The proposal document should address the following points. Use these points as headers in your document.

* **Basic Info.** The project title, your names, e-mail addresses, UIDs, a link to the project repository.

**Wrangling Social Determinants of Health**

**Group Info**

|  |  |  |  |
| --- | --- | --- | --- |
| Chase Maughan | Kalpana Simhadri  [u1473696@umail.utah.edu](mailto:u1473696@umail.utah.edu)  u1473696 | Rachel Buchleiter  [U1414409@umail.utah.edu](mailto:U1414409@umail.utah.edu)  U1414409 | Xincen Xi |

Project repository: <https://github.com/r-buch/social-covid>

**Background and Motivation.**

Social Determinants of Health (SDOH) are recognized as important factors influencing health outcomes of the population. SDOH is primarily measured in five domains as defined by the CDC. It can be challenging to measure these insecurities directly. While there are assessments that can be used when patients have a medical encounter, this is not standardized nationwide and are inconsistently used. Since this priority area has high impact but can be elusive to identify, more effort is needed to understand related data. One proxy highly correlated with these insecurities is census tract data. It is a priority for national health organizations and funding. One organization prioritizing research that may use these types of data sets is the National Institute on Minority Health and Health Disparities (NIMHD). The flexibility and numerous research questions that could leverage this data set, or similar social determinants of health sources, made it an appealing learning opportunity for the group.

**Project Objectives.** Provide the primary questions you are trying to answer with your data. What would you like to learn and accomplish? List the benefits of how the data could be useful.

* Identify publicly available data can be used in analyses focused on social determinants of health
* Clean publicly available data associated with domains of social determinants of health
* Identify and merge relevant open data into a data set suitable for analysis identifying social determinants of health. Using the FACETS project as an example.

Reference

FACETS: using open data to measure community social determinants of health. doi: 10.1093/jamia/ocx117

* **Data.** From where and how are you collecting your data? If appropriate, provide a link to your data sources.

**Data Processing.** Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

Some key points that can be considered for data processing are:

* **Data cleanup**: identifying and handling missing data, standardizing formats, dealing with the outliers, checking the data consistency.
* **Data transformation**: includes normalizing data where necessary to allow comparison across different variables and creating indices by combining multiple indicators into composite indices, if needed.
* **Data integration**: merging data with other publicly available datasets if necessary, ensuring proper alignment of different geographic regions and time periods. Also ensuring that the variables from different sources are harmonized to be directly comparable, involving adjustments to variable scales, definitions, or methodologies
* **Develop scripts**: using Python that automate as much data cleaning and transformation process as possible.
* **Version control**: Use version control for the code and data processing steps to track changes and ensure reproducibility
* **Documentation**: thoroughly document all data cleaning and processing steps, including decisions made and their rationale.

**Design.** How will you display your data? Provide some general ideas that you have for the design. Develop **one alternative prototype design for your data**.. Describe your designs and justify your choices of visual encodings.

**Must-Have Features.** List the features without which you would consider your project to be a failure.

**Optional Features.** List the features which you consider to be nice to have, but not critical.

**Project Schedule.** Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

Week 1:

Week 2:

This proposal is the first part of your process book. As a ballpark number: your proposal should contain about 3-4 pages of text.

You will schedule a project review meeting with a staff member during regular lecture times of the week marked in the schedule. Make sure all of your team members are present at the meeting.

The proposal will be submitted by uploading it to your team’s GitHub repository.