

# Capstone Project - Where to open a health facility in Sindh

Rahim Hasnani

August 2020

## Introduction & Background

In a developing country like Pakistan, health facilities are not available and at times people have to travel a great deal to see a medical practitioner or expert pertaining to their needs. With scarce resources, it is important to make a right decision when opening a health facility to ensure maximum accessibility and coverage. From a business point of view, one would open a facility where population density is higher and there is little competition. In contrast, from an accessibility point of view, the facility should be located where there is little or no alternative available nearby.

Therefore, the factors to consider are:

- Population in the vicinity (within x miles/kilometers)
- Presence of facilities in the vicinity

## Data

The scope is restricted to districts of Sindh Province. Pakistan has four provinces (and other administrative units). Sindh Province has 29 districts.

FourSquare API is used to explore presence of Medical centres; It is one of the parent categories and include various sub-categories. Geopy would be used to get geo coordinates of districts.

## Methodology

### 1. Get Data about Districts of Sindh

Data about districts of Sindh province was found here:

[https://en.wikipedia.org/wiki/Districts\\_of\\_Sindh](https://en.wikipedia.org/wiki/Districts_of_Sindh)

### 2. Use Geopy to get coordinates

Geopy was used for coordinates. Some of them were incorrect and were manually corrected. Nomination data address search provided wrong coordinates. Some were fixed by changing the spelling of the district names while others were manually corrected by reconfirming the coordinates from google maps manually.

Map	Sr. No.	District	Headquarters	Area (km²)	Population (in 2017)	Density (people/km²)	Ing	lat	
0	NaN	1	Badin	Badin	6470.0	1804516.0	279.0	68.840151	24.655167
1	NaN	2	Dadu	Dadu	8034.0	1550266.0	193.0	67.771833	26.732137
2	NaN	3	Ghotki	Mirpur Mathelo	6506.0	1647239.0	253.0	69.314110	28.002629
3	NaN	4	Hyderabad	Hyderabad	1022.0	2201079.0	2155.0	68.375038	25.380102
4	NaN	5	Jacobabad	Jacobabad	2771.0	1006297.0	363.0	68.436436	28.281309
5	NaN	6	Jamshoro	Jamshoro	11250.0	993142.0	88.0	68.266172	25.400723
6	NaN	7	Karachi Central	Karachi	62.0	2972639.0	48336.0	66.978425	24.944166
7	NaN	8	Kashmore	Kandhkot	2551.0	1089169.0	427.0	69.581475	28.432292
8	NaN	9	Khairpur	Khairpur	15925.0	2405523.0	151.0	68.763411	27.529540
9	NaN	10	Larkana	Larkana	1906.0	1524391.0	800.0	68.210151	27.556480
10	NaN	11	Matlari	Matlari	1459.0	769349.0	527.0	68.445487	25.597186
11	NaN	12	Mirpur Khas	Mirpur Khas	3319.0	1505876.0	454.0	69.011062	25.526282
12	NaN	13	Naushahro Feroze	Naushahro Feroze	2027.0	1612373.0	369.0	68.128692	26.849123
13	NaN	14	Naushahro Feroze	Naushahro Feroze	2027.0	1612373.0	369.0	68.128692	26.849123

### 3. Find Medical Centers near the districts

Foursquare API was used to search medical centres nearby. All medical centers near 100 KM of each district were extracted. When searched within 10 KM a lot of district did not have any data. This may be because of incomplete data in Foursquare in this part of the world.

	District	Latitude	Longitude	Venue Address	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Badin	24.655167	68.840151		Afzal medical centre	25.401089	68.321594	Hospital
1	Badin	24.655167	68.840151		Saddar Doctor's Line	25.398275	68.368510	Medical Center
2	Badin	24.655167	68.840151		Wali Bhai Rajputana Hospital	25.417656	68.345596	Hospital
3	Badin	24.655167	68.840151	Near Tower Market	Liaquat University Hospital	25.403516	68.367846	Hospital
4	Badin	24.655167	68.840151	LUMHS Colloney	Civil Hospital Jamshoro	25.431655	68.276126	Hospital
5	Badin	24.655167	68.840151		Dr Azam Clinic	25.402332	68.375298	Doctor's Office
6	Badin	24.655167	68.840151	Latifabad #6	Hilal-e-Ahmer Hospital	25.365412	68.352836	Hospital
7	Badin	24.655167	68.840151		Civil Hospital Hyderabad	25.401903	68.367101	Hospital
8	Badin	24.655167	68.840151	Near CTS Coaching Centre	G.G Jhagrani Medicare	25.400939	68.370070	Medical Center
9	Badin	24.655167	68.840151	Autobhan	Dr Safdar	25.388852	68.349872	Hospital
10	Badin	24.655167	68.840151	Saddar	Doctor's Line	25.401572	68.368609	Doctor's Office
11	Dadu	26.732137	67.771833	Civil Hospital Rd	Civil Hospital Nawabshah	26.244447	68.406361	Medical Center
12	Dadu	26.732137	67.771833		JOHI	26.692457	67.609588	Doctor's Office
13	Dadu	26.732137	67.771833	near Ladies Jail	national medical centre	27.553698	68.213364	Medical Center
14	Dadu	26.732137	67.771833		LINAR	27.560718	68.201699	Hospital

	Venue	Venue Category
District		
Badin	11	3
Dadu	5	3
Ghotki	13	4
Hyderabad	13	3
Jacobabad	9	4
Jamshoro	47	6
Karachi Central	49	6
Karachi East	49	6
Karachi South	49	6
Karachi West	49	6
Kashmore	5	3
Khairpur	9	4
Korangi	49	6
Larkana	9	4
Malir	40	6

#### 4. Compute distance to medical centers

Using the venue latitude and longitude against district coordinates, distance to medical centers was calculated.

In addition, the proximity was converted to discrete clusters by identifying whether the medical center is sub 10KM, sub 20KM, sub 50KM from the district.

	District	Latitude	Longitude	Venue Address	Venue	Venue Latitude	Venue Longitude	Venue Category	distance	Sub10K	Sub20K	Sub50K
0	Badin	24.655167	68.840151		Afzal medical centre	25.401089	68.321594	Hospital	98.056911	0	0	0
1	Badin	24.655167	68.840151		Saddar Doctor's Line	25.398275	68.368510	Medical Center	95.349407	0	0	0
2	Badin	24.655167	68.840151		Wali Bhai Rajputana Hospital	25.417656	68.345596	Hospital	98.371765	0	0	0
3	Badin	24.655167	68.840151	Near Tower Market	Liaquat University Hospital	25.403516	68.367846	Hospital	95.887845	0	0	0
4	Badin	24.655167	68.840151	LUMHS Colloney	Civil Hospital Jamshoro	25.431655	68.276126	Hospital	103.392746	0	0	0
5	Badin	24.655167	68.840151		Dr Azam Clinic	25.402332	68.375298	Doctor's Office	95.402671	0	0	0
6	Badin	24.655167	68.840151	Latifabad #6	Hilal-e-Ahmer Hospital	25.365412	68.352836	Hospital	93.026424	0	0	0
7	Badin	24.655167	68.840151		Civil Hospital Hyderabad	25.401903	68.367101	Hospital	95.769651	0	0	0
8	Badin	24.655167	68.840151	Near CTS Coaching Centre	G.G Jhagrani Medicare	25.400939	68.370070	Medical Center	95.528016	0	0	0
9	Badin	24.655167	68.840151	Autobhan	Dr Safdar	25.388852	68.349872	Hospital	95.402497	0	0	0
10	Badin	24.655167	68.840151	Saddar	Doctor's Line	25.401572	68.368609	Doctor's Office	95.662177	0	0	0
11	Dadu	26.732137	67.771833	Civil Hospital Rd	Civil Hospital Nawabshah	26.244447	68.406361	Medical Center	83.264131	0	0	0
12	Dadu	26.732137	67.771833		JOHI	26.692457	67.609588	Doctor's Office	16.713772	0	1	1
13	Dadu	26.732137	67.771833	near Ladies Jail	national medical centre	27.553698	68.213364	Medical Center	101.294414	0	0	0
14	Dadu	26.732137	67.771833		LINAR	27.560718	68.201699	Hospital	101.509525	0	0	0

#### 5. Preparing data for Visual Analysis and Clustering

The data was grouped and pivoted to show venue categories as Medical Center is a parent category of Foursquare API. Count of each type of medical center for each district were identified.

Furthermore, count of each medical center against the proximity clusters were also identified, as shown below.

District		Venue							Sub10K		...
Venue Category		Dentist's Office	Doctor's Office	Emergency Room	Eye Doctor	Hospital	Medical Center	Medical Lab	Dentist's Office	Doctor's Office	...
0	Badin	0.0	2.0	0.0	0.0	7.0	2.0	0.0	0.0	0.0	...
1	Dadu	0.0	1.0	0.0	0.0	2.0	2.0	0.0	0.0	0.0	...
2	Ghotki	0.0	2.0	1.0	0.0	7.0	3.0	0.0	0.0	0.0	...
3	Hyderabad	0.0	2.0	0.0	0.0	8.0	3.0	0.0	0.0	2.0	...
4	Jacobabad	0.0	2.0	1.0	0.0	4.0	2.0	0.0	0.0	0.0	...

5 rows × 29 columns

## 6. Clustering of districts

Based on the pivoted data, clustering was done with 4 clusters:

District	lat	lng	Cluster Labels	Venue Dentist's Office	Venue Doctor's Office	Venue Emergency Room	Venue Eye Doctor	Venue Hospital	Venue Medical Center	...
0	Badin	24.655167 68.840151	1.0	0.0	2.0	0.0	0.0	7.0	2.0	...
1	Dadu	26.732137 67.771833	1.0	0.0	1.0	0.0	0.0	2.0	2.0	...
2	Ghotki	28.002629 69.314110	1.0	0.0	2.0	1.0	0.0	7.0	3.0	...
3	Hyderabad	25.380102 68.375038	1.0	0.0	2.0	0.0	0.0	8.0	3.0	...
4	Jacobabad	28.281309 68.436436	1.0	0.0	2.0	1.0	0.0	4.0	2.0	...
5	Jamshoro	25.400723 68.266172	2.0	5.0	7.0	0.0	1.0	23.0	9.0	...
6	Karachi Central	24.944166 66.978425	3.0	5.0	6.0	1.0	0.0	33.0	3.0	...
7	Kashmore	28.432292 69.581475	1.0	0.0	1.0	0.0	0.0	3.0	1.0	...
8	Khairpur	27.529540 68.763411	1.0	0.0	2.0	1.0	0.0	4.0	2.0	...
9	Larkana	27.556480 68.210151	1.0	0.0	2.0	1.0	0.0	4.0	2.0	...
10	Matlari	25.597186 68.445487	1.0	0.0	2.0	0.0	0.0	8.0	3.0	...
11	Mirpur Khas	25.526282 69.011062	1.0	0.0	2.0	0.0	0.0	8.0	3.0	...
12	Naushahro Feroze	26.849123 68.128692	1.0	0.0	2.0	1.0	0.0	5.0	3.0	...
13	Naushahro Feroze	26.245202 68.404022	1.0	0.0	2.0	0.0	0.0	8.0	3.0	...

# Results

## Districts without any medical centers

Out of the 29 districts, 2 districts Tharparker and Umarkot did not have any medical centers based on Foursquare data in the vicinity of 100KM.

index	District		lat	Ing	Venue
0	22	Umarkot	25.365530	69.740126	0.0
1	20	Tharparkar	24.943186	70.241012	0.0
2	1	Dadu	26.732137	67.771833	5.0
3	7	Kashmore	28.432292	69.581475	5.0
4	14	Qambar	27.589243	67.999675	9.0
5	4	Jacobabad	28.281309	68.436436	9.0
6	8	Khairpur	27.529540	68.763411	9.0
7	9	Larkana	27.556480	68.210151	9.0
8	17	Sukkur	27.696188	68.858875	9.0
9	16	Shikarpur	27.957798	68.646551	9.0
10	19	Tando Muhammad Khan	25.123574	68.533519	11.0
11	18	Tando Allahyar	25.293473	68.609599	11.0
12	0	Radin	24.655167	68.840151	11.0

### Metropolitan areas have more centers

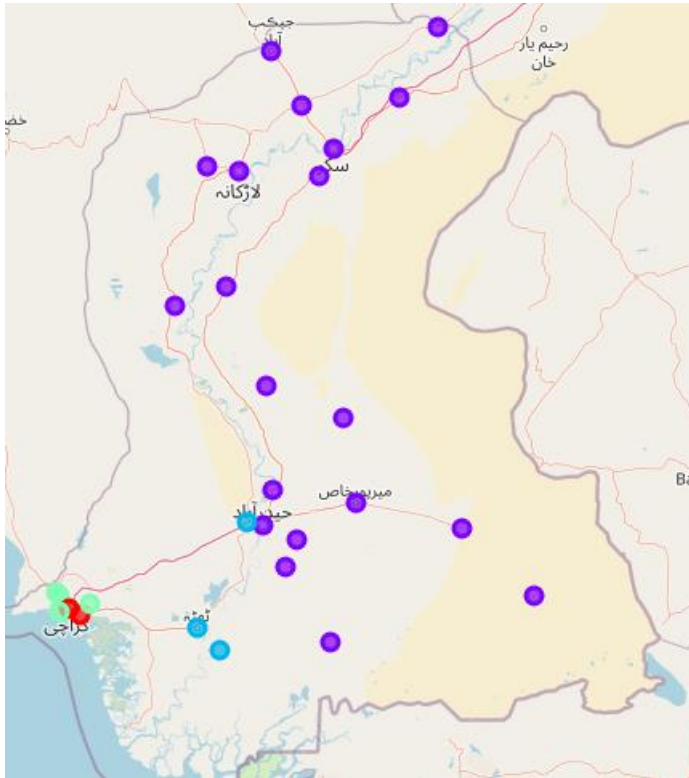
As expected, the big metropolis (districts of Karachi) and the districts nearby had more medical centers:

index	District		lat	Ing	Venue
28	28	Malir	24.895919	67.196919	49.0
27	26	Karachi West	24.969503	66.959243	49.0
26	25	Karachi South	24.844128	66.980173	49.0
25	24	Karachi East	24.870894	67.046654	49.0
24	23	Sujawal	24.603377	68.078900	49.0
23	21	Thatta	24.746900	67.924028	49.0
22	6	Karachi Central	24.944166	66.978425	49.0
21	27	Korangi	24.826819	67.129813	49.0
20	5	Jamshoro	25.400723	68.266172	47.0
19	13	Nawabshah	26.245292	68.404023	14.0
15	11	Mirpur Khas	25.526282	69.011062	13.0
18	2	Ghotki	28.002629	69.314110	13.0
17	3	Hyderabad	25.380102	68.375038	13.0
16	10	Matiari	25.597186	68.445487	13.0

## Clustering Results

As seen on the cluster map, the biggest city of the country, Karachi forms two clusters. The other cluster (blue) is formed by the second biggest city (Hyderabad) as well as districts near Karachi. Rest of the districts are made into one cluster.

Note: For the purpose of clustering, two districts with no data were assigned the purple cluster for visualization.



## Discussion

### Severe lack of facilities in rural areas

There are only 12 districts where medical centers are present within 20KM radius. Unless the data available is grossly incomplete, the situation of medical facilities in rural center is grave!

	index	District	Sub10K	Sub20K
0	7	Karachi East	43	49
1	12	Korangi	27	49
2	8	Karachi South	15	45
3	3	Hyderabad	10	11
4	14	Malir	7	47
5	6	Karachi Central	5	47
6	5	Jamshoro	5	9
7	11	Khairpur	4	4
8	9	Karachi West	2	45
9	13	Larkana	2	2
10	18	Nawabshah	2	2
11	23	Sukkur	2	2
12	1	Dadu	0	1

### Clustering interpretation

When looking at the three clusters (all but the purple one) below, we see that Cluster zero has the most hospitals in 10KM radius, whereas cluster three has more hospitals in 10KM and 20KM radius than cluster two districts.

The biggest city, Karachi is broken down into six districts (Central, East, South, West, Korangi and Malir) and they exclusively form cluster zero and three whereas the districts near Karachi form cluster two. Rest of the province is in cluster one.

Cluster one districts need more medical facility. This is mostly consistent with the table above.

	District	Cluster Labels	Sub10K Hospital	Sub20K Hospital	Venue Hospital
5	Jamshoro	2.0	5.0	7.0	23.0
6	Karachi Central	3.0	4.0	31.0	33.0
21	Thatta	2.0	0.0	0.0	33.0
23	Sujawal	2.0	0.0	0.0	33.0
24	Karachi East	0.0	27.0	33.0	33.0
25	Karachi South	3.0	11.0	29.0	33.0
26	Karachi West	3.0	1.0	30.0	33.0
27	Korangi	0.0	17.0	33.0	33.0
28	Malir	3.0	5.0	31.0	33.0

# Limitations

The results of the work depend on the accuracy of foursquare API data.

# Future Directions

The analysis done is very limited and can be extended in various directions.

Better illustrations of maps with marked districts would add to readability and visualization. The number of medical facilities can be compared against the population and the area of the district.

More analysis of different types of medical centers could be carried out.

In addition, data provided by foursquare is limited to near a geo location. Perhaps, location based on area of a district or a province would be more useful.