Analysing Residential Neighborhoods in Boston

Applied Data Science: Capstone Project Report

Sudhir Nair, May 2021

Introduction

Moving to any new city can be quite daunting. Finding a place to live is among the many challenges that newcomers to a city face. It starts with identifying the right area to search for a house. There can be multiple factors that drive this decision but generally people would prefer a neighborhood that allows them access to most amenities while being affordable in terms on the rent. This selection can get quite complicated when one has very little understanding about the new city. My project is attempting to layer in data science to help make this decision making easier. Boston has been considered as the target city in this project but the principles can be replicated for any city of choice.

Boston - The City of Neighborhoods

Boston is a diverse city and the capital of the state of Massachusetts in US. The city covers 48.4 square miles (125 km2) with an estimated population of 692,600 in 2019. Boston is sometimes called a "city of neighborhoods"

because of the profusion of diverse subsections; the city government's Office of Neighborhood Services has officially designated 23 neighborhoods. In this project we will be using the neighborhood definition as per Analyze Boston database.

In this project we will start with ranking of neighborhoods in Boston based on the rental prices and then layer in data of venues in the surrounding to build a blended scoring matrix that allows us to see which neighborhoods offer advantage of having numerous venues in the vicinity while still being affordable.

This analysis will be of interest to anyone looking to move to a new city as it'll help shortlist potential areas that can be considered for house-hunting. Apart from potential residents, this analysis can also be of interest to real estate developers as it could highlight areas where people are interested in living and also areas which are underserved by amenities.

Data

The key to a robust analysis is obtaining reliable data for all the parameters. For this project we will be leveraging data for a diverse set of sources and then integrating them to allow us to see the full picture. We will be referencing the following data sources:

- Geospatial data and neighborhood classification from the Analyze Boston website. Analyze Boston is the City of Boston's open data hub managed by Citywide Analytics Team. Specifically we will be using the Boston Neighborhoods geojson file which will provide the name, polygon coordinates and area for each Boston neighborhood.

- Rental prices in each neighborhood in Boston from Real Estate Boston. This website uses data from the rental website Zumper to present the median rental price for a 1BR apartment in each neighborhood of Boston. The data is presented for multiple time periods but we will be using the latest data (Winter 2020) in the analysis.

- Foursquare Places API to pull the list of venues in each neighborhood. This API returns a set of venues based on the latitude and longitude queried along with other variables such as venue category, radius of search etc.