Meeting Notes from 2/25/2021:

- Primary final outputs:
  - **Clean dataset** - easily readable and organized
  - A reproducible, with as much automation as possible, **process**
    - So that as the input datasets change, we can run the changes in the dataset through the process, so we can keep the dataset updated
  - A stretch: how to share this dataset/process → interactive mapping tool database **website**
    - So people can explore number of housing units and residential density in any neighborhood
- Python to R challenge
  - R package: reticulate
    - Allows you to download and use python packages within R
  - https://rstudio.github.io/reticulate/
- Team work structure/flow schedule
  - Github project board - basic project track records
    - Checklist
  - Working together is suggested
  - Version control
    - Creating one shared repository of the forked spark repository
  - Data processing
    - Documentation = important
    - Jupetyer notebook - markdown
- Importing datasets
  - GitLFS - synchronize datasets and data
  - Running google drive with the most updated version of data
  - Jupyter notebooks in google colab - google docs for coding ***
- Pull requests
  - Add team name and project name
  - Basic description
- Important to have a common key (LOC_ID)
- Geodatabase (read by rqis)
- Its okay for the final product can be in pandas and they can do whatever with it after
- Once you export as JSON, mongo can take a JSON data dump as input