1. **Please provide a one paragraph description of the goals of your project. You can list the same description from the previous deliverable or provide new details about aspects that have changed since Week 4.**

We plan to build a heat map that highlights regions (district/ward level) in Chicago that shows the relationship between social investment and crime outcomes. The explanatory variables would be proxies for social investment spending in a region (referenced in following table). We hope to run a multivariate regression analysis and use the magnitude of the error term to display intensity of the degree by which predictions are off from the actual observed crime levels in the heatmap at ward-level, given the level of a specific ward's social investment.

In a perfect scenario, the errors will be close to zero, meaning the relationship between investment and crime would be homogeneous along the wards so policymakers are directing more funding to places with more "disadvantages", to equate opportunities. Otherwise, the "big positive" errors could give signals for a policymaker because those wards will have more crime than social investment compared to the average.

To add an interactive element to the application, we will allow users to build visualizations between pairs of variables from the 4 datasets so that they can see patterns to explain the result of regression analysis/heat map.

For each source of data that you expect to use, please list the source of data, who will be responsible for collecting data from that source, and a date by which you expect the work of gathering the data from that source to be complete.

Data	**Source**	**Responsible**	**Expected date of collection**	**Method of extraction**
Chicago daily crimes report - 2020	**Chicago Portal**	**Julian Varon**	**02/28/2022**	**API**
Family and Support Services Delegate Agencies	**Chicago Portal**	**Julian Varon**	**02/28/2022**	**API**
Small Business and Improvement Fund	**Chicago Portal**	**Julian Varon**	**02/28/2022**	**API**
Chicago Microlending Institute Microloans	**Chicago Portal**	**Julian Varon**	**02/28/2022**	**API**
Chicago wards shape file	**Chicago Portal**	**Pedro Ramonetti**	**Done**	**Downloaded zip file**

2. **Please give a brief sketch of the work that needs to be done to complete your project (other than data collection), include a description of which team member(s) will be responsible for completing this work and the expected timeline for completion**

Task	Description	Responsible	Expected date of completion
Data, gathering, merging and cleaning	**In charge of downloading the several datasets, and merging them using "wards" as keys.**	**Pedro Ramonetti & Julian Varon**	**04/03/2022**
Regression Analysis	**Create a linear regression module that takes the crime outcomes, as a dependent variable, and the explanatory variables related to social investment.** **From this model, the security dependent variable for each ward would be predicted, and the estimated error will help us to see the position of the security-investment relationship of each ward in comparison with the average ratio. The errors will be displayed on heatmap (high error will be shaded darker than low error for each ward).**	**Diego Martinez**	**04/03/2022**
Visualizations	**Exploratory analysis by doing cross-tabs of different variables to see which variables can be made available to the user to select from a dropdown (across all datasets) to explore meaningful relations to add context to the regression results**	**Brinda Sapra & Julian Varon**	**05/03/2022**
GIS Implementation of heat map	**Use shapefiles to draw the map boundaries for Chicago and use**	**Brinda Sapra & Diego Martinez & Pedro Ramonetti**	**07/03/2022**
Application	**Connecting all previous work to the user command-line interface**	**Pedro Ramonetti**	**16/03/2022**

3. **Include any additional information you wish to provide about your project.**

   None.

4. **If you need additional feedback from me then please make sure to include that in your report.**

   We are planning to use Geopandas for creating the heat map but we have not previously worked with shapefiles/GIS data. Is there any code snippet you can share with us to see how shapefiles are used with Geopandas?