

<b>The Bottom Line</b>	The City of Cincinnati and the Data Science for Social Good (DSSG) team are helping code inspectors prioritize properties likely to have a significant code violation. By predicting future violations, they can reduce unnecessary inspections and get to problem properties early, before significant damage to the property occurs. Preliminary results indicate that the "discovery rate," or the rate at which inspections result in discovery of an actual violation, could increase by more than 30% by using the model developed by the team.
<b>Problem</b>	To date, most code inspections occur as the result of complaints from neighbors. About 40% of these inspections do not result in a code violation, while other violations go unreported and eventually lead to health and safety risks.
<b>Leadership</b>	The City's Chief Performance Officer and the Chief of the Code Inspection division drove the project forward, marshaling the necessary resources from the City's side and helping refine the project's scope.
<b>Staffing</b>	Other than the leaders identified above, the project required staff time from the data owners and aggregators, with some assistance from the inspectors doing the enforcement work on the front lines.
<b>Data</b>	The project used data collected by the City of Cincinnati, including parcels, code enforcement, assessments, and water shutoffs. Census data was also used. The data science team also relied on GIS data from Hamilton County, of which Cincinnati is a part.
<b>Technology</b>	The project uses open-source software. Data is provided to the data science team in a variety of formats, including spreadsheets and access to databases, and the list of priority buildings for inspection is served back to the inspectors in spreadsheets.
<b>Methodology</b>	First, city officials scoped the problem. Second, data owners aggregated and delivered the data. Third, the data science team analyzed and processed the data to create a model. Finally, the list of priorities was delivered to the City, for testing with a select group of inspectors. A final phase will involve embedding the prioritization tool into the city's system of record.
<b>Results</b>	This project is ongoing. To date, the team has increased the "discovery rate," or the rate at which properties inspected are found to indeed have a code violation. Although the results are still preliminary, the model has the potential to improve the discovery rate by 30%. The Cincinnati team has noted that its model does a good job of identifying neighborhoods or areas with large numbers of violations, but it hopes to improve the model to better identify

troubled individual parcels in strong neighborhoods.

**Replication**

The University of Chicago team noted that the data was a key ingredient to this project. All of the datasets that were used had at least three to five years of historical data, allowing the team to test its prediction models against actual violations received. The City of Cincinnati team also emphasized that it was important to have someone embedded within the department who was extremely familiar with the processes the project targets and who understands the potential value of predictive analytics.

**Learn More**

Data Science for Social Good | Center for Data Science and Public Policy |  
University of Chicago | [www.dssg.uchicago.edu](http://www.dssg.uchicago.edu) | [datascifellows@gmail.com](mailto:datascifellows@gmail.com)