Meghan Tighe

**Text Mining & Analysis Mini Project: Write Up**

Project Overview:

I performed sentiment analysis on twitter posts related to the four houses of Hogwarts: Slytherin, Gryffindor, Hufflepuff and Ravenclaw.

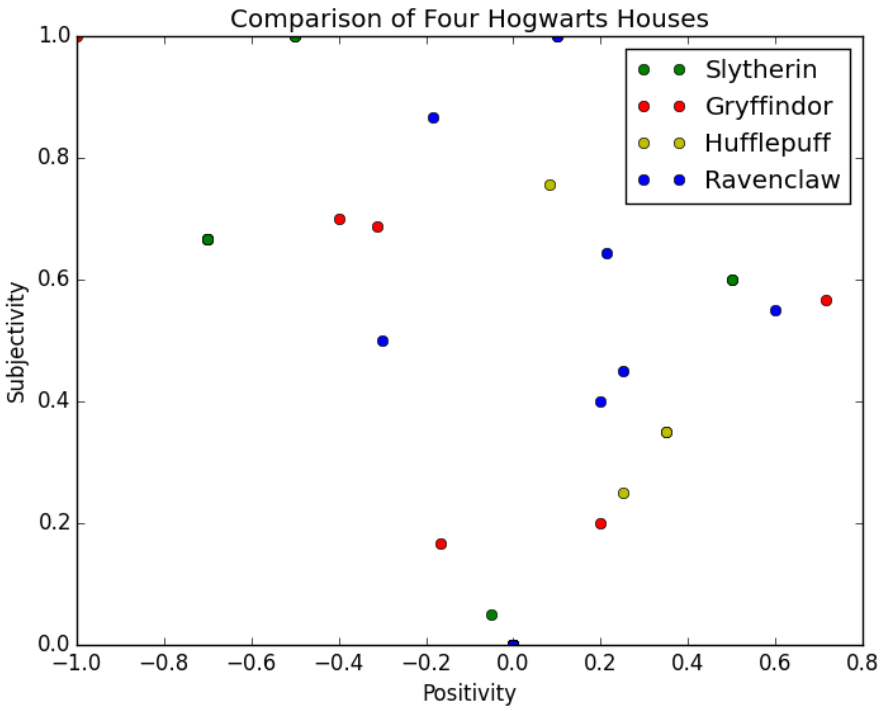
Implementation:

Initially, I tried to access twitter in the manner described in the project description. After working a few times, my code kept giving me weird errors and I determined (with the help of a ninja) that this was due to Twitter limiting requests for pulling tweets. I really felt that Twitter would work best for this though, so I found another way to do it (awesome description here if you're curious: http://adilmoujahid.com/posts/2014/07/twitter-analytics/ ). After finding this workaround, I was able to use tweepy to pull tweets that mention the four houses.

Once I had the tweets, I saved them into a text file. I wrote another script that opened the text file and extracted only the body of the tweet (excluding author, time, location, language, etc.) I was then able to iterate through this list and calculate the sentiment analysis of each tweet using pattern.

Finally, I plotted all of the tweets for each of the four houses, in separate colors. The resultant scatterplot is shown below.

Results:




Unfortunately, it seems that there isn't an overwhelming consensus on the sentiment of Twitter users towards the houses of Hogwarts. I'm somewhat disappointed but after scanning

through some of the collected tweets I think that this graph does somewhat accurately describe the sentiments.

When I originally got my code working, I tested with things that would have a more clear polarity. To be specific, I plotted the difference between tweets containing #yay and #shit. There was a clear correlation where the tweets containing the “yay” hashtag scored higher on polarity. At the time, I did not realize that this data or the corresponding graph would be useful results so I did not save them.

Reflection:

Learning how to use Twitter in spite of its limitations was actually quite challenging. To be honest, I think I'm glad that I didn't know how hard it would be because although the amount of work would have intimidated me, I now believe that it was worth it. I could have improved this by using pickling instead of saving the tweets directly as a text file. Although I didn't really use doctests, I carefully tested as I went through the process. I learned a lot through this project and think that the result is really cool!