Outline: Analysing COVID-19 pandemic using Python

**Keyword:** COVID-19

**Author: Shivam Kumar**

**Publish Date:**

Analysing COVID-19 pandemic using Python

## Introduction

The COVID-19 outbreak since late January in India has caused a dystopia and affected most parts of the country. It has been difficult to analyse the SARS-CoV-2 (Coronavirus) because of lack of data and the innate nature of the virus itself. Since the first case of this virus infection back in November, 2019 the world has put in lots of research and efforts to gather neccesary data. This blog attempts to analyse some data and draw some conclusions about the COVID-19 pandemic in India.

The code

The version of Python used is 3.8.2 on an Ubuntu Machine. The code has been split in to a number of segments listed below with the code included.

1. Libraries used

import numpy as np

import seaborn as sns

import pandas as pd

import matplotlib.pyplot as plt

## 2. Importing dataset

x = pd.read\_csv("covid2.csv").head(40)

3. Sorting data

This section seems long and cumborsome but all it does is extract data from the dataset and store it in python lists using for loops.

cases = x.loc[:,"Active Cases"]

populationlist = x.loc[:,"Population"]

cases2 = x.loc[:,"Total Cases"]

literacyrate = x.loc[:,"Literacy rate "]

deaths = x.loc[:,"Deaths"]

gdspc = x.loc[:,"GDP per capita"]

gdspt = x.loc[:,"Total GDP"]

gdphealth = x.loc[:,"Healthcare expenditure"]

deaths2 = x.loc[:,"Death2"]

#Initialising lists

pop = list()

active = list()

total = list()

literacy = list()

death = list()

gdp = list()

gdpt = list()

gdph = list()

death2 = list()

#looping lists to get data from the dataset

for index in cases:

active.append(index)

for index in populationlist:

pop.append(index)

for index in cases2:

total.append(index)

for index in literacyrate:

literacy.append(index)

for index in deaths:

death.append(index)

for index in gdspc:

gdp.append(index)

for index in gdspt:

gdpt.append(index)

for index in gdphealth:

if index > 0.0:

gdph.append(index)

for index in deaths2:

if index > 0.0:

death2.append(index)

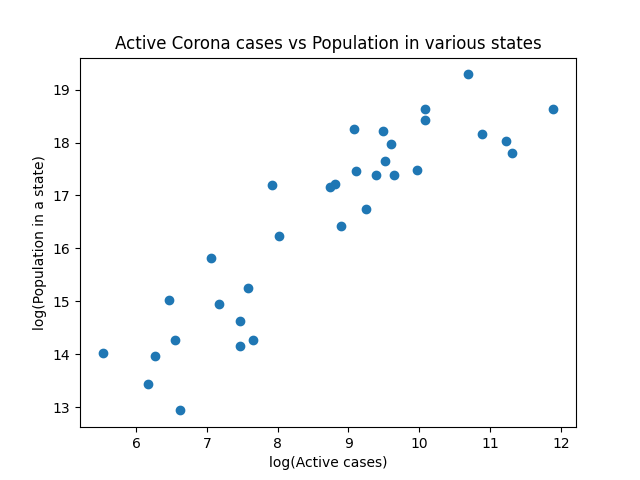
4. Plotting scatter graphs

# I) Active cases vs Population

plt.scatter(np.log(active),np.log(pop))

plt.title("Active Corona cases vs Population in various states")

plt.xlabel("log(Active cases)")

plt.ylabel("log(Population in a state)")

plt.show()

# II) Total cases vs Population

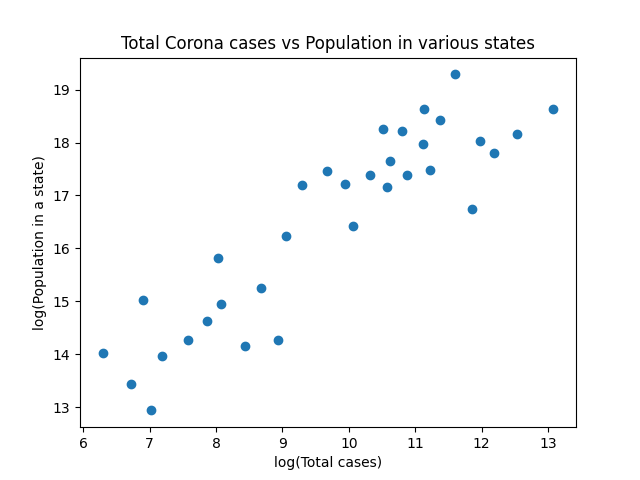
plt.scatter(np.log(total),np.log(pop))

plt.title("Total Corona cases vs Population in various states")

plt.xlabel("log(Total cases)")

plt.ylabel("log(Population in a state)")

plt.show()



# III) Active cases vs Death

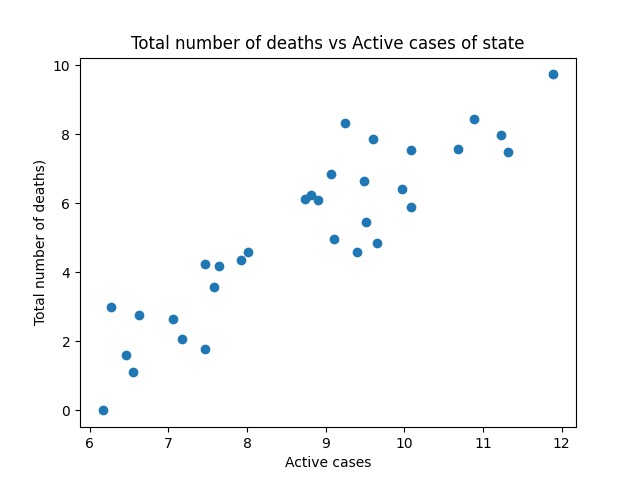
plt.scatter(np.log(active),np.log(death))

plt.title("Total number of deaths vs Active cases of state")

plt.xlabel("Active cases")

plt.ylabel("Total number of deaths)")

plt.show()



# IV) Total cases vs Death

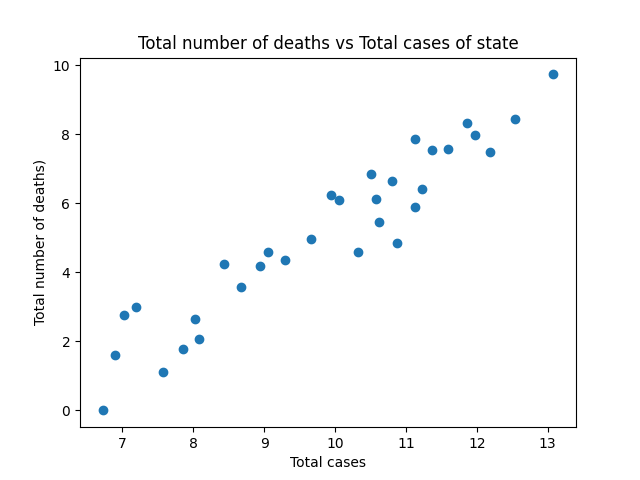
plt.scatter(np.log(total),np.log(death))

plt.title("Total number of deaths vs Total cases of state")

plt.xlabel("Total cases")

plt.ylabel("Total number of deaths)")

plt.show()



# V) Total cases vs Literacy rate

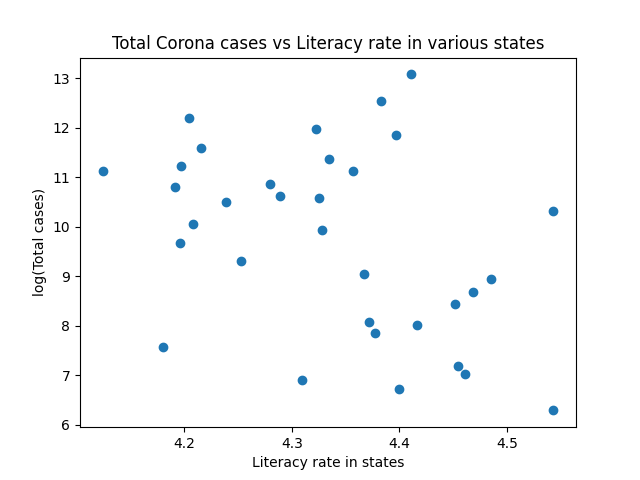
plt.scatter(np.log(literacy),np.log(total))

plt.title("Total Corona cases vs Literacy rate in various states")

plt.xlabel("Literacy rate in states")

plt.ylabel("log(Total cases)")

plt.show()



# VI) Active cases vs literacy

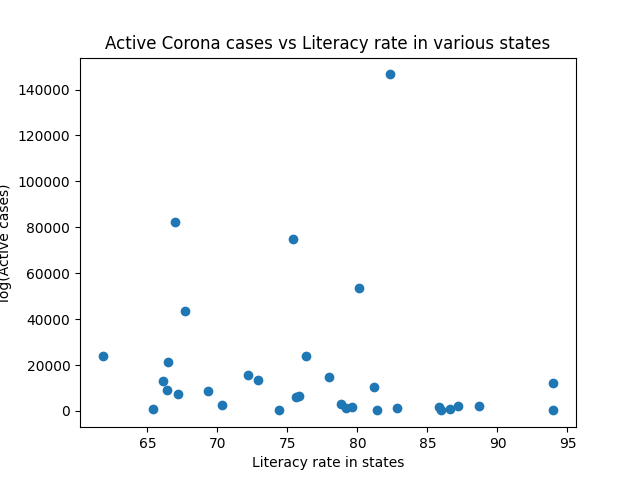
plt.scatter(literacy,active)

plt.title("Active Corona cases vs Literacy rate in various states")

plt.xlabel("Literacy rate in states")

plt.ylabel("log(Active cases)")

plt.show()



# VII) Total deaths vs GSDP

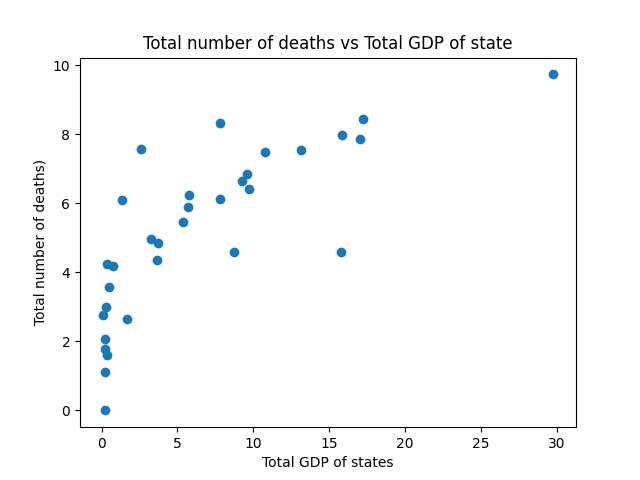
plt.scatter(gdpt,np.log(death) )

plt.title("Total number of deaths vs Total GDP of state")

plt.xlabel("Total GDP of states")

plt.ylabel("Total number of deaths)")

plt.show()



# VIII) Healthcare expenditure vs Death

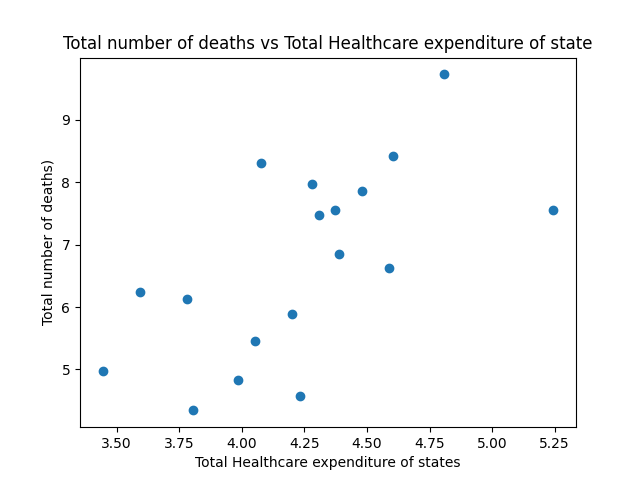
plt.scatter(np.log(gdph),np.log(death2))

plt.title("Total number of deaths vs Total Healthcare expenditure of state")

plt.xlabel("Total Healthcare expenditure of states")

plt.ylabel("Total number of deaths)")

plt.show()



Conclusions

* The scatter graphs of active and total number of infection cases vs Population (Graphs I and II) of states clearly indicate that both the active and total number of cases increase if the states are massively populated.
* Similarly, the number of deaths occuring due to the SARS-CoV-2 virus infection increases with the increase in both active and total number of cases as shown by graphs III and IV. This result is both expected and obvious.
* Graphs V and VI do not show any definite pattern which suggests that Literacy rate doesn’t play any role in the number of active or total number of infection cases.
* Graph VII is an unexpected result and probably a coincidence. It suggests that the number of deaths occuring due to the COVID-19 pandemic increase expinentially for states with higher GDP.
* It is intuitive to think that states which spend more on healthcare should have lesser number of COVID-19 casualties but graph VIII suggests oherwise and no pattern is oberserved,

Data and References

The dataset “covid2.csv” was compiled through various resources and references as listed below :

1. COVID-19 cases, deaths and cured/discharged data as of August 7th, 2020 “<https://www.mygov.in/corona-data/covid19-statewise-status/>”
2. Population of Indian states as reported in the 2011 census http://statisticstimes.com/demographics/population-of-indian-states.php
3. Literacy rate of Indian states https://www.jagranjosh.com/general-knowledge/literacy-rate-among-the-indian-states-census-2011-1476094746-1
4. GDP and GDP per capita of Indian states http://statisticstimes.com/economy/gdp-capita-of-indian-states.php
5. Healthcare expenditure of major states in India https://www.statista.com/statistics/685200/india-highest-public-health-expenditure-by-state/