Case Study: Presidential Speech Modeling Rubric

Due: TBD
Submission format: Upload link to github repo to canvas
Individual Assignment

General Description: Submit to canvas a link to your case study repository and prepare a brief presentation regarding your process for class

Preparatory Assignments: All previous Data Science curriculum and practice

Why am I doing this? Case studies are an excellent way to check-in on the skills you've gained in your time as a data science student. We use case studies to practice being a data scientist, by starting and finishing projects with real-world applications. In this case study, you will learn about how you can use data science to analyze text-based data and create useful insights with political and practical implications. You will also have a chance at practicing a very commonly used but tricky modeling approach: sentiment analysis.

- **Course Learning Objective:** Analyze text-based data from a JSON file
- **Course Learning Objective:** Develop practical insights from data science
- **Course Learning Objective:** Prepare findings for presentation to your peers
- **Course Learning Objective:** Translate jargon for public presentation

What am I going to do? You will begin by reading through the materials in this folder, starting with the one-page introductory document that summarizes the prompt for this project. Make sure to thoroughly read through the attached third-party materials that will offer you more background both into the qualitative aspects of this case study and the intricacies of topic modeling with python. After getting a full understanding of the skills you will need to employ, produce a few items to be presented in a github repo and to your class. You will also need to create a brief, layman-friendly powerpoint presentation.

Tips for success:

- **Have fun!** This project is not meant to be stress inducing. If you're having trouble with it, there are an abundance of online resources at your disposal that may not be included in this case study. The data science community is your friend!
- **Keep an eye on the bigger picture.** It is easy to get bogged down in the world of model-making, but the best product will be one that keeps in mind that this is *not* a project about the numbers. It's about how you contextualize what the data science tells you!

How will I know I have Succeeded? You will meet expectations when you follow the criteria in the rubric below.

Spec Category	**Spec Details**
Formatting	● Repository: An online folder on github.com containing all the materials necessary for the project ● Contents ○ README.md ○ SRC folder ○ Figures folder ○ License
README.md	● Goal: An overview of the repository. What would somebody need to know before taking a look at all the materials you've included? ● Use plain, simple language. ● Sections: ○ Goal of the project ○ Explanation of the data (source, format) ■ Did you need to clean up the data in any way? ○ Model and analysis plan ■ Did you include any stop words? ○ Challenges ■ What obstacles did you face, if any? ○ Figures table of contents ■ If you've included figures, contextualize them here. ○ References ■ If you've used anything beyond what is included in this folder, cite them here using IEEE documentation style.
Figures folder	● Goal: Visualize the key aspects of the project. ● This is *optional*, but highly suggested. ● Include any exploratory plots made. ● If you've put your final results into a graph or chart, make sure to include these here in an image format.
SRC folder	● Goal: This folder contains all the source code for your project. ● Include all code files you produce. ○ You will not be graded on the percent accuracy of your model, nor will you be graded on whether or not you followed the exact methods used in the case study materials. ○ Rather, this will be graded on whether or not you produce a successful model using data science tools that extracts topics from the corpus of presidential speeches.
License	● Goal: This section will explain the terms of using your repository to people who may view it. ○ You may use the MIT license, which is the default!
Presentation	● Goal: Contextualize and explain your project to an audience. ● This presentation is *not* about the intricacies of the model you've produced. Rather, focus on the big picture process you followed to understand the model you created and to develop practical conclusions from your project!