

Saving Lives, Time, and Money: Using Data to Find Unsafe and Unhealthy Buildings Faster

Housing inspections are traditionally done periodically, in response to complaints, or when owners or renters turn over. A new study shows that machine learning can help inspectors identify risky homes nearly twice as fast as conventional practices.

By Katharine Robb, Jorrit de Jong, and Lisa Cox

Is it possible to save lives, time, and money, all at once? Could you use data that you already have to do it? As we describe in our recent peer-reviewed paper, the [answer is yes](#). We found that machine learning could help a city identify properties with health-threatening housing code violations nearly twice as fast as conventional practices.

More and more cities—large and small—are harnessing machine learning and existing municipal data to help power charge a city's efforts to deliver services to those who need them most. In Chelsea, Massachusetts, a city of more than 40,000 people packed into less than two square miles just north of Boston, our research team worked with the city to do just that.

A New Look at Old Problems

Chelsea residents are predominantly recent immigrants, people of color, and low-income families. With skyrocketing rental costs and stagnant wages, overcrowded conditions in substandard housing are common. The city's housing inspectors encounter families doubling or tripling up together, living in unfinished basements, windowless closets, and open-air porches. Some lack access to bathrooms, kitchens, and even plumbing. Setups like these harken back to an earlier era, an era when we learned that unacceptable living conditions not only devastate families but offer fertile ground for [disease to flourish](#) and fires to blaze, putting us all at risk.

Setups like these harken back to an earlier era, an era when we learned that unacceptable living conditions not only devastate families but offer fertile ground for disease to flourish and fires to blaze, putting us all at risk.

In 2019, our research team (that holds expertise in public health, public policy, and data science) partnered with Chelsea's Inspectional Services Department (ISD) to see if we could use city data to [root out long-hidden, health-threatening housing code violations and nip others cropping up in the bud](#). More specifically, could ISD use a data-driven model to be better and faster at finding health-threatening housing code violations?



Finding Threats More Quickly

Housing codes represent the absolute minimum standards for safe housing—such as protection from the elements, a way to escape from fire, and working smoke detectors—and violations **can be hard to find**.

Many violations are never called in for fear of landlords raising rent, language barriers, or residents simply not knowing that they can. Often, the most vulnerable residents are the least likely to complain.

Many violations are never called in for fear of landlords raising rent, language barriers, or residents simply not knowing that they can. Often, the most vulnerable residents are the least likely to complain. On the other hand, cities can only inspect so many properties in a day. Finding—and prioritizing—the people that live in the riskiest places could help inspectors enforce housing codes not only more quickly and effectively, but more equitably, too.

To gather clues on where to find code violations, our team used data from inspections that the city had done proactively from 2015 to 2018 and the expertise of the city’s housing inspectors. We then trained a computer (using a process called machine learning) to make educated guesses about which properties should be inspected first (predictive analytics). We found that “listening” to these predictions could nearly double the number of productive inspections. That is, the “hit rate” for inspections identifying health-threatening housing code violations could increase by a factor of 1.8. Chelsea could now identify and resolve more housing code violations without increasing inspection resources. In essence, the city could “do more with less.”

And, as always, there is still more to do. Now that the city has become better at finding risky conditions, our team is helping to look at how to resolve those risky conditions. Currently, Chelsea is partnering with a local social service agency so **inspectors can do more than issue citations**. They can now make referrals for those who face challenges complying with housing codes because they experience mental illness or poverty, connecting people to the services they need to live healthier lives.

That is, the “hit rate” for inspections identifying health-threatening housing code violations could increase by a factor of 1.8.

The Upshot for Other Cities

While our study demonstrated that machine learning could help a city save lives, time, and money by more accurately identifying unsafe and unhealthy housing earlier than conventional code enforcement, we believe that many more machine learning applications are waiting to be discovered. To help fulfill that potential, we offer three lessons from the Chelsea study:

- 1 Machine learning does not substitute the knowledge of city officials but leverages and expands it.** More and more, cities can pull data from tax assessors’ offices, utilities, public works, and other sources to help train computers to pinpoint where city staff’s attention is most needed. However, on-the-ground knowledge is imperative for checking and balancing decisions, ensuring that the actions that are based on data are human-approved. Used conscientiously, machine learning can help cities improve a wide range of operations, garnering faster, more effective, and more equitable results.

- 2 Investing in data-analytics takes time, resources, and effort, but when cities leverage partnerships a lot is possible.** While some cities may be quite far along in their data capabilities, others will need to [invest substantial resources of time, money, and expertise in data integration](#) and data norms. But it's important to note that the efforts shouldn't be shouldered for just one cause—these benefits extend far beyond single city initiatives or departments. Furthermore, if cities are willing and creative, they can forge partnerships with universities and others to integrate and analyze data, as Chelsea did. Moreover, models don't need to be developed from scratch. Code from local government machine learning projects are often published online, which can be adapted for others to use. (For instance, our code is available on [GitHub](#).)
- 3 Using data can make cities more effective, equitable, and efficient but only if protections are in place, biases are corrected regularly, and human judgement remains part of the process.** Data-driven models are not perfect. After all, they contain assumptions that may be incorrect in certain cases and biases of the data used to create them. To ensure that models are not punitive or invasive and to protect citizens from racial profiling and other destructive biases, models need to be piloted and assumptions need to be questioned. Cities need to actively manage the models by regularly engaging diverse partners to question how they work, then using that feedback to update the models accordingly. For example, you may ask your community the following: is the model moving inspection resources to more needy parts of the city? Is it putting an unfair burden on landlords? How is it impacting tenants?

No matter the size or state of the city, machine learning and predictive analytics can help cut wasted energy by identifying high priorities faster, catching problems earlier, and finding the most vulnerable residents before it's too late. And in the world of city services, being more effective, efficient, and equitable not only saves time and taxpayer dollars, but it can save lives.

No matter the size or state of the city, machine learning and predictive analytics can help cut wasted energy by identifying high priorities faster, catching problems earlier, and finding the most vulnerable residents before it's too late.

Further Reading

- “Using Integrated City Data to Identify and Intervene Early on Housing-related Public Health Problems”
[Journal of Public Health Management and Practice](#)
- “Hiding in Plain Sight: Tackling housing-related public health problems using existing city data and machine learning”
[Data-Smart City Solutions](#)
- “Can We Predict Where Housing-Related Public Health Problems Exist Using City Data?” | [Data-Smart City Solutions](#)
- “Aggregate to Innovate: Lessons from Chelsea, Massachusetts” | [Harvard Ash Center](#)
- “On Further Inspection: Engaging housing inspectors to improve public health in Chelsea, Massachusetts”
[Harvard Ash Center](#)
- “How Cities Can Use Housing Data to Predict COVID-19 Hotspots: Lessons from Chelsea, MA” | [Data-Smart City Solutions](#)
- “At the Center of the Outbreak: Researcher Katharine Robb details how housing policies affect social and health crises, like the current pandemic” | [The Harvard Gazette](#)
- “State Capabilities for Problem-Oriented Governance” | [Perspectives on Public Management and Governance](#)
- “Integrating Housing Code Enforcement with Social Service Provision to Improve Public Health in Chelsea, MA”
| Submitted March 2021, [Pre-Print Available](#)