



IBM DATA SCIENCE CAPSTONE PROJECT

AKSHAY SHARMA

Introduction

- The outbreak of COVID-19 in different parts of the world is a major concern for all the administrative units of the respective countries. Delhi is also facing this very tough task for controlling the virus outbreak and has managed its growth rate through some strict measures.
- Centralizing the isolation centers creates some issues for the general public:
 - Accessibility issues for the large chunk of the populations living at the extreme corners of the state.
 - Transportation Issues: those who are transporting these infected people will need PPE and other protections and there is always a possibility of infection for them.
 - Social Distancing Problem: Centralizing the testing and isolation centers can cause huge problems for a state like Delhi

- So Decentralising the Testing and isolation centers will surely help to tackle this pandemic more efficiently. And This Work can be used by both Govt Agencies as well as private firms.

Data:

I used the Wikipedia website to extract the information for the Delhi Region subdivisions and other demographics such as Area in square km and population in each of the region was also available on that page.

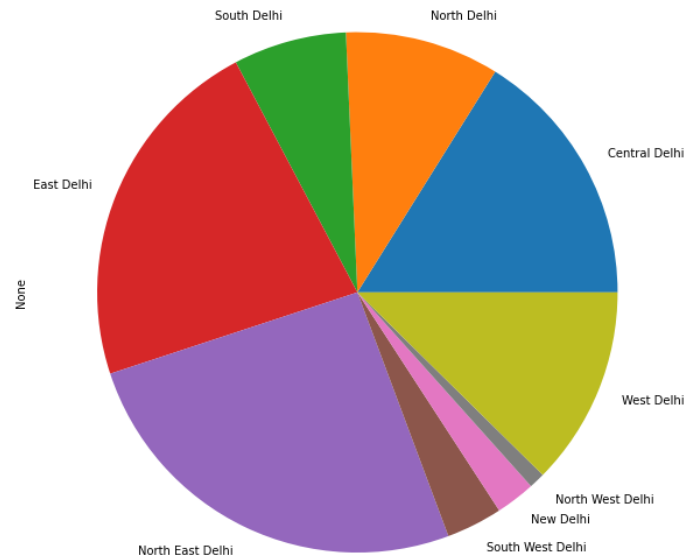
Out[81]:

	District	Area SqKM	Population	Latitude	Longitude
0	Central Delhi	23	578671	28.698548	77.219391
1	North Delhi	59	883418	28.613895	77.209006
2	South Delhi	249	2733752	28.485169	77.196380
3	East Delhi	49	1707725	28.620477	77.309181
4	North East Delhi	56	2240749	28.723308	77.266857
5	South West Delhi	421	2292363	28.586448	76.979153
6	New Delhi	35	133713	28.613895	77.209006
7	North West Delhi	2344	3651261	28.613895	77.209006
8	West Delhi	131	2531583	28.647952	77.085565

Exploratory Data Analysis:

```
In [82]: import matplotlib.pyplot as plt
dfp_a=df["Population"]/df["Area SqKM"]
dfp_a.index=df["District"]
dfp_a.plot(kind="pie",figsize=(10,15))
plt.xlabel("PIE CHART FOR POPULATION PER AREA KM^2")
```

Out[82]: Text(0.5, 0, 'PIE CHART FOR POPULATION PER AREA KM^2')



```
In [83]: print("Correlation B/w Population and Area :",df[["Population","Area SqKM"]].corr())
```

Correlation B/w Population and Area :			Population	Area SqKM
Population	1.000000	0.666108		
Area SqKM	0.666108	1.000000		

An exploratory data analysis on the above data frame to gain some small insights into the dataset. And as we can infer from the above pie chart, is that the area is not evenly distributed and this can be a problem while setting up the isolation and testing centers. I then also checked if the population has some correlation with the area.

Methodology and Results:

So the data is further used to generate the data for the nearby hospitals, medical suppliers, medical labs, and emergency centers and to see how they're clustered around the city. Local hospitals, labs can provide help to setup Isolation camps and test centers.

Out[84]:

	name	categories	lat	lng
0	Primus super speciality hospital	Hospital	28.592961	77.181430
1	Dr. Ram Manohar Lohia Hospital	Hospital	28.624654	77.200609
2	Sir Gangaram Hospital सर गंगाराम अस्पताल	Hospital	28.638601	77.188884
3	Loknayak Hospital	Hospital	28.638303	77.238829
4	B.L kapoor memorial hospital	Hospital	28.643368	77.179899
5	BLK Super Specialty Hospital	Hospital	28.643392	77.179582
6	Safdarjung hospital	Hospital	28.566744	77.204870
7	St. Stephen's Hospital	Hospital	28.666454	77.214489
8	Aashlok Hospital	Hospital	28.566571	77.198902
9	AIIMS	Medical School	28.567922	77.209255
10	Moolchand Hospital	Hospital	28.564782	77.243883
11	Moolchand medcity	Hospital	28.566345	77.235344

Figure: Dataframe generated for Nearby Hospitals.

Out[85]:

	name	categories	lat	lng
0	Draeger India Private Limited	Medical Supply Store	28.565078	77.188793
1	Draeger India Private Limited	Medical Supply Store	28.565068	77.188774
2	Draeger India Private Limited	Medical Supply Store	28.565032	77.188851
3	Draeger India Private Limited	Medical Supply Store	28.565052	77.188774
4	Modern Times Helpline Pharma	Medical Supply Store	28.645798	77.243843
5	kirpashakti foundation	Medical Supply Store	28.615955	77.080628

Figure: Dataframe generated for Nearby Suppliers

Generating Map to Visualise all the Hospitals, Medical Supply Stores, Medical Labs and Emergency Centres in Delhi

```
In [88]: Df_Delhi_Medical_infra = pd.concat([nearby_hospitals, nearby_medicalsupply, nearby_medicallabs, nearby_emergencycentres], ignore_index=True)
```

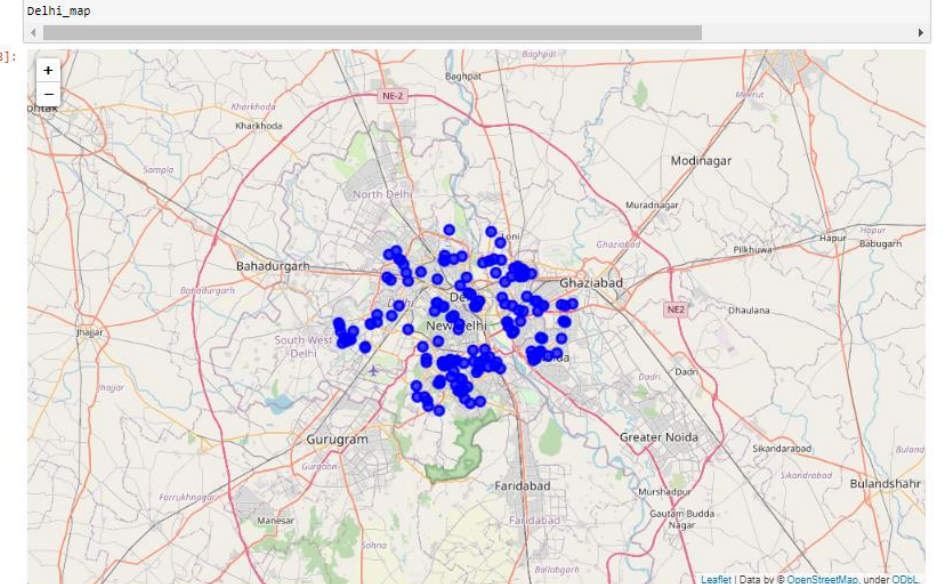
```
In [89]: Df_Delhi_Medical_infra
```

```
Out[89]:
```

	name	categories	lat	lng
0	Primus super speciality hospital	Hospital	28.592981	77.181430
1	Dr. Ram Manohar Lohia Hospital	Hospital	28.624654	77.200609
2	Sir Gangaram Hospital सर गंगाराम अस्पताल	Hospital	28.638601	77.188884
3	Loknayak Hospital	Hospital	28.638303	77.238829
4	B.L kapoor memorial hospital	Hospital	28.643368	77.179899
5	BLK Super Specialty Hospital	Hospital	28.643392	77.179582
6	Safdarjung hospital	Hospital	28.568744	77.204870
7	St. Stephen's Hospital	Hospital	28.668454	77.214489
8	Aashlok Hospital	Hospital	28.568571	77.198902
9	AIIMS	Medical School	28.567922	77.206255
10	Moolchand Hospital	Hospital	28.564782	77.243883
11	Moolchand medcity	Hospital	28.568345	77.235344

Figure: Merging all the Dataframe into one to see the spread of the data.

```
for lat, lng, name, categories in zip(Df_Delhi_Medical_infra['lat'], Df_Delhi_Medical_infra['lng'], Df_Delhi_Medical_infra['name'], Df_Delhi_Medical_infra['categories']):  
    label = '{}'.format(name)  
    label = folium.Popup(label, parse_html=True)  
    folium.CircleMarker(  
        [lat, lng],  
        radius=5,  
        popup=label,  
        color='blue',  
        fill=True,  
        fill_color='blue',  
        fill_opacity=0.7,  
        parse_html=False).add_to(Delhi_map)
```



Then I used the Folium.map function and I passed the Longitude and Latitude Data to the function as the parameters

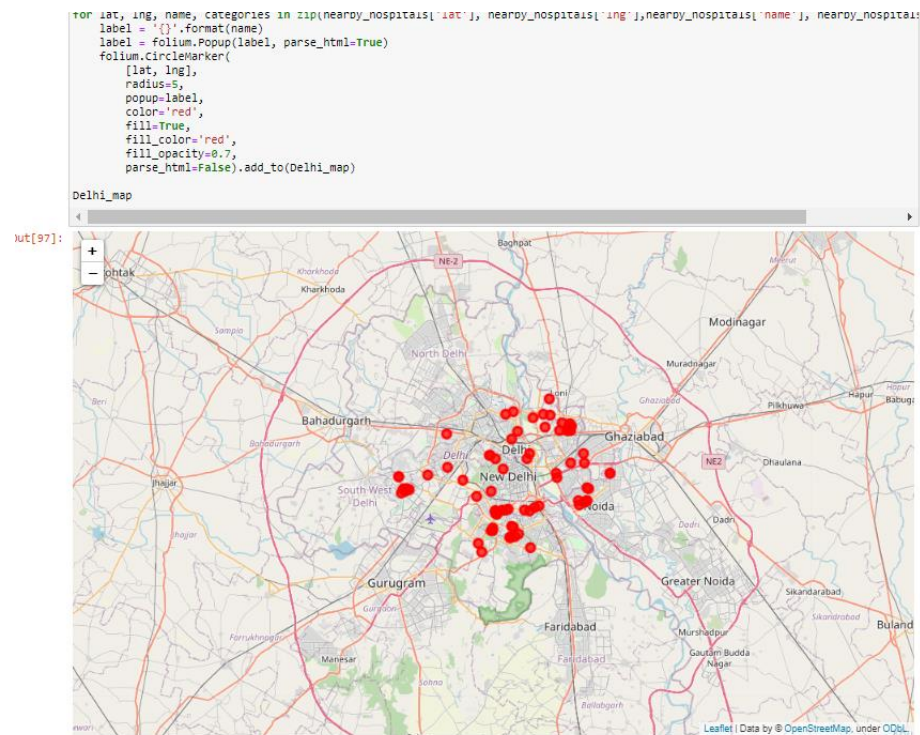


Figure: Spread of Hospitals.

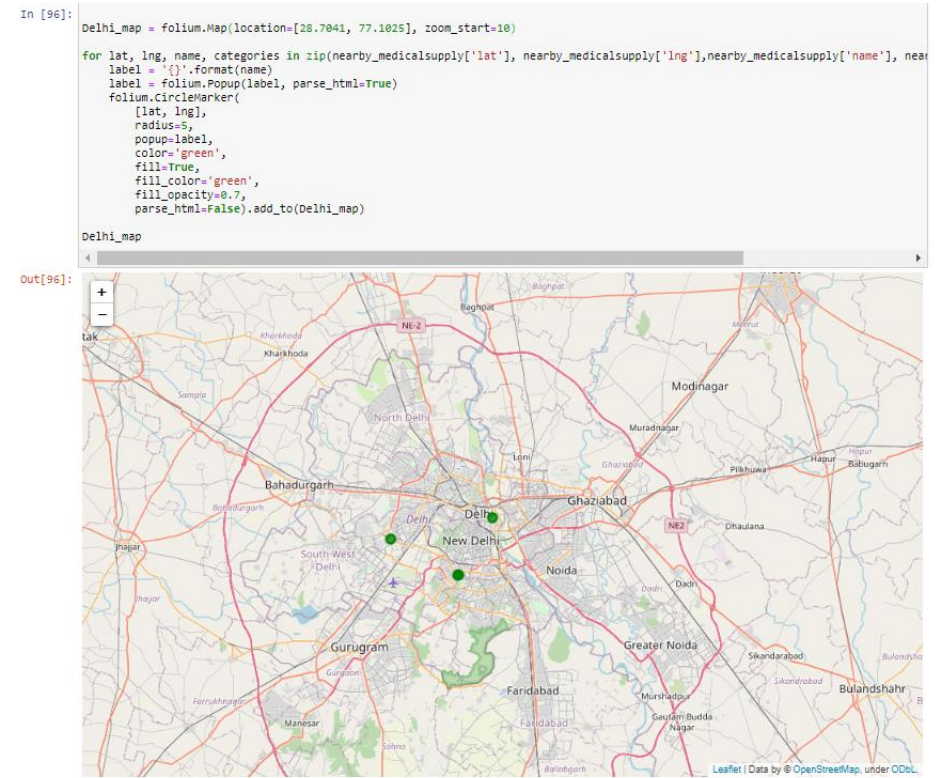


Figure: Spread of Medical Suppliers.

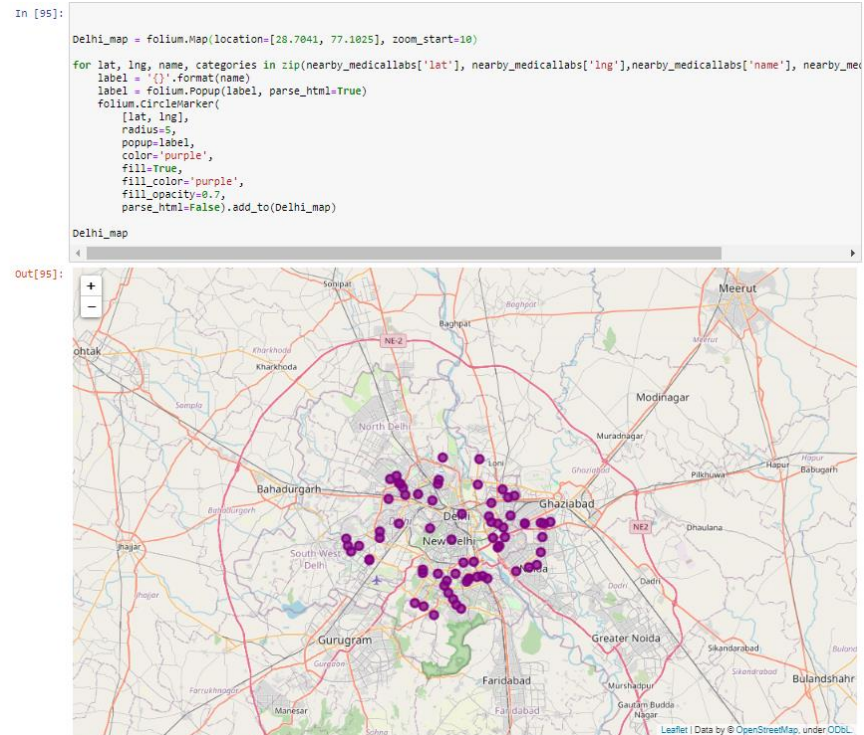


Figure: Spread of Medical Labs

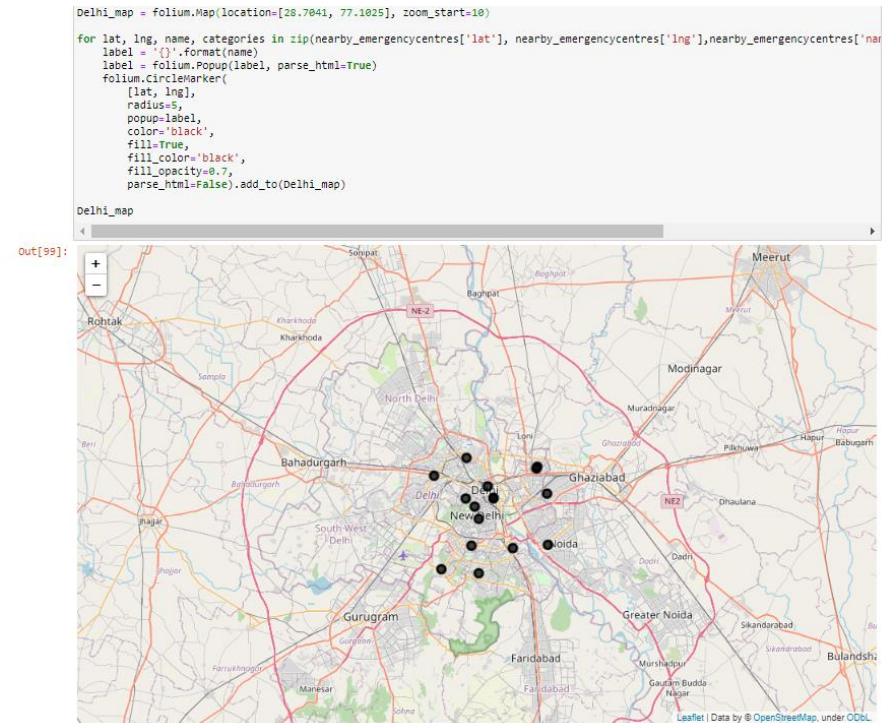


Figure: Spread of Emergency Centers.

we can infer from the spreads, the Facilities are not evenly spread throughout the city and are majorly clustered around the center of the city

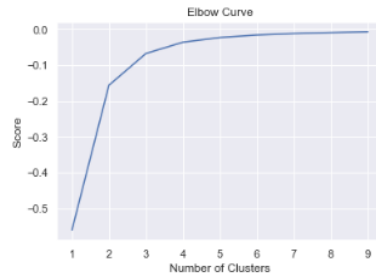
So Kmeans Clustering is used to find out the actual clustering

Kmeans Clustering and Elbow Evaluation Method:

Using K-Means and Evaluation Measures to find how the Infra is actually clustered

```
In [103]: import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import seaborn as sns; sns.set()
import csv
df_kmeans=Df_Delhi_Medical_infra
```

```
In [104]: K_clusters = range(1,10)
kmeans = [KMeans(n_clusters=i) for i in K_clusters]
Y_axis = df_kmeans[['lat']]
X_axis = df_kmeans[['lng']]
score = [kmeans[i].fit(Y_axis).score(Y_axis) for i in range(len(kmeans))]
# Visualize
plt.plot(K_clusters, score)
plt.xlabel('Number of Clusters')
plt.ylabel('Score')
plt.title('Elbow Curve')
plt.show()
```



```
In [106]: kmeans = KMeans(n_clusters = 3, init = 'k-means++')
kmeans.fit(df_kmeans[df_kmeans.columns[3:5]]) # Compute k-means clustering.
df_kmeans['cluster_label'] = kmeans.fit_predict(df_kmeans[df_kmeans.columns[2:4]])
centers = kmeans.cluster_centers_ # Coordinates of cluster centers.
labels = kmeans.predict(df_kmeans[df_kmeans.columns[2:4]]) # Labels of each point
df_kmeans
```

Out[106]:

	name	categories	lat	lng	cluster_label
0	Primus super speciality hospital	Hospital	28.592951	77.181430	2
1	Dr. Ram Manohar Lohia Hospital	Hospital	28.624654	77.200609	2
2	Sir Gangaram Hospital सर गंगाराम अस्पताल	Hospital	28.638801	77.188884	2
3	Loknayak Hospital	Hospital	28.638303	77.238829	1
4	B.L kapoor memorial hospital	Hospital	28.643368	77.178899	2
5	BLK Super Speciality Hospital	Hospital	28.643392	77.179582	2
6	Safdarjung hospital	Hospital	28.566744	77.204870	2
7	St. Stephen's Hospital	Hospital	28.666454	77.214489	1
8	Aashlok Hospital	Hospital	28.566571	77.198902	2
9	AIIMS	Medical School	28.567922	77.209255	2
10	Moolchand Hospital	Hospital	28.564782	77.243883	2
11	Moolchand medcity	Hospital	28.566345	77.235344	2

Cluster Label is added to the corresponding Hospitals and Other facilities

Conclusion:

In this study, I analyzed the data for Delhi's Health Infrastructure and its demographics. And then used the data to view the spread of the hospital's facilities. Then I built the Kmean Clustering model to see the actual clustering of medical facilities in the state. And This model is very effective to see which of the state needs more attention in terms of new hospitals, testing centers, and isolation camps. For Example, I would highly recommend to setup more isolation centers and test on the outskirts of Delhi as there have relatively less no of medical facilities which are evident from the plots.

The above study can also be fine-tuned by adding by taking following into the consideration:

- Use of population data for clusterings.
- Taking into account the medical infrastructure of the individual facilities such as No of beds available and No of doctors, nurses available. Etc

These data are more difficult to extract and quantify, but if optimized, could bring significant improvements to the models.