INSY 5378: Homework 2

**DUE DATE** : Oct 5th, Wednesday 11:59 p.m.

This is the third Python assignment. There are 5 problems and each problem is 20 points and the total is 100 points.

Instructions:

1. **Late submissions will not be accepted.**
2. Submit your Python code (.py file, not the output .txt files) in Blackboard. DO NOT email your homework to me or TA.
3. [20 points] Create a file with name “hw2.py” and download “nasdaq.txt” , “raven.txt” , “frankenstein” files into the same directory where your “hw3.py” code is located.
4. [20 points] In “hw2.py” file, define a function make\_length\_wordcount() that:
   1. Takes input as the filename in the current directory (as a string)
   2. Returns a dictionary, which each key is *word length* and its value is the number of words with that length.
      1. Words should be translated to lowercase.
      2. For example, if the input file’s text is “Hello Python people Welcome to the world of Python”, then the dictionary should be:

{2: 2, 3: 1, 5: 2, 6: 3, 7: 1}

1. [20 points] In “hw2.py” file, define a function make\_word\_count() that:
   1. Takes input as the filename in the current directory (as a string)
   2. Returns a dictionary, which each key is a word and its value is the number of occurrences of that word.
      1. Words should be translated to lowercase.
      2. For example, if the input file’s text is “Hello Python people Welcome to the world of Python”, then the dictionary should be:

{'hello': 1, 'of': 1, 'people': 1, 'python': 2, 'the': 1, 'to': 1, 'welcome': 1, 'world': 1}

1. [20 points] In “hw2.py” file, define a function analyze\_text() that:
   1. Takes input as the filename in the current directory (as a string)
   2. Uses the two functions above – make\_length\_wordcount() and make\_word\_count() - to construct (i) length-wordcount dictionary and (ii) word count dictionary.
   3. Opens a new output file “FILE\_analyzed\_FIRST\_LAST.txt” and write two dictionaries into this file (in the format below).
   4. For example, if the input file is “test.txt”, the output file name is “test\_analyzed\_GENE\_LEE.txt” and it should contain the following lines:

Words of length 2 : 2

Words of length 3 : 1

Words of length 5 : 2

Words of length 6 : 3

Words of length 7 : 1

to : 1

of : 1

people : 1

the : 1

python : 2

welcome : 1

hello : 1

world : 1

1. [20 points] In “hw2.py” file, run the analyze\_text() function three times with the following inputs:
   1. “nasdaq.txt”
   2. “raven.txt”
   3. “frankenstein.txt”

Your hw2.py code should generate the following three files:

“nasdaq\_analyzed\_FIRST\_LAST.txt”

“raven\_analyzed\_FIRST\_LAST.txt” “frankenstein\_analyzed\_FIRST\_LAST.txt”

Please check test.txt and test\_analyzed\_GENE\_LEE.txt to check the expected results.

Again, please submit your Python code, not the output files.