Geographical Analysis of Hospital Location In Merida, Yucatan, Mexico

William Gruintal

June 10, 2021

1 Capstone Project - The Battle of the Neighborhoods (Week 2)

1.0.1 Applied Data Science Capstone by IBM/Coursera

1.1 Table of contents

1.2 Introduction: Business Problem

In this project, we will try to find an optimal location for a hospital. Specifically, this report will be targeted to stakeholders interested in opening a **small hospital** in **Merida, Yucatan**, Mexico. Since this small hospital will be competing with other hospitals and doctor's offices, we would refer as "hospital" to the sum of hospitals plus doctor's offices in a given area.

Since there are many medical centers in Merida we tried to detect **locations that are not already crowded with medical centers**. We were also particularly interested in **areas with no hospitals in vicinity**. We would also prefer locations **north to city center**, assuming that first two conditions are met.

We will generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

1.3 Data

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

• Number of existing medical centers in the neighborhood (any type of medical center)

• Number of and distance to hospitals in the neighborhood, if any

• Distance of neighborhood from city center in the north direction (just as a reference)

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources were needed to extract/generate the required information:

• Centers of candidate areas were generated algorithmically and approximate addresses of centers of those areas were obtained using **Geocoder library reverse geocoding**

• Number of medical centers and their type and location in every neighborhood were obtained using **Foursquare API**

• Coordinates of Merida center were obtained using **Geocoder library** of well-known Merida location (Merida Centro)

1.3.1 Neighborhood Candidates

We created latitude & longitude coordinates for centroids of our candidate neighborhoods. We created a grid of cells covering our area of interest, which is approx. 24x24 kilometers centered around Merida city center.

We first found the latitude & longitude of Merida city center, using specific, well-known address and Geocoder library:

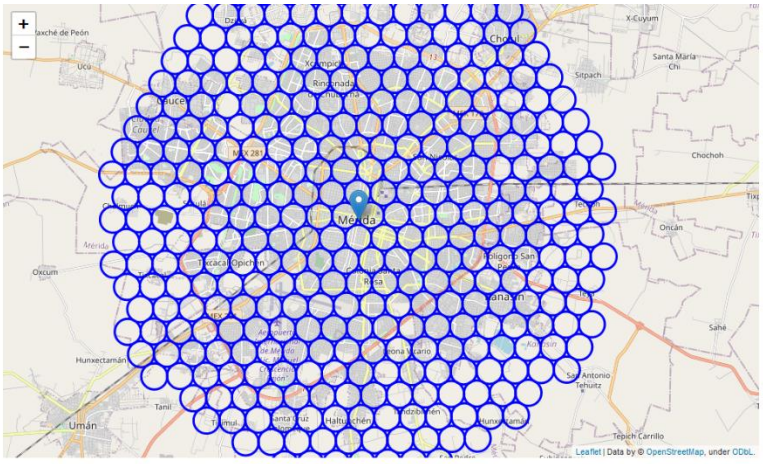
Merida, Yucatan coordinates are: 20.96670000000006, -89.62316999999996.

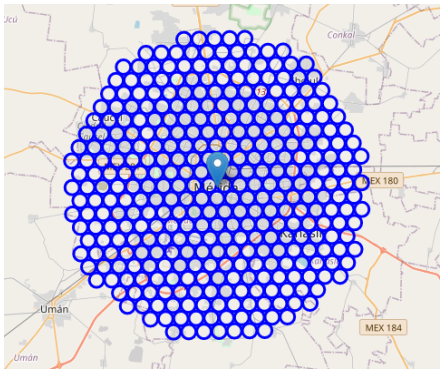
Then we created a grid of area candidates, equally spaced, centered around city center and within **~12km** from Merida Centro, which covers the entire city and some adjacent inhabited areas. Our neighborhoods were defined as circular areas with a radius of 600 meters, so our neighborhood centers were 1200 meters apart.

To accurately calculate distances we needed to create our grid of locations in Cartesian 2D coordinate system which allowed us to calculate distances in meters (not in latitude/longitude degrees). Then we would project those coordinates back to latitude/longitude degrees to be shown on Folium map. For these reasons, we created functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

In order to create a **hexagonal grid of cells**: we offset every other row, and adjust vertical row spacing so that **every cell center is equally distant from all it's neighbors**. After this step, 360 candidate neighborhood centers were generated.

Finally, we visualized the data we had so far: city center location and candidate neighborhood centers:





Now we had the coordinates of centers of neighborhoods/areas to be evaluated, equally spaced (distance from every point to it's neighbors was exactly the same) and within **~12km from Merida Centro**.

Next step was to use Geocoder library to get approximate addresses of those locations, place them all into a Pandas dataframe and save this data into local file:

	Address	Latitude	Longitude	X	Y	Distance from center
0	Diagonal 24, Dzununcán, Mérida, Yucatán, 97315	20.865108	-89.655983	223645.261253	2.309501e+06	11760.689991
1	Mérida, Yucatán	20.865286	-89.644459	224845.261253	2.309501e+06	11450.494709
2	Privada 56, Xmatkuil, Mérida, Yucatán, 97315	20.865464	-89.632935	226045.261253	2.309501e+06	11260.276598
3	Mérida, Yucatán	20.865641	-89.621411	227245.261253	2.309501e+06	11196.152423
4	Mérida, Yucatán	20.865817	-89.609887	228445.261253	2.309501e+06	11260.276598
5	Mérida, Yucatán	20.865993	-89.598363	229645.261253	2.309501e+06	11450.494709
6	Mérida, Yucatán	20.866167	-89.586839	230845.261253	2.309501e+06	11760.689991
7	Mérida, Yucatán	20.874218	-89.673434	221845.261253	2.310541e+06	11503.176225
8	Mérida, Yucatán	20.874398	-89.661910	223045.261253	2.310541e+06	10991.044685
9	Mérida, Yucatán	20.874577	-89.650385	224245.261253	2.310541e+06	10590.706457

1.3.2 Foursquare

After previous steps, we used Foursquare API to get info on medical centers in each neighborhood.

We were interested in venues in 'medical center' category, but only those that are proper medical centers - veterinaries, quiropractors, medical labs, etc. are not direct competitors so we did not care about those. So we included in our list only venues that have some key words in category name, and made sure to detect and include all the subcategories of specific 'hospital' category (hospital and doctor's office), as we needed info on hospitals in the neighborhood.

We went over our neighborhood locations, got nearby hospitals, and maintained a count of all found medical centers and all found hospitals obtaining the following totals:

Total number of medical centers: 547

Total number of hospitals: 258

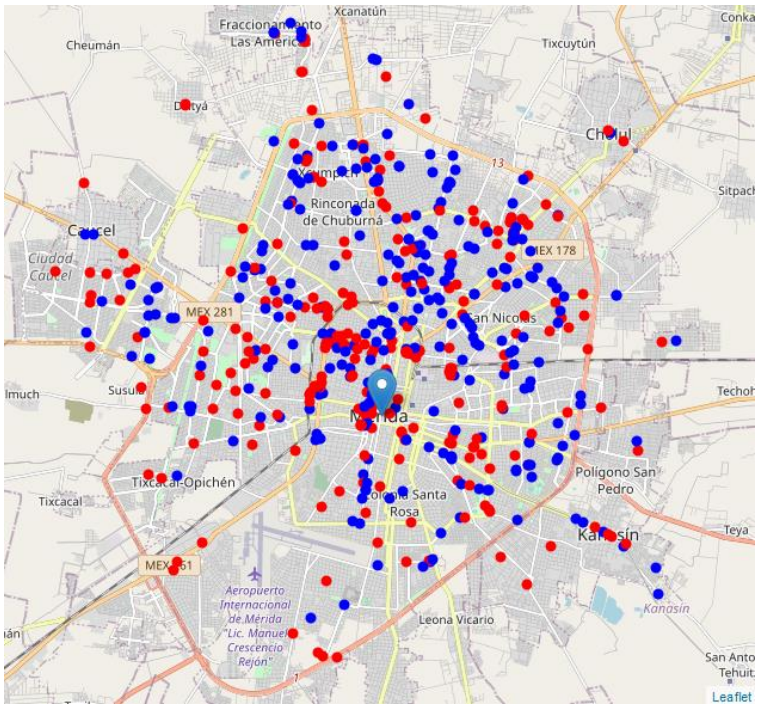
Percentage of hospitals: 47.17%

Average number of medical centers in neighborhood: 0.2777777777777778

After some exploration of our data we found that venue 'Opticas Devlyn' was incorrectly categorized as "Doctor's Office" since it is an eye care center. Nonetheless, we considered that this error in data from source was not relevant and would not bias our conclusions:

```
[{'reasons': {'count': 0,
   'items': [{'summary': 'This spot is popular',
     'type': 'general',
     'reasonName': 'globalInteractionReason'}]},
  'venue': {'id': '50256af321c61aebc69bc8cd',
   'name': 'Opticas Devlyn',
   'location': {'address': 'Calle 60 # 299 Carretera Merida - Progreso Km 8.5',
    'lat': 21.0395158093409,
    'lng': -89.63091360677139,
    'labeledLatLngs': [{'label': 'display',
      'lat': 21.0395158093409,
      'lng': -89.63091360677139}],
    'distance': 4736,
    'postalCode': '97110',
    'cc': 'MX',
    'city': 'Mérida',
    'state': 'Yucatán',
    'country': 'México',
    'formattedAddress': ['Calle 60 # 299 Carretera Merida - Progreso Km 8.5',
     '97110 Mérida, Yucatán',
     'México']},
   'categories': [{'id': '4bf58dd8d48988d177941735',
     'name': "Doctor's Office",
     'pluralName': "Doctor's Offices",
     'shortName': "Doctor's Office",
     'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/building/medical_doctorsoffice_',
```


We then proceeded to visualize all the collected medical centers (blue) in our area of interest on a map, and showed hospitals in different color (red).



As the map shows we had all the medical centers in area within 12 kilometers from Merida Centro, and we knew which ones were hospitals. We also knew which medical centers exactly were in vicinity of every neighborhood candidate center.

This concluded the data-gathering phase. In the next phase, we would use this data for analysis to produce the report on optimal locations for a new hospital!

1.4 Methodology

We directed our efforts on detecting areas of Merida that have low medical center density, particularly those with low number of hospitals. We limited our analysis to area ~12km around city center, which covers the entire city and some adjacent inhabited areas.

In first phase we had collected the required **data: location and type (category) of every medical center within 12km from Merida center** (Merida Centro). We have also **identified hospitals** (according to Foursquare categorization and our operational definition).

Second phase in our analysis would be calculation and exploration of 'medical center density' across different areas of Merida. We used heatmaps to identify a few promising areas around north to center with low number of medical centers in general (and no hospitals in vicinity) and focused our attention on those areas.

In third and final phase we focused on most promising areas and within those created **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we take into consideration locations with **no more than 4 medical centers in radius of 500 meters**, and we wanted locations **without hospitals in radius of 500 meters**. We presented a map of all such locations but also created clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

1.5 Analysis

We performed some basic explanatory data analysis and derived some additional info from our raw data. First, we counted the **number of medical centers in every area candidate**:

Average number of medical centers in every area with radius = 600m: 0.2777777777777778

	Address	Latitude	Longitude	X	Y	Distance from center	Medical centers in area
0	Diagonal 24, Dzununcán, Mérida, Yucatán, 97315	20.865108	-89.655983	223645.261253	2.309501e+06	11760.689991	0
1	Mérida, Yucatán	20.865286	-89.644459	224845.261253	2.309501e+06	11450.494709	0
2	Privada 56, Xmatkuil, Mérida, Yucatán, 97315	20.865464	-89.632935	226045.261253	2.309501e+06	11260.276598	0
3	Mérida, Yucatán	20.865641	-89.621411	227245.261253	2.309501e+06	11196.152423	0
4	Mérida, Yucatán	20.865817	-89.609887	228445.261253	2.309501e+06	11260.276598	0
5	Mérida, Yucatán	20.865993	-89.598363	229645.261253	2.309501e+06	11450.494709	0
6	Mérida, Yucatán	20.866167	-89.586839	230845.261253	2.309501e+06	11760.689991	0
7	Mérida, Yucatán	20.874218	-89.673434	221845.261253	2.310541e+06	11503.176225	0
8	Mérida, Yucatán	20.874398	-89.661910	223045.261253	2.310541e+06	10991.044685	0
9	Mérida, Yucatán	20.874577	-89.650385	224245.261253	2.310541e+06	10590.706457	0

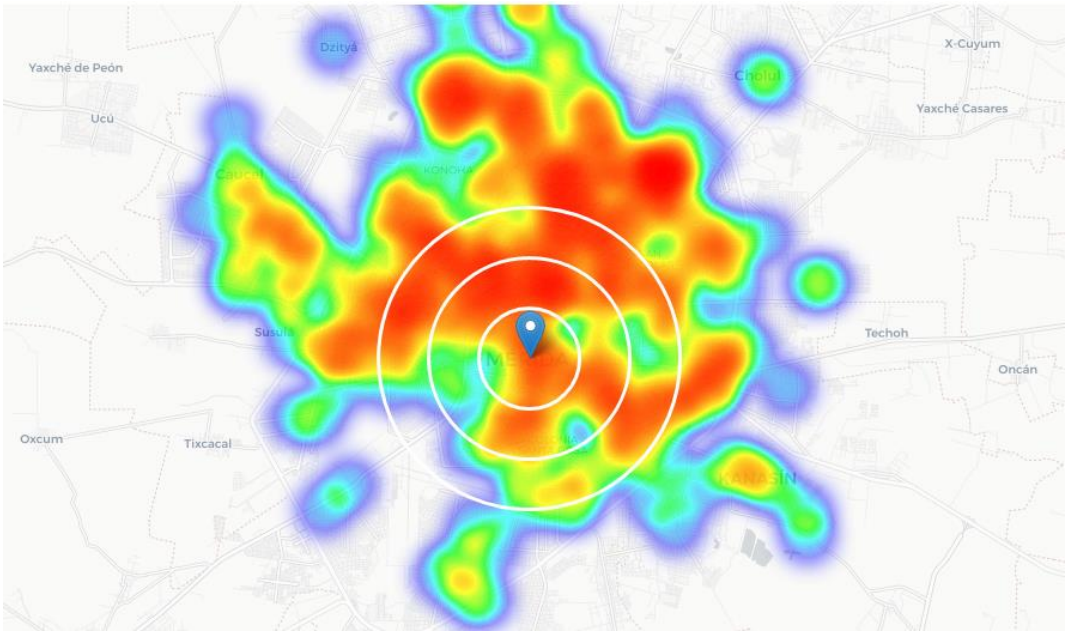
After that we calculated the **distance to nearest hospital from every area candidate center** (not only those within 600m as we wanted distance to closest one, regardless of how distant it would be).

	Address	Latitude	Longitude	X	Y	Distance from center	Medical centers in area	Distance to hospital
0	Diagonal 24, Dzununcán, Mérida, Yucatán, 97315	20.865108	-89.655983	223645.261253	2.309501e+06	11760.689991	0	4461.567451
1	Mérida, Yucatán	20.865286	-89.644459	224845.261253	2.309501e+06	11450.494709	0	4141.800051
2	Privada 56, Xmatkuil, Mérida, Yucatán, 97315	20.865464	-89.632935	226045.261253	2.309501e+06	11260.276598	0	4072.771797
3	Mérida, Yucatán	20.865641	-89.621411	227245.261253	2.309501e+06	11196.152423	0	4305.955595
4	Mérida, Yucatán	20.865817	-89.609887	228445.261253	2.309501e+06	11260.276598	0	4834.773734
5	Mérida, Yucatán	20.865993	-89.598363	229645.261253	2.309501e+06	11450.494709	0	5575.734977
6	Mérida, Yucatán	20.866167	-89.586839	230845.261253	2.309501e+06	11760.689991	0	6176.400829
7	Mérida, Yucatán	20.874218	-89.673434	221845.261253	2.310541e+06	11503.176225	0	4596.592117
8	Mérida, Yucatán	20.874398	-89.661910	223045.261253	2.310541e+06	10991.044685	0	3861.056780
9	Mérida, Yucatán	20.874577	-89.650385	224245.261253	2.310541e+06	10590.706457	0	3273.634526

Average distance to closest hospital from each area center: 1528.8526993296232

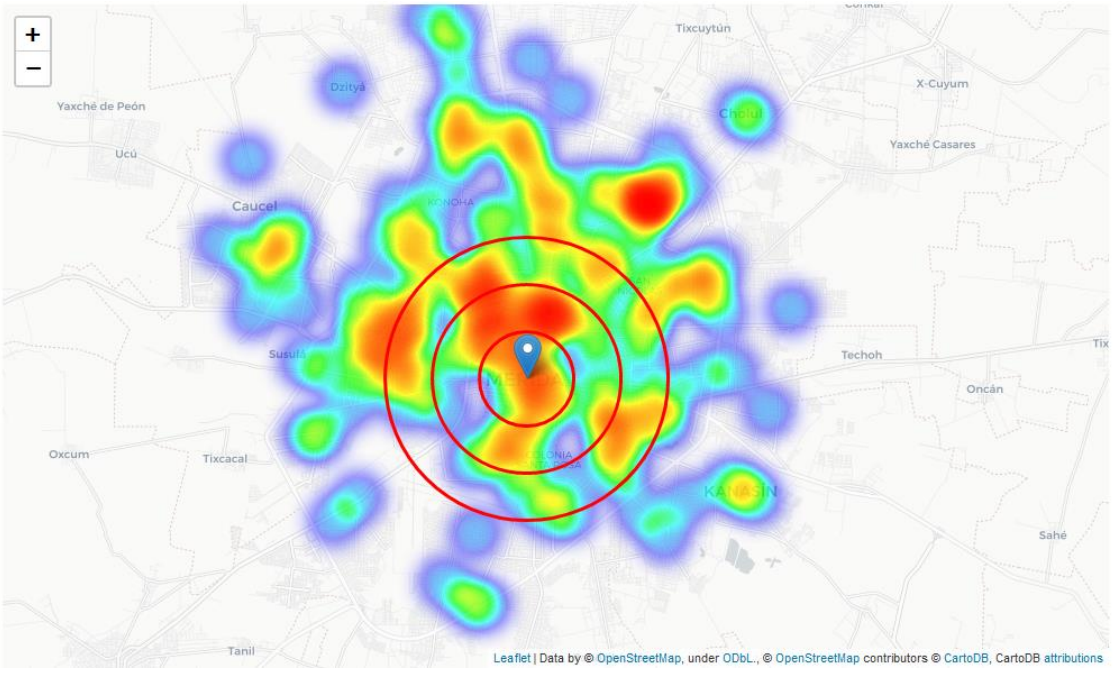
So **on average a hospital can be found within ~1500m** from every area center candidate. That seems to be fairly apart even though the average is pulled up by the distances in areas south and outside the city.

Here we create our first map showing **heatmap / density of medical centers** and trying to extract some meaningfull info from that. Also, we showed few circles indicating distance of 1.5km, 3.0km and 4.5km from Merida Centro as a visual reference.



Here we noticed that an interesting pocket of low medical center density north to city center was found **north-west from Merida Centro**.

We created another heatmap map showing **heatmap/density of hospitals** only.



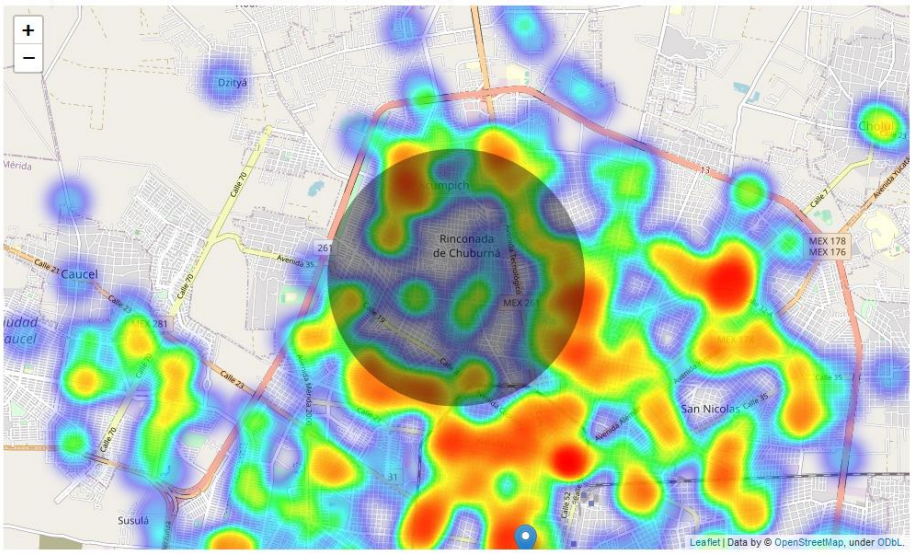
This map was not so 'hot' (hospitals represent a subset of ~47% of all medical centers in Merida) but it also indicated higher density of existing hospitals directly north from Merida Centro, with biggest pocket of **low hospital density positioned north-west from city center**. Most south side of city was 'cold' meaning low hospital density but this area was not of interest to stakeholders.

Based on this we focused our analysis on areas north-west from Merida center - we moved the center of our area of interest and reduce it's size to have a radius of **2.5km**. This placed our location candidates mostly in neighborhoods near to the area known as **'Barrio de Chuburna'** and near to neighborhood **Camara de la Construccion**.

1.5.1 Barrio de Chuburna and Camara de la Construccion

Analysis of some market intelligence sites showed that Barrio de Chuburna could be categorized as a 'type C' which can be roughly equivalent to 'middle class' while Camara de la Construccion could be classified as 'type D+' which made it a bit less desirable than the former. Nonetheless, both 'regions of interest' deserved further analysis.

We defined a new, narrower region of interest, which included parts of Barrio de Chuburna and Camara de la Construccion neighborhood with low medical centers count north-west to Merida Centro and showed in a map:



This covered the biggest pocket of low medical center density in Barrio de Chuburna and Camara de la Construccion noorth-west to Merida center.

We then created a new, more dense grid of location candidates restricted to our new region of interest (we made our location candidates **200m apart**).

We obtained 565 candidate neighborhood centers.

After this, we calculated the two most important things for each location candidate: **number of medical centers in vicinity** (we used radius of **500 meters**) and **distance to closest hospital**.

	Latitude	Longitude	X	Y	Medical centers nearby	Distance to hospital
0	20.993585	-89.636138	225945.261253	2.323698e+06	30	257.007515
1	20.995022	-89.644335	225095.261253	2.323871e+06	19	129.901041
2	20.995052	-89.642413	225295.261253	2.323871e+06	15	115.980142
3	20.995082	-89.640490	225495.261253	2.323871e+06	24	130.628347
4	20.995111	-89.638568	225695.261253	2.323871e+06	27	39.423670
5	20.995141	-89.636646	225895.261253	2.323871e+06	25	199.793286
6	20.995171	-89.634724	226095.261253	2.323871e+06	21	283.856613
7	20.995201	-89.632801	226295.261253	2.323871e+06	20	164.764716
8	20.995231	-89.630879	226495.261253	2.323871e+06	20	231.776286
9	20.995260	-89.628957	226695.261253	2.323871e+06	15	337.737751

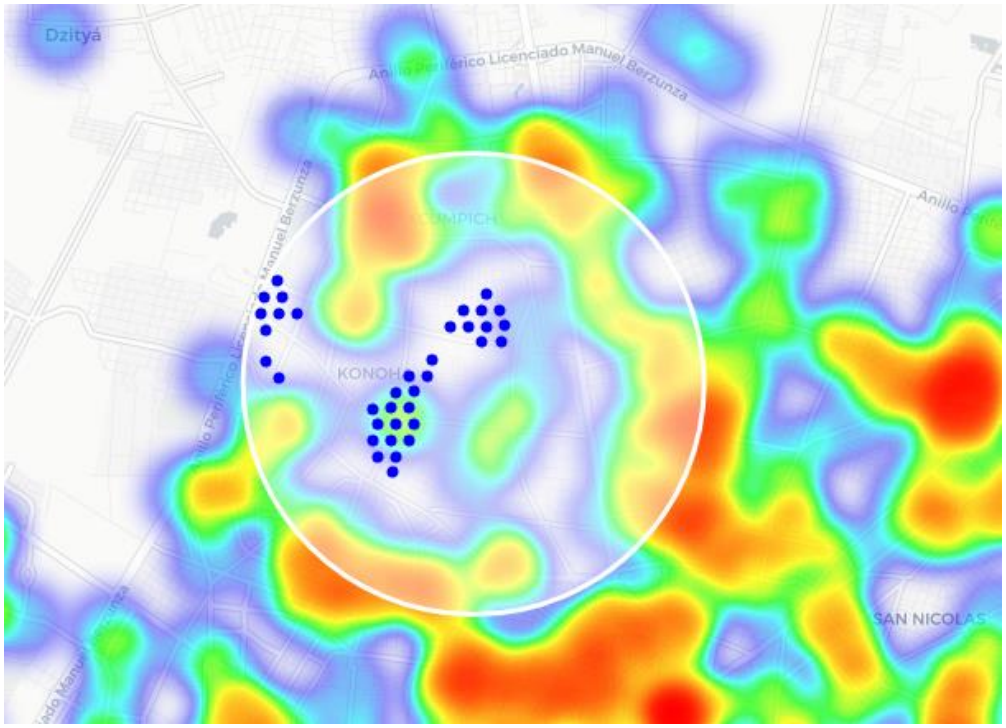
Then we **filtered** those locations: we were interested in only in **locations with no more than 4 medical centers in radius of 500 meters**, and **no hospitals in radius of 500 meters**. We obtained the following totals:

Locations with no more than four medical centers nearby: 43

Locations with no hospitals within 500m: 242

Locations with both conditions met: 36

Then we showed them on a map:

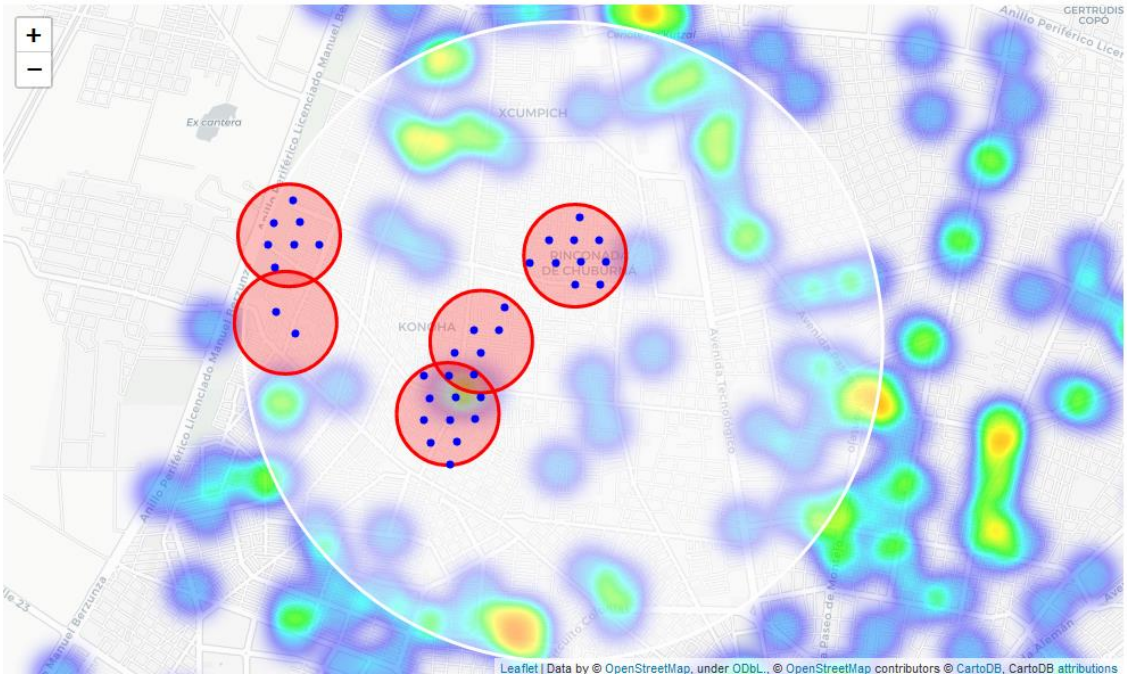


We now had a bunch of locations north-west to Merida Centro, mostly in **Barrio de Chuburna (Rinconada de Chuburna, San Pedro Uxmal and San Vicente Chuburna), Camara de la Construccion and some parts south from it**. We know that each of those locations has no more than four medical centers in radius of 500m, and no hospitals closer than 500m. Any of those locations is a potential candidate for a new hospital, at least based on nearby competition.

Again, we showed those good locations in a form of heatmap:

What we had at that point was a clear indication of zones with low number of medical centers in vicinity, and no hospitals at all nearby.

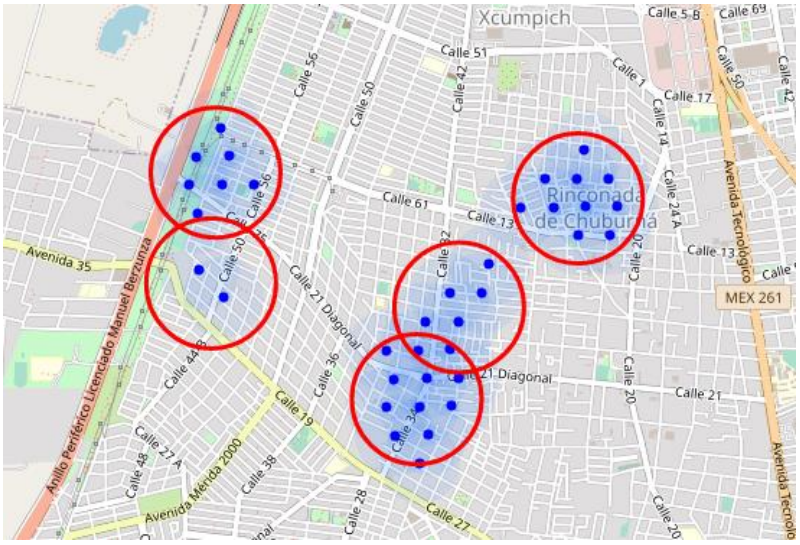
We then **clustered** those locations to create **centers of zones containing good locations**. Those zones, their centers and addresses would be the final result of our analysis.



Our clusters represented groupings of most of the candidate locations and cluster centers were placed nicely in the middle of the zones ‘rich’ with location candidates.

Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

We decided to see those zones on a city map without heatmap, using shaded areas to indicate our clusters:

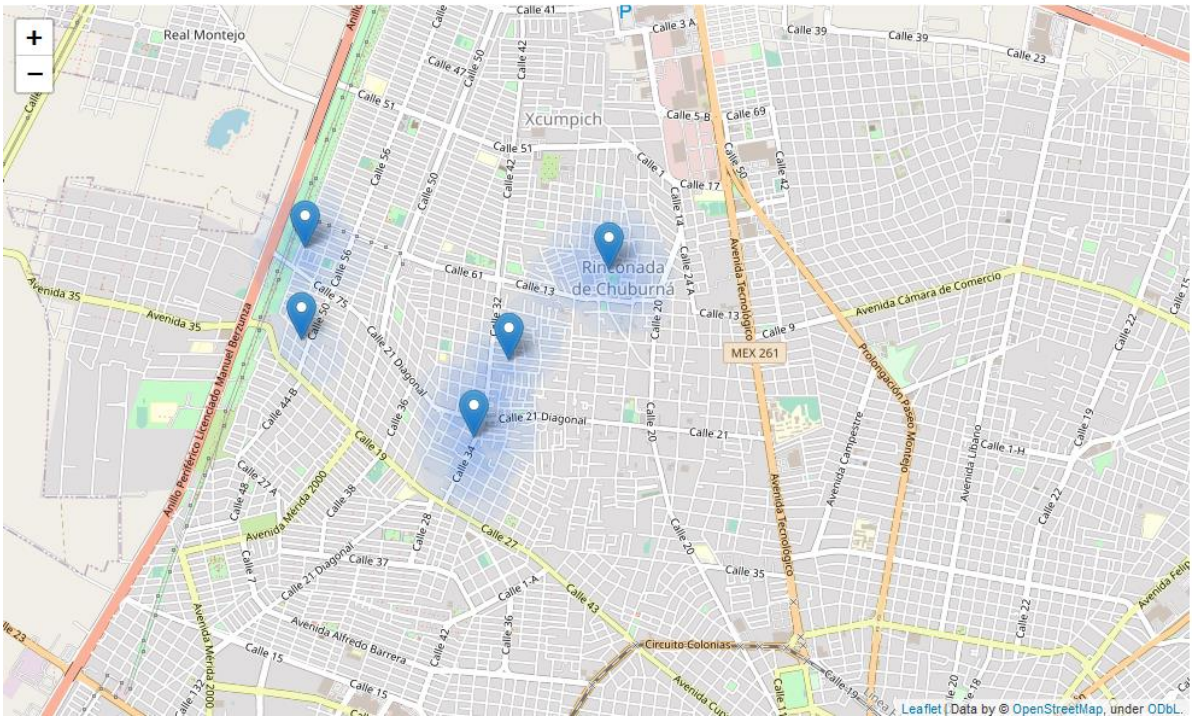


Finally, we **reverse geocoded those candidate area centers to get the addresses** which can be presented to stakeholders. We obtained the following addresses as centers of areas recommended for further analysis:

- **Calle 60 460, Fracc Camara de la Construcción, Mérida, Yucatán, 97203**
- **Calle 17A 101D, Fraccionamiento Chuburna de Hidalgo, Mérida, Yucatán, 97208**
- **Calle 21B 308, San Vicente Chuburna, Mérida, Yucatán, 97206**
- **Calle 9 99-99, Fracc Rinconada de Chuburna, Mérida, Yucatán, 97208**
- **Calle 29 Diagonal 412, Merida o Elefante Grande, Mérida, Yucatán, 97203**

This concluded our analysis. We had created **5 addresses** representing centers of zones containing locations with low number of medical centers and no hospitals nearby, all zones being north to city center (all between 5.0km and 7.5km from Merida Centro). Although zones were shown on a map with a radius of ~400 meters (red circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential hospital locations. Most of the zones are located in Barrio de Chuburna and Camara de la Construccion neighborhood, which we have identified as interesting due to being 'middle class' and north to city center.

We could prioritize our search beginning with the three addresses located in Barrio de Chuburna since it is slightly more interesting than the other ones due to its sociodemographic profile (type 'C').



1.6 Results and Discussion

Our analysis showed that although there is an important number of medical centers in Merida, Yucatan (~550 in our initial area of interest, which was 24x24km around Merida Centro), there are pockets of low medical center density north-west and south to city center. Highest concentration of medical was detected exactly north from Merida Centro, so we focused our attention to areas north-west excluding south, corresponding to Barrio de Chuburna (Rinconada de Chuburna, San Pedro Uxmal and San Vincente Chuburna), Camara de la Construccion neighborhood and a couple of areas south of it.

Something notable was the fact that south of the city and some areas near but outside the city have a really low offer of medical services in general and hospital services in particular. This might be an interesting opportunity for another study and a business opportunity in itself.

After directing our attention to this more narrow area of interest (covering approx. 5x5km north-west from Merida Centro) we first created a dense grid of location candidates (spaced 200m apart); those locations were then filtered so that those with more than 4 medical centers in radius of 500m and those with a hospital closer than 500m were removed.

Those location candidates were then clustered to create zones of interest, which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is five zones containing largest number of potential new medical center locations based on number of and distance to existing venues - both medical centers in general and hospitals particularly. This, of course, does not imply that those zones are actually optimal locations for a new hospital! Purpose of this analysis was to only provide info on areas north to Merida center but not crowded with existing medical centers (particularly hospitals). It is entirely possible that there is a very good reason for small number of medical centers in any of those areas, reasons that would make them unsuitable for a new medical center regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis, which could eventually result in location, which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

1.7 Conclusion

Purpose of this project was to identify Merida areas north to center with low number of medical centers (particularly hospitals) in order to aid stakeholders in narrowing down the search for optimal location for a new hospital. By calculating medical center density distribution from Foursquare data we have first identified general areas that justify further analysis (Barrio de Chuburna and Camara de la Construccion), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby medical centers. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal medical center location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to basic services like water, electricity, garbage collection, etc.), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood, etc.