

***A Social Media Analysis Case Study:
Scraping and Analyzing Subreddits related to Disney***

**Jonathon Scroggins
DSC 680 Fall 2021
<https://github.com/JDScroggins>**

Background

Social media as defined by Merriam-Webster is “forms of electronic communication (such as websites for social networking and microblogging) through which users create online communities to share information, ideas, personal messages, and other content (such as videos)”. The biggest keys to social media are the ability to communicate and build community. Society today is relying on social media more than ever before for information on a variety of topics, from current events to how to get the most out of your pantry space. Social media has been around since the first people drew on caves. However, in the late 1990s to the early 2000s is when online social media really took off. Social media sites that provided users the ability to blog became a popular way to communicate with the world. Sites like Facebook, MySpace, Twitter and YouTube gave users the ability to build a community through posts and comments. In the year 2005, two guys from the University of Virginia came up with a site called Reddit. Reddit is a site where people can create accounts, post videos or commentary on the latest trend and they can comment on other people’s post. This creates a large community that can post and comment frequently. Reddit can be broken down into pages called subreddits. Subreddits are pages that are about a particular topic and can be managed by an administrator and create a community of followers that can post and comment. There are many posts that turn into conversations about a particular topic that a post is about. For example, this project will look at two subreddits that are related to Disney. One of the subreddits is r/DisneyWorld, which has over 36,000 followers and the other subreddit is r/disney which has over 613,000 followers.

Problem Statement

The analysis I will perform will be composed of multiple layers, however it will all be focused on the problem statement: How can a Social Media Analysis be beneficial a business?

Methods

To perform a social media analysis, one must first do a process called web scraping to get the data from the website. For this project, the two subreddits mentioned above will be scrapped. Web scraping of Reddit allows the analyst to pull posts and comments from a subreddit page. To do this in Python like this project, analysts use the PRAW package. First, like most cites that would be scrapped, one must set up an account and get access to the data on the subreddits. Once that is done the data can be pulled from the cite. In the case of this project, the top ten hot topics were pulled and put into a dataframe like the one below:

	title	score	id	subreddit	url	num_comments	body	created
0	Monthly Self-Promotion Thread (October 2021)	5	pzm15j	WaltDisneyWorld	https://www.reddit.com/r/WaltDisneyWorld/comme...	18	Got a blog, YouTube channel, Etsy page, podcas...	1.633140e+09
1	Weekly FAQ & Reopening Discussion Thread	14	q5nqgm	WaltDisneyWorld	https://www.reddit.com/r/WaltDisneyWorld/comme...	162	Please post all your **general** WDW comments ...	1.633925e+09
2	Just got to our room at the boardwalk and had ...	497	q94w5q	WaltDisneyWorld	https://i.redd.it/f7p6iqd1qqt71.jpg	40		1.634359e+09
3	Very Long Disney Trip Report 9-26 to 10-1, or,...	17	q9evx2	WaltDisneyWorld	https://www.reddit.com/r/WaltDisneyWorld/comme...	3	Before we begin, I would like to apologize for...	1.634400e+09
4	Anyone Taken the Train to WDW?	18	q9ccv8	WaltDisneyWorld	https://www.reddit.com/r/WaltDisneyWorld/comme...	9	I'm taking a trip next year, and considering t...	1.634392e+09
5	Any former cast members here? Just stumbled on...	393	q8wmvg	WaltDisneyWorld	https://i.redd.it/v3hvwq19aot71.jpg	43		1.634329e+09
6	Jellyrolls Is Back!	175	q90qel	WaltDisneyWorld	https://i.imgur.com/70Qqqjx.jpg	32		1.634343e+09
7	One little spark of Inspiration	26	q97yjj	WaltDisneyWorld	https://i.redd.it/g9t7qjdzvrt71.jpg	1		1.634373e+09
8	As requested: Full parts list and assembly ins...	38	q94cr4	WaltDisneyWorld	https://i.redd.it/xt689f3qiq71.png	4		1.634357e+09
9	Best Vegetarian Signature Dining	3	q9gr3y	WaltDisneyWorld	https://www.reddit.com/r/WaltDisneyWorld/comme...	7	Which signature dining experience at WDW do yo...	1.634406e+09

The dataframe above includes the title of the posts, the Reddit score, what subreddit it is from, and the number of comments the post has received. It shows that there is a weekly FAQ post that has 162 comments. For this subreddit we will use that one to see the sentiment people have when asking their questions. The other subreddit we scrapped from produced a dataframe that looks like this:

	title	score	id	subreddit	url	num_comments	body	created
0	Disney's Encanto Teaser Trailer	1054	og6y7b	disney	https://www.youtube.com/watch?v=togmdDHG3Pw	190		1.625750e+09
1	Another warm up sketch. This time, Ariel. Than...	591	q7zl9d	disney	https://i.redd.it/6el5y5em2ft71.jpg	21		1.634218e+09
2	Disney Movies : summarised by Olaf	258	q81lyn	disney	https://youtu.be/2Stzk2aopxs	22		1.634224e+09
3	Oogie looming over California Adventure	1199	q7dod8	disney	https://i.redd.it/iv93cl70l8t71.jpg	16		1.634139e+09
4	It's a work in progress but I'm trying to pain...	198	q7ldjs	disney	https://i.redd.it/6vwf1tw5fat71.jpg	6		1.634161e+09
5	Had a great time at the Oogie Boogie Bash last...	309	q7cyr0	disney	https://i.redd.it/q3ce9k2se8t71.jpg	5		1.634137e+09
6	Thanks Disney! Really enjoyed The Scariest Sto...	30	q7p328	disney	https://i.redd.it/ysqk2brwfbt71.jpg	2		1.634174e+09
7	My fanart of Remy from Ratatouille! What it he...	57	q7hvx8	disney	https://i.redd.it/zjvy745ck9t71.jpg	3		1.634151e+09
8	Found these cups at my grandma's house. She ha...	15	q7ku10	disney	https://www.reddit.com/gallery/q7ku10	5		1.634160e+09
9	My kitten and my 2 year old enjoying Luca	6	q7p5sx	disney	https://www.reddit.com/gallery/q7p5sx	0		1.634174e+09

Again, here you can see what the top topics are right now. The one thing to remember for this type of scrapping is that it can change quickly, so if there is a particular post that is desired to be scrapped, the id column in the dataframe needs to be taken note of because it is the piece that will be used to scrape the comments from that post. For example, we will be looking at the feedback for the new movie Encanto's trailer that has 190 comments, we will also be scraping a post on the feelings that people have for a new product called Genie+ that will be released soon. This post is no longer on the top 10 but the id will still scrape even as more comments are added.

After the initial scrape to find the top posts for the day or to find a particular topic an analyst can look at a particular post and get the comments from that post. After scrapping those comments an analyst can perform a sentiment analysis. A sentiment analysis is used to analyze the overall sentiment from a piece of text, in this case the comments on a post within a subreddit. The sentiment analysis will analyze the words used and classify them as positive, negative, and neutral. This allows a for a gauge on the customer's response to a new product or the overall business. For this project, a VADER sentiment analysis was done. According to the website GeeksforGeeks.com a "VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis is lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media." To perform the sentiment analysis there are a few things

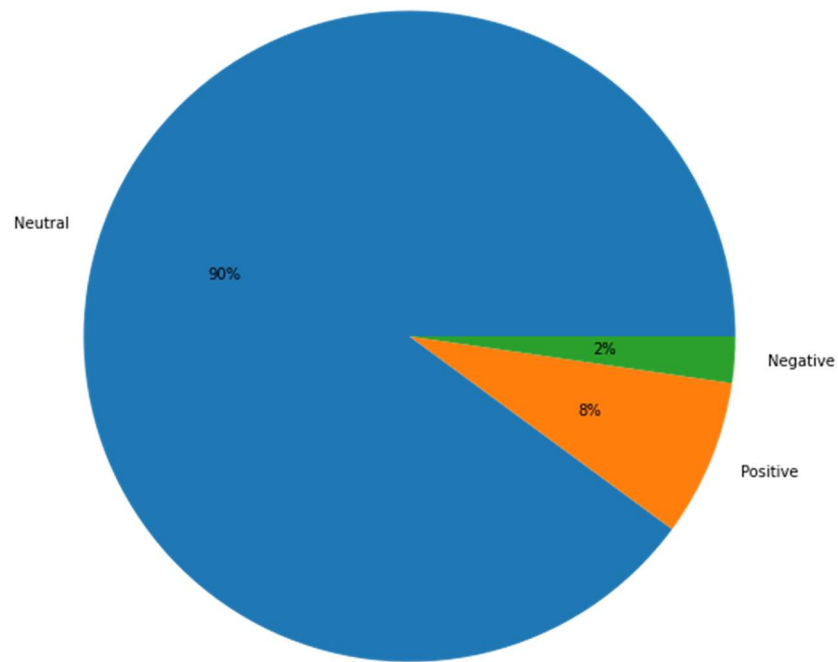
that need to be done to the comments before they are ready to be analyzed. First all the non-words need to be removed, things like emojis and urls. Then we will tokenize the comments, which basically means we will break the comments into individual words. After tokenization, the stop words are removed. Stop words are the most common words that do not affect the sentiment like the articles (the, a, an). Lastly, before running through the analyzer, the words are normalized which basically means to take them down to their root words. Now the comments are ready for the analyzation.

Results

As a reminder for this analysis, three posts were scrapped, the FAQs on the DisneyWorld subreddit, the Encanto trailer release post and the Genie+ announcement reactions on the Disney subreddit. These were chosen to show three ways that a social media analysis can be beneficial to a business. The first shows how a sentiment analysis can give insight into the customer's feelings towards the company, the second shows how feedback can be scrapped when a product is released and the third can show how feedback can be analyzed after the announcement of a new product in development or about to be released.

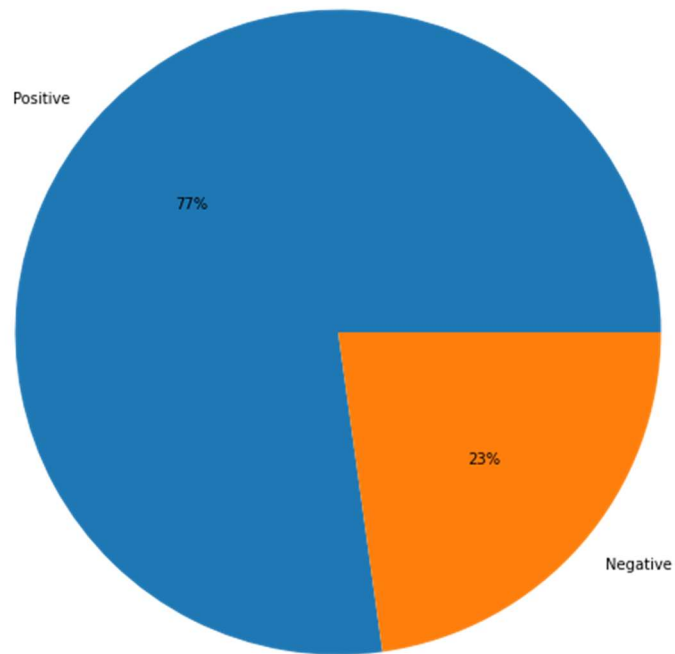
For the FAQs post there were 162 comments that were analyzed. Within those comments there were 3954 words that were used in the sentiment analysis after cleaning the data. The pie chart shows the initial results for the analysis.

Percent of Word Sentiment of Words in Comments



As you can see here, the highest percentage of words in the comments would be labeled as neutral. To get a better idea of whether the sentiment is more positive or negative, we can drop the neutral words as well.

Percent Positive and Negative Words in Comments

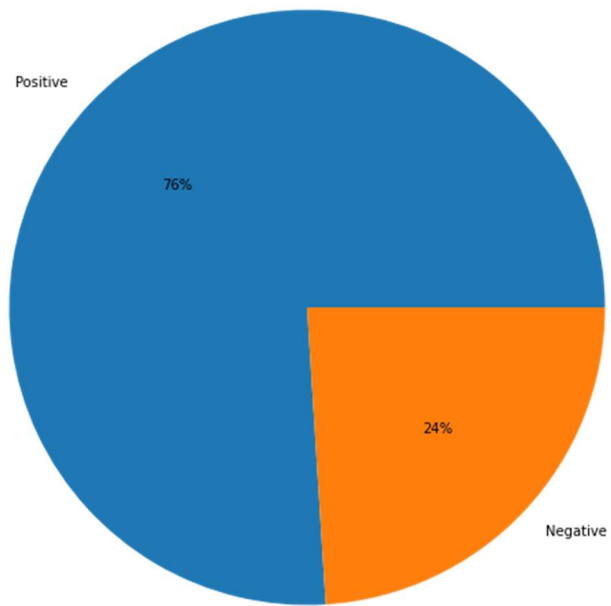


Based on this analysis, the sentiment towards Disney World right now would be considered significantly more positive than negative. To get an idea of the word choices that were used and what words were used most often positively or negatively a word cloud can be created. A word cloud will present all the words used that were part of a particular sentiment and based on the size of the word in the cloud shows the frequency of the word usage, the larger the word the more often it was used.

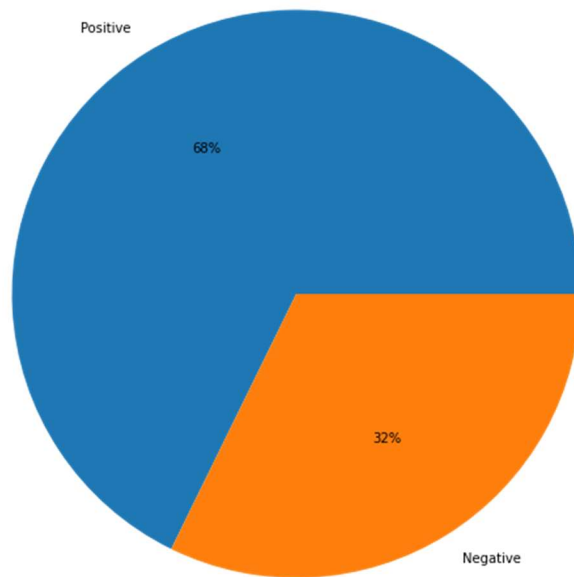
[illegible]

We can do the same analysis technique for the other two posts. The results of the sentiment analysis are represented in these pie charts.

Percent Positive and Negative Words in Comments to Post on Encanto



Percent Positive and Negative Words in Comments to Post on Genie+



Based on the results of the sentiment analysis the overall sentiment towards the Encanto trailer would be significantly positive. However, for the release on Genie+ announcement the results though still leaning toward positive they are not as significantly positive as the others. Word clouds and other visualizations were created for these as well and can be found in the Appendix. Doing the sentiment analysis on Genie+ also showed one of the positives when it comes to a sentiment analysis but also one of the cautions. Let's look at the word clouds for Genie+.

Word Cloud of Positive words about Genie+



Word Cloud of Negative words about Genie+



The first thing that stands out in these two word clouds is the word 'pay'. This would communicate that one of the major things influencing the negative sentiment toward the product is something to do with paying for it. Looking further at this specific product could indicate that since the product that it replaced was free and the customer must pay for it now, is a significant influencing factor. Most of the words that are in the positive word cloud make sense that they are positive, however, a word that stands out would be attraction. This is a word that can have multiple meanings, and, in this case, it might mean a ride or show and a sense of love or emotion. That could be influencing the percentages and one might want to investigate the frequency of that word and if it skews the data at all. In this case, when the word attraction is changed to neutral it does shift the percentages slightly by about 1%, which normally would not be highly significant.

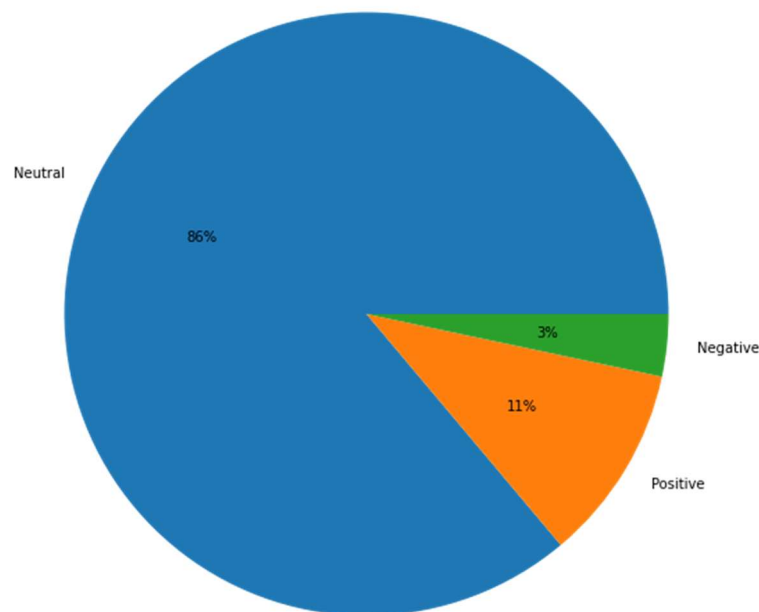
Discussion/Recommendations

Social media is everywhere and as this case study shows, can be a beneficial tool for a business. When a business wants to know the honest reactions of the public to either the business itself or its products, social media is a very beneficial resource. Having a social media presence is almost a requirement for a thriving business today. That presence can allow the business to communicate to the public in ways that it has not been able to before. It can allow for instant feedback and then that feedback can be analyzed for the sentiment as shown in this case study. It can allow a business to make sure that the public's opinions are being heard and used to influence the business's direction. For example, the sentiment analysis showed a very positive reaction to the Encanto trailer which can be an indication that the advertising campaign and strategy is building good buzz towards that movie. Also, the sentiment for the Genie+ product is not as significantly positive, so the business can look at that sentiment and see if there are ways

that things can be adjusted to increase the positive sentiment. One other benefit to an analysis of social media is that it can be tracked daily since it may change daily. While doing this project I noticed that the sentiments did shift slightly each time I scrapped the comments. As more comments are added, the sentiment can shift slightly positive or slightly negative which shows that social media perception can change and is something that can be monitored as marketing campaigns and product releases are done. A social media analysis is beneficial to a business. The business can use the analysis to gauge the reaction of its consumers and adjust accordingly to make sure people are positive toward their products and services.

Appendix

Percent of Word Sentiment of Words in Comments to Post on Encanto



Response	Percentage
Agree	86%
Disagree	9%
Don't know	4%

[illegible]

[illegible]

Complete history of social media: Then and now. Small Business Trends. (2021, January 22). Retrieved October 15, 2021, from <https://smallbiztrends.com/2013/05/the-complete-history-of-social-media-infographic.html>.

Merriam-Webster. (n.d.). *Social Media*. Merriam-Webster. Retrieved October 15, 2021, from <https://www.merriam-webster.com/dictionary/social%20media>.

Python: Sentiment analysis using Vader. GeeksforGeeks. (2021, June 10). Retrieved October 2, 2021, from <https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/>.

Reddit. reddit. (n.d.). Retrieved October 2, 2021, from <https://www.reddit.com/>.

Selvaraj, N. (2020, September 12). *A beginner's guide to sentiment analysis with python*. Medium. Retrieved October 2, 2021, from <https://towardsdatascience.com/a-beginners-guide-to-sentiment-analysis-in-python-95e354ea84f6>.

Tanner, G. (2019, February 12). *Scraping reddit data*. Medium. Retrieved October 2, 2021, from <https://towardsdatascience.com/scraping-reddit-data-1c0af3040768>.

Textblob vs vader for sentiment analysis in Python. Analytics Vidhya. (2021, January 11). Retrieved October 2, 2021, from <https://www.analyticsvidhya.com/blog/2021/01/sentiment-analysis-vader-or-textblob/>.

The history of Reddit. Honor Society. (2020, August 17). Retrieved October 15, 2021, from <https://www.honorsociety.org/articles/history-reddit>. Wikimedia Foundation. (2021, September 22). *Social Media Analytics*. Wikipedia. Retrieved October 2, 2021, from https://en.wikipedia.org/wiki/Social_media_analytics.

The python reddit api wrapper. PRAW. (2021). Retrieved October 2, 2021, from <https://praw.readthedocs.io/en/stable/>.

What is social media analytics? IBM. (n.d.). Retrieved October 15, 2021, from <https://www.ibm.com/topics/social-media-analytics>.