

Santiago de Chile neighborhood similarity

Beltrán Larraín Budge

May 8, 2020

IBM Coursera Capstone Project

Introduction

- Knowing similarities between neighborhoods could help you after the COVID-19 pandemic.
- Most known places will be crowded after quarantine.
- If a neighborhood is known for having a lot of good places to eat, can we know if there is a substitute in case this one has too much demand?
- Can help you try new thing that you already like.

Data acquisition and preparation

- Postal Codes and Neighborhood names from https://es.wikipedia.org/wiki/Anexo:C%C3%B3digos_postales_de_Chile
- Latitude and Longitude from Geolocator package.
- Fixing a couple of points manually to make sense with reality. Some neighborhood coordinates were out of the city since they are huge and have rural areas, for those the location of the municipality building was considered as new center.
- Foursquare API for nearby venues request.
- Eliminating the incomplete and insignificant data.

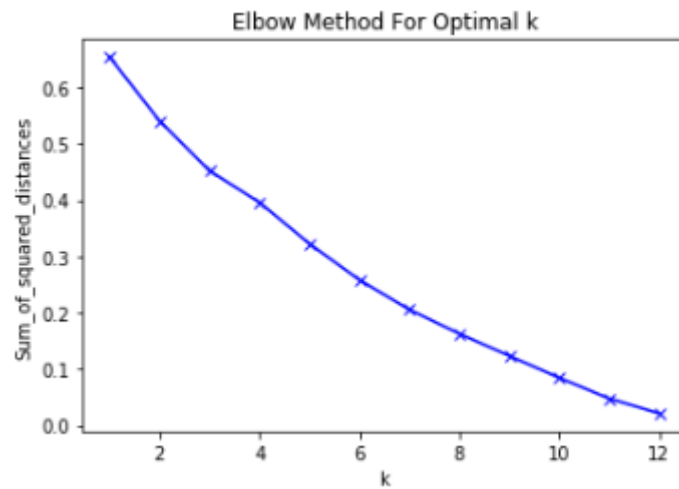
Data acquisition and preparation

- Foursquare API for nearby venues request.
- Number of nearby venues per neighborhood:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
La Cisterna, CL	14	14	14	14	14	14
La Florida, CL	11	11	11	11	11	11
Las Condes, CL	30	30	30	30	30	30
Lo Barnechea, CL	15	15	15	15	15	15
Lo Prado, CL	16	16	16	16	16	16
Macul, CL	11	11	11	11	11	11
Peñalolén, CL	16	16	16	16	16	16
Providencia, CL	39	39	39	39	39	39
Puente Alto, CL	21	21	21	21	21	21
San Bernardo, CL	40	40	40	40	40	40
San Miguel, CL	25	25	25	25	25	25
Santiago, CH	61	61	61	61	61	61
Ñuñoa, CL	31	31	31	31	31	31

Clustering

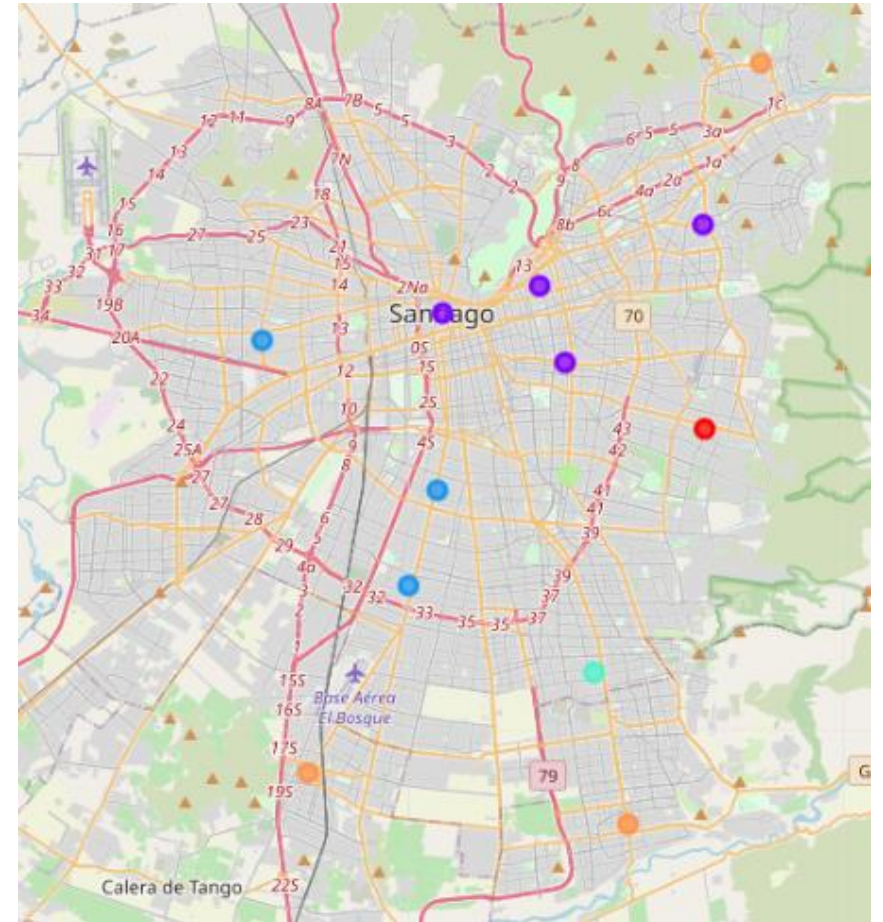
- Since we are working with unlabeled data: K-means
- Best K based on elbow method using Sum of Squared Distance Error.



- Optimal K=6

Clustering

- Cluster 0 has only Peñalolén assigned.
- Cluster 1 has Las Condes, Ñuñoa, Providencia and Santiago.
- Cluster 2 has La Cisterna, Lo Prado and San Miguel.
- Cluster 3 has La Florida.
- Cluster 4 has Macul.
- Cluster 5 has Lo Barnechea, Puente Alto and San Bernardo.



Conclusions

1

Cluster 1 has Las Condes, Ñuñoa, Providencia and Santiago, which the 4 are known for being the best places to go to restaurants, and the data shows that for all of them, the top 10 venue types are almost exclusively eating spots. This shows that the most concentration of restaurants is there.

2

Cluster 2, with La Cisterna, Lo Prado and San Miguel which are shown in the data as more rounded places with a lot to offer, there are places to eat but also a lot of pharmacies, bakeries, shops, offices and metro stations. This shows that their neighborhoods are less specialized and have many different venues to offer, making them great for living since you have almost anything nearby and the areas are less commercial.

3

Cluster 5 shows Lo Barnechea, Puente Alto and San Bernardo, the 3 most peripheral neighborhoods of the city, so having them in the same cluster makes a lot of sense. Since they are the most far away neighborhoods from the center, they need to be self-sufficient, and the data shows the greatest variety in venues, with gyms, jewelry stores, pet shops, outdoor activities, and plazas.