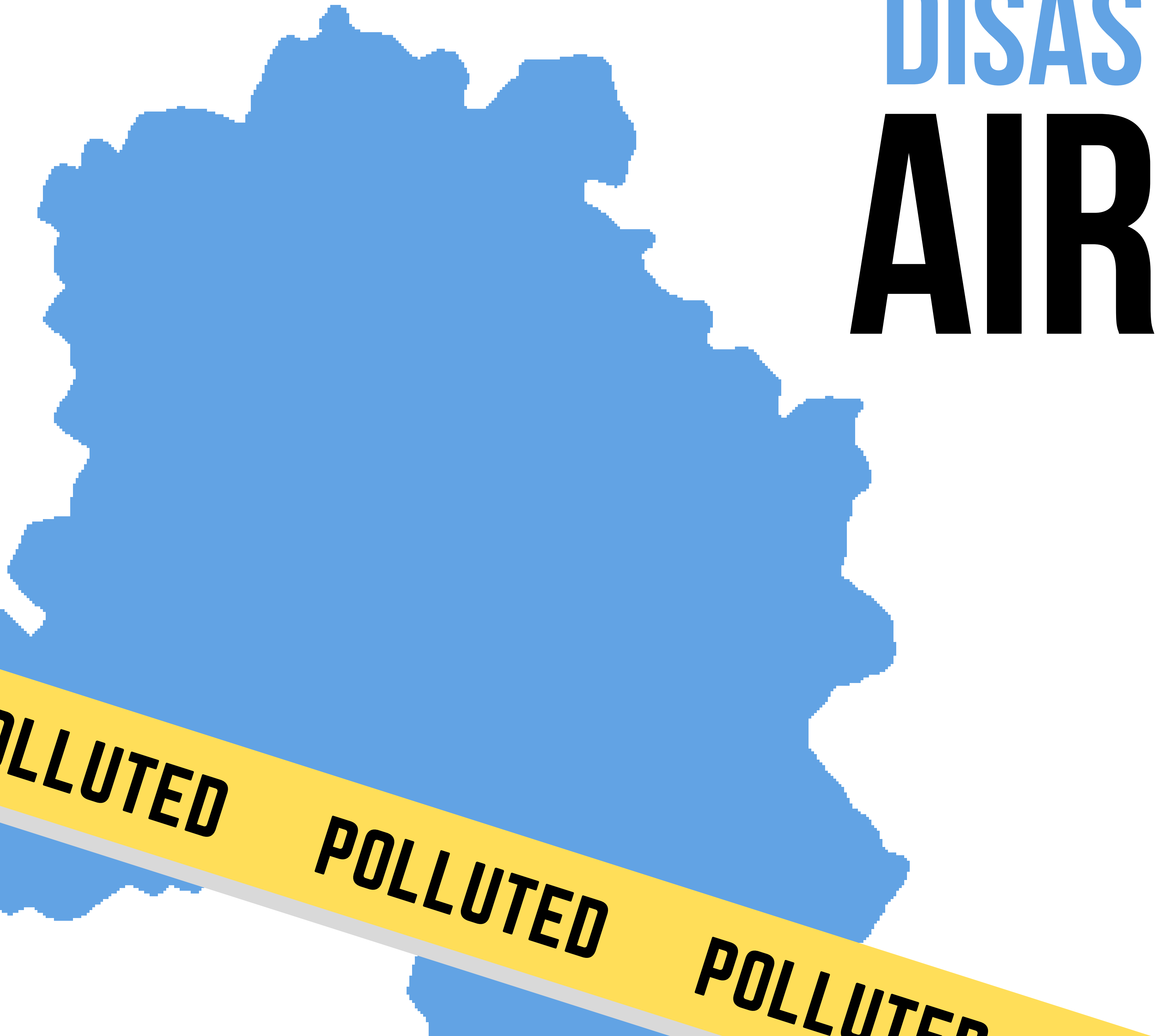
DISASTER MANAGEMENT

AIR POLLUTION IN DELHI



Team 3

PROJECT OBJECTIVES

The 5 Objectives

AQI and Pollutants

Analyse pollutants and AQI over the years in Delhi

Trend and Seasonality

Analyse the trend and seasonality in AQI and pollutants

Factors Responsible

Study reasons behind the concentrations and seasonality

Effect of Lockdown

Analyse the effect of Lockdown on air quality and pollution

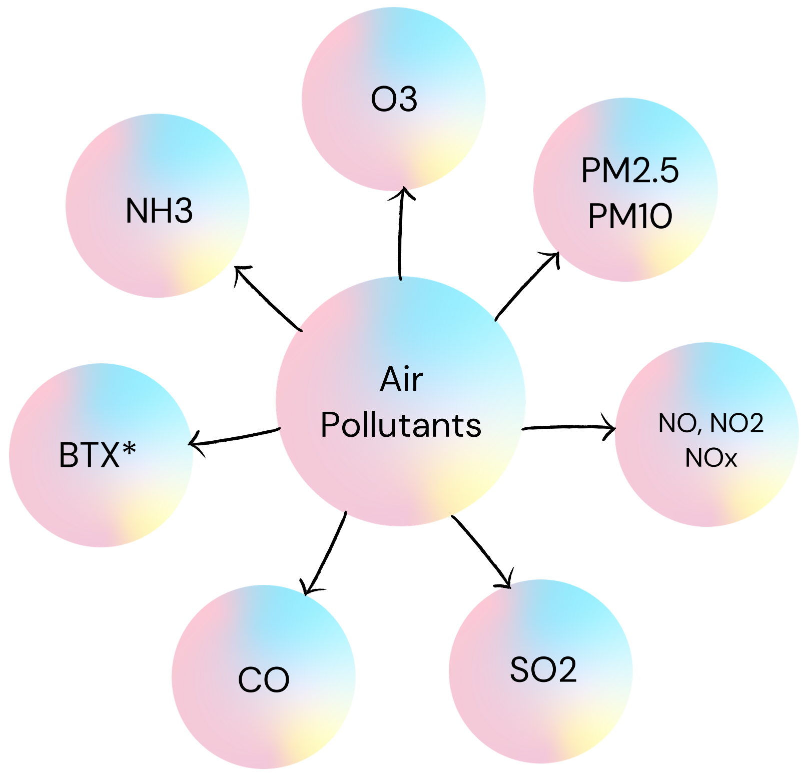
Future Air Quality

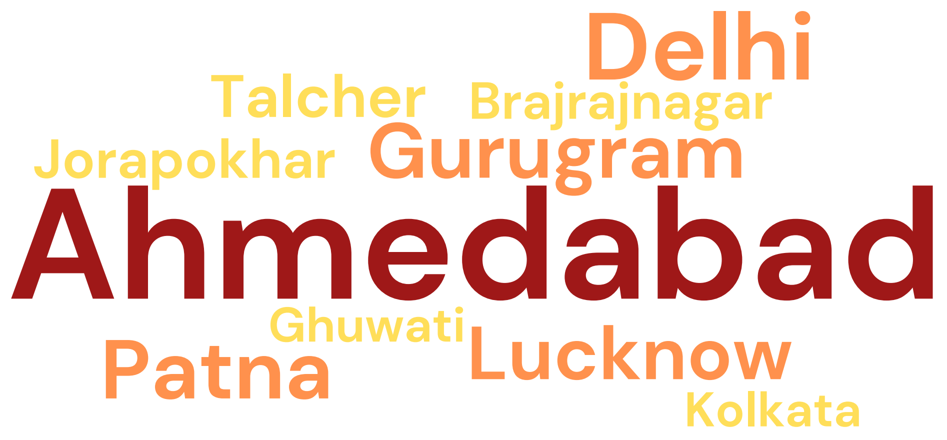
Predict future concentrations of pollutants and AQI

RECAP OF PHASE 1

Phase 1: Recap

The available data is between 2015-01-01 and 2020-07-01





PM2.5	117.2	2nd
PM10	232.8	1st
NO2	50.79	2nd
SO2	15.9	6th
O3	51.32	2nd
BTX	26.86	2nd
CO	1.98	3rd
AQI	259.5	2nd

Data was collected from Central Pollution and Control Board

Phase 1: Recap

The air quality index (AQI) is an index for reporting air quality on a daily basis. It is a measure of how air pollution affects one's health within a short time period.

Good **0-50** Minimal impact	**Satisfactory** **51-100** Minor breathing discomfort to sensitive people	**Moderate** **101-200** Breathing discomfort to people with lung/heart problems
Poor **201-300** Discomfort to people on prolonged exposure	**Very Poor** **301-400** Repiratory effects on prolonged exposure	**Severe** **>401** Respiratory effects on even healthy people

Calculation of AQI: AQI is defined as ratios of the measured concentration of the atmospheric pollutants to their standard prescribed values.

$$\text{AQI pollutant} = \left(\frac{\textit{pollutant concentration reading}}{\textit{Standard Concentration}}\right) x\ 100$$

The AQI calculation uses 7 measures: PM2.5, PM10, SO2, NOx, NH3, CO and O3. Final AQI is the maximum Sub-Index with the condition that at least one of PM2.5 and PM10 should be available and at least three out of the seven should be available.

OBJECTIVE 1

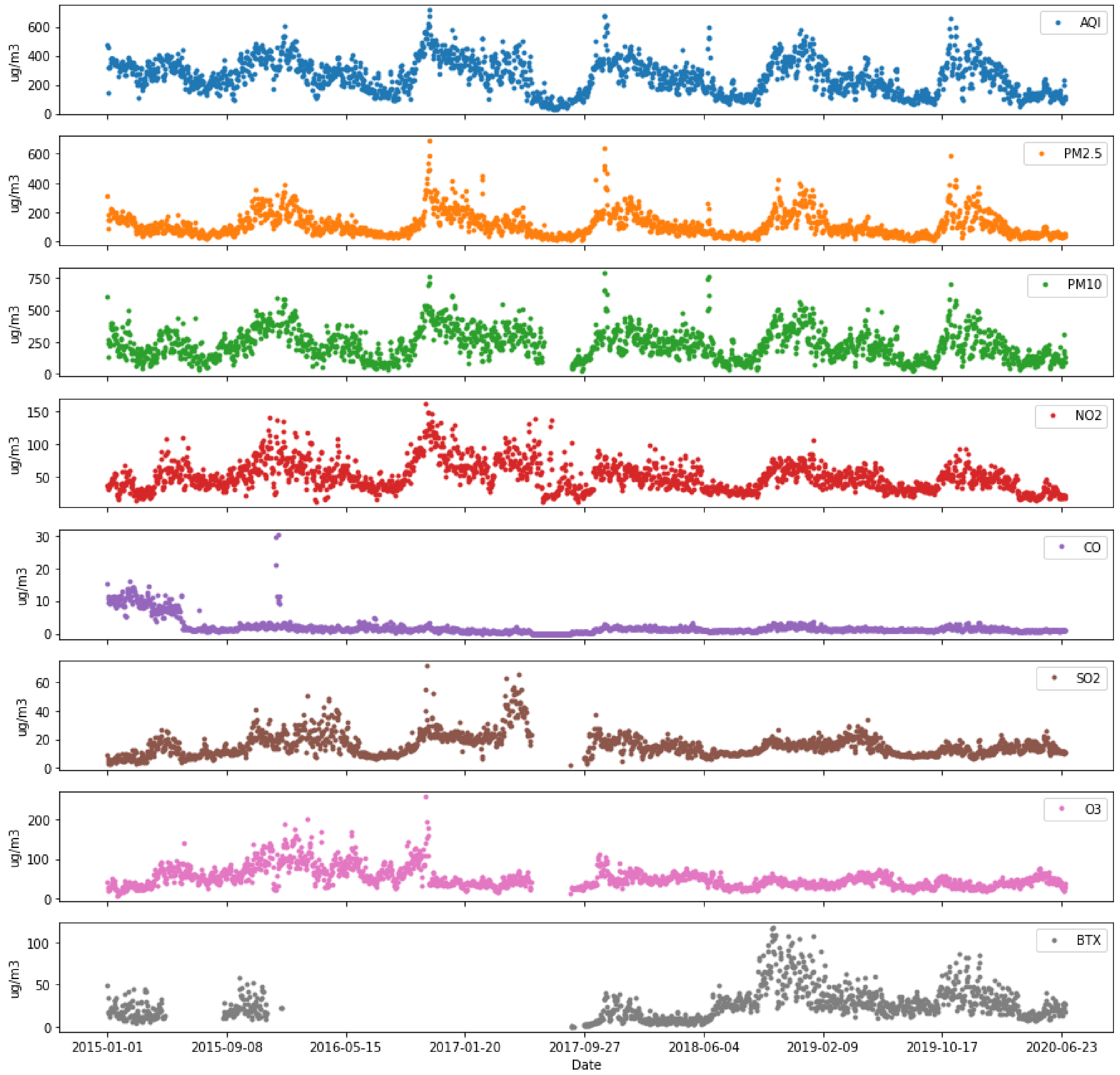
Analysis of AQI and Pollutants

Analyse pollutants and AQI over the years in Delhi

Objective 1: Analyse pollutants and AQI over the years in Delhi

AQI and pollutants over the years

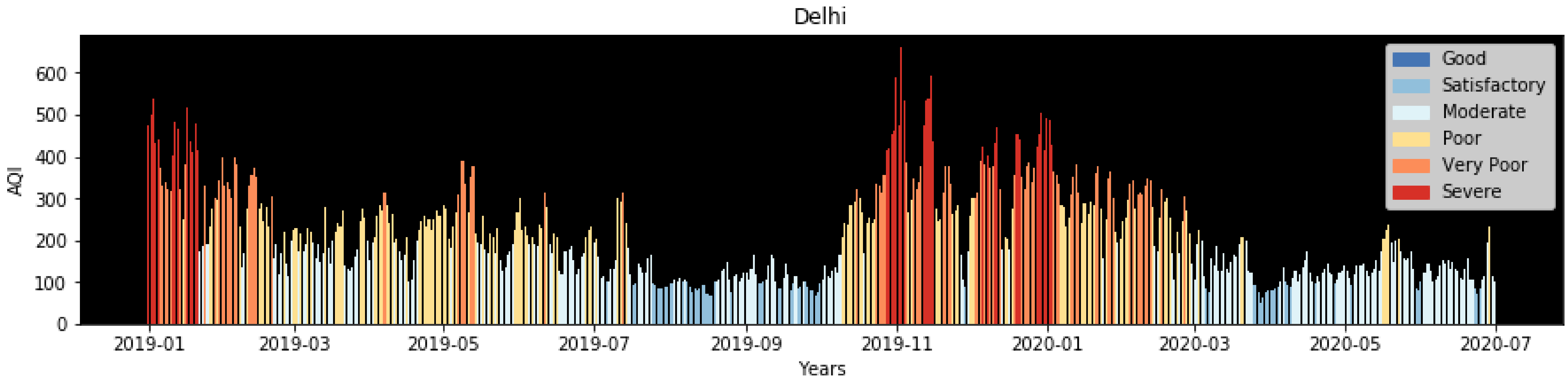
- increasing trend: AQI, PM2.5, PM10, BTX
- decrease in values in 2020
- presence of seasonality
- missing Plots



Objective 1: AQI and Pollutants

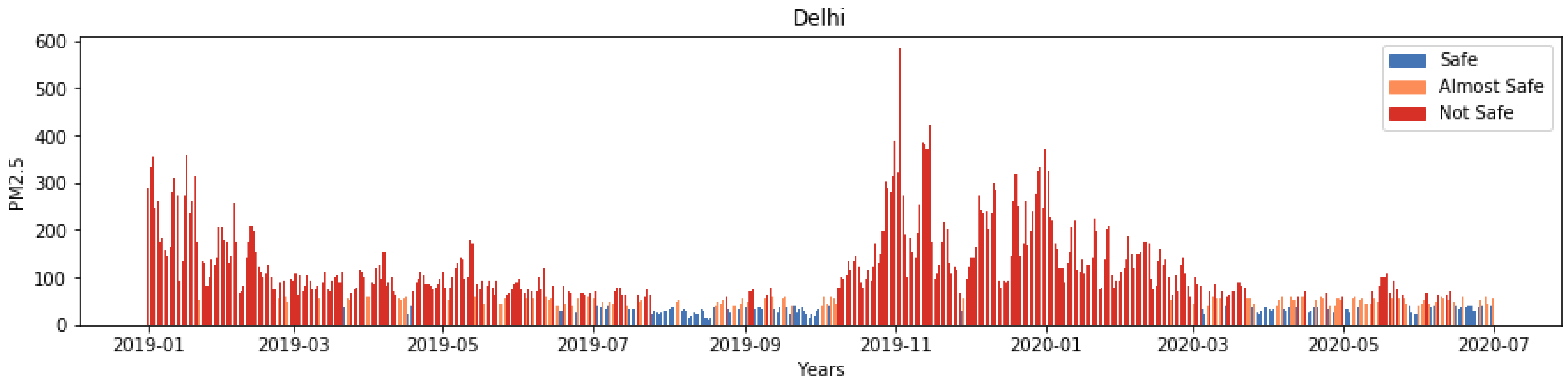
AQI and pollutants over 2019-2020: AQI

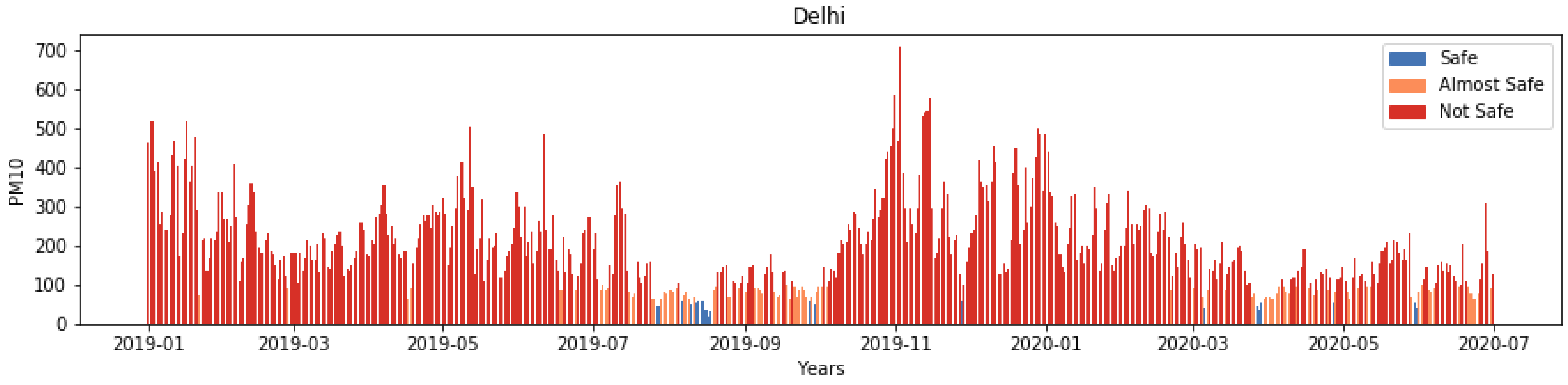
- AQI goes up to more than 600: Severe
- Highest AQI in November



Objective 1: AQI and Pollutants

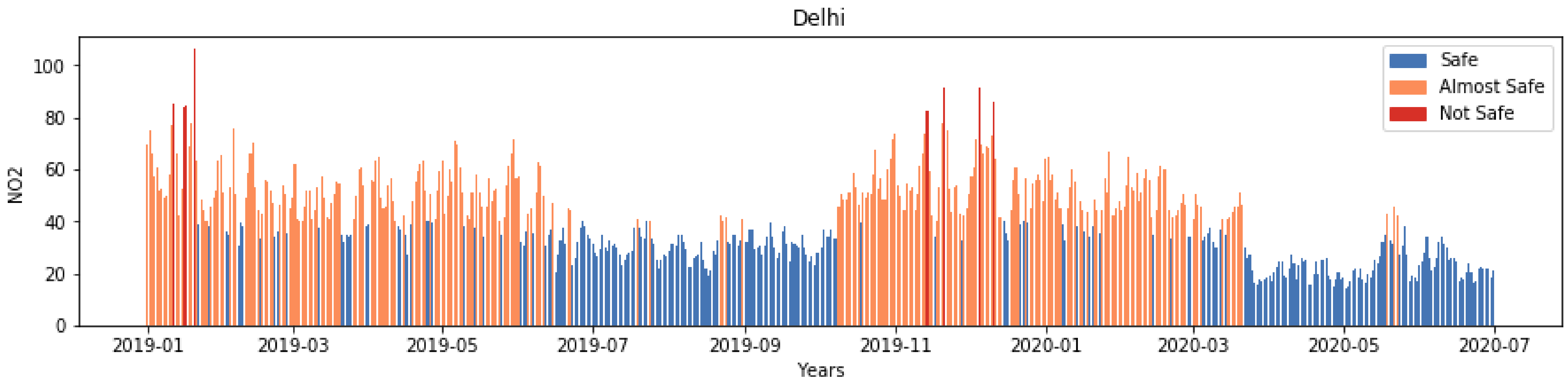
AQI and pollutants over 2019-2020: PM2.5 and PM10

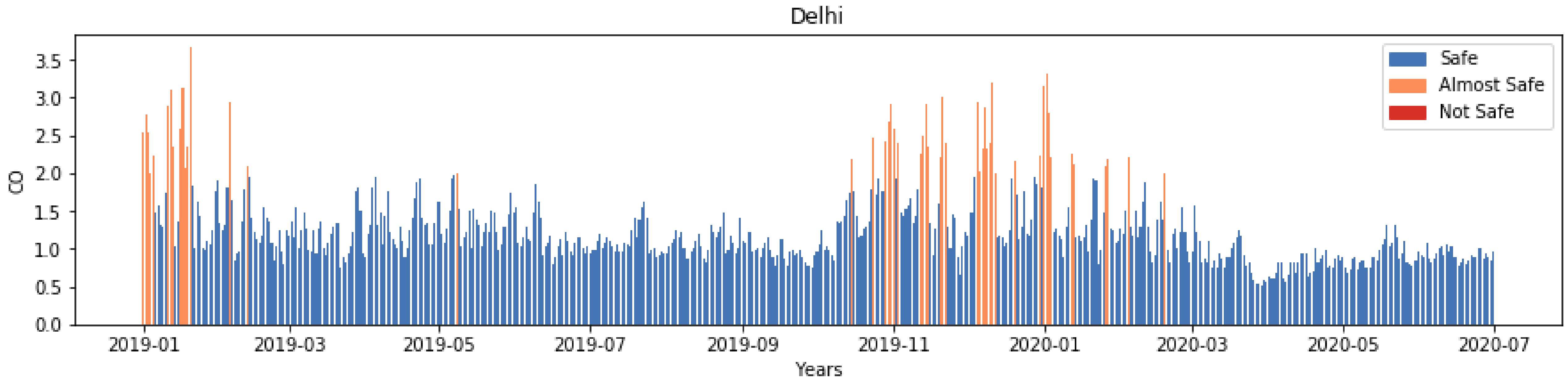




Objective 1: AQI and Pollutants

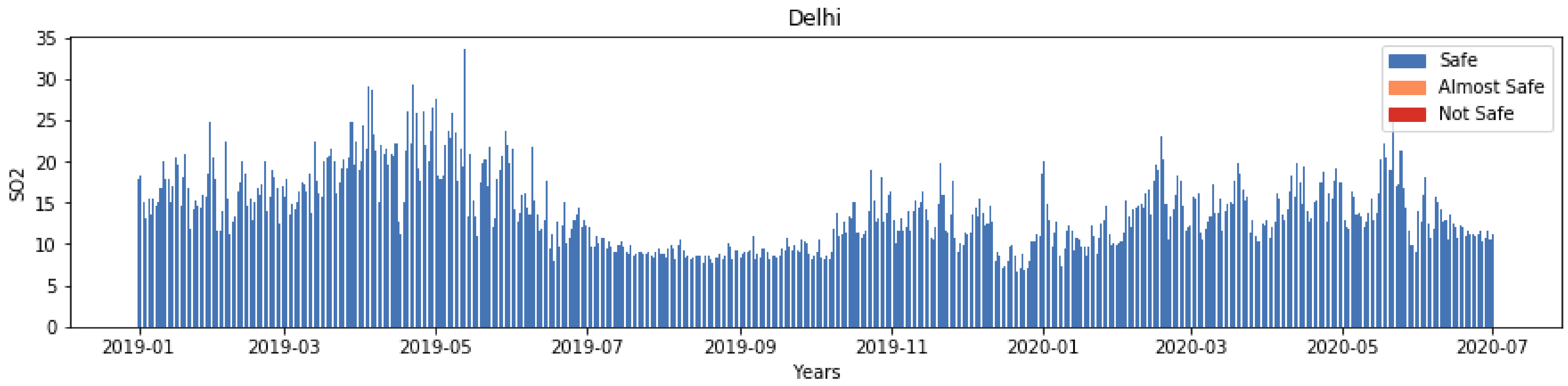
AQI and pollutants over 2019-2020: NO2 and CO

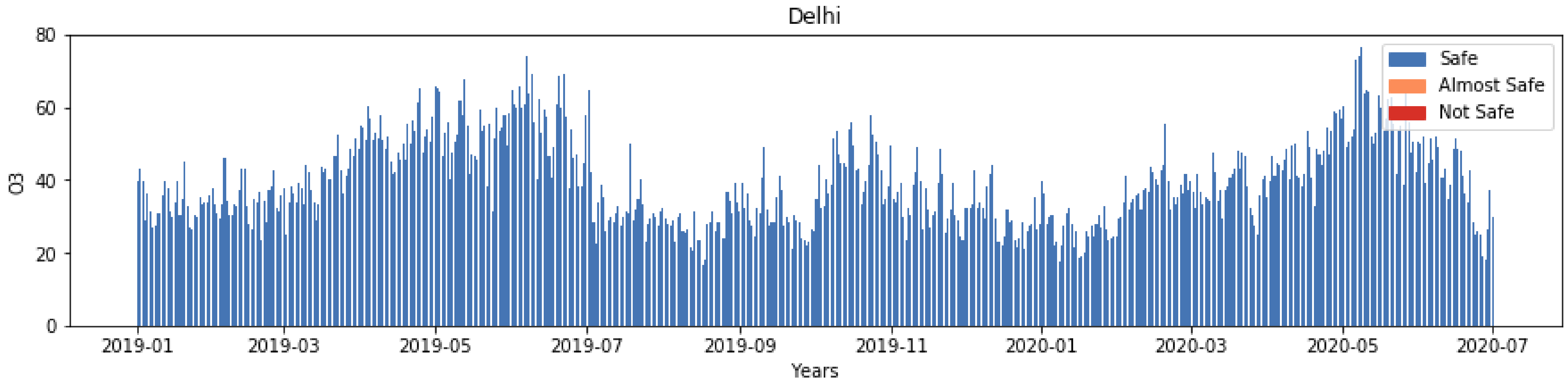




Objective 1: AQI and Pollutants

AQI and pollutants over 2019-2020: SO2 and O3





OBJECTIVE 2

Trend and Seasonality

Analyse the trend and seasonality in AQI and pollutants

Objective 2: Trend and Seasonality

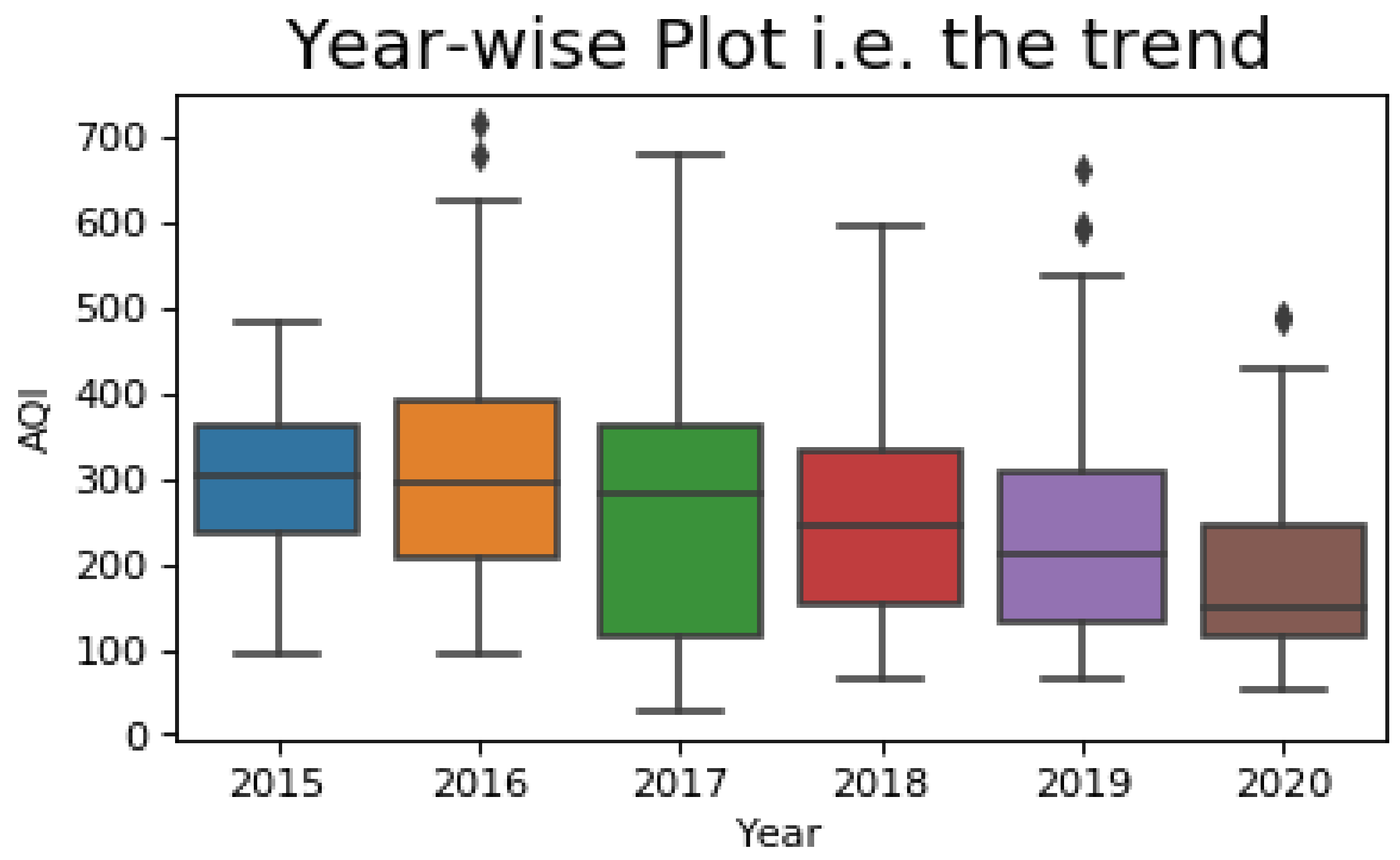
Trend

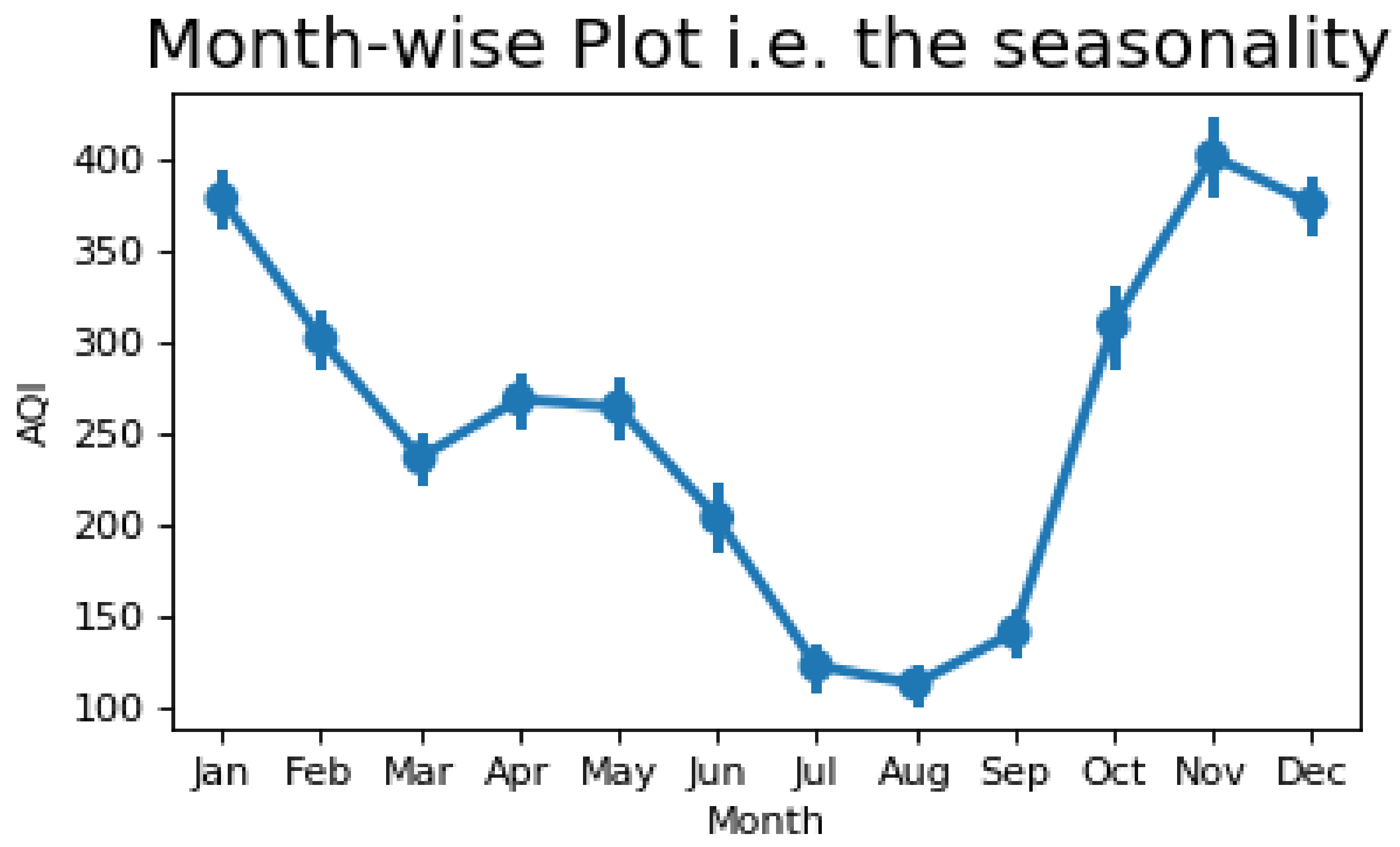
The trend shows the general tendency of the data to increase or decrease during a long period of time.

Seasonality

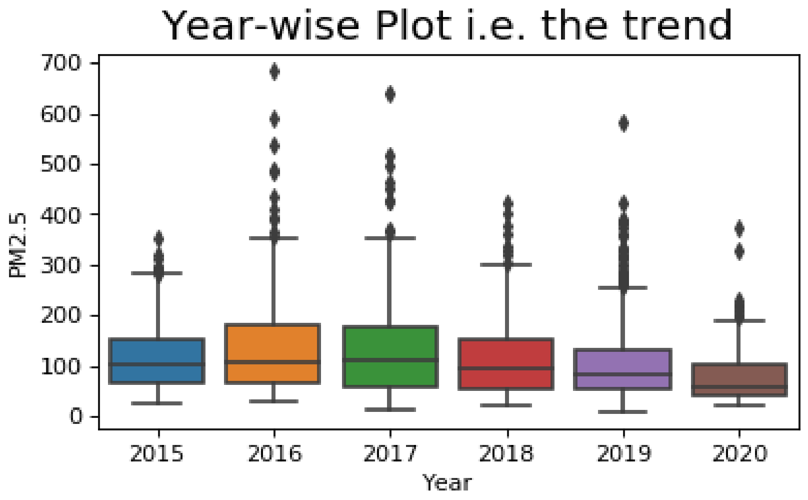
Seasonality shows the regular and predictable changes in the data that recur every calendar year.

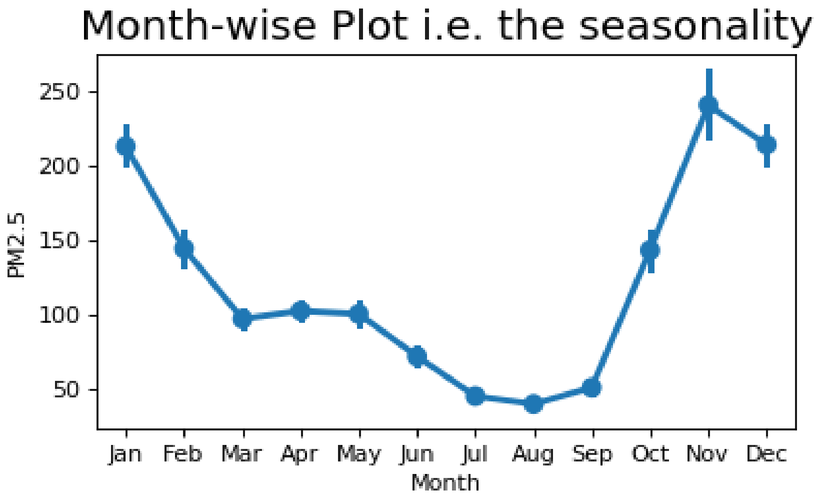
Trend and Seasonality of Delhi AQI

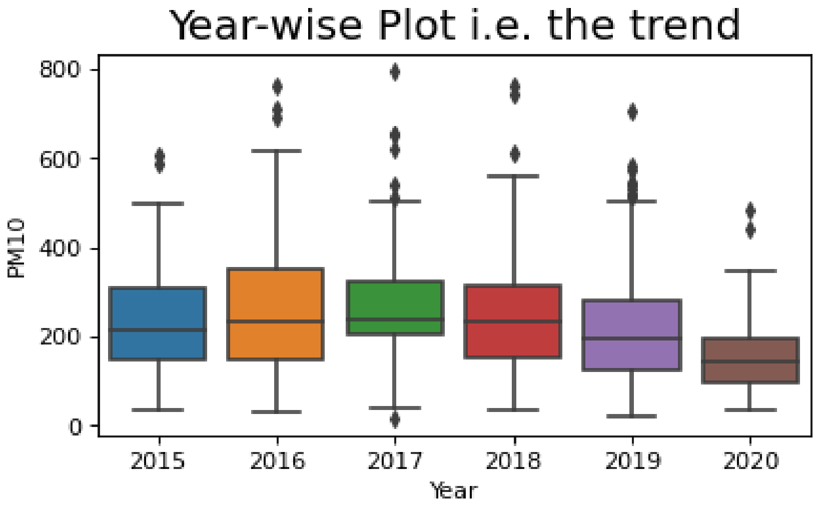


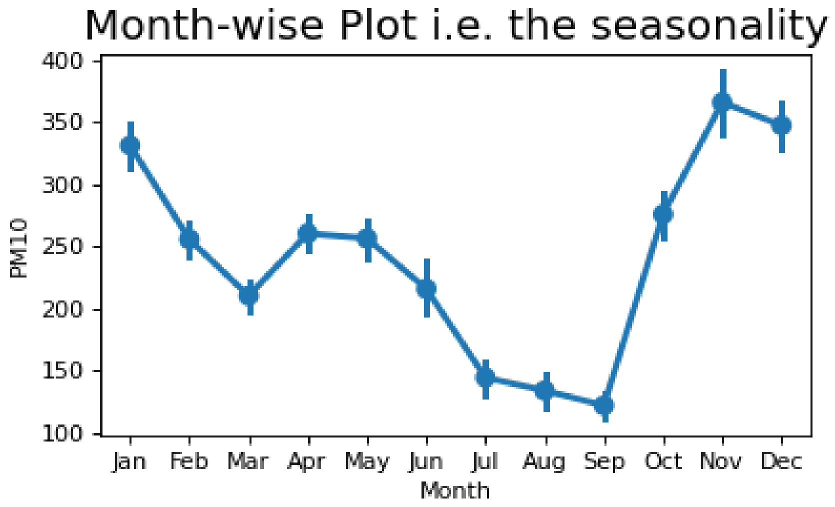


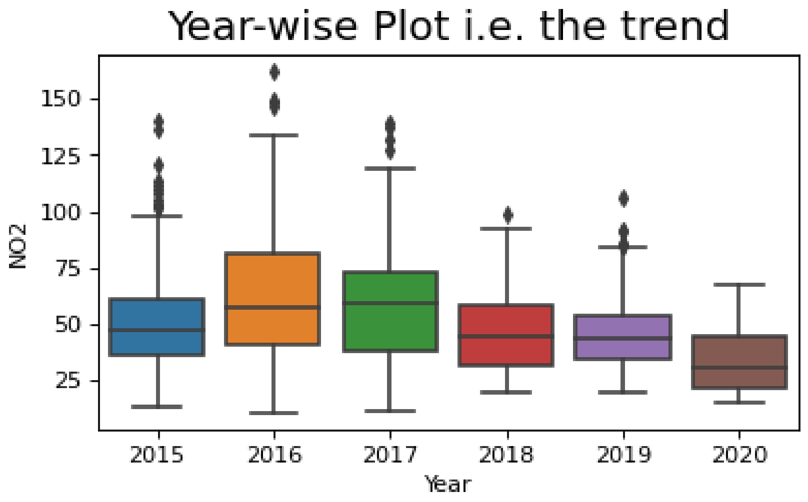
Trend and Seasonality of Delhi Pollutants

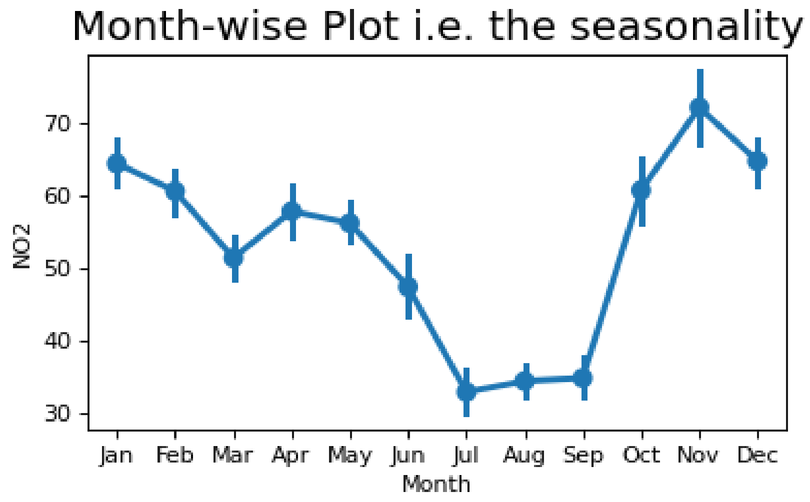


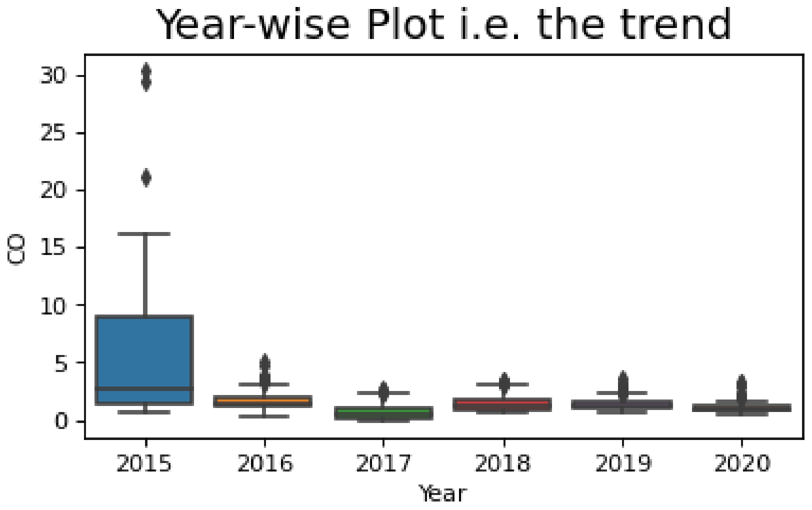


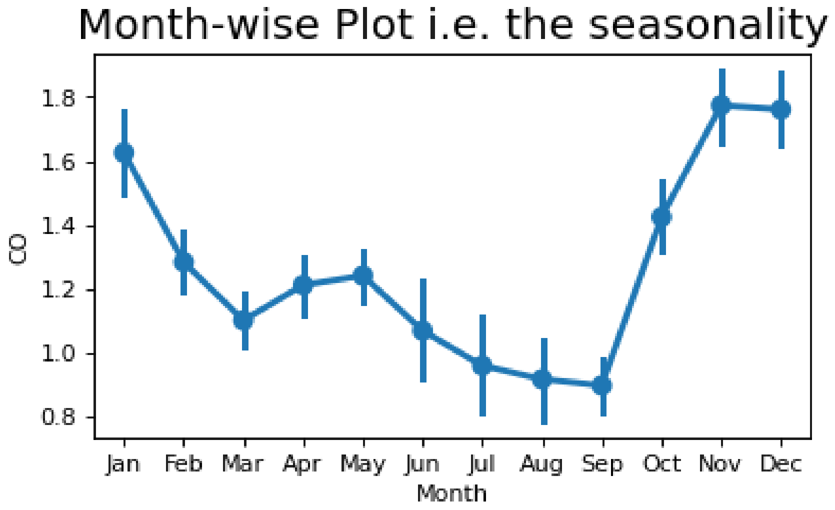












Trend and Seasonality of Delhi Pollutants

There is an abnormal rise in the AQI values and pollutant concentrations during the winter. Why?

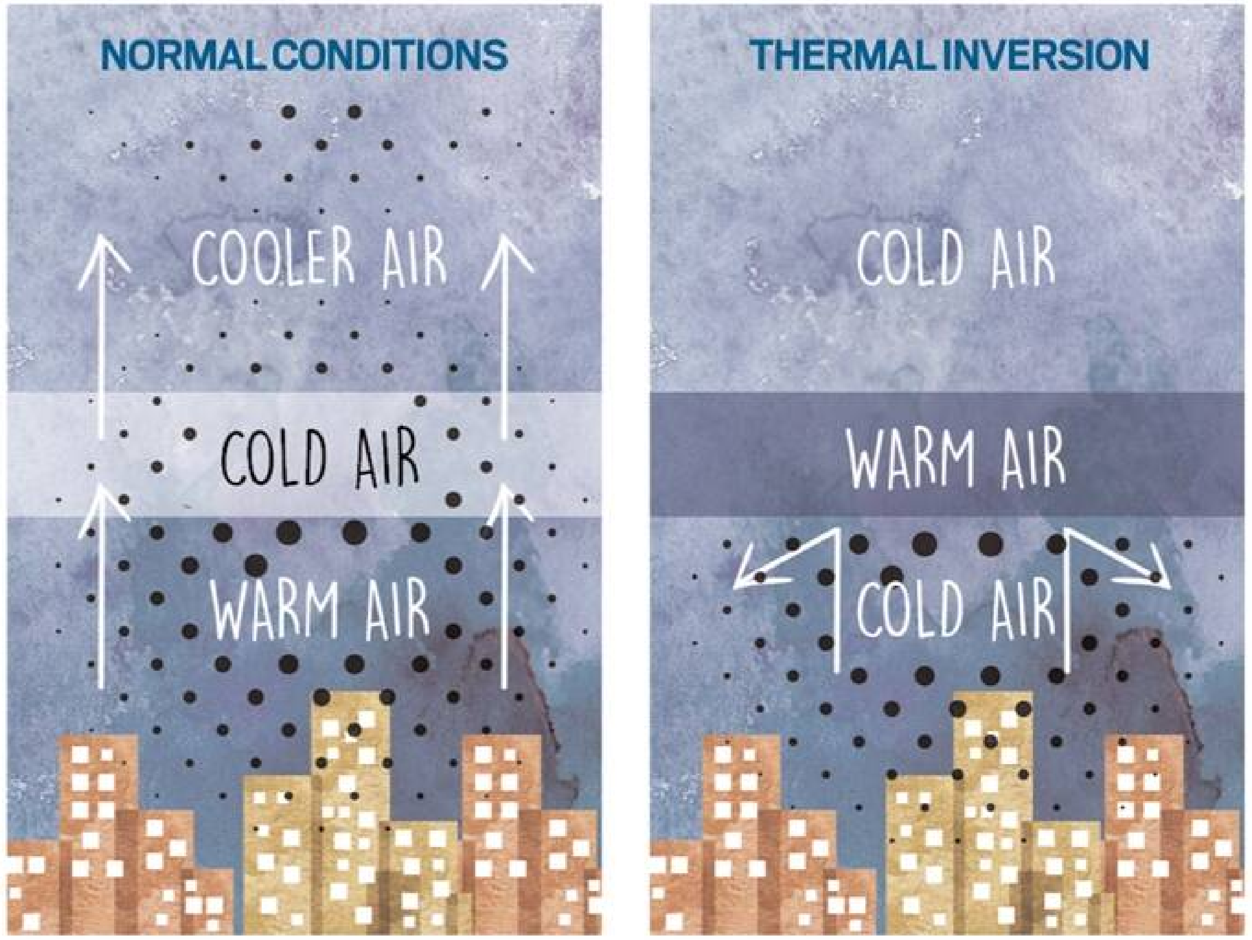
OBJECTIVE 3

Factors Responsible

Study reasons behind the pollutant concentrations and seasonality

Winter Inversion

In summer, air in the planetary boundary layer is warmer and lighter, and rises upwards more easily. This carries pollutants away from the ground and mixes them with cleaner air in the upper layers of the atmosphere however during winter, the planetary boundary layer is thinner as the cooler air near the earth's surface is dense. The cooler air is trapped under the warm air above it which forms a kind of atmospheric 'lid'. This phenomenon is called Winter Inversion.



Dip in temperature
When the temperature dips, it lowers the inversion height, which is the layer beyond which pollutants cannot disperse into the upper layer of the atmosphere. The concentration of pollutants in the air increases when this happens

Objective 3: Factors Responsible

Valley Effect

When concentration of pollutants increases in low lying areas such as valleys because of certain weather conditions such as winter when cold air containing pollutants generated from vehicular emission becomes trapped by a layer of warmer air above the valley. This phenomenon is known as Valley Effect.

Wind Speeds

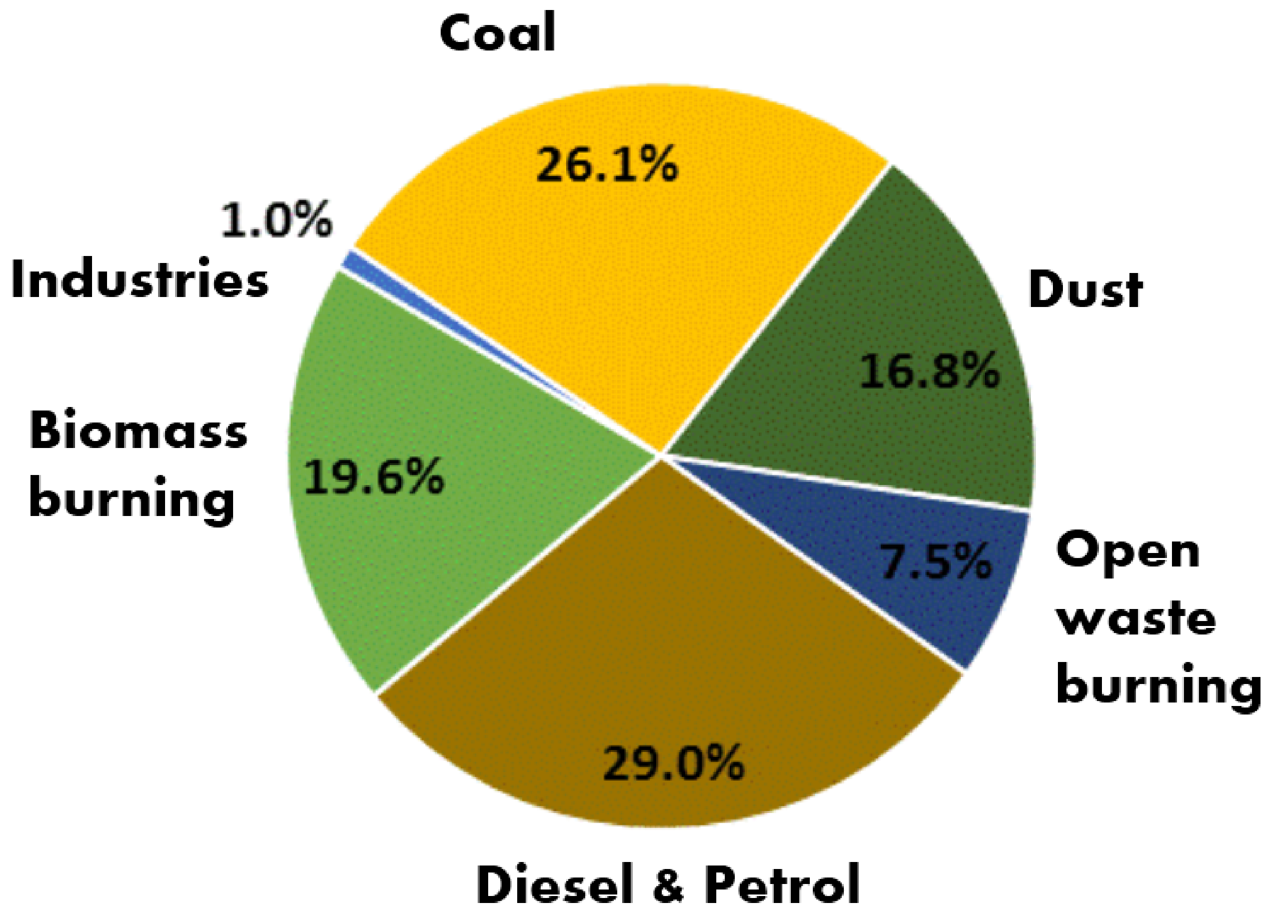
High-speed winds are very effective at dispersing pollutants, but winters bring a dip in wind speed overall as compared to summers.

The combination of these meteorological factors makes a region prone to pollution.

Objective 3: Factors Responsible

Reasons for the high concentration of pollutants in Delhi

- Dust storms
- Crop fires: Crop fires have been an easy way to get rid of paddy stubble quickly and at low cost for several years.
- Burning of solid fuels for heating
- Firecracker-related pollution during Diwali
- Non-linear city structure which inhibits displacement of air.
- Tough to take the solid waste out of the city causes pollution when burnt within the city.
- Non-linear city structure means more transport routes leading to congestion.



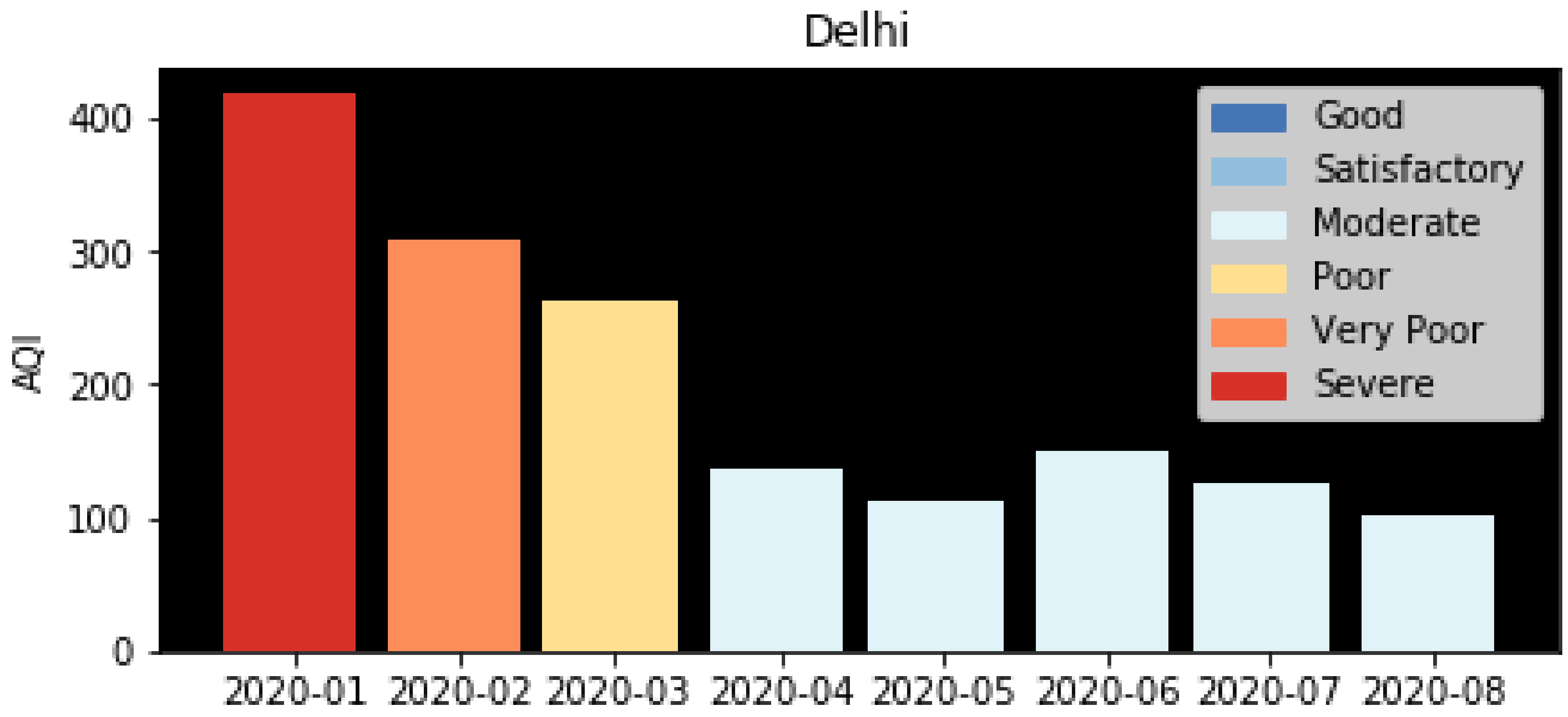
OBJECTIVE 4

Effect of Lockdown

Analyse the effect of Lockdown on air quality and pollution

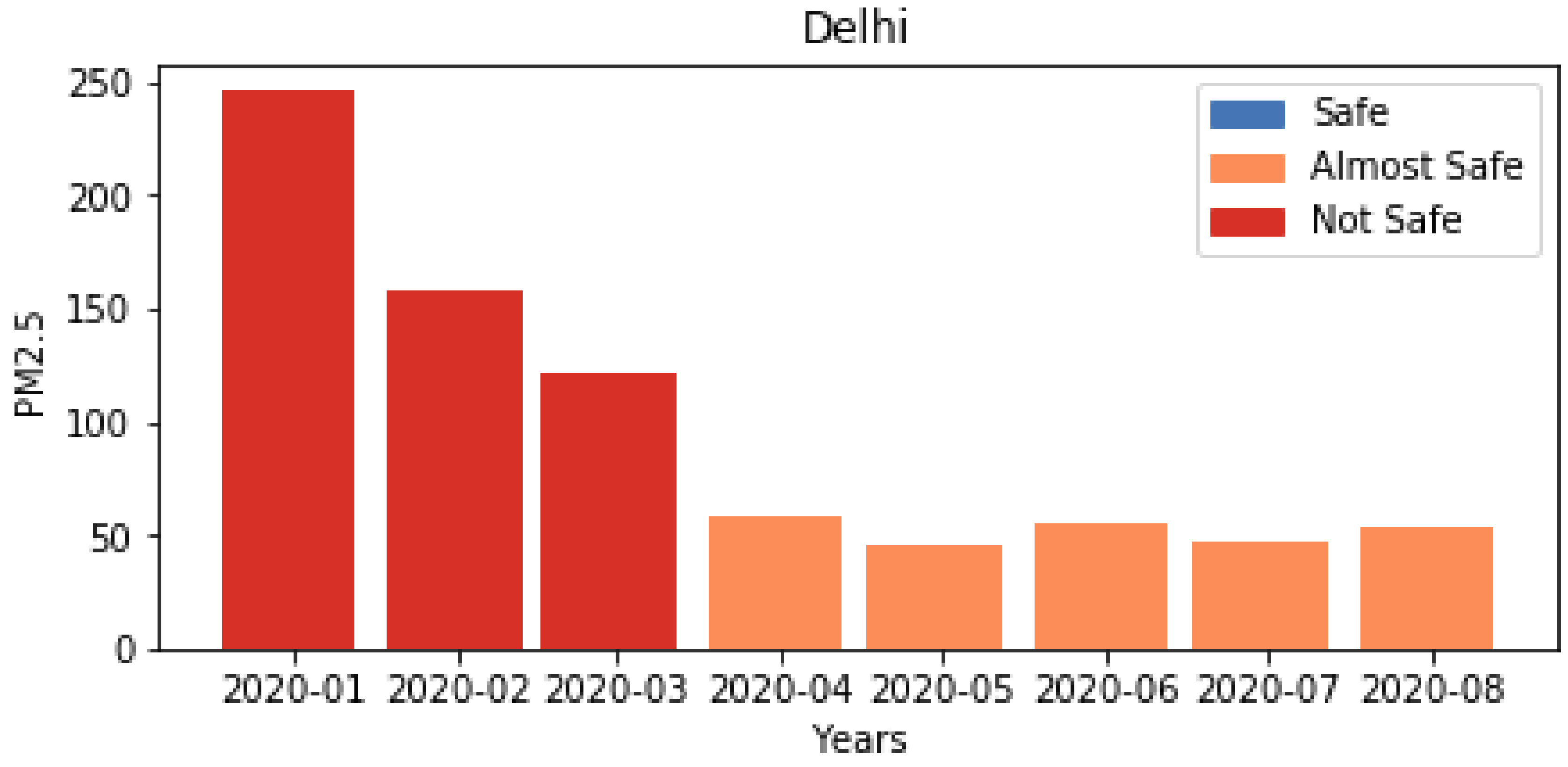
Objective 4: Effect of Lockdown

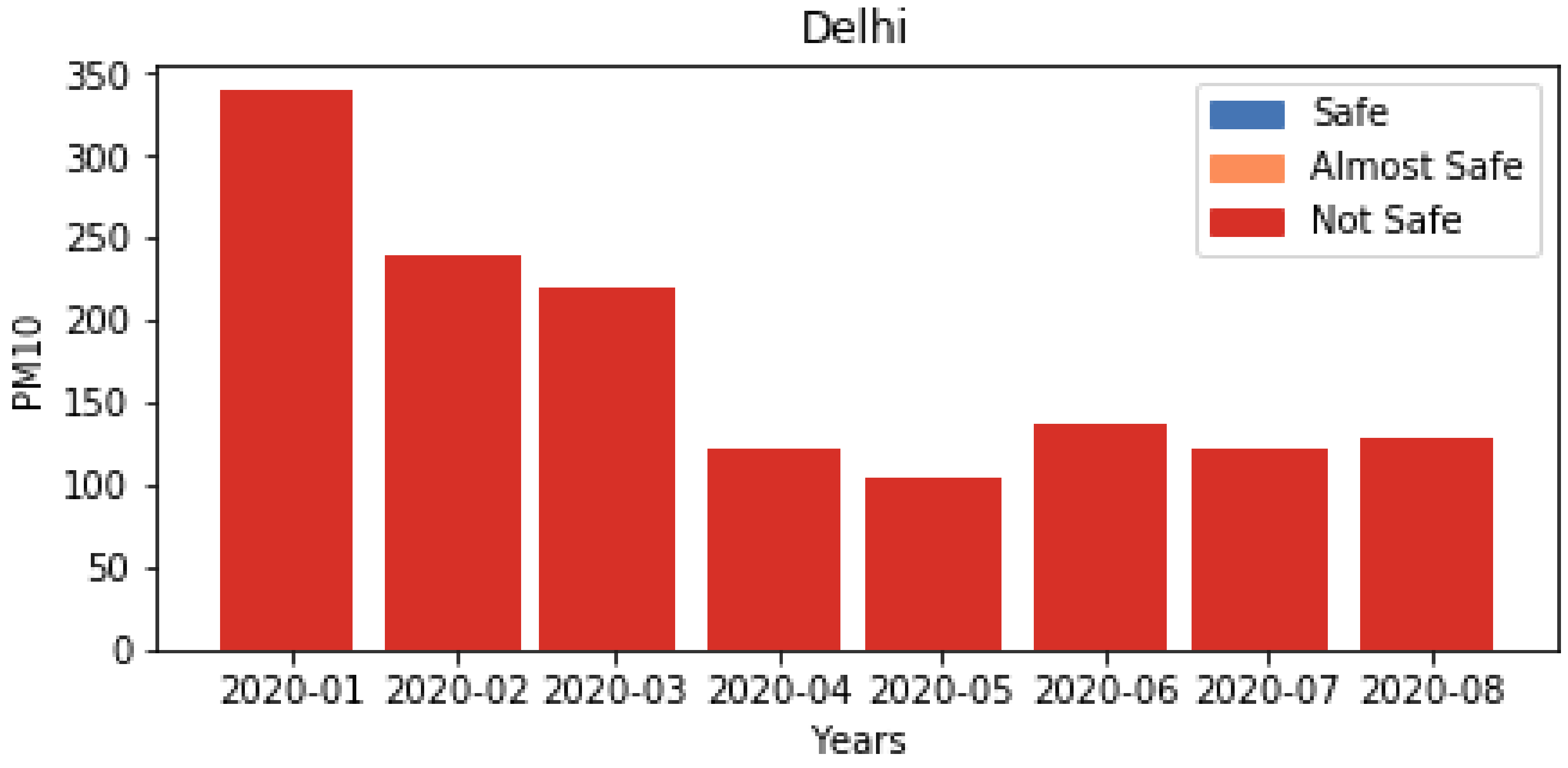
AQI and pollutant concentrations in the year 2020: AQI



Objective 4: Effect of Lockdown

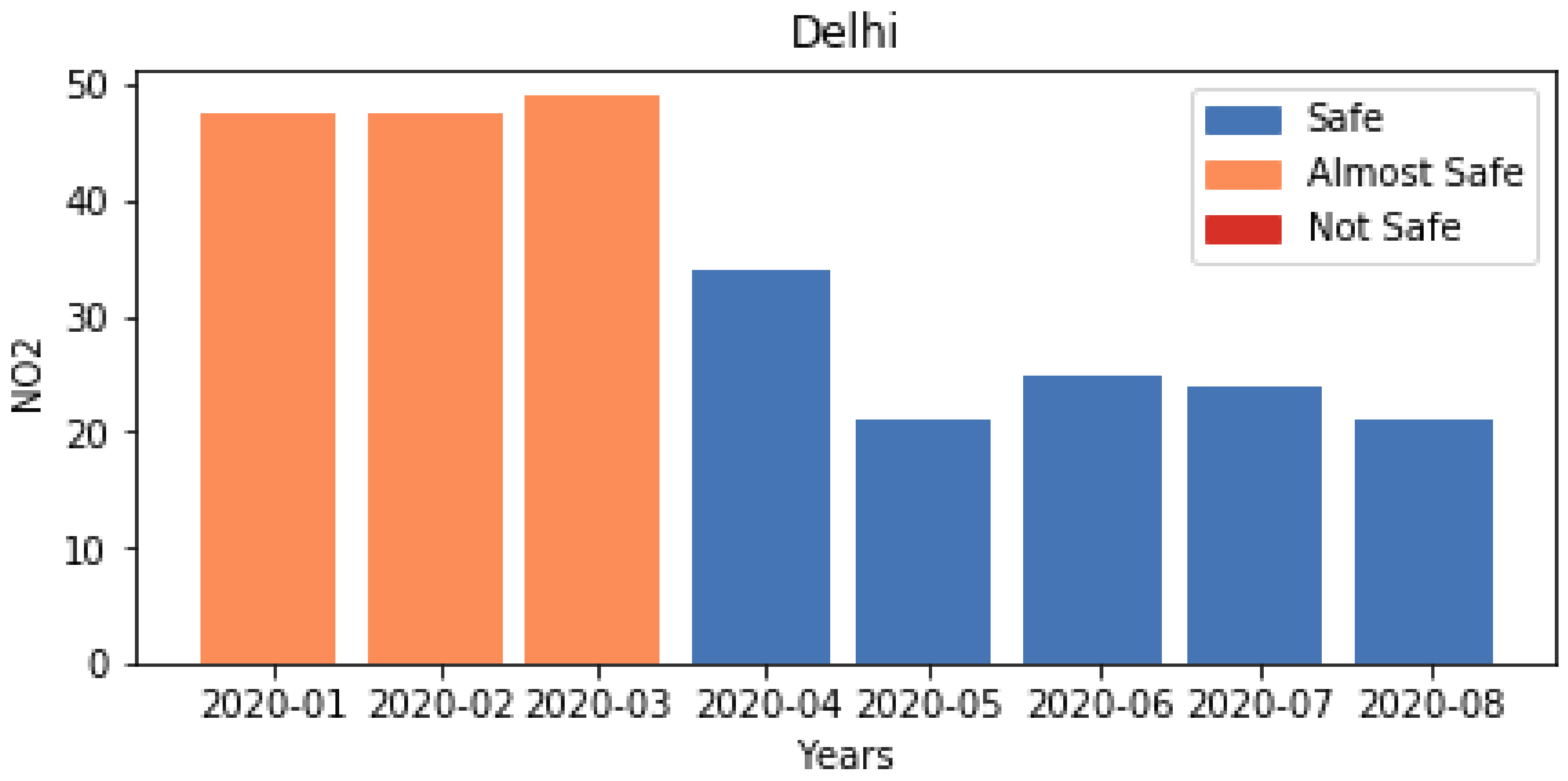
AQI and pollutant concentrations in the year 2020: PM 2.5 and PM10

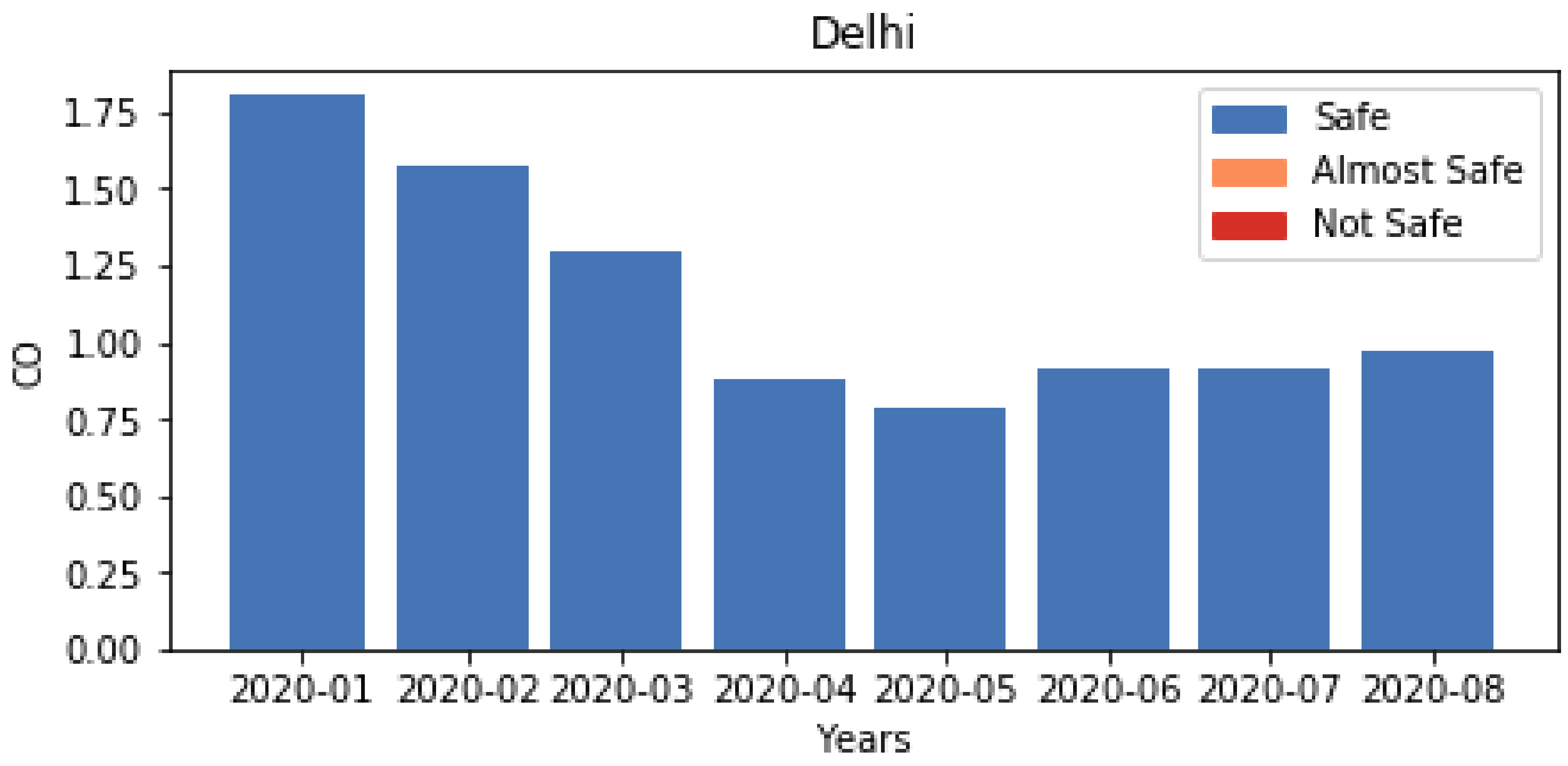




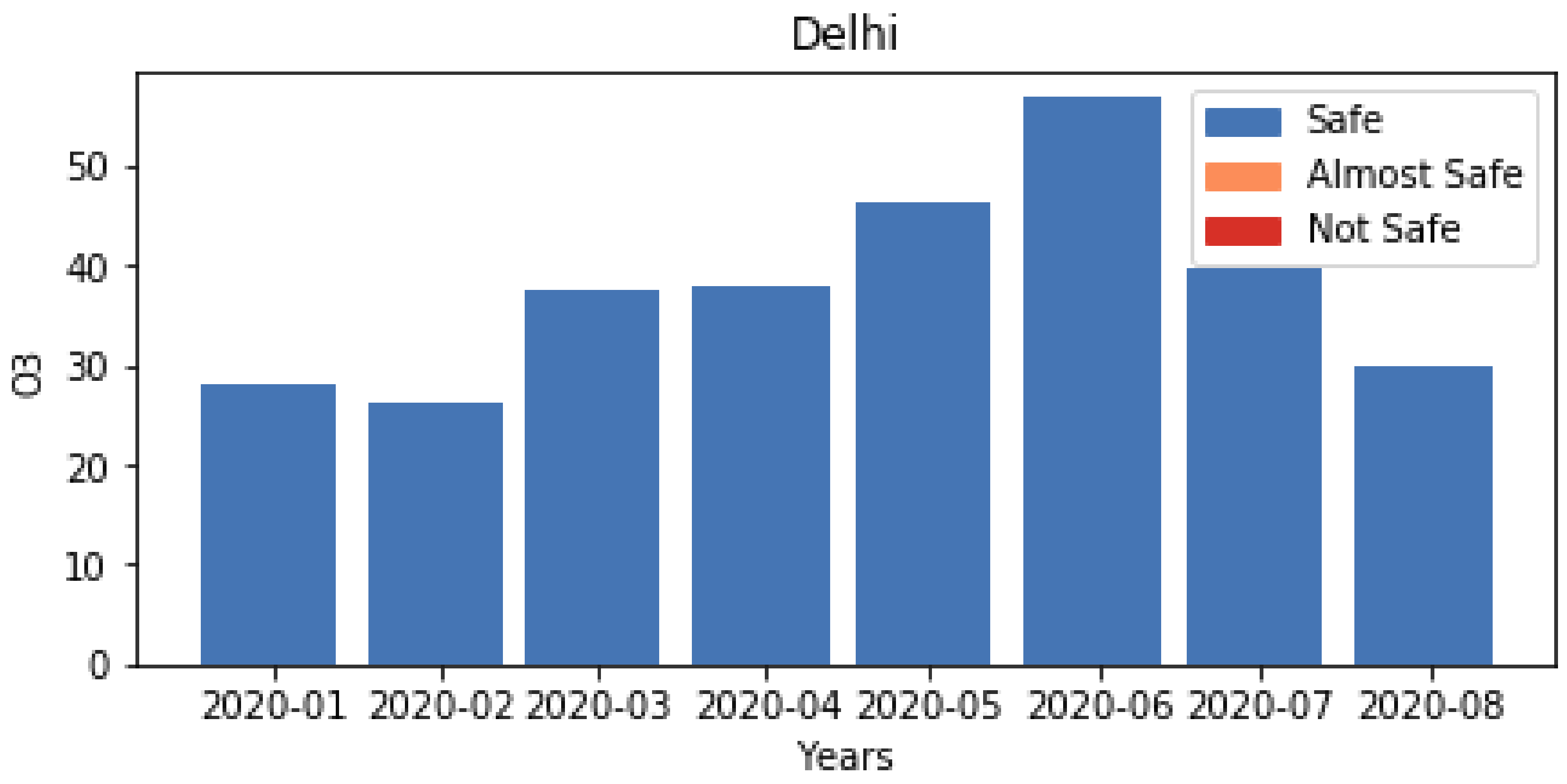
Objective 4: Effect of Lockdown

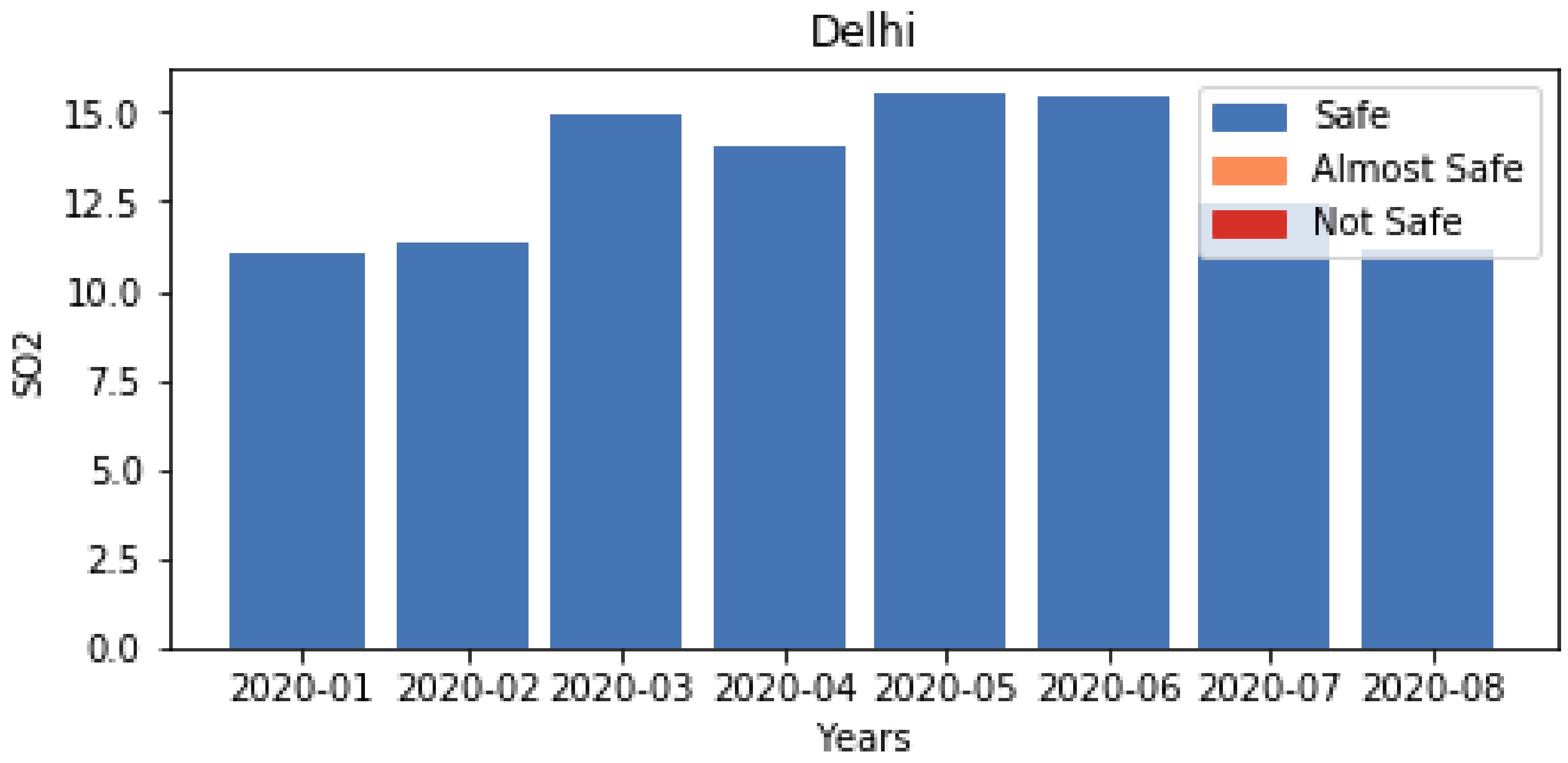
AQI and pollutant concentrations in the year 2020: NO2 and SO2





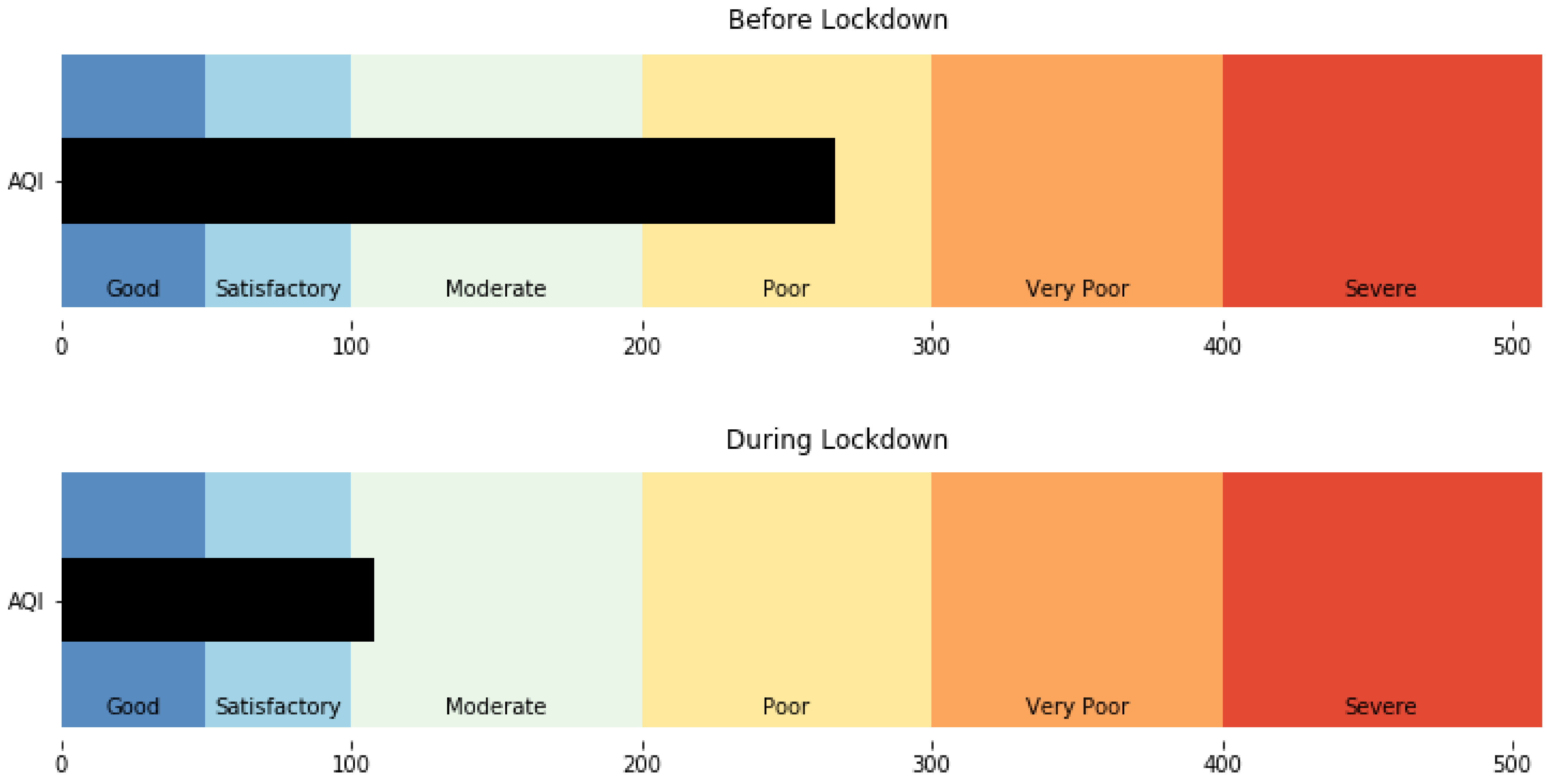
AQI and pollutant concentrations in the year 2020: O3 and CO





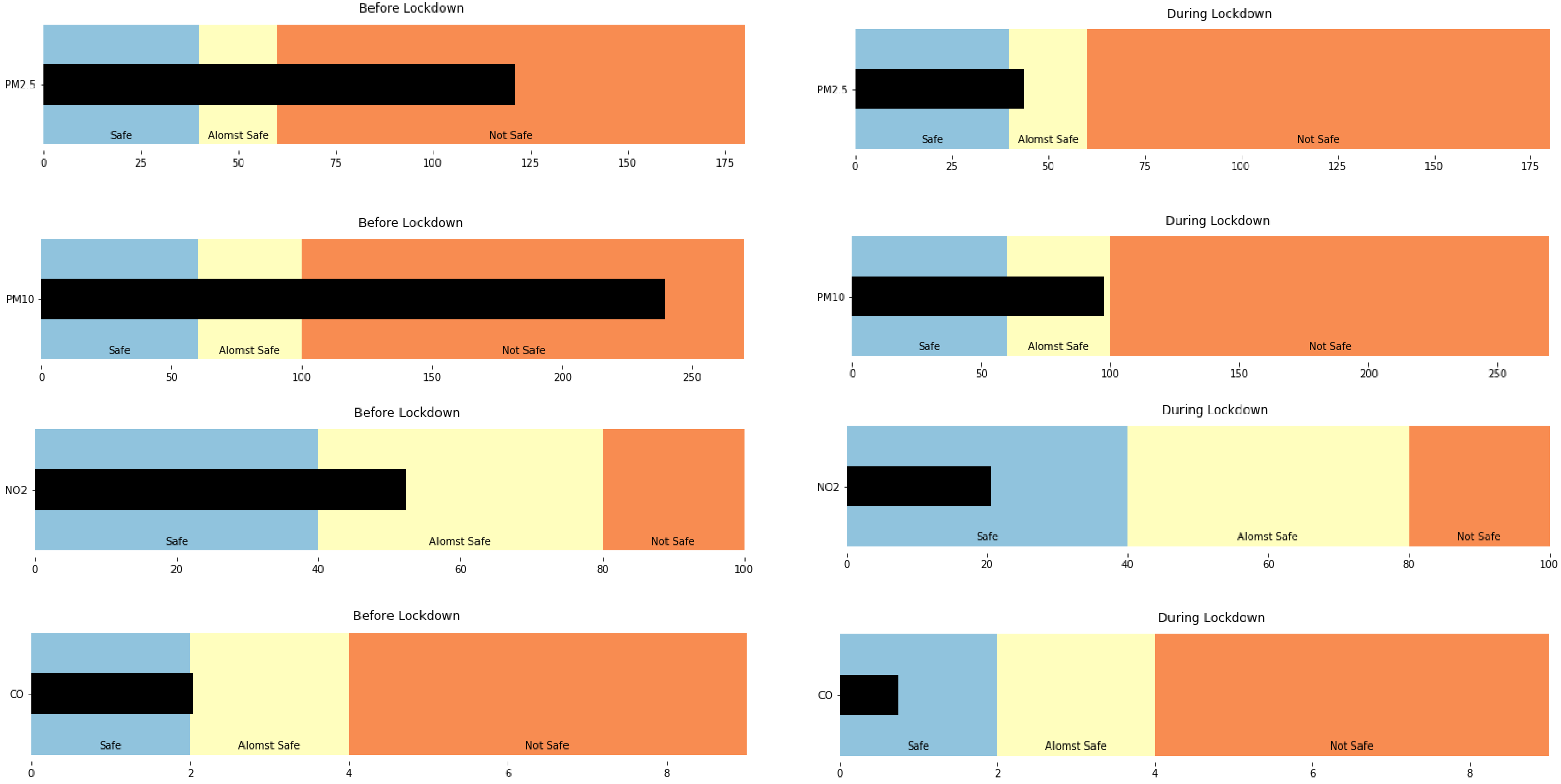
Objective 4: Effect of Lockdown

AQI and pollutant concentrations before and during lockdown



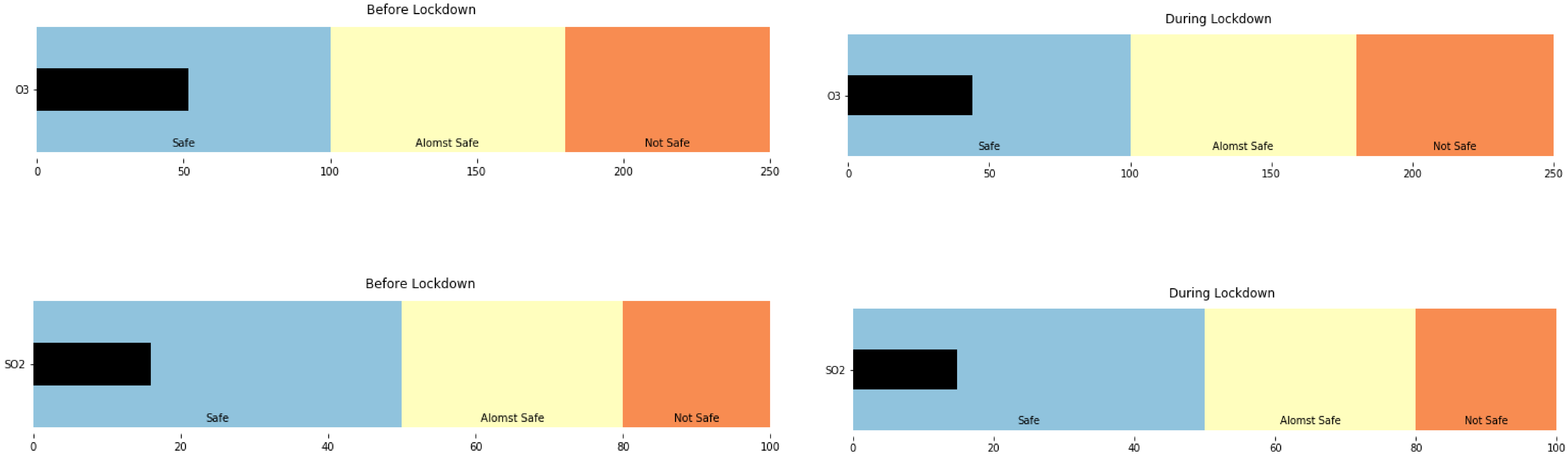
Objective 4: Effect of Lockdown

AQI and pollutant concentrations before and during lockdown



Objective 4: Effect of Lockdown

AQI and pollutant concentrations before and during lockdown



OBJECTIVE 5

Future Air Quality

Predict future concentrations of pollutants and AQI

Methodology: Seasonal Decomposition

- To check if the data has seasonality throughout.
- It also describes the trends in the data.
- The idea here was to use a Time Series Model. We used a multiplicative model for seasonal decomposition.
- It suggests that the components are multiplied together as follows:

y(t) = Level * Trend * Seasonality * Noise

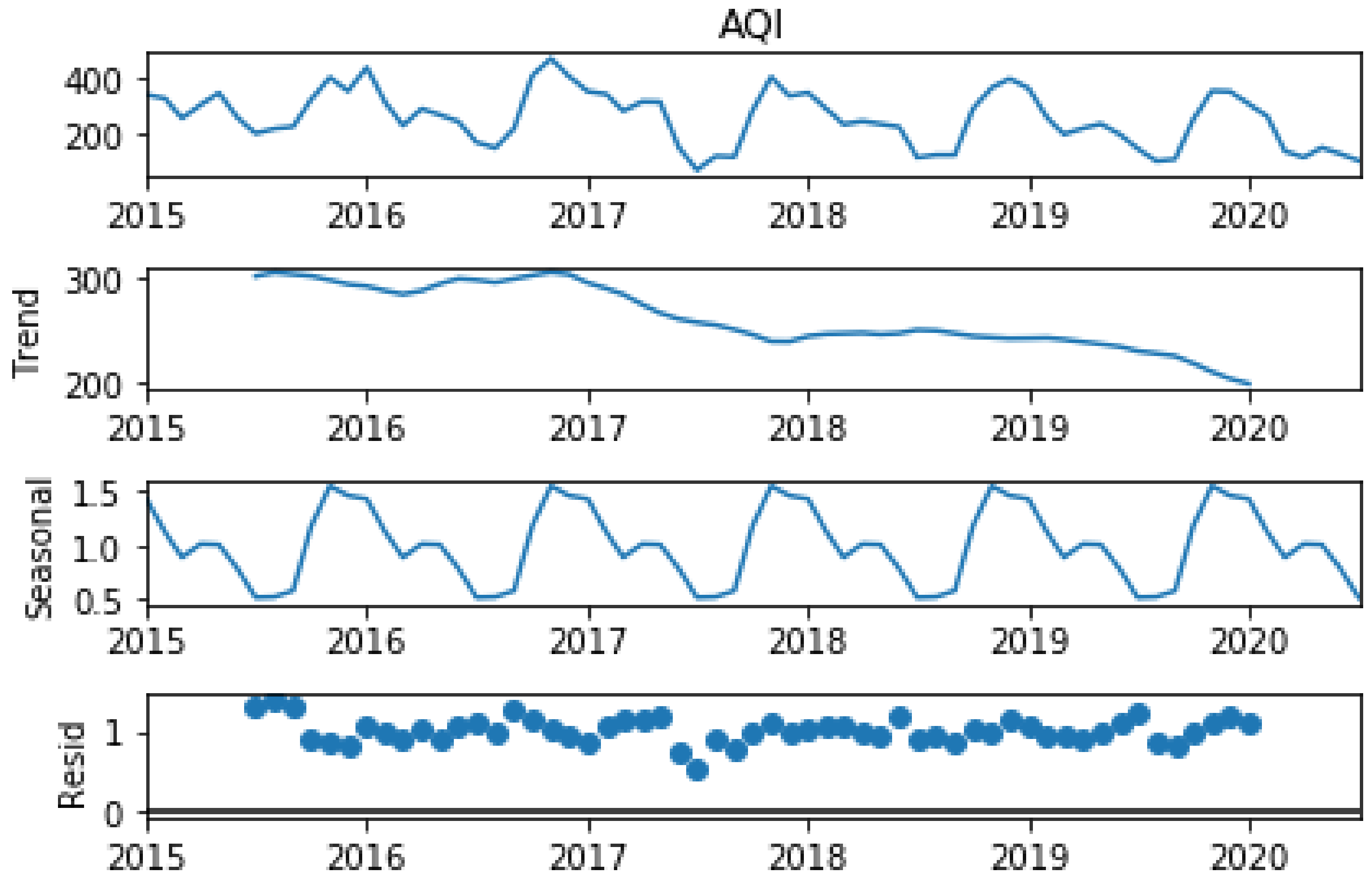
- Nonlinear, such as quadratic or exponential.
- Changes increase or decrease over time.
- **Why not a linear model?** If linear model, dimensions will increase hence complexity of the model will also increase

Objective 5: Future Air Quality

Methodology: Seasonal Decomposition

We identified all required parameter such as:

- **Level:** The average value in the series.
- **Trend**: The increasing or decreasing value in the series.
- **Seasonality**: The repeating short-term cycle in the series.
- **Noise/Residue**: The random variation in the series.



Model 1: LSTM

- Long-short term memory is a common artificial recurrent neural network structure that is often used for time-series predictions.
- It stores past observations in memory, and while training it learns how to use this memory so as to not lose track of longer-term patterns.
- Being as vanilla as they are, some measures need to be taken to account for inconsistencies in data and accounting for random variations.
- Hence, some more configuration is required to set it up as compared to other models like ARIMA, SARIMAX, or Prophet.

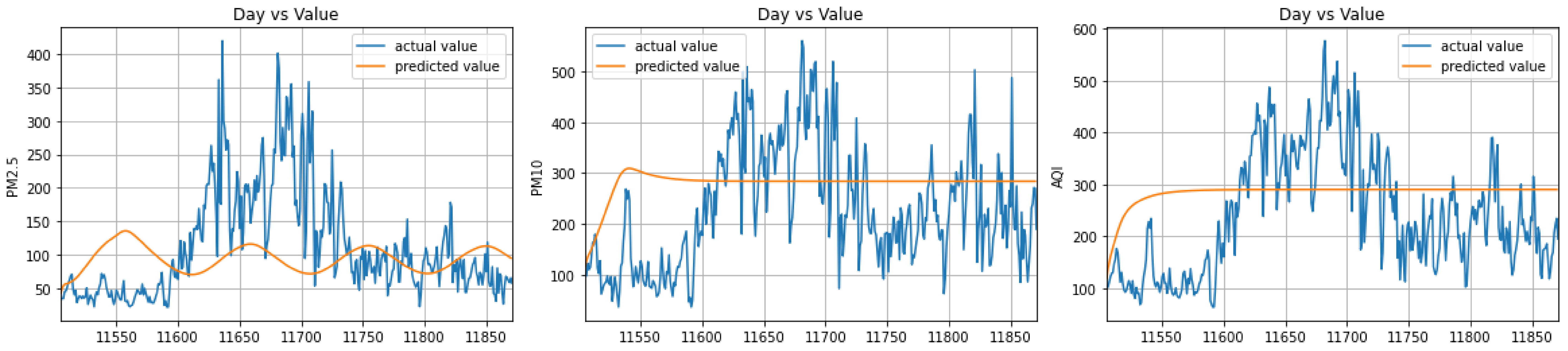
Model 1: LSTM - Handling Missing Values

- In order to fix that, first, we chose to simply eliminate the missing values altogether.
- This way, there are no missing values in the dataset when we run our training or eval loops.
- Quick experiments with PM2.5 concentration did not show a substantial difference, and so we opted to keep it as it was.
- Hence, this method was finalised.

Objective 5: Future Air Quality

Model 1: LSTM - Training and Testing

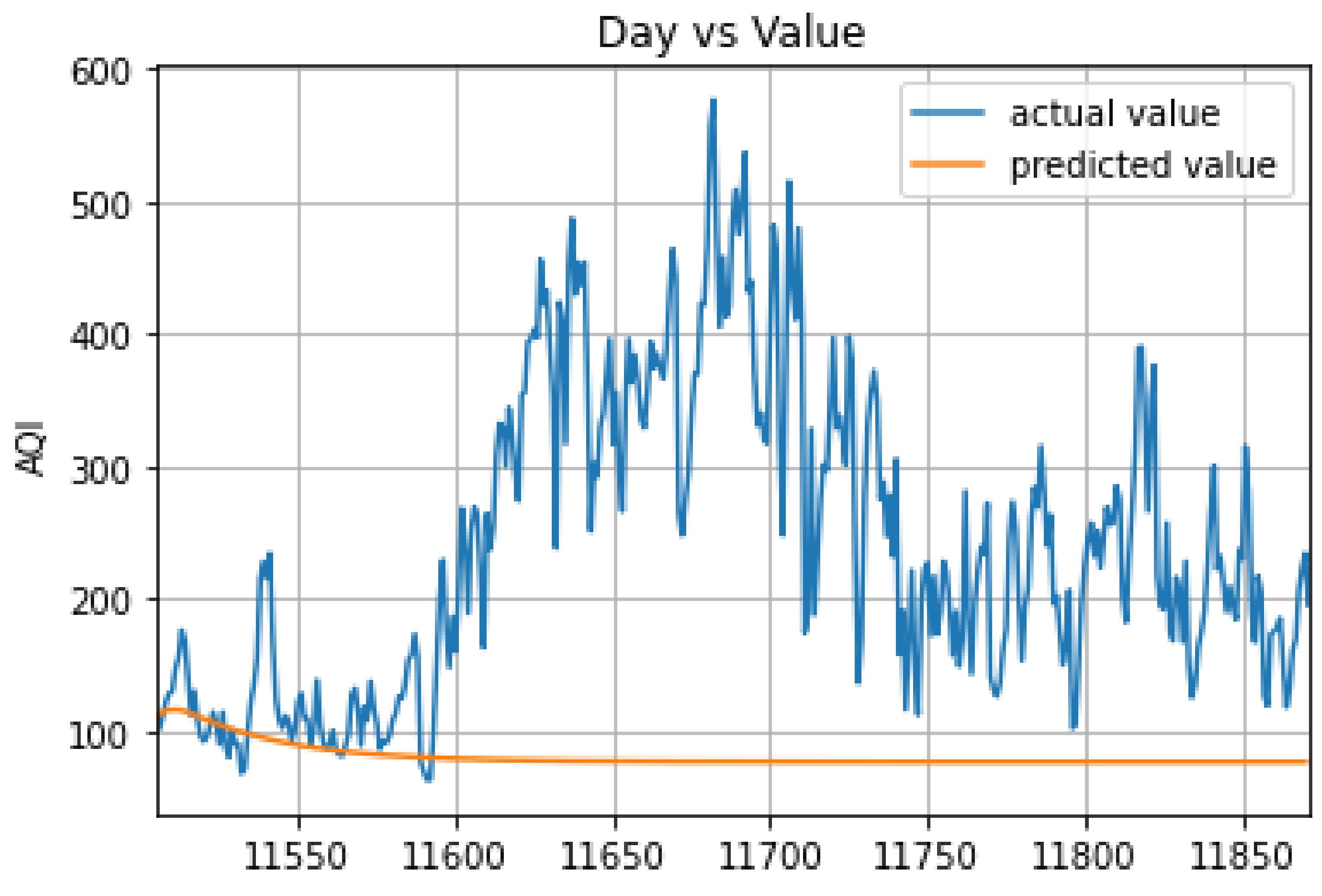
We trained the LSTM model on data from 2015-2018 and tested it on 2019 data to test the accuracy.

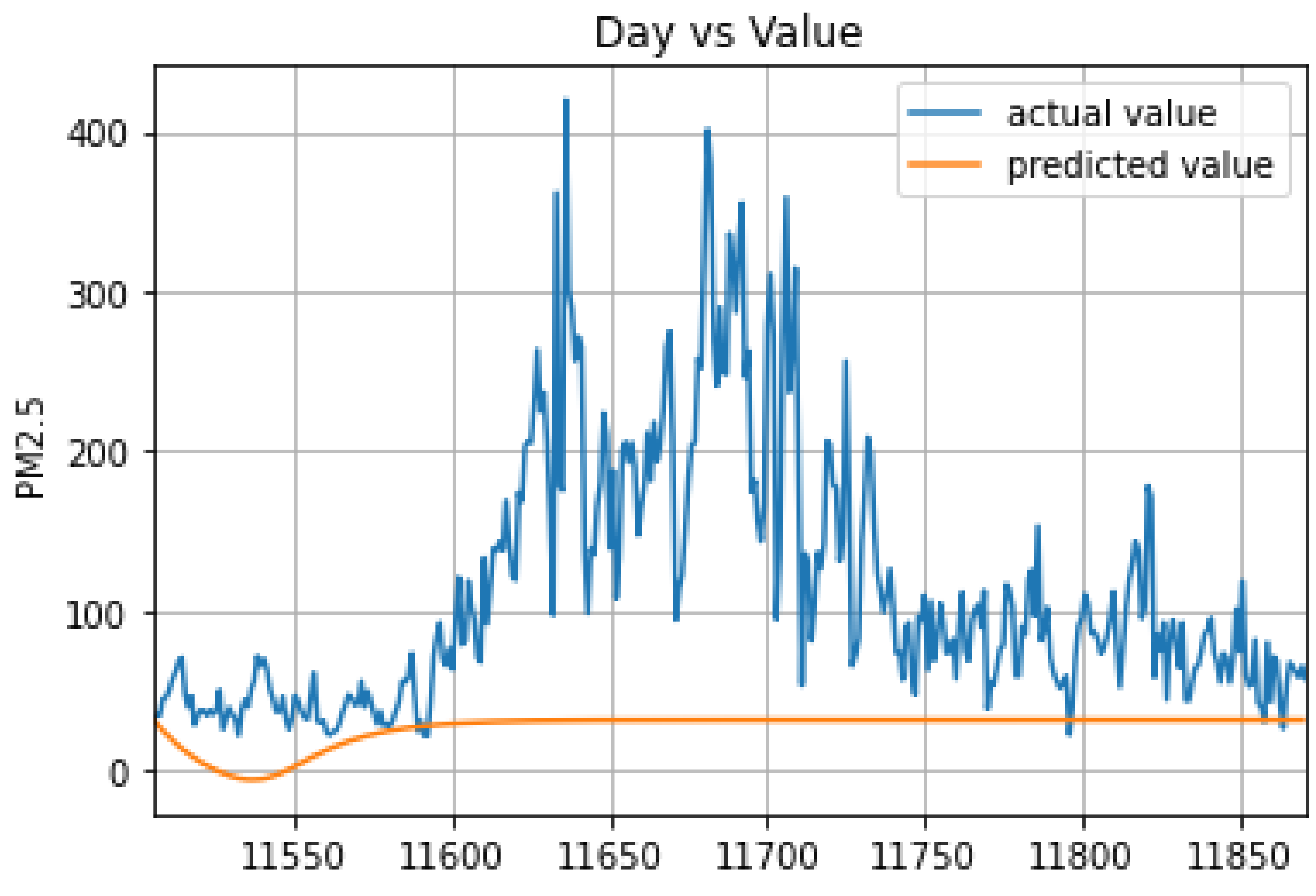


Objective 5: Future Air Quality

Model 1: LSTM - Predicting the unknown: 2021 values

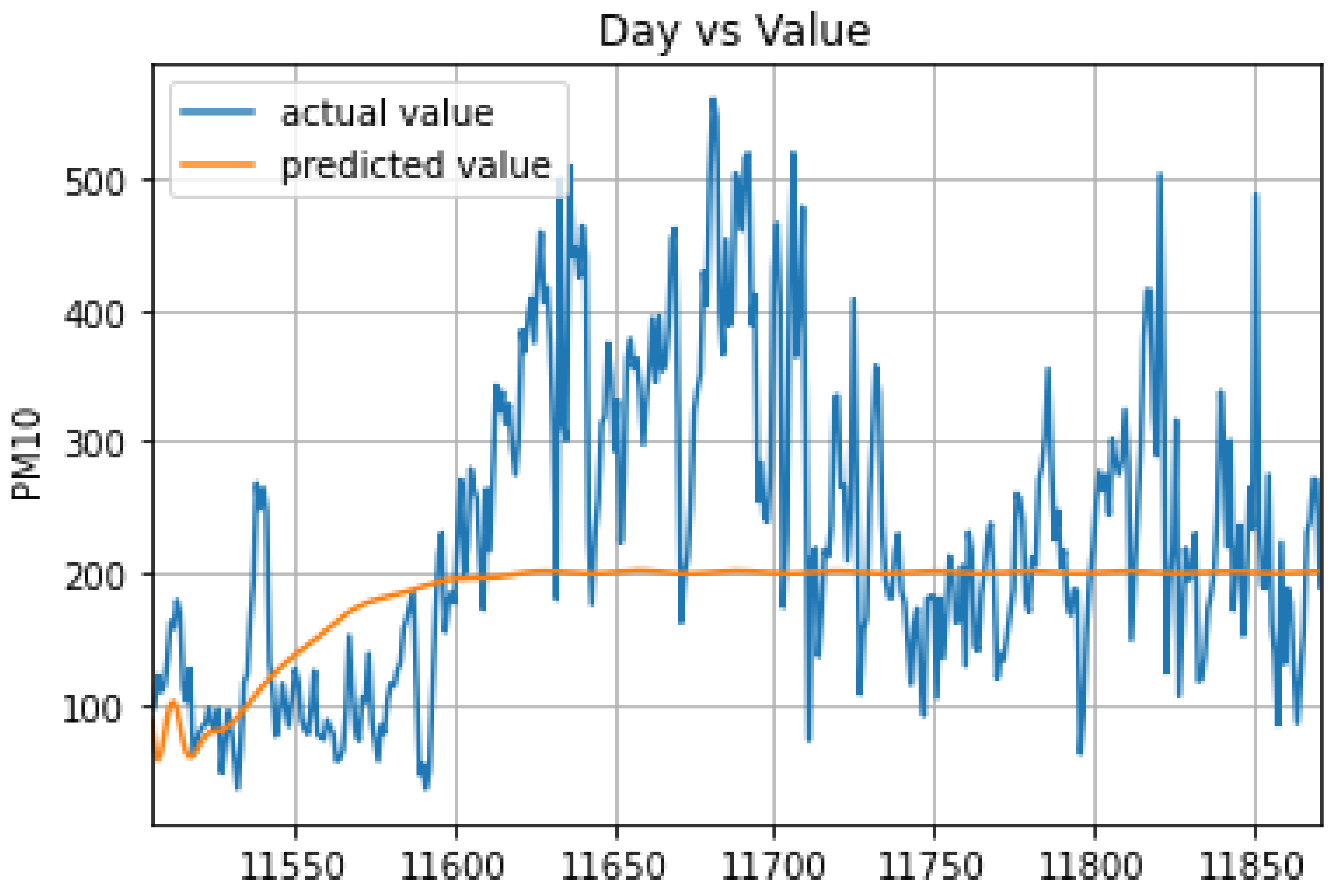
We trained the LSTM model on entire available and predicted the 2021 values

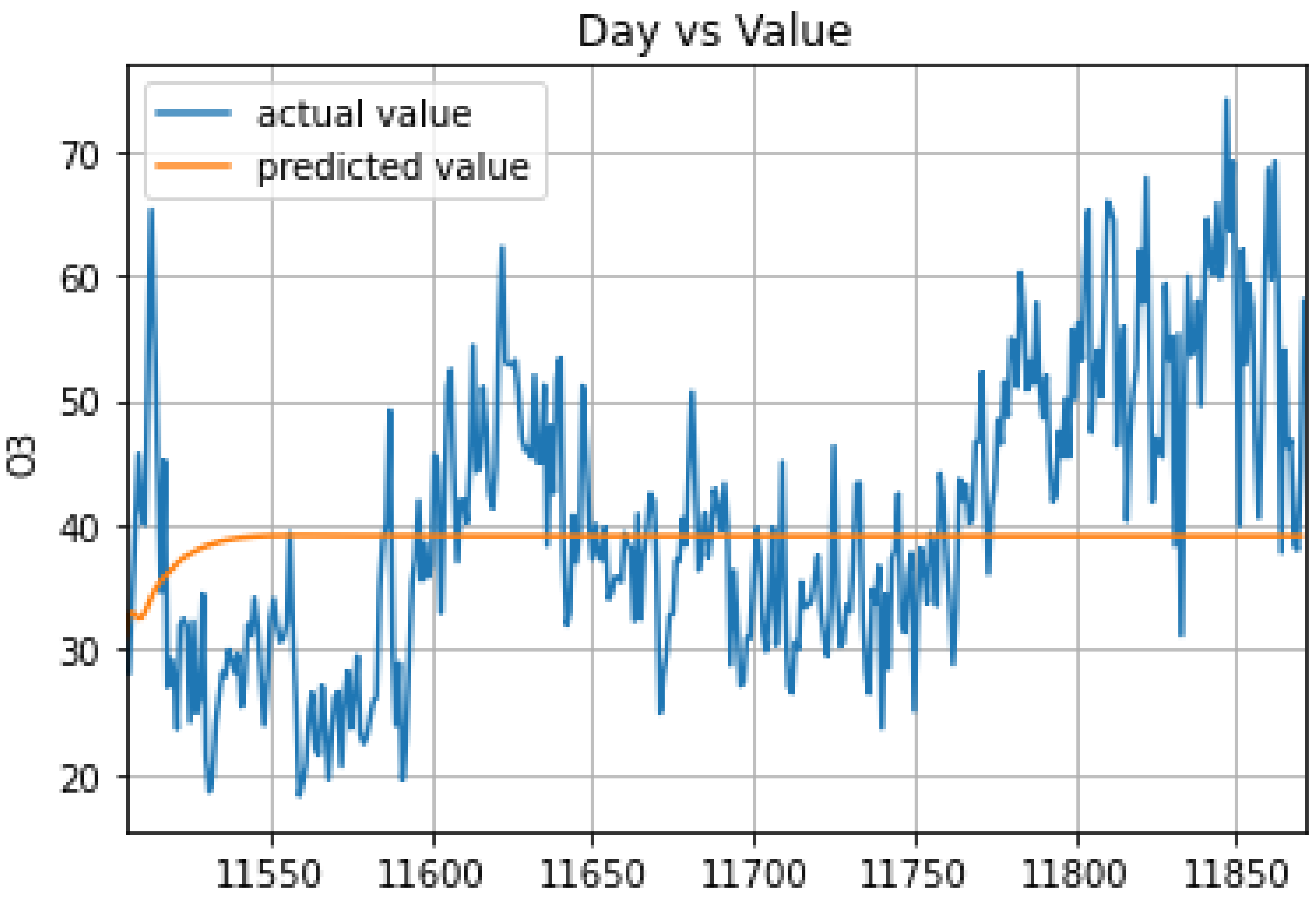




Model 1: LSTM - Predicting the unknown: 2021 values

We trained the LSTM model on entire available and predicted the 2021 values





Model 2: SARIMAX

- Autoregressive Integrated Moving Average, or ARIMA, is one of the most widely used forecasting methods for univariate time series data forecasting.
- Although the method can handle data with a trend, it does not support time series with a seasonal component.
- An extension to ARIMA that supports the direct modelling of the seasonal component of the series is called SARIMA.
- SARIMAX is an extension of the SARIMA model that also includes the modelling of exogenous variables (covariates and can be thought of as parallel input sequences that have observations at the same time steps as the original series).

Model 2: SARIMAX - Handling Missing Values

- Deleting rows with NULL values: The error generated was less but the data got distorted and the AQI values strayed away from the trend.
- Another issue with the approach was that some months did not have any values at all - so when we grouped the data month-wise, it still increased the NULL values - hence the model failed for most of the pollutants.
- So this approach was discarded.

Model 2: SARIMAX - Handling Missing Values

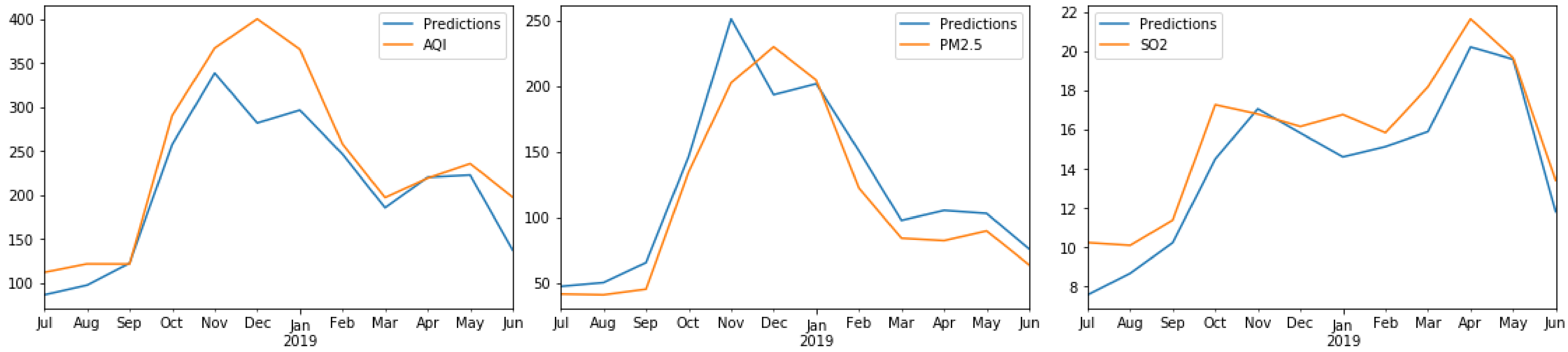
- Filling Missing Values in Database with most frequent values (i.e. Mode)
- ***Filling Missing Values in Database with an average of all the values (i.e. Mean)***
- Filling Missing Values in Database with values separating the higher half from the lower half of a data sample (i.e. Median)

Pros and Cons:

- Prevent data loss which results in the removal of the rows and columns
- Imputing the approximations add variance and bias

Model 2: SARIMAX - Training and Testing to tune parameters

Phase 1: Without considering the effect of lockdown. For this phase, we trained the model on data from 2015-2018 and *tested on 2019 data* to test the accuracy and tune the parameters.

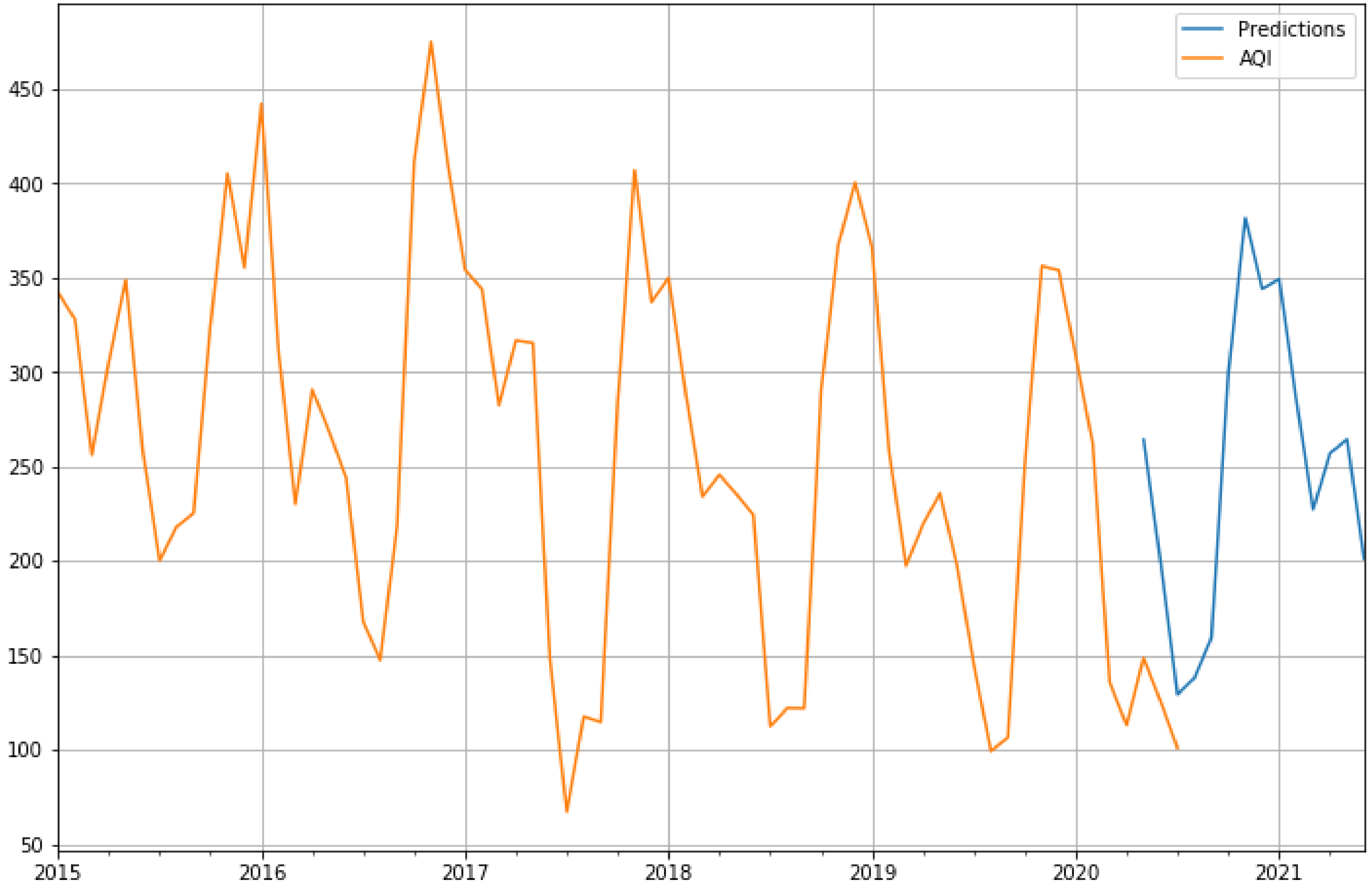


Objective 5: Future Air Quality

Model 2: SARIMAX - Predicting the unknown: 2021 values

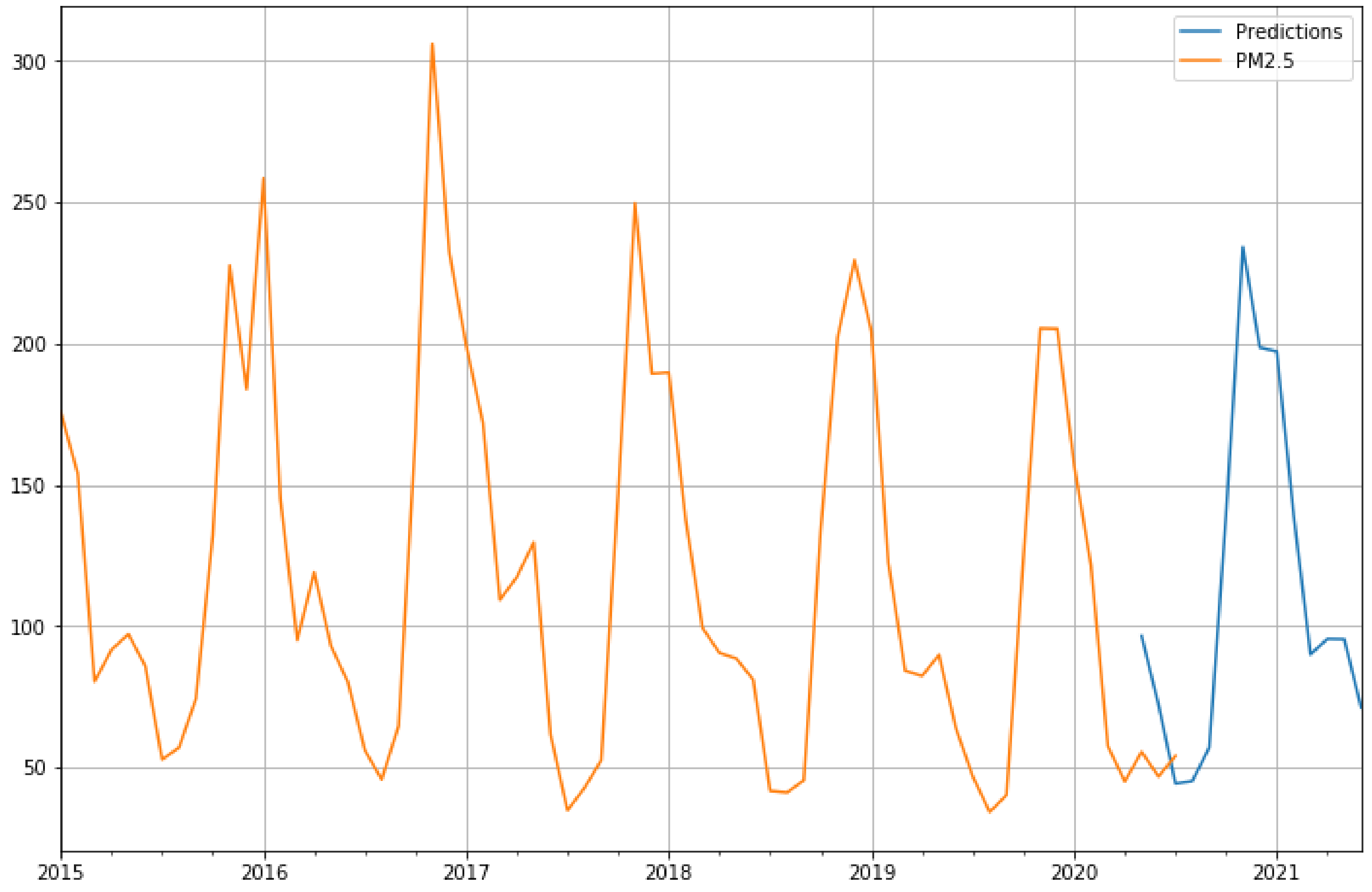
Phase 1: Without considering the effect of lockdown. Trained on 2015-July 2019

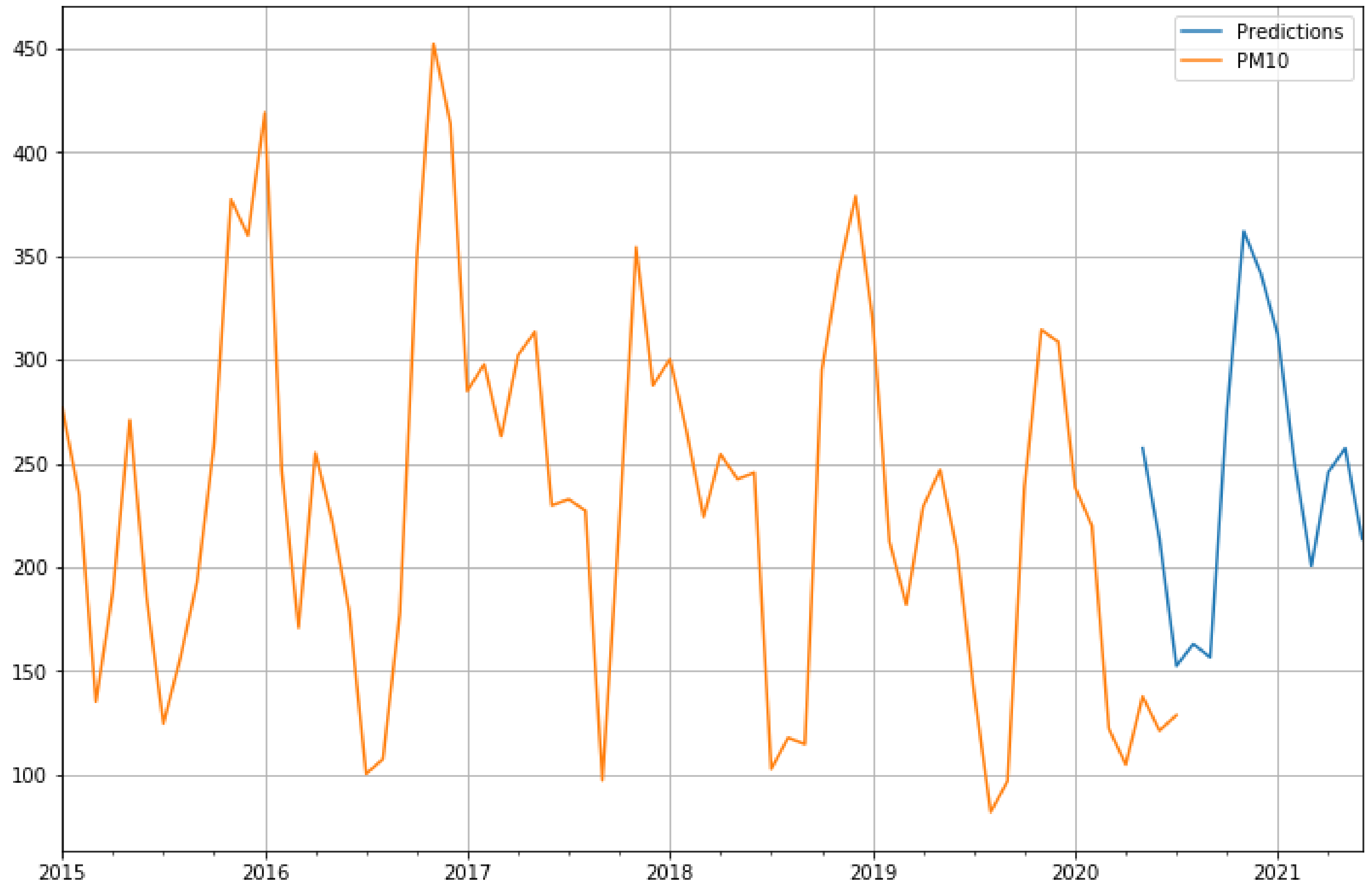
- This plot represents the quality of air if there was no lockdown, to begin with, or if the lockdown conditions were to completely disappear.
- We note that the 2021 prediction is that of very poor air quality i.e. approximately 380 AQI which is on the severe side of the Very Poor category.
- Even the lower bound on the AQI has increased from previous years.



Model 2: SARIMAX - Predicting the unknown: 2021 values

Phase 1: Without considering the effect of lockdown. Trained on 2015-July 2019

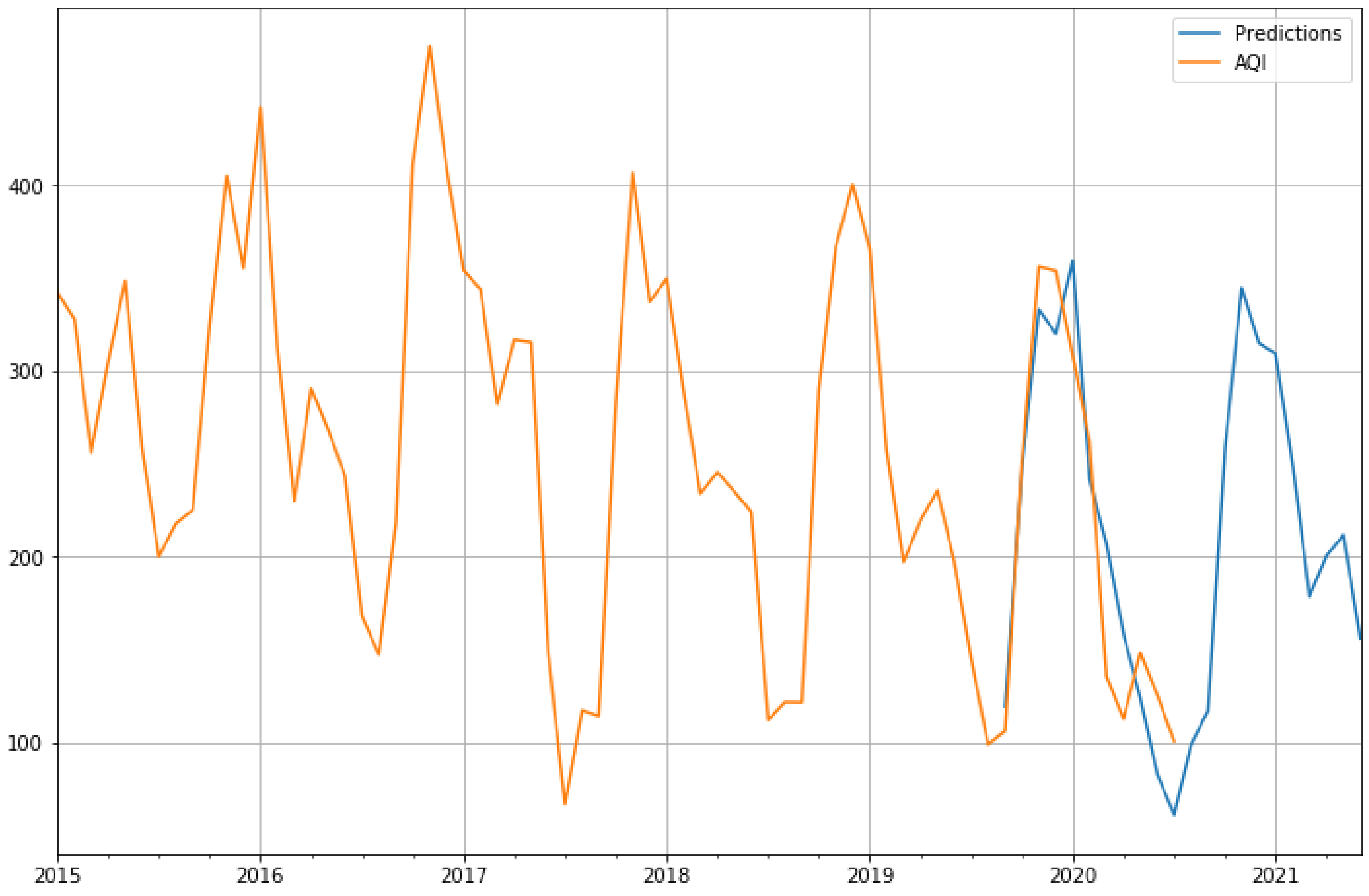




Model 2: SARIMAX - Predicting the unknown: 2021 values

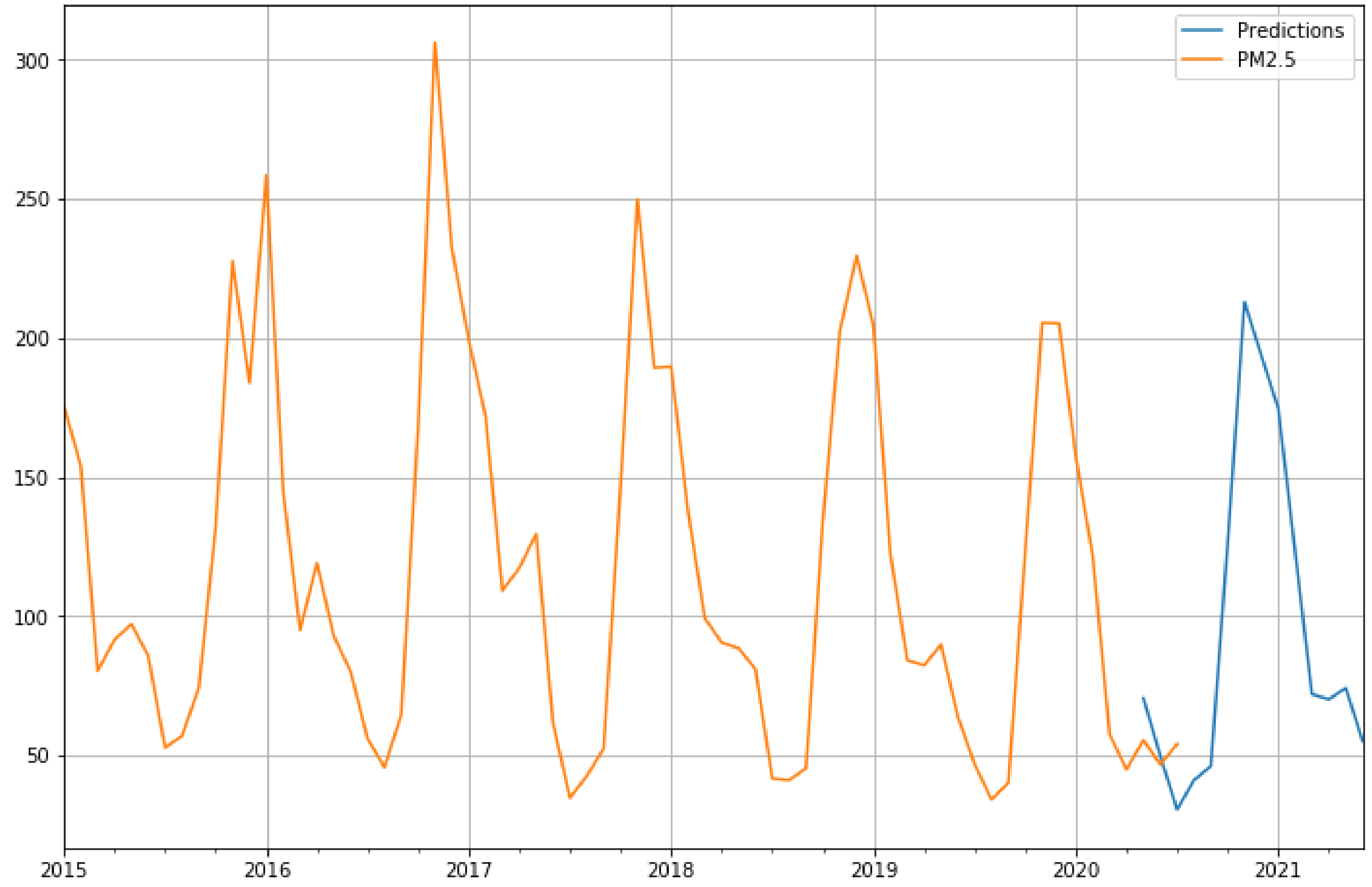
Phase 2. Considering the effect of lockdown.
For this phase, we trained the model on data from 2015-2020 and predicted the unknown values for 2020-2021.

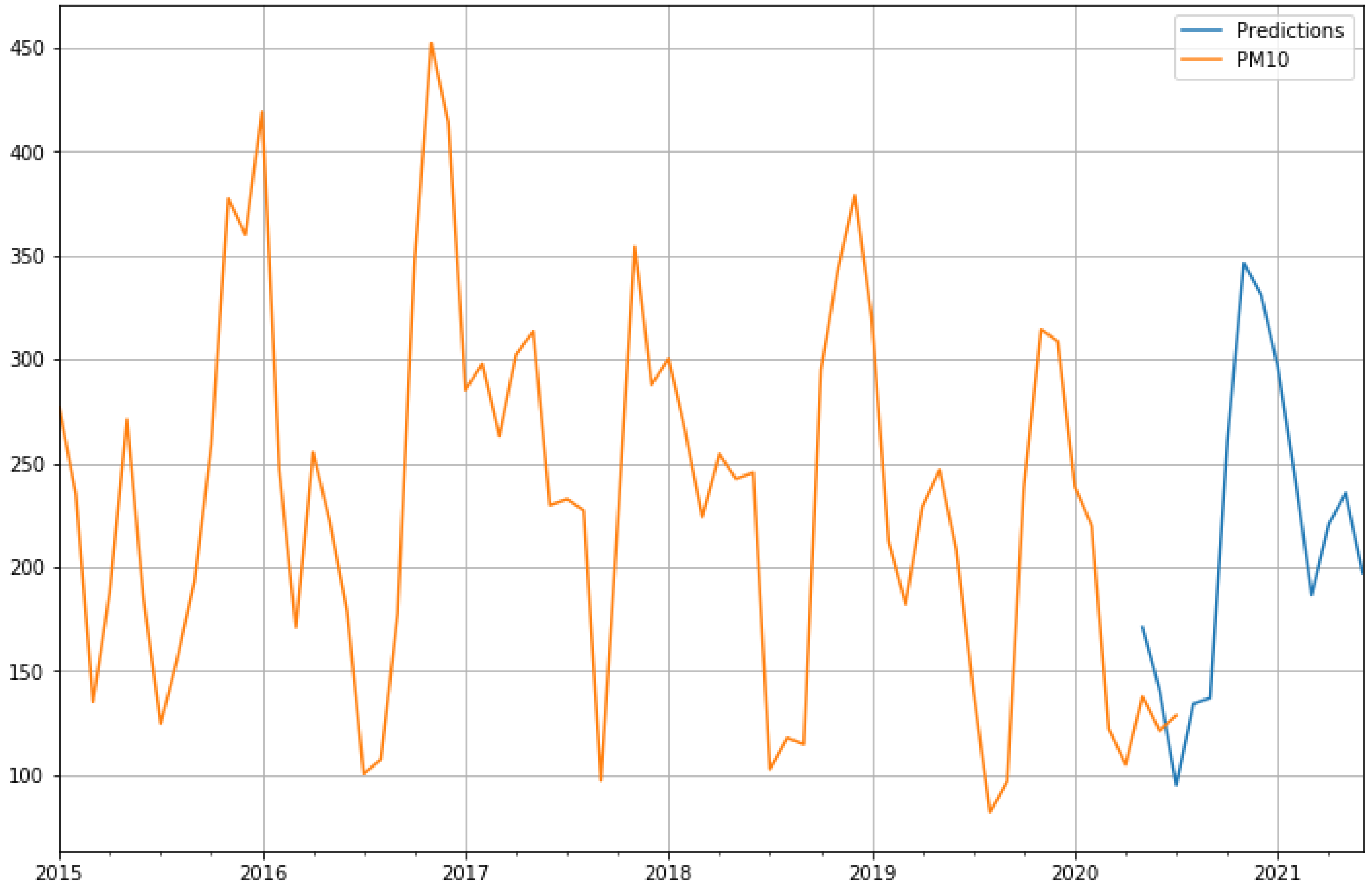
- This plot represents the quality of air if the nature of human activities remains the same as it was during the lockdown.
- We note that the 2021 prediction has improved, however, still lies on the poor side of the AQI.
- Even the lower bound on the AQI have decreased a lot from previous years.



Model 2: SARIMAX - Predicting the unknown: 2021 values

Phase 2. Considering the effect of lockdown. For this phase, we trained the model on 2015-2020





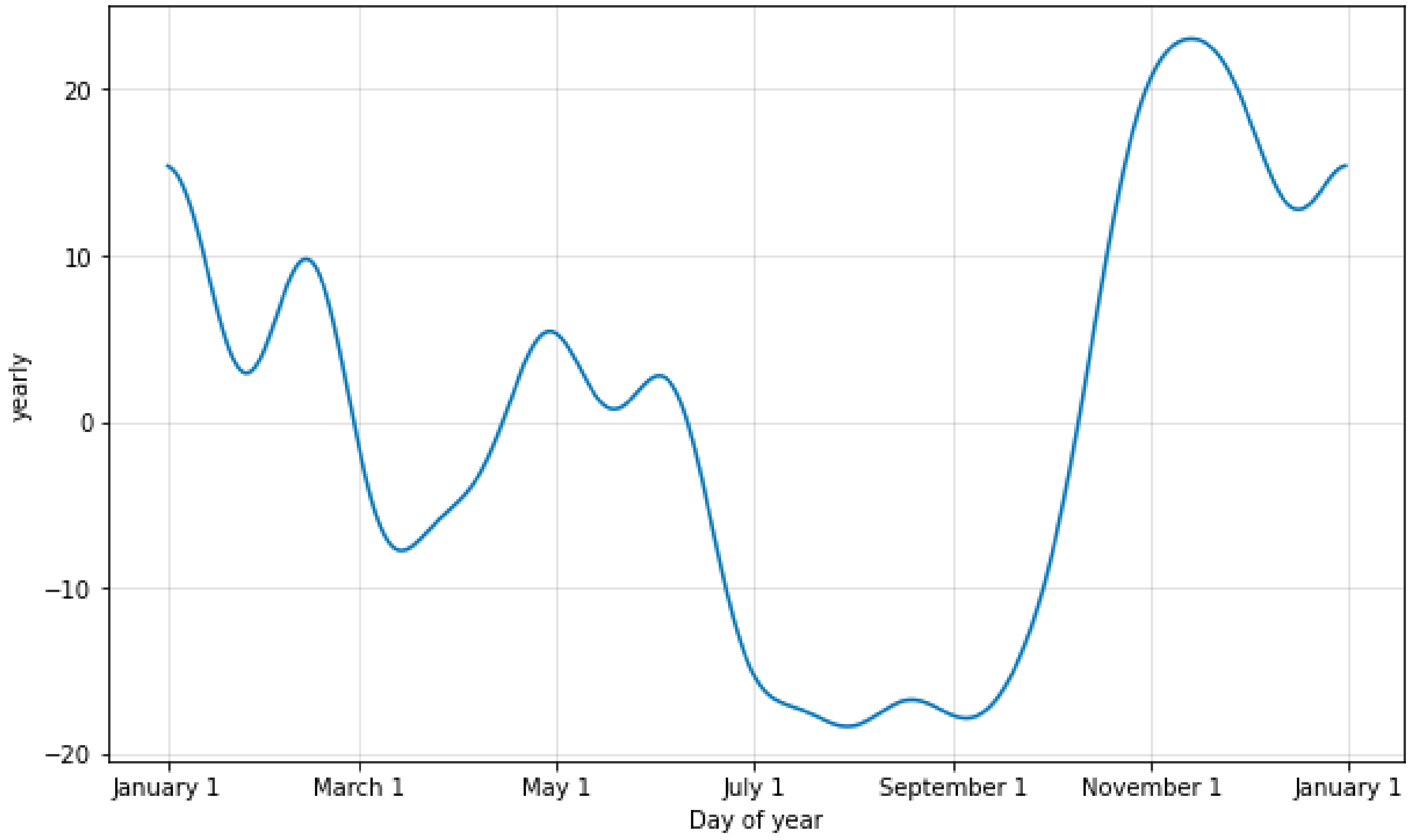
Model 3: Prophet

- Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects.
- It works best with time series that have strong seasonal effects and several seasons of historical data.
- Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.
- It uses a simple, modular regression model that often works well with default parameters.

Note: We do not need to impute the data since Prophet understands there are missing points and fills the gaps as it trains.

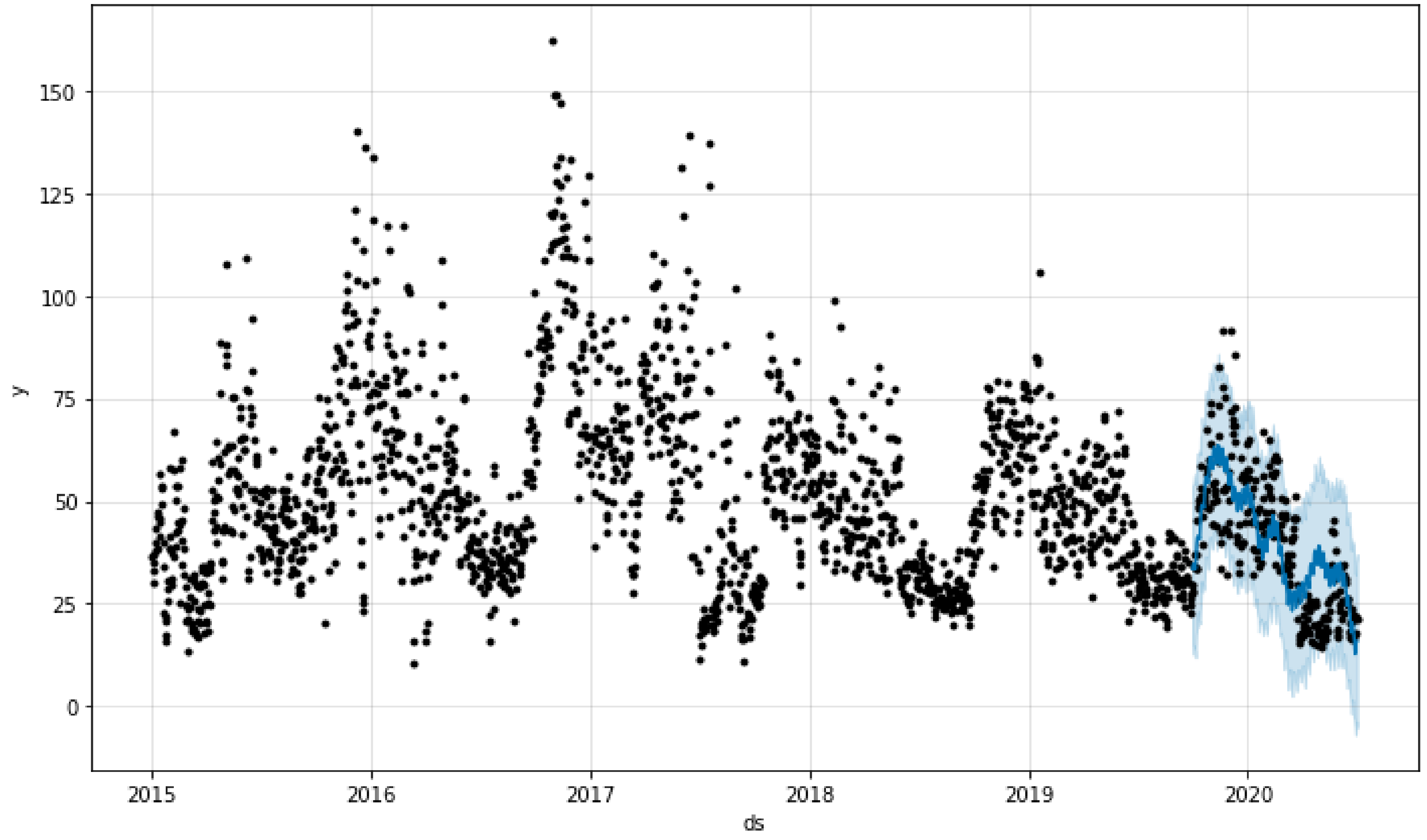
Model 3: Prophet: Training and Testing to tune parameters

- **In-sample forecast:** Even though we predict values for the last 12 months, we pass the whole data for the training process.
- Ideally, the model has seen the data before and would make a perfect prediction. Nevertheless, this is not the case as the model tries to generalize across all cases in the data.
- The plot shows the seasonal component identified by Prophet



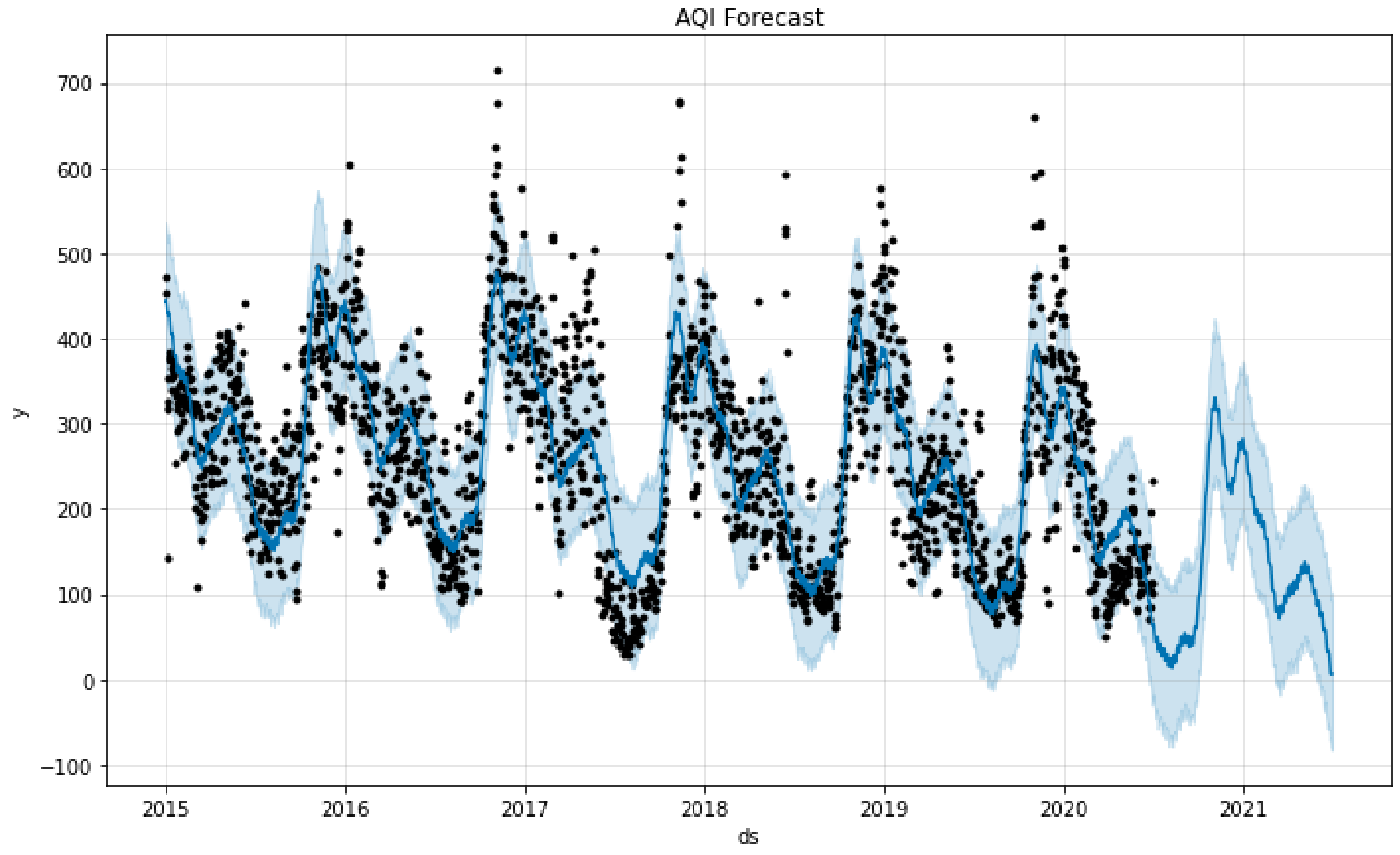
Model 3: Prophet - Training and Testing to tune parameters

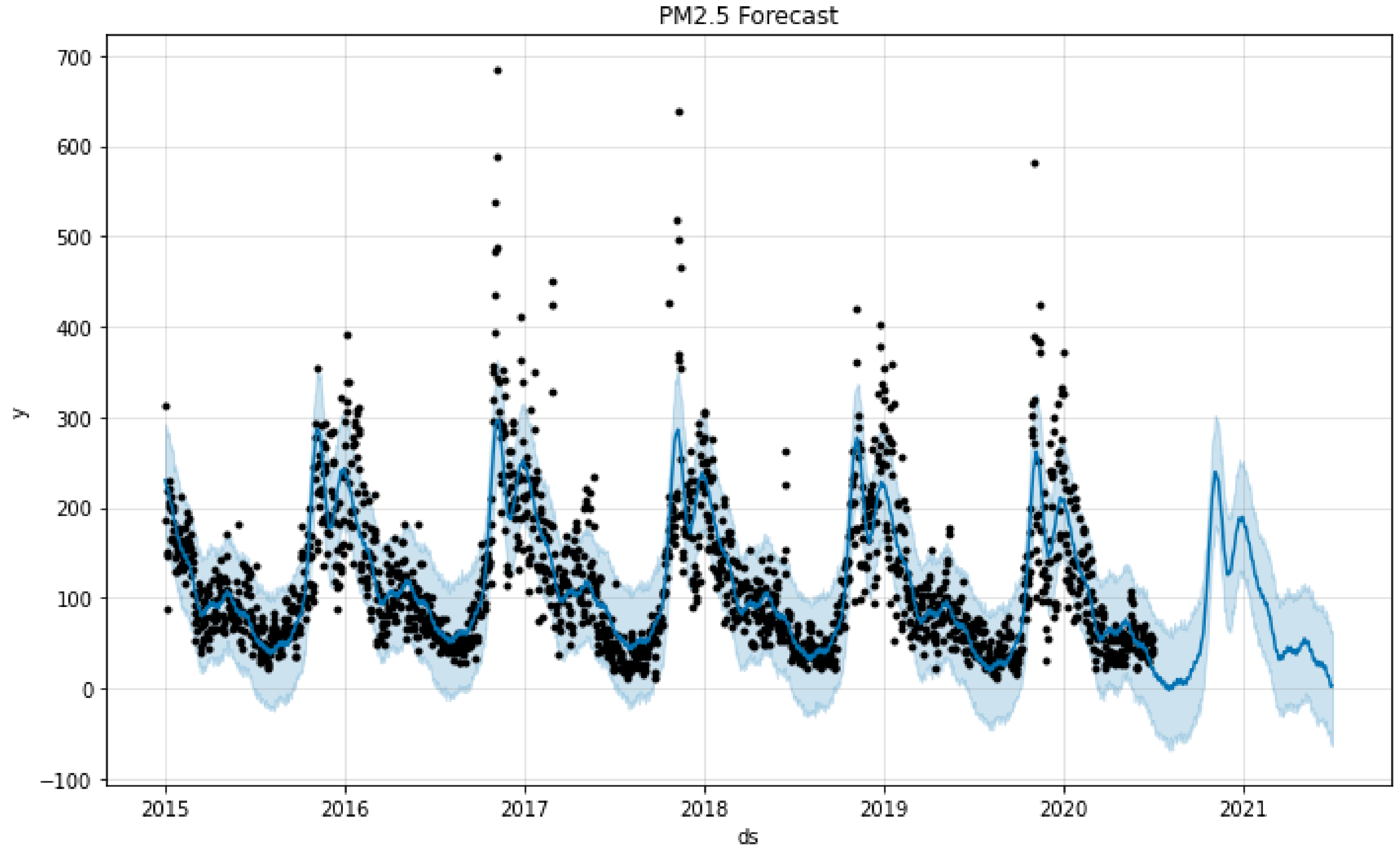
- The model was trained on data from 2015 to 2020
- Predicted values for 2019-2020
- The black points represent the actual values.
- The blue line represents the predicted values.
- The light blue shading represents the range of values.



Model 3: Prophet - Predicting the unknown: 2021 values

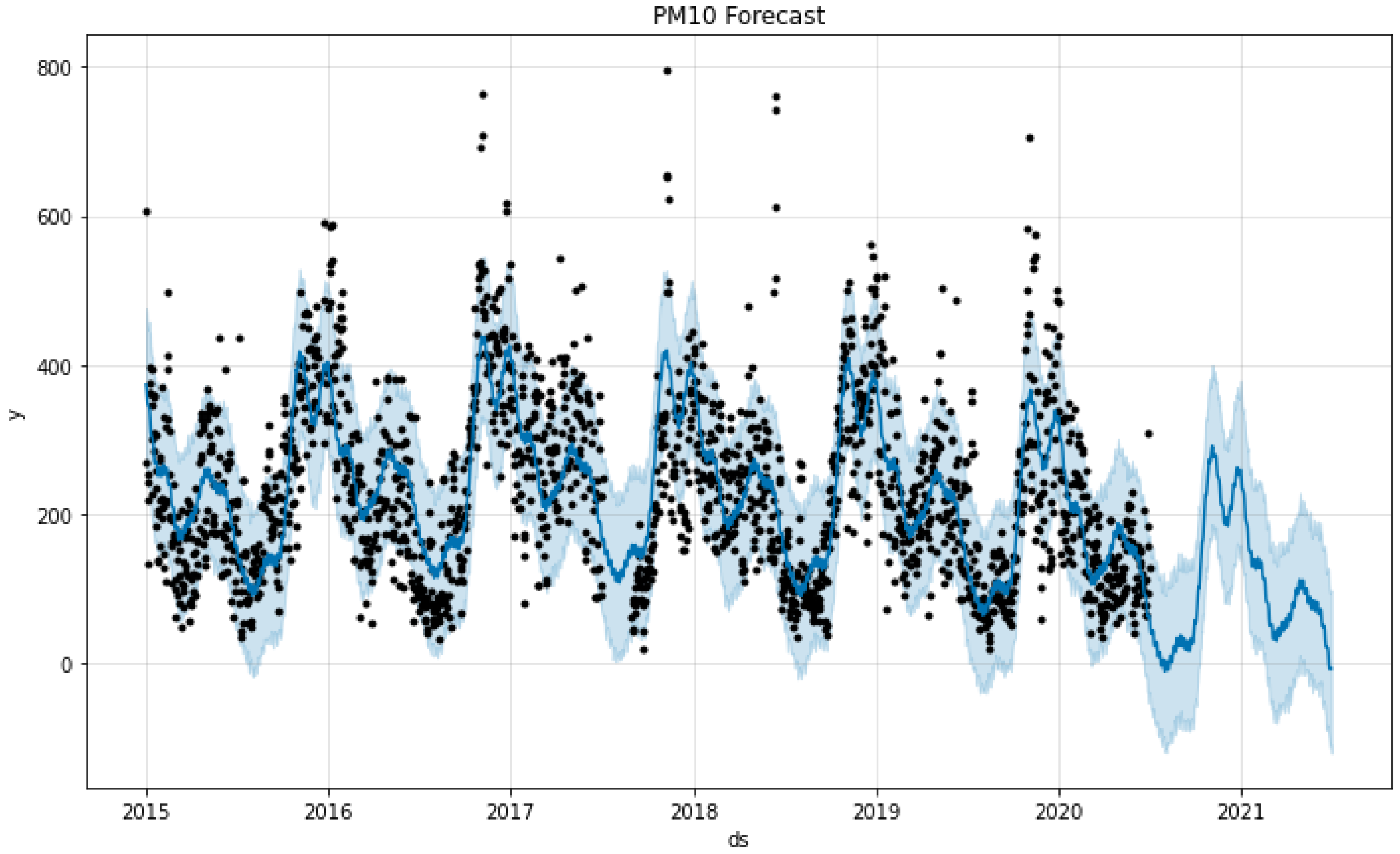
- **Out-of-sample forecast:** Predicting values beyond our dataset

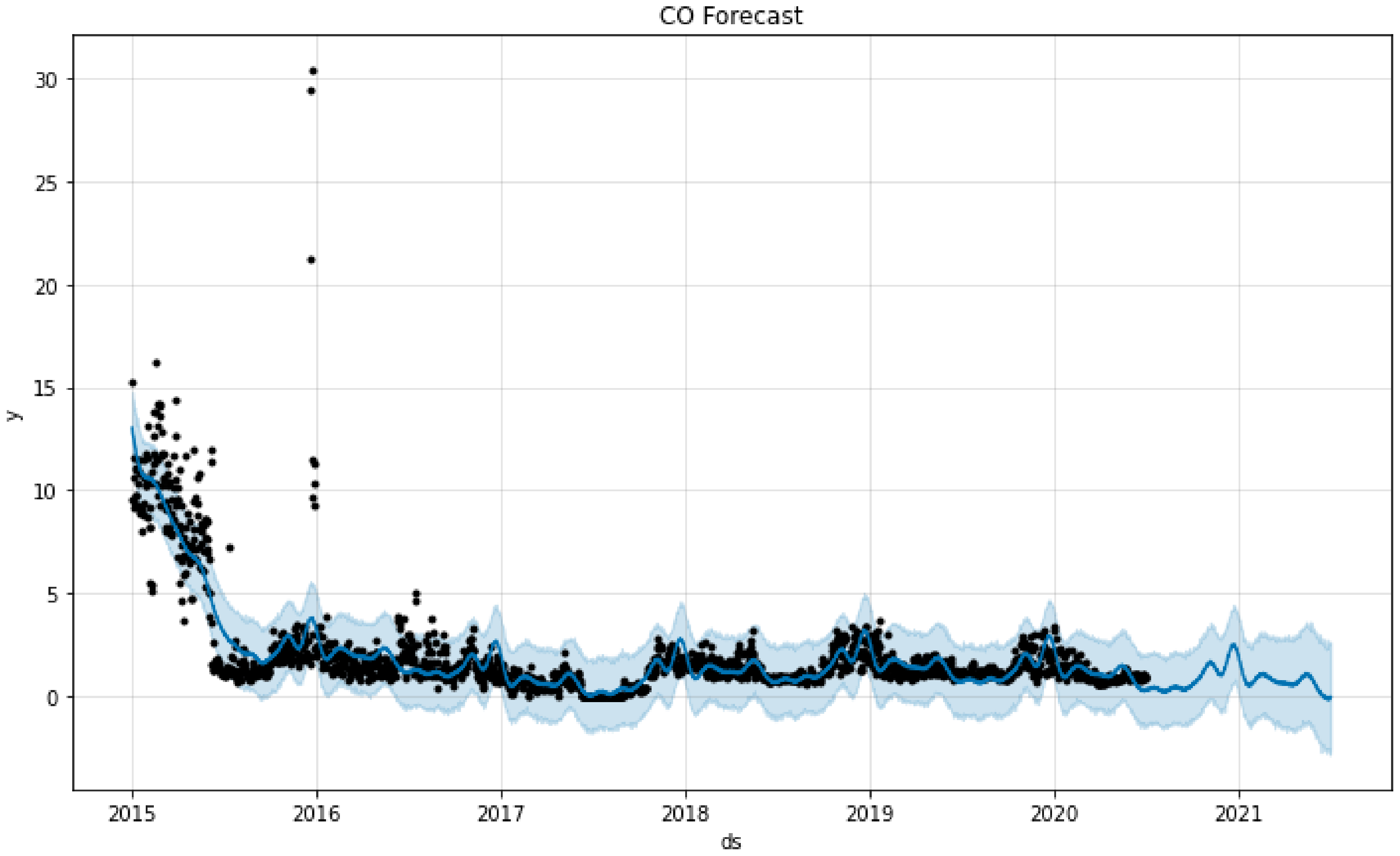




Model 3: Prophet - Predicting the unknown: 2021 values

- **Out-of-sample forecast:** Predicting values beyond our dataset

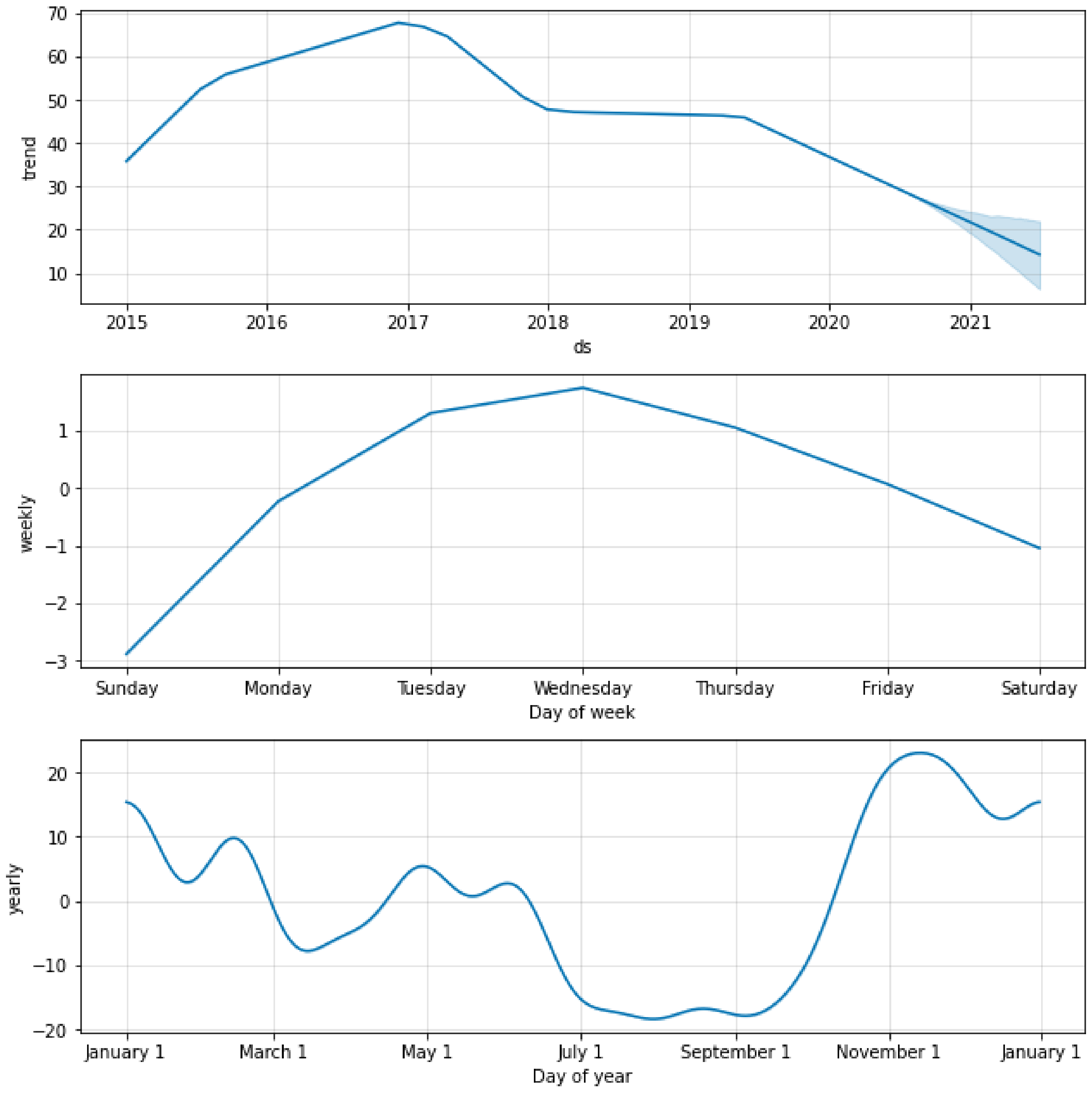




Objective 5: Future Air Quality

Model 3: Prophet - Predicting the unknown: 2021 values

- Weekly seasonality, Yearly Seasonality and trend predicted by Prophet for Delhi



DISCUSSION AND CONCLUSION

Discussion

- LSTMs provide the worst result of the lot, probably due to poor hyperparameters and their generalised nature.
- We can see that SARIMAX and Prophet trade blows when it comes to performance on predicting the future data.
- This performance does not seem to vary significantly across categories from the 2018-19 data, and the 2019-20 data: SARIMAX performs better than Prophet on PM2.5, SO2, and CO, while the reverse is true for PM10, O3, AQI. The only value that switches around is NO2.

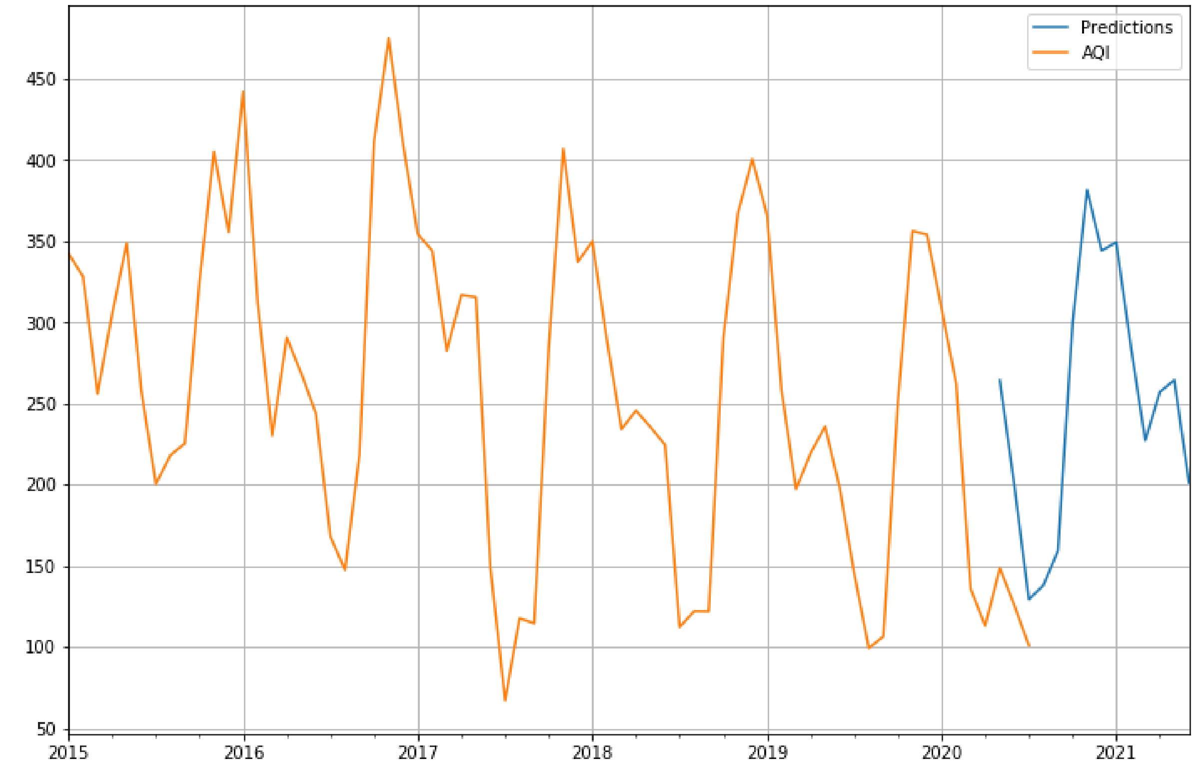
RSME 2019-2020	SARIMAX	LSTM	Prophet (s=2000)
PM2.5	28.744024821660677	123.07858102806442	33.030065352871304
PM10	76.52449137021613	138.53707223468663	55.526877872613944
NO2	11.286551147549964	28.806266720461227	9.054781512949596
SO2	4.3720068036053545	12.596179453761032	3.180200406436058
CO	0.2939873443351936	0.7343326671575078	1.3013884855513544
O3	24.952767741928778	18.713872628622738	14.218913942476874
AQI	69.23578918181191	201.1118009411132	47.014884350286245

RSME 2018-2019	SARIMAX	LSTM	Prophet (s=2000)
PM2.5	22.732492085387353	88.73265959287475	18.539922803525208
PM10	39.106712968850374	120.70019656609585	37.83781766940039
NO2	5.984225553581485	19.130738016389728	6.47750859505361
SO2	1.6676166027082113	14.521046452483802	1.770003651146963
CO	0.7235001719782407	0.7196393589289877	0.8667088674269894
O3	17.7873256051094	15.879392381766214	5.832394777954752
AQI	46.525126697298184	135.46684959995457	31.910006075802748

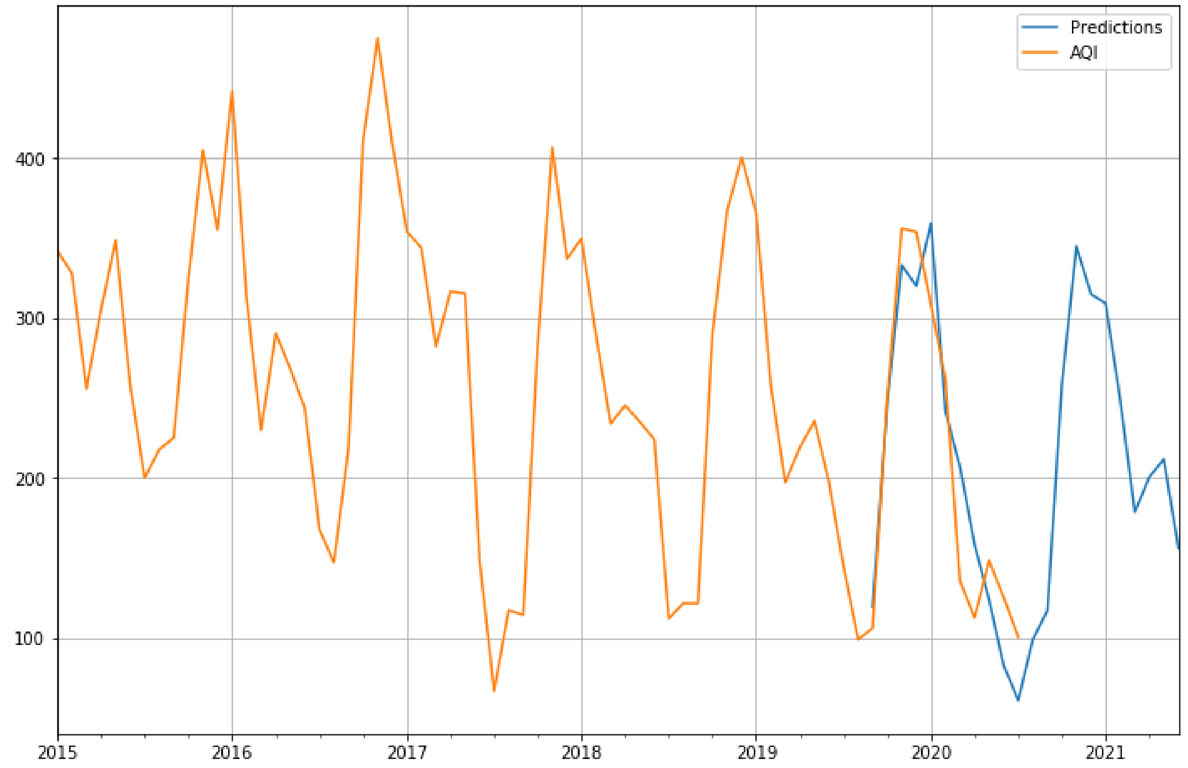
Conclusion

- Lockdown has impacted the pollution levels of the environment to improve the air quality in a short span owing to very few human activities. The measures taken to ward off the threat of the virus are virtually identical to the measures that climate activists have been demanding for decades: less travel, less work and less environmental expropriation.

Without lockdown's effect

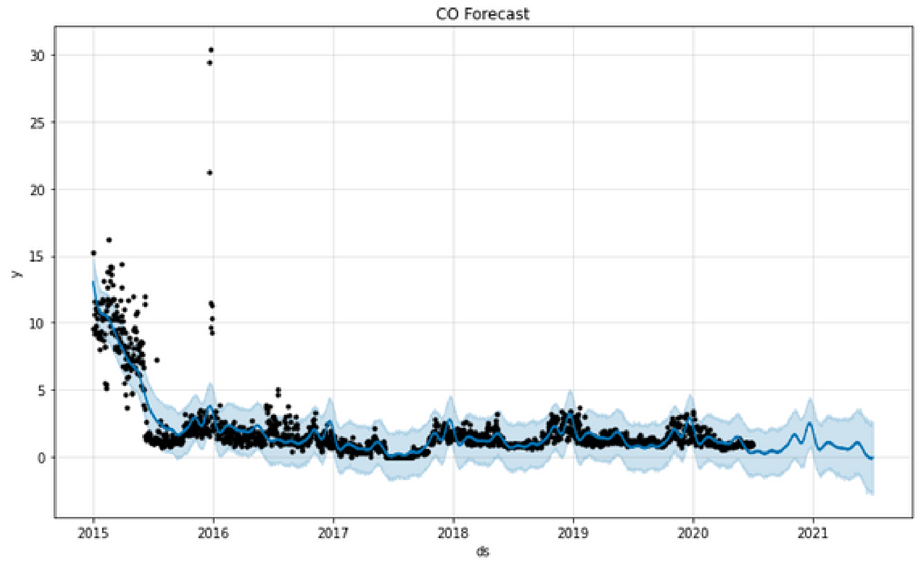


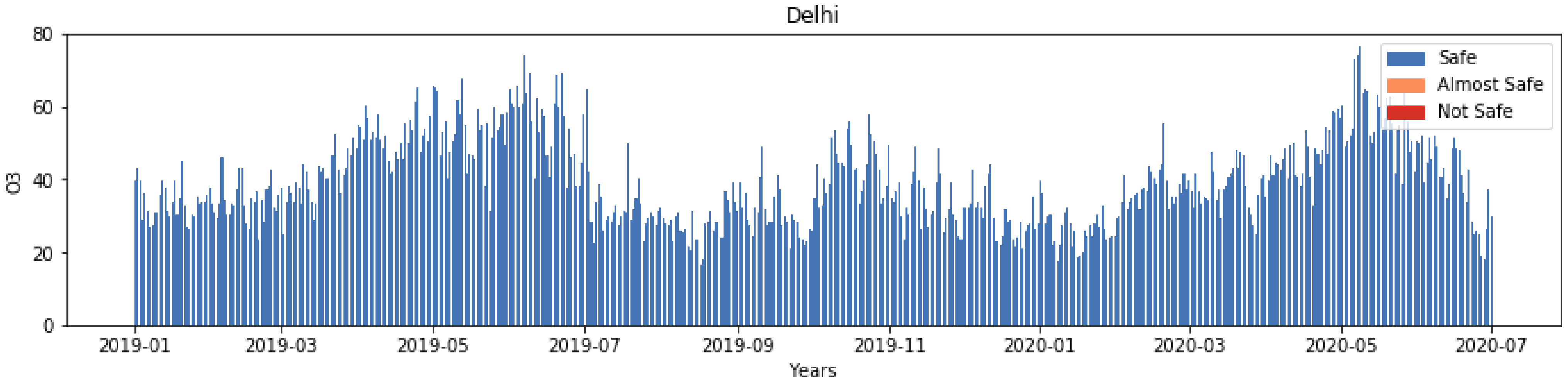
With lockdown's effect



Conclusion

- For pollutants that have seen almost constant levels, the error is low and the prediction is also similar and mostly lies in the safe category.





- It is also evident how poor the quality of our air is. With or without lockdown, the concentrations of harmful pollutants like PM2.5 and PM10 remain far too high for comfort.
- The AQI tends to the high side of 200 or more, often reaching above 400 at times which lies in the severe category, where it is recommended that “good air quality” be below 50 to ensure people have no troubles or health issues that arise from just breathing.

Thank you! Questions?

Harshita Sharma
Abhigyan Ghosh
Zubair Abid