

U.S. Election Case Study Rubric

DS 4002 - Fall 2024 - Instructor: Britney Hoang

Due: December 16, 2024

Submission format: Upload link to Github repository on Canvas.

Individual Assignment

Why am I doing this?

The goal of this project is to explore text-data and improve your understanding of a data science project framework. You will learn how to use sentiment analysis and other techniques to analyze text-data.

- Course Learning Objective: Employ previous data science knowledge to apply a new technique on text-data
- Course Learning Objective: Analyze and interpret results and draw meaningful findings.

General description: Submit a link to your case study repository to Canvas.

What am I going to do?

You will utilize the sentiments in tweets about Joe Biden and Donald Trump during the 2020 presidential election to see if there are any notable patterns in positive or negative emotions towards each candidate and identify any keywords associated with them. In the end, you will prepare a deliverable that hits all the requirements. Through analyzing whether tweets about Joe Biden and Donald Trump were predominantly positive or negative, and identifying the major topics of discussion, one can potentially reveal how public sentiment was formed and how it might impact future elections. Thus, this project highlights the importance of analyzing social media data to better understand public opinion trends.

Final deliverables should include:

- Github Repository link - including data, data dictionary, cleaning, and code

Tips for success:

- Remember the basics. Do not overcomplicate the steps or project.
- Stay organized and make sure to go through each step carefully. It is important to make sure to properly clean the dataset and import all packages correctly.
- Do not be afraid to ask for help. Your instructor is here to help and answer any questions.

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	<ul style="list-style-type: none">• Repository

	<ul style="list-style-type: none"> ○ Submit a GitHub repository including folders for: <ul style="list-style-type: none"> ■ README.md ■ LICENSE <ul style="list-style-type: none"> ● Use MIT as default ■ SCRIPTS ■ OUTPUT ■ REFERENCES ○ Use pdf format when possible
README.md	<ul style="list-style-type: none"> ● Include a brief overview of your results/findings for the case study <ul style="list-style-type: none"> ○ Quick summary but still allow readers to understand the goal and main idea of the project
SCRIPTS folder	<ul style="list-style-type: none"> ● Include all the source code for the project <ul style="list-style-type: none"> ○ Execute sentiment analysis and frequent word analysis ○ Code should be well-documented on how the sentiment analysis was performed and TF-IDF was applied ○ There should be many comments throughout code explaining the purpose of each command or section does ○ Perform statistical significance test
OUTPUT folder	<ul style="list-style-type: none"> ● Upload any information, graphs, figures, etc. <ul style="list-style-type: none"> ○ Anything generated should be included in this section ● Use informative names for all of the files.
REFERENCES.md	<ul style="list-style-type: none"> ● References should be included in IEEE citation style

Acknowledgements: Special thanks to Professor Alonzi from UVA DS 4002 for the rubric structure.