



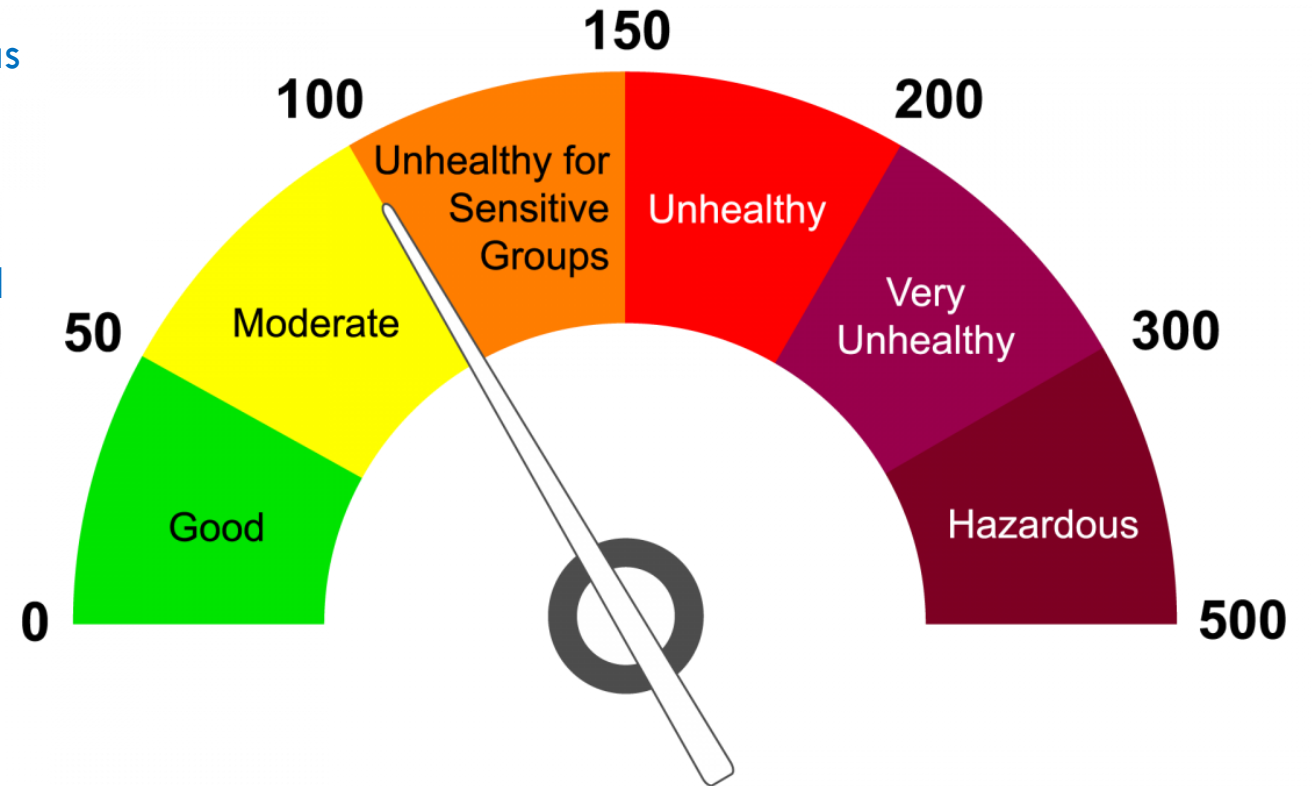
AIR QUALITY INDEX ANALYSIS & FORECASTING : BEFORE / AFTER COVID

Time Series Forecasting
Chiranjit Pathak

AIR QUALITY INDEX

An air quality index is used by government agencies to communicate to the public how polluted the air currently is or how polluted it is forecast to become. Public health risks increase as the AQI rises. Different countries have their own air quality indices, corresponding to different national air quality standards.

Air is what keeps humans alive. Monitoring it and understanding its quality is of immense importance to our well-being.





PROBLEM STATEMENT

Based on **AQI** of some cities in India, how **COVID-19** has influenced the **air quality** and how it can be forecasted post **COVID-19** using time series modeling technique?

DATA LOADING AND DESCRIPTION

AQI calculation:

1. The Sub-indices for individual pollutants at a monitoring location are calculated using its **24-hourly average concentration value (8-hourly in case of CO and O₃) and health breakpoint concentration range**. The worst sub-index is the AQI for that location.

2. All the eight pollutants may not be monitored at all the locations. **Overall AQI is calculated only if data are available for minimum three pollutants out of which one should necessarily be either PM_{2.5} or PM₁₀**. Else, data are considered insufficient for calculating AQI. Similarly, a minimum of 16 hours' data is considered necessary for calculating sub index.

Source of the data:

<https://www.kaggle.com/rohanrao/air-quality-data-in-india>

Acknowledgements:

The data has been made publicly available by the Central Pollution Control Board:
<https://cpcb.nic.in/> which is the official portal of Government of India.

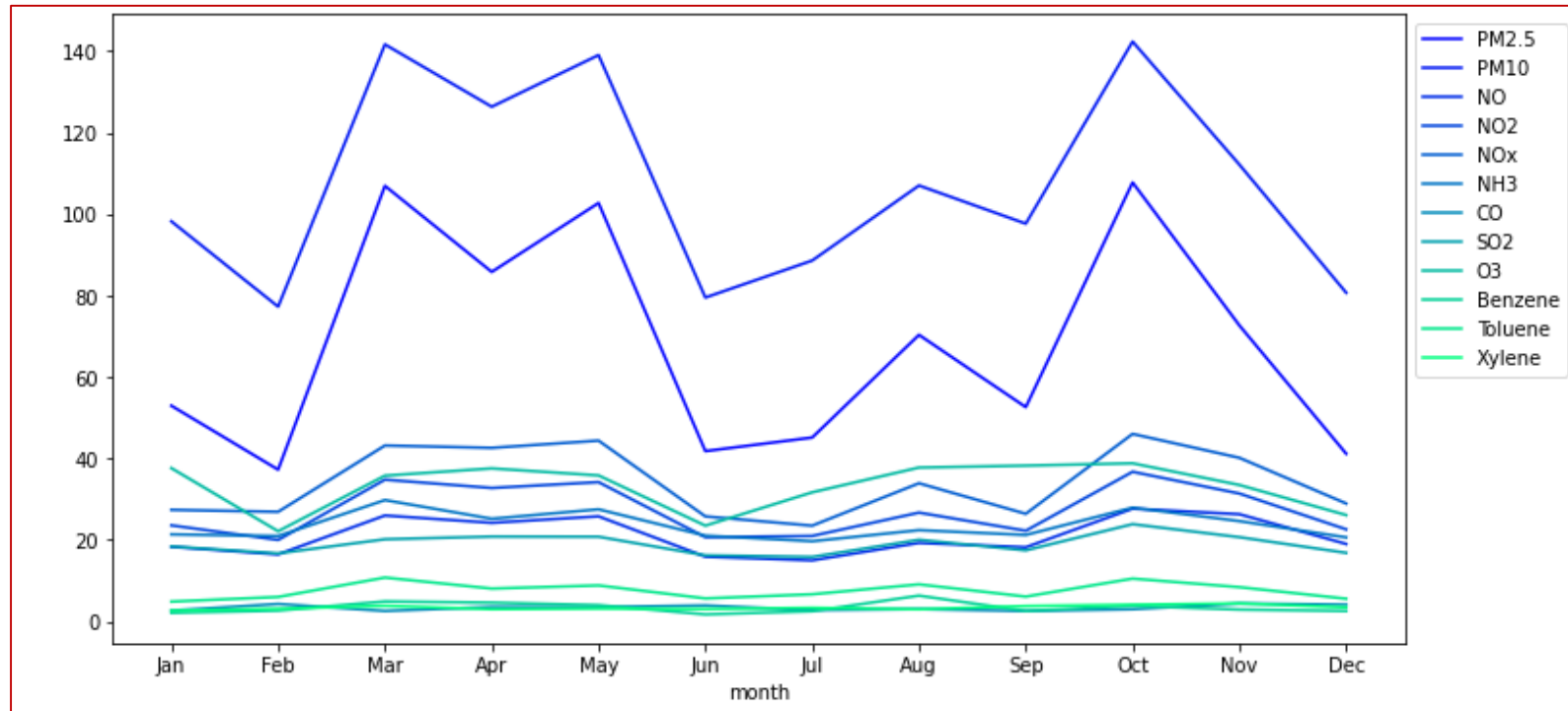
They also have a real-time monitoring app:
https://app.cpcbccr.com/AQI_India/



EXPLANATORY DATA ANALYSIS

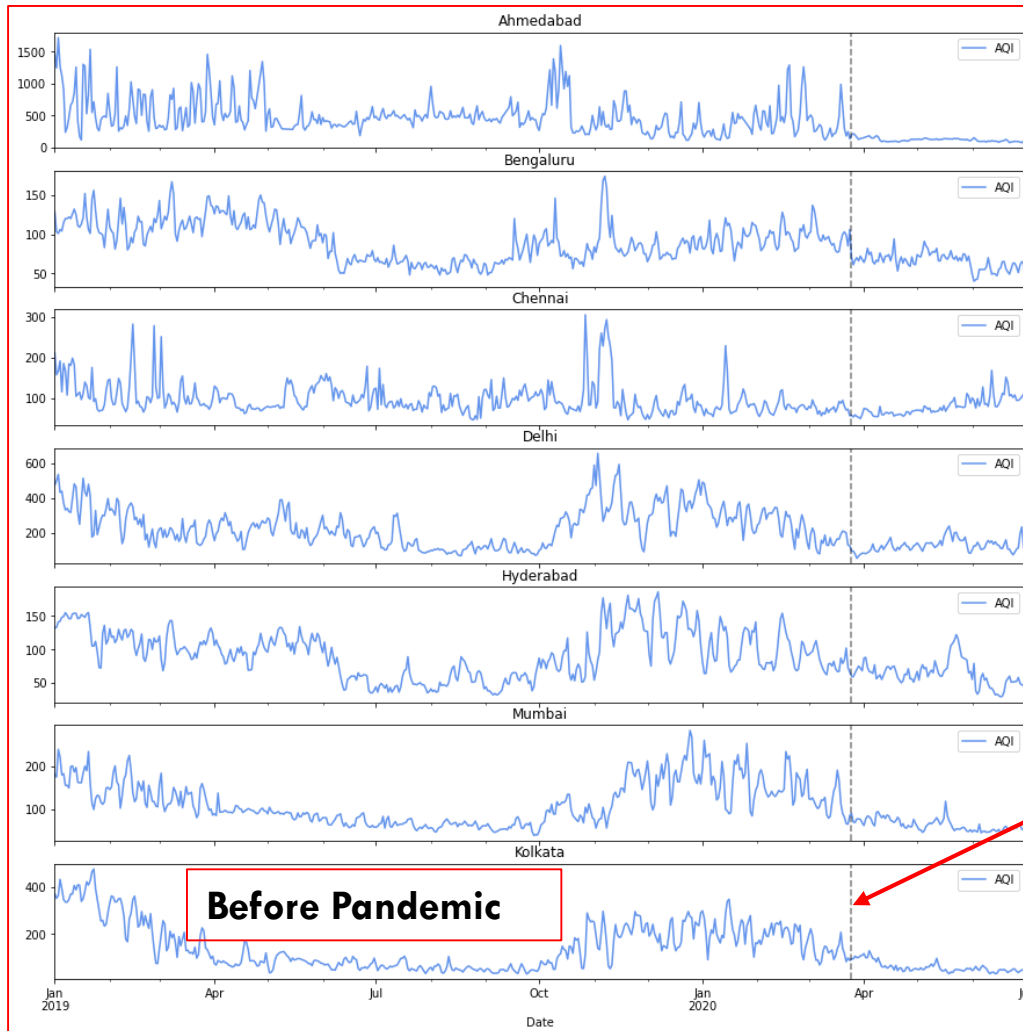
**Processing, Profiling and
Analysis**

PROCESSING AND EDA



The mean Concentration of basic component in air across the time period of 2015-2019 and how they are varied over months in each year.

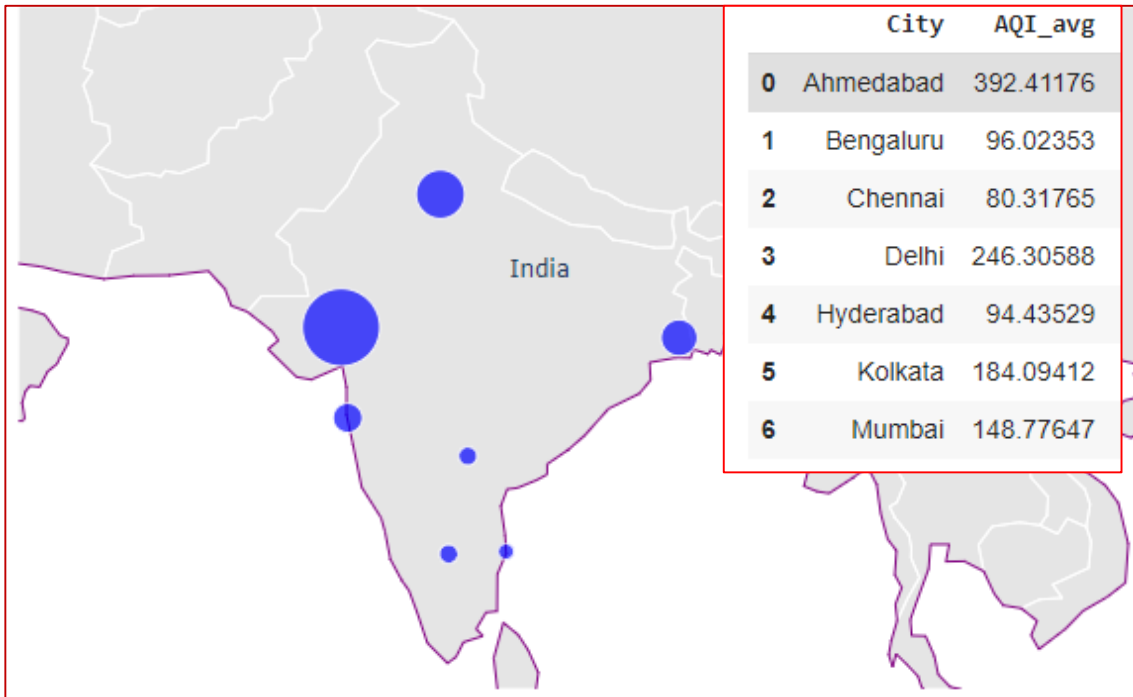
PROCESSING AND EDA



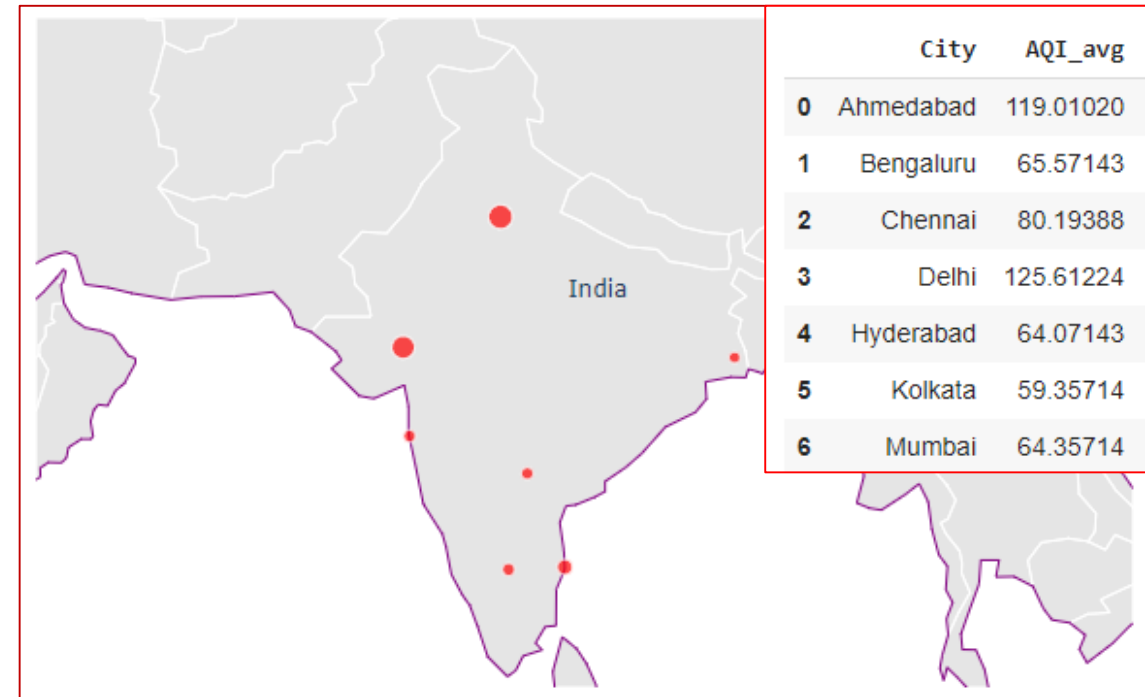
The City wise AQI values during
1st January 2019 to 1st July 2020.

The Lockdown due to COVID has
significantly helped to improve the Air
quality almost in all Metro cities in
India.

PROCESSING AND EDA



The City wise AQI average values during **1st January 2020 to 25th March 2020.**



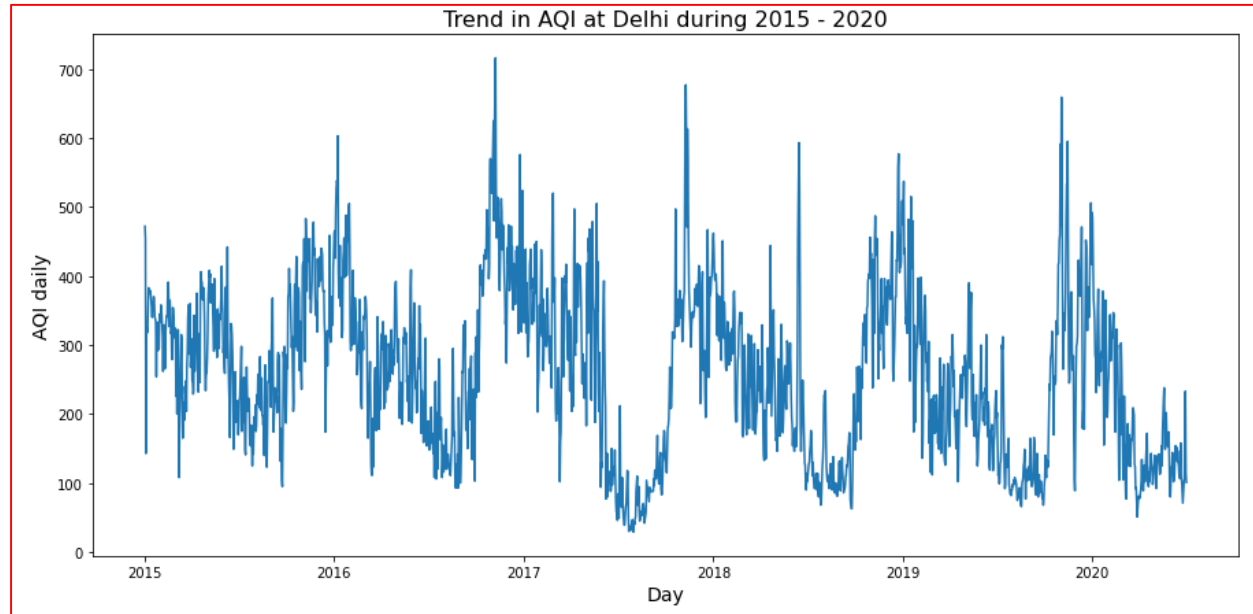
The City wise AQI average values during **26th March 2020 to 1st July 2020.**



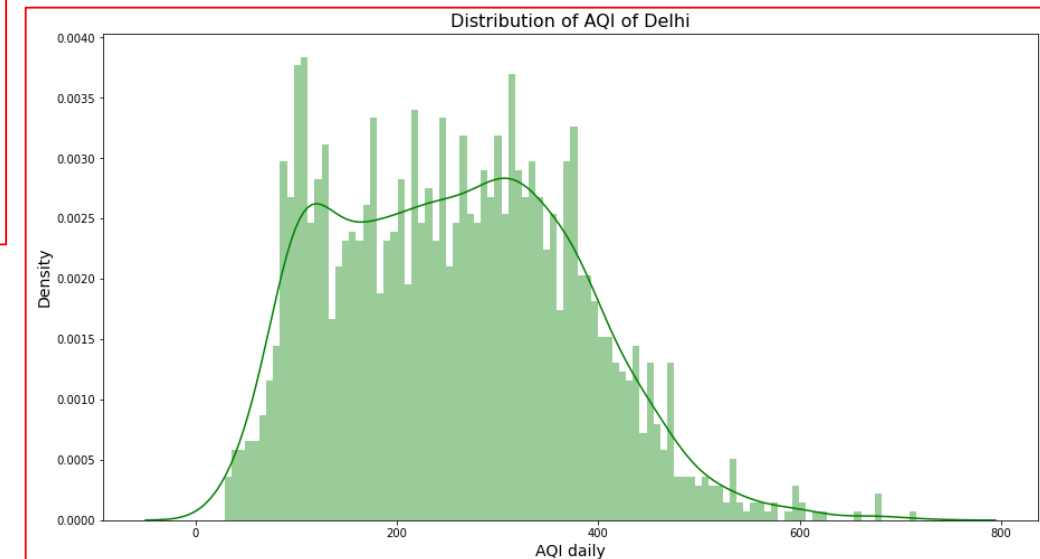
MODEL FORMULATION

Time Series Analysis
Model Evaluation

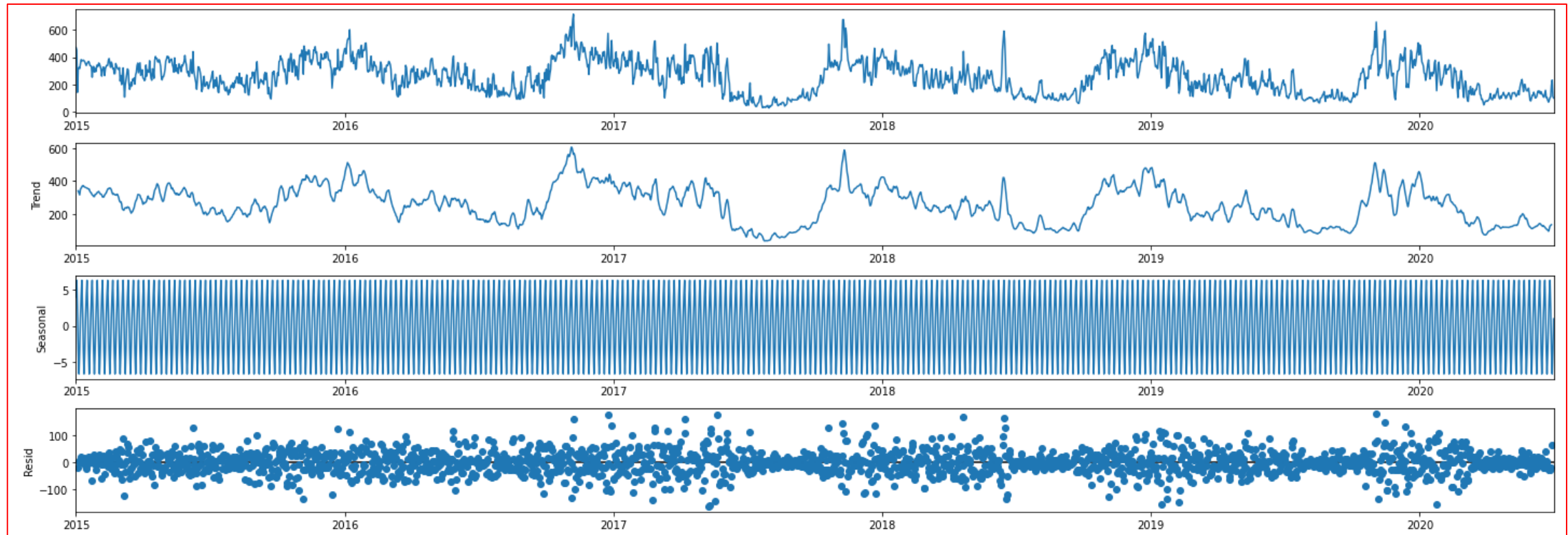
TIME SERIES ANALYSIS ON AQI FOR DELHI



We will analyze the time series considering the AQI of the city “Delhi”

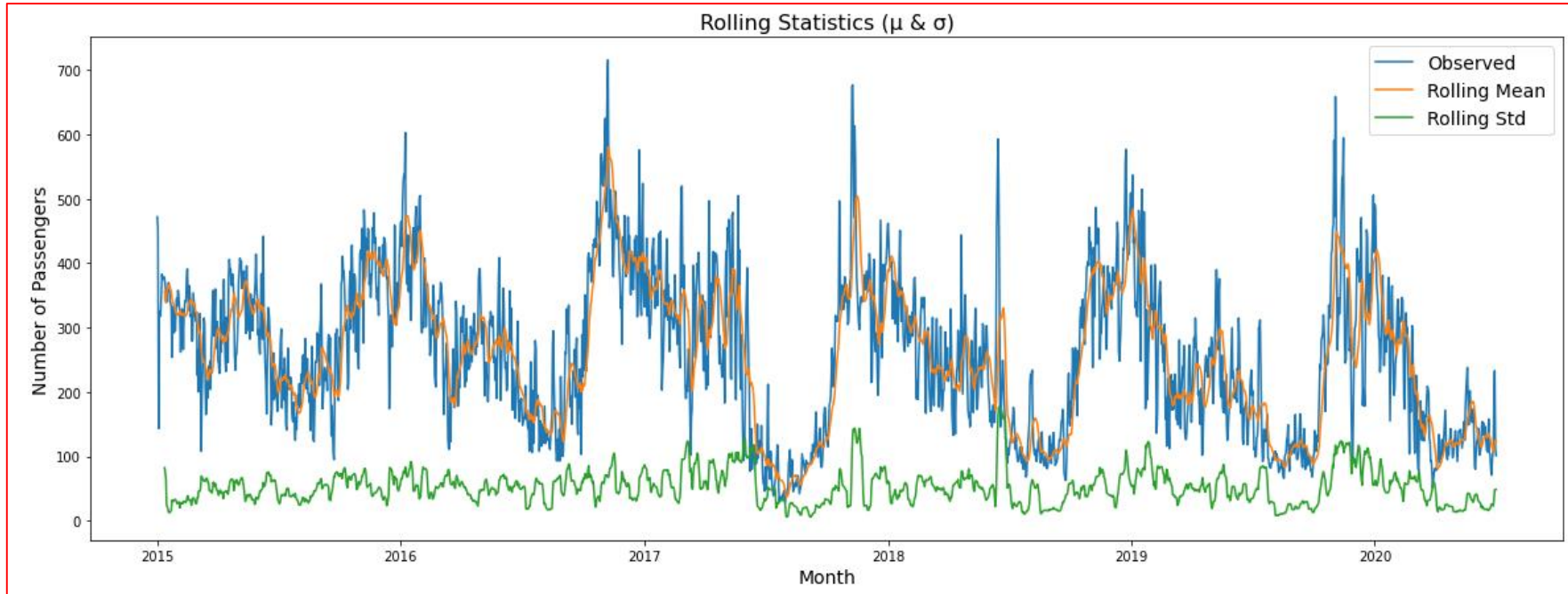


TIME SERIES DECOMPOSITION



2015 to 2020 time series data set of AQI has been decomposed using **“seasonal-decompose”** function

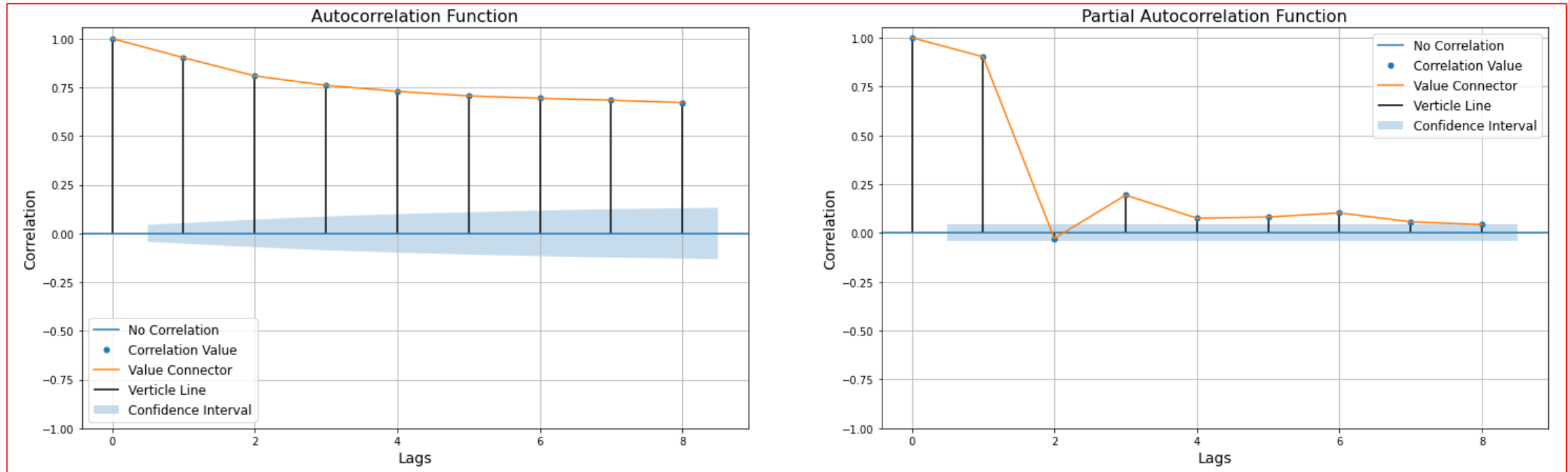
STATIONARITY CHECK : ADF TEST



ADF Test result shows that the original dataset is stationary (using BIC as auto lag)

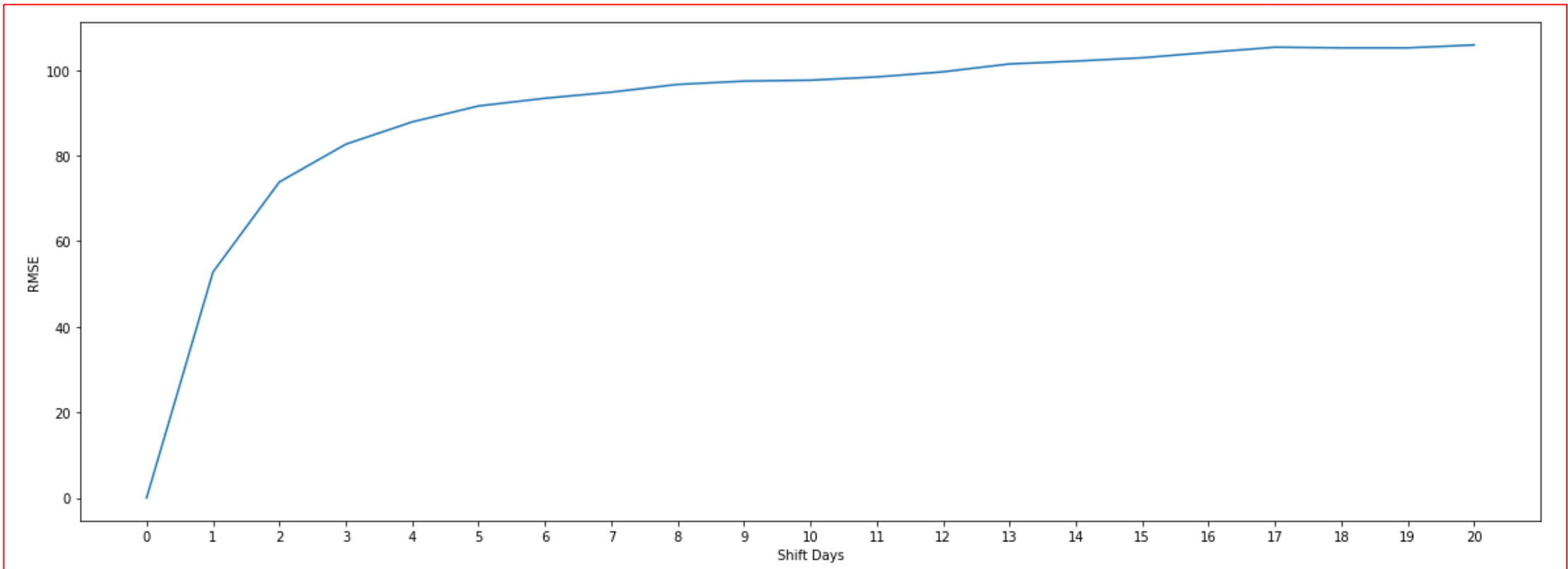
	ADF_Test_Statistics	p-value	Used_Lags	Number_Of_Observations	Critical_Value (1%)	Critical_Value (5%)	Critical_Value (10%)
Original	-4.83883	0.00005	8	2000	-3.43362	-2.86299	-2.56754

ACF AND PACF STUDY : ORIGINAL DATA



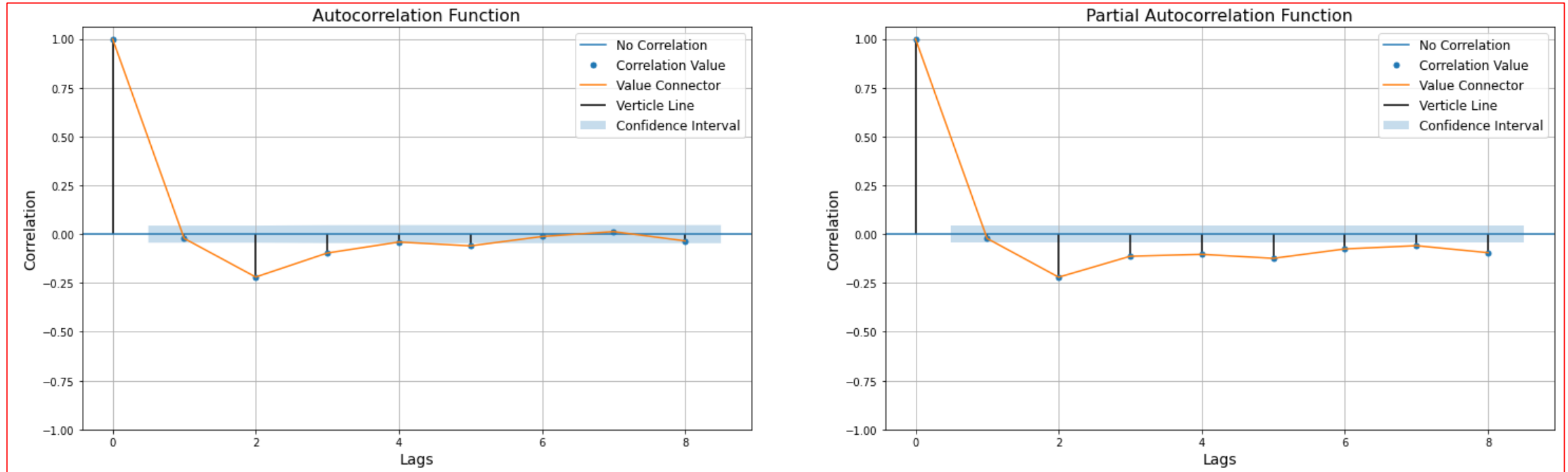
ACF plot suggest that there is high degree of Auto correlation presents inside the series which needs to be removed before model building.

DIFFERENCING AND BASE MODEL



Differencing smoothing has been employed and RMSE measured → based on the plot **difference period = 1** can be used to prepare a base model.

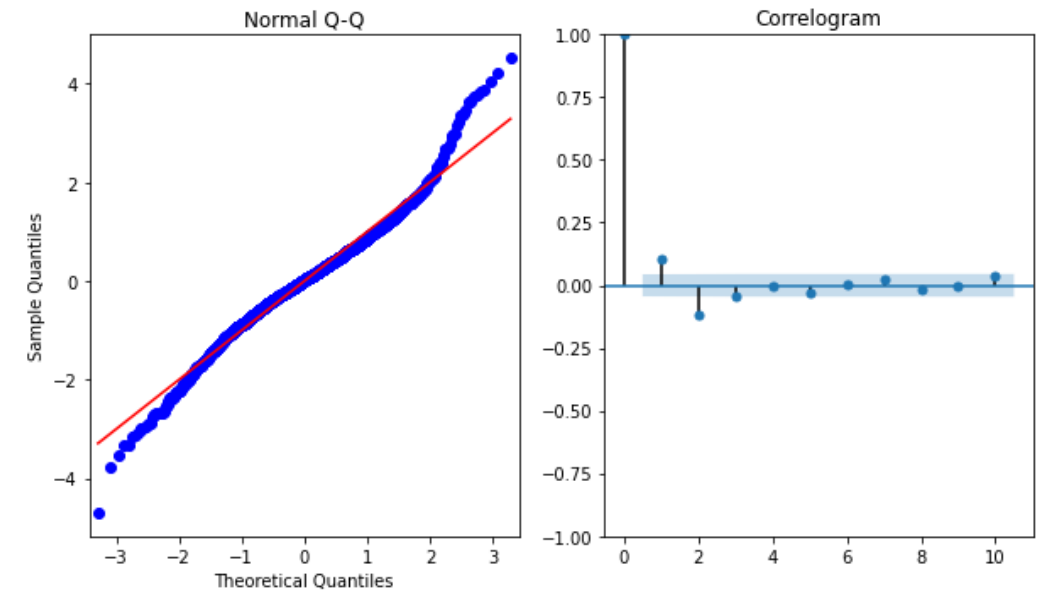
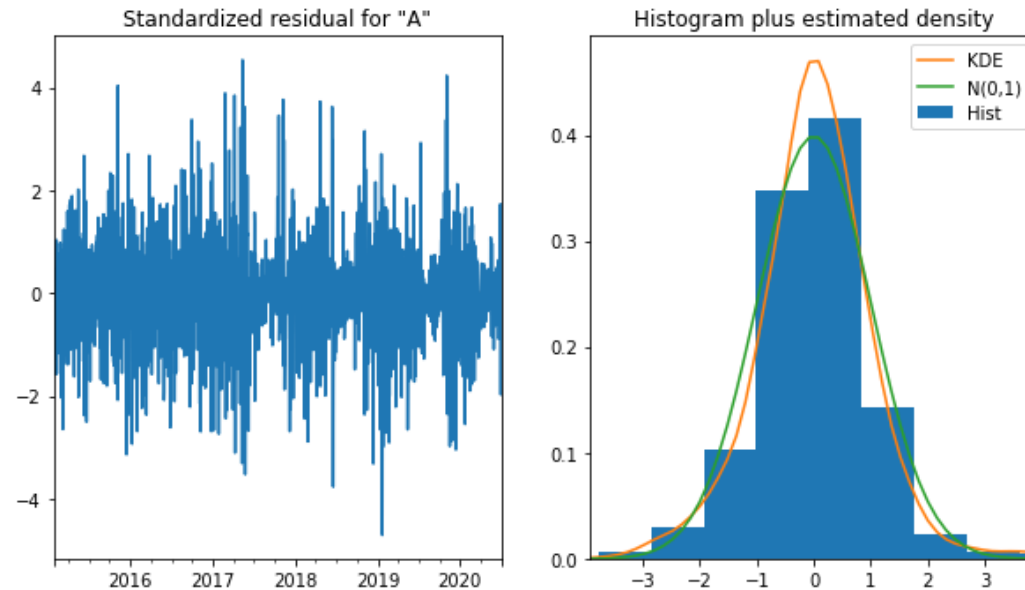
ACF AND PACF STUDY : DIFFERENCED DATA



ACF plot suggest that higher degree of Auto correlation has been removed with differencing; **ACF = 1** and **PACF = 1**

SEASONAL ARIMA MODEL EVALUATION

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.6945	0.022	31.420	0.000	0.651	0.738
ma.L1	-0.9152	0.013	-72.489	0.000	-0.940	-0.890
ar.S.L12	0.0520	0.018	2.834	0.005	0.016	0.088
ma.S.L12	-1.0000	11.652	-0.086	0.932	-23.838	21.838
sigma2	2513.4776	2.93e+04	0.086	0.932	-5.49e+04	5.99e+04



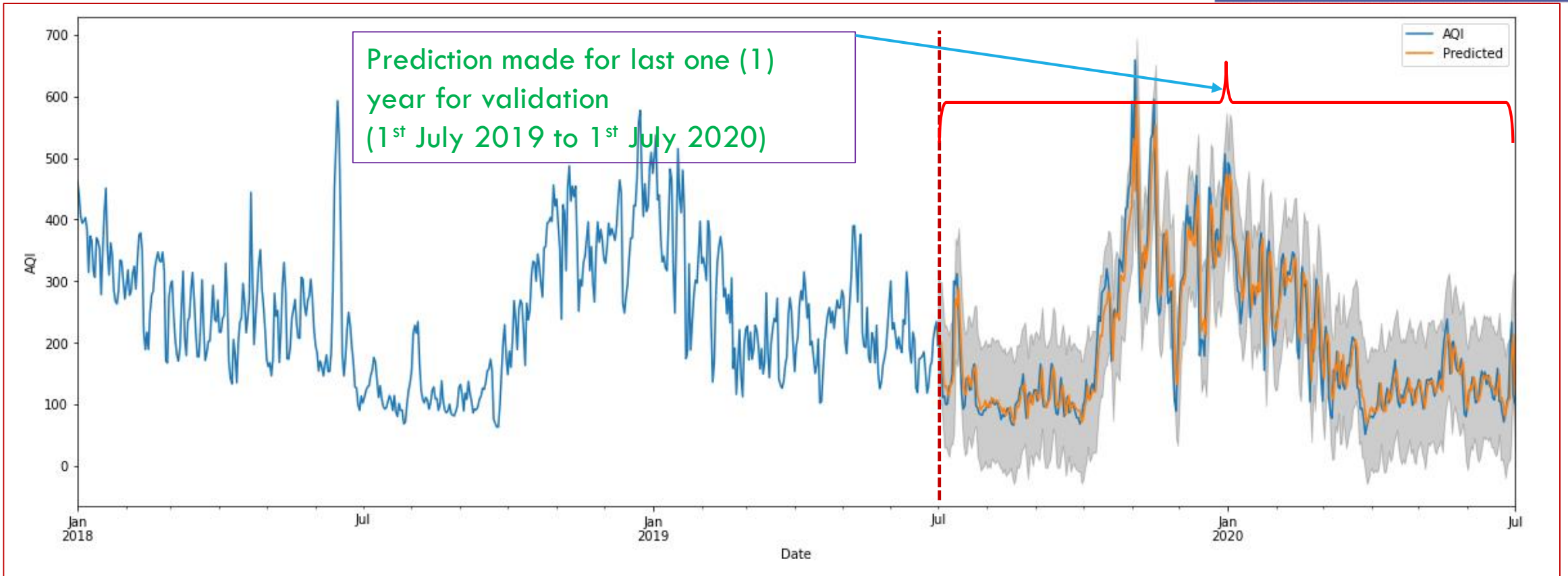
Those observations lead us to conclude that our model produces a satisfactory fit that could help us understand our time series data and forecast future values



CONCLUSION

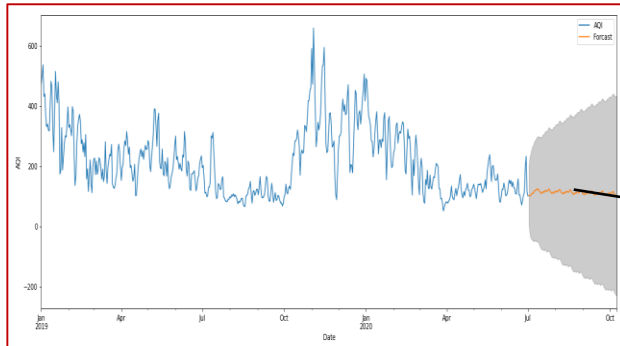
**Prediction & Forecasting
Conclusion**

PREDICTION AND VALIDATE DATASET

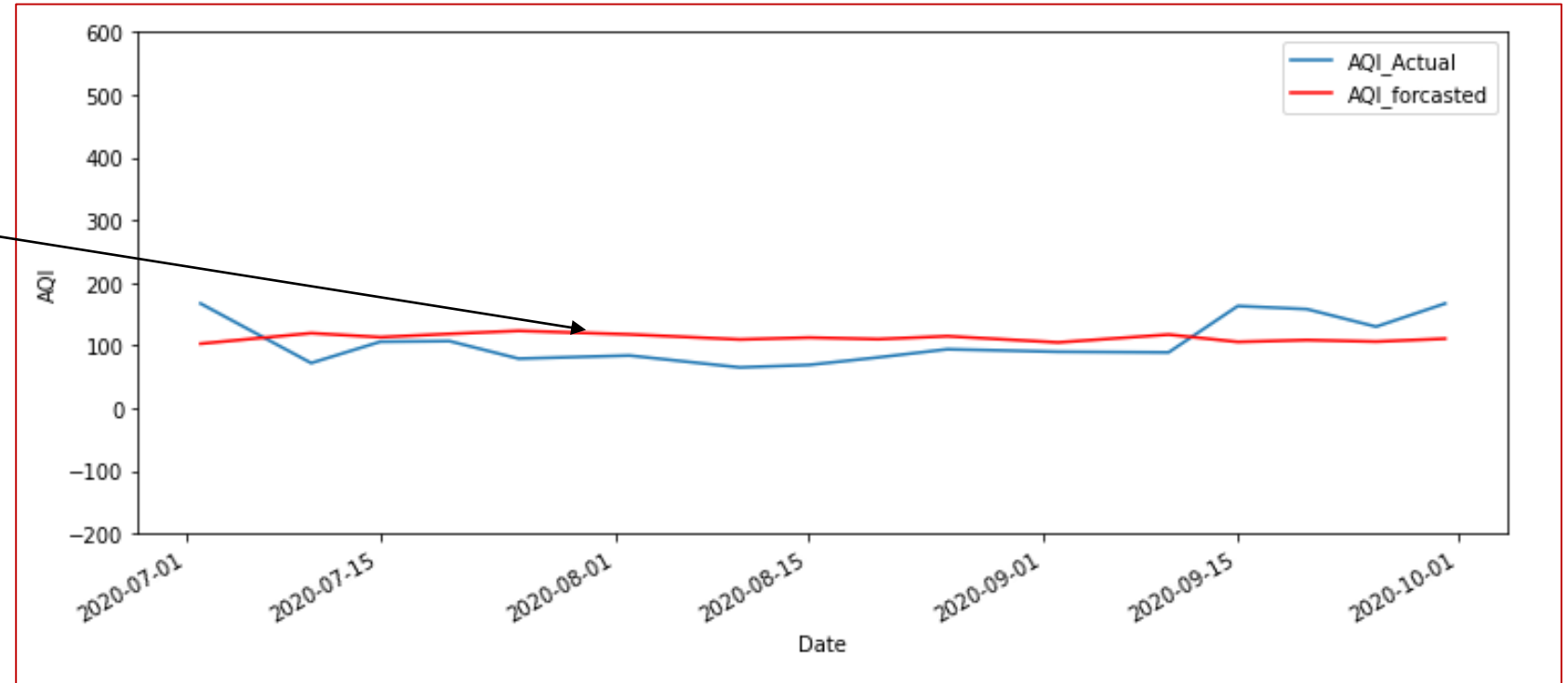


Model performance seems okay with 44 RMSE → Forecasting to be done

FORECASTING NEXT 100 DAYS



Prediction after 1st July 2020
made for next 100 days



AQI of Delhi during July 2020 to October 2020 has been collected separately from CPCB website and are being plotted along with “**Model forecasts**” and it can be concluded that the performance of the model is satisfactory.



Thanks for reading!



Lets collaborate and happy to receive any
feedback/suggestion/comment at..

pathak.chiranjit@gmail.com