

# Assignment 1 – Using NLTK and Vader Sentiment Analysis

In this assignment you will be using [NLTK](#), [PyPDF2](#), [py-readability-metrics](#) and [wordcloud](#) to compare two news stories on the same topic, from two different sources. Complete your analysis and commentary in a Jupyter notebook. The notebook should be pushed to your personal Github class repository and a URL to the notebook submitted for the assignment of Canvas

1. Install PyPDF2 and py-readability-metrics
2. In a new notebook, import nltk and run the following commands:
  - a. `nltk.download('punkt')`
  - b. `nltk.download('stopwords')`
  - c. `nltk.download('vader_lexicon')`
3. Read the two news stories found in this assignment on Canvas. In a markdown cell in your notebook, briefly characterize the issue at hand and your impressions from a news reporting objectivity perspective.
4. Use PyPDF2 to extract the text out of each document
  - a. Display the text
  - b. Briefly note any anomalies in your notebook
5. Use NLTK to complete sentiment analysis for each of the articles
  - a. Compare the results from the two articles
  - b. Compare the results to your reading of the articles; do you agree with the outcome?
6. Run sentiment analysis on the article headlines
  - a. Do those results agree with the overall sentiment?
7. Using py-readability-metrics complete 2-3 different readability scores
  - a. Discuss what the chosen metric is measuring
  - b. Compare the scores from both articles
8. Using wordcloud, create word clouds for both articles
  - a. Briefly comment on the word clouds