Rating neighborhoods to find optimal place to raise children in Boston

By Abdullah Babgi, Yan Spinelli, Matthew Pinheiro, Hester van der Laan

CS 591 Data Mechanics Spring 2017

Professor Lapets

Boston University

Introduction

In quick developing metropolitans, like Boston, it becomes harder to find an affordable place to raise kids. Housing prices continue to increase and not every neighborhood is the right environment for a kid to grow up and prepare for a successful college career. This project aims to answer the question, "If a family with presumably one or two children who are preparing to go to college were to move to Boston, what would be the most ideal area to move to?" To rate the areas in Boston to answer the posed question, we obtained and scraped data from multiple online sources that included Boston and Cambridge housing rates, crime rates, population by age, corresponding neighborhoods to zip codes and education information. The dataset can be pulled from datamechanics.io with the links below:

- `http://datamechanics.io/data/jspinell_mpinheir/AgeRanges.json`
- `http://datamechanics.io/data/jspinell_mpinheir/CrimeRate.json`
- `http://datamechanics.io/data/jspinell_mpinheir/EducationCosts.json`
- `http://datamechanics.io/data/jspinell_mpinheir/HousingRates.json`
- `http://datamechanics.io/data/jspinell_mpinheir/neighborhoods.json`

These datasets were derived from bestplaces, zipatlas, and clrsearch:

- `http://www.bestplaces.net/transportation/zip-code/massachusetts/cambridge/02141`
- `http://zipatlas.com/`
- `http://www.clrsearch.com/Boston-Demographics/MA/02215/Crime-Rate`

Methods

All of the datasets are organized by zip code, which allowed the creation of 3 datasets that found correlations between 2-bedroom apartment rates, crime rates, college graduation an population by age. To transform the data tiers were created of housing prices and mapped averages of crime data, graduation rates and populations of ages 15-19 for those respective areas. Initially, the housing prices were broken up into tiers for "average Rent 2 bedroom", based on zipcodes. The crime score for Boston and Cambridge were normalized and assigned by tier. A similar approach was used for the

graduation rate data. A MapReduce algorithm was used to average graduation rates per tier, to create education rates per tier. Finally, to indicate what the average age of the population of 15-19 year olds was another MapReduce was used and AgeByTier was created.

For the second part of the project statistical analysis was performed to check correlation between housing prices and crime rates, college graduation rates, and the age of the population.

Lastly, Prices of houses in a given area were predicted; using linear regression values for graduation rates, age by population and crime rates. The results were compared with the value of the houses against the actual prices of the houses to determine whether or not the price reflected the value of the house. This data can be used to advise families that are looking into buying a house in Boston. The results tell whether or not the house they are looking at is a good buy for the asking price.
A limitation of the data, which simplified our job, but additionally is a restriction to the precise location of the house, is that all the data is organized by zip code. Within a single zip code, one can find very different looking 2 bedroom apartments, generalizing by zip code is thus a limitation. Furthermore, crime rate is similarly very generalized data. A family moving to Boston might be interested in violence on the streets, robbery, etc., crime that influences their child safety, rather than for example crime like fraud.

Results

The scores of average graduation rate, crime rate and average population of the age 15-19 per zip code were visualized in a map of Boston. Figure (1)-(3) display their scores using a color scale.

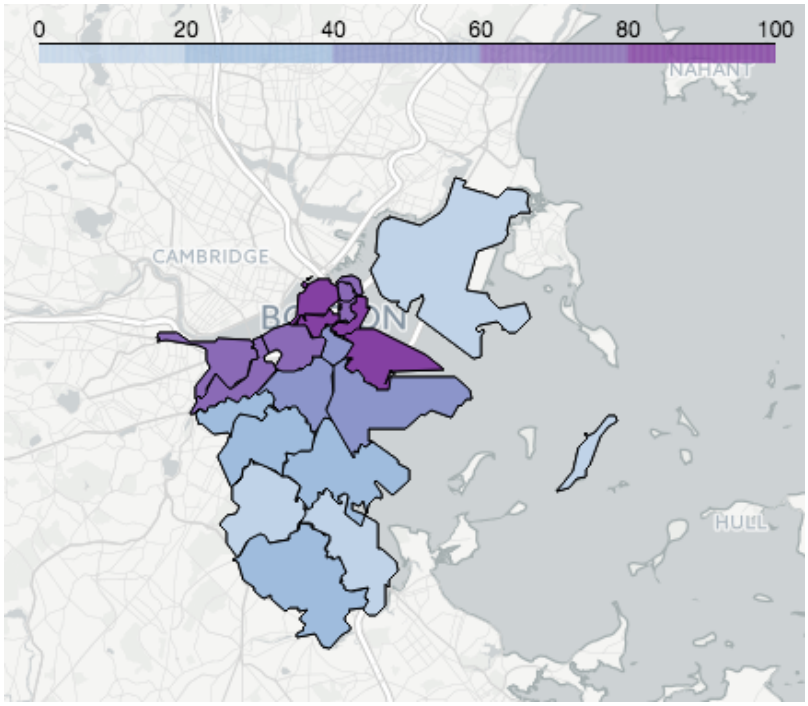

Figure (1): Average Graduation Score per Tier

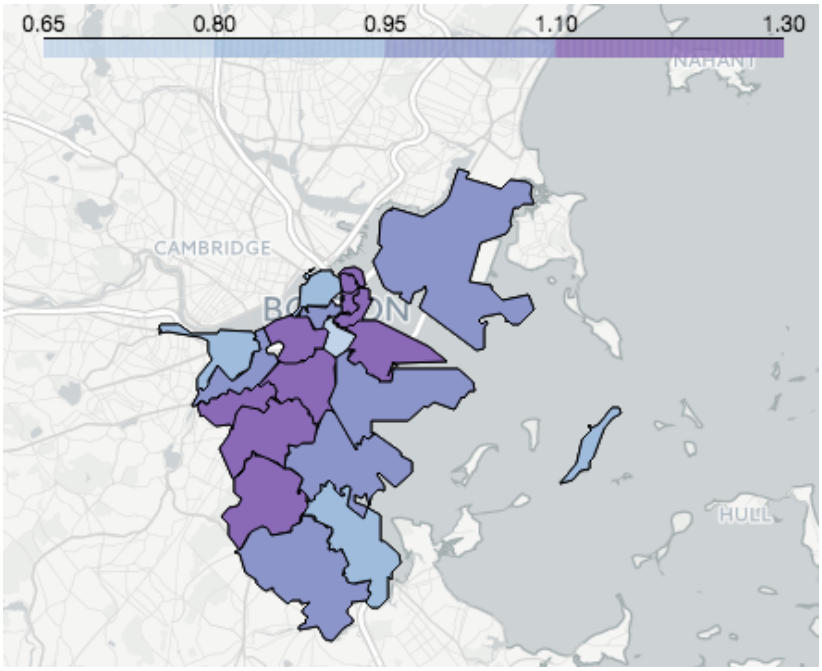

Figure (2): Crime Rate Score per Tier

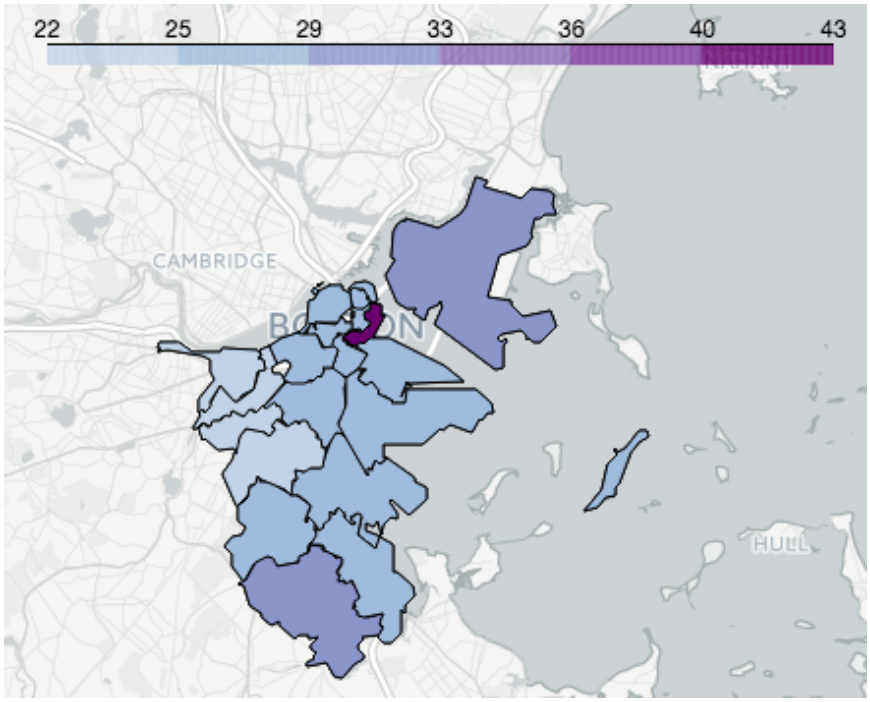

Figure (3): Average Population of the age 15-19 per Tier

Lastly, figure (4) displays the housing prices per zip code.

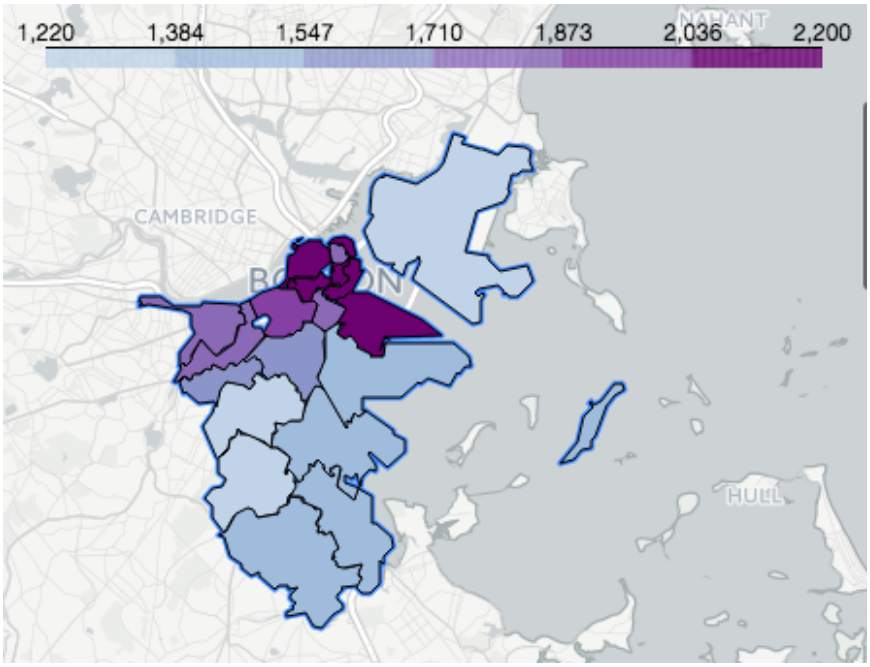

Figure (4): Housing prices per zip code

The correlation coefficient of crime rates and housing prices was surprisingly

low(-0.11), as well as the age population and housing prices (0.24). This indicates crime rates barely correlate to housing prices, neither does the age population. The college graduation strongly correlates positively to the housing prices (0.87). Figure (5) visualizes the difference in correlation values.

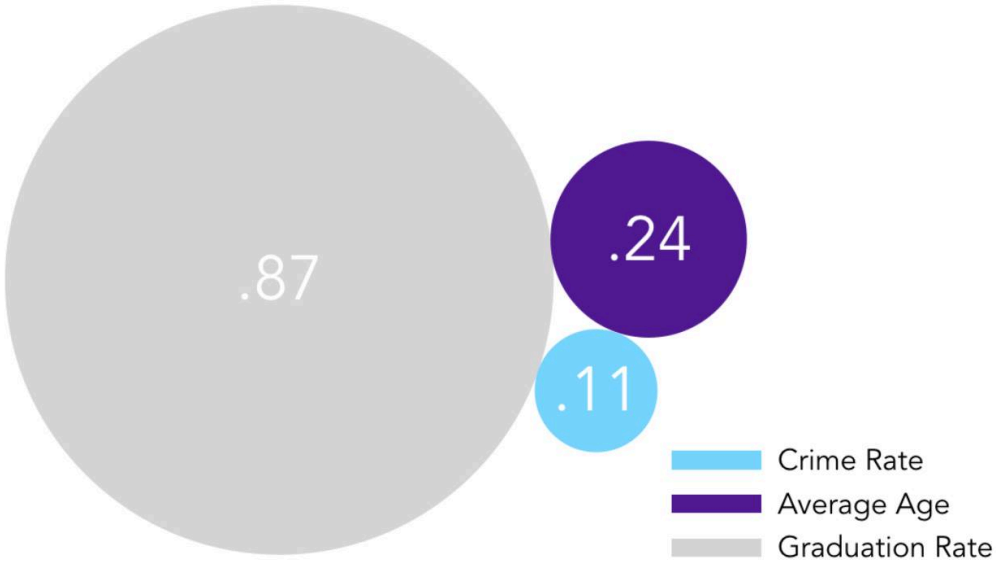

Figure (5): Correlation between housing prices and crime rate, average age, graduation rate

Linear regression was performed on each attribute to predict housing prices per zip code; graduation rate, average age and graduation rate. Only the graduation rate score had a linear regression with housing prices as expected from the correlation value. Figure (6) shows the linear regression between graduation rate and housing price.

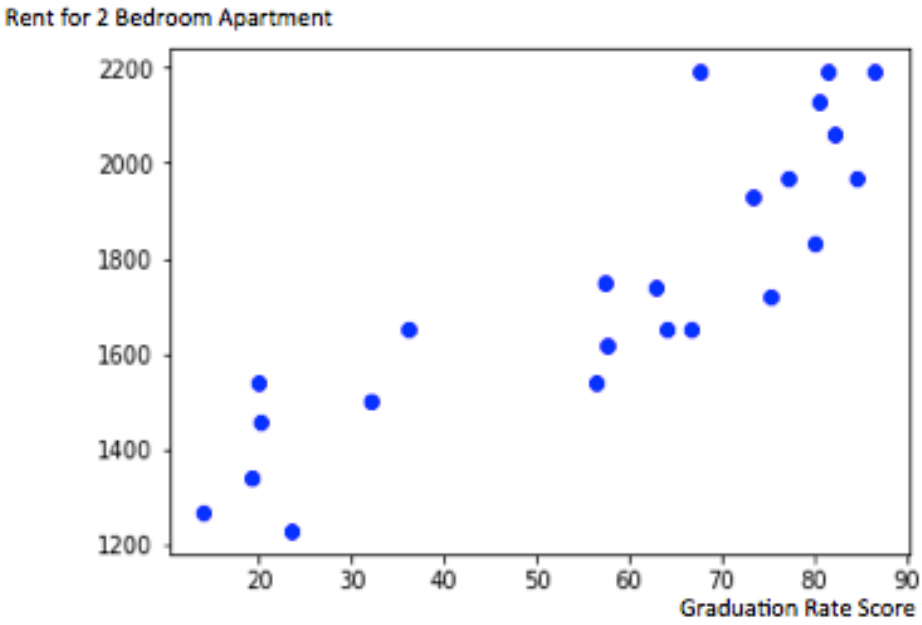

Figure (6): Linear Regression between graduation rate and housing prices

The predicted housing prices are slightly different then the actual housing prices. However, depending on the neighborhood the houses are rated higher or lower than the actual housing price inconsistently. Figure (7) displays the actual and the predicted housing prices per zip code.

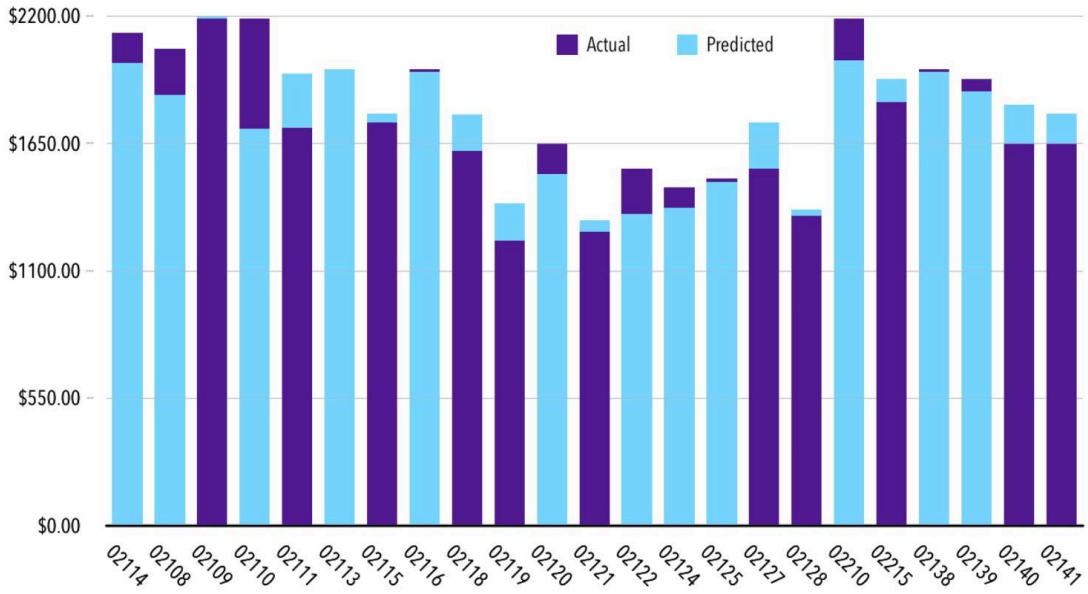

Figure (7): Predicted values versus actual values of housing prices

Conclusion

Using the figures in the result section a family moving to Boston would be able to figure out the crime rate, graduation rate, and average age of each zip code. Additionally, they could determine whether the house they are looking at, is overpriced, underpriced or correctly valued. This data simplifies the complicated decision when moving into a new metropolitan. Interestingly enough, our analysis pointed out that crime rates do not correlate to housing prices, an unexpected result. The actual prices versus the predicted values seemed to vary randomly. For future projects one could try to find a better method of predicting housing prices, using different attributes, such as location, age of the buildings, or other.