**Project Proposal:**

* Presentation link (will export to the google drive once i figure it out): <https://docs.google.com/presentation/d/1SmEMpsq19QhiZFhDzLf3RKm1eREx8pIVcP4h2rvnxmE/edit?usp=sharing>
* What the problem statement is: Use sentiment analysis of presidential speeches to predict future gas prices, approval ratings, labor force participation, and/or GDP growth.
  + Topics in speeches
  + Just US Presidents
  + Topic modeling
    - Trump vs. Bush, Clinton, etc. – would need to compare recent presidents b/c of advancements in society (energy, gas, etc.)
  + Similarity between president and speeches
  + How the topics have evolved over time (or look at the latest 4-5 presidents) – probably more interesting if start with washington
    - Has there ever been a president similar to Trump?
    - Name of the President and speech (2 column file)
* Data source(s): Kaggle and others
  + https://www.kaggle.com/datasets/littleotter/united-states-presidential-speeches
* Type of analytics to run on the data:
  + Sentiment analysis (could be interesting)
  + Similarity analysis - between presidents or pick Trump and
  + Lift analysis - see which presidents are most similar in terms of public speaking
* Expected outcomes:
  + Be able to input more recent speeches to try and predict outcome variables in the future.

To-Do List:

* **8 minutes per presentation (7-8 slides, says he will give us a template)**
* Scrape the archives for speeches Parthiv
  + <https://gist.github.com/kboghe/4a0aa1b656a1be0674d765ce0a0c3542>
* Data pre-processing - reference part C table in Assignment 2 Jason
* Perform similarity analysis between presidents and picking Trump (word2vec similarity) Anthony
  + Cosine similarity (secondary priority to word2vec)
* Perform lift analysis to see which presidents are most similar to Trump (find his top 4 attributes and see the lift between that and maybe 10 presidents, just like the beer example we just did) Nicole
* Topic modeling – cluster the presidents together Parthiv, Jason, Nicole
  + Example of MDS map code: <http://harrywang.github.io/document_clustering/> Kyle
    - Create MDS using only 7-8 presidents at a time, otherwise the MDS plot will be confusing Kyle
  + Topic modeling example code: <https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0>
  + Barua said he would post his topic modeling code. Email him if not posted by tonight