Lab 4: July 31, 2019

Agenda: quick set up of JupyterHub, review of classes and functions & testing+redeveloping broken code, more data input practice & visualization, coding exercise / practice python interview ?s

Part I: JupyterHub

For Thursday's (8/1/2019) guest lecture, we're going to make use of JupyterHub. Our lecturer has requested that everyone be up and running before class, so we're going to take (hopefully) <5min to get set up on everyone's personal computers (famous last words ☺)

Best we can tell, the first necessary step to access the AAE hub from your laptop is to connect to the campus VPN. If you've never done this, directions below.

Windows users: https://kb.wisc.edu/aae/81242
Mac users: https://kb.wisc.edu/aae/81248

From here, open your terminal/shell. The first time you access the Hub, you must connect to the server using your NetID. Type at command line:

```
ssh {YOUR NET_ID HERE}@jupyter.aae.wisc.edu
```

If you're having connection issues at this point, raise your hand and we'll get your sorted out.

If you've successfully connected to the server (this should require you to enter your NetID password), go to https://jupyter.aae.wisc.edu in your browser of choice. Create a new notebook, play around a bit, make sure everything works. When you're everything is good for tomorrow, close your notebook in the browser and you can close your *ssh* remote connection at the command line with:

```
ENTER~.
```

(That is ENTER TILDE PERIOD, all in succession). Note: from now on, you should just be able to access the Hub at the link above without logging into the server.

Part II: Class/function review + Testing (<30 minutes)

Overall, you guys did a nice job on Lab 8.2. It was a challenging lab – there was a lot that went into building up the script from scratch. That said, there were some common problems in your code. Several of you wrote scripts that passed Zybooks' tests, but nonetheless result in undesirable logical errors if the parameters of the problem change just a little bit…

Your task today is to review some code that is built from mistakes that several of you made. This code is available on GitHub (Lab4) in the file *troubleshoot_me.py*.

In particular, **you now need this code to run when there are *n* customers (so an unspecified number) who have a shopping cart**. You have two objectives:

1. Come up with a test that *breaks* the code as currently written when there are more than one customer instances.
2. Explain in detail any errors or bad Python practices that you see in the code
3. Correct the code so that it works in this more general setting.

What we're doing here is essentially *test-driven development*. We haven't done much of this, but you should 100% be aware of the principle. Please skim this article!: https://medium.com/@bethqiang/the-absolute-beginners-guide-to-test-driven-development-with-a-practical-example-c39e73a11631. This kind of code development has become the norm in data engineering, but a bit less common for data scientists/analysts.

Part III: Data input + Visualization (1 hour)

I'm providing you with some data on wine reviews that have been scraped from the magazine WineEnthusiast[1]. You can access them on the lab's GitHub – note that the file is provided both in *.csv* and *.json* formats. Make of that what you will, but the files should be identical in terms of content.

Here are your tasks:

1. Write python code to input the data as a:
   a. List
   b. Dictionary
   c. DataFrame
2. Create **two novel visualizations/tables** from this data that highlight features you find interesting or curious. A histogram of wine prices is not novel nor interesting… please make more effort than that. If you put in some thought into this exercise, you should walk away with a bit more confidence creating visualizations!

Part IV: Coding exercise @ the whiteboard. (<1 hr)

To be discussed during lab!

PLEASE SUBMIT YOUR CODE (.py files or jupyter notebook) FOR TODAY'S LAB IN YOUR PERSONAL FOLDER ON GITHUB.

---

[1] Source: https://www.kaggle.com/zynicide/wine-reviews/downloads/wine-reviews.zip/4