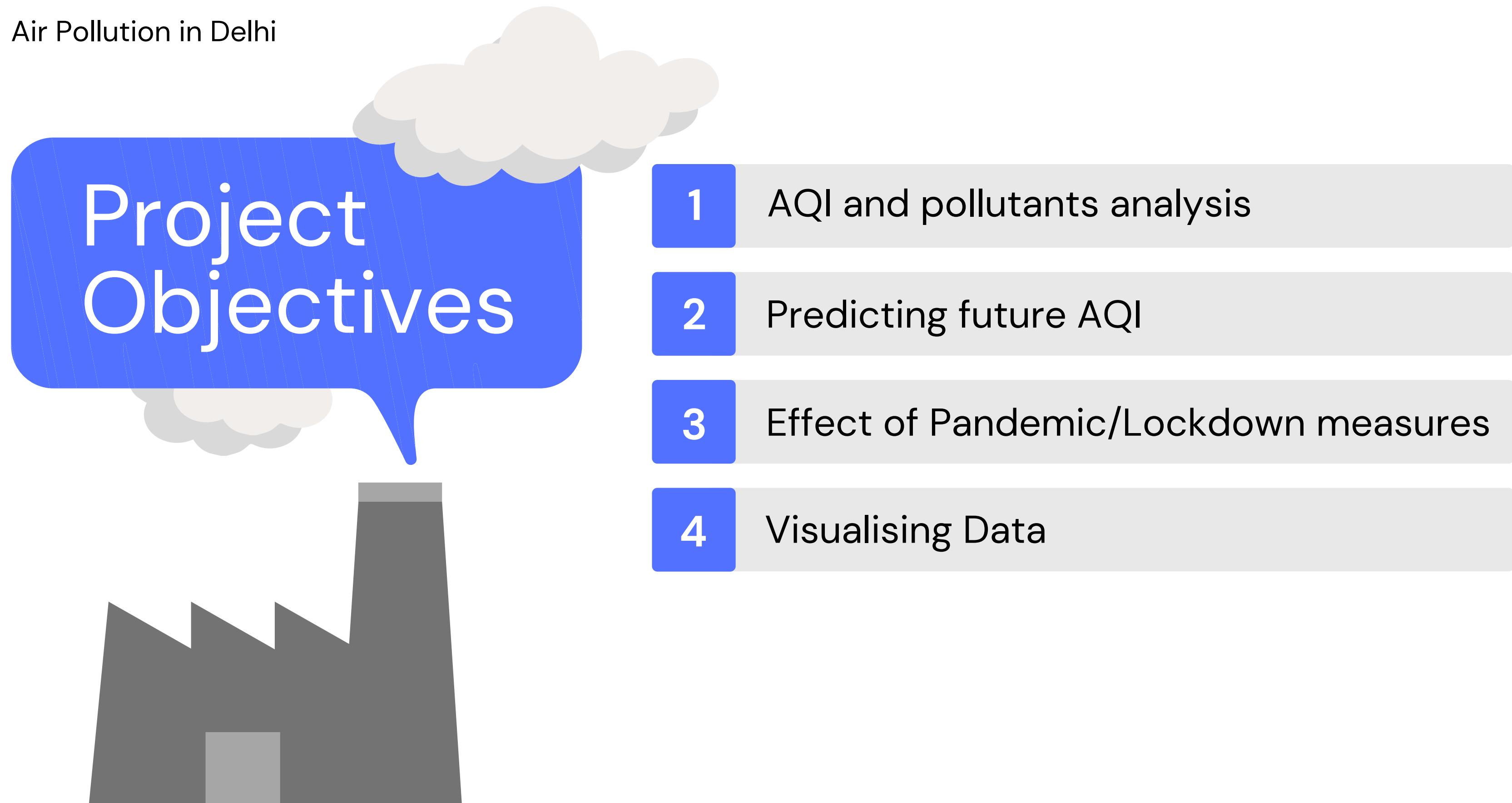


AIR POLLUTION IN DELHI

POLLUTED POLLUTED POLLUTED POLLUTED POLLUTED POLLUTED POLLUTED

Harshita Sharma
Abhigyan Ghosh
Zubair Abid



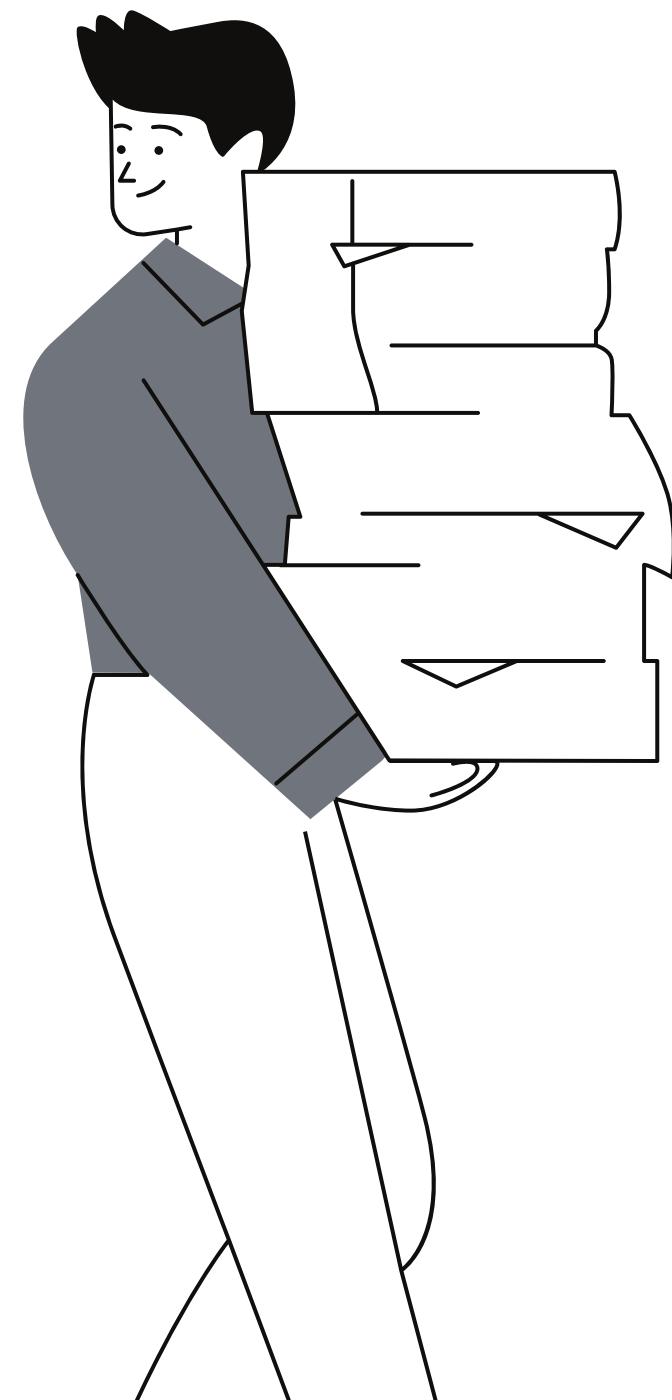
Air Pollution in Delhi

Data Collection

CPCB collects air quality data which is updated every week. At present, the Real-Time Ambient Air Quality Monitoring Network consists of 261 Continuous Ambient Air Quality Monitoring Stations in 134 Cities covering 23 States & Union Territories which are connected to the web-based system.

National Air Quality Index

Data was collected from **CPCB**: Central Pollution and Control Board.



Air Pollution in Delhi

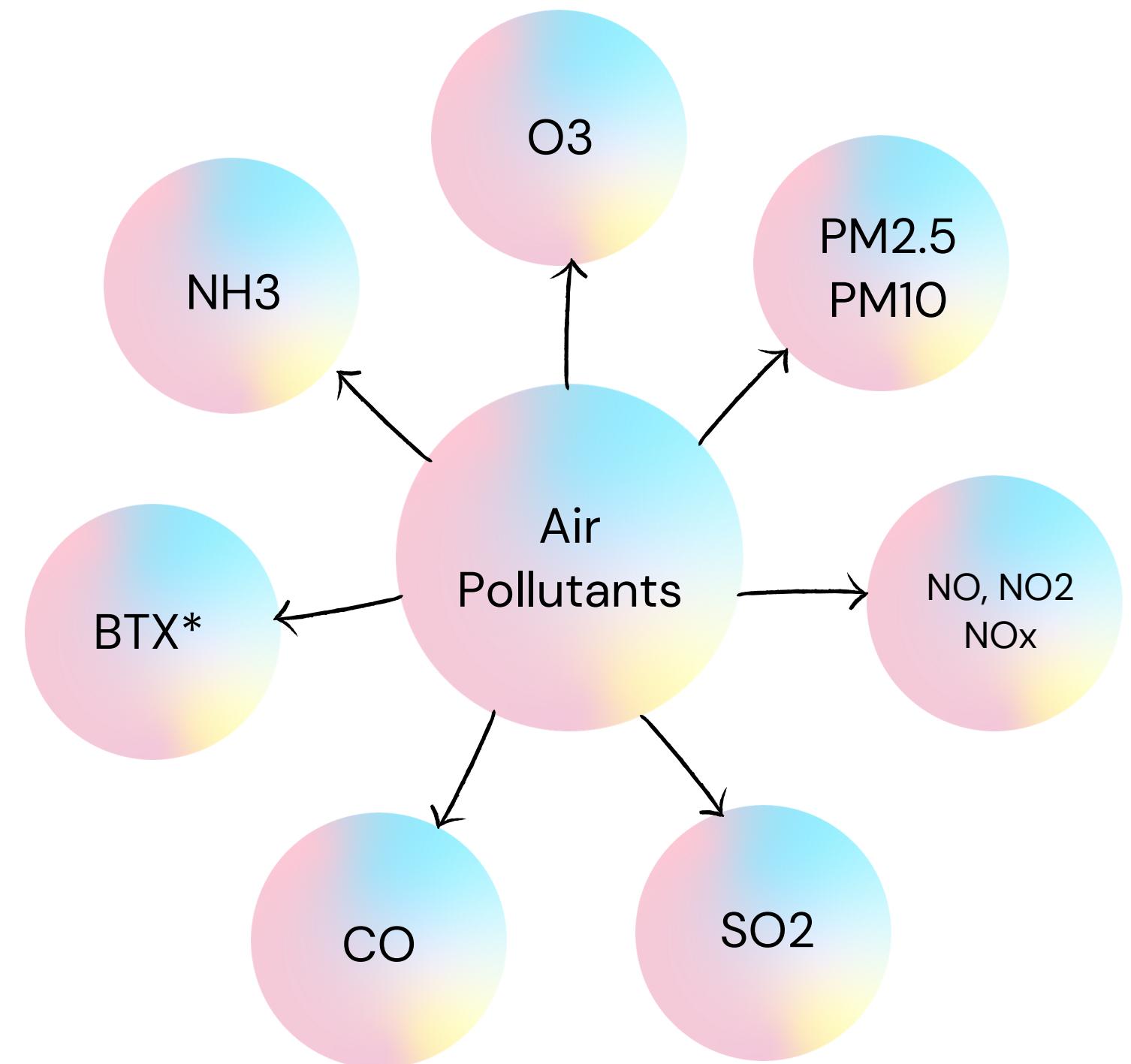
Data Analysis

Time series data: The available data is between 2015-01-01 and 2020-07-01 with pollutant concentrations in each city for each day.



Data Analysis

Types of Air Pollutants in the dataset.



*Benzene, Toulene and Xylene are grouped under BTX

Data Analysis

**10 most polluted cities
(worst AQI) in the dataset.**

26 cities

Chennai Mumbai **Delhi** Bengaluru Lucknow Ahmedabad Hyderabad Patna Gurugram Visakhapatnam Amritsar Jorapokhar Jaipur Thiruvananthapuram Amaravati Brajrajnagar Talcher Kolkata Guwahati Coimbatore Shillong Chandigarh Bhopal Kochi Ernakulam Aizawl

Delhi
Talcher Brajrajnagar
Jorapokhar Gurugram
Ahmedabad
Ghuwati Patna Lucknow
Kolkata



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.

A web-based system is designed to provide AQI on a real-time basis. It is an automated system that captures data from continuous monitoring stations without human intervention and displays AQI based on running average values.

E.g. AQI at 6 am on a day will incorporate data from 6 am on the previous day to the current day.

For manual monitoring stations, an AQI calculator is developed wherein data can be fed manually to get AQI value.

Air Pollution in Delhi



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.

6 categories of the air have been created in this air quality index.

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
Respiratory 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.

The AQI calculation uses 7 measures: PM2.5, PM10, SO2, NOx, NH3, CO and O3.

Sometimes measures are not available due to lack of measuring or lack of required data points.

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.

Delhi

Data Analysis

1. The city-level data of Delhi was extracted from the dataset.
2. Benzene, Toluene and Xylene were grouped together as BTX.

Air Pollution in Delhi



Pollutant-wise most polluted cities in the dataset.

AQI
Delhi: 259.49

2nd
highest concentration in India

Air Pollution in Delhi



PM 2.5
Delhi: 117.2*

**Pollutant-wise most polluted
cities in the dataset.**

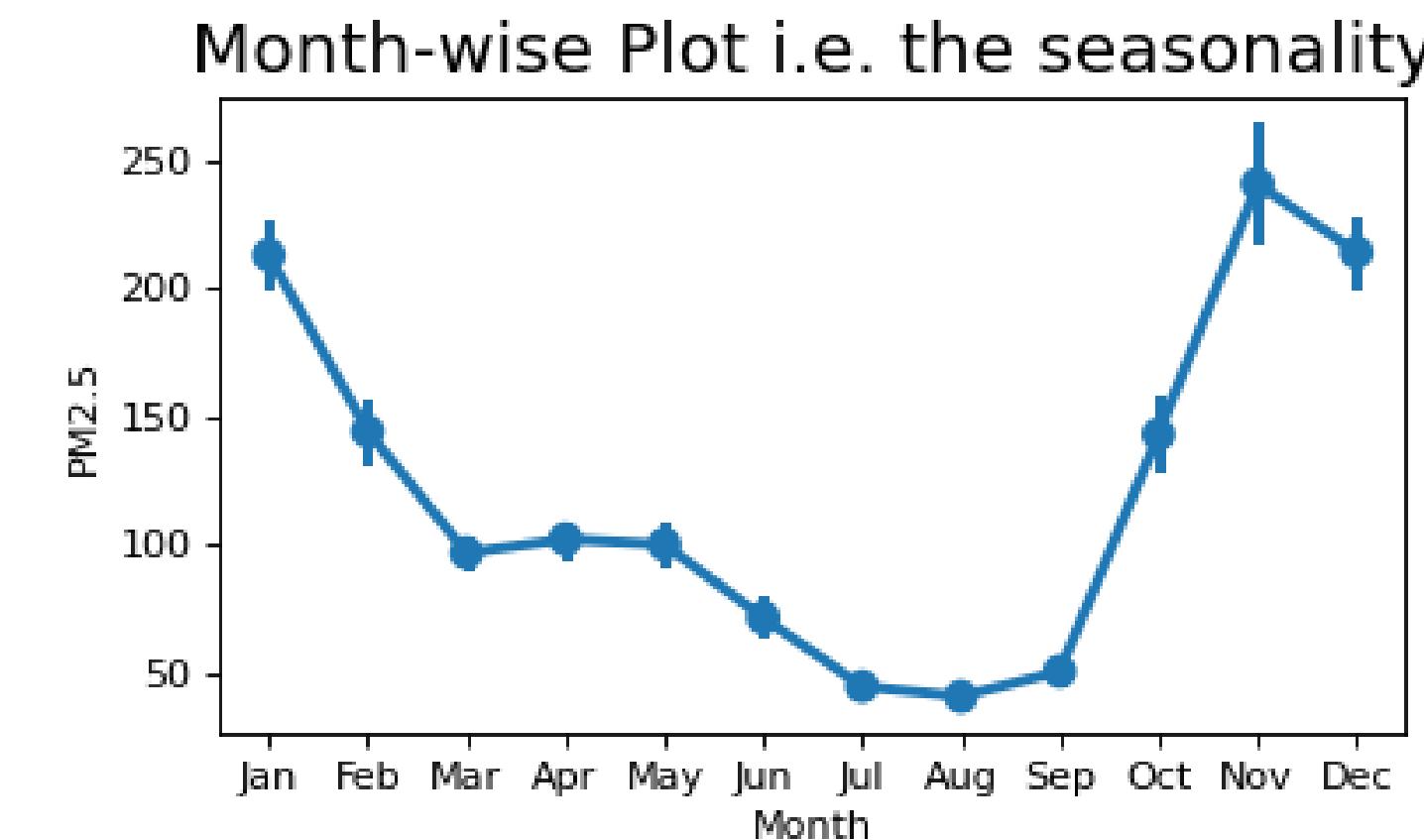
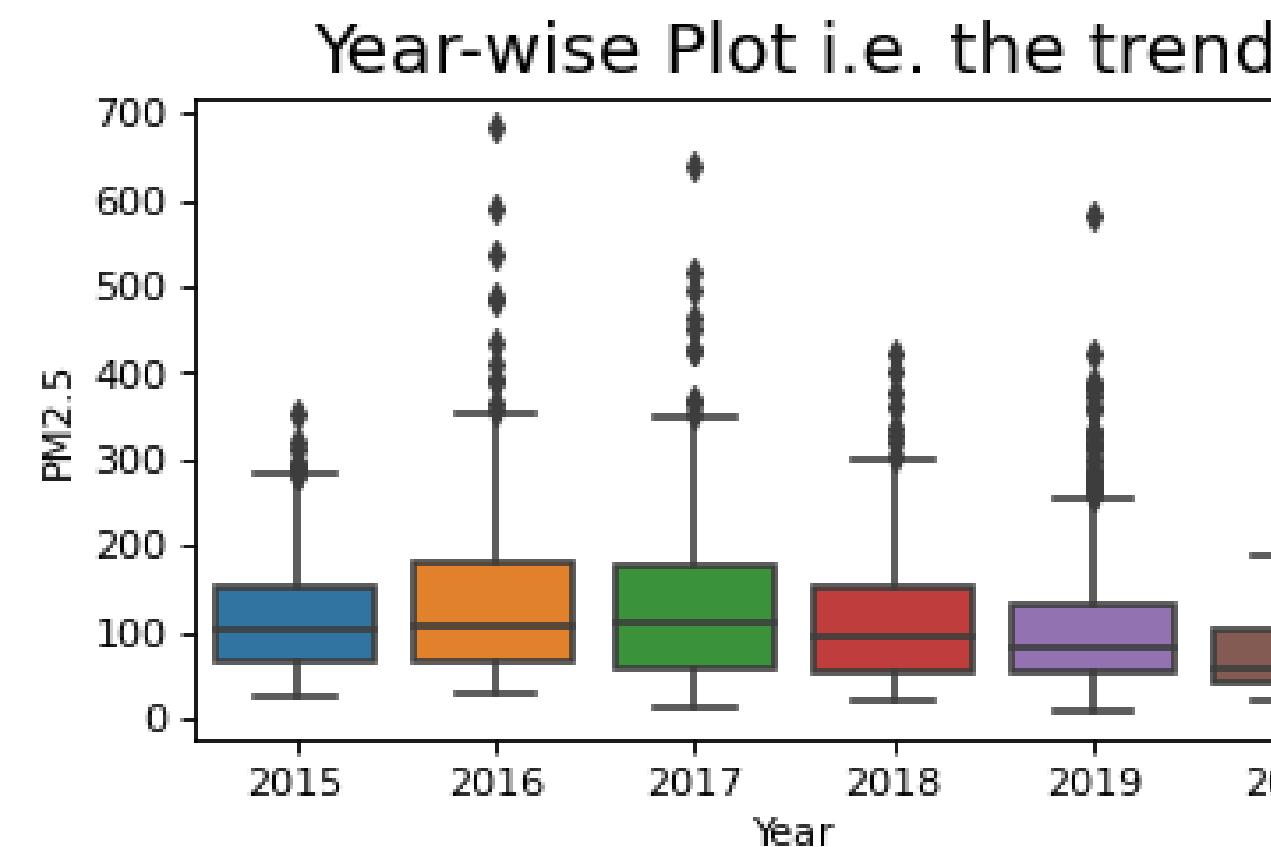
2nd
highest concentration in India

* $\mu\text{g}/\text{cubic meter}$

Air Pollution in Delhi

Data Analysis

PM 2.5
Delhi: 117.2



Air Pollution in Delhi



PM 10
Delhi: 232.81*

**Pollutant-wise most polluted
cities in the dataset.**

1st
highest concentration in India

* $\mu\text{g}/\text{cubic meter}$

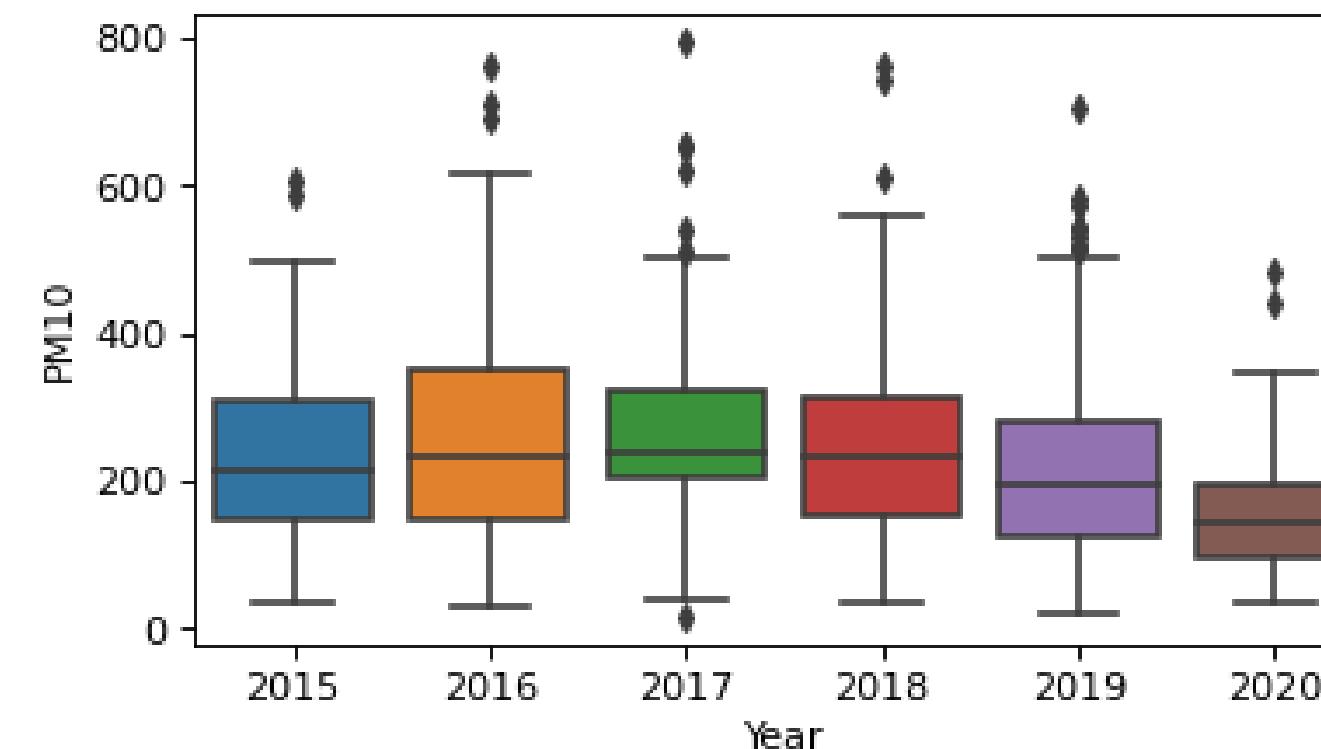
Air Pollution in Delhi

Data Analysis

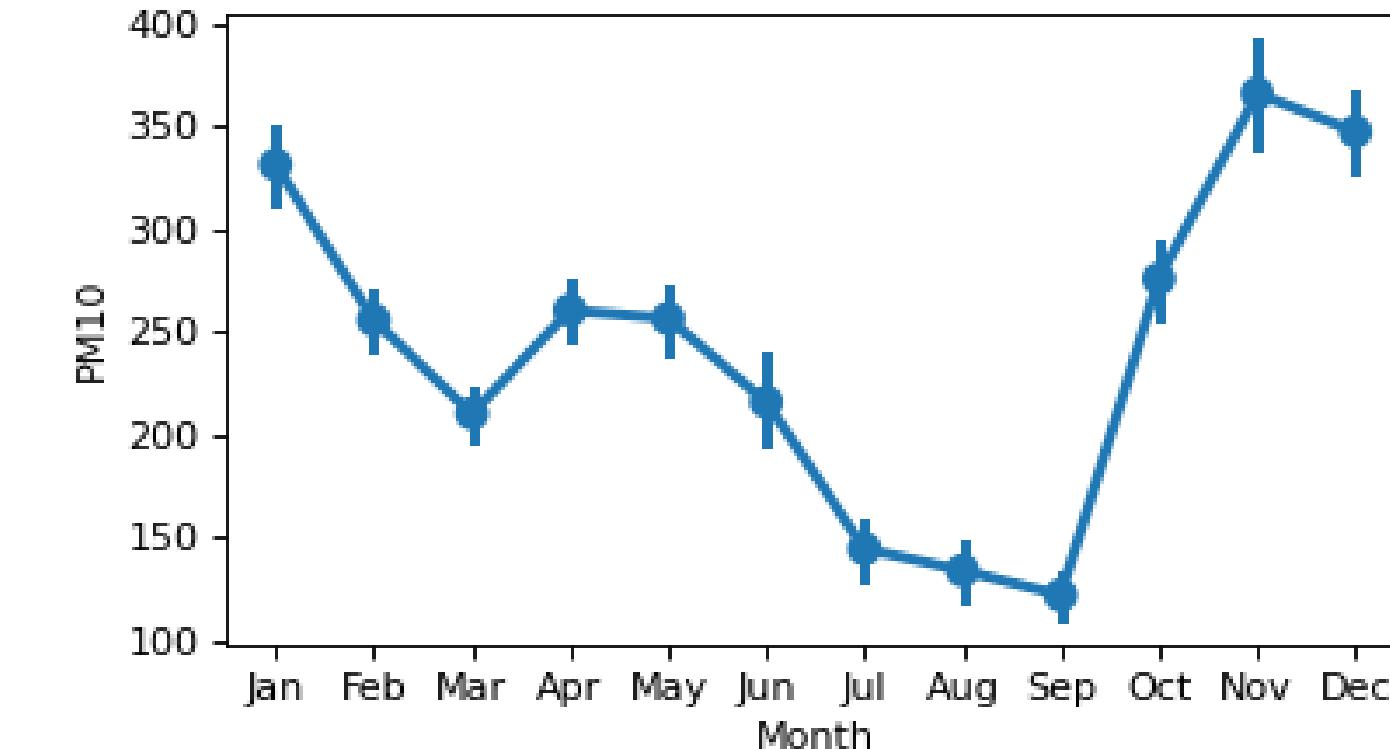
PM 10

Delhi: 232.81

Year-wise Plot i.e. the trend



Month-wise Plot i.e. the seasonality



Air Pollution in Delhi



NO₂
Delhi: 50.79*

Pollutant-wise most polluted cities in the dataset.

2nd
highest concentration in India

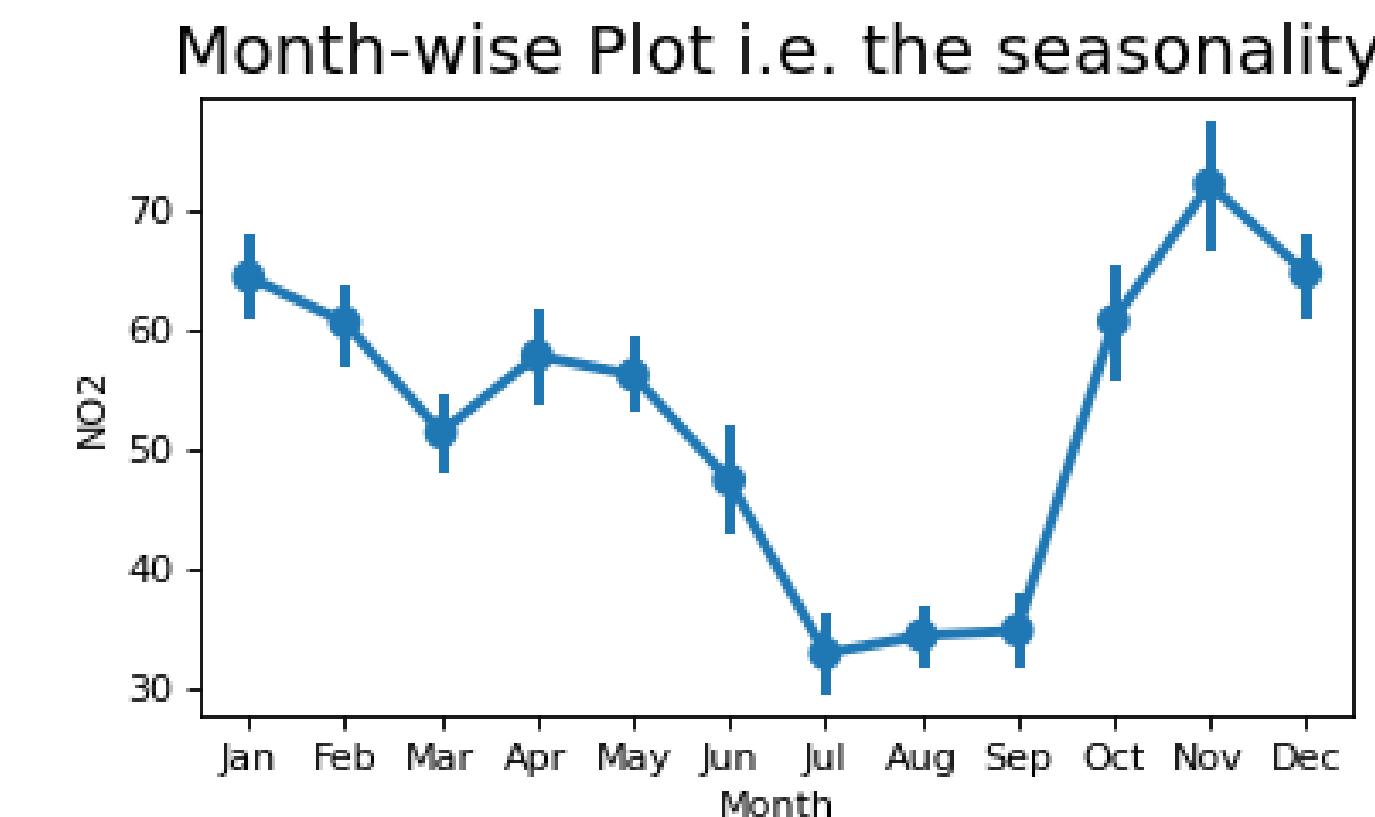
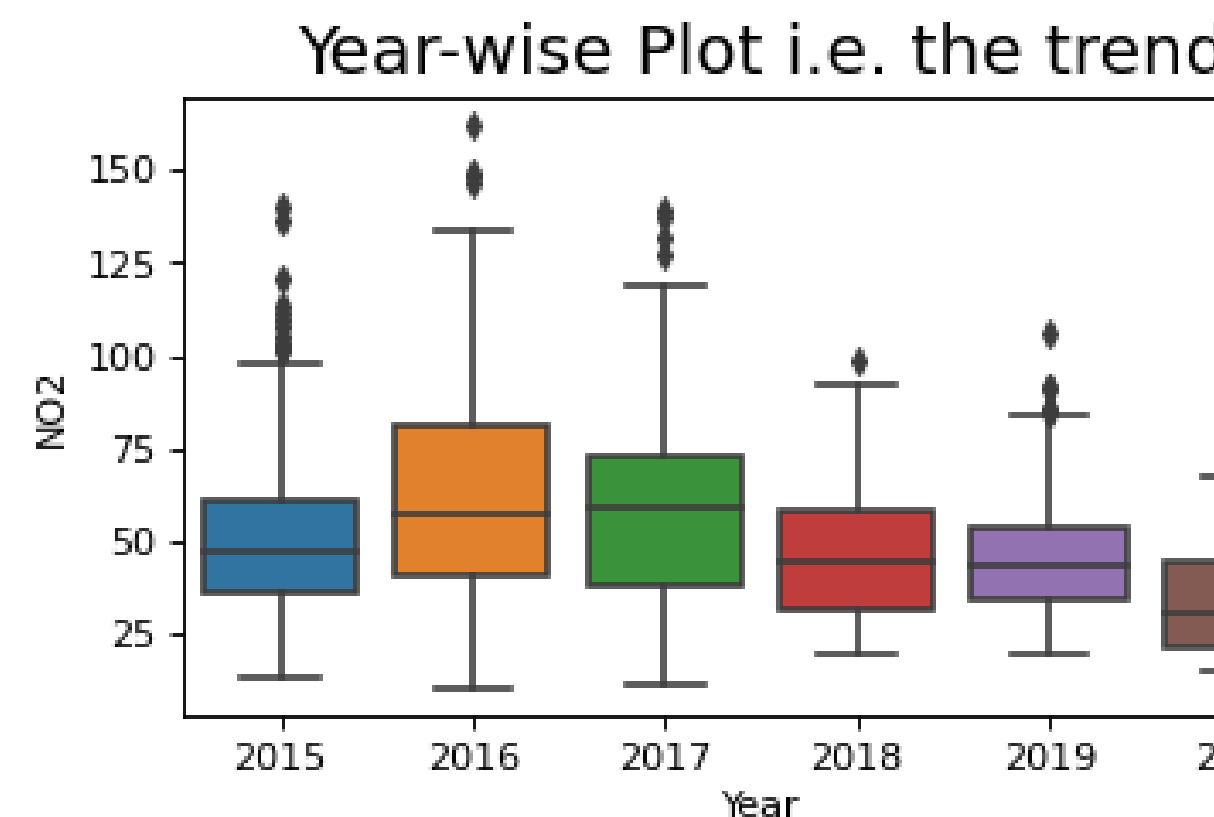
* $\mu\text{g}/\text{cubic meter}$

Air Pollution in Delhi

Data Analysis

NO₂

Delhi: 50.79



Air Pollution in Delhi

Data Analysis

SO₂

Delhi: 15.9*

Pollutant-wise most polluted cities in the dataset.

6th

highest concentration in India

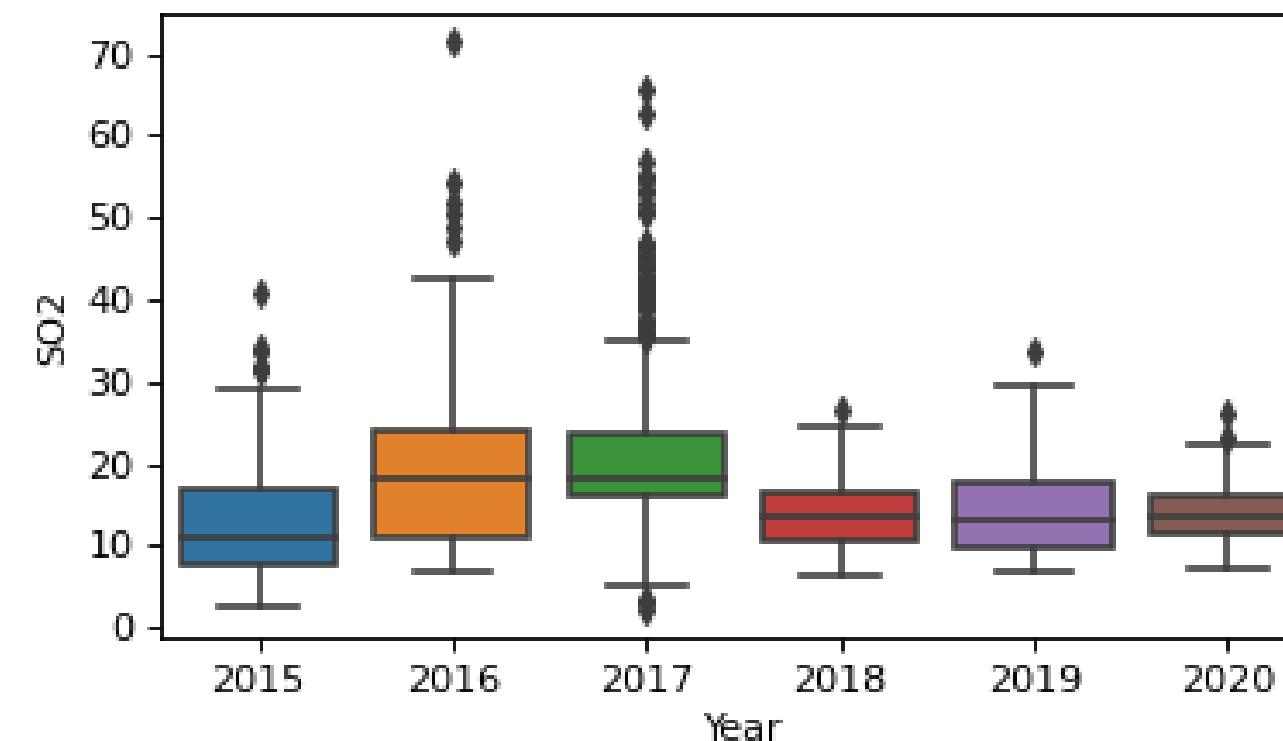
* $\mu\text{g}/\text{cubic meter}$

Data Analysis

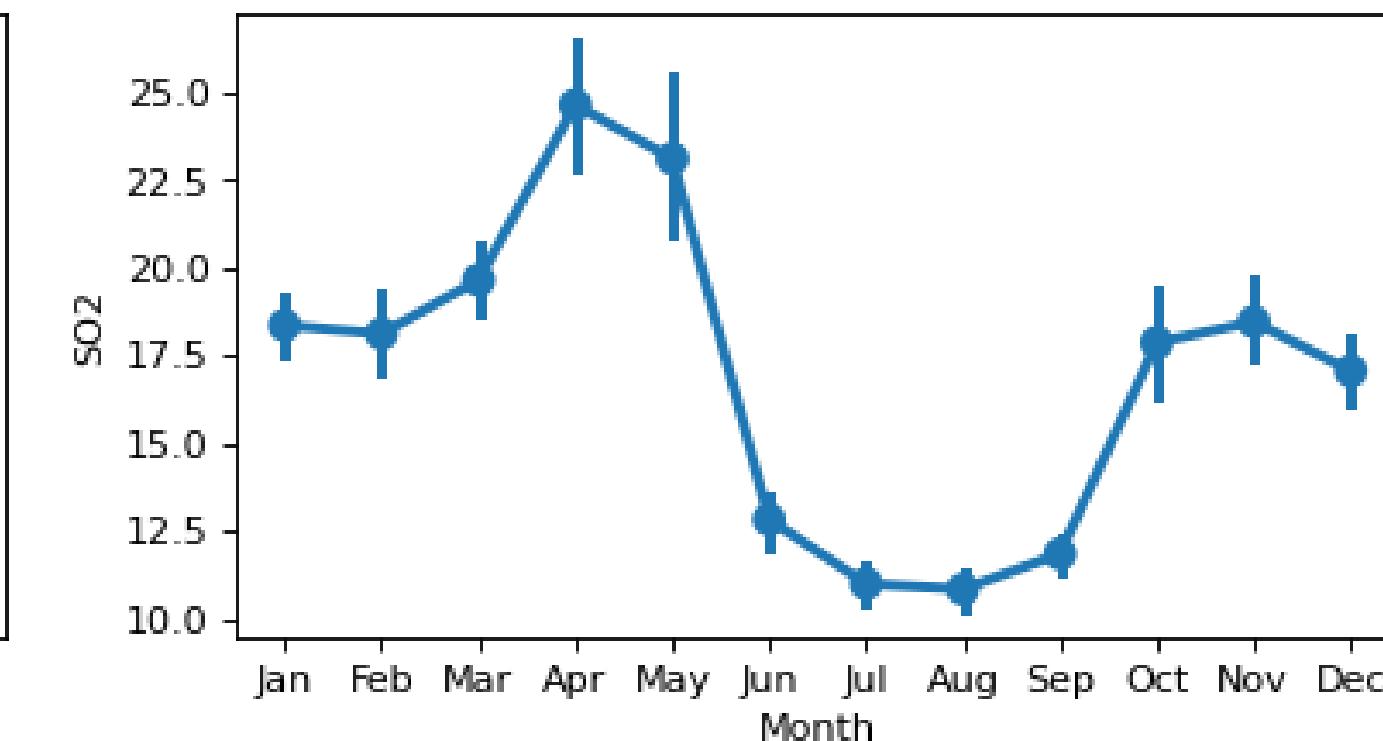
SO₂

Delhi: 15.9

Year-wise Plot i.e. the trend



Month-wise Plot i.e. the seasonality



Air Pollution in Delhi



Pollutant-wise most polluted cities in the dataset.

CO
Delhi: 1.98*

3rd

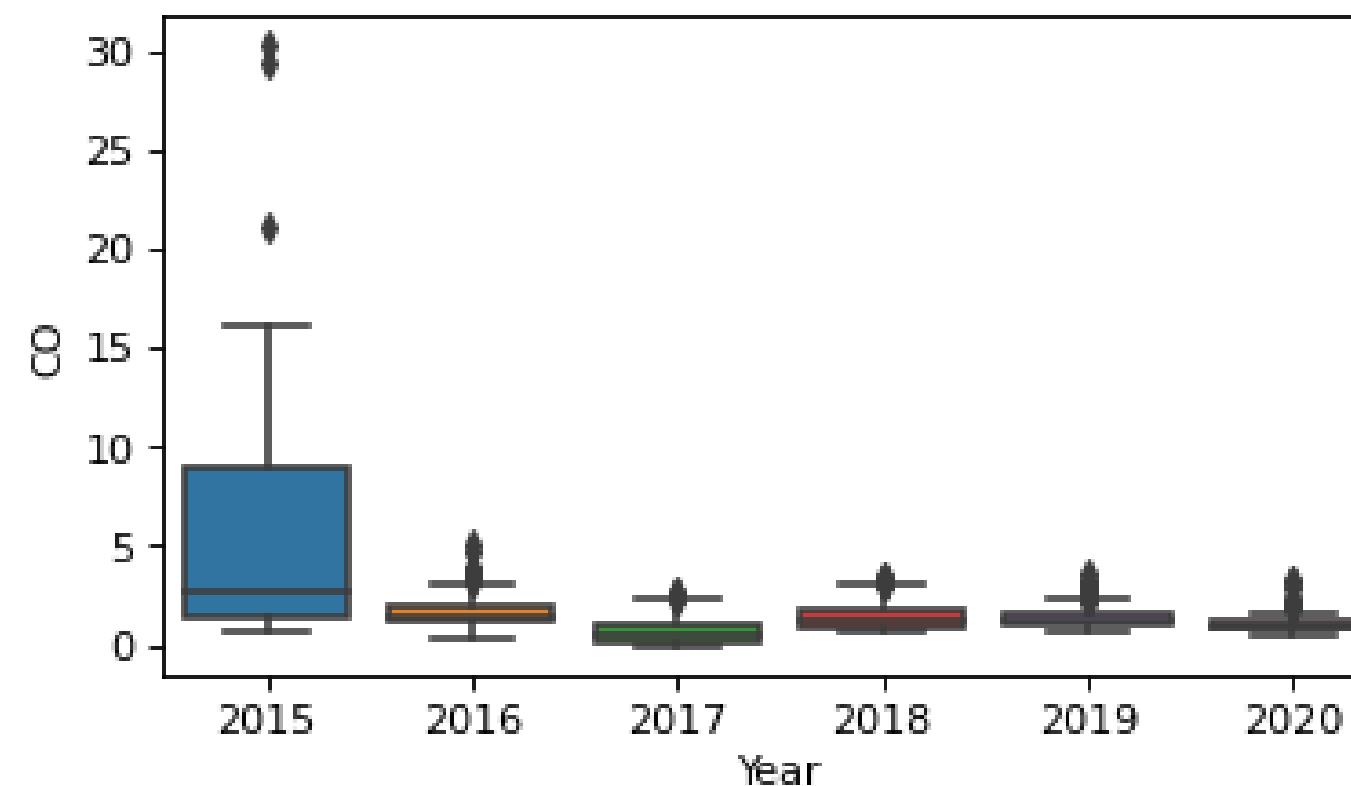
highest concentration in India

* $\mu\text{g}/\text{cubic meter}$

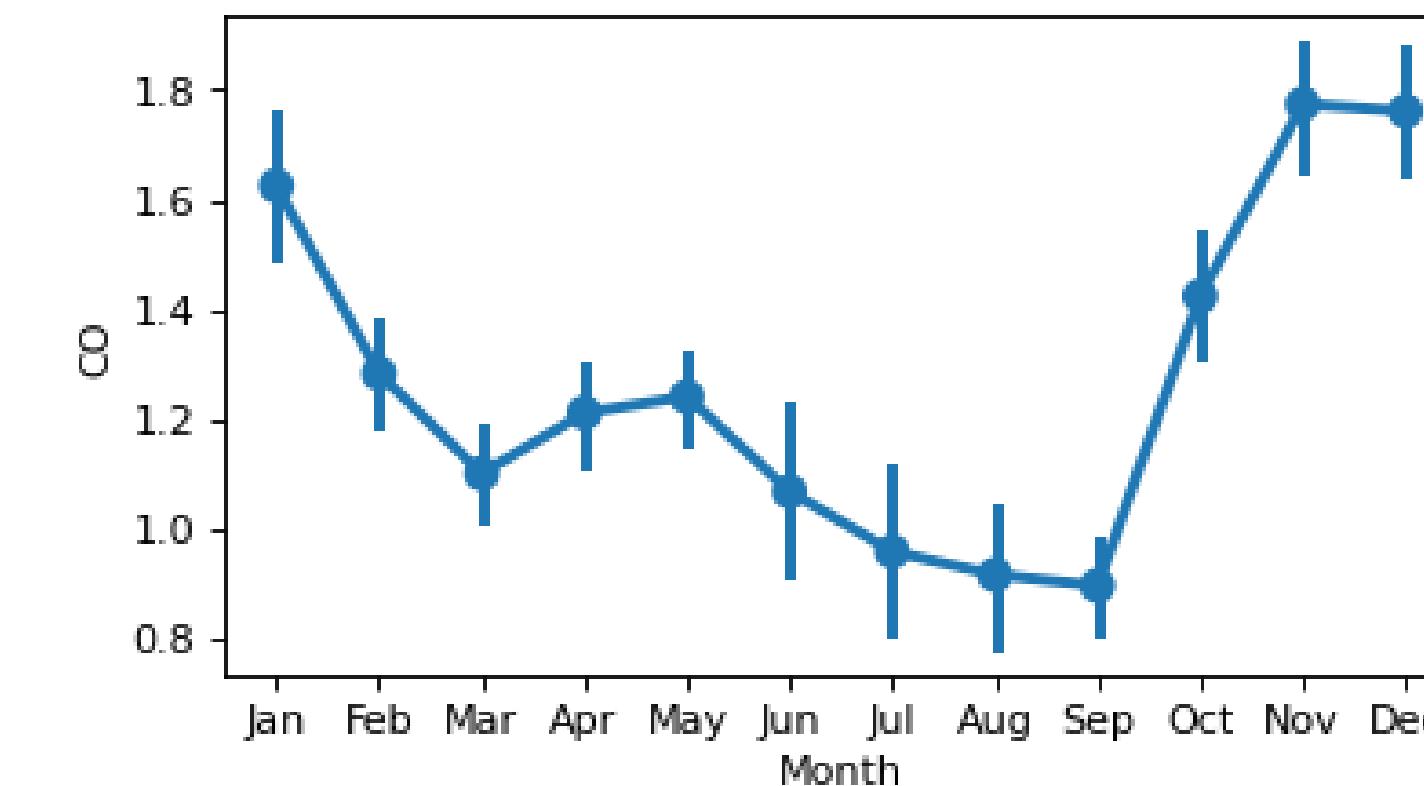
Air Pollution in Delhi

Data Analysis

Year-wise Plot i.e. the trend



Month-wise Plot i.e. the seasonality



CO
Delhi: 1.98



O3

Delhi: 51.32*

Pollutant-wise most polluted cities in the dataset.

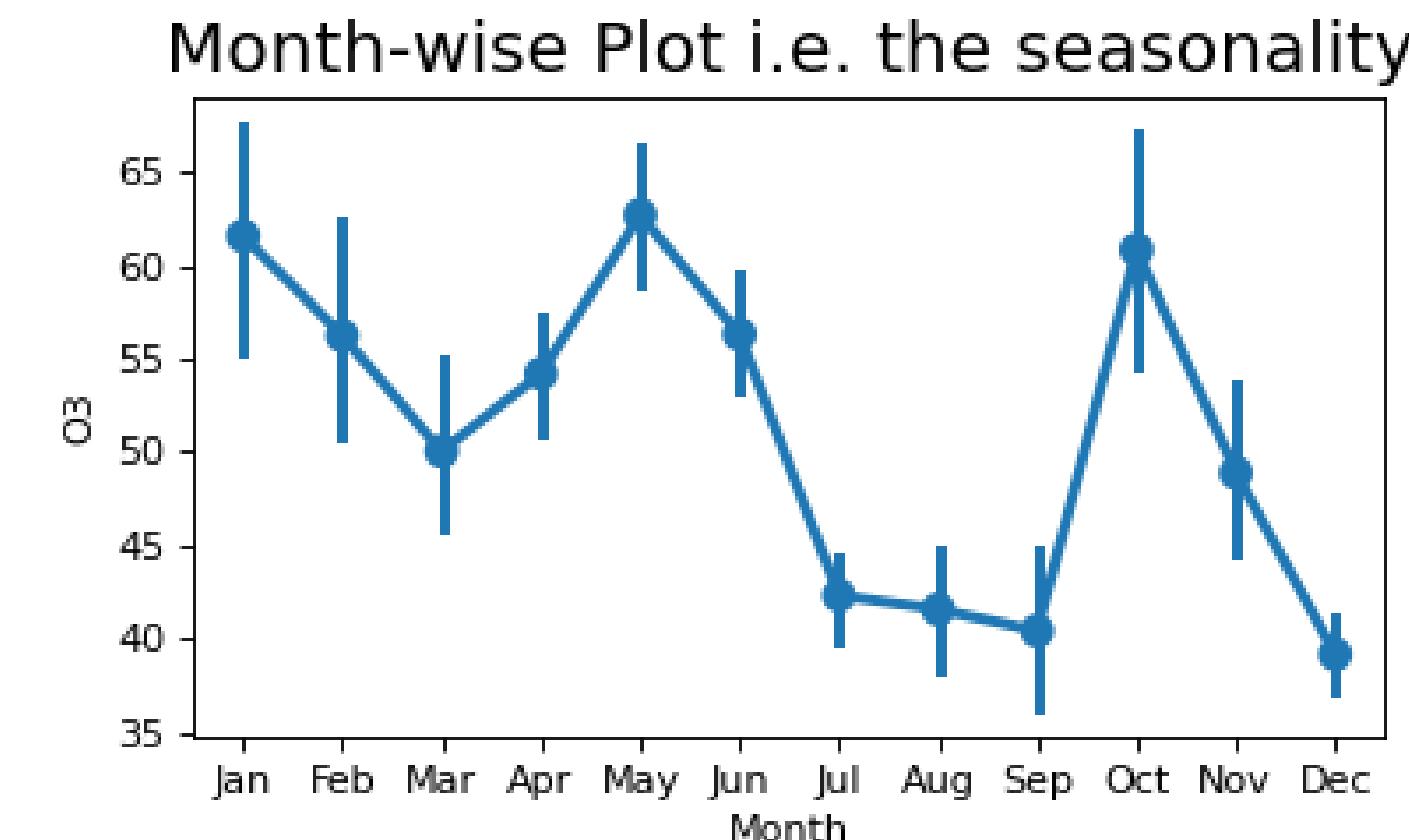
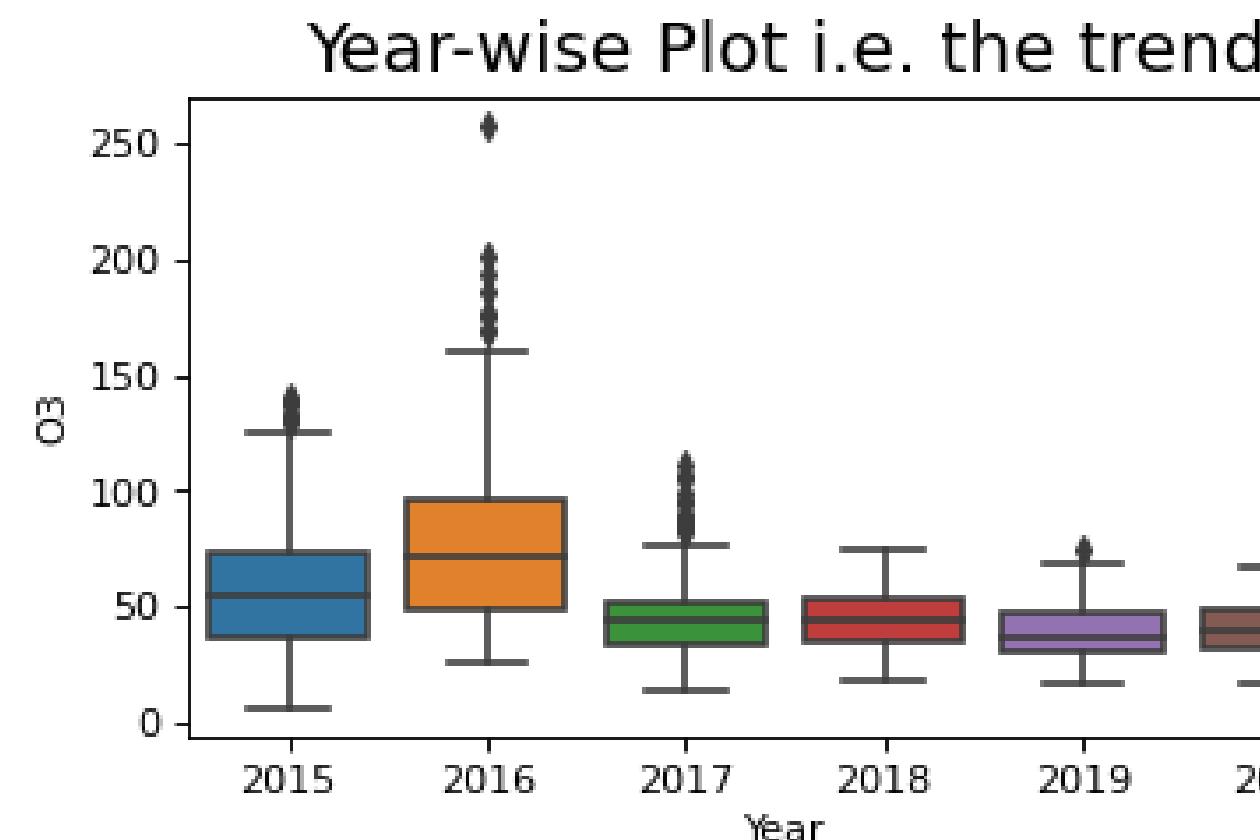
2nd

highest concentration in India

* $\mu\text{g}/\text{cubic meter}$

O3
Delhi: 51.32

Data Analysis



Air Pollution in Delhi



Pollutant-wise most polluted cities in the dataset.

BTX
Delhi: 26.86*

2nd
highest concentration in India

* $\mu\text{g}/\text{cubic meter}$

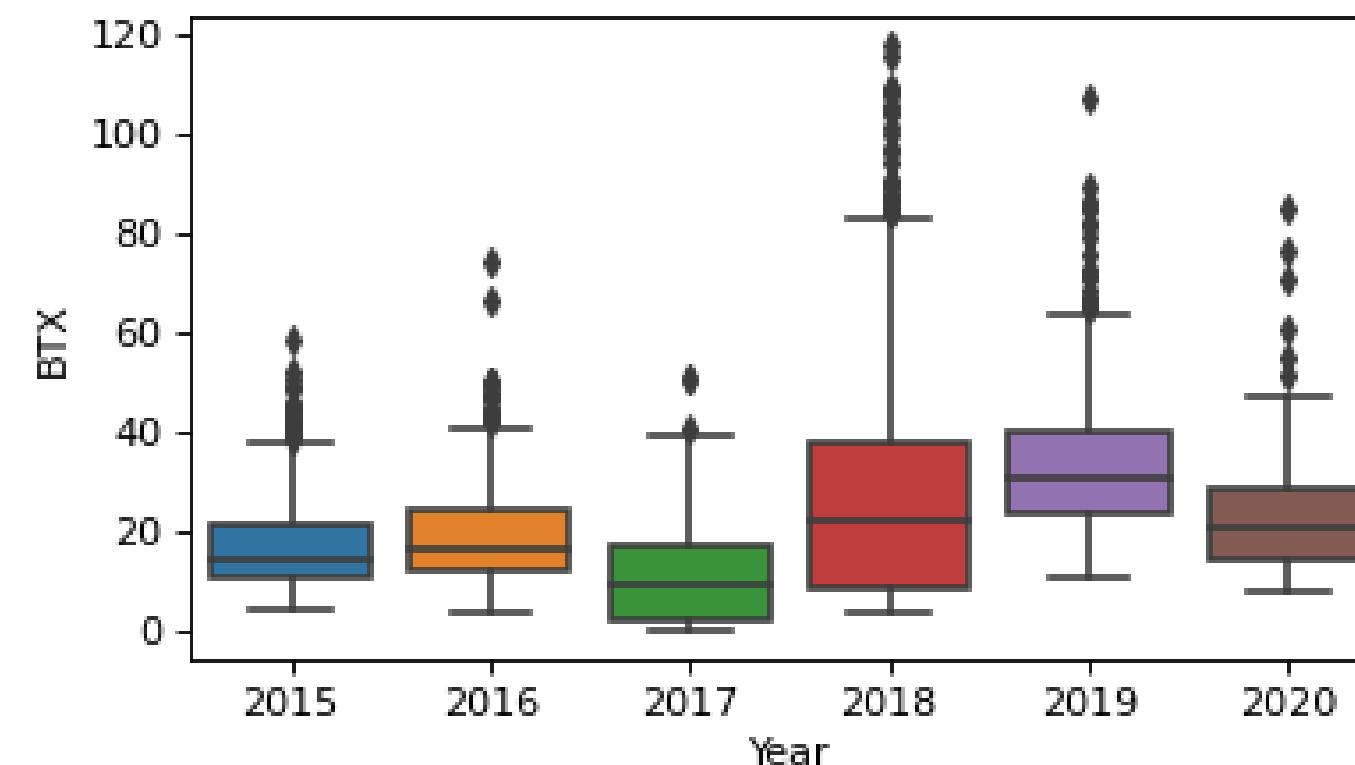
Air Pollution in Delhi

Data Analysis

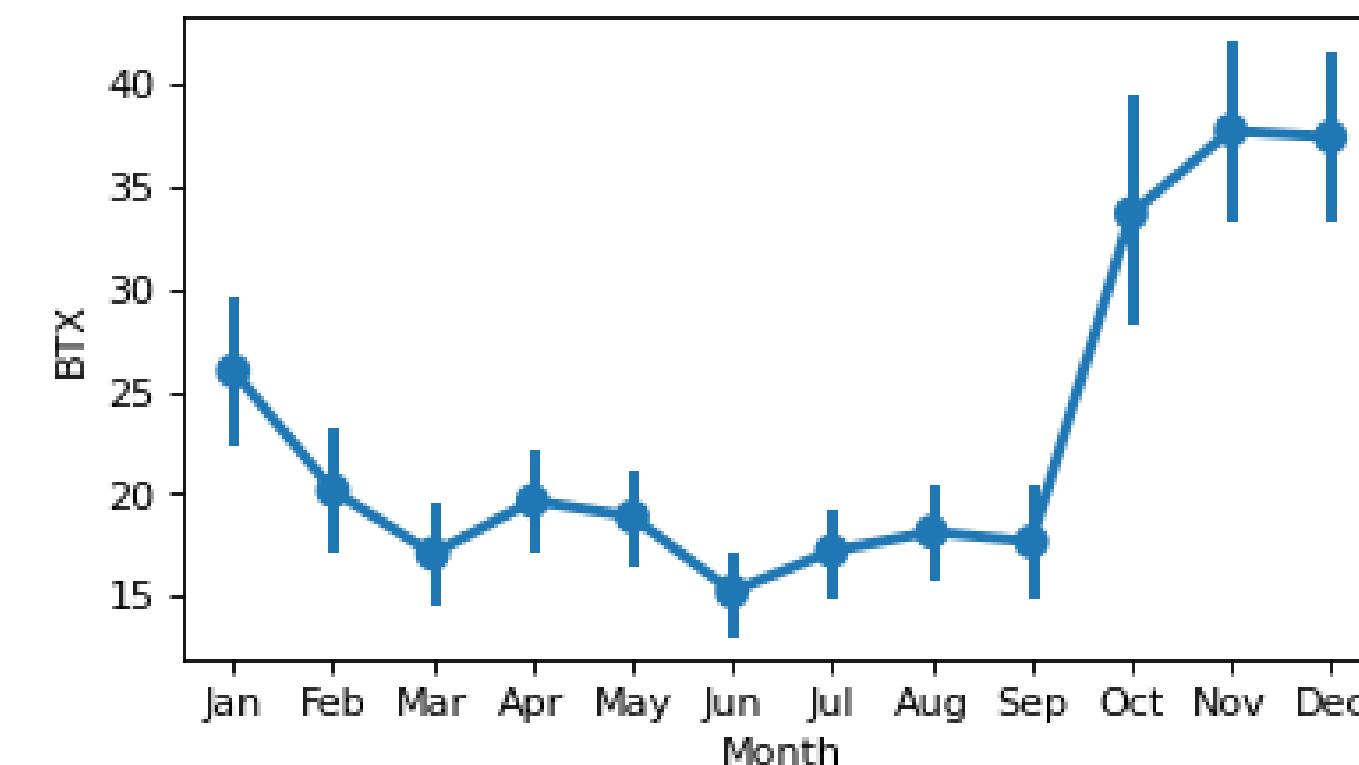
BTX

Delhi: 26.86

Year-wise Plot i.e. the trend



Month-wise Plot i.e. the seasonality



Air Pollution in Delhi

AQI

Data Analysis

It was evident from experience that **Delhi AQI has improved during Lockdown.**

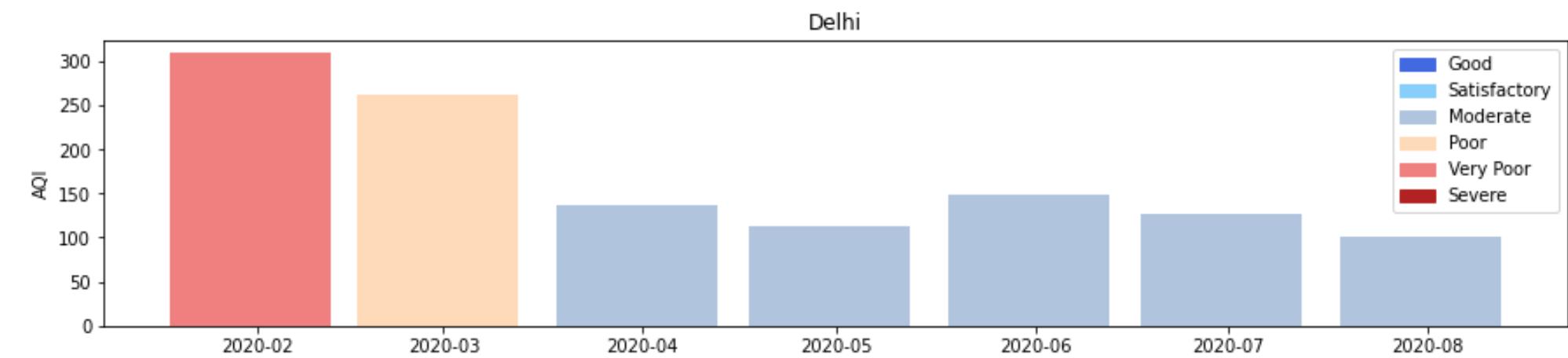


Air Pollution in Delhi

AQI

Data Analysis

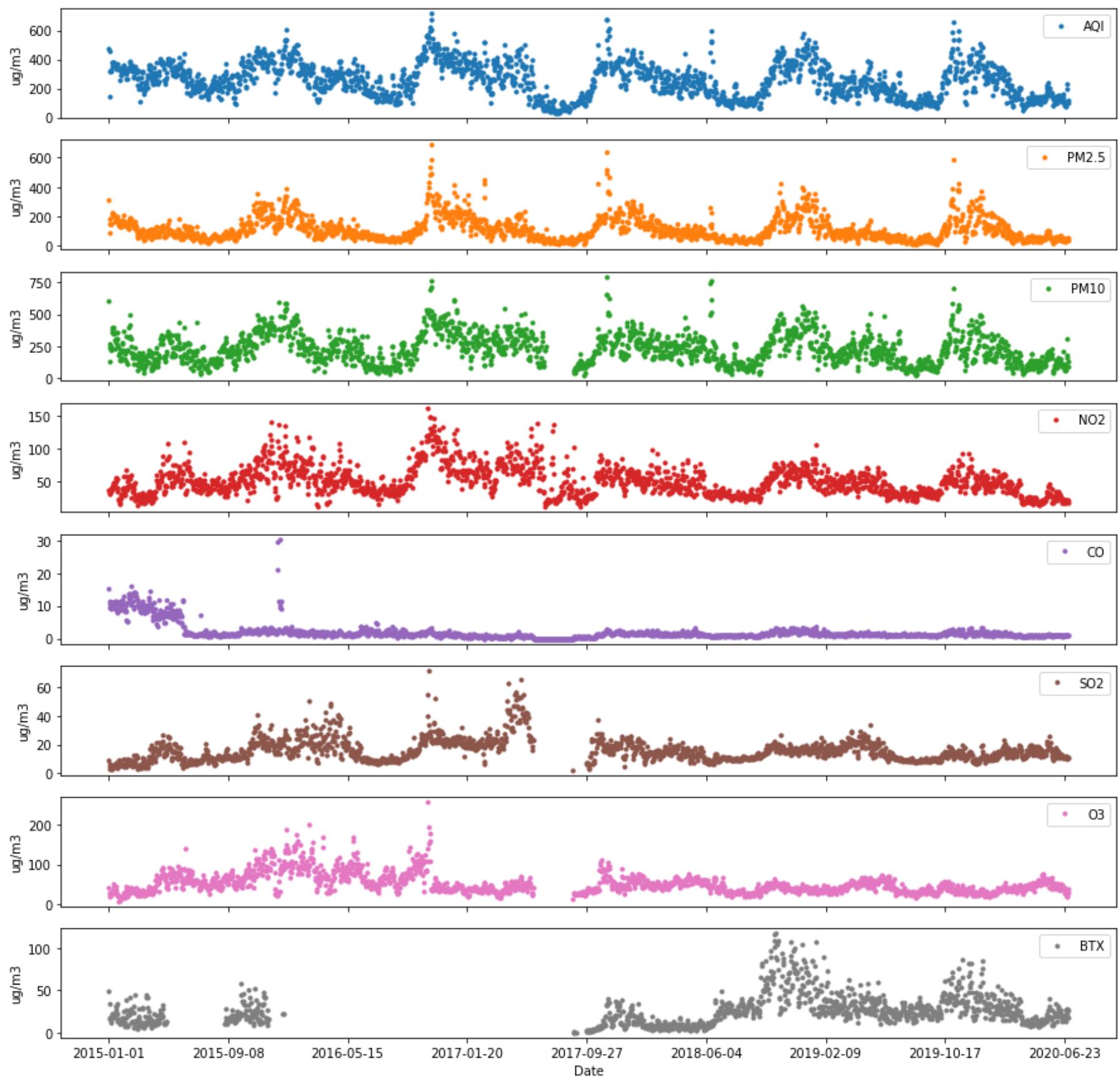
To understand the effects on AQI during **Lockdown**, we studied the **AQI levels in 2020 in Delhi.**



Air Pollution in Delhi

Preprocessing

Plotting entire data for Delhi for all pollutants.



Preprocessing

Data was checked for any **missing or NULL values**.

	Missing Values	% of Total Values
BTX	781	38.9
SO2	110	5.5
O3	84	4.2
PM10	77	3.8
AQI	10	0.5
AQI_Bucket	10	0.5
PM2.5	2	0.1
NO2	2	0.1

Preprocessing

Handling missing values in the dataset

- Deleting Rows with NULL Values
- Filling Missing Values in Database with the 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)

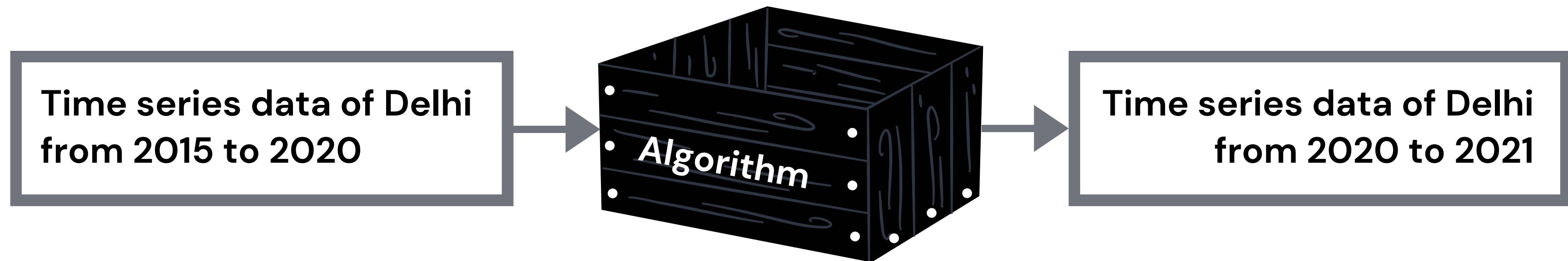
Air Pollution in Delhi

Coming Up...

- Handling Missing Values
- Predicting pollutant concentrations and AQI values for the upcoming year
- Looking into the effect of Lockdown on AQI values
- Visualisation of results
- Understanding the trend and seasonality in the data using environmental phenomena



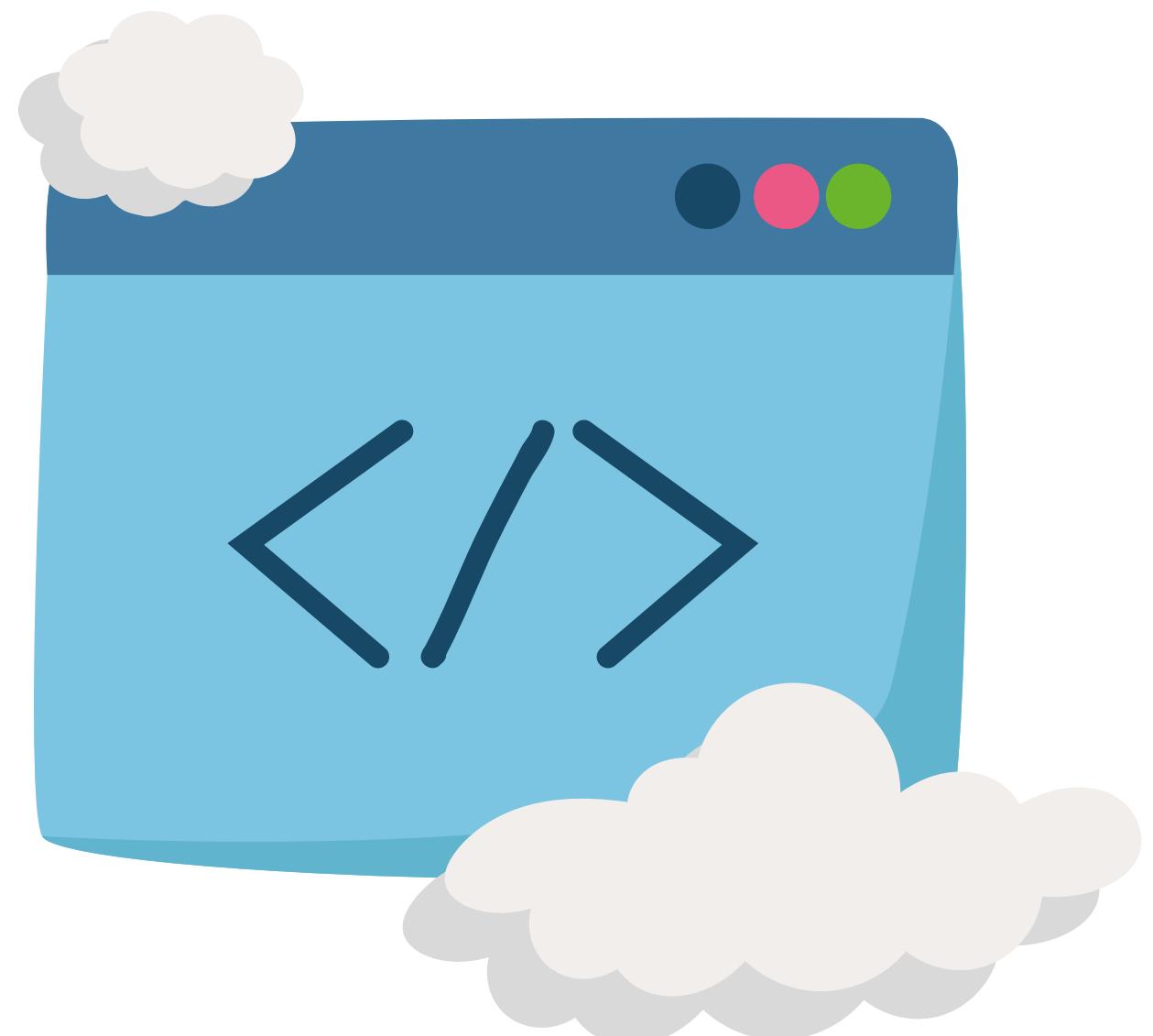
Input/Output: Predicting the unknown



Algorithms/Models

Play around with time series algorithms like:

- SARIMAX
- LSTM



Thank you!

Questions?

Harshita Sharma (20171099)
Abhigyan Ghosh (20171089)
Zubair Abid (20171076)