IBM DATA SCIENCE CAPTONE PROJECT
Javier Alejandro Rendon Carrillo

HOW TO CHOOSE THE BEST LOCATION FOR YOUR MEDICAL PRACTICE

Table of contents

1. Introduction

When it comes to purchasing a home or investment property, it's all about **location, location, location**. The same rule applies when you're looking to buy or rent a space for your medical practice. According to a July 2014 report by The Associated Press-NORC Center for Public Affairs Research, 50 percent of patients consider the location of medical practice when choosing a doctor. Another report similarly found that 70 percent of healthcare consumers deem location either critical or very important when selecting a provider or healthcare system.

2. Objective

The aim of this project is find an optimal location for a Medical practice. For most of Healthcare Professionals (HPC) is it always a headache find a suitable place to open practice or even change the actual facility location. Scope of this project is help HCPs which Neighborhoods in Queretaro city, México are best suited for their business. Location of the facility will have a significant impact on the practice outcome, (adjacent and nearby shops and offices) play a very important role in building a positive early impression of the clinic. We will use our data science techniques to detect the most promising neighborhoods based on criteria selected in background section. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

3. Background and significance¶

HEALTHCARE SYSTEM IN MEXICO¶

Mexico has achieved universal health coverage and its public healthcare is acceptable for most Mexican residents. Despite this, the private healthcare sector has grown considerably and is driven by increasing disposable income, the growth of medical tourism, and ease of access to higher quality private healthcare services. Mexico's public healthcare operates through the Instituto Mexicano de Seguro Social (IMSS) and Seguro Popular systems. These cover patients for most medical services and prescription drugs. Those employed in Mexico are automatically enrolled in the IMSS system and their contribution to the scheme is deducted from their salary. Those who are not formally employed may voluntarily enroll in the IMSS system, in which case they will have to pay an annual contribution fee. People who cannot afford the IMSS system must enroll with the Seguro Popular system. Fees for the Seguro Popular system are charged on a sliding scale depending on a resident's income. While public healthcare in Mexico is relatively good, the quality of services varies between hospitals. Most mexicans above mid income opt for private health care, which they finance through private health insurance. Although private hospitals are more expensive, they are better equipped, provide greater access to specialized procedures and generally provide higher quality care.

HOW PATIENTS CHOOSE THEIR PRACTITIONER¶

Since there are lots of HCPs in Queretaro we will try to detect locations that fulfill the next 5 points:

---

1. Demographics

We need to define our demographic data such as population age, net income, education. We also have to consider whether or not the population is growing or

declining, age is a demographic trait that can have several financial impacts on the doctor's office.– it is usually easier to break into newer communities than mature communities where you would have to take patients away from practitioners who have been in the area many years. Another aspect to take into account is the economic level of the population. An area with a high rate of low-income residents will likely have more patients going to social security than doctors in the private sector.

---

2. Accessibility

How Long and How Far Do Adults Travel and Will Adults Travel for Primary Care? Accordingly to Washington State Health Services adults are willing spend 28.4 minutes and travel a distance of 32 kilometers (1). The location choose for the medical practice must be accessible and convenient for patients. For example, a good rule of thumb is to choose a location within 20 minutes of the residential area you hope to serve.

When comparing locations, consider the availability and amount of parking. Free parking is always preferable. And aim for a location with a spacious entryway where elderly, injured or disabled patients can be dropped off and picked up without difficulty.

Choice of Physician: Consumers who have selected a physician in the past three years are more concerned about convenient location (62%); friendly office staff (56%); than success rates (22%).

Choice of Hospital: Consumers who have selected a hospital in the past three years are more likely to choose a hospital based on convenient location.

Source: 2014 Healthgrades American Hospital Quality Report to the Nation

3. Competition

We need to determine how many providers are in the area, how big their practices are, and what their specialties are. Finding a space that's well-known as site for medical practitioners can work to our advantage since people are accustomed to traveling there. "It's a lot easier to tap into an existing behavior than to create a behavior all by itself."

4. Visibility

A location in a remote part of town might seem cost-effective, but having low visibility will mean spending more money on marketing to get patients in the door. "Think about marketing costs as part of the rent equation." A medical office that's located on a major road or thoroughfare, or in a busy shopping center, can give you maximum visibility.

5. Nearby Hospitals, Pharmacies and other business

Speaking of proximity to other businesses, medical practices benefit from operating close to places such as: -Pharmacies & drug stores -Hospitals -Urgent care centers -Fitness centers Beyond the obvious convenience of locating close to the hospital, the clinic also will benefit from the patient perception that is located in a recognized healthcare area. Let's think where are popular businesses, such as supermarkets and banks? The more popular businesses attract more potential clients. Also, upscale businesses attract upscale clients – think Starbucks.

In order to identify success keys factors or attractors like anchor stores that have a large influx of people, such as banks, shopping malls, etc. ; versus repulsors like parking absence or low visibility we need to plot them on a map; this map will contain

data associated with the demand, offer, attractors and repulsors, with this information proceeds to define the most optimal proposal (Figure 1).

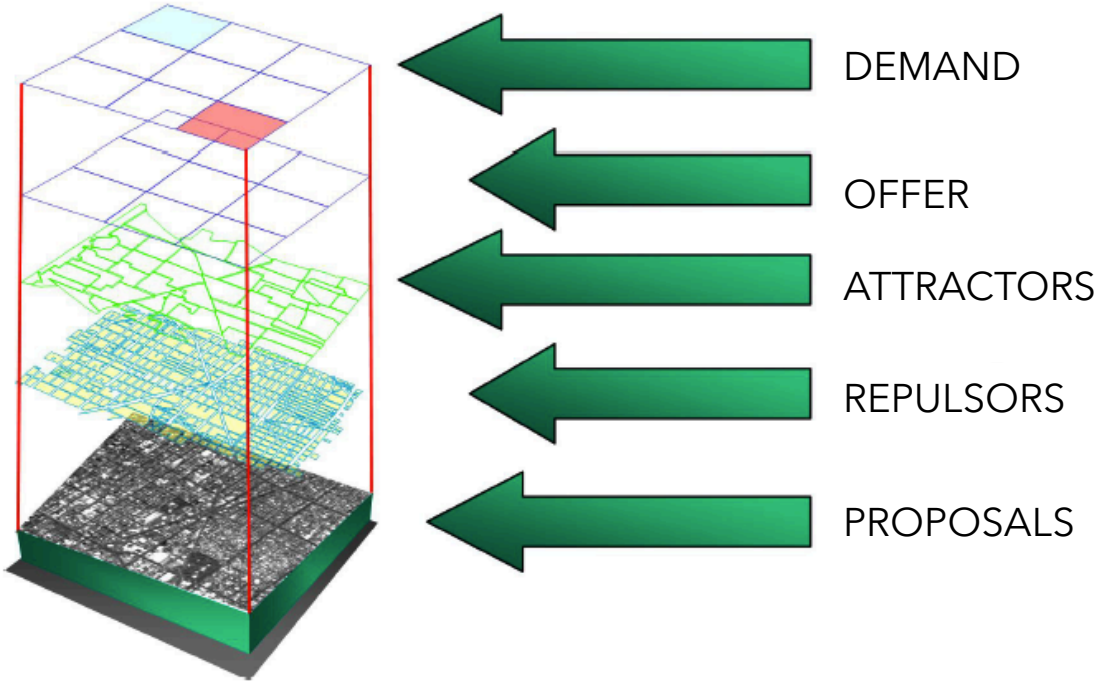

Fig. 1: Procedure to identify optimal proposal for medical practice.

NEIGHBORHOODS AND ACCESS TO HEALTH CARE¶

Studies have shown a strong relationship between neighborhood of residence and a number of health outcomes including low birth weight and infant mortality (Buka et al., 2003; O'campo et al., 1997; Szwarcwald et al., 2002), cardiovascular disease and other chronic conditions including coronary heart disease (Sundquist et al., 2004; Diez-Roux et al., 1997), stress, and depression (Matheson et al., 2006; Boardman et al., 2001). In addition, neighbourhood contextual characteristics have also been

shown to influence health related behaviors, such as smoking, alcohol consumption, diet, and physical activity. The availability of, and access to, primary health care is one neighbourhood characteristic that has the ability to directly impact health. Research has demonstrated that increased distance to health care services results in reduced utilization of the healthcare system, and increase area-based inequities in health status (Hiscock et al., 2008; Kordaet al., 2007; Haynes, 2003).

When conducting neighbourhood-level research, the choice of units of analysis is of utmost importance. In a recent review,DeMarco and De Marco (2010) identify a range of methods used for measuring neighbourhoods including the use of administrative units (e.g., census tracts, block groups), the use of specified distances (e.g., radius from homes) and the use of resident-defined boundaries. Much of the research uses census based units (e.g., census tracts) to represent neighbourhoods (Wang and Luo,2005; Wang, 2007; Pearce et al., 2006; Guagliardo et al., 2004).

Zonation effect, relates to how empirical results are dependent upon where area boundaries are drawn. A shift in the location of boundaries can easily cause results to change between positive and negative in terms of health outcomes or service availability, depending on whether boundaries include or exclude data.

4. Data

Based on definition of our problem, factors that will influence our decision are:

Demographics: Age and income. Accessibility: Radius of 15 km from city center. Competition: Number of proximity clinics. Visibility: Proximity to principal avenues and from city center Nearby Hospitals, Pharmacies and other business: Hospitals, pharmacies, restaurants, coffee shops, will be taken in account.

Neigbourhoods segmentation by postal code will be inaccurate for statical analysis then we use AGEB. According to the definition of INEGI an Área Geoestadística Básica (AGEB) - Subdivision of the municipalities or delegations that make up the country, which allows the formation of primary sampling units and the organization of statistical information. It has three fundamental attributes: a) Perfectly recognizable on the ground, as it is delimited by identifiable and durable topographic features. b) Generally homogeneous in terms of its geographical, economic, and social characteristics. c) Its extension is such that it can be covered walking by a single person. AGEBs are classified as more urbanized and less urbanized. A more urbanized AGEB can vary in size between 20 and 80 blocks.

Following data sources will be needed to extract/generate the required information: -Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using https://github.com/marioalbertodev/colonias-queretaro/blob/master/colonias.json -Demographics will be obtained using INEGI (National Institute of Statistics and Geography) https://www.inegi.org.mx/app/indicadores/?t=0200&ag=22#D02000070. https://www.inegi.org.mx/programas/enigh/nc/2018/default.html. -Number of Practices, Hospitals, pharmacies, restaurants, coffee shops, and their type and location in every neighborhood will be obtained using Foursquare API.

5. Design Research and Data processing

Libraries

List of python libraries imported

Library	Reason
Folium	Mapping library, to create accurate map visualisations
Pandas	Data Analysis library, to create DataFrames and process data.
Matplotlib	Plotting library, to show visualisations of the data.
Seaborn	Plotting library, works on top of Matplotlib.
urllib.request	URL request library, to make web requests.
bs4	BeautifulSoup library, to scrape the data from web requests.
geocoder	GeoCoder library, to get the latitude and longitude using an address.
Numpy	Scientific library, we will be using it for Arrays.
JSON	JSON library, used to load and parse JSON files.
Requests	Web Requests library, to access websites and parse HTML
Pygeodesy	Geo Point library, provides functions for latitude and longitude calculations
Scikit-Learn	Machine Learning library, provides functions we will use for Clustering and Normalising data
Scipy	Maths library, provides math algorithms for statistics
osmnx	Spatial geometries library, model, project, visualize, and analyze real-world street networks from OpenStreetMap's APIs.
kneed	Elbow method calculation library, provides math algorithms for calculate elbow

5.1.Download and Explore Dataset

Data was downloaded from the following URLs

I. https://raw.githubusercontent.com/Alexrendon/Capstone-Project-Notebook/master/Queretaro_MedicalCenters.csv
II. https://github.com/marioalbertodev/colonias-queretaro/blob/master/colonias.json
III. https://www.inegi.org.mx/app/indicadores/?t=0200&ag=22#D02000070
IV. https://www.inegi.org.mx/programas/enigh/nc/2018/default.html
V. https://raw.githubusercontent.com/Alexrendon/Capstone-Project-Notebook/master/Queretaro_Coordinates.csv
VI. https://raw.githubusercontent.com/Alexrendon/Capstone-Project-Notebook/master/AGEB_DATA.csv
VII. https://raw.githubusercontent.com/Alexrendon/Capstone-Project-Notebook/master/SOCIOECONOMIC%20INDEX.csv
VIII. https://github.com/Alexrendon/Capstone-Project-Notebook/blob/master/qro_ageb_urb.geojson

5.2. Data Cleansing

Data was preprocessed, duplicated information was deleted as outliers, missing values were completed, and typography errors were corrected, some health-centers have branches, for them the name of main avenue from their location was added to their name.

5.3. Create a map of Queretaro Health Centers

Queretaro city was used for this analysis. The geographical coordinates of Queretaro are 20.5954708, -100.3970593. Using Folium we create a map, Figure 2 shows a Choroplet map of the number of health entitled population in each AGEB, and the

location of each health center, as we can see majority of the Health Centers are clustered in the city center.

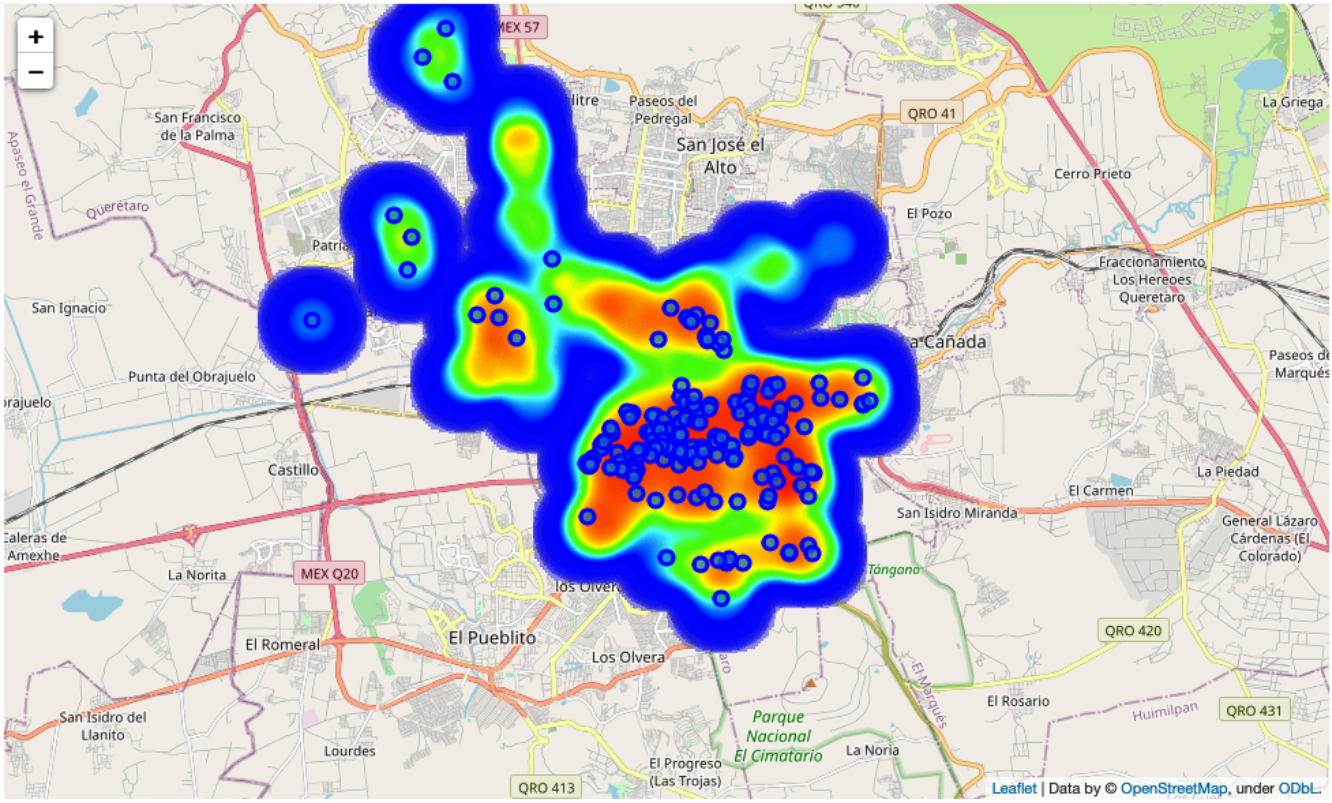

Fig. 2: Heath map of Querétaro City Health centers, each blue dot is a unique health-center.

5.4.Explore Querétaro with Foursquare API

For our venue data we used the Foursquare API. This allows us to make calls to the API to search/explore a region depending on the latitude and longitude, categories, radius and limit when calling the API.

As we wanted to use multiple categories when searching the decision was to use the 'search' API endpoint, 'explore' will only allow a single category parameter when calling the API. Radius was set to 500 meters and limit was 50 venues around the point of interest.

8626 venues were returned by Foursquare with 231 uniques categories, most frequently occurring venues within 500 meters of Health Centers are shown in figure 3, Restaurants are the most common venue followed by coffee shops, bars, hotels, pharmacies and fitness centers.

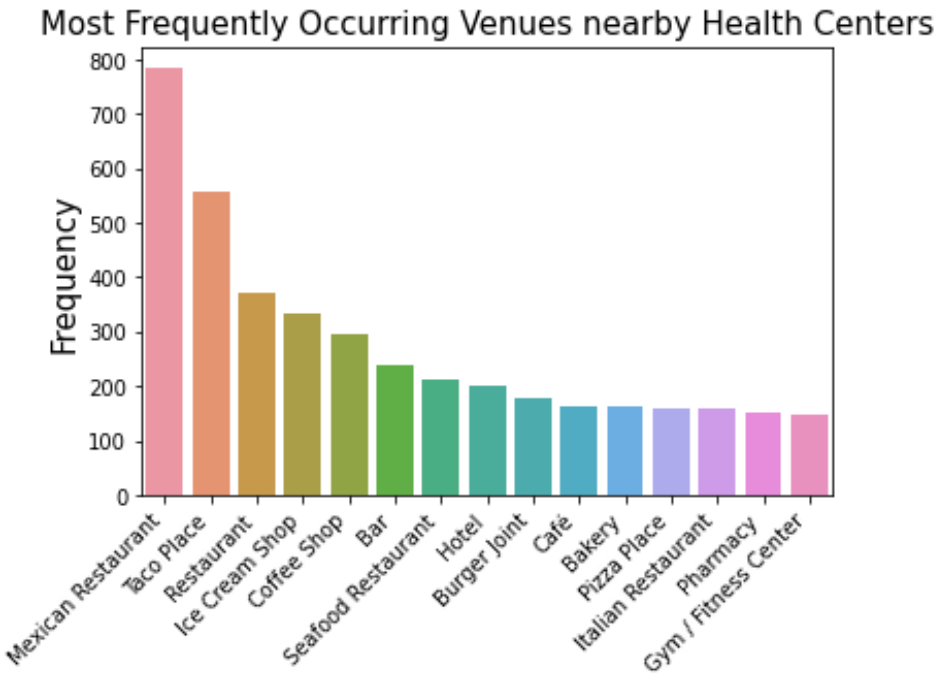

Fig. 3 : Most frequent venues nearby Health centers

6. Methodology

Now that we have gathered the data from scraping, and from our geojson file we can start to analyze the results.

Data gathered in the Data section includes:

- Medical Centers
- INEGI AGEBs
- DataFrame for All Venues within each Health Center
- DataFrame for AGEBS within each Health Centers

- DataFrame for Socioeconomic Indicators within each AGEBs
- DataFrame for Health Centers within each Neighbourhood

We also visualized data retrieved in the form of Choropleth Maps, and then added a Heath map to so we could see if there are any obvious clusters.

Our goal is to find the a suitable location for a new Medical Practice in a location where there are minimal competition and a populous neighbourhood.

In the Analysis section we will use the following methods:

Clustering using K-Means our Health Centers using the data we have for all venues within each AGEB. This will help us to find the most populous AGEBs and the socioeconomic data within each Neighbourhood, we will then use this clusters to find appropriate locations.

Matrix Plot using to show the correlation between venues.

DBSCAN Clustering will be used to compare clusters against K-means clusters.

The final product of the research will be to provide a couple of address within the most populous, accesible and visible AGEBS, Nearby Hospitals, Pharmacies and other business, and medium competence. The address will be located in the centroid position of the most promising clusters for each Neigbourhood or AGEB accordingly to results.

7. Analysis

In this section we hope to use our analysis to show suitable locations for Medical Practices in Queretaro City.

To start we will asses Nearby business venues distribution and find if follows a trend, then we will cluster the Health Centers into similar clusters using by K-Means.

After that we will see if there is a correlation between the frequency of a kind of venue and health center. From socioeconomic data and demographics we get insights and data to determine optimal AGEBs and optimal market for the medical practice.
Following on from that using DBSCAN we will find the cluster distribution on health centers by network spatial based clustering. This will be used later to find the ideal location for the Medical practice by analyzing accessibility and visibility.
Lastly the final analysis will be to find the ideal locations. This will get us a centralized latitude and longitude for the locations

7.1.Analyze Each Health Center

One-Hot-Encoding will be used to normalize the categories from Foursquare API, 231 uniques categories were found on the searching step. Normalizing the category field makes it easier to count the frequency of categories.

	Healthcenter	ATM	Accessories Store	American Restaurant	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	A
0	Laboratorio Corregidora	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Laboratorio Corregidora	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Laboratorio Corregidora	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Laboratorio Corregidora	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Laboratorio Corregidora	0	0	0	0	0	0	0	0	0	0	0	0	0	

We group in descending order the venues found and create a table that shows the top 10 categories per venue, data is show in notebook.

click to expand output; double click to hide output

	Healthcenter	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	9th Most Common Venue	10th Most Common Venue
0	A&K Medical	Restaurant	Pharmacy	Mexican Restaurant	Pizza Place	Cosmetics Shop	Convenience Store	Bus Station	Bar	German Restaurant	Coffee Shop
1	ADMINISTRACIÓN MEDICA INDUSTRIAL	Mexican Restaurant	Taco Place	Park	Hot Dog Joint	Burger Joint	Sandwich Place	Restaurant	Pizza Place	Convenience Store	Food & Drink Shop
2	ALFADENT	Coffee Shop	Restaurant	Taco Place	Pizza Place	Other Great Outdoors	Pool Hall	Dance Studio	Recreation Center	Café	Sushi Restaurant
3	Alfadent	Coffee Shop	Mexican Restaurant	Gym	Restaurant	Bakery	Ice Cream Shop	Taco Place	Salad Place	Italian Restaurant	Burger Joint
4	Aquaskin	Ice Cream Shop	Restaurant	Clothing Store	Hot Dog Joint	Donut Shop	Bubble Tea Shop	Sausage Shop	Burger Joint	Salad Place	Coffee Shop

7.2.Cluster Health Centers by K-Means

Health Centers were clustered by venues found in the surroundings of each of them, K-means was the algorithm used to process the data. To determine the optimal value of K for our dataset we used Elbow method and Silhouette Coefficient as shown in figure 4 and 5 .

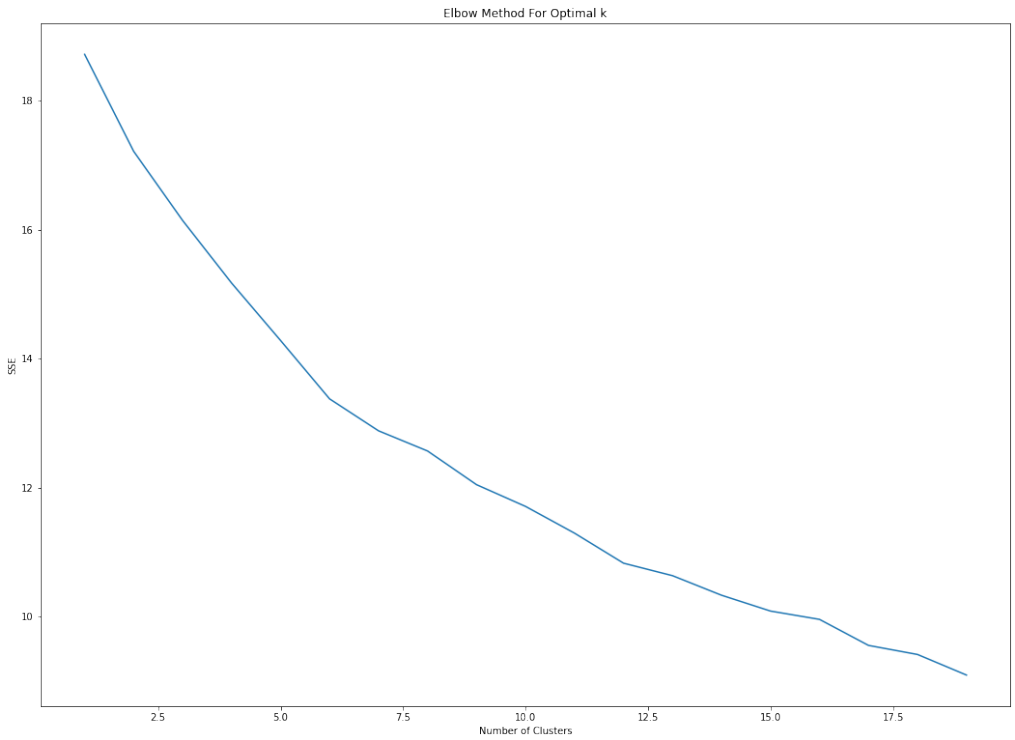

Fig. 4: Elbow method for determine K value.

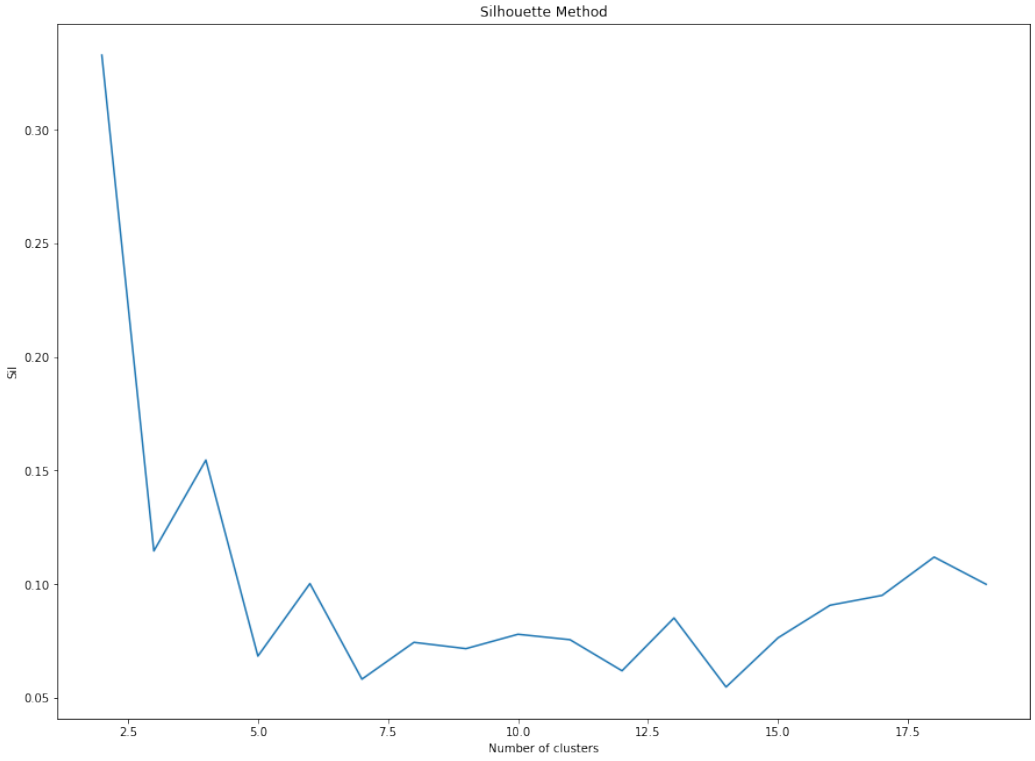

Fig. 5: Silhouette method for determine K value.

After getting our K-Elbow diagram, we can see that there is not a huge elbow in either the distortion or inertia methods.

As K-Means works by selecting a random centroid for each cluster, if you repeat the functions it will change every time. After running our function several times the most consistent K when we have calculated on numerous occasions.

7.3.Cluster Health Centers and most common venues

After performing our K-Means function on our data using a K of 4, Health centers were clustered and labeled for further analysis, as show in table 1.

Table 1: Cluster Venues

CLUSTER 0		CLUSTER 1		CLUSTER 2		CLUSTER 3	
Venue kind	Total venues	Venue kind	Total venues	Venue kind	Total venues	Venue kind	Total venues
Mexican Restaurant	37	Mexican Restaurant	59	Taco Place	38	Soccer Field	1
Restaurant	19	Hotel	2	Convenience Store	5		
Coffee Shop	17	Seafood Restaurant	1	Shopping Mall	2		
Ice Cream Shop	14	Flower Shop	1	Seafood Restaurant	1		
Clothing Store	14	Soccer Stadium	1	Pharmacy	1		
Taco Place	8	BBQ Joint	1				
Burger Joint	6	Taco Place	1				
Café	6	Park	1				
Nightclub	5	Movie Theater	1				
Pizza Place	3						
Hotel	3						
Food Truck	2						
Convenience Store	2						
Park	2						
Bus Station	2						
Bar	1						
Asian Restaurant	1						
Plaza	1						
Fried Chicken Joint	1						
Gym / Fitness Center	1						
Seafood Restaurant	1						
Movie Theater	1						
Soccer Field	1						
Flower Shop	1						
Steakhouse	1						
BBQ Joint	1						
Pharmacy	1						
Health & Beauty Service	1						
Italian Restaurant	1						
Paper / Office Supplies Store	1						
Playground	1						
Sandwich Place	1						
Yoga Studio	1						

Private system practices are 85% of Medical Practices in Queretaro as shown in table 2, Cluster 0 and 1 are the most frequent kind. Nearby Hospitals, Pharmacies and other business are more frequent in this clusters, so this will be assessed as a probably optimal point for the the Medical Practice.

Table 2

Health System	Cluster Labels	Healthcenter	Percent
Private	0	138	50.364964
Private	1	56	20.437956
Private	2	41	14.963504
Public	0	20	7.299270
Public	1	12	4.379562
Public	2	6	2.189781
Public	3	1	0.364964

7.4.Demographics Analysis

The optimal profile for this market is, the population that maximizes these variables as show in figure 6 and figure 7:

A. Population from 31 to 45 years old and from 0 to 5.
B. 18 years old or more married Population.
C. Employed population.
D. Population with higher education (18 years old or more).

Socioeconomic index also relevant, main reason is that as the socioeconomic indicator is higher the area is wealthier, and less prone to be attended in the Public Health system. Probably one of the AGEBS is above the mean in one of the characteristics listed above but the population is from low income, so the will seek medical attention in Public Healthcare system. In this section we determine that

AGEBS with socioeconomic index from 5 to 8 are the best zones for the medical practices as show in figure 8.

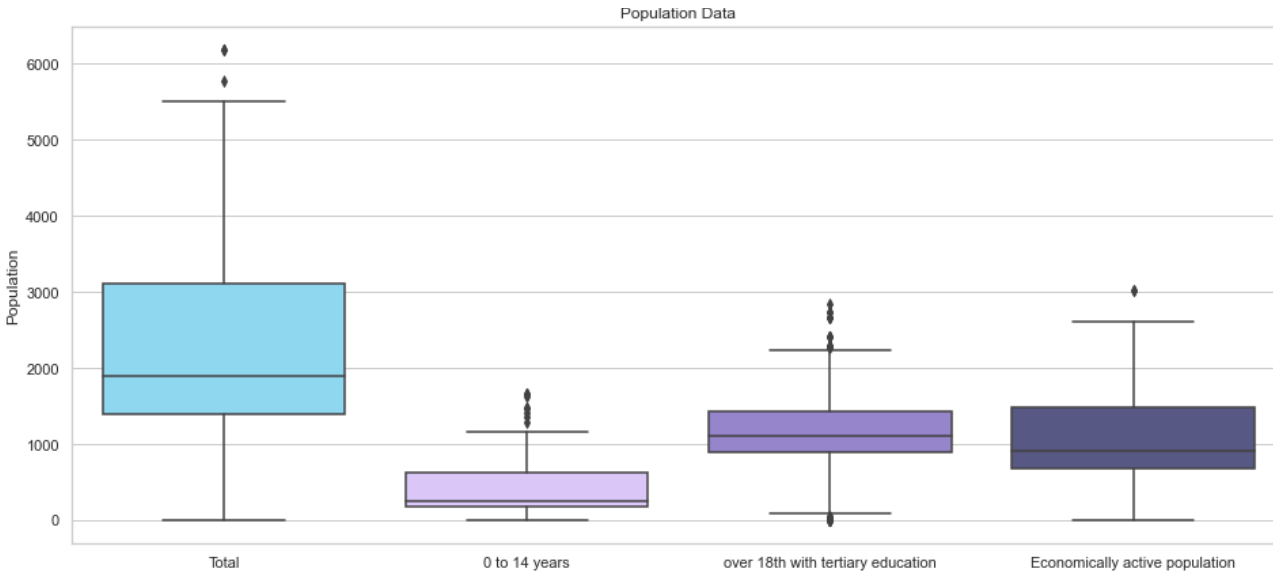

Fig. 6: Target population demographics.

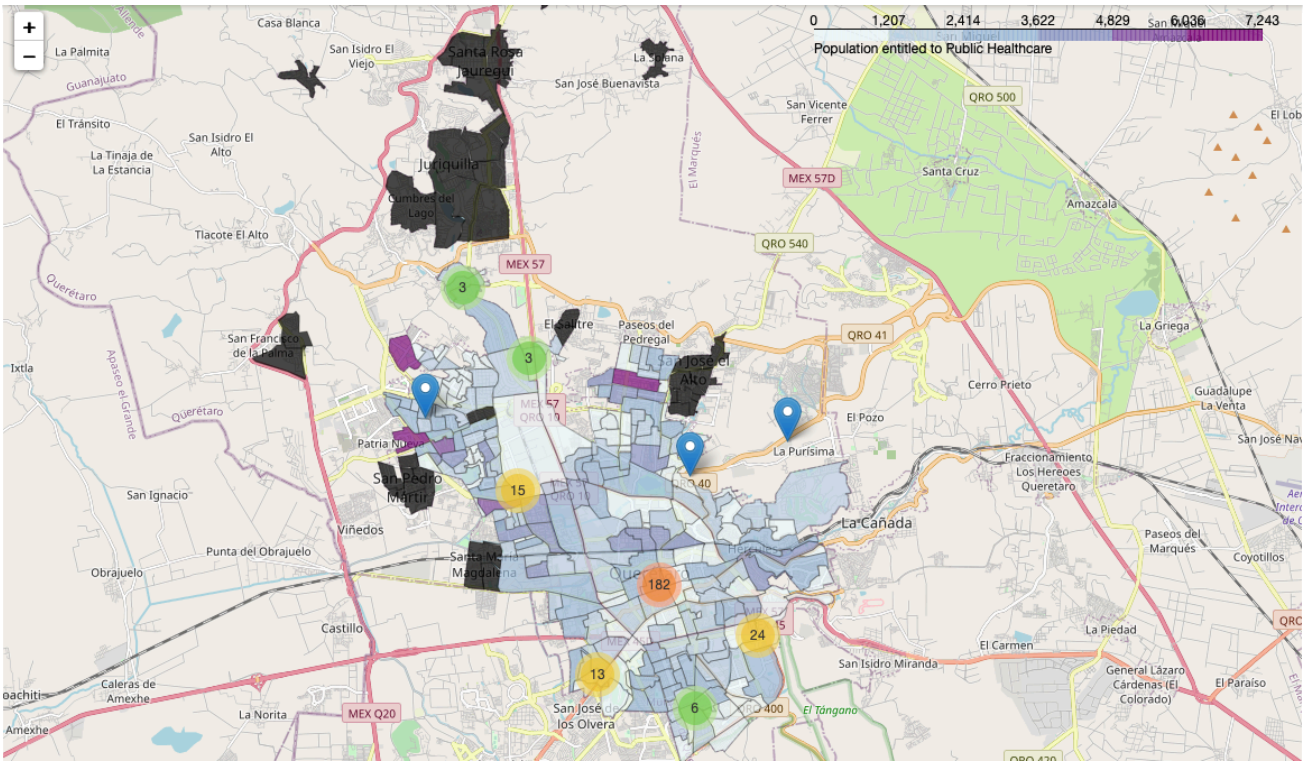

Fig. 7 : Map showing total population entitled to healthcare within Querétaro AGEBS, and location of each health center.

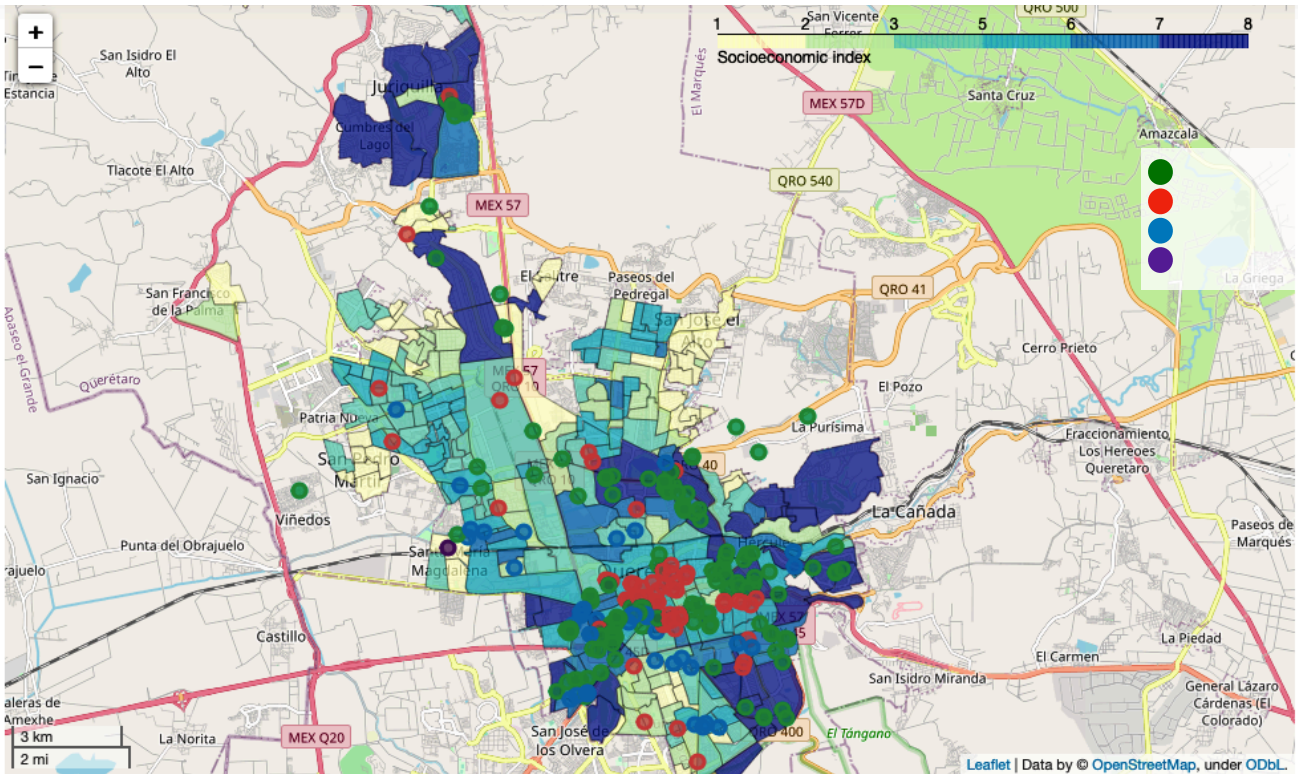

Fig. 8: Choropleth map of Querétaro City, socioeconomic index of each AGEB is displayed by gradient, health-centers are colored by K-means cluster.

7.5.Analyze Nearby Business

Medical practices benefit from operating close to places such as: -Pharmacies & drug stores -Hospitals -Urgent care centers -Fitness centers, we need to identify the most crowed clusters. Most frequently occurring venues within 500 meters of Health Centers are shown in figure 3, Restaurants are the most common venue followed by coffee shops, bars, hotels, pharmacies and fitness centers. There is not a correlation between the most frequent venues around health centers as shown in figure 9.

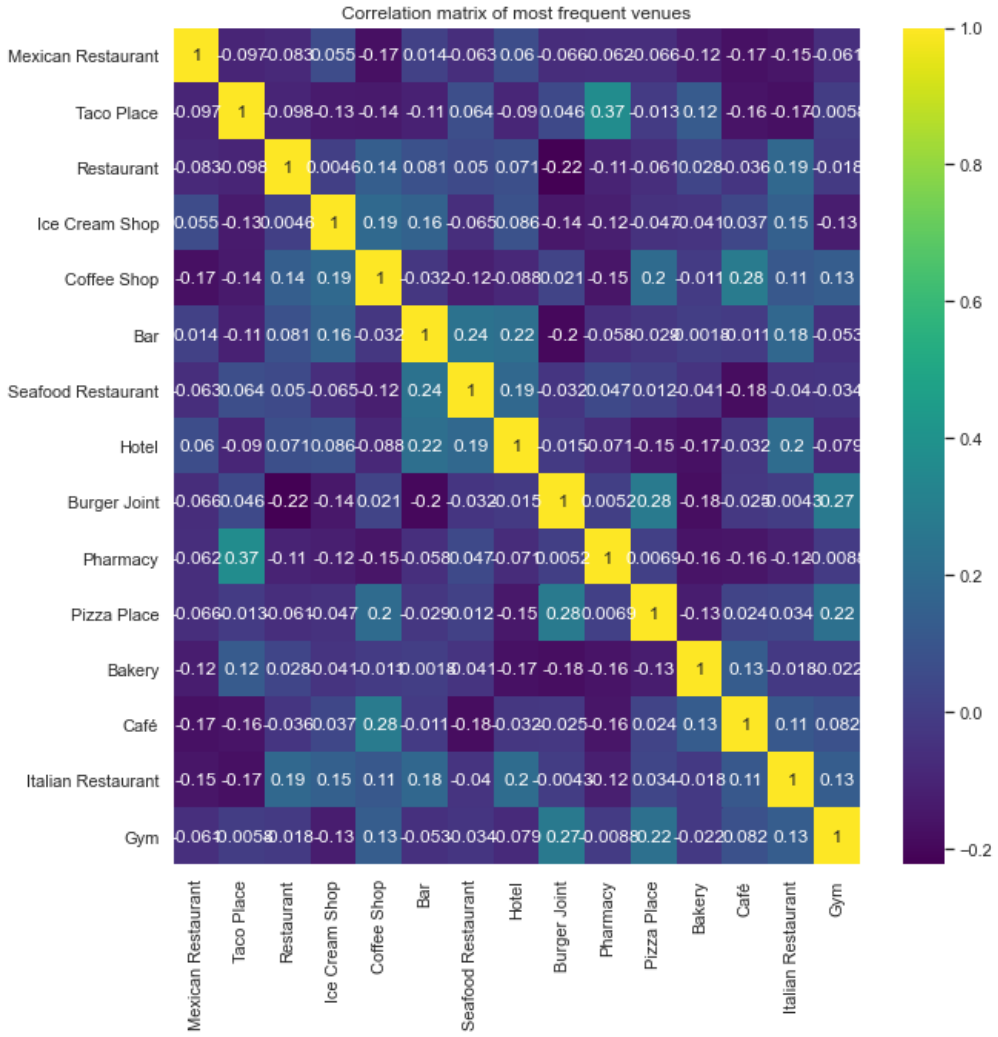

Fig 9. Correlation matrix between venues nearby health centers

Cluster 0 is the most frequent, 120 are from private health system and 18 from public health system, 138 in total, being the most frequent means being close this areas could imply more success.

7.6.Analyze Accessibility & Visibility

Medical practice location must be within 20 minutes of residential areas, main reason is because 90% of patients come from 15 km radius. It is important to observe the micro-local aspect (visibility, image, safety, roads, travel time, location on road axes, parking lots, etc.). And aim for a location with a spacious entryway where elderly, injured or disabled patients can be dropped off and picked up without difficulty.

As shown in figure 10, four radius are plot from the centroid of the city: 2.5 km, 5 km, 10 km and 15 km, to demonstrate the maximum distance that patients will travel for healthcare.

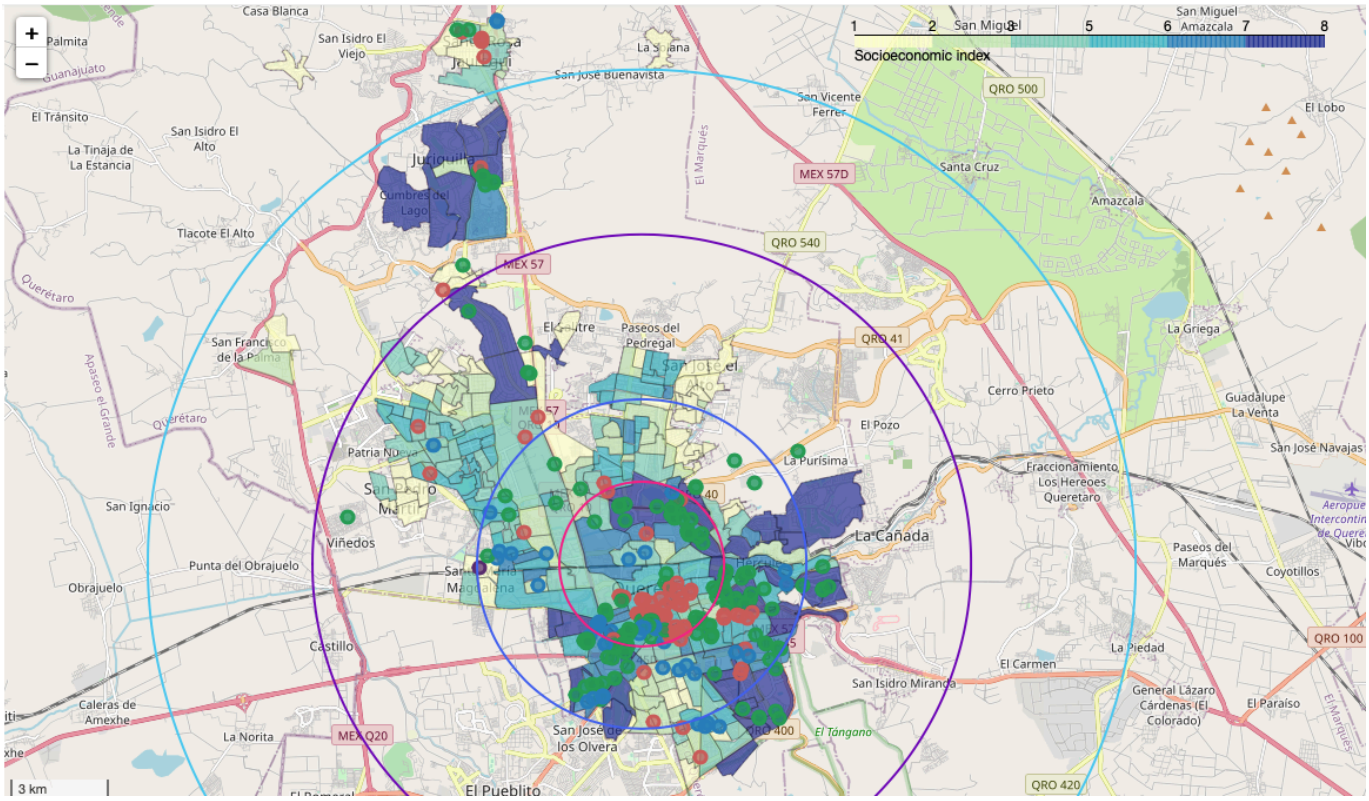

Fig. 10: Choropleth map showing maximum distance that patient travel for basic healthcare, pink circle 2.5km radius, navy blue 5 km radius, purple circle 10 km radius and light blue 15 km radius.

Cluster zero (red points) and one (green points) meet accessibility & visibility criteria, Most of them are within the 5km radius area, this clusters are in proximal to residential communities and highway routes, and they are relative close to busy retail zones. Majority of Queretaro Health centers are closer to important avenues, we can highlight 4:

Boulevard Bernardo Quintana

Avenida Ignacio Zaragoza

Avenida Constituyentes

Carretera Federal 57 Mexico Queretaro

So our ideal location has to be close to these main highways and inside the 5 km circle radius, also close to cluster 1 or 0 and in the most wealthy AGEBS.

7.7.Network-Based Spatial Clustering

As proposed in the previous section, human mobility is network-constrained. To properly model agglomeration along a city's street network, we must use network-based spatial clustering, we clearly defined 4 main avenues were most of the Health-centers are clustered.

Urban circulation is constrained to networks of streets and paths. Due to terrain and hydrological or even geological formations, even adjacent land parcels may not interface with each other except through a long trip along the street network.

To apply network based spatial clustering, first we need to attach every establishment to its nearest network node by OSMnx library. thanks to gboeing for his tutorial.

After download and assemble the street network from Queretaro City with OSMnx. We visualize how our health-centers cluster into distinct districts or neighbourhoods, as many establishments and industries tend to do, map of Queretaro streets network and health-centers is show in figure 11.

Fig. 11: Map showing street network and health-centers distribution in Queretaro City.

In order to explore how these health centers agglomerate, we can identify spatial clusters using DBSCAN algorithm. DBSCAN identifies points as members of a cluster if each is within epsilon distance of another and if this cluster contains at least ***n*** minpts - number of points. For this example we parameterize it with an epsilon = 500 and minpts = 3. That is, points must be within 500 meters of each other and a cluster must contain at least 3 points.

We compute DBSCAN by converting everything to radians, fitting it, then getting cluster labels for each establishment. We can visualize our establishments, coloring them by spatial cluster label as show in figure 12.

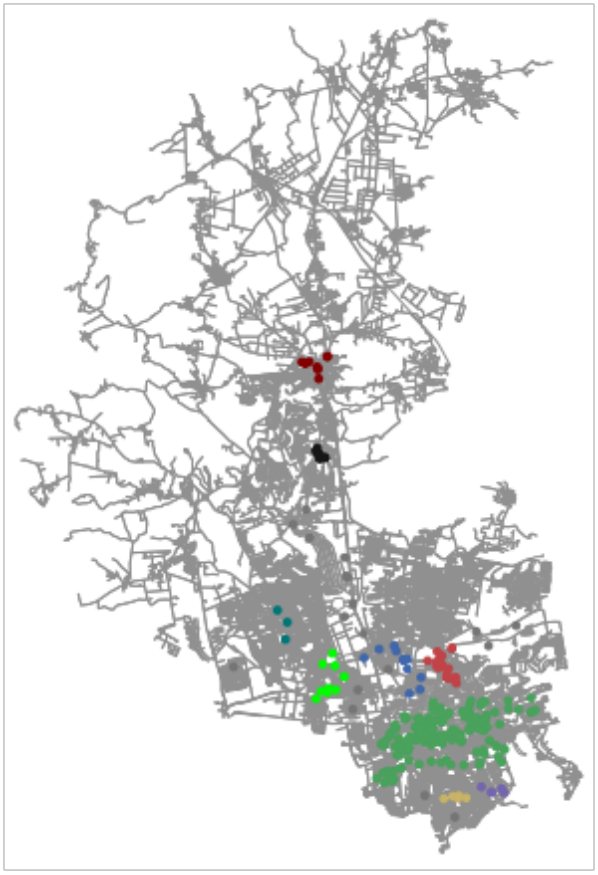

Fig. 12: DBSCAN clustering of health-centers and street network.

DBSCAN find 11 clusters, each cluster show in figure 10 contains at least 3 health-centers and no health-center is more than 500 meters away from another.

As mentioned early urban circulation constrained to networks of streets and paths, and some health-centers even adjacent may not interface with each other except through a long trip along the street network.

To identify how these health-centers cluster together in a meaningful way given that urban circulation is network-constrained, we must re-label them using network-based spatial clustering.

Steps for this methodology are as follow:

I. Identify if the graph is strongly connected or weakly connected, we use a strong connected graph.
II. Use OSMnx to attach health-centers to its nearest network node.
III. Create a function to calculate a node-based network distance matrix.
IV. Calculate the distance matrix using our OSMnx street network, G, and re-index to make the matrix establishment-based.
V. Re-compute DBSCAN fitting it to our precomputed sparse matrix to speed up computation and use same parameters.

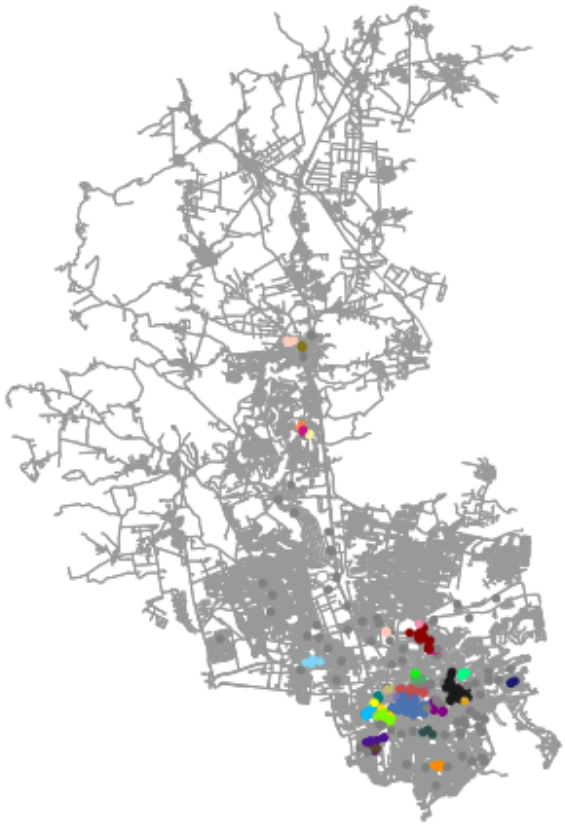

Fig. 13: Street network map and Health center clustering by Network based spatial clustering

In the initial step spatially DBSCAN find 10 clusters. Network-constrained density-based spatial clustering find 31 clusters as show in figure 13. This more accurately reflects circulation and agglomeration in real-world urban space, for example IMSS Hospital General Regional 1 is a huge facility and occupies a street block, but there's only one entrance for patient in Av. 5 de Febrero, in the 4 streets that surround them its the only way to get in.

At this point we know from k-means clustering that cluster 0 and 1 health-centers are the best places due to nearby business area, and Health-centers from this cluster are most promising areas for the medical practice., also we know the health-centers that are in the most wealthy AGEBS have more possibility to succeed and attract patients, and finally the most accessible and visible health-centers are close to these avenues: Boulevard Bernardo Quintana, Avenida Ignacio Zaragoza, Avenida Constituyentes, Carretera Federal 57 Mexico Queretaro.

For a closer examination of top rated areas we zoom in the street network to analyze distribution as shown in figure 14, central clusters were more prone to success due to accomplish most of the requisites. As shown in figure 14, network based spatial clusters are more homogeneous due to the fact that most of them are inside the same AGEB and share also some similarities due to the area, for example blue cluster

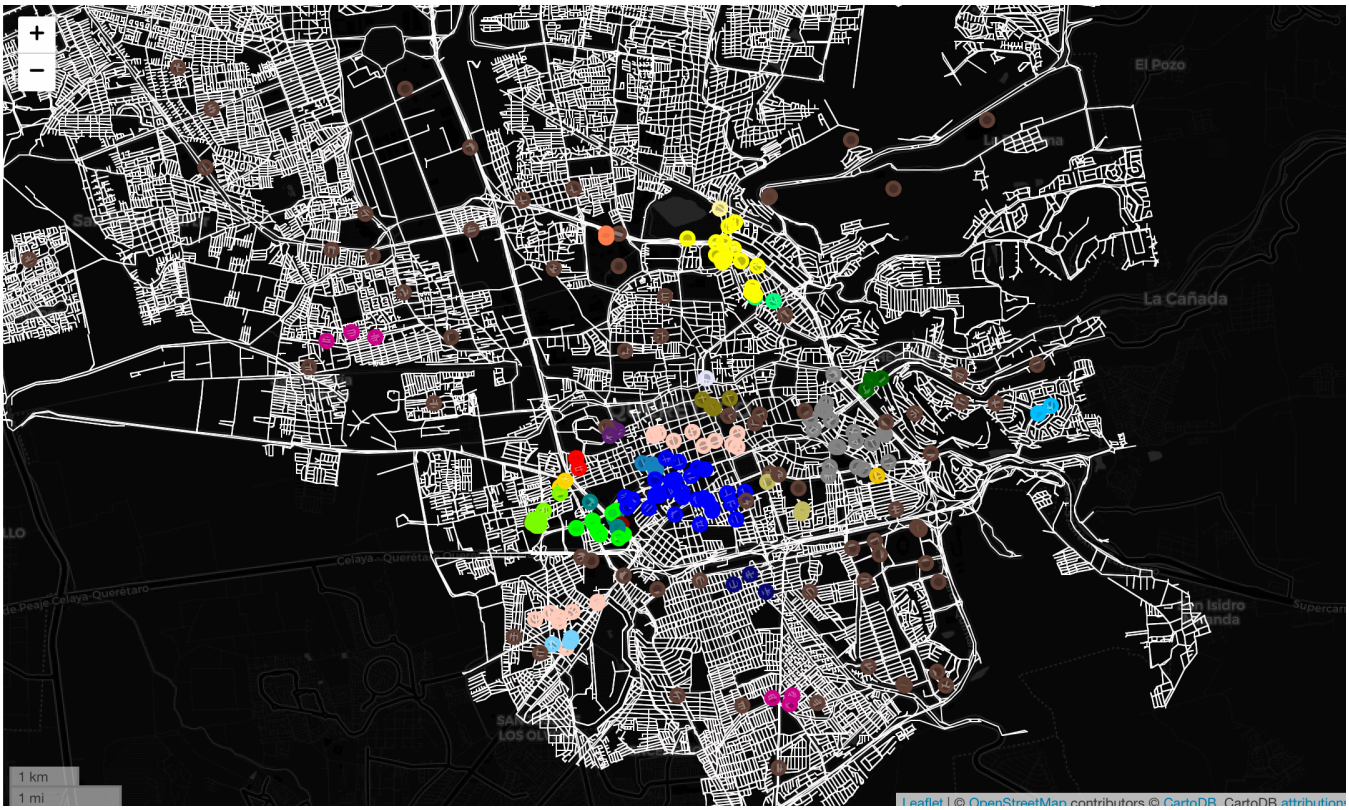

Fig. 14: Network based spatial clustering of Health-centers in Querétaro City.

7.8.Get Suitable Locations

From the K-Means Clustering , we selected the following locations for further analysis: Cluster zero and one Health-centers close to these principal avenues

1. Boulevard Bernardo Quintana
2. Avenida Ignacio Zaragoza
3. Avenida Constituyentes
4. Carretera Federal 57 Mexico Queretaro

So our ideal location has to be close to these main highways and inside the 5 km circle radius, also close to cluster 1 or 0 and in the most wealthy AGEBS. After

Network spatial based clustering by DBSCAN, we define 31 clusters in these principal avenues, from those the clusters we create a function to show the best locations.

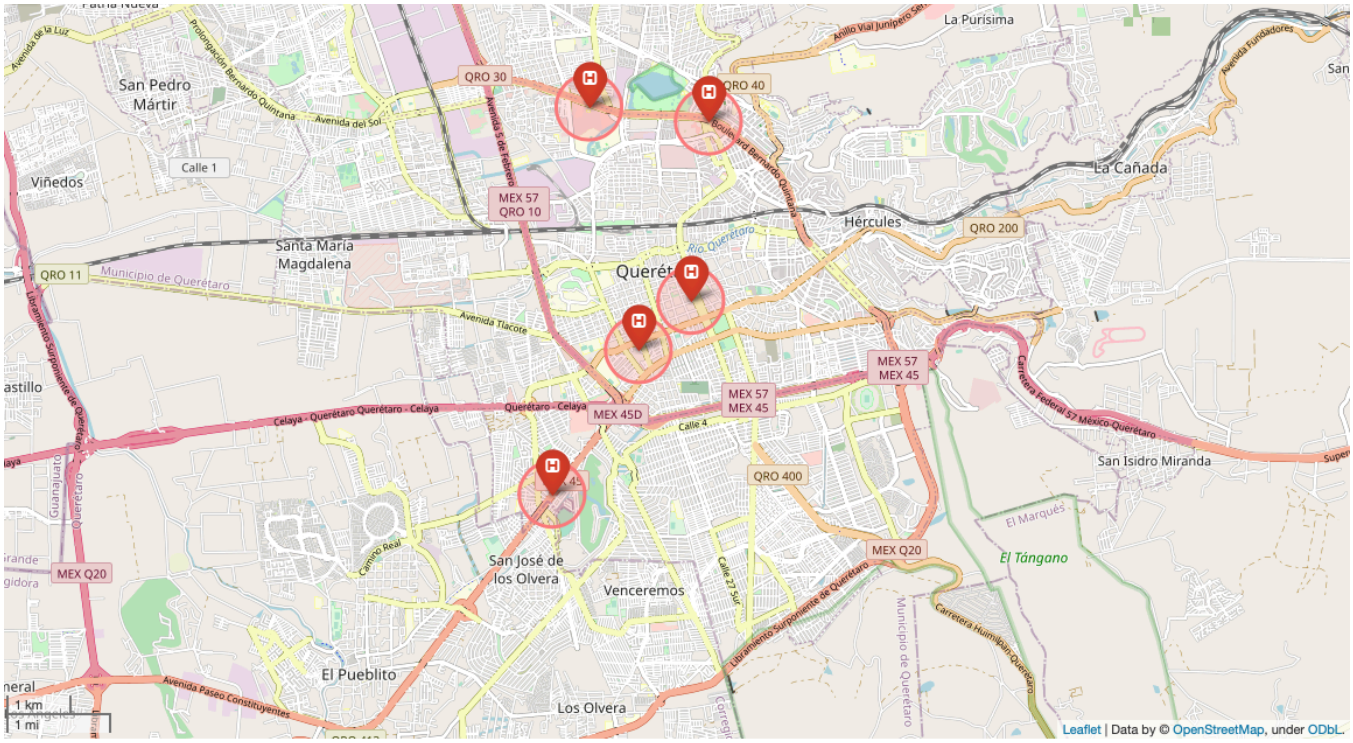

Fig. 15: Ideal location for Medical Practices in Querétaro City.

5. Results

In this market study we were able to forecast five possible success points for opening of a new specialty clinic for the city of Querétaro Mexico. All points have the same probability of success, since they comply with the main key factors, such as: Closeness to private hospitals in the city, which allows the fast and efficient transfer of patients and the medical staff within the hospital area; the proximity to quick and easy access to the road network in the city. Location of hospitals and private clinics was appreciated according to the socioeconomic level of Querétaro by the mapping tool; great concentration is appreciated of the private health-care system in the highest

socioeconomic zone levels, while the public health-care system is located in the lowest economic levels areas.

We were able to find 4 clusters from K-means and 31 from Network based spatial clustering DBSCAN, Cluster zero and one from k-means were the most promising and most frequent clusters on the city, Private system practices labeled as cluster 0 and 1 were 86% of Health-centers in Querétaro. Nearby Hospitals, Pharmacies and other business are more frequent in this clusters. Also Network Based Spatial Clustering show how these main clusters distribute and how they compare with k-means label, this lead to identify the best DBSCAN clusters which are labeled as 1 on k-means were also the Health-centers close to these principal avenues:

1. Boulevard Bernardo Quintana
2. Avenida Ignacio Zaragoza
3. Avenida Constituyentes
4. Carretera Federal 57 Mexico Queretaro

It was determined that the new high-specialty offer should be located very close to the main road axes with adequate access. Most adequate locations must be within 500m from this addresses or in one of these venues as show in table 2.

Table 2: Ideal address for Medical Practice

	Ideal Address
1	Star Médica, Boulevard Bernardo Quintana, Delegación Centro Histórico, Santiago de Querétaro, Municipio de Querétaro, Querétaro, 76130, México
2	Privada Ignacio Zaragoza, Villas del Refugio, Delegación Centro Histórico, Santiago de Querétaro, Municipio de Querétaro, Querétaro, 76030, México
3	Lateral Bernardo Quintana, Rinconada Álamos, Delegación Centro Histórico, Santiago de Querétaro, Municipio de Querétaro, Querétaro, 76160, México
4	Avenida Paseo Constituyentes, Mansiones del Valle, Delegación Josefa Vergara y Hernández, Municipio de Querétaro, Querétaro, 76185, México
5	32, Calle Benito Juárez Sur, Centro, Delegación Centro Histórico, Santiago de Querétaro, Municipio de Querétaro, Querétaro, 76000, México

5. Discussion

Querétaro city is a strong business and economic centre, its economic growth is above the national average, the city is expected to increase in size 35% over the next 20 to 25 years. This means socioeconomic data may change, COVID 19 pandemic will inflict a change over the economy dynamics but on the other side the private health-care system will keep growing.

From the results achieved it can be seen that there are other factors we could use to find a more suitable location such as price per square meter and ratings of close venues, if the area is booming or decaying, we could look at the mean lease price for clinics within these AGEBS so that the stakeholder can see the financial implications too.

Another important service that must be on the same premises, or at least close enough, is the clinic diagnostics laboratory, physicians rely on laboratory testing to make timely and evidence based decisions, this allows physicians to apply the most appropriate, cost effective use of expensive drugs and other therapies. Clinic

laboratories services have a direct impact on many aspects of patient care including, length of stay, patient safety, resource utilization and customer satisfaction.

6. Conclusion

Our research goal was to find suitable areas for medical practices, and to locate and area within the AGEB which would be most suitable, and do not be to far or to hard to get by.

From our results we can see the following:

Querétaro follows the national rule in terms of the most popular restaurant types. This could help us with further research if required. This research has provided valuable information on potential access to primary care.This information may be used in future research to further the dialogue of neighbourhood-level access to care, as well as to refine the methods used to examine potential access.

Of the addresses that we retrieved from our final analysis, the most suitable appears to be in Bernardo Quintana and Avenida Ignacio Zaragoza with the highest cluster of venues situated in close proximity. The Bernardo Quintana and Avenida Ignacio Zaragoza area also has a high density of similar clinics and laboratories, and we can make an assumption that due to the densities of both venues and medical practices venues the lease price in the area will be the highest.

Additional individual characteristics including age, gender, ethnicity, socioeconomic-status, beliefs about health and the actual need for care will also determine whether and where an individual seeks care (Gatrell, 2002; Aday andAndersen,1974).