Classifying Airbnb Listings in New York City

Jonathan Pichot, Fernando Melchor, and Avikal Somvanshi
Center for Urban Science + Progress
New York University
New York, NY

***Abstract*—**

I. Introduction

New York City is one of America's most visited cities, with iconic monuments like the Statue of Liberty and the Empire State Building, a vibrant arts and cultural scene, and an enviable conentration of bars and restaurants catering to every taste and budget. It's a vast city including hundreds of neighborhoods each with their unique history, economy, and character. The city attracted nearly 60 million visitors in 2015, with a year over year growth of about 2 million vistors a year since 2010.[1] Of course, all these visitors need a place to stay.

Airbnb provides a platform for residents to rent space in their homes and apartments. Founded in 2008, Airbnb's mission is to help people "monetize their extra space." They've been very successful, with over 3 million listings in over 65,000 cities worldwide.[2]

II. Literature Review

III. Data and Methods

The four datasets collected for the analysis include:

A. Asthma Hospitilization Discharge Rate

Asthma hospitilization data was collected from New York State's Department of Health (NYDOH). The NYDOH publishes hospital discharge numbers collected by the Statewide Planning and Research Cooperative System (SPARCS). For this dataset, they focus only on hospital discharges where the principal diagnosis was asthma. This dataset is a three-year average over the years 2012 to 2014. The discharge rate per zip code is the total number of discharges divided by the average population of that zip code over the same three-year span. Certain zip codes with particularly low discharge rates are redacted for privacy reasons.

NYDOH publishes the data on their website as tables per county but they do not provide a direct download. The data was collected by scraping the NYDOH website for the five counties in New York City: Kings, Richmond, New York, Bronx, and Queens. Zip codes without enough discharges to be published because of privacy and those with such a low count to have a Relative Standard Error of greater than 30% were also dropped. This resulted in a total of 173 zip codes of asthma data.

B. Building Energy Usage

Local Law 84 in New York City requires owners of large buildings (and soon to be mid-sized buildings) to benchmark the energy usage of their properties and submit this data in a consistent way to the city. These benchmarks have come to cover many different energy and pollution related attributes. The two attributes relevant to this study are Site EUI (kBtu/ft2) and Direct GHG Emissions (MtCO2e). Site Energy Use Intensity is the amount of energy consumed by the property in British thermal units per gross square foot. Direct GHG Emissions are the total greenhouse gases emitted by the property in metric tons of carbon dioxide equivalent.

Site EUI is used to understand the energy intensity of a building, while Direct Emissions will be used to determine if there is a relationship between greenhouse gas emitting buildings and asthma rates within an area. The data collected was for the year 2013. This year was chosen to fit in the middle of the three-year average (2012-2014) collected for asthma hospitilizaion.

The dataset comes as a list of individual buildings and their performance in many energy related metrics. The buildings were grouped by zip code and their metrics summed to a list of 181 zip codes with the total Site EUIs and Direct Emissions.

C. Neighborhood Income and Population

Neighborhood and Income data were collected through the American Census Bureau's website. The data comes from the American Community Survey 5-year average over the years 2011-2015. These years were selected to overlap with the asthma hospitalization rates as much as possible. The attributes collected include household median income, income per capita, and total population per zip code tabulation area.

D. Merging

The dataset was merged by zip code. As much as zip codes are not the ideal demographic tabulation unit, they were necessary in this analysis as they provided the smallest level of resolution available for asthma hosptilization rates. The American Community Survey has its own proxy for zip code knows as a Zip Code Tabulation Area. ZCTAs allow for easy merging with other zip code indexed data. After merging all datasets, and dropping any zip codes that did not have all attributes available, primarily because of low number of asthma cases, 164 zip codes remained from the original 181.

E. Mapping

An important way to understand data of this kind is to see it on a map. To map the data, a shapefile of zip code boundaries was downloaded from NYC Open Data.

IV. Results

The two primary questions this project aims to answer is is what, if any, relationship there is between building energy use and direct emissions to local asthma hospitilization rates. These results will be compared to how income affects energy use and asthma hospitilization rates.

A. Visualization

When comparing the maps of energy use and emissions, it is also not surprising to see similar patterns emerge as well. It makes sense to see Manhattan being the most energy intensive, but it is somewhat surprising to see that Manhattan also has some of the highest concentration of direct emissions.

B. Outliers

C. Asthma Discharge Rate

D. Building Energy Use

V. Discussion and Conclusion

The results above are not what one might intuitively expect. There is essentially no relationship between energy use of buildings or their direct emissions of greenhouse gases on the asthma hospitilization rates of the zip code in which they are located. Instead, a much better indicator of the severe asthma rate in a region is the median household income. This might lead one to believe that the healthcare and living conditions provided by higher wealth have a much larger impact in New York City on developing and treating asthma than environmental factors. This analysis is not comprehensive enough to make that kind of conclusion, though, as there very well may be other environmental factors that are not taken into account by the building energy dataset that cause asthma. It may also be that greater wealth clusters people in regions with lower environmental contagions, which result in the relationship between wealth and asthma. Futher analysis is needed.

It does appear that wealthier neighborhoods use less energy than those that are less wealthy, though this relationship is not particularly strong. This may be the result of new buildings built with higher efficiency HVAC systems and better insulation. Regardless, the trend is not very strong and would require further investigation to understand how higher wealth may result in lower building energy use.

As it stands, the strongest insight provided by this analysis is that the wealthier a neighborhood, the lower the number of asthma-related hospitilizations take place.

References

[1] NYC & Company, “NYC Visitation Statistics.” [Online]. Available: http://www.nycgo.com/research/nyc-statistics-page

[2] “About Us - Airbnb.” [Online]. Available: https://www.airbnb.com/about/about-us