

# COVID DATA ANALYTICS CASE STUDY ON INDIA

---



SUMMER  
OF MACHINE  
LEARNING  
AT SKOLTECH

*Vaibhav Anand*

*Vineet Kumar*

Indian Statistical Institute

[contact@covid-isical.tech](mailto:contact@covid-isical.tech)

# COVID-19 AND INDIA

---

- First COVID-19 case: 30<sup>th</sup> January, 2020
- First fatality: 12<sup>th</sup> March, 2020
- Current status (19/08/2020):
  - Confirmed: 2,834,755 (+68,129 in a day)
  - Active: 685,037 (+8,658 in a day)
  - Recovered: 2,095,211 (+58,508 in a day)
  - Deceased: 53,978 (+963 in a day)
  - Total tests: 30,941,264
  - Proportion positive amongst tested (current week): 7.22%

Heatmap

# **DASHBOARD**



# COVID Dashboard

*This will aid in unlock strategy design and intervention monitoring*

India | States | Comparison | Districts | FAQs | Disclaimers and Explanations | Contact Us

Confirmed	Active	Recovered	Deceased	Cumulative Tested	Proportion Positive (this wk)	Red Zones	Orange Zones	Green Zones
2834755 +68129	685037 +8658	2095211 +58508	53978 +963	30941264	7.22%	130	326	277

## Risk Summary

Data	
ICU Beds	32420.00
Oxygen Supported Beds	119637.00
ICU Beds/ 100 Active Cases	4.79
O2 Beds/ 100 Active Cases	17.68
Avg. Traffic Intensity of Ongoing Week *	1.08
Weekly Avg. Confirmed Cases	64999.00
Weekly Avg. Recovered Cases	60455.00
Weekly Avg. Deceased Cases	1099.00

\*Traffic Intensity is the ratio of number of confirmed cases and number of recoveries in a given period. Values > 1 indicate faster arrival than recovery. Large values beyond 1 would inundate the health care system)

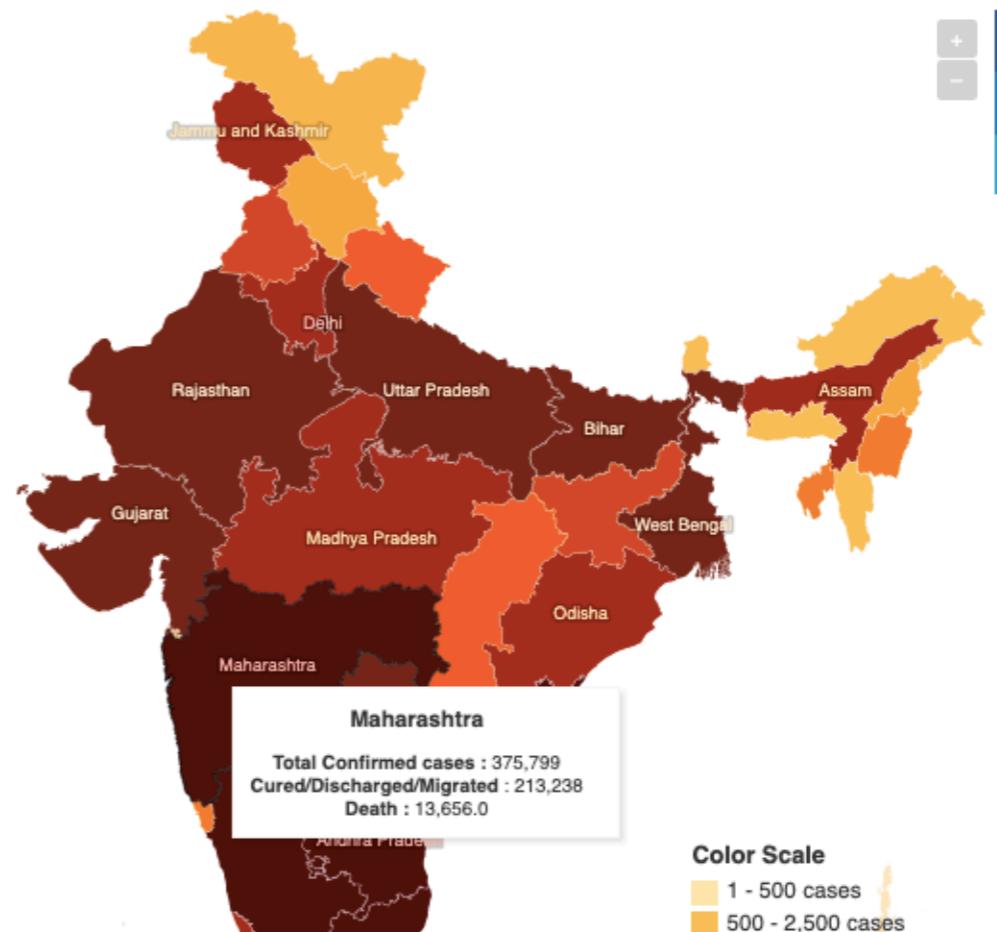
[Click Here](#) to download a risk assessment report.

[Click Here](#) for past reports.

Each week has been denoted by a date falling on the Monday when the week ends. For instance, '08/06/2020' represents the week that ends on '08/06/2020'.

Please refer the FAQs and Disclaimer and Explanations sections to better understand the graphs.

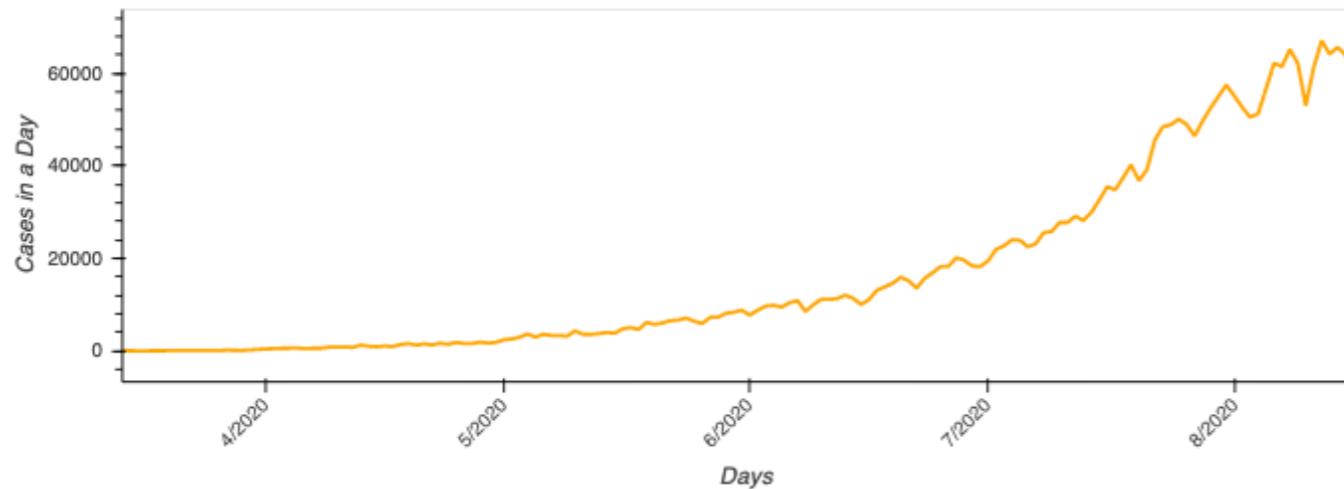
Heatmap of Confirmed Cases



# SPREAD ASSESSMENT

---

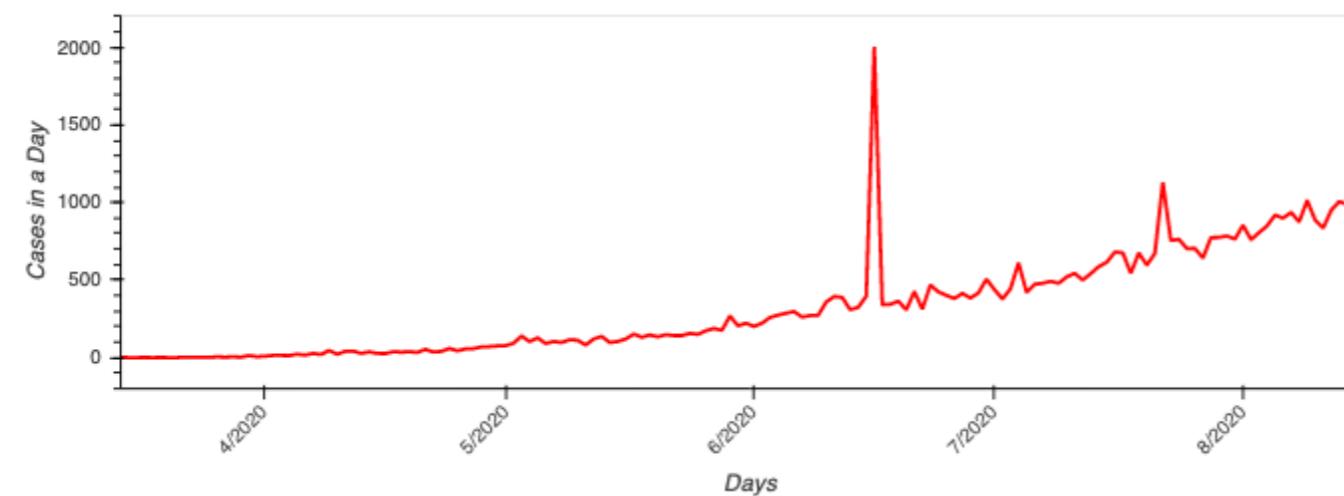
An increasing trend in weekly average and daily confirmed cases indicates spread



*Daily Confirmed Cases*



*Daily Recovered Cases*

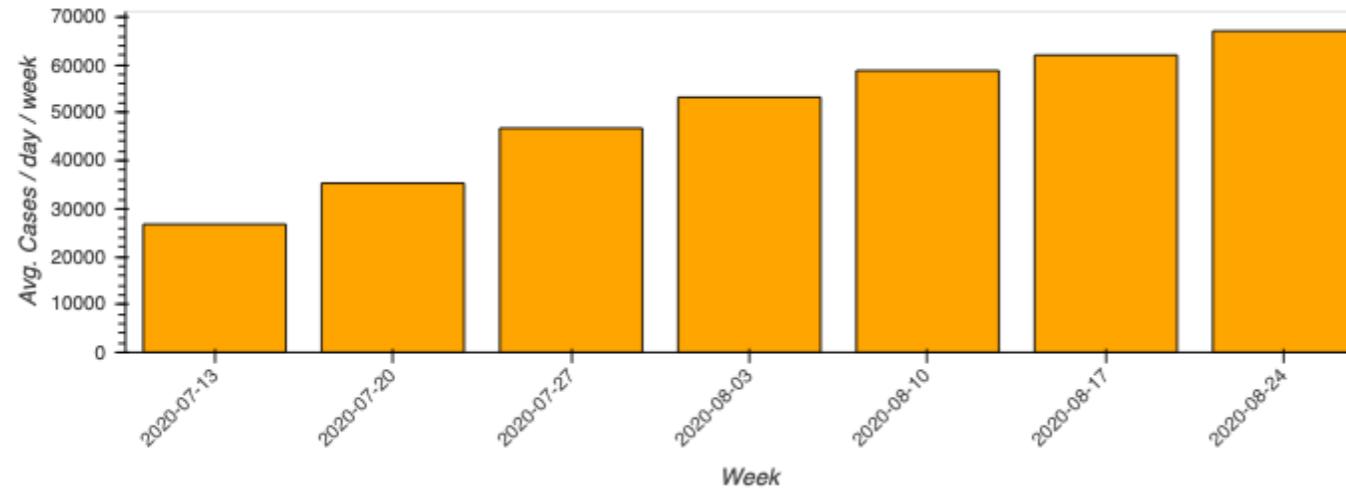


*Daily Deceased Cases*

# SPREAD ASSESSMENT

---

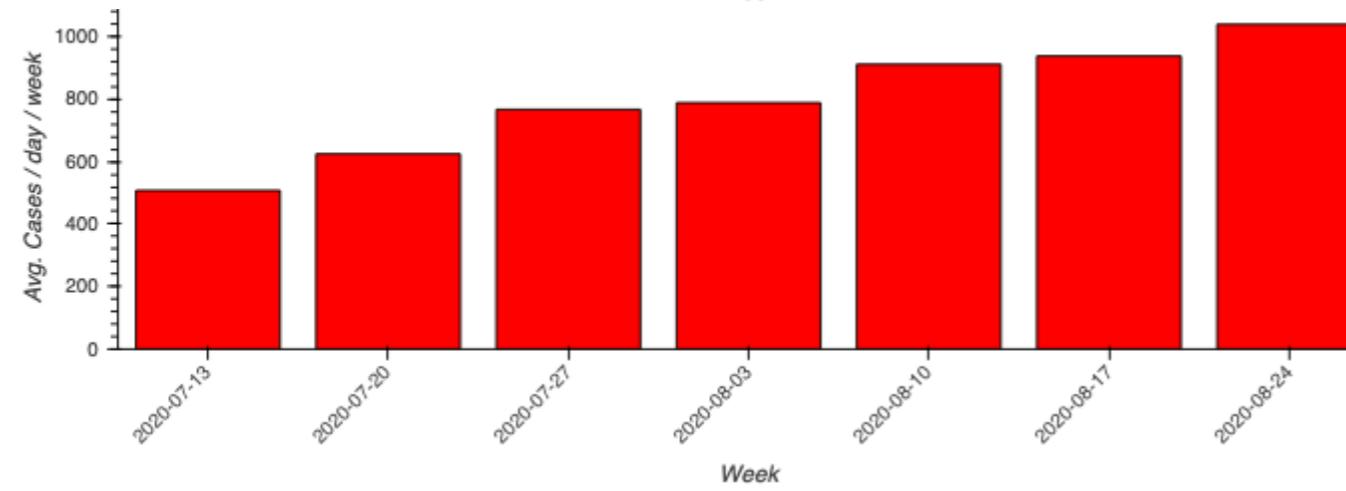
An increasing trend in weekly average and daily confirmed cases indicates spread



*Weekly Confirmed Cases*



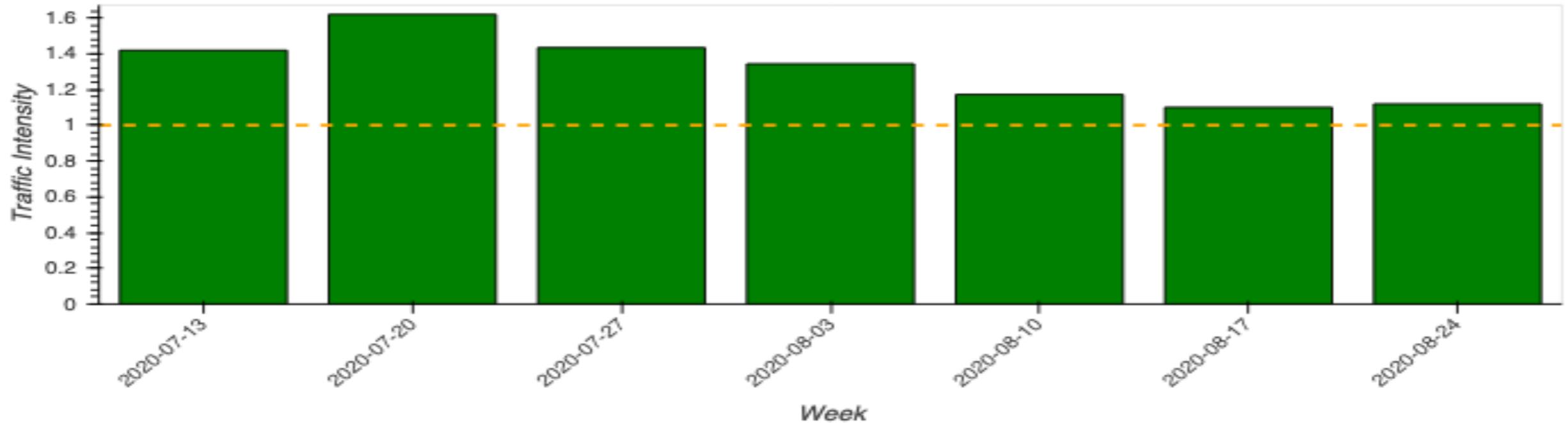
*Weekly Recovered Cases*



*Weekly Deceased Cases*

# RISK ASSESSMENT

---



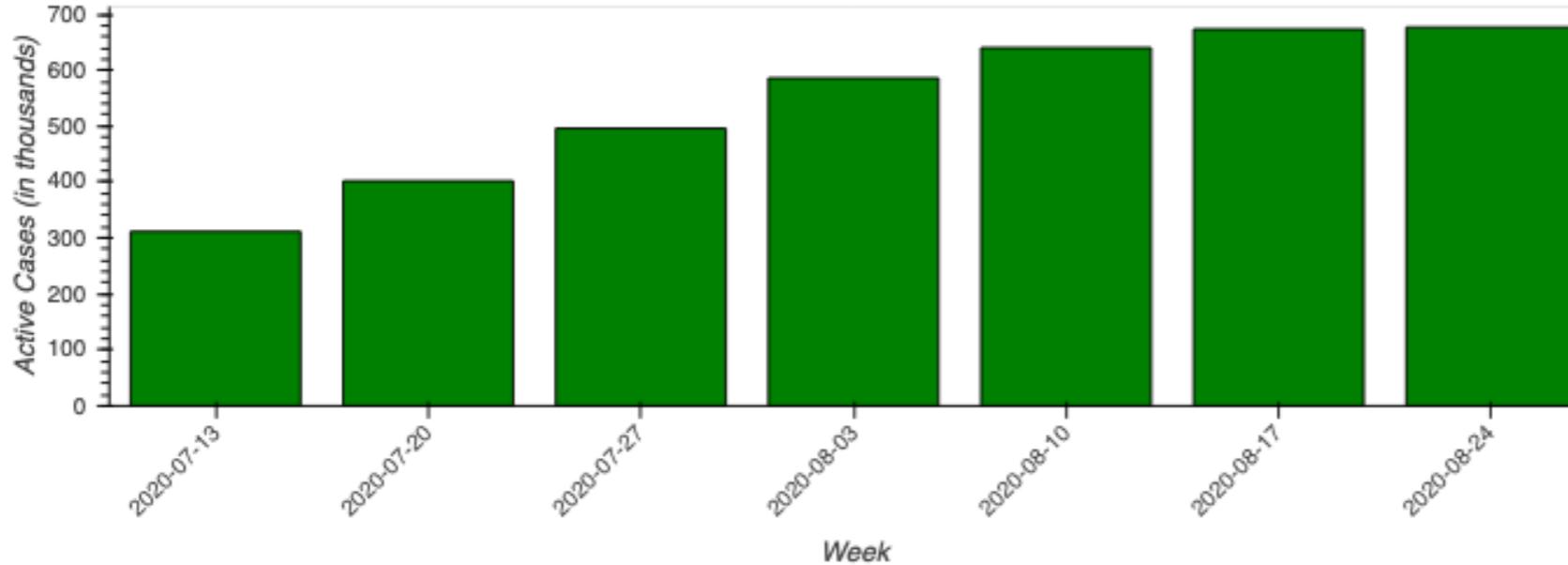
*Traffic Intensity per Week*

$$\text{Traffic Intensity} = \frac{\text{Avg. Confirmed Cases/day/week}}{\text{Avg. Recovered Cases/day/week}}$$

- Traffic intensity is an indication of resource-overwhelm risk. It basically shows how fast are new cases occurring vis-a-vis recoveries
- Just like if rate of inflow > outflow results in a traffic jam at a traffic junction in future, similarly Traffic Intensity > 1 is risky and Traffic Intensity > 2 needs corrective action

# RISK ASSESSMENT

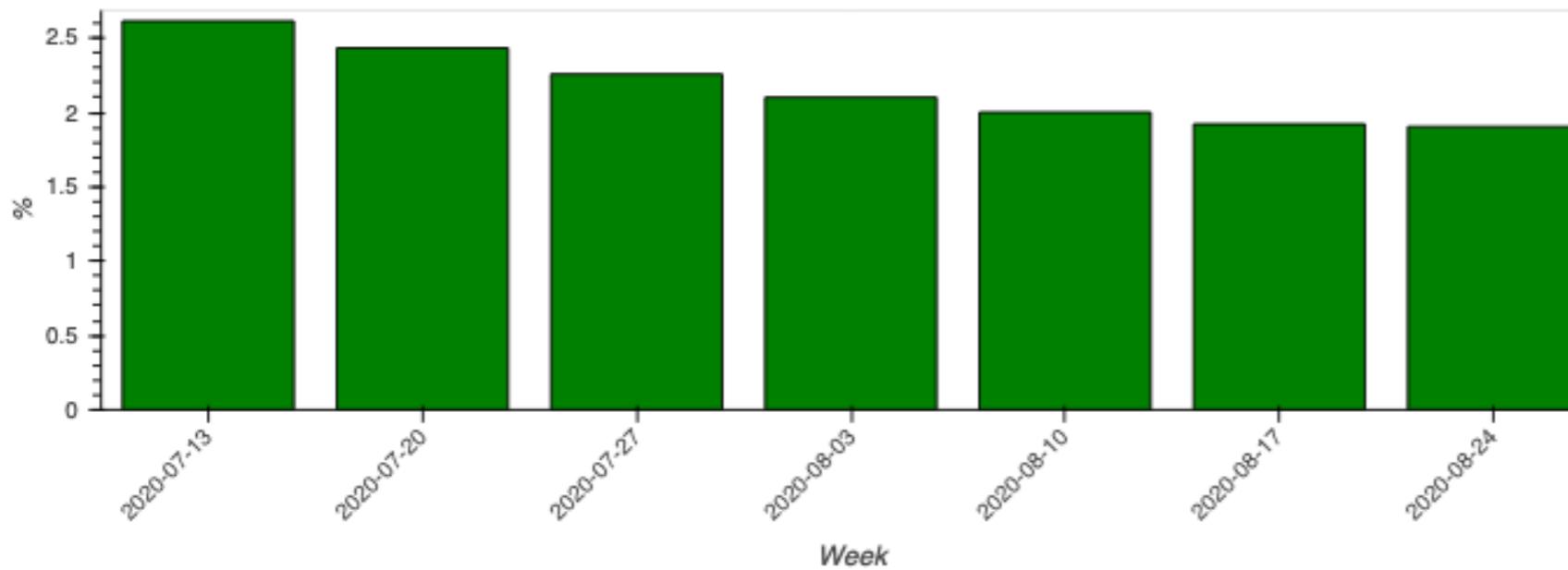
---



*Active Cases per Week*

Active cases / week = Cumulative Confirmed - Cumulative Recovered - Cumulative Deceased

---

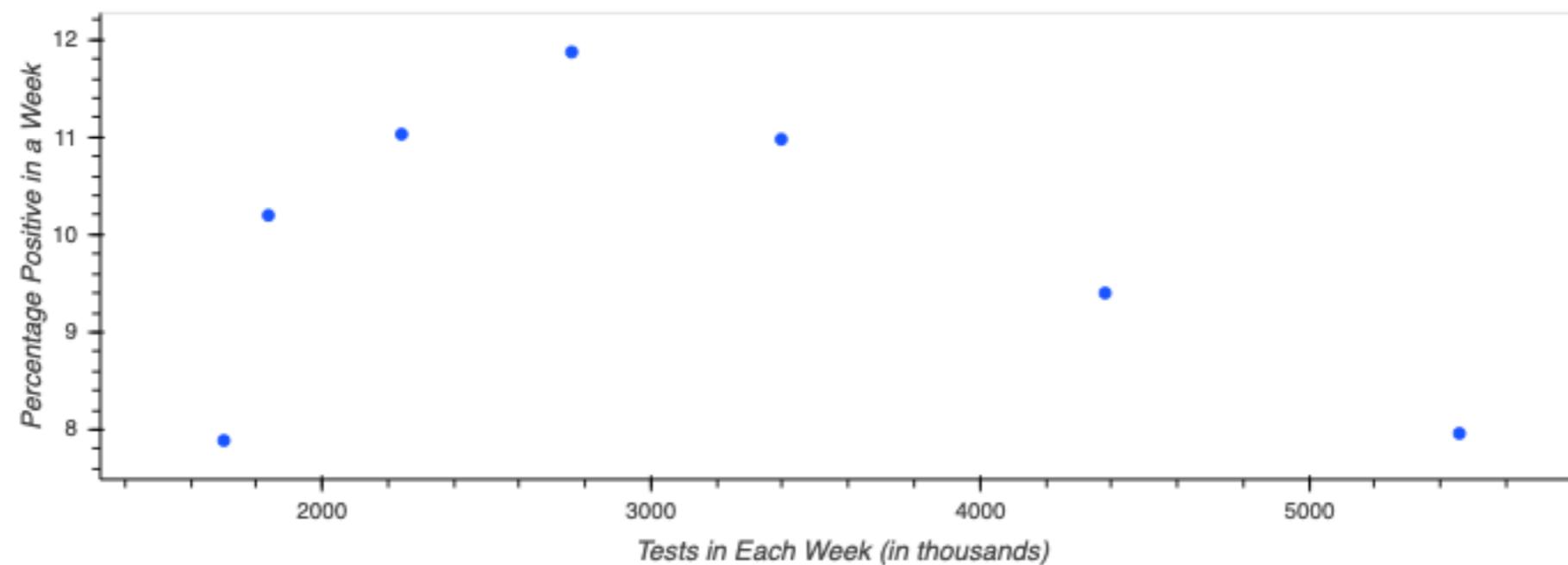
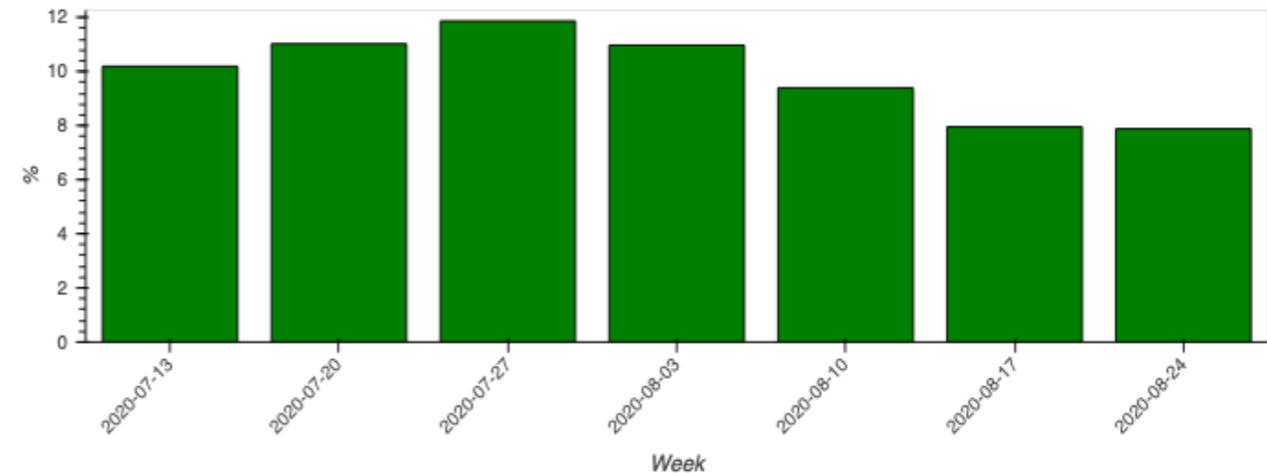
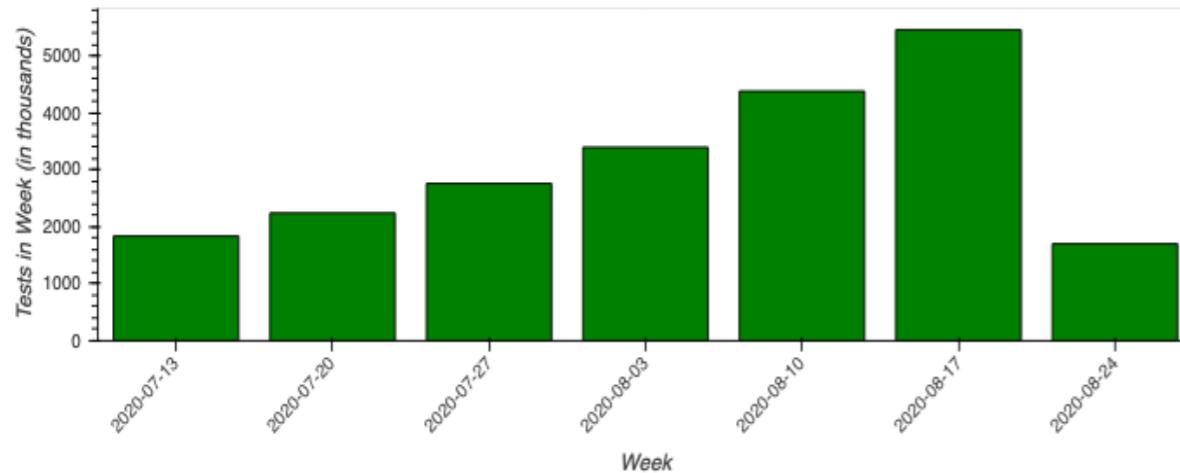


*Death Proportion till a Week*

Increasing trend is a cause of concern. Large variation across weeks may indicate systemic issues related to reporting.

# RISK ASSESSMENT

---



## *Testing Adequacy*

- Large variation across consecutive weeks requires investigation
- Increasing trend when number of tests are increasing may either indicate rapid spread or issues with testing



# COVID Dashboard

This will aid in unlock strategy design and intervention monitoring

[India](#) [States](#) [Comparison](#) [Districts](#) [FAQs](#) [Disclaimers and Explanations](#) [Contact Us](#)

Choose State

Karnataka

Confirmed	Active	Recovered	Deceased	Cumulative Tested	Proportion Positive (this wk)	Red Zones	Orange Zones	Green Zones
249590 +8642	81097 +1315	164150 +7201	4327 +126	2197259	0.0%	3	20	7

## Risk Summary

Data	
ICU Beds	2075.000000
Ventilators	999.000000
ICU Beds/ 100 Active Cases	2.560000
Ventilators/ 100 Active Cases	1.231737
Avg. Traffic Intensity of Ongoing Week *	1.050000
Weekly Avg. Confirmed Cases	8153.500000
Weekly Avg. Recovered Cases	7794.000000
Weekly Avg. Deceased Cases	132.500000

\*Traffic Intensity is the ratio of number of confirmed cases and number of recoveries in a given period. Values > 1 indicate faster arrival than recovery. Large values beyond 1 would inundate the health care system).

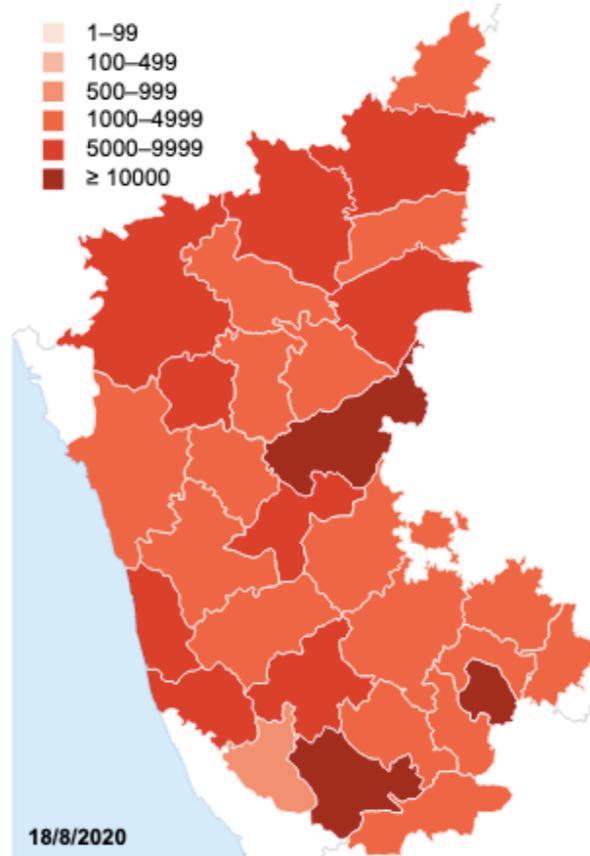
[Click Here](#) to download a risk assessment report.

[Click Here](#) for past reports.

Each week has been denoted by a date falling on the Monday when the week ends. For instance, '08/06/2020' represents the week that ends on '08/06/2020'.

Please refer the [FAQs](#) and [Disclaimer and Explanations](#) sections to better understand the graphs.

## Heatmap of Confirmed Cases



*Similar Analysis at State Level*

# COMPARATIVE ASSESSMENT

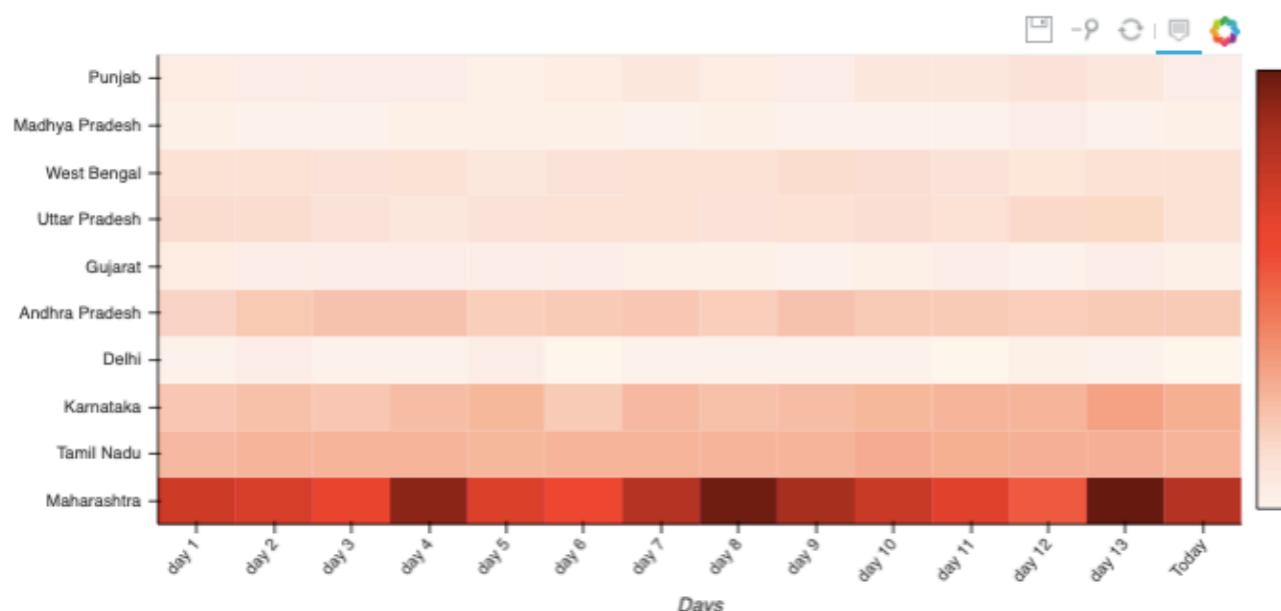
## COVID Dashboard

*This will aid in unlock strategy design and intervention monitoring*

Choose Criteria

Deceased

### Heatmap



Interpretation Tip: This heatmap shows daily distribution of various kinds of cases for top 10 States across last 14 days

### Risk Profile

The table can be sorted according to State name / metric by clicking on the respective column names

States	Proportion +	Total Tested	Traffic Intensity	Active Cases	Delta Confirmed	Delta Recovered	Delta Deceased
Maharashtra	13.98	850,858.0	1.14	160,413.0	13,165.0	9,011.0	346.0
Tamil Nadu	8.55	892,809.0	1.02	53,155.0	5,795.0	6,384.0	116.0
Karnataka	11.15	616,184.0	1.09	81,097.0	8,642.0	7,201.0	126.0
Delhi	4.37	237,871.0	1.07	11,137.0	1,398.0	1,320.0	9.0
Andhra Pradesh	5.38	719,964.0	1.05	86,725.0	9,742.0	8,061.0	86.0
Gujarat	2.48	574,847.0	0.98	14,320.0	1,145.0	1,120.0	17.0
Uttar Pradesh	1.39	1,277,487.0	1.13	49,645.0	5,076.0	5,620.0	53.0
West Bengal	9.27	388,305.0	1.13	27,678.0	3,169.0	2,973.0	53.0
Madhya Pradesh	4.44	243,455.0	1.23	10,717.0	976.0	762.0	18.0
Punjab	7.29	185,386.0	1.68	12,460.0	1,683.0	941.0	22.0

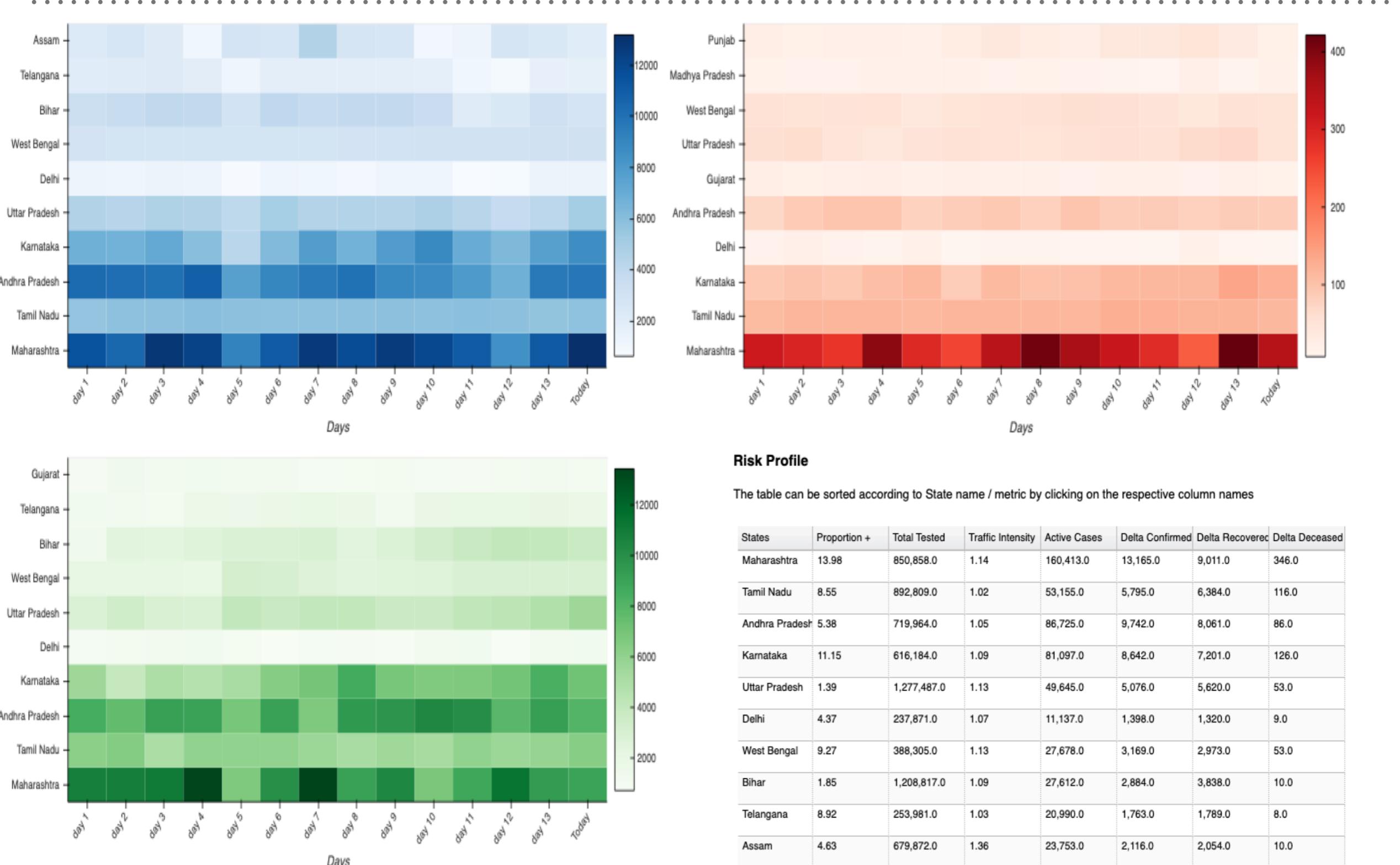
Interpretation Tip: Higher values of three major parameters – Traffic Intensity, Active Cases, and Proportion of Tests Positive signify risk. May need to be specially investigated.

### Daily Cases of the Chosen Type

The table can be sorted as per State name / cases on a day by clicking on the respective column names

States	day 1	day 2	day 3	day 4	day 5	day 6	day 7	day 8	day 9	day 10	day 11	day 12	day 13	Today
Maharashtra	316	300	275	390	293	256	344	413	364	322	288	228	422	346
Tamil Nadu	110	119	118	119	114	118	119	119	117	127	125	120	121	116
Karnataka	93	100	93	107	114	86	113	103	104	114	116	115	139	126
Delhi	15	23	16	13	20	8	14	14	11	10	8	18	12	9
Andhra Pradesh	72	89	97	97	80	87	93	82	97	87	88	82	88	86
Gujarat	27	22	23	25	20	23	18	18	15	19	20	15	20	17
Uttar Pradesh	61	63	47	41	51	56	54	50	55	58	56	66	70	53
West Bengal	56	52	51	54	41	49	54	56	60	58	51	45	55	53

# COMPARATIVE ASSESSMENT



# **DEATH AND SPREAD ANALYSIS**

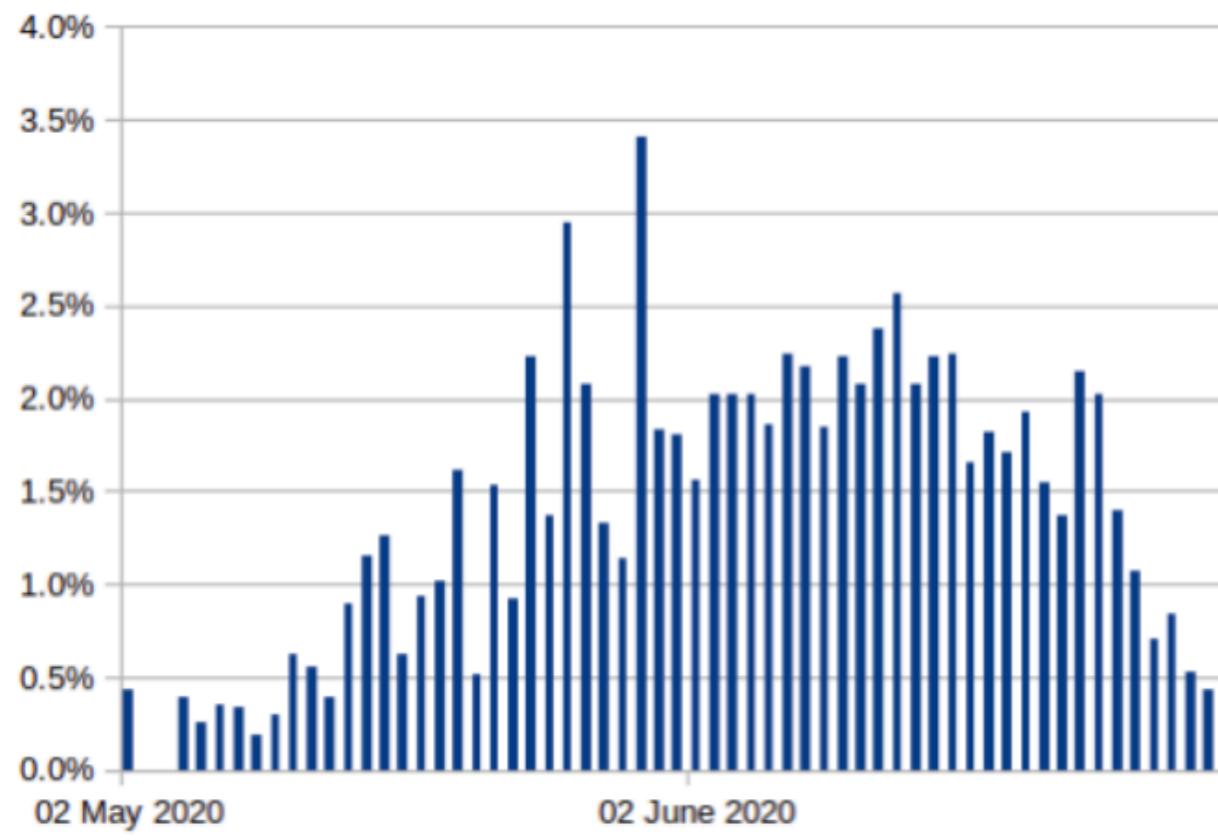
# BOUNDS ON DEATH RATE

---

- We arrived at two sets of bounds:
  - Crude Fatality Rate (CFR) :: Upper Bound
  - Kaplan-Meier Survival Analysis (KM) :: Lower Bound
- Data was used from the State of Tamil Nadu since it was the only State providing detailed information related to date of admission and death

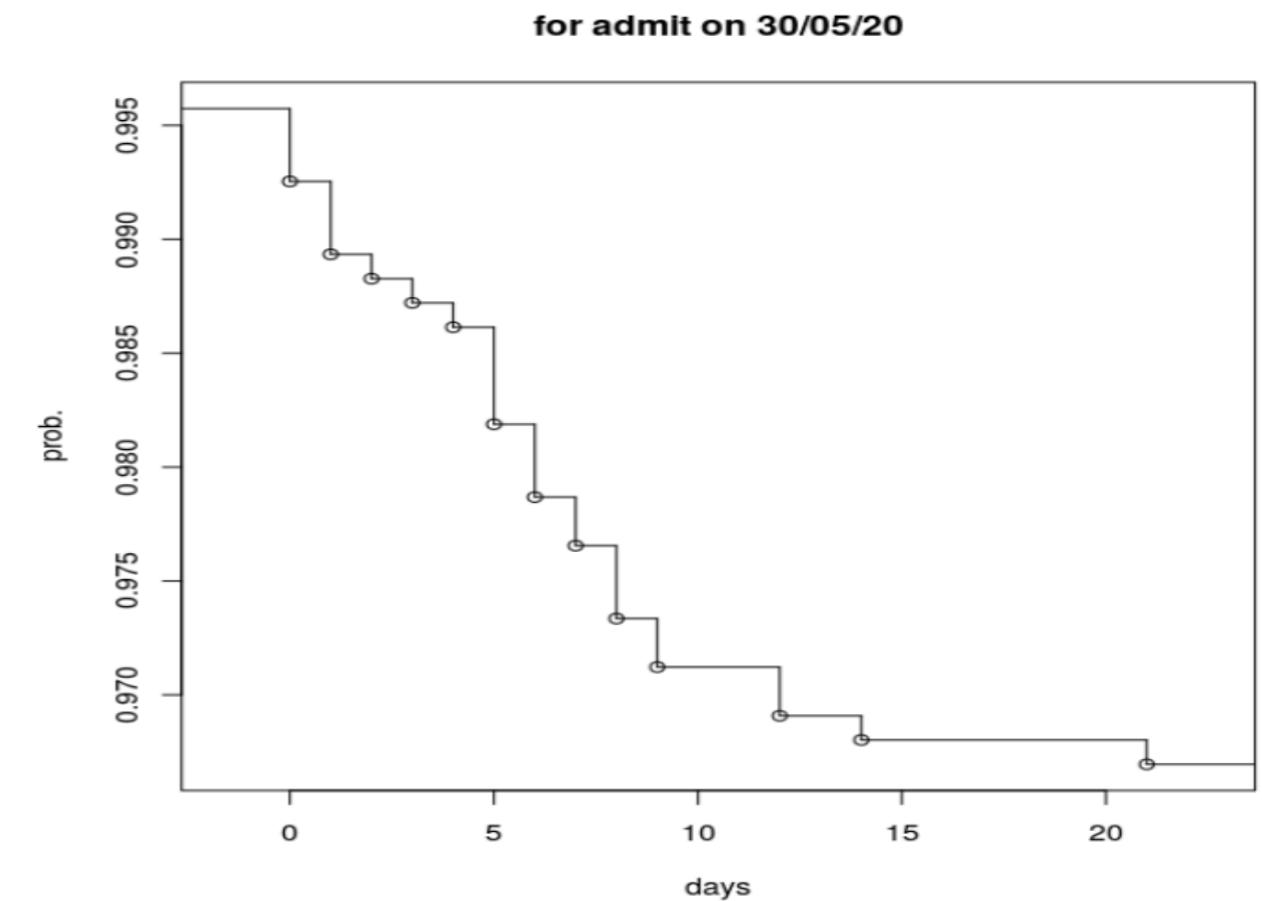
# BOUNDS ON DEATH RATE

---



*Crude Fatality Rate*

$$CFR = \frac{\sum \text{death}}{\sum \text{confirmed}}$$

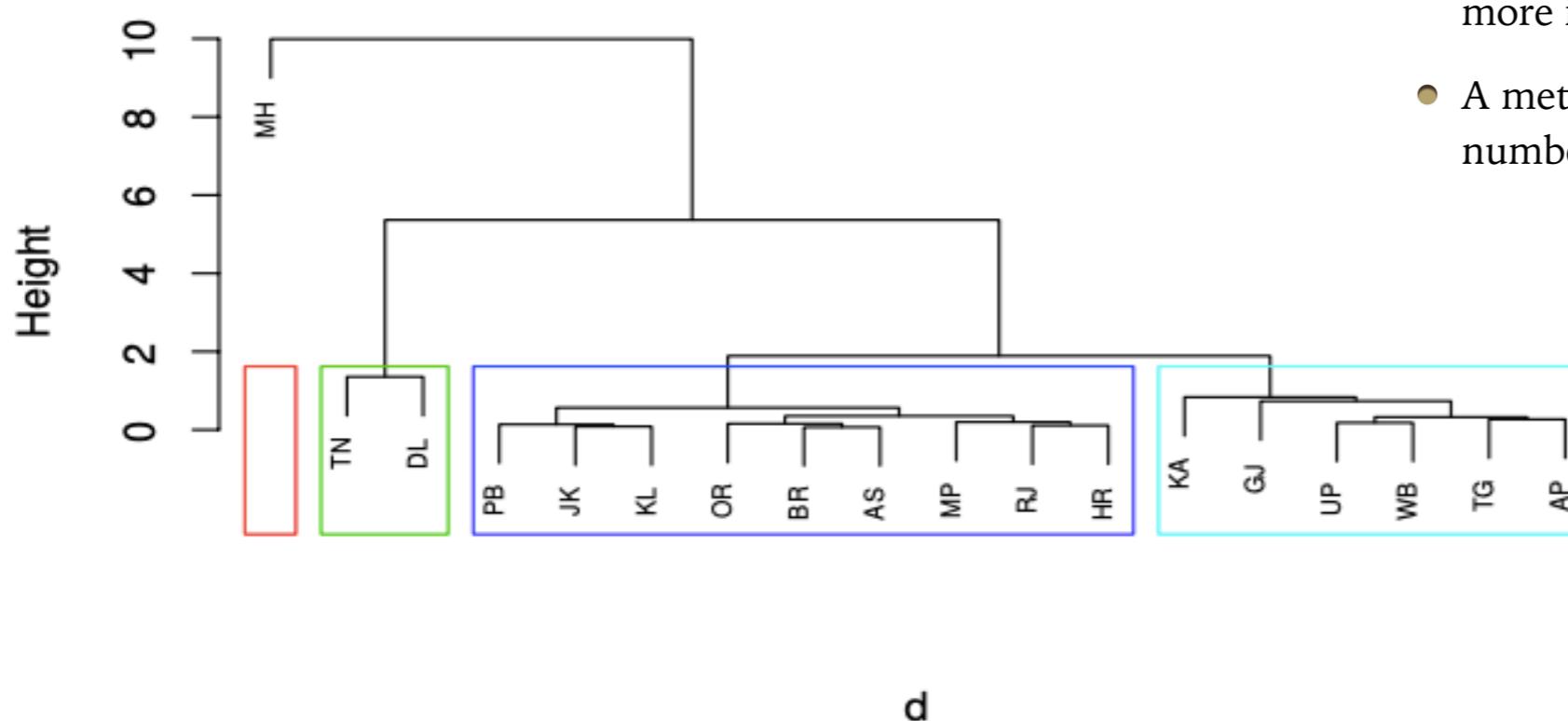
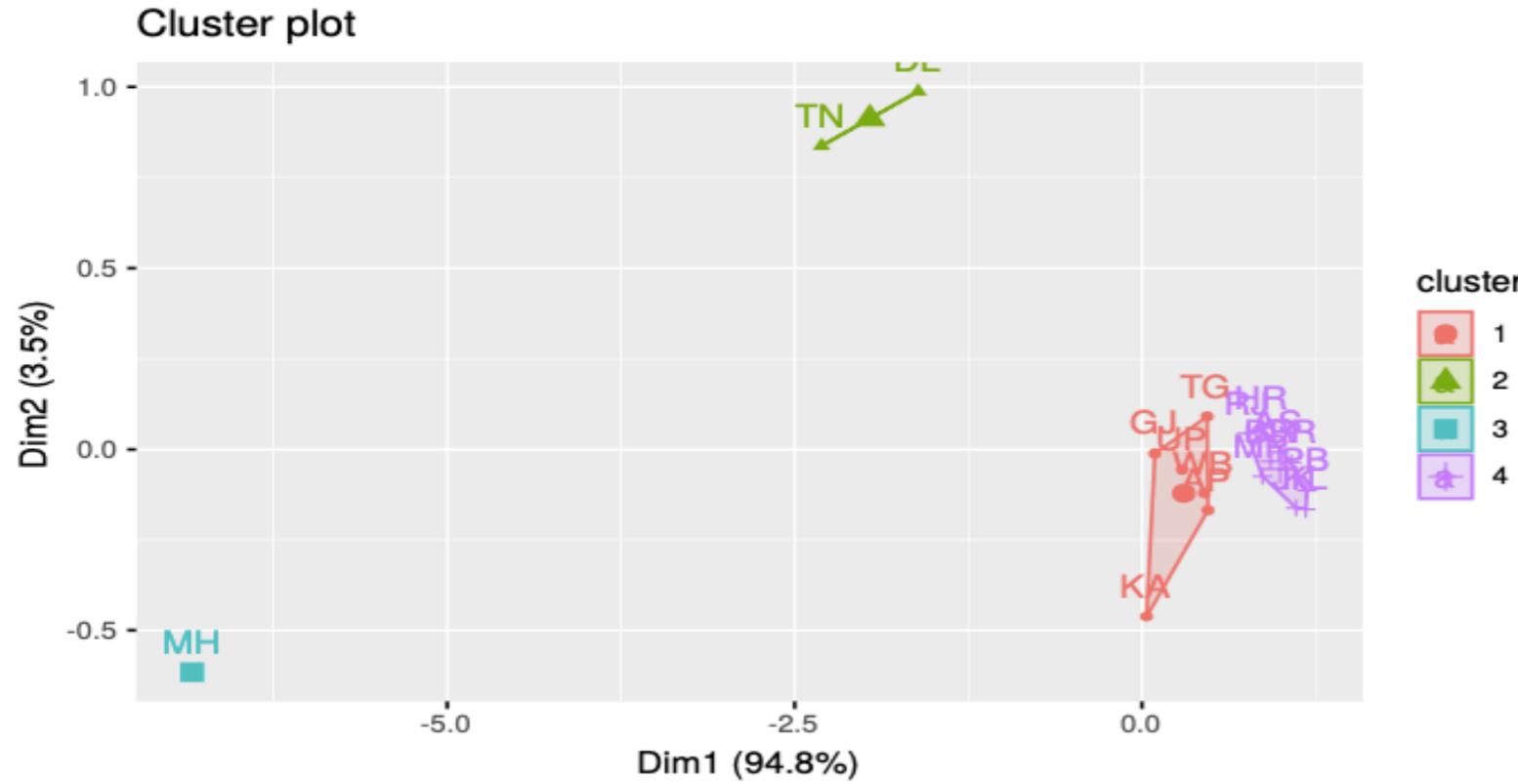


*Kaplan Meier Analysis*

$$\hat{S}_i = \prod_{i:t_i \leq t} \left( i - \frac{d_i}{n_i} \right)$$

# SPREAD BASED CLUSTERING

---



- Clustered States and Districts based on number of active cases and deaths
- Districts having a large city were found to have more infection
- A method needs to be devised to normalize number of cases across districts

# **FUTURE WORK**

# FUTURE WORK

---

- Behaviour of Recovery Time
- Assessment of efficacy of Contact Tracing
- Short-term Forecasting



**THANK YOU!**