Presidential Sentiment Analysis
MSDS692 – Data Science Practicum 1

Progress Report for Week *5*

Jeremy Beard

Project Details
This project is centered around a dataset which contains speeches from the State of the Union of all presidents. The project will utilize a variety of natural language processing techniques in order to answer questions that have been created surrounding the dataset. Sentiment analysis will be utilized, general word commonality will be explored and word frequency will be analyzed. The final output will be a visualization comparing all the presidents to each other within the lens of the State of the Union.

Project Timeline:

Week 1 – Project definition and submit proposal (DONE)

Week 2 – Datasets selected and former related work collected (DONE)

Week 3 – Initial data loading and data cleaning performed (DONE)

Week 4 – Initial output and analysis for single speeches completed (DONE)

Week 5 – Output and analysis for all speeches / all presidents completed (DONE)

Week 6 – Result congregated, summary visualizations created, presentation began

Week 7 – Presentation completed, dry runs completed

Week 8 – Final Project Presentation

Planned Work for the Week:
From last week, my planned work for this week was to finish outputting all the charts and data relative to the questions I have come up with. This list of questions is a living list but right now the questions are as follows:

- *What is the quantitative positive/negative sentiment between all the speeches?*
- *Which presidents use the widest variety of words?*
- *What is the quantitative positive/negative sentiment between all the presidents?*
- *What are the themes or buzzwords among the different speeches / presidents?*
- *What were the most common words used in each speech?*
- *Which presidents gave the longest speeches?*
- *Which presidents gave the shortest speeches?*
- *Which presidents used the most unique words?*
- *Which presidents used the least unique words?*

This week I planned to finish the majority of the analysis that I planned to do related to the data. Finishing all/most analysis this week will set me up for success when it comes to starting the report/presentation and creating the story which was talked about this week during the course. I have no plans for including n-grams or lemmatization right now but I may also add those concepts in.

Progress for the Week:
This week I didn't make a ton of progress, but I did build out the script a bit more to create initial visualizations of the sentiment from each president as a whole, as well as year-to-year. I was a bit disappointed in the sentiment analysis results as the sentiment scores were generally low and didn't really display any trend. To create a story out of this, it may be a bit difficult. I may try to perform a different form of sentiment analysis later, maybe it will provide different results. As of now, I am mostly done with answering all the questions above. There are a couple stragglers I still have to answer but I don't expect this to take much effort. Looking good so far! Just have to create the story.

Roadblocks/Issues:
No roadblocks or major issues to report yet. I think I will email you/the professor this week about a review even though they are not mandatory. Just to make sure I'm on track and to get any other thoughts. I am most free on Fridays so I will probably be requesting this time Thursday evening or Friday sometime. I was a bit disappointed in my sentiment analysis scores like I mentioned previously, so maybe I will mention that. Lemmatization may help with the sentiment scores so maybe I will perform that this week to see if I can dial in the scores a bit more.

Plan for next Week:
Next week I plan to touch up the charts I have created to make them a bit more unique from each other and also stand out to the reader a bit more, using the knowledge gained from my former Data Visualization course. I plan to finalize these charts and begin some form of report/presentation. The story begins!

Resources for the Week:

State of the Union Corpus (1790 - 2018). (2018, October 19). Kaggle.

https://www.kaggle.com/datasets/rtatman/state-of-the-union-corpus-1989-2017