**Wildfire Smoke Impact Analysis**

**And its effect on asthma**

**Project Report**

**By Priyam Gupta**

# Introduction

The relationship between wildfires and public health represents a critical area of study, particularly for communities in the western United States. This analysis investigates the correlation between wildfire events and respiratory health outcomes in Glendale, Arizona, using a novel Smoke Impact Score methodology. The research is particularly timely given the observed increase in wildfire frequency and intensity over recent decades, as evidenced by historical data showing a significant upward trend in total acres burned since the 1990s.

The analysis addresses a pressing public health concern: understanding how wildfire smoke affects respiratory health in urban populations. While previous studies have examined general air quality impacts, this research specifically focuses on quantifying the relationship between wildfire characteristics (size and distance) and health outcomes. This understanding is crucial for city officials and healthcare providers to develop targeted interventions and preparedness strategies.

# Background/Related Work

**Existing Research and Its Influence**

The development of the Smoke Impact Score is rooted in prior research on wildfire smoke dispersion and its impact on air quality and health outcomes. Existing studies have highlighted the complexity of wildfire smoke patterns, which are influenced by multiple environmental factors such as wind speed, direction, atmospheric pressure, and topography, rather than being solely dependent on fire size or proximity. These insights revealed that traditional distance-based models, such as inverse-square relationships, often fail to adequately capture smoke impacts, especially at medium-to-long distances from the fire source.

This research informed our hypothesis that a refined metric could better quantify smoke impacts on respiratory health outcomes, such as asthma hospitalizations, by incorporating additional variables like fire size and distance. The key hypothesis driving this analysis was:

* **Hypothesis:** Wildfire size and proximity directly impact respiratory health outcomes, but a simple distance-based approach is insufficient. A modified, context-sensitive scoring model is required to improve predictive accuracy.

This understanding shaped the development of the Smoke Impact Score and guided the analysis framework, ensuring the model captured the nuances of wildfire smoke dispersion patterns.

**Research Questions**

Based on existing studies and our analysis objectives, the following research questions guided this project:

1. How well does a Smoke Impact Score correlate with PM2.5 air quality measurements and respiratory health outcomes?
2. What is the relationship between wildfire size, distance, and health outcomes such as asthma hospitalizations?
3. Can a simplified statistical model effectively explain observed health outcomes, or are more complex machine learning models required?

**Model Selection**

During the development of the Smoke Impact Score, several modeling approaches were considered, including traditional distance-based models and more complex machine learning algorithms. However, the analysis revealed no strong inherent correlation between the Smoke Impact Score and asthma hospitalizations, likely due to confounding factors like healthcare access, socioeconomic conditions, and external interventions.

Given these limitations, a **simple linear regression model** was selected for its interpretability and robustness to overfitting, especially with datasets showing limited direct correlation. More complex models, such as random forests or neural networks, were deemed unsuitable because they would likely overfit the data without improving predictive power. The linear regression model provided a baseline understanding of the relationship between the Smoke Impact Score and asthma hospitalizations, balancing simplicity with explanatory utility.

# Methodology

**1. Smoke Impact Score Development**

The Smoke Impact Score was designed as a composite metric to quantify the relationship between wildfire events and their impact on air quality and health outcomes. This score incorporates two primary variables: wildfire size and distance from the affected population.

* **Custom Scoring Function**: A custom scoring function was developed to account for the non-linear relationship between wildfire characteristics and their impacts.
* **Cube Root Normalization:** Wildfire size (measured in acres) was normalized using a cube root transformation to reduce the influence of extremely large fires, which could disproportionately skew the score.
* **Linear Distance Penalty:** A linear distance penalty was applied for distances up to 650 miles, reflecting the observed dispersion patterns of wildfire smoke. This cutoff was determined based on prior research showing that smoke impacts decline sharply beyond this range.

**Why These Methods Were Chosen:**

* The cube root transformation ensures that the Smoke Impact Score reflects both the size of the wildfire and its relative significance without being overwhelmed by outliers.
* The linear distance penalty aligns with real-world observations and improves the interpretability of the model, making it suitable for public health applications.

**2. Health Data Analysis**

To evaluate the relationship between the Smoke Impact Score and respiratory health outcomes, the analysis focused on hospital admissions for asthma-related cases.

* **Focus on Asthma**: Asthma cases were chosen as the primary health outcome due to their sensitivity to particulate matter (PM2.5) and their clear link to air quality changes.
* **Aggregated and De-identified Data:** To ensure privacy and compliance with ethical standards, the study utilized aggregated, de-identified hospital records obtained from the Arizona Department of Health Services.
* **Multi-Year Analysis:** Trends were analyzed across multiple years to account for temporal variations in wildfire activity, air quality, and healthcare access.

**Why These Methods Were Chosen:**

* Asthma-related hospitalizations are a well-documented indicator of air quality impacts, making them a reliable metric for assessing the health effects of wildfire smoke.
* De-identification of health data ensures compliance with ethical guidelines while allowing for meaningful analysis of public health trends.

**3. Ethical Considerations**

Human-centered considerations were integral to the study's design, ensuring that the analysis prioritized public health and ethical transparency.

* **Public Health Focus:** The Smoke Impact Score and health analysis were designed to provide actionable insights for public health interventions, such as early warning systems and targeted advisories for vulnerable populations.
* **Data Privacy:** To protect individual privacy, all health data were aggregated and de-identified. The study adhered to relevant data protection standards, ensuring that no personally identifiable information was used.
* **Transparency and Reproducibility:** The methodology was documented in detail to enable reproducibility and transparency. Acknowledging the limitations of the model ensures that stakeholders understand its scope and potential biases.

**Why These Ethical Considerations Were Chosen:**

* Prioritizing public health ensures that the findings are actionable and directly benefit the affected population.
* Transparency and acknowledgment of limitations build trust and ensure that the model's insights are used appropriately.

**Human-Centered Approach**

Throughout the study, the methodology was informed by a human-centered approach that emphasized ethics, privacy, and the practical application of findings. By focusing on actionable insights and ethical considerations, the study aimed to balance scientific rigor with real-world relevance, providing a foundation for public health planning and wildfire management.

# Findings

**1. Increasing Wildfire Intensity**

* Total acres burned by wildfires have increased significantly since the 1990s, with peaks observed in 2006 (3.8M acres) and 2020 (4.0M acres).
* Larger wildfires produce more severe smoke impacts, correlating with a rise in Smoke Impact Scores over the same period.

चित्र में निम्न शामिल है टेक्स्ट, प्लॉट, लाइन, आरेख

Description automatically generated

**2. Smoke Impact Score and Air Quality**

* Higher Smoke Impact Scores generally correspond to elevated AQI values, although the relationship is not linear.
* The most significant smoke impacts occurred within 650 miles of Glendale, Arizona, with peak frequency observed around 900 miles.

चित्र में निम्न शामिल है टेक्स्ट, लाइन, आरेख, प्लॉट

Description automatically generated

**3. Asthma Hospitalization Trends**

* A notable inverse trend was observed between asthma hospitalizations and the Smoke Impact Score from 2005 to 2020:
  + Asthma-related hospitalizations declined significantly after 2015, even as Smoke Impact Scores rose.
  + The spike in Smoke Impact Scores in 2020 coincided with a sharp drop in hospitalizations, potentially influenced by pandemic-related healthcare access changes or successful mitigation efforts.

चित्र में निम्न शामिल है टेक्स्ट, लाइन, आरेख, प्लॉट

Description automatically generated

# Discussion/Implications

**1. Public Health Impacts**

* The inverse relationship between asthma hospitalizations and Smoke Impact Scores highlights the potential effectiveness of mitigation measures, such as:
  + Air quality advisories and increased use of air filtration systems.
  + Public awareness campaigns targeting vulnerable populations.

**2. Response to Increasing Wildfire Intensity**

* The rise in wildfire activity and size demands immediate and long-term public health interventions:
  + **Immediate Actions:**
    - Development of an early warning system based on Smoke Impact Scores.
    - Allocation of healthcare resources during high-risk wildfire seasons.
  + **Long-term Strategies:**
    - Enhanced wildfire risk assessment and air quality modeling.
    - Investment in forest management and climate change mitigation measures.

**3. Implications of 2020 Anomalies**

* The decline in hospitalizations in 2020 may reflect reduced healthcare access during the COVID-19 pandemic rather than a reduction in actual asthma exacerbations caused by wildfire smoke.
* Future studies should incorporate behavioral and healthcare access data to refine these findings.

# Limitations

1. **Data Constraints:**
   * Lack of detailed atmospheric data (e.g., wind direction and speed) in the Smoke Impact Score model.
   * Gaps in historical wildfire and AQI data.
2. **Modeling Simplifications:**
   * Linear distance relationships may oversimplify smoke dispersion.
   * Lack of validation against real-time smoke plume data.
3. **Scope Limitations:**
   * Focus on a single metropolitan area may limit generalizability.
   * Inability to fully account for all confounding variables, such as preexisting health conditions or socioeconomic factors.

# Conclusion

**Restating Research Questions and Hypotheses**

This study set out to explore the following key research questions:

1. How well does the Smoke Impact Score correlate with PM2.5 air quality measurements and respiratory health outcomes?
2. What is the relationship between wildfire size, proximity, and health outcomes, particularly asthma-related hospitalizations?
3. Can a simplified statistical model effectively capture the observed relationship, or are more complex models required?

The primary hypothesis was that wildfire size and proximity directly impact respiratory health outcomes, but traditional distance-based models are insufficient for accurately capturing smoke impacts. A modified Smoke Impact Score, incorporating fire size and linear distance penalties, was proposed as a solution.

**Summary of Findings**

The study revealed several key insights:

1. **Smoke Impact Score Effectiveness**: The Smoke Impact Score captured general trends in air quality degradation and its correlation with respiratory health outcomes. However, the relationship between the score and asthma hospitalizations was not strong, suggesting the presence of confounding factors such as healthcare access, socioeconomic disparities, and public health interventions.
2. **Wildfire Trends**: The increasing scale and intensity of wildfires over recent decades are clear contributors to air quality degradation. Larger fires, combined with their proximity to urban centers, represent significant public health risks.
3. **Asthma Hospitalizations**: Despite rising Smoke Impact Scores, asthma hospitalizations showed a declining trend in recent years. This may be attributed to effective public health measures, increased awareness, and protective behaviors, as well as potential changes in healthcare-seeking patterns during events like the COVID-19 pandemic.

**Human-Centered Data Science Implications**

This study exemplifies the principles of human-centered data science by focusing on the intersection of environmental events and public health. Key aspects include:

* **Ethical Design**: By ensuring data privacy and prioritizing the needs of vulnerable populations, the analysis adhered to ethical standards that safeguard individuals while delivering actionable insights.
* **Transparency and Practicality**: The simple regression model and the detailed explanation of its limitations make the findings accessible and usable for public health planning.
* **Focus on Actionable Insights**: The findings inform strategies for early warning systems, health advisories, and resource allocation, providing direct value to policymakers, healthcare providers, and community leaders.

**Final Remarks**

The Smoke Impact Score provides a novel framework for quantifying the relationship between wildfire smoke and respiratory health outcomes. While the findings emphasize the importance of proactive public health measures, they also highlight the need for further refinement of the model to incorporate additional environmental and behavioral variables. This study advances our understanding of human-centered data science by demonstrating how data-driven insights can directly benefit public health and guide strategies for mitigating the growing impact of wildfires on urban populations. Continued vigilance, interdisciplinary collaboration, and iterative refinement of methodologies are essential to address these complex challenges effectively.

# References

* [CDC National Environmental Public Health Tracking Network](https://ephtracking.cdc.gov)
* [Arizona Department of Health Services - Public Health Data](https://pub.azdhs.gov/health-stats/)
* [National Interagency Fire Center - Historical Wildfire Data](https://www.nifc.gov/fire-information/statistics/wildfires)
* [Environmental Protection Agency - Air Quality System Data](https://www.epa.gov/aqs)

# Data Sources

* Wildfire Perimeter Data: ESRI:102008 coordinate system
* [Air Quality Data (EPA AQI Database)](https://www.epa.gov/aqs)
* [Health Records (Arizona Department of Health Services, aggregated, de-identified)](https://data.azdhs.gov/reports-and-catalogs/home)
* [Asthma hospitalizations](https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm)