Analysis of Wildfire Smoke and Its Impact on Asthma in Philadelphia, PA

DATA 512 AU 2024 - Course Project Report

ABHINAV DUVVURI

Table of Contents

1. Introduction

Wildfire smoke is more than just a hazy inconvenience—it's a growing public health concern. Over the past few years, there have been many news reports in my assigned location - Philadelphia, PA suggesting that smoke might be causing an increase in asthma cases, especially in vulnerable populations like children and older adults.

Philadelphia County, while not a hotspot for wildfires itself, frequently experiences the effects of wildfire smoke from other regions. Smoke from Canadian wildfires, in particular, often drifts southward during wildfire season, carried by prevailing winds. These distant fires can release massive amounts of particulate matter and pollutants, significantly impacting air quality in Philadelphia. This smoke lingers in the atmosphere, reducing visibility and posing health risks, especially for people with asthma or other respiratory conditions. The periodic influx of wildfire smoke makes it a critical concern for public health planning in the county.

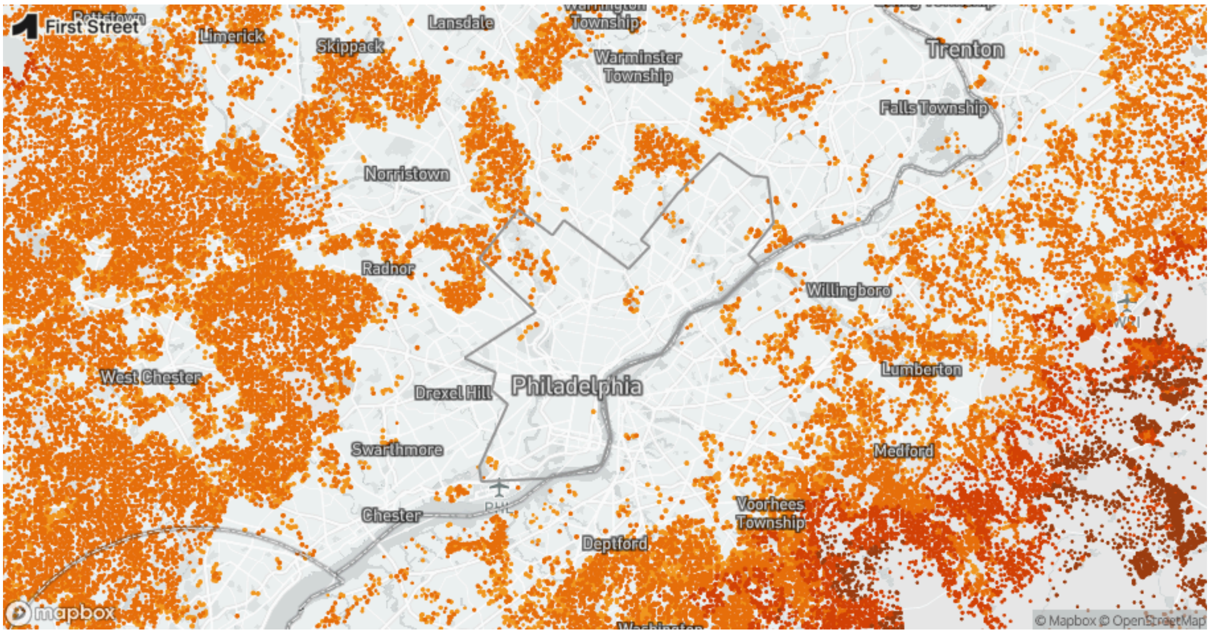

Fig 1: Wildfires near Philadelphia county

This inspired me to dig deeper into the issue. I wanted to verify if there was a real connection between wildfire smoke and asthma rates in our area. If this connection exists, what could it mean for the future? Will asthma cases rise as smoke levels increase, or could interventions help reduce the impact? These were some of the questions I had and wanted to answer.

To begin, I developed an initial smoke estimate using data from the US Geological Survey. This estimate combines acres burned, fire severity, and the distance of fires around the Philadelphia region. Analysis revealed that local smoke levels are influenced by both nearby wildfires and distant events, such as Canadian wildfires, which contributed to noticeable spikes in smoke levels. This provided a foundation to explore how these smoke patterns relate to asthma trends.

Using this smoke data, I analysed asthma-related hospitalization and death records to identify patterns and build forecasts. My goal was to understand the problem and create scenarios that show how changes in wildfire smoke—whether increases or decreases—might affect hospitalizations and deaths caused by asthma.

This project isn’t just about the numbers. It’s about understanding the serious health risks associated with wildfire smoke and providing actionable insights that could help communities prepare, protect vulnerable groups, and create strategies for a healthier future.

2. Background

Wildfire smoke has emerged as a major public health concern, with its far-reaching effects on air quality and respiratory health becoming increasingly evident. The smoke produced by wildfires contains a complex mixture of harmful pollutants, including fine particulate matter (PM2.5), carbon monoxide, volatile organic compounds (VOCs), and other toxic substances. These pollutants are particularly dangerous due to their ability to penetrate deep into the lungs and even enter the bloodstream, where they can cause or worsen respiratory and cardiovascular issues. Liu K et al, has linked exposure to PM2.5, increased risks of asthma attacks, chronic obstructive pulmonary disease (COPD), and other respiratory ailments. Further, Wilgus, M et al, talk about how wildfire smoke is causing an increase in the number of COPD and asthma related cases.

While Philadelphia County is geographically distant from fire-prone regions, the county is not insulated from the effects of wildfire smoke. The increasing frequency and intensity of wildfires, driven by climate change, have led to longer-lasting and more severe smoke events around the area, with pollutants traveling hundreds or even thousands of miles. As a densely populated urban area with existing air quality challenges, Philadelphia is particularly vulnerable to these distant environmental events. Vulnerable populations, such as children, older adults, and those with pre-existing respiratory conditions, face higher risks, as prolonged smoke exposure can lead to increased hospitalizations, worsening of chronic illnesses, and in some cases, premature death.

By analyzing historical trends in wildfire smoke exposure and its impacts on asthma cases and hospitalizations, this study aims to shed light on the broader implications of wildfire activity for urban populations. The findings will not only help in understanding the specific risks to Philadelphia but also inform strategies to mitigate the health effects of worsening air quality in the face of increasing wildfire activity.

Asthma is particularly relevant to this study because it disproportionately affects vulnerable populations, including children, older adults, and individuals with pre-existing health conditions. In children, asthma is one of the leading causes of hospitalizations and missed school days, while in older adults, it increases the risk of severe complications and mortality. Even short-term exposure to elevated PM2.5 levels from wildfire smoke has been shown to trigger asthma attacks, increase emergency room visits, and, in severe cases, lead to respiratory failure.

The CDC Multiple Causes of Death dataset is valuable for analyzing asthma-related deaths as it provides detailed mortality data, including specific causes of death, across the U.S. This helps identify trends and understand the impact of smoke on asthma fatalities. Similarly, the PHC4 dataset is useful for studying asthma hospitalizations in children under 18, as it includes comprehensive hospitalization records for Pennsylvania. This data allows for a focused analysis of how smoke exposure affects pediatric health outcomes in Philadelphia, giving insights into patterns and potential interventions.

Philadelphia County is impacted by smoke from distant sources such as Canadian wildfires (see figure 1), which often drift southward during wildfire season. Data from the US Geological Survey and air quality monitoring agencies have revealed spikes in smoke levels in the region, correlating with wildfire activity in Canada and surrounding areas. This transboundary smoke can exacerbate local air pollution, compounding health risks for residents.

Addressing asthma-related risks is crucial for understanding and mitigating the broader health impacts of wildfire smoke as well. As climate change drives more frequent and severe wildfires, the likelihood of increased smoke exposure is rising. Proactive measures to study these effects can provide valuable insights for public health strategies, particularly in urban areas like Philadelphia, where population density amplifies the potential for widespread health consequences.

For forecasting asthma hospitalizations and deaths, many models were available, each with distinct strengths. Linear regression could provide a simple relationship between smoke levels and health outcomes, while machine learning models like Random Forests or Gradient Boosting could capture complex nonlinear patterns. Time series models like ARIMA or Prophet are particularly suited for trends and seasonality, with ARIMA handling univariate data and Prophet excelling in flexibility and interpretability. However, I chose the SARIMAX model because it combines the strengths of ARIMA for time series forecasting with the ability to incorporate exogenous variables, such as smoke levels, that directly influence asthma outcomes. This allowed for a more nuanced understanding of how changes in smoke estimates impact health trends over time. Key considerations included the availability of historical data, the interpretability of the model for public health stakeholders which will be useful while explaining to the city council. SARIMAX met these criteria by effectively capturing both temporal dependencies and external factors, making it ideal for this analysis.

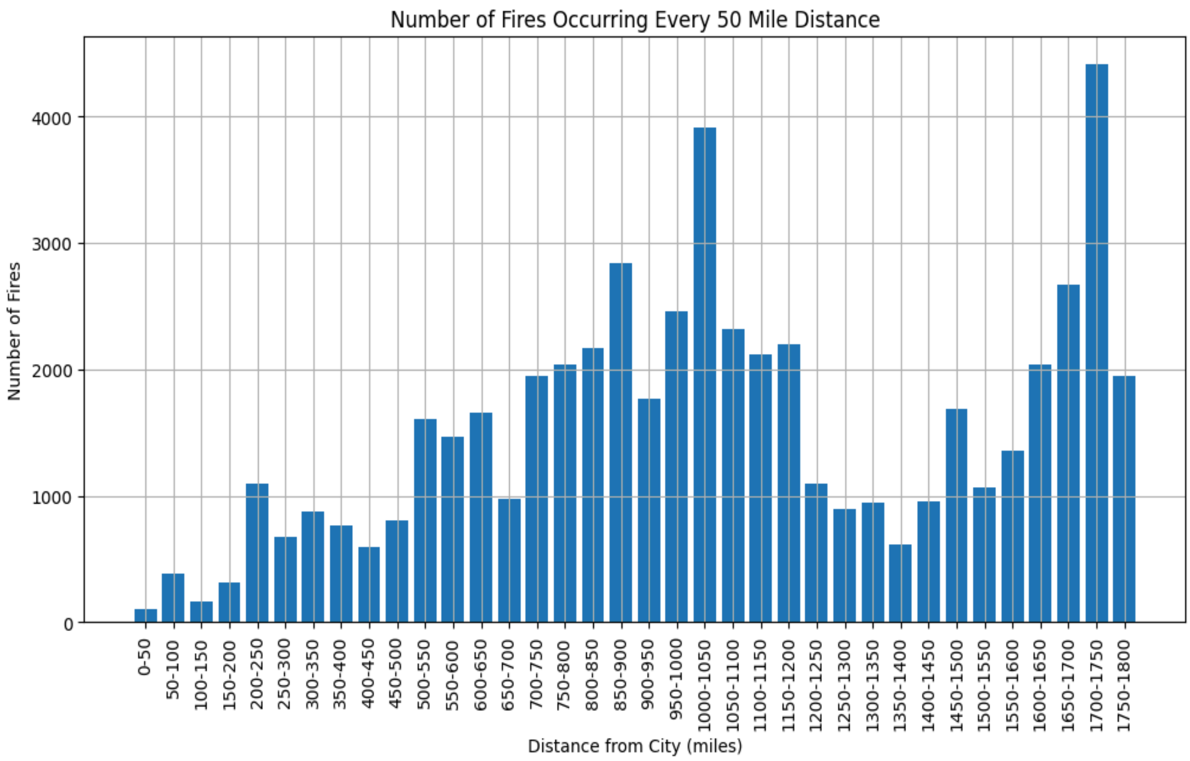

Figure 2: Histogram showing the number of fires occurring every 50 miles in Philadelphia

3. Methodology

For this analysis, the methodology used was primarily statistical modelling and forecasting while adhering to human centered principles to explore the relationships between wildfire smoke, air quality and Asthma hospitalizations and deaths.

a. Handling wildfire smoke data

The city assigned for this project is Philadelphia, PA. To analyze the relationship between wildfire smoke and air quality in Philadelphia, the first step was to calculate the distance between the city and the location of various wildfires. The latitude and longitude coordinates of Philadelphia were sourced from Wikipedia, and the distances were computed. Wildfires, being irregular in shape, can make the concept of "distance" complex. Therefore, two approaches were employed. The first calculated the shortest distance from the city to the fire's perimeter, while the second considered the average distance of all points on the perimeter to the city. Both methods were retained during initial calculations, with the final choice made later based on analysis requirements.

b. Handling AQI Data

Next, the Air Quality Index (AQI) data for Philadelphia was gathered using an EPA-provided API. AQI is a standardized measure of air quality based on pollutant levels, categorized from "Good" to "Hazardous." Historical data from 1961 to 2024 was collected. The AQI data was carefully processed, and invalid records were excluded using the "validity indicator" field. Both gaseous and particulate data were merged into a single dataset, and further data cleaning included grouping by site and date, then aggregating pollutant AQI by their maximum values, as the most toxic pollutant determines the site's overall air quality.

c. Creating a smoke estimate

To estimate smoke in Philadelphia caused by wildfires, a formula incorporating multiple factors was developed. This included the total area burned (GIS_acres), fire severity (weighted by squaring its value), and the inverse of the shortest distance from the fire to the city. The formula aimed to capture the scale and proximity of fires, as well as their severity, to estimate smoke levels effectively. Records of fires farther than 650 miles from Philadelphia were excluded, in line with the assignment guidelines.

```
SmokeEstimate = (AcresBurnt * FireSeverity^2) / Distance
```

d. Comparing AQI data and smoke estimates

Next, the data was visualized to understand trends. A histogram showed the number of fires (figure 2) occurring within different distance intervals from Philadelphia, revealing more fires farther from urban areas. Time-series graphs depicted total acres burned per year and the estimated smoke levels alongside AQI data for Philadelphia. These visualizations highlighted historical trends, including the limited data accuracy before 1985 and potential correlations between AQI and smoke estimates in later years.

e. Forecasting smoke

To predict future smoke estimates, an ARIMAX model was chosen. This forecasting model combines historical trends with external variables, such as wildfire data, to enhance prediction accuracy. The ARIMAX model was trained on data between 1960 and 2020, incorporating variables like fire area and severity. Model diagnostics, including residual plots and p-values, confirmed the model's validity, enabling reliable forecasts of smoke estimates through 2050. An ARIMAX model is very easy to interpret and hence will be easier to explain to non-technical folks. Hence, it was chosen. Explainability of models is also an important human-centric design consideration.

f. Extending smoke estimate to asthma analysis

To extend my previous smoke forecast from 2020-2050, I examined its social impacts by focusing on asthma-related trends. Specifically, I wanted to understand how changes in smoke levels might affect hospitalizations for children under 18 and deaths due to asthma. I used three scenarios for smoke levels: current trends, a 50% reduction, and a 70% reduction. All data containing aggregation counts less than 10 were dropped from the dataset since that would aid in better data privacy and anonymity - an important human centric consideration.

g. Handling Asthma hospitalization Data and Death Data

I collected data on asthma deaths from the CDC Wonder database and hospitalization data for children from PHC4, which is hosted by Pennsylvania’s EDDIE platform. Both datasets were cleaned, combined, and prepared for analysis. I then merged this information with my smoke forecast data to explore how changes in smoke levels could influence asthma outcomes. As before, All data containing aggregation counts less than 10 were dropped from the dataset since that would aid in better data privacy and anonymity - an important human centric consideration.

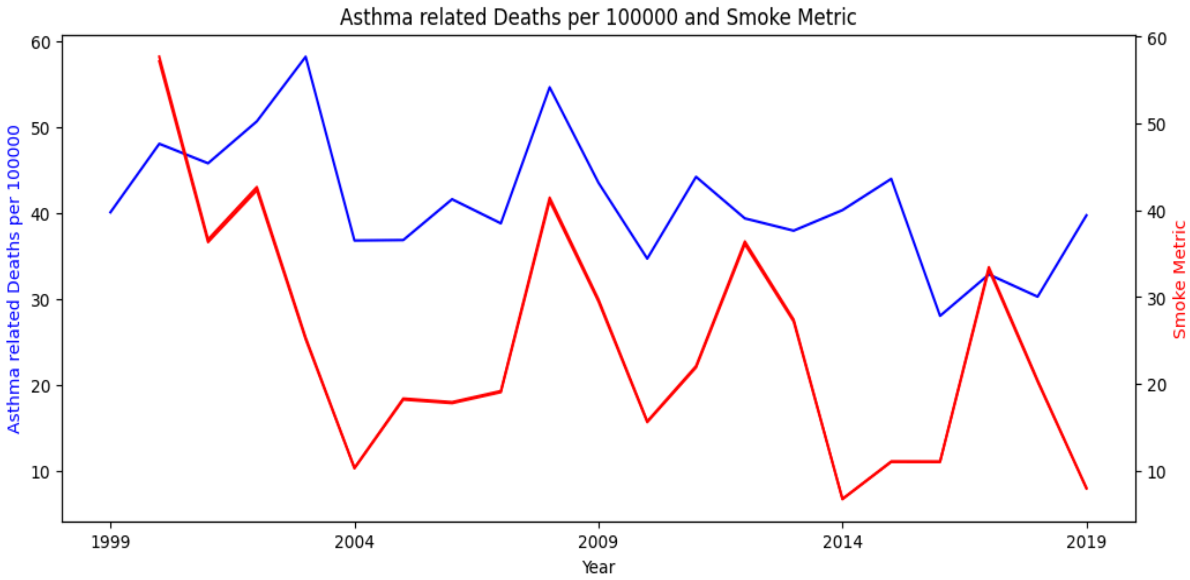

Figure 3: Asthma Deaths and Smoke metric in the same plot. There seems to be some correlation

h. Modelling and forecasting asthma hospitalizations and Deaths

To analyze the relationship between smoke exposure and asthma outcomes, I developed two SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors) models for both hospitalizations and deaths. I incorporated smoke estimates as exogenous variables, with asthma-related hospitalizations for children under 18 and asthma-related deaths serving as dependent variables.

I split the data into training and testing sets to ensure model robustness. I chose SARIMAX models for their capacity to capture seasonality, trends, and the influence of external variables. After model validation, I generated forecasts under three scenarios: current smoke trends, 50% smoke reduction, and 70% smoke reduction, enabling assessment of potential public health impacts under different smoke mitigation strategies.

A SARIMAX model is very easy to interpret and hence will be easier to explain to non-technical folks and the city council. Hence, it was chosen. Explainability of models is also an important human-centric design consideration.

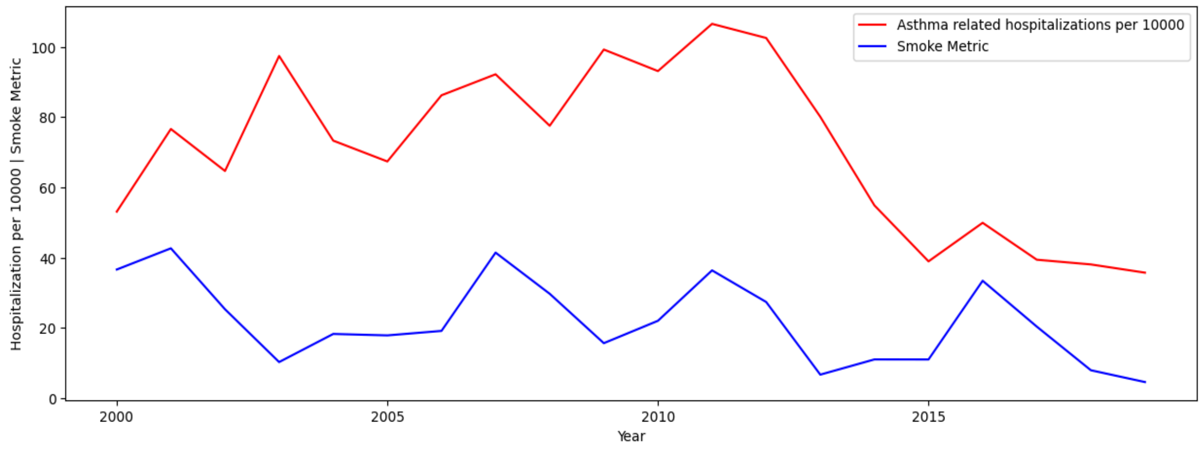

Figure 4: Asthma Hospitalization in kids(<18 years) and Smoke metric in the same plot. There seems to be some correlation

4. Findings

The analysis revealed several key interesting patterns. The histogram of fire occurrences (figure 2) showed that most fires occurred at greater distances from the city, likely due to urban areas having less forest cover and fewer wildfires. Most of the fires seem to have occurred in Canada as can be seen in the spikes in the graph. Fires within 650 miles of Philadelphia were included in the smoke estimation model, as specified in the assignment guidelines.

The time-series graph of acres burned per year highlighted a significant increase in wildfires since the late 1980s, aligning with improved data recording from that period onward and showing the growing issue of wildfires. Peaks in wildfire activity were observed between 1985 and 1995, though the underlying causes remain unclear. This increase in wildfire activity also coincided with a rise in smoke estimates for Philadelphia, demonstrating a clear link between wildfire scale and potential air quality impacts.

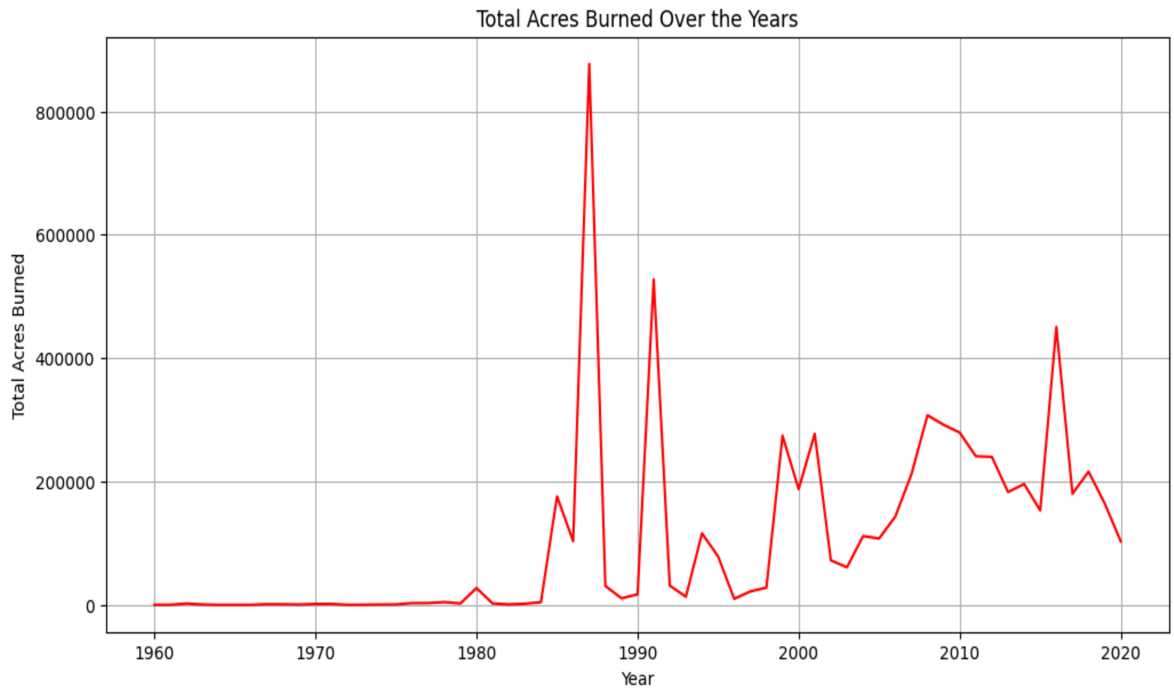

Figure 5: Total number of acres burnt over the years

When comparing the smoke estimates with AQI data, it became evident that early data (1960–1985) lacked sufficient reliability for meaningful correlations. However, from 1995 onward, the trends between smoke estimates and AQI began to align more closely (figure 6), suggesting that wildfires significantly influenced air quality during this period. Interestingly, in some cases, increases in smoke estimates corresponded to AQI rises in subsequent years, potentially reflecting the delayed effects of smoke on urban air quality.

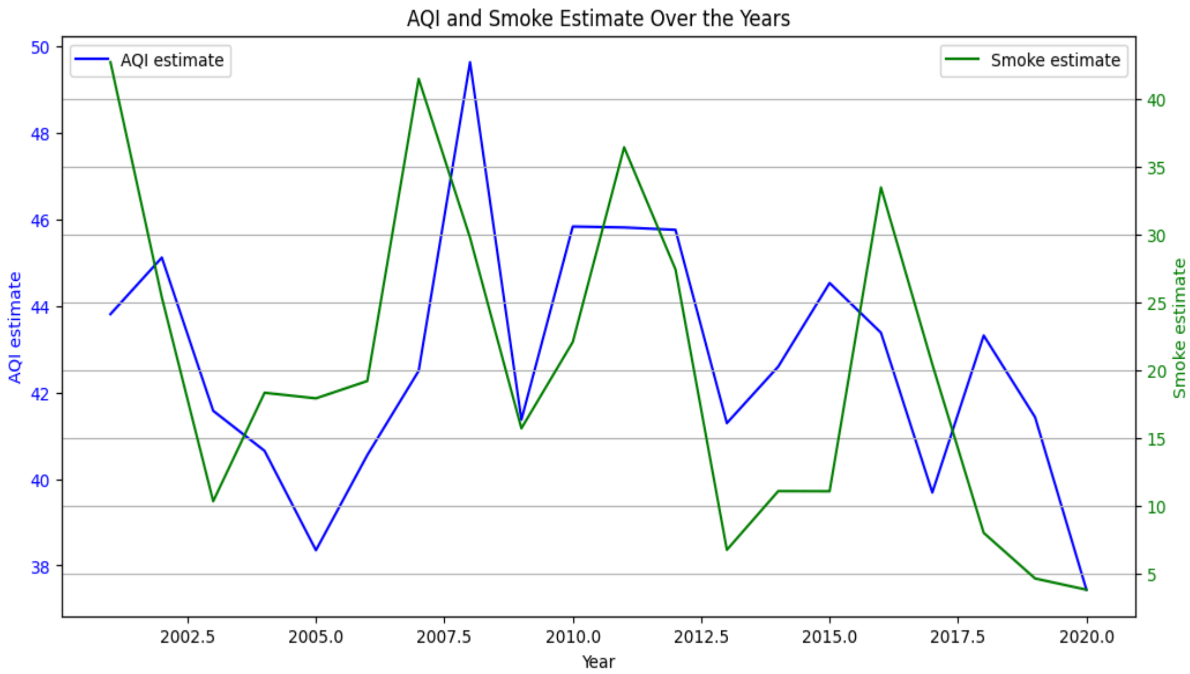

Figure 6: AQI and smoke estimate over the years.

Figure 7 shows the forecast for smoke levels in Philadelphia from 2021 to 2050 using an ARIMAX model. The blue line shows the actual smoke levels from the past, with noticeable changes, especially between 1980 and 2000, when wildfires had a big impact on air quality. Starting in 2021, the red line shows the predicted smoke levels, which steadily rise over time. This increase is likely due to more wildfires caused by climate change.

The pink shaded area around the forecast shows the range of uncertainty, meaning the actual levels could be higher or lower than predicted. The pink area gets wider over time, showing that predictions for the distant future are less certain. This chart highlights the growing smoke problem and the challenges it could bring for health and the environment in the coming years.

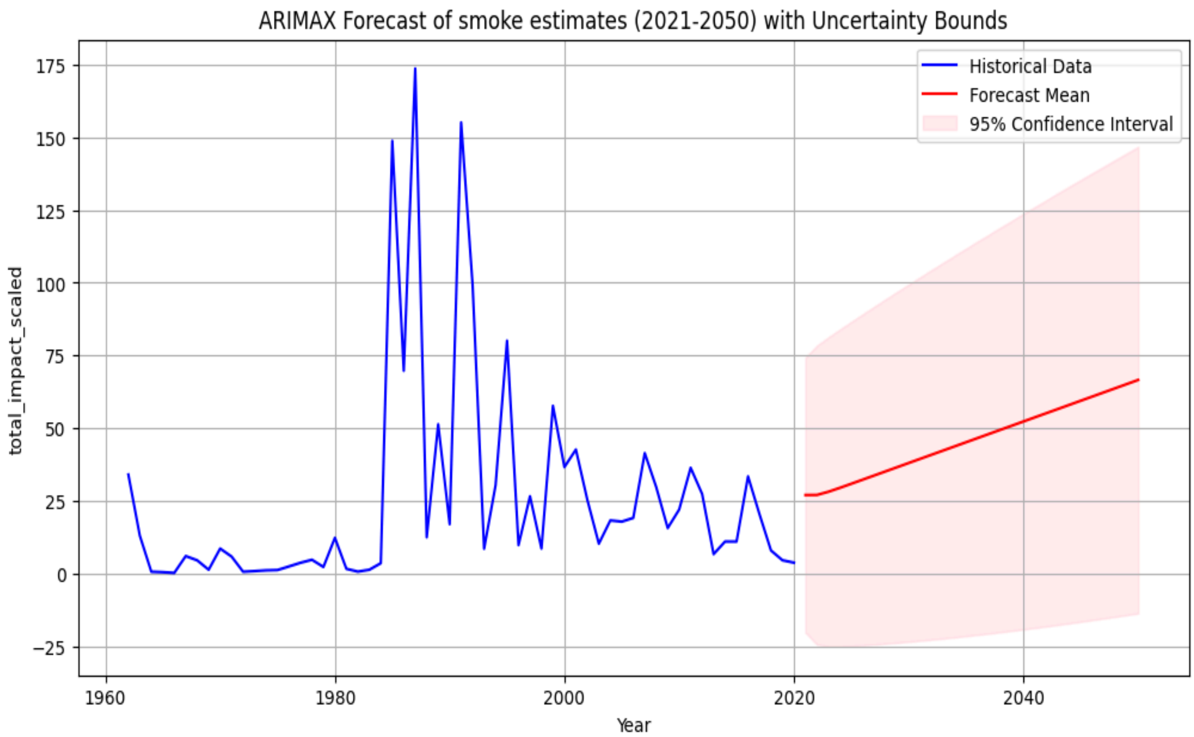

Figure 7: Forecast of the smoke estimate for the next 30 years.

Next, looking at the asthma data, the analysis of asthma data (as shown in Figure 7) uncovered critical insights into how smoke exposure affects health outcomes, particularly for children and older adults. For child hospitalizations due to asthma, the data showed a small but clear relationship between smoke levels and the frequency of hospital visits.

The forecast (Figure 8) indicated that if current smoke trends persist, hospitalizations for children under 18 will likely increase steadily over time. This is concerning, as it highlights the vulnerability of children to respiratory issues exacerbated by prolonged exposure to airborne particulate matter from wildfire smoke.

However, there is some hope: a 50% reduction in smoke levels resulted in a significant decrease in projected hospitalizations, while a 70% reduction showed an even greater impact. These findings emphasize that lowering smoke exposure has the potential to directly improve children's health, reducing the burden on families and healthcare systems.

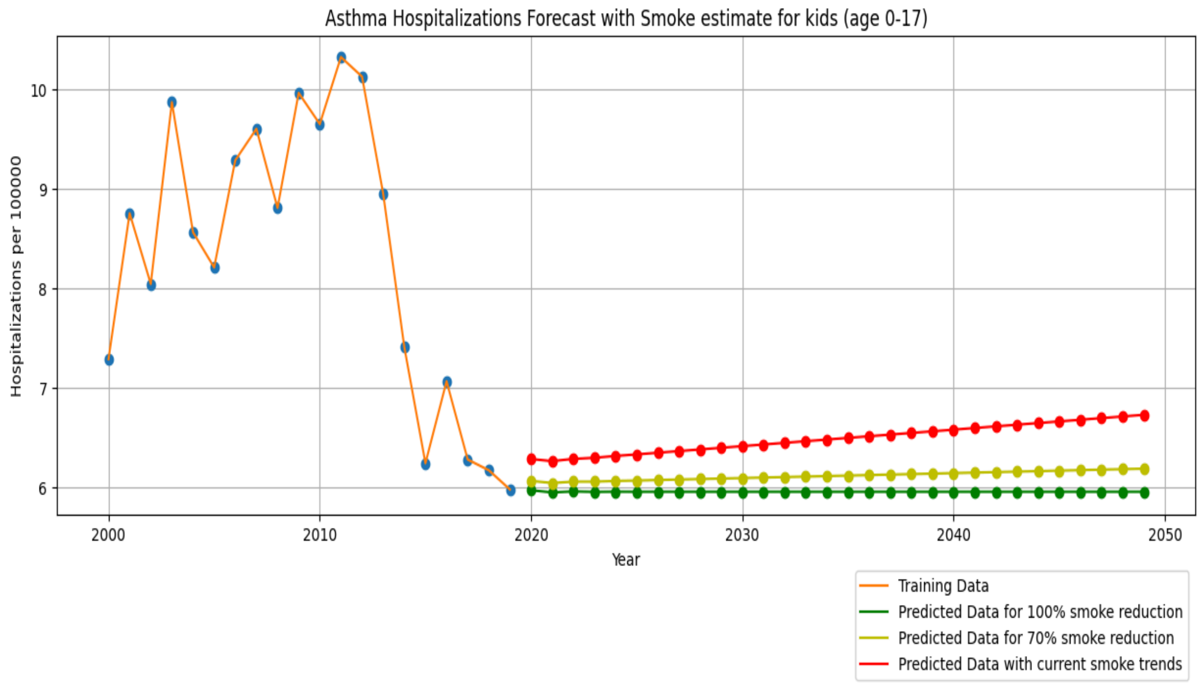

Figure 8: Forecast of asthma hospitalizations for kids (<18 years) for the next 30 years using different smoke levels - 50% of current smoke estimate, 70% of current smoke estimate and 100% of current smoke estimate

When examining asthma deaths, the connection to smoke levels was even stronger. The correlation was around 0.5 with a statistically significant result at alpha as 0.05. The data also revealed that most asthma-related deaths occur in older adults, underscoring their heightened susceptibility to the effects of smoke exposure. The forecast projected that if current smoke trends continue, deaths due to asthma will rise significantly in the coming decades. However, similar to the hospitalization trends, reducing smoke levels by 50% led to a noticeable decline in predicted deaths, and a 70% reduction resulted in even greater improvements. These findings highlight the immense potential of smoke reduction strategies to save lives, particularly for vulnerable populations such as the elderly, who are at higher risk of severe respiratory complications.

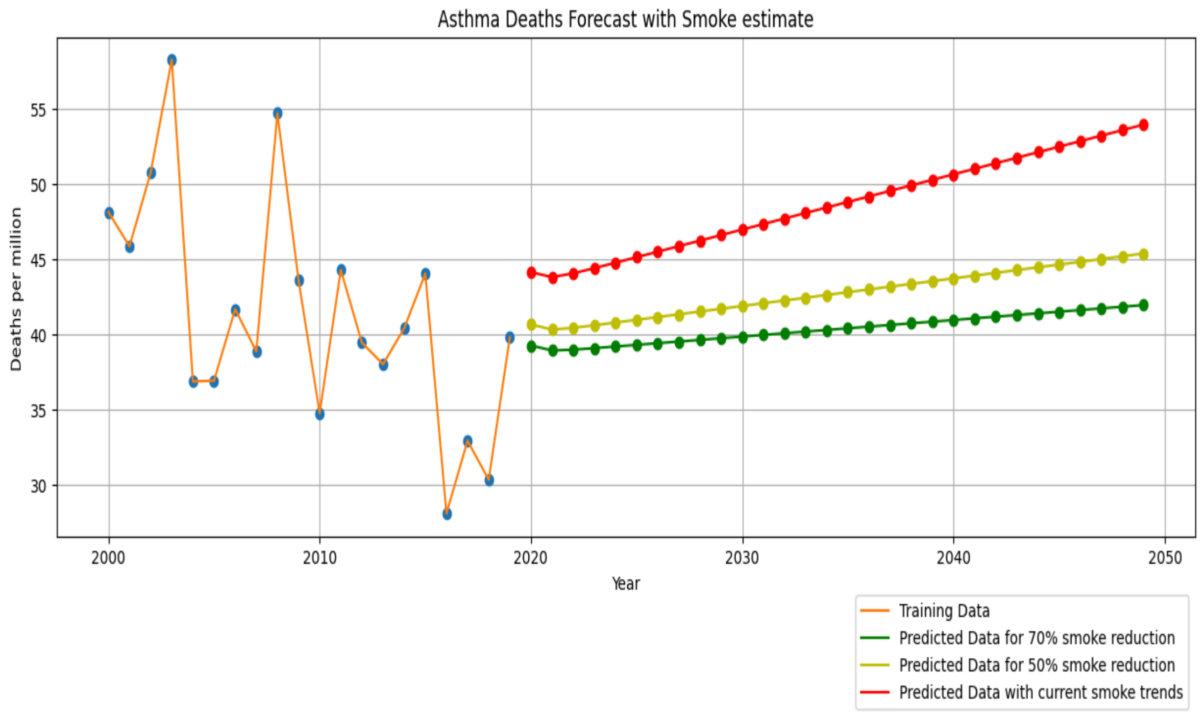

Figure 8: Forecast of asthma hospitalizations for kids (<18 years) for the next 30 years using different smoke levels - 50% of current smoke estimate, 70% of current smoke estimate and 100% of current smoke estimate

Overall, the results demonstrate a clear relationship between smoke exposure and adverse asthma outcomes. They underscore the importance of implementing smoke mitigation policies, as even moderate reductions in smoke levels could lead to substantial health benefits. Lowering smoke exposure not only protects children and older adults from immediate respiratory distress but also reduces the long-term burden on healthcare resources, illustrating the widespread societal benefits of cleaner air.

5. Discussion

This study highlights the growing impact of wildfires on air quality in Philadelphia, with a focus on understanding how smoke from distant fires affects urban areas. The research combined data cleaning, feature engineering, and modeling to generate reliable smoke estimates based on wildfire activity. Although earlier data had some gaps, the trends from 1985 onward offered important insights into the relationship between wildfires and air quality. These trends underline how wildfire activity has become a critical environmental issue, even for cities located far from fire-prone regions.

The histogram (figure 2) revealed that Philadelphia, like many urban areas, is relatively shielded from nearby wildfires due to the limited availability of fire-prone vegetation in heavily urbanized environments. However, the increasing frequency and intensity of wildfires over the past decades have led to a significant rise in smoke exposure, even in distant cities. This highlights the interconnectedness of environmental issues and their far-reaching effects on urban populations.

The forecasting model proved highly effective in forecasting smoke levels by combining historical data with external variables such as fire severity, burned area, and proximity. While some factors, like average distance, were less influential, the model successfully captured the key drivers of smoke levels, providing a reliable tool for long-term forecasting. This reinforces the importance of integrating diverse datasets and environmental variables to understand complex systems like air quality dynamics.

This study also shows the importance of accurate, consistent data for tracking and analyzing long-term environmental changes. The observed correlation between wildfire smoke and poor air quality (as measured by AQI) highlights the pressing need for cities to prepare for the health and environmental impacts of worsening wildfire activity due to climate change. The insights from this study provide a foundation for making data-driven decisions to protect public health and improve air quality in urban areas.

From a **human-centric perspective**, these findings can guide policies aimed at reducing the impact of wildfire smoke on vulnerable populations, such as children, the elderly, and those with asthma. Local authorities, city councils, and public health officials can take actionable steps to mitigate these risks.

One issue for a place like Philadelphia where most of the smoke comes from faraway places is that smoke mitigation strategies won't be very effective. However, there are a lot of precautionary measures that can be taken. **Recommendations to the city council include** strengthening communication systems to provide timely air quality alerts. This will help people get ready for any future air quality dips. Another recommendation is to improve the air quality monitoring infrastructure. This will increase the accuracy of the systems. Finally, targeting interventions for at-risk communities. For instance, providing access to cleaner indoor air environments, distributing masks during high-smoke events, and offering public health education can significantly reduce health risks.

Further, the data was handled with a **human centric mindset**. This was done by following the rules enforced by the data owners and by making sure to use aggregations everywhere and making sure that any aggregation with less than 10 individuals were dropped for additional privacy.

Human-centered data science principles learnt in this course guided many decisions in this project. The main focus was on understanding real health problems, like asthma hospitalizations and deaths, which affect people's well-being. The datasets used - the CDC Multiple Causes of Death and PHC4 data, were chosen because they provide reliable information about groups like children and older adults who are most at risk. The SARIMAX model was used to create clear and easy-to-understand predictions so that healthcare workers, policymakers, and others could use the results to make decisions. Testing scenarios, like reducing smoke levels, helped show how actions could improve public health. Overall, the goal was to make sure the findings were useful, easy to understand, and focused on helping people.

6. Limitations

Smoke estimates might not be perfect. We used data from the US Geological Survey and a formula that looks at burned areas, how bad fires were, and how far away they were. But this doesn't capture everything about how smoke moves in cities. Things like wind, hills, and weather can change smoke levels, and our method might miss these.

Further, I didn't have all the data I would have liked. I used information about kids going to hospitals and people dying from asthma. But this doesn't show us every asthma case. Some milder cases or deaths that asthma made worse but weren't directly caused by it might be missing. Another issue is that we focused a lot on smoke, but asthma is complicated. Many things besides air quality can cause asthma, like genes, allergies, jobs, and how much money people have. My study doesn't look at all these other factors, so I might be saying smoke causes more problems than it really does.

Another limitation of this study is the availability of hospitalization and asthma death data, which only spans from 1999 to 2020. This relatively short timeframe may not fully capture long-term trends or variations in asthma outcomes that could provide a more comprehensive understanding. Additionally, the limited time frame might affect the accuracy of predictions for future scenarios, as the model relies on a smaller historical dataset. Expanding the data range in future studies could help strengthen the analysis and provide more robust insights.

Also, predictions for the future use math (SARIMAX models), but these are based on what's happened before and what we think will happen with smoke. If big changes happen with climate, how we manage fires, or how we protect people's health, our predictions might not be right and the model needs to be fit accordingly.

Finally, most smoke coming into Philadelphia seems to originate in the Canadian wildfires. However, given that we have restricted our analysis only to the 650 mile radius. This means we have missed out on some key wildfires which might have affected the area.

7. Conclusion

In this analysis, the initial question was to understand how wildfire smoke levels impact air quality and health outcomes in Philadelphia, particularly concerning asthma-related hospitalizations and deaths. The findings clearly demonstrate a significant correlation between smoke exposure and increased health risks, especially for vulnerable populations such as children and older adults. The analysis revealed that rising smoke levels, driven by increased wildfire activity, are expected to lead to more hospitalizations and deaths if current trends continue. However, the study also highlighted the potential benefits of smoke reduction strategies, with substantial improvements in health outcomes through even moderate reductions in smoke levels. These insights emphasize the importance of addressing wildfire smoke as a public health issue and implementing mitigation measures to protect communities and reduce healthcare burdens. A few recommendations for the same have also been provided.

8. References

- GeoJSON. GeoJSON Python Library Documentation. https://geojson.readthedocs.io/en/latest/
- Pyproj. Pyproj Library Documentation. https://pyproj4.github.io/pyproj/stable/
- PHC4 county level health data - https://www.phc4.org/reports/county-level-health-care-reports/
- EDDIE - Enterprise Data Dissemination Informatics Exchange - https://www.phaim1.health.pa.gov/EDD/WebForms/HospitalDischargeCntySt.aspx
- SARIMAX Model Documentation. https://www.statsmodels.org/stable/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html

- Asthma in Pennsylvania - https://www.cdc.gov/national-asthma-control-program/media/Asthma_in_PA.pdf
- CDC Wonder - Multiple causes of Death - https://wonder.cdc.gov/mcd.html
- Asthma visits during air quality events - https://site-dr.mainlinehealth.org/-/media/files/pdf/research/population-health/air-quality-population-health-brief.pdf
- Liu K, Hua S, Song L. PM2.5 Exposure and Asthma Development: The Key Role of Oxidative Stress. Oxid Med Cell Longev. 2022 Apr 4;2022:3618806. doi: 10.1155/2022/3618806. PMID: 35419163; PMCID: PMC9001082.
- Wilgus, M. -L., & Merchant, M. (2024). Clearing the Air: Understanding the Impact of Wildfire Smoke on Asthma and COPD. *Healthcare*, *12*(3), 307. https://doi.org/10.3390/healthcare12030307

9. Data Sources

- AQS - Air Quality System API - https://aqs.epa.gov/aqsweb/documents/data_api.html
- USGS Wildfire Dataset - https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81
- CDC Wonder Multiple Causes of Death - https://wonder.cdc.gov/mcd.html
- PHC4 Asthma hospitalization data (EDDIE) - https://www.phaim1.health.pa.gov/EDD/WebForms/HospitalDischargeCntySt.aspx
- CDC BRFSS Asthma survey - https://www.cdc.gov/brfss/acbs/index.htm