Correlation between Air Pollution and NHS Emergency Hospital Admissions

By
Jabed Hossain 23057838
Aidin Miralmasi 23053472
Jan Michael Corpus 23044709
Ella Pournezhadian 23048129
Anita Lamichhane 23053864

SUPERVISOR
Xiaojun Luo

University of the West of England
Frenchay, Bristol
[September 2024]

ABSTRACT

This study investigates the correlation between air pollution and emergency hospital admissions in Bristol, UK, with a focus on particulate matter (PM) and nitrogen pollutants. Using data from the NHS and the UK-AIR database, we performed statistical analyses to understand how fluctuations in pollutant levels relate to spikes in emergency medical cases. Methodologies included data preprocessing, correlation matrices, and predictive modelling through imputation and regression techniques. Our findings indicate that PM pollutants have a significant impact on respiratory and cardiovascular admissions, underscoring the public health risks of urban air pollution. This research highlights the importance of environmental policies aimed at reducing pollutant exposure to improve health outcomes.

Key Words: [Air pollution, Emergency hospital admissions, Predictive modelling, Public health, Correlation].

Table of Contents

Table of Figures

Table of Tables

1. Introduction

Air pollution is a pressing global issue with implications for public health and environmental sustainability. The adverse effects of air pollution on human health, ranging from respiratory diseases to cardiovascular disorders, have been extensively documented in scientific literature. Understanding the complex interplay between air pollution levels and health outcomes is essential for informing evidence-based policies and interventions aimed at mitigating the health risks associated with poor air quality.

In light of these concerns, this project aims to investigate the correlation between air pollution and Emergency Hospital Admissions, with the overarching goal of explaining the relationship between environmental exposure and immidiate health outcomes. By leveraging advanced data analysis techniques and interdisciplinary approaches, we seek to uncover insights that can inform public health strategies, urban planning initiatives, and environmental regulations.

This introduction sets the stage for the subsequent sections of the project, which will dive into the methodology, data analysis, results, and implications of the study. Through a thorough investigation of the existing evidence and machine learning techniques and algorithms, we aim to explain the connection between air pollution and Emergency Hospital Admissions. Our goal is to employ rigorous data-driven methodologies to contribute to a better understanding of this complex relationship, ultimately working towards a healthier and more sustainable future for all.

Github link to project repository: IGP-Repo

1.1 Project Objectives

Primary Objectives:

The primary objective of this research is to **distinguish the correlation between specific air pollutants' concentration levels and the frequency of emergency hospital admissions** within Bristol. By examining these associations, we aim to provide insights into the immediate health impacts of air pollution exposure on vulnerable populations.

Secondary Objectives:

This study aims to explore the health effects of air pollution. Specifically, it will analyse the time interval between exposure to elevated levels of air pollutants and the subsequent need for emergency medical attention. Understanding the lagged effects of air pollution exposure on health outcomes is crucial for informing preventive measures and healthcare resource allocation. Through this investigation, the study seeks to contribute to a deeper understanding of the complex relationship between air pollution and human health, supporting evidence-based policies and interventions aimed at mitigating its adverse effects.

1.2 Litreture Review

1.

Katsouyanni, K., Schwartz, J., Spix, C., Touloumi, G., Zmirou, D., Zanobetti, A., Wojtyniak, B., Vonk, J. M., Tobias, A., Pönkä, A., Medina, S., Bachárová, L., & Anderson, H. R. (1996, April). **Short-term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol**. Journal of Epidemiology and Community Health, 50(Suppl 1), S12-S18.

https://doi.org/10.1136/jech.50.suppl_1.s12

The research article investigates the link between air pollution and the frequency of emergency hospital visits due to cardiovascular and respiratory ailments in London, UK. Employing a time-series analytical approach, the study scrutinizes the day-to-day fluctuations in air pollution and corresponding hospital admissions spanning a period of five years. The analysis reveals a noteworthy

association between elevated levels of specific air pollutants, notably sulphur dioxide and particulate matter, and a surge in hospital admissions for heart and lung diseases. These observations highlight the critical need for interventions aimed at improving air quality to alleviate the detrimental health impacts of pollution.

Data Collection: In terms of data collection, the study leverages the comprehensive APHEA database, which encompasses data from 15 European cities, each presenting unique socio-environmental profiles and air pollution challenges. The collected data comprises daily records of various air pollutants, including black smoke, sulfur dioxide, fine particulates, nitrogen dioxide, and ozone, alongside daily statistics of overall and specific causes of mortality and emergency hospital admissions. The study meticulously adjusts for potential confounding elements, such as weather conditions and time-related variables, to ensure the integrity of the findings.

Air Pollution Measurements: To assess the impact of air pollution on public health, data was meticulously gathered from public monitoring networks strategically positioned across each participating city. For a robust analysis of urban pollution, measurements were exclusively sourced from stations situated in areas with high traffic density or those representing general urban background levels. Ozone, a pollutant with distinct suburban concentration patterns, was an exception, where data from suburban stations were also considered. To ensure the consistency and comparability of data across different urban environments, average readings from these multiple stations were calculated.

Health Data: The study's health impact assessment was anchored on the comprehensive collection of daily health statistics. This included the total number of deaths and hospital emergency admissions, with a particular emphasis on respiratory diseases, which are known to be influenced by air pollution exposure. By analyzing both overall health outcomes and specific causes, the study aimed to provide a thorough understanding of the health ramifications of air pollution. This approach ensured that the analysis captured the full spectrum of potential health effects attributable to air pollution.

Addressing Potential Confounders: The study meticulously incorporated meteorological variables such as temperature and humidity, recognizing their potential to confound the analysis. Seasonal and chronological factors, including the day of the week and holidays, were systematically controlled for. Additionally, the influence of influenza outbreaks and pollen concentrations was considered, further strengthening the robustness of the results.

Unified Methodology: A harmonized methodology was employed across all participating groups to ensure the comparability of effect estimates. Autoregressive Poisson models, capable of accommodating overdispersion, were utilized to report health effects as relative risks. This methodological consistency was crucial in adapting the analysis to the varied local climates, pollution levels, and social conditions encountered across European settings.

Comprehensive Findings: The study's findings revealed that the uniform approach facilitated a collective analysis of data from diverse European locales. The models yielded relative risk estimates, shedding light on the seasonally varying effects of air pollutants and their interactions. This comprehensive analysis underscores the importance of considering both individual and combined effects of pollutants on health.

Analytical Rigor: Poisson regression, accounting for autocorrelation and overdispersion, served as the backbone of the statistical analysis. This approach allowed for the control of long-term trends and seasonal fluctuations, ensuring a focused examination of the relationship between air pollutants and health outcomes. The use of relative risks as a measure provided a clear understanding of the potential health implications of air pollution.

Meta-Analytical Synthesis: To enhance the analytical power and facilitate cross-city comparisons, a meta-analysis was conducted. This synthesis combined relative risk estimates from various cities, assessing the heterogeneity and exploring synergies among different predictive variables. The meta-analytical technique not only bolstered the study's statistical strength but also provided a comprehensive view of the health impacts of air pollution across Europe.

Geographical and Seasonal Insights: The study's air pollution data was strategically sourced from monitoring stations located in high-traffic urban areas and background urban zones to accurately reflect the urban pollution profile. For ozone, a pollutant with lower concentrations in urban centers, suburban stations were also utilized. This geographical strategy ensures a comprehensive capture of the urban pollution landscape. The analysis highlighted distinct seasonal variations in pollution levels,

with some cities experiencing heightened pollution in winter due to pollutants like sulfur dioxide and particulates, while others faced elevated ozone levels in summer. This seasonal differentiation is crucial for a nuanced understanding of pollution trends and their health implications.

Collaborative Strength and Biological Foundation: The project's collaborative nature, involving multiple cities, allowed for a substantial sample size and a wide range of exposure scenarios, thereby enhancing the generalizability and robustness of the findings. The statistical analyses were not only statistically sound but also biomedically informed, ensuring that the results are anchored in biological plausibility and relevance.

Public Health Impact and Future Directions: The research underscores the pressing need to address air pollution's health effects, as even modest but consistent associations across different urban settings can have significant public health implications. These findings can guide policymakers in devising targeted interventions to improve air quality and public health.

Concluding Remarks on Research Standardization: The study identifies a critical research gap: the absence of a standardized method to evaluate the short-term health effects of air pollution. The conclusion calls for the establishment of such a standardized approach, which is essential for ensuring comparability and adaptability in future epidemiological research, particularly when dealing with aggregated time series data.

2.
Milojevic, A., Wilkinson, P., Armstrong, B., Bhaskaran, K., Smeeth, L., & Hajat, S. (2014). **Short-term effects of air pollution on a range of cardiovascular events in England and Wales: Case-crossover analysis of the MINAP database, hospital admissions, and mortality.**
Original Article. Retrieved from ResearchGate:
https://www.researchgate.net/publication/263288440_Short-term_effects_of_air_pollution_on_a_range_of_cardiovascular_events_in_England_and_Wales_Case-crossover_analysis_of_the_MINAP_database_hospital_admissions_and_mortality

The study investigates the short-term effects of air pollution on various cardiovascular events in England and Wales, including the MINAP database for acute coronary syndrome/MI, HES for over 2 million CVD emergency hospital admissions, and mortality data from the Office for National Statistics for over 600,000 CVD deaths. Employing a time-stratified case-crossover design, the research examines the relationship between air pollution levels and cardiovascular outcomes. Results indicate that air pollution is associated with increased rates of cardiovascular events, underscoring the importance of addressing air quality to safeguard public health. Three national databases from England and Wales were utilized for data analysis.

Approach: The study adopts a case-crossover analysis to investigate the short-term effects of air pollution on cardiovascular events. This approach allows for a detailed examination of the temporal relationship between exposure to pollutants and subsequent health outcomes.

Research : The central problem addressed by the research is the impact of short-term exposure to air pollutants, specifically NO2 and PM2.5, on the incidence of cardiovascular events such as arrhythmia. The study seeks to determine the extent to which these pollutants contribute to the risk of cardiovascular incidents.

Myocardial Ischaemia National Audit Project (MINAP) Database:
- MINAP is a comprehensive national register of admissions to hospitals for patients with acute coronary syndrome/MI.
- Data included patient characteristics, diagnostic information, comorbidities, acute treatment, and outcomes.
- Analysis encompassed events occurring during 2003–2009, with location information rounded to a resolution of 100 m.
- Hospital Episode Statistics (HES):
- HES provided data on over 2 million CVD emergency hospital admissions.

- Specific cardiovascular events were identified within the database.
- Mortality Data (Office for National Statistics):
- Mortality data included over 600,000 CVD deaths, allowing for analysis of mortality outcomes.

Pollutants and Exposure Assessment:
- The research linked **daily mean concentrations** of air pollutants such as **CO, NO2, PM10, PM2.5, SO2, and O3** with cardiovascular events.
- **Exposure assessment** was meticulously conducted using the daily maximum of 8-hourly running mean of **O3**, measured at the nearest air pollution monitoring site to the patient's residence, providing a precise measure of the individual's exposure.

Statistical Analysis:
- The study modeled the **effects of pollutants** on cardiovascular health using lags of up to **4 days** to capture the immediate and short-term impact.
- **Adjustments** were made for **ambient temperature** and **day of the week**, accounting for potential confounding factors that could influence the results.

Comparison of Associations:
- A thorough examination of the **associations between air pollution** and various cardiovascular outcomes, including **MI, arrhythmias, atrial fibrillation, heart failure, and pulmonary embolism**, was conducted.
- **Effect estimates** were provided as percentage increases in risk for a **10th–90th centile** increase in pollutant concentration, offering a clear quantification of the health risks associated with air pollution.

Discussion:
- The study's discussion aimed to enhance the understanding of the **pathophysiological mechanisms** by which air pollution impacts cardiovascular disease.
- It specifically investigated whether **thrombotic events**, such as acute MI and stroke, have stronger associations with air pollution compared to **non-thrombotic outcomes**, suggesting that air pollution may affect cardiovascular health through multiple pathways.

Algorithms:
The specific algorithms used for statistical modeling and analysis were not explicitly mentioned in the provided excerpt. However, typical approaches for case-crossover designs may involve conditional logistic regression or other time-series analysis methods.

Table 1: Summary of exposure data in 2003 - 2009

Pollutant	Number of monitors	Median (IQR)	10th–90th centile range
CO (mg/m^3)	61	0.2 (0.2–0.4)	0.4
NO_2 (µg/m^3)	93	24 (13–37)	45
O_3 (µg/m^3)	82	61 (46–76)	61
PM_{10} (µg/m^3)	62	20 (15–27)	26
$PM_{2.5}$ (µg/m^3)	46	10 (7–15)	16
SO_2 (µg/m^3)	71	3.1 (2–6)	10.4
Mean temperature (d/C)	717	9.85 (6.2–13.9)	13.85

CO, carbon monoxide; NO_2, nitrogen dioxide; $PM_{2.5}$, particulate matter less than 2.5 µm in aerodynamic diameter; PM_{10}, particulate matter less than 10 µm in aerodynamic diameter; SO_2, sulfur dioxide.

Findings and Research Gap:
- The findings indicate a significant association between air pollution and increased rates of cardiovascular events.
- The research identifies a gap in understanding the mechanisms by which air pollution affects cardiovascular health, particularly in relation to thrombotic events.

Conclusion:

- The study concludes that there is a need for further research to fully understand the health impacts of air pollution and to develop interventions that can mitigate these effects.
- The findings have significant implications for public health policies and clinical practices aimed at reducing the burden of cardiovascular diseases.

3.

Bristol City Council. (2017, February). **Health impacts of air pollution in Bristol**. Retrieved from Bristol City Council website:

https://www.bristol.gov.uk/files/documents/599-health-impacts-of-air-pollution-in-bristol-february-2017/file#:~:text=The%20new%20results%20show%20that,being%20attributable%20to%20air%20pollution

This document is a report on the health impacts of air pollution in the City of Bristol. It was prepared by Air Quality Consultants Ltd on behalf of Bristol City Council. The report provides an update on a previous study conducted in 2014, which concluded that 188 deaths in Bristol in 2010 were attributable to air pollution, with 24 of these deaths being caused by local road traffic emissions. The new results show that around 300 deaths each year in Bristol can be attributed to exposure to both nitrogen dioxide and fine particulate matter, representing about 8.5% of deaths in the city being attributable to air pollution. The proportions of deaths attributable to air pollution vary across the city, with concentrations and deaths being highest in the center of the city. Road traffic is identified as the dominant local source of emissions contributing to these deaths.

The report also provides information on the context of air quality in Bristol. The city operates a number of monitoring sites for nitrogen dioxide and particulate matter, and an Air Quality Management Area (AQMA) has been declared for exceedances of the annual mean and hourly nitrogen dioxide objectives. Background concentrations of PM2.5 and nitrogen dioxide in Bristol are shown in figures 2 and 3 of the report.

The sources of air pollution in Bristol are discussed in the document. Particulate matter (PM2.5) has both natural and anthropogenic sources, including road vehicles, industrial sources, power stations, domestic heating, and shipping. The population-weighted total PM2.5 concentration, with 81% being anthropogenic, 50% being secondary PM2.5, and 23% being regional primary PM2.5. This suggests that approximately 27% of the anthropogenic fraction is effectively from local sources and potentially locally controllable. Nitrogen dioxide concentrations are mainly determined by emissions from combustion processes, with road transport and the electricity supply industry being the main sources in the UK. Within Bristol, road transport is the main contributor to nitrogen dioxide emissions.

The report also explains the contribution of air pollution to deaths. The Committee on the Medical Effects of Air Pollution (COMEAP) has quantified the effects of exposure to particulate matter and nitrogen dioxide on deaths. The risk coefficients used to quantify deaths attributable to PM2.5 and nitrogen dioxide are provided in Table 1 of the report. The calculations use the relative risk (RR) for a 10 µg/m3 concentration increment. The RR for PM2.5 is 1.06 (range 1.01 to 1.12), while the RR for nitrogen dioxide is 1.025 (range 1.01 to 1.04).

Key Findings:
- Approximately **300 deaths annually** in Bristol are attributable to air pollution, specifically to nitrogen dioxide and fine particulate matter (PM2.5).
- This figure represents about **8.5% of all deaths** in the city, with the highest concentrations and death rates in the city center.
- Road traffic emissions are identified as the **primary local source** of these harmful pollutants.

Health Impact Assessment:
- The Committee on the Medical Effects of Air Pollution (COMEAP) has provided risk coefficients to quantify the impact of PM2.5 and nitrogen dioxide on mortality.
- The relative risk (RR) for a 10 µg/m3 increase in PM2.5 concentration is **1.06** (range 1.01 to 1.12), and for nitrogen dioxide, the RR is **1.025** (range 1.01 to 1.04).

Uncertainties and Research Needs:

- **Acknowledgment of Calculation Uncertainties**: The report recognizes uncertainties in the estimations, especially concerning pollutant concentrations and risk coefficients, with wide confidence intervals for both PM2.5 and nitrogen dioxide, highlighting the necessity for ongoing research and refinement of these estimates.

Demographic Considerations:
- **Age and Sensitivity Variations**: Emphasizing the need to consider the age structure and vulnerability of different populations when evaluating health effects distribution, indicating that interventions should be tailored to the unique demographics of each ward.
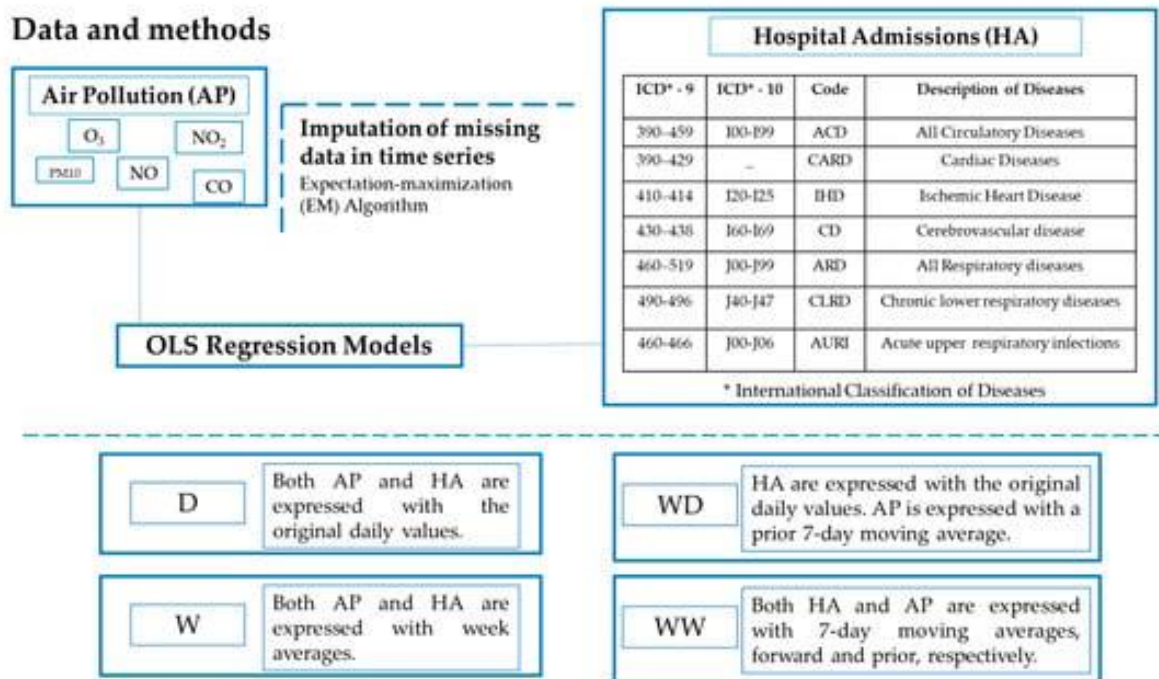
Comparative Analysis:
- **Comparison with Road Traffic Fatalities**: The report draws a comparison between air pollution-related deaths and fatalities from road traffic collisions, which were numbered at 12 in 2013, to illustrate the significant and often underrecognized impact of air pollution on public health.

In summary, this report provides updated information on the health impacts of air pollution in Bristol. It shows that around 300 deaths each year in the city can be attributed to exposure to both nitrogen dioxide and fine particulate matter. The report highlights the variations in deaths attributable to air pollution across the city and identifies road traffic as the main source of emissions contributing to these deaths.

4.
Franco, P., Gordo, C., Marques da Costa, E., & Lopes, A. (2020). Air Pollution and Emergency Hospital Admissions—Evidences from Lisbon Metropolitan Area, Portugal. Applied Sciences, 10(22), 7997.

Data and methods

Air Pollution (AP)
O_3 NO_2 PM10 NO CO

Imputation of missing data in time series
Expectation-maximization (EM) Algorithm

OLS Regression Models

Hospital Admissions (HA)

ICD* - 9	ICD* - 10	Code	Description of Diseases
390-459	I00-I99	ACD	All Circulatory Diseases
390-429	–	CARD	Cardiac Diseases
410-414	I20-I25	IHD	Ischemic Heart Disease
430-438	I60-I69	CD	Cerebrovascular disease
460-519	J00-J99	ARD	All Respiratory diseases
490-496	J40-J47	CLRD	Chronic lower respiratory diseases
460-466	J00-J06	AURI	Acute upper respiratory infections

* International Classification of Diseases

D	Both AP and HA are expressed with the original daily values.	WD	HA are expressed with the original daily values. AP is expressed with a prior 7-day moving average.
W	Both AP and HA are expressed with week averages.	WW	Both HA and AP are expressed with 7-day moving averages, forward and prior, respectively.

Air Pollution and Emergency Hospital Admissions—Evidences from Lisbon Metropolitan Area, PortugalThe relevance of air pollution in the public health agenda has recently been reinforced—it is known that exposure to it has negative effects in the health of individuals, especially in big cities and metropolitan areas. In this article we observed the evolution of air pollutants (CO, NO, NO2, O3, PM10) emissions and we confront them with health vulnerabilities related to respiratory and circulatory diseases (all circulatory diseases, cardiac diseases, cerebrovascular disease, ischemic heart disease, all respiratory diseases, chronic lower respiratory diseases, acute upper respiratory infections). The study is supported in two databases, one of air pollutants and the other of emergency hospital admissions, in the 2005–2015 period, applied to the Lisbon Metropolitan Area. The analysis was conducted through Ordinary Least Squares (OLS) regression, while also using semi-elasticity to quantify associations. Results showed positive associations between air pollutants and admissions,

1.3 Objective

The research paper is a study that evaluates the impact of air pollution on emergency hospital admissions in the Lisbon Metropolitan Area (LMA), Portugal, over a period from 2005 to 2015. It specifically investigates the correlation between air pollutants such as carbon monoxide (CO), nitrogen oxides (NO, NO_2), ozone (O_3), and particulate matter (PM_{10}) with respiratory and circulatory diseases. The study employs Ordinary Least Squares (OLS) regression and semi-elasticity to quantify the associations and also considers demographic and socioeconomic factors that may influence vulnerability to air pollution.

Research Approach and Data Analysis: A retrospective analysis was conducted, encompassing a decade of data from 2005 to 2015. The research focused on five major air pollutants: carbon monoxide (CO), nitrogen oxides (NO and NO_2), ozone (O_3), and particulate matter (PM10). The approach aimed to discern patterns and correlations between the presence of these pollutants and the incidence of health emergencies.

Methodology: The study utilized Ordinary Least Squares (OLS) regression and semi-elasticity as its primary analytical tools. These methods were employed to quantify the relationship between pollutant levels and hospital admissions, taking into account the variability in demographic and socioeconomic factors that could affect susceptibility to air pollution.

1.4 Findings

The study's findings indicate a positive association between air pollutants and emergency hospital admissions, with respiratory diseases being more impacted than circulatory diseases. CO and O_3 showed the highest number of associations with hospital admissions. Notably, O_3 was identified as a significant predictor for hospital admissions in the under-15 age group, while PM_{10} was particularly predictive for the over-64 age group. This suggests that different age groups may be affected differently by various pollutants.

Geographical Disparities: The research also highlights the distinction between urban and suburban areas in terms of air pollution and its relation to emergency hospital admissions. The urban city core showed higher levels of pollution and related health outcomes compared to the suburban areas.

Research Gap and Need for Comprehensive Data: The study identified a significant gap in the availability of comprehensive air pollution data across the LMA. This limitation points to the necessity for more extensive data collection efforts to achieve a more accurate and holistic understanding of air pollution's health impacts.The document acknowledges that there are uncertainties in the calculations, particularly regarding the concentrations used and the risk coefficients. It cites wide confidence intervals for both PM2.5 and nitrogen dioxide, which points to the need for further research and more precise data to refine these estimates.

Overall, the study concludes that air pollution, especially CO and c, is associated with an increased risk of emergency hospital admissions for respiratory and circulatory diseases in the LMA. The findings emphasize the importance of addressing air pollution as a public health concern, especially in urban areas, and highlight the need for targeted interventions to protect vulnerable age groups from the negative effects of air pollution. The study also identifies a research gap in the form of a lack of comprehensive air pollution data across all territories in the LMA, suggesting a need for expanded data collection and analysis to better understand the full impact of air pollution on health.

2. Data Collection and Preprocessing

This report is to document the data cleansing and preprocessing steps taken on air pollution and Emergency Hospital Admissions in Bristol.
All the codes related to this part of the project, belong to the team git repository, in which the link to the related codes has been mentioned in the body.

2.1 Emergency Hospital Admissions data

Github link to the related code: Health data cleansing code1 & Health data cleansing code1

Datasets to achieve emergency hospital admissions have been downloaded from NHS. These datasets are time series which contain the monthly count of different types of UK hospital admissions for a year.
Based on the description in the provided NHS site, the weekly and monthly A&E Attendances and Emergency Admissions collection collects the total number of attendances in the specified period for all A&E types, including Urgent Treatment Centres, Minor Injury Units and Walk-in Centres, and of these, the number discharged, admitted or transferred within four hours of arrival. Also included are the number of Emergency Admissions, and any waits of over four hours for admission following the decision to admit. Data are shown at provider organisation level, from NHS Trusts, NHS Foundation Trusts and Independent Sector Organisations.
The table illustrated in Figure 1 shows the schema of the raw data file.

Table 2: Column description of Health Dataset. Github link to the related file: HealthData

#	Column Name and Description
1	Date (Year-Month)
2	Code (of the NHS Trust, NHS Foundation Trust or the Independent Sector Organisation)
3	Region (of the NHS Trust, NHS Foundation Trust or the Independent Sector Organisation)
4	Name (of the NHS Trust, NHS Foundation Trust or the Independent Sector Organisation)
5	A&E attendances Type 1 Departments - Major A&E
6	A&E attendances Type 2 Departments - Single Specialty
7	A&E attendances Type 3 Departments - Other A&E/Minor Injury Unit
8	A&E attendances Total attendances
9	A&E attendances less than 4 hours from arrival to admission, transfer or discharge Type 1 Departments - Major A&E
10	A&E attendances less than 4 hours from arrival to admission, transfer or discharge Type 2 Departments - Single Specialty
11	A&E attendances less than 4 hours from arrival to admission, transfer or discharge Type 3 Departments - Other A&E/Minor Injury Unit
12	A&E attendances less than 4 hours from arrival to admission, transfer or discharge Total Attendances < 4 hours
13	A&E attendances greater than 4 hours from arrival to admission, transfer or discharge Type 1 Departments - Major A&E
14	A&E attendances greater than 4 hours from arrival to admission, transfer or discharge Type 2 Departments - Single Specialty
15	A&E attendances greater than 4 hours from arrival to admission, transfer or discharge Type 3 Departments - Other A&E/Minor Injury Unit
16	A&E attendances greater than 4 hours from arrival to admission, transfer or discharge Total Attendances > 4 hours

17	Emergency Admissions Emergency Admissions via Type 1 A&E
18	Emergency Admissions Emergency Admissions via Type 2 A&E
19	Emergency Admissions Emergency Admissions via Type 3 and 4 A&E
20	Emergency Admissions Total Emergency Admissions via A&E
21	Emergency Admissions Other Emergency admissions (i.e not via A&E)
22	<mark>Emergency Admissions Total Emergency Admissions</mark>
23	Emergency Admissions Number of patients spending >4 hours from decision to admit to admission
24	Emergency Admissions Number of patients spending >12 hours from decision to admit to admission

Cleansing and Preprocessing steps done on this dataset are:
1. **Handling invalid values:** Utilized Python for cleaning each dataset to drop NaN and invalid values.
2. **Merging data files**: Merged all NHS data files for each year from 2015 to 2024 into a single file, and developed a merging mechanism in Python to handle variations in column names and structures.
3. **Cropping data:** Cropped data into a table containing total emergency admissions for Bristol City NHS organizations. (Figure 2)
4. **Total value:** Calculated total value of Bristol city emergency admissions. (Figure 3 and Figure 4)
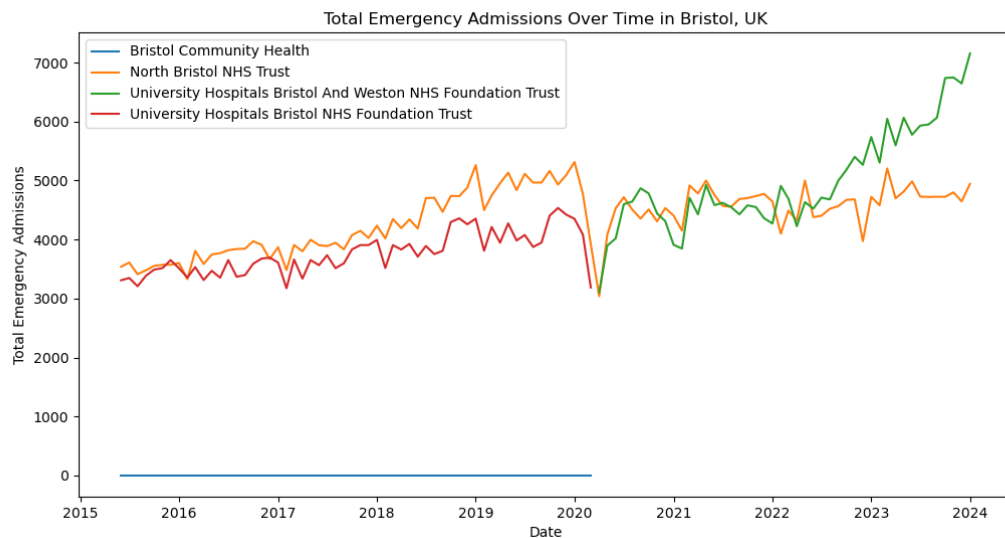

Figure 1: Merged, cleaned and cropped data that only shows count of emergency hospital admissions for Bristol City health organizations during period 2015 to 2024.

As shown in the Figure 2, Bristol Community Health ended their reporting in 2020, and "University Hospitals Bristol NHS Foundation Trust" has been changed to "University Hospitals Bristol and Weston NHS Foundation Trust".
All hospitals do not have all types of patients. As a result, number of emergency admissions of Bristol Community Health is zero.

Final look of the data which is used for the project is shown in the Figure 3 and visualized in Firgure 4.

Table 3: Final health data schema. Github link to the related file: FinalHealthData

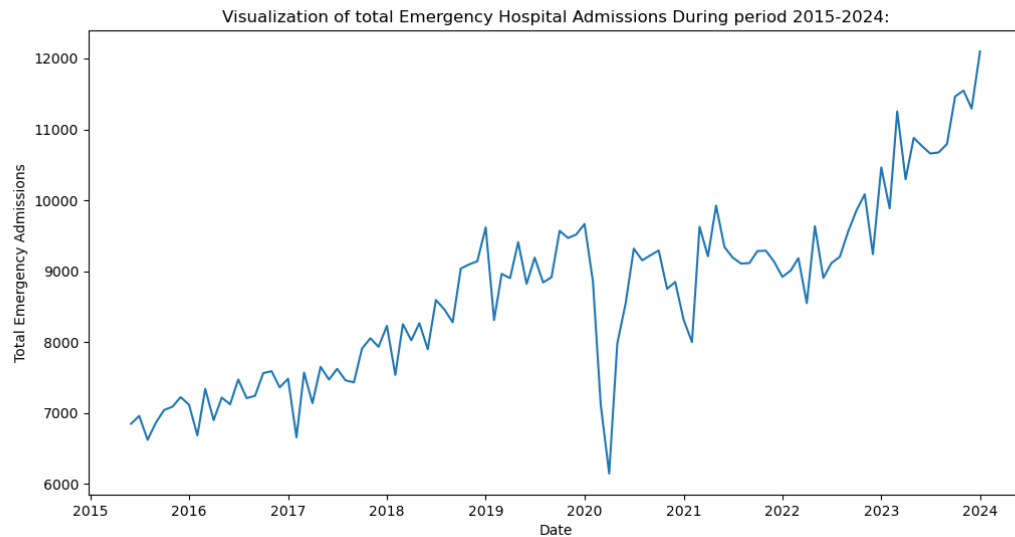| #	Column Name	Unit
1	Date	Date
2	Total Emergency Admissions	Number

Figure 2: Final look of total emergency hospital admissions in Bristol, which is the sum value of emergency admissions of all the health organizations in this city.

2.2 Air polution data

Github link to the related code: AP data cleansing code

For the air pollution data, we turned to the UK-AIR database, maintained by the UK DEFRA, to acquire datasets pertaining to air quality monitoring stations across the UK, including those in Bristol City. Each dataset corresponds to a specific monitoring station, providing hourly time series data on various pollutants.

Upon retrieval, the datasets underwent extensive preprocessing to ensure their suitability for analysis and integration with the health data. This preprocessing involved several steps:

1. **Cleaning Invalid Values:** We addressed inconsistencies in the data by converting 'No data' values to NaN (Not a Number) and removing rows containing negative pollutant values attributed to sensor faults.
2. **Consolidation:** To streamline analysis, all air pollution data files were consolidated into a single comprehensive file, facilitating easier management and manipulation.
3. **Temporal Alignment**: Given that the air pollution data is reported hourly while the health data is aggregated monthly, we needed to synchronize the two datasets effectively. This synchronization process involved aligning timestamps and aggregating the hourly data to match the monthly frequency.

To achieve this synchronization, we considered several approaches:

- **Resampling:** By resampling the hourly data to the monthly frequency, we aggregated the data into monthly intervals using statistical functions such as mean or sum.

- **Rolling Window Aggregation:** Alternatively, we employed rolling window statistics to calculate aggregated values over monthly intervals, offering greater flexibility in capturing short-term variations.
- **Interpolation:** In cases where the hourly data was dense, interpolation was utilized to fill missing values before aggregating to the monthly frequency, preserving the original granularity.
- **Timestamp Alignment:** Aligning the timestamps of the hourly data with the monthly data allowed for aggregation within each month, typically by taking the mean or another relevant aggregation function.

Ultimately, we opted to retain the minimum, mean, and maximum values of each pollutant within a given month. This decision was informed by the belief that both maximum and minimum values could hold significance for future analysis and inform subsequent research endeavors.
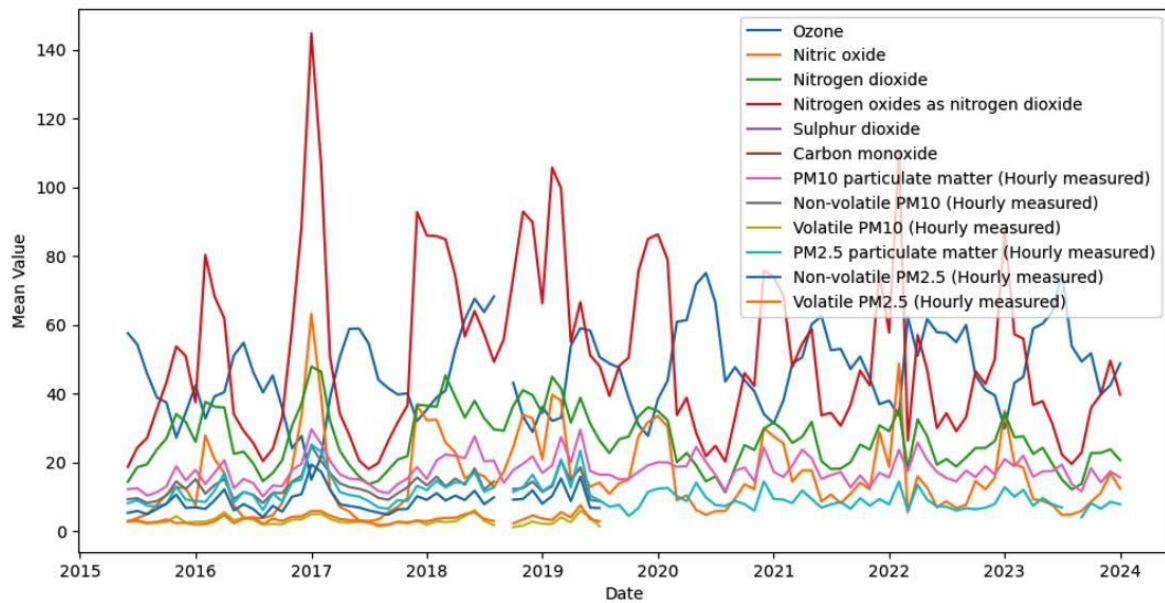

Figure 3: Look of Air Pollution merged dataset.

Final Look of Air Pollution Data:

Table 4: Final Look of Air Pollution Data:

	Column Name	Sub-column	Scale
1	Date		date
2	Time		time
3	StationName		
4	Ozone	Min	V ugm-3
		Mean	
		Max	
5	Nitric oxide	Min	V ugm-3
		Mean	
		Max	
6	Nitrogen dioxide	Min	V ugm-3
		Mean	

		Max	
7	Nitrogen oxides as nitrogen dioxide	Min	V ugm-3
		Mean	
		Max	
8	Sulphur dioxide	Min	V ugm-3
		Mean	
		Max	
9	Carbon monoxide	Min	V ugm-3
		Mean	
		Max	
10	PM10 particulate matter (Hourly measured)	Min	V ugm-3 (Ref.eq)(TEOM FDMS)
		Mean	
		Max	
11	Non-volatile PM10 (Hourly measured)	Min	V ugm-3
		Mean	
		Max	
12	Volatile PM10 (Hourly measured)	Min	V ugm-3
		Mean	
		Max	
13	PM2.5 particulate matter (Hourly measured)	Min	V ugm-3
		Mean	
		Max	
14	Non-volatile PM2.5 (Hourly measured)	Min	V ugm-3 (TEOM FDMS)
		Mean	
		Max	
15	Volatile PM2.5 (Hourly measured)	Min	V ugm-3 (TEOM FDMS)
		Mean	
		Max	
16	Modelled Wind Direction	Min	N deg
		Mean	
		Max	
17	Modelled Wind Speed	Min	N ms-1
		Mean	
		Max	
18	Modelled Temperature	Min	N degC
		Mean	
		Max	

Github link to the related Files: AP raw data & AP final data

3. Methodology

In this section, we outline the methodological approach adopted to investigate the correlation between air pollution and emergency hospital admissions. The methodology encompasses a systematic process of data processing, exploratory analysis, predictive modelling, and correlation analysis. Emphasizing transparency and reproducibility, the methodology prioritizes code-driven data preprocessing to minimize manual interventions and ensure the integrity of the analysis. Leveraging insights from existing literature, we designed our approach to address key gaps identified in previous studies, aiming to contribute novel findings to the field of air pollution epidemiology. Through a combination of statistical techniques and machine learning algorithms, our methodology seeks to uncover nuanced relationships between air pollutants and health outcomes, providing valuable insights for public health decision-making. By detailing our methodological framework, we aim to provide a comprehensive understanding of the analytical processes employed in this study, facilitating reproducibility and enabling future research endeavours in this critical area of study.

3.1 Correlation Calculation between Air pollution and Weather conditions

The inclusion of weather variables in the correlation analysis allowed us to discern potential confounding factors that could obscure the true relationship between air pollution and health outcomes. For instance, high levels of air pollution during colder months may coincide with increased respiratory illnesses, but without accounting for temperature variations, it would be challenging to ascertain whether air pollution itself is the primary driver of health effects or if cold temperatures play a significant role.

To investigate the correlation between air pollution and emergency hospital admissions, the project utilized statistical methods, including correlation and regression analyses. A correlation matrix was initially generated between pollutants and weather conditions, including monthly mean temperature and wind speed, to gain insights into their interrelationships.
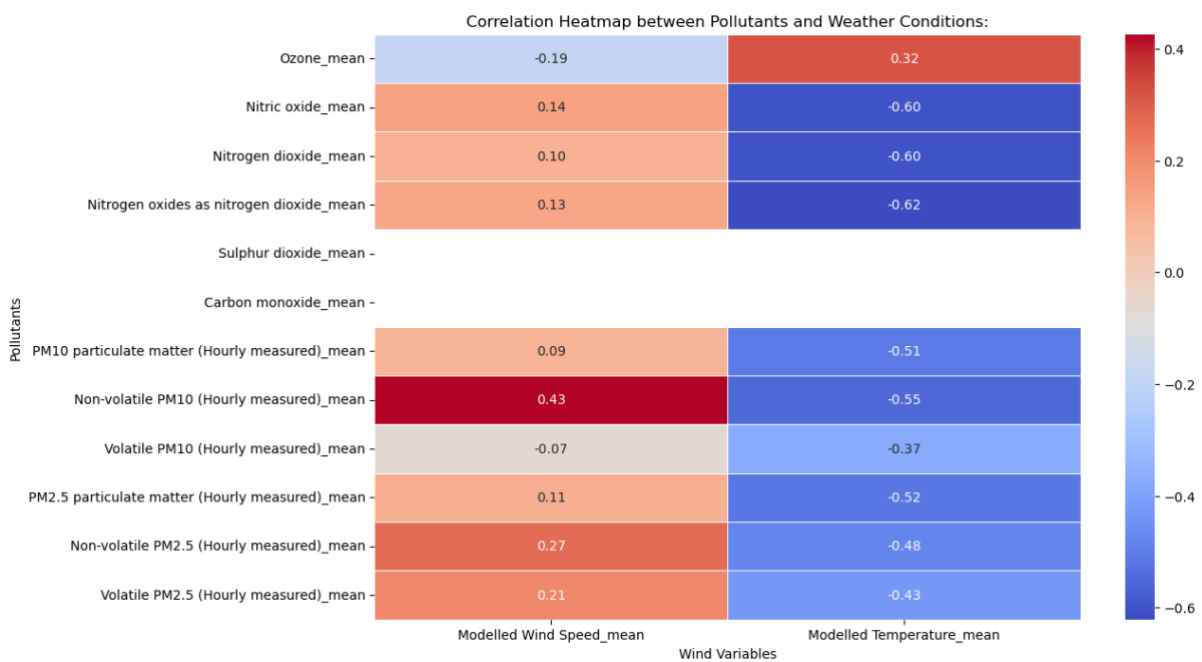

Figure 4: Correlation Heatmap between air pollutant and weather conditions

Overall, the correlation calculation between air pollution and weather conditions served as a foundational step in our analysis, providing valuable insights into the complex interplay between environmental factors and health outcomes. By systematically examining these relationships, we laid the groundwork for subsequent analyses aimed at elucidating the direct impact of air pollution on emergency hospital admissions.

3.2 Correlation Calculation between Categories and Emergency Hospital Admissions

Subsequently, simple correlation matrices were constructed for each pollutant category with total emergency admissions. These matrices provided foundational insights into the individual and collective impacts of pollutants on emergency admissions, laying the groundwork for more detailed analyses of their correlation with health outcomes.
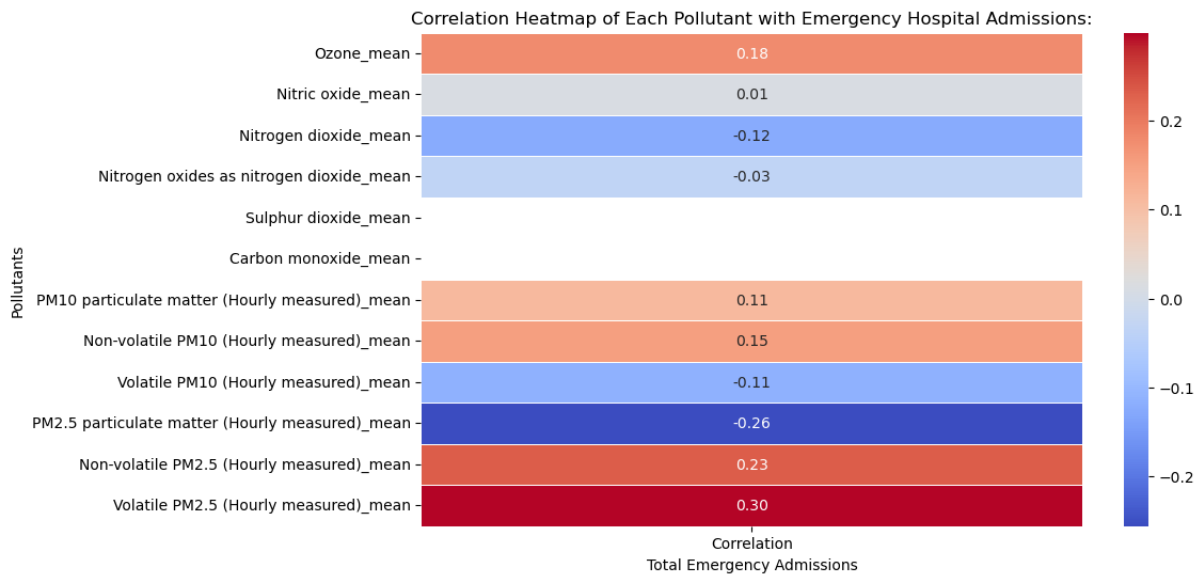

Figure 5: Correlation Heatmap of Each Pollutant with Emergency Hospital Admissions.

3.3 Categorized Air pollutants approach

Upon identifying weather conditions that exerted similar influences on certain pollutants, the project proceeded to categorize pollutants into three distinct groups:

Nitrogen pollutants: Nitric oxide, Nitrogen dioxide, Nitrogen oxides as nitrogen dioxide. Originating primarily from human activities such as industrial processes, agriculture, transportation, and energy production, nitrogen pollutants can exert profound impacts on ecosystems, human health, and atmospheric composition. By grouping these pollutants together, we aimed to discern common trends in their correlation with emergency hospital admissions and assess their collective influence on public health.
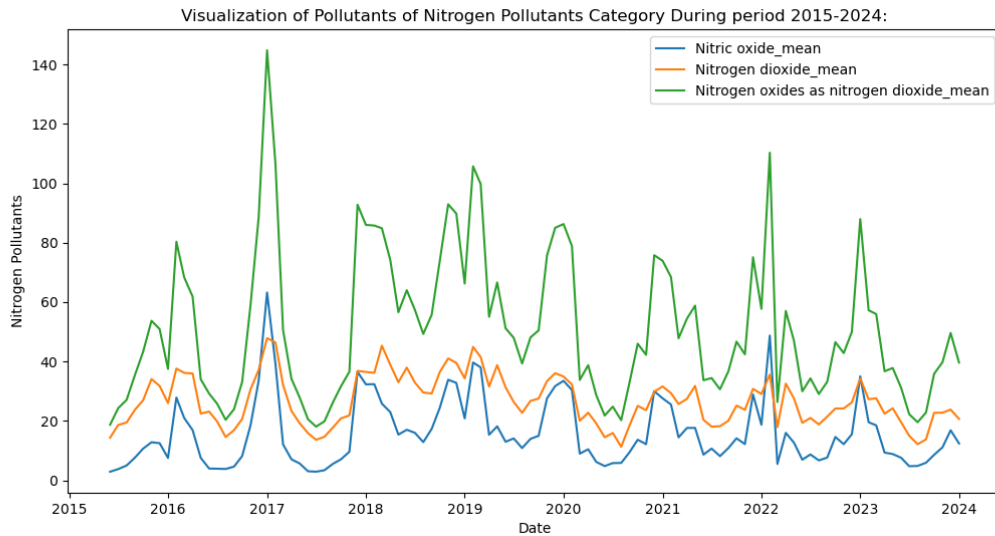
Figure 6: Visualization of Nitrogen Pollutants Category During period 2015-2024

PM pollutants: PM10 particulate matter (Hourly measured), Non-volatile PM10 (Hourly measured), Volatile PM10 (Hourly measured), PM2.5 particulate matter (Hourly measured), Non-volatile PM2.5 (Hourly measured), Volatile PM2.5 (Hourly measured).

PM pollutants are a significant component of air pollution and can originate from various sources, including vehicle emissions, industrial processes, and natural phenomena like dust storms and wildfires. By categorizing PM pollutants, we aimed to explore their distinct contributions to emergency hospital admissions and elucidate their implications for public health.
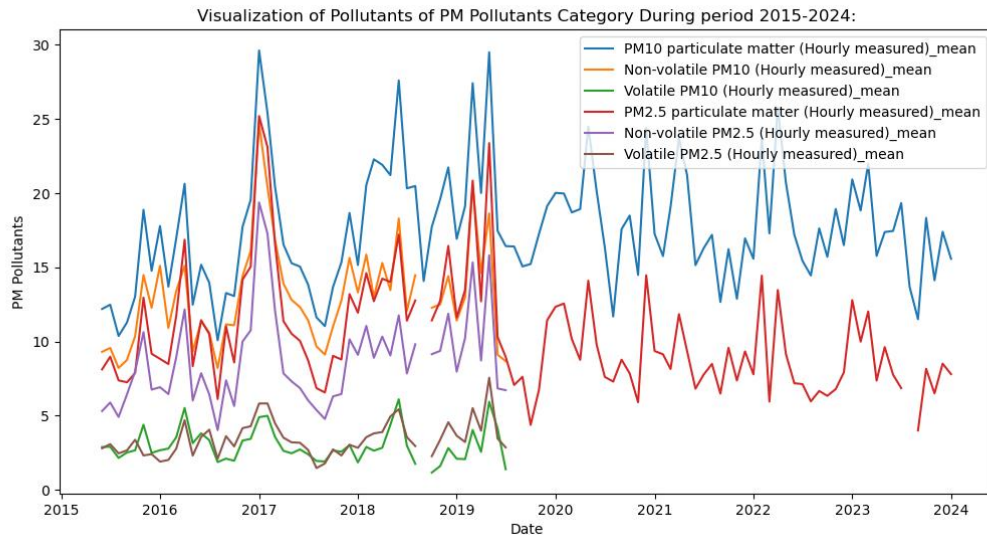

Figure 7: Visualization of PM Pollutants Category During 2015-2024.

Other pollutants: Ozone, Sulphur dioxide, Carbon monoxide.

While initially included in the analysis, sulphur dioxide and carbon monoxide were excluded due to data unavailability. Ozone, sulphur dioxide, and carbon monoxide are key air pollutants with diverse sources and impacts on human health and the environment. By examining their correlation with emergency hospital admissions, we sought to comprehensively assess the relationship between air pollution and public health outcomes.
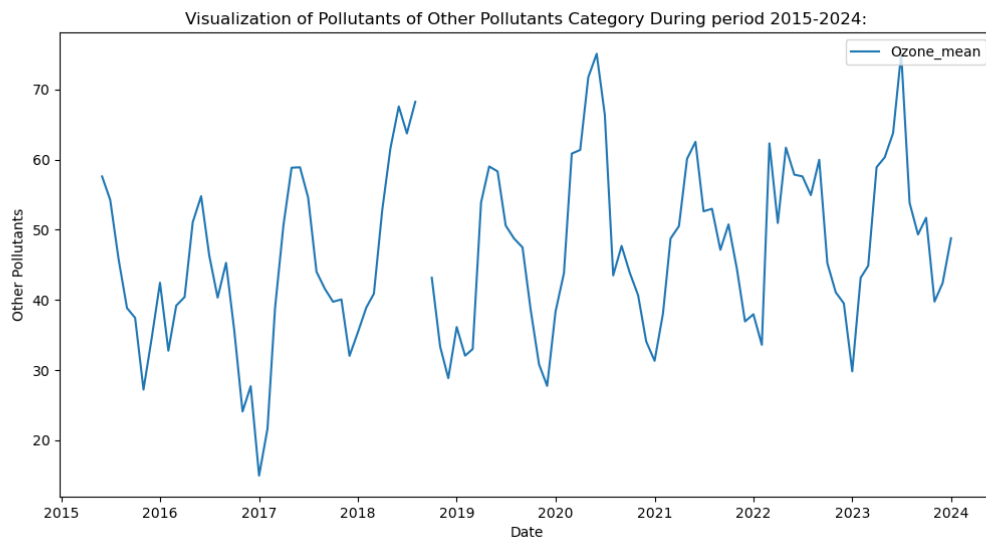
Figure 8: Visualization of Other Pullutants Cateory During 2015-2024

3.4 Analyzing Correlation of Emergency Hospital Admissions and Pollutants Categories using Comparative Analysis

In this section, we embark on a comparative analysis aimed at evaluating the performance of two distinct imputation methods, namely SimpleImputer and HistGradientBoostingRegressor, in predicting total emergency hospital admissions based on categorized air pollutants. This comparative approach serves to assess the efficacy of each method in capturing the intricate relationship between air pollution and public health outcomes, thereby providing valuable insights for informed decision-making in healthcare management.

The comparative analysis entails fitting predictive models using both SimpleImputer and HistGradientBoostingRegressor to forecast total emergency hospital admissions. Each method employs unique imputation techniques to address missing values within the dataset before training the predictive models. Subsequently, the predictions generated by both approaches are juxtaposed against the actual number of hospital emergency admissions, facilitating a thorough comparison of their respective performances.

To execute the comparative analysis effectively, correlation coefficients will be computed to quantify the strength and direction of the relationship between predicted and actual admissions for each pollutant category. This correlation analysis is pivotal in validating the predictive models and assessing their ability to accurately capture the impact of air pollution on public health outcomes.

3.5 Key aspects of this approach include:

Visualisation: Visualizing the actual and predicted emergency admissions for each pollutant category enables the identification of temporal trends and patterns. This visualization aids in interpreting the performance of the models, providing valuable insights into their predictive capabilities.
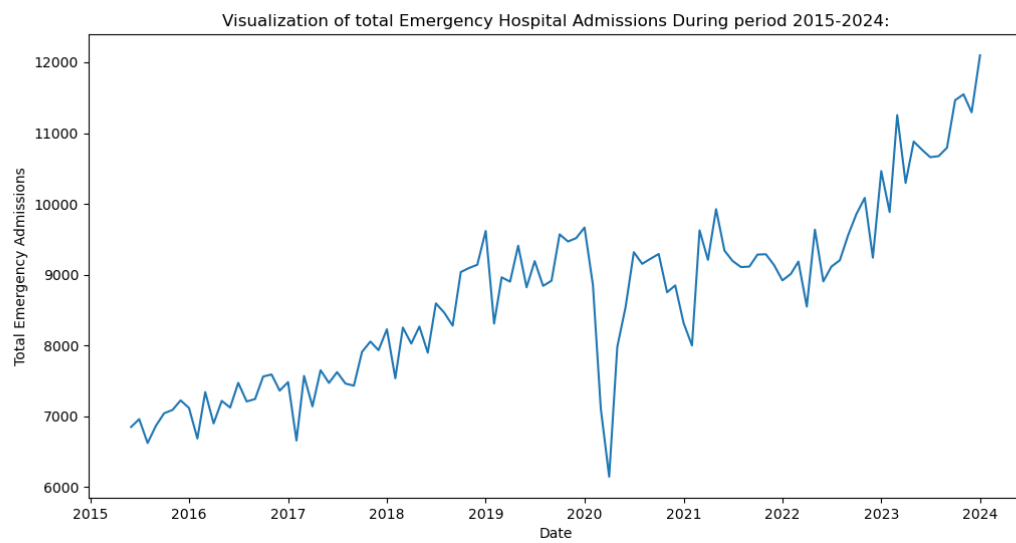
Figure 9: Visualisation of total Emergency Hospital Admissions During 2015-2024
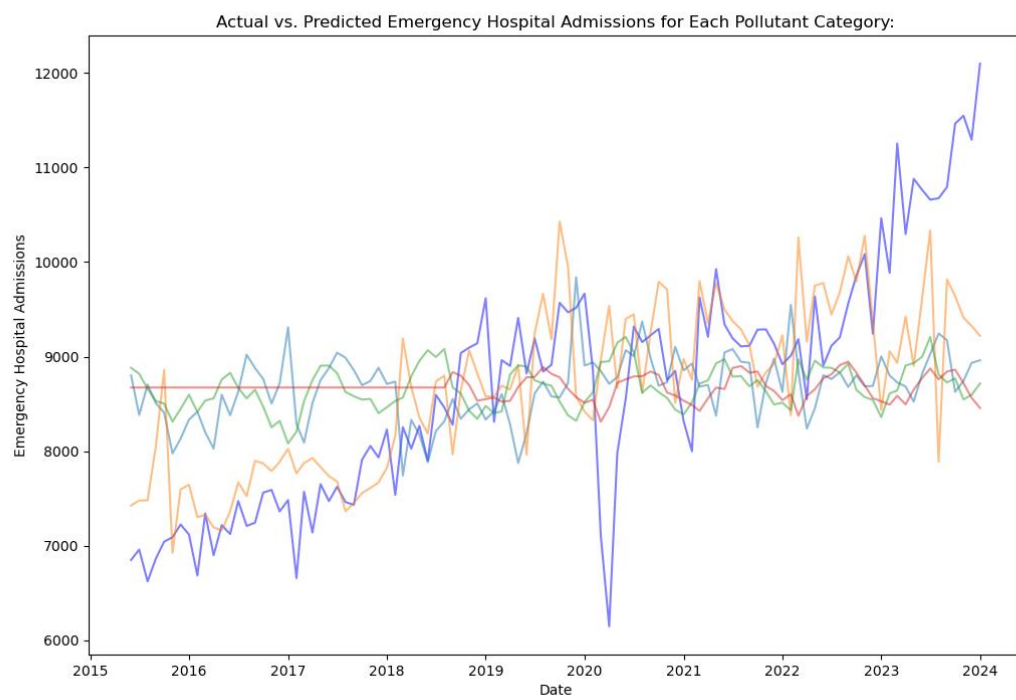

Figure 10: Actual vs Predicted Emergency Hospital Admissions for Each Pollutant Category
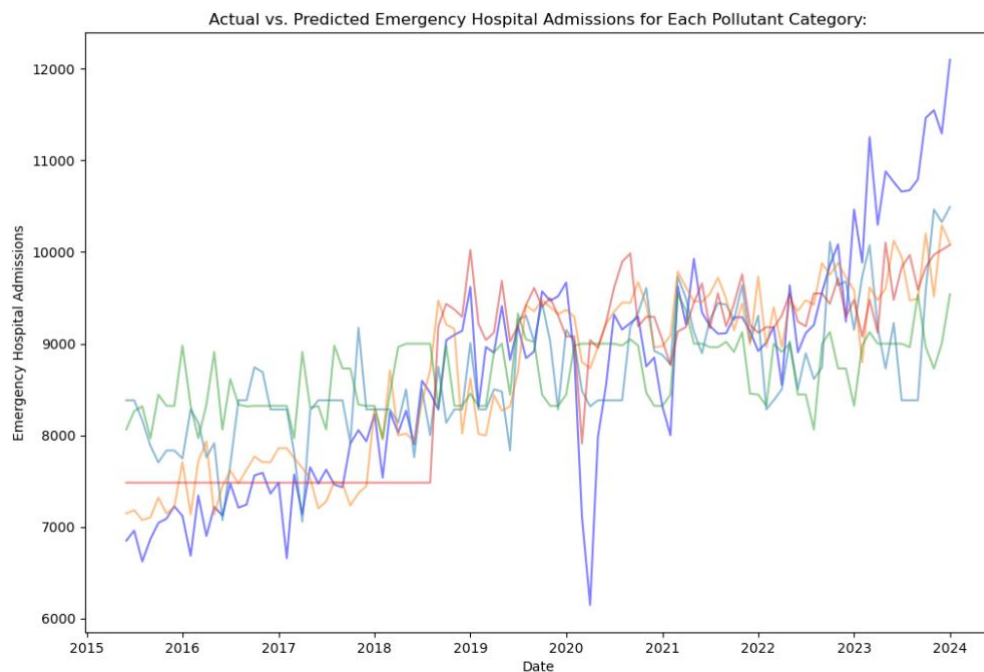
Figure 11: Actual vs Predicted Emergency Hospital Admissions for Each Pollutant Category

Correlation Analysis: Calculation of correlation coefficients between actual and predicted emergency admissions for each pollutant category facilitates the evaluation of the relationship between air pollution and emergency admissions. By assessing the strength and direction of these correlations, we gain valuable insights into the influence of different pollutant categories on emergency hospital admissions.
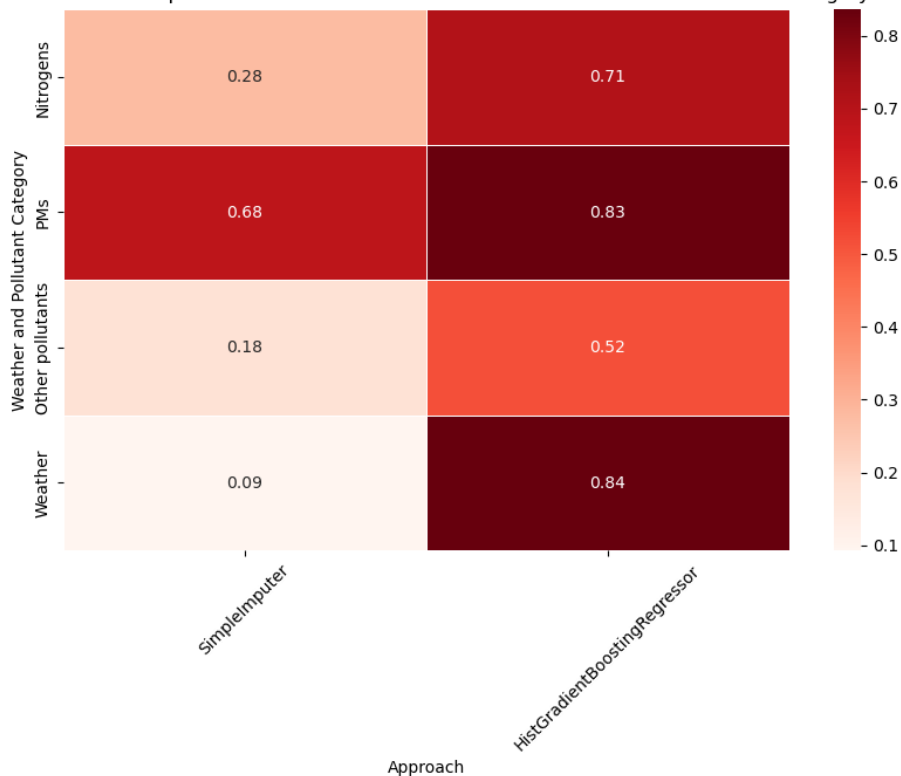

Figure 12: Correlation Heatmap between Actual and Predicted Admissions for Each Pollutant Category

Comparison of Error Metrics: Evaluation of Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2) scores across different modeling approaches facilitates the assessment of prediction accuracy. These error metrics provide quantitative measures of the models' performance, allowing for informed comparisons and guiding the selection of the most effective predictive models. Before delving into the comparative analysis, we introduce the imputation methods utilized to address missing values in the dataset. Imputation techniques are crucial for ensuring the integrity and accuracy of predictive models by replacing missing values with estimated values based on available data.

```
Metrics for SimpleImputer Approach:
Category: Nitrogens
MAE: 988.9102837674844
MSE: 1468699.8912541205
R2: 0.0763194713384393

Category: PMs
MAE: 645.1426501504344
MSE: 845239.0379525489
R2: 0.4684204403700646

Category: Other pollutants
MAE: 1015.3936817464821
MSE: 1538991.3936527423
R2: 0.03211241958974376

Category: Weather
MAE: 1022.6988752665341
MSE: 1576306.0841328474
R2: 0.008644825403381629

Metrics for HistGradientBoostingRegressor Approach:
Category: Nitrogens
MAE: 717.7851780645237
MSE: 838652.930171196
R2: 0.47256251156754214

Category: PMs
MAE: 518.9803440561483
MSE: 500369.62388529
R2: 0.6853123762936384

Category: Other pollutants
MAE: 887.2375987148886
MSE: 1235448.057235386
R2: 0.22301395851085315

Category: Weather
MAE: 509.77799765976647
MSE: 486435.7090715541
R2: 0.6940755592135245
```

Figure 13: Evaluation metrics for each category

SimpleImputer:
This straightforward imputation technique replaces missing values with a central tendency measure, such as the mean, median, or most frequent value of the respective feature. Its simplicity and ease of implementation make it suitable for preliminary data preprocessing tasks.

HistGradientBoostingRegressor:
Unlike SimpleImputer, HistGradientBoostingRegressor is a machine learning algorithm capable of handling missing values inherently. By integrating missing value handling directly into the training process, this algorithm enables more accurate predictions by considering the relationships between features and target variables.

Both imputation methods were employed to address missing values before fitting predictive models. While SimpleImputer employed mean imputation, median imputation, and KNN imputation, HistGradientBoostingRegressor was utilized to train predictive models directly on the dataset with missing values, leveraging its inherent capability to handle missing data.

In summary, the utilization of SimpleImputer and HistGradientBoostingRegressor for handling missing values underscores our project's commitment to robust data preprocessing and predictive modeling. These methods ensure the integrity and accuracy of the predictive models, thereby facilitating a comprehensive investigation into the correlation between air pollution and emergency hospital admissions. Through their complementary roles in data imputation and predictive modeling, SimpleImputer and HistGradientBoostingRegressor contribute to our project's success in uncovering meaningful insights into public health outcomes related to air pollution.

3.6 Results

Both approaches employed in the analysis converge on a notable correlation between predicted hospital emergency admissions within the PM pollutants category and the actual number of hospital emergency admissions. This robust correlation underscores the significance of particulate matter pollutants in influencing public health outcomes, particularly in the context of emergency hospital admissions.

The consistent agreement between the predictive models and actual admissions highlights the efficacy of the methodology in accurately capturing the relationship between air pollution and public health outcomes. These findings not only contribute to our understanding of the complex interplay between environmental factors and human health but also have practical implications for healthcare management and policy-making.

By identifying PM pollutants as a key determinant of emergency hospital admissions, this study provides actionable insights for health organizations and policymakers. Such insights can inform targeted interventions and mitigation strategies aimed at reducing air pollution levels and mitigating its adverse effects on public health.

Overall, the methodology employed in this project demonstrates its effectiveness in elucidating the correlation between air pollution and emergency hospital admissions, thereby advancing our knowledge and understanding of the intricate relationship between environmental factors and public health outcomes.

4. Conclusion

In conclusion, this project aimed to investigate the correlation between air pollution and emergency hospital admissions, addressing critical gaps in the existing literature while providing actionable insights for public health organisations. Through meticulous data preprocessing and cleansing, emphasis was placed on maintaining reproducibility and code reusability, ensuring that modifications to the data were minimised and implemented programmatically. This approach enhances the transparency and sustainability of the project, enabling future researchers to build upon our work seamlessly.

Our project sought to fill gaps identified in previous literature, such as the absence of a standardized method for evaluating short-term health effects of air pollution and the need for comprehensive air pollution data across different territories. By synthesising findings from studies like Katsouyanni et al. (1996), Milojevic et al. (2014), and Franco et al. (2020), our work contributes to a deeper understanding of the health impacts of air pollution. It highlights the urgency of addressing this public health concern.

Methodologically, our project leveraged advanced algorithms, including correlation and regression analyses, to explore the relationship between air pollution and emergency hospital admissions. The utilization of techniques like imputation and HistGradientBoostingRegressor facilitated accurate predictions and robust analysis, enabling us to uncover a significant correlation between PM pollutants and emergency hospital admissions in Bristol.

Furthermore, the project's outcomes have practical implications for health organisations in Bristol and beyond. By identifying peak periods of PM pollutant concentration, health organizations can enhance preparedness and resource allocation to address potential spikes in emergency hospital admissions. Additionally, our methodology's modularity and reusability lay the foundation for future analyses, including investigations into the temporal relationship between air pollution peaks and emergency hospital admissions.

This project underscores the importance of interdisciplinary collaboration in addressing complex public health challenges. By combining data science methodologies with domain-specific knowledge, we have shed light on the intricate relationship between air pollution and public health outcomes, paving the way for informed decision-making and targeted interventions.
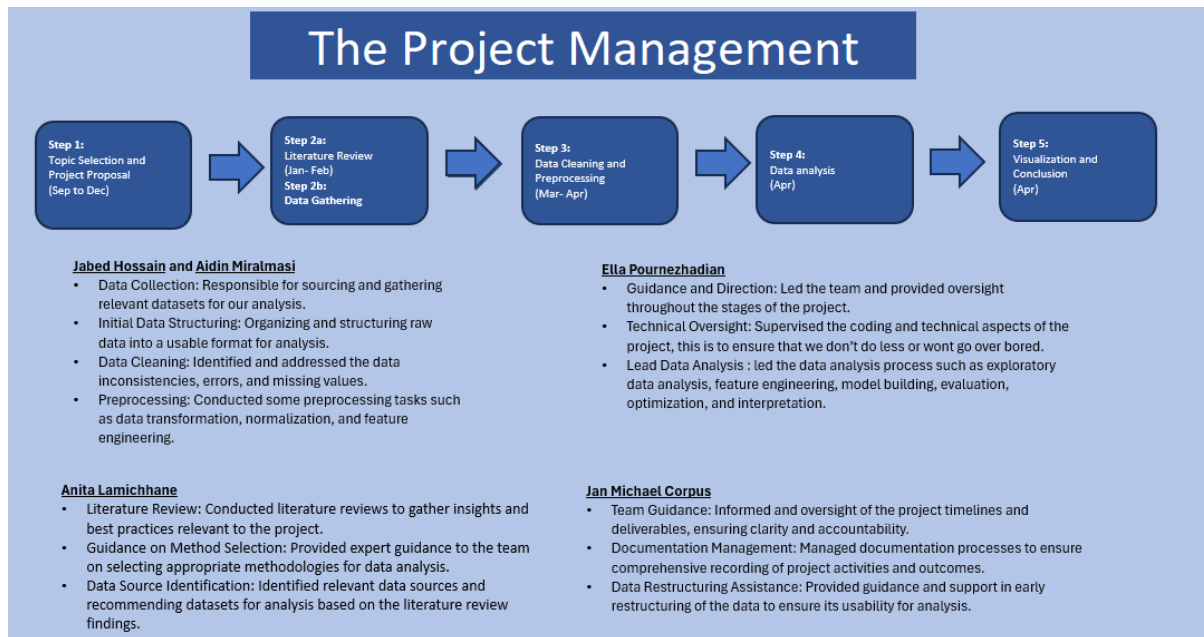
5. Project Management

Figure 14: Project management

Reflecting on our project's journey, it's evident that our methodology not only allowed us to explore the link between air pollution and emergency hospital admissions but also helped fill some key gaps in existing literature.

Initially, our goal was straightforward: find a dataset to study how air pollution affects emergency hospital visits. However, finding the right data wasn't easy. We searched through papers and even reached out to experts, but we kept hitting dead ends. Despite these challenges, we were determined to find a way forward.

As we regrouped and reassessed our approach, we realized we needed a more structured plan. We divided tasks among team members, focusing on research, data collection, and analysis. This helped us stay organized and on track.

While we faced some hurdles along the way, like figuring out how to handle the data and conduct our analysis, we didn't let them deter us. With guidance from our leaders and insights from literature reviews, we refined our methods and pressed on.

As our understanding of the dataset grew and our analytical skills improved, things started to fall into place. In a matter of weeks, we had a clear plan for processing and analysing the data, bringing us closer to answering our research question.

Looking back on our project, it's clear that teamwork, perseverance, and a structured approach were key to our success. By overcoming obstacles and staying focused on our goals, we not only addressed our research question but also contributed to the body of knowledge on air pollution and public health.

6. References:

Franco, P., Gordo, C., Marques da Costa, E., & Lopes, A. (2020). Air Pollution and Emergency Hospital Admissions—Evidences from Lisbon Metropolitan Area, Portugal. Applied Sciences, 10(22), 7997.

Katsouyanni, K., Schwartz, J., Spix, C., Touloumi, G., Zmirou, D., Zanobetti, A., ... & Anderson, H. R. (1996). Short-term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol. Journal of Epidemiology and Community Health, 50(Suppl 1), S12-S18.

Milojevic, A., Wilkinson, P., Armstrong, B., Bhaskaran, K., Smeeth, L., & Hajat, S. (2014). Short-term effects of air pollution on a range of cardiovascular events in England and Wales: Case-crossover analysis of the MINAP database, hospital admissions, and mortality.