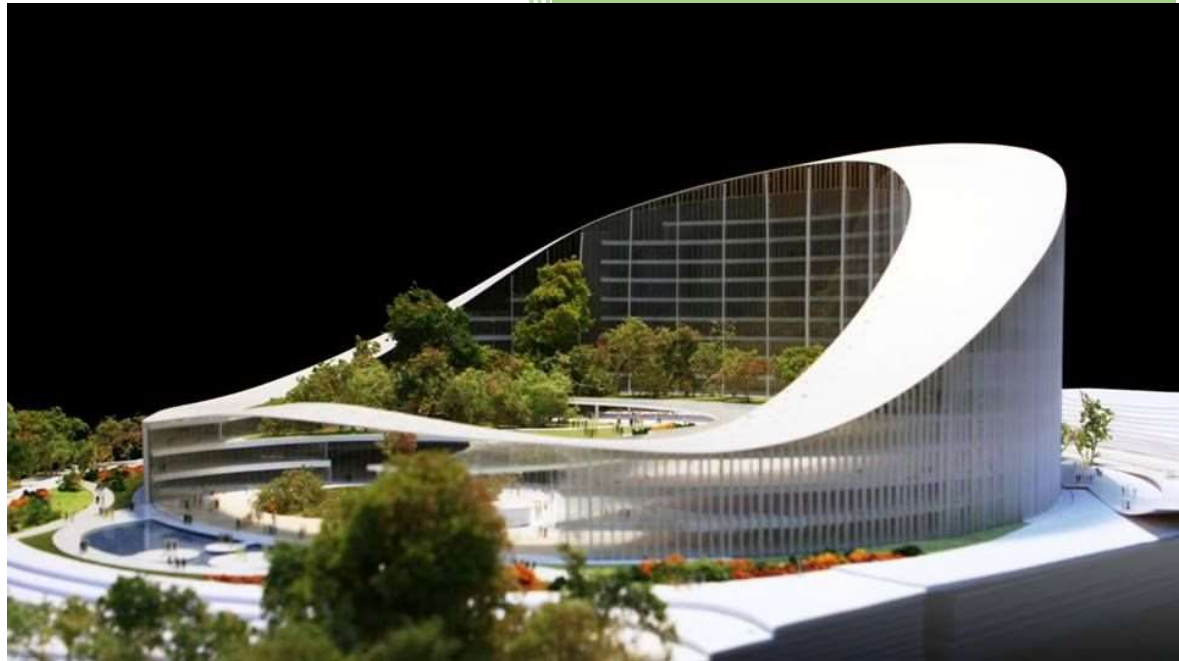


2020

Coursera Capstone IBM Applied Data Science



Abdallah Ghebriou

Coursera

Introduction:

For many patients

Algeria is the largest country in Africa and is estimated to have a population of around 40 million people. Algeria has a public health care system, which is accessible and free of charge to all citizens of Algeria. The public health care system is financed by the government of Algeria.

The COVID-19 push us to think about a new hospital with more beds including technology to provide a great service to our citizens.

However, opening a new hospital requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decision that will determine whether the hospital will be a success or failure.

Business problem:

The Objective of this capstone project is to analyze and select the best locations in the City of Algiers to open new Hospital

Using data science methodology and machine learning techniques like clustering, this project aims to provide solution to answer the business question:

Is it the perfect location? Where would you like that they open it?

Data:

To solve the problem, we will need the following data:

1-population of the area where hospital is recommended to be build.

2-latitude and longitude coordinates of those neighborhoods, this is required in order to plot the map and also get the venue data

3-get the venue data, which is related to hospital, we will use this data to perform clustering on the area.

Source of data and methods to extract:

Source of the data will be from an excel file, excel file provides the data of that area.

We will use web scraping techniques to extract the data from the excel file using the python requests and beautiful soup package,

After we will get the geographical coordinates of the area using the python Geocoder package.

We will use the foursquare API to get the venue data for that area

foursquare has one of the largest databases of 4 million places and used by over 80000 developers.

Foursquare API will provide many categories of the venue data, we are interested in the Hospital.

Using the methodology section where we will discuss the steps taken in this project.

Data analysis that we did and the machine learning technique that was used.

Create Data Table of Venues in St. Charles

In [260]:

```
venues = results['response']['groups'][0]['items']
stc = json_normalize(venues) # flatten JSON
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
stc = stc.loc[:, filtered_columns]
stc['venue.categories'] = stc.apply(get_category_type, axis=1)
stc.columns = [col.split(".")[0] for col in stc.columns]
stc.insert(0, 'County', 'St. Charles')
print('{} venues were returned by Foursquare.'.format(df.shape[0]))
stc
```

Co de	Provinc e		Num ber of distri cts	Number of munici palities	Area		Popula tion (2008 ^[4])	Density (2008)	
					km ²	sq mi		pe r k m ²	per sq mi
01	Adrar		6	16	254,471	98,252	261,258	1.03	2.7
02	Chlef		13	35	4,795	1,851	1,002,088	209	540
03	Laghouat		10	24	25,057	9,675	455,602	18	47

Co de	Provinc e		Num ber of distri cts	Number of munici palities	Area		Popula tion (2008 ⁽¹⁾)	Density (2008)	
					km ²	sq mi		pe r k m ²	per sq mi
04	Oum el-Bouaghi		12	29	6,783	2,619	621,612	81	210
05	Batna		21	61	12,192	4,707	1,119,791	92	240
06	Béjaïa		19	52	3,268	1,262	912,577	279	720
07	Biskra		10	27	9,576	3,697	547,137	5,714	14,800
08	Béchar	بشار		11	60,850	23,490	219,898	3.61	9.3
09	Blida	البلدية		25	1,575	608	1,002,937	591	1,530
10	Bouïra	البويرة		45	4,439	1,714	695,583	157	410
11	Tamanghasset	تمنراست		5	336,839	130,054	115,043	0.34	0.88
12	Tébessa	تبسة		28	14,227	5,493	648,703	46	120
13	Tlemcen	تلمسان		53	9,061	3,498	949,135	105	270

Co de	Provinc e		Num ber of distri cts	Number of munici palities	Area		Popula tion (2008 ^[1])	Density (2008)	
					km ²	sq mi		pe r k m ²	per sq mi
14	Tiaret			42	20,673	7,982	846,823	41	110
15	Tizi Ouzou			67	2,956	1,141	1,127,608	316	820
16	Algiers			57	1,190	460	2,988,145	2,511	6,500
17	Djelfa			36	66,415	25,643	1,092,184	46	120
18	Jijel			28	2,577	995	636,948	247	640
19	Sétif			60	6,504	2,511	1,489,979	229	590
20	Saïda			16	6,764	2,612	330,641	49	130
21	Skikda			38	4,026	1,554	898,680	223	580
22	Sidi Bel Abbès			52	9,096	3,512	604,744	66	170
23	Annaba			12	1,439	556	609,499	424	1,100
24	Guelma			34	4,101	1,583	482,430	118	310

Co de	Provinc e		Num ber of distri cts	Number of munici palities	Area		Popula tion (2008 ⁽¹⁾)	Density (2008)	
					km ²	sq mi		pe r k m ²	per sq mi
25	Constantine			12	2,187	844	938,475	427	1,110
26	Médéa			64	8,866	3,423	819,932	92	240
27	Mostaganem			32	2,175	840	737,118	325	840
28	M'sila			47	18,718	7,227	990,591	53	140
29	Mascara			47	5,941	2,294	784,073	132	340
30	Ouargla			10	194,552	75,117	311,337	1.6	4.1
31	Oran			26	2,121	819	1,454,078	688	1,780
32	El Bayadh			22	78,870	30,450	228,624	3.2	8.3
33	Illizi			4	198,815	76,763	34,715	0.17	0.44
34	Bordj Bou Arréridj			34	4,115	1,589	628,475	160	410

Co de	Provinc e		Num ber of distri cts	Number of munici palities	Area		Popula tion (2008 ^[1])	Density (2008)	
					km ²	sq mi		pe r k m ²	per sq mi
35	Boumerdès			32	1,356	524	802,083	504	1,310
36	El Taref			24	3,339	1,289	408,414	122	320
37	Tindouf			2	159,000	61,000	49,149	0.31	0.80
38	Tissemsilt			22	3,152	1,217	294,476	93	240
39	El Oued			22	45,738	17,660	485,281	1,061	2,750
40	Khenchela			21	9,811	3,788	386,683	40	100
41	Souk Ahras			26	4,541	1,753	438,127	95	250
42	Tipaza			28	1,605	620	591,010	273	710
43	Mila			32	3,407	1,315	766,886	220	570
44	Aïn Defla			36	4,891	1,888	766,013	156	400

Co de	Provinc e		Num ber of distri cts	Number of munici palities	Area		Popula tion (2008 ^[1])	Density (2008)	
					km ²	sq mi		pe r k m ²	per sq mi
45	Naâma			12	29,950	11,560	192,891	6.5	17
46	Aïn Témouchent			28	2,379	919	371,239	156	400
47	Ghardaïa			9	23,890	9,220	306,322	12.82	33.2
48	Relizane			38	4,870	1,880	726,180	152	390
49	El M'ghair			8	8,835	3,411	162 267	0.94	2.4
50	El Menia			4	62,215	24,021	57,276	0.92	2.4
51	Ouled Djellal			6	11,410	4,410	174,219	15	39
52	Bordj Baji Mokhtar			2	120,026	46,342	16,437	014	36
53	Béni Abbès			10	101,350	39,130	50,163	049	130

Co de	Provinc e		Num ber of distri cts	Number of munici palities	Area		Popula tion (2008 ⁽¹⁾)	Density (2008)	
					km ²	sq mi		pe r k m ²	per sq mi
54	Timimoun			10	65,203	25,175	122,019	19	49
55	Touggourt			11	17,428	6,729	247,221	14	36
56	Djanet			2	86,185	33,276	17,618	02	5.2
57	In Salah			3	131,220	50,660	50,392	038	98
58	In Guezzam			2	88,126	34,026	11,202	013	34
Total				1541	2,381,741	919,595	34,080,030	14	36

Create map of venues Algeria

In [261]:

```
stc_map = folium.Map(location=[county_latitude, county_longitude], zoom_start=12)

# add markers to map
for lat, lng, name, categories in zip(stc['lat'], stc['lng'], stc['name'], stc['categories']):
    label = '{}{}'.format(categories, name)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7).add_to(stc_map)
```

Tiaret	5
Tizi Ouzou	5
Algiers	3
Djelfa	3
Jijel	5
Sétif	2
Saïda	4
Skikda	22

Sidi Bel Abbès	54
Annaba	5
Guelma	6
Constantine	7
Médéa	55
Mostaganem	45
M'sila	47
Mascara	44
Ouargla	7
Oran	4
El Bayadh	45
Illizi	8
Bordj Bou Arréridj	8
Boumerdès	9
El Taref	5

Tindouf	10
Tissemsilt	47
El Oued	25
Khenchela	54
Souk Ahras	1
Tipaza	4
Mila	44
Aïn Defla	45
Naâma	44
Aïn Témouchent	44
Ghardaïa	10
Relizane	11
El M'ghair	22
El Menia	33
Ouled Djellal	32

Bordj Baji Mokhtar	12
Béni Abbès	24
Timimoun	11
Touggourt	12
Djanet	14
In Salah	12
In Guezzam	12

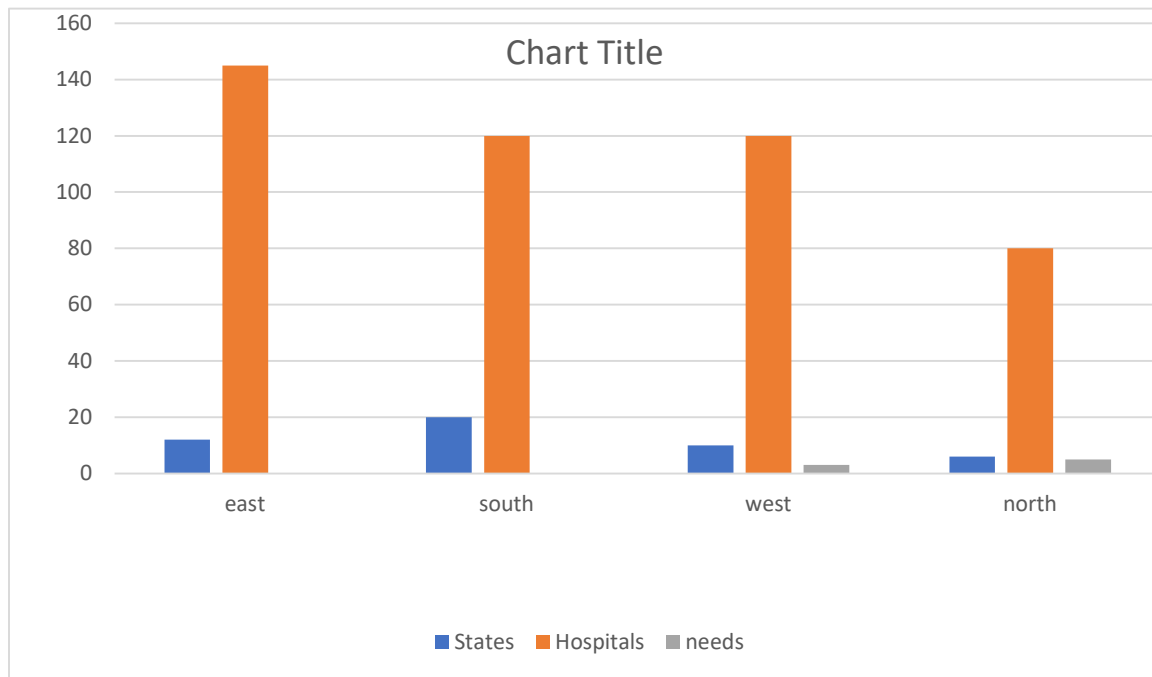
Barplot showing Hospitals in locations

In [268]:

```
df_venues_final.groupby('County')['Venue Type']\
    .value_counts()\
    .unstack(level=1)\
    .plot.bar(stacked=True)
```

Out[268]:

<matplotlib.axes._subplots.AxesSubplot at 0x1f7c8356b08>



Barplot showing Hospitals in locations (east, south and west)

Results

As can be seen from the data, there are not many hospitals on north and needs to increase in the other locations

Conclusion

In conclusion, the data obtained from mo.gov was very useful in indicating counties within the state. As well as foursquare for obtaining information about specific venues within the counties. This data was put together to form tables and maps showing where the best locations for future potential bre hospitals might be. Again, it seems that the best locations for a north needs the hospital