Boston Neighborhood Livability

by Daren McCulley and Jasper Burns

Rent & Assessment Study

Introduction

In an effort to characterize Boston neighborhoods we decided to take a closer look at the available data on the rental market and combine it with the property assessment data made public by the city. The city's assessing website describes three approaches to determining an assessed value.[1] Our focus is on the Income Approach, which employs data to determine what a property might earn. One of our main objectives is to reduce the uncertainty inherent in the word might. In a city like Boston, where nearly 40% of all households rent and rental vacancy has decreased year over year for the past decade towards 3%, the data necessary to support the Income Approach is widely available.[2]

According to a Boston Redevelopment Authority report, more than half of the city's income comes from property taxes, and property taxes account for over 80% of the money Boston is capable of raising on its own, i.e. not coming from the state.[3] With over 50% of Boston's acreage enjoying a tax-exempt status, Boston needs a property tax policy grounded in the data-centric Income Approach to assessing. Not just to ensure that everyone pays their fair share, but to help guide decisions about where to invest Boston's limited resources in order to maintain the housing diversity that is essential to vital neighborhoods. Understanding when, where, and how a neighborhood is gentrifying may allow for affordable housing initiatives before they become financially or politically intractable to enact.

Methods

Web-based real estate companies like Zillow offer rent *estimates* mapped to specific addresses. However, mining this data without their consent violates their API's terms of use. Our rental data was sourced from the popular apartment hunting website PadMapper.[4] Rather than aggregate the rental data in an attempt to reproduce what Zillow (et al.) already provide we attempted something more ambitious. Since assessments are the foundation for property taxes the dataset is both detailed and fine-grained enough to attempt to join with another dataset at per unit level. We used address as the common attribute to tie rental data to assessment data,

---

[1] http://www.cityofboston.gov/assessing/default.asp

[2] http://www.deptofnumbers.com/rent/massachusetts/boston/

[3] http://www.bostonredevelopmentauthority.org/getattachment/f0ce585f-6943-4cfe-9ac0-20126ad5f14c

[4] http://www.padmapper.com/

necessitating a collapse of all records associated with that location into a single assessed value per unit (for assessment data) or an average advertised rent (for PadMapper data).

Results

Our results show some interesting patterns. In general, for the high-end neighborhoods (e.g. Back Bay and South End) rental prices track with assessed values leading to data that presents linearly with a positive slope. However, in developing neighborhoods (e.g. Brighton and East Boston) the data is grouped into two clusters. Very few of the properties are assessed for more than $500,000 per unit and roughly half of them rent for more than $3,000 per unit. For certain this clustering could be an artifact of PadMapper or the way the data was aggregated for each location, however, if that isn't the case than it suggests that developers and landlords in Allston, Brighton, and East Boston enjoy a much better rent to assessed value ratio than landlords in South End.

It is important to understand that the PadMapper data is a single snapshot of the rental market captured at the time the scripts were run. An aggregated dataset, especially one that captures the high season for rental turnovers, is more appropriate for this analysis.

Future Work

Our hope is that by joining these datasets on a per unit basis, policy makers will have the means to ask more interesting questions based on data that would otherwise be obscured by painting with a broader brush. For example, how do developer owned properties compare with privately owned ones? Would looking at the same data over time reveal a trend that in part explains the way neighborhoods like West Roxbury and South End have developed? In what corners of Boston is the goal of the Income Approach seemingly out of balance with the ratio of assessment to rental income? Finally, for anyone subscribing to Neil Smith's Rent-Gap Theory the ability to accurately plot assessment data against rent prices is an important indicator of gentrification.[5]

While the assessment data is by definition complete, up-to-date, detailed, and fine-grained, PadMapper is just the opposite. Nothing prevents brokers from listing the same apartment dozens of times at different rents or fabricating/omitting the key piece of information, the address. Additionally, listing price is not necessarily the eventual rent agreed to by the future tenant. However, any person or business earning rental income in the state of Massachusetts must fill out a Schedule E declaring what that income was for each property each year. The software we've developed to scrape, scrub, and aggregate a low fidelity data source like PadMapper could easily be used with a high fidelity source like tax records to provide a complete and accurate picture of Boston's rent-to-assessment landscape based on tying earned rent to assessed value at a per unit level.

---

[5] N. Smith, Annals of the Association of American Geographers, Vol. 77, No. 3 (Sep., 1987), pp. 462-465

The other untapped wealth of data comes from the fact that Boston must have assessment data going back to 1985. Their assessing website shows the complete assessment history for any parcel going back to 1985 or whenever the parcel was first assessed if after 1985. For this study, we only had access to the 2014 and 2015 assessment data through the Socrata API. However, armed with 30 years of data the possibilities to explore that data for patterns that help explain how Boston developed and the direction the city is headed are endless.

Vitality & Composition Study

Introduction

While the first study focused on where Boston had sensible rents, this study focuses on where those rents are worthwhile, i.e. how vibrant each neighborhood in the city is. This study draws inspiration from sociologist Jane Jacobs's book, *The Death and Life of Great American Cities*. Jacobs concluded that a vibrant city life only happens in physically diverse environment, and one of her conditions to meeting this diversity was that city districts must serve more than two functions. So this is the hypothesis we tried to test, namely that the vitality of a neighborhood correlates with the diversity of its environment.

Methods

Since Boston's data was most frequently organized by zip code (rather than by a neighborhood), our analysis was on a zip code basis. By analyzing Boston's 2015 Property Assessment Data, we were able to assign each zip code a diversity score, which we said was equal to the percent commercial of a neighborhood (Commercial Units / (Commercial Units + Residential Units)).

We then had to define a vitality score for the y-axis. We define the vitality as the average reviews per business for the top ranked 300 businesses in that zip code in Yelp's API.[6] The thinking here was that vitality implies a lot of visitation to local businesses, and the more visitations implies more people writing reviews. By making it an average *per business*, this helps normalize for the fact that zip codes with greater percent commercial would have had more total reviews.

Results

The results were underwhelming. We had a positive correlation which was good, but our r^2 was pretty weak at 0.283. The mean diversity score (commercial property prevalence) was 7.5%, with a standard deviation of 5%. Our average vitality score was 117.2, with a standard deviation of 96.7.

---

[6] https://www.yelp.com/developers/documentation/v2/search_api

Looking at the graph, we can see that between 2-12% commercial, there is essentially no correlation with vitality, but once neighborhoods have >12% commercial properties, they never have a vitality score below the mean. We surmise that there might be a stronger correlation with better metrics used to for measuring both axes, outlined below.

Future Work

This type of analysis has been done before to greater effect by using more granular data. Several researchers at the University of Trento developed a similar but more sophisticated experiment for a number of cities in Italy.[7] Inspiration from their project to draw upon for future work on this study could be to include the ratio of high and low income renters into the diversity score, as Jacobs's book also had the condition that "buildings must be diverse in terms of age and form to support a mix of low-rent and high-rent tenants."[8]

Besides amplifying the diversity score, if given greater access to more revealing datasets, one could rewrite the vitality function in the script to further qualify what it means to be of "high vitality." The Italian study used information Foursquare data and mobile-phone records showing the number of calls in an area to build its vitality. If given access to more high fidelity sources like these, or perhaps the Twitter fire hose, one could a more accurate vitality rating.

---

[7] https://www.technologyreview.com/s/601107/data-mining-reveals-the-four-urban-conditions-that-create-vibrant-city-life/
[8] Ibid.