Proposal for Opening a Bar-Restaurant in Caracas City.

The Battle of the Neighborhoods
Applied Data Science Capstone- Capstone Project
Jhonny Calvo
calvorondon@gmail.com

Introduction

Problem Background

- Caracas, is Venezuela's Capital. Here is the main commercial and cultural center of the country. Most of governmental and private institutions has their main office in this city.

Problem Description

- A Bar-Restaurant is a commercial establishment where foods, alcoholic and non-alcoholic drinks are sold. They are usually aimed at adults, especially young people.

- The problem lies in finding a suitable place to establish a new Bar-Restaurant. It must be located in an accessible place, within a busy area with low competition.

Objective of the project

- The client of this project wishes to establish a new bar-restaurant, aimed at people between 25 and 50 years old, who are executives and professionals who wish to have fun after leaving the office. The client also requires the bar restaurant to be located in an accessible place where you can get customers who meet the profile.

Data Description

We will use two datasets to analyze Caracas City.

Dataset 1:

- This Dataset is taken from the Antipodas.Net website.

- Contains the principal neighborhood's coordinates of Caracas City.

- We use this datasets to explore the neighborhoods distribution.

- To obtain this data, we have used **BeautifulSoup**, LXML and requests, which are libraries for web scraping.

	Neighborhood	Distance	Coordinates
1	Parroquia San Pedro	1 Km	N10°29'20.36" O66°53'20.29"
2	Parroquia El Recreo	2 Km	N10°30'18.11" O66°53'15.83"
3	Chacao	3 Km	N10°29'45.53" O66°51'12.49"
4	Parroquia Candelaria	3 Km	N10°30'21.46" O66°54'17.78"
5	Parroquia San Agustín	3 Km	N10°29'38.76" O66°54'36.94"

Venue	Venue Latitude	Venue Longitude	Venue Category
Estadio Olímpico Universitario	10.490449	-66.885710	Soccer Stadium
Plaza del Rectorado	10.492148	-66.890817	College Quad
Video San Pedro	10.485993	-66.891776	Video Store
Estadio Universitario de Caracas	10.489607	-66.884255	Baseball Stadium
Plaza Cubierta	10.491076	-66.890764	College Quad

Dataset 2:

- We use **Foursquare** to explore the principal Venues of Caracas City.

- Contains all Venues for Caracas City, its coordinates, and the Neighborhoods where the Venue belongs.

- We have search places with the follow tags: Bar, Restaurant (chinese, spanish, american, etc..), Lounge, and many others.

- The coordinates found in **Dataset 1**, have been utilized as input for the Foursquare API.

Methodology

The aim in this project is to classify all Bar-Restaurant venues in Caracas city, so we can evaluate the density of clusters by neighborhoods, and provide the best places to open a new Bar-Restaurant.

Part 1: ordering and cleaning Dataset 1. We have use **Foursquare**, with Dataset 1 as input, and so we can obtain all Bar Restaurant venues in Caracas city (Dataset 2). **Folium** library is used to visualize neighborhoods and venues in a map.

Part 2: Consist in the analysis of each neighborhood, so we can **know the frequency** of venues and the common venues per neighborhoods.
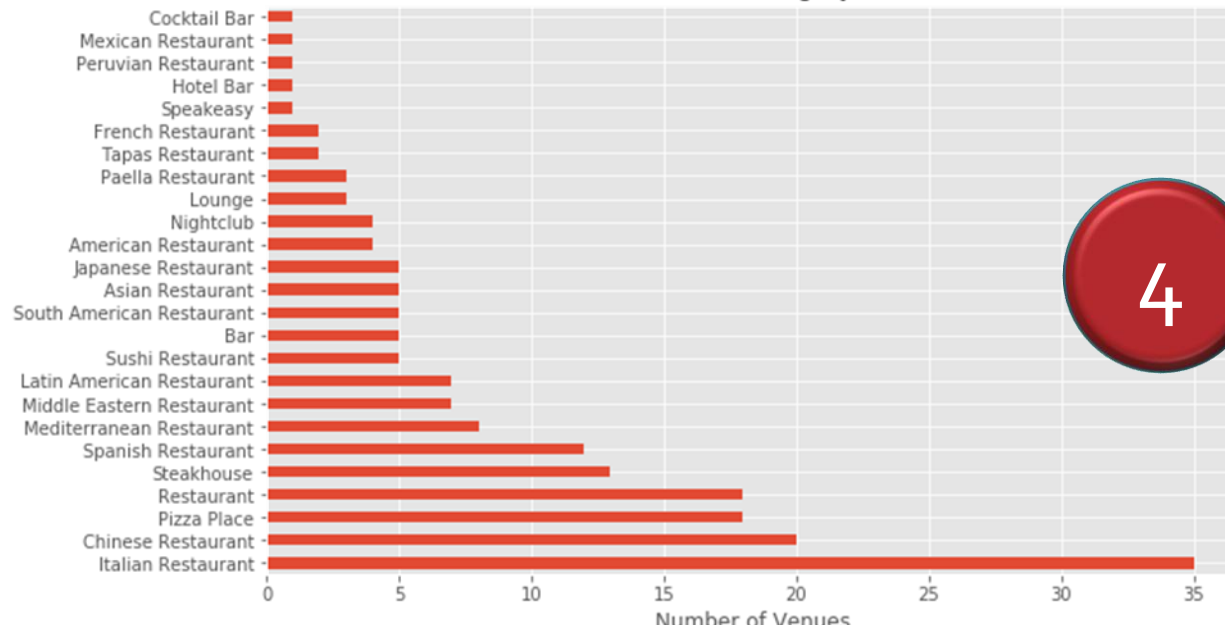
Part 3: Clustering neighborhoods.
We used **k-means** to classify neighborhoods by venues, and observe it in a map, and know how the clusters are distributed in the city.

Part 1

1. Data Ordering.
2. Conversion from DMS to DD (handle for foursquare).
3. Data Visualization (Neighborhoods)
4. Filtering
5. **Venues**

	Neighborhood	Lat	Long
0	Parroquia San Pedro	N10°29'20.36"	O66°53'20.29"
1	Parroquia El Recreo	N10°30'18.11"	O66°53'1"
2	Chacao	N10°29'45.53"	O66°51'
3	Parroquia Candelaria	N10°30'21.46"	O66°54'17
4	Parroquia San Agustín	N10°29'38.76"	O66°54'36.94"

1

	Neighborhood	Latitude	Longitude
0	Parroquia San Pedro	10.489889	-66.889694
1	Parroquia El Recreo	10.505306	-66.88980
2	Chacao	10.497306	-66.8546
3	Parroquia Candelaria	10.507111	-66.906889
4	Parroquia San Agustín	10.496000	-66.912611

2

3

Venue Category

4



5
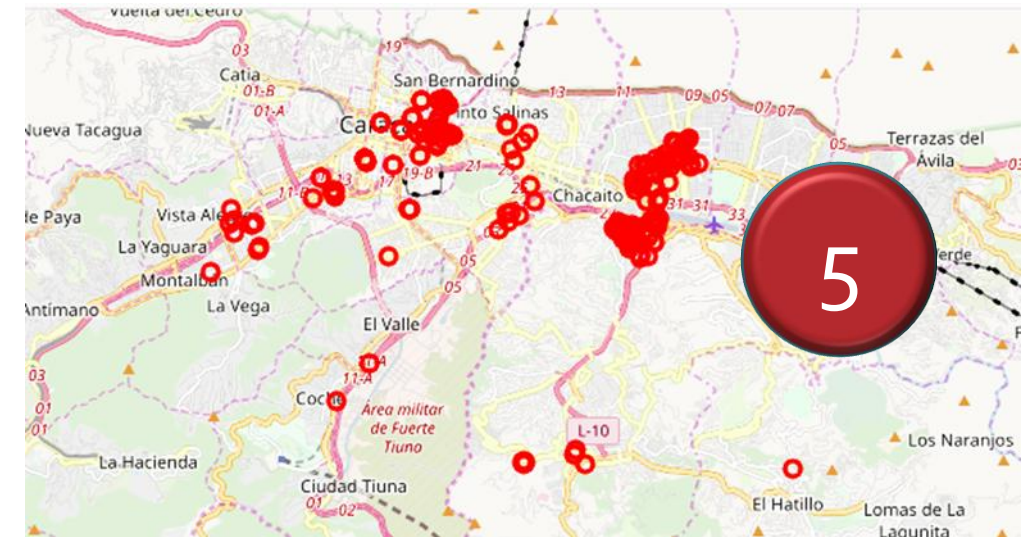
Analyzing Each Neighborhood

- The k-means algorithm isn't directly applicable to categorical variables because Euclidean distance function isn't really meaningful for discrete variables.

- Using pandas.get_dummies, we have convert categorical variable into dummy/indicator variables.

- This process help to group row by neighborhoods and taking the mean of the frequency of occurrence of each venue category, and most common venue. It also help to run the clustering.

```
----Baruta----
                  venue  freq
0   Italian Restaurant   0.43
1         Pizza Place    0.29
2   American Restaurant  0.14
3      Tapas Restaurant  0.14
4         Cocktail Bar   0.00


----Carrizal----
                         venue  freq
0                    Steakhouse   1.0
1            American Restaurant  0.0
2     Middle Eastern Restaurant  0.0
3              Sushi Restaurant  0.0
4                     Speakeasy  0.0
```
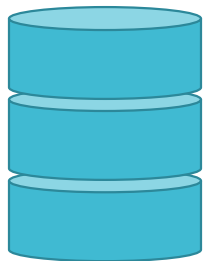
	Neighborhood	1st Most Common Venue	2nd Most Common Venue
0	Baruta	Italian Restaurant	Pizza Place
1	Carrizal	Steakhouse	Tapas Restaurant
2	Chacao	Italian Restaurant	Pizza Place
3	Las Mercedes	Italian Restaurant	Steakhouse
4	Parroquia Altagracia	South American Restaurant	Italian Restaurant
5	Parroquia Candelaria	Italian Restaurant	Spanish Restaurant
6	Parroquia Catedral	Chinese Restaurant	South American Restaurant
7	Parroquia Coche	Chinese Restaurant	Tapas Restaurant
8	Parroquia El Paraíso	Chinese Restaurant	Italian Restaurant
9	Parroquia El Recreo	Steakhouse	Spanish Restaurant
10	Parroquia El Valle	Chinese Restaurant	Tapas Restaurant

Part 3: Clustering neighborhoods.

- Dataset 2, can be viewed as an unlabeled data, and we want to classify venues in different groups or clusters, in order to observe the most common Bar-Restaurants by neighborhoods and make this analysis easier.

- We have use **sklearn.cluster – Kmean Clustering** methods to obtain clusters of Bar-Restaurant venues.
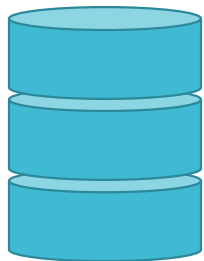
	Neighborhood	American Restaurant	Asian Restaurant	Bar	Chinese Restaurant	Cocktail Bar	French Restaurant	Hotel Bar	Italian Restaurant	Japanese Restaurant	...
0	Baruta	0.142857	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.428571	0.000000	...
1	Carrizal	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	...
2	Chacao	0.071429	0.047619	0.047619	0.023810	0.02381	0.023810	0.02381	0.142857	0.047619	...
3	Las Mercedes	0.000000	0.046512	0.069767	0.023256	0.00000	0.023256	0.00000	0.186047	0.046512	...
4	Parroquia Altagracia	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.333333	0.000000	...

Results

We have selected 6 as the number of cluster. We obtained that clusters labeled as 'o', contain 135 venues, representing 72.6% of total clusters.
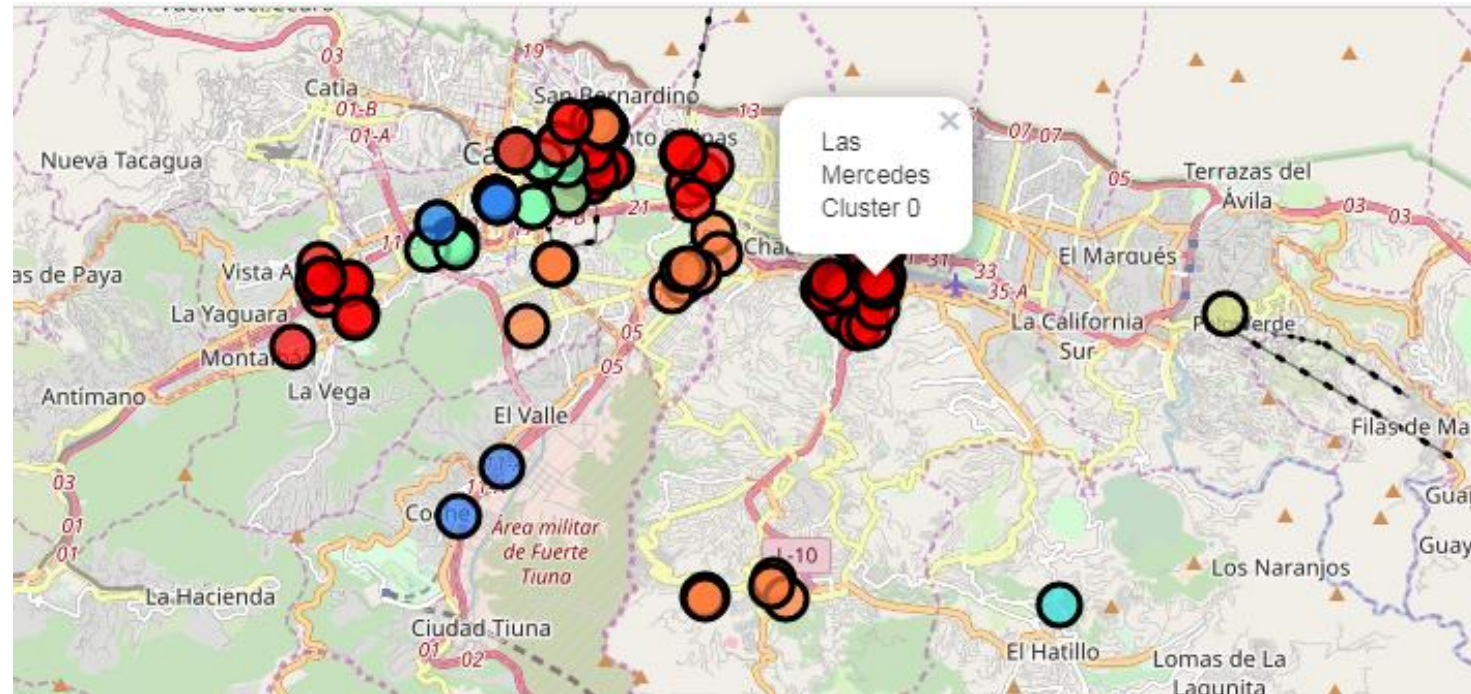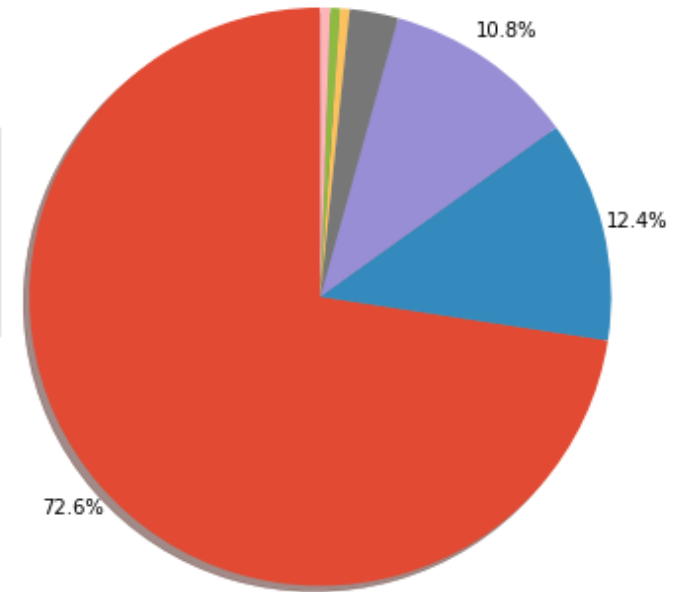

Cluster Labels

|	
0	135
6	23
4	20
2	5
5	1
3	1
1	1

Places with Bars in Caracas

10.8%

12.4%

72.6%

Results
Cluster 'o'

We have found that "Italian Restaurants" are the most common Bar-Restaurant venues in the city,

1st Most Common Venue	
Italian Restaurant	125
Steakhouse	7
South American Restaurant	3

2nd Most Common Venue	
Steakhouse	43
Pizza Place	42
Spanish Restaurant	37
Latin American Restaurant	10
Italian Restaurant	3

5th Most Common Venue	
Mediterranean Restaurant	45
Spanish Restaurant	43
Chinese Restaurant	30
Pizza Place	10
Tapas Restaurant	7

Cluster "0" 1st Most Common Venue

- Italian Restaurant
- Steakhouse
- South American Restaurant

2.2%   5.2%

Cluster "0" Neighboorhoods
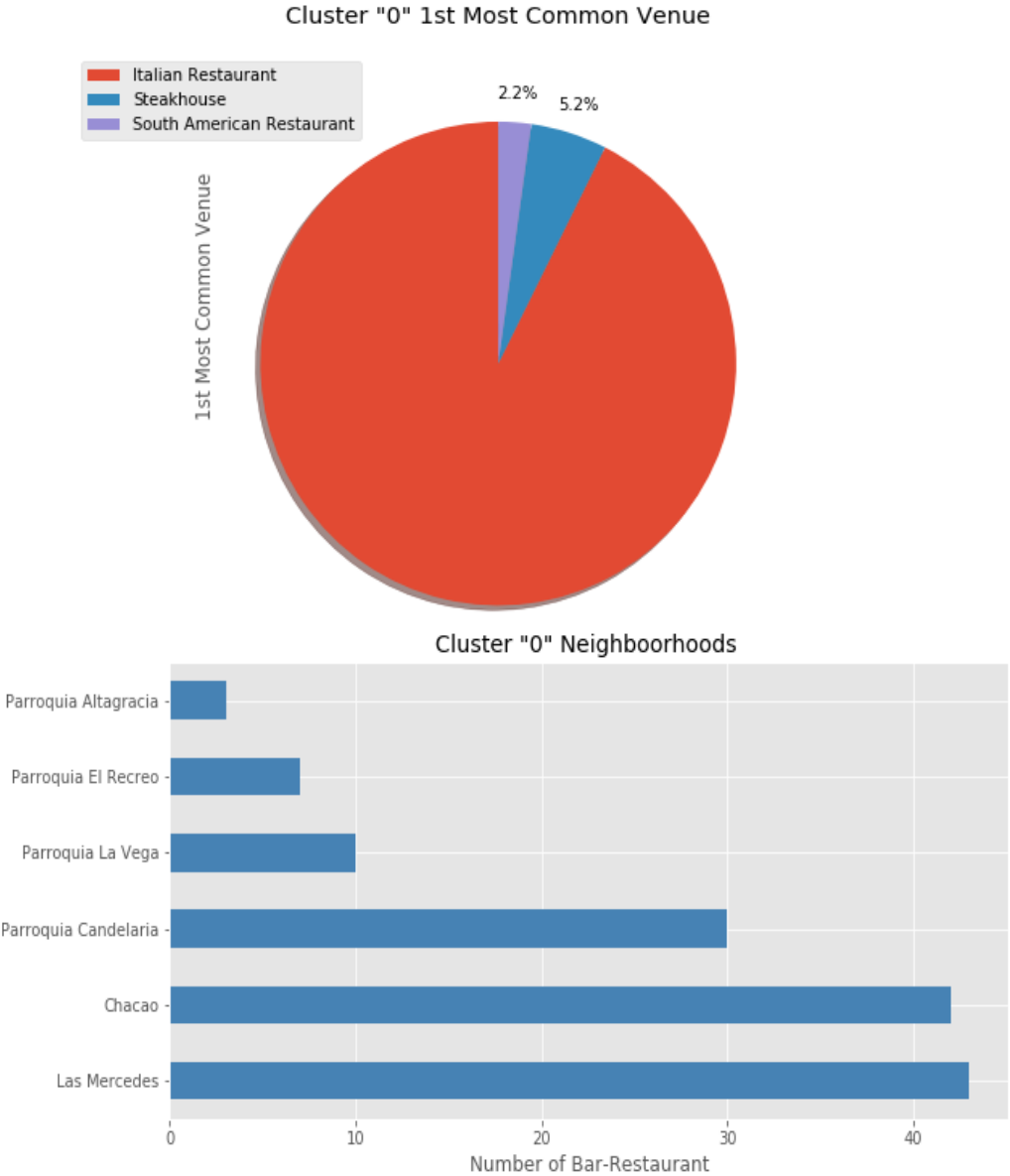
Most of these venues are located in "Las Mercedes", "Chacao", and "La Candelaria" neighborhoods.
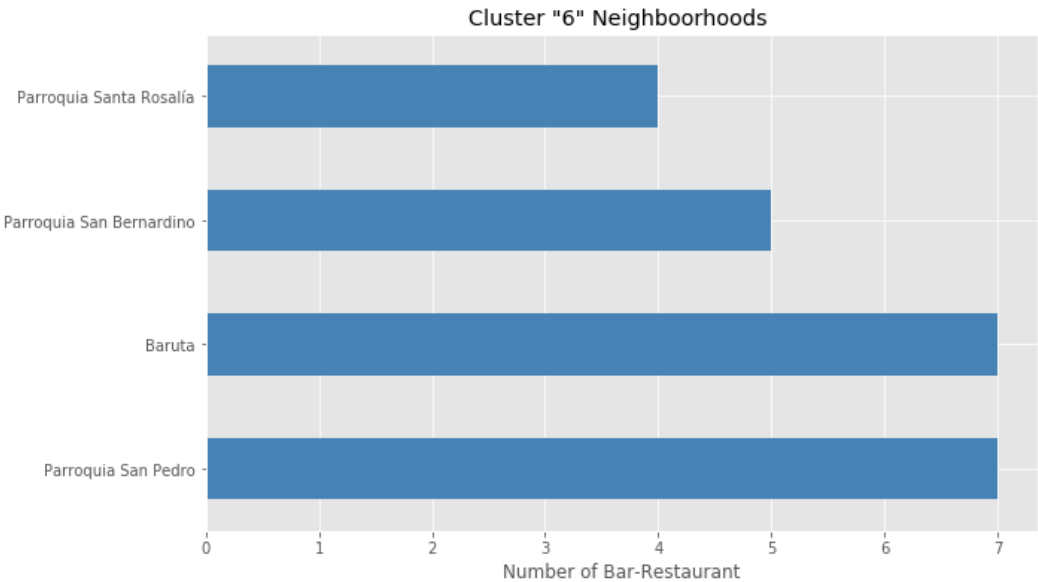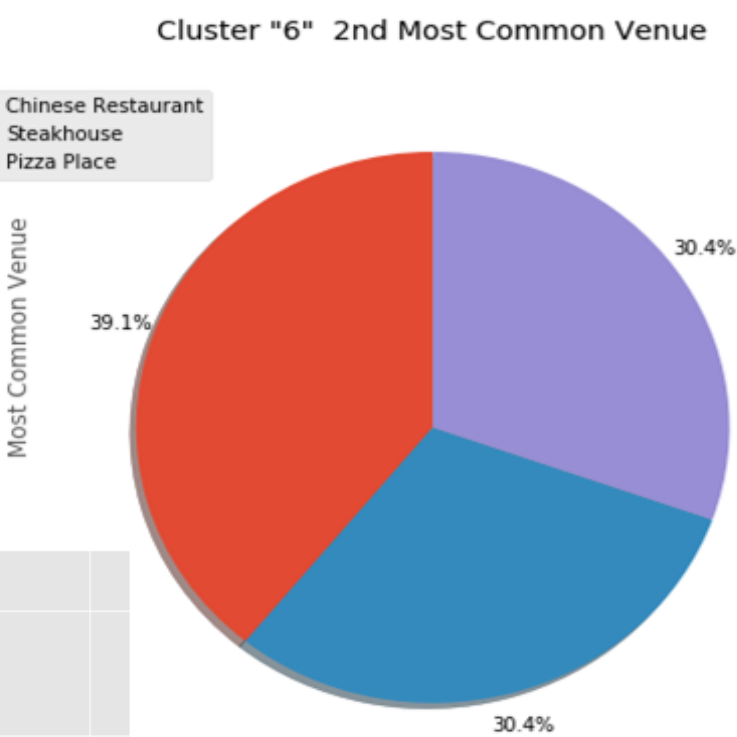
Results
Cluster '6'

The Most common, again, is the Italian Restaurant together with Pizza Place.
We have explored the second most common place in order to know the other venues we can frequently find with this label.

Most Common Venue	
Italian Restaurant	12
Pizza Place	11

2nd Most Common Venue	
Chinese Restaurant	9
Steakhouse	7
Pizza Place	7

Cluster "6" 2nd Most Common Venue
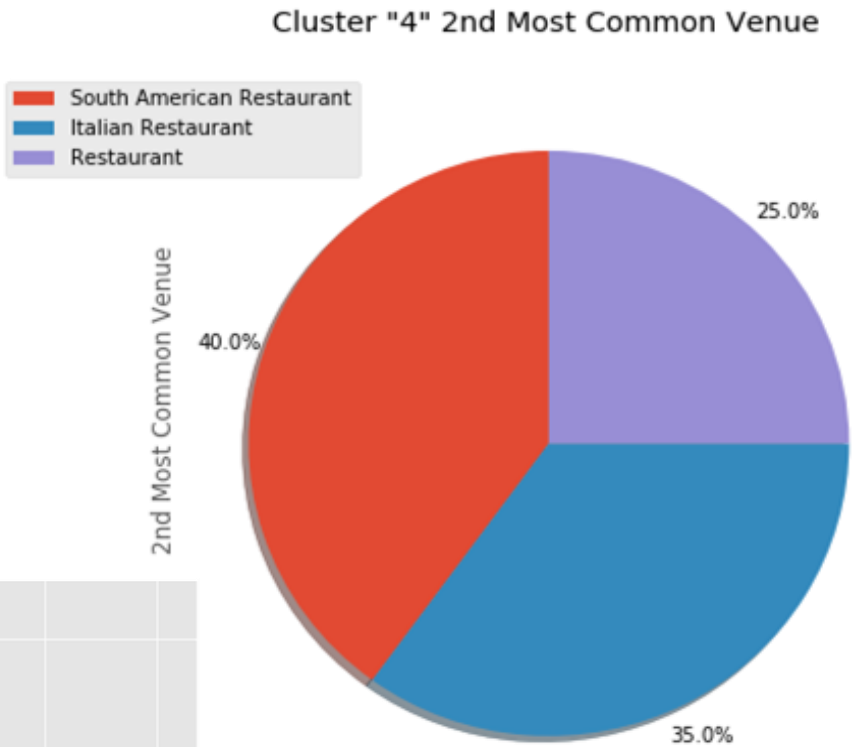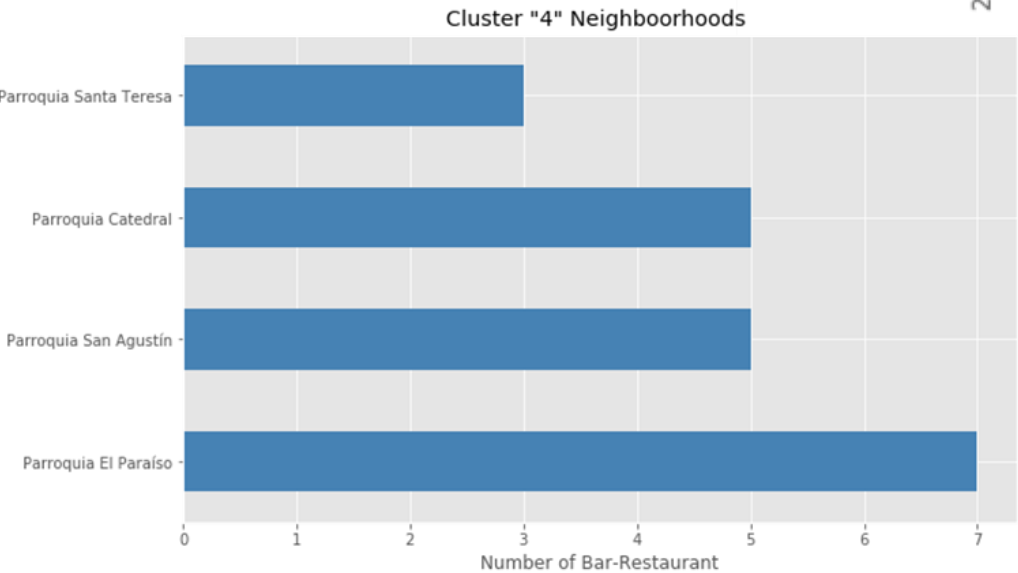
Cluster "6" Neighboorhoods

Chinese Restaurant has an important presence in the neighborhoods San Pedro, Baruta, San Bernardino and Santa Rosalía.

Results Cluster '4'

The most common places with '4' label cluster, is the Chinese Restaurant. The second most common places are South American Restaurant, Italian Restaurant and 'generic' restaurants.

2nd Most Common Venue	
South American Restaurant	8
Italian Restaurant	7
Restaurant	5

Cluster "4" 2nd Most Common Venue

Cluster "4" Neighboorhoods

These Bar-Restaurants are located in El Paraiso, San Agustin, Catedral and Santa Teresa neighborhoods.

Clusters 1, 2, 3 and 5

They are interesting to explore given the little competition that these places may have.

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Petare	5	Pizza Place	Tapas Restaurant	Mediterranean Restaurant	Asian Restaurant	Bar
Pueblo el Hatillo	3	Middle Eastern Restaurant	Tapas Restaurant	Mediterranean Restaurant	Asian Restaurant	Bar
Parroquia San Juan	2	Chinese Restaurant	Tapas Restaurant	Mediterranean Restaurant	Asian Restaurant	Bar
Parroquia San Juan	2	Chinese Restaurant	Tapas Restaurant	Mediterranean Restaurant	Asian Restaurant	Bar
Parroquia San Juan	2	Chinese Restaurant	Tapas Restaurant	Mediterranean Restaurant	Asian Restaurant	Bar
Parroquia El Valle	2	Chinese Restaurant	Tapas Restaurant	Mediterranean Restaurant	Asian Restaurant	Bar
Parroquia Coche	2	Chinese Restaurant	Tapas Restaurant	Mediterranean Restaurant	Asian Restaurant	Bar
Carrizal	1	Steakhouse	Tapas Restaurant	Mediterranean Restaurant	Asian Restaurant	Bar

We have observed that Petare, El Hatillo and Carrizal neighborhoods have Asian restaurants and Bars as the least common places in the city.

Discussion

- The results shows that Italian Restaurant, Pizza Place, Steakhouse, Spanish Restaurant, Mediterranean Restaurant and Chine Restaurant are the most common venues or Bar-Restaurant with high presence in the city.

- In contrast, we have observed that Steakhouse and Tapas Restaurant are the least common venues near the city and business center.

- We recommend evaluating these neighborhoods more, since they are in places near the city and the business center.

Conclusions

- ✓ We have found that most of the restaurants near the business center are Italian Restaurant, Pizza Place, Steakhouse, Spanish Restaurant, Mediterranean Restaurant and Chine Restaurants.

- ✓ They are mostly found in Chacao, Las Mercedes and Candelaria neighborhoods.

- ✓ We recommend locating a new bar-restaurant within these areas, as long as it is of a different type from the mentioned categories, as a Tape Restaurant.