Olga Zaychykova

Capstone Project - The Battle of Neighborhoods

Relocation Decision Assistance

1. Introduction

1.1. Background

According to the US Census Bureau about 40 million people (or 14% of US population) move every year. People are moving for different reasons: job, family, school, cost of living, etc. But despite the reason the questions people face when they are looking for a new place to live are very common. Is the housing price in the new location affordable? Will I have same or better amenities in my new neighborhood? What is the quality of public schools? The purpose of this project is to help answer those questions with the use of the data science.

1.2. Problem

These days, Internet is great source of information on the topic of interest. But despite wide resources or sometimes because of the resource abundance, the answers may not be so easy to find. For example, a person would need to spend hours researching a single neighborhood and may not even investigate other close by areas that might be more suitable for relocation. In some cases, a person may not even know where to start looking. This project is aimed to guide the relocation decision making by showing clusters of suitable neighborhoods according to provided criteria.

To demonstrate the results, I will study a specific case. A hypothetical family is happy with their current location but want to be closer to their family. Therefore, they want to relocate to the state of MA. Their criteria are good public schools, affordable housing and good amenities – just like their current place.

1.3. Interest

The results of the project will be equally interesting for real estate agents as well as their clients.

2. Data acquisition and cleaning

2.1. Data sources

These project uses three data sources:

- USZipCode database available on PyPI. The database provides geographical, statistical, real-estate data for every US zip code.

  Here is information provide by USZipCode database for the target zip code:

```
{'area_code_list': ['510', '707', '925'],
 'bounds_east': -121.925567,
 'bounds_north': 37.942412,
 'bounds_south': 37.849998,
 'bounds_west': -122.05489,
 'common_city_list': ['Walnut Creek'],
 'county': 'Contra Costa County',
 'housing_units': 10756,
 'land_area_in_sqmi': 15.25,
 'lat': 37.91,
 'lng': -122.05,
 'major_city': 'Walnut Creek',
 'median_home_value': 719200,
 'median_household_income': 121067,
 'occupied_housing_units': 10390,
 'population': 25818,
 'population_density': 1693.0,
 'post_office_city': 'Walnut Creek, CA',
 'radius_in_miles': 6.0,
 'state': 'CA',
 'timezone': 'Pacific',
 'water_area_in_sqmi': 0.01,
 'zipcode': '94598',
 'zipcode_type': 'Standard'}
```

- FourSquare data for information on amenities. FourSquare API is used to collect the data.

  Exploration of venues for the target zip code:

	name	categories	lat	lng
0	Heather Farm Park	Park	37.918570	-122.041768
1	Gardens at Heather Farm	Garden	37.918820	-122.044038
2	Lottie's Creamery	Ice Cream Shop	37.899577	-122.060806
3	Sports Basement	Sporting Goods Shop	37.918327	-122.036975
4	Montecatini Ristorante	Italian Restaurant	37.901636	-122.062617

- GreatSchools web site for public schools rating. I wrote a function to scrape the web site and to get the average schools rating for the provided zip code.

  Result of GreatSchools scraping function for the target zip code:

```
Average elementary schools rating:  6
Average middle schools rating:  7
Average high schools rating:  7
```

2.2. Data Cleaning

The data for the desired relocation area (state of MA) is collected from the USZipCode data. I don't need locations that are lacking information about median house price. Therefore, those locations are excluded.

To narrow down the potential relocation areas, I have removed zip codes that have median house price above the desired target. For the target price, I have taken the house median price in the current location of the client. Also, after obtaining average school rating, I have eliminated those locations that have average school rating below the desired outcome.

2.3. Feature Selection

I will use clustering to find location that are like the current place of the client. The following features are be used: average school rating (obtained from GreatSchools), median home price (obtained from USZipCode database), and list of venues (obtained from FourSquare).

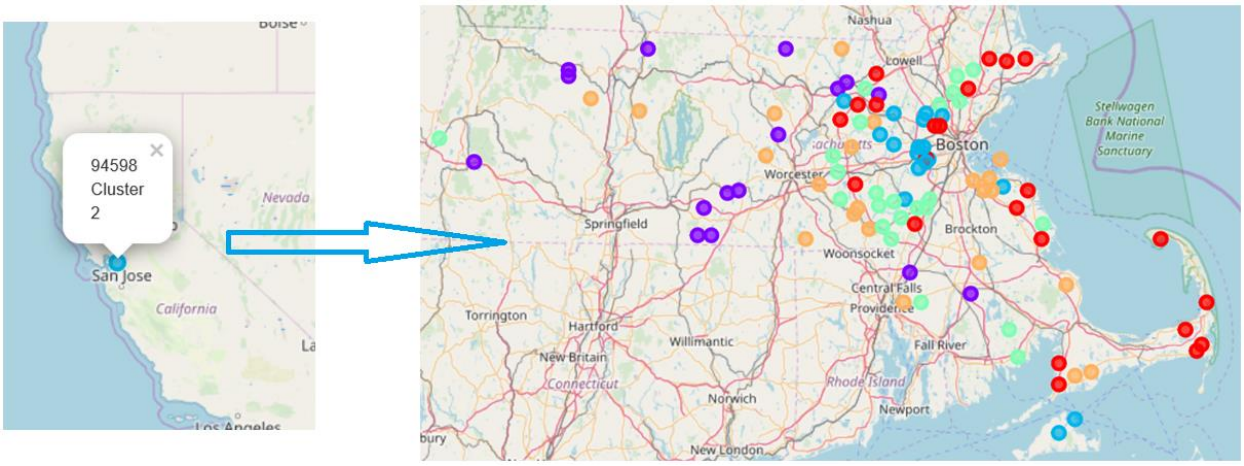
I have used one hot encoding and the frequency mean appearance of each venue as well as average school rating in the 5 miles radius of the zip code to produce numerical features suitable for clustering.

Here is an example of the prepared dataframe:

	Zipcode	Accessories Store	African Restaurant	Airport	American Restaurant	Amphitheater	Antique Shop	Apres Ski Bar	Aquarium	Arcade	...	Warehouse Store	Waterfall	Waterfront
0	01010	0.000000	0.00	0.000000	0.035714	0.00	0.071429	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
1	01054	0.000000	0.00	0.000000	0.037037	0.00	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
2	01081	0.000000	0.00	0.000000	0.000000	0.00	0.086957	0.000000	0.00	0.00	...	0.000000	0.043478	0.000000
3	01238	0.034884	0.00	0.000000	0.069767	0.00	0.011628	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
4	01254	0.000000	0.00	0.000000	0.023810	0.00	0.011905	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
5	01338	0.000000	0.00	0.000000	0.100000	0.00	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
6	01341	0.000000	0.00	0.000000	0.066667	0.00	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
7	01360	0.000000	0.00	0.000000	0.062500	0.00	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
8	01370	0.000000	0.00	0.000000	0.120000	0.00	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
9	01431	0.000000	0.00	0.000000	0.037037	0.00	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
10	01432	0.000000	0.00	0.000000	0.053191	0.00	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
11	01434	0.000000	0.00	0.000000	0.050847	0.00	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
12	01451	0.000000	0.00	0.000000	0.013699	0.00	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000
13	01460	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.00	0.01	...	0.000000	0.000000	0.000000
14	01463	0.000000	0.00	0.000000	0.035714	0.00	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.000000

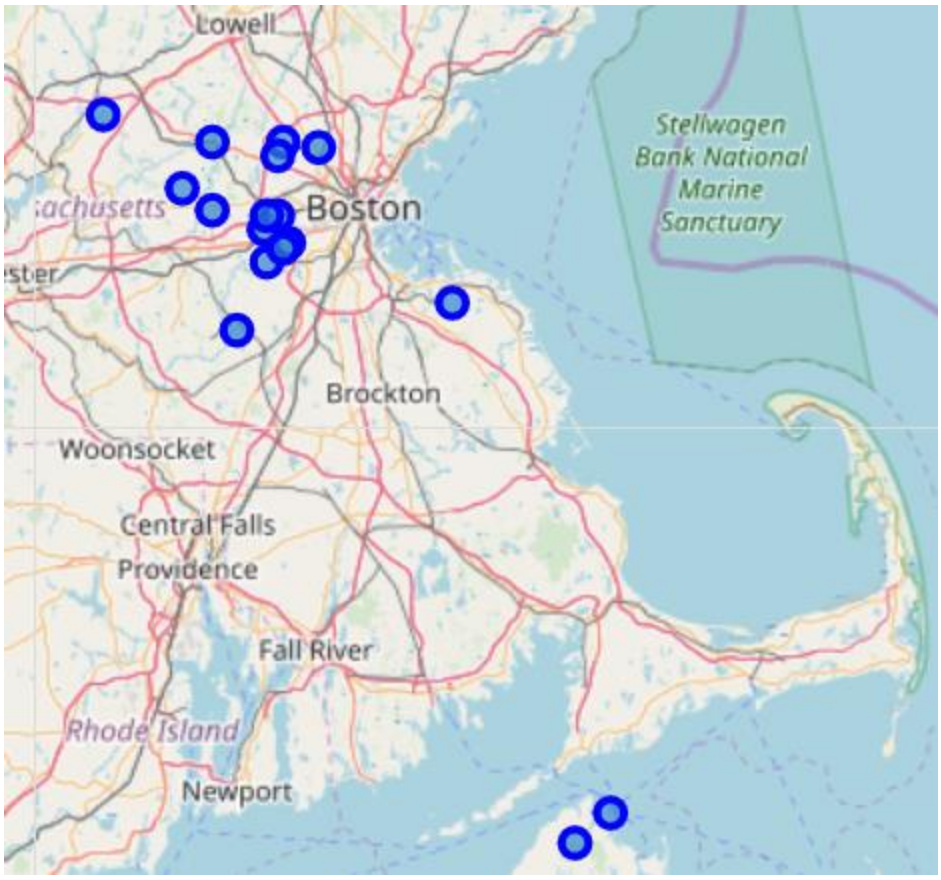
3. Methodology

To find potential relocation areas, I have used k-Means Clustering algorithm. This simple machine learning algorithm allows to group locations by their similarities: near by venues, school rating, median house price.



4. Results

Current location ended up in the cluster 2. That cluster contains 17 potential relocation cities that are similar to the current location. Here are those 17 locations on the map:



The list of the cluster 2 cities with their median home price and average school rating:

	post_office_city	median_home_value	e_school
12	Harvard, MA	590900	6
30	Concord, MA	686700	7
38	Sudbury, MA	632800	6
39	Wayland, MA	589700	6
45	Winchester, MA	676800	6
54	Hingham, MA	646700	6
57	Medfield, MA	582800	6
74	Lexington, MA	688500	6
75	Lexington, MA	718300	6
76	Newton Highlands, MA	682600	6
77	Newton Lower Falls, MA	643300	6
79	West Newton, MA	646700	6
80	Auburndale, MA	633600	6
83	Needham, MA	678500	7
84	Needham Heights, MA	631000	6
88	Oak Bluffs, MA	627300	6
89	Vineyard Haven, MA	630600	7
102	Walnut Creek, CA	719200	6

The bottom city is our current location that we are comparing to.
Three locations are standing up from the list because of the higher school rating: Concord, Needham and Vineyard Haven.

The list of top 10 venues in the cluster 2:

	zipcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
12	01451	Donut Shop	Pizza Place	Sandwich Place	Farm	Italian Restaurant	Intersection	Hotel	Café	Golf Course	Fast Food Restaurant
30	01742	American Restaurant	Historic Site	Pharmacy	Pizza Place	History Museum	Ice Cream Shop	Park	Café	Coffee Shop	Liquor Store
38	01776	Sandwich Place	American Restaurant	Pizza Place	Gym / Fitness Center	Gym	Golf Course	Coffee Shop	Farm	Donut Shop	Bakery
39	01778	Coffee Shop	Grocery Store	American Restaurant	Pizza Place	Italian Restaurant	Bakery	Golf Course	Department Store	Furniture / Home Store	Japanese Restaurant
45	01890	Italian Restaurant	Trail	Ice Cream Shop	Mexican Restaurant	Grocery Store	Park	Deli / Bodega	Bakery	Vegetarian / Vegan Restaurant	Supermarket
54	02043	American Restaurant	Italian Restaurant	Seafood Restaurant	Grocery Store	Pizza Place	Golf Course	Café	Beach	Department Store	Breakfast Spot
57	02052	Donut Shop	Pizza Place	Grocery Store	Bar	Pharmacy	Convenience Store	Coffee Shop	American Restaurant	Restaurant	Sandwich Place
74	02420	American Restaurant	Pizza Place	Italian Restaurant	Grocery Store	Hotel	Coffee Shop	Sandwich Place	Spa	Gym	Supermarket
75	02421	American Restaurant	Italian Restaurant	Grocery Store	Seafood Restaurant	Hotel	Pizza Place	Sandwich Place	Park	Gym	Historic Site
76	02461	Ice Cream Shop	Pizza Place	American Restaurant	Park	Italian Restaurant	Grocery Store	Burger Joint	Liquor Store	Coffee Shop	Japanese Restaurant
77	02462	American Restaurant	Grocery Store	Ice Cream Shop	Pizza Place	Coffee Shop	Sandwich Place	Liquor Store	Bakery	Mexican Restaurant	Golf Course
79	02465	Ice Cream Shop	Italian Restaurant	Grocery Store	Liquor Store	Pizza Place	Bakery	Café	Mexican Restaurant	Burger Joint	Pub
80	02466	Italian Restaurant	Ice Cream Shop	Pizza Place	Grocery Store	Liquor Store	Bakery	Burger Joint	Park	American Restaurant	New American Restaurant
83	02492	Italian Restaurant	American Restaurant	Coffee Shop	Pizza Place	Grocery Store	Bagel Shop	Sandwich Place	Japanese Restaurant	Trail	Golf Course
84	02494	Ice Cream Shop	American Restaurant	Italian Restaurant	Pizza Place	Grocery Store	Liquor Store	Burger Joint	Japanese Restaurant	Bagel Shop	Sandwich Place
88	02557	Beach	American Restaurant	Seafood Restaurant	Bakery	Coffee Shop	Bar	Bed & Breakfast	Ice Cream Shop	Golf Course	Brewery
89	02568	Beach	American Restaurant	Bakery	Seafood Restaurant	Coffee Shop	Ice Cream Shop	Clothing Store	Bed & Breakfast	Farm	Movie Theater
102	94598	Ice Cream Shop	Bakery	Grocery Store	Gym / Fitness Center	Park	Spa	BBQ Joint	Sporting Goods Shop	Burger Joint	American Restaurant

5. Observation & Conclusion

The results allow to narrow down the potential relocation cities. With 5 centroids, we ended up with 17 similar locations to look at. Out of those 17 locations some may be selected as preferred, based on lower home price and higher school rating (for example Concord, MA). Or preference may be given based on top venues in the area.