

# PANDEMIC DATA SCIENCE

**Prof. Anirban Dasgupta**

**Harshit Kumar**

**18110063**

**Pranshu Kumar Gond**

**18110124**

**Sagar Bisen**

**18110124**



# INDEX:

1. Goal Statement
2. Datasets explained
3. EDA (Vulnerability Metric)
  - a. State vulnerability
  - b. Rail vulnerability
4. Results
5. References

Can the currently available pandemic data help in fighting emergent situations in the near future?



# Datasets

- Demographic Dataset
- Vaccination Dataset
- Health Infrastructure Dataset
- Indian Railway Network
- Case Counts

# O1

## DEMOGRAPHIC DATASET

Contains the data related to general population.

Features: Illiterate population, Working population, marginal workers, male/female

District-wise distribution



## 02 VACCINATION DATASET

Contains the vaccination-related details.

Features: Total vaccination sites,  
Doses registered. (till 25 April, 2021)

District-wise distribution of two vaccines(Covishiled and Covaxin)



# 03 HEALTH INFRASTRUCTURE DATASET

Health facilities in hospitals of India

Features: Total govt. hospitals, No. of beds in hospitals maintained etc.

State-wise distribution of hospitals and their provided facilities.



# 04 CASE COUNT

District wise case counts related to covid.

Features: Active cases, Recovered cases, Death count etc.

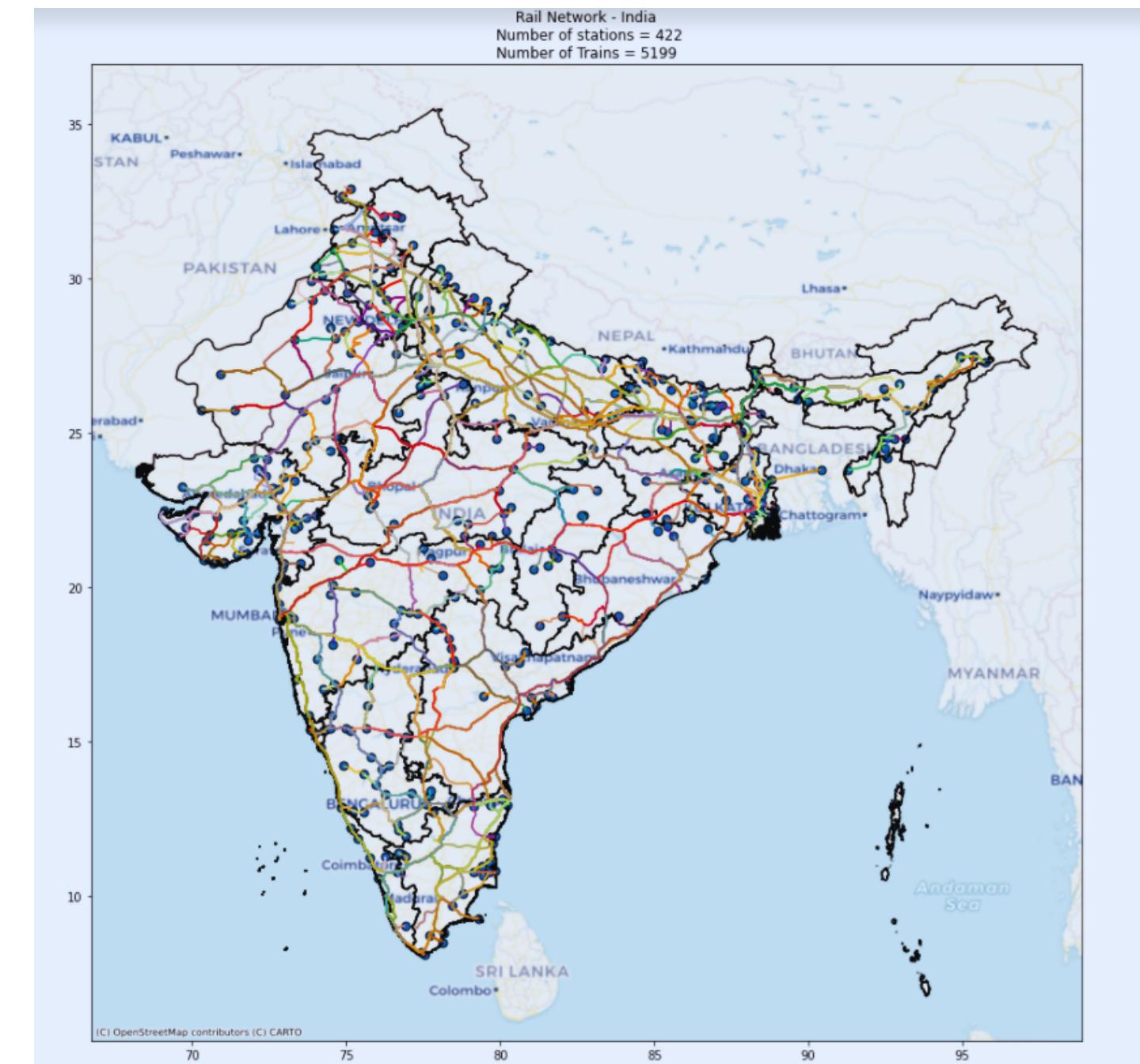
Depict the current situation of India



# 05 INDIAN RAILWAY NETWORK

State-wise distribution of network.

A total of 5199 trains and 422 stations.

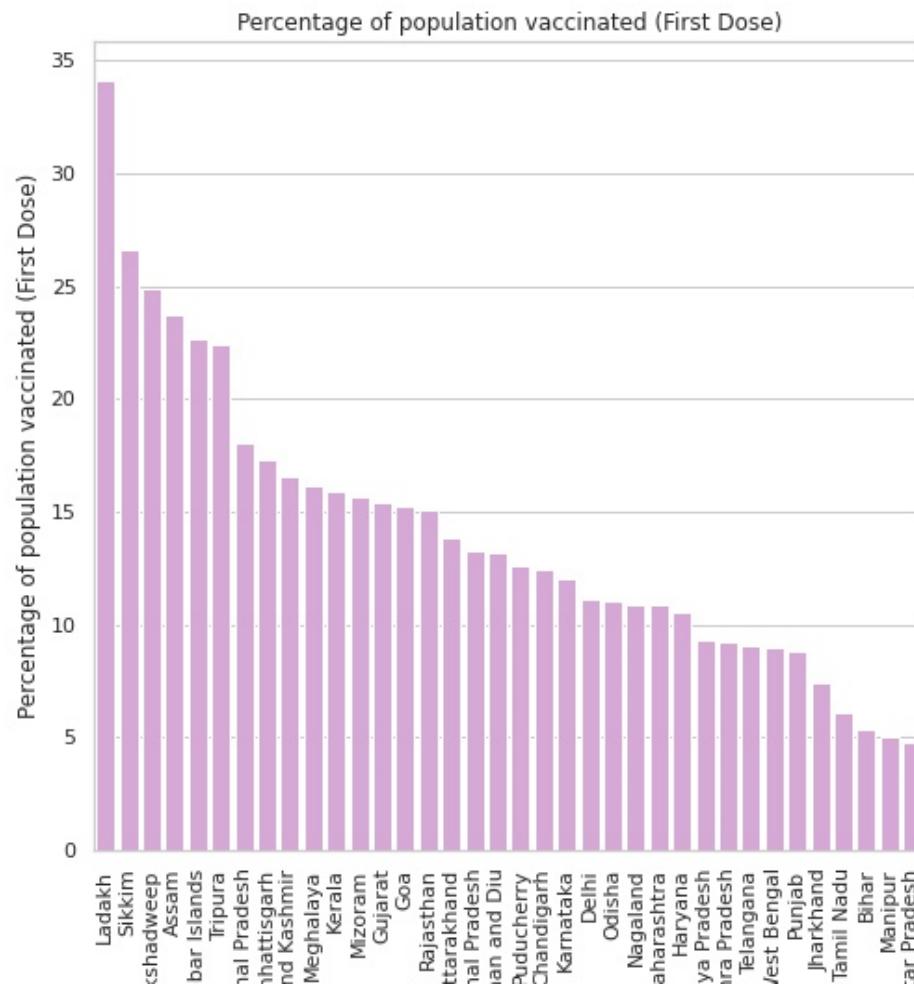


# EDA

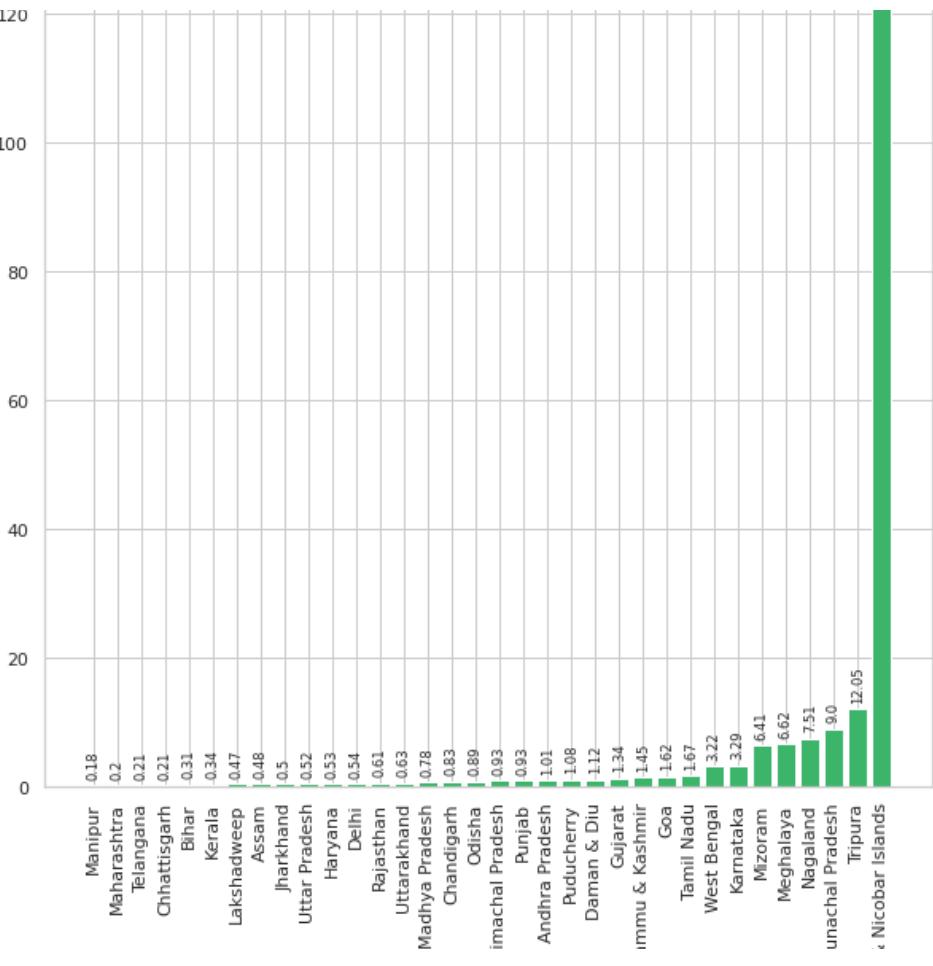


## Exploratory Data Analysis to understand the data better

% of population  
vaccinated



Active Beds/  
active cases



# EDA



## Exploratory Data Analysis to understand the data better

### Correlation Matrix

Feature	Correlation to the Number of Active Cases
Marginal workers population (others)	0.801850
Literate population	0.761175
Number of households	0.735023
Female Population	0.697729
Male Population	0.691201

Table 1: The Correlation Matrix showing the strong correlations of demographic features to the number of Active Cases

# Vulnerability Metric



Performed EDA to find Vulnerability Metrics

→ Performed on District-wise Datasets for Demographics and Railway Network.

## The State Vulnerability Score

- To show the effect (score) of the pandemic on each state of India
- Calculated by the following factors:
  - % of Population vaccinated
  - The ratio of total available beds and active cases
  - No. of active cases in each state.

## The Rail-line Vulnerability Score

- To show the amount of risk involved travelling through that particular route.
- Calculated by the following factors:
  - Duration and distance of the route.
  - State Vulnerability Score: Stations of vulnerable states are riskier.
  -

# DBSCAN

Density-based spatial clustering of applications with noise (DBSCAN)

→ Performed on District-wise Datasets for Demographics and Vaccination.

## Theory:

- To capture arbitrarily shaped clusters or detect outliers, DBSCAN is used.
- Implemented to determine whether the point:
  - Core Point
  - Border Point
  - Outlier

## Algo:

- Random start point.
- If, not a core point, mark it as noise.
- Else, cluster formation starts. Point marking and addition to the cluster are done recursively.
- For unvisited points, go to step 1

# Results (DBSCAN):

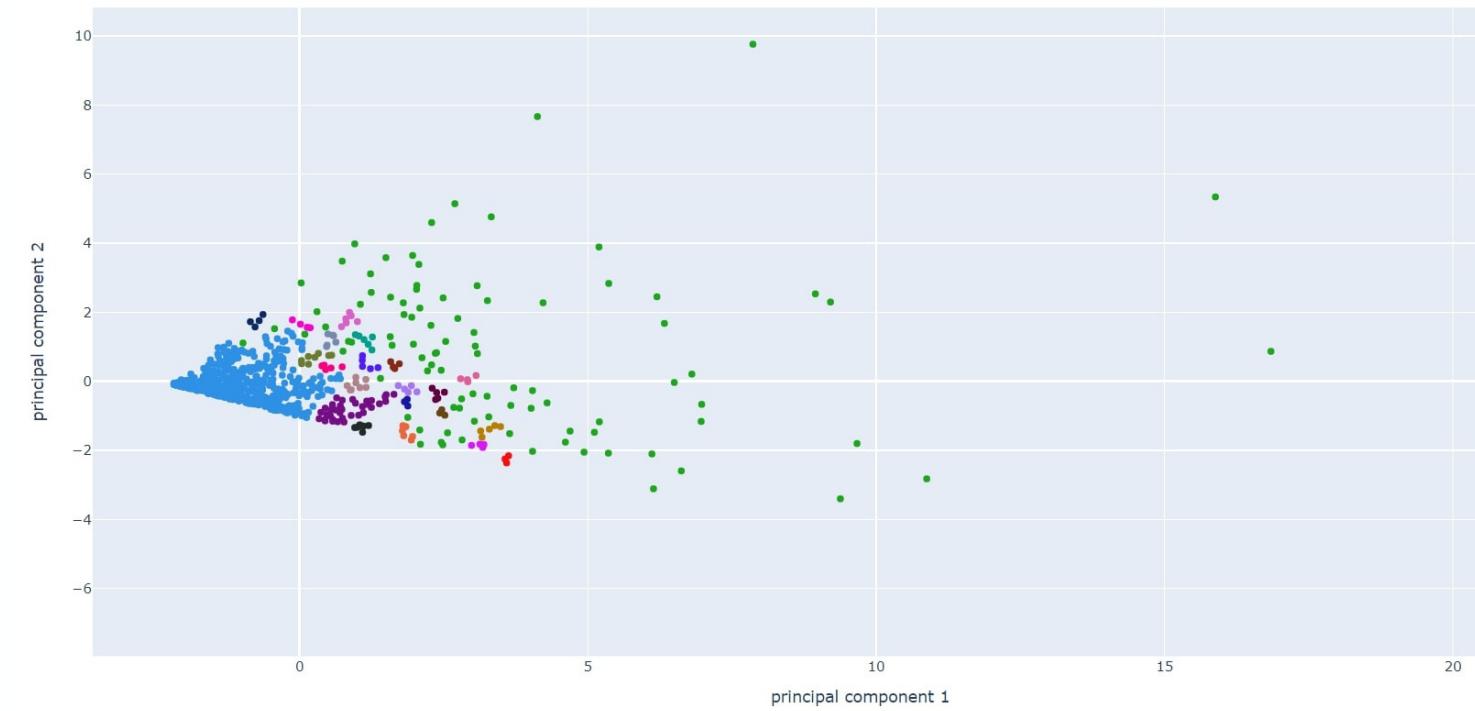
## Demographic Dataset:

- Dense cluster consists of smaller town/cities
- Outliers are large districts/cities



## Vaccination Dataset:

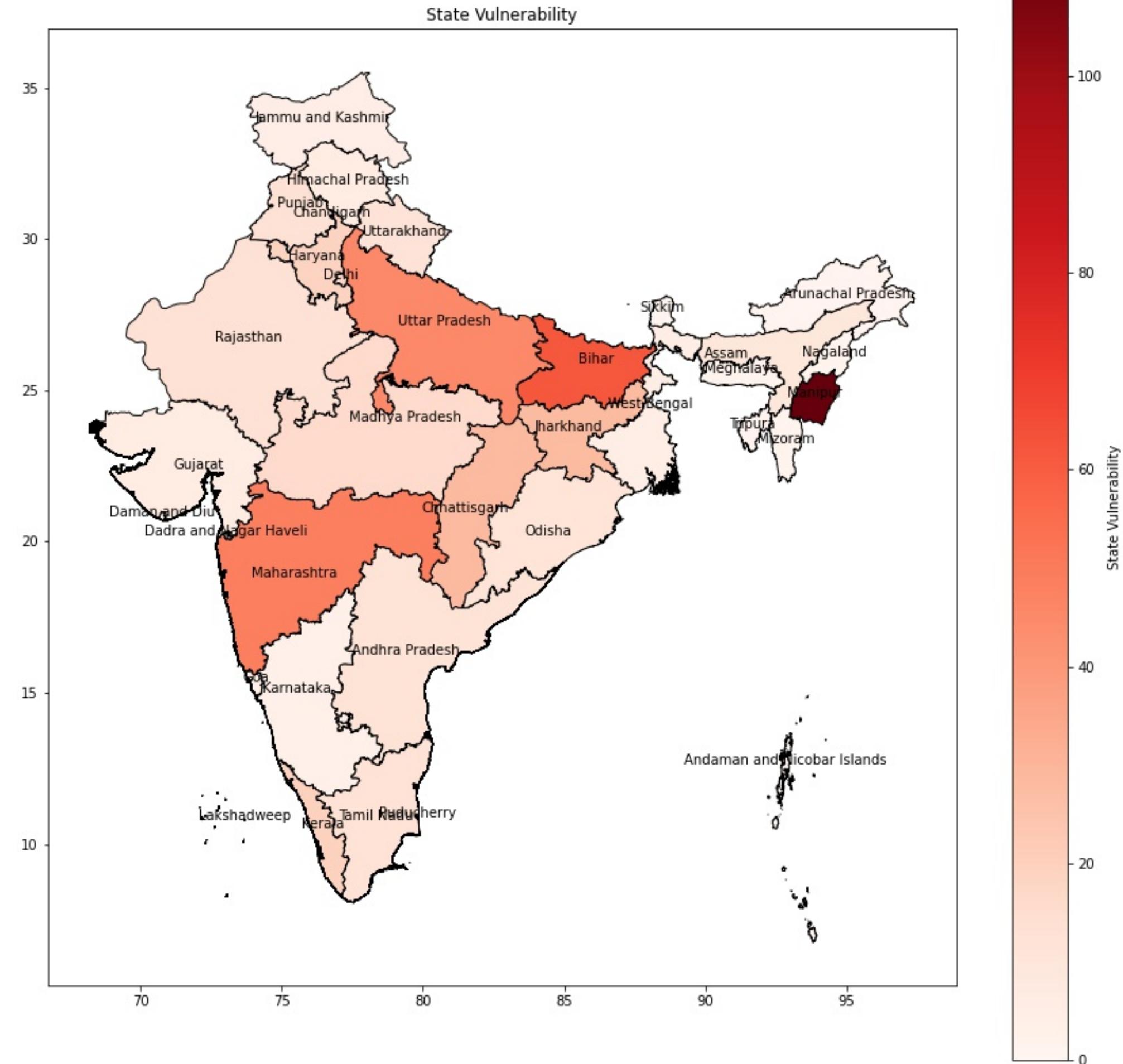
- Dense cluster consists of smaller town/cities
- Outliers are large districts/cities



# Results (EDA):

## 1. The State vulnerability metric

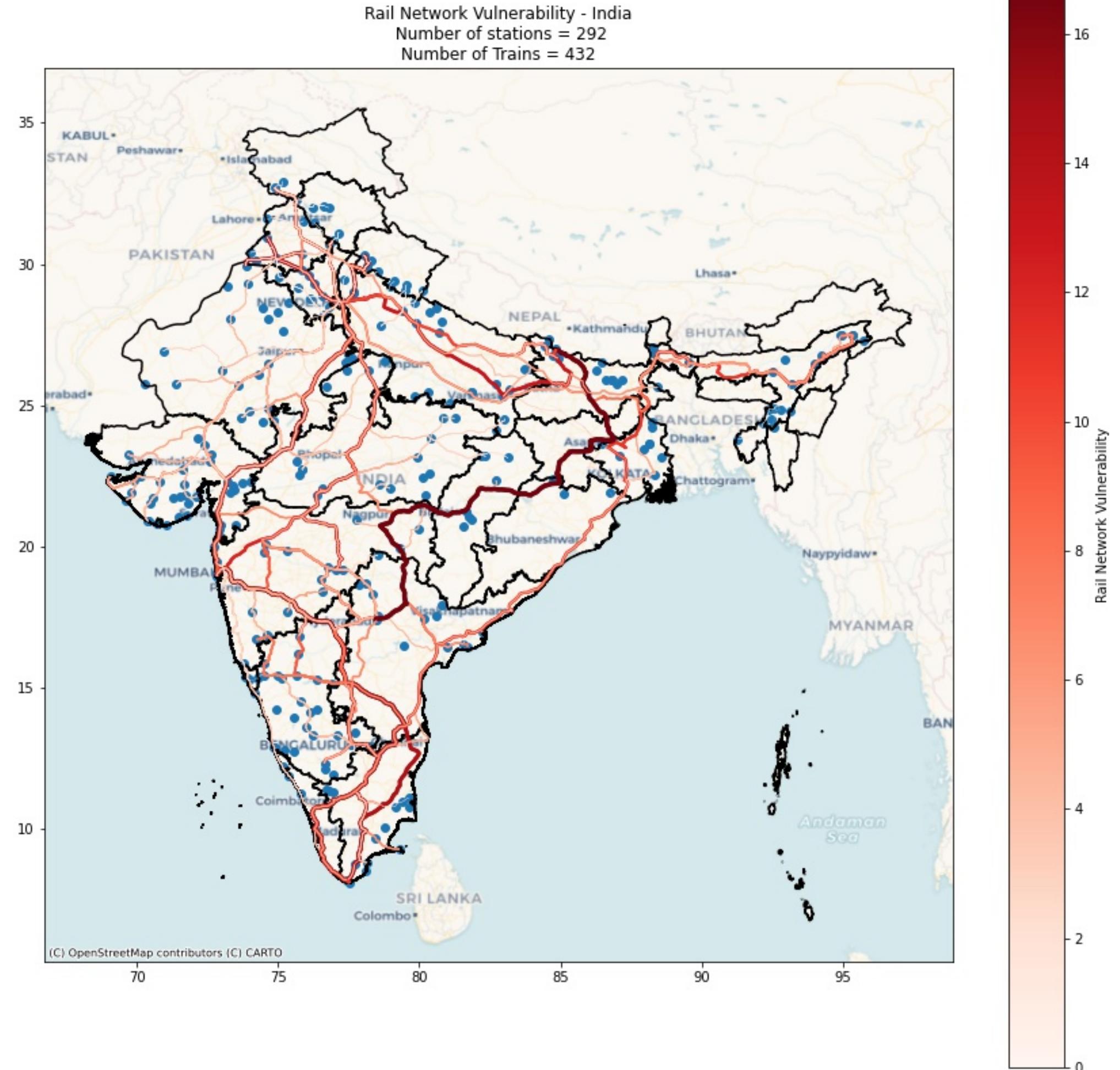
- Manipur has highest State Vulnerability Score, because the percentage of population vaccinated is quite less in Manipur (only 5.031%). Also, the ratio of total beds and the active cases in Manipur is the lowest (0.177)
- Some of the other states with high State Vulnerability Scores are Bihar, Uttar Pradesh, Maharashtra and Chhattisgarh.



# Results (EDA):

## 2. The Railway Network vulnerability metric

- As expected, we can see the positive correlation of state vulnerability of states with trains travelling from those states
- Trains with higher duration tend to have higher vulnerability metric



# DEMO

# References



- Demographics Dataset
  - <https://livingatlas.esri.in/server/rest/services/LivingAtlas/INDDemography/MapServer/0>
- Vaccination Dataset
  - <https://dashboard.cowin.gov.in/>
- Health Infrastructure Dataset
  - <https://pib.gov.in/PressReleasePage.aspx?PRID=1539877>
- Case Counts
  - <https://github.com/covid19india/api>
- Indian Railway Network
  - <https://github.com/datameet/railways>