Report by B-34 GROUP . Pragati Sinha(B19CSE065) ,

Parmar Devanshi(B19CSE062), Surbhi Dua (B19CSE088),

Sriramoju Sreeja(B19CSE086).

Descriptive Data Analysis

COVID-19 Data in India

The SD is just a measurement to tell how a set of values spread out from their mean.

The smallest possible value of SD is zero. If SD is zero, all the numbers in a dataset share the same value.

Reporting the SD along with the mean gives one the impression of how valid that mean value actually is (i.e. if the SD is huge, the mean is totally invalid - it is not an accurate measure of central tendency, because the data is so widely scattered.)
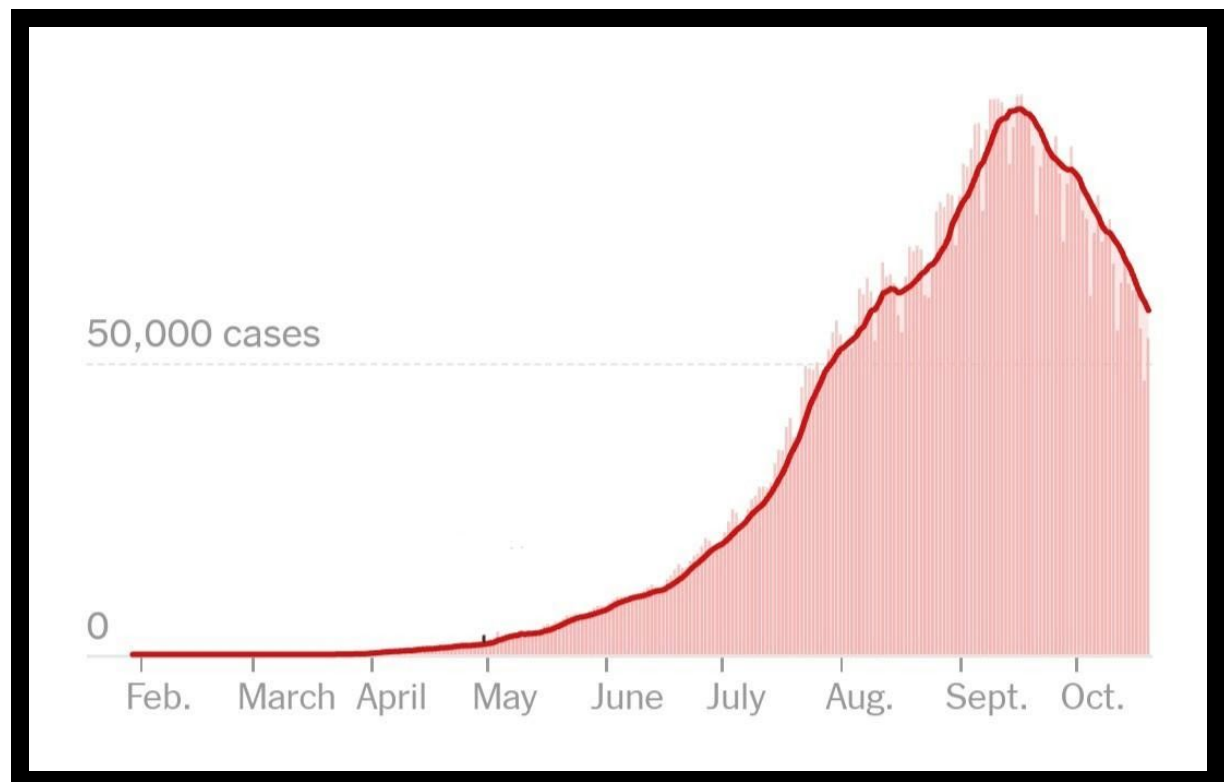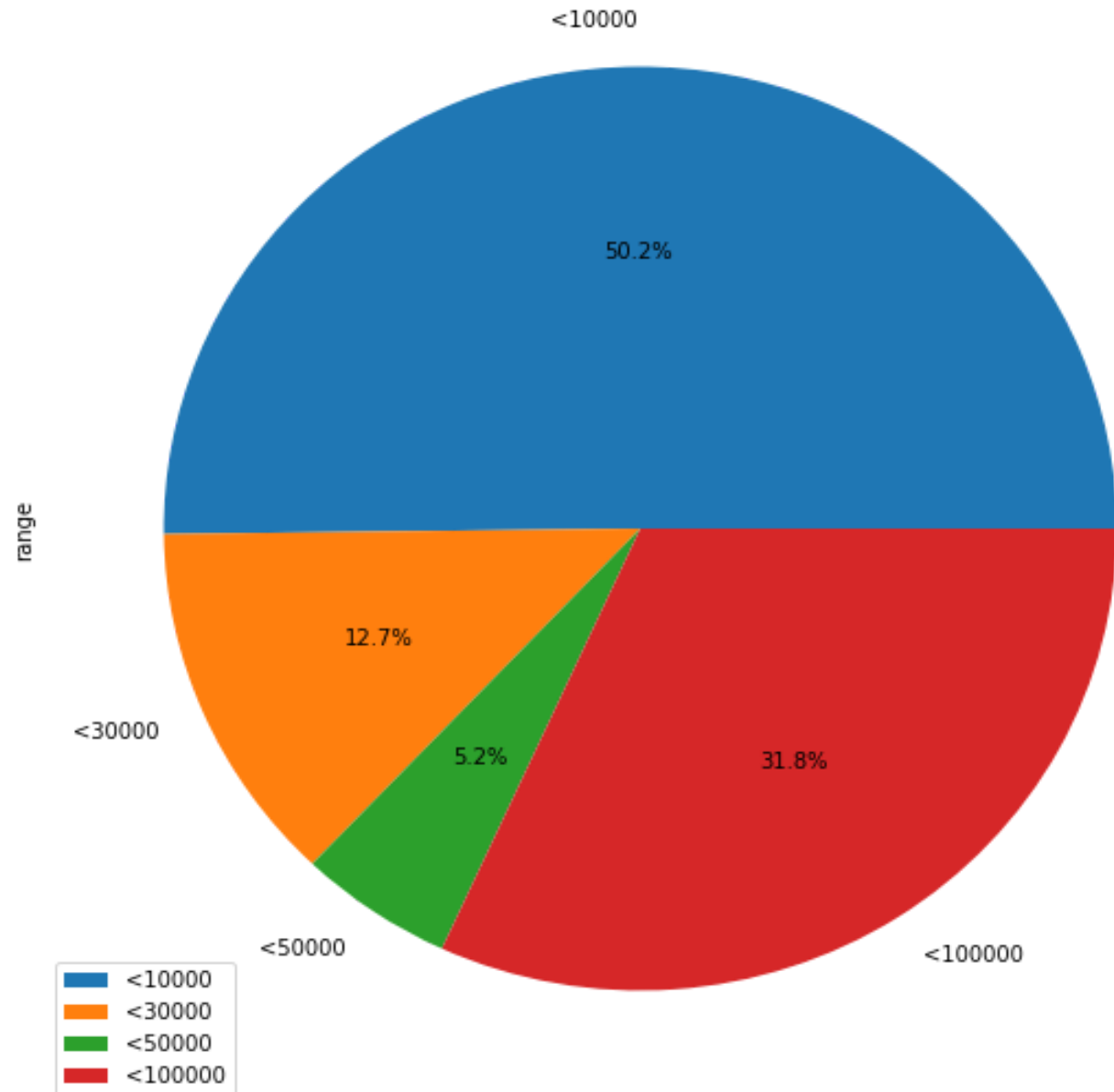
deviation = exact value - average

a) From the first case till date :

❖ **Data:**
From first case till 22nd October 2020

❖ **Graphical Representation:**

❖ **Measures of Central Tendency and Dispersion**

i) Mean=28973.48

ii) Median= 10331.5(middle of June)

iii)   Mode = <10000 (50.2%)

iv)   Standard Deviation=32630.76

v)   max= 97894 (17/09/2020),deviation=68920.58 (very high)

vi)   Range : [0,97894]

❖ <u>Interpretation</u> :-

The first case in our country was recorded on 30th Jan 2020. According to the mean value 28,973 cases are recorded in India daily but the standard deviation is also very high implying that the data is widely scattered ,daily recorded COVID-19 cases in India have drastically increased from Feb to Sept. However, they have decreased steeply from Sept to Oct (as observed from the graph). The median for data is way less than mean and maximum implying that the virus was very much under control till June 2020.From pie chart from 1st case recorded on 30th jan till oct 22 for 50.2 % of the days coronavirus recorded was less than 10000. However these days were till 11th June 2020 only . For 31.8 % of the days covid-19 range was 50000-100000 starting from 30 July implying how fastly corona grew from June to July.
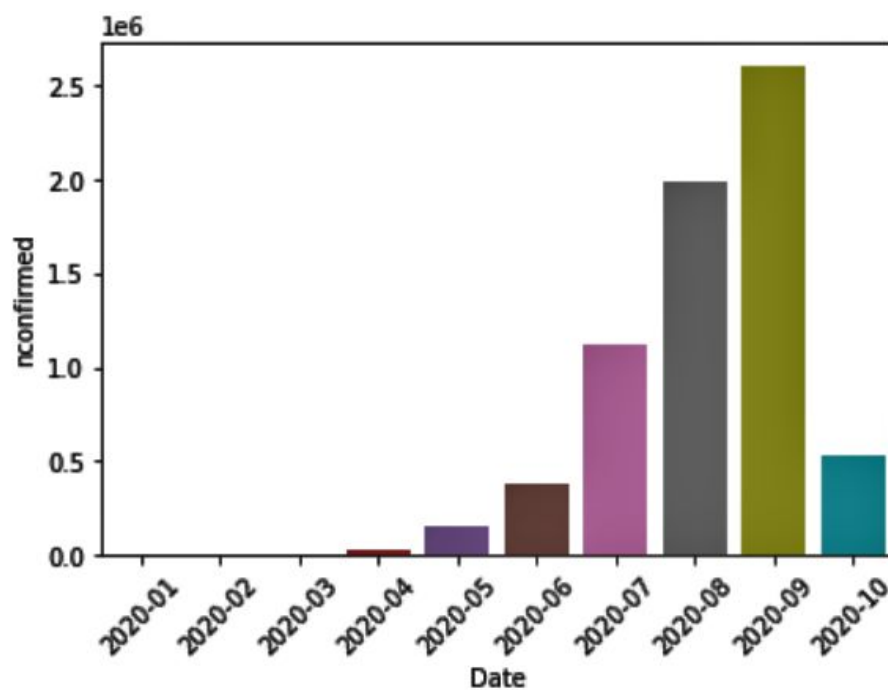
<u>Apt : Mean.</u>
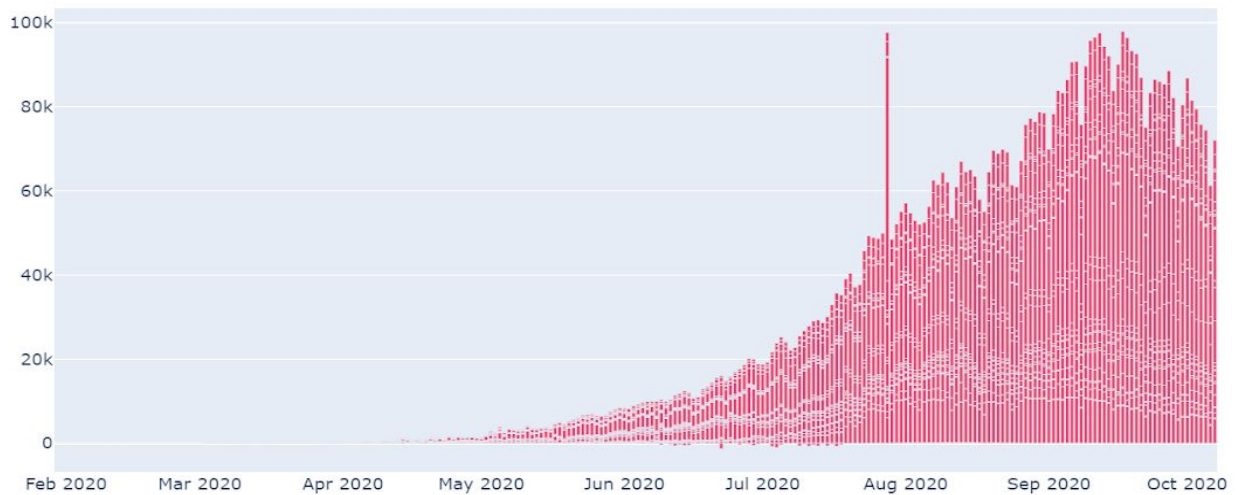
b) **Average Per month:**

❖ **Data:**

Given is the data for average per month:

Avg per month

❖ **Graphical representation:**

❖ Measures of Central Tendency and Dispersion

i) Mean =691751.45

ii) Median = 149093.0 (in May)

iii) Standard Deviation =982797.39(very high)

iv) Max = 2604518 (September 2020) , deviation=1912766.5(very high)

v) Min = 1 (Jan 2020)

vi) Range = [1,2604518]

❖ Interpretation:

Clearly from the graph itself we can see that covid-19 in the country had

reached its peak in September 2020. According to the mean every month

691751.45 people are confirmed corona positive but at the same time S.D

val is also very high hence this mean data is of no significance . Corona in our country has drastically increased from January to September.Median is 149093 in may way less than mean implying corona was in control in May 2020.
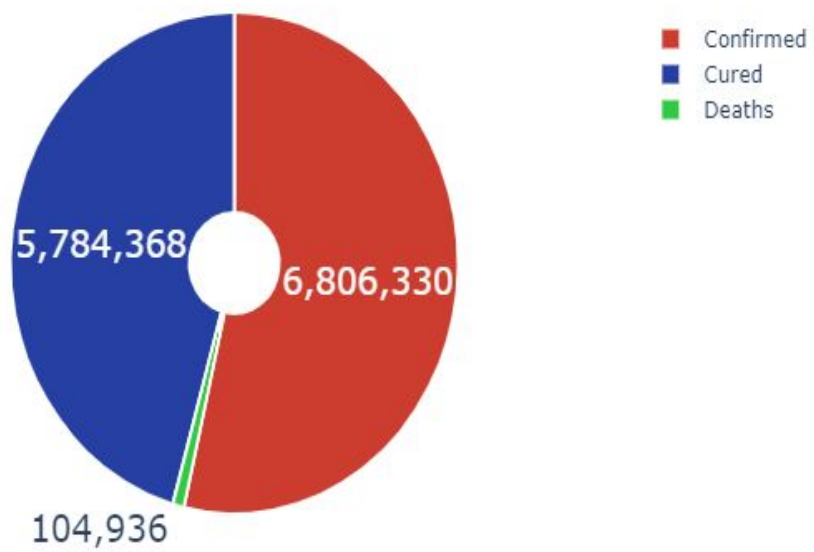
<u>Apt : mean</u>

c) <u>State wise average and recovery rate</u>

❖ Data:

Given is the data for state wise confirmed, cured and death in each state and union territory of India till 7th October 2020:
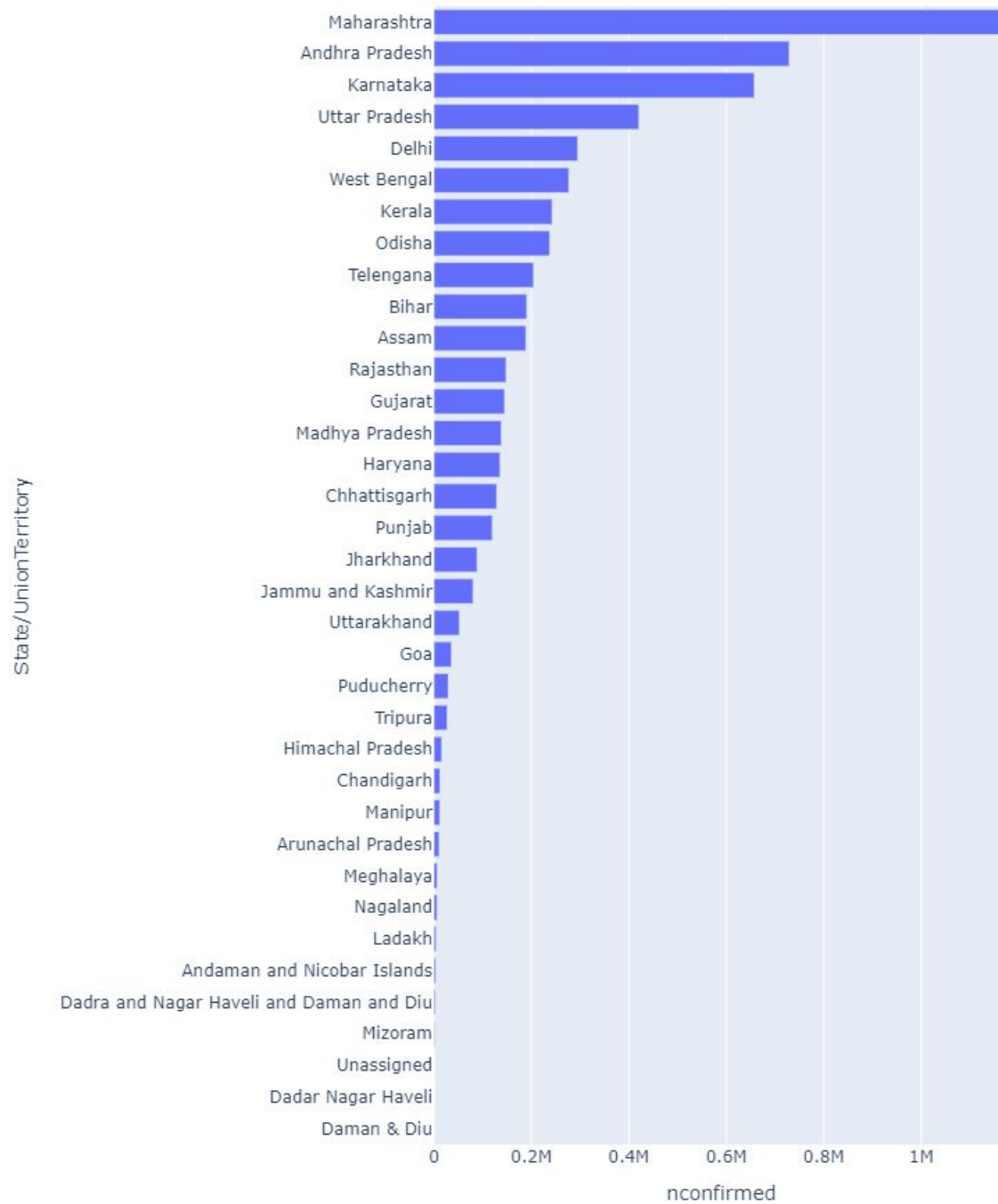
Statewise_cured_death_confirmed

Current Situation in India

a) **AVERAGE**

Total nconfirmed Cases

❖ **Measures of Central Tendency and Dispersion**

i) Mean=193060.886

ii) Median = 120016 (Punjab)

iii) Standard deviation= 292539.72 (very high)

iv) Max = 1465911(Maharashtra) , Deviation (Maharashtra)=1273020.67

iv) Min = 2148(Mizoram) , Deviation (Mizoram)= -190742.33

Vi) Range= [2148 , 1465911]

Interpretation: According to the average that is 193060.886 every state in our country has encountered this many cases till now.But the actual picture is completely different ! As we can see in the graph only a few states have only maintained this average some are extremely high some are extremely low. The high value of standard deviation implies the same that how dispersed has been the reach of corona so far in the states of the country . The median value for Punjab is way less than max achieved by Maharashtra .

<u>APT : MEAN</u>

b) RECOVERY RATE :

Graph:

Recovery Rate

Recovery Rate
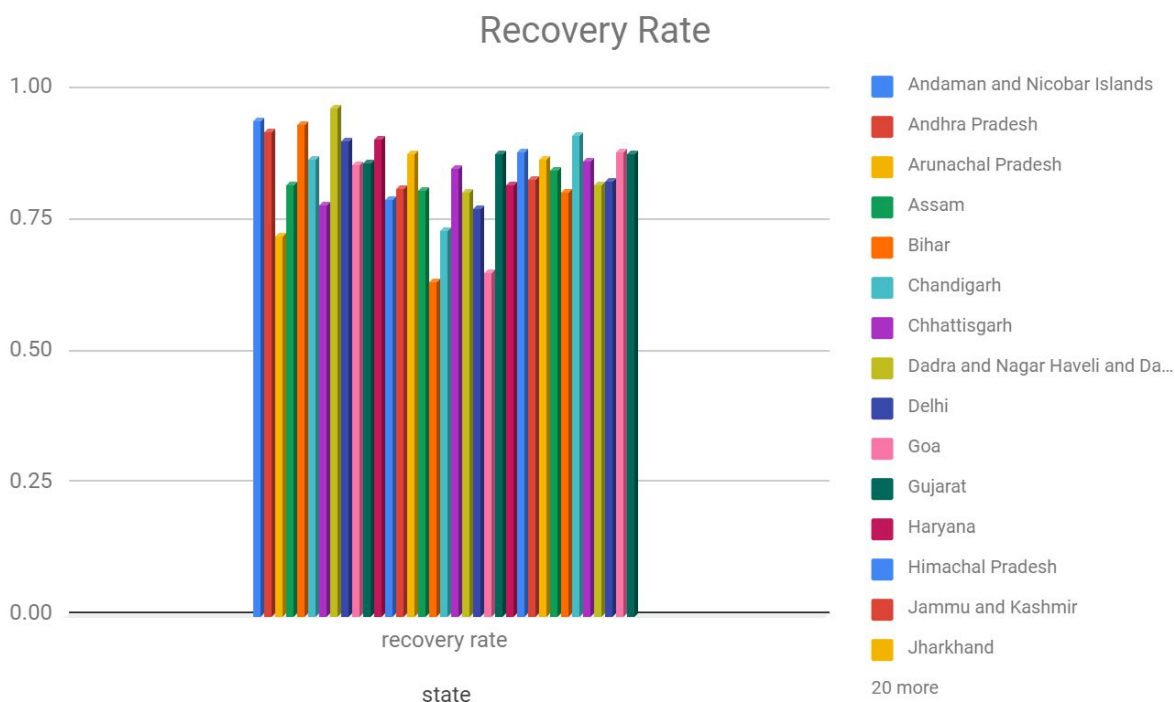
❖ **Measures of Central Tendency and Dispersion**

i) Mean= 83.87%

ii) Median = 85.12 % (Madhya Pradesh)

iii) Standard deviation = 7.25 % (less)

iv) Max = 96.68%(Dadra and Nagar Haveli and Daman and Diu), Deviation= 12.81%

v) Min = 63.46% (Kerala), Deviation= -20.41 % (high !!)

vi) Range =[63.46%,96.68%]

INTERPRETATION :

Every state of our country has on an average 83.87% of the cases recovered from the disease . The value of sd being quite low (7.25%) implies that this data is not dispersed. Hence covid-19 has not proven to be a deadly disease in the country . However states like kerala and meghalaya have lagged behind in the recovery rate (owing to deviation of 20.41 % for kerala)as compared to other states making the average lower as compared to the median value of mp (85.12%).While other states and union territories like AP, bihar , delhi , haryana , TN have recovery rate more than 90 % with highest being achieved by Dadra and Nagar Haveli and Daman and Diu (96.68%).
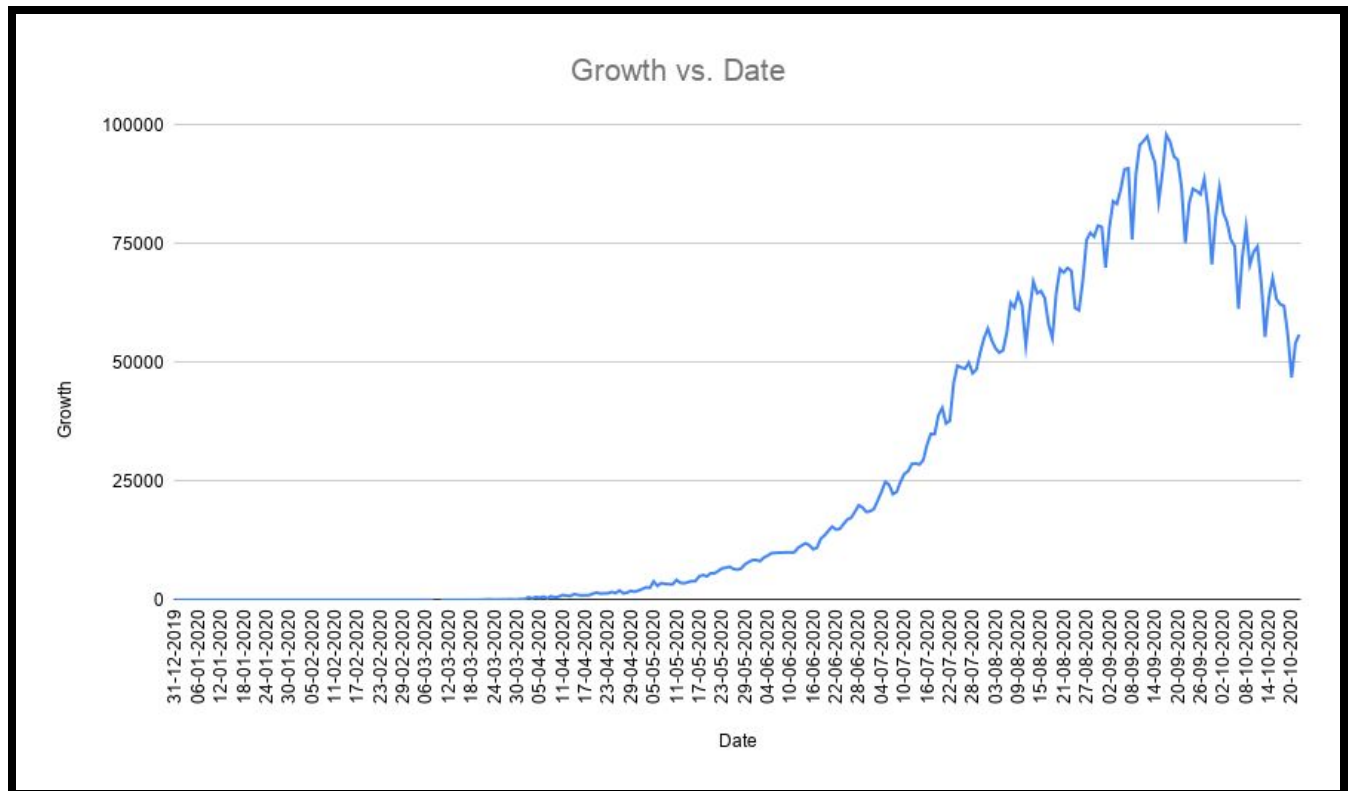
<u>Apt : mean</u>

d) <u>Growth</u>

Data:

Given is the data for growth rate:

<u>Growth</u>

Graphical Representation :

Measures of central tendency and dispersion:

i) Mean=28973.48

ii) Median= 10331.5(middle of June)

iii) Mode = <10000 (50.2%)

iv) Standard Deviation=32630.76

v) max= 97894 (17/09/2020),deviation=68920.58 (very high)

vi) Range : [0,97894]

Interpretations:

Clearly covid-19 in our country started increasing fastly in may 2020 with reaching its median in the middle of June which is (10331.5 only). However the real upsurge started in the month of july and reached its peak in september 2020 with 17/09/2020 recording the highest coronavirus cases for a day in the country.After the upsurge cases have decreased but still maintained an avg of 73,962.66 every day .

e)Death Rate

Data : Death rate

Graphical representation :
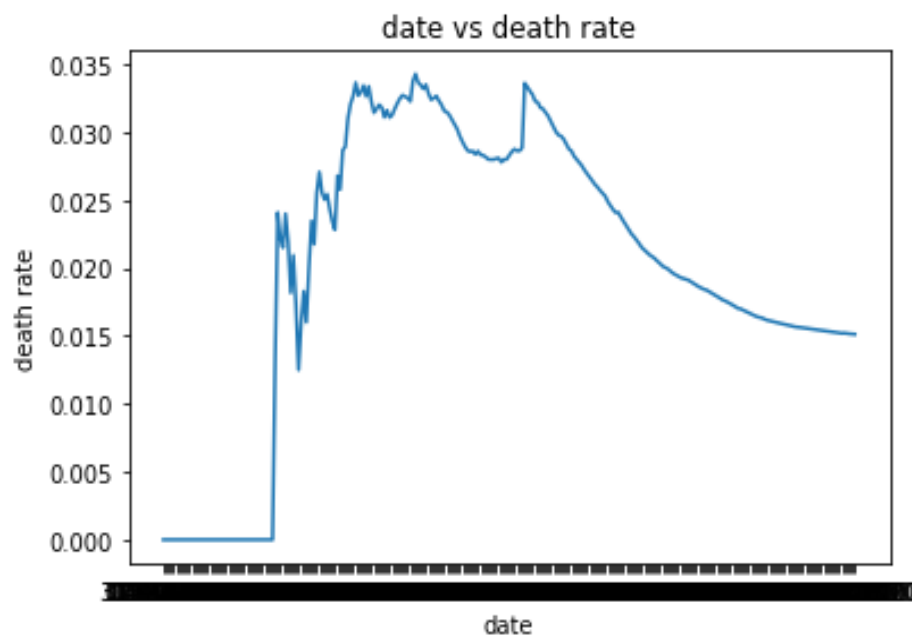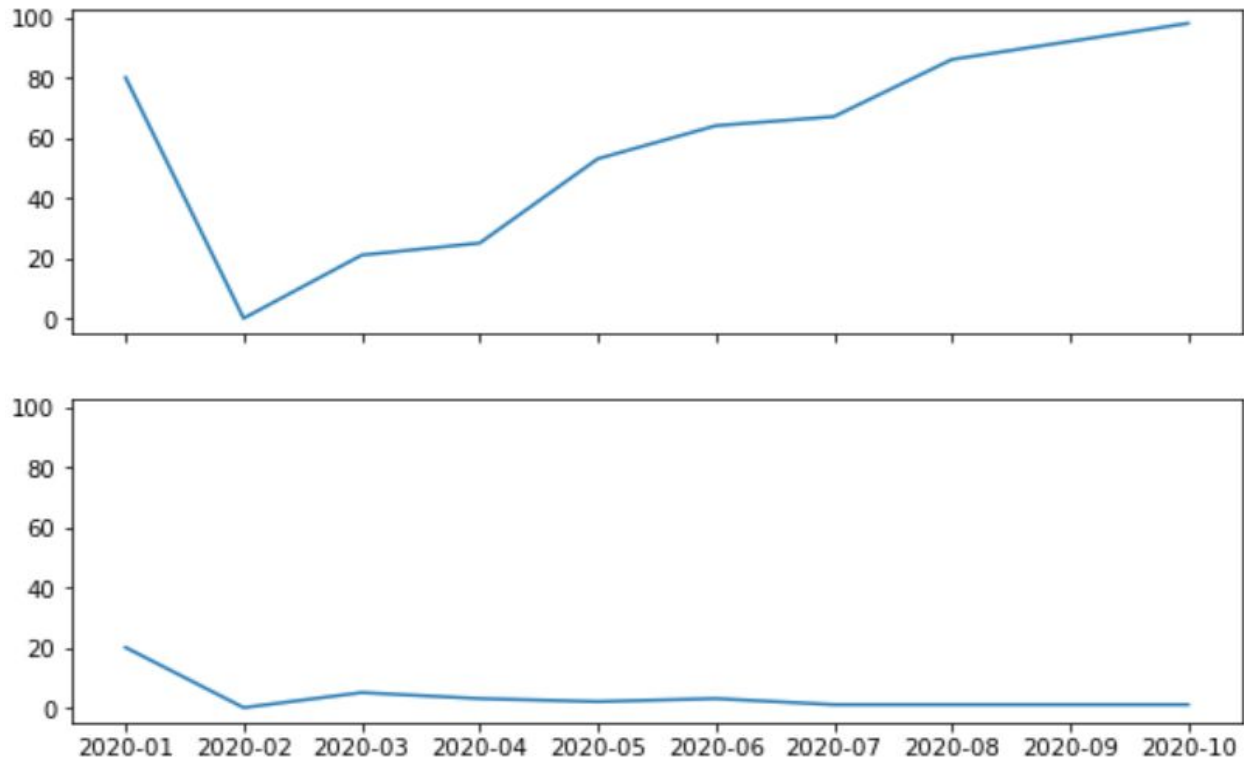
First graph shows cured rate month wise and the 2nd graph shows death rate month wise.

MEASURE OF CENTRAL TENDENCY AND DISPERSION

i) Mean = 2 %

ii) Median = 2.13 % (2-08-2020)

iii) Standard deviation = 1% (very low)

iv) Max= 3.43 % (6-05-2020) , Deviation= 1.43%

v) Range=[0,3.43%]

INTERPRETATION

2% is the death rate of the country on an average which can be considered as the general d.r of our country so far because of the sd val being very low. Despite a sudden upsurge in the no. of cases between august and september our death rate has decreased continuously after 17-06-2020 with max being recorded on 6/5/2020 (only 3.43% with deviation 1.43%). Median value of 2.13% is greater than mean which further gives us a positive outcome that average has gone down because of the continuous decrease in d.r after the upsurge. The cured rate graph also shows that after the big boom we have been able to increase our cured rate.

APT : MEAN

CONCLUSION:

- We used different types of analysis to classify Indian states and UTs on the basis of the various status of COVID-19. Here, grouped 27 states and 5 UTs into six clusters (I−VI). Except the areas under cluster II, all the areas were affected much with COVID-19, where the state Maharashtra has a high number of confirmed cases. Box plot shows variations among different clusters in the three cases. The trend in box plots showed a good trend of cured cases in Jharkhand and Kerala but worst condition of death cases in the state Maharashtra. It was found that the areas under clusters III and VI were required optimization of

monitoring techniques (screening, closedown, curfews, lockdown, evacuations, legal actions etc.) which will be very valuable to the government, doctors, the police and others involved in understanding seriousness of the spread of novel coronavirus (COVID-19) to improve government policies, decisions etc. Areas under clusters I, IV and VI needed more medical facilities (ventilators, testing kits, masks etc.), treatment etc. to reduce the number of deceased persons.