The secrets behind different categories in US Youtube videos

By Xinyu Liu

Since the foundation of Youtube online video in 2005, it has made a huge set of breakthroughs that surprises everyone. It is one of the world-famous video sharing website and maintains a list of the top trending videos on the platform. There are various kinds of videos on it and we can't even possible to know actually about its trending unless we do the analysis from the data. Here the data I use is called US videos, which is a daily record of the top trending YouTube US videos and was collected using the YouTube API, including the data of the video title, channel title, publish time, tags, views, likes and dislikes, description, and comment count from year 2006 to 2018.

Figure1: WordCloud of US videos titles and tags

When you take a first look at the data and may eager to know what kind of videos has the tendency to be popular, you may first refer to a WordCloud of the titles and tags of the video shown in Figure1. In the WordCloud of ‘tag’, we may see the most frequent items are among 'makeup tutorial', 'music video', 'star wars' and etc. It probably gives us a primary and direct feeling of what is most popular in the tags of the music videos in United States, though is a good understanding of this dataset but actually kind of limited to tell the whole story. More need to be analyzed.

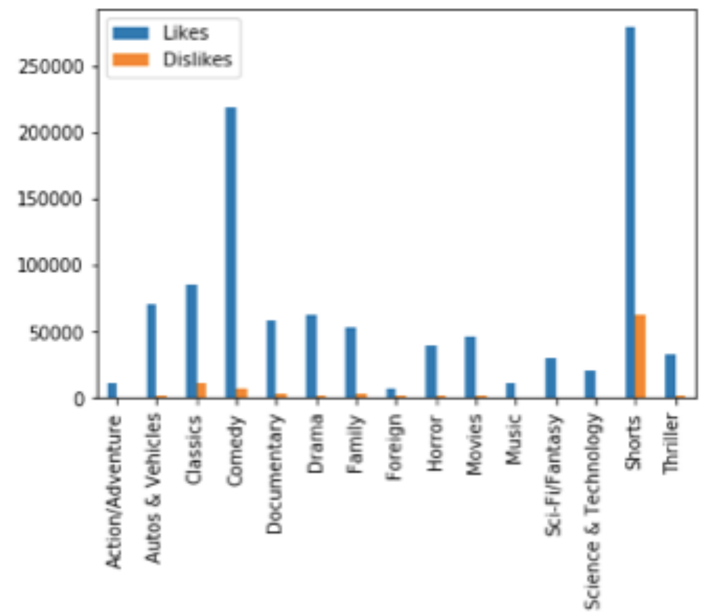

Figure2: Average Likes and Dislikes of different US video categories

Then we go further to take use of the category of the videos that like or view or comment most by people, for example the like&dislike of different categories in Figure2. At first glance, we are surprised to see Shorts gets the most average likes and dislikes at the same time and Comedy gets the second most likes but not so many dislikes. It is an interesting start and you are likely to think there must be something behind it.

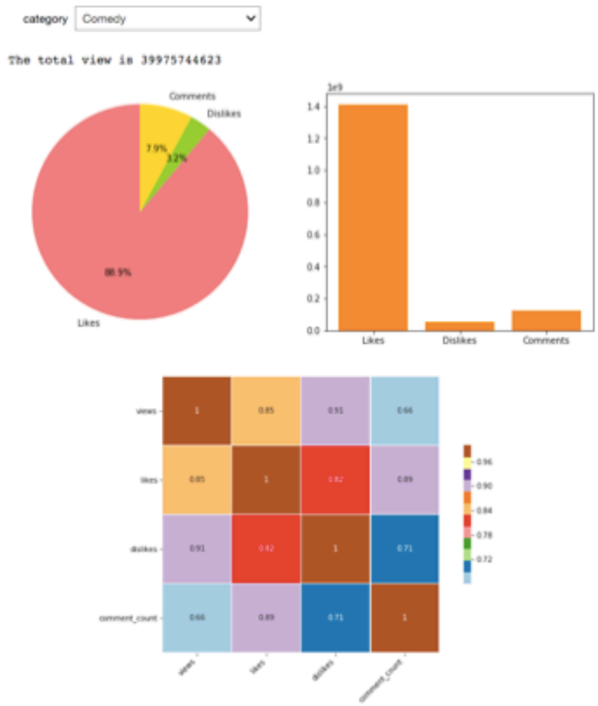

Figure3: Analysis and Correlation of reviews in the category Comedy

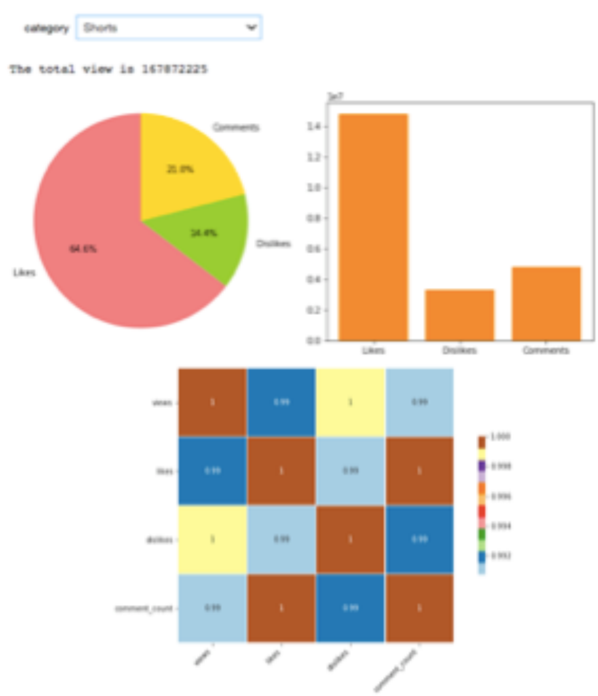

Figure4: Analysis and Correlation of reviews in the category Shorts

Let’s just focus for a while on the two categories and combine them with their analysis and correlation of reviews in the two categories in Figure3 and Figure4. In the correlation heat-map of Comedy, the relation between likes or dislikes and views is high while between views and comment_count is not of the same level. However, in Shorts, all relations between each other is quite high and close to linear relation and see its statistical plot: in the total 167872225 views, compared to the statistical plot of Comedy whose views size is about 2.5 times of Shorts, the dislikes and comments really take up a relatively big part. You may say that investing in Shorts is just like the condition of stock, high income and high risk and Comedy is more like securities, more stable to manage. However, there is still a very import key point we have’t include in, which is time, especially the time duration between publish time and trending time.

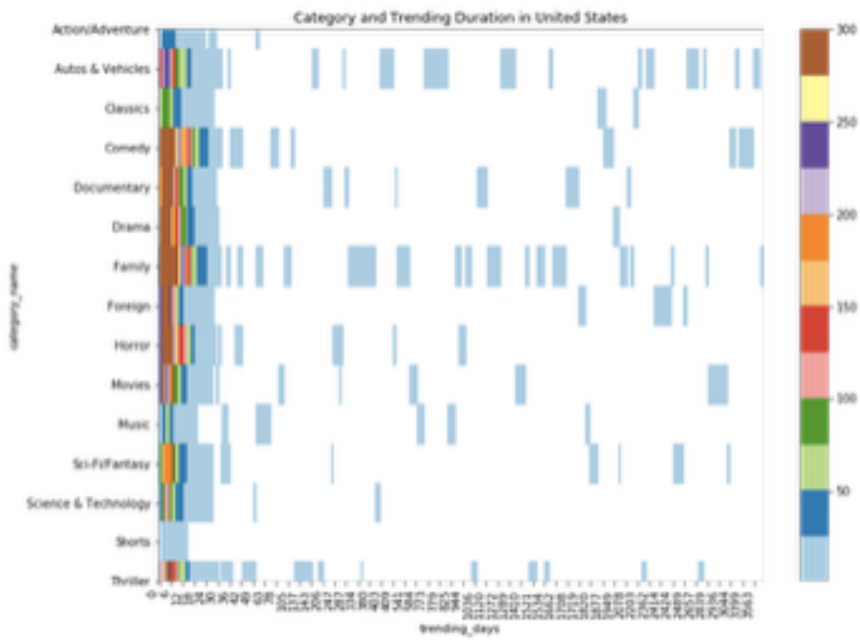

Figure5: heat-map of different categories and trending duration

So, after computing the subtraction between publish date and trending date(omit a lot of trivial details), we get what I called the ‘trending days’. Here is a heat-map of the Category and its trending duration in Figure 5. Still take a look at the Comedy and Shorts, we can see that although they both for most part have short trending days like within one month, Comedy generally has the good possibility to be discover and popular again while the Shorts mostly popular for a short time and never come back. If you are an investor, you may be in a dilemma of whether choosing the long-term popularity or short-term explosive heat. The attracting part and what I am emphasizing is not only about the two categories I have analyzed here, but also other similar or totally different categories, there is a lot to discover. And broadly no matter what you want to do about this, there is always something to do.

Back to the WordCloud in the beginning, there are striking words 'Music Videos', 'Star World', 'Makeup Tutorial' and so on and they are likely from the category of movies, music and others. After all of what I have discussed, I swear you must have got more interesting ideas and welcome to communicate with me.

Source:

Information:

https://en.wikipedia.org/wiki/YouTube

Data:

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

Writing examples:

https://fivethirtyeight.com/features/media-coverage-doesnt-actually-determine-public-opinion-on-the-economy/

https://fivethirtyeight.com/features/a-comic-strip-tour-of-the-wild-world-of-pandemic-modeling/?cid=referral_taboola_feed