GERMAN DOMINGUEZ

DATA SCIENCE PROJECT CASPTONE

Opening a restaurant/bar in Monterrey Metropolitan Area

Opening a restaurant/bar in Monterrey Metropolitan Area

Introduction where you discuss the business problem and who would be interested in this project.

Background

Monterrey Metropolitan area is currently the third largest city in México, behind México City and Guadalajara. Monterrey is a modern city growing a fast pace, it is home of Tec de Monterrey, the best university in Mexico(I work there) also has automotive and manufacturing clusters, cement and construction clusters, IT cluster, tourism clusters, restaurant and mall plazas. Due to particular cultural aspects of society, people love to go out to restaurants and bars, from Wednesday nights to Sunday afternoon.

Clients interested

Considering the background, it represents an attractive place for investors, particularly for in the restaurant/bar industry. Now, due to the current pandemic, confinement restrictions and physical distancing, it doesn't seem like a good idea to open a restaurant or a bar, however, it is clear to see that people is eager to go out to these kind of places, now even more than ever.

Problem or business case

The intention of this project is make a recommendation of a place to open a restaurant or bar in Monterrey Metropolitan area which is integrated by 12 boroughs. To achieve this, the project will have particular objectives: explore the Monterrey Metropolitan area neighborhoods looking up for zones of interest like universities or plazas. Search for common venues and explore its surroundings. Segment and cluster neighborhoods and particular venues to find those with restaurants or bars. And finally, try to find the current trends in restaurants or bars. (This may or may not be possible given that most of these venues are closed and people is not going out)

Data where you describe the data that will be used to solve the problem and the source of the data.

For this particular project, geographical information about the Monterrey metropolitan area is needed in the first place. Given that there is no budget to buy this information, it is obtained for free from a page called, GeoNames, a geographical database which covers all countries and contains over eleven million placenames that are available for download free of charge.

The dataset contains México's geographical information of all states, including the following attributes: 'Country', 'Postal Code', 'Neighborhood', 'State', 'State_No', 'Borough', 'Borough_No', 'City', 'City_No', 'Latitude', 'Longitude', 'Coo_No'.

From this dataset not all information is useful and needs to be discarded, but the attributes to be analyzed are: 'Country', 'Neighborhood', 'State', 'Borough', 'Latitude', 'Longitude'. First it's important to make sure to get the right Country, Mexico in this case, then the State of Nuevo León, which is where Monterrey is located. All of these attributes are *object or string* except for the Postal Code and coordinates which area *int* and *float*. Getting the right information is key but it is also essential to perform a quick check for null values, and drop if any.
The main attributes 'Neighborhood', 'Borough' are needed to filter the three most important of the metropolitan area, and 'Latitude', 'Longitude' to make the Foursquare queries. It is important to note that some neighborhoods may have the same Postal Code, therefore all duplicates needs to be eliminated.

The next piece of information required regarding the location is extracted from a Wikipedia page containing information about Monterrey Metropolitan Area. The main table has: 12 Boroughs, population, total area, human development index, etc. From here it is easy to identify the 3 most important Boroughs by human development index: 'San Pedro Garza García', 'Monterrey', 'San Nicolás de los Garza'

Finally, the venue information is obtained from the Foursquare API using queries with the selected and cleaned information before. The attributes needed for this are 'Neighborhood', 'Latitude', 'Longitude'.
Some examples: use coordinates to search particular venues like bars near a university, students love to drink and dance.
Segment and cluster neighborhoods with restaurants and bars downtown Monterrey.
Or find the most common venues in a residential area in Monterrey.

All the data and additional information for the project is generated based on these datesets, pages and queries. This may be more than enough data to analyze and achieve the defined objectives: explore the Monterrey metropolitan area neighborhoods looking up for zones of interest like universities or plazas. Search for common venues and explore its surroundings. Segment and cluster neighborhoods and particular venues to find those with restaurants or bars. And finally, try to find the current trends in restaurants or bars.

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

The methodology to wrangle and analyze the data was pretty straight forward and to facilitate the understanding and development of the project it was divided into the following sections:

Data load: Load into a pandas dataframe the obtained dataset for free from a page called, GeoNames, a geographical database which covers all countries and contains over eleven million placenames that are available for download free of charge.

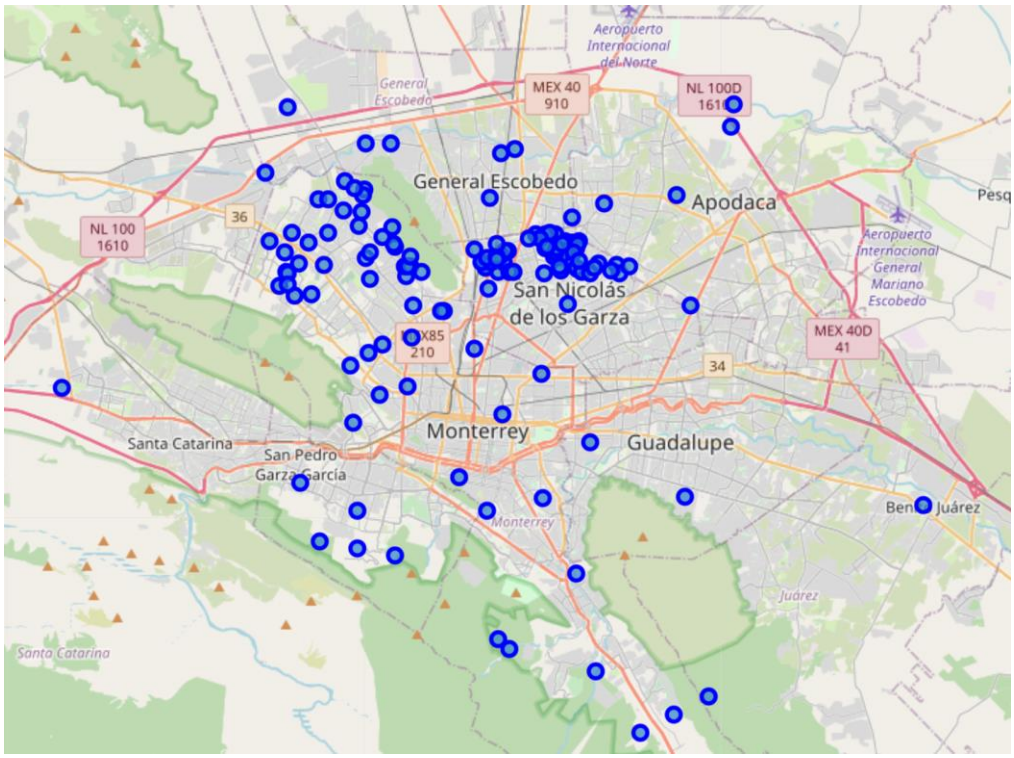
Data cleaning: Droping duplicates values and features not needed 'State_No','Borough_No', 'City', 'City_No','Coo_No'
Data selection: Selecting the right data of state Nuevo León and boroughs 'San Pedro Garza García', 'Monterrey', 'San Nicolás de los Garza'
Additional data: Wikipedia page containing information about Monterrey Metropolitan Area. The main table has: 12 Boroughs, population, total area, human development index, etc. From here it is easy to identify the 3 most important Boroughs by human development index: 'San Pedro Garza García', 'Monterrey', 'San Nicolás de los Garza'
Data exploration: Know the type of data, shape, descriptions and some counts of the data of interest:

```
Neighborhoods(['Abasolo', 'Agualeguas', 'Los Aldamas', 'Allende',
'Anáhuac',
       'Apodaca', 'Aramberri', 'Bustamante', 'Cadereyta Jiménez',
       'El Carmen', 'Cerralvo', 'Ciénega de Flores', 'China',
       'Doctor Arroyo', 'Doctor Coss', 'Doctor González', 'Galeana',
       'García', 'San Pedro Garza García', 'General Bravo',
       'General Escobedo', 'General Terán', 'General Treviño',
       'General Zaragoza', 'General Zuazua', 'Guadalupe', 'Los Herreras',
       'Higueras', 'Hualahuises', 'Iturbide', 'Juárez',
       'Lampazos de Naranjo', 'Linares', 'Marín', 'Melchor Ocampo',
       'Mier y Noriega', 'Mina', 'Montemorelos', 'Monterrey', 'Parás',
       'Pesquería', 'Los Ramones', 'Rayones', 'Sabinas Hidalgo',
       'Salinas Victoria', 'San Nicolás de los Garza', 'Hidalgo',
       'Santa Catarina', 'Santiago', 'Vallecillo', 'Villaldama'],

      dtype=object)
```

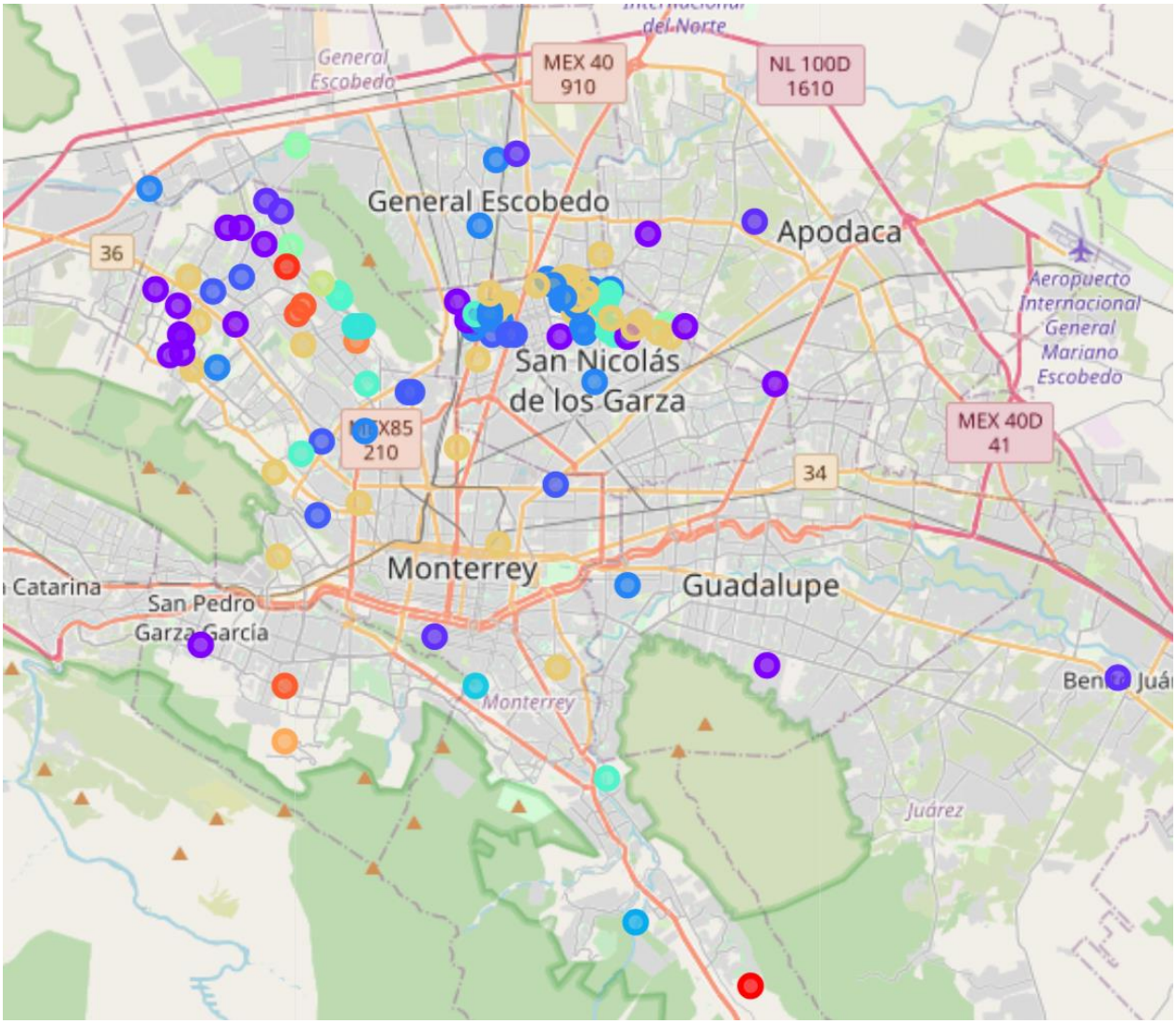
Borough	Neighborhood
Monterrey	75
San Nicolás de los Garza	63
San Pedro Garza García	8

Exploration of the neighborhoods in Monterrey: Visualization of neighborhoods in a map:



Foursquare API queries: use geopy library to get the latitude and longitude values of Monterrey and the venue information is obtained from the Foursquare API using queries with the selected and cleaned information before. The attributes needed for this are 'Neighborhood', 'Latitude', 'Longitude'.
Analyze data of each neighborhood: dataframe with the top 10 venues for each neighborhood.
Look up for venues in a certain neighborhood (near Tec de Monterrey university): indicate a neighborhood to lookup for its 10 most common venues
Search a venue category in that neighborhood (we are interested in restaurant or bars): get geographical coordinates of the indicated neighborhood and get the names of particular venues in the results (competition)
Search for trending venues in that neighborhood: it is possible that due to covid-19 restrictions the result was 'No trending venues are available at the moment!'
Cluster Neighborhoods: first evaluate to find the best Ks, plot the inertias. Test 60 ks and run 18 as it seems a good balance between clustering reliability and inertia score. Create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

Map of Monterrey's clustered neighborhoods: visualization of the clusters in a map of Monterrey



Cluster examination: count the total number of venues in each cluster, examine some clusters and assign a name based on its most common venues and look up for the cluster label of our neighborhood.

Results section where you discuss the results.

- It is interesting to see that the neighborhood with more venues are the wealthy ones.

Neighborhood	Venue
Cumbres Elite	68
La Boquilla	44
Punto Central	36
Rincón del Valle	34
Urdiales	31
Anáhuac Sendero	31
Los Mezquites	31
Viejo Roble	30
Balcones de Galerias	30
Real de Cumbres	30

How many unique categories can be curated from all the returned venues: 162

- Taco Place, Convenience Store, Mexican Restaurant, Burger Joint and Park categories are the most popular

Venue Category	Venue
Taco Place	157
Convenience Store	156
Mexican Restaurant	116
Burger Joint	81
Park	56
Pharmacy	55
Pizza Place	47
Seafood Restaurant	42
Gym	41
Ice Cream Shop	40

- La Boquilla neighborhood near the university does not have bars in the 10 most popular venues, however, it has more than 30 bars near (1000 meters).

	Neighborhood	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
53	La Boquilla	Mexican Restaurant	Taco Place	Convenience Store	Restaurant	Burger Joint	Hot Dog Joint	Park	Ice Cream Shop	Cosmetics Shop	Dance Studio

- Also, it does have less than a dozen restaurants near (1000 meters).

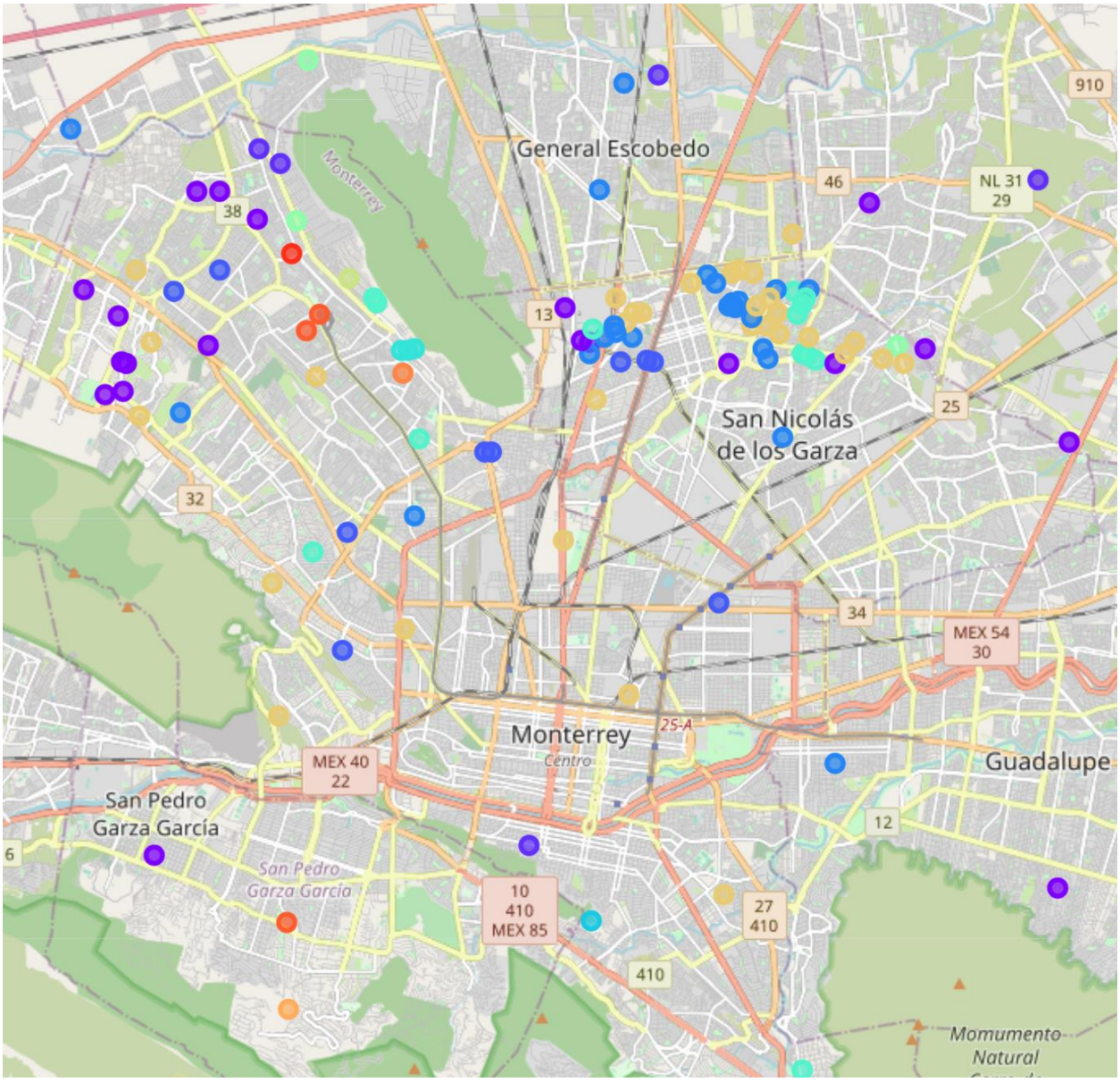
	name	categories	address	lat	lng
9	La Ezquina	Mexican Restaurant	Junco de la Vega 8	25.649799	-100.286705
0	Jerry's home restaurant	Restaurant	Av.Ricardo Covarrubias 3327	25.652403	-100.280247
5	Restaurante Chely	Restaurant	NaN	25.657289	-100.284155
4	Olive Garden	Italian Restaurant	Av. Revolución 2703	25.653209	-100.275706
7	P.F. Chang's	Chinese Restaurant	Av. Revolución 2703, Local R-08	25.653416	-100.275009
3	restaurante Hare Krishna TEC	Vegetarian / Vegan Restaurant	NaN	25.650409	-100.292515
2	Yummy's restaurant	Mexican Restaurant	av del estado 202a	25.650688	-100.293364
8	La Charanda	Beer Garden	Filósofos 106	25.650812	-100.293401
1	Báltica Restaurant	Seafood Restaurant	Junco de la Vega Col. Roma	25.654656	-100.292334
10	Big Ben Restaurante	Wings Joint	Avenida Luis Elizondo	25.645744	-100.292572
6	Restaurante Fiesta Inn	Restaurant	Paseo Tec	25.654741	-100.292959

- Near this neighborhood it wasn't possible to find trends in venues. Probably due the current pandemic.

```
[ ]  # display trending venues
     trending_venues_df
```

```
'No trending venues are available at the moment!'
```

- Finally, the clusters show an interesting image as most of the neighborhoods have taco places and convenience stores, some others had a variety of restaurants and other parks and athletics venues.



	1st Most Common Venue
Cluster Labels	
13.0	32
4.0	24
1.0	20
8.0	13
3.0	11
2.0	7
16.0	3
7.0	3
10.0	3
0.0	1

Examples of clusters:

Cluster 2: Mexican, Sea food and Wings restaurants

```
[ ] monterrey_merged.loc[monterrey_merged['Cluster Labels'] == 2, monterrey_merged.columns[[1] + list(range(2, monterrey_merged.shape[1]))]].head()
```

	Postal Code	Neighborhood	State	Borough	Latitude	Longitude	Cluster Labels	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
23	64208	Conquistadores	Nuevo León	Monterrey	25.7907	-100.3807	2.0	Mexican Restaurant	Food Truck	Wings Joint	Dessert Shop	Exhibit	Event Space	Electronics Store	Dog Run	Dive Bar	Diner
24	64209	Tierra Propia (F-35)	Nuevo León	Monterrey	25.7877	-100.3761	2.0	Mexican Restaurant	Pizza Place	Wings Joint	Department Store	Event Space	Electronics Store	Dog Run	Dive Bar	Diner	Dessert Shop
27	64217	PROVILEON San Bernabé	Nuevo León	Monterrey	25.9379	-100.3574	2.0	Plaza	Mexican Restaurant	Seafood Restaurant	Wings Joint	Department Store	Event Space	Electronics Store	Dog Run	Dive Bar	Diner
52	64429	Heriberto Jara	Nuevo León	Monterrey	25.6472	-100.0958	2.0	Mexican Restaurant	Burrito Place	Seafood Restaurant	Wings Joint	Event Space	Electronics Store	Dog Run	Dive Bar	Diner	Dessert Shop
66	64723	Unión Loma Larga	Nuevo León	Monterrey	25.6592	-100.3242	2.0	Mexican Restaurant	Dessert Shop	Convenience Store	Wings Joint	Exhibit	Event Space	Electronics Store	Dog Run	Dive Bar	Diner

Cluster 3: Movie theaters, Mexican restaurants and Candy stores

```
[ ] monterrey_merged.loc[monterrey_merged['Cluster Labels'] == 3, monterrey_merged.columns[[1] + list(range(2, monterrey_merged.shape[1]))]].head()
```

	Postal Code	Neighborhood	State	Borough	Latitude	Longitude	Cluster Labels	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
8	64070	El Mirador Centro	Nuevo León	Monterrey	25.7334	-100.3321	3.0	Movie Theater	Afghan Restaurant	Candy Store	Juice Bar	Motel	Dessert Shop	Fast Food Restaurant	Exhibit	Event Space	Electronics Store
10	64100	San Bernabe	Nuevo León	Monterrey	25.7333	-100.3333	3.0	Movie Theater	Afghan Restaurant	Candy Store	Juice Bar	Motel	Dessert Shop	Fast Food Restaurant	Exhibit	Event Space	Electronics Store
12	64102	Mirasol	Nuevo León	Monterrey	25.7676	-100.3891	3.0	Mexican Restaurant	Candy Store	Brewery	Wings Joint	Dessert Shop	Exhibit	Event Space	Electronics Store	Dog Run	Dive Bar
40	64339	Valle Verde 3er Sector	Nuevo León	Monterrey	25.7183	-100.3620	3.0	Mexican Restaurant	Juice Bar	Candy Store	Convenience Store	Seafood Restaurant	Big Box Store	Bus Station	Sushi Restaurant	BBQ Joint	Gym

Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

- The dataset with geographical coordinates had several duplicated coordinates, this showed in the map. With more time, it can be possible to create a function or query to get more accurate data. It is clear to see that several important neighborhoods were missing.
- Based on the exploration and the search, opening a bar in the desired neighborhood is not recommended, as there are many around. However, opening a restaurant or taco place could work as these are the most common venues and there are not a lot of them around.
- For the clusters, more work needs to be done to find the Ks, because as it is, some clusters only have one neighborhood, and others are really far away from each other's. This may be because of the missing data from the original dataset.

Conclusion section where you conclude the report.

This project was really fun, drilling down into the data, creating groups and subtotals of venues, getting a valuable insight of the most common venues in a particular neighborhood, having the ability to search for a particular venue in a certain place and finally clustering and discovering how the landscape looks like to make the right recommendations for clients in your home city.