Housing Search in Massachusetts Supplemented by Foursquare Venue Data

IBM Applied Data Science Capstone, Coursera

Chris Grace

May 21, 2021

# Introduction

In the search for a new home, it is generally easy to identify various characteristics such as square footage or number of bedrooms and search for properties using well-known sites such as Zillow.com, Realtor.com, or Redfin.com. However, if you desire to investigate nearby amenities of each neighborhood or city, that typically involves looking up each address individually to judge suitability. This project aims to streamline that process to narrow down the number of properties which will need to be investigated.

Our clients are homeowners who came to our real estate office in Massachusetts. They would like to relocate to a larger house due to a growing family. They like the nearby amenities of their current neighborhood and would like to find a house in a neighborhood which offers similar nearby amenities. Due to limited houses on the market, they also don't want to limit themselves to their current neighborhood. Therefore, we will provide a means to input search parameters for the house, such as cost, square footage, number of bedrooms, etc. in a specified county. Next, each returned address will be used as the center point of an amenity search to determine if the required amenities are within the required distance for that amenity.

The filtered list which meets the required criteria should output relevant summary information which may help to narrow houses down further such as price, square footage, number of bedrooms, and names and proximity to required amenities.

The required amenities are:

|  |  |  |  |
| --- | --- | --- | --- |
| **Amenity** | **Distance** | **Amenity** | **Distance** |
| Park | 0.5 miles | Train Station | 2 miles |
| Grocery Store | 2 miles | Golf Course | 4 miles |
| Fire Station | 2 miles | Emergency Room | 5 miles |
| Police Station | 2 miles |  |  |

While this project is being developed for a single client, this situation is encountered enough that this script is being developed to help future clients and to be used as a marketing differentiator from other real estate offices.

# Data

This project will use a few sources of data.

* **Redfin API**

Redfin has an unofficial API which can be used to create a csv file of houses which meet criteria. The process to download a csv file with the API is briefly described here: <https://support.redfin.com/hc/en-us/articles/360016476931-Downloading-Data>. When the link for this option is examined, it can be seen that the parameters can be altered to produce a csv file of our choosing. An example of an API search query is below.

https://www.redfin.com/stingray/api/gis-csv?al=1  
&market=boston&max\_listing\_approx\_size=3500  
&max\_price=1000000&min\_listing\_approx\_size=2000  
&min\_price=500000&min\_stories=1&num\_beds=3  
&num\_homes=350&ord=redfin-recommended-asc  
&page\_number=1&region\_id=1338  
&region\_type=5&sf=1,5,6,7&status=1  
&time\_on\_market\_range=30-&uipt=1&v=8

This will return houses:  
-From 500,000 to 1,000,000 USD  
-From 2,000 to 3,500 square feet  
-Single Family  
-In Essex County, MA (region\_id=1338)  
-At least 3 bedrooms  
-Less than 30 days on market

It is important to note that redfin will only return the top 350 results in the csv file. Therefore, it will be important to make sure the filtering parameters are appropriate to return a number below this threshold.

* **Foursquare API** (<https://developer.foursquare.com/>)

Foursquare has an API which enables the user to query nearby venues of interest when the latitude and longitude of an address is entered. We will use the "Venue Search" call. The information for this API can be found here: <https://developer.foursquare.com/docs/api-reference/venues/search/>. It is important to note that the free unverified API only allows 950 search calls a day. Therefore, having a list of 100 houses to search for 7 distinct amenities will result in 700 calls to the API. This makes it important to narrow down the list of homes searched by using well selected criteria. If this is an issue, the number of calls can be increased to 99,500 per day by verifying the account with a credit card.

# Methodology

The way this data will be processed is:

1. Define current residence address and lookup geolocation.

The client currently lives at:



1. Retrieve data from Redfin based on desired house attributes.

The desired house attributes are:

* Max Price: $800,000
* Min Price: $649,000
* Max Square Footage: 3,500 sq ft
* Min Square Footage: 2,000 sq ft
* Min Bedrooms: 3
* County: Essex County, Massachusetts
  + When looked up on Redfin this maps to region ID 1338
* Single Family Home

These attributes resulted in the following call to the Redfin API:

<https://www.redfin.com/stingray/api/gis-csv?al=1&market=boston&max_listing_approx_size=3500&max_price=800000&min_listing_approx_size=2000&min_price=649000&min_stories=1&num_beds=3&num_homes=350&ord=redfin-recommended-asc&page_number=1&region_id=1338&region_type=5&sf=1,5,6,7&status=1&time_on_market_range=30-&uipt=1&v=8>

This query returned 39 properties. The full list as a csv file is available in Appendix A.

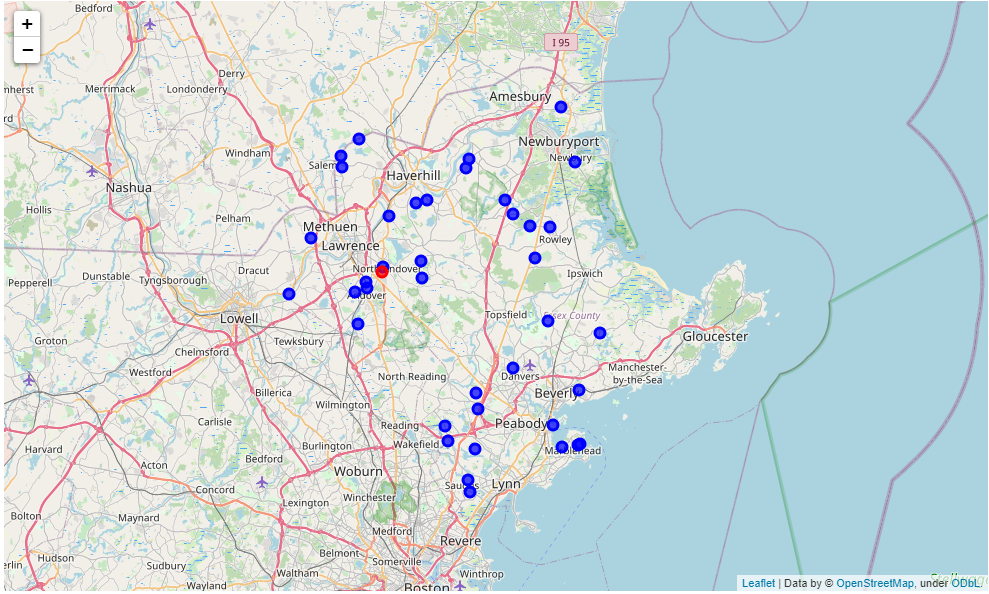


Figure 1: Redfin Property Query, All Returned Properties

1. Define desired neighborhood amenities and required distance.

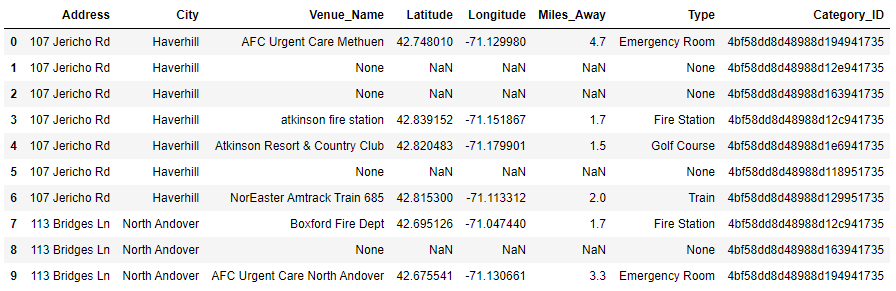
The desired neighborhood amenities are:

Table 1: Desired Amenities and Resulting Category ID

|  |  |  |
| --- | --- | --- |
| **Amenity** | **Distance** | **Category ID** |
| Park | 0.5 miles | 4bf58dd8d48988d163941735 |
| Grocery Store | 2 miles | 4bf58dd8d48988d118951735 |
| Fire Station | 2 miles | 4bf58dd8d48988d12c941735 |
| Police Station | 2 miles | 4bf58dd8d48988d12e941735 |
| Train Station | 2 miles | 4bf58dd8d48988d129951735 |
| Golf Course | 4 miles | 4bf58dd8d48988d1e6941735 |
| Emergency Room | 5 miles | 4bf58dd8d48988d194941735 |

1. Retrieve neighborhood amenities from Foursquare which meet desired criteria.

The first 10 lines of the output are shown below in table x, but the entire results are in a CSV in



1. Combine Redfin and Foursquare data.
2. Filter to only include houses which meet all criteria.
3. Visualize the neighborhood of one of the houses on the acceptable list.
4. Provide a way to visualize other neighborhoods which are also acceptable.

# Appendix

1. Appendix A



1. Appendix B
2. Appendix C