SPOTLE.AI

Exploratory Data Analysis of COVID-19 Outbreak in India

Group Member:

PRIYANK JHA
VIDHYA SUBRAMANIUM
DEVLEENA BANERJEE
KARTHIK
MODABBIR

PROBLEM STATEMENT:

Developed a detailed Exploratory Data Analysis on the Covid19 Outbreak in India.

In this project, we have used the datasets provided in the below website for our analysis.

Dataset : https://api.covid19india.org/data.json

INTRODUCTION:

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus.

Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness.

The best way to prevent and slow down transmission is be well informed about the COVID- 19 virus, the disease it causes and how it spreads. Protect yourself and others from infection by washing your hands or using an alcohol based rub frequently and not touching your face.

The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, so it’s important that you also practice respiratory etiquette (for example, by coughing into a flexed elbow).

At this time, there are no specific vaccines or treatments for COVID-19. However, there are many ongoing clinical trials evaluating potential treatments.

Artificial intelligence applied to the medical domain can have very real consequence

COVID-19 tests are currently hard to come by — there are simply not enough of them and they cannot be manufactured fast enough, which is causing panic.

Given that there are limited COVID-19 testing kits, we need to rely on other diagnosis measures.

Since COVID-19 attacks the epithelial cells that line our respiratory tract, the outbreak is quiet robust and spreads through human contact who is a carrier of the virus.

And given that all hospitals are tracking and maintaining a record of the tested cases we are able to build the EDA of COVID-19 which gives us a statistical numbers on where we stand.

DATASET PREPARATION:

Dataset was taken from following link: https://api.covid19india.org/data.json

We have gathered all the possible information provided in the dataset to build the EDA of Covid19

We ended up with more than 10+ statistical graphs and arrive at a clear picture as on the data provided by 10th May 2020.

DETAILED DESCRIPTION AND ANALYSIS

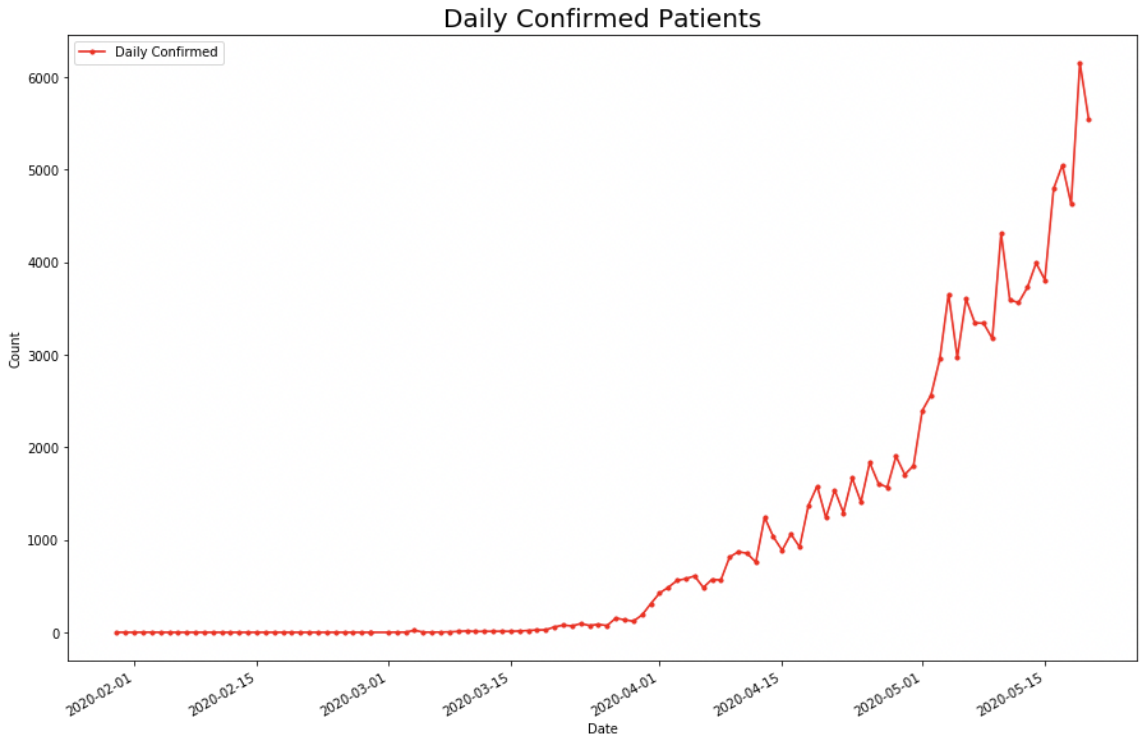
- Firstly, we pre-process the given datasets.
- A high level data exploration was conducted to understand the details of the datasets provided in the website.
- The data was provided in the 'Object' type wherein the specific columns had to be converted to 'numeric' type of data.
- More conversion of the type of data was carried out on the 'Date' column which helped in plotting the data in the graphs.
- In depth analysis of the data was carried out for the respective columns of the dataset:

    1. Daily Confirmed Patients
    2. Total Confirmed Patients
    3. Daily Recovered Patients
    4. Total Recovered Patients
    5. Daily Deceased Patients
    6. Total Deceased Patients

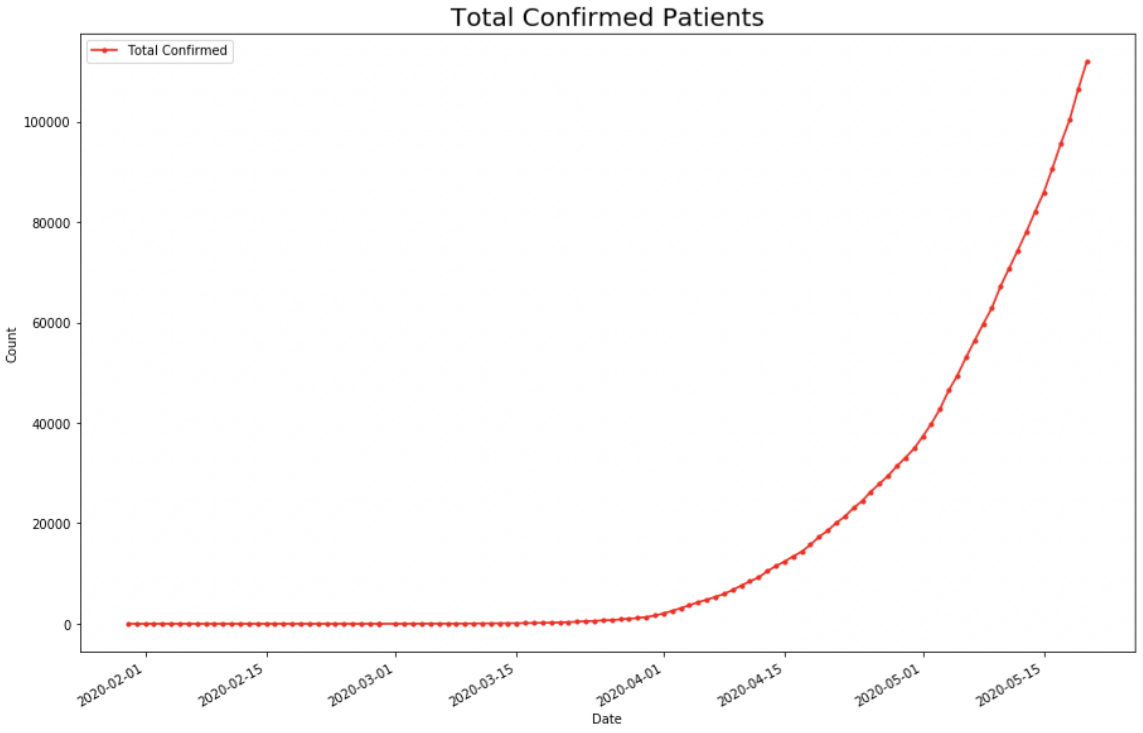
- A final pair-plot was used to give an overall picture of the cases.

DESCRIPTION OF EDA USING GRAPHS:

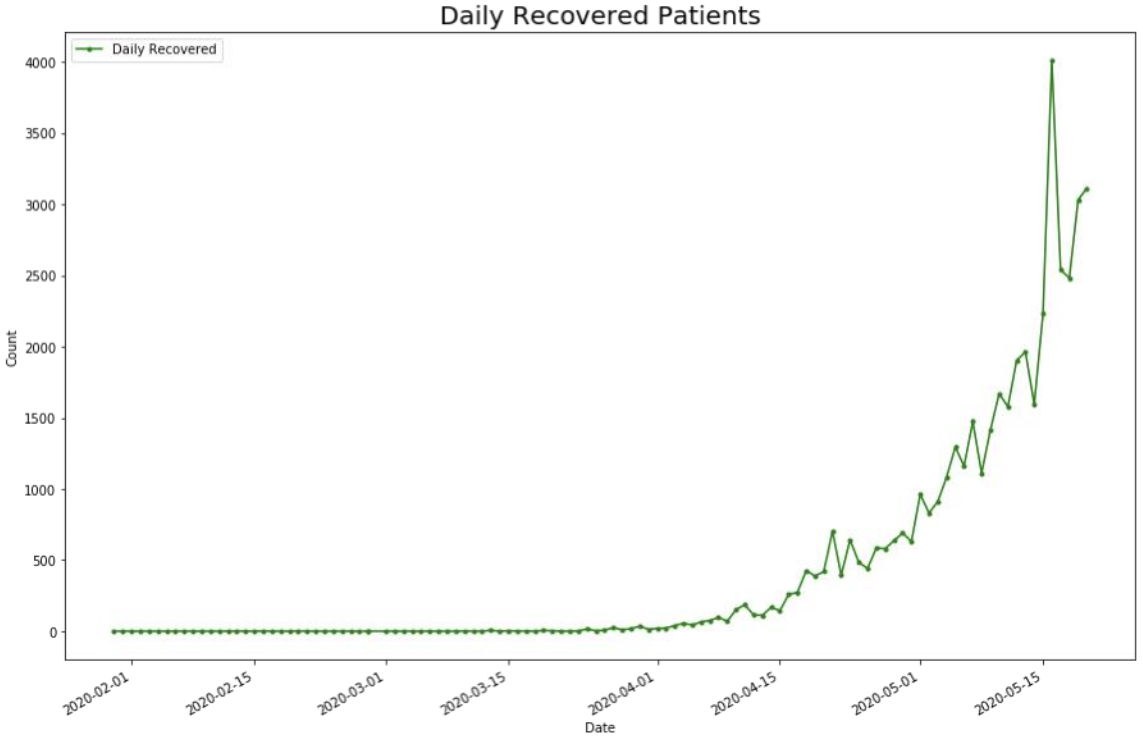
On doing an everyday analysis we clearly see that the number of cases raising was flat until the beginning of April 2020. A slow and steady raise is observed in the number of cases peaking at few intervals – around 15th of April, 5th of May which clearly indicates that the numbers are doubling every 10-15 days or so

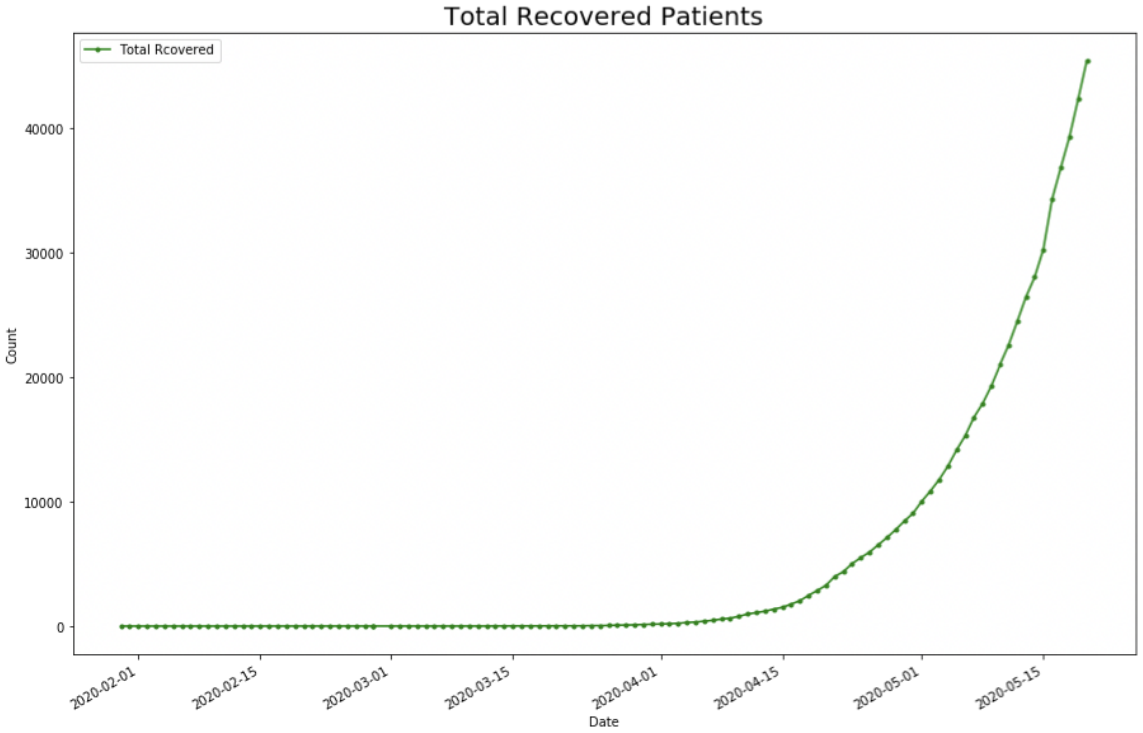


The raise is very steady and peaking in the month of May is very visible in this graph.

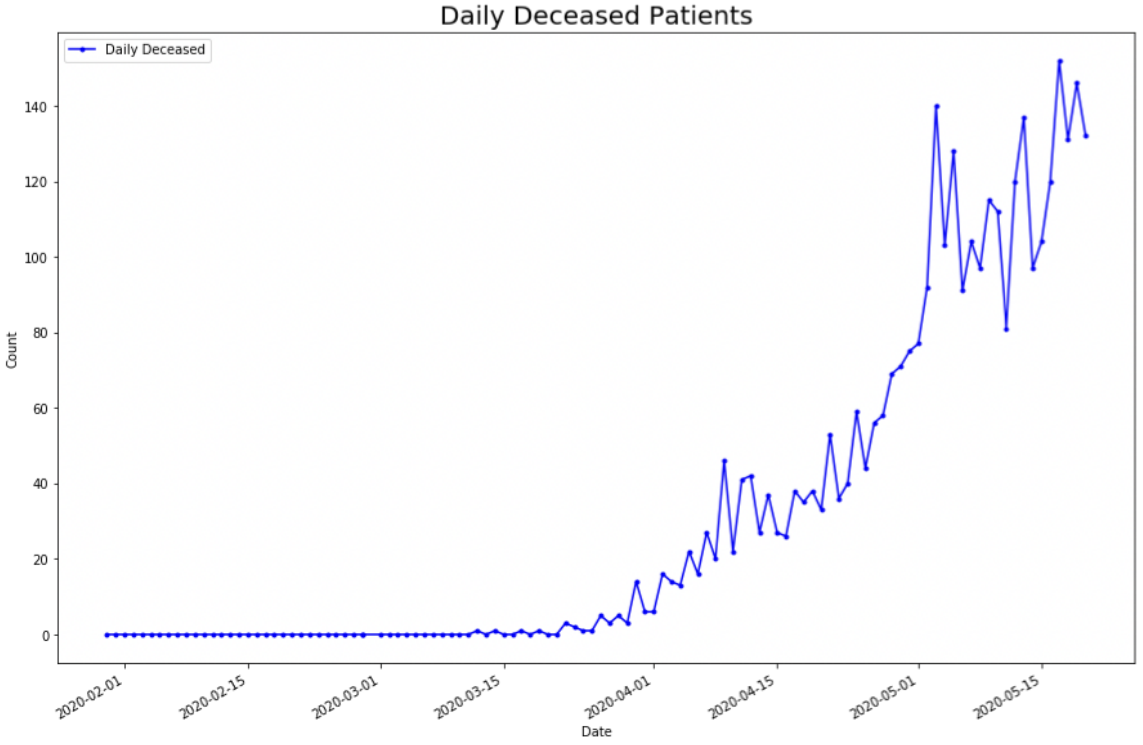


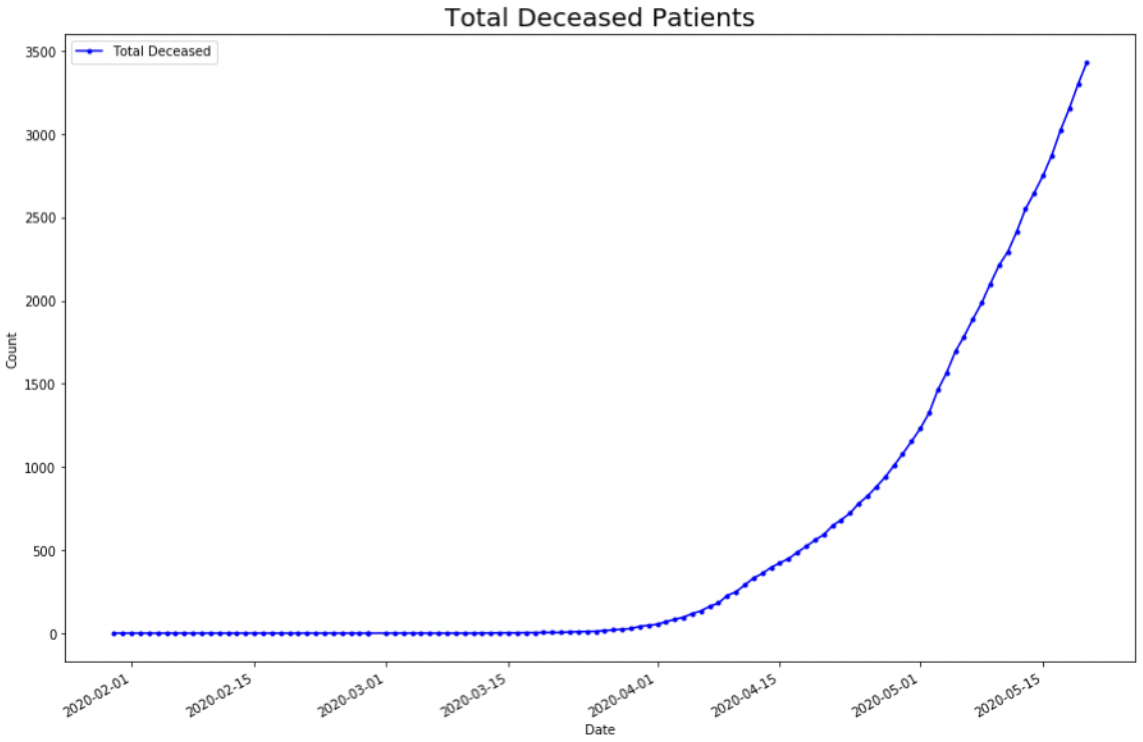
The no of cases recovered from the virus is visibly equal to the number of cases found positive.



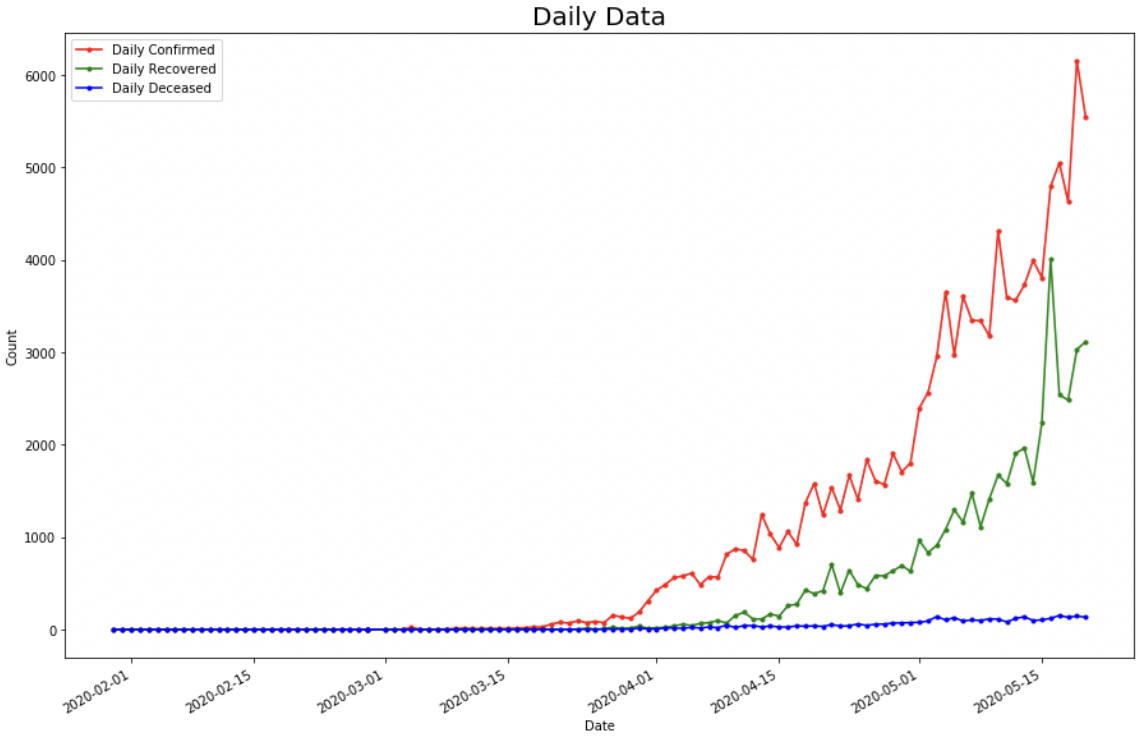


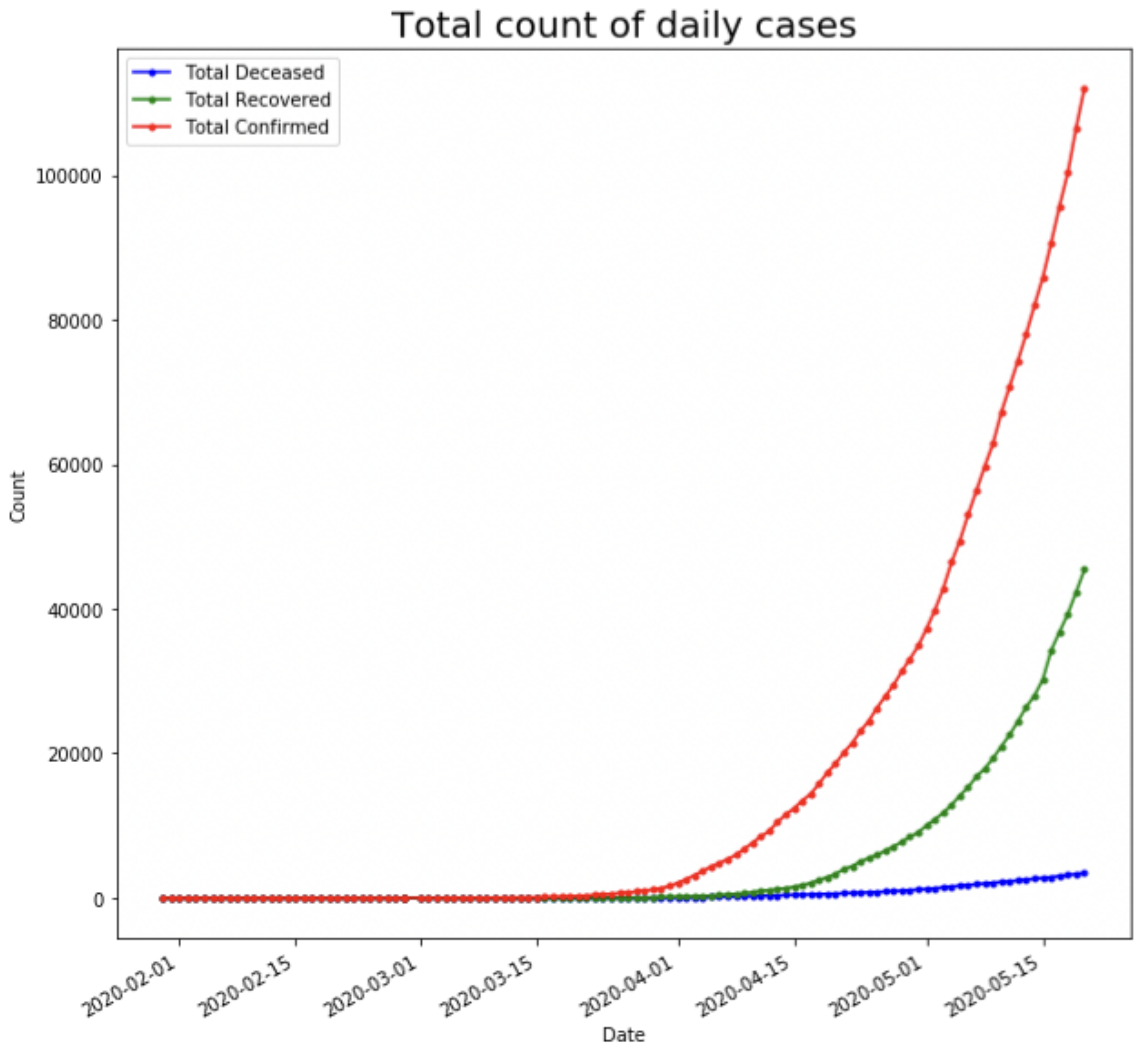
Though the deceased graph individually stands equating with the number of cases tested positive and recovered the overall % of deceased number stands very low globally in India.





The graph below describes the overall cases of tested positive, recovered and deceased. In this the blue line indicates the deceased which clearly stands lower compared to recovery and tested positive.





As we can see from the above results of the COVID-19 EDA Analysis its evident that the number of positive cases are quiet steeply increasing but the deceased is comparatively very low

LIMITATIONS AND FUTURE IMPROVEMENTS

- Lack of enough reliable data on major transmission case.
- Of all the cases, if there is specific information on the cause of transmission captured it would help for future analysis

REFERENCES:

WHO – Health topics/ Coronavirus

Ministry of Health and Family Welfare - https://www.mohfw.gov.in/