

Homework 2

(*Individual Homework)

MSAN 694, Diane Woodbridge

Due: Nov 2nd, 11:59 PM

Description

Given the *input_file* and *app_name* in “user_definition.py”, complete the python code “hw2.py” to:

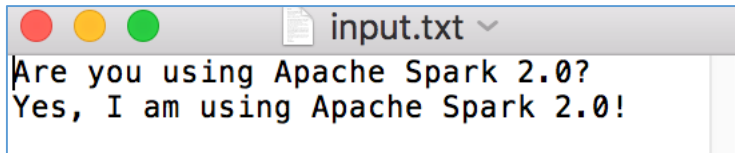
1. Print the distinct words with alphabetical characters in alphabetical order (lowercase).
 - Ignore/Remove special characters including : ; . , [] { } () “ ’ ? !
 - Assume a hyphenated word are 2 or more words. (**off-campus** → **off campus**)
 - Assume different forms of a same word (plural vs singular nouns, present vs past tense vs past participle verbs, etc.) are different. (**study**≠**studies**, **do**≠**did**)
2. Print the sum of all numerical values in the text as a float. Some things to keep in mind:
 - Numbers may or may not include commas after every third digit from right to left (e.g., **3,000** and **3000** are valid and equal). You may ignore numbers that mix comma and non-comma formatting (i.e., 3000,000 can be ignored since it’s not properly formatted).
 - Numbers may be floats and may or may not have a leading 0 (e.g., **.25** and **0.25** are valid and equal)
 - Don’t evaluate equations and remove operations (**3*3** → **3 3**)
3. Print an empty line between step 1 and 2 (example below).

Submit the hw2.py file (**ONLY**) - the name of your file should be **hw2_LastName_Firstname.py** on Canvas. Make sure it runs in **Python 2.7**.

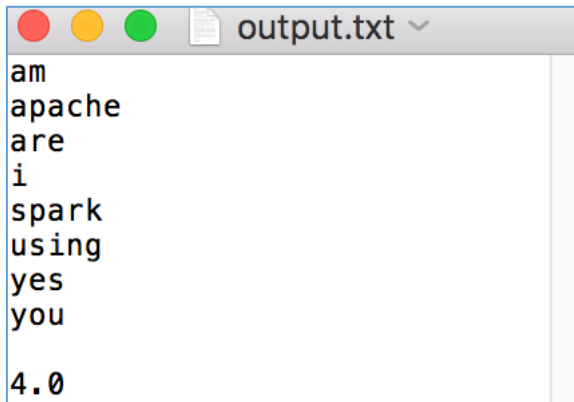
This assignment counts **6%** of your final grade.

We provide three example input files (input_1/input.txt, input_2/input.txt, input_3/input.txt) and corresponding output.txt.

If you run `spark-submit hw2_Woodbridge_Diane.py > output.txt`, the output should be:

A screenshot of a text editor window titled 'input.txt'. The window has a standard macOS-style title bar with red, yellow, and green window control buttons. The text inside the editor consists of two lines: 'Are you using Apache Spark 2.0?' followed by 'Yes, I am using Apache Spark 2.0!' on the next line.

```
input.txt
Are you using Apache Spark 2.0?
Yes, I am using Apache Spark 2.0!
```

A screenshot of a text editor window titled 'output.txt'. The window has a standard macOS-style title bar with red, yellow, and green window control buttons. The text inside the editor is split across eight lines, each containing a single word or number: 'am', 'apache', 'are', 'i', 'spark', 'using', 'yes', 'you', and '4.0' on the final line.

```
output.txt
am
apache
are
i
spark
using
yes
you
4.0
```