Machine Learning Project Proposal
Group: Liza Soriano, Luise San Martin, and Jessica Langston
CAPP 30254 | April 30, 2019

Background

According to our own analysis of data from the Eviction Lab at Princeton University, eviction rates and eviction filing rates across block groups have returned to pre-great recession levels in the City of Chicago. Nevertheless, they disproportionately affect African American population and families living in poverty[1].

According to Matthew Desmond, the author of Pulitzer Prize-winning book "Evicted", evictions are both a symptom and a source of homelessness and poverty in America. In Cook County, many residents are displaced from their homes as rents rise and more high-income earners move into Chicago and the surrounding suburbs.[1] Evictions because of property foreclosures, demolitions, or evacuations due to building neglect are also occurring in the county.

This housing crisis in Cook County is disproportionately affecting low-income, minority residents. The vast majority of low-income households in Chicago are rent-burdened (i.e. rent is more than 30% of household income): 84% of households earning 30-50% of Chicago's median income are rent-burdened, while those earning less than 30% of the median income are severely rent-burdened (rent is more than 50% of their household income).[2] As a consequence, higher numbers of evictions occur in areas with more poverty, frequently coinciding with locales in the South and West sides of the city where African American and Latino populations traditionally reside. A study shows that the predominantly Latino neighborhood of Pilsen has lost about 10,000 Latino families to gentrification since 2000.[3]/[4] More than 200,000 African Americans have left Chicago during the same time period.[5]

Lack of affordable housing along with stagnant income rates contribute to conditions that force families to leave their homes. Evictions can further exacerbate this problem because they are put on an individual's court records, which many landlords screen for, thus making it harder for evicted families to secure housing in the future. Therefore, minimizing the adverse impacts of evictions should be a priority for policy makers and organizations tasked with ensuring housing justice for Cook County's residents. Addressing this problem would also be in the interest of entities that want to reduce racial inequalities in the county.

Illinois laws and several Chicago ordinances decently protect tenants' rights in Cook County. These guidelines limit acceptable reasons for evictions, prohibit discriminatory or retaliatory actions by the landlords towards tenants, and in many cases place the burden of following these stringent rules upon the landlords.[6] However, when tenants are unaware of their rights or do not have the resources to adequately deal with landlord conflicts, improper and illegal evictions can occur. Free resources provided by the city, along with outreach programs by many housing rights advocates and NGOs therefore try to fill this information gap and assist tenants in dealing with abusive landlords. These services can help tenants stay in their homes or at least lessen the impact of displacement.

---

[1] See Appendix 1 for more information.

<u>Goals</u>

Our proposed solution consists of building a Machine Learning algorithm to identify areas of the Cook County (at the block level) that are most vulnerable to evictions. We are imagining we are working for the Metropolitan Tenants Organization (MTO), an actual tenants' rights organization that operates in Chicago. Through their Tenant Stabilization program, MTO works with tenants to document and resolve housing issues, equipping them with the tools and resources they need to prevent eviction and/or determine their next steps if they are evicted. By identifying areas of Chicago that are most vulnerable to evictions, we will be able to help MTO focus their resources on the people who need it the most.

In order to identify the block groups most susceptible to eviction, we will use this variable for our machine learning models: the ratio of evictions to eviction notices for a block group in a year[2]. By looking at the ratio of evictions to eviction notices, we will be able to identify where eviction notices are most likely to result in an eviction. We imagine that a block with many eviction notices but not many actual evictions may not be as high of a priority for MTO to target in their outreach because perhaps they are already knowledgeable about what to do if given an eviction notice (whether it means they are able to dispute an unfair eviction, find other housing quickly, etc.). Our actual label will be a binary variable indicating whether or not a block group falls in the highest 10% of eviction filings resulting in evictions. This threshold will be a parameter of our model that we can adjust if we see fit.

<u>Analysis</u>

We have data from the Eviction Lab at Princeton University. It includes demographic and housing spatial information at the block group level for the City of Chicago, which we define as Cook County. The timespan of this dataset ranges from 2000 to 2016. We have decided to focus on the five most recent years, 2011 to 2016, but this date range will be a parameter of our data extraction code that we can adjust if we see fit.

The first data exploration we conducted on the Eviction Lab dataset showed it is plausible to use it for the purpose of our project. The dataset has columns for evictions and eviction filings, which will allow us to calculate the ratio of eviction to eviction notices and the number of eviction notices that end up being actual evictions in all the block groups across Chicago in a certain year, during the chosen timeframe. The dataset contains a total of 3,993 block groups for each of the years of analysis. None of them has missing values for the columns evictions and eviction filings during 2012-2016. Nonetheless, some of the potential predictor features of this dataset present missing values for this timeframe, like rent burden (missing 18% of data points), median property value (missing 5%), median household income (missing 3%) and median gross rent (missing 20%).

Our data exploration also helped illustrating some of the ideas mentioned in the problem definition. In 2016, the eviction filing rate across all Chicago was 3.4%, while in block groups where the African American population was higher than 50% it reached 6.7% and in block groups where the poverty rate was higher than 30% it was 5.1%. Eviction rates have decreased during the last five years, while eviction filing rates have maintained similar levels.

---

[2] We define this indicator as the ratio between the number of evictions and the number of eviction filings in a given area during a certain length of time.

Table 1: Eviction and eviction filing rates in the Cook County, 2012-2016

Year	Eviction rate	Eviction filing rate
2012	1.6%	3.9%
2013	1.4%	3.6%
2014	1.2%	3.5%
2015	1.0%	3.2%
2016	0.9%	3.8%

Data source: Eviction Lab at Princeton University

We will explore using the rest of the variables included in the dataset as features of our Machine Learning model, for example: population, median household income, percentage of renter occupied units, percentage of population living in poverty, and others. Additionally, we will augment this dataset to explore if adding additional demographic and spatial features from other data sources can improve the proposed Machine Learning algorithms we will define. Two datasets we will include these additional features from are the American Community Survey and Open Street Map.

We will use the Machine Learning pipeline to generate different types of models we have learned about, including but not necessarily limited to stacking methods, bagging methods and simple classifiers (decision trees, logistic regression and support vector machines). Then we will select the best model based on our evaluation criteria (specified below).

We will validate every one of our models using temporal hold outs, since we are working with data over time. To achieve this, we'll randomly split a number of training and testing datasets and examine key metrics of accuracy, precision, and recall on each model and test dataset combination, to determine how each model performs compared to each other and compared to a baseline we define as a simple 5-level decision tree model.

Caveats

Since we only have data through 2016, the models that we build will make predictions of which Cook County blocks are most susceptible to evictions as if it were the beginning of 2017. We will build our pipeline in a generalized way such that we could easily apply our models to more recent data if we had it.

The eviction rate figure comes with some caveats, in that it does not always paint the full picture of what happens when someone is evicted. It does not count scenarios where tenants agree to move out before facing judge, which is what happens before an eviction officially occurs. For example, the tenant might find alternate housing quickly, might accept cash from landlord in exchange for keys, or might not show up at mandatory arbitration at courthouse.

<u>Policy Outcomes</u>

Our goal is to help MTO target their outreach and interventions to the Chicago residents most vulnerable of eviction, so if we are able to reduce the time and effort they, as a non-profit with limited resources, spend prioritizing where to focus their work, we are at least partially fulfilling our desired impact. The model/s we produce would also be useful to government agencies and any other non-profit groups working to address affordable housing in Chicago.

Ideally, we would like to see a reduction in both the number of evictions in targeted block groups, as well as a reduction in the proportion of eviction filings that result in eviction, because that would indicate that MTO's work is having an impact in helping people avoid eviction.
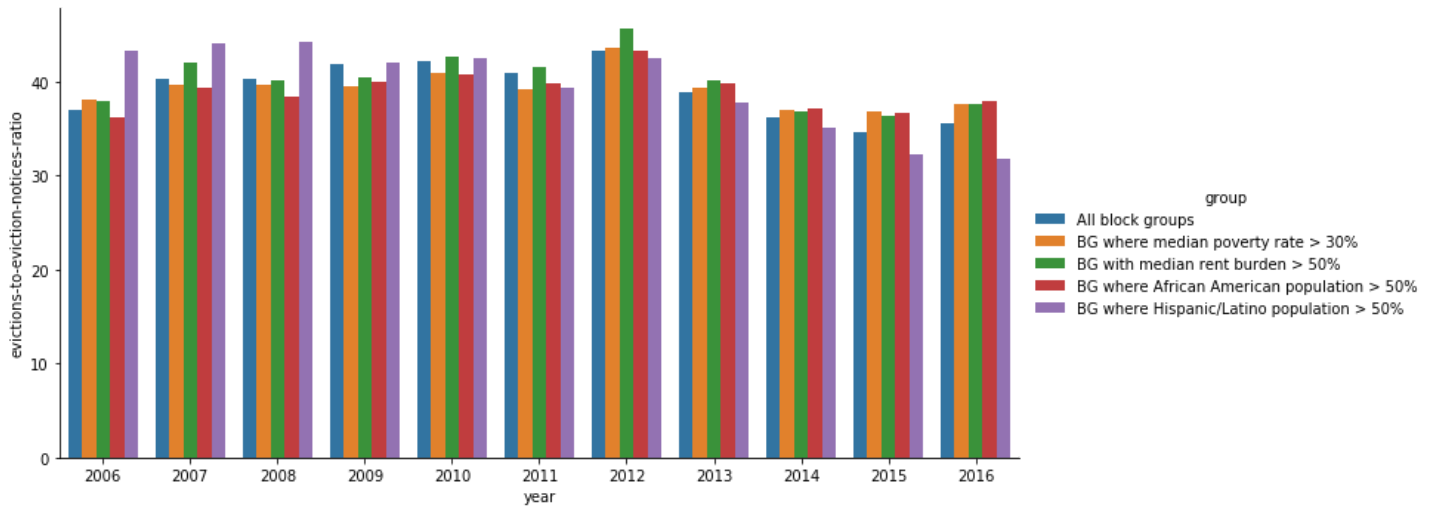
Appendix 1

Graph 1: Eviction filing rates (percentages) across years, for selected block groups
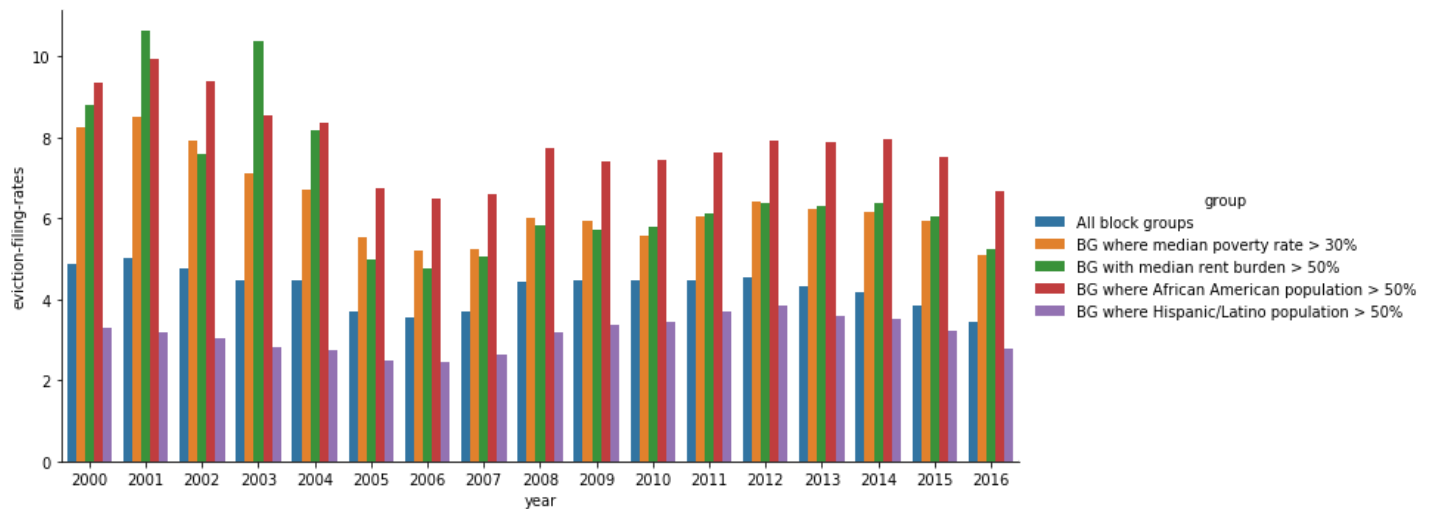
Graph 2: Eviction rates (percentages) across years, for selected block groups
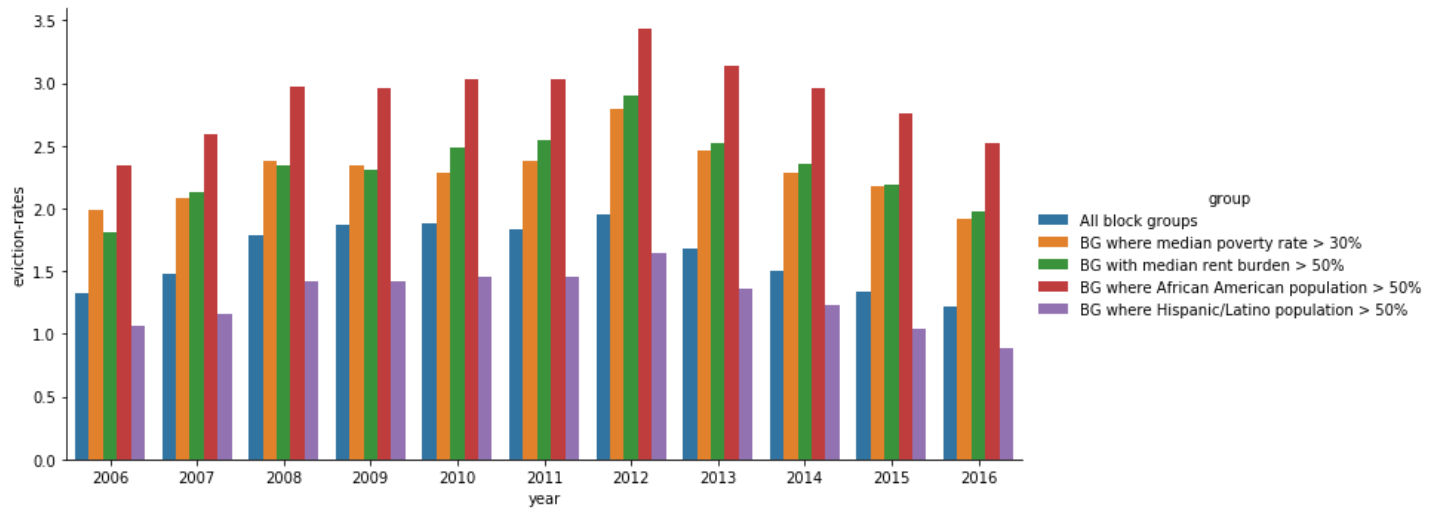
Graph 3: Eviction to eviction notice ratio (percentage) across years, for selected block groups

A more in-depth data exploration is available in a Jupyter Notebook of our project repository:
https://github.com/luisesanmartin/ml-eviction-chicago/tree/master/notebooks