Clustering apartment communities by surrounding venues

Introduction

Background

When it comes to renting an apartment, one of the most important factors people consider is the surrounding venues. In other words, people care a lot about what kinds of venues are in the vicinity of the community. For example, some people would like to live in a busy area with supermarkets and restaurants while others prefer a quiet neighborhood. Therefore, it will be significantly helpful if the information of surrounding venues can be collected and analyzed. Moreover, considering the large number of apartment communities in big cities, such as Houston, it will be overwhelming to go through all the communities of the entire city. So, it will make people's life much easier to segment the apartment communities into a few clusters, so that the communities in a single cluster share similar types of surrounding venues. In this way, a person that is interested in a particular kind of neighborhood can simply look for the corresponding cluster and focus only on the communities in that cluster. Or one can conveniently compare different apartments by checking which cluster each apartment belongs to.

Business Problem

Can we cluster the apartment communities in Houston on the basis of the neighboring venues of each apartment community?

Target audience

The target audience of this project will be the people who are looking for new apartments to rent. Besides, it may also provide useful information for the managers and leasing agents of the apartment communities.

Data

The data needed for this project, along with the sources of the data are described as follows.

List of apartment communities

A full list of the apartment communities in Houston can be obtained by web scraping using Requests and BeautifulSoup. Specifically, the website of Apartment List (https://www.apartmentlist.com/tx/houston) will be scraped. And the name and the address of each community will be retrieved.

Geospatial data (latitude and longitude) of apartment communities

The geospatial data of each apartment community can be acquired by converting the address of each community into the corresponding coordinates using the geocoders provided by the geopy package. If an address in our apartment list cannot be recognized by geopy, that apartment community will be deleted from our dataset. Google Map may also be referred to for data validation.

Information of surrounding venues

Information of the surrounding venues can be conveniently obtained through the Foursquare API. The explore endpoint will be used. And venues within 750 meters of each community will be retrieved. The retrieved information includes the name of the venue, the category of the venue and the geo coordinates of the venue.

Methodology

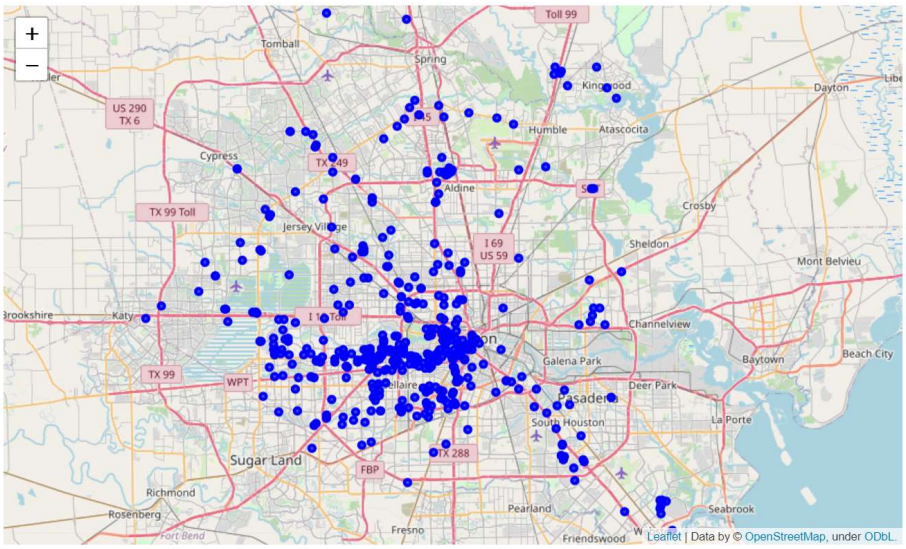
Data cleaning & Exploratory analysis

The initial dataset containing the name and the address of each apartment was obtained from https://www.apartmentlist.com/tx/houston by web scraping. Then, duplicated records were deleted. And a few incorrect addresses were revised based on the information provided by Google Map.

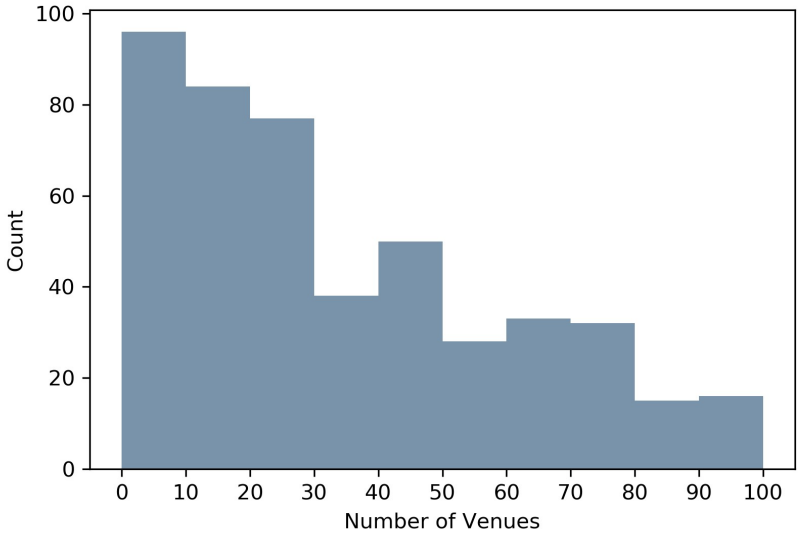
Geopy was employed to convert the address of each apartment community to the corresponding coordinates. In particular, since the number of records in the data set was quite large, the rate limiter class (geopy.extra.rate_limiter.RateLimiter) was used to add delays between geocoding calls to reduce the load on the Geocoding service, and to retry failed requests and swallow errors for individual rows. After collection, the geospatial data were combined with the initial apartment dataset. Some addresses were not recognized by the geocoder, and no coordinates were retrieved for those records. Several other addresses were converted to coordinates, but the coordinates were not correct because they are outside the geo boundary of Houston. The records with the above invalid values were removed from the dataset. The clean dataframe is as follows.

	Apartment Name	Address	Latitude	Longitude
0	Timber Ridge	12200 Fleming Dr, Houston, TX	29.775678	-95.220987
1	Montierra	2345 Sage Rd, Houston, TX	29.743883	-95.466360
2	Ashton on West Dallas	1616 W Dallas St, Houston, TX	29.758375	-95.383325
3	Summervale	9221 Pagewood Ln, Houston, TX	29.725929	-95.527784
4	Remington Park	5510 S Rice Ave, Houston, TX	29.722547	-95.468239

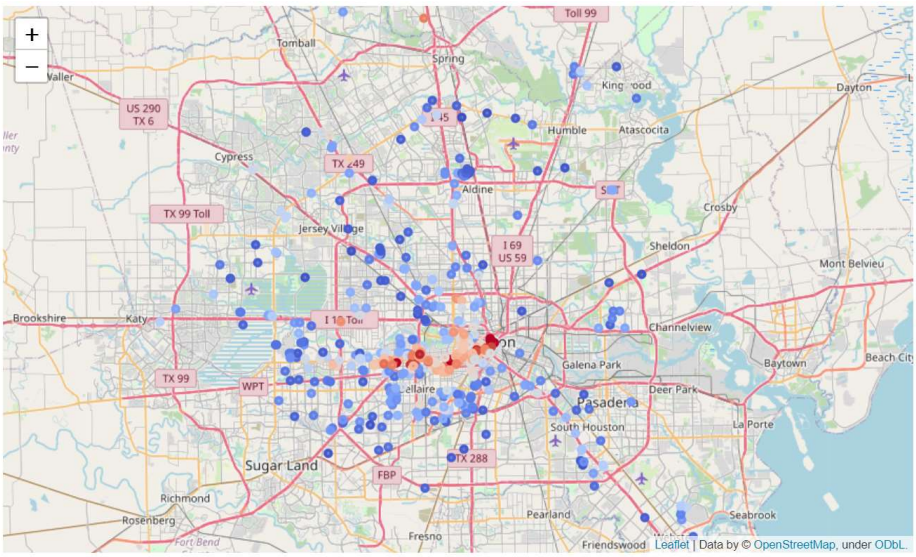
Then, the locations of the apartment communities were visualized using the folium package.



The information of the surrounding venues were collected through the Foursquare API. Venues within 750 meters of each community were retrieved. The data were then transformed into a dataframe. The venues were grouped by the apartment they surround, and the venues within each group were counted. The distribution of the venue number for each community is shown below.



The apartment communities were again visualized with the color as the indicator of the amount of venues around each community. Warmer colors means larger numbers of venues around.



In addition to the number of venues, the top 10 common venue categories were highlighted for each apartment community.

	Apartment Name	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	1300 North Post Oak	Gym / Fitness Center	Coffee Shop	Pool	American Restaurant	Pizza Place	Dry Cleaner	None	None	None	None
1	1624 Holman St	Coffee Shop	Bar	Bakery	Nightclub	American Restaurant	Gas Station	Park	Bookstore	Sandwich Place	Café
2	1900 Yorktown	Bank	Park	Gym / Fitness Center	Chinese Restaurant	Mexican Restaurant	Shipping Store	Bakery	Sandwich Place	Gym	Athletics & Sports
3	1901 Richmond Ave	Park	Food Truck	Mexican Restaurant	Sandwich Place	Pizza Place	Coffee Shop	Pub	Italian Restaurant	Thai Restaurant	Record Shop
4	1919 Portsmouth St	Park	Mexican Restaurant	Food Truck	Seafood Restaurant	Bookstore	Pizza Place	Pub	Coffee Shop	Record Shop	Café

Clustering by the venue categories

KMeans was employed for clustering the apartment communities. Specifically, the clustering was based on the occurrence of each venue category around one community. Regarding the data preparation, the categorical data in the ‘venue category’ column of the venue dataframe were converted into numeric values through one hot encoding. Then the data were grouped by the name of the community. Finally, the occurrence of each category was normalized within every group.

The clustering was conducted with the number of clusters set to 5. After clustering, the apartment communities were grouped by the cluster label. And the top 10 common venue categories of all the communities in each single cluster were examined.

Cluster 0

	Apartment Name	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	Timber Ridge	Seafood Restaurant	Fast Food Restaurant	Fried Chicken Joint	Mexican Restaurant	Men's Store	Grocery Store	BBQ Joint	Sandwich Place	Gym	Clothing Store
4	Remington Park	Fast Food Restaurant	Hotel	Furniture / Home Store	Pool	Health & Beauty Service	Electronics Store	Big Box Store	Grocery Store	Coffee Shop	Gym
16	The Preakness	Fast Food Restaurant	Bank	Mobile Phone Shop	Mexican Restaurant	Rental Car Location	Sandwich Place	Tex-Mex Restaurant	Buffet	Restaurant	Chinese Restaurant
17	La Casita	Furniture / Home Store	Mexican Restaurant	Music Venue	Fast Food Restaurant	Sandwich Place	Gas Station	Video Store	Mattress Store	None	None
33	Lakeshore	Fast Food Restaurant	Vietnamese Restaurant	Spa	Pharmacy	Mexican Restaurant	Shipping Store	Restaurant	Bank	Tea Room	Sandwich Place

Cluster 1

	Apartment Name	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
95	Windmill Landing	Business Service	None	None	None	None	None	None	None	None	None

Cluster 2

	Apartment Name	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
1	Montierra	Hotel	Sandwich Place	New American Restaurant	American Restaurant	Department Store	Jewelry Store	Health & Beauty Service	Burger Joint	Sporting Goods Shop	Hotel Bar
2	Ashton on West Dallas	Pub	Vegetarian / Vegan Restaurant	Bar	Gay Bar	Park	Café	Dry Cleaner	Gym	Auto Garage	River
5	Woods on Lamonte	Dog Run	Dive Bar	Bar	Bank	Sandwich Place	Furniture / Home Store	Park	Grocery Store	Thrift / Vintage Store	Storage Facility
9	Adobe Springs	Dog Run	Dive Bar	Fried Chicken Joint	Pet Store	Mexican Restaurant	Storage Facility	Furniture / Home Store	Bar	Bank	Grocery Store
10	Harvest Hill	Rental Car Location	Pizza Place	Indian Restaurant	Smoke Shop	Grocery Store	Liquor Store	Baseball Stadium	Gas Station	Men's Store	Moving Target

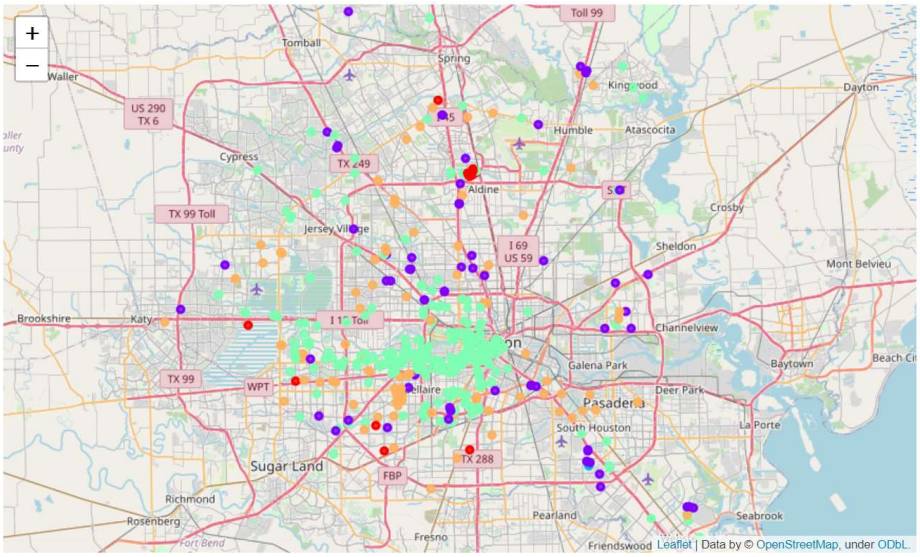
Cluster 3

	Apartment Name	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
3	Summervale	Nightclub	South American Restaurant	Liquor Store	Furniture / Home Store	Gourmet Shop	Caribbean Restaurant	Ethiopian Restaurant	Latin American Restaurant	Taco Place	Mexican Restaurant
6	The Gables at Richmond	South American Restaurant	Indian Restaurant	Latin American Restaurant	Fried Chicken Joint	Nightclub	Gourmet Shop	Caribbean Restaurant	Liquor Store	Ethiopian Restaurant	Food Truck
7	Towne Lake Apartments	Mexican Restaurant	Donut Shop	Bar	Seafood Restaurant	Cosmetics Shop	Movie Theater	Tennis Court	Bakery	Taco Place	Gas Station
8	Sterling Point	Mobile Phone Shop	Mexican Restaurant	Grocery Store	Sandwich Place	Asian Restaurant	Gas Station	Pharmacy	Building	Seafood Restaurant	Fried Chicken Joint
11	Arbor On Richmond	South American Restaurant	Indian Restaurant	Latin American Restaurant	Fried Chicken Joint	Nightclub	Gourmet Shop	Caribbean Restaurant	Liquor Store	Ethiopian Restaurant	Food Truck

Cluster 4

	Apartment Name	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
24	The Millennium Kirby	Gym	Sandwich Place	Park	Soccer Field	None	None	None	None	None	None
34	Breckenridge	Gas Station	Wine Bar	Clothing Store	Discount Store	Plaza	Park	Cosmetics Shop	None	None	None
38	Parc at 505	Park	Furniture / Home Store	Gas Station	Burger Joint	None	None	None	None	None	None
106	San Paloma	Park	Cosmetics Shop	Playground	Bus Station	Home Service	None	None	None	None	None
208	Westbury Crossing	Discount Store	Grocery Store	Mexican Restaurant	Park	Bar	Gas Station	None	None	None	None

At last, the clusters of apartment communities were visualized with different colors representing different clusters.



Results

The average number of surrounding venues for one apartment community is about 33. About 50% of the communities have 26 or fewer venues around. 25% of the communities display a venue number larger than 50.

Communities in the downtown, midtown and uptown districts have the most neighboring venues. In other areas, the number of surrounding venues is considerably lower.

Based on the clustering result, the representative venue categories for each cluster are as follows.

Cluster 0 Fast Food Restaurant, Sandwich Place, Chinese Restaurant, Clothing Store
Cluster 1 Business Service

Cluster 2	Hotel, Coffee Shop, Sandwich Place, Cosmetics Shop
Cluster 3	Mexican Restaurant, Gas Station, Pizza Place, Fast Food Restaurant
Cluster 4	Park, Gas Station, Wine Bar, Plaza

Discussion

The clustering is only based on the ratio between different venue categories. One should also consider the absolute number of venues to get a better idea about the distribution of venues around different apartment communities.