

# COSC 481 – Miniproject #2

Task: create a twitter scraper, create a JSON parser, and do some rudimentary data visualization on a bunch of recent tweets

For this project you will create a python program that will do the following things:

1.) Pull tweets from Twitter for the following hashtags and store the resulting JSON data:

```
#csforall  
#equality  
#stem  
#yolo
```

Do this only for the two-day window from 2/27-2/28. Get as many tweets as you can. Keep in mind that the standard API will only let you look 7 days back, so make sure you do this data collection near the window! It is fine to store this data as a plain text file. Multiple plain text files are fine.

2.) Determine and print out for each hashtag:

```
Most frequent user  
User with most followers  
Tweet with most retweets  
Tweet with most favorites
```

3.) Create a bar chart for each hashtag with time along the x axis and frequency along the y axis. Use the data pulled in #1, and collect the frequency in 6 hour chunks. For instance, for the first element in your #yolo chart, you will figure out the frequency of #yolo tweets from 00:00 eastern to 05:59 eastern on 2/27. Draw the scatter plots to compare each frequency with each other frequency across all of the other hashtags. For instance, you will need to draw the scatter plots for yolo 0-5:59 on 2/27 with the 0-5:59 on 2/27 for csforall, equality, and stem. You don't need to compare them across the other time brackets. There will be a lot of scatter plots. Are there any that are correlated? Put your answer in the comments of your python source.

4.) Submit your sufficiently commented source file and data file(s) via Blackboard by 11:59pm on 3/4. This project will be worked on in teams of two or three. Teams will be assigned in class on February 19<sup>th</sup>.

Potential deductions:

- 100 Does not interpret/crashes
- 25 Missing data for one or more hashtags
- 10 Missing barchart (each)
- 5 Missing barchart element (each)
- 5 Missing scatterplot (each)
- 10 Missing correlation comment
- 25 Lack of comments