Reminder to all: **Set small attainable goals for yourself!**

**PART 1: This is for the reviewer to complete**

Reviewer Name: **\_\_\_Stephanie Schefer\_\_\_\_\_\_\_\_\_**

Who are you reviewing: **\_\_\_\_\_Shradha Dinesh\_\_\_\_\_\_**

1) Provide a brief (~1-2 paragraph) description of the goal of their project:

Shradha's project aims to replicate and expand upon a 2019 analysis that discovered a significant number of bills introduced by state lawmakers were directly copied from model legislation proposed by industry groups, think tanks, and corporations. To achieve this, the project will leverage the LegiScan API to obtain up-to-date legislative data from various statehouses and the U.S. Congress. Additionally, model bill data from the American Legislative Exchange Council (ALEC) will be collected. The next step involves standardizing the diverse data sources, considering the variations in file formats such as PDFs, Word files, and web pages. Once the data is prepared, the project will focus on analyzing text similarity among a subset of bills, identifying common topics and keywords. Finally, the project will visualize the semantic similarity of these bills, potentially in an interactive manner, to provide valuable insights into legislative trends and the influence of model legislation across different policy areas. Through these steps, Shradha's project aims to shed light on the prevalence and impact of model bills on state-level policymaking, utilizing updated data and advanced analysis techniques in a more streamlined manner.

2) Is it clear how you would obtain the data needed for their project? I.e., if the data is collected from an online source, is there code and/or clear instructions on how to obtain it? If the data is simulated, is it clear how you would run the simulation? Etc.

The data is obtained from ALEC (American Legislative Exchange Council). When investigating the site, one can find legislative bills by filtering searches by year, issues, types, policy status, and tags. Shradha’s project aims to reproduce results for 2019 in which there were over 1 million bills. While the data is downloaded in a .csv file, Shradha has written python code to store the data in a SQL lite database. She has a big jumpstart on the project by creating a database creation file, a fetch data file, a fill database file, and a SQL queries file.

The database creation file’s main purpose seems to be to convert a .csv file into a single table database with a tBills table. Here the variables include, the bill id, number, title, description, state, session, filename, status, date, url, error, and content. The .csv file is obtained through the fetch\_data.py file. It uses an LEGISCAN\_API\_KEY to interact with the ALEC page. The .csv that gets created includes the same variables that are stored in the database, except it does not include the error and content. These seem to be variables that Shradha creates for analysis of the data. After the database structure is initialized, and the data .csv file is obtained, the fill database file can be run. This file populates the database with the given data.

3) Do you have any suggestions or questions for the person you are reviewing about what their final product will look like? If not, then please describe how their current plans are in alignment with what you would want to see.

The structure of Shradha’s GitHub is clear from the file names. I would suggest she adds some description about the files and their purpose within the readme.md file as well. I am assuming this is already a part of her plan for future iterations, so it is understandable that it has not been done yet. This may include a timeline about which files were created first, second, etc. and the order in which one would run and replicate her work.

As for the overall goal of the project, I understand the project is to create a smaller scale dataset resulting from the overwhelming amount of bills currently listed, but I am a little confused on how the bills will be consolidated. I am under the impression that the bills that are copied off model legislation will all be classified as the same bill? But what about if they are model legislation that relates to different topics? I think my misunderstanding is regarding legislation itself. Maybe defining model legislation or an example would help. Is the model legislation something that lawmakers can write from the same model template and apply it to various topics, such as the same template being used to produce a tax bill vs a voting policy. These would be two different topics so combining them into one subset of bill does not seem to make sense. However, if the model of bill is just adopted by various towns/states/legislative districts, and it truly is the same topic and speech, then that does in fact make sense to combine for a simplified subset of the data.

You have done a lot of work on the project so far! It looks great and I am excited to see what the final result looks like! Do you have a target dataset size for the final result? Is it just reducing the 10k copies of bills, or is your goal to further reduce the dataset results?

**PART 2: This is for the person being reviewed to complete**

1) Has your reviewer accurately described your project?

2) Please summarize, in your own words, the suggestions your reviewer has provided. If they have suggested something that is beyond the scope/intention of your project, have you communicated this to them? Otherwise, please indicate which suggestions you plan to incorporate into your project.