

# An Analysis of Covid-19 In India

---

Name:- Lushna Samantaray

Redg No:- 1801289164

Trident Academy Of Technology,  
Bhubaneswar



# Problem Statement

- The aim is to analyze the dataset of covid-19 i.e (“Covid -19 in India”) which we got from Kaggle.
- And while analyzing the dataset we will get some brief ideas of the situation of covid-19 in India.
  - Like the total number of (confirmed, recovered, death, active) cases reported per day
  - Also find out when cases are less and what are the steps taken for it
  - And we will also get to know that the measures taken now a days are sufficient or some new measures should be taken to fight against covid.

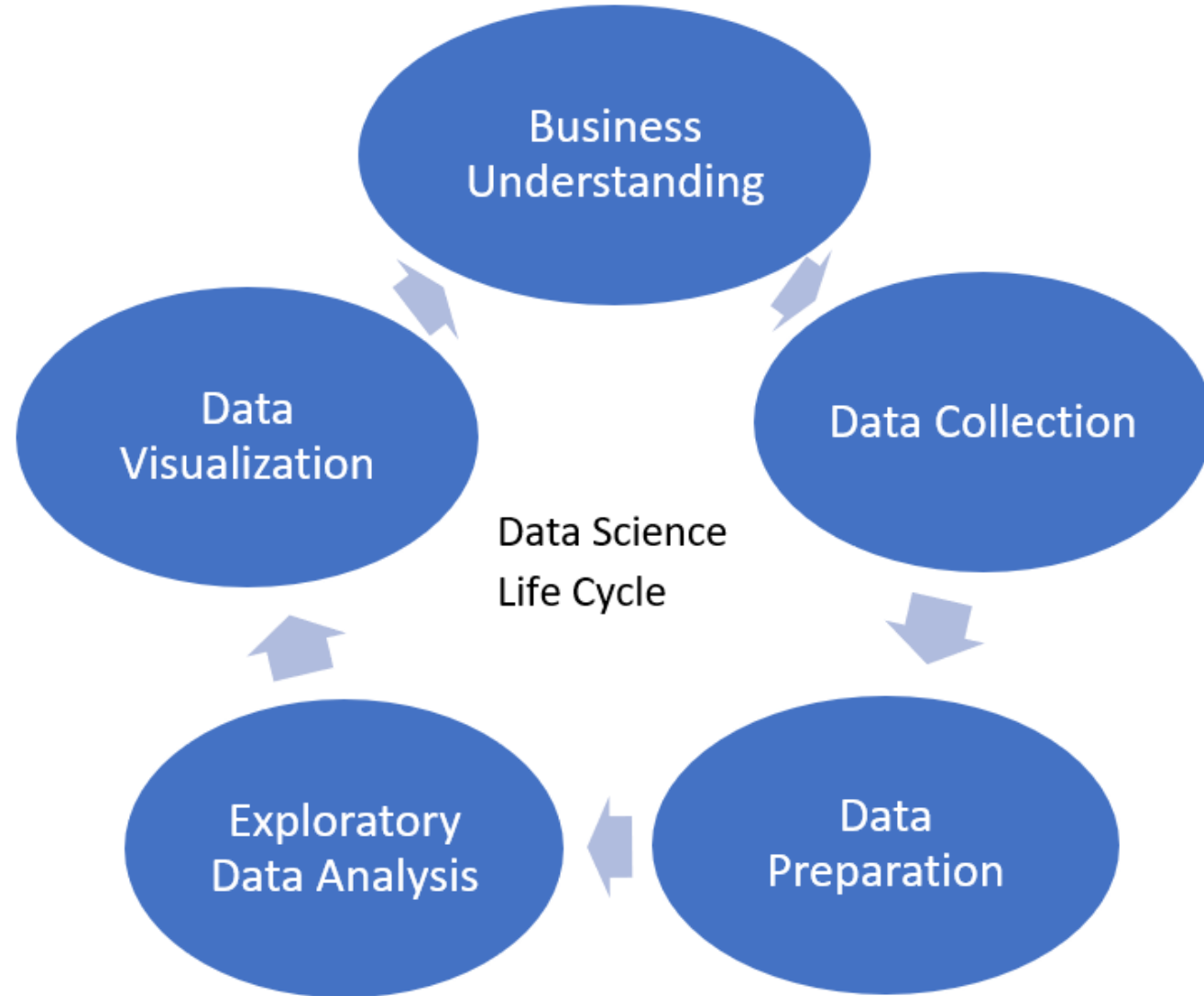
# Introduction

- COVID-19 is an infectious disease caused by the Coronavirus, biologically known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease was first identified in Wuhan, the capital of China's Hubei province in December 2019 and has spread all over the world since then. As of writing this, on 28th April 2020, 00:55 IST, there are 3 million confirmed cases throughout the world and has resulted in 208,000 deaths according to Google. Coronaviruses are a large family of viruses which may cause illness in animals or humans. In humans, several coronaviruses are known to cause respiratory infections ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS).

# Business Understanding

- The pandemic is sweeping through India at a pace that has staggered scientists. Daily case numbers have exploded since early March: the government reported 273,810 new infections nationally on 18 April. High numbers in India have also helped drive global cases to a daily high of 854,855 in the past week, almost breaking a record set in January.
- Researchers in India are now trying to pinpoint what is behind the unprecedented surge, which could be due to an unfortunate confluence of factors, including the emergence of particularly infectious variants, a rise in unrestricted social interactions, and low vaccine coverage. Untangling the causes could be helpful to governments trying to suppress or prevent similar surges in india.

# Approach



# Data Understanding

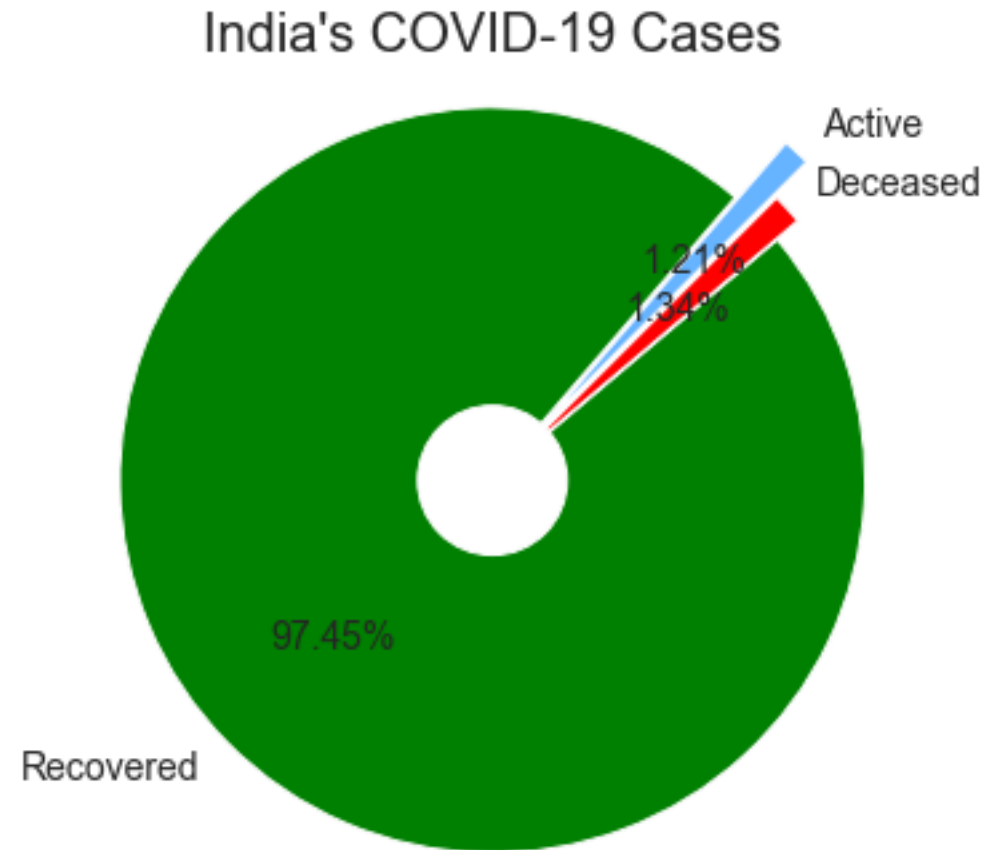
S.no	Column Name	Description
1	Date	Date of the number cases reported
2	Time	Time of the reports published
3	State/UnionTerritory	Individual States in India
4	ConfirmedIndianNational	No. of reported confirmed case of Indian nationals
5	ConfirmedForeignNational	No. of reported confirmed case of Foreign nationals
6	Cured	No. of people already recovered or discharged
7	Deaths	No. of people died
8	Confirmed	Total no. of confirmed cases reported
9	States	State/UnionTerritory renamed to States
10	Recovered	Cured renamed to Recovered
11	Deceased	Deaths renamed to Deceased
12	Active_case	No. of people still being treated / not yet recovered
13	New Cases	No. of confirmed cases reported per day
14	New Recoveries	No. of people recovered per day
15	New Deceased	No. of people died per day
16	Mortality Rate (per 100)	No. of deaths in a particular population w.r.t size of the population, per unit of time
17	Recovery Rate (per 100)	No. of Recoveries in a particular population w.r.t size of the population, per unit of time
18	Confirmed_Cases	New Cases renamed to Confirmed_Cases
19	Recovered_Cases	New Recoveries renamed to Recovered_Cases
20	Deaths	New Deceased renamed to Deaths
21	India_Recovered	Total no. of Recovered_Cases (sum total)
22	India_Active	Total no. of Active_case (sum total)
23	India_Deceased	Total no. of Deaths (sum total)
24	State_wise	All the data's were grouped w.r.t States
25	Dayswise	All the data's were grouped w.r.t Date
26	Incidence Rate (per 100)	No. of incidence taking place per day

*Features of covid-19 cases*

# Summary of the Dataset

- The number of rows in the dataset :- 18047
- The number of columns in the dataset :- 11
- The dataset has a mixture of Discrete and Continuous features.
- Number of Numerical features :- 8
- Number of Categorical Features :- 3
- There are lots of NULL values present in the dataset.

# Data Visualization



Covid-19 Situation in India

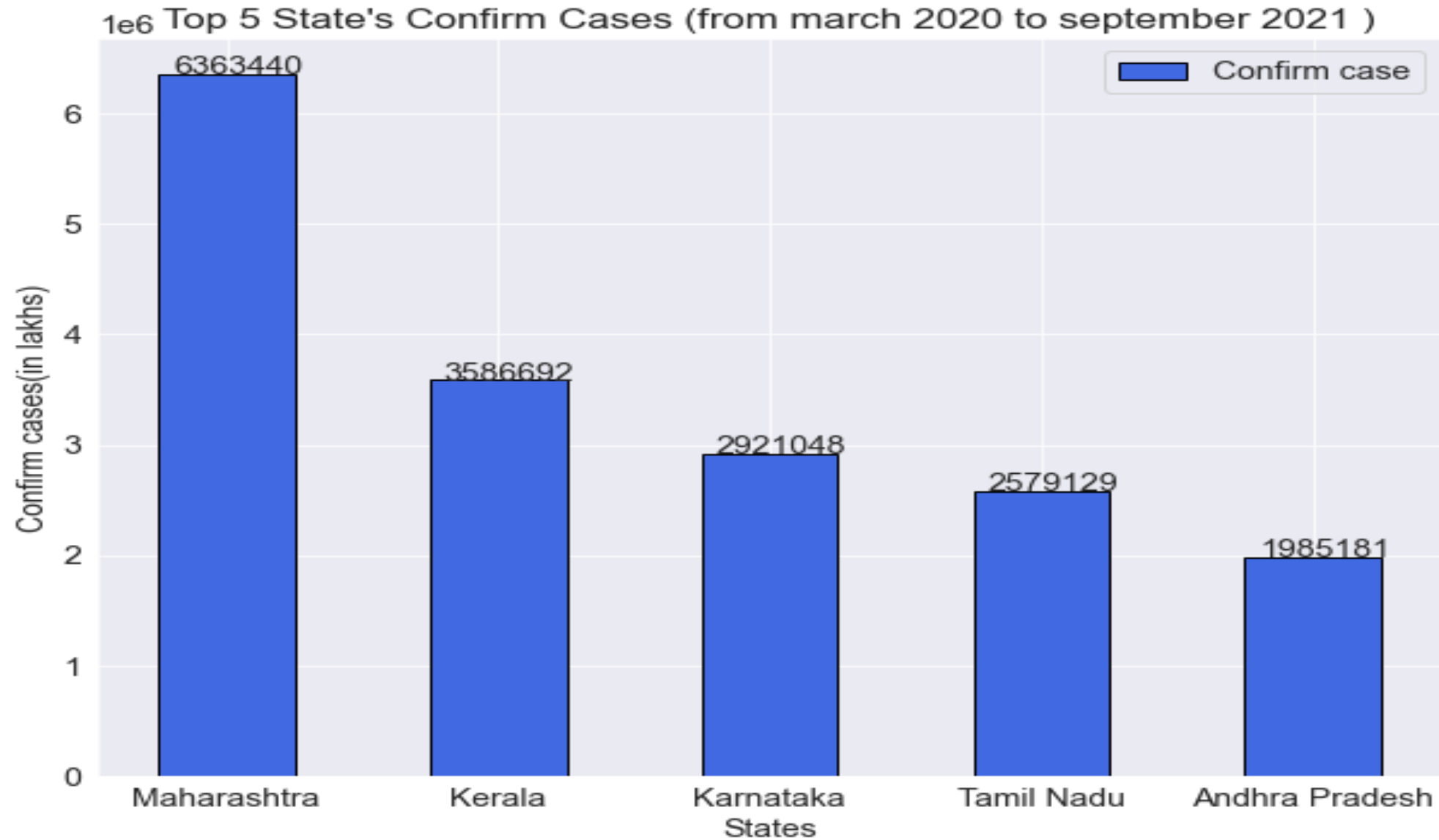


# Data Visualization

States	Confirmed_Cases	Deaths	Recovered_Cases	Active_case	Mortality Rate(per 100)	Recovery Rate(per 100)
Maharashtra	6363440	134201	6159676	69563	2.110000	96.800000
Kerala	3586692	18004	3396184	172504	0.500000	94.690000
Karnataka	2921048	36848	2861499	22701	1.260000	97.960000
Tamil Nadu	2579129	34367	2524400	20362	1.330000	97.880000
Andhra Pradesh	1985181	13564	1952736	18881	0.680000	98.370000
Uttar Pradesh	1708806	22775	1685492	539	1.330000	98.640000
West Bengal	1534998	18252	1506532	10214	1.190000	98.150000
Delhi	1436851	25068	1411280	503	1.740000	98.220000
Chhattisgarh	1003355	13544	988189	1622	1.350000	98.490000
Odisha	988996	6565	972710	9721	0.660000	98.350000
Rajasthan	953850	8954	944700	196	0.940000	99.040000
Gujarat	825080	10077	814802	201	1.220000	98.750000
Madhya Pradesh	791976	10514	781330	132	1.330000	98.660000
Haryana	770112	9652	759790	670	1.250000	98.660000
Bihar	725277	9645	715352	280	1.330000	98.630000
Telangana	650352	3831	638410	8111	0.590000	98.160000
Punjab	599572	16322	582791	459	2.720000	97.200000
Assam	576148	5420	559684	11044	0.940000	97.140000
Jharkhand	347439	5130	342102	207	1.480000	98.460000
Uttarakhand	342461	7368	334650	443	2.150000	97.720000
Jammu and Kashmir	322770	4392	317081	1297	1.360000	98.240000
Himachal Pradesh	208614	3537	202761	2316	1.700000	97.190000
Goa	172082	3164	167978	940	1.840000	97.620000
Puducherry	121765	1800	119115	850	1.480000	97.820000
Manipur	105423	1664	96776	6983	1.580000	91.800000
Tripura	80659	773	77811	2075	0.960000	96.470000
Meghalaya	69768	1185	64157	4426	1.700000	91.960000
Chandigarh	61991	811	61150	30	1.310000	98.640000
Arunachal Pradesh	50604	248	47821	2535	0.490000	94.500000
Mizoram	46319	171	33722	12426	0.370000	72.800000
Nagaland	28810	585	26852	1373	2.030000	93.200000
Sikkim	28017	356	25095	2566	1.270000	89.570000
Ladakh	20409	207	20130	72	1.010000	98.630000
Dadra and Nagar Haveli and Daman and Diu	10653	4	10646	3	0.040000	99.930000
Lakshadweep	10263	51	10165	47	0.500000	99.050000
Andaman and Nicobar Islands	7547	129	7412	6	1.710000	98.210000

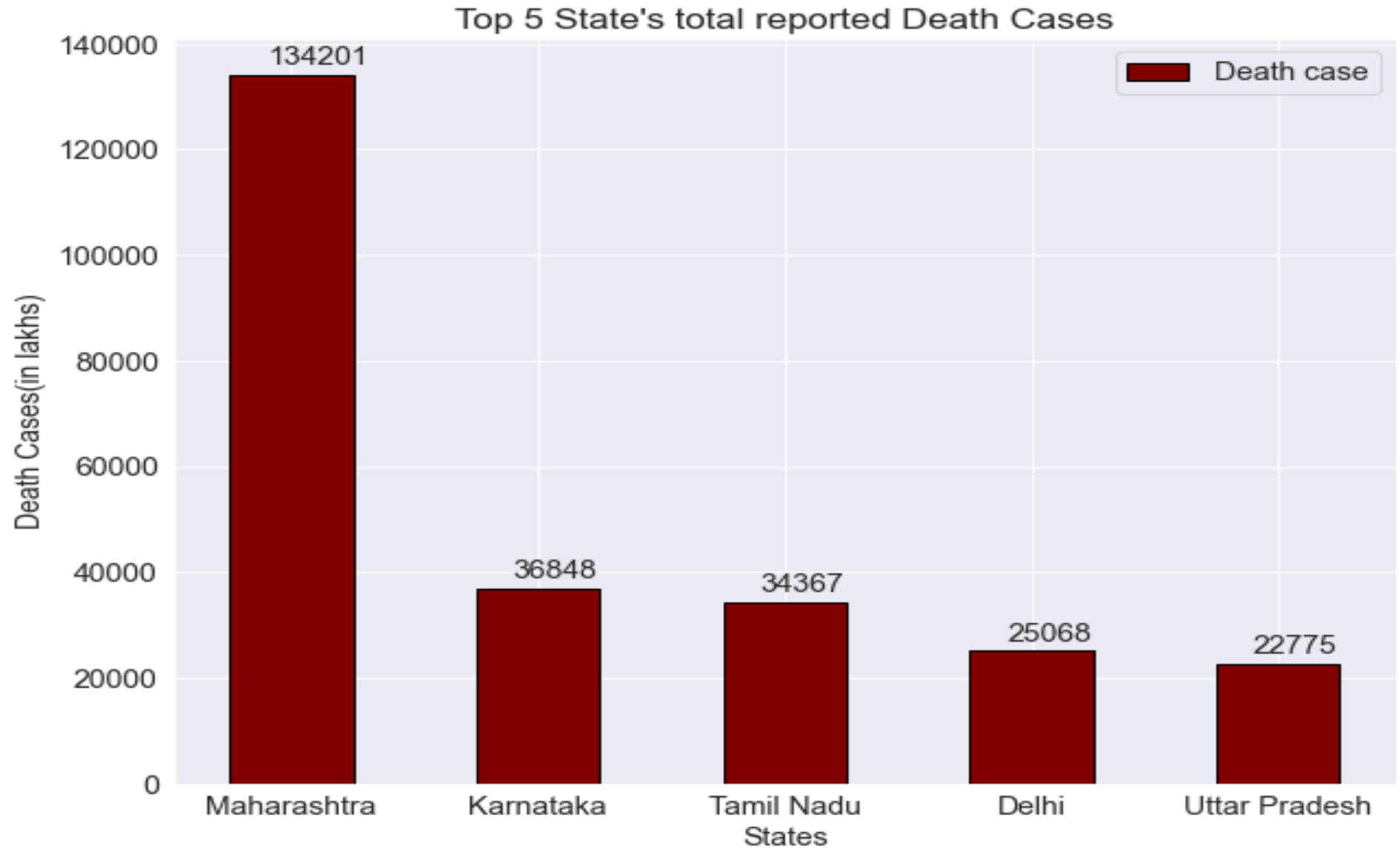
State Wise total Cases Reported

# Data Visualization



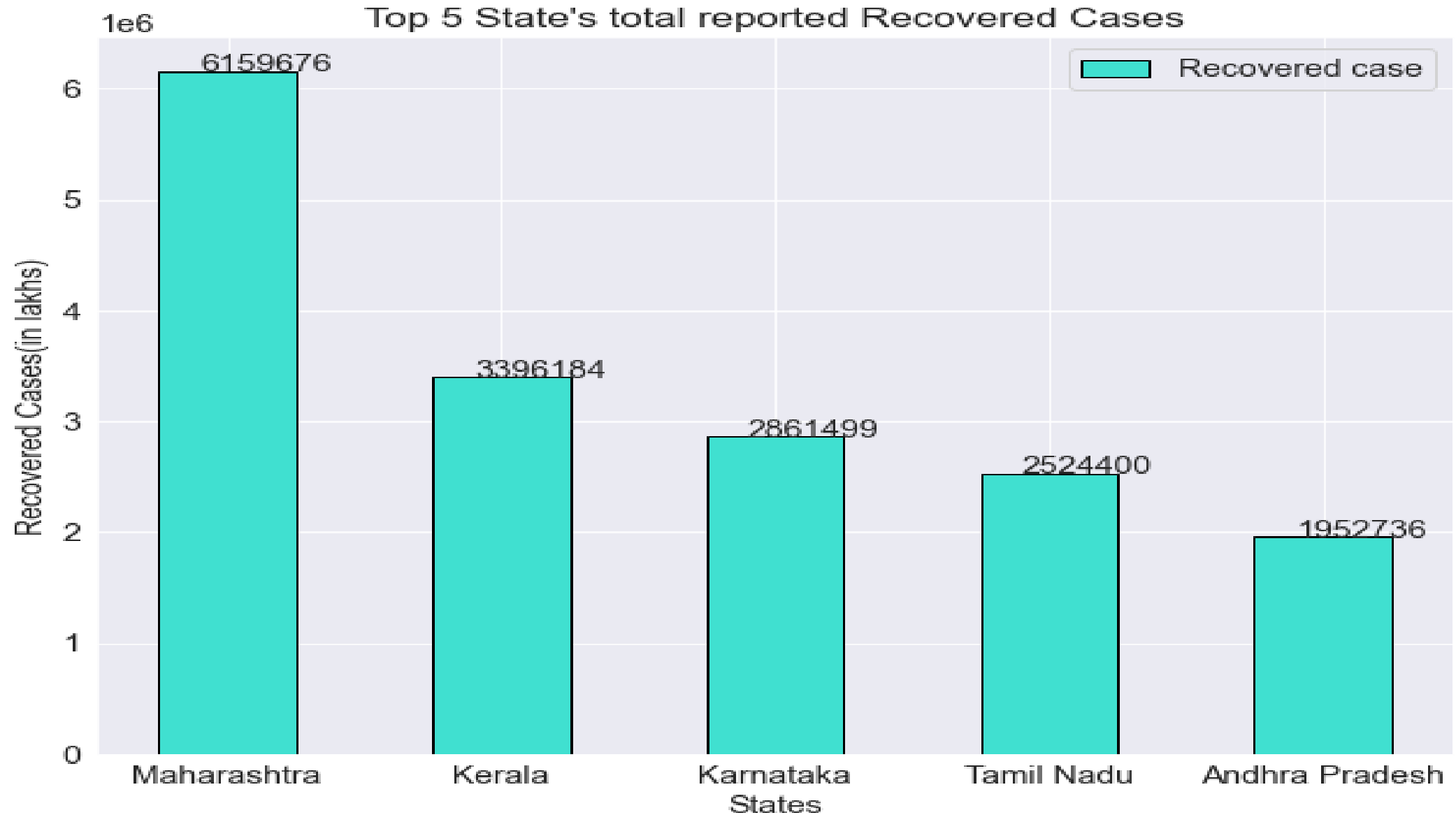
State VS Confirm Case

# Data Visualization



State VS Death Case

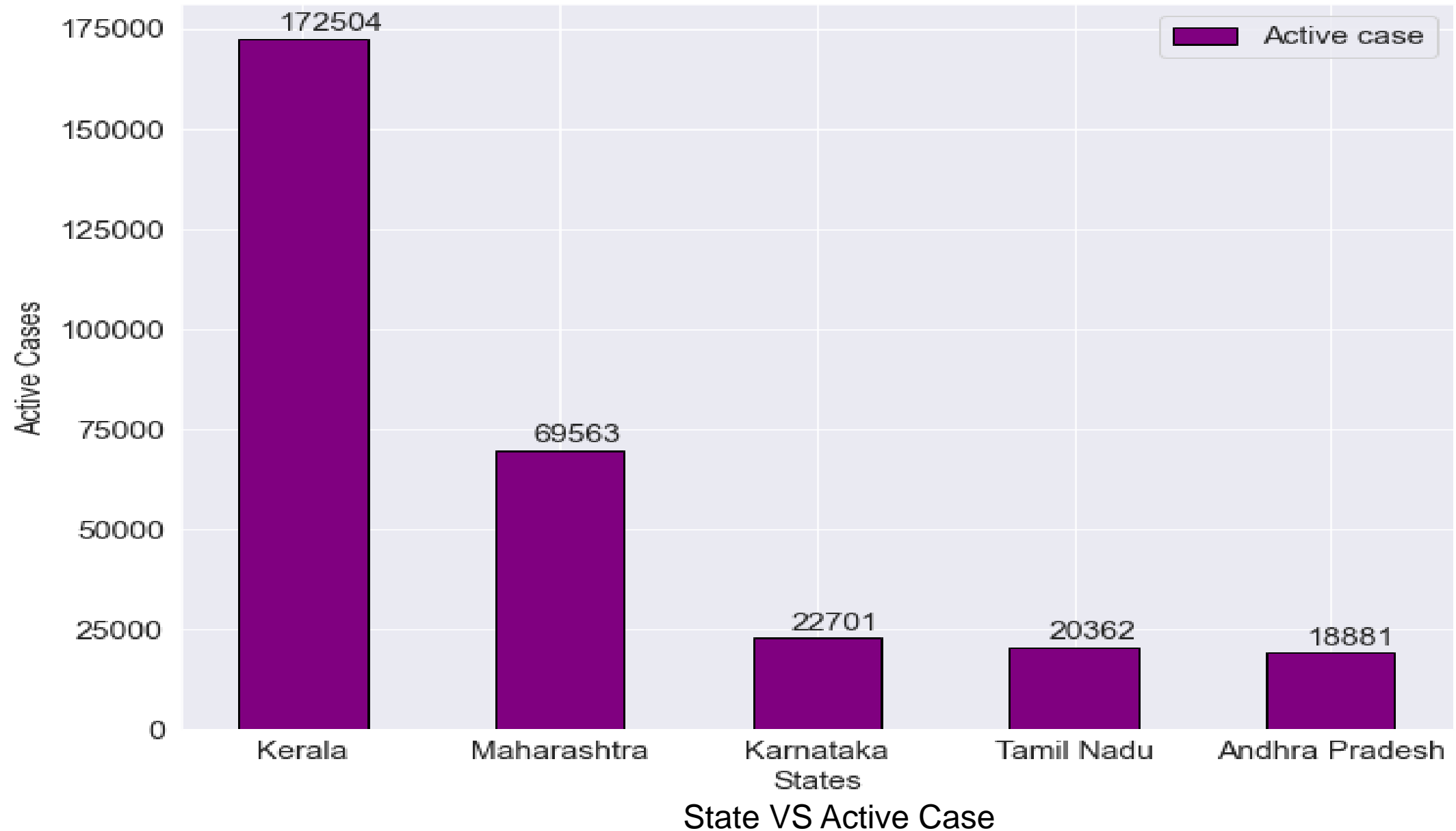
# Data Visualization



State VS Recovered Case

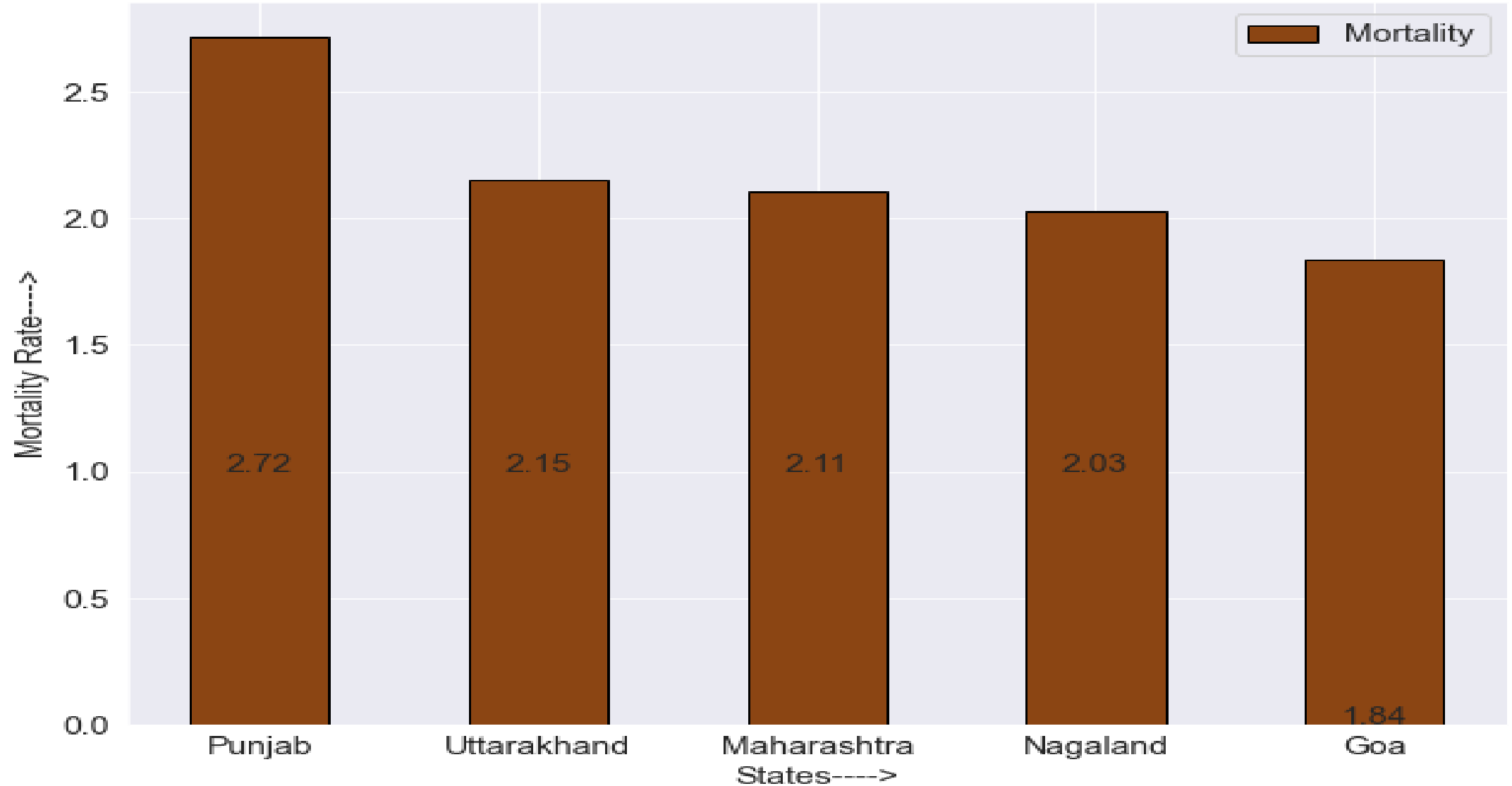
# Data Visualization

Top 5 State's total reported Active Cases



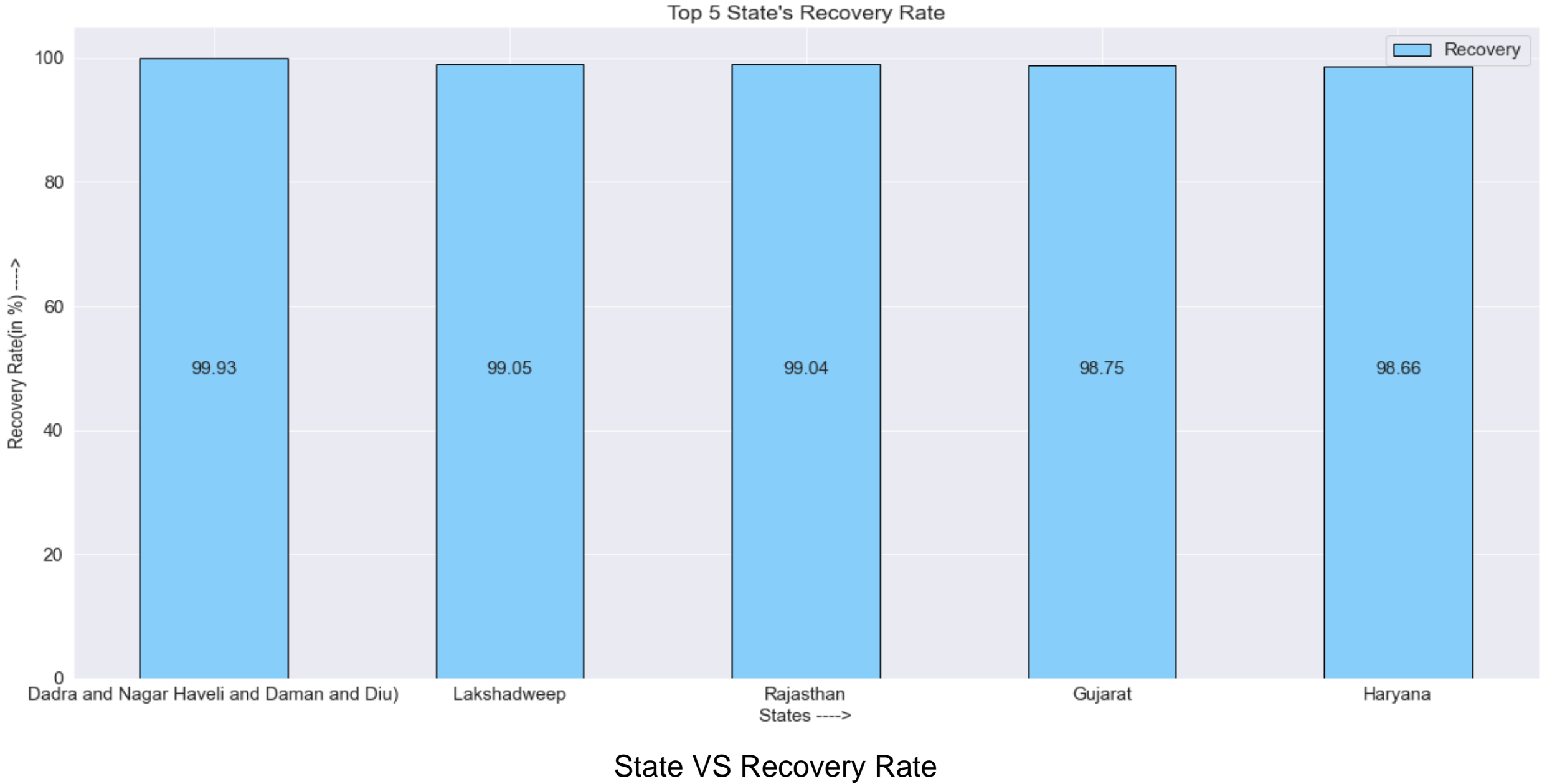
# Data Visualization

Top 5 State's Mortality Rate



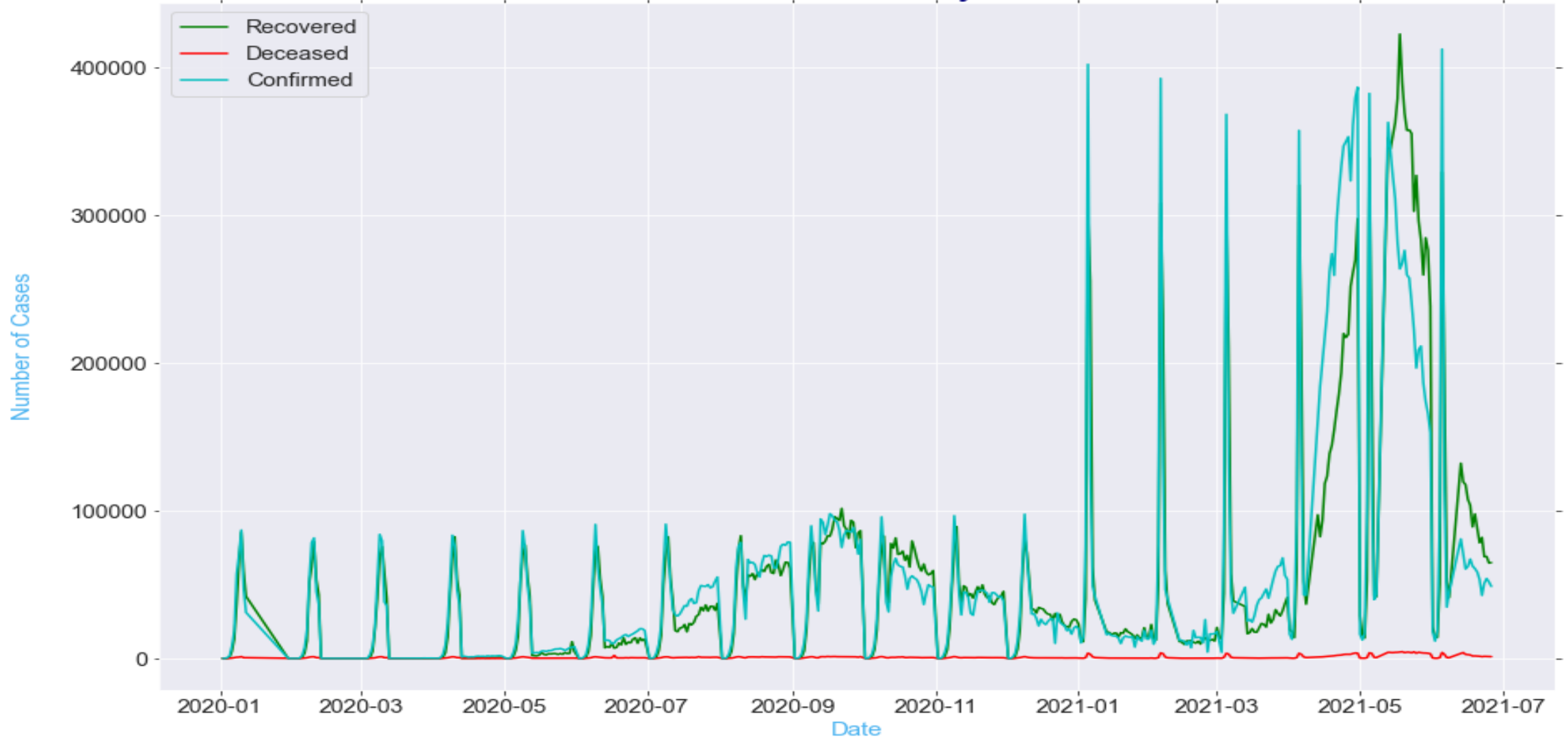
State VS Mortality Rate

# Data Visualization



# Data Visualization

## India Covid-19 Daily Cases

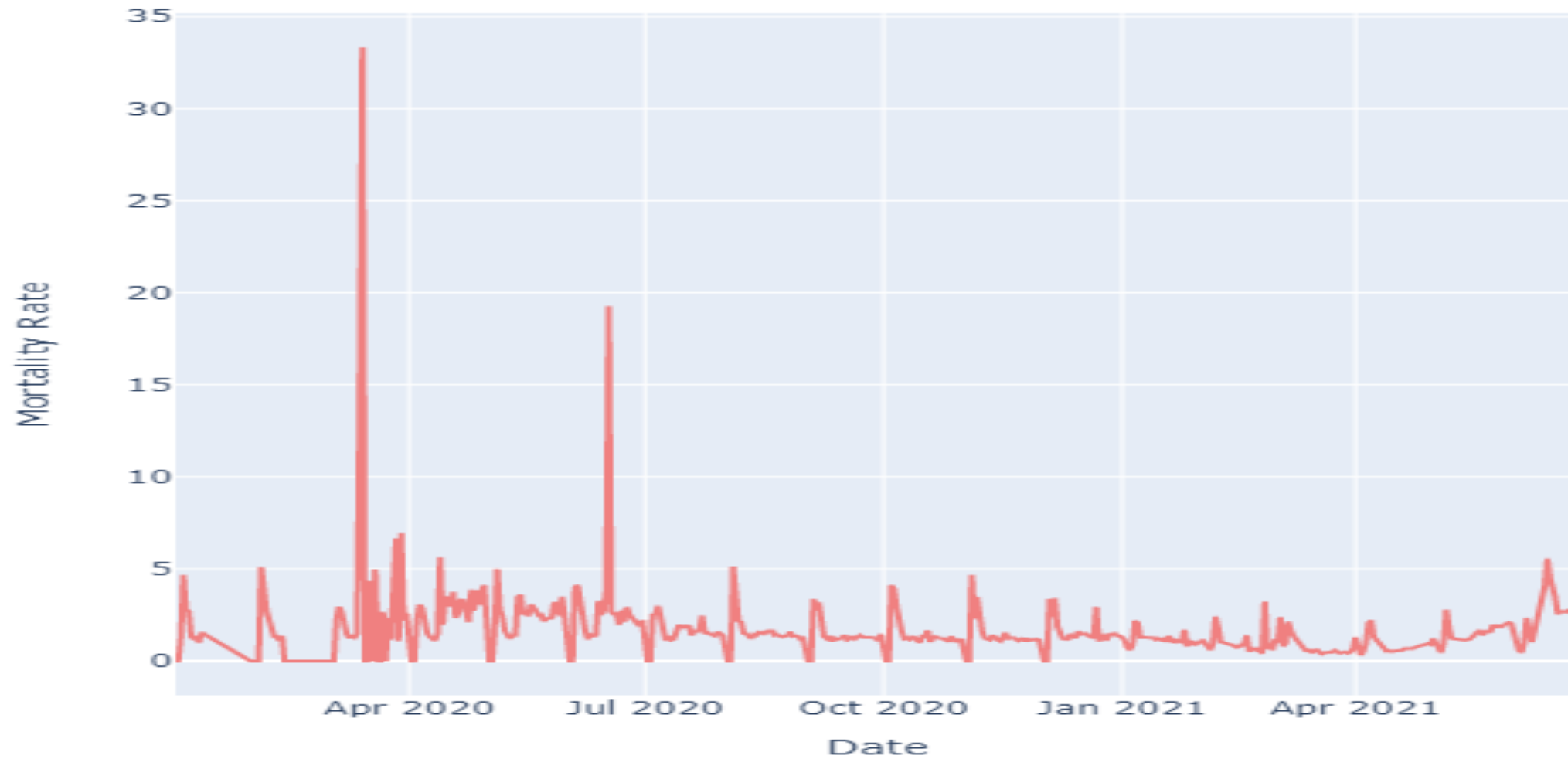


Date VS Total Number of reported cases



# Data Visualization

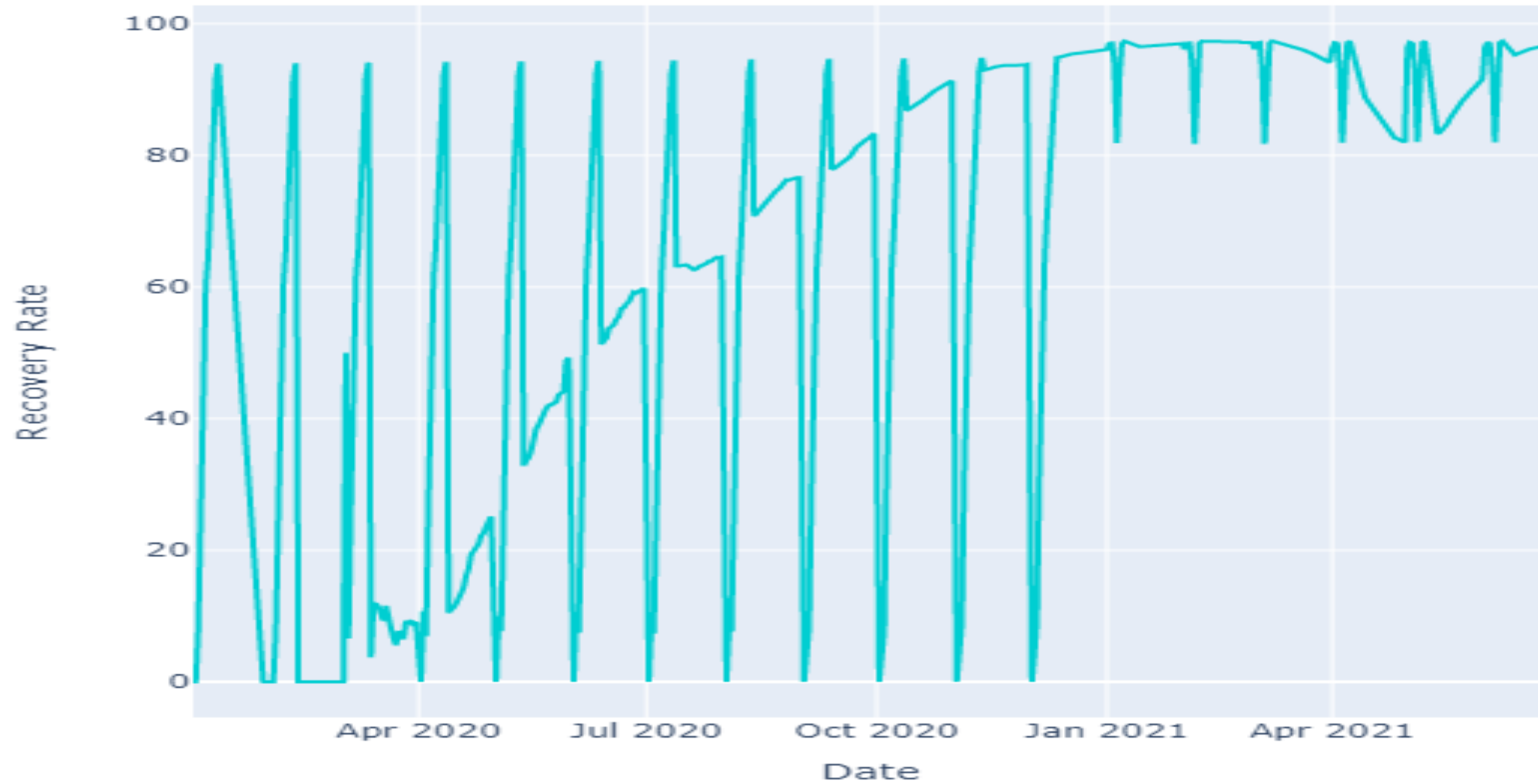
Covid-19 Mortality Rate (per 100) in India



Date VS Mortality Rate

# Data Visualization

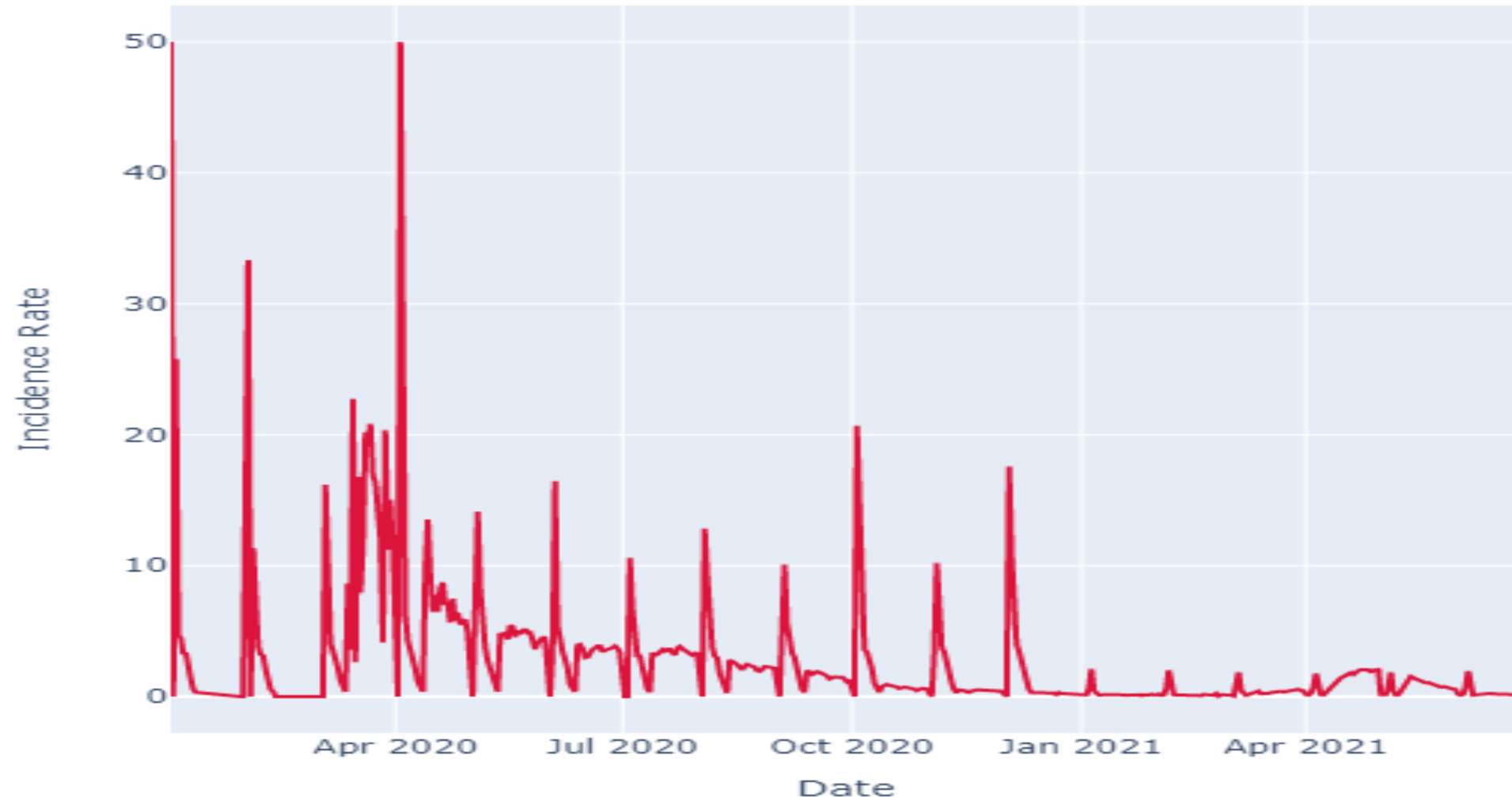
Covid-19 Recovery Rate (per 100) in India



Date VS Recovery Rate

# Data Visualization

Covid-19 Incidence Rate(per 100) in India



Date VS Incidence Rate

# Data Preparation

- Dropping the duplicate data's
- Combining all the duplicate rows into one
- Remove all the Null and unassigned values
- Dropping all the missing features

# Data Transformation

- Dropping some column for easy access
- Renaming some columns
- Adding some more important columns for visualization
- Grouping the data for future use

# Conclusion

- Assuming that the current measures are inefficient and inadequate, the number of estimated cases will continue to increase by the end of Sept.
- The current control measures( lockdown and Quarantine where necessary) should be taken, also avoid formation of crowds, so that the estimated rise of cases will go down.
- At the time of the outbreak of COVID-19, the best and most urgent steps must be taken to overcome the coronavirus epidemic. The fight against coronavirus infection should be an emergency and should not be delayed.