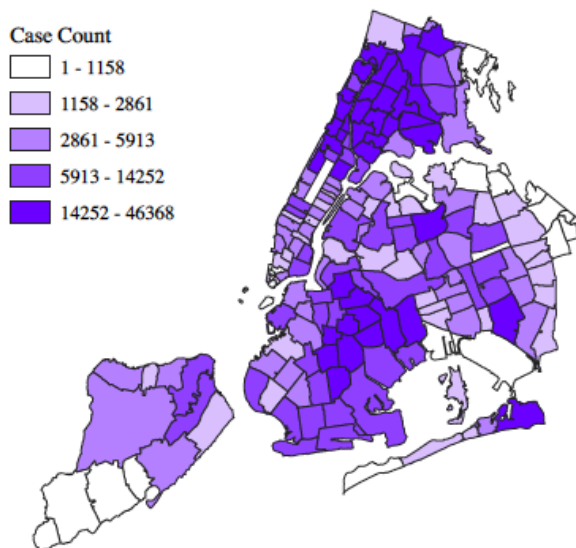## *Tenant Outreach Memorandum*

The eviction crisis is one of the primary contributors to the urban housing problem, and those who are most at-risk are residents already living on low income. Through a combination of supervised learning, causal inference techniques, and spatial analysis, I study the effects of occupied housing and renter demographic characteristics on residential housing case filings in New York City. The most relevant indicator areas with higher likelihoods of eviction are home size, the duration that the tenant has lived in their home, and race/ethnicity. Namely, those who are more inclined to face eviction are individuals without a proper kitchen, those who have moved in more recently, renters living in physically smaller homes. In terms of race and ethnicity, Black residents followed by Latinx residents, are more targeted for eviction than White or Asian residents. Furthermore, other factors such as median rent, age and education appear to be less significant in comparison. Nonetheless, rent is still positively associated with eviction. The age population most affected is between 55-64, and the education group that has completed some college or associate's degree for education. Based on the spatial analysis, however, the most significant factor is location. The spatial cluster map and spatial statistics, shows high indication of spatial dependence and clustering, with the Bronx and west Brooklyn having the biggest clusters and most density of eviction filings, while Staten Island has lower housing court cases. As a further analysis, I inspect the spatial correlation between the Black population density and court cases, and find a significant positive relationship in the figures below. Given these important findings, it is recommended that city policymakers prioritize the location of outreach to neighborhoods in the Bronx and West Brooklyn. In terms of individual characteristics, I urge them to work more with members of the Black and Latinx communities, adults nearing retirement, and those who have not completed their college program.

Residential Housing Cases by NYC ZIP Code   **Black Population Density**
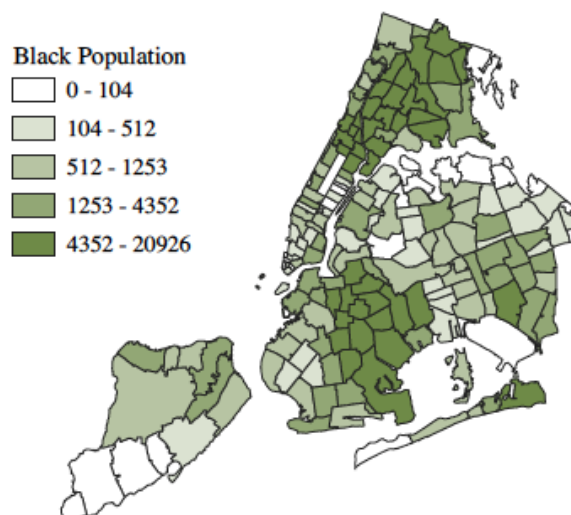
Problem

While some New Yorkers have trouble finding housing as a result of the rise in rent prices, others are being displaced from their homes. Those who are involuntarily displaced through eviction even continue to face financial troubles in the future, as getting evicted can also hurt their credit score or prevent them from finding other housing options. Moreover, there are large disparities in evictions. Studies in the past have shown that there is a high risk of eviction for low-income women and Blacks and Hispanics/Latinx[1]. As previously discussed, the residential housing cases in NYC also show clear signs of eviction discrimination by location and between the pre and post-pandemic. I contribute to these findings further by determining what types of communities and housing situations show are more at-risk of eviction, as well as what kinds of indicators determine the outcome of an eviction over a denied court decision.

Data and Methods

First, I investigate if there is spatial dependence for residential housing cases in NYC with spatial data analysis software (QGIS and GeoDa). With the index datatable within the OCA Housing Court Records, I obtain a count of all housing cases per postal code. I then use the ZIP Code Boundaries shapefile from NYC Open Data to display the density of residential cases by ZIP code. Due to the abnormality of ZIP code shapes in NYC, I set the spatial matrix to Queen's. From there, I analyze spatial dependence with the Moran's I statistic, then run a Local Univariate Moran's I clustering analysis to detect any areas with clusters of spatial correlation.

For the second analysis, I identify which factors can determine the outcome of an eviction case by running a logistic regression model on 1,039,894 cases, and do so in the following manner. I first define my target variable as the court outcome. The data for this comes from the decisions table within the OCA Housing Court Records. I specifically use the highlights column to classify the case outcomes into a binary judgment or denied outcome. Judgment and default judgment are assigned a 1, whereas a null, stipulation, or denied is assigned a 0. For the feature variables, I collect data from the Census Bureau's 2021 data on demographic characteristics describing the tenant population by ZIP code. I am particularly interested in the following demographic characteristics of renters: race/ethnicity, education, age, and move-in year. All variables are provided in count values. Studying the renters population, rather than the general population, gets us closer to understanding the types of individuals that are at higher risk of eviction. This data is then merged with each court case such that every row represents a court case/verdict and its feature variables related to tenants. Lastly, I run a logistic regression model to determine the most meaningful features, with my target variable being a binary indicator of facing eviction judgment or not.

I also conduct a generalized linear regression with Python, but instead of comparing denied from judgment cases, I predict the number of total eviction cases for each census tract and highlight the meaningful predictors. This method in some ways is a better assessment of eviction compared with my previous model because considering that the decisions data shows a

---

[1] Desmond, Matthew. "Eviction and the Reproduction of Urban Poverty." *American Journal of Sociology*, vol. 118, no. 1, 2012, pp. 88–133., https://doi.org/10.1086/666082.

disproportionate amount of cases favoring the plaintiff over the defendant in eviction cases, I assume that the characteristic differences between a person becoming a defendant for a case and not being one could be substantially greater than than being a defendant for eviction and receiving judgment. For my feature variables, I take the same renter characteristic data, collected at the census tract level. Furthermore, I add features regarding the conditions of occupied houses from the 2021 Census Bureau's selected housing characteristics, including the number of rooms and type of in-home facilities. To obtain my target variable - the raw number of eviction cases - I use the NYC Open Data's Eviction Data. Finally, I merge the target data with my feature data and find results strongly consistent with the previous regression model in terms of demographic relationships.

Limitations and Future Research

When plotting the residuals of our linear model, I noticed small indications of clustering between eviction cases and certain population variables. Additionally, I became concerned with the presence of outliers after conducting Grubbs tests on several feature variables. Thus, as a robustness measure, I include an additional robust regression model with Huber-White clustered standard errors in R and results that resonate with the previous models.

One limitation pertaining to our second analysis was how the pandemic has impacted eviction. While this is out of the scope of this analysis, I encourage scholars to consider studying how the pandemic has changed the dynamics of eviction and how it affected different groups of people differently. While this research analyzes spatial autocorrelation of housing cases, others could consider going a step further and conducting spatial lag or error models to control for the spatial autocorrelation that was found when determining what factors indicate high likelihood of eviction.

Another shortcoming for our study is that I did not analyze the demographics of the evicted persons due to the ethical concerns of individual characteristics being tied to eviction cases. I attempt to navigate around this by investigating the relationship between eviction cases and renter population characteristics and do so at various areal units. However, we cannot be entirely confident that the renters described by the level of eviction cases is exactly reflective of the individual tied to each case. Nonetheless, we can be more careful about assuming the relationship and still use the information as a tool to help craft a tenant outreach plan.

Our models and maps analyze the eviction and demographic rates and counts rather than eviction rates. Although in many cases it is more beneficial to use rate for a more normal distribution, we preserve the raw count of our data because many of our feature variables were initially presented in counts, some of which were unable to be converted to rates. We encourage scholars to study this matter further, using rates rather than counts.

Lastly, while this study analyzes the spatial relationship between two significant indicators - Black population and location - it does not go further to investigate the interaction effects between each indicator, such as race and education or tenant duration and median rental

costs. Thus, I encourage future studies to include two way (or more) interaction effects between certain renter and housing characteristics.