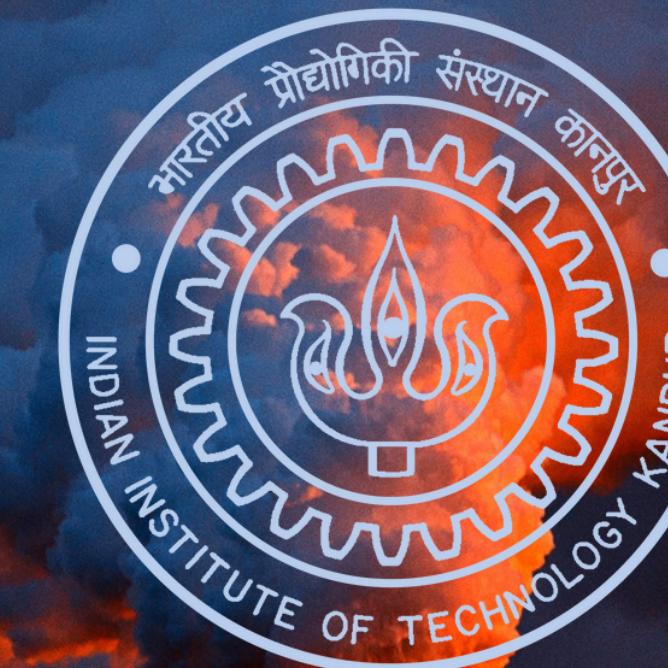


DECadal POLLUTION ANALYSIS AND IMPACT OF COVID-19



INDIAN INSTITUTE OF TECHNOLOGY
KANPUR

Presented by
Group-14

ABHISHEK SAHU	21111002
ALOK KUMAR	21111008
AMAR RAJA	21111009
ARCHIT GUPTA	21111015
T SAI KIRAN	21111061



TODAY'S OUTLINE

OUR MAIN POINTS

Problem Statement

Datasets

Analysis tools and Observation

- **Chi-Squared Test for finding Outlier**
- **Z-Score Test for spot detection**
- **Correlation**
- **Heatmaps**
- **Barplots**
- **Clustering**

Impact Of Covid19 on Pollution

Results



PROBLEM STATEMENT

OBJECTIVES OF OUR ANALYSIS

- Performing Decadal analysis of air pollution in India using various pollutant levels.
- Finding outlier states which are Hotspots and Coldspots during 2011-2020.
- Correlation between different pollutants with Number of industries, Number of motor vehicles and Population of the state.
- Finding the most polluted states in India with respect to SO₂, NO₂ and RSPM levels.
- Clustering States based on their pollution levels.

A photograph showing a vast expanse of white and light blue cumulus clouds against a clear blue sky. The clouds are fluffy and vary in size, creating a sense of depth and texture.

PROBLEM STATEMENT

OBJECTIVES OF OUR ANALYSIS

- Clustering States based on their pollution levels.
- Monthly analysis of Pollutants(NO₂, SO₂, PM) over last 5 years
- Cities with Highest Pollutant Concentrations
- Impact of Covid: Analyzing AQI of different cities over 2020.
- Comparison of AQI before and after lockdown
- Effect of Lockdown on individual pollutant levels

DATASETS

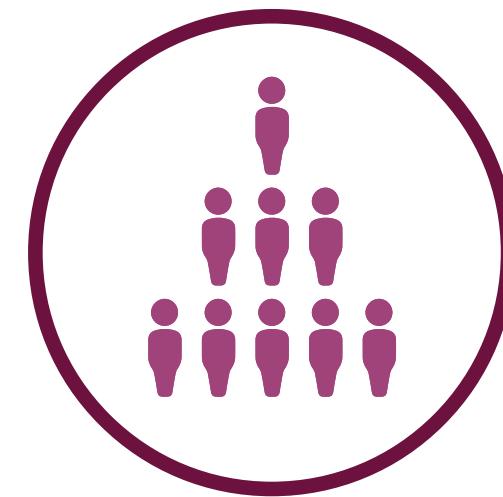
USED IN OUR ANALYSIS



AIR QUALITY DATA

- 1) 1990-2015
- 2) 2015-2020

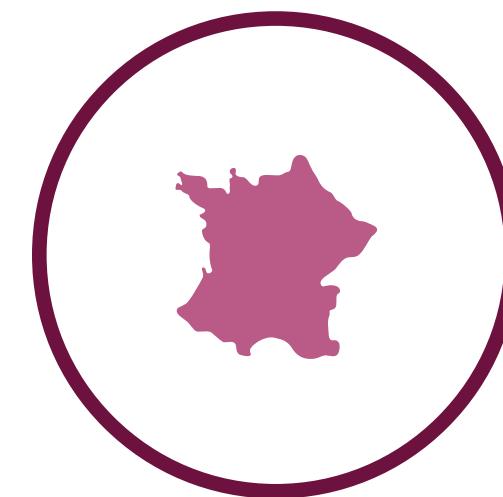
Source: Kaggle, cpcbccr



CENSUS DATA

- 1) 2001
- 2) 2011

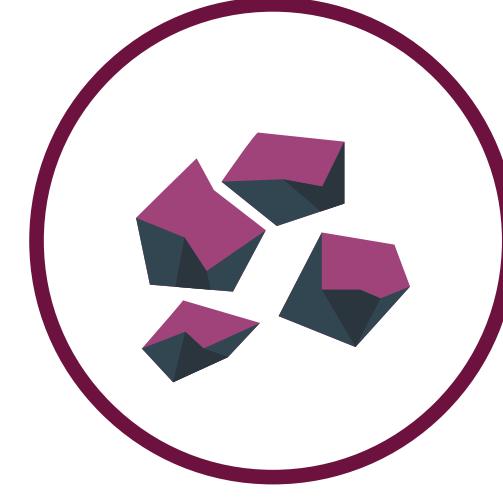
Source: censusindia.gov.in



STATE-WISE INDUSTRY DATA

- 1) 2001-2006
- 2) 2007-2008
- 3) 2009-2015

Source: labourbureau.goc.in



COAL PRODUCTION

2001-2016

Source:mospi.nic.in

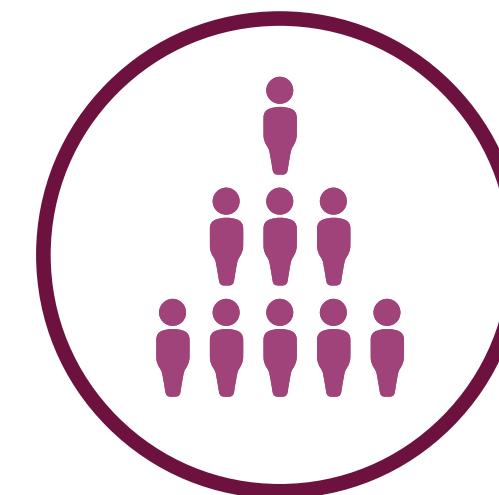
DATA PREPROCESSING

STEPS USED TO CONVERT RAW DATA TO USEFUL DATA



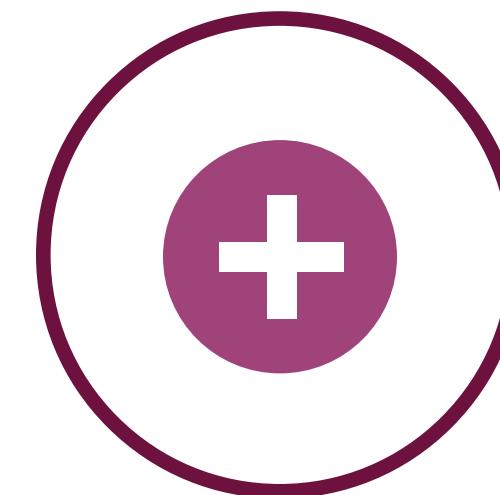
DATA CLEANING

- 1) Missing Value Handling
- 2) Data Type Consistency



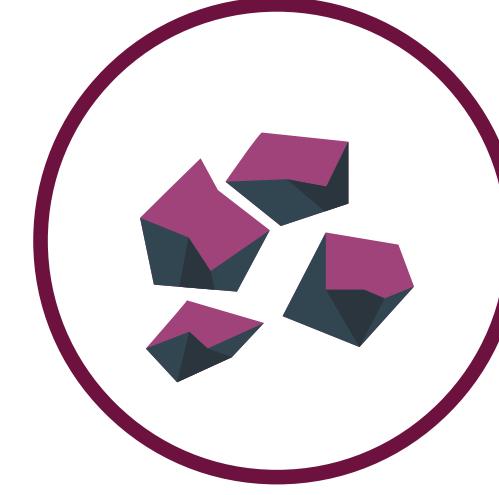
INTERPOLATION

- 1) Use of Interpolation technique to Handle Missing Data



EXTRAPOLATION

- 1) Use of Extrapolation technique to generate new Data based on the growth rate



DATA INTEGRATION

- Merge data with proper attributes

Chi-Squared Test

- Used Chi-Squared Test to detect outliers
- Formula used:

$$\chi^2 = \sum_{i=1}^N \frac{(x_i - e_i)^2}{e_i}$$

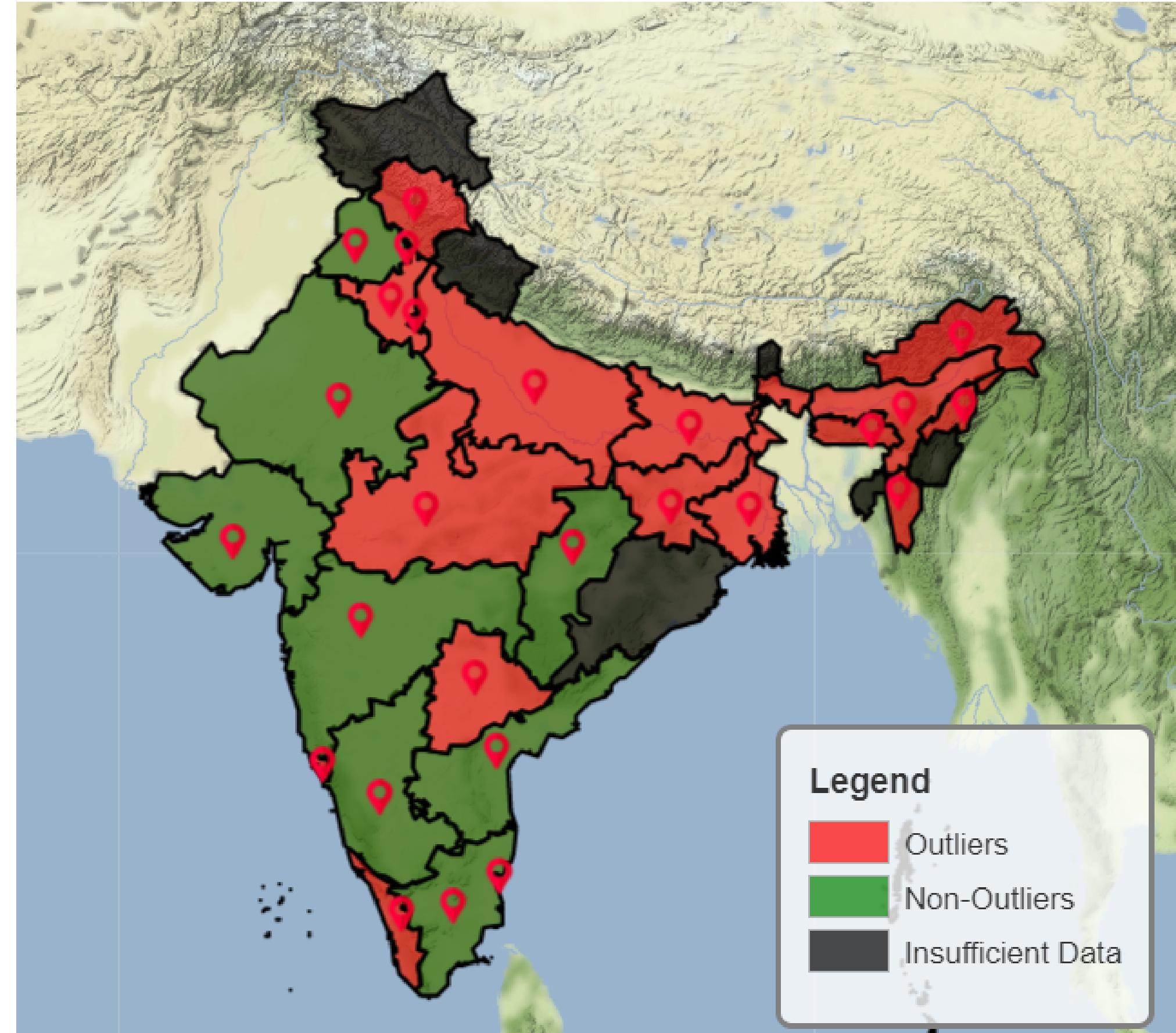
where x: object to be tested

xi: value of x in ith dimension

Ei: mean value on ith dimension among all objects

- H0: State is not an outlier
 - H1: State is an outlier
-
- A State is considered as an outlier if its p-value is less than it's level of significance (1%).

Map To Visualize Outlier using Chi-Squared Test



Observation From Chi-Squared Test

- The Map of India shown in the Figure shows that **Himachal Pradesh, Haryana, Uttar Pradesh, Madhya Pradesh, Bihar, Jharkhand, West Bengal, Assam, Meghalaya, Mizoram, Nagaland, Arunachal Pradesh, Telangana, Goa, Kerala** are outliers. But using the Chi-Square test for outlier detection, we can't determine whether the outlier state is Hotspot or Coldspot
- A lot of the Northern part fall under this category. States like **Uttar Pradesh, Haryana, Madhya Pradesh** may be hotspots because they are **highly polluted States** of India, and the average temperature of the States are on higher sides.
- On the other hand, states like **Himachal Pradesh and Northeastern states like Assam, Meghalaya, Arunachal Pradesh**, etc., can be coldspots as they are the Himalayas and Mountain regions.
- The states like **West Bengal and Kerala** might be an outlier as recently they have been hit by multiple **natural calamities like Cyclone, Flood**, etc. To get an accurate picture of the Hotspot and Coldspot, we will test with a Z-score.

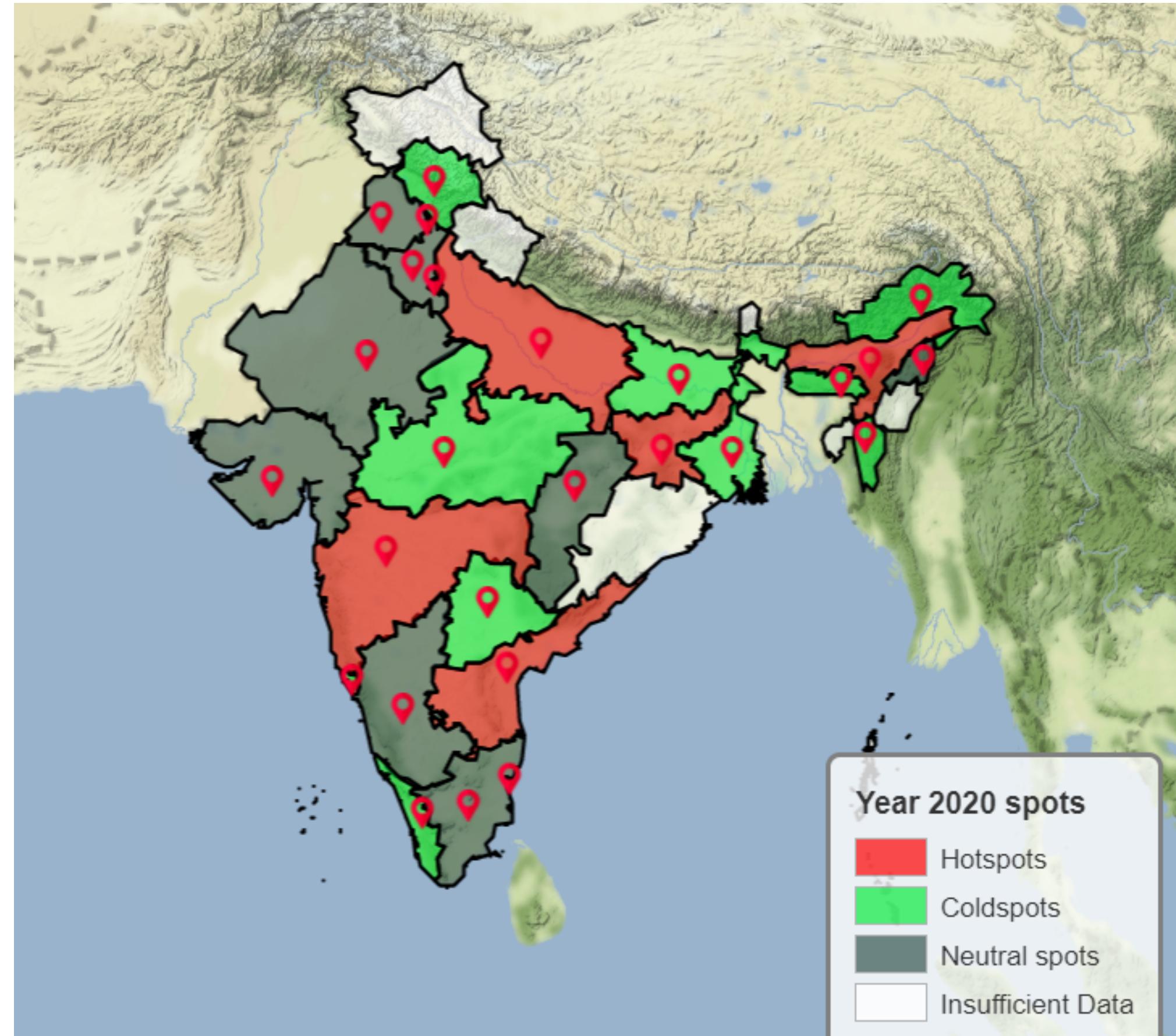
Z-Score Test

- Used Z-Score Test to detect Spot (i.e Hotspot or Coldspot)

$$\text{Mean Pollutant Concentration} = \frac{SO_2 \text{ conc.} + NO_2 \text{ conc.} + RSPM \text{ conc.}}{3}$$

- Since it's possible that air pollution of a state is also affected by the increasing air pollution level of its neighboring states, we have defined a state as hotspot or coldspot taking into consideration the pollution level of its neighbors.
- A state is Hotspot if: $MPC_{\text{state}} > \text{Mean}_{\text{neighbor}} + \frac{1}{2} \text{ std}_{\text{neighbor}}$
- A state is Coldspot if: $MPC_{\text{state}} < \text{Mean}_{\text{neighbor}} - \frac{1}{2} \text{ std}_{\text{neighbor}}$

Map To Visualize Outlier using Z-score Test

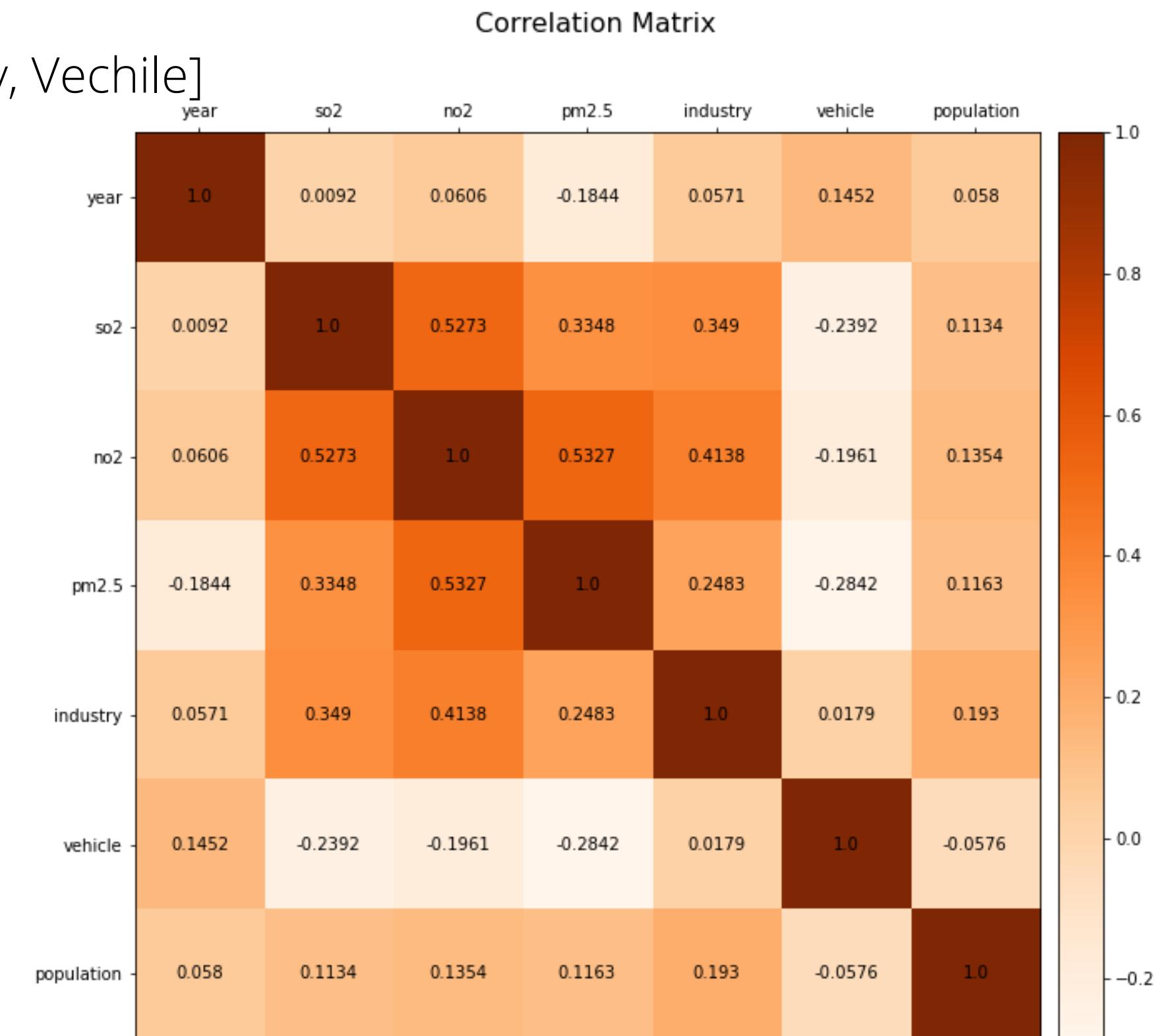


Observation From Z-Score Test

- The Map of India shown in the Figure shows that **Uttar Pradesh, Maharashtra, Jharkhand, Assam, Andhra Pradesh** are the Hotspots. A possible reason could be the large population and the ongoing Industrial works which might increase the temperature.
- Whereas in **Assam**, it might happen that due to a **low-pressure system** over the Bay of Bengal, there is less moisture which results in **dry weather**.
- Also, States like **Himachal Pradesh, Bihar, West Bengal, Arunachal Pradesh, Meghalaya, Mizoram, Nagaland, Madhya Pradesh, Telangana, Goa and Kerala** are some of the cold spots. States having mountains and Himalayas are known to be cold spots, whereas, in **Goa, Kerala**, the overall temperature becomes low due to heavy **rainfall**.
- Other states like **Rajasthan, Gujrat, Karnataka, Tamil Nadu**, etc., are in neutral spots as the average temperature of these States remains almost constant because they are hotter in the day and colder at night.
- Again we also have some States like **Odisha, Sikkim**, etc., where we don't have enough pieces of evidence and data to support our hypothesis.

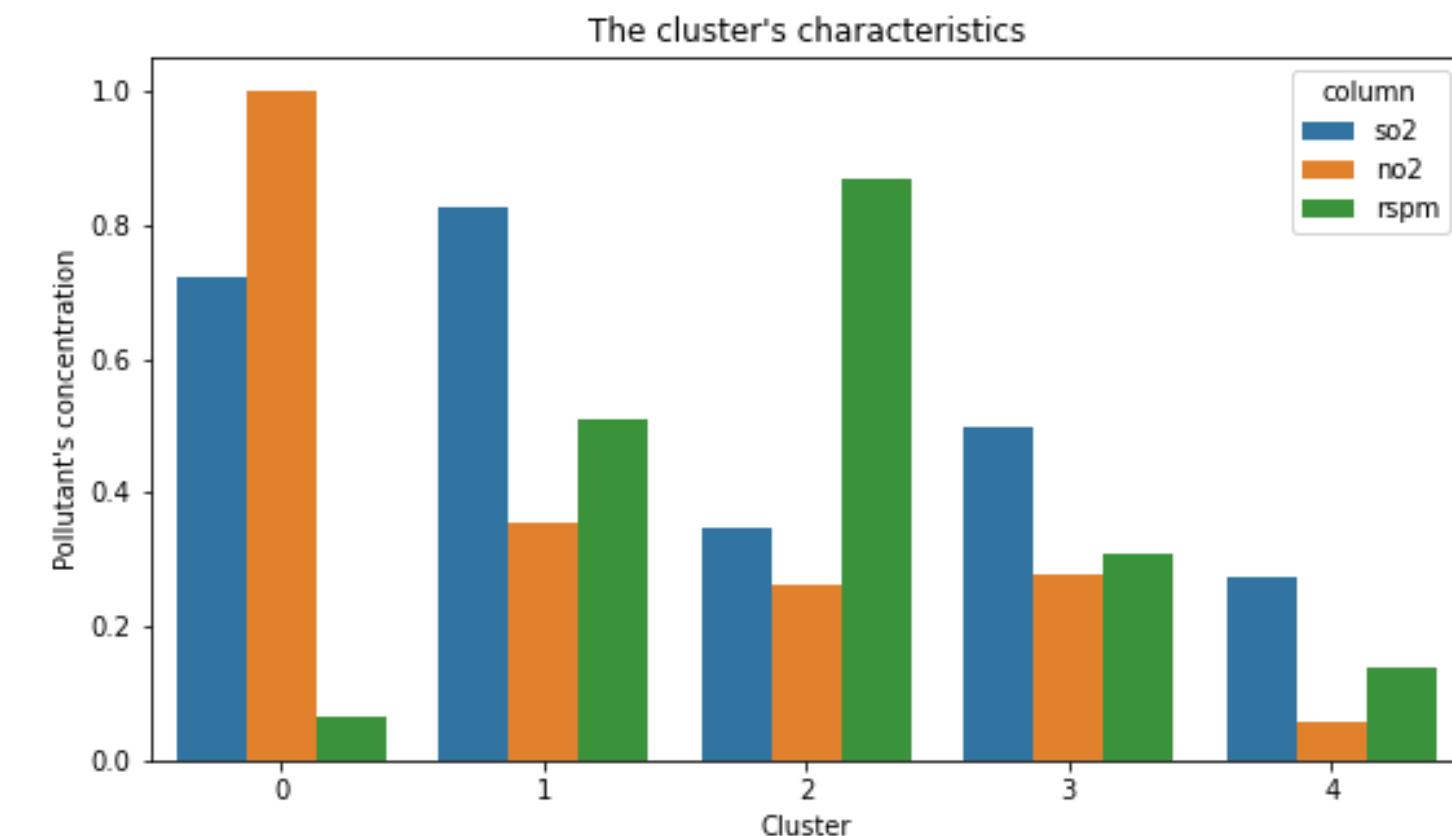
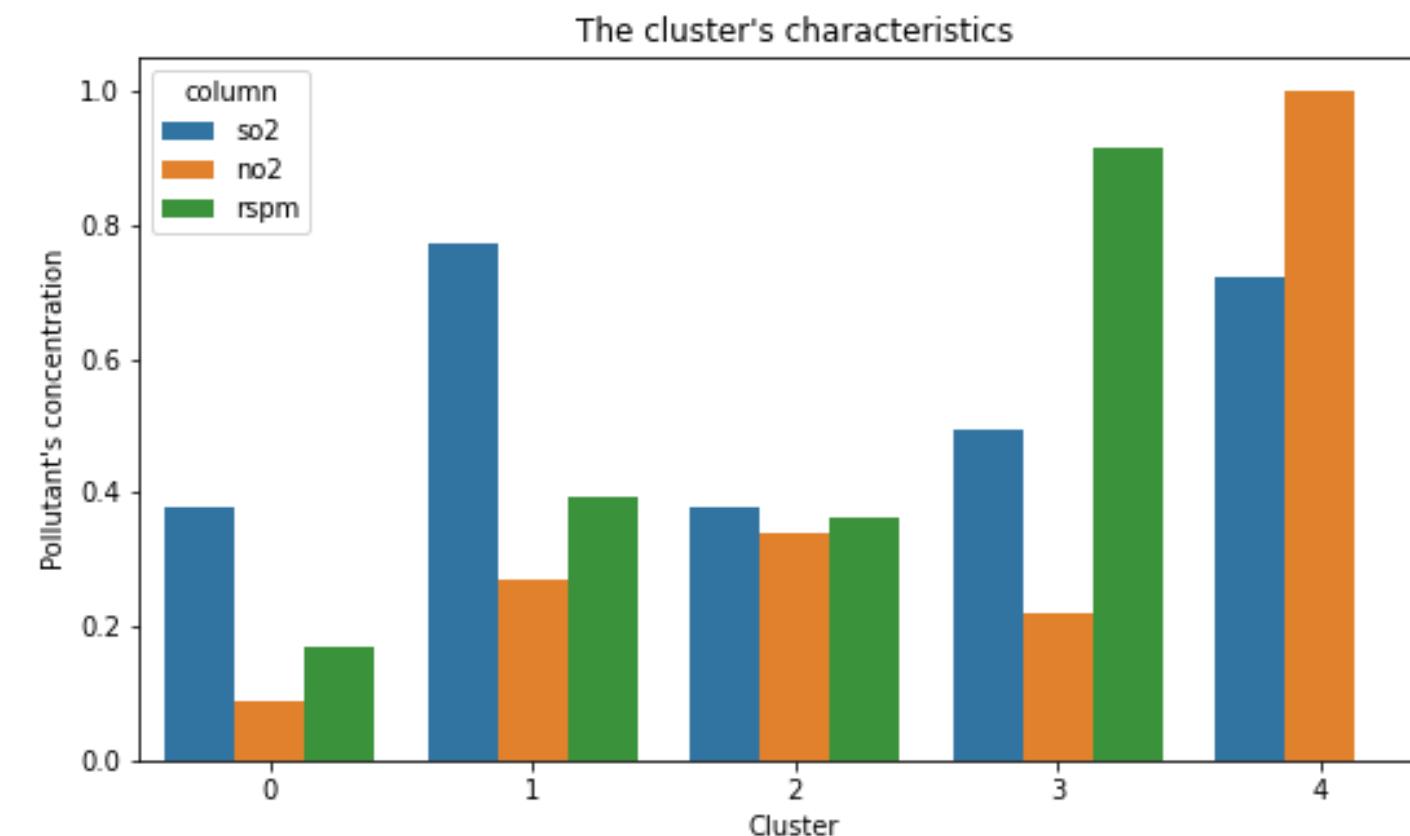
Spearman's Correlation

- More generic than Pearson
- Correlation of [SO₂,NO₂,PM] vs [Industry, Vechile] also seems natural.
- Strong Correlation observed between:
 1. PM and NO₂
 - 2.SO₂ and NO₂
 - 3.NO₂ and Industry



Clustering

- We grouped the states based on SO₂, NO₂, and PM_{2.5} levels using Air Pollutant Concentration data from 2014.
- This is done using K-means algorithm with Euclidean distance to group data with identical features.



Clustering

Cluster	States-2019	States-2020	Cities-2019	Cities-2020
0	Andhra Pradesh Chandigarh, Chhattisgarh Madhya Pradesh Maharashtra, Odisha Punjab, Rajasthan Tamil Nadu, Telangana West Bengal	Andhra Pradesh, Goa Himachal Pradesh Karnataka, Kerala Maharashtra, Meghalaya, Mizoram Puducherry, Rajasthan West Bengal	Visakhapatnam Bhopal, Jaipur Hyderabad, Kolkata	Chandigarh, Kochi Thiruvananthapuram Shillong, Aizawl Amritsar, Chennai
1	Arunachal Pradesh Bihar Dadra & Nagar Haveli Daman & Diu Delhi Gujarat Jharkhand Uttarakhand	Arunachal Pradesh,Dadra & NagarHaveli Daman & Diu,Delhi Gujarat, Jharkhand Uttarakhand	Bengaluru Kochi Thiruvananthapura m, Mumbai, Shillong Aizawl Chennai Coimbatore	Guwahati, Talcher
2	Haryana	Bihar, Chandigarh, Chhattisgarh, Madhya Pradesh, Odisha, Punjab, Tamil Nadu, Telangana	Ahmedabad	Visakhapatnam Bengaluru, Bhopal Mumbai, Jaipur Coimbatore Hyderabad, Kolkata
3	Assam, Jammu & Kashmir, Nagaland, Uttar Pradesh	Haryana	Guwahati, Chandigarh, Talcher, Amritsar	Ahmedabad
4	Goa, Himachal Pradesh, Karnataka Kerala Meghalaya Mizoram, Puducherry	Assam, Jammu & Kashmir, Nagaland, Uttar Pradesh	Patna, Delhi, Lucknow	Patna, Delhi, Lucknow

HeatMaps for SO₂ Concentration City

From the heatmap of cities, we can say

- Ahmedabad has the highest SO₂ concentration in the year 2019.
- Talcher of Odisha has higher in 2016.
- Shillong of Meghalaya and Aizwal have the most negligible SO₂ concentration.



HeatMaps for SO2 Concentration State

From 2011-2020,

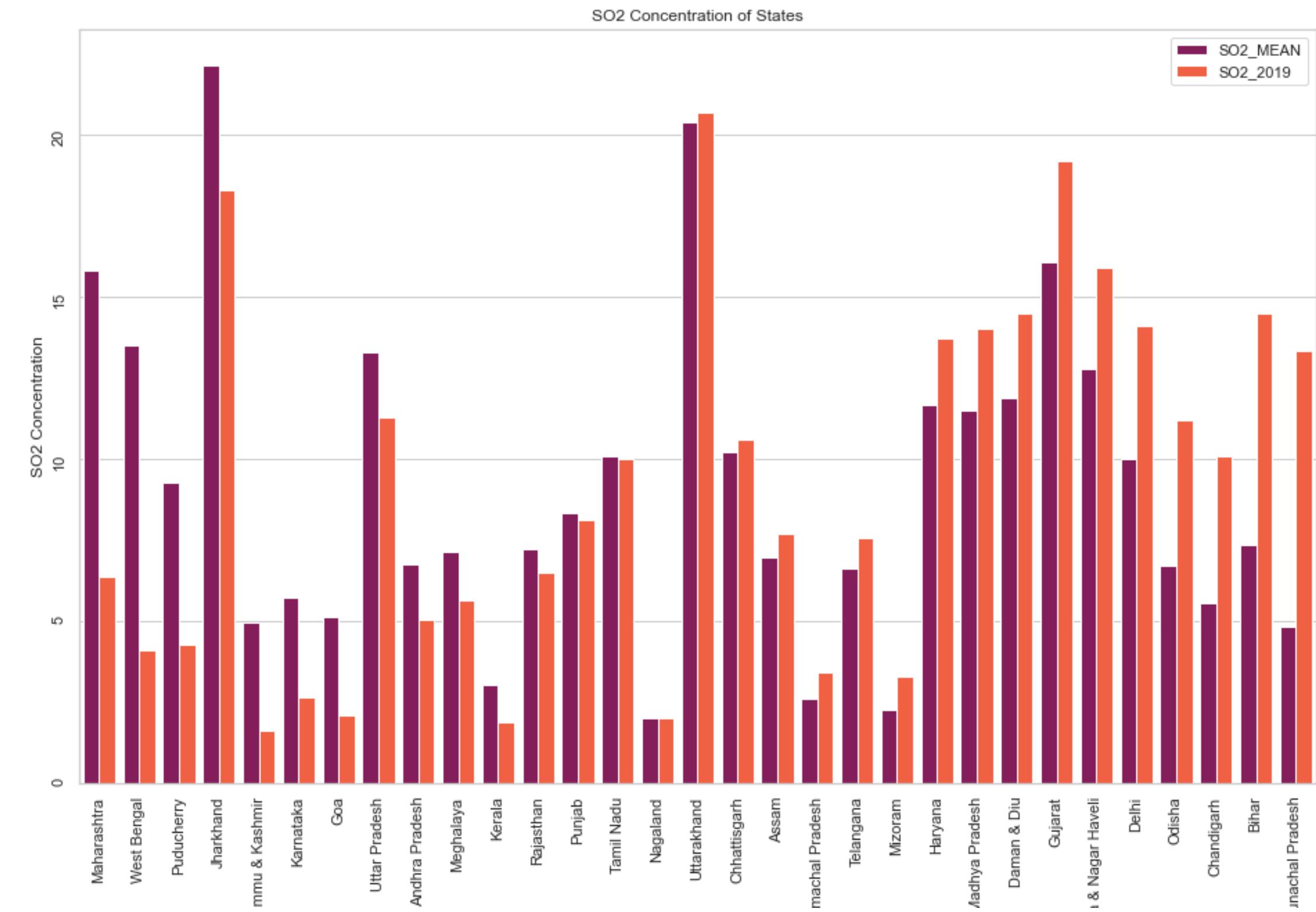
- Uttarakhand, Maharashtra, and Jharkhand paid the most attention to SO2.
- For Meghalaya, we can see a lot of bounce within the SO2 attention in 2013, 2018, and 2011, and a lot of attention in 2011.
- In the last ten years, Himachal Pradesh, Manipur, Mizoram, and Nagaland have paid substantially less attention to SO2.
- For ten years, Himachal Pradesh, Manipur, Mizoram, and Nagaland have displayed much less attention for SO2. Himachal Pradesh, Manipur, Mizoram, and Nagaland have paid far less attention to SO2 over the last ten years.



HeatMaps for SO₂ Concentration

In the above barplot,

- Jharkhand has the best SO₂ attention for 2019.
- Nagaland has the least SO₂ attention for 2019.
- We can see that the degrees of SO₂ attention for Jharkhand and Uttarakhand are nearly identical, putting Uttarakhand in second place.
- Nagaland, Mizoram, and Kerala have SO₂ levels that are almost identical.



HeatMaps for NO₂ Concentration City

From the heatmap of cities,
we can say

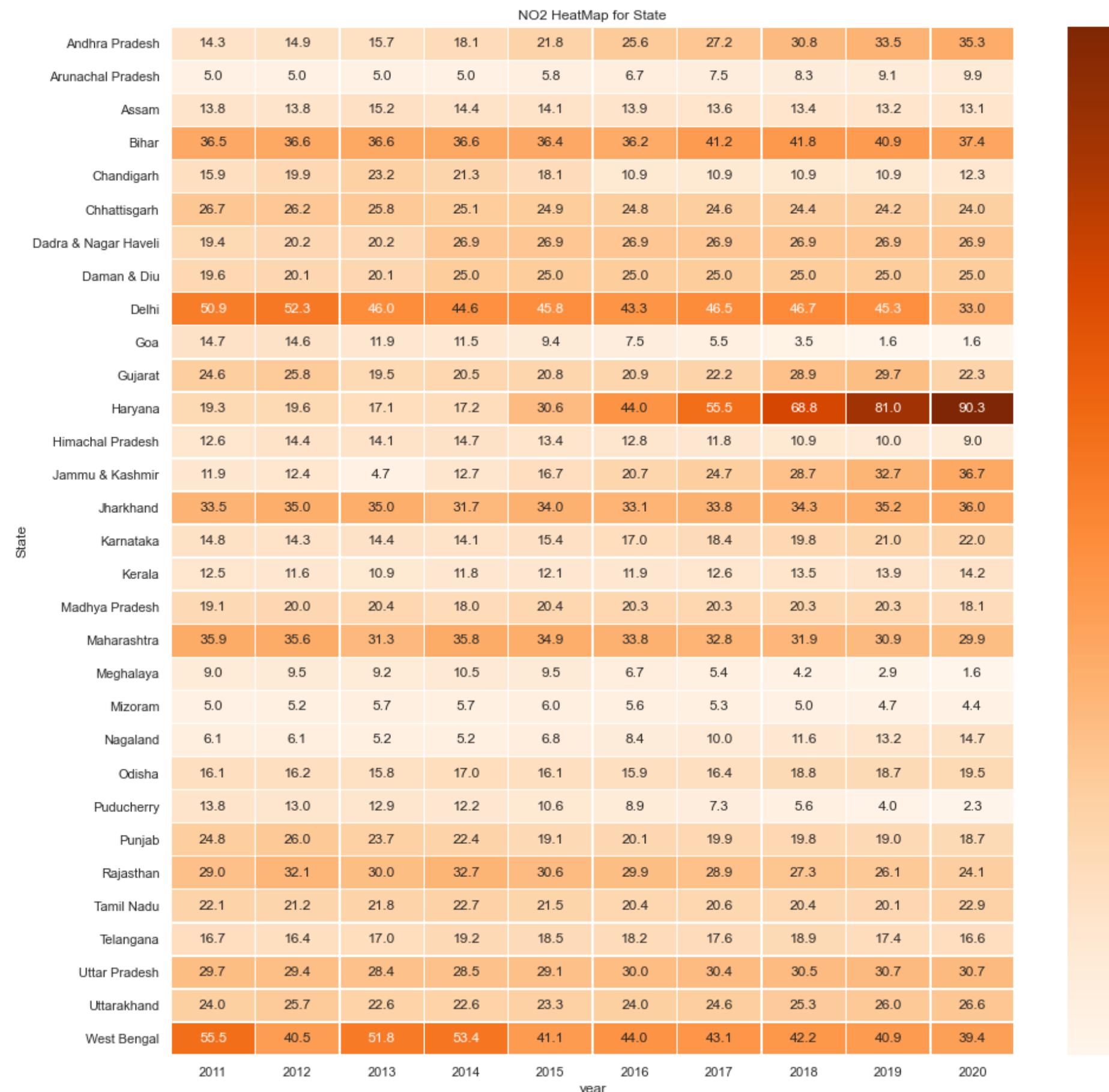
- Ahmedabad has the highest NO₂ concentration in the year 2019.
- Shillong of Meghalaya and Aizawl have the most negligible NO₂ concentration.



HeatMaps for NO₂ Concentration State

From 2011-2020,

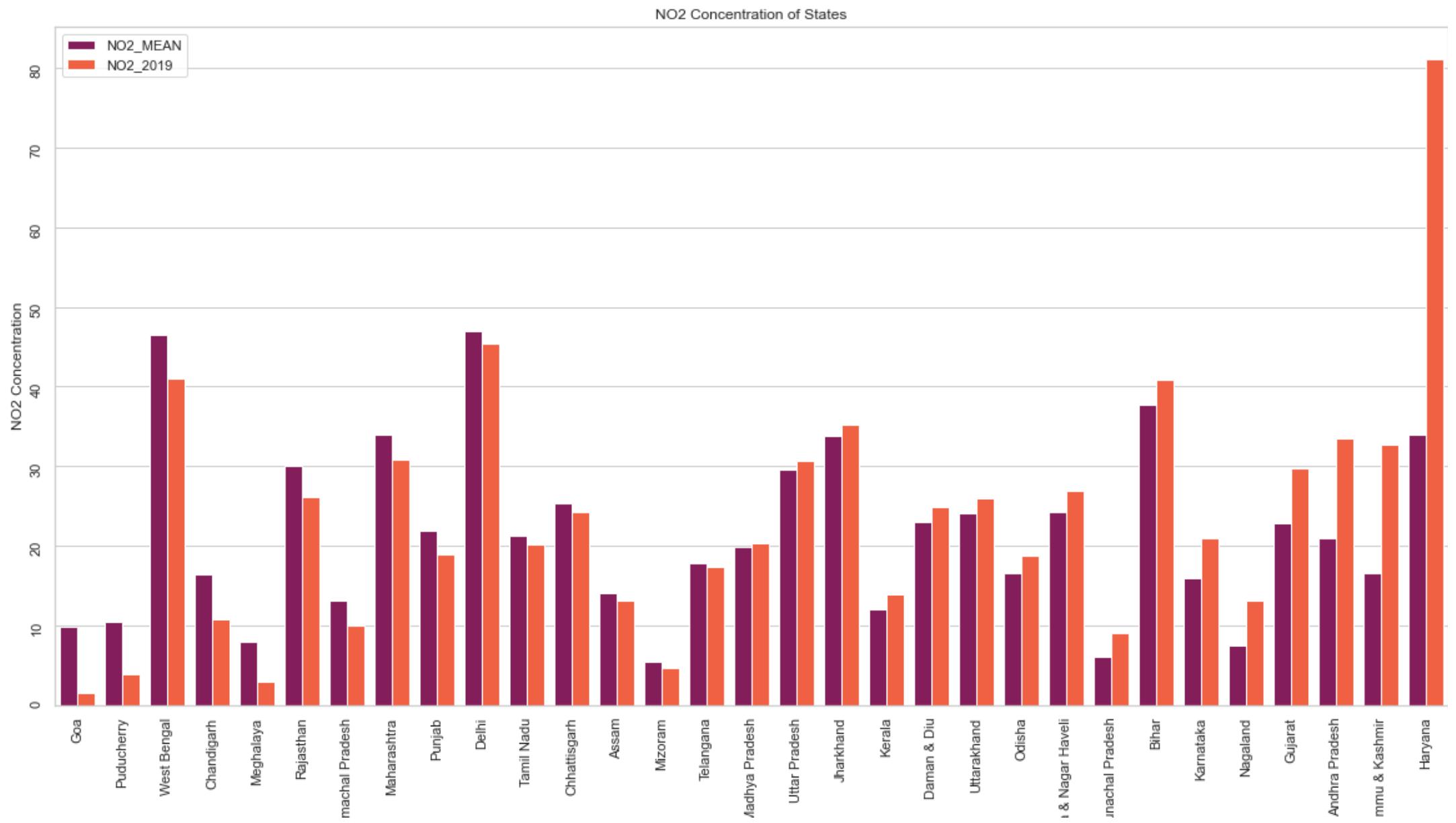
- Haryana has a somewhat high NO₂ concentration from 2015-2020.
- We can see that the NO₂ concentrations in Jammu and Jharkhand are very similar to those in Delhi.
- For the years 2016-2020, Nagaland, Mizoram, and Meghalaya have the lowest NO₂ concentrations.
- For the year 2019, NO₂ levels in Jharkhand have risen dramatically.



HeatMaps for NO₂ Concentration

In the above barplot,

- Haryana has initiated NO₂ concentration in 2019 more significant than the mean.
- In terms of NO₂ concentrations, Delhi and West Bengal are in second place.
- NO₂ concentrations are modest in Jharkhand, Maharashtra, and Bihar.
- In the same way that Goa has the lowest SO₂ concentration, it also has the lowest NO₂ concentration.



HeatMaps for PM Concentration City

From the heatmap of cities,
we can say

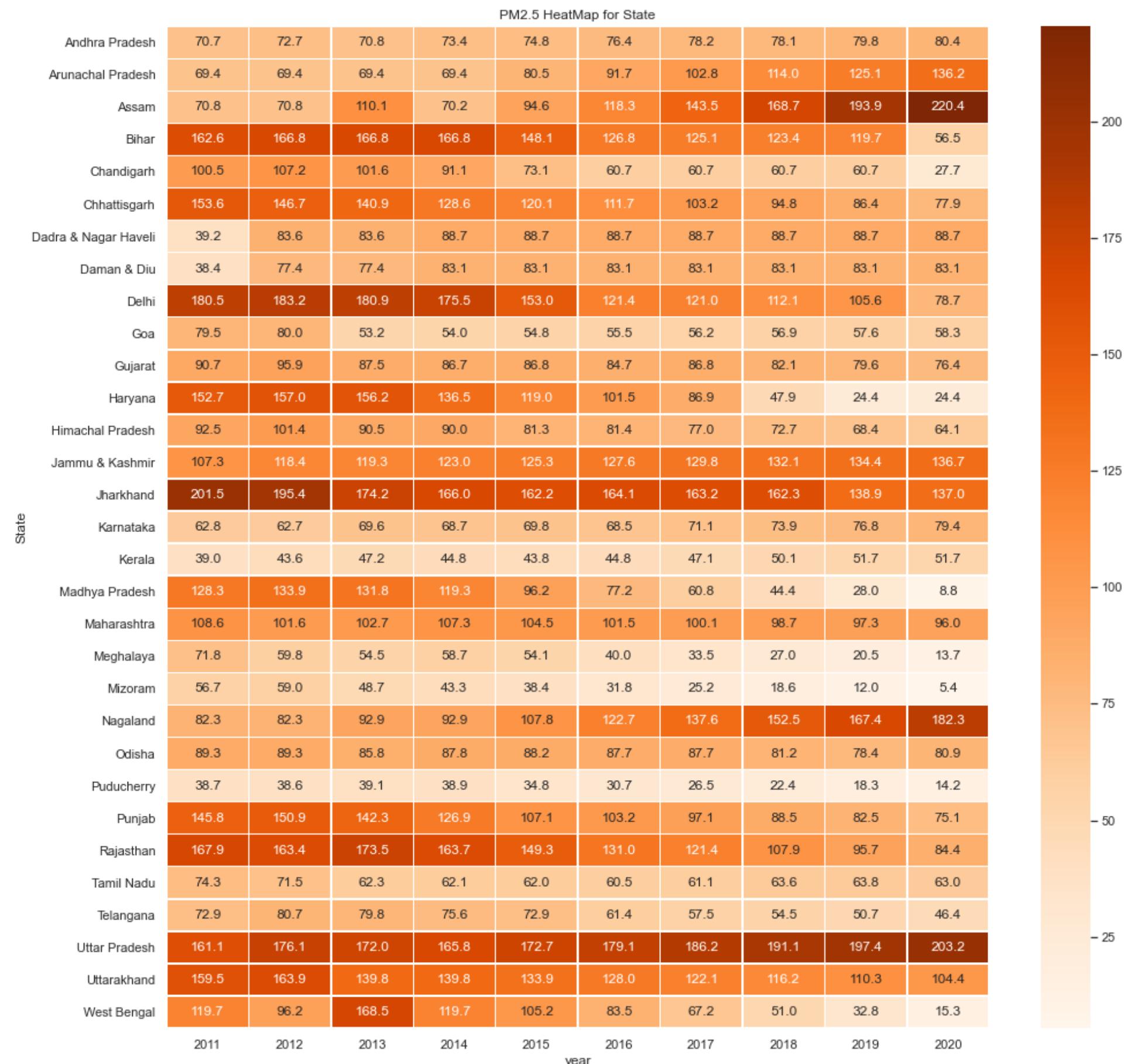
- Bhopal had the highest PM2.5 concentration in 2013
- Aizawal had the lowest.



HeatMaps for PM Concentration State

From 2011-2020,

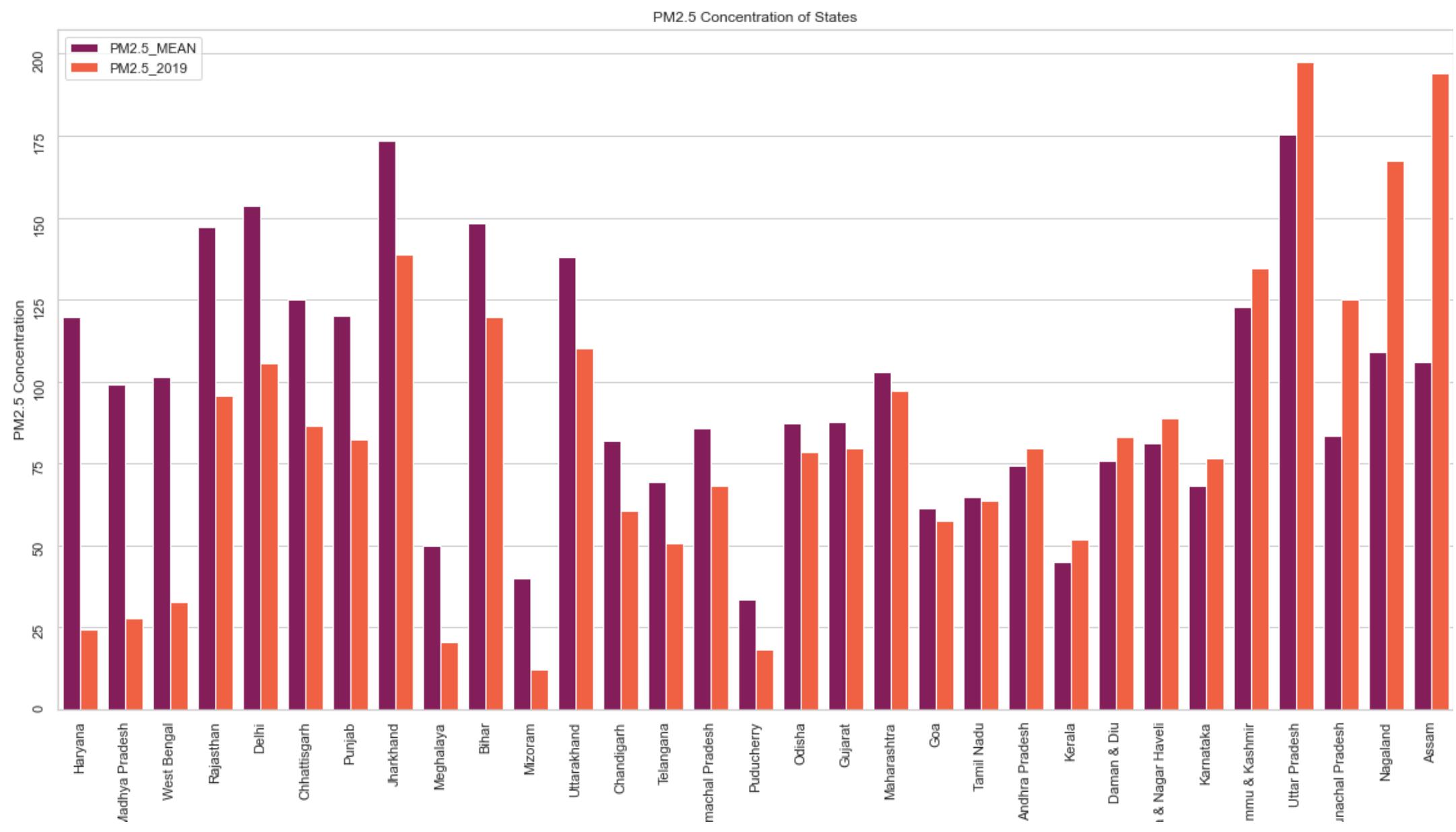
- Assam and Uttar Pradesh have had high PM2.5 concentrations for the past ten years.
- After 2015, the concentration of PM2.5 in Nagaland increased dramatically.
- In Jharkhand, there was a sharp increase in PM2.5 concentration in 2012, followed by moderate to high PM2.5 concentration.
- In addition, PM2.5 concentrations in Punjab, Uttarakhand, and Puducherry have gradually decreased.



HeatMaps for PM Concentration

In the above barplot,

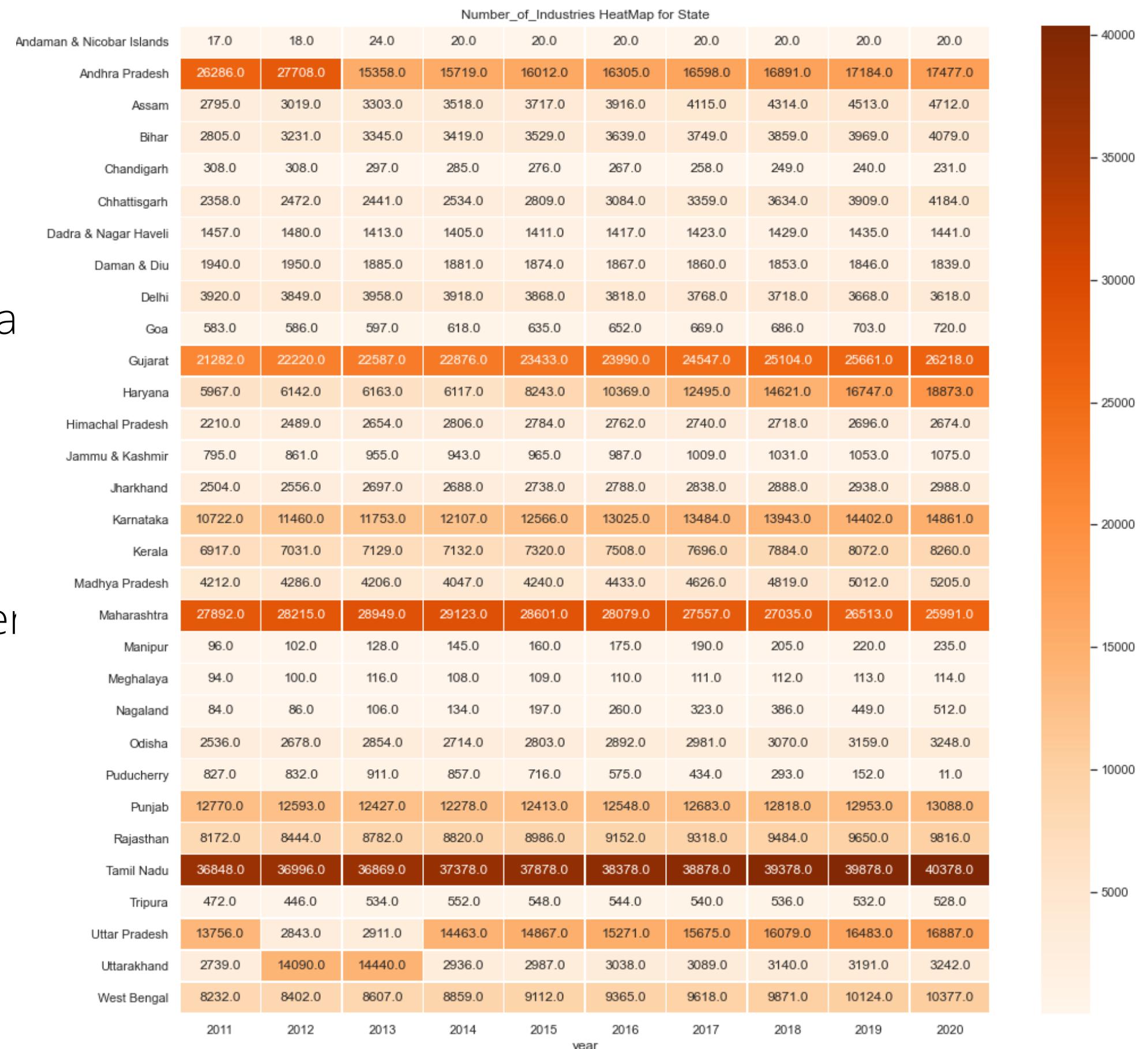
- Uttar Pradesh has the highest PM2.5 levels.
 - Uttar Pradesh and Jharkhand, which are placed second and third, have practically identical PM2.5 mean concentration values but they are in opposite directions through the years while Uttar Pradesh kept on increasing till 2019 while Jharkhand gradually decreased to 2019.
 - Uttar Pradesh is the most densely inhabited, immediate action is required to control pollution levels in these states. It's also worth noting that Puducherry has the lowest PM2.5 levels.



HeatMaps for Industry

From the heatmap,
we can say

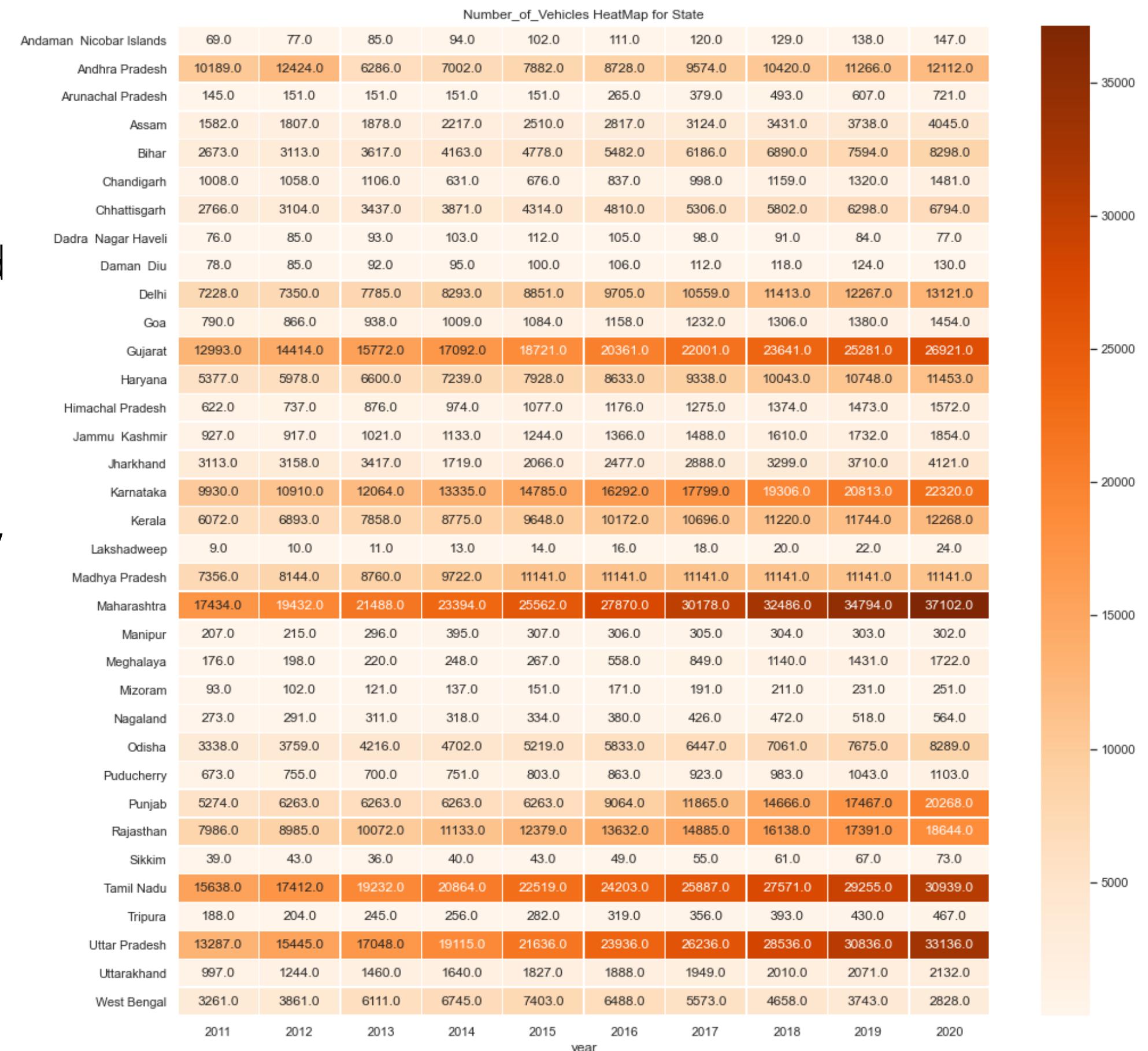
- Tamil Nadu, Gujrat and Maharashtra have the highest number of Industries.
- Puducherry, Manipur, Chandigarh, Meghalaya have the lowest number of industries.



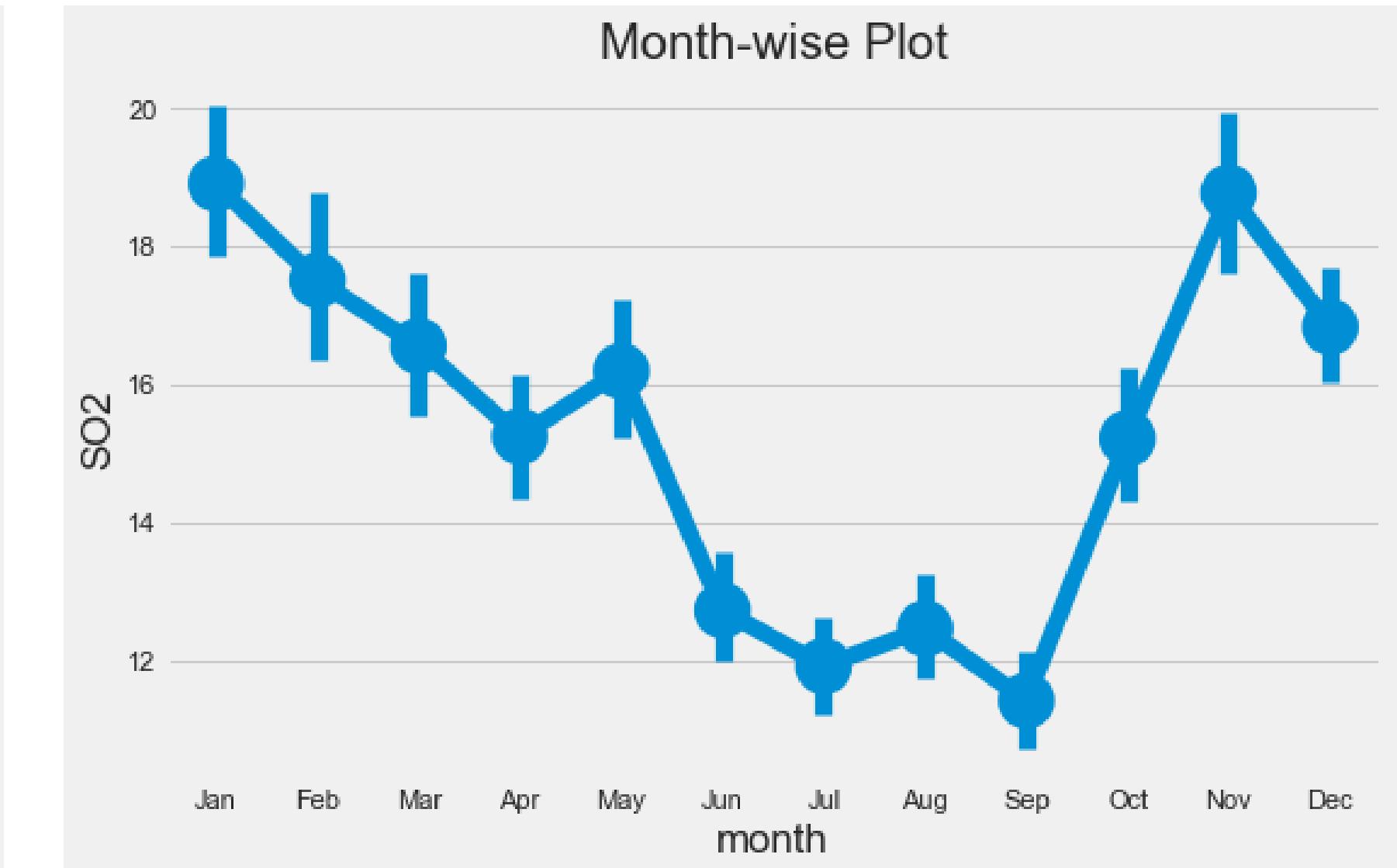
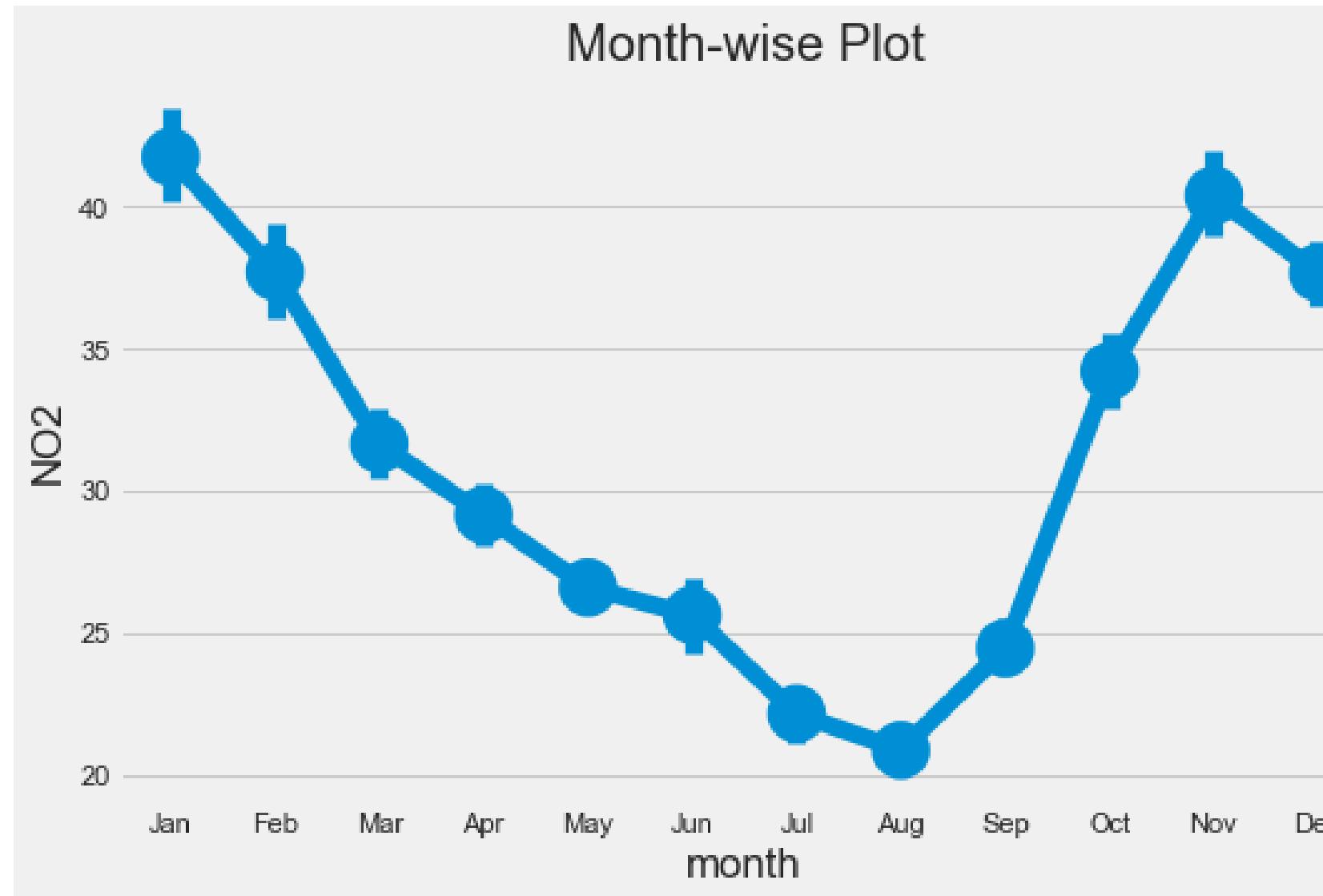
HeatMaps for Vehicles

From the heatmap,
we can say

- Tamil Nadu, Andhra Pradesh and Maharashtra have the highest number of Vehicles.
- Lakshadweep, Dadra Nagar Haveli, Sikkim have the lowest number of Vehicles.

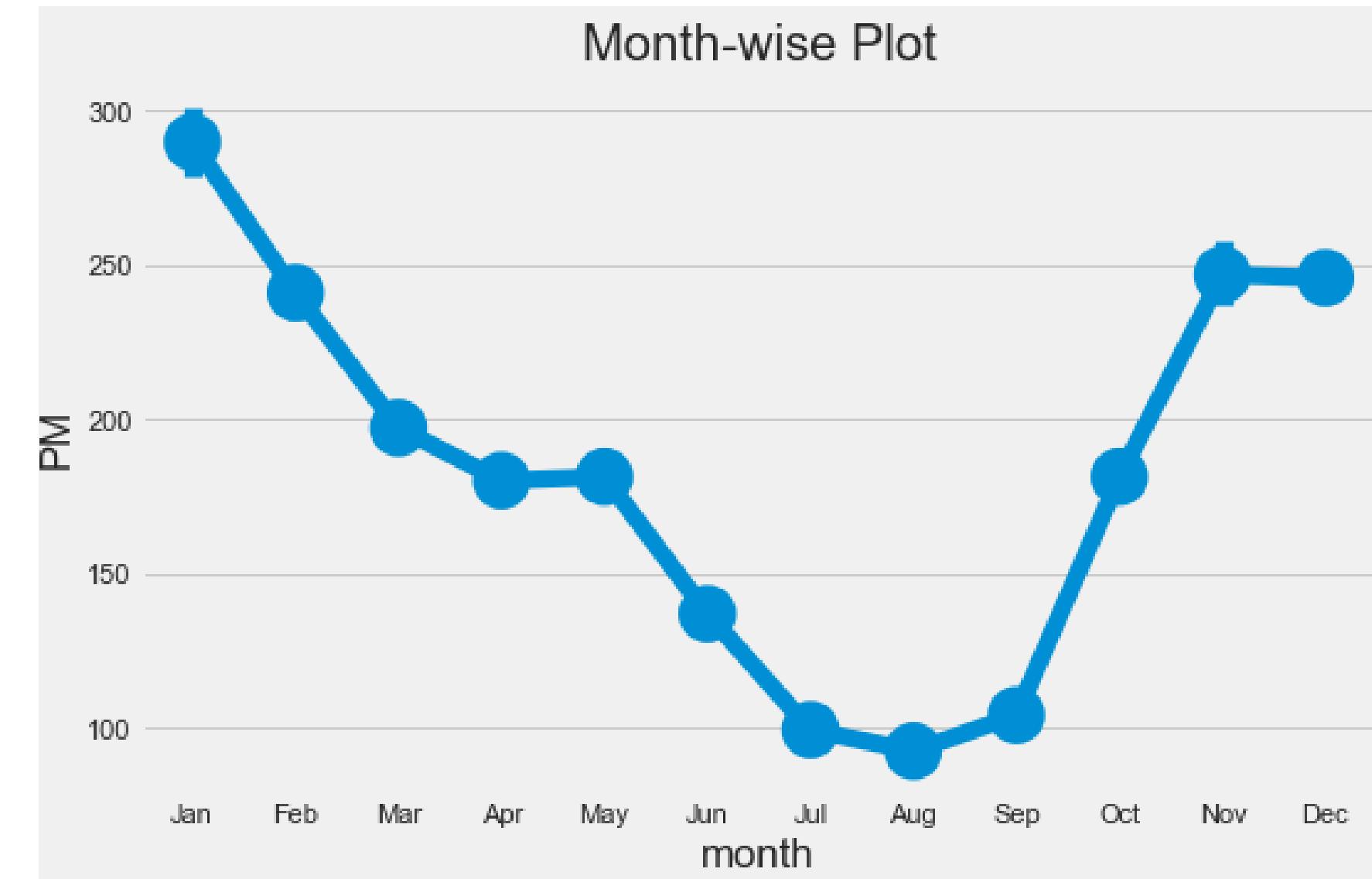


Monthly analysis of Pollutants



Monthly analysis of Pollutants

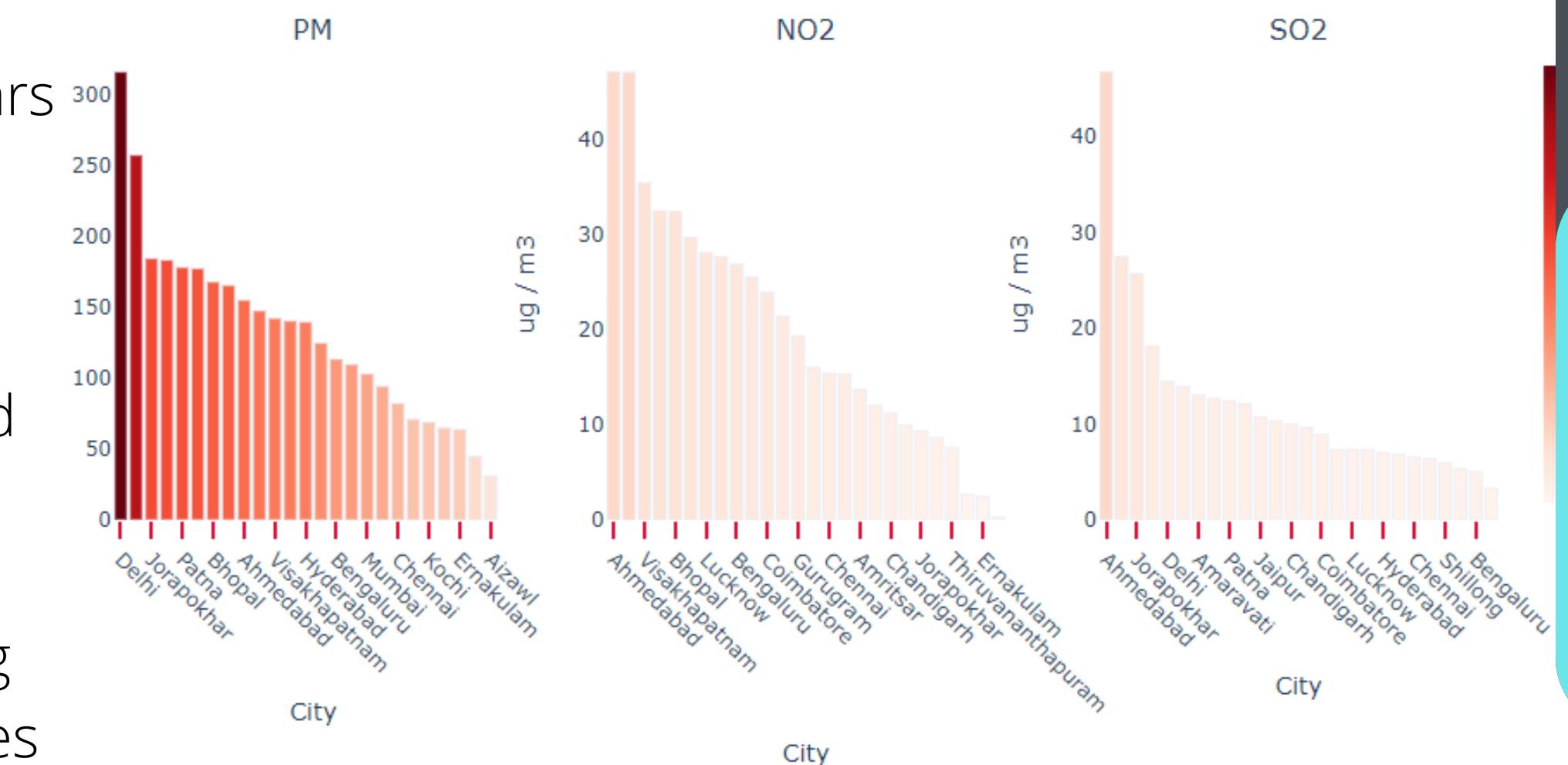
- From these three plots, we can clearly see that the pollution level in India falls in the Month of July-September. This might happen due to the monsoon season. During these days, less number of people move from their home and overall Traffic in the cities is reduces.
- The pollution level then starts rising and reach the highest levels in the winter months. Again, it's during these months that a lot of crop residue burning takes place, especially in northern parts of India.



Cities with Highest Pollution Concentration

From the above plot,

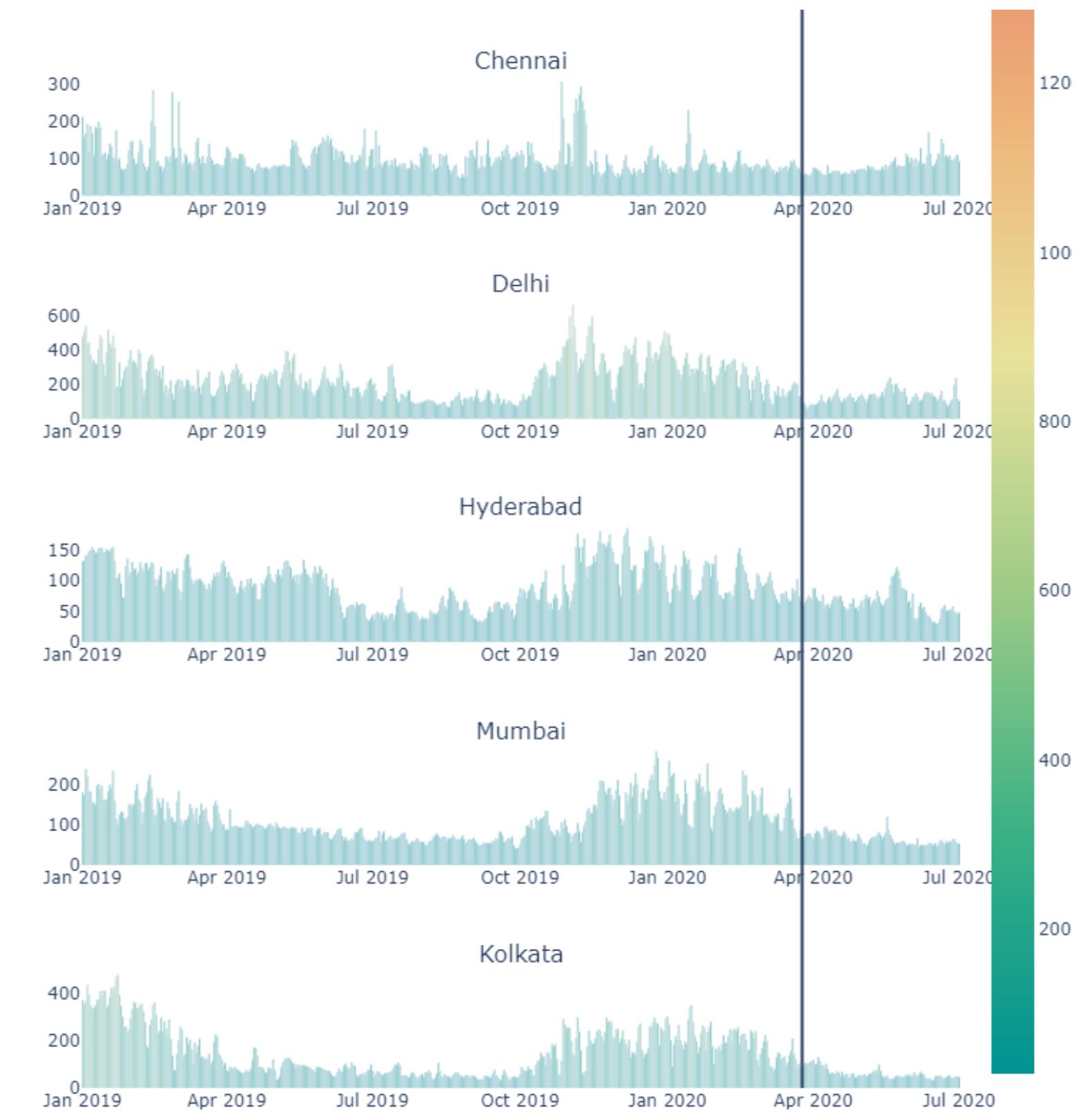
- It is evident that Delhi has had the maximum Particulate Matter average over the past 5 years
- Ahmedabad topped both NO₂ and SO₂ levels over 5 years.
- There are nearly 11 Thermal power plants around the NCR firing coal continuously round the clock to generate electricity for NCR.
- Fly ash produced from these thermal power plants and stubble burning in the surrounding regions of NCR contributes to the higher values of particulate matter in Delhi.



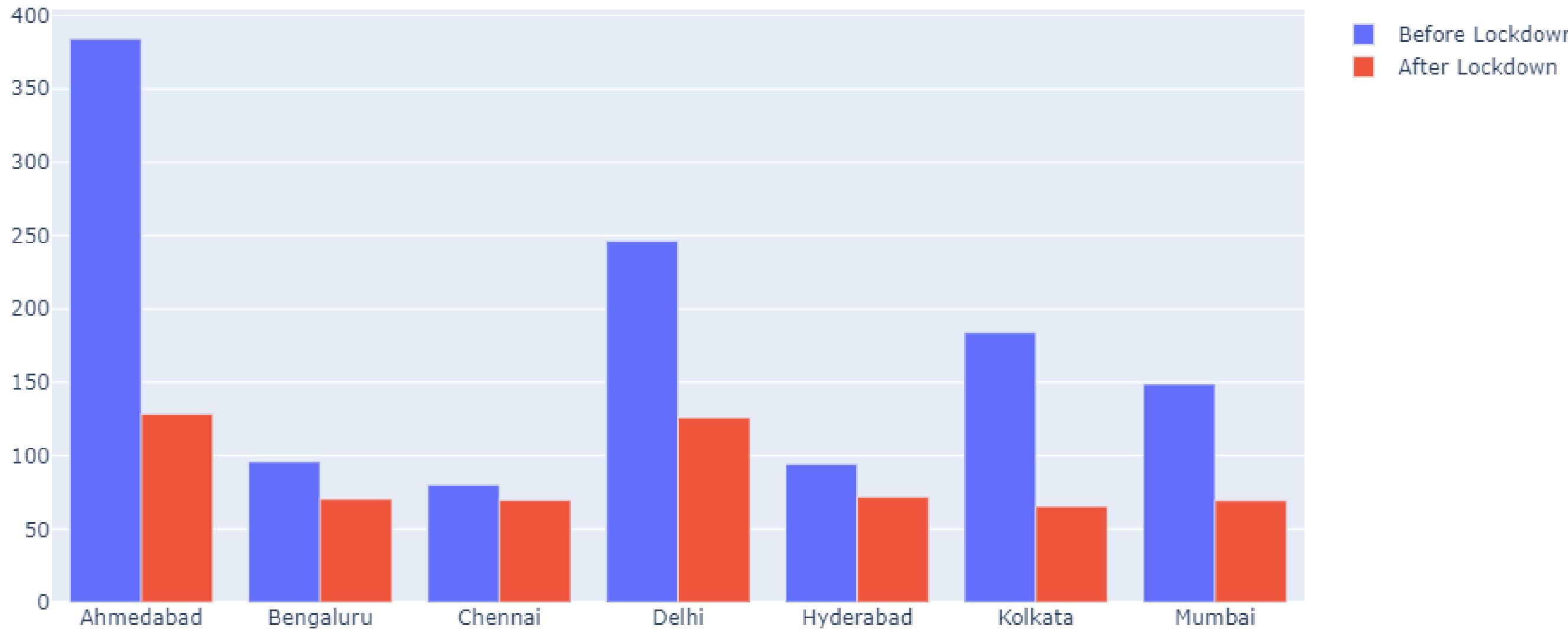
Effect Of Lockdown

From the above plot,

- The black line indicates the start date of Lockdown.
- Clearly, after the lockdown AQI of all the States has reduced significantly.



Comparison of AQI before and after lockdown

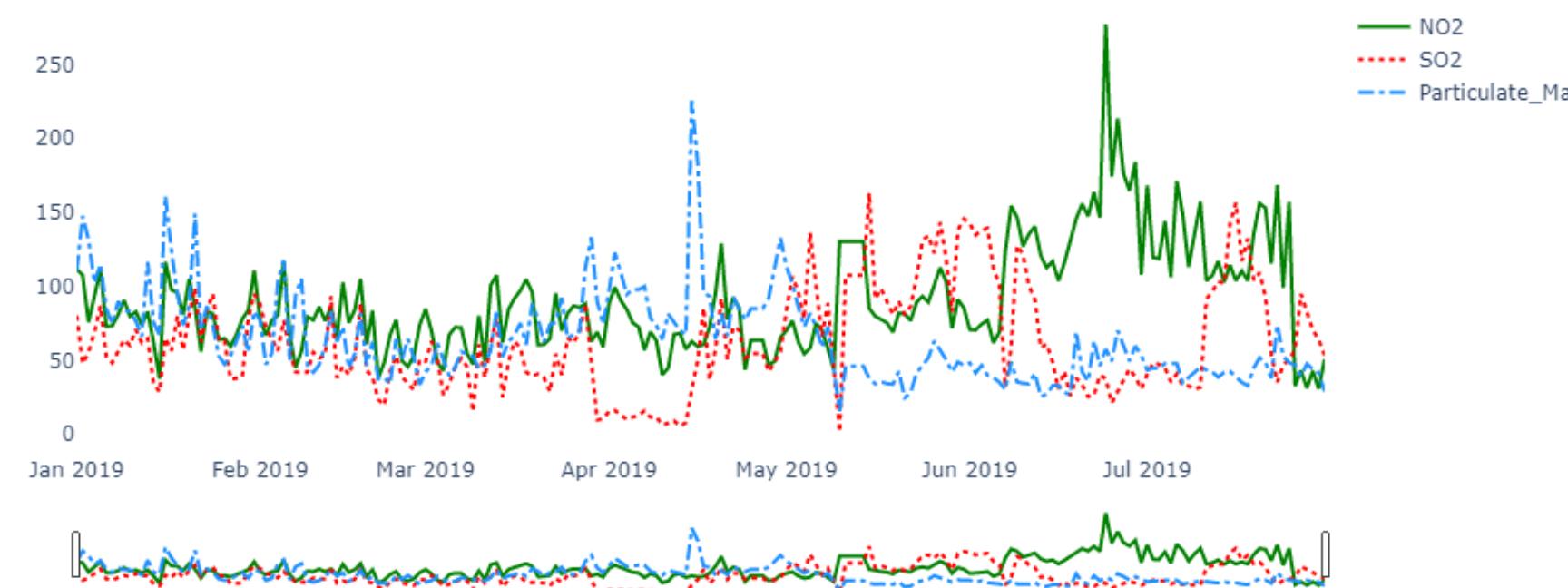


From the above plot,

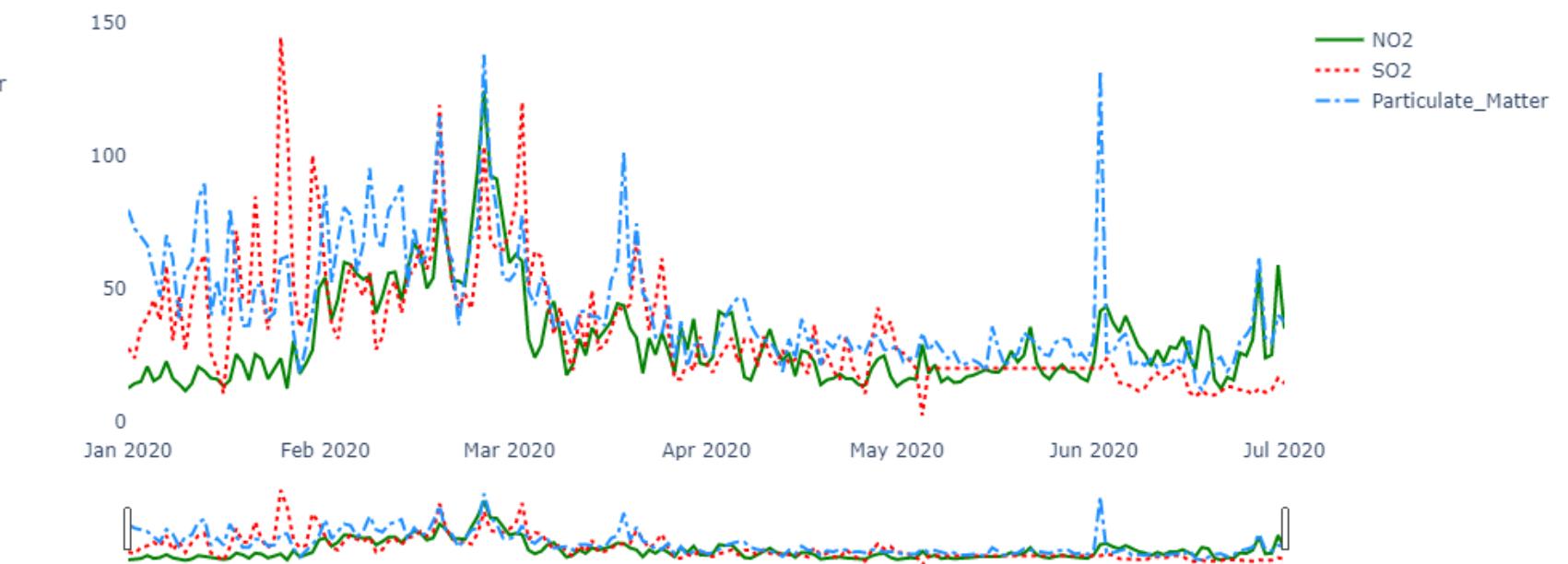
- We can observe a significant reduction of AQI of all the Cities when compared the result with Before Lockdown

Effect of Lockdown on individual pollutant levels

Ahmedabad 2019

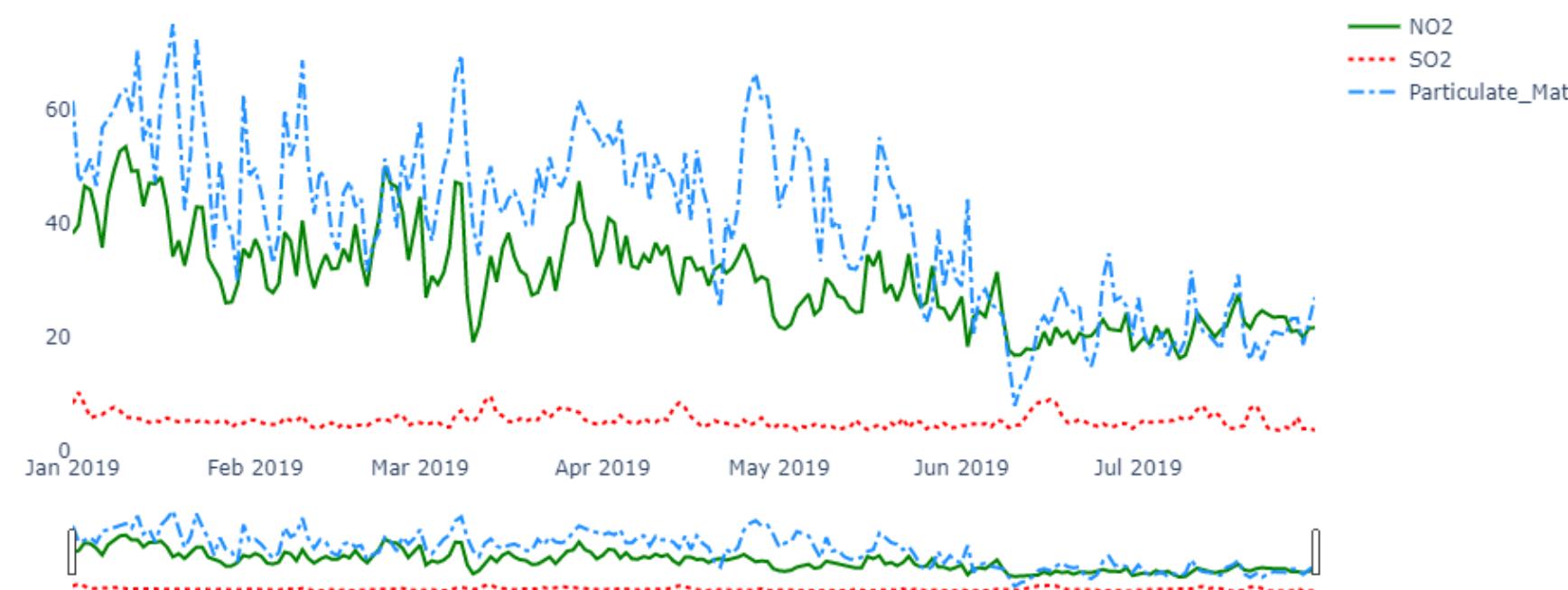


Ahmedabad 2020

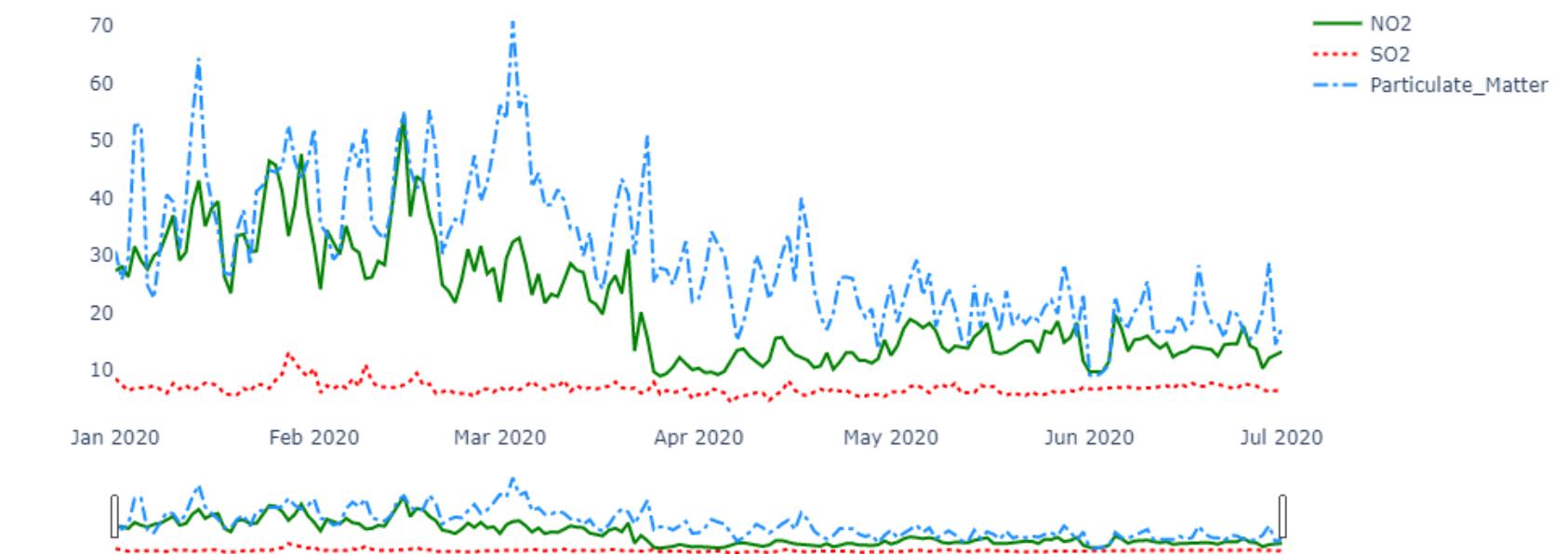


Effect of Lockdown on individual pollutant levels

Bengaluru 2019

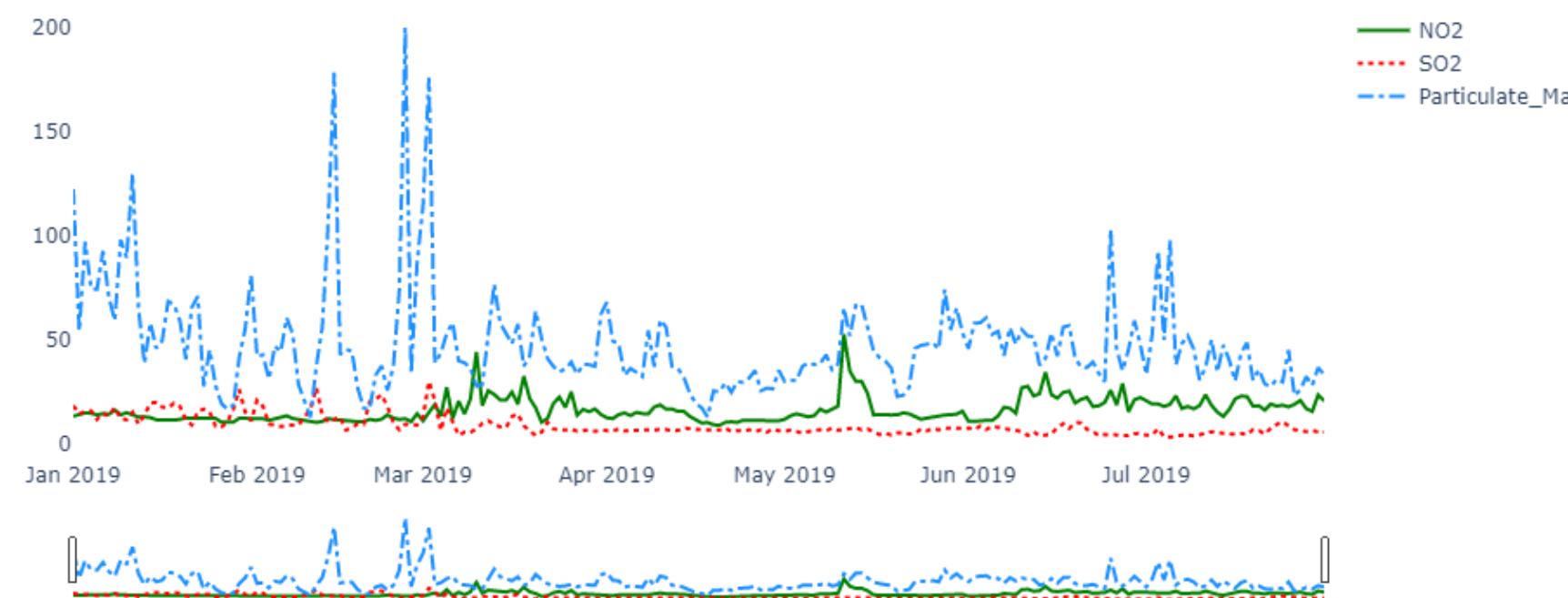


Bengaluru 2020

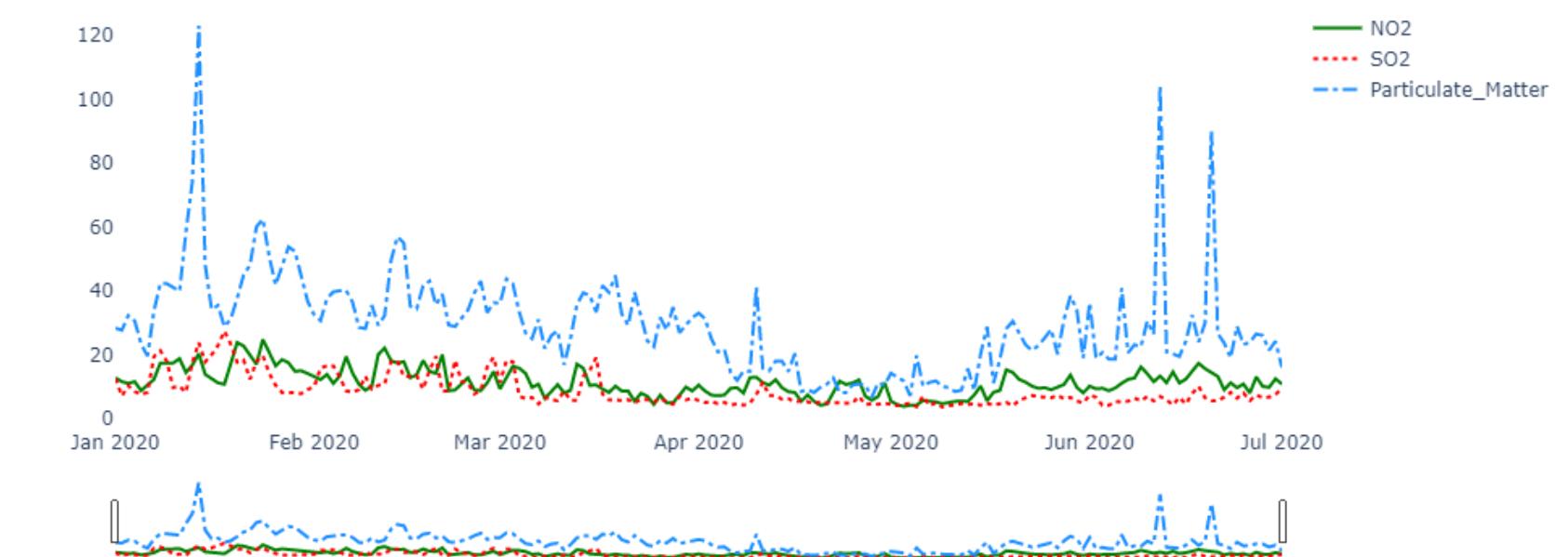


Effect of Lockdown on individual pollutant levels

Chennai 2019

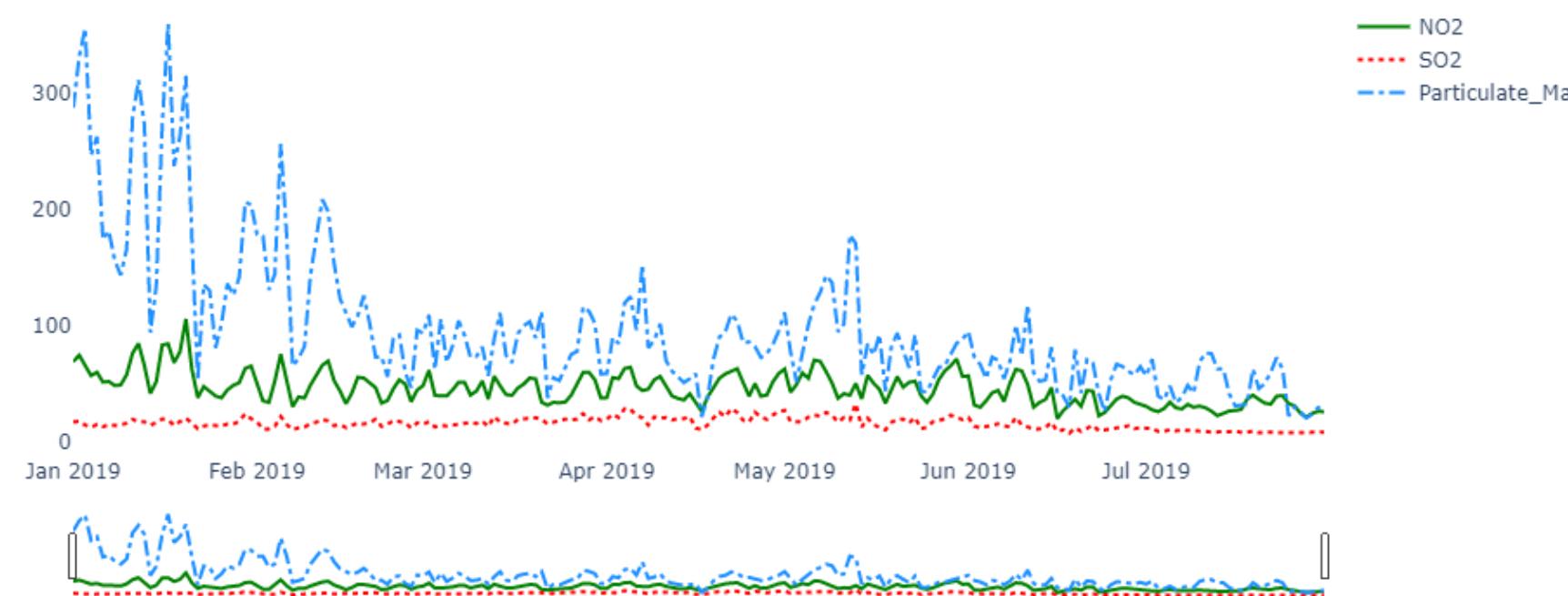


Chennai 2020

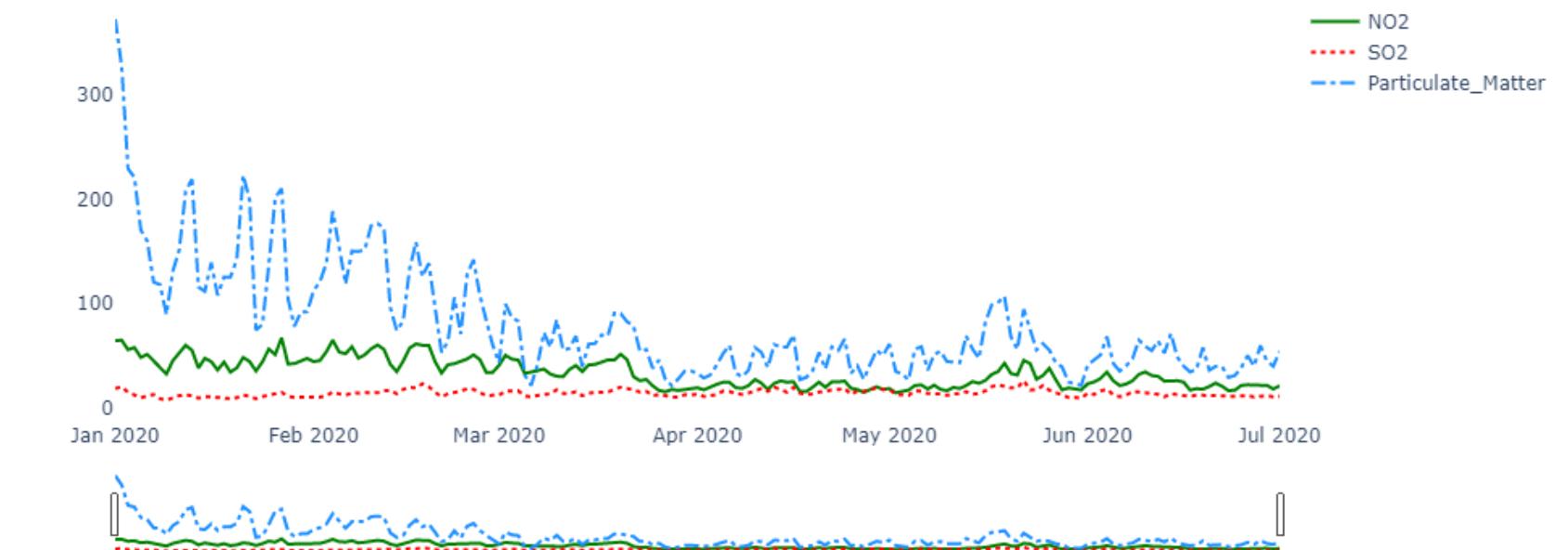


Effect of Lockdown on individual pollutant levels

Delhi 2019

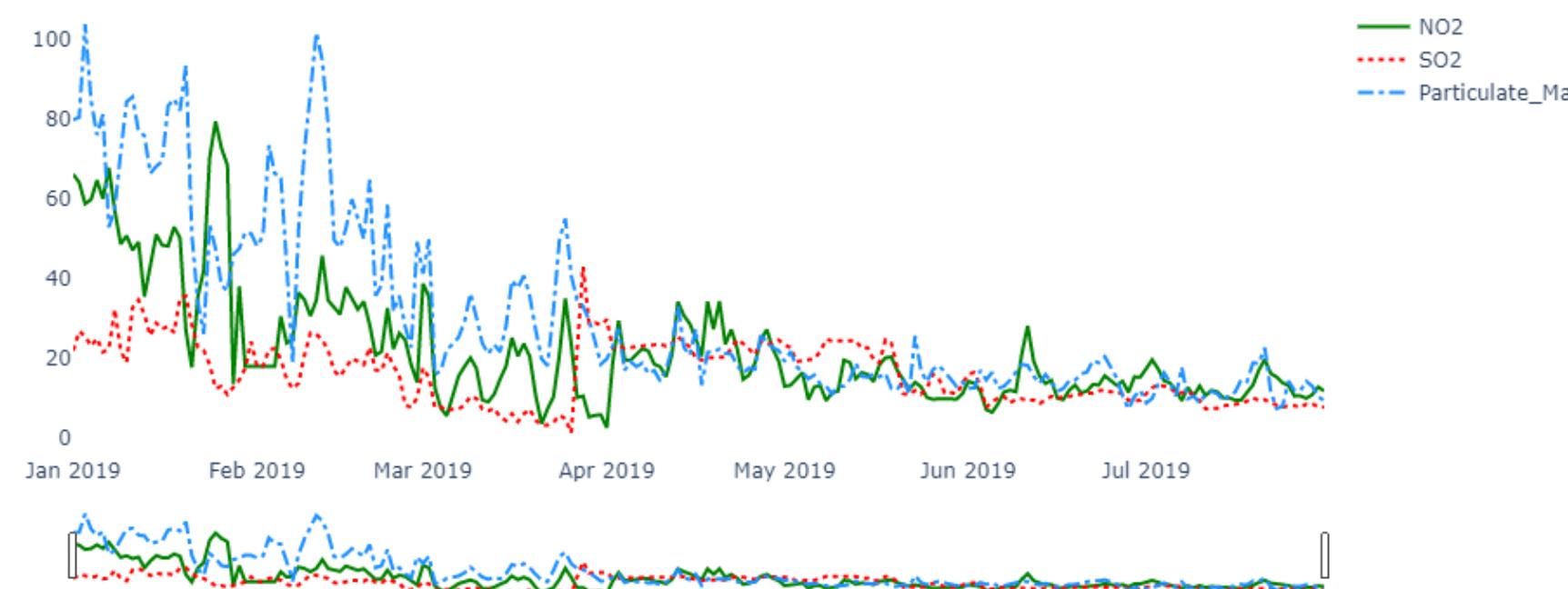


Delhi 2020

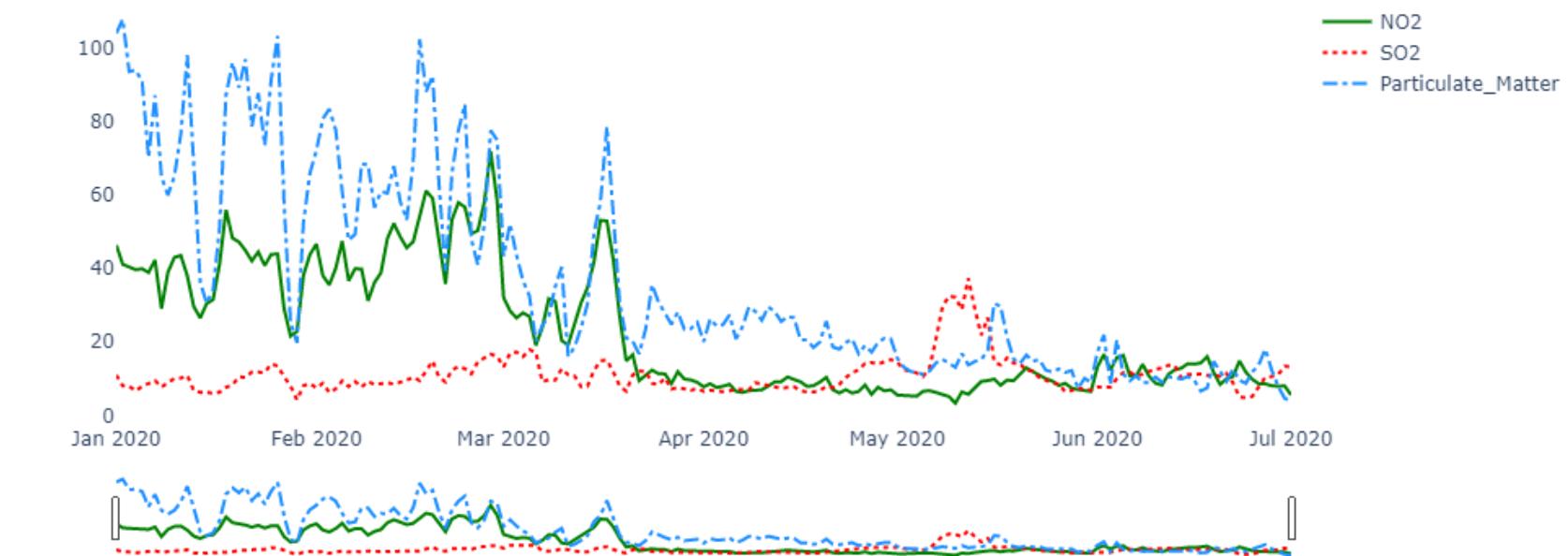


Effect of Lockdown on individual pollutant levels

Mumbai 2019

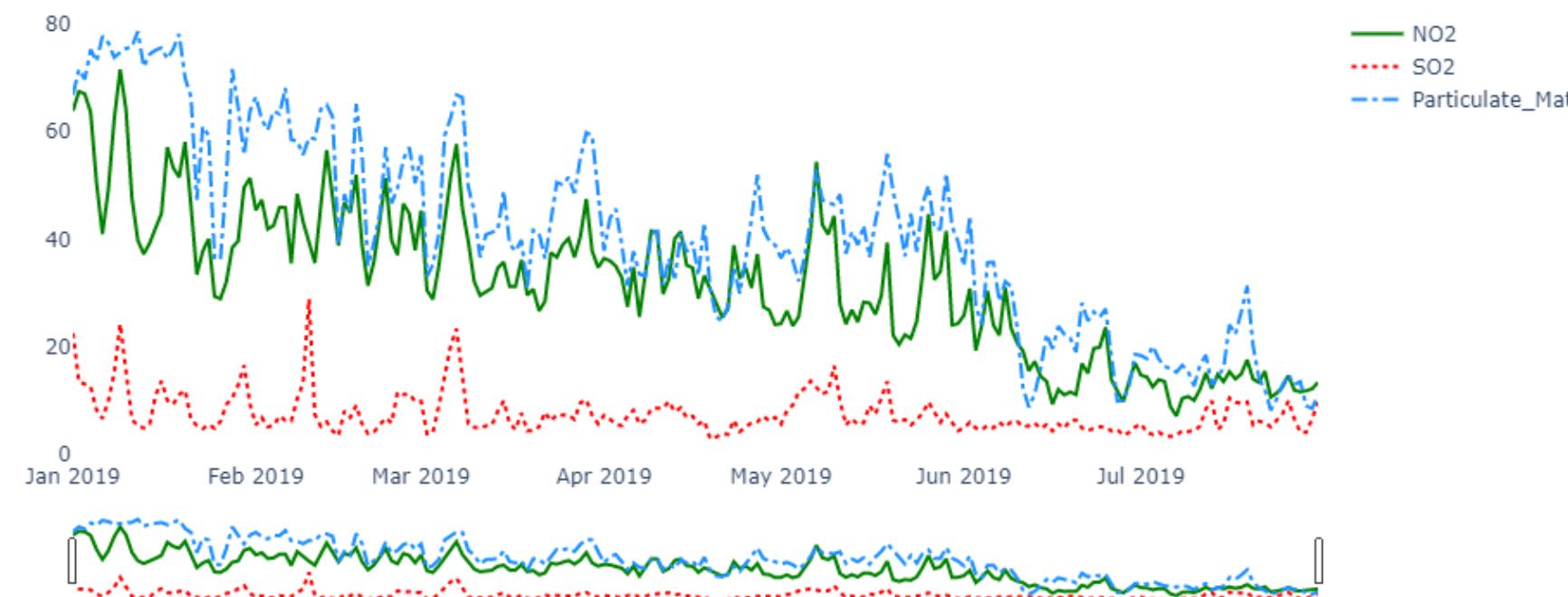


Mumbai 2020

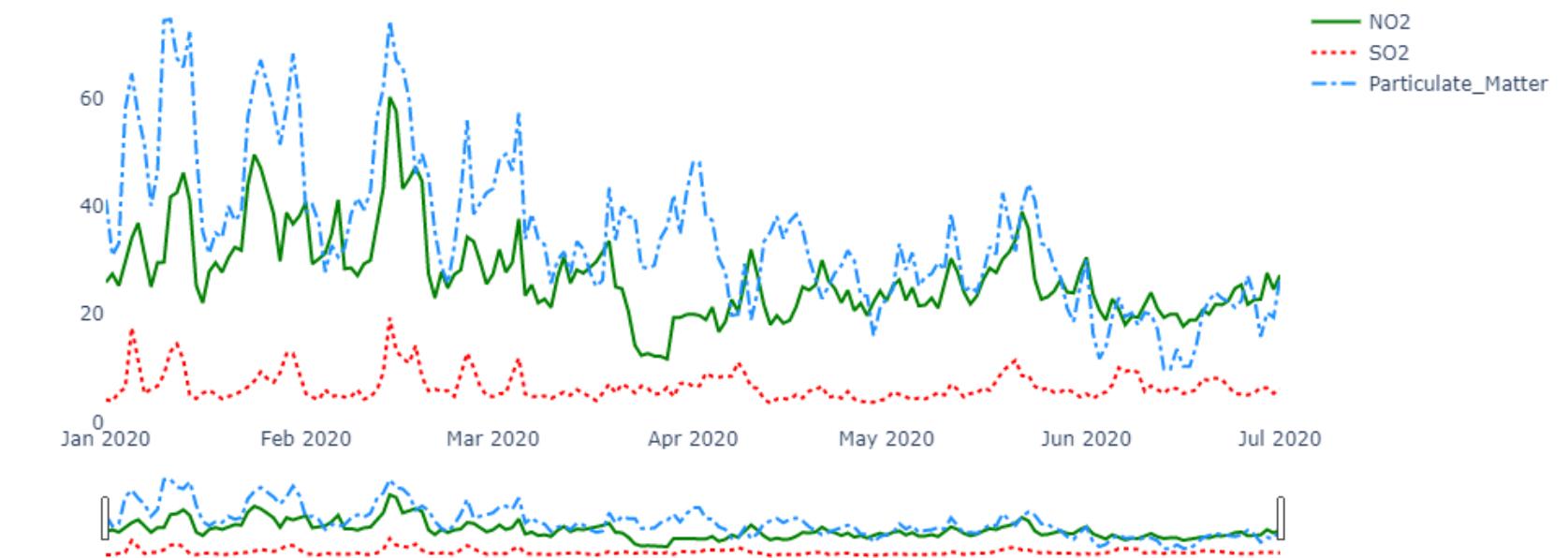


Effect of Lockdown on individual pollutant levels

Hyderabad 2019

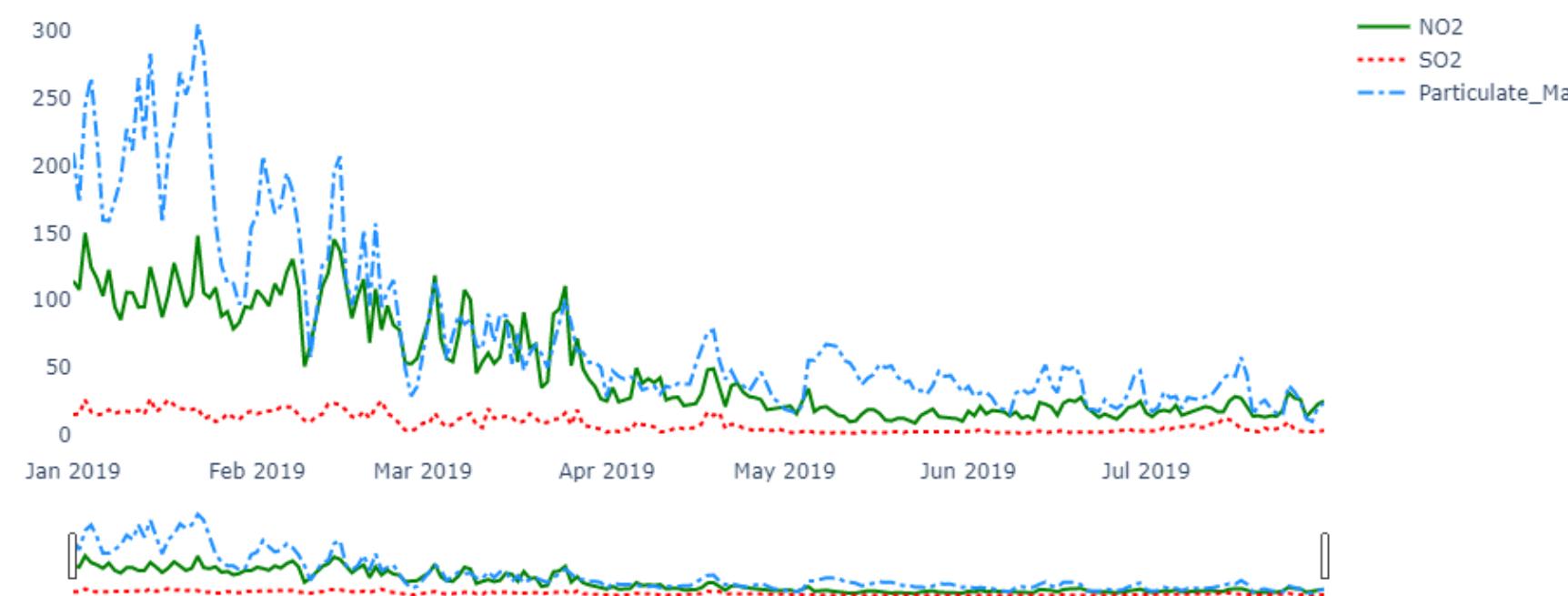


Hyderabad 2020

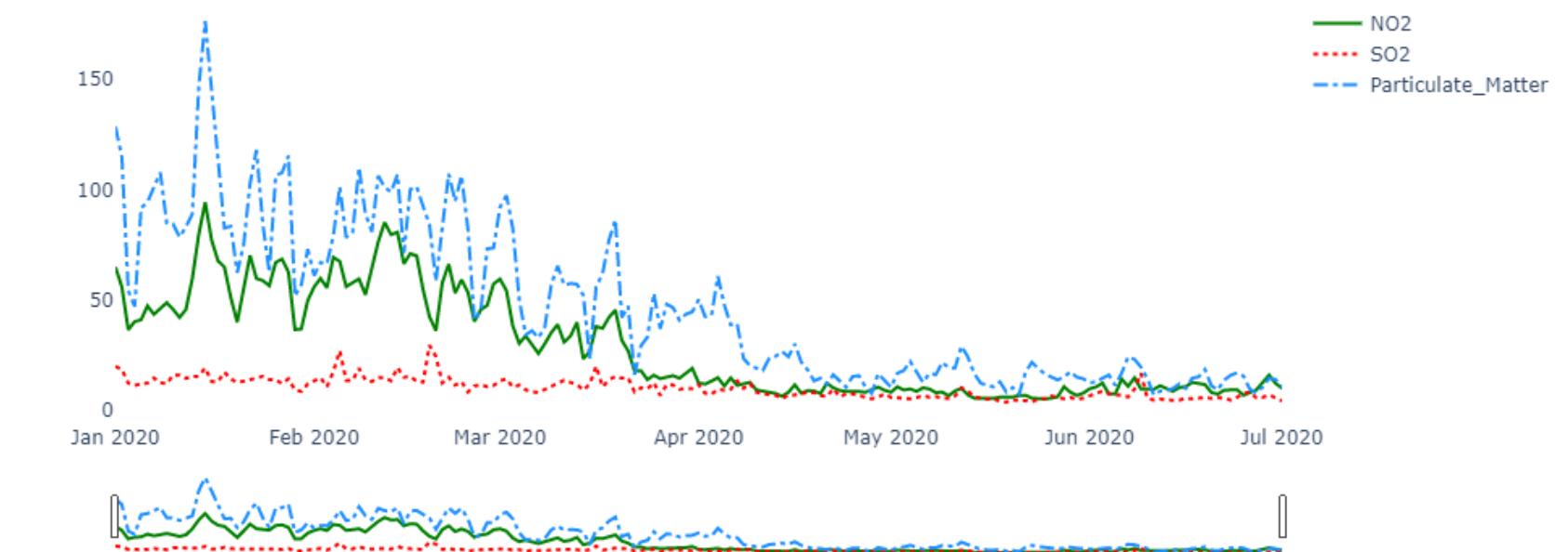


Effect of Lockdown on individual pollutant levels

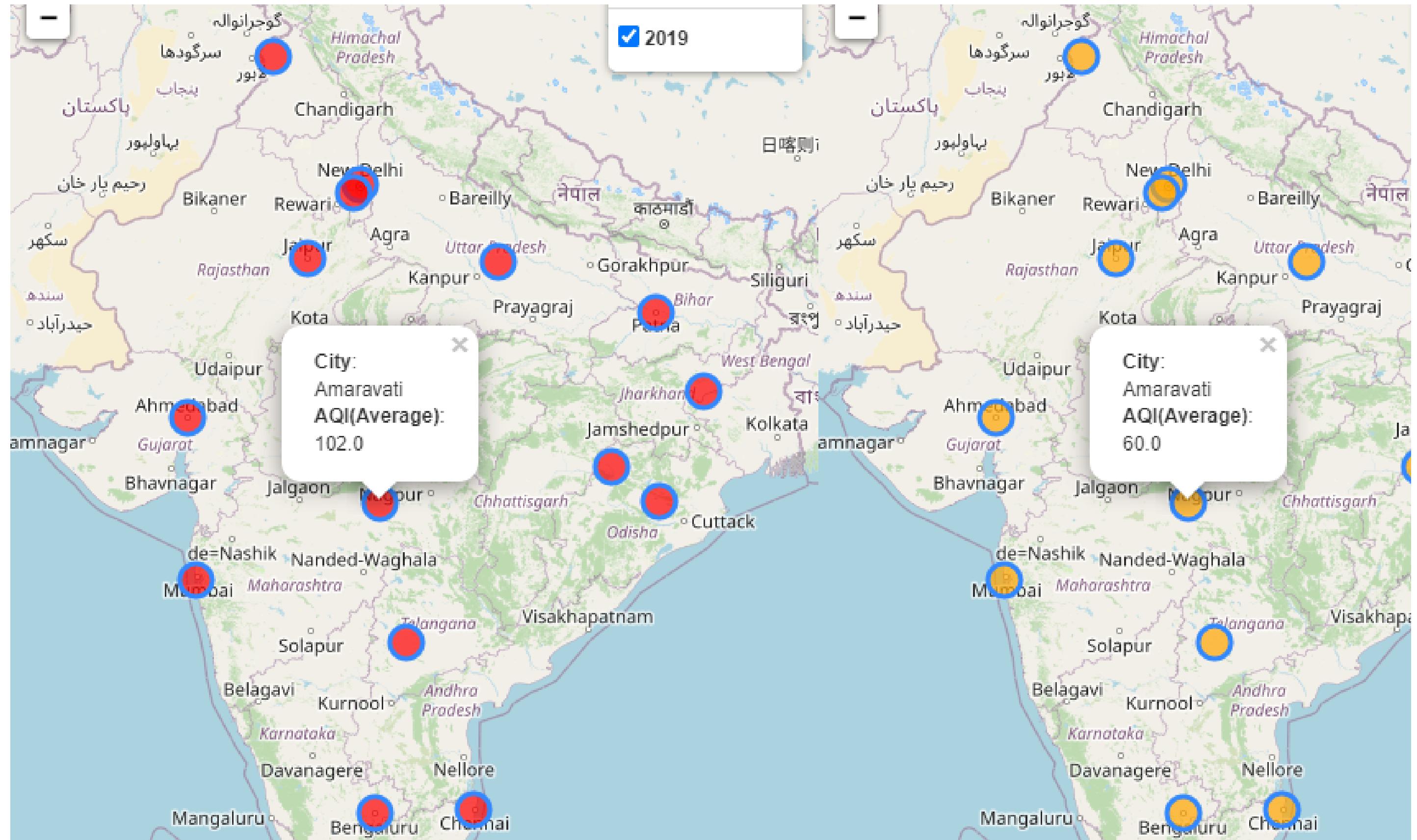
Kolkata 2019



Kolkata 2020



Dual map to visualize the AQI change during 2019 and 2020





RESULTS AND CONCLUSION

DECADAL POLLUTION ANALYSIS

- States with large populations and high industrial activity are hotspots. Such states include Uttar Pradesh, Maharashtra, Andhra Pradesh etc.
- Other states like Himachal Pradesh, Goa, Kerala and almost all the seven sister states are coldspots.
- NO₂ levels are highly correlated with the number of vehicles and the number of industries.
- The number of vehicles is also highly correlated with the number of industries. This shows that vehicular registration is more in industrialized places.
- A significant correlation between PM2.5 and Industry/vehicles is observed.
- Uttarakhand, Maharashtra, and Jharkhand paid the most attention to SO₂. For Meghalaya, we can see a lot of bounce within the SO₂ in 2013, 2018, and 2011. For ten years, Himachal Pradesh, Manipur, Mizoram, and Nagaland have displayed much less change in SO₂.



RESULTS AND CONCLUSION

DECADAL POLLUTION ANALYSIS

- Ahmedabad has the highest SO₂ concentration in the year 2019 Talcher of Odisha has higher in 2016. In contrast, Shillong of Meghalaya and Alwal have the most negligible SO₂ concentration.
- Haryana has a somewhat high NO₂ concentration from 2015-2020. We can see that the NO₂ concentrations in Jammu and Jharkhand are very similar to those in Delhi. For the years 2016-2020, Nagaland, Mizoram, and Meghalaya have the lowest NO₂ concentrations. For the year 2019, NO₂ levels in Jharkhand have risen dramatically.
- Ahmedabad has the highest NO₂ concentration, and Aizawl has the least.
- Haryana has NO₂ concentration in 2019 more significant than the mean. Delhi and West Bengal are in second place. NO₂ concentrations are modest in Jharkhand, Maharashtra, and Bihar. Whereas Goa has the lowest SO₂ and NO₂ concentrations.



RESULTS AND CONCLUSION

DECADAL POLLUTION ANALYSIS

- Assam and Uttar Pradesh have had high PM2.5 concentrations for the past ten years. After 2015, the concentration of PM2.5 in Nagaland increased dramatically. In Jharkhand, there was a sharp increase in PM2.5 concentration in 2012, followed by moderate to high PM2.5 concentration. In addition, PM2.5 concentrations in Punjab, Uttarakhand, and Puducherry have gradually decreased.
- Bhopal had the highest PM2.5 concentration in 2013 while Aizawl had the lowest.
- Uttar Pradesh has the highest PM2.5 levels in 2019. Uttar Pradesh and Jharkhand, which are placed second and third, have practically identical PM2.5 mean concentration values.
- PM2.5 in Uttar Pradesh kept on increasing till 2019 while in Jharkhand, it gradually decreased.
- Uttar Pradesh is the most densely inhabited, immediate action is required to control pollution levels in these states.



RESULTS AND CONCLUSION

IMPACT OF COVID 19 ANALYSIS

- **Delhi has the maximum Particulate Matter average over the past 5 years whereas Ahmedabad topped both NO₂ and SO₂ levels over 5 years.**
- **Pollution in all the Indian cities is more than the safer limits before the lockdown.**
- **Starting June-July, till October, AQI during the monsoon generally falls due to monsoonal rains washing the pollutants from the air.**
- **Clearly, there is a rapid downfall in AQI levels after 25th March 2020(after lockdown) in all the cities under consideration. Of all the cities, Ahmedabad had the steepest decline followed by Mumbai and Delhi. Whereas, Bengaluru hasn't recorded such steeper falls compared to other cities.**
- **Ahmedabad and Delhi experienced a greater decrease in AQI**



That Ends The Presentation

If you have any questions feel free to ask]
