Deliverable 2: Extension Project Proposal

Ciaran Hikaru Ueda Fitzgerald
Boston University
cueda@bu.edu

Haoran Hu
Boston University
hhran@bu.edu

Daniel Scrivener
Boston University
dscriv@bu.edu

Julie Le
Boston University
tjle@bu.edu

1 Summary

From the viewpoint of a prospective tenant, some affordable housing options are likely to appear more desirable than others. Up to this point, our analysis has been agnostic to desirability; by incorporating this metric, we seek a more holistic understanding of the Boston affordable housing market as viewed by the target population. Although the attractiveness of any given property is not quantifiable by any objective standard, records of past property incidents and landlord violations serve as a natural starting point for estimating unit desirability. Crucially, this data is accessible to us, and can be cross-referenced with our existing affordable housing datasets. Our confidence in the effectiveness of this measure is based on the hypothesis that prospective tenants' decisions are likely to be influenced by feedback from the surrounding community. To that end, official records of landlord violations provide a reasonable data-driven approximation of the community's view of a particular property or landlord.

2 Motivation

From the viewpoint of a prospective tenant, some affordable housing options are likely to appear more desirable than others. Up to this point, our analysis has been agnostic to desirability; by incorporating this metric, we seek a more holistic understanding of the Boston affordable housing market as viewed by the target population. Although the attractiveness of any given property is not quantifiable by any objective standard, records of past property incidents and landlord violations serve as a natural starting point for estimating unit desirability. Crucially, this data is accessible to us, and can be cross-referenced with our existing affordable housing datasets. Our confidence in the effectiveness of this measure is based on the hypothesis that prospective tenants' decisions are likely to be influenced by feedback from the surrounding community. To that end, official records of landlord violations provide a reasonable data-driven approximation of the community's view of a particular property or landlord.

3 Questions for Analysis

1. At any given residential address, what is the relative likelihood that the property will be subject to a future property violation (based on prior landlord violations at or around that location?)
2. Is there any correlation between neighborhood density of property violations and neighborhood density of property violations? If so, how might that affect the options available to prospective tenants seeking affordable housing?
3. (Stretch) Based on the timeframe presented by the data, are there any emergent trends in which neighborhoods seem to be accruing more or fewer property violations?

4 Data Sets Additional Sources

The *Analyze Boston: Building and Property Violations* dataset inspired this extension project and will serve as our primary source of information. This dataset is particularly convenient, as it readily provides the necessary metrics for cross-referencing entries between other datasets: namely, latitude and longitude in addition to street addresses with consistent formatting. In other words, there will be a limited need for time-intensive library functions (such as those provided by Geopy and Google Cloud) in preprocessing the dataset. We will also be able to maximize our use of entries without discarding those for which the relevant identifying geographic information cannot be found. Additionally, this dataset is regularly updated with recent information on property violations, strengthening the coherency of our assumption that the data reflects current community viewpoints.

5 Visualizations

We aim to visualize desirability as a property of Boston neighborhoods as well as the individual units they comprise. As such, a heatmap will be generated to show broader trends in desirability over a large geographic space. Histograms broken down by neighborhood/ZIP code facilitate analysis of this information. It may also be possible to assign a unique identifier to each landlord and plot individual violation counts, giving us a sense of potential "bad actors" who are more likely to commit property violations in the future. As a stretch goal, a web-based front end where users can query the desirability of properties — either as a location-based approximation, or as a lookup of known properties in the database — poses an interesting application of the analysis.