McKenzie Skrastins

Dr. Ford

DIDA-280A Twitter Analysis

Due: March 24, 2024

As a child, one of my favorite T.V. shows was, "*Avatar: The Last Airbender*". Consequently, when the 2024 live action show came out, I was very excited. I'm an avid tik tok, instagram, and snapchat user, and recently, my feed on the former two social media platforms have been filled with ATLA content. I almost never use Twitter, and so when I found out about this assignment, I became curious as to what Twitter users had been saying about ATLA. I knew I wanted "*Avatar: The Last Airbender*" to be the topic I used.

People are opinionated, and people love to share their opinions with others. The rise of social media has allowed the population to freely express their thoughts to the whole world. Not surprisingly, when a new show, especially one with a lot of controversy and hype surrounding it, is released, many people flock to social media to share their judgments. *Avatar: The Last Airbender* is an example of this; many people took to social media when it was first announced, right before it came out, and people have continued to talk about the show even after it was released.

The keywords, "*Avatar*", "*Airbender*", "*ATLA*", "*Aang*", "*Katara*", "*Sokka*", "*Zuko*", "*Azula*", "*Toph*", "*Iroh*", "*Ozai*", "*Appa*", "*Waterbender*", "*Firebender*", and "*Earthbender*" were used when using Twitter's Advanced Search. The hashtags, "*#ATLA*", "*#AvatarTheLastAirbende*r", "*#Aang*", "*#Katara*", "*#Sokka*", "*#Zuko*", "*#Azula*", "*#Toph*", "*#Iroh*", and "*#Ozai*" were also used. The data was collected from Twitter and then uploaded to

Communalytic. Various visualizations and conclusions were drawn using the tools on this website.

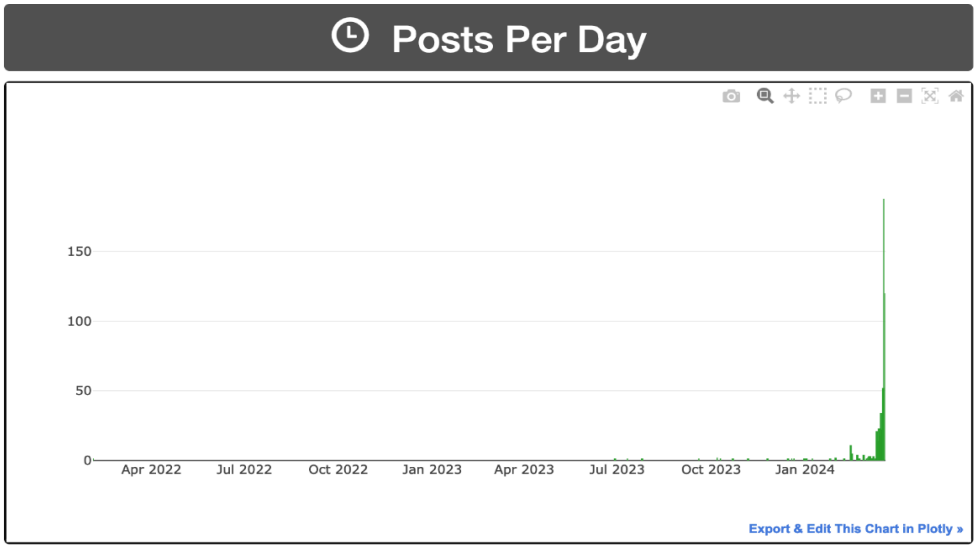
One visualization that was made was a word cloud; this image demonstrated the most common words found in the Twitter scrape. See figure 1 below;

Figure 1: ATLA Twitter Scrape Word Cloud

The only aspect of this visualization that surprised me was the word "*zuko*" being smaller than words such as "*free*" and "*like*". The size of each word correlated to how prominent the word was in the web scrape; the larger the word, the more often it appeared in the Twitter search. Zuko was the name of a character in the show, and so it would have made sense for the name to be as large as the other character names, but it was not.

Another plot that was generated with Communalytic was a Posts Per Day histogram. See Figure 2;

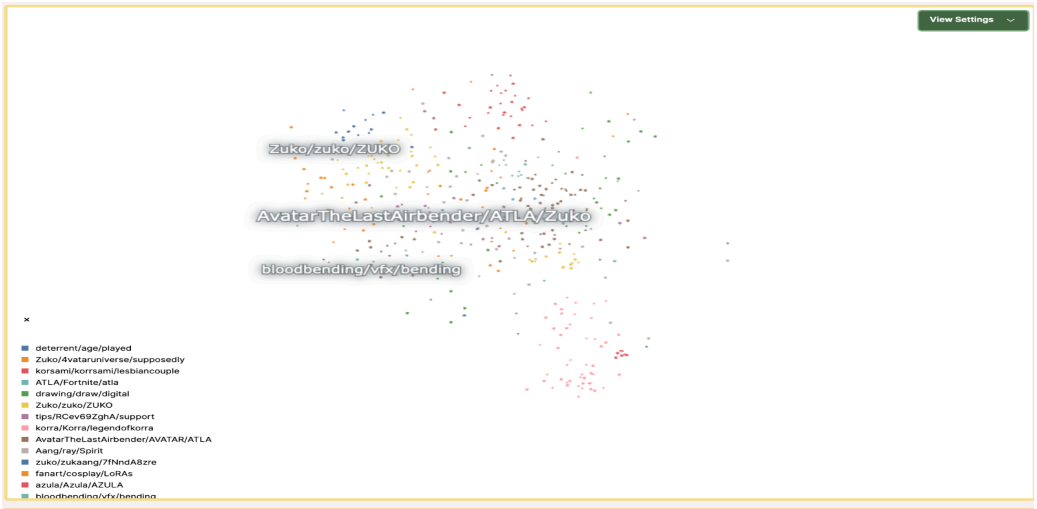
Figure 2: Histogram of Posts Per Day vs Users



This graph shows a peak in the amount of users posting, with the keywords and hashtags mentioned above, in late January. This was most likely due to the live action T.V. show coming out in February. The graph also showed small peaks around July 2023, followed by a flat line. This made me wonder if some news about the live action show was announced in July, causing fans to post on Twitter.

A topic analysis was performed on the data, and one attribute I found interesting was that although "*Zuko*" was not prominent in the word cloud, it is in the topic analysis. See figure 3;

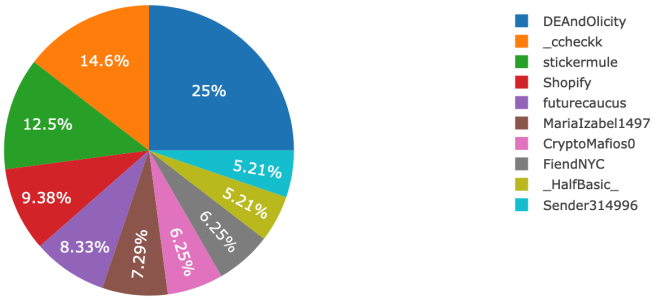
Figure 3: Topic Analysis Based On the Keywords and Hashtags Specified



I would love to explore the differences between these two types of analyses, and figure out what has caused the gap in the importance of these words.

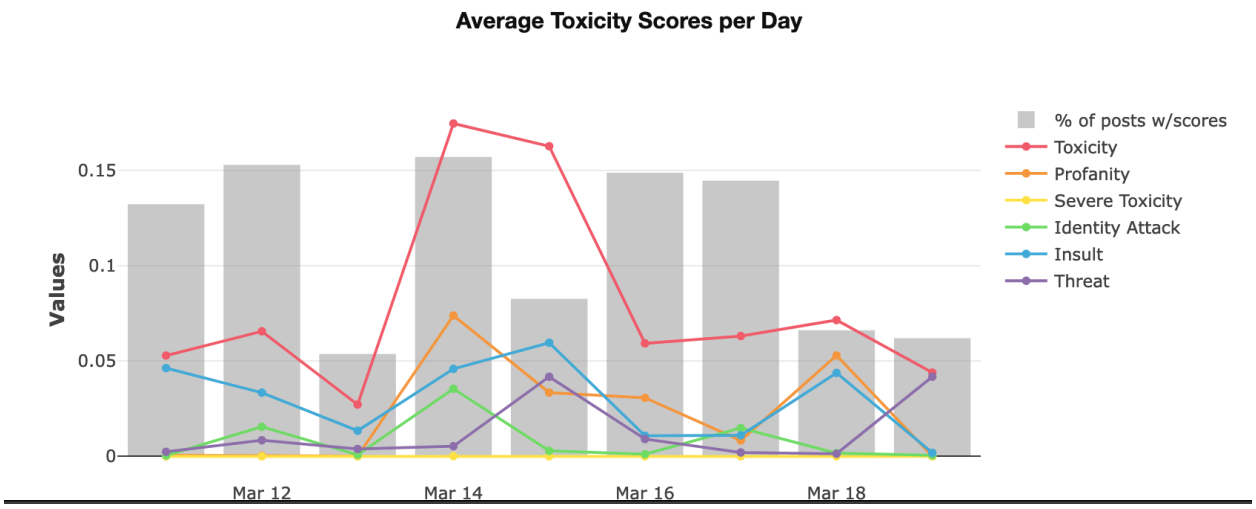
The following pie chart showed the top ten posters using any of the keywords or hashtags mentioned previously. An in depth analysis of "*DEAndOlicity*" was performed, as they were found to be the top poster.

Figure 4: Pie Chart of Top Posters Using Keywords and Hashtags Specified



The following chart visualizes DEAndOlicity's toxicity over the past week. See figure 5;

Figure 5: Bar Plot of DEAndOlicity's Toxicity Scores Per Day, Recorded Over a Week



On average, DEAndOlicity's toxicity scores were low, with an overall average toxicity score having been 0.02667. Furthermore, DEAndOlicity's highest toxicity score was in the "Insult" category; see Figure 6;

Figure 6: DEAndOlicity's Toxicity Scores

	Average for Dataset	Highest value
Toxicity		
	0.08473	0.99700
Severe Toxicity		
	0.00005	0.00213
Identity Attack		
	0.01076	0.64304
Insult		
	0.0303	0.99713
Profanity		
	0.02377	0.90877
Threat		
	0.01043	0.72416
Overall Toxicity		
	0.02667	--

DEAndOlicity's highest scoring Insult was, "*RT @AdoraSchuyler: you're such an idiot ☺.*" The smiley face leads me to believe that DEAndOlicity was being sarcastic with their comment. Based on context, DEAndOlicity's second highest Insult did not seem toxic, "*RT @jinxedhater: i don't think anyone should be this handsome it's, quite frankly, ridiculous and offensive….*" I found it interesting that Communalytic detected these comments as insults, when in actuality they are harmless.

A sentiment analysis was also performed on DEAndOlicity's posts, and the following chart was constructed;

Figure 7: DEAndOlicity's Posts Sentiment Ratings

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	243	69 (28.40%)	43 (17.70%)	131 (53.91%)
TextBlob (English/EN)	243	43 (17.70%)	64 (26.34%)	136 (55.97%)
TextBlob (German/DE)	1	0 (0.00%)	1 (100.00%)	0 (0.00%)

This graph showed that slightly over half of DEAndOlicity's posts were found to be positive. Even though a decent amount were found to be negative, the findings from the toxicity report may be applied here; Communalytic detects harmless comments as toxic, and so most of the negative sentiments were most likely harmless as well.

What I found most interesting about this project was my realization about how easy it was to collect data about people from the internet. I do not know who DEAndOlicity is, and although they are anonymous on Twitter, I found out several oddly specific things about them because of this project. I recently deleted all social media apps from my phone, and this acts as another motive to keep them off my phone; if I could easily extract information from the internet, how much information could professionals extract?