Impact of Air Quality on Urban Health Outcomes: Exploring the Effects of PM2.5 and PM10 on Respiratory Health

Leon Smyth - Dundalk Institute of Technology

February 2025

Contents

1 Background

Airborne pollution and it's affects on public health outcomes has been extensively researched, particularly in countries such as China and India with some of the worst air pollution in the world. Health outcomes such as increased mortality, have been linked with air pollution, in diseases such as cardiovascular disease and respiratory issues, such as COPD, Asthma and reduced lung function. Particular matter (PM) are elements of the air pollution that are believed to impact health outcomes. In particular, PM2.5 (fine particular matter) and PM10 (coarse particle matter) are two such elements that are believed to increase the risk in the above mentioned diseases. Understanding and reducing both short and long-term exposure to air pollution and mitigating the impact of it on respiratory health is crucial to the health systems. This project looks at air pollution data in Ireland and compares it with India while cross referencing it with respiratory health outcomes for the same locations, in an effort to investigate the relationship between PM2.5 and PM10 in respiratory health outcomes.

1.1 Research Questions

The research questions that this paper aims to answer are:

- Is there a relationship between PM2.5 and respiratory health outcomes?
- Is there a relationship between PM10 and respiratory health outcomes?
- Can we predict the respiratory health outcomes in 2035 based on particulate matter (PM2.5 and PM10)?

1.2 Problem Definition

Despite numerous studies highlighting the impact of air pollution, there remains a need for more targeted research on the specific impact of PM2.5 and PM10 on urban respiratory health outcomes. The problem lies in the lack of localised, real-time data that can accurately capture exposure levels and their direct correlation with respiratory diseases. Additionally, disparities in society exacerbate the issue, as lower-income communities often experience higher pollution exposure and reduced access to healthcare. This research aims to investigate the relationship between PM2.5 and PM10 concentrations and respiratory health outcomes in urban populations, identifying key risk factors and mitigation strategies, and also predicting the future air quality and health outcomes.

1.3 Project Relevance in the Field of Data Analytics

Air quality and its impact on respiratory health presents a data-driven problem that is very much relevant to the field of Data Analytics. This project uses data analytics techniques such as data collection via web-scraping, data wrangling and processing, predictive modelling, and visualisation to address a critical environmental and public health issue. The following data analytics techniques will be used on this project, showing it's relevance to the field of Data Analytics:

- Data Collection - Web-scraping to extract data from environmental website, to create an air pollution dataset necessary to carry out the investigations, along with filtering from publicly available health-related data, are one of the many data analytics processes for gathering data.
- Data Wrangling and EDA - Air quality data comes from multiple sources such IoT sensors in monitoring stations, and satellite imagery. Data analytics allows for joining many datasets from multiple sources and allows for reshaping the data into the format needed for exploratory data analysis.
- Predictive Modelling for Health Risks - By analysing historical air pollution and health records, predictive models can be developed to predict respiratory health outcomes based on PM2.5 and PM10 exposure levels.
- Geospatial and Temporal Analysis - Using spatial analytics, heat-maps, and time-series forecasting, it is be possible to identify pollution dense areas and/or seasonal variations in air quality, such as that seen in Figure 1 below, which in turn could assist with planning and environmental policy.

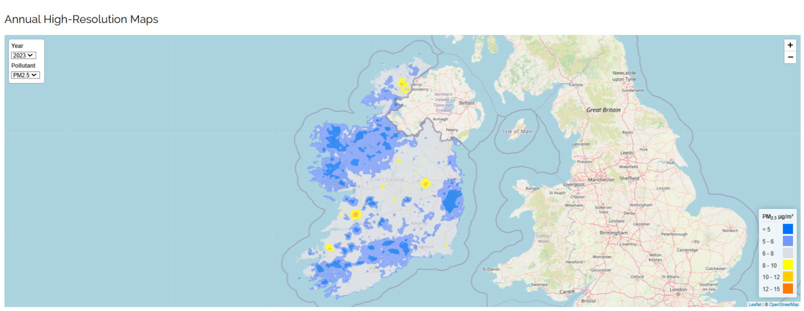

Figure 1: Annual Air Quality Map Ireland 2023((EPA) 2025)

1.4 Core Technology, Architecture and Research Processes Used

1.4.1 Core Technology

This project will predominantly involve Python as the core technology:

- For Data Collection, data is be retrieved from publicly available health datasets from the CSO and Python is used to extract the air quality data from the air quality website and parse the scraped data into a manageable dataset (BeautifulSoup). Python (Pandas) is also used to manage the large datasets from both air quality and health outcomes.

- Machine Learning Algorithms such as Regression models (e.g., linear regression, random forests) will be used to predict respiratory health outcomes based on PM2.5 and PM10 exposure.
- Statistical Analysis will be used to understand correlations between air quality levels and respiratory conditions.
- Time-Series Forecasting will be used for predicting future air quality and health outcomes.
- Python (Seaborn/Matplotlib) is used for creating visualisations and Shiny can be used for more interactive visuals, to help understand the data insights better.

1.4.2 Architecture

The main data architecture is a Python pipeline and Python visualisations. The datasets are constructed with Python using a python webscrpaer-parser, and saved to CSV files. From there they are uploaded into a second Python script that analyses and visualises the data.

1.4.3 Research Process

The research process will include the following stages:

- Literature Review: Existing research on the effects of PM2.5 and PM10 on respiratory health will be reviewed to build an understanding of existing research.
- Data Collection: Gathering real-time air quality data and health records from urban Ireland initially.
- Data Cleaning: Cleaning and preprocessing data to ensure accuracy and consistency. This includes handling missing data, outliers, and incorrect entries.
- Exploratory Data Analysis (EDA): Performing descriptive statistics, visualisations and correlations to identify patterns and relationships.

1.5 Hypothesis

The hypothesis is that increased exposure to PM2.5 and PM10 directly correlates with worsened respiratory health outcomes in urban populations. Based on this, the project aims to develop predictive models that can forecast the future air quality and estimate its impact on respiratory health.

1.6 Structure of the Report

The report will contain the following:

- Introduction overview of air pollution and its health impact
- Research questions, problem definition and objectives
- Literature Review of existing studies on PM2.5 and PM10 and their effects on health
- Data Collection and Methods, Data description, Cleaning and Exploration, Visualisation and Initial Analysis
- Future Analysis and Weekly Planner
- Appendix

2 Literature Review

2.1 Introduction

Air pollution is a global environmental health risk, especially in urban areas where industrial and vehicle emissions, among other pollutants, contribute significantly to poorer air quality. It is estimated that, each year exposure to such pollutants, contributes to millions of deaths and reduced years of healthy life (WHO 2021). According to the World Health Organisation (WHO) the impact of air pollution on human health is now comparable to risks such as unhealthy diets and smoking, and suggests that it worsens cardiovascular and respiratory diseases especially in low and middle income countries with very densely populated urban areas and high levels of industrialisation. With countries such as China and India having some of the worst air pollution in the world, there is a lot of research into particulate matter (PM) such as fine particulate matter (PM2.5) and coarse particulate matter (PM10) and it's affects on health. This literature review examines existing research on the relationship between PM2.5, PM10, and respiratory health outcomes, particularly in countries such as China and India.

2.2 Understanding PM2.5 and PM10

PM2.5 refers to fine particulate matter with a diameter equal to or less than 2.5 micrograms, while PM10 refers to particles with a diameter equal to or less than 10 micrograms, which is often referred to as coarse particulate matter (Environmental Protection Agency 2023). These airborne particles originate from sources such as vehicle emissions, industrial activities, and natural sources such as wildfires and dust storms (WHO 2021). In Ireland the main source of particulate matter is solid fuel heating (Environmental Protection Agency 2023) caused by burning solid fuel for fire where it is expelled into the air by the chimneys of a fire or solid fuel stove. Figure 2 below shows the comparable size of PM2.5 and PM10 to human hair. PM10 is at least 5-7 times smaller than human hair and PM2.2 is at least 20-28 times smaller, making them almost invisible to the naked eye.

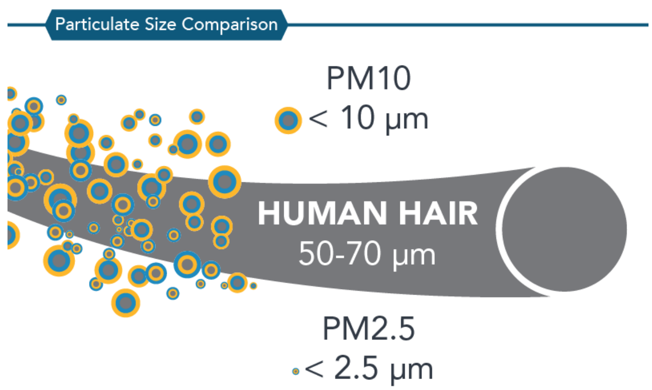

Figure 2: Particulate Size Comparison(CARB 2025)

These tiny particles can remain in the air for long periods of time, which increases the likelihood of inhalation and deep lung penetration (Sturm 2020). There have been many studies into particulate matter due to their impact on human health. The U.S. Environmental Protection Agency (EPA) introduced PM10 standards in 1987 to control the inhalation of coarse particles with PM2.5 standards added in 1997 (United States Environmental Protection Agency 2025b), when there was greater knowledge of the health risks associated with them. The most recent update was in 2024 reducing the exposure standard from 12 micrograms to 9.0 micrograms per meter cubed of air. Events such as the Great Smog of London in 1952, which lasted for five days but is known to have killed up to 12,000 people (Polivka 2018), the Chinese Airpocalypse in 2013, where it was found that there was an increased risk in all-cause, cardiovascular and respiratory emergencies (Ferreri et al. 2018), and The Great Delhi Smog in 2016/2017, which was likened to smoking 50 cigarettes a day or being in a gas chamber (Basu 2019), which also continues to have ongoing issues today, have helped to highlight the risks of particulate pollution which have all lead to stricter air quality regulations worldwide.

2.3 Health Effects of PM2.5 and PM10 on the Respiratory System

Given the tiny size of PM2.5, these particles can get deep into the respiratory system and bloodstream, causing severe health risks (Pope and Dockery

2006), which is demonstrated in their pathway linking PM exposure to cardiopulmonary morbidity and mortality as seen in Figure 3. Although PM10 is larger, inhalation can lead to increases in asthma, bronchitis, COPD, reduced lung function, including significant respiratory distress, especially among vulnerable people such as children and elderly (Brunekreef and Holgate 2002). Numerous studies have demonstrated a strong correlation between exposure to particulate matter and respiratory diseases. Studies suggest that prolonged exposure to PM2.5 can reduce lung function and exacerbate respiratory symptoms(Yang et al. 2020).

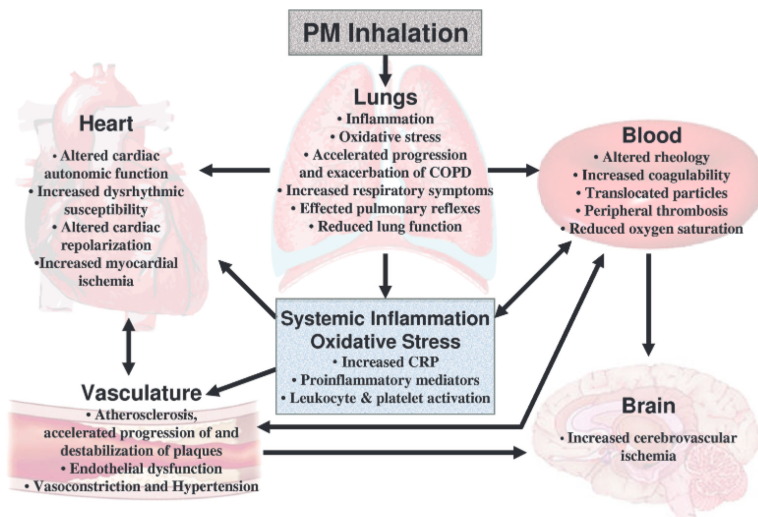

Figure 3: PM Exposure on Cardiopulmonary Morbidity/Mortality

2.3.1 Short-term Exposure Effects

Short-term exposure to high levels of PM2.5 and PM10 can trigger acute respiratory events. Studies have shown that during periods of increased air pollution, emergency department visits for asthma attacks and COPD significantly rise (Han et al. 2022). A study found that even short-term exposure to elevated PM concentrations showed an increased risk in cardiovascular, respiratory and cerebrovascular mortality (Orellano et al. 2020). In Delhi, India during the Diwali Fest the fireworks alone cause excessive air pollution, and has been reported to cause between 32 and 75 early deaths per day during this period (Y. Chen et al. 2020).

2.3.2 Long-term Exposure Effects

Long-term exposure to particulate matter has more profound and chronic implications. Research by (Pope and Dockery 2006) revealed that long-term exposure to particulate matter is associated with reduced life expectancy due to its many respiratory and cardiovascular implications, including but not limited to, rapid progression of COPD, and exacerbating existing pulmonary disease, cardio-vascular disease, lung damage, decreased lung function, respiratory distress, and hypoxaemia. Another study by (Guo et al. 2019) found that higher levels of PM2.5 in Taiwan, over a two year period, was linked to worsening lung function and more cases of poor lung health in children and adolescence. However, more research is needed to look into the relationship between the specific sources of the particulate matter and health outcomes to identify if some sources have more or less favourable health outcomes.

2.4 Mitigation Strategies and Policy Interventions

Urban populations are disproportionately affected by air pollution due to higher population density, higher traffic density, industrial zones, and less green spaces. There have been efforts to mitigate the effects of PM2.5 and PM10 that include things such as Air Quality Standards; such as the EPA National Ambient Air Quality Standards (NAAQS) (United States Environmental Protection Agency 2025a) in the US and the EU Ambient Air Quality Directive (European Union 2024). Stricter emissions regulations, as seen in the NAAQS, urban green space development, as seen by the CARB, and improved public transportation systems to reduce vehicle pollution are all strategies taken in an effort to reduce air pollution.

Cities that implement air quality monitoring systems and public awareness campaigns have seen measurable reductions in pollution-related health issues, such as Toronto, where public alerts reduced ER visits by 25% per day for asthma-related issues (H. Chen et al. 2018). However, there are shortcomings in the number of locations that monitor air quality, with the majority of the focus being on major cities, leaving many other locations without air quality monitoring. There have been further measures to attempt to mitigate air pollution, such as replacing smokey coal with smokeless, providing alternative solutions to fossil fuel heating, such as heat pumps. Additionally, individual protective measures such as air purifiers, wearing masks, and avoiding outdoor activities during high-pollution periods, have been recommended (Carlsten et al. 2020).

2.5 Research Gaps

Research gaps exist regarding long-term effects, and specific pollutant components and morbidity of PM2.5 and PM10 exposure. While there is a large amount of research to link the impact of PM2.5 and PM10 exposure to health implications, it is lacking in specific classes of air pollution that make up the particulate matter and the specific affects of each individual component of it.

There is also a lack of complete nationwide air pollution monitoring in each country, (WHO 2021), as the main cities are more closely monitored, which in return limits the data available for research.

2.6 Conclusion

In conclusion, this literature review highlights the significant health risks associated to air pollution, particularly from fine and coarse particulate matter (PM2.5 and PM10). Research shows a strong link between exposure to these pollutants and respiratory diseases, with short-term exposure leading to acute health events like asthma attacks and long-term exposure contributing to chronic respiratory and cardiovascular conditions. Countries with high levels of industrialisation and urban density, such as China and India, are especially vulnerable, experiencing severe air pollution and widespread health impacts. Historical events like the Great Smog of London and more recently the air pollution crises in Delhi, emphasise the need for air quality regulations and ongoing research.

Strategies to mitigate air pollution include government policies, such as stricter emissions regulations and the promotion of greener urban areas, have shown positive impacts on public health. However, significant research gaps remain, particularly in the areas of long-term effects of specific pollutant components and issues with gaps in air quality monitoring, specifically outside major cities. Addressing these gaps will require expanding monitoring networks, targeted studies on pollutant sources, and a focus on vulnerable populations.

References

(EPA), Environmental Protection Agency (2025). *Annual High-Resolution Air Quality Maps*. Accessed: 2025-02-23. URL: `https://airquality.ie/annual-maps`.

Basu, Mausumi (May 2019). "The great smog of Delhi". In: *Lung India* 36.3, pp. 239–240. DOI: `10.4103/lungindia.lungindia_363_18`.

Brunekreef, Bert and Stephen T. Holgate (2002). "Air pollution and health". In: *The Lancet* 360.9341, pp. 1233–1242. ISSN: 0140-6736. DOI: `10.1016/S0140-6736(02)11274-8`. URL: `https://doi.org/10.1016/S0140-6736(02)11274-8`.

CARB, California Air Resources Board (2025). *Inhalable Particulate Matter and Health*. Accessed: 15 February 2025. URL: `https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and-health`.

Carlsten, Christopher et al. (2020). "Personal strategies to minimise effects of air pollution on respiratory health: advice for providers, patients and the public". In: 55.6. DOI: `10.1183/13993003.02056-2019`.

Chen, Hong et al. (2018). "Effect of air quality alerts on human health: a regression discontinuity analysis in Toronto, Canada". In: *The Lancet Planetary Health* 2.1, e19–e26. DOI: `10.1016/S2542-5196(17)30185-7`.

Chen, Ying et al. (2020). "Local characteristics of and exposure to fine particulate matter (PM2.5) in four indian megacities". In: *Atmospheric Environment: X* 5, p. 100052. ISSN: 2590-1621. DOI: https://doi.org/10.1016/j.aeaoa.2019.100052. URL: https://www.sciencedirect.com/science/article/pii/S2590162119300553.

Environmental Protection Agency (2023). *Air Quality in Ireland 2023*. Accessed: 2025-02-17. URL: https://www.epa.ie/publications/monitoring--assessment/air/Air_Quality_Report_23_v14.pdf.

European Union (2024). *DIRECTIVE (EU) 2024/2881 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2024 on ambient air quality and cleaner air for Europe*. Accessed: 2025-02-20. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202402881.

Ferreri, Joshua M et al. (2018). "The January 2013 Beijing "Airpocalypse" and its Acute Effects on Emergency and Outpatient Visits at a Beijing Hospital". In: *Air Quality, Atmosphere, Health* 11.3, pp. 301–309. DOI: 10.1007/s11869-017-0538-0.

Guo, Cui et al. (2019). "Long-Term Exposure to Ambient Fine Particulate Matter (PM2.5) and Lung Function in Children, Adolescents, and Young Adults: A Longitudinal Cohort Study". In: *Environmental Health Perspectives* 127.12, p. 127008. DOI: 10.1289/EHP5220.

Han, Chang Hoon et al. (2022). "Short-term Effects of Exposure to Particulate Matter on Hospital Admissions for Asthma and Chronic Obstructive Pulmonary Disease". In: *Medicine* 101.35, e30165. DOI: 10.1097/MD.0000000000030165.

Orellano, Pablo et al. (2020). "Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: Systematic review and meta-analysis". In: *Environment International* 142, p. 105876. ISSN: 0160-4120. DOI: https://doi.org/10.1016/j.envint.2020.105876. URL: https://www.sciencedirect.com/science/article/pii/S0160412020318316.

Polivka, Barbara J. (Apr. 2018). "The Great London Smog of 1952". In: *AJN, American Journal of Nursing* 118.4, pp. 57–61. DOI: 10.1097/01.NAJ.0000532078.72372.c3.

Pope, C. Arden and Douglas W. Dockery (2006). "Health Effects of Fine Particulate Air Pollution: Lines that Connect". In: *Journal of the Air & Waste Management Association* 56.6, pp. 709–742. DOI: 10.1080/10473289.2006.10464485. eprint: https://doi.org/10.1080/10473289.2006.10464485. URL: https://doi.org/10.1080/10473289.2006.10464485.

Sturm, Robert (2020). "Modelling the deposition of fine particulate matter (PM2.5) in the human respiratory tract". In: *AME Medical Journal* 5.0. ISSN: 2520-0518. URL: https://amj.amegroups.org/article/view/5437.

United States Environmental Protection Agency (2025a). *National Ambient Air Quality Standards (NAAQS) Table*. Accessed: 2025-02-20. URL: https://www.epa.gov/criteria-air-pollutants/naaqs-table.

— (2025b). *Timeline of Particulate Matter (PM) National Ambient Air Quality Standards (NAAQS)*. Accessed: 2025-02-18. URL: https://www.epa.gov/

pm-pollution/timeline-particulate-matter-pm-national-ambient-air-quality-standards-naaqs.
WHO, World Health Organisation (2021). *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide.* Accessed: 2025-02-14. URL: https://iris.who.int/bitstream/handle/10665/345334/9789240034433-eng.pdf.
Yang, Mo et al. (2020). "Is PM1 similar to PM2.5? A new insight into the association of PM1 and PM2.5 with children's lung function". In: *Environment International* 145, p. 106092. ISSN: 0160-4120. DOI: https://doi.org/10.1016/j.envint.2020.106092. URL: https://www.sciencedirect.com/science/article/pii/S016041202032047X.

3 Exploration of Data and Methods

3.1 Data Collection and Methods of Data Collection used

There are a number of data sources needed for this project, a data source that measures air quality and another that measures health outcomes for both Ireland and India. The data collection method for health outcomes for Ireland was complete through the CSO wesbite at https://data.cso.ie/, where the data for "Total Hospital Discharges by Principle Diagnoses and Area" (County) was located. The data was initially filtered on diseases relevant to the study and the full dataset from 2017 - 2013 was downloaded to a CSV file.

The data collection method for air quality for Ireland was more difficult to extract. Using the EPA wesbite at https://airquality.ie/readings, it was noticed when applying the dates to the calendar that the following URL:

'https://airquality.ie/readings?station=EPA-54&dateFrom=01+Jan+2023&dateTo=30+Jun+2023' was used to extract 6 months data to a web page line chart for one EPA location, which meant the parameters were visible which allowed a web scraper to be built to capture and parse the data over all stations over longer periods. The Web Scraper involved building a Python script using the beautifulSoup package to capture the HTML and JavaScript on the webpage. Then, searching the response for the location name and extracting it, along with all PM2.5 and PM10 data, where applicable, and saving to CSV files which could be read back into the analysis script.

3.2 Procedure of Ethical Approval Undertaken

Ethical approval was complete by submitting the ethical clearance declaration form, co-signed by the project supervisor. External approval was not necessary as the data was already publicly available data and no studies were complete on actual participants.

3.3 Data Description

There are two sources of data in the initial exploration and analysis (for Ireland initially), one from the EPA on air quality and the other extracted from the CSO, for Respiratory Health Hospital Discharges.

Table 1: Air Pollution Monitoring (PM2.5 and PM10 values) - Ireland

Variable	Description	Data Type
Timestamp	Date-time when measures were recorded	dateTime
Value	Measure value	Float
PM	Particulate Matter Type (PM2.5 or PM10)	Float
Location	Location of Air Pollution Monitor	Object

Table 2: Respiratory Health Diagnosis - Ireland

Variable	Description	Data Type
Statistic Label	Description of data (All Hospital Discharges Or Rate per 1,000 Population)	Object
Year	Year of the value reported	Int
Sex	Sex of the value reported	Object
ISHMT	Primary Diagnosis in Hospital	Object
Area	County	Object
UNIT	Number	Object
VALUE	Total Numbers discharged or discharged rate per 1,000 population	Float

3.4 Data Cleaning and Exploration

The Air Pollution Monitoring data was extracted to 4 CSV files, one for the first 6 months of 2023, the last 6 months in 2023, the year of 2022 and the last one for the year of 2021. All four files were read into a python script and merged together into one data frame. Timestamp was converted into an actual timestamp data type and year was extracted as a variable from Timestamp for later aggregation by year. County was extracted from the location, but needs more work as the initial exploration was done on Dublin and so the focus was on the Dublin data. To do a more widespread analysis the county field would need additional cleaning and preparation. Given that Dublin had many different locations, Dublin 15 (the Blanchardstown location) had the most data points and was used to represent Co. Dublin in this initial exploration. Dropped the timestamp, location and county fields to aggregate the data by PM and mean value to get a yearly mean value for PM2.5 and PM10 for Dublin.

The data for the Hospital Discharges by Principle Diagnoses and Area, was first filtered by the following respiratory health outcomes as these were the health outcomes that were deemed most appropriate for respiratory health analysis:

- Acute upper respiratory infections & influenza
- Asthma
- Chronic obstructive pulmonary disease & bronchiectasis
- Other acute lower respiratory infections
- Other diseases of the respiratory system
- Other diseases of upper respiratory tract
- Pneumonia

The data was then filtered by 'All Hospital Discharges' to get the total number of discharges by diagnosis as it is better for the aggregation that would come later in the analysis. All the aggregated totals along with any totals that belonged an 'Area' that was not a county as this could not be joined to the air quality data. This data was then aggregated with the air pollution data to explore the findings.

3.5 Data Visualisation

Figure 4 below shows the total number of hospital discharges for male and female patients, where each year represents data points across all seven diagnosis, showing a clear drop in discharges for these specific diagnosis in 2020 and 2021 during the 'Covid' years. The chart also shows that there are more males than females discharged from hospital with these diagnosis.

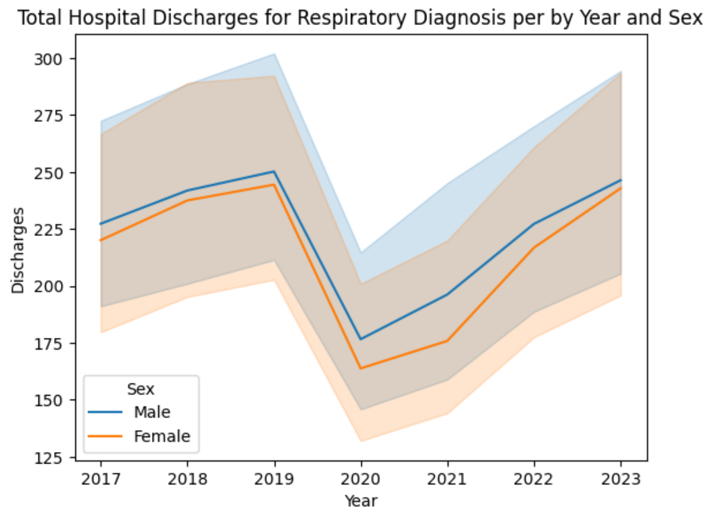

Figure 4: Hospital Discharges 2017 - 2023 (Respiratory Diagnosis)

Looking at Co. Dublin only, Figure 5 below shows the total number of hospital discharges for all patients from 2017 to 2023 for all seven diagnosis. Pneumonia and Chronic Obstructive Pulmonary Disease (COPD) are the two diagnosis with the most discharges in each year. Acute upper respiratory infections and influenza have the least discharges in each year. Pneumonia spiked during 2020 and even more so during 2021, during the COVID years, whereas the other diagnosis reduced during both of these years. During both 2020 and 2021, Pneumonia diagnosis accounted for more than double any of the other diagnosis. While COPD is one of the main diagnosis each year, it also reduced during the COVID years.

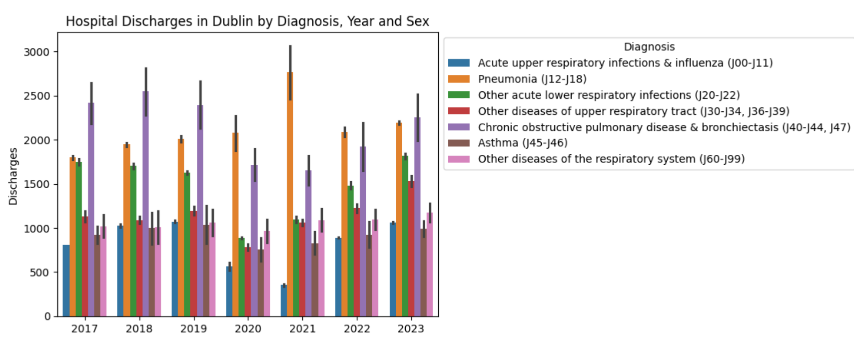

Figure 5: Breakdown of Total Hospital Discharges 2017 - 2023 (Respiratory Diagnosis)

Figure 6 shows the mean particulate matter concentrations for PM2.5 and PM10 in micrograms per meter cubed (µg/m$^{\mathbf{3}}$) in Dublin from 2021 to 2023. In 2022 both PM2.5 and PM10 concentrations increased on 2021 and in 2023 they reduced again on both 2021 and 2022 figures.

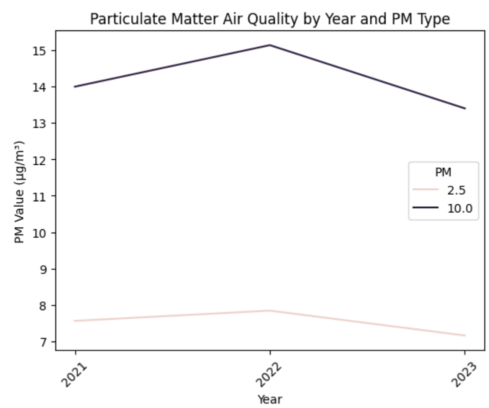

Figure 6: Mean Particulate Matter PM2.5 and PM10 in Dublin - 2023

3.6 Preliminary Analysis

Figure 7 shows the mean particulate matter concentrations for PM2.5 and PM10 in micrograms per meter cubed (µg/m$^{\mathbf{3}}$) in Dublin from 2021 to 2023 with the total number of hospital discharges for respiratory diagnosis. On initial inspection it would appear that as PM2.5 and PM10 concentrations rise, so to do the hospital discharges where the primary diagnosis was one of the seven diagnosis specified, however during 2023 it would appear that both PM2.5 and PM10 concentrations dropped while the total number of hospital discharges increased.

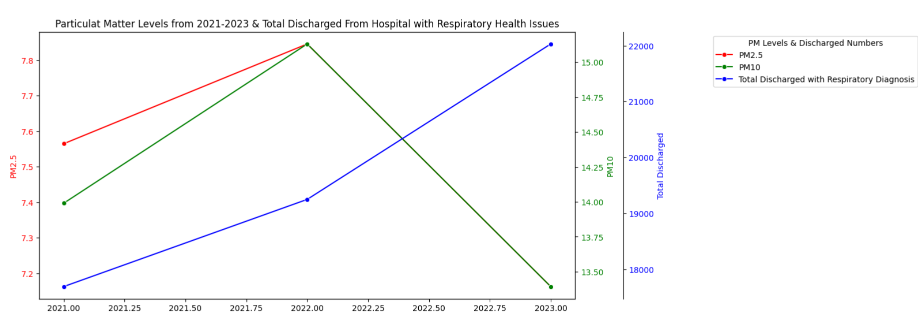

Figure 7: Mean Particulate Matter PM2.5 and PM10, with Respiratory Hospital Discharges, Dublin - 2023

4 Proposed Future Analysis

As the initial exploratory data analysis (EDA) was preliminary in Dublin and only in Ireland, the future proposed analysis will be more comprehensive, to include other counties for a comparison within Ireland. It will also include both health outcomes and air quality data from India. This will enable a direct comparison between Ireland and India in both PM2.5 and PM10 concentrations and investigate the relationship between these particulate matters and hospital diagnosis in both countries. Extracting additional air quality data for Ireland to allow the data modelling to go back to 2017 in line with the hospital diagnosis data, would allow forecasting to a future date where predicting the health outcomes for 2035 would be based on the predicted air quality.

In addition the hospital data is aggregated at source to a yearly figure, it would be more appropriate to have this data by day, week or month granularity as the air quality data is per hour and it would be easier to see the health outcomes if they were more aligned on a similar timeline. Should this data become available it would allow for a more accurate analysis of a potential link between PM2.5 and PM10. Table 3 below contains the plan to progress the project, which is a week-by-week task planner, included weekly are supervisor check-ins.

4.1 Weekly Planner

Table 3: Weekly Planning Tasks

Week No.	Description
1	Welcome project feedback and update weekly planner to include tasks that may have been missed. Add additional Air Quality data to existing dataset to go back as far as 2017, investigate if possible to get the health data to a more granular level.
2	Get data on India Air Quality and Health Outcomes from 2017 - 2023
3	Look at dissertation and outline the final report structure considering feedback given
4	Clean the new data and wrangle it into a useable format
5	Additional Exploratory Analysis, considering all-Ireland and India, to explore the impact of PM2.5 and PM10 on respiratory health outcomes
6	Start working on data model to predict the air quality and respiratory health outcomes in 2035
7	Fine Tune and Finish Analysis
8	Continue to Write Dissertation
7	Iterate over any additional Analysis needed
9	Continue to Write Dissertation
10	Complete Dissertation
11	Screen-cast and anything else outstanding
12	Allow last week for delays or any unforeseen issues

5 Appendix

5.1 ChatGPT Prompts

- give me the python regex to parse this date and value, into year, month, day, hour, minute, second, [Date.UTC(2023,0,9,4,0,0), 2.86]
- how to offset a third axis on an sns linechart
- how do I make a legend for all 3 axis(ax1, ax2, ax3) on one chart
- how do I hide legends for axis 1 and 2

5.2 Ethics Form