Positive Aspect of COVID-19

Abstract— **The novel COVID-19 virus has left the world reeling with thousands of deaths and lakhs of cases. The entire human life has come to a standstill with the majority of the countries enforcing shutdown/lockdown and travel bans. Amidst this chaos, there is an unlikely beneficiary in the form of mother nature. There is a drastic reduction in air pollution all over the world which has been a major concern over the years. This positive outcome has inspired us to research more on the datasets available and predict the impact of improved air quality on the overall ecosystem, such as by how far the melting of polar ice will be pushed which was originally predicted at 2100. The positive aspect of this pandemic could be an eye-opener to reanalyze the parameters which play a vital role in a healthier environment. This might encourage people to keep the air immaculate once this lockdown is over.**

I. Introduction

A. Motivation

Air pollution poses one of the biggest threats to human health. 90% of the global population breathes unsafe air. While coronavirus might be dominating the world news as the hideous killer, a silent killer is contributing to nearly 7 million more deaths a year: Air Pollution[1]. If we consider first world countries like USA which have a greater will and infrastructural capability to curb these deaths, we find that in 2018 alone eroding air quality was linked to nearly 10,000 additional deaths relative to the 2016 benchmark according to researchers at Carnegie Mellon University[2].

With the onslaught of this worst pandemic causing a global health crisis, COVID-19 has brought the earth to a standstill.In U.S. more than three-quarters of annual greenhouse gases are produced by transportation, industry, and power generation as reported by the wall street journal[3]. All of the three sectors mentioned above have been worst hit and virtually shut down because of the social distancing measures taken by governments around the world to stop the COVID-19 spread.

B. Objective

We as part of our research have analyzed the data of Air travel, Traffic on roads, and Industrial production along with the concentration of air pollutants in various cities that have been impacted severely by the virus impact. We have tried to visualize and analyze data to bring out the drastic contrast which we are witnessing worldwide in all walks of life, through our application of prediction models and graphical representation of data.

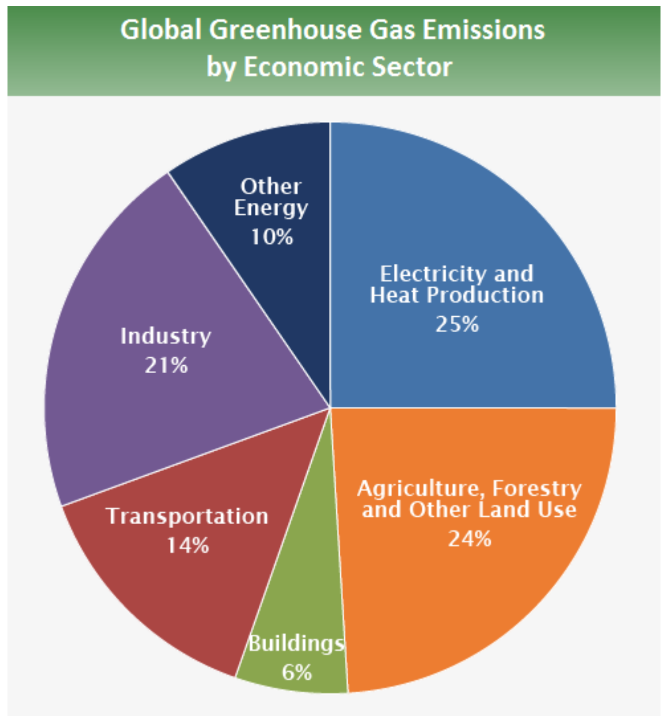

Fig. 1. Global emissions by each sector

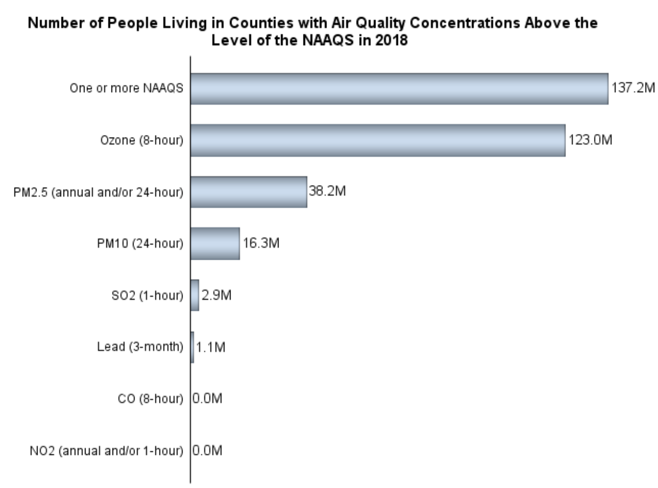

Fig. 2. No. of people living in US counties with high pollution

The effects of lockdown are different for individual cities. It is evident from air pollution data (PM2.5) that cities which are richer, more industrialized, and colder experience larger reductions in Air Pollution as compared to cities with low manufacturing output. For example Delhi and Wuhan. These are also the ones who have maximum restrictions and draconian laws in place to stop the spread of coronavirus. Predicting and depicting these kinds of findings with the data is our objective here. We have also predicted pollution levels for these cities for the near future assuming that the lockdown procedures are followed for the next few months.

II. System Design and Implementation

A. Algorithms considered

Statsmodel library of python was used to perform time series prediction. Statsmodel have numerous models, 3 of which we used are described below.

1) Auto-regression: An auto-regressive (AR) model predicts future behavior based on past behavior. The process is basically a linear regression of the data in the current series against one or more past values in the same series. AR model usually gets “close enough” for it to be useful in most scenarios. Figure 5 shows the prediction made for ’PM 2.5’ for year 2020 for city New York with SARIMA model.

AR Model

```
from statsmodels.tsa.ar_model import AR

def predict(data,value):
    ts = data[' pm25'].dropna()
    model = AR(ts.astype(float))
    model_fit = model.fit(trend='nc')
    yhat = model_fit.predict(len(ts), len(ts)+value)
    result[' pm25'] = yhat

result = pd.DataFrame()
predict(train,len(val)-1)
print('MSE',mean_squared_error(val[' pm25'], result[' pm25']))
print('RMSE',sqrt(mean_squared_error(val[' pm25'], result[' pm25'])))
```

```
MSE 91.19712838402361
RMSE 9.54971875942028
```

Fig. 3. AR Model Python

2) Moving Average: A moving average is a technique to get an overall idea of the trends in a data set; it is an average of any subset of numbers. Figure 4 shows the prediction made for ’PM 2.5’ for year 2020 for city New York with Moving Average model.

MA Model

```
result = pd.DataFrame()
from statsmodels.tsa.arima_model import ARMA
def predict(data,value):
    ts = data[' pm25'].dropna()
    model = ARMA(ts.astype(float), order=(0, 1))
    model_fit = model.fit(disp=False)
    yhat = model_fit.predict(len(ts), len(ts)+value)
    result[' pm25'] = yhat

result = pd.DataFrame()
predict(train,len(val)-1)
print('MSE',mean_squared_error(val[' pm25'], result[' pm25']))
print('RMSE',sqrt(mean_squared_error(val[' pm25'], result[' pm25'])))
```

```
MSE 77.69495058447953
RMSE 8.814473925565808
```

Fig. 4. ’MA Model Python

3) Seasonal Autoregressive Integrated Moving Average: Seasonal Autoregressive Integrated Moving Average, SARIMA or Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series as well as an additional parameter for the period of the seasonality. Figure 5 shows the prediction made for ’PM 2.5’ for year 2020 for city New York with SARIMA model.

SARIMA Model

```
result = pd.DataFrame()
from statsmodels.tsa.statespace.sarimax import SARIMAX
def predict(data,value):
    ts = data[' pm25'].dropna()
    model = SARIMAX(ts.astype(float), order=(1, 1, 1), seasonal_order=(1, 1, 1, 1))
    model_fit = model.fit(disp=False)
    yhat = model_fit.predict(len(ts), len(ts)+value)
    result[' pm25'] = yhat

result = pd.DataFrame()
predict(train,len(val)-1)
print('MSE',mean_squared_error(val[' pm25'], result[' pm25']))
print('RMSE',sqrt(mean_squared_error(val[' pm25'], result[' pm25'])))
```

```
MSE 120.75916486396369
RMSE 10.989047495755203
```

Fig. 5. ’SARIMA Model Python

B. Technologies and Tools Used

1) Programming: Python was used for programming. Python provides support to libraries like statsmodel which make data analysis easy. Statsmodel have various time-series models to choose from.

2) Visualization: Seaborn and Matplotlib was used for visualising different graphs. Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

C. Architecture

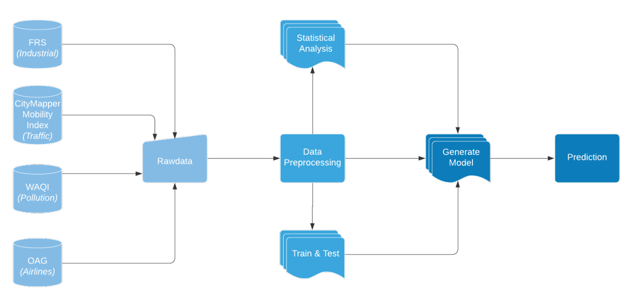

Fig. 6. Architecture of time series prediction

III. Experiments

We have collected the data from heterogenou sources, performed some statistical analysis. Details about each section are given below.

A. Dataset Used

WAQI project provides the unified and world-wide air quality information. All the air quality data is the official data collected from each country respected Environmental Protection Agency (EPA).	
About WAQI	https://aqicn.org/contact/
Data source	https://aqicn.org/data-platform/register/
Type of Data	Air pollutants data of different cities is collected in .csv format.
Size of Data	72KB*7 csv files
No. of instances/Statistics:	Records: 2154*7 cities Features: 6 (PM2.5, PM10, O3, NO2, SO2, CO)

TABLE I

WORLD AIR QUALITY INDEX (WAQI)

CityMapper provides road traffic data integrated from various open sources including local transit authorities.	
Data source	https://citymapper.com/cmi
Type of Data	Road Traffic Data is collected in .csv format
Size of Data	16 KB for each city
No. of instances/Statistics:	2020 Daily data across 41 global cities

TABLE II

CITYMAPPER MOBILITY INDEX

OAG is data supplier that provides the digital flight and travel information over the world.	
About WAQI	https://www.oag.com/about-oag
Data source	https://www.oag.com/coronavirus-airline-schedules-data
Type of Data	Air Traffic Data is collected in .csv format
Size of Data	168 KB
No. of instances/Statistics:	2019-2020 Weekly data across 15 countries

TABLE III

OFFICIAL AIRLINE GUIDE (OAG)

Federal Reserve System is the central banking system of US.	
About WAQI	https://www.oag.com/about-oag
Data source	https://www.oag.com/coronavirus-airline-schedules-data
Type of Data	USA Operational industrial data collected in.csv format
Size of Data	212 KB/file
No. of instances/Statistics:	Monthly data for USA 2018 onwards spanned across 835 sectors

TABLE IV

FEDERAL RESERVE SYSTEM

.

B. Data Preprocessing

Data had lot of blank values which had to be filled by putting in the weighted average values. Also daily data had to be converted to weekly data for predicting values.

C. Data Processing

The final data was not stationary and to successfully implement Time-Series Prediction, it had to be made strict stationary. We used Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Augmented Dickey–Fuller(ADF) test to check for Stationarity of data.

Result of KPSS and ADF can be summed up to 4 cases as follows [4]:

Case 1: Both tests conclude that the series is not stationary then, series is not stationary.

Case 2: Both tests conclude that the series is stationary then, series is stationary.

Case 3: KPSS is stationary and ADF is not stationary then, trend stationary, remove the trend to make series strict stationary.

Case 4: KPSS is not stationary and ADF is stationary then, difference stationary, use differencing to make series stationary.

By KPSS and ADF, data was found to be difference stationary. To make the data strict stationary from difference stationary ”Differencing” and then ”Log Transformation” was used.

Differencing is the method to compute the difference of consecutive terms in the series. Differencing is typically performed to get rid of the varying mean. Whereas Log Transformations are used to stabilize the non-constant variance of a series.

D. Evaluation Methodology

Given data was split into the period before shelter in place started and after it. The data after shelter in place was set aside for comparison with the predicted data. Data from 2014 to shelter in place was used to train the time series model, which resulted in prediction of the content of pollutants if shelter in place would not have happened.

Three models naming Auto-Regression, Moving Average, Seasonal Autoregressive Integrated Moving Average were used for time series prediction.

Results from these models was tested by keeping aside known testing values from January to the date of Lock-down and then later comparing this to the predicted value by the model.

Eventually, Moving Average model came to perform best with repeat to others. RMSE value was used to check the accuracy. Finally, Moving Average algorithm was used to predict unknown values for all other cities using complete known dataset. Figure 7 shows the final predicted graph for the PM2.5 content with Shelter in place and without it for predicted valus of 2020.

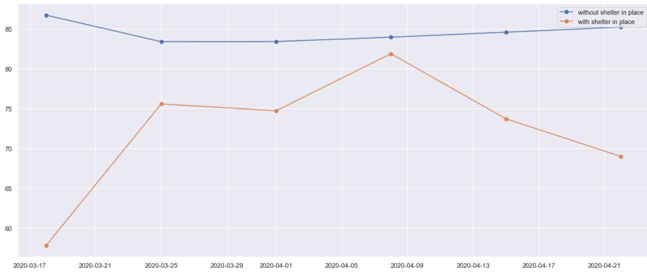

Fig. 7. Year 2020 with and without shelter in place

E. Graphs

We have performed the statistical analysis on the pollution dataset to identify the impact of shelter in place on air pollution in various cities worldwide. Using the available dataset, we identified the cities with major impact of lockdown indeed has shown the improvement in the air quality. The effect of lockdown varied across cities and showed a greater reduction in air pollution levels in regions that are rich in industrialization and heavily congested with traffic. Different series of graphs has been shown below to support our study:

1) Impact of Pollutant: As shown in Figure 8, we analyzed and limited the project scope to PM2.5 pollutant as it has higher concentration in comparison to other pollutants.

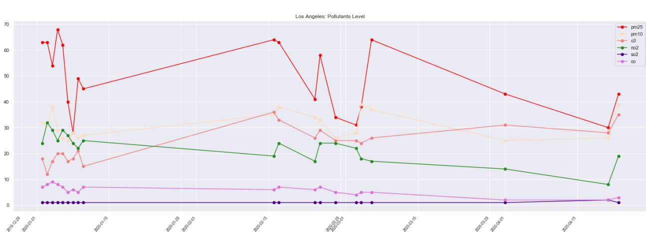

Fig. 8. Los Angeles- Pollutants Level

2) Impact of Lockdown on PM2.5: Figure 9 shows the severe decline that has been observed in year 2020 in comparison to previous years for City: Los Angeles.

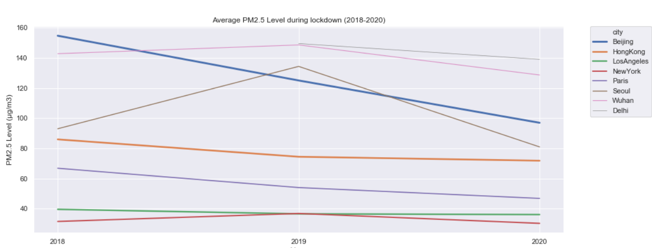

Fig. 9. Average PM2.5 Level during lockdown (2018-2020)

3) Comparison of PM2.5: Significant decrease in PM2.5 Level is observed in major cities in year 2020 post lockdown as shown in Figure 10

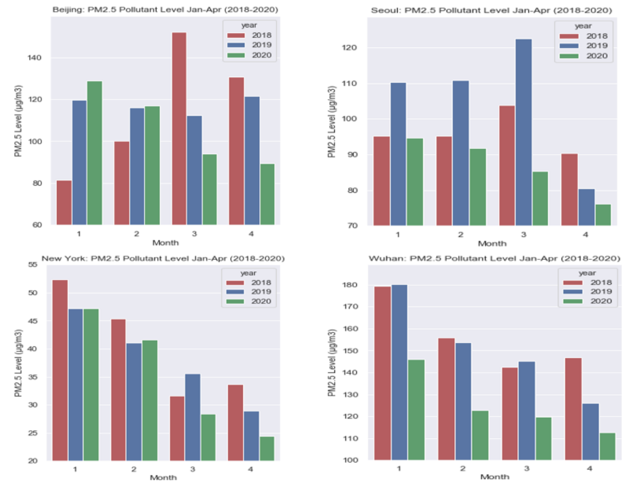

Fig. 10. PM2.5 Pollutant Level Jan-Apr (2018-2020)

4) PM2.5 transition: As shown in Figure 11 Considerable decline in PM2.5 is observed in Year 2020 post lockdown.

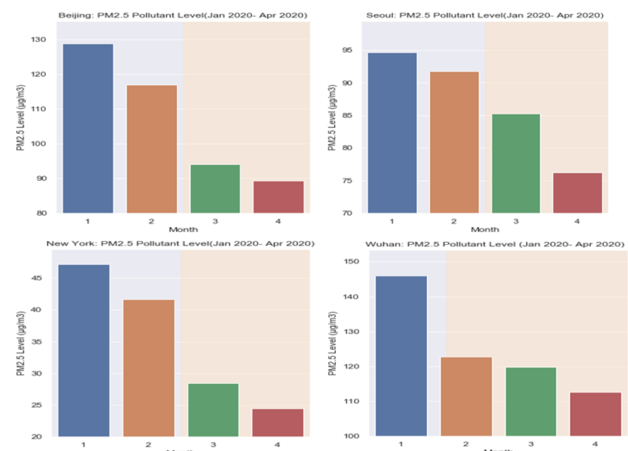

Fig. 11. PM2.5 Pollutant Level(Jan 2020- Apr 2020)

5) Lockdown duration analysis: During lockdown Year 2020 has shown minimum air pollution in comparison to previous two years as shown in Figure 12.

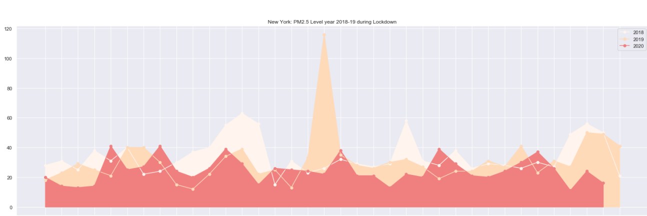

Fig. 12. New York- PM2.5 Level year 2018-19 during Lockdown

6) PM2.5 drill-down during lockdown : Figure 13 shows drill down analysis of PM2.5 level before and after imposed lockdown.

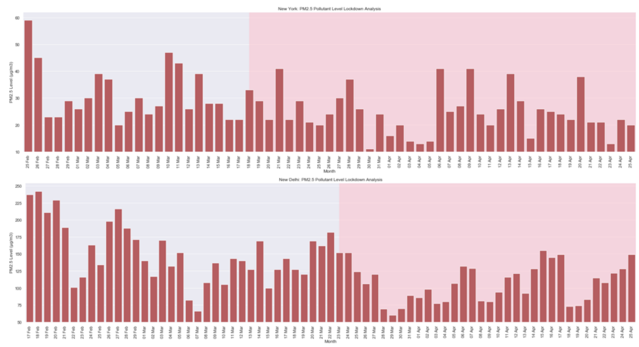

Fig. 13. PM2.5 Pollutant Level Lockdown Analysis

7) Air Traffic: Air Traffic trend has shown significant decrease post lockdown in various countries. Figure 14 shows the same.

Fig. 14. Air Traffic Trend (Year-2020)

8) Road Traffic: City Traffic trend has shown immense decline post lockdown in different cities. This includes data for both public and private transit.

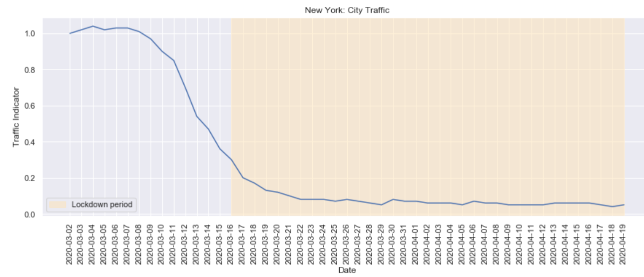

Fig. 15. New York- City Traffic

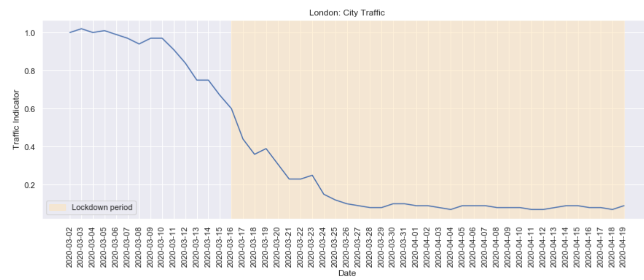

Fig. 16. London- City Traffic

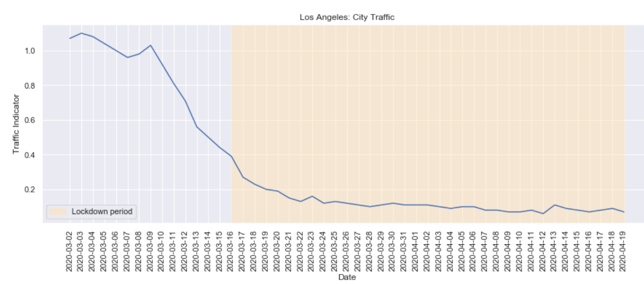

Fig. 17. Los Angeles- City Traffic

9) Industrial operations: Figure 18 shows industrial operations were severely impacted by the lockdown.

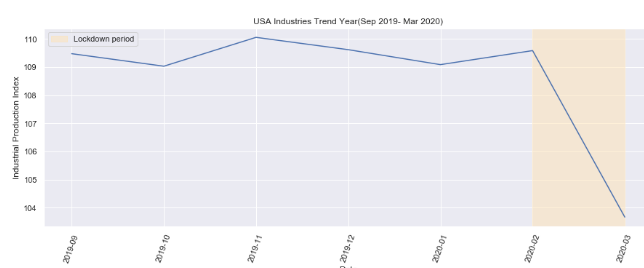

Fig. 18. USA Industries Trend Year(Sep 2019- Mar 2020)

F. Result Analysis

We have performed impact of major pollution contributor over the air pollution data. We were able to corelate that decline in operations of flights, vehicles and industries have positively impacted the environment by reducing the PM2.5 level. Figure 19 demonstrates the same.

Fig. 19. Comparative Analysis of PM2.5 Pollutant with AirTraffic, Industry CityTraffic

IV. Conclusion Discussion

Decisions made and difficulties faced have been described below.

A. Decisions made

Using the lockdown timeline and understanding the measures taken by the government helped us understand the nature of data and helped us to better plot it on a graph, helping us in decisions like averaging the pollution data per week because there are sudden spikes recorded in sensors as people tend to react collectively to the news and announcements by government thereby increasing road traffic and hence increasing the pollution level making it difficult to predict a trend.

B. Difficulties

The dataset had some outliers which needed to be preprocessed and hence had to be discarded. Some models which we expected to show drastic contrast did not show much change. Such as the pollution level change in Seoul and New York. This when we researched found out was because these are flagship cities of first world countries and hence have better environment protection policies in place when compared to other cities of developing countries.

We also faced difficulty in finding datasets of air quality for various countries and cities around the world, this poses a serious problem because what cannot be measured cannot be managed. According to a report of IQAir[1] areas that lack air quality information are often estimated to have some of the world's most severe air pollution, putting huge populations at risk. Africa, a continent of 1.3 billion people, currently has less than 100 monitoring stations that make PM2.5 data available to the public in real-time.

C. Conclusion

Using the timeline of lockdowns and comprehensive dataset of air pollution we found out that COVID-19 despite its heavy economic toll on society has a very positive effect on the global air quality for the time being. According to this year's Air Quality Index cities with historic levels of PM 2.5 have witnessed a dramatic drop in air pollution since enforcing lockdown; 44% in Wuhan, 54% in Seoul, and 60% in New Delhi[5]. According to researchers led by Xiao Wu at Harvard University[6] an increase of just 1 microgram per cubic meter corresponded to a 15% increase in Covid-19 deaths. So we can safely say that this quarantine period has indeed been more life saving than life-threatening.

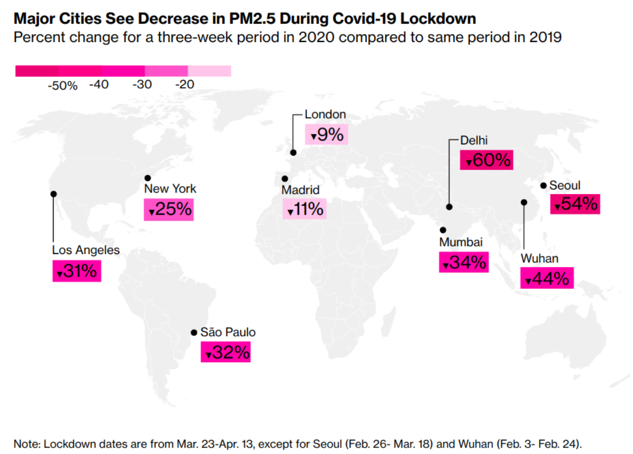

Fig. 20. Changes throughout the world

All true, perhaps. But falling emissions driven by economic distress are rarely sustainable, and easily reversible. Enforced systems change, imposed without public consent, will never last. A decade ago the emissions rebounded sharply after the housing crisis 2008 when the governments rolled out economic stimulus packages that often focused on infrastructure and industrialization[3]. We need a long term approach to control the emissions which is sustainable. We need to achieve the Paris target decided on earth day 2016 i.e. to keep earth's temperature well below 2C. The current prediction by En-ROADS/Climate Interactive says emissions getting back to normal earth's temperature will be +4.1C from the pre-industrial levels by the year 2100[7]. One long term solution might be that some of our remote working habits remain[8].

REFERENCES

[1] IQAir-report. [Online]. Available: https://www.iqair.com/blog/report-over-90-percent-of-global-population-breathes-dangerously-polluted-air

[2] CMU-paper. [Online]. Available: https://www.nber.org/papers/w26381

[3] WSJ-article. [Online]. Available: https://www.wsj.com/articles/coronavirus-lockdowns-clear-the-air-but-the-green-effect-could-be-fleeting-11586095204

[4] Analytic-Vidhya. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/09/non-stationary-time-series-python/

[5] BBC-article. [Online]. Available: https://www.bbc.com/future/article/20200427-how-air-pollution-exacerbates-covid-19

[6] Harvard-paper. [Online]. Available: https://www.medrxiv.org/content/10.1101/2020.04.05.20054502v2

[7] Bloomberg-article. [Online]. Available: https://www.bloomberg.com/graphics/2020-global-warming-simulator/

[8] en-roads. [Online]. Available: https://en-roads.climateinteractive.org/scenario.html?v=2.7.15