Capstone Project Report

Introduction

The city of Santiago in Chile is known for its huge seggregation. The city is divided into 31 different communes. The upper class is mainly concentrated in the north eastern communes. The lower class lives mainly in the southern ones, and little to none mix of classes is seen across the city. However, the country has experienced a great economic growth throughout the last decades, lifting the lower class into a higher economic status. This, together with a growth in population, raised the prices of housing across the whole city.

In this project I will focus on trying to distinguish the economic status of a commune based on the type of venues that are located within it. Knowing the type of venues that form part of the communes with higher economic status could be of great importance for real estate agencies. With this information the will be able to determine if a lower class commune is getting a higher economic status if it has similar venue types as higher class communes. If this is the case, they will wnat to invest on this places.

Data

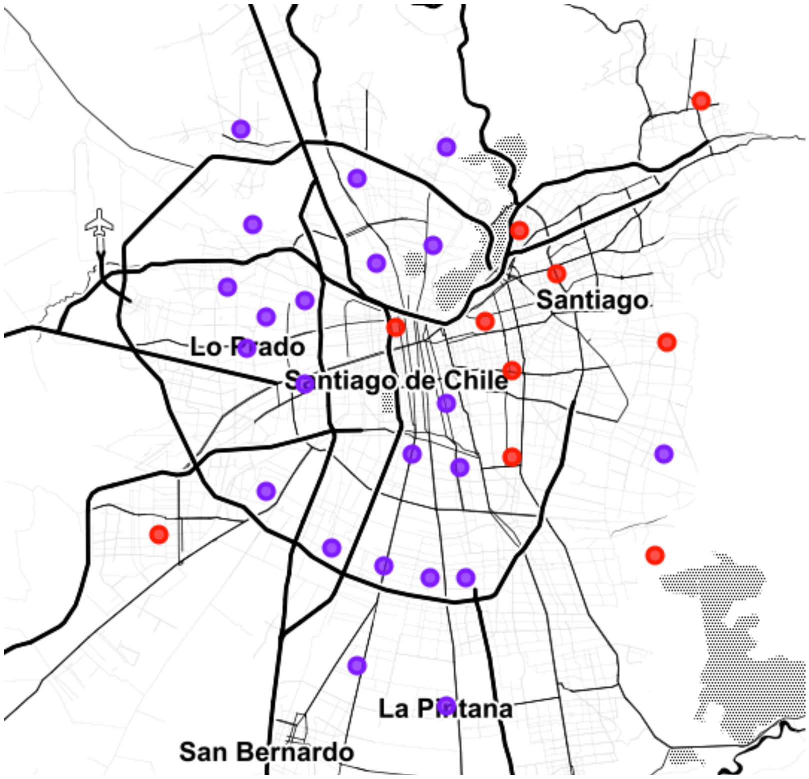
For this project I will require data about location and development indeces for each commune. Fortunately, I found this information on this link: https://es.wikipedia.org/wiki/Anexo:Comunas_de_Chile. I loaded this data into a dataframe and preprocessed it, obtaining a dasaet called santiago_data that contains the commune name, HDI (human development index and coordinates for each commune in santiago. Together with this information, I obtained data for venues in each Commune using the foursquare API, this information is contained in the santiago_venues dataset.

Methodology

At first, I focused on differentiating the communes with higher HDI from the one with lower HDI. Then, using the data about venues in each commune, I applied a K-Means methodology over the data frame, obtaining 5 different clusters that grouped communes with similar venues. Using this information, I was able to obtain which communes with lower income where similar, according to the type of venues, to ones with higher income. With this data, I was able to extract the type of venues that differentiate these communes.

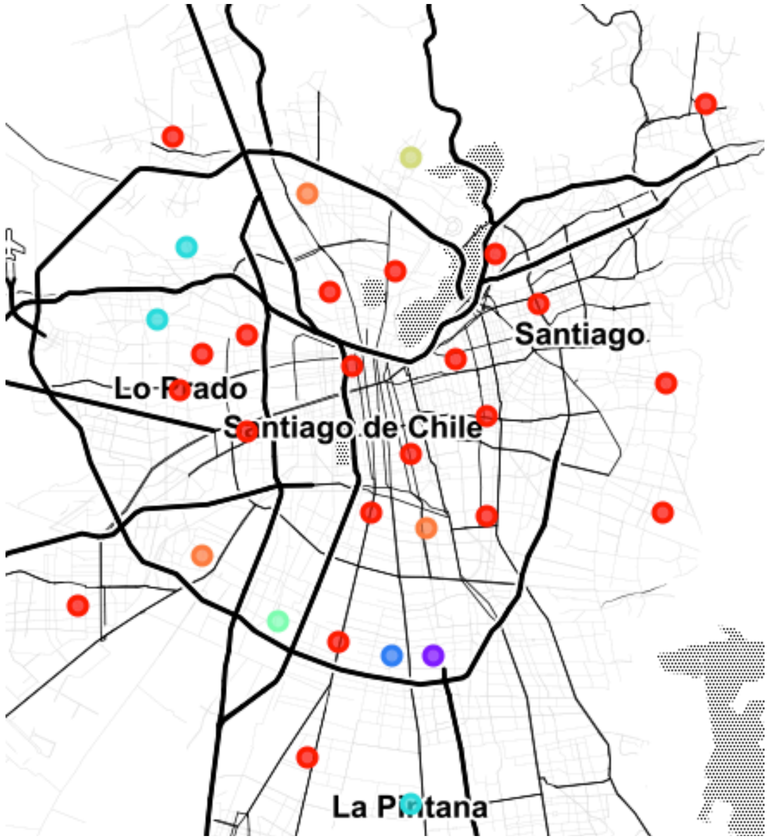
Results

As stated before, I divided de data frame into communes with higher HDI and ones with lower HDI. These two groups are displayed in the following map:



The red dots represent communes that are more developed in comparison to the ones represented by the purple dots. As I said before, the wealthier neighborhoods mainly concentrate in the upper east side of the city.

With the information about type of venues in each commune I applied a K-Means algorithm that allowed me to form 5 different clusters of communes according to their similarity of venues. Each one of these clusters is represented with a different color in the following map:



We can observe that the communes with higher income became part of the same cluster. This shows us that they have similar types of venues. However, these are not the only communes that form part of this cluster. Other communes with lower income like “Quinta Normal”, “Pudahuel” or “Lo Prado” also have similar types of venues. This could show us that this neighborhoods have the potential of becoming more developed and lifting the prices of houses among them. We can also observe some of the most common types of venues in these communes in the following table:

Neighborhood	Cluster	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
Santiago	1.0	0.0	Coffee Shop	Peruvian Restaurant	Pizza Place	Sandwich Place	Chinese Restaurant	Burger Joint	Asian Restaurant	Japanese Restaurant	Bakery	Sushi Restaurant
El Bosque	2.0	0.0	Pet Store	Pizza Place	Food & Drink Shop	Diner	Farmers Market	Falafel Restaurant	Electronics Store	Donut Shop	Dive Bar	Dessert Shop
Estación Central	2.0	0.0	Residential Building (Apartment / Condo)	Gym	Argentinian Restaurant	Food Truck	Restaurant	Japanese Restaurant	Yoga Studio	Farmers Market	Falafel Restaurant	Electronics Store
Huechuraba	2.0	0.0	Market	Outdoors & Recreation	Ice Cream Shop	Dive Bar	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Donut Shop
Independencia	2.0	0.0	Food	Fried Chicken Joint	Park	Plaza	Asian Restaurant	Sandwich Place	Diner	Farmers Market	Falafel Restaurant	Electronics Store
La Cisterna	2.0	0.0	Sushi Restaurant	Pizza Place	Chinese Restaurant	Bakery	Fast Food Restaurant	Middle Eastern Restaurant	Basketball Court	Farmers Market	Falafel Restaurant	Electronics Store
La Reina	1.0	0.0	Sushi Restaurant	Coffee Shop	Café	Gourmet Shop	General Entertainment	Italian Restaurant	Liquor Store	Fish & Chips Shop	Cupcake Shop	Chinese Restaurant
Las Condes	1.0	0.0	Restaurant	Fast Food Restaurant	Bakery	Creperie	Plaza	Dessert Shop	Sandwich Place	Coffee Shop	Bike Shop	Salad Place

We can observe that the most common venues in these locations are mainly established restaurants. The main difference I observed with the other clusters of communes is that the last ones have more fast-food restaurants, meaning cheaper options of food. This could be the reason why they don't form part of the cluster with high income communes.

Discussion

To be more certain about how useful this information to real estate agencies is, it would be desirable to obtain the information about house prices in Santiago. It would be interesting if further studies focused on whether communes similar to the ones with higher income, have raised their prices in the last years.

It would also be useful if more analysis is put into the model of K-means obtained in this work, to check how accurate it is.

Conclusion

Using a K-Means algorithm, I was able to distinguish communes similar to ones with higher HDI. This data could be of great interest for real estate agencies, because it shows wich communes are best to invest in.