

Report

Project Goals:

The booming social media and influencer industry has radically changed the game for both marketers and consumers as shown in the two linked¹ articles. This industry boom has created a new "Career goal" for a generation of young minds. This young generation has a real opportunity to earn serious amounts of money just by having a lot of followers and getting a lot of likes on sites such as Instagram, Twitter, and YouTube. However, to become an influencer and really break into this industry you need to be able to generate the likes and fan engagement statistics both in insane amounts and at an insane rate. This is why we decided to explore the YouTube dataset in an attempt to gauge what type of content and what posting habits may lead to greater viewership/likes and, therefore, opportunity of revenue generation for the potential influencer.

The particular YouTube dataset we are exploring includes several variables that could help us in our exploration. These variables might be able to show us some insight into what brings success in the influencer lifestyle. Some of the variables are the date the video was trending, video title, channel title, video category, time the video was published, views, likes, dislikes, comment counts, etc.

A particularly important feature of our dataset is that it included data for several different countries. So, we chose to focus on four countries in particular; USA, Canada, Great Britain, and Mexico. This decision was made with the idea of making comparisons and discovering major similarities/differences or hidden patterns in the type of videos that trend in each of these countries.

Exploratory Analysis:

While we had four different countries to explore, we decided to approach this exploration in the same way for all four in order to have a better point of comparison.

1

<https://www.forbes.com/sites/forbestechcouncil/2019/02/13/calculating-the-true-size-of-the-influencer-marketing-industry/#580068ec658d>
<https://www.businessinsider.com/the-2019-influencer-marketing-report-2019-7>

The different types of analysis we explored focused on category popularity (by average number of views), channel popularity (by average number of views), time to trending analysis, publish time analysis, and text analysis (specifically, tags and descriptions).

In respect to category popularity, we found that on average the "Entertainment" category was the most popular across 3 of the four countries. Great Britain was the outlier here, who's most popular category was music. Additionally, "People & Blogs", "Music", "How To & Style", and "News & Politics" always being near the top in each of the countries.

When it came to channel popularity, the overwhelming majority of Top 3 spots were taken by music artists' official channels. The US had its Top 4 consist of music artists. Canada had six out of the Top 10 spots consist of music artist channels. Great Britain had seven out of its Top 10 spots consist of music artist channels. In contrast to these three countries and their affinity for music artist related channels, Mexico did not have a music artist related channel in its Top 4 (the fifth top channel belonging to BTS) and ChildishGambinoVEVO came in at #8 in terms of popularity by average number of views. This shows a definite difference between the video viewing habits of USA, Canada, and Great Britain versus those of Mexico.

Particularly, ChildishGambinoVEVO had an overwhelming dominance in the USA where the channel averaged over 14 million views. Compared to the second-best channel (BeckyGVEVO, which averaged around 6 million views), we can certainly see the enormous gap between these channels. ChildishGambinoVEVO was number one and number two in Canada and Great Britain, respectively.

The results from our Time to Trending Analysis focused on both categories and channels. For the US, the fastest to trending category on average was "Nonprofits & Activism", while the slowest to trending was "Autos and Vehicles". Additionally, the fastest to trending channel was of Bill Gates. In Canada the fastest category on average was "Sports" (with the fastest channel on average being "CanucksHub" a non-official hockey team channel) and the slowest was "Science & Tech". In Great Britain the fastest category was "Music", while the slowest was "Comedy". Finally, in Mexico the fastest category to trend on average was "Pets & Animals" and the slowest category was "Music". Somewhat surprisingly "The View" was the fastest to trending on average channel in Mexico.

From our publish time analysis we were able to explore publishing trends amongst the different countries to try and make inferences on the possible effect this might have on time to trending patterns. For the US, the

most common publish time for a video was 4 PM. However, the videos fastest to trending were those posted at 11 PM. In Canada, the most popular time to post videos that would trend quickly was 5 PM. These quick-trending videos were mostly comedy videos and would on average take about a day and two hours to reach trending status. Great Britain had videos posted at around 12 PM - 1 PM. Additionally, the top category videos are also posted at around this time. Finally, in Mexico, the most popular time to post videos that would trend quickly was 9 PM. Interestingly, the most popular time to post videos in general was 2 AM. However, these videos posted at 2 AM would often be near the bottom of the list for the fastest to trending videos.

In regards to text analysis we were able to separate the strings within titles/tags/descriptions to create lists of these words and also create lists of the most popular words being used in those video titles, tags, and descriptions using regular expressions and string functions. Then, using list methods and pandas functions, we were able to filter for videos in which words related to a user-selected video/topic were present. And finally, we are working on a method to tag each word accordingly using the Natural Language Toolkit in python to get the desired "Part of Speech"² out of our created word-lists (this is still a WIP, as of this writing).

Solution and Insights:

While we set out to take a predictive approach to our analysis, our dataset did not lend itself well to the task. Even though we were unable to truly predict a variable, we were able to extract several insights from our investigation of the YouTube dataset. In order to attempt a more predictive approach, our dataset would have been better served including data from videos that did not reach trending status.

Some of the most important insights we were able to extract revolve around publishing time and our text analysis of tags and descriptions. Specifically when talking about publishing times, we were able to identify what hour of the day is best for each of the countries in order to achieve a faster time to trending. These times differ from country to country and are most likely due to cultural and societal differences within each region. While the US and Mexico have their fastest to trending videos at night (11 PM and 9 PM respectively), Canada has their fastest trending videos posted at around 5 PM and Great Britain at around noon. Through this publish time

² <https://www.nltk.org/book/ch05.html>

analysis we are able to make recommendations to aspiring influencers about what time they should post their videos in order to maximize their number of views and minimize the time it takes them to get to trending depending on their specific country. Moreover, assuming we could get more localized data, the same approach would work to make recommendations for more specific locations.

Additionally, the text analysis approach that was described in the previous section also allows us to provide recommendations to aspiring influencers as to what tags would ensure that their videos are seen by the audience they are truly targeting.

Unfortunately, we also found that if your name is not Kylie Jenner or Childish Gambino you may have a hard time rising to the top of the YouTube trending list. Meaning that, while the influencer industry is certainly growing and becoming a very good source of income, it still is not quite as mainstream as some would have us believe. Additionally, one has to take into account that the YouTube trending algorithm has been a hotly debated topic in recent years due to several controversial videos making it to the top of the trending list³. This creates a lot more uncertainty regarding what exactly a content creator must do in order to reach the trending status. Thus, it is safe to conclude that being an influencer can still be a somewhat unreachable career goal without the proper drive and motivation. ⁴

³ <https://www.polygon.com/2018/2/21/17035758/how-does-youtube-trending-work-parkland>

⁴ Pertinent charts, plots, and data will be included in the Presentation deliverable.