Who Should I Follow on Twitter?
By Alisha Pegan
Feb. 23rd, 2017

My project shows the top Twitter usernames who posted positive tweets containing a relevant hashtag. The function in the program takes two parameters: a hashtag and an integer, n, for the top 'n' list. The goal of the program is to inform fellow tweeters, like myself, who want to follow other tweeters that post positive things about a certain topic and are well-liked. For instance, I want to know who I should follow on Twitter regarding urban design.

The key components of my code are:

1. Twitter search for tweets with hashtag
2. Store the data into an organized list with dictionaries for username, timestamp, text, and followers.
3. Pickle dump and load
4. Run a sentiment analysis for every tweeters using the Vader Sentiment Analyzer
5. Gather all the tweets with a positive sentiment, and sort it by most followers for each user
6. Print top 'n' usernames

I applied for a Twitter app, and made a python script containing my keys and secrets. I used the tweepy package which had the Cursor function that ensured I did not go over my rate limit since I am gathering a max of 1000 of tweets at once.

Next, I created a list where every tweet is a dictionary containing a username, timestamp, tweet text, and the number of other people following the user. The tricky part was understanding the data I got from Twitter, which was a dictionary with sub dictionaries and contained a total of 77 keys. Finding the keys was confusing, for instance, the username is tweet.user.screen_name. However, once the set-up was understandable, it was easy to store the data I wanted.

I pickled the data using the instructions given to use in the Mini Projects page.

Using the sentiment analyzer stated in the Text Mining Mini Projects page, I stored a new key in every tweet dictionary for the sentiment. That key contained a dictionary for the sentiment polarity score for the tweet text. There are four scores: compound, positive, negative, and neutral.

Because I like positive people. I gathered a list of all the tweets that had a positive sentiment and stored it into a new list. Then, I sorted the list according to the user's followers, from highest to lowest.

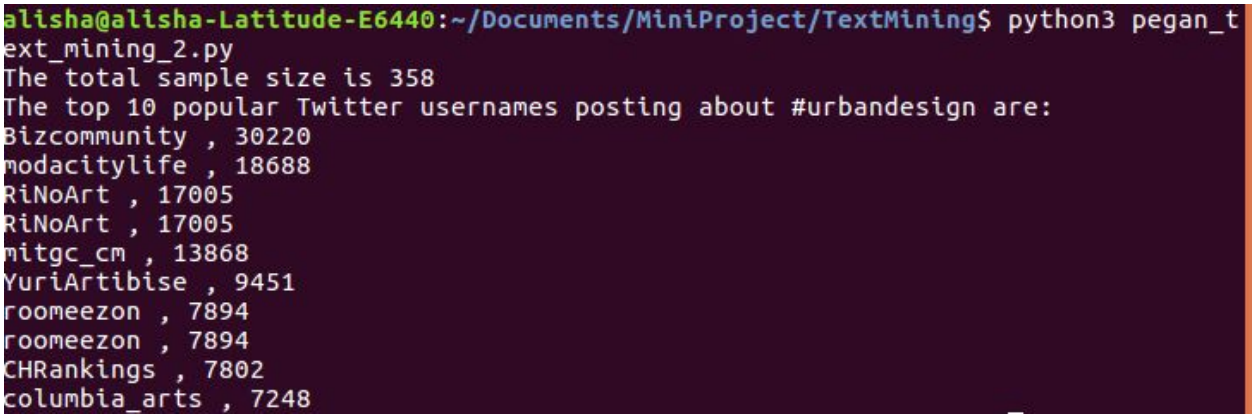
Lastly, I printed the statement for the top n usernames by printing the usernames of the first n elements in the sorted list.

Results

pop_tweeters('#virtual reality, 30)

```
alisha@alisha-Latitude-E6440:~/Documents/MiniProject/TextMining$ python3 pegan_text_
mining_2.py
The total sample size is 1000
The top 30 popular Twitter usernames posting about #virtualreality are:
TheVegasEscorts , 46881
815wrldtrvlr , 28183
EdTech_HigherEd , 24778
StartUpRealTime , 24359
digiindie , 21751
adirado29 , 21654
CeciliaGunadia , 17231
1hourpro , 13631
wiredprworks , 12723
RobinPelleck , 11981
TheTechMuseum , 10727
robertnsills , 9047
startapp , 7888
virtrealitytime , 7501
virtrealitytime , 7501
AHAMeetings , 6216
RockAtDell , 6003
YourFutureSA , 5778
juliastockings , 5680
hummingbigbird , 5455
ownaspace , 4694
LiquidHub , 4626
AuntMinnieEuro , 4556
yulbiz , 4533
timreha , 4531
SkarredGhost , 4360
SkarredGhost , 4360
viarbox , 4278
viarbox , 4278
Nick_Kokkinos , 4263
```

pop_tweeters('#urbandesign', 10)



Reflection

I found this to be an amazing project. I took me about three days (5-8 hours each day), which was a bit rough for my other assignments, but I enjoyed the entire process. I did not have a good plan for unit testing because I was confused on how to do that. I think having the results and seeing the code work was quite empowering, especially because I am very into data analytics for urban planning. It made me think about all the other data I could analyze. I wish I knew syntax for other modules like tweepy. I had a friend that helped me every step of the way. Without his help, I would have burned my computer a long time ago. I really really like this! I am would love to do this related to geographical location. My friend mapped all the colleges that stood in solidarity with the University of Missouri students who were fighting for racial justice in their university system. She did this by using Twitter and mapping the colleges that tweeted in support. I could see myself finding data like that easily now.