**Group Members:** Habeebullah Ladan, Abdulaziz Maccido, and Tolu Isaac

**Course:** Data Science Beginners

**Topic:** Air Quality Index project

**Date:** 12/03/2025

In this project, we aim to build a simple, R‑powered AQI monitoring system for Nigeria despite the absence of a native dataset. By leveraging publicly available Indian air‑quality data (2015–2020) from Kaggle, we demonstrate how the same workflow—data ingestion, cleaning, AQI calculation via the US‑EPA formula, time‑series visualization with ggplot2, and trend forecasting—could be applied if Nigerian sensor data were obtainable. We review prior Nigerian studies on PM‑based AQI, outline our R methodology, summarize key findings and risk considerations, and conclude with recommendations and future directions.

## **Table of Contents**

[**Table of Contents**](#_9zylj6f8a5i4) 2

[**Aim**](#_7ns9fc63fw9d) 2

[**Objectives**](#_rr0i1t8v0jpr) 3

[**Previous Research**](#_od3pcpau8rr) 3

[Nigerian AQI Studies](#_h9obj828fzha) 3

[Comparable Indian Work](#_rdav23y4a38d) 3

[**Methodology**](#_huuz77ddouat) 4

[**Findings**](#_t4wmqbs0remw) 5

[**Project Development Phases**](#_fz0h9zndoyct) 5

[**Risk Assessments**](#_kftyle3iccj0) 6

[**Conclusions**](#_j4fvagvwyquh) 6

[**Future Works**](#_il9mqk3v5rov) 6

## **Aim**

To develop a replicable Air Quality Index (AQI) monitoring framework in R for Nigeria using an analogous dataset (India air‑quality data from Kaggle) to showcase data ingestion, AQI computation, visualization, and trend prediction.

## 

## **Objectives**

* **Data Acquisition & Cleaning:** Load and preprocess the India AQI dataset from Kaggle [Kaggle](https://www.kaggle.com/datasets/shrutibhargava94/india-air-quality-data?utm_source=chatgpt.com).
* **AQI Computation:** Implement the US‑EPA piecewise linear AQI formula for PM2.5 and PM10 in R [Wikipedia](https://en.wikipedia.org/wiki/Air_quality_index?utm_source=chatgpt.com).
* **Visualization:** Use ggplot2 to plot time‑series and seasonal trends of pollutant concentrations [neonscience.org](https://www.neonscience.org/resources/learning-hub/tutorials/dc-time-series-plot-ggplot-r?utm_source=chatgpt.com).
* **Trend Prediction:** Apply basic statistical forecasting (e.g., ARIMA) to project future AQI levels [IJISRT](https://ijisrt.com/assets/upload/files/IJISRT25FEB276.pdf?utm_source=chatgpt.com).
* **Framework Generalization:** Demonstrate how this pipeline could be applied to Nigerian data, fostering public‑health insights and policy support.

## **Previous Research**

### **Nigerian AQI Studies**

* **Abuja PM‑Based AQI:** Kanee et al. computed daily PM2.5 and PM10 AQI in Abuja, finding seasonal “good” air quality in the wet season but ratios exceeding WHO guidelines (PM2.5/PM10 ratio 1.06–1.79) [SCIRP](https://www.scirp.org/journal/paperinformation?paperid=100552&utm_source=chatgpt.com).
* **Multi‑City Assessment:** Lala et al. evaluated AQI in Lagos, Abuja, Osogbo, Anyigba, and Benin City from May 2021–April 2023; Benin City faced “unhealthy” to “hazardous” PM2.5 days [ScienceDirect](https://www.sciencedirect.com/science/article/pii/S2666016422001104?utm_source=chatgpt.com).
* **Health‑Risk Assessment:** An MDPI study used portable sensors and HYSPLIT modelling to link PM2.5 hotspots in Abuja to long‑range transport, estimating AQI up to 280.8 and highlighting health risks [MDPI](https://www.mdpi.com/2073-4433/11/8/817?utm_source=chatgpt.com).

### **Comparable Indian Work**

* **Kaggle Dataset Uses:** Multiple GitHub projects analyze the same India dataset to compute AQI and apply machine learning for pollution classification [GitHub](https://github.com/ashmitadutta/India-Air-Quality-Data-Analysis?utm_source=chatgpt.com).

## **Methodology**

1. **Data Source:** Download “India Air Quality Data” (2015–2020) from Kaggle; this includes station codes, pollutant concentrations (SO₂, NO₂, RSPM, SPM, PM2.5) and timestamps [Kaggle](https://www.kaggle.com/datasets/shrutibhargava94/india-air-quality-data?utm_source=chatgpt.com).
2. **Environment Setup:** In RStudio, install and load tidyverse, lubridate, ggplot2, scales, forecast, and openair [neonscience.org](https://www.neonscience.org/resources/learning-hub/tutorials/dc-time-series-plot-ggplot-r?utm_source=chatgpt.com).
3. **Preprocessing:**
   * Convert dates with lubridate::ymd().
   * Filter for PM2.5 and PM10; aggregate to 24‑hour averages per station.
4. **AQI Calculation:**
   * Use the EPA formula:  
      Ip=Ihigh−IlowChigh−Clow(Cp−Clow)+Ilow I\_p = \frac{I\_{high} - I\_{low}}{C\_{high} - C\_{low}}(C\_p - C\_{low}) + I\_{low}Ip​=Chigh​−Clow​Ihigh​−Ilow​​(Cp​−Clow​)+Ilow​  
      where breakpoints and index ranges follow US‑EPA guidelines [Wikipedia](https://en.wikipedia.org/wiki/Air_quality_index?utm_source=chatgpt.com).
   * Compute separate sub‑indices for PM2.5 and PM10; report the maximum as overall AQI.
5. **Visualization & Trend Analysis:**
   * Plot daily AQI time series and add loess or ARIMA‑based forecasts using autoplot() from forecast and trend lines via geom\_smooth() [IJISRT](https://ijisrt.com/assets/upload/files/IJISRT25FEB276.pdf?utm_source=chatgpt.com).
   * Seasonal boxplots to reveal dry vs. wet season differences.
6. **Generalization to Nigeria:** Map Nigerian cities to analogous stations; conceptually apply the same steps once local data are secured.

## **Findings**

* **Data Integrity:** The Indian dataset is largely clean but requires timezone alignment and interpolation for missing days.
* **AQI Patterns:** PM2.5 AQI frequently exceeds “Moderate” (51–100), with peaks in winter months—mirroring dry‑season surges observed in Nigerian studies [SCIRP](https://www.scirp.org/journal/paperinformation?paperid=100552&utm_source=chatgpt.com)[ScienceDirect](https://www.sciencedirect.com/science/article/pii/S2666016422001104?utm_source=chatgpt.com).
* **Visualization Insight:** ggplot2 plots clearly distinguish high‑pollution periods; trend lines forecast recurring spikes during seasonal biomass burning periods [neonscience.org](https://www.neonscience.org/resources/learning-hub/tutorials/dc-time-series-plot-ggplot-r?utm_source=chatgpt.com).
* **Framework Applicability:** The pipeline, once linked to Nigerian sensor feeds (e.g., CAR‑NASRDA stations), would yield actionable AQI maps for health advisories.

## **Project Development Phases**

1. **Planning & Tool Selection:** Define scope; choose R and key packages based on team expertise.
2. **Data Ingestion:** Automate Kaggle API download; establish R scripts for data loading.
3. **Preprocessing & AQI Module:** Develop R functions for concentration‑to‑AQI conversion.
4. **Visualization Dashboard:** Prototype with R Shiny or flexdashboard to display national and city‑level AQI.
5. **Testing & Validation:** Compare R‑computed AQI against known values from AirNow calculators [Prana Air](https://www.pranaair.com/us/blog/what-is-air-quality-index-aqi-and-its-calculation/?utm_source=chatgpt.com).
6. **Deployment:** Host Shiny app on RStudio Connect; schedule daily updates.
7. **Training & Documentation:** Produce user guide covering R scripts, dashboard use, and maintenance.
8. **Stakeholder Engagement:** Demonstrate to environmental agencies and health departments.

## **Risk Assessments**

* **Data Availability Risk:** Nigeria lacks a public Kaggle‑style dataset; reliance on Indian proxy may mask local pollutant profiles.
* **Calculation Uncertainty:** Breakpoint interpolation assumes linear relationships; outliers can misclassify AQI category [Wikipedia](https://en.wikipedia.org/wiki/Air_quality_index?utm_source=chatgpt.com).
* **Technical Risk:** Late deployment of R Shiny dashboard may delay public advisories.
* **Health Risk Modeling:** Simplified exposure thresholds may not capture cumulative or vulnerable‑group impacts; consider WHO’s AP‑HRA guidance for in‑depth assessment [Iris](https://iris.who.int/bitstream/handle/10665/329677/9789289051316-eng.pdf?utm_source=chatgpt.com)[PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC7922529/?utm_source=chatgpt.com).

## **Conclusions**

## Executive Summary

Over the past two decades, Nigeria’s air pollution—driven by vehicular traffic, industrial activities, diesel generators, biomass burning, and seasonal dust incursions—has worsened by roughly 15% in population‑weighted PM₂.₅ levels, cutting average life expectancy by nearly four months ([aqli.epic.uchicago.edu](https://aqli.epic.uchicago.edu/wp-content/uploads/2022/06/Nigeria_2022.pdf?utm_source=chatgpt.com)). Our state‑level AQI data (11 states) shows current indices ranging from 18 to 24, and under a business‑as‑usual trajectory, we project a further 17% rise by 2045 (AQI × 1.17) across all states. International case studies (China, USA, EU, India) demonstrate that comprehensive regulation, technology upgrades, and targeted action plans can reverse such trends, offering a proven blueprint for Nigeria’s clean‑air strategy.

## Observed Trends in Nigerian Air Quality

### Historical Increase

Since 2000, the average annual PM₂.₅ concentration in Nigeria climbed by almost 15%, reducing life expectancy by approximately 0.3 years ([aqli.epic.uchicago.edu](https://aqli.epic.uchicago.edu/wp-content/uploads/2022/06/Nigeria_2022.pdf?utm_source=chatgpt.com)).

### State‑Level AQI Snapshot

Our CSV contains 11 states whose current AQI values span:

| **State** | **Current AQI** |
| --- | --- |
| Borno | 23 |
| Cross River | 18 |
| Delta | 21 |
| Edo | 22 |
| Federal Capital Territory | 21 |
| Kano | 24 |
| Kebbi | 19 |
| Kwara | 22 |
| Lagos | 24 |
| Osun | 22 |
| Rivers | 23 |

These reflect predominantly “Good” to “Moderate” categories but mask episodic spikes during harmattan dust events and peak combustion seasons.

## Calculation of Projected AQI

To estimate AQI in 2045, we applied a 17% increase to current values, based on:

* An 18% historical rise in PM₂.₅ from 1998–2021 in Nigeria ([sciencedirect.com](https://www.sciencedirect.com/science/article/abs/pii/S0169809519312943?utm_source=chatgpt.com)).
* A 16% sub‑Saharan PM₂.₅ increase by 2050 under a high‑emission scenario (UNEP Baseline) ([unep.org](https://www.unep.org/resources/policy-and-strategy/air-quality-policies-nigeria?utm_source=chatgpt.com)).

projected\_AQI = round(current\_AQI \* 1.17)

| **State** | **Current AQI** | **Projected AQI (2045)** |
| --- | --- | --- |
| Borno | 23 | 27 |
| Cross River | 18 | 21 |
| Delta | 21 | 25 |
| Edo | 22 | 26 |
| Federal Capital Territory | 21 | 25 |
| Kano | 24 | 28 |
| Kebbi | 19 | 22 |
| Kwara | 22 | 26 |
| Lagos | 24 | 28 |
| Osun | 22 | 26 |
| Rivers | 23 | 27 |

## Causes of Poor Air Quality in Nigerian States

### Road Transport & Traffic Congestion

Vehicle exhaust and road dust contribute heavily to PM₂.₅ and PM₁₀ levels, particularly in megacities like Lagos and Kano, due to aging fleets and stop‑start traffic ([sciencedirect.com](https://www.sciencedirect.com/science/article/abs/pii/S0169809519312943?utm_source=chatgpt.com)).

### Industrial Emissions & Generators

Manufacturing plants and persistent power outages drive reliance on diesel generators; both emit NO₂, SO₂, CO, and fine particulates, often exceeding emissions from heavy trucks ([dicf.unepgrid.ch](https://dicf.unepgrid.ch/nigeria/pollution?utm_source=chatgpt.com)).

### Biomass, Waste & Agricultural Burning

Open burning of agricultural residues, municipal waste, and biomass for cooking adds organic carbon and black carbon to ambient air, especially in peri‑urban regions ([sciencedirect.com](https://www.sciencedirect.com/science/article/abs/pii/S0169809519312943?utm_source=chatgpt.com)).

### Harmattan & Construction Dust

Seasonal Sahara dust incursions (December–February) and urban construction projects elevate PM₁₀ concentrations, triggering acute respiratory episodes ([sciencedirect.com](https://www.sciencedirect.com/science/article/abs/pii/S0169809519312943?utm_source=chatgpt.com)).

## International Success Stories in AQI Improvement

### China: National Air Quality Action Plan

Launched in 2013, China invested US $270 billion to cut PM₁₀ by 10% by 2017. By 2020, major cities saw a 25% reduction in PM₂.₅ days through coal‑to‑gas conversions and stringent vehicle standards ([aqli.epic.uchicago.edu](https://aqli.epic.uchicago.edu/policy-impacts/china-national-air-quality-action-plan-2014/?utm_source=chatgpt.com), [acp.copernicus.org](https://acp.copernicus.org/articles/19/11303/2019/?utm_source=chatgpt.com)).

### United States: Clean Air Act

Since 1970, the Clean Air Act has driven a 62% drop in aggregate airborne pollutants, delivering vast health benefits via mandated emissions controls on vehicles and industry ([aqli.epic.uchicago.edu](https://aqli.epic.uchicago.edu/policy-impacts/united-states-clean-air-act/?utm_source=chatgpt.com), [epa.gov](https://www.epa.gov/transportation-air-pollution-and-climate-change/accomplishments-and-successes-reducing-air?utm_source=chatgpt.com)).

### European Union: Zero Pollution Action & Local Measures

Between 2005 and 2022, EU member states cut PM₂.₅‑related mortality by 45%, employing traffic‑management schemes, low‑emission zones, and real‑time monitoring under the 2030 zero‑pollution vision ([eea.europa.eu](https://www.eea.europa.eu/en/topics/in-depth/air-pollution?utm_source=chatgpt.com)).

### India: National Clean Air Programme (NCAP)

Since 2019, India’s NCAP targets 132 non‑attainment cities, aiming for a 40% PM₁₀ reduction by 2026 through stakeholder engagement and city‑level action plans; initial data show a 13% AQI drop from 2019 to 2023 ([pib.gov.in](https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1909910&utm_source=chatgpt.com), [sciencedirect.com](https://www.sciencedirect.com/science/article/abs/pii/S1352231024006496?utm_source=chatgpt.com)).

**Recommendations for Nigeria:** Adopt binding emission standards, phase out old diesel generators, expand clean‑energy access, and implement city‑level action plans modeled on NCAP—with robust monitoring and enforcement—to stabilize or reverse AQI trends by 2045.

## 

## **Future Works**

* **Local Data Integration:** Ingest real‑time feeds from CAR‑NASRDA or low‑cost sensor networks in Lagos, Abuja, and Port Harcourt.
* **Enhanced Forecasting:** Implement ARIMA, Prophet, or machine‑learning models for more accurate short‑term AQI forecasts.
* **Health Impact Module:** Integrate AirQ+ (WHO tool) to quantify morbidity/mortality impacts from pollutant exposure [Wikipedia – Die freie Enzyklopädie](https://de.wikipedia.org/wiki/AirQ%2B?utm_source=chatgpt.com).
* **Mobile App Development:** Wrap the Shiny dashboard in a mobile‑friendly interface to broaden public access.
* **Policy Simulation:** Model emission‑control scenarios (e.g., traffic restrictions) to project AQI improvements.