Blue Sky Analytics Assignment

This set of slides present the approach I took for the assignment, the outcome of the results and the implication of these results on the problem case given in the assignment.

NOTE: Please use present mode to view the presentation in the clearest format

Blue Sky Analytics Assignment

In this solution, I:

1. Identified the most affected developing country during the COVID-19 pandemic and examined the impact COVID-19 had on the water and air quality of the country (India)
2. Used multiple objective time-series data sources to back my results of analysis
3. I correlated the results to the environmental, urban and climatic changes of the chosen country during the period
4. Applied multiple geospatial techniques to examine the effect on satellite data of the world and region in question.
5. Employed machine learning to predict the effect of COVID-19 in the air quality during the time-frame

Blue Sky Analytics Assignment

Blue Sky Analytics

Identifying most affected developing country during the COVID-19 pandemic

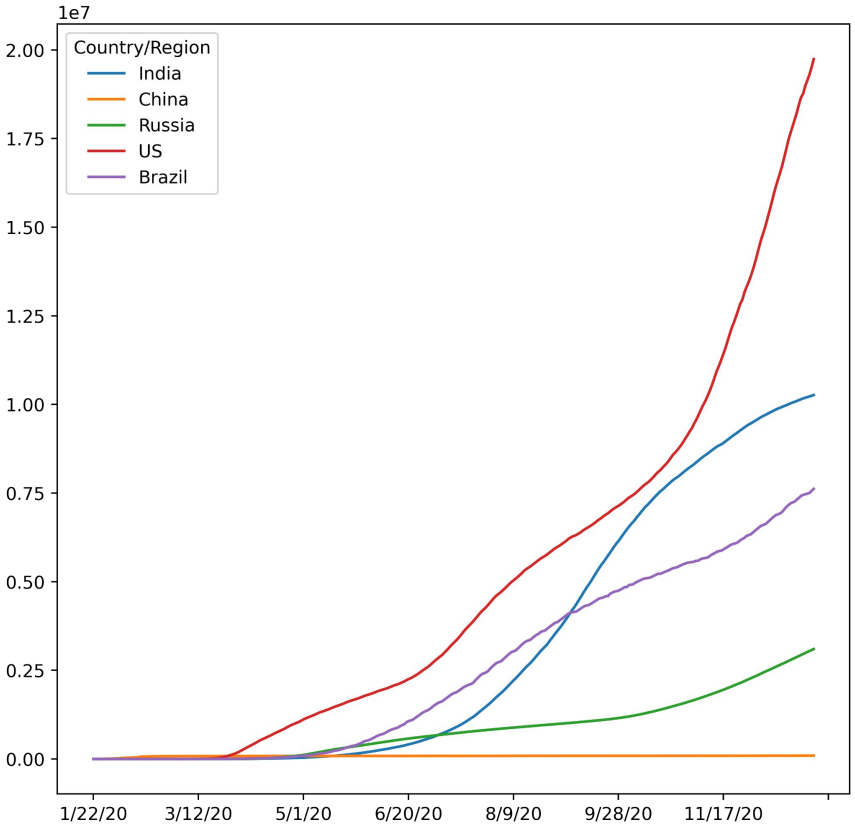
I used the COVID-19 cases data obtained from John Hopkins university GitHub repo. I obtained the data for total cases, recovered and deaths. For this assignment, I decided to focus on the year 2020 as it represents the critical period of outbreak.

We can analyse the most affected countries and see which one to choose for the assignment in the next slide:

Blue Sky Analytics Assignment

Blue Sky Analytics

Identifying most affected developing country during the COVID-19 pandemic



Blue Sky Analytics Assignment

Blue Sky Analytics

Identifying most affected developing country during the COVID-19 pandemic

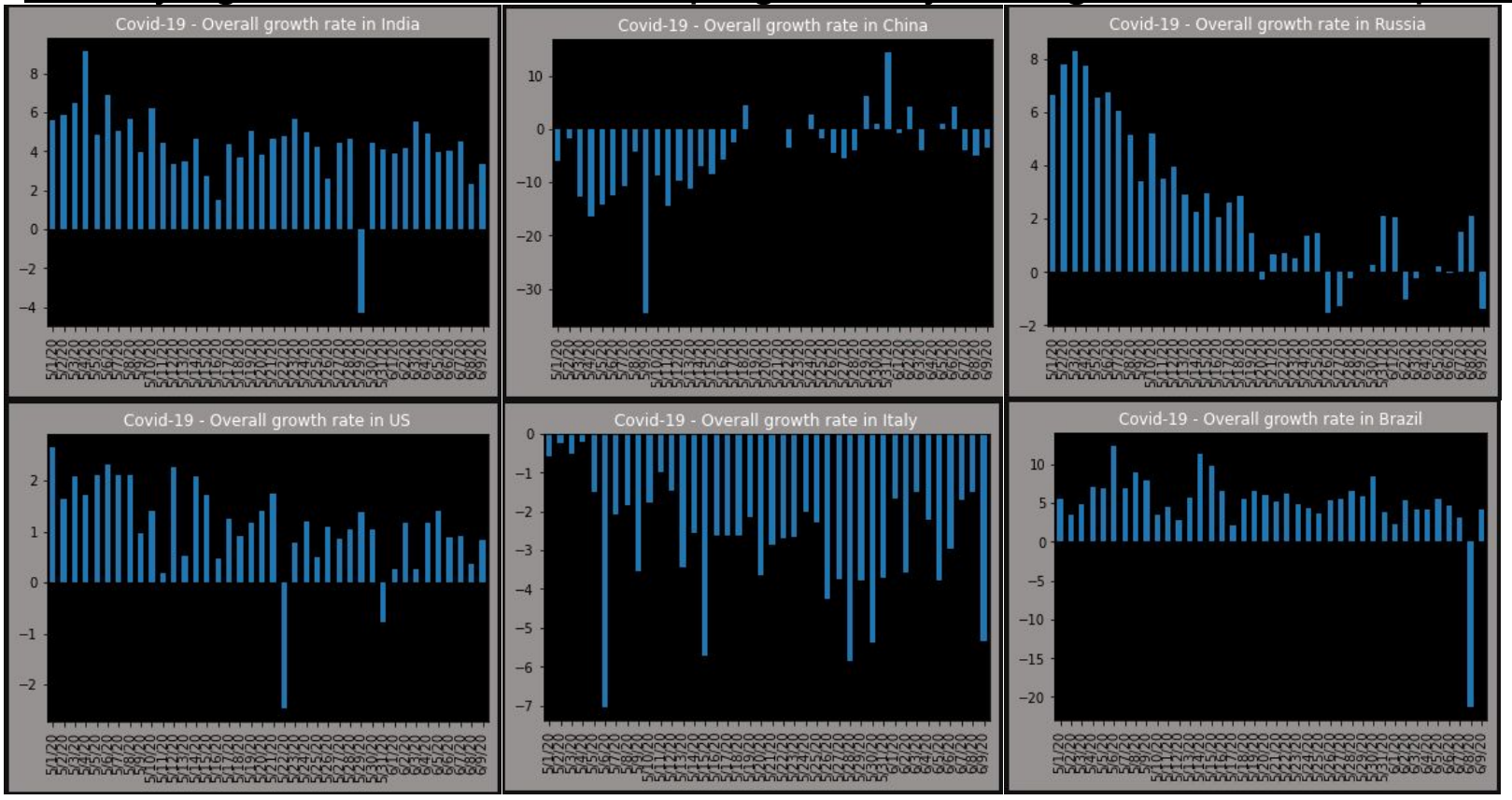
We can see from the previous plot that the most affected countries are the United States, India and Brazil with India being the most affected developing country.

This can also be seen in India’s growth rate visualised among the other critically hit countries during the month of May in the next set of figures

Blue Sky Analytics Assignment

Blue Sky Analytics

1. Identifying most affected developing country during the COVID-19 pandemic



Blue Sky Analytics Assignment

Blue Sky Analytics

Identifying most affected developing country during the COVID-19 pandemic

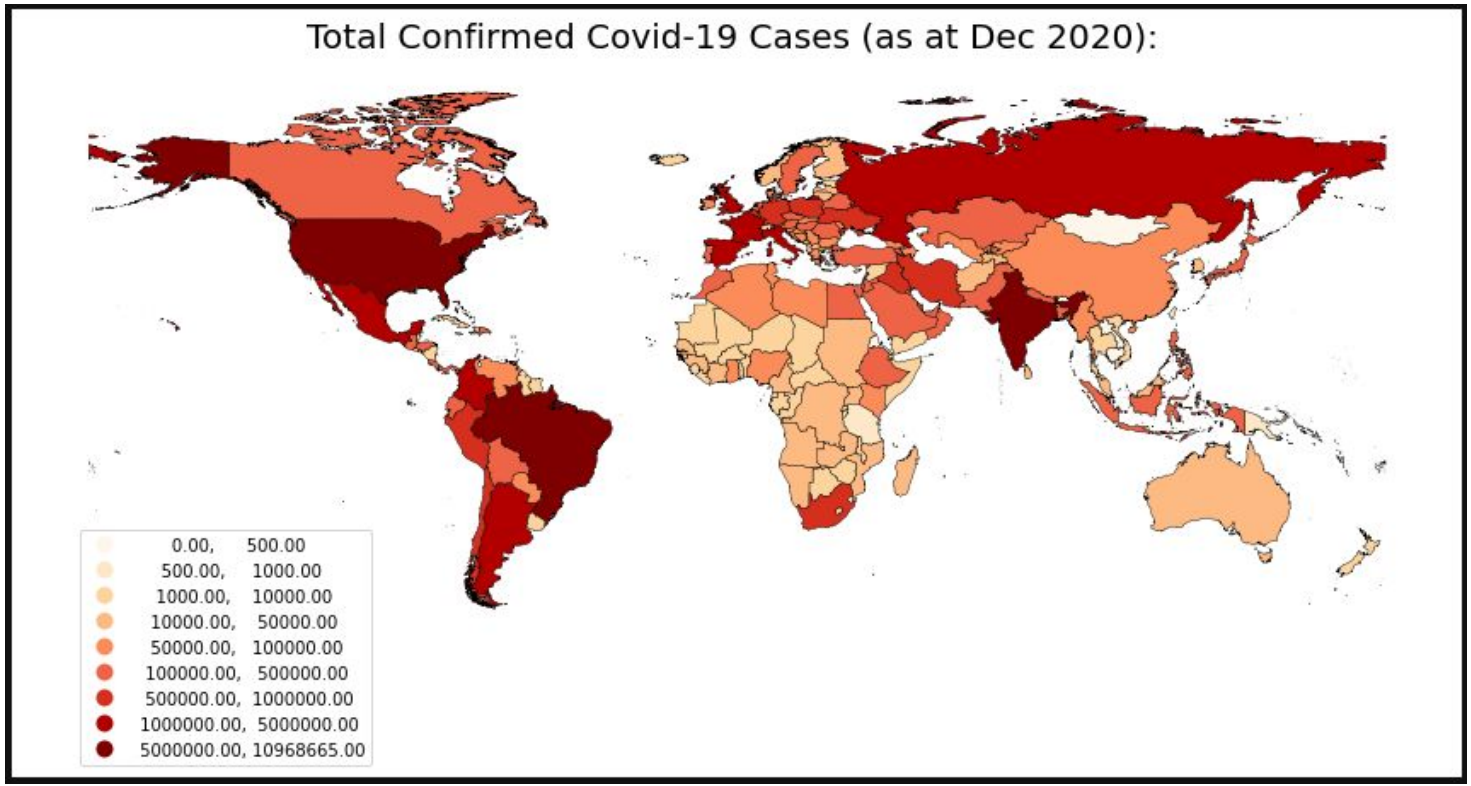
Geospatial analysis warmup:

I used Geopandas with the world shapefile data to visualise the heatmap of all countries in the world with the palette legend representing the confirmed cases to cement my choice of India for the assignment

Blue Sky Analytics Assignment

Blue Sky Analytics

We can see India was one of the “hottest” zones during COVID-19



Blue Sky Analytics Assignment

Blue Sky Analytics

Environmental Changes In Movement Caused by COVID-19

Since we will be discussing about the air and water quality, identifying the effect of COVID-19 on dwellers' movements would be crucial in making necessary assumptions.

I made use of Google's COVID-19 mobility reports for the year 2020 to analyse mobility.

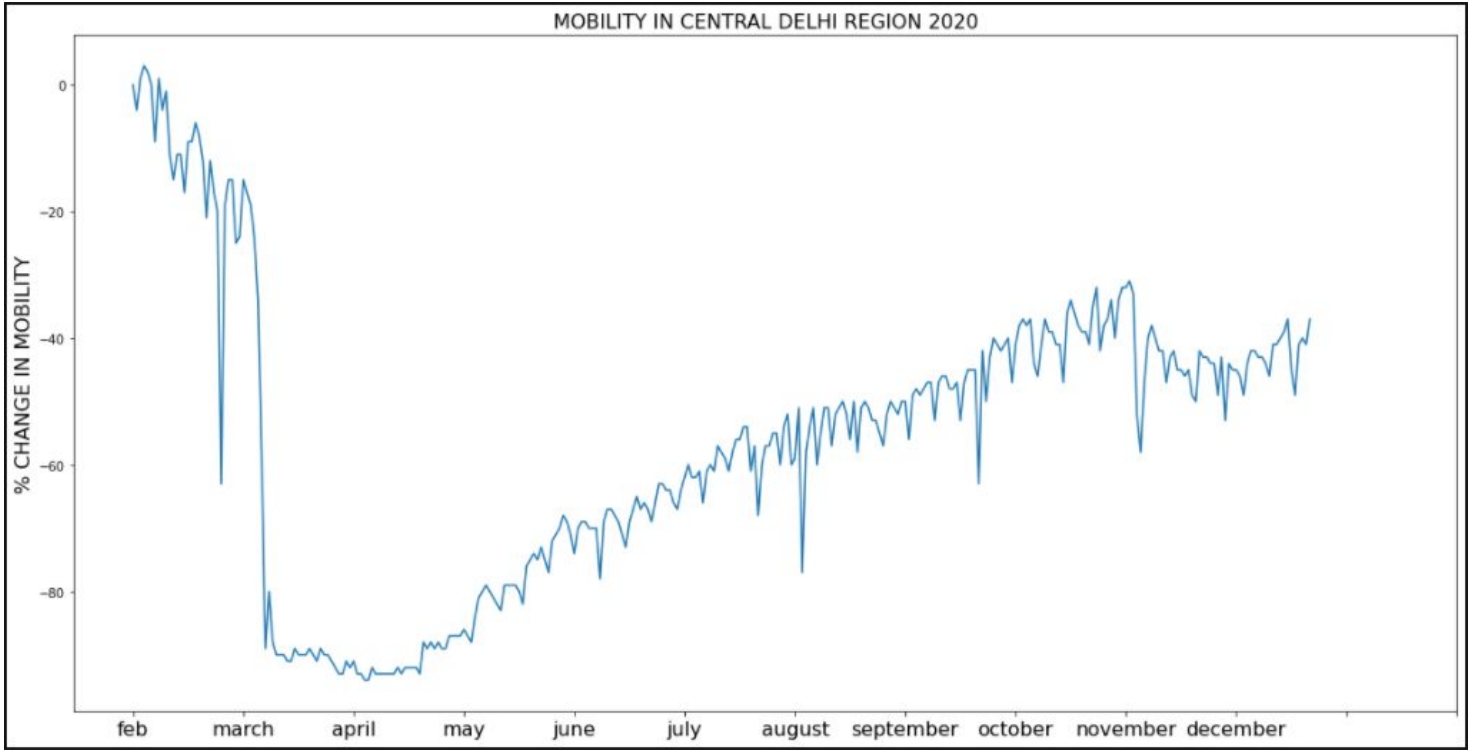
I used Central Delhi as a case study, it represents the typical mobility of India for that year, which we can see in the next slide.

I used the retail and recreation rate of change to factor in mobility.

Blue Sky Analytics Assignment

Blue Sky Analytics

Environmental Changes In Movement Caused by COVID-19



Blue Sky Analytics Assignment

Blue Sky Analytics

Environmental Changes In Movement Caused by COVID-19

We can see that as expected, movements sharply dropped by March due to the lockdown and increased over time as a result of the relaxation of lockdown measures, but movements still fell short of the typical movement before lockdown.

We will then analyse the air quality during this period, particularly the pollutant level and Air Quality Index (AQI).

We expect the trend to follow suit too with this.

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India

I used Sentinel-5p time-series satellite data obtained from Google's earth engine to obtain data for the air quality of every location around the world. Filtered for India as it's our case study.

With this data, we can analyse air pollutants like Nitrogen dioxide (NO_2), Carbon monoxide (CO), Methane (CH_4), Sulphur dioxide (SO_2), Ozone (O_3).

For simplicity, I analysed NO_2 and CO pollutants.

I also used GIS of South-East Asia geojson vector file alongside this satellite data to visualise a heatmap.

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India

south_asia — Features Total: 17, Filtered: 17, Selected: 0

	featurecla	NAME_LONG	cartodb _2
1	Admin-0 country	Afghanistan	Co-sponsor of ...
2	Admin-0 country	Bangladesh	Co-sponsor of ...
3	Admin-0 country	Bhutan	Co-sponsor of ...
4	Admin-0 country	China	Co-sponsor of ...
5	Admin-0 country	Indonesia	Co-sponsor of ...
6	Admin-0 country	Iran	Co-sponsor of ...
7	Admin-0 country	Kyrgyzstan	Co-sponsor of ...
8	Admin-0 country	Sri Lanka	Co-sponsor of ...
9	Admin-0 country	Maldives	Co-sponsor of ...
10	Admin-0 country	Myanmar	Co-sponsor of ...
11	Admin-0 country	Nepal	Co-sponsor of ...
12	Admin-0 country	Pakistan	Didn't co-spon...
13	Admin-0 country	Thailand	Co-sponsor of ...
14	Admin-0 country	Tajikistan	Co-sponsor of ...
15	Admin-0 country	Turkmenistan	Co-sponsor of ...
16	Admin-0 country	Uzbekistan	Co-sponsor of ...
17	Admin-0 country	India	Co-sponsor of ...

India is the 17th country on the vector file, simply extracting the 17th multipolygon and using the satellite data would thus provide real-time analysis of the air quality in the country.

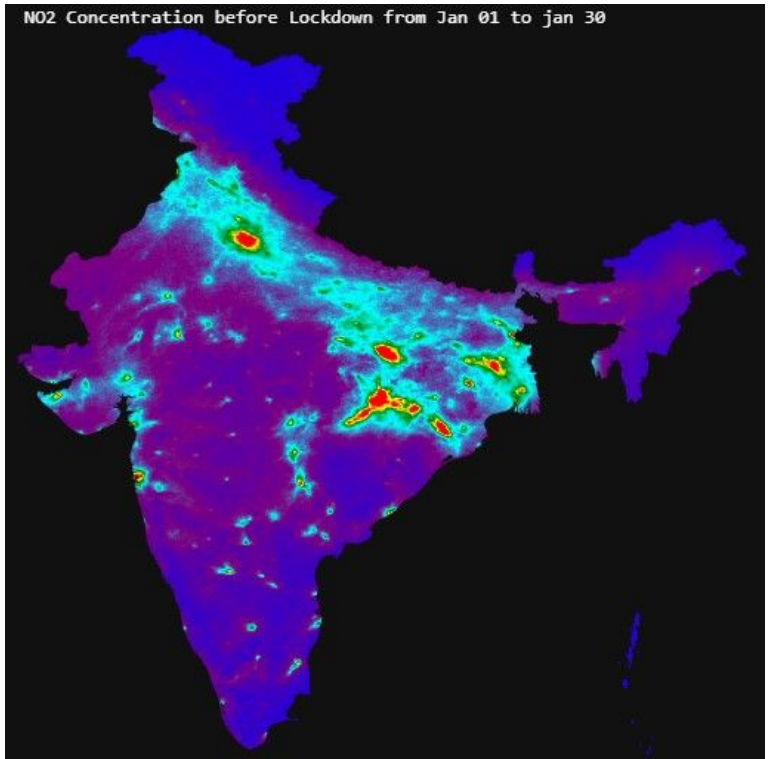
(image at the side is obtained from using QGIS software)

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (NO2)

First, we will analyze the NO2 levels in India before, during and after lockdown



- Before Lockdown

We can see the high NO2 concentration levels, especially in Delhi, Bengal and Jharkhand areas. Where **darker blue** heat maps represent higher level, **lighter blue** heat maps mean lower concentration and **red** heat maps represent the highest concentration levels.

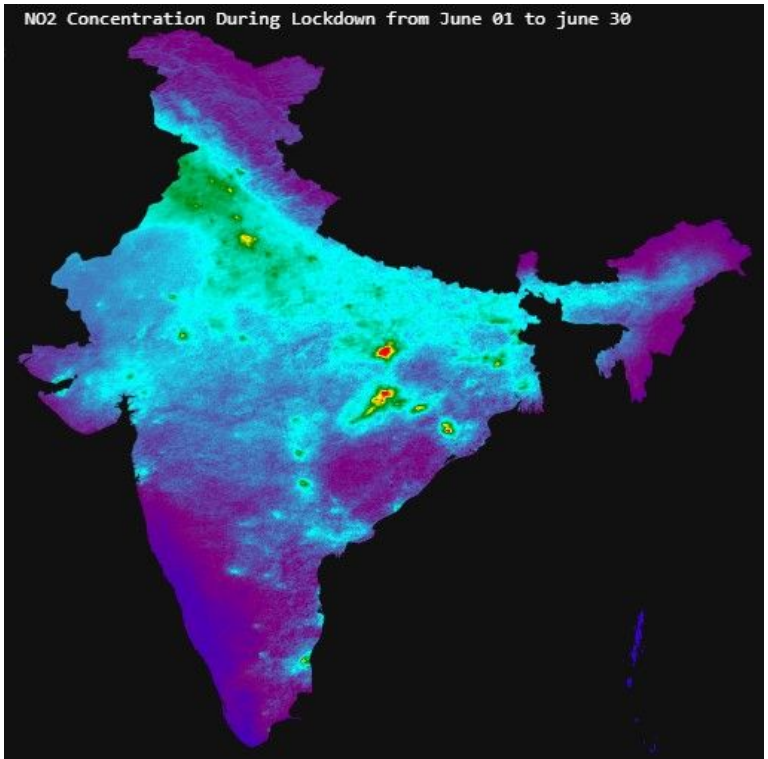
Next, we will see the NO2 levels during lockdown

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (NO2)

During Lockdown



We can see the reduced NO2 concentration levels during lockdown, represented by the **lighter blue** heat maps.

The **red** heat map of “hot zones” also declined.

Hence, the air had better quality during lockdown.

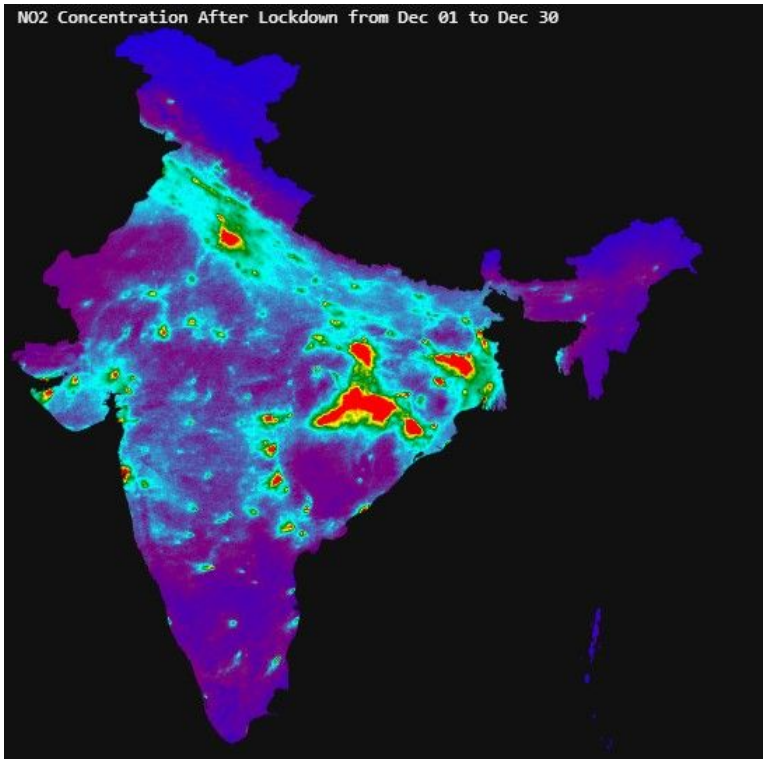
Next, we will see the NO2 levels after lockdown

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (NO2)

After Lockdown



We can see the increased NO2 concentration levels after lockdown again.

The **red** heat map of “hot zones” grew stronger as well, even stronger than the levels before Lockdown.

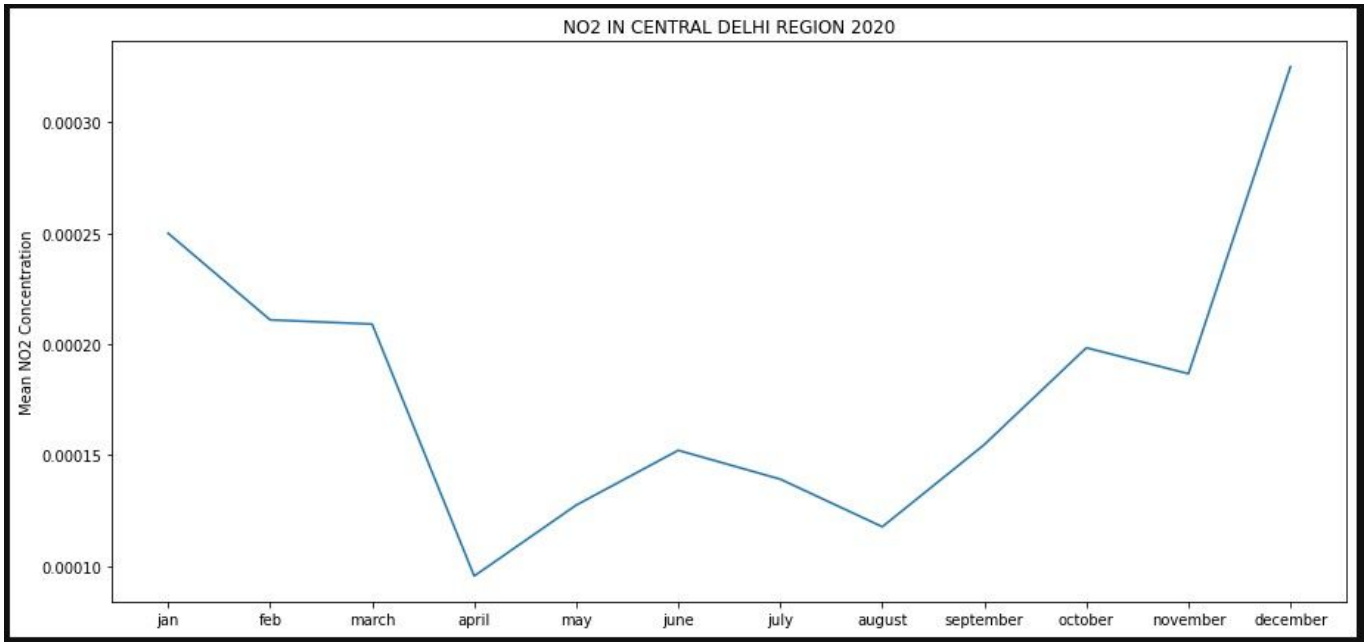
Let’s then see the time-series plot of this data

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (NO2)

- Time-series plot summarising the explained points. High NO2 levels followed by decreased levels during lockdown and even stronger levels after lockdown.

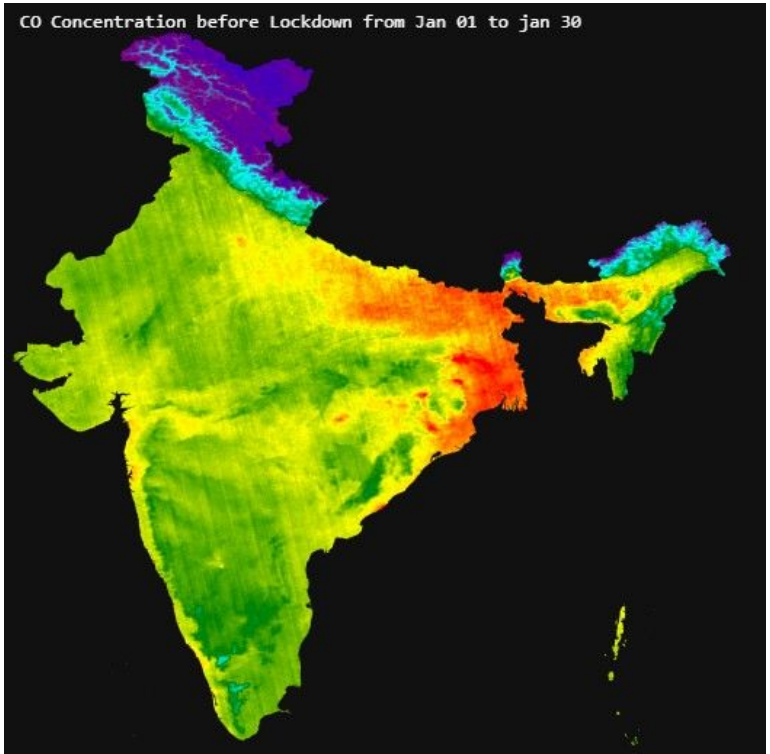


Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (CO)

We will analyze the CO levels in India before, during and after lockdown



- Before Lockdown

We can see a similar pattern like the NO2 levels, with the dark green and red heatmaps signalling the zones with increased CO levels.

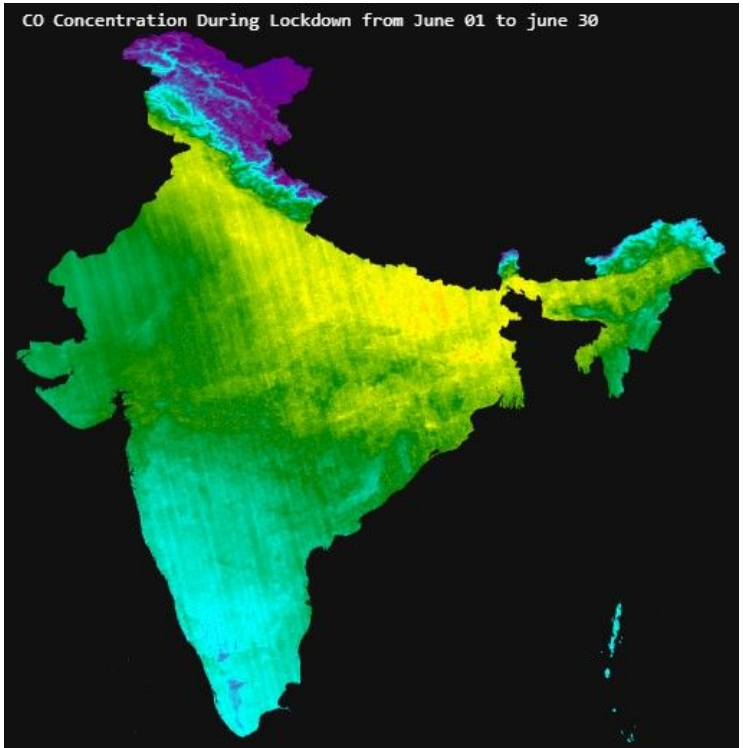
We can visualise the CO concentration during lockdown

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (CO)

During Lockdown



We can see a similar pattern like the NO2 levels during lockdown. The country's heat map looks **"cooler"**, implying reduced CO levels during lockdown, no red zones identifiable.

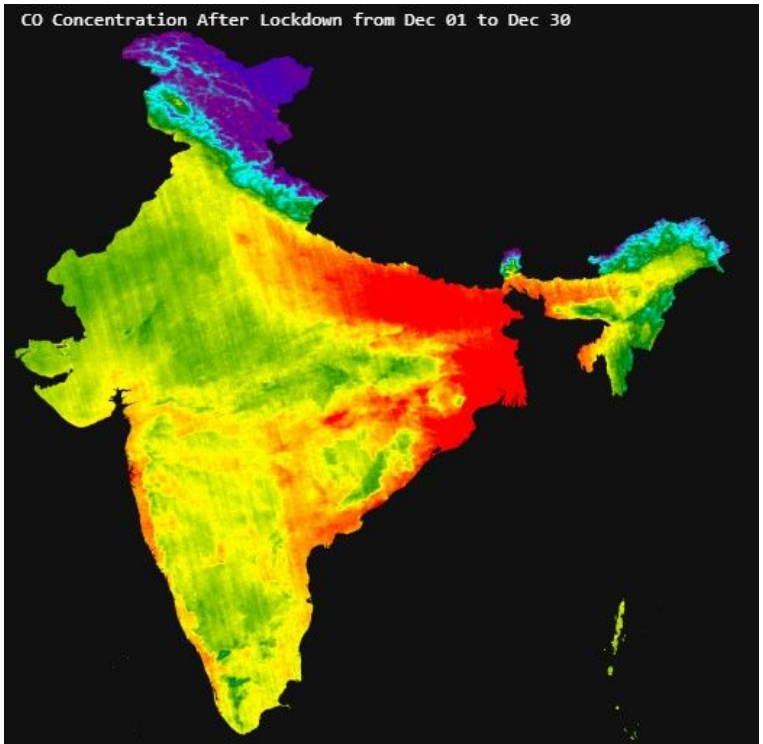
We can visualise the CO concentration after lockdown

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (CO)

After Lockdown



We can see a similar pattern like the NO2 levels after lockdown. Country's heat map looks even **"hotter"** than before lockdown implying increase CO concentration.

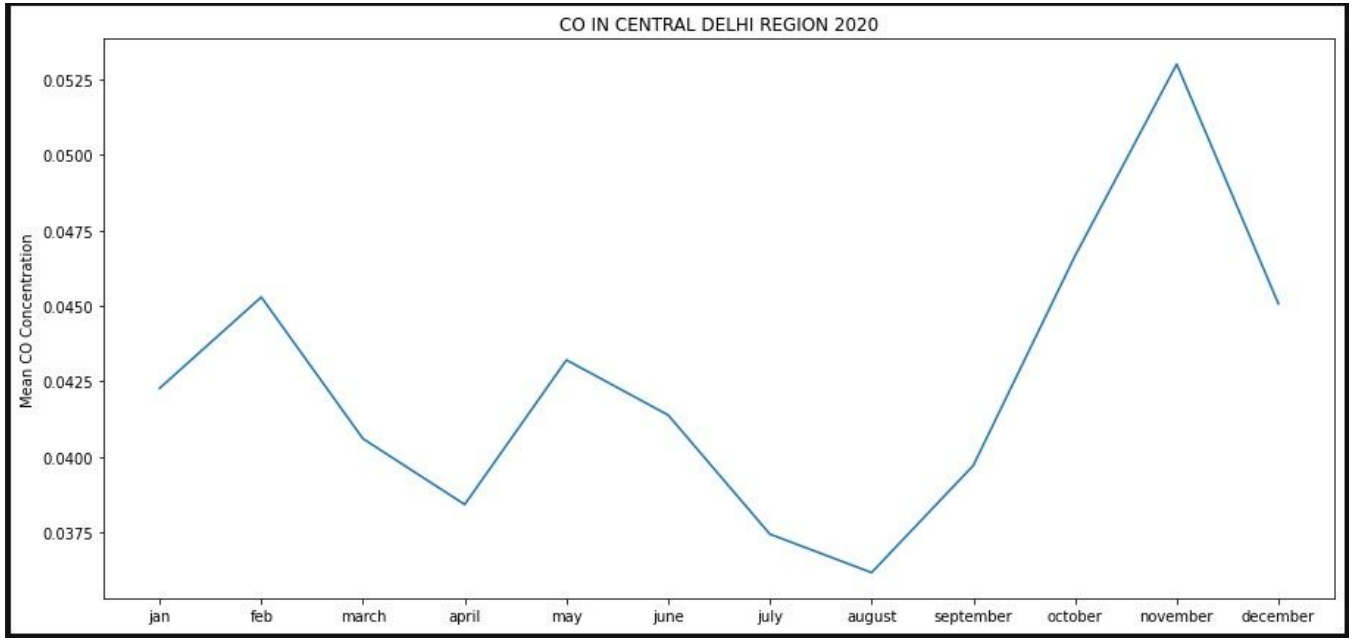
More **red** zones identifiable.

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (CO)

- Time-series plot summarising the explained points. High CO levels followed by decreased levels during lockdown and even stronger levels after lockdown.



Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (Inference)

From the NO2 and CO trends observed over the Indian Peninsula (given in the previous images), it can be seen that there was a reduction in the monthly average NO2 and CO after lockdown, and the reduction continued through April.

More than half of the signal was due to the lockdown, with only the highest concentrations of NO2 (due to industrial activity) remaining near the orissa region, where most of the refinery and industrial capability is located.

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (Inference)

Even during the lockdown, it would be expected that there would be some industrial activity to support essential services although the activities would be far less. The trend in NO2 is correlated to the mobility pattern observed during the lockdown which is shown in slide 10 (Environmental Changes In Movement Caused by COVID-19).

We will further use another objective data source to inspect the air quality of India, focusing on the air quality index too here:

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (AQI Analysis)

Data Sources:

1. Air Quality of India (2015-2020):

   Central Pollution Control Board (Central Pollution Control Board: CPCB)
2. Indian Cities Database:

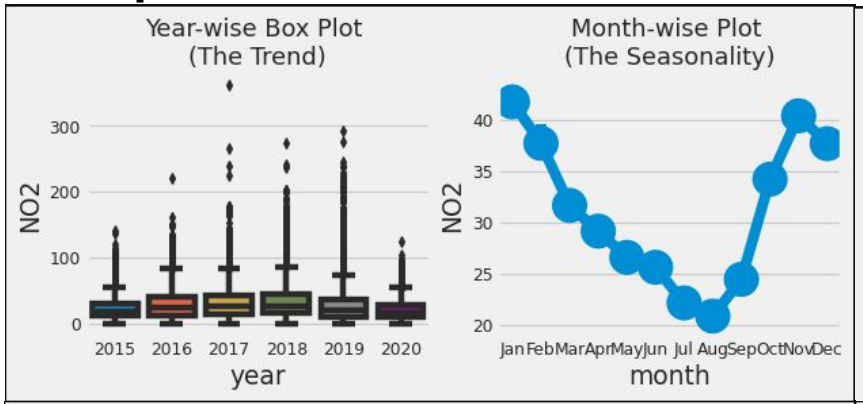
   simplemaps.com

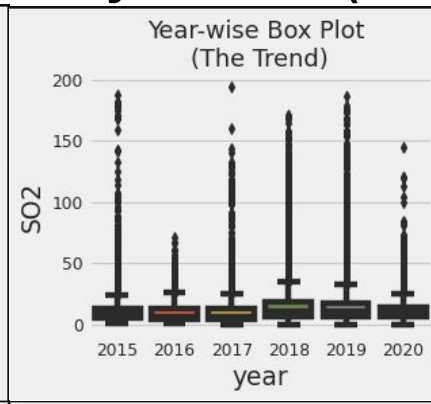
We will then analyse the air quality with respect to the pollutants (NO2, SO2, Benzene, Toluene and Xylene (BTX), Particulate matter (PM10) during 2015-2020 in the next figure

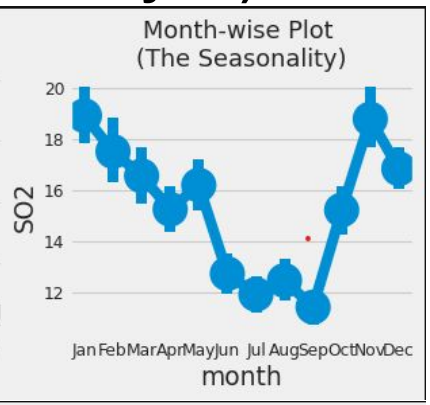
Blue Sky Analytics Assignment

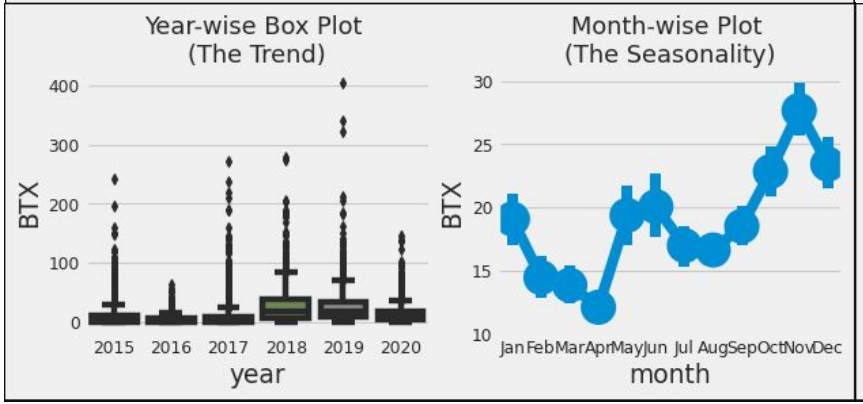
Blue Sky Analytics

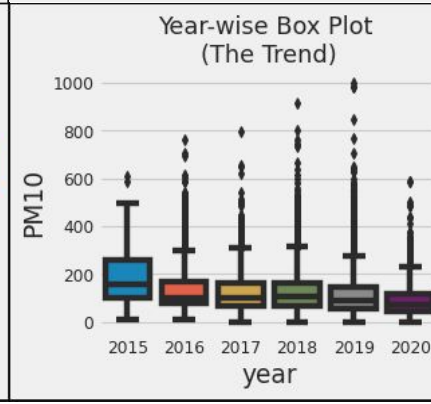
Impact of COVID-19 on the Air Quality of India (AQI Analysis)

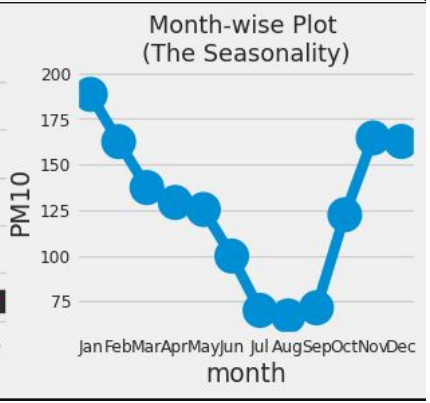












Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (AQI Analysis)

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.

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (AQI Analysis)

We can see that over the years, the pollution levels declined in 2020, during the COVID-19 pandemic lockdown.

Let us then inspect the AQI for the biggest cities in India before lockdown measures and after lockdown measures. The AQI has certain rankings from good, satisfactory, moderate, poor, very poor and extreme.

I then used the AQI values to visualise a pointer for these cities based on the scales.

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (AQI Analysis)

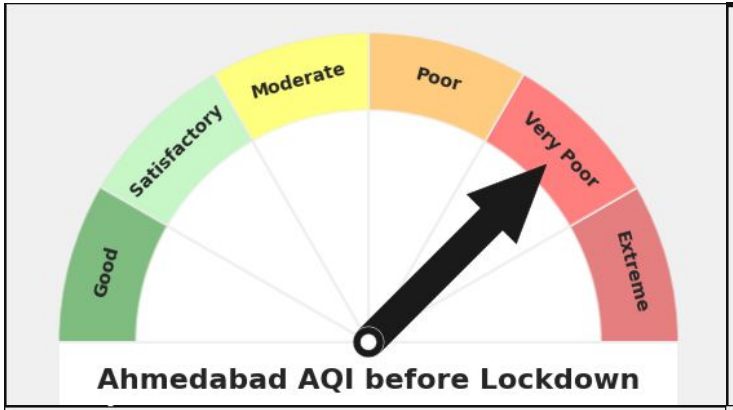
As seen, the air quality was way better after lockdown, indicated by a lower AQI (the lower the AQI), the better the air quality in a region.

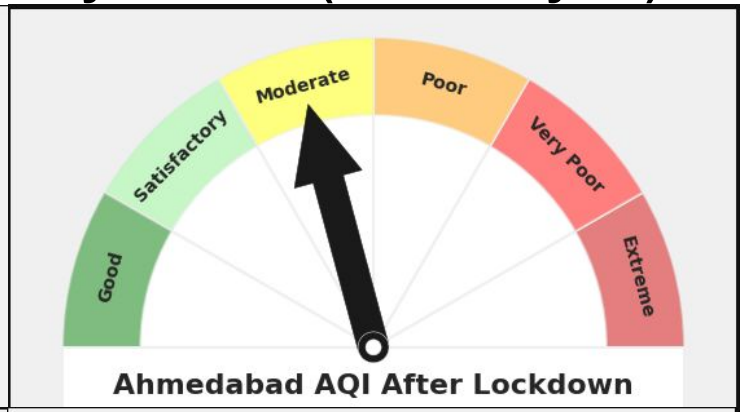
City	Before Lockdown (January - March 2020)	After Lockdown (March - May 2020)
Ahmedabad	384	128
Bengaluru	96	68
Chennai	80	62
Delhi	246	107
Hyderabad	94	66
Mumbai	149	74

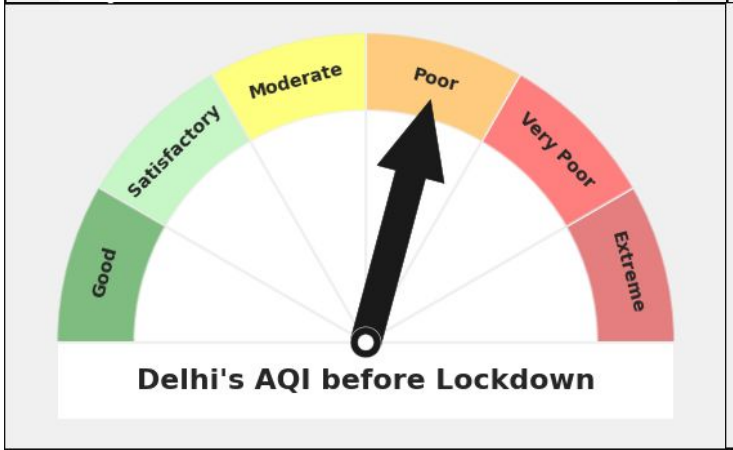
Blue Sky Analytics Assignment

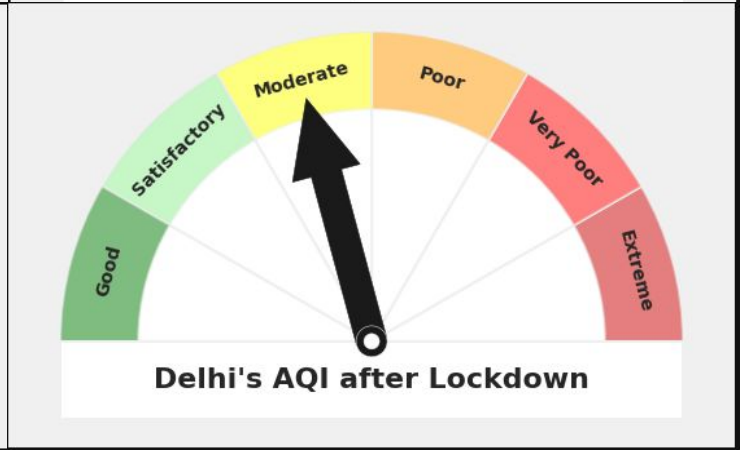
Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (AQI Analysis)





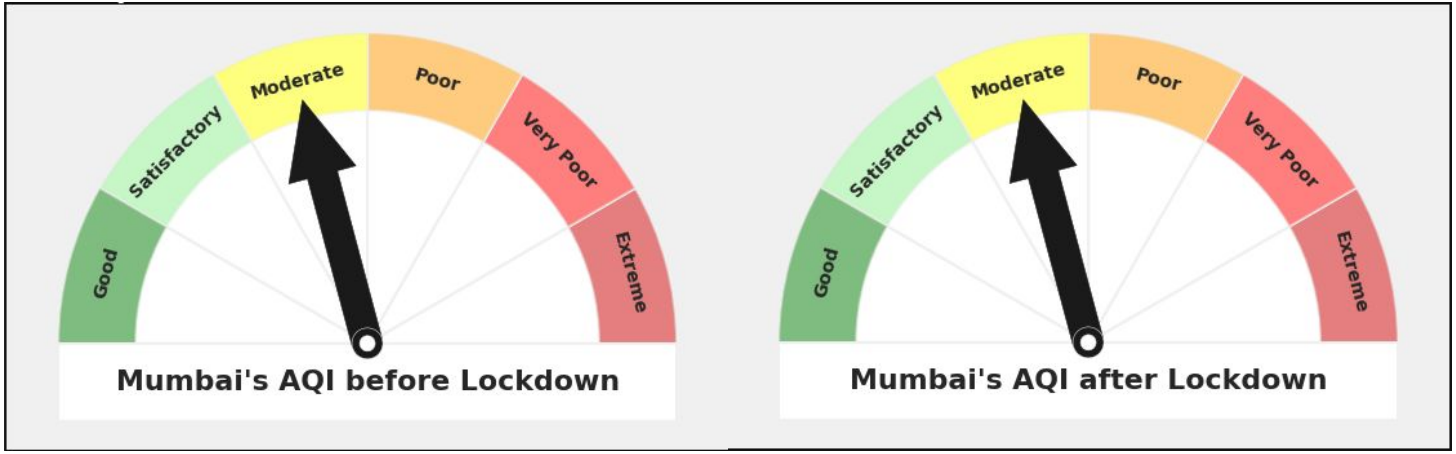




Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (AQI Analysis)



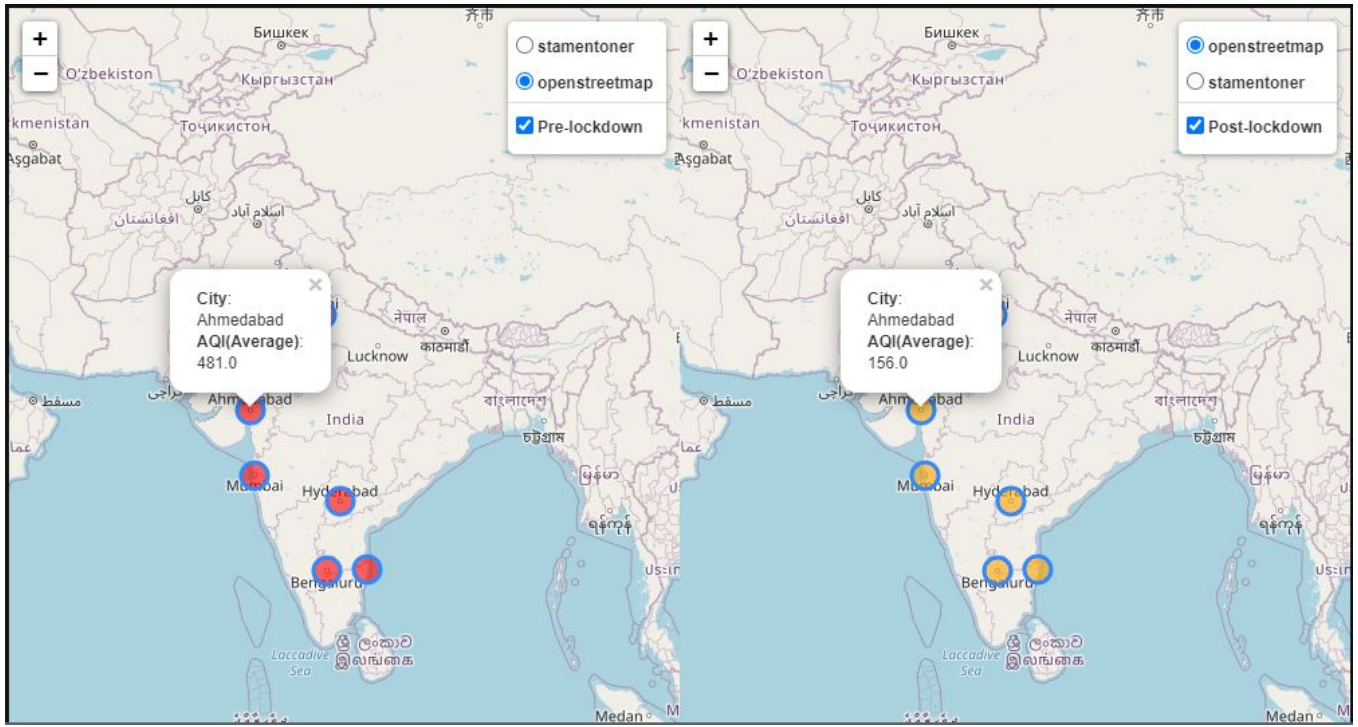
I used geospatial analysis on openstreetmap and stamen toner data and obtained results for the AQI both pre-lockdown and post-lockdown

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (AQI Analysis)

- Openstreetmap view

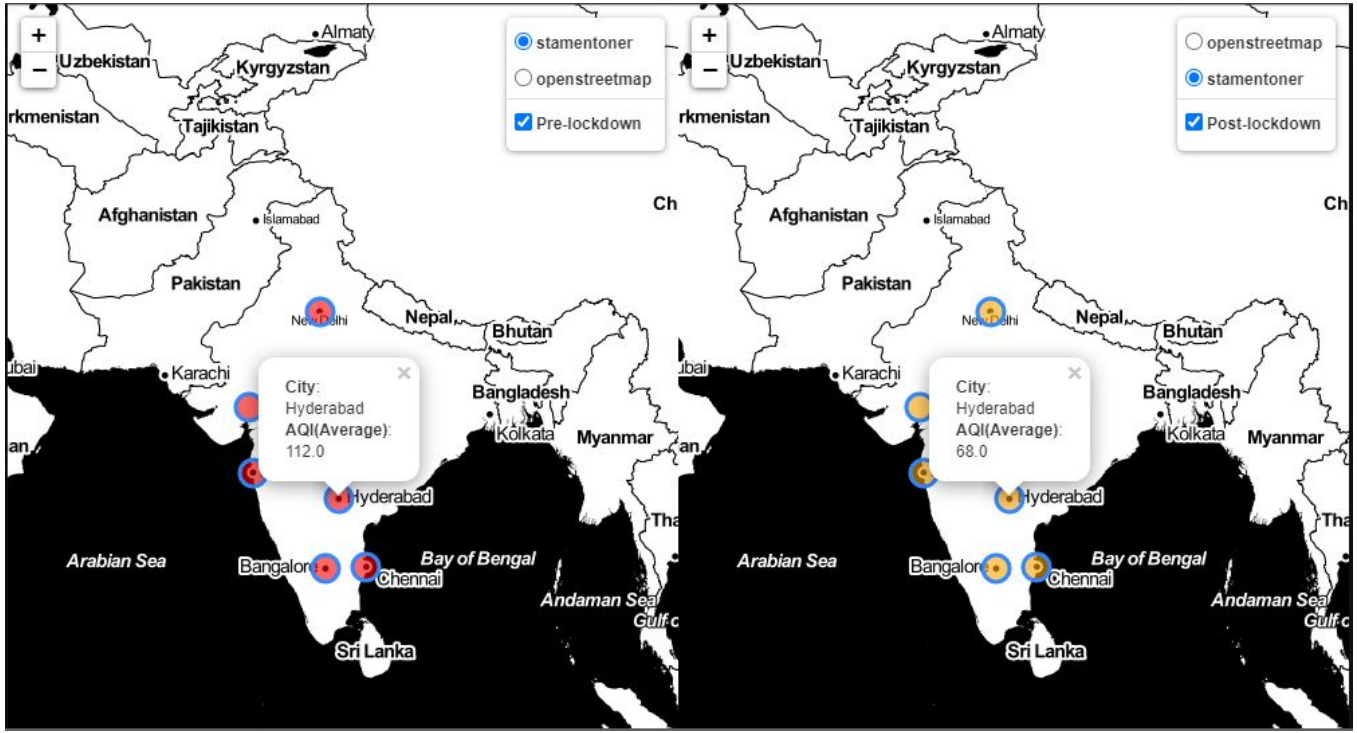


Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (AQI Analysis)

- Stamentoner view



Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (ML Prediction)

I then applied machine learning to predict the Air Quality Index (AQI) of a region in India, from 2020 to 2021.

I used a deep learning model and time-series data of the AQI from 2019 to 2021.

I decided to build the model around Delhi's data as a case study

Data Source:

1. Daily Air Pollution Data - India & USA

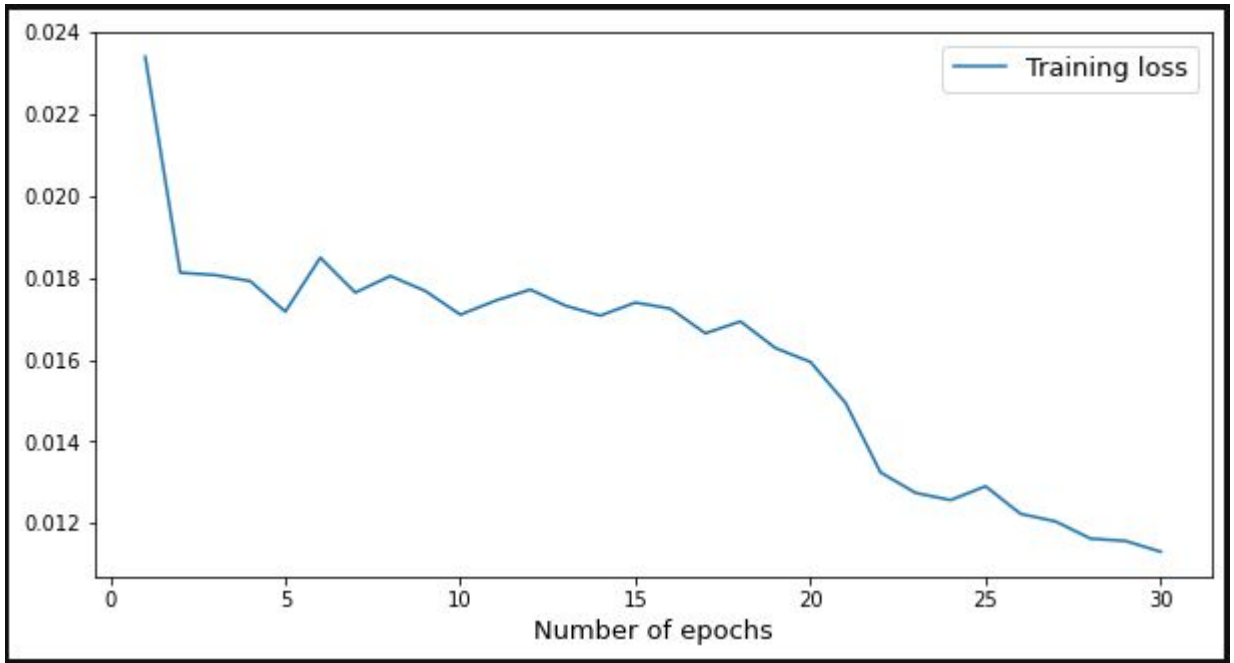
   https://aqicn.org

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (ML Prediction)

Training loss against number of epochs is visualised below. The model does a good job of making predictions, seen by the low, declining loss with number of epochs



Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (ML Prediction)

NOTE:

Adam optimiser was used with the mean squared error loss function, the appropriate function for regression type of tasks.

Since it was a prediction problem of time-series data, I used a Recurrent Neural Network (RNN) model.

As usual, Delhi's data was split into training and test data. Training data was then further split into training and validation data.

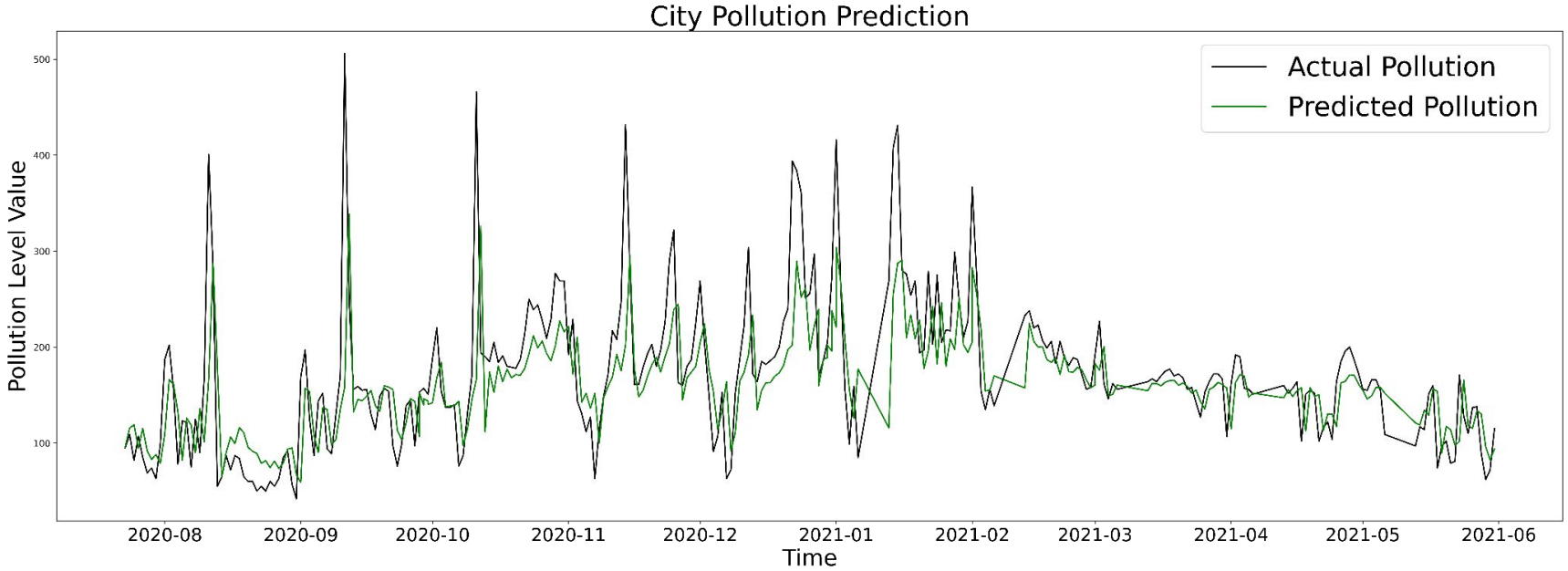
The architecture of the model and details can be found in the Jupyter notebook in my shared GitHub repo.

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Air Quality of India (ML Prediction)

We can see the great prediction results of the model, despite the low number of epochs. Thus, with more training, we will get even better results!



Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Water Quality of India

I took a brief look at how COVID-19 affected the water quality of India.

Water pollution remained a major concern in the last few decades as these were gradually deteriorating in many spheres including the hydrosphere.

As the nation-wide lockdown period in India completed more than two months, these images in the next figures helped understand the short-term environmental changes.

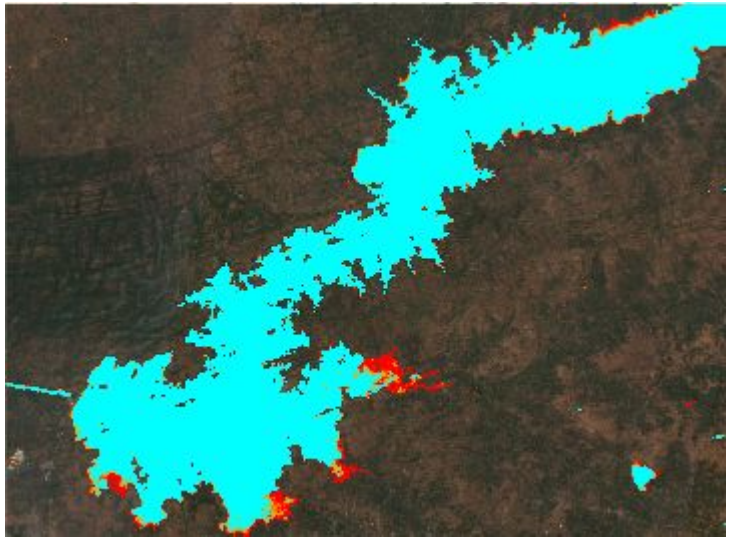
Tapi River was my study area to assess its water quality. Tapi River is one of the longest rivers of Western India

Blue Sky Analytics Assignment

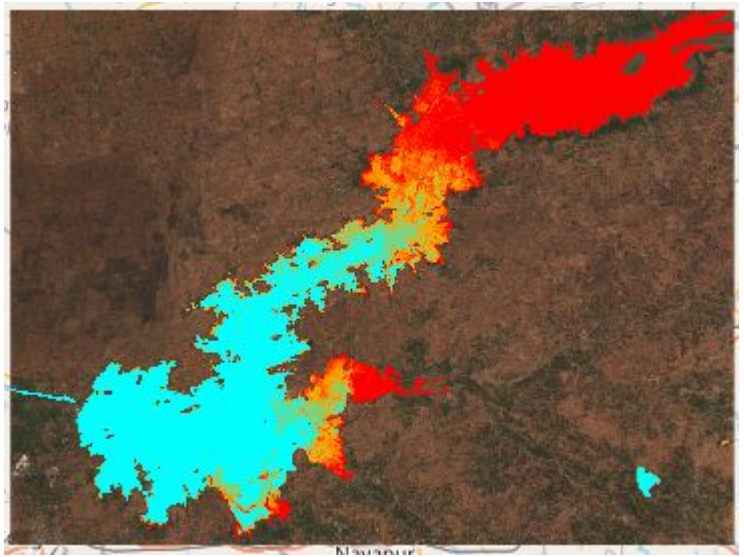
Blue Sky Analytics

Impact of COVID-19 on the Water Quality of India

Observing the water quality during January and April 2020 (before and during lockdown)



January 2020



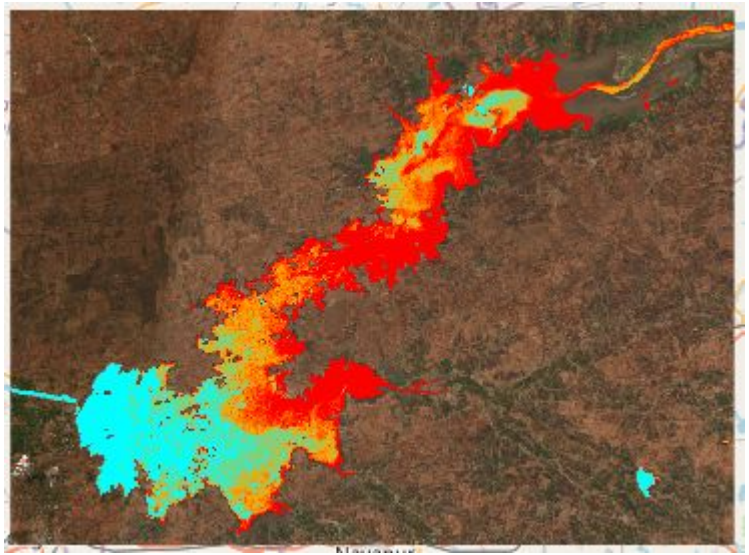
April 2020

Blue Sky Analytics Assignment

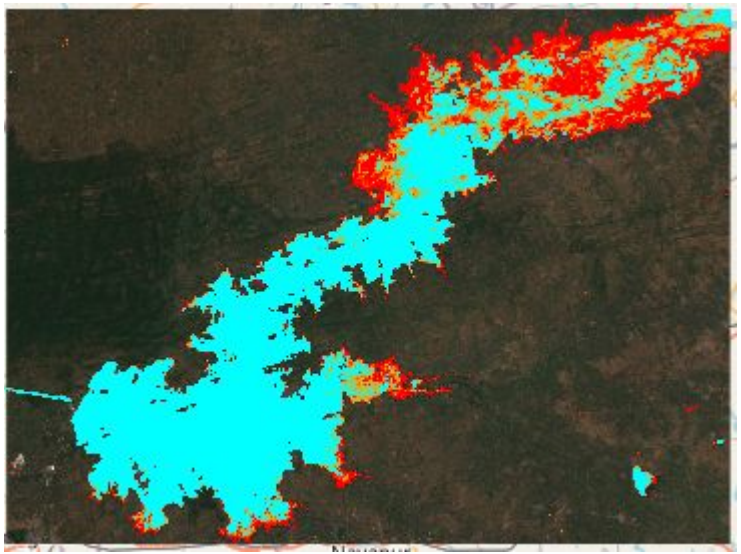
Blue Sky Analytics

Impact of COVID-19 on the Water Quality of India

Observing the water quality during May and December 2020 (during and after lockdown)



May 2020



December 2020

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Water Quality of India

Using remote sensing data, it is displayed the improvements in ambient water quality in terms of decreased chlorophyll-a levels for a section of the Tapi River in the Gujrat region of India.

Thus, the closure of anthropogenic and industrial activities has resulted in the decline of pollutants concentration in the Tapi River. The upsurge in Chlorophyll-a concentration observed may be supported by the fact that the natural river sedimentation can also alter the those levels in water bodies. The flow of sediments is governed by many factors such as precipitation, water and soil conservation projects, construction of reservoirs, carbonate weathering, dilution, etc.

Blue Sky Analytics Assignment

Blue Sky Analytics

Impact of COVID-19 on the Water Quality of India

An average decrease of chlorophyll-a (NDCI) concentrations during lockdown compared to the pre-lockdown period despite the domestic wastewater continuing to drain into the Tapi River shows that the major deteriorating factor of the river is the industrial sewage discharge and other outdoor activities, which severely declined during the lockdown measures, hence severe pollution stopped.

Thus, the changes pre and during lockdown are very significant to understand the short-term, as well as the long-term effect of anthropogenic activities on the hydrosphere

Blue Sky Analytics Assignment

Blue Sky Analytics

Conclusion

I have successfully implemented every task on the shared assignment. I analysed the effect of COVID-19 on the air and water quality of India, using ground data and satellite data, while correlating these effects with the environment. I used multiple data sources to ensure consistency of results in different geographic regions and explored changes in earth-related attributes.

I also applied machine learning and geospatial analysis as needed.

There were no major issues in the data or methodologies but further training on the machine learning would improve prediction results as well as completeness of the air quality index data.

Blue Sky Analytics Assignment

All tasks in the assignment have been completed, with some added useful tasks.

Feedback and comments on improvements would be welcomed as well, thank you!