International Journal of
Technology and Systems
(IJTS)

Exploratory Data Analytics of Air Pollutant Data for Air Quality Management Application

Oluwatomisin Ajayi

IPRJB

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024

www.iprjb.org

Exploratory Data Analytics of Air Pollutant Data for Air Quality Management Application

[1*] Omoniyi Samuel Aiyenuro and
[2]Oluwatomisin Ajayi

Article History

Received 14th May 2024

Received in Revised Form 19th June 2024

Accepted 22nd July 2024

How to cite in APA format:

Omoniyi, S., & Ajayi, O. (2024). Exploratory Data Analytics of Air Pollutant Data for Air Quality Management Application. *International Journal of Technology and Systems*, *9*(2), 82–107. https://doi.org/10.47604/ijts.2790

Abstract

Purpose: With escalating concerns over the detrimental impacts of air pollution on public health and the environment, this study embarks on a comprehensive exploration of air quality dynamics, focusing on Ogun state metropolis of Nigeria. The primary objective is to contribute to the burgeoning field of air quality management by harnessing the power of data analytics, exploratory data analysis, and advanced python programming libraries. The study seeks to address critical questions regarding the current state of air quality, major sources of pollution, government interventions, and individual contributions to better air quality.

Methodology: The research employs Exploratory Data Analysis to unveil intrinsic patterns and trends within historical air quality datasets extracted from IoT devices located across Ogun state. This initial phase aims to discern key statistical parameters, including mean concentrations, variations, and correlations among various pollutants. Subsequently, the study employs Topic Modeling to extract latent themes and sentiments from qualitative data, specifically focusing on public opinions gathered through digital surveys. Resulting topics provide insights into public perceptions regarding air quality, pollution sources, and the efficacy of governmental interventions.

Findings: Experimental result is revealing of the diversity of the dataset. The EDA returned a mean Particulate Matter concentration of approximately 12.65 µg/m³, while the mean Nitric Oxide concentration is approximately 13.56 µg/m³. Notable correlations include a strong positive correlation between PM2.5 and PM10 (0.69), indicating a substantial association between fine and coarse particulate matter. Additionally, there is a noteworthy positive correlation between NO and Nox (0.97), suggesting a high degree of correlation between nitric oxide and nitrogen oxide levels. However, negative correlations are observed, such as the substantial negative correlation between RH (relative humidity) and PM2.5 (-0.46), implying an inverse relationship between humidity and fine particulate matter levels. The study develops an innovative air quality management application.

Unique Contribution to Theory, Practice and Policy: The application aims to empower both the public and policymakers with actionable insights, fostering informed decision-making and collaborative efforts towards improving air quality. It is highly recommended that industry experts should continue to imbibe ethical standards in data acquisition and deployment practices.

Keywords: *Air Quality Management, Topic Modelling, Exploratory Data Analysis*

INTRODUCTION

In recent years, the importance of effective air quality management has garnered increasing attention, both globally and within academic circles. The significance of maintaining high air quality is paramount, given its direct and indirect impacts on human health, environmental sustainability, and overall well-being (Arroyo, Herrero, Suárez, & Lozano, 2019). As urbanization and industrialization continue to rise, so do concerns about air pollution, prompting a surge in academic interest and research endeavors to comprehensively understand and address this multifaceted challenge (Odekanle, Fakinle, Odejobi, Akangbe, Sonibare, Akeredolu, Oladoja, 2022) Academia plays a pivotal role in advancing our knowledge of air quality dynamics, pollution sources, and the potential consequences for ecosystems and human populations.

This heightened interest stems from a recognition of the intricate interplay between air quality, public health, and sustainable development (Popović, Radovanovic, Vajs, Drajic, & Gligorić, 2022). The academic community's engagement in researching and developing effective air quality management strategies reflects a commitment to mitigating the adverse effects of air pollution and fostering a healthier and more sustainable future for communities worldwide. The dynamic nature of air quality issues demands interdisciplinary collaboration and a continuous scholarly pursuit, positioning academia as a driving force in the quest for innovative solutions to ensure breathable, clean air for all (Asha, Natrayan, Geetha, Beulah, Rene, Sumathy, Varalakshmi, Neelakandan, 2022)

In addressing the challenges posed by air quality management, the integration of cutting-edge technologies has emerged as a transformative and indispensable approach. Technological advancements have played a pivotal role in revolutionizing the monitoring, assessment, and management of air quality. The deployment of state-of-the-art sensors, Internet of Things (IoT) devices (Asha, et al., 2022), and cloud computing solutions (Arogundade, Abayomi, Abayomi-Alli, Misra, Alonge, Olaleye, Ahuja, 2021) has enabled real-time data collection and analysis, offering a more comprehensive and timely understanding of air pollution dynamics. The use of data science algorithms and data analytics tools has further facilitated pattern recognition (Olaleye, Arogundade, Johnbosco, Sadare, Azeez, Opatunji, Tewogbade, Akintunde, Saminu, 2023) and the identification of pollution sources, prediction of air quality trends, and development of targeted mitigation strategies (Asha, et al., 2022).

The marriage of technology with air quality management has not only enhanced the accuracy and precision of monitoring systems but has also expanded the scope of research in this field. Remote sensing technologies and satellite-based observations provide a broader spatial perspective, allowing for the monitoring of air quality on regional and global scales. Additionally, the development of sophisticated air quality models, coupled with the processing power of modern computers, has enabled researchers to simulate complex atmospheric processes, assess the impacts of different emission scenarios (Wang, Du, Zhao, Zhou, Russo, Xi, Zhang, Zhou, Chengjun, 2022) and evaluate the effectiveness of various regulatory measures.

Furthermore, the advent of artificial intelligence (AI) has significantly contributed to the refinement of predictive models and the extraction of valuable insights from vast datasets. AI-driven systems can identify patterns, correlations, and outliers, offering a more nuanced understanding of air quality variations and potential health risks (Zhang, Pan, Yu, & Liu, 2022). The seamless integration of technology into air quality management has not only accelerated the pace of research but has also facilitated the development of innovative and adaptive solutions to combat air pollution. The synergy between technology and air quality management has marked a paradigm shift in our ability to comprehend, monitor, and address the complexities of air pollution. This technological renaissance underscores the interdisciplinary nature of contemporary air quality research and emphasizes the pivotal role of technology in fostering sustainable practices and safeguarding environmental health (Wang T. , et al., 2022).

Against the backdrop of the increasing significance of effective air quality management and the pivotal role played by technology in advancing understanding of air pollution dynamics, this study adopts a comprehensive and innovative approach. Leveraging the capabilities of advanced data analytics, our research aims to delve into the intricate facets of air quality within the context of Ogun state metropolis of Nigeria. The study seeks to build upon the existing body of knowledge by exploring novel insights into the major sources of air pollution, visualizing air quality trends, and offering a one-stop application for air quality managers to get summary statistics of primary data at their disposals. By employing topic modelling for analyzing citizens opinions on the prevalent challenges and future prospects, this research endeavors to contribute valuable findings that can inform evidence-based policies and initiatives for sustainable air quality improvement.

Existing studies fails to adopt the incorporation of an exploratory data analysis functionality in their approach, for a use case study that is hugely data-driven. This problem is to be addressed in this study. The interdisciplinary nature of our approach, encompassing aspects of data science, environmental science, and technology integration, reflects a commitment to addressing the complexities of contemporary air quality challenges through a holistic lens. The rest of the paper is presented in the following ways. Section 2 reviews and discusses existing literatures, while section 3 introduces the conceptual framework of the study. The experimental results are discussed in section 4 while the study is concluded in section 5.

LITERATURE REVIEW

Air quality issues pose significant challenges across the African continent, with Nigeria, and more specifically, Ogun State, grappling with the complex consequences of air pollution. In Africa, rapid urbanization, industrialization, and population growth contribute to heightened levels of air pollutants. Urban centers, often characterized by increased vehicular emissions, industrial activities, and inadequate waste management, experience elevated concentrations of pollutants such as particulate matter (PM), nitrogen dioxide (NO2), and sulfur dioxide (SO2) (Kumar & Pande, 2023). Nigeria, as one of Africa's most populous nations and a hub of industrial activities, faces severe air quality challenges (Odekanle, et al., 2022). Industries, vehicular traffic, and open waste burning contribute substantially to the emission of pollutants, impacting both urban and rural areas (Chukwu, Morse, & Murphy, 2022).

Ogun State, located in the southwestern region of Nigeria, shares in these broader challenges. The state's industrial zones, including areas with manufacturing and processing facilities, contribute to localized air pollution (Odekanle, et al., 2022). The use of generators, common in the face of unreliable power supply, further adds to pollutant levels (Chukwu, Morse, & Murphy, 2022). Moreover, agricultural practices, including land clearing and burning, contribute to the release of pollutants into the atmosphere and the consequences of these air quality issues are far-reaching, affecting public health, the environment, and overall quality of life (Chauhan & Shah, 2021).

Inadequate regulatory frameworks and enforcement mechanisms have also been mentioned to exacerbate the problem (Odekanle, et al., 2022), allowing for unchecked emissions. While efforts to address air quality concerns are underway, a comprehensive and coordinated approach is crucial. This involves strengthening regulatory measures, promoting sustainable industrial practices, enhancing public awareness, and investing in advanced air quality monitoring systems (Almalawi, Alsolami, Khan, Asif, Alkhathlan, Fahad, Irshad, Qaiyum, Alfakeeh, Ahmed, 2022). Collaborative initiatives at the national and regional levels are essential to effectively combat air pollution in Ogun State, Nigeria, and throughout the African continent, while addressing these challenges is pivotal for safeguarding public health, preserving ecosystems, and fostering sustainable development in the region (Odekanle, et al., 2022). The health impacts of poor air quality is far reaching, and are discussed in the following section.

Health Impacts of Poor Air Quality

The detrimental health effects of poor air quality pose a significant and escalating challenge in various regions across the world, with Africa, Nigeria, and Ogun State being no exception. Exposure to air pollution, characterized by elevated levels of particulate matter (PM), nitrogen dioxide (NO2), sulfur dioxide (SO2), and other pollutants, has been linked to a range of adverse health outcomes, affecting respiratory, cardiovascular, and overall well-being (Odekanle, et al., 2022).

In Africa, where rapid urbanization and industrialization often outpace environmental regulations, the health impacts of poor air quality are particularly pronounced. Urban centers grappling with increased vehicular emissions, industrial activities, and biomass burning contribute to elevated pollution levels. The population, especially in densely populated urban areas, faces heightened risks of respiratory diseases, cardiovascular complications, and other health issues attributed to prolonged exposure to polluted air (Chukwu, Morse, & Murphy, 2022). Nigeria, being one of the most populous countries in Africa, contends with significant air quality challenges. Urban areas such as Lagos, Port Harcourt, and the federal capital territory, Abuja, experience elevated pollution levels due to a combination of vehicular emissions, industrial activities, and inefficient waste management practices. These urban centers bear the brunt of health consequences, including an increase in respiratory infections, aggravated asthma, and cardiovascular diseases among the exposed population (Kumari & Toshniwal, 2022).

In Ogun State, the impact of poor air quality is intertwined with industrialization and urban development. Areas with a concentration of industrial facilities often experience higher pollution levels, impacting the health of nearby residents (Odekanle, et al., 2022). Chronic exposure to pollutants like particulate matter can lead to respiratory issues such as bronchitis and asthma, particularly affecting vulnerable populations such as children and the elderly. Respiratory diseases, including chronic obstructive pulmonary disease (COPD) and lung cancer, are among the most documented health impacts of poor air quality (Liu, Wang, Li, Wen, & Deng, 2022). Prolonged exposure to fine particulate matter and other pollutants can cause inflammation of the respiratory tract, compromising lung function over time. Cardiovascular effects, such as increased risk of heart attacks and strokes, are also associated with air pollution, contributing to a heightened burden of non-communicable diseases.

It is important to note that the health impacts are not uniform across different demographic groups. Vulnerable populations, including children, the elderly, and individuals with pre-

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024
www.iprjb.org

existing health conditions, are more susceptible to the adverse effects of poor air quality. Additionally, socio-economic factors often play a role in determining the level of exposure and access to healthcare resources, further exacerbating health disparities. Addressing the health impacts of poor air quality in Africa, Nigeria, and Ogun State requires a multi-faceted approach. Implementing and enforcing stringent environmental regulations, promoting sustainable urban planning, investing in cleaner technologies, and raising public awareness are critical steps toward mitigating the health risks associated with air pollution (Odekanle, et al., 2022). Public health interventions that target vulnerable groups and enhance healthcare infrastructure can contribute to building resilient communities in the face of ongoing environmental challenges.

Technological Trends in Air Quality Monitoring

In recent years, the landscape of air quality monitoring has undergone a profound transformation, marked by the integration of cutting-edge technologies that promise unprecedented precision, scalability, and accessibility in assessing environmental air conditions. The traditional methods of air quality measurement, relying on stationary monitoring stations with limited coverage, are giving way to a new era of technological innovation (Asha, et al., 2022). The rise of mobile applications dedicated to air quality monitoring represents a paradigm shift in public engagement. Individuals can now access real-time air quality information through user-friendly apps, fostering awareness and encouraging collective efforts towards pollution reduction. These applications often incorporate features like personalized health recommendations based on current air quality conditions, creating a direct link between data and individual well-being.

Sensors

One of the notable trends is the widespread adoption of sensor technologies (Kumari & Toshniwal, 2022), particularly low-cost and portable sensors that enable real-time monitoring across diverse locations. These sensors leverage advances in miniaturization, connectivity, and data analytics, allowing for the deployment of monitoring networks in urban areas, industrial zones, and even on individual vehicles (Chukwu, Morse, & Murphy, 2022). This democratization of monitoring empowers communities to actively participate in assessing and addressing local air quality concerns.

Internet of Things (IoT)

The IoT has emerged as a cornerstone in modern air quality monitoring (Asha, et al., 2022). IoT-enabled devices facilitate seamless data collection, transmission, and analysis (Liu, Wang, Li, Wen, & Deng, 2022). These interconnected devices, often embedded in urban infrastructure or wearable gadgets, provide a dynamic and comprehensive understanding of air quality variations. The data generated by IoT devices contribute to robust datasets, enhancing the accuracy of pollution mapping and supporting informed decision-making.

Remote Sensing Technologies

Remote sensing technologies, including satellite-based monitoring and aerial drones, have expanded the spatial scope of air quality assessments. Satellites equipped with advanced sensors offer a macroscopic view, capturing regional and global air quality trends (Asha, et al., 2022). Drones, on the other hand, provide a finer resolution and the flexibility to navigate challenging terrains, enabling detailed studies of localized pollution sources.

Machine learning and Artificial Intelligence (AI) algorithms

These technologies play a pivotal role in processing the vast datasets generated by modern monitoring technologies (Olaleye T. O., Arogundade, Misra, Abayomi-Alli, & Kose, 2023). These intelligent systems can identify complex patterns, predict pollution levels, and offer insights into the correlation between different pollutants. The integration of AI enhances the predictive capabilities of ICT-based models (Arogundade, et al., 2021), thereby contributing to more effective pollution management strategies use case.

Review of Existing Studies

Existing studies in the study area are discussed in this section of the study. Their methodologies, case studies, aims and strengths are discussed in succession.

Zhang and colleagues (Zhang, Pan, Yu, & Liu, 2022) conducted a comprehensive study in China, employing data mining techniques to analyze air quality on a large scale. Their work focused on big data analytics, utilizing a resource orchestration approach. The authors generated a process model to understand the dynamics of air quality in China. The strength of their study lies in the application of advanced data mining methods to handle large datasets, providing valuable insights into the complex interactions influencing air quality. The primary data source for their study comprised extensive datasets related to China's air quality.

Asha et al. (2022) contributed to the field by designing an automated environmental toxicology-based air pollution monitoring system. Their study incorporated IoT, cloud computing, and AI, specifically the Elman Neural Network. The methodology included parameter tuning and the implementation of a real-time alarming system. The strength of their work lies in the integration of cutting-edge technologies for effective real-time monitoring. The primary data source for their study was real-time data, enhancing the accuracy and timeliness of their findings. (Kumar & Pande, 2023) focused on air pollution predictive analytics in India, utilizing machine learning techniques such as Naïve Bayes, Support Vector Machine, and XGBoost. They implemented exploratory data analysis (EDA) and addressed class imbalance through minority oversampling. However, their work lacked consideration of the bias-variance tradeoff. The strength of their study lies in the comprehensive analysis of a 6-year dataset, providing a robust foundation for predictive modeling. The primary data source for their study was a secondary dataset spanning six years. Kumari and Toshniwal (2022) study case study was in India. They evaluated the effect of lockdown on air quality in three Indian cities. Employing central tendencies and variation analysis, they adopted a phased-primary data collection approach.

The strength of their work lies in capturing the temporal variations in air quality during the lockdown period. The primary data source for their study was collected in phases from March 2020 to April 2020**.** Wang et al (2022) conducted a study in China focusing on long-term air quality predictions during the Beijing Winter Olympics. They employed a T-mode objective circulation classification method and the EC-Earth climate prediction model. The hybridization of climate prediction data with weather circulation and meteorological elements strengthened their predictive capabilities. Their study's strength lies in the innovative integration of climate and meteorological data for more accurate predictions. The primary data source for their study comprised secondary data spanning from 2015 to 2021.

In Arroyo Herrero, Suarez, & Lozano (2019), the team implemented an IoT-based air quality measurement system, incorporating cloud computing and AI for data visualization, along with machine learning using Multilayer Perceptron. Their focus on a low-cost, low-size, and low-power consumption method highlights the practicality of their approach. The study's strength lies in its ability to balance efficiency and cost-effectiveness in air quality monitoring. Almalawi, et al., (2022) worked on the development of an air qualification index, employing Linear Regression, Support Vector Regression, and Gradient Boosted Decision Tree (GBDT) Ensembles. Their ensemble methodology optimized the model's performance and addressed bias issues. The study's contribution lies in the creation of a robust air quality index, combining the strengths of multiple regression techniques. Wang, et al., (2022) contributed to air pollution prediction using a Graph Attention Network (GAT) and Gated Recurrent Unit (GRU). Their ensemble approach combined concentration data and meteorological data, providing a holistic understanding of the factors influencing air pollution. The study's integration of advanced neural network architectures showcases a sophisticated approach to air quality prediction. In Popovic et al. (2022) (Popović, Radovanovic, Vajs, Drajic, & Gligorić, 2022), authors utilized fog computing for air quality monitoring in urban areas. Their approach involved sensor-side processing encapsulated in the form of microservices, allowing for efficient data handling. The study's emphasis on the discovery of erroneous measurements and sensor faults enhances the reliability of air quality monitoring systems. Mengara park, Jang, & Yoo (2022) focused on air quality forecasting using a convolutional BiLSTM autoencoder. Their distributed framework, named data parallelism, ensures efficient processing of air quality data. The study's use of advanced neural network architectures highlights its commitment to accurate and timely air quality forecasting.

In the work of Mi, Sun, Zhang, & Liu (2022), authors worked on air quality forecasting using a Hierarchical Graph Neural Network, incorporating historical readings and various urban contextual factors. The study utilized secondary data from a public repository, contributing valuable insights into the long-term dynamics and contextual influences on air quality. Liu et al. (2022) on the other hand focused on IoT with cloud computing for air pollution monitoring, employing a low-cost multimodality sensor-based system. The study's real-time primary data collection ensures responsiveness in monitoring air pollution. The emphasis on cost-effectiveness and real-time data acquisition makes this approach practical for widespread deployment in monitoring systems.

Main gap identified in literature is the non-inclusion of a robust data analysis functionality that would provide an instant statistical summary of the selected region. This will give an instant view of the current state of a location of interest, as soon as the application is launched. This study therefore employs the functionality of an exploratory data analysis tool for the purpose.

METHODOLOGY

The methodological approach employed in this chapter is discussed in this section. The phases of the conceptual framework include air quality secondary data acquisition from the Ministry of Environment (Ogun State), and the qualitative survey primary data from acquired from residents of Ogun state. The exploratory data analysis of the air quality data is conducted with python programming, and consequently, topic modelling of the survey data is implemented. The air quality visualization system is then designed and implemented for the use of critical stakeholders at the ministry of environment in Ogun state.

Dual Data Acquisition

The first data is the air quality data acquired through the aid of IoT devices from the Ogun state environs. The data is made up of 23 attributes and 263 instances between October and

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024
www.iprjb.org

December 2023. The attributes of the data are described in Table 1. A survey is also conducted between the same period where Google digital form is designed to acquire opinions of Ogun state residents on the quality of air and their perspectives on some germane issues. The form link is sent out on WhatsApp groups and a total of 59 respondents were gotten on the four open-ended questions listed below:

Question 1: What are your thoughts on the current state of air quality in the Ogun metropolis?

Question 2: In your opinion, what are the major sources of air pollution in Ogun?

Question 3: How well do you think the government are addressing air quality issues in Ogun?

Question 4: What role can individuals play in contributing to better air quality in Ogun?

The qualitative primary data acquired from the survey, and the air quality data are analyzed in the subsequent sections of this chapter, and the experimental results are explained.

Table 1: Description of Attributes Contained in the Ogun Air Quality Dataset

S/N	Attributes	Description
1	Start Date	Data collection starting date
2	End Date	Data collection ending date
3	PM2.5 (ug/m3)	2.5 Particulate Matter concentration
4	PM10 (ug/m3)	10 Particulate Matter concentration
5	NO (ug/m3)	Nitric Oxide concentration
6	NO2 (ug/m3)	Nitrogen Dioxide concentration
7	NOx (ppb)	Nitrogen Oxides concentration
8	NH3 (ug/m3)	Ammonia concentration
9	SO2 (ug/m3)	Sulfur Dioxide concentration
10	CO (mg/m3)	Carbon Monoxide concentration
11	Ozone (ug/m3)	Ozone concentration
12	Benzene (ug/m3)	Benzene concentration
13	Toluene (ug/m3)	Toluene concentration
14	Temp (degree C)	Temperature
15	RH (%)	Relative Humidity
16	WS (m/s)	Wind Speed
17	WD (deg)	Wind Direction
18	SR (W/mt2)	Solar Radiation
19	BP (mmHg)	Barometric Pressure
20	VWS (m/s)	Vertical Wind Speed
21	Xylene (ug/m3)	Xylene concentration
22	RF (mm)	Rainfall
23	AT (degree C)	Apparent Temperature

Exploratory Data Analysis

The Exploratory Data Analysis (EDA) is a crucial step in the data analysis process that involves summarizing, visualizing, and understanding the main characteristics of a dataset (Olaleye T. O., et al., 2023). The summary statistics (mean, max, min, 25%, 50%, 75%) provide valuable insights into the distribution and central tendency of the data. These summary statistics collectively provide a comprehensive view of the data's central tendency, variability, and distribution. EDA involves examining the air quality data parameters to identify patterns,

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024
www.iprjb.org

outliers, and potential relationships in the dataset, guiding further analysis and decision-making processes. The heat map of the correlation coefficient visualization plot aid in the exploration of the complex dataset. The various EDA tasks employed on the air quality data are discussed below:

1. Mean (Average)

The mean represents the average value of the dataset. It provides an overall measure of central tendency.

$$\text{Mean}=\frac{Sum\ of\ all\ values}{Number\ of\ values} \qquad (1)$$

2. Min (Minimum)

The minimum value is the smallest observation in the dataset, indicating the lowest point in the distribution.

Min=Smallest value in the dataset (2)

3. 25%, 50%, 75% (Percentiles)

Percentiles provide insights into the distribution of data. The median (50th percentile) is the middle value, and quartiles (Q1 and Q3) divide the data into four equal parts.

$$\text{Q1 = Value at } \frac{1}{4}(n+1) position\ in\ ordered\ dataset \qquad (3)$$

$$\text{Q2 = Value at } \frac{1}{2}(n+1) position\ in\ ordered\ dataset \qquad (4)$$

$$\text{Q3 = Value at } \frac{3}{4}(n+1) position\ in\ ordered\ dataset \qquad (5)$$

For Q1, value at the position that separates the lowest 25% of the data is gotten.

For Q2 (Median), value at the position that separates the lowest 50% of the data is gotten.

For Q3, value at the position that separates the lowest 75% of the data is gotten.

4. Max (Maximum)

The maximum value is the largest observation in the dataset, indicating the highest point in the distribution.

Max=Larg

est value in the dataset (6)

5. Standard Deviation:

The standard deviation measures the amount of variability or dispersion in a set of values. A low standard deviation indicates that the values are close to the mean, while a high standard deviation indicates greater variability.

$$\text{Standard deviation} = \sqrt{\frac{\Sigma(Xi\ X')^2}{N}} \qquad (7)$$

The correlation coefficient is a statistical measure that quantifies the degree to which two variables are related in a linear manner. It assesses the strength and direction of a linear relationship between two continuous variables. The most commonly used correlation coefficient is Pearson's correlation coefficient (r).

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024
www.iprjb.org

6. Pearson's Correlation Coefficient (r):

$$r = \frac{\sum(Xi\,X')(Yi\,Y')}{\sqrt{\sum(Xi\,X')^2\sum(Yi\,Y')^2}} \qquad (8)$$

- Xi and Yi are individual data points in the two variables.
- X' and Y' are the means of the two variables, respectively.

The numerator computes the covariance of the two variables, measuring how they vary together. The denominator is the product of the standard deviations of the two variables, normalizing the covariance.

- r ranges from -1 to 1:
  i. r=1 indicates a perfect positive linear relationship.
  ii. r=−1 indicates a perfect negative linear relationship.
  iii. r=0 indicates no linear relationship.
- The sign of r indicates the direction of the relationship:
  i. Positive r implies a positive association (as one variable increases, the other tends to increase).
  ii. Negative r implies a negative association (as one variable increases, the other tends to decrease).
- $|r| < 0.3$: Weak correlation
- $0.3 \leq |r| < 0.7$: Moderate correlation
- $|r| \geq 0.7$: Strong correlation

Topic Modelling of the Survey Data

Topic modeling is a technique employed to automatically unveil latent thematic structures embedded within extensive collections of documents or texts (Asmussen & Møller, 2019). In the specific context of scrutinizing users' perspectives on the quality of air they breathe in their environment, topic modeling serves to distill meaningful topics or themes from user-generated content, including surveys or reviews. This methodology facilitates the extraction of qualitative insights from textual data, enhancing the depth of understanding in this study. An established approach in topic modeling is Latent Dirichlet Allocation (LDA), which posits that documents consist of mixtures of topics and, in turn, topics are comprised of mixtures of words. Essentially, LDA endeavors to reverse engineer the document creation process, contributing to a nuanced analysis of user sentiments and opinions (Chauhan & Shah, 2021). LDA is used in this study to analyze the responses of Ogun state residents based on the following questions posed to them:

To accomplish this, LDA adheres to a systematic methodology:

i. Preprocessing: Text data undergoes preprocessing through the removal of stop words (common words like "and," "the," etc.), punctuation, and stemming (reducing words to their root form).
ii. Tokenization: Sentences and documents are disassembled into individual words or tokens.
iii. Document-Term Matrix: LDA represents the data in a matrix format, where rows correspond to documents and columns to terms (words). The matrix values typically denote word frequencies within documents.
iv. Selecting the Number of Topics: Researchers must specify the number of topics to extract, often requiring a subjective decision.

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024
www.iprjb.org

v. Running LDA: The LDA algorithm iteratively assigns words in documents to topics and topics to documents, optimizing the explanation of the corpus structure.
vi. Keyword Extraction: For each topic, LDA assigns a probability distribution over words. The words with the highest probabilities within a topic are deemed keywords for that topic. vii.
vii. Topic Labeling: Researchers scrutinize the extracted keywords and allocate meaningful labels to each topic based on their understanding of the keywords.
viii. Interpretation: With labeled topics, researchers qualitatively analyze the content within each topic, gaining insights into users' opinions about the air quality monitoring system.

The assigned weights to keywords within each topic indicate the probability of a word being associated with that specific topic. Topics are essentially groupings of words sharing similar weight distributions, forming coherent and interpretable thematic clusters. This framework enables the detection of recurring themes in user opinions, such as apprehensions regarding pollution sources, contentment with system features, or recommendations for enhancement.

Implementation of Air Quality Monitoring System

Python programming is used to implement a graphical user interphase that reads air quality data from its data base for the development of a monitoring system which employs visualization techniques. Users could select the particular year and the choice air attributes under investigation. The system will then plot three different data visualizations for the chosen attributes and timeline. The summary statistics of the chosen information is also displayed on the GUI interphase to aid informed decisions. The application is designed on the Jupyter environment on the Anaconda Navigator integrated development environment.

FINDINGS

Analysis of the qualitative data is done, which is a topic modelling of Ogun residents' perception of the quality of air and other associated matters. The four open-ended questions posed to them triggered textual opinions which were preprocessed and analyzed by the LDA, in order to infer actionable insights from the five topics generated by LDA on each of the four categories of responses. Table 2, Table 3, Table 4 and Table 5 presents the result of the LDA analysis on the 59 responses to each of the four questions respectively. Five topics (0 to 4) are generated for each of the 59 responses to each of the four questions, while the identified keywords for each of the topics are presented. The weights assigned to each of the keywords are presented and the likely logical deductions from each topic is also presented on the table.

Table 2 presents the outcome of the topic modelling on responses to question 1 which asks "*What are your thoughts on the current state of air quality in the Ogun metropolis?*" The results from the topic modeling analysis offer insightful deductions regarding public sentiments on the current state of air quality in the Ogun metropolis. The first identified topic suggests a negative sentiment towards social amenities and the overall state of the region. This indicate concerns or dissatisfaction with the infrastructure and public services related to air quality management. Conversely, the second topic highlights positive emotions towards air quality, with an acknowledgment of existing pollution concerns. This dual perspective underscores the nature of public opinions, recognizing both the positive aspects and ongoing challenges. The third topic emphasizes negative sentiments related to poor air quality and its impact, explicitly connecting it to pollution issues. This topic suggests a heightened awareness and concern among respondents regarding the adverse effects of pollution on air quality and, consequently, public health. The fourth topic underscores the need for improvement and regulation in response to perceived poor conditions. This aligns with a call for proactive measures and regulatory interventions to address the identified air quality challenges. Lastly, the fifth topic captures opinions on moderate air quality and its association with cars. This suggests that respondents may attribute a moderate level of air quality to vehicular emissions, pointing towards a recognition of specific sources of pollution.

Table 3 presents the outcome of the topic modelling on responses to question 2 which asks "*In your opinion, what are the major sources of air pollution in Ogun?*". The results of the topic modeling analysis shed light on the public's perceptions regarding the major sources of air pollution in Ogun, as expressed in response to the second question. The first identified topic reflects concerns about various pollution sources, including generators, waste, cars, vehicles, and exhaust emissions. This suggests a broad awareness among respondents about the diverse contributors to air pollution, ranging from individual activities, such as generator usage, to broader issues like vehicular emissions and waste management practices. The second topic delves into discussions surrounding industrial activities, burning, vehicle emissions, and waste as significant contributors to pollution. This highlights a recognition of the role that industrial processes and combustion-related activities play in deteriorating air quality. The emphasis on both industrial and vehicular sources indicates a multifaceted understanding of pollution origins within the community. The third topic centers on concerns related to health, air quality, carbon monoxide, and their impact on the Ogun atmosphere. This topic underscores the link between air pollution and public health, with a specific focus on the potential dangers associated with carbon monoxide and the overall quality of the air in Ogun. It suggests a heightened awareness among respondents about the consequences of air pollution on their well-being. The fourth topic revolves around discussions on pollution from generators, industrial areas, and vehicles in special areas. This topic suggests that respondents may perceive certain locations or zones within Ogun as particularly vulnerable to pollution, warranting specific attention and intervention measures. The fifth topic brings attention to talks surrounding industrial activities, refuse, cars, and vehicle emissions contributing to pollution. This reinforces the recurring theme of industrial processes and vehicular activities as key sources of air pollution, complemented by concerns about waste management practices.

Table 4 presents the outcome of the topic modelling on responses to question 3 which asks "*How well do you think the government are addressing air quality issues in Ogun?*" The outcomes of the topic modeling analysis for the question assessing public perceptions of the government's efforts in addressing air quality issues in Ogun reveal distinct themes. The first identified topic centers on discussions related to preventing air pollution, emphasizing the role of work and regulations. This suggests that respondents perceive proactive measures, such as regulatory frameworks and workplace practices, as crucial in preventing the onset or exacerbation of air pollution. The second topic highlights a focus on waste management, government policies, and issues specific to metropolitan areas.

This implies that respondents are attuned to the significance of proper waste disposal and the need for tailored policies in densely populated urban centers. The attention to metropolitan areas suggests a recognition that urban environments may present unique challenges in air quality management. The third topic revolves around conversations regarding government involvement, air quality issues, and guidelines. This indicates that respondents are engaged in discussions regarding the active role of the government in addressing air quality concerns and the establishment of guidelines to regulate and monitor air quality. The fourth topic delves into talks emphasizing the need for effective pollution control, better dispositions, and expressing concerns. This suggests that respondents are looking for tangible and impactful actions from the government in terms of pollution control measures and an overall improvement in their approach to addressing air quality concerns. The final topic involves deliberations on the need for government improvements and ensuring the well-being of the masses. This broader theme suggests a holistic expectation from the government, encompassing not only specific air quality measures but also a commitment to the general welfare and health of the public.

Table 5 presents the outcome of the topic modelling on responses to question 4 which asks "*What role can individuals play in contributing to better air quality in Ogun?*". The outcomes of the topic modeling analysis for the question exploring the role of individuals in contributing to better air quality in Ogun reveal several key themes. The first identified topic underscores the importance of individuals ceasing waste burning, emphasizing the role of the government and raising concerns about waste administration issues. This implies that respondents recognize the detrimental effects of waste burning on air quality and view government intervention and effective waste administration as crucial elements in addressing this issue. The second topic centers on discussions related to environmental policies, cleanliness, and discipline, highlighting their implications for better air quality.

This suggests that respondents acknowledge the impact of individual behaviors, adherence to environmental policies, and maintaining cleanliness and discipline as essential components in the collective effort toward improving air quality. The third topic places emphasis on individual accountability for ensuring good air quality and expresses worries about poor air quality. This theme suggests a recognition among respondents that each individual plays a role in maintaining air quality standards and underscores concerns about the consequences of poor air quality on public health and well-being. The fourth topic revolves around talks about green initiatives, tree planting, and the impact of waste burning on communities. This indicates an awareness among respondents of the positive contribution of environmentally friendly practices, such as tree planting and green initiatives, in mitigating air quality issues, particularly those arising from activities like waste burning. The fifth topic speaks to car pollution, avoiding burning practices, and responsible car usage. This theme highlights the role of individuals in reducing vehicular pollution by avoiding harmful burning practices and adopting responsible car usage habits.

Table 2: Modelling of Users' Responses to Q1

Topic	Identified Keywords	Weights of Keywords	Likely Consensus
0	'bad'	0.133	Negative expressions towards 'social' 'amenities' and 'state'
	'social'	0.056	
	'amenity'	0.056	
	'area'	0.031	
	'state'	0.031	
1	'good'	0.153	Positive emotions on air quality, notwithstanding some pollution concerns
	'polluted'	0.123	
	'highly'	0.065	
	'Better'	0.035	
	'managed'	0.035	
2	'air'	0.107	Concerns on 'poor' 'air' 'quality' and its 'impact' due to 'pollution'
	'poor'	0.092	
	'pollution'	0.063	
	'impact'	0.048	
	'quality'	0.048	
3	'need()',	0.153	Emphasis on the need for 'improvement' and 'regulation' due to 'poor' conditions.
	'improvement'	0.124	
	'poor'	0.035	
	'serious'	0.035	
	'regulation'	0.035	
4	'okay'	0.049	Opinions on 'moderate' 'air' quality and its relation to 'cars'.
	'air'	0.049	
	'car'	0.049	
	's'	0.049	
	'moderate'	0.027	

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024

Table 3: Modelling of Users' Responses to Q2

Topic	Identified Keywords	Weights of Keywords	Likely Consensus
0	'generator'	0.049	Concerns on pollution sources like 'generators', 'waste', 'cars', 'vehicles', and 'exhaust' emissions.
	'waste'	0.048	
	'car'	0.039	
	'vehicle'	0.039	
	'exhaust'	0.030	
1	'burning'	0.051	Discussion on 'industrial' activities, 'burning', 'vehicle' 'emissions', and 'waste' contributing to pollution.
	'industry'	0.051	
	'vehicle'	0.028	
	'waste'	0.028	
	'emission'	0.028	
2	'health'	0.027	Concerns related to 'health', 'air' quality, carbon 'monoxide', and their impact on Ogun atmosphere
	'air'	0.027	
	'carbon'	0.027	
	'monoxide'	0.027	
	'ogun'	0.027	
3	'pollution'	0.039	Discussions on 'pollution' from 'generators', 'industrial' 'areas', and 'vehicles' in special 'areas'.
	'generator'	0.039	
	'vehicle'	0.039	
	'industrial'	0.027	
	'area'	0.027	
4	'industry'	0.061	Talks around 'industrial' activities, 'refuse', 'cars', and 'vehicle' 'emissions' contributing to pollution.
	'refuse'	0.059	
	'car'	0.047	
	'emission'	0.047	
	'vehicle'	0.047	

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024
www.iprjb.org

Table 4: Modelling of Users' Responses to Q3

Topic	Identified Topics	Weights of Keywords	Likely Consensus
0	'preventing'	0.024	Discussions bothering on 'preventing' 'air' pollution, 'work', 'regulations'
	'air'	0.024	
	'work'	0.024	
	'ogun'	0.024	
	'regulation'	0.024	
1	'waste'	0.056	Focus on 'waste' 'management', 'government' policies, and issues in 'metropolitan' areas.
	'management'	0.042	
	'much'	0.029	
	'metropolitan'	0.029	
	'government'	0.029	
2	'government'	0.039	Conversations bothering on 'government' involvement, 'air' 'quality' 'issues', and guidelines
	'issue'	0.020	
	'air'	0.020	
	'quality'	0.020	
	'ogun'	0.020	
3	'effective'	0.027	Talks on the need for 'effective' 'pollution' 'control', 'better' dispositions, and concerns.
	'better'	0.027	
	'control'	0.027	
	'n'	0.027	
	'pollution'	0.027	
4	'need'	0.057	Deliberations on the 'need' for 'government' 'improvements' and 'ensuring' the 'well-being' of the masses.
	'well'	0.043	
	'government'	0.030	
	'improve'	0.030	
	'sure'	0.030	

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024
www.iprjb.org

Table 5: Modelling of Users' Responses to Q4

Topic	Identified Topics	Weights of Keywords	Likely Consensus
0	'waste()',	0.046	Highlighting 'stopping' 'waste' 'burning', 'government's role, and waste administration issues.
	'refuse(0.046)',	0.046	
	'stop(0.046)',	0.046	
	'government()',	0.032	
	'burning(0.032)'	0.032	
1	'policy()',	0.044	Discussion centers on 'environmental' 'policies', 'cleanliness', 'discipline', and their implications
	'clean()',	0.030	
	'discipline(0.030)',	0.030	
	'environmental(0.030)',	0.030	
	"s(0.016)'	0.016	
2	'air	0.058	Emphasis on 'individual' accountability for 'good' 'air' 'quality' and worries about 'poor' air quality.
	'quality	0.058	
	'individual	0.044	
	'good	0.044	
	'poor	0.030	
3	'green	0.030	Talks about 'green' initiatives, 'tree' 'planting', and 'waste' 'burning's impact on communities
	'tree	0.030	
	'burning	0.030	
	'planting	0.030	
	'waste	0.018	
4	'car	0.044	Speaks to 'car' 'pollution', 'avoiding' 'burning' 'practices', and responsible 'car usage'
	'pollution	0.044	
	'burning	0.030	
	'avoid	0.030	
	'usage	0.030	

The result of the summary statistics likewise speaks volume. The provided summary statistics in Table 6 and Table 7 (count, mean, std, min, 25%, 50%, 75%, max) offer valuable insights into the air quality data for different pollutants in the study. These statistics provide an initial overview of the central tendency, spread, and range of the air quality attributes.

- PM2.5 (Particulate Matter 2.5)

The mean PM2.5 concentration is approximately 12.65 µg/m³. The data has a relatively low standard deviation (6.80), indicating that the values are not very dispersed. The minimum and maximum concentrations are 1.25 and 38 µg/m³, respectively.

- PM10 (Particulate Matter 10):

The mean PM10 concentration is around 40.65 µg/m³. The standard deviation is 15.29, suggesting moderate variability in the data. The minimum and maximum concentrations are 6.25 and 86 µg/m³, respectively.

- NO (Nitric Oxide):

The mean NO concentration is approximately 13.56 µg/m³. The data has a relatively high standard deviation (21.47), indicating considerable variability. The minimum and maximum concentrations are 1.45 and 174.73 µg/m³, respectively.

- NO2 (Nitrogen Dioxide):

The mean NO2 concentration is around 38.20 µg/m³. The standard deviation is 8.46, suggesting moderate variability. The minimum and maximum concentrations are 16.92 and 59.82 µg/m³, respectively.

- Nox (Nitrogen Oxides):

The mean Nox concentration is approximately 28.69 µg/m³. The standard deviation is 16.96, indicating moderate variability. The minimum and maximum concentrations are 9.80 and 146.33 µg/m³, respectively.

- NH3 (Ammonia):

The mean NH3 concentration is approximately 6.62 µg/m³. The standard deviation is 1.34, suggesting relatively low variability. The minimum and maximum concentrations are 4.43 and 10.97 µg/m³, respectively.

- SO2 (Sulfur Dioxide):

The mean SO2 concentration is around 5.25 µg/m³. The standard deviation is 1.00, indicating relatively low variability. The minimum and maximum concentrations are 1.78 and 8.15 µg/m³, respectively.

- CO (Carbon Monoxide):

The mean CO concentration is approximately 0.75 mg/m³. The standard deviation is 2.25, indicating considerable variability. The minimum and maximum concentrations are 0.11 and 20.72 mg/m³, respectively.

- Ozone:

The mean Ozone concentration is around 14.28 µg/m³. The standard deviation is 2.66, suggesting moderate variability. The minimum and maximum concentrations are 2.25 and 22.80 µg/m³, respectively.

- Benzene:

The mean Benzene concentration is approximately 0.56 µg/m³. The data has a moderate standard deviation (0.28), indicating some variability. The minimum and maximum concentrations are 0.10 and 1.63 µg/m³, respectively.

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024
www.iprjb.org

Table 6: Summary Statistics of the First Ten Air Quality Attributes

	PM2.5	PM10	NO	NO2	Nox	NH3	SO2	CO	Ozone	Benzene
Count	262.000000	262.000000	262.000000	262.000000	262.000000	256.000000	251.000000	262.000000	261.000000	262.000000
Mean	12.650458	40.653969	13.562939	38.203626	28.694695	6.619492	5.245259	0.749809	14.283487	0.557443
Std	6.796394	15.291250	21.470527	8.460976	16.963210	1.342072	1.003785	2.249692	2.658160	0.278573
Min	1.250000	6.250000	1.450000	16.920000	9.800000	4.430000	1.780000	0.110000	2.250000	0.100000
25%	7.542500	29.062500	5.662500	31.947500	20.055000	5.615000	4.730000	0.220000	13.780000	0.300000
50%	12.125000	40.750000	9.830000	39.110000	26.500000	6.410000	5.250000	0.320000	14.520000	0.610000
75%	17.000000	50.937500	14.422500	43.915000	32.330000	7.300000	5.930000	0.430000	15.170000	0.700000
Max	38.000000	86.000000	174.730000	59.820000	146.330000	10.970000	8.150000	20.720000	22.800000	1.630000

- Toluene:

Mean concentration of Toluene is approximately 4.48 µg/m³. Moderate standard deviation (1.54) indicates some variability. Minimum and maximum concentrations are 0.67 and 8.47 µg/m³, respectively.

- Temp (Temperature):

Mean temperature is around 32.65 °C. Low standard deviation (0.74) suggests low variability. Temperature ranges from 31.73 to 37.63 °C.

- RH (Relative Humidity):

Mean relative humidity is approximately 68.11%. Standard deviation (6.25) suggests some variability. Relative humidity ranges from 53.75 to 83.33%.

- WS (Wind Speed):

Mean wind speed is approximately 2.06 m/s. Moderate standard deviation (0.51) indicates some variability. Wind speed ranges from 0.83 to 3.3 m/s.

- WD (Wind Direction):

Mean wind direction is approximately 214.42 degrees. High standard deviation (48.74) indicates considerable variability. Wind direction ranges from 11.33 to 332.25 degrees.

SR (Solar Radiation):

- Mean solar radiation is approximately 106.85 W/m².

High standard deviation (152.94) suggests considerable variability. Solar radiation ranges from 6 to 724 W/m².

- BP (Barometric Pressure):

The summary statistics are not provided in the given data (NaN).

- VWS (Vertical Wind Speed):

Mean vertical wind speed is approximately 0.08 m/s. Standard deviation (0.26) indicates some variability. Vertical wind speed ranges from -0.1 to 1.28 m/s.

- Xylene:

Mean concentration of Xylene is approximately 0.14 µg/m³. Low standard deviation (0.13) suggests low variability. Xylene ranges from 0.1 to 0.8 µg/m³.

- RF (Rainfall):

Mean rainfall is approximately 0.20 mm. Moderate standard deviation (0.31) indicates some variability. Rainfall ranges from 0 to 1.62 mm.

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024
www.iprjb.org

Table 7: Summary Statistics of the Next Ten Air Quality Attributes

	Toluene	Temp	RH	WS	WD	SR	BP	VWS	Xylene	RF
Mean	4.475725	32.646298	68.110725	2.058244	214.421908	106.845382	NaN	0.077977	0.141183	0.200611
Std	1.541343	0.744416	6.247122	0.512970	48.742280	152.939344	NaN	0.260009	0.127185	0.313173
Min	0.670000	31.730000	53.750000	0.830000	11.330000	6.000000	NaN	-0.100000	0.100000	0.000000
25%	3.300000	32.135000	63.750000	1.722500	188.000000	7.000000	NaN	-0.100000	0.100000	0.000000
50%	4.300000	32.570000	68.500000	2.040000	217.000000	13.500000	NaN	-0.100000	0.100000	0.000000
75%	5.550000	32.950000	72.627500	2.465000	241.000000	166.625000	NaN	0.200000	0.100000	0.250000
Max	8.470000	37.630000	83.330000	3.300000	332.250000	724.000000	NaN	1.280000	0.800000	1.620000

The heat map in Figure 1 and Figure 2 visualizes the correlation between the twenty air quality attributes contained in the dataset. Figure 1 visualizes for the first ten attributes while Figure 2 depicts that of the next ten attributes. For the complete dataset (df), notable correlations include a strong positive correlation between PM2.5 and PM10 (0.69), indicating a substantial association between fine and coarse particulate matter. Additionally, there is a noteworthy positive correlation between NO and Nox (0.97), suggesting a high degree of correlation between nitric oxide and nitrogen oxide levels. However, negative correlations are observed, such as the substantial negative correlation between RH (relative humidity) and PM2.5 (-0.46), implying an inverse relationship between humidity and fine particulate matter levels. Figure 1 shows the correlation coefficient between NO and Nox stands out prominently at 0.97, signifying an exceptionally strong positive correlation between nitric oxide (NO) and nitrogen oxides (Nox). This observation is consistent with the inherent chemical processes where nitric oxide is a precursor to nitrogen oxides. The high correlation suggests a synchronized increase or decrease in the concentrations of these pollutants, indicating a shared source or common atmospheric transformation. Another noteworthy correlation involves the pair of Benzene and NO2. The correlation coefficient between Benzene and NO2 is substantial at 0.74, denoting a strong positive correlation between these two pollutants. This correlation is of particular concern as it suggests a significant association between the presence of benzene, a known hazardous air pollutant, and nitrogen dioxide (NO2), a key component of traffic-related emissions and industrial activities. Such correlations emphasize the need for targeted monitoring and regulatory measures to address potential co-occurring sources of these pollutants, ensuring comprehensive strategies for air quality management and public health protection.

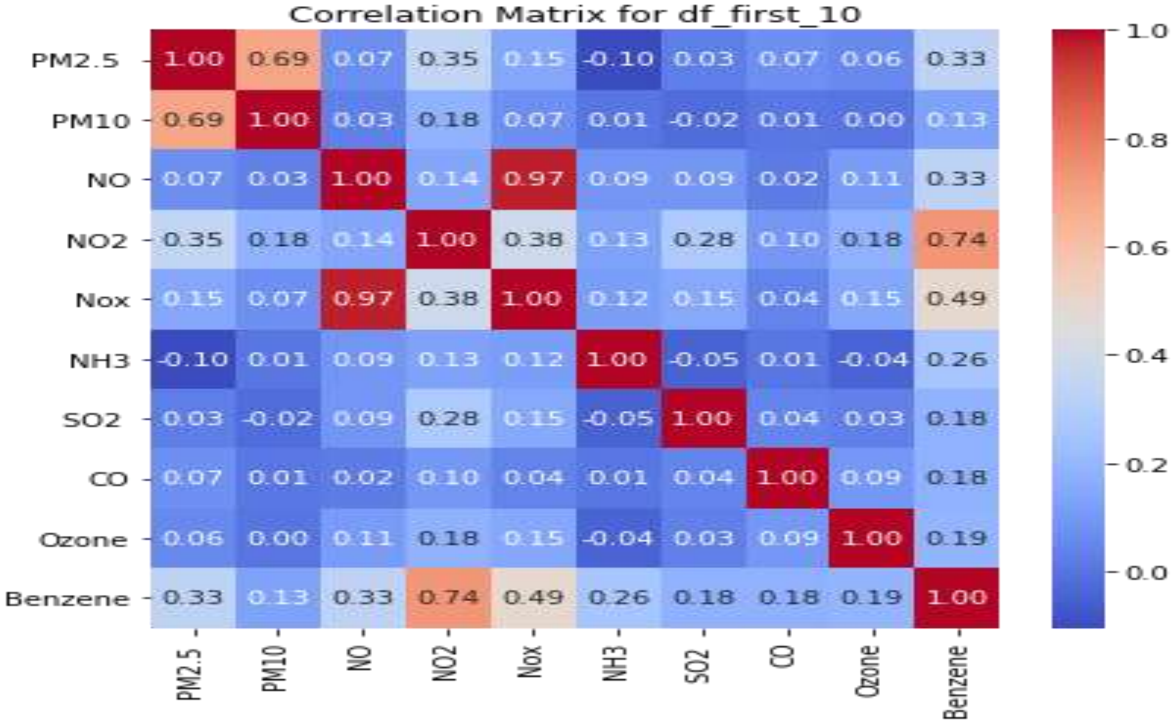

Figure 1: Heat Map of the Correlation Coefficient Plot for the First Ten Attributes

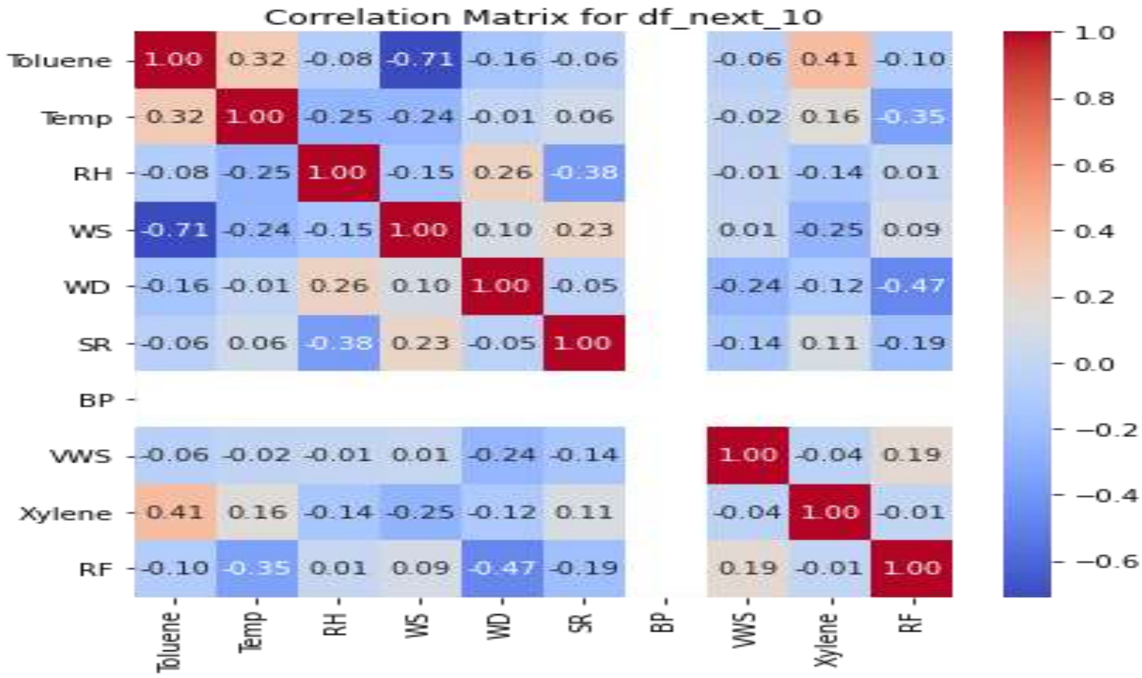

Figure 2: Heat Map of the Correlation Coefficient Plot for the Next Ten Attributes

Figure 3 depicts the user interface of the Ogun State air monitoring system post-implementation, comprising four distinct sections: the Title Bar, Application Choice Area, Plot Area, and Summary Statistics Pane. The Application Choice Area features four buttons,

International Journal of Technology and Systems
ISSN 2518-881X (Online)
Vol.9, Issue 2, No.5, pp 82 – 107, 2024
www.iprjb.org

including three facilitating the selection of the desired year, air attribute, and plot type. Upon making these selections, the Plot Data button in green initiates the retrieval of chosen data attributes from the database and generates the specified visualization as per the user's preferences. Concurrently, the summary statistics for the chosen air attribute within the selected year are computed and displayed on the pane beside the plot area. Figure 4 showcases the outcome of interacting with the application. The plot area visually represents data trends, patterns, dispersion, or time series of data points based on user choices, while the summary statistics offer a succinct overview of crucial information about the selected dataset. This includes measures such as mean, standard deviation, minimum, maximum, and quartiles for the chosen attribute, like PM2.5 concentration, within the specified year. These statistics provide insights into the central concentration, variability, and potential outliers in the air quality data, aiding users in comprehending the overall distribution and characteristics of the selected attribute's values for their chosen timeframe. Users have the flexibility to switch to another plot, such as a scatter plot, while maintaining the chosen year and air attribute. Figure 5 illustrates the automatic update of the plot from a box plot in Figure 4 to a scatter plot. If a different year (e.g., 2016), air attribute, or plot type (Time series) is selected (Figure 5), the summary statistics will dynamically adjust to reflect the new data attributes retrieved from the database instantly.

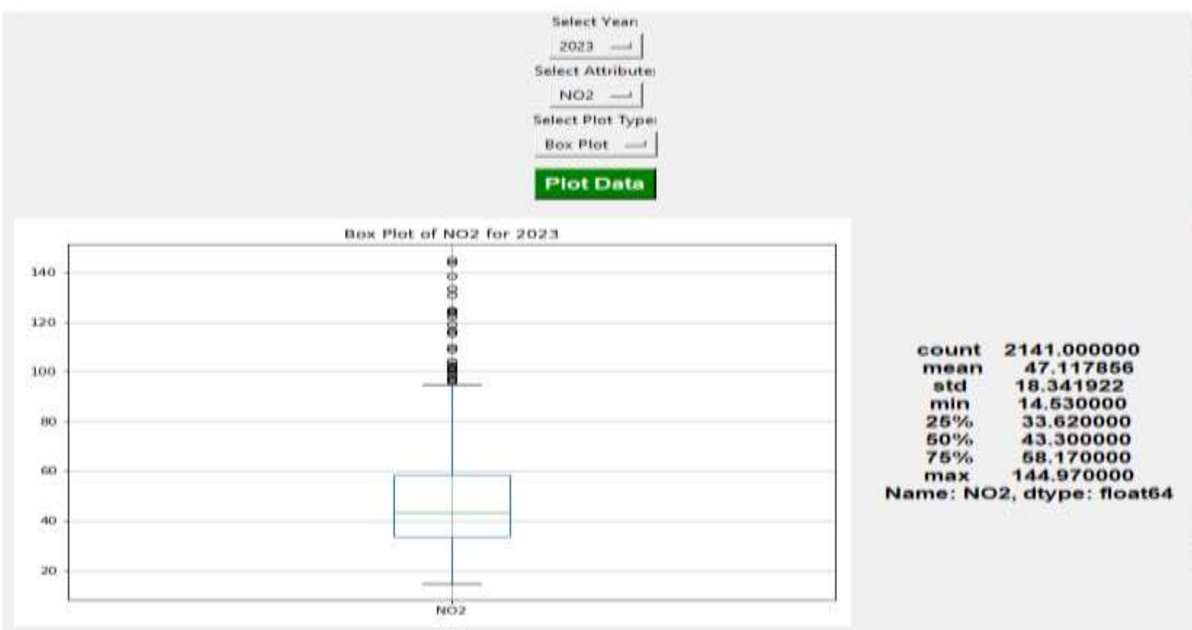

Figure 3: Interphase of Box Plot on the Monitoring System

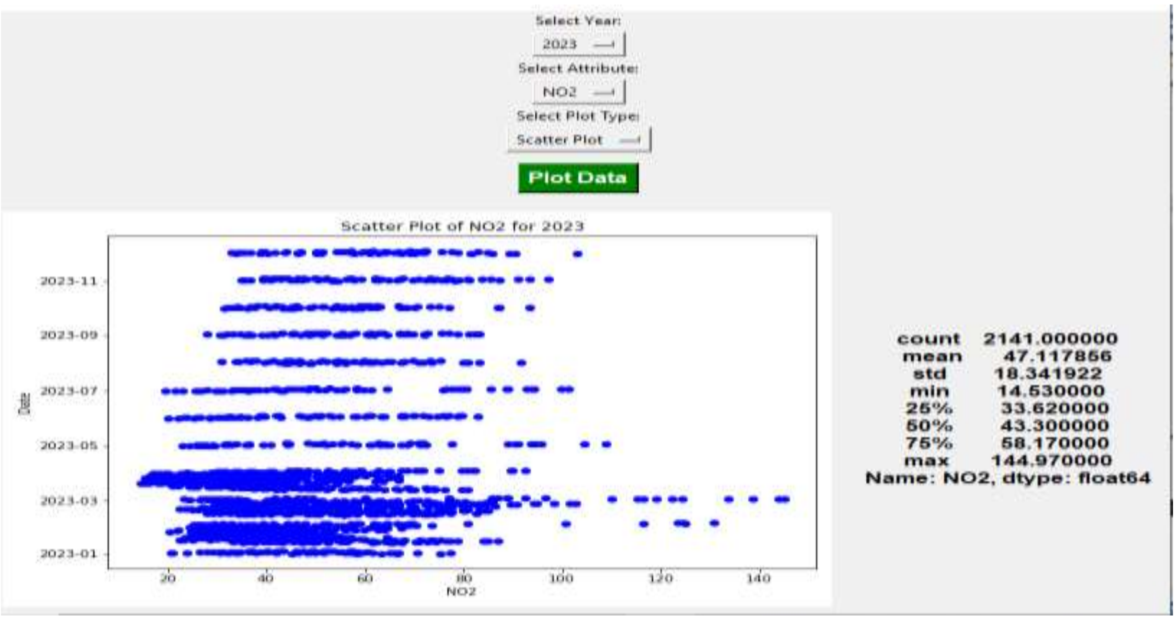

Figure 4: Interphase of Scattered Plot on the Monitoring System

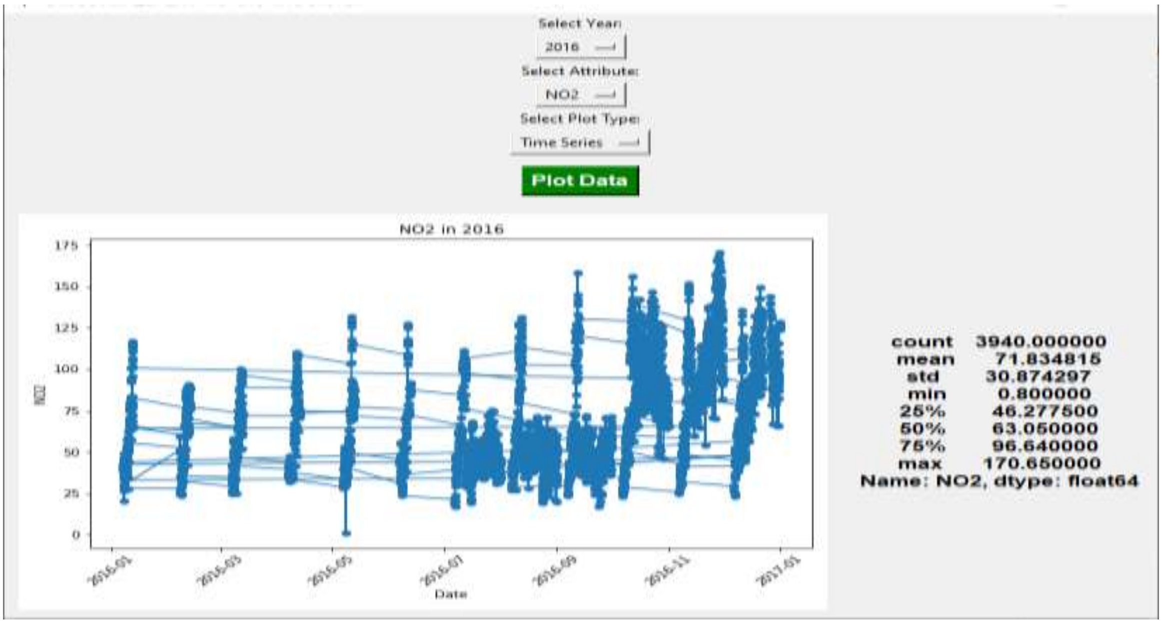

Figure 5: Interphase of Time Series Plot on the Monitoring System

CONCLUSION AND RECOMMENDATION

In conclusion, this study presents a multifaceted exploration of air quality management, blending traditional statistical analyses with innovative data-driven methodologies and technological solutions. The insights derived from Exploratory Data Analysis (EDA) provide a foundational understanding of the statistical parameters, correlations, and variations within historical air quality datasets. These findings lay the groundwork for informed decision-making regarding air quality improvement strategies. The integration of Topic Modeling techniques unveils latent themes within qualitative data, capturing nuanced public sentiments and opinions related to air quality. The resulting topics shed light on diverse perspectives concerning pollution sources, government interventions, and individual responsibilities. Such qualitative insights enrich the overall understanding of air quality challenges and provide a basis for targeted policy recommendations. The development of an advanced air quality management application marks a pivotal contribution, bridging the gap between data analysis and actionable outcomes. The application, designed for user-friendly access, empowers both the public and policymakers to interact with real-time air quality data, visualize trends, and explore thematic patterns derived from Topic Modeling. This technological solution fosters a collaborative approach, encouraging collective efforts towards sustainable air quality management. By amalgamating statistical rigor, qualitative insights, and cutting-edge technology, this study not only deepens our comprehension of air quality dynamics but also presents practical tools for stakeholders to actively engage in the improvement of air quality. As we navigate the challenges of increasing urbanization and industrialization, the holistic approach presented in this study underscores the importance of collaborative, data-informed strategies for achieving and maintaining clean and healthy air for all.

Based on the comprehensive findings of this study, a key recommendation for future studies is the implementation of targeted regulatory measures to address pollution sources identified through the Topic Modeling analysis. Specifically, industries contributing significantly to air pollution, as highlighted by public sentiments, should be subject to stricter emission controls and monitoring. Furthermore, the integration of the developed air quality management application into public health initiatives and government policies can enhance real-time monitoring and facilitate proactive decision-making. Ultimately, a multidimensional approach that combines regulatory actions, community engagement, and technological solutions is crucial for sustainable air quality improvement in the study area.

REFERENCES

Almalawi, A., Alsolami, F., Khan, A. I., Alkhathlan, A., Fahad, A., Irshad, K., . . . Alfakeeh, A. S. (2022). An IoT based system for magnify air pollution monitoring and prognosis using hybrid artificial intelligence technique. *Environmental Research*, *206*, 112576.

Arogundade, O., Abayomi, F., Abayomi-Alli, A., Misra, S., Alonge, C., Olaleye, T., & Ahuja, R. (2021). Cloud computing offered capabilities: threats to software vendors. *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020* (pp. 227-239). Springer.

Arroyo, P., Herrero, J. L., Suárez, J. I., & Lozano, J. (2019). Wireless Sensor Network Combined with Cloud Computing for Air Quality Monitoring. *Sensors*, *19*(3). doi:https://doi.org/10.3390/s19030691

Asha, P., Natrayan, L., Geetha, B., Beulah, J. R., Sumathy, R., Varalakshmi, G., & Neelakandan, S. (2022). IoT enabled environmental toxicology for air pollution monitoring using AI techniques. *Environmental Research*, *205*, 112574. doi:https://doi.org/10.1016/j.envres.2021.112574

Asmussen, C. B., & Møller, C. (2019). Smart literature review: a practical topic modelling approach to exploratory literature review. *Journal of Big Data*, *6*(1), 1-18.

Chauhan, U., & Shah, A. (2021). Topic modeling using latent Dirichlet allocation: A survey. *CM Computing Surveys (CSUR)*, *54*(7), 1-35.

Chukwu, T. M., Morse, S., & Murphy, R. (2022). Poor air quality in urban settings: A comparison of perceptual indicators, causes and management in two cities. *Sustainability* , *14*(3), 1438.

Kumar, K., & Pande, B. P. (2023). Air pollution prediction with machine learning: a case study of Indian cities. *International Journal of Environmental Science and Technology*, *20*, 5333–5348. doi:https://doi.org/10.1007/s13762-022-04241-5

Kumari, P., & Toshniwal, D. (2022). Impact of lockdown measures during COVID-19 on air quality– A case study of India. *International Journal of Environmental Health Research*, *32*(3), 503-510. doi:10.1080/09603123.2020.1778646

Liu, Y., Wang, P., Li, Y., Wen, L., & Deng, X. (2022). Air quality prediction models based on meteorological factors and real-time data of industrial waste gas. *Scientific Reports*, *12*(1), 9253.

Mengara, M., Park, E., Jang, J., & Y.Yoo. (2022). Attention-based distributed deep learning model for air quality forecasting. *Sustainability*, *14*(6), 3269.

Mi, J., Sun, X., Zhang, S., & Liu, N. (2022). Residential Environment Pollution Monitoring System Based on Cloud Computing and Internet of Things. *Residential Environment Pollution Monitoring System Based on Cloud Computing and Internet of Things*, *2022*, 8 pages. doi:https://doi.org/10.1155/2022/1013300

Odekanle, E. L., Fakinle, B. S., Odejobi, O. J., Akangbe, O. E., Sonibare, J. A., Akeredolu, F. A., & Oladoja, O. M. (2022). COVID-19 induced restriction in developing countries and its impacts on pollution load: case study of Lagos mega city. *Heliyon*, *8*(8).

Olaleye, T. O., Arogundade, O. T., Johnbosco, A., Sadare, O. O., Azeez, A. M., Opatunji, A. A., . . . Akintunde, S. (2023). Phishing Attack and Defense: An Exploratory Data Analytics of Uniform Resource Locators for Cybersecurity. In A. Khang, V. Abdullaye, B. Jadhav, S. Gupta, & G. Morris, *AI-Centric Modeling and Analytics* (pp. 304-323). Taylor & Francis .

Olaleye, T. O., Arogundade, O. T., Misra, S., Abayomi-Alli, A., & Kose, U. (2023). Predictive Analytics and Software Defect Severity: A Systematic Review and Future Directions. *Scientific Programming , 2023*, 1-18. doi:https://doi.org/10.1155/2023/6221388

Popović, I., Radovanovic, I., Vajs, I., Drajic, D., & Gligorić, N. (2022). Building Low-Cost Sensing Infrastructure for Air Quality Monitoring in Urban Areas Based on Fog Computing. *Sensors, 22*(3). doi:https://doi.org/10.3390/s22031026

Wang, T., Du, H., Zhao, Z., Zhou, Z., Russo, A., Xi, H., & C. Z. (2022). Prediction of the Impact of Meteorological Conditions on Air Quality during the 2022 Beijing Winter Olympics. *Sustainability, 14*(8), 11616. doi:https://doi.org/10.3390/su14084

Wang, T., Du, H., Zhao, Z., Zhou, Z., Russo, A., Xi, H., . . . Zhou, C. (2022). Prediction of the Impact of Meteorological Conditions on Air Quality during the 2022 Beijing Winter Olympics. *Sustainability, 14*(8), 11616. doi:https://doi.org/10.3390/su14084574

Zhang, D., Pan, S. L., Yu, J., & Liu, W. (2022). Orchestrating big data analytics capability for sustainability: A study of air pollution management in China. *Information & Management, 59*(5), 103231.