Capstone Project – The Battle of Neighborhoods

1) Introduction

"Prospects of a Restaurant, where to install and which kind of food to sell in Santiago City, Chile."

Description of the problem:

This is an analysis for an entrepreneur who asked which neighborhood in Santiago would be the better to install a Gourmet Restaurant and what kind of competence can find around this place. To answer the requirement this project begin with an analysis of Santiago's Neighborhoods.

Discussion of the Background:

Santiago is the capital city of Chile, a country located in Latin America. Chile has 15 states, and Santiago is located in the "Metropolitan Region" State, this state has 7.073 millions of inhabitants at 2017 and represents nearly 40 of Chile population. Each State in Chile is divided in "Comunas", Santiago city that contains many "comunas", this Project will analyze this "comunas" and will try to define where is the best comuna to install a Gourmet restaurant, but also which kind of restaurant it can be found in those comunas, and what are their main venues.

People would be interested in this Project:

- Entrepreneurs who are interested in start a Gourmet Restaurant business in Comunas in Santiago.
- People who would like to live and know which kind places can be found on Comunas that are described in this project.
- Data Scientist who would like to extend this analysis to different Comunas in Santiago or different states either in a demographic or business approach

2) Data Acquisition

2.1 Wikipedia to get Comuna's Data

The Wikipedia page of "Santiago's Comunas" contains a table with information of Comunas and Demographic Indexes, this will be used to limit the Comunas that will be analized.

The Wikipedia page of "Chile's Comunas" contains a table with information of all Chile Comunas and geographic information, this will be used to limit the search venues limit.

To get this info it was used Beautifulsoup 4 and pandas library.

2.2 Geolocator to get Comuna's Latitude and Longitude

It was used Geopy to get Latitude and Longitude of the final Comunas, but after plot in Seaborn, some coordinates where wrong, for those cases Wikipedia page "Chile's Comunas" was consulted and added manually.

2.3 Foursquare API: to get venues of each Comuna and compare them

Foursquare API contributed to get Venues of each Comuna, no filter was used in foursquare query. For each Comuna 100 any kind of venues were obtained.

3) Methodology

The methodology that was applied in this project consisted in 3 steps

3.1 Literature review to understand social indexes in Santiago, in order to get richer Comunas to install a Gourmet Restaurant. Define a selection of Comunas to work with and get all data needed.

3.2 Get venues of each Comuna from Foursquare to:
   - Analyze top 10 more frequent Venues for Comuna.
   - Analyze Category Venues to get only that are restaurants, sort and classify by Comuna

3.3 Use of K-means algorithm to Compare Comunas and know which are similar between them.

4) Results

4.1 Literature Review

In literature review step to understand how political divisions is used in Chile, this map of Metropolitan Region was found, Santiago's Comunas are divided by range of Human Develoxptment index at 2017, from Dark Green (higher index) to Dark Red (lower index) the comunas where catalogued.

This was the first approach to look for the richer Comunas, those with index higher than 0.9 are 8 and are located in the north-east side of Santiago, their names sort by HDI score are Vitacura, Las Condes, Providencia, Ñuñoa, La Reina, Santiago Centro, Lo Barnechea and Macul.
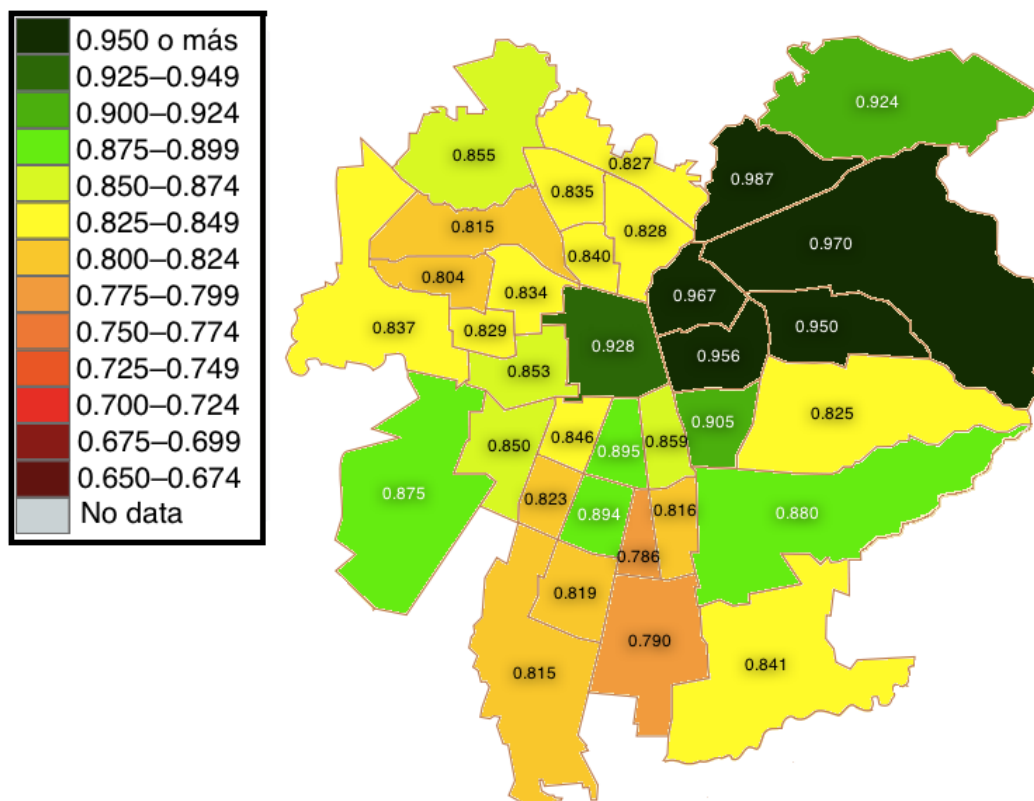
In the same Wikipedia website it can be found Urban Quality index (ICVU in Spanish). Following Dataframe shows Santiago's Comunas sorted by this Index.

```
df_comunas
```

Out[11]:

	Comuna	Population_2017	Houses	Population_growth	ICVU_index	Poberty
0	Vitacura	85384	23878	6.2%	75.59	2.8
1	Las Condes	294838	82099	19.6%	74.54	4.8
2	La Reina	92787	25768	-3.8%	70.45	9.8
3	Lo Barnechea	105833	16124	43,2%	69.09	17.3
4	Providencia	142079	51183	21.4%	68.79	4.6
5	Ñuñoa	208237	54692	28.2%	66.28	10.7
6	Macul	116534	29870	4.1%	61.92	17.5
7	Santiago	404495	99155	88.9%	59.27	11.6
8	San Miguel	107954	22655	38.6%	58.53	11.6
9	La Florida	366916	97137	0.6%	57.23	17.0
10	Maipú	521627	125939	11.5%	55.87	12.5
11	Quilicura	210410	35113	66.6%	55.86	18.5
12	Huechuraba	98671	16386	33.4%	55.70	23.8
13	Peñalolén	241599	51542	11.8%	54.28	20.7
14	La Granja	116571	32035	-12.2%	51.97	24.5
15	San Joaquín	94492	24233	-2,1%	51.88	23.7
16	La Cisterna	90119	22817	6%	51.63	20.0
17	Recoleta	157851	36606	7.2%	51.62	26.2
18	Pudahuel	230293	47902	18%	51.11	20.5
19	San Bernardo	297262	60723	20.4%	51.00	22.0
20	Estación Central	147041	32357	16.6%	49.96	14.5

The 8 Comunas with better ICVU Index are the same with HDI Index adobe 0.9 points, they also have the lowest Poverty percentage except by Lo Barnecha Comuna, but also has one of the highest Population Growth percentage. Finally this Project chose this 8 Comunas to analyze.

As Comunas don't have the same size, Square Kilometer size was added from another Wikipedia Website, this allowed to calculate the "Radius" Variable, a input for foursquare API.

```
df_merge1
```

Out[21]:

	Comuna	Population_2017	Houses	Population_growth	ICVU_index	Poberty	Square_KM2	Radius
0	Vitacura	85384	23878	6.2%	75.59	2.8	28.3	3001
1	Las Condes	294838	82099	19.6%	74.54	4.8	99.0	5613
2	La Reina	92787	25768	-3.8%	70.45	9.8	23.0	2705
3	Lo Barnechea	105833	16124	43,2%	69.09	17.3	1024.0	18054
4	Providencia	142079	51183	21.4%	68.79	4.6	14.3	2133
5	Ñuñoa	208237	54692	28.2%	66.28	10.7	16.9	2319
6	Macul	116534	29870	4.1%	61.92	17.5	12.9	2026
7	Santiago de Chile	404495	99155	88.9%	59.27	11.6	23.2	2717

Next step consisted in ask to Geopy for Latitude and Longitud, after get results and plotting Coordinates, these changes were applied

➢ Santiago as a normal city name in Latin America, was changed to Santiago de Chile to get correct coordinates
➢ Lo Barnechea and Las Condes were wrong in their coordinates and were changed manually.

This is the Final Data frame that was used for this project.

`df_final`

Out[26]:

	Comuna	Population_2017	Houses	Population_growth	ICVU_index	Poberty	Square_KM2	Radius	Address	Latitude	Longitude
0	Vitacura	85384	23878	6.2%	75.59	2.8	28.3	3001	Vitacura, CL	-33.380206	-70.565796
1	Las Condes	294838	82099	19.6%	74.54	4.8	99.0	5613	Las Condes, CL	-33.416667	-70.600000
2	La Reina	92787	25768	-3.8%	70.45	9.8	23.0	2705	La Reina, CL	-33.447373	-70.533412
3	Lo Barnechea	105833	16124	43,2%	69.09	17.3	1024.0	9027	Lo Barnechea, CL	-33.350000	-70.516667
4	Providencia	142079	51183	21.4%	68.79	4.6	14.3	2133	Providencia, CL	-33.428838	-70.611337
5	Ñuñoa	208237	54692	28.2%	66.28	10.7	16.9	2319	Ñuñoa, CL	-33.454330	-70.600582
6	Macul	116534	29870	4.1%	61.92	17.5	12.9	2026	Macul, CL	-33.491943	-70.599732
7	Santiago de Chile	404495	99155	88.9%	59.27	11.6	23.2	2717	Santiago de Chile, CL	-33.437797	-70.650445

This is a map of Comunas location in Santiago City.
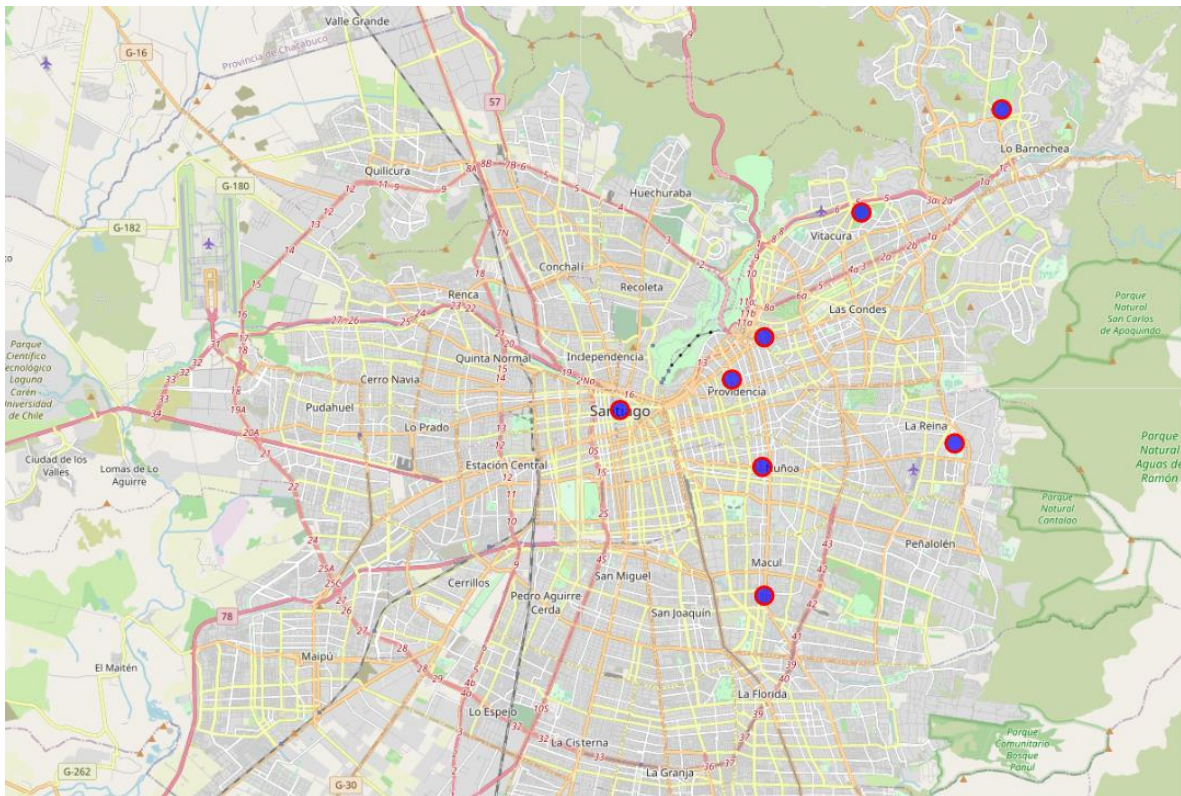
4.2 Fourquare API query

After Foursquare query top 10 Venues for these Comunas were found

10 Most Frequently Occuring Venues in the 8 Selected Comunas of Santiago
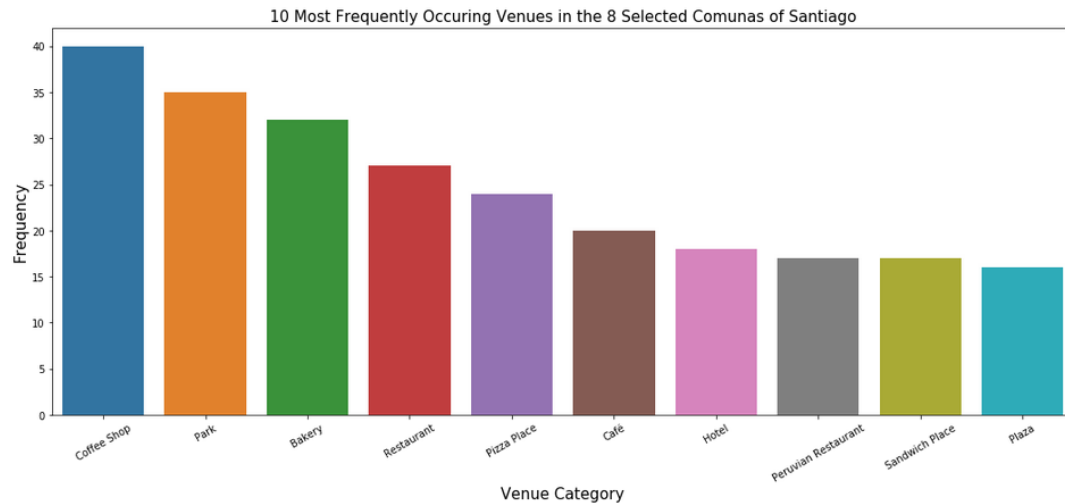
In top 10 categories there are basic food places like Coffee, Sandwich and Pizza Places, also some Parks and Plazas, there are Restaurant but without category and Peruvian Restaurant, but it's necessary to look for each Comuna which are Top 10 venues category.

	Comuna	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	La Reina	Soccer Field	Restaurant	Park	Pizza Place	Coffee Shop
2	Las Condes	Park	Bakery	Coffee Shop	Hotel	Peruvian Restaurant
3	Lo Barnechea	Bakery	Coffee Shop	Shopping Mall	Gym	Park
4	Macul	Chinese Restaurant	Restaurant	Soccer Field	Pharmacy	Hot Dog Joint
5	Providencia	Coffee Shop	Hotel	Park	French Restaurant	Indian Restaurant
6	Santiago de Chile	Coffee Shop	Café	Park	Pizza Place	Theater
7	Vitacura	Café	Park	Hotel	Gym / Fitness Center	Bakery
8	Ñuñoa	Coffee Shop	Bakery	Peruvian Restaurant	Pizza Place	Plaza

6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Plaza	Pet Store	Farmers Market	Sushi Restaurant	Racetrack
Scenic Lookout	Ice Cream Shop	Japanese Restaurant	Mediterranean Restaurant	Liquor Store
Supermarket	Motorcycle Shop	Trail	Golf Course	Gym / Fitness Center
Sushi Restaurant	Pizza Place	Sandwich Place	Food Truck	BBQ Joint
Ice Cream Shop	Bakery	Café	Scenic Lookout	Restaurant
Hotel	Sushi Restaurant	Bookstore	Plaza	Scenic Lookout
Restaurant	Italian Restaurant	Seafood Restaurant	Gym	Coffee Shop
Restaurant	Italian Restaurant	Sandwich Place	Café	Park

Comunas with more Restaurant also were analyzed:

Number of Restaurants as Top-Most Venues
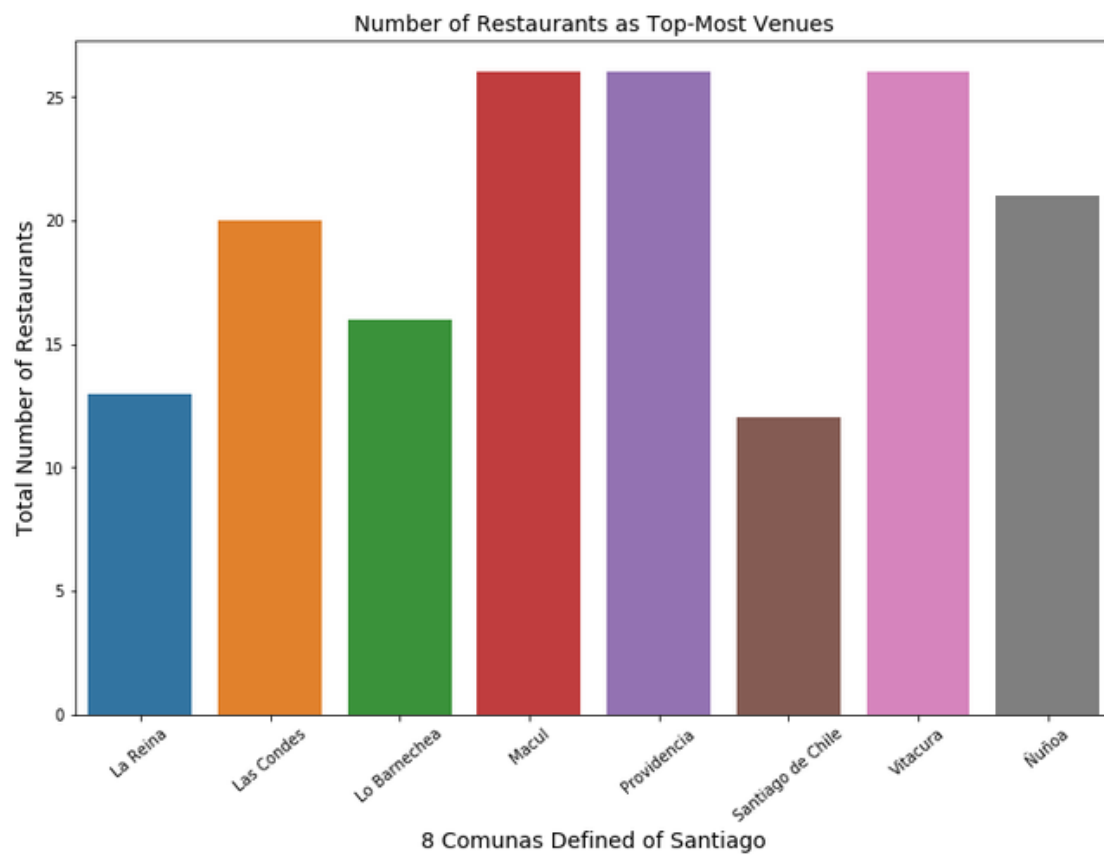
4.3 Cluster of Comunas

The clustering process gave those results for k = 4.
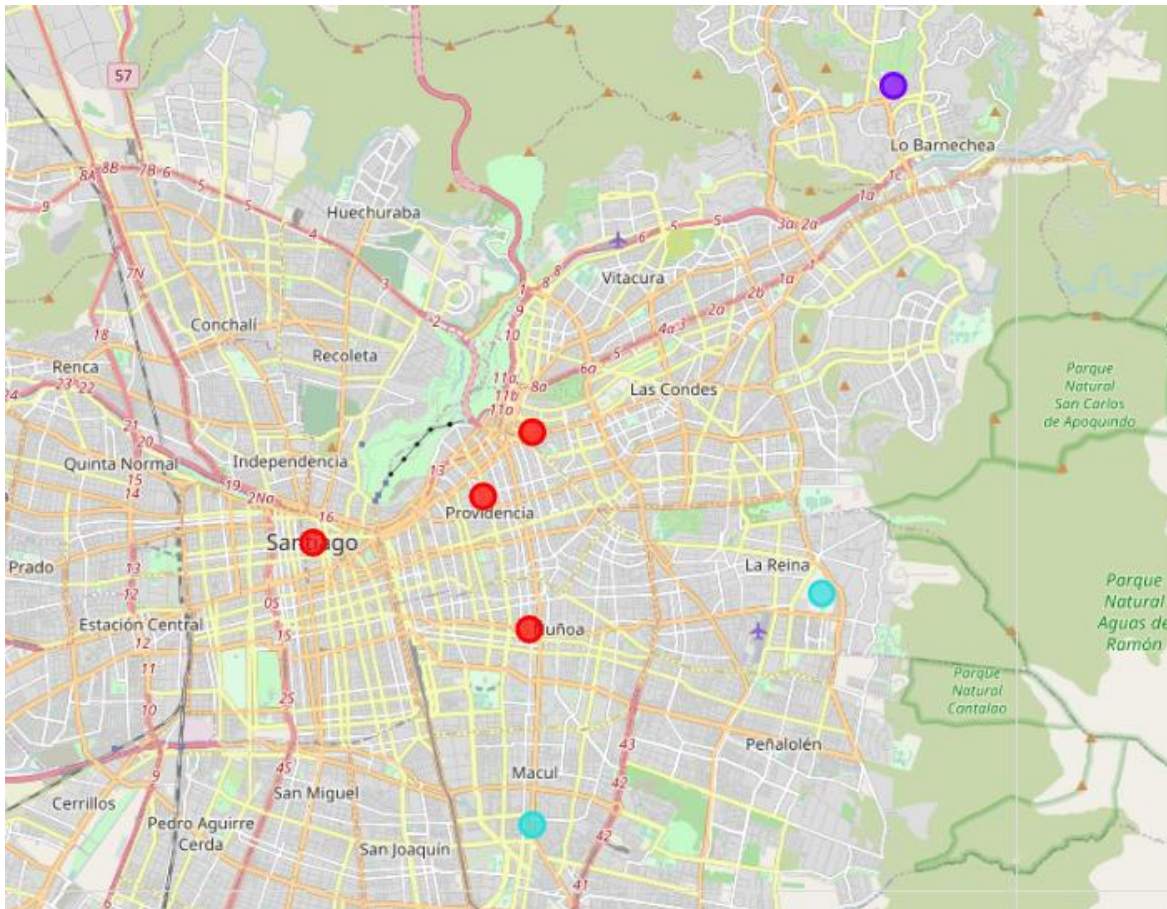
5) Discussion

Since the clustering was based only on the most common venues o each Comuna, Clusters can be defined in these term.

- ➢ **Cluster 0 (Las Condes – Providencia – Ñuñoa – Santiago):** Cluster with several Coffee shops at top category venue, bakeries and pizza places followed by Hotels and parks and some Restaurants in the end of this top.
- ➢ **Cluster 1 (Lo Barnechea):** Cluster without Restaurants in top 10 Venues.
- ➢ **Cluster 2 (Macul – La Reina):** Cluster of Fast Food places as Pizza, Hot Dog, Sandwiches places and Sushi Restaurants in top 10 venues.
- ➢ **Cluster 3 (Vitacura)** Cluster with more Restaurants venues in top 10 Category also are several parks, gyms and coffee shops venues.

Looking the other attributes of comunas are also grouped by:

- ➢ **Cluster 0:** Comunas close to city center, with the highest number of houses and in the middle-high Cluster Social indexes.
- ➢ **Cluster 1:** Cluster of Comunas far from City Center, biggest Square meters and highest poverty.
- ➢ **Cluster 2:** Cluster of Comunas at southeast of Santiago, low population growth.
- ➢ **Cluster 3:** Cluster of Comunas with highest ICVU index and lowest poverty Percentage.

According to these results the recommendations to answer the question where to install a Gourmet Restaurant

Vitacura would be good Comuna to install a Gourmet Restaurant, but also the Comuna with more Restaurant, Italian and Sea Food kind of restaurant are already positioned in top venues, would be better to look for a different kind of Food.

Comunas of Cluster 0 would be a second option to install the Restaurant, less Restaurant competitors can be found and more people lives in this Comunas, but these people could look for less elaborated kind of food.

Lo Barnechea would be third option for entrepreneur, there are no competitors in this Comuna looking at top venues, this comuna has the highest poverty percentage of the 8 richest Comunas of Santiago, also the Population growth is high, this Comuna will be the riskiest place to install a Gourmet Restaurant but can be the first in doing it. A deeper study of this comuna is suggested.

Comunas of Cluster 2 are not a recommended for a Gourmet Restaurant, many basic food can be found in these comunas.

6) Conclusion

After analyze 8 richest comunas of Santiago de Chile, 4 cluster of comunas can be found by their more frecuent Venues Category, it was also interesting to find out that how venues categories are related to social indexes such as poberty and ICVU, more restaurants are placed in comunas with better indexes and more basic food places in comunas with lower social indexes.

If a richest comuna is chosen it is possible to find a lot of competitors, as you decrease in comuna money spend capacity less restaurant are found but basic food places competitors take their place like Pizza, Sushi and Sandwiches sellers.

A different strategy is also recommend, trying to install a Restaurant in comunas with nice indexes, high population growth that has no competitors (Lo Barnechea), this could be a risky decision but risk can be handled with deeper analyzes focused in a particular area.

For the reason exposed adobe, is not easy to define the place to install a Gourmet Restaurant, marketing strategy, type of food to be sold, number and kind of competitors the characteristics of people who lives in the comuna and the risk aversion of the entrepreneur needs to be defined for next steps, but this first approach of Santiago de Chile Comunas gives to this entrepreneur a better picture of what can be expected.