Paw Print of Inequality:

SOCIO ECONOMIC CATEGORIZATION OF SAN FRANCISCO DE QUITO NEIGHBORHOODS

Business Problem/Research Objective

The actual study wants to pinpoint the living and socio-economic conditions in every neighborhood around the metropolitan district of San Francisco de Quito. With the help of Geolocation tools and some data science techniques like clustering, we can classify Neighborhoods based on socio economic information of each neighborhood. The result of the study could help to answer some important question about the inequality around the city:

- Which neighborhoods have better or worst living conditions (House construction features and basic services like drinking water and home waste management)?
- Difference of Education skills and conditions between neighborhoods
- Teen pregnancy rates
- Branch and job category
- Neighborhoods with similar commercial characteristics
- Health coverage

Data and Sources

Fourscore API

From this resource I am going to take the information of venues of each neighborhood like name and classification (restaurant, coffee shop, gym, etc)

Geo Data from "http://www.codigopostalecuador.com/quito-1896:

This page has information of zip code and location of Quito

I will use this web page to make a data frame of latitude and longitude information, with the help of beautifullsoup.

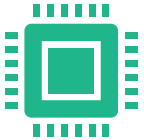
Socio-Economic Data base (Instituto de la Ciudad, QUITO)

In this base we can find some economic demographic and social indicators, like teen pregnancy rate, education attendance, population, density, health coverage, living conditions.

METHODOLOGY

Data Acquisition and cleaning

- Import the Geo data from the web page: http://www.codigopostalecuador.com/quito-1896

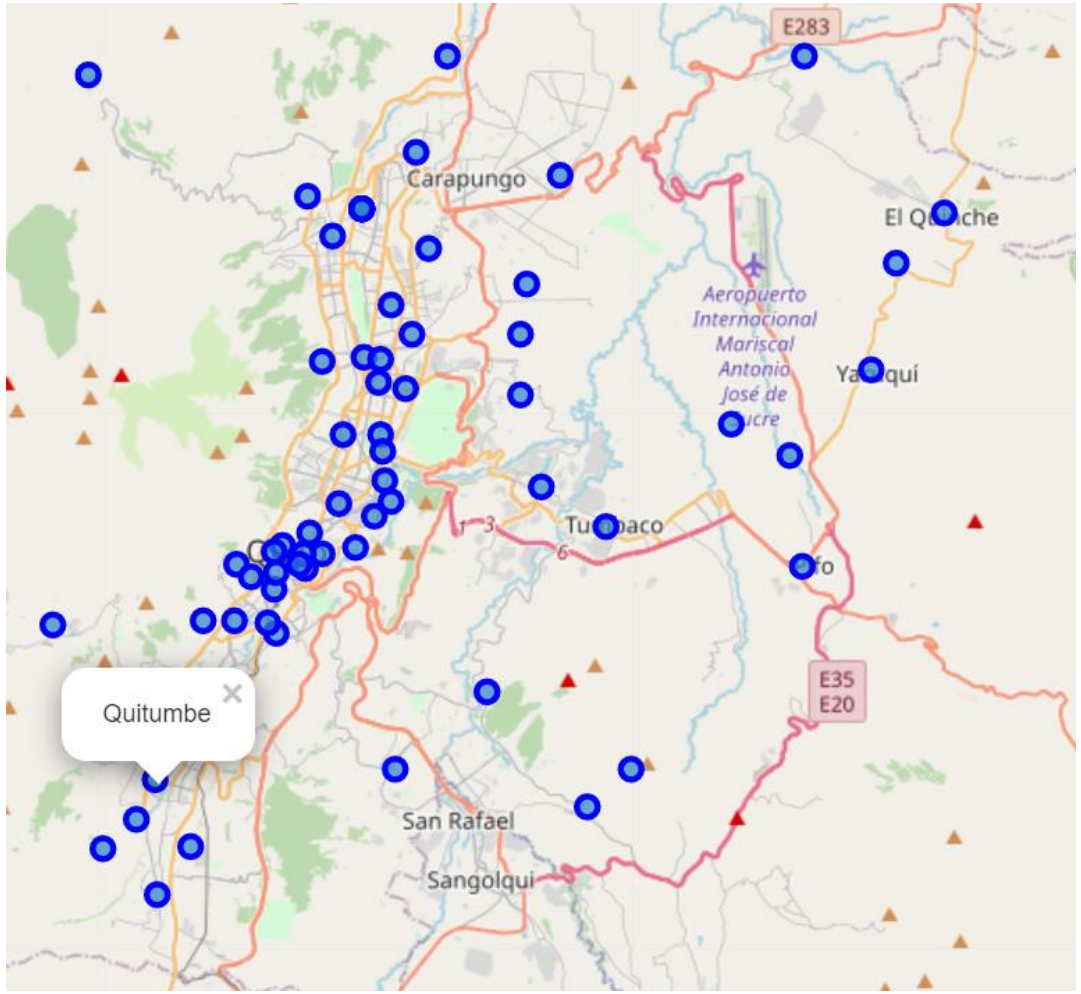


- Read and decode the web information with “beutyfullsoup” and transform to “pandas” data frame.

- Develop a geo map with the latitude and longitude information of each neighborhood, to check if every single point is locating correctly.

- There is a little percent of elements inside de data frame that have wrong geo location information

- Wrong geo location information were replaced with the right one.



	Codigo Postal	Lugar	Latitud	Longitud
1	170151	Alangasi	-0.305050	-78.413440
2	170101	Alfaro (Chimbacalle)	-0.233330	-78.516670
3	170152	Amaguaña	-0.380840	-78.515440
4	170153	Atahualpa (Chabaspamba)	0.167761	-78.360507
5	170129	Belisario Quevedo	-0.108559	-78.487738
6	170102	Benalcazar	-0.182620	-78.481220
7	170154	Calacali	-0.001140	-78.513550
8	170155	Calderon (Carapungo)	-0.097490	-78.422510
9	170120	Carcelen	-0.089710	-78.469920
10	170130	Centro Historico	-0.218822	-78.513549

Foursquare Venues (API)

Using clean base of geo location and some specifications like radius 500 and limit of requirements (1000), we can extract the venue information of each districts of San Francisco de Quito city from Foursquare page and put it in a data frame.

	Lugar	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alangasi	-0.305050	-78.413440	Parque De Alangasi	-0.307701	-78.415412	Park
1	Alangasi	-0.305050	-78.413440	Cancha Vieja	-0.306081	-78.415789	Soccer Field
2	Alfaro (Chimbacalle)	-0.233330	-78.516670	Museo Interactivo de Ciencia	-0.236313	-78.516309	Science Museum
3	Alfaro (Chimbacalle)	-0.233330	-78.516670	MIC	-0.231879	-78.515172	Science Museum
4	Alfaro (Chimbacalle)	-0.233330	-78.516670	Mesón de la Recoleta	-0.231333	-78.512786	South American Restaurant

Generate dummies of every single venue category for every single venue

	Lugar	Advertising Agency	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Alangasi	0	0	0	0	0	0	0	0
1	Alangasi	0	0	0	0	0	0	0	0
2	Alfaro (Chimbacalle)	0	0	0	0	0	0	0	0
3	Alfaro (Chimbacalle)	0	0	0	0	0	0	0	0
4	Alfaro (Chimbacalle)	0	0	0	0	0	0	0	0

Calculate the percent that the venue category represents in the total number of categories in the neighborhood.

	Lugar	Advertising Agency	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint
0	Alangasi	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000
1	Alfaro (Chimbacalle)	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000
2	Amaguaña	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000
3	Belisario Quevedo	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000
4	Benalcazar	0.0	0.000000	0.000000	0.015152	0.000000	0.0	0.000000	0.000000	0.0	0.015152

	Lugar	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	C
0	Alangasi	Park	Soccer Field	Seafood Restaurant	Food & Drink Shop	Flea Market	Fire Station	Fast Food Restaurant	Farmers Market	
1	Alfaro (Chimbacalle)	South American Restaurant	Science Museum	Comedy Club	Food & Drink Shop	Food	Flea Market	Fire Station	Fast Food Restaurant	
2	Belisario Quevedo	Bus Station	Soccer Stadium	Seafood Restaurant	Farmers Market	Food & Drink Shop	Food	Flea Market	Fire Station	
3	Benalcazar	Italian Restaurant	Hotel	Bakery	Coffee Shop	Ice Cream Shop	French Restaurant	Japanese Restaurant	Fast Food Restaurant	
4	Calderon (Carapungo)	Pizza Place	Chinese Restaurant	Park	Wings Joint	Farm	Food	Flea Market	Fire Station	

10 more common venues for each neighborhood.

Socio-economic Data

Import Socio-economic Excel data base that was provided by the ‘Instituto de la ciudad’

	Parroquias	Poblacion	Densidad	Proporción_Mujer	Promedio_edad	Promedio_per_hogar	Madres_Adolscentes	PEA
0	Alangasí	0	0.000000	0.000000	0.000000	0.000000	0.000000	0
1	Amaguaña	31106	44.363608	0.505079	28.307336	3.830000	0.032300	14158
2	Atahualpa	1901	15.200114	0.501841	34.276696	3.358657	0.044444	840
3	Belisario Quevedo	45370	8345.680223	0.527551	32.628675	3.100000	0.022811	24008

Divide the base in different segments or topics

Education

	Parroquias	Alfabetismo	Años_Escolaridad	Asistencia_basica	Asistencia_Bachiller	Asistencia_Superior
0	Alangasí	0.000000	0.000000	0.000000	0.000000	0.000000
1	Amaguaña	0.993509	9.110869	0.969914	0.845095	0.301080
2	Atahualpa	0.986971	7.467333	0.953039	0.722222	0.191045
3	Belisario Quevedo	0.997355	12.262085	0.979246	0.892808	0.454538
4	Calacalí	0.981928	8.243497	0.957659	0.792627	0.218310

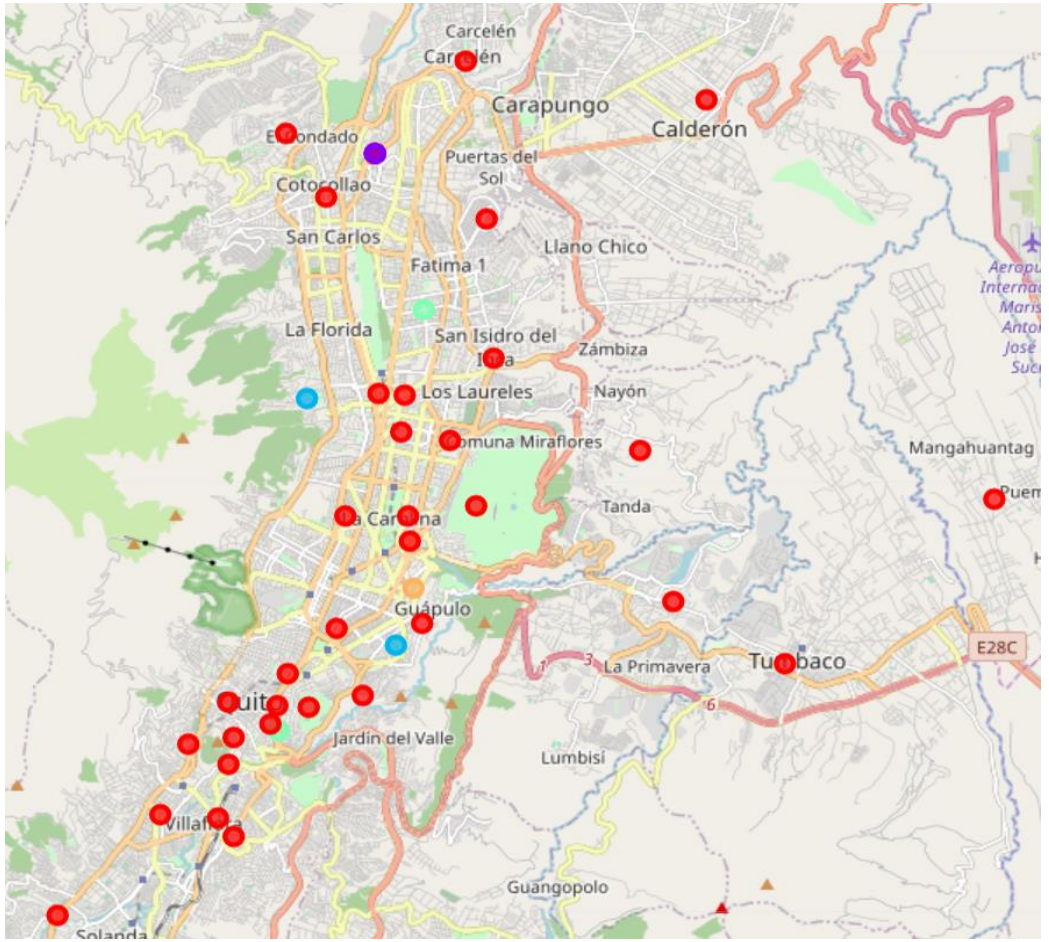
Living Conditions

	Parroquias	Vienda_Mal	Vivienda_Buen	Agua_potable	Viviendas_inadecuadas	Salubridad_inadecuadas
0	Alangasí	0.000000	0.000000	0.000000	0.000000	0.000000
1	Amaguaña	0.029371	0.424822	0.815523	0.085461	0.139845
2	Atahualpa	0.066071	0.276786	0.667857	0.132509	0.367491
3	Belisario Quevedo	0.007102	0.582747	0.945456	0.014964	0.033686
4	Calacalí	0.037500	0.322115	0.715385	0.054545	0.216268

K Means Cluster

K means cluster by venues

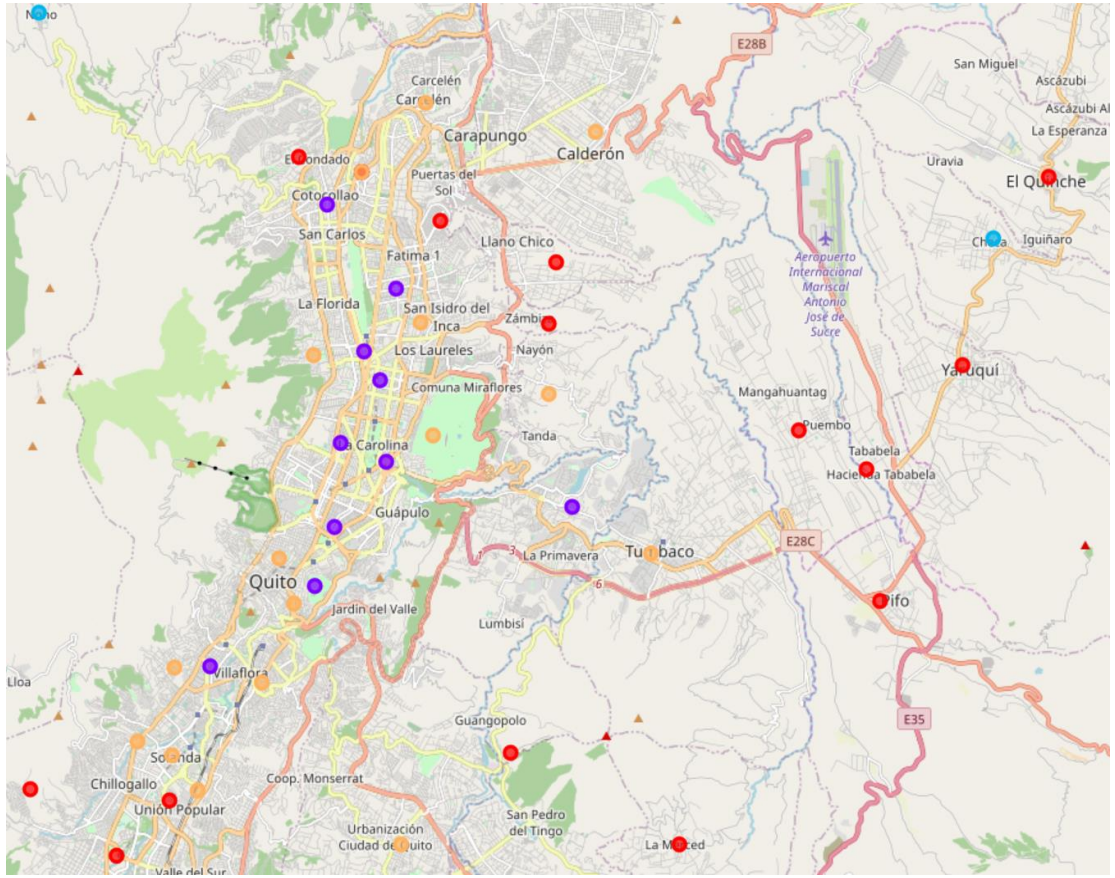
	Lugar	Latitud	Longitud	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Alangasi	-0.305050	-78.413440	0	Park	Soccer Field	Seafood Restaurant	Food & Drink Shop	Flea Market	Fire Station	Fast Food Restaurant	Farmers Market
1	Alfaro (Chimbacalle)	-0.233330	-78.516670	0	South American Restaurant	Science Museum	Comedy Club	Food & Drink Shop	Food	Flea Market	Fire Station	Fast Food Restaurant
2	Amaguaña	-0.380840	-78.515440	3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Belisario Quevedo	-0.108559	-78.487738	0	Bus Station	Soccer Stadium	Seafood Restaurant	Farmers Market	Food & Drink Shop	Food	Flea Market	Fire Station
4	Benalcazar	-0.182620	-78.481220	0	Italian Restaurant	Hotel	Bakery	Coffee Shop	Ice Cream Shop	French Restaurant	Japanese Restaurant	Fast Food Restaurant



K means cluster by Education Skills

	Parroquias	Latitud	Longitud	Cluster Labels	Alfabetismo	Años_Escolaridad	Asistencia_basica	Asistencia_Bachiller	Asistencia_Superior
0	Alangasí	-0.305050	-78.413440	3	0.000000	0.000000	0.000000	0.000000	0.000000
1	Amaguaña	-0.380840	-78.515440	0	0.993509	9.110869	0.969914	0.845095	0.301080
2	Atahualpa	0.167761	-78.360507	2	0.986971	7.467333	0.953039	0.722222	0.191045
3	Belisario Quevedo	-0.108559	-78.487738	1	0.997355	12.262085	0.979246	0.892808	0.454538
4	Calacalí	-0.001140	-78.513550	0	0.981928	8.243497	0.957659	0.792627	0.218310

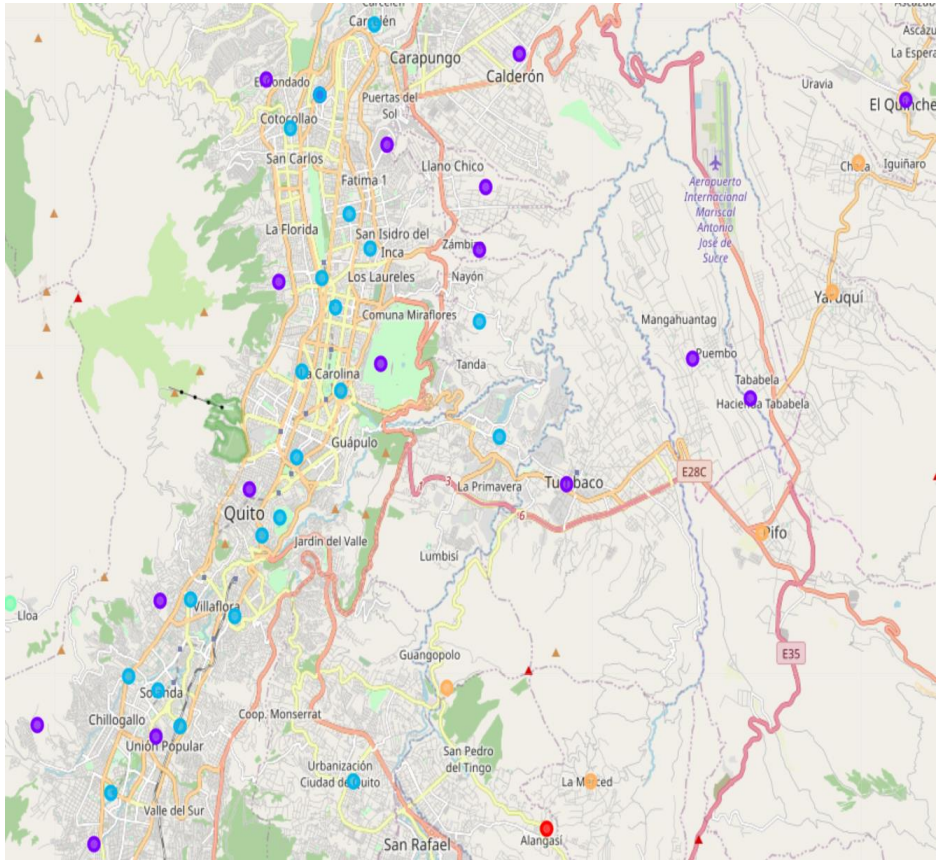
Number of Cluster	Education Level	Color in the map
1	High Level	Purple
4	Medium-High Level	Orange
0	Medium Level	Red
2	Low Level	Blue
3	Very Low Level	Green



K means cluster by Living Conditions

	Parroquias	Latitud	Longitud	Cluster Labels	Vienda_Mal	Vivienda_Buen	Agua_potable	Viviendas_inadecuadas	Salubridad_inadecuadas
0	Alangasí	-0.305050	-78.413440	0	0.000000	0.000000	0.000000	0.000000	0.000000
1	Amaguaña	-0.380840	-78.515440	4	0.029371	0.424822	0.815523	0.085461	0.139845
2	Atahualpa	0.167761	-78.360507	3	0.066071	0.276786	0.667857	0.132509	0.367491
3	Belisario Quevedo	-0.108559	-78.487738	2	0.007102	0.582747	0.945456	0.014964	0.033686
4	Calacalí	-0.001140	-78.513550	4	0.037500	0.322115	0.715385	0.054545	0.216268

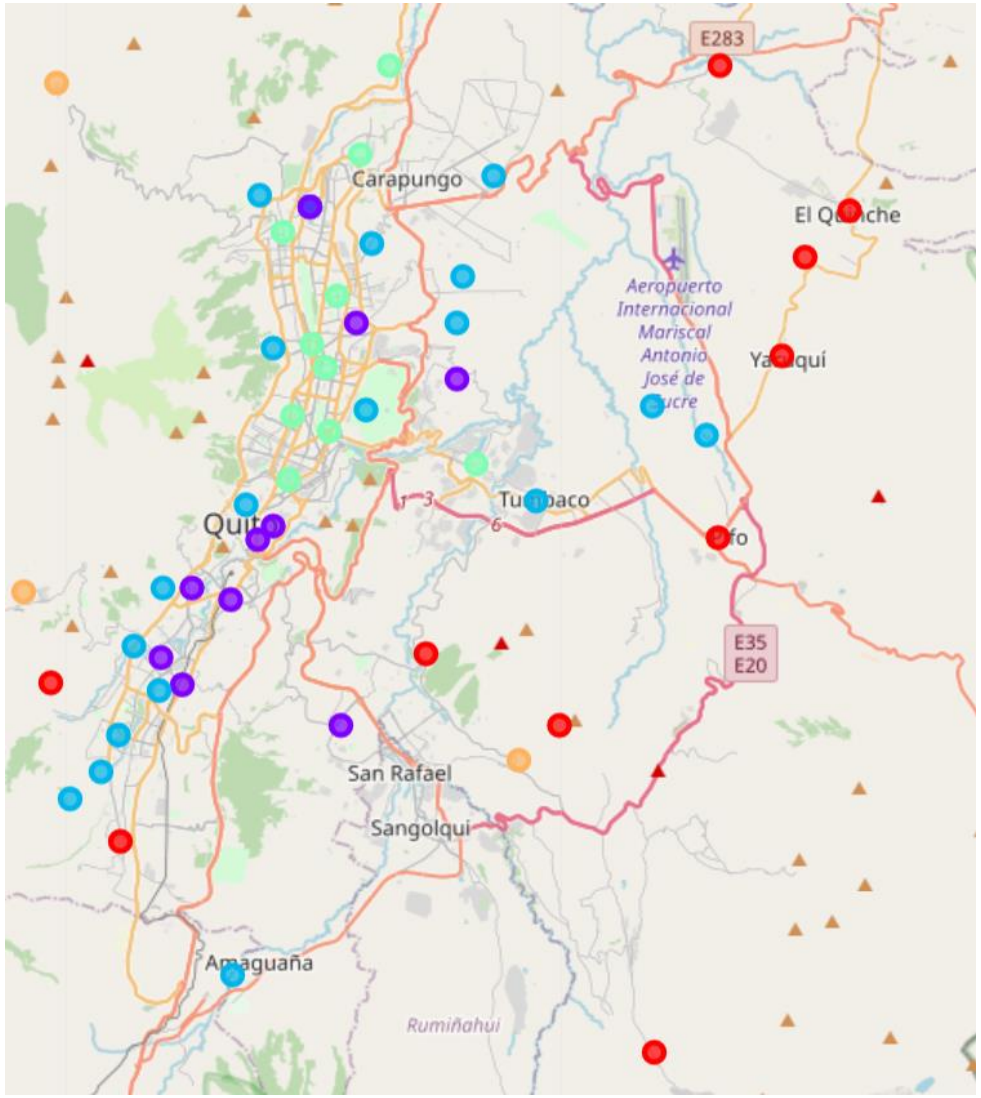
Number of Cluster	Living Conditions	Color in the map
2	High Conditions	Blue
1	Medium-High Conditions	Purple
4	Medium Conditions	Orange
3	Low Conditions	Green
0	Very Low Conditions	Red



K means cluster by Health Coverage

	Parroquias	Latitud	Longitud	Cluster Labels	Seguro_Privado	IESS	IESS_campesino	Jubilados
0	Alangasí	-0.305050	-78.413440	4	0.000000	0.000000	0.000000	0.000000
1	Amaguaña	-0.380840	-78.515440	2	0.047515	0.436346	0.001245	0.302326
2	Atahualpa	0.167761	-78.360507	4	0.026302	0.172121	0.344242	0.147368
3	Belisario Quevedo	-0.108559	-78.487738	1	0.128477	0.530531	0.001318	0.390399
4	Calacalí	-0.001140	-78.513550	0	0.046213	0.304348	0.006293	0.113772

Number of Cluster	Health Coverage	Color in the map
2	High Coverage	Blue
1	Medium-High Coverage	Purple
4	Medium Coverage	Orange
3	Low Coverage	Green
4	Very Low Coverage	Red



Conclusions

The study found that the districts that are more isolated from the center of the metropolitan city have less educations skills, living conditions and health coverage. Most of this neighborhood are in the rural or urban-rural area.

It is very alarming the fact that in most of this remote areas people are living without appropriate conditions. They do not have drinking and clean water; good waste manages or structurally safe houses.

The municipality could focus and redirect the budget of every year to develop more efficient projects