CS Homework #23: Clean word list

Deadline: 4/18/2020, 9:00 pm.
*Save your code as lastname_homework23.py and submit on **GOOGLE CLASSROOM***

Task 1
Repeat the first task from the previous homework. You will continue to work with the same text file.

Task 2
Once again, open the file in Python. Makes sure to implement proper error handling and use "with open" approach similar to what we did in class.

Task 3
Create a word list and PRINT all words that contain "happy" and "sad". (Notice that "contain" is not the same as "equal to").

Task 4
Now create a CLEAN word list. That is, make sure that you get rid of ALL punctuation from the text and all words are converted to the lower case format.

Task 5
Using the CLEAN word list, find the number of "sad" and "happy" words in your text.

Task 6 (required for CS 201 and CS 101B)
Using collections.Counter, find 50 most frequent words in the text.

Task 7 (required for CS 201; optional challenge for CS 101B)
Create a list of 20 most frequent words that are 5 or more characters long.

Task 8 (required for CS 201; optional challenge for CS 101B)
"Stopwords" are the words that are usually removed from text analysis (words such as "we", "for", "the", and so on). Use the "stopwords" list posted on Google Classroom to skip those words.

Create a new list of 20 most frequent words (any number of characters) which do NOT include the stopwords.