Presidential Inauguration Speeches Case Study Rubric

DS 4002 - Karan Rawat
Due: May 12, 2025
Submission Format: Upload link to Github Repository on UVA Canvas

Individual Assignment

Why am I doing this? This case study aims to help you develop your data science skills by performing data analysis with real life data. It also aims for you to develop your skills with text analysis and scraping as both can be key aspects of data science. Specifically, you will learn how to scrape data from a website, perform sentiment analysis, model your data, and write about your findings. The deliverable will show your ability to go through all phases of a data science project.

What am I going to do? In this assignment you will find out if there is a relationship between the party of the presidential elect (Democratic or Republican) and the primary type of language used in their inaugural speech (polarizing or unifying). You will do this by first going to the github repository (link: https://github.com/KaranRawatUVA/cs3). Once there, you must read the hook pdf, and the readings in various materials. Once you have an understanding of the assignment, start working on scraping the data. There will be an example scraping in the file path: Various Materials/Example Code. Then once you have your speeches scraped, run a sentiment analysis on it. It is up to you to figure out how you want to do this. Either you must implement it yourself or find some method or API to do it. Research all options before committing to one. Then do some exploratory data analysis (EDA) on your data, so you can learn more about your data. This will also help you strengthen your story about the data. Next start making some models based on your data. Use Point Biserial Correlation to see the correlation between party and the primary type of language used in their inaugural speech (polarizing or unifying). Write down all your findings in a five page paper (including models). The paper should have an introduction, a section focused on how you got your data, your EDAs, your findings and models, and lastly a conclusion.

Deliverables include:
- A five page paper on your case study
- A data dictionary
- Documented code
- Github repository containing all of the code and data you used

Tips for success:
- Read all of the papers and code provided to you! They will give you context on what this project is all about. The provided code especially will save you time in trying to get the project done in a timely manner.
- Research parts of the project that you do not fully understand. This will make the product better as you will actually know what you are doing.
- Don't presume what the data will show you. Look at the data to give you insights, not the other way around.

How will I know I have Succeeded?

You will meet expectations on this case study when you follow the criteria in the rubric below.

Spec Category	Spec Details

Formatting	Submit a link to a Github repo with the following items:
	● README.md
	    ○ This will give an outline of what you did in your project and how to run the Github
	● LICENSE.md
	● Code
	    ○ Have all your Python code saved in a folder called "Scripts"
	● Data and Data Dictionary
	    ○ Have your data and data dictionary saved in folder called "Data"
	● Five Page Paper

Code	Your Python code should meet the following guidelines:
	● Make sure that the code is well documented with a lot of comments explaining why you did it that way
	● Make sure that variables names make sense and that your code is organized in an easy to follow way
	● Cite where you got help from outside resources
	● They should include:
	    ○ Data scraping and Sentiment Analysis
	    ○ Exploratory Data Analysis
	        ■ Graphs to show insights before the modeling
	    ○ Point Biserial Correlation and other Modeling (include graphs)

Data Dictionary	Your data dictionary must be in a CSV format and meet the following guidelines: ● Must explain each column of your data in a way that anyone can pickup and understand it

Paper	The paper must be five pages in a PDF format and have the following parts:
	● Introduction
	    ○ This part of the paper must introduce the case study's main questions and give background as to why it is important
	● Data Explanation
	    ○ Here you must explain how you got your data and how you performed sentiment analysis on the speeches
	● Exploratory Data Analysis
	    ○ Present your findings about the data from the EDAs
	    ○ Include multiple graphs
	● Findings
	    ○ Explain the findings you got from your models and what it means with multiple visuals
	    ○ Answer the case study's questions with your findings
	● Conclusion
	    ○ Wrap up the paper by restating why and what you have done and how it relates to answering the case study's questions
	● Sources
	    ○ Make sure to include any sources you may have used in MLA format