Restaurant Location

Contents

1 Introduction

Santiago is the capital and largest city in Chile, South America. With a population of over 6 million people and the most important city in terms of industry and finances, generating 45% of the country GDP. It attracts a lot of entrepeneurs to develop a new business here.

Santiago is divided in 35 Boroughs called *Comunas*, all connected through public transport. Each Comuna is different socially and economically, so it's important to make a difference between each of them.

In recent years the food scene has flourished in Santiago, with hundreds of new restaurants showing a modern view of chilean cuisine as well as from international heritage.

1.1 Business Project

Being such a large city with so much to offer, our main goal is to help future entrepeneurs looking to put a profitable restaurant, of any kind in Santiago, to choose the best location and improving the probability of success, in order to accomplish that we will take into consideration variables such as poverty index, population density and how well connected is through the city subway *Metro*

1.2 Interests

Fellow entrepeneurs looking to develop a new business or expand an existing one, making decisions about location based in data, also this analysis can be aplicable to other kind of business such as pubs or clubs.

2 Data acquisition and Cleaning

2.1 Data Sources

The information about the list of Comunas, poverty index, population density and postcode were fetched from Wikipedia which has the latest info about population density taken from the last Census in the year 2017. With the postcode and name of each Comuna was possible to obtain the geolocation of each borough through the webpage *download.geonames*.
The information of Venues in Santiago was obtained using Foursquare API, and the information of each subway stations and its geolocation was fetched using GoogleMaps API.

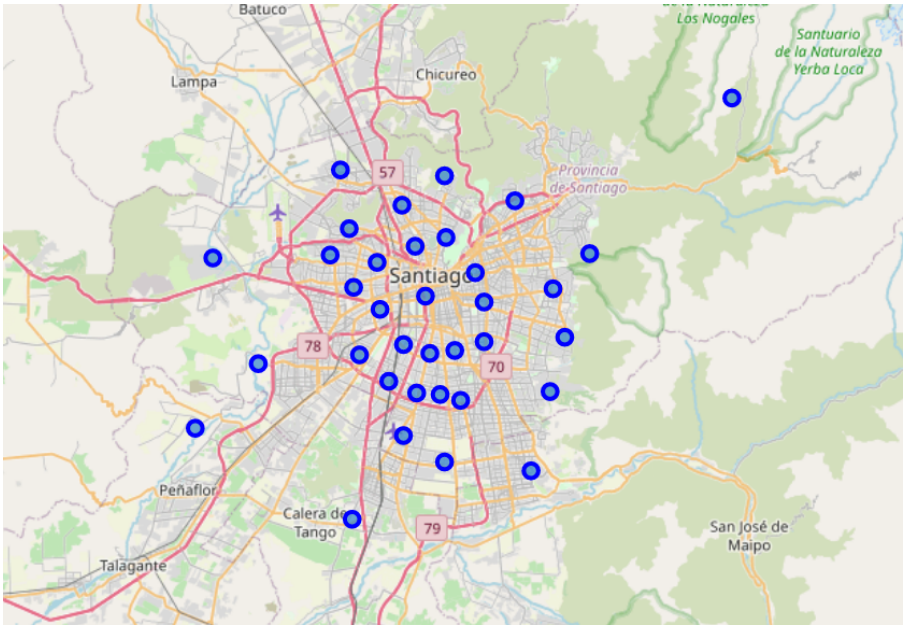
2.2 Data Cleaning

Data obtained from multiple sources were combined into one table that had all the information about each Comuna. There were no missing values from the information from Wikipedia and the webpage *download.geonames*, so we didn't have to look for other extra sources of information. From the Foursquare API we acquire the information about venues in each Comuna, then we chose the five Comunas with the largest number of venues i.e. the more commercial areas, and from GoogleMaps API we took the information of subway stations on the five Comunas and an important data to choose from here was how many station are in each borough to make the analysis. This information is in the table below.

	Comuna	Population Density	Poverty index	lat	lng	Venue	Metro count
0	Nunoa	9698.16	10.7	-33.4593	-70.6003	22	11
1	Providencia	8429.15	4.6	-33.4362	-70.609	29	11
2	San Miguel	8122.76	11.6	-33.4991	-70.6517	24	5
3	Santiago	8654.83	11.6	-33.4541	-70.656	23	21
4	Vitacura	2846.63	2.8	-33.3799	-70.5724	25	0

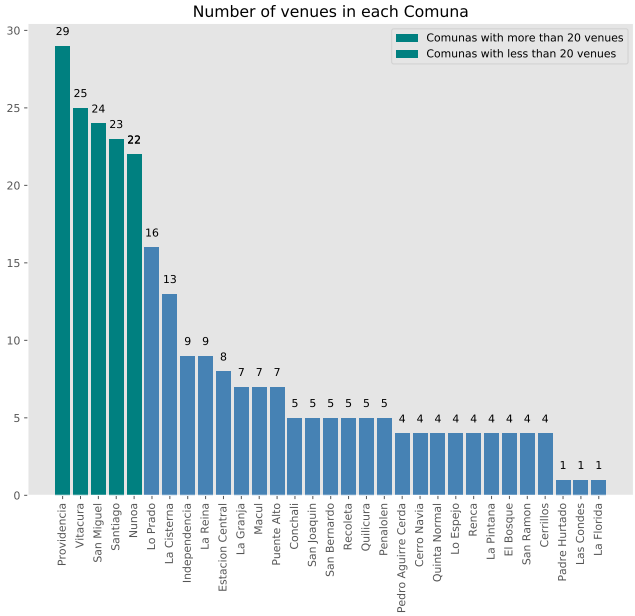
3 Exploratory Data Analysis

Using Folium we can visualize where is each Comuna in a map with its geolocation.



3.1 Best commercial Comuna

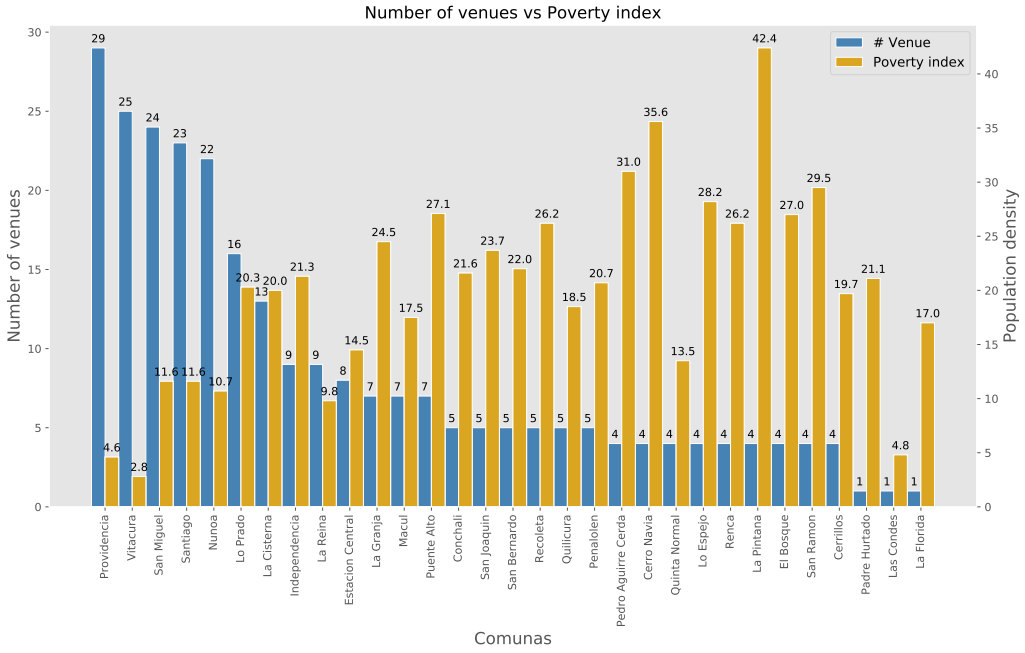
Using the Foursquare API, we obtained the information of all venues in Santiago city grouped by each Comuna.The main goal of this analysis is to locate the five most commercial Comunas in Santiago. We construct a bar plot with the number of venues in each Comuna for comparison.



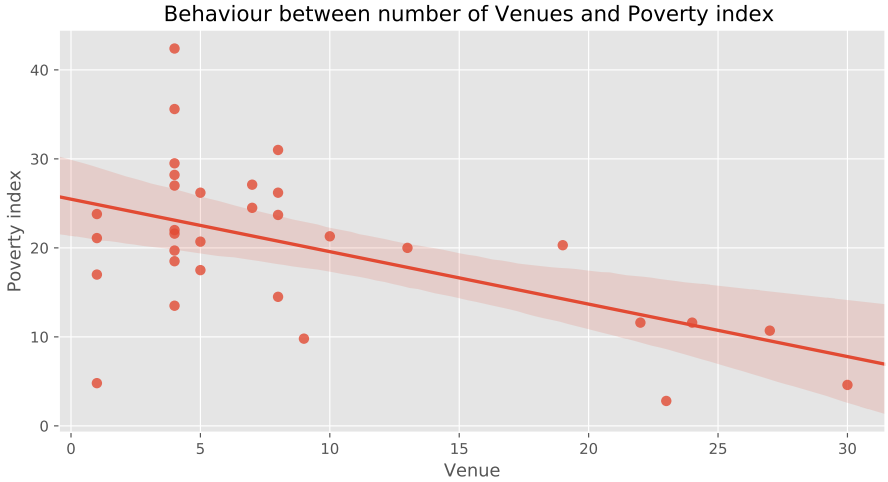
The chosen Comunas are clearly more popular with respect of number of venues than the others, so certainly is a good parameter to choose where people is more willing to invest money in a new venue.

3.2 Comparison with poverty index

With the data obtained from Wikipedia concerning the poverty index, we made a comparison between the number of venues of each comunas, and the poverty index. We will focus our atention on the wealthiest Comunas.



As we can see in the barplot, the chosen comunas present a low to medium poverty index, but to make sure if there is any relation between them we present a scatterplot with a regression line to have an intuition of this.



From this last graph is difficult to make conclusions, but with a correlation matrix we can make a more informed decision.

	Venue	Poverty index	Population Density
Venue	1.000000	-0.538381	0.089366
Poverty index	-0.538381	1.000000	0.395554
Population Density	0.089366	0.395554	1.000000

Figure 1: Correlation matrix between number of venues, poverty index and population density

As we can see from the plots and the correlation matrix there is no linear relation between the number of venues, population density and poverty index.

4 Results

4.1 Predicting modelling

We will use the simplest model to cluster our Comunas, *k-means* will be extremely useful to determine which placement will be best for our hypothetical restaurant. To help processing the information in a more uniform distribution, we need to normalize our data, as the features we are analysing are in different scales.

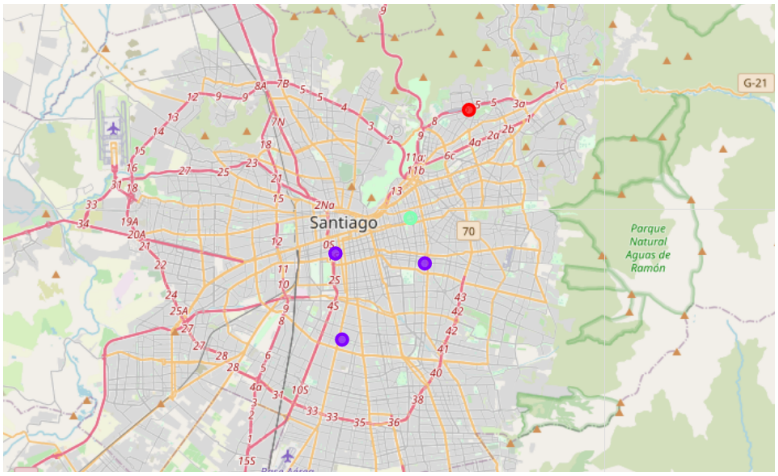
4.2 K-means clustering

As we don't have a big dataframe to analize and separate between train and test, just the information of five comunas, we choose to separate the information en three clusters, $k = 3$. The information used to put in the model is:

	Population Density	Poverty index	Venue	Metro count
0	9698.16	10.7	22	11
1	8429.15	4.6	29	11
2	8122.76	11.6	24	5
3	8654.83	11.6	23	21
4	2846.63	2.8	25	0

4.3 K-means results

Using Folium we can visualize where is each chosen Comuna and in which Cluster there are.



4.4 Clusters Summary

And the results of using *K-means* are summarize in the tables below.

Cluster 1

	Comuna	Population Density	Poverty index	lat	lng	Venue	Metro count	Cluster Labels
4	Vitacura	2846.63	2.8	-33.3799	-70.5724	23	0	0

- Low Population Density
- Low Poverty index
- Low number of subway stations

Figure 2: Summary of Cluster 1

Cluster 2

	Comuna	Population Density	Poverty index	lat	lng	Venue	Metro count	Cluster Labels
0	Nunoa	9698.16	10.7	-33.4593	-70.6003	27	11	1
2	San Miguel	8122.76	11.6	-33.4991	-70.6517	24	5	1
3	Santiago	8654.83	11.6	-33.4541	-70.656	22	21	1

- High Population Density
- High Poverty index
- High and low number of subway stations

Figure 3: Summary of Cluster 2

Cluster 3

	Comuna	Population Density	Poverty index	lat	lng	Venue	Metro count	Cluster Labels
1	Providencia	8429.15	4.6	-33.4362	-70.609	30	11	2

- High Population Density
- Low Poverty index
- Medium number of subway stations

Figure 4: Summary of Cluster 3

5 Conclusions

For this project we labeled the most commercial Comunas in Santiago, according to, how many subway station are, its poverty index and population density.

We chose comunas with a large number of venues for its commercial interest, even though this can be counterintuitive because if there is a high level of competition they can have customer loyalty to them, on the other hand a high number of venues can imply customers looking for a more diverse kind of option. We will work with the assumption of the latter.

Also another important aspect is the number of subway stations. We are looking for places with easy access. Metro stations is the most prefered form of transportation of *Santiaguinos*, although Santiago has an integrated transport system that includes, besides the subway, an extensive network of buses, however they are not as well mantained as the subway and also its very slow in comparison due to the high traffic in a big city as Santiago.

The next aspect in our analysis is the poverty index of each Comuna. We would like to place our restaurant in a Comuna where people have a high spending capacity and can afford to indulge themselves in eating out. The target of the price restaurant itself its not specified, but no matter if its a cheap, midprice or expensive restaurant, this high spending capacity works for all of them.
And the last aspect is the population density. At first, one can think that it can be a relation between the number of venues in each Comuna and the population density, but our statistical analysis showed that the problem is far more complex than this, and there is no strong lineal relation between the two variables, but still is preferable a Comuna with high population density for the possibility of the incresed visibility this can bring.

Looking at our *K-Means* analysis the cluster that fits all the criteria described above is cluster number 3. This cluster is conformed by the Comuna:

- Providencia

For our future entrepeneurs looking for developing a new restaurant in Santiago or looking to expand an existing business, our Client would like to consider the Comuna Providencia as a location, preferable in a direction around 300 meters around a subway station.

6 Future Directions

An interesting way to improve this model is to refine the target and study a specific type of restaurant, this can be achived by using the GoogleMaps API. Foursquare did not have an up to date information of venues in Santiago resulting in not too many information to work with and not too many options of venues to analize. Also it could be useful to include the information of Household income and to refine the search in smaller neighborhoods instead of Comunas.

7 References

1. $https://es.wikipedia.org/wiki/Anexo:Comunas_de_Santiago_de_Chile$
2. $https://es.wikipedia.org/wiki/Anexo:CB3digos_postales_de_Chile$
3. $http://download.geonames.org/export/zip/CL.zip$
4. $https://es.foursquare.com/$
5. $https://developers.google.com/maps/documentation$