DATA VISUALIZATION

# TOPIC:
# AIR QUALITY MONITORING AND FORECASTING

BY CHARAN SB
27PG00072

UNDER THE GAIDENCE OF
DR BHANU.KN

# THE FUTURE EARTH/LOCATIONS?

# INTRODUCTION

- Air pollution has become one of the most pressing global concerns due to its serious impact on human health, the environment, and climate change. Monitoring and predicting air quality is essential for public safety, policy planning, and environmental sustainability

- This project aims to perform a comprehensive analysis of air quality using a real-world dataset containing pollutant concentrations and meteorological parameters such as temperature and humidity.

- The dataset includes time-stamped measurements of key pollutants PM2.5, PM10, $O_3$, $NO_2$, $SO_2$, CO along with temperature, humidity, and the Air Quality Index (AQI).followed by exploratory visualizations to understand pollutant patterns and relationships.

- Linear Regression and Random Forest Regressor models to predict AQI based on pollutant concentrations and weather features. Furthermore, we use Facebook Prophet, a time series forecasting model, to project future AQI trends, enabling proactive environmental planning.

- we visualize historical pollutant trends and AQI distribution using multi-line plots, subplots, and pie charts. This gives deeper insight into the temporal dynamics of pollution levels and the proportion of days falling into different AQI categories such as Good, Moderate, and Poor.

# Key Insights

- Missing data handled smoothly without dropping rows.
- Pollutants like PM2.5, $NO_2$, and $O_3$ showed significant influence on AQI in pairplot.
- Random Forest model outperformed Linear Regression, indicating the importance of non-linear relationships in air quality data.
- Prophet forecasting provided a reliable method to predict future AQI trends with uncertainty bounds.
- Pollutant trends over time showed fluctuations likely due to environmental or seasonal changes.
- Most AQI values fall into Moderate to Poor, indicating potential public health concerns.

# METHODOLOGY

Exploratory Data Analysis (EDA)

- Display basic info (shape, nulls).
- Use Seaborn Pairplot to visualize relationships between pollutants and AQI.

Linear Regression

- Train a model using pollutants as input and AQI as output.
- Predict AQI and plot actual vs predicted results.

Random Forest Regressor

- Use a more robust model (Random Forest).
- Compare prediction accuracy with Linear Regression.

Time Series Forecasting (Using Prophet)

- Use Facebook Prophet to forecast future AQI values.
- Generate future AQI predictions

Pollutant Trend Visualization

- Plot $O_3$, $NO_2$, $SO_2$, CO trends over time using subplots

AQI Classification & Pie Chart

- Categorize AQI into: Good, Moderate, Poor.
- Count each category and visualize using a pie chart.

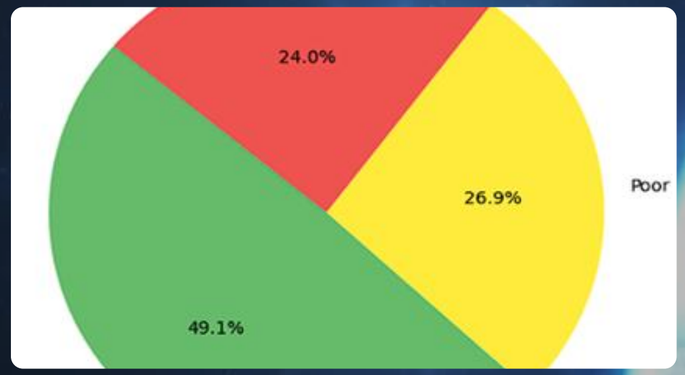
# DATASET OVERVIEW : AIR QUALITY MONITORING DATA

**The dataset is a time-series environmental dataset that captures atmospheric conditions and pollution levels across timestamps and locations. It includes key air pollutants and meteorological factors that directly influence the Air Quality Index (AQI).**

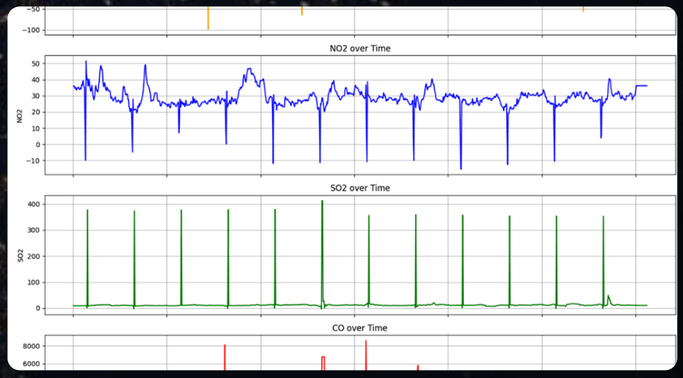
| Column Name |
| --- |
| location_id |
| datetimeUtc |
| pm2.5 |
| pm10 |
| o3 |
| no2 |
| so2 |
| co |

EVERY SMALL ACTION MATTERS

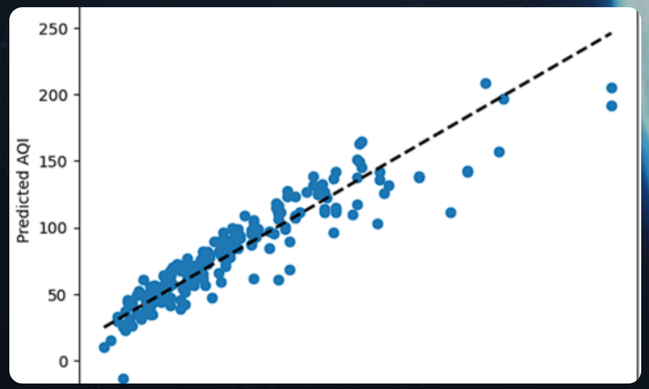
# PROTECT OUR PLANET



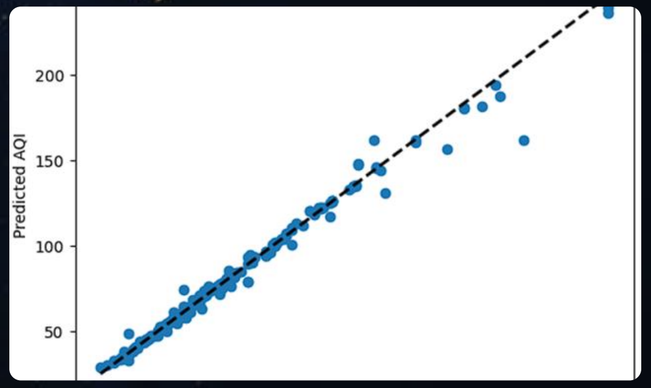

PAI CHART AQI Distribution — Good vs Moderate vs Poor



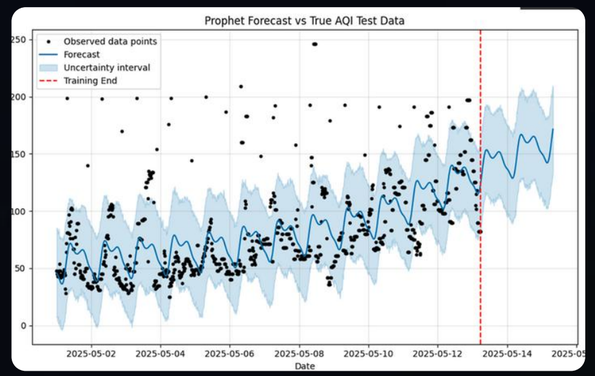

Time Series Analysis: $O_3$, $NO_2$, $SO_2$, and CO




Scatter plot you provided shows Actual vs Predicted AQI




Prophet Forecast: Future AQI Predictions



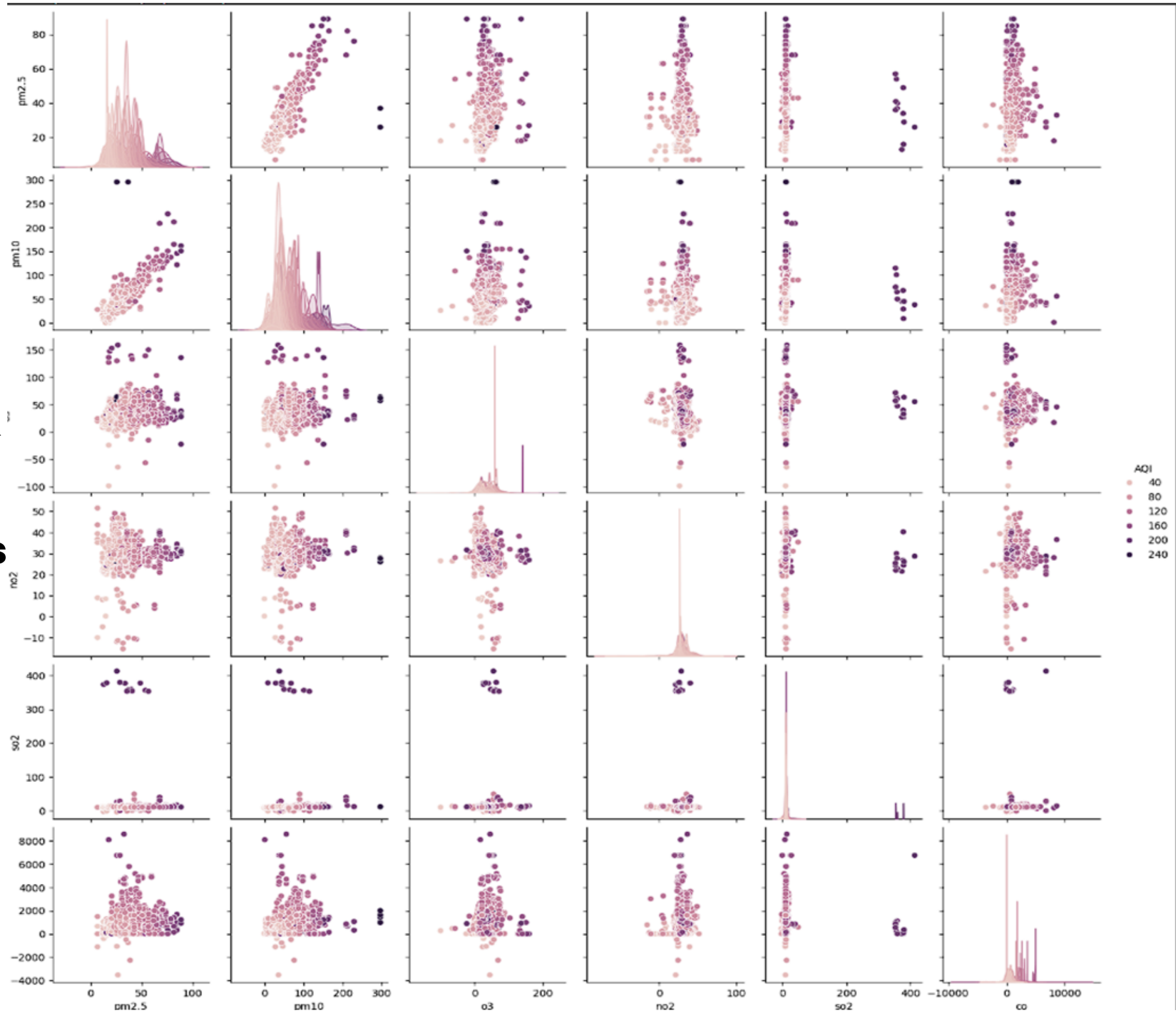

Prophet Forecast vs True AQI Test Data

# Multivariate Pairplot of Air Quality Indicators

- The pairplot shown provides a multivariate correlation analysis between key air pollutants
- PM2.5, PM10, $O_3$, $NO_2$, $SO_2$, CO with the Air Quality Index (AQI) represented by color intensity.
- The pairplot helps identify how strongly different air quality variables are related to each other and how they relate to the Air Quality Index (AQI) (shown by color intensity).

## How the Pairplot Helps in AQI Analysis

- Identifying Key Pollutant Relationships with AQI
- The color hue in the pairplot represents AQI levels.
- By observing which pollutants show darker color points at higher values, we can identify which pollutants influence AQI the most.
- For example:
- PM2.5 and PM10 show strong correlation with AQI higher values are associated with poor air quality.
- This helps in feature selection when building

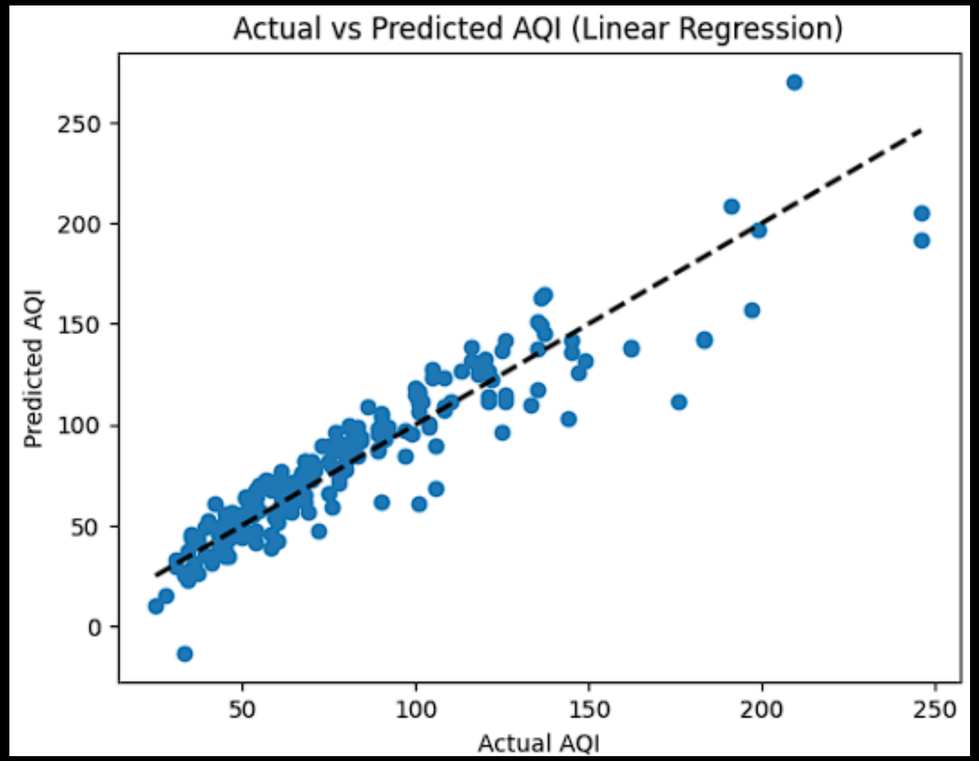


# ACTUAL VS PREDICTED AQI (LINEAR REGRESSION)

This scatter plot shows the performance of a Linear Regression model in predicting Air Quality Index (AQI) based on pollutant and weather data.

The x-axis shows the actual AQI (real values from the dataset).
The y-axis shows the predicted AQI (what the model predicted).
Ideally, if predictions are perfect, all dots would lie exactly on the dashed black line.

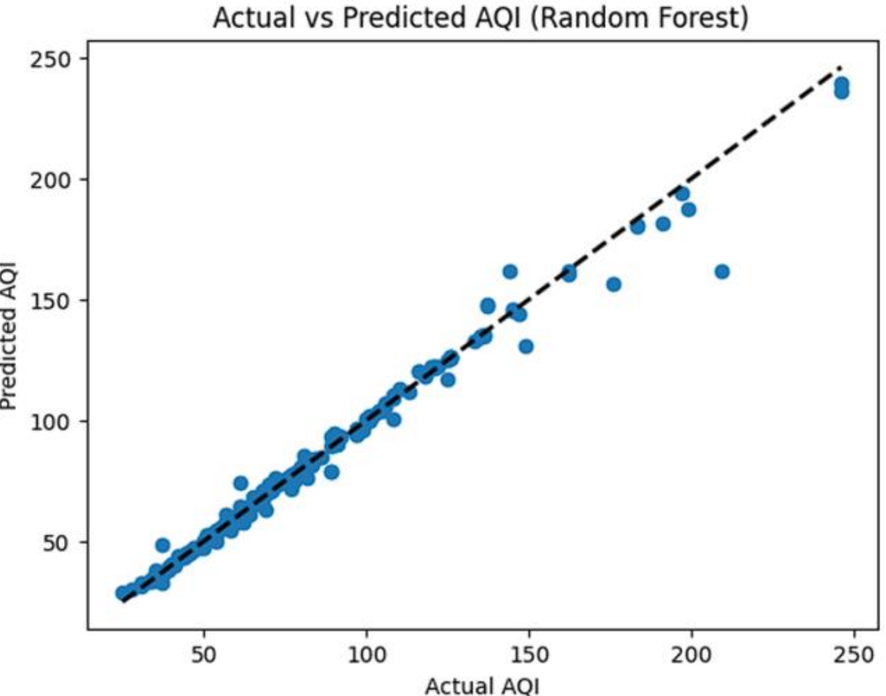
| Benefit |
| --- |
| 🎯 Model Evaluation |
| 🚨 Health Safety |



# Actual vs Predicted AQI (Random Forest)

- This plot visually compares the actual Air Quality Index (AQI) values with the predicted AQI values generated by a Random Forest Regressor model. It's a key tool for evaluating model accuracy and reliability in air quality forecasting.

- X-axis: Actual AQI values (true values from the dataset).
- Y-axis: Predicted AQI values (from the model).
- Blue dots: Individual predictions by the model.
- Dashed line: Ideal scenario (perfect prediction)



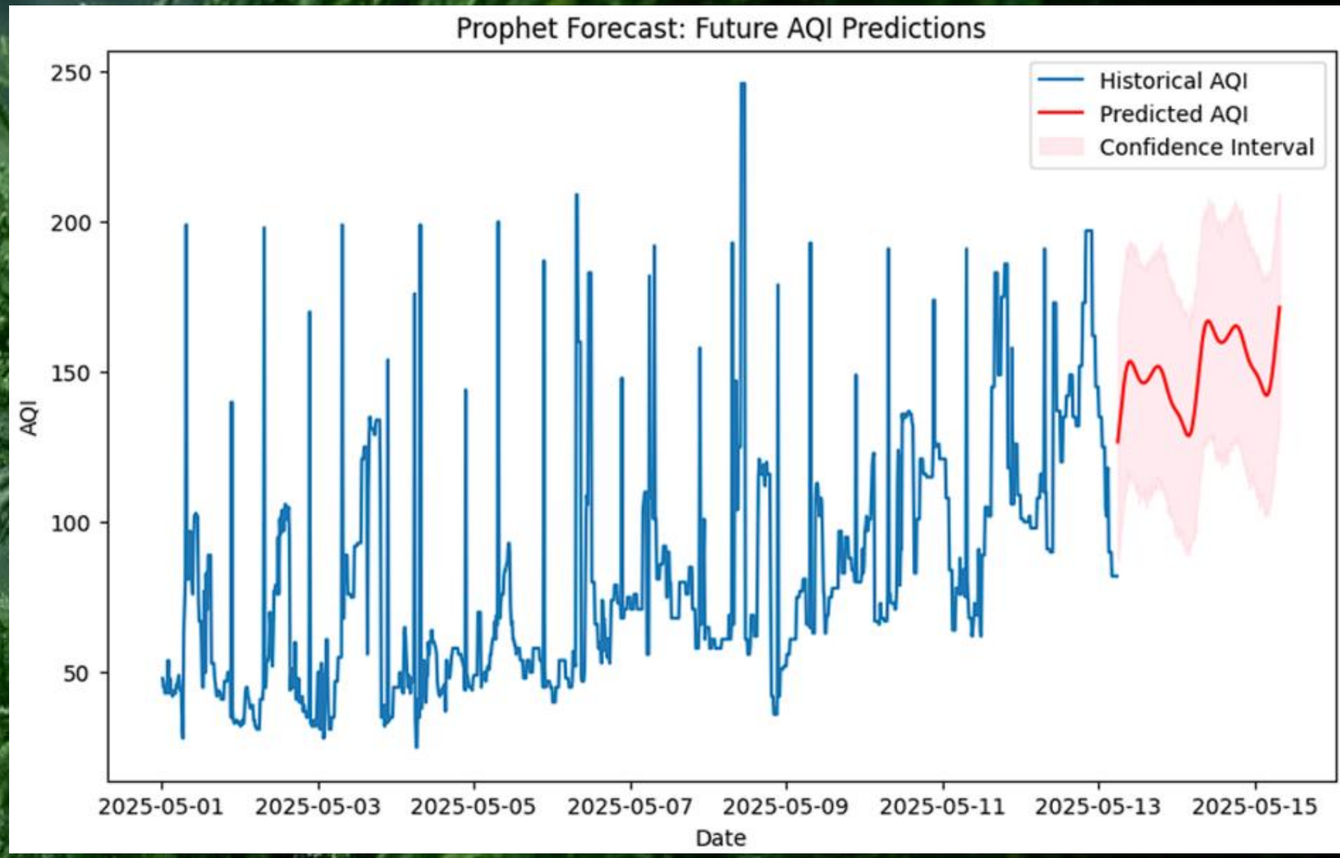

Usage of the Random Forest AQI Prediction Plot

- This plot is used to visually assess how accurately the Random Forest model predicts AQI values.
- predicted values align closely with actual ones,
- This is crucial before deploying the model in a real-world air monitoring system.
- Air Quality Forecasting
- Public Health Alerts
- Urban and Environmental Planning

# Prophet Forecast: Future AQI Predictions

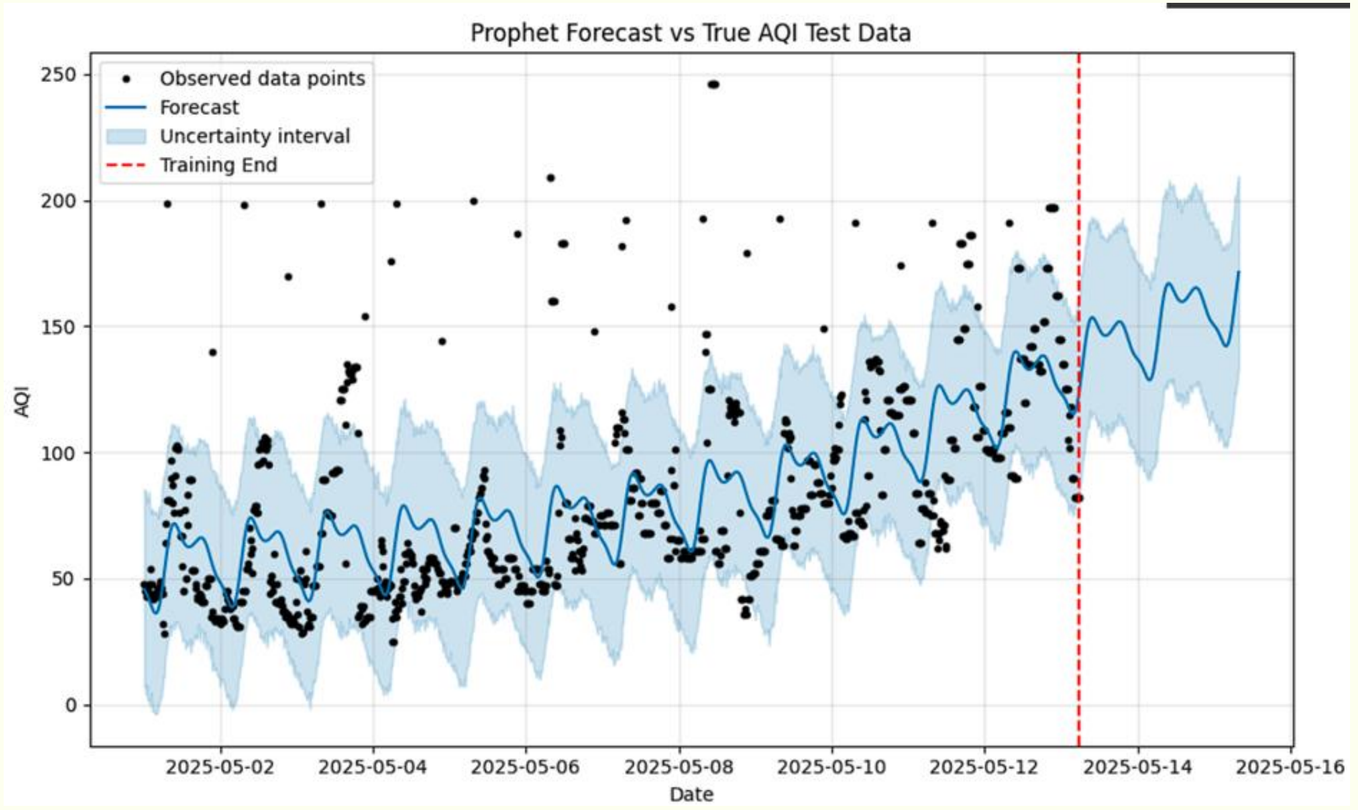
Time series plot visualizes the forecasted Air Quality Index (AQI) values using

- Blue Line: Historical AQI data (actual measurements up to May 12, 2025).
- Red Line: Forecasted AQI values (from May 12 to May 15, 2025).
- Shaded Pink Area: Confidence interval (range within which the true AQI is likely to fall, with 95% probability).

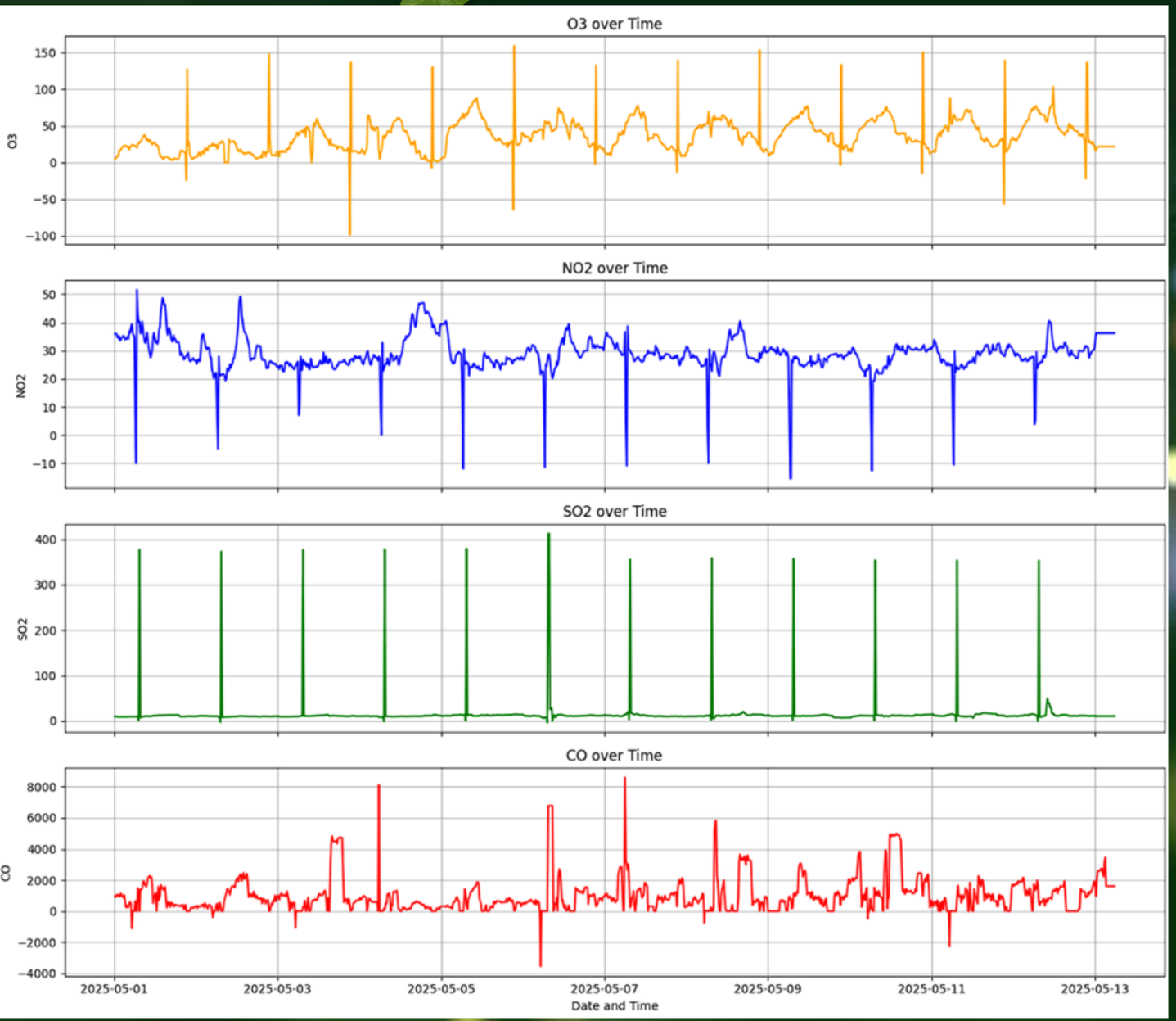



Why It's Important

- Accurate short-term forecasting of air quality
- Helps in early health alerts for poor air days
- Useful for policy planning and pollution control
- Provides uncertainty range to prepare for worst-case scenarios


Prophet Forecast vs True AQI Test Data
Observed data points
Forecast
Uncertainty interval
Training End
AQI
Date
250
200
150
100
50
0
2025-05-02
2025-05-04
2025-05-06
2025-05-08
2025-05-10
2025-05-12
2025-05-14
2025-05-16

# POLLUTANTS OVER TIME



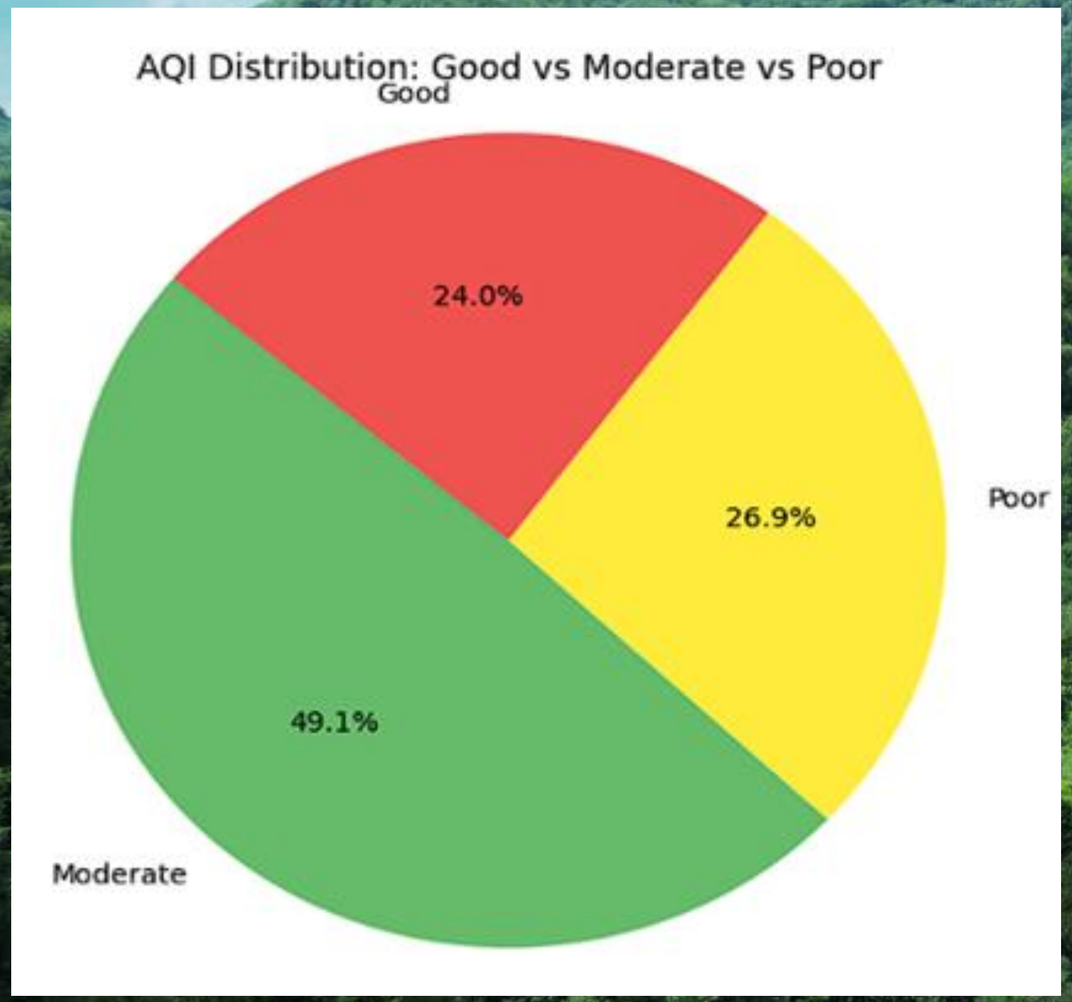

- time series plots for four major air pollutants:
- $O_3$ (Ozone) – 🟧 Orange
- $NO_2$ (Nitrogen Dioxide) – 🔵 Blue
- $SO_2$ (Sulfur Dioxide) – 🟢 Green
- CO (Carbon Monoxide) – 🔴 Red
- Ozone ($O_3$) is like the sun's shadow—it follows a daily rising and falling smoothly, mostly reacting to sunlight during the day.

- $NO_2$ (Nitrogen Dioxide) behaves like traffic—it stays steady but dips suddenly, when traffic slows

- $SO_2$ (Sulfur Dioxide) is quiet most of the time, but occasionally bursts loudly—likely from industrial emissions or power plants firing up.

- CO (Carbon Monoxide) is unpredictable spiking up and down like a wild heartbeat, possibly from vehicles, burning trash, or cooking fires.

# PAI CHART AQI DISTRIBUTION GOOD VS MODERATE VS POOR




| AQI Category | Percentage |
| --- | --- |
| Good | |

# EARTH IS OUR COMMON HOME

**INSIGHT:**

- CONTROLLING PM2.5 AND PM10 CAN HAVE THE MOST SIGNIFICANT IMPACT ON IMPROVING AQI.
- RANDOM FOREST IS THE BEST-PERFORMING MODEL FOR AQI PREDICTION IN THIS SCENARIO
- LINEAR REGRESSION IS SUITABLE FOR BASIC TREND ANALYSIS, BUT NOT FOR PRECISE AQI PREDICTION, ESPECIALLY DURING POLLUTION SPIKES
- PROPHET IS VALUABLE FOR EARLY WARNINGS AND PROACTIVE ENVIRONMENTAL PLANNING.
- TIME-BASED VISUALIZATIONS HELP IN IDENTIFYING POLLUTION SOURCES AND TIMING, AIDING REGULATORY ACTION.

# CONCLUSIO
# N

THE PROJECT SUCCESSFULLY DEMONSTRATES A COMPREHENSIVE APPROACH TO AIR QUALITY ANALYSIS AND FORECASTING. BY APPLYING LINEAR REGRESSION AND RANDOM FOREST, IT PREDICTS AQI WITH REASONABLE ACCURACY. THE INTEGRATION OF PROPHET ENABLES SHORT-TERM AQI FORECASTING, WHILE VISUALIZATIONS LIKE PAIRPLOTS, TIME SERIES GRAPHS, AND PIE CHARTS HELP INTERPRET POLLUTANT TRENDS AND AQI CATEGORIES EFFECTIVELY. OVERALL, THIS WORKFLOW PROVIDES A SOLID FOUNDATION FOR DATA-DRIVEN ENVIRONMENTAL MONITORING AND DECISION-MAKING.


Thanks YOU