Data Analysis for a Market penetration strategy in Peru for a mental healthcare company

Juan de Dios Huarancca Cahuana

Nov 2019

Data Acquisition

GEODIR - Ubigeo de Perú (INEI)

A web page where I found information about, departments, provinces and districts of Peru and their surface in kilometers and their population.

Foursquare API

I have to create an account to use this platform from I got information about venues of departments in Peru.
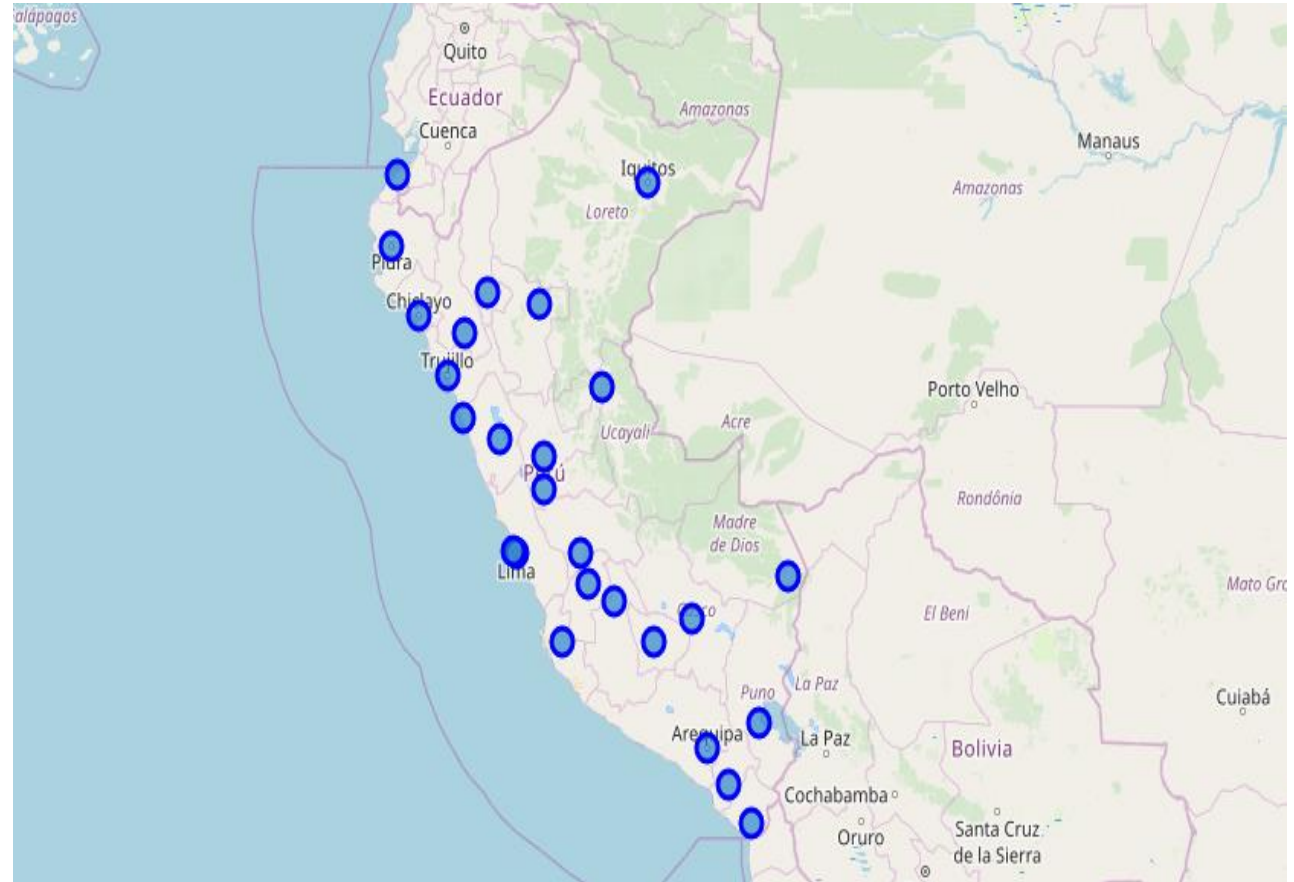
Open data portal of Peru

Statistics about mental health attention in Peru in 2017. These information is provided by Health Ministry of Peru.

Methodology
Cluster departments of Peru

- Information from geodir-ubigeo-department.

- I used Google Map, 'Search Nearby' option, to get the latitude and longitude of the each departments

- I used folium library for python to visualized geographic details of Peru and Its departments

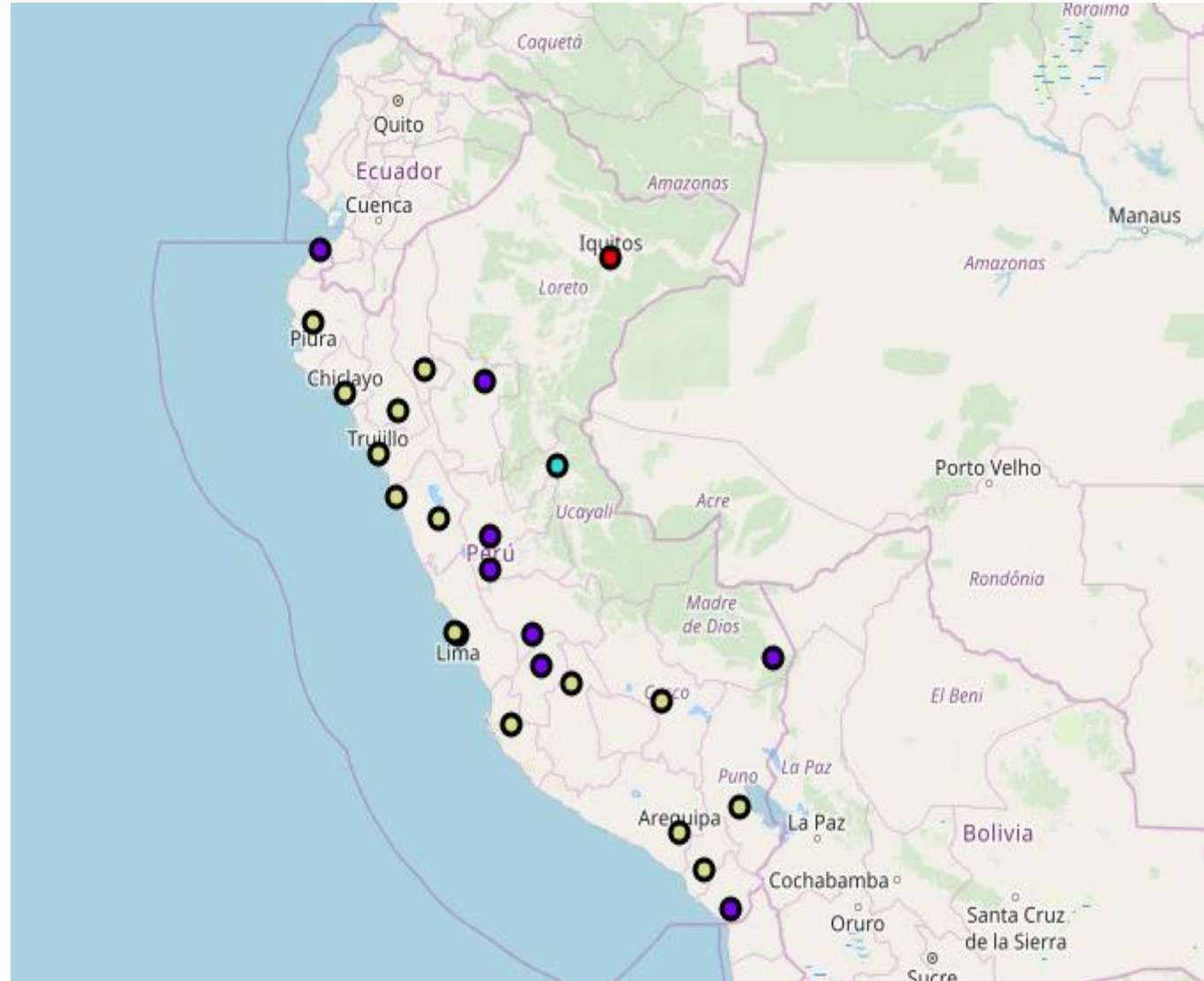|	Borough	Neighborhood	Latitude	Longitude
0	2534	Amazonas	-6.231690	-77.869034
1	2625	Ancash - Huaraz	-9.526000	-77.529000
2	2812	Ancash - Chimbote	-9.063000	-78.589000
3	2900	Apurimac	-14.050000	-73.088000
4	3020	Arequipa	-16.398867	-71.536961
5	3143	Ayacucho	-13.160427	-74.225697

Methodology
Cluster departments of Peru

- Using K-Means algorithm, an unsupervised learning, we cluster departments using the common venues that departments shared. I will run K-Means to cluster departments into 4 clusters.

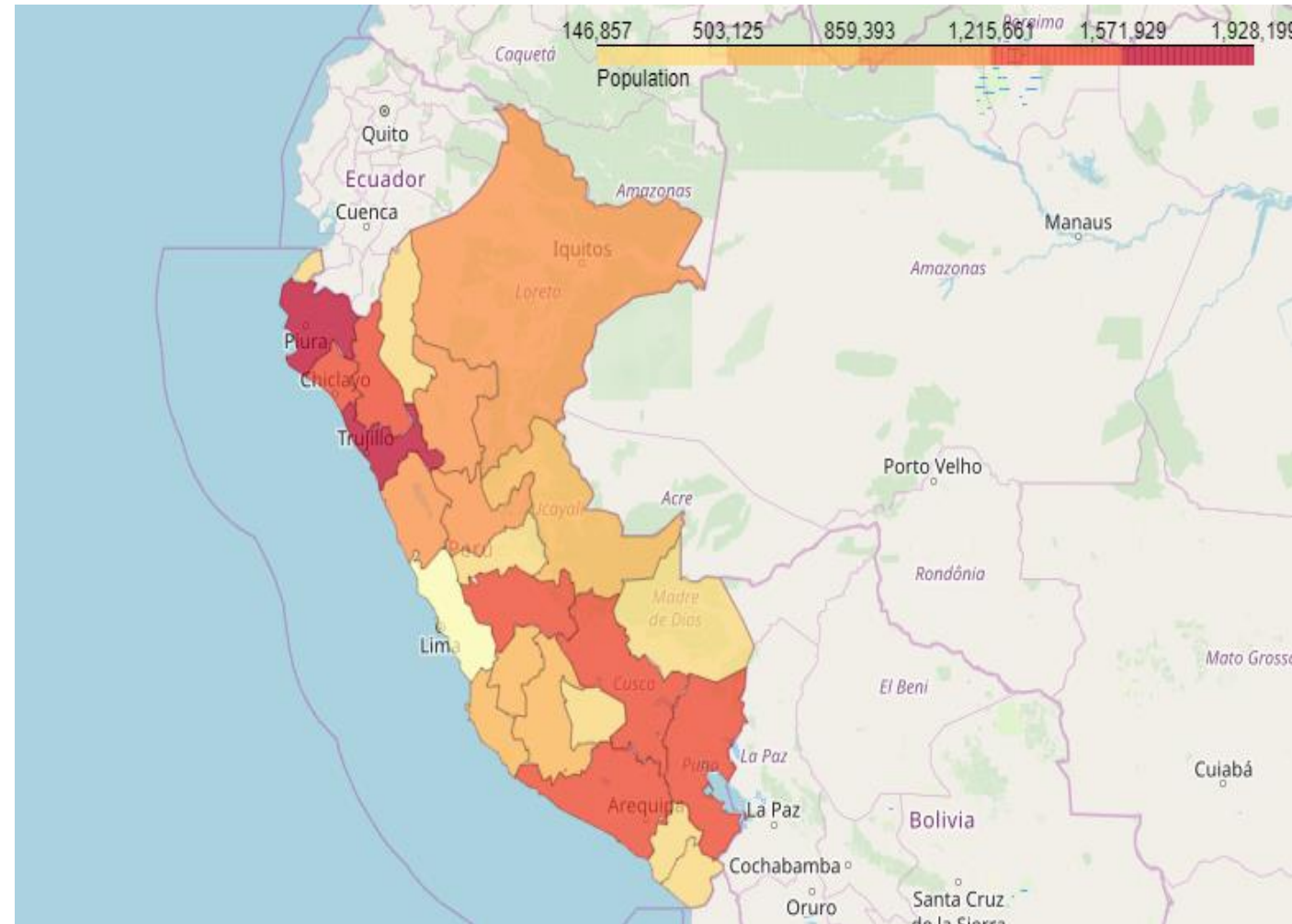|	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	2534	Amazonas	-6.231690	-77.869034	3	Café	Park	Peruvian Restaurant	Plaza	Bar	Bed & Breakfast	Restaurant
1	2625	Ancash - Huaraz	-9.526000	-77.529000	3	Pizza Place	Restaurant	Café	Bar	Hotel	Mountain	Farmers Market
2	2812	Ancash - Chimbote	-9.063000	-78.589000	3	Zoo	Breakfast Spot	Harbor / Marina	Outdoors & Recreation	Seafood Restaurant	Himalayan Restaurant	Bakery
4	3020	Arequipa	-16.398867	-71.536961	3	Hotel	Peruvian Restaurant	Bed & Breakfast	Sandwich Place	Bar	Italian Restaurant	Latin American Restaurant
5	3143	Ayacucho	-13.160427	-74.225697	3	Restaurant	Plaza	Shopping Mall	Flea Market	Market	Breakfast Spot	Pizza Place
6	3292	Cajamarca	-7.162000	-78.513000	3	Plaza	Coffee Shop	Pizza Place	Fried Chicken Joint	Steakhouse	Japanese Restaurant	Bus Station
7	3414	Cusco	-13.522839	-71.967627	3	Plaza	Hotel	Market	Soccer Stadium	Bed & Breakfast	Pizza Place	Seafood Restaurant
8	3518	Huancavelica	-12.787280	-74.973095	1	Park	Hotel	BBQ Joint	Plaza	Electronics Store	Concert Hall	Convenience Store
9	3606	Huanuco	-9.929325	-76.239485	1	Hotel	Pizza Place	Coffee Shop	Gym	Plaza	Chinese Restaurant	Fish & Chips Shop

Methodology
Data Analysis of Peru's population

- Data from "GEODIR - Ubigeo of Peru" and I loaded data into a data frame, I got data about population and surface in kilometers. I did not consider information of Lima department because I analyzed departments where the company does not operate.

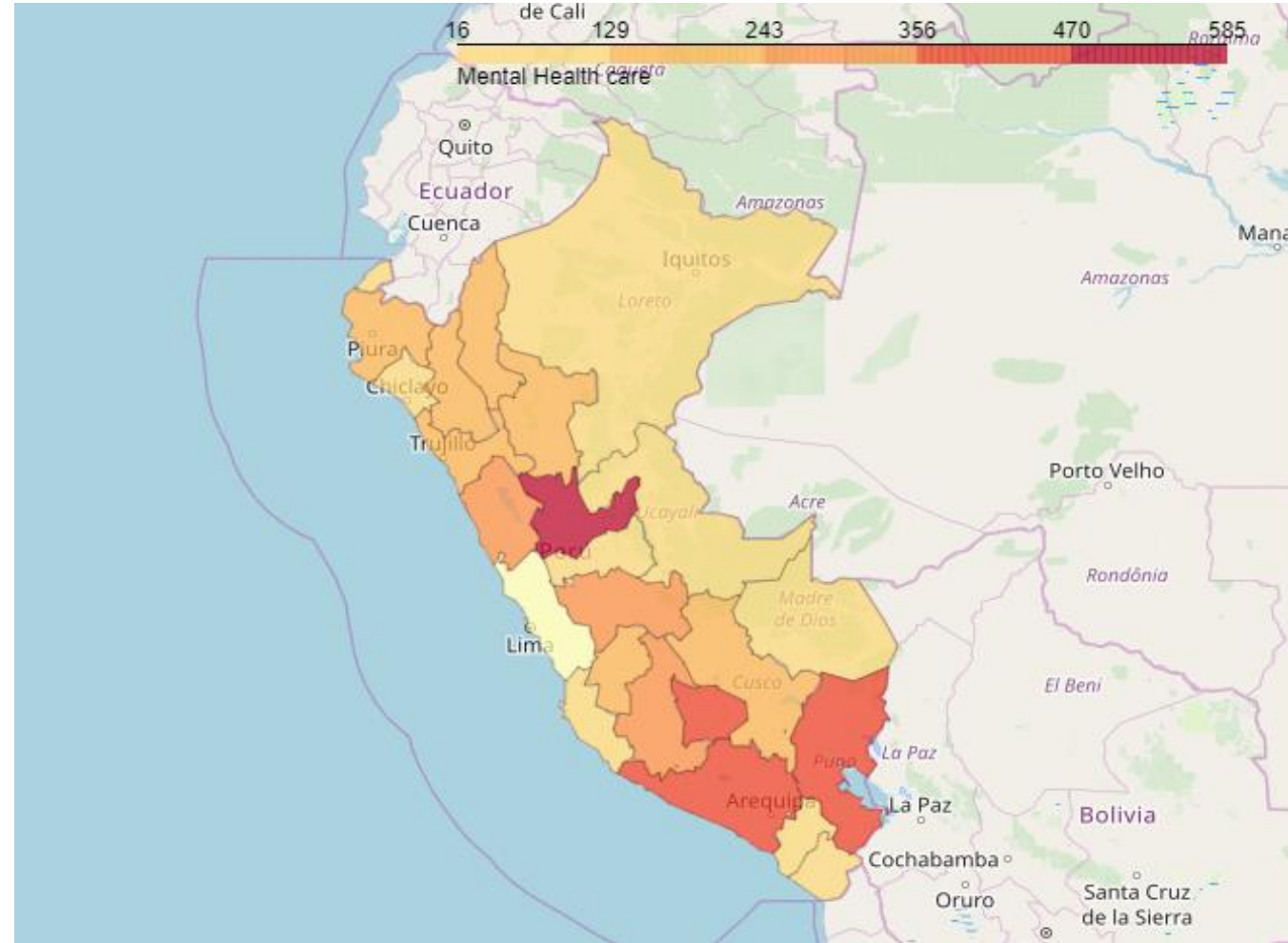|	Departamento	Ubigeo	Poblacion	Superficie	Y	X	PoblacionxKm2
0	AMAZONAS	872418	425832.0	39249.13	-513.09100	-6546.49210	10.849463
1	ANCASH	3500411	1166180.0	35881.30	-1547.78840	-12878.17160	32.501052
2	APURIMAC	2553828	464583.0	21037.99	-1169.15150	-6139.68930	22.083051
3	AREQUIPA	4402800	1329801.0	63343.95	-1740.19090	-7886.44640	20.993339
4	AYACUCHO	6022013	711058.0	44357.00	-1653.16360	-8804.64560	16.030345
5	CAJAMARCA	7705419	1540009.0	33304.32	-840.59490	-9994.51770	46.240518
7	CUSCO	9041166	1338897.0	82694.66	-1535.97104	-8051.65712	16.190852
8	HUANCAVELICA	9042953	505495.0	22159.22	-1286.26270	-7491.87530	22.811949
9	HUANUCO	8446193	878196.0	37867.36	-807.24810	-6417.32250	23.191371
10	ICA	4741337	810214.0	21305.51	-600.35930	-3256.05840	38.028379
11	JUNIN	14932491	1379934.0	44328.80	-1456.73560	-9347.29340	31.129514
12	LA LIBERTAD	10841207	1928198.0	25495.42	-654.32250	-6509.57630	75.629191
13	LAMBAYEQUE	5327109	1290617.0	14461.52	-252.23490	-3029.61160	89.244906
15	LORETO	8502149	1068130.0	368799.48	-243.87230	-3937.19010	2.896235

Methodology
Data Analysis of Peru's population

- Data from statistics about mental health attention in Peru from Peruvian open data portal. This data was delivered by Health ministry of Peru. Then I cleaned and prepared the data adding latitude and longitude of departments.

- I used folium library to show the departments where Public Health Institution attends mental health patient.

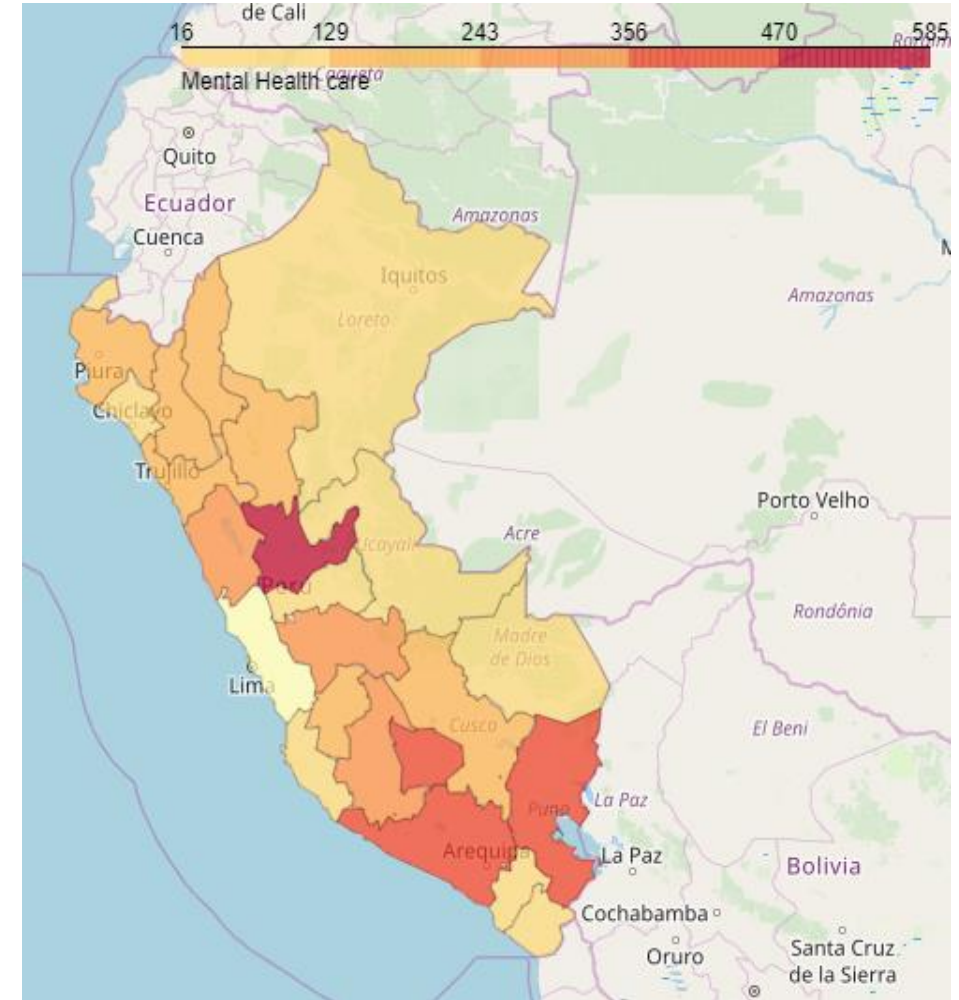|	Departamento	Atendidos	Atenciones	Poblacion	Superficie	Y	X	PoblacionxKm2	AtencionXPoblacion
0	AMAZONAS	114.0	148.0	425832.0	39249.13	-513.09100	-6546.49210	10.849463	3.475549
1	ANCASH	219.0	283.0	1166180.0	35881.30	-1547.78840	-12878.17160	32.501052	2.426727
2	APURIMAC	322.0	391.0	464583.0	21037.99	-1169.15150	-6139.68930	22.083051	8.416150
3	AREQUIPA	279.0	378.0	1329801.0	63343.95	-1740.19090	-7886.44640	20.993339	2.842531
4	AYACUCHO	206.0	306.0	711058.0	44357.00	-1653.16360	-8804.64560	16.030345	4.303446
5	CAJAMARCA	163.0	220.0	1540009.0	33304.32	-840.59490	-9994.51770	46.240518	1.428563
7	CUSCO	163.0	203.0	1338897.0	82694.66	-1535.97104	-8051.65712	16.190852	1.516173
8	HUANCAVELICA	143.0	192.0	505495.0	22159.22	-1286.26270	-7491.87530	22.811949	3.798257
9	HUANUCO	498.0	584.0	878196.0	37867.36	-807.24810	-6417.32250	23.191371	6.649996

Results

- Clustering analysis show departments that share same characteristic according venues they have, Lima is in the cluster number 3 with these departments.

- According the population analysis, it does not include Lima, the departments of Piura and Trujillo are the more populate departments in Peru and these two department are in the cluster where is Lima

Results

- Finally, analyzing the mental health care data reported by Health Ministry of Peru, the departments of the north of Peru (Piura, Tumbes, Lambayeque y Trujillo) does not have attention so It show that is a good opportunity to open a branch in these area.

Conclusion

- The north zone of Peru has a good place to install a new business, including the departments of Trujillo, Lambayeque, Piura and Tumbes.

- I recommend to collect data about private mental health institution in the north of Peru to better the analysis.

References

Demografía del Perú
https://es.wikipedia.org/wiki/Demograf%C3%ADa_del_Per%C3%BA

GEODIR - Ubigeo of Peru (INEI)
http://www.geodir.co/recursos/ubigeo-inei-peru.html

Foursquare API
https://developer.foursquare.com/

GITHUB – GeoJson Files about Peru.
 https://github.com/juaneladio/peru-geojson

Peru's Open data portal – statistics about mental health attention in Peru: 2017.
https://www.datosabiertos.gob.pe/dataset/minsa-salud-mental

Internal Gross Product, According to departments: 2018
https://www.inei.gob.pe/media/MenuRecursivo/boletines/pbi_departamental2018.pdf

Juan de Dios Huarancca's Github repository
https://github.com/jhuarancca/Coursera_Capstone/tree/master/FinalCapstone