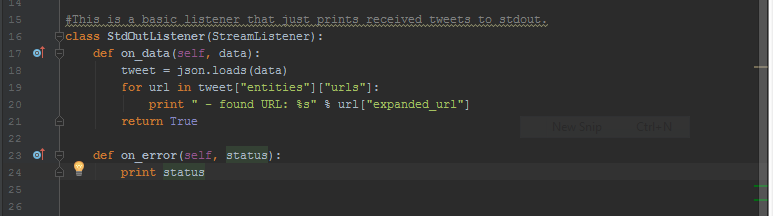
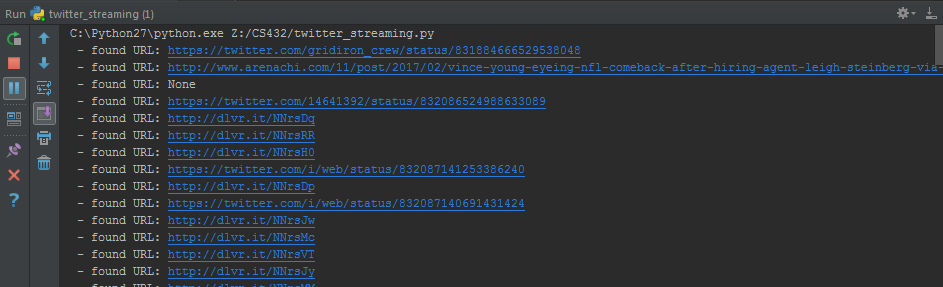
Write a Python program that extracts 1000 unique links from Twitter.

1.For the first question, I utilized the template given in the prompt, (<http://adilmoujahid.com/posts/2014/07/twitter-analytics/> and made a few modifications. When the template is ran we get the desired data, however it comes with excess information- we only required the URI from the tweets that were pooled. I inquired another classmate on how they were going about simplifying the output and I was told to check out the tutorial on this link, <http://code.runnable.com/Us9rrMiTWf9bAAW3/how-to-stream-data-from-twitter-with-tweepy-for-python> . I inserted a slightly modified version of the portion involving the JSON formatting and ‘bad character elimination’ into the template and was able to get a much cleaner output with only ‘found URL: xxxxxxxxxxxx’ being output if one was found in the tweet. Otherwise the output was ‘-found URL: None.

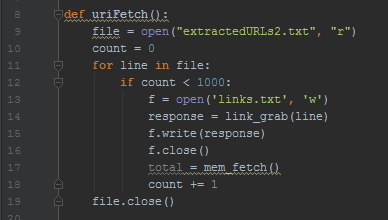


I took some time going through the links manually at random, they included a nice mix of twitter pictures, YouTube videos, and external links to other various sites. I had difficulty running the program through the Putty via the commands given in the tutorial on the first link so I simply copied the output window into a text file. The output of the program look like this-

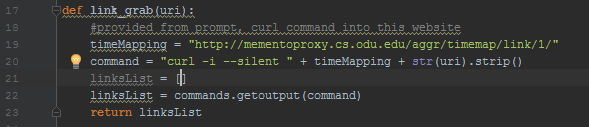


I was able to mine 5000+ URI’s using this method in a relatively short amount of time. For the purpose of this assignment however I trimmed the list down to 1000 unique links, and got rid of the portion that says ‘- found URL:’ in my final increment of my twitter\_streaming.py. It now just outputs URLs or None if none were found in the queried tweet.

2.For the second question, I created a new .py file, memento\_counter.py, and imported a few libraries that I thought I would need, including json, re, sys, & commands. I then read in the contents of my extractedURLs2.txt file. From there the program writes into a file the 1000 links from my input text file.



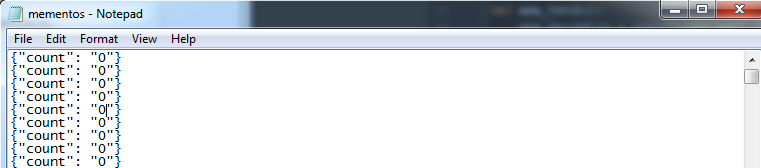
Next I attempted to make use of the time mapping link provided on the prompt. I did addition research on the use of timemaps, most of which were being used in conjunction with twitter- including <https://github.com/maxcw/time-maps> and <https://districtdatalabs.silvrback.com/time-maps-visualizing-discrete-events-across-many-timescales> . I used the curl command to send the data from my text file into the data field on the site using-



And stored the return values in an out-file title ‘mementos.txt’.

The vast majority(all) of the mementos were 0, as we discussed in class I believe.

The output of my text file looked like-



It is impossible for all of the links to have mementos of 0. When I began reworking possible solutions I accidentally saved over my previous work in my memento\_counter.py therefore I lost the code that generated the 0 error. I ran out of time reworking this portion of the question.