International Home Searching Using Web Scraping and Foursquare

Alejandro Aylwin

May 23, 2020

Index

Introduction

Buying a house is among the biggest and most important decisions people make during their life. In an ideal situation, a future home owner will visit all potential properties multiple times and evaluate the benefits and shortcomings of each one before making a decision. But how can this be done from thousands of miles away? What are the alternatives for a person from Asia moving to South America, or someone from Australia moving to Europe? The following project will attempt to provide a solution for people living far away from their future homes that do not have the chance to evaluate their options in person. For this we will consider our first case consisting of a young couple from South America moving to Santa Barbara, California for business.

Data

The first set of data used will be the characteristics of the available homes. These will include:

- Price
- Address
- Type (Condo or Townhouse)
- Number of bedrooms
- Number of bathrooms
- Area (Square feet)

This first set of data will be obtained from the real estate website zillow.com.

The second set of data will describe the main venues in proximity to the available homes that the future owners can benefit from. This second set of data will be obtained using the Foursquare API to retrieve the necessary information from the database.

Both sets of information will be used as criteria in order to come up with the best alternative.

Methodology

1. **Conduct Interview with Clients**

The first stage in our methodology consists of interviewing the clients to determine their budget, requirements and priorities. From the interview we gather the following information:

- Budet: Maximum $1.000.000
- Requirements: The couple are pet owners so their main requirement is to have either a park or a dog run within the top venues closeby.
- The order of priorities after their initial requirement are: number of bedrooms, number of bathrooms and then price.

2. **Web Scraping**

We can now scrape the real estate website zillow.com for properties within the budget. To do this we use ParseHub .We create and save a csv file with all the information we need. The csv file contains the following table with 34 available homes.

Type	Price $	Address	Latitude	Longitude	Bedrooms	Bathrooms	Area sqr_ft
Townhouse for sale	995,000	282 N La Cumbre Rd, Santa Barbara, CA 93110	34.4439	-119.75038	3	2	1,472
Condo for sale	990,000	105 W De La Guerra St UNIT D, Santa Barbara, CA 93101	34.41782	-119.70016	1	2	1,116
Townhouse for sale	959,000	262 Calle Esperanza, Santa Barbara, CA 93105	34.43687	-119.74597	3	3	1,646
Townhouse for sale	949,000	105 W De La Guerra St UNIT E, Santa Barbara,	34.41782	-119.70016	1	2	1,066

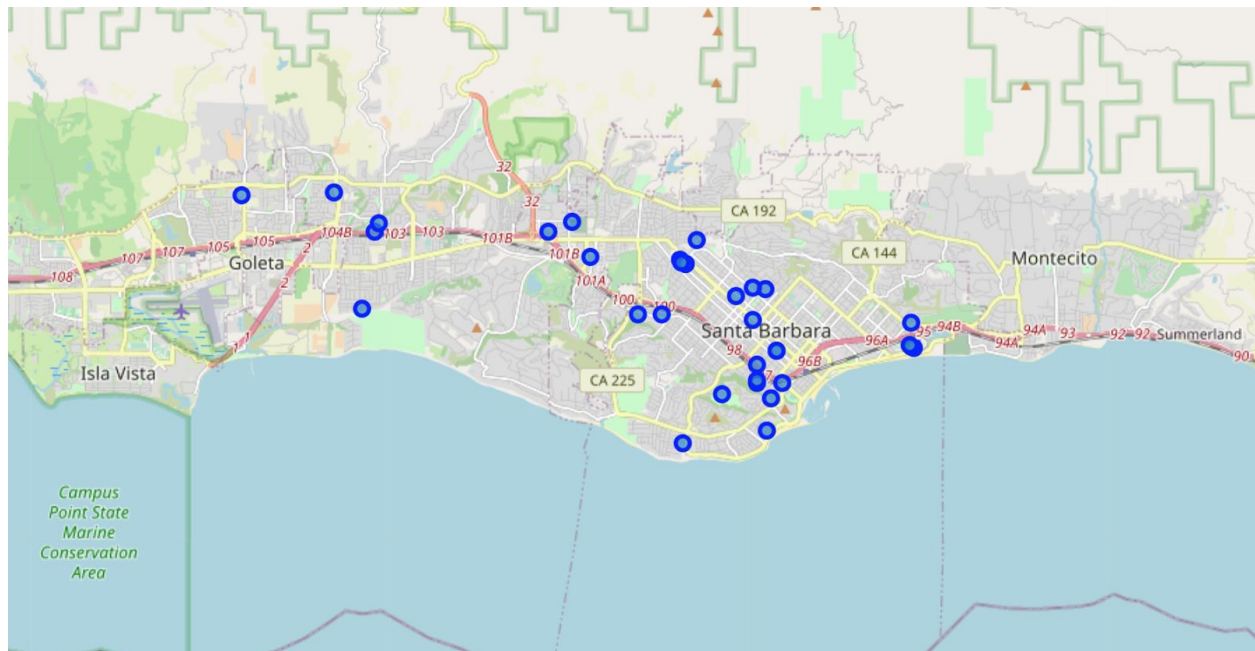|							
		CA 93101					
Condo for sale	925,000	4004 Via Lucero APT 13, Santa Barbara, CA 93110	34.44212	-119.75612	2	3	2,010
Condo for sale	899,000	1220 Coast Village Rd APT 212, Santa Barbara, CA 93108	34.0928092	-118.3286614	2	2	1,209
Townhouse for sale	895,000	2605 Hacienda Ct, Santa Barbara, CA 93105	34.425312	-119.734112	3	3	1,840
Condo for sale	895,000	18 W Victoria St APT 107, Santa Barbara, CA 93101	34.4241051	-119.7060586	1	2	1,066
Condo for sale	869,000	5463 Tree Farm Ln, Santa Barbara, CA 93111	34.449453	-119.831174	2	2	1,758
Townhouse for sale	835,000	1426 Laguna St #A, Santa Barbara, CA 93101	34.430339	-119.7028853	1	2	1,049
Townhouse for sale	825,000	211 Reef Ct, Santa Barbara, CA 93109	34.399264	-119.7233002	2	2	1,077
Condo for sale	799,500	66 Barranca Ave APT 4, Santa Barbara, CA 93109	34.401684	-119.7026889	2	2	1,155
Townhouse for sale	782,000	323 Ladera St #2, Santa Barbara, CA 93101	34.408441	-119.7014211	3	2	1,425
Townhouse for sale	780,000	621 W Ortega St APT A, Santa Barbara, CA 93101	34.411379	-119.7049699	2	3	1,464
Condo for sale	775,000	2330 Vista Madera, Santa Barbara, CA 93101	34.425404	-119.7283002	2	1	1,065
Townhouse for sale	735,000	633 W Ortega St, Santa Barbara, CA 93101	34.412033	-119.7049768	2	2	1,118
Condo for sale	699,000	25 Ocean View Ave #C5, Santa Barbara, CA 93103	34.423699	-119.6671551	2	2	1,246
Condo for sale	695,000	2627 State St APT 3, Santa Barbara, CA 93105	34.436321	-119.7238334	2	2	1,396
Townhouse	690,000	5305 Traci Dr, Santa	34.449893	-119.80875	2	2	1,154

e for sale		Barbara, CA 93111		5			
Condo for sale	690,000	415 W Gutierrez St APT 6, Santa Barbara, CA 93101	34.411378	-119.698834	3	2	1,278
Condo for sale	650,000	316 Por La Mar Cir, Santa Barbara, CA 93103	34.418657	-119.666584	1	1	700
Condo for sale	639,000	2525 State St APT 25, Santa Barbara, CA 93105	34.435492	-119.722377	1	2	1,352
Condo for sale	636,000	2525 State St APT 11, Santa Barbara, CA 93105	34.435492	-119.722377	2	2	1,250
Condo for sale	619,900	1600 Garden St APT 35, Santa Barbara, CA 93101	34.430708	-119.705986	2	2	891
Condo for sale	619,000	1701 Anacapa St UNIT 3, Santa Barbara, CA 93101	34.429078	-119.710015	2	2	1,242
Condo for sale	615,000	2525 State St APT 31, Santa Barbara, CA 93105	34.435492	-119.722377	1	1	1,072
Condo for sale	614,000	969 Miramonte Dr APT 5, Santa Barbara, CA 93109	34.409071	-119.713444	2	2	1,120
Condo for sale	610,000	27 N San Marcos Rd #B, Santa Barbara, CA 93111	34.442138	-119.798848	2	1	1,221
Condo for sale	605,000	432 Por La Mar Cir, Santa Barbara, CA 93103	34.418717	-119.667328	1	1	700
Townhouse for sale	599,000	5034 Birchwood Rd, Santa Barbara, CA 93111	34.443791	-119.79765	2	3	1,104
Townhouse for sale	599,000	5095 Rhoads Ave APT C, Santa Barbara, CA 93111	34.426364	-119.801748	2	2	1,260
Condo for sale	598,000	125 Por La Mar Cir, Santa Barbara, CA 93103	34.419174	-119.667404	1	1	700
Condo for sale	569,000	30 W Constance Ave UNIT 2, Santa Barbara,	34.435721	-119.723643	1	1	1,009

		CA 93105					
Townhous e for sale	555,000	425 Transfer Ave APT A, Santa Barbara, CA 93101	34.415081	-119.70503 8	2	2	1,340
Condo for sale	519,000	2727 Miradero Dr APT 306, Santa Barbara, CA 93105	34.440243	-119.71975 4	2	2	1,063

Table 1: Available homes information

3. Pandas Dataframe and Mapping Available Homes

We then open the csv file in our Jupyer Notebook using Pandas and map the available homes using Folium.

Map 1: Available Homes

4. Obtaining and Formatting Second Dataset Using Foursquare API

Now that we have our first set of data we can now move on to the second set. In order to do this we use our Foursquare credentials and version to request the nearby venues for every address on the available homes dataframe. From our request we make sure to gather only data that will be useful to us which include the names of the names, locations (latitude and longitude) and categories of the venues. Once we have all the requested information we now must format the data. To do this we use one hot encoding and group by Address. After this we can sort the rows to show the top 5 nearby venues for each address. By doing so, we format the information into a dataframe that matches our first dataset. The result is the following dataframe:

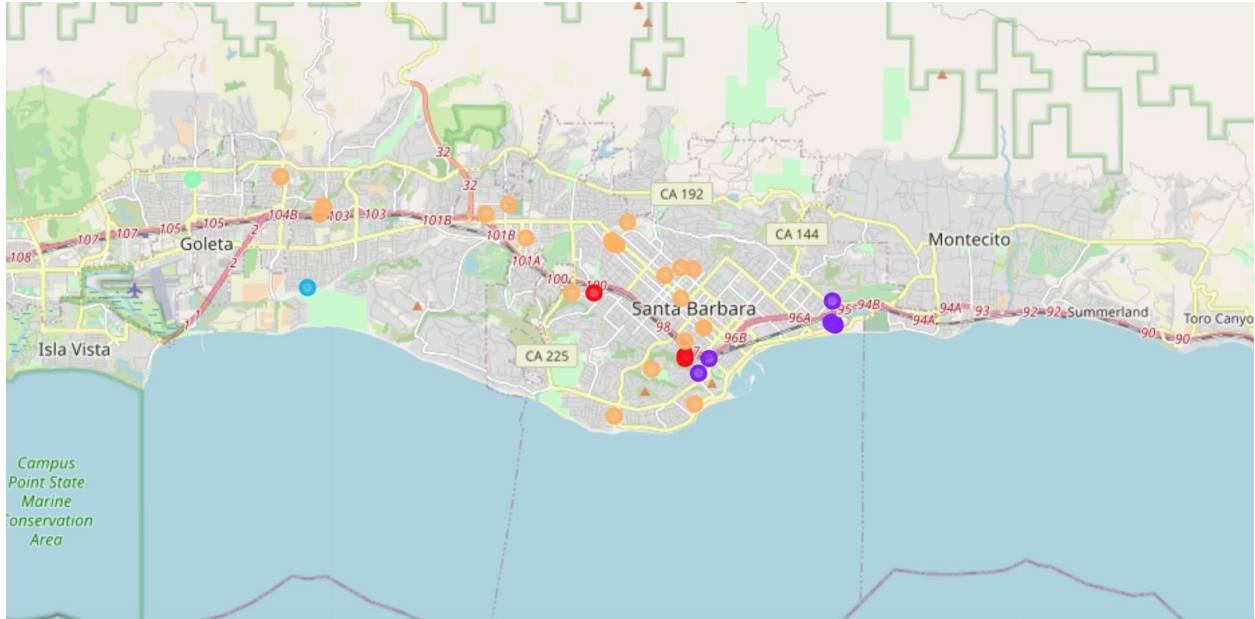| Address	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
282 N La Cumbre Rd, Santa Barbara, CA 93110	Fast Food Restaurant	Hotel	Bakery	Pizza Place	Coffee Shop
105 W De La Guerra St UNIT D, Santa Barbara, CA 93101	Sushi Restaurant	Pizza Place	Bar	Ice Cream Shop	Burger Joint
262 Calle Esperanza, Santa Barbara, CA 93105	Pizza Place	Bank	Accessories Store	Mobile Phone Shop	Shipping Store
105 W De La Guerra St UNIT E, Santa Barbara, CA 93101	Sushi Restaurant	Pizza Place	Bar	Ice Cream Shop	Burger Joint
4004 Via Lucero APT 13, Santa Barbara, CA 93110	Coffee Shop	Sporting Goods Shop	Fast Food Restaurant	Home Service	Food Truck
1220 Coast Village Rd APT 212, Santa Barbara, CA 93108	Gym	Comedy Club	Vegetarian / Vegan Restaurant	Sandwich Place	Theater
2605 Hacienda Ct, Santa Barbara, CA 93105	Tanning Salon	Home Service	Chinese Restaurant	Dance Studio	Liquor Store
18 W Victoria St APT 107, Santa Barbara, CA 93101	Bakery	Italian Restaurant	Deli / Bodega	Paper / Office Supplies Store	erican Restau
5463 Tree Farm Ln, Santa Barbara, CA 93111	Farm	Zoo Exhibit	Department Store	Farmers Market	Donut Shop
1426 Laguna St #A, Santa Barbara, CA 93101	Sushi Restaurant	Lounge	Playground	Diner	Bed & Breakfast
211 Reef Ct, Santa Barbara, CA 93109	Grocery Store	Pizza Place	Pharmacy	Fast Food Restaurant	nvenience Sto

Address					
66 Barranca Ave APT 4, Santa Barbara, CA 93109	Beer Garden	Theater	Snack Place	Beach	College Library
323 Ladera St #2, Santa Barbara, CA 93101	Hotel	College Library	Grocery Store	Snack Place	Park
621 W Ortega St APT A, Santa Barbara, CA 93101	Park	Convenience Store	Playground	Gym	Tour Provider
2330 Vista Madera, Santa Barbara, CA 93101	Park	Zoo Exhibit	Department Store	Exhibit	Donut Shop
633 W Ortega St, Santa Barbara, CA 93101	Gym	Tour Provider	Park	Deli / Bodega	Donut Shop
25 Ocean View Ave #C5, Santa Barbara, CA 93103	Zoo Exhibit	Liquor Store	Restaurant	Campground	Exhibit
2627 State St APT 3, Santa Barbara, CA 93105	Coffee Shop	Mexican Restaurant	Italian Restaurant	Hotel	Gastropub
5305 Traci Dr, Santa Barbara, CA 93111	Home Service	Spa	Zoo Exhibit	Deli / Bodega	Donut Shop
415 W Gutierrez St APT 6, Santa Barbara, CA 93101	Hotel	Bed & Breakfast	Gym	Motel	nvenience Sto
316 Por La Mar Cir, Santa Barbara, CA 93103	Hotel	Zoo Exhibit	Tennis Court	Zoo	Volleyball Court
2525 State St APT 25, Santa Barbara, CA 93105	Italian Restaurant	Hotel	Bakery	Motel	xican Restaur
2525 State St APT 11, Santa Barbara, CA 93105	Italian Restaurant	Hotel	Bakery	Motel	xican Restaur
1600 Garden St APT 35, Santa Barbara, CA 93101	Motel	Playground	Bed & Breakfast	Garden	Park
1701 Anacapa St UNIT 3, Santa Barbara, CA 93101	Hotel	Breakfast Spot	Park	Deli / Bodega	nvenience Sto
2525 State St APT 31, Santa Barbara, CA 93105	Italian Restaurant	Hotel	Bakery	Motel	xican Restaur
969 Miramonte Dr APT 5, Santa Barbara, CA 93109	Pool	Recruiting Agency	Zoo Exhibit	Department Store	Exhibit
27 N San Marcos Rd #B, Santa Barbara, CA 93111	Home Service	Pool	Dog Run	Zoo Exhibit	Deli / Bodega
432 Por La Mar Cir, Santa Barbara, CA 93103	Hotel	Zoo Exhibit	Harbor / Marina	Restaurant	Campground
5034 Birchwood Rd, Santa Barbara, CA 93111	Park	Pool	Dog Run	Zoo Exhibit	Exhibit
5095 Rhoads Ave APT C, Santa Barbara, CA 93111	Home Service	Department Store	Seafood Restaurant	Business Service	Zoo Exhibit

125 Por La Mar Cir, Santa Barbara, CA 93103	Hotel	Zoo Exhibit	Restaurant	Tennis Court	Zoo
30 W Constance Ave UNIT 2, Santa Barbara, CA 93105	Coffee Shop	Mexican Restaurant	Italian Restaurant	Hotel	Gastropub
425 Transfer Ave APT A, Santa Barbara, CA 93101	Mexican Restaurant	Coffee Shop	Thrift / Vintage Store	Pizza Place	Pharmacy
2727 Miradero Dr APT 306, Santa Barbara, CA 93105	History Museum	General Entertainment	Rental Service	Planetarium	Zoo Exhibit

Table 2: Top 5 Nearby Venues

5. Clustering and Labeling

Now we can cluster and label our second dataset based on the nearby venue characteristics. We use this Unsupervised Machine Learning Method because the highest priority of this case is to deliver a specific venue category requirement. After we have assigned and labeled the addresses to their clusters we can now merge the first and the second datasets by the address resulting in a dataframe with all the information gathered. This dataframe includes the Type, Price, Address, Latitude, Longitude, Bedrooms, Bathrooms, Area, 1st Most Common Venue, 2nd Most Common Venue, 3rd Most Common Venues, 4th Most Common Venue and 5th Most Common Venue. We can also now map the available homes according to their clusters.

Map 2: Available Homes Clustered

Results

The first cluster (label 0) includes 2 properties:

- 621 W Ortega St APT A, Santa Barbara, CA 93101
- 633 W Ortega St, Santa Barbara, CA 93101

The second cluster (label 1) is the largest and includes 27 properties:

- 282 N La Cumbre Rd, Santa Barbara, CA 93110
- 105 W De La Guerra St UNIT D, Santa Barbara, CA 93101
- 262 Calle Esperanza, Santa Barbara, CA 93105
- 105 W De La Guerra St UNIT E, Santa Barbara, CA 93101
- 4004 Via Lucero APT 13, Santa Barbara, CA 93110
- 2605 Hacienda Ct, Santa Barbara, CA 93105
- 18 W Victoria St APT 107, Santa Barbara, CA 93101
- 5463 Tree Farm Ln, Santa Barbara, CA 93111
- 1426 Laguna St #A, Santa Barbara, CA 93101
- 211 Reef Ct, Santa Barbara, CA 93109
- 66 Barranca Ave APT 4, Santa Barbara, CA 93109

- 323 Ladera St #2, Santa Barbara, CA 93101
- 25 Ocean View Ave #C5, Santa Barbara, CA 93103
- 2627 State St APT 3, Santa Barbara, CA 93105
- 415 W Gutierrez St APT 6, Santa Barbara, CA 93101
- 316 Por La Mar Cir, Santa Barbara, CA 93103
- 2525 State St APT 25, Santa Barbara, CA 93105
- 2525 State St APT 11, Santa Barbara, CA 93105
- 1600 Garden St APT 35, Santa Barbara, CA 93101
- 1701 Anacapa St UNIT 3, Santa Barbara, CA 93101
- 2525 State St APT 31, Santa Barbara, CA 93105
- 432 Por La Mar Cir, Santa Barbara, CA 93103
- 5095 Rhoads Ave APT C, Santa Barbara, CA 93111
- 125 Por La Mar Cir, Santa Barbara, CA 93103
- 30 W Constance Ave UNIT 2, Santa Barbara, CA 93105
- 425 Transfer Ave APT A, Santa Barbara, CA 93101
- 2727 Miradero Dr APT 306, Santa Barbara, CA 93105

The third cluster (label=2) includes 1 property:

- 2330 Vista Madera, Santa Barbara, CA 93101

The fourth cluster (label=3) includes 3 properties:

- 969 Miramonte Dr APT 5, Santa Barbara, CA 93109
- 27 N San Marcos Rd #B, Santa Barbara, CA 93111
- 5034 Birchwood Rd, Santa Barbara, CA 93111

The fifth and last cluster (label=4) includes 1 property:

- 5305 Traci Dr, Santa Barbara, CA 93111

By observing with more detail we see that the first and fourth clusters (labels 0 and 3) include venues consisting of outdoor activities. On the other hand we observe that the second cluster (label 1) groups addresses surrounded mainly by hotel and gastronomical services. Finally we see that the third and fifth cluster (labels 2 and 4) group seemingly unrelated venue categories.

We find our 4 potential candidates in clusters 1 and 4 (labels 0 and 3) since they are the only properties with "Parks" and "Dog Runs" within their top 5 nearby venues. These properties correspond to the following addresses:

From cluster 1:

- Option 1: 621 W Ortega St APT A, Santa Barbara, CA 93101 (Park as 1st Most Common Venue)
- Option 2: 633 W Ortega St, Santa Barbara, CA 93101 (Park as 3rd Most Common Venue)

From cluster 2:

- Option 3: 27 N San Marcos Rd #B, Santa Barbara, CA 93111 (Dog Run as 3rd Most Common Venue)
- Option 4: 5034 Birchwood Rd, Santa Barbara, CA 93111 (Park as 1st Most Common Venue and Dog Run as 3rd Most Common Venue)

Recommendation

In order to select the best choice for our clients we follow the selection criteria using their order of priorities.

Option	Bedrooms	Bathrooms	Price
Option 1 ⇨⇨⇨	2 ⇨⇨⇨	3 ⇨⇨⇨	$780,000
Option 2 ⇨⇨⇨	2 ⇨⇨⇨	2	-
Option 3 ⇨⇨⇨	2 ⇨⇨⇨	1	-
Option 4 ⇨⇨⇨	2 ⇨⇨⇨	3 ⇨⇨⇨	$599,000

Table 3: Selection Criteria

In Table 3 we see that all options have the same number of bedrooms. So we pass to our second criteria. We see how options 2 and 3 are eliminated due to the number of bathrooms. Since Option

1 and 4 both have the same number of bathrooms we move to the next criteria which is the price. Since Option 4 has a lower price than Option 1, we conclude that the best option for the young couple is Option 4:

Address: 5034 Birchwood Rd, Santa Barbara, CA 93111

Type: Townhouse

Price: $599,000

Bedrooms: 2

Bathrooms: 3

Area: 1,104 square feet

1st Most Common Venue: Park

2nd Most Common Venue: Pool

3rd Most Common Venue: Dog Run

4th Most Common Venue: Zoo Exhibit

5th Most Common Venue: Exhibit

Conclusion

In conclusion, we see how basic data can be transformed into useful and valuable information by following a structured methodology. These series of steps are scalable, applicable and improvable to fit different requirements and cases. This example as shown can help at least in some proportion to guide the decision making process.