# Covid19 Case Study India

Author: Arnab Roy

Date: 25/05/2020

# Introduction

Covid19 cases are on rise in India; under this assignment, we have performed some preliminary analysis for Covid19 cases in India. In this case study we have typically worked on the below problem statement,

**Part 1**

* Identification of the distribution of Covid19 cases across various states in India. Analyzing them based using map visualization.
* State segmentation using clustering based on the below parameters
  + % Confirmed
  + % Active
  + % Recovered
  + % Death
* Represent the clusters in a map for visualization

**Part 2**

* Try to figure out the number of hospitals out in the state using the Foursquare API, and then compare them across states.

# Data acquisition and cleaning

## Data Sources

In order to do this analysis, below data has been identified. Both datasets (CSV) can be downloaded from [**https://api.covid19india.org/csv**](https://api.covid19india.org/csv)

* **State wise cases data:** which contains data for 34 states in India. For each state it has recovered, Deaths, active , confirmed, increase from previous day
* **State wise test count data:** which contains the cummulative covid19 test count for each day for each state, along with all other information e.g. number of beds, population for that state, % positive, % negative e.t.c.
* **Visualization for India state map:** Below json has been used to create Choropleth maps, the json is specific to Indian States.

<https://gist.github.com/ProProgrammer/781d5fbcb1d4364616c5#file-states2-json>

* **Foursquare API Documentation:** All documentation related to foursquare API end point present in the below location

<https://developer.foursquare.com/docs/places-api/endpoints/>

Also since we are only working towards retrieving the list of Hospitals we have used category specific information’s from the below link which is also part of foursquare API documentation.

<https://developer.foursquare.com/docs/build-with-foursquare/categories/>

# Data Cleaning and manipulation

For state wise cases data only below columns has been retained rest all columns, which are not required for the analysis, has been deleted.

* State name
* Confirmed
* Recovered
* Deaths
* Active

State wise test count data, contains date wise cummulative count of test conducted for each state .Hence in order to retrieve the latest state level test count , only the row with last data collection (for each state) date has been retained. As part of the analysis only below information has been retained and rest all has been deleted.

* State name
* Test count

The above 2 dataset has been merged based on state name to create a complete dataset that has following information, using the **geopy** we have also added the latitude and longitude information for each states.

**Covid19 Full Stat**

* State name
* Confirmed
* Recovered
* Deaths
* Active
* Test count
* Latitude
* Longitude

After invoking Foursquare API to retrieve the list of hospitals another dataset has been created with the below attributes

**Covid19 Hospital Info**

* State name
* Longitude
* Latitude
* Hospital Name
* Hospital Latitude
* Hospital Longitude

The above dataset can also be clubbed with the cluster information retrieved while performing the clustering on **Covid19 Full Stat**. This merging will create the below DB which can then further used for visualizing the hospitals on Map.

**Covid19 Hospital Clusters**

* State name
* Longitude
* Latitude
* Hospital Name
* Hospital Latitude
* Hospital Longitude
* Clusters

# Problem solving approach

**Part 1**

Once we have the merged information below, columns can be calculated

* + % Confirmed = Confirmed \* 100 / Total Tested
  + % Active = Active \* 100 / Confirmed
  + % Recovered = Recovered \* 100 / Confirmed
  + % Death = Deaths \* 100 / Confirmed

Based on above information the exploratory analysis can be made and clusters can be formed. Once the clusters are created, we can use Choropleth to represent them in the India Map.

**Part 2**

This is slightly difficult compare to Part 1. In order to solve this we should first use the Foursquare API with Venues/Search and then invoke the API with the Category ID = ‘4bf58dd8d48988d196941735’ to retrieve the list of hospitals present in the state.

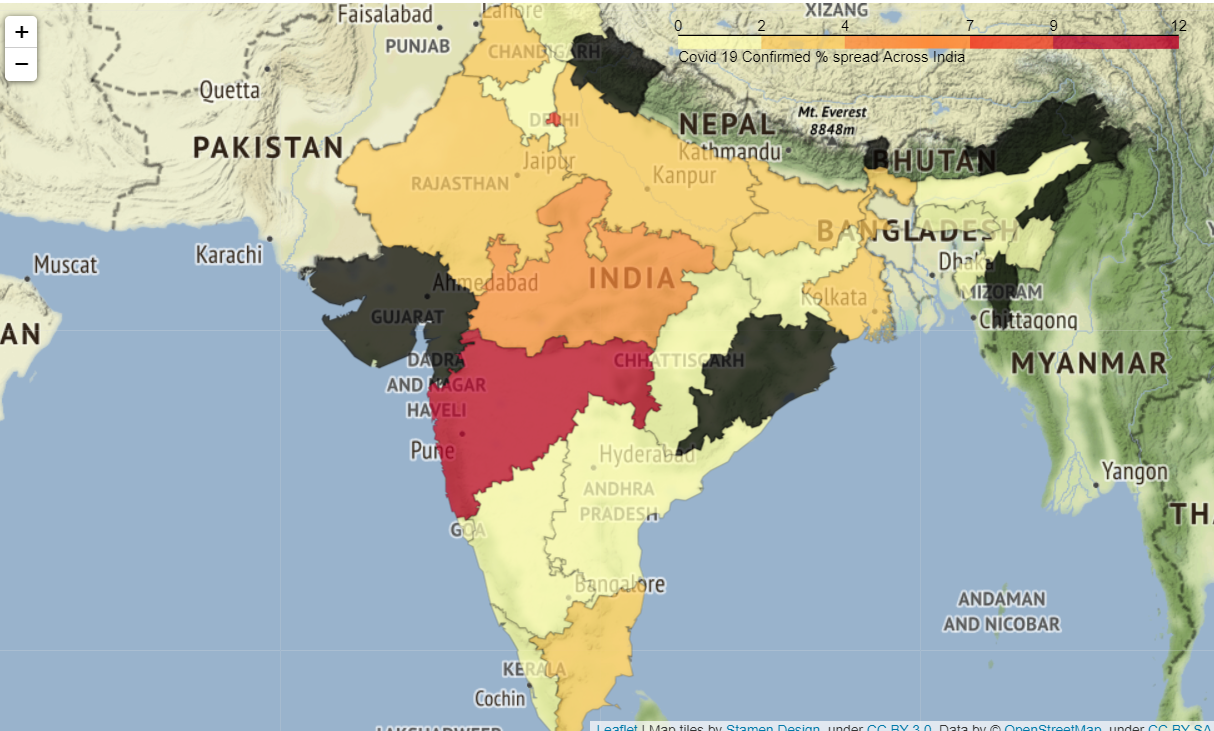
The same information can then grouped by State to get an indicated count on Number of Hospitals accorss the state. A barchart can then be formed to represent the state wise comparison.

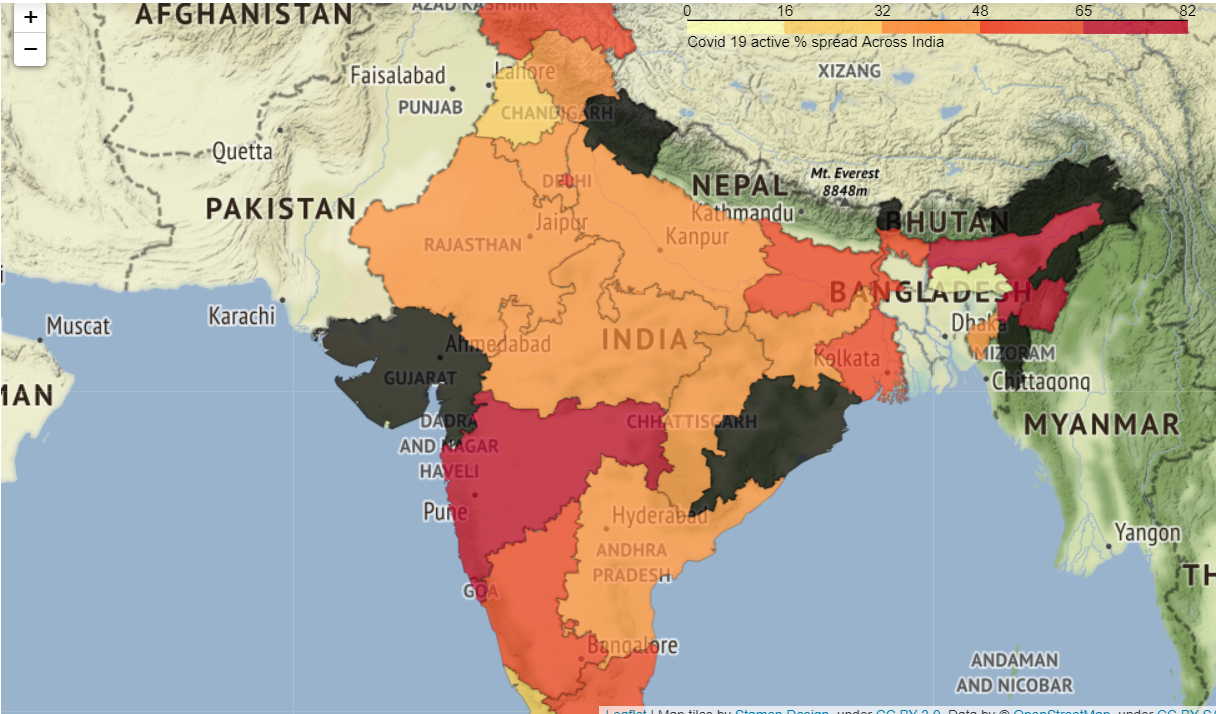
# Full Solution and Analysis

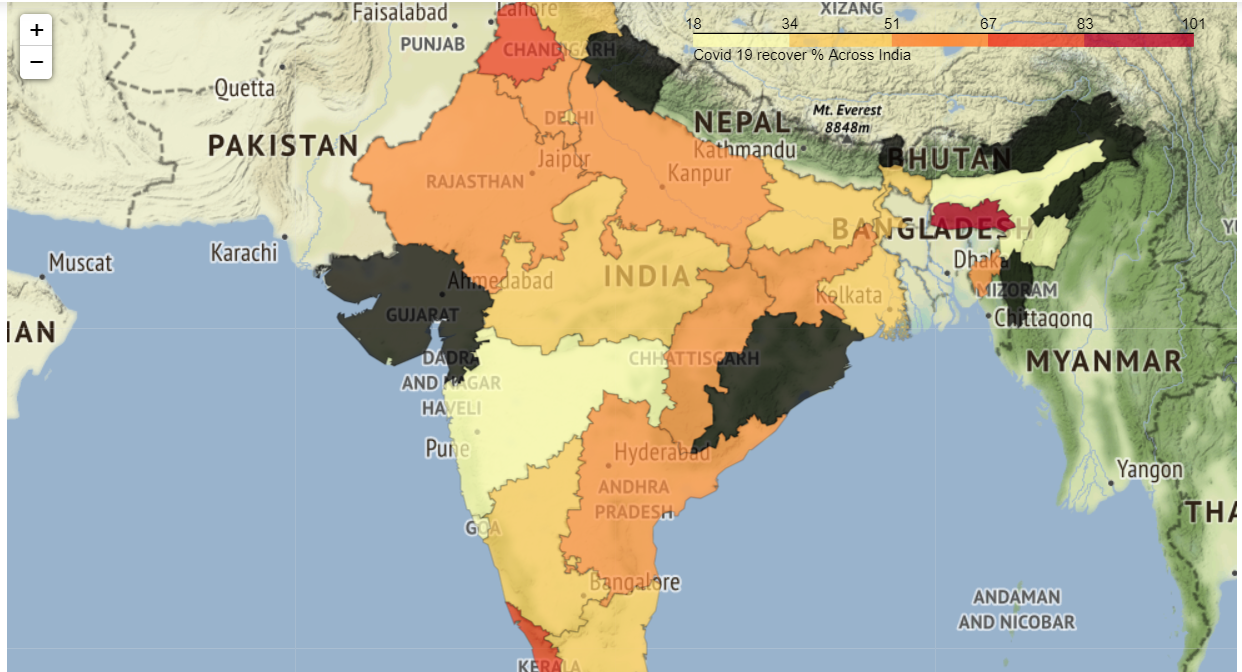
**Part 1**

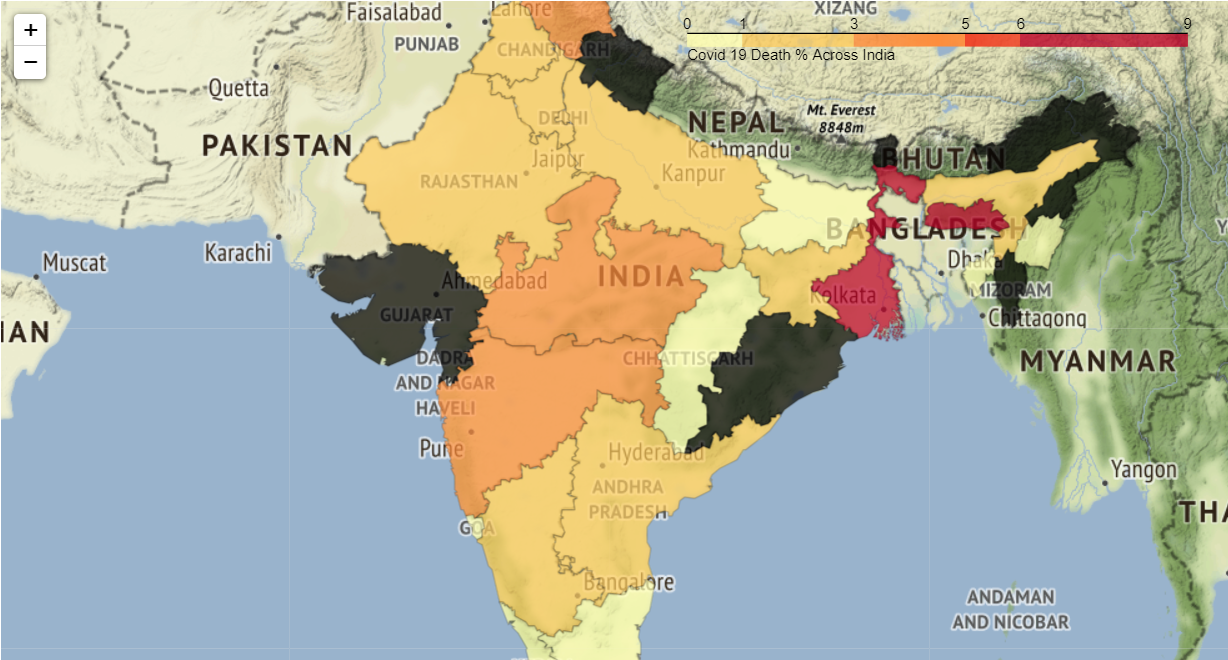
A quick map plotting indicates the below distribution of % confirmed, Active, recovered and death cases in India. Visualization clearly shows that the states like Maharashtra, Gujrat and Rajasthan are worst impacted in terms of %confirmed and %active cases. Whereas most of the states also shows a moderate (30%-50%) and good (55%-70%) recovery rate.

In terms of death WB being the worst with death % 6%-9%, which is higher than most of the other states mostly having death % 1-5% which is almost at per India’s average.





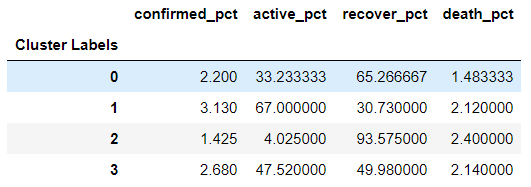




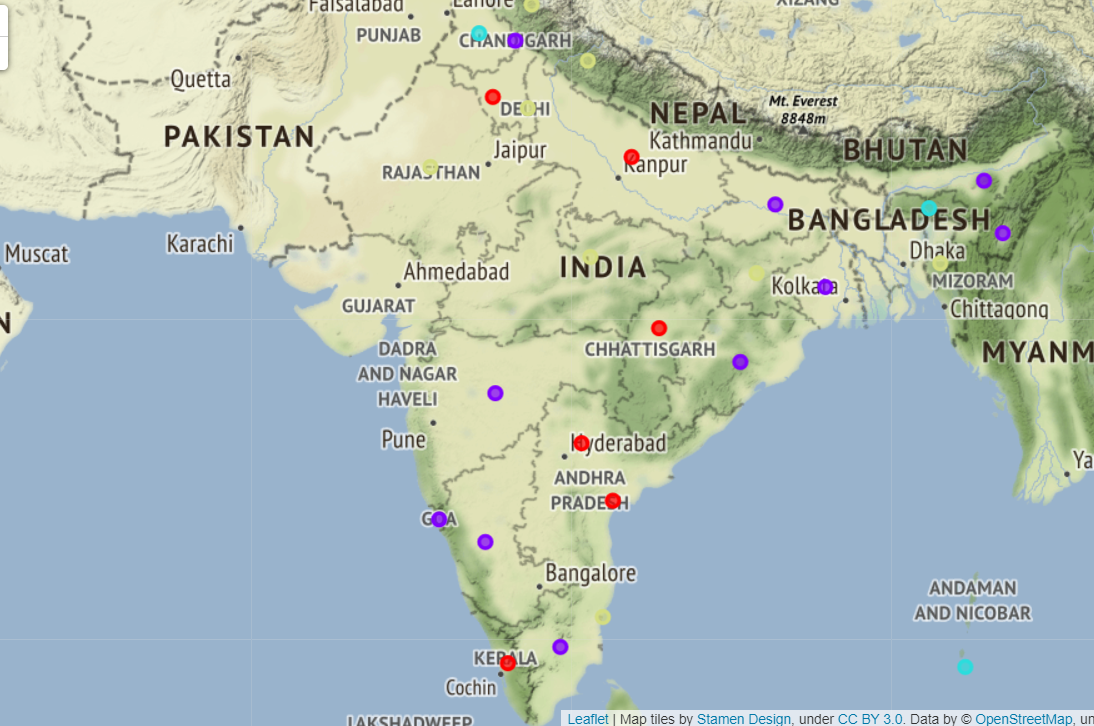
Clustering mean analysis shows the below, on each clusters separately below has been revealed

**Cluster 0,2:** Group of 10 states on total where average Recovery % has been far better which helps reducing the active %.Cluster 2 comprises of the set of small states where the testing % is not very good hence both Confirmed & active % are very low.

**Cluster 1, 3:** Groups 20 states on total where testing count is good which results slightly high-confirmed cases and moderate recovery %.

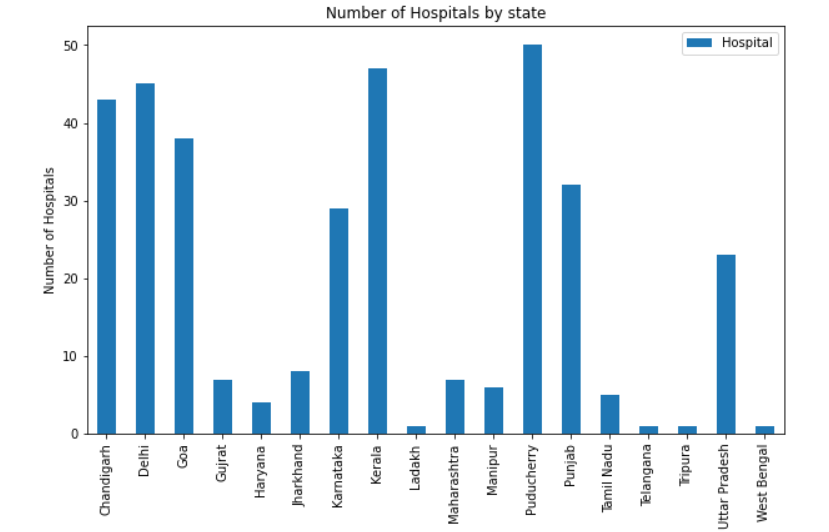


As part of further analysis, have tried to represent the clusters in the map, which shows below.



**Part 2**

Foursquare API with Venues/Search method has been used with the category ID = ‘4bf58dd8d48988d196941735’ (specifically represents the hospitals) to retrieve the list of hospitals along with their latitude and longitude information available within the state.

A quick visualization shows that the number of hospitals are more for southern states (Karnataka/Kerala/Puducherry/Goa) than Northern States.

# Challenges

Foursquare API ‘s shows different number of hospitals with the slight change of latitude and longitude information, e.g. while invoking with the actual lat, long (19.531932, 76.055457) information for Maharashtra it returns 0 number of hospitals but return 7 hospital details while invoking with (19,76). For analysis purpose we have always passed on the actual latitude and longitude information for states, so for this analysis this limitation has been accepted.