Presidential Approval Rating – Case Study

DS 4002 – Spring 2024 - Instructors: Javier Rasero; TA: Mercedes Mora-Figueroa

General Description: This rubric outlines the goals of the case student project and what is needed to succeed on the assignment of replicating a case student concerning measuring presidential approval rating via natural language processing from a previous DS4002 student.

Preparatory Assignments/Knowledge – Knowledge of Python, GitHub

Why am I doing this? This is your opportunity to highlight knowledge of implementing data analysis in python. In addition, it is an opportunity for you to gain new experiences with unknown data or analysis tools. It also gives experience replicating projects and validating results, a very important skill for testing novel research.

- Course Learning Objective: Reimplementing data analysis plans

What am I going to do? Access the GitHub repository https://github.com/ArjunaBazaz/CS3_2020_Presidential_Approval_Rating_Analysis.git from a previous DS4002 student and replicate the code in the SCRIPTS folder. Follow the instructions in the reproduction instructions section of the README in the GitHub and then follow the instructions in the rubric below. After learning about the project, you are going to report your conclusions.

Tips for success:

- Read all of the instructions before beginning
- Make sure to research the packages that is being used
- Review the DATA folder and the data appendix before beginning to familiarize yourself with the data
- Make sure directory information is clear and code is changed to match that

How will I know I have Succeeded? You will meet expectations on if you successfully follow the steps, reimplement the project, and detail your findings.

Data Acquisition	• Goal: download all of the data • Follow reproduction instructions if you want to download data directly from the source • go to the SCRIPTS folder and download all of the data to your computer
Data Analysis	• Goal: Conduct Basic analysis of the data • Run through 1_data_cleaning.ipynb to create one dataset with the entire csv file of both candidates around the election date

	• Open up the SCRIPTS folder and run through the file 2_initial_data_exploration.ipynb and run the file, pay close attention to how the matplotlib files are constructed and how they are implemented • Gain initial understanding on the data from the data appendix • Report findings in conclusion, decide what confounding variables could arise in the conclusion based on what we see from the data here
Results Gathering	• <u>Goal:</u> Conduct natural language processing • Open the csv file debate_tweets_cleaned.csv • Understand how the Vader package works in Python • Implement the natural language processing to score the sentiment of every tweet and save the results in a csv file called debate_tweets_with_sentiment.csv. In this case, positive values mean a positive sentiment and negative values mean a negative sentiment
Statistical Analysis	• <u>Goal:</u> Conduct statistical analysis • Gather all of the data around the election into separate lists of sentiment scores. One for Biden before the debate, one for Trump before the debate, one for Biden after the debate, and one for Trump after the debate. • Conduct a two sample T-test on the results to see where there was significance in the distribution of sentiments. • Plot the results to see the changes in sentiment for each candidate before and after the debate.

Conclusions	• <u>Goal:</u> Summarize the results
	• Write a summary of the results you found and the impact of it
	• Questions to consider:
	  ○ Which candidate had a better debate according to Twitter?
	  ○ Did combined sentiment become more positive or more negative following the debate?
	  ○ Did this match polling results at the time?
	  ○ How can candidates use this information to use twitter in their campaigns?

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled from Streifer & Palmer (2020).