Computational Social Science 1A: Human Psychology and Social Technologies

Masters in Computational Social Science
University of California, Berkeley
wdtmacss@berkeley.edu

Assignment Three: Class Project

Did Twitter's Deplatforming intervention work?

Assignment Instructions

Write a computational essay that provides a data-driven answer to the question: *Did the Post-January 6th Deplatforming Intervention reduce misinformation sharing on Twitter?*

Your primary source of data is the dataset accompanying the McCabe et al. study we have examined in class.

Your computational essay will:

- ☐ Explore the structure of the dataset (Section 2: Data exploration)
- ☐ Perform a Difference in Differences (DiD) analysis of the McCabe dataset that conceptually replicates the DiD used in the study to support their primary conclusions (Section 3: Replication).
- ☐ Perform additional DiD or other analyses that examine key assumptions and study decisions or otherwise offer insight into the robustness of the conclusions reached on the basis of DiD analysis applied to this dataset (Section 4: Extensions).
- ☐ Conclude with a short written answer to the question, clearly stating your conclusions and general reflections in relation to the strengths and limitations of the methods we have applied and the dataset we have examined (Section 5: Conclusions).

I will provide a template notebook that you will extend. The notebook includes a shared version of the dataset.

Assignment Guidance

- General Targets (more detail in the *Grading* section)
    - A fair grade (e.g. C+) will be assigned to essays that evidence a clear attempt at appropriate analysis but are incomplete or slightly inaccurate (e.g the code doesn't run or the regression model is mis-specified).

    - A good grade (e.g. B+) will be assigned to essays that evidence a fully functioning and appropriate DiD analysis with appropriate interpretation (a successful *conceptual replication* of the McCabe analysis).
    - An excellent grade (e.g. A) will be assigned to essays that go beyond a single conceptual replication to include creative and exploratory additional DiD or other analyses on different subsets of the data which are well motivated by relevant questions.

- Your notebook should include **code, analysis, results, and data visuzaliations**.
- It can also include **simple narrative commentary** explaining the analyses you are undertaking and the motivation. The class notebooks can be used as an example.
- Strong essays will include consideration of the **limitations of the dataset and of the difference in differences analysis method**, either verbally or using data.
- The class project is supposed to offer an opportunity to write a simple computational essay that you enjoy and feel proud of. Feel free to be creative and exploratory.

Grading

Instructions to graders:

- Use the grading rubric below to assign each essay to one of the score range categories
- Generally an essay should satisfy criteria **for all three columns** (product, process, interpretation) to be assigned that category.
- In cases where an essay doesn't meet the criteria for a score range category in one of more columns (e.g. the essay has good process and product but only fair interpretation), choose the lower scoring category, but assign higher scores within that range if appropriate.
- If an essay has very mixed score range categories for the different components, you can use your judgment to assign the appropriate score range category. For example, an essay that has Excellent process, Good product, but Fair interpretation might be appropriate for the Good score range category, but would not be appropriate for the Excellent score range category.
- The goal here is to use the grading rubric as guidelines that support your overall assessment of whether a student demonstrates competence in the programming and data analysis skills required for this assignment (DiD analysis expressed as linear regressions, implemented and applied to the McCabe dataset using Python).
    - An essay demonstrating this competence should be at or above the Good score range category.
    - Only essays that go beyond evidencing competence and in addition evidence more in-depth analyses or advanced methods and clear conceptual understanding should be assigned to the Excellent category

Score Range	**Letter Grade Range**	**Product** (code, results, data visualizations)	**Process** (use of coding assistants and conversational AI systems)	**Interpretation**
93-100	A+, A (Excellent)	☐ Creative use of in-depth or advanced analysis or visualization techniques ☐ Uses data or visualization to criticize the DiD method or the dataset effectively	☐ Attests no use of external resources ☐ Clear independence of external resources for multiple complex tasks	☐ Thoughtful reflection on methods, data, and results ☐ Clear integration of data and narrative
84-92	A-, B+, B (Good)	☐ Extends DiD analysis to address one or more of key assumptions or study design choices ☐ Successfully applies DiD analysis to McCabe et al dataset – conceptual replication	☐ Some reliance on external resources for complex tasks ☐ Clear independence of external resources for basic tasks	☐ Some reflection on methods, data, and results ☐ Some integration of data and narrative
74-83	B-, C+, C (Fair)	☐ Attempts to apply DiD analysis to McCabe et al dataset ☐ Code incomplete or doesn't run ☐ Uses incorrect variables or model structure ☐ No visualizations	☐ Clear dependence on external resources for complex tasks ☐ Some reliance on external resources for basic tasks	☐ Answers questions
Below 74	Below C	☐ Clear difficulty implementing analysis	☐ Clear dependence on external resources for basic tasks (e.g. reading a csv, inspecting data, generating summary statistics)	☐ Little or no written answers or conclusions included.
Grader Comments				