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 price, 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 as requested by the client 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

1. Define Current Residence and Geolocation

   The client currently lives at: 201 Andover ST, North Andover 01845

   After geolocating the address and storing the information in a DataFrame, the summary is below:

	ADDRESS	CITY	STATE	ZIP	PRICE	BEDS	BATHS	LATITUDE	LONGITUDE
0	201 Andover ST	North Andover	MA	01845	0	NA	NA	42.680095	-71.120048

2. Retrieve Redfin Data

   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

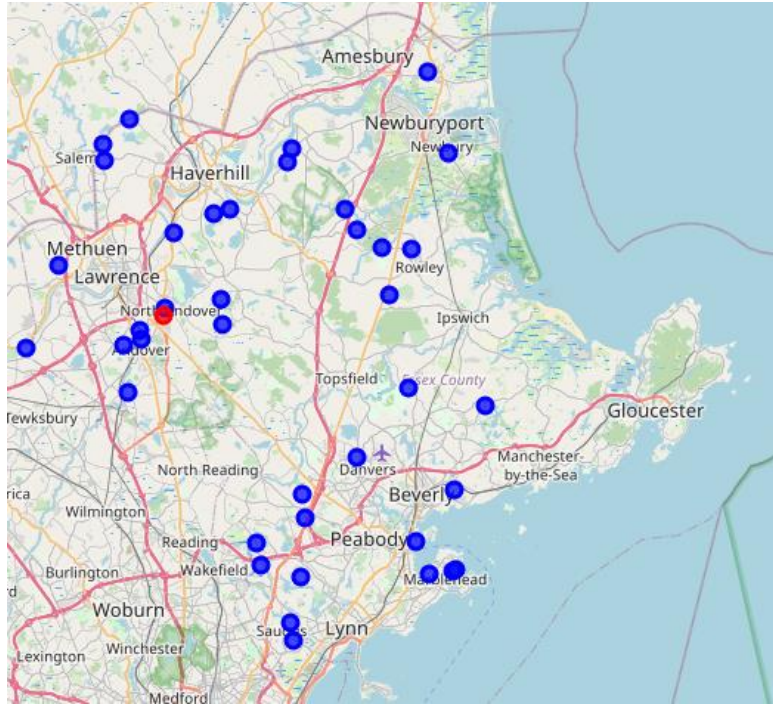   This query returned 39 properties. The full list as a csv file is available in Appendix A.

Figure 1: Redfin Property Query, All Returned Properties

3. Define Neighborhood Amenities and Distance

    The client desired neighborhood amenities and required distances are listed in Table 1. The category ID was manually obtained from the following Foursquare website:
    https://developer.foursquare.com/docs/build-with-foursquare/categories/

Table 1: Desired Amenities and Resulting Category ID

Amenity	Max 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

4. Retrieve Neighborhood Amenities

    The requirements in Table 1 were used along with the 39 properties previously identified to create queries to the Foursquare API to return the closest venue which met the criteria.

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

Table 2: All Returned Properties

	Address	City	Venue_Name	Latitude	Longitude	Miles_Away	Type	Category_ID
0	107 Jericho Rd	Haverhill	AFC Urgent Care Methuen	42.748010	-71.129980	4.7	Emergency Room	4bf58dd8d48988d194941735
1	107 Jericho Rd	Haverhill	None	NaN	NaN	NaN	None	4bf58dd8d48988d12e941735
2	107 Jericho Rd	Haverhill	None	NaN	NaN	NaN	None	4bf58dd8d48988d163941735
3	107 Jericho Rd	Haverhill	atkinson fire station	42.839152	-71.151867	1.7	Fire Station	4bf58dd8d48988d12c941735
4	107 Jericho Rd	Haverhill	Atkinson Resort & Country Club	42.820483	-71.179901	1.5	Golf Course	4bf58dd8d48988d1e6941735
5	107 Jericho Rd	Haverhill	None	NaN	NaN	NaN	None	4bf58dd8d48988d118951735
6	107 Jericho Rd	Haverhill	NorEaster Amtrack Train 685	42.815300	-71.113312	2.0	Train	4bf58dd8d48988d129951735
7	113 Bridges Ln	North Andover	Boxford Fire Dept	42.695126	-71.047440	1.7	Fire Station	4bf58dd8d48988d12c941735
8	113 Bridges Ln	North Andover	None	NaN	NaN	NaN	None	4bf58dd8d48988d163941735
9	113 Bridges Ln	North Andover	AFC Urgent Care North Andover	42.675541	-71.130661	3.3	Emergency Room	4bf58dd8d48988d194941735

As you can see, not every criteria was met for each property. For the 107 Jericho Rd property, only 4 of the 7 criteria have been met and would therefore not be a suitable property.

5. Retrieve Current Neighborhood Amenities

The process was repeated for the amenities in the current residence at 201 Andover ST for future comparison to any property being considered. As can been seen, the current property meets all the required neighborhood amenities, but of course is too small which is why the current owners are looking at alternatives. These neighborhood amenities are shown in the table below.

Table 3: Amenities at Current Residence, 201 Andover ST

	Address	City	Venue_Name	Latitude	Longitude	Miles_Away	Type	Category_ID
0	201 Andover ST	North Andover	North Andover Town Common	42.684259	-71.115001	0.4	Park	4bf58dd8d48988d163941735
1	201 Andover ST	North Andover	Market Basket	42.684600	-71.136469	0.9	Grocery Store	4bf58dd8d48988d118951735
2	201 Andover ST	North Andover	North Andover Fire Department	42.682369	-71.110961	0.5	Fire Station	4bf58dd8d48988d12c941735
3	201 Andover ST	North Andover	North Andover Police Department	42.692820	-71.119249	0.9	Police Station	4bf58dd8d48988d12e941735
4	201 Andover ST	North Andover	MBTA Andover Station	42.659738	-71.144101	1.9	Train Station	4bf58dd8d48988d129951735
5	201 Andover ST	North Andover	Andover Country Club	42.672571	-71.167067	2.4	Golf Course	4bf58dd8d48988d1e6941735
6	201 Andover ST	North Andover	AFC Urgent Care North Andover	42.675541	-71.130661	0.6	Emergency Room	4bf58dd8d48988d194941735

This table was also written to a csv file in Appendix C.

6. Filter Houses Which Meet Criteria

Next, we'll combine the summation of the criteria met in the Foursquare data with the Redfin data to allow filtering to only include those houses that meet all 7 amenity criteria. This results in 8 properties which are deemed suitable. These 8 properties are listed below.

Table 4: Acceptable Properties

	ADDRESS	CITY	STATE	ZIP	PRICE	BEDS	BATHS	LATITUDE	LONGITUDE	Total_Criteria_Met
0	9 Allison Cir	Haverhill	MA	1835	695000	4	2.5	42.749880	-71.073292	7
1	8 Rock O Dundee Rd	Andover	MA	1810	649000	3	1.5	42.669856	-71.141811	7
2	76 High Rd	Newbury	MA	1951	649900	3	2.0	42.790985	-70.855265	7
3	50 Cobblestone Cir	North Andover	MA	1845	739900	3	2.5	42.685932	-71.119205	7
4	37 Biscayne Ave	Saugus	MA	1906	649900	5	2.0	42.469664	-71.002022	7
5	2 Robin Rd	Beverly	MA	1915	799900	4	2.5	42.561573	-70.848841	7
6	19 Emerton St #19	Salem	MA	1970	684900	4	3.5	42.525309	-70.885780	7
7	119 Shawsheen Rd	Andover	MA	1810	799900	3	2.5	42.660239	-71.156665	7

Results

There are 8 properties which meet all 7 amenity criteria and all pricing and size criteria. They are the following properties:

1. 9 Allison Cir, Haverhill
2. 18 Rock O Dundee Rd, Andover
3. 276 High Rd, Newbury
4. 350 Cobblestone Cir, North Andover
5. 437 Biscayne Ave, Saugus
6. 52 Robin Rd, Beverly
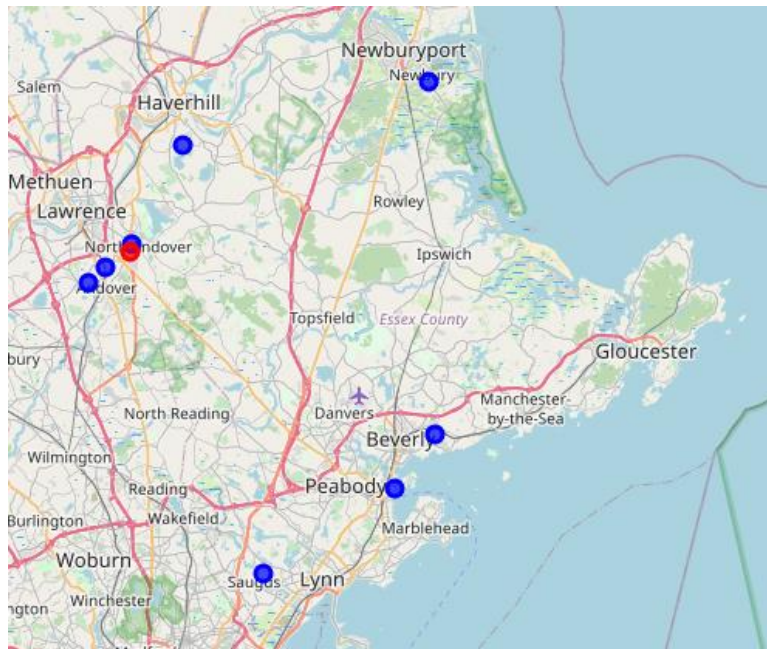7. 619 Emerton St #19, Salem
8. 7119 Shawsheen Rd, Andover

Figure 2: Suitable Properties (Red = Current Property, Blue = For Sale)

The Jupyter Notebook is coded to allow each house to be viewed individually. For example, when the index for 8 Rock O Dundee RD in Andover is entered, you'll see the name of the venue for each required category and the specs of the house for sale. You will also see a map of the neighborhood and will be able to hover over each mark for details (This report contains a static version of the map with added labels).

Table 5: Available Required Amenities Near 8 Rock O Dundee RD

	Address	City	Venue_Name	Latitude	Longitude	Miles_Away	Type	Category_ID
231	8 Rock O Dundee Rd	Andover	Doctors Park	42.675334	-71.135925	0.5	Park	4bf58dd8d48988d163941735
232	8 Rock O Dundee Rd	Andover	Whole Foods Market	42.661179	-71.144769	0.6	Grocery Store	4bf58dd8d48988d118951735
233	8 Rock O Dundee Rd	Andover	Indian Ridge Country Club	42.641669	-71.170020	2.4	Golf Course	4bf58dd8d48988d1e6941735
234	8 Rock O Dundee Rd	Andover	AFC Urgent Care North Andover	42.675541	-71.130661	0.7	Emergency Room	4bf58dd8d48988d194941735
235	8 Rock O Dundee Rd	Andover	Central Station	42.658373	-71.143096	0.8	Fire Station	4bf58dd8d48988d12c941735
236	8 Rock O Dundee Rd	Andover	Andover Police Dept	42.657890	-71.142580	0.8	Police Station	4bf58dd8d48988d12e941735
237	8 Rock O Dundee Rd	Andover	MBTA Andover Station	42.659738	-71.144101	0.7	Train Station	4bf58dd8d48988d129951735

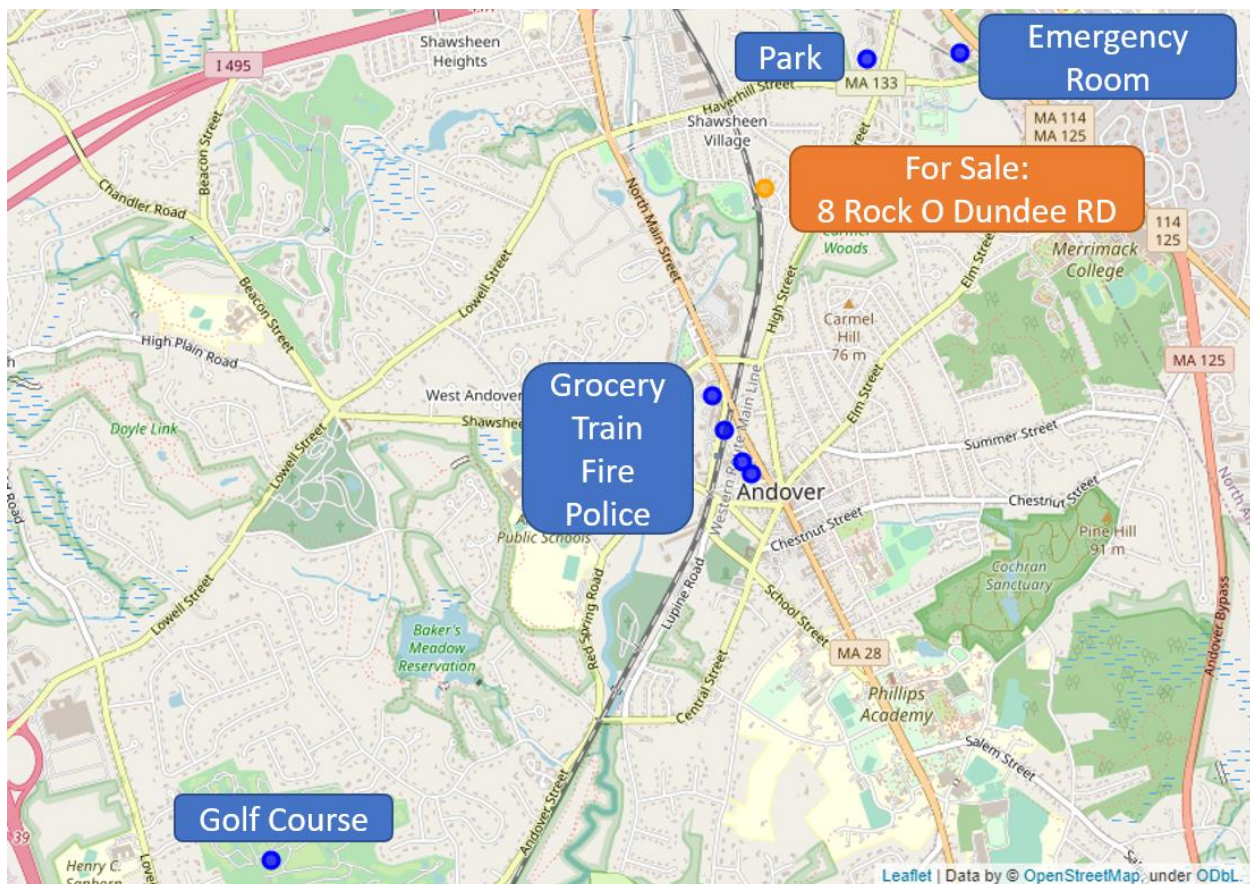|	ADDRESS	CITY	STATE	ZIP	PRICE	BEDS	BATHS	LATITUDE	LONGITUDE
0	8 Rock O Dundee Rd	Andover	MA	1810	649000	3	1.5	42.669856	-71.141811

Figure 3: Neighborhood of 8 Rock O Dundee RD

Discussion

Difficulties encountered while going through this data:

- The Redfin API is not officially supported, I did need a workaround to provide a user agent to allow the request to be made from within the Jupyter Notebook environment.
- The foursquare API is not perfect and sometimes returns results which may be a bit miscategorized or mislocated. For instance, in the 8 Rock O Dundee Rd example, instead of a hospital emergency room, it returned an urgent care clinic which won't have the same capability. I also saw another case where a doctor's office was returned instead of an emergency room. Since there weren't too many of these cases it's easy enough to manually check if this arises. If it became more of a problem, two results from each category could be returned to provide a backup option in case one of the results is not desirable.
- The query limit of the foursquare API can become a problem. For this academic use, it worked out ok, but for a more serious use, it would have to be upgraded to at least a verified account.
- I did some simple plotting of the data for this course, but having the data output to a more user-friendly interface such as Tableau would be nice to visualize the results output to the csv files.

Conclusion

I was able to retrieve a list of for sale houses from Redfin, match them with nearby amenities which met the customer's criteria from Foursquare and filter them to only show houses which met all the requirements. In the end we took a list of 39 properties and reduced it down to only 8. Of the 8 which need to be explored further, the code makes plotting the neighborhood with required amenities easy to allows decisions to be made efficiently. This will significantly reduce the burden for both the real estate office and the customer in searching out suitable homes.

Appendix

1. Appendix A

   redfin_2021-05-21.c
   sv

2. Appendix B

   search_results.csv

3. Appendix C

   search_results_resid
   ence.csv