Civic Analytics Fall 2017: Problem Identification Paper

Hao Xi, Christian Moscardi, Sarah Schoengold

Overview

A report from McKinsey Global Institute released in 2014 pointed out that the global urban residents spend more than 650 billion US dollars in the housing expenditure per year (McKinsey, 2014). New York, Tokyo, London and other global cities, are the most affected cities.

Housing issues go beyond just supply and demand. At its most basic, property cost and rents rise when demand is higher than the supply, pushing key members of the workforce — teachers, nurses, bus drivers, cleaners — into regions with longer commutes and into less-desirable areas. This has led to a well-known and well-documented urban affordability crisis across the globe.

City of NYC Context and background

Opinions abound as to why housing is so expensive in New York - many of which boil down to simple rules of supply and demand. However, a few key important factors exist. For example, one factor that we know to be correlated with housing costs is accessibility to public transportation. New York's infrastructure, without question, directly relates to the city's day-to-day operations and economic vitality. Recent subway difficulties, according to one NYT analysis, will cost the city $389 million (Hu, 2017). To quote StreetEasy's Data Science team, "Transportation will rule real estate conversation in 2017" (StreetEasy, 2016). We go into more detail regarding other factors below.

Description of challenge or opportunity

High-quality, affordable housing is one of New York's perennial challenges, both for individual residents looking for their next home and for policymakers attempting to create equitable housing policy for all New Yorkers (and probably paying obscene rent themselves). According to Street Easy's survey, rents have increased twice as fast as wages in the past seven years (Walker, 2017). According to StreetEasy, this has hit low-income earners the hardest - not only are members of this group typically the least economically secure, but rents in the bottom fifth of

New York's price range have increased disproportionately to rents in other quintiles (Walker, 2017).

Of course, this all presents a challenge. On the other hand, there's substantial opportunity to develop affordable housing, increase equity, and keep New York City humming as one of the country's most vital economic engines.

Problem statement

Affordable housing is hard to come by in New York. Furthermore, the city's least expensive private housing is often riddled with various issues - lack of access to transit, food, and quality green space are a few potential ones. Compounding this is the growing demand for more housing in the city.

We seek to better understand where New York's best - and, perhaps more important, worst - housing stock currently lies. From this analysis, we hope to provide tools and recommendations for policymakers to implement positive change to the end of providing higher-quality, affordable housing for all New Yorkers.

Conceptual approach and methods

There are many possible approaches to determining where the city's best and worst housing lies. To start, we aim to set a baseline of current affordable housing stock by the numbers. As described in the data section below below, this is mostly a data wrangling problem.

From there, we hope to develop a "score," not dissimilar to Zillow's Zestimate, based on a variety of factors that are 1) within policymakers' control and 2) of interest to residents (Zillow, ???). A few sources we've begun to consider:

- Transit Accessibility, proximity to major roads/highways, and commuter patterns
- Air Quality
- Park / green space accessibility

- Food accessibility

There are a few ways to measure the efficacy of this score. For example, we could measure how strongly these factors explain current prices in the housing market. This may give us some insight into what the market (and by extension, individual homeseekers) values in housing.

We also hope to make this score malleable - for example, if we (or policymakers) decide that air quality and transit access are the end-all-be-all for quality housing, we want to be able to identify high-and-low scoring housing along those two dimensions.

From this point, policymakers can ask a variety of interesting questions. For example, based on evolving commuter patterns, where could we augment transportation service? We hope to identify some potential interesting interventions ourselves, as well.

Description of end projects (including format)

The above methodology leads to two natural outputs:

1) An analysis / report identifying potential areas for intervention - including specific areas, important factors, and potential remedies. This analysis will have a particular focus on New York's current affordable housing stock.
2) An interactive map-based tool where one can weigh various factors and see a "heatmap" of potentially good and bad housing stock based on that user's needs. This tool could also be used by policymakers to visualize housing issues along specific dimensions.

Data types and sources

Our question requires a comprehensive look at affordable housing in the New York City area. Affordable housing is fairly well documented and accessible on NYC open data portals. To understand the complete view, we'll merge three main resources: a) PLUTO, b) subsidized housing data from Coredata.nyc, and c) rent stabilized data from Taxbills.nyc (DCP, 2017), (Furman Center, 2017) (Krauss, 2015).

The PLUTO dataset contains over seventy fields maintained by several different agencies across the city (DCP, 2017). We're particularly interested in the location and classification of residential buildings. Coredata.nyc standardizes indicators into a housing and neighborhoods datahub (Furman Center, 2017). Using their online tools, we can extract data ranging from properties subsidised by federal agencies, like HUD, to state agencies, like NYS Division of Housing and Community Renewal. Finally, taxbills.nyc is a database that shows where landlords receive tax breaks from the city for rent stabilization programs (Krauss, 2015).

After compiling a dataset that provides an overview of housing costs in the city, we'll focus on datasets that describe factors on which housing prices may depend. The factors we're initially most interested in include: transit accessibility and commuter patterns, air quality, park/green space availability, and food accessibility. There are several different approaches when looking at transit accessibility and its relationship to housing prices. Some existing datasets we know we want to work with include:

- TLC data
- MTA data
- Walk Score
- NYC bus data

(NYC Open Data, 2017), (NYC Open Data, 2017), (The Open Bus, 2017), (Walk Score, 2017). In addition to these available datasets, we'll look at housing areas' proximity to major roads and highways using The NYC Street Centerline (CSCL), which contains data from NYC streets, traffic directions, and road types (NYC Open Data, 2017).

From these public datasets, several studies have been done to infer commuter patterns in the city. Perhaps most relevant to our questions, the U.S. Census Bureau outputs county-to-county worker flow files on a quarterly basis which can help us understand the movement of New York populations (US Census Bureau, 2016). Additionally, the Federal Reserve Bank of New York

published a study in 2005 commenting on the evolution of commuter patterns in New York (Federal Reserve, 2005). While the study only looks at data at the county level, their analysis may help us hone in on trends born in the early 2000s that persist today. For example, the study identified that the New York population had a "reduced sensitivity to long commuting distances over time" (Bram, McKay, 2005).This trend was established before disruptive events such as ridesharing technologies were mainstream or knowledge of impending disruptions like autonomous vehicles or L-Train closure. With 2010 census data, we could potentially replicate aspects of these studies to see if these trends hold true.

As for our secondary factors, both air quality data and green space accessibility are readily available on Open Data NYC. The Air Quality data set include surveillance data from sensors around the city that record levels of toxins in the air (NYC Open Data, 2017). The Green Spaces data set is a basemap that looks at not only parks, but also courts, cemeteries and other open areas throughout the New York metro area (NYC Open Data, 2017). Finally, we can use Retail FoodStores data set maintained by data.ny.gov to determine proximity of housing to local food options (NY State Data, 2017).

Data access and availability

The data for both regarding housing costs and public transportation is widely available online through NYC Open Data. The areas in which we may encounter difficulties are incorporating features that are more qualitative by nature. For example, we'd like to consider some informal transportation methods in our analysis — such as Dollar Van ridership for commuter patterns — but understand that they may be difficult to embrace in a meaningful way due to a lack of standardized (and digitally available) records.

In addition to undocumented transportation routes, food accessibility may be particularly difficult to include in our analysis. While the Retail Food Store data set does give us an understanding of where food is available in certain areas, it doesn't inform us *what kind* of food

is available. It's difficult to parse out what is a bodega and what is a farmers market. Others have studied food deserts in New York and have come across the same issues (Seegers, 2015).

Legal and privacy considerations

As with any data project, we want to be careful to ensure individuals remain anonymous while looking at housing data. Since the datasets we're using are largely from open data sources, we're not too concerned about exposing anything sensitive with what we're considering in our analysis currently.

Works Cited

Bram, J., & McKay, A. (2005). The Evolution of Commuting Patterns in the New York City Metro Area . Federal Reserve Bank of New York, 11(10). Retrieved October 2, 2017, from https://www.newyorkfed.org/medialibrary/media/research/current_issues/ci11-10.pdf.

Woetzel, J., Ram, S., Mischke, J., Garemo, N., & Sankhe, S. (2014, October). A blueprint for addressing the global affordable housing challenge (Rep.). Retrieved October 2, 2017, from

McKinsey Global Institute website: http://globalhousingindicators.org/sites/globalhousingindicators.org/files/McKinsey%20Global%20Institute%20Full%20Report.pdf

StreetEasy. (2016, December 21). StreetEasy's 2017 New York Housing Market Predictions [Web log post]. Retrieved October 2, 2017, from https://streeteasy.com/blog/2017-nyc-housing-predictions/

Curbed. (2017, August 16) Rents are increasing twice as fast as wages in New York. Alice Walker. [Web log post]. Retrieved October 2, 2017, from https://ny.curbed.com/2017/8/16/16154956/nyc-rent-prices-wage-increase-comparison

Seegers, M. (2015, April 24). No Food Deserts? Mapping Retail Food Stores in NYC. Retrieved October 2, 2017, from http://research.prattsils.org/blog/coursework/information-visualization/no-food-deserts-mapping-retail-food-stores-in-nyc/

Zillow, I. (??). What is a Zestimate? Zillow's Zestimate Accuracy. Retrieved October 02, 2017, from https://www.zillow.com/zestimate/#wha

Hu, W. (2017, October 01). New York's Subway Delays Could Cost City $389 Million a Year. Retrieved October 02, 2017, from https://www.nytimes.com/2017/10/01/nyregion/new-yorks-subway-delays-could-cost-city-389-million-a-year.html?_r=0

Data Sources Cited

TLC Trip Record Data. (n.d.). Retrieved October 02, 2017, from http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

NYC Open Data. (2017). NYC Street Centerline (CSCL). Retrieved October 2, 2017, from https://data.cityofnewyork.us/City-Government/NYC-Street-Centerline-CSCL-/exjm-f27b/data

NYC Open Data. (2017). MTA Turnstile Data. Retrieved October 2, 2017, from http://web.mta.info/developers/turnstile.html

NYC Open Data. (2017). Green Spaces Data. Retrieved October 2, 2017, from https://data.cityofnewyork.us/Recreation/Green-Spaces/mwfu-376i/data

NYC Open Data. (2017). Air Quality Data. Retrieved October 2, 2017, from https://data.cityofnewyork.us/Environment/Air-Quality/c3uy-2p5r

NY State data.ny.gov. (2017). Retail Food Stores. Retrieved October 2, 2017, from https://data.ny.gov/Economic-Development/Retail-Food-Stores/9a8c-vfzj

United States Census Bureau: County-To-County Worker Flow Files. (2016, September 28). Retrieved October 2, 2017, from https://www.census.gov/data/tables/2000/dec/county-to-county-worker-flow-files.html

The Open Bus. (2017). New York City Buses. Retrieved October 2, 2017, from http://www.theopenbus.com/nyc-bus.html

DCP, N. (2016, September). PLUTO Dataset. Retrieved October 02, 2017, from https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page

Furman Center, N. (2017). CoreData.nyc. Retrieved October 02, 2017, from http://coredata.nyc/

Krauss, J. (2015, July 1). Whither Rent Regulation. Retrieved October 02, 2017, from http://blog.johnkrauss.com/where-is-decontrol/

Walk Score. (2017). New York Walk Score. Retrieved October 2, 2017, from https://www.walkscore.com/NY/New_York