Air Quality Index (AQI) Predictor of Karachi City

1.Introduction

One of the burning environmental issues facing modern urban communities relates to air pollution. The Air Quality Index (AQI) is a unified scale representing the level of several air pollutants present in the air, such as particulate matter (PM 2.5 / PM 10), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3) ,carbon dioxide(CO2), aerosol optical depth, dust, uv index and methane. The system will assign the numeric value to every concentration of a pollutant and will target the categories of readings that can be read and understood by the general population that include, but are not limited to, the warnings of the health hazard and possible air quality conditions, such as, Good, Moderate, Unhealthy, and Hazardous.

Being the biggest metropolitan geographical location in Pakistan, Karachi has especially been vulnerable to the mass air pollution. Rapid urbanization, high vehicular emissions, active industrial work as well as gaps in stringent regulations on the environment are some of the key contributors of this condition. It leads to the worsening of the air quality, which is highly noticeable in worn-out traffic routes and in industrial zones, leaving a population of nearly two hundred million souls at risk.

An AQI Predictor is a computational framework that, by merging modern weather and environmental predictors, provides real-time monitoring the ambient-air quality together with a potential airliner pollutant rank. These possibilities allow policymakers, environmental regulatory agencies, and the general population to implement pre-emptive measures--not only momentarily delaying outdoor recreational activities when air quality is poor but also strengthening emission-limitation systems.

2.Data Collection

The analysis performed on the present study used the data about atmospheric pollutants which had been acquired through **Open-Meteo Air Quality API**. This API offers high quality historical and real time data that does not require authentication keys. The data set focuses on Karachi city with the coordinates being **at 24.8608 and 67.0104**. Using the API endpoint ***https://air-quality-api.open-meteo.com/v1/air-quality***, **hourly** records for the **past 61 days** were collected, along with a one-day forecast.

The request included a comprehensive set of air quality and environmental parameters such as particulate matter concentrations (**PM_{10}** and **$PM_{2.5}$**), **carbon monoxide (CO)**, **carbon dioxide (CO_2)**, **nitrogen dioxide (NO_2)**, **sulphur dioxide (SO_2)**, and **ozone (O_3)**, as well as **aerosol optical depth**, **dust concentration**, **ammonia (NH_3)**, and **methane (CH_4)**. In addition to pollutant measurements, the dataset also included pollen counts for alder, birch, grass, mugwort, olive, and ragweed, along with **UV index** values under both current and clear-sky conditions.

During the creation of a predictive model of the air quality of Karachi, the **United States Air Quality Index (AQI)** standard was used as the primary measurement, thus, allowing compatibility with established health threats characteristics around the globe. Considering such a large assembly of variables enables the model to take into account not just direct concentrations of pollutants but also secondary environmental factors thus increasing the accuracy that the air quality trends of Karachi can be forecasted.

3.Data Analysis

The hourly air quality data that were accessed through **Open-Meteo Air Quality API** was also downloaded locally in **CSV** format to carry forward the process of processing and statistical analysis. This kind of formatting makes it easier to inspect, detect, and reproduce data as well as integrate with commonly used data analysis libraries like Pandas. A basic analysis showed that some of the features had substantial percentages of missing numbers in the data as **NaN**. These features included **ammonia**, **alder_pollen**, **birch_pollen**, **grass_pollen**, **mugwort_pollen**, **olive_pollen**, and **ragweed_pollen.**

The data used in this research had some variables that are underrepresented and thus cannot be analyzed meaningfully. They could have drawn **bias and noise i**nto the prediction model. These variables were therefore **dropped off** in the dataset. Retention files of particulate material, gas concentrations, UV index values, dust values and other outcomes of weather developments indicated a more reliable data-matrix to make an air-quality index prediction model. This data-cleaning process served to make the rest of variables complete and consistent and, therefore, strengthened the strength and interpretability of further works such as analysis and modelling.

3.1. AQI Calculation

In the calculations carried out in this case, an additional computational approach was needed since the initial data set failed to provide an air-quality index (AQI) of every hour sampled. In particular, the AQI of the US Environmental Protection Agency that is largely based on particulate matter (PM) concentrations was used. In turn, specific Python functions were developed to convert the measured PM2.5 and PM10 concentration values up to the AQI scores by means of using the officially approved breakpoint tables.

Air Quality	AQI
Good	0-50
Moderate	51-100
Unhealthy for sensitive groups	101-150
Unhealthy	151-200
Very unhealthy	201-300
Hazardous	>300

The total AQI at an hour was computed as the greatest of the PM 2.5 AQI and the PM 10 AQI and therefore reflected the most severe pollutant at that point in time. These values were subsequently rounded off to obtain to the nearest integer to enhance ease of understanding and to save in the new column, namely **us_aqi**. The process maintained the raw concentration of pollutant value in the dataset but also the standardized, intuitive measure giving the severity of the air-quality allowing its monitoring and prediction values to be accurate in Karachi.

Mathematical Process

Let:

- $C_{PM2.5}$ = $PM_{2.5}$ concentration in $\frac{\mu g}{m3}$
- C_{PM10}= PM_{10} concentration in $ug/m3$

- $AQI_{PM2.5}$= AQI value based on $PM_{2.5}$
- AQI_{PM10}= AQI value based on PM_{10}
- I_{high} , I_{low}= AQI breakpoint values
- C_{high}, C_{low}= concentration breakpoint values
- $AQI_{final}\}$ = overall AQI

1. AQI for a single pollutant ($PM_{2.5}$ or PM_{10})

The AQI for a given pollutant is calculated using the **linear interpolation formula**:

$$AQI_p=\frac{I_{high}-I_{low}}{C_{high}-C_{low}}\times(C_p - C_{low})+ I_{low}$$

Where:

- p represents the pollutant ($PM_{2.5}$ or PM_{10})
- Cp is the measured concentration of pollutant p
- $[C_{high}, C_{low}]$ and $[I_{high}, I_{low}]$ are taken from the **US EPA AQI breakpoint table**

2. Determining Overall AQI

The overall AQI is determined as the **maximum** of the individual AQI values:

$$AQI_{final}=\max(AQI_{PM2.5}, AQI_{PM10})$$

3. Final Adjustment

After calculation, the AQI is rounded to the nearest integer:

$$AQI_{final}=\mathrm{round}(AQI_{final})$$

4.Data Visualization

The phenomenon of the Karachi air quality was studied over the period of 61 days based on the implementation of Python-based visualization tools, namely **Matplotlib and Seaborn**. The focus was made on 3 main indices PM 2.5, PM 10 and respective values of the US Air Quality Index (AQI) .

To visualize how the calculated US Air Quality Index (AQI) has changed over time based on the measure taken on a period of 61 days in Karachi the **time series plot was developed** in *Figure II*. It is evident that the chart shows daily variations of air pollutants with strong spikes marking the periods of increased pollution. There exists an evident daily cyclicality, and it is put on top of less regular fluctuations. The time distribution exposes **recurrent spikes**, which could be related to in-rush conditions in traffic, manufacturing release and weather transformation. The graph, therefore, provides a qualitative picture of the time change in **air quality** and allows identifying the periods in need of some attention and possible pollution sources.

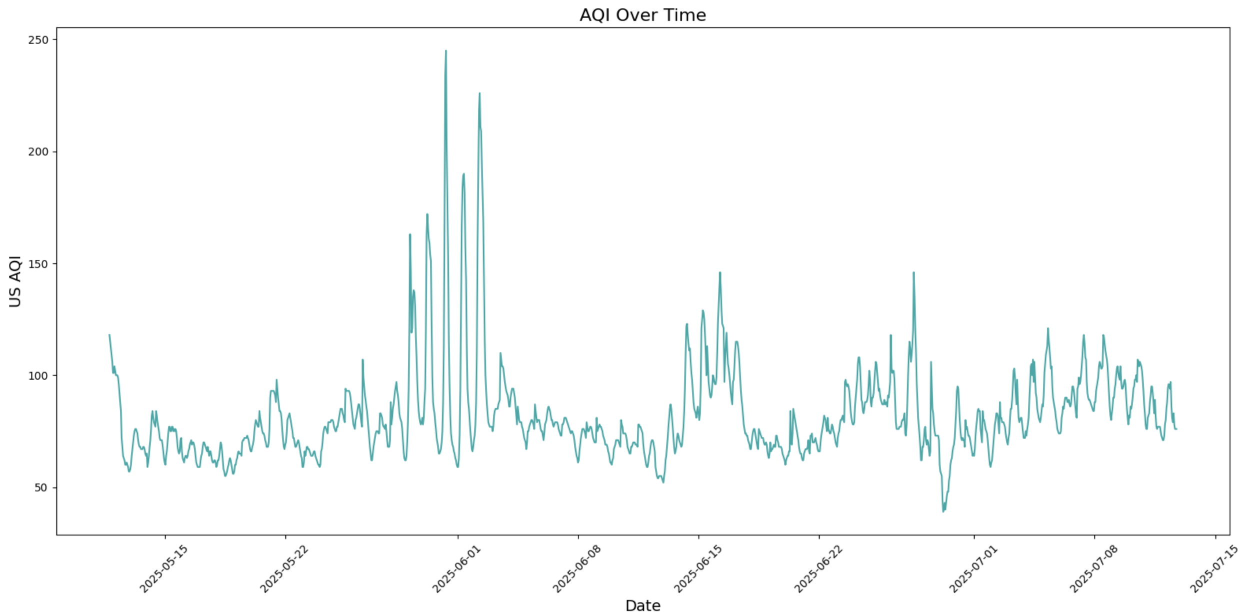


Fig II

A distribution plot of the US AQI values was generated in *Figure III* to understand the frequency and spread of different air quality levels in Karachi over the collected period. The histogram, enhanced with a KDE curve, reveals that a significant portion of the recorded hours fell into the 50 to 100, with fewer instances of "Good" air quality. The presence of multiple peaks suggests varying pollution intensity across different days or times, likely influenced by environmental and anthropogenic factors.

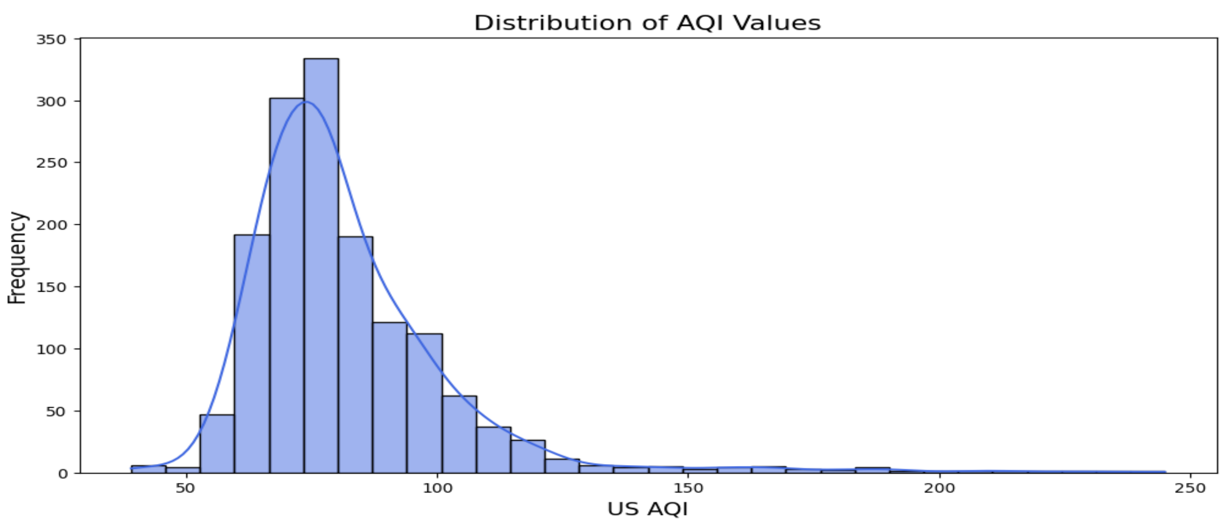


Fig III

In order to explain the overall trend and outliers, a box plot in *Figure IV* of the United States Air Quality Index (AQI) of Karachi was drawn. The graph marks the **median** of the level of air quality, the **interquartile range (IQR)** and a set of extreme values that are related to pollution increases. The long upper whisker and the great number of outlier points indicate most of the hours consisted of **moderate AQI situations**, but occasional spikes reached Unhealthy or worse levels. In this way, the visualization not only represents the variability and skew of quality inherent in Karachi based on the air quality but also shows the periodicity of the pollution incidents though their severity can be extreme.

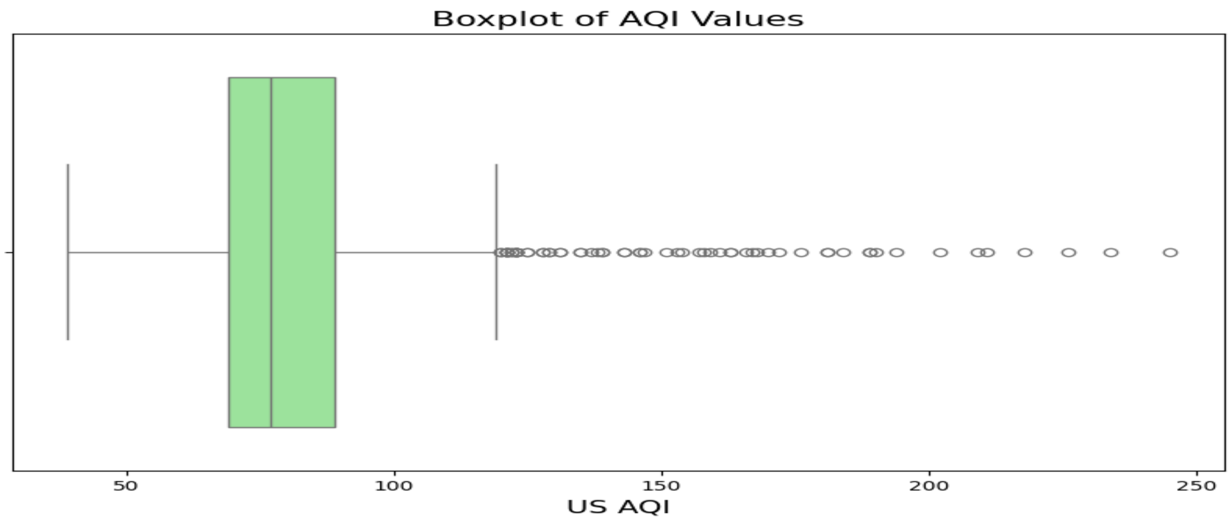


Fig IV

Hourly observation air quality indices (AQI) in the United States were reclassified into the six US Environmental Protection Agency (EPA) standard categories: Good, Moderate, Unhealthy for Sensitive Groups, Unhealthy, Very Unhealthy and Hazardous. This reclassification makes a visual judgment of these time trends in air quality interpretable. A count plot of the categories shows that the Most frequently occurring category is Moderate air quality whereas the next most common refers to Unhealthy for Sensitive Groups and with lesser but not insignificant counts are Unhealthy and Very Unhealthy conditions. In many cases where Hazardous classified episodes occur, they are uncommon; however, it can illustrate why critical pollution episodes that present severe health risks are possible. The resulting distribution provides a clear image on how much of the time Karachi population is facing different levels of pollution intensity. So, the counts of categories can be seen in Figure V.

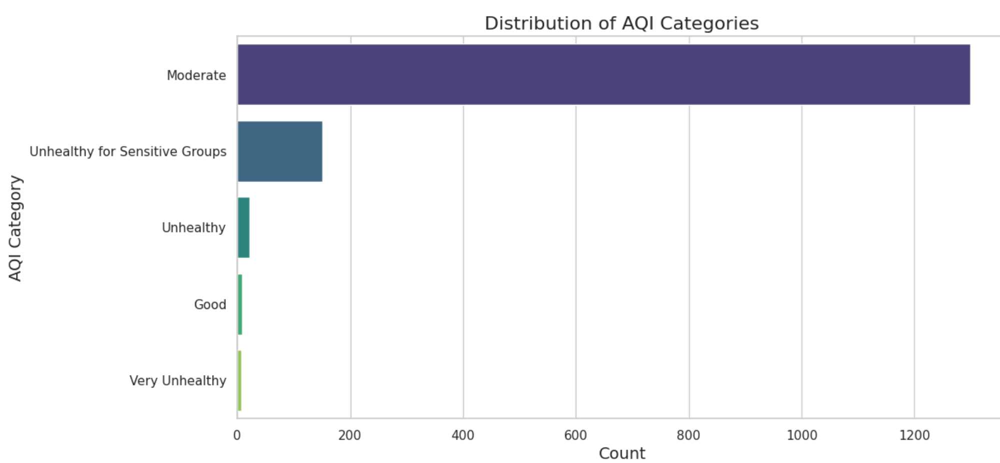


Fig v

The composition of Air Quality Index (AQI) categories incurred during the 61-days observation in Karachi is depicted in the given **pie chart**. It is a graphical interpretation of proportionate occurrence since the predominant

Moderate conditions took the larger proportion of about 8**7.3%** and then Unhealthy for Sensitive Groups before Unhealthy, very Unhealthy and lastly Hazardous episodes occupied smaller proportions and can be shown in figVI. The chart, therefore, effectively conveys the scale of each of the categories in comparison to that of another showing that though hazardous situations remained a rare occurrence, a significant part of the registered period was devoted to a situation that could impact vulnerable groups.

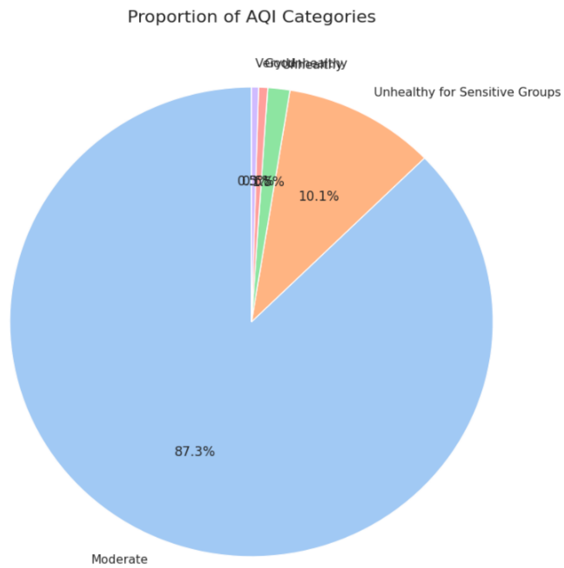


Fig VI

An illustration of the relationships between air-quality parameters and the computed US AQI were to be clarified by producing a **correlation heatmap** in Figure VII. The visualization focused on only pollutant concentration and continuous variables since all the non-numeric and categorical columns such as **date** and *aqi_category*, were removed. The resultant map revealed that there were robust positive associations between the PM 2.5, PM 10, dust,sulpur dioxide, ozone, carbon monoxide and aerosol optical depth thus substantiating their leading effect in controlling the air-quality matters. The given analysis determines the pollutants that have the greatest impact on the AQI of Karachi and, thus, informs efforts that rely on predictive modeling and prioritize environmental policies.

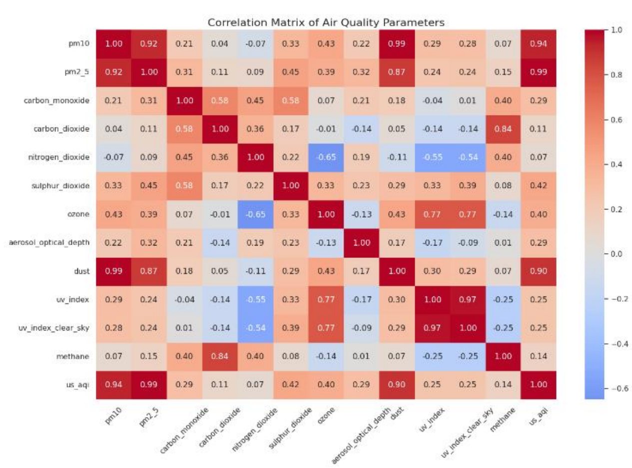


Fig VII

A scatter plot was used to visualize the relationship between **$PM_{2.5}$ concentrations** and the calculated **US AQI**, with data points color coded by AQI category. The plot in Figure VIII clearly shows a strong positive relationship, where higher $PM_{2.5}$ levels correspond to significantly higher AQI values, often crossing into *Unhealthy* or worse categories. The clustering of points in the lower $PM_{2.5}$ range reflects periods of better air quality, while the steep upward trend highlights the pollutant's critical role in driving AQI spikes in Karachi.

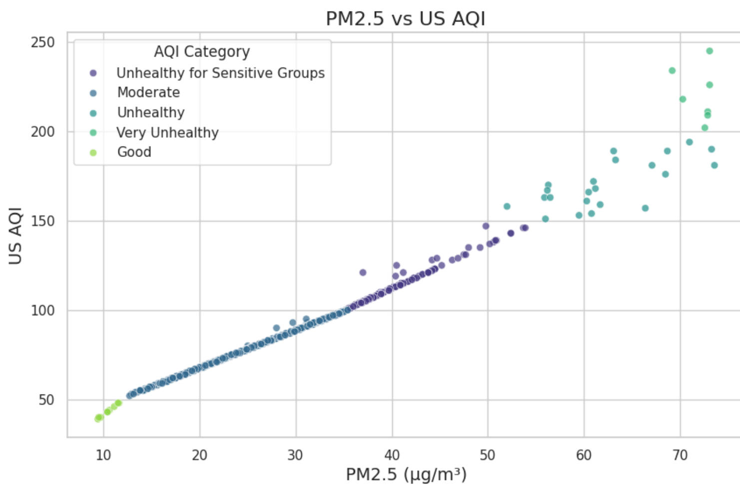


Fig VIII

A scatter plot was created to examine the relationship between **PM_{10} concentrations** and **US AQI**, with points color coded according to AQI category in Figure IX. The visualization reveals a clear upward trend, indicating that higher PM_{10} levels generally lead to elevated AQI values. While moderate PM_{10} concentrations often fall within the *Good* or *Moderate* ranges, spikes in PM_{10} are strongly associated with *Unhealthy* and *Very Unhealthy* air quality levels.

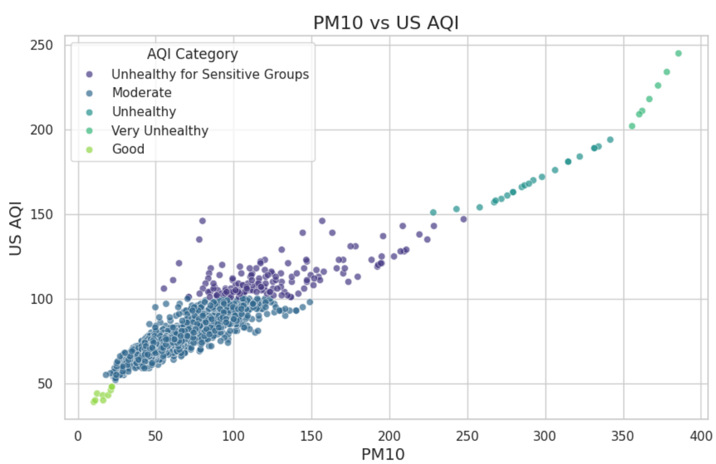


Fig IX

The scatter plot between dust concentration and US AQI highlights how airborne dust particles influence air quality in Karachi. The visualization in Figure X shows a positive correlation, where higher dust levels tend to correspond with increased AQI values, often pushing the air quality into *Unhealthy* or even *Very Unhealthy* categories. However, the spread of points also indicates periods where elevated dust does not significantly raise AQI, suggesting that dust is only one of several pollutants driving poor air quality.

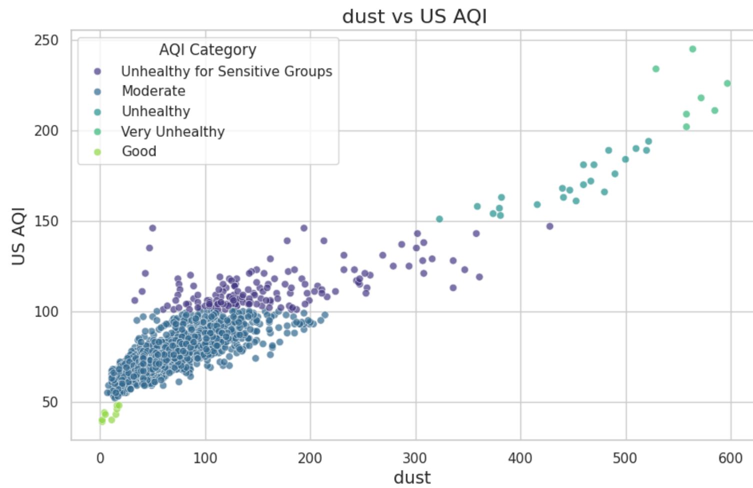


Fig X

The scatter plot comparing sulphur dioxide (SO_2) concentrations with US AQI reveals how this gaseous pollutant impacts overall air quality. While SO_2 alone may not always be the dominant pollutant, the plot shows that higher concentrations often coincide with elevated AQI values, particularly in the *Moderate* to *Unhealthy* range. The clustering of points at lower SO_2 levels suggests that Karachi's AQI is frequently influenced by a combination of pollutants rather than SO_2 in isolation.

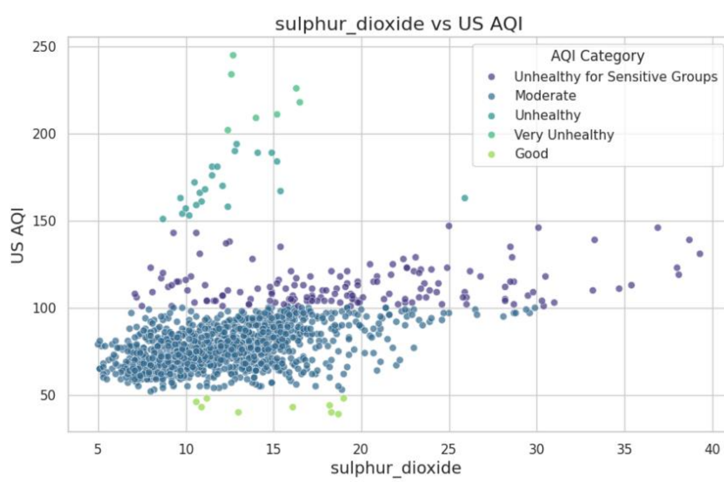


Fig XI

The diurnal cycle of air quality obtained on the **hour-by-hour time scale** shows the daily pattern of air quality varying with time of the day, called a line plot, shown in Figure XII, as average AQI by Hour of the day. The data reflects that the level of the pollutants alternates during the day according to human activity patterns, weather conditions and macro-atmospheric processes. It is worthy to note that, the AQI measurements tend to rise during early morning and evenings- times when there are acute traffic jams and lack of dispersion, as opposed to when the AQI measurements are relatively lower at midpoints due to the photo chemical reactions and dilution of these pollutants' activities under the solar effect. This detailed study can be useful to determine when to issue a notification to the public and the development of pollution control plans by place.

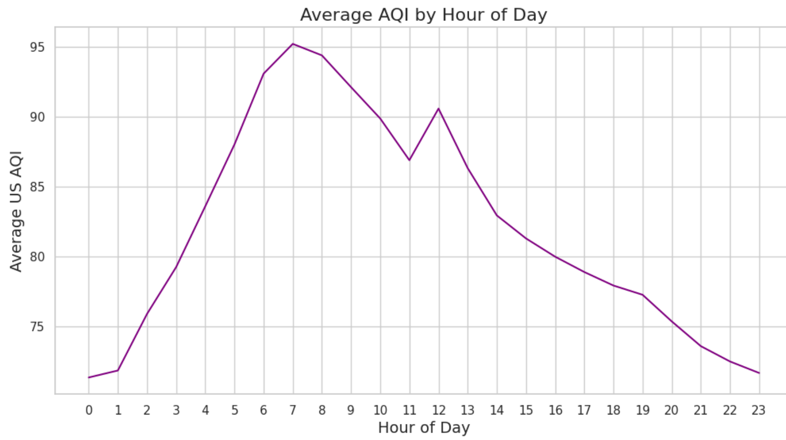


Fig XII

The **AQI Distribution by Month** boxplot in Figure XIII presents the seasonal variability of air quality in Karachi. Each box represents the spread, median, and outliers of AQI values for a specific month, allowing clear observation of periods with consistently higher or lower pollution levels. Seasonal peaks in AQI may align with factors such as increased industrial activity, reduced wind dispersion, or climatic changes like winter inversions that trap pollutants closer to the ground. This analysis is essential for understanding long-term pollution trends and planning seasonal mitigation strategies.

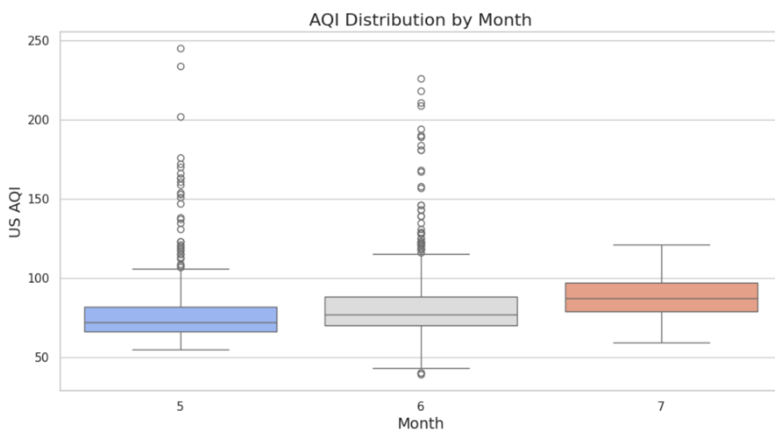


Fig XIII

5. Selection of Highly Correlated Features for AQI Prediction

To enhance the predictive performance of the AQI model, a correlation analysis was conducted between all available features and the target variable, *US AQI*. Only those features exhibiting a Pearson correlation coefficient greater than 0.25 with respect to *US AQI* were retained for further modeling. Based on this criterion, the selected predictors were PM_{10}, carbon monoxide (CO), sulphur dioxide (SO_2), dust, aerosol optical depth (AOD), and ozone (O_3). These features were extracted and stored in a new CSV file to ensure a clean and concise dataset for subsequent modeling steps.

For model training, the dataset was divided into **input features (X)** and **target variable (y)**. Here, *X* consisted of the selected predictors: ['pm10', 'carbon_monoxide', 'sulphur_dioxide', 'dust', 'aerosol_optical_depth', 'ozone'], while *y* corresponded to the *US AQI* values. This separation ensured a clear distinction between explanatory variables and the response variable, enabling the application of supervised machine learning algorithms for AQI prediction.

In the feature selection, a strong and highly obvious association with the air quality variable (PM2.5) was avoided, even though it had a correlation coefficient of an uncharacteristically high score of 0.99 with respect to the target variable (US AQI). Its elimination relieves the model of bias, reduces overfitting and guarantees that the consequent predictions will be based on a larger set of features, thus both mitigates generalizability and improves the interpretability of generated predictions.

6.Model Development and Evaluation for AQI Prediction

6.1 Train Test Split

To prepare the dataset to model, the features were first separated into training and testing as I used the 80:20 ratio of split of features using the *train_test_split* function offered by scikit-learn. This data were further divided into a training set (80 percent of the available data) and the testing set (20 percent of the available data) with the training set later used to train machine-learning models and testing set only used to assess the performance of these models on unseen samples. A constant *random_state (42)* controlled reproduction of the data manipulation phases down the line, to ensure the same results of the experiment.

6.2 Feature Standardization

The chosen characteristics are heterogeneous when it comes to the units and scales of measurements. To normalize this imbalance, the **StandardScaler** feature was used to standardize the feature. The transformation standardizes each feature by making it have a mean of 0 and a standard deviation of 1, hence ensuring that each of the features has equal contribution during training of the model and that when constructing the model, there is no case where the magnitudes of individual characteristics inadvertently introduce model bias. The training set only was fitted on the scale, to avoid information leakage; the same parameters that were obtained during this fit were subsequently used on the test sets.

6.3 Model Implementation

After standardization I applied different Machine Learning models which include the Linear Regression, Decision Tree, K-nearest neighbor (KNN), Random Forest, XGBoost and Artificial Neural Network (ANN). Each model evaluated with metrices like Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R-Square (R^2) error. The mathematical representation of each metric can be seen as following:

I. Mean Square Error (MSE)

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE = mean squared error

n = number of data points

Y_i = observed values

$\hat{Y}_i$ = predicted values

II. Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

$$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n \text{ are predicted values}$$
$$y_1, y_2, \ldots, y_n \text{ are observed values}$$
$$n \text{ is the number of observations}$$

III. Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y_i}|$$

where,

n: *number of observation*

y_i: *the actual value of the* i^{th} *observation*

$\widehat{y_i}$: *the predicted value of the* i^{th} *observation*

IV. R- Square (R^2)

Coefficient of Determination (R Square)

$$R^2 = \frac{SSR}{SST}$$

$$SSR = \sum_i (\hat{y}_i - \bar{y})^2$$

$$SST = \sum_i (y_i - \bar{y})^2$$

Where,

- SSR is Sum of Squared Regression also known as variation explained by the model
- SST is Total variation in the data also known as sum of squared total
- y_i is the y value for observation i
- y_bar is the mean of y value
- y_bar_hat is predicted value of y for observation i

www.ashutoshtripathi.com

The comparative performance *Table I* summarizes the results of six machine learning models applied to predict US AQI based on selected air quality parameters. Among these**, XGBoost** achieved the best performance, with the lowest MSE (15.12), RMSE (3.89), and MAE (2.62), alongside the highest R^2 value of 0.9477, indicating excellent predictive accuracy. **Random Forest** closely followed, showing similarly strong results. **KNN** Regressor and **ANN** also performed well, outperforming **Linear Regression and Decision Tree** in most metrics. **Decision Tree** yielded the lowest accuracy, with the highest error values, highlighting its relatively limited generalization capability for this dataset. Overall, ensemble-based models (XGBoost and Random Forest) proved most effective for AQI prediction in this study.

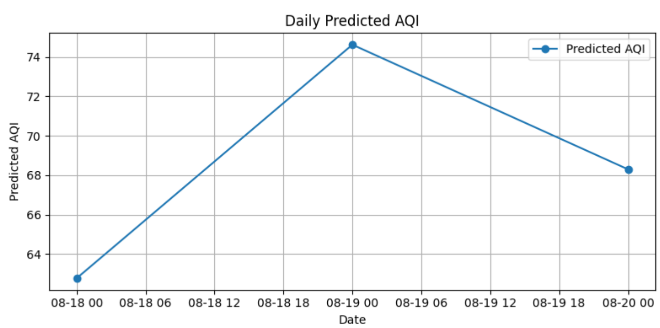
Model	MSE	RMSE	MAE	R^2
Linear Regression	26.54	5.15	3.61	0.9082
Decision Tree	39.54	6.29	4.66	0.8633
KNN Regressor	22.32	4.72	3.55	0.9228
Random Forest	15.83	3.98	2.76	0.9452
XGBoost	15.12	3.89	2.62	0.9477
ANN	18.61	4.31	3.07	0.9356

Table I

From above table it can be determined that XGBOOST performs well as compared to other models. Storing the xgboost model allows for future reuse without retraining, enabling quick deployment or integration into prediction pipelines. JSON format is particularly useful for portability and easy loading in other environments or applications.

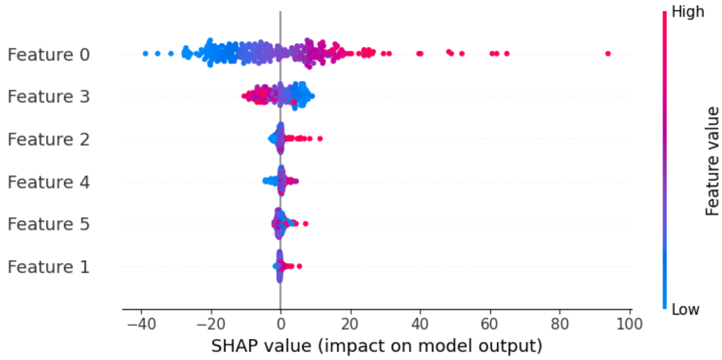
6.4 Deployment of an XGBoost-Powered Web Application for 3-Day Air Quality Forecasting

This code implements an Air Quality Prediction Dashboard using Python, XGBoost, and Gradio. It fetches real-time air quality data (such as PM10, CO, SO_2, dust, AOD, and ozone) from the Open-Meteo Air Quality API for a given latitude and longitude, scales the features with a pre-trained scaler, and then predicts the U.S. Air Quality Index (AQI) using a trained XGBoost regression model. The predicted AQI values are aggregated into daily averages for three days, displayed in a table, and visualized as a line plot. A Gradio-based interface allows users to input coordinates, run predictions, and view results interactively.

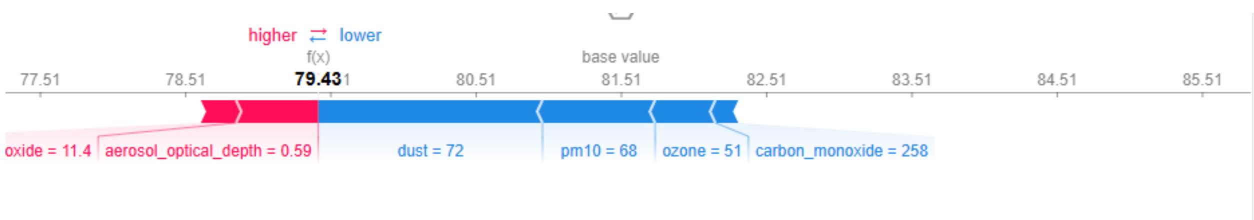


6.5 SHapley Additive explanations SHAP and Local interpretable model-agnostic explanations LIME

Applied **SHAP (SHapley Additive exPlanations)** to interpret the predictions of the trained XGBoost model. An explainer is initialized with the model and the scaled training data, then SHAP values are computed for the scaled test set to measure each feature's contribution to the model's output for every prediction. Finally, a **summary plot** is generated, which visualizes the average impact and direction (positive or negative) of each feature on the predicted AQI, helping to identify the most influential factors driving the model's decisions.

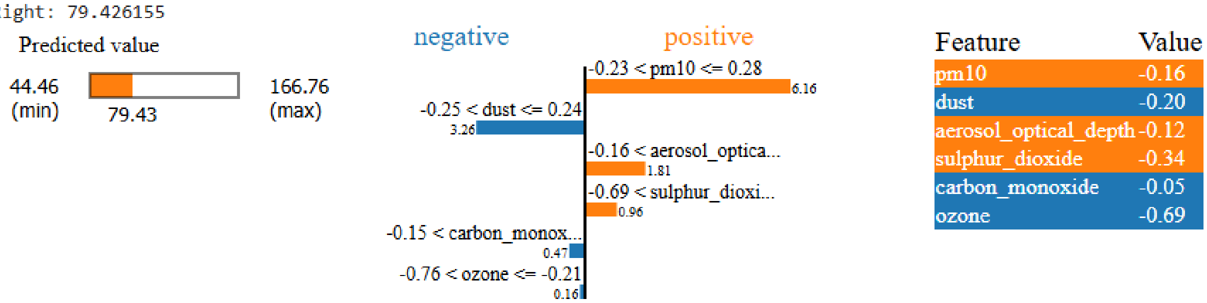


This snippet creates an interactive SHAP force plot to explain a single AQI prediction. It first initializes SHAP's JavaScript support for visualizations, selects a specific test instance (here, index 0), and then uses the computed SHAP values to display how each feature pushed the prediction higher or lower compared to the model's baseline (expected value). The force plot visually shows feature contributions in red (increasing the prediction) and blue (decreasing the prediction), offering a clear, instance-level explanation of the model's decision.

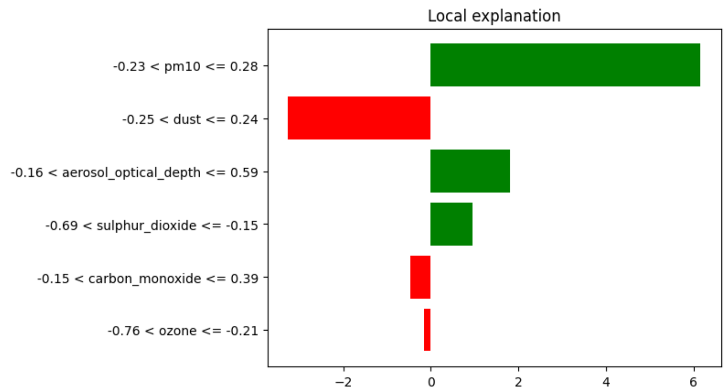


This graph **LIME (Local Interpretable Model-agnostic Explanations)** to interpret the AQI prediction for a single test instance. It starts by creating a Lime Tabular Explainer with scaled training data, providing feature names and specifying **regression mode** since AQI is a continuous value. It then selects a test sample (i = 0) and generates an explanation by showing how changes in each feature locally influence the prediction made by the **XGBoost model**. The output is a clear, human-readable breakdown of the most influential features and their weights for that specific prediction.

```
Intercept 84.87063377692391
Prediction_local [89.90149554]
Right: 79.426155
```



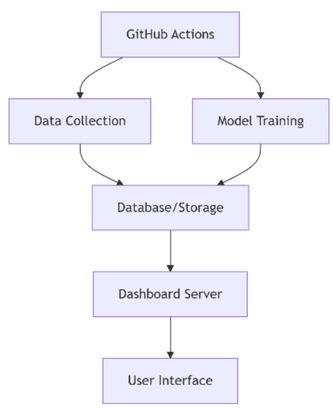
In this plot the **x-axis** shows the magnitude of each feature's contribution (positive or negative) to the AQI prediction. The **y-axis** lists the features, ranked by their local importance for the chosen instance. **Green bars** indicate factors pushing the predicted AQI higher, while **red bars** indicate factors pulling it lower.



6.6 Forecasting

7.CI and CD Pipeline Implementation

In this project I implement a fully automated CI/CD pipeline for an air quality monitoring system using GitHub Actions. The pipeline consists of four interconnected workflows that handle data collection, model training, dashboard deployment, and fallback scenarios as shown in following figure. Each workflow is triggered on a scheduled basis (with manual override capability) and produces versioned artifacts that feed into subsequent stages.



Data Collection Workflow

The data collection workflow serves as the pipeline's foundation, running hourly at the top of each hour. It begins by setting up a Python environment and installing necessary dependencies including openmeteo_requests for API access and pandas for data processing. The workflow then executes collect_features.py, which fetches real-time air quality metrics from designated APIs, performs data cleaning and transformation, and outputs a structured CSV file.

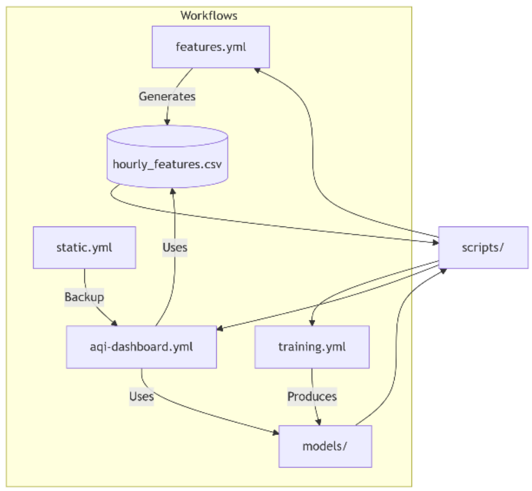
Model Training Workflow

Triggered 30 minutes past each hour to allow fresh data availability, the model training workflow implements the machine learning CI process. After retrieving the latest data artifact, it runs train_model.py which implements the full model lifecycle - including feature engineering, train-test splitting, XGBoost model training with predefined hyperparameters, and performance evaluation against test data. The workflow enforces quality gates by validating key metrics like RMSE and R^2 scores against predefined thresholds.

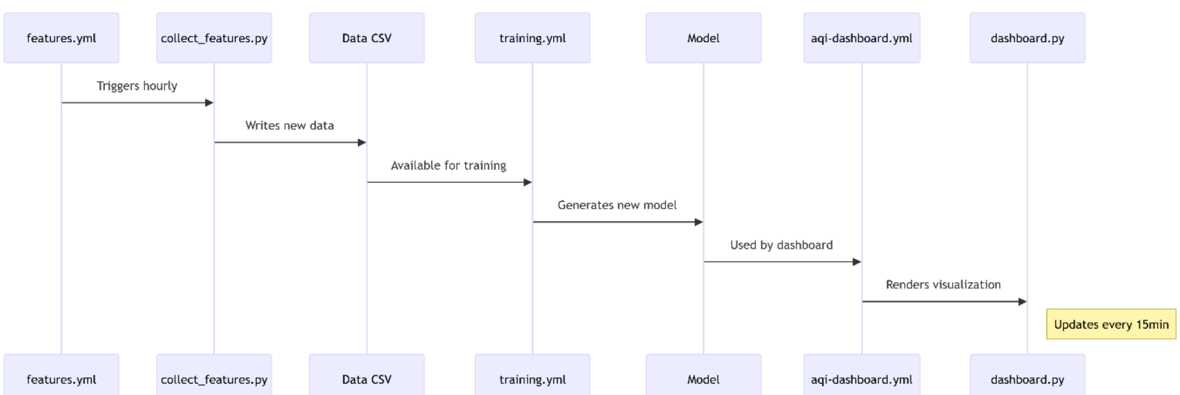
Dashboard Deployment

Serving as the CD component, the dashboard workflow runs every 15 minutes to provide near-real-time visibility. It combines the latest data and model artifacts with visualization logic from dashboard.py, which uses Streamlit and Plotly to create interactive charts showing current conditions and forecasts. The workflow employs several optimization techniques: it caches dependencies between runs, implements incremental updates when data hasn't changed, and generates both interactive elements and static fallback content.

The whole process of above workflows is present in following figure:



Data Pipeline Flow



References

1. World Health Organization. (2021). *Ambient (outdoor) air pollution.* WHO. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

2. U.S. Environmental Protection Agency. (2023). *Air Quality Index (AQI) Basics.* https://www.airnow.gov/aqi/aqi-basics/

3. Chen, L., Bai, Z., Kong, S., Han, B., You, Y., & Ding, X. (2018). Prediction of PM2.5 concentrations using a hybrid model of XGBoost and LSTM. *Atmospheric Environment*, 200, 264–273. https://doi.org/10.1016/j.atmosenv.2018.12.045

4. Zhang, J., & Ding, W. (2017). Prediction of Air Pollutants Concentration Based on an Extreme Learning Machine: The Case of Hong Kong. *Environmental Modeling & Assessment*, 22(6), 583–601. https://doi.org/10.1007/s10666-017-9550-8

5. Kumar, A., & Singh, P. (2020). Air quality prediction using deep learning techniques. *Journal of Ambient Intelligence and Humanized Computing*, 11(9), 4039–4048. https://doi.org/10.1007/s12652-019-01523-4

6. Bai, Y., Zeng, B., Li, C., Zhang, J., Zhang, M., & Gong, P. (2018). Air Pollution Forecasts: An Overview. *International Journal of Environmental Research and Public Health*, 15(8), 1571. https://doi.org/10.3390/ijerph15081571

7. GitHub Docs. (2024). *Configuring a publishing source for GitHub Pages.* https://docs.github.com/en/pages/configuring-a-publishing-source-for-your-github-pages-site

8. Plotly. (2024). *Plotly Python Open Source Graphing Library.* https://plotly.com/python/

9. Pandas Development Team. (2024). *pandas: Python Data Analysis Library.* https://pandas.pydata.org/

10. OpenAQ. (2024). *Open Air Quality Data Platform.* https://openaq.org/