**: Uses ALL CAPS or !!!!! to overemphasize titles.
- Capture the breadth of what's happening on YouTube and in the world
- Ideally, are surprising or novel

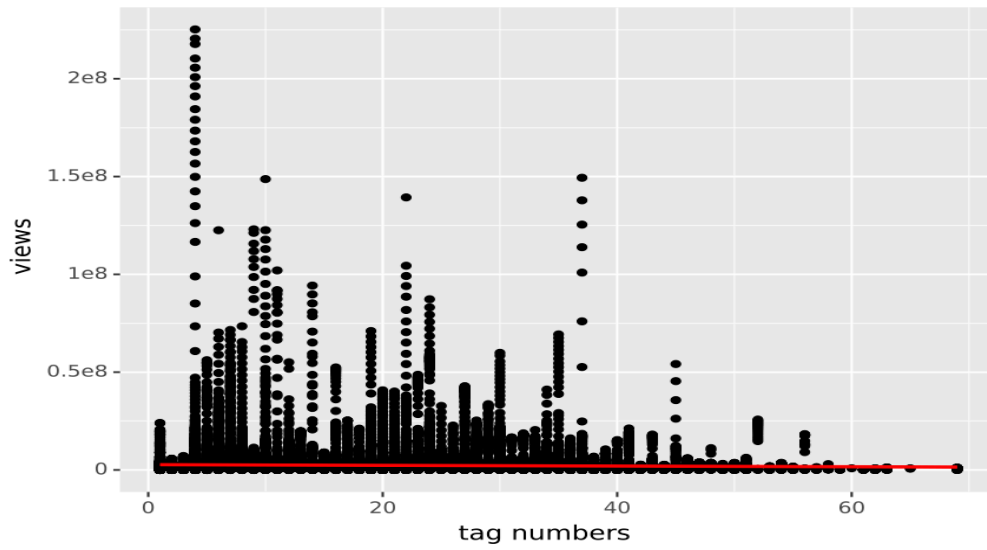We will be examining two of the measurements above.

3.1: Relationship between video tags & topics and heated issues

For this part, we are taking a step on analyzing the tags and categories for the trending videos. According to the academy lesson stated by the YouTube for video creators, the categorization helps viewers find content, allows advertisers to refine their targeting and provides creators with common best practices. Whereas the tags for a particular video can help YouTube's searching engine to identify and rank your videos. They are crucial for optimizing one's channel and helping promote their videos. Since the dataset our team used was all about trending videos from Nov.2017 to Jun.2018, we are about to find out how the tags and categories influence their trending status. The following problems were designed to discover various perspectives for both factors:
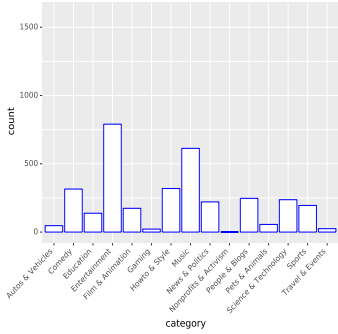
3.1.1: What topic of video do people prefer over time?

The word cloud shows the most trending video tags from Nov.2017 to June.2018. In this time period, the most popular topics were Star Wars, makeup tutorial, and super bowl. From this figure, it seems
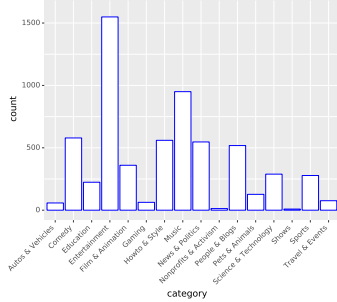
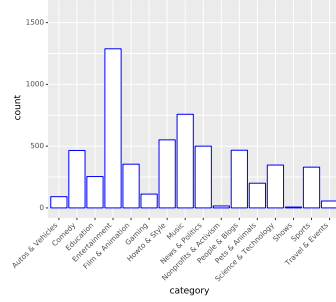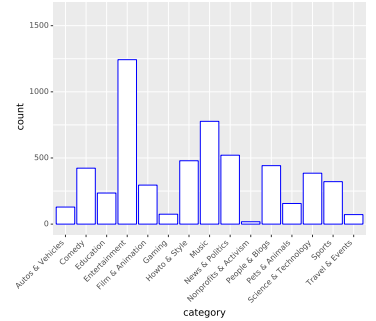that the related trending categories should be movies, how to & style, sports, people & blog. There might be a chance that at some point in the month, a nationwide event pushed these topics to become the trends. In the next few parts, we'll take a look at the categories and tags in different months.

3.1.2: Are tags helping a video to be trending and getting views? Do category of the video and the tags matches/relevant?

Since all the videos we analyze are trending videos, the tags attached to them are helpful by default. However, does the number of tags used on videos affect view amount? The figure below shows a weak relation between them with an R-value of -0.02894 and a small p-value, which infers no linear correlation between them. Nonetheless, we observe there were videos have extremely high views with around five tags and relatively low views when the number of tags is above 40. By glancing over the tags without filtering, I found that the all-time trending tags are related to entertainment, how to & style, music, comedy, people & blogs, film & animation and game. Just like how the categories rank in the bar plot with entertainment, music, how to & style and comedy were leading all months, it shows a strong relation between tags and categories. This suggests video producers include the tags under the thread of their categories. In such a manner, the YouTube matches the video with users' searching request more accurately. Foremost, if the content of the video has attributes that match these tags, they are likely to be shown to the audiences in order to get to the trending list.

Views against Tag Numbers
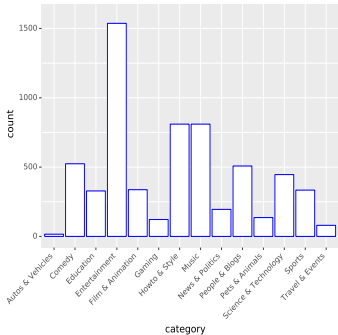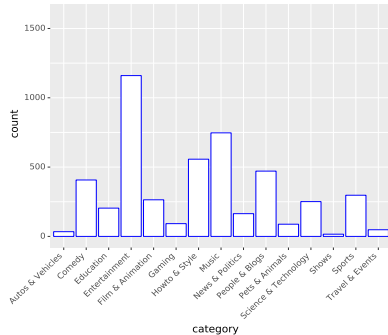
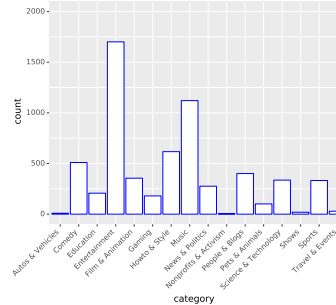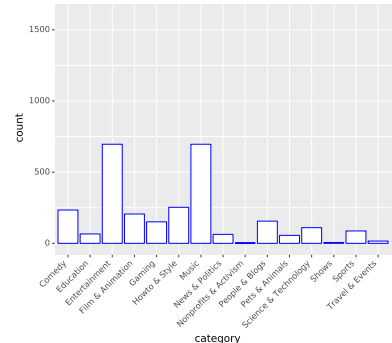### 3.1.3: Do tags and popular video topic relate to the news on that time?

After I filtered out the most shared tags among all months, some unique tags appeared from most shown to less occurred in the frequency plot. The following are the list of events related to each

month. Events or YouTube channels were identified by using the YouTube API v3 search as well as google search.

- **2017/11**: *iphone x*: The phone was released on November 3; Liza Koshy: a youtuber; The Late Late Show channel; justice league: film released in America on November 17; Thanks Giving, November 23rd; Black Friday, November 24[th;]
- **2017/12**: *christmas*: upcoming holiday; star wars last jedi: upcoming film;
- **2018/01**: The topics were mostly repeated from the last month.
- **2018/02**: *super bowl*: on Febuary.4th; black panther: film released on Feb.16th; winter olympics: holding on February 9th; day: Valentine's day on February 14th; star game: NBA All – Star Game on February 18th; commercial: Sky Kids Toy channel;
- **2018/03**: *galaxy s9*: Samsung phone released on March 16; cake recipe;
- **2018/04**: foil ball challenge: Japanese trend; james: NBA player and WWE player; challenge: try not to laugh challenge; Coachella: music festival
- **2018/05**: *iphone*: iPhone SE 2018; wedding: Prince Harry and Ms.Meghan Markle marriage; challenge: try not to laugh challenge; google: keynote; infinity war: film Avengers: Infinity War; fortnite: online game
- **2018/06**: *shawn mendes*: a song creator; smith: Sam Smith: Baby, You Make Me Crazy; Will smith: FIFA Live It Up.
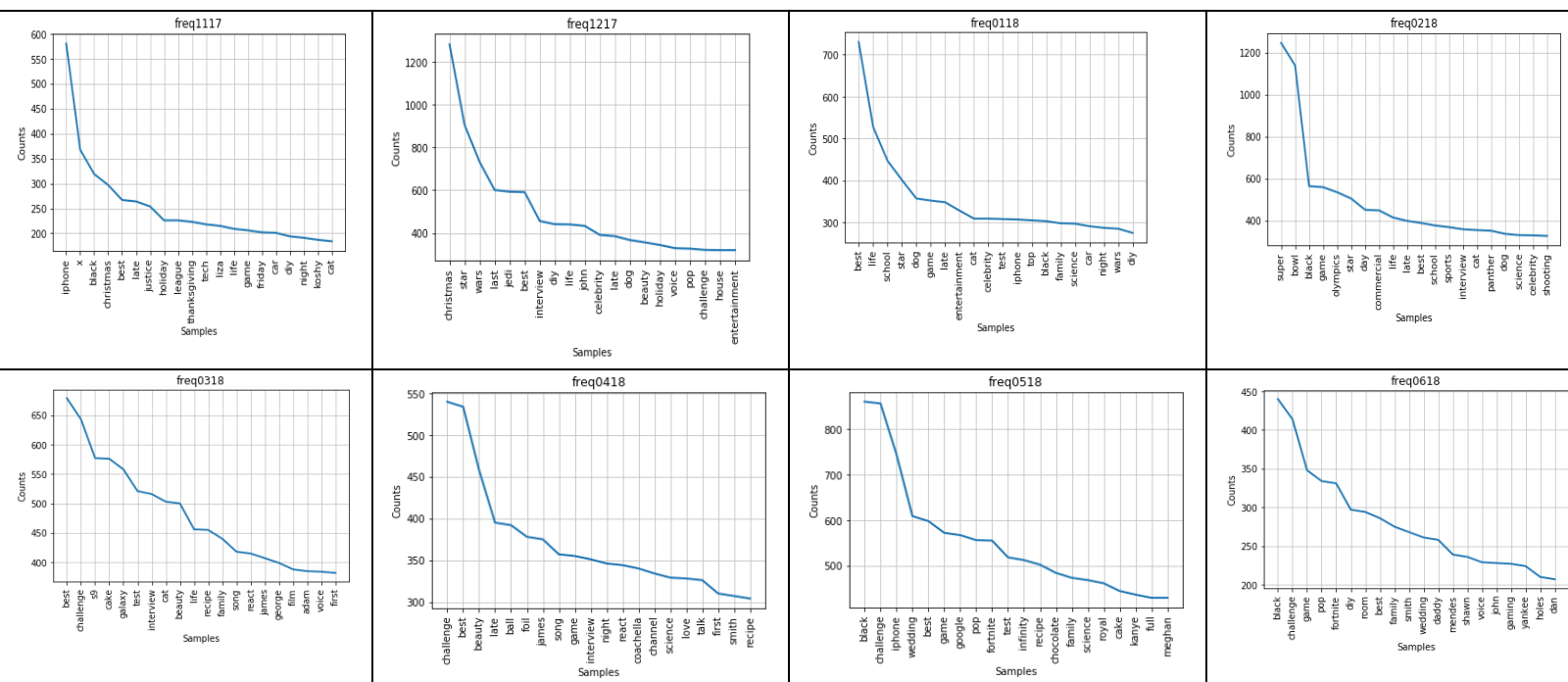

The trending tags were mostly determined by some big events, Youtuber, movie, or songs that were on smash hit. It implies that viewers tend to spend more time on entertainment and follow others

on current significant events. It is a good idea to add elements to the video content that stick with hot topics.

Overall, this part about tags and categories gives an insight to video producers the importance of how they define their category and tags. And more specifically, to make the video in trending, their video contents should reflect the updated tags, and their category needs to be defined well.

3.2: Fully-capitalized video title

As previously mentioned, one feature that YouTube considers when determining if a video is ranked on trending is to avoid ALL CAPS titles. Title is the first thing viewers see when a video is suggested to them. However, while some creators try to use catchy titles to drive clicks and Watch Time, YouTube states that videos are less likely to be recommended to new viewers if their titles are too "**loud**," or overemphasized; therefore, creators need to be careful to use all capitals to emphasize titles. To prove that using fully-capitalized titles is not ideal to attract viewers, we compare the average views between videos with capitalized titles and videos with non-capitalized titles.
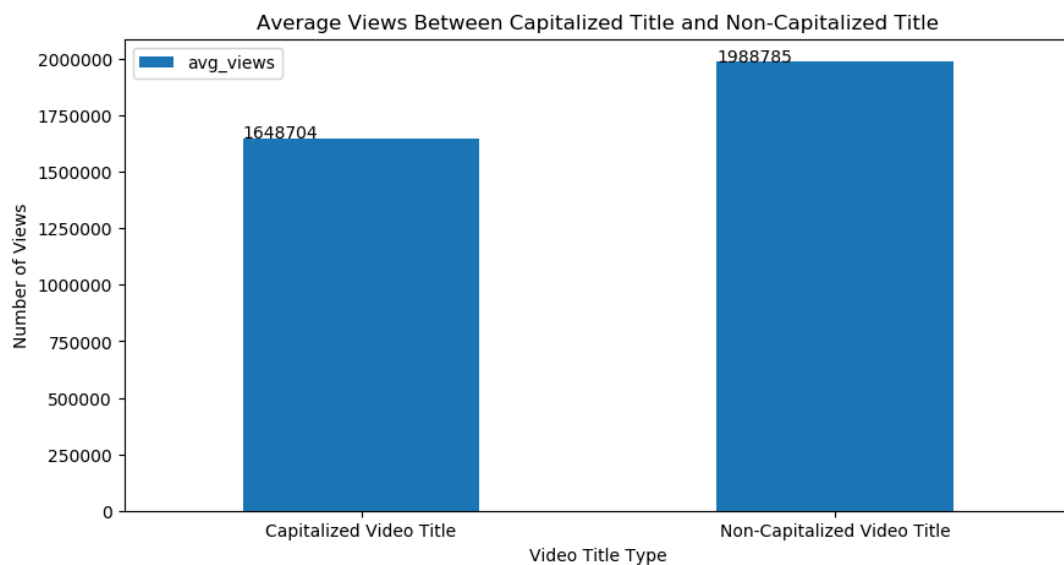
The result shows that capitalized-title videos have an average of 1.6 million views, which is 3 hundred thousand lowers than the average views of non-capitalized-title videos. Thus, creators should avoid using all CAPs titles to optimize their videos for discovery.

Part 4: Discovery of other components

Within this part, we will show one of the discoveries we found that will make video to be trending.

The Most Influential Channels and Their Categories

A personal YouTube channel is available to everyone who joins YouTube as a member. The channel serves as the home page for the user's account, and it's where content creators upload and publish their videos. YouTube provides a place to browse through YouTube channels where you can check out popular channels and subscribe to them if you choose to do so.

Next, we want to explore what are the most influential YouTube channels and see if there are any patterns on their categories. One way to determine the most influential channels is to find the channels that are dominant in the trending section.
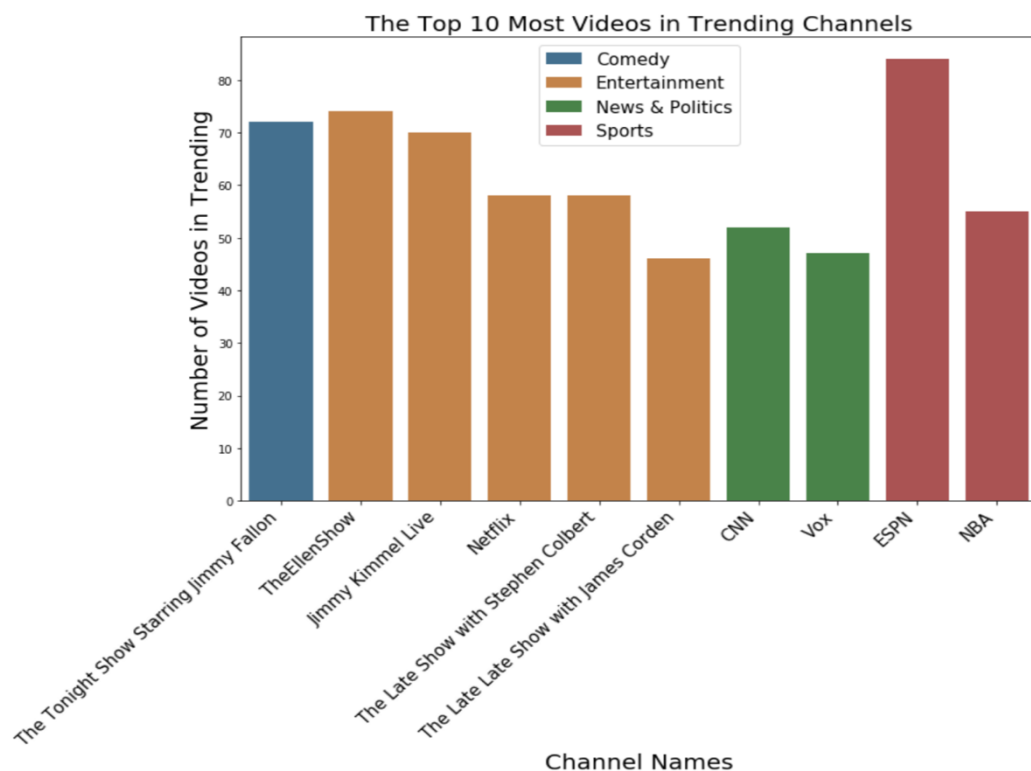
As the bar plot above shown, the channel that has the most videos in trending is ESPN, an official YouTube channel for the ESPN television channel that features up-to-the-minute sports news. Besides the sport category, another category that dominate the trending section is entertainment. Among the top 10 channels with most videos in trending, half of the channels are categorized to entertainment and they are all celebrity interviews and talk shows according to our research. The category comedy is very similar to entertainment, so we can conclude that the most popular video types are entertainment and sports. Thus, this is a good reference for a content creator to know what kinds of videos appealing to a wide range of viewers so that they can decide their video contents.

Conclusion and Considerations:

In summary, our analysis fulfilled our goal to prove that there are patterns/similarities within the trending videos and between each other, to show how YouTube determine trending videos, and to discover new possibility to make a video to be trending.

Since we have proven there is a relationship between video tags & topic and heated issues, avoid using misleading video title, and discovered that focusing on certain topic according to the category, will maximize the chance the video to be trending, Video content creators can take this report as a reference to benefit themselves when creating video contents.

Despite there are many more stated and unstated reasons of how videos can be trending on YouTube, we believe that by looking at this project, readers can get a sense of how the system works and how the measurement is applied within the Trending. In future project, we can look more deeply into the data and explore more components of a video, to provide a more precise, informative and thorough results within the report.

Limitation and Dataset:

The most difficult part of this research was getting a suitable dataset. We started this project with two options of dataset, YouTube-8M and YouTube API. However, YouTube-8M was too large for us to download and handle since its smallest dataset is 31 Gigabytes. Also, it was not appropriate using the YouTube API because we could only retrieve the trending videos of the current day. We wanted to have videos from a period of time, and have a large number of videos to analyze from.

We then searched a new dataset and decided to use the one from Kaggle created by Mitchell J, which she collected the trending videos daily for months using the API.

The other difficulty of this research was the limitation of the dataset, we brainstormed many different approaches of the topic and came up with many questions, however, due to the limited categories within the dataset, we had to approach based on the conditions we have.

Sources and Links:

https://support.google.com/youtube/answer/7239739?hl=en

https://support.google.com/youtube/answer/141805

https://www.kaggle.com/datasnaek/youtube-new

https://creatoracademy.youtube.com/page/lesson/overview-categories?cid=platform&hl=en#strategies-zippy-link-2