Homework 2

The files hw1_1.txt and hw1_1_test_dat1.txt contain partial results from a model estimation at different iterations (it). In each iteration, multiple parameters are updated. The value of v tells which parameter is updated. An update involves multiple runs to iteratively recompute the estimated parameter value. The value of epoch tells which run the iterative computation produced the output. For all questions except first part of part a), consider only lines that contain 'U_beta**2 sum'.
Write a python code to do the following:

a). Read in the content of one file. Create a dictionary to store the following three objects as key:value pairs (Note: All keys of the dictionary should be string and all values are numerical. The data type for it, v, and epoch are integer. 'U_beta**2 sum' is float.):

- count the number of lines in the entire file and store it in variable n_lines
- count the number of unique values of the variable 'it' and store its value in n_it
- Create a dictionary with key 'it' and value being a dictionary, which is again a dictionary of Key:Val pairs where the Key gives the value of 'it' and the Val is the list of unique 'v' values under the same 'it' value.

b). Write a python code to extract it, v, epoch, and U_beta**2 sum from the file. Compute log10 of U_beta**2 sum. Let the code save the extracted data into a data frame with column names it, v, epoch, log10Ubeta2sum. (Note: Use np.log10( ) to get log10Ubeta2sum from 'U_beta**2' sum.)

c) Define a function to take three input arguments (it, v, and the entire data frame) and return the minimum log10Ubeta2sum. This function finds the smallest log10Ubeta2sum among all rows with the same it and v. The name of the function is findmin.