Most suitable place to open a new venue in San Pedro Garza García, N.L., México

Eduardo J. Mendoza

July 6, 2020

1 Introduction

1.1 Background

San Pedro Garza García is the municipality with more income per capita in whole latin america and its seated right next to Monterrey, Nuevo León, México, a city rich in industry and known for having an entrepreneurship culture and rich development in all type of industries. San Pedro, SPG from now on, is home for around 150,000 habitants in 69 square kilometers. SPG is also known for is rich diversity in all type of cuisines and recreational centers, and plenty night life in the city center.

1.2 Problem

The main goal is to find out a suitable niche to start a new business, it can be a recreational center, a restaurant, and where to place it in a city rich in diversity of venues and with high competitiveness between business.

1.3   Interest

Angel investors and entrepreneurs who are already in the area or wish to expand to this area will be interested in this information since it will give an overview of the venues in each neighborhood and based on this we can spot niches or opportunities in other neighborhoods; with this information we hope to reduce the risk of such an investment by making an educated decision after analyzing the neighborhoods in the city.

2   Data acquisition and cleaning

2.1   Data Sources

Postal codes were obtained from the official Mexican postal service, the problem is that this data set does not contain the latitude and longitude of the postal codes, so this other needed information was gathered from the free online database known as GeoNames. The venues information was gathered from the Foursquare API.

2.2   Data Cleaning

The data downloaded from different places was scraped into a single database; the Mexican postal services database didn't have the information of the longitude and latitude directions of each postal code so that is where GeoNames database came to the rescue.

The main problem with GeoNames database is that it had several entries of the same Postal Code with the same latitude and longitude; it was fixed by eliminating duplicates and having a single latitude and longitude per postal code.

When merged together there were columns that were dropped since they didn't contain valuable information for or analysis, these columns were: 'Pais' which is country, and 'Duplicate' which was a column with Boolean values to determine if the Postal Code is duplicated in the database.

After during these the procedure was follow until an error was encountered, for some reason Foursquare API does not have any information for two specific Postal codes in our database, so this two Postal Code rows were dropped in order to remove noise from the database that could mess with the clustering procedure. Below there is a screenshot of the head of the database in order to show how it was settled for analysis.

	CP	ZONA	MUNICIPIO	ESTADO	LATITUD	LONGITUD
0	66200	San Pedro Garza Garcia Centro	San Pedro Garza Garcia	Nuevo Leon	25.661658	-100.410000
1	66210	La Leona	San Pedro Garza Garcia	Nuevo Leon	25.684703	-100.414234
2	66214	El Obispo	San Pedro Garza Garcia	Nuevo Leon	25.679291	-100.417176
3	66215	San Pedro	San Pedro Garza Garcia	Nuevo Leon	25.679291	-100.417176
4	66216	El Obispo	San Pedro Garza Garcia	Nuevo Leon	25.681186	-100.403745

3 Exploratory data analysis

For this project we are going to do an analysis of every neighborhood in the municipality of San Pedro Garza García, we will map every Postal Code in the city and analyze the top venues around each neighborhood. The first step was to collect the required data which is the information of the postal codes in the municipality, and the latitude and longitude in order to identify them on a map; this locations were mapped using Folium Maps in our notebook and gave us a visual idea of the location of every postal code inside the municipality.

Second step was the one mention in the previous section where we realized the gathering, filtering, and preprocessing of the databases until reaching the fil data frame presented in the screenshot above; then with the locations the code was written to extract the main venues around each postal code.

Once with the information of the principal venues per postal code we could make an idea of the amount of venues registered nearby and created the following summary:

```
CP      MUNICIPIO                ZONA
66200   San Pedro Garza Garcia   San Pedro Garza Garcia Centro        23
66210   San Pedro Garza Garcia   La Leona                              3
66214   San Pedro Garza Garcia   El Obispo                            14
66215   San Pedro Garza Garcia   San Pedro                            14
66216   San Pedro Garza Garcia   El Obispo                            12
66217   San Pedro Garza Garcia   Zona Industrial                       4
66218   San Pedro Garza Garcia   Valle del Seminario 1 Sector          6
66219   San Pedro Garza Garcia   Revoluci√≥n 1er Sector                5
66220   San Pedro Garza Garcia   Del Valle                            68
66224   San Pedro Garza Garcia   Fuentes del Valle Sector Colinas      6
66225   San Pedro Garza Garcia   La Joya                              39
66226   San Pedro Garza Garcia   Bugambilias                          39
66227   San Pedro Garza Garcia   Las Capillas                         25
66228   San Pedro Garza Garcia   Residencial San Carlos                2
66230   San Pedro Garza Garcia   San Pedro                             2
66233   San Pedro Garza Garcia   Santa Elena                           5
66235   San Pedro Garza Garcia   Los Olmos                             6
66236   San Pedro Garza Garcia   Residencial Palo Blanco              15
66237   San Pedro Garza Garcia   Los Sauces 2 Sector                  18
66238   San Pedro Garza Garcia   Lucio Blanco 3er Sector               9
66239   San Pedro Garza Garcia   Volkswagen                           12
66240   San Pedro Garza Garcia   La Cooperativa                       19
66244   San Pedro Garza Garcia   Lomas Valle Sector Convento          11
```

```
66245   San Pedro Garza Garcia   Barranca Del Pedregal                  4
66246   San Pedro Garza Garcia   Hacienda Del Valle                    19
66247   San Pedro Garza Garcia   Mansi√≥n Del Rosario                   5
66249   San Pedro Garza Garcia   Villas de Terrasol                     6
66250   San Pedro Garza Garcia   Zona Jer√≥nimo Siller                  5
66254   San Pedro Garza Garcia   Carrizalejo                           18
66256   San Pedro Garza Garcia   Lomas Del Valle                        4
66257   San Pedro Garza Garcia   Del Valle Sector F√°tima              44
66259   San Pedro Garza Garcia   La Ca√±ada                            39
66260   San Pedro Garza Garcia   Corporativo Proser                   100
66263   San Pedro Garza Garcia   Zona Campestre                         5
66264   San Pedro Garza Garcia   Jardines del Campestre                28
66265   San Pedro Garza Garcia   Valle Del Campestre                   35
66266   San Pedro Garza Garcia   La Diana                              39
66267   San Pedro Garza Garcia   Santa Engracia                        64
66268   San Pedro Garza Garcia   Del Valle Sect Norte                  30
66269   San Pedro Garza Garcia   Del Valle Sect Oriente                80
66270   San Pedro Garza Garcia   Colinas de San Agustin                 3
66273   San Pedro Garza Garcia   Zona Montebello                       58
66274   San Pedro Garza Garcia   Jardines de San Agustin                2
66275   San Pedro Garza Garcia   Mesa de la Corona                      1
66276   San Pedro Garza Garcia   Hacienda San Agustin                  18
66277   San Pedro Garza Garcia   Privada San Roberto                    4
66278   San Pedro Garza Garcia   Flor de Mayo                           3
66279   San Pedro Garza Garcia   Punto Central                          7
66280   San Pedro Garza Garcia   Villa Del Pedregal                     6
66285   San Pedro Garza Garcia   Sierra Nevada                          4
66286   San Pedro Garza Garcia   Colonial La Sierra                     1
66287   San Pedro Garza Garcia   Lomas Del Rosario                      1
66295   San Pedro Garza Garcia   Lomas de San Angel                     3
66296   San Pedro Garza Garcia   Colinas de San Angel 1er Sector        3
66297   San Pedro Garza Garcia   Villa Chipinque                       24
```

The venue with most venues registered was 66260 with 100 venues, even though the limit of the search was top up to 100 but not many others postal codes have similar amount of venues, so it was determined raising the limit value was not necessary for our analysis.

	CP	MUNICIPIO	ZONA	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
52	66295	San Pedro Garza Garcia	Lomas de San Angel	American Restaurant	Forest	Golf Course	Zoo Exhibit	Dog Run	Flower Shop	Flea Market	Fast Food Restaurant	Farmers Market	Event Space
51	66287	San Pedro Garza Garcia	Lomas Del Rosario	Beach Bar	Zoo Exhibit	Food Court	Food	Flower Shop	Flea Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
42	66274	San Pedro Garza Garcia	Jardines de San Agustin	Botanical Garden	Dance Studio	Dog Run	Food	Flower Shop	Flea Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
7	66219	San Pedro Garza Garcia	Revolución 1er Sector	Burger Joint	Dog Run	Park	Department Store	Zoo Exhibit	Flea Market	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
10	66225	San Pedro Garza Garcia	La Joya	Café	Pizza Place	Paper / Office Supplies Store	Fast Food Restaurant	Supermarket	Convenience Store	Boutique	Ice Cream Shop	Gym	Mexican Restaurant

Resuming our analysis, a K-means procedure was realized in order to cluster the neighborhoods, but similarities based on the top 10 venues on each neighborhood. Right above you can see the head of the data frame that shows top 10 venues of 5 postal codes, it is possible to begin drawing conclusions but is way more easier if we see the neighborhoods clustered in a map first and then start seeing the characteristics of each cluster.
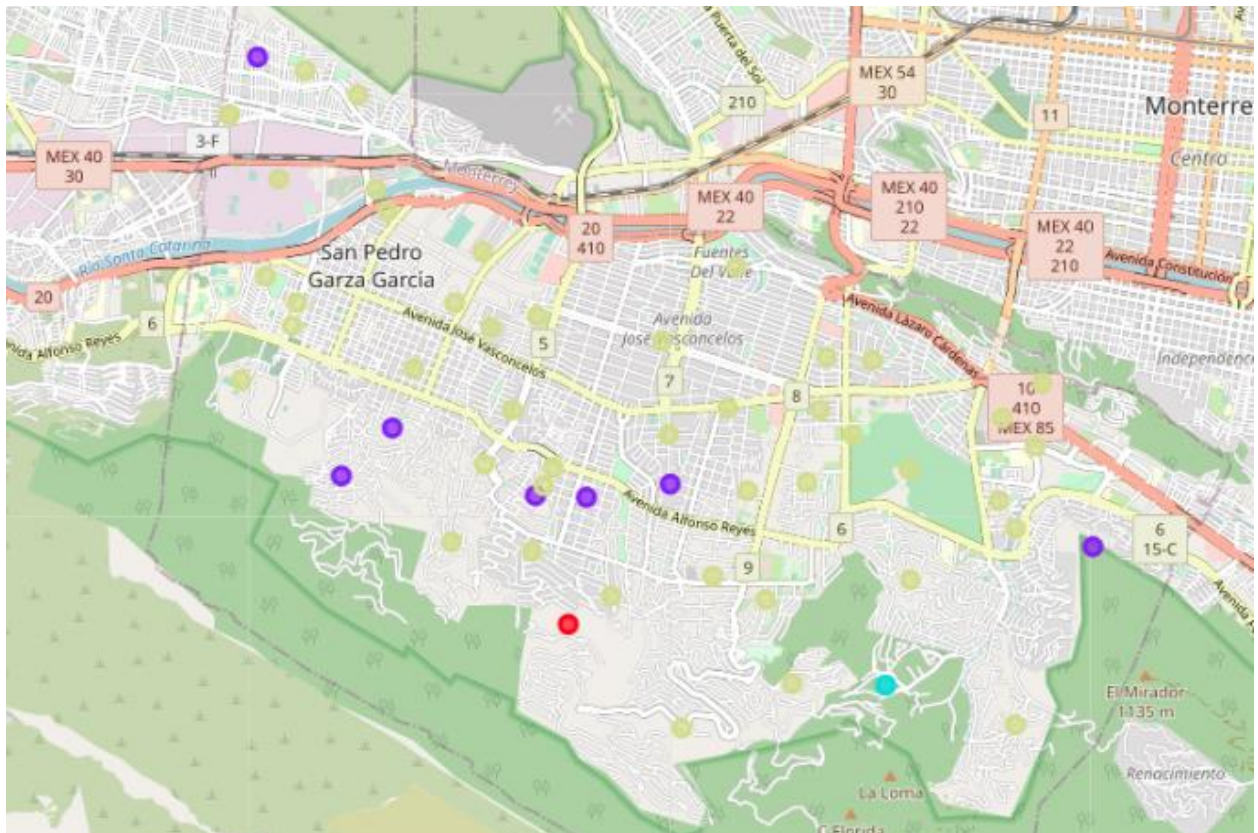
Cluster	Color	Characteristics
0	Red	This neighborhood principal venues are department stores, zoo and dive bars
1	Purple	Around these neighborhoods there are a lot of parks, school buildings and recreational places for sports.
2	Blue	Recreational places, zoo exhibits and diners
3	Yellow	Burger joints, parks, and department stores.

As you can see above, we have processed the data from API Foursquare and have created a more complex view of the neighborhood by analyzing and ranking the top venues on each neighborhood and finally clustering the neighborhoods by similarity; one thing to take into account is that data is constantly changing since everyday people keep uploading information at Foursquare.

Another important point is that the quality of this study depends also on how much information does Foursquare have to share; I personally think that Foursquare is not that popular in Latin America and by personally analyzing venues downloaded from the API I can say that some of the information is outdated... nevertheless this is a good exercise to demonstrate the power of data science and how we can use this tool for such type of complex analysis.

Analyzing the info we can see that in cluster number 3 there are already many restaurants, parks, a different stores ranging from flower shops, to supermarkets or even; blue cluster also has diners as one of their top 3 venues and the red cluster seems to be more about department stores and dive bars; which leaves us with cluster number one that has a lot of recreational places such as parks, gyms, and school buildings.

4 Conclusion and following steps

After analyzing the previous information, there is a detail that cached my eye: cluster number one. Cluster number one has really interesting characteristics which I think it makes it suitable to start a restaurant or a juice bar; this cluster main characteristics are that is surrounded by recreational places such as parks, gyms and also school buildings. Having this information plus the uprising trends during the last decade about a healthier life style and taking care of your body, I think that starting a restaurant of healthy food, or a smoothie/juice bar or even a different type of gym in this neighborhood is very promising since people who go to this recreational places in the neighborhood would find it convenient to have a healthy restaurant or a smoothie bar where they have a snack or a drink after a long walk in the park. Also you can take into account all the teenagers who go the schools around this neighborhoods and their parents who will frequent these places and would have to wait frequently for their sons and would prefer to wait in a nice healthy bar or restaurant instead of just being park in their cars.

For next steps I would propose studying this neighborhoods in order to scout what type of venues are there since as I said before, foursquare information is dependent on what the users upload to the platform and its popularity in every location; after this I would start creating a business plan for a healthy lifestyle venue and see if it fits as the data hints it will.