

CS2 Rubric – Gerrymandering Case Study

DS 4002 - Fall 2023 – Jillian Barta

Due: December 2023

Submission format: GitHub repository (submitted by link to canvas)

Individual Assignment

General Description: Submit to collab assignments a link to your GitHub repository and a summary of results in PDF format.

Preparatory Assignments

- Hook document and attached supplemental materials.

Why am I doing this? The goal of this assignment is to allow you to walk through the steps of the data science lifecycle to practice doing the work of real data science! Although the scenario and objective of this case study are already provided to you, you will still get to apply your creativity to the rest of the project. Through this process, you will gain the ability to understand the problem at hand, clean the data, explore the data, create a model(s), and practice data visualization. One of the most important aspects of data science is the ability to communicate your results effectively. After all, your project will not be able to help anyone if you are unable to explain your findings to both technical and non-technical stakeholders!

- Course Learning Objective: translate a real-world scenario into a functioning model.
- Course Learning Objective: effectively communicate your findings to a general audience.

What am I going to do? You are tasked with building a model that can detect and predict the severity of gerrymandering in the 50 states. You have been given the link to a website that provides the district boundaries of each state along with the 2022 congressional partisan lean within each district. Use these in any way you must to confirm or deny the presence of gerrymandering in each state. Then, select the most gerrymandered and the least gerrymandered state. Finally, explain/summarize your model and results in a short write-up. All code and materials should be submitted to a single GitHub repository.

Tips for success:

- Take your time and analyze your options – there are many possibilities to achieve a successful model.
- Keep yourself organized – follow the data science pipeline and document along the way.
- Stay focused and do not give up – talk to your professor or TA if you need help!
- Have fun and get creative!

How will I know I have Succeeded? You will meet expectations on CS2 when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none"> One Github Repository (submitted via link on collab) containing: <ul style="list-style-type: none"> A README.md file (which auto displays) A LICENSE.md file (use MIT as default) A SRC folder A DATA folder A FIGURES folder A SUMMARY folder
GitHub Repository	<ul style="list-style-type: none"> Goal: This repository serves to organize all your code, data, materials, and findings into one organized and digestible platform for others to view. Use markdown headers to divide content. README.md <ul style="list-style-type: none"> Make an H2 (##) section explaining the contents of the repository. SRC section – Installing/Building your code and usage of your code. DATA section – Data dictionary, data files (or link). FIGURES section – All figures produced with summary of takeaways. REFERENCES section – List all references at end of README.md using IEEE citation style. Licenses.md file – use MIT licensing. SRC folder – contains annotated markdown file. Data folder – contains all data sets including cleaned and uncleaned versions. Figures folder – contains all figures produced in your project with takeaways. Summary folder – see below.
Summary Folder	<ul style="list-style-type: none"> Goal: This folder serves to hold a summarized version of your model and results that effectively communicates to non-technical individuals. <ul style="list-style-type: none"> PDF file Submit to your GitHub repository and separately to the canvas assignment. A written explanation of your model and an analysis of the results (1-2 paragraphs). List which states did and did not have the presence of gerrymandering. Specify the most gerrymandered state and least gerrymandered state.

Acknowledgements: Special thanks to Jess Taggart, Professor Alonzi and Harsh Anand. This structure is pulled direction from Streifer & Palmer (2020).