**YouTube Trending: Coronavirus Comments**

Julianna Cole (jec56@njit.edu)

Rhea Pavithran (rp553@njit.edu)

New Jersey Institute of Technology

IS 465-002: Advanced Information Systems

Professor Xinyue Ye

May 12, 2020

**Introduction**

After conducting research about the YouTube Trending page and its relationship to the coronavirus outbreak, we wanted to continue looking further into this topic. In our first research project, we discovered that there was a relationship between the increase in the number of coronavirus-related videos featured on the YouTube Trending page and the increase in the number of coronavirus cases in the United States. We decided to go a step further with this research and examine the comments posted to these coronavirus-related videos. By looking at these comments, we hoped to gain insights into how people were reacting to the coronavirus outbreak.

Our first research question is, “What are the most commonly used words in the top comments on coronavirus-related videos featured on the YouTube Trending page during the coronavirus outbreak?” Our hypothesis was that the most commonly used words might include “coronavirus”, “virus”, or “quarantine”. It would make sense that these words would be popular in the comment sections of the coronavirus-related videos we were examining because users may be commenting on the content of these videos. However, we were surprised to identify the word “people” as the most frequently used word in the collected comments.

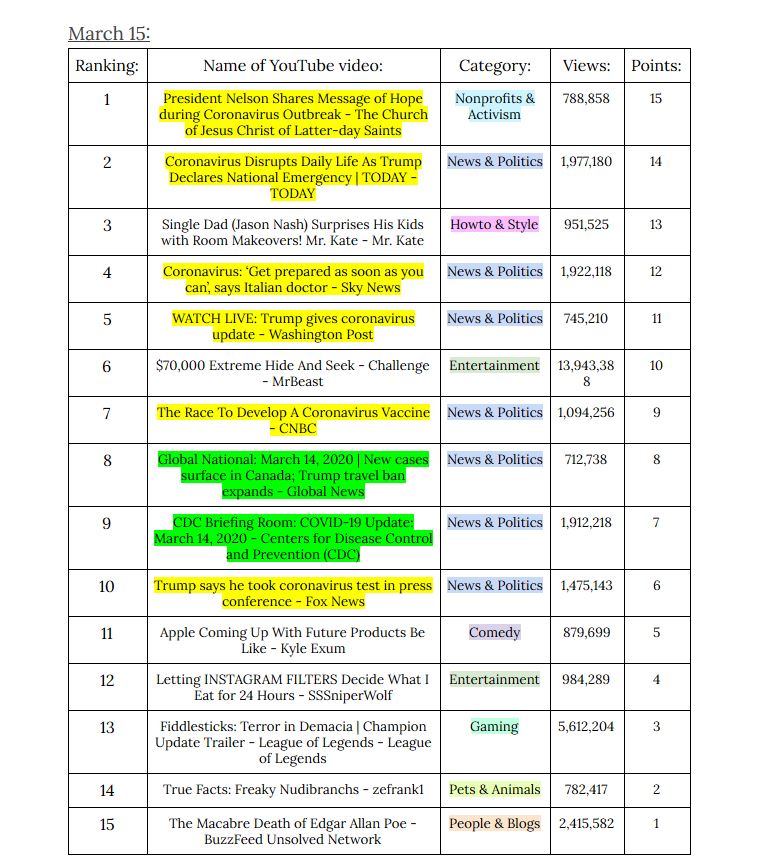
After answering our first research question, we wanted to discover the reason why “people” might be the most popular word used in the collected comments. Our second research question is, “In what context is the word ‘people’ used in the top comments on coronavirus-related videos featured on the YouTube Trending page during the coronavirus outbreak?” We wanted to know the specific context in which this word was being used because it would allow us to better understand users’ reaction and feelings toward the coronavirus outbreak.

Our third and final research question is, “What is the most upvoted type of comment on coronavirus-related videos featured on the YouTube Trending Page during the coronavirus outbreak?” After reading through the comments containing the word “people” and determining the context in which it was being used, we wanted to continue this research with all of our collected comments. By determining the types of comments that are most popular on YouTube, we can get a better idea of the way that users communicate on YouTube during this extremely trying time for our society. Our hypothesis was that, like the comments containing the word “people”, humorous comments would be highly-upvoted and positively-received by the users on YouTube. Ultimately, examining these collected comments is valuable to us because it shows us the way average people are coping with the pandemic and how they communicate during a global emergency.

**Explanation of Data**

In order to find the most commonly used words in the comments section on coronavirus-related videos, we first had to collect comments to compare. We went through all 60 videos that we previously identified in our YouTube Analysis as directly related to coronavirus and copied the top three comments that we saw, totaling to 180 comments. We put all of them into a table so that we could see which comments corresponded to which video, as some of the top comments related directly to the video that they were under rather than to coronavirus. After putting all the comments into a text file, we were able to generate a word cloud in order to visualize the top words and see if they met our hypothesis of the top words being “coronavirus”, “virus”, or “quarantine”. From there, after we found that the most popular word was “people”, we wanted to see in which context the word “people” was used, so we searched for all comments containing the word people, and went through them individually, splitting them into 7 categories including Social Awareness Joke, Political Statement, Anecdote, Social Awareness Statement, Comment on Video, Joke, and Positivity. The top three categories were Social Awareness Joke, Political Statement, and Anecdote. We found that a lot of people were making jokes, but many of the jokes were shedding light on the severity of the situation as well as the importance of social distancing and enforcing the quarantine, so we labeled that category Social Awareness Jokes. There were some personal stories from people all over the world, labeled Anecdotes. There were also several people angered by the lack of enforcement and leadership displayed by politicians, so we labeled this category Political Statement. This helped us determine the public’s reactions and feelings towards the virus and outbreak. From there, we were able to go through all the comments and record the total number of comments in each category as well as the total number of likes each comment got in order to further understand the public’s reactions to this outbreak.

**Project Steps**

**Step 1: Video Data Collection**

|  |  |
| --- | --- |
| (Figure 1) The YouTube Trending page | (Figure 2) Table of Top 15 YouTube Trending videos  Link: [YouTube Trending Dataset](https://docs.google.com/document/d/17Fseyp6EAlSGZSEJfbVsO1-NqkfcXmm3iE6ZQVgjBfM/edit?usp=sharing) |

The process of collecting our data starts on the YouTube Trending page. Figure 1 shows an example of the Trending page, which is frequently updated with the videos that are most popular on the site. Factors that determine a video's popularity and rank on the Trending page include number of views, comments, and like, and the rate at which the video is gaining views.

In order to record daily video rankings from March 6 to March 22, we first created the table shown in Figure 2, which records the rankings of the first fifteen videos on the YouTube Trending Page, as well as the video title, the category it was tagged under, and the number of views it got. Figure 2 shows a sample table we filled out on March 15th.

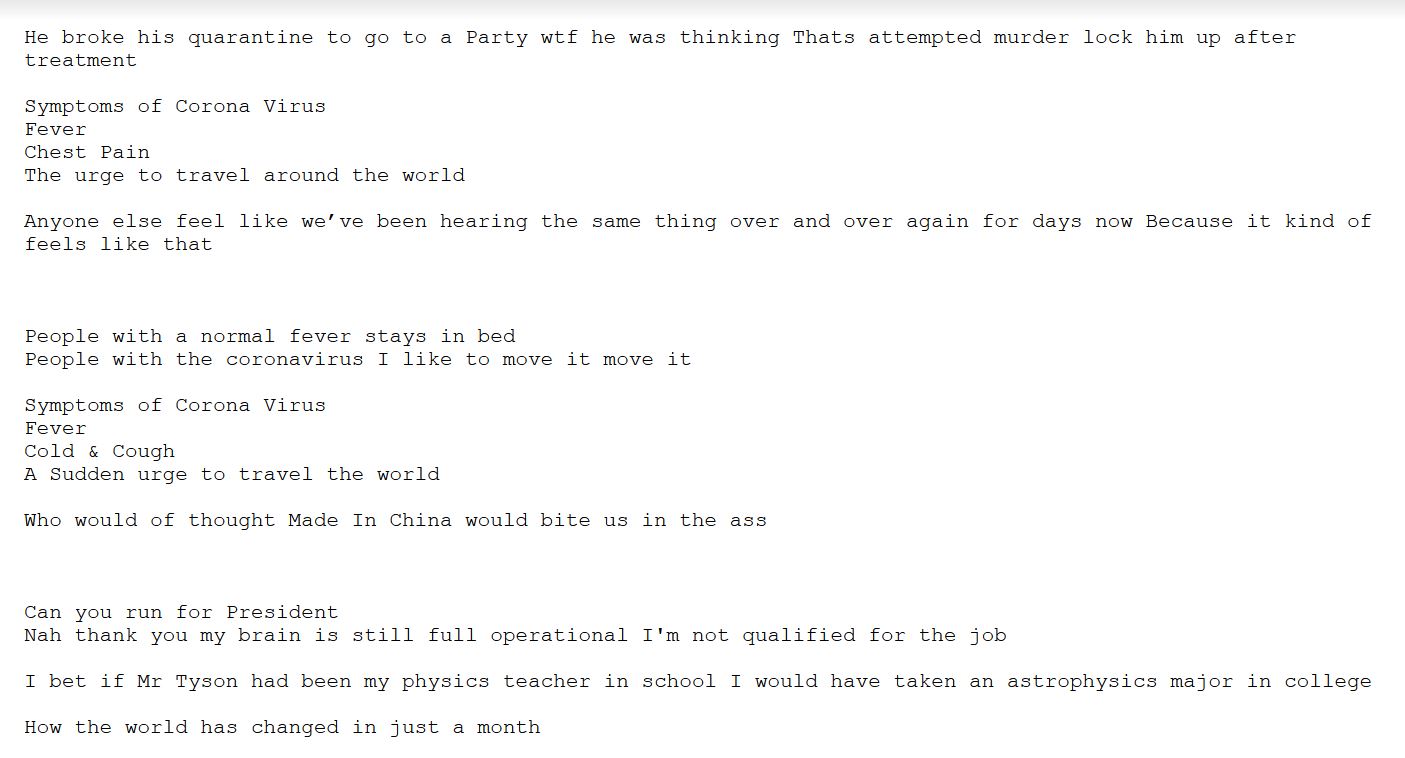
When we finished collecting the video information, we went through the table and identified the videos that were related to the coronavirus. The videos highlighted with yellow are videos with the word “coronavirus” in the title, and the videos highlighted with green are the videos still related to coronavirus without the term in the title.

**Step 2: Comment Data Collection**

|  |  |
| --- | --- |
| (Figure 3) Comments on a coronavirus-related YouTube video | (Figure 4) Table of comments from top 15 YouTube Trending videos  Link: [YouTube Trending Dataset](https://docs.google.com/document/d/17Fseyp6EAlSGZSEJfbVsO1-NqkfcXmm3iE6ZQVgjBfM/edit?usp=sharing) |

After identifying the coronavirus-related videos, we looked at the top three comments from each video. As shown in Figure 3, each comment on YouTube has an “upvote” and “downvote” option. Comments are usually sorted based on the amount of upvotes they receive from users. This means that comments with the most upvotes will be displayed at the top of the comment section in descending order.  
 We collected the top three comments from each video because we wanted to use the most popular and highly-rated comments for our data collection. We copied each comment and stored it in a table along with the title of the associated video. With 60 coronavirus-related videos, we collected a total of 180 comments.

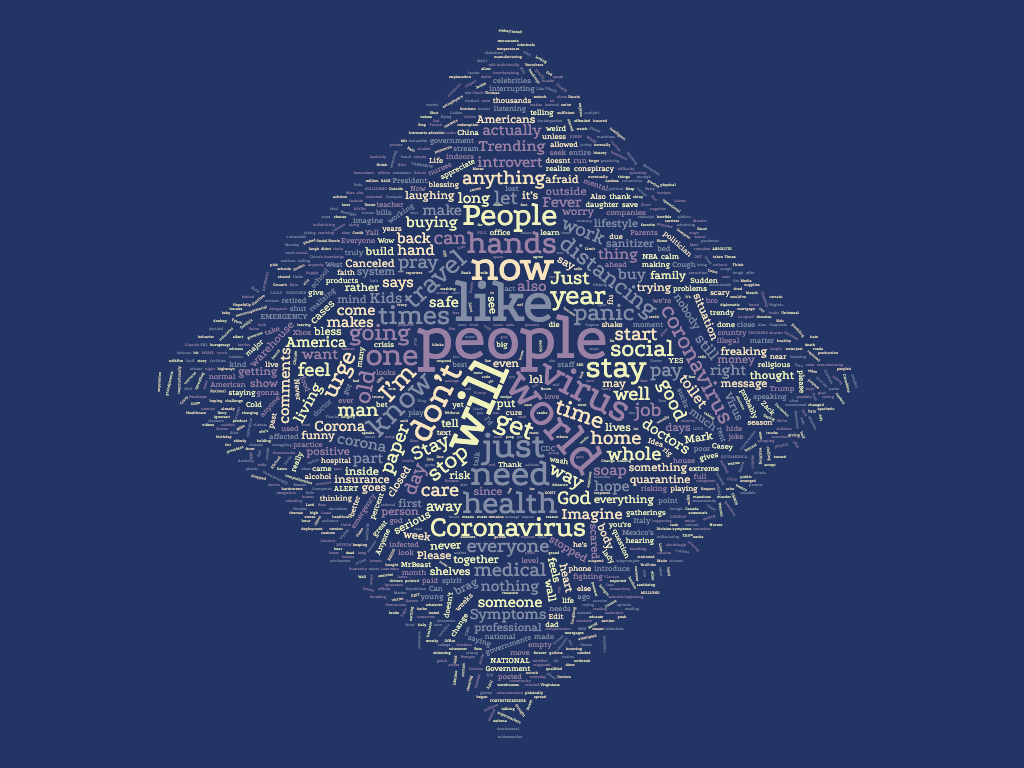
**Step 3: Creating a file of comments**

****

(Figure 5) Text file containing all YouTube comments from selected coronavirus-related videos

After completing our comment data collection, we created the text file shown in Figure 5. The file contains the text from each of the 180 comments we collected from the 60 coronavirus-related videos. We intended to use this file to generate a word cloud, which would visualize the most common words found throughout the comments of the videos. In order to ensure accuracy of our data, we eliminated all punctuation and symbols from this file. This would reduce the number of duplicate words in our word cloud and ultimately provide a more accurate visualization of our data.

**Step 4: Generating a word cloud**

****

(Figure 6) Word cloud generated from top words found in coronavirus-related videos

In order to determine the most frequently used words in the comments, we generated a word cloud. We used WordClouds.com to create the image shown in Figure 6. This word cloud provides a visualization of the most common words found throughout the collected comments by scaling words based on the number of instances of that word. The more instances of the word in the comment section, the larger its scale in the word cloud.

Using this word cloud, we were able to identify the word “people” as the most commonly used word in our collected comments. We also identified words such as “virus”, “stay”, “coronavirus”, and “world” as popular terms used in the comments. In order to keep our data as accurate as possible, we decided to exclude words that we determined to be filler words, including “will”, “like”, and “now”.

**Step 5: Counting instances of top words**

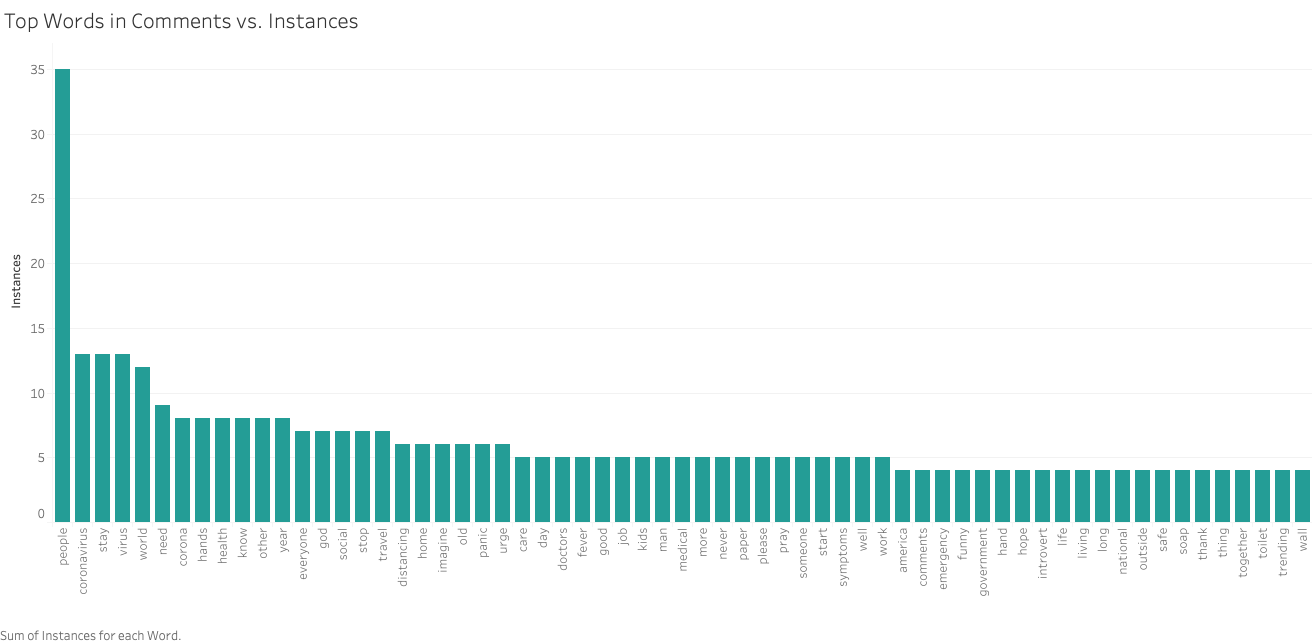
****

(Figure 7) Spreadsheet of top words and instances

Link: [Top Words in Comments](https://docs.google.com/spreadsheets/d/1ZMnyATg7AZ2zgH9lOSI8pvUeb16VvZZVBFAINUkMm5w/edit?usp=sharing)

After creating the visualization of our comment data, we wanted to determine the number of instances for each of the most popular words. As shown in Figure 7, we created a list of the most popular words, excluding filler words such as “will”, “like”, “one”, and “now”. We felt that these words would not contribute any valuable information to our research question.

Using this list, we obtained the number of instances of each word by searching our file containing the text from the collected comments. We created a spreadsheet that listed each word along with its number of instances throughout the comments. By looking at this figure, we can see that the most popular word, “people”, is mentioned 35 times throughout the comments of our 60 coronavirus-related videos.



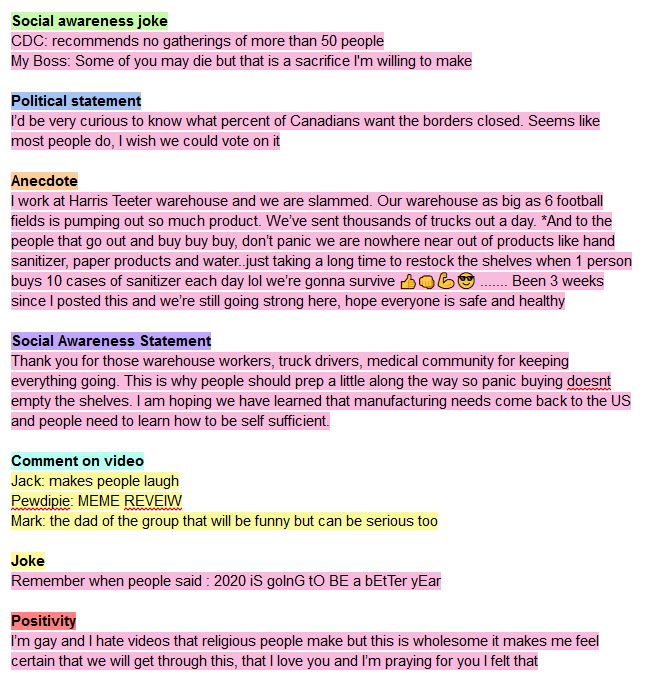
(Figure 8) Bar graph of instances of top words

Our spreadsheet shown in Figure 7 was used to create the bar graph shown in Figure 8. This graph provides another visualization of the most popular words mentioned in our collected comments. By looking at this graph, we can see that the word “people” is used more than twice as frequently as any other word, excluding filler words.

Other popular words include: “virus”, “stay”, “coronavirus” with 13 instances each; “world” with 12 instances; “need” with 9 instances; “year”, “other”, “know”, “health”, “hands”, “corona” with 8 instances each; “travel”, “stop”, “social”, “god”, “everyone” with 7 instances each; “urge”, “panic”, “old”, “imagine”, “home”, “distancing” with 6 instances each.

Going into this research project, we had hypothesized that the word “coronavirus” would be the most popular word. However, “people” is the most popular word, even when adding up the instances of the words “coronavirus”, “corona”, and “virus”. We found this interesting, so we decided to look deeper into the context in which the word “people” was being used.

**Step 6: Determining the topic of each comment containing “people”**

****

(Figure 9) Categorization of comments containing “people”

Link: [YouTube Comments](https://docs.google.com/document/d/1PbL32br_GmFIs7j1ULT7SMVeyvjxzUr6V0_V3AmMSGI/edit?usp=sharing)

To determine the context in which the word “people” was being used, we went back to our collected comments. First, we highlighted all comments containing the word “people” in pink, as shown in Figure 9. Although there are 35 instances of the word, there are 28 comments containing the word “people” because some comments used the word more than once. Then, we highlighted any comments that were unrelated to the coronavirus outbreak in yellow.

Next, we read through all of the comments and looked for common themes between them. We summarized each comment with a description that helped us determine the category it belonged to. We found that there were similarities between a lot of these comments, and we came up with seven categories which the comments would be sorted into.



(Figure 10) Spreadsheet of “people” comment categories

Link: ["People" Comment Categories](https://docs.google.com/spreadsheets/d/1S2B9gB9CZjEpsqrIKEAUHC85sv0paiqNetOXF1TqY1M/edit?usp=sharing)

After reading through and sorting our collected comments containing the word “people”, we counted the number of comments in each category. We created the spreadsheet shown in Figure 10 which listed each of the seven categories along with the number of comments in each.

The most popular category of comments containing the word “people” is “Social awareness joke” with 11 comments. In this category, we see users commenting on things like social distancing, quarantine rules, and supply shortages in a humorous way. All of these comments are related to the coronavirus outbreak.

The next most popular category is “Political statement” with 5 comments. This category contains comments about things like closing country borders, the economy, and President Trump. All of these comments are related to the coronavirus outbreak.

The “Anecdote” category has 3 comments. The comments in this category are personal stories shared by the comment poster. Two of the posters are essential workers sharing information about supply shortages. The third poster is a teenager from Italy sharing his experience during the coronavirus outbreak. All of these comments are related to the coronavirus outbreak.

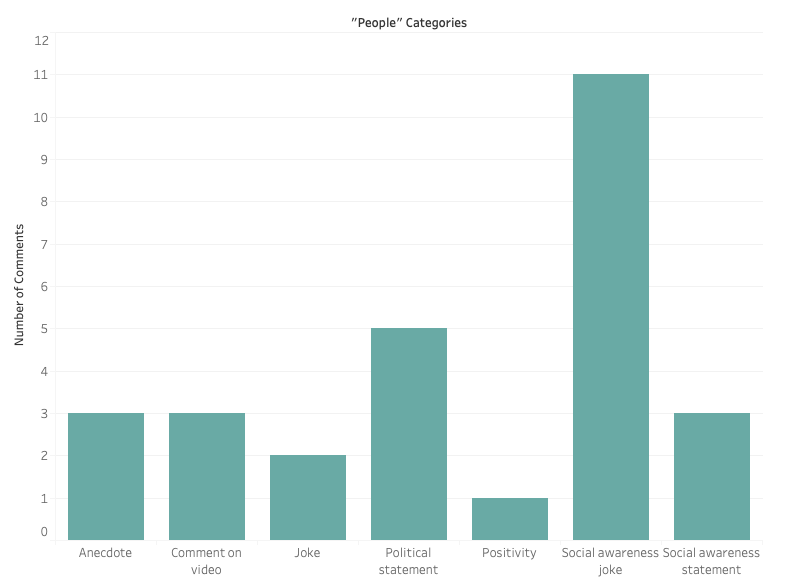
The “Social awareness statement” category has 3 comments. These comments cover topics such as washing hands, spreading the virus, and preparing for the pandemic. All of these comments are related to the coronavirus outbreak.

The “Comment on video” category has 3 comments. Although all three of these comments are posted to coronavirus-related videos, none of the comments are related to the coronavirus. Instead, they comment on the specific people in the videos without referring to the coronavirus in any way.

The “Joke” category has 2 comments. Both of these jokes are loosely related to the coronavirus. The first comment is a joke about the irony of staying away from “positive” people instead of “negative” people. The second comment jokes about everyone’s expectations for 2020 to be a better year than 2019.

The “Positivity” category has 1 comment. In this comment, the poster talks about how he was sure we would get through the pandemic. He also added that he is praying for everyone. This comment is related to the coronavirus outbreak.

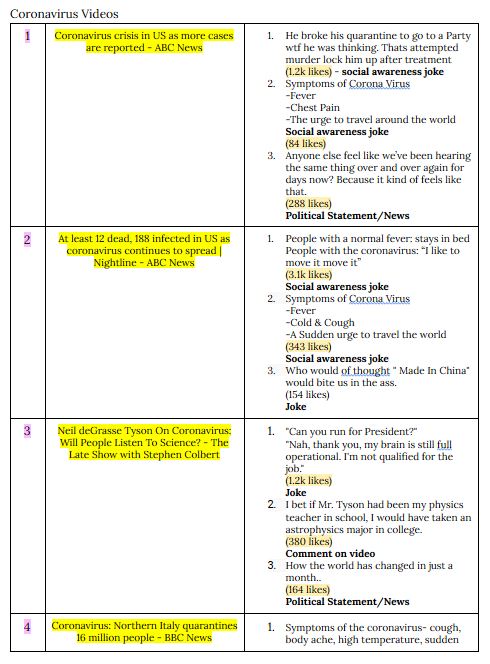
Overall, we determined that 25 of the 28 collected comments were related to the coronavirus outbreak.



(Figure 11) Graph of “people” comment categories

The graph shown in Figure 11 is a visual representation of our spreadsheet shown in Figure 10. By looking at this graph, we can see that the “Social awareness joke” category contains more than twice as many comments than any other category. This means that the most popular type of comment containing the word “people” is a social awareness joke. Some of the common topics of these jokes include social distancing, quarantine rules, and supply shortages.

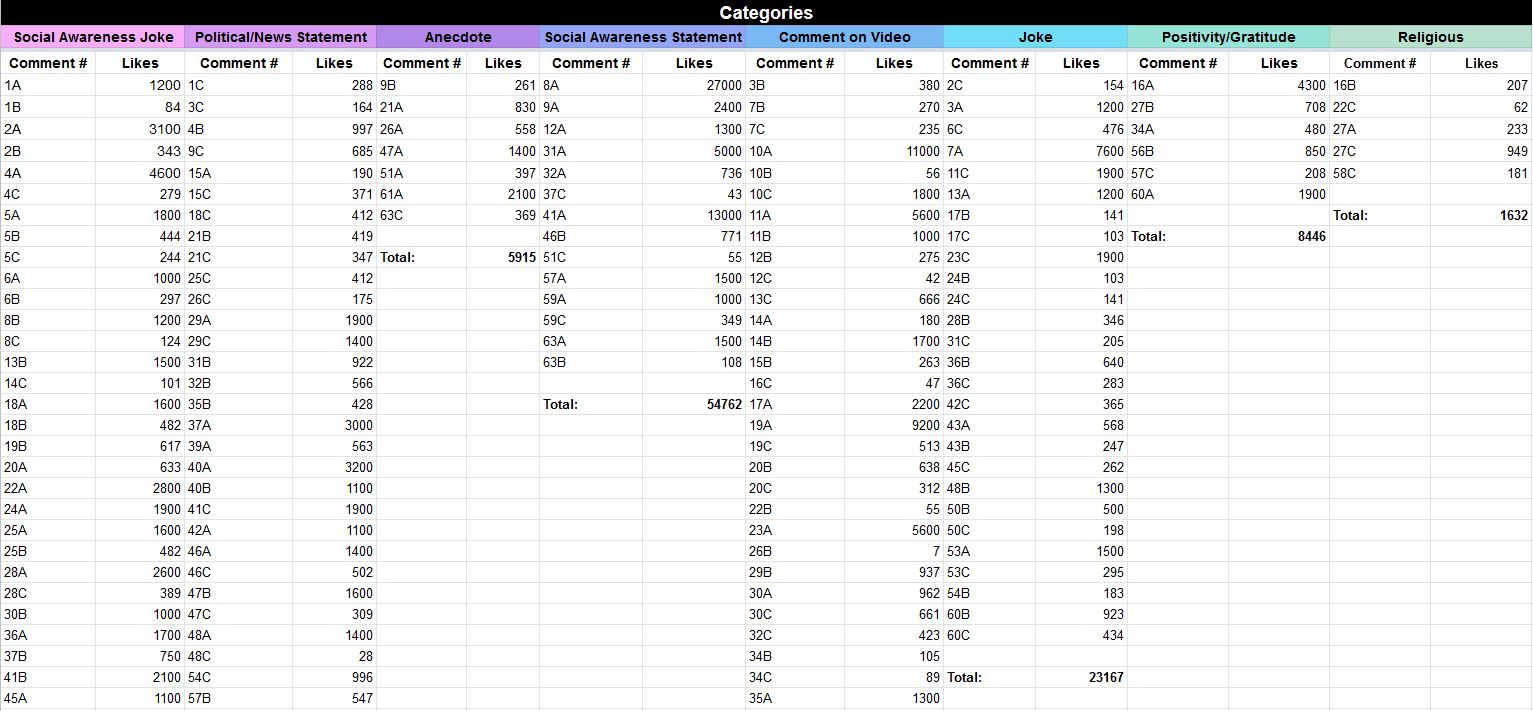
**Step 7: Collecting upvote data and categorizing all comments**

****

(Figure 12) Document of comments with upvote data and assigned category

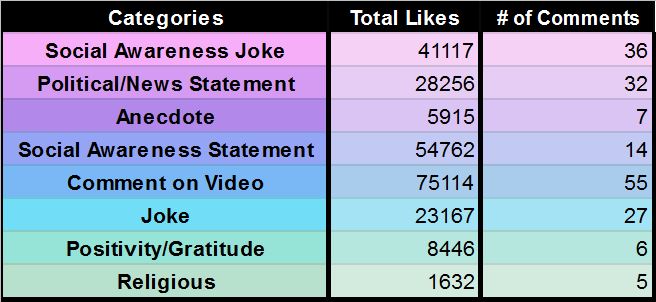
Link: [YouTube Trending Dataset](https://docs.google.com/document/d/17Fseyp6EAlSGZSEJfbVsO1-NqkfcXmm3iE6ZQVgjBfM/edit?usp=sharing)

To answer our third research question of “What is the most upvoted type of comment on coronavirus-related videos featured on the YouTube Trending Page during the coronavirus outbreak?”, we went back to our document containing all of our collected comments. For each comment, we collected the number of upvoted the comment received, which is highlighted below each comment in Figure 12. After collecting the number of upvotes for each comment, we read through the comments and assigned it to a category using the same categorization process as the comments containing the word “people”. We had only one new category called “Religious”, which contains comments about God, praying, or religion in general. These comments were typically found on videos related to religion.

****(Figure 13) Spreadsheet of each comment’s upvotes and category assignment

Link: [Comments vs. Categories and Views](https://docs.google.com/spreadsheets/d/1LAwsBH6qyWeHsOsMWZd4yUdlETaXUtpWoc7R-cHetdQ/edit#gid=0)

After categorizing each comment, we created the spreadsheet shown in Figure 13. Under each category, the comment is listed with the number of upvotes it received. Once we had gone through all of the 180 comments and assigned them each to one of our eight categories, we calculated the total number of upvotes each category received. We also recorded the number of comments each category contained. **Step 8: Calculating comment category totals**



(Figure 14) Spreadsheet of comment category upvote totals

Link: [Comments vs. Categories and Views](https://docs.google.com/spreadsheets/d/1LAwsBH6qyWeHsOsMWZd4yUdlETaXUtpWoc7R-cHetdQ/edit#gid=0)

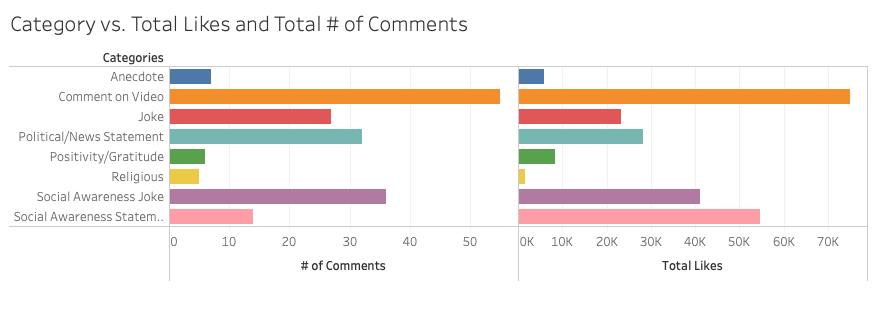
Using the spreadsheet in Figure 13, we created another spreadsheet that was much simpler to read and contained the data that we were really looking for. Figure 14 shows a spreadsheet of the total upvotes (likes) each comment category received along with the number of comments in each category.

We found that the comment category that received the most upvotes was “Comment on Video” with 75,114 total upvotes. Comments in this category are users’ thoughts or opinions on a specific part of the video. They may be referencing a specific person, line, or topic within that video, and these comments generally do not make sense outside of the context of that video. This category also had the most number of comments out of any category with 55 comments.

The next most upvoted category was “Social Awareness Statement” with 54,762 total upvotes. These are comments that talk about topics such as social distancing, quarantine, washing hands, and other lockdown guidelines that have been put into place during the coronavirus outbreak. It is interesting that this category received so many upvotes with one of the least amount of comments with 14 comments. We found that this category had some of the most liked comments out of all of the collected comments, with the most upvoted comment having 27,000 upvotes.

The “Social Awareness Joke” category received a total of 41,117 upvotes with 36 comments. These are comments about social awareness topics such as social distancing or quarantine, but they are more humorous than the social awareness statements. This is the category that was the most popular for comments containing the word “people”, but across all collected comments it is the third most popular.

Thinking back to our original hypothesis, we examined both the “Social Awareness Joke” and “Joke” categories. The comments in the “Joke” category are more broad than “Social Awareness Joke” because they contain all other jokes that do not fit into that specific category. However, if we combine these categories, we can see how popular joking in comments is in general on these YouTube videos. These two categories have a combined total of 64,284 upvotes across 63 comments. This makes all “Jokes” the second most upvoted comment category across these coronavirus-related videos. It also makes it the most common type of comment because it contains a combined total of more comments than any other category. While this is not exactly what we had originally hypothesized, these results are similar to what we expected to see with this research question. Like the comments containing the word “people”, humorous comments and jokes are among the most popular types of comments across these coronavirus-related videos.

(Figure 15) Graph of comment category upvote totals

The graphs shown in Figure 15 is a visual representation of the spreadsheet shown in Figure 14. The graph on the left shows the number of comments in each category. The graph on the right shows the total upvotes received by each category. By looking at these graphs, we can clearly see that the “Comment on Video” category contains the most comments and received the most upvotes.

**Results**

By researching the number of coronavirus related videos on the YouTube trending page, we were able to find how widespread and dangerous this virus really could be. Looking specifically at the 180 top comments of these videos allowed us to see how the public was reacting to the virus, and how people are learning to live with this sudden shift in lifestyle. Our first research question questioned what the most commonly words in the top comments that we collected would be, and we had hypothesized that because the videos relate specifically to the coronavirus outbreak, the most commonly used words might be “coronavirus”, “virus”, or “quarantine”. However, after generating a word cloud and counting instances of the most popular words, excluding filler words like “will” and “like”, we found that the most popular word was “people”, mentioned 35 times throughout these top comments. Because this result was different from the one we hypothesized, we decided to go further into our research to find the context in which the word “people” was used in order to determine reasons why the word was so popular.

After going into each of the comments containing the word “people”, we found that there were 28 separate comments in total. From there, we went into each one and categorized them based on common themes that we found. Examples of comments from the three top categories are found below.

**“People” Categories**

**Social Awareness Joke: 11 comments**

The top category that we found was “Social Awareness Joke”, where users were writing humorous jokes whilst still shedding light on social distancing and quarantine.

*“People with flu: ima stay home; People with corona: I will travel across the land searching far and wide” --on the video titled, “Fast moving developments in the coronavirus outbreak” by ABC News*

**Political Statement: 5 comments**

The next most popular category was “Political Statement”. This category discussed the United States Administration, closing country borders, and the state of the economy. We found that most of these comments felt negatively about the way the United States is handling the virus, especially in comparison with other countries that have much lower death rates.

*“I’d be very curious to know what percent of Canadians want the borders closed. Seems like most people do, I wish we could vote on it “ --on the video titled, “Global National: March 14, 2020 | New cases surface in Canada; Trump travel ban expands” by Global News*

**Anecdote: 3 comments**

The third most popular category was titled “Anecdote”, with personal stories from all around the world. Two of these stories were from essential workers sharing vital information about supply shortages, and one was from a teenager in Italy sharing his experience surviving the outbreak.

*“I work at Harris Teeter warehouse and we are slammed. Our warehouse as big as 6 football fields is pumping out so much product. We’ve sent thousands of trucks out a day. \*And to the people that go out and buy buy buy, don’t panic we are nowhere near out of products like hand sanitizer, paper products and water..just taking a long time to restock the shelves when 1 person buys 10 cases of sanitizer each day lol we’re gonna survive....... Been 3 weeks since I posted this and we’re still going strong here, hope everyone is safe and healthy” --on the video titled, “Trump, White House Coronavirus Task Force hold a press briefing” by Fox News*

From there, we went through all of our categories to see which categories accumulated the most comments and from there, which of those categories accumulated the most points. This allowed us to see the manner in which the public is communicating with each other and what methods they are using to stay in touch with one another. From here, we saw that the “Comment on Video” category was the most popular in terms of both number of comments and total likes, as there were 75,114 total likes for the 55 comments that we saw. This showed us that a lot of people were taking heed of what was being said in the video, either through jokes or other types of comments that are not always coronavirus-related, but still popular to the general public nonetheless. The next popular category in terms of likes was Social Awareness Statement, with 54,762 likes. This category did not contain the second most number of comments, and we think this shows us that although there weren’t many comments in this category, a lot of them were very powerful, resulting in many likes. One of the comments even got 27,000 likes, which surprised both of us. The second most popular video in terms of number of comments was Social Awareness Joke, with 36 total comments, which showed us that many of the ways that people tried to gain popularity and relate to others in the comments section was by comedy and humor.

Although we saw that the top word and top category was not directly related to the coronavirus as we had predicted, we found that the public was speaking more about coming together, making jokes about those that should be taking the quarantine more seriously, and applauding those who are essential workers and putting their lives on the line. This showed us that not a lot of people were discussing the virus specifically, but were using the comments section to interact with one another, talk about their own government and share their opinions, and share experiences with people across the country, letting them know that we are all going through this together, which we found uplifting and necessary during such a vital time like this.

**References**

“Wordcoulds.com” (2020). Retrieved from https://www.wordclouds.com/

“YouTube Trending” (2020). Retrieved from https://www.youtube.com/feed/trending