Vandana Anand

Krystal Grant

Daniel Mcdonough

Youtube Trending Videos

Introduction:

Youtube is a famous website in which users can share and watch videos ranging from any topic. In addition, to provide users with updated news and trending videos, Youtube has a platform to show the list of recent, viral videos. Moreover, social media sites such as Youtube give individual opinions and false information a level of credibility comparable to established media outlets. Youtube itself has a massive social significance as almost 1.3 billion people use it everyday. Since so many people use the site, the most relevant and popular videos spread like wildfire to other sites. These videos (if news related) are often regarded as definitive sources even if they are meant as an opinion/reaction/satire piece. This has become increasingly significant to analyze since the “fake news” age is upon us and it’s important to know where the information for trending videos come from, as it could relay echo chambers or false information. We hope to answer the big question: are people switching away from mainstream media in favor of an independent secondary information source?

Hypothesis:

We hypothesize that US trending videos on Youtube about current events come from and are viewed mostly from independent second hand new sources such as vloggers and other channels (Philip Defranco, Late Night Shows, etc), rather than mainstream news organizations (NBC, CBS, FOX, etc).

Definitions:

Mainstream News Sources: A source that most have heard of and consider reputable. May have other mediums of News than just social media.

Independent News Sources: A source that has no vested interest in a written topic and therefore it is commonly expected to describe the topic from a disinterested perspective. There are a small number of individuals that use their own video as a source.

Methodology:

1. Obtain Raw Data From: https://www.kaggle.com/datasnaek/youtube-new
2. Organized Data to be between November 2017 and February 2018
3. Extracted the Videos from the chosen 3 categories: Travel and Events (19) ,News and Politics (25), Science and Technology (28)
4. Parse for what qualifies as News
    a. (ie remove Superbowl Commercials, and Movie Trailers)
5. Research the uploaders' channel to determine classification as a Mainstream or Independent news source.
6. Calculate and compare the averages of views of Independent and Mainstream news.

Assumptions:

- Youtube has not explicitly stated what dictates a view, however it is known that Youtube verifies all views after 300 and check proportional watch time to overall video time.
- Views are from all over the world but are trending in America
- To constitute as a current event/news the video has to be within the following boundaries
    - Have a relevant Category ID:

19 - Travel & Events
21 - Videoblogging
22 - People & Blogs
23 - Comedy
24 - Entertainment
25 - News & Politics
27 - Education
28 - Science & Technology
29 - Nonprofits & Activism

    - Must affect society as a whole
        - ie. DIY and Unboxing videos do not count
- The category given to the video are absolute and accurate
- The difference between Independent and mainstream news Organizations are determined by all the following criteria:
    - Mainstream news organizations are “grand” in the sense that they cover mediums other than Social Media
    - Independent news sources are smaller in that they may only have Social Media and one other medium and are run by a few members
    - Mainstream news is frequently updated

Raw Data:

https://docs.google.com/spreadsheets/d/1sqtnjjxNOB9SFXKp0s7ee0KG0dnvUTHc8R8mlN8GzVQ/edit#gid=1964268307

Sorted Data:

https://docs.google.com/spreadsheets/d/1yx8ALrBK8MU2be3H7RKKJaBcslIXcED2YMbr8E8E16o/edit#gid=0

The following are the summarized data tables:

Category: Travel and Events	Independent News Source	Mainstream News Source
Amount of videos for each source	**24**	**15**
Avg. Views for each source	**429,483.6667**	**2,030,372.4**
Did the category pass our hypothesis?	NO (P < 0.0001)	

Explanation:

We can see that, on average, views for mainstream news sources are more than the views for independent news sources for the travel and events category. About 1.6 million more people have viewed mainstream news source videos than independent news source videos. We can observe that there is no correlation between the amount of views and the amount of videos for each type of news source because the difference in the amount of views is very large even though mainstream sources have less videos. Therefore, our hypothesis is disapproved.

Category: News and Politics	Independent News Source	Mainstream News Source
Amount of videos for each source	**7**	**172**
Avg. Views for each source	**167,298.4286**	**284,794.1221**
Did the category pass our hypothesis?	NO (P = 0.0027)	

Explanation:

From the data summary, it can be observed that there is a greater average amount of views for mainstream news sources than independent news sources in the news and politics category. About 117,000 more people have viewed mainstream news sources than independent

news sources. Although there are much more videos for mainstream videos in this category, the number of views for each category doesn't differ by a very large amount. Again, this shows there is no correlation between the amount of views and the amount of videos for each type of news source. Therefore, our hypothesis is disapproved.

Category: Science & Tech	Independent News Source	Mainstream News Source
Amount of videos for each source	**13**	**26**
Avg. Views for each source	**448,922.1538**	**1,941,732.923**
Did the category pass our hypothesis?	NO (P < 0.0001)	

Explanation:

From the data summary it can be observed that there is a greater average amount of views for mainstream news sources than independent news sources in the science and technology category. About 1.5 million more people have viewed mainstream news sources than independent news sources. Although this shows there may be a correlation between the amount of views and the amount of videos for each type of news source, the average number of views is much greater for mainstream sources than independent sources. Thus, our hypothesis is again disapproved.

Overall Results:

We observed that in all three of our selected categories, the average number of views for the mainstream news videos is significantly greater than the average number of views of independent videos, thus disproving our hypothesis. The P values were determined by comparing the two means via the z index calculated by the following equation:

$$z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

In all 3 cases, the z index corresponding p value is greater than the a statistically significant confidence value for disproving a null hypothesis. So, we have very little confidence to prove our hypothesis accurate.

Conclusion: Did we approve or disapprove our hypothesis?

Since all of our three categories failed our hypothesis, we would say that our hypothesis was disproved meaning that mainstream news sources are more prefered than independent sources. All of the data from each of the three categories selected shows that on average, the number of views on the mainstream news videos is significantly greater than that of the independent videos. Some variables of our experiment that we found may have affected our results are some of the categories that we selected to investigate. For example a lot of travel videos come from independent YouTubers rather than companies and is intrinsically more opinion based than neutral and news related. Another variable that could affect the overall experiment is determining what qualifies as news and that some mainstream news sources could also be biased and so may not be as trustworthy in general. If we were to later expand on the project, we would analyze all of the relevant data to collect more information on whether news sources are from official platforms or not. We would also go deeper into the content of each video to truly understand whether the videos are truthful to their described title and description.

Hindsight: Could someone arrive at a different conclusion by using the same raw data?

Using the same raw data there are many different conclusions that could be made by looking at different aspects of the data. One could use more or all categories in testing our hypothesis which may result in a different outcome. Someone could also classify Independent and Mainstream differently where we would have conflicting data and possibly different results. Some may use different metrics to analyze this data. The likes and dislikes, for example, were ignored in our analysis. Including this data could give a representation of US viewers' opinions on Mainstream and Independent sources.

Other Interesting Data:

Overall Data	Independent News Source	Mainstream News Source
# of Unique Sources On Trending	**37**	**58**
What Could This Mean?	Mainstream News is more likely to get on to trending?	

Resources:

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

Youtube

Google Sheets