POVERTY VERSUS SERVICES IN SANTIAGO'S COMMUNES

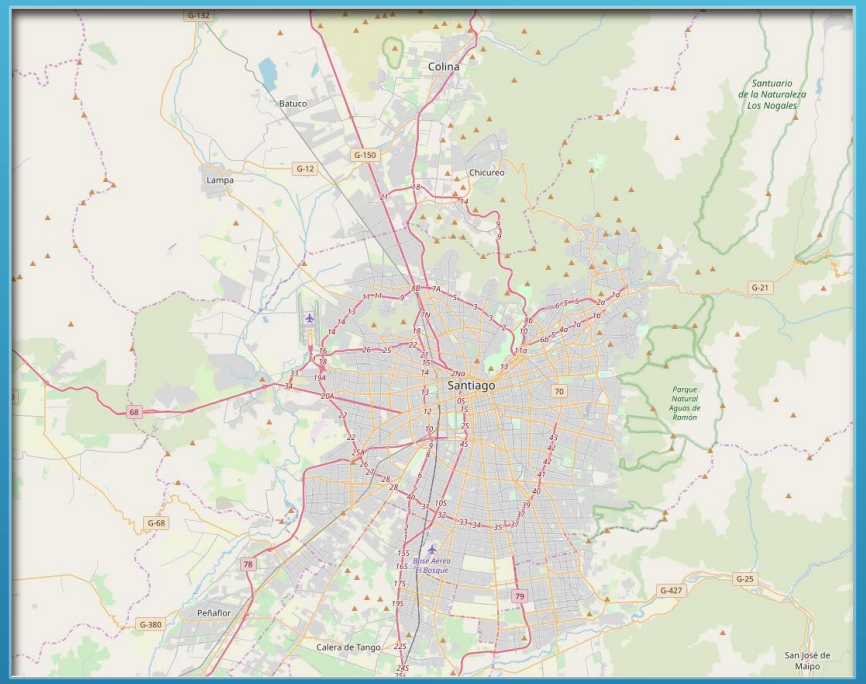
Osvaldo Gac P
Data Scientist
Dec 12, 2019

SANTIAGO OF 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 LARGEST AND MOST DENSELY POPULATED CONURBATION, THE SANTIAGO METROPOLITAN REGION, WHOSE TOTAL POPULATION IS 7 MILLION. THE CITY IS ENTIRELY LOCATED IN THE COUNTRY'S CENTRAL VALLEY. MOST OF THE CITY LIES BETWEEN 500 M (1,640 FT) AND 650 M (2,133 FT) ABOVE MEAN SEA LEVEL.

SANTIAGO IS THE CULTURAL, POLITICAL AND FINANCIAL CENTER OF CHILE AND IS HOME TO THE REGIONAL HEADQUARTERS OF MANY MULTINATIONAL CORPORATIONS. THE CHILEAN EXECUTIVE AND JUDICIARY ARE LOCATED IN SANTIAGO, BUT CONGRESS MEETS MOSTLY IN NEARBY VALPARAÍSO.

THE WHOLE OF GREATER SANTIAGO DOES NOT FIT PERFECTLY INTO ANY ADMINISTRATIVE DIVISION, AS IT EXTENDS INTO FOUR DIFFERENT PROVINCES AND 37 **COMMUNES**. THE MAJORITY OF ITS 641.4 KM2 (247.65 SQ MI) (AS OF 2002) LIE WITHIN SANTIAGO PROVINCE, WITH SOME PERIPHERAL AREAS CONTAINED IN THE PROVINCES OF CORDILLERA, MAIPO, AND TALAGANTE.

CHILE IS INTO THE 10 COUNTRIES MOST INEQUALITY IN THE WORLD, REFLECTED IN GINI'S INDEX 0.47 IN 2018, IN SANTIAGO YOU CAN FIND **POOR COMMUNES** AND **RICH COMMUNES**, NOWADAYS DIFFICULT TO UNDERSTAND IN TERM OF WHAT KIND OF SERVICES AND SHOPS THERE ARE IN EACH COMMUNE.



THE OBJECTIVE OF THE PROJECT IS TO EXPLORE THE POSSIBILITY OF, TAKING ADVANTAGE OF STATE OF THE ART MACHINE LEARNING ALGORITHMS, TO SEGMENT THE COMMUNES USING FOURSQUARE API TO GET ALL THE VENUES AND SERVICES AND COMPARE WITH THE POVERTY INDEX OF EACH COMMUNE, TO RESOLVE THE QUESTIONS:

“A MORE POOR THE COMMUNE LESS SERVICES?”

“WHAT TYPE OF SERVICES CHARACTERIZES THE POOR COMMUNES AND THE RICH COMMUNES”

THIS WILL HELP TO STAKEHOLDER TO DECIDE THE SERVICES THEY WANT TO DEVELOP IN A COMMUNE CONSIDERING THE LACK OF SERVICES PER COMMUNE AND THE LEVEL OF RICHNESS

DATA

TO CONSIDER THE PROBLEM WE CAN LIST THE DATA AS BELOW:

I FOUND THE INDEX OF POVERTY OF EACH COMMUNE TO 2015[1]

I USED FOURSQUARE API TO GET THE MOST COMMON VENUES OF GIVEN COMMUNES OF SANTIAGO[2] .

THERE ARE NOT TOO MANY MAPS OF SANTIAGO IN GEOPANDAS OR SHAPEFILE MAP, BUT NOT IN GEOJSON, SO I BUILT A MAP IN GEOJSON USING THE INFORMATION OF THE BOUNDER POINT IN CARTO MAPS[3].

METHODOLOGY

THE METHODOLOGY USED IN THE PROJECT CONSIST IN: DATA COLLECTION, PRINCIPALLY IS A SECONDARY SOURCE GET FROM SEARCHING ON INTERNET AND USING FOURSQUARE API

SEARCH FOR POVERTY INDEX BY COMMUNES AND MAP IT TO OBSERVE THE DIFFERENT ZONES IN SANTIAGO. TO GET THE DATA I SCRAPING THE WIKIPEDIA PAGE WHERE I FOUND A TABLE WITH THE INFORMATION.

I USED PYTHON FOLIUM LIBRARY TO VISUALIZE GEOGRAPHIC DETAILS OF SANTIAGO AND HOW LOOKS EACH COMMUNES WITH DIFFERENT INDEX: POVERTY AND NUMBER OF SERVICES.

USING FOURSQUARE GET THE VENUES PER COMMUNES

TO GET THE SEGMENTATION IS USED K-MEANS ALGORITHM, THE REASON TO USE THIS ALGORITHM IS BECAUSE SEGMENTATION IS THE PRACTICE OF PARTITIONING A COMMUNES BASE INTO GROUPS THAT HAVE SIMILAR CHARACTERISTICS.

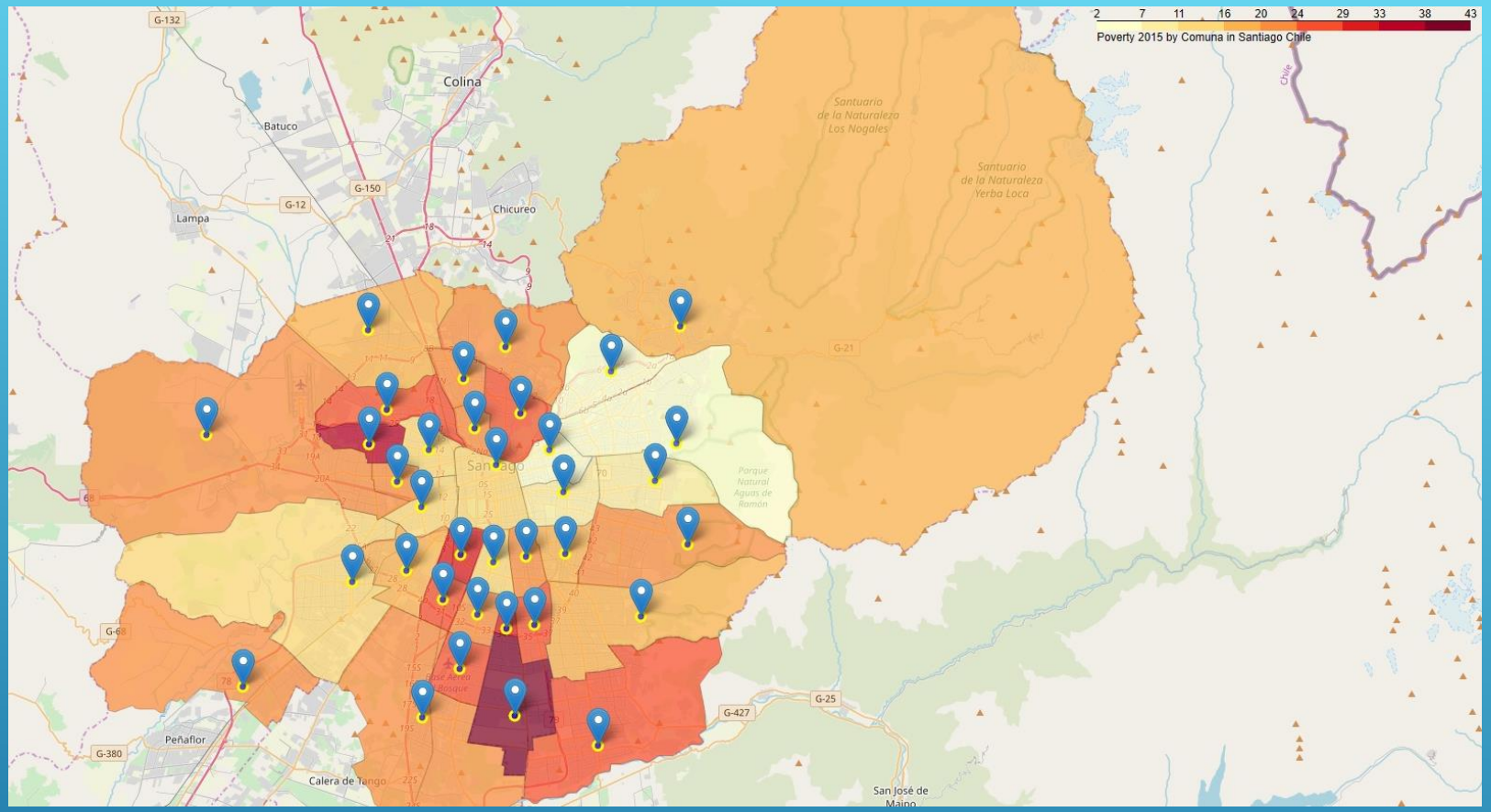
TO DETERMINE THE BETTER NUMBER OF SEGMENTS IS USED: ELBOW POINT, SILHOUETTE SCORE OR DAVIES BOULDIN SCORE AND GIVE US THE POSSIBILITY TO COMPARE ALL OF THIS METHODS
ANALYZES EACH GROUP OR SEGMENT AND THE RELATION WITH THE POVERTY INDEX FOR EACH COMMUNE

DATA PREPARATION

SCRAPING TO WIKIPEDIA PAGES, I CREATE A DATAFRAME INCLUDING THE NAME OF EACH COLUMN, POPULATION, POVERTY INDEX AND SO ON. THE TABLE WAS PROCESSED TO ELIMINATE LATIN CHARACTERS, ROW WITH BLANK NEAR ALL VALUES, CHANGE "," BY "." , ELIMINATE THE UNNECESSARY COLUMNS AND CHANGE TO ENGLISH THE COLUMNS NAMES AND VALUES TYPES.

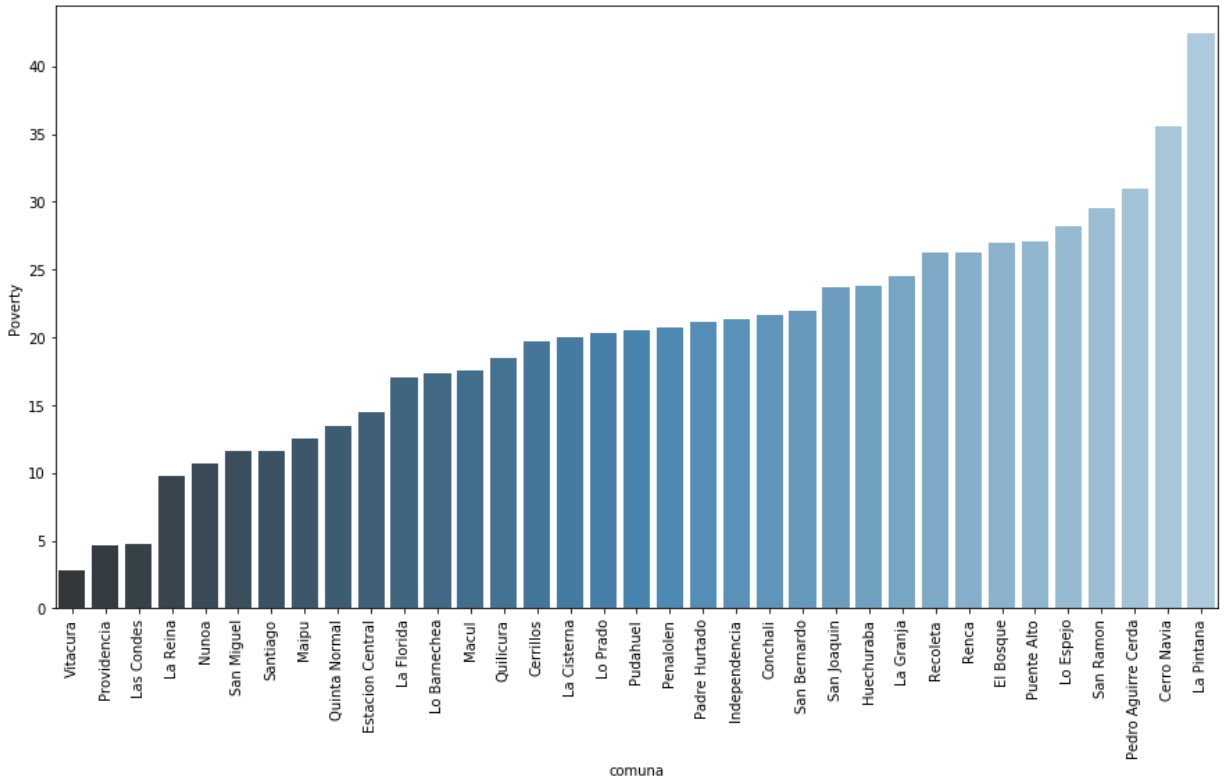
USING NOMINATIM I GOT THE GEOGRAPHICAL LOCATION OF EACH COMMUNES OF SANTIAGO

	comuna	Location	Population_(2017)	Population_density_(2002)	Population_growth_(2002-2017)	Poverty	Population_growth_(2002-2017)_%	Latitude	Longitude
0	Cerrillos	surponiente	80832	4329.08	12.9%	19.7	12.9	-33.502503	-70.715918
1	Cerro Navia	norponiente	132622	13482.91	-10.7%	35.6	-10.7	-33.425145	-70.743954
2	Conchali	norte	126955	12070.29	-4.4%	21.6	-4.4	-33.385096	-70.674491
3	El Bosque	sur	162505	12270.72	-7.3%	27.0	-7.3	-33.562352	-70.676820
4	Estacion Central	surponiente	147041	9036.31	16.6%	14.5	16.6	-33.463658	-70.704966



THE SANTIAGO CHOROPLETH MAP

THE CHOROPLETH MAP ALLOWS TO SEE THE GEOGRAPHICAL LOCATIONS OF THE POOR AND RICH COMMUNES, MORE DARK, MORE POOR



POVERTY IN EACH COMMUNE

SHOW THE RICHEST TO THE POOREST COMMUNES

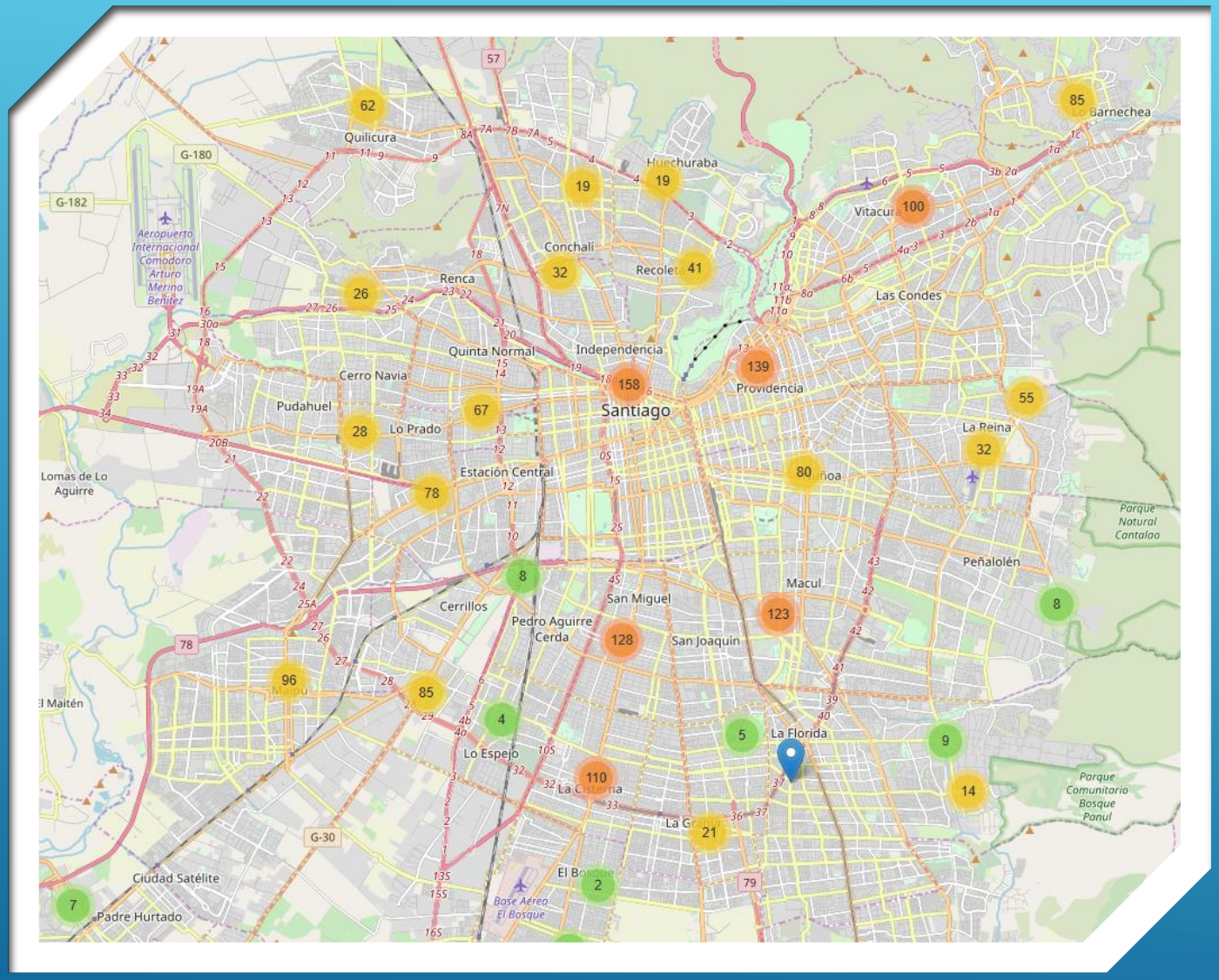
THE RICHEST ARE VITACURA, PROVIDENCIA AND LAS CONDES

THE POOREST LA PINTANA, CERRO NAVIA AND PEDRO AGUIRRE CERDA

VENUES

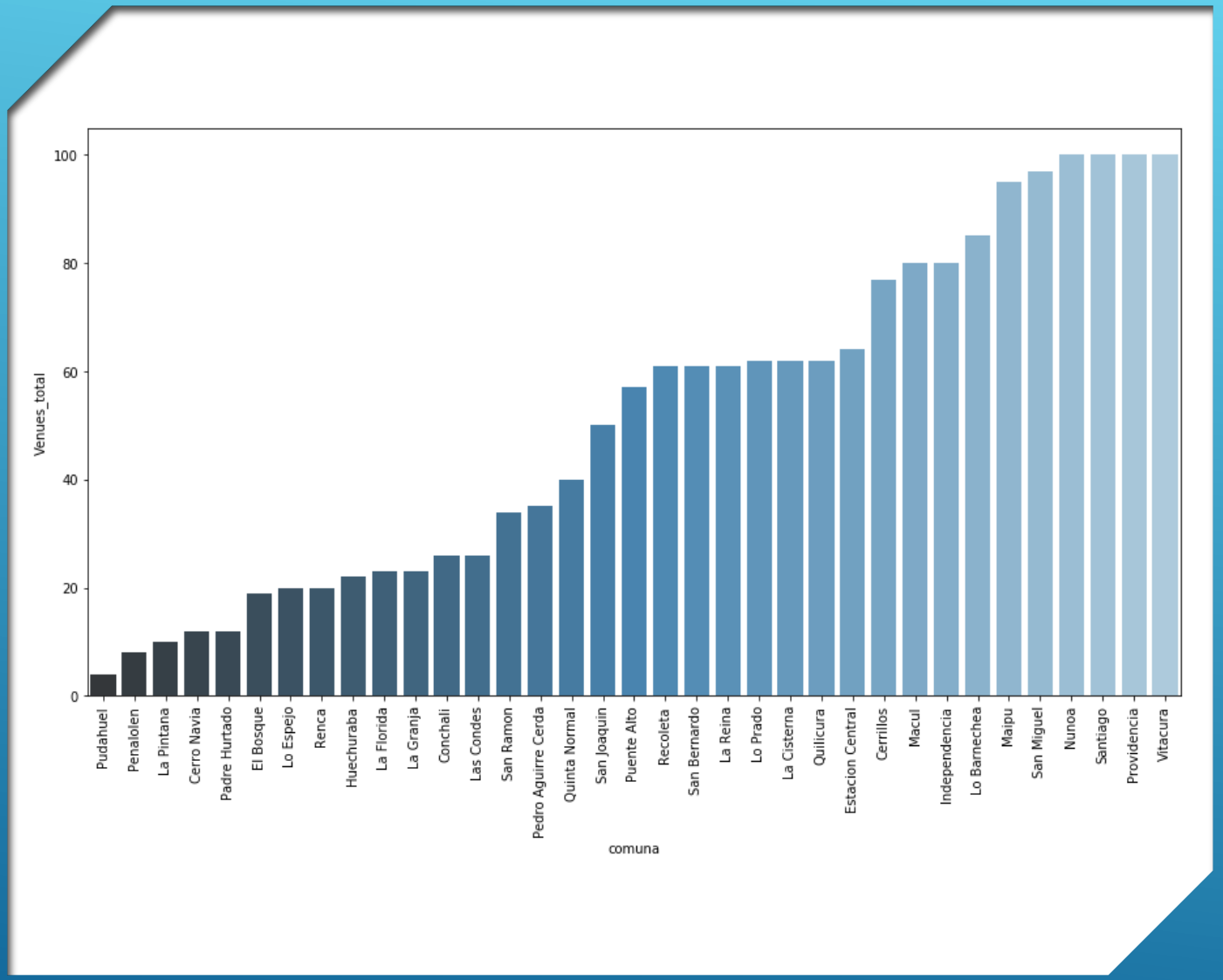
1,788 VENUES FOR 35 COMMUNES, USING RADIUS 2000 AND LIMIT 100.

THE MAP IS SO CROWDED



CLUSTER VENUES MAP

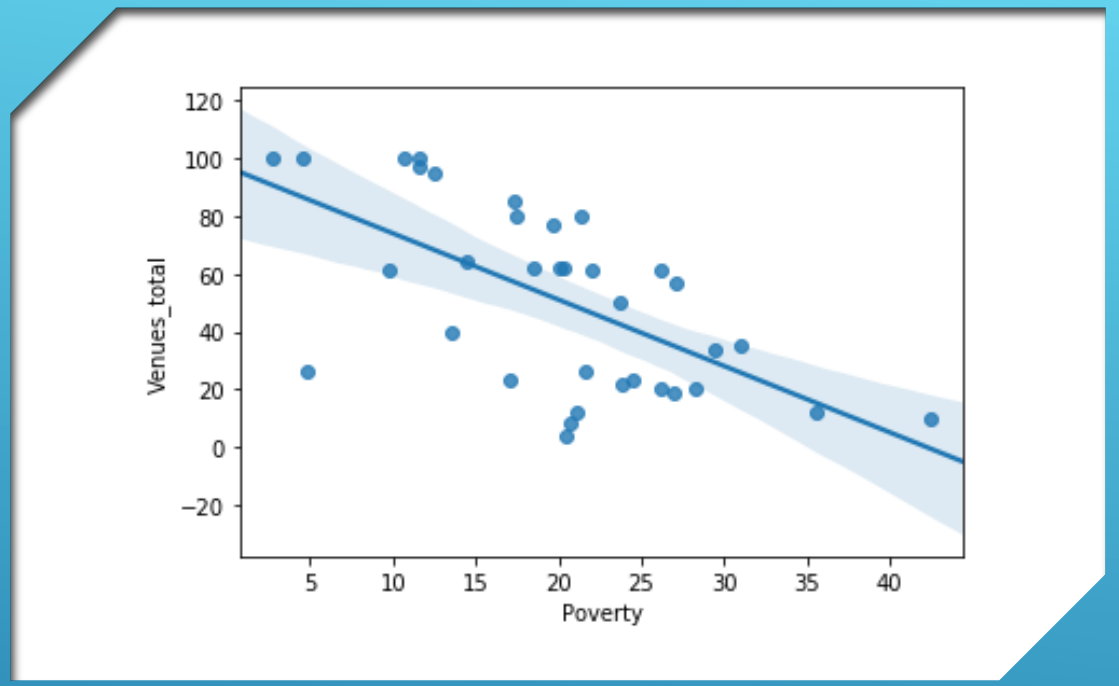
ALLOWS TO OBSERVE THE MAP MORE SORTED



VENUES IN EACH COMMUNE

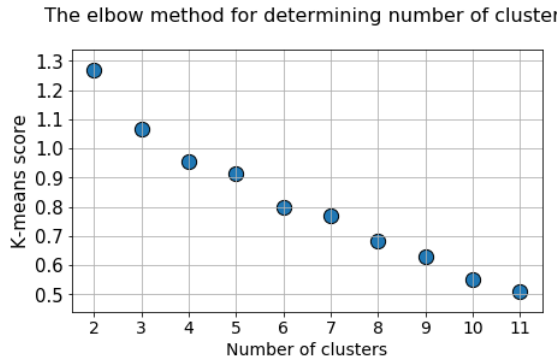
AS WE CAN SEE THE POOREST COMMUNES HAVE LESS VENUES

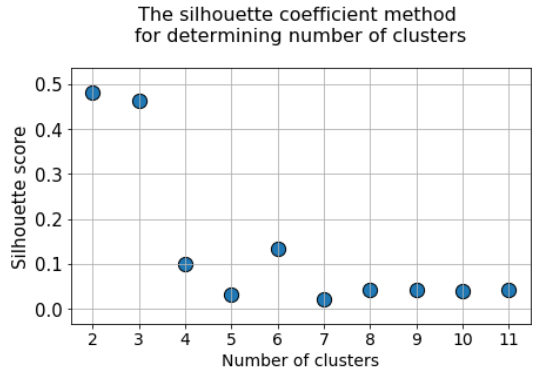
WHILE PUDAHUEL HAS 2 VITACURA HAS PROBABLE MORE THAN A 100 (THIS IS THE LIMIT IN THE STUDY) IN THE RADIUS OF 2000

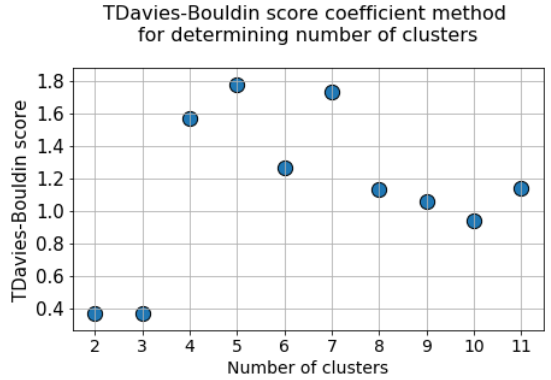


POVERTY AND TOTAL VENUES TREND

THERE IS A DOWNWARD TREND BETWEEN POVERTY AND THE TOTAL VENUES IN EACH COMMUNE



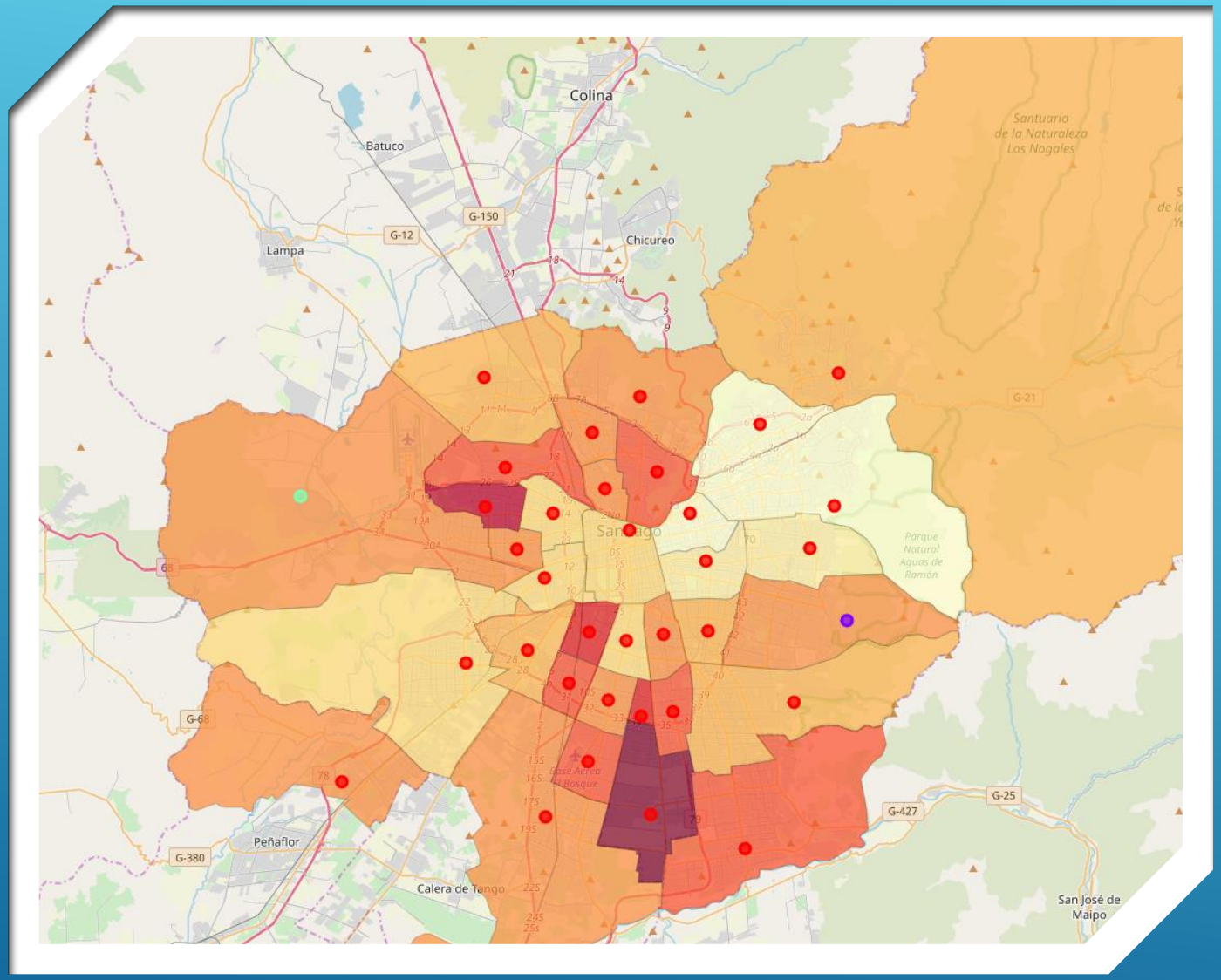




METHODS TO DETERMINE THE CLUSTER NUMBER

ELBOW POINT IS NOT CLEAR TO GET THE CLUSTER NUMBER, THERE IS NOT A CLEAR CHANGE OF SLOPE

HOWEVER THE SILHOUETTE AND DAVIES-BOULDIN SCORE SHOW 2 OR 3 LIKE A BETTER NUMBER OF CLUSTERS. I WILL USE 3 CLUSTERS



SANTIAGO SEGMENTED

APPLYING THE SEGMENT OVER A CHOROPLETH MAP, WE OBSERVE NO DIFFERENCE BETWEEN THE POOR AND RICH COMMUNES

CLUSTER 1 HAVE A LOT OF VENUES OF FOOD

CLUSTER 2 INCLUDE PARK AND GARDEN CENTER

CLUSTER 3 INCLUDE BODEGAS AND ZOO EXHIBIT

RESULTS AND DISCUSSION

SANTIAGO OF CHILE IS A BIG CITY AND VERY POPULATED, WHEN TRAVEL ALONG THE CITY IS VERY CLEAR WHAT ARE THE POOR COMMUNES AND THE RICH COMMUNES, BECAUSE THE SIZE OF HOUSES, GARDENS AND THE CARS, BUT IT SEEM WHEN COMPARE THE COMMUNES WITH THE MOST COMMON VENUES, THERE IS NO DIFFERENCES BETWEEN THEM.

THERE IS A DOWNWARD TREND BETWEEN POVERTY AND THE TOTAL VENUES IN EACH COMMUNE, SO IT IS EXPECTED THAT THERE ARE MORE VENUES IN RICH COMMUNES, WHERE PEOPLE CAN AFFORD MORE RESTAURANT DINNER, GO SHOPPING AND SO ON.

I USED K-MEANS TO DO THE SEGMENTATION OF THE COMMUNES, BUT THERE ARE OTHERS ALGORITHM TO DEVELOP THIS TASK AND I ALSO INCLUDED 35 COMMUNES OF SANTIAGO, BUT IT COULD BE INCLUDED IN THE STUDY THE FARTHEST COMMUNES, SINCE THOSE MAINTAIN RURAL LIFE TO GET ANOTHER SEGMENT.

WHEN ANALYZE THE DIFFERENT SCORE, THE ELBOW POINT IS NOT CLEAR TO DETERMINE THE NUMBER OF CLUSTERS, PROBABLY THE CLUSTER ARE NOT CLEARLY SEPARATED, BUT WHEN IS ANALYZE THE SILHOUETTE OR DAVIES BOULDIN SCORE IS CLEARER TO CONCLUDE THERE ARE 2 OR 3 CLUSTERS. I TOOK 3 CLUSTER, HOWEVER THE DIFFERENCE IS LITTLE

IT IS VERY KNOWN THE DIFFERENCE BETWEEN THE POOR AND RICH COMMUNES, TO DETERMINE THE DISSIMILARITY IS NECESSARY TO UNDERSTAND THE BEHAVIOR OF THE PEOPLE, SO IF THE ANALYSIS IS FOCUS IN THE SERVICES LIKE BANK'S BRANCH OFFICES, CARS SHOPS AND MALL IS POSSIBLE TO FIND MORE DIFFERENCE.

OTHER IMPORTANT PROBLEM IS DETERMINING THE EXACT BORDER OF EACH COMMUNE, TO AVOID INCLUDE A VENUE IN THE COMMUNE THAT DOES NOT BELONG TO. I TOOK 2 KM, BUT IS VERY PROBABLE THERE ARE SOME VENUES IN 2 COMMUNES AND OTHERS THAT WERE INCLUDED, BECAUSE THE BORDER IS BIGGER THAN 2 KM.

CONCLUSION

USING THE ACTUAL DATA ANALYTICS AND ML TOOLS TO GET AND ANALYZE INFORMATION MAKE THE LIFE EASIER AND ALLOW TO TAKE BETTER DECISIONS.

IN THIS STUDY I FOUND A DOWNWARD IN THE AMOUNT OF VENUES THERE ARE IN THE POOREST VERSUS THE RICHEST COMMUNES BUT THERE IS NO DIFFERENCE COMPARING THE KIND OF VENUES, WHEN THE COMPARISON IS HAVING DONE TAKING THE COMMON VENUES, BUT FOR MY SURPRISE THE FOOD VENUES ARE A LOT IN ALL THE COMMUNES.