Social Media and Computer Science

PLEASE READ ALL OF THE BELOW BEFORE ATTEMPTING THE CHALLENGES

In this workshop we will be investigating data from social media using some python code. I have tried to make sure that the workshop is accessible to everyone, regardless of how much experience you have with python.

There are two parts to this workshop: coding and data exploration. If you are more interested in one or the other then please do concentrate on that. If you are more interested in coding, see what you can do to improve the code provided. If you are more interested in data and investigation, feel free to skip to the solutions and use those to investigate the questions provided. I would suggest having a look at both – coding and investigation are both extremely useful skills! The **Extension Challenges** are suggestions for what more experienced coders might want to attempt with the code provided.

Open Section1-IntroToData in replit.

This simple bit of code loads the 3000(ish) most recent tweets by Barack Obama and displays the text of the most recent tweet.

Read through the code. Notice that each line of code is numbered on the left-hand side. Lines with a # at the beginning are greyed out – these are comments, notes to help us understand the code. They are not executed when the code runs, so they have no impact on what the code actually does.

We won't need to edit line 2, this is just adding some functions to python that we need to work with twitter data.

Read the comments on lines 4 and 7 to understand what the code is doing.

Can you figure out how to:

- Load tweets from Taylor Swift instead of Barack Obama (Hint: look carefully at the code and the files available in replit. If you have problems, remember that python is very sensitive to capital letters being in the right places)
- Print the number of **likes** that a tweet got instead of the **text** of the tweet. What about the number of **retweets**
- Print the second most recent tweet instead of the most recent. What about the 432nd most recent?
- Add an extra line of code so that you're printing the text and the likes of a tweet when you run the script. (HINT: remember you'll need to change the tweet number in both lines to get data from the same tweet)
- Try to find some tweets from each person that are more or less popular (as measured by the number of likes and retweets). What topics make for a popular tweet - what are people responding to?
- Compare this for Barack Obama and Taylor Swift.

NOTE: I would suggest manually looking through the data a little to get used to it, but the solution to the extension challenge will show you just the most popular tweets.

Extension Challenge: Write a script that will print tweets to the console if they have higher likes than a certain threshold (e.g. print all tweets that have more than 2 million likes) [you will need to use an if statement to do this]