Modeling the Topics of U.S. Presidential Speeches

A UVA Data Science Case Study by Graham Twente, Spring 2023

Prompt: Welcome! The Executive Office of the President thanks you for applying for the role of Data Analyst under the White House Director of Speechwriting. Your first task for this prestigious position is to analyze the major topics addressed in U.S. presidential speeches over time. To do this, you will implement topic modeling, a form of natural language processing, to identify the most important themes addressed by past presidents. Your analysis will provide historical grounds to shape the President's upcoming State of the Union Address. This task will give you the opportunity to demonstrate your skills in data analysis, critical thinking, and communication of results, as you work with our nation's highest office. Good luck!

Deliverable: Produce a methodology to identify and describe the top topics contained in presidential speeches throughout U.S. history. You have been given a dataset of presidential speeches (1796-2023) compiled by the University of Virginia's Miller Center, a leading nonpartisan institute on presidential scholarship and political history. Your work will be assessed by the depth of thought and curiosity with which you design your methodology, conduct analysis, and communicate conclusions.