Jameyh_narrative

Week 10, HW 9

Discussion Problem 1 (in py file)

Discussion Problem 2

The length is 1000 because there are 1000 records in the file. It's also 1000 because it's only a sample of all the horse files. We are looking at a data sample from the wiki horse query. The data structure is a list. Structurally, the dataset contains 1000 items in the list which happen to be dictionaries. Each dictionary then holds the information for each horse in key/value pairs. The typical keys within each dictionary (if not missing) are "horse", "horseLabel", "mother", "father", "birthyear", and "genderLabel."

Coding Problem 2

For this problem, the goal was to build a dictionary with frequency counts for a given key/value pair of appearances of values and/or missing value. The key used was "birthyear." I created an empty dictionary and started a for loop where I looped over the dictionaries in the list and stored the returned value from the clean _missing function to the variable <year>. This value represents either the value itself or the string "missing" if the value is missing. Then within the loop, the program checks to see if the key is in the dictionary (checks for membership). If so, it increments the count for that value. If not (base case), it adds it in to start the count for that value. The program outputs how many horses were born in the years included in the dataset and how many horses were missing a birthyear value.

Coding Problem 3

This problem gave me some practice working with .items() again and getting used to accessing key and value by position in a list. I also started to get my mind around the concept of a row around this time working on this problem partially, thinking about it at this basic level. Here, the row is being made in the loop and it's a list with 2 items and there will be as many rows as there are counts of birthyears (including the missing value one). Csvout.writerow(row) is used in the loop and I started to understand how this is different from Coding Problem 4. I was at first confused by the writerow vs writerows.

Coding Problem 4

For whatever reason, I had a bit of a time putting together the pieces even though I thought individually they made complete sense to me. Initially, I saved the value returned from the clean_missing(key, record) function and passed that variable to get_id_from_url (url_string). While that works, I wanted the practice of passing the function to the function and also it made

for easier reading when it came to the values that required that extra step. I did have to make sure I was performing those in the right order and that I could visualize what was happening at each step. Once I saw that clean_missing (key, record) returns “missing,” I could understand how it was being used in the other function. Initially, I reviewed the functions in the opposite order and I kept asking myself, how would it get the string “missing”? Also, I’m further practicing the adding rows within a loop and understanding the 2d list building mechanisms. I’m feeling much more comfortable with that process now. I also started to get familiar with converting to the csv writer object which is useful.

Weekly Summary

This was another one of those revelatory weeks overall. I had a lot of practice between finishing up some tweaks with the midterm and figuring out the extra credit and doing this homework. This is definitely the week I got practice with 2 dimensional lists and have tried to really make sense of how rows are constructed as data gets more and more complex. I was stuck for quite a while on understanding how to portion out the data the way I wanted it to read out and composing that within the loop properly. I think now I have a much better sense of it and as I practice it more, it will become ingrained. At least, that is my hope!