Analysis of Air Quality in the COVID Era

AC-102

MTE Innovative Project

Ansh Anurag 2K20/B2/71
Ananya Ranjan 2K20/B2/57

Introduction

On March 25 2020, the Indian government placed its population of more than 1.3 billion citizens under lockdown in an effort to reduce the spread of the COVID-19 disease. All non-essential shops, markets and places of worship were closed with only essential services including water, electricity and health services remaining active.

Citizens started to experience better air quality so much so that the scenic Dhauladhar Peaks of Himachal Pradesh became visible from neighbouring states. On normal days, these peaks lie hidden behind the film of smog.

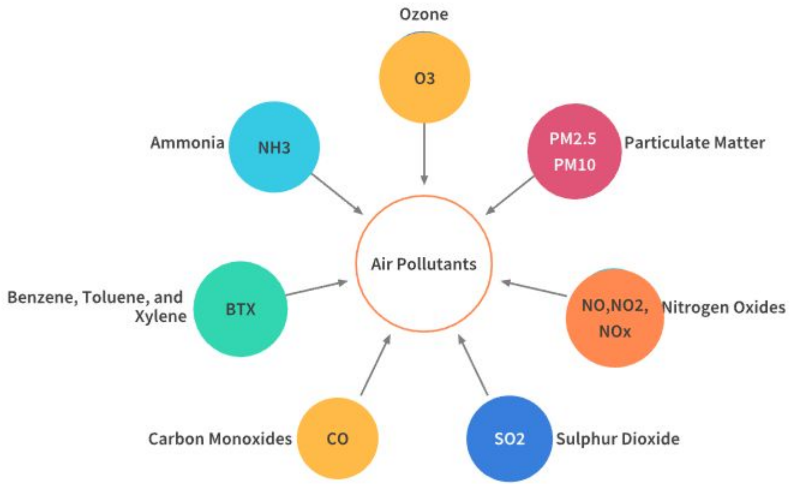
Objective

Here we have access to a large amount of granular data relating to the concentration of major air pollutants in India and it will be interesting to see if the claim of reduced air pollution is being actually backed by data. Before going further let's understand about the constituents of Air Pollution.

Types of Air Pollution

- **Particulate matter (PM2.5 and PM10)** > Particulate matter is a mix of solids and liquids, including carbon, complex organic chemicals, sulphates, nitrates, mineral dust, and water suspended in the air. PM varies in size. Some particles, such as dust, soot, dirt or smoke are large or dark enough to be seen with the naked eye. But the most damaging particles are the smaller particles, known as PM10 and PM2.5.Source. The following diagram will help to understand the concept more concretely.
- **Nitrogen Oxides** (NO, NO2, NOx) > Nitrogen oxides are a group of seven gases and compounds composed of nitrogen and oxygen, sometimes collectively known as NOx gases.The two most common and hazardous oxides of nitrogen are nitric oxide(NO) and nitrogen dioxide(NO2)
- **Sulphur Dioxide**(SO2) > Sulfur dioxide, or SO2 is a colorless gas with a strong odor, similar to a just-struck match. It is formed when fuel containing sulfur, such as coal and oil, is burned, creating air pollution.
- **Carbon Monoxide**(CO) > Carbon monoxide is a colorless, highly poisonous gas. Under pressure, it becomes a liquid. It is produced by burning gasoline, natural gas, charcoal, wood, and other fuels.

- **Benzene, Toluene and Xylene** (BTX) > Benzene, toluene, xylene, and formaldehyde are well-known indoor air pollutants, especially after house decoration. They are also common pollutants in the working places of the plastic industry, chemical industry, and leather industry
- **Ammonia**(NH_3) > Ammonia pollution is pollution by the chemical ammonia (NH3) – a compound of nitrogen and hydrogen which is a byproduct of agriculture and industry.
- **Ozone**(O_3) > Ground-level ozone is a colorless and highly irritating gas that forms just above the earth's surface. It is called a "secondary" pollutant because it is produced when two primary pollutants react in sunlight and stagnant air. These two primary pollutants are nitrogen oxides (NOx) and volatile organic compounds (VOCs).



Methodology

In this project, the analysis has been done in 2 parts:

- **Analysis of the pollution level in India, over the years - from 2015 to 2020**
  This will a holistic view of how the pollutant levels have been rising in India and what is the current situation
- **Effect of Lockdown on the Pollution level in India**
  Here we shall examine the pollution level in India before and after the first stage of Lockdown. Also we shall compare the pollution level around the same months in 2019, to see the the difference, if any. Additionally we could also examine the difference between the the current dates and the winter months(October, November) of 2019 when the pollution levels are generally the highest in Northern parts of India.

Python Libraries Used

Numpy, Pandas, Matplotlib, Folium, Seaborn

A First look at City Data

We have daily and hourly city data as well as daily and hourly Station data. Station refers to the continuous pollution monitoring stations operated and maintained by the Central Pollution Control Board (CPCB) and the State Pollution Control Boards. We began by analyzing the various cities' daily data to get a big picture. First of all we imported the dataset and the necessary libraries for the analysis.

	Missing Values	% of Total Values
Xylene	16807	64.100000
PM10	10766	41.100000
NH3	9847	37.600000
Toluene	7555	28.800000
Benzene	5287	20.200000
PM2.5	4289	16.400000
AQI	4282	16.300000
AQI_Bucket	4282	16.300000
NOx	4043	15.400000
O3	3660	14.000000
SO2	3544	13.500000
NO	3233	12.300000
NO2	3217	12.300000
CO	1961	7.500000

Points to note :

- It seems that we have a lot of null values in some columns e.g.Xylene and NH3.There could be a couple of reasons for the missing data:
  - Station does not have device to capture it.
  - Issue in the meter reading.

Handling the missing Data

There were a lot of missing data points in the given dataset. So we followed the idea of data imputation in which we fabricated the data points using the K-Nearest Neighbours Algorithm which used the adjacent data points to complete the dataset if the data loss was less than 10%. If the missing data was more than 60%, we did not analyse the aspects since the analysis and predictions will not have a good accuracy due to lack of data points.

Cities included in the dataset

Given below is the code snippet showing the 24 cities included:

```
cities = city_day['City'].value_counts()
print(f'Total number of cities in the dataset : {len(cities)}')
print(cities.index)
```

```
Total number of cities in the dataset : 24
Index(['Mumbai', 'Chennai', 'Delhi', 'Ahmedabad', 'Bengaluru', 'Lucknow',
       'Hyderabad', 'Patna', 'Gurugram', 'Amritsar', 'Jorapokhar', 'Jaipur',
       'Thiruvananthapuram', 'Amaravati', 'Brajrajnagar', 'Talcher', 'Kolkata',
       'Guwahati', 'Shillong', 'Chandigarh', 'Bhopal', 'Ernakulam', 'Kochi',
       'Aizawl'],
      dtype='object')
```

Combining Data Points

In order to group the Pollutants on the basis of the characteristics, we grouped:

- Benzene, Toluene and Xylene as BTX.
- PM 10 and PM 2.5 as particulate matter.

In [9]:

```
city_day['BTX'] = city_day['Benzene']+city_day['Toluene']+city_day['Xylene']
city_day.drop(['Benzene','Toluene','Xylene'],axis=1);
```

```
city_day['Particulate_Matter'] = city_day['PM2.5']+city_day['PM10']
```

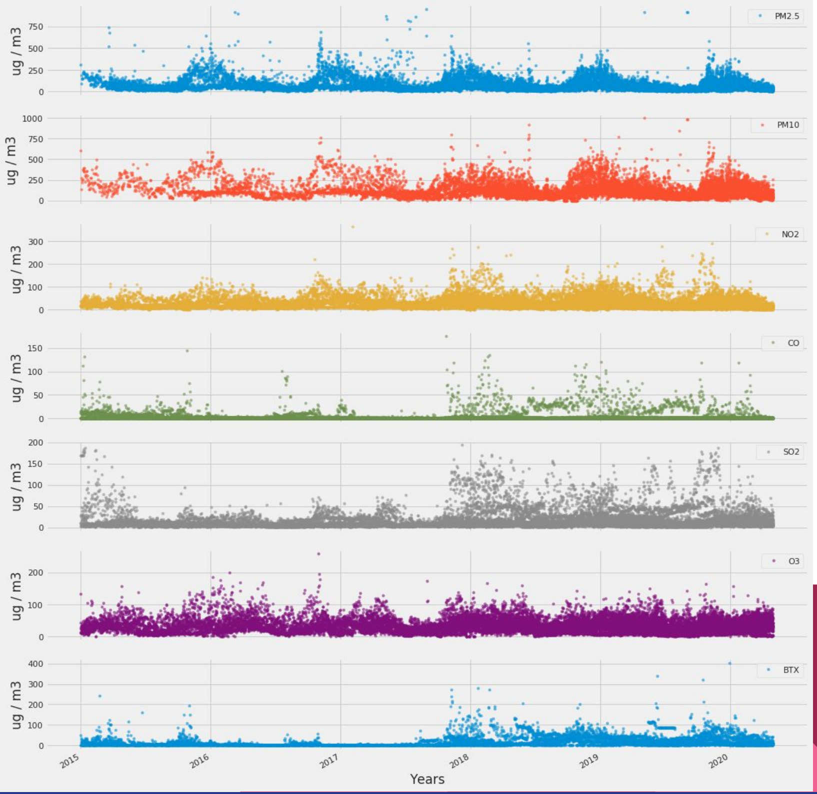
Even though a lot of columns have been provided in the dataset, we shall select a few prominent ones. We created a new dataframe called `pollutants` containing the major pollutants responsible for air pollution.

```
pollutants = ['PM2.5','PM10','NO2', 'CO', 'SO2','O3', 'BTX']
```

Visualising Yearly Data

Given below are the visual depictions of the amount of pollutants released against the timeframe of our dataset.

```
In [12]:
city_day.set_index('Date',inplace=True)
axes = city_day[pollutants].plot(marker='.', alpha=0.5, linestyle='None', figsize=(16, 20), sub
plots=True)
for ax in axes:

    ax.set_xlabel('Years')
    ax.set_ylabel('ug / m3')
```



Points to note :

- PM2.5 and PM10 pollution show a seasonal effect, with pollution being higher in winter months as compared to the summer ones.
- SO2 level has started increasing after 2017, although it had also seen a sudden rise in 2015 also. The same pattern is also reflected in BTX levels also.

This is just an overview. We will have to look deeper to get specific answers.

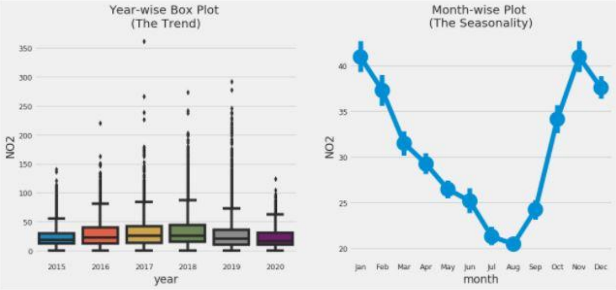
Year and Month-wise Distribution

We further explored the seasonality of our data with box plots. We shall use boxplots to group the data by different time periods and display the distributions for each group. We grouped the data by year and month. All measurements are in `(ug / m3)`.

2.4.1 NO_2

In [14]:

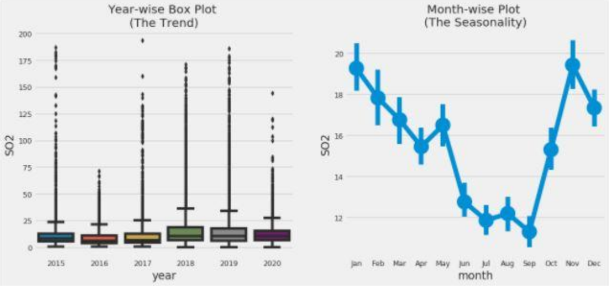
```
city_day.reset_index(inplace=True)
df = city_day.copy()
value='NO2'
trend_plot(df,value)
```



2.4.2 SO_2

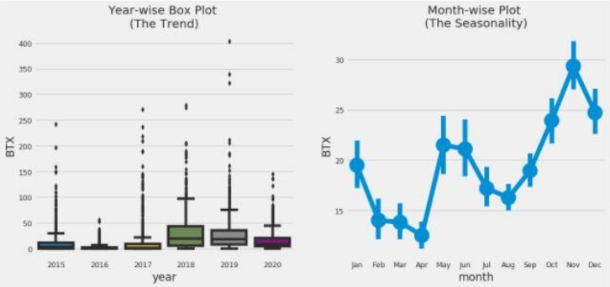
In [15]:

```
#city_day.reset_index(inplace=True)
df = city_day.copy()
value='SO2'
trend_plot(df,value)
```



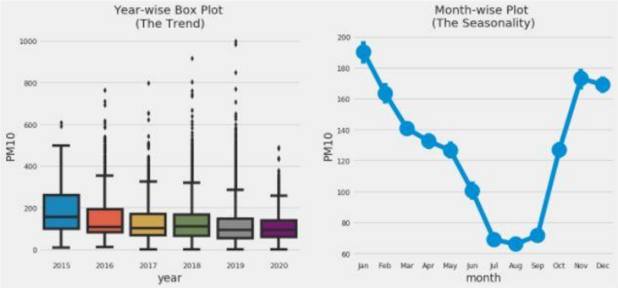
2.4.3 BTX

In [16]:
```
value='BTX'
trend_plot(df,value)
```



2.4.5 PM10

In [18]:
```
value='PM10'
trend_plot(df,value)
```



Points to note :

- There is a clear trend that pollution level in India falls in the month of July and August. This might be majorly because monsoon season sets in during these months.The BTX levels additionally show a major decline around April.
- The pollution level then start rising and reach highest levels in winter months. Again, its during these months that a lot of crop residue burning takes place,especially in northern parts of India.
- SO2 level has started increasing after 2017, although it had also seen a sudden rise in 2015 also. The same pattern is also reflected in BTX levels also.
- The median values of 2020 are generally less as compared to other years giving us a sense that there might be a reduction on pollution lately.

Effect of Lockdown

Let us now see how has the Lockdown affected the AQI levels in the prominent cities of India.For this we shall consider the data from 2019 onwards only. But before that we shall understand what AQI is:

AQI : Air Quality Index

An air quality index (AQI) is used by government agencies[1] to communicate to the public how polluted the air currently is or how polluted it is forecast to become.There are six AQI categories, namely Good, Satisfactory, Moderately polluted, Poor, Very Poor, and Severe. The proposed AQI will consider eight pollutants (PM10, PM2.5, NO2, SO2, CO, O3, NH3, and Pb) for which short-term (up to 24-hourly averaging period) National Ambient Air Quality Standards are prescribed.[23] Based on the measured ambient concentrations, corresponding standards and likely health impact, a sub-index is calculated for each of these pollutants. The worst sub-index reflects overall AQI. Likely health impacts for different AQI categories and pollutants have also been suggested, with primary inputs from the medical experts in the group. The AQI values and corresponding ambient concentrations (health breakpoints) as well as associated likely health impacts for the identified eight pollutants are as follows:

AQI	Remark	Color Code	Possible Health Impacts
0-50	Good		Minimal impact
51-100	Satisfactory		Minor breathing discomfort to sensitive people
101-200	Moderate		Breathing discomfort to the people with lungs, asthma and heart diseases
201-300	Poor		Breathing discomfort to most people on prolonged exposure
301-400	Very Poor		Respiratory illness on prolonged exposure
401-500	Severe		Affects healthy people and seriously impacts those with existing diseases

AQI for some of the major cities of India

The cities that will be the subject of our study are - Ahmedabad, Delhi, Bengaluru, Mumbai,Hyderabad and Chennai.

In [24]:

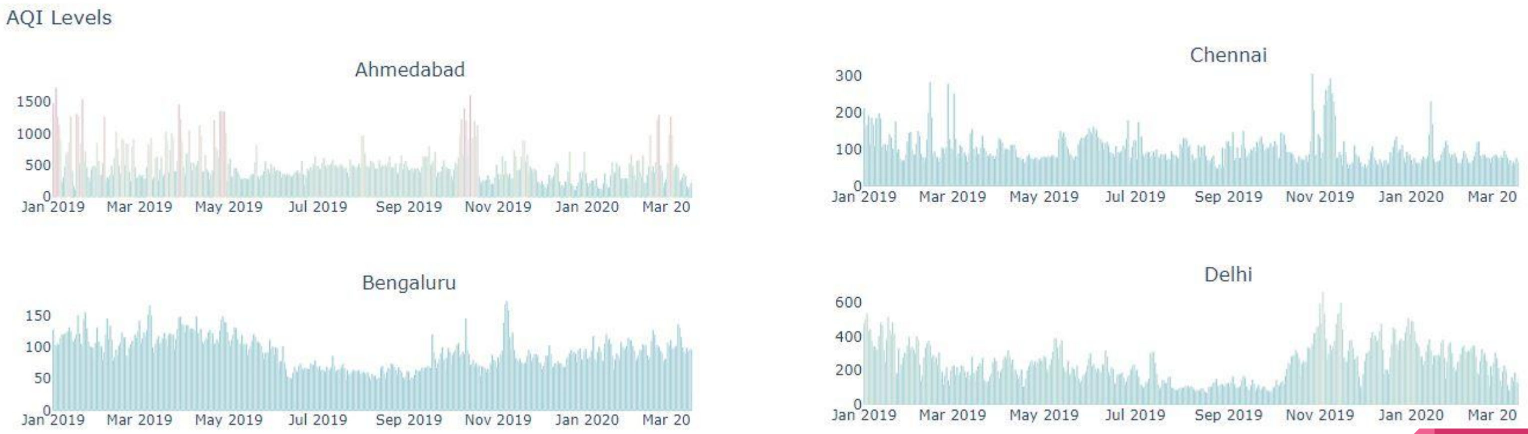
```
cities = ['Ahmedabad','Delhi','Bengaluru','Mumbai','Hyderabad','Chennai']

filtered_city_day = city_day[city_day['Date'] >= '2019-01-01']
AQI = filtered_city_day[filtered_city_day.City.isin(cities)][['Date','Cit
y','AQI','AQI_Bucket']]
AQI.head()
```

Out[24]:

	Date	City	AQI	AQI_Bucket
1461	2019-01-01	Ahmedabad	1474.0	Severe
1462	2019-01-02	Ahmedabad	1246.0	Severe
1463	2019-01-03	Ahmedabad	1719.0	Severe
1464	2019-01-04	Ahmedabad	1264.0	Severe
1465	2019-01-05	Ahmedabad	1127.0	Severe

Making Bar Graphs

You can find the Source Code on the link.



AQI Before and After the Lockdown

In [27]:

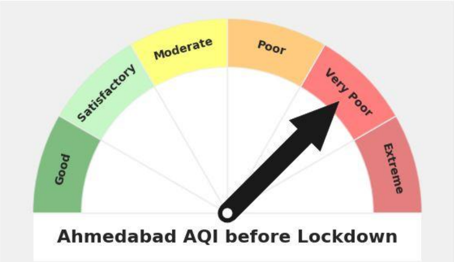
```
AQI_beforeLockdown = AQI_pivot['2020-01-01':'2020-03-25']
AQI_afterLockdown = AQI_pivot['2020-03-26':'2020-05-01']
```

In [28]:

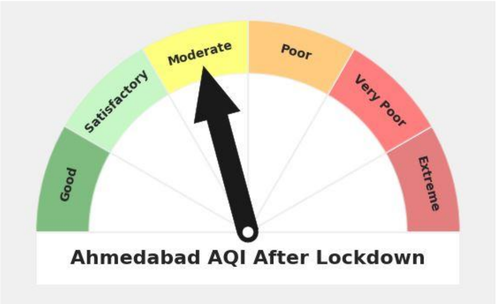
```
print(AQI_beforeLockdown.mean())
print(AQI_afterLockdown.mean())
```

```
City
Ahmedabad    383.776471
Bengaluru     96.023529
Chennai       80.317647
Delhi        246.305882
Hyderabad     94.435294
Mumbai       148.776471
dtype: float64
City
Ahmedabad    127.972973
Bengaluru     68.513514
Chennai       62.189189
Delhi        107.378378
Hyderabad     65.675676
Mumbai        73.972973
dtype: float64
```

"Ahmedabad's AQI levels"



Ahmedabad AQI before Lockdown



Ahmedabad AQI After Lockdown

"Delhi's AQI levels"

Delhi's AQI before Lockdown

Delhi's AQI after Lockdown

Template Code for the Representation

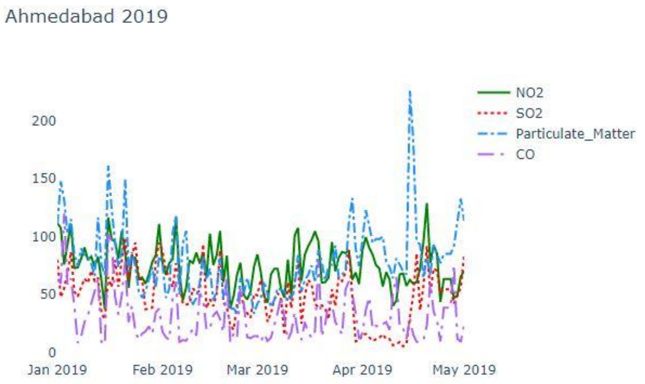
In [29]:

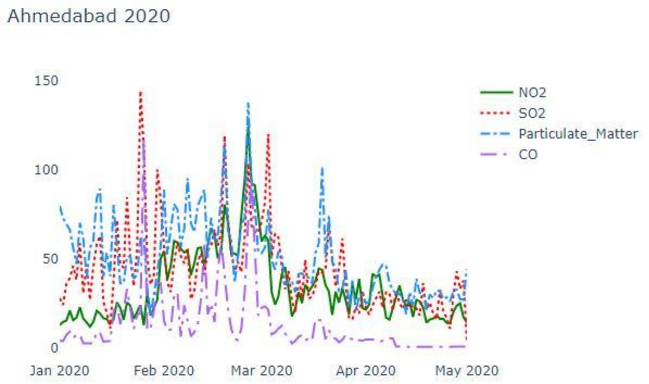
```
# Helper functions

#source: http://nicolasfauchereau.github.io/climatecode/posts/drawing-a-gaug
e-with-matplotlib/
from matplotlib.patches import Circle, Wedge, Rectangle

def degree_range(n):
    start = np.linspace(0,180,n+1, endpoint=True)[0:-1]
    end = np.linspace(0,180,n+1, endpoint=True)[1::]
    mid_points = start + ((end-start)/2.)
    return np.c_[start, end], mid_points

def rot_text(ang):
    rotation = np.degrees(np.radians(ang) * np.pi / np.pi - np.radians(90))
    return rotation
```

Effect of Lockdown on Individual levels of Pollutants

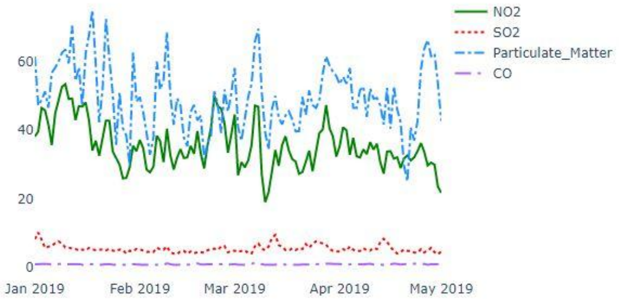
We compared the level of pollutants between January and April between between 2020 and 2019. This will give an idea whether the pollution levels have actually subsided or the pollution actually remains low during the onset of summer in India.

```
In [36]: pollution_comparison('Ahmedabad')
```

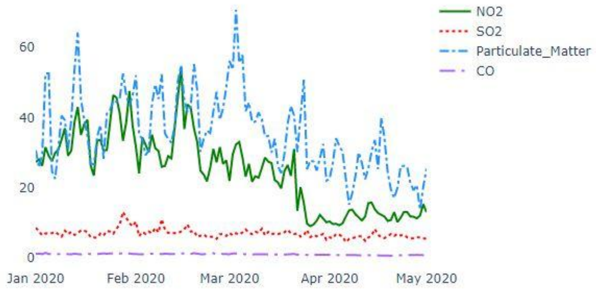




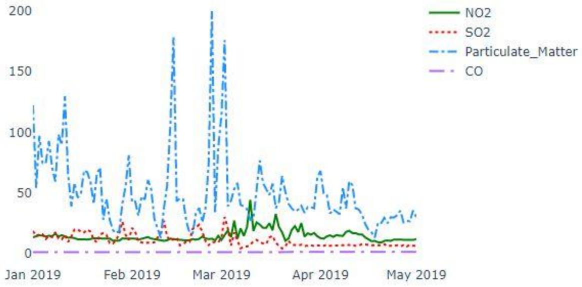
Bengaluru 2019



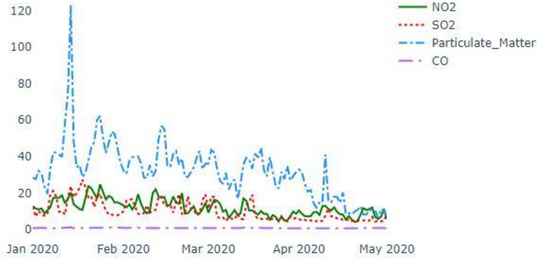
Bengaluru 2020

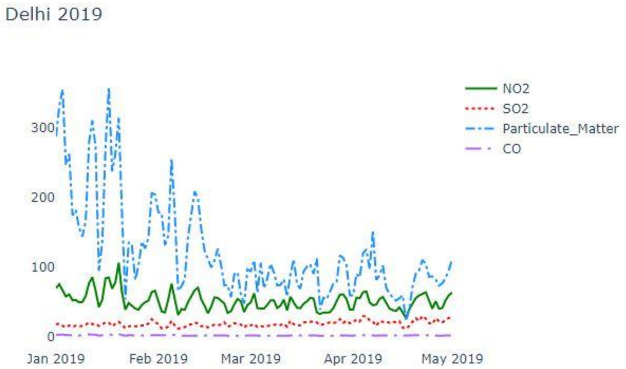


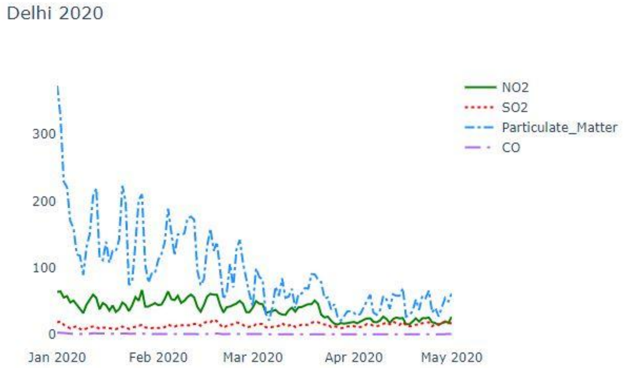
Chennai 2019

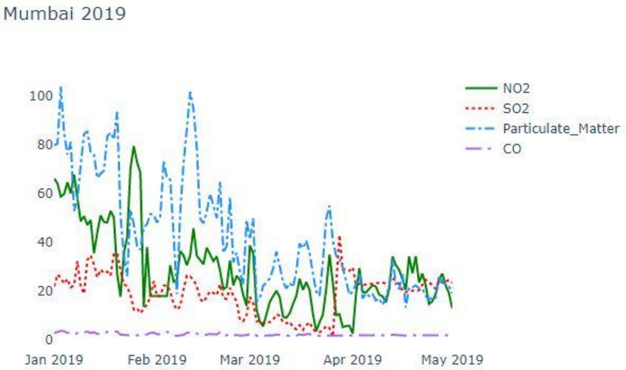


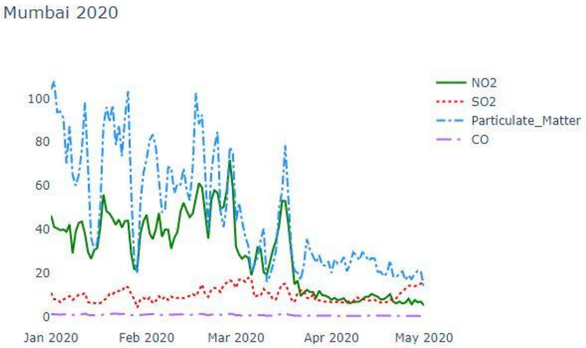
Chennai 2020

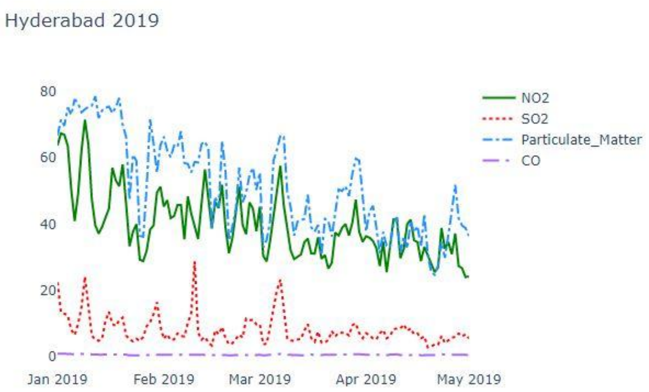


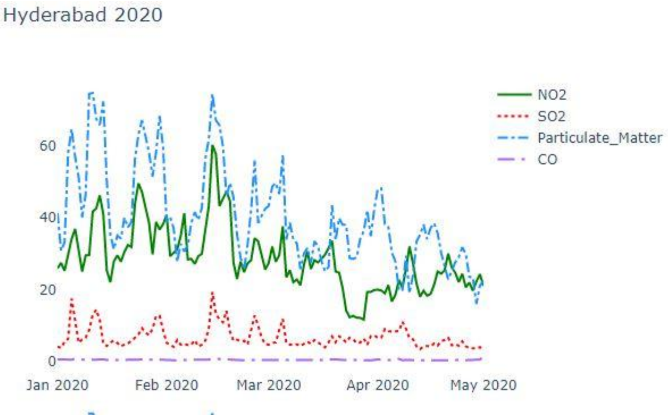

Delhi 2019
NO2
SO2
Particulate_Matter
CO
300
200
100
0
Jan 2019
Feb 2019
Mar 2019
Apr 2019
May 2019


Delhi 2020
NO2
SO2
Particulate_Matter
CO
300
200
100
0
Jan 2020
Feb 2020
Mar 2020
Apr 2020
May 2020


Mumbai 2019
NO2
SO2
Particulate_Matter
CO
100
80
60
40
20
0
Jan 2019
Feb 2019
Mar 2019
Apr 2019
May 2019


Mumbai 2020
NO2
SO2
Particulate_Matter
CO
100
80
60
40
20
0
Jan 2020
Feb 2020
Mar 2020
Apr 2020
May 2020





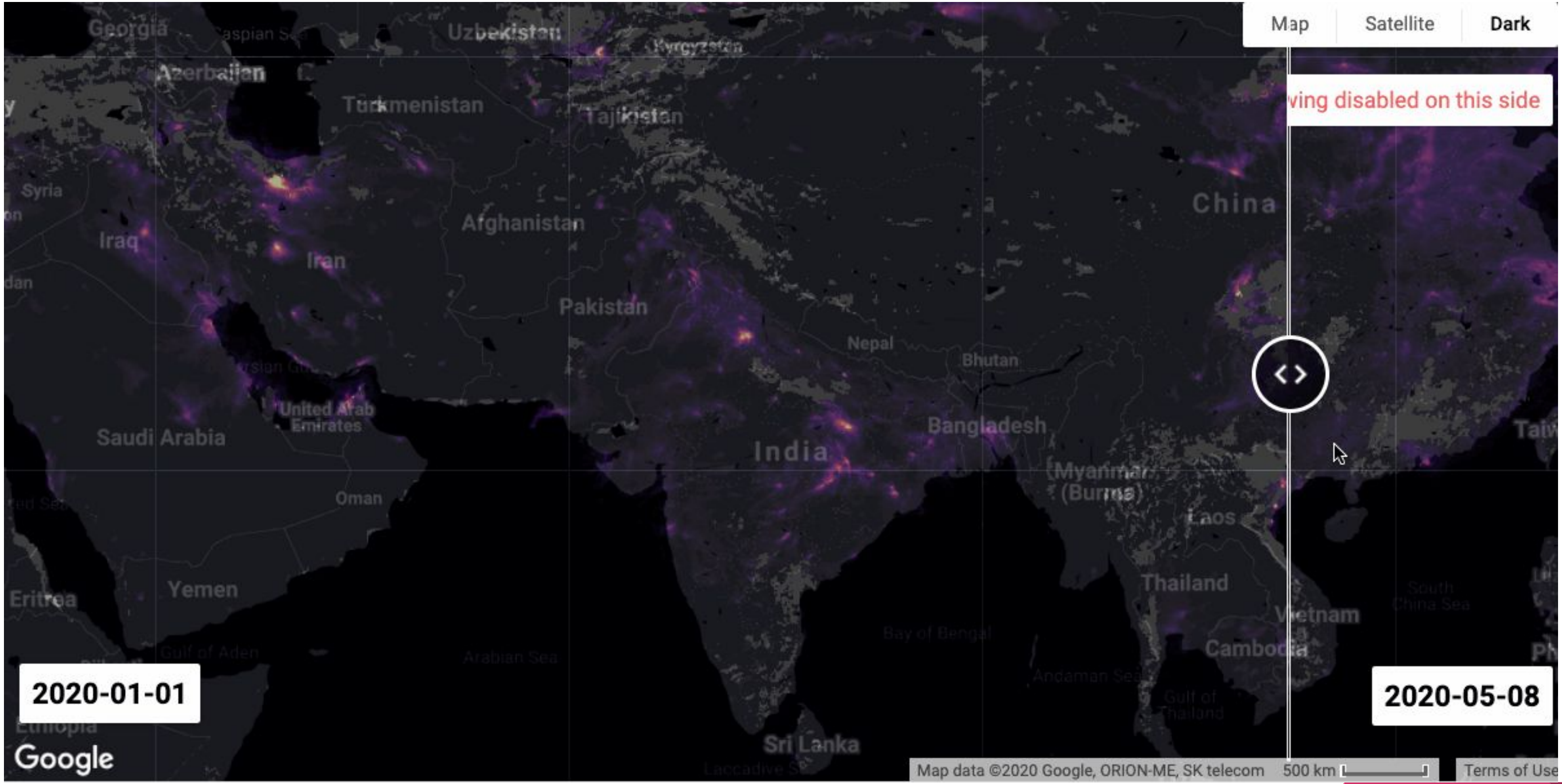
Points to note :

- It is interesting to note that the Pollution level in India generally drops down as summer approaches. This can also be corroborated by the graphs above.
- However, the reduction in march 2020 is more pronounced as compared to march 2019

A Look at Pollution levels through Satellites

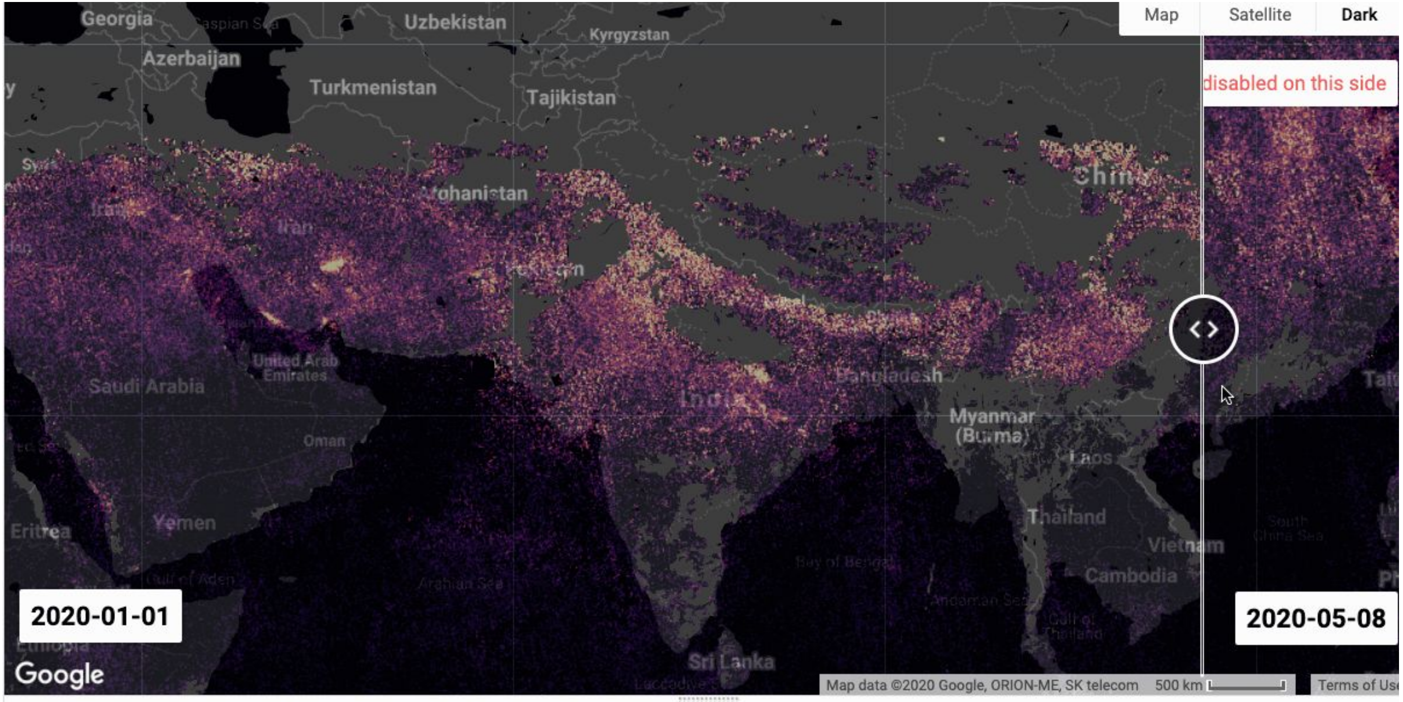
Satellite Imagery is the image of Earth(or other planets) which are collected by imaging satellites. Satellites have been collecting Earth Observation data for decades. Some changes in the concentration of air pollutants are visible from satellites. One such instrument collecting these measurements is the Tropospheric Monitoring Instrument (TROPOMI) sensor onboard the Sentinel 5P OFFL NO2 mission satellite currently in orbit.

Sentinel-5 Precursor (**Sentinel-5P**) is an Earth observation satellite developed by ESA as part of the Copernicus Programme. The Copernicus Programme is dedicated to monitoring air pollution and Sentinel 5P Precursor is its first mission. It consists of an instrument called Tropomi (TROPOspheric Monitoring Instrument) which is a spectrometer to monitor ozone, methane, formaldehyde, aerosol, carbon monoxide, NO2 and SO2 in the atmosphere.



NO2 (μmol/m²), tropospheric vertical column 9-day mean

20 200 400



SO2 (μmol/m²), vertical column 9-day mean

0 500 1000