Boston Neighborhood Analysis

John Hennessey
May 22, 2020

I.I Introduction

This study will inspect and compare types of venues in different neighborhoods throughout Boston. It will result in venue data clustered by neighborhood. Graphs will display where a certain cuisine is most popular. The top venues in each neighborhood will also be considered and displayed.

I.II. Problem

How does a person decide where to live when they move to a new city? For this project I will assume that types of venues indicate types of culture and that a neighborhood with a person's favorite type of venues will be the best place for them to live. To solve this problem, I will analyze and group venue data by neighborhood and present each neighborhood's top cuisine choices.

I.III. Discussion

When someone decides to move to Boston, how will they know where to live? Most people who move to a new city attempt to live close to their job. Besides this requirement though, many who do not have a familiarity with the different boroughs and neighborhoods find themselves lost. After moving to a certain neighborhood, many people end up spending most of their time in a different neighborhood as it appeals to their interests and most importantly, it contains their favorite venues.

II. Data Acquisition and Cleaning

Unfortunately, I could not find all the necessary data in one place. I had to create an excel file containing the different neighborhoods of Boston and their coordinates which I found on Google. These coordinates were linked to zip codes, however some neighborhoods had multiple zip codes and some zip codes contained multiple neighborhoods. In some cases, I had to estimate the coordinates for a given neighborhood. I estimated coordinates for these neighborhoods which, when mapped, were close enough for the sake of visualization and venue prevalence estimation.

Using this excel file, I input the coordinates for each neighborhood into Foursquare to attain different venues within 1000m from the center of each neighborhood. The Foursquare Data combined with the Neighborhood Data offers insight into the venue options in each neighborhood. Using this information in combination with an individual's venue preferences, I can determine the best place for an individual to live using a model.

III. Exploratory Data Analysis

The first data exploring phase required learning about venue types. I downloaded an excel file with all of the venue types and how many of each type were found. I did the same for number of venues per neighborhood to gain more insight into the general amount of opportunity available in each neighborhood. The most popular venues through out Boston are Pizza Place, Park, Coffee Shop, and Italian Restaurant occurring 96, 88, 86, and 80 times respectively. Foursquare capped my venues per neighborhood at 100 and many neighborhoods hit this threshold. North Cambridge, Roxbury, and West Cambridge had the least amount of venues with 27, 27, and 26 respectively while neighborhoods close to and including Downtown all hit the 100 venue threshold. Venues in each neighborhood were organized by type. One hot encoding was performed and the top venues for each neighborhood were placed in a table. These visualizations give some insight into the differences between neighborhoods in Boston. More analysis is required (and detailed in the next section) to gain actionable and clear results showing the best neighborhoods for a person given their venue category criterion.

IV. Predictive Modeling

Ultimately, I simplified this study to accept the following user inputs: five most important venues to have in the neighborhood including one “top priority” venue. Hopefully, a neighborhood exists that will have at least one instance of each of these five venues and many instances of the “top priority” venue. In order to determine if a neighborhood satisfies the “top priority” venue portion, I will check where the venue is ranked in each neighborhood. If it falls in the top 3 the requirement is satisfied (score = 1), top 5 (score = .8), top 10 (score = .6), top 20 (score = .4), outside of the top 20 (score = 0). Combining these scores with the proportion of the five venues found in the neighborhood and dividing this number by 2 will give a neighborhood ranking between 0-1. Functions were created to score each neighborhood. These rankings will make it easier to visualize how suited each neighborhood is to a person’s lifestyle.

I asked a friend to give me his top 5 venues which were 'Italian', 'Pizza', 'Ice Cream', 'Gym', and 'Coffee' with his top choice being ‘Italian’. Venue Types that match his criteria are 'Italian Restaurant', 'Pizza Place', 'Ice Cream Shop', 'Gym', 'Gym / Fitness Center', 'Coffee Shop'. Back Bay and Assembly Square both had all 5 venue types in their top 20, but neither had ‘Italian Restaurant’ in their top 3. However, the South End scored better than these neighborhoods because it has the top criterion ‘Italian Restaurant’ as its most prevalent venue category. It is interesting to note that these neighborhoods are all close enough to have overlapping venues. This make sense especially if key venues (such as Italian Restaurants) appear at the overlaps but that is out of the scope of this project.

V. Conclusion

It is often difficult to determine the best neighborhood to live in when moving to a new city on your own. Venue data when organized by neighborhood, grants insights which can be helpful when making such a decision. Initial visualizations give more insight into neighborhood trends, but a more refined model is required to gain actionable insights into which neighborhood a person should move to. A person’s venue interests can be prioritized and compared to the

venues available in each neighborhood in order to make a data driven decision. In this example, with this person's criterion, this person should look at moving to the South End. If they have other requirements such as distance to work that might affect their decision, they could look at the other top neighborhoods recommended to them such as the Back Bay and Beacon Hill. Regardless, this model helps a person to focus their apartment search on a few neighborhoods and prevents them from getting overwhelmed by the plethora of options.