AIT526 Individual Lab 4

Due Date: Please check the class schedule on blackboard.

Sentiment Analysis with VADER in NLTK

Tools (as shown in the class):

1) **Jupyter Lab** (Desktop or online) or Desktop **Jupyter Notebook** or any **Python IDEs**
2) **Python 3**
3) **NLTK VADER** (https://www.nltk.org/_modules/nltk/sentiment/vader.html)
   - Installation: pip install vaderSentiment

Coding Resources*:

1) Source of Internet

Note that you must include **reference(s) in the code comments when you refer others' work.*

Text Data Location: /blackboard/Assignments/Individual Labs/ Lab4/textdata.zip
Please unzip this file into four different text files to load the task.

Tasks (Extra Credit 4 points):

Please follow the **step-by-step instructions** with code examples, tutorials, and hints/code snippets to implement the following tasks:

1 **Sentiment Analysis with VADER in NLTK for Shakespeare's Poems**:

There are five Shakespeare's poems in five different text files: Shakespeare_*.txt. Please conduct sentiment analysis for each poem and rank the poem sentiments from positive to negative. **Note that** you can use *list*, *dictionary*, or *dataframe* or any data structure for this assignment.

1.1 (2.0 points) Calculate the overall sentiment score for each file and list the filename with corresponding scores.

***Hints**: 1) tokenize sentences, 2) use SentimentIntensityAnalyzer() to compute each sentence sentiment using compound, and 3) then use a certain statistic function to get an overall score. 4) write a function to get sentiment for each text. The function and results may be similar to:*

```
def GetSentiment(text):
    """
    Input: # a text with multiple sentences
    Usage: Sentiment Analysis with NLTK Vander
    Return: a score of overall sentiments
    """
```

The overall scores for all five text files may be similar to:

```
0.27688
0.4231
-0.4098538461538461
-0.1808625
0.1604
```

The filenames with corresponding scores may be similar to:

```
{'Shakespeare_Blow, Blow, Thou Winter Winda.txt': 0.27688,
 'Shakespeare_Sonnet 130.txt': 0.4231,
 "Shakespeare_Juliet's Soliloquy.txt": -0.4098538461538461,
 'Shakespeare_Fear No More.txt': -0.1808625,
 'Shakespeare_A Fair Song.txt': 0.1604}
```

1.2 (0.5 points) Rank the overall sentiment score for each file from the **positive** score to the **negative** score. Print all the filenames with corresponding scores.
***Hints**: Use sorted() to automatically rank. The outputs may be similar to:*

```
[('Shakespeare_Sonnet 130.txt', 0.4231),
 ('Shakespeare_Blow, Blow, Thou Winter Winda.txt', 0.27688),
 ('Shakespeare_A Fair Song.txt', 0.1604),
 ('Shakespeare_Fear No More.txt', -0.1808625),
 ("Shakespeare_Juliet's Soliloquy.txt", -0.4098538461538461)]
```

1.3 (0.5 points) Your program should be robust and able to dynamically process any number of files as inputs to generate sentiment values for each file and then automatically rank all of them.

1.4 (1.0 points) Briefly describe what you can find from the results.

- **1.4.1 (0.3 points)** Double check each overall score if they make sense for each poem based on your understanding for the poem.
- **1.4.2 (0.3 points)** Briefly explain the statistical methods you choose to get an overall sentiment score are feasible.
- **1.4.3 (0.4 points)** Briefly describe more possible extended applications based on this assignment.

You are strongly suggested to follow Python coding convention to write the code. The program should be robust and will be rerun for grading.

SUBMISSION

1. Write all your code and answers with explanation in the Notebook.
2. In the code file, please do not forget to write your name, course #, and date in the comments.
3. **Run ALL Cells**:

Open your IPython file in Jupyter, go to **Run**->**Run All Cells**. Please make sure all of you code has been run and print out the results.

4. **Save to HTML**:
   Go to **File**-> **Export Notebook As…**->**Export Notebook to HTML**, and save your work into HTML file.
5. **Submission**:
   a. Write your work with **four** file names **"AIT526_YourFullName_Lab4.ipynb"** and **"AIT526_YourFullName_ Lab4.HTML"**.
   b. **Zip** all files to **ONE zipped file** since blackboard does not allow you to submit HTML file separately.
   c. Go to the Blackboard **/Course Content/Optional Individual Labs/** to submit **ONE zipped file.**