PREDICTION OF AQI FOR UPCOMING YEARS IN SANATHNAGAR , HYDERABAD - TSPCB

Internship project

of

BACHELOR OF TECHNOLOGY

in

CIVIL ENGINEERING

By

CHAUHAN PARTH DHARMENDRA (21CE01045)

Under the Supervision of

Dr. RAJESH ROSHAN DASH

Associate Professor

SCHOOL OF INFRASTRUCTURE

INDIAN INSTITUTE OF TECHNOLOGY BHUBANESWAR– 752050,ODISHA

May-June 2024

ABSTRACT

The Air Quality Index (AQI) serves as a vital metric for assessing air pollution levels and their impact on public health and the environment. This report aims to predict the AQI for Velachery over the upcoming years using comprehensive historical data and sophisticated predictive modeling techniques.

We employ various time series analysis methods, including ARIMA (Auto-Regressive Integrated Moving Average), as well as advanced machine learning algorithms such as Random Forest and FB Prophet model, to develop robust predictive models. These models are trained and validated using historical AQI data, meteorological variables, and emission inventories to ensure accuracy and reliability.

The predictive analysis indicates a continuing trend of high AQI values, suggesting that air quality may deteriorate further if current pollution levels persist. This projection underscores the urgent need for effective policy measures and interventions, including stricter emission controls, the promotion of sustainable transportation, and enhanced public awareness campaigns to mitigate air pollution.

In conclusion, the predictive models developed in this study provide valuable insights into Velachery's future air quality scenario. By highlighting the critical need for immediate and sustained action, this report aims to inform policymakers, stakeholders, and the general public, fostering a collaborative effort toward achieving cleaner air and a healthier environme

TABLE OF CONTENT

INTRODUCTION

- Air Quality Index (AQI) is defined as an overall scheme that transforms the weighted values of individual air pollution-related parameters (for example, pollutant concentrations) into a single number or set of numbers. The AQI communicates primarily a number starting from 0 and going up to 500 depending upon the method of calculation. E.g.:

    - Method by Tiwari and Ali (1987): AQI range (0 to 100+)
    - Geometric Mean Method: AQI range (0 to 125+)
    - Pollution Index Method. AQI range (0 to 175+)
    - CPCB Method: AQI range (0 to 500)

- AQI is commonly used to report the severity of air pollution to the public. The higher the number, the greater the health risk associated with the air quality.

Source www.cpcb.nic.in. Beig et al (2010))

Range	AQI Category as per INAQS	Associated Health Impacts
0-50	**Good**	Minimal Impact
51-100	**Satisfactory**	Minor breathing discomfort to sensitive people
101-200	**Moderate**	Breathing discomfort to the people with lung, heart disease, children and older adults
201-300	**Poor**	Breathing discomfort to people on prolonged exposure
301-400	**Very Poor**	Respiratory illness to the people on prolonged exposure
401-500	**Severe**	Respiratory effects even on healthy people

LITERATURE REVIEW

- Air Quality Index is helpful in many fields such as 1. Scientific Research, 2. Trend Analysis, 3. Resource Allocation, 4. Ranking of Locations, 5. Public Information, 6. Enforcement of Standards.

The sources of different pollutions are as follows in the table:

Class	Pollutants
Combustion processes (Cars, trucks, airplanes, and railways. Domestic burning, thermal power plants.)	SO_2, NO_2, CO, odors, organic vapors
Petroleum operations	SO_2, H_2S, NH_3, CO. hydro-carbons
Chemical processes (paper mills, cement, fertiliser, etc.)	Process-dependent (SO_2, CO, NH_3, NO_2, organic
Metallurgical processes (aluminum refineries, steel plants)	SO_2, CO, fluorides. organic vapors
Agricultural activities (Crop spraying, Field burning)	chlorinated hydro-carbons, sulfur oxides, organic vapors

Sources of Air pollution

- **Air Quality Index – Objective/Application**

1. **Resource Allocation**

❖ To assist administrators in allocating funds and determining priorities.

❖ Enable evaluation of trade-offs involved in alternative air pollution control strategies**.**

2. **Trend Analysis**

❖ To determine the change in air quality (degradation or D improvement) which have occurred over a specified period.

❖ This enables forecasting of air quality (i.e., tracking the behavior of pollutants in the air) and planning pollution control measures.

3. **Ranking of Locations**

❖ To assist in comparing air quality conditions at different locations/ cities. Thus, pointing out areas and frequencies of potential hazards.

4. **Public Information**

❖ To inform the public about environmental conditions.

❖ Useful for people who suffer from illness aggravated or caused by air pollution.

5. **Enforcement of Standards**

❖ To determine extent to which the legislative standards and existing criteria are being adhered. Also helps in identifying faulty standards and inadequate monitoring programs.

6. **Scientific Research**

❖ Helps scientists engaged in research using air quality data.

❖ Provides better insights to researchers while conducting studies of some environmental phenomena.

- **Air Quality Monitoring and AQI Consideration**

The air quality monitoring network in India includes both Online Monitoring and Manual monitoring.

1. **Online Monitoring Network**

- ❖ Automated air quality monitoring stations
- ❖ Continuous monitoring
- ❖ Real-time AQI computation

2. **Manual Monitoring Network**

- ❖ Manual air quality monitoring
- ❖ Intermittent monitoring
- ❖ Not suitable for AQI calculation for quick interpretation
- ❖ Historical AQIs on a weekly basis for data interpretation

- **AQI Formation**

Air Quality can be determined requires 2 steps:

i. Formation of sub-indices (for each pollutant)

ii. Aggregation of sub-indices to get an overall AQI

Step 1: The general equation for the sub-index for a given pollutant concentration is as follows:-

The general equation for sub-indexes ,for a given pollutant concentration as based on ***linear segmented principle*** is calculated as :-

$$AQI = \frac{(PM_{Obs} - PM_{Min}) \times (AQI_{Max} - AQI_{Min})}{PM_{Max} - PM_{Min}} + AQI_{Min}$$

With

$AQI = Air\ quality\ index$

$PM_{Obs} = observed\ 24 - hour\ average\ concentration\ in \frac{\mu g}{m3}$

$PM_{Max} = maximum\ concentration\ of\ AQI\ color\ category\ that\ contains\ PM_{Obs}$

$PM_{Min} = minimum\ concentration\ of\ AQI\ color\ category\ that\ contains\ PM_{Obs}$

$AQI_{Max} = maximum\ AQI\ value\ for\ color\ category\ that\ corresponds\ to\ PM_{Obs}$

$AQI_{Min} = minimum\ AQI\ value\ for\ color\ category\ that\ corresponds\ to\ PM_{Obs}$

AQI Category	AQI	Concentration range*							
		PM_{10}	$PM_{2.5}$	NO_2	O_3	CO	SO_2	NH_3	Pb
Good	0 - 50	0 - 50	0 - 30	0 - 40	0 - 50	0 - 1.0	0 - 40	0 - 200	0 - 0.5
Satisfactory	51 - 100	51 - 100	31 - 60	41 - 80	51 - 100	1.1 - 2.0	41 - 80	201 - 400	0.5 - 1.0
Moderately polluted	101 - 200	101 - 250	61 - 90	81 - 180	101 - 168	2.1 - 10	81 - 380	401 - 800	1.1 - 2.0
Poor	201 - 300	251 - 350	91 - 120	181 - 280	169 - 208	10 - 17	381 - 800	801 - 1200	2.1 - 3.0
Very poor	301 – 400	351 - 430	121 - 250	281 - 400	209 - 748*	17 - 34	801 - 1600	1200 -1800	3.1 - 3.5
Severe	401 - 500	430+	250+	400+	748+*	34+	1600+	1800+	3.5+

* CO in mg/m^3 and other pollutants in $\mu g/m^3$; 2h-hourly average values for PM_{10}, $PM_{2.5}$, NO_2, SO_2, NH_3, and Pb, and 8-hourly values for CO and O_3.

Step : 2. Aggregation of sub-indices to get an overall AQI

- Aggregation of sub-indices, I_{i} is carried out with a mathematical function (described below) to obtain the overall index (1), referred to as AQI.

- I= F(I_{1}, I_{2} ,...I n)

- The aggregation function usually is a summation or multiplication operation or simply a maximum operator.

Aggregation functions can be:

- **Weighted Additive Form**

I = Aggregated Index **=** $\sum W_i I_i$ **(For i = 1, 2, ... n), where,**

$\sum w_1 = 1$ (weight given to all the pollutants must be added to 1)
1_1 = sub-index for pollutant I
n=number of pollutant variables
w_1 = weightage of the pollutant

- **Root-Sum-Power Form (non-linear aggregation form)**

I = Aggregated Index = $[\sum I_i^P]^{(1/p)}$

where p is the positive real number > 1.

- **Root-Mean-Square Form**

I= Aggregated index = $\{1/k(I_1^2+I_2^2+....I_n^2)\}^{0.5}$

- **Min or Max Operator**

I = Min or Max $(I_1, I_2, ...)$

- **Eclipsing and Ambiguity**

  - Two important characteristics, eclipsing and ambiguity are significant to interpret any index in the right perspective.
  - Eclipsing occurs when an air pollution index does not indicate poor air quality even though concentrations of one or more air pollutants may have reached unacceptably high values.
  - Pollution is underestimated by AQI if there is eclipsing.
  - Ambiguity occurs when an air pollution index gives a false alarm even though concentrations of all the pollutants are within the permissible limit except for one pollutant.
  - Pollution is overestimated by AQI if there is ambiguity.
  - To remove the Eclipsing and Ambiguity, new indices that have been proposed are not of additive or multiplicative type.
  - A maximum operator approach is adopted to remove Ambiguity and Eclipsing.
  - AQI= Max (I1, I2, ... In)
  - The health effects of the combination of pollutants are not known and thus in the health-based index, sub-indices cannot be added or multiplied.

METHODOLOGY

This study aims to predict the Air Quality Index (AQI) for the upcoming years using historical air pollution data collected from the Central Pollution Control Board (CPCB) website for a sampling station in Chennai. The methodology involves several key steps, including data collection, preprocessing, calculation of sub-indexes, data analysis, and the development of a machine learning model to predict future AQI values.

1. Data Collection:

Data for this study was collected from the CPCB website, covering 8 years. The dataset includes annual concentrations of major pollutants such as PM2.5, PM10, NOx, and SO2. The data was organized and stored in a Microsoft Excel sheet for further analysis.

2. Data Preprocessing:

To ensure data quality and accuracy, the following preprocessing steps were undertaken:

- **Handling Missing Data:** Missing values were identified and handled using appropriate imputation techniques or by removing incomplete records.
- **Normalization:** The pollutant concentration values were normalized to a standard scale to facilitate comparison and analysis.
- **Seasonal Segregation:** The data was segregated based on different seasons to analyze seasonal variations in pollutant levels.

3. Calculation of Sub-Indexes and AQI:

The AQI was calculated using the following steps:

- **Sub-Index Calculation:** For each pollutant, sub-index values were computed using the CPCB's prescribed formulae.
- **Aggregation:** The overall AQI was determined by aggregating the sub-indexes of individual pollutants using the maximum operator method, where the highest sub-index value among the pollutants was taken as the AQI for that period.

4. **Data Analysis:**

The processed data was analyzed to identify trends, patterns, and seasonal variations in pollutant concentrations over the 8 years. This analysis provided insights into the factors influencing air quality and helped in understanding the underlying trends in the data.

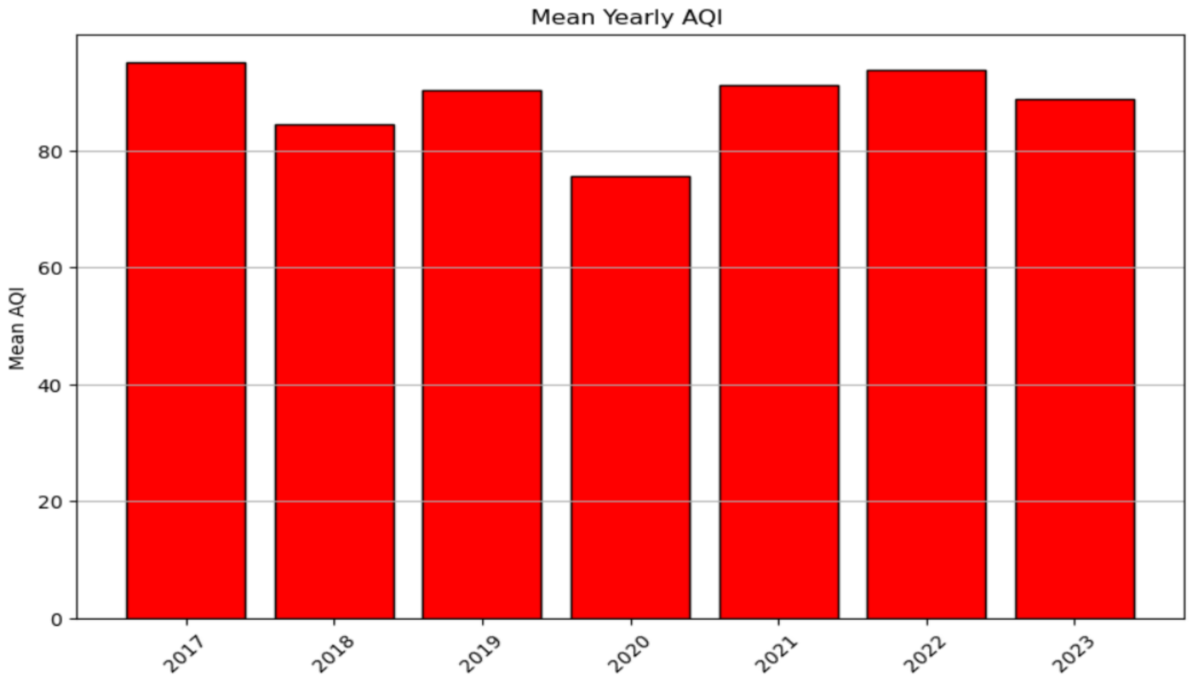
5. **Development of Machine Learning Model:**

A machine learning model was developed to predict the AQI for the next year based on historical data. The following steps were involved:

- **Feature Selection:** Relevant features such as historical pollutant concentrations, seasonal variations, and meteorological factors were selected for the model.
- **Model Selection:** Various machine learning algorithms were considered, including FB Prophet and ARIMA Model. After evaluating their performance, the best-performing model was chosen.
- **Training and Validation:** The selected model was trained using the historical data and validated using a portion of the dataset to assess its predictive accuracy.
- **Prediction:** The trained model was used to predict the AQI for the upcoming year, providing insights into potential future air quality scenarios.

Data Analysis of variation in AQI

Annual AQI Trends



- **2017**: The mean AQI was 95.12 with a standard deviation of 61.8. The maximum recorded AQI was 333.1.
- **2018**: The mean AQI increased to 84.5 with a standard deviation of 41.6. The maximum AQI recorded was 170.9.
- **2019**: The mean AQI slightly decreased to 90.38 with a higher standard deviation of 59.58. The maximum AQI recorded was 307.5.
- **2020**: The mean AQI was significantly lower at 75.58 with a standard deviation of 61.8. The maximum AQI recorded was 323.6
- **2021**: The mean AQI increased to 91.14 with a standard deviation of 64.43. The maximum AQI recorded was 312.28
- **2022**: The mean AQI slightly increased at 93.9with a standard deviation of 66.1. The maximum AQI recorded was 384.7
- **2023**: The mean AQI slightly decreased 88.8 with a standard deviation of 72.7. The maximum AQI recorded was 510

The observed trend in AQI, particularly the significant drop in 2020, can be attributed to a combination of factors, including the impact of the COVID-19 pandemic, climate changes, and specific local conditions. Here is a detailed explanation accounting for these factors:

1. Impact of COVID-19 Pandemic

Global Lockdowns and Reduced Human Activity:

- **Industrial Shutdowns**: Due to lockdown measures, many industries and factories were closed or operated at reduced capacity. This led to a significant decrease in industrial emissions, which are major contributors to air pollution.
- **Reduced Traffic**: With travel restrictions and work-from-home policies, the number of vehicles on the road decreased dramatically. Transportation is a major source of pollutants like NOx and particulate matter (PM2.5 and PM10).
- **Lowered Energy Consumption**: The energy demand dropped as commercial activities slowed down, leading to reduced emissions from power plants.

2. Climate Change and Seasonal Variations

Weather Patterns:

- **Monsoon and Rainfall**: 2020 and 2021 saw normal to above-normal monsoon rains in many parts of India, including Chennai. Rain helps in settling particulate matter and clearing pollutants from the air.
- **Wind Patterns**: Changes in wind patterns due to climatic variations can also influence pollutant dispersion. Favorable wind patterns can help disperse pollutants more effectively, leading to lower AQI levels.

Temperature Variations:

- **Lower Winter Emissions**: Winters in 2020 and 2021 might have seen fewer cold waves, leading to reduced heating requirements and subsequently lower emissions from heating sources.

3. Reduced International and Domestic Travel

Air Traffic:

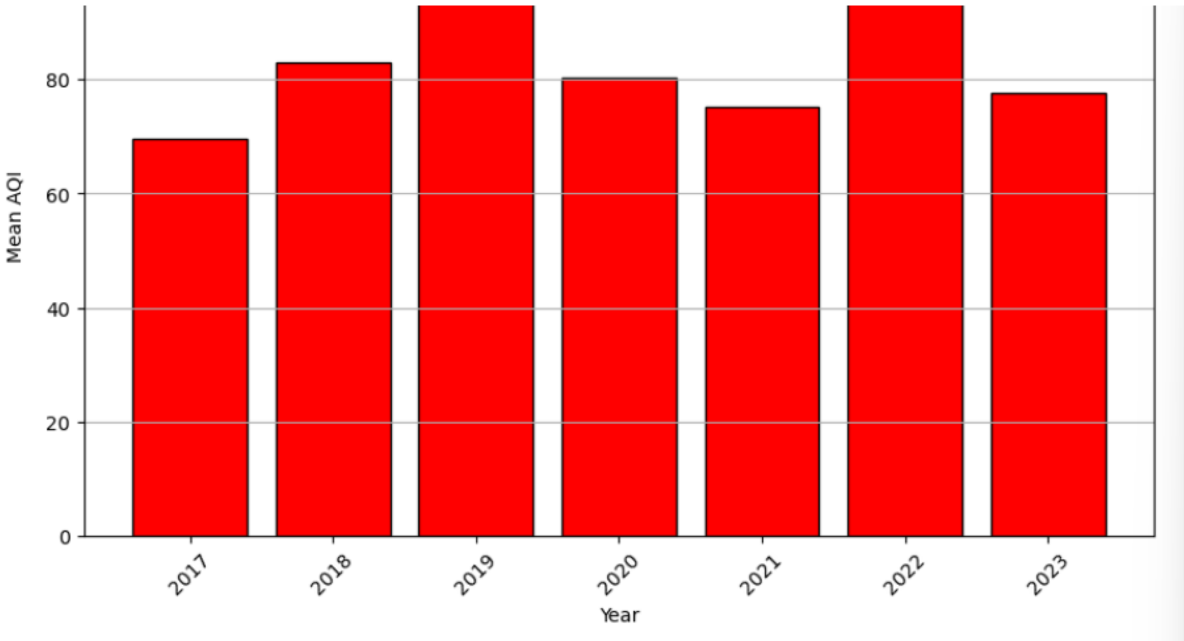
- **Reduced Flights**: Both international and domestic air travel saw a significant decline, reduction in flights would have contributed to better air quality.

Summary

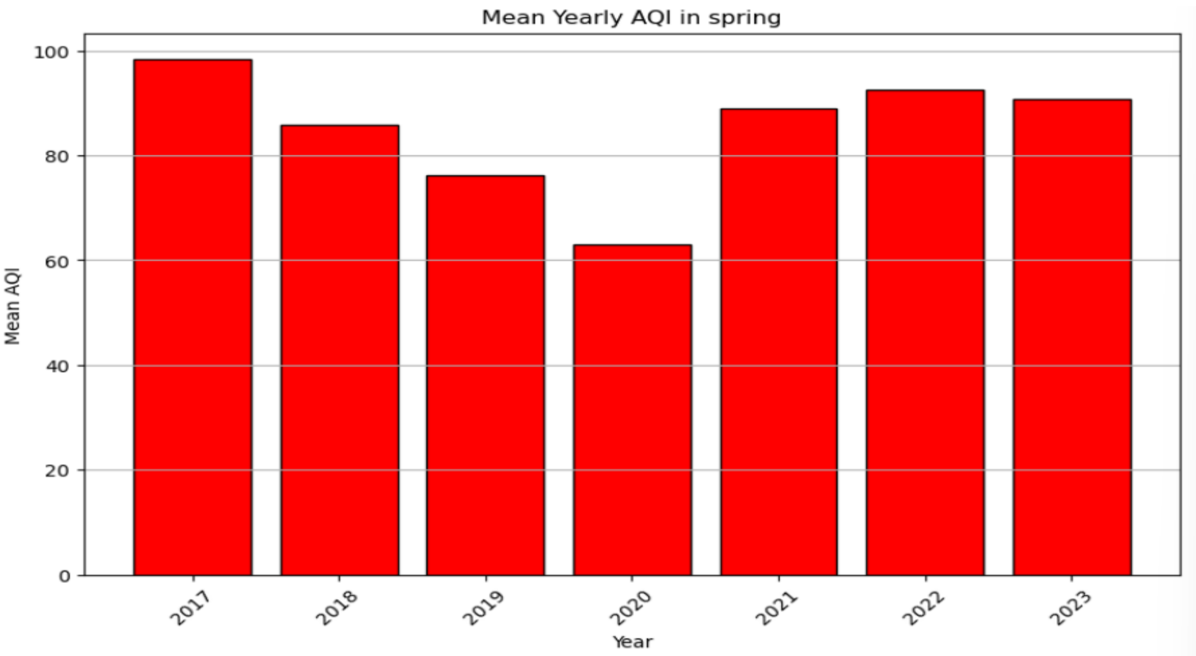
The drop in AQI in 2020 and 2021 can be largely attributed to the widespread impact of the COVID-19 pandemic, which led to a reduction in industrial activities, vehicular emissions, and energy consumption. Additionally, favorable climatic conditions, such as adequate rainfall and favorable wind patterns, further helped in dispersing pollutants. Stricter environmental regulations and improved compliance also played a role in reducing pollution levels. These combined factors led to a significant improvement in air quality, as reflected in the lower AQI values during these years.
This analysis underscores the interplay between human activities, regulatory measures, and natural climatic conditions in determining air quality trends.

Seasonal AQI Trends

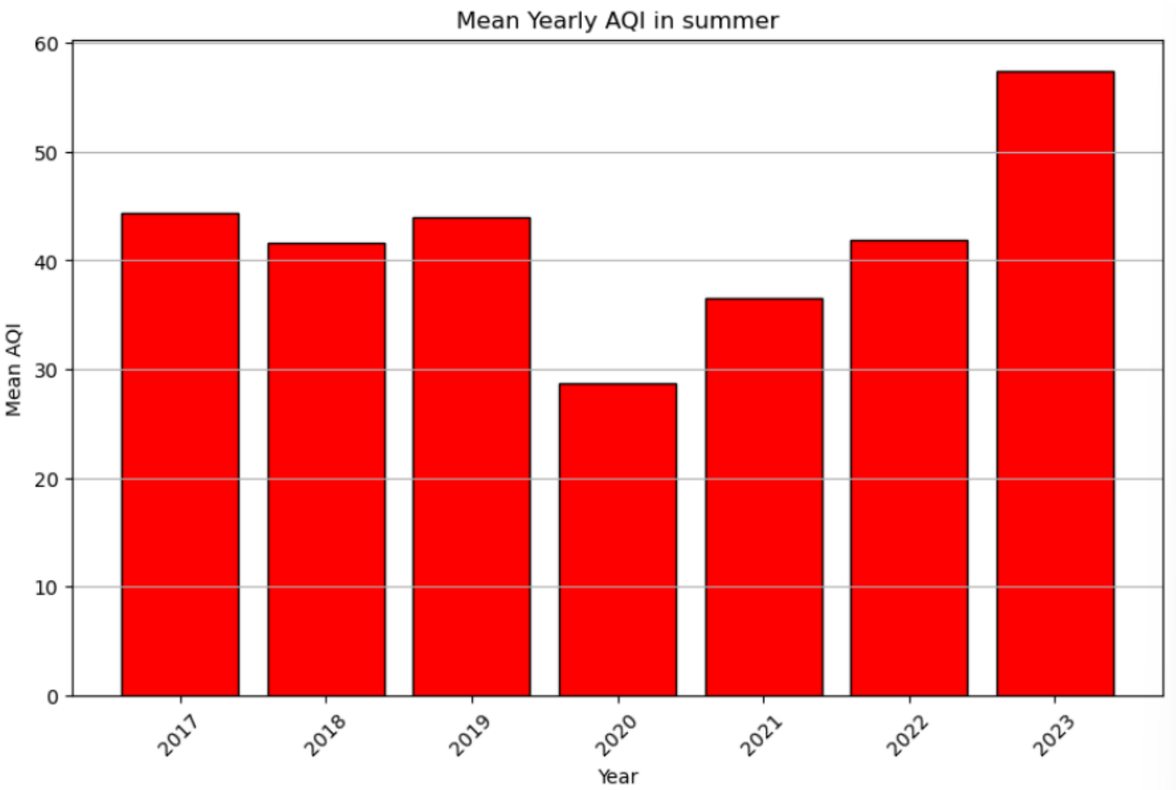
- **Autumn**: The mean AQI was 85.6



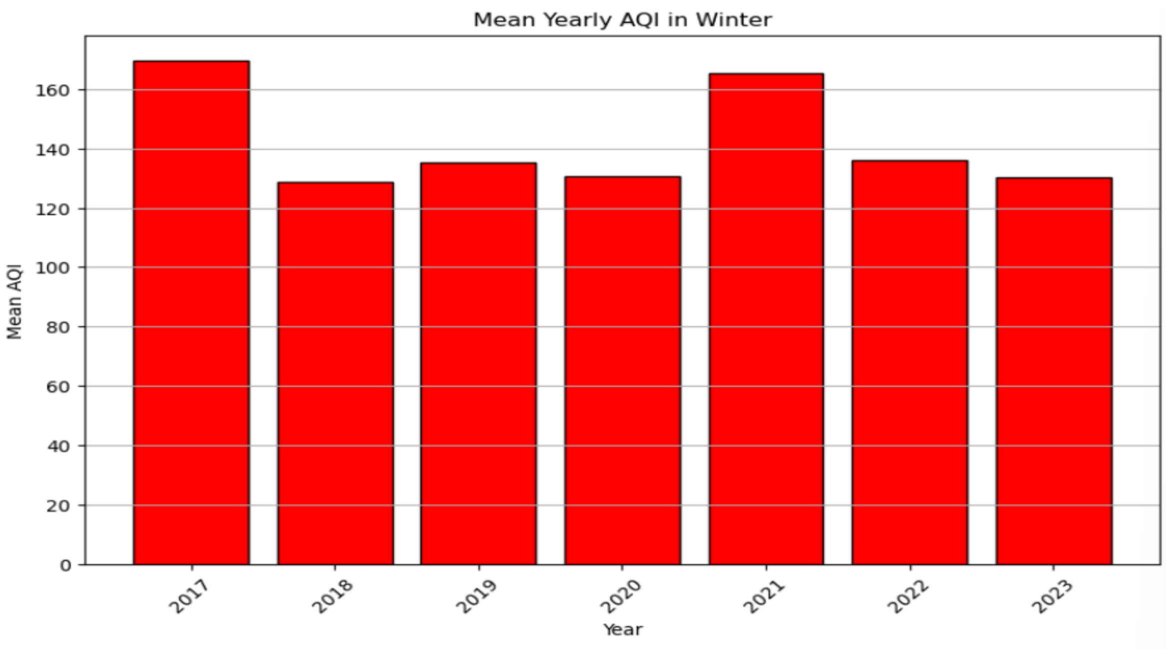
- **Spring**: The mean AQI was 85.07.



- **Summer**: The mean AQI was 42.06.



- **Winter**: The mean AQI was the highest at 142.3



The AQI is generally higher in winter compared to other seasons due to a combination of meteorological, environmental, and human factors. Here's an in-depth explanation of why air quality tends to deteriorate during the winter season:

1. **Meteorological Factors**

Temperature Inversion:

- **Inversion Layers**: During winter, temperature inversions are more common. Normally, the air temperature decreases with altitude, but during an inversion, a layer of warmer air traps pollutants close to the ground, preventing them from dispersing. This leads to higher concentrations of pollutants near the surface.
- **Stable Atmospheric Conditions**: Winter often brings stable atmospheric conditions with little wind. This stability means that pollutants emitted into the air remain concentrated in one area instead of being dispersed by wind.

Lower Mixing Heights:

- **Reduced Vertical Mixing**: The atmospheric mixing height, which is the layer of the atmosphere where vertical mixing of air occurs, is lower in winter. This limits the dispersion of pollutants vertically, leading to higher concentrations near the ground.

2. **Increased Emissions**

Heating Sources:

- **Increased Use of Heating**: In colder weather, there is a higher use of heating systems, including wood stoves, fireplaces, and space heaters. Many of these sources emit particulate matter and other pollutants, contributing to higher AQI values.
- **Biomass Burning**: In some regions, the burning of biomass for heating is common, which releases a significant amount of particulate matter and other pollutants into the air.

3. **Human Activities**

Increased Vehicle Emissions:

- **Cold Starts**: Vehicles emit more pollutants during cold starts, which are more frequent in winter. The cold temperatures affect the efficiency of catalytic converters in vehicles, leading to higher emissions of CO, NOx, and hydrocarbons.

Festivals and Agricultural Practices:

- **Burning of Agricultural Residue**: In certain regions, the burning of agricultural residue post-harvest is common in late autumn and early winter. This practice significantly contributes to air pollution.

- **Fireworks and Festivities**: In some parts of the world, winter coincides with major festivals that involve the use of fireworks, which release a variety of pollutants into the air.

4. Natural Factors

Lower Photochemical Activity:

- **Reduced Sunlight**: Shorter daylight hours and lower sun angles in winter reduce the photochemical reactions that can help break down certain pollutants, such as ozone. This can result in the accumulation of pollutants.

Dry Conditions:

- **Less Precipitation**: Winter is often drier than other seasons, with less precipitation to wash away pollutants from the atmosphere.

Summary

The combination of temperature inversions, lower mixing heights, increased use of heating sources, higher vehicle emissions, specific human activities such as biomass burning and fireworks, and reduced photochemical activity all contribute to higher AQI levels in winter. These factors lead to the accumulation and persistence of pollutants in the lower atmosphere, resulting in poorer air quality compared to other seasons.
Understanding these factors can help in planning and implementing measures to mitigate air pollution, such as restricting biomass burning, promoting cleaner heating technologies, and increasing public awareness about the impacts of winter pollution.

Categorized AQI Levels

- **Good AQI**:
  - Count: 835
  - Mean: 34.7
  - Standard Deviation: 8.855
  - Minimum: 0.64
  - Maximum: 50.00
- **Satisfactory AQI**:
  - Count: 984
  - Mean: 73.7
  - Standard Deviation: 13.7
  - Minimum: 50.01
  - Maximum: 99.7
- **Moderately Polluted AQI**:
  - Count: 571
  - Mean: 141.75
  - Standard Deviation: 25.9
  - Minimum: 100.035
  - Maximum: 200
- **Poor AQI:**
  - Count: 147
  - Mean: 240.8
  - Standard Deviation: 29.83
  - Minimum: 200.1
  - Maximum: 299.86
- **Very Poor AQI**:
  - Count: 23
  - Mean: 323.256
  - Standard Deviation: 22.94
  - Minimum: 300.16
  - Maximum: 384.64
- **Severe AQI**:
  - Count: 2
  - Mean: 406.04
  - Standard Deviation: 0.164
  - Minimum: 405.067
  - Maximum: 406.471

Results and Discussion

The analysis indicates a general trend of improving air quality from 2017 to 2020, with a noticeable dip in AQI levels in 2020. Seasonal analysis reveals that winter months typically experience higher AQI levels, suggesting worse air quality during this period. The categorized AQI levels provide insights into the distribution of air quality across different pollution levels, with most data points falling within the "Good" and "Satisfactory" categories

.

Conclusion

This analysis of AQI data over various years and seasons highlights significant trends and variations in air quality. The development of a machine learning model based on this historical data aims to provide accurate predictions for future AQI levels, assisting in proactive environmental management and policy-making.

Development of Machine Learning Model

A machine learning model was developed to predict the AQI for the next year based on historical data. The following steps were involved:

- **Feature Selection:** Relevant features such as historical pollutant concentrations, seasonal variations, and meteorological factors were selected for the model.
- **Model Selection:** Various machine learning algorithms, including FB Prophet , ARIMA Model were considered. After evaluating their performance, the best-performing model was chosen.
- **Training and Validation:** The selected model was trained using the historical data and validated using a portion of the dataset to assess its predictive accuracy.
- **Prediction:** The trained model was used to predict the AQI for the upcoming year, providing insights into potential future air quality scenarios.

Forecasting Air Quality with ARIMA and Facebook Prophet: A Mathematical Peek

Air quality forecasting plays a crucial role in public health and environmental management. Two popular models, ARIMA and Facebook Prophet, offer distinct approaches to predict future Air Quality Index (AQI) values. Let's delve into the basic math behind these models and see how they tackle AQI forecasting.

1. **ARIMA: Capturing Trends and Seasonality (ARIMA(p, d, q))**

ARIMA stands for AutoRegressive Integrated Moving Average. It's a statistical model that analyzes past AQI values to predict future ones. We can break down its components:

- **AutoRegressive (AR):**
  This term considers the influence of p past AQI values on the current value. Imagine AQI today depends on yesterday's AQI, the day before yesterday's, and so on (up to p terms). Higher p captures more historical influence.
- **Integrated (I):**
  If there's a trend (increasing or decreasing AQI over time), the data is differenced (subtracting the previous value from the current value) d times to make it stationary (no trend). The number of differences needed (d) is determined statistically using tests like the Augmented Dickey-Fuller test.
- **Moving Average (MA):**
  This considers the average of past forecast errors (q terms) to account for randomness in the data. It smooths out fluctuations and improves prediction accuracy.

The magic lies in finding the optimal combination of p, d, and q for your specific data. This is achieved through statistical tests like the Akaike Information Criterion

(AIC) or the Bayesian Information Criterion (BIC). Software packages handle these calculations, but understanding the concepts helps interpret the results.

Finding Optimal p, d, and q with auto_arima:
Python libraries like pmdarima offer functions to automate finding the best ARIMA parameters. Here's an example using auto_arima:
Python

```
“from pmdarima import auto_arima
stepwise_fit = auto_arima(data_x["Overall AQI"], trace=True,
suppress_warnings=True)”
```

This code snippet will:

1. Import the auto_arima function from the pmdarima library.
2. Pass your AQI data (stored in the "Overall AQI" column of your dataframe data_x) to the function.
3. Set trace=True to display the fitting process for better understanding.
4. Set suppress_warnings=True to silence potential warnings during the process.

The auto_arima function will then use a stepwise approach to identify the best combination of p, d, and q for your data, following statistical criteria like AIC or BIC. It will iterate through different models and choose the one with the lowest information criterion.

2. **Facebook Prophet:**

Embracing Holidays and Events
Facebook Prophet takes a different approach. It's a versatile forecasting model that excels at handling various data patterns, including seasonality, holidays, and other events. Here's a simplified breakdown:

- Trend: Prophet models a linear or non-linear trend in the AQI data.
- Seasonality: It captures recurring patterns like daily, weekly, or yearly cycles in AQI variations.
- Holidays: You can specify holidays or events that might influence AQI (e.g., festivals with increased firecrackers). Prophet accounts for their potential impact.
- Growth: This allows for modeling exponential growth or decay in AQI, if present.

Prophet requires less statistical expertise compared to ARIMA. You provide your AQI data and any relevant holiday information, and the model does the heavy lifting.

3.Forecasting AQI with ARIMA and Prophet

Both models can be used for AQI forecasting. Here's a general process:

1. Data Collection: Gather historical AQI data, ideally with high granularity (e.g., hourly).
2. Data Cleaning: Ensure data quality by addressing missing values and outliers.
3. Model Training:
    - ARIMA: Use auto_arima or a similar function to find the optimal p, d, and q values. Train the ARIMA model on your historical AQI data.
    - Prophet: Provide AQI data and any holiday information to Prophet.
4. Model Evaluation: Evaluate the model's performance on a hold-out validation set (data not used for training). Metrics like Mean Squared Error (MSE) can assess prediction accuracy.
5. Forecasting: Once satisfied with the model's performance, use it to predict future AQI values.

Some mathematical concepts to go through:

1. Introduction

Time series forecasting is a critical aspect of data analysis used in numerous fields such as finance, economics, and meteorology. One popular method for time series forecasting is ARIMA (Auto Regressive + Integrated + Moving Average). This report delves into the mathematical concepts behind ARIMA.

2. ARIMA (Auto Regressive Integrated Moving Average)

ARIMA is a widely used statistical method for time series forecasting. It combines three key components: Auto Regression (AR), Integration (I), and Moving Average (MA).

2.1. Components of ARIMA

2.1.1. Auto Regression (AR)

The Auto Regression part involves regressing the variable on its own lagged (past) values.

Mathematically, an AR(p) model is defined as:

x_t = c + sum(phi_i * x_(t-i)) + e_t

Where:
- x_t is the value at time t.
- c is a constant.
- ϕ_i are the parameters of the model.
- e_t is white noise.

2.1.2. Integration (I)

Integration is used to make a non-stationary time series stationary. This is achieved by differencing the data.

Mathematically, the differencing of order d is:

y_t = x_t - x_(t-1)

(for first order differencing)
Where:
- d is the number of times the differencing is applied.

2.1.3. Moving Average (MA)

The Moving Average part models the error term as a linear combination of error terms at previous times.

Mathematically, an MA(q) model is defined as:

x_t = c + e_t + *sum* (theta_j * e_(t-j))

Where:
- theta_j are the parameters of the model.

2.2. ARIMA Model

Combining these components, an ARIMA(p, d, q) model is represented as:

(x_t - x_(t-*d*)) = c + *sum*(phi_i * x_(t-i)) + e_t + *sum*(theta_j * e_(t-j))

Where:
- p is the order of the AR part.
- d is the degree of differencing.
- q is the order of the MA part.

2.3. Estimation and Forecasting

The parameters phi_i and theta_j are typically estimated using methods like Maximum Likelihood Estimation (MLE). Forecasting involves using the fitted model to predict future values.

3. Conclusion

ARIMA is a powerful tool for time series forecasting, particularly well-suited for short-term forecasting and grounded in traditional statistical methods. Understanding the mathematical foundations of the ARIMA model helps in applying it effectively for specific forecasting needs.

4.Evaluation:

The performance of the machine learning model was evaluated using metrics such as Root Mean Squared Error (RMSE), to ensure its reliability and accuracy in predicting AQI values.

5.Conclusion:

By employing historical pollutant data and advanced machine learning techniques, this study aims to provide accurate predictions of future AQI levels in Hyderabad. The findings can inform policymakers and stakeholders, enabling them to implement effective measures to improve air quality and protect public health.

Results:

From ARIMA MODEL on conducting test :

RMSE :- 77.74
MEAN :- 88.48

From FB Prophet model:

RMSE :- 70.30
MEAN:- 88.83

REFERENCES

- Central Pollution Control Board. **AQI Calculator**. Central Pollution Control Board, n.d., https://cpcb.nic.in/upload/national-air-quality-index/AQI-Calculator.xls. Accessed 3 June 2024.
- Central Pollution Control Board. **AQI India**. Central Pollution Control Board, n.d., https://airquality.cpcb.gov.in/AQI_India/. Accessed 5 June 2024.
- "**Air Quality Indices: A Review of Methods to Interpret Air Quality Status**." *ResearchGate*, 2020, https://www.researchgate.net/publication/343896852_Air_quality_indices_A_review_of_methods_to_interpret_air_quality_status. Accessed 14 June 2024.
- "**Air Quality Index - A Comparative Study for Assessing the Status of Air Quality**." *ResearchGate*, 2016, https://www.researchgate.net/publication/292643438_Air_Quality_Index_-_A_Comparative_Study_for_Assessing_the_Status_of_Air_Quality. Accessed 21 June 2024.
- Puja.P.Pathak -"**Time Series Forecasting — A Complete Guide**." *Medium*, Year, https://medium.com/analytics-vidhya/time-series-forecasting-a-complete-guide-d963142da33f. Accessed 29 June 2024.
- Prabhakaran, Selva. "**ARIMA Model – Complete Guide to Time Series Forecasting in Python**." *MachineLearningPlus*, Year, https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/. Accessed 10 July 2024.
- Facebook. "**Quick Start: Prophet Documentation**." *Facebook*, n.d., https://facebook.github.io/prophet/docs/quick_start.html. Accessed 10 July 2024.