Presidential Sentiment Analysis
MSDS692 – Data Science Practicum 1

Progress Report for Week *2*

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 utilizing, 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*

> *Week 4 – Initial output and analysis for single speeches completed*

> *Week 5 – Output and analysis for all speeches / all presidents completed*

> *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:
This next week, I plan to begin loading in the data to a Python script and organizing it according to examples I've used in the past, collected this week. I want to begin to finalize the list of question I'd like to answer during this project. So far, the questions I have are:

- *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?*

Progress for the Week:
This week the dataset that's going to be used was finalized. Other president-related datasets were explored but the State of the Union Corpus dataset seems to be the most interesting and usable. This week projects from the past were gathered in order to act as references to the

current project. Other natural language processing assignments have been completed in the past and these will be referenced in order to complete aspects of this project.

Roadblocks/Issues:
No roadblocks yet! Next week we will enter the main exploratory phase and I may have roadblocks after that begins. Right now the initial project and status workflow is being determined so there are no major roadblocks to this type of phase.

Plan for next Week:
Next week I plan to have an initial script created which loads in all the data and begins organizing it according to questions or objectives listed above. Initial data cleaning and filtering for stopwords and more will be undergone. The output of next week will be initial data exploration and results for the data.

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