Twitter Sentiment Towards LGBTQ vs Related Hate Crimes

Github: https://github.com/kumar-Ranjith/Sentiment-Analysis-With-Twitter-LGBTQ-

Abstract:
Social media platforms like Twitter provide a vast amount of data that can be analyzed to gain insights into various topics and issues. One such topic is the LGBTQ+ community, which has been a subject of discussion and debate on social media platforms. Twitter has become a platform for people to express their opinions and sentiments on a wide range of issues related to the LGBTQ+ community. Therefore, analyzing the sentiment of tweets related to the LGBTQ+ community can provide valuable insights into the opinions and attitudes of people towards this community. The purpose of this project is to analyze the sentiments towards the LGBTQ+ community and to analyze its relationships with current events, namely the amount of hate crimes in the community over a seven year period. This is done through the acquisition of Twitter posts over a seven year period, analysis of the general attitude towards the LGBTQ+ community within the posts, and examination of the relationship between the attitude towards the LGBTQ+ community and the frequency of hate crimes towards the community in the same time interval.

This project focuses on the text data from each tweet to create a model that can predict a sentiment score that can define the general attitude towards the LGBTQ+ community. The relationship between this score as well as the frequency of hate crimes within key regions are analyzed as well. The project focuses on using the twitter data as our key sample and acts as our representation of the overall community so that we can make a connection between the community's sentiment and the region's known hate crimes specific to the LGBTQ+. After conducting the experiment, we saw little correlation between hate crimes and twitter sentiment on LGBTQ+, however there are key notable dependencies among features. From our data, we explored the region (by state) and the time (by day, month, and year) and used a plethora of statistical analysis tools to reach a conclusion.

Related Work

One of the relevant methods and algorithms that have been used in the past for problems such as the question we are answering is sentiment analysis. Sentiment analysis, at its most basic level, is taking a sentence or paragraph and analyzing whether it is positive, negative, or simply neutral. It is not focusing on solely what words are being said, but the intent of the author when saying them. Sentiment analysis takes the form of a handful of steps, first recognizing that there is an opinion being expressed, second being whether the opinion is positive or negative, and finally who or what the sentiment is directed towards. In our case, the focus is on sentiment directed towards the LGBTQ+ community, and we will be using a prebuilt sentiment analysis engine to assign scores to each tweet. More information on the subject can be found in Bing Liu's 2020 textbook called Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, and examples on how sentiment analysis has been used for other subject can be shown in Yasen and Tedmori's 2019 study on movie reviews called Movies Reviews Sentiment Analysis and Classification.

Data Sets

The focal dataset of this project is a large collection of tweets relating to the LGBTQ+ community over seven years, from 2015 to 2022, sourced with the help of UCI faculty. These tweets were chosen if they contained any of the following words: LGBT, LGBTQIA, Gay, Lesbian, Bisexual, Queer, Trans, Transgender, Homophobia, Transphobia, Homophobe, Transphobic, Homophobic, Transphobe, Asexual, Aromantic, Cisgender, Homosexual, and Pansexual. These words were specifically chosen to include as much as the LGBTQ+ community as possible as well as cover the prejudices against the community.

However, due to the size and certain properties of the dataset, we had to conduct some data wrangling before further analysis. Firstly, although we had hoped to analyze tweets from over 2015 to 2022, the tweets from 2020 to 2022 were limited, so we might have to settle with performing analyses from only 2015 to 2019. Furthermore, the size of the entire dataset, (over 2 million observations), was too large to sufficiently analyze, so we decided to randomly sample the data with 50 observations from each day over the time period. The resulting dataset ended with about 72,000 observations. Each observation consists of a single tweet with the following features:

1. Create at: A timestamp marking when the tweet was created.
2. Geotag State Name: The US State where the tweet was created
3. Geotat City Name: The US city where the tweet was created
4. Id: A unique identifier for the tweet
5. Text: The content of the tweet.
6. User Description: The user biography of the tweet's author
7. User followers Count: the amount of followers the tweet's author has
8. User Id: a unique identifier for the author

For example, the following tweet can be an example row of our data.

Create at	**Getoag City Name**	**Geotag State Name**	**Id**	**Text**	**User Description**	**User followers count**	**User Id**
2015-12-07T00:52:16.000Z	Bayshore	NC	669027872309436416	Queer squad hype!	HAPPY AS HELL	452.0	308576924

We did take a brief look at the distribution of keywords across our smaller dataset, and found that gay was the most common of our keywords by a landslide, and while some of the others did see some overall frequency, they were incomparable to how common gay was. Surprisingly enough, negative keywords like homophobia and transphobia were significantly less common than expected, only making up a small fraction of tweets. While some more specific terms such as aromantic and pansexual were less common than these keywords, it was interesting nonetheless to see the negative keyword count be so low.

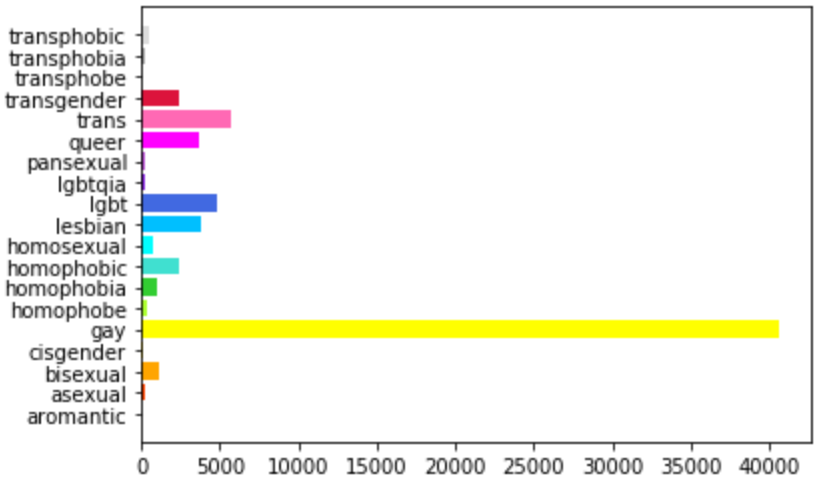

Figure 1: Keyword Count Distribution

The other dataset was sourced from the official FBI government website at https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/downloads#datasets, and it contains US hate crimes committed from 1991 to 2021. It contains about 226,000 observations, and each observation represents a single crime and has 28 features, but this project is interested in only the following:

1. Incident Id: A unique identifier for the crime.
2. State Abbr: The US state where the crime took place, abbreviated.
3. Incident Date: the date of the crime
4. Offense Name: The type of crime committed, e.g. Intimidation, Assault, etc.
5. Bias Description: Describes the victim of the hate crime. E.g. anti-gay, anti-trans, etc.

An example row of data would look like this:

Incident Id	State Abbr	Incident Date	Offense Name	Bias Description
180579	CA	2021-12-09	Simple Assault	Anti-Transgender

This dataset contains data for all hate crimes, but we are only interested in data pertaining to the LGBTQ+ community from 2015 onwards, so we reduced the dataset to contain only crimes from 2015 onwards with Bias Description of any of the following: Anti-Lesbian, Anti-Transgender, Anti-Gender Non-Conforming, Anti-Bisexual, Anti-Gay, or, Anti-Lesbian, Gay, Bisexual, or Transgender (Mixed Group). The resulting dataset contained about 51,000 observations. These two datasets, the tweets dataset containing 72,00 observations and the hate crime dataset containing 51,000 observations, are stored locally on csv files and can be easily manipulated using Pandas, R, SQL, or any other data management system.

When looking at United States data specifically, we can see which regions the tweets are most abundant in. Comparing the counts per state by the population of each state, we can understand how dispersed the data is over the entirety of the United States. Figure 2 shows a choropleth of the United States with the color scheme growing darker as the tweet per person value increases.

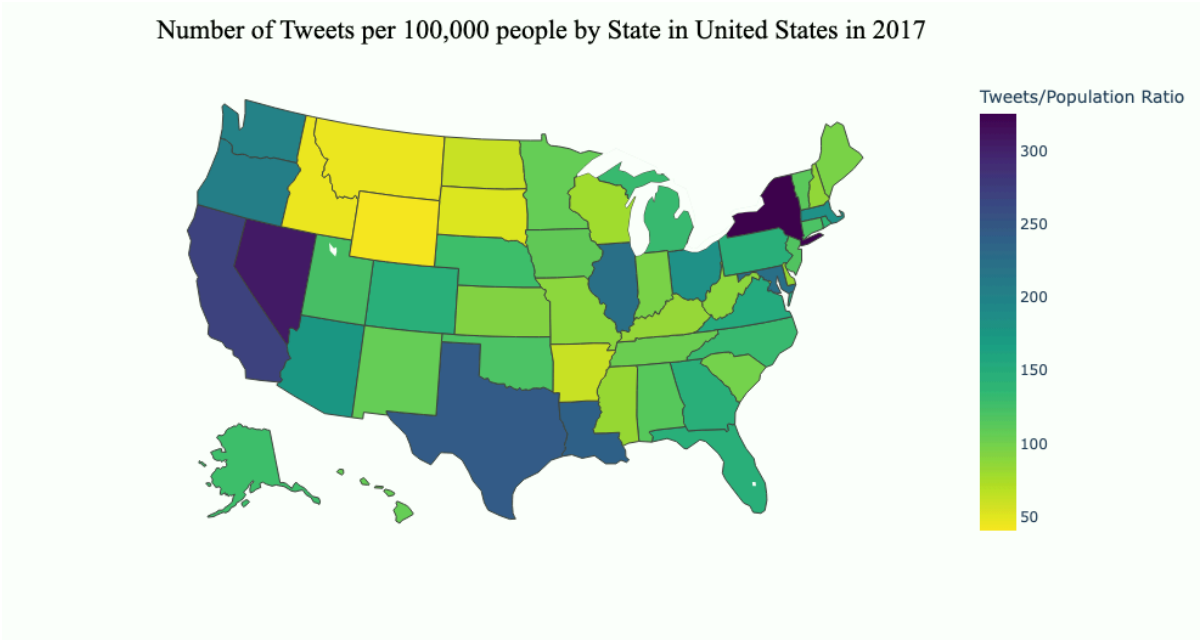

Figure 2: Choropleth of Tweets per 100,000 people by state in 2017

Overall Technical Approach

In order to gain a better understanding of the datasets, we first perform exploratory data analysis. We first wanted to look for the existence of any trends between the raw numbers of the Twitter and hate crimes datasets. In other words, we wanted to see if there were any obvious trends between the frequencies of hate crimes and the frequencies of tweets using our selected LGBTQ+ keywords across all 50 U.S. states. We decided to create two bubble plots for two separate years, 2015 and 2019, to also highlight any changes in the frequencies over a substantial period of time. These plots take into account three dimensions of the data: number of tweets, number of hate crimes, and the corresponding U.S. state. We chose to focus only on five keywords that could be directly translated into a reported bias description. For example, there are currently no bias descriptions present in the hate crime dataset relating to the keyword "aromantic." However, there is a bias description titled "Anti-Gay (Male)" which we used to search for tweets containing the keyword "gay." The five selected keywords are "transgender", "lgbt", "lesbian", "gay", and "bisexual."

In Figure 3, we can see that "gay" is both the most prominently used keyword and the most frequently occuring hate crime in 2015. Interestingly, we can see that a significant number of the states do not have data points plotted for some of the keywords. We discovered that although a state may have had tweets containing a keyword, they had 0 related hate crimes occur in 2015. Thus, they are not present in the graph. We can see that most states that did have data points had hate crime counts of less than 50, as well as a lower number of tweets, less than 100, for the associated word. Overall, there does not appear to be any visual trends with this plot alone.

Much of what can be described about 2015's results can be applied to the results of 2019, as seen in Figure 4. The most frequently occurring keyword and bias still remains "gay" across all 50 states. However, we can also see from the two scales for number of crimes and number of tweets that the minimum and maximum values have all nearly doubled over the past 4 years. There is evident growth

seen for “transgender” and “bisexual,” of which the former is particularly understandable with the recent increase of awareness regarding the transgender community. Once we have our correlation scores, we will delve into exploring the possible reasons behind the increasing rates of both variables.

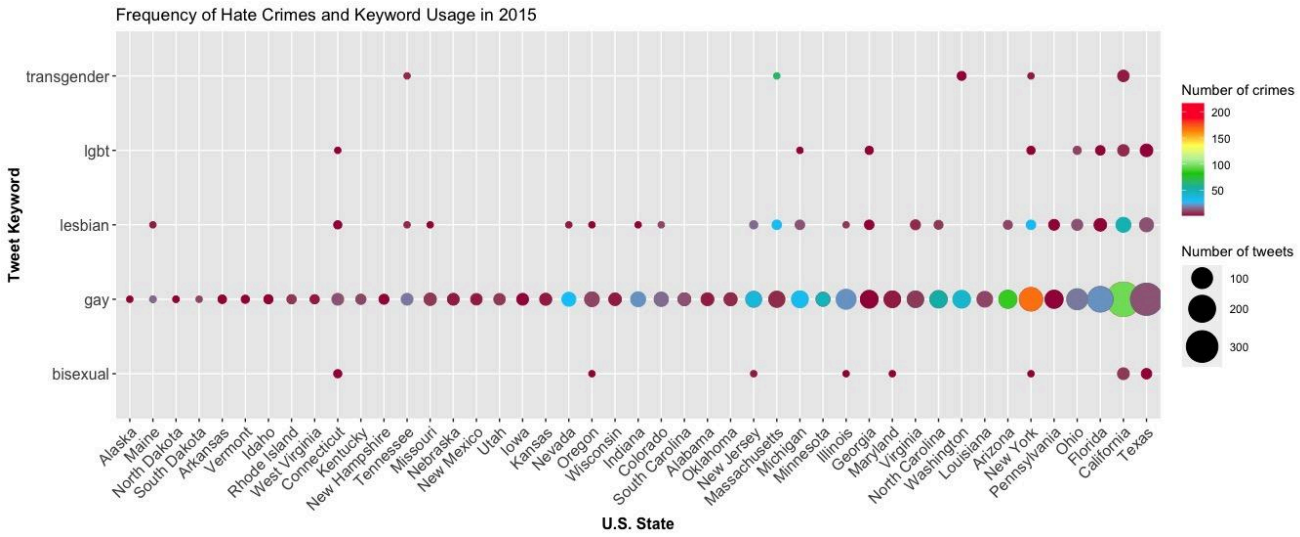

Figure 3: A bubble chart of the frequencies of tweets containing specific keywords and hate crimes within all 50 U.S. states in 2015.

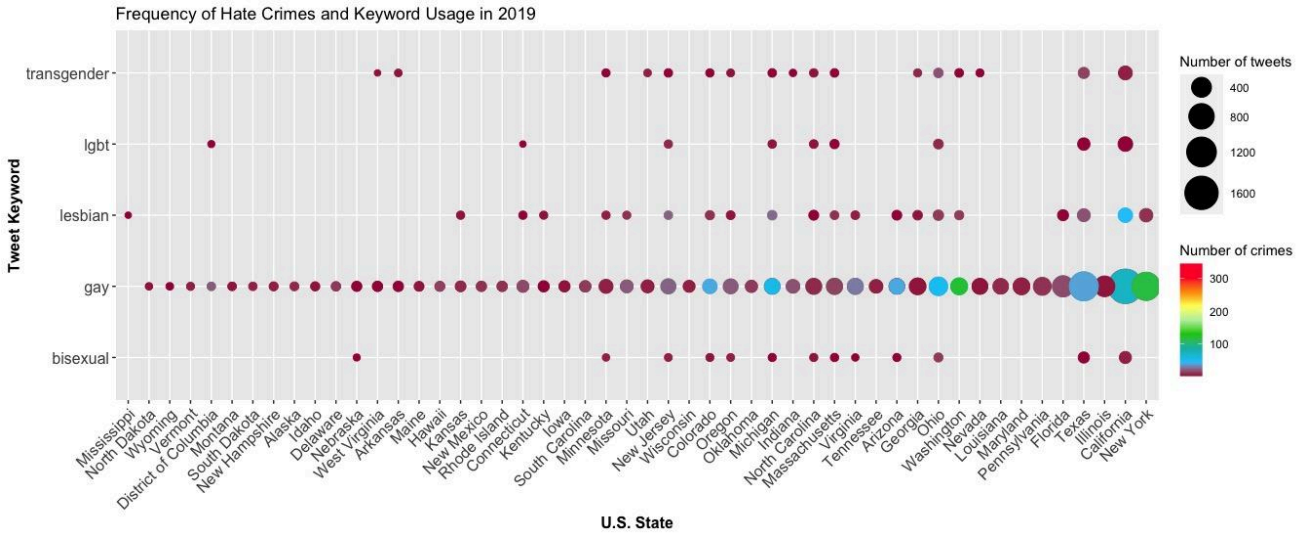

Figure 4: A bubble chart of the frequencies of tweets containing specific keywords and hate crimes within all 50 U.S. states in 2019.

In our initial glimpse of the data, we found a handful of bot tweets, which left us concerned as to whether a significant portion of our data would be made up of nonhuman responses. However, when we subset a portion of the data to go over it by hand to denote which tweets were made by bots in order to gain a more complete understanding of what we were working with, we found that only 3% of the tweets subset were

made by bots. A means to remove the bot tweets ultimately was not necessary due to the very small proportion of tweets made by bots.

Software

We were fortunate enough to only need a handful of various software for our analyses, using primarily python's pandas library in conjunction with rstudio's regression analysis, and microsoft excel's easy csv editing. We used Microsoft Excel in order to take an initial glance at the data, seeing the structure it took and what variables were important to our analysis, as well as which ones we believed we could discard. Additionally, we used it to manually filter bot tweets for our subset, ultimately finding that bots took up a surprisingly small portion of the data. Bot tweets, though with some variation, generally consisted of tweets using a large amount of hashtags or keywords in order to gain more engagement redirecting the reader either to a link within the tweet itself or in the bot's bio.

For the bulk of our data wrangling, analysis, and tokenizing, we used python and its Pandas library. It made it easy to both randomly select data by a particular characteristic, in our case by date, as well as adding a column with our sentiment score for each tweet. Additionally, pandas data visualization was robust and clean enough for us to not need to turn to alternative data visualization programs, such as R's ggplot library. To generate our sentiment scores, we trained a logistic classifier model using tweets from the NLTK library, discussed in more detail in the next section.

Finally, when working on our overall analyses for predicting the data, we used rstudio for our regression analysis. We found the outputs it generated to be useful for gauging how well crimes could predict sentiment, as well as how much variation was present in the data in regards to our linear regressions.

Experiments and Evaluation

In order for us to analyze the relationship between sentiment and hate crime data, we must first score our tweets by sentiment. Initially, we used an unsupervised model by TextBlob to score the sentiments, but upon reviewing the results, we found them to be unsatisfactory. Although there was no way to accurately measure the error due to the data not having a sentiment column previously, the model seemed to be overly biased in rating tweets as positive, even when the tweet itself was not positive by our standards. Because of this, we decided to train our own model. The model was trained off of 10,000 tweets from the Python NLTK library (https://www.nltk.org/howto/twitter.html) . The tweets are tokenized, lemmatized, vectorized, and inserted into a Logistic Classifier that either classifies the tweet as Positive or Negative. Essentially this means that the tweets were broken into words, removed of noise (such as links, handles, and hashtags), had each word turned into a base form (such as a noun) and turned into a matrix that a logistic classifier could understand (Added in response to comments from draft 2). Two example tweets can be shown below: one tweet contains positive language like "respect" and "ally" and was classified as positive, and the other has negative language like "homophobic" and "racist" and was classified as negative.

An ally can make a difference. Stand up for the LGBT+ community because they deserve respect too.	Positive
I always forget how homophobic, racist, and	Negative

transphobic my relatives can be, but then Facebook kindly reminds me....	

Currently, there is no way to accurately determine which model performs better since our original dataset does not have sentiments as a variable, but at a first glance, our group determined that the new model seems to record tweets more accurately. In the future if we have extra time, we might try subsetting a portion of our data to train the model on, to train accuracy. Alternatively, we could conduct a blind survey on volunteers to measure the accuracy of the models. One note about this model is that it is quite pessimistic, however. It is much more likely to record a tweet as negative than positive as shown in the figure below. This is important to keep in mind when analyzing our models. Even if we do not take into account any factors we can expect the odds of a tweet being positive over negative to be below one.

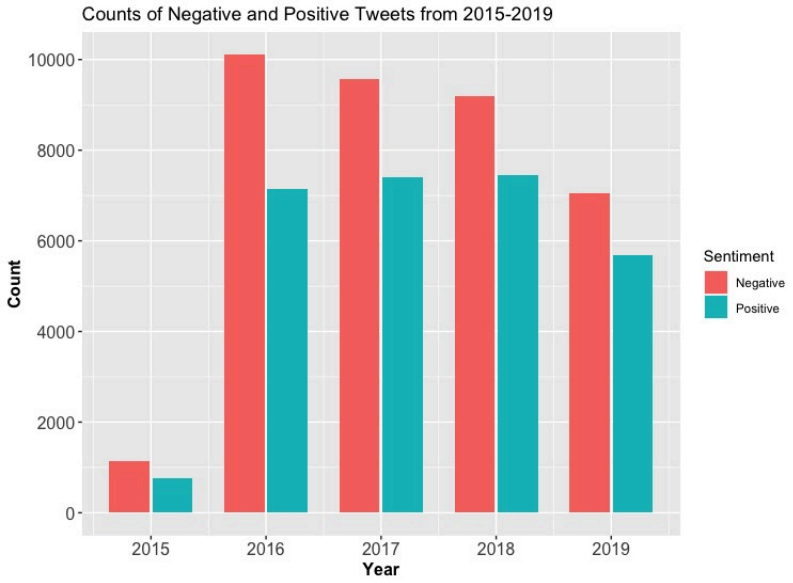

Figure 5: Sentiment Counts by Year

To analyze the relationship between sentiment and hate crime rate, we first create a linear regression model as our baseline. We chose this as our initial model because the response variable to be measured is the number of hate crimes in any given U.S. state. The only direct explanatory variables in the model are the counts of negative and positive sentiment tweets from each state. This model helps us gain insight on whether or not it is possible to predict hate crime count in a state from tweet sentiment count alone, without other possible factors involved. Our null hypothesis is that none of the predictor variables have a significant relationship with the response variable. The completed model equation is as follows, along with a table of the model's results:

$$CrimeCount = 74.193 - 8.1423Negative + 27.0233Positive$$

	Estimate	**P-value**
Intercept	74.19301	0.2675
Number of negative tweets	-8.14225	0.9281

Number of positive tweets	27.02326	0.0994

We find that if a state has 0 negative and positive tweets, it has about 74 hate crimes on average. For every additional crime committed in any given state, the number of positive tweets increases by about 27 and the number of negative tweets decreases by about 8. These numbers are small compared to the intercept and are not drastic enough in value to suggest significance. Using our chosen significance level of 0.05, we also find that since the p-values for the negative count and positive count variables are greater and, in general, are high, this indicates that we would fail to reject the null hypothesis. Thus, we can conclude that there is little to no relationship between negative or positive sentiment, and the hate crime frequencies. We also obtain a relatively low R-squared value of 0.3392, which suggests that the model does not fit the data well. This would mean that only about 33.92% of the variability observed in the hate crime counts is explained by this model.

In this model alone, there are a number of reasons why the correlation between the two is low. One reason is that the data contains a number of irrelevant tweets that can affect the accuracy of the scores. These tweets are usually not related to the LGBTQ+ community at all, but since they contain one or more of the keywords, they were still included in the model. Another reason would be confounding variables. For example, despite the increase in positive tweets as seen in the labeled data, this could be due to the influence of the news rather than the hate crimes specifically. A third reason relates to the lack of other factors being included, such as location and date, which is what we addressed in the following logistic regression models.

We then made a logistic model that analyzes the odds of a tweet being positive when taking the number of hate crimes that day. The reason we chose a logistic model was because our response variable, the sentiment, has only two categories: positive and negative. To begin, we chose the number of hate crimes against the LGBTQ+ community of that day and the number of followers the author has as factors, and analyze how these affect the odds of a tweet about the community being positive. These factors might be significant for the following reasons: the number of crimes might suggest an outgoing support for the community, and the number of followers might encourage the user to behave differently. For instance, a user with a large following might incentivize them more positively to cater to a larger audience. The completed model equation was as follows:

$$Logit(tweetPositive) = -0.3356 + 4.6*10^{-7}userFollowersCount + 3.6*10^{-3}numCrimes$$

where *tweetPositive* represents the odds of the tweet being positive over negative, *userFollowersCount* represents the number of users the tweet author has, and *numCrimes* being the number of crimes committed against the LGBTQ+ population that day. The p-values for the number of user followers and the number of crimes is about 0.1925 and 0.008, respectively, and the AIC is 89269.

	Estimate	**P-value**
Intercept	-0.3366	$2*10^{-16}$
User Followers	$4.601*10^{-7}$	0.1925
Number of crimes per day	3.609*10-^{3}	0.0076

It is immediately noticeable that the number of followers is not statistically significant, with a high p-value of 0.19. Not only is the p-value quite high, but the log odds is very low, meaning that the number of followers seems to have no effect on the tweet sentiment, at least when taking the number of crimes that day into account. Surprisingly, the number of crimes is a statistically significant predictor of Tweet sentiment. However, the actual log odds for this factor are also quite low. Doing the math, for every additional crime committed that day there is about an $e^{0.0036} = 1.003$ increase in the odds of a tweet being Positive rather than Negative, which is negligible. Despite being a significant predictor given these variables, it fails to be of practical significance.

There are several reasons why this might be the case. Firstly, the number of hate crimes on a daily basis might be too small of a time scale to have any effect. If we want to analyze the relationship between the trend in hate crimes and the tweet sentiment, a longer time period such as a month might capture the trend better. Secondly, not all the tweets in the dataset are related to the LGBTQ+ community; the word "Gay" in particular has many connotations and is not always used to describe the community. Thirdly, not all hate crimes are reported or recorded, so these crimes are unable to be a factor in this model, even if they are significant. Unfortunately there is little we can do about the third reason: if a crime is unrecorded we have no way of obtaining that data, but we decided to see what we could do with the other two reasons.

The following datasets are an attempt to address these issues. The first is one with the same factors as the first model, except that instead of having the number of crimes for that day as a factor, it has the number of crimes for that month. The second is the same except that it only includes data about tweets about the "LGBT" keyword specifically. This is done to ensure that the tweets we analyze talk only about the LGBT community.

Crimes per month

	Estimate	**P-value**
Intercept	-.4727	$2*10^{-16}$
User Followers	$4.614*10^{-7}$	0.192
Number of crimes per month	$3.519x10^{-4}$	$6.64x10^{-6}$

Containing "LGBT" only*

	Estimate	**P-value**
Intercept	0.753	0.0005
User Followers	$-6.782*10^{-7}$	0.5396
Number of crimes per month	$-4.857x10^{-5}$	0.8948

The dataset in the presentation was incorrect - this is the corrected version*

The first gives us similar results as our original model, significant p-values but low estimates. Our second model, however, gives us very high p-values, meaning that when filtering by keywords more specific to the LGBT community, the number of crimes becomes no longer statistically significant. Ultimately what

these models tell us is that the LGBTQ+ community is largely unaffected by the number of crimes and the Date, at least within the confines of the model and time span.

We further explored this by grouping rows with their state and year features. We created a view that showed the counts of positive and negative sentiments, the hate crimes, the overall sentiment for year and year and state, and the difference between positive and negative sentiments for that instance.

Year	State	Positive	Negative	crimes	Overall
2015	California	118	183	837	0

This view was useful in determining two things: the significance of the difference between Positive and Negative sentiments and the significance of the difference between the number of hate crimes and the negative sentiments. The first test is important because we want to see if one of the sentiments are repeatedly larger or smaller than the other, which makes sure that the data is not skewed in nature. The second test is important because we want to see if the number of hate crimes are constant or changing, since if they are constant, then hate crimes are independent of each other and there is no reason to do any further analysis on the data. To remove confounding data from the test, we tested the ratios of hate crimes over population per 1000 people. We want the number of hate crimes to be dependent on data, so we can justify the creation of a story and explain any correlation between hate crimes and twitter LGBTQ+ sentiment scores. We will create two hypothesis tests to test the significance of these means:

Test 1:

Null (H0): Positive - Negative = 0

Alterative(HA): Positive - Negative is not equal to 0.

Test Statistic = -1.537, Degrees of Freedom = 480, P-value = 0.1249

Test 2:

Null (H0): All ratio of hate crime over population sizes are equal

Alternative (HA): At least one is not equal

F-value = 56.27 P-value = 1.09e-12

Here we can see that the difference between Positive and Negative sentiments are not significant with an 0.05 significance level, therefore we fail to reject the null, and prove that the data is not inherently skewed by dependence on features unrelated to the tweet. In the second test, we proved that the means of the number of hate crimes are not constant, so we reject the null and conclude that there is a significant difference in hate crimes. Therefore we can conclude from the data that the number of hate crimes are dependent on the sentiment differences.

We did also briefly create a handful of linear regression models using the keywords present in the text of the tweet as a factor for sentiment prediction in addition to location, date, and monthly crimes: investigating whether the addition of that keyword was an influential factor in predicting the sentiment of the LGBTQ+ community for a particular day. Only two keywords from our set were considered statistically significant: LGBT and LGBTQIA. These may be because those using those terms are already fairly likely to be pledging their support for the community, as well as offering a positive message as

opposed to a negative one. Unfortunately, date and location were once again significantly more important predictors than the keyword predictors, which while they did fall below our 0.05 alpha level, were still significantly higher than date. We also tested to see if predicting for a rolling average would see any significant changes, because we were concerned that the single day sentiment was too variable.

Notebook Description

We implemented each model in either an RMarkdown file, standard R file, or Jupyter Notebook. For each of them, the general process was as follows: download and read in the data locally, filter data by the desired columns, adjust variables to ensure they can be properly used in the model (for example, converting ‘Positive’ and ‘Negative’ to 1 and 0 respectively for the logistic regression model), and plugging in the data to create the model. From there, we use the resulting output for analysis and creating data visualizations.

Discussion and Conclusion

From our methods and algorithms, we learned that when there is a lot of noise in data like ours, it is difficult to get a clear correlation between two possibly correlated factors. While we expected there to be some noise, we also expected it to be a bit clearer than how little overall there was. The correlation present seems to be primarily in regards to environmental factors such as population and location rather than in response to the number of crimes, which makes sense for our data.

It was more difficult than expected to filter out bot tweets, and given the small proportion of them present within our dataset, we ultimately decided against filtering them out as it was not worth possibly filtering out possible human tweets alongside them. Additionally, it was more difficult than expected to get statistics on the full dataset instead of our subset due to the sheer number, as our machines struggled to create even basic graphs or histograms from it. It was also somewhat difficult to even draw conclusions from our data, as there was so much noise present that it made it difficult to try and draw any overall conclusions.

We learned that despite how robust pandas is at handling large amounts of data, it still struggles to visualize it at such a high scale as the full 2-million tweet dataset. It was able to subset it to a smaller proportion well, as well as return some overall counts, but when it came to displaying even basic histograms or graphs, the software really struggled.