Indian Institute of Technology Jodhpur
Department of Mathematics

Impact of COVID-19 lockdown on the Air Quality in Indian cities

Instructor:
Dr. Kiran Kumar Hiremath

Group 4:
Annem Khyathi Reddy
Ashish Ledalla
Harshita
Saptarshi Sardar
Yash Prakash Mishra

1 Abstract

The COVID-19 pandemic has had a substantial impact on global economic activity. Despite the fact that it consumed a large number of human breaths and increased unemployment, it has a favourable impact on the environment. To slow the spread of the sickness, the government placed a rigorous lockdown on its residents, which had a positive impact on the environment. The influence of lockdown on the environment was evaluated using air pollutant concentrations in this study. The Air Quality Index (AQI) is calculated based on the concentration of air pollutants. The Air Quality Index shows which cities in our country are the most and least polluted. A higher AQI number indicates a more polluted city, while a lower AQI value indicates a less polluted city. The impact of lockdown on air quality was investigated in this study, and it was discovered that during the lockdown period, air pollutant concentrations decreased in every city across the country. The main impacting air concentrators that control the air quality of all the selected sites before, during, and after lockdown have also been identified as PM 2.5 and PM 10.

2 Objective

Our objective for this analysis is understand the various effects of the COVID-19 pandemic and the lockdown due to it on various cities in India.

3 Introduction

Air contamination in India is an intense issue and it can't be overlooked. The developing air contamination in the nation has been one of the significant worries for both the public authority and the residents and it adds to the unexpected losses of 2 million Indians consistently. Many moves have been made by the public authority to control contamination, yet at the same time the issue stands.The starter air polluters in metropolitan regions involves carbon dioxide (CO_2), carbon monoxide (CO), nitrogen oxides (NO_2), nitrogen monoxide (NO), and particulate matter $PM_{2.5}$, PM_{10}. All things considered, the significant air contamination factor is $PM_{2.5}$ or particulate matter which is up to 2.5 microns in breadth.

Corona virus or COVID-19 is a sporadic type virus that infects human beings, , and normally causes respiratory contamination that varies from the normal cold to novel extreme illnesses. The simple and fast spread nature of COVID-19 compromised the entire world. As safeguards, maximum country have found a way different ways to control the pace of spread of COVID-19. Social distancing ought to be kept up with and individuals are proposed to remain at home as COVID-19 is a contagious sickness. This pandemic has been the most exceedingly awful involvement with human existence which influenced their ways of life too. The greater part of the nations have applied severe lockdown all through the nation as a careful step.

Coronavirus began affecting India in February 2020. India experienced a jump in COVID-19 affirmed cases crossing 100. The WHO proclaimed the COVID-19 episode a pandemic on 12 March 2020. Accordingly, the public government of India (GOI) started on 22 March a 14 hours intentional public time limitation. This was trailed by a total lockdown of 21 days in significant urban communities and areas where COVID-19 cases were identified. On 14 April, the lockdown was stretched out for stage II until 3 May 2020 and was additionally reached out from 4 to 17 May as stage III of the lockdown. The limitations on friendly separating and self-isolate measures authorized during the three-stage lockdown caused diminished emanations from mechanical sources, vehicle exhaust, and coal-terminated force plants all through the country.

The lockdown significantly impacted social and economic activities but it has temporarily improved the air quality in most of the polluted cities in the country. Due to the mandatory lockdown implementation in all major cities in the world, the $PM_{2.5}$ level is drastically decreasing compared to the scenario before lockdown in the same places. Lockdown restricts the people movement and industry shutdown which results in a 25–60% reduction in $PM_{2.5}$ level. In maximum studies, $PM_{2.5}$, PM_{10}, NO_2, and CO have been taken for calculating the Air Quality Index (AQI). From all the studies, it is clear that air quality is improved after the implementation of lockdown (during the COVID pandemic) which directly improves the health conditions of human beings. Being the data science students, we decided to analyze the Indian air quality data before, during and after COVID-19 to find something meaningful. The main contribution of this analysis is as follows:

1. The eighteen most polluted cities have been chosen for this study. This selection is totally based on the availability of data for the month of April in the years 2019, 2020 and 2021.
2. The air pollutant data (i.e., particle pollution, ground level ozone, nitrogen dioxide, carbon monoxide, and sulphur dioxide) before, during and after lockdown has been taken from several sources for selected 18 cities.
3. The AQI for different pollutants are compared across three mentioned years.
4. The levels of health concern in the 18 cities which are considered for the analysis is displayed.
5. Principal Component Analysis was performed to understand and visualize the similarity among the selected cities.
6. Correlation Analysis was done to check about the relation between all the air pollutants.
7. This study presents the air quality scenario around the country for three conditions: before, during and after lockdown.

4 Air Quality Index(AQI)

The government agencies use an Air Quality Index (AQI) to give information about air pollution to the public. The higher value of AQI indicates the higher air pollution level and higher health alarming condition.The air quality basically depends on the concentration of pollutants and atmospheric conditions. AQI equation is used to convert air pollutant concentration into Air Quality Index. The mathematical expression of AQI is as follows:

$$A_{QI} = \frac{I_{max} - I_{min}}{C_{max} - C_{min}}(P_C - C_{min}) - I_{min}$$

Here,
A_{QI} denotes Air Quality Index for pollutants.
P_c denotes concentration of pollutants.
C_{max} denotes concentration break point that is greater than or equal to P_c.
C_{min} denotes concentration break point that is less than or equal to P_c.
I_{max} denotes index break point corresponding to C_{max}.
I_{min} denotes index break point corresponding to C_{min}.

5 Data Obtaining

Particle pollution (known as particulate matter comprising $PM_{2.5}$ and PM_{10}), ground-level ozone (O_3), nitrogen dioxide (NO_2), carbon monoxide (CO), and sulfur dioxide (SO_2) are the five air pollutant concentrations used to construct the Air Quality Index (AQI) (SO_2). For this study, the AQI values of $PM_{2.5}$,PM_{10} ,O_3, NO_2 , CO, and SO_2 air pollutants are considered. 18 major cities in India based on the availability of air quality index data have been selected for the study.

A countrywide lockdown was imposed in India in several phases due to the COVID-19 pandemic. For this study, we considered the month of April in each year which had its all 30 days under lockdown in the year 2020. The years 2019, 2020, and 2021 are considered for the comparative study. This will give us an idea of AQI in these 18 cities of India before COVID-19 (April 2019), during COVID-19 lockdown (April 2020), and after COVID-19 (April 2021).

The relevant data was obtained from the World Air Quality Index portal (available at www.aqicn.org). The data for all three years was available on the portal in the CSV format. Though it seemed simple to just download the data it was not in a structured format. The original data for each year contains the data for around 380 major cities around the globe with all kinds of all pollutant species in a single column and all the corresponding dates of the second quadrant of a year(April, May, June) in a single column. This representation can be understood using a simple analogy of unrolling a three-dimensional tensor into a matrix. The three dimensions here are 1. Time (dates), 2. Locations (Cities), 3. Features (pollutant species). The raw data obtained was all scrambled and unrolled. This has to be filtered and arranged in a proper way. The challenge was to transform the available data rather than collection. The following is the format of the raw data.

Date	**Country**	**City**	**Species**	**Count**	**Min**	**Max**	**Median**	**Variance**
2019-04-10	MO	Macau	O_3	117	1.4	42.9	15.7	1129.1
2019-04-14	MO	Macau	O_3	114	1.5	39.4	10.1	734.48
2019-04-15	MO	Macau	O_3	117	1.7	24.1	9.8	367.55
2019-04-22	MO	Macau	O_3	118	10.3	70.9	36	1859.4

We experimented with both median and max AQI values for each pollutant as a max value of a pollutant indicates the extent to which the pollutant was able to reach in a given location. The following is the format of transformed data where the values represent the median AQI values corresponding to each specie.

City	**Date**	$C0$	NO_2	O_3	PM_{10}	$PM_{2.5}$	NO_2
Hyderabad	2019-04-01	4.2	11.4	14.6	82	93	3
Hyderabad	2019-04-02	5.5	14.3	14.9	82	125	2.4
Hyderabad	2019-04-03	5.9	14	13	82	24.1	2.6
Hyderabad	2019-04-04	5.2	12.9	12	72	106	2

6 Data Preparation and Cleaning

After transforming the data into the required format the problem now at hand is the missing data. Initially, the data contained 23 Indian cities but the data available across all three years restricted us to use only 18 of them as the data for these cities was not collected in one or the other years.

Once the problem on available data for respective cities is solved the next challenge is that the data was not available for all the days of April. These missing values are filled in using interpolation of available data. This method may not be robust enough but in the case of time-series data, a simple moving average interpolation imputation was used which works better compared to mean/median imputation. This cleaned data is used for further analysis and visualization.

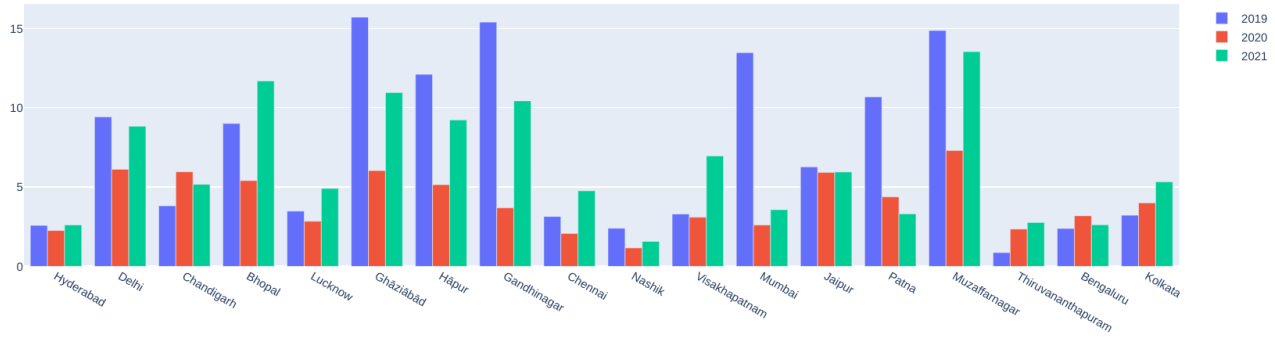
7 Data Set

The data set is obtained from Air Quality Open Data Platform Worldwide COVID-19 data set (https://aqicn.org/data-platform/covid19/verify/1964ab4d-36e8-4a0f-ac83-522b17ff47f2). The final data set contains the following features :

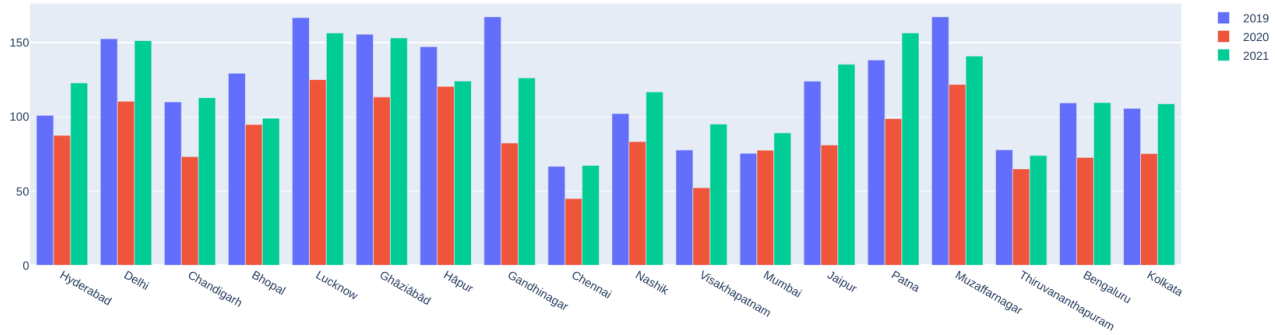
S.No.	Attribute	Description
1.	city	Name of the city
2.	Date	Date of Sample Collection
3.	co	Carbon monoxide(AQI)
4.	no2	Nitrogen dioxide(AQI)
5.	o3	Ozone(AQI)
6.	pm10	Particles with aerodynamic diameter smaller than 10 μm (AQI)
9.	pm10	Particles with aerodynamic diameter smaller than 2.5 μm (AQI)
10.	so2	Sulphur dioxide(AQI)

8 Data Visualization

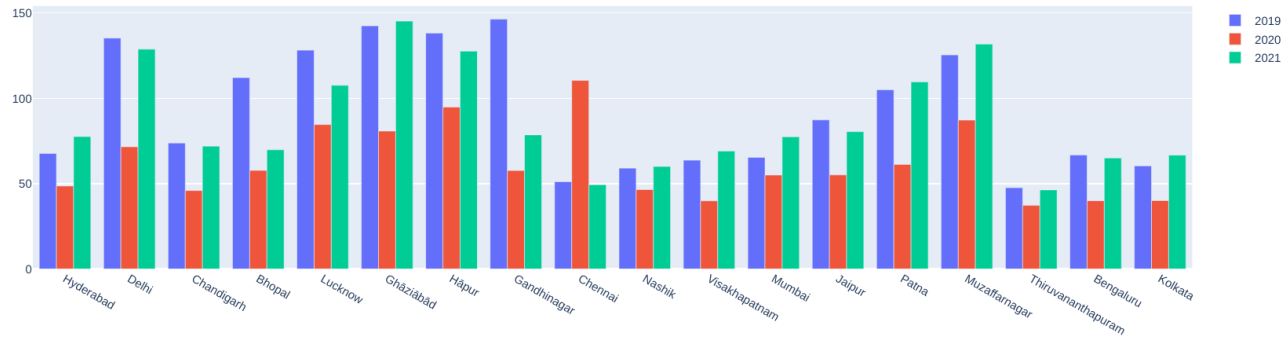
- Average SO_2 AQI in April 2019, April 2020 and April 2021.



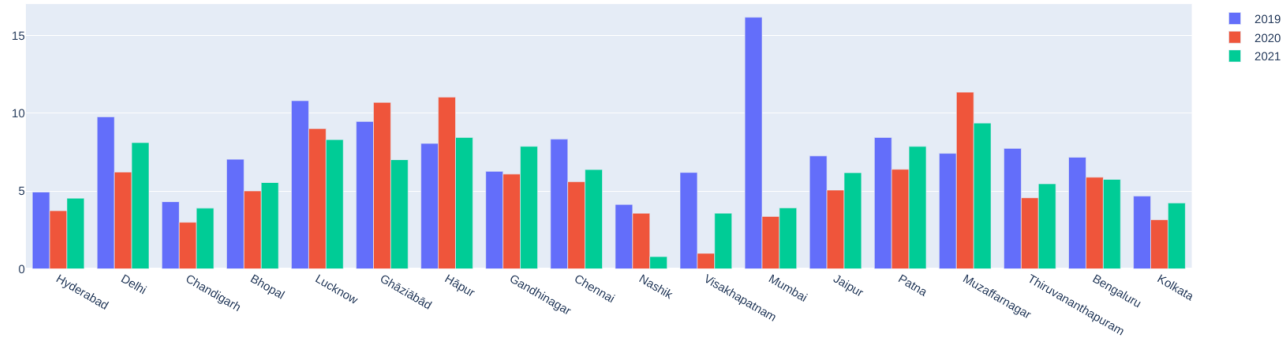
- Average $PM_{2.5}$ AQI in April 2019, April 2020 and April 2021.



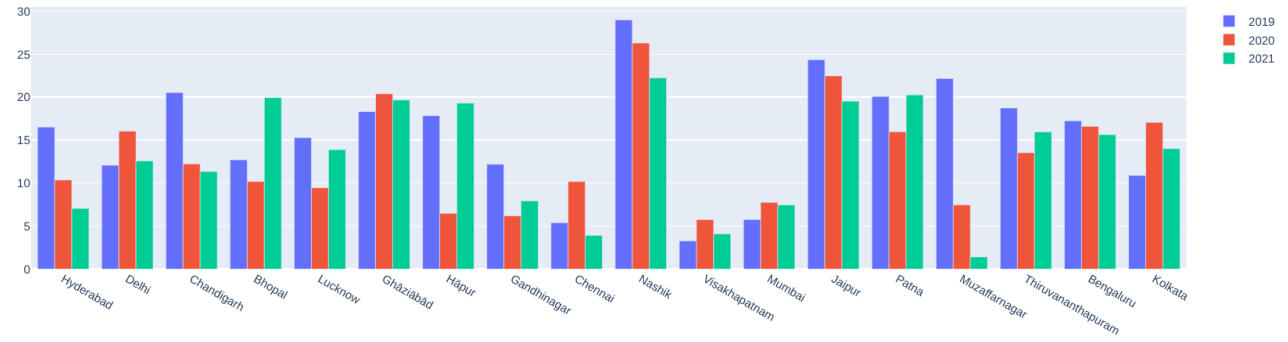
- Average PM_{10} AQI in April 2019, April 2020 and April 2021.



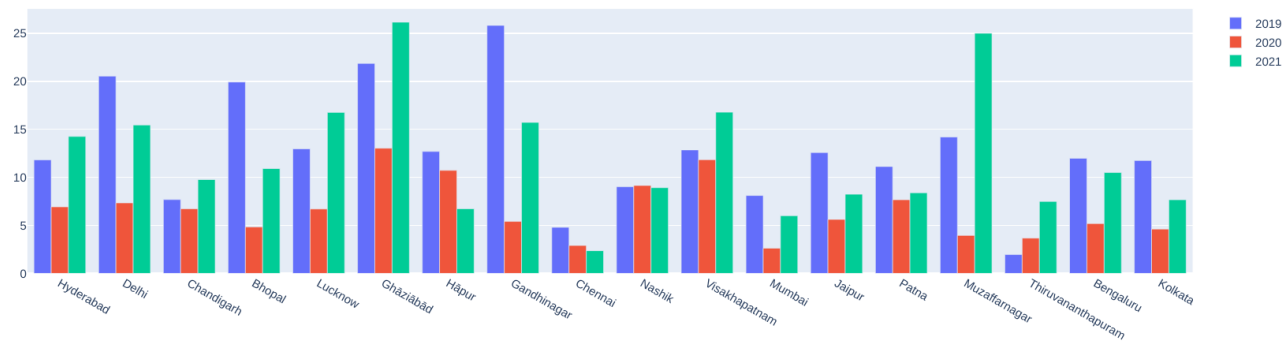
- Average CO AQI in April 2019, April 2020 and April 2021.



- Average O_3 AQI in April 2019, April 2020 and April 2021.



- Average NO_2 AQI in April 2019, April 2020 and April 2021.

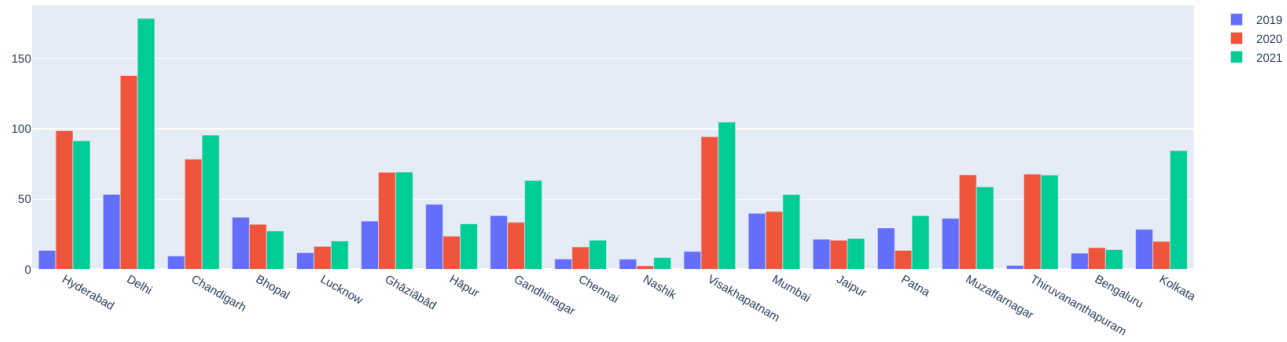


As we can see in the above graphs, there is a decline in the AQI values in all the pollutants in almost all cities in the year 2020. Low the AQI value better is the quality of air. This shows how the inactivity during lockdown has improved the quality of air in all most all cities of India.

This can also be compared to to the year 2021 where the activity resumed and so did the pollution. This shows a significant improvement in the air quality when the country was under lockdown.

- SO_2 Anomaly

The below graph is based on the average maximum AQI value per day of SO_2 in the month of April for three years. We experimented on average and max values and the pattern was similar for both in case of all the pollutants except SO_2. This is an anomalous behaviour especially shown by SO_2 which was also found by other studies.

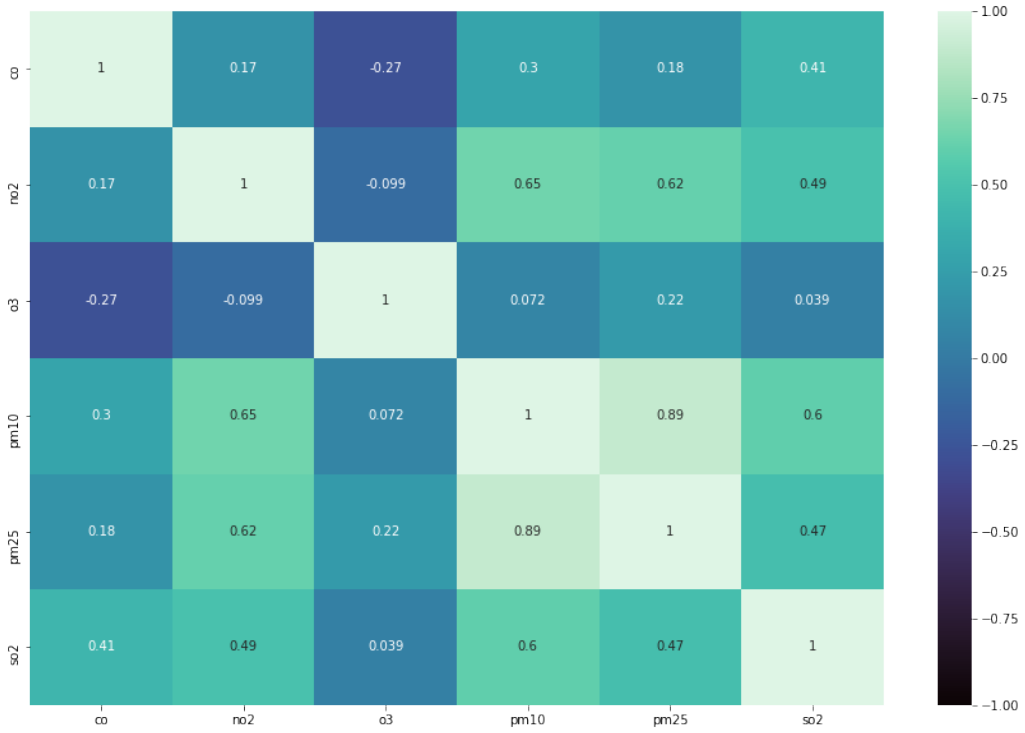


9 Correlation Analysis

Correlation Coefficient is used to find how two or more variables are related to one another, also referred to as linear dependence. Correlation coefficient(say r) always lies between -1 to +1. If the value of r = 1, it means the two data set are positive correlated(means if one data is increasing then other is also increasing), if r = -1, it means data is negatively correlated(means one data is increasing the other is decreasing) and if r = 0, it means no correlation(means one is independent from other). So, we use this method to find the correlation between all the air pollutants.

- Correlation Matrix and heat map for 2019.

From below heat map we can see that, the correlation coefficient of PM_{10} and $PM_{2.5}$ is 0.89 which is nearby 1. So, we can conclude that these two pollutants are most correlated. And the most negative correlation coefficient is of O_3 and CO which is equal to -0.27. Correlation Coefficient of O_3 and $S0_2$ is equal to 0.039 which is near by 0 , so these two pollutants are hardly dependent to each other.

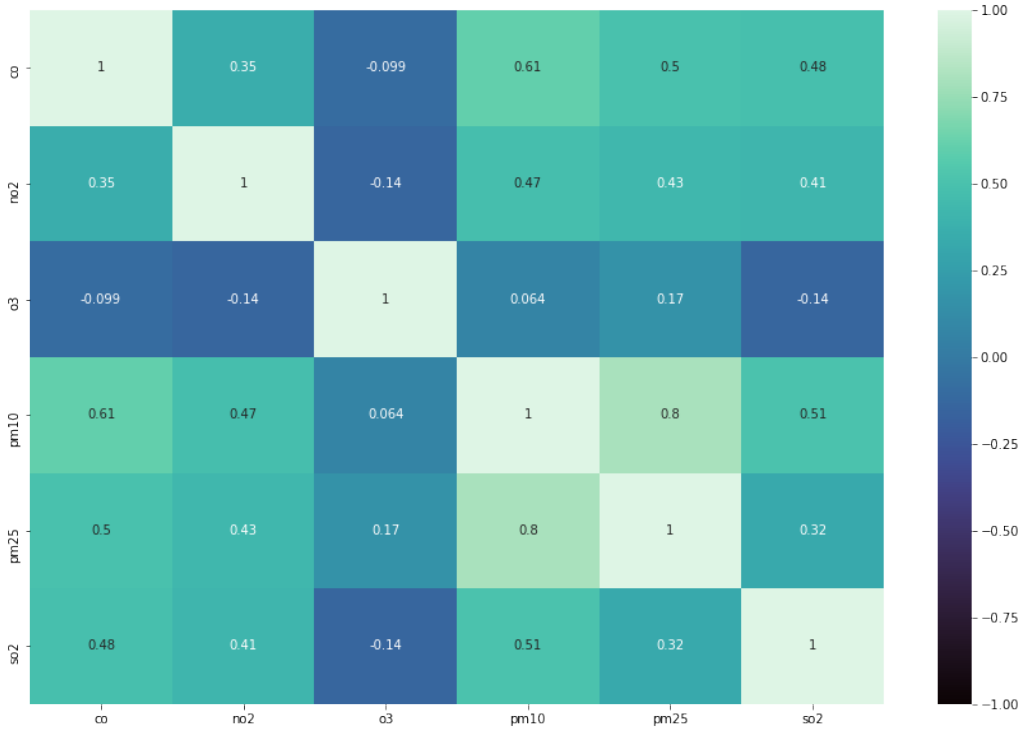


- Correlation Matrix and heat map for 2020.

  During lockdown in April 2020, we can see that the highest correlation coefficient is between CO and $PM_{2.5}$ which is equal to 0.66. And the most negative correlation coefficient is of O_3 and PM_{10} which is equal to -0.16. Correlation Coefficient of O_3 and $PM_{2.5}$ is equal to -0.0088 which is near by 0, so these two pollutants are hardly dependent to each other.

- Correlation Matrix and heat map for 2021.

  In April 2021, we observe that the highest correlation coefficient is between PM_{10} and $PM_{2.5}$ which is equal to 0.8 . And the most negative correlation coefficient is between O_3 and SO_2 which is equal to -0.14. Correlation Coefficient of O_3 and PM_{10} is equal to 0.064 which is near by 0, so these two pollutants are less effect each other.



  Finally, comparing the correlation matrices in three years, we see that the pattern of 2019 and 2021 are a bit more similar compared to 2020 which is the effect of lockdown.

10 AQI and Health

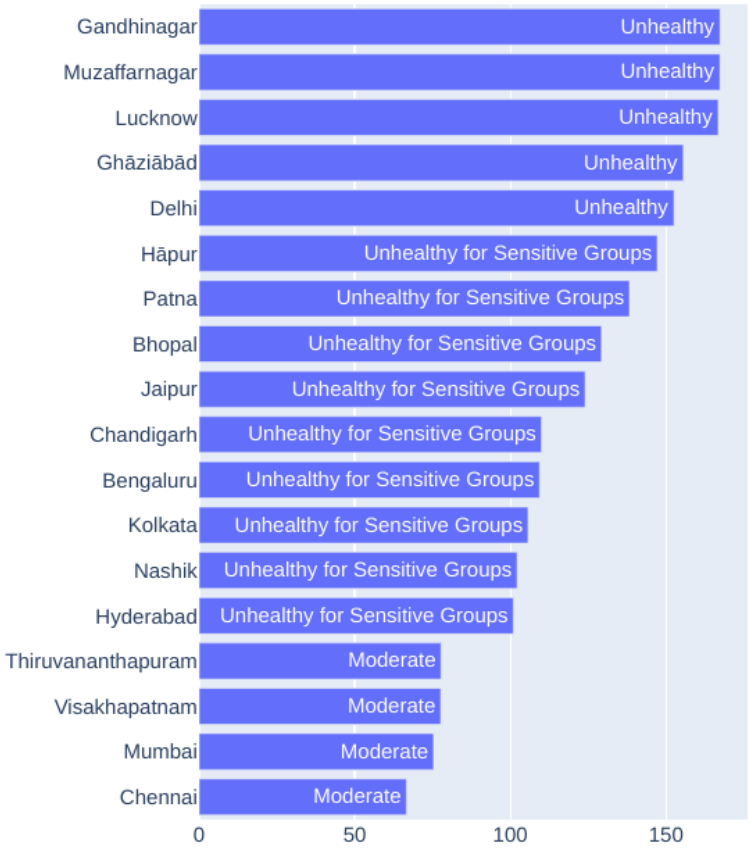
Here cities are labelled as the good, moderate, unhealthy for sensitive groups, unhealthy, very unhealthy , hazardous based on their AQI values. The below attached table mentions the criteria for the classification of the cities.

Quality of Air is becoming a serious concern for every section of the society day by day. Based on the AQI values for 2019,2020,2021 we have labelled cities like moderate, unhealthy, hazardous etc. This labelling indicates the level of the city's health concern.

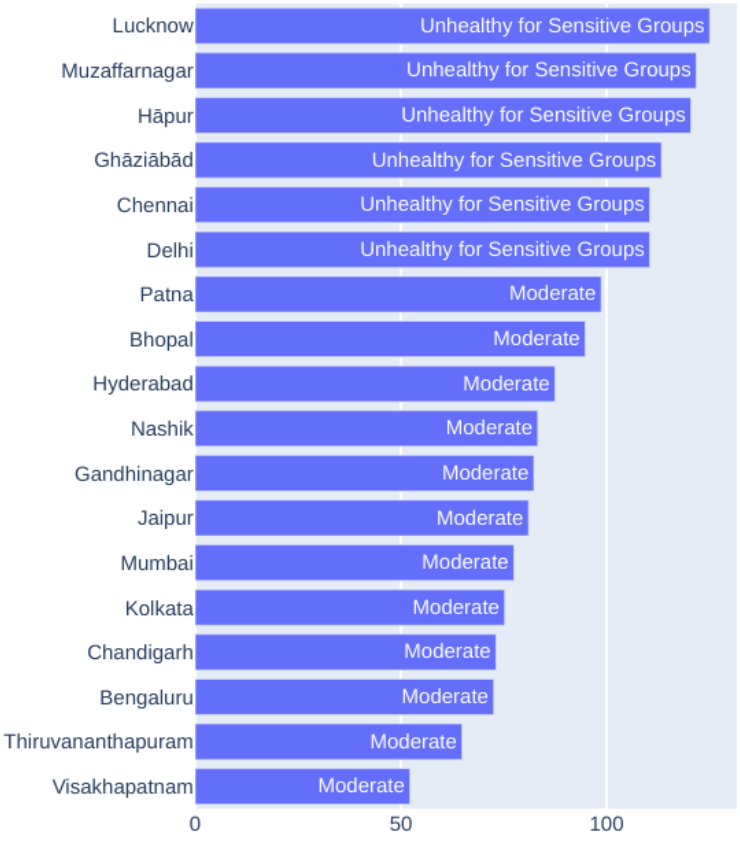
Due to COVID-19 lockdown imposed in 2020, AQI of the many cities across the world has reduced. Even in India due to strict lockdown imposed throughout the April month AQI values have been reduced to some extent. Because of this some cities in the 'Unhealthy for senstive groups' class have came to 'Moderate' class and also some cities in 'Unhealthy' class have came to 'Unhealthy for sensitive groups'. Which is a very good sign for the quality of air in the city. Even though the air quality index has been increasing from 2020 to 2021 it is little lesser than what it is has been before the lockdown.

Air Pollution Level	AQI
Good	0 - 50
Moderate	51 to 100
Unhealthy for Sensitive Groups	100 to 151
Unhealthy	151 to 200
Very Unhealthy	201 to 300
Hazardous	301 to 500

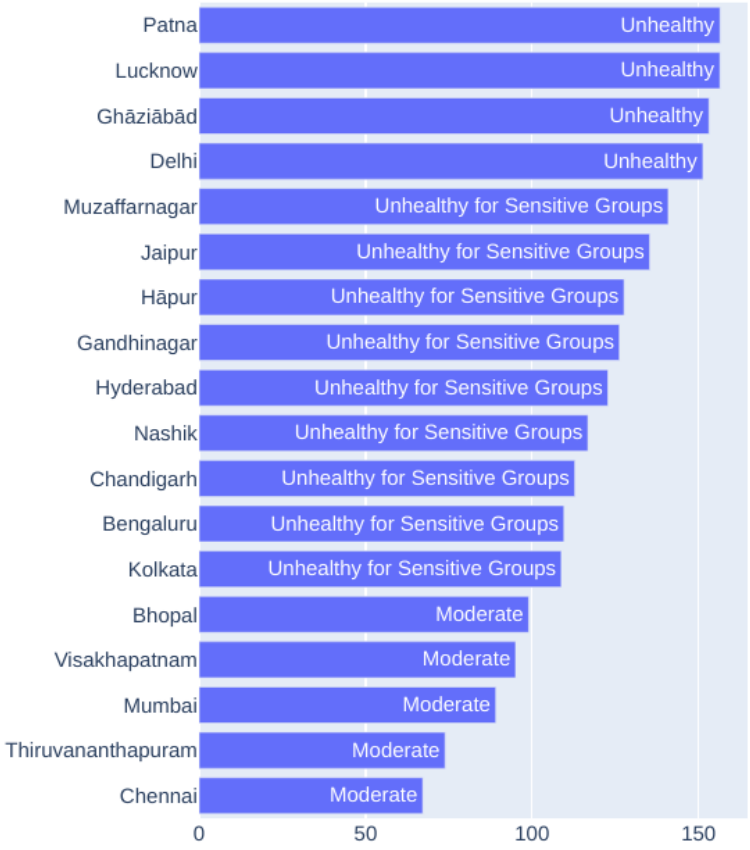
- Figure indicating health concern of cities based on AQI values in April 2019.



- Figure indicating health concern of cities based on AQI values in April 2020.



- Figure indicating health concern of cities based on AQI values in April 2021.

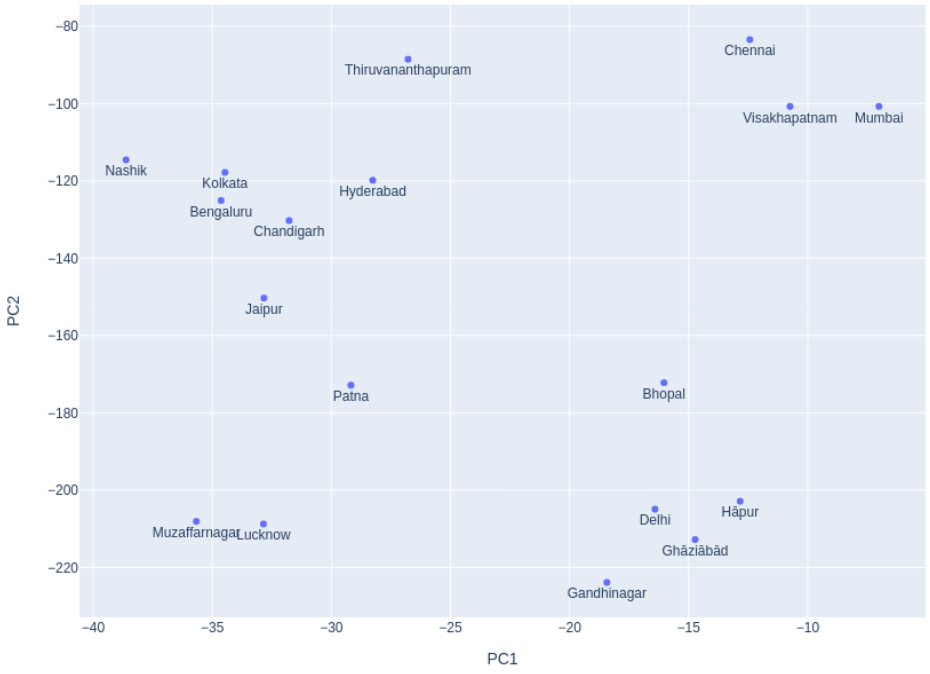


11 Data Analysis using Principal Component Analysis

Principal Component Analysis, or PCA, is a dimensionality-reduction approach for reducing the dimensionality of large data sets by transforming a large collection of variables into a smaller one that retains the majority of the information in the large set. Since our ability to visualize is limited to only three dimensions, in order to visualize the spread of different cities based on the pollutants, we used PCA to first reduce the 6 pollutants(dimensions) to two dimensions. That is, we take only the first two principal components and transform the 6D data to 2D data. The first two principal components comprise more than 90 percent of the variation in all three years and thus can be considered to be accurate.

This 2D data is then used in a scatter plot to visualize the spread of cities. We need to note that the scatter plot only represents the spread and clusters of cities and the axes are not interpretable. So comparing the axes of two different years is not valid, only the closeness of cities and grouping the cities is valid. Essentially, the scatter plots help us understand the closeness of cities based on their respective air quality.

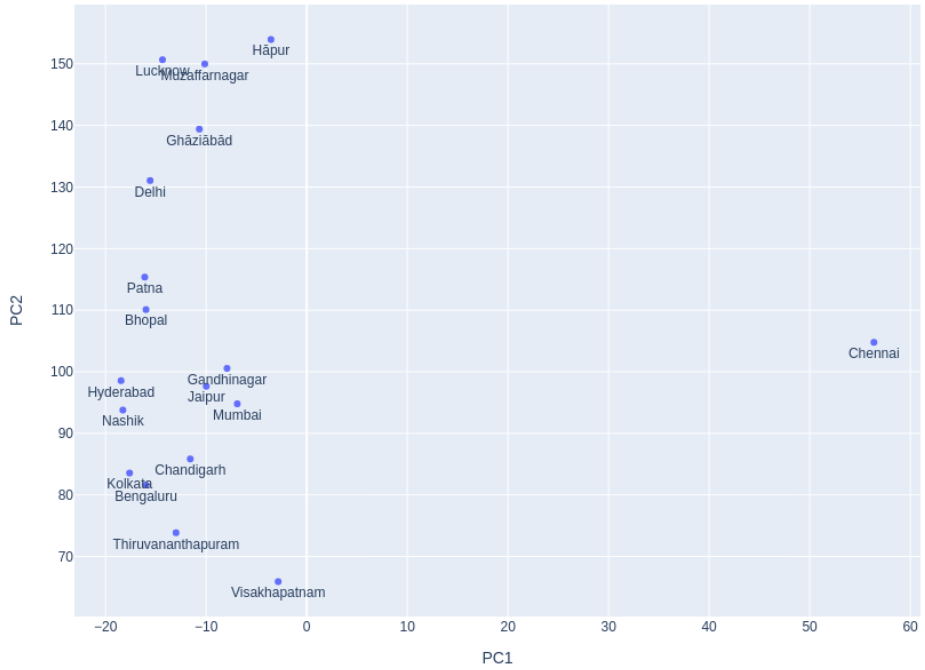
- Scatter plot showing the spread of cities using first two principal components for 2019.



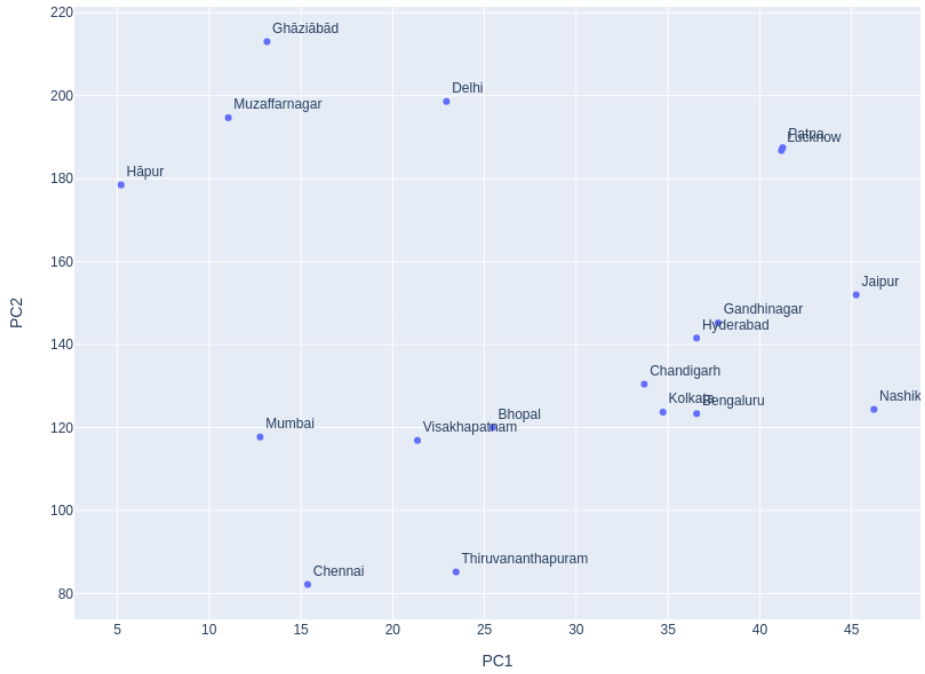
We can observe the clusters of cities based on the extent of pollution. The cities that are close show similar patterns of pollutant concentrations and thus similar air quality.

In the year 2020 we can clearly observe the shift in the pattern compared to the year 2019. This is the effect of lockdown. Chennai is an anomaly in the year 2020, this is due to the non-availability of proper data for Chennai in 2020.

- Scatter plot showing the spread of cities using first two principal components for 2020.



- Scatter plot showing the spread of cities using first two principal components for 2021.



12 Results And Conclusion

Thus by working on this project and performing various tasks we got a hands-on experience of CRSIP-DM in a small scale real world application. Here, in this mini project we got exposure to various data science activities like formulating and understanding problem statement, data collection, data cleaning, data modelling, data analysis etc.

From this study, we can conclude that there is a significant difference in air quality when there is a lockdown and when there is no lockdown. The restrictions due to lockdown like restrictions on transportation and shutdown of institutions and industries have caused the inactivity of major air pollution contributors such as vehicles and factories. This study indicates that we are causing major damage to the environment and the quality of air has again come down in 2021 as we can see. This temporary improvement in air quality is not useful in the long term. We should thus try to prevent further damage by using sustainable energy resources rather than fossil fuels which are the major contributors to degrading the air quality.

13 Contributions

- M20MT013 Yash Prakash Mishra - Problem Statement formulation
- B18EE008 Ashish Ledalla - Data obtaining and Data preparation
- B18CSE049 Saptarshi Sardar - Data visualization
- M19DCS003 Harshita - Correlation Analysis
- B18CSE005 Annem Khyathi Reddy - AQI and Health, Data Analysis using principal component analysis

14 Tools and Methods Used

- python
- pandas/numpy
- Time series interpolation
- matplotlib/Plotly
- PCA

Bibliografia

[1] https://link.springer.com/article/10.1007/s11356-021-14462-9

[2] https://aqicn.org/data-platform/covid19/verify/1964ab4d-36e8-4a0f-ac83-522b17ff47f2

[3] https://airqualitynews.com/2020/09/23/sulphur-dioxide-pollution-doubled-during-lockdown/