# COMP 598 Homework 6 – Shell Customization and Data Collection

30 pts
Assigned Oct 30, 2020
Due Nov 9, 2020 @ 11:59 PM

**This is an INDIVIDUAL Assignment – each student's work must be their own, each student completes this assignment, there are no teams for homework 6.**

Non-standard (i.e., built-in) python libraries you can use:

- pandas
- requests

## Task 1: Inspiration on login

What if every time you logged into your EC2, it printed out a nice, inspirational word of wisdom?  Now that we know about .bash_profile and .bashrc, we can do that!

Add code to your .bash_profile script that picks a random quote out of the inspirational quotes file below and prints it to out (i.e., echos it).  The location of the inspirational quotes file:

https://gist.github.com/JakubPetriska/060958fd744ca34f099e947cd080b540

To do this assignment, you'll need to download this file.  Store it as ~/.data/quotes.csv (you can assume that it will be here as well when we test your script).

Your script should print the quote out on one line (it should be surrounded by quotes) and then the author underneath following a tilde.  For example:

    "An apple a day keeps the doctor away."
        ~ Some Body

Spacing on the line itself is entirely up to you.

A couple tips as you work on this:
- In the file, some of the quotes don't have an author on that line.  In this case, it's fine – you should just print a blank author (but the tilde will still be there).
- bash <script_name> will run a script file.  For example, "bash .bash_profile" will run your bash script.  "source .bash_profile" will also do this.
- You'll need to write this script entirely in bash shell script… no python here!  This also means no supporting libraries or methods.  The only thing you'll use are (1) shell script, (2) pipes, and (3) standard unix commands (e.g., cut, head, tail, etc…)
- To select the line, take a look at the $RANDOM bash environment variable, head, and tail.
- To get the quote, check out the cut command.
- One thing you should read up on as you work on this are the shell evaluation techniques `…` and $(…).

## Task 2: Election Data Analysis (20 pts)

We're in the final days of the presidential election, let's collect some data about this from Reddit.  We'd like to get a sense for the hottest content in the /r/politics subreddit over three consecutive days.  You're welcome to pick whichever three days you want, but it will be the most interesting on Nov 2, Nov 3, and Nov 4.

Write a script "collect_hottest.py" that collects the 500 hottest posts in the subreddit specified.  It should run as follows:

COMP 598, Fall 2020

```
python3 collect_hottest.py -o <output_file> <subreddit>
```

For this task, you'll use the /r/politics subreddit. So it would be run with <subreddit> set to "/r/politics".

Run it at roughly the same time on three consecutive days, saving the files as <yyyy><mm><dd>_politics.json. Save one post per line - where the post is the dictionary under the "data" key.

Keep this data around - we'll use it in our assignment next week (HW 7).

## Submission Instructions

Your MyCourses submission must be a single zip file entiled HW6_<studentid>.zip.  It should contain the following items:

- scripts/
  - o   .bash_profile - the script for Task 1.  This script ONLY has to print out a random inspirational quote when run.  It should assume that the file ~/.data/quotes.csv is available.
  - o   collect_hottest.py – the script for Task 2
- data/
  - o   <date1>_politics.json – the 500 posts you collected from the first date
  - o   <date2>_politics.json – the 500 posts you collected from the second date
  - o   <date3>_politics.json – the 500 posts you collected from the third date