Positive Aspect of COVID-19

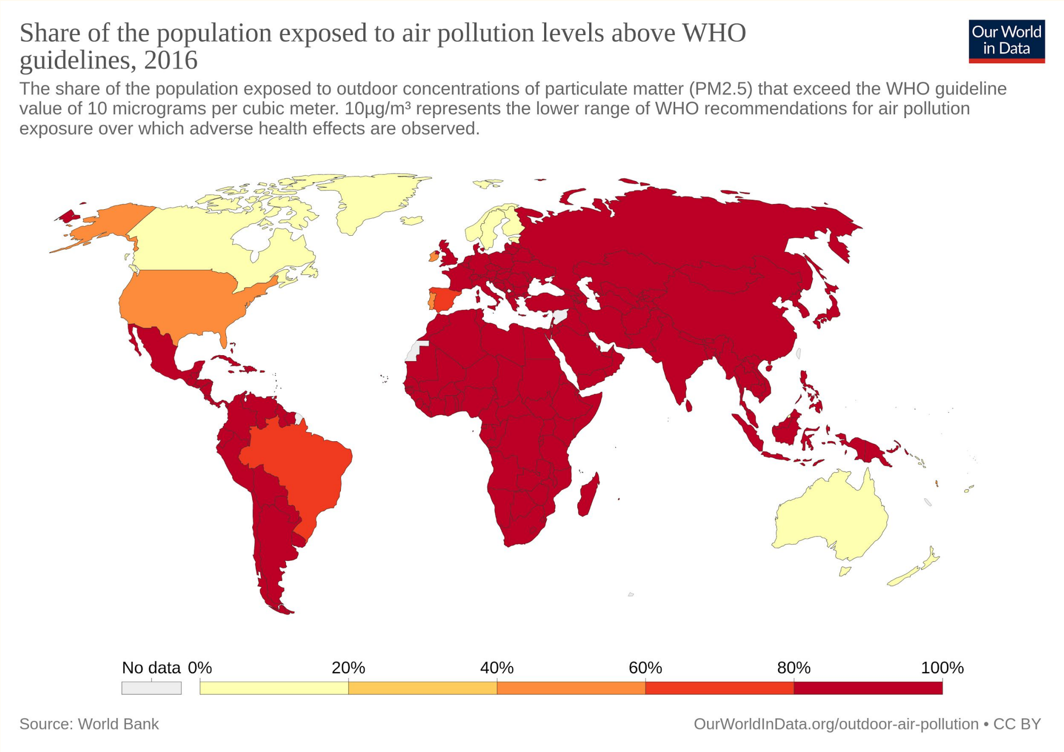
Team: Aether(Group 10)
Team members: Aditya Agarwal
Naman Agrawal
Pooja Kataria

Introduction: Motivation

- 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.
- in 2018 alone eroding air quality was linked to nearly 10,000 additional deaths in the U.S. relative to the 2016 benchmark
- In U.S. More than three-quarters of annual greenhouse gases are produced by transportation, industry, and power generation
- With the onslaught of this worst pandemic causing a global health crisis , COVID-19 has brought the earth to a stand still.

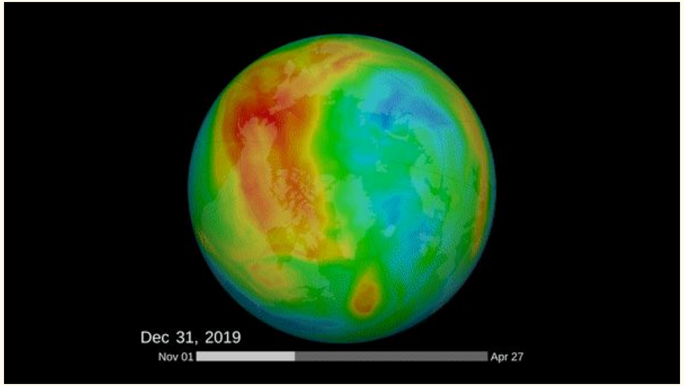
Objective : We as part of our research have studied, analysed and predicted the data of Air travel , Traffic on roads and Industrial production along with the concentration of air pollutants in various cities which have been impacted severely by the disease.Our objective is to capture this unprecedented event vividly.

Problem Statement



Impact of Air pollution

- Deaths
- Ozone depletion
- Low Immunity



Project Scope

Impact of various parameters

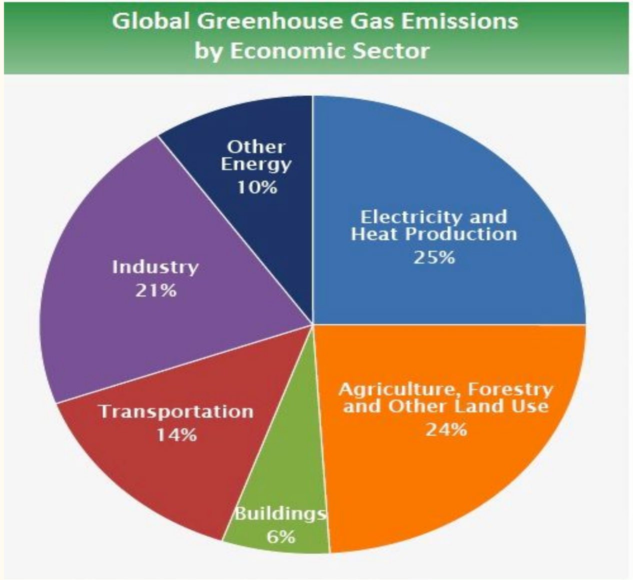


Image Source:

Major contributor to air pollution(PM2.5)

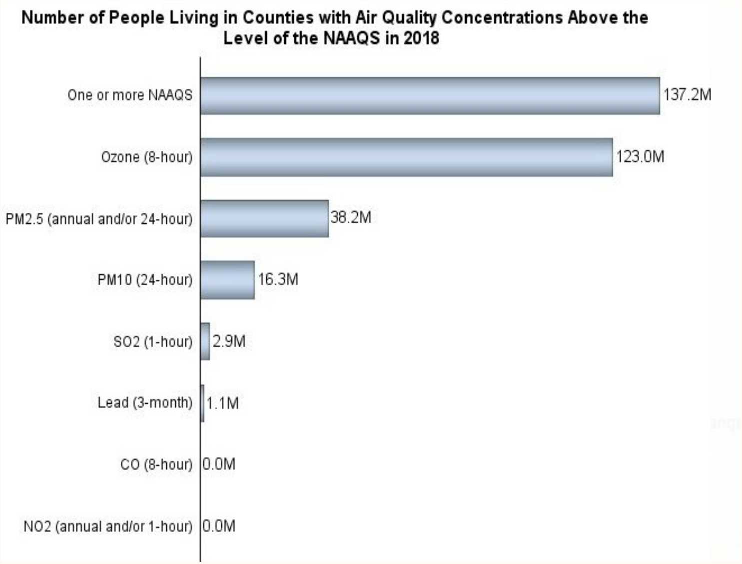


Image Source:

Dataset

WAQI(Major)

- Official EPA air quality dataset collected globally
- **Format :** .csv
- **Size:** Past four year daily data of air pollutants
- **Features:** PM2.5, PM10, O3, NO2, SO2, CO

CityMapper Mobility Index

- Road traffic data collected from transit authorities
- Size: Year 2020 daily data across countries

OAG

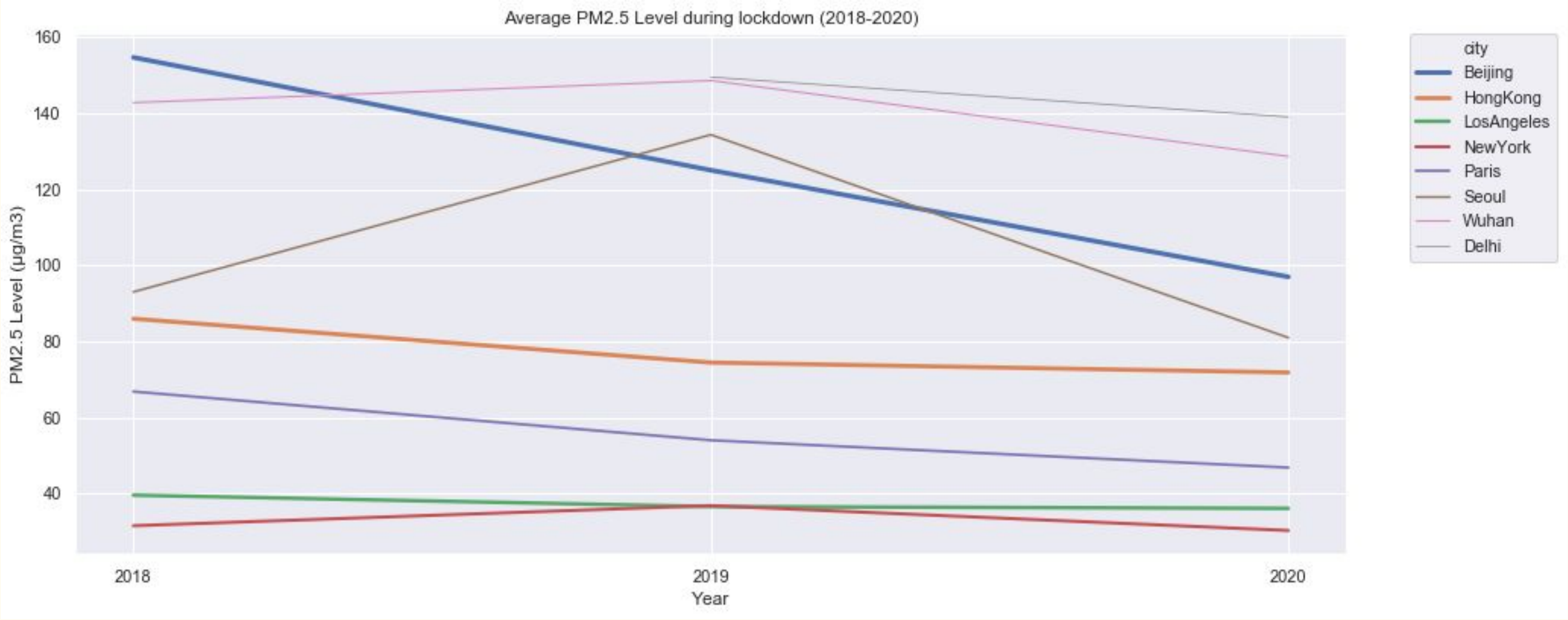
- Global digital flight and travel data
- Size: Last two year data across 15 countries

Federal Reserve System

- Operational Industrial data
- Size: Monthly data for USA 2018 onwards spanned across 835 sectors

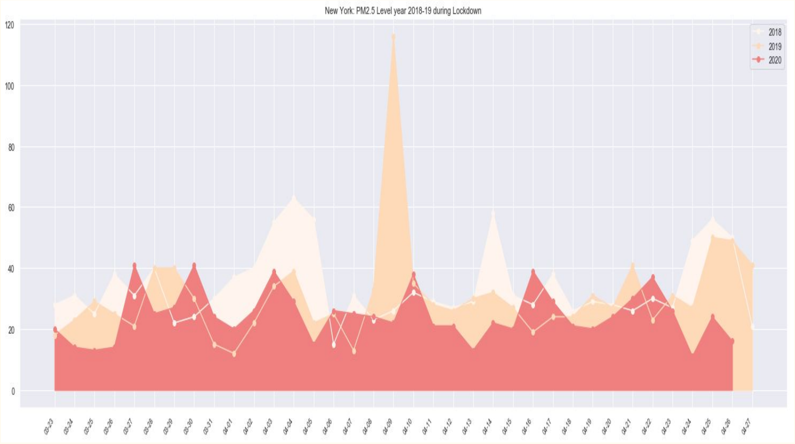
Statistical Analysis

Average PM2.5 Level comparison cities globally [2018-2020]: 🔻 17% (2020)

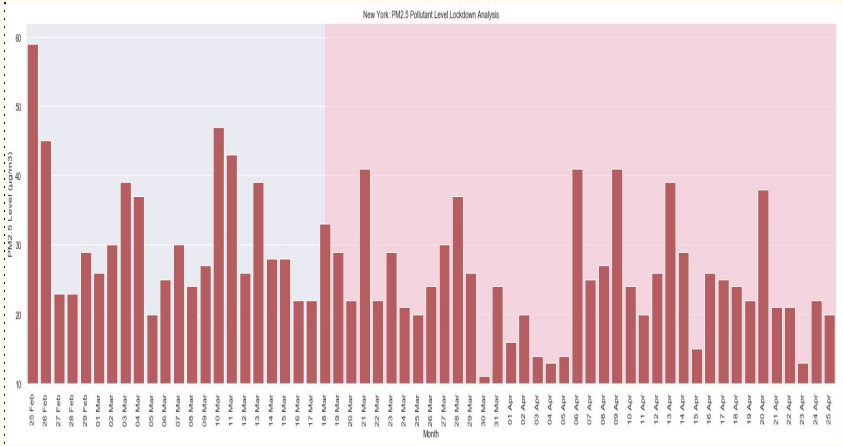


Statistical Analysis(Cont..)

🔻 21% Pollutant level during lockdown [2018-2020]: New York

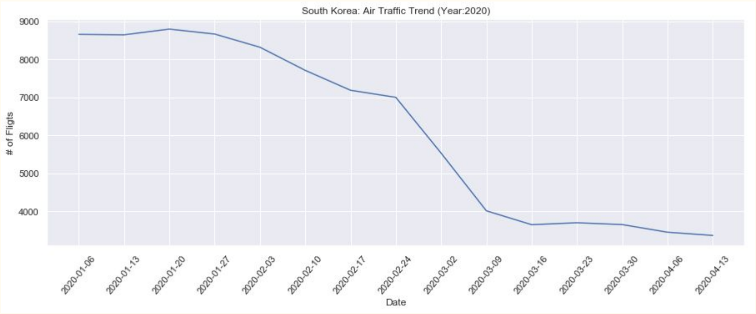


🔻 22%: Impact of lockdown (Highlighted Area):New York

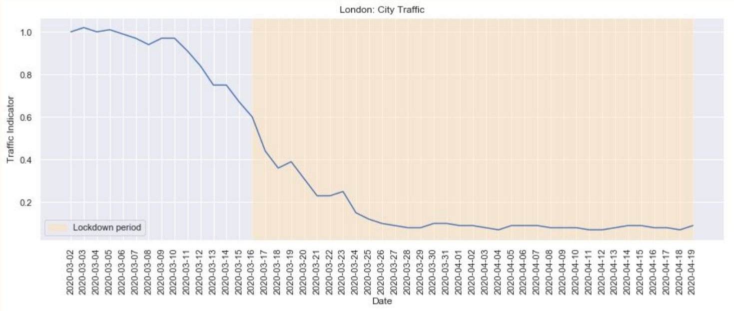


Statistical Analysis(Cont..): Measures

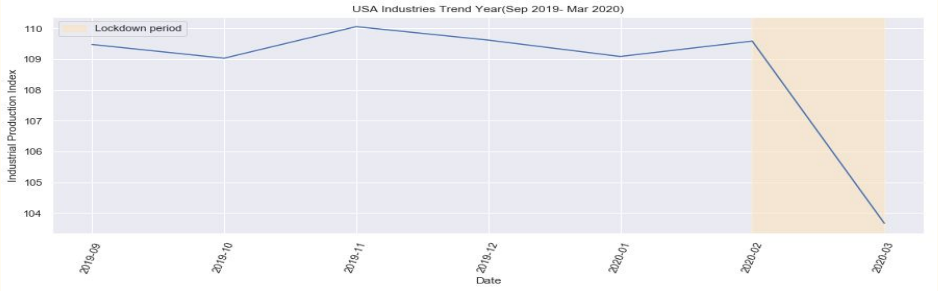
Air Traffic(South Korea): ▼ 52%



City Traffic(London): ▼ 85%

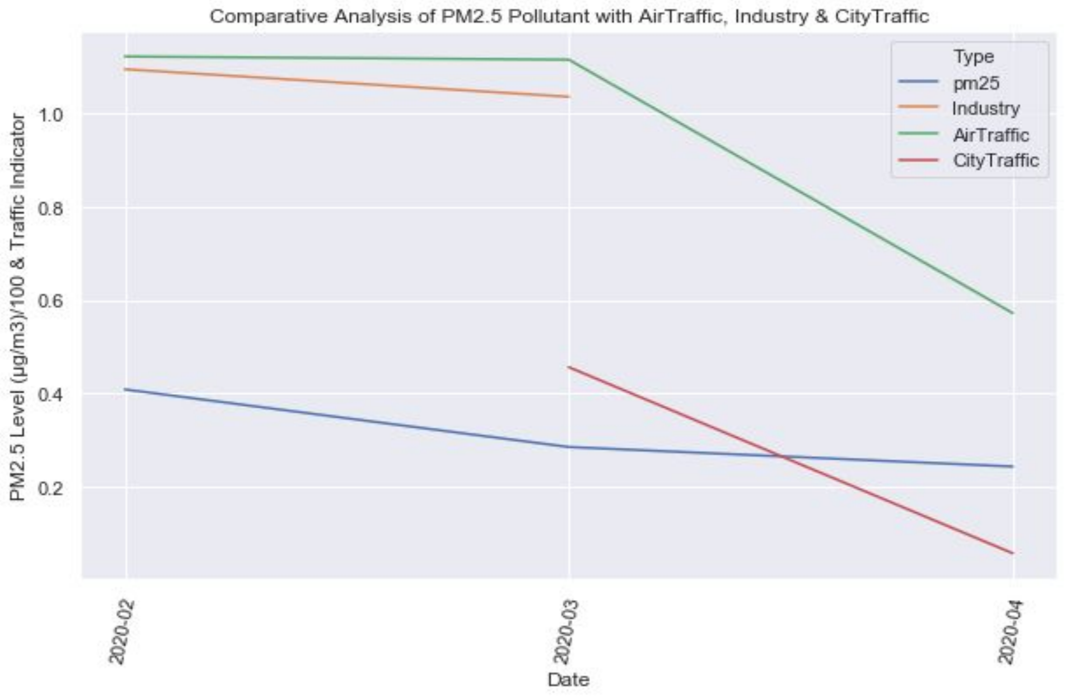


Industrial Operations Trend(USA): ▼ 5%

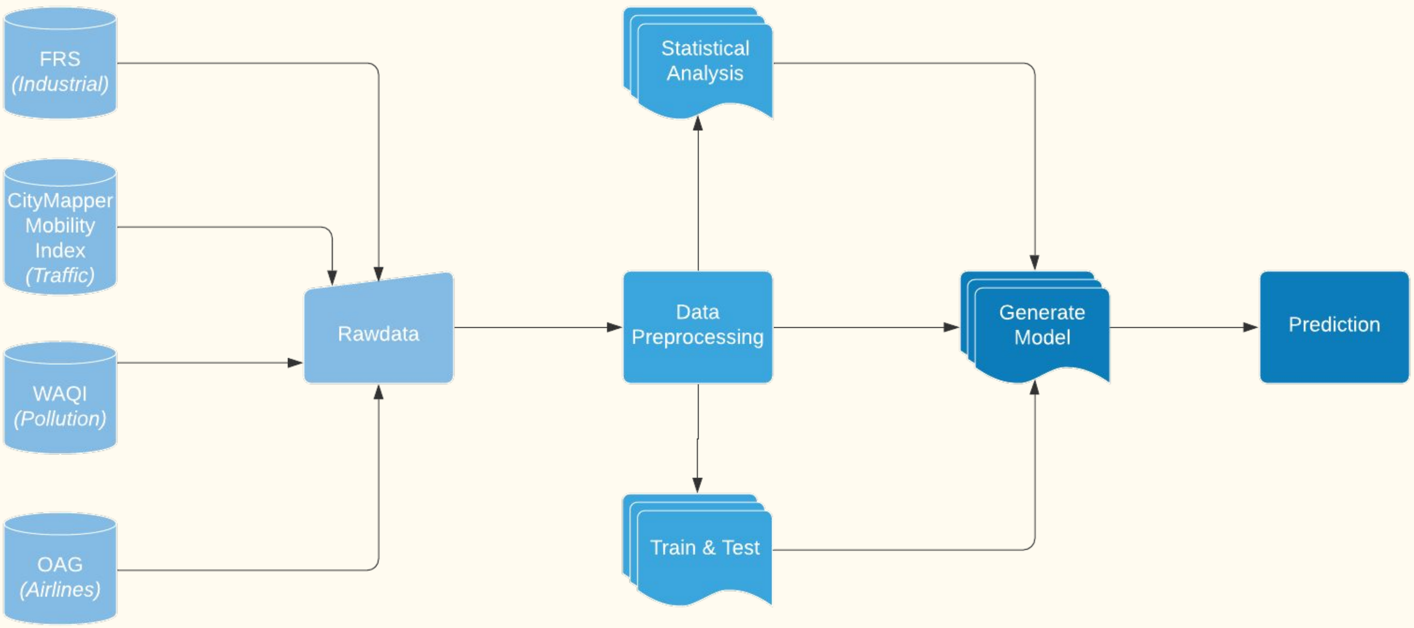


Results :Comparative Analysis

Successfully correlated that decline in flights operation, vehicles and industries positively contributed to environment health.



Proposed Solution



Algorithms Variations

- Auto Regression: An auto-regressive (AR) model pre-dicts future behavior based on past behavior. The processis basically a linear regression of the data in the currentseries against one or more past values in the same series. AR model usually gets “close enough” for it to be useful in most scenarios.

Algorithms Variations

- 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.

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
```

Algorithms Variations

- 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.

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.

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.

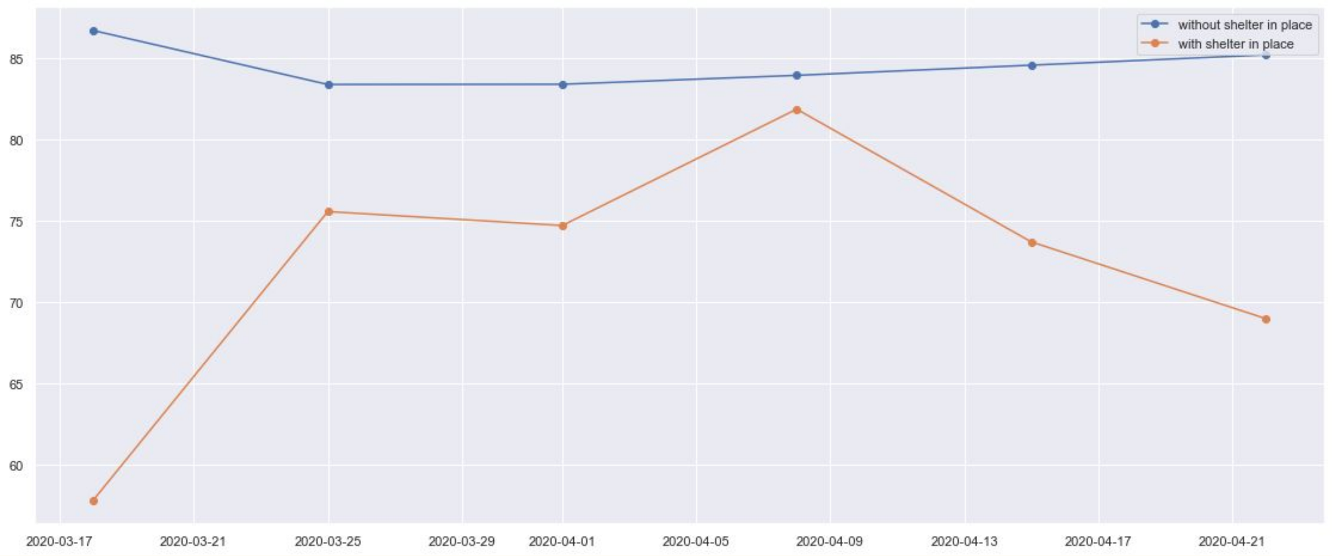
Differencing is the method to compute the difference of consecutive terms in the series. Differencing is typically performed to get rid of the varying mean. Whereas Log Transformations are used to stabilize the non-constant variance of a series.

Evaluations(2-3 slides)

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 ?? shows the final predicted graph for the PM2.5 content with Shelter in place and without it for predicted valus of 2020.



Things that work

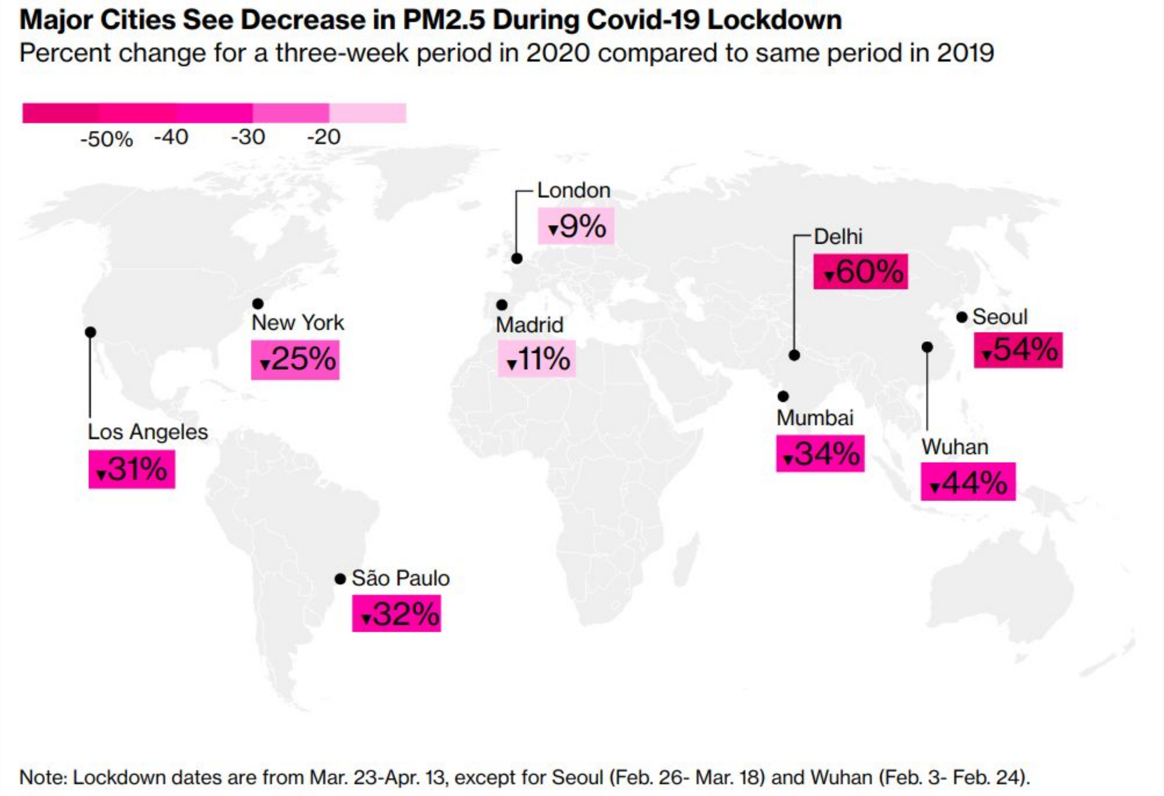
- Use of lockdown timeline to understand the data and find out a trend
- Averaging of pollution data weekly because it was found out that people react collectively to news announcement and start panic buying and hoarding and there by spiking the pollution for that single day.

Things that didn't work

- Dataset had outliers which needed to be weeded out.
- Some cities did not show a drastic change in pollution which was unexpected, researching about them it was found out that they had good environmental policy in place all year round
- Data set for African countries was not abundant.

Conclusions

- Just 1 microgram per cubic metre 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.
- 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.



Future Suggestions

1) Electric Vehicles:Make electric vehicles the new normal and out number the conventional cars.

2) Renewable Power:Sink tax dollars and resources into clean-energy subsidies.

3) Afforestation:Regrow enough trees to cover India Twice.

4) Carbon Tax:Charge polluters $75 for every metric ton of CO2 they produce.

5) Nuclear Power:Mandate government handouts tosource 11% of energy production through nuclear fuel.