

PROJECT REPORT

ON

Analysis of Covid-19 Data

Submitted By

P Charith

Ajit



Submitted in partial fulfillment of the requirements for internship

International Institute of SDG's & Public Policy Research

1st floor Azad Market, Near AMU, Aligarh, India

Abstract

The novel coronavirus (COVID-19) that was first reported at the end of 2019 has impacted almost every aspect of life as we know it. This paper focuses on the incidence of the disease in Italy and Spain—two of the first and most affected European countries. Using two simple mathematical epidemiological models—the Susceptible-Infectious-Recovered model and the log-linear regression model, we model the daily and cumulative incidence of COVID-19 in the two countries during the early stage of the outbreak, and compute estimates for basic measures of the infectiousness of the disease including the basic reproduction number, growth rate, and doubling time. Estimates of the basic reproduction number were found to be larger than 1 in both countries, with values being between 2 and 3 for Italy, and 2.5 and 4 for Spain. Estimates were also computed for the more dynamic effective reproduction number, which showed that since the first cases were confirmed in the respective countries the severity has generally been decreasing. The predictive ability of the log-linear regression model was found to give a better fit and simple estimates of the daily incidence for both countries were computed.

Objective

The main objective of this project is to gather useful insights from Covid-19 Data. Covid-19 has had and is still having a huge negative impact on every individual. It is important to keep a track of the number of Active cases so as understand how efficient Covid-19 vaccine .The following are the steps to achieve the objective:

1. Analysis of the effect of Covid-19 on each state in India.
2. Analysis of the effect of Covid-19 on Age Groups
3. Analysis on Tests conducted in India and in Major Affected States

Introduction

Coronaviridae is a family of viruses with a positive-sense RNA that possess an outer viral coat. When looked at with the help of an electron microscope, there appears to be a unique corona around it. This family of viruses mainly cause respiratory diseases in humans, in the forms of common cold or pneumonia as well as respiratory infections. These viruses can infect animals as well (1, 2). Up until the year 2003, coronavirus (CoV) had attracted limited interest from researchers. However, after the SARS (severe acute respiratory syndrome) outbreak caused by the SARS-CoV, the coronavirus was looked at with renewed interest (3, 4). This also happened to be the first epidemic of the 21st century originating in the Guangdong province of China. Almost 10 years later, there was a MERS (Middle East respiratory syndrome) outbreak in 2012, which was caused by the MERS-CoV (5, 6). Both SARS and MERS have a zoonotic origin and originated from bats. A unique feature of these viruses is the ability to mutate rapidly and adapt to a new host.

The zoonotic origin of these viruses allows them to jump from host to host. Coronaviruses are known to use the angiotensin-converting enzyme-2 (ACE-2) receptor or the dipeptidyl peptidase IV (DPP-4) protein to gain entry into cells for replication (7–10).

In December 2019, almost seven years after the MERS 2012 outbreak, a novel Coronavirus (2019-nCoV) surfaced in Wuhan in the Hubei region of China. The outbreak rapidly grew and spread to neighboring countries. However, rapid communication of information and the increasing scale of events led to quick quarantine and screening of travelers, thus containing the spread of the infection. The major part of the infection was restricted to China, and a second cluster was found on a cruise ship called the Diamond Princess docked in Japan.

Origin

The new virus was identified to be a novel Coronavirus and was thus initially named 2019-nCoV; later, it was renamed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and the disease it causes is now referred to as Coronavirus Disease-2019 (COVID-19) by the WHO. The virus was suspected to have begun its spread in the Huanan seafood wholesale market in the Wuhan region. It is possible that an animal that was carrying the virus was brought into or sold in the market, causing the spread of the virus in the crowded marketplace. One of the first claims made was in an article published in the *Journal of Medical Virology* (14), which identified snakes as the possible host. A second possibility was that pangolins could be the wild host of SARS-CoV-2, though the most likely possibility is that the virus originated from bats. Increasing evidence and experts are now collectively concluding the virus had a natural origin in bats, as with previous such respiratory viruses.

Similarly, SARS and MERS were also suspected to originate from bats. In the case of MERS, the dromedary camel is an intermediate host (5, 10). Bats have been known to harbor coronaviruses for quite some time now. Just as in the case of avian flu, SARS, MERS, and possibly even HIV, with increasing selection and ecological pressure due to human activities, the virus made the jump from animal to man. Humans have been encroaching increasingly into forests, and this is true over much of China, as in Africa. Combined with additional ecological pressure due to climate change, such zoonotic spillovers are now more common than ever. It is likely that the next disease X will also have such an origin (25). We have learned the importance of identification of the source organism due to the Ebola virus pandemic. Viruses are unstable organisms genetically, constantly mutating by genetic shift or drift. It is not possible to predict when a cross-species jump may occur and when a seemingly harmless variant form of the virus may turn into a deadly strain. Such an incident occurred in Reston, USA, with the Reston virus (26), an alarming reminder of this possibility. The identification of the original host helps us to contain future spreads as well as to learn about the mechanism of transmission of viruses. Until the virus is isolated from a wild animal host, in this case, mostly bats, the zoonotic origin will remain hypothetical, though likely. It should further be noted that the virus has acquired several mutations, as noted by a group in China, indicating that there are more than two strains of the virus, which may have had an impact on

its pathogenicity. However, this claim remains unproven, and many experts have argued otherwise; data proving this are not yet available. A similar finding was reported from Italy and India independently, where they found two strains. These findings need to be further cross-verified by similar analyses globally. If true, this finding could effectively explain why some nations are more affected than others.

Covid-19 Pandemic in India

The COVID-19 pandemic in India is a part of the worldwide pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). As of 27 September 2021, according to official figures, India has the second-highest number of confirmed cases in the world (after the United States of America) with 33,678,786 reported cases of COVID-19 infection and the third-highest number of COVID-19 deaths (after the United States and Brazil) at 470,115[4] deaths. However these figures exhibit severe under-reporting .The spread of COVID-19 in India was initially characterized by lower case numbers and fewer deaths compared with numbers in many developed countries. This was mainly due to a stringent lockdown and demographic factors. However, economic constraints forced a staggered lockdown exit strategy, resulting in a spike in COVID-19 cases in June 2020. Subsequently, India became the third-worst affected country worldwide. Low spending on health as a percentage of gross domestic product (GDP) meant there was a shortage of hospital beds and ventilators and a lack of medical personnel, especially in the public health sector. Nevertheless, technological advances, supported by a strong research base, helped contain the health and economic damage resulting from the pandemic. In the future, measures such as asymptomatic testing, public–private partnerships, and technological advances will be essential until a vaccine against COVID-19 can be developed and rolled-out in India.

The first cases of COVID-19 in India were reported on 30 January 2020 in three towns of Kerala, among three Indian medical students who had returned from Wuhan, the epicenter of the pandemic.[11][12][13] Lockdowns were announced in Kerala on 23 March, and in the rest of the country on 25 March. On 10 June, India's recoveries exceeded active cases for the first time.[14] Infection rates started to drop in September, along with the number of new and active cases.[15] Daily cases peaked mid-September with over 90,000 cases reported per-day, dropping to below 15,000 in January 2021. A second wave beginning in March 2021 was much more devastating than the first, with shortages of vaccines, hospital beds, oxygen cylinders and other medical supplies in parts of the country. By late April, India led the world in new and active cases. On 30 April 2021, it became the first country to report over 400,000 new cases in a 24-hour period. Experts stated that the virus may reach an endemic stage in India rather than completely disappear.

Dataset Considered

The Dataset used here is from 2 sources:

1. https://www.kaggle.com/imdevskp/covid19-corona-virus-india-dataset?select=patients_data.csv
2. <https://github.com/charithcherry/covid-19-data-analysis>

Dataset Description :

1.

- **complete.csv** - Day to day state wise no. of cases (From MoHFH website)
- **patients_data.csv** - Raw patient level data
- **nationleveldaily.csv** - Day by day nation level numbers
- **statelevellatest.csv** - State level latest numbers
- **districtlevellatest.csv** - District level latest numbers
- **tests_daily.csv** - Day by day no. of tests
- **testslateststate_level** - Latest state level tests

2. **Data is dated till 20th July 2020**

- AgeGroupDetails.csv – Details based on Age Groups
- HospitalBedsIndia.csv – Data regarding availability of beds in Hospitals of India
- IndividualDetails.csv – Individual patient data
- StatewiseTestingDetails.csv - State level testing numbers
- covid_19_india.csv – Complete covid-19
- population_india_census2011.csv – India's population census in the year 2011

Package

Here we use:

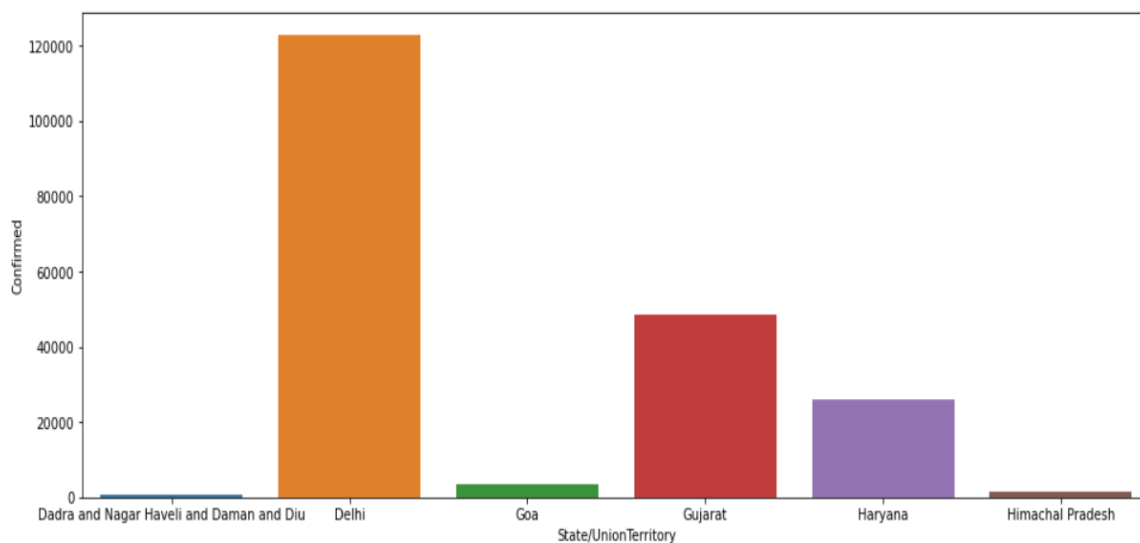
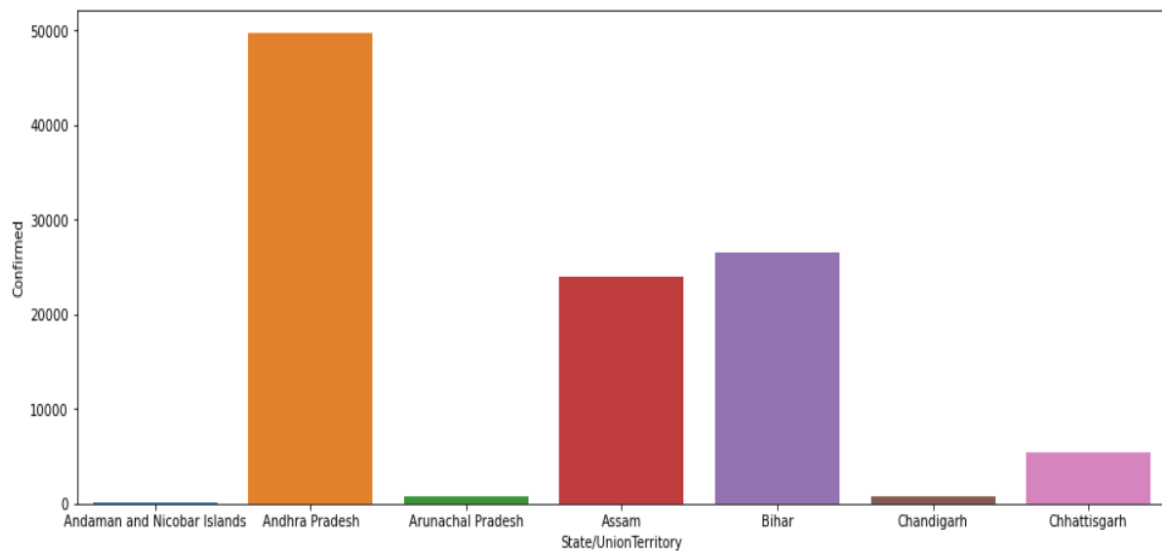
- **python**
Install python from here <https://www.python.org/downloads/>
- **Jupyter Notebook**
Install using command on command prompt: pip install notebook . To Run command on command prompt: jupyter notebook .
- Other Necessary packages are: **Numpy, Pandas, Seaborn and Matplotlib.**

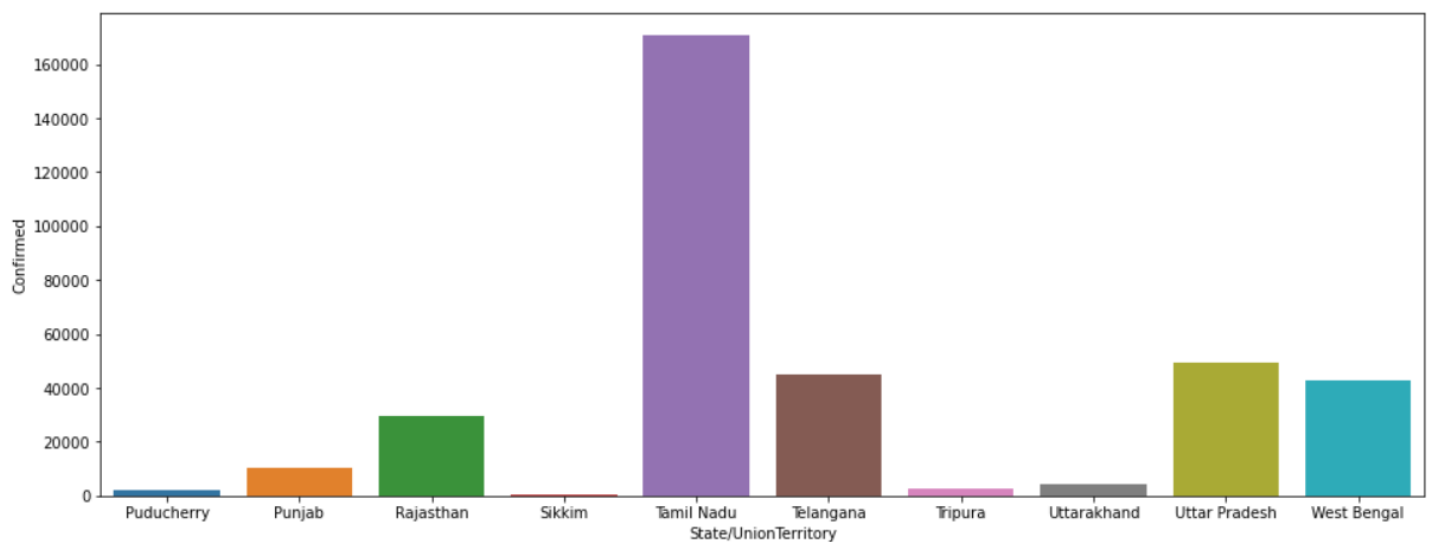
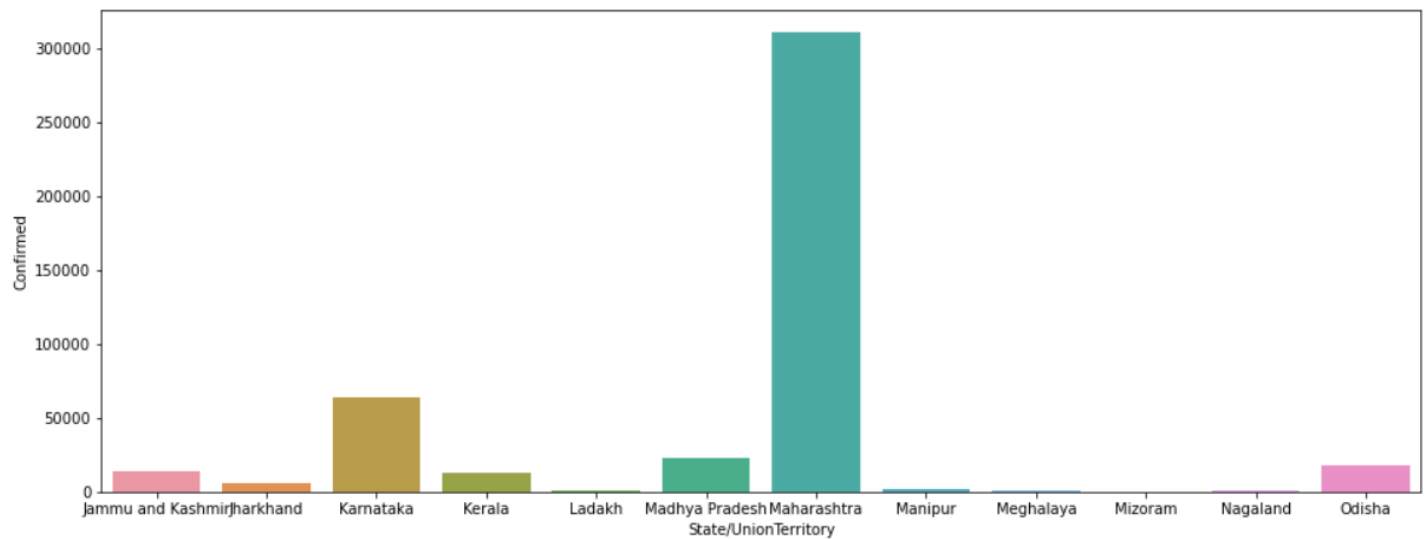
Data Analysis and Visualizations

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions. Data visualization is the representation of information in the form of a chart, diagram, picture, etc.

Lets begin

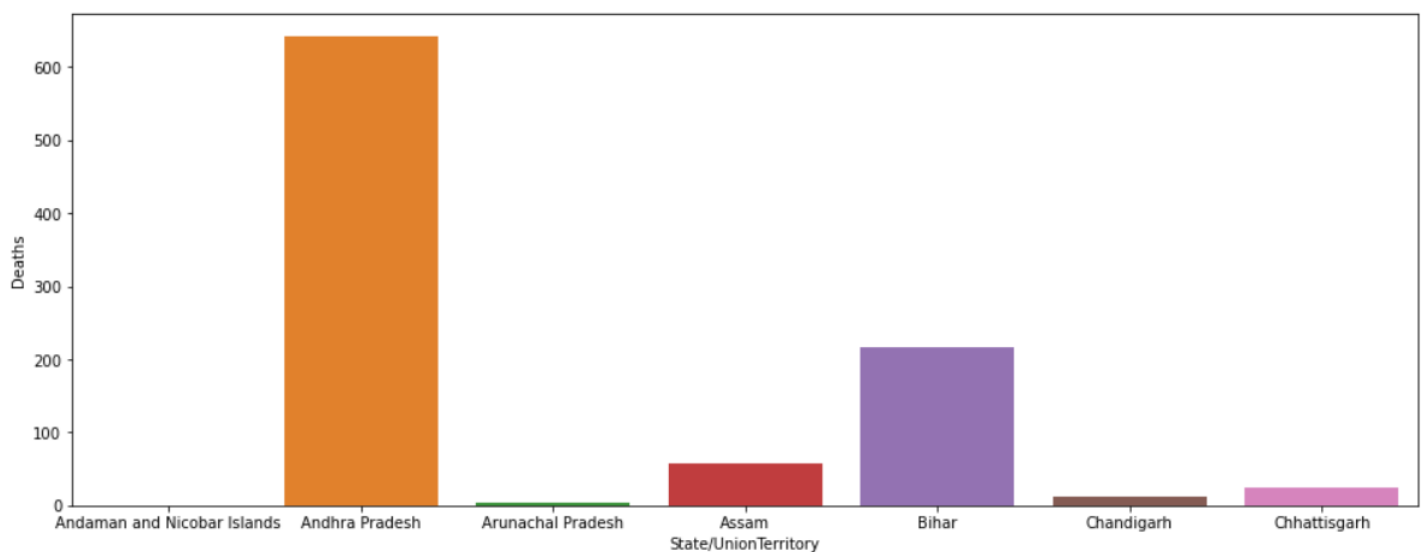
1) State wise plots for number of confirmed cases:

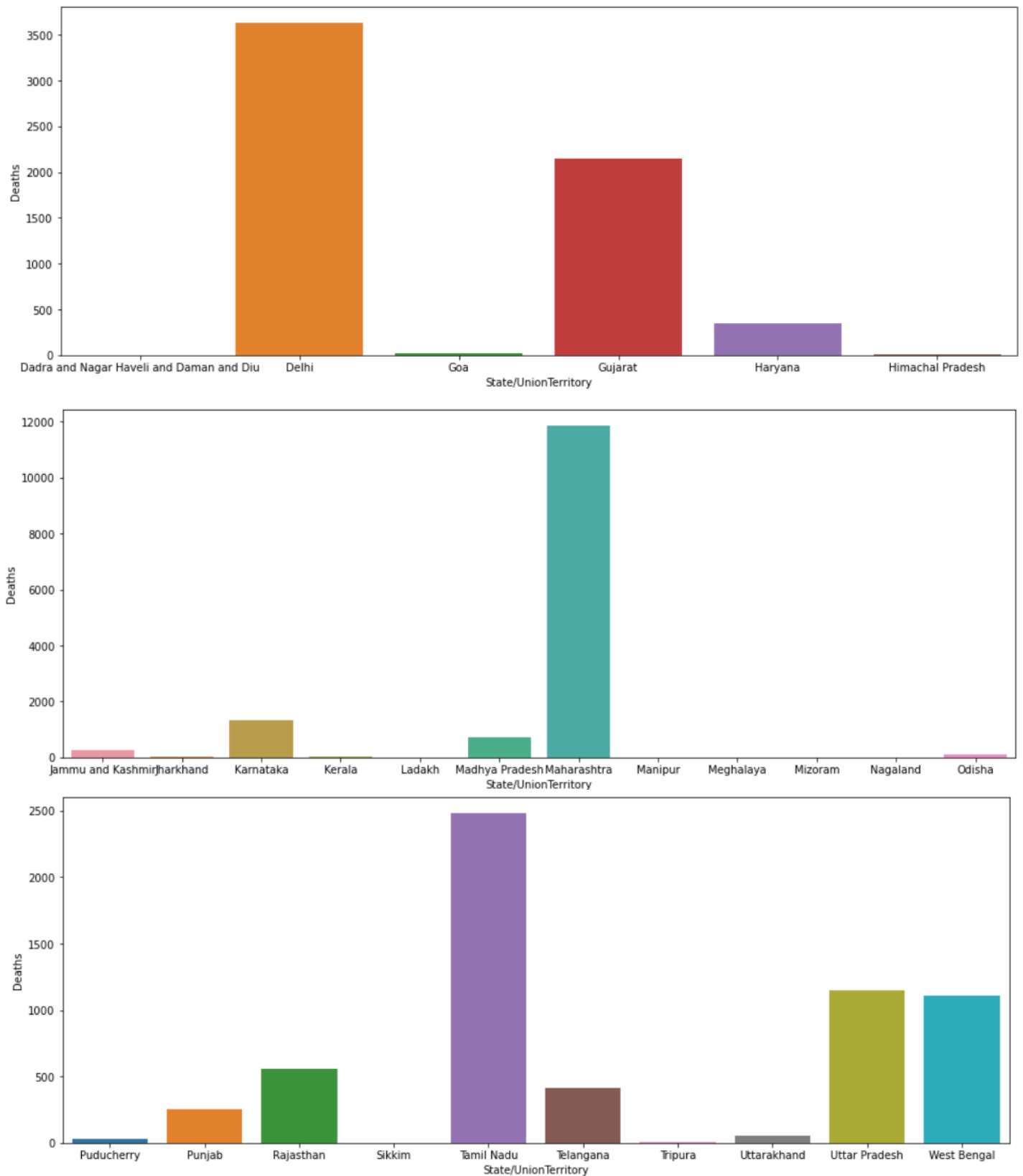




From the above we can conclude that Maharashtra is the state with maximum number of active cases.

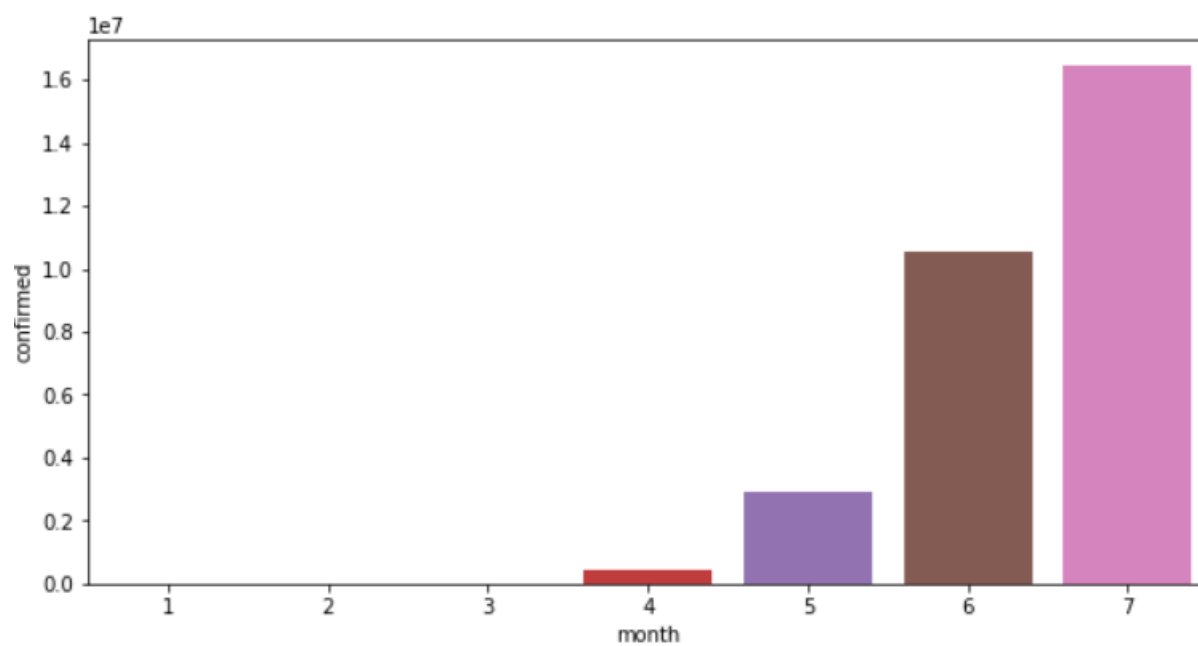
2) State wise plots for number of death cases:



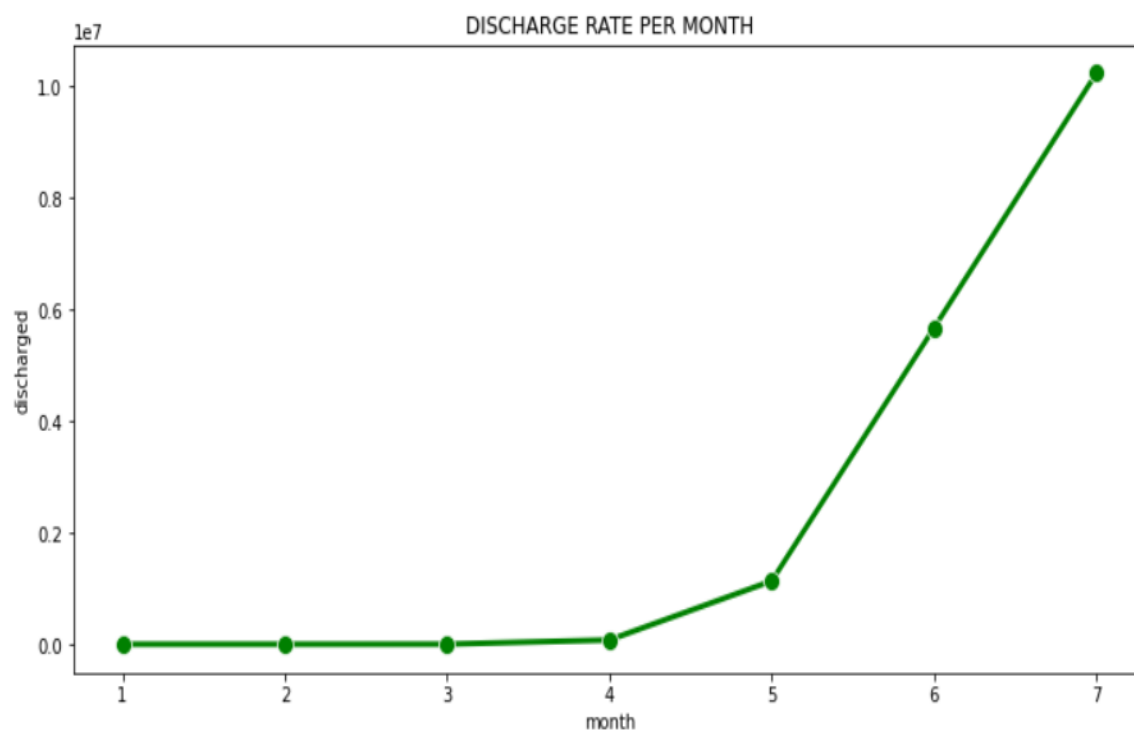


From the above we can conclude that Maharashtra is the state with maximum number of deaths.

3) Confirmed cases Month wise



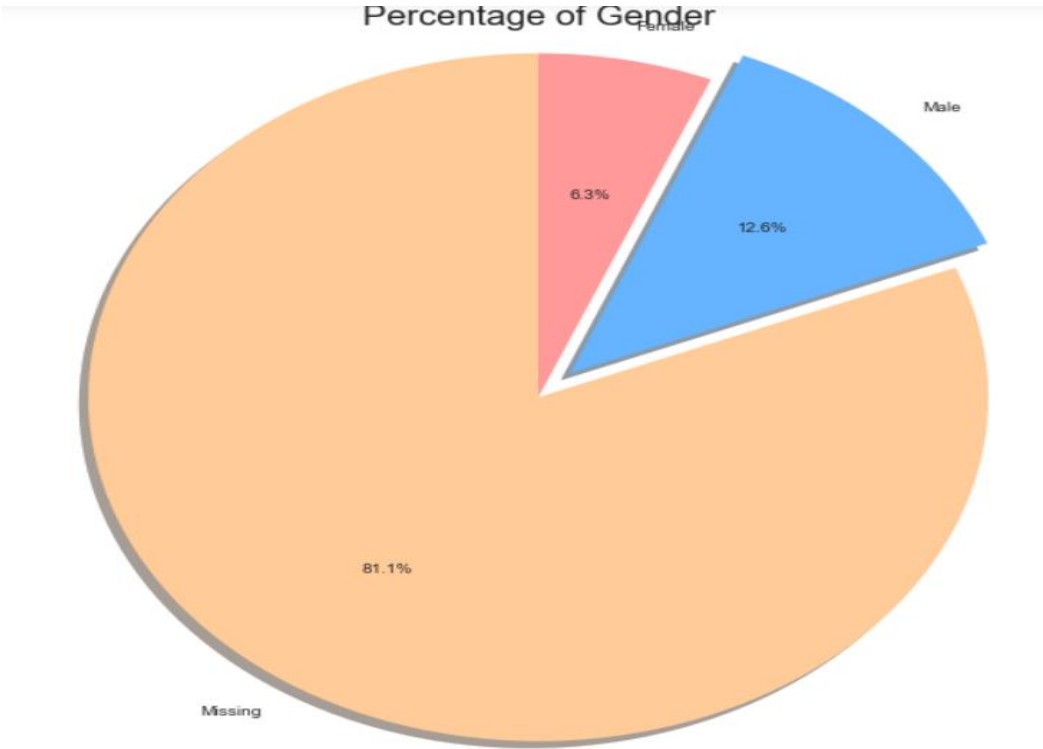
4) Discharge rate per month



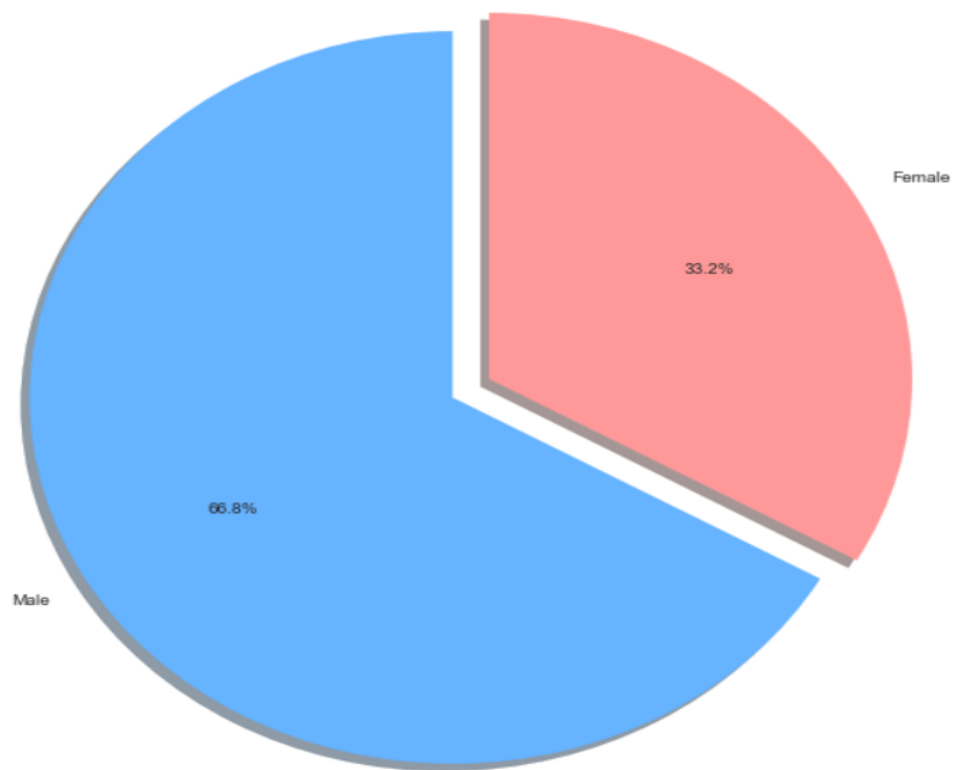
5) Death rate per month



6) Covid-19 affect crowd based on gender

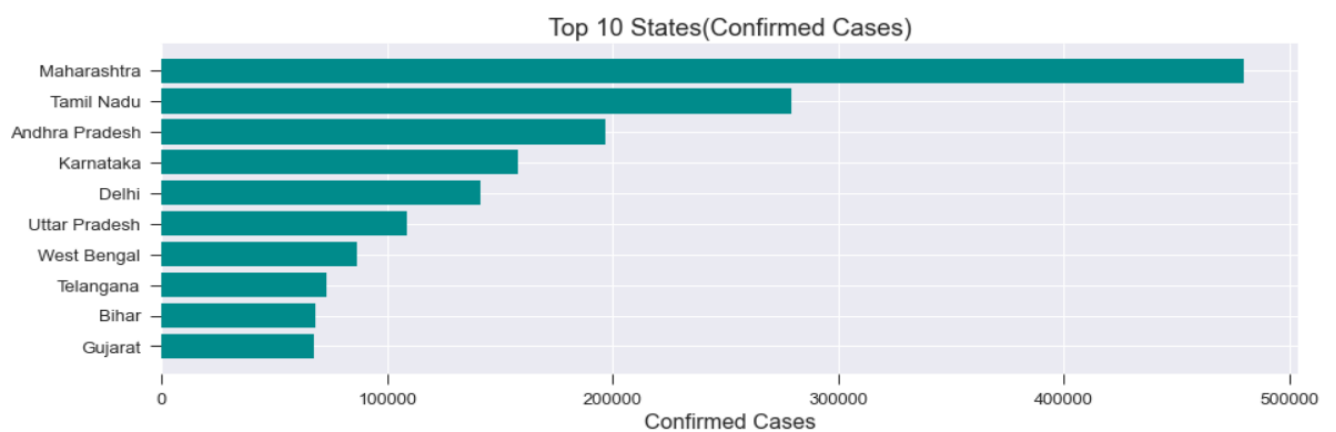


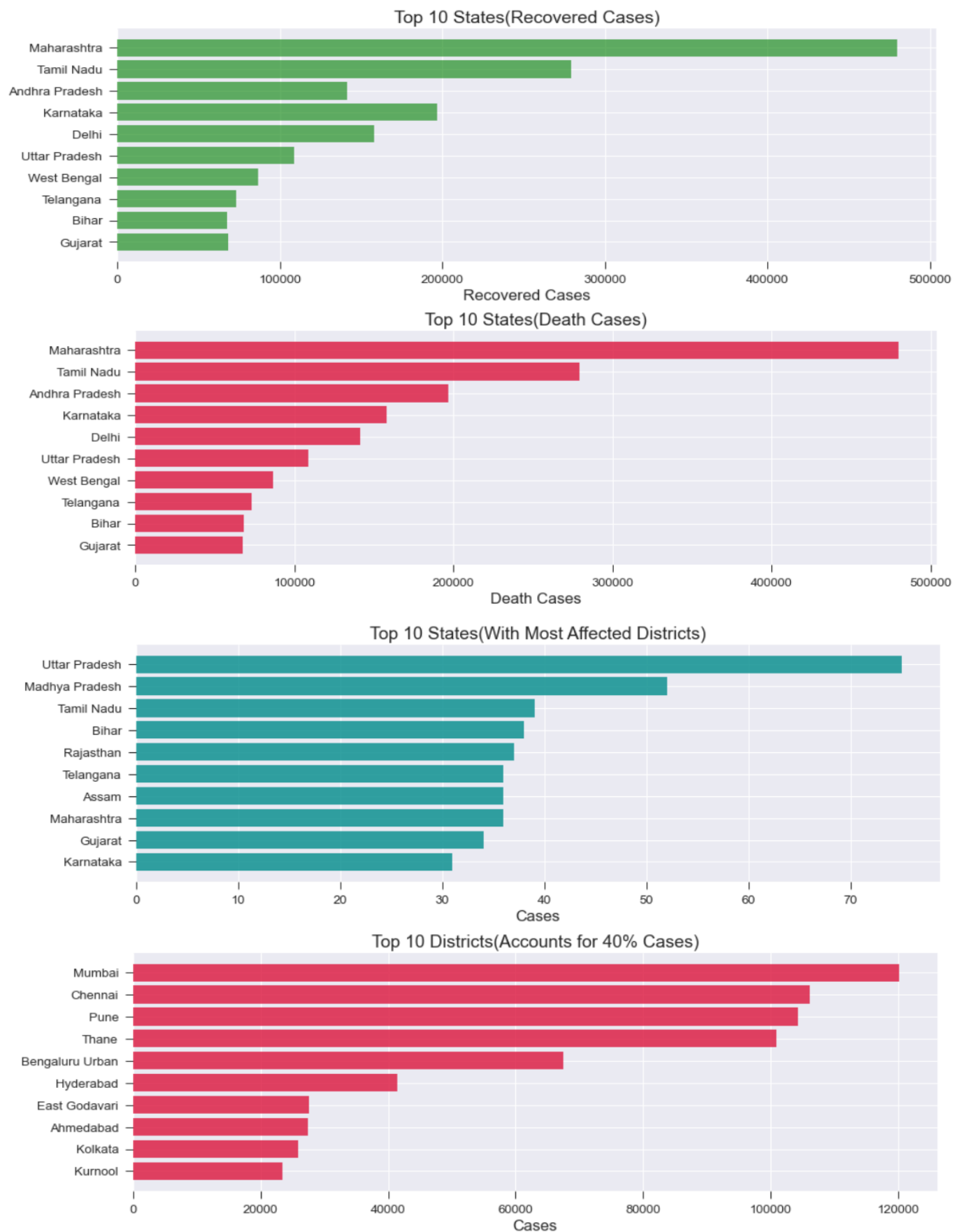
Percentage of Gender (Ignoring the Missing Values)



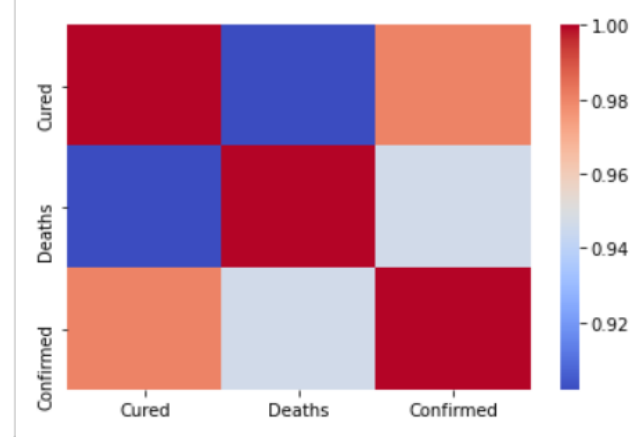
From the above Pie chart we can conclude that the male gender was the most affected one.

7) Top 10 States plot

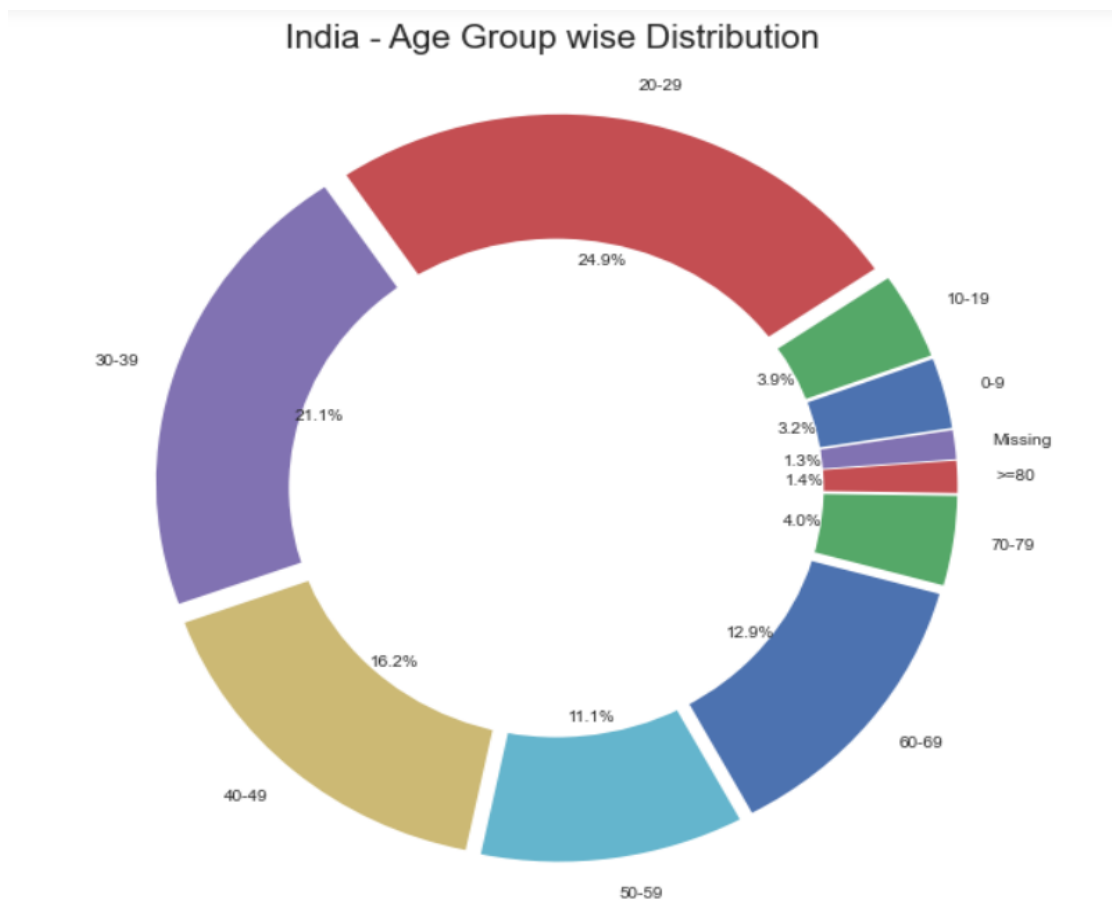




8) HeatMap correlation between confirmed, Deaths and cured



9) Age Group wise Distribution



From the above distribution we can conclude that age 20-29 were affected the most followed by age group 30-39.

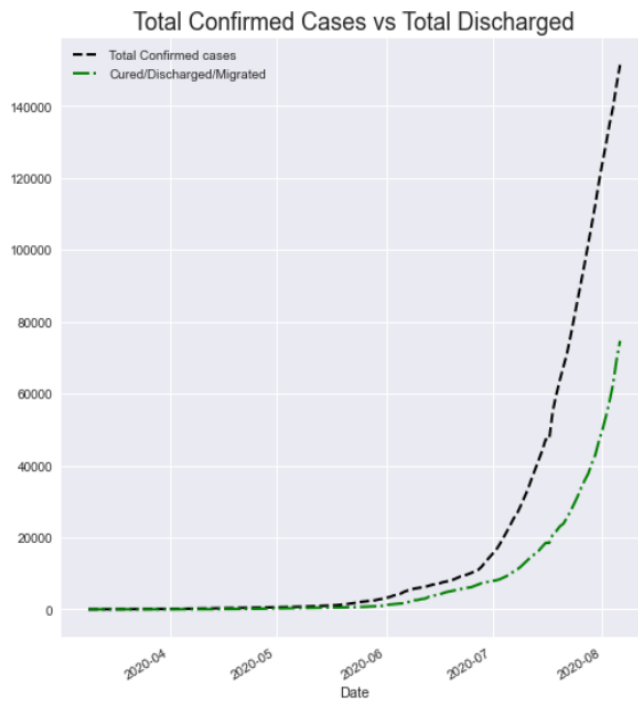
10) Active vs Recovered vs Deaths for states

[‘karnataka’,’Tamil Nadu’,’Andhra Pradesh’,’Telegana’]

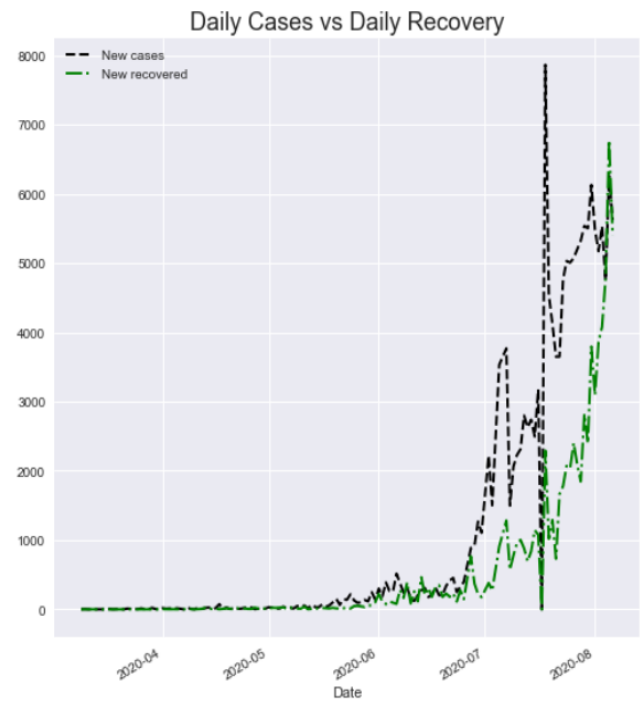


11) Daily Cases vs Daily Recovery and Total Confirmed Cases vs Total Discharged
state=['Karnataka']

Analysis of Indian Poverty



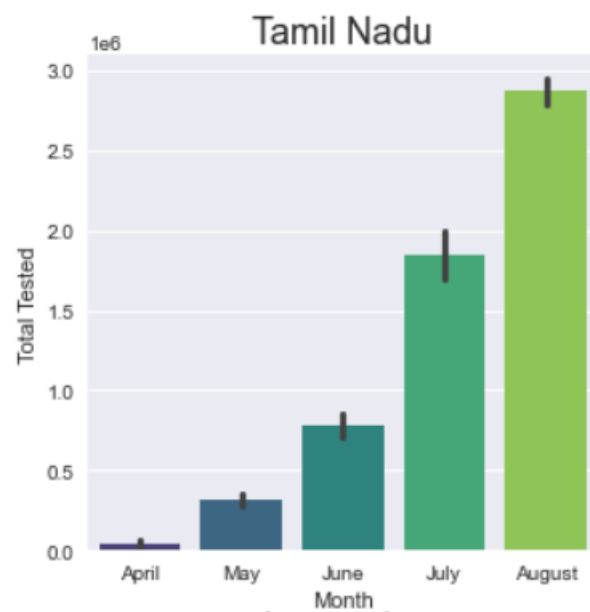
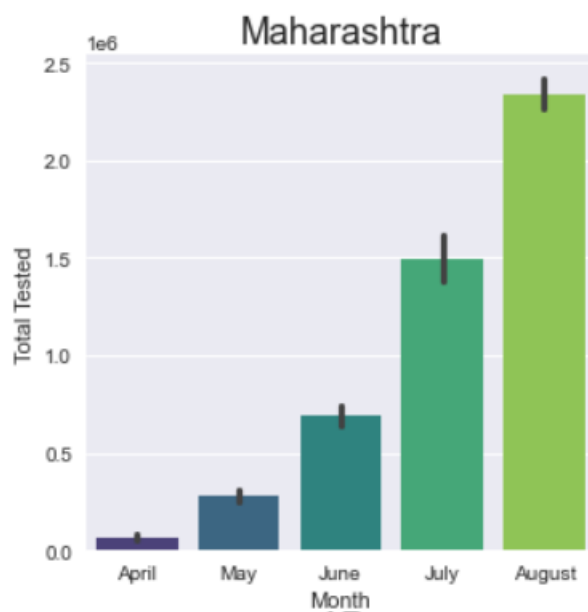
IISPPR

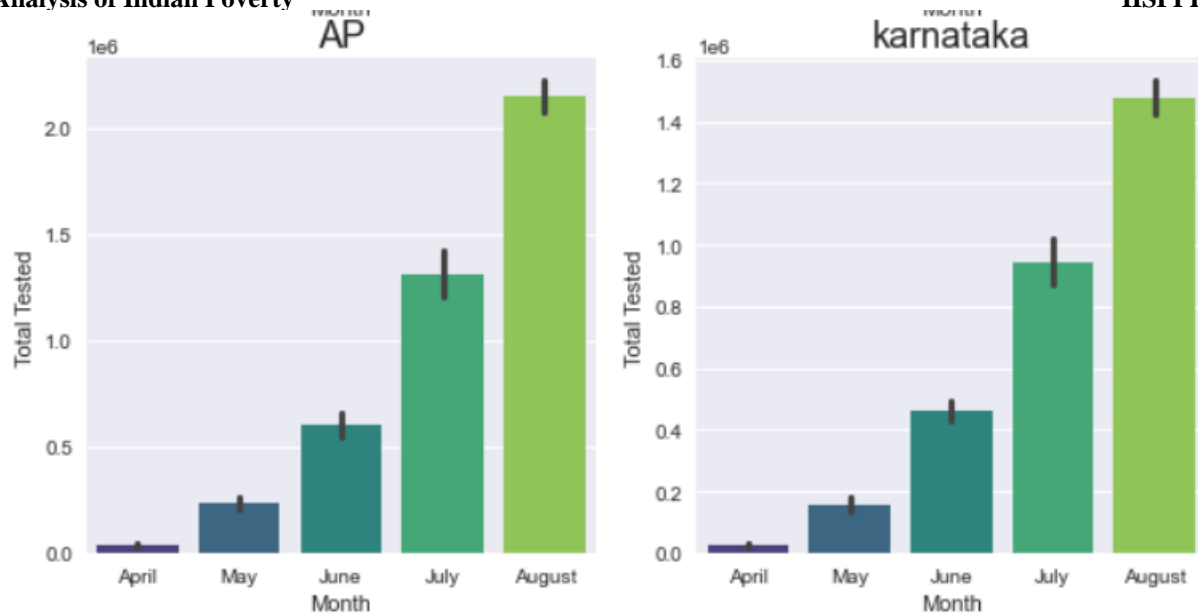


12) Tests Conducted by the Major Affected States

[‘Maharashtra’, ‘Tamil Nadu’, ‘Andhra Pradesh’, ‘Karnataka’]

Total Tests Conducted





13) Trends : Number of cases vs date

