Project Documentation

COVID-19

**DATA ANALYSIS**

**By:** Rikshita Dasgupta

B.Sc Data Science(2nd Year)

**Netaji Subhash Engineering College, Kolkata**

PROJECT OBJECTIVE

During the current coronavirus pandemic, monitoring the evolution of COVID-19 cases is of utmost importance for the authorities to make informed policy decisions (e.g., lock-downs), and to raise awareness in the general public for taking appropriate public health measures. This is done by understanding the analysis of outbreak of COVID -19 with death rates in different areas.

**Keywords**

Virus, WHO, COVID-19, Environmental pollution, Global Pandemic, SARS-CoV-2

**Abstract**

The aim of the project is to provide data analysis of covid-19 (a pandemic started in December 2019). Through plotting of data, various cases have been studied like most affected countries due to this pandemic. Study of data from various countries is combined to show the growth of cases and recovery graph. Comparison graphs has also been plotted to analyse how much INDIA is getting affected/recover day by day.

**Introduction to Covid-19**

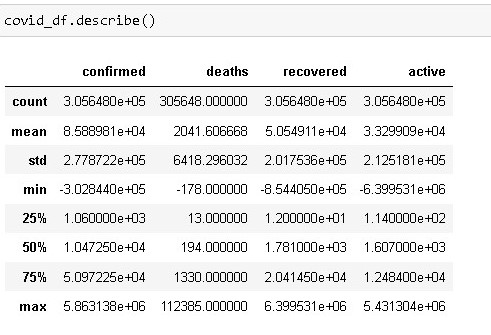
On 31st December 2019, in the city of Wuhan (CHINA), a cluster of cases of pneumonia of unknown cause was reported to WHO. In January 2020, a previously unknown new virus was identified, subsequently named 2019 novel corona virus. WHO declared the COVID-19 as a pandemic.

**Problem Statement**

The pandemic has already taken grip over peoples’ life. Since the start of the pandemic, some countries are facing problem of ever-increasing cases. Through the data analysis of cases one can analyse how countries all over the world are doing in terms of controlling the pandemic. Analysing data leads to adapt the prevention model of the countries that are doing great in terms of lowering the graph. This project gives an insight of how a country is doing in terms of limiting the spread.

**CODE**

1. **Importing the libraries**
2. **Data Pre-processing :**
3. This section includes loading of dataset into a proper readable format and the parsing of dates to datetime.
4. Renaming some columns into short and descriptive names.
5. Adding new column ‘active cases’ with the help of other cases available in the dataset.
6. Checking null values and removing the unnecessary columns (>10% null values)
7. Checking and dropping duplicate values
8. Statistical description of the dataset



**We observe some negative values in the statistical description of the data and remove those values according to our need.**

1. Removing garbage values (negative values)
2. Finding out the first date and last date of the dataset.

Display total number of **Confirmed, Deaths** and **Recovered**:



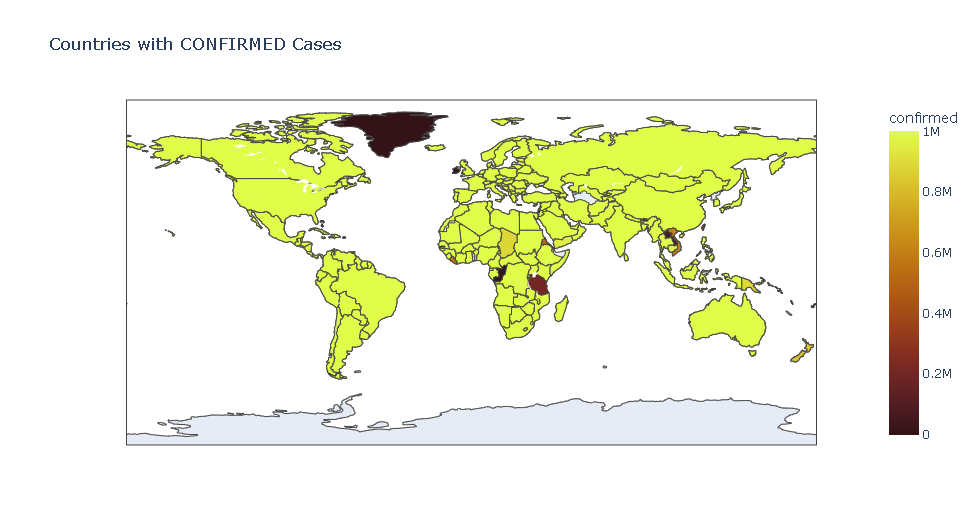
Countries with most Confirmed Cases (Interactive DataFrame):



Interactive Layout showing countries with most number of confirmed cases with 10 as default value.

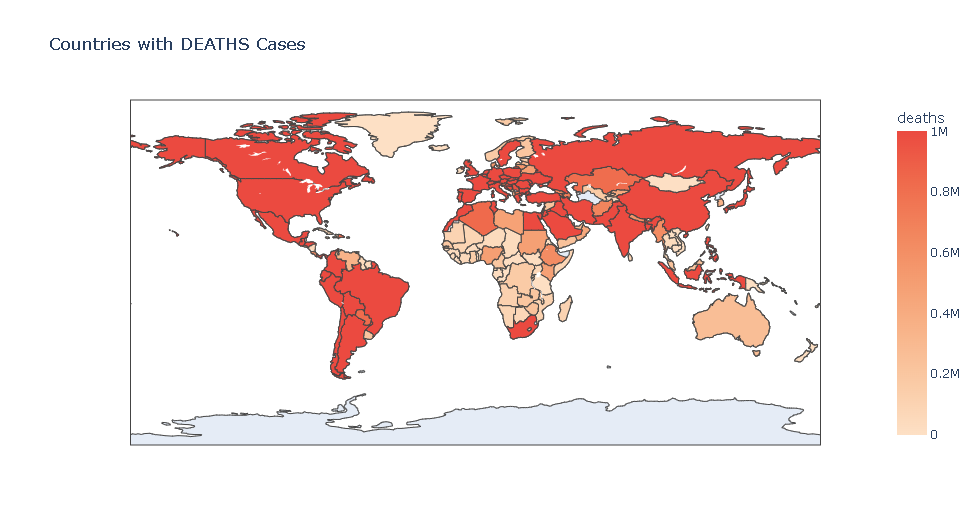
**Plotting on world map :**

Confirmed cases:



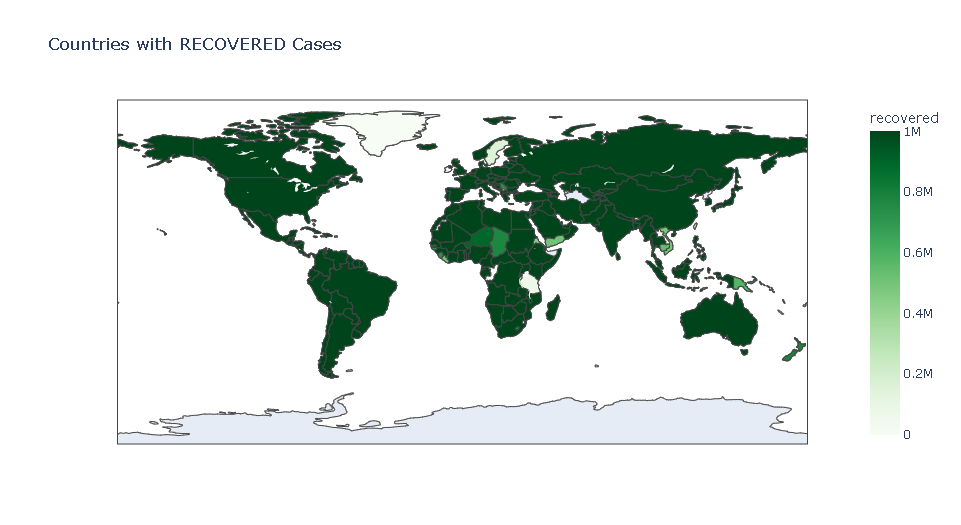
**Observation**: The confirmed cases are less in few countries of Africa.

Death cases:



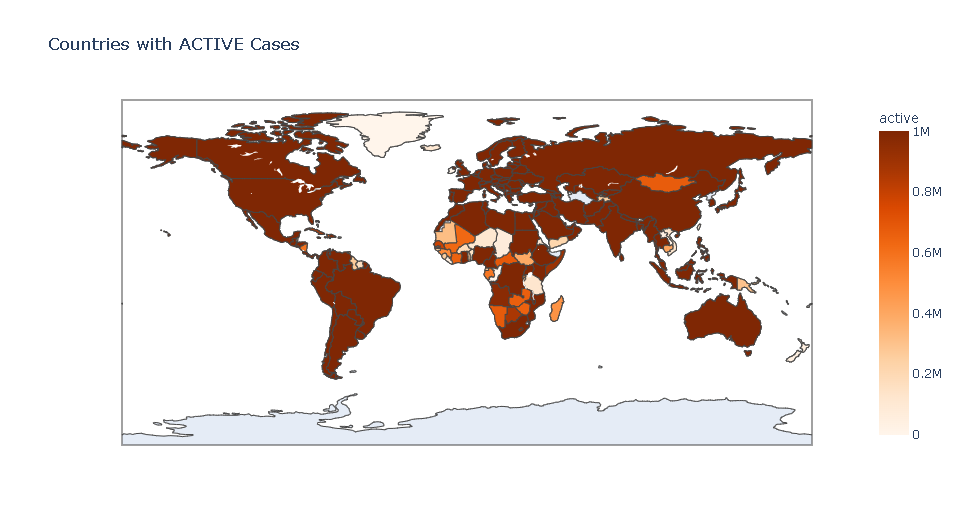
**Observation**: The death cases are least in most countries of Africa and in Mongolia in Asia.

Recovered cases:



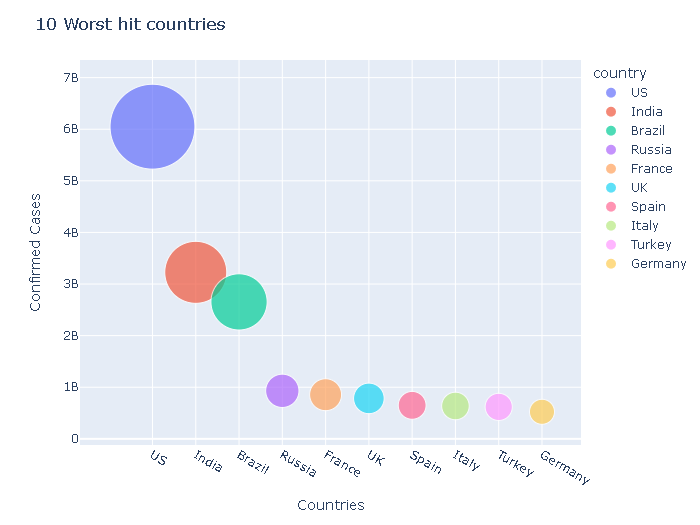
**Observation**: The recovered cases are less in few countries of Africa.

Active cases:

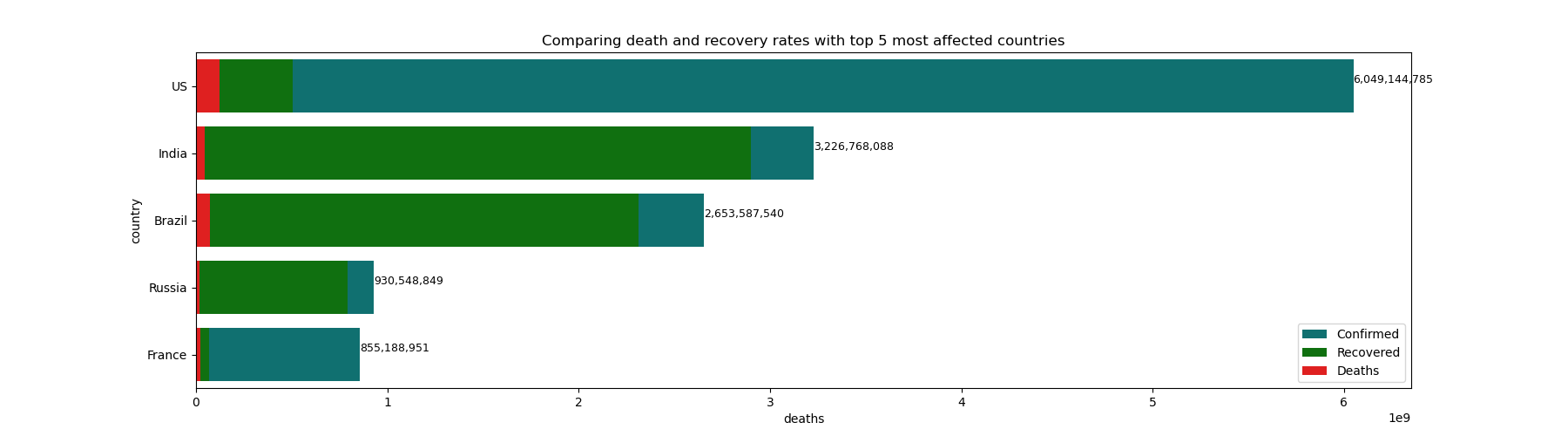


**Observation**: The active cases are less in most countries of Africa and Mongolia in Asia.

Top Worst hit countries (bubble chart):

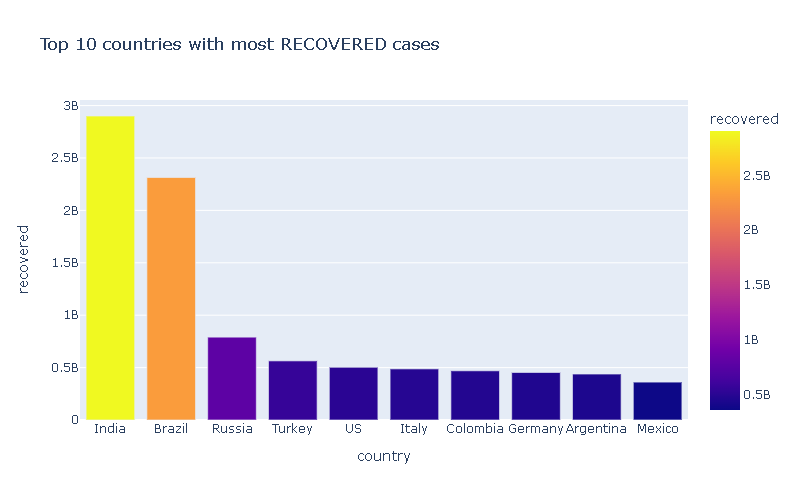
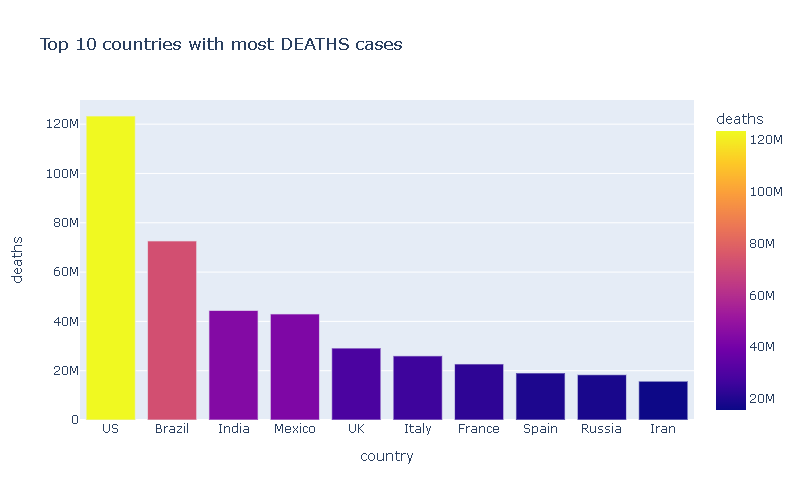
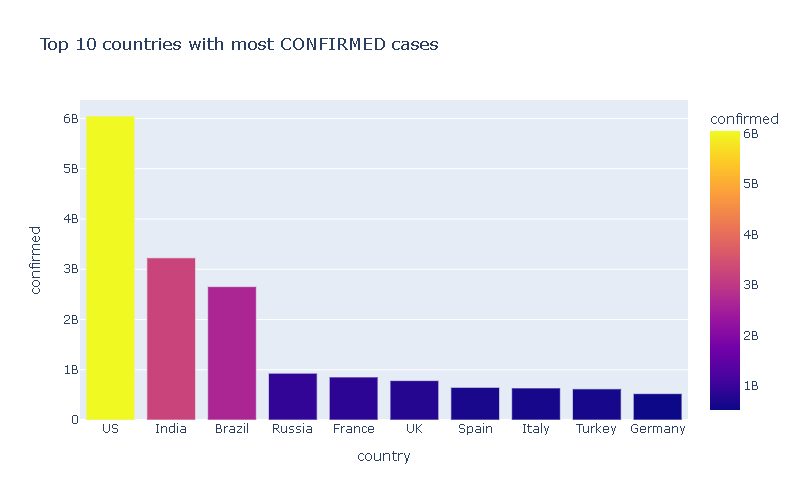
****

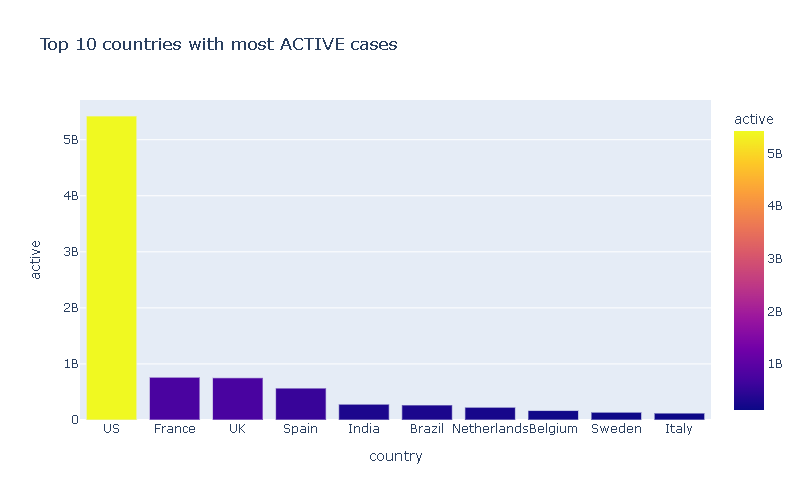
**Worldwide top 5 confirmed cases vs their death and recovery :**

****

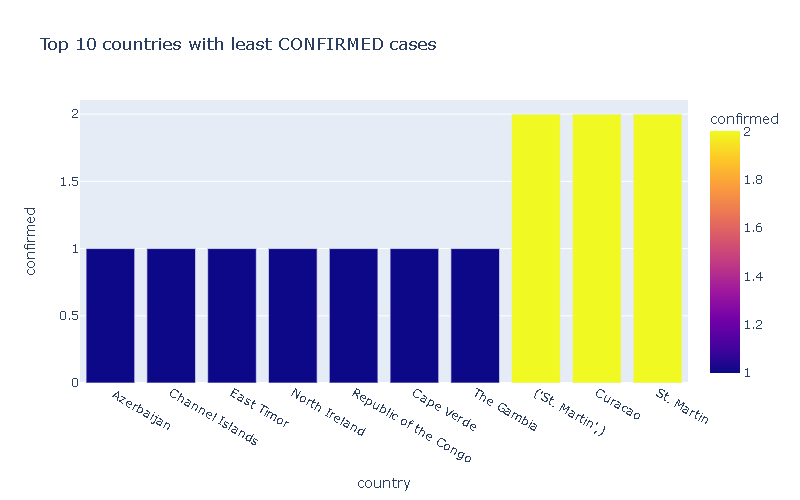
**Observation**: The recovery rate in US is least compared to its confirmed cases.

**Top 10 countries with most confirmed cases:**

****

****

**Countries with least number of confirmed cases:**

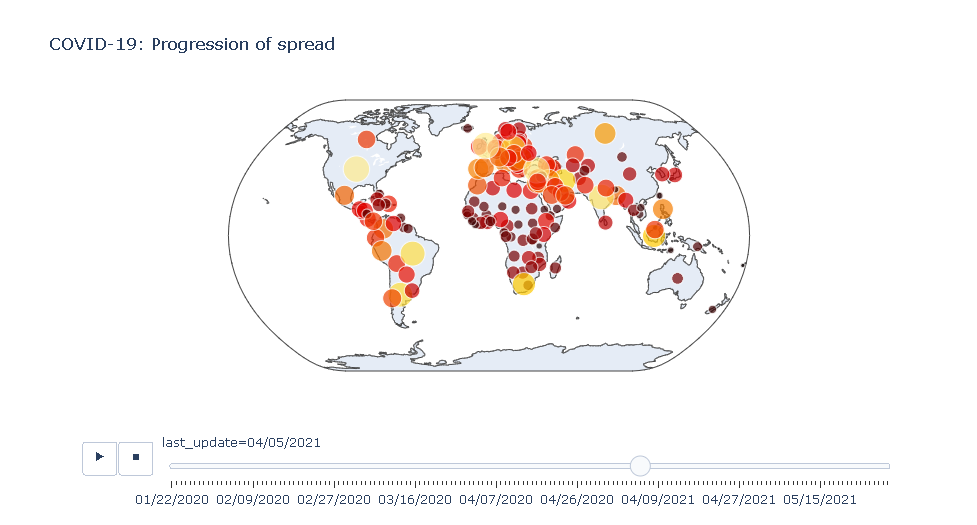
****

**Observation**: Countries like Azerbaijan, Channel Islands, East Timor, North Ireland, Republic of the Congo, Cape Verde, The Gambia are the most well maintained with the least number of cases.

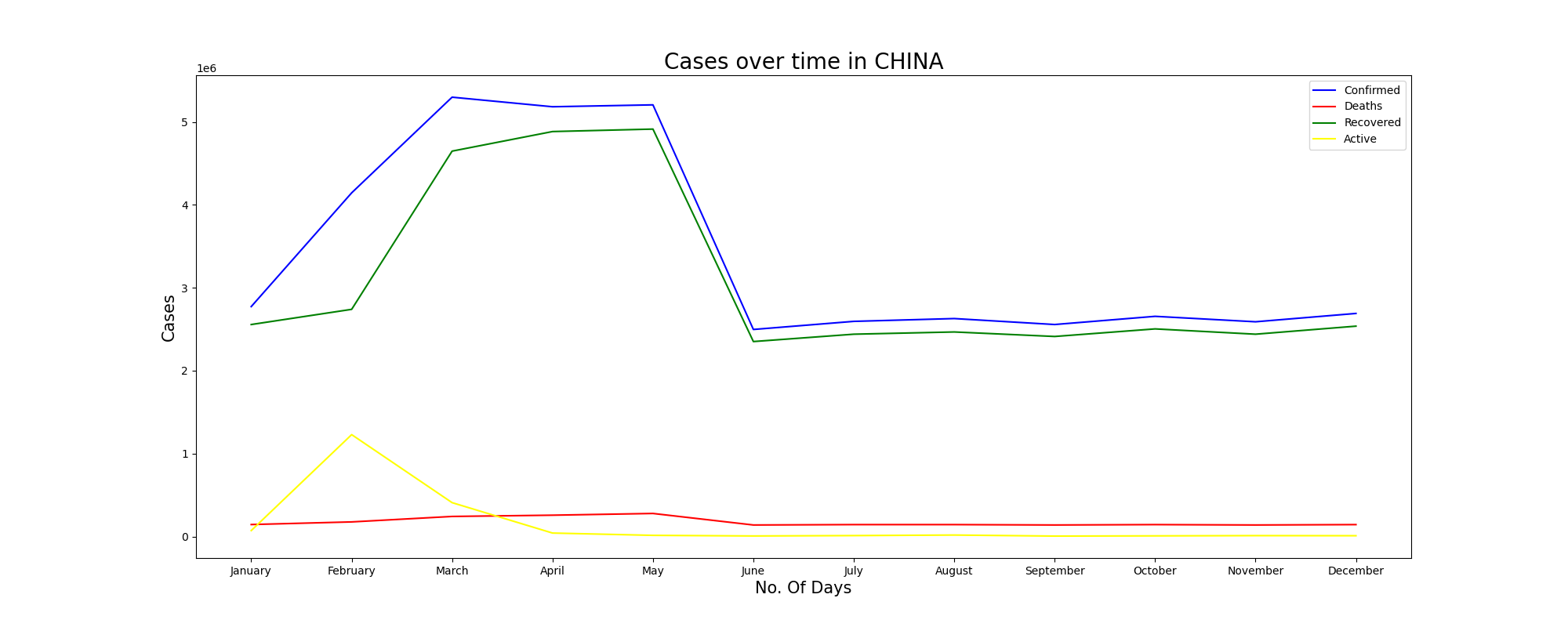
**Global Spread of Covid-19:**

****

**Covid-19: Progression of spread**

****

**Covid-19 Analysis on Mainland China:**

****

**Observation**: In China the increase of confirmed cases and recoveries took place from March to May. The recoveries were slow in February, while the confirmed kept increasing and there were very less deaths. Thus there was drastic increase in the number of active cases in February.

**Covid-19 Analysis on India:**

1. Import the datasets
2. Rename required column
3. Dropping unnecessary columns,

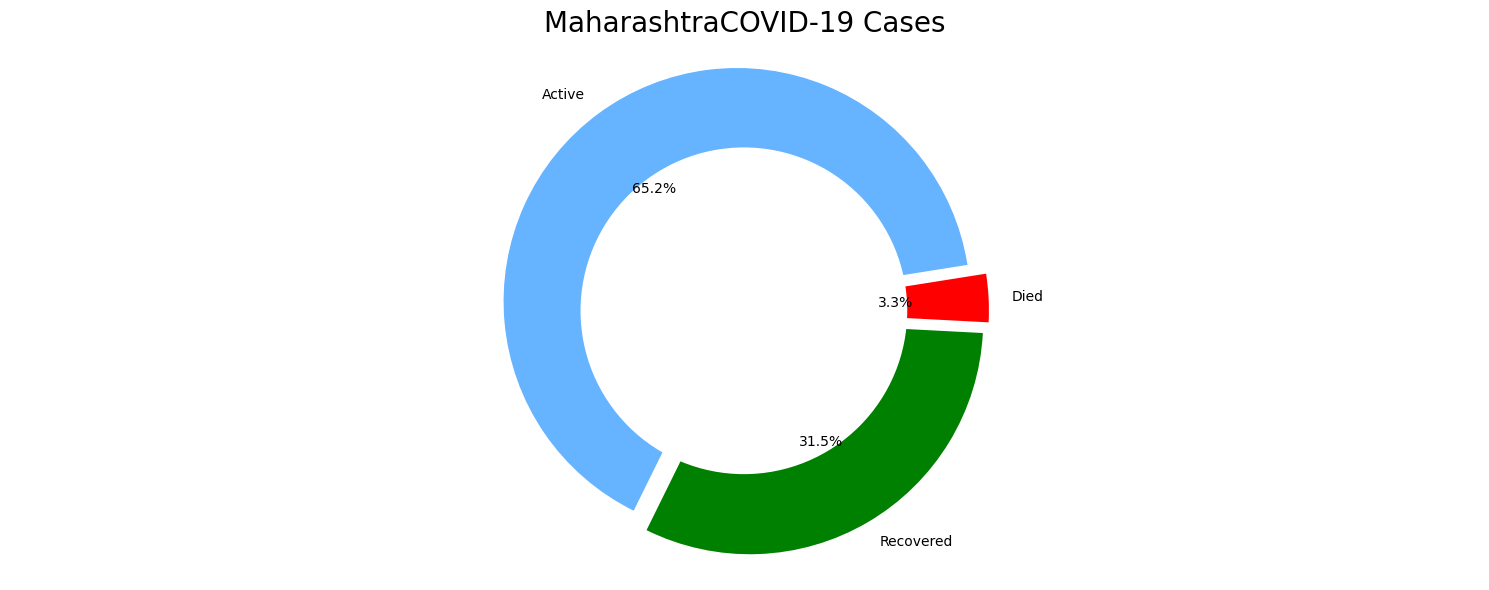
Converting ‘date’ column to datetime in specific format,

Replacing the dash with zeros in required columns

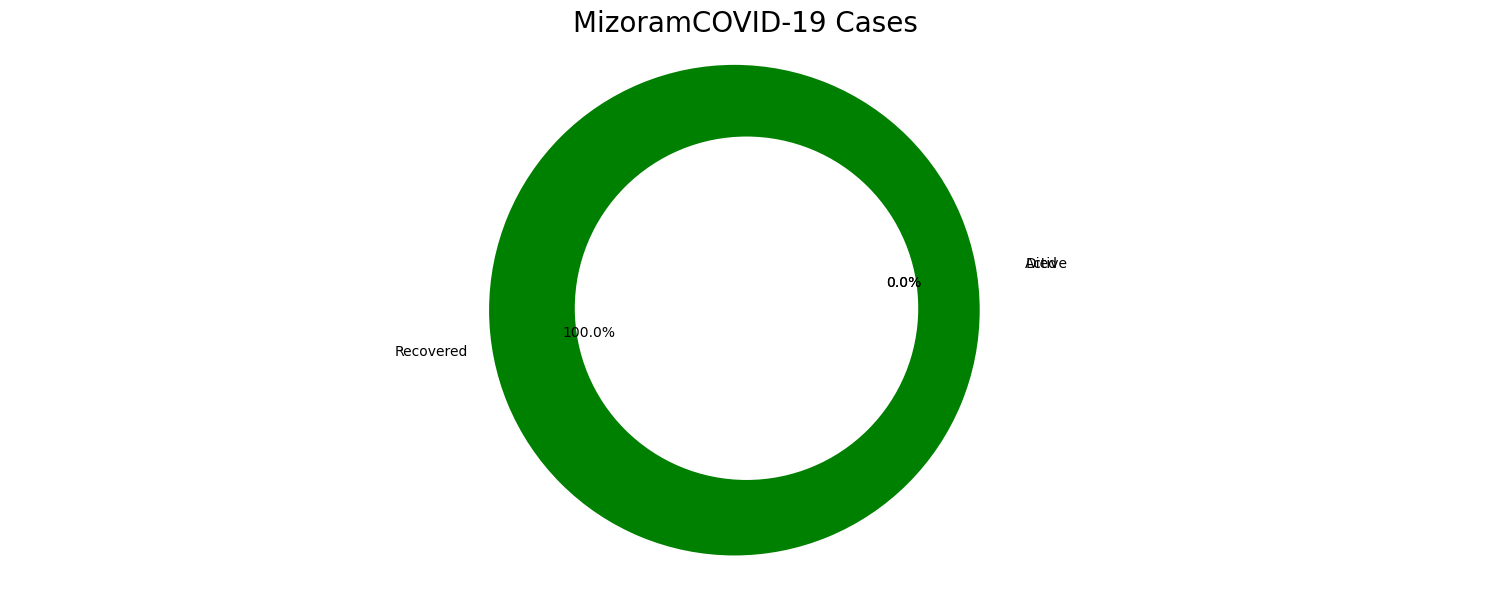
1. Finding the difference between the two dataframes with respect to states and correcting it.

India: Data Visualization

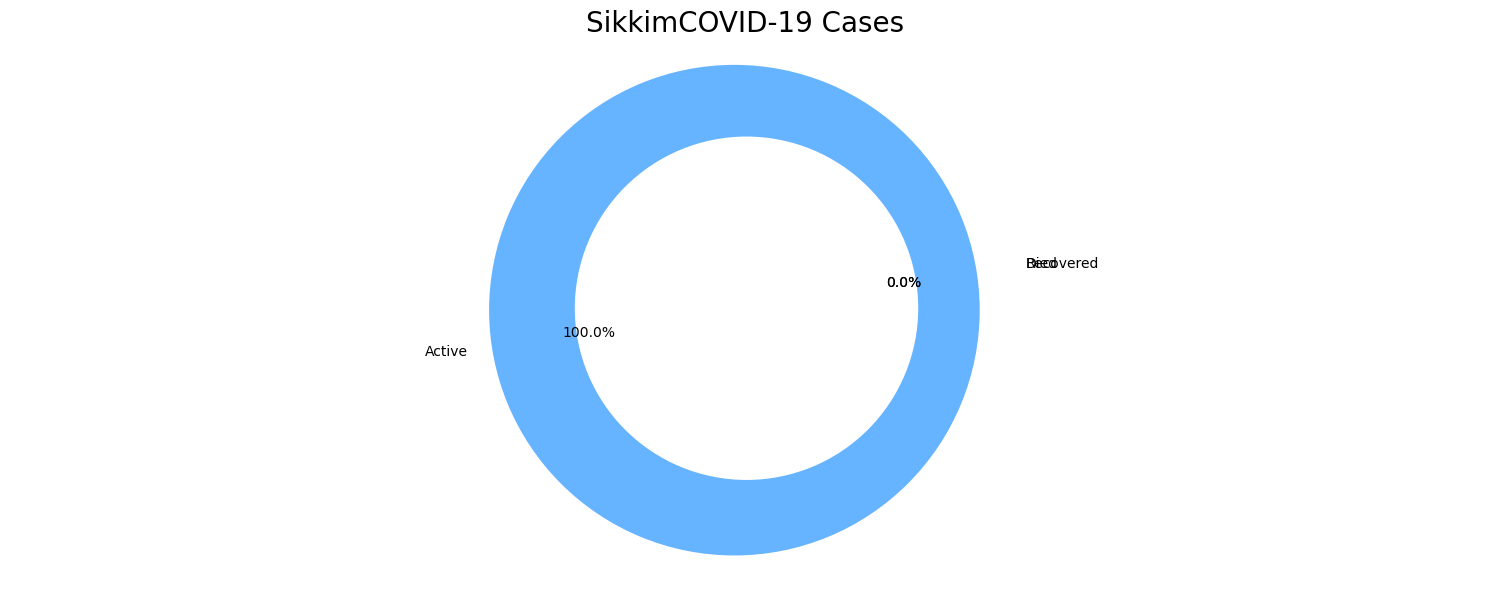
State-wise covid-19 status in India

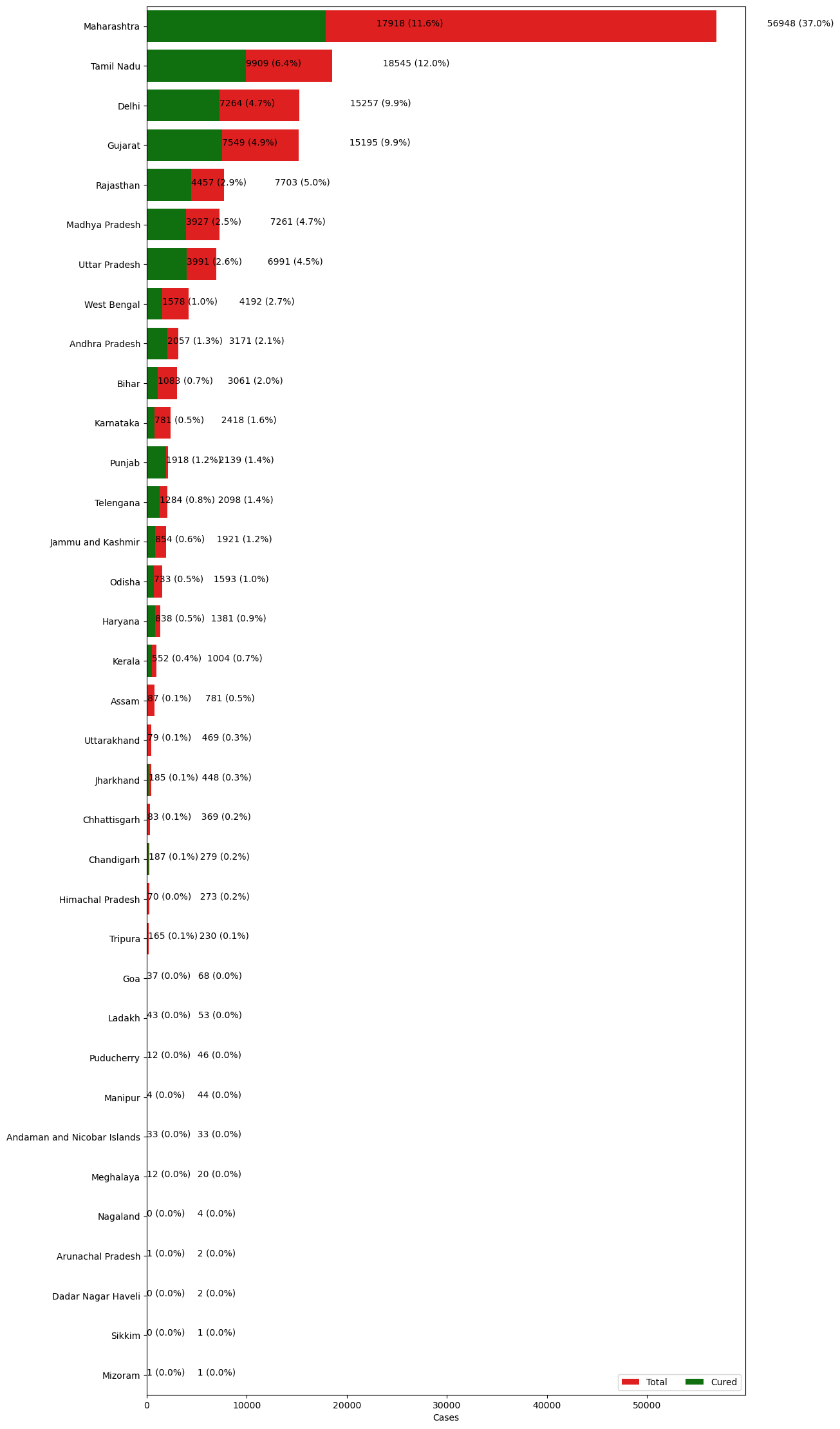


State with all Recovered Cases



State with all Active Cases

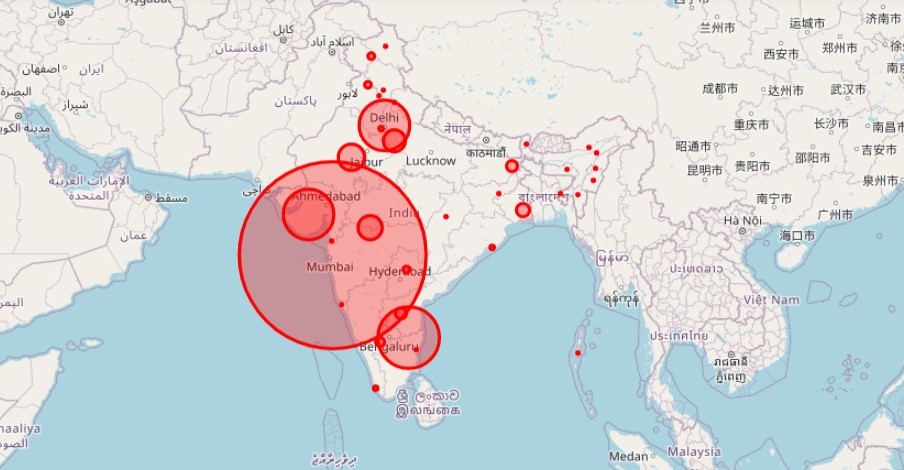




**Maharashtra has most confirmed cases (56948 i.e. 37.0%) and Mizoram has least confirmed cases (1 i.e. 0.0%).**

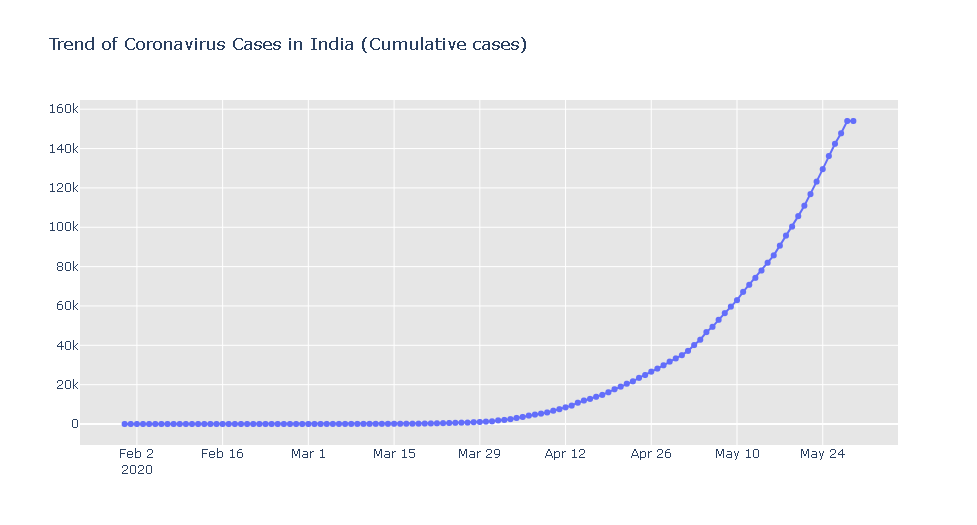
**Spread in India**

**Geographical Representation**

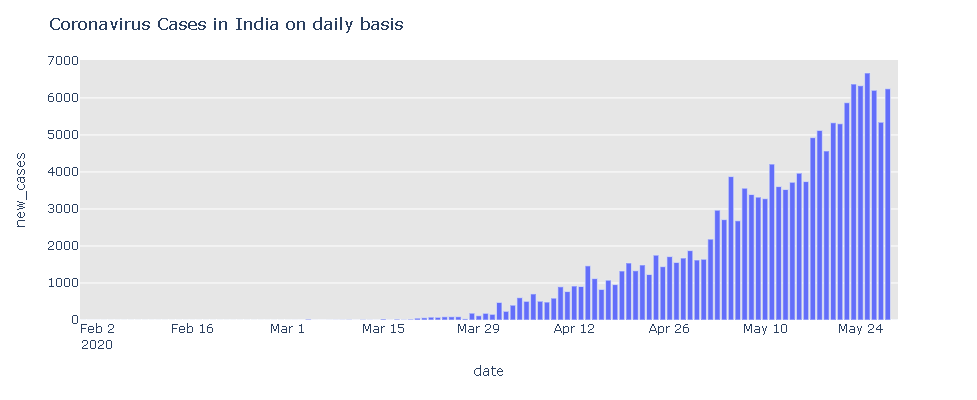
****

**Observation:** Most of the confirmed cases are found in south-west India.

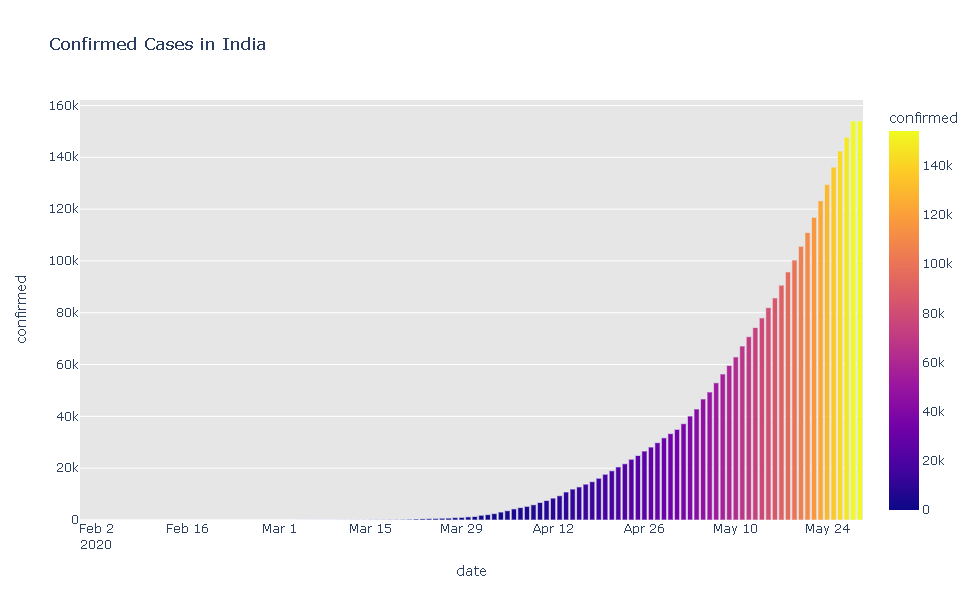
**Trend of cases in India**

**Observation:** The spread of coronavirus cases increased drastically from beginning of April 2020

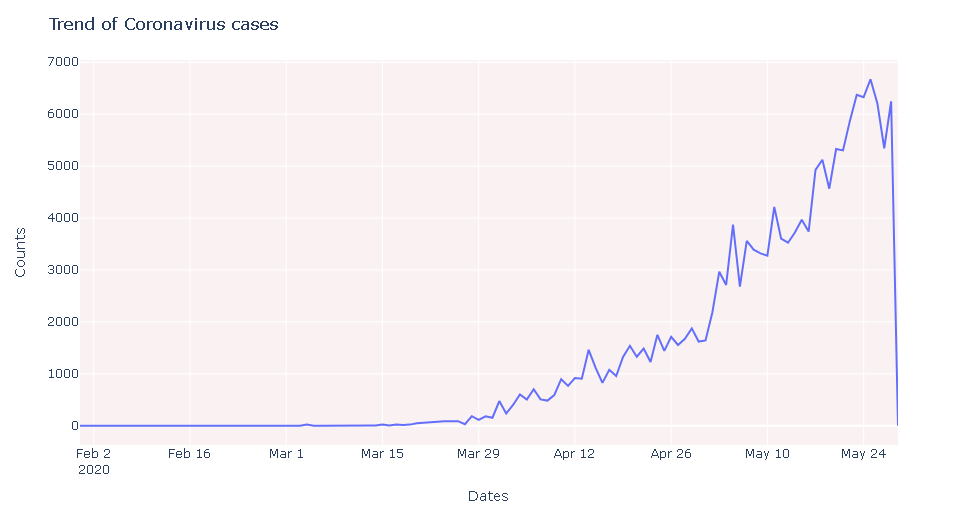
**Daily cases**

****

**Confirmed Cases in India**

****

**Trend of Coronavirus cases**

****

**THANK YOU**