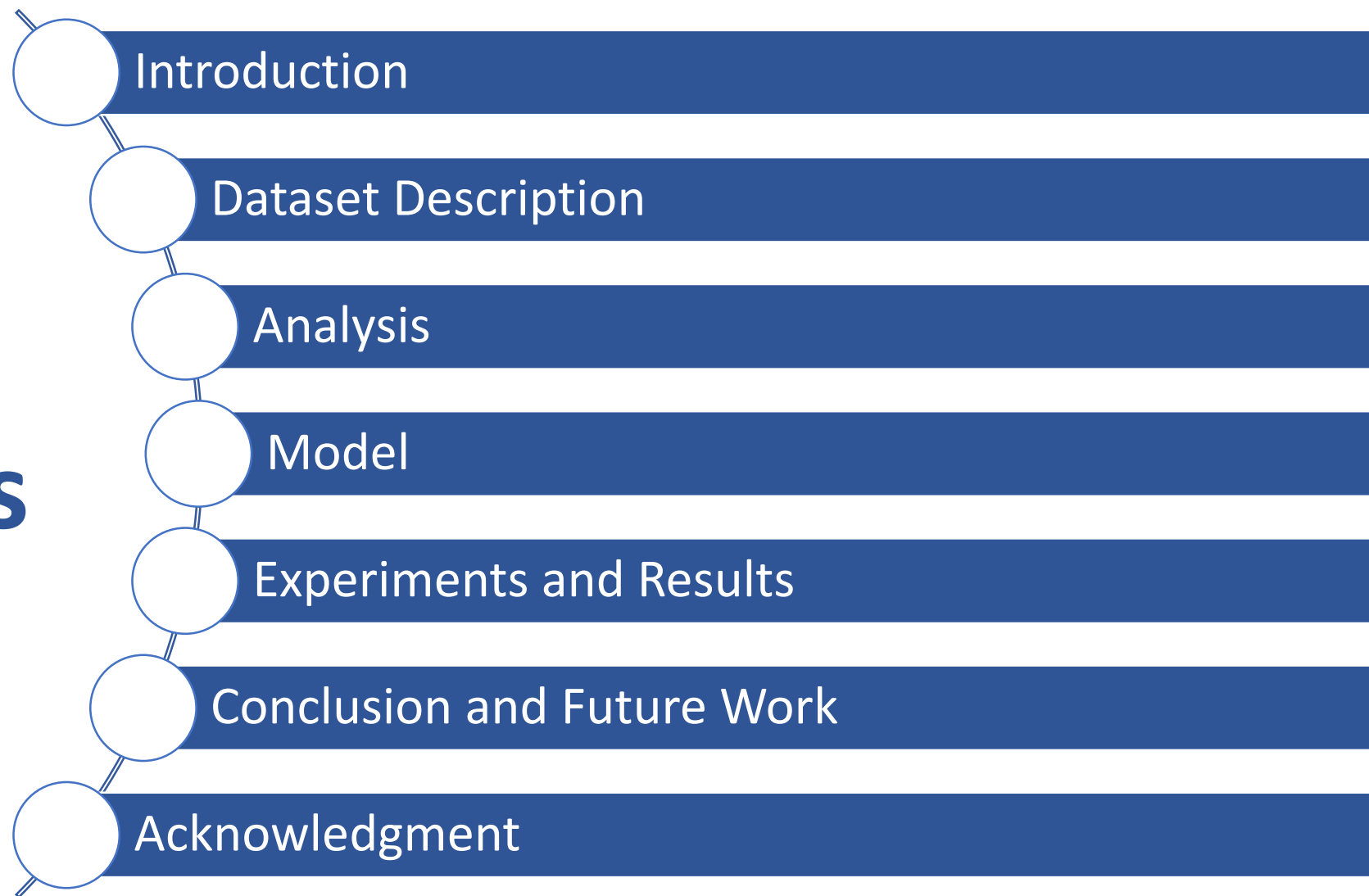


**Clustering model for Saudi Arabia cities
according to its population, Covid 19 cases and
total of hospitals**

Capstone Project - The Battle of Neighborhoods (Week 2)

Abdulwahed Salam

Outlines



Introduction

Saudi Arabic is a country located in middle east. it consists of many states and each state consist of many cities. KSA play a major role to control the covid 19 epidemic. Nowadays, the total number of cases around 250 up to 350. Now it is time of getting the vaccine and distributed to all people in different locations. As we know the vaccine is limited and to distributed it fairly is important. we have many cities with different features. Based on that we have to build a model to cluster the cities. Also, the vaccine needs to be store in specific conditions. one of them is cold and must to be minus 85. KSA like many countries, there are small cities or region does not have hospital and we want to consider this issue.

Dataset

Daily cases of covid 19 in Saudi Arabia

Day	Date	Case	Cure	Death
0	الخميس 20-12-17	174	208	10
1	الأربعاء 20-12-16	181	173	11
2	الثلاثاء 20-12-15	180	199	11
3	الاثنين 20-12-14	142	201	10
4	الأحد 20-12-13	125	243	11

KSA Cities Hospital according to Wiki Data

Hospital	City	State	lat	long
0	مستشفى الملك خالد	نجران	منطقة نجران	17.545267 44.233758
1	مستشفى نجران العام	نجران	منطقة نجران	17.495299 44.144086
2	مستشفى الولادة والأطفال	نجران	منطقة نجران	17.553579 44.272491
3	مستشفى شروره	شروره	منطقة نجران	17.484406 47.101262
4	مستشفى حيونا العام	حيونا	منطقة نجران	17.847307 44.019830

Covid 19 Cases according to KSA Cities

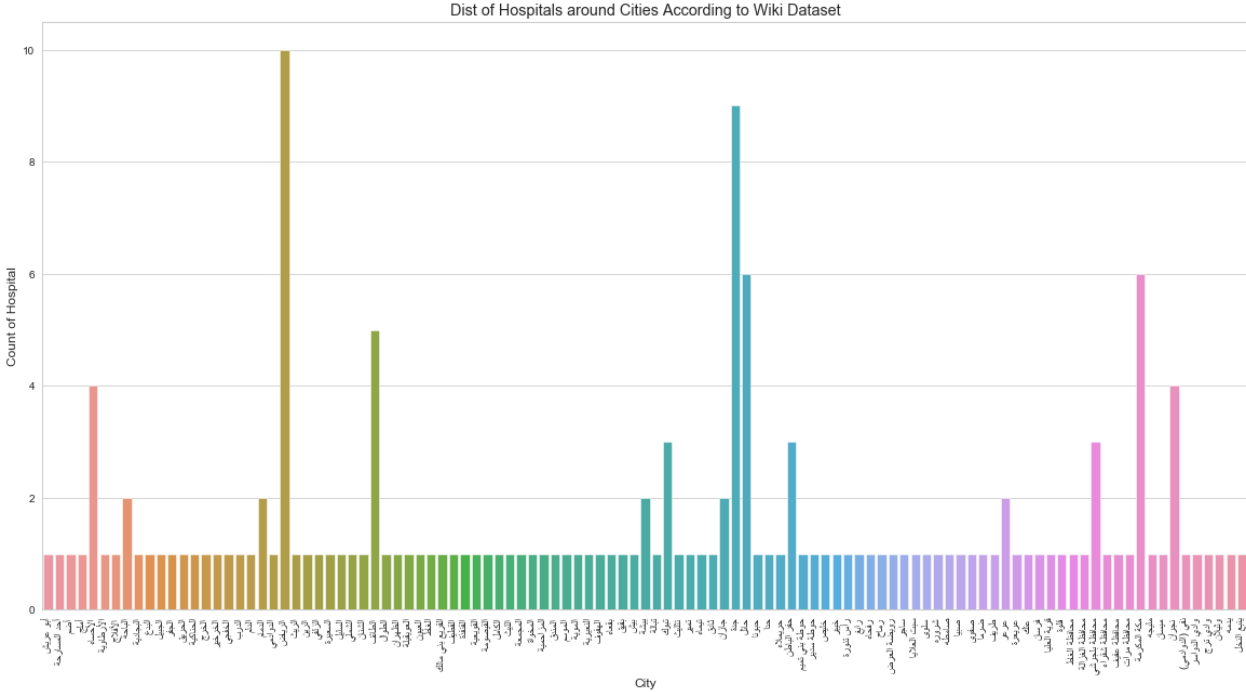
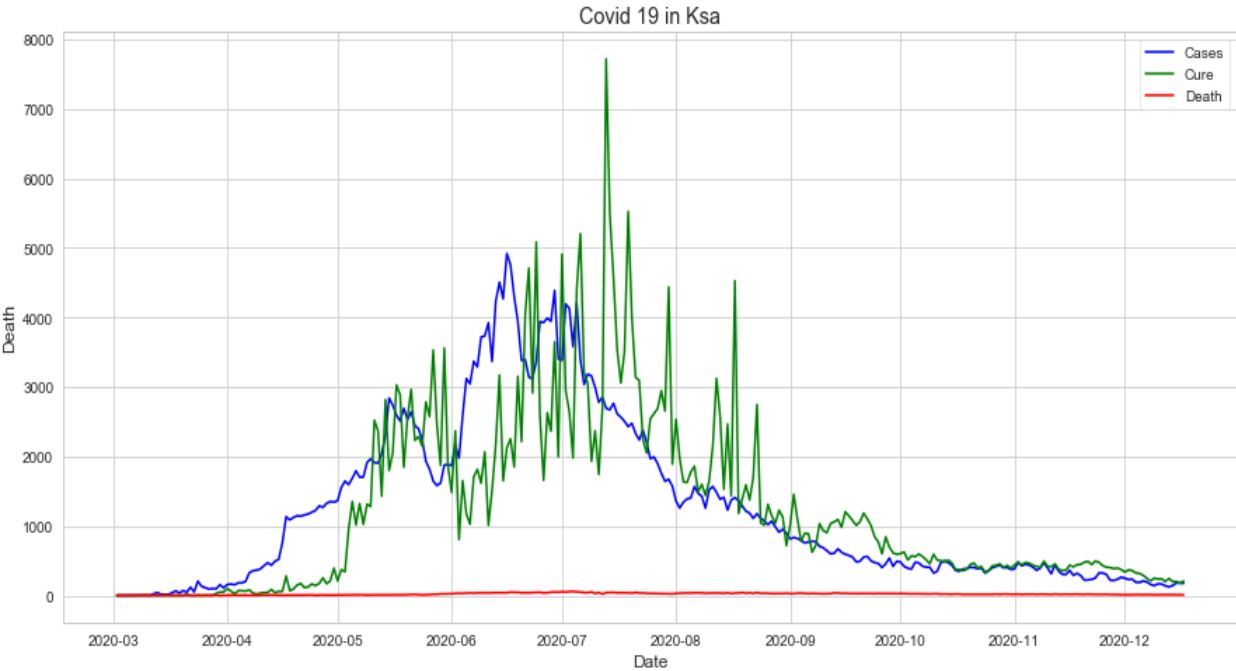
	City	State	population	lat	long
0	الرياض	الرياض	5188286	24.713552	46.675296
1	جدة	مكة المكرمة	3430697	21.485811	39.192505
2	مكة المكرمة	مكة المكرمة	1534731	21.389082	39.857912
3	المدينة المنورة	المدينة المنورة	1100093	24.524654	39.569184
4	الدمام	المنطقة الشرقية	903312	26.420683	50.088794

Hospital information according to Foursquare API

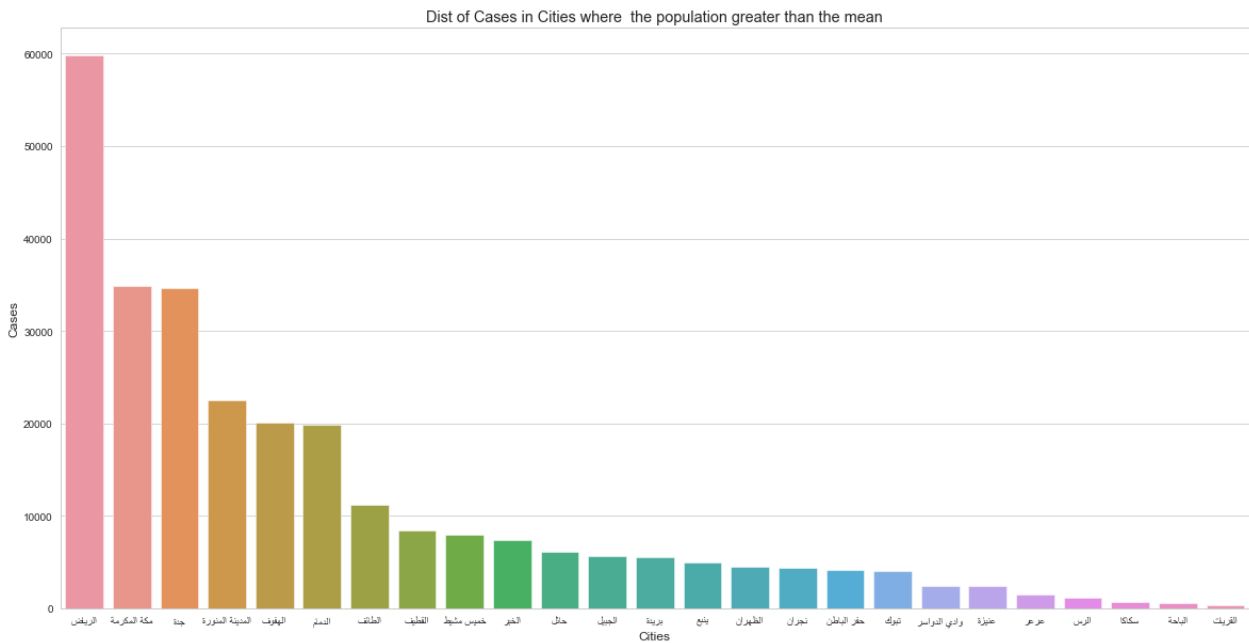
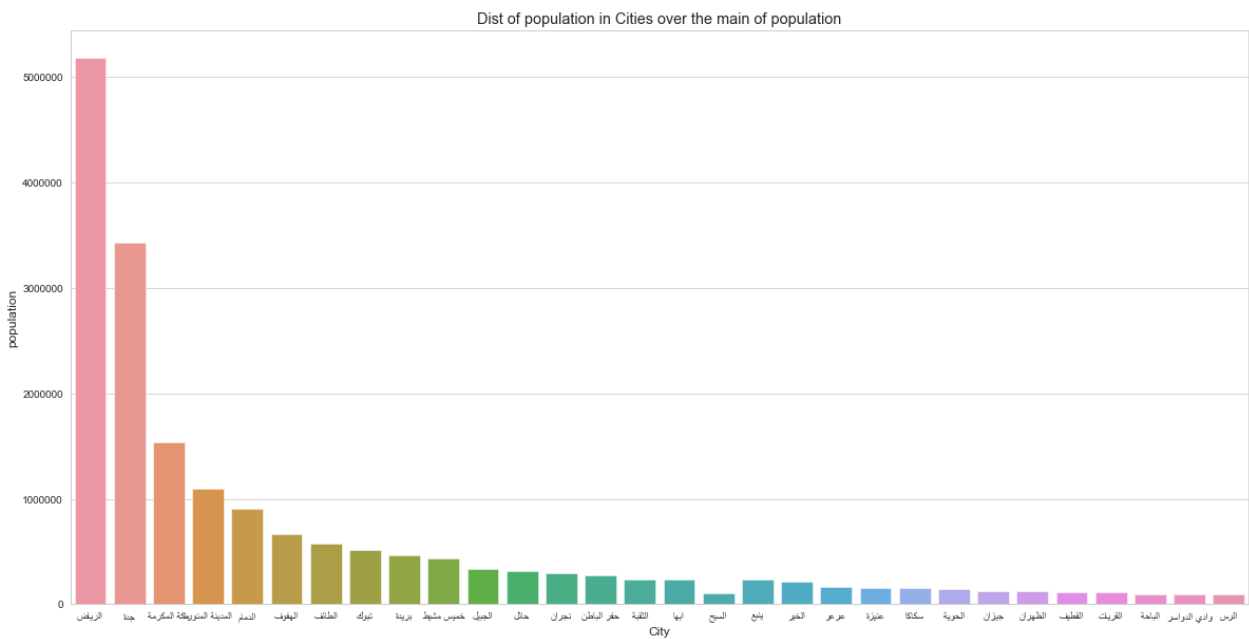
	categories	hasPerk	id	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat
0	Hospital	False	4f5e9dd1e4b033732b5efaf6	Ar-Rahmanyah	SA	الرياض	المملكة العربية السعودية	Takhassusi Street	1961	[Ar-Rahmanyah (Takhassusi Street), الرياض 1234...	[[{"label": "display", "lat": 24.719242, "long": 46.675296}]]	24.719242
1	Hospital	False	4bdf3036e75c0f47f033ca03	Takhassusi St.	SA	الرياض	المملكة العربية السعودية	Khurais Rd.	4661	[Takhassusi St. (Khurais Rd.), الرياض 11211...	NaN	24.671692
2	Hospital	False	4df9d62ad22d964c6b9b133a	King Abdullah Rd.	SA	الرياض	المملكة العربية السعودية	King Khalid Rd.	5422	[King Abdullah Rd. (King Khalid Rd.), الرياض 1...	NaN	24.712649
3	Hospital	False	4b9ba5c4f964a5207f1636e3	King Fahd Rd.	SA	الرياض	المملكة العربية السعودية	NaN	4383	[King Fahd Rd., الرياض 12381, المملكة العربية...	[[{"label": "display", "lat": 24.746909, "long": 46.675296}]]	24.746909
4	Hospital	False	4eeca9a29c28028ddd9bdcdb	Khurais Road	SA	الرياض	المملكة العربية السعودية	NaN	10089	[Khurais Road, الرياض 11635, المملكة العربية...	[[{"label": "display", "lat": 24.721248, "long": 46.675296}]]	24.721248

Analysis

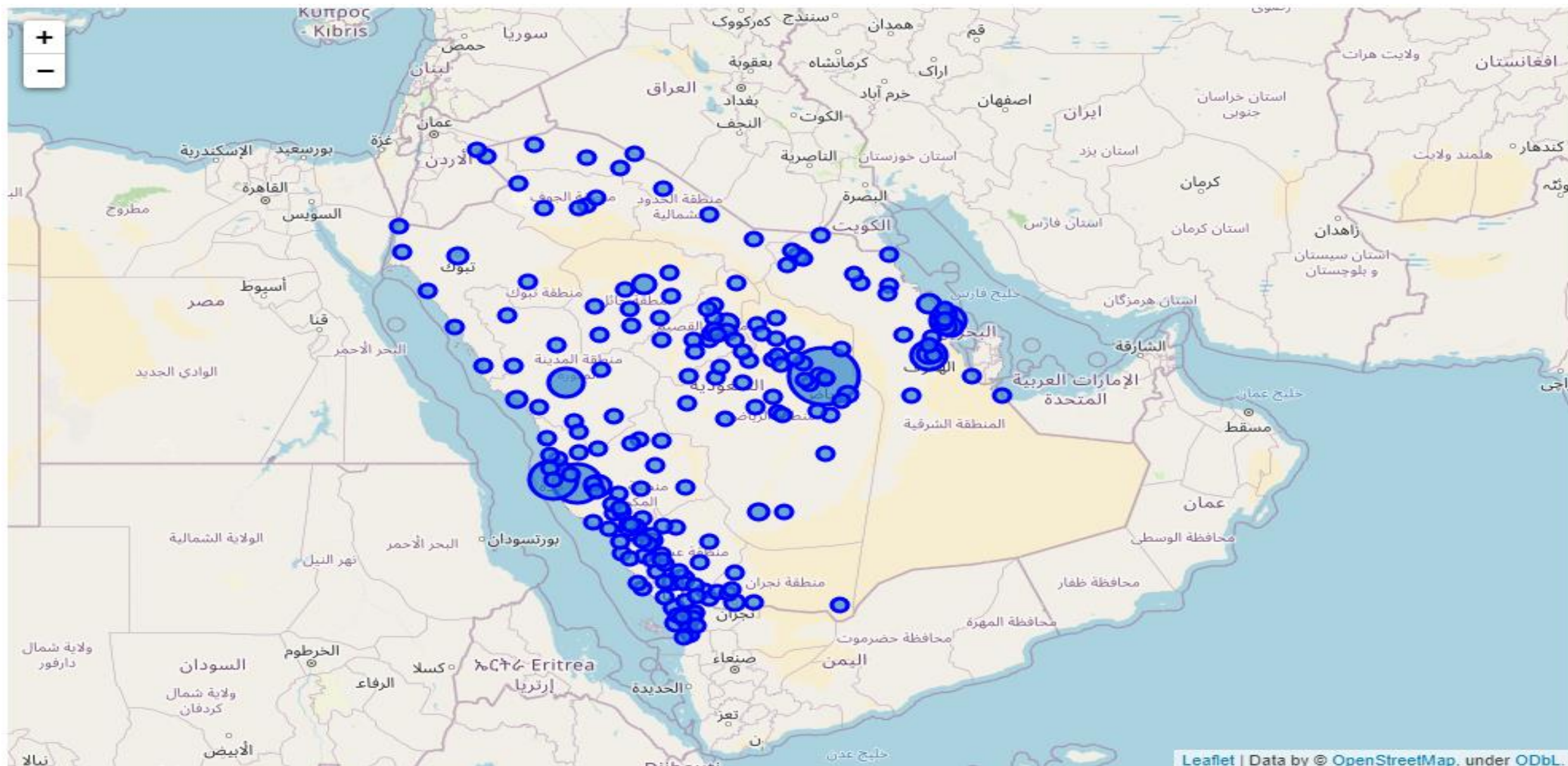
Case	Cure	Death
360690	351573	6101



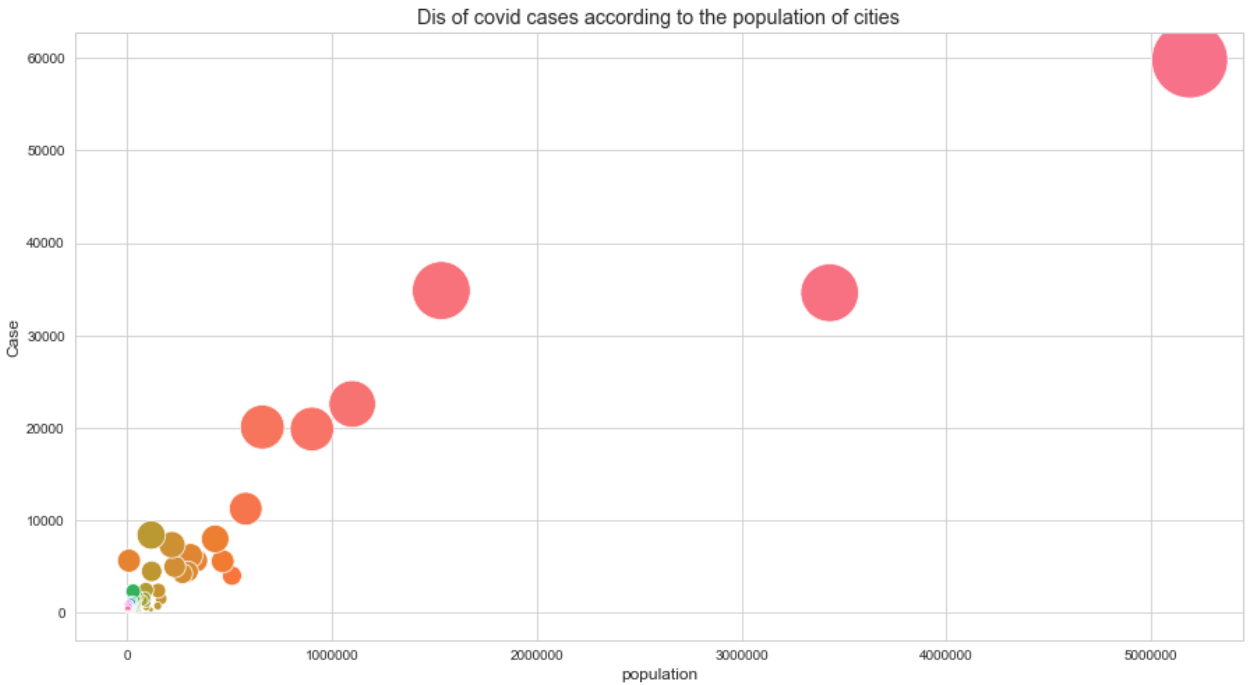
Analysis



Analysis



Analysis

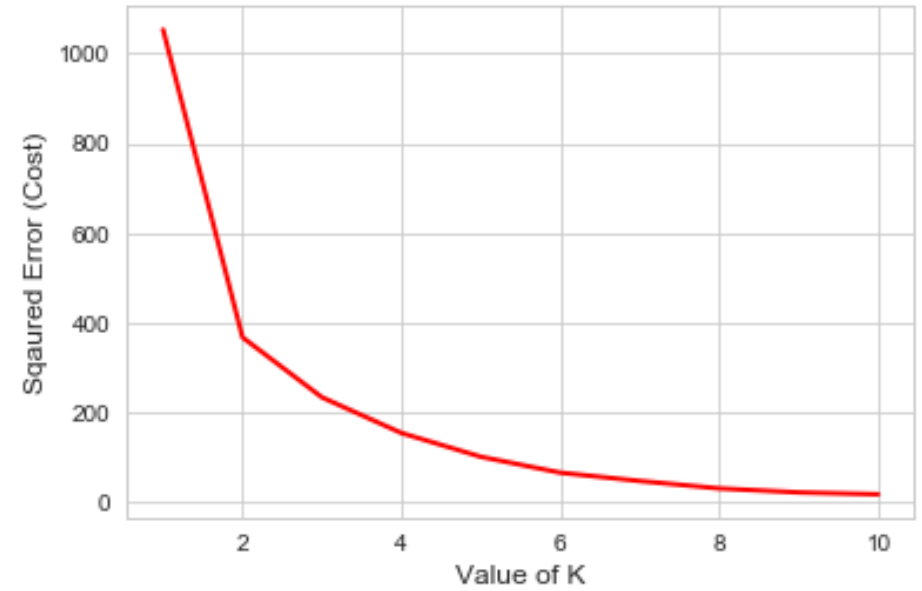


Proposed Model

We use Kmean cluster method to do that. It depends of the three different inputs Covid 19 Cases, number of hospitals and population.

We set the parameters of the Kmean model as follow:

- Init:k-means++
- n_clusters:4
- n_init:12



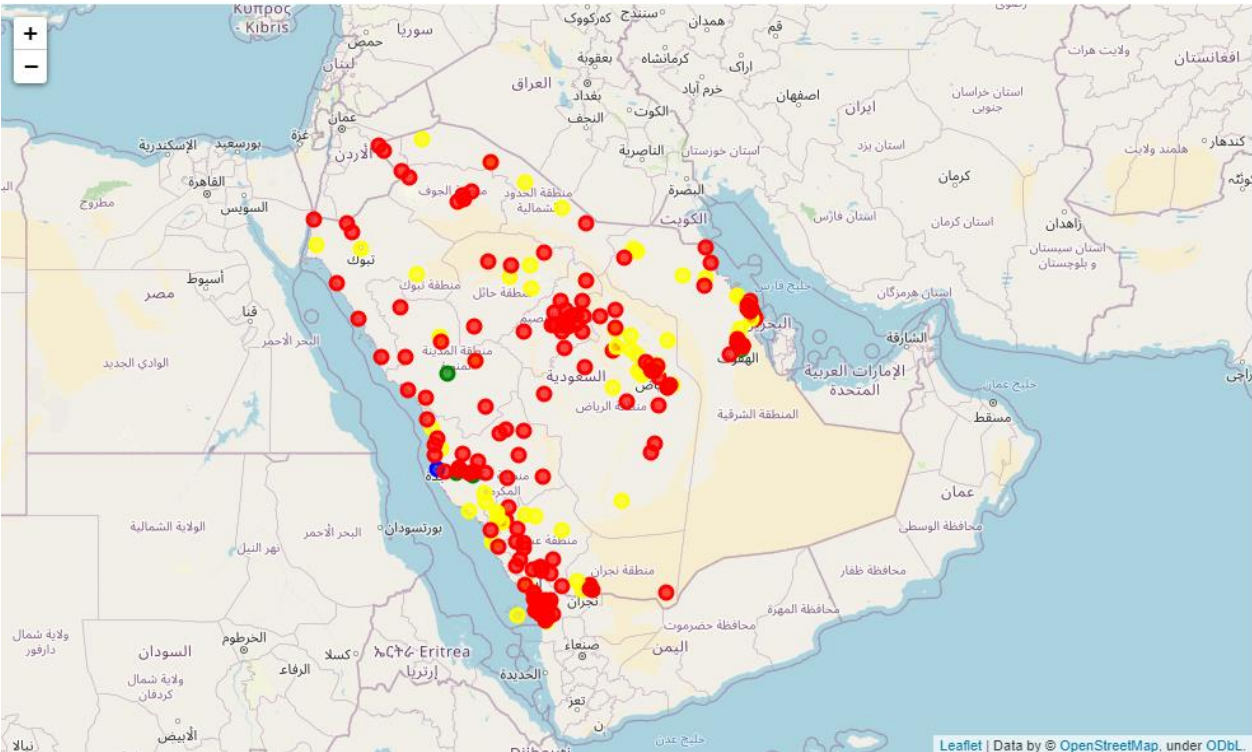
Experiments and Results

The mean of input parameters according to each cluster

cluster	population	Case	Hospital
0	2.279911e+04	328.020661	0.000000
1	4.309492e+06	47202.500000	9.500000
2	9.557788e+05	21709.400000	2.800000
3	3.313381e+04	819.950980	1.205882

Cluster	Number
0	242
1	2
2	5
3	102

Distribution of Clusters



Experiments and Results

Cluster 1 information

	City	State	population	lat_x	long_x	disp_x	Case	disp_y	lat_y	long_y	Hospital	cluster
0	الرياض	الرياض	5188286.0	24.713552	46.675296	ض.البرلا	59805.0	ض.البرلا	24.713552	46.675296	10.0	1
1	جدة	مكة المكرمة	3430697.0	21.485811	39.192505	منج	34600.0	منج	21.485811	39.192505	9.0	1

Top 20 of cluster 3

	City	State	population	Case
7	بيوك	بيوك	512629.0	3978.0
10	الجبيل	المنطقة الشرقية	337778.0	5600.0
11	الجبيل	المنطقة الشرقية	10241.0	5600.0
12	حائل	حائل	310897.0	6141.0
13	نجران	نجران	298288.0	4423.0
14	حفر الباطن	المنطقة الشرقية	271642.0	4192.0
20	عرعر	الحدود الشمالية	167057.0	1474.0
25	الطهران	المنطقة الشرقية	120521.0	4447.0
26	القطيف	المنطقة الشرقية	118327.0	8377.0
28	اليابحة	اليابحة	95089.0	558.0
29	وادي الدواسر	الرياض	93036.0	2429.0
31	بيشة	عسير	86201.0	1571.0
43	رابع	مكة المكرمة	55304.0	274.0
45	رفحاء	الحدود الشمالية	52712.0	0.0
46	صفوى	المنطقة الشرقية	50447.0	1683.0
48	طريف	الحدود الشمالية	48108.0	209.0
49	المجمعة	الرياض	47743.0	1013.0
54	الدلم	الرياض	40114.0	428.0
56	بقيق	المنطقة الشرقية	36207.0	1487.0
57	الحيون	المنطقة الشرقية	33042.0	679.0

Cluster 2 information

	City	State	population	Case
2	مكة المكرمة	مكة المكرمة	1534731.0	34839.0
3	المدينة المنورة	المدينة المنورة	1100093.0	22563.0
4	الدمام	المنطقة الشرقية	903312.0	19851.0
5	الظهر	المنطقة الشرقية	660788.0	20066.0
6	الطائف	مكة المكرمة	579970.0	11228.0

Top 20 of cluster 0

	City	State	population	Case
8	بريدة	القصيم	467410.0	5532.0
9	خميس مشيط	عسير	430828.0	7956.0
15	الغنية	المنطقة الشرقية	238066.0	0.0
16	ايها	عسير	236157.0	0.0
17	السيح	الرياض	103216.0	0.0
18	ينبع	المدينة المنورة	233236.0	4948.0
19	الخبر	المنطقة الشرقية	219679.0	7325.0
21	عنيزة	القصيم	152895.0	2357.0
22	سكاكا	الجوف	150257.0	699.0
23	الحوية	مكة المكرمة	148151.0	0.0
24	جيزان	جازان	127743.0	0.0
27	القرينات	الجوف	116162.0	304.0
30	الرس	القصيم	92501.0	1098.0
32	تاروت	المنطقة الشرقية	77757.0	0.0
33	سوهاج	المنطقة الشرقية	75794.0	14.0
34	شروخ	نجران	75237.0	1286.0
35	بحره	مكة المكرمة	75213.0	0.0
36	الخفجي	المنطقة الشرقية	67012.0	0.0
37	صبياء	جازان	63143.0	0.0
38	الدوادمي	الرياض	61834.0	0.0

Conclusion and Future Work

In this work, I analyzed the Covid 19 of Saudi Arabia and the situation. Also, I analyze and show how Saudi Arabia control the epidemic. In this study we collect the data from many resources like ministries, Wiki, foursquare api and, etc. we combine the data and manipulate it and fill it with required data using different python libraries like geopy, foursquare and BeautifulSoup libraries. Finally, we build a cluster model using Kmean algorithm and set its parameters in correct way. The input data of the model were population, Covid 19 cases and number of hospitals. The main reason of this study is to help Stakeholders to distribute the vacancies to the people in fairly way.

Future work

I plan to build a model of classification algorithm to help the authorities to give priorities to people according to the people in need more than others according to the study of the affection of people.

Acknowledgment

Thanks to IBM for their great Course