

Project Overview

For this project, I decided to scrape tweets from Twitter with the #guncontrol and #gunsense to analyze their sentiment and determine which political party the person could have belonged to. I was hoping to learn how people felt in regards to the issue and which parties used what hashtags because those who use #guncontrol are typically against it, and #gunsense was typically for people for the issue. I wanted to see if the data matched the current day thoughts that the media portrayed. Additionally, I thought it would be interesting to determine if there was a correlation between sentiment and objectivity.

Implementation

The first step to executing this project was scraping the tweets from Twitter based on a certain hashtag. I used the pattern module and modified some code that I found in the documentation to scrape 100 tweets from each hashtag. I had to create a list for each hashtag where I added each tweet to. After that, I picked the data so that I didn't always have re-run the query and ensure I was working with the same data each time. I saved the 100 tweets from each hashtag into their respective objects: `reloaded_guncontrol_tweets` and `reloaded_gunsense_tweets`. I chose these names so it was easy to identify which group of tweets I was working with.

The next step once I had all the tweets in an object, I wanted to do some analysis to see if I could determine which political party the people who tweeted were. I thought that developing the program myself would probably be too complicated so I started to do some research online and I found that Indico.io had a package that actually was able to breakdown a saying to determine what percent of a political party the person was based on what they said. So after signing up for Indico to get the API key, I ran each tweet through the `indico.political` function. The result was that each tweet had a dictionary where the political party was the index. I made a design decision here to put all the code for analyzing and separating the party percentages under one for loop rather than running the political analysis, putting all the dictionaries into a list and then working to separate them from them. After separating the party percentages into a list, I then found the mean using the numpy package. This was done for both tweets with the #guncontrol and #gunsense.

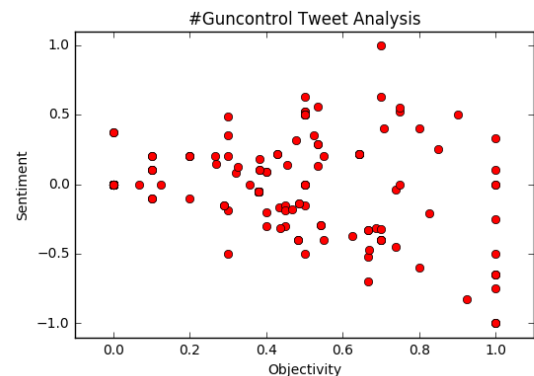
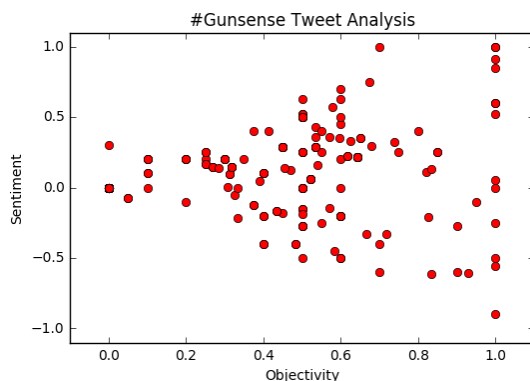
While I found the analysis for the political party breakdown to be quite interesting, I wanted to also do a sentiment analysis to determine the feelings people had and which hashtags they used. An additional part of the sentiment analysis was that you could determine the objectivity of a statement. It's important to note that for the sentiment analysis a 1 means positive and -1 means negative, and for the objectivity, the 1 means subjective and 0 means objective. So after running the analysis, I knew that the only way to further analyze the data I would have to separate the tuple so that the sentiment scores (the first element) were in one list and the objectivity score(the second element) was in another. After doing that, I felt the best way

to understand the impact would be to visualize the data, rather than just viewing the raw numbers so I created a scatter plot using pyplot. Finally, I also conducted a linear regression model to further understand if there was a relationship between sentiment and objectivity. This again was done for both tweets with the #guncontrol and #gunsense.

Results

As I mentioned above, I used the Indico.io package to do a political analysis of the tweets I found on twitter. From this analysis, I determined that there seemed to be a pretty uniform distribution among the political parties for both #guncontrol and #gunsense tweets, but there were some slight differences that analysis could be derived from. For example, for #guncontrol, there are more tweets that exhibit Libertarian and Green parties as opposed to Liberal and Conservative. This was something I found to be interesting because typically, in the news and such you hear about conservatives and liberals fighting about this issue. After looking into the raw data that the package produced, it was apparent that there was a whole range of political parties that talked about this issue, which is why it's so popular.

So my hypothesis because I started this project was that tweets with #guncontrol would be positive because they are for gun control and that #gunsense tweets would be more negative because they are typically against gun control. I wasn't sure how objectivity would play into my analysis and was curious to see that. These were the two visualizations I created from running the sentiment analysis using the pattern module.



I found this analysis to be quite interesting because there is no real pattern here but there are some interesting insights that can still be determined. First of all, it seems that tweets that were more factual were determined to also be more positive in terms of sentiment. As the level of subjectivity increased, the variation in terms of sentiment increased as well, suggesting that as people share their opinions its more obvious whether a tweet was more positive or negative. I went one step further to see if there was a strong correlation between sentiment and objectivity by running a linear regression model however when I did so, the R-squared value for models for both #guncontrol and #gunsense were extremely low, indicating that a linear regression isn't a good fit for this model and that no real conclusion can be drawn from this graph. I did find it interesting that there was such a fluctuation in terms of objectivity and sentiment. For sentiment

it makes sense because people have such highly contended views in regards to this issue, but for objectivity it means that some people share just facts in regards while others their views. Finally, by taking the average of sentiment and objectivity, I learned that for #gunsense, the tweets were more positive which was different than I expected, but more factual, while #guncontrol were more negative but again more factual. The average of that suggests more objectivity in the tweets is good because it means people are using facts to back their ideas.

Reflection

From a process stand point, I believe that my code is easy to understand and quite clean due to the comments and object names that make sense. I think I could improve in using more techniques to try to get a better analysis of the data that was at hand. I think that my project had a good scope, but also want to recognize that 100 tweets is quite a small sample size concerning the number of tweets out there with these hashtags, so analysis from this project may not be applicable to all tweets with these hashtags. I've learned a lot in regards to the resources online and writing code that is clean and makes sense to the person looking at it. I wish I had known more techniques that I could have used. It would have loved to do further analysis on the political party breakdown but the code took quite a while to run so that was a hurdle that I wish I knew how to overcome. Additionally, I wish there had been a way to correlate the political party with the sentiment and objectivity scores to develop a deeper analysis.