Python for Text Analysis Workshop Plan

4 sessions

Topics will include

* Intro
  + Reading text files
  + Cleaning—punctuation, stopwords, numbers
  + Tokenization—word, sentence
  + Basic word frequency analysis
  + Sentiment analysis with VADER!
* Bag-of-words analysis:
  + Part of speech tagging
  + NER
  + Tf-idf
* Unsupervised ML methods for text analysis
  + topic modeling
  + clustering with doc2vec
* Supervised ML methods for text analysis
  + Random Forest
  + If time—zero-shot with BERT?