Research online Publishing
www.ronpub.com

Open Access

Open Journal of Internet of Things (OJIOT)
Volume 9, Issue 1, 2024
http://www.ronpub.com/ojiot
ISSN 2364-7108

Towards Sustainable Smart Cities: Monitoring Air Quality and Comfort for Citizen Health

Jaime Fabian Arias Aguilar, Everton Gomede, Leonardo de Sousa Mendes

School of Electrical and Computer Engineering, University of Campinas.
Albert Einstein Avenue, n° 901 - Cidade Universitária Zeferino Vaz, Campinas, Brazil.
ajaimefabian@gmail.com, [lmendes, gomede]@unicamp.br

Abstract

In smart city development, addressing air pollution and climate change through advanced environmental monitoring systems is crucial for enhancing urban quality of life and public health. This study focuses on the architecture of an intelligent system designed for real-time environmental monitoring in smart cities. The system aims to improve urban quality of life by addressing air pollution and climate change and assessing their impact on public health. The proposed system uses Arduino technology and integrated sensors to monitor PM10, PM2.5, toxic gases, and temperature. It incorporates a database in InfluxDB and Node-RED for efficient data management and visualization. The analysis employs the Knowledge Discovery in Data (KDD) methodology, Principal Component Analysis (PCA), and the DBSCAN algorithm for clustering high-pollution areas. The findings highlight the significant impact of air quality variables on environmental comfort. The system effectively identifies areas with high pollution levels, enabling informed urban planning and decision-making. In conclusion, this study emphasizes the need for effective air quality management and cross-sector collaboration to create healthier urban environments. The intelligent system demonstrates the potential for enhancing environmental comfort and addressing the environmental challenges of modern cities.

Type of Paper and Keywords

Regular research paper: *smart city, environmental comfort, air pollution, public health, air quality, sensors, IoT*

1 Introduction

A smart city can be conceptualized as a traditional urban environment augmented with innovative features designed to enhance the quality of life for its citizens and promote overall well-being. Economically, integrating technology within a smart city can stimulate and revitalize the local economy, fostering the growth of the digital economy. Socially, a smart city encourages the formation of online communities and enriches social life by providing platforms for interaction and engagement. Politically, a smart city enhances citizen participation and engagement through online decision-making, fostering a more inclusive and responsive governance system. Environmentally, a smart city integrates urban living with ecological responsibilities, promoting sustainability and encouraging lifestyles that preserve natural elements, such as green corridors for flora and fauna. Technologically, a smart city is characterized by its advanced infrastructure and services, leveraging cutting-edge technology to improve urban living conditions and efficiency [4].

To further understand the implications of smart cities, it is crucial to examine the environmental system, which provides valuable insights into the functioning of a city. As noted by Cueva, Lopera, and Torner [40],

humans have historically sought to understand the interplay between climatic conditions and human well-being, particularly regarding bio-climatic comfort. In his study, Hippocrates posited that health and well-being are intrinsically linked to climate. His treatise "On Air, Water, and Places" emphasized that elements such as air, water, and climatic conditions are crucial for the health of a city's inhabitants. Consequently, it is evident that environmental factors significantly impact an individual's health, well-being, and happiness [40], [4].

Given the significance of environmental factors, defining comfort is inherently complex due to its subjective nature and the necessity to encompass various perspectives. Several researchers describe comfort primarily as a state of climatic or thermal well-being; however, it also encompasses other forms of material satisfaction. This well-being is emphasized due to a harmonious balance between humans and their environment, making it a topic of ongoing interest and diverse interpretations [40].

Building on these concepts, Max Sorre, in his study The Foundations of Human Geography, discusses the concept of climatic comfort and its relationship with the micro-climate of cities and human-induced modifications [46]. The sustainability and livability of urban environments are pivotal aspects of this discourse. Consequently, smart cities endeavor to integrate advanced technologies across environmental sectors to address these concerns. Key initiatives include reducing CO2 emissions, controlling airborne pollutants, and managing water resources more efficiently. The incorporation of greenery into urban areas enhances both residents' well-being and the health of ecosystems. Furthermore, efforts to minimize waste generation and improve waste management are crucial in advancing the circular economy. Metrics such as temperature, humidity, toxic gases, particulate matter, and noise pollution are vital for assessing human comfort in these settings.

However, migrating people from rural to urban areas leads to a scarcity of natural resources and environmental challenges in both cities and resource extraction areas. Uncontrolled urban growth exacerbates climate change, as urban areas account for 70% of greenhouse gas (GHG) emissions, with urban transport contributing 40% of these emissions. The primary issues associated with this growth include high resource consumption, pollution, and waste generation, with significant pollutants being carbon dioxide, nitrogen oxide, and tropospheric ozone.

To address these environmental challenges, CETESB DE PORTAS ABERTAS[1] presented a vehicle pollution control plan for 2023-2025, responding to Brazil's vehicle fleet reaching 947,743 vehicles in 2022[2]. This plan includes (a) regulation and control of new and existing vehicles and (b) institutional and technological actions. Additionally, there is an emphasized need for advanced monitoring systems for specific locations. This project aims to develop a real-time response system to address these challenges.

Building on the regulatory and technological frameworks, this study aims to explore various researchers' perspectives on the concepts of intelligence and environmental comfort to develop an intelligent system that assesses environmental impact. This system will enable individuals to monitor the state of the environment in real-time and travel safely. The research will delve into multiple viewpoints within the field of environmental comfort and create a monitoring system to track levels of hazardous gases such as carbon monoxide (CO), ozone (O3), nitrogen dioxide (NO2), and particulate matter (PM2.5 and PM10), using Arduino and integrated sensors as the primary tracking mechanism.

Furthermore, sensor data will be gathered and analyzed to establish correlations and determine the environmental comfort level. The study will also expand the scope of applications for environmental intelligence and evaluate different data mining techniques to enhance the system's effectiveness. The proposed research project investigates whether the combination of temperature, humidity, toxic gases, and particulate matter can indicate human comfort. The study is driven by specific inquiry questions, such as (a) identifying elements that influence individuals' comfort within their surroundings, (b) evaluating existing technological solutions for assessing environmental comfort impacts, (c) analyzing how short-term exposure to environmental pollutant concentrations (measured in $\mu g/m^3$) affects the health and comfort of Campinas residents, and (d) determining minimal contamination levels that may pose health risks. This research aims to comprehensively understand the factors contributing to human comfort and health, assess current technologies, and identify critical pollutant thresholds.

Viewed through the lens of smart cities, this project initiates discussions involving multiple stakeholders to foster a secure, hygienic, and conducive environment for a high quality of life. This study provides insights into how the environment influences urban settings. While previous studies have primarily focused on technological interventions to enhance citizen safety, none have addressed the measurement of environmental comfort.

[1] https://cetesb.sp.gov.br/

[2] https://cetesb.sp.gov.br/veicular/wp-content/uploads/sites/6/2023/04/Plano-de-Controle-de-Prevencao-Veicular-2023-2025.pdf

This research endeavor introduces pioneering efforts in environmental monitoring through sensor-based data collection, processing, and dissemination of real-time information pertinent to environmental comfort.

The following sections will outline the proposed research methodology and system architecture, discuss the theoretical foundations of our approach, and present an analysis to evaluate the effectiveness of the modified method. Additionally, we will provide a comprehensive review of the environmental landscape in Brazil. Finally, we will discuss the practical implications of this approach and suggest avenues for future research.

1.1 Key Contributions

In the growing concerns about sustainability and well-being in smart cities, this research comprehensively addresses the challenges and opportunities in air quality management and urban comfort. Through an innovative approach that combines advanced technologies with rigorous analytical methodologies, the study aims to provide practical and effective solutions to enhance public health and quality of life in urban environments. The following highlights the key contributions of this work, which reflect both the impact of the research on understanding atmospheric pollution and its practical application in the design and management of sustainable smart cities:

1. **Impact of Atmospheric Pollution on Public Health:** The study underscores the severe threat that air pollution poses to public health. The research significantly contributes by addressing the need for constant and efficient air quality monitoring in smart cities, which is crucial for mitigating health risks associated with atmospheric pollution and protecting vulnerable communities.

2. **Challenges of Toxic Gas and Particulate Matter Pollution:** The analysis addresses the challenges associated with pollution from toxic gases and particulate matter (PM), which adversely affect human health, such as respiratory problems and cardiovascular diseases. The research provides a practical solution by developing an intelligent system that uses advanced sensors for real-time detection and monitoring of these pollutants, offering a valuable tool for proactive air quality management.

3. **Urban Comfort and Climate Well-being:** The work highlights the importance of thermal comfort and humidity in the perception of well-being in urban settings. An integrated approach is proposed that considers air quality and other environmental factors such as temperature and humidity. This holistic approach is essential for designing sustainable and pleasant cities for their inhabitants, enhancing their quality of life.

4. **Intersection of Urban Planning and Environmental Health:** The research emphasizes the need to integrate urban planning and environmental health. It presents a framework for analyzing the environmental impact on public health within the context of smart cities. This provides crucial data for urban planners and promotes urban development that prioritizes the well-being of citizens.

5. **Relevance of Real-Time Monitoring:** The study highlights the importance of systems that enable continuous air quality monitoring and other environmental parameters. The contribution of designing an IoT-based system for real-time monitoring allows for proactive environmental management, facilitating early intervention and improving response capacity to pollution events.

6. **Application of Knowledge Discovery in Data (KDD) Methodology:** The research rigorously applies the KDD methodology to extract valuable insights from the collected data. The KDD process in this study includes data selection, statistical analysis, and pattern identification using advanced techniques such as histograms and key statistical measures. This structured approach ensures a deep understanding of air quality data, allowing for the identification of extreme values that could represent risks to public health. The KDD methodology provides a solid foundation for improving future air quality management technologies and policies.

7. **Analysis of Low-Cost Sensors for Measuring Environmental Variables:** The study also examines the application of low-cost sensors to measure key environmental variables, which reduces monitoring costs and expands the accessibility of these technologies in resource-limited areas. This strategy facilitates the implementation of monitoring systems in various contexts, promoting sustainability and equity in environmental management.

8. **Presentation of a Reference Framework for IoT Architecture:** Finally, the research offers a reference framework for the IoT architecture used in the monitoring system. This framework is crucial for effectively integrating sensors, communication modules, and analysis platforms, providing a solid

basis for future developments in environmental quality management in smart cities.

The research presents a comprehensive technological solution that addresses the challenges of air quality and urban comfort, highlighting the application of the KDD methodology for in-depth data analysis. The study significantly contributes to creating healthier and more sustainable cities through advanced technologies and real-time monitoring, improving citizens' quality of life, and promoting environmental well-being.

2 Related Work

In this section, we emphasize the critical importance of the environment for public health and the severe consequences of environmental degradation. We describe several polluting variables and analyze their impact on citizens' health. Additionally, we explore fundamental concepts related to data mining, sensors, and architectures presented in similar projects.

2.1 Particulate Matter and Air Quality

Several environmental studies underscore the significance of particulate matter (PM10 and PM2.5) suspended in the atmosphere. For instance, [25] provides a spatio-temporal analysis of PM10 and PM2.5, elucidating the direct impact of factors such as wind direction, road infrastructure, and the spatial dispersion of pollutant sources on their concentrations.

Related research by [17] and [35] characterizes air quality through the lens of PM10 and PM2.5 levels. [17] reveals average PM2.5 values of 30.11 µg/m³ and PM10 values of 39.82 µg/m³, indicating a direct correlation between ambient temperature and particulate matter concentration, while relative humidity shows an inverse relationship.

In Colombia, studies on air quality and particulate matter by [37] and [9] provide significant insights. [37] discusses variations in particle diameters attributed to emission sources and examines the physical factors affecting particle re-suspension. Concurrently, [9] investigates the influence of travel patterns, traffic volume, and street layout on pollutant exposure, employing a methodology involving area analysis, particle monitoring, inhalation rate adjustment, and data collection.

The application of data science and computational techniques in air pollution analysis is also noteworthy. [44] employs the CART method to partition observations into homogeneous terminal nodes. Similarly, [49] demonstrates the significant influence of meteorological and air pollution factors on fine dust using machine learning algorithms, such as linear regression and decision trees. Furthermore, [1] advocates using MLP-NN neural networks to model particulate matter due to its non-linear characteristics.

Researchers have also addressed monitoring toxic gases such as HCN, CO, H2S, NO2, NH3, and SO2. For example, [50] utilizes IoT technology with components like the STM32F103 chip and a wireless transmission module. At the same time, [30] proposes a sensor system that sends alerts via GSM when pollutant levels exceed predefined limits.

For thermal comfort determination, [34] and [28] analyze relative humidity and temperature using programs like Labview for monitoring, data acquisition, and parameter calculation.

Big data technologies have proven valuable for environmental monitoring. [45] discusses using Apache, MapReduce, HBase, data mining, and visualization for Environmental Monitoring and Management Systems (EMMS). Additionally, [16] proposes novel data visualization dashboards to support anomaly prediction and enhance decision-making.

In summary, modern technologies and analytical techniques are essential for understanding and mitigating the impact of air pollution on public health.

2.2 Architecture and Components

The architecture described in [45] leverages sensor technology to collect data, which is then processed in an upper layer using various big data technologies and data mining techniques. This process transforms raw data into useful information for different applications, addressing the needs of Environmental Monitoring and Management Systems (EMMS). For environmental monitoring using the Internet of Things (IoT), authors in [20] implemented a 4-layer architecture:

- *Application Layer*: This layer stores, organizes, processes, and shares environmental data and other information obtained from sensors, devices, and web services. It also includes decision-making functions through monitoring various variables.

- *Middleware Layer*: Positioned between the network layer and the application layer, this layer utilizes software, tools, models, and platforms for data management.

- *Network Layer*: Responsible for transmitting information and interconnecting systems and platforms. This layer is divided into access networks (short-range wireless networks) and transport networks (WAN networks).

- *Perception Layer*: Collects data and information about physical factors related to environmental monitoring and management. Real-time data collection based on IoT in this layer involves using multiple sensors, including RS platforms, on-site instruments, mobile technology, RFID, and other sensors.

The EMMS architecture is depicted in Figure 1.

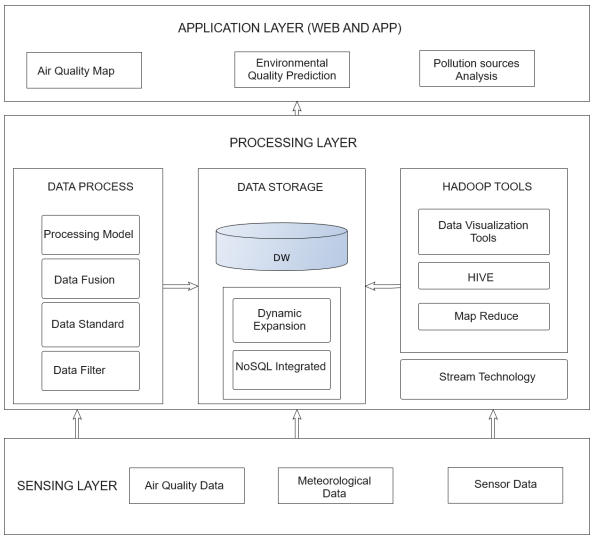

Figure 1: Detailed Architecture of the Environmental Monitoring and Management System (EMMS), illustrating the interactions between the sensing, processing, and application layers.

The Environmental Monitoring and Management System (EMMS) architecture comprises three primary layers: sensing, processing, and application. The sensing layer is responsible for collecting various types of data, including air quality data, meteorological data, and sensor data. This data is then transmitted to the processing layer, which comprises several components: data processing, data storage, and Hadoop tools. The data process component handles the processing model, data fusion, standardization, and filtering. The data storage component features a Data Warehouse (DW) that supports dynamic expansion and integrates NoSQL databases for scalable data management. Hadoop tools[3] within the processing layer facilitate data visualization, data processing with Hive[4], and large-scale data analysis using MapReduce and stream technology. Finally, the application layer utilizes the processed data to provide user-facing services such as air quality maps, environmental quality prediction, and pollution sources analysis through web and mobile applications. This layered architecture ensures efficient data collection, processing, storage, and visualization, enabling comprehensive environmental monitoring and management.

[3] https://hadoop.apache.org/
[4] https://hive.apache.org/

2.3 Environment, Society, and Health

From its inception to today, humanity has contended with myriad challenges to survive as a species on Earth. From addressing hunger to adapting to extreme temperatures, humans have continually sought new ways to navigate their environment.

As society has evolved, diverse cultural groups have emerged, each characterized by distinct behaviors and values. Human interaction with biotic and abiotic elements has led to adverse effects, some of which carry irreversible consequences. Such issues, as detailed by [12], are recognized as ecological problems.

Human activities generate particles, gases, vapors, and other atmospheric pollutants, crucial in transmitting airborne diseases and significant for public health. Scholars have noted that air pollution disproportionately impacts impoverished communities, even within affluent nations. Karti Sandilya, a study author from the NGO Pure Earth, underscores the interconnectedness of pollution, poverty, health, and social injustice.

In 2015, pollution in Brazil accounted for 101,739 deaths, accounting for 7.49% of the country's total fatalities. Of these, 70,685 deaths were attributed to air pollution alone. Brazil ranked 148th globally regarding pollution-related deaths, trailing behind neighboring South American countries such as Uruguay, Chile, and Ecuador. This analysis, conducted by The Lancet, encompassed 188 countries [33].

2.4 Atmospheric Pollution

Toxic gas pollution occurs when the Earth's atmospheric composition is altered by solid or liquid gases or suspended particles in proportions different from natural levels, posing significant risks to human health. These gases can damage various materials, reduce visibility, and emit unpleasant odors, constituting detrimental environmental changes [26].

Certain industries release hazardous gases into the atmosphere. Pollution from specific sources can disperse across extensive geographic regions with relatively minimal localized impact. However, diffuse pollution can be concentrated by wind patterns and topographical features, leading to significant effects in urban areas. Since the 1970s, chlorofluorocarbons (CFCs) have been identified as potent greenhouse gases, significantly

contributing to the depletion of the ozone layer in the stratosphere.

Indoor air pollution arises from various sources, including tobacco use, specific construction materials, cleaning agents, and household furnishings. Many researchers posit that this pollution also contributes to global warming [3]. Chlorofluorocarbons (CFCs) are primarily responsible for damaging the ozone layer. Typically, pollutants disperse from their origins without accumulating to hazardous levels. However, weather patterns can transport pollutants from land sources, posing risks to previously clean environments. Health issues associated with increased air pollution include impaired lung function and elevated risks of heart attacks [18]. The dispersion of pollutants in the atmosphere and their subsequent effects play a critical role in the overall impact of air pollution.

2.5 Toxic Gases

Toxic gases are substances that, when inhaled over time, can exert various effects on human health, ranging from unconsciousness to potentially fatal outcomes if untreated. These gases pose significant risks and challenges for the environment and living organisms at high atmospheric concentrations. While some pollutant gases originate naturally, such as those emitted by volcanic activity, industrial activities, including fossil fuel combustion and heavy road traffic, significantly contribute to their production [11]. Table 1 displays the Air Quality Guideline levels the World Health Organization set.

Table 1: Air Quality Guideline levels and interim targets for CO, NO_2, and NH_3 (in g/m^3, ppm)

Recommendation	CO	NO_2	NH_3
Interim target 1	7, 0.006	120, 0.064	800, 1.149
Interim target 2	N/A	50, 0.027	400, 0.574
AQG level	4	25, 0.013	200, 0.287

Common pollutant gases include sulfur compounds, carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2), and ammonia (NH3). Sulfur compounds, released during the combustion of coal or oil, irritate the eyes and respiratory system, exacerbating respiratory diseases, particularly in vulnerable populations. Carbon monoxide (CO), a colorless and odorless gas, displaces oxygen in the bloodstream, causing symptoms ranging from headaches to death at high concentrations. Sulfur dioxide (SO2), known for its pungent odor, is emitted from combustion processes and industrial activities, contributing to air pollution and acid rain formation. Nitrogen dioxide (NO2), produced from fuel combustion in vehicles and industrial processes, is a toxic gas characterized by a strong odor and brown color. Ammonia (NH3), originating from the decomposition of organic matter and agricultural practices, contributes to nitrogen oxide formation, posing health risks to humans [11].

2.6 Particulate Material

Particulate matter (PM) in the atmosphere is a mix of solid and liquid particles, with diameters ranging from 20 microns to less than 0.05 microns. Due to its complex nature, PM is characterized not only by its concentration but also by its diameter, chemical composition, phase, and morphology [43]. Table 2 classifies types of particulate matter harmful to human health. Smaller particles are particularly dangerous as they can penetrate deeper into the respiratory tract, potentially causing severe health issues and even death [22].

Table 2: Particulate Material Types

Particulate Matter	Diameter	Name
PM 0.1	up to 0.1 µm	Ultrafine
PM 2.5	0.1 to 2.5 µm	Fine
PM 10	2.5 to 10 µm	Coarse
PTS	10 to 50 µm	Thick Inhalable

Fine particles, often resulting from combustion processes in vehicles, industries, and biomass burning, differ from coarse inhalable particles generated by mechanical processes like wind erosion, sea waves, and grinding operations. Particle size significantly influences their health impact, with smaller particles such as PM2.5 being linked to respiratory and cardiovascular issues on a global scale [8].

In its 2005 global update, the World Health Organization (WHO)[5] established interim targets for air quality to reduce high concentrations of air pollutants that pose severe health risks. Table 3 presents the short-term (24-hour) exposure levels for PM10 and PM2.5.

Table 3: AQG level and interim targets for PM_{10} and $PM_{2.5}$

Recommendation	PM_{10} ($\mu g/m^3$)	$PM_{2.5}$ ($\mu g/m^3$)
Interim target 1	150	75
Interim target 2	100	50
Interim target 3	75	37.5
Interim target 4	50	25
AQG level	45	15

[5] https://www.who.int/

The recommended short-term PM2.5 Air Quality Guideline (AQG) level is 15 µg/m³, and for PM10, it is 45 µg/m³. These levels are defined as the 99th percentile of the annual distribution of 24-hour average concentrations. Exposure to PM2.5 at these levels is associated with varying mortality rates at different interim target levels.

2.7 Comfort, Humidity, and Heatwave

In the 20th century, studies on the relationship between humans, cities, and climate began with the pioneering work of [32] and [13]. These studies initiated urban climate research, highlighting the importance of understanding energy and humidity balances for bio-climatic comfort. Urban environments can positively and negatively impact comfort, prompting recent research to focus on the intersection of urban climate, comfort, and urban planning.

Humidity, influenced by rainfall, proximity to the sea, vegetation, and air temperature, affects human comfort by altering air quality perception, thermal sensation, and skin moisture. Thermal comfort, defined as satisfaction with the thermal environment [19], involves multiple factors that affect sensory and physiological responses [24]; [5]. Mechanical comfort relates to the direct impact of wind [10]. Most individuals feel comfortable at temperatures between 21°C and 26°C and relative humidity levels between 30% and 70%.

2.7.1 Humidex

In 1979, Masterton and Richardson introduced the *humidex*, a temperature-humidity index designed to correlate external thermal discomfort in Canada's temperate zones [36]. This index utilizes air temperature and relative humidity as its sole meteorological parameters. The *humidex* is based on two key hypotheses: first, the human body's neutral point for heat balance falls within the range of 27°C to 30°C; second, the body struggles to dissipate heat effectively when the temperature exceeds 32°C and the relative humidity surpasses 75%. The *humidex* is mathematically defined as:

$$HD = t_a + \frac{5}{9}(p_{as} - 10) \tag{1}$$

where t_a represents the temperature in degrees Celsius, and p_{as} denotes the water vapor pressure in millibars, calculated according to the equation:

$$p_{as} = 6,112 \left(10^{\frac{7,5t_a}{237,7+t_a}}\right) \frac{RH}{100} \tag{2}$$

The Relative Humidity (RH) index provides insights into human comfort levels amidst varying weather conditions, emphasizing the impact of temperature and humidity on thermal discomfort.

Table 4 displays the thermal sensation ranges based on *humidex* values. Discomfort is not typically felt when the *humidex* is below 29. However, values of 40 or higher indicate significant thermal discomfort due to factors such as air currents, temperature differences, or surface variations.

Table 4: Humidex levels

Thermal sensation	Humidex value
No discomfort	< 29
Mild discomfort	From 30 to 39
Discomfort, avoid exertion	From 40 to 45
Risk	From 46 to 54
Impending heat stroke	> 54

Table 4 specifies four ranges: Humidex values below 29 indicate no discomfort, values between 30 and 39 suggest mild discomfort, and values from 40 to 45 indicate discomfort where exertion should be avoided. Humidex values between 46 and 54 represent a risk of heat-related health issues, and values exceeding 54 signal an impending heat stroke, underscoring severe thermal stress. This classification helps assess and respond to different levels of heat discomfort for public health and safety.

2.7.2 Heatwave

Heatwave definitions vary in the literature, with some considering temperatures exceeding 35ºC as a criterion. Heat waves can also be defined as consecutive days with temperatures surpassing the 95th percentile.

This study utilizes the CTXP90 index (cutoff value from the P90 of maximum temperatures over 15 days) and the Heat Wave Magnitude Index (HWMI). The HWMI defines a heat wave as two or more consecutive days with maximum temperatures above the 90th percentile in a 30-day window centered on the assessment day. Adapted from [42] and [23], this algorithm identifies hot days based on historical data spanning 22 years. The algorithm involves:

1. Calculating the 90th percentile (P90(d)) of maximum temperatures for each day d over a 15-day range.

2. Iterating through observations until a heat wave condition is met: a heat wave begins if maximum temperatures on days d, d+1, and d+2 exceed their respective P90 values.

3. Marking consecutive days as part of the heat wave and incrementing the counter (i) until all observations are processed.

Since the data is time-dependent, machine learning algorithms are adapted to recognize temporal patterns and behaviors.

2.7.3 Database Architectures for IoT

Many database systems with varying speed, latency, and scalability characteristics have been developed to manage massive volumes of real-time data, mostly in smart city applications. These developments are a result of the expansion of IoT infrastructures. Common solutions, each tailored for certain purposes and scenarios, include distributed databases, time series databases, and cloud-based methods.

Time series databases (TSDB) are particularly useful for managing sensor data in the Internet of Things (IoT) space because they maximize temporal data querying and storage. One of the most widely used TSDBs in this context is InfluxDB, which has demonstrated efficacy in gathering and evaluating substantial amounts of real-time data. Because its design enables quick searches and effective storage, it is a desirable choice for Internet of Things applications that need continuous and real-time data handling [38][29].

Though TSDBs are very helpful for cloud-based and centralized applications, they can present some difficulties when used in large-scale distributed infrastructures. For instance, sensor networks in smart cities produce vast amounts of data that are scattered geographically, which can lead to significant transfer costs and latency problems when utilizing a fully centralized cloud-based architecture. Under these circumstances, comparing solutions that improve performance without sacrificing system scalability and economy is required.

The flexibility and scalability of cloud-based systems have made them popular for managing sensor data in Internet of Things applications. Under these methods, sensor data is gathered and sent to cloud servers for processing, storing, and querying. With this strategy, businesses don't have to worry about the underlying hardware and can swiftly scale their infrastructures to manage growing data volumes. That being said, latency is a major problem with cloud infrastructure. The delays that come with transmitting data to the cloud for processing might lead to subpar performance in applications like smart cities, where quick decision-making is essential [48].

Furthermore, if sensors' quantity and frequency increase, the data transport cost between dispersed devices and the cloud may rise significantly. To lessen dependency on centralized cloud services, distributed architectures—one more effective alternative—have been investigated in response to these difficulties. Distributed databases have become a viable way to overcome the drawbacks of centralized designs in large-scale IoT infrastructures, such as smart cities. As a result of these databases' ability to process and store data across different geographical locations, latency and transfer costs are decreased as all data is not required to pass via a single central cloud server [39][2].

Distributed architectures, in particular, have been shown to greatly increase performance in wireless sensor networks in recent research, such as A SPARQL Benchmark for Distributed Databases in IoT Environments. This method more effectively distributes data, enabling queries to run more quickly in geographically scattered situations. According to the study, RDF3X and Hexastore are two partitioning strategies and topologies that can optimize network traffic and enhance the management of scattered data [48][6].

Depending on the context of the application, cloud-based techniques and distributed architectures are comparable, and each has its advantages. Cloud solutions are perfect for smaller or centralized applications since they are simple to implement and scale. On the other hand, distributed architectures provide advantages over centralized IoT infrastructures, such as smart cities, in terms of lowering latency and data transfer costs [31].

While time series databases like InfluxDB are optimized for quick queries, their use in distributed infrastructures may be limited by the costs associated with transferring large volumes of data to the cloud, according to recent studies like the Application of Time Series Database for IoT Smart City Platform. Distributed designs, on the other hand, enable more localized data processing and storage, which lowers network traffic and boosts overall system efficiency [27]

To summarize, selecting database designs for the Internet of Things necessitates careful consideration of factors such as infrastructure size, data type, and performance demands. Large-scale dispersed situations may limit TSDBs like InfluxDB in centralized designs, even though these databases are incredibly efficient at managing massive amounts of temporal data in real-time. In these situations, distributed databases provide a more scalable and economical option, improving performance by cutting down on latency and network expenses. For IoT infrastructures like smart cities, where scalability and efficiency are critical, combining distributed methods with technologies like InfluxDB may offer the perfect mix.

3 Materials and Methods

We designed, developed, and implemented computational intelligence technologies to monitor air quality and noise levels, establishing the comfort levels associated with various environmental conditions.

3.1 Materials

The system initially used sensors and a programming board, with plans for a database and cloud connectivity to store collected data. The core hardware, an ESP32 SIM800L T-CALL V1.3, gathered data from integrated sensors and communicated with the server via a GSM module, efficiently managing data collection and transmission.

3.1.1 LILYGO® TTGO T-Call V1.4 ESP32

The TTGO T-Call is an advanced ESP32 development board featuring a SIM800L GSM/GPRS module. In addition to Bluetooth and Wi-Fi capabilities, it allows communication via SMS, phone calls, and internet connectivity. It utilizes the ESPRESSIF-ESP32 chipset, which includes a 240MHz Xtensa® processor with one or two 32-bit LX6 cores, ensuring robust performance. The board has 4MB of QSPI flash memory, 8MB of PSRAM, and 520 kB of SRAM.

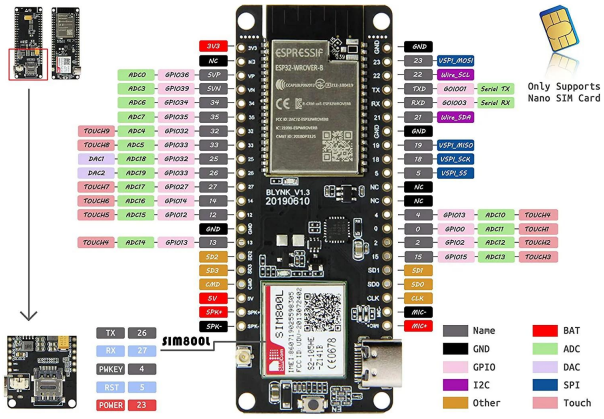

Figure 2: Pinout Diagram of the ESP32 SIM Card SIM800L T-Call V1.4, showcasing the various pin configurations and functionalities.

Figure 2 illustrates the pinout and specifications of the pins. This comprehensive information facilitates informed decision-making and implementation for various applications. Key hardware features of the TTGO T-Call include a reset button and the CP2104 chip for USB to TTL communication. Its modular interface supports various connectivity options such as UART, SPI, SDIO, I2C, PWM, TV PWM, I2S, and IRGPIO, making it highly versatile. The board operates within a voltage range of 2.7V to 3.6V, with a current consumption of approximately 70mA during operation and around 1.1mA in sleep mode.

Table 5: LILYGO® TTGO T-Call V1.4 ESP32

Hardware	Details
Chipset	ESPRESSIF-ESP32 240MHz Xtensa® single-/dual-core 32-bit LX6
FLASH	QSPI flash 4MB / PSRAM 8MB
SRAM	520 kB SRAM
Button	Reset
USB to TTL	CP2104
Modular interface	UART, SPI, SDIO, I2C, PWM, TV PWM, I2S, IRGPIO
Working voltage	2.7V-3.6V
Working current	About 70mA
Sleep current	About 1.1mA

Table 5 provides the hardware specifications for the LILYGO® TTGO T-Call V1.4 ESP32 module. It features an ESPRESSIF-ESP32 240MHz Xtensa® single-/dual-core 32-bit LX6 microprocessor. The module includes 4MB of QSPI flash memory and 8MB of PSRAM, and 520 kB of SRAM. It has a reset button and utilizes the CP2104 chip for USB to TTL communication. The modular interface supports protocols such as UART, SPI, SDIO, I2C, PWM, TV PWM, I2S, and IRGPIO. The working voltage ranges from 2.7V to 3.6V, with a working current of approximately 70mA and a sleep current of about 1.1mA. These specifications highlight the module's versatility and suitability for various IoT applications.

3.1.2 SDS011 Particulate Matter Sensor

The SDS011 sensor measures air quality using laser scattering to detect particles in the air, with a range of 0.3 to 10 micrometers, providing quick and accurate results.

Figure 3: SDS011 Particulate Matter Sensor is utilized to measure air quality.

Figure 3 illustrates the SDS011 Particulate Matter Sensor, which measures air quality using laser scattering to detect particles in the air. The sensor has a detection

range of 0.3 to 10 micrometers, providing quick and accurate results. This capability makes it a valuable tool for monitoring air quality and assessing particulate matter concentrations in various environments.

3.1.3 DHT22 Temperature and Humidity Sensor

The DHT22 sensor measures temperature and relative humidity with high precision and low cost, making it ideal for scientific data acquisition systems.

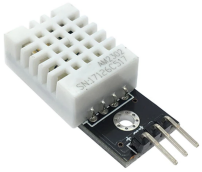

Figure 4: DHT22 Sensor, designed for measuring temperature and relative humidity.

Figure 4 illustrates the DHT22 Temperature and Humidity Sensor, which measures temperature and relative humidity with high precision and low cost. This makes it ideal for scientific data acquisition systems. The DHT22 sensor is widely used in various applications due to its accuracy and affordability, providing reliable environmental monitoring and analysis data.

3.1.4 MiCS-6814 Air Quality Sensor

The MiCS-6814 sensor detects concentrations of various gases (CO, NO_2, NH_3, among others) with independent channels for each gas, making it suitable for precise multi-gas detection.

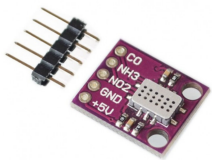

Figure 5: MiCS-6814 Air Quality Sensor detects concentrations of various gases including CO, NO2, and NH3.

Figure 5 illustrates the MiCS-6814 Air Quality Sensor, which detects concentrations of various gases, including CO, NO2, and NH3, using independent channels for each gas. This multi-gas detection capability makes it suitable for precise air quality monitoring, allowing for the accurate measurement of different pollutants in the environment. The MiCS-6814 sensor is a versatile tool for comprehensive air quality analysis in various applications.

3.1.5 Ublox Neo-6m Gy Neo6mv2 GPS Module

The GPS module monitored geographic coordinates in real-time, ensuring precise location tracking. It transmitted data to a server for processing, making it available for visualization and interpretation. Integration with Google Maps allowed users to view tracked locations on an interactive map, enabling applications like navigation, asset tracking, and environmental monitoring. This real-time system enhanced spatial data analysis, response to location-based events, and informed decision-making based on accurate positioning information.

3.1.6 GSM Network Module

The GSM 900 system utilizes 900 MHz frequencies for data transmission, employing Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) technologies. These technologies manage the allocation of frequencies and facilitate efficient communication. TDMA divides each frequency into time slots, allowing multiple users to share the same frequency without interference. FDMA assigns individual frequencies to different users, ensuring clear and reliable communication channels. This combination of technologies enables the GSM 900 system to support robust and seamless communication in urban and rural areas, providing extensive coverage and enhancing connectivity.

3.2 Methods

The project starts with a use case that includes three distinct blocks. The first stage analyzes human comfort based on a review of scientific articles [14][47]. Then, real data is collected for analysis, selecting relevant environmental variables and preparing a dataset for research [41].

Figure 6 illustrates the growth of the vehicle fleet in Campinas, as recorded by the IBGE, from 2006 to 2022. The graph shows a steady vehicle increase, from approximately 550,000 in 2006 to over 900,000 in 2022. This significant growth highlights the expanding urban mobility and potential challenges related to traffic congestion, air pollution, and infrastructure demands. The data underscores the need for comprehensive urban planning and environmental monitoring to address the implications of this increasing vehicle density on human comfort and urban sustainability.

The analysis of comfort in urban spaces is based on previous studies on urban heat islands and other factors

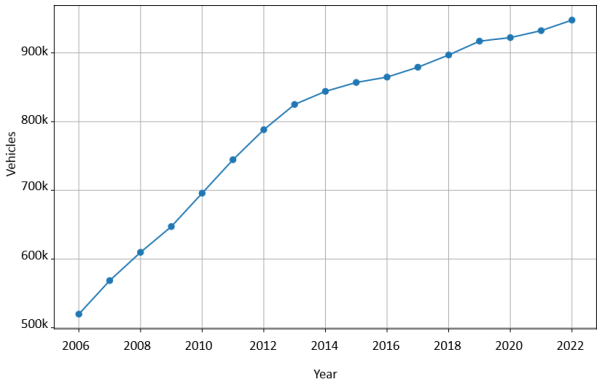

Figure 6: Growth of the Vehicle Fleet in Campinas from 2006 to 2022, as recorded by IBGE.

influencing users' psychological and physiological satisfaction in public spaces [14][47][41]. Users and stakeholders are identified through discussions, interviews, and questionnaires, characterizing areas of interest and concerns [15].

3.3 IoT Architecture System

Creating a quality environment for citizens is essential in a smart city system. The proposed system monitors environmental conditions in real-time, employing the IoT Reference Model to develop a comprehensive monitoring solution [7]. This architecture consists of four layers: sensing, network, processing, and application, as illustrated in Figure 7.

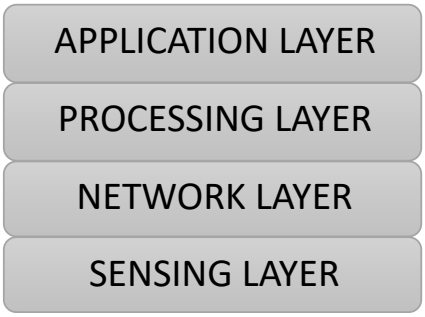

Figure 7: IoT-based System Architecture, illustrating the four key layers: Sensing Layer, Network Layer, Processing Layer, and Application Layer.

The ***sensing layer*** collects data via sensors and GPS during a person's journey. It also handles data preprocessing, including filtering and edge computing. This functionality is implemented on the Unit Measurement and Communication On-Board (UMCO) hardware, specifically the *ESP32 SIM Card SIM800L*. The hardware components used in this project include:

- DHT-22 Sensor
- MICS-6814 Air Quality Sensor
- SDS011 Particulate Matter Sensor
- Ublox Neo-6m GPS Module
- LILYGO TTGO ESP32

Part of the UMCO hardware also supports the ***network layer*** by providing data transmission via a mobile network using GPRS/GSM. Standard communication protocols for data transmission are encompassed within this layer. In the ***processing layer***, data management activities such as storage, cleaning, filtering, and analysis are conducted. This layer employs data preprocessing techniques and machine learning algorithms, like clustering, to identify patterns in the data generated by the previous layers. The ***application layer*** focuses on management and decision-making functions. Here, algorithms utilize pre-processed data to provide relevant information for specific applications. The proposed model aims to establish a comprehensive real-time environmental monitoring system. Key features of the model include:

- Analyzing comfort levels by correlating air quality and temperature data collected along the paths traveled by individuals.
- Monitoring traffic conditions through the Google Maps API.

Current IoT technologies are integrated to collect data, transmit it to the cloud, analyze it, and produce useful results.

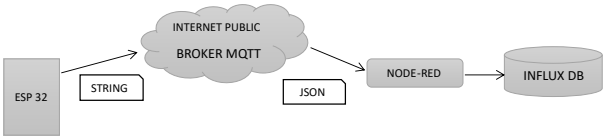

Figure 8: IoT System Operation Block Diagram, depicting data flow from the ESP32 module through a public MQTT broker to Node-RED and finally to InfluxDB for storage.

Figure 8 presents the block diagram of the system operation. Sensors connected to the ESP32 capture environmental variables transmitted to the cloud via a GSM module. Data is sent to a public broker (HiveMQ[6])

[6] https://www.hivemq.com/

using the MQTT protocol, interfacing with NODE-RED. The information, transmitted in JSON format, is stored in an InfluxDB database.

3.3.1 ESP32

The ESP32 is the project's central component, seamlessly integrating sensor functionalities and internet connectivity through its GSM module. Upon initialization, the program calibrates the sensors to ensure accurate data collection. It then establishes GSM connections to enable reliable data transmission. The ESP32 continuously retrieves sensor values, including particulate matter measurements, temperature, humidity, and various gases. This data is subsequently sent to a public broker for cloud processing, where it can be analyzed and visualized in real-time. By leveraging the ESP32's capabilities, the system ensures efficient and continuous environmental monitoring, facilitating informed decision-making and timely responses to changing environmental conditions.

3.3.2 NODE-RED

NODE-RED[7] is a graphical development tool that utilizes flows to receive, process, and transmit data efficiently. In this project, the ESP32 collects environmental data from various sensors and sends this data to a public broker using the MQTT protocol. NODE-RED then accesses this data from the broker, processes it, and stores it in a database. This setup allows for seamless integration and real-time data handling, enabling continuous monitoring and analysis of environmental conditions. The graphical interface of NODE-RED simplifies the development and management of data flows, enhancing the system's overall efficiency and usability.

3.3.3 InfluxDB

InfluxDB[8] stores time-series data, ensuring each data point is synchronized with a timestamp. This allows for precise tracking and analysis of temporal trends. InfluxDB categorizes and stores sensor data in this project into air quality, climatic temperature, spatial location, and time. This organization facilitates efficient data retrieval and evaluation, enabling comprehensive monitoring and analysis of environmental conditions across different dimensions.

[7] https://nodered.org/

[8] https://www.influxdata.com/

3.4 Decisions and Technological Choices

Several critical design choices were made when creating the Internet of Things (IoT) system described in this work to ensure its scalability, robustness, and suitability for monitoring urban air quality. The system integrates data storage, communication protocols, sensor data gathering, and analytic tools to provide real-time environmental insights. The selection of the GSM module for reliable data transfer and the ESP32 microcontroller for its flexible sensor integration capabilities were pivotal in these decisions.

InfluxDB was chosen for time-series data storage because it can rapidly handle large volumes of timestamped sensor data, facilitating synchronized data processing and retrieval. This design feature ensures the system can efficiently monitor and assess various environmental characteristics, including air quality indices, meteorological conditions, and geographic locations.

The materials and methods section outlines the comprehensive approach used in this study. The system's central component is the ESP32, which integrates sensor functionalities and internet connectivity via its GSM module. The system employs various sensors to monitor environmental parameters such as particulate matter (PM10 and PM2.5), temperature, humidity, and gases (CO, NO2, NH3). The data collected by these sensors is transmitted using the MQTT protocol to a public broker, where NODE-RED processes and routes it to InfluxDB. This time-series database stores the data with precise timestamps. The methodology involves calibrating the sensors, establishing GSM connections, and employing data analysis techniques such as Knowledge Discovery in Data (KDD), Principal Component Analysis (PCA), and the DBSCAN algorithm for clustering. This integrated approach ensures real-time monitoring, accurate data collection, and effective analysis to support environmental assessment and decision-making.

4 Results and Discussions

The collected data is initially analyzed to assess particulate matter concentrations (PM10 and PM2.5), temperature, humidity, and various gases (CO, NO2, NH3). Statistical analyses and visualization techniques are employed to identify patterns and anomalies within the dataset. Furthermore, applying data mining techniques, such as Principal Component Analysis (PCA) and the DBSCAN clustering algorithm, reveals significant insights into the spatial and temporal variations in air quality. The discussion contextualizes these findings within the broader framework of urban environmental health, highlighting the implications for

public health and urban planning. By comparing the results with established air quality guidelines and previous studies, this section underscores the effectiveness of the proposed intelligent system in providing real-time, actionable insights to mitigate environmental risks and enhance urban sustainability.

4.1 Data Analysis

After implementing the data collection and monitoring system, we conducted initial analyses to extract valuable insights, aiming to identify behavioral patterns of the primary variables. This process adhered to the Knowledge Discovery in Data (KDD) methodology, a well-established approach for leveraging raw datasets to enable intelligent system learning and decision-making [21].

The KDD process began with the *selection stage*, where techniques such as histograms and calculating key statistical measures were employed to understand and manipulate the database thoroughly. The dataset comprised 32,836 records, and Table 6 summarizes the statistical measures for each variable. Notably, the average particulate matter concentrations (PM10 and PM2.5) were 20.2516 µg/m³ and 11.8457 µg/m³, respectively, within the regular Air Quality Guideline (AQG) levels. However, these variables occasionally exhibited extreme values, which pose potential health threats according to AQG standards.

Table 6: Statistical Measures of the Dataset

Index	CO	NH3	NO2	Humid.	Temp.	P10	P2.5
Count	32836	32836	32836	32836	32836	32836	32836
Mean	4.385	0.683	0.719	53.972	23.374	20.252	11.846
Std. dev	0.009	0.038	2.423	17.797	4.657	13.284	8.111
Min	4.161	0.419	0.006	1.000	10.100	1.400	1.000
25%	4.385	0.670	-1.000	40.400	20.200	11.300	5.400
50%	4.385	0.680	0.140	51.900	22.700	16.900	9.200
75%	4.385	0.681	0.346	65.200	26.500	27.300	16.800
Max	6.011	1.946	103.525	99.900	52.800	261.100	78.800

Table 6 provides a comprehensive overview of the statistical measures for the dataset, encompassing 32,836 records for various environmental parameters, including CO, NH3, NO2, humidity, temperature, PM10, and PM2.5. The mean values indicate average concentrations and conditions, with CO at 4.385 ppm, NH3 at 0.683 ppm, and NO2 at 0.719 ppm. The dataset reveals a mean humidity of 53.972% and the average temperature of 23.374°C. The particulate matter concentrations show average values of 20.252 µg/m³ for PM10 and 11.846 µg/m³ for PM2.5. Standard deviations indicate variability, with NO2 showing the highest variability (2.423 ppm) among the gases. The maximum recorded values highlight potential extremes in the data, such as 261.100 µg/m³ for PM10 and 78.800 µg/m³ for PM2.5, which could pose significant health risks. The dataset captures various environmental conditions, providing a solid foundation for analyzing air quality and its implications on public health.

4.2 Detailed Analysis

This section aims to uncover the underlying patterns and relationships within the data, offering a comprehensive understanding of the environmental conditions. By employing statistical and visual analysis techniques, we can interpret the data's behavior and variability, essential for informed decision-making and effective environmental management. The following subsections provide a focused analysis of temperature, humidity, particulate matter, and gaseous pollutants, highlighting key findings and their implications.

4.2.1 Temperature and Humidity

Histograms were constructed for temperature and humidity, as shown in Figure 9, revealing a normal distribution and appropriate data dispersion. These visualizations illustrate these variables' potential ranges and behaviors in the observed environment, providing valuable insights into their variability and trends.

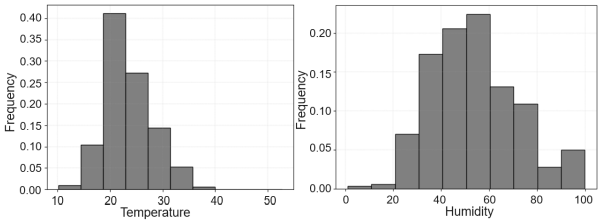

Figure 9: Histograms of Temperature and Humidity, displaying the frequency distribution of temperature and humidity readings.

Figure 9 presents histograms for temperature and humidity, revealing a normal distribution and appropriate data dispersion. These visualizations illustrate these variables' potential ranges and behaviors in the observed environment, providing valuable insights into their variability and trends. By analyzing these histograms, we can understand the typical conditions of temperature and humidity, which are crucial for assessing environmental comfort and planning appropriate interventions.

The temperature histogram shows a normal distribution centered around 20 to 25 degrees Celsius, indicating that most temperature readings fall within this range, with a few observations extending to 50 degrees Celsius. The humidity histogram also exhibits a normal distribution, with most readings clustered between 40%

and 80%, highlighting the prevalent humidity conditions in the observed environment. These visualizations reveal appropriate data dispersion and provide insights into the typical ranges and variability of temperature and humidity, essential for understanding the environmental context and assessing comfort levels. The normal distribution patterns suggest stable environmental conditions, valuable for predicting and managing urban climate and air quality.

4.2.2 Particulate Matter (PM10 and PM2.5)

The journey data, illustrated in Figure 10, highlights PM10 and PM2.5 concentrations over a specific route, demonstrating their impact on air quality. Geo-location data was also used to map this behavior spatially, providing a comprehensive view of how particulate matter concentrations vary across different locations.

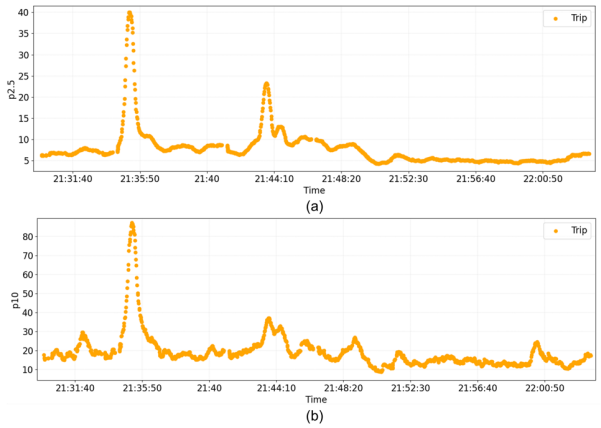

Figure 10: PM10 and PM2.5 Concentrations During a Journey, (a) the PM2.5 concentration levels are shown, and (b) the PM10 concentration levels are displayed over time.

Figure 10 illustrates the concentrations of PM10 and PM2.5 particulate matter during a specific journey, with time on the x-axis and concentration levels on the y-axis. Graph (a) shows the PM2.5 concentrations, while graph (b) depicts the PM10 concentrations. Both graphs reveal notable spikes in particulate matter levels at specific times during the journey, particularly around 21:35:50 and 21:44:10, indicating periods of higher pollution exposure. These spikes could be attributed to passing through areas with increased emissions, such as busy intersections or industrial zones. After these peaks, the concentrations gradually decrease, suggesting a return to areas with cleaner air. The data highlights the variability of air quality encountered during travel and underscores the importance of monitoring particulate matter to identify pollution hotspots and assess exposure risks.

4.2.3 PM10 Analysis

Analyzing the concentration of particulate matter PM10, as illustrated in Figure 11, provides valuable insights into the distribution and variability of PM10 levels in the observed environment. The histogram reveals the frequency of different PM10 concentration ranges, helping to identify common and extreme values. According to Table 6, the distribution of PM10 concentrations has a lower limit of 1.4 µg/m³, with the first quartile at 11.3 µg/m³, indicating that 25% of the observations fall below this value. The median concentration is 16.9 µg/m³, meaning half of the observations are below this level. The third quartile is 27.3 µg/m³, showing that 75% of the data points are below this concentration. The highest observed regular concentration is the upper limit, 52.5 µg/m³. Notably, the dataset includes a maximum outlier of 261.1 µg/m³, which indicates a significant pollution event or an anomaly in the data collection process. These statistics help understand the typical PM10 exposure levels and the potential for extreme pollution events, which are crucial for assessing air quality and formulating environmental policies.

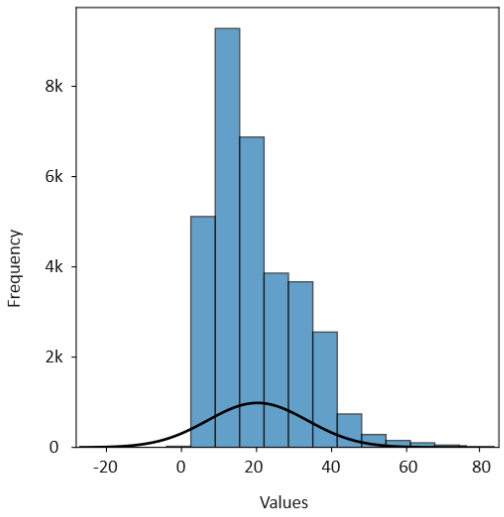

Figure 11: Histogram of PM10 Concentrations, displaying the frequency distribution of PM10 values.

Figure 11 shows the mode of PM10 concentration at 14 ($\mu g/m^3$), the most frequent value in the dataset. The mean concentration is 20.25 ($\mu g/m^3$) with a confidence interval of $\pm$13.284 ($\mu g/m^3$). These values fall within

the Acceptable Air Quality Guidelines (AQG) threshold, indicating that the recorded PM10 levels do not pose significant health risks to citizens.

4.2.4 PM2.5 Analysis

Figure 12 indicates that the most common PM2.5 concentration during data collection is 4 ($\mu g/m^3$). The average concentration is 11.84 ($\mu g/m^3$) with a margin of ± 8.111 ($\mu g/m^3$). These figures are within AQG limits, indicating no significant health risks for the population. Figure 12 illustrates a histogram of PM2.5 concentrations, visually representing the distribution and frequency of PM2.5 values within the dataset. The histogram reveals a skewed distribution, with most PM2.5 concentrations clustered between 0 and 20 µg/m³. The highest frequency is observed around the lower concentration levels, indicating that most recorded PM2.5 values are relatively low. There are fewer instances of higher PM2.5 concentrations, with a gradual decline in frequency as the values increase, and only a few observations exceed 40 µg/m³. This skewed distribution suggests that while the average PM2.5 levels are generally within a manageable range, occasional spikes could pose significant health risks. These higher values highlight the need for continuous monitoring and targeted interventions to mitigate the impact of elevated PM2.5 levels on public health.

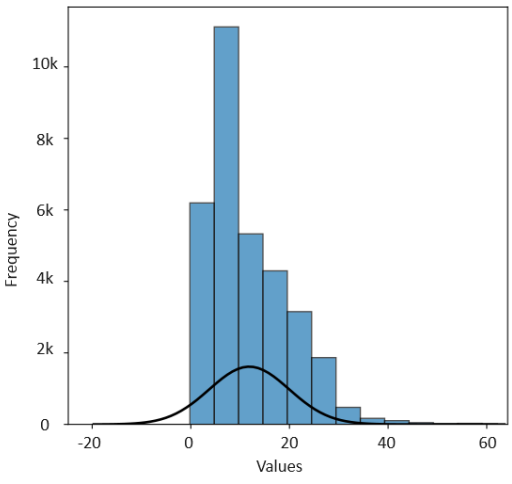

Figure 12: Histogram of PM2.5 Concentrations with the frequency distribution of PM2.5 values.

PM10 and PM2.5 particles are released from various sources, including exhaust from factories, vehicles, and road dust. These fine particles are significant pollutants that pose health risks, especially when penetrating deep into the respiratory system. The SDS011 sensor, a high-precision device with a measurement range of 0.0-999.9 µg/m³, was employed to monitor these particulate matter levels.

During the monitoring period, the sensor recorded peak values of 261 µg/m³ for PM10 and 78.8 µg/m³ for PM2.5. Such high concentrations, although rarely observed, indicate severe pollution episodes. These peak values were specifically measured on the highway, where vehicle emissions and road dust are predominant sources of particulate matter. The spatial distribution of these readings, as illustrated in Figure 13, highlights the critical points along the highway where particulate matter concentrations were exceptionally high.

These findings underscore the importance of continuous air quality monitoring, especially in high-traffic areas, to identify pollution hotspots and implement effective mitigation strategies. The data collected can inform policy decisions and contribute to developing regulations to reduce emissions from industrial and vehicular sources, ultimately protecting public health and improving urban air quality.

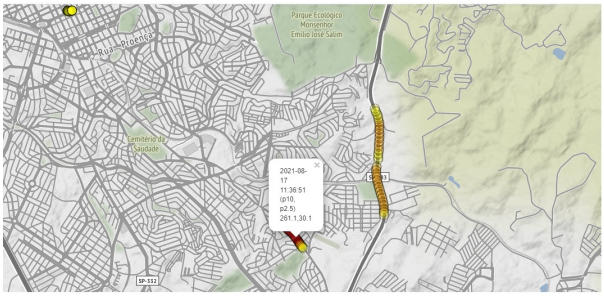

Figure 13: Geographic Visualization of Maximum PM10 Values, highlighting the locations with the highest recorded PM10 concentrations.

Figure 13 provides a geographic visualization of maximum PM10 values recorded in a specific urban area. The map highlights various locations where PM10 concentrations peaked, with color-coded markers indicating the severity of the readings. One notable point shows a maximum PM10 value of 261.1 µg/m³ recorded on August 17, 2021, at 11:30:51, significantly exceeding typical safety thresholds and suggesting a severe pollution event. This spatial representation identifies pollution hotspots and helps understand the spatial distribution of air quality issues. Such data is crucial for urban planners and environmental authorities to develop targeted strategies for pollution mitigation, enhance air quality management, and protect public health in the affected areas.

4.3 Toxic Gases (NH3, CO, NO2)

Certain gases naturally occur in soil, air, and water. Nitrogen exists in the atmosphere as gaseous N2, and one way to produce ammonia is through nitrogen-fixing microorganisms that capture atmospheric nitrogen and convert it into ammonia. Another method involves the bacterial conversion of nitrogenous compounds from dead organisms or their waste into ammonia. Ammonia is corrosive to the skin, eyes, and lungs, even at low concentrations in the air. Exposure to anhydrous ammonia can cause skin and eye burns, severe sore throat, and coughing/wheezing.

Analyzing the concentration of ammonia (NH3) in parts per million (ppm) provides valuable insights. According to Figure 14, the most frequent concentration (mode) is 0.678 ppm. The mean concentration is 0.683 ppm with a margin of ±0.038 ppm, which falls within the interim target 1 and interim target 2 levels, indicating minimal health risks for the population.

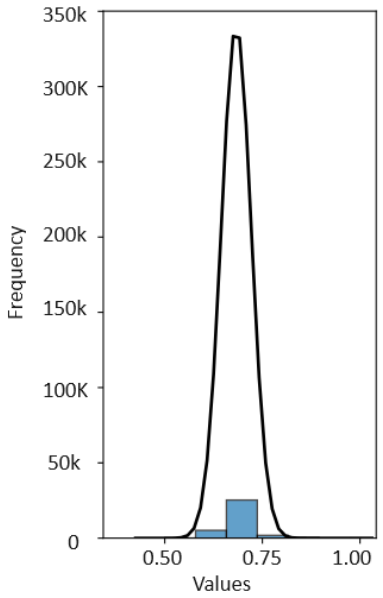

Figure 14: The Histogram of NH3 Concentrations with the frequency of ammonia (NH3) values.

Carbon monoxide (CO) is a byproduct of burning fuels like gas, oil, coal, and wood. It is also produced naturally during chlorophyll synthesis, plant decomposition, forest fires, and atmospheric methane oxidation. Inhaled CO reduces the blood's oxygen-carrying capacity, leading to oxygen deprivation in organs and tissues, causing cardiac and nervous system issues, headaches, dizziness, and fatigue, affecting humans and wildlife.

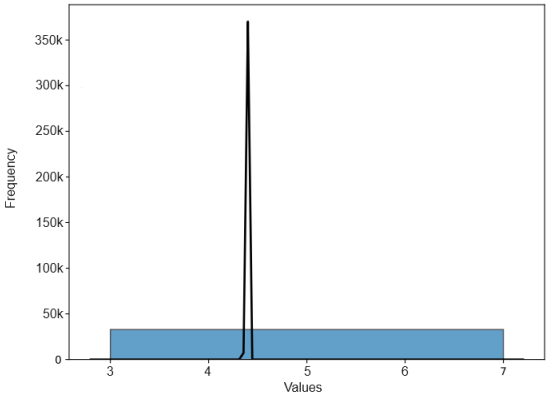

Figure 15: The Histogram of CO Concentrations with the frequency of carbon monoxide (CO) values.

Figure 15 presents the data collection results, showing a mode of 4.385 ppm and a mean of 4.385 ppm with a ±0.009 ppm range. These values exceed Air Quality Guideline (AQG) standards, posing significant health risks to citizens.

Nitrogen dioxide (NO2) is produced through the oxidation of atmospheric nitrogen, particularly at high temperatures. Its primary source is the intense traffic on major highways, which generates significant combustion processes.

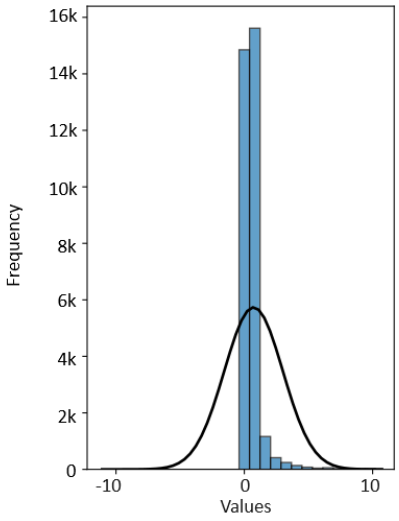

Figure 16: The Histogram of NO2 Concentrations displays the frequency distribution of nitrogen dioxide (NO2) values.

In examining nitrogen dioxide (NO2) concentrations measured in parts per million (ppm), Figure 16 reveals a mode of 0.63 ppm and a mean of 0.7191 ppm, with a range of ±2.283 ppm. These readings surpass Air Quality Guidelines (AQG), indicating substantial health hazards to citizens.

4.4 Humidex Index Analysis

To thoroughly analyze the Humidex Index, we apply Formula 1, which integrates the collected temperature and relative humidity data to calculate the index. This formula allows us to quantify the combined effects of heat and humidity on perceived temperature. The resulting Humidex values measure thermal comfort, which is crucial for assessing environmental conditions, particularly during warm weather.

In Figure 17, the distribution of Humidex values is visualized, revealing a mode of 13.5, indicating that this value appears most frequently in our dataset. The mean Humidex value is calculated at 14.91781, suggesting that, on average, the perceived temperature is slightly higher than the mode. The standard deviation, with a range of ±3.19068, indicates the variability in the data, reflecting fluctuations in the combined heat and humidity levels.

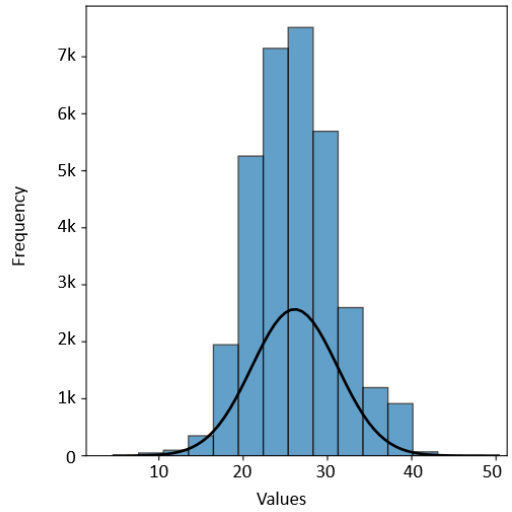

Figure 17: Histogram of Humidex Index, illustrating the frequency distribution of Humidex values. The histogram shows a roughly normal distribution, with most Humidex values clustered around the central range of 25 to 35.

Referring to Table 4, we observe that the Humidex values consistently remain below 29 throughout the dataset. According to standard interpretation guidelines, Humidex values under 29 are typically associated with a comfortable thermal sensation, where most individuals do not experience significant heat-related discomfort. This observation suggests that the environmental conditions during the data collection were generally favorable, with little heat stress or discomfort risk. The analysis underscores the importance of monitoring the Humidex Index in various applications, including weather forecasting, public health assessments, and climate studies, to maintain thermal comfort and mitigate potential heat-related hazards.

The analyses of key environmental variables indicated that while average values mostly stayed within safe levels, extreme values occasionally breached AQG thresholds, potentially posing health risks. This comprehensive data exploration sets the foundation for developing predictive models and more refined interventions to enhance urban air quality monitoring and management.

4.5 Multivariable Analysis

To analyze variables with different units of measurement, we standardize them using the formula:

$$Z = \frac{\text{Raw Score} - \text{Mean}}{\text{Standard Deviation}}$$

Table 7 presents the variance-covariance matrix concerning the principal components (PC). Each *eigenvalue* represents the variance a principal component explains. For instance, the eigenvalue for the first principal component (PC1) is 2.165, indicating that it accounts for 2.165 units of variance.

Table 7: Matrix Variance - Covariance

PC	Eigenvalue	% Variance
1	2.165	30.924
2	1.642	23.452
3	1.062	15.175
4	1.000	14.286
5	0.904	12.914
6	0.168	2.402
7	0.059	0.848

The "% variance" column indicates the percentage of total variance explained by each principal component. For example, PC1 explains 30.924% of the total variance. This metric helps understand each principal component's importance in the overall data structure. The first two principal components explain approximately 54.376% of the total variance. This highlights their significance in capturing the main variability in the data. Table 8 and Figure 18 analyze

axes 1 and 2, representing components 1 and 2. In PC1, Humidex (0.543) and Temperature (0.534) are the most significant variables critical for evaluating air quality and influencing comfort in the study area.

Table 8: Loadings

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
CO	-0.002	0.0046	-0.0021	0.999	0.016	-0.007	-0.009
NH3	0.062	0.19	0.643	0.012	-0.739	0.033	-0.004
NO2	-0.063	0.063	0.746	-0.009	0.66	-0.004	-0.004
P 10	0.451	0.526	-0.104	-0.003	0.114	0.697	-0.103
P 2.5	0.455	0.528	-0.076	-0.003	0.076	-0.699	0.117
Temp.	0.534	-0.455	0.086	0.004	0.003	0.112	0.697
Humidex	0.543	-0.445	0.08	0.003	-0.001	-0.108	-0.67

Humidex, a composite index that combines temperature and relative humidity, is a crucial indicator of thermal comfort and plays a significant role in assessing the environmental conditions that impact human well-being. By factoring in temperature and humidity, the Humidex provides a more comprehensive measure of perceived heat, particularly in regions where high humidity can exacerbate the effects of warm temperatures. A higher Humidex value corresponds to a stronger sensation of heat, which can significantly influence an individual's comfort levels, potentially leading to discomfort, decreased productivity, and even health risks such as heat exhaustion or heatstroke.

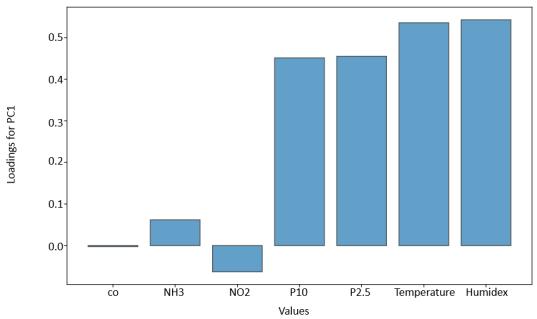

Figure 18: Loadings of Each Variable for Principal Component 1 (PC1), illustrating the contribution of each variable to PC1. The bar chart shows that temperature and Humidex have the highest loadings, indicating they are the most influential variables in explaining the variance captured by PC1.

In the context of the study area, both Humidex and temperature are closely intertwined with the quality of life. These indicators reflect the environmental conditions that the local population experiences daily. High Humidex values can diminish the quality of life by creating oppressive heat conditions that limit outdoor activities, reduce work capacity, and increase the need for cooling measures, affecting both physical and mental well-being. On the other hand, maintaining moderate temperature levels contributes to a more favorable living environment, enhancing comfort, productivity, and overall satisfaction with daily life.

Thus, the interplay between Humidex and temperature is not merely a matter of meteorological interest but a critical factor in public health, urban planning, and community well-being. Understanding and managing these variables can lead to better-informed decisions that improve living conditions, mitigate the adverse effects of extreme weather, and ultimately enhance the quality of life for the population in the study area.

4.6 Clustering and Data Decomposition Analysis

To apply the DBSCAN algorithm, it is crucial first to determine the values of *epsilon* and *min_samples* for clustering. Given the number of records, we decided to automate the search for these values using the *KneeLocator* function from the *kneed* library. Additionally, we defined a simple function to set *min_samples* based on the dataset size.

The automated process yielded the following results:

- Automatically determined *epsilon* value: 0.0021286850401138644
- Automatically determined *min_samples* value: 10

With the above parameters, we obtained 26 clusters, ranging from cluster 0 to cluster 25. We will analyze clusters 0, 13, and 25.

4.6.1 Cluster 0 Analysis

The analysis of Cluster 0 begins with a focus on the loadings of variables within Cluster 0, as shown in Table 9. These loadings indicate the variables' influence on each principal component, offering insights into the data's structure and highlighting the most significant variables contributing to the cluster's formation.

Table 9: Loadings of Cluster 0

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
CO	-0.002	0.007	-0.06	0.99	-0.009	-0.006	-0.009
NH3	0.088	0.24	0.498	0.036	0.828	0.04	0.005
NO2	-0.073	0.123	0.833	0.044	-0.531	-0.001	0.001
P 10	0.559	0.401	-0.103	-0.009	-0.146	0.693	0.118
P 2.5	0.571	0.394	-0.076	-0.007	-0.098	-0.698	-0.13
Temp.	0.416	-0.56	0.144	0.014	0.029	0.134	-0.688
Humidex	0.42	-0.545	0.131	0.012	0.036	-0.114	0.704

In PC1, variables such as "P 2.5" and "P 10" exhibit high loadings, indicating their significant contributions to the variability captured by PC1. This is visualized in Figure 19.

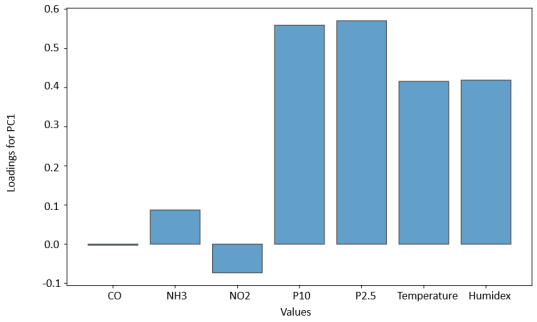

Figure 19: Loadings of Each Variable for Principal Component 1 (PC1) for Cluster 0, highlighting the influence of each variable within this specific cluster. The bar chart reveals that PM10 variations and PM2.5 have the highest loadings, indicating they are the most significant contributors to the variance captured by PC1 in Cluster 0.

These findings highlight the complex interplay of environmental factors shaping air quality dynamics within Cluster 0. Recommendations address the identified health risks associated with high concentrations of pollutants, emphasizing the need for stringent regulatory measures and increased public awareness. Implementing these measures can help mitigate the adverse health effects and improve air quality.

4.6.2 Cluster 13 Analysis

The PCA results in Table 10 offer key insights into the data structure of Cluster 13. Although the analysis spanned 26 clusters, our focus on Cluster 13 emphasizes the variables most significantly influencing its data. PM10, with an absolute loading of -0.549, emerged as the most influential variable on PC1, indicating its substantial contribution to the variability within this cluster. Other variables, such as PM2.5 and NO2, also displayed significant loadings on PC1, underscoring their role in explaining the variance in Cluster 13.

Table 10: Loadings of Cluster 13

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
CO	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	1.0
NH3	0.138	0.004	-0.99	-0.01	-0.033	0.06	0.0
NO2	-0.497	0.218	-0.055	-0.832	-0.094	0.037	0.00
P 10	-0.549	0.187	-0.057	0.457	-0.672	0.015	0.0
P 2.5	-0.545	0.231	-0.098	0.313	0.732	0.074	0.0
Temp.	-0.217	-0.682	0.01	-0.017	-0.01	0.7	0.0
Humidex	-0.32	-0.632	-0.09	-0.019	0.046	-0.71	0.0

Figure 20 visually depicts the variable loadings for Principal Component 1 (PC1) within Cluster 13. This plot clearly illustrates each variable's contribution to PC1, simplifying the interpretation of the PCA results. By identifying the variables that most influence the variance captured by PC1, the visualization provides deeper insights into the underlying structure of the data.

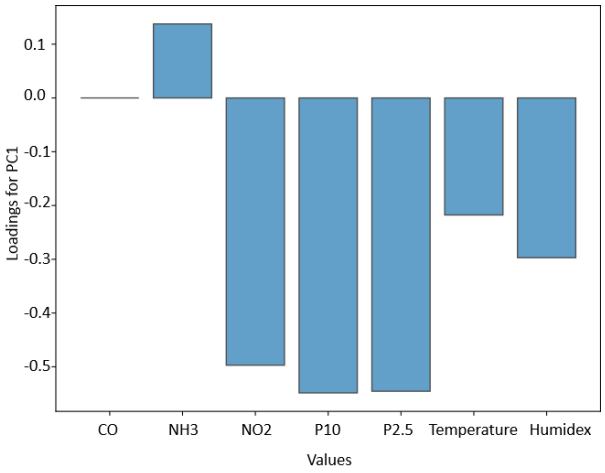

Figure 20: Loadings of Each Variable for Principal Component 1 (PC1) for Cluster 13 illustrate the contribution of each variable within this specific cluster.

The recorded concentrations of PM10 and PM2.5 in Cluster 13 were within the moderate range, indicating potential health risks. To protect public health, it is essential to implement measures to reduce particulate matter exposure. Continuous air quality monitoring and pollution control are also crucial, as weather conditions significantly impact pollutant dispersion.

In conclusion, tackling air pollution and promoting sustainable practices are vital to safeguarding public health and improving environmental quality in Cluster 13. Implementing pollution control measures and sustainable urban development policies will help mitigate adverse health effects and ensure the local population's well-being.

4.6.3 Cluster 25 Analysis

Table 11 shows the loadings of variables on PC1 and PC2, with P10 and Humidex being the most significant contributors. P10 has the highest loading on PC1, strongly influencing this component, while Humidex has a substantial loading on PC2, indicating its impact on the second component. These results emphasize the importance of P10 and Humidex in the overall variance structure captured by the first two principal components.

Table 11: Loadings of Cluster 25

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
CO	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	1.0
NH3	0.499	0.031	-0.295	-0.633	0.382	-0.341	0.000
NO2	0.202	-0.405	0.776	-0.387	-0.189	0.094	0.000
P 10	-0.549	0.056	-0.046	-0.361	-0.454	-0.598	0.000
P 2.5	-0.519	-0.166	-0.213	-0.496	0.239	0.595	0.000
Temp.	-0.290	0.541	0.514	-0.008	0.578	-0.157	0.000
Humidex	0.235	0.715	0.017	-0.273	-0.470	0.371	0.000

Figure 21 visually represents the variable loadings for the first principal component, PC1. The figure demonstrates the substantial influence of P10 on this component, with P10 exhibiting the highest loading among all the variables analyzed. This high loading indicates that P10 is a key factor driving the variance captured by PC1, highlighting its critical role in shaping the underlying data structure associated with this principal component.

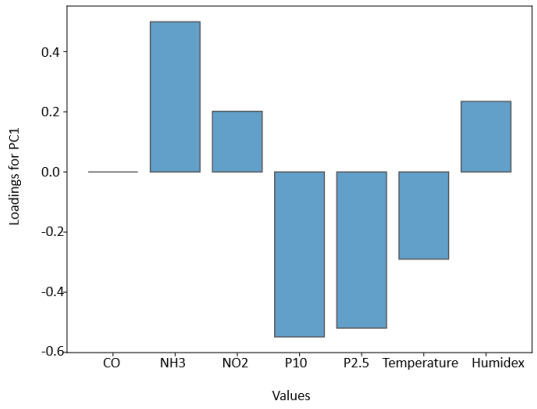

Figure 21: Loadings of Each Variable for Principal Component 1 (PC1) for Cluster 25, displaying the influence of each variable within this cluster. The bar chart reveals that CO and NH3 have significant positive loadings, indicating a strong positive contribution to PC1.

Regarding health implications, the high loading of PM10 on PC1 suggests significant contributions to variability, which is associated with respiratory and cardiovascular risks. Similarly, the high loading of Humidex on PC2 indicates potential discomfort and heat stress, affecting overall well-being. These findings underscore the importance of addressing particulate matter pollution and thermal comfort to mitigate adverse health effects in Cluster 13.

The maximum values of humidity (21.5), temperature (20°C), PM10 (19.1 µg/m³), and PM2.5 (5.8 µg/m³) in Cluster 25 reveal important insights. The humidity level is within the "Mild discomfort" range on the Humidex scale (Table 4), while the temperature is considered comfortable. However, the PM10 and PM2.5 levels exceed the World Health Organization's interim targets, indicating potential health risks.

Although the concentrations of PM10 and PM2.5 in Cluster 25 are below the interim targets set by the World Health Organization (WHO) for air quality standards (Table 3), they still pose potential health risks. The interim target for PM10 is 150 µg/m³, and the observed value is 19.1 µg/m³. PM2.5's target is 75 µg/m³, while the recorded level is 5.8 µg/m³.

High levels of particulate matter like PM10 and PM2.5 can adversely affect health, as they can penetrate deeply into the respiratory system, leading to respiratory and cardiovascular diseases and exacerbating conditions such as asthma and bronchitis.

To address these health risks, it is crucial to implement measures to reduce air pollution. This includes promoting cleaner energy sources, enhancing emission controls, adopting stricter vehicle emission standards, and encouraging sustainable urban planning. Public health initiatives are also essential, such as raising awareness about air quality and encouraging personal protective measures during periods of poor air quality.

In summary, the elevated PM10 and PM2.5 levels in Cluster 25 highlight the need for effective air quality management strategies. By addressing air pollution and promoting sustainable practices, we can create healthier and more livable environments for the residents of Cluster 25 and beyond.

Clusters 0, 13, and 25 analysis reveals differences in air pollution levels and related health risks. Clusters 0 and 13 show moderate pollution, whereas Cluster 25 has significantly higher particulate matter concentrations, posing serious health concerns. Improving air quality involves regulatory measures, public education, and community involvement.

To safeguard public health, it is crucial to prioritize air quality management and implement interventions recommended by organizations like the WHO. Collaborative efforts across various sectors and levels—local, national, and international—are necessary to create healthier environments and ensure the well-being of future generations.

4.7 Comparative Analysis

Our review of existing research identified several Internet of Things (IoT)-based systems designed to monitor urban air quality. These studies emphasize sensor deployment strategies, data collection approaches, and analysis methodologies. However, our work stands out through the comprehensive integration

of Node-RED for efficient data flow visualization and analysis alongside InfluxDB for real-time data management.

Our methodology differentiates itself by not merely focusing on individual components or specific types of sensors. Instead, we integrate multiple technological elements into a cohesive IoT architecture tailored to monitor urban environments. This integration significantly enhances data accuracy and system dependability. Moreover, our system facilitates advanced analytical methods, such as Knowledge Discovery in Databases (KDD), enabling sophisticated pattern recognition and anomaly detection.

Our results underscore the system's robustness in handling diverse environmental variables. For instance, the SDS011 sensor recorded peak PM10 and PM2.5 values of 261 µg/m³ and 78.8 µg/m³, respectively, highlighting its sensitivity and accuracy in capturing critical pollution levels. Using clustering algorithms like DBSCAN further refined our analysis, identifying high-pollution areas that require targeted interventions.

We achieved a more precise air quality assessment by deploying sensors in strategic locations, such as highways and industrial neighborhoods, aligning with studies emphasizing location's significance in environmental monitoring. However, our integration of real-time data processing and advanced visualization tools provides a more dynamic and responsive approach than the more static methods observed in other research.

Our work contributes to the field by demonstrating a scalable, reliable, and comprehensive IoT-based system that can enhance urban air quality monitoring and management, offering actionable insights for urban planners and public health officials.

4.8 Practical Implications

Deployment in various locations, including vehicles and public spaces, enhances the adaptability of this system, enabling real-time monitoring of climate and air quality in places frequented by large numbers of people, such as restaurants and sporting events. This capability is crucial for safeguarding public health by promptly identifying environmental factors that could impact participants' well-being.

By leveraging this mobile dataset, our approach expands environmental monitoring beyond fixed locations, significantly advancing over traditional stationary systems. This methodology enhances the granularity and coverage of data collection and facilitates flexible decision-making for health authorities, urban planners, and event organizers. It provides stakeholders with actionable insights to mitigate health risks associated with fluctuating environmental conditions in densely populated areas.

Furthermore, our system's capability to proactively manage environmental quality in dynamic urban settings is underscored by its advanced data analytics, including real-time anomaly detection and predictive modeling. These capabilities contribute to the discourse on smart cities and sustainable urban development by emphasizing the importance of data-driven strategies in enhancing environmental resilience and public health outcomes.

5 Conclusion

This research provides critical insights into modern cities' environmental and climate challenges, focusing on the smart city concept aimed at enhancing urban quality of life. It emphasizes the development of an intelligent system for real-time environmental monitoring, which is crucial for informed decision-making regarding urban environmental safety.

The study explores various aspects of environmental comfort, specifically creating a toxic gas monitoring system using Arduino and integrated sensors. It addresses key research questions about the factors influencing environmental comfort and the existing technological solutions to foster a safe, clean, and healthy urban environment within smart city frameworks.

To achieve these objectives, the research methodology includes hardware implementation for data collection and establishing an IoT system architecture, enabling comprehensive data analysis. Techniques such as Knowledge Discovery in Data (KDD), univariate and multivariate analyses, and algorithms like DBSCAN are employed to provide a detailed understanding of air quality and its public health implications.

Moreover, key findings highlight the significance of Humidex, temperature, and particulate matter concentrations in evaluating environmental comfort and air quality. Clustering and data decomposition analyses reveal areas with varying pollution levels and associated health risks, emphasizing the need for targeted interventions to safeguard residents' health.

Furthermore, the study recommends integrating the IoT architecture into central smart city monitoring systems to improve real-time data accuracy and enhance responses to air quality warnings. Given that increased vehicle numbers are a primary pollution source, deploying the dataset in strategic locations such as industrial neighborhoods and main avenues is suggested. This approach would enable precise air quality assessments and identification of health-impacting factors. Encouraging drivers to use devices

that record real-time data can help maintain an up-to-date and comprehensive database.

In conclusion, this research underscores the urgency of managing air quality and implementing evidence-based interventions recommended by global health organizations like the WHO. Collaboration among stakeholders at various levels is essential to create healthier and more sustainable urban environments for current and future generations.

Building on the findings of this research, future work could focus on enhancing the IoT system's capabilities by integrating machine learning algorithms for predictive analytics and real-time response automation. This includes developing advanced models to accurately forecast air quality trends and identify potential pollution sources.

Additionally, expanding the network of sensors to cover more diverse geographic areas, including rural and suburban regions, would provide a more comprehensive understanding of air quality dynamics. Future studies could also explore the impact of different urban planning strategies on environmental comfort and health outcomes.

Engaging in cross-disciplinary collaborations with urban planners, public health officials, and policymakers will be essential to design effective interventions and policies that promote sustainable urban development. Finally, investigating the social and economic implications of implementing these smart city technologies could offer valuable insights into optimizing resource allocation and maximizing public benefit.

REFERENCES

[1] S. Abdullah, M. Ismail, A. N. Ahmed, and W. N. W. Mansor, "Big data analytics and artificial intelligence in air pollution studies for the prediction of particulate matter concentration," in *Proceedings of the 3rd International Conference on Telecommunications and Communication Engineering*, 2019, pp. 90–94.

[2] A. Alelaiwi, "Evaluating distributed iot databases for edge/cloud platforms using the analytic hierarchy process," *Journal of Parallel and Distributed Computing*, vol. 124, pp. 41–46, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0743731518307603

[3] M. A. Arbex, U. d. P. Santos, L. C. Martins, P. H. N. Saldiva, L. A. A. Pereira, A. L. F. Braga *et al.*, "A poluição do ar e o sistema respiratório," *Jornal Brasileiro de Pneumologia*, 2012.

[4] J. F. A. Arias and L. Mendes, "Conceptual theoretical approach about smart cities," in *2017 IEEE First Summer School on Smart Cities (S3C)*, Aug 2017, pp. 132–136.

[5] A. H. F. Ashrae and G. Atlanta, "American society of heating," *Refrigerating and Air-Conditioning Engineers*, 2009.

[6] C. Asiminidis, G. Kokkonis, and S. Kontogiannis, "Database systems performance evaluation for iot applications," *International Journal of Database Management Systems (IJDMS) Vol*, vol. 10, 2018.

[7] M. Bauer, N. Bui, J. De Loof, C. Magerkurth, A. Nettsträter, J. Stefa, and J. W. Walewski, "IoT Reference Model," in *Enabling Things to Talk*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 113–162.

[8] P. I. D. BELO, "Quantificação dos níveis de partículas finas (mp2, 5) no município de vitória," *Trabalho de Conclusão de Curso de Graduação em Engenharia Ambiental, Universidade Federal do Espírito Santo, Vitória/ES*, 2011.

[9] R. M. Betancourt, B. Galvis, S. Balachandran, J. Ramos-Bonilla, O. Sarmiento, S. Gallo-Murcia, and Y. Contreras, "Exposure to fine particulate, black carbon, and particle number concentration in transportation microenvironments," *Atmospheric environment*, vol. 157, pp. 135–145, 2017.

[10] B. Blocken and J. Carmeliet, "Pedestrian wind environment around buildings: Literature review and practical examples," *Journal of Thermal Envelope and Building Science*, vol. 28, no. 2, pp. 107–159, 2004.

[11] M. Brauer, M. Amann, R. T. Burnett, A. Cohen, F. Dentener, M. Ezzati, S. B. Henderson, M. Krzyzanowski, R. V. Martin, R. Van Dingenen *et al.*, "Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution," *Environmental science & technology*, vol. 46, no. 2, pp. 652–660, 2012.

[12] G. Castro, J. Myers, and R. Ricklefs, "Ecology and energetics of sandlerlings migrating to four latitudes," *Ecology*, vol. 73, no. 3, pp. 833–844, 1992.

[13] T. J. Chandler, *The climate of London*. Hutchinson, 1965.

[14] J. Clarke and W. Bach, "Comparison of the comfort conditions in different urban and suburban microenvironments," *International journal of biometeorology*, vol. 15, no. 1, pp. 41–54, 1971.

[15] A. Cooper, R. Reimann, D. Cronin, and C. Noessel, *About face : the essentials of interaction design*, 4th ed. John Wiley & Sons, 2014.

[16] T. A. de Almeida, F. de Franco Rosa, and R. Bonacin, "Designing data visualization dashboards to support the prediction of congenital anomalies," in *International Conference on Human-Computer Interaction*. Springer, 2021, pp. 143–162.

[17] L. G. I. Domiguez, E. J. H. Alonso, J. A. C. Guarnizo, J. A. Carrillo, and J. A. V. Guativa, "Monitoreo de material particulado pm10 y pm2.5 en la ciudad de villavicencio," in *2019 Congreso Colombiano y Conferencia Internacional de Calidad de Aire y Salud Pública (CASP)*, 2019, pp. 1–5.

[18] F. C. Drumm, A. E. Gerhardt, G. D. Fernandes, P. Chagas, M. S. Sucolotti, and P. Kemerich, "Poluição atmosférica proveniente da queima de combustíveis derivados do petróleo em veículos automotores," *Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental*, vol. 18, no. 1, pp. 66–78, 2014.

[19] Y. Epstein and D. S. Moran, "Thermal comfort and the heat stress indices," *Industrial health*, vol. 44, no. 3, pp. 388–398, 2006.

[20] S. Fang, L. D. Xu, Y. Zhu, J. Ahati, H. Pei, J. Yan, and Z. Liu, "An integrated system for regional environmental monitoring and management based on internet of things," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1596–1605, May 2014.

[21] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI magazine*, vol. 17, no. 3, pp. 37–37, 1996.

[22] S. Friendlander, "Smoke, dust and haze: Fundamentals of aerosol dynamics," 2000.

[23] J. L. Geirinhas, R. M. Trigo, R. Libonati, C. A. Coelho, and A. C. Palmeira, "Climatic and synoptic characterization of heat waves in brazil," *International Journal of Climatology*, vol. 38, no. 4, pp. 1760–1776, 2018.

[24] B. Givoni, *Climate considerations in building and urban design*. John Wiley & Sons, 1998.

[25] J. S. Guzmán, H. T. Hernández, J. M. Mora, J. E. P. Johan, and S. Vanegas, "Análisis espacio-temporal de concentraciones de material particulado en bogotá: Un año de operación de una red independiente: Spatial and temporal analysis of particulate material concentrations in bogota: A year of operation of an independent network," in *2019 Congreso Colombiano y Conferencia Internacional de Calidad de Aire y Salud Pública (CASP)*, 2019, pp. 1–7.

[26] R. Haddad and J. J. Bloomfield, "La contaminacion atmosferica en america latina," 1964.

[27] Y. Hao, X. Qin, Y. Chen, Y. Li, X. Sun, Y. Tao, X. Zhang, and X. Du, "Ts-benchmark: A benchmark for time series databases," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, 2021, pp. 588–599.

[28] Z. Hu, H. Liu, X. Wang, Y. Dang, Z. Wang, M. Yu, and M. Zhang, "Remote sensing and gis application in environmental effect monitoring of great engineering: Indicators and method," in *2010 18th International Conference on Geoinformatics*. IEEE, 2010, pp. 1–5.

[29] InfluxData, "Influxdb documentation," 2024, accessed: 2024-08-14. [Online]. Available: https://docs.influxdata.com/influxdb/v2.0/

[30] H. Jamal, M. Huzaifa, M. A. Sodunke, and J. O. Odiete, "Smart heat stress and toxic gases monitoring instrument with a developed graphical user interface using iot," in *2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*. IEEE, 2019, pp. 1–6.

[31] P. John, J. Hynek, T. Hruška, and M. Valný, "Application of time series database for iot smart city platform," in *2023 Smart City Symposium Prague (SCSP)*, 2023, pp. 1–6.

[32] P. Kratzer, "Das stadtklima, 143 pp," *Friedrich Vieweg, Wiesbaden, Germany*, 1937.

[33] P. J. Landrigan, R. Fuller, N. J. Acosta, O. Adeyi, R. Arnold, A. B. Baldé, R. Bertollini, S. Bose-O'Reilly, J. I. Boufford, P. N. Breysse *et al.*, "The lancet commission on pollution and health," *The Lancet*, vol. 391, no. 10119, pp. 462–512, 2018.

[34] A.-M. Lovin, Ş. Silion, and A.-M. Nicuţa, "Thermal comfort specific conditions in vehicles and real-time calculation of pmw and ppd thermal comfort indices," in *2013 8TH International Symposium on Advanced Topics in Electrical Engineering (ATEE)*. IEEE, 2013, pp. 1–4.

[35] G. Luna, M. Andrés, V. Gonzalez, J. Mario, T. Muñoz, B. Ceron, L. Carlos *et al.*, "Evaluación espacial y temporal de pm 10 y pm 2.5 en colombiautilizando información satelital (cams, modis-aod) y mediciones de calidad del aire en superficie [not available in english]," in *2019 Congreso Colombiano y Conferencia Internacional*

de Calidad de Aire y Salud Pública (CASP). IEEE, 2019, pp. 1–5.

[36] J. Masterton and F. Richardson, "A method of quantifying human discomfort due to excessive heat and humidity. report cli 1-79," *Atmospheric Environment Service, Environment Canada, Downsview*, 1979.

[37] P. Montes-Muñoz, C. Zafra-Mejía, and J. Pachón-Quinche, "Contaminación atmosférica por material particulado resuspendido: factores físicos condicionantes en áreas urbanas," in *2019 Congreso Colombiano y Conferencia Internacional de Calidad de Aire y Salud Pública (CASP)*, 2019, pp. 1–4.

[38] S. N. Z. Naqvi, S. Yfantidou, and E. Zimányi, "Time series databases and influxdb," *Studienarbeit, Université Libre de Bruxelles*, vol. 12, pp. 1–44, 2017.

[39] K. Pawar and V. Attar, "A survey on data analytic platforms for internet of things," in *2016 International Conference on Computing, Analytics and Security Trends (CAST)*, 2016, pp. 605–610.

[40] A. J. Pérez Cueva, F. Gómez Lopera, and J. Tornero, "Ciudad y confort ambiental: estado de la cuestión y aportaciones recientes," 2006.

[41] I. A. Raja and G. S. Virk, "Thermal comfort in urban open spaces: a review," *Proceedings of moving thermal comfort standards into the 21st century*, pp. 342–52, 2001.

[42] S. Russo, A. Dosio, R. G. Graversen, J. Sillmann, H. Carrao, M. B. Dunbar, A. Singleton, P. Montagna, P. Barbola, and J. V. Vogt, "Magnitude of extreme heat waves in present climate and their projection in a warming world," *Journal of Geophysical Research: Atmospheres*, vol. 119, no. 22, pp. 12–500, 2014.

[43] J. H. Seinfeld, S. N. Pandis, and K. Noone, "Atmospheric chemistry and physics: from air pollution to climate change," *Physics Today*, vol. 51, p. 88, 1998.

[44] M. Stoimenova-Minova, S. Gocheva-Ilieva, and A. Ivanov, "Pm10 prediction using cart method depending on the number of observations," in *Proceedings of the 2020 3rd International Conference on Mathematics and Statistics*, 2020, pp. 65–70.

[45] M. Tayab, W. Zhou, M. Zhao, and S. Li, "Big data and public services for environmental monitoring system," in *2016 11th International Conference on Computer Science & Education (ICCSE)*. IEEE, 2016, pp. 139–143.

[46] G. V. Ulate, "Espacio y territorio en el análisis geográfico," *Reflexiones*, vol. 91, no. 1, pp. 313–326, 2012.

[47] J. Unger, "Comparisons of urban and rural bioclimatological conditions in the case of a central-european city," *International Journal of Biometeorology*, vol. 43, no. 3, pp. 139–144, 1999.

[48] B. Warnke, J. Mantler, S. Groppe, Y. C. Sehgelmeble, and S. Fischer, "A sparql benchmark for distributed databases in iot environments," in *Proceedings of the International Workshop on Big Data in Emergent Distributed Environments*, ser. BiDEDE '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3530050.3532929

[49] J. Yoo, D. Shin, and D. Shin, "Prediction system for fine particulate matter concentration index by meteorological and air pollution material factors based on machine learning," in *Proceedings of the Tenth International Symposium on Information and Communication Technology*, 2019, pp. 479–485.

[50] J. Zhu, Y. Fu, Y. Xing, Y. Zhang, and Q. Qiao, "Multi-component toxic gas monitoring system based on internet of things," in *2019 IEEE International Conference on Mechatronics and Automation (ICMA)*. IEEE, 2019, pp. 769–773.

AUTHOR BIOGRAPHIES

Jaime Fabian Arias Aguilar is a Cyber Security Engineer at Foundever - Sophos LATAM in Brazil with extensive industry and academic experience. Formerly held roles in engineering and academia in Brazil and Ecuador. Currently pursuing a PhD in Telematics at UNICAMP, focusing on smart cities, IoT, and cybersecurity, with published work on smart cities and urban mobility.

Professor Dr. Everton Gomede is a Computer Scientist and Data Analyst with over 20 years of experience developing algorithmic processes and data-driven solutions. Specializing in machine learning, data science, and software development, he has a proven track record of enhancing business strategies and operational efficiency. His academic roles include positions at the University of British Columbia, University of São Paulo, FEEC/Unicamp, and others. He also serves as a Visiting Professor, teaching AI, machine learning, deep learning, and data science courses. Additionally, he is a Reviewer for several scientific journals and has authored multiple technical books and papers.

Professor Dr. Leonardo de Souza Mendes holds degrees in Electrical Engineering from Universidade Gama Filho (1984), Pontifícia Universidade Católica do Rio de Janeiro (1987), and Syracuse University (PhD, 1991). Currently a Full Professor at UNICAMP, he also serves as a project reviewer for the Fundação de Amparo à Pesquisa do Estado de São Paulo. His expertise includes Smart Cities, Smart Enterprises, and cognitive systems for Smart Grids. As the Scientific Director of the Communication Networks Laboratory (LaRCom) at UNICAMP, he has been pivotal in developing Municipal Infoways, Digital Cities, and Smart Cities since 2001. His research focuses on intelligent systems for urban applications, including Education, Health, and Public Administration.