Control the Spread of Coronavirus in London

1. Introduction:

1.1 Background:

In the current scenario of global pandemic, the coronavirus spread in more than 180 countries around the world. The pandemic started from the seafood market of Wuhan city in Hubei province in January as per the existing records available. The pandemic has now affected most of the world including major European countries like Italy, Spain, UK, France among the worse effected regions, and also United States.

The coronavirus infections result into disease name Covid-19 which starts with mild flu or influenza and later on has different impact on people with different age group. The infection actually is considered to be highly contagious and spreads through coming in contact with the person having related symptoms which are sometime undetectable. Hence, the population getting infected can be divided among two types: developing symptoms and asymptotic i.e. not showing any symptoms. It is therefore sometimes difficult to trace people with no symptoms and it is quite inevitable that they become the potential careers of coronavirus.

The recovery period is considered to be 14 days among the healthy people with high immunity whereas resulting into fatalities for some of the older as well as people with pre-existing illness. Henceforth, the mortality rate has come out to be 5-6% globally.

1.2 Problem:

In the current scenario of global pandemic, the coronavirus spread in more than 180 countries around the world. The pandemic started from the seafood market of Wuhan city in Hubei province in January as per the existing records available. The pandemic has now affected most of the world including major European countries like Italy, Spain, UK, France among the worse effected regions, and also United States.

Over 143,000 people have tested positive for coronavirus in the UK.

In London, the figure for positive tests stands at 23,063.

So, **the problem before us is how to control the spread of coronavirus in London.**

In order to control the situation they need to first implement strict lockdown in the top most worst effected boroughs and then figure out the neighbourhood of the borough with hospitals and other heallth care facilities. If sufficient hospitals are

not there then plan is to convert the other common public venues which are closed as of now to be converted to healthcare and quaratine facilities.

We will focus on the worst effected borough and explore its neighborhoods and the hospitals along with 10 most common venues in each neighborhood so that the best neighborhood suited to and health ministry of London needs can be selected.

For this we have to analyse the data closely and find out the most worst effected region of London using Foursquare API, so that the government and health ministry of the city can implement certain precautionary measures in order to curb down the random spread and try to control the situation to certain degree.

1.3 Interest:

Government, Doctors and Health Care workers who are real audience at this time of crisis. This project will provide location of the hospitals and near testing centres for COVID-19 patients.

Data for this project can be fetched from Foursquare API which provide latitudes and longitudes which can help cluster the areas containing hospitals or other places which can contribute to such health care facilities.

2. Data acquisition and cleaning

2.1 Data sources

The data acquired for the project is the combination data from three sources. The first source of the project uses London Coronavirus Infected cases in London.

The second source of data is scraped from the Wikipedia page that contains the list of London Boroughs. The page contains additional information about the borough.

The third data source is the list of Neighbourhoods in the Royal Borough of Brent as found on a Wikipedia page. This dataset is created from scratch using the list of neighbourhoods available on the site.

2.2 Data Cleaning

The data preparation for each of the three sources of data is done separately. Initially, we took the data from showing total number of confirmed infections in each borough.

The second data is scraped from a Wikipedia page using the Beautiful Soup library in python. Using this library, we can extract the data in the tabular format as shown in the website. After the web scraping, string manipulation is required to

get the names of the boroughs in the correct form. This is important because we will be merging the two datasets together using the Borough names.

The two datasets are merged on the Borough names to form a new dataset that combines the necessary information in one dataset. The purpose of this dataset is to visualize the coronavirus infection rates in each borough and identify the worst effected borough with the maximum number of coronavirus cases recorded so far.

After visualizing the coronavirus cases in each borough, we can find the borough with the maximum number of cases and hence tag that borough as the worst effected borough. The third source of data is acquired from the list of neighbourhoods in the worst affected borough on Wikipedia. This dataset is created from scratch, the pandas data frame is created with the names of the neighbourhoods and the name of the borough with the latitude and longitude are obtained using Google Maps API geocoding to get the final dataset.

2.3 Relevant Solution

This is a critical problem that happened globally so the only solution suggested by the experts to maintain social distancing and to exhibit lockdown in the cities which are worst affected by the pandemic. This lockdown prohibits the unnecessary travel of the people as well as going out of the homes for the non-essential reasons. Also, building of hospitals along with quarantine facilities and employing the healthcare workers so that critical patients effected by the disease can be better taken care of.

One way is to analyse the worst regions of London and practice strict Lockdown there. Along with that the regions have to be critically monitored for the spread.

People should not be allowed to go out without masks and only one person per house should permitted to go out for grocery items. No unnecessary travels should be availed for citizens of that region. The infected people should be home quarantine and asked to stay away from other members. Senior Citizens and children below age of 5 should be given extra precautions of that region. The complete region should be declared red zone and strict lockdown should be enforced in the region.

The government should observe serious measures to handle the situation by maintaining required number of hospitals and medical facilities for the severely and critically ill patients infected by the disease. Currently, whatever number of hospitals and health facilities are there should be dedicated mostly to take care of above Covid-19 patients until unless there is an emergency of some other serious illness.

Hence, using the **Foursquare API** we can help locate common venues with hospitals in the neighbourhoods. Since hospitals and other healthcare facilities seems to be very less in the neighbourhoods of Brent which is worst effected, therefore plan should be to convert certain places like malls and restaurants which are closed for now to quarantine places or temporary hospitals where large number of patients can be taken care.

So, that where home quarantine for some people is not possible their institutional quarantine should be provided so as to curb down the further spread of the coronavirus infections.

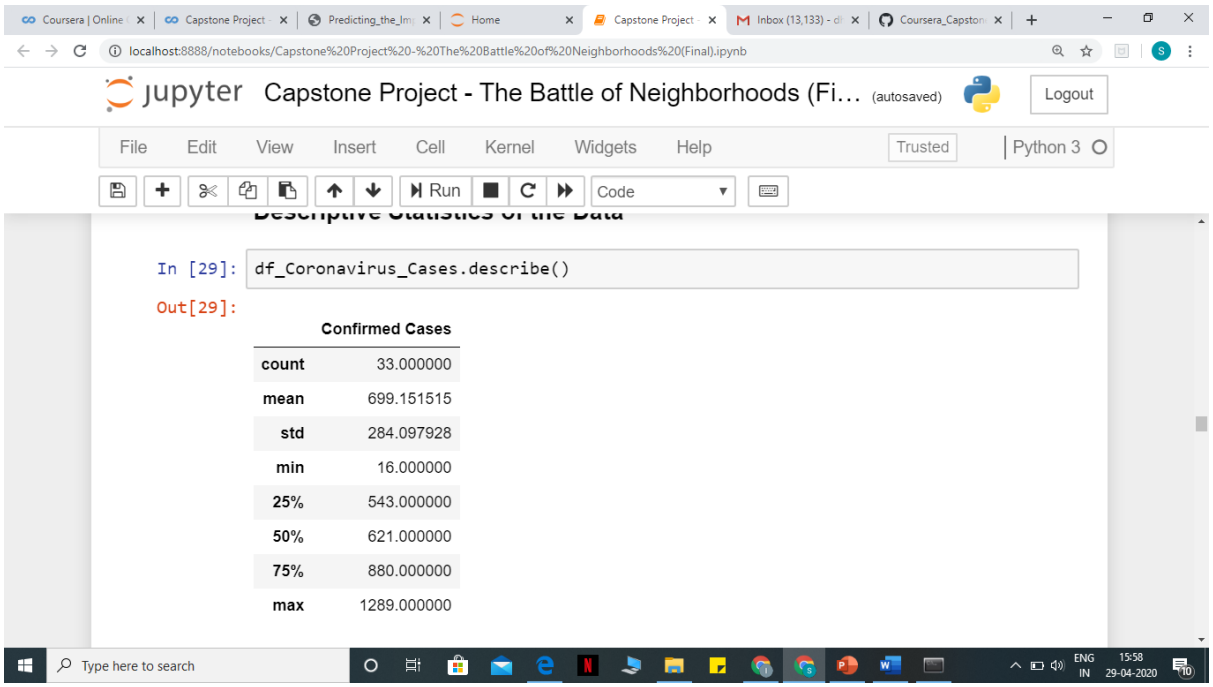
We can select the cluster with maximum no. of restaurants and studios available which can be utilized to fight the coronavirus pandemic in the region.

3. Methodology

3.1 Exploratory Data Analysis

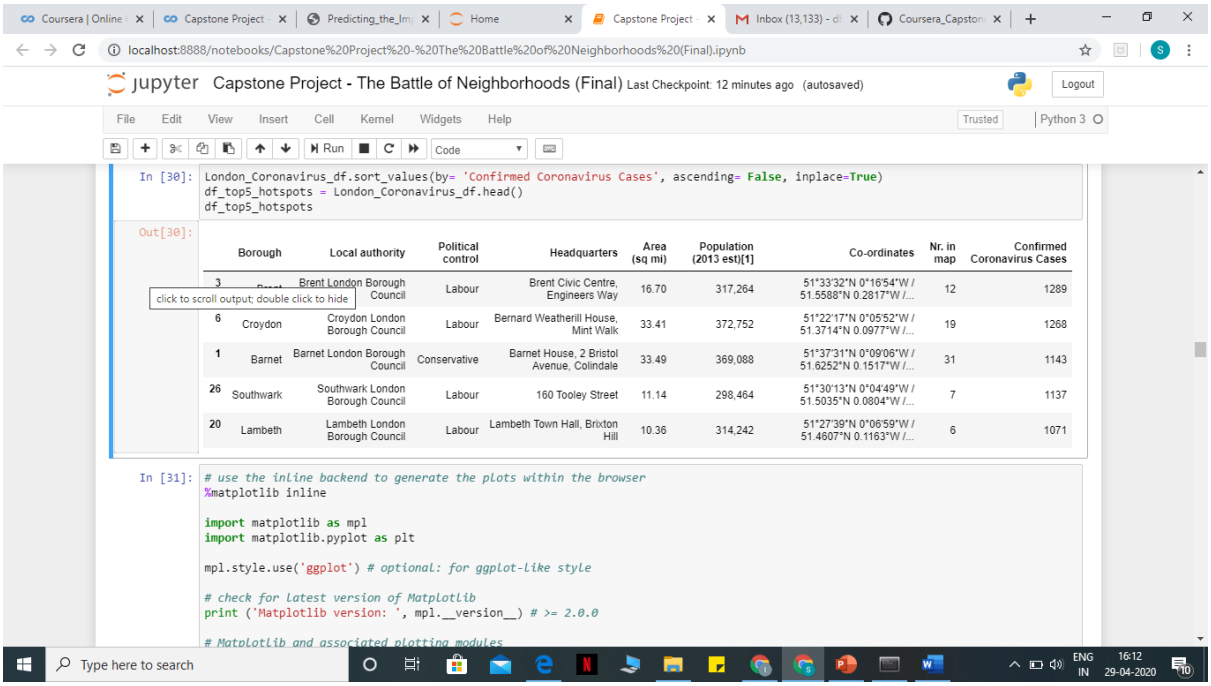
3.1.1 Statistical Summary of Data

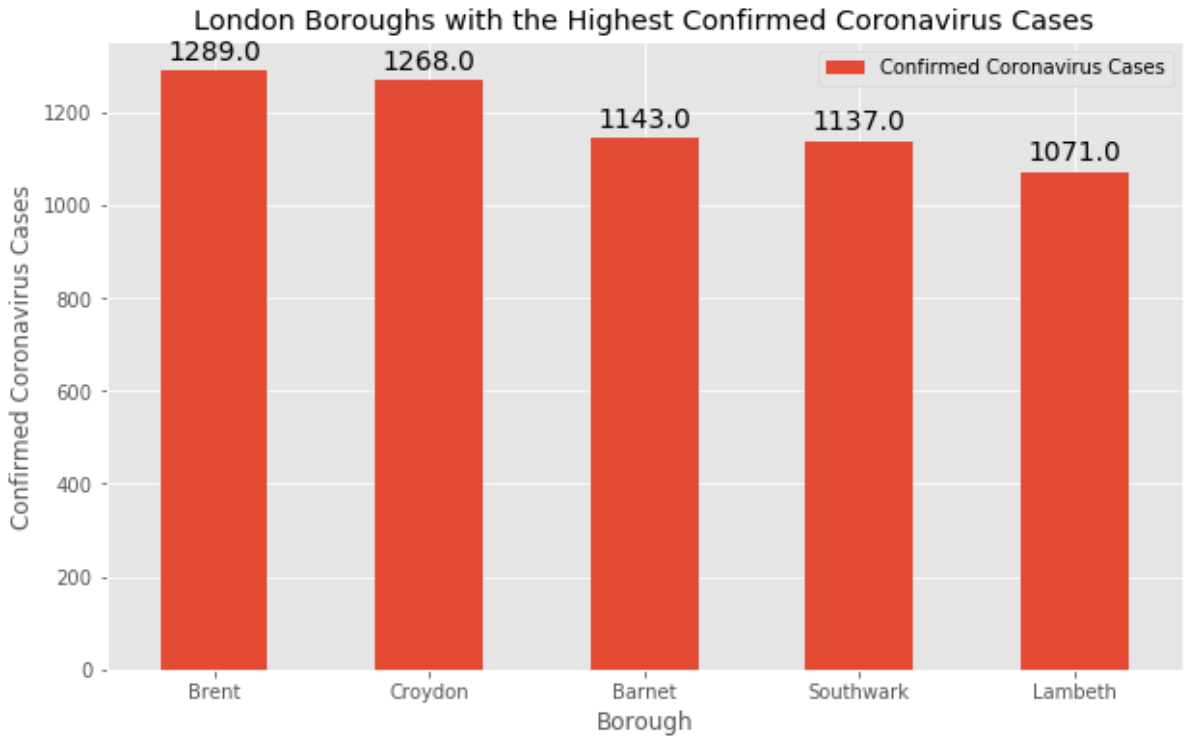
We can get the statistics of London Coronavirus data by using the describe function. The function returns mean, standard deviation, minimum, maximum, 1st quartile(25%), 2nd quartile(50%) and third quartile(75%) and total number of boroughs effected by coronavirus infections which is 33.



3.1.2 Boroughs with maximum number of Coronavirus infected cases

Figure out the top five 5 boroughs with maximum number of coronavirus positive cases.

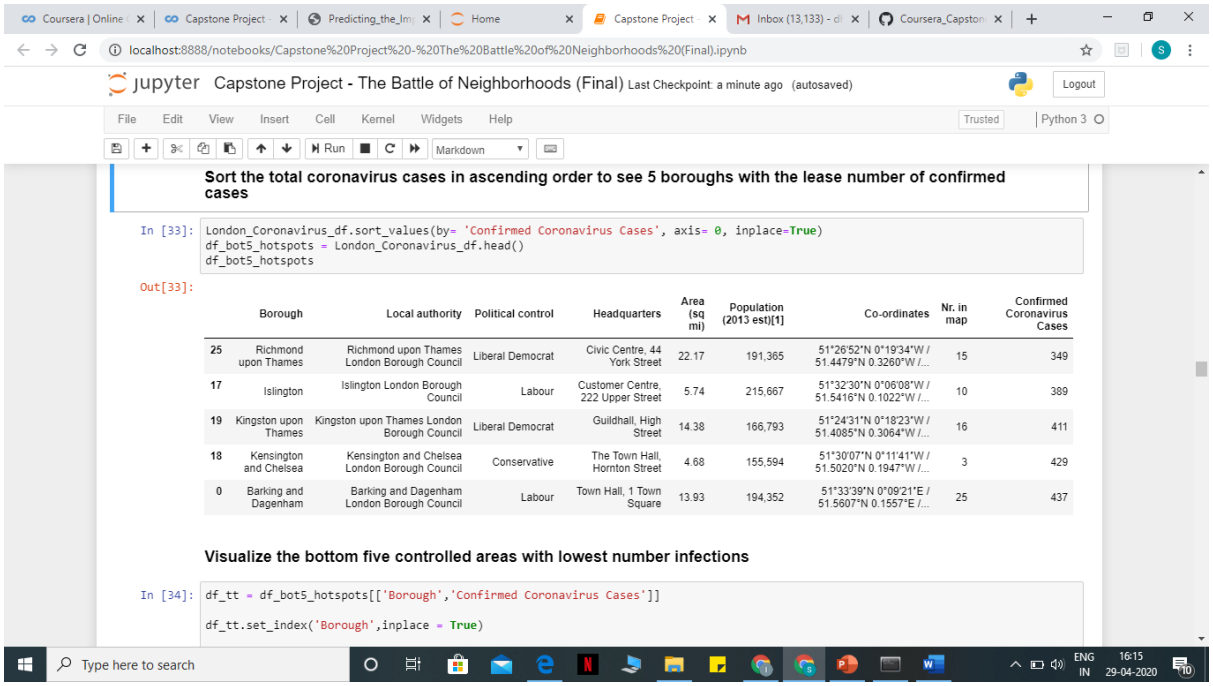




So, Brent is the Borough with the highest number of confirmed positive cases.

3.1.3 Boroughs with minimum number of Coronavirus infected cases

Figure out the top five 5 boroughs with lowest number of coronavirus positive cases.



Sort the total coronavirus cases in ascending order to see 5 boroughs with the lease number of confirmed cases

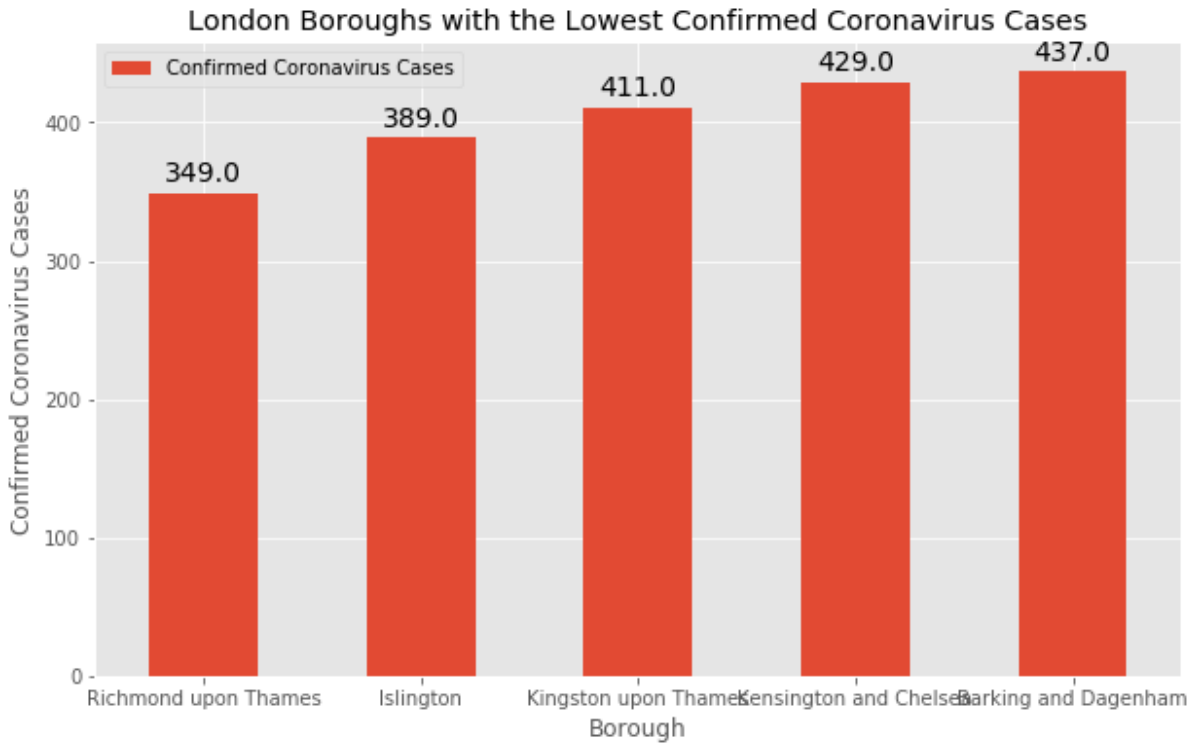
```
In [33]: London_Coronavirus_df.sort_values(by= 'Confirmed Coronavirus Cases', axis= 0, inplace=True)
df_bot5_hotspots = London_Coronavirus_df.head()
df_bot5_hotspots
```

Out[33]:

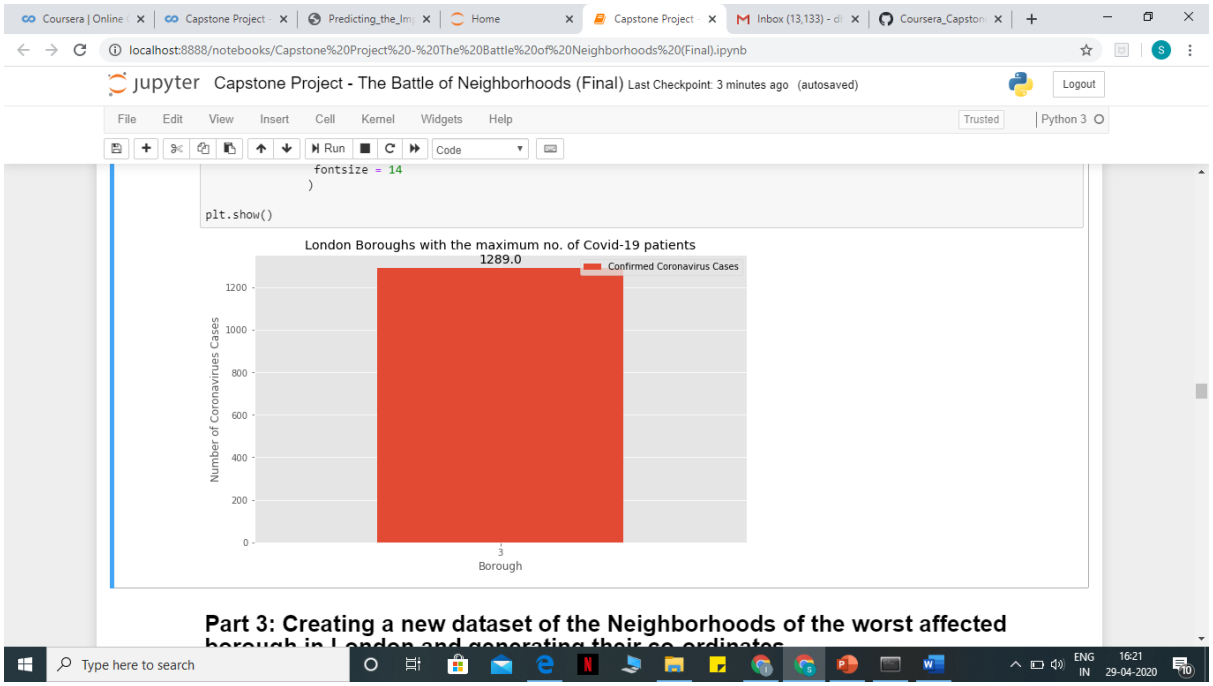
	Borough	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Nr. in map	Confirmed Coronavirus Cases
25	Richmond upon Thames	Richmond upon Thames London Borough Council	Liberal Democrat	Civic Centre, 44 York Street	22.17	191,365	51°26'52"N 0°19'34"W / 51.4479°N 0.3260°W /...	15	349
17	Islington	Islington London Borough Council	Labour	Customer Centre, 222 Upper Street	5.74	215,667	51°32'30"N 0°06'08"W / 51.5416°N 0.1022°W /...	10	389
19	Kingston upon Thames	Kingston upon Thames London Borough Council	Liberal Democrat	Guildhall, High Street	14.38	166,793	51°24'31"N 0°18'23"W / 51.4085°N 0.3064°W /...	16	411
18	Kensington and Chelsea	Kensington and Chelsea London Borough Council	Conservative	The Town Hall, Hornton Street	4.68	155,594	51°30'07"N 0°11'41"W / 51.5020°N 0.1947°W /...	3	429
0	Barking and Dagenham	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194,352	51°33'39"N 0°09'21"E / 51.5607°N 0.1557°E /...	25	437

Visualize the bottom five controlled areas with lowest number infections

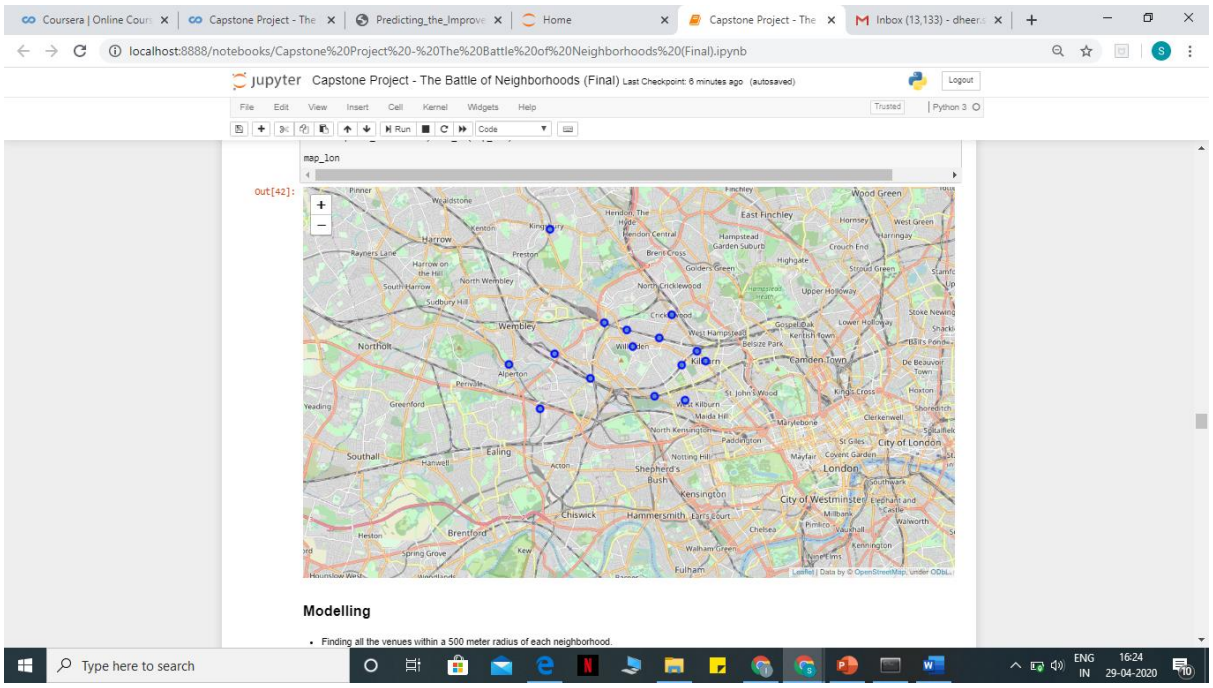
```
In [34]: df_tt = df_bot5_hotspots[['Borough','Confirmed Coronavirus Cases']]

df_tt.set_index('Borough',inplace = True)
```



After looking at the above data we found that Brent is the place with the maximum number of coronavirus cases and this is the place where the strict government guidelines to be followed and measures to be taken to control the current spread of the virus.



3.1.4 Visualize the neighbourhood of Brent Borough

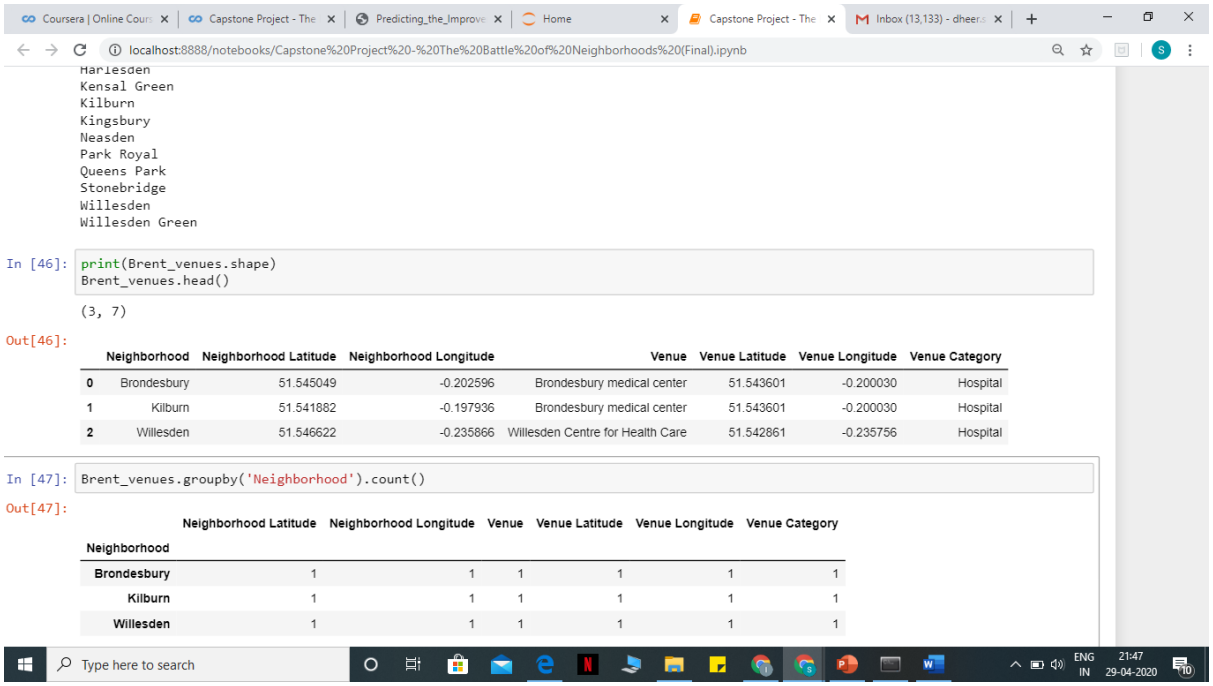


3.2 Modelling

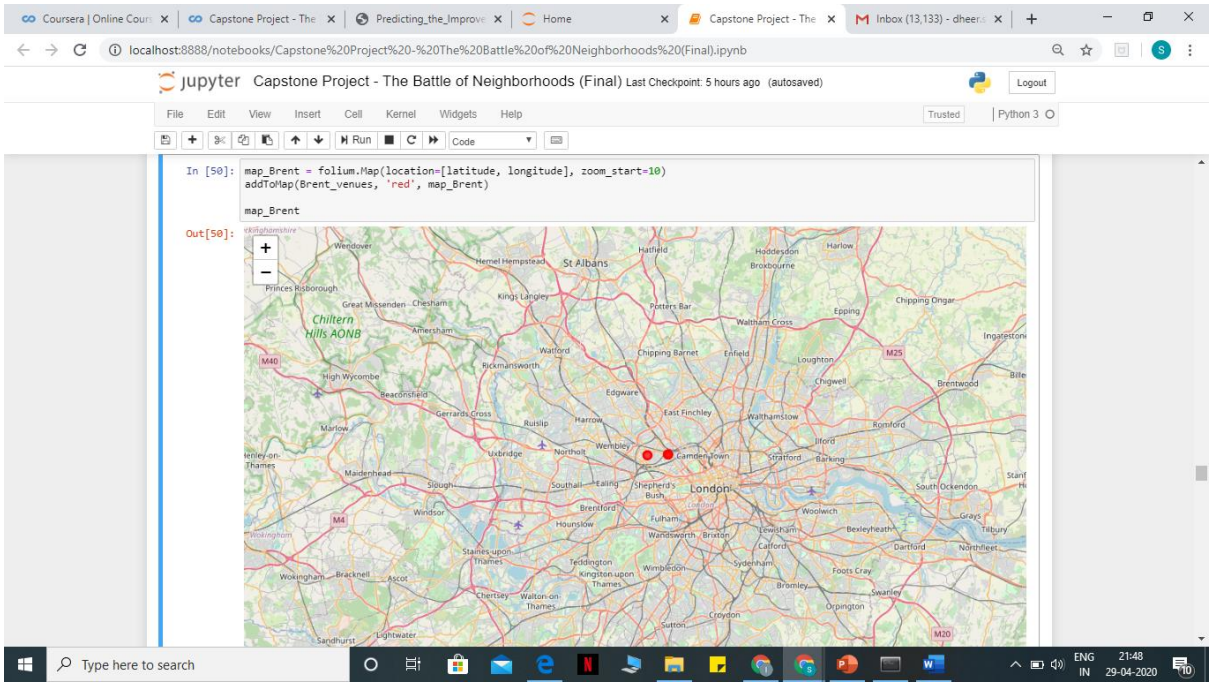
To help people find similar neighborhoods in the worst affected borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. Initially we will figure out the hospitals in the neighbourhood of that borough for which we will use 2 clusters, since there are very limited hospitals. Then

we will try to find out all the public venues which are closed for now and can be voluntarily converted to health care and quarantine facilities. Here we will use a cluster size of 5 that will cluster the neighbourhoods into 5 clusters. The reason to conduct a K- means clustering is to cluster neighborhoods with similar venues together so that they can be shortlisted by the health care workers to control the current spread of the virus.

Using the final dataset containing the neighbourhoods in Brent along with latitudes and longitudes, we can get the hospitals located within 500 metre radius using Foursquare API. This returns the json file containing the hospitals in each neighbourhood. This json file id converted to dataframe containing hospitals with its co-ordinates and category.



Visualizing the hospitals in neighbourhood we gat below map:



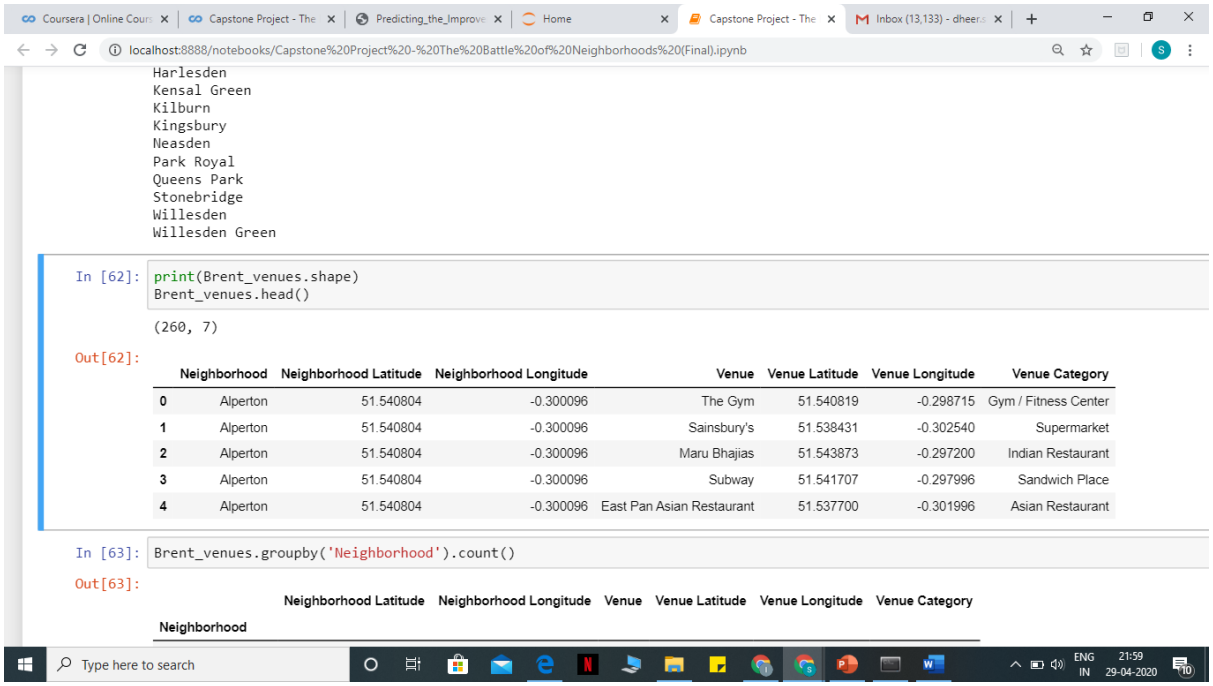
Next step is to Analyse each neighbour using one hot encoding on hospital data.

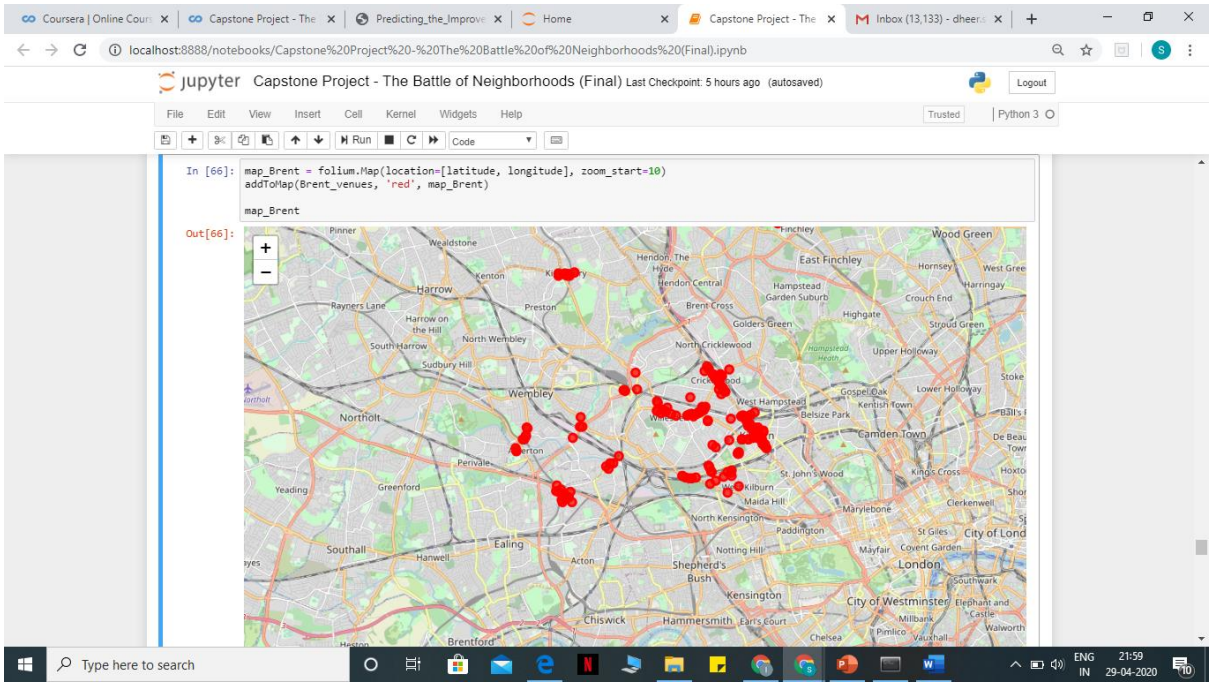
(One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithm which can do a better job in prediction.) The venue data is later grouped by neighbourhood and means of the hospitals are calculated.

Later on, clustering is done using k-means clustering i.e. unsupervised ML algorithm that cluster data based predefined cluster size. We are choosing 2 for hospitals since not many hospitals are not there located in the map.

We perform the similar steps to figure out other venues which are voluntarily contributing converting themselves to health care centres and quarantine facilities for which we are choosing cluster size to 5.

Below are the details of that.

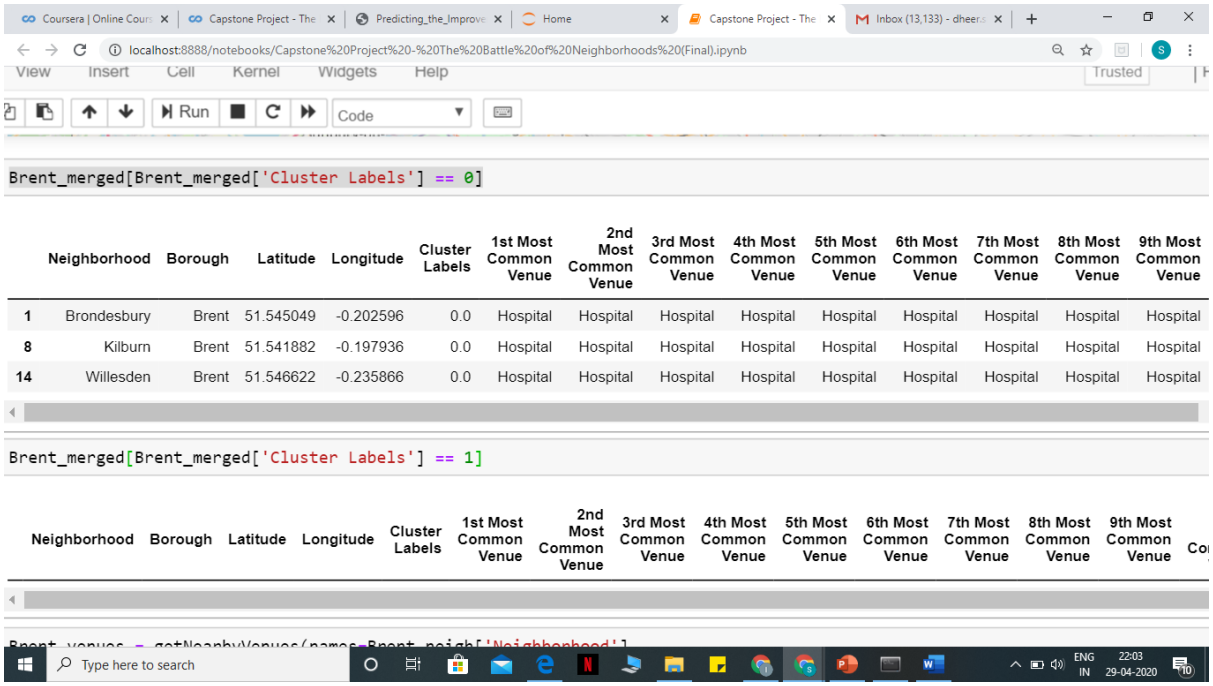




4. Results

After running the k-means clustering we can access each cluster to see neighbourhoods with hospitals and other venues belonging to each cluster.

1st Cluster of neighbourhoods with hospitals.



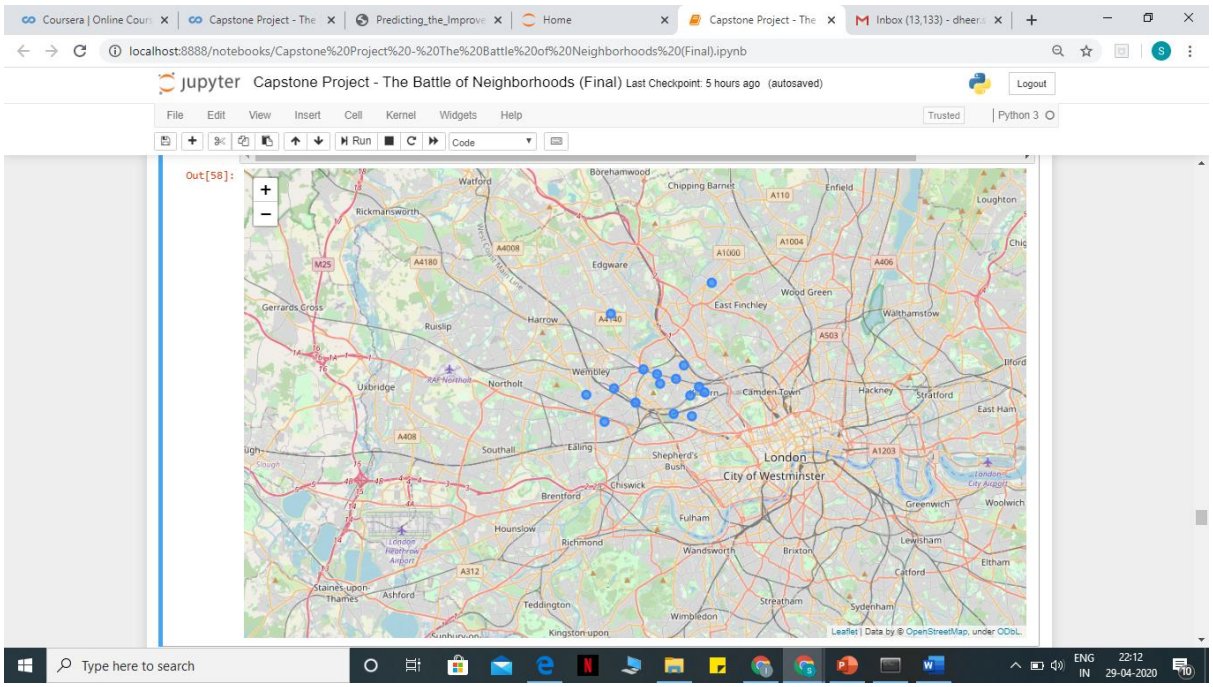
```
Brent_merged[Brent_merged['Cluster Labels'] == 0]
```

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
1	Brondesbury	Brent	51.545049	-0.202596	0.0	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital
8	Kilburn	Brent	51.541882	-0.197936	0.0	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital
14	Willesden	Brent	51.546622	-0.235866	0.0	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital	Hospital

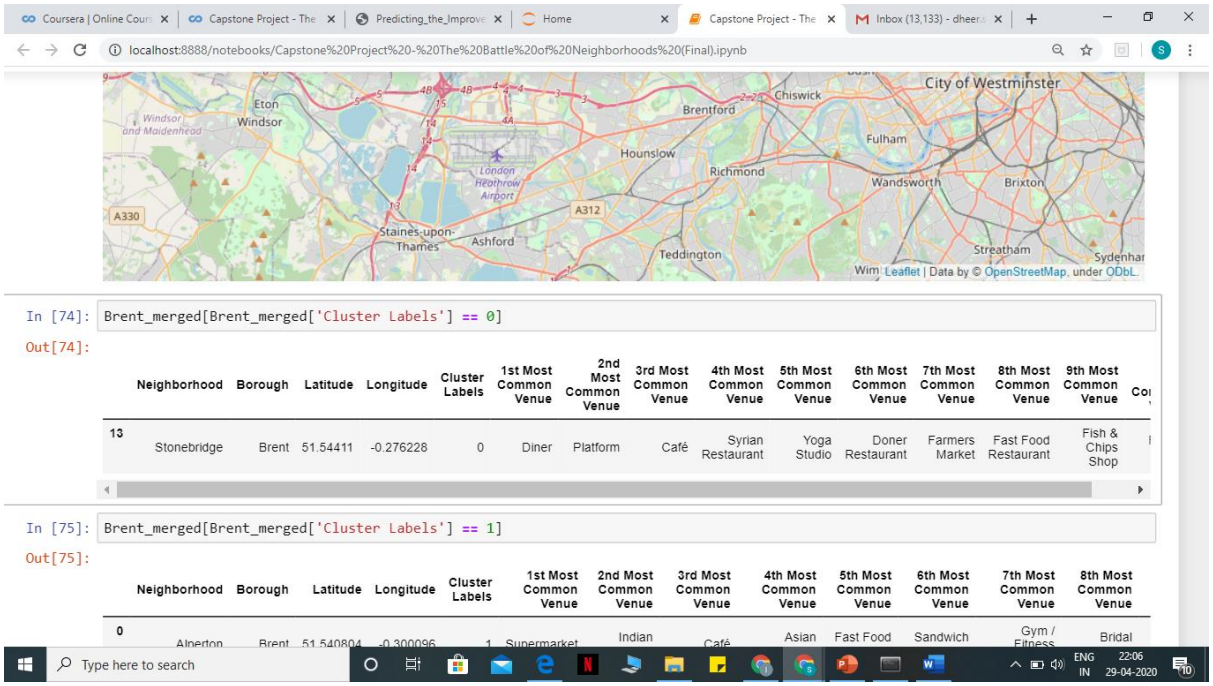
```
Brent_merged[Brent_merged['Cluster Labels'] == 1]
```

Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	Co

Visualizing the clustered neighbourhoods on a map using folium library



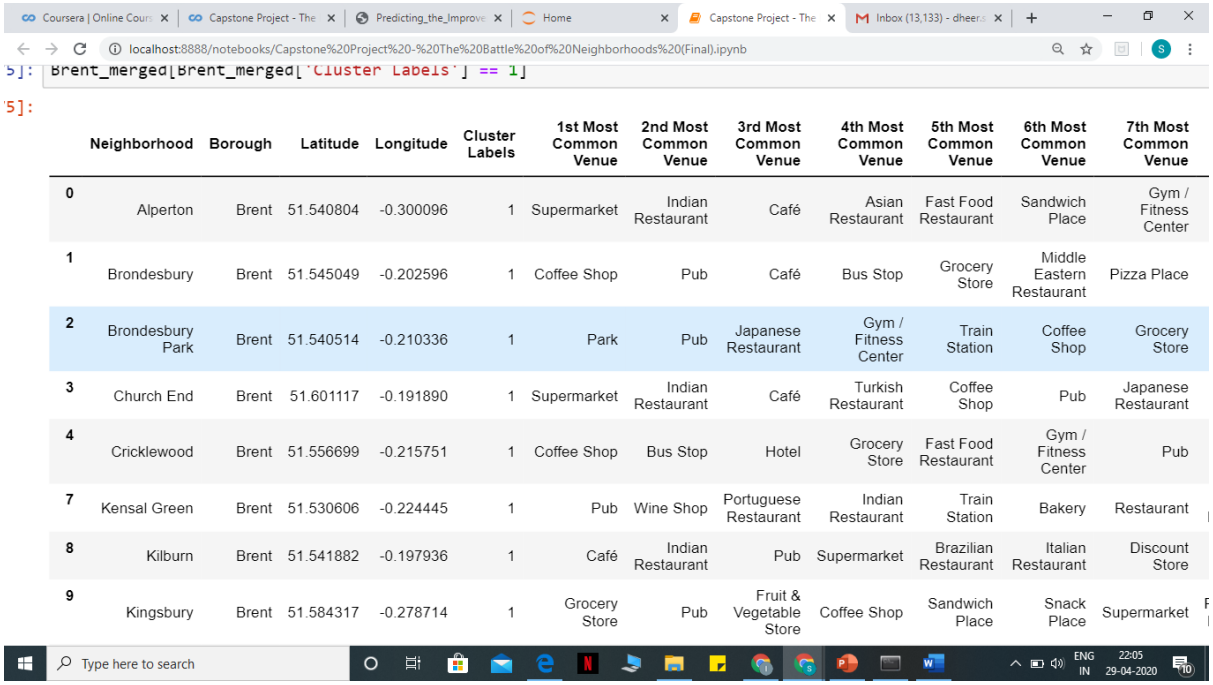
First Cluster with other venues



	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	Co
13	Stonebridge	Brent	51.54411	-0.276228	0	Diner	Platform	Café	Syrian Restaurant	Yoga Studio	Doner Restaurant	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	

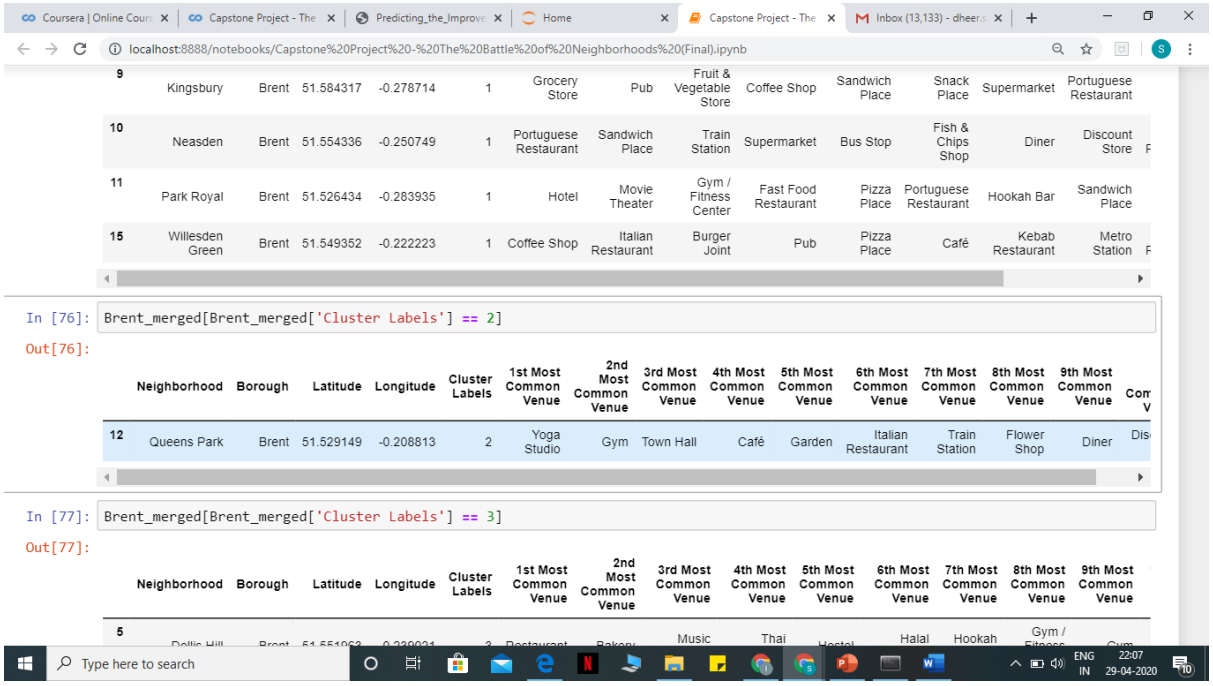
	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Alperton	Brent	51.540804	-0.300096	1	Supermarket	Indian Restaurant	Café	Asian Restaurant	Fast Food Restaurant	Sandwich Place	Gym / Fitness	Bridal

Second Cluster with other venues, we will choose this cluster due to maximum number of venues available and also neighbourhoods with hospitals will be situated nearer i.e in Brondesbury and Kilburn which will help the doctors to traverse from hospitals to the health care facilities.



	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	Alperton	Brent	51.540804	-0.300096	1	Supermarket	Indian Restaurant	Café	Asian Restaurant	Fast Food Restaurant	Sandwich Place	Gym / Fitness Center
1	Brondesbury	Brent	51.545049	-0.202596	1	Coffee Shop	Pub	Café	Bus Stop	Grocery Store	Middle Eastern Restaurant	Pizza Place
2	Brondesbury Park	Brent	51.540514	-0.210336	1	Park	Pub	Japanese Restaurant	Gym / Fitness Center	Train Station	Coffee Shop	Grocery Store
3	Church End	Brent	51.601117	-0.191890	1	Supermarket	Indian Restaurant	Café	Turkish Restaurant	Coffee Shop	Pub	Japanese Restaurant
4	Cricklewood	Brent	51.556699	-0.215751	1	Coffee Shop	Bus Stop	Hotel	Grocery Store	Fast Food Restaurant	Gym / Fitness Center	Pub
7	Kensal Green	Brent	51.530606	-0.224445	1	Pub	Wine Shop	Portuguese Restaurant	Indian Restaurant	Train Station	Bakery	Restaurant
8	Kilburn	Brent	51.541882	-0.197936	1	Café	Indian Restaurant	Pub	Supermarket	Brazilian Restaurant	Italian Restaurant	Discount Store
9	Kingsbury	Brent	51.584317	-0.278714	1	Grocery Store	Pub	Fruit & Vegetable Store	Coffee Shop	Sandwich Place	Snack Place	Supermarket

Third Cluster with venues



	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	Com V
9	Kingsbury	Brent	51.584317	-0.278714	1	Grocery Store	Pub	Fruit & Vegetable Store	Coffee Shop	Sandwich Place	Snack Place	Supermarket	Portuguese Restaurant		
10	Neasden	Brent	51.554336	-0.250749	1	Portuguese Restaurant	Sandwich Place	Train Station	Supermarket	Bus Stop	Fish & Chips Shop	Diner	Discount Store		
11	Park Royal	Brent	51.526434	-0.283935	1	Hotel	Movie Theater	Gym / Fitness Center	Fast Food Restaurant	Pizza Place	Portuguese Restaurant	Hookah Bar	Sandwich Place		
15	Willesden Green	Brent	51.549352	-0.222223	1	Coffee Shop	Italian Restaurant	Burger Joint	Pub	Pizza Place	Café	Kebab Restaurant	Metro Station		

```
In [76]: Brent_merged[Brent_merged['Cluster Labels'] == 2]
```

Out[76]:

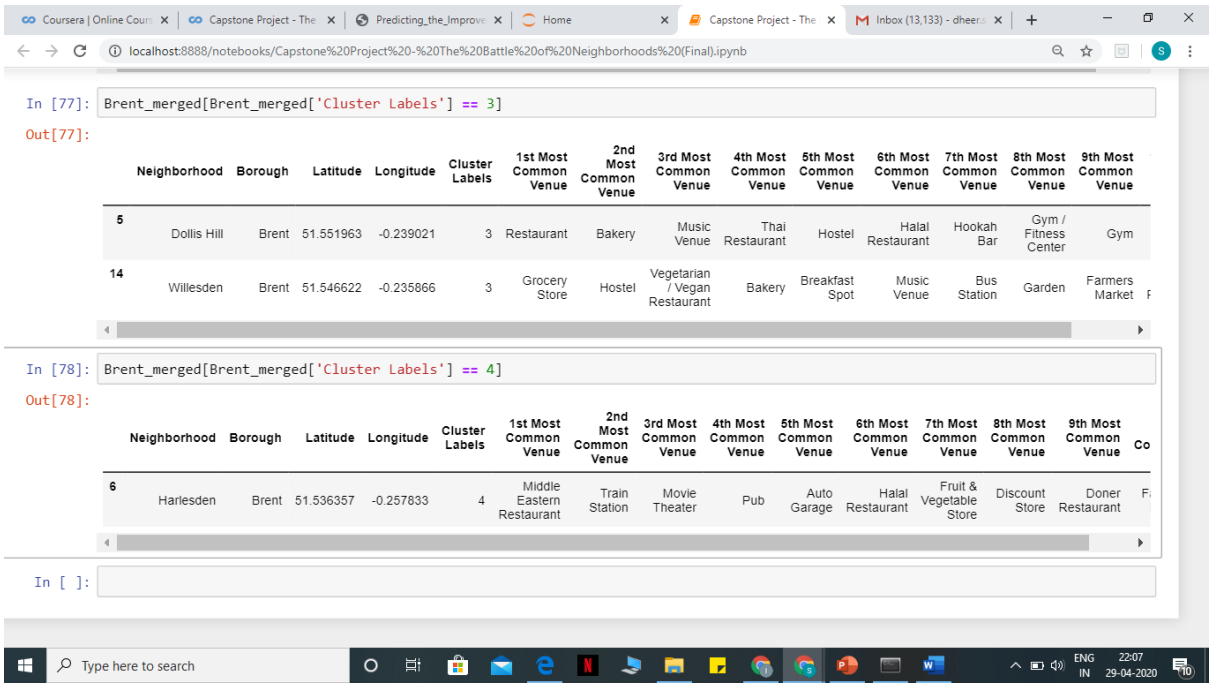
	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	Com V
12	Queens Park	Brent	51.529149	-0.208813	2	Yoga Studio	Gym	Town Hall	Café	Garden	Italian Restaurant	Train Station	Flower Shop	Diner	Dis

```
In [77]: Brent_merged[Brent_merged['Cluster Labels'] == 3]
```

Out[77]:

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
5	Dollis Hill	Brent	51.551963	-0.239021	3	Restaurant	Bakery	Music Venue	Thai Restaurant	Hostel	Halal Restaurant	Hookah Bar	Gym / Fitness Center	Gym

Fourth and Fifth Cluster with Venues, we can consider this fourth cluster also since it has the neighbourhood i.e Willesden with hospital nearby.



```
In [77]: Brent_merged[Brent_merged['Cluster Labels'] == 3]
```

Out[77]:

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
5	Dollis Hill	Brent	51.551963	-0.239021	3	Restaurant	Bakery	Music Venue	Thai Restaurant	Hostel	Halal Restaurant	Hookah Bar	Gym / Fitness Center	Gym
14	Willesden	Brent	51.546622	-0.235866	3	Grocery Store	Hostel	Vegetarian / Vegan Restaurant	Bakery	Breakfast Spot	Music Venue	Bus Station	Garden	Farmers Market

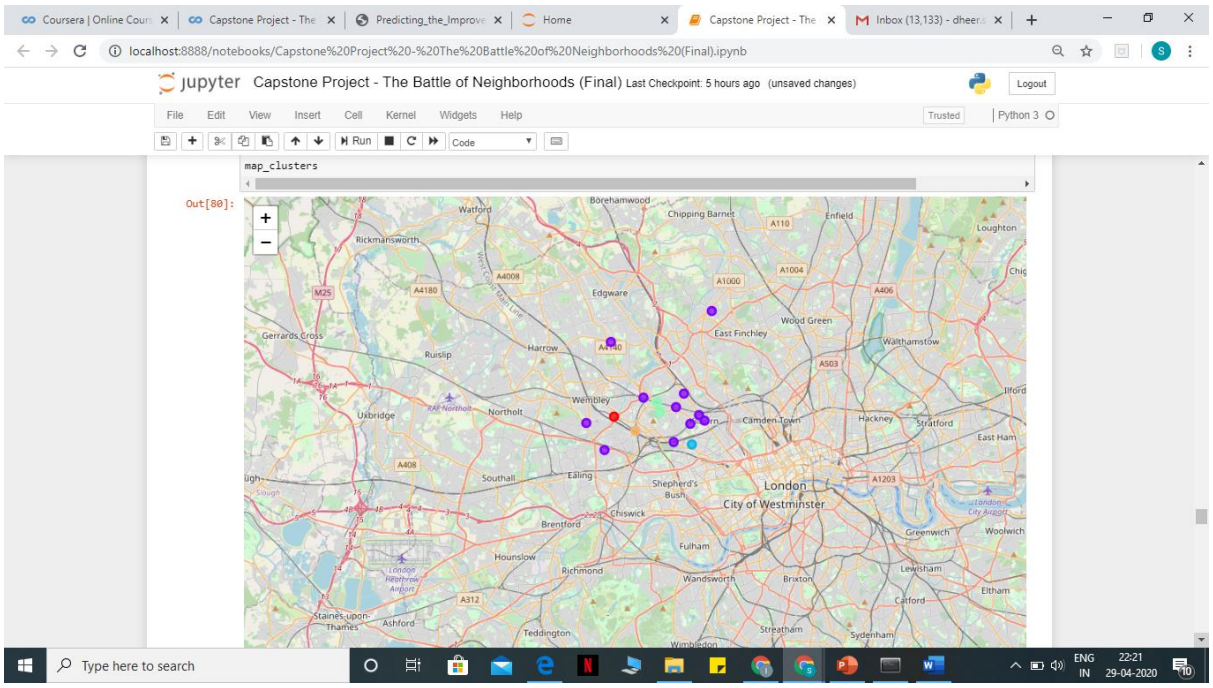
```
In [78]: Brent_merged[Brent_merged['Cluster Labels'] == 4]
```

Out[78]:

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	Co
6	Harlesden	Brent	51.536357	-0.257833	4	Middle Eastern Restaurant	Train Station	Movie Theater	Pub	Auto Garage	Halal Restaurant	Fruit & Vegetable Store	Discount Store	Doner Restaurant	F

```
In [ ]:
```

Visualizing a clustered map using folium library



Each cluster is color coded for the ease of presentation. We can see that majority of neighbourhood falls in purple which is second cluster. First, Third and fifth cluster are in Red, Yellow and blue. Whereas fourth cluster is lemon green in color.

5. Discussion

The project is created by looking at the current global scenario of coronavirus pandemic. No wonder how long it prevails in the global market and how worst effect it can have to the economy of the countries round the world. The aim of the project is to assist the present government of London as well the health care ministry in tackling the spread of coronavirus in the worst effected areas(Boroughs) of London.

The availability of hospitals and health care facilities should be the prior need of all people living there. Hence considering that the 3 neighbourhoods have been figured out for the worst affected region there. Hence cluster 1 should be preferred first.

For creating institutional quarantine facilities, second cluster and fourth cluster with maximum number of venues as well close proximity of the hospitals should be chosen. Apart from that other facilities like voluntarily contribution of restaurant owners for providing online food deliveries at this time of crisis should be considered as one major criteria as a preferred cluster. Apart form that there are certain clusters like 3rd & 4th which has townhall & hostel facilities and are closed for now and hence can be converted to Institutional quarantine facilities.

First and fifth cluster should be considered for there containment of food and vegetable supplies to the places which are strictly under lockdown.

Similarly, third cluster has a good transportation facility which is suitable for movement of the required emergency medicines as well as medical equipment to meet the enormous demand.

6. Conclusion

This project provides a better understanding of the clustering of the neighbourhoods in terms of common venues or specific places based on certain categories mentioned like hospitals, gyms, restaurant, grocery stores, etc. Hence, we can efficiently use data science to solve the problem of coronavirus spread in certain hotspots of the country where the infections have gone out of control and risking the life of people in those areas. The project can be further enhanced to building of more and more hospitals and transfer of medicines and medical equipment for the treatment of Covid-19 critically ill patients.