

# **The Capstone Project - Battle of the neighborhoods**

## **Identifying areas for real estate investment**

### **1. Introduction**

#### **1.1. Background**

Real estate is one of the fastest growing markets in the United States today. The residential housing market has seen unprecedented growth in the last year, fueled by demands caused by the pandemic (people opting to work from home, moving to cheaper cities etc). As the residential real estate market grows, commercial real estate market follows suit. The upcoming housing communities need supporting commercial establishments such as coffee shops, restaurants, office spaces, preschools etc. around them. Therefore, this market area presents a great avenue to invest and grow for someone who is looking for opportunities in commercial real-estate.

#### **1.2. Problem**

The potential investors might not be physically present at the location of interest to know where such growth pockets exist. Usually, either they spend a lot of time exploring the areas on their own or send a real-estate agent to these locations. Both of which are time consuming and expensive affairs. In this project, I aim to present a data driven analysis to identify the upcoming communities and the commercial establishments around them (using Foursquare API). The goal of this project is to give a bird's eye view of the growth of a particular geographical location and also identify location with high growth potential for investment.

#### **1.3. Interest**

This study will be of interest to anyone who wants to invest in real-estate commercial establishments in an upcoming market. It could be an individuals, group of individuals or a real-estate company. This study is aimed to present the landscape of investment opportunities of a location remote to the investor and is most useful to someone who wants to learn about an upcoming remote city. In this study I focus on commercial establishments but it can easily be extended to other areas of interests such as land, residential dwellings etc.

### **2. Data Collection**

For this study I picked Williamson county in Texas, which is the third fastest growing county in Texas. I did not want to pick the top most counties as those counties would already have seen enough growth. The idea is to pick markets that have high potential to grow. Further, I looked at

the fastest growing cities in Williamson county, this information was obtained from wikipedia. Again to ensure highest growth potential opportunities, I left the top 2 cities and picked the third largest city in Williamson county, 'Georgetown' for this study. Next step was to look at where new housing communities were coming up in Georgetown. This information was obtained from Williamson County Association of Relator's website, it had names and addresses of the new and upcoming communities. Once I had these addresses, I cleaned them and used google API to obtain the location coordinates for those addresses. As the last step , I used the address coordinates of the new housing communities in Foursquare API to obtain venues around them.

## 2.1. Data Sources

### 2.1.1.Wikipedia

I used Beautiful soup to scrape the wikipedia page for Williamson County. From there I obtained the information of top 5 largest cities in Williamson County. The information was from 2018 and the present day numbers might be higher. The following table lists the top 5 cities that I got from scraping the wikipedia page for Williamson county.

Largest Cities in Williamson County		1950	1960	1970	1980	1990	2000	2010	2018 (estimate)
#									
1	Round Rock	1,438	1,878	2,811	11,812	30,923	61,136	99,887	128,739
2	Cedar Park	202	385	692	3,474	5,161	26,049	48,937	76,999
3	Georgetown	4,951	5,218	6,395	9,468	14,842	28,339	47,400	74,180
4	Leander	-	-	-	2,179	3,398	7,596	26,521	56,111
5	Hutto	n/a	400	545	659	630	1,250	14,698	26,434

### 2.1.2. New housing communities in Georgetown

The new and upcoming housing communities of Georgetown were obtained from Williamson County Association of Relator's website '[wclife.org](http://wclife.org)'. The website lists 18 new and upcoming communities in Georgetown.

### 2.1.3.Foursquare API

The scale of development around the new housing communities is gathered by using Foursquare API. The venue data surrounding each community is used to asses the existing businesses and the potential opportunities to invest in that part of the town.

## 2.2. Formatting the data

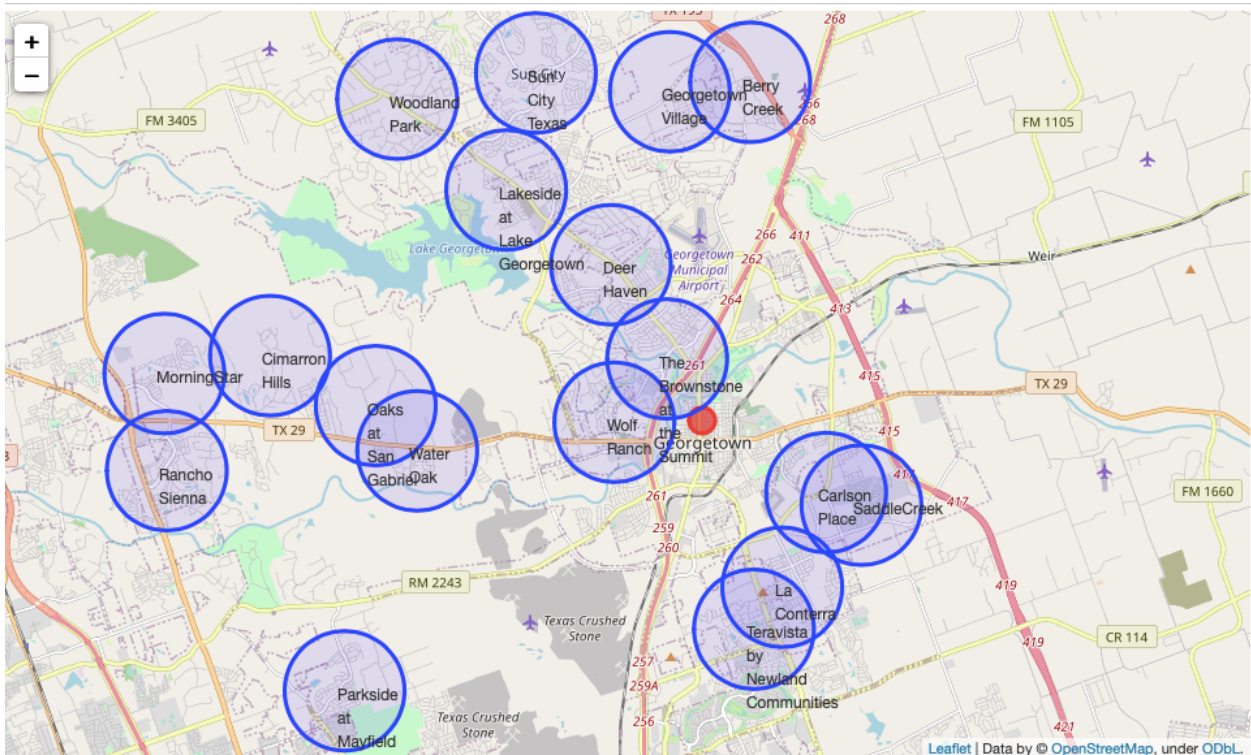
The new housing community data was scraped from [wclife.org](http://wclife.org) website and a dataframe was created as shown below:

	Communities	Address
0	Berry Creek	30325 Tiger Woods Dr.   Georgetown, TX 78628
1	Carlson Place	Rockride Lane and Carlson Cove   Georgetown,...
2	Cimarron Hills	307 Flint Ridge Trail   Georgetown, TX 78628
3	Deer Haven	3816 Williams Drive   Georgetown, TX 78633
4	Georgetown Village	Shell Road   Georgetown, TX 78626

The 'Address' column was cleaned to remove the special characters and using Google API, the location information for each address is obtained. The dataframe was updated with location information as below:

	Communities	Address	latitude	longitude
0	Berry Creek	30325 Tiger Woods Dr. , Georgetown, TX 78628	30.714406	-97.664709
1	Carlson Place	Rockride Lane and Carlson Cove , Georgetown,...	30.621265	-97.644688
2	Cimarron Hills	307 Flint Ridge Trail , Georgetown, TX 78628	30.652358	-97.791276
3	Deer Haven	3816 Williams Drive , Georgetown, TX 78633	30.672989	-97.701429
4	Georgetown Village	Shell Road , Georgetown, TX 78626	30.712256	-97.685876
5	La Conterra	358 Grand Junction Trl , Georgetown, TX 78626	30.599753	-97.656252
6	Lakeside at Lake Georgetown	113 Lake Spring Circle , Georgetown, TX 78628	30.689969	-97.729009
7	MorningStar	101 Ridgeview Court , Georgetown, TX 78628	30.648283	-97.819130
8	Oaks at San Gabriel	1211 Terrace Drive , Georgetown, TX 78628	30.641011	-97.763402
9	Parkside at Mayfield Ranch	101 Garner Park Cove , Georgetown, TX 78628	30.576220	-97.771617
10	Rancho Sienna	404 Via De Sienna Blvd , Georgetown, TX 78628	30.626230	-97.818482
11	SaddleCreek	8605 Daisy Cutter Crossing , Georgetown, TX ...	30.618414	-97.635369
12	Sun City Texas	1501 Sun City Blvd , Georgetown, TX 78633	30.716527	-97.721194
13	Teravista by Newland Communities	1000 Wildflower Park Drive , Georgetown, TX ...	30.590221	-97.663704
14	The Brownstone at the Summit	204 Adams St , Georgetown, TX 78628	30.651620	-97.686597
15	Water Oak	2609 Water Oak Parkway , Georgetown, TX 78628	30.630742	-97.752472
16	Wolf Ranch	Wolf Ranch Parkway , Georgetown, TX 78626	30.637230	-97.700509
17	Woodland Park	101 W. Red Bud Meadow , Georgetown, TX 78633	30.710592	-97.757756

The image below is the visualization of the communities on a map, created using folium. The red center in the middle is downtown Georgetown. As expected the new and upcoming housing communities are all coming around the downtown area.



### 3. Methodology and Analysis

Now that we have the location data for the new communities, in this section we will discuss the methodology and analysis performed. I started my analysis from the downtown area so as to not to exclude investment opportunities in the downtown area. To evaluate the kind of businesses present in the downtown area and to get a basic idea of investment opportunities present in that part of town. This is further narrowed by selecting a particular business, a coffee shop around downtown Georgetown.

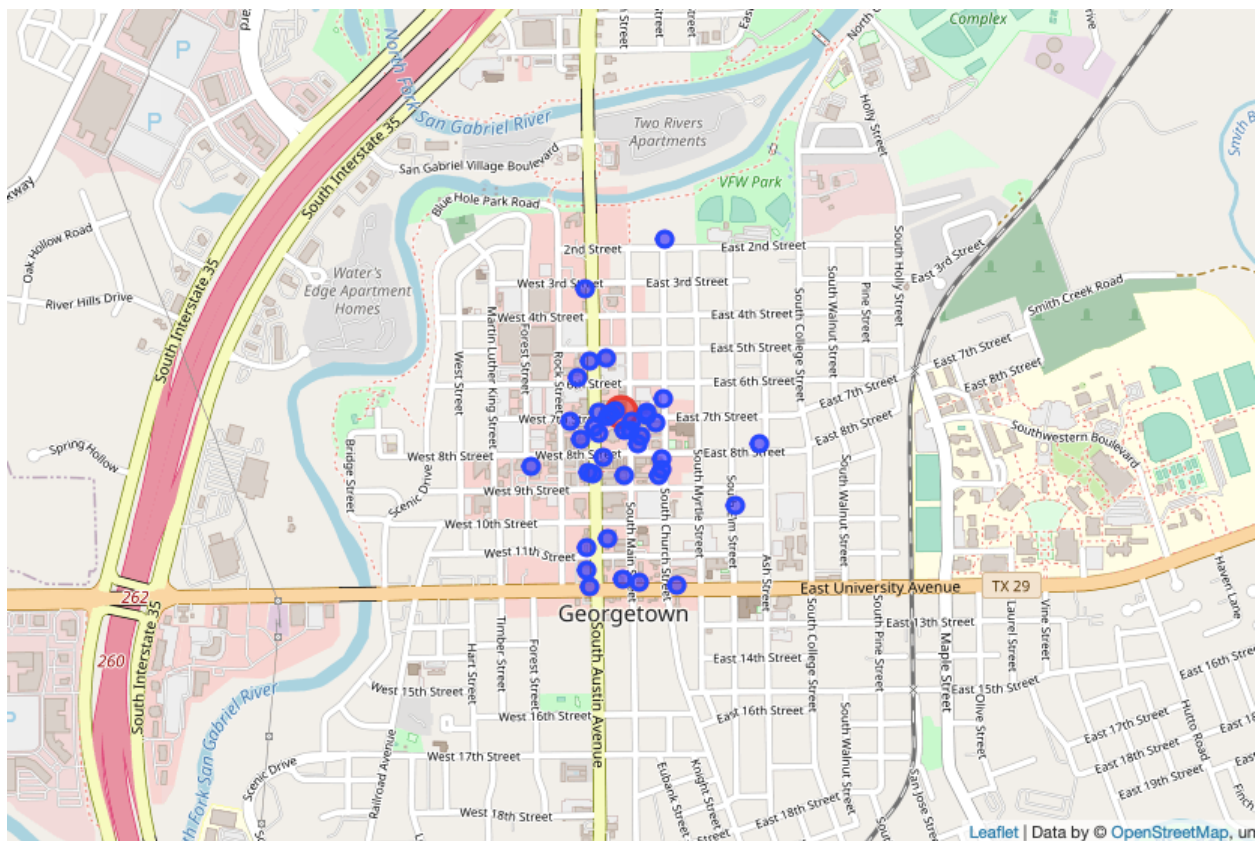
The same analysis is carried out for the new housing communities. First I gathered all the venues near the new and upcoming communities in Georgetown using Foursquare API. As we are looking for any commercial real-estate opportunities, this part includes all the venues. Next I applied K-mean clustering to these venues to divide them into clusters. This is done to identify the areas that have clusters with less number of venues. The clusters/areas with less number of venues are the areas of interest for us because: a. A business is considered to have high potential

if it has less competition around. B. Real-estate values go up as the city develops, so one would like to invest before this growth happens and maximize the return from the investment. Because of these two reasons and based on the classification of venues I identified communities that have less (relatively) venues and can be marked as the favorable places to invest. Then I sorted the venues and picked two specific business categories, coffee shops and pizza places. Then with the help of Foursquare API I checked how one would fare if one were to invest either in a coffee shop or pizza place in those areas.

## 4. Results and Discussion

### 4.1. Georgetown Downtown

As the new housing communities cover pretty much the entire Georgetown area except for the downtown. To keep the analysis comprehensive venues in the downtown area are explored using Foursquare API with the radius of 500 meters. The following is the image of venues from the downtown area.

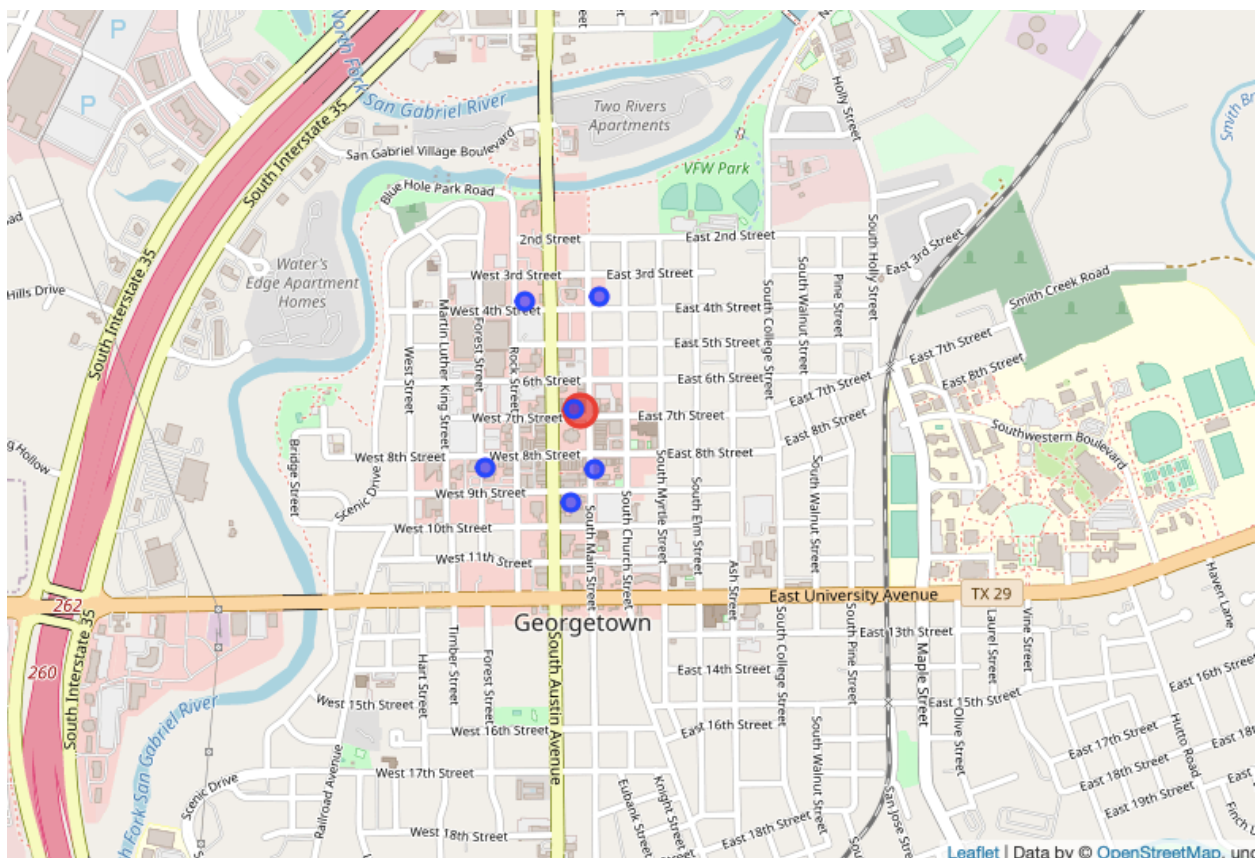


There is not much scope for new construction in the downtown area as it is limited by land availability. So there may not be many investment opportunities for new properties but if someone is interested in an existing space with a specific business, we can further explore this area. To this end I picked Coffee shops as the business of choice and carried out a Foursquare



API venue query for “Coffee”. Below is the list of venues that offer coffee, followed by the map of the venues.

```
0      Red Poppy Coffee Co.  
1      Cianfrani Coffeehouse  
2      Lamppost Coffee Georgetown  
3      Je Suis Coffee  
4      Jailhouse Coffee Shoppe  
5      309 Coffee  
Name: name, dtype: object
```



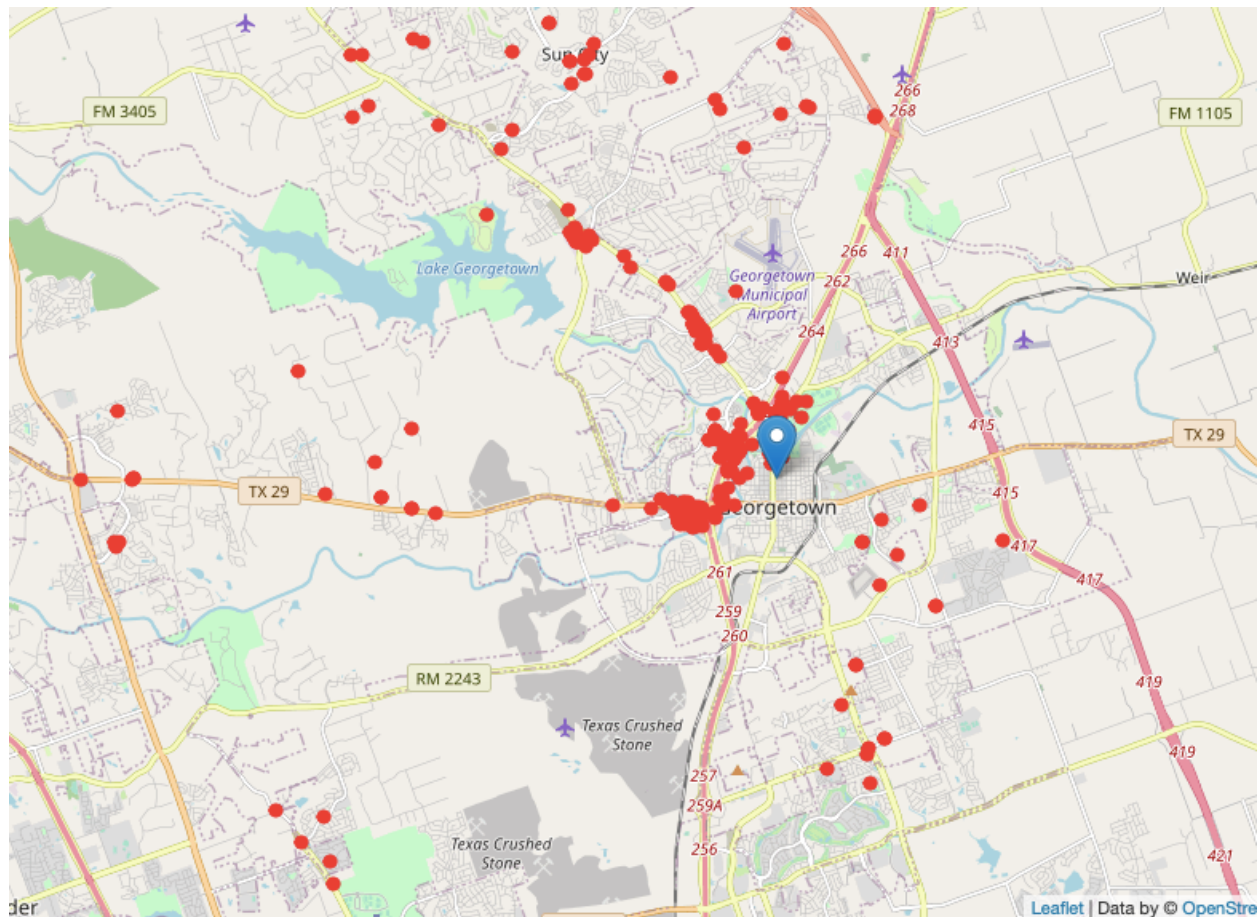
There are a few coffee shops present in the downtown Georgetown area but still there is scope to invest. Any investment opportunity, either to own or rent a coffee shop in that part of the town will be profitable assuming with growth (down the lane) the demand will also increase. Since

this is in the developed part of the town, the real estate prices to own/rent will be high due to limited availability and the growth margins may be lower.

## 4.2. New Housing Communities

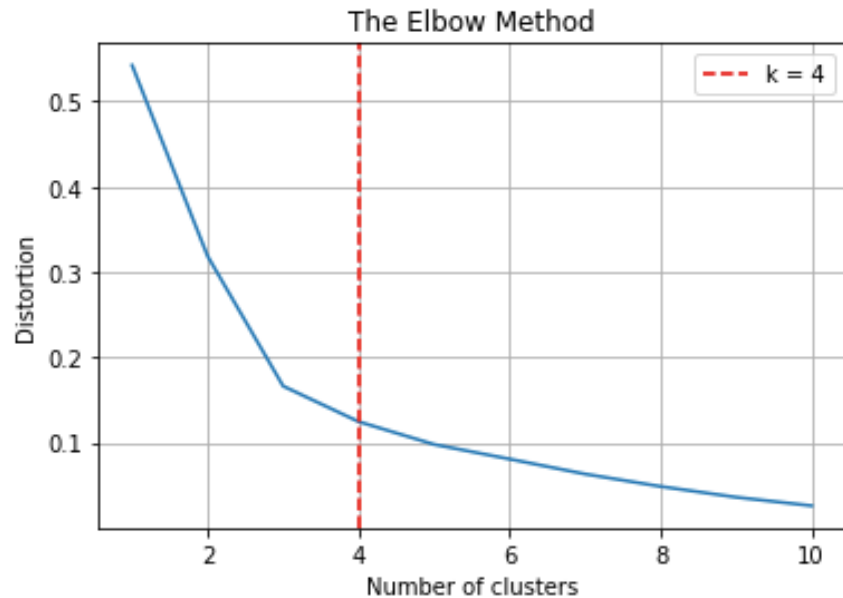
### 4.2.1. Foursquare API venue data for new housing communities

The next step to carry out a similar analysis around the new housing communities of Georgetown city. The venue data for the housing community from Foursquare API is obtained by giving the location information of each community with a radius of 1500, to make sure that no venues are left out. I tried different values of 'radius', these being new communities there were not many venues available and Foursquare did not give results for smaller radii. After trying a few 'radius' values and checking the results, 1500 meters seemed optimum for this study. Although there were some overlaps in the areas the venue data was not compromised too much. With this 'radius' a total of 263 venues were found around all the 18 communities put together, which are shown in the figure below.

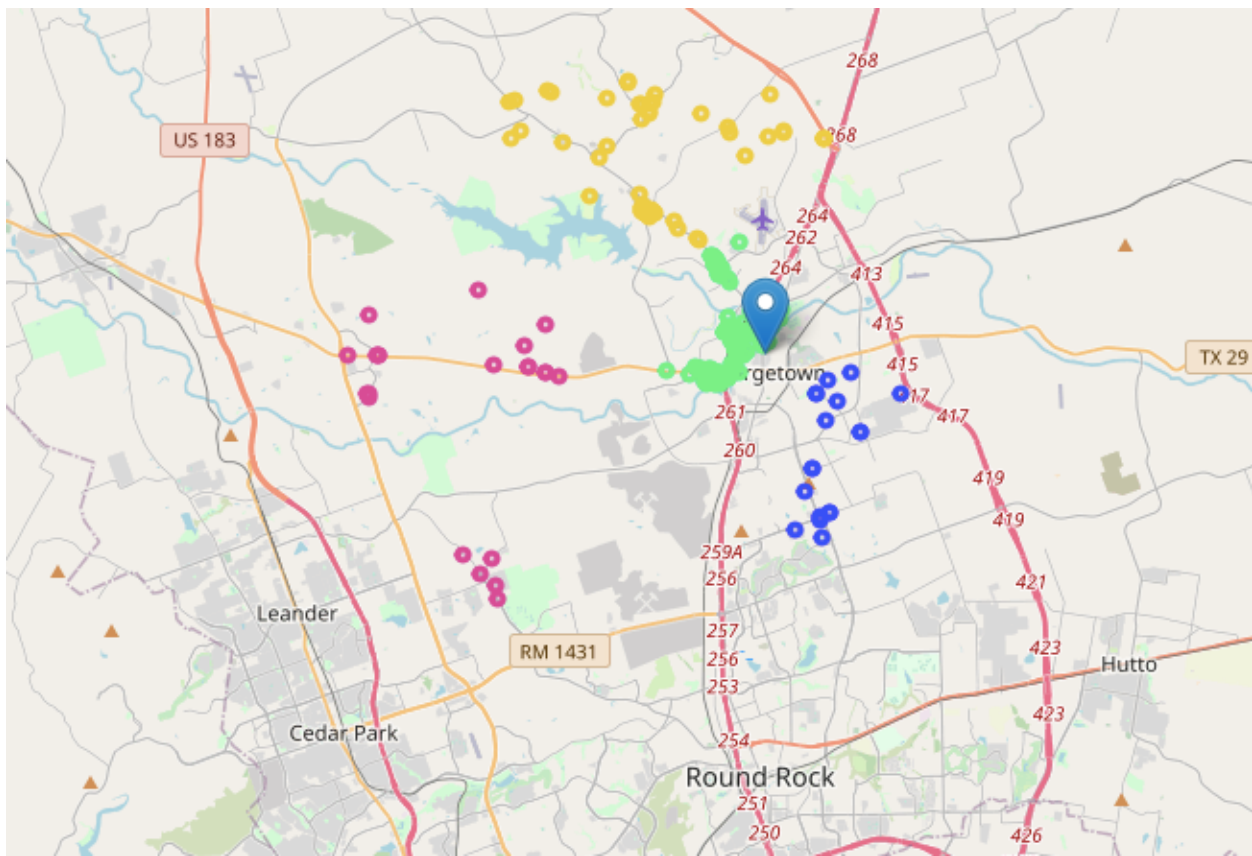


#### 4.2.2. K-Means clustering

Before applying the K-Means clustering to the venue data, an optimum value of 'k' was found using the 'Elbow Method' as shown in the picture below. Based on this, a value of 'k' = 4 is picked and the venues are classified using K-Means.

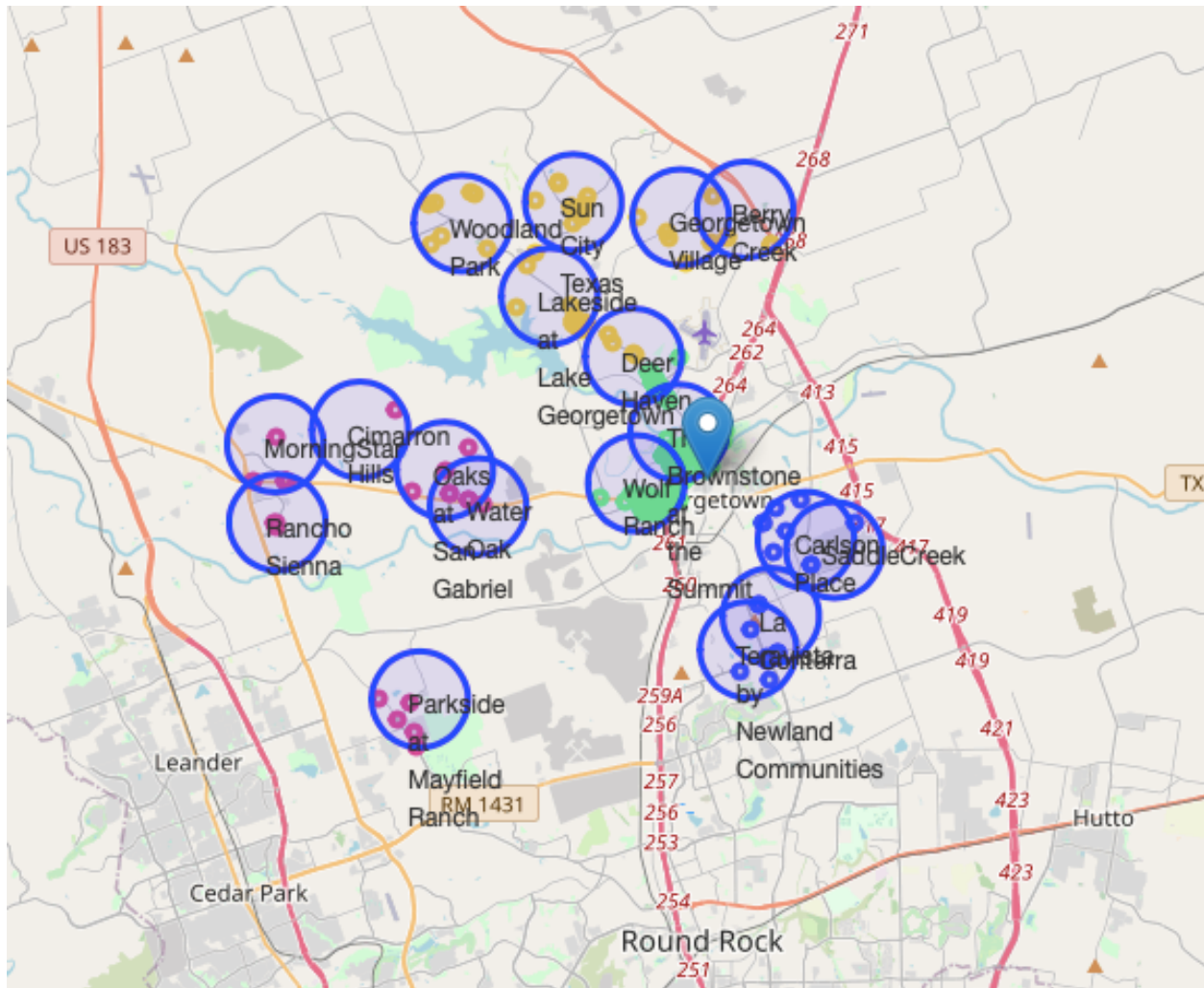


The map of the venues after classification is as follows:





As expected the cluster near downtown Georgetown (pin) is the most populated, making it the least favorable cluster for investment (with high growth potential). The “Sun city” area northwest of downtown (cluster in yellow) has relatively more venues compared to the other two clusters, showing that it is in developing stage and there is still opportunity to invest. The other two clusters are sparsely populated making them the most favorable to invest for a good return on investment. Now let’s look at how these clusters map to the new housing communities.



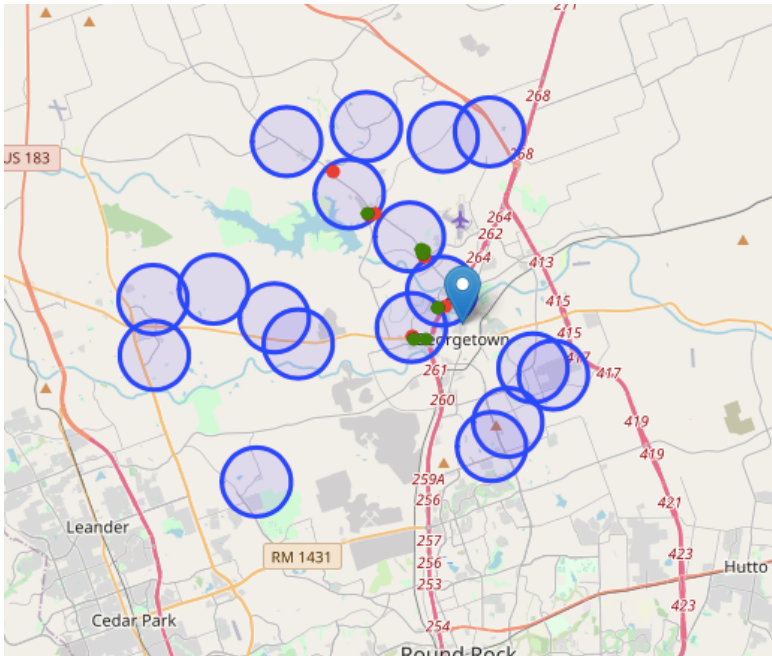
These clusters when overlapped with housing communities clearly show which communities has sparsely populated venues and which communities are near densely populated venues, implying which communities are more favorable for an investment.

The venue data is further sorted based on the category of businesses and top 25 categories are listed below. It gives a fair overview of the businesses available in the area.

Fast Food Restaurant	11
Pizza Place	10
Sandwich Place	9
Restaurant	8
Pharmacy	7
American Restaurant	6
Mexican Restaurant	6
Salon / Barbershop	6
Gas Station	5
Park	5
Cosmetics Shop	5
Convenience Store	5
Burger Joint	5
Construction & Landscaping	5
Gym	4
Seafood Restaurant	4
Department Store	4
Home Service	4
Asian Restaurant	4
Video Store	4
Ice Cream Shop	3
Golf Course	3
Fried Chicken Joint	3
Chinese Restaurant	3
BBQ Joint	3

Name: Cat, dtype: int64

If an investor is interested in a specific retail business space, for example pharmacy and pizza place. I wanted to show what the prospects of opening/owning these businesses in the Georgetown area. The venue locations are sorted to pick the above two categories and the map of these businesses are shown in the map below, red dots are pharmacies and green dots are pizza places.



As both pizza places and pharmacies are located in the downtown and 'Sun city' area, there is still a lot of potential possibilities to invest in a pizza place or a pharmacy near other housing communities.

## 5. Conclusion

In conclusion, the areas around the new housing communities of Georgetown has a very good growth potential for real-estate investment of any kind. If someone is looking for an opportunity to invest in land, commercial establishment or a retail space this analysis can be helpful in identifying the potential geographical locations without having to visit Georgetown. Based on the results it can be concluded that the area around the communities west and south of Georgetown have a lot of potential to grow. The "Sun city" north of Georgetown has more venues compared to other clusters but based on the nature of the business there still might be investment opportunities. All in all Georgetown is a great city if someone is looking to invest in new business developments as there are quite a few new and upcoming housing communities but not many supporting businesses.

The another aspect that can enhance this analysis is office location data such as doctor's office, insurance etc., which could not be obtained from Foursquare. Going forward to make the analysis more complete this information can be added too. Another aspect that will make this analysis more meaningful is to add the commercial real estate listing data. So an investor can see the listings of available commercial spaces along with the businesses that are already present in a given area. Unfortunately this data was not available for free and a lot of real estate companies make it illegal to scrape their websites. This is another piece that will improve this study.