The Battle of The Neighborhood Santiago vs Buenos Aires

Lesly Vilches

1.Introduction

1.1 Background

Santiago and Buenos Aires are two cities in South America, located in opposite areas of the region, the first in the western part and the second in the eastern part of the continent.Both cities are very diverse and are the capitals of their respective countries (Chile and Argentina), they have some things in common, multiple panoramas, versatile and cosmopolitan, full of contrasts, combining the traditional history and modernity,and importants tourist destination in Latin America.

Given all of the above, interest arises in discovering how similar they are.

This Project seeks to segment each city and to explore what services are available in each of these divisions, to know which are the most common category in each of them, using the Foursquare API, and to then classify them and perform a clustering analysis to determine how similar or different they are.It will use the k-means clustering algorithm to complete this task.

1.2 Business Problem

This can be useful for stakeholders in moving from one city to another, and who want to maintain a similar lifestyle when changing places of residence.

Some questions to answer are:

-In each city, the amount of venues available and categories are similar?

-What are the 10 most common categories in each city and in each division?

 -What characteristics does each cluster have?

- Are there neighborhoods with similar characteristics between both cities?

2.Data

2.1 Data Acquisition

The data necessary to carry out the analysis of the problem are:

-**Administrative division of each city and its geographical coordinates. For** Santiago de Chile, the data was extracted from: https://es.wikipedia.org/wiki/Anexo:Comunas_de_Chile. While the information from Buenos Aires was extracted from: https://es.wikipedia.org/wiki/Anexo:Barrios_de_la_ciudad_de_Buenos_Aires

- **The most common venues in each area**. These will be obtained using the Forsquare API

- **Value of the square meter of the housing**. These were obtained from:

https://www.latercera.com/pulso/noticia/precio-del-m2-departamentos-supera-los-2-millones-siete-comunas-santiago/654495/, for Santiago and from:https://blog.properati.com.ar/reporte-del-mercado-inmobiliario-properati-oikos-bs-as-mayo-2019/, for Buenos Aires.

- **Files json with the divisions of each city.** For Santiago file was downloaded from: https://github.com/jlhonora/geo and for Buenos Aires from: https://data.buenosaires.gob.ar/dataset/barrios/archivo/1c3d185b-fdc9-474b-b41b-9bd960a3806e.

2.2 Data Preparation

The first thing is to get the data, scrapping every website, already mentioned and transform the data into a pandas data framework.

Before carrying out any analysis, a data wrangling is carried out, to obtain clean data frames and only with the necessary information.

First, a change of column names is made for clarity and to standardize the names of each table. Then a data cleanup is performed, where unnecessary data is removed. Then the decimal geographic coordinate data for each commune or neighborhood is loaded or append to its respective data frame. Resulting dataframes: stgo, for the city of Santiago (Figure 1) and data_ar, for the city of Buenos Aires (Figure 2). Both data frames contain the neighborhoods (communes in the case of Santiago) of each city and their geographic coordinates.

	Communes	Province	Superficie(km2)	Población2017	Densidad(hab./km2)	IDH 2005	latitud dms	latitude	longitude	UF/m2	USD/m2
0	Santiago	Santiago	23.2	404495	17435.1	0.807	-33°26'14"	-33.437222	-70.657222	64.4	2261.728
1	Cerrillos	Santiago	21	80832	3849.1	0.743	-33°30'0"	-33.500000	-70.716667	39.7	1394.264
2	Cerro Navia	Santiago	11	132622	12056.5	0.683	-33°25'19.2"	-33.422000	-70.735000	NaN	NaN
3	Conchalí	Santiago	10.7	126955	11865.0	0.707	-33°22'48"	-33.380000	-70.675000	51.5	1808.680
4	El Bosque	Santiago	14.2	162505	11444.0	0.711	-33°34'1.2"	-33.567000	-70.675000	NaN	NaN
5	Estación Central	Santiago	15	147041	9802.7	0.735	-33°27'32.4"	-33.459000	-70.699000	52.8	1854.336
6	Huechuraba	Santiago	44.8	98671	2202.5	0.737	-33°22'4.8"	-33.368000	-70.634000	54.6	1917.552
7	Independencia	Santiago	7	100281	14325.9	0.709	-33°24'46.8"	-33.413000	-70.666000	57.9	2033.448
8	La Cisterna	Santiago	10	90119	9011.9	0.775	-33°31'44.4"	-33.529522	-70.664253	49.8	1748.976
9	La Florida	Santiago	70.2	366916	5226.7	0.804	-33°31'30"	-33.533300	-70.583300	52.9	1857.848
10	La Granja	Santiago	10	116571	11657.1	0.689	-33°31'60"	-33.533333	-70.625000	36.7	1288.904
11	La Pintana	Santiago	30.6	177335	5795.3	0.679	-33°34'58.8"	-33.591389	-70.636423	NaN	NaN

Figure 1: stgo data frame

	Nombre del barrio	Superficie[1]	Habitantes (año 2007)	Densidad de habitantes (año 2007)	Comuna de pertenencia de la Ciudad de Buenos Aires	latitud	longitud
0	Agronomía	2,1 km²	34.580	8.645	Comuna 15	-34.591516	-58.485385
1	Almagro	4,1 km²	139.262	33.960	Comuna 5	-34.609988	-58.422233
2	Balvanera	4,4 km²	152.198	34.950	Comuna 3	-34.609215	-58.403140
3	Barracas	7,6 km²	77.474	10.194	Comuna 4	-34.645285	-58.387562
4	Belgrano	6,8 km²	138.942	20.433	Comuna 13	-34.561308	-58.456545

Figure 2: data_ar data frame

The data on house prices per square meter, in the case of the city of Buenos Aires, only the information found is read and loaded, in the data frame called precio_ar (Figure 3). In the case of the city of Santiago we have cells without values, as there are few cases, these missing data were replaced with the average value of the data in the UF / m2 and USD / m2 column, respectively (Figure4)

The resulting dataframes are shown below:

	Barrio	Precio m^2(USD)	% Δ
0	PUERTO MADERO	6665.175940	0.020713
1	PALERMO	4093.694284	0.016272
2	BELGRANO	3923.967975	0.009917
3	NÚÑEZ	3902.073708	0.021539
4	RECOLETA	3525.404238	-0.011297
5	COLEGIALES	3525.301526	0.029862
6	RETIRO	3491.433211	-0.032615
7	COGHLAN	3487.932087	0.075715
8	VILLA URQUIZA	3395.696608	-0.027234
9	SAAVEDRA	3212.155568	0.019856
10	VILLA ORTÚZAR	3206.412478	0.011211
11	VILLA DEVOTO	3093.976496	0.017263
12	CHACARITA	3087.818937	0.016525
13	CABALLITO	3009.393981	0.007905

Figure 3: Precio_ar data frame

	Communes	Province	Superficie(km2)	Población2017	Densidad(hab./km2)	IDH 2005	latitud dms	latitude	longitude	UF/m2	USD/m2
0	Santiago	Santiago	23.2	404495	17435.1	0.807	-33°26'14"	-33.437222	-70.657222	64.400000	2261.728
1	Cerrillos	Santiago	21	80832	3849.1	0.743	-33°30'0"	-33.500000	-70.716667	39.700000	1394.264
2	Cerro Navia	Santiago	11	132622	12056.5	0.683	-33°25'19.2"	-33.422000	-70.735000	58.796296	2085.680
3	Conchalí	Santiago	10.7	126955	11865.0	0.707	-33°22'48"	-33.380000	-70.675000	51.500000	1808.680
4	El Bosque	Santiago	14.2	162505	11444.0	0.711	-33°34'1.2"	-33.567000	-70.675000	58.796296	2085.680
5	Estación Central	Santiago	15	147041	9802.7	0.735	-33°27'32.4"	-33.459000	-70.699000	52.800000	1854.336
6	Huechuraba	Santiago	44.8	98671	2202.5	0.737	-33°22'4.8"	-33.368000	-70.634000	54.600000	1917.552
7	Independencia	Santiago	7	100281	14325.9	0.709	-33°24'46.8"	-33.413000	-70.666000	57.900000	2033.448
8	La Cisterna	Santiago	10	90119	9011.9	0.775	-33°31'44.4"	-33.529522	-70.664253	49.800000	1748.976
9	La Florida	Santiago	70.2	366916	5226.7	0.804	-33°31'30"	-33.533300	-70.583300	52.900000	1857.848
10	La Granja	Santiago	10	116571	11657.1	0.689	-33°31'60"	-33.533333	-70.625000	36.700000	1288.904
11	La Pintana	Santiago	30.6	177335	5795.3	0.679	-33°34'58.8"	-33.591389	-70.636423	58.796296	2085.680
12	La Reina	Santiago	23	92787	4034.2	0.883	-33°26'34.8"	-33.443000	-70.532000	79.400000	2788.528

Figure 4: stgo resulting data frame.

3. Methodology

After obtaining the necessary data to carry out this project, the methodology used is detailed below.

The procedure for conducting the analysis consists of the following parts:

1. Segment each city: Santiago City was divided by communes belonging to the province of Santiago, plus the peripheral communes of Puente Alto and San Bernardo. Giving a total of 34 communes to analyze. The names and geographic coordinates of each commune are in the stgo data frame. Buenos Aires City was divided by neighborhoods, which correspond to a total of 48 divisions to analyze, data are in data_ar data frame.

To standardize the names of the divisions of each city, the term neighborhood is used.

2. Explore Every Neighborhood: The Foursquare API was used to explore the nearby locations of each neighborhood within a 1100-meter radius. The maximum number of venues is limited to 30. Extracting venues names, venues category (eg coffee shop, restaurant, etc.) and their coordinates.

After all the venues are obtained, the total number of venues per neighborhood and for each category is counted. To visualize this information, bar graphs are created with the number of venues per neighborhood, and the 10th most common categories in the city, and a table is constructed with the 10 most frequent venues categories for each neighborhood.

This step, first is carried out for each city separately, to know its individual characteristics, such as a unique number of categories by city and by neighborhood. And the 10 most common categories of the city. Then it is done with the total of neighborhoods in both cities, giving a total of 82 neighborhoods to explore.

3. Clustering Neighborhoods: After get the most common categories and locations in each neighborhood, these data will be used to group the neighborhoods into clusters. It was used the k-means clustering algorithm, with a number of clusters equal to 5, to complete this task.

Like the previous step, a clustering analyze was carried out for each city separately and then another analysis considering both cities. To obtain characteristic clusters of each city and then to finally know which are the similar neighborhoods between both cities and if there are neighborhoods with unique characteristics.

4. Mapping Cities: To visualize the neighborhoods and clusters, a map of each city is made, then another is made to visualize the clusters of each city and the clusters of both cities.

To complement the information obtained after clustering, a choropleth map is made to visualize the price, in average dollars of housing per m2 in each neighborhood, to extract more details from each city.

4. Results
4.1 Santiago de Chile

The city is divided into 34 communes, which to standardize the vocabulary will be called neighborhoods. The distribution of these in the city is shown in the following figure.
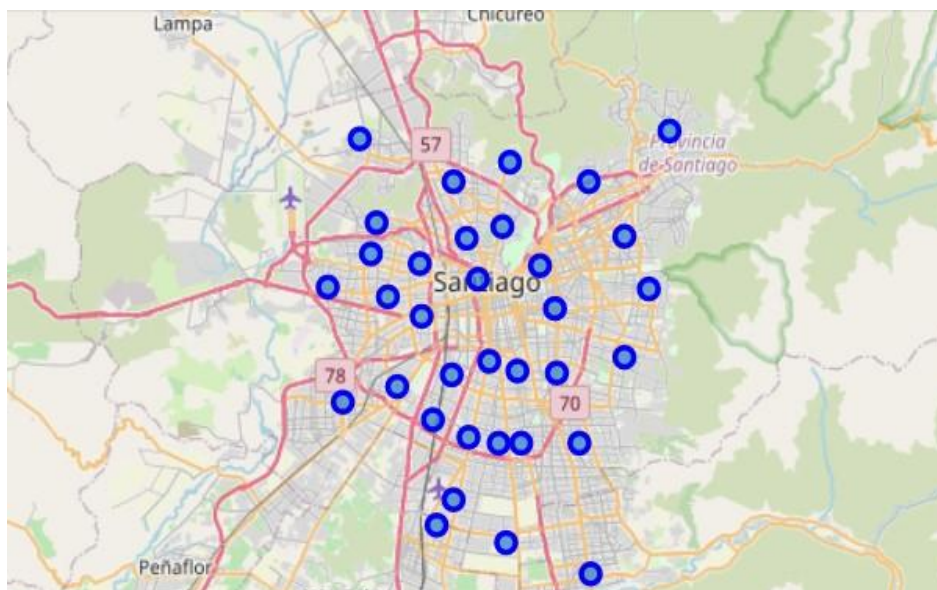

Figure 5: Santiago neighborhoods.

It was set to explore a limit of 30 venues per neighborhood. However, only 18 of the 34 neighborhoods reach this limit, there are even 8 neighborhoods that have less than 10 venues available in their surroundings. As shown in figure 6.
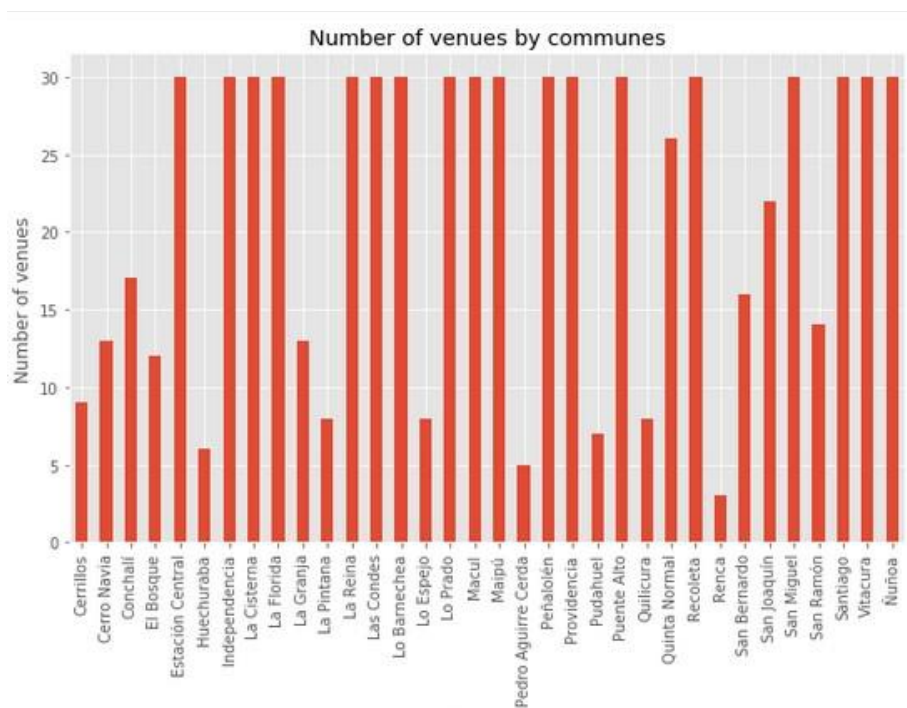

Figure 6: Number of venues per neighborhood(commune).

Despite the above result, 163 uniques venues categories are obtained. The 10th most common are summarized below:

	Venue
Venue Category	
Sandwich Place	16
Fast Food Restaurant	16
Gym	18
Pharmacy	21
Plaza	22
Pizza Place	26
Restaurant	27
Chinese Restaurant	28
Bakery	29
Sushi Restaurant	42
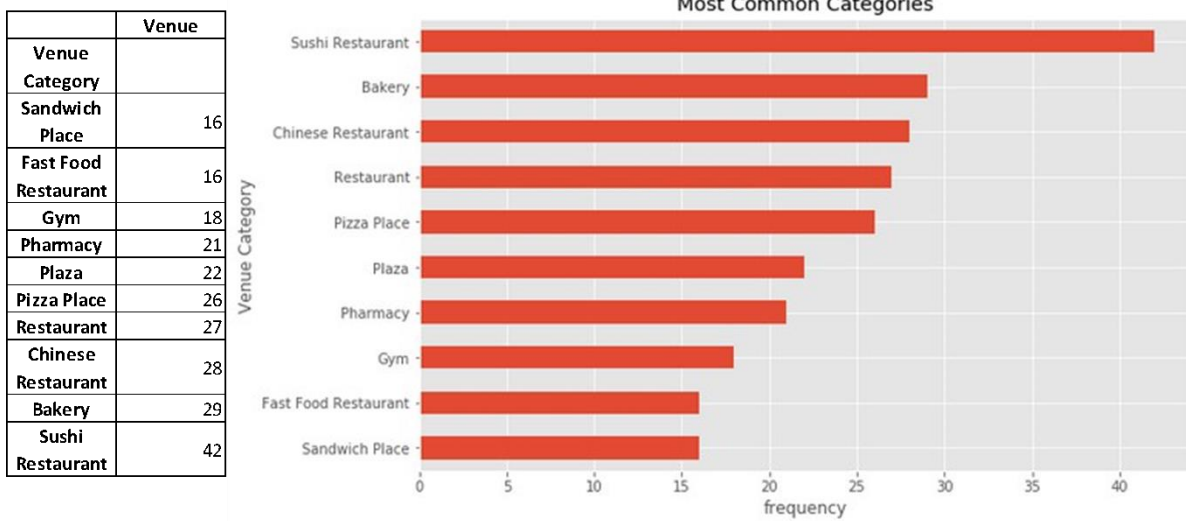
Figure 7: 10th Most Common Venue
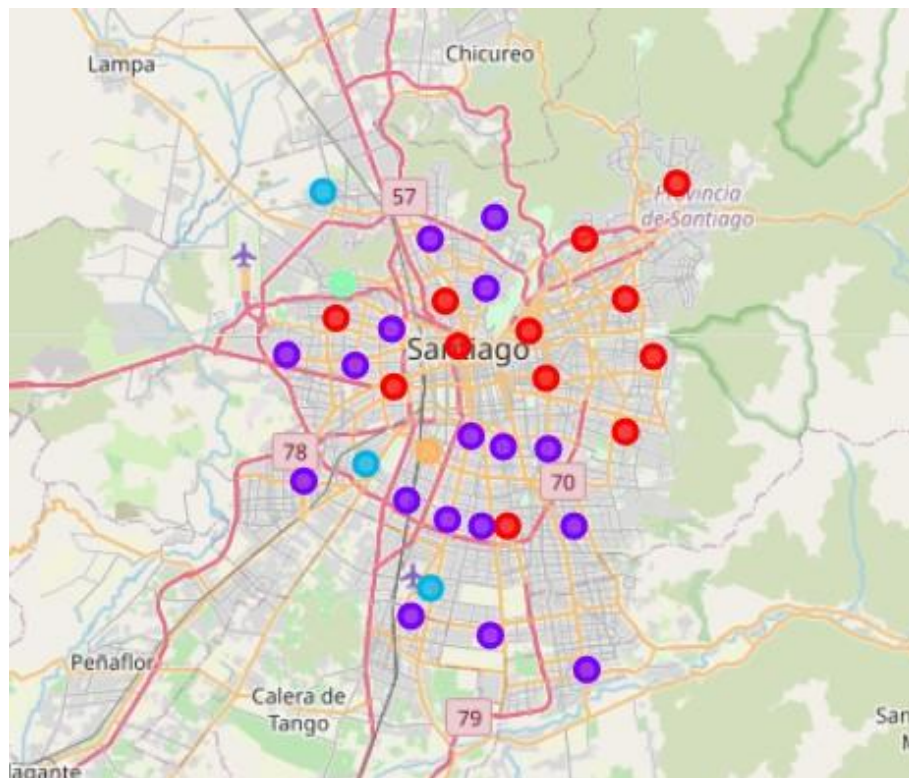
Figure 8: Mapping Cluster, Santiago de Chile

After applying the k-means clustering algorithm, with a number of clusters equal to 5, the result can be seen in figure 8, where most of the neighborhoods belong to cluster 0 or cluster 1, the other

clusters have 3 or fewer neighborhoods, which is why they can be defined as outliers, that is, unusual areas in the city.

Focusing on cluster 0 and cluster 1, it is possible to notice that the city is divided into two large blocks: cluster 0 is more frequent to the east and cluster 1 is more frequent to the west. The neighborhoods that make up each cluster are summarized below, in addition to their most common places. The details of each neighborhood can be consulted at: The Battle of Neighborhoods.ipynb

-**Cluster 0(Rojo):** Restaurant, Bakery and Coffe Shop

Neighborhoods:Santiago, Cerro Navia, Estación Central, Independencia, La Granja, La Reina, Las Condes, Lo Barnechea, Ñuñoa, Peñalolen, Providencia, Vitacura.

-**Cluster 1(morado):** Chinese Restaurant, Sushi Restaurant, Park, Farmers Market

Neighborhoods:Conchalí, Huechuraba, La Cisterna, La Florida, La Pintana, Lo Espejo, Lo Prado, Macul, Maipú, Pudahuel, Quinta Normal, Recoleta, San Joaquín, San Miguel, San Ramón, Puente Alto, San Bernardo.

-**Cluster 2(Turquesa):**Flea Market and Park, bus station -Fast food restaurant, liquor store)

Neighborhoods:Cerrillos, El Bosque, Quilicura.

-**Cluster 3 (verde):** Outdoors & Recreation and Brewery

Neighborhoods:Renca

-**Cluster 4 (naranjo):** Big Box Store, Hot dog joint

Neighborhoods:Pedro Aguirre Cerda

When making the choropleth map of the average house prices per square meter, the highest prices are located in the eastern sector of the city, exceeding 2100 dollars per m2, in addition, these neighborhoods belong to cluster 0 of the city.
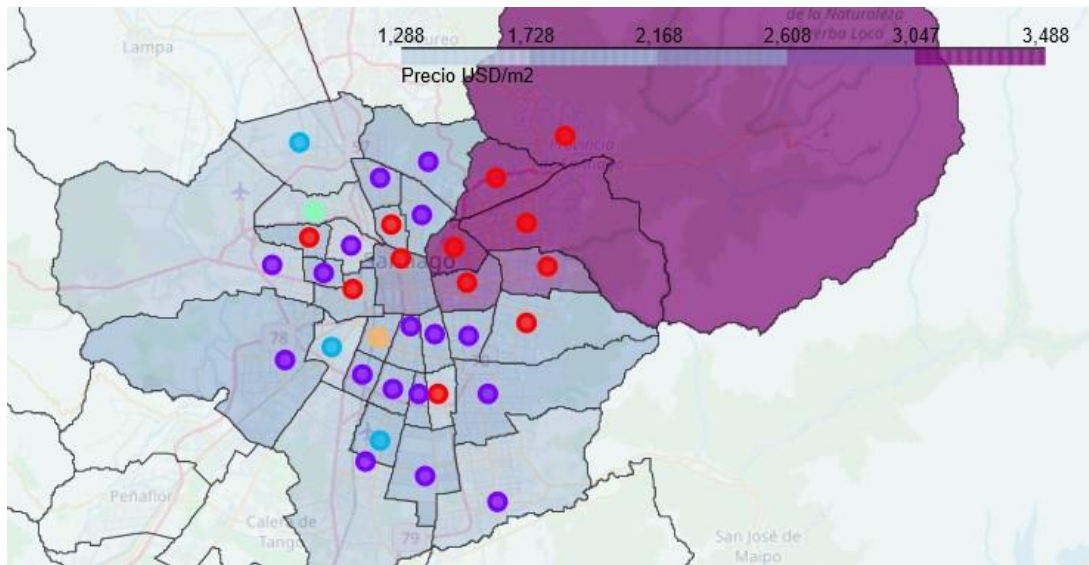
Figure 9: Choropleth map clusters and average house prices per square meter, Santiago, Chile.

4.2 Buenos Aires

The city is divided into 48 neighborhoods. The distribution of these in the city is shown in the following figure.
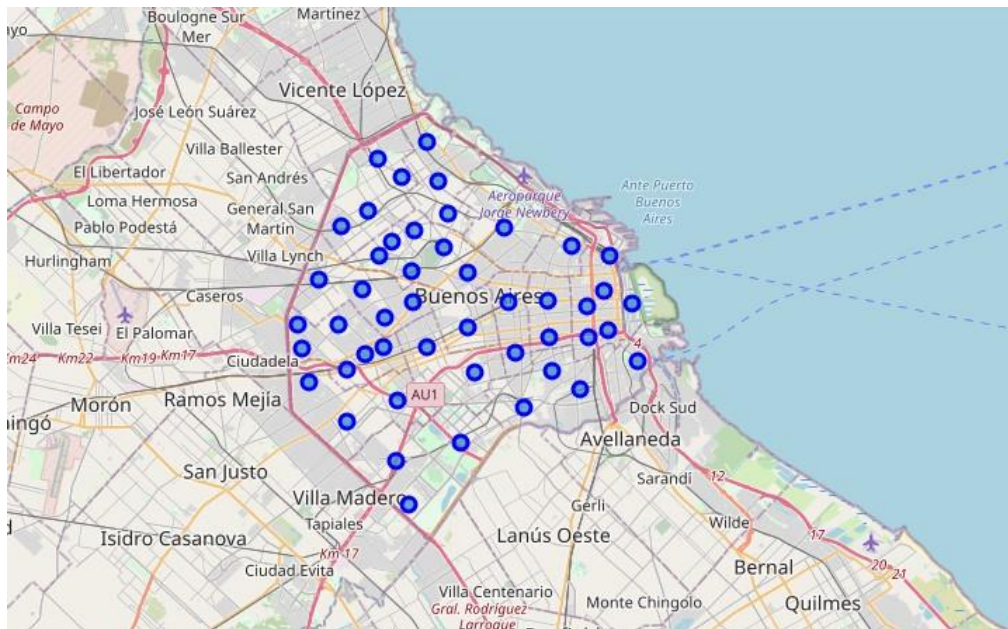

Figure 10: Neighborhoods Buenos Aires, Argentina.

It was set to explore a limit of 30 venues per neighborhood. Most neighborhoods reach this limit, only 2 neighborhoods have less than 10 venues available in their surroundings. As shown in figure 11.
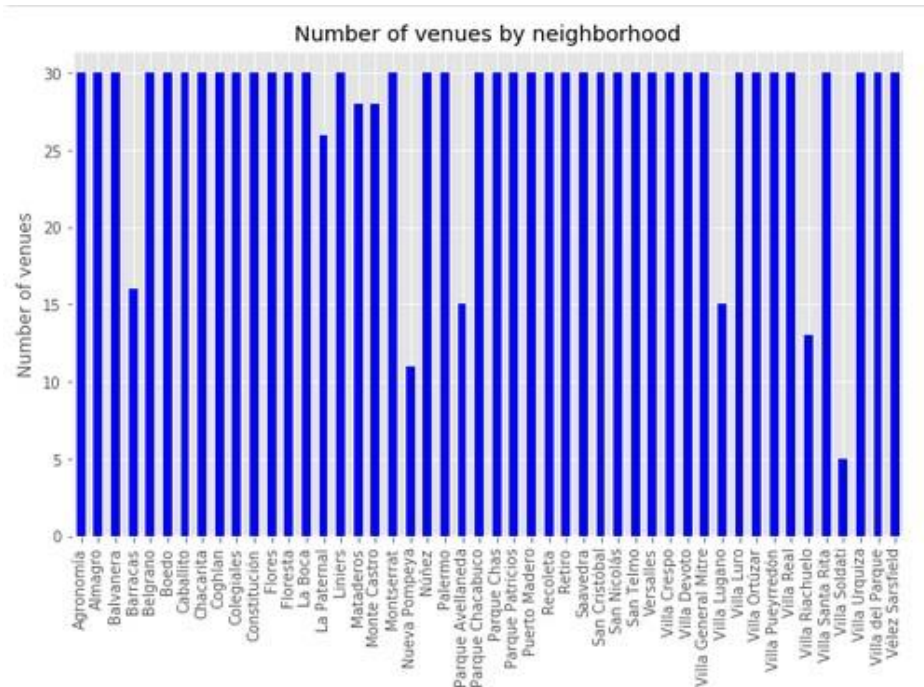
Figure 11: Number of venues per neighborhood

In the city there are 200 uniques venues categories . The 10th most common are summarized below:
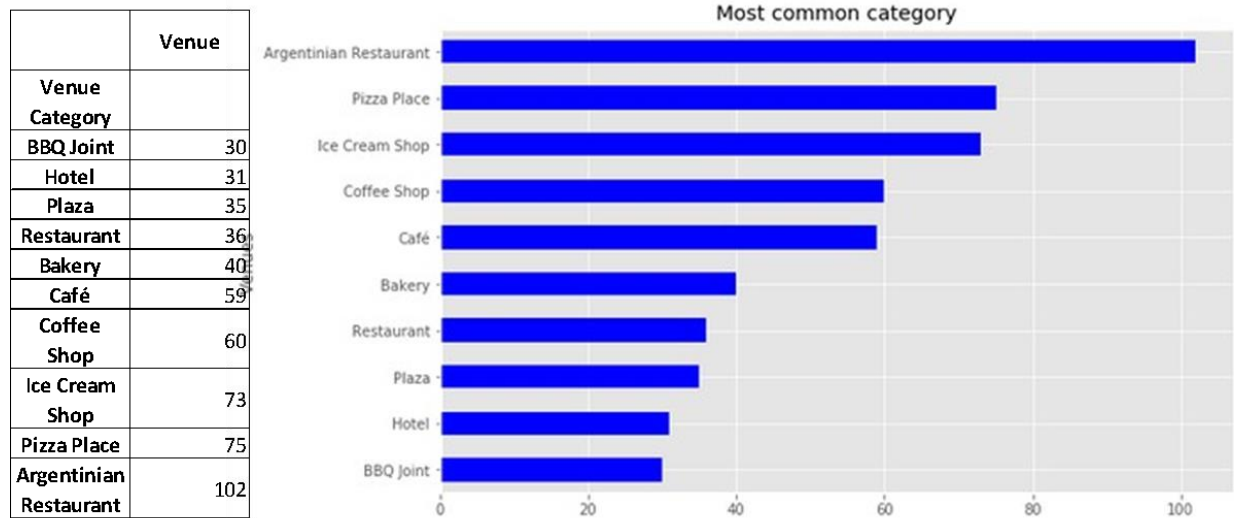
	Venue
Venue Category	
BBQ Joint	30
Hotel	31
Plaza	35
Restaurant	36
Bakery	40
Café	59
Coffee Shop	60
Ice Cream Shop	73
Pizza Place	75
Argentinian Restaurant	102

Figure 12: 10th Most Common Venue

After applying the k-means clustering algorithm, with a number of clusters equal to 5, the result can be seen in figure 13, where the most unusual neighborhood is Villa Soldati in Cluster 0. Cluster 4 is the one with the highest number of neighborhoods and It is distributed throughout almost the entire city. Cluster 2 is more frequent in the western part of the city, while cluster 3 is more frequent in the southeast and closely resembles cluster 4 and cluster 1.
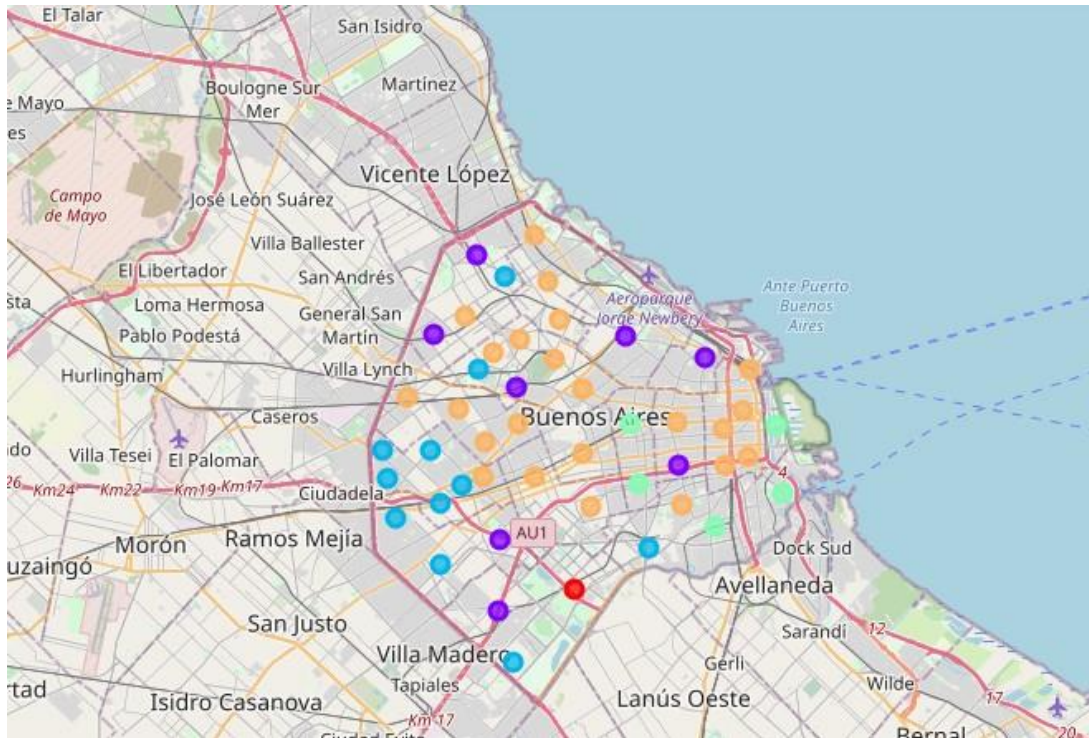
Figure 13: Mapping Cluster, Buenos Aires, Argentina

The neighborhoods that make up each cluster are summarized below, in addition to their most common places. The details of each neighborhood can be consulted at: The Battle of Neighborhoods.ipynb

-**Cluster 0(rojo):**Recreation Center, Moving Target

Neighborhoods: Villa Soldati.

-**Cluster 1(morado):**Argentinian Restaurant, Ice Cream Shop

Neighborhoods: La Paternal, Palermo, Parque Avellaneda, Recoleta, Saavedra, San Cristóbal, Villa Lugano, Villa Pueyrredón

-**Cluster 2(turquesa):**Pizza Place, Coffe Shop (plaza,restaurant)

Neighborhoods: Agronomia, Coghlan,Liniers, Mataderos,Monte Castro, Nueva Pompeya,Vélez Sarsfield, Versalles, Villa Luro, Villa Real, Villa Riachuelo.

-**Cluster 3(verde):** Argentinian Restaurant,(Bar,Cheese Shop,Ice cream Shop,Coffe Shop)

Neighborhoods: Almagro, Barracas,Boedo, La Boca, Puerto Madero

-**Cluster 4 (naranjo):** Restaurant, Coffe Shop, Ice Cream Shop (Bar, Theather, Hotel)

Neighborhoods: Balvanera, Belgrano, Caballito, Chacarita, Colegiales, Constitución, Flores, Floresta, Montserrat, Nuñez, Parque Chacabuco, Parque Chas, Parque Patricios, Retiro, San Nicolás, San

Telmo, Villa Crespo, Villa del Parque, Villa Devoto, Villa General Mitre, Villa Ortúzar, Villa Santa Rita, Vila Urquiza.

When making the choropleth map of the average house prices per square meter, the highest prices are located in Puerto Madero with 6666 dollars per m2, however, in general, a homogeneous price range per m2 is observed for the city, with the range between 2141 and 3197 dollars being common, whose neighborhoods belong to cluster 4 of the city in greater quantity.
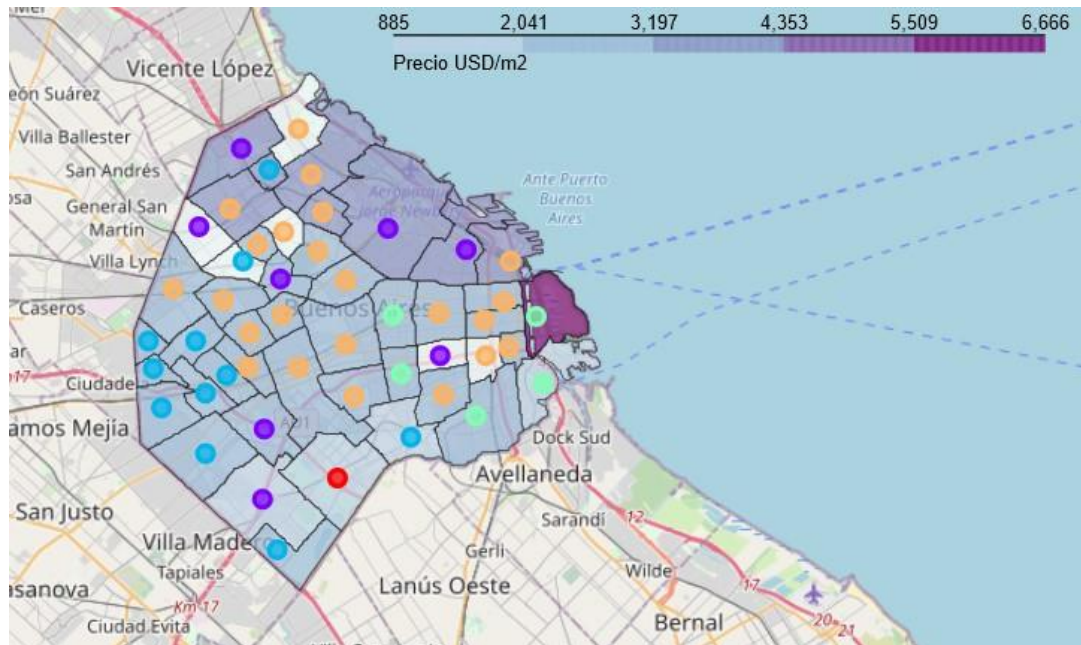

Figure 14: Choropleth map clusters and average house prices per square meter, Buenos Aire, Argentina.

4.3 Both Cities

Exploring the 82 neighborhoods belonging to each city and maintaining the limit of 30 maximum venues per neighborhood, it is obtained that there are 247 uniques categories and the 10th most common are summarized below:
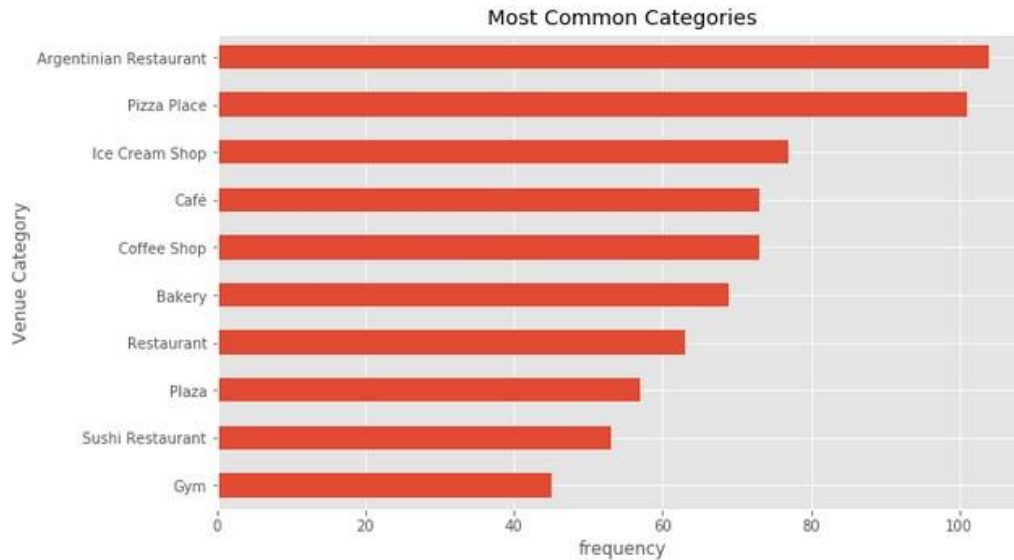
Figure 15: The 10th most common categories, for both cities.

After applying the k-means clustering algorithm, with a number of clusters equal to 5, the result can be seen in figure 16, where only clusters 3 and 0 are present in both cities. The most unusual neighborhood is Renca, the only member of Cluster 4. Cluster 2, is the same as for the analysis carried out for the city of Santiago. Both cluster 0 and 2 are only present in the city of Santiago, while cluster 1 is only present in Buenos Aires.

Cluster 0 is located in the northeast part of Santiago, while in Argentina it is distributed almost homogeneously in the city and is the most frequent along with cluster 1. Cluster 3 has more neighborhoods in the city of Santiago, and is located in the western part of the city. In Buenos Aires it is unusual and is located in the southwestern part.
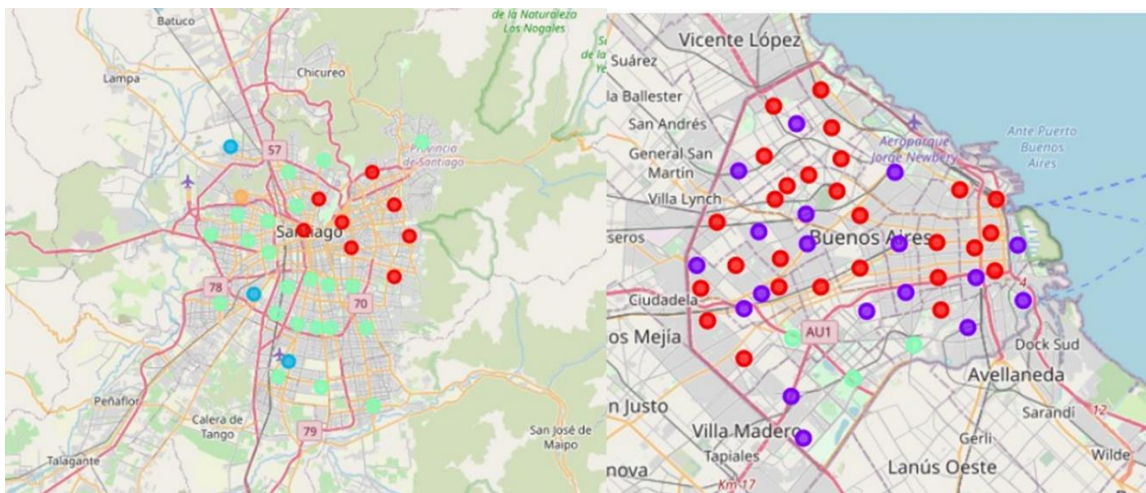

Figure 16: Mapping Cluster, Santiago de Chile and Buenos Aires, Argentina

The neighborhoods that make up each cluster are summarized below, in addition to their most common places. The details of each neighborhood can be consulted at: The Battle of Neighborhoods.ipynb

-**Cluster 0 (Red) :** Restaurant, Coffe Shop (Bakery, Pizza Place, Bar)

Neighborhoods: Santiago, La Reina, Las Condes, Ñuñoa, Peñalolén, Providencia, Recoleta, Vitacura, Agronomia, Balvaner, Belgrano, Caballito, Chacarita, Colegiales, Flores, Floresta, Liniers, Mataderos, Montserrat, Monte Castro, Núñez, Parque Chas, Parque Patricios, Recoleta, Retiro, Saavedra, San Cristóbal, San Nicolás, San Telmo, Versalles, Villa Crespo, Villa Devoto, Villa Ortúzar, Villa Santa Rita, Villa Urquiza.

-**Cluster 1 (purple):** Argentinian Restaurant, Pizza Place, Ice Cream Shop, (Coffe Shop)

Neighborhoods: Almagro, Barracas, Boedo, Coghlan, Constitución, La Boca, La Paternal, Palernom Parque Chacabuco, Puerto Madero, Vélez Sarsfield, Villa del Parque, Villa General Mitre, Villa Lugano, Villa Luro, Villa Pueyrredón, Villa Real, Villa Riachuelo.

-**Cluster 2(Turquesa):** Flea Market and Park, bus station (Fast food restaurant, liquor store)

Cerrillos, El Bosque, Quilicura

-**Cluster 3 (green):** Sushi, Chinese, Fast Food Restaurant (Farmers Market)

Cerro Navia, Conchalí, Estación Central, Huechuraba, Independencia, La Cisterna, la Florida, La Granja, La Pintana, Lo Barnechea, Lo Espejo, Lo Prado, Macul, Maipú, Pedro Aguirre Cerda,Pudahuel, Quinta Normal, San Joaquin, San Miguel,San Ramón, Puente Alto,San Bernardo, Nueva Pompeya, Parque Avellaneda, Villa Soldati.

-**Cluster 4 (orange):**Outdoors & Recreation and Brewery

Neighborhoods: Renca

When making the choropleth map of the average house prices per square meter, the highest prices are located in Puerto Madero with 6666 dollars per m2, however, this neighborhood belongs to cluster 1, only present in Buenos Aires.

For cluster 0, the average house price per m2 for neighborhoods in Buenos Aires ranges between 2041 and 4353 dollars, while for neighborhoods in Santiago it ranges between 1728 and 3488 dollars, both ranges are similar, for which a person at Moving from one city to the next can find a similar place to where you leave and maintain a similar lifestyle.

For Cluster 3, the price ranges for the neighborhoods of Buenos Aires are between 885 and 3197 dollars, while for the neighborhoods of Santiago these range between 1228 and 3488 dollars, for which it is also possible to find a home with similarly priced in a similar neighborhood.
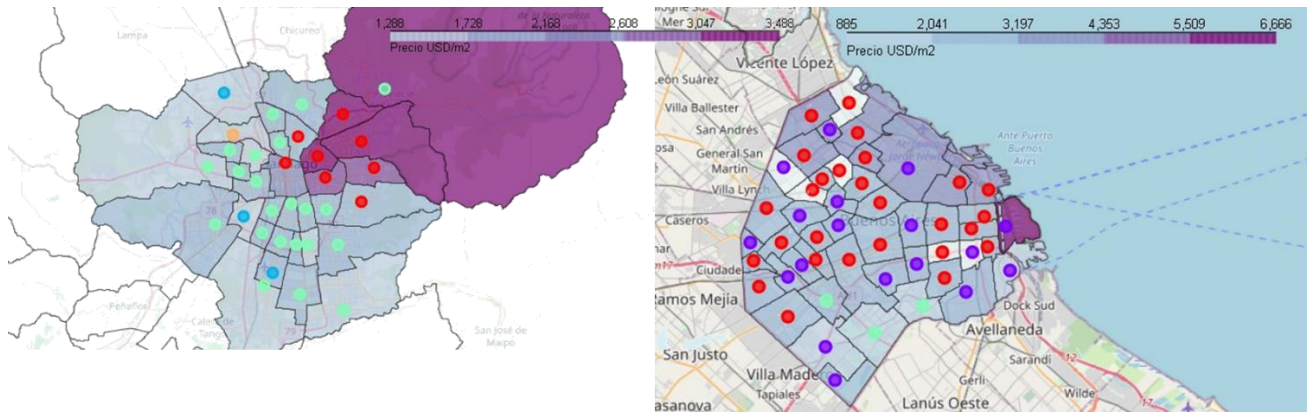
Figure 17: Choropleth map clusters and average house prices per square meter, Santiago de Chile and Buenos Aires, Argentina.

5. Discussion

In the city of Santiago it is observed that the amount of venues available is not the same for each neighborhood, there are even some that have less than 10 venues. The 18 neighborhoods that reach the maximum limit established are characterized by being tourist, commerce and bohemian sectors. In contrast, most neighborhoods in Buenos Aires reach this limit and a more homogeneous distribution of the venues and their categories is observed. Probably, this difference is due to the fact that, the venue information provider Foursquare has more information on file for the city of Buenos Aires than for Santiago. Therefore, to make a more detailed grouping of data, it would be necessary to obtain information from other sources for future analysis.

Despite the above it is possible to extract useful information.

In both cities the most common categories are related to food, in the city of Santiago sushi and bakery restaurants dominate, while in Buenos Aires, argentine restaurants are the majority, surpassing the number of 100 places, followed by pizza places.

Both cities have unique neighborhoods. As Renca in the case of Santiago and Villa Soldati, both with the least amount of information available.

Based on the results, both cities have neighborhoods with similar characteristics, present in clusters 0, characterized by venues Restaurant, Coffe Shop (Bakery, Pizza Place, Bar) and cluster 3 with venues, Sushi, Chinese, Fast Food Restaurant (Farmers Market). By adding the house price per square meter factor, these neighborhoods also fall within a similar range.

Other interesting data to add to future analyzes would be the crime rate for each neighborhood and the distance to your workplace or from the city center.

6. Conclusion

163 unique categories were found in the city of Santiago, while 200 were found in Buenos Aires. When analyzing both cities, the unique categories increase to 247.

The most common categories of venues are related to food. For Santiago it corresponds to Sushi restaurants and for Buenos Aires, Argentine restaurants. When analyzing both cities, the most common venues are Argentine restaurants, pizza places, ice cream shop and coffee shop.

When analyzing both cities if there are similar neighborhoods, by clustering analysis two clusters were obtained, cluster 0 with the most common venues of Restaurant, Coffee Shop (Bakery, Pizza Place, Bar) and made up of the neighborhoods of: Santiago, La Reina , Las Condes, Ñuñoa, Peñalolén, Providencia, Recoleta, Vitacura, Agronomia, Balvaner, Belgrano, Caballito, Chacarita, Colegiales, Flores, Floresta, Liniers, Mataderos, Montserrat, Monte Castro, Núñez, Parque Chas, Parque Patricios, Recoleta, Retiro , Saavedra, San Cristóbal, San Nicolás, San Telmo, Versalles, Villa Crespo, Villa Devoto, Villa Ortúzar, Villa Santa Rita, Villa Urquiza and cluster 3, with the most common venues of sushi restaurant, Chinese restaurant and Fast Food Restaurant (Farmers Market), the neighborhoods that comprise it correspond to: Cerro Navia, Conchalí, Central Station, Huechuraba, Independencia, La Cisterna, Florida, La Granja, La Pintana, Lo Barnechea, Lo Espejo, Lo Prado, Macul, Maipú, Pedro Aguirre Cerda , Pudahuel, Quinta Normal, San Joaquin, San Miguel, San Ramón, Puente Alto, San Bernardo, Nueva Pompeya, Avellaneda Park, Villa Soldati.

By adding the average price factor of the square meter of housing, the ranges are similar for all the neighborhoods present in each cluster.

In conclusion, it is possible to generate some useful information for people who are considering moving among Buenos Aires and Santiago or vice versa.