Capstone Project - The Battle of Neighborhoods
Report

Shasank M. Dave April,2020

1. Introduction

1.1 Background

The outbreak was identified in Wuhan, China, in December 2019, declared to be a Public Health Emergency of International Concern on 30 January 2020, and **recognized as a pandemic by the World Health Organization on 11 March 2020**. As of 16 April 2020, more than 2.1 million cases of COVID-19 have been reported in 210 countries and territories, resulting in more than 140,000 deaths. More than 532,000 people have recovered, although there may be a possibility of relapse or reinfection. The deaths per diagnosed cases varies significantly between countries

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Most people who fall sick with COVID-19 will experience mild to moderate symptoms and recover without special treatment.

1.2 Problem

Fortunately we are in times of social media and internet to keep us update with all the information, as social media has negative side to, with all fake news making rounds on social media which ill-informs manys

1.3 Interest

So with this project it is my attempt to plot and visualize correct information from genuine sources to spread correct information.

Scope of this project is to do **analysis & clustering** on COVID-19 zones in **Mumbai, Maharashtra, India(The most effected city in India)** & further Map testing labs/Hospitals(**Get information using foursquare**) for COVID-19 in mumbai

2. Data Acquisition and Cleaning

2.1 Data Acquisition

Based on definition of our problem, For COVID-19 following are the main data scores:

- https://www.kaggle.com/sudalairajkumar/covid19-in-india
- Information tracked on daily basis https://api.covid19india.org/data.jso
- Number of Active Cases/Cured Cases/Deaths
- Data for mumbai covid-19 cases is sourced from http://stopcoronavirus.mcgm.gov.in/ https://github.com/shasaankdave/Coursera_Capstone/blob/master/Containment_Zones_BMC_Mumbai.pdf
- Details about mumbai Pincodes, Area,Borough(Called BMC Wards) are sourced from https://data.gov.in/resources/all-india-pincode-directory-along-contact-details & only mumbai's pincodes & details are extracted.

- We decided to use folium maps with latitude & longitude to display covid-19 zones, cluster & testing facilities Latitude & Longitude based on Pincodes of mumbai are obtained using **Google Maps API geocoding**

- Details about COVID-19 testing labs in mumbai are collect as follows**

- Testing Labs in mumbai are selected using https://www.kaggle.com/sudalairajkumar/covid19-in-india

- Further Details about Hospitals/testing labs are got from **Foursquare API** using categoryId parameter** in the request URL to select hostipal as venue category
  example venue categoryId for hospitals : “4bf58dd8d48988d196941735”
  Refer for more details :
  https://developer.foursquare.com/docs/build-with-foursquare/categories/

2.2 Data Cleaning

Data for mumbai city is extracted from an pdf on http://stopcoronavirus.mcgm.gov.in/
Mumbai’s pincodes are selected from all india pincode directory.
latitude & Longitude data for hospital data was reported to be wrongm hence we use **Geolocator API** as as follows to get co-ordinates:

Found Latitude on Longitude reported Above as not correct.

Hence Use GeoLocator API to get authentic Latitude & Longitude

```
In [35]: lat=[]
         long=[]
         pincd=[]

         #address = 'Mumbai, India'
         for pincode in mumbai_labs_ll['Pincode']:
             add= str(pincode) + ',india'
             address = add
             geolocator = Nominatim(user_agent="can_explorer")
             location = geolocator.geocode(address,timeout=15)
             latitude = lat.append(location.latitude)
             longitude =long.append(location.longitude)
             pincd.append(pincode)
             print('The geograpical coordinate found for {}.'.format(add))

         The geograpical coordinate found for 400012,india.
         The geograpical coordinate found for 400034,india.
         The geograpical coordinate found for 400012,india.
         The geograpical coordinate found for 400008,india.
         The geograpical coordinate found for 400012,india.
         The geograpical coordinate found for 444001,india.
         The geograpical coordinate found for 400012,india.
         The geograpical coordinate found for 400053,india.
         The geograpical coordinate found for 400070,india.
         The geograpical coordinate found for 400062,india.
         The geograpical coordinate found for 411004,india.
         The geograpical coordinate found for 400053,india.
         The geograpical coordinate found for 400072,india.
         The geograpical coordinate found for 400012,india.
         The geograpical coordinate found for 400092,india.
         The geograpical coordinate found for 400080,india.
         The geograpical coordinate found for 400016,india.
         The geograpical coordinate found for 400064,india.
         The geograpical coordinate found for 400016,india.
```

Further data is merged with pincode details:

Out[38]:

	lab	address	Pincode	city	state	type	Latitude	Longitude
0	Seth GS Medical College & KEM Hospital, Mumbai	Seth GS Medical College & KEM Hospital, Achary...	400012	Mumbai	Maharashtra	Government Laboratory	19.000795	72.840147
1	Kasturba Hospital for Infectious Diseases, Mumbai	Kasturba Hospital for Infectious Diseases, San...	400034	Mumbai	Maharashtra	Government Laboratory	18.976686	72.812367
2	National Institute of Virology Field Unit, Mumbai	National Institute of Virology Field Unit, Haf...	400012	Mumbai	Maharashtra	Government Laboratory	19.000795	72.840147
3	Grant Medical College & Sir JJ Hospital, Mumbai	Grant Medical College & Sir JJ Hospital, J J M...	400008	Mumbai	Maharashtra	Government Laboratory	18.970566	72.833956
4	Haffkine Institute, Mumbai	Haffkine Institute, Acharya Donde Marg, Parel,...	400012	Mumbai	Maharashtra	Government Laboratory	19.000795	72.840147

Foursquare Data to get hospital details, venue categoryId parameter used for hospital:

details on venue category ids Category id for **Hospitals** – 4bf58dd8d48988d196941735

In [41]:
```
#FourSquare API Parameter SetUp

LIMIT = 100

CLIENT_ID = 'ZVLUP1PU41P2XVZN1R15EWMRBMOUU5MA3HTHOYKFKZTHWMYJ' # your Foursquare ID
CLIENT_SECRET = 'ST0MKT1KNGXTV5PXYGRDHNUHA5XRYKTM5V4NGUQG4FHEP245' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version
CATEGORYID='4bf58dd8d48988d196941735' #venueid for hospitals

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
lat='19.25023195'
lng='73.16017493'
radius=500
```

```
Your credentails:
CLIENT_ID: ZVLUP1PU41P2XVZN1R15EWMRBMOUU5MA3HTHOYKFKZTHWMYJ
CLIENT_SECRET:ST0MKT1KNGXTV5PXYGRDHNUHA5XRYKTM5V4NGUQG4FHEP245
```

In [42]:
```
 # create the API request URL
radius=500
latitude=19.0025 #Mumbai
longitude=72.8421 #Mumbai
#for latitude,longitude in zip(mumbai_covid_clusters['Latitude'],mumbai_covid_clusters['Longitude']):
url = 'https://api.foursquare.com/v2/venues/search?categoryId={}&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit=
results = requests.get(url).json()["response"]#['groups'][0]['items']
    #results = requests.get(url).json()
results
```

3. Methodology

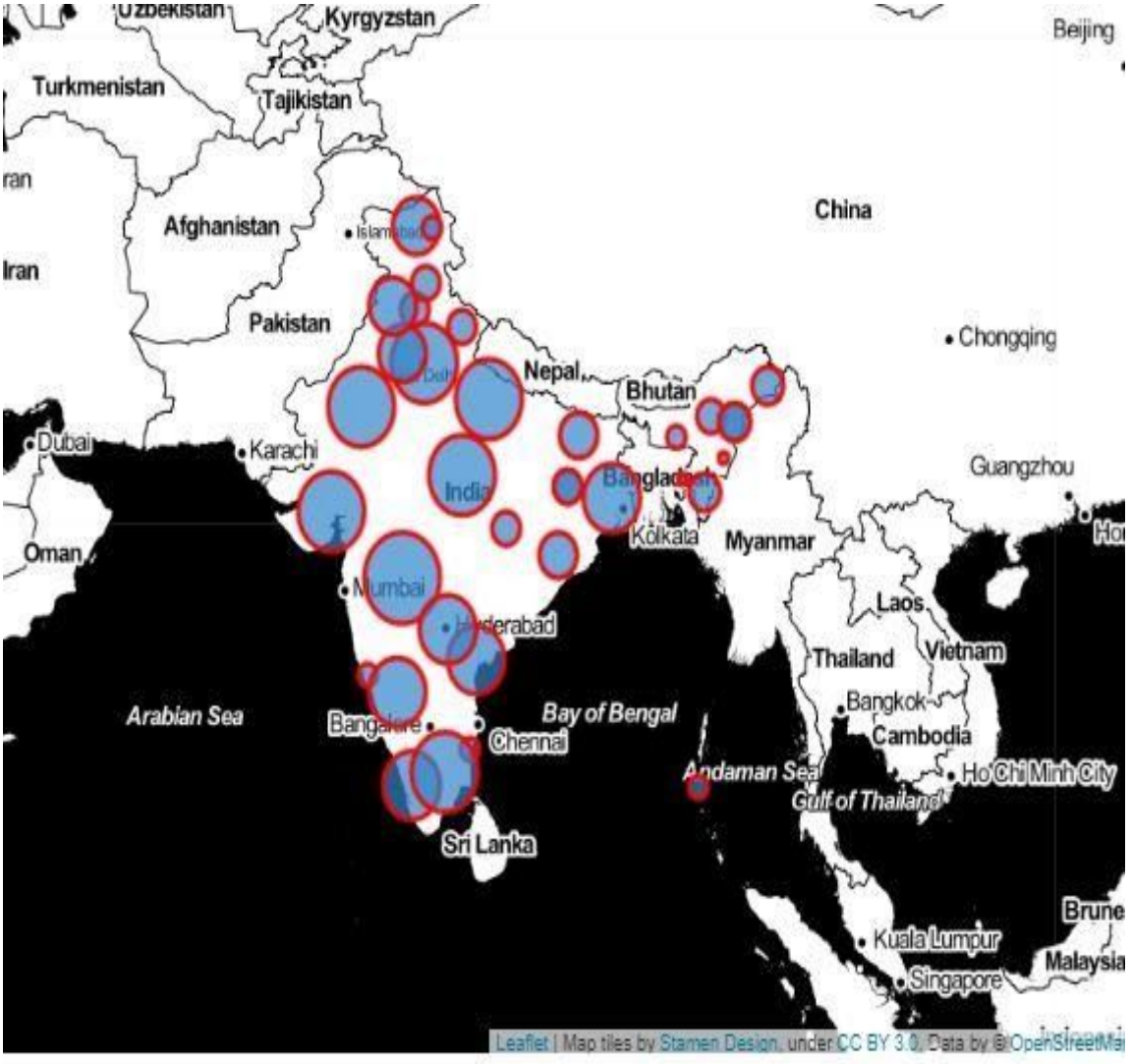
3.1 Statistics & Neighbourhood exploration/Selection:

In this project we will direct our efforts on finding covid-19 zones in mumbai and further find details and plot folium maps for:

- Covid-19 zones in Mumbai, India
- Covid-19 testing facilities in the neighbourhood
- Coivd-19 Clustered for mumbai neighbourhood.

In first step with "**COVID-19:INDIA**" above we tried to analyse covid-19 situation in INDIA with bar plot, line plot & folium maps to get the top effected city, after analysis we found **Mumbai, Maharashtra, India** is the most effected city in india, Hence we selected Mumbai as our neighbourhood for this capstone project.

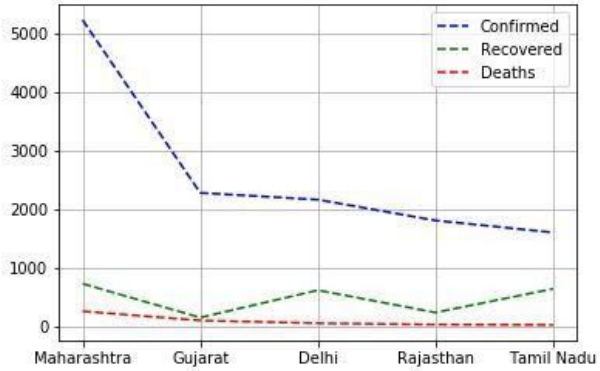
Map of India with COVID-19 Intensity:

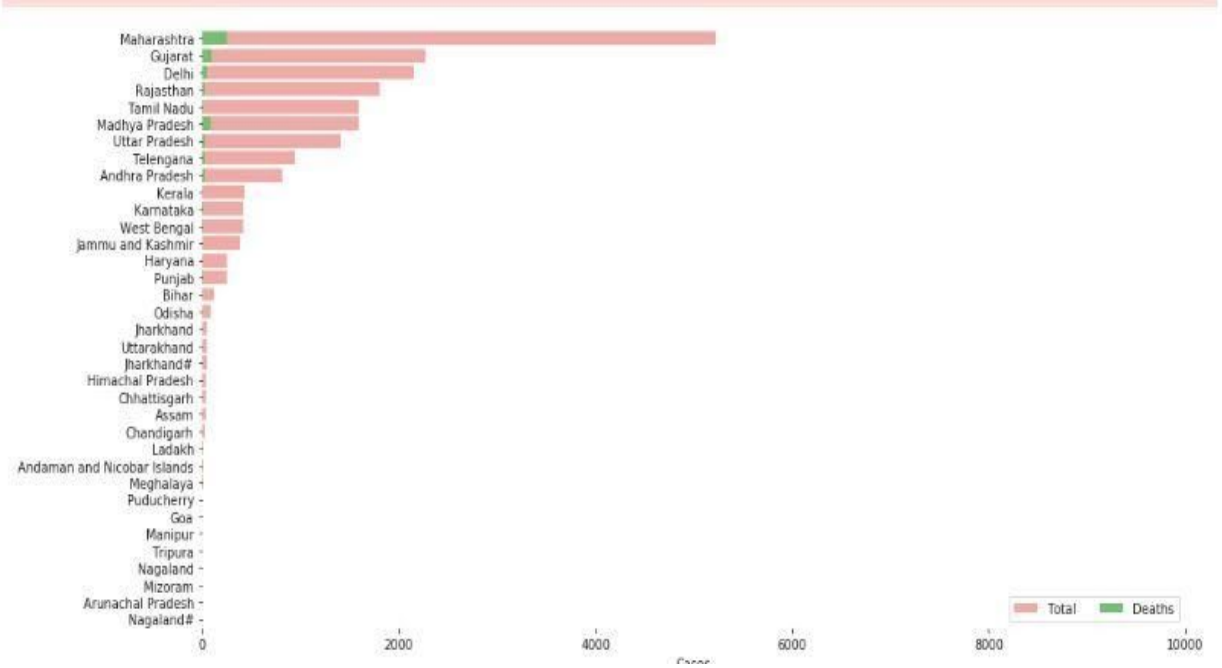


Exploration of data from india's dataset helped us find maharashtra state & mumbai city as epicenter for covid-19 in india, used following exploration::

States with Most Number of cases ¶

```
In [16]: plt.plot(india_head['States'], india_head['Confirmed'], 'b--', label='Confirmed')
         plt.plot(india_head['States'], india_head['Recovered'], 'g--', label='Recovered')
         plt.plot(india_head['States'], india_head['Deaths'], 'r--', label='Deaths')
         plt.grid()
         plt.legend() # add legend based on line labels
         plt.show()
```





By the end of step1 above we were able to identify mumbai as epicenter & used as neighbourhood for this project.

In Second step we tried get data for hospitals/testing labs in mumbai from source & tried to get details using **foursquare API** and all information for latitude & longitude was obtained using **Geolocator API**

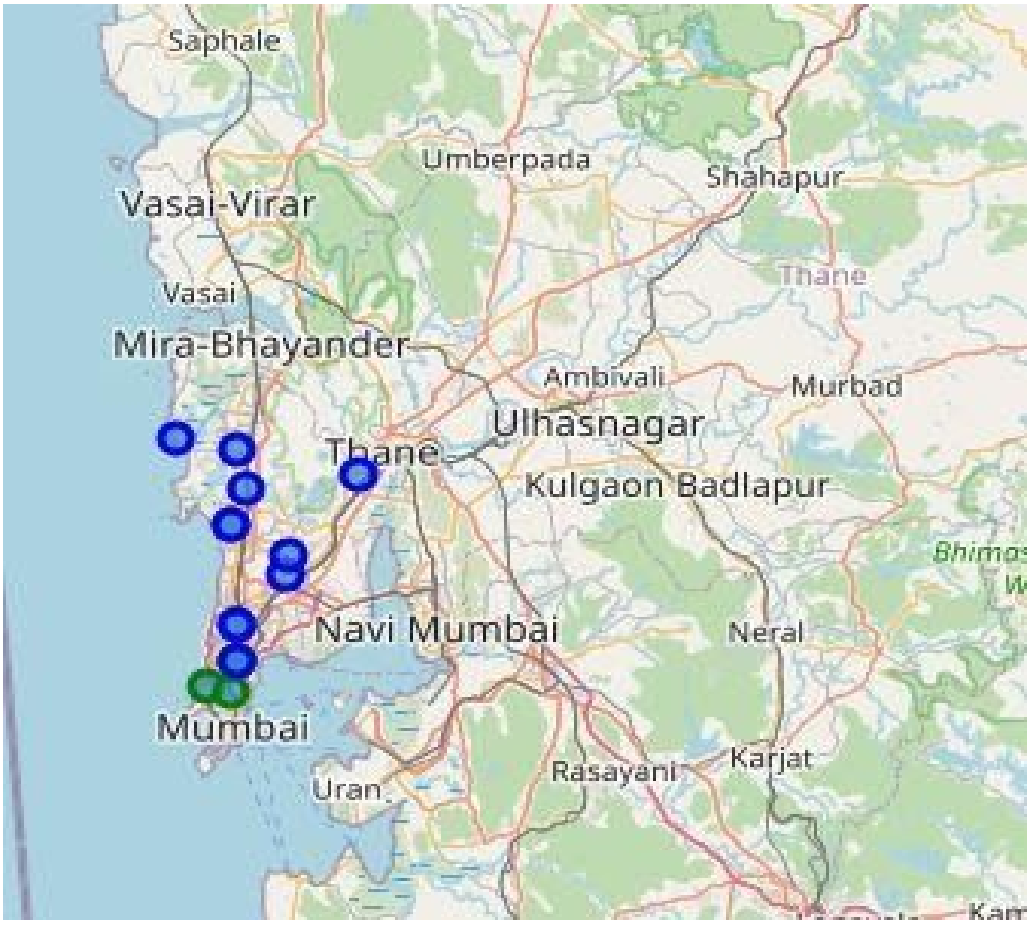
details on venue category ids Category id for Hospitals – 4bf58dd8d48988d196941735

In [41]:
```
#FourSquare API Parameter SetUp

LIMIT = 100

CLIENT_ID = 'ZVLUP1PU41P2XVZN1R15EWMRBMOUU5MA3HTHOYKFKZTHWMYJ' # your Foursquare ID
CLIENT_SECRET = 'ST0MKT1KNGXTVSPXYGRDHNUHA5XRYKTM5V4NGUQG4FHEP245' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version
CATEGORYID='4bf58dd8d48988d196941735' #venueid for hospitals

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
lat='19.25023195'
lng='73.16017493'
radius=500
```

```
Your credentails:
CLIENT_ID: ZVLUP1PU41P2XVZN1R15EWMRBMOUU5MA3HTHOYKFKZTHWMYJ
CLIENT_SECRET:ST0MKT1KNGXTVSPXYGRDHNUHA5XRYKTM5V4NGUQG4FHEP245
```

In [42]:
```
 # create the API request URL
radius=500
latitude=19.0025 #Mumbai
longitude=72.8421 #Mumbai
#for latitude,longitude in zip(mumbai_covid_clusters['Latitude'],mumbai_covid_clusters['Longitude']):
url = 'https://api.foursquare.com/v2/venues/search?categoryId={}&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit=
results = requests.get(url).json()["response"]#['groups'][0]['items']
    #results = requests.get(url).json()
results
```

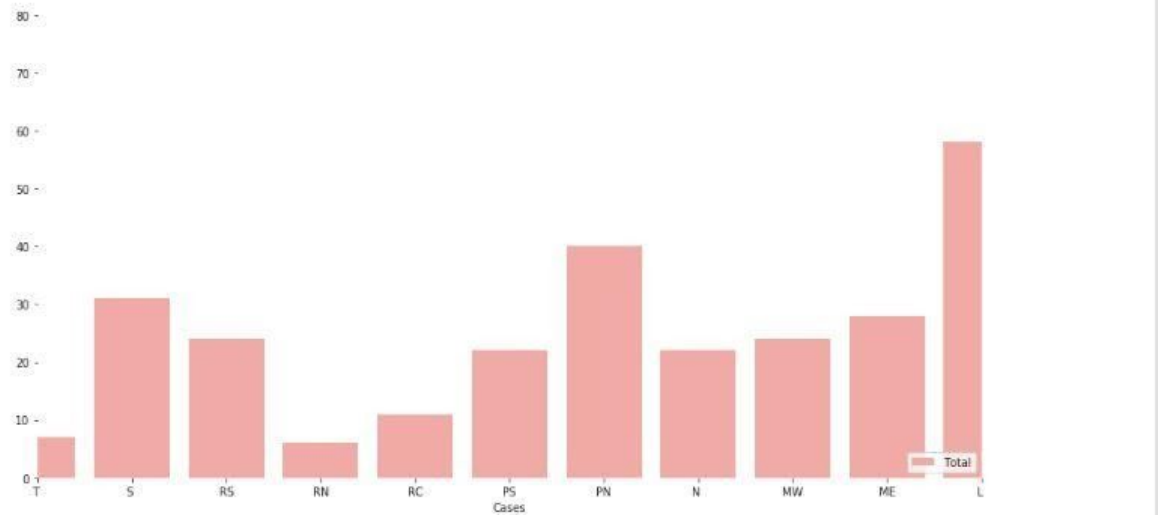
Mumbai Map for Hospital/Testing Lab, ColorCoded→Green for Government hospital & Blue for Non-Government Hospital.



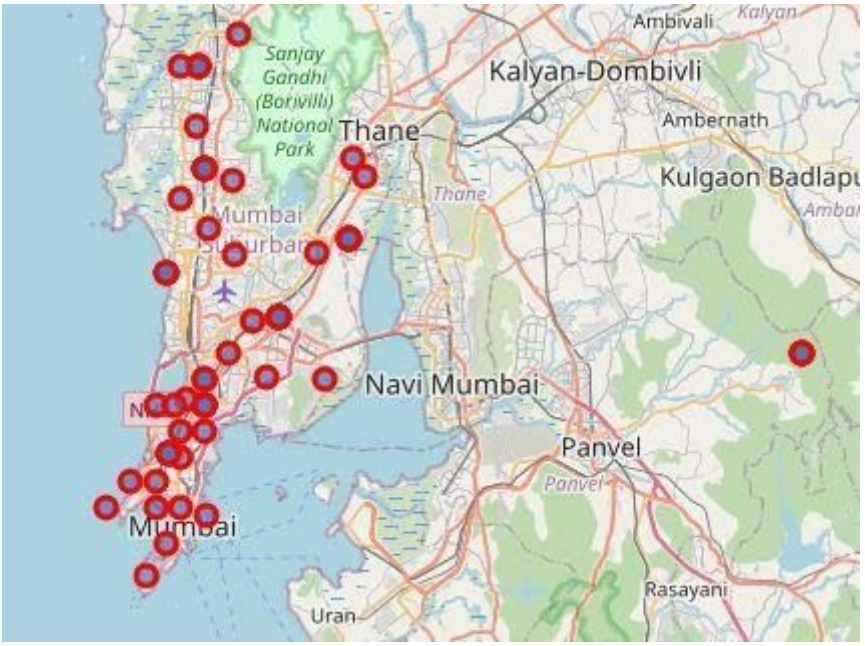
In Third step we started with getting information regarding covid-19 zones in mumbai, information is sourced from mumbai municipal corportation website, other needed information regarding postal code was found from government of india site and to have a clear picture of cases in mumbai we grouped mumbai covid zones as per boroughs(Called Wards in Mumbai) and plotted covid zones.

Bar Plot for wards with top cases:

```
            label= Total , color= r' )


# Add a legend and informative axis label
ax.legend(ncol=2, loc="lower right", frameon=True)
ax.set(xlim=(0, 10), ylabel="",
       xlabel="Cases")
sns.despine(left=True, bottom=True)
plt.show()
```



Mumbai COVID-19 Hotspot:



From the Map & bar chart we derive that south mumbai has maximum number of cases.

3.2 Modelling

In fourth and final step we tried to cluster mumbai covid-19 zone using k-mean clustering, to have clear label on clustering we plot a folium maps for clusters color-coded by cluster numbers., k-value for clustering was 3 as we have in total 21 wards.

4. Results

After running the K-means clustering we can access each cluster created to see which ward was assigned what cluster[0,1,2]

final result set for cluster data has following columns:

- Pincode
- Latitude
- Longitude
- Cluster_label(Refer fig4)

```
In [101]: mumbai_covid_clusters=pd.DataFrame({"Pincode": pincd_lst,
                                              "Latitude": covid_lat_lst,
                                              "Longitude":covid_long_lst,
                                              "Cluster_label":cluster_lst,
                                               })
          mumbai_covid_clusters.head()
```

Out[101]:

	Pincode	Latitude	Longitude	Cluster_label
0	400001	18.6291	72.8919	2
1	400002	17.0509	73.2910	2
2	400003	18.9500	72.8333	2
3	400004	18.9500	72.8167	2
4	400005	18.9069	72.8108	2

Fig 4 Cluster

Cluster 2 is the biggest cluster with following wards total 34 points:

- L Ward
- PN Ward
- S Ward
- ME Ward
- MW Ward

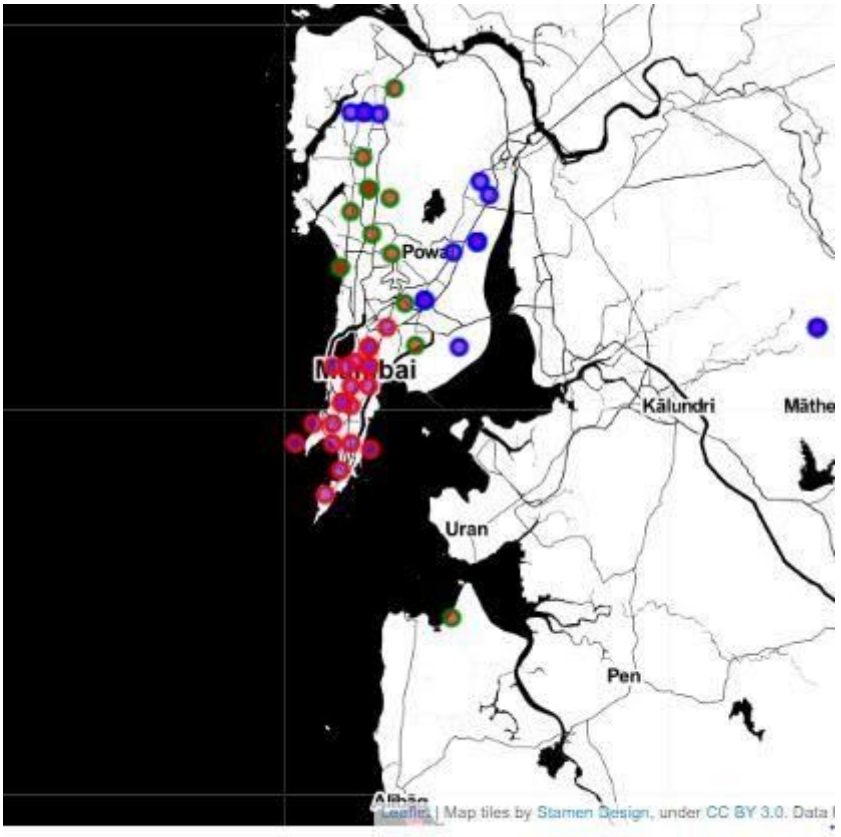
Mumbai K-mean Clusters plotted on Map:

Color Coded:

Red→ Cluster 2

Green→ Cluster 1

Blue → Cluster 0



5. Discussion

Our Analysis show following are the insights:

- Total Number of covid-19 cases in India : 27892
- Total Number of covid-19 cases in mumbai: 5407 src: https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Maharashtra
- Around 20% of totals india's cases are in mumbai with around 17 Hospitals/testing labs(3 Government/14 Non-Government Labs)

Mumbai City is most populated city in india with population of **20,411,274** Source:https://worldpopulationreview.com/world-cities/mumbai-population/ , with very densily populated slums around and as **Mumbai** is epicenter for COVID-19 in india currently with around 20% cases of india's number, it is highly on the risk of getting into phase 3(Community transmission).

Mumbai city has around 1000 diagnostic centre and only 17 are catering for COVID-19 which is very few in number & for such pandemic situation and for epicenter like mumbai testing facilities should be increased for quick assessment, rapid testing and stop the spread of virus.

6.Conclusion

Purpose of this project was to start analysing COVID-19 situation in India & pickup the city which has the most number of corona positive cases, Further to put the analysis & visuzlation sourced from authentic sources to avoid fake information & inform people well.

Mumbai was identified as epicenter city for current corona pandemic with around 20% share alone by this city for total number of case in india, further information about testing center was mapped with cases to know relation between number of cases vs testing labs.

Based upon this analysis it was found that for common man testing facilities for COVID-19 in the city of mumbai should be increases looking at the current number.

Finally I would say!!

STAY HOME #STAY SAFE