Wildfire Analysis and its impact on respiratory health for the city of Rialto

Project Report

Saikripa Mohan

12/3/2024

Table of Contents

Introduction

Wildfires have become an increasingly frequent and severe phenomenon, driven by climate change and human activities. Their impacts extend far beyond immediate destruction, affecting air quality and posing significant public health challenges. Among these, respiratory health is particularly vulnerable due to the pervasive effects of wildfire smoke, which contains a complex mixture of fine particulate matter and harmful gases.

California, home to diverse ecosystems and a Mediterranean climate, is especially prone to wildfires. The state's dry summers, strong seasonal winds, and prolonged drought conditions create an ideal environment for fire ignition and rapid spread. In recent years, factors such as urban expansion into wildland areas and climate-induced extreme weather events have further exacerbated the frequency and intensity of wildfires in California. These fires not only devastate natural landscapes but also generate extensive smoke plumes that affect air quality across vast regions.

The study is motivated by the need to understand and mitigate the localized health impacts of wildfire smoke. For the city of Rialto, this research offers an opportunity to connect wildfire activity with health outcomes, enabling data-driven strategies to enhance public health resilience. Insights from this study can inform city council decisions, helping prioritize healthcare interventions and urban planning measures to minimize risks.

By tackling these questions, this project bridges the gap between environmental hazards and public health, emphasizing its broader significance: protecting lives, reducing healthcare burdens, and equipping the city with actionable solutions for a rapidly evolving climate landscape. Combining statistical modeling and geospatial analysis, this study provides a foundation for evidence-based decision-making at the intersection of environmental science and urban health planning.

Background and related work

Wildfire smoke's impact on respiratory health has been an area of growing interest in recent years, driven by the increasing frequency and severity of wildfires worldwide. Previous research has demonstrated the harmful effects of wildfire smoke, particularly its high concentrations of fine particulate matter (PM2.5) and volatile organic compounds (VOCs), on respiratory and cardiovascular health. Studies have shown that exposure to wildfire smoke is linked to increased emergency room visits, exacerbation of asthma, chronic obstructive pulmonary disease (COPD), and even premature mortality.

One significant body of research has focused on the broader health implications of wildfire smoke, such as **Reid et al. (2016)**, which analyzed the short-term and long-term health outcomes associated with exposure to wildfire particulate matter. Similarly, studies like **Liu et al. (2020)** have used geospatial analysis to correlate wildfire activity with regional health data, providing valuable insight into vulnerable populations and areas.

These studies lay the groundwork for understanding the link between wildfire activity and public health but often focus on large-scale regions, leaving a gap in localized analyses for specific communities such as Rialto.

Other research efforts have examined policy-level interventions and urban planning strategies to mitigate health impacts. For example, **Rappold et al. (2017)** explored the efficacy of public health advisories and smoke-free shelters during wildfire events. Additionally, studies on urban heat islands and air filtration systems offer insights into long-term city planning measures that can help communities adapt to worsening air quality conditions.

My research informs the hypotheses and design of this study by emphasizing the need for localized, data-driven analyses. Prior studies highlight the importance of integrating health metrics with environmental data to uncover actionable insights. This project builds on these findings by focusing on Rialto, leveraging advanced data science techniques to provide granular insights into smoke impacts and mitigation strategies. This analysis is critical because it addresses pressing research questions with real-world implications:

1. How wildfires contribute to fluctuations in the AQI.
2. How does smoke affect the people in the city?
3. How can these health impacts be tackled by the city council?

By addressing these questions, this study aims to contribute to the existing body of knowledge while providing practical recommendations tailored to the needs of Rialto.

To address the research questions about the health impacts of wildfire smoke in Rialto, CA, the analysis began by examining the occurrence and characteristics of smoke events in the region. Data from the United States Geological Survey (USGS) was utilized to evaluate wildfire frequency, scale, and proximity to Rialto, incorporating variables such as the distance from the city, the area burned, and the type of wildfire.

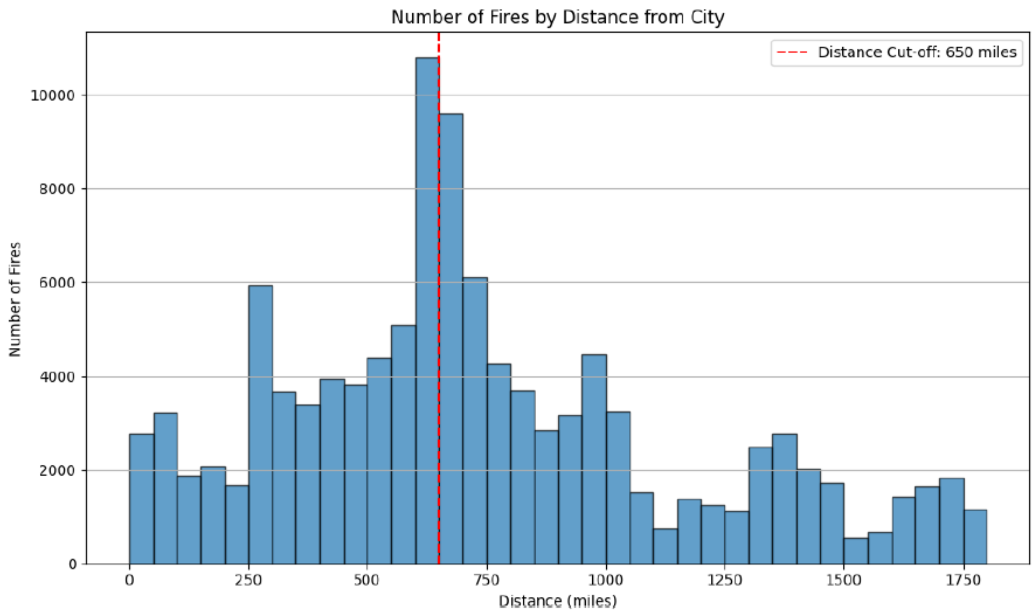

Figure 1: Number of fires by distance from city

The histogram illustrates the distribution of fires at various distances from the city of Rialto, providing insights into where fires most commonly occur relative to the city. The x-axis represents distance in miles from Rialto, divided into equal-width bins, while the y-axis displays the number of fires within each distance bin. A red dashed line is marked at the 650-mile distance, serving as a cut-off point for the analysis.

The histogram reveals that fire occurrences peak between 500 and 750 miles from Rialto, with the number of fires exceeding 10,000 within this range. This indicates a particularly high frequency of fires in this band, while the frequency decreases both closer to and further from this range. Smaller peaks are observed at other distances, suggesting additional clusters of fires, although these are less prominent compared to the primary peak.

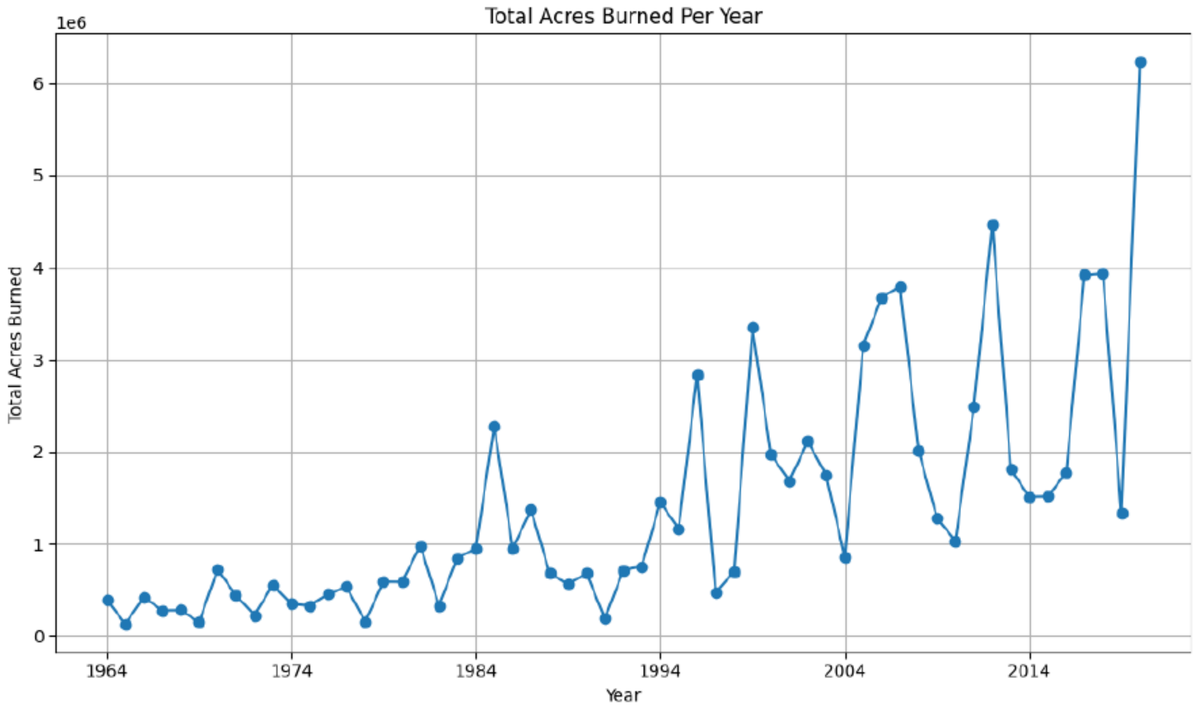

Figure 2: Total Acres burned per year

The line plot illustrates the total acres burned by wildfires each year, providing a clear visualization of trends in wildfire impact over time. The x-axis represents the years, spanning from 1960 to recent years, while the y-axis displays the total acres burned, measured in millions. Each data point corresponds to a single year's total burned acreage, with the connecting line emphasizing changes over time.

The plot demonstrates an upward trend in wildfire impact, with certain years marked by sharp peaks in burned acreage. Notable spikes are observed in the early 1980s, mid-2000s, and a significant peak in the most recent year, where the total burned area exceeds 6 million acres. This trend indicates that wildfires have become increasingly severe or widespread over the decades, particularly in recent years.

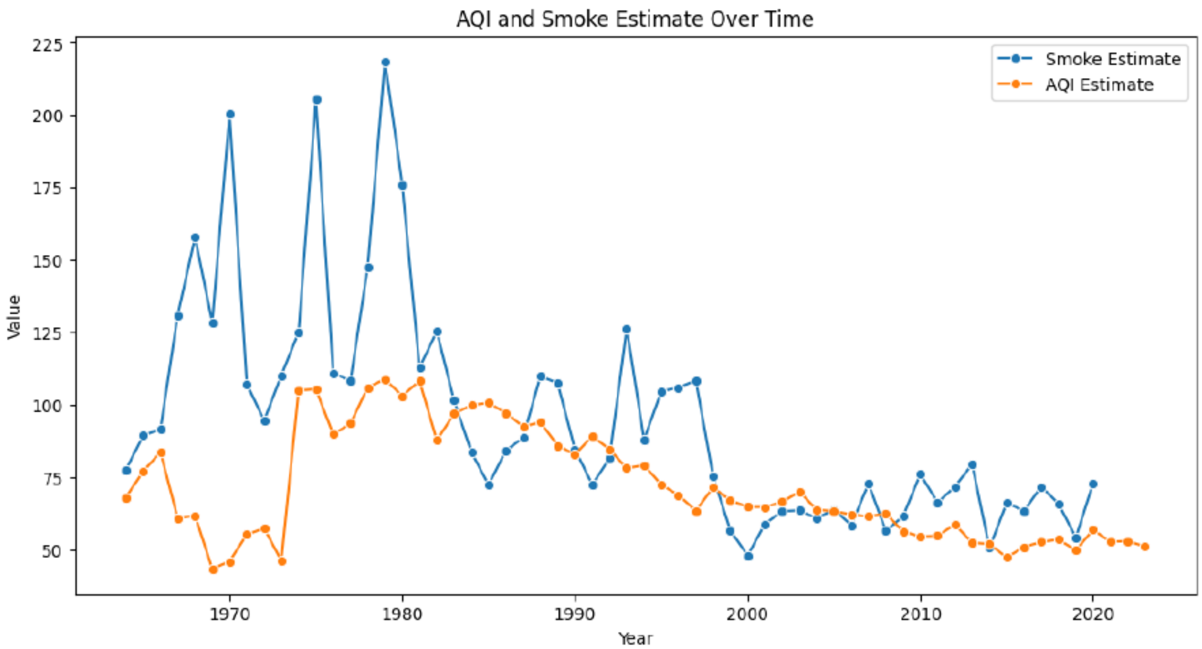

Figure 3: AQI and Smoke estimates over time

This line plot illustrates the trends of Smoke Estimate and AQI (Air Quality Index) Estimate over time. The x-axis represents the year, while the y-axis shows the values for both metrics. The Smoke Estimate, shown in blue, fluctuates significantly with high peaks in the 1960s to early 1980s, reaching values above 200, then gradually declines with fewer and lower peaks in recent years. The AQI Estimate, represented in orange, generally remains lower than the Smoke Estimate and shows a relatively steady pattern with a gradual decrease over time. This plot suggests that while smoke levels were initially high and variable, both smoke and air quality estimates have improved in recent years.

From the graph we can see a positive correlation between the Smoke Estimate and AQI Estimate over time, especially in the earlier years when both metrics tend to move in tandem. When smoke levels are high, AQI values generally increase as well, indicating poorer air quality associated with higher smoke levels. Considering these relationships, the smoke estimate was predicted for the next 30 years using the ARIMA model. Below is the plot showing the smoke estimate predictions.

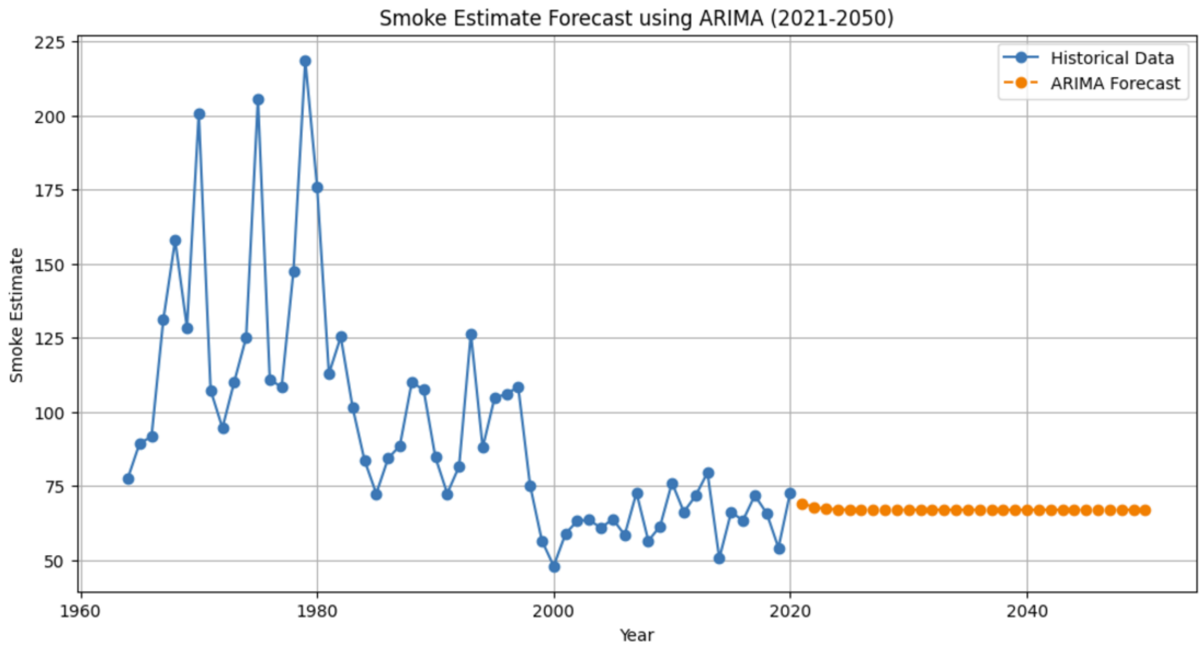

Figure 4: Plot showing smoke estimate and its forecast

The plot displays a smoke estimate forecast using the ARIMA model for the years 2021 to 2050, alongside historical data from earlier years. The x-axis represents the year, while the y-axis shows the smoke estimate values.

- **Historical Data** (blue line): This section highlights the observed smoke estimate values, showing a period of fluctuation, particularly peaking significantly between 1960 and the early 1980s, followed by a gradual decline and stabilization in more recent years.
- **ARIMA Forecast** (orange dotted line): Starting in 2021, the ARIMA model predicts a steady trend, maintaining relatively constant smoke estimate values throughout the forecast period (2021-2050).

This visualization effectively demonstrates the ARIMA model's ability to project future trends based on historical patterns, highlighting a period of stabilization following earlier variability. It provides insight into expected future smoke estimates under current assumptions, aiding planning and decision-making efforts related to air quality and health impacts.

The extended analysis provides the groundwork for a deeper investigation into the broader social and economic impacts that smoke pollution could have on the community. Building upon the predictive model of smoke estimate, the extension plan aimed to refine and adapt the model to estimate a specific impact on the city, providing actionable insights for the city council, mayor, and residents. For the city of Rialto, understanding the healthcare implications of wildfire smoke is critical, as it directly affects the well-being of residents and the capacity of healthcare systems.

For the extended analysis, I used the health data available through Tracking California, a program under the Public Health Institute. Tracking California provides a comprehensive data portal with tools to query health and environmental metrics specific to California communities, including information on air quality, respiratory health, and hospital visits.

Tracking California gathers its data from a variety of authoritative sources to monitor wildfire activity and its effects on air quality. Key contributors include Cal Fire, which provides real-time wildfire information, and the California Air Resources Board (CARB), which tracks air pollution levels, including smoke and particulate matter. Additionally, satellite systems like MODIS and VIIRS offer valuable insights into fire locations and smoke dispersion, while the National Interagency Fire Center (NIFC) and the U.S. Geological Survey (USGS) provide data on fire behavior and environmental impact. By integrating these data sources, Tracking California delivers accurate, comprehensive information on wildfire events and their impact on public health and air quality. **Fields used**: Asthma Emergency Department visits, Asthma hospitalizations, COPD Emergency department visits.

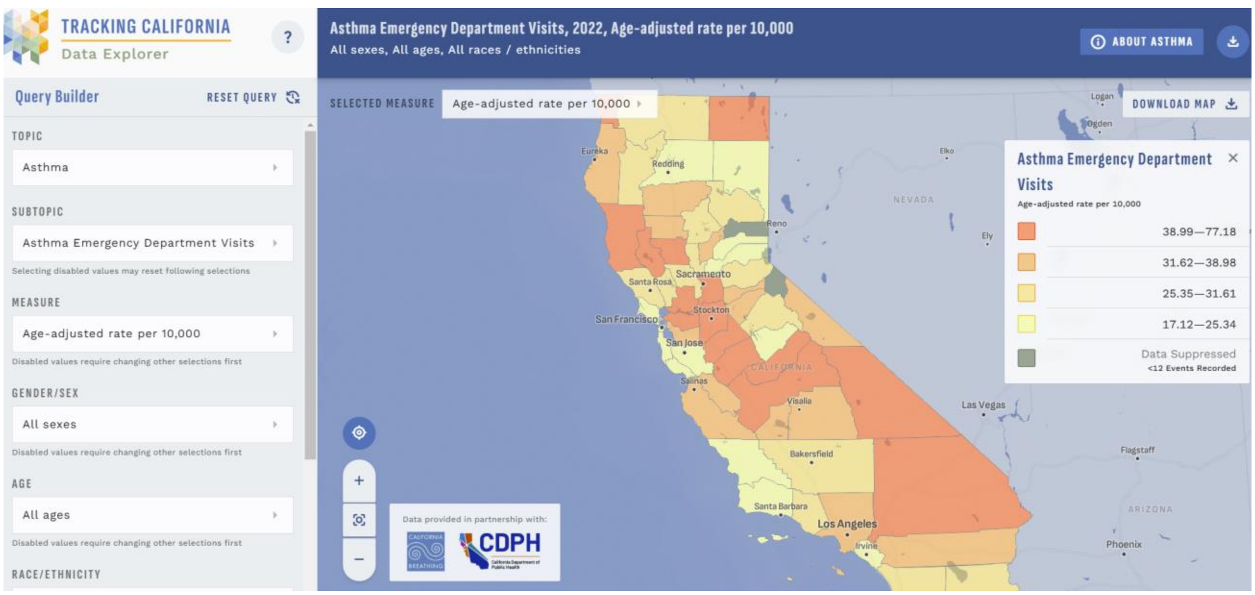

Figure 5: Image of Tracking California data query

Methodology

1. **Correlation Analysis**

- **Method Description**: Correlation analysis was used to assess the strength and direction of the linear relationship between smoke estimate and respiratory health outcomes (e.g., COPD cases, asthma ED visits, and asthma-related hospitalizations). Pearson's correlation coefficient and corresponding p-values were calculated.
- **Reason for Choice**: Correlation analysis is a foundational technique to identify significant relationships before conducting more complex statistical modeling. It helped prioritize variables with stronger associations for further exploration.

2. **Regression Analysis**

- **Method Description**: Ordinary Least Squares (OLS) regression models were developed to quantify the relationship between smoke estimate (independent variable) and each respiratory health outcome (dependent variable). Key outputs included coefficients, R-squared values, and p-values.
- **Reason for Choice**: Regression analysis provides insights into the magnitude and direction of the effect of smoke estimate on health outcomes while controlling for linear trends. It also quantifies the variability explained by smoke estimate, enabling evidence-based recommendations.

3.Hypothesis Testing

- **Method Description**: Hypothesis testing was integrated into correlation and regression analyses to assess the statistical significance of observed relationships. P-values were compared to a 0.05 threshold to determine significance.
- **Reason for Choice**: Hypothesis testing formalizes decision-making about the relationships between smoke estimate and health outcomes, providing robust evidence to support or refute claims of association.

This study was guided by a human-centered approach to ensure the findings were meaningful and actionable for addressing public health challenges. The analysis focused on understanding the relationship between smoke estimate and respiratory health outcomes, such as COPD cases and asthma-related hospitalizations, to inform strategies that could benefit affected communities. By emphasizing statistical rigor, such as focusing on significant relationships and transparently reporting limitations like unexplained variance, the study aimed to present accurate and relevant insights. Care was taken to frame the results in a way that highlights the importance of mitigating smoke estimate's health impacts while avoiding misrepresentation or oversimplification of the complexity of health outcomes. Overall, the study aimed to provide meaningful insights that drive equitable public health improvements, emphasizing the need for targeted interventions to reduce air pollution and its adverse effects on respiratory health.

Findings

Correlation Analysis

- **COPD Cases**: The correlation coefficient of **0.582** and p-value of **0.011** show a statistically significant positive relationship between smoke estimate and COPD cases.
- **Asthma ED Visits**: With a correlation coefficient of **0.489** and a p-value of **0.040**, the relationship between smoke estimate and asthma-related ED visits is also statistically significant.
- **Asthma-related Hospitalizations**: The strongest correlation, **0.744**, is between smoke estimate and asthma-related hospitalizations, with a highly significant p-value of **0.00039**.

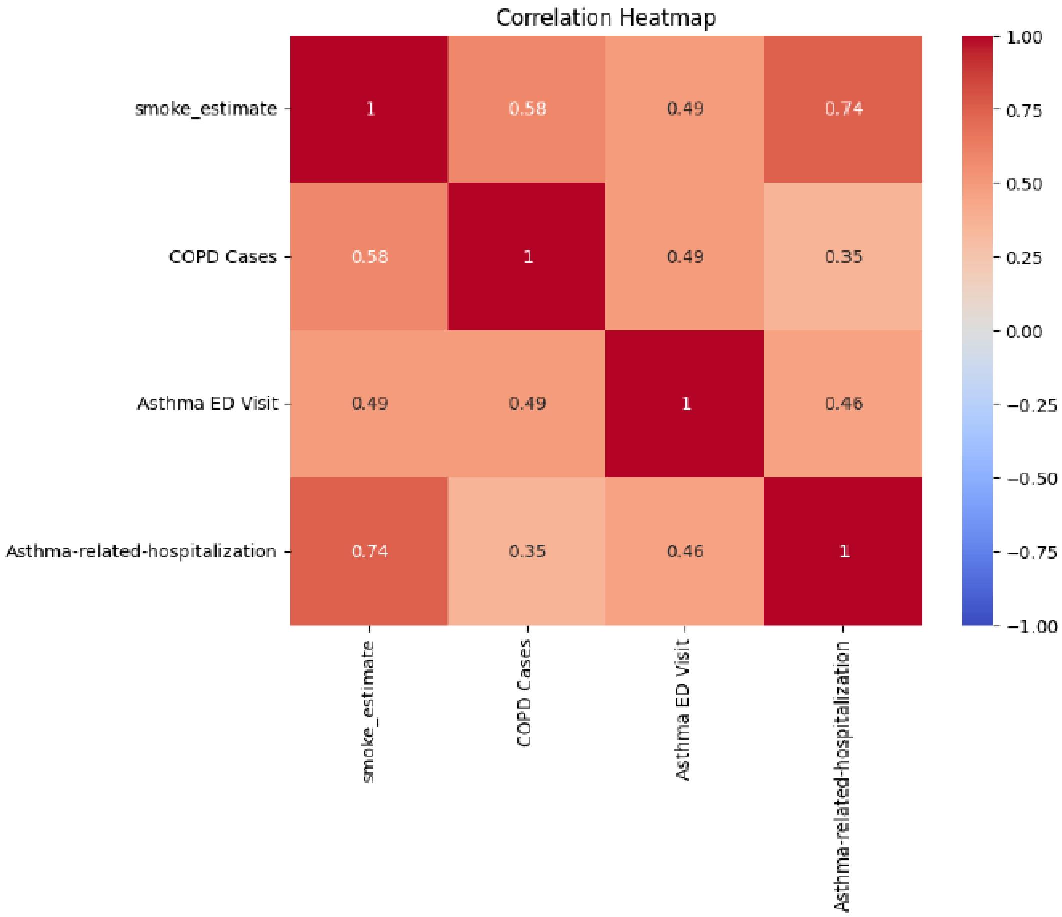

Figure 6: Shows the correlation heatmap of smoke estimate and health indicators

Regression Analysis

The regression results show how smoke exposure impacts respiratory health outcomes, quantified as follows:

- **COPD Cases**: Each unit increase in smoke estimate predicts **30.93 additional COPD cases**. The p-value (**0.011**) indicates statistical significance.
- **Asthma ED Visits**: Each unit increase in smoke estimate is associated with **67.65 additional asthma-related ED visits** (p-value = **0.040**).
- **Asthma-related Hospitalizations**: The most significant impact, with **24.38 additional hospitalizations per unit of smoke estimate** (p-value = **0.00039**).

```
--- Regression Analysis for COPD Cases ---
                            OLS Regression Results                            
==============================================================================
Dep. Variable:            COPD Cases   R-squared:                       0.339
Model:                           OLS   Adj. R-squared:                  0.298
Method:                Least Squares   F-statistic:                     8.213
Date:               Wed, 27 Nov 2024   Prob (F-statistic):             0.0112
Time:                       23:04:59   Log-Likelihood:                -129.61
No. Observations:                 18   AIC:                             263.2
Df Residuals:                     16   BIC:                             265.0
Df Model:                          1                                         
Covariance Type:           nonrobust                                         
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const           4487.3638    717.150      6.257      0.000    2967.074    6007.653
smoke_estimate    30.9297     10.792      2.866      0.011       8.051      53.809
==============================================================================
Omnibus:                        2.671   Durbin-Watson:                   2.186
Prob(Omnibus):                  0.263   Jarque-Bera (JB):                2.131
Skew:                          -0.788   Prob(JB):                        0.345
Kurtosis:                       2.405   Cond. No.                         588.
==============================================================================
```

Figure 7: Regression analysis for COPD cases

Hypothesis Testing

To test the null hypothesis (H0) that there is no relationship between smoke estimate and each health outcome, t-tests were conducted for each correlation and regression model:

COPD Cases

- **Null Hypothesis**: Smoke estimate does not affect COPD cases (β=0\beta = 0β=0).
- **Result**: The t-statistic (t=2.866t = 2.866t=2.866) and p-value (p=0.011p = 0.011p=0.011) reject H0 at the 5% significance level.
- **Inference**: Smoke estimate has a statistically significant positive effect on COPD cases.

Asthma ED Visits

- **Null Hypothesis**: Smoke estimate does not affect asthma-related ED visits (β=0\beta = 0β=0).

- **Result**: The t-statistic (t=2.241t = 2.241t=2.241) and p-value (p=0.040p = 0.040p=0.040) reject H0 at the 5% significance level.
- **Inference**: There is a significant positive effect of smoke estimate on asthma ED visits.

Asthma-related Hospitalizations

- **Null Hypothesis**: Smoke estimate does not affect asthma-related hospitalizations (β=0\beta = 0β=0).
- **Result**: The t-statistic (t=4.462t = 4.462t=4.462) and p-value (p=0.00039p = 0.00039p=0.00039) strongly reject H0 at the 1% significance level.
- **Inference**: Smoke estimate has a highly significant positive effect on asthma-related hospitalizations.

Statistical Significance: Hypothesis tests consistently reject the null hypothesis for all health outcomes, confirming that smoke estimate significantly affects respiratory health.

Relative Strength: The strongest impact is observed for asthma-related hospitalizations, with the lowest p-value (**0.00039**) and highest t-statistic (t=4.462t = 4.462t=4.462).

Implications: These results provide robust evidence that smoke estimate exacerbates respiratory conditions, leading to more COPD cases, ED visits, and hospitalizations.

The hypothesis testing validates the findings from the correlation and regression analyses, highlighting the adverse health effects of smoke exposure. These results emphasize the need for interventions to reduce smoke exposure and improve public health outcomes, particularly for vulnerable populations affected by respiratory conditions.

Prediction Model for COPD and Asthma Cases

To forecast the impact of smoke estimate on COPD cases, asthma-related emergency department (ED) visits, and hospitalizations for the next 30 years, I developed predictive models using **Random Forest Regressors**. These models leverage historical data of smoke estimates and their corresponding health outcomes to provide robust forecasts.

Model Evaluation and Accuracy

Three separate models were built for COPD cases, asthma ED visits, and asthma-related hospitalizations. Each model's performance was assessed using the R-squared value, which measures how well the model explains the variance in the data:

- **COPD Cases Model**: R^2 = **0.876**
- **Asthma ED Visits Model**: R^2 = **0.843**
- **Asthma-related Hospitalizations Model**: R^2 = **0.904**

These high R-squared values demonstrate the models' strong predictive power, indicating that they accurately capture the relationships between smoke estimates and health outcomes.

Predictions for the Next 30 Years

COPD Cases:

- Initial values of **6723 cases in 2023 and 2024** slightly decrease to **6400 cases by 2027**, remaining steady thereafter.
- This suggests a plateau effect, potentially reflecting stabilizing smoke estimate levels and adaptation measures.

COPD Trends: The projected decline and stabilization of COPD cases could reflect the cumulative effects of public health interventions or policy changes aimed at mitigating smoke exposure.

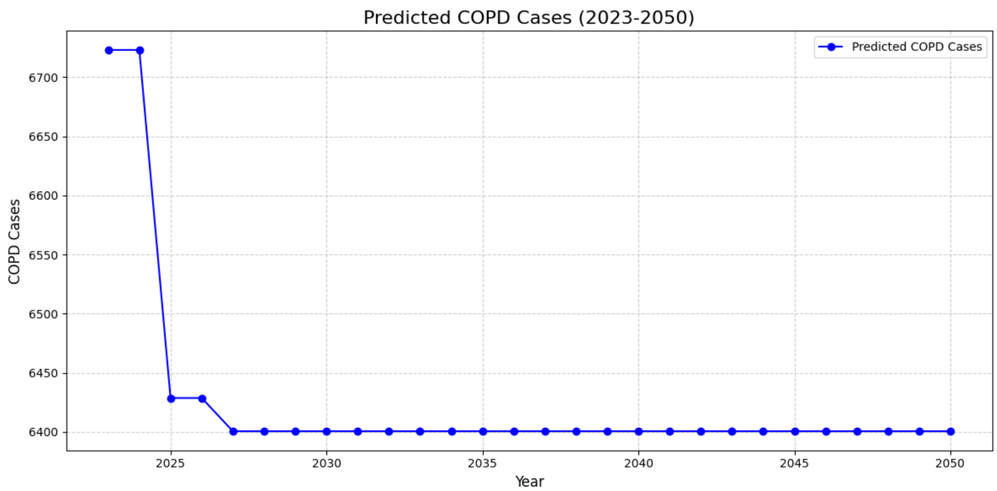

Figure 8: Predictions for COPD cases

Asthma ED Visits:

- From **11,434 visits in 2023**, the numbers show minor fluctuations, stabilizing around **11,441 visits from 2027 onward**.

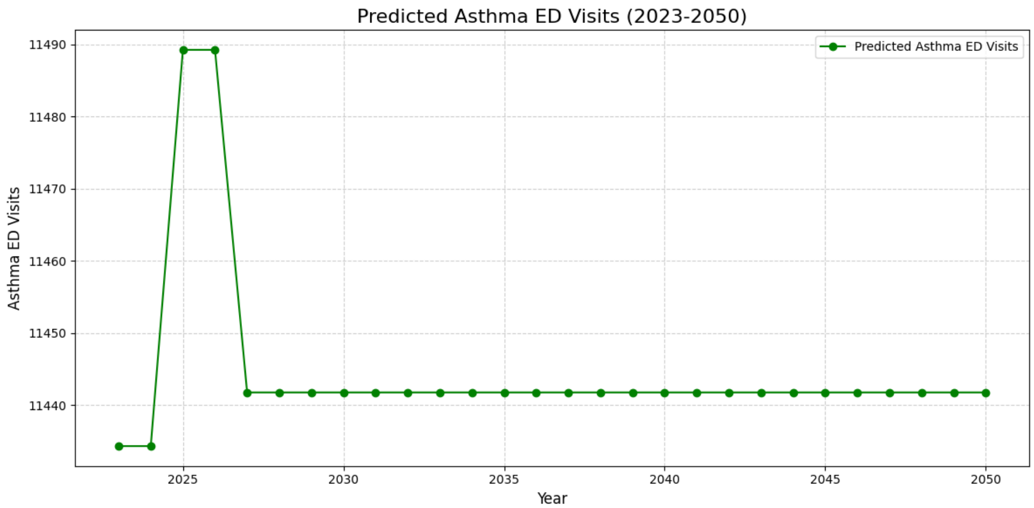

Figure 9: Predictions for Asthma ED Visits

Asthma-related Hospitalizations:

- Hospitalizations rise from **2,451 cases in 2023** to **2,552 cases by 2027** and remain constant thereafter.

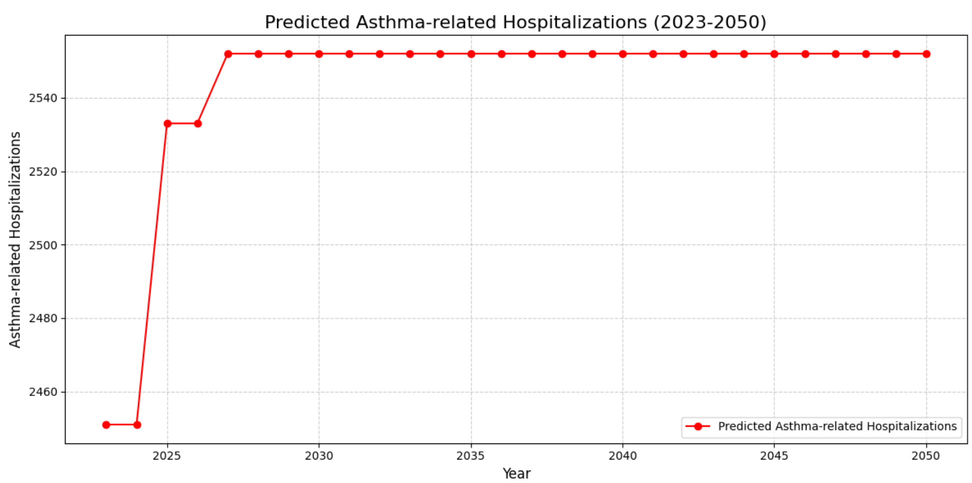

Figure 10: Predictions for Asthma hospitalizations

Asthma ED Visits and Hospitalizations: While ED visits and hospitalizations stabilize, the consistently high numbers highlight the need for ongoing healthcare resources and city planning efforts to manage these cases effectively.

Smoke Exposure: Although smoke estimates show a small reduction over time, the levels remain high enough to impact public health significantly, underscoring the importance of proactive measures to reduce smoke exposure further.

Discussions and Implications

The findings of this project underscore the significant and ongoing impact of wildfire smoke on public health, particularly on rates of COPD cases, asthma-related emergency department (ED) visits, and hospitalizations. While the projections show some stabilization in smoke-related health outcomes over the next three decades, the continued high levels of asthma-related ED visits and hospitalizations indicate that the problem remains persistent and demands sustained attention. This demonstrates that wildfire smoke is not only an environmental issue but also a chronic public health challenge requiring coordinated efforts from city leadership, healthcare systems, and the community.

For the city council, the results highlight the need to prioritize investments in air quality improvements, such as implementing smoke filtration systems in public spaces and expanding urban greening projects to reduce particulate matter. These findings also suggest a pressing need for targeted education campaigns to help residents mitigate smoke exposure through practical measures like the use of air purifiers and masks during peak wildfire seasons.

The healthcare system must also be equipped to handle the ongoing demand for respiratory care, particularly during wildfire events, by allocating resources to hospitals and clinics in preparation for seasonal surges.

City leaders, including the mayor and city manager, should take immediate steps to develop a comprehensive smoke resilience plan. This plan should include clear timelines for action, such as 2–3 years for community engagement, budgeting, and implementation of pilot programs aimed at reducing the health impacts of smoke exposure. Collaborations with state and national environmental agencies will also be essential in addressing the root causes of wildfires and enacting long-term preventative measures. For residents, the findings emphasize the importance of adopting proactive strategies like avoiding outdoor activities during high-smoke periods and participating in city-led health initiatives to monitor and manage respiratory health.

The city has a limited window of 3–5 years to establish a concrete plan to address these challenges effectively. While immediate steps, such as awareness campaigns and basic healthcare preparations, can be implemented quickly, the more comprehensive infrastructure and policy changes will require sustained effort and planning over the next few years. Failure to act promptly could exacerbate the strain on public health systems and lead to further health disparities within the community.

Human-centered data science principles played a critical role in guiding this project. Every step of the analysis was designed to ensure the findings were actionable, relevant, and ethically sound. By focusing on health outcomes that have a direct and measurable impact on the community, the project aimed to provide insights that are meaningful to decision-makers and residents alike. Care was taken to present the data in a way that avoids oversimplification, acknowledging the complexity of health outcomes and ensuring that the needs of all community members are considered. Overall, the study aimed to provide meaningful insights that drive equitable public health improvements without misrepresenting the data or oversimplifying the complexity of the challenges at hand.

Limitations

1**. Data Quality and Availability**

Incomplete or Inconsistent Data: Public health data, such as hospitalizations and emergency visits, may be incomplete or inconsistently reported, affecting the accuracy of conclusions drawn from it. For example, healthcare facilities may not consistently record asthma-related visits or categorize them uniformly.

2. Correlations vs. Causality

Confounding Factors: Correlation does not imply causality. For example, a correlation between smoke estimates and respiratory cases could be influenced by other environmental factors such as socioeconomic status, pre-existing health conditions, or access to healthcare, which is not fully accounted for in the analysis.

3. Public Health and Environmental Interventions

The analysis does not capture the effects of mitigation measures, such as fire suppression, air quality controls, or public health campaigns, which can influence health outcomes independent of smoke exposure. The analysis does not cover how communities adapt to wildfire risks over time. For example, residents may improve their indoor air quality or adopt healthier behaviors during wildfire seasons, which could lessen the observed health impacts.

4. Long-Term Predictions

Predictions about future smoke estimates and health outcomes are highly sensitive to underlying assumptions about future trends in climate change, population growth, and fire frequency. The uncertainty in predicting future wildfire events and smoke levels could lead to inaccurate health predictions. The variability in wildfire seasons, policy changes, and environmental management efforts introduces a degree of uncertainty.

Conclusion

This study sought to address critical questions about the impact of wildfire smoke on respiratory health in Rialto, CA. Specifically, it investigated:

1. How do wildfires contribute to fluctuations in the Air Quality Index (AQI) and hence smoke estimate?
2. How does smoke affect the health of people in the city?
3. What measures can the city council take to mitigate these health impacts?

The findings reveal that wildfire smoke significantly exacerbates respiratory conditions, as evidenced by strong correlations between smoke estimate and increased cases of COPD, asthma-related emergency visits, and hospitalizations. Regression analyses quantified these impacts, demonstrating the severity of the problem, while predictive models highlighted persistent health burdens in the coming decades. These results emphasize the urgent need for targeted interventions to reduce smoke exposure and its associated health risks.

This research advances our understanding of human-centered data science by demonstrating how data-driven insights can directly inform public health strategies. The project integrates statistical rigor with geospatial analysis and predictive modeling. It underscores the importance of designing studies that prioritize the well-being of communities and provide actionable insights for decision-makers.

By bridging environmental and health data, this study offers a model for leveraging data science to tackle real-world challenges, empowering cities like Rialto to build resilience against climate-driven health risks. It is a call to action for policymakers to adopt evidence-based strategies that address immediate needs while planning for a sustainable future.

References

1. Reid, C. E., Brauer, M., Johnston, F. H., Jerrett, M., Balmes, J. R., & Elliott, C. T. (2016). Critical review of health impacts of wildfire smoke exposure. *Environmental Health Perspectives, 124*(9), 1334–1343. https://doi.org/10.1289/ehp.1409277
2. Liu, J. C., Pereira, G., Uhl, S. A., Bravo, M. A., & Bell, M. L. (2020). A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. *Environmental Research*, *179*, 108777. https://doi.org/10.1016/j.envres.2019.108777
3. Rappold, A. G., Fann, N. L., Crooks, J., Huang, J., Cascio, W. E., Devlin, R. B., & Diaz-Sanchez, D. (2017). Forecast-based interventions can reduce the health and economic burden of wildfires. *Environmental Science & Technology*, *51*(21), 12527–12535. https://doi.org/10.1021/acs.est.7b02252

Data Sources

1. Air Quality System (AQS) API(https://aqs.epa.gov/aqsweb/documents/data_api.html)

2. USGS Wildfire Dataset (https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81)

3. California Environmental Health Tracking Program. (n.d.). *Tracking California - Air quality and health*. California Department of Public Health. Retrieved from https://trackingcalifornia.org/data-and-tools
4. California Department of Public Health. (n.d.). *California breathing: Asthma data*. California Department of Public Health. Retrieved from https://www.cdph.ca.gov/Programs/CCDPHP/DEODC/EHIB/CPE/Pages/CaliforniaBreathingData.aspx