Homework 2
(*Individual Homework)

Description
Given the *input_file* and *app_name* in "user_definition.py", complete the python code "hw2.py" to:
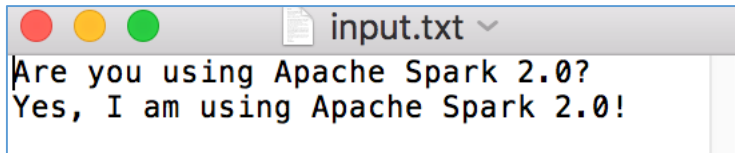
1. Print the distinct words with alphabetical characters in alphabetical order (lowercase).
    o   Ignore/Remove special characters including : ; . , [] {} () " ' ? !
    o   Assume a hyphenated word are 2 or more words. (**off-campus ➔ off campus**)
    o   Assume different forms of a same word (plural 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:
    o   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).
    o   Numbers may be floats and may or may not have a leading 0 (e.g., **.25** and **0.25** are valid and equal)
    o   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**.
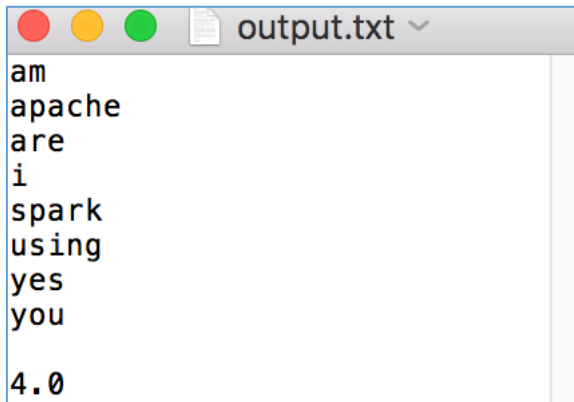
This assignment counts **6%** of your final grade.

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:

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

```
am
apache
are
i
spark
using
yes
you

4.0
```