

- Applied Data Science Capstone
- March 2021
- Pedro Reyes O.

Exploring Best locations for Bakery store in Santiago, Chile Part 1/2

Introduction

- Santiago de Chile, is the capital and largest city of Chile as well as one of the largest cities in the Americas. It is the center of Chile's most densely populated region, the Santiago Metropolitan Region, whose total population is 7 million, of which more than 6 million live in the city's continuous urban area.
- Commercial activity is one of the drivers of economic growth and has generated many investors wanting to compete.
- Location, location is the key for success



The Problem

Small investors and innovators want to explore specialized bakery chains whose requirements are for convenience and differentiating products

The locations must combine an adequate balance between the rental cost per square meter, that there are complementary attractive places and that there is a minimum volume of population residing in the vicinity

Data acquisition and cleaning

The following sources of information are identified that will be helpful in addressing the problem:

- Latitude, longitude, Surface (km²), Population, 2020 Density (hab./km²) of districts in Santiago: https://es.wikipedia.org/wiki/Anexo:Comunas_de_Chile
- Lease values per square meter in districts in Santiago:

https://www.buenainversion.cl/blog/valor-metro-cuadrado/

Venues in Santiago:

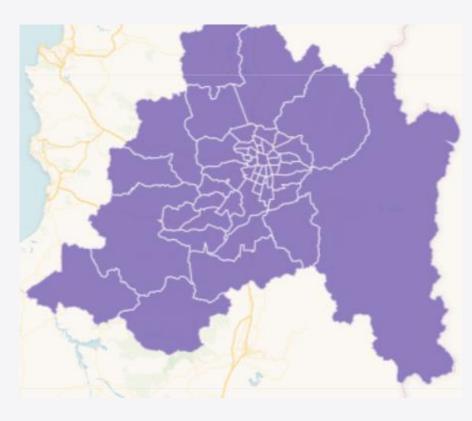
https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll
={},{}&radius={}&limit={}'

GeoJson file for districts in Santiago:

https://cephei.carto.com/tables/comunas_santiago/public

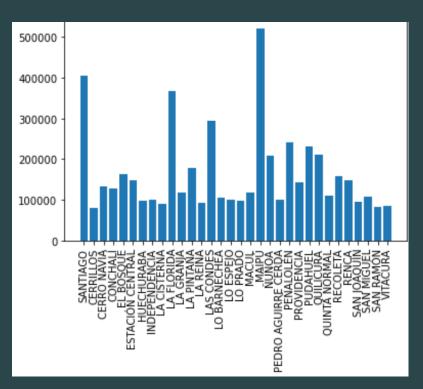
The processes necessary to obtain the data and its preparation include:

• web scrapping, direct loading, elimination of duplicates, homologate names of districts, join bases and normalize (standardize) the data before starting machine learning processes.

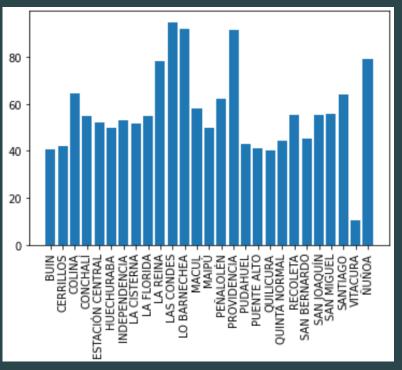


Featured exploratory views

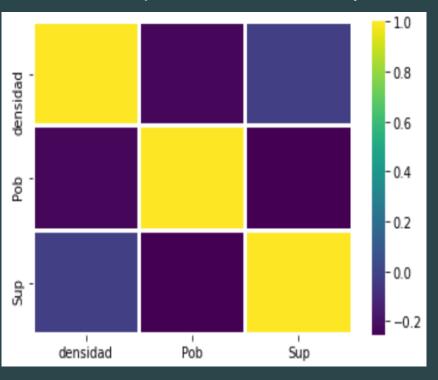
Population



Value UF/ m^2 lease

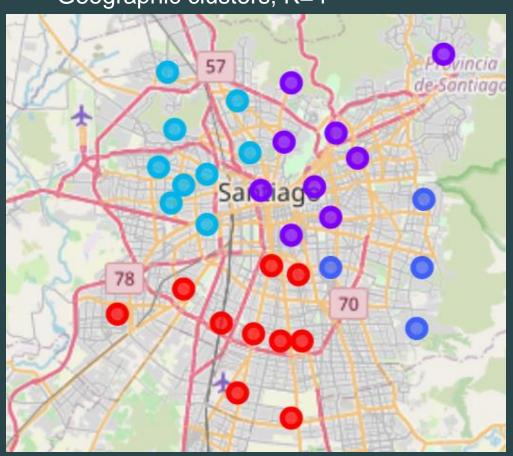


HeatMap: Correlation of Surface, poulaton and density

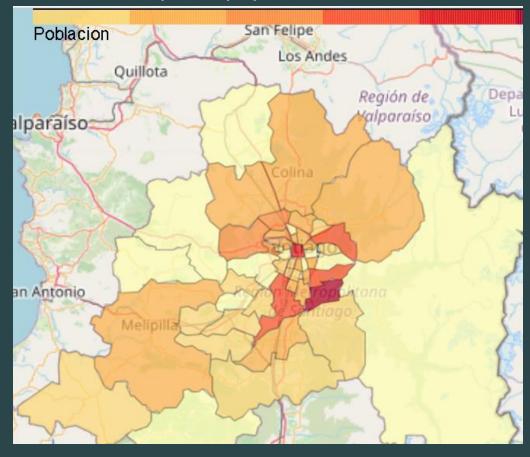


Preliminary results

Geographic clusters, K=4



Choropleth, population 2017



Methodology

Busines understanding defining the problem and its utility for the stakeholder

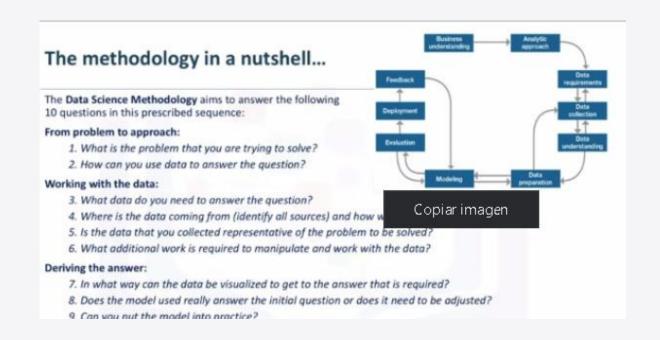
Data and processing.

Choose analytical approach: in this case the data will be explored through bar graphs, heat maps and choropleths.

Combining the data and normalizing it, an unsupervised learning technique will be applied to obtain clusters (Kmeans) and visualize them through folium.

The results will be presented graphically and tabularly with the necessary detail to be able to replicate them,

Finally, the conclusions and discussions will be analyzed, its scope and possible improvements will be seen.





- **Applied Data Science Capstone**
- March 2021
- Pedro Reyes O.

Code avaliablel, in this link:

Exploring Best locations for Bakery store in Santiago, Chile Part 1/2