INF 510
Homework 4 (20 points)
Due Sunday, March 31st at 11:59pm (via blackboard)

Find three (3) data sets on the web that are of interest to you! (If there isn't any data that's interesting to you, you're in the wrong class/should drop):

1. One must be "scrape-able" (i.e. not available via an external API)
    a. "Data set" is not just a simple web page
    b. Needs to be something that requires automation to obtain
    c. If you could just cut-and-paste the data, it's not a data set
    d. Data should be somewhat structured (maybe). If what you want to scrape are large text blurbs or images, that's fine, as long as there is an algorithmic way of obtaining a large number of these items, while attaching meta-data to them.
    e. Again, if you could obtain the data manually vs. writing a script, it doesn't count! Good rule of thumb: If it would take you less than 10 minutes to manually obtain this data, then it doesn't require automation.
2. One must be available via external public API
    a. You have to be able to access it without a ton of trouble
    b. You can use OAuth if needed, but if you can't get at the API, it doesn't count
    c. If you decide to use an API that requires OAuth or some other authentication mechanism, ***you must test that you can access this before hand***
    d. The API must allow a "reasonable" number of free accesses per day/hour/lifetime, etc. Enough for you to be able to test and deploy assignments and projects based on it
    e. There is a list (by no means comprehensive!) of public APIs here: https://github.com/toddmotto/public-apis/blob/master/README.md
        i. Note which require authentication, and which do not!
3. The third can be either (API or scape)

You don't need to be in love with all three data sets, but at least one should be interesting to you! If you're a Ph.D. student, something related to your research would be great! If you're a MS student, here's your chance to work on data of your choosing!

For *each* data set, you need to include:

1. The URL/API endpoint of the data set. For APIs, include links to the API endpoint **and** the documentation for the API

2. A brief description of what the data set is/contains
3. A brief (2-4 sentence) of what you might actually *do* with the data, if you had it

EXTRA CREDIT (10 points)**:** You can receive 10 points extra credit for picking three data sets that all integrate together, and describing how this interaction would work (use Lab 9 as an example; we scrape football data, and then reference two APIs to enrich that data).