

Sandy Ta

Project Overview:

I used Twitter as a datasource using the Twitter Rest API. My code extracts a certain number of tweets from Twitter and then performs a sentiment analysis on them. In my example, I was specifically looking at 100 tweets directed towards Jet Blue to gauge how users felt towards the company. I hoped to create a tool for companies to be able to assess how well they are doing in terms of customer satisfaction based off of what their customers were saying about them on Twitter.

Implementation:

The pattern Python package was particularly useful for accessing both the Twitter API as well as for conducting sentiment analysis. In addition to Python, I also used matplotlib which allowed me to create graphs with my tool. While I used my code to analyze JetBlue, I decided to wrap my code in a function so that it can be tested with a variety of companies taking any number of tweets as a sample size. While pattern allows for each tweet to be easily analyzed for their sentiment, I realized it is not very helpful for companies to see the sentiment score for individual tweets. Additionally, the sentiment that pattern's sentiment package displays is given on a scale from 1 (highly positive sentiment) to -1 (highly negative sentiment). The scale does not allow for people to quickly determine whether a tweet was positive or not.

In order to address these problems, I divided the sentiment scores into three types: negative, neutral, and positive to make it easier for companies to determine a tweet's sentiment. From these categories, I created a counter for each sentiment type which allows for the viewer to gain a holistic view of the total number of tweets for each sentiment category for a given sample size. To complement the total number of tweets for each sentiment category, I implemented a bar graph component to give a visualization of the results. Finally, I implemented code to display the sentiment type's results as a percentage as well. I decided to display the results as percentages as well since the percentages adds more meaning than purely just the numbers. For example, in a sample size of 40, a result displaying that there were 20 tweets categorized as having positive sentiment is not as impactful as knowing that 50% of the total tweets were positive.

One design decision that I made included wrapping my sentiment analysis tool in a function. Originally, instead of using a function, my code started with a ‘for’ statement that looked like this:

```
“for tweet in Twitter().search(‘@JetBlue’, count = 100)”
```

My thinking was that if someone wanted to use this tool for other companies, they could simply switch ‘@JetBlue’ with another company and the count with a different number. However, I ultimately decided that designing my tool as a function instead of as a for statement was the better decision. One of the reasons behind this decision was that I wanted for others to have the ability to build on top of this tool. It would be easier and take up less lines of code to simply have to call on a function when implementing it with other projects particularly if someone needed to call on the function multiple times than to copy and paste the entire sentiment tool’s code, which could quickly become redundant.

Results:

JetBlue:

the number of tweets that have a positive sentiment is 22 or 22.0 %

the number of tweets that have a neutral sentiment is 67 or 67.0 %

the number of tweets that have a negative sentiment is 11 or 11.0 %

United:

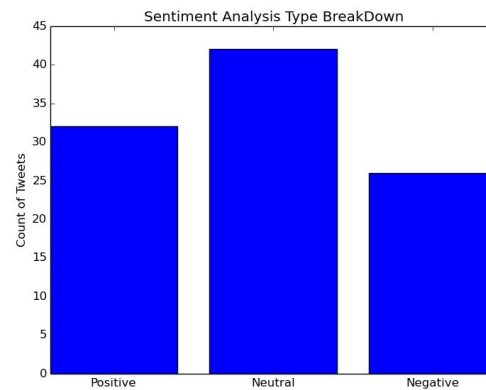
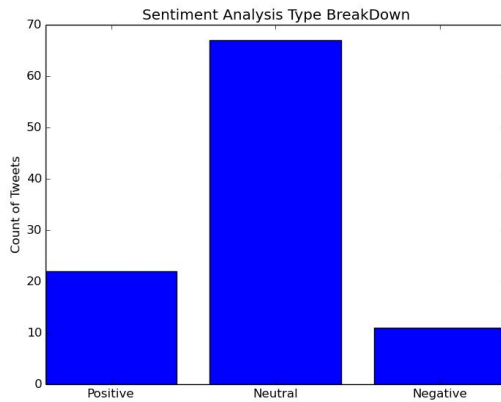
the number of tweets that have a positive sentiment is 32 or 32.0 %

the number of tweets that have a neutral sentiment is 42 or 42.0 %

the number of tweets that have a negative sentiment is 26 or 26.0 %

@JetBlue

@United



I tested my function by performing a sentiment analysis on the past 100 tweets directed towards Jet Blue. My results show that out of 100 tweets to JetBlue, 22% of them were classified as positive tweets while 11% of them were classified as negative tweets with the rest being neutral. I decided to compare Jet Blue's results to United Airlines to see how Jet Blue was doing in comparison to another airline. For the past 100 tweets directed towards United Airlines, 32% were classified as positive tweets while 26% of them were classified as negative tweets with the rest being neutral. In comparing Jet Blue to United, while United had a higher number of positive tweets directed towards it, it also had a much larger number of negative tweets directed towards it than Jet Blue. Overall, I would say Jet Blue is performing better than United for their customer satisfaction since they had much less complaints directed towards them whereas United's customers clearly have very polarized opinions on the airline.

In using my tool, companies such as Jet Blue are able to determine how well they are performing in comparison to their competitors. If it turns out Jet Blue is outperforming other airlines in terms of customer satisfaction as presented on social media sites such as Twitter, then the company can use this to their competitive advantage when marketing themselves towards potential customers. On the other hand, after using my sentiment analysis tool, a company such as United might uncover that there is a large number of negative comments directed towards them on social media. These results show that United needs to take extra actions in keeping their customers happy and that their current performance is unsatisfactory in the eyes of a large portion of their customers based on this sample size.

Hillary Clinton:

the number of tweets that have a positive sentiment is 38 or 38.0 %

the number of tweets that have a neutral sentiment is 47 or 47.0 %

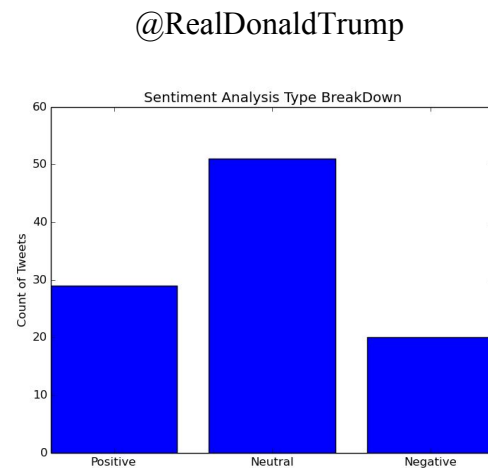
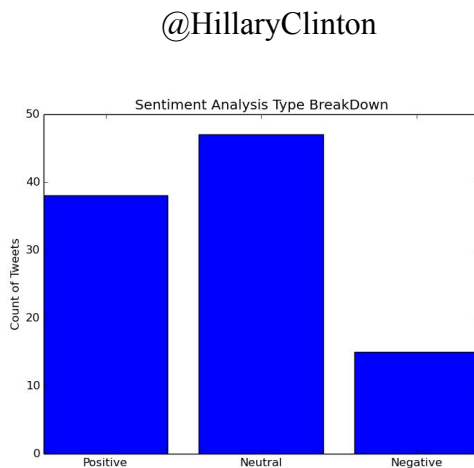
the number of tweets that have a negative sentiment is 15 or 15.0 %

Donald Trump:

the number of tweets that have a positive sentiment is 29 or 29.0 %

the number of tweets that have a neutral sentiment is 51 or 51.0 %

the number of tweets that have a negative sentiment is 20 or 20.0 %



While I intended for my tool to be used by companies to assess their customer satisfaction performance using Twitter, I was looking at other use cases for where my sentiment analysis tool can be applied. One example I came up with outside of company customer satisfaction is determining people's attitudes towards the presidential candidates. I used Hillary Clinton and Donald Trump's twitter handles to mine data on what perspective voters and media outlets were saying about them. Overall, I discovered that Hillary Clinton had 9% more positive

tweets directed at her compared to Donald Trump. Additionally, she also had 5% less negative tweets directed at her compared to Donald Trump as well. Despite Hillary Clinton having less negative tweets, both candidates still have high levels of negative posts directed at them. The results of this analysis suggest that Hillary Clinton has a better perception among voters than Donald Trump does. Overall, my sentiment analysis tool can be used in a variety of cases outside of just determining company and customer relations performance.

Reflection:

When creating my sentiment analysis tool, I think I did a good job at making my code as concise as I possibly can. I believe each component of my tool serves a purpose in being included. In creating my project, I consistently kept in mind my tool's audience (companies) and what kind of information would be most beneficial for my audience to know, which really helped me to vet what my tool should and should not include. While I used Pattern's sentiment module when developing my tool, one of the limitations of this is that the sentiment tool may mislabel a tweet and put it in the wrong category type. Particularly, some of the tweets I saw had connotations of sarcasm that would be difficult for a computer to pick up. I wish I had known this before I had started the project in order to look into more robust word banks for negative/positive sentiment analysis or even create one myself. In expanding upon this project, an interesting next step would be to see how the company's respond to comments with negative sentiment directed at them.