The Battle of Neighborhoods in Santiago, Chile

Patricia Avaroma

November 26, 2020

1 Introduction

1.1 Background

When you must move to a new city for many people it is an amazing stage of life but stressful as well. Should you look for something near to your new job, school, or university? or maybe look for quality life, or price, or a mix of everything?

Therefore, it is beneficial to have segmented the city considering a series of features and let future residents to take their decision of leaving in an easier way, limiting some scopes regarding the features.

1.2 Problem - Objective

This project aims to cluster the city of Santiago considering the top 10 neighborhoods with the best Urban Quality life's Index to show some options to future residents, and I will explain what distinguishes each cluster.

1.3 Stakeholders

Families, workers, or students would be interested in this segmentation to better take the decision of moving related to their interests.

2 Data source and usage

To resolve the problem, we need information about:

- Types of venues around the neighborhoods of the city of Santiago, like restaurants, health venues, parks, schools, universities, and so on.
- The most recent ranking with the Urban Quality Life's Index of the city of Santiago.
- The home purchase and rental prices to complement the features to be analyzed (in UF, "Unidad de Fomento", and indexed unit of account).

The source for venues is Foursquare location data, and for the Index and prices I decided to scrap the data from public Chilean newspapers with official information.

3 Data gathering and wrangling

Before we start to gather all the information, it is important to mention the political division of Chile for a better understanding of the scope in this project.

For the government and the internal state management, Chile is divided into 16 regions, which at the same time are divided in 56 provinces. For local management, the provinces are divided into 346 Comunas. Given that in Santiago the people take home living decision from Comunas, this will be the definition of Neighborhoods for this project.

By the effect of this project we will focus to take a look to the province of Santiago in the "Región Metropolitana (Metropolitan Region)", that is the capital of Chile. Santiago has 32 Neighborhoods.

Table 1: Neighborhoods

	Province	Capital	Neighborhood
0	Santiago	Santiago	Cerrillos
1	Santiago	Santiago	Cerro Navia
2	Santiago	Santiago	Conchalí
3	Santiago	Santiago	El Bosque
4	Santiago	Santiago	Estación Central
5	Santiago	Santiago	Huechuraba
6	Santiago	Santiago	Independencia
7	Santiago	Santiago	La Cisterna
8	Santiago	Santiago	La Granja
9	Santiago	Santiago	La Florida
10	Santiago	Santiago	La Pintana
11	Santiago	Santiago	La Reina
12	Santiago	Santiago	Las Condes
13	Santiago	Santiago	Lo Barnechea
14	Santiago	Santiago	Lo Espejo
15	Santiago	Santiago	Lo Prado
16	Santiago	Santiago	Macul
17	Santiago	Santiago	Maipú
18	Santiago	Santiago	Ñuñoa
19	Santiago	Santiago	Pedro Aguirre Cerda
20	Santiago	Santiago	Peñalolén
21	Santiago	Santiago	Providencia
22	Santiago	Santiago	Pudahuel
23	Santiago	Santiago	Quilicura
24	Santiago	Santiago	Quinta Normal
25	Santiago	Santiago	Recoleta
26	Santiago	Santiago	Renca
27	Santiago	Santiago	San Miguel
28	Santiago	Santiago	San Joaquín
29	Santiago	Santiago	San Ramón
30	Santiago	Santiago	Santiago
31	Santiago	Santiago	Vitacura

Once the Neighborhoods are identified, as these are too many, we will focus the analysis in the main 10 Neighborhoods that make up the Top 10 of the Urban Quality Life's Index (ICVU- Indice de Calidad de Vida Urbana, in Spanish).

The ICVU is an index that measure and compare in relative terms the urban quality life of Neighborhoods and cities in Chile, from a set of variables referred to six dimensions that express the state of situation in the provision of services and goods to the population, both public and private, and their socio-territorial impact, according to the “Cámara Chilena de la Construcción”. The six variables are: Living places, sociocultural conditions, business environment, labor conditions, health and environment, connectivity and mobility.

Table 2: Top 10 Neighborhoods

	Pos.	Neighborhood	Region	ICVU Score
0	1	Vitacura	Metropolitana	75.59
1	2	Las Condes	Metropolitana	74.54
2	3	La Reina	Metropolitana	70.45
3	4	Lo Barnechea	Metropolitana	69.09
4	5	Providencia	Metropolitana	68.79
5	6	Ñuñoa	Metropolitana	66.28
6	8	Macul	Metropolitana	61.92
7	16	Santiago	Metropolitana	59.27
8	17	San Miguel	Metropolitana	58.53
9	22	La Florida	Metropolitana	57.23

Then, we got the information regarding sales prices by m2 of the apartments from the Neighborhoods selected to be incorporated to the analysis.

This information is expressed in UF (“Unidad de Fomento”, and indexed unit of account).

Table 3: Sale’s prices by m2 (UF)

	Neighborhood	m2 price
0	Buin	38.3
1	Cerrillos	45.5
2	Colina	56.9
3	Conchalí	49.2
4	Estación Central	47.2
5	Huechuraba	53.5
6	Independencia	52.1
7	La Cisterna	47.7
8	La Florida	52.3
9	Lampa	28.7
10	La Reina	76.0
11	Las Condes	97.6
12	Lo Barnechea	91.3
13	Macul	54.8
14	Maipú	49.7
15	Peñalolén	60.1
16	Providencia	92.5
17	Pudahuel	44.5
18	Puente Alto	43.8
19	Quilicura	43.4
20	Quinta Normal	51.0
21	Recoleta	49.0
22	Renca	44.5
23	San Bernardo	42.4
24	San Joaquín	53.5
25	San Miguel	52.9
26	Santiago	63.4
27	Vitacura	103.1
28	Ñuñoa	75.3

Then, we merged the 3 tables to have a unique data frame by the moment.

Table 4: ICVU score and Prices by Neighborhoods

	Pos.	Region	Province	Capital	Neighborhood	ICVU Score	m2 price
0	1	Metropolitana	Santiago	Santiago	Vitacura	75.59	103.1
1	2	Metropolitana	Santiago	Santiago	Las Condes	74.54	97.6
2	3	Metropolitana	Santiago	Santiago	La Reina	70.45	76.0
3	4	Metropolitana	Santiago	Santiago	Lo Barnechea	69.09	91.3
4	5	Metropolitana	Santiago	Santiago	Providencia	68.79	92.5
5	6	Metropolitana	Santiago	Santiago	Ñuñoa	66.28	75.3
6	8	Metropolitana	Santiago	Santiago	Macul	61.92	54.8
7	16	Metropolitana	Santiago	Santiago	Santiago	59.27	63.4
8	17	Metropolitana	Santiago	Santiago	San Miguel	58.53	52.9
9	22	Metropolitana	Santiago	Santiago	La Florida	57.23	52.3

Before finishing, it only was missing the coordinates from every Neighborhood.

Table 5: Coordinates by Neighborhoods

	Neighborhood	latitude	longitude
0	Vitacura	-33.380206	-70.565795
1	Las Condes	-33.424788	-70.517498
2	La Reina	-33.447373	-70.533412
3	Providencia	-33.428838	-70.611337
4	Ñuñoa	-33.454330	-70.600582
5	Macul	-33.491943	-70.599732
6	Santiago	-33.437776	-70.650450
7	San Miguel	-33.497552	-70.652158
8	La Florida	-33.530817	-70.544076
9	Lo Barnechea	-33.362323	-70.513423

Finally, we merged our first data frame with this information and created our principal data frame, the one that will be used to explore the information in Foursquare and to visualize some data.

Table 6: Coordinates, ICVU and Prices by Neighborhoods

	Pos.	Region	Province	Capital	Neighborhood	latitude	longitude	ICVU Score	m2 price
0	1	Metropolitana	Santiago	Santiago	Vitacura	-33.380206	-70.565795	75.59	103.1
1	2	Metropolitana	Santiago	Santiago	Las Condes	-33.424788	-70.517498	74.54	97.6
2	3	Metropolitana	Santiago	Santiago	La Reina	-33.447373	-70.533412	70.45	76.0
3	4	Metropolitana	Santiago	Santiago	Lo Barnechea	-33.362323	-70.513423	69.09	91.3
4	5	Metropolitana	Santiago	Santiago	Providencia	-33.428838	-70.611337	68.79	92.5
5	6	Metropolitana	Santiago	Santiago	Ñuñoa	-33.454330	-70.600582	66.28	75.3
6	8	Metropolitana	Santiago	Santiago	Macul	-33.491943	-70.599732	61.92	54.8
7	16	Metropolitana	Santiago	Santiago	Santiago	-33.437776	-70.650450	59.27	63.4
8	17	Metropolitana	Santiago	Santiago	San Miguel	-33.497552	-70.652158	58.53	52.9
9	22	Metropolitana	Santiago	Santiago	La Florida	-33.530817	-70.544076	57.23	52.3

There is missing the last feature to be analyzed, the venues categories of our group of Neighborhoods. To collect it I will use Foursquare API. To get the information I decided to limit the radius of search to 2,000 meters and 500 values as results by Neighborhoods. The result was that exist 791 venues around the ten (10) Neighborhoods.

4 Methodology and exploratory data analysis

4.1 Methodology

This Project is focused on the exploration of the Top 10 Neighborhoods of Santiago's Province according to the ICVU score.

The first step was to inform how is Chile made administratively and based on it indicate that given the case any foreign person is planning to move to Santiago, Chile, it will be clustered only the Top 10 Neighborhoods based on the ICVU score.

The second step was the gathering of the information. To show all the Neighborhoods in Santiago, gather information about ICVU score, retrieve only the Top 10 Neighborhoods with the best scores in Santiago, then we gathered the coordinates and merged all the information. With the coordinates we could explore in Foursquare to get all the Venues' Categories of Santiago Neighborhoods.

Once we explored all the venues, we obtained a data frame with the following information to make the model:

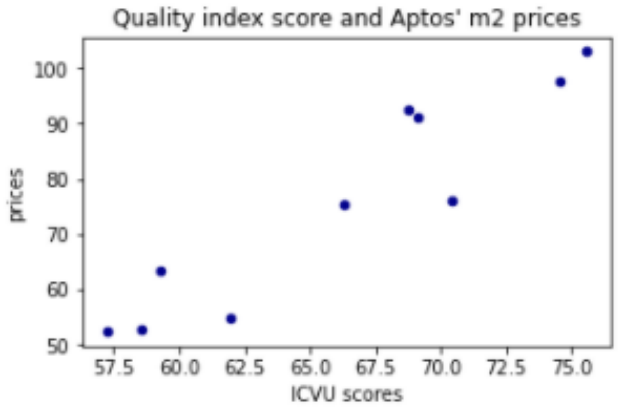
- Neighborhood = Or city (Comuna).
- Neighborhood Latitude = Neighborhood Latitude.
- Neighborhood Longitude = Neighborhood Longitude.
- Venue = Name of the venue.
- Venue latitude = Venue latitude.
- Venue Longitude = Venue Longitude.
- Venue Category = Type of venue.
- Pos. = Position in the ICVU Ranking.
- Region = Maximum administrative unit. Limited for this Project to "Región Metropolitana"
- Province = Intermediate administrative unit.
- Capital = Province's capital.
- ICVU Score = Score of the Urban Quality Life's Index.
- latitude = Coordinate.
- longitude = Coordinate.

The third and last step is to create the clusters with some of the features to be selected.

4.2 Exploratory data analysis

Let us visualize if there is any relation between the ICVU score and the m2 prices.

Graph 1: Quality life score and Aptos' m2 prices.



We can see the relation between ICVU score and apartments' prices. It looks like the biggest the score is, the price will be higher.

Apart from these features, we will explore the venues around.

Table 7: Count of venues in every Neighborhood

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Pos.	Region	Province	Capital	latitude	longitude	ICVU Score	m2 price
La Florida	24	24	24	24	24	24	24	24	24	24	24	24	24	24
La Reina	60	60	60	60	60	60	60	60	60	60	60	60	60	60
Las Condes	28	28	28	28	28	28	28	28	28	28	28	28	28	28
Lo Barnechea	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Macul	86	86	86	86	86	86	86	86	86	86	86	86	86	86
Providencia	100	100	100	100	100	100	100	100	100	100	100	100	100	100
San Miguel	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Santiago	99	99	99	99	99	99	99	99	99	99	99	99	99	99
Vitacura	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Ñuñoa	100	100	100	100	100	100	100	100	100	100	100	100	100	100

As we can see, Lo Barnechea, Providencia, Vitacura and Ñuñoa have the highest number of venues around. Las Condes and La Florida have the smaller number of venues around.

Exploring the venues, we can see that it exists around 185 categories of venues.

Table 8: 5 most common venues by Neighborhoods

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	La Florida	Park	Gym	Restaurant	Big Box Store	Coffee Shop
1	La Reina	Restaurant	Pizza Place	Sushi Restaurant	Chinese Restaurant	Playground
2	Las Condes	Soccer Field	Plaza	Park	Racetrack	Pharmacy
3	Lo Barnechea	Gym	Restaurant	Coffee Shop	Sushi Restaurant	Pharmacy
4	Macul	Chinese Restaurant	Restaurant	Soccer Field	Hot Dog Joint	Sushi Restaurant
5	Providencia	Coffee Shop	Park	Hotel	French Restaurant	Ice Cream Shop
6	San Miguel	Sushi Restaurant	Pizza Place	Restaurant	Plaza	Park
7	Santiago	Coffee Shop	Café	Pizza Place	Hotel	Hostel
8	Vitacura	Café	Gym / Fitness Center	Restaurant	Bakery	Italian Restaurant
9	Ñuñoa	Coffee Shop	Bakery	Plaza	Restaurant	Pizza Place

The category most repeated in 1st place is Coffee (café) shop and then restaurants. Then we can see a mix of parks and gyms.

5 Clustering model

To determine the clusters of Neighborhoods in Santiago it will be applied the K-means model, it is more efficient than the Hierarchical and allows to define the number of optimal clusters.

This algorithm works iteratively to assign every explored object to a cluster (k) based on their features. Objects are grouped based on the similarity of their features.

Entries must be numeric values. Given that the values of the features to be analyzed do not have an order of importance, I will use one-hot encoding over the “Venue Category” feature.

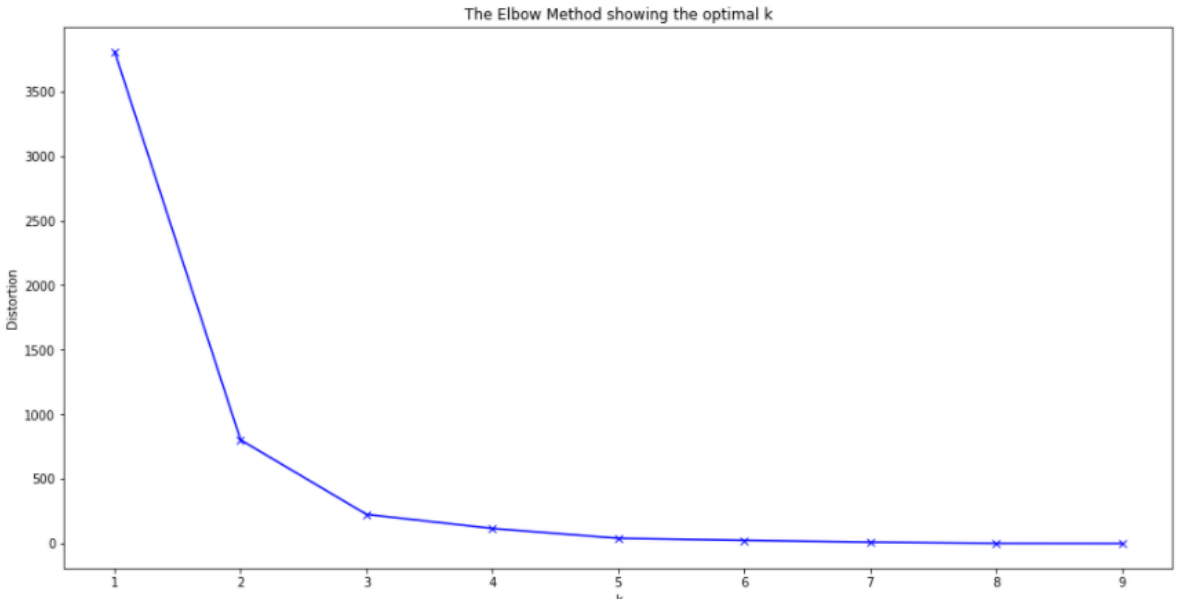
To run the algorithm, we need the features which will work into the model and a k value. The model iterates around two steps: (1) to every row of the data set will be assigned a centroid, the nearest and randomly, based on the Euclidean distance. (2) The centroids will be updated, based on the measure of every point assigned in the last step. The distances will be updated, the rows of data are reassigned again, and the centroids recalculated until the error is minimized, it means the total distance (Distortions).

So, we need to define the features and normalize their values with standard deviation. Features will be:

- ICVU score
- M2 prices
- Venues categories

Then, we need to find the best k (number of clusters) for the model. For that, we used the Elbow model.

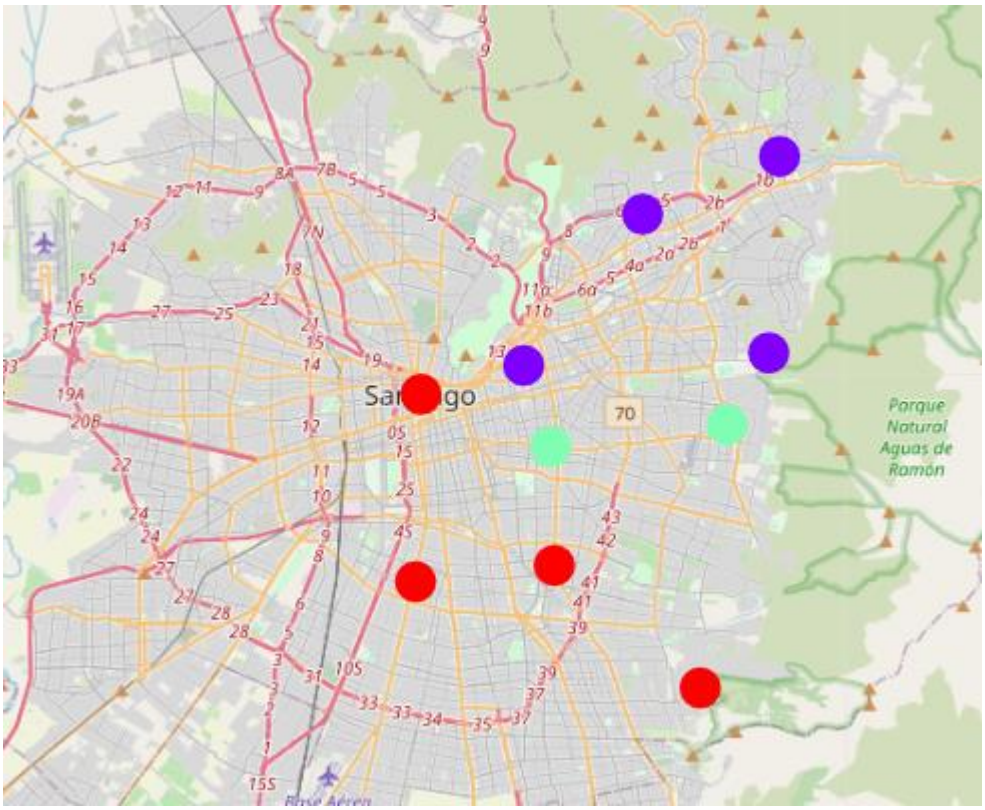
Graph 2: Best k



Best k is 3.

Now we could run the algorithm. The result is:

Graph 3: Map of Santiago and clustering of top 10 Neighborhoods



We can see clearly the three clusters:

Table 9: Cluster 1

	Neighborhood	ICVU Score	m2 price	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
6	Macul	61.92	54.8	0	Chinese Restaurant	Restaurant	Soccer Field	Hot Dog Joint	Sushi Restaurant
7	Santiago	59.27	63.4	0	Coffee Shop	Café	Pizza Place	Hotel	Hostel
8	San Miguel	58.53	52.9	0	Sushi Restaurant	Pizza Place	Restaurant	Plaza	Park
9	La Florida	57.23	52.3	0	Park	Gym	Restaurant	Big Box Store	Coffee Shop

This is characterized by having the smaller m2 prices and ICVU scores. It also has many restaurants, coffee shop and hotels as venues in common.

Table 10: Cluster 2

	Neighborhood	ICVU Score	m2 price	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Vitacura	75.59	103.1	1	Café	Gym / Fitness Center	Restaurant	Bakery	Italian Restaurant
1	Las Condes	74.54	97.6	1	Soccer Field	Plaza	Park	Racetrack	Pharmacy
3	Lo Barnechea	69.09	91.3	1	Gym	Restaurant	Coffee Shop	Sushi Restaurant	Pharmacy
4	Providencia	68.79	92.5	1	Coffee Shop	Park	Hotel	French Restaurant	Ice Cream Shop

This is characterized by having the highest m2 prices and ICVU scores. It also has many coffee shop, gym, and parks as venues in common. It seams to be mor green.

Table 11: Cluster 3

	Neighborhood	ICVU Score	m2 price	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	La Reina	70.45	76.0	2	Restaurant	Pizza Place	Sushi Restaurant	Chinese Restaurant	Playground
5	Ñuñoa	66.28	75.3	2	Coffee Shop	Bakery	Plaza	Restaurant	Pizza Place

This is characterized by having intermediate m2 prices and ICVU scores. It also has many restaurants as venues in common.

The algorithm cluster the neighborhoods mainly about the ICVU score and m2 prices. Restaurant is the type of venue that repeat on the three clusters.

6 Results and discussion

The model indicates that there are three cluster of neighborhoods regarding three kind of features: "Venue category", "ICVU score" and "Square meter's price (m2 price)".

Exploring the data, we could see that those Neighborhoods which are more expensive, have more venues around (100, based on the Radius and limit specified). We could see as well that the higher the score, the higher the price is.

The cluster are as follows:

- Cluster 1 (0): This groups the Neighborhoods with the lowest ICVU scores in the group defined and the lowest price by m2. This cluster has neighborhoods with Restaurants, Coffees and hotels as most in common places.
- Cluster 2 (1): This group the Neighborhoods with the highest ICVU score and the highest prices by m2. You can find Coffee shops, Parks and Gym as the most common places.
- Cluster 3 (2): This group the Neighborhoods with the intermediate ICVU score and prices by m2. You can find here Restaurants.

As we can see, the three clusters have in common the Restaurants, and its segmented mainly by the ICVU score and properties prices by m2. The ICVU include an evaluation about services and goods provision, both public and private. It gives a score regarding living places, sociocultural conditions, business environment, labor conditions, health and environment, connectivity, and mobility.

This result does not imply those clusters are perfect or that will not change in the future with more information that could be considered.

7 Conclusion

The purpose of this project was to identify some clusters based on the top 10 Neighborhoods from Santiago city based on the ICVU score, to give some information about prices, quality life and venues around to foreign people thinking about moving to Santiago, Chile. By gathering information about the Neighborhoods, their ICVU score, apartments' prices and exploring their venues' category around from Foursquare, then I generated a complete data frame to be used for the model. Clustering of these Neighborhoods was performed to identify profiles of the Neighborhoods on Santiago.

Families, workers, or students could take decision based on the characteristics identified in every cluster.