DSC680 - Depression or Cyberbullying Sentiments from Twitter

Joshua Burden

Bellevue University

04/02/2023

Abstract

In the first term project I chose to examine the topic of depression and cyberbullying using various research papers and datasets found from Kaggle and other resources. The purpose would be to extract meaningful insights into the data by exploring the datasets, and to build a meaningful model to find NLP sentiments that can help predict cyberbullying and early warning indications of depression through social media.

Business Problem

Depression is one of the most prevalent mental illnesses that can be very debilitating, disabling, defeating, but is very curable. It is underwhelmingly undiagnosed, and has a tendency to be created or otherwise increased in intensity due to either online harassment, or negative feedback loops perpetuated from various forms of social media such as Facebook, Instagram, or twitter. With the current downsizing and purchases of platforms, there is a combination of conversation being had if social media is free, and safe. I am looking at the data points of social media’s safety as there is plenty of evidence to suggest that social media helps increase the amount of depression that could be encouraged due to the amount of cyber bullying.

Datasets

There are several datasets that were found and believe that they contain enough data to be able to form an educated opinion based on facts of whether social media is a breeding ground of perpetual cyberbullying resulting in an increased bout of depression. One dataset that was found contains a set of over 1,600,000 tweets extracted using the twitter API. There seems to be potential means of creating a classification system of positive, neutral, or negative comments from the tweets.

Another set of data shows another set of tweets that could be used to create an NLP classification system of cyberbullying tweets using 46,017 tweets. These two datasets could be used to build quite a bit of different models such as a text sentiment analysis engine, and exploratory analysis using word clouds, predictive text models, Long Short-term memory networks, and other such Recurrent neural networks.

Summary of Methods

As this is still exploratory, this section will need to be created once the data has been cleaned, parsed, explored, and the methods will need to go here later. There is a high likelihood since there will be sentiment analysis, Keyword extraction, Topic Modeling, and other methods that relate to NLP. With the data that is extracted there is also a high likelihood of using graphs such as histograms, bar charts, and word clouds to drive a more visual aspect of the data.

Ethical Considerations

Depression and cyberbullying examination through datasets using API’s such as Twitter/Reddit/Facebook need to consider the following metrics and ethical questions:

1. Respect for laws and rules.
2. Justice and integrity of the individuals or groups being targeted and those who are persecuting the victims.
3. Honesty and altruism in the dataset manipulations and storytelling.
4. Respecting the mental and bodily autonomy and confidentiality of those using the platforms.

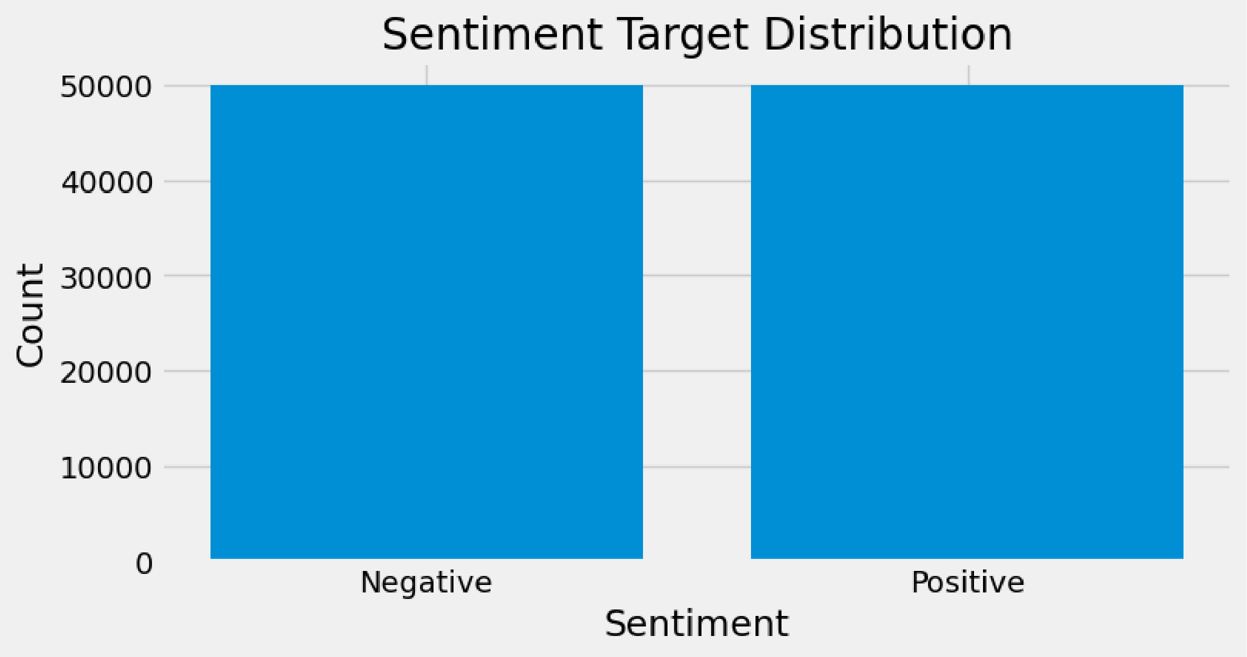
The data should be anonymized in such a way that would respect the individual's privacy. Though the 1st Amendment allows for the internet spaces such as Facebook, Twitter, Reddit, and other platforms to be a public forum, it is imperative that these tweets and posts are anonymized in such a fashion as to protect any victims from further points of contention while also respecting an individual’s right to free speech and self-expression. There is no plan for diagnosis either, so performing or using these datasets as means to help either prevent causing more harm or perpetuating more bullying or depression.

Challenges/Issues

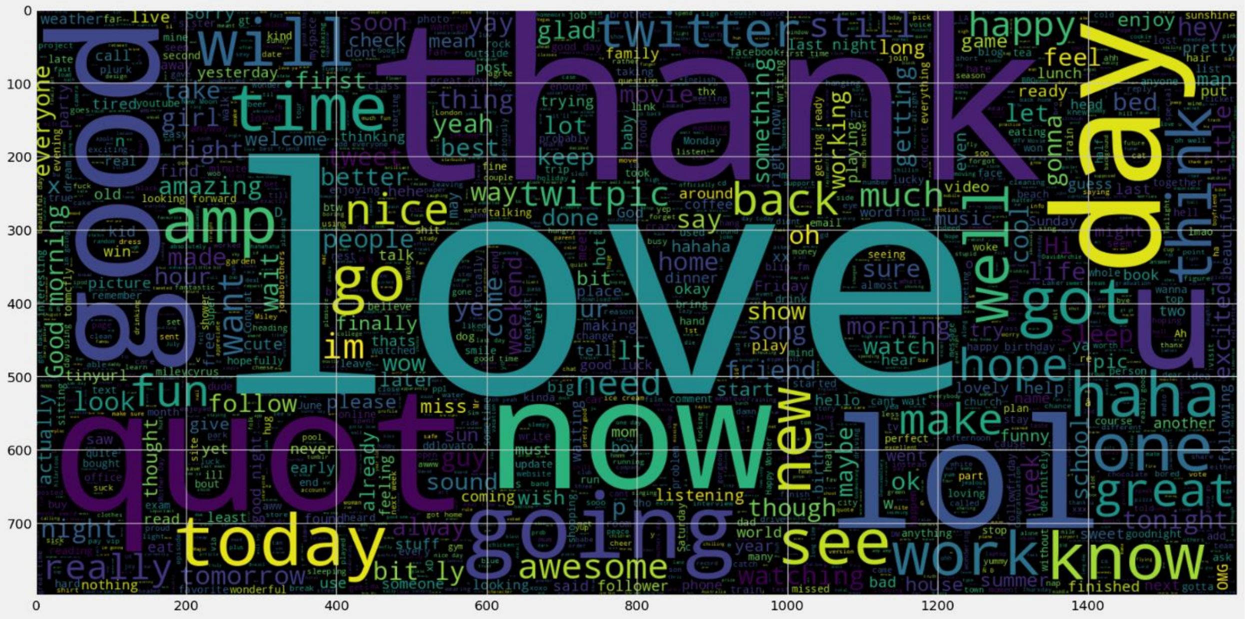
Some of the data sets seem to be older than 2020. There might be an issue getting a more current data set, as API’s that were available in the past might not be available for the public now. There is also a challenge of sorting through the data and finding meaningful insights into the data that produce the kinds of data this project is looking to produce. Additionally, there could also be plenty of opportunities that produce a sentiment that seems like either a cyberbullying or depressive message but could in fact not be an accurate assessment. A training model with wrong metrics that it is testing against could produce results that are not what the story is trying to tell.

Findings

When evaluating the data there were many discoveries found throughout the data. The first thing that needed to occur to make this data work was to clean it properly. It was prudent that the data be checked for proper format and to begin analysis on the different datasets. From three data sets, an extraction of information was gathered. From the first dataset a sentiment target distribution was created using 100,000 tweets and was equally divided between values of Negative and Positive.

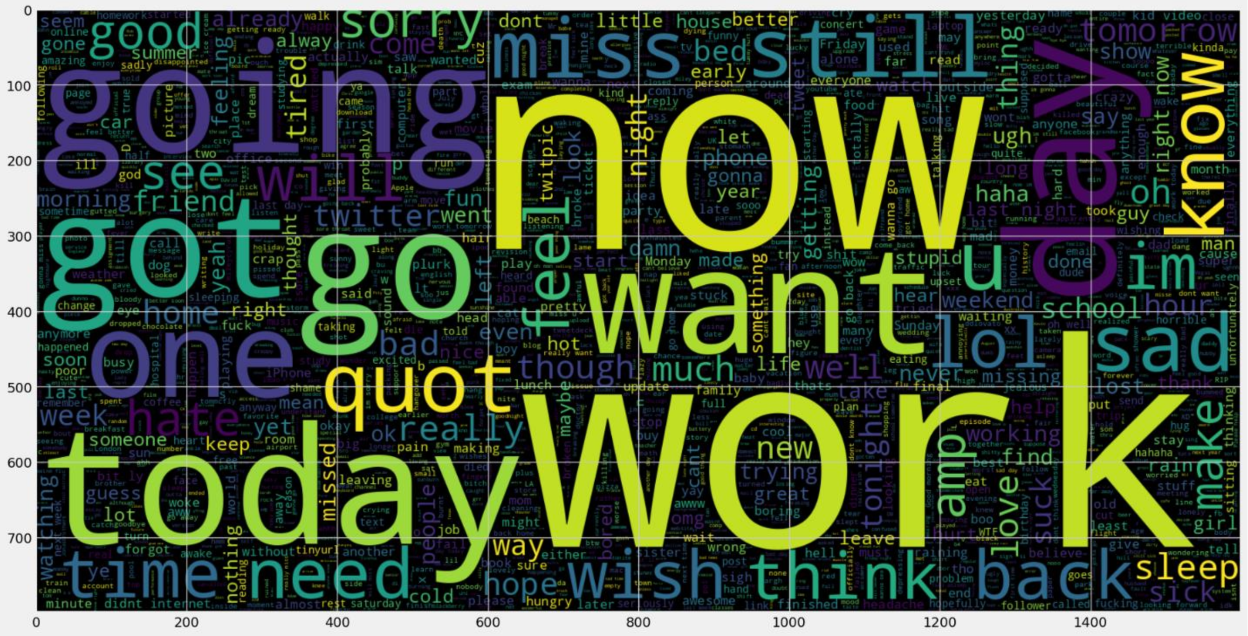


From the Twitter sentiments a word cloud was formed to see the top words that were found in Positive tweets. This produced a word cloud of the top words used in all the positive sentiment tweets. Commonly found words in tweets that were considered positive, or convey a positive sentiment were ‘Love’, ‘thank’, ‘now’, ‘lol’, ‘good’, ‘going’, ‘day’, ‘think’, ‘awesome’, and so many more.

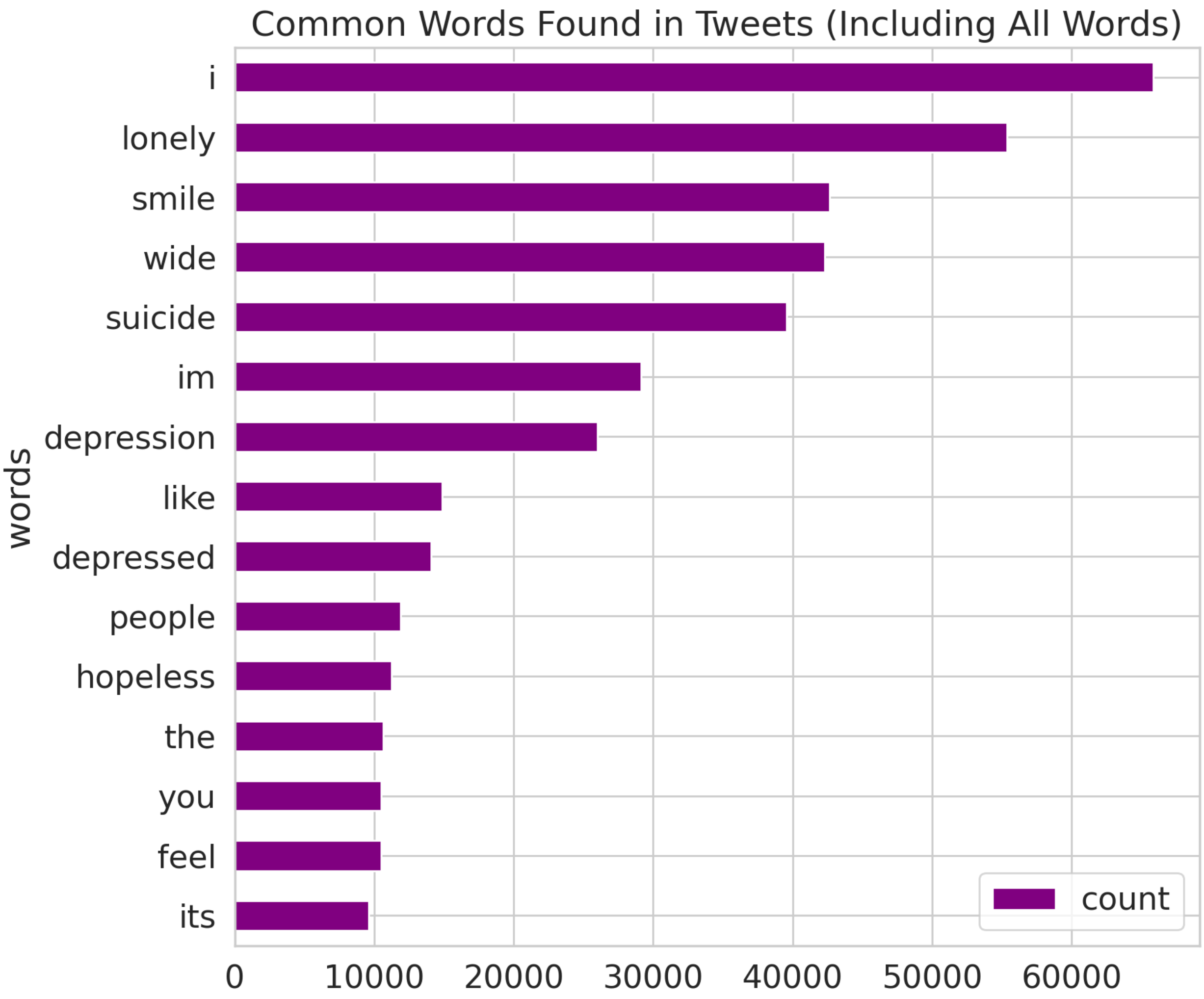


Negative tweet sentiment

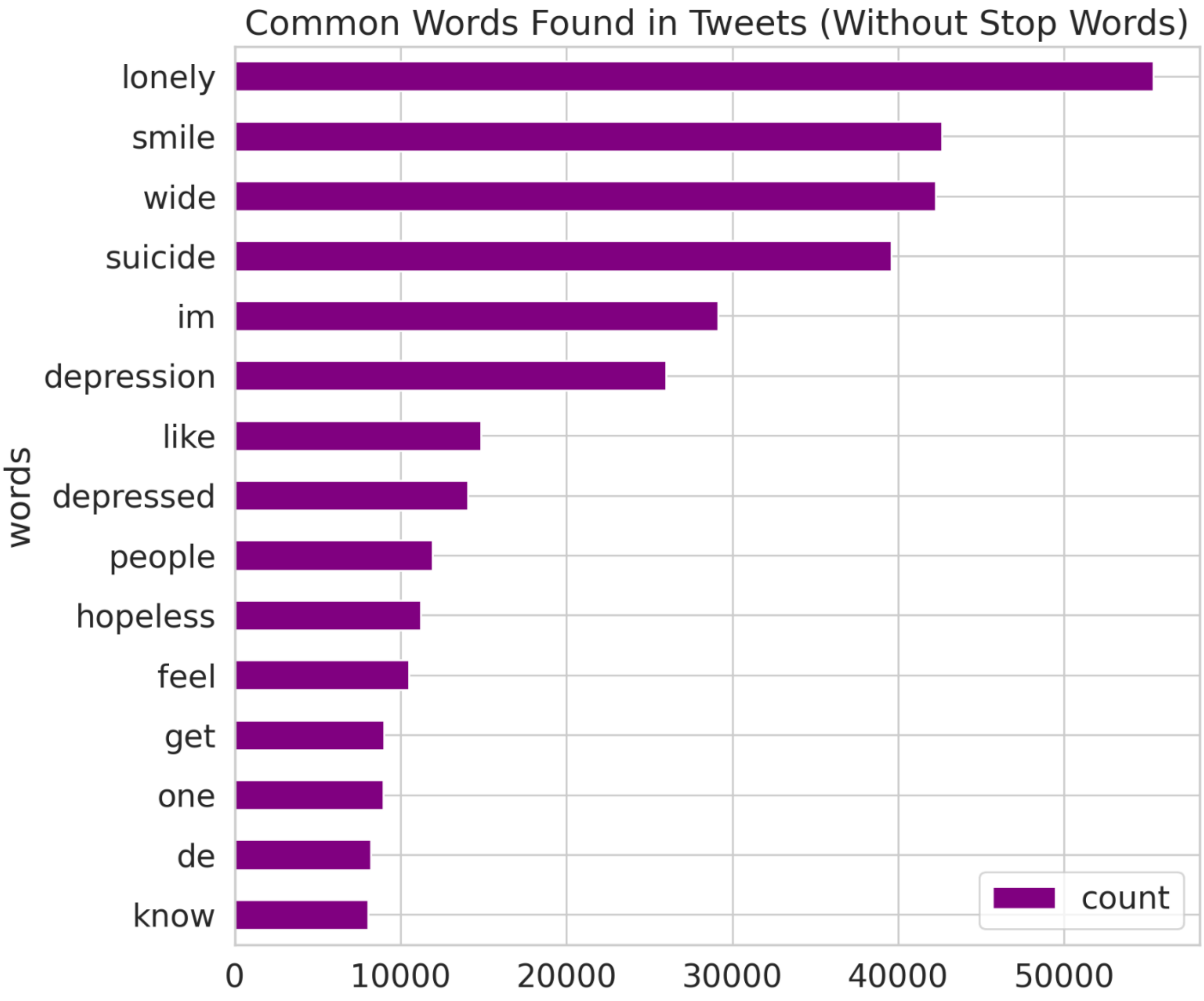
Negative sentiment analysis word cloud produced results that had a more bleak outlook. Common words were now, work, going, got, go, today, miss, think, need, and many others. A word cloud was produced to represent the negative tweet sentiments.



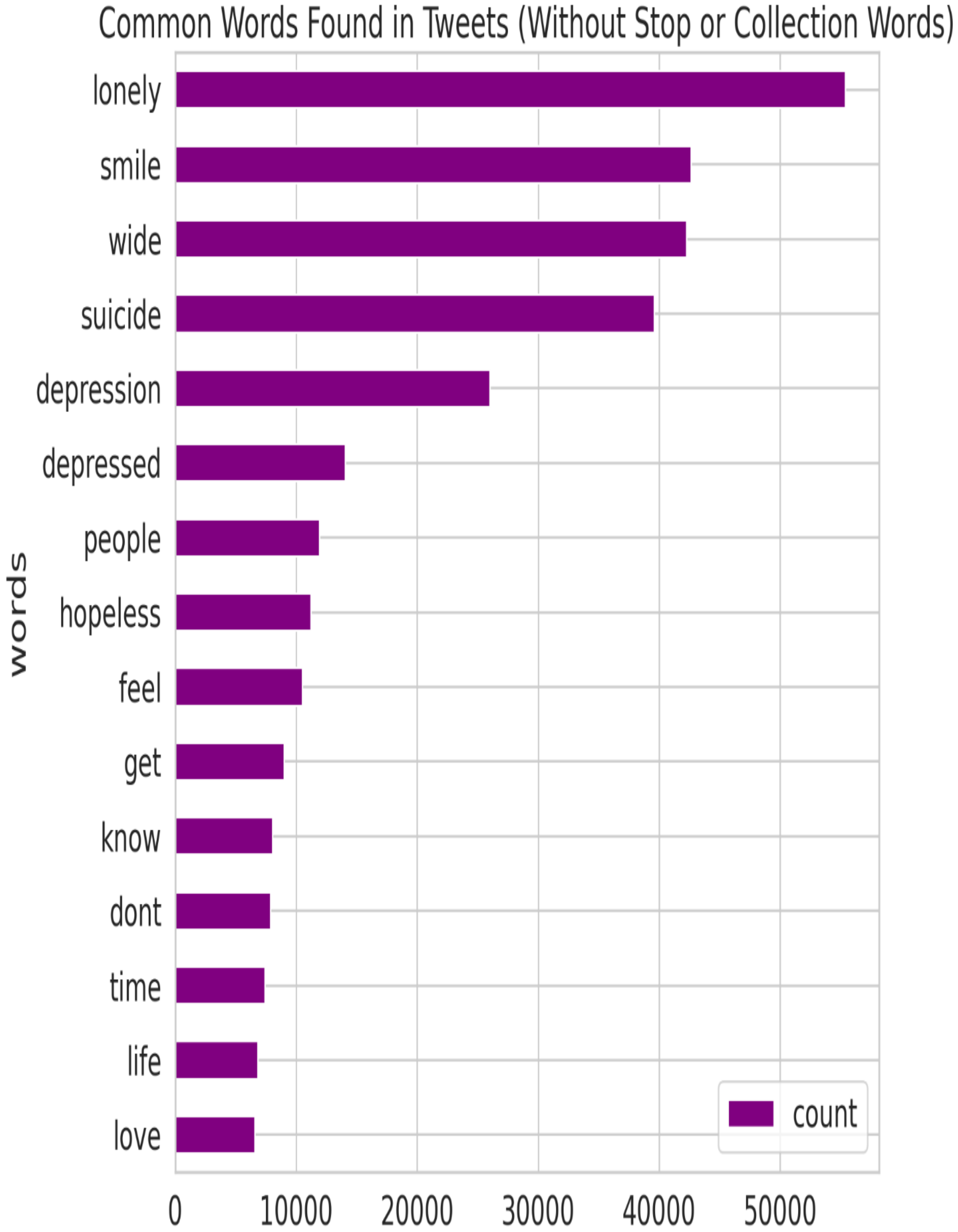
Some commonalities were found among all tweets as well. These words were found often all the tweets regardless of the stop words.



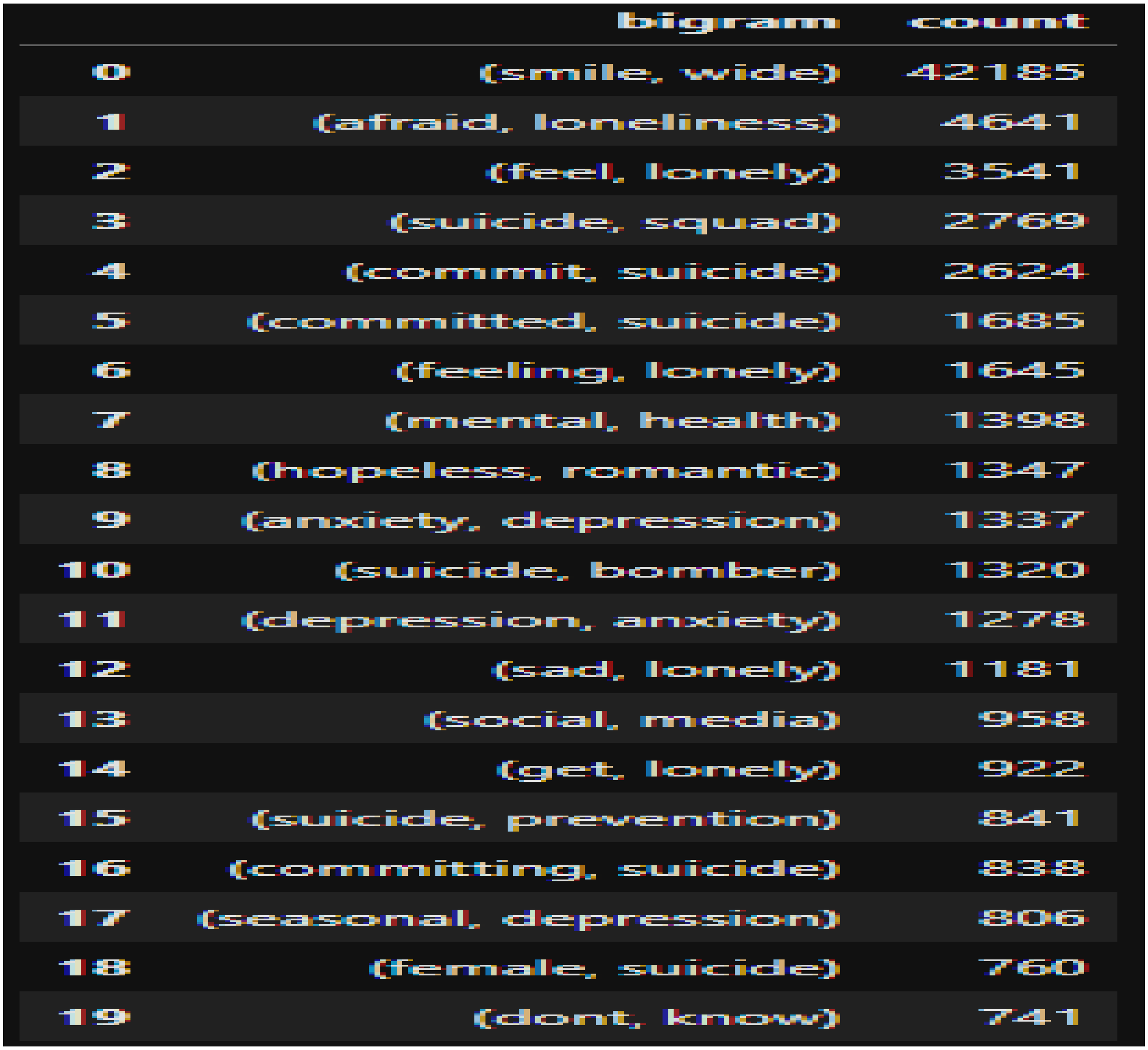
When we begin to filter out "stop words", words that are commonly found in a language, we start to see a different set of words that are more common. If we exclude the words 'im' and 'wide,' we are left with a top five words in tweets: lonely, smile, suicide, depression, depressed. This indicates that there is more likelihood of a depressive person on the end of a tweet than there are a person who is experiencing a good disposition.



Without stop and collection words, filtering out the collection words creates the same type of appearance as above but without the filler words or words such as ‘de’. It also added more words after ‘know’, such as ‘don’t’, ‘time’, ‘live’, and ‘love’.   The rational for this is that though most are depressed and tend to be a larger makeup of the tweets created, the opposite side of depressed tweets begin to bubble up as either neutral words, descriptive, or positive sentiments.


Common Words Found in Tweets (Without Stop or Collection Words)
words
lonely
smile
wide
suicide
depression
depressed
people
hopeless
feel
get
know
dont
time
life
love
0
10000
20000
30000
40000
50000
count

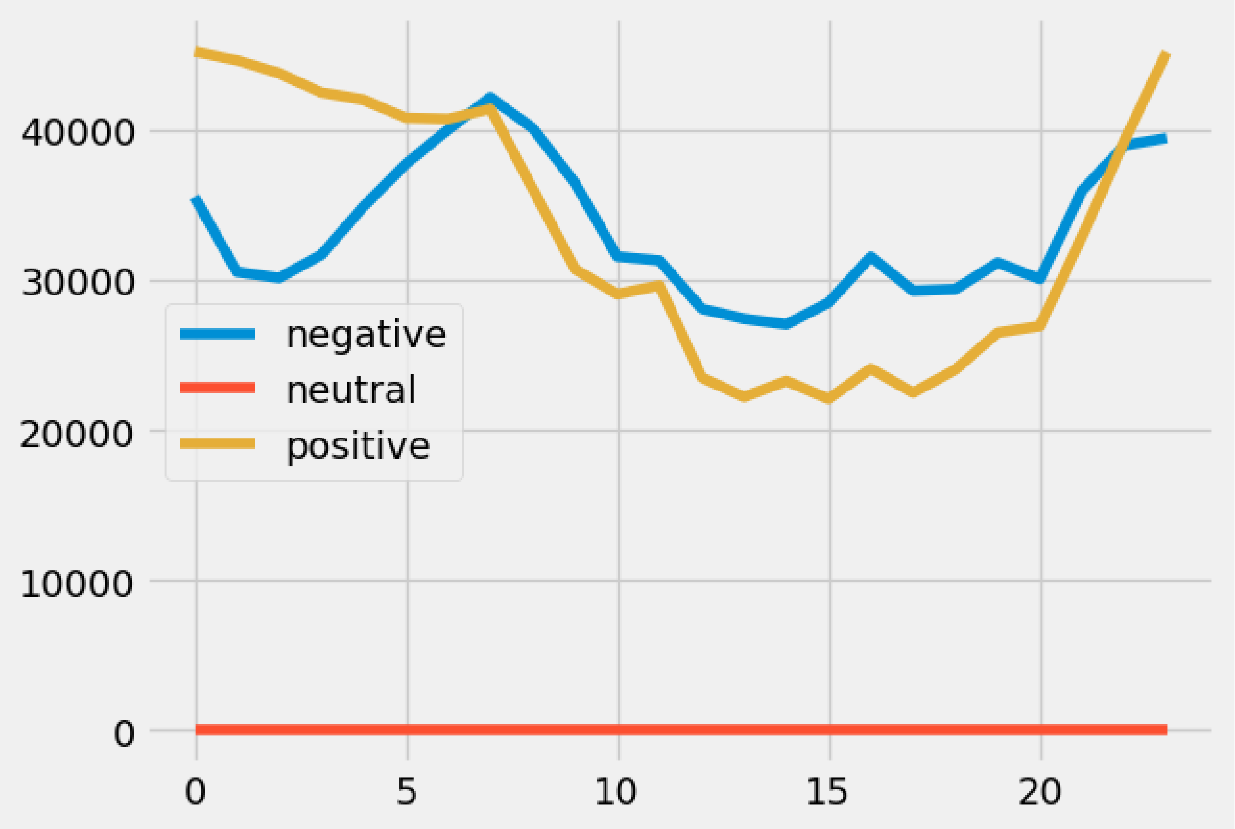
A bigram is considered a sequence of two adjacent elements from a string of tokens, which are typically letters, syllables, or words. Using the dataset and generating a bigram and extracting the top twenty words we find that there are two connections between the following:



	bigram	count
0	(smile, wide)	42185
1	(afraid, loneliness)	4641
2	(feel, lonely)	3541
3	(suicide, squad)	2769
4	(commit, suicide)	2624
5	(committed, suicide)	1685
6	(feeling, lonely)	1645
7	(mental, health)	1398
8	(hopeless, romantic)	1347
9	(anxiety, depression)	1337
10	(suicide, bomber)	1320
11	(depression, anxiety)	1278
12	(sad, lonely)	1181
13	(social, media)	958
14	(get, lonely)	922
15	(suicide, prevention)	841
16	(committing, suicide)	838
17	(seasonal, depression)	806
18	(female, suicide)	760
19	(dont, know)	741

Generalization of works

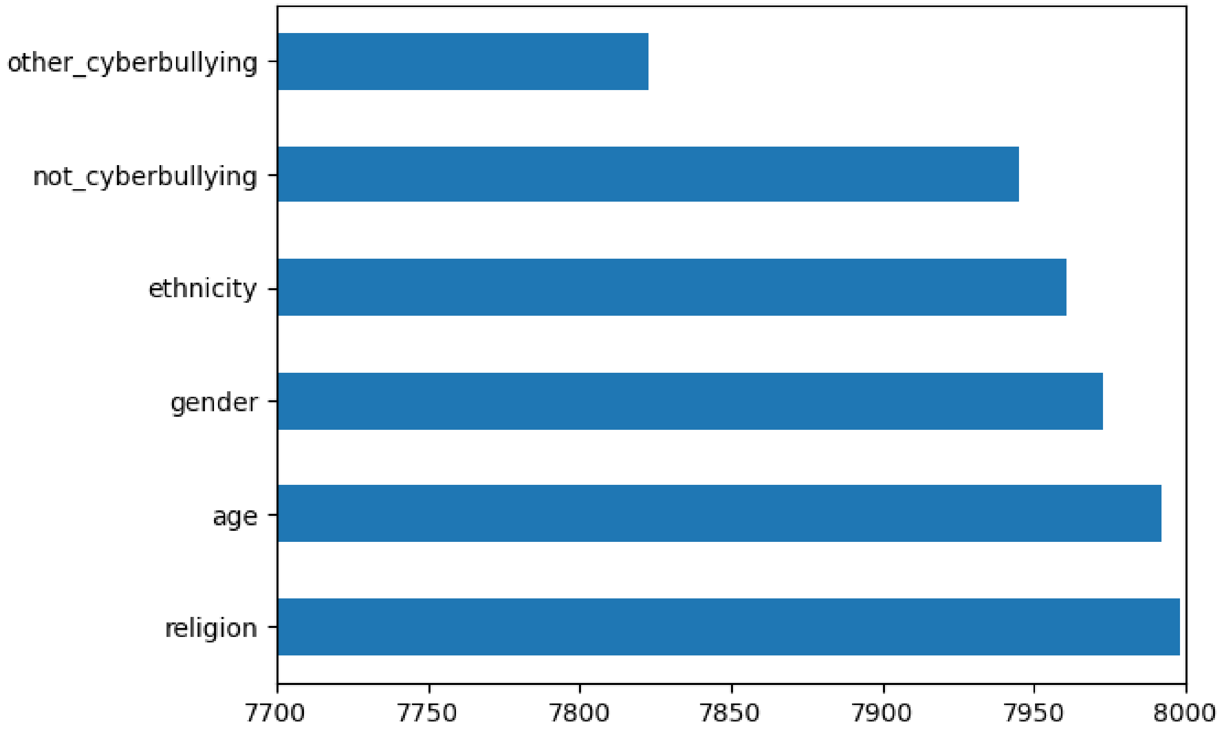
When the data is generalized and the number of tweets is considering and giving a positive, neutral, or negative value of sentiment based on time, we can see something interesting happen. From adding the data in the range of a 24-hour period, positive tweets start at a high point at the 00:00 mark, while negative tweets while lower than positive tweets, trend downward, and then begin having a massive uptick right around 02:00, and when most people are beginning their day, around 6:00, positive and negative tweets are around the same point. Throughout the day, we see that negative tweets and positive tweets both tend to track downward in count, but negative tweet sentiments continue to stay about positive tweets. At 20:00 we see something interesting again, we see a massive uptick in positive tweet sentiments, and a massive uptick in negative uptick sentiments as well.



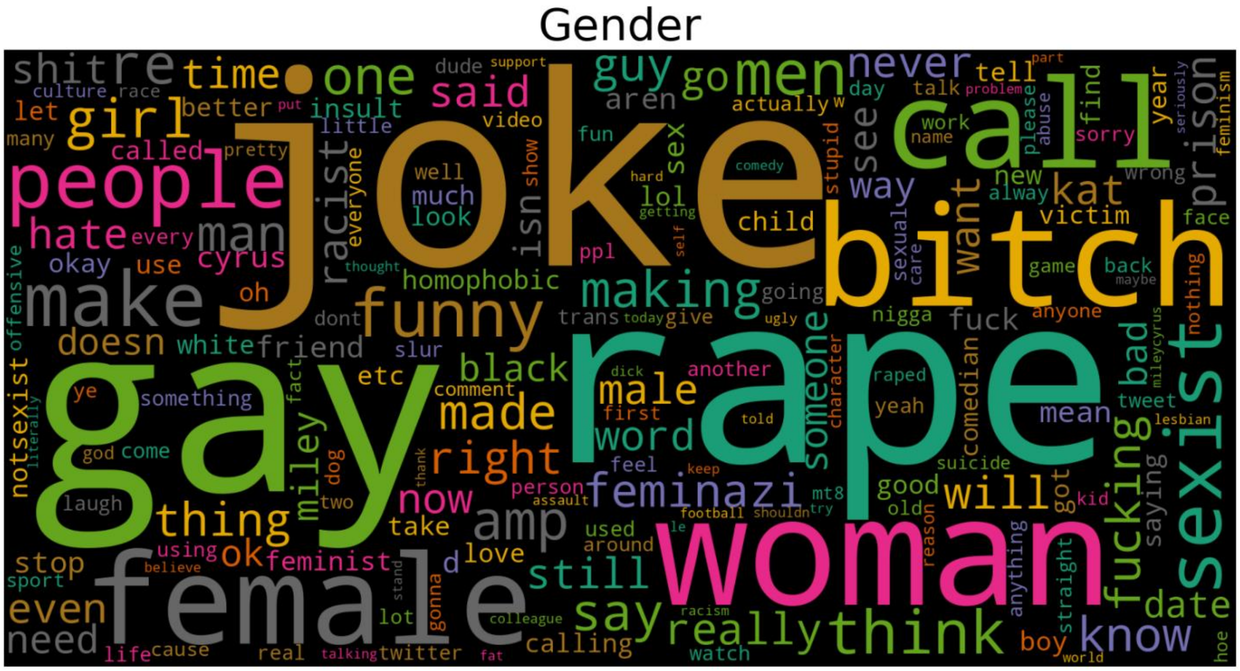
It would be easy to conclude that during the day, people are experiencing stress at work, while also having little wins, but the negatives always tend to be more profound than the positive tweets. Around the time that most people are getting off work, we see a skyrocketing value of positive tweet sentiments while also having a large increase in negative tweet sentiments.

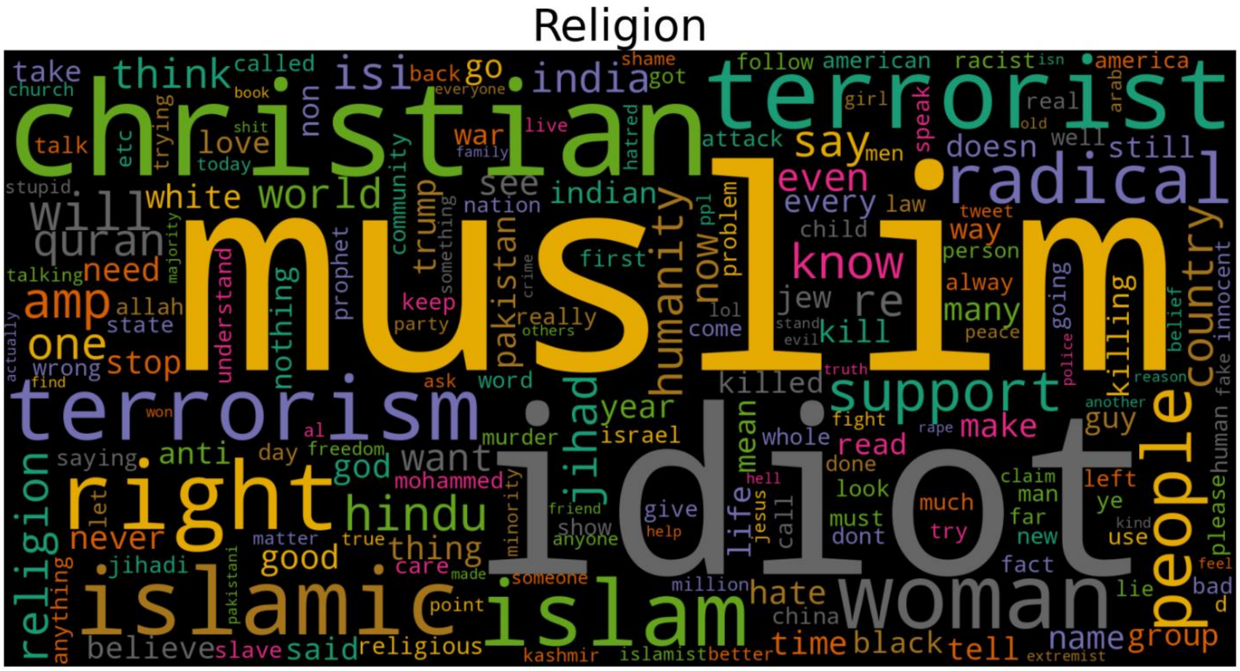
Cyberbullying

When it comes to cyberbullying the data set produced several different sentiments broken up by Gender, Race, Religion, Ethnicity, and what would be classified as 'other' kinds of cyberbullying. From exploring this dataset, there were a fair share of cyberbullying values that were often used in negative tweet sentiments. Using the dataset cyberbullying_tweets.csv, a list of labels was found to be unique in the cyberbullying_type column:

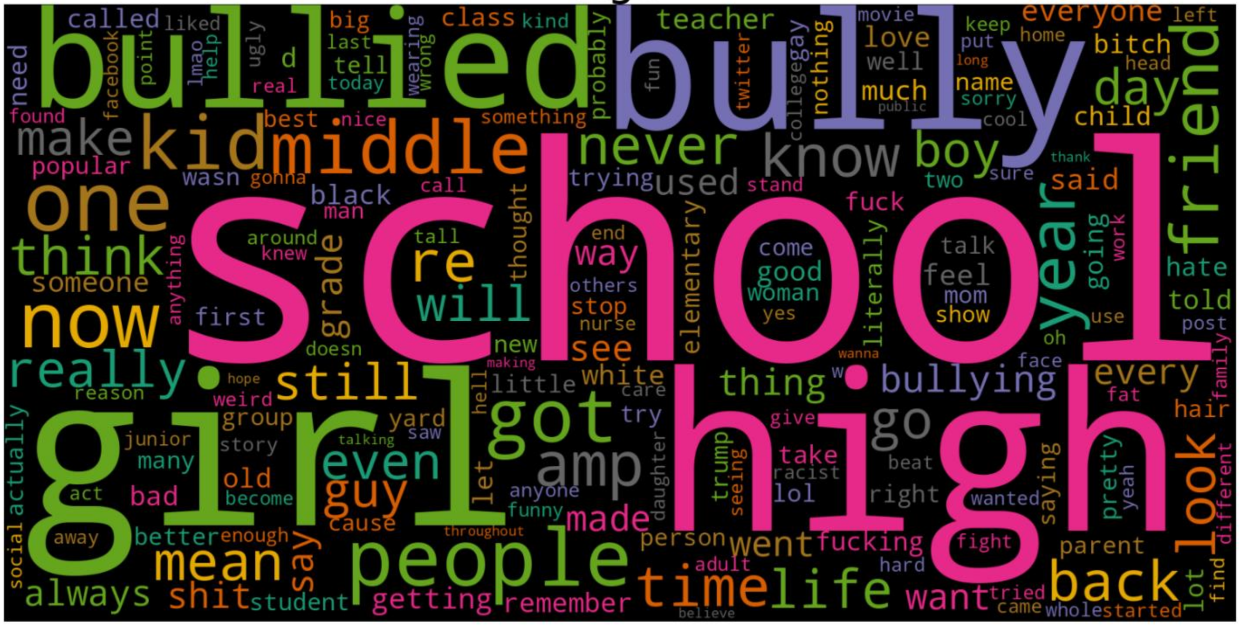


Of the different labels created, it was easy at this point to build some word clouds of the most common words that were used based on the unique label’s column. After adding an update to the stopwords, we were left with a word cloud for each unique type.

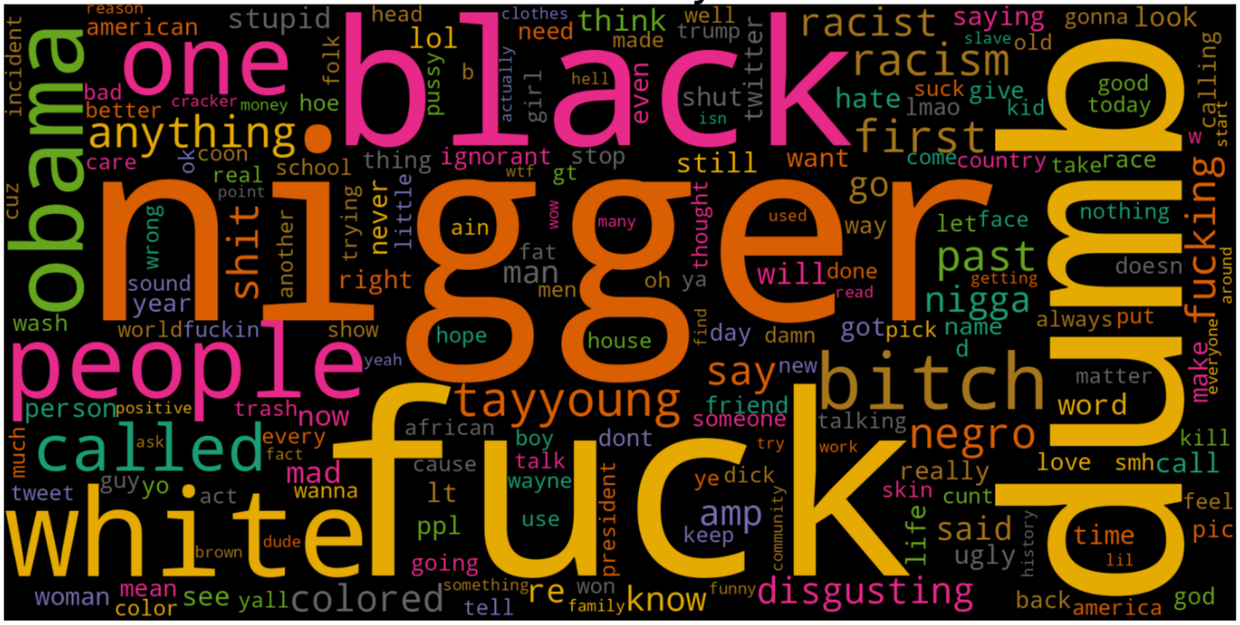


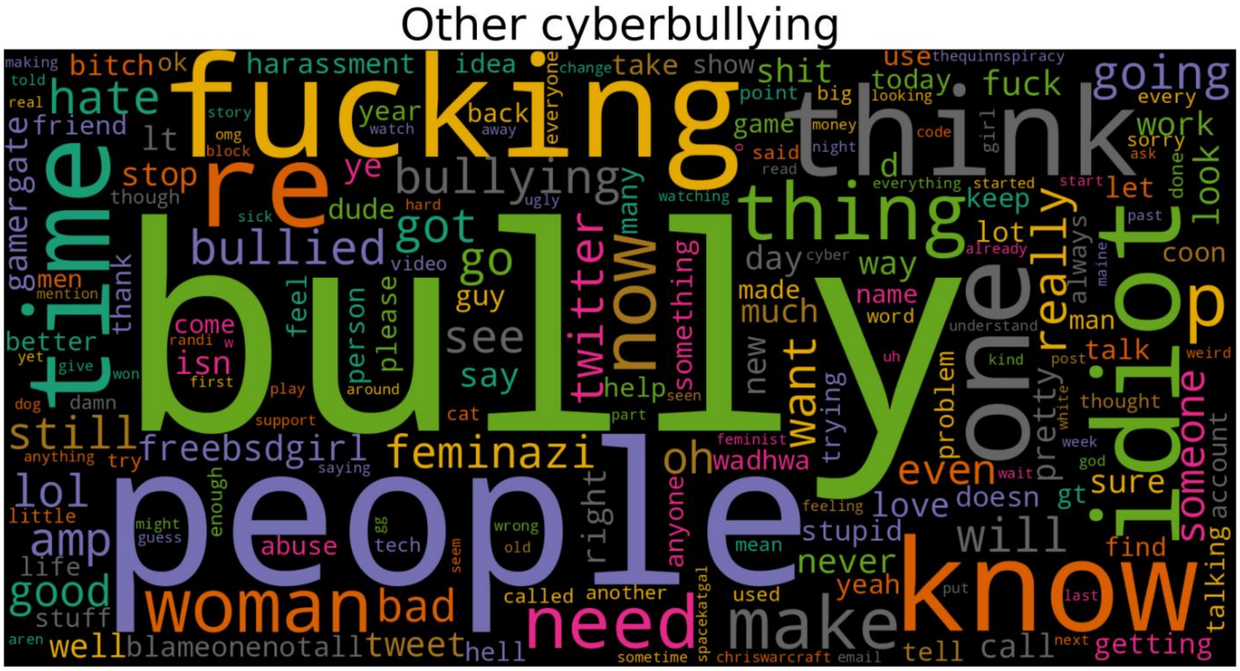


Age

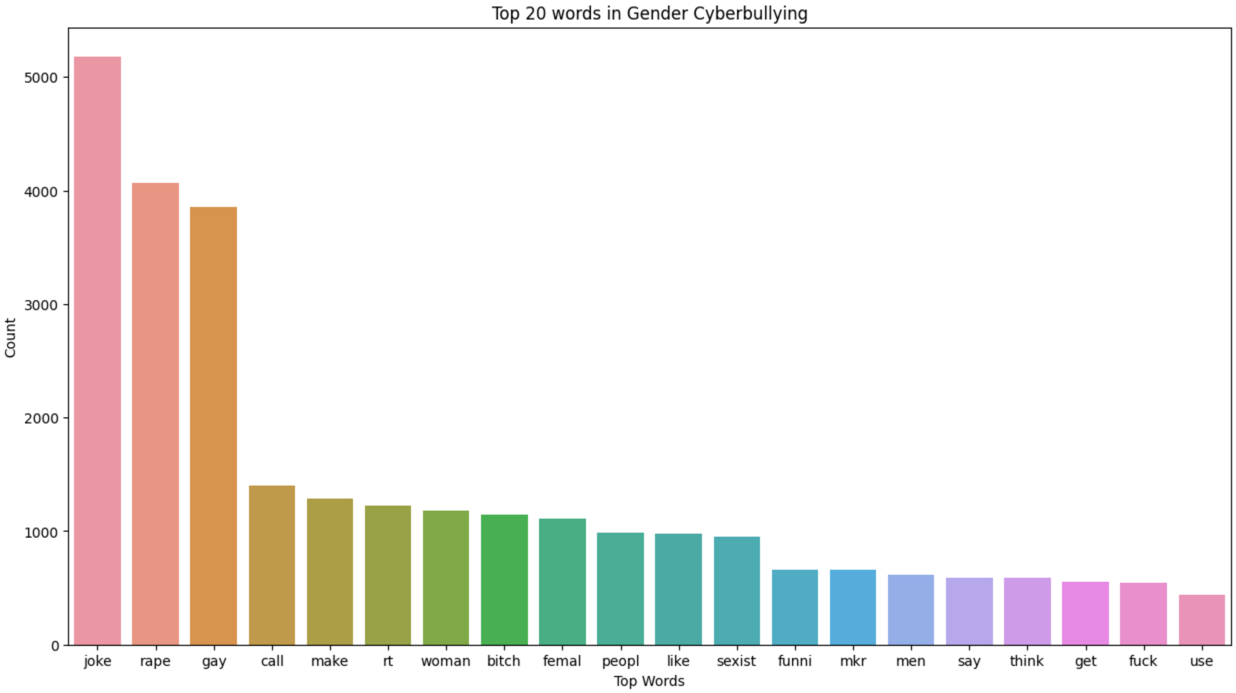


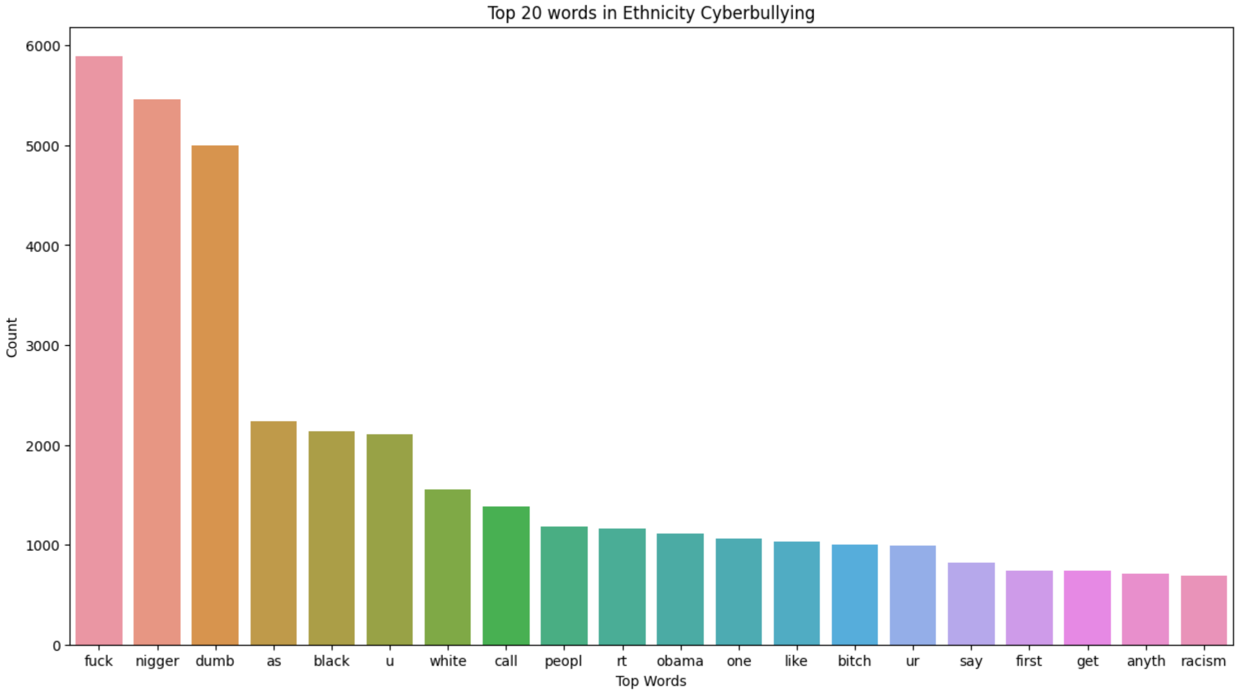
Ethnicity

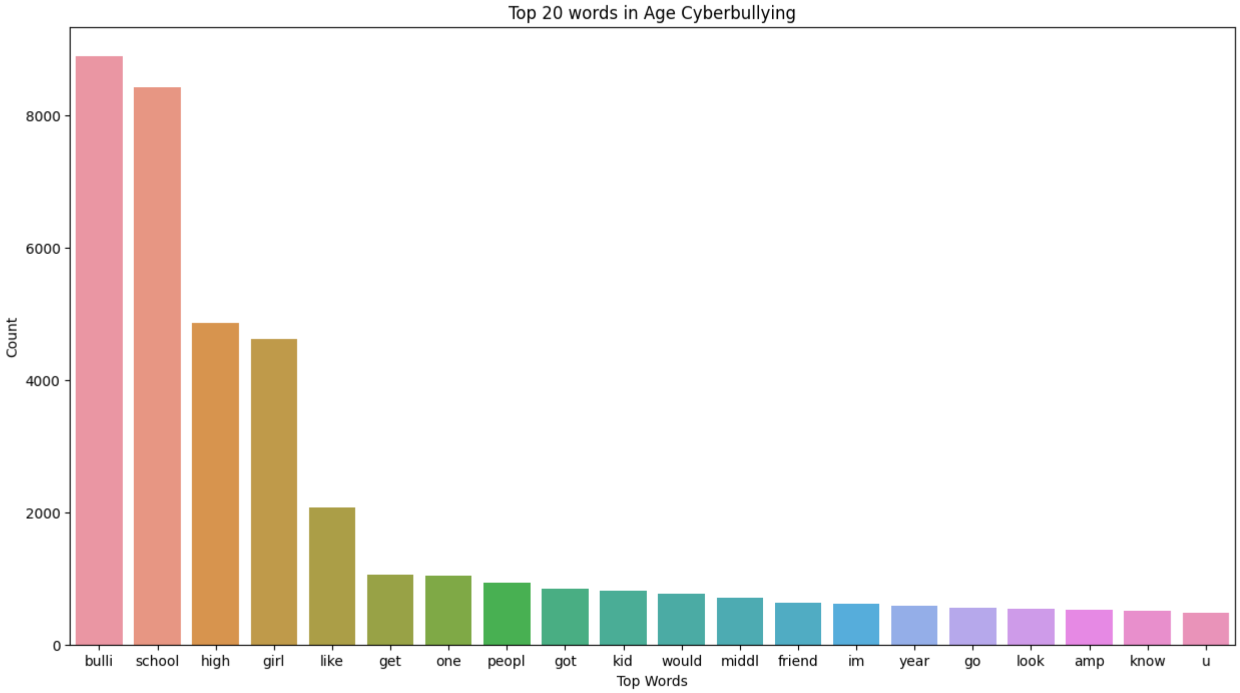


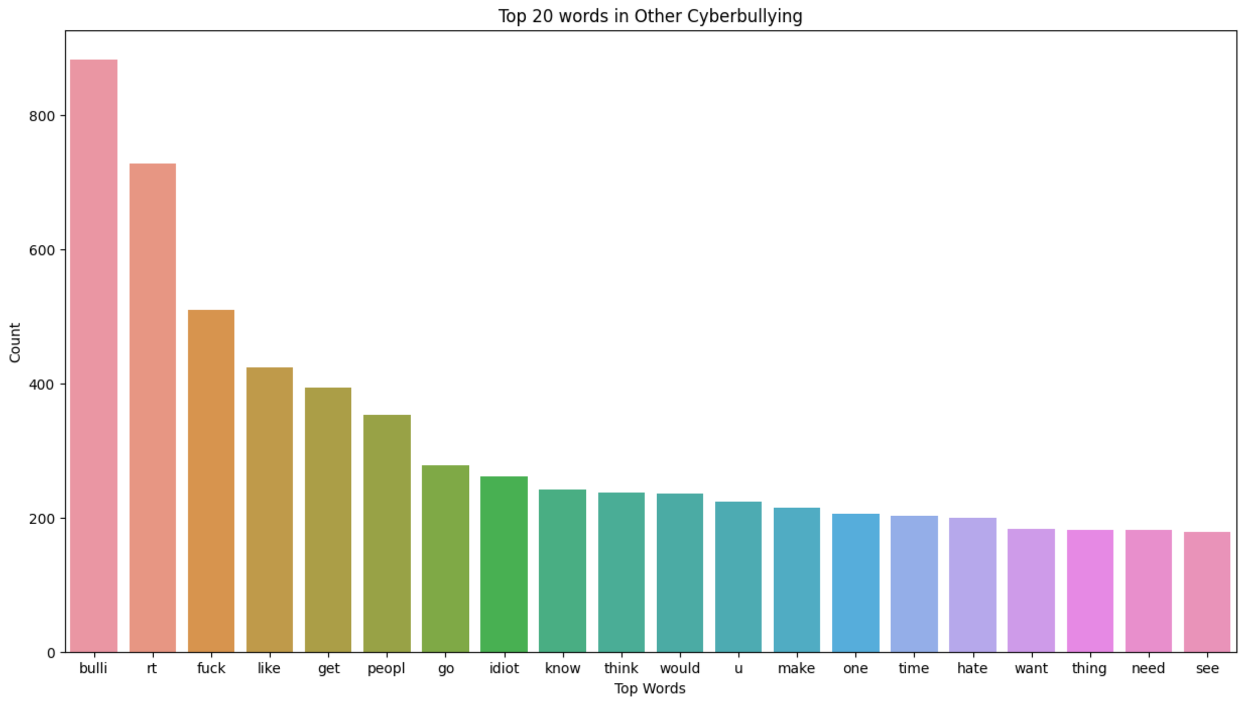


After initially looking at the word clouds, a bar graph was used to create a list of the top 20 words that were used in the unique labels.




Top 20 words in Ethnicity Cyberbullying


Top 20 words in Age Cyberbullying



From what it would look like, there is quite a bit of negative sentiment towards gays, females, blacks, Muslims, Christians, but unsurprising was a large set of words from sets that had the word ‘bulli,’ and the next word in Age was ‘School.’ There is a safe estimation that cyber bullying tends to target those who are women, gay, black, of a certain religion such as Christian or Muslim, but when it came to age, those who are bullied the most are in a school setting and are female.

Conclusion

From exploring the data, most people on the internet are depressed. While bullying online does seem to have an impact on some demographics, the takeaway for depression was less about bullying and more about time of day, and what people experience at or after work. Those who are online and tweet tend to tweet negatively over the course of a day, which could be caused by a myriad of reasons, but the words that appear most are related to ‘work’, ‘going’, having a relation to time such as ‘yesterday’, ‘today’, ‘hour’, and depressive words like ‘bad’, ‘miss’, ‘sad’, ‘want’, and ‘sorry’ or states of mind such as ‘bored’, ‘great’, and ‘suck’. Though not conclusive, depressive states of mind could be an early warning sign of stress caused by work or time related stress.

References

https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification

https://www.kaggle.com/datasets/gargmanas/sentimental-analysis-for-tweets

https://www.kaggle.com/datasets/kazanova/sentiment140

https://medium.com/swlh/detecting-depression-in-social-media-via-twitter-usage-2d8f3df9b313

https://www.who.int/news-room/fact-sheets/detail/depression

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0012948

https://www.semanticscholar.org/paper/Depressive-Moods-of-Users-Portrayed-in-Twitter-Park-Cha/8dd58913bd343f4ef23b8437b24e152d3270cdaf?p2df

https://www.pnas.org/doi/10.1073/pnas.1802331115

https://arxiv.org/pdf/1804.07000.pdf

http://www.munmund.net/pubs/icwsm_13.pdf

https://www.nature.com/articles/s41598-017-12961-9

https://www.pewresearch.org/internet/fact-sheet/social-media/