Homework 5

Intro to Data Science Fall 2024

Due Wednesday 10/9

Name:

Submit your work to me in two steps: Problems 1 and 2 should be written on a piece of paper and turned in to me. Problem 3 should be submitted as a text file in Brightspace containing your code so I can copy and paste it to run it.

Problem 1. The zoo has a feeding regiment for it's animals written out as:

"The mammals need to be fed as followed: the elephants needs 375 pounds of food a day, the chimpanzee needs 3.75 pounds of food per day, and the hamsters need 2 ounces of food a day. The reptiles need: the cobra needs half a pound of food once a week, and the crocodile needs 17 kilograms feed every three days. Birds: The ostritch needs to eat 10 ounces of food a day, and the parakeet needs to eat 2 ounces of seed three times a day."

Perform data munging to turn the data presented in these instructions into a structured table containing the following information: species, animal type, feeding frequency, and amount of food. Be sure your units are all consistent with each other!

Problem 2. The following table represents data from a 10 day hiking trip on the Apalachian trail. Clean the data by removing obvious noise. Then complete the missing data, following any pattern if there is any.

Day	Net elevation gain (ft)	Weight of backpack (lbs)
1	200	30
2	52	
3	-420	26.5
3	-360	24.1
5		-23
6	121	22.5
7	–200	212
8		20
9	-600	
10	800	18.8

Problem 3. Okay, so I ran a 100 mile race (really 101 miles, but who is counting). The data on the length of time each mile took me is included in a .csv file uploaded on Brightspace. It took me about 30 hours and 47 minutes to complete. However, that 30 hours 47 minutes is my entire time on the race course, but not necessarily the time that I spent running. There were many aid stations throughout the course where I stopped to rest, eat food, have my feet worked on, etc. This added a nontrivial amount of noise

in the form of extra sitting time to some of the miles if I want to know how quickly I ran each mile.

I believe the shortest period of time I spent sitting at an aid station was 7 minutes. I want this data cleaned to modify the time for the miles that appear to have a 7 minute or longer rest break.

a) Use Python to import this .csv file, and compute the total time of the race (in hours) and the average mile time (in minutes). Have Python print out a message with these results.

b) Have your program go through each item in the mile time, and determines if the length of time is more than 7 minutes longer than the average of the adjacent mile times. If it is, replace the value with the average of the adjacent mile times. If it is not, leave the entry alone.

c) Print out a message containing the new total time (in hours) and the new average time.