Part 4 - Project Repository

The Course Project

The course project consists of five parts:

- Part 1 - Common Analysis sets the stage for the subsequent assignments. In Part 1 you conduct a base analysis. All of the students in the class will conduct the same analysis, but with a slightly different data subset.
- Part 2 - Extension Plan will require you to ask a human centered data science question that extends the work in Course Project Part 1 - Common Analysis.
- Part 3 - Presentation will require you to give a modified (shorter) PechaKucha presentation of your completed project.
- Part 4 - Project Repository, creation of a fully documented repository and also requires the submission of a written project report.
- Part 5 - Peer Feedback, complete semi-structured evaluation and feedback on in-class presentations of your peers.

Project Repository

For this part, you will create a repository for your course project. The goal is to demonstrate that you can incorporate all of the human centered data science considerations you learned in this course and create research artifacts that are understandable, impactful, and reproducible.

A repository should include:

- A written document (a paper, a report) that fully covers all of the results for your assigned US city, including your work for Part 1 - Common Analysis, as well as what you proposed to complete in your Part 2 extension. An outline of possible sections to include in your written report is provided below.
- All your code, organized in one or more notebooks, fully documented.
- Complete data descriptions, and associated links to data sources for all data that you used. Remember to cite your sources.
- Complete documentation of any adopted or adjunctive models used. Remember to cite your sources.
- Documentation of any intermediate data files that you create for your analysis.
- Appropriate README file in .md or .txt format.
- Appropriate LICENSE file

Submission

There are two submissions required to complete this assignment. You will make both submissions on Canvas.

You will first submit a link to your repository that contains all of the pieces of your project (code, data descriptions, links, supporting materials, written report - everything - as described above). Make sure that the repository or drive is shared with the instructional staff so that we can read and score your submission.

Additionally, you will submit a stand alone version of your written report.

Canonical Written Report Sections

The written portion of your report should be clear prose. We expect this to be written in proper english. You should proofread your report to remove misspellings and make sure it is grammatical. Your written document should have the following sections.

1. Introduction
   Why is this analysis interesting or important? What motivates asking and answering the question? Does it solve a real problem or tackle an unresolved research question? Why does it matter?

2. Background/Related Work
   What other research has been done in this area? How does this research inform your hypotheses, your analysis, or your system design? What are your hypotheses or research questions?

   During the development of Course Project - Part 2 you probably considered existing models that you could have adapted or adopted to use with your existing data. The most relevant ones should be covered here and potentially why one was selected. As well, any datasets that you used to extend your Part 1 results should be covered. All of these things are prior and/or related work.

3. Methodology
   Not just your analytical methods, but also, why you chose them, and how human-centered considerations such as ethics informed the way you designed your study.

4. Findings
   What did you find? Use words and figures, don't just point to code.

5. Discussion/Implications

Why are your findings important or interesting; What should the city council, city manager/mayor, and city residents do to address your findings? How long do they have to make a concrete plan?

This section should include a thoughtful reflection that describes the specific ways that human centered data science principles informed your decision-making in this project.

6. Limitations
   This is a required section for your report.There are often many, many limitations for any study. If you honestly tried to list them all, this might end up being the longest section. You should prioritize and list the ones that are most likely to have a significant impact on your results. Specific license issues could be a limitation, depending on what data sources you used. Flaws in the data, data cleaning techniques, potential assumptions and/or how you handled missing values could be a limitation. Statistical techniques often have specific assumptions and preconditions; if you're not certain all of the data meets those requirements - this is a good place to make that clear.

7. Conclusion
   Restate your research questions/hypotheses and summarize your findings. Explain to the reader how this study informs their understanding of human centered data science.

8. References
   A list of publications (blogs, articles, research papers) that you refer to in your text.

9. Data Sources
   A list of links to the relevant data sources that you used.