CICS 397A – Fall 2019

Homework 2

Due Monday, October 21st at 11:59pm

You are encouraged to discuss the assignment in general with your classmates, and may collaborate closely with 1-2 other students on the design and logic of your solutions. If you choose to do so, you must indicate with whom you worked. In addition, the code you submit must be entirely your own; two students submitting the same code will be considered plagiarism.

Code must be written in a reasonably current version of Python (>3.0), and be executable from a Unix command line. You are free to use Python's standard modules for data structures and utilities.

Part I: Data Scraping

For this part of the assignment, your task is to create your own data set using content scraped from the web using the `scraper.py` code that we looked at in class. The data set you create can be drawn from any source that interests you — for example, you might scrape movie ratings, sports scores, headlines, financial indicators, political polls, etc.

The data set you create must have at least 200 instances (rows) and 3 attributes (columns). Once assembled, the data should be exported to a csv file Using Python's `csv` module.

Part II: Data Wrangling with Pandas

The `mustard_analytics_pandas.py` file contains stub code for carrying out different aggregations and manipulations of the car refueling data from Homework 1. As with that assignment, your job is to fill in the function bodies to produce the output described in the comments. This time, however, you will do all the data manipulation using the `pandas` module. In most cases, your solution should only take a couple of lines of code. That said, getting the hang of the syntax can be tricky, so start early.

Grading

We will run your program and examine the output for correctness. Note that we may run it on a data file with a different number of rows (but the same column format), so your code should be

able to handle data files of arbitrary lengths. Your grade will be determined by how many of the exercises achieve the correct output, with partial credit being awarded.

What to Submit

You should submit:

- Your scraping code, contained in a file called `scrape_data.py`
- The data file you created, called `awesome_data.csv`
- A modified `mustard_analytics_pandas.py`
- A readme.txt, containing
    - Your name
    - Your favorite ice cream flavor
    - Anyone with whom you worked with on the assignment (see note above)
    - Notes or warnings about what you got working, what is partially working, and what is broken