

Where should I buy a house in Austin

Both a nice home and an wise investment

Hongxu Dong
March 13, 2021

1. Introduction

1.1. background

My husband and I work and live in Houston, Texas. We both have a well-paid job in oil&gas companies. Starting from last year, we began to seriously think about how to invest our money, with high return rate and low risk. No matter how many different options we have, buying a house as an investment will definitely be a great one, especially under the circumstance that the interest rate for mortgage is historically low in U.S. However, where we should buy a house becomes an essential topic. None of an ambitious investor would want to buy a house in Houston; with the seemingly unlimited land to use, no one would like to pay much to buy an older, smaller, or worse school-district house when better options are everywhere available. Plus, the house prices have been fluctuating up and down in recent years, and no one is super optimistic about its future trend with the downhill of oil&gas industry. So our target, naturally becomes Austin. Austin is only two hours drive from us, and it's all known to be a booming city, where more and more high-tech companies originally based in bay area have either moved or planned to move to. And who knows, we may also move there in the near future. With the Austin house price rocketing, more friends like us are considering buying a house there, before it's too late.

1.2. Problem&Audience

For young professionals living nearby who plan to buy a house as both an investment method and a potential future home, we want the houses to meet the following three main requirements: 1) it should be located in a convenient area, close to either offices of big companies or transportation stations, as well as good rating restaurants and bars. This is important not only because we may want to live there in the future, more importantly it will make it a popular location for renting; 2)

the neighborhood should be safe with low crime rate; 3) the house price should be increasing (which is easy to meet in Austin), and also in a range that we can afford.

In summary, we want to find one or a few good neighborhoods in Austin, where life is convenient and house price is still affordable while having good increasing potential. How do we want to define 'affordable'? Having a rough collection of salaries information of friends of mine ,at similar age and with comparable education background, I believe a house in the price range of 400 to 1000 thousands would be affordable, of course with a mortgage which means our down payment would be about 100 to 200 thousands.

The final recommendation of this analysis can be used as a guide for myself, and those young professionals who are considering buying a house in Austin, as an investment, or as their own home.

2. Data

2.1. Data needed

In order to evaluate the neighborhoods in Austin from the above three different aspects, we will need to utilize the data from three resources: 1) Foursquare for the venues in each neighborhoods in Austin; 2) crime rate data of neighborhood in Austin; 3) house price statistics of neighborhoods in Austin.

2.2. Data sources

2.2.1. Austin neighborhoods and coordinates

We can find a list of neighborhoods in Austin here: https://en.wikipedia.org/wiki/List_of_Austin_neighborhoods. The list of neighborhoods doesn't come with coordinates, so we need to read in the list first and then get the coordinates.

2.2.2. Venues data from Foursquare

Using Foursquare API, we can request all the venues in all neighborhoods in Austin. The neighborhoods can be further clustered according to the most popular

venue types, which will provide an indicator as to which neighborhood is convenient for specific needs.

2.2.3. House price and crime data

The average or median house price by neighborhood in Austin will be obtained from the report in Texas Real Estate Research Center (<http://www.city-data.com/nbmaps/neigh-Austin-Texas.html#N121>). The information shown for each neighborhood is slightly different one from another and we may need to manually read in the data needed. Note that the data is only for the year of 2016 and is not up to date.

Crime data of each neighborhoods in Austin can be found from a variety of different websites, including austintexas.com (The search will be limited to recent one year (from March 1st 2020 to March 1st 2021) and the list will be scraped using beautifulsoup library and key information including address and crime type will be scraped to form a table.) and <https://www.neighborhoodscout.com/tx/austin/crime>.

3.Methodology

3.1. Clustering neighborhood using nearby venues

By using the Foursquare API, we can acquire the venues around a certain coordinate within a certain radius. Then we can cluster the neighborhoods based on the numbers/popularities of different venue categories each neighborhood has.

3.1.1. Collect neighborhood list

We can find a list of neighborhoods in Austin here: https://en.wikipedia.org/wiki/List_of_Austin_neighborhoods. We first read in the website content as text using requests module, and then find the neighborhood names using BeautifulSoup package. The list of neighborhood is read into a Dataframe and the total number of neighborhoods is 64.

3.1.2. Obtain coordinates of each neighborhood

When we have the list of neighborhood names, we can import geopy library and use the Nominatim module to get the latitude and longitude. Looping over the list, we read and save the coordinates and attach them to the Dataframe. The updated Dataframe looks like below.

	Neighborhood	Latitude	Longitude
34	Mueller, Austin, Texas	30.296524	-97.700203
35	Far West, Austin, Texas	30.351668	-97.748642
36	Tarrytown, Austin, Texas	30.303125	-97.770485
37	West Congress, Austin, Texas	30.215382	-97.772770
38	West End, Austin, Texas	32.780870	-96.805660
39	Barton Creek, Austin, Texas	30.265638	-97.763966
40	Barton Hills, Austin, Texas	30.251571	-97.784106
41	Bouldin Creek, Austin, Texas	30.255667	-97.755481
42	Dawson, Austin, Texas	30.232926	-97.761418

3.1.3. Clustering neighborhoods

Using Foursquare API, we can request all the venues in all neighborhoods in Austin. Among all the information requested, we only need venue name and venue category for clustering; each venue will become a row in the dataframe as below. The categories include Cafe, Restaurants, Bakery, etc.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bryker Woods, Austin, Texas	30.305246	-97.754585	Tiny Boxwoods	30.306058	-97.749789	American Restaurant
1	Bryker Woods, Austin, Texas	30.305246	-97.754585	Kerbey Lane Café	30.308030	-97.750470	Café
2	Bryker Woods, Austin, Texas	30.305246	-97.754585	Olive & June	30.307450	-97.751046	Italian Restaurant
3	Bryker Woods, Austin, Texas	30.305246	-97.754585	Tiny's Milk And Cookies	30.305971	-97.749950	Bakery
4	Bryker Woods, Austin, Texas	30.305246	-97.754585	Austin Flower Delivery	30.307787	-97.751224	Flower Shop
5	Bryker Woods, Austin, Texas	30.305246	-97.754585	Anderson's Coffee Co	30.308382	-97.750355	Coffee Shop
6	Bryker Woods, Austin, Texas	30.305246	-97.754585	Brykerwood Veterinary Clinic	30.305978	-97.749611	Veterinarian

Realizing that some neighborhoods may have very few venues which can not be fairly clustered, we need to check the number of venues obtained by each neighborhood. Below is the counts of venues for each neighborhood sorted in ascending order.

Neighborhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Spyglass-Barton's Bluff, Austin, Texas	3		3	3	3	3
The Ridge at Lantana, Austin, Texas	4		4	4	4	4
Onion Creek, Austin, Texas	5		5	5	5	5
Shady Hollow, Austin, Texas	6		6	6	6	6
Southeast Austin, Austin, Texas	6		6	6	6	6
Canyon Creek, Austin, Texas	7		7	7	7	7
Granada Hills, Austin, Texas	8		8	8	8	8
Dove Springs, Austin, Texas	10		10	10	10	10
South Manchaca, Austin, Texas	13		13	13	13	13
Copperfield, Austin, Texas	13		13	13	13	13

we can see that there are 11 neighborhoods which have only acquired less than 15 venues. The least number of venues is 3 which is of Spyglass-Barton's Bluff;and The Ridge at Lantana only has 4. Checking the map, we can see that these neighborhoods are far away from Austin city. So let's exclude these 11 neighborhoods from our list and now we have 53 neighborhoods to cluster.

We can convert the venue list into the frequency for each venue category which can be used to cluster. In order to do it, we first one-hot encoding the table so that each venue category is represented in one column; then by grouping by neighborhood and calculate the group mean we obtain the frequency of each venue category for each neighborhood per row.

	Neighborhood	ATM	Adult Boutique	African Restaurant	American Restaurant	Antique Shop	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Astrologer	Athletics & Sports
0	Allandale, Austin, Texas	0.000000	0.000000	0.000000	0.000000	0.021739	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Balcones Woods, Austin, Texas	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.026316
2	Barrington Oaks, Austin, Texas	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.048780	0.000000	0.000000
3	Barton Creek, Austin, Texas	0.000000	0.000000	0.000000	0.030000	0.000000	0.00	0.000000	0.00	0.010000	0.000000	0.000000	0.000000	0.000000	0.020000

Now the data is ready to be clustered! Using the KMeans module in sklearn library, we can now cluster the neighborhoods based on the frequencies of different venue types. The proper number of clusters need to be tested since too many or too less would make the clustering less useful and representative. After some trial, 6 clusters are decided as the best one.

3.2. Analyze each cluster

By checking the most popular venues in neighborhoods, we can have a better idea what kind of neighborhoods are being clustered into one group and decide on the cluster to choose. We can first limit the venue categories into a narrow list of interest and see the typical frequencies of these categories in each cluster.

3.3. Collect house price information

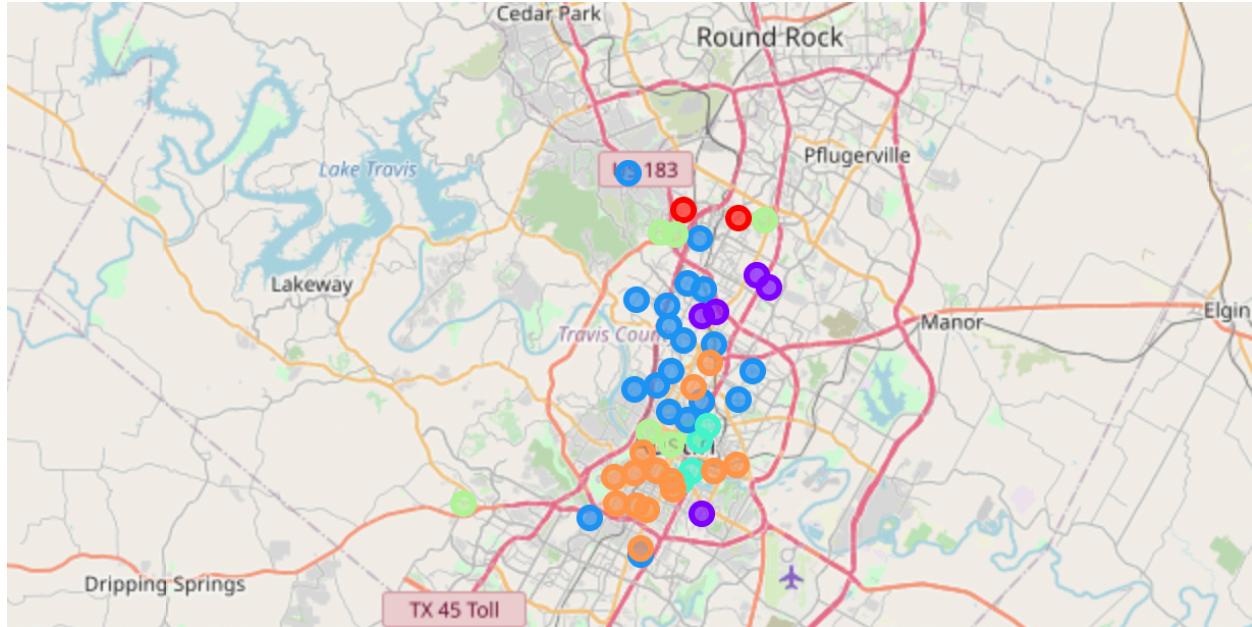
We end up choosing the cluster 2 and also narrow to the few neighborhoods along the Colorado River which will have a good view. The house price information in a table format ready to be read in and use are not easily available after some search; so I decided to manually read in the needed data from <http://www.city-data.com/nbmaps/neigh-Austin-Texas.html#N121>. Since we have narrowed our pool into less than 10 neighborhoods, we can easily collect different information about them from more resources and have a direct comparison.

4. Results and Discussion

4.1. Neighborhood clustering

Using KMeans method we cluster the 53 neighborhoods into 6 groups. The clustering was also tested using all 64 neighborhoods and was found to result in 5 groups with only one neighborhood inside while most other neighborhoods were in one large group. It is not effective to cluster the neighborhoods with some extremely different neighborhoods. So the neighborhoods with Plotting the location of each neighborhood in the map, we can see the neighborhood with the cluster

marked as different colors in the map, shown below. Each group has at least 2 neighborhoods and no single color or group dominates. This could be a good sign.



Cluster 0 is the red dot which only have 2 and are located in the far north.

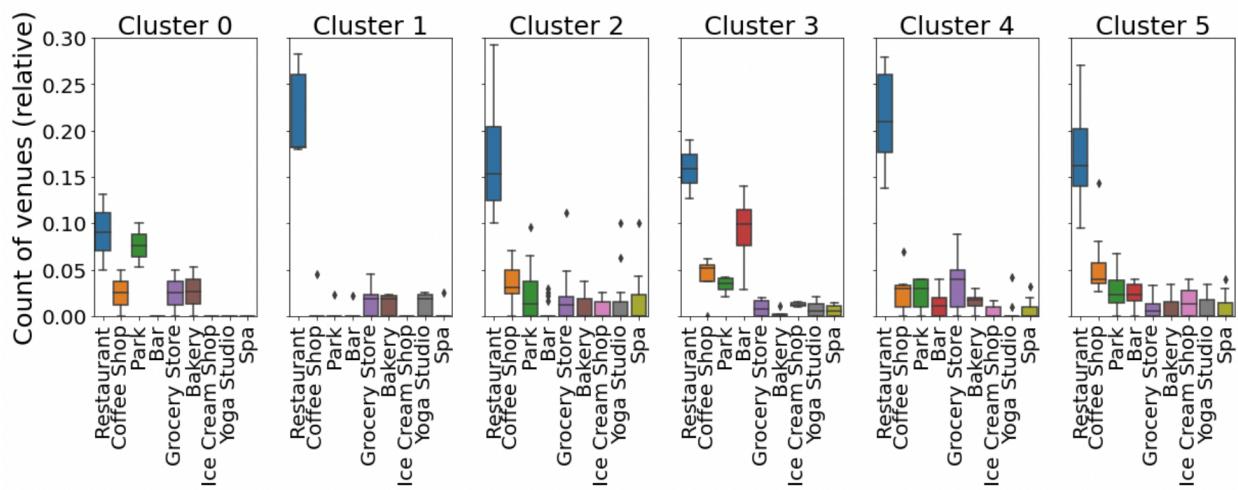
Cluster 1 is the purple one. The light blue and green are cluster 3 and 4, respectively, which mostly located around city center. Cluster 2 and 5, the blue and the orange, have the most number of neighborhoods and scatter everywhere around the city. Now we can go deeper and analyze the characteristics of each cluster to decide which cluster meets our interest and need.

4.2. Cluster analysis

We can group by venue category and count venue numbers for each category, then find the most common venue Caterogies for each neighborhood and put them in a DataFrame as below.

Neighborhood	Latitude	Longitude	Cluster Labels	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
Bryker Woods, Austin, Texas	30.305246	-97.754585	2	Sandwich Place	Park	American Restaurant	Coffee Shop	Café	Spa	Pharmacy	Italian Restaurant	Trail	Men's Store
Hancock, Austin, Texas	30.295896	-97.724768	2	Sandwich Place	Park	Mexican Restaurant	Grocery Store	Fast Food Restaurant	Coffee Shop	Music Store	Salon / Barbershop	Gym / Fitness Center	American Restaurant
Pemberton Heights, Austin, Texas	30.249730	-97.738826	3	Food Truck	Taco Place	Park	Indian Restaurant	Convenience Store	Bar	Lake	Pizza Place	Boutique	Vegetarian / Vegan Restaurant
Rosedale, Austin, Texas	30.313390	-97.744898	2	Coffee Shop	Mexican Restaurant	Clothing Store	Italian Restaurant	Bakery	Sandwich Place	Gift Shop	Mobile Phone Shop	Pharmacy	Food Truck
Clarksville, Austin, Texas	30.277680	-97.759807	4	American Restaurant	Yoga Studio	Grocery Store	Mexican Restaurant	Furniture / Home Store	Spa	Clothing Store	Pet Store	Park	Trail

However, it's still not convenient to check each neighborhoods individually and we cannot directly conclude the characteristics of different clusters. A better way to show it is by plotting out the numbers of major venue categories. The 'major venue categories' here being selected are the ones that I personally care the most about, while in the same while I think it can represent the interests of similar age and background (young professionals, from Asian). The 10 venue categories got selected here are restaurant, coffee shop, park, bar, grocery store, bakery, ice cream shop, yoga studio, and spa. Here below shows the box plot of these categories for the 6 clusters.



First we see that restaurant always appears as the most popular venues in all clusters; this is majorly because all different types of restaurants are being combined into one group for counting. And it is also true that restaurants of course are the most popular type of venues if you go around. Similarly, grocery stores exist in all 6 clusters at similar level so is not effective in distinguishing different clusters.

Second, it can be seen that cluster 0 have the least restaurants (relatively) and don't have many ice cream shops, yoga studios, or spas; in the same while it has

the highest number of parks. It makes sense that the neighborhoods are located far away from the city center; it's close to nature but may not provide a convenient life. Similarly, it can be seen that cluster 1 lacks coffee shops, ice-cream shops, or spas; while in the contrary to cluster 0, cluster 1 doesn't have many parks either. Checking the locations of cluster 1 neighborhoods (purple) in the map, they are also too far away from city center and many are close to cluster 0 neighborhoods. Since coffee, ice-cream, and yoga mean a lot to us, either these two clusters may not be a good option.

Cluster 3 doesn't have many bakeries and cluster 4 lacks yoga which also decrease their attractiveness especially under the circumstance where cluster 2 and 5 don't have anything missing. Even though cluster 2 doesn't have many bars, living very close to bars may not be a wise idea to some people. Cluster 3 has the highest number of bars, which also corresponds to the fact that these neighborhoods are all around downtown area and are the most busy places. It is a good place to hang out with friends but living nearby may not be very safe; plus houses there may be too expensive.

So cluster 2 and 5 show up as the potentially best clusters to go for. They provide all the venues we most need, and if we check their locations in the map they are everywhere around the city which leaves us a lot of options. Now we can take a closer look at only these two neighborhoods. Using the same method, we can find the neighborhoods in each cluster and list their most common 10 venue types. Note that in there the venue Categories are the original venue categories labeled by Foursquare and restaurant types are detailed divided into Mexican restaurants, Chinese restaurants, American restaurants, etc. And the table has some limitation since only the most 10 common venue types are displayed and the frequencies are not shown. We can have a glance here to guide some impression. Here below shows the table for cluster 2 neighborhoods. It impressed me that

Neighborhood	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
0 Bryker Woods, Austin, Texas	Sandwich Place	Park	American Restaurant	Coffee Shop	Café	Spa	Pharmacy	Italian Restaurant	Trail	Men's Store
1 Hancock, Austin, Texas	Sandwich Place	Park	Mexican Restaurant	Grocery Store	Fast Food Restaurant	Coffee Shop	Music Store	Salon / Barbershop	Gym / Fitness Center	American Restaurant
3 Rosedale, Austin, Texas	Coffee Shop	Mexican Restaurant	Clothing Store	Italian Restaurant	Bakery	Sandwich Place	Gift Shop	Mobile Phone Shop	Pharmacy	Food Truck
6 Allandale, Austin, Texas	Food Truck	Pizza Place	Park	Pharmacy	Mexican Restaurant	Mobile Phone Shop	Spa	Business Service	Supermarket	Storage Facility
8 Barrington Oaks, Austin, Texas	Korean Restaurant	Asian Restaurant	Grocery Store	Chinese Restaurant	Pizza Place	Mexican Restaurant	Ramen Restaurant	Cajun / Creole Restaurant	Big Box Store	Mongolian Restaurant
9 Battle Bend Springs, Austin, Texas	Hotel	Mexican Restaurant	Spa	Furniture / Home Store	Steakhouse	Gas Station	Tex-Mex Restaurant	Rental Car Location	Mobile Phone Shop	Sandwich Place
10 Brentwood, Austin, Texas	Coffee Shop	Taco Place	Food Truck	Pet Store	Burger Joint	Pizza Place	Gas Station	Sandwich Place	Mexican Restaurant	Liquor Store
12 Highland, Austin, Texas	Chinese Restaurant	Korean Restaurant	Bakery	Liquor Store	Mexican Restaurant	Automotive Shop	Toy / Game Store	Burger Joint	Clothing Store	Movie Theater
13 North Burnet, Austin, Texas	Mexican Restaurant	Shoe Store	Furniture / Home Store	Sandwich Place	Food Truck	Asian Restaurant	Sporting Goods Shop	Spa	Brewery	Burger Joint
17 North Shoal Creek, Austin, Texas	Burger Joint	Sandwich Place	American Restaurant	Coffee Shop	Mexican Restaurant	Pharmacy	Yoga Studio	Taco Place	Bakery	Gym

Barrington Oaks and Highland both have a lot of assian food. And Westgate has Szechuan restaurants as the 8th most common venue while West Campus has Bubble Tea shop as the 7th most common. I also saw that many have bakery, park, spa in the top 10 list. I like them!

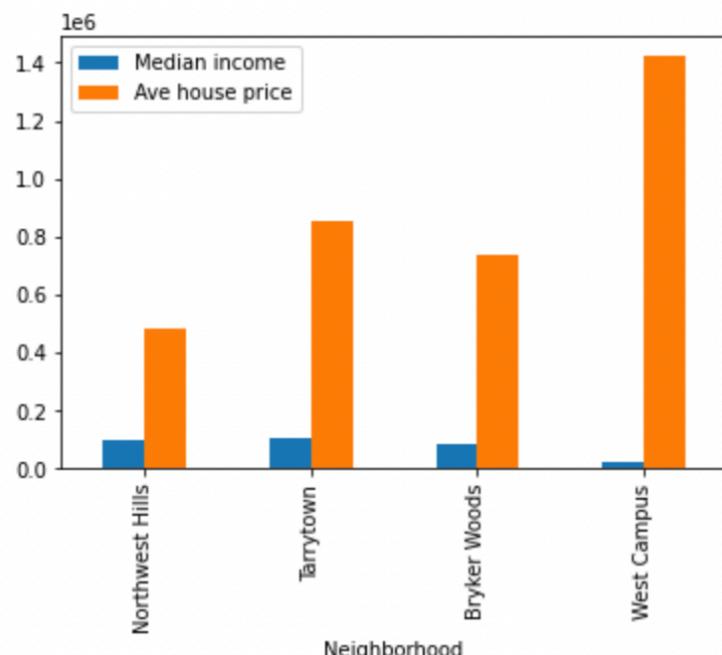
Now let's check the list for cluster 5, shown below. I saw a lot of Mexican restaurants, bars and stores; but not many Asian restaurants I may like better and also not many bakeries. After checking the common venue lists of the cluster 5, I decide that they are much less attractive to me compared to cluster 2, which have a lot of assian food, sweet shops, as well as parks and yoga studios. So I am going to limit my selection to cluster 2.

Neighborhood	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
5 Hyde Park, Austin, Texas	Park	Pharmacy	Coffee Shop	Gym / Fitness Center	Pool	Mexican Restaurant	Bar	Bus Stop	Ice Cream Shop	Art Gallery
16 North Loop, Austin, Texas	Coffee Shop	Pizza Place	Vegetarian / Vegan Restaurant	Convenience Store	Hotel	Cocktail Bar	Thrift / Vintage Store	Clothing Store	Vietnamese Restaurant	Breakfast Spot
29 East Congress, Austin, Texas	Coffee Shop	Liquor Store	Food Truck	Convenience Store	Taco Place	Hardware Store	Gym / Fitness Center	Gym	Mexican Restaurant	Music Store
31 Goolville, Austin, Texas	Mexican Restaurant	Art Gallery	Farm	Coffee Shop	Brewery	Food Truck	New American Restaurant	Gym / Fitness Center	Garden Center	Print Shop
33 Holly, Austin, Texas	Mexican Restaurant	Food Truck	Bar	Convenience Store	Pizza Place	Coffee Shop	BBQ Joint	Brewery	Italian Restaurant	Pub
39 Barton Creek, Austin, Texas	Food Truck	Gym / Fitness Center	Park	Trail	American Restaurant	Coffee Shop	Spa	Theater	Yoga Studio	Clothing Store
40 Barton Hills, Austin, Texas	Taco Place	Burger Joint	Yoga Studio	Thrift / Vintage Store	Tex-Mex Restaurant	Bar	Tapas Restaurant	Bookstore	Scenic Lookout	Sandwich Place
41 Bouldin Creek, Austin, Texas	Coffee Shop	Burger Joint	Park	Salon / Barbershop	Food Truck	Ice Cream Shop	Restaurant	Gym / Fitness Center	Japanese Restaurant	Mexican Restaurant
42 Dawson, Austin, Texas	Mexican Restaurant	Food Truck	Convenience Store	Dive Bar	Taco Place	Coffee Shop	Grocery Store	Gym	Yoga Studio	Café
43 Galindo, Austin, Texas	Convenience Store	Food Truck	Residential Building (Apartment / Condo)	Mexican Restaurant	Platform	Trail	Taco Place	Lawyer	Tennis Court	Flower Shop

Cluster 2 includes still a large number of neighborhood. Let's see if we can find some good options near the Colorado river; closing to Colorado river means a good view and it will be pretty convenient if we want to take a walk or have a relaxing Sunday afternoon in the river. The few neighborhoods in cluster 2 that are along the river are Northwest Hills, Tarrytown, Bryker Woods, West Campus. Let's further compare among these four neighborhoods with more house price information.

4.3. House price comparison

House price information can be collected from different resources including government database and real estate sales websites. We decided to use the data provided by <http://www.city-data.com/nbmaps/neigh-Austin-Texas.html#N121>. After checking the website in detail, we can see that inside each neighborhood, different types of information are provided, which can make the flow complicated. Since we have narrowed into just a few neighborhoods. We can easily collect the information and manually read the in. Among the available information, average house price and median house income may be some good indicators to check. So let's plot them here in a simple bar chart as below.



It can be seen that for Northwest, Tarrytown, and Bryker Woods, the average house prices are proportional with median house income. This is pretty normal. For West campus, most houses are for renting to students at university, which explains why the median income is low while house price is so high. While it's a good place for investment, the price is too high, plus it's not economic to live there ourselves. Let's consider it a good investment opportunity when we get richer. And with the same reason, Northwest Hills are most affordable for us! Lastly I checked the crime rate, it's also relatively safe compared to other neighborhoods near city center. That's our final target area--Northwest Hills!

6. Conclusion

The neighborhoods of Austin are being clustered into 6 groups. After analyzing each cluster, we have targeted to one single cluster, which provide the services we care the most as young professionals (who are passion about bakery, sweets, and yoga). Adding a little bit more requirement, we narrow it down to the few neighborhoods along the river. And eventually by comparing the house price, we select the best potential neighborhood to buy house in. The neighborhood, Northwest Hills, have all the venues we need, close to the river, and with more affordable house prices.

Note that the house price information used here is year of 2016 and may not correctly reflect the current condition; based on the fact that the house price are rocketing lately, a detailed search and analysis on the house prices including appreciation rate should be more detailedly conducted before a final decision is made. The project is used here as an initial guide to narrow your choices and hope it helps!

