IBM Data Science Capstone

Clustering Apartment Communities by Surrounding Venues

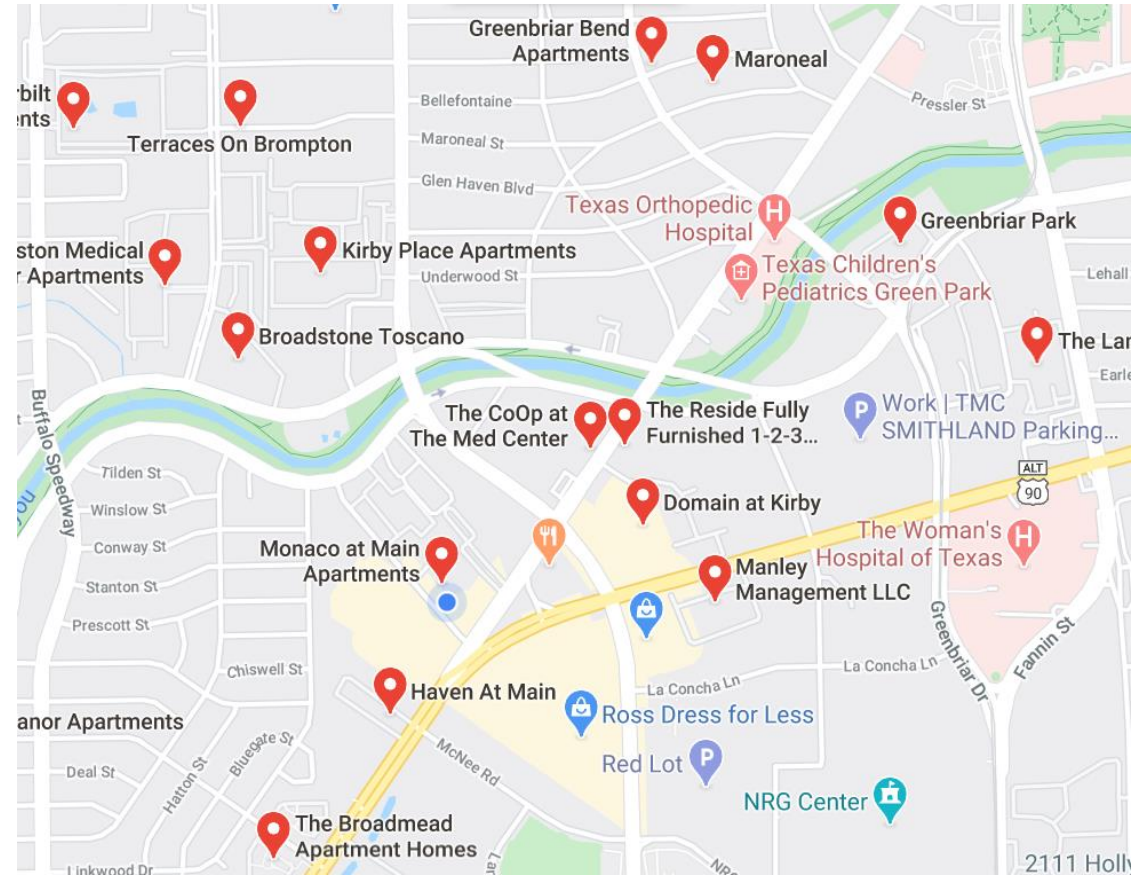**Yan Ding**
April, 2020

Introduction

When it comes to renting an apartment…

What kinds of venues are in the vicinity of the community?

How many venues are there?

Large amount of information is overwhelming. Clustering can help.

Introduction

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 Acquirement

List of apartment communities

A full list of the apartment communities in Houston can be obtained by scraping the website below using Requests and BeautifulSoup.
https://www.apartmentlist.com/tx/houston

Geospatial data (latitude and longitude) of apartment communities

The geospatial data can be acquired by converting the address of each community into the corresponding coordinates using the geocoders provided by the geopy package.
geopy.extra.rate_limiter.RateLimiter
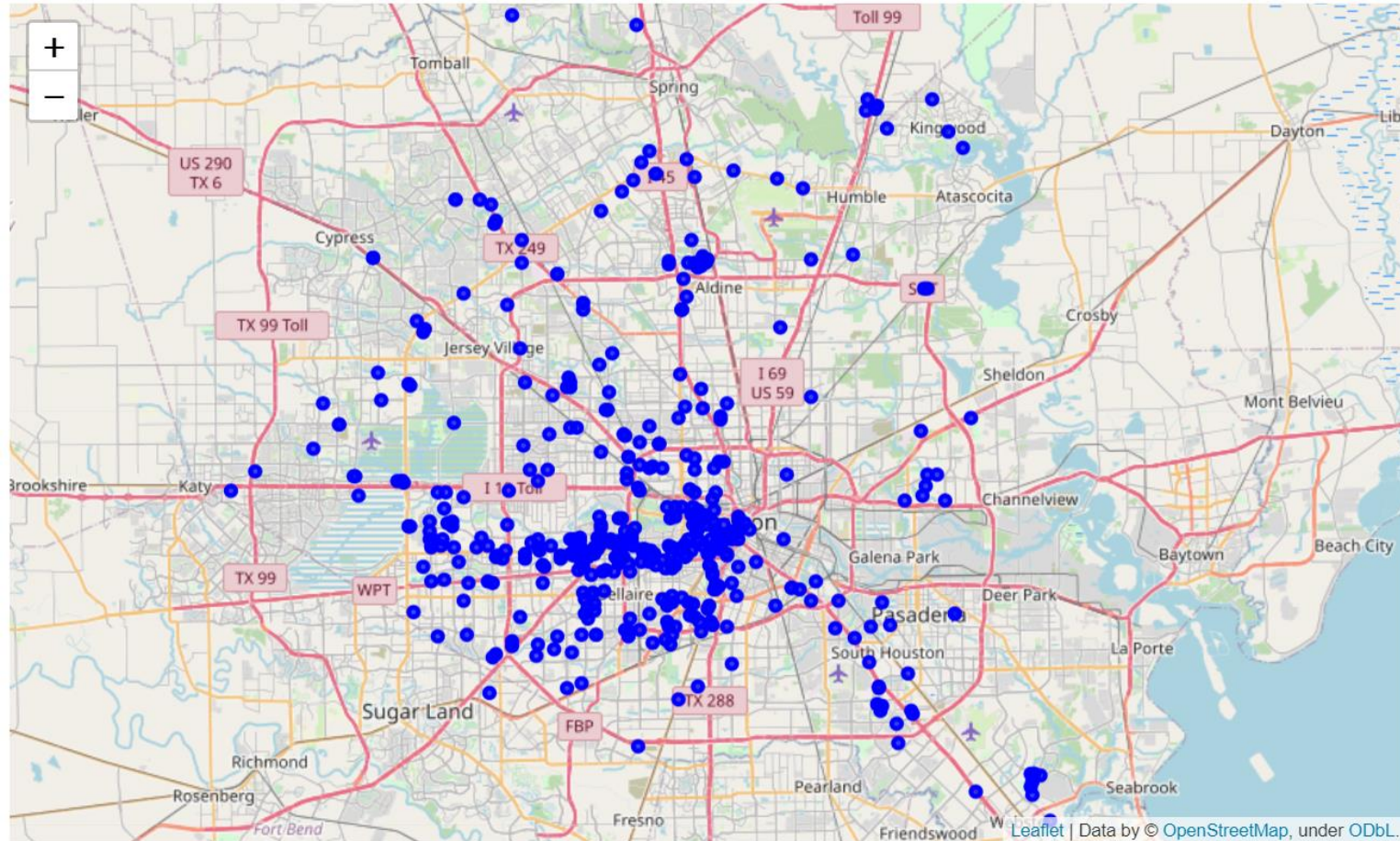
Information of surrounding venues

Information of the surrounding venues can be conveniently obtained through the Foursquare API. Venues within 750 meters of each community will be retrieved.
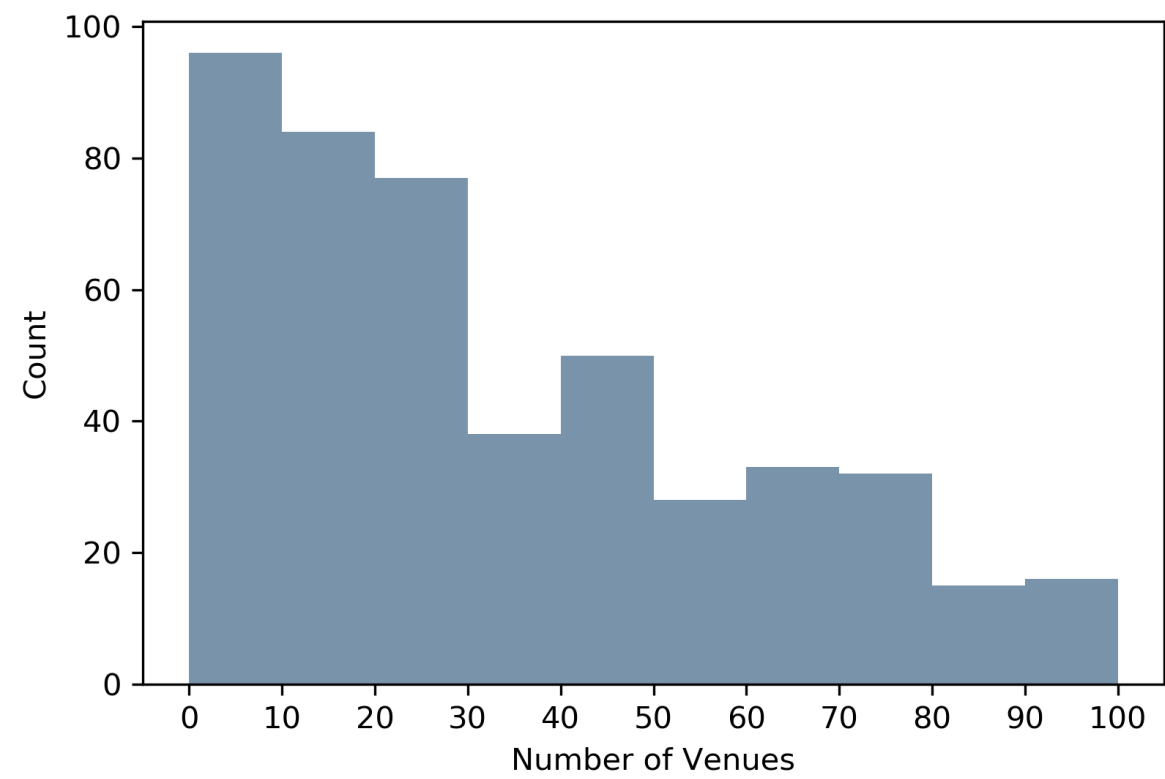Including the name, the category the geo coordinates of the venue.

Exploratory Analysis

Apartment Communities on the Map

Exploratory Analysis

Distribution of the Number of Surrounding Venues

Venue Count	
count	469.000000
mean	33.921109
std	26.119142
min	1.000000
25%	12.000000
50%	26.000000
75%	51.000000
max	100.000000

Exploratory Analysis

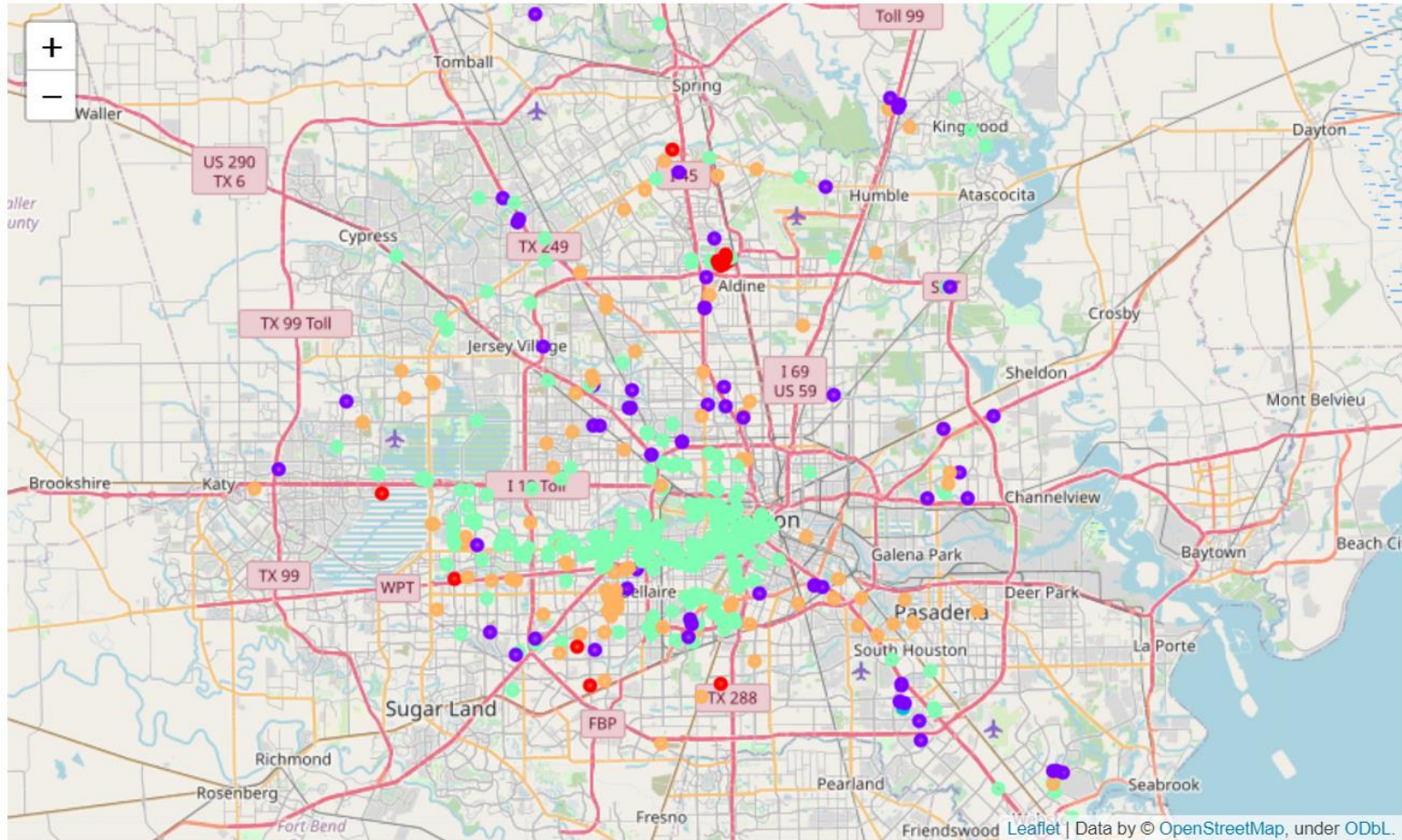## Number of Surrounding Venues against Apartment Location

Exploratory Analysis

Top 10 Common Venue Categories 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 with KMeans

Visualization of the Clusters

Clustering with KMeans

Most Common Venue Categories within Each Cluster

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