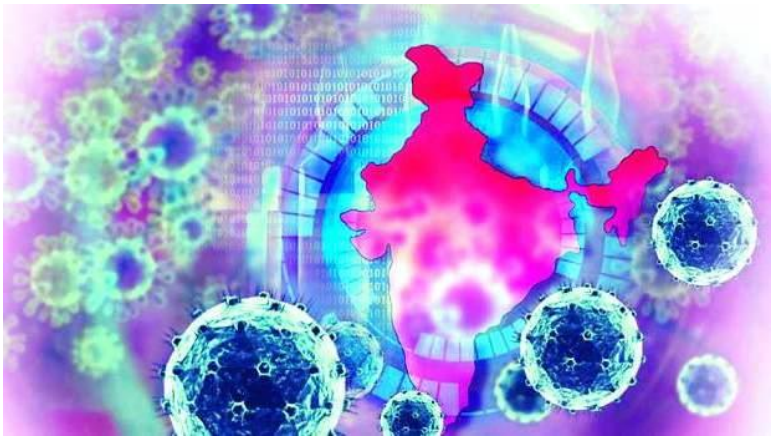
Predictive modelling project

Project report of Exploratory data analysis on Covid data for India



Introduction

This project aims to study the prevailing trends and patterns of the Covid-19 pandemic in India through exploratory data analysis.

The COVID-19 pandemic, also known as the coronavirus pandemic, is an ongoing global pandemic of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The outbreak was first identified in Wuhan, China, in December 2019. The World Health Organization declared the outbreak a Public Health Emergency of International Concern on 30 January 2020, and a pandemic on 11 March.

India currently has the largest number of confirmed cases in Asia, and has the third highest number of confirmed cases in the world after the United States and Brazil.

Table of contents

About the data set

Name of the data set: covid_19_India.csv

Data source: Kaggle (https://www.kaggle.com/sudalairajkumar/covid19-in-india)

Data span: 30th January 2020 to 24th July 2020

Data dictionary:

The data set consists of 4461 observations and 9 columns

Sl.no	Variable name	Variable description
1.	Sl.no	Serial number
2.	Date	Date of observation
3.	Time	Time of observation
4.	State.Union territory	Name of the State / Union territory
5.	Confirmed Indian National	Cumulative number of confirmed Indian nationals
6.	Confirmed Foreign National	Cumulative number of confirmed foreign nationals
7.	Cured	Cumulative number of cured people
8.	Deaths	Cumulative number of death cases
9.	Confirmed	Cumulative number of confirmed cases

Software used for analysis: R studio and Tableau

All the inferences and analysis is limited to 24th July 2020

Descriptive Analytics using R studio

```
rm(list=ls(all=T))

covid<-read.csv(choose.files(),header=TRUE)
```

Installing all the required packages and importing their respective libraries

Viewing the data set

```
View(covid)
```

	X	ï..Sno	Date	Time	State.UnionTerritory	Cured	Deaths	Confirmed
1	1	1	30/01/20	6:00 PM	Kerala	0	0	1
2	2	2	31/01/20	6:00 PM	Kerala	0	0	1
3	3	3	01/02/20	6:00 PM	Kerala	0	0	2
4	4	4	02/02/20	6:00 PM	Kerala	0	0	3
5	5	5	03/02/20	6:00 PM	Kerala	0	0	3
6	6	6	04/02/20	6:00 PM	Kerala	0	0	3
7	7	7	05/02/20	6:00 PM	Kerala	0	0	3
8	8	8	06/02/20	6:00 PM	Kerala	0	0	3
9	9	9	07/02/20	6:00 PM	Kerala	0	0	3
10	10	10	08/02/20	6:00 PM	Kerala	0	0	3
11	11	11	09/02/20	6:00 PM	Kerala	0	0	3
12	12	12	10/02/20	6:00 PM	Kerala	0	0	3
13	13	13	11/02/20	6:00 PM	Kerala	0	0	3
14	14	14	12/02/20	6:00 PM	Kerala	0	0	3
15	15	15	13/02/20	6:00 PM	Kerala	0	0	3

Data dimension

```
dim(covid)
```

The raw data set consists of 9 variables and 4461 observations

Variable names

```
names(covid)
```

The 9 variables are ;

"Sno" "Date" "Time"
"State.UnionTerritory" "ConfirmedIndianNational" "ConfirmedForeignNational"
"Cured" "Deaths" "Confirmed"

Head and Tail of the data set

```
View(head(covid,10))
View(tail(covid,10))
```

Data structure

```
str(covid)
```

```
'data.frame':    4461 obs. of  9 variables:
 $ ï..Sno             : int  1 2 3 4 5 6 7 8 9 10 ...
$ Date                : chr  "30/01/20" "31/01/20" "01/02/20" "02/02/20" ...
$ Time                : chr  "6:00 PM" "6:00 PM" "6:00 PM" "6:00 PM" ...
$ State.UnionTerritory    : chr  "Kerala" "Kerala" "Kerala" "Kerala" ...
$ ConfirmedIndianNational : chr  "1" "1" "2" "3" ...
$ ConfirmedForeignNational: chr  "0" "0" "0" "0" ...
$ Cured               : int  0 0 0 0 0 0 0 0 0 0 ...
$ Deaths              : int  0 0 0 0 0 0 0 0 0 0 ...
$ Confirmed             : int  1 1 2 3 3 3 3 3 3 3 ...
```

The raw data set has 5 ‘character’ variables and 2 ‘integer’ variables

Excluding columns 'Confirmed Indian national' and 'Confirmed foreign national' since they were applicable when 13-14 people came from italy and are no longer relevant.

```
covid$ConfirmedIndianNational<-NULL
covid$ConfirmedForeignNational<-NULL
```

```
dim(covid)
```

The data set now has 4461 observations and 7 variables

Checking for missing values in the data

```
sum(is.na(covid))
```

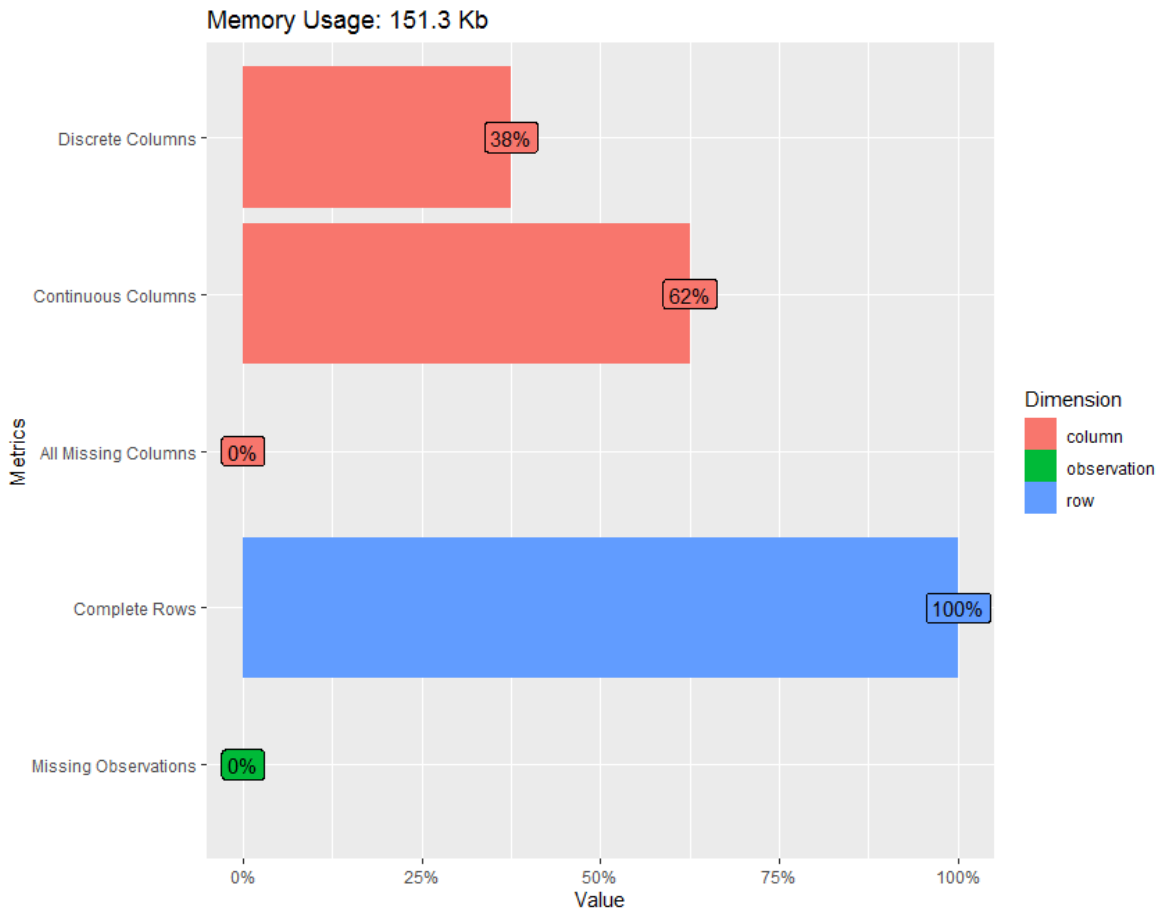
```
[1] 0
```

Checking for column wise missing values, if any

```
colSums(is.na(covid))
```

```
ï..Sno          Date          Time State.UnionTerritory         Cured 
     0             0             0             0             0 
        Deaths     Confirmed 
             0             0 
```

```
plot_intro(covid)
```



Clearly,there are no missing values in the data set.

The above plot also confirms the presence of both discrete (38%) and continuous (62%) Columns in the data set

To get the number of unique values in the column ‘State.UnionTerritory’

```
length(unique(covid[["State.UnionTerritory"]]))
[1] 40
```

Raw data shows the presence of 40 unique states/union territories

Viewing the unique state names

```
as.data.frame(unique(covid$State.UnionTerritory))
```

	unique(covid$State.UnionTerritory)
1	Kerala
2	Telengana
3	Delhi
4	Rajasthan
5	Uttar Pradesh
6	Haryana
7	Ladakh
8	Tamil Nadu
9	Karnataka
10	Maharashtra
11	Punjab
12	Jammu and Kashmir
13	Andhra Pradesh
14	Uttarakhand
15	Odisha
16	Puducherry
17	West Bengal
18	Chhattisgarh
19	Chandigarh
20	Gujarat
21	Himachal Pradesh
22	Madhya Pradesh
23	Bihar
24	Manipur
25	Mizoram
26	Andaman and Nicobar Islands
27	Goa
28	Unassigned
29	Assam
30	Jharkhand
31	Arunachal Pradesh
32	Tripura

33 Nagaland
34 Meghalaya
35 Dadar Nagar Haveli
36 Cases being reassigned to states
37 Sikkim
38 Daman & Diu
39 Dadra and Nagar Haveli and Daman and Diu
40 Telangana

The data set shows the presence of 40 unique states. But from the above output its is clear that due to difference in spellings ‘Telangana’ has been mentioned twice as ‘Telangana’ and ‘Telengana’. Same goes with ‘Dadra and Nagar Haveli and Daman and Diu’.

Dadra and Nagar Haveli and Daman and Diu, (DNHDD) is a union territory in western India.It was created through the merger of the erstwhile union territories of Dadra and Nagar Haveli and Daman and Diu. This merger happened in 2019. Hence I renamed Dadar Nagar Haveli and Daman and Diu, which were 2 separate names in the dataset as ‘Dadra and Nagar Haveli and Daman and Diu’.

Also there were 2 separate categories called ‘unassigned’ and ‘Cases being reassigned to states’ which I renamed as ‘unassigned’

Checking for the summary of the data after making the necessary changes using the describe function

```
library(Hmisc)
describe(covid)
```

```
covid

 7  Variables      4461  Observations
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
i..Sno
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
    4461        0     4461        1     2231     1487      224      447     1116     2231     3346     4015     4238

lowest :    1    2    3    4    5, highest: 4457 4458 4459 4460 4461
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
Date
       n  missing distinct
    4461        0      177

lowest : 01/02/20 01/03/20 01/04/20 01/05/20 01/06/20, highest: 30/05/20 30/06/20 31/01/20 31/03/20 31/05/20
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
Time
       n  missing distinct
    4461        0        7

lowest : 10:00 AM 5:00 PM  6:00 PM  7:30 PM  8:00 AM , highest: 6:00 PM  7:30 PM  8:00 AM  8:30 PM  9:30 PM

Value      10:00 AM  5:00 PM  6:00 PM  7:30 PM  8:00 AM  8:30 PM  9:30 PM
Frequency        27       899       600        56      2823        28        28
Proportion    0.006     0.202     0.134     0.013     0.633     0.006     0.006
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
State.UnionTerritory
       n  missing distinct
    4461        0       36

lowest : Andaman and Nicobar Islands Andhra Pradesh              Arunachal Pradesh           Assam                       Bihar
highest: Tripura                     Unassigned                  Uttar Pradesh               Uttarakhand                 West Bengal
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
Cured
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
    4461        0     1905    0.996     4525     7951        0        0        5      123     2008    10655    20705

lowest :      0      1      2      3      4, highest: 169569 175029 182217 187769 194253
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
Deaths
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
    4461        0      726    0.958    220.1    400.7        0        0        0        3       49      405      983

lowest :     0     1     2     3     4, highest: 11854 12030 12276 12556 12854
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
Confirmed
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
    4461        0     2370        1     7904    13617        1        3       35      470     4080    17437    35451

lowest :      0      1      2      3      4, highest: 310455 318695 327031 337607 347502
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
```

The above output shows a detailed summary of each column in the data set. It shows the number of observations , number of distinct values , number of missing values and some basic descriptive statistics for each column.

The above output tells us that,

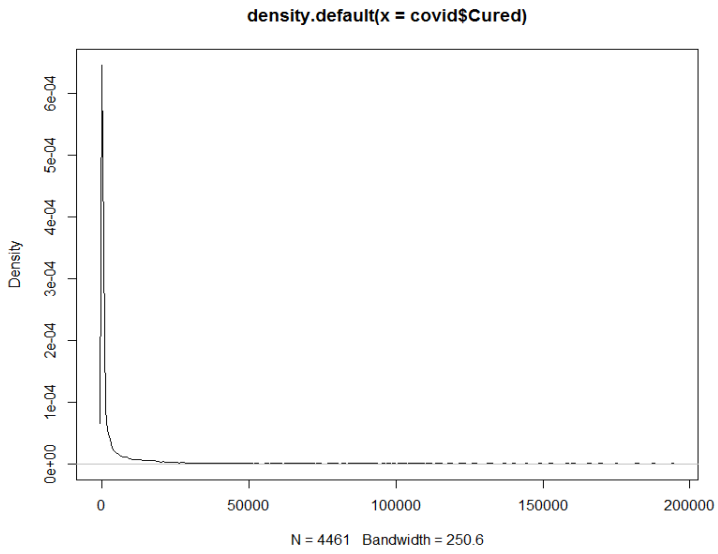
The data consists of 36 distinct states/union territories (including the unassigned category),177 distinct dates and 7 distinct time periods.

From the ‘highest’ of the confirmed,cured and deaths column we can say,

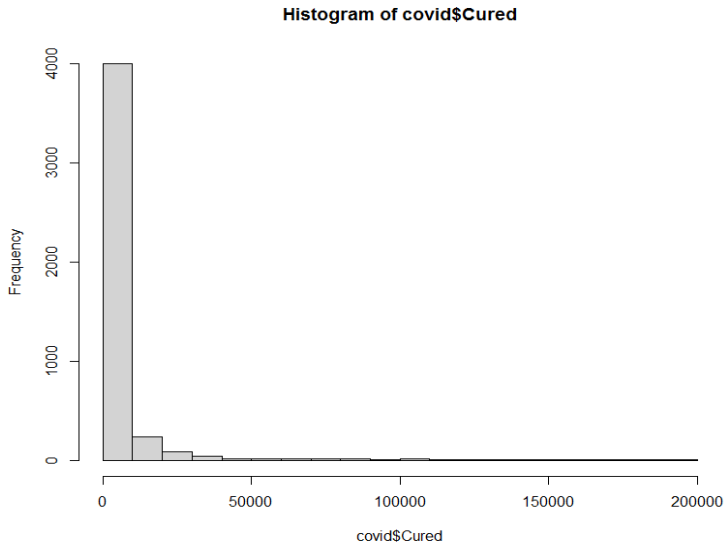
There are a maximum of 347502 confirmed cases,a maximum of 12854 deaths,and a maximum of 194253 cured cases in the data.

Density plot and histogram for all numeric variables

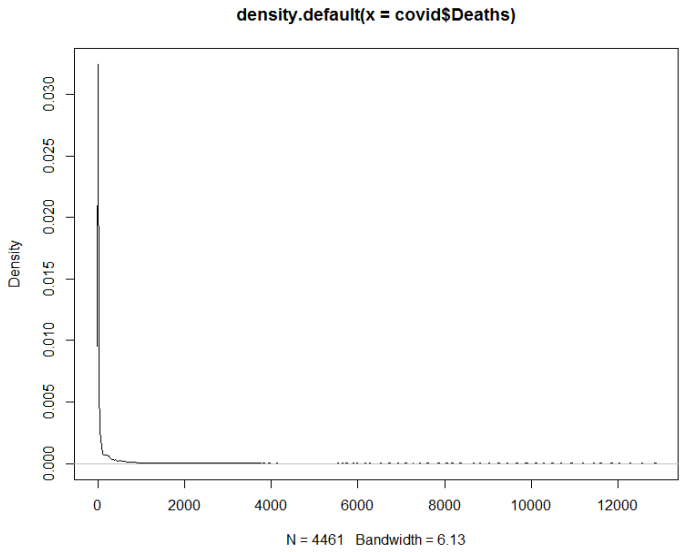
```
library(ggplot2)

c<-density(covid$Cured)
plot(c)
```



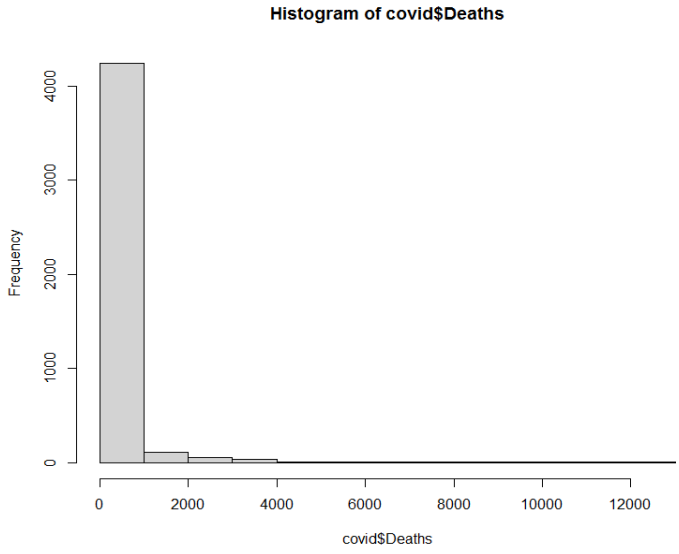

```
hist(covid$Cured)
```



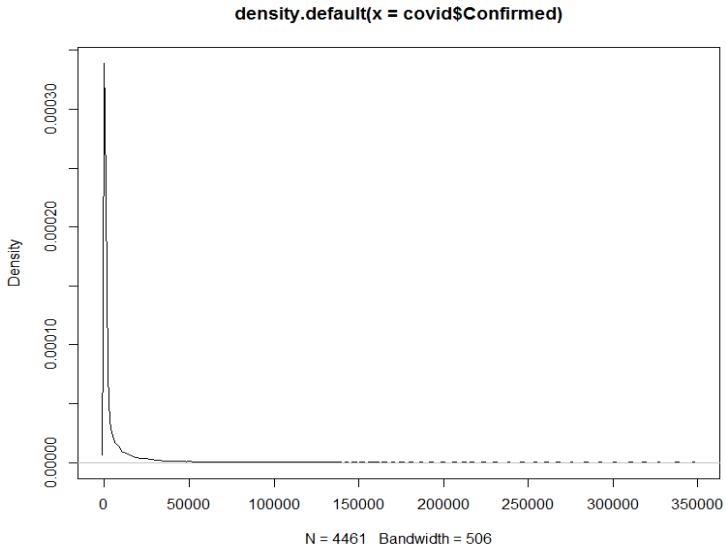
```
d<-density(covid$Deaths)
plot(d)
```



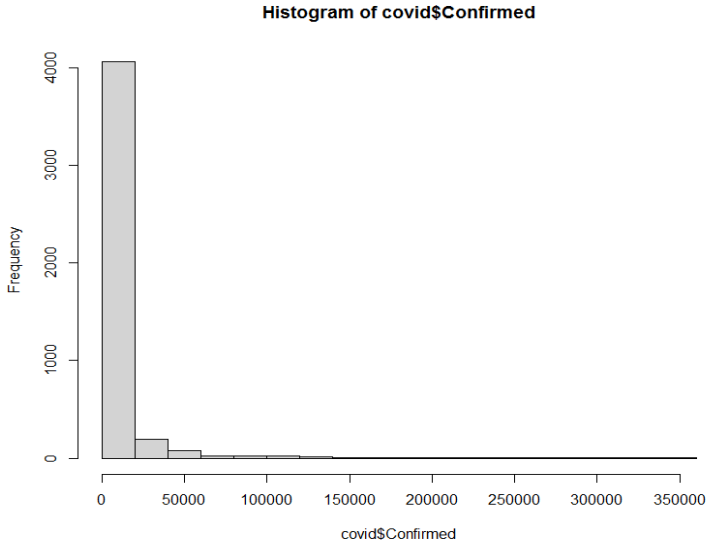

```
hist(covid$Deaths)
```



```
co<-density(covid$Confirmed)
plot(co)
```




```
hist(covid$Confirmed)
```



Skewness co-efficient for all numeric variables

Interpretation : When the skewness co-efficient is <-1 or >+1 then a variable is said to be highly skewed, when it ranges between -1 to -0.5 or 0.5 to 1 then a variable is said to be moderately skewed, when it ranges between -0.5 to 0.5 then a variable is said to be symmetric in its distribution.

```
install.packages("e1071")
library(e1071)

skewness(covid$Cured)
[1] 6.552309
```

The variable cured is right/positively skewed and the same is depicted in the above density plot and histogram

```
skewness(covid$Deaths)
[1] 8.402251
```

The variable Deaths is right/positively skewed and the same is depicted in the above density plot and histogram

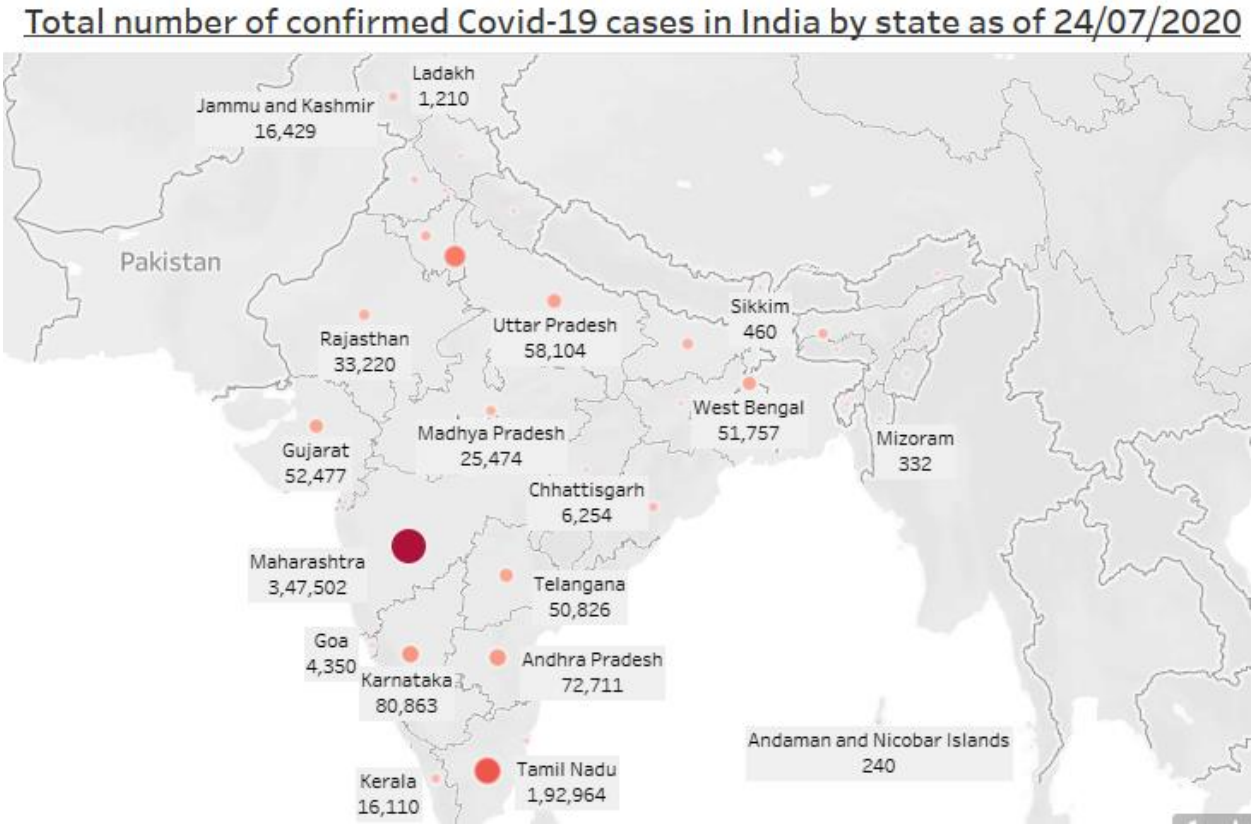
```
skewness(covid$Confirmed)
[1] 6.843354
```

The variable Confirmed is right/positively skewed and the same is depicted in the above density plot and histogram

Exporting the corrected data

```
write.csv(covid,"covid_corrected_final.csv")
```

Exploring the corrected data through visuals on Tableau

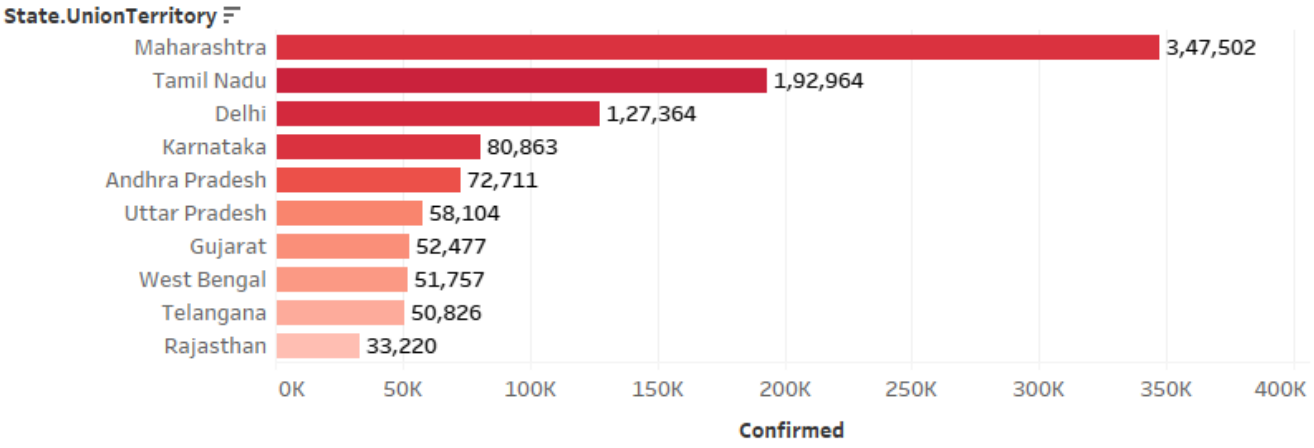


Total number of confirmed Covid-19 cases in India by State/Union Territory

Total number of confirmed cases in India : 12,97,210
Highest number of confirmed cases : Maharashtra (3,47,502)
Lowest number of confirmed cases : Andaman and Nicobar Islands (240)

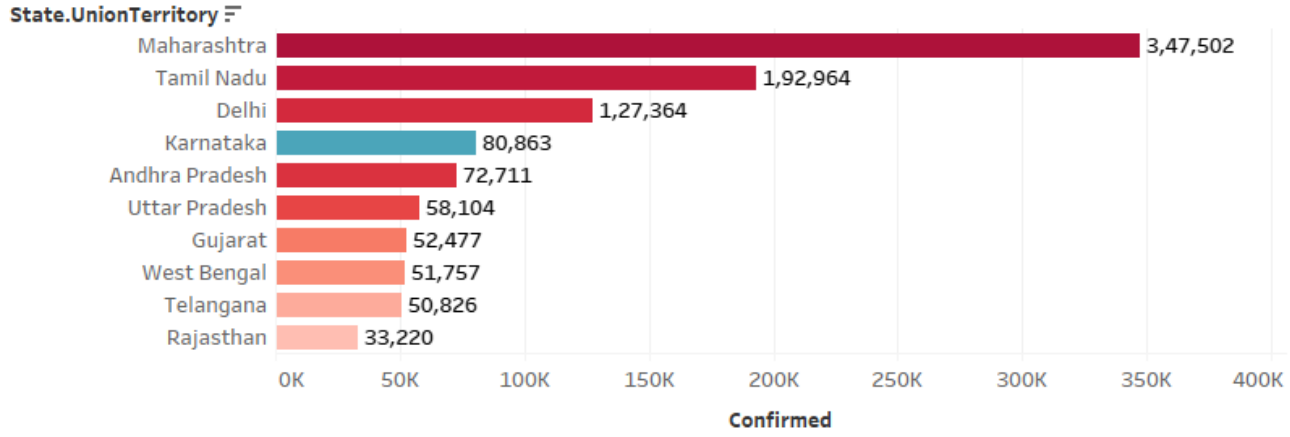
State.UnionTerritory	
Maharashtra	3,47,502
Tamil Nadu	1,92,964
Delhi	1,27,364
Karnataka	80,863
Andhra Pradesh	72,711
Uttar Pradesh	58,104
Gujarat	52,477
West Bengal	51,757
Telangana	50,826
Rajasthan	33,220
Bihar	31,980
Haryana	28,975
Assam	28,791
Madhya Pradesh	25,474
Odisha	21,099
Jammu and Kashmir	16,429
Kerala	16,110
Punjab	11,739
Unassigned	9,265
Jharkhand	6,975
Chhattisgarh	6,254
Uttarakhand	5,445
Goa	4,350
Tripura	3,656
Puducherry	2,420
Manipur	2,115
Himachal Pradesh	1,834
Ladakh	1,210
Nagaland	1,174
Arunachal Pradesh	991
Chandigarh	800
Dadra and Nagar Haveli and Da..	770
Meghalaya	534
Sikkim	460
Mizoram	332
Andaman and Nicobar Islands	240
Grand Total	12,97,210

Top 10 Indian States/Union Territories by the total number of confirmed Covid-19 cases



Where does Karnataka stand when it comes to the total number of confirmed Covid-19 cases?

Karnataka is at the 4th position as of 24/07/2020



Total number of deaths due to Covid-19 in India by State/Union Territory

Out of the 12,97,210 confirmed cases there have been 30,601 deaths in India,Maharashtra reporting the highest number of deaths

Case fatality rate in India as of 24/07/2020 - 2.4%
Mortality rate in India as of 24/07/2020 - 0.0022%

State.UnionTerritory	
Maharashtra	12,854
Delhi	3,745
Tamil Nadu	3,232
Gujarat	2,252
Karnataka	1,616
Uttar Pradesh	1,289
West Bengal	1,255
Andhra Pradesh	884
Madhya Pradesh	780
Rajasthan	594
Telangana	447
Haryana	378
Jammu and Kashmir	282
Punjab	277
Bihar	217
Odisha	114
Assam	70
Jharkhand	67
Uttarakhand	60
Kerala	50
Puducherry	34
Chhattisgarh	30
Goa	29
Chandigarh	13
Himachal Pradesh	11
Tripura	10
Meghalaya	4
Arunachal Pradesh	3
Dadra and Nagar Haveli and Daman and Diu	2
Ladakh	2
Andaman and Nicobar Islands	0
Manipur	0
Mizoram	0
Nagaland	0
Sikkim	0
Unassigned	0
Grand Total	30,601

Case fatality rate

Case fatality rate (CFR) is the proportion of the number of deaths divided by the number of confirmed patients of a disease.

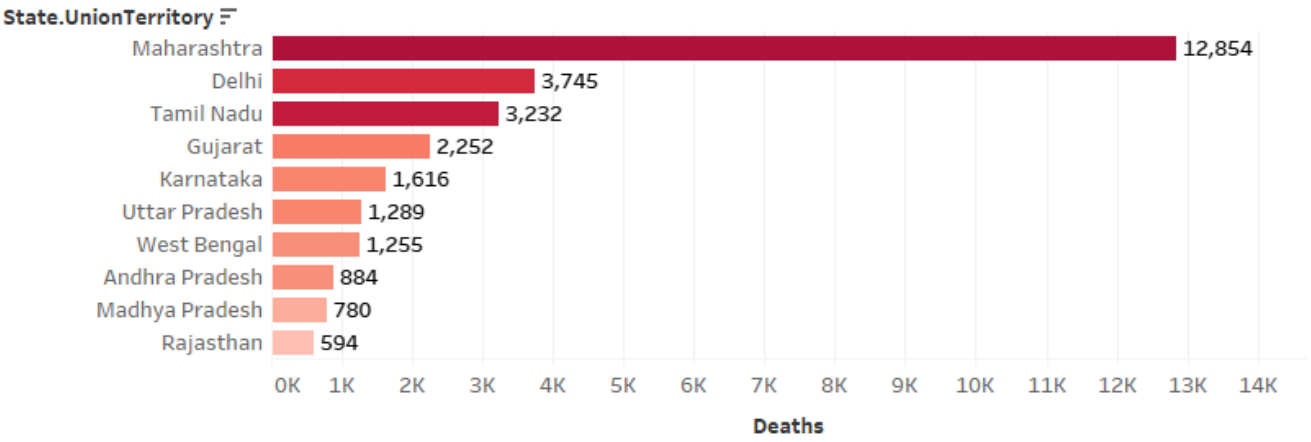
Interpretation : Out of 100 people who have been infected with coronavirus, 2.4 (or 2) people have died.
Expressing case fatality rate out of 100000 (since the number of confirmed cases are in lakhs) - out of 100000 people who have been infected with coronavirus 2,359 people have died.

Mortality rate

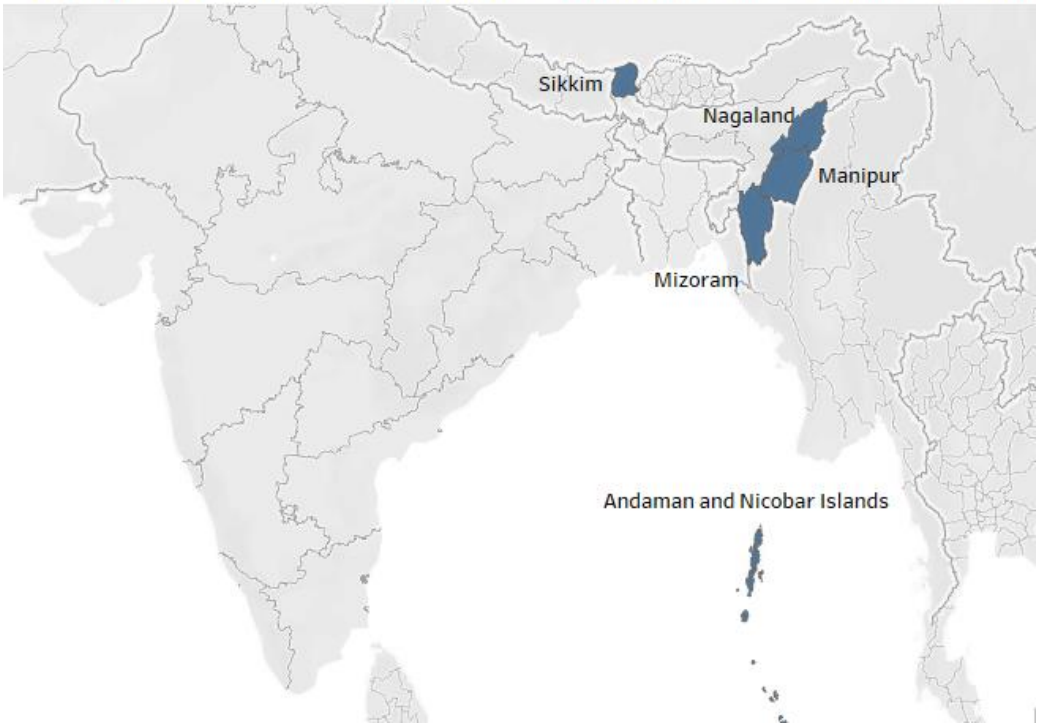
Mortality rate is the proportion of the number of deaths divided by the total population at risk (the entire population of India in our case)

Interpretation : Out of 100 people of the total population at risk of coronavirus in India 0.0022 have died of coronavirus.

Top 10 states by the number of reported deaths



Indian states that have reported zero deaths due to Covid-19 so far



State-wise case fatality rate

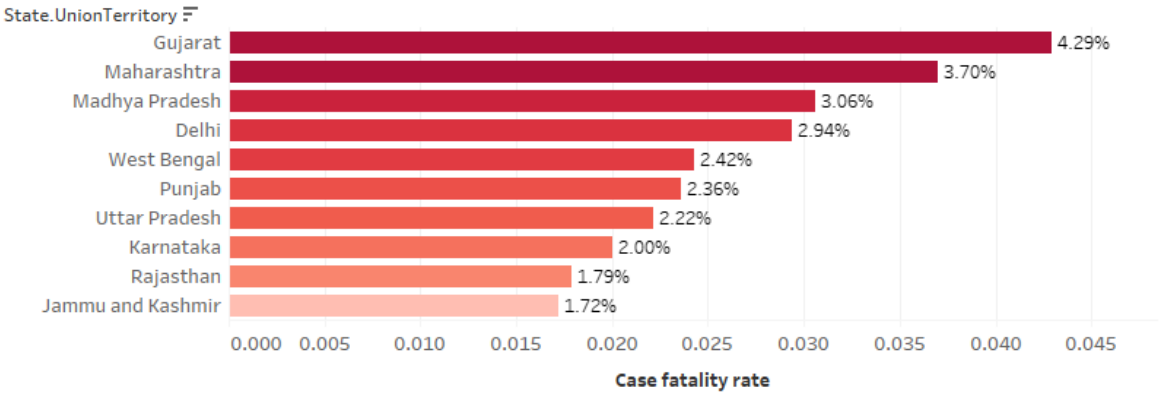
Gujarat has the highest case fatality rate (4.29%) and Ladakh has the least (0.17%)

State.UnionTerritory	
Gujarat	4.29%
Maharashtra	3.70%
Madhya Pradesh	3.06%
Delhi	2.94%
West Bengal	2.42%
Punjab	2.36%
Uttar Pradesh	2.22%
Karnataka	2.00%
Rajasthan	1.79%
Jammu and Kashmir	1.72%
Tamil Nadu	1.67%
Chandigarh	1.63%
Puducherry	1.40%
Haryana	1.30%
Andhra Pradesh	1.22%
Uttarakhand	1.10%
Jharkhand	0.96%
Telangana	0.88%
Meghalaya	0.75%
Bihar	0.68%
Goa	0.67%
Himachal Pradesh	0.60%
Odisha	0.54%
Chhattisgarh	0.48%
Kerala	0.31%
Arunachal Pradesh	0.30%
Tripura	0.27%
Dadra and Nagar Haveli and Daman and Diu	0.26%
Assam	0.24%
Ladakh	0.17%
Andaman and Nicobar Islands	0.00%
Manipur	0.00%
Mizoram	0.00%
Nagaland	0.00%
Sikkim	0.00%
Unassigned	0.00%

- Sikkim ,Nagaland, Mizoram, Manipur, Andaman and Nicobar islands have a case fatality rate of 0% since no deaths have occurred in these states so far (as depicted by the previous chart).

- When it comes to death rate ,Gujarat has overtaken Maharashtra, even though.Maharashtra has the highest number of

Top 10 states by case fatality rate (Death rate)



States by the number of recovered Covid-19 cases

India has reported a total of 8,17,209 recovered cases as of 24/07/2020

India's recovery rate as of 24/07/2020 - 63%

State.UnionTerritory	
Maharashtra	1,94,253
Tamil Nadu	1,36,793
Delhi	1,09,065
Telangana	39,327
Gujarat	37,978
Andhra Pradesh	37,555
Uttar Pradesh	35,803
West Bengal	31,656
Karnataka	29,310
Rajasthan	23,815
Haryana	22,249
Bihar	20,769
Assam	20,699
Madhya Pradesh	17,359
Odisha	14,393
Jammu and Kashmir	8,709
Punjab	7,741
Kerala	6,594
Chhattisgarh	4,377
Uttarakhand	3,399
Jharkhand	3,174
Goa	2,655
Tripura	2,072
Manipur	1,466

Out of the total number of people who have been infected by coronavirus (1297210), **63% have recovered.**

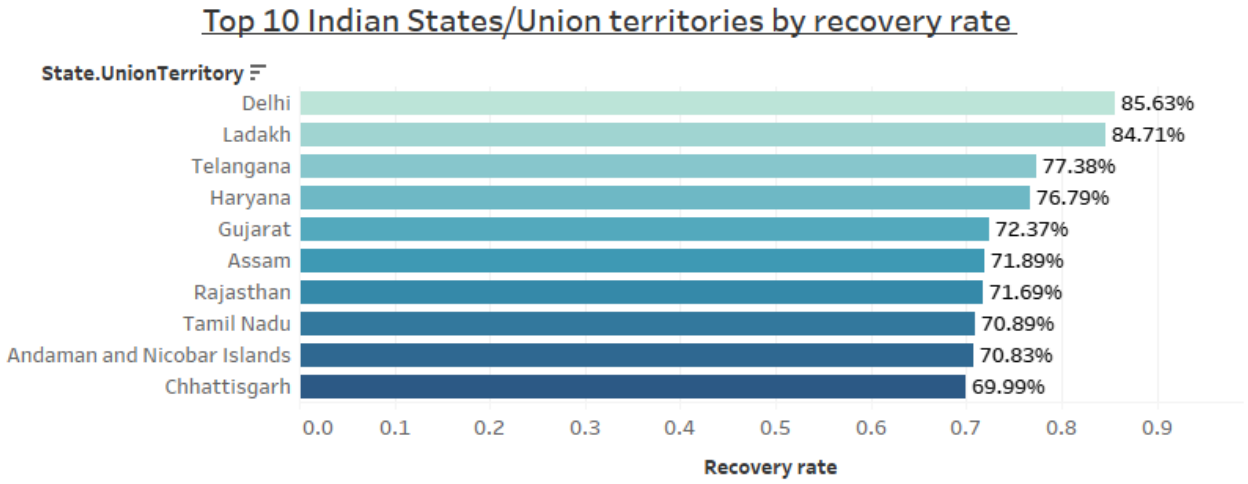
Puducherry	1,400
Himachal Pradesh	1,136
Ladakh	1,025
Chandigarh	531
Nagaland	530
Dadra and Nagar Haveli and Daman and ..	489
Arunachal Pradesh	334
Mizoram	183
Andaman and Nicobar Islands	170
Sikkim	122
Meghalaya	78
Unassigned	0
Grand Total	**8,17,209**

Indian states in order of their recovery rate as of 24/07/2020

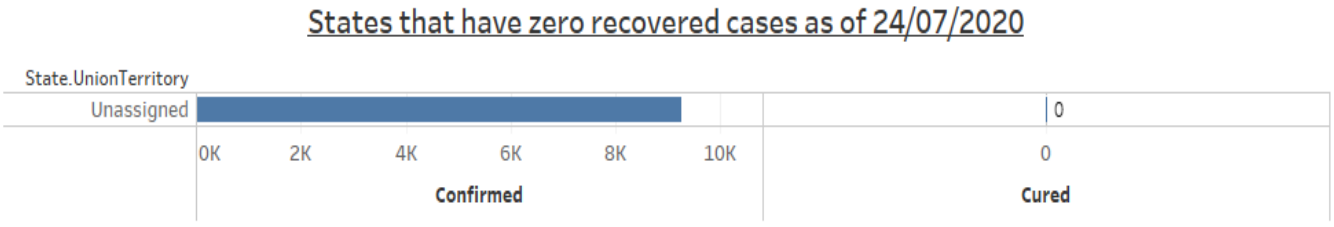
Delhi has the highest recovery rate (86%) and Meghalaya has the least (15%)

State.UnionTerritory	Confirmed	Cured	Recovery rate
Delhi	127,364	109,065	85.63%
Ladakh	1,210	1,025	84.71%
Telangana	50,826	39,327	77.38%
Haryana	28,975	22,249	76.79%
Gujarat	52,477	37,978	72.37%
Assam	28,791	20,699	71.89%
Rajasthan	33,220	23,815	71.69%
Tamil Nadu	192,964	136,793	70.89%
Andaman and Nicobar Islands	240	170	70.83%
Chhattisgarh	6,254	4,377	69.99%
Manipur	2,115	1,466	69.31%
Odisha	21,099	14,393	68.22%
Madhya Pradesh	25,474	17,359	68.14%
Chandigarh	800	531	66.38%
Punjab	11,739	7,741	65.94%
Bihar	31,980	20,769	64.94%
Dadra and Nagar Haveli and Daman and Diu	770	489	63.51%
Uttarakhand	5,445	3,399	62.42%
Himachal Pradesh	1,834	1,136	61.94%
Uttar Pradesh	58,104	35,803	61.62%
West Bengal	51,757	31,656	61.16%
Goa	4,350	2,655	61.03%
Puducherry	2,420	1,400	57.85%
Tripura	3,656	2,072	56.67%

Maharashtra	347,502	194,253	55.90%
Mizoram	332	183	55.12%
Jammu and Kashmir	16,429	8,709	53.01%
Andhra Pradesh	72,711	37,555	51.65%
Jharkhand	6,975	3,174	45.51%
Nagaland	1,174	530	45.14%
Kerala	16,110	6,594	40.93%
Karnataka	80,863	29,310	36.25%
Arunachal Pradesh	991	334	33.70%
Sikkim	460	122	26.52%
Meghalaya	534	78	14.61%
Unassigned	9,265	0	0.00%



Even though Delhi is among the top ten states (3rd) by the number of confirmed cases ,it ranks **top in terms of its recovery rate.**



All states have at least one recovery. There is no state where the recoveries are 0 except for those observations in the data that belong to the unassigned category

I subtracted the total number of cured cases from the total number of confirmed cases to obtain the total number of active cases in India.

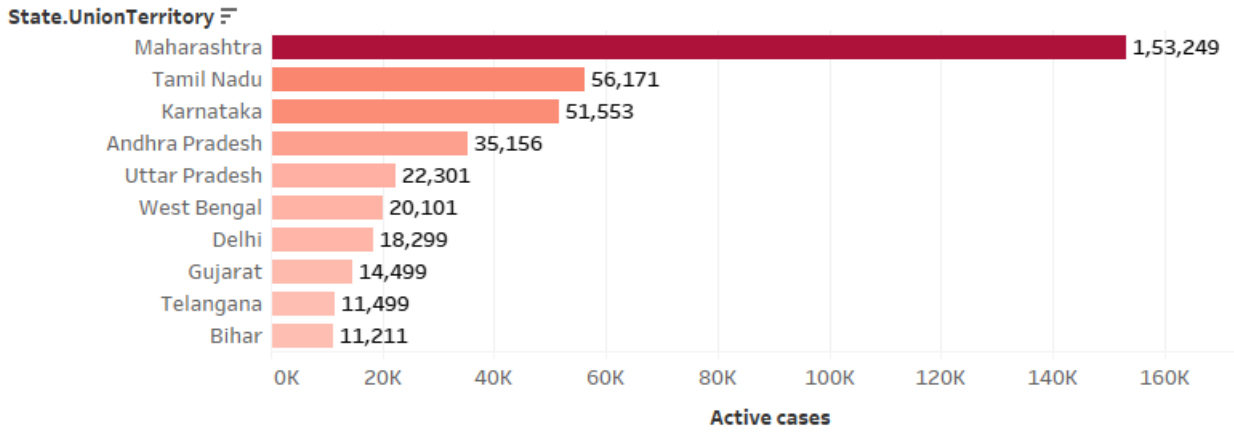
Total number of active cases by state/Union territory

Maharashtra has the highest number of active cases(1,53,249) and Andaman and Nicobar Islands the least(70)

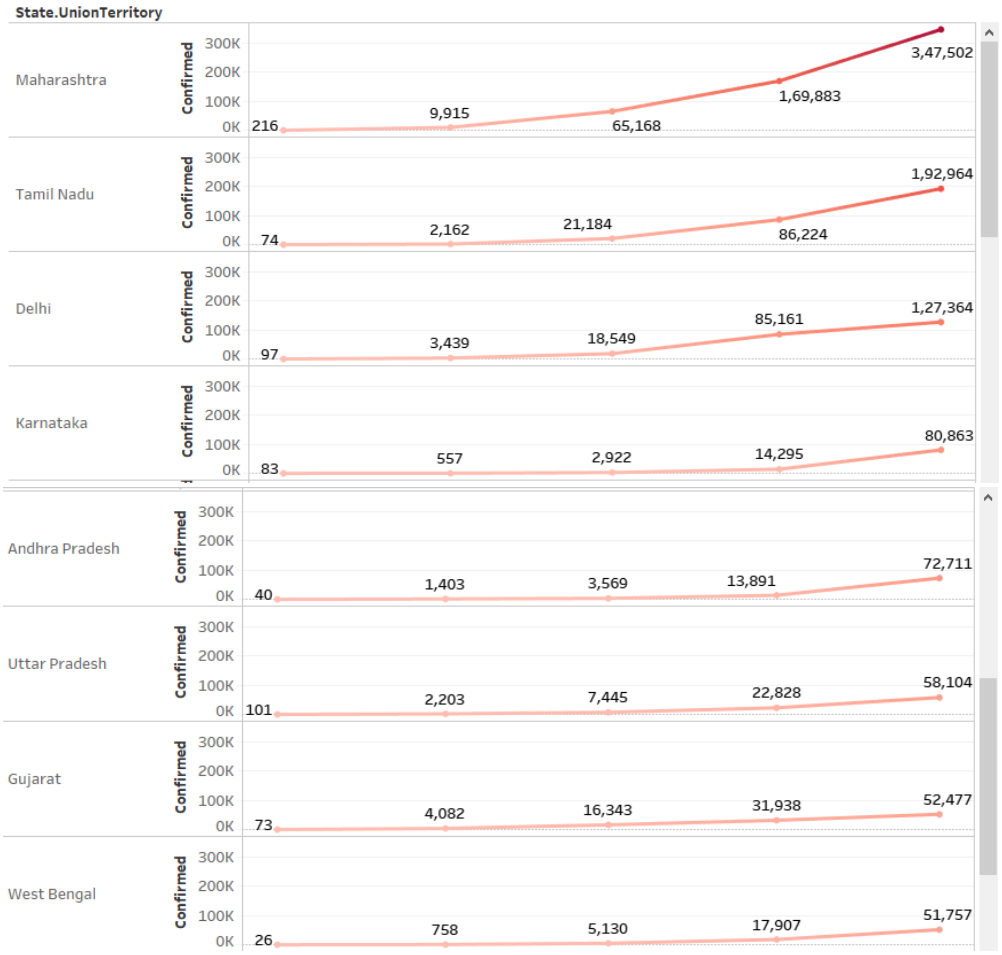
State.UnionTerritory	
Maharashtra	1,53,249
Tamil Nadu	56,171
Karnataka	51,553
Andhra Pradesh	35,156
Uttar Pradesh	22,301
West Bengal	20,101
Delhi	18,299
Gujarat	14,499
Telangana	11,499
Bihar	11,211
Kerala	9,516
Rajasthan	9,405
Unassigned	9,265
Madhya Pradesh	8,115
Assam	8,092
Jammu and Kashmir	7,720
Haryana	6,726
Odisha	6,706
Punjab	3,998
Jharkhand	3,801
Uttarakhand	2,046
Chhattisgarh	1,877
Goa	1,695
Tripura	1,584
Puducherry	1,020
Himachal Pradesh	698
Arunachal Pradesh	657
Manipur	649
Nagaland	644
Meghalaya	456
Sikkim	338
Dadra and Nagar Haveli and Daman and Diu	281
Chandigarh	269
Ladakh	185
Mizoram	149
Andaman and Nicobar Islands	70
Grand Total	**4,80,001**

- There are 4,80,001 active cases in India as of 24/07/2020
- The total number of active cases in India are 4,80,001 and the total number of recoveries are 8,17,209. Clearly the number of recoveries are greater than the number of active cases.

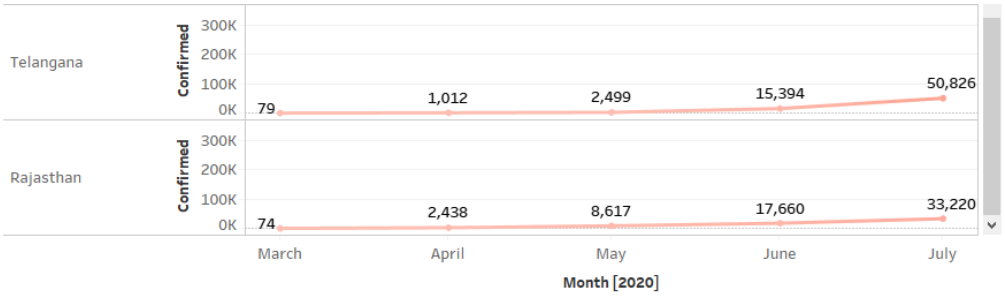
Top 10 States/Union territories by total number of active cases



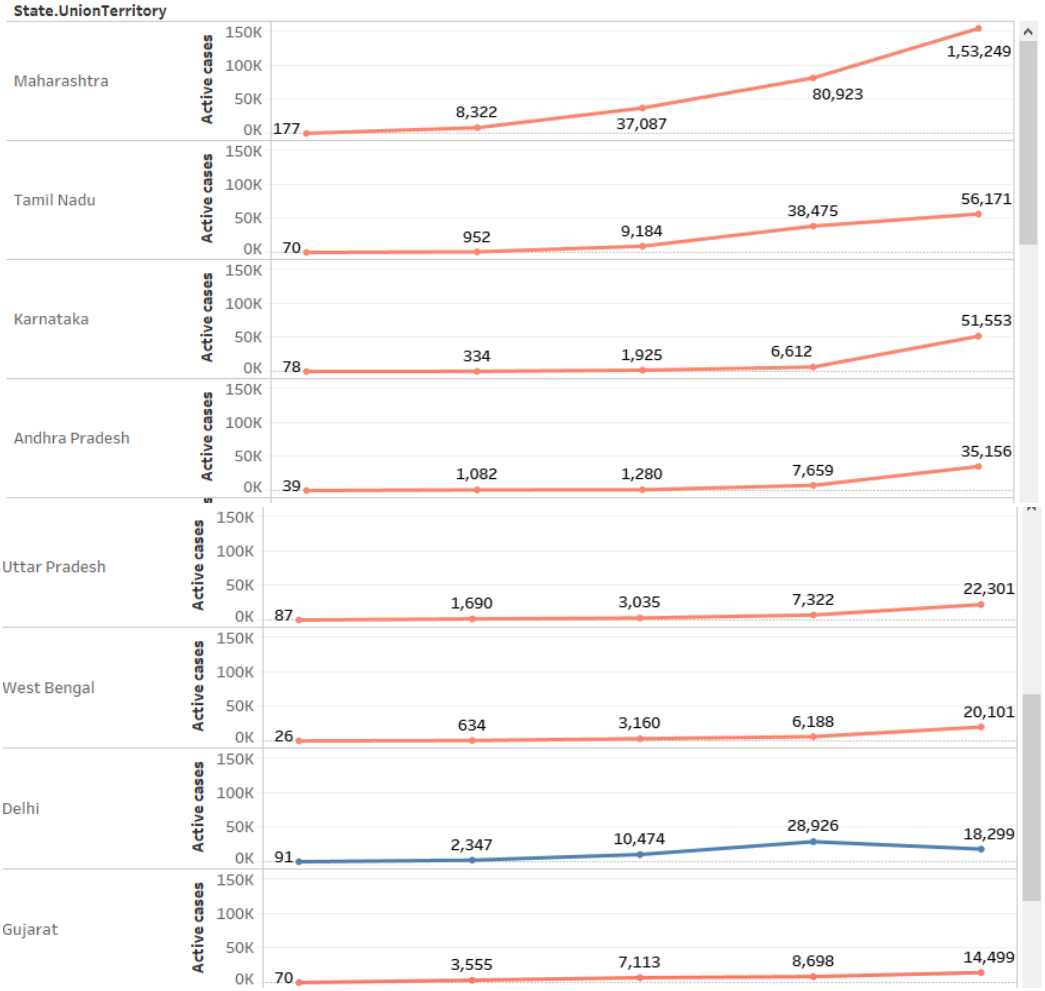
Month-wise trend in the number of confirmed Covid-19 cases in the top 10 states/union territories in India(by most number of cases)



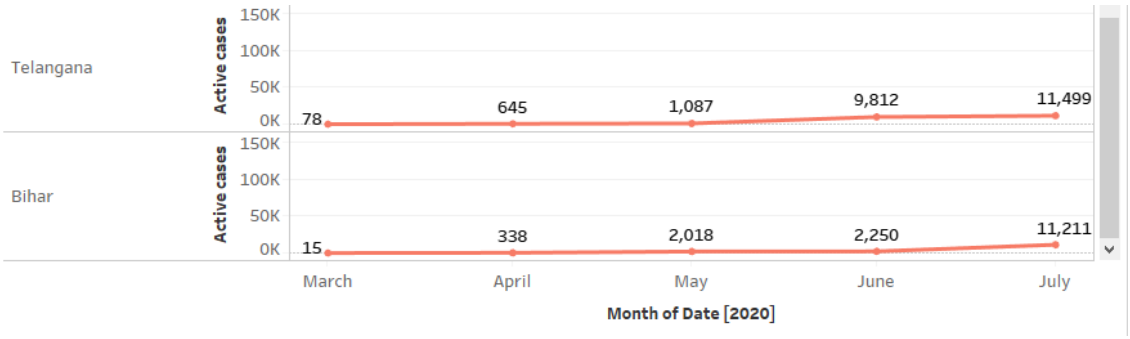
There is a **constant upward trend** in the number of cases across all 10 states.



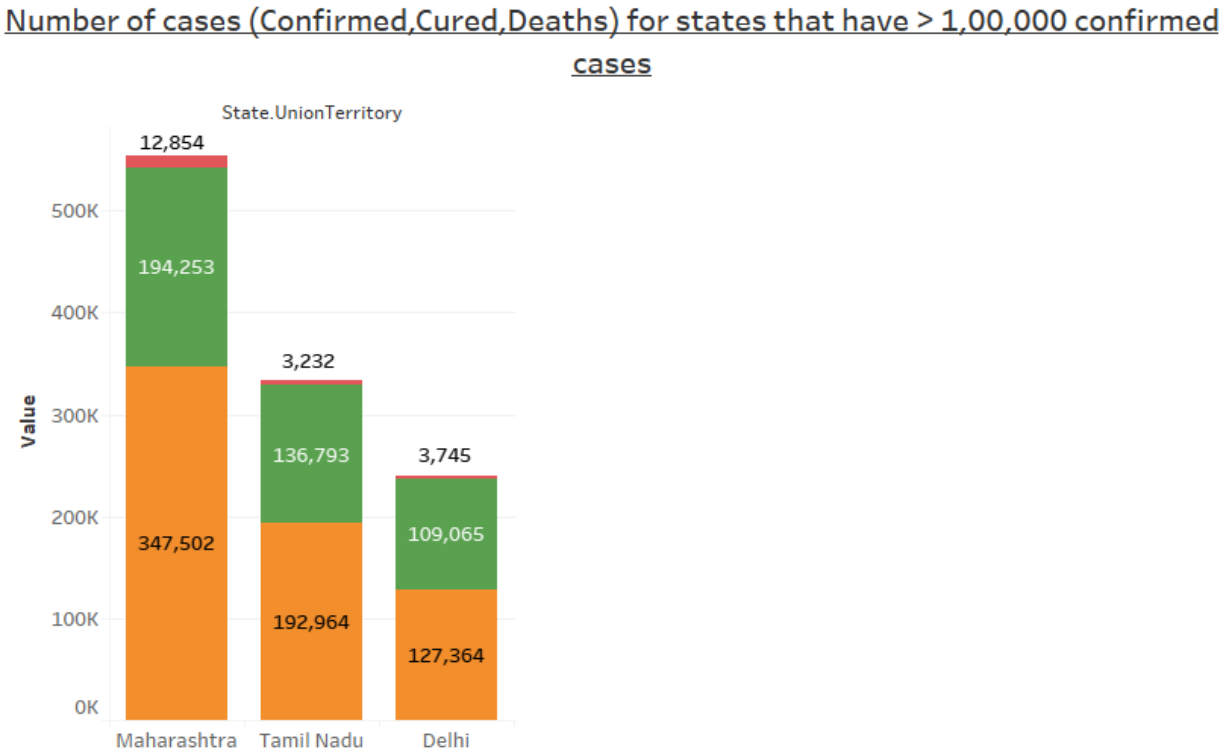
Month-wise trend in the number of active Covid-19 cases in the top 10 states/union territories in India(by the number of active cases)



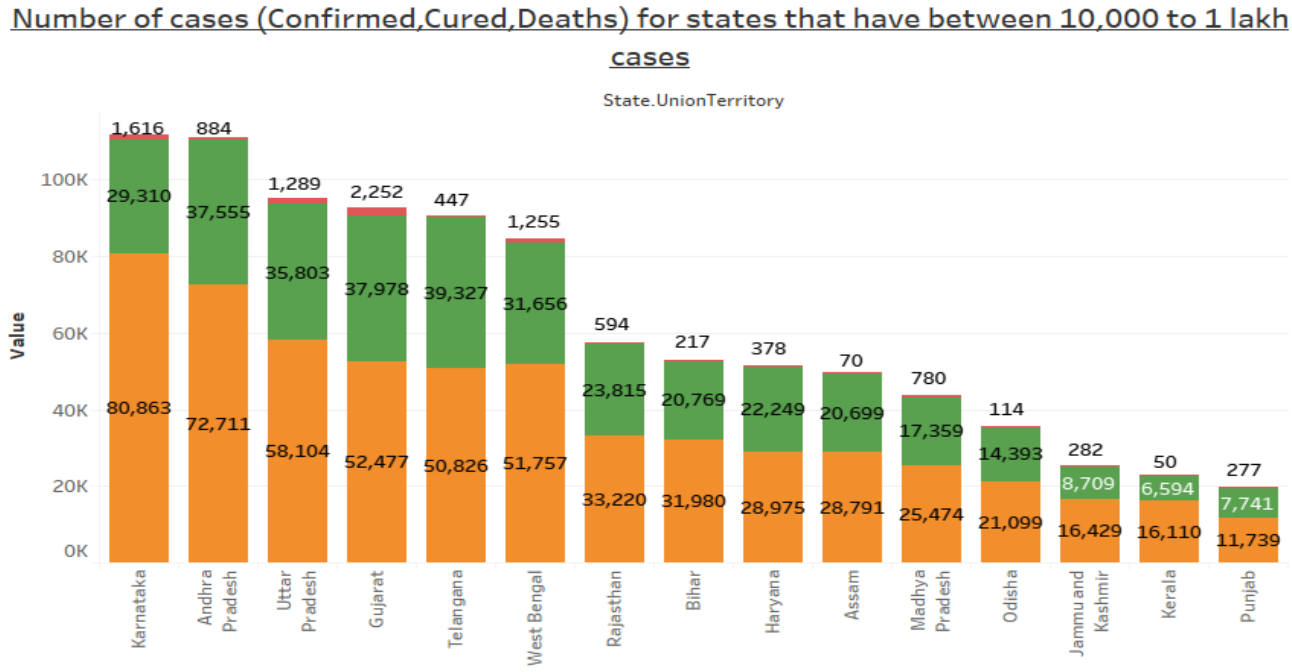
Among the top 10 states by active cases, **Delhi** is the only state which has witnessed a dip in the number of active cases from the month of June (28,926 active cases) to July (18,299 active cases). All other states have only witnessed a rise in these figures.



Which Indian states have crossed the 1 lakh mark in terms of the number of confirmed covid19 cases?

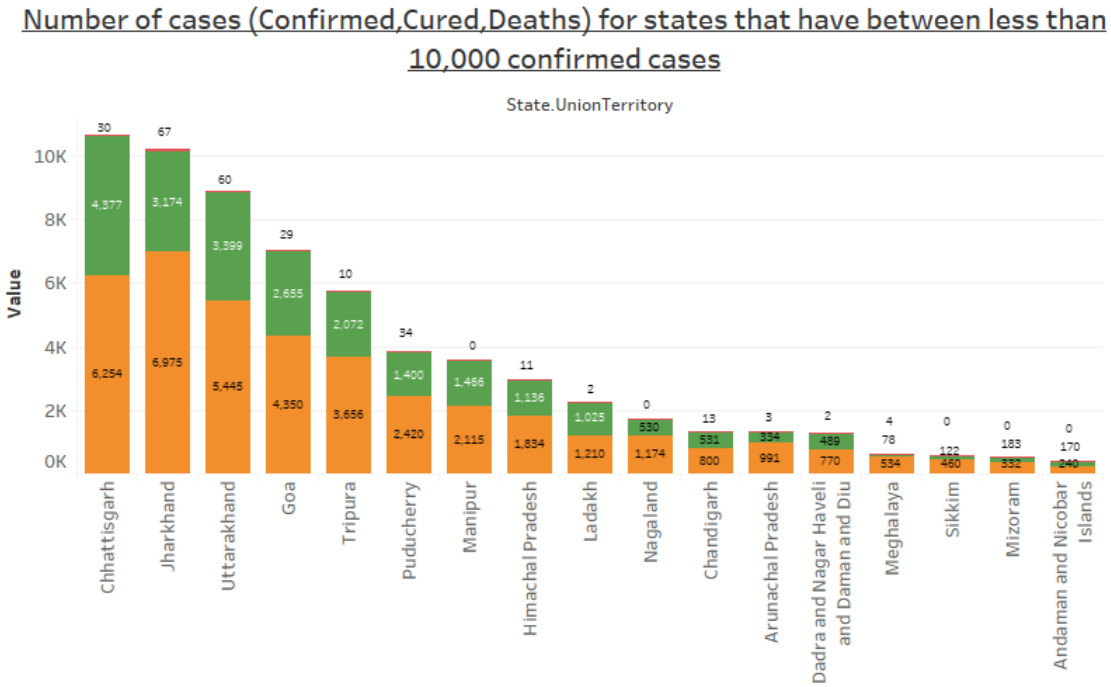


Which Indian states have crossed the 10,000 mark in terms of the number of confirmed covid19 cases?



Karnataka is **less than 20000 cases** (19137) away from crossing the 1 lakh mark.

Which Indian states have less than 10,000 confirmed covid19 cases?



Top 10 states/Union terrirtories by confirmed cases

State.UnionTerritory	Confirmed.	Deaths.	Cured	Active cases
Maharashtra	347,502	12,854	194,253	153,249
Tamil Nadu	192,964	3,232	136,793	56,171
Delhi	127,364	3,745	109,065	18,299
Karnataka	80,863	1,616	29,310	51,553
Andhra Pradesh	72,711	884	37,555	35,156
Uttar Pradesh	58,104	1,289	35,803	22,301
Gujarat	52,477	2,252	37,978	14,499
West Bengal	51,757	1,255	31,656	20,101
Telangana	50,826	447	39,327	11,499
Rajasthan	33,220	594	23,815	9,405
Grand Total	**1,067,788**	**28,168**	**675,555**	**153,249**

- The top 10 states by confirmed cases contribute to a total of 10,67,788 cases out of 12,97,210(the total confirmed cases in India). These states account for **82.31%** of the total confirmed covid19 cases in India.

- The top 10 states by confirmed cases contribute to a total of 28,168 deaths out of 30,601(the total deaths in India). These states account for **92.05%** of the total covid19 deaths in India.

- The top 10 states by confirmed cases contribute to a total of 6,75,555 recoveries out of 8,17,209(the total recoveries in India). These states account for **82.67%** of the total covid19 recoveries in India.

- The top 10 states by confirmed cases contribute to a total of 1,53,249 active cases out of 4,80,001(the total active cases in India). These states account for **32%** of the total active covid19 cases in India.

A comparison of the status of Covid-19 in India

June Vs July

	24 - June - 2020	24 - July - 2020
Confirmed cases	4,57,307	12,97,210
Active cases	1,98,622	4,80,001
Deaths	14,476	30,601
Recoveries	2,58,685	8,17,209
Case fatality rate(Death rate)	3.17%	2.4%
Recovery rate	57%	63%

What has changed over the last one month?

- In a span of one month 8,39,903 new confirmed cases have been reported.
- In a span of one month the number of active cases have increased by 2,81,379
- In a span of one month 16,125 new deaths have been reported.
- In a span of one month 5,58,524 new recoveries have been reported.
- In a span of one month, case fatality rate has decreased by 0.77%
- In a span of one month , recovery rate has increased by 6%

Where does Karnataka stand ?

A comparison of June Vs July

	24 - June - 2020	24 - July - 2020
Confirmed cases	9,721	80,863
Active cases	37,171	51,553
Deaths	150	1,616
Recoveries	6,004	29,310
Case fatality rate(Death rate)	1.54%	2.00%
Recovery rate	61.76%	36.25%

- In a span of one month Karnataka has jumped to the 4th position with 80,863 confirmed cases in July from being on the 11th position with 9,721 confirmed cases in June.
- 71,142 new cases have been confirmed in Karnataka in a span of just one month.
- In a span of one month the number of active cases have increased by 13,836
- In a span of one month 1,466 new deaths have been reported.
- In a span of one month 23,306 new recoveries have been reported.
- Case fatality rate has is pretty much the same with a small increase of 0.46%
- Recovery rate has decreased drastically by 25.51%

The Levitt model

1. Levitt model on cumulative confirmed cases for the state of Maharashtra

Model output

SUMMARY OUTPUT

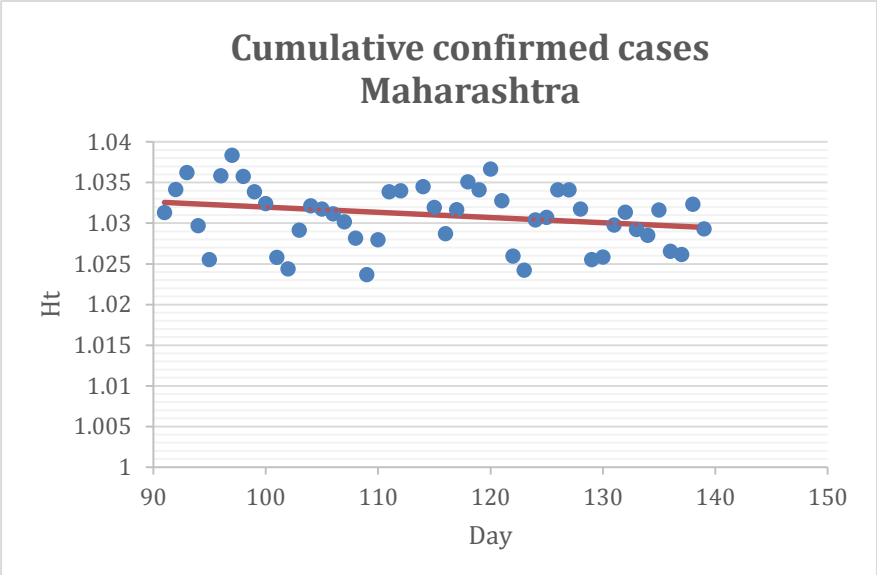
Regression Statistics	
Multiple R	0.232473236
R Square	0.054043806
Adjusted R Square	0.033917078
Standard Error	0.003876145
Observations	49

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	4.03434E-05	4.03434E-05	2.685176	0.107964098
Residual	47	0.000706152	1.50245E-05		
Total	48	0.000746495			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.038402518	0.004536743	228.8872344	2.8E-73	1.029275768	1.047529267	1.029275768	1.047529267
X Variable 1	-0.000064161	3.9155E-05	-1.638650662	0.107964	-0.000142931	1.46083E-05	-0.000142931	1.46083E-05

Scatter plot



Regression equation :

Y=(-0.000064)(X) + 1.0384

Value of X : 584

On the 584th day from the day the first case was reported the number of cases will go to single digit figures.

2. Levitt model on cumulative deaths for the state of Maharashtra

Model output

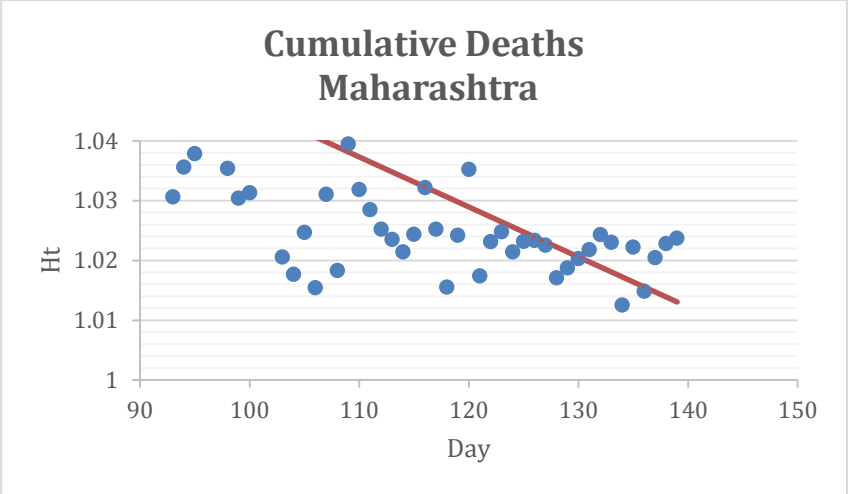
Regression Statistics	
Multiple R	0.260759274
R Square	0.067995399
Adjusted R Square	0.048165514
Standard Error	0.04470375
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.006852471	0.006852471	3.428935589	0.070351037
Residual	47	0.093925986	0.001998425		
Total	48	0.100778457			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1.129285214	0.052322449	21.58318733	4.93639E-26	1.024026024	1.234544404	1.024026024	1.234544404
X Variable 1	-0.000836201	0.000451576	-1.851738532	0.070351037	-0.001744655	7.22531E-05	-0.001744655	7.22531E-05

Scatter plot



Regression equation :

Y=-(-0.0008)(X) + 1.1293

Value of X : 160

On the 160th day from the day the first death was reported the number of deaths will go to single digit figures.

3. Levitt model on cumulative confirmed cases for the state of Karnataka

Model output

SUMMARY OUTPUT

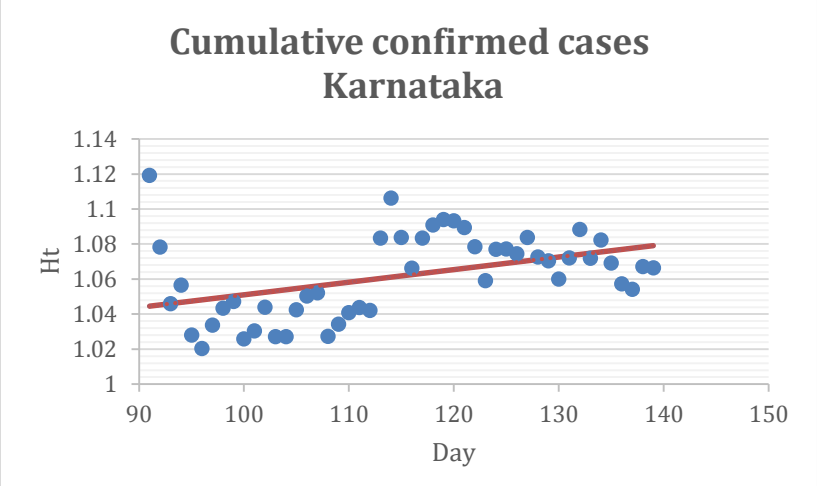
Regression Statistics	
Multiple R	0.435024806
R Square	0.189246581
Adjusted R Square	0.171996509
Standard Error	0.021546654
Observations	49

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.005093271	0.005093271	10.97077006	0.001785739
Residual	47	0.02182014	0.000464258		
Total	48	0.026913411			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.978961925	0.025218772	38.81877778	2.32193E-37	0.928228299	1.029695551	0.928228299	1.029695551
X Variable 1	0.000720917	0.000217654	3.312215279	0.001785739	0.000283054	0.001158781	0.000283054	0.001158781

Scatter plot



Regression equation :

Y= 0.0007(X) + 0.9789

Value of X : 32

On the 32nd day from the day the first case was reported the number of cases will go to single digit figures.

4. Levitt model on cumulative deaths for the state of Karnataka

Model output

SUMMARY OUTPUT

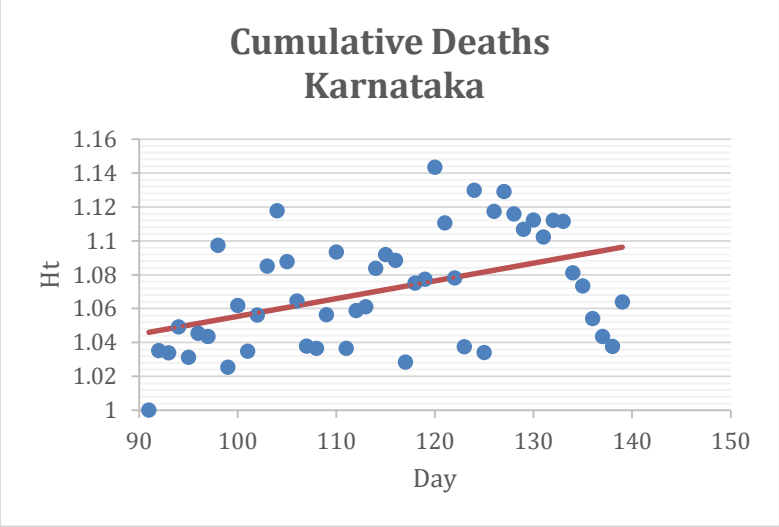
Regression Statistics	
Multiple R	0.440227199
R Square	0.193799987
Adjusted R Square	0.176646795
Standard Error	0.030809737
Observations	49

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.010724691	0.010724691	11.29818808	0.001548747
Residual	47	0.044614275	0.00094924		
Total	48	0.055338965			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.950864014	0.036060529	26.36855394	8.18203E-30	0.878319587	1.023408442	0.878319587	1.023408442
X Variable 1	0.001046115	0.000311225	3.361277745	0.001548747	0.00042001	0.001672219	0.00042001	0.001672219

Scatter plot



Regression equation :

Y= 0.0010(X) + 0.9509

Value of X : 50

On the 50th day from the day the first death was reported the number of deaths will go to single digit figures.