New restaurant location

Qianqian Liu

November 14, 2019

Introduction

In this project we will try to find the ideal location for opening a new restaurant in San Antonio area. San Antonio is the 7th most populous city in the United States. The city is famous for its tourist attractions include the River Walk, the Tower of Americas, SeaWorld. Commercial entertainment like Six Flags and Morgan's Wonderland amusement parks.

We would like to choose a place where there is already has many residents or commercial facilities but not too many restaurants as competition.

To start, we need data of zip code and their population in San Antonio, and we will also use data from Foursquare for information about venues like restaurants, hotels, and other stores.

Firstly, we will explore the population and density distribution by zip code in San Antonio. We will exam the existing restaurants, how they distribute and how many people they serve currently. Then we will also explore the locations of restaurants and other places like gyms, hotels in San Antonio. Finally, we will build a regression model to identify the important factors that could affect the number of restaurants.

Data

Following data sources will be needed to extract/generate the required information:
coordinate of San Antonio counties zip code will be obtained by scaping from website
http://sanantonio.areaconnect.com/zip2.htm?city=San%20Antonio&qs=TX&searchtype=bycity

population of San Antonio counties will be obtained by scaping from website
http://zipatlas.com/us/tx/san-antonio/zip-code-comparison/population-density.htm

number of venues and their categories and location in every zip code will be obtained using Foursquare API.

Methodology

In this project, we assume that there would be many factors that could affect the decision to open a restaurant such as population, number of hotels, gyms, shops etc.
In the first step, we will explore the distribution of the number of restaurants and population by visualization.

Secondly, we will try to find the areas which have similar characteristics by using DBSCAN clustering algorithm.
In the final step, we will use linear regression model to find the significant factors and based on those factors we make recommendation the ideal location for opening a new.

Results

- Population and existing restaurants

The population and density map with the number of restaurants are shown as Figure 1 and Figure 2.
In the figure 1, red represent population and blue represent restaurants. We can see there is imbalance of distribution of restaurants and population. Most restaurants are in the center of San Antonio, especially cumulated in 78205, which is downtown. 78227 also has a lot of restaurants. In some parts there are high in population but has not many restaurants such as 78250, 78228, and 78240.
In the figure 2, red represent population density and 78250 has still low number of restaurants. 78201 and 78228 also have similar situation. 78205, 78215 and 78227 seems to have plenty of restaurants compared to its population density.
From the both maps, we can see that 78205 has most of the restaurants. Riverwalk is the famous tourist attraction in San Antonio and it is located in 78205 and 78215 so the restaurants are very crowded there. 78227 and 78232 also have many restaurants. 78227 is along the Loop 410 and 78232 is in stone oak.
Then we calculated the ratio of restaurant number and population, and sorted the ratio values so that we can avoid the areas that are already crowded with existing restaurants.
We will use 0.0001 as a cutoff value for ratio, so we would exclude zip code 78205, 78215, 78236, 78208, 78234, 78226, 78204, 78227, 78225, 78232, 78259.

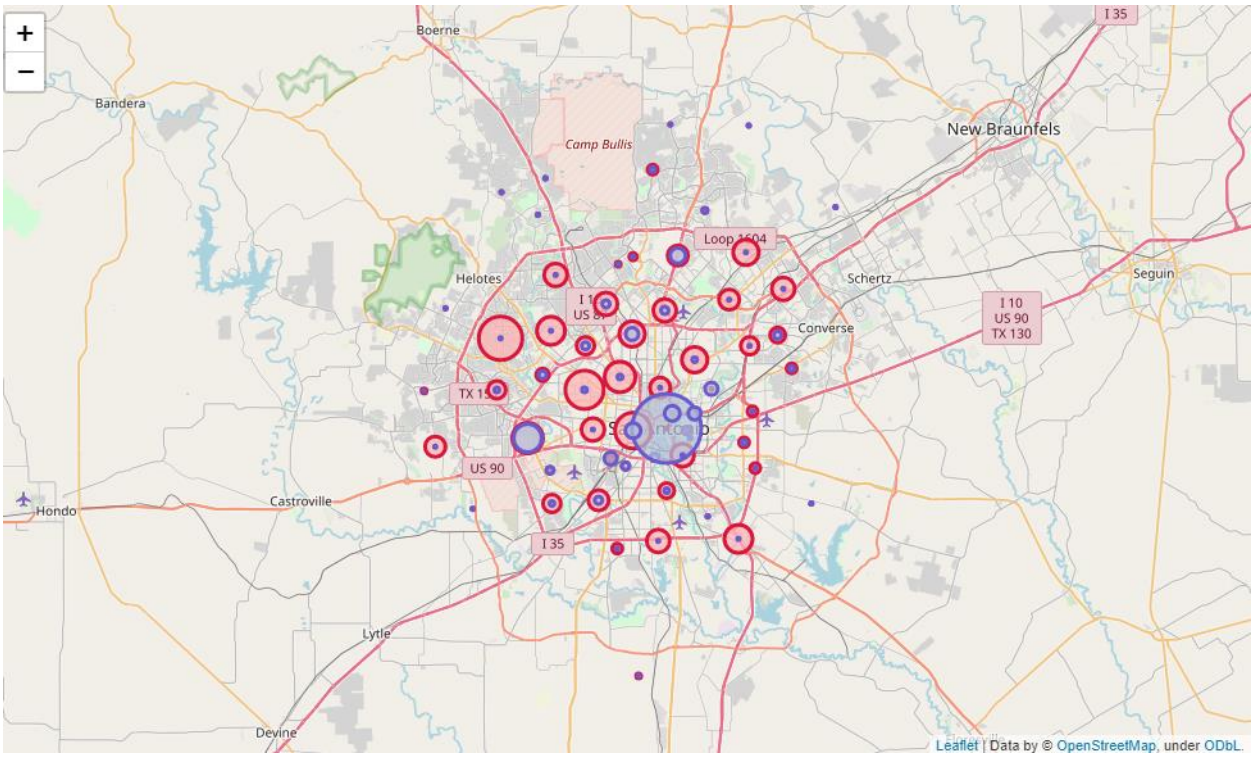

Figure 1. Population and restaurants

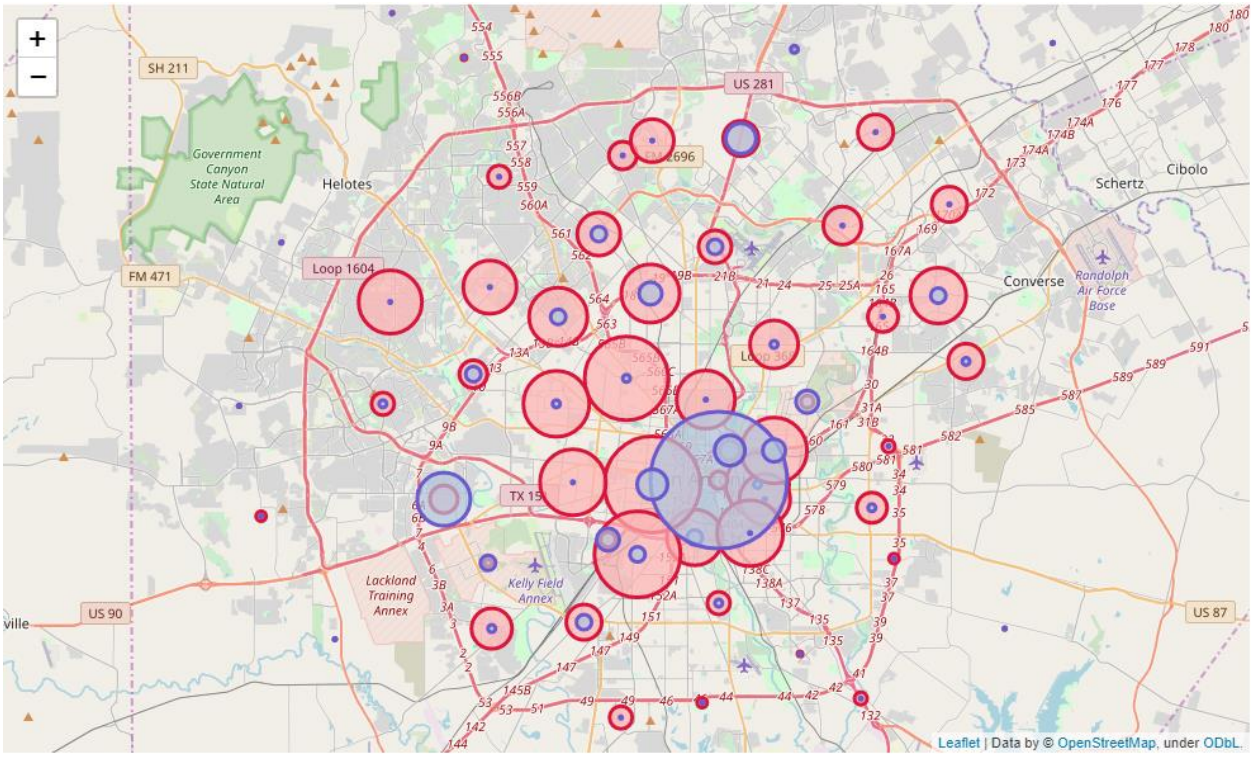

Figure 2. Density and restaurants

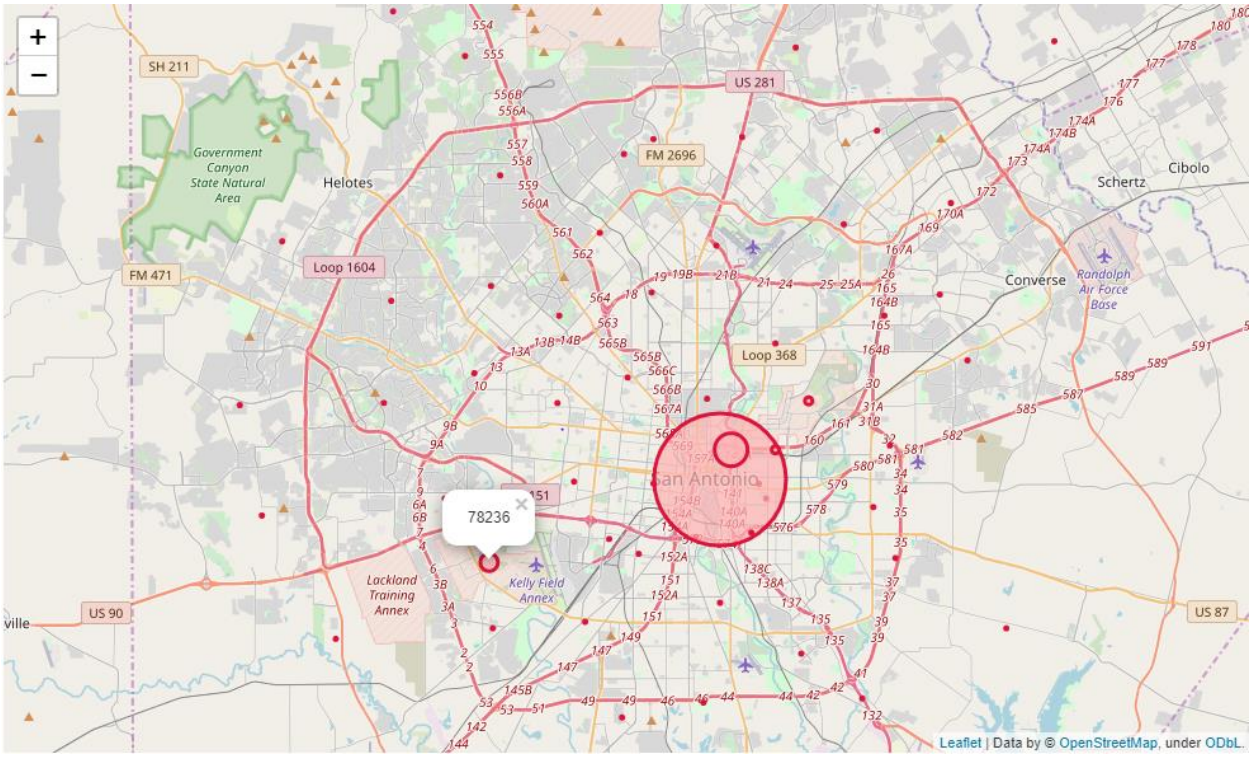

Figure 3. Ratio of number of restaurants and population

- Clustering the zip code

We choosed DBSCAN which can identify the outliers or abnormal points. Our dataset was clustered into 3 categories, part of which are shown in the Table 1-3.
Cluster label "1" contains only 2 areas 78219 and 78222. These two areas have little population and density, and the no ohter commercial facilities like hotel, gym, grocery stores etc.
Cluster label "-1" which is counted noise the places like grocery store or hotel are very little, no more than 1 and the sum of popular places no more than 2.
Cluster label "0" contains areas which have most residents or commercial facilities.
Considering we prefer to open the new restaurant in a commercial place but not too many restaurants existing, we would like to choose cluster 0 but exclude the zip code with high ratio of restaurants and population.
Figure 4 shows the clusters in the map.

	Cluster Labels	zipcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	zip_code	county	lat	...	population	density	gym	hotel
2	-1	78203	Park	Historic Site	Fried Chicken Joint	Gym / Fitness Center	Wings Joint	78203	Bexar	29.415	...	5845	4073.78	1.0	0.0
10	-1	78212	Locksmith	Wings Joint	Entertainment Service	Food & Drink Shop	Food	78212	Bexar	29.462	...	31220	4625.20	0.0	0.0
15	-1	78217	Advertising Agency	Gas Station	Gym	Pet Store	Wings Joint	78217	Bexar	29.544	...	32502	2981.79	1.0	0.0
18	-1	78220	Fried Chicken Joint	Pharmacy	Mexican Restaurant	Wings Joint	Farmers Market	78220	Bexar	29.411	...	16668	2339.54	0.0	0.0
23	-1	78228	Mexican Restaurant	Wings Joint	Farmers Market	Food & Drink Shop	Food	78228	Bexar	29.460	...	58091	5240.63	0.0	0.0
26	-1	78231	Pool	Grocery Store	Wings Joint	Entertainment Service	Food	78231	Bexar	29.577	...	7615	2107.95	0.0	0.0
28	-1	78233	Warehouse Store	Wings Joint	Electronics Store	Food	Flower Shop	78233	Bexar	29.554	...	36065	2691.19	0.0	0.0
30	-1	78235	Gym	Basketball Court	Wings Joint	Farmers Market	Food & Drink Shop	78235	Bexar	29.342	...	692	339.62	2.0	0.0
32	-1	78237	Convenience Store	General Entertainment	Wings Joint	Entertainment Service	Food	78237	Bexar	29.423	...	36273	5270.83	0.0	0.0
35	-1	78242	Fast Food Restaurant	Wings Joint	Food Truck	Food & Drink Shop	Food	78242	Bexar	29.354	...	28786	3180.00	0.0	0.0
37	-1	78247	Gas Station	Video Store	Wings Joint	Entertainment Service	Food	78247	Bexar	29.588	...	39773	2709.29	0.0	0.0

Table 1. Cluster with label “-1”

	Cluster Labels	zipcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	zip_code	county	lat	...	population	density	gym	hotel	p
0	0	78201	Donut Shop	BBQ Joint	Video Store	Convenience Store	Pharmacy	78201	Bexar	29.472	...	47387	6664.58	0.0	0.0	
1	0	78202	Electronics Store	Food	Grocery Store	Clothing Store	Mexican Restaurant	78202	Bexar	29.422	...	11746	5026.90	0.0	0.0	
3	0	78204	BBQ Joint	Seafood Restaurant	Convenience Store	Fast Food Restaurant	Food	78204	Bexar	29.397	...	11905	4332.32	0.0	0.0	
4	0	78205	Hotel	Bar	Steakhouse	Ice Cream Shop	Theater	78205	Bexar	29.424	...	1564	1472.21	0.0	19.0	
5	0	78207	Fast Food Restaurant	Mexican Restaurant	Gym / Fitness Center	Wings Joint	Entertainment Service	78207	Bexar	29.422	...	56348	7554.55	1.0	0.0	
6	0	78208	Fast Food Restaurant	Hotel	BBQ Joint	Burger Joint	Convenience Store	78208	Bexar	29.438	...	5079	5130.21	0.0	2.0	
7	0	78209	Ice Cream Shop	Spa	Gift Shop	Pizza Place	Organic Grocery	78209	Bexar	29.488	...	40675	3848.79	1.0	0.0	
8	0	78210	Pizza Place	Convenience Store	Steakhouse	Furniture / Home Store	Gourmet Shop	78210	Bexar	29.399	...	37345	5148.26	0.0	0.0	
9	0	78211	Fast Food Restaurant	Gas Station	Donut Shop	Bank	Mexican Restaurant	78211	Bexar	29.357	...	31214	2699.50	0.0	0.0	
11	0	78213	Fast Food Restaurant	Wings Joint	Grocery Store	Gas Station	Mexican Restaurant	78213	Bexar	29.512	...	39477	4671.63	1.0	0.0	
12	0	78214	Boutique	Ice Cream Shop	Pool Hall	Athletics & Sports	Mexican Restaurant	78214	Bexar	29.366	...	23705	1745.92	0.0	0.0	
13	0	78215	Coffee Shop	Gym	Park	Bar	Bakery	78215	Bexar	29.438	...	1395	1609.24	2.0	1.0	

Table 2. Cluster with label “0”

	Cluster Labels	zipcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	zip_code	county	lat	...	population	density	gym	hotel	park	cstore
17	1	78219	Business Service	Wings Joint	Farmers Market	Food & Drink Shop	Food	78219	Bexar	29.440	...	14169	934.68	0.0	0.0	0.0	0.0
19	1	78222	Business Service	Wings Joint	Farmers Market	Food & Drink Shop	Food	78222	Bexar	29.387	...	15203	744.23	0.0	0.0	0.0	0.0

Table 3. Cluster with label "1"

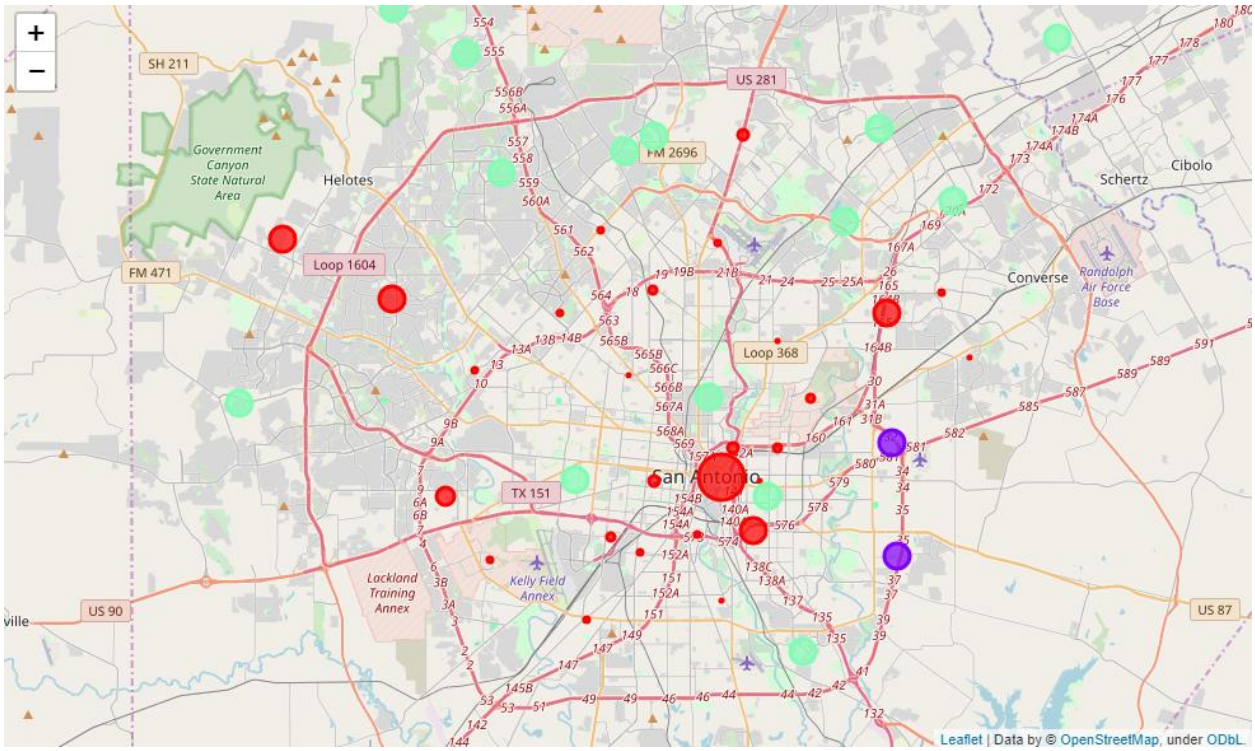

Figure 4. Clusters on map

- Linear regression.
  By fitting multiple linear regression model using covariates population, density, hotel, park, convenient store, grocery store, gym, and fast food, we got the R square 0.81 and hotel and fast food are significant. The coefficients for hotel is about 0.90 and for fast food is 0.94, meaning that there are positive linear relationship between the number of restaurants and hotel as well as fast food store. About one unit increasing of hotel and fast food store associate with one more restaurants in each zip code area. The fitted model is:

$$Restaurant = 0.488 + 0.377Population - 0.137Density + 0.897Hotel + 0.825Park + 0.508ConvenientStore + 0.203GroceryStore + 0.181Gym + 0.94FastFood$$

The significance is shown in the table 4:

	var	coef	pv
0	intercept	0.488036	1.987890e-02
1	population2	0.377308	9.005459e-02
2	density2	-0.137000	5.702830e-01
3	hotel	0.896773	8.241567e-17
4	park	0.824746	2.144398e-01
5	cstore	0.507679	2.705105e-01
6	gstore	0.202952	6.376952e-01
7	gym	0.181149	5.010400e-01
8	fast	0.940187	8.714539e-03

Table 4. Regression table

Based on the analysis we did before, we filtered the zip code by the number of hotels and fast food stores and we got 5 final zip codes and showed in the Figure 5. The sizes of dot represent the total number of the hotels and fast food stores in that zip code. As we can see on the map, the ideal place to open a new restaurant we will recommend is **78207** and **78213**, and then 78211, 78244, 78239 would be our second recommendation.

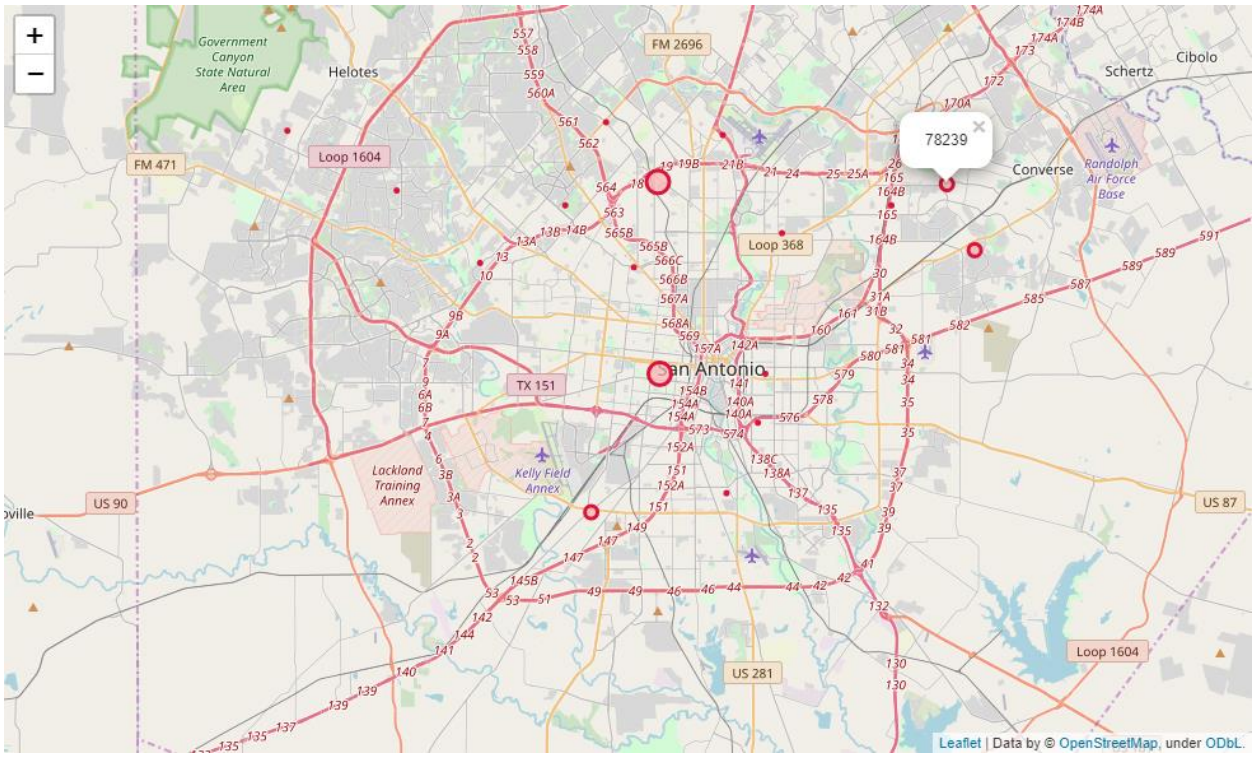

Figure 5. Final filtered zip code

Discussion

In the future work, we can add more potential factors that could affect the decision to open a new restaurant such school rating, house price, rental price, or customer reviews of other restaurant. We may also consider what type of restaurant that is going to open. We can build the model to predict if a restaurant is open in a certain area, is it going to be successful or not.

Conclusion

The purpose of this project is to find a location for openning a new restaurant in the city of San Antonio. We considered the distribution of San Antonio population and the existing restaurant density. We decided to avoid the area with too many restaurant, but at the same time we recommend the area with plenty residents and commerical facilities. Based on those considerations, we filtered the zip codes with high restaurant density, and then cluster the zip codes by venues, and choose the cluster with most commercial facilities. We found the most significant factors are hotels and fast food stores by fitting regression model, so we made recommendation for the ideal areas with the most hotels and fast food stores located in.
In conclusion, 78207 and 78213 will be our first choice, then followed by 78211, 78244, and 78239.