

HW Extra Credit (HW #6 Dropped)

[Re-submit Assignment](#)

Due Nov 22, 2019 by 11:59pm **Points** 4 **Submitting** a file upload **File Types** html

Complete these exercises and submit a single Jupyter Notebook file (in .html format, not .ipynb) that contains your responses by **11:59PM on Friday 11/22**. Because this is extra credit, late assignments will not be considered. Additionally, only .25 of 1 point will be given for exercises with any errors (see rubric).

The notebook should be well organized. Each section should be **clearly labeled with the exercise (and part) that it addresses** (e.g., Exercise #1a, #1b, #2) in a Markdown cell block. Use (clear and concise) comments as needed to help describe each step of your process. All notebook cells that contain essential steps should be executed and the output should be visible, so as to demonstrate your successful completion of the exercise. If you cannot complete an exercise in its entirety, you should make an effort to demonstrate your intermediate progress in order to maximize partial credit, and move forward as best as possible. You may submit any written answers to the exercises in the notebook as text cells.

Academic Integrity: Each student is expected to submit his or her own original work. You may collaborate with your classmates on the concepts of the homework assignment, but you should not submit the same documentation for any part of the assignment. Submissions that contain significant similarities will be reported directly to the Office of Student Conduct.

Background

This assignment builds off of HW #4 scraping data from <https://pokemondb.net> (<https://pokemondb.net/>).

Exercise #1 - All the photos! Sort of...

Step 1.1: Start with your deduplicated pokedex data. Just like in **Step 2.4 from HW #4**, create a new column in the DataFrame called *sample2* that tags every 12th pokemon. (This will reduce our sample size.)

Step 1.2: Scrape the photo URLs for each pokemon in sample2 like in **Step 3.1 from HW #4**.

Step 1.3: Display the photos for all of the sample2 pokemon in your Jupyter notebook. Example code for how to display multiple images in a loop is below.

```
Display(Image(url=____, unconfined=True))
```

Exercise #2 - Rescrape and Clean the Location Table

Step 2.1: Using BeautifulSoup, scrape the Location table **for Bulbasaur** properly and add it to a DataFrame. When multiple columns are combined, separate the columns and duplicate the location information. For example, the column "Red Blue" that contains "Pallet Town" will become two columns-- "Red" with location "Pallet Town" and "Blue" with location "Pallet Town." Transpose the DataFrame so the video game (e.g. "Red") is the column name.

Step 2.2: Generalize Step 2.1 for all pokemon in the original sample (the sample from **Step 3.1 from HW #4**). Append all the data together. If information is not available for a pokemon in a particular game, the column should contain NaN. Display the entire table.

Exercise #3 - Location Analysis Done Right

Step 3.1: Join the pokedex data to the location DataFrame created in Step 2.2 above. (Exclude pokemon that are not in the sample.) For the locations in pokemon game X, calculate the average total points for each location. Which location has the highest average total point score?

HW Extra Credit Rubric				
Criteria	Ratings			Pts
Exercise #1 - All the photos	1.0 pts Correct	0.25 pts One or more errors	0.0 pts No attempt or output not displayed	1.0 pts
Exercise #2.1 - Location table for Bulbasaur	1.0 pts Correct	0.25 pts One or more errors	0.0 pts No attempt or output not displayed	1.0 pts
Exercise #2.2 - Location table for all pokemon	1.0 pts Correct	0.25 pts One or more errors	0.0 pts No attempt or output not displayed	1.0 pts
Exercise #3 - Analysis	1.0 pts Correct	0.25 pts One or more errors	0.0 pts No attempt or output not displayed	1.0 pts
				Total Points: 4.0