Lab #2

- Due Apr 5 by 11:59pm
- Points 100
- Submitting a website url or a file upload
- Available Jan 25, 2024 at 12am - Apr 7 at 11:59pm

This assignment was locked Apr 7 at 11:59pm.

5018- Advanced Analytic Techniques

Lab #2 Assignment ~ Conduct Text Analysis

1. Collect some texts. Compare them in a number of ways.
2. You will likely want to have them be "bags of words." Prepare the text through removing upper case, white space, punctuation, and consider stemming the words, if appropriate for you purpose.
3. Generate relative word frequencies for each bag of words, and compare them to each other.
4. Articulate what differences (if any) you notice and whether this comports with a theory of why these bags of words should be similar or different.
5. Run statistical tests of association between the bags of words (correlation, cosine similarity, regression or Chi-squared), and explain what they indicate.
6. Do one more big thing-- either a sentiment analysis of the bags of words; rerun your analysis but using bigrams and/or trigrams; consider the role of negation words ("not," "no", etc.) on your earlier analysis; run a parts of speech tagger; look at the temporal unfolding of your words; or do a topic modelling exercise. For whichever thing you choose, explain what you are doing and whatever you find makes sense in some way theoretically.
7. Extra credit: do some wordclouds of your texts

Lab 2

Criteria	Ratings		Pts
			Total Points: 100