

IBM Data Science Professional Certificate

Capstone Project

Santiago's Communes and Infrastructure

Cristian Castro Álvarez

Introduction and objectives

Santiago de Chile is the capital, largest and most populated city of Chile. The city is divided into communes, that are quite diverse and different in terms of mean monthly income, facilities, infrastructure and venues. In this regard, Santiago can be quite segregated and the quality of life between different communes can be very different. Due to these big differences between each commune, choosing a place to live can be challenging to foreigners or tourists.

In this project I'll try to cluster Santiago's communes in order to give advice on where to live or stay for a time depending on your income and interests (of venues obtained through Foursquare data). This project is aimed at giving advice to foreigners that want to visit or live in Santiago de Chile (Chile, 2020) and intends to provide information of the city beforehand.

Data and Methodology

The methodology and the data to be used are summarized in the figure below.

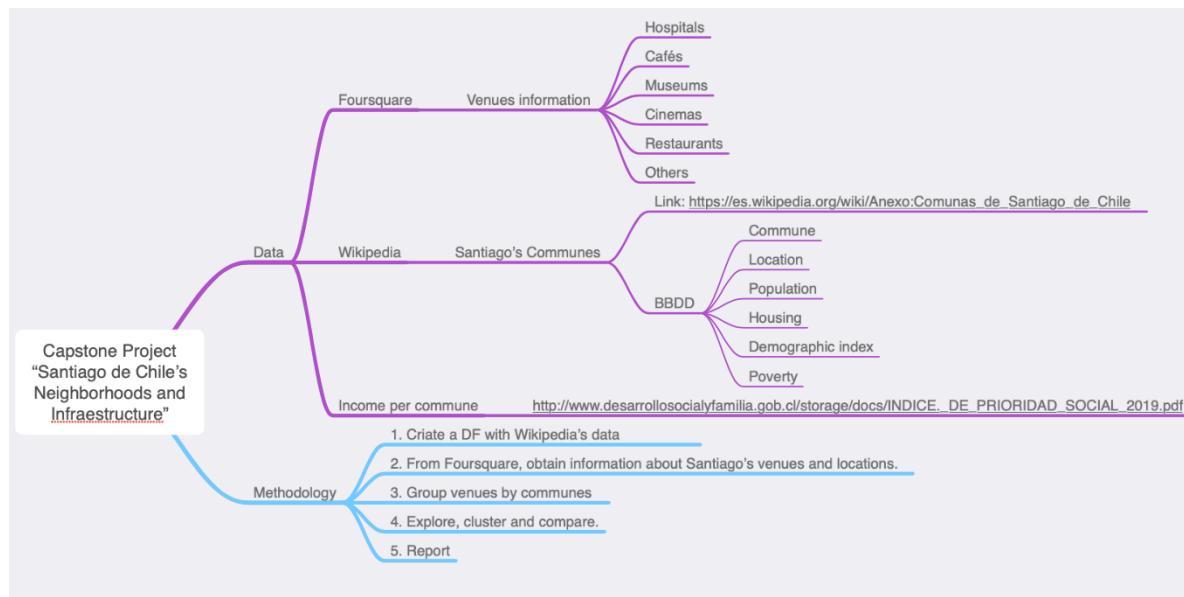


Figure 1 Data and Methodology – Capstone Project.

The data will be obtained from:

- Foursquare (Venues, description and location)

- Wikipedia (Communes description and location)
- Government of Chile (Average monthly income per commune)

With all of the above and considering the methodology shown, good advices can be done based on interests and budget.

Results

The communes of Santiago de Chile are located within the area according to the figure below.

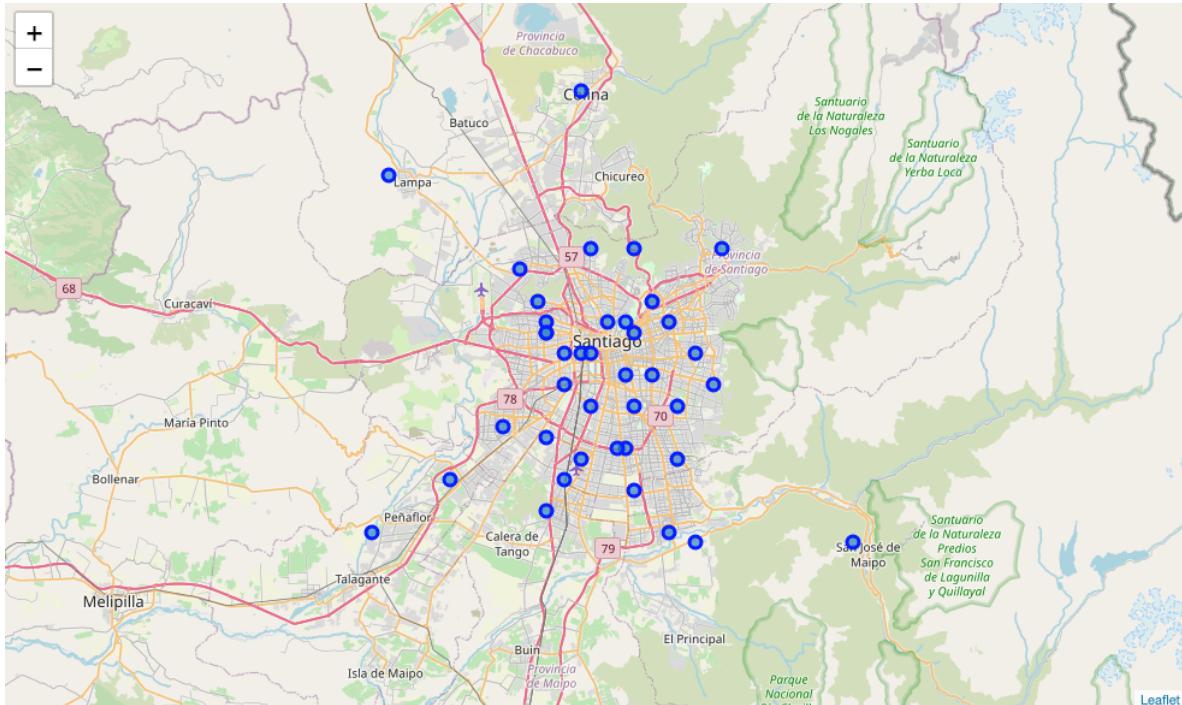


Figure 2 Plan View of Location of Communes in Santiago de Chile.

As the methodology stated, the communes were clustered based on the information provided by Foursquare Data and also including their average income – provided by the Government of Chile.

The methodology for clustering was based on an unsupervised machine learning method called K-Means. To obtain these results, the independent variables were defined as: type of most common venues in the commune, normalized average income of the commune and location of the commune. The results are shown in the figure below and detailed in Table 1.

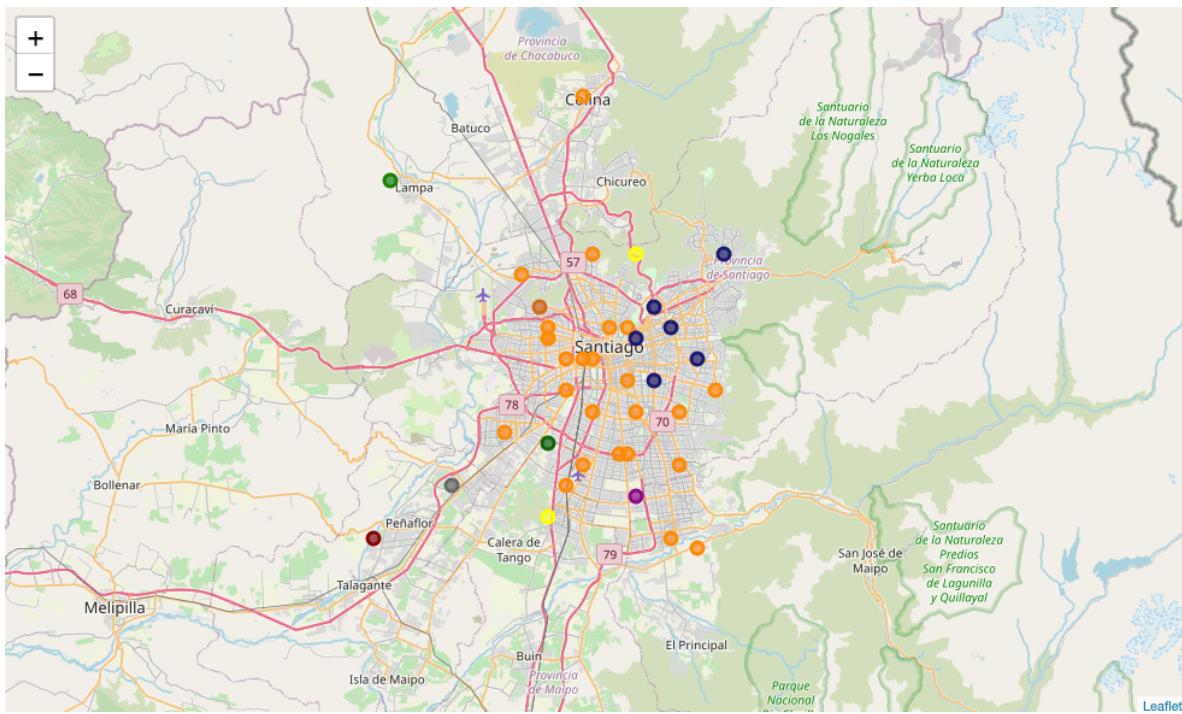


Figure 3 Plan View of Communes Clustering in Santiago de Chile.

Table 1 Detail of Cluster in Santiago de Chile.

| Cluster | Communes in the Cluster | AVG Income (CLP) | 5 Main Venue Types in the Cluster |
|-----------|---|------------------|---|
| Cluster 0 | Conchali | \$695.487 | Japanese Restaurants, Yoga Studios, Dive Bars, Flea Markets and Fish Markets. |
| Cluster 1 | Cerrillos, Cerro Navia, Colina, El Bosque, Estacion Central, Huechuraba, Independencia, La Cisterna, La Florida, La Granja, Lo Prado, Macul, Maipu, Pedro Aguirre Cerda, Peñalolen, Pirque, Pudahuel, Puente Alto, Quilicura, Quinta Normal, Recoleta, San Joaquin, San Miguel, San Ramon and Santiago. | \$789.217 | Bakeries, Sushi Restaurants, Restaurants, Pharmacies and Plazas. |
| Cluster 2 | La Reina, Las Condes, Lo Barnechea, Nunoa, Providencia and Vitacura. | \$1.588.920 | Coffee Shops, Restaurants, Pizza Places, Pharmacies and Plazas. |
| Cluster 3 | Lampa | \$793.962 | Mountains, Yoga Studios, Foods, Flea Markets, Fish Markets. |
| Cluster 4 | Penaflor | \$728.946 | Fried Chicken Joints, Fast Food Restaurants, Yoga Studios, Foods, Flea Markets. |
| Cluster 5 | La Pintana | \$561.901 | Others, Plazas, Furniture/Home Stores, Big Box Stores and Yoga Studios. |
| Cluster 6 | Renca | \$652.198 | Grocery Stores, Convenience Stores, Plazas, Gardens and Dive Bars. |
| Cluster 7 | Lo Espejo | \$599.175 | Theme Parks, Others, Soccer Fields, Gym Centers and Delis. |
| Cluster 8 | Padre Hurtado | \$700.422 | Lakes, Big Box Stores, Basketball Courts, Soccer Courts and Sporting Goods. |
| Cluster 9 | San Bernardo | \$682.645 | Bakeries, Intersections, Parks, Shop & Services and Bars. |

Besides the results from the clustering process itself, we can see a plot contrasting the average income of each cluster in Santiago de Chile.

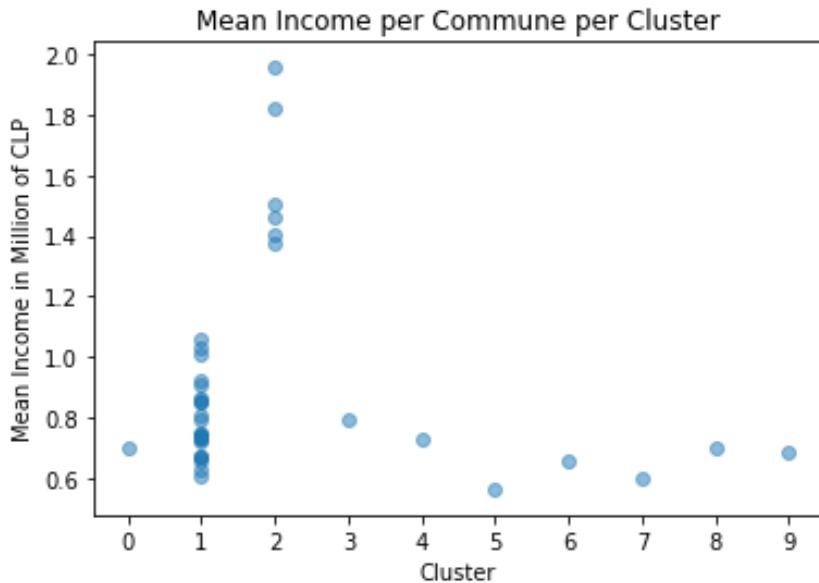


Figure 4 Average Income per Cluster in Santiago de Chile.

From Figure 4 it can be seen that Cluster 2 outstands not only by the main type of venues that it holds (Coffee Shops, Restaurants, Pizza Places, Pharmacies and Plazas) but also but a significant different average income. Although correlation doesn't mean causality, in this topic it can be assumed that because of the attractive venue types and the high income relative to the rest, the communes in Cluster 2 can be understood as safe and fun for the average foreign tourist.

Discussion

The unsupervised machine learning method of K-Means clustering was used to cluster the 37 different communes of Santiago de Chile in terms of average income and venue types of each communes.

From the above, the method gave 10 clusters where the cluster 2 highlights as the most attractive for foreign tourist. This is due to fun and attractive common venues such as: Coffee Shops, Restaurants, Pizza Places, Pharmacies and Plazas, and also high average income. Although correlation doesn't mean causality, this can be understood as a safe and fun place for foreign tourists.

Conclusion

Based on the Foursquare Data for the city of Santiago de Chile and the economic data provided by the Government of Chile, if a foreign tourist visits Santiago de Chile, it is highly recommended that he or she stays in communes of cluster 2, such as: La Reina, Las Condes, Lo Barnechea, Nunoa, Providencia and Vitacura. These communes guarantee fun venues to attend, and safe environment.