



Which category of Youtube is most likely to create long term success?

Oscar Sucre

Data 101

Data Driven Blog

3/5/2023

<https://www.kaggle.com/datasets/themrityunjaypathak/most-subscribed-1000-youtube-channels>

Overview:

I am a senior at Rutgers in my last semester, my major is Economics and my minor is Statistics. I always had an interest in data science but I wasn't aware of the data science minor. One of the biggest platforms today is Youtube which has allowed millions of people to become creators of many types of video content. Being on Youtube is one of my favorite pastimes and tools for learning, whether it be learning a new guitar song or how to fix a car part. The creators on Youtube, called "Youtubers", have been able to have very successful careers by being on this platform, top creators being multimillionaires. This has invited me to investigate what the most successful type of youtubers are by looking at the data of the 1000 most subscribed. I will gauge how many videos are made per top channel and see how many views they have garnished. What is the best performing category on youtube?

Dataset:

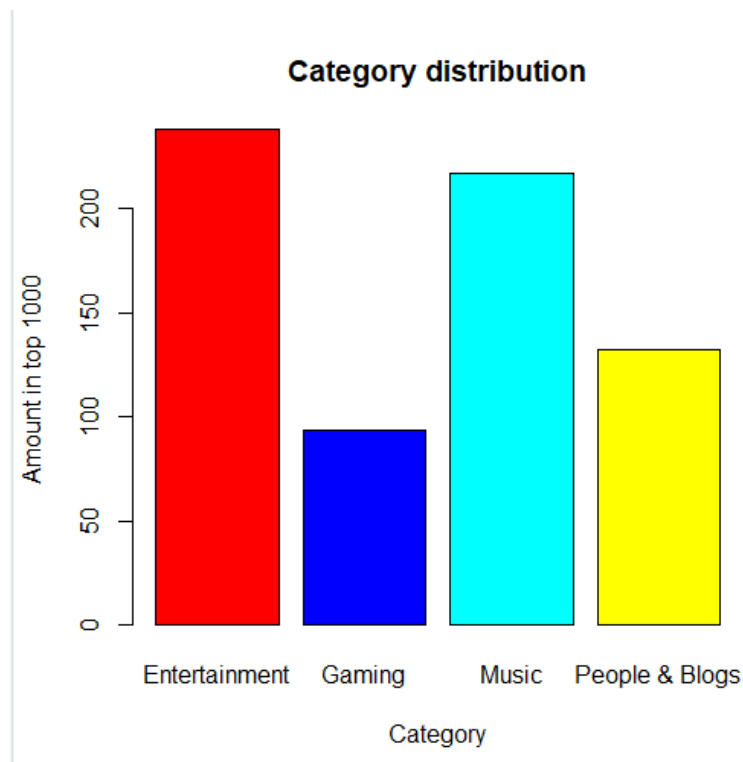
I found this dataset from Kaggle. It is reliable as it links the website that scraps the data for the dataset and is taken directly from Youtube. This data also has a 10 usability score which is ideal. Although I did have to clean the data marginally to be able to manipulate it. The only downside to this data is that it only has 1000 sample sizes so it will be difficult to subset further. But it would be very difficult to find data on more youtube channels because there are so many.

Narrowing down:

Originally there are 19 Categories to deal with, including N/A. This is definitely too many to be dealing with, especially because there are so many with an incredibly small amount.

| | | | | | | |
|----------------------|--------|-----------|-----------------|-----------------------|----------------|----------------|
| Autos & Vehicles | Comedy | Education | Entertainment | Film & Animation | Gaming | Howto & Style |
| 1 | 68 | 45 | 238 | 50 | 94 | 43 |
| Movies | Music | N/A | News & Politics | Nonprofits & Activism | People & Blogs | Pets & Animals |
| 2 | 217 | 30 | 27 | 2 | 132 | 6 |
| Science & Technology | Shows | Sports | Trailers | Travel & Events | | |
| 18 | 14 | 10 | 2 | 1 | | |

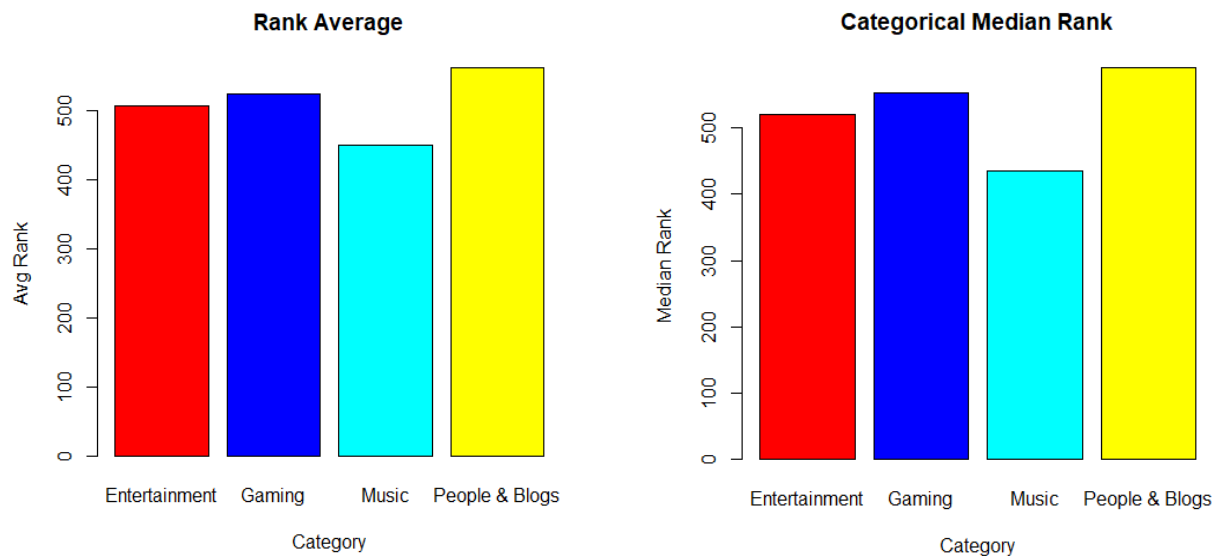
Out of the 19 I decided to pick the ones with a significant amount of the standings, being Entertainment , Gaming, People & Blogs, and Music.



Investigating which is the best:

Investigating with Subscriber Rank.

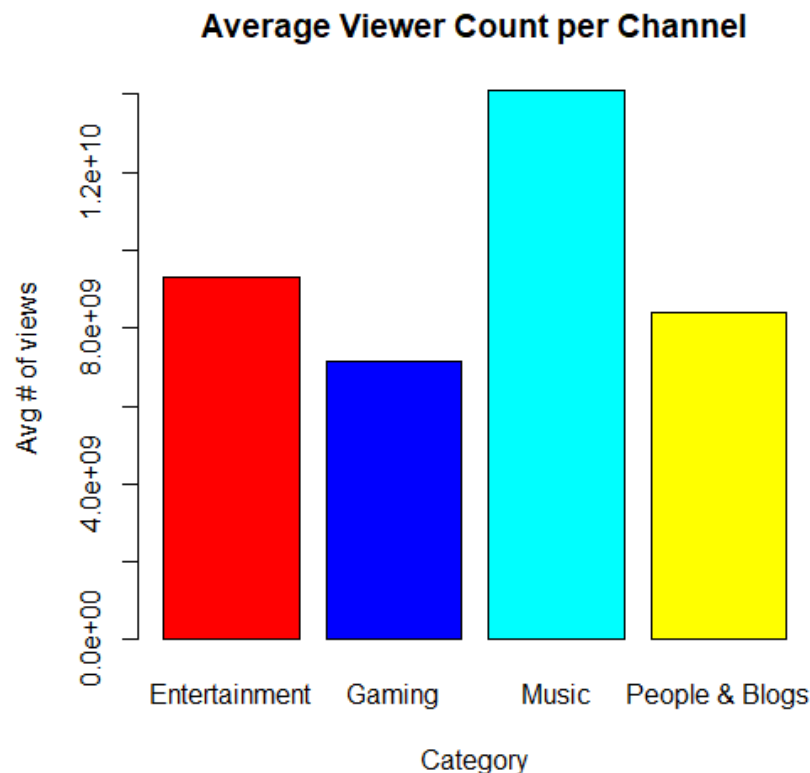
For the best category to get into for youtube I firstly looked at the overall rank of each of the 4 categories. The lower the ranks should imply which is the best.



Here it shows that both average and median ranks imply that Music is the highest ranking overall. But it isn't by a large margin, so I decided to investigate further.

Investigating with View Count.

When working with the viewer count, I had to manipulate the data to work with the viewer count number because it was actually in char format and included commas. After changing it into numeric form I took the mean view count for each category and found much clearer results than last time.



Music is clearly on top by average viewers by far. Supporting the earlier finding that music had better mean and median scores.

Using Z- Test to find P-Value:

My Hypothesis from using my previous discoveries is that Music is the best category and most likely to be at the top.

First, I tested the differences of each category using Rank. The null hypothesis for each of these is that Music mean Rank is equal to the other mean ranks.

The Alternative Hypothesis are:

Z-test for Music mean Rank < Entertainment mean Rank:

P-Value = 0.016

Reject the Null Hypothesis

Z-test for Music mean Rank < Gaming mean Rank:

P-Value = .0172

Reject the Null Hypothesis

Z-test for Music mean < People & Blogs mean:

P-Value = .00027

Reject the Null Hypothesis

These are great results for my hypothesis showing that the mean Music rank is lower than all the other ones. They are all lower than the P-Value .05 by a wide margin making them significant.

The second null hypothesis I am testing is that mean Music video views are equal to the other mean video views.

The Alternative Hypothesis are:

Z-test for Music mean video views > Entertainment mean video views :

P-Value = 5.655236e-05

Reject the Null Hypothesis

Z-test for Music mean video views > Gaming mean video views :

P-Value = 1.150877e-08

Reject the Null Hypothesis

Z-test for Music mean video views > People & Blogs mean video views:

P-Value = 7.477492e-05

Reject the Null Hypothesis

These are great results for my hypothesis showing that the mean Music video views are higher than all the other ones. These P-Values are all incredibly close to 0 making it easy to reject at any level of significance especially .05.

Conclusion:

When looking at the most successful videos they are clearly ones that focus on Music. These videos thrive because of their replay ability and popularity outside of just the Youtube platform. This was proven by both their subscriber rank and their videos viewed.