Bad Landlords II - Councilor Breadon
Team 4

Maria Shevchuk '23 (maria410@bu.edu)

Eddie Jones '24 (ewj327@bu.edu)

Aaron Liauw '24 (aliau@bu.edu)

Francesco Ciraolo PhD 1 (fciraolo@bu.edu)

Table of Contents

1. Introduction

In collaboration with councilor Breadon, we are performing data analysis on Boston's property and renting datasets to determine which landlords in the area are so-called "bad landlords". The stated aim of the project is to build a trackable system for property violations throughout the Boston area and to create a matrix that determines whether a given landlord is a "bad landlord". Finding a specific ordinance that can be used to define a threshold for a "bad landlord" (also called a "scofflaw" landlord) was another stated goal of the project. The overall motivation and impact of the project is to increase transparency and accessibility for the housing development process in the greater Boston area so that the public can make informed decisions about where to live and understand the shifting landscape of their city.

2. Base Analysis
2.1. Initial Data Analysis

Our initial step when starting the project was to familiarize ourselves with the results and findings from the previous fall semester project. At this point, we were posed the following key questions:

1. What are the top 3-5 most common violation types from the fall project?
2. Can we verify these findings with the data available to us?

After consulting some of the deliverables of the previous teams from the fall semester, we concluded that the most common violation types they outlined were Enforcement Violations, followed by Housing Complaints and Sanitization Requests, with most violations classified in the "low severity" category. (Source: team 2 deliverable 4, team 4 deliverable 5)

Based on our own preliminary analysis, we were able to draw the following conclusions: Our findings from the RentSmart dataset analysis agree with those of team 4 (deliverable 5). We found Enforcement Violations to be the most common type of violation, followed by Sanitation Requests and Housing Complaints (figure on the left). Within the Enforcement Violations, the most common violation is for "improper trash storage" (figure on the right).
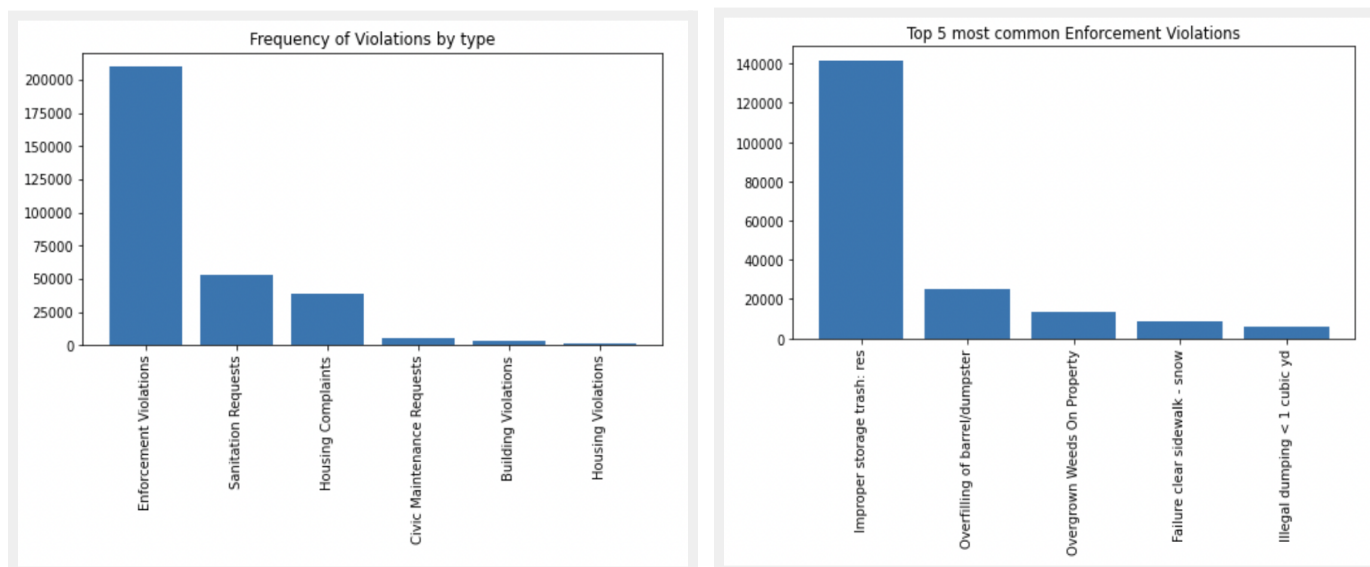
After consulting with the client, we learned that the violations we are interested in do not include any violations enforced by the city/public, nor those unrelated to a specific unit. Upon receiving this feedback, we changed our strategy, and identified the "Housing Violations" in the RentSmart dataset to be our primary category of interest. In this violation category, the most common violations are "Smoke Detectors & Carbon Monoxide Alarms", "Owners Responsibility to Maintain Structural Elements", and "Owners Installation/Maintenance Responsibility".

According to the [Building and Property Violations](#) dataset, the most common violation for fitness of "human habitation" by the Department of Public Health is violation CMR410.500, "Owner's Responsibility to Maintain Structural Elements"; it is then followed by CMR410.351 "Owner's Installation and Maintenance Responsibilities." Other violations include CMR410.550 and CMR410.482, which are "Extermination of Insects, Rodents and Skunks" and "Smoke Detectors and Carbon Monoxide Alarms" respectively. Because the dataset focuses on buildings and property, it makes sense that maintenance takes up a large number of violations, especially in lower-income communities.

	code	case_count
58	CMR410.500	2472
30	CMR410.351 A	1379
70	CMR410.550 B	935
53	CMR410.482 A	794
52	CMR410.481	263
9	CMR410.200 A	214
82	CMR410.602 D	190
11	CMR410.201	174
8	CMR410.190	157
59	CMR410.501 A	149
45	CMR410.451	131
79	CMR410.602 A	125

After completing our preliminary analysis, we were provided with additional key questions:
1. What is the spectrum of violations and severity in regards to "worst landlord" classifications?
2. What landlords are non-compliant? (Overall volume, severe violations)
3. Which landlords are impacting vulnerable communities the most? Is there a correlation between vulnerable communities and violation frequency?

Initially, answering the first question posed some challenges since Boston does not currently have a classification or categorization system for violations and their severity. We brought this concern to the client's attention and were later provided with an ordinance - a working document highlighting specific violations of interest in regards to worst landlords classification. Please refer to section 2.2 (Scofflaw Property Owner List) for continued exploration and reworking of questions 1 and 2.

As part of our extension project, we planned to tackle the third key question, investigating whether there is a relationship between vulnerable communities (based on social vulnerability index, climate resilience, and amount of college students) and housing violations. It is important to note that, as instructed by the client, all properties owned by the city of Boston are omitted from the data analysis. Please refer to section 3 (Extension Analysis) for further insight into our exploration of key question 3.

2.2. Scofflaw Property Owner List

After receiving the ordinance and scofflaw landlord definition from the city council, we have adjusted our final goal to better fit the given criteria. In summary, the ordinance identified 6 violations as the threshold for a landlord being designated as a 'scofflaw', and outlined 3 specific violation codes that need to be considered: State Building Code (780 CMR), the State Fire Code (527 CMR), and the State Sanitary Code (105 CMR 410).

At this point in our project, new key questions were posed:
1. Since data merging is not possible at this time, which available dataset contains the most relevant data to identify scofflaw property owners based on the ordinance?
2. What features, in addition to the scofflaw property owner list, could provide useful insights for tenants and future renters?

Since we don't have access to a dataset that contains all violation codes of interest in addition to landlord/owner names, we decided to work with the housing violations in the RentSmart dataset, which seemed to be of most relevance to our new goal. From this dataset, we are able to extract various housing violations, including fire safety and sanitation-related violations. One of the advantages of this dataset is the fact that it includes the landlord/owner information for each property, while many others don't. On the other hand, a downside of the RentSmart dataset is the violation category labeling, which is inconsistent with standardized city violation codes, and is thus inconsistent with the codes outlined in the ordinance. However, we have been able to identify that the housing violations subsections in the dataset are largely representative of all 3 state codes of interest.

We have been able to match each subsection in the Housing Violations table of the RentSmart dataset to a CMR code. We have identified that the Housing Violations category in the dataset relates specifically to the State Sanitary Code (105 CMR 410). Our code matching can be found in this spreadsheet and a preview is demonstrated in the figure below.

Subcategory in RentSmart's Housing Violations	Code of Massachusetts Regulations (CMR) Number
Smoke Detectors & Carbon Monoxide Alarms - Owner shall provide, install and maintain in operable condition smoke detectors and carbon monoxide alarms	105 CMR 410.482: Smoke Detectors and Carbon Monoxide Alarms
Owners Responsibility to Maintain Structural Elements - Structural elements shall be maintained free from holes, cracks, loose plaster, or other defects.	105 CMR 410.353: Asbestos Material
Owners Installation/Maintenance Responsibility - All Owner-installed optional equipment	105 CMR 410.351: Owner's Installation and Maintenance Responsibilities

We have also identified that the Building Violations category in the RentSmart dataset corresponds to the State Building Code (780 CMR) and State Fire Code (527 CMR). Our code matching can be found in this spreadsheet and a preview is demonstrated in the figure below.

Subcategory in RentSmart's Building Violations	Code of Massachusetts Regulations (CMR) Number
Unsafe Structures	780 CMR 116: Unsafe Structures and Equipment
Maintenance of Means of Egress	780 CMR 102.8: Maintenance of Existing Buildings and Structures
Failure to Obtain Permit	780 CMR 105.1: Required Permits

While a thorough review of our manual code matching results is necessary, we believe that this can be very useful for future work on this project. Violation categories in the RentSmart dataset correspond to the CMR codes of interest highlighted in the ordinance.

After identifying violation categories in the RentSmart dataset that correspond to the CMR codes of interest highlighted in the ordinance, we proposed our solution to the client. We clarified that, at this stage, our plan would be to focus exclusively on the aforementioned violation categories, and that we would implement the proposed threshold of 6 violations to compile a list of scofflaw landlords/owners. The figure below shows the landlords/owners that exceed the proposed threshold of 6 housing violations along with the violation count.

Along with this figure and a list of identified scofflaw landlords, we have added a search feature to our dashboard that will allow the user to select a landlord from a dropdown list and view their violation history along with which specific properties those violations are associated with.

2.3. Dashboard

As part of our project, we created an interactive dashboard to house data from all our different sources in one space. The aim is to help users gain a bigger picture of the housing market when seeking a place to rent in Boston. Our dashboard contains the following pages:

- Index - Our homepage. Provides a summary of our project and explains how to navigate to the other tabs.
- Map - Displays violations from various data sources through openstreetmap. Can select the following: Top 5 311 dataset, RentSmart dataset, Property Violations dataset.
- Search - Utilizes RentSmart dataset to look up violations with "scofflaw landlord" names. Displays owner's buildings and violations (includes information on type of violation, a brief description, and address of the violation)
- Extension - A short description of our extension project, along with graphs the model produced.

The dashboard can be found at the following link:
https://francescociraolo.github.io/bad-landlords-team4/.

2.4. Challenges and Future Suggestions

As discussed on multiple occasions with the PM, in the first presentation to the client, and eventually during the client meeting last week, there are two issues, with no easy solution, that prevent us from achieving maximally relevant results.
- Several datasets, curated by the town or in general by the public administrations, are provided but they are not mergeable in any way;
    - this is true also for the datasets derived from other datasets such as the RentSmart based on the 311 complainants
    - the datasets are also inconsistent with each other (address format, etc)
- The expected and planned goal of classifying the landlords based on number and relevance of their violations is impossible since no dataset contains the required data

After several calls to support and promised tools, we had to clarify how the goal is not reachable and decide with the client and PM how to direct the efforts toward the creation of a reasonable product. Further, we believe that our manual code matching of the RentSmart violations is a good starting point for addressing some of the aforementioned inconsistencies and can be helpful for teams working on this project in the future.

3. Extension Analysis

For the extension analysis, we plan to review the *Climate Ready Boston Social Vulnerability* dataset and see if there existed a high correlation between any of its predictors (disabilities, low income, English proficiency, POC, children, older adults, medical illness) and the response variable, the number of CMR410 Public Health Violations in a region. In essence, we aimed to perform an inference task (as opposed to prediction) to find the most influential factors for public health violations. The reason we chose the number of CMR410 Public Health Violations as a response variable is because the Boston council is most interested in them as violations associated with "bad landlords." By finding these factors, we could potentially find what communities are most vulnerable to these "bad landlords" and a higher risk of violations.

Here are the two key questions we had for the extension analysis:
1. What are the most influential demographic predictors for violation count?
2. Which groups are more vulnerable to property violations?

It is important to note that results of the analysis are not prescriptive; in other words, they do not perform methods of scientific inquiry and testing in the real world. So we should not infer anything about these demographic groups based on the results of the analysis. However, the analysis could shed some light on what groups we might what to consider, for example, when tracking resource allocation or vulnerability.

Process

The first order of business is finding the number of violations per region. To do this, we grouped and sorted data from the *Building and Property Violations Dataset* to find the number of CMR410 violations per Boston neighborhood. We chose this dataset because the *RentSmart* dataset did not have enough city labels for us to perform analysis on in conjunction with the *Climate Ready Boston Social Vulnerability Dataset*. We then merged the resulting data from the *Building and Property Violations Dataset* with the *Climate Ready Boston Social Vulnerability Dataset* by city. After some data cleaning, we applied a machine learning technique called lasso regularization, which applies a shrinkage term, called *lambda*, on coefficients of the model. We selected the best shrinkage term through cross-validation. Although the shrinkage is used to minimize the error in a prediction task, we are more interested in the effect it has on the coefficients as this is what gives insights on the predictors.

Below is a graph demonstrating the effect shrinkage has on the predictors. As the shrinkage term *lambda* increases, we see some predictors (such as 'medical illness' and 'limited English proficiency') merge to 0.00. How to interpret this graph is that terms that merge earlier (towards the left) has less of an impact/importance on the model, while terms that merge later (towards the right) has more of an impact/importance on the model. On the other hand, terms that lie above zero are related to positive rate (related to increased numbers of violation counts), while terms that lie below zero are related to negative rate (related to decreased numbers of violation counts).

Coefficient weights as lambda increases
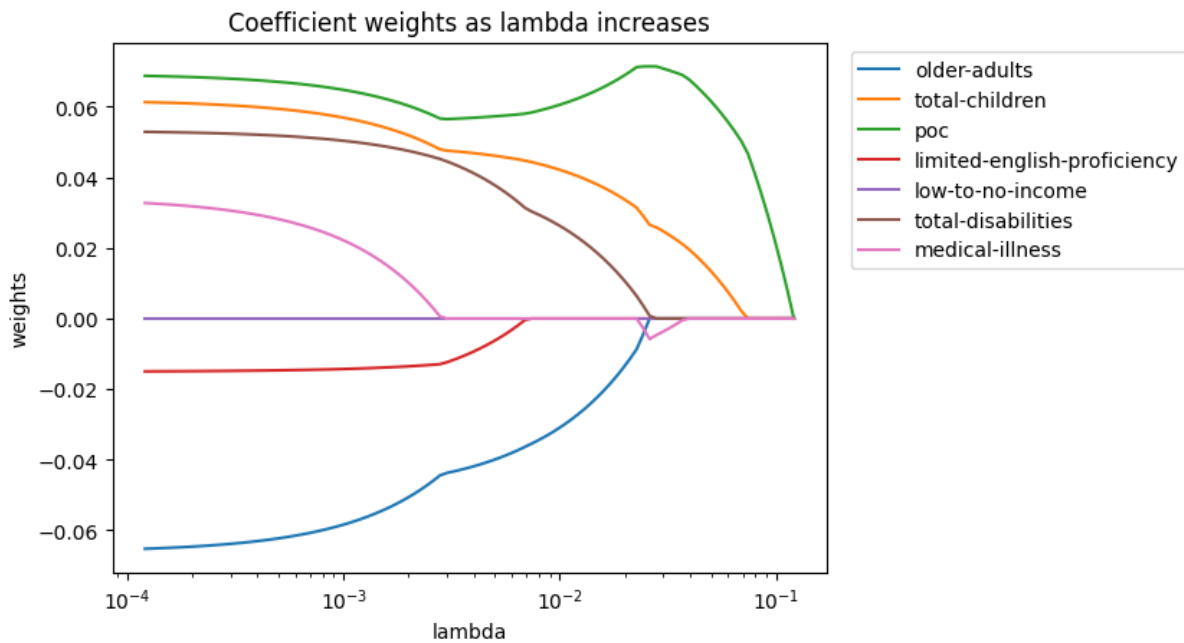
If we take a slice of the graph at the best *lambda* term (where the model makes the least amount of error), we can see the following chart which provides a more intuitive representation of each predictor's effect on the model.

Coefficient weights at best lambda

Since this is not a prescriptive task, we should not interpret the data alone without further experimentation and real-life surveying. However, it is interesting to find that older adults and limited english proficiency have lower rates of violation counts. This could be a sign that there may not be enough resources for these groups to report property violations. On the other hand people of color, cities with a high-children makeup, and cities with more medical/physical disabilities have higher rates of violation counts. For those with disabilities and children, it could be that more attention is put in the community to address public health violations. For people of color, it could be that data is skewed because a majority of POC are renters in Boston instead of owners. Again, data alone cannot explain these relationships and further investigation must be made to find why these relationships exist. We recommend that the Boston city council perform a deeper dive into these communities to determine the causes of these occurrences.
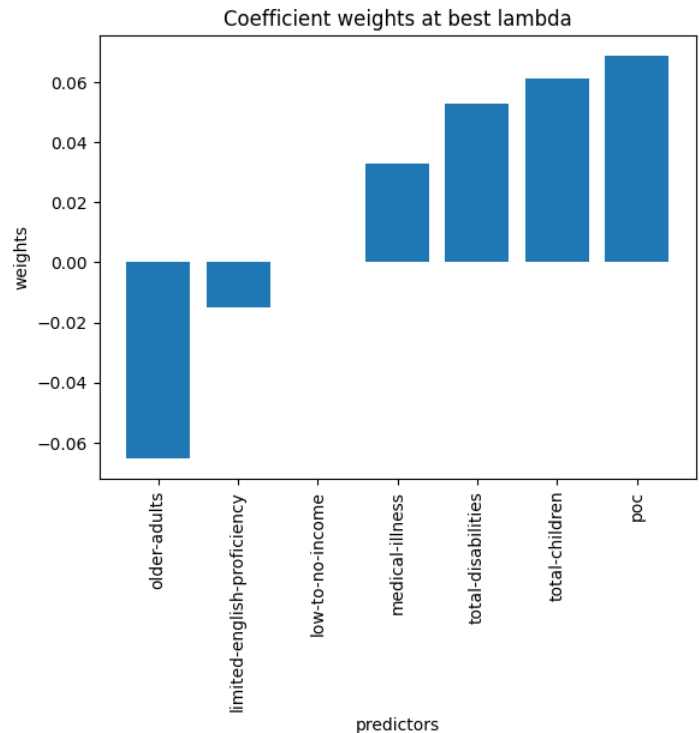
4. Member Contributions

Although all members have been core contributors to this project throughout the semester, below is a more detailed breakdown of each member's contribution.

Maria Shevchuk - cover page and sections 2.1, 2.2 (report), repo readme(s), any analysis related to *RentSmart* dataset, CMR code matching for *RentSmart* dataset

Eddie Jones - section 1 (report), preliminary analysis of *Income-Restricted Housing Inventory*, assistance with cleaning + merging datasets for extension project

Aaron Liauw - extension project section 3 (report), analysis related to *Building and Property Violations Dataset* and *Climate Ready Boston Social Vulnerability*

Francesco Ciraolo - sections 2.3 and 2.4 (report), datasets analysis and cleanup, dashboard creation, optimization and improvement