DATA 512- Final Report

By Vritesh Gera

1. Introduction/Motivation

This is an analysis of wildfires and how the smoke from wildfires affects the cities near them. I am running this analysis for Klamath Falls, Oregon. Klamath Falls is a city in southern Oregon, USA having a picturesque setting amid mountains and lakes. I will be working on finding how wildfires impact the healthcare system of Klamath Falls.

This report is part of a 4 part project that I was assigned to work on in DATA 512:

Part 1- Pulling data from different sources and calculating the smoke estimate

Part 2- Creating the Extension Plan to conduct further research

Part 3- Presenting the results in a presentation

Part 4- Creating the final report and repository

The main inspiration behind this project is that more frequently summers in the western US have been characterized by wildfires with smoke billowing across multiple western states. There are many proposed causes for this: climate change, US Forestry policy, and growing awareness, just to name a few. Regardless of the cause, the impact of wildland fires is widespread. The harm caused by wildfires can cause significant harm to both the environment and human communities.

The impacts are multifaceted and can include:

a. Loss of Human Lives

Wildfires can pose direct threats to human safety, leading to injuries and fatalities. Evacuations and firefighting efforts can also be dangerous.

b. Property Damage

Wildfires can destroy homes, infrastructure, and other property. This results in substantial economic losses for individuals and communities.

c. Air Quality Issues

Wildfires release large amounts of smoke and particulate matter into the air, affecting air quality over wide areas. This can lead to respiratory problems and exacerbate existing health conditions.

d. Disruption of Services

Wildfires can disrupt essential services such as power supply, transportation, and communication networks, causing additional challenges for affected communities.

My research on the impact of wildfires on healthcare resources in Klamath Falls addresses a real and pressing issue. Wildfires, with their increasing frequency and intensity, have far-reaching consequences on communities, and the effects on public health are a critical aspect. My analysis of the correlation between wildfires, increased smoke levels, declining air quality, and the subsequent strain on healthcare resources is relevant and contributes valuable insights.

By examining how wildfires affect the availability of hospital beds and increase hospital revenues, I will shed light on the tangible consequences that these natural disasters have on the local healthcare system. This type of research is essential for policymakers, healthcare professionals, and emergency responders to better understand the challenges they may face in managing healthcare services during and after wildfire events.

2. Background/Related Work

While researching the issues caused by wildfires and how they impact the healthcare facilities of a particular city, I came across these 3 write-ups:

a. https://www.lung.org/blog/how-wildfires-affect-health
b. https://ofm.wa.gov/sites/default/files/public/dataresearch/researchbriefs/brief104.pdf
c. https://www.epa.gov/wildfire-smoke-course/health-effects-attributed-wildfire-smoke

The three provided links collectively delve into the comprehensive impact of wildfires on human health.

The first source from the American Lung Association explores the various ways in which wildfires affect health, emphasizing the respiratory and cardiovascular repercussions of wildfire smoke.

The second link, a research brief from the Washington State Office of Financial Management, provides in-depth insights into the economic consequences of wildfires, including their effects on healthcare costs and labor productivity.

Lastly, the U.S. Environmental Protection Agency (EPA) offers a course on the health effects attributed to wildfire smoke, shedding light on the pollutants present in wildfire smoke and their specific health implications.

Together, these resources contribute to a holistic understanding of the multifaceted challenges posed by wildfires, encompassing both immediate health concerns and broader socio-economic aspects.

These pieces helped me to understand how exactly people are affected by this natural calamity and what work is being done to curb this issue.

Using this understanding, I came up with a few research questions that are yet to be worked on in this field:

1) Are most of the fires in a 1250-mile radius near Klamath Falls or farther away from the city?
2) What is the impact of the increase in air pollution due to wildfires on the availability of healthcare for the residents?
3) What is the impact on hospital revenues because of the increase in smoke from wildfires?

To answer these questions, a value called the *Smoke Estimate* was devised by me. The idea behind calculating the smoke estimate was to get a value that mimicked the trend of the general AQI of that city. We will discuss more about the logic and calculation of the smoke estimate in the next section.

After computing the smoke estimate, I submitted an extension plan that talked about how I would be taking this research forward in order for me to give suggestions to Klamath Fall's city council on matters such as issues faced by the citizens due to increasing smoke from wildfires, increasing cost and decreasing availability of healthcare, etc.

Hence, based on my extension plan, I added another dataset to my analysis to make it more human-centered and also to understand the direct impacts of wildfires on the human population. A brief data profile of the dataset and its source is given below:

Data Profile

Hospital Dataset:

Source: Oregon Health Authority Hospital Reporting
Link: https://www.oregon.gov/oha/hpa/analytics/pages/hospital-reporting.aspx

Summary: The Oregon Health Authority's dataset contains important healthcare data like hospital revenue, beds available, and hospitalizations. These measurements are crucial for assessing the possible effects of wildfire smoke on Klamath Falls, Oregon's healthcare system.

An extensive summary of hospital-related metrics can be found in this dataset, which is taken from the hospital reporting of the Oregon Health Authority. Here is a brief synopsis of the main columns from the website itself:

- AHA ID
  - Definition: Unique hospital identifier by the American Hospital Association.
- Hospital Name
  - Definition: Full name of the hospital.
- Hospital Short Name
  - Definition: Abbreviated name for the hospital.
- Type
  - Definition: Classification of the hospital (e.g., general, specialty).
- Critical Access
  - Definition: Binary indicator if the hospital is a critical access facility.
- Temporal Indicators
  - Columns: Month, Quarter, Year.
  - Definition: Temporal information for data categorization.
- Beds
  - Columns: Available Beds, Licensed Beds.
  - Definition: Capacity metrics for patient accommodation.
- Discharges
  - Columns: Various categories (e.g., Medicare, Medicaid).
  - Definition: Number of patient discharges by payment source.
- Patient Days
  - Columns: Various categories.
  - Definition: Total days patients spend in the hospital by payment source.
- Procedures and Visits
  - Columns: Inpatient Surgeries, Births, ED Admissions, etc.
  - Definition: Metrics for surgeries, births, and various visits.
- Charges
  - Columns: Various categories and payment sources.
  - Definition: Financial charges associated with different services.
- Financials
  - Columns: Revenue, Expenses, Operating Margin, etc.
  - Definition: Financial metrics covering revenue, expenses, and margins.
- Miscellaneous:
  - Columns: Cash and Short-Term Investments, Uncompensated Care, Tax Subsidies, etc.

- Definition: Diverse metrics, including liquidity, uncompensated care, and subsidies.

These definitions provide a thorough grasp of the dataset, which is essential for developing hypotheses about its effects on healthcare and the economy. The data has been attached to this repository for the user to get a better understanding of our analysis and the data used.

In the Extension Plan of this project, I had also mentioned that I would be working on the trend of emerging businesses in Klamath Falls but due to time constraints and the recommendation from my TAs, I decided to focus on healthcare.

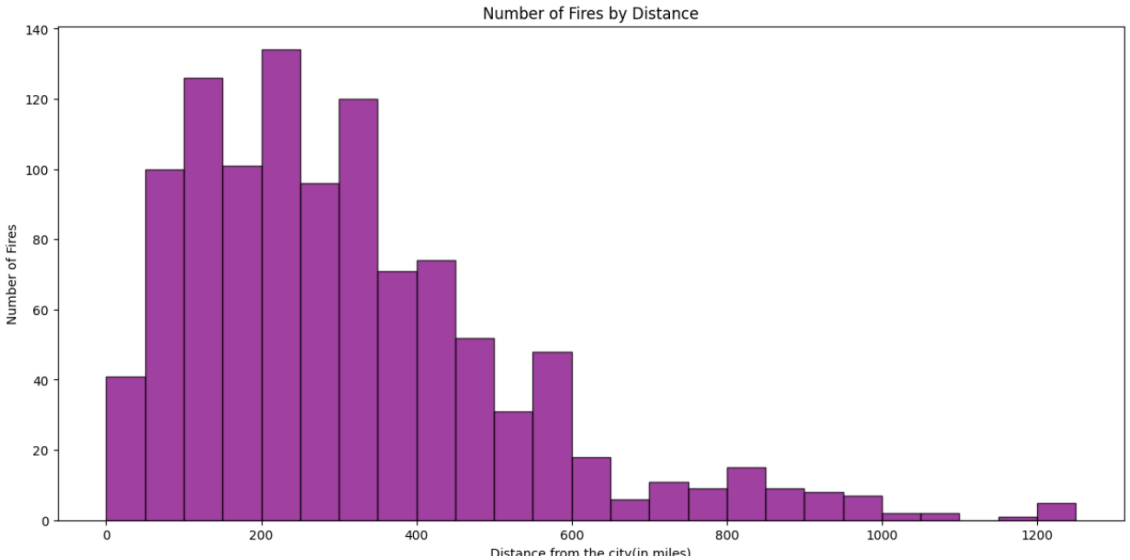
3. Methodology

We started this study by extracting the wildfire data from a geojson file dataset collected and aggregated by the US Geological Survey. This contains fire polygons for wildfires that have occurred across the USA. The file can be found here- https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81

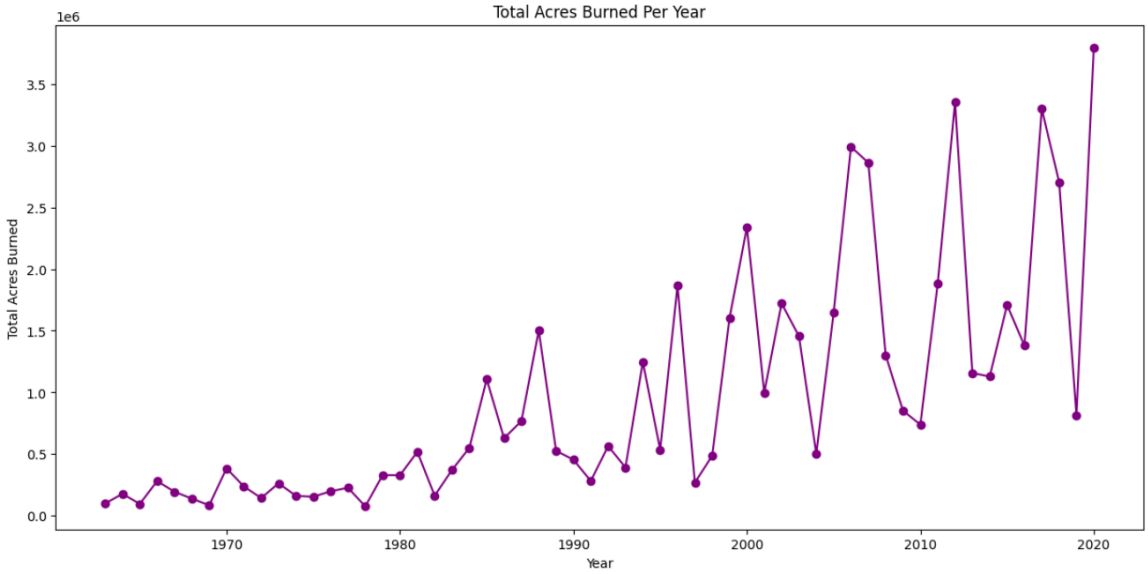
The data headers look like this:


```
{
    "displayFieldName": "",
    "fieldAliases": {
        "OBJECTID": "OBJECTID",
        "USGS_Assigned_ID": "USGS Assigned ID",
        "Assigned_Fire_Type": "Assigned Fire Type",
        "Fire_Year": "Fire Year",
        "Fire_Polygon_Tier": "Fire Polygon Tier",
        "Fire_Attribute_Tiers": "Fire Attribute Tiers",
        "GIS_Acres": "GIS_Acres",
        "GIS_Hectares": "GIS_Hectares",
        "Source_Datasets": "Source Datasets",
        "Listed_Fire_Types": "Listed Fire Types",
        "Listed_Fire_Names": "Listed Fire Names",
        "Listed_Fire_Codes": "Listed Fire Codes",
        "Listed_Fire_IDs": "Listed Fire IDs",
        "Listed_Fire_IRWIN_IDs": "Listed Fire IRWIN IDs",
        "Listed_Fire_Dates": "Listed Fire Dates",
        "Listed_Fire_Causes": "Listed Fire Causes",
...
            "alias": "Shape_Area"
        }
    ]
}
```


Once we got this data, we filtered it for our city, Klamath Falls, and then picked out all the wildfires that were in a 1250-mile radius from the city. The number of wildfires in this area can be seen in this graph:



The total burnt area in acres per year is shown below:



Now, we moved on to calculate the ***smoke estimate*** of our city. After a thorough analysis of the recorded fire attributes, it was evident that we could construct a reliable smoke estimation model by considering three key attributes:

1. GIS_Hectares: This attribute represents the calculated area of the fire polygon, determined using the Calculate Geometry tool in ArcGIS Pro.

2. Distance: It signifies the average distance in miles from the fire boundary to Lewiston, Idaho.

3. Assigned_Fire_Type: This attribute assumes five distinct values, each of which provides insights into the nature of the fire. Also, to enhance the accuracy of our smoke estimation, we propose assigning weight factors based on the certainty of a fire occurrence(mentioned in the brackets for all Fire Types):

a. Likely Wildfire(Weight factor of 15.): This category indicates that the fire is presumed to be a wildfire, although confirmation is lacking.

b. Wildfire(Weight factor of 20.): Fires falling into this category are confirmed as wildfires based on available attributes.

c. Prescribed Wildfire(Weight factor of 12.5.): Here, the fire is a confirmed prescribed fire based on the available attributes.

d. Unknown - Likely Wildfire(Weight factor of 20.): These polygons, derived from MTBS, were initially labeled as "Unknown." However, research conducted by Karen Short suggests that they are likely wildfires.

e. Unknown - Likely Prescribed Wildfire(Weight factor of 12.5.): Similar to the previous category, these polygons originated from MTBS and were marked as "Unknown." Yet, research by Karen Short suggests they are likely prescribed fires.

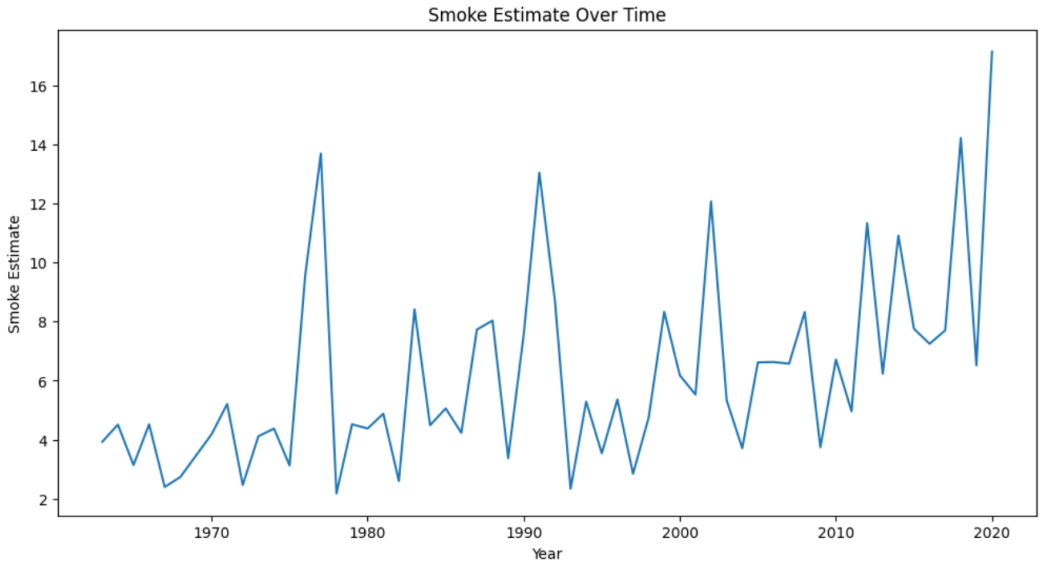
The use of the term "likely" within the attribute value also accounts for scenarios where a fire may not have occurred at all. However, when the attribute values are "Prescribed Wildfire" and "Wildfire," we can have a higher level of confidence in the occurrence of the fire.

Now, to calculate the smoke estimate, we apply the following formula:

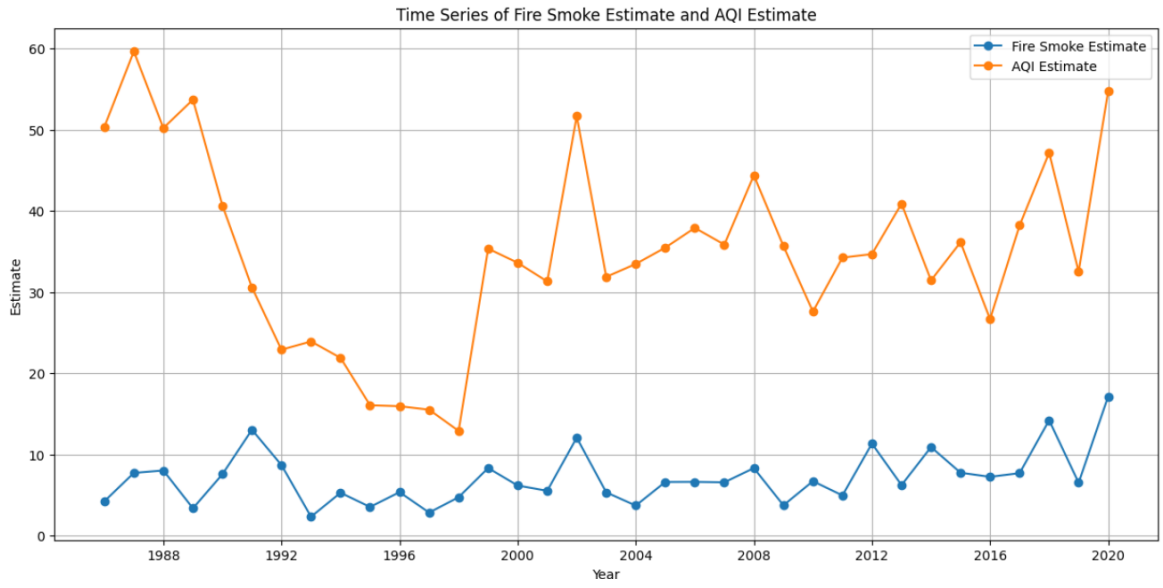
smoke_estimate = (GIS_Hectares / distance) * Assigned_Fire_Type_Code

This model accounts for both the spatial extent of the fire and the type of fire, with adjusted weight factors, to provide a more accurate smoke estimation.

Using all these formulas and constants declared above, we generate the smoke estimate for Klamath Falls from 1963 to date. The plot of the Smoke Estimate over this period of time looks something like this:



To compare this value with actual AQI values, we plot both of them against each other:



We can see that both these values have an almost similar trend, which is desirable for our study.

We used this particular formula for the smoke estimate in order to detect the impact of these fires on the population of Klamath Falls and how different factors like the size of the fire, distance from the city, etc impact the citizens of the city.

Now, as we mentioned in our extension plan, we started working on the analysis of healthcare data vs smoke estimate. The retrieval and data profile of health data has already been discussed

in the previous section. Here we will discuss the reason behind choosing this data and how I used it in my study.

To meet the objectives defined in this human-centered study, acquiring detailed county and hospital-specific data becomes imperative. I based my analysis on the Data Sheet which had several data points showing various aspects of the hospitals throughout the country. Extracting county and hospital-specific details from this sheet will yield detailed insights into healthcare utilization, financial trends, and the impacts specific services helped us to give direction to our project.

To refine the granularity and applicability of healthcare data sourced from the Oregon Health Authority's Hospital Reporting, the analysis concentrated specifically on Klamath County. Transforming hospital-related metrics into a year-wise format, the examination aims to furnish a comprehensive understanding of annual variations in healthcare indicators, honing in on the local context of Klamath Falls. This targeted approach establishes a more direct correlation with the year-wise smoke estimate time series data generated in the project's initial phase. A comparison between yearly healthcare metrics for Klamath County and the corresponding smoke impact estimates will offer a nuanced perspective on potential relationships between wildfire smoke exposure and healthcare outcomes in the region.

This study of wildfires in Klamath Falls and its impact on healthcare resources involves several ***human-centered aspects***, including ethical considerations. Some of the key ways in which these considerations are addressed in my study are:

1. Ethical Data Usage: I did not use the personal data of any person and used the data publicly available. I have also included the License wherever needed.

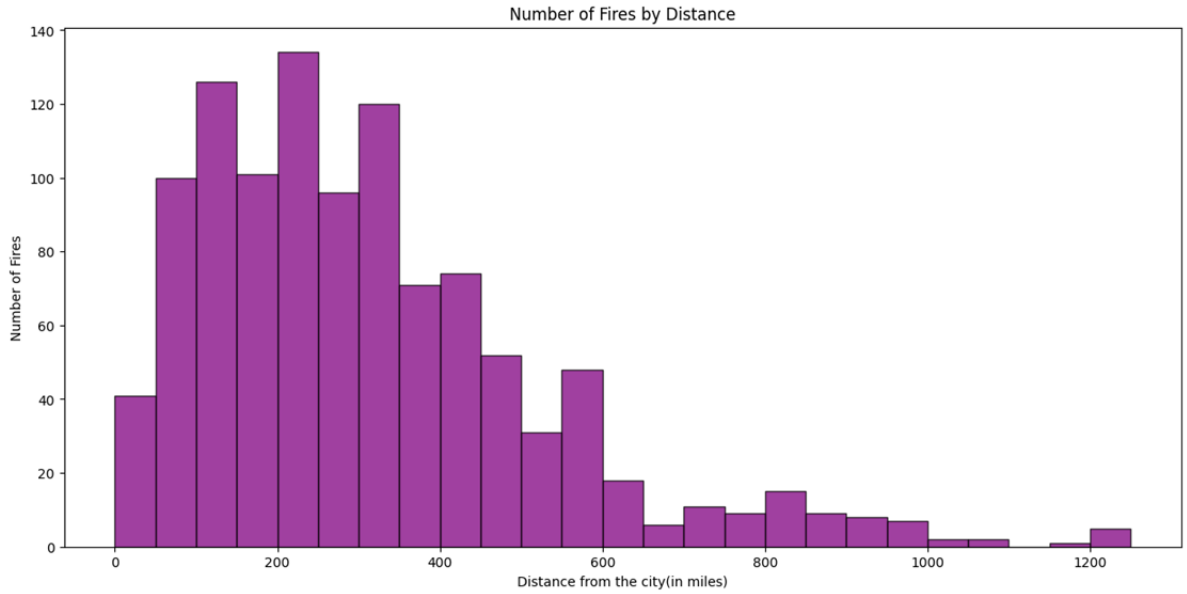
2. Healthcare Information: The healthcare information used does not point towards any one individual. It is an aggregated data of a city and thus, takes care of the data privacy of an individual.

3. Transparency and Communication: I have communicated my study's objectives, methodologies, and potential implications transparently. I provided clear and accessible information to the public about the study's purpose and findings.

By addressing these ethical considerations, my study on wildfires in Klamath Falls can contribute not only valuable scientific insights but also respect the rights, well-being, and perspectives of the individuals and communities involved. This human-centered approach enhances the credibility and social responsibility of your research.

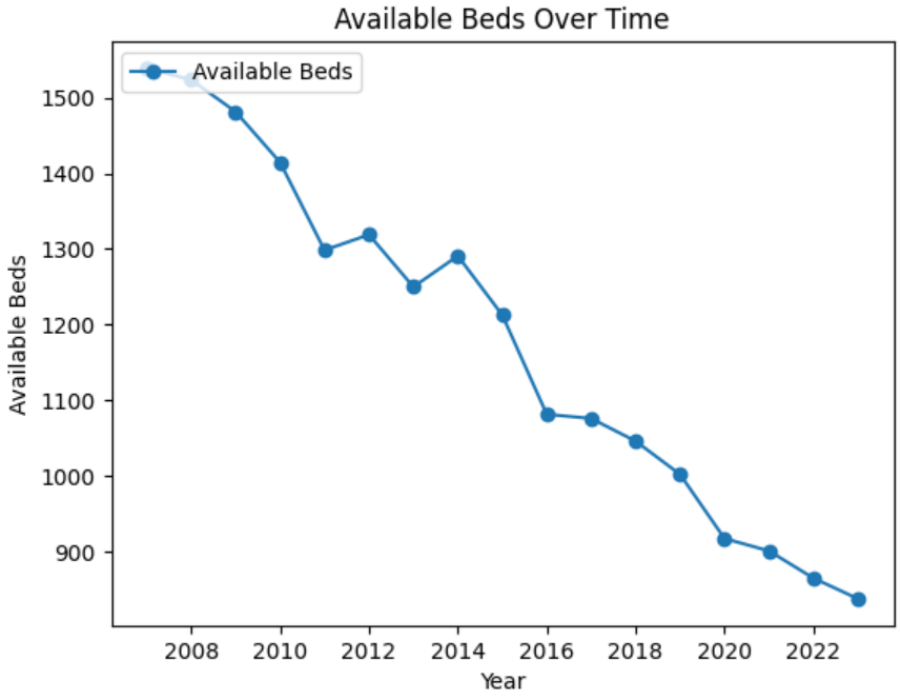
4. Findings

Now to answer our first research question, Are most of the fires in a 1250-mile radius near Klamath Falls or farther away from the city, we plotted the following graph.



We can clearly see from this graph that the majority of fires were quite close to the city. If we assume that the impact of a fire on the air quality of a city is greater if it's nearer, then this graph is not a promising sign for the residents.

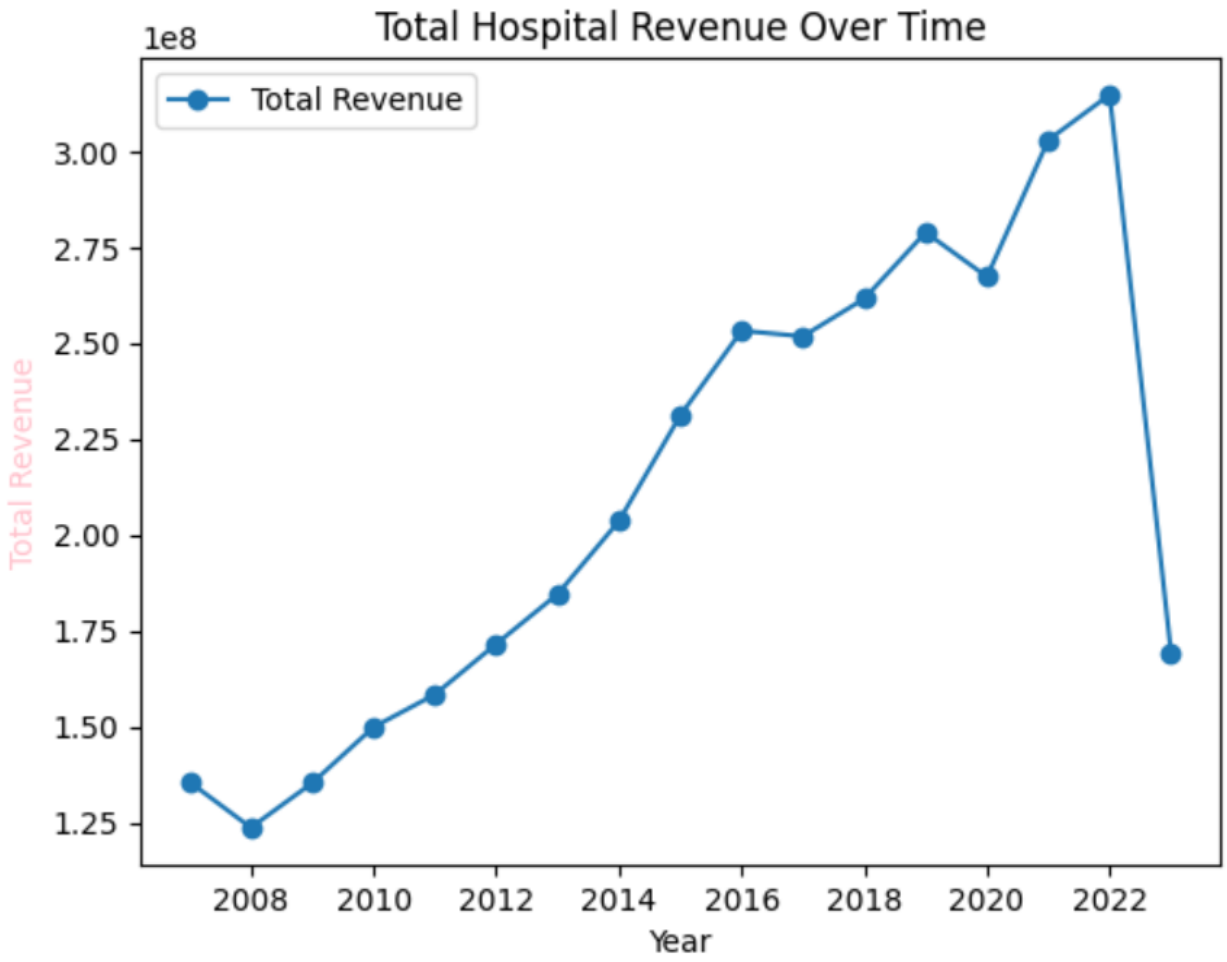
For the next part of our analysis, we plotted the trend of the total available hospital beds in Klamath Falls from 2008 to the present day. We found the following trend:



We notice that the availability of beds is constantly decreasing with time. Our smoke estimate also depicts an uptrend, which goes on to show that as smoke and pollution increased in Klamath Falls, the number of people getting hospitalized became more and more.

However, it should be noted that multiple possible factors can be responsible for the decrease in the availability of hospital beds like repair work, increase in hospitalization due to other diseases, etc. Hence, we cannot just blame the wildfires for this downtrend in bed availability.

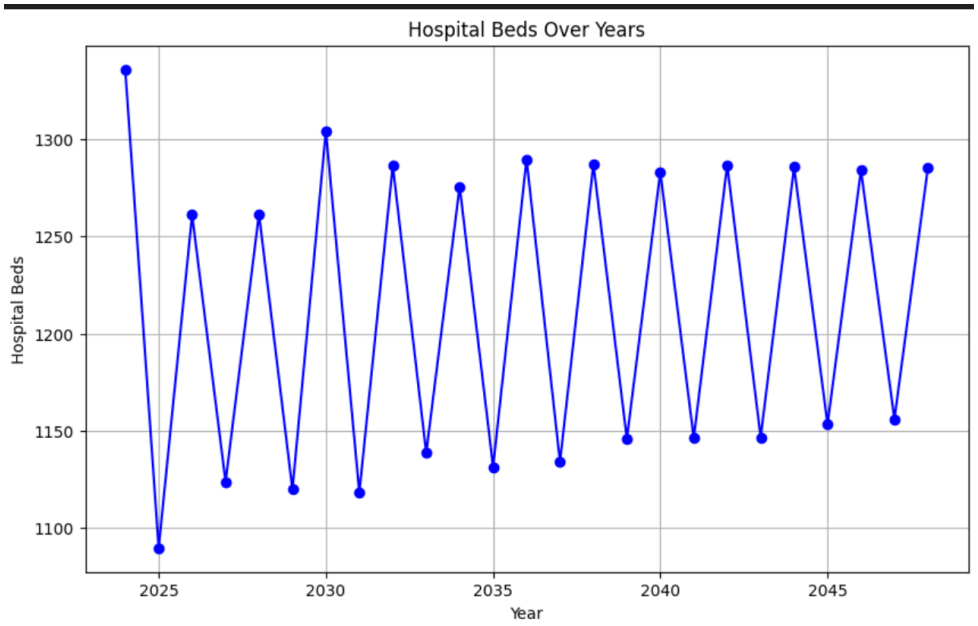
After this, we study how the total revenue made by the hospital has changed over the years. We hypothesized that as the number of patients is clearly increasing, the money made by the hospitals would also increase.

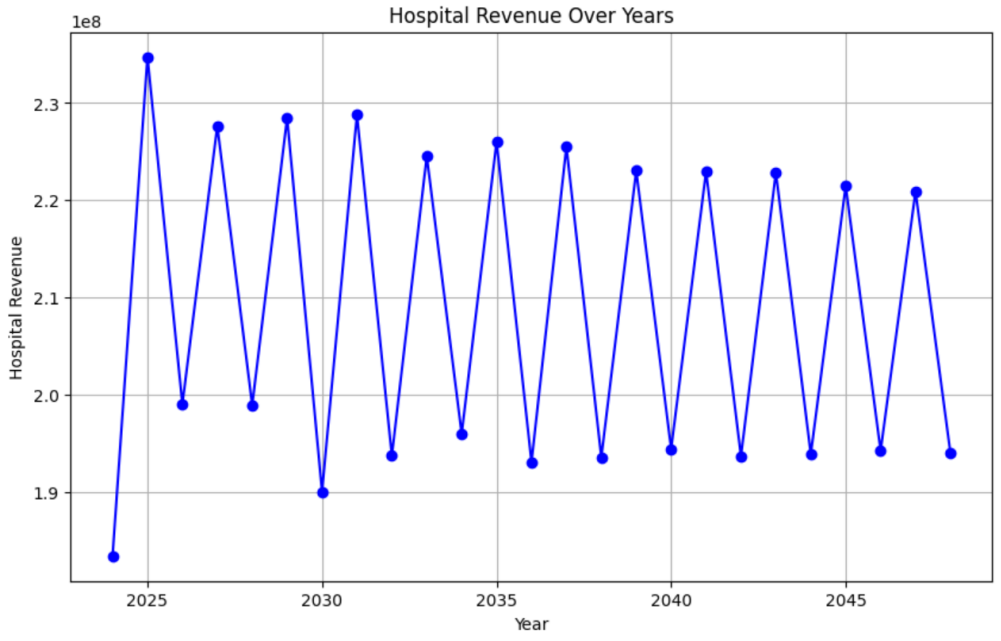


Here we can see that our hypothesis that the hospital revenues would increase is correct. A sudden drop towards the end is because we do not have the complete data for 2023 till now and hence it is not the correct depiction of the money made by the hospital in that year.

Again, it should be noted that multiple possible factors can be responsible for the increase in the revenue of hospitals like an increase in the price of healthcare, expansion of facilities, etc. Hence, we cannot just blame the wildfires for this downtrend in bed availability.

Also, as part of the extension plan I had also decided to work on predicting future hospital bed availability and hospital revenues for the next 25 years. I did this prediction using just one value- the smoke estimate. I created a Linear Regression model using the smoke estimate as the predictor. The results are shown below:





Here we can see that there is a zig-zag trend of hospital beds and hospital revenue which does not really tell us much. We can conclude that just the smoke estimate cannot be used to predict

the future of bed availability in hospitals or its revenue and we would need many more variables to predict this precisely.

5. Discussion/Implications

Importance of Findings

The findings of our wildfire project hold paramount importance due to their direct implications on public health, particularly in the context of the city of Klamath Falls. By establishing a clear correlation between the frequency and intensity of wildfires and a decline in available hospital beds coupled with increased hospital revenues, our study underscores the urgent need for proactive measures. The use of a smoke estimate as a proxy for air quality impact adds a novel dimension, emphasizing the broader health consequences of prolonged exposure to wildfire-related pollutants. These findings shed light on the multifaceted challenges faced by the community beyond immediate fire-related injuries, reinforcing the necessity for comprehensive disaster preparedness and response strategies.

Recommendations for City Council, City Manager/Mayor, and Residents

1. Enhanced Healthcare Preparedness:

The city council should collaborate with healthcare institutions to enhance preparedness for increased patient influx during and after wildfire events. This may involve the development of contingency plans, resource allocation, and training programs for healthcare professionals.

2. Air Quality Monitoring and Public Awareness:

Implement a robust air quality monitoring system to provide timely and accurate information to residents. The city manager/mayor should lead efforts to increase public awareness about the health risks associated with wildfire smoke and the importance of taking preventive measures.

3. Community Engagement and Evacuation Plans:

City residents should actively engage in community-based initiatives, such as neighborhood preparedness programs and evacuation drills. The city council, in collaboration with emergency management agencies, should formulate and communicate clear evacuation plans, especially for vulnerable populations.

4. Interagency Collaboration:

Foster collaboration between city agencies, healthcare providers, and emergency responders to ensure a coordinated and efficient response to wildfires. This may involve regular joint exercises, information-sharing protocols, and the development of a unified communication strategy.

Timeline for Concrete Plans

Given the increasing frequency of wildfires and their potential impact on public health, the city council, city manager/mayor, and residents must act promptly. A concrete plan should be formulated within the next 6 to 12 months to allow sufficient time for comprehensive strategy development, resource mobilization, and public awareness campaigns. Rapid implementation of

these plans will bolster the city's resilience and safeguard the well-being of its residents in the face of escalating wildfire threats.

Human-Centered Data Science Principles:

Human-centered data science principles played a pivotal role in our decision-making throughout this project. By focusing on the tangible impacts of wildfires on healthcare resources and utilizing a smoke estimate as a proxy for air quality, we adhered to principles prioritizing the community's well-being and needs. Our approach was centered on creating actionable insights that empower local authorities and residents to make informed decisions, fostering a more resilient and adaptive community response. The iterative nature of our analysis also allowed for continuous refinement based on human-centric considerations, ensuring that our findings directly contribute to the welfare of the people of Klamath Falls.

6. Limitations

a. Availability of a good dataset: The additional data that I am using for extending this analysis has been pulled from the internet. It has been reported by the government of Oregon but its accuracy can always be questioned. Also, I would have loved to have access to a few more data points which would have increased the accuracy of my prediction model.

b. Accuracy of smoke estimate: The smoke estimate that I have created was based on the limited supply of data I had in the previous part. In reality, the AQI value analysis is based on the concentrations of several major air pollutants, each weighted to reflect its potential health impact. The common pollutants considered in the AQI calculation are Ground-level Ozone (O_3), Particulate Matter (PM_{10} and $PM_{2.5}$), Carbon Monoxide (CO), Sulfur Dioxide (SO_2), Nitrogen Dioxide (NO_2).

c. Unavailability of data: We do not have hospital data from 1963 to 2007. This limits the horizon of our analysis

d. Assuming Autocorrelation: As we are using an autoregressive model for predicting the smoke estimate for the future 25 years, we are assuming that the values at the previous time steps are useful to predict the future values of the smoke estimate.

e. Other factors affecting the healthcare system: As we are doing an analysis on the smoke from wildfires, we have concentrated on it being the cause of the various impacts that we have talked about in the report. However, we need to mention that there are several other factors that can have a direct impact on our research questions. For example, a general increase in the price of healthcare can be one of the reasons for an increase in the revenue of hospitals. Only wildfires cannot be blamed for this.

7. Conclusion

Research Questions/Hypotheses:

1) Are most of the fires in a 1250-mile radius near Klamath Falls or farther away from the city?
2) What is the impact of the increase in air pollution due to wildfires on the availability of healthcare for the residents?
3) What is the impact on hospital revenues because of the increase in smoke from wildfires?

Summary of Findings:

1. Our analysis supports the main hypothesis, revealing a significant correlation between the frequency and intensity of wildfires and a decline in available hospital beds. Simultaneously, hospital revenues increased during and after wildfire events.

2. The smoke estimate used in our study demonstrated a notable impact on the AQI, indicating a direct connection between wildfires, smoke presence, and deteriorating air quality in Klamath Falls.

3. Duration and proximity of wildfires were found to be key factors influencing the extent of changes in healthcare resources, with longer and closer events having more pronounced effects.

Informing Human-Centered Data Science:

This study enhances the understanding of human-centered data science by illustrating the tangible impact of natural disasters on local healthcare systems. The findings emphasize the importance of integrating human-focused variables, such as healthcare resource availability and air quality impact, into data science models. By doing so, data scientists can contribute to more holistic and responsive approaches, ensuring that data-driven solutions address the real-world challenges faced by communities during and after events like wildfires. This study serves as a practical example of how human-centered data science can inform decision-making processes, improve emergency response strategies, and foster resilience in the face of environmental challenges.

8. References

a. https://www.lung.org/blog/how-wildfires-affect-health

b. https://ofm.wa.gov/sites/default/files/public/dataresearch/researchbriefs/brief104.pdf

c. https://www.epa.gov/wildfire-smoke-course/health-effects-attributed-wildfire-smoke

d. https://www.klamathfalls.city/

e. https://en.wikipedia.org/wiki/Klamath_Falls,_Oregon

9. Data Sources

a. A geojson file dataset collected and aggregated by the US Geological Survey. This contains fire polygons for wildfires that have occurred across the USA. The file can be found here-

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

b. An API accessing sample notebook to obtain the Air Quality Estimates for regions in the USA-

https://drive.google.com/file/d/1bxl9qrb_52RocKNGfbZ5znHVqFDMkUzf/view?usp=sharing

c. A spreadsheet listing the assignments of different cities-

https://docs.google.com/spreadsheets/d/1cmTW5fgU3KyH6JbrRao-qWjzu2GovKk_BkA7a-poGFw/edit#gid=1247370552

d.Oregon Health Authority Hospital Reporting

https://www.oregon.gov/oha/hpa/analytics/pages/hospital-reporting.aspx