

Data Science Capstone

Rahim Hasnani

August 2020

Problem: Where to open a health facility
in Sindh

Problem Statement

- ▶ Sindh Province is in Pakistan
- ▶ It has 29 districts
- ▶ Being a developing country, health facilities are scarce
- ▶ People have to travel great distance to access healthcare
- ▶ The motivation behind the project is to identify districts where health facilities should be opened/provided first.

Data Sources

- ▶ Wikipedia to get list of districts of Sindh:
https://en.wikipedia.org/wiki/Districts_of_Sindh
- ▶ Geopy or google map for geo coordinate of the 29 districts
- ▶ FourSquare API to provide venue data for 'Medical Center' category near the districts

Methods & Tools

- ▶ Scrapping of data from Wikipedia page using pandas
- ▶ Correction & validation of geo coordinates
- ▶ Exploratory data analysis on healthcare facilities data from Foursquare API
- ▶ Grouping of data for exploratory data analysis
- ▶ Calculation of distance from district to medical facility
- ▶ Distance partitioning on the basis of within 10KM, within 20KM, withing 50M and more than 50KM
- ▶ Pivoting/reframing data for clustering analysis

Districts and geo locations

	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	Matiali	Matiali	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	Naushahro	4619.0	2012047.0	240.0	68.404022	26.245202

Facilities data from Foursquare

	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

Summary of total facilities and distinct categories

	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

Distance to facility and further categorization (within 10KM, 20KM....)

	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

Results & Conclusions

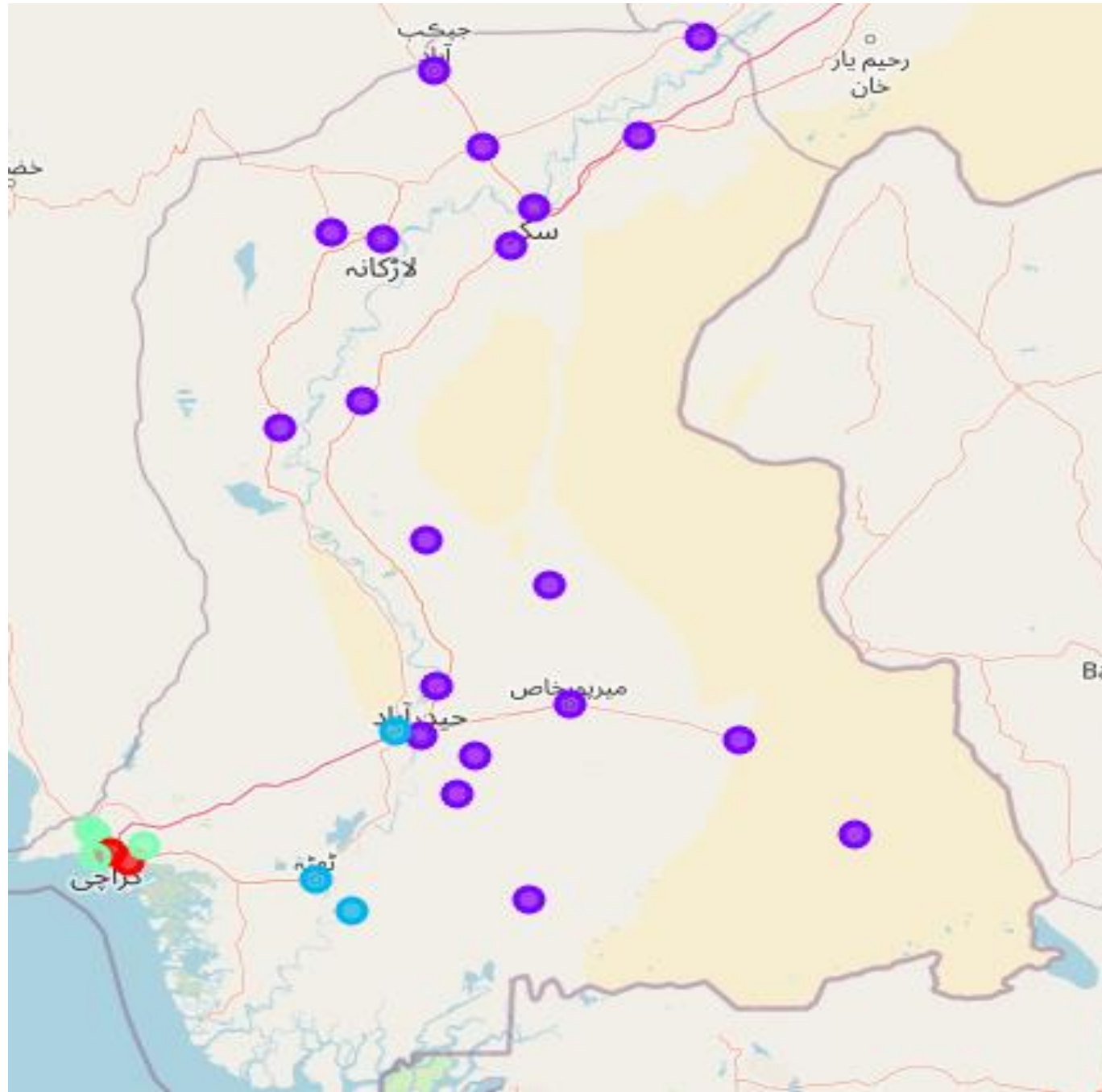
Districts without any medical facility

	index	District		lat	lng	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	Badin		24.655167	68.840151	11.0

Metropolitan areas have more 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	Matari	25.597186	68.445487	13.0

Clustering Results



Discussion: Severe lack of facilities in rural areas

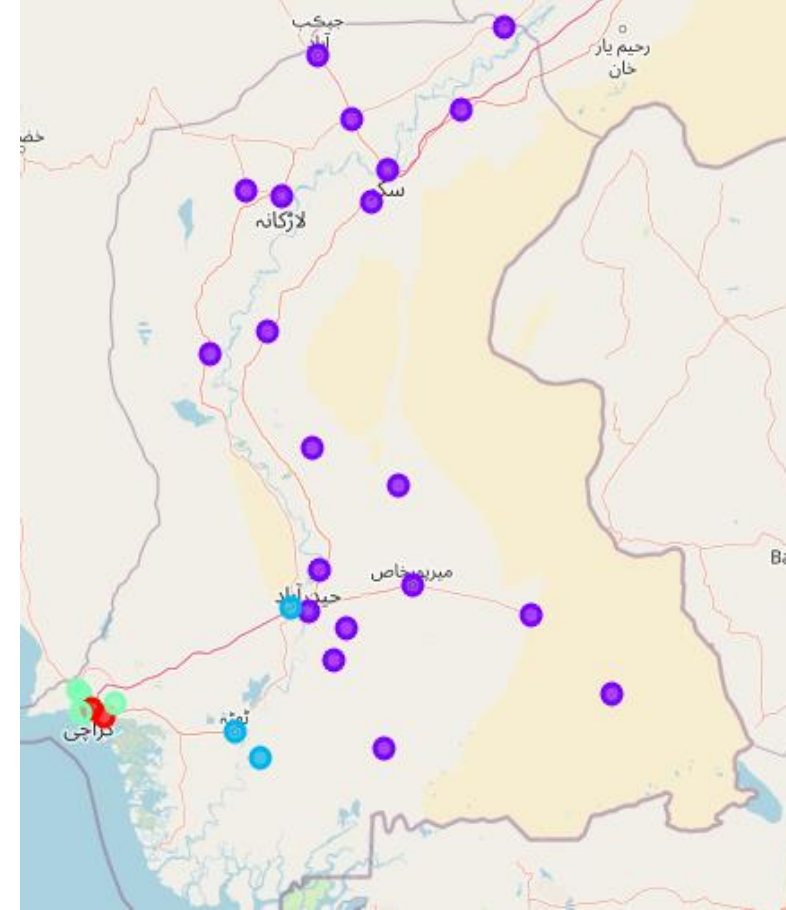
- ▶ ONLY 12 DISTRICTS WITH MEDICAL FACILITIES WITHIN 20km RADIUS!

	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.
- ▶ Purple districts need more medical facility. This is mostly consistent with last slide

	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



THANK YOU!