CAPSTONE PROJECT

The Battle of neighborhoods | Moving to Santiago by Francisco Vásquez Pozo.

Content.

- 1. Discussion of the problem and problem background.
  - 1.1. Introduction.
  - 1.2. Background of Santiago.
  - 1.3. Target audience.
  - 1.4. Problem to solve.
- 2. Description of the data.
- 3. Methodology.
  - 3.1. Map od Santiago de Chile.
  - 3.2. Nearby venues and its categories.
  - 3.3. K-means Clustering Unsupervised Approach
- 4. Results.
  - 4.1. Map of Clusters.
- 5. Discussion.
- 6. Conclusion.

But first, lets have a look…

1. Discussion of the problem and problem background.

- **1.1. Introduction.¶**
- The main purpose of this project is to help people moving to Santiago de Chile to explore the facilities around different neighborhoods of the city, in order to settle in one of them.
- **1.2. Background of Santiago.¶**
- Santiago de Chile, is the capital and largest city of Chile as well as one of the largest cities in the Americas. It is the center of Chile's most densely populated region, the Santiago Metropolitan Region, the total population of which is 7 million, and more than 6 million of them live in the city's continuous urban area.
- Founded in 1541 by the Spanish conquistador Pedro de Valdivia, Santiago has been the capital city of Chile since colonial times. The city has a downtown core of 19th-century neoclassical architecture and winding side-streets.
- Santiago is the cultural, political and financial center of Chile and is home to the regional headquarters of many multinational corporations. The Chilean executive and judiciary branches are located in Santiago, but Congress itself meets mostly in nearby Valparaíso.

1. Discussion of the problem and problem background.

- **1.3. Target audience.**
- The project aims to aid people moving to Santiago de Chile to select the suitable neighborhood that suits their necessities. Particularly young professionals and newly graduated students. Since the biggest companies are located in the capital, fresh grads from different regions of the country are likely to relocate in a city that is mostly unknown for them. Therefore, it’s important for them to get access to the services that provide information necessary to select a new place to settle for a new life. And basically, this could be of use to anyone looking to relocate in the capital.
- **1.4. Problem to solve.**
- Provide a clustering algorithm that suggests anyone looking to relocate in Santiago de Chile with the amenities included in the neighborhoods.

2. Description of the data.

- The zip codes of the Metropolitan Region were obtained from: (The Santiago Metropolitan Region is made up of 6 provinces and 52 communes, from these I selected the 11 most important communes from the website) http://www.codigopostalchile.com/santiago-436
- Data about different venues is necessary, about different neighborhoods of the specific borough. In order to gain that information, we will use "Foursquare" location information. As previously discussed, Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even picures. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.
- The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue is as follows:

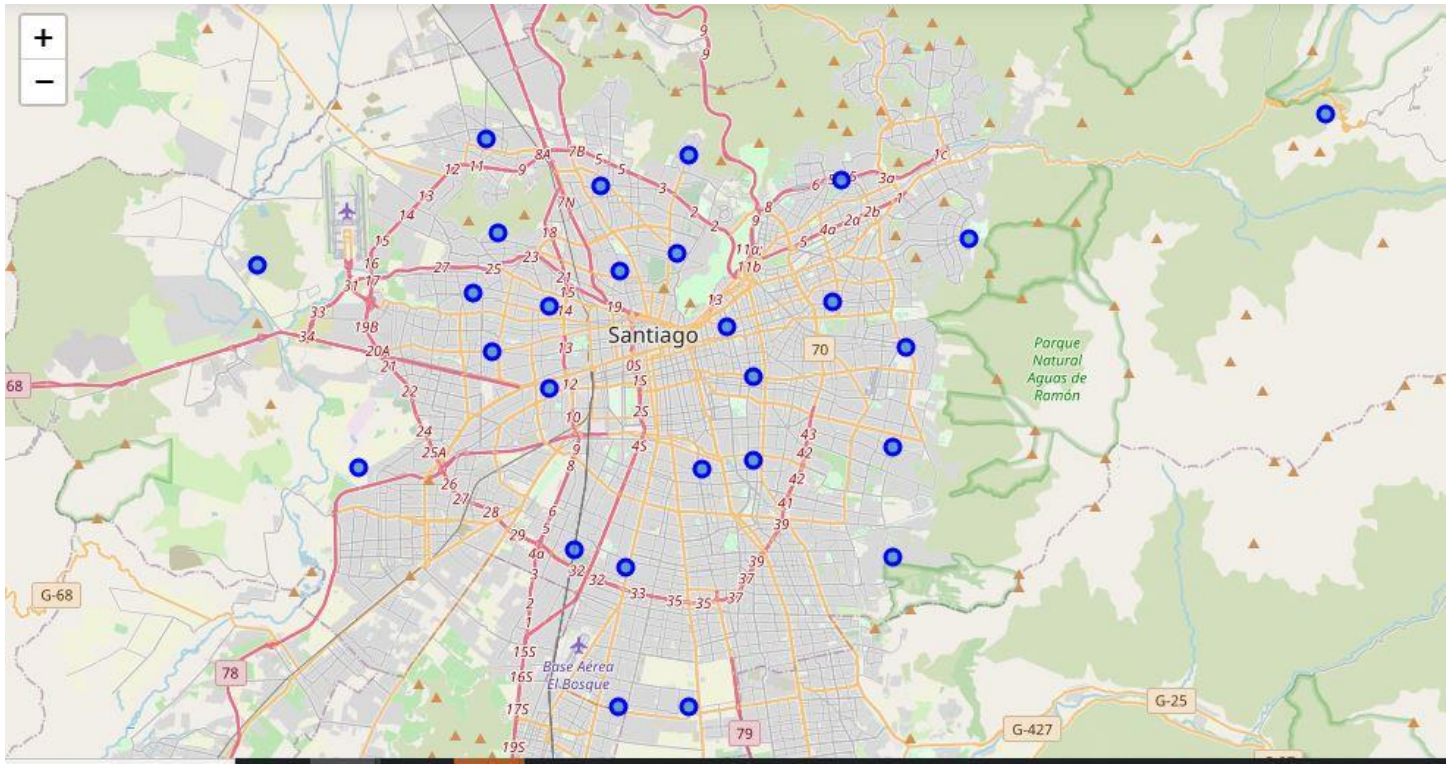
1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue
6. Venue Latitude
7. Venue Longitude
8. Venue Category

	Código	Neighborhood	Latitud	Longitud
0	7500000	Providencia	-33.43485	-70.61573
1	7550000	Las Condes	-33.40033	-70.50269
2	7630000	Vitacura	-33.37763	-70.56219
3	7690000	Lo Barnechea	-33.35153	-70.33607
4	7750000	Ñuñoa	-33.45449	-70.60383

2. Description of the data.

The head of the final dataset obtained is the following:

3.1. Map of Santiago de Chile



3. Methodology

First off, is important for us to have a clear vision of the location we are working on. The last census of 2017 indicates that 7,112,808 inhabitants live in the Metropolitan Region of Santiago. With 52 communes, the Metropolitan Region has a majority of urban areas, in fact only 18 are rural. Composed of 32 communes, the province of Santiago concentrates 78% of the regional population.

3.2. Nearby Venues and its categories.

Using the API Foursquare obtained the categories of the nearby venues. We can have a peek of the head of the resulting dataset as follows:

	name	categories	lat	lng
0	KTM	Motorcycle Shop	-33.362120	-70.507150
1	Camino A Farellones Km.0	Mountain	-33.366907	-70.498399
2	La Divina Comida	Italian Restaurant	-33.359570	-70.507337
3	El Mesón de la Patagonia	Restaurant	-33.357492	-70.506567
4	Pollo Al Cognac	Restaurant	-33.360354	-70.506882

The most common venues founded are visualized as follows:

	Código	Neighborhood	Latitud	Longitud	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
0	7500000	Providencia	-33.43485	-70.61573	0.0	Pizza Place	Coffee Shop	Sandwich Place	Restaurant	Burger Joint	Chinese Restaurant	Peruvian Restaurant	Bakery	Martial Arts School
1	7550000	Las Condes	-33.40033	-70.50269	0.0	Soccer Stadium	Pharmacy	Athletics & Sports	Hockey Field	Fast Food Restaurant	College Cafeteria	Soccer Field	Stables	Gym / Fitness Center
2	7630000	Vitacura	-33.37763	-70.56219	0.0	Coffee Shop	Sushi Restaurant	Pharmacy	Park	Café	Farmers Market	Board Shop	Plaza	Supermarket
3	7690000	Lo Barnechea	-33.35153	-70.33607	0.0	Ski Area	Coffee Shop	Mountain	Rock Climbing Spot	Snack Place	Latin American Restaurant	Yoga Studio	Fast Food Restaurant	Farmers Market
4	7750000	Ñuñoa	-33.45449	-70.60383	0.0	Bakery	Chinese Restaurant	Coffee Shop	Pizza Place	Bar	Sandwich Place	Sushi Restaurant	Yoga Studio	Gymnastics Gym

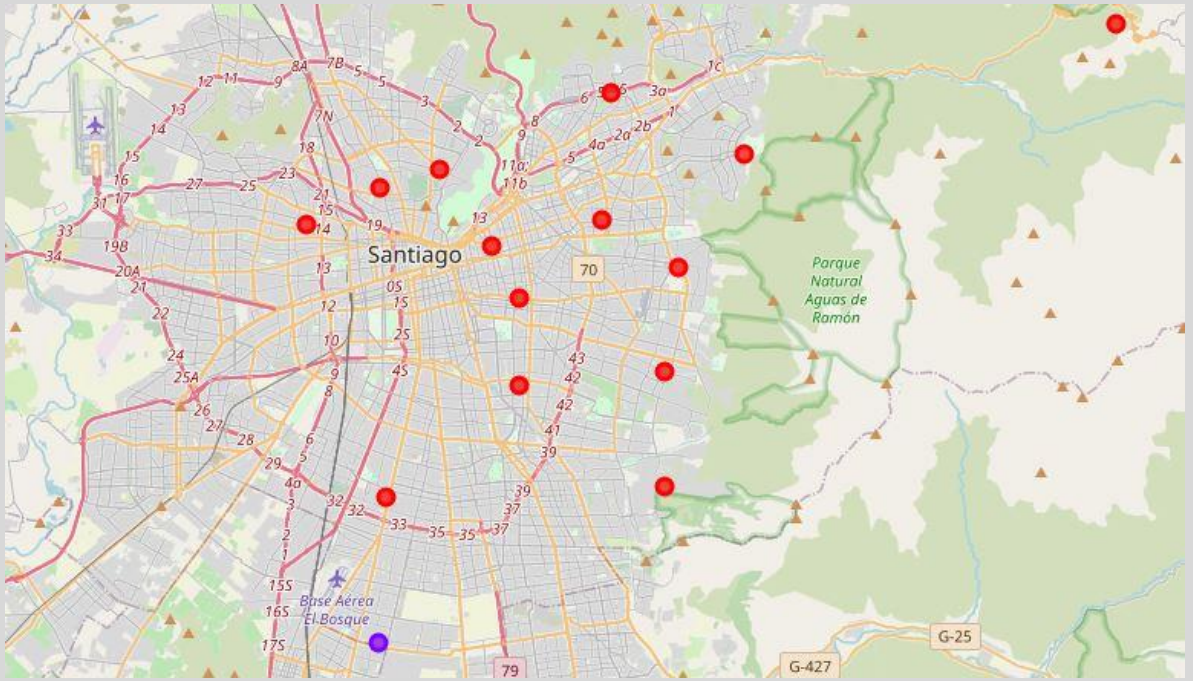
3.3. K-means Clustering Unsupervised Approach.

- There are various types of clustering algorithms such as partitioning, hierarchical, or density based clustering. K-means is a type of partitioning clustering. That is, it divides the data into k non-overlapping subsets or clusters without any cluster internal structure or labels. This means, it's an unsupervised algorithm. Objects within a cluster are very similar, and objects across different clusters are very different or dissimilar.
- the objective of K-means is to form clusters in such a way that similar samples go into a cluster and dissimilar samples fall into different clusters, it can be shown that instead of a similarity metric, we can use dissimilarity metrics. In other words, conventionally, the distance of samples from each other is used to shape the clusters. So, we can say, K-means tries to minimize the intra-cluster distances and maximize the inter-cluster distances.
- Thus utilizing this algorithm the cluster labels for venues of each Commune were found. As depicted in the following fiugre.

	Código	Neighborhood	Latitud	Longitud	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
0	7500000	Providencia	-33.43485	-70.61573	0.0	Pizza Place	Coffee Shop	Sandwich Place	Restaurant	Burger Joint	Chinese Restaurant	Peruvian Restaurant	Bakery	Martial Arts School
1	7550000	Las Condes	-33.40033	-70.50269	0.0	Soccer Stadium	Pharmacy	Athletics & Sports	Hockey Field	Fast Food Restaurant	College Cafeteria	Soccer Field	Stables	Gym / Fitness Center
2	7630000	Vitacura	-33.37763	-70.56219	0.0	Coffee Shop	Sushi Restaurant	Pharmacy	Park	Café	Farmers Market	Board Shop	Plaza	Supermarket
3	7690000	Lo Barnechea	-33.35153	-70.33607	0.0	Ski Area	Coffee Shop	Mountain	Rock Climbing Spot	Snack Place	Latin American Restaurant	Yoga Studio	Fast Food Restaurant	Farmers Market
4	7750000	Ñuñoa	-33.45449	-70.60383	0.0	Bakery	Chinese Restaurant	Coffee Shop	Pizza Place	Bar	Sandwich Place	Sushi Restaurant	Yoga Studio	Gymnastics Gym

4. Results

- Here there is a graphical representation of the results founded by the algorithm. Only two clusters were found, as you can see.
- Is clear to appreciate that the results are not quite conving as expected. The initial algorithm was set to find 4 clusters but only two were found. Moreover, there is one cluster that agglomerates al the venues in one big section that covers almost the entire City, and there is one lonely cluster with just one location in the south. From that it can be concluded that due to the similarities in the venues founded by the API Foursquare in Santiago and probably due to people not activily giving their input about the venues available in the app, there were not richness of data; so the algorithm clustered all the venues in one big cluster, not particularly what we were looking for.



5. Discussion.

- As seen, the results we not as expected, only 2 about of 4 clusters were found and one cluster in particular agglomerates the whole batch of the data, probably due to people not actively giving their input about the venues available in the app, there were not richness of data; so the algorithm clustered all the venues in one big cluster, not particularly what we were looking for.
- In the other hand, this was just only one attempt. The possibility to find more clusters still remains possible. the procedure used to cluster is still open to discussion without taking into account the code itself. Other algorithms could also be used and this is an approach that I am more interested in evaluating for future endeavors.

6. Conclusion.

- In despite the results were not as convincing as we hoped in the beginning, this project, and the whole set of courses opened the doors for me to the world of machine learning and more prominently, to data science itself.

Title Lorem Ipsum

LOREM IPSUM DOLOR SIT AMET, CONSECTETUER ADIPISCING ELIT.

NUNC VIVERRA IMPERDIET ENIM. FUSCE EST. VIVAMUS A TELLUS.

PELLENTESQUE HABITANT MORBI TRISTIQUE SENECTUS ET NETUS.