Indian Air Pollution data Analysis

A report by the Health Effects Institute on air pollution in India (2018) reports that air pollution was responsible for 1.1 million deaths in India in 2015.

**AIR POLLUTION** refers to the release of **pollutants** into the **air** that is detrimental to human health and the planet as a whole .

Today air pollution has been one of the significant problems to deal with for any nation. In South Asia, it is ranked as the sixth most dangerous killer.  
One does not realize the harmful effects of a problem if he/she has not experienced it in the first place.  
Take Delhi, for instance, we all have experienced what it feels like inhaling in the ‘deadly’ smog that remained for about a week, after Diwali. Citizens were advised not to leave their homes and were asked to wear masks whenever going outside. Looking outside the window made me feel like I was living in a gas chamber. Low visibility, a high number of deaths, etc. were the effects of pollution.

The dataset contains the following features :

1. **stn\_code** : Station code. A code is given to each station that recorded the data.
2. **sampling\_date**: The date when the data was recorded.
3. **state**: It represents the states whose air quality data is measured.
4. **location**: It represents the city whose air quality data is measured.
5. **agency**: Name of the agency that measured the data.
6. **type**: The type of area where the measurement was made.
7. **so2**: The amount of Sulphur Dioxide measured.
8. **no2**: The amount of Nitrogen Dioxide measured
9. **rspm**: Respirable Suspended Particulate Matter measured.
10. **spm**: Suspended Particulate Matter measured.
11. **location\_monitoring\_station**: It indicates the location of the monitoring area.
12. **pm2\_5**: It represents the value of particulate matter measured.
13. **date**: It represents the date of recording (It is a cleaner version of ‘sampling\_date’ feature)

**Some Defintions :**

**NO2:** Combustion from power sources or Transport.

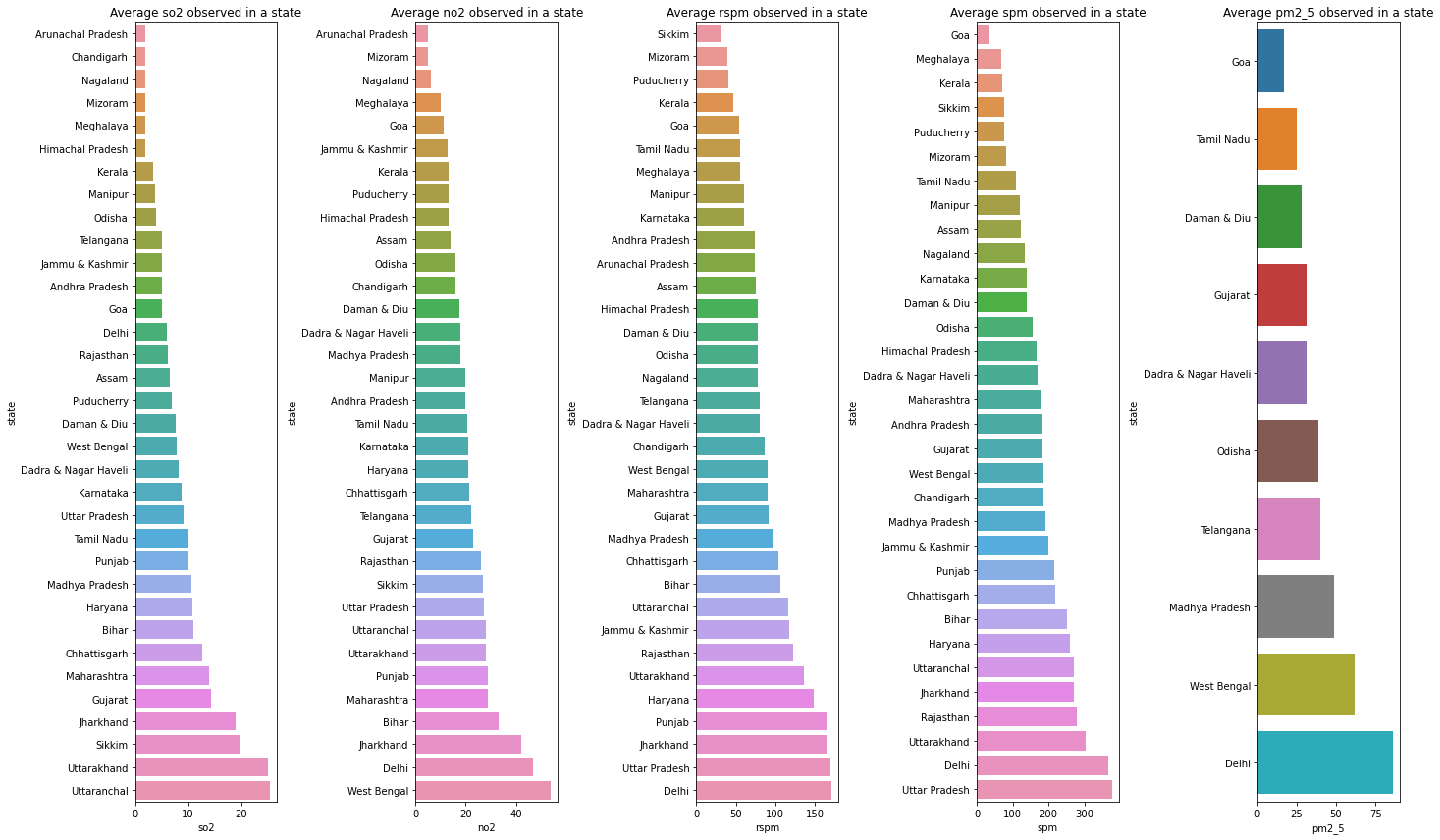
**SO2:** Coal burning, Oil burning, Manufacturing of Sulphuric acid.

**spm:** Suspended particulate matter and are known to be the deadliest form of air pollution. They are microscopic in nature and are found to be suspended in earth's atmosphere.

**rspm:** Respirable suspended particulate matter. A sub form of spm and are responsible for respiratory diseases.

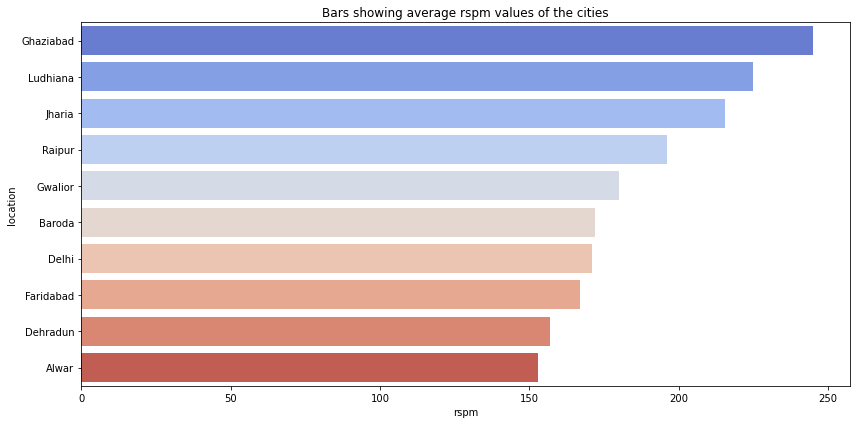
**pm2\_5:**Suspended particulate matter with diameters less than 2.5 micrometres. They tend to remain suspended for longer durations and potentially very harmful.

**Question 1**:  Find out the state which least polluted.



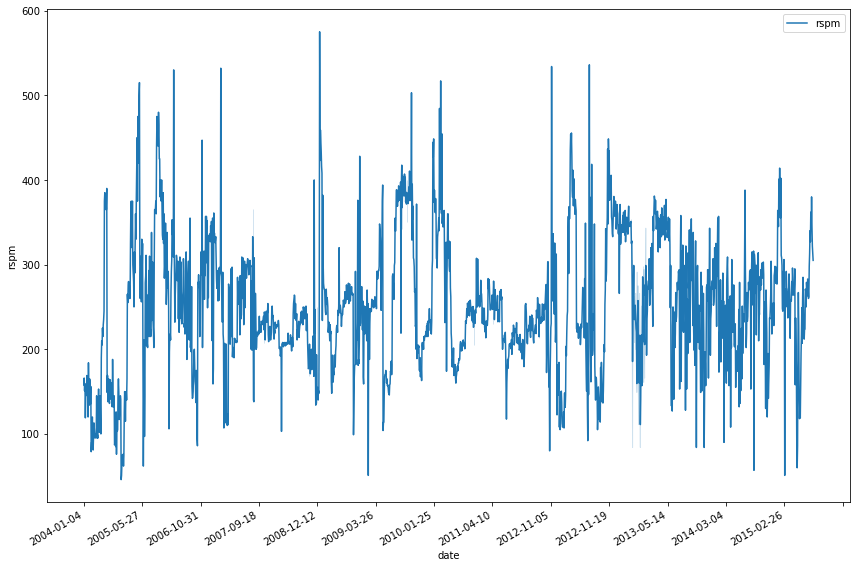
From above barplot we can conclude that Goa is the least polluted city.

Question 2: Top TEN Cities have highest risk of respiratory diseases.



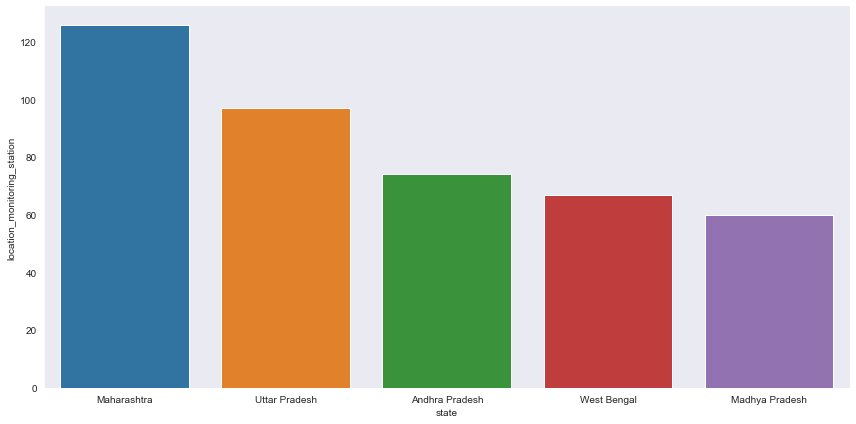
The above plot shows all the 10 cities which have highest risk of respiratory disease.

Question 3: Show RSPM variation in ghaziabad over time.



The above lineplot shows variation of RSPM in Ghaziabad overtime

Question 4: Top five states with highest number of monitoring stations.



|  |
| --- |
| **state** | **location\_monitoring\_station** |
| **Maharashtra** | 126 |
| **Uttar Pradesh** | 97 |
| **Andhra Pradesh** | 74 |
| **West Bengal** | 67 |
| **Madhya Pradesh** | 60 |