

IMPACT OF WILDFIRES ON DEATHS DUE TO RESPIRATORY ISSUES

MOHAMMAD DANISH NADEEM

COURSE: DATA 516 HUMAN CENTERED DATA SCIENCE

FINAL SUBMISSION

Table Of Contents

- 1. Introduction**
- 2. Methodology**
 - a. Data Sources**
 - b. Analysis**
- 3. Findings**
- 4. Discussion and Recommendations**
- 5. Limitations and Considerations**
- 6. References**
- 7. Data Sources**

Section 1 Introduction: *Understanding the Impact of Wildfires on Health in Ector County*

In recent years, the surge in the frequency and intensity of wildfires has emerged as a pressing global concern, contributing to a range of environmental, economic, and health challenges. The United States, in particular, has experienced a notable escalation in wildfire occurrences, bringing about diverse health implications for communities exposed to these environmental hazards.

This analysis seeks to delve into the intricate relationship between the increasing incidence of wildfires and its potential health repercussions, focusing specifically on chronic respiratory diseases and chronic obstructive pulmonary diseases (COPD). The motivation behind this project stems from the urgent need to comprehend the health consequences of escalating wildfires, given their multifaceted impact on public well-being.

Within the confines of the United States, the state of Texas, and more specifically Ector County, has not been immune to the growing threat of wildfires. This study zeroes in on Ector County, an area encompassing West Odessa, situated within 1250 miles of the region. By narrowing our focus to this locality, we aim to provide a comprehensive understanding of the relationship between wildfires and health outcomes within the specific context of Ector County.

West Odessa, a significant part of Ector County, is characterized by its unique weather patterns, which play a crucial role in influencing the dynamics of wildfires. A succinct exploration of West Odessa's climate will be undertaken to elucidate the environmental conditions that contribute to the prevalence of wildfires in the region. As we delve into this analysis, our goal is to contribute valuable insights that can inform public health strategies, thereby mitigating the adverse health effects associated with the increasing wildfire occurrences in Ector County.

Section 2: *Methodology*

A: *Data Sources & Data Profile*

The comprehensive analysis conducted in this study relies on data procured from three primary sources, each contributing a crucial aspect to the overarching investigation.

1. Wildfire Data (1963-2023):

The foundational wildfire data utilized in this study originates from www.sciencebase.gov, covering the extensive timeframe from 1963 to 2023. This dataset serves as the cornerstone for evaluating the temporal patterns, frequencies, and geographic distributions of wildfires within the specified period.

Data Description:

This dataset was initially complex, containing nearly 30 columns, all of which are documented in this [notebook](#). Following preprocessing, only the pertinent columns were retained, specifically:

1. Area burned in Acres
2. Distance (representing the distance of the fire from West Odessa)

2. Air Quality Index (AQI) Data (1973-2020):

The evaluation of air quality, a pivotal component of this analysis, draws upon data sourced from the US Environmental Protection Agency's (EPA) website at www.epa.gov. The AQI data spans the years 1973 to 2020, offering insights into the variations in air quality that coincide with wildfire events. This information aids in understanding the potential correlation between wildfire occurrences and air quality fluctuations.

Data Description:

In the case of Ector County, data for the period 1973-2020 was available for only one of the following parameters in any given year. Additionally, the data was accessible on a daily basis, prompting the calculation of yearly averages for the sake of simplicity. To address potential gaps in data for any specific Air Quality Index (AQI) type, a comprehensive dataset was constructed by merging the available information for the entire period from 1973 to 2020.

3. Healthcare Data (1980-2024):

The health-related dimension of the analysis relies on data obtained from the Institute for Health Metrics and Evaluation (IHME) website, accessible at ghdx.healthdata.org. Covering the period from 1980 to 2024, this dataset contributes critical information regarding the prevalence and trends of chronic respiratory diseases and chronic obstructive pulmonary diseases (COPD). The integration of healthcare data allows for a nuanced examination of the potential health impacts resulting from increased wildfire activity.

Data Description:

This dataset covers the state of Texas from 1980 to 2014 and includes a comprehensive set of columns such as measure ID, measure name, location ID, location name, FIPS code, cause ID, cause name, sex ID, sex, age ID, age name, year ID, metric, and upper/lower bounds. The dataset offers a broader context for the region, facilitating comparisons and contextualizing the impact of wildfires on healthcare in West Odessa within the larger state scenario.

License/Terms of Use: Data made available for download on IHME Websites can be used, shared, modified or built upon by non-commercial users in accordance with the [IHME FREE-OF-CHARGE NON-COMMERCIAL USER AGREEMENT](#). For more information (and inquiries about commercial use), visit IHME [Terms and Conditions](#).

Sr.No	Column Name	Description
1	Measure_id	Unique Numeric Identifier for the measure in an IHME database of data dimensions
2	Measure_name	The indicator of the estimate
3	Location_id	Unique Numeric Identifier for the location and stored in an IHME database of data dimensions
4	Location_name	Location Name for which the data is represented
5	FIPS	FIPS (Federal information processing standards code)
6	Cause_id	Unique Numeric Identifier for the cause of injury and stored in an IHME database of data dimensions
7	Cause_name	Reason of disease
8	Sex_id	Numeric representation of gender
9	Sex_name	Gender
10	Age_id	Unique Numeric Identifier for the age group generated and stored in an IHME database of data dimensions
11	Age_name	The group to which the age is assigned
12	Year_id	Year of the estimate
13	Metric	Unit of measure for Mx, Lower, and Upper
14	Mx	Posterior mean estimate
15	Lower	2.5 Percentile
16	Upper	97.5 Percentile

These three primary data sources collectively form the backbone of the study, enabling a holistic exploration of the interconnections between wildfires, air quality, and public health outcomes. The meticulous curation of data from reputable platforms ensures the reliability and relevance of the findings presented in this analysis.

B: Analysis

After obtaining and cleaning the data, the initial phase of the analysis involved the calculation of a crucial metric known as the "wildfire score." This score was derived using the formula:

$$\text{Wildfire Score} = 2 \times \text{Area Burned in Acres} / \text{Distance}$$

The rationale behind this calculation lies in the intuitive understanding that the impact of a wildfire is directly proportional to the area it engulfs, and inversely related to the distance from the source. This approach aligns with a human-centric perspective, acknowledging that the farther the fire, the lesser its potential impact.

Subsequently, the analysis delved into the correlation between the wildfire score and the Air Quality Index (AQI) over the years. This step aimed to validate the developed score by establishing its relationship with a recognized environmental indicator, ensuring that it aligns with broader trends in air quality and wildfire impact.

Having established a correlation, the analysis progressed to predicting the wildfire score for the subsequent 25 years. The predictive model, applied from 2020 onwards, utilized an ARIMA (AutoRegressive Integrated Moving Average) model. This forecasting methodology allowed for a projection of the wildfire score, offering valuable insights into the potential trends and patterns in wildfire impact in the forthcoming years.

Building upon the initial methodology, the analysis extended to incorporate healthcare data, focusing specifically on Chronic Respiratory Diseases (CRD) and Chronic Obstructive Pulmonary Diseases (COPD). This phase involved a nuanced examination of mortality trends for both genders individually and combined.

Utilizing the healthcare data, time series plots were generated to meticulously analyze the progression of deaths attributed to CRD and COPD over the years. The temporal patterns unveiled by these plots provided valuable insights into the evolving health landscape and trends in respiratory-related mortalities.

Further, percentage changes in mortality rates were calculated for males, females, and the combined dataset. This quantitative approach allowed for a comprehensive understanding of the relative shifts in mortality trends, enabling a more nuanced interpretation of gender-specific impacts.

To discern potential correlations between deaths caused by CRD and COPD, heat maps were generated. These visual representations facilitated the identification of patterns and relationships between these respiratory conditions, contributing to a more holistic comprehension of the intricate dynamics at play.

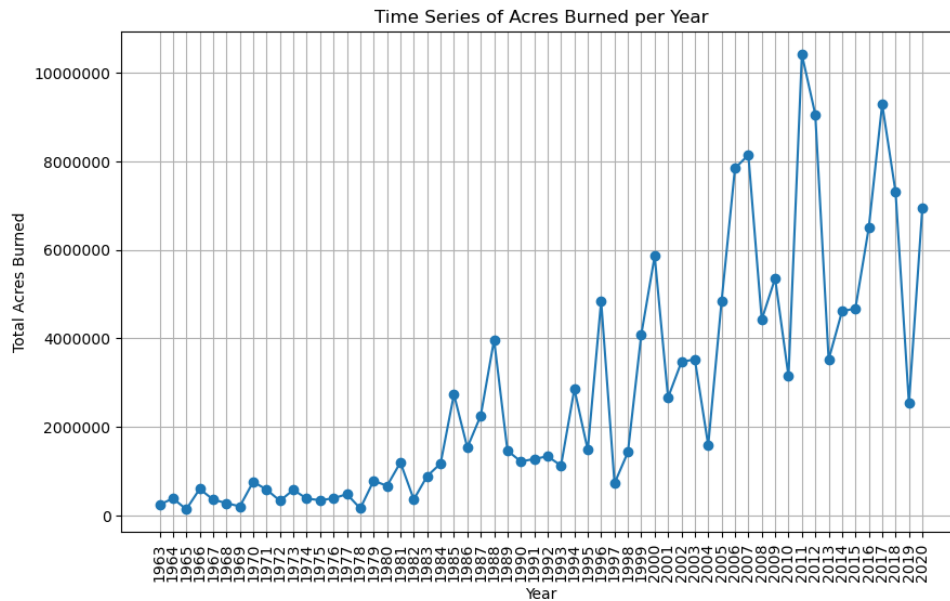
Building upon the predictive modeling approach employed for the wildfire score, ARIMA forecasting was extended to project deaths due to CRD. This temporal forecasting model offered a glimpse into future mortality trends, aiding in the anticipation of potential challenges and the formulation of targeted healthcare strategies.

In essence, the extended methodology encompassed a comprehensive analysis of healthcare data, employing diverse statistical techniques to unveil patterns, correlations, and forecast future trends in mortality rates related to Chronic Respiratory Diseases. This multifaceted approach contributes to a more holistic understanding of the interplay between environmental events, health outcomes, and the implications for public health management.

Section 3: *Findings*

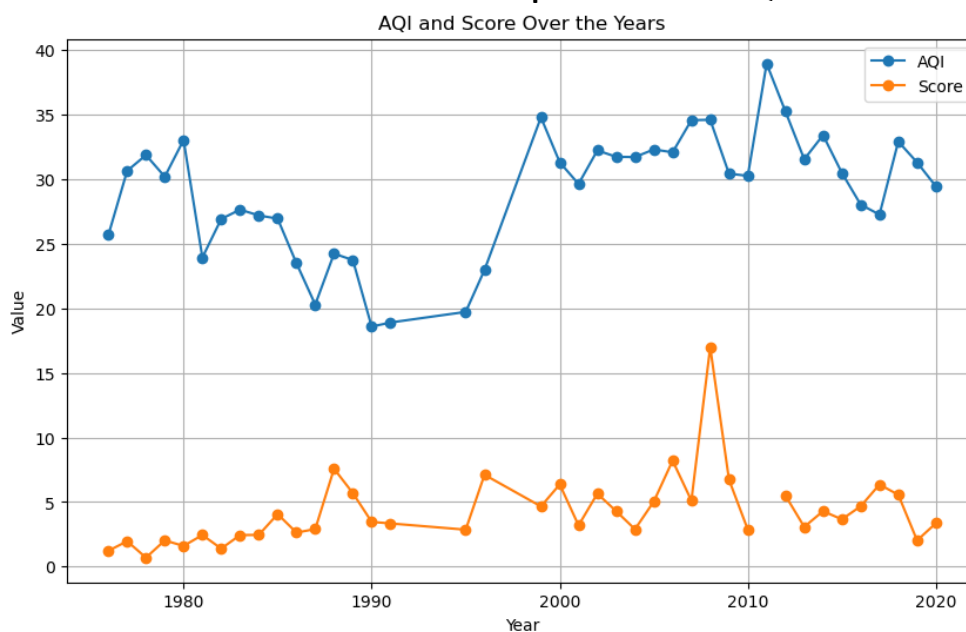
In this section, I will present key findings from my analysis, highlighting significant insights that underscore the complex interplay between wildfires, air quality, and environmental impact. These findings aim to provide a comprehensive understanding of the consequences of wildfires, serving as valuable inputs for informed decision-making in both environmental management and public health initiatives.

1. Acres burned has increased over the year



The graph is a time series representation that provides an insightful look into the annual dynamics of fires occurring within a 1250-mile radius of West Odessa. It focuses on the total acres burned each year, making it a valuable tool for understanding the patterns and trends in fire activity in this particular geographic region. By plotting time (years) on the x-axis and the total acres burned on the y-axis, this graph allows us to track the variations in fire intensity over time. This information can be invaluable for assessing the impact of fires on the local environment, as well as for developing strategies to manage and mitigate fire-related risks. The geographical criterion of 1250 miles from West Odessa narrows the scope of the data, making this graph a valuable resource for understanding the dynamics of fires in this specific area over the years.

2. Positive correlation between wildfire impact score and AQI 0.29

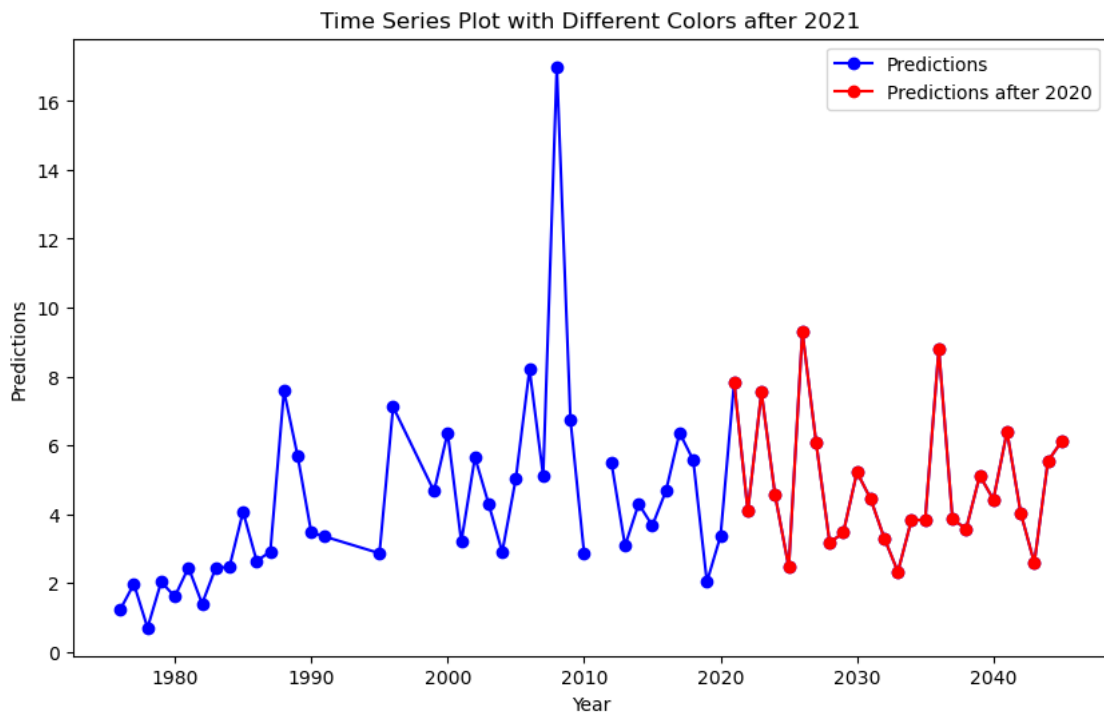


The time series graph provides a comprehensive overview of air quality conditions in West Odessa, integrating two significant components: a customized fire smoke score and the official Air Quality Index (AQI) retrieved from authoritative government resources. The custom-created fire smoke score is a unique estimate developed from specific data sources and methodology, serving as an indicator of localized fire-related air quality. In contrast, the AQI is derived from official government data, encompassing a broader range of pollutants and environmental factors.

By plotting these two distinct metrics over time, the graph allows for a comparative analysis of the custom fire smoke score and the standardized AQI. This visualization offers valuable insights into how your localized fire smoke estimates align with the broader government-recognized air quality assessments. It becomes a valuable resource for discerning trends and discrepancies, thereby enhancing our understanding of air quality dynamics in West Odessa and the specific impact of fire events on local air quality.

Note that the **correlation** between the Score and the AQI is **0.2325**.

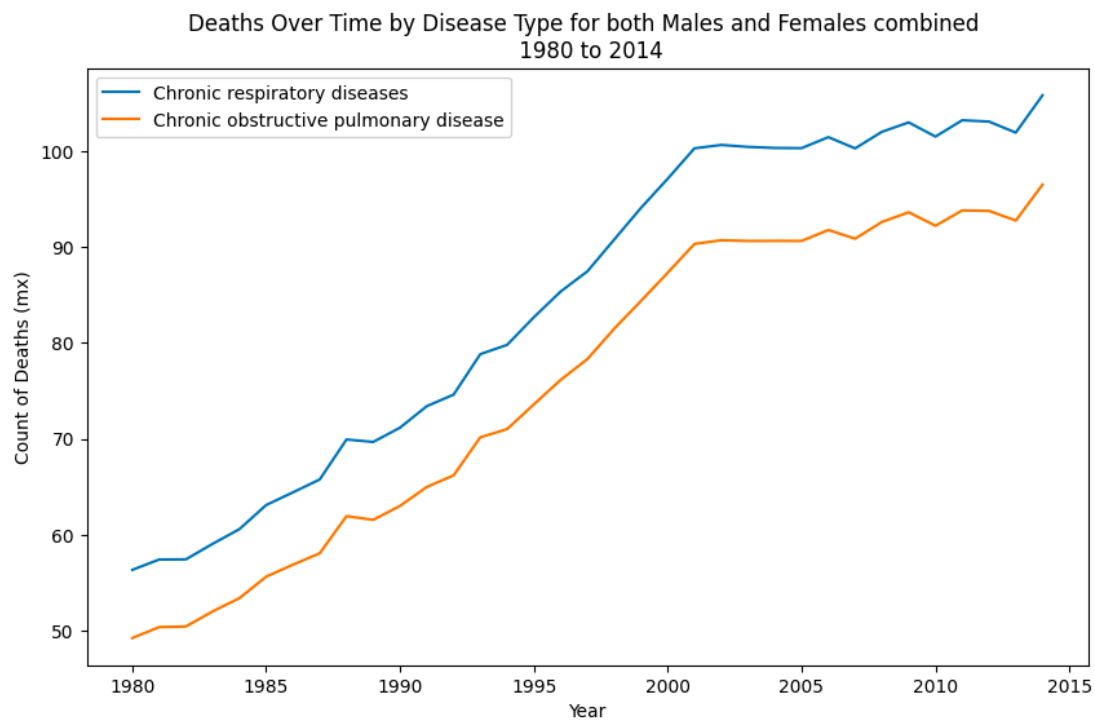
3. Increasing trend in Wildfire Impact Score



Over here, I conducted predictions using ARIMA for the period beyond 2020, extending to the next 25 years. The y-axis signifies the impact score, while the x-axis delineates the respective years. ARIMA, renowned for its efficacy in time series forecasting, serves as a robust tool for capturing temporal patterns and projecting future trends. Notably, the graphical representation reveals an anticipated and sustained increase in the wildfire score post-2020, persisting for the subsequent two decades. This ascending trend signifies a foreseen intensification in the negative impacts of wildfires. The discerned trajectory underscores the importance of proactive measures and adaptive strategies to

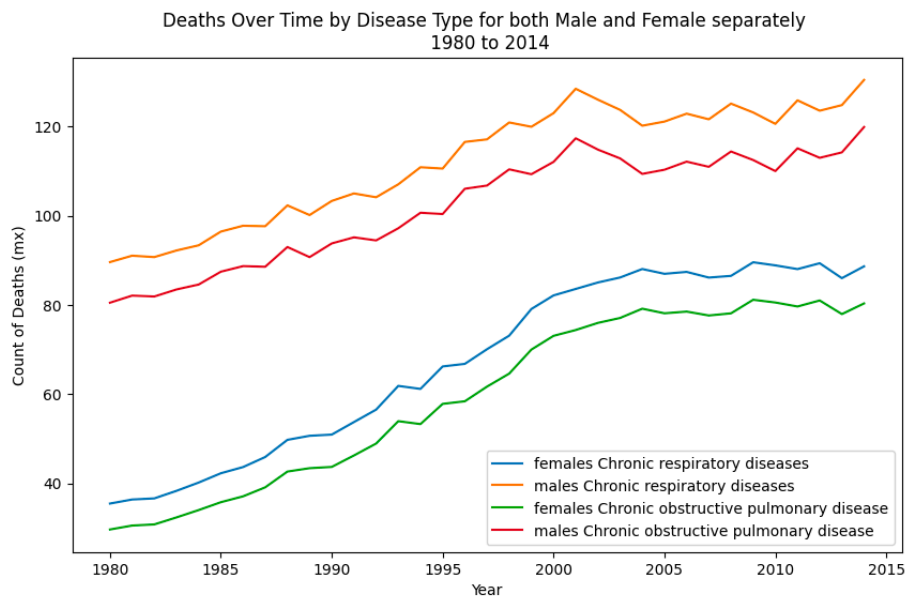
address the potential escalation of consequences posed by wildfires on both the environment and public health.

4. Increase for the number of deaths for both males and females from chronic respiratory disease and chronic obstructive pulmonary disease



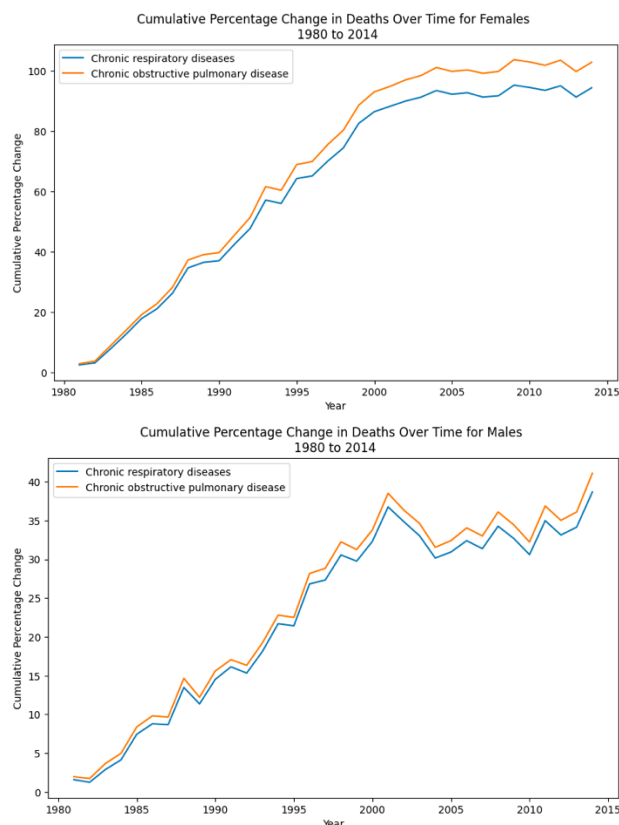
In this time series graph, the y-axis represents the count of deaths attributed to chronic respiratory diseases and chronic obstructive pulmonary disease (COPD), while the x-axis denotes the corresponding years. The visualization distinctly illustrates an upward trend in the number of deaths related to both chronic respiratory diseases and COPD from 1980 to 2014. Notably, a pronounced and steep incline is observed in the period spanning from 1980 to 2000, suggesting a substantial rise in mortality rates during this specific timeframe. These findings underscore the critical nature of respiratory health concerns and emphasize the need for targeted interventions and healthcare strategies to address the escalating impact of chronic respiratory diseases and COPD.

5. Independently for males and females



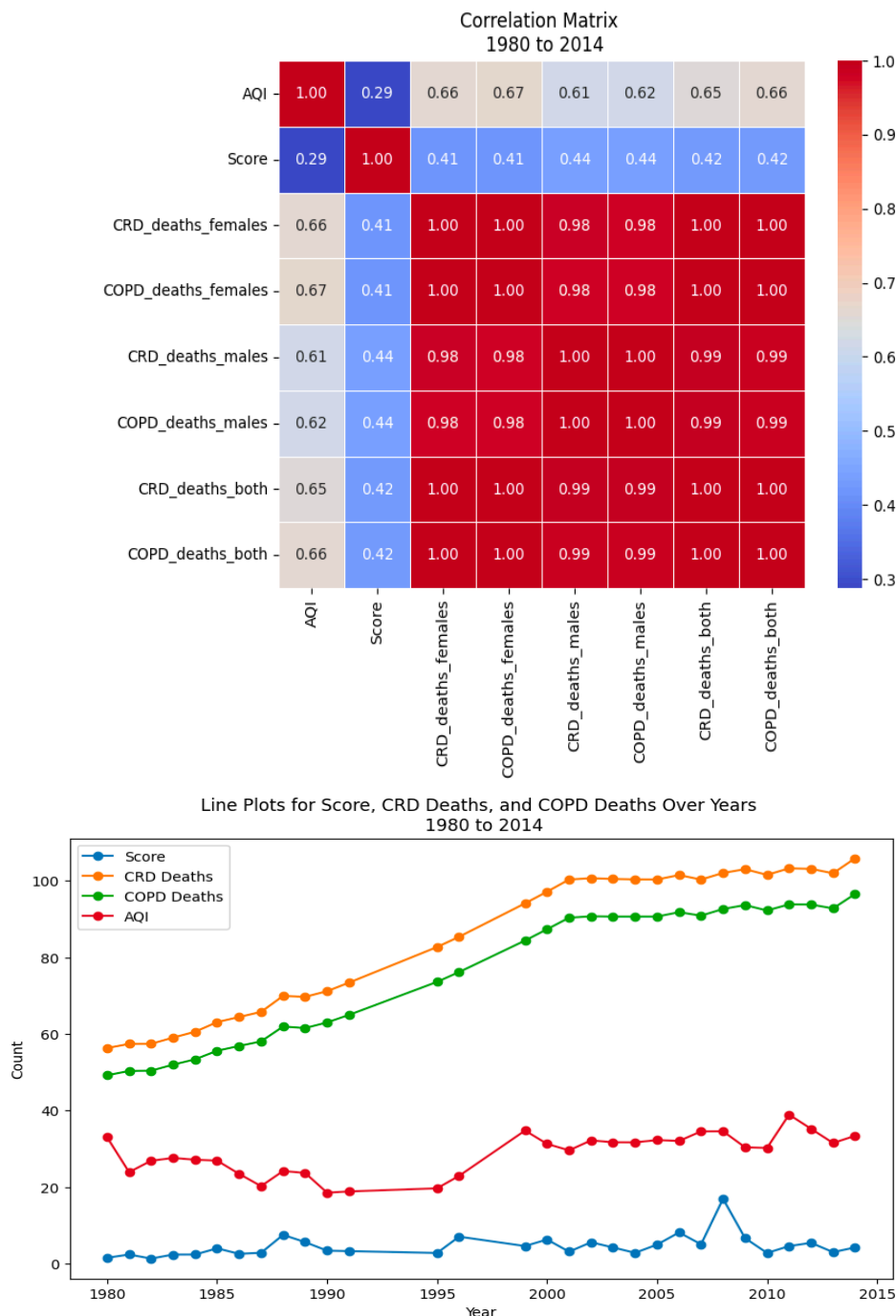
This is an extension of the above findings where we can see the count of deaths individually for males and females over the years. We can observe that more deaths on males have occurred as compared to females for both the diseases. However, the number of deaths have sharply risen for both.

6. In percentage terms female deaths have doubled whereas has had lesser impacts on male deaths



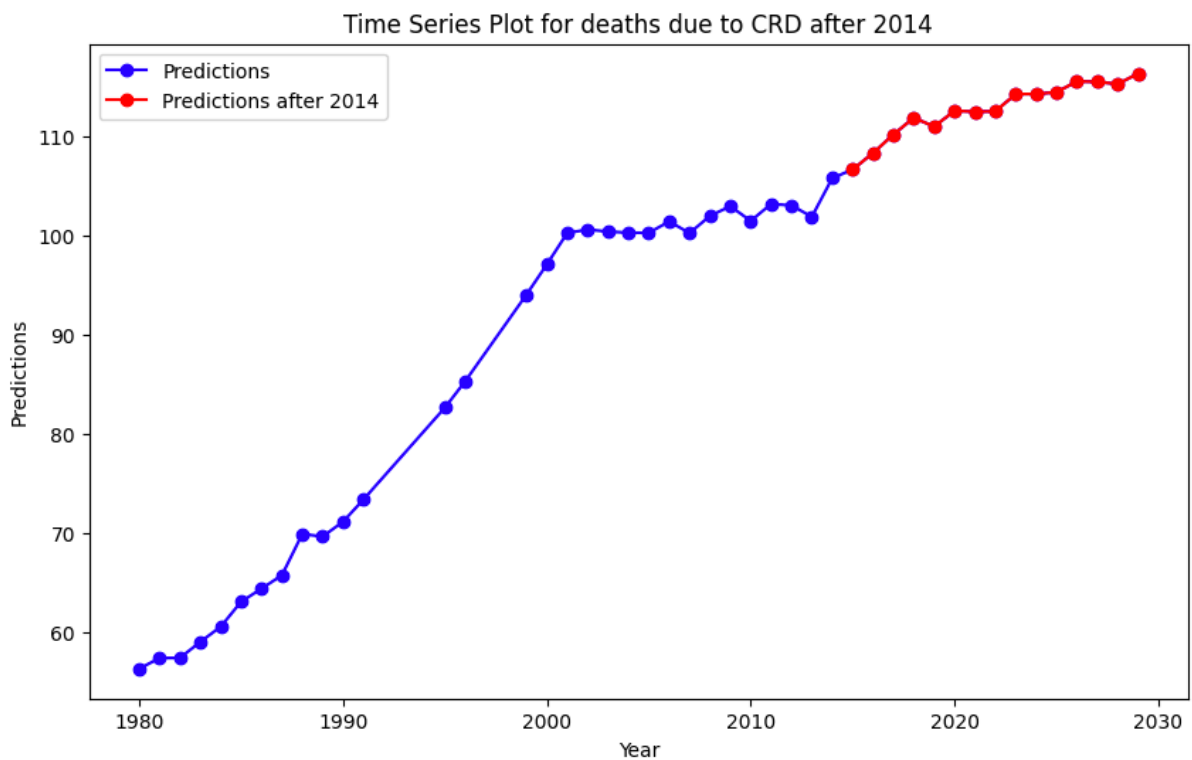
This comparative graph illustrates the percentage change in deaths between males and females, revealing a surprising outcome: the number of deaths in females has doubled. Specifically, there is a notable 100% increase in female deaths, contrasting with a comparatively lower 50% rise in male deaths. This unexpected discrepancy prompts the proposition that wildfires have exerted a more pronounced impact on the mortality of women attributed to respiratory diseases. The discerned gender-specific variation underscores the importance of delving deeper into the nuanced effects of wildfires on public health, necessitating targeted interventions and strategies to mitigate the disparate impacts observed across different demographic groups.

7. Positive relation between deaths due to respiratory issues, AQI and Impact Score



In both the heatmap and the time series plot, a discernible positive correlation emerges between deaths attributed to respiratory issues and both the Air Quality Index (AQI) and the calculated impact score. The correlation coefficients for deaths due to respiratory issues among males, females, and the combined genders exceed 0.4, indicating a substantial positive relationship. The graphical representation further substantiates this correlation, notably during the period from 1990 to 2000. In this timeframe, a pronounced increase is evident in both AQI levels and deaths due to respiratory issues, emphasizing a simultaneous escalation in environmental pollution and respiratory health concerns. These observations underscore the interconnectedness of air quality, wildfire impact, and respiratory health outcomes, calling for targeted interventions to address the potential health ramifications.

8. Increasing deaths forecast for future



I conducted a time series prediction using ARIMA for deaths attributed to chronic respiratory diseases. Notably, this analysis excludes COPD, as both chronic respiratory diseases and COPD exhibit a robust correlation, making predictions for one condition informative for the other. This approach enhances user comprehension. The prediction reveals a continuous increase in deaths due to chronic respiratory issues from 2014 to 2030. Aligning this observation with the earlier wildfire impact score prediction, which similarly indicated an escalating trend, suggests a robust positive relationship between the two. This assertion implies a critical need for immediate interventions to curtail the rising impact of wildfires and mitigate their associated health consequences, particularly in the context of chronic respiratory diseases.

Section 4: Discussion and Recommendations

The findings of this study hold crucial importance due to their direct implications for public health and environmental management. The observed positive correlation between wildfire impact, air quality, and respiratory health outcomes underscores the urgent need for informed decision-making at both the administrative and community levels. As we navigate the complex web of factors contributing to heightened mortality rates related to chronic respiratory diseases, the following considerations emerge:

City Council and Administrators:

In light of Ector County's arid climate, it is recommended that the city council explores innovative solutions to mitigate the risk of wildfires. Implementing artificial rain initiatives during the dry fire-prone months can be a proactive measure to dampen the environment and reduce the likelihood of wildfires spreading. Furthermore, initiating a "Go Green" campaign can be pivotal in promoting sustainable practices and encouraging residents to adopt low-carbon-impact products. This multifaceted approach not only addresses the immediate risk of wildfires but also contributes to long-term environmental sustainability.

City Residents:

The findings emphasize the role of individuals in safeguarding their health during wildfire events. City residents should stay informed about air quality conditions, follow prescribed safety measures, and be proactive in adopting measures to reduce their vulnerability to respiratory issues. Public awareness campaigns, community engagement initiatives, and the provision of resources for vulnerable populations are essential components of a comprehensive strategy.

Timeline for Action:

Given the escalating trends identified in our predictive modeling, there is an immediate need for action. A concrete plan should be formulated and implemented within the next few years to address the foreseeable risks associated with increasing wildfire impact. Swift action will help minimize the long-term health consequences and environmental degradation resulting from these events.

Human-Centered Data Science Principles:

Human-centered data science principles played a pivotal role in shaping the decision-making process throughout this project. The interdisciplinary nature of the analysis, incorporating environmental, health, and demographic dimensions, aligns with the principles of inclusivity and relevance to diverse stakeholders. The emphasis on data-driven insights ensures that policies and interventions are grounded in empirical evidence, enhancing their efficacy and adaptability.

In conclusion, the implications of our findings underscore the immediate need for collaborative, data-informed strategies to address the rising impact of wildfires on public health. The integration of human-centered data science principles ensures that the decision-making process prioritizes the well-being of the community and remains responsive to the dynamic challenges posed by environmental events.

Section 5: Limitations and Considerations

1. ARIMA Model Constraints:

While ARIMA is a powerful time series forecasting model, it has its limitations. One significant constraint is the need for more granular data for enhanced prediction accuracy. Incorporating monthly or weekly data could potentially improve predictions. Additionally, forecasting over an extended period, such as 25 years, may introduce uncertainties and reduce accuracy. Exploring alternative models like Long Short-Term Memory networks (LSTMs) within the realm of deep learning might be advantageous if additional data becomes available.

2. Wildfire Impact Score Accuracy:

The accuracy of the wildfire impact score could be further refined given more time and comprehensive research. Incorporating additional determining factors and exploring diverse datasets for extracting correlations would contribute to a more precise analysis. The limitations of available time and data sources might have influenced the comprehensiveness of the wildfire impact score calculation.

3. Unexplored Contributing Factors:

Certain significant contributing factors such as global warming, dietary habits, and cultural practices like vaping or smoking were not explicitly considered in our analysis. These elements could potentially have a substantial impact on respiratory health and might contribute to higher mortality rates. Future research could benefit from a more expansive exploration of these external factors.

4. Correlation vs. Causation:

Despite the depth of our analysis and the strength of the results, it is essential to acknowledge the inherent challenge of establishing causation. While correlations between variables have been identified, causal relationships cannot be definitively ascertained. This limitation underscores the importance of interpreting findings with caution and the need for continued research to uncover causal links.

In conclusion, this study, like any research endeavor, is subject to limitations. Prioritizing and acknowledging these limitations is critical for a transparent and comprehensive interpretation of the results.

Section 6: Conclusion

In this study, we embarked on an exploration of the intricate relationships between wildfires, air quality, and respiratory health, employing a human-centered data science approach. The research questions aimed to unravel the impacts of wildfires on chronic respiratory diseases, utilizing a multidimensional analysis that incorporated environmental factors, health outcomes, and predictive modeling.

Our investigation commenced with the development of a wildfire impact score, integrating area burned and proximity metrics. The ARIMA model was then employed to forecast the escalating trends in the wildfire impact score, providing insights into future implications. The time series analyses unveiled an alarming surge in deaths related to chronic respiratory diseases, particularly among females, accentuating the urgent need for targeted interventions.

Crucially, our findings revealed a positive correlation between wildfire impact, air quality, and respiratory health outcomes. The nuanced analysis brought attention to gender-specific disparities in mortality rates, emphasizing the importance of considering diverse demographic groups in public health strategies.

However, the study is not without limitations. Factors such as data constraints, model assumptions, and unexplored external influences underscore the complexity of interpreting results. The challenge of establishing causation, despite robust correlations, necessitates cautious interpretation.

This study contributes to the understanding of human-centered data science by exemplifying the intricate interplay between environmental events, health outcomes, and predictive modeling. The multidimensional analysis showcases the relevance of data-driven insights in shaping proactive public health strategies. As we confront the escalating challenges posed by wildfires and their implications on human health, this study underscores the significance of data science in informing policies and interventions for a resilient and adaptive society.

Section 7: References

Jaffe, D. A., O'Neill, S. M., Larkin, N. K., Holder, A. L., Peterson, D. L., Halofsky, J. E., & Rappold, A. G. (2020). Wildfire and prescribed burning impacts on air quality in the United States. Journal of the Air & Waste Management Association (1995), 70(6), 583–615. <https://doi.org/10.1080/10962247.2020.1749731>

Section 8: Data Sources

1. www.sciencebase.gov (Wildfire Data)
2. www.epa.gov (AQI Data)
3. ghdx.healthdata.org (Health Care Data)
4. <https://github.com/Mdanishnadeem/Wildfire-Analysis> (Git hub Repo)