Unifying or Dividing? Analyzing Presidential Inaugural Speeches with Text Data

Scenario:
You are a data scientist tasked with analyzing whether U.S. presidential inaugural addresses contribute more to national unity or division - and whether political party plays a role in shaping this tone. As part of a civic data journalism initiative, you must scrape, clean, and analyze presidential speech data to uncover trends in unifying versus polarizing language. Your client wants clear, reproducible results and expects a professional data product.

Topic/Context/Motivation:
Recent years have heightened public concern about political polarization. Yet political rhetoric has long played a role in shaping national identity. This case study invites you to engage with historical and modern speeches alike, use basic Natural Language Processing (NLP) techniques, and answer the question: **Are Democratic and Republican presidents more likely to use unifying or polarizing language?**

You will build your skills in data scraping, text feature engineering, descriptive analysis, and basic modeling.

What You Will Produce:
You will submit a Jupyter Notebook that:

- Scrapes inaugural addresses from the web,
- Cleans and processes text data,
- Creates quantitative features (counts and ratios),
- Visualizes key patterns,
- Conducts basic statistical or predictive analysis,
- Reflects critically on bias and uncertainty.

Deliverables and technical expectations are fully detailed in the attached Rubric.

GitHub Repository: https://github.com/emilymacris/DS4002-CS3