

Identification of zones with a high density of green areas in Santiago, Chile

Mauricio Santelices

July, 2020

1. Introduction

1.1 Background

Chile is among the countries with the highest obesity levels in the world, which, together with the high rate of sedentary lifestyle, represents a serious risk to the health of the population. In Santiago, the capital, lives about a third of the country's population, so it's expected that the largest number of patients with sedentary lifestyle and obesity-related diseases come from this city.

In recent times, plans have been enacted to promote a healthy lifestyle, changes in food habits and the practice of outdoor physical activity in the various parks of the country, however, the city of Santiago presents an uneven distribution of green areas (GA) that could be used for this purpose, so it is important to determine the distribution of these areas in order to plan improvements in access to parks or in the construction of new green areas to the citizens.

1.2 Problem

The purpose of this project is to identify the areas with the highest density of GAs in the Santiago Metropolitan Area (SMA), made up of 36 communes (the name that each administrative unit of the city receives) through a geospatial analysis based on the coordinates (latitude and longitude) of each important GA in the city.

1.3 Interest

Organizations that develop plans that promote healthy lifestyle can use the distribution of GA in the city in order to plan according to the conditions of access to them for residents of a given area. Entities responsible for proposing improvements to public transport or construction can use this information to improve access to existing GAs or to create new GAs in low-density areas.

2. Data acquisition and cleaning

2.1 Data sources

The list of communes that make up the SMA can be found [here](#), section "Comunas del Área Metropolitana de Santiago de Chile". The table lists each commune, its general location on the map and its population as of the last census (year 2017), among other data, however, it does not have the coordinates of each commune. This list was obtained using scraping techniques and the coordinates of each commune through geolocation libraries.

Additionally, I proceeded to obtain a list of venues through the [Foursquare API](#) for each commune, using a wide search radius in order to collect as many venues as possible.

2.2 Data cleaning

Once the list of all possible venues for each commune was obtained, I proceeded to select only those categorized as "Plaza", "Park" or "Garden", in addition to removing any duplicate value, since in the collection phase, I chose to use a very wide radius for each commune in the city, in order to obtain as many venues as possible.

2.3 Feature selection

After cleaning the data, a list of 37 unique GAs registered on Foursquare with their coordinates (latitude and longitude) was obtained. Because the objective of the study is to determine high-density areas and that these areas can encompass different neighboring communes, I decided to discard the commune name, in which each GA is located, so the final features only include the name of each GA, its latitude and its longitude.

	Venue	Venue Latitude	Venue Longitude
0	Parque El Rosal	-33.485326	-70.765686
1	Running Fernández Albano	-33.523544	-70.660307
2	Viveros de Renca	-33.403115	-70.734140
3	Parque Forestal	-33.436429	-70.640126
4	Parque Bicentenario	-33.400747	-70.601693

Tabla 1: Extract of the final table obtained after cleaning the data.