

cs109a Final Project, Milestone 2
Project C: Predicting Types of Crime

Project team #109

Contributors:

Christopher Campion
Sheraz Choudhary
Fabio Pruneri
Michael Sedelmeyer

1. What is your group #?

- 109

2. Have you met/communicated with your fellow teammates?

- Yes, we discussed the project and milestone 2 in person as a group on campus on October 15th. In addition, we have set up a GitHub repository and Google Drive project folder for remote collaboration on the project.

3. Have you met/communicated with your assigned TF? If not, please provide a reason.

- Yes, Evan MacKay is our assigned TF. Two of us were able to meet with him via Zoom on October 22nd to introduce our team, address questions identified by all of our team-members, and discuss project expectations and supplemental resources related to the project.

4. Has your team formulated a well-defined question to address in your project, based on the project description and references? If so, please write down the question. If your team hasn't done this yet, that's okay!

- No. We are currently still working with the general research questions as they were posed in the project's "Problem Statement". Once our team has had more time to explore and cleanse the starting datasets, identify and inspect additional available datasets, and conduct a cursory exploration of the literature related to crime prediction and inference, we will establish a better defined research question.

5. Briefly describe your team's plans for work to be completed by Nov 20th (tentative milestone three). Please assign specific tasks to team members and deadlines for when these tasks are to be completed.

- Select which features from the properties dataset are potentially useful and identify and execute any initial feature engineering, handling of missing data, and assessment of potential measures of growth by aggregating different years. Sheraz by Nov. 1st.

- Conduct a brief review of academic literature to identify prior findings related to crime prediction and inference to help guide our search for additional data sources and our eventual EDA and model prototyping. Mike by Nov. 1st.
- Convert street names/locations in the property dataset into LAT, LON coordinates. Fabio by Nov. 1st.
- Clean the crime report dataset, handle missing values, recategorize variables where required, and inspect diversity stats. Fabio by Nov. 1st.
- Refine and establish a well-defined question to address in our project based on the available data, project requirements, and literature findings. Together Nov. 5th
- Find at least one new dataset related to policing activity (policing statistics, calls police responded to , etc.), or any other dataset that could be useful and clean that dataset. Mike collaborating with Sheraz by Nov. 5th.
- Find one new dataset that could be useful from any other area and clean as needed. Cristopher by Nov. 5th.
- Convert additional data source observations for geo-locational analysis (i.e. enrich with LAT, LON). Mike by Nov. 6th.
- After initial data collection, cleaning, and exploration, decide whether to treat space as a continuous or discrete variable. If discrete, then decide schema for dividing spatial units (e.g. square miles/grid, districts, streets, etc). Together by Nov. 7th.
- Create features for every area that count the number of (functioning) street lights and the same thing for existing GPS-based datasets. Multiple team-members based on additional datasets found and individuals' familiarity with each by Nov. 10th.
- Create a visual map that summarizes the information collected and analyzed to-date. Chris and Sheraz by Nov. 14th.
- Attempt initial prototype models and other descriptive EDA. Mike and Fabio by Nov. 14th.
- Refactor code and python workflows where needed for reproducibility of results and clarity of interpretation for plots and graphs. Multiple team-members, divided by section as data manipulation and EDA evolves by Nov. 16th
- Write, review, and edit the final version of Milestone 3. Multiple team-members, divided by section and with multiple team-member's review for each section by Nov. 19th

- Submit our final version of Milestone 3. Sheraz by Nov. 20th