INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

A Pyspark-Sparksql Pipeline for Urban Air Quality Monitoring: 4-Year Trends and Who Assessment (2021–2024)

[1]Prince Nweke Onyeka., [2]Chinoso Job., [3]Onwe, Festus Chijioke

[1]Department of Data Analytics and Technology, University of Greater Manchester Manchester, United Kingdom

[2]School of Built Environment, Engineering, and Computing, LEEDSBECKETT University

[3]Information Technology Department, University of Port Harcourt, Rivers State, Nigeria

DOI: https://doi.org/10.51584/IJRIAS.2026.11070068

Received: 17 July 2026; Accepted: 22 July 2026; Published: 01 August 2026

ABSTRACT

Air quality monitoring datasets are inherently temporal, multi-pollutant, and long-horizon in character, making them well suited to scalable distributed computing frameworks that can handle the volume, velocity, and variety of multiyear daily observations while supporting transparent, SQL accessible analytical workflows. This paper presents a reproducible PySpark–SparkSQL analytical pipeline applied to four years (2021–2024) of daily urban air quality monitoring data, addressing five structured research questions about seasonal and temporal variability, pollutant drivers of the Air Quality Index (AQI), human-activity effects on ambient concentrations, and compliance with the revised WHO 2021 Air Quality Guidelines. Our findings document: consistently higher AQI values in winter months, with November peaking at a mean AQI of 342.13 across the four-year period; $PM_{2.5}$ and PM_{10} as the dominant AQI drivers with Pearson correlations of 0.87 and 0.91 respectively; marginal but consistent reductions in pollutant levels during holiday and weekend periods, with the strongest effects for trafficrelated NO_2; and a chronic, near-daily exceedance of WHO 24-hour mean thresholds for $PM_{2.5}$ (343–364 exceedance days per year) and PM_{10} (347–359 days). A previously unreported SO_2 exceedance spike in 2024—110 exceedance days compared to fewer than 30 in each prior year—is identified as an anomaly of policy significance indicating an emerging industrial emission source. The pipeline, implemented in Python (PySpark, Pandas, Matplotlib, Seaborn) on Google Colab, provides a reproducible, computationally scalable framework adaptable to other urban air quality monitoring contexts.

Keywords: Air quality, PySpark, SparkSQL, temporal analytics, $PM_{2.5}$, AQI, WHO guidelines, seasonal analysis, big data, urban environment.

INTRODUCTION

URBAN air quality monitoring generates continuous streams of multi-pollutant daily observations that, over multi-year periods, accumulate into datasets of sufficient volume and temporal depth to support the identification of systematic patterns, anomalies, and trends that are not visible in shorter-horizon analyses. The revised WHO 2021 Air Quality Guidelines, which substantially tightened the recommended exposure thresholds for $PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2, and CO N. P. Onyeka is with the School of Creative Arts and Technologies, University of Greater Manchester, Manchester, United Kingdom. Manuscript prepared for journal submission; corresponding author e-mail to be inserted prior to submission. relative to the 2005 guidelines [1], [2], amplify the analytical value of such datasets: by establishing stricter thresholds, they create a new population of exceedance events that requires systematic characterisation to support evidence-based policy and intervention planning.

Distributed computing frameworks are well suited to this analytical challenge. Apache Spark's in-memory distributed processing architecture, accessed through the PySpark Python API and the SparkSQL declarative query interface, enables scalable aggregation, filtering, and transformation of large tabular datasets at speeds and at data volumes that exceed the practical limits of single-machine Pandas operations. Critically, SparkSQL's ability to express complex analytical queries in standard SQL syntax lowers the barrier for adaptation and replication of the analysis by public health and environmental science practitioners who may have SQL fluency without specialised distributed computing expertise.

This paper presents a complete, reproducible PySpark– SparkSQL analytical pipeline applied to four years of daily urban air quality data (2021–2024), framing the analysis around five research questions motivated by the companion literature review [3]–[5] and directly relevant to the evidence needs of public health agencies, urban planners, and environmental regulators. The pipeline covers data preprocessing, feature engineering, exploratory data analysis, WHO-threshold exceedance benchmarking, and visualisation.

Related Work And Positioning

Temporal analysis of air quality monitoring data is well established in the environmental science literature, with approaches ranging from single-site time-series analysis to satellite-derived global spatiotemporal maps. Wei et al.'s global daily $PM_{2.5}$ study provides a systematic, high-resolution characterisation of spatial and temporal variability at a global scale [6]; Liu et al.'s satellite-based analysis of NO_2 seasonal variation documents urban-scale seasonal cycles at 2001–2018 temporal depth [7]; and Pan et al.'s INLA-based joint analysis of ozone and $PM_{2.5}$ introduces spatial statistical methods for characterising extreme pollution events [8].

What is less systematically represented in the published literature is the transparent, reproducible, accessible analytical pipeline perspective: a complete, documented workflow from raw multi-pollutant monitoring data through cleaning, feature engineering, EDA, WHO-threshold benchmarking, and visualisation, implemented in a scalable open-source computational framework with practical adaptability to monitoring contexts beyond the study site. This paper addresses that gap, with specific contributions differentiating it from the existing literature: (i) a four-year, daily-resolution temporal coverage enabling year-on-year change detection that shorter studies cannot support; (ii) an IQR-based Winsorisation outlier treatment that preserves dataset completeness compared to outlier-removal approaches; (iii) a structured five-question analytical framing directly aligned with the evidence needs of public health and regulatory stakeholders; and (iv) the identification of a previously unreported 2024 SO_2 exceedance spike warranting further industrial emission source investigation.

Data And Study Design

A. Dataset

The dataset, Airquality.csv, contains daily air quality observations covering the period 1 January 2021 to 31 December 2024, encompassing 1,461 observations across seven measured variables: $PM_{2.5}$ (μg/m^3), PM_{10} (μg/m^3), NO_2 (μg/m^3), SO_2 (μg/m^3), CO (mg/m^3), ozone (μg/m^3), and the composite Air Quality Index (AQI). Additional categorical variables include a daily public holiday indicator and the date field from which temporal features are engineered. The dataset provides sufficient multi-year temporal coverage to detect both within-year seasonal patterns and inter-annual trends.

B. Research Questions

The analysis is structured around five research questions (RQ), each mapping to a specific analytical output:

- RQ1: How does the AQI vary across months and seasons?
- RQ2: Which pollutants most strongly influence AQI, and when do they peak?
- RQ3: Does AQI differ significantly between holiday and non-holiday periods?
- RQ4: Are pollutant levels higher on weekdays compared to weekends?

- RQ5: Which pollutants most frequently exceed WHO 2021 24-hour mean thresholds, and how does this vary by month and year?

METHODOLOGY

A. Data Preprocessing Pipeline

The preprocessing pipeline was implemented in Python using Pandas for initial data manipulation and PySpark DataFrames for subsequent scalable operations. The column naming convention was standardised at the outset: the $PM_{2.5}$ column was renamed to PM2_5 to ensure SparkSQL compatibility, since column names containing decimal points are not valid SQL identifiers.

Missing values were detected and handled via PySpark's na.fill method, using forward-fill imputation to preserve temporal continuity in the time series, since complete case deletion would disproportionately affect winter months when sensor gaps are more frequent due to adverse weather.

Date transformation and feature engineering: The date column was converted to PySpark TimestampType using to_timestamp, enabling extraction of Year, Month, Day, and Day Of Week. Season indicators were created from month groups: Winter (December–February), Spring (March–May), Summer (June–August), Autumn (September– November). Weekend and weekday boolean flags were derived from Day Of Week.

Outlier detection and treatment: Outliers in all numeric pollutant columns were identified using the Interquartile Range (IQR) method. For each column, the lower bound was defined as $Q_1-1.5\times$ IQR and the upper bound as $Q_3+1.5\times$ IQR. Rather than removing flagged observations—which would reduce the dataset and potentially bias seasonal aggregations by disproportionately eliminating winter pollution events— extreme values were capped at these bounds via Winsorisation, preserving dataset completeness while reducing the influence of isolated measurement spikes. Histograms and boxplots confirmed that post-Winsorisation distributions were substantially less right-skewed than pre-treatment, particularly for $PM_{2.5}$ and PM_{10}.

B. Analytical Methods

Descriptive statistics: Summary statistics (mean, min, max, standard deviation) were generated for all pollutant and AQI variables using PySpark's describe method and SparkSQL GROUP BY aggregations.

Correlation analysis: Pearson correlation coefficients between AQI and each pollutant were computed using PySpark's stat.corr function over the clean, Winsorised dataset. Results were visualised as a heatmap using Seaborn after converting SparkSQL results to Pandas DataFrames.

Temporal aggregation: SparkSQL SELECT ... GROUP BY Month, Season queries computed mean AQI and mean pollutant concentrations for each month and season across the four-year study period. These grouped means were visualised as bar plots and line plots to reveal within-year patterns.

Comparative analysis: Pollutant and AQI levels were compared across holiday/non-holiday and weekday/weekend groups using SparkSQL GROUP BY IsHoliday and GROUP BY IsWeekend queries, supplemented by boxplots to display distributional differences rather than only mean differences.

WHO threshold exceedance analysis: Binary exceedance flags were computed for each pollutant on each day, comparing the daily observation against the WHO 2021 24-hour mean thresholds (Table I). SparkSQL SUM(exceedance_flag) aggregations by month and year generated the exceedance frequency counts underlying RQ5.

C. Justification of Methodological Choices

PySpark and SparkSQL were selected over single-machine Pandas for three reasons. First, multi-year daily data with multiple pollutant columns and multiple derived features generate

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Table I Who 2021 Air Quality Guideline Values (24-Hour Mean) Used For Exceedance Benchmarking

Pollutant	Threshold	Units
PM2.5	15	$\mu g/m^3$
PM_{10}	45	$\mu g/m^3$
NO_2	25	$\mu g/m^3$
SO_2	40	$\mu g/m^3$
CO	4	mg/m^3
O_3	100	$\mu g/m^3$

a data volume and computational complexity that benefits from Spark's lazy evaluation and in-memory distributed processing, even on a single-node Google Colab environment. Second, SparkSQL enables transparent, readable, version-controlled analytical queries that can be inspected and replicated by nonspecialist practitioners. Third, the PySpark framework scales directly to larger monitoring networks, longer time horizons, and higher-frequency data without code changes.

Winsorisation was chosen over outlier deletion because air quality outliers—unusually high $PM_{2.5}$ concentrations during pollution episodes or instrument calibration events—are often scientifically meaningful and deleting them would distort both the sample size and the seasonal distribution of extreme events. IQR-based bounds were preferred over standarddeviation-based bounds because the pollutant distributions are right-skewed, violating the normality assumption implicit in standard-deviation bounds.

RESULTS

A. RQ1: Seasonal and Monthly AQI Variation

The monthly mean AQI (averaged across 2021–2024) reveals a strongly winter-dominant pattern. November recorded the highest monthly mean AQI of 342.13, followed by January (305.24) and December (297.26), while the summer months

June, July, and August exhibited the lowest values in the 180–220 range. The seasonal means confirm this pattern: winter (December–February) has the poorest air quality overall, followed by spring and autumn, with summer consistently recording the best air quality.

This winter dominance is consistent with the epidemiological and monitoring literature describing enhanced $PM_{2.5}$ and PM_{10} concentrations during cold months due to stable anticyclonic meteorological conditions that reduce atmospheric mixing, combined with elevated heating emissions from residential and commercial combustion sources [3]. The interannual consistency of the November peak—observed in each of the four years—suggests a structural, meteorologicallydriven pattern rather than an isolated anomaly.

B. RQ2: Pollutant Influence on AQI and Seasonal Peaks

Pearson correlations between AQI and individual pollutants across the full four-year dataset identify $PM_{2.5}$ ($r=0.87$) and PM_{10} ($r=0.91$) as the dominant AQI drivers. NO_2 shows a moderate positive correlation ($r=0.52$). O_3, SO_2, and CO each exhibit weak correlations with AQI ($r<0.3$), indicating that they are not primary drivers of the composite index under the conditions observed in this dataset.

The correlation heatmap visualises these relationships, with the strong $PM_{2.5}$–AQI and PM_{10}–AQI associations visually dominant. Monthly trend plots show that $PM_{2.5}$ and PM_{10} peak in November–December and reach their lowest values in June–August, consistent with RQ1's seasonal pattern. By contrast, O_3 shows a spring–summer peak (April–May) driven by photochemical reactions with UV radiation, consistent with the established inverse seasonal relationship between O_3 and primary combustion pollutants [8]. SO_2 peaks in April–May (spring), while NO_2 peaks in November, reflecting enhanced combustion activity and reduced atmospheric mixing.

C. RQ3: Holiday versus Non-Holiday AQI

Grouping days by the public holiday indicator shows that mean AQI is marginally lower on holidays compared to nonholidays (approximately 3–5% reduction in mean across the four-year period). $PM_{2.5}$ and PM_{10} remain persistently high on holidays—only 10–15% below non-holiday means—while NO_2 shows the most pronounced reduction, consistent with its predominantly traffic-related source profile [4]. These results indicate that while traffic-related emission reductions on public holidays are detectable, they are insufficient to bring the AQI or particulate concentrations close to WHO daily guideline levels. Background particulate concentrations from residential heating and long-range transport maintain high PM levels regardless of daily activity patterns.

This finding echoes the conclusion from COVID-19 lockdown studies: traffic reduction alone, while measurably reducing NO_2, cannot resolve the chronic PM exceedance problem [5], [9].

D. RQ4: Weekday versus Weekend Pollutant Levels

Weekday mean PM_{10} and AQI are both higher than weekend means, by approximately 5–8% across the four-year sample. $PM_{2.5}$ shows a similar weekday elevation. NO_2 also exhibits higher weekday levels, consistent with its trafficdominated source. The boxplot distributions for all pollutants show overlapping interquartile ranges between weekday and weekend groups, indicating that the distributional overlap is large even where the means differ—weekdays produce more frequent extreme values rather than a uniform concentration shift.

The pattern confirms that weekday–weekend activity cycles impose a detectable signature on ambient concentrations, but the magnitude is modest compared to the seasonal variation documented in RQ1. The practical implication for urban air quality management is that short-term activity modulation produces limited improvements relative to the structural emission source reductions required to address the persistent winterseason PM burden.

E. RQ5: WHO Threshold Exceedances

The exceedance analysis documents a chronic exceedance burden for $PM_{2.5}$ and PM_{10} that is qualitatively different from the other pollutants in severity and persistence. Table II summarises annual exceedance day counts for all pollutants.

Table II Annual Who 2021 24-Hour Mean Exceedance Day Counts By Pollutant (2021–2024)

Pollutant	2021	2022	2023	2024	Trend
PM2.5	364	343	358	361	Persistently near-daily
PM_{10}	359	347	352	356	Persistently near-daily
NO_2	277	146	182	300	Rising in 2024
SO_2	0	24	29	110	Spike in 2024
CO	0	0	0	0	No exceedances

O_3	0	2	6	4	Minimal

$PM_{2.5}$ and PM_{10} chronic exceedance: Both pollutants exceed their WHO 24-hour mean thresholds (15 and 45μg/m^3 respectively) on 343–364 days per year throughout the study period. This near-daily chronic exceedance leaves essentially no days per year within WHO safe limits for either pollutant, representing the most critical air quality challenge in this dataset.

NO_2 variable exceedance: NO_2 exhibits pronounced yearto-year variability (277 days in 2021; 146 in 2022; 182 in 2023; 300 in 2024), with the highest exceedance burden in 2021 and 2024. The 2024 elevation—corresponding to approximately 82% of days exceeding the 25μg/m^3 threshold— marks a deterioration warranting investigation of emission source changes.

SO_2 2024 spike: The most striking finding from the exceedance analysis is the dramatic increase in SO_2 exceedances in 2024: from 0 days in 2021, 24 days in 2022, and 29 days in 2023, to 110 days in 2024—a nearly fourfold increase over 2023 that represents a qualitative rather than incremental change in SO_2 ambient levels. This pattern is inconsistent with natural meteorological variability and strongly suggests an industrial emission source change (e.g. new industrial facilities, changes in fuel composition or quality, or enforcement failures in desulphurisation requirements) that warrants urgent regulatory investigation [7].

December and November exceedancedominance:

Monthly aggregation of exceedance counts identifies December and November as the months with the highest multipollutant exceedance burden, consistent with the winterdominant AQI pattern in RQ1.

DISCUSSION

A. Synthesis of Findings

The five research questions produce a coherent picture of air quality dynamics in the study location. Particulate matter ($PM_{2.5}$ and PM_{10}) is both the dominant AQI driver and the dominant source of WHO exceedance burden, with chronic near-daily exceedance throughout all four years of the study period. This represents a public health emergency in the epidemiological sense: the evidence linking $PM_{2.5}$ exposure above 15μg/m^3 to cardiovascular and respiratory mortality is well established [3], [10], [11], and the near-daily exceedance documented here implies that the entire resident population is experiencing chronic excess risk every year.

NO_2 exceedances are highly variable year-on-year, suggesting sensitivity to factors that PM exceedances do not exhibit. This may reflect sensitivity to meteorological conditions (mixing height, wind direction) that vary inter-annually, or to policy-driven changes in traffic emission standards.

The SO_2 2024 anomaly is the most actionable finding from a regulatory perspective. The abrupt rise from fewer than 30 exceedance days per year to 110 suggests an identifiable, specific source change rather than a gradual trend. Industrial source auditing and enhanced monitoring of SO_2 point-source emissions in 2024–2025 is the appropriate regulatory response.

B. Human Activity Effects

The consistent but modest weekday/weekend and holiday/non-holiday differences confirm that traffic and activity-related emission reductions are detectable in ambient monitoring data but are insufficient as standalone pollution control strategies. This is directly consistent with the COVID19 lockdown literature: even the extreme emission reductions of 2020 lockdowns reduced NO_2 substantially but left $PM_{2.5}$ only modestly improved, because heating, agriculture, and long-range transport continue to contribute PM regardless of local traffic activity [4], [5], [9].

C. Methodological Contributions

The pipeline's primary technical contribution is the combination of SparkSQL-based temporal aggregation with Winsorisation-preserved dataset completeness, which enables seasonal analyses that outlier-deletion approaches would bias. The open-source, Google-Colab-implementable design ensures that the framework is accessible to practitioners without highperformance computing infrastructure, which is particularly relevant for public health monitoring agencies in resourceconstrained settings.

RECOMMENDATIONS AND POLICY IMPLICATIONS

Based on the five research questions and the cross-cutting findings, four evidence-based policy recommendations follow directly from the analysis.

Recommendation 1 (Particulate reduction priority): The chronic near-daily exceedance of WHO $PM_{2.5}$ and PM_{10} thresholds demands prioritisation of particulate reduction strategies above all other pollution control objectives. These should include stricter vehicle emission standards, residential solid fuel combustion restrictions, industrial dust control, and clean-fuel transition support programmes. The universality of the exceedance burden—no month in any year meets WHO guidelines—indicates that incremental improvements are insufficient and structural policy change is required.

Recommendation 2 (Winter action plans): The concentration of exceedance burden in November–February motivates targeted seasonal interventions: winter low-emission zones, temporary factory emission curtailments during stable weather episodes, and enhanced monitoring-based public health advisories on poor-air-quality days.

Recommendation 3 (SO_2 industrial audit): The 2024 SO_2 spike requires immediate regulatory investigation: industrial emission source audits, verification of desulphurisation equipment compliance, and high-frequency monitoring of pointsource emissions in the vicinity of the monitoring network.

Recommendation 4 (Data infrastructure investment): The analytical value demonstrated by four years of daily monitoring data—particularly the ability to detect the SO_2 2024 anomaly, to quantify year-on-year NO_2 variability, and to establish the baseline exceedance burden against which future intervention effects can be measured—underscores the value of sustained, systematic, high-frequency monitoring and accessible data infrastructure. The PySpark framework described here provides a scalable technical foundation for such infrastructure.

CONCLUSION

This paper has presented a complete, reproducible PySpark– SparkSQL analytical pipeline for multi-year urban air quality temporal analytics, addressing five structured research questions and generating findings with direct policy relevance. The principal empirical contributions are: the quantification of winter-dominant AQI seasonality (November peak of 342.13); the identification of $PM_{2.5}$ (r=0.87) and PM_{10} (r=0.91) as overwhelmingly dominant AQI drivers; the confirmation that human-activity cycles (holidays, weekends) produce detectable but modest NO_2 reductions while leaving the PM burden largely unaffected; the documentation of chronic near-daily WHO exceedances for $PM_{2.5}$ and PM_{10} throughout 2021– 2024; and the identification of an anomalous SO_2 exceedance spike in 2024 requiring urgent regulatory investigation. A companion literature review provides the epidemiological and policy context establishing the public health significance of these findings [1], [4], [10], [12].

LIMITATIONS

The dataset covers a single monitoring location or monitoring network area; spatial heterogeneity in concentrations within the study area cannot be addressed with these data. The public holiday indicator does not distinguish between holiday types (national vs. local), which may affect the interpretability of RQ3 findings.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

The SO_2 2024 anomaly is identified as an empirical observation warranting investigation; the absence of point-source emission data in this dataset means that source attribution cannot be confirmed within the analysis. No predictive model was developed; the analysis is descriptive and exploratory, providing the baseline from which predictive and prescriptive analytics could subsequently be developed.

REFERENCES

1. World Health Organization, "WHO global air quality guidelines: Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide," World Health Organization, Geneva, Tech. Rep., 2021. [Online]. Available: https://www.who.int/publications/ i/item/9789240034228
2. B. Hoffmann, H. Boogaard, A. de Nazelle, Z. J. Andersen, M. Abramson, M. Brauer, B. Brunekreef, F. Forastiere, W. Huang, H. Kan, J. D. Kaufman, K. Katsouyanni, M. Krzyzanowski, N. Kuenzli, F. Laden, M. Nieuwenhuijsen, A. Mustapha, P. Powell, M. Rice, A. Roca-Barcelo,´ C. J. Roscoe, A. Soares, K. Straif, and G. Thurston, "WHO air quality guidelines 2021: Aiming for healthier air for all – a joint statement by medical, public health, scientific societies and patient representative organisations," International Journal of Public Health, vol. 66, p.1604465, 2021.
3. C. Li, M. S. Hammer, A. van Donkelaar, A. I. Lyapustin, A. M. Sayer, N. C. Hsu, J. S. Apte, L. Zhang, Q. Zhang, R. T. Burnett, and R. V. Martin, "Reversal of trends in global fine particulate matter air pollution," Nature Communications, vol. 14, p. 5150, 2023.
4. Z. S. Venter, K. Aunan, S. Chowdhury, and J. Lelieveld, "COVID19 lockdowns cause global air pollution declines," Proceedings of the National Academy of Sciences, vol. 117, no. 32, pp. 18984–18990, 2020.
5. L. Saha, A. Kumar, S. Kumar, J. Korstad, S. Srivastava, and K. Bauddh, "The impact of the COVID-19 lockdown on global air quality: A review," Environmental Sustainability, vol. 5, no. 1, pp. 53–69, 2022.
6. J. Wei, Z. Li, A. Lyapustin, Y. Zhan, L. Liu, Z. Yang, Y. Zhang, R. Liu, and Z. Ma, "First close insight into global daily gapless 1 km $PM_{2.5}$ pollution, variability, and health impact," Nature Communications, vol. 14, p. 8349, 2023.
7. S. Liu, P. Valks, G. Curci, Y. Chen, L. Shu, J. Jin, S. Sun, D. Pu, X. Li,
8. J. Li, X. Zuo, W. Fu, Y. Li, P. Zhang, X. Yang, T.-M. Fu, and L. Zhu, "Satellite NO_2 retrieval complicated by aerosol composition over global urban agglomerations: Seasonal variations and long-term trends (2001– 2018)," Environmental Science & Technology, vol. 58, no. 18, 2024.
9. J. Pan, K. He, K. Wang, Q. Mu, and C. Ling, "Spatio-temporal joint analysis of extreme ozone and moderate $PM_{2.5}$ in california with INLA approach," arXiv preprint arXiv:2404.14446, 2024.
10. C. He, S. Hong, L. Zhang, H. Mu, A. Xin, Y. Zhou, J. Liu, N. Liu, Y. Su, Y. Tian, B. Ke, Y. Wang, and L. Yang, "Global, continental, and national variation in $PM_{2.5}$, O_3, and NO_2 concentrations during the early 2020 COVID-19 lockdown," Atmospheric Pollution Research, vol. 12, no. 3, pp. 136–145, 2021.
11. Y. Zhang, Z. Li, J. Wei, Y. Zhan, L. Liu, Z. Yang, Y. Zhang, R. Liu, and Z. Ma, "Long-term exposure to ambient NO_2 and adult mortality: A nationwide cohort study in china," Journal of Advanced Research, vol. 41, pp. 13–22, 2022.
12. X. Chen, L. Qi, S. Li, and X. Duan, "Long-term NO_2 exposure and mortality: A comprehensive meta-analysis," Environmental Pollution, vol. 341, p. 122971, 2024.
13. C. Tonne, E. Puzzolo, N. Kunzli, Z. J. Andersen, I. Annesi-Maesano,¨
14. F. Forastiere, A. Haines, M.-E. Heroux, F. J. Kelly, M. Nieuwenhuijsen,´ L. Perez, and S. Vardoulakis, "Promoting health through climate change mitigation in Europe," The Lancet Planetary Health, vol. 9, no. 1, pp. e15–e25, 2025.