

IST652 Project Deliverable 2

Phase 2: Project Progress

In this step you should have a road map of the steps you will be taking to complete your analysis. In addition, at this stage you should also complete the following

- fine tune your research questions.
- upload your dataset into jupyterhub and conduct some preliminary cleaning and transformation.
- provide coding activities conducted so far.
- have a better sense of team members responsibilities.
- set a schedule to meet

Step 1: What is Your Idea and Research Questions, Again?

Please reiterate your project idea below (you can copy it from the project proposal if there were no changes).

--== Double-click and put the title team members and brief description of your project below ==--

About the research questions on Code violations dataset

This dataset records every housing code violation in Syracuse, as well as where it occurred, when it occurred, what the violation was, as well as the name and location of who is responsible for the violation.

Potential questions could be:

1. Is there any correlation between neighborhood/location and type of violation?
2. Are the building owners generally local? Are any of them repeat offenders?
3. How did COVID impact the number of housing violations (if at all)?
4. Have the kinds of violations given out changed over time?
5. Will we see a relationship between demographic and the types/amount of violations?

Step 2: Problem Analysis - Roadmap

What are the preliminary major steps you will be completing? Include the research question and steps taken to answer that question? Are there any unique functions you will be incorporating which we have not covered in the classroom? Describe below.

--== Double-click and describe steps towards analysis of your project below ==--

Step 1. Done: We have read the dataset and did data cleaning and transformation on several columns. This step included converting column types (string to datetime), checking date range and filtering bad records (date recorded as back to year 1900), merging our data with API-sourced census data, and doing some string cleaning in the violation column (a shorter version of violation name is created excluding redundant information), all of which we covered in class.

Step 2. Now: We are working on some descriptive analysis regarding several attributes in the data table using seaborn and also by grouping the data.

Step 3. Next: We will solve the potential questions listed in idea and research questions section

Step 3: Preliminary Code

Include coding that has been completed at this preliminary stage.

In []: *# Step 3: Write code here. Add additional cells as necessary.*
There is too much code to insert here we have included the .ipynb and .pdf of our analysis s

Step 4: Meeting Schedule

We are all busy. You and your team should be working towards this project on a weekly basis. Share your proposed schedule below.

--== Double-click and describe your project schedule below ==--

We alternate, each working with the data for a few days at a time with a specific goal in mind. When we reach our goal, we communicate with our team member so they can look over our work and continue on with the analysis. All data cleaning, transformation, and most of our preliminary analysis has been completed, and we will be handling our main analysis questions with the time remaining.