Rthvik Raviprakash
11th December 2023

FINAL REPORT

1. Introduction

This analysis is significant as it sheds light on wildfire patterns and impacts near Helena, Montana, over several decades specifically over the years 1963-2023. It gains importance in the context of climate change and its role in mitigating wildfire frequency and intensity. The study is particularly relevant in the context of climate change and the impact it has on the Health, economy, education etc, where wildfires are becoming more frequent and intense. Further, the study is not only academically intriguing but also practically vital in guiding policies and emergency responses.

The primary motivation is the urgent need to understand and mitigate the effects of wildfires, a growing concern in the era of climate change and as mentioned to lessen the impacts it has on humans as a whole and in this study specifically the health impacts. This study aims to protect communities and the environment by providing a comprehensive understanding of wildfire trends and effects, guiding preparedness and resilience strategies.

This analysis tackles real-world issues by offering insights into wildfire behaviors and impacts, assisting in risk assessment, health impact understanding, and healthcare planning. It also addresses research gaps by exploring wildfire occurrences, their proximity to urban centers, and their temporal changes, contributing valuable knowledge to wildfire management and research. The study is crucial for community safety, public health, and environmental conservation. Its findings are pivotal for informed decision-making in land management, emergency preparedness, and healthcare.

2. Background/Related Work

Significant research has been conducted in the area of wildfire analysis, particularly in the context of its impact on urban areas and public health. Historical data, such as the case of forested areas adjacent to Helena, Montana, demonstrates how changes in land use and fire suppression tactics have altered the natural fire regime, leading to denser forest growth and increased risk of high-intensity crown fires[1]. Studies have shown that the encroachment of conifers in areas that were historically grasslands has changed the fire dynamics, posing greater threats to urban-interface areas. This shift emphasizes the importance of understanding wildfire behavior and its implications on urban planning and public health. Prior research informs this study by providing a baseline understanding of fire behavior, risk factors, and mitigation strategies, which are crucial for developing effective wildfire hazard reduction plans.

Research questions in this domain often focus on understanding the long-term impacts of wildfires on communities, ecosystem health, and public safety. This study is informed by

previous findings and aims to extend the knowledge base by addressing specific research questions:

Research Questions:

1. What are the estimated smoke impacts on your assigned city for the last 60 years?

    - This question seeks to quantify the historical impact of wildfire smoke on Helena, Montana, providing insights into trends and changes over time.

2. What is the estimate of the child mortality rates in the next 25 years?

    - Focusing on the future, this query aims to predict the impact of wildfire smoke on child mortality rates, a critical aspect of public health planning.

3. How has the number of people with poor health and the number of patients seen by primary care physicians changed over the past few years?

    - This question examines the broader health impacts of wildfires, exploring the correlation between fire events and the health status of the local population.

4. Based on the estimate of the child mortality rate due to smoke exposure and the general health, what is the estimate of the number of patients to be seen and the number of doctors/health care providers required in the future?

    - This question aims to translate health impact estimates into actionable insights for healthcare planning, ensuring the community is well-equipped to handle future challenges posed by wildfires.

These research questions are critical for understanding the long-term effects of wildfires on urban centers like Helena, Montana. They aim to bridge the gap between historical data and future planning, offering a comprehensive view of the challenges and necessary responses to mitigate the impact of wildfires on public health and safety.

I have used these additional datasets:

Dataset 1 - Poor/Fair Health Percentage:

- Content: This dataset encompasses the percentage of adults reporting poor/fair health in Lewis and Clark County, the location of Helena, Montana. It includes Confidence Interval values, specifying both lower and upper bounds for the reported percentages. The dataset also provides Geography ID information for Lewis and Clark County, which, for the purpose of our analysis, will be omitted. The relevant information comprises the percentage values of adults reporting poor/fair health for each year, along with the corresponding upper and lower bounds of the confidence intervals. The data consists of data points of each year from 2014 to 2022.

Link to the dataset: (Note: the datasets are also in the repository)

https://drive.google.com/file/d/1lFeh0jp0kdy4K3GgyNj4ZvAE2lT-vBGM/view?usp=sharing

Dataset 2 - Under-18 Mortality Rate: (Note: the datasets are also in the repository)

- Content: This dataset focuses on the number of deaths among residents under the age of 18 per 100,000 population in Lewis and Clark County. Similar to Dataset 1, it includes high and low confidence intervals. The Geography ID information is present but will not be utilized in the analysis. The essential data includes the mortality rate for residents under 18, along with the associated high and low confidence intervals. The data consists of data points of each year from 2014 to 2022.

Link to the dataset:
https://drive.google.com/file/d/1hykffk7zsYiaMtt83lLCBXZNID-r25rQ/view?usp=sharing

Dataset 3 - Population-to-Provider Ratio: (Note: the datasets are also in the repository)

- Content: This dataset provides the ratio of the population to the number of primary care providers other than physicians in Lewis and Clark County. Similar to the preceding datasets, it includes Geography ID information, which will be disregarded in our analysis. The pertinent data consists of the population-to-provider ratio for each year from 2014 to 2022.

Link to the dataset:
https://drive.google.com/file/d/18B5sXk8JNz5Rf2vwoqbjJgUMDeloezKb/view?usp=sharing

I have downloaded all the three dataset from here:
https://datausa.io/profile/geo/lewis-and-clark-county-mt

Licensing: https://www.countyhealthrankings.org/terms-use

3. Methodology:

My preliminary analysis aimed to analyze the number of wildfires occurring for every 50 mile radius within 1250 miles of Helena, Montana. The reason for doing this is to understand the distance at which we see the most number of wildfires and if the smoke impact is due to fires nearby or does even fires from further distances have a significant impact. Also, for us to understand how safe of a city Helena is for its citizens to live in.

The method for it is as follows:

Data Source and Attributes:

The data, in JSON format can be found here Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons), pertains to wildfires, specifically designated under the 'Wildfire' feature. It includes various attributes like fire year, size in acres and hectares, fire type, and information on the mapping and recording of the fire events.

Data Processing:

The processing involves a Python script which can be found [here] that iterates through this wildfire data, focusing on fires between 1963 and 2023. Key steps include:

- Data Preparation: This step categorizes the data, like counting fires within specific distance ranges and aggregating fire sizes by year.
- Geographic Analysis: Using the pyproj library, the script converts the fire perimeter data from the ESRI:102008 coordinate system to the more universal EPSG:4326 (latitude/longitude format). This step is crucial for accurately computing distances and geographical analyses.
- Distance Calculation: The script calculates the shortest distance from a city to the fire perimeter. This is done using the Geod class from the geographic library, which operates on the WGS84 ellipsoid model. The method considers the irregular shape of fires, opting to find the nearest perimeter point to the city.
- Smoke Impact Estimation: For fires within a 1250-mile range from a city, the script estimates smoke impact based on fire size, type, and distance, aggregating this impact by year.

I also wanted to analyze the total acres of fires burned within a 1250 mile radius of Helena to understand how much area is being burned over the years and to have an idea of what to expect in the future. This analysis also allows us to understand the intensity of the fires over the years.

Data Processing:

The processing involves a Python script which can be found [here] that iterates through this wildfire data, focusing on fires between 1963 and 2023. This is very similar to the data processing step for the previous figure of the number of fires based on the distance from Helena over the years. Key steps include:

- Data Preparation and aggregation: This step categorizes the data, like aggregating the acres by year, since for each of the features (i.e a fire) after the data has been loaded it consists of the year and the acres of land that was affected by the fire. So for every iteration I use a dictionary that consists of the years as keys and for every fire of a given year that is being iterated will be added to the total count of the acres of that year in the dictionary. So, we will finally have a dictionary of the years as keys and the values to be the total acres of land burned due to the fires for that particular year in over the years 1963-2020 within a 1250 mile radius of Helena, Montana.

Then to answer my first research question which is to estimate the smoke impact over the past 60 years, I needed to come up with a value or a metric that can give me an estimate of the smoke impact for Helena. The methodology is as follows.

How are the values calculated?

- Smoke impact estimate: First, I calculate a value for a given fire using the formula (weight*acres)/(distance**2) where the weight is a predefined value from 0.0-1.0, based on the type of fire (i.e Wildfire, Prescribed Fire etc.) which you can refer to in the

ReadMe and my commented code in the notebooks for the specifics of it. The acres is the value for the acres of land for that fire that was affected and the distance is the value of the distance of that fire from Helena, Montana. After calculating this for each of the fires and for a particular year, I then take the average for that particular year and define that to be my smoke impact estimate. The decision to use the inverse of the distance squared in smoke impact calculations is based on principles of physics, specifically the inverse-square law. This law states that a specified physical quantity or intensity is inversely proportional to the square of the distance from the source of that physical quantity. In practical terms, for smoke impact, this means the impact (or intensity) of smoke diminishes with the square of the distance from the source of the fire.

- AQI estimate: For the AQI estimate I have taken the AQI values based on the AQI data collected based on the Gaseous pollutants and based on the Particulate Pollutants. For each year I have taken the average of the AQI values for both the gaseous pollutants and the Particulate pollutants. I noticed that particularly for Helena, Montana only for the years 1986-2023 I could find that data for AQI for both gaseous pollutants and the particulate pollutants which is what I have calculated and shown in my visualization. Since there are two values for each year one from the gaseous pollutants and one from the particulate pollutants I have taken the max value as my estimate between the two as my estimate.

Finally, I have normalized the values for both my smoke impact estimate and the AQI estimate to be a value from 0.0 to 1.0 using the min-max normalization and have plotted as above. So, you can read the graph as the value of the normalized smoke impact estimate and the normalized AQI estimate over the years 1986-2023 based on the data collected from the county of Lewis and Clark which is where Helena, Montana is located. The y- axis represents the value and the x-axis represents the years.

Data Processing:

The data for this process was requested data from the US Environmental Protection Agency (EPA) Air Quality Service (AQS) API. You need to use your email to get an API key and further for a city that you pick you can get a summary data of your county for both the Gaseous AQI pollutants (I have picked CO, SO2, NO2 and O3) and the Particulate AQI pollutants (I have picked PM10, PM2.5 and Acceptable PM2.5). Then you can preprocess the data to extract only the relevant information which in our case the AQI values based on the date. Then based on the dates I have taken the average values of the AQI for a given year and stored it in a dictionary separately for Gaseous and Particulate AQI pollutants. You can find detailed information and code on how to do this [here].

To answer my second research question, using the data of the smoke estimate generated previously I have used the Holt-Winters Exponential Smoothing model, a time series forecasting technique. The model is applied to historical average smoke impact data. The data is structured into a DataFrame, indexed by year, and the model incorporates both additive trends and seasonal components to predict the next 25 years of average smoke impact. The historical data and

forecasted values are then visually represented in a plot, highlighting past trends and future projections.

To answer my next two research questions that focus on the health impacts and to give recommendations for the city council.

First, I collected data over the past few years specifically over the years 2014-2022 for the number of patients seen by primary care providers over the years and visualized the trend. The reason for this was to understand the overall trend in the number of people requiring general healthcare checkups. This would also give me an idea of if the number of people falling ill has been increasing/decreasing over the years or if the number of healthcare professionals have increased/decreased leading to the trend in the number of patients seen.

Further, to support that analysis I also collected data of the percentage of people in Lewis and Clark county which is where Helena is situated who reported poor or fair health over the years 2014-2022. Visualizing the trend in this data will give me an idea to support my previous analysis of the number of patients seen over those years by looking at the trend of the number of people falling ill.

Now that we have looked at a couple of methods to see the trend in general health, let's move on to health impacts due to wildfire smokes then further combine that with the general healthcare in Helena for our final conclusions and recommendations to the city council.

The reason that I wanted to look at child mortality was that based on research it has been found that inhaling the wildfire smoke has a direct impact which increases the risk of child mortality in your children. Each 1 μg/m3 increment of fire-sourced PM2·5 was associated with a 2·31% (95% CI 1·50–3·13) increased risk of child mortality [2]. I was able to find data for the child mortality rates over the years of 2014-2020 for which I first looked at the most recent trend. Then I built a Random Forest Regressor model to predict the child mortality rates for the next 25 years or so to see the trends, fluctuations, peaks etc.

I combined all the three datasets into a single pandas dataframe and checked the correlation between each of the features (i.e the number of patients seen by primary care providers over the years, the percentage of people with poor or fair health over the years and the child mortality rates over the years). The reason for me to understand the correlation between each of these features is to see if the child mortality rates are correlated with the general health of the population and what measures need to be taken especially in terms of talking about the wildfire impact in this regard and the decisions that need to be made.

<u>Human Centered Considerations that informed my study design:</u>

Feedback and Adaptability: I designed the study with the understanding that community feedback is valuable. This openness to adapt and revise the study based on community input reflects a commitment to a participatory, ethical research process.

Consideration of Long-term Effects: Understanding that the impact of wildfires is not just immediate but also long-term, I designed the study to consider both short-term and extended effects on public health and environmental quality. This comprehensive view ensures that the study's recommendations are relevant both for immediate response and for future community resilience planning.

Long-term Health Effects: Beyond immediate health impacts, consider the long-term effects of smoke exposure on the community's health. This includes chronic respiratory issues, mental health concerns due to the trauma of wildfires, and other prolonged health consequences.

Community Impact Focus: The primary design consideration was the potential impact of wildfires on the Helena community. This focus directed the choice of data points, such as distance of fires from the city, smoke impact, and health-related consequences, ensuring that the study's outcomes would be directly relevant to the residents' safety and well-being.

4. Findings

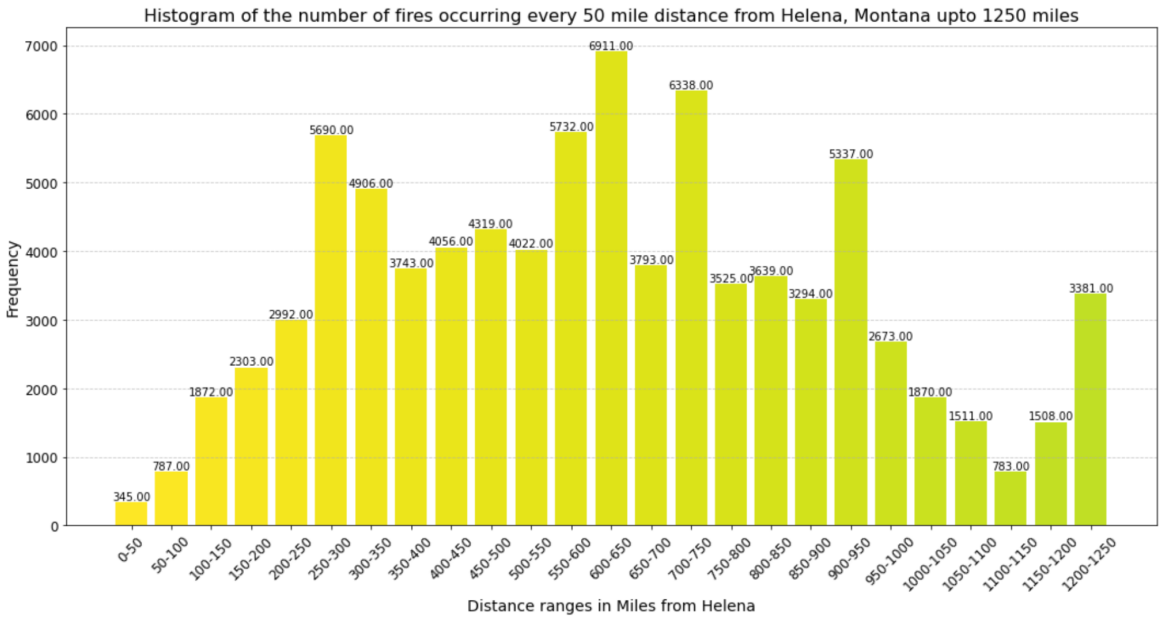

Figure 1: Histogram of the Number of fires vs the distance ranges from Helena, Montana over the years 1963-2023

As we can see from the figure above the total number of fires within a 250 mile radius of Helena over the years 1963-2023 is about 8299 which is a lot less than the number of fires in the next 500 mile radius which shows about 49510 fires. This shows us that when it comes to Helena that

a lot of the fires occur further away from but that doesn't stop from the wildfires having an impact in the city.

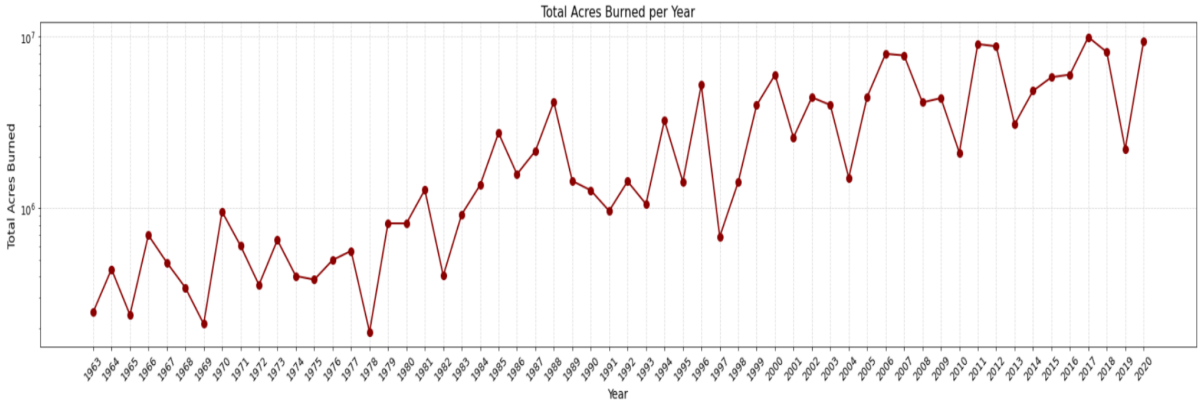

Figure 2.1: Total number of acres of land burned over the years 1963-2020 within a 1250 mile radius of Helena, Montana (y-axis is log scaled)

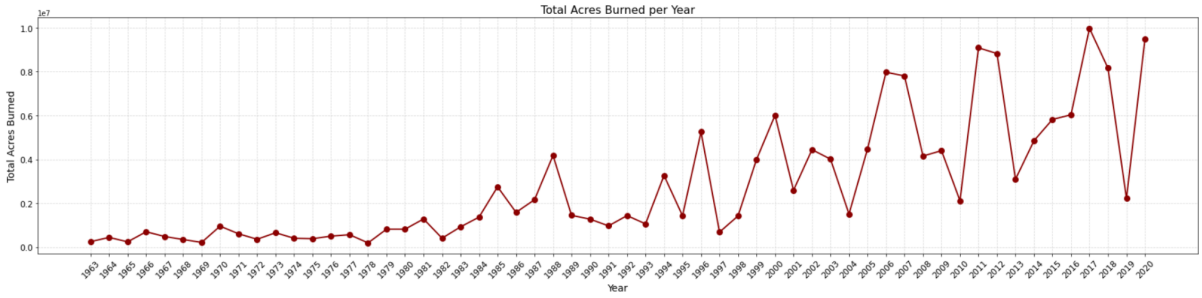

Figure 2.2: Total number of acres of land burned over the years 1963-2020 within a 1250 mile radius of Helena, Montana (without the y-axis being log scaled)

The above two graphs are basically the same but the first one being log scaled. It can be seen that in the past 25 years or so the total acres of fires within a 1250 mile radius of Helena log scale has not gone below 10^6. This shows that even though there are fluctuations compared to the years before about 25 years ago it has shown an increase and has not gone below a certain threshold.

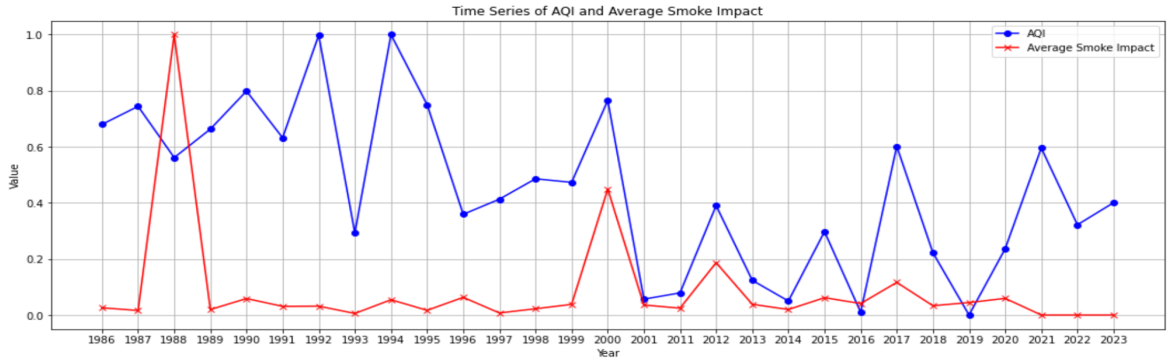

Figure 3: The time series graph of the Average Smoke impact estimate and the AQI estimate vs years for AQS stations in Lewis and Clark county which is where Helena, Montana is located.

As you can see from the graphs above it can be seen that in the past 10 years or so the smoke impact has stabilized and has not shown any notable increases but the AQI shows fluctuations with a few peaks here and there showing that the AQI has seen these peaks from the fires that has occurred at greater distances. This can be concluded since as mentioned in the methods the smoke estimate is inversely proportional to the square of the distance of the wildfire occurrence. So, we can say that wildfires occurring at further distances still have an impact on the AQI.

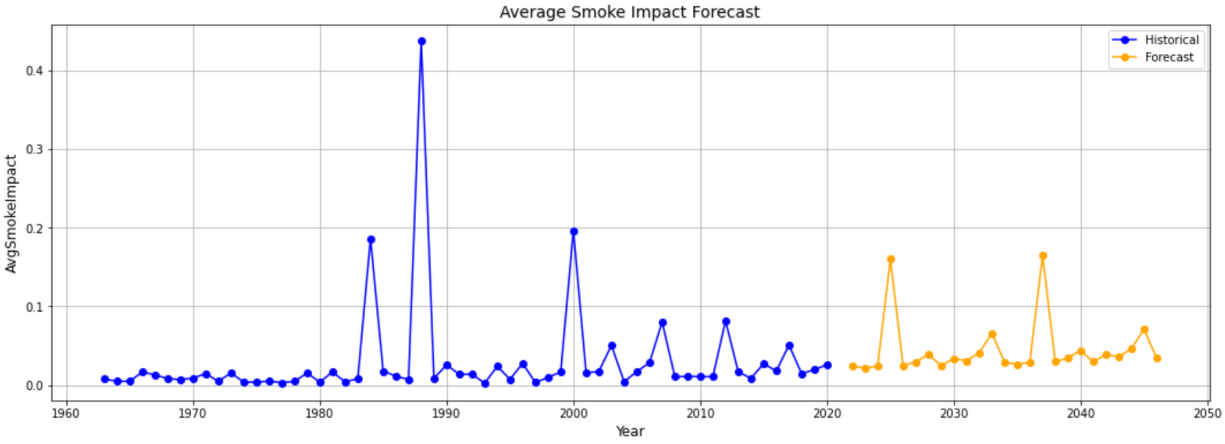

Figure 4: Smoke impact estimate forecast for the next 25 years.

The above graph shows the smoke impact estimate of Helena for the next 25 years and we can see that there are a couple of major peaks in the year 2024 and 2037 which is something to keep in mind for emergency preparedness in the future while making future plans.

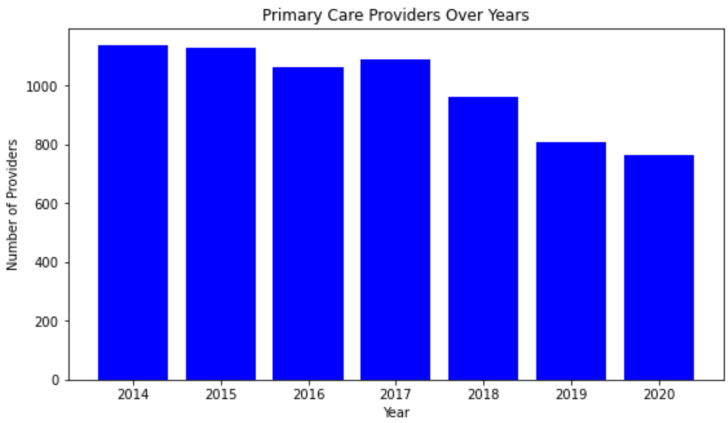

Figure 5: The number of patients seen(i.e number of providers) by primary care providers over the years 2014-2020

As we can see over the years the number of patients seen(i.e number of providers) by primary care providers have been decreasing. The reason for that could be the number of people getting sick are decreasing or the fact that the number of healthcare professionals/doctors have decreased over the years.

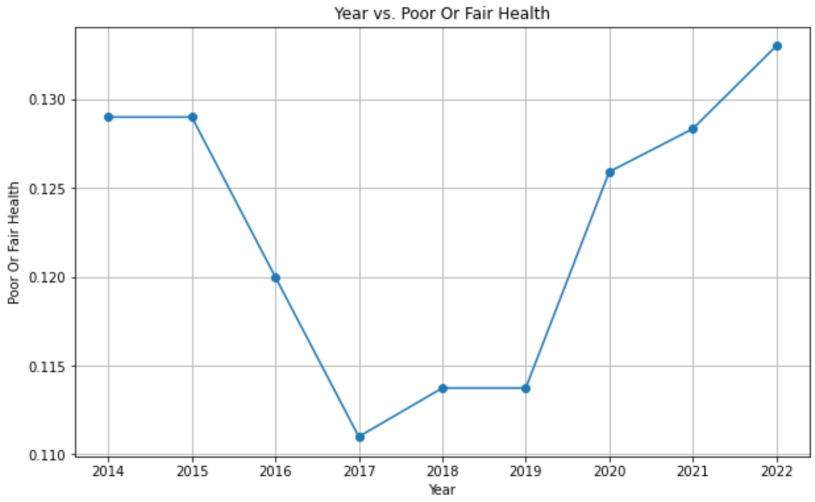

Figure 6: The percentage of people in Helena with poor or fair health over the years 2014-2022

This above graph shows us that in the past 4 years or so the number of people falling sick or the number of people with poor or fair health has increased which contradicts our previous graph showing us that there is a much higher likelihood that the number of doctors/healthcare professionals have decreased.

Based on the above two graphs we see a trend in increasing low or poor health by about 0.05% of the population, and the number of patients seen has decreased at an average of 10% each year.

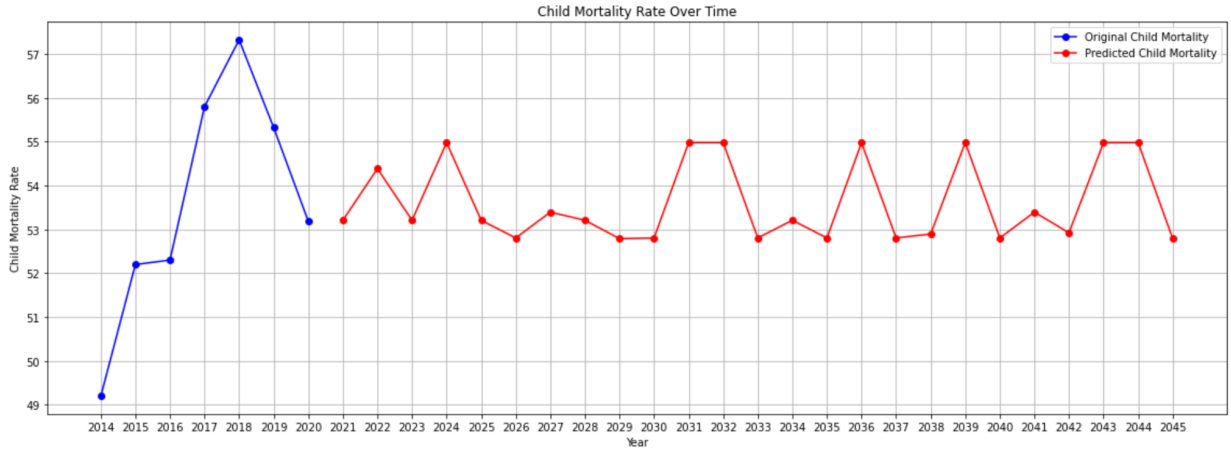

Figure 7: The Child Mortality Rate Forecast for the next 25 years.

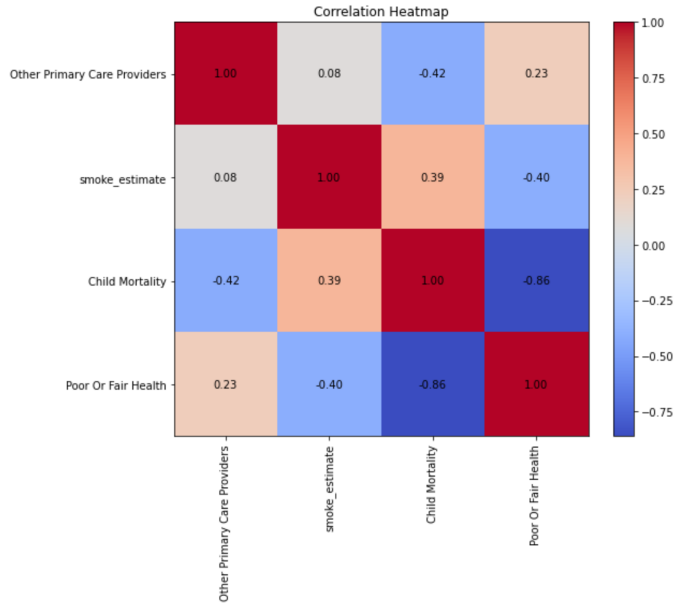

Figure 7: The correlation heatmap between smoke_estimate, child mortality rate, number of people with poor or fair health and the number of patients seen by primary care providers over the years.

Based on the correlation plots we see the above correlations. The correlation between the percentage of people with poor or fair health over the years is -0.85 with the child mortality rates. You can also see the correlation between the other features. There is a positive correlation between the child mortality rates with the smoke impact.

5. Discussion/Implications

Importance and Interest of Findings:

My analysis provided vital insights into the wildfire impacts on Helena, Montana. The identification of a higher frequency of wildfires in a specific radius highlights areas needing focused mitigation efforts. The stable smoke impact versus fluctuating AQI levels indicates that distant wildfires still significantly influence Helena's air quality. Critically, the negative correlation between general health and child mortality rates, juxtaposed with the positive correlation with smoke impact, emphasizes the complex interplay between environmental factors and public health. The negative correlation of poor health with the child mortality rates did not make sense since they go hand in hand which is why I recommend the following for the council. Given that there is going to be some peaks and fluctuations in the child mortality rate in the future based on our model it is important to note the following recommendations.

Recommendations for City Council and Residents:

1. Healthcare Resource Expansion: In light of the negative correlation between poor health and child mortality rates, I recommend increasing healthcare resources, particularly for regular health checks for children and mothers especially to ensure that the child mortality rates can be controlled which can be due to the smoke inhaled from the wildfires. This is crucial for early detection and prevention of health issues exacerbated by wildfire smoke.

2. Enhancing Doctor Availability: To manage the increasing healthcare demands, I propose a 30-40% increase in the number of patients seen annually by primary care providers. This will necessitate calculating and subsequently augmenting the number of healthcare professionals, ensuring adequate coverage for both general health and specific checks for babies and their mothers to ensure that the child mortality rates can be controlled which can be due to the smoke inhaled from the wildfires.

3. Emergency Preparedness and Public Health Initiatives: The city should strengthen its emergency response plans for peak wildfire years (i.e 2024 and 2037 from our results) and implement public health initiatives to address the trend of increasing poor health reports, despite fewer visits to primary healthcare providers.

4. Timely Action Plan Development: It's imperative for the city council and mayor to start formulating a concrete action plan within 6-12 months, considering the urgency highlighted by the predicted peaks in smoke impact and child mortality rates.

Human-Centered Data Science Approach:

In this project, human-centered data science principles significantly influenced my decision-making process. My primary focus was on how the data and its subsequent interpretation could impact the lives of Helena's residents. This approach led me to delve deep into not just the environmental data but also the health-related outcomes. Such an integrated analysis ensured that my study's recommendations directly addressed the real human needs and

concerns in Helena. Key to this methodology were ethical considerations like the privacy of health data and inclusivity in data interpretation. My analysis was driven by the objective to provide actionable insights that would enhance the well-being and safety of the Helena community, reflecting my commitment to using data science in a responsible and empathetic manner.

Throughout the project, the principles of human-centered data science remained at the forefront of my methodology. This wasn't just about analyzing environmental data; it was about understanding its direct implications on public health, especially for vulnerable groups such as children and mothers. In handling health data, I placed utmost importance on ethical considerations such as ensuring privacy and inclusivity. My analysis was aimed at providing actionable, empathetic insights to enhance the well-being and safety of Helena's residents, exemplifying responsible and impactful data science usage. These insights included recommendations for increasing healthcare resources, particularly focusing on regular health checks for children and mothers, in response to the observed negative correlation between poor health and child mortality rates. I also proposed a significant increase in patient handling capacity for healthcare providers, aiming to address both general health concerns and specific health risks due to wildfire smoke. This holistic approach underscores my commitment to applying data science principles in a manner that prioritizes human health and welfare, setting a precedent for future studies in this critical area.

6. Limitations

This study, while comprehensive in its scope and analysis, encounters several limitations that are important to acknowledge, as they could significantly impact the results and interpretations.

1. Data Temporal Range: The datasets used have temporal limitations. The child mortality rate data covers only the years 2014-2020, and the data for the percentage of people reporting poor or fair health, as well as the number of patients seen by primary healthcare providers, is limited to 2014-2022. This restricted timeframe may not fully capture longer-term trends and patterns that are crucial for a more comprehensive understanding of the impacts of wildfires.

2. Geographical Scope of Data: The data is at the county level (Lewis and Clark County) and not specific to the city of Helena. This broader geographical scope could potentially dilute the specificity of the findings as they relate to the city, leading to generalizations that may not precisely reflect the city's unique situation.

3. Forecasting Assumptions: The study assumes the accuracy of the forecasts for child mortality rates and smoke impact over the next 25 years. These forecasts are based on historical data and predictive modeling, which inherently include uncertainties and potential for error. Consequently, the predictions should be interpreted with caution and understood as estimations rather than definitive future outcomes.

4. Correlation Assumptions: A critical assumption made is the expected positive correlation between poor health and child mortality rates. However, the analysis revealed a high negative correlation, which could indicate either an unexpected trend or a limitation in the data or its interpretation. This discrepancy emphasizes the need for careful consideration of underlying assumptions in data analysis.

5. External Influences and Data Quality: External events and variables, such as policy changes, unexpected healthcare crises, or extreme weather events, could influence the accuracy of the study's predictions. Additionally, unforeseen ethical or legal considerations related to data usage or results dissemination may arise, necessitating adjustments. The reliability and completeness of datasets are also crucial; any data inconsistencies, gaps, or inaccuracies encountered would impact the study's accuracy and might require additional time for data cleaning and validation.

Acknowledging these limitations is essential for a clear understanding of the study's scope and the reliability of its conclusions. It also highlights the need for ongoing evaluation and adaptability in response to new data and changing circumstances.

7. Conclusion

In this study, I set out to address key research questions regarding the impact of wildfires within and around Helena, Montana. The primary questions explored were the estimated smoke impacts on Helena over the last 60 years, the projected child mortality rates in the next 25 years, the changes in the health status of the population, and the healthcare resources required to address these impacts.

My findings revealed a higher concentration of wildfires within a 250-750 mile radius of Helena, indicating a broad area of impact. Despite a stable smoke impact in recent years, fluctuating AQI levels suggest that distant wildfires significantly influence Helena's air quality. Notably, a negative correlation was found between general health and child mortality rates, contrasting with the positive correlation between smoke impact and child mortality. This indicates a complex relationship between environmental factors and public health. Additionally, an increase in reported poor health, despite a decrease in primary healthcare visits, calls for enhanced healthcare resources and strategies.

This study is a prime example of human-centered data science in action. By integrating environmental data with public health outcomes, it highlights the importance of considering the human element in data analysis. Ethical considerations in data handling and a focus on inclusivity and privacy were crucial in shaping the study. The insights provided by this research are not just numbers and trends; they reflect the real-life implications of environmental changes on a community's health and well-being. In conclusion, this study not only contributes to the understanding of wildfire impacts but also underscores the vital role of human-centered principles in data science, ensuring that research outcomes are relevant, ethical, and beneficial to society.

8. References

[1]https://www.montana.edu/extension/forestry/publications/fact-sheets/FF_Developing%20Fire%20Hazard%20Reduction%20Plan_PK.pdf

[2]https://www.thelancet.com/journals/lanplh/article/PIIS2542-5196(21)00153-4/fulltext#:~:text=Findings,larger%20fires%20was%20more%20toxic.

9. Data Sources

https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81

https://datausa.io/profile/geo/lewis-and-clark-county-mt