**Exploratory Air Quality Analysis in India**

To analyse the effect of the pandemic on the climate, I wanted to look at Air quality in India, since, it has been well documented that the quality has taken a hit due to factors like industrialization and pollution in the country. I got city wise and daily measures of air quality (AQI) and levels of all the components in it from the following link : <https://www.kaggle.com/rohanrao/air-quality-data-in-india/version/12>

After processing the data and initially looking at the kind of data I was working with, it was important to look at the lockdown and non-lockdown distributions of AQI for a couple of cities I wanted to analyse, namely, Delhi and Patna. Since, the distributions were more or less normally distributed, we can say, that the mean would be an accurate indicator of the entire population. I took the months of March, April and May in 2018 and 2020, which were non-lockdown and lockdown periods respectively.

For geo-spatial analysis, I obtained geo-spatial data, i.e. shape files for all cities in India from the following link : <https://www.diva-gis.org/gdata> . For the purposes of efficiency and so that I don’t have to load data for all cities into memory every time I re-run my code, I recreated the structure of the file and saved only data for Delhi and Patna. After some initial preprocessing and making the data analysis ready, I used visualizations to look at some relationships.

***Below are the key takeaways:***

1. From the visualizations it was clear that the 2020 AQI in Patna and Delhi was much lower (around 134 and 142 respectively) as compared to that in 2018, when it was much higher (around 235 and 215 respectively). Thus, the impact of the lockdown is clearly visible.
2. If we compare the two cities, we can see that the AQI in Delhi was higher than that of Patna in 2018 but during the lockdown period in 2020, Delhi's AQI fell even below that of Patna.
3. The AQI dropped quite a lot in both cities during the lockdown, almost more than 30% and 40% in Patna and Delhi respectively.
4. It seems that vegetative areas and built-up areas have an effect on the air quality. Delhi has a higher AQI than Patna but during lockdown, its AQI is lower than that of Patna. This change can be attributed to the amount of vegetative cover. As is already well-known, more vegetative areas have purer air and I found that Delhi has a larger area covered by vegetation as compared to that of Patna. Also, the built-up area of Delhi is slightly less than that of Patna. Thus, higher vegetative cover and a lower built up area could contribute to improved air quality.
5. NO2, CO and SO2 are fairly strongly correlated to AQI as is time. During the lockdown period, NH3 was positively correlated to AQI whereas it was negatively correlated during the non-lockdown period.
6. For 2018, Visakhapatnam, Bangalore, Chennai, Hyderabad and Thiruvananthapuram seem to be very strongly correlated to the AQI, whereas, during 2020, Mumbai, Hyderabad, Kolkata, Chennai, Amritsar, Kochi, Visakhapatnam, Emakulam, Jaipur, Bangalore and Thiruvananthapuram are strongly correlated. Thus, as compared to 2018, a lot more cities during the lockdown period have had an impact on air quality and this could be a good indicator that the activities taking place in these cities during lockdown had an effect on the AQI.
7. On looking at the proportions of Carbon Monoxide (CO) specifically, it was evident that there was a significant drop in the levels in the months when lockdown was imposed. And not only that, the levels seemed to be gradually dropping month on month as well.
8. In case of Nitrogen Dioxide (NO2), the levels actually increased in the first month of lockdown, i.e. March, 2020 as compared to that in March, 2018. But post this, the levels start dropping.
9. Looking at week on week averages of CO levels for Bangalore and Lucknow, we can see that though they lie on different scales, there is a consistent gradual pattern from the non-lockdown period to the lockdown period and we can see an evident dip as we enter the weeks of 2020.
10. We can also conclude from this periodic graph that before 2020, the CO levels don’t seem to drop and thus, this can only be observed during the lockdown.
11. In terms of weekly averages of Nitrogen Dioxide (NO2) levels during the lockdown months, the patterns are inconsistent as some weeks record higher levels than others and so there are dips as well as peaks but when comparing these levels to those in 2018, there has been a percentage decrease at every week.
12. If we compare the average Ammonia (NH3) levels as well as the number of operating factories during the two periods in Chennai, we can see that the levels decrease during lockdown and to add to this, a lower number of factories also seem to be in effect. Thus, there is possibly a correlation between AQI and the number of operating factories.
13. All in all, there seems to be sufficient evidence that there has been improvement in air quality during the lockdown. This can be down to lesser number of factories operating during this period. Similarly, we can say that increased vegetative cover or decreased built up land can also result in improved air quality.

In terms of developments or further analysis, I feel I could have used correlation tests if I had data for factories/traffic or any other activities taking place in the cities at a more granular level. Further, a regression model can be created or time series can be performed to predict AQI for the future, which, in turn, could be an indicator of the relationship between the pandemic and the air quality. Also, more modifications could be done to the geo-spatial visualizations so as to incorporate more details as well as aesthetics.