ECONOMIC PROFILE OF NYC HOUSING VIOLATIONS

Anushka Kataruka, Arman Deendar, Shravya Sompalli, Tabitha Lynn

Background

Understanding the distribution and persistence of housing violations is crucial for ensuring safe and equitable living conditions in cities. Historically, disinvestment in poor areas of major cities has led to worse building conditions. In addition, housing violations have historically reflected this disinvestment or signalled collaboration between municipalities and private investors to displace poorer residents by citing buildings and subsequently tearing them down.

We aim to shed light on the various contributing factors to housing violations in NYC. In particular, we wanted to explore the relationship between housing violations across bureaucratic organizations in NYC and socioeconomic factors.

Problem Statement

We want to investigate the relationships between housing violations and income levels. Specifically, we wish to determine if violations are more likely to occur and more likely to remain ongoing in lower income neighborhoods.

Data

Our data was obtained from the Census Bureau and NYC Open Data APIs. We merged two datasets for the basis of our analysis: Housing Maintenance Code violations and Department of Buildings (DOB) violations. After cleaning the merged dataset, we also merged median income data for each zip code in the New York City area. The final dataset contains the following columns: house_number, street, zipcode, boro, latitude, longitude, type_of_violation, class, status, year_of_violation, med_income, and above_median.

Results & Methodology

Test 1: We used chi-square independence test to check if there is a relationship between whether a housing violation case is open or closed (the case's status) and the income classification (high income vs low income) of the violation zipcode. With a T-statistic of **0.41059** and p-value of **0.5217**, we fail to reject the null hypothesis. There is no sufficient evidence to prove that there is a relationship between the housing violation case's status and the income classification of that zipcode.

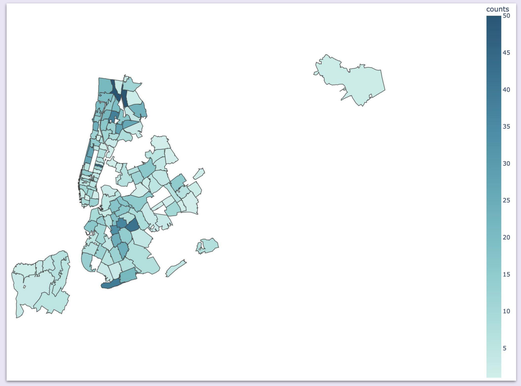

Figure 1: chloropleth map of housing violations in NYC by zipcode

Test 2: We used a one-sample t-test to test whether the median income across the zipcodes where housing violations, i.e., the zipcodes in our dataset, is the same as the median income in New York in 2021. With a T-statistic of **-3.4522** and a p-score of **0.0005725**, we reject the null hypothesis.

Test 3: we used a two-sample t-test to determine if the average number of violations per year in high income zip codes are statistically different from the average number of violations per year in low income zip codes. We receive a T-statistic of **-14.09345** and p-value of **7.5536e-12**

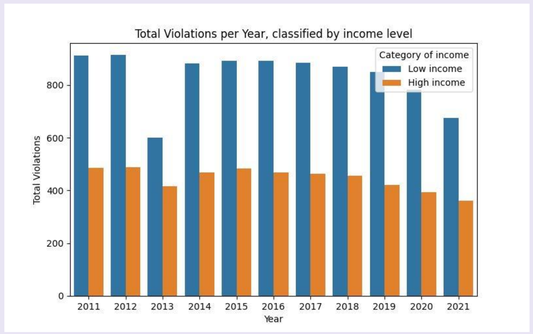

Figure 2: Bar chart depicting the disparities in housing violations between high income versus low income populations, 2011-2021.

The p-value is much smaller than 0.05, thus we reject the null hypothesis. Based on these results, we can conclude that there is a statistically significant difference between high income zip codes and low income zip codes in the average number of violations per year.

Limitations

One major limitation on our analysis comes from our merged Census Bureau dataset. Since we are basing income levels on zip code, there is an immediate assumption that our results could be biased. Because of the variability in characteristics per each zip code, especially in New York where entire buildings have their own zip codes, relying on zip code as a geographical standard for which then to base median income off of, presents us with the possibility of discrepancies as not all zip codes are consistent for a given demographic characteristic (population). In addition, zip codes are problematic in that they do not accurately represent human behavior or real boundaries; they tend to be less representative of income disparity in large metropolitan areas. As such, these findings skew our analysis as we are assuming some sort of standardization to base our classification of zip codes as higher income or lower income (based on the median NYC data).

Conclusion

With the given dataset, we conclude that there is not enough evidence to claim a relationship between whether a violation case is open or closed and the median income classification. Based on our second two tests, we can prove a connection between having more housing violations in lower income areas. The median income across households with housing violations is not similar/equal to the median income in New York and there is a statistically significant difference between high income zip codes and low income zip codes in the average number of violations per year