# 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 use case,

**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.

## 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.

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

# 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 get/Venues and then try to figure out some way to find categories that represents the hospitals. Once that is found then we can calculate the count on the state level and analyze further.