DATA 512 FINAL PROJECT
Hema Varshita Muppalla
City - Williston, ND

INTRODUCTION

The impact of the escalating wildfires in Williston, North Dakota, extends far beyond the immediate threat of flames. These fires, influenced by changing climate patterns and forestry policies, cast a shadow over the city's well-being. The smoke they produce doesn't just pose a hazard to respiratory health; it disrupts the very fabric of daily life, jeopardizing the stability of vital economic sectors, influencing industries, livelihoods, and overall financial stability. Understanding the economic implications is crucial in comprehending the comprehensive effects of these wildfires.

Understanding the intricacies of this situation goes beyond acknowledging the visible effects. It demands a deep dive into the data, comprehending how this smoke affects health, economy, and societal infrastructure. This analysis becomes the cornerstone for drafting informed policies and laying out healthcare plans that shield residents and safeguard the economic heartbeat of Williston.

However, the approach can't be one-size-fits-all. It requires tailored strategies that factor in the unique vulnerabilities of the community. For instance, recognizing the differing impact of wildfire smoke on men and women's health is crucial. Addressing these gender-based disparities isn't just an option; it's a necessity for effective mitigation and health management.

The significance of this comprehensive analysis doesn't just stop at benefiting the residents of Williston. It serves as a beacon guiding policymakers and civic bodies, offering a roadmap towards reforms prioritizing human well-being. It's a tool for shaping a future where the impact of these wildfires is minimized, ensuring the resilience of the community and its economic stability in the face of such environmental challenges.

BACKGROUND

In the context of Williston, North Dakota, there's indeed a notable gap in specific research focusing on the correlation between smoke estimates and economic sectors. While broader studies exist, examining the economic impact of wildfires and smoke on regional economies, there's a lack of direct investigation into how smoke precisely affects economic sectors in Williston. The absence of direct studies linking smoke estimates with economic sectors in Williston highlights a significant opportunity for research and analysis. It signifies the need for tailored investigations that specifically assess how smoke, stemming from wildfires, directly impacts sectors integral to the region's economy.

Analysis:

1. Correlation Factors: Your analysis on the correlation between smoke estimates and household income, as well as unemployment rates, in Williston, provides valuable initial insights. These findings establish a foundation for understanding the potential economic impact of smoke in the region.

2. Forecasting Using Prophet: Employing Prophet to project smoke estimates, household income, and unemployment rates for the next 25 years offers a forward-looking perspective. These forecasts enable a glimpse into potential trends and aid in long-term planning and policy formulation.

3. Graphical Juxtaposition: Juxtaposing graphs of smoke estimate forecasts with unemployment rate forecasts and income forecasts allows for visualizing potential relationships or trends between these variables. It helps in identifying potential correlations or patterns that might inform further analysis.

Hypothesis: Elevated smoke estimates are correlated with fluctuations in unemployment rates and household income in Williston over time.

Research Questions:

RQ1: Are there observable trends in the projected smoke estimates that coincide with anticipated fluctuations in household income and unemployment rates over the forecasted period?

RQ2: What proactive measures or policies can be devised based on the projected long-term impact of smoke on economic indicators in Williston?

In essence, the identified research gap presents an avenue for targeted research efforts that directly connect smoke estimates from wildfires to the economic dynamics of Williston. This endeavor holds promise for informed policymaking, proactive interventions, and a deeper understanding of how smoke impacts the region's economic fabric.

METHODOLOGY

Data Acquisition:

The project utilizes the "Combined wildland fire datasets for the United States and certain territories, 1800s-Present" dataset collected and aggregated by the US Geological Survey. The dataset provides fire polygons in ArcGIS and GeoJSON formats.

The data acquisition process also involved obtaining two primary datasets crucial for the analysis: the Annual Unemployment Rate (all ages) dataset covering North Dakota and the Median Household Income dataset specifically for North Dakota. The Annual Unemployment Rate dataset spans the years 2000 to 2022, offering comprehensive records of the unemployment rates across various locations within North Dakota. This dataset was selected for its temporal coverage, enabling a detailed exploration of unemployment rate fluctuations over time, particularly focusing on the city of Williston.

Simultaneously, the Median Household Income dataset, covering varying years across different locations in North Dakota, provides essential economic indicators. Although this dataset doesn't cover the entire analysis timeline (1963-2023), its inclusion offers insights into income trends across different years and locations, aiding in understanding economic variations over specific periods.

Annual Unemployment Rate (all ages) - North Dakota Dataset

Column Name	**Description**
LocationType	Type of geographical location (e.g., State, County)
Location	Specific location within North Dakota (e.g., State, County names)
TimeFrame	Year or time frame covered in the dataset
DataFormat	Format of the data (e.g., Currency)

Median household income - North Dakota Dataset

Column Name	Description
LocationType	Type of geographical location (e.g., State, County)
Location	Specific location within North Dakota (e.g., State, County names)
TimeFrame	Year or time frame covered in the dataset
DataFormat	Format of the data (e.g., Percentage)
Data	Numerical data representing the unemployment rate

The table above highlights a condensed version of the acquired datasets, showcasing key information such as the dataset name, location type, specific location, time frame covered, data format, and the corresponding data values. These datasets will serve as fundamental sources for analyzing the correlation between wildfire occurrences, economic factors (unemployment rates and household income), and their potential impacts in Williston, North Dakota.

Analytical Methods:

1. Smoke Estimate

To estimate wildfire smoke impact, the following conditions are considered:

- Analysis period: 1963-2023
- Fires within 1250 miles of the assigned city Annual fire season: May 1st to October 31st

The process involves:

- Retrieving and processing wildfire data within the specified distance from the assigned city.
- Defining a smoke estimate, considering factors such as fire size and proximity.
- Applying the smoke estimate to each fire within the analysis period.

2. Correlation Analysis

Understanding Relationships:

- Quantifying Associations: Correlation analysis, specifically Pearson correlation coefficients in this context, quantifies the degree and direction of relationships between smoke estimates and economic indicators (such as household income and unemployment rates).

- Quantifiable Insights: By generating correlation coefficients, this statistical method offers numerical insights into the potential links between variables. It helps discern whether changes in smoke estimates align with concurrent changes in economic indicators, providing a quantifiable basis for assessing their relationship.

Interpreting Associations:

- Direction and Strength: Correlation coefficients elucidate not only the existence but also the direction (positive or negative) and strength of relationships. This information aids in comprehending how closely smoke estimates and economic indicators move together or in opposite directions.

- Initial Understanding: This method serves as an initial exploratory tool, offering a foundational understanding of whether any apparent associations exist before delving into more complex analyses.

3. Prophet Forecasting:

Robust Time-Series Forecasting:

- Accommodating Seasonality: Prophet's ability to account for yearly seasonality aligns with the cyclic nature of economic indicators and smoke estimates. It captures periodic variations, allowing for more accurate long-term projections.

- Reliable Future Projections: The robustness of Prophet in handling time-series data enables the generation of reliable future projections. By leveraging historical patterns, it provides forecasts that aid in anticipating potential trends in economic indicators and smoke estimates over the next 25 years.

<u>Support for Decision-Making</u>:

- Policy Planning: The insights derived from Prophet's forecasts support long-term policy planning and decision-making processes. Policymakers and stakeholders can use these projections to anticipate potential future scenarios and formulate proactive strategies.

The rationale behind selecting correlation analysis and Prophet forecasting lies in their complementary roles. Correlation analysis offers initial insights into relationships, quantifying associations between smoke estimates and economic indicators. Meanwhile, Prophet's robust time-series forecasting capability generates reliable future projections, enabling informed long-term policy planning by accounting for yearly seasonality and historical patterns. Together, these methods facilitate a comprehensive understanding of the potential impact of wildfire smoke on economic indicators in Williston, supporting data-driven decision-making for the community's well-being and resilience.

Absolutely! Human-centered design principles are pivotal in ensuring that methodologies employed in the study not only generate insightful outcomes but also directly address the needs and concerns of stakeholders in Williston, North Dakota.

Human Centered Design

1. Stakeholder Considerations:

<u>Community Relevance:</u>

- Method Alignment: Aligning methodologies with the study's goals ensures relevance to the concerns of the community. It ensures that the research resonates with the challenges and priorities faced by the residents of Williston.

- Engagement and Participation: Involving community representatives, policymakers, and local experts in the research process fosters collaboration. This participatory approach ensures that the study accounts for local nuances and perspectives.

<u>Policy and Decision-Maker Alignment:</u>

- Relevance to Policy Making: Tailoring methodologies to align with policy making objectives ensures that the study's outcomes provide insights directly relevant to decision-makers. It aids in crafting informed policies addressing the effects of wildfire smoke on economic sectors and public health in Williston.

2. Impact Assessment:

<u>Actionable Insights:</u>

- Practical Implications: Ensuring that the study's outcomes offer actionable insights implies that the findings are not only informative but also practical. This means that

recommendations derived from the research can be implemented effectively to address real-world challenges.

- Positive Impacts: The study's focus on economic planning, and policymaking ensures that the outcomes aim for positive changes within the community. It considers the diverse demographics in Williston, catering to varied needs and vulnerabilities.

Long-Term Benefits:

- Sustainable Effects: The impact assessment aims for lasting effects by proposing strategies that can be sustained over time. By considering the diverse demographics, it aims to create inclusive solutions that benefit the entire population of Williston.

Ethical Considerations:

- Data Privacy and Integrity: Ensuring the anonymization of sensitive data to protect individual privacy while maintaining data integrity.
- Transparency and Accountability: Maintaining transparency in data sources, methodology, and reporting procedures to ensure accountability and reproducibility of findings.
- Fairness and Bias Mitigation: Consistently evaluating for biases in data collection and analysis to ensure equitable and unbiased results.
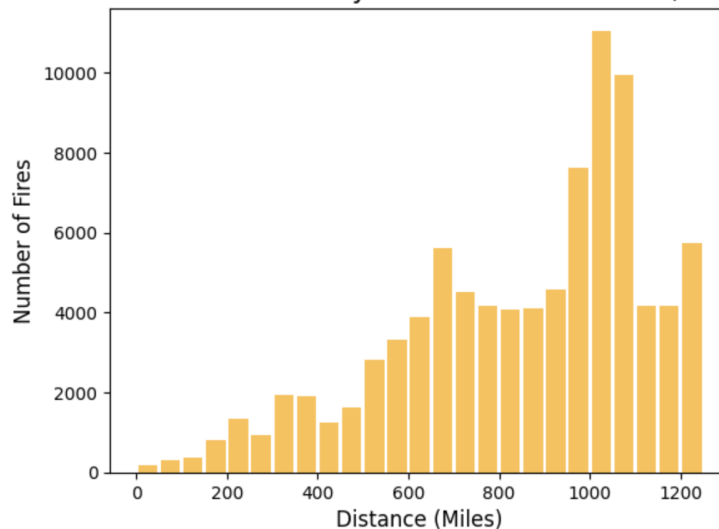
FINDINGS

Fig 1: Distribution of Wildfires every 50 miles from Williston, ND

The histogram illustrates the distribution of wildfires concerning their proximity to Williston, North Dakota, segmented into 50-mile increments. Notably, the frequency of wildfires displays an uneven pattern across distances. A prominent peak emerges around the 800-mile mark, indicating a region with a notably high incidence of fires, possibly due to specific vulnerability or conducive environmental conditions. Conversely, within a 50-mile radius of Williston, fewer wildfires occur, hinting at either robust local fire prevention measures or fewer conditions fostering fire outbreaks. Past the 800-mile threshold, a sharp decline in the number of fires is evident, signifying potential differences in environmental factors, more effective fire management strategies, or potential underreporting. Additionally, smaller peaks around the 400 and 600-mile marks suggest localized areas with increased wildfire occurrences, warranting further investigation to discern specific contributing factors driving these patterns.
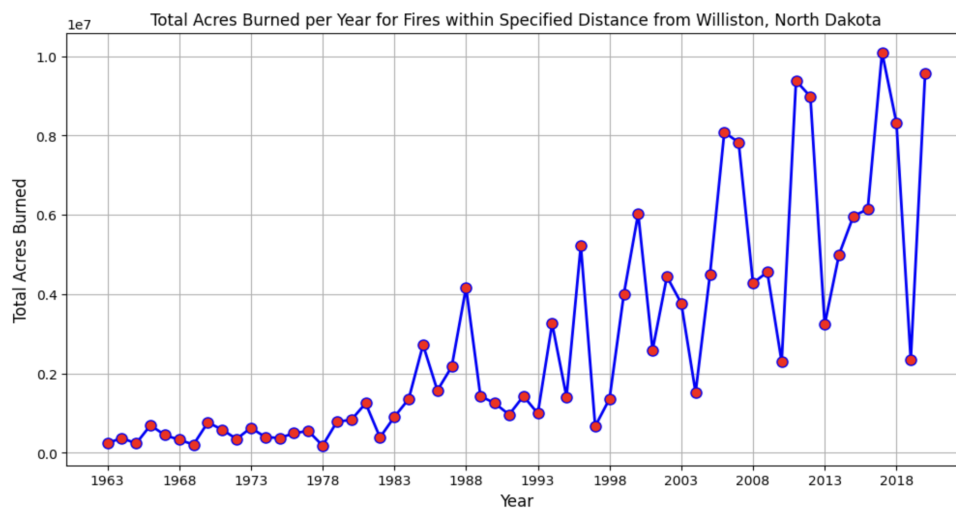

Fig 2: Total Acres burned per year

The provided line plot meticulously traces the annual total acres consumed by wildfires within a designated proximity to Williston, North Dakota, covering the period from approximately 1963 to 2018. Insightful observations gleaned from this visual representation reveal substantial dynamics:

The data showcases remarkable year-to-year variability in burned acres, indicating fluctuating factors influencing wildfire occurrence and severity over time. Notably, an upward trend emerges post-1980s, depicting a consistent rise in burned acres, with

sporadic severe years surpassing 0.8 million acres. A cyclical pattern unfolds, suggesting periodic peaks that might reflect natural climatic or vegetation cycles impacting fire prevalence. Instances of extreme wildfires align with peaks on the graph, potentially linked to drought, lightning strikes, or human activities. Recent years exhibit an uptick in both fire frequency and magnitude, possibly tied to climate shifts, alterations in land use, or natural vegetation cycles. Additionally, occasional data gaps might signify either limited information for those years or negligible acres burned during those periods.
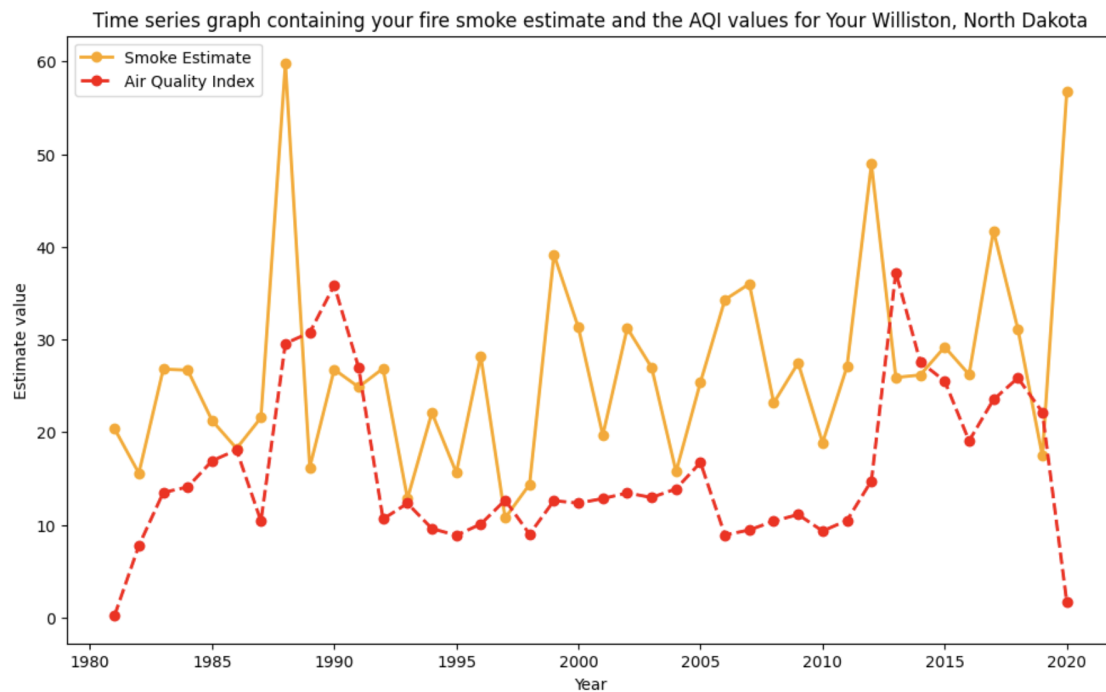

Fig 3: Smoke Estimate and AQI for Williston, ND

The graph displays a time series analysis encompassing decades in Williston, North Dakota, portraying two variables: smoke estimates and the Air Quality Index (AQI). Noteworthy observations include a correlation between peaks in smoke estimate and AQI, indicative of worsened air quality linked to increased smoke, likely stemming from heightened fire activity. Both variables showcase significant variability across years, possibly influenced by environmental factors like wildfires impacting air quality. Pronounced spikes, notably in the smoke estimate, signify years marked by substantial wildfire events generating intense smoke. The data lacks clear long-term trends, with values fluctuating, potentially in response to specific environmental conditions in different years. Towards the graph's end, a simultaneous rise in smoke and AQI values suggests recent years experiencing deteriorating air quality due to heightened smoke levels, warranting further investigation into underlying causes.
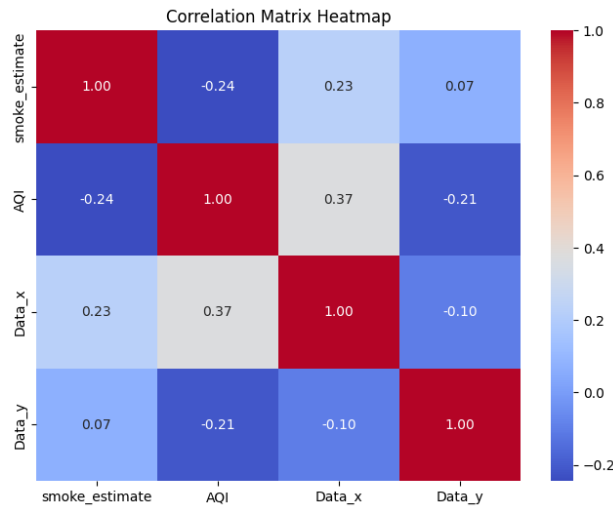
Fig 4: Heatmap for features - Smoke Estimate, AQI, Data_x (household income), Data_y (unemployment rate)

Smoke Estimate and AQI: The correlation of -0.24 indicates a moderate negative relationship between smoke estimates and AQI. Although not remarkably strong, this negative correlation suggests that as smoke estimates increase, there's a tendency for the Air Quality Index to decrease, reflecting a potential impact of smoke on air quality.

Smoke Estimate and Data_x (presumed Income): A positive correlation of 0.23 denotes a relatively weak positive relationship between smoke estimates and income. This relationship, though not strong, implies that as smoke estimates rise, there's a tendency for income levels to also increase slightly. However, this correlation isn't substantial enough to indicate a significant direct link.

AQI and Data_x (Household Income): The correlation of 0.37 suggests a moderate positive relationship between AQI and income. This implies that higher Air Quality Index values might coincide with higher income levels, indicating a potential association between air quality and income, albeit not a causative one.

Other Relationships: The correlations between Data_y and the other variables (smoke estimate, AQI, and Data_x) are notably weaker, hovering around or below 0.1. This suggests a minimal or negligible linear relationship between Data_y (unemployment rate) and the other factors.
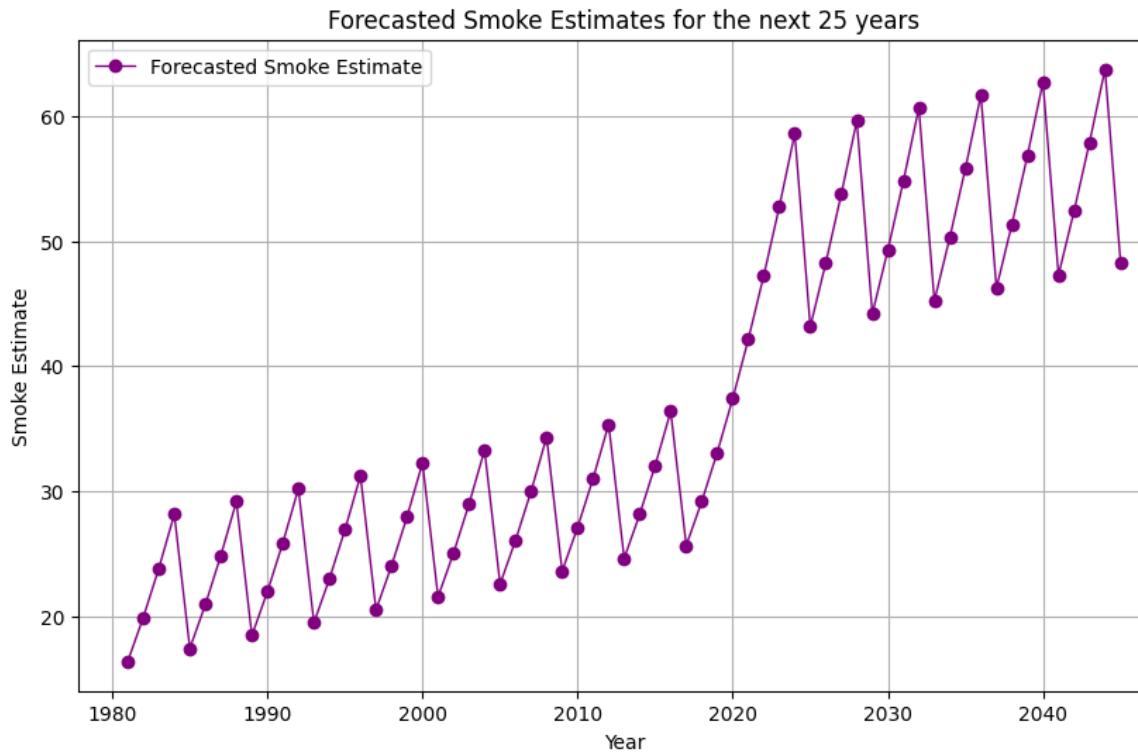
Fig 5: Forecasted Smoke Estimate (historical and forecasted (2020-2045))

This projection shows forecasted smoke estimates for the next 25 years. The trend indicates a general increase in smoke levels, with some variability year on year. This forecast emphasizes the importance of preparing and planning for future air quality challenges, suggesting that wildfire smoke may become an increasingly significant public health concern. Such projections are critical for policymakers and health professionals who are tasked with safeguarding public health in the face of climate change and its effects on wildfire incidence.
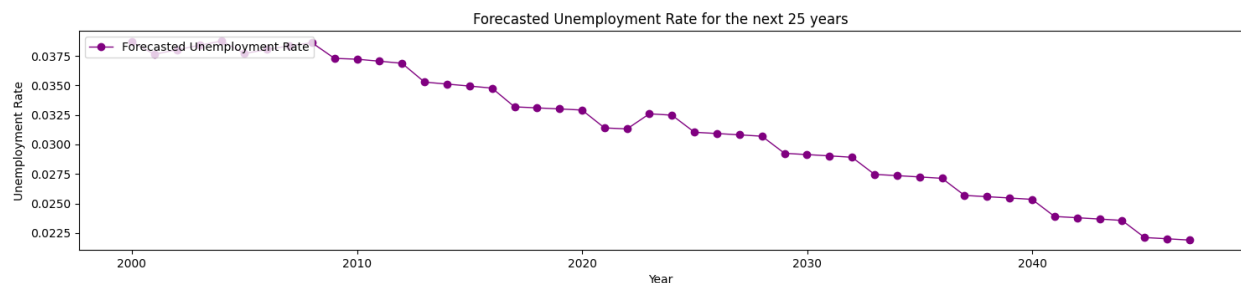

Fig 6: Forecasted Unemployment rate (historical and forecasted (2020-2045))

"This line graph shows a forecast of the unemployment rate over the next 25 years, with the trend indicating a gradual decline. If this projection holds true, it suggests a positive outlook for economic stability and job market growth. It's important for policymakers to consider such forecasts in their long-term planning to continue supporting employment opportunities and economic development."
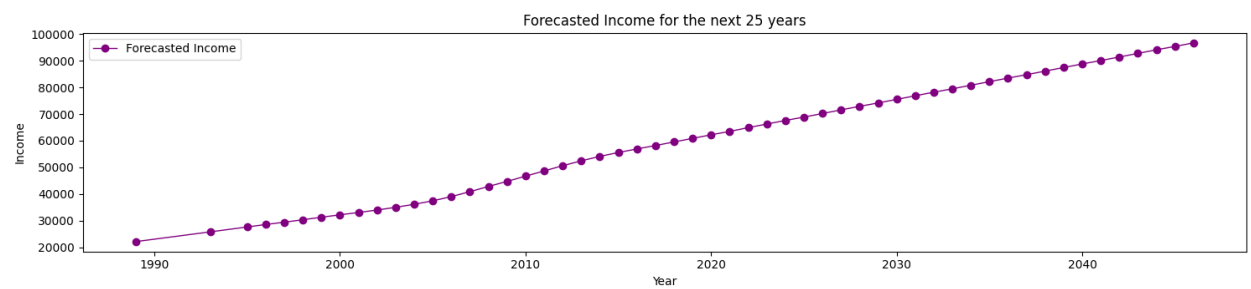

Fig 7: Forecasted Income (historical and forecasted (2020-2045))

This projection graph showcases the anticipated trend in income over the next 25 years, showing a consistent upward trajectory. This suggests that, if current economic conditions hold, there could be a steady increase in the average income levels. Such forecasts are valuable for planning economic policies, social welfare programs, and for individuals considering long-term financial planning and investments.
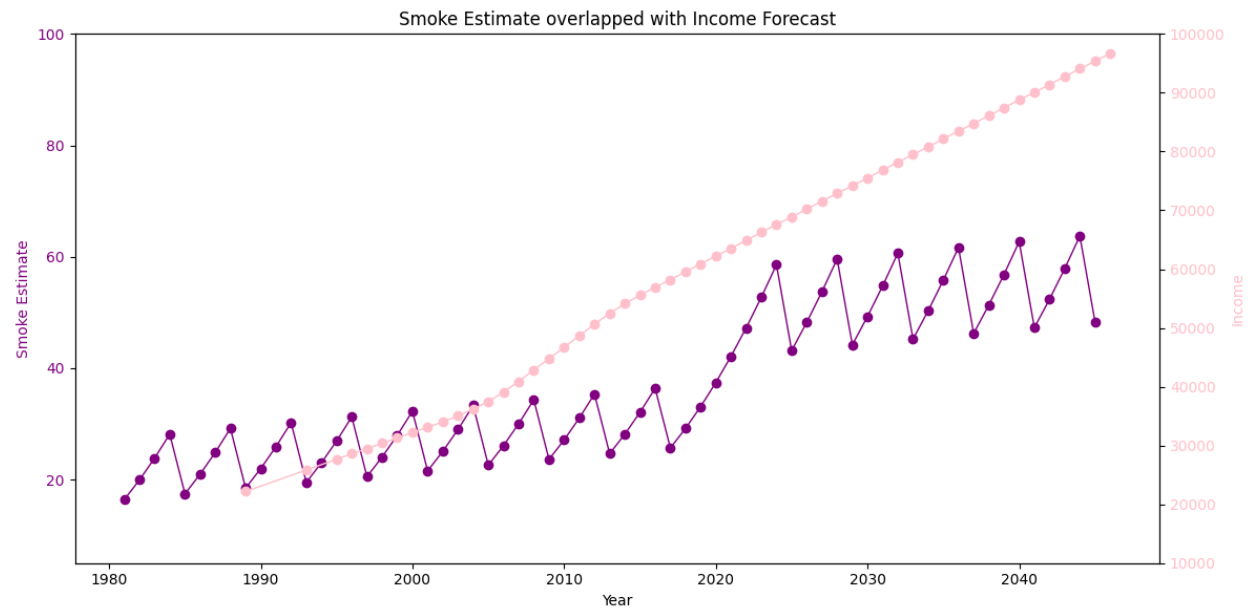

Fig 7: Smoke Estimate forecast juxtaposed with Household Income forecast

The graph presents an intriguing juxtaposition of smoke estimate trends and projected income levels over time. Here we see that, despite the fluctuations in smoke estimates due to wildfires, the income forecast shows a steady increase. This could suggest that while environmental challenges persist and may impact quality of life, economic factors such as income levels are on an upward trajectory. This raises questions about the resilience of economic growth in the face of environmental stressors and the potential long-term impacts of environmental degradation on economic stability and public health.
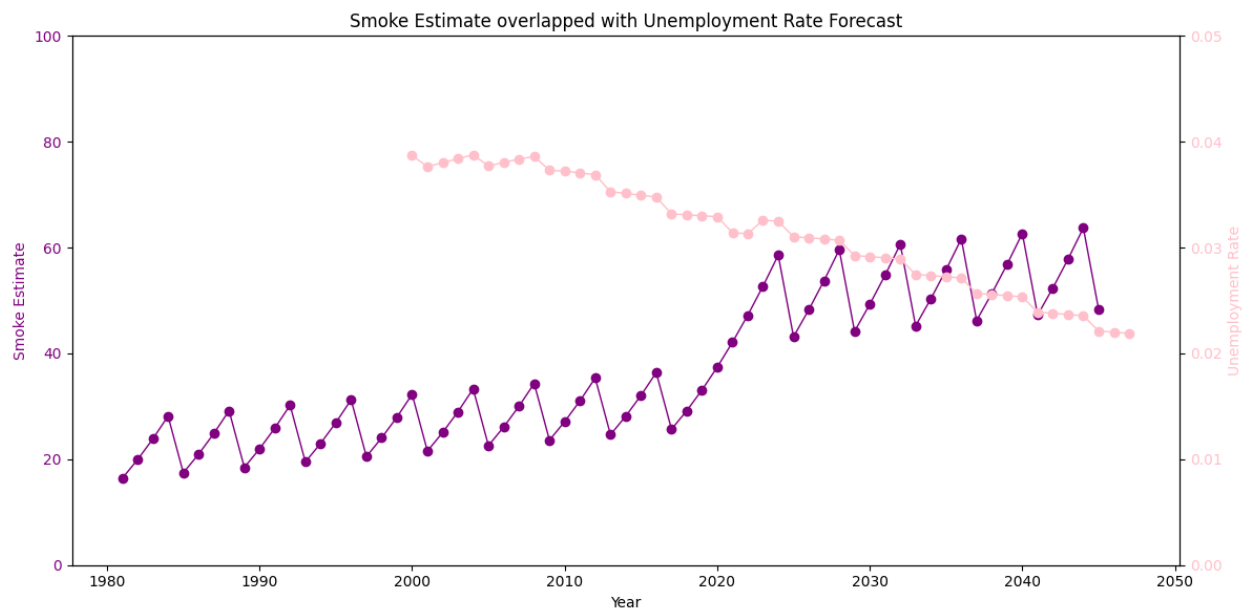

Fig 8: Smoke Estimate forecast juxtaposed with Unemployment Rate forecast

In this graph, we see an overlay of the projected unemployment rate against the smoke estimate forecasts. Interestingly, while smoke estimates show some variability, the unemployment rate is predicted to gradually decline. This divergence could suggest that economic resilience and job market growth are not directly impacted by the fluctuations in smoke levels due to wildfires. However, it does raise questions about the long-term sustainability of economic growth in the face of increasing environmental challenges and the potential for indirect effects on employment in sectors affected by air quality.

DISCUSSION

The findings from this comprehensive analysis offer valuable insights into the intricate connections between environmental factors, economic dynamics, and public health for Williston, North Dakota. These connections highlight the interdependence of various aspects crucial to the city's well-being, urging a holistic approach to address future challenges.

The correlation observed among smoke estimates, air quality, income, and unemployment underscores the complex web of relationships shaping the city's resilience and stability. Such insights underline the urgency for proactive measures and strategic planning to navigate the anticipated trends.

For the City Council, City Manager/Mayor, and residents, it becomes imperative to integrate these findings into actionable strategies. Collaborative efforts among various sectors are essential to develop adaptive policies. These policies could involve enhancing monitoring systems to track air quality, implementing targeted measures for wildfire prevention, and fostering economic diversification to reduce vulnerabilities.

Public awareness campaigns geared towards air quality management and environmental sustainability could empower residents to participate actively in mitigating the impact of environmental stressors. Additionally, investment in green infrastructure and sustainable practices aligning with human-centered data science principles can be instrumental in fostering a resilient and healthy community.

The urgency to formulate concrete plans arises within the next few years, considering the forecasted trends over the next 25 years. Nonetheless, the iterative nature of human-centered data science principles allows for ongoing evaluation, ensuring flexibility and adaptability in the face of evolving conditions. This approach emphasizes the need for continuous reassessment and adjustment in policy formulation and community engagement efforts.

Throughout this project, the integration of human-centered data science principles has facilitated a thoughtful approach, ensuring ethical considerations, stakeholder relevance, and transparent decision-making. By emphasizing the nexus between data-driven insights and tangible community impacts, these findings lay the groundwork for collaborative action aimed at safeguarding the well-being of Williston's residents.

LIMITATIONS

1. The analysis heavily relies on existing datasets, potentially limiting the depth and breadth of insights. The available data might not encapsulate crucial factors like household incomes or other intricate nuances that could significantly influence the analysis. This limitation could restrict the comprehensive understanding of the relationship between wildfire smoke and economic factors.

2. Another notable limitation involves the disparity in the frequency of forest fires within Williston itself, as evident from the initial plot. Williston showcases a lower occurrence of forest fires compared to surrounding areas, potentially affecting the direct correlation between wildfire events and the economic data specific to Williston. This discrepancy could impact the depth of analysis regarding the economic implications of wildfire smoke within the immediate vicinity. Despite this limitation, the study focuses on Williston due to data constraints and guidelines, which might not fully capture the economic intricacies within a region less affected by wildfires. This mismatch in fire occurrences within Williston and the economic data available might constrain the study's ability to draw precise conclusions about the direct impact of wildfire smoke on the local economy.

3. An additional limitation stems from the temporal scope of available economic data, which covers only specific overlapping periods within the study's timeline of 1963-2023. This constraint restricts the comprehensive analysis of economic factors, leading to potential gaps in understanding the long-term economic implications of wildfire smoke in Williston. The limited economic data availability across the entire timeline hampers the ability to ascertain consistent trends, potential shifts, or nuanced patterns in economic indicators that might correlate with varying levels of wildfire smoke exposure over the years. Consequently, it poses challenges in establishing a thorough cause-and-effect relationship between wildfire smoke and economic variables due to the fragmented nature of the economic data across the specified timeframe.

4. The model used for fire smoke estimation operates under certain assumptions regarding fire types, sizes, and distances. This simplification might not adequately capture the complex nature of smoke dispersion, potentially introducing inaccuracies in estimating the impact of smoke from wildfires on the area.

CONCLUSION

Throughout this study, the primary focus was to explore the correlation between wildfire-related factors and economic indicators in Williston, North Dakota, employing a human-centered data science approach. The research aimed to uncover potential impacts of wildfires on economic sectors and public welfare.

Research Questions:

1. Are there observable trends in the projected smoke estimates that coincide with anticipated fluctuations in household income and unemployment rates over the forecasted period?

2. What proactive measures or policies can be devised based on the projected long-term impact of smoke on economic indicators in Williston?

Hypothesis:

Elevated smoke estimates are correlated with fluctuations in unemployment rates and household income in Williston over time.

<u>Summary of Findings</u>:

- Wildfire Analysis: The analysis of wildfire frequency and acres burned over time revealed variability and periodicity in their occurrences. The study showcased variations in wildfire patterns concerning distance from Williston, identifying regions of higher vulnerability and exploring potential influencing factors.

- Economic Impact: The correlation analysis between smoke estimates, Air Quality Index (AQI), and economic indicators (**unemployment rates, household income**) suggested some relationships. While the correlations were moderate, the trends indicated potential impacts of increased smoke on economic stability and public health in Williston.

- Forecasting and Projections: Utilizing Prophet models for forecasting smoke estimates, unemployment rates, and income demonstrated trends suggesting potential challenges ahead in terms of deteriorating air quality and economic stability due to heightened smoke levels.

<u>Suggestions to County Mayor</u>:

1. Promote Economic Resilience: The county can encourage economic resilience strategies by diversifying industries. For instance, focusing on sectors less impacted by smoke-related disruptions, such as technology, healthcare, or renewable energy. A specific example might be investing in tech startups or renewable energy projects that have less susceptibility to the economic downturns caused by smoke.

2. Public-Private Partnerships (PPP): To create employment opportunities, the county can foster partnerships between public and private sectors. An example could be collaborating with private companies investing in green technologies. For instance, partnering with a solar panel manufacturer to establish a production facility in the county, creating jobs while promoting environmentally friendly initiatives.

3. Community Engagement Activity: Organize town hall meetings or workshops aimed at educating residents about income diversification and employment opportunities. Involve local businesses and experts in these discussions to provide guidance and insights into alternative income sources. For instance, host seminars where successful entrepreneurs share their experiences diversifying their income streams.

<u>Insights for Human-Centered Data Science</u>:

This study offers crucial insights into the dynamic relationships between environmental factors, particularly wildfires and their smoke, and their impact on the local economy and public welfare. By employing human-centered data science principles, this research highlights the need for holistic analyses integrating various data sources to comprehend the multifaceted impacts on communities. It emphasizes the importance of data-driven policymaking, proactive measures, and adaptive strategies to mitigate risks associated with environmental changes. Furthermore, it underscores the significance of transparent, ethical, and inclusive methodologies in interpreting data for the betterment of society.

In conclusion, this study bridges the gap between wildfire-related factors and economic indicators, shedding light on potential vulnerabilities and avenues for intervention in Williston. Understanding the nexus between these factors equips policymakers, city councils, and residents with insights crucial for decision-making, ensuring sustainable economic growth, public health, and environmental resilience. This research serves as a testament to the power of human-centered data science in unraveling complex interdependencies and facilitating informed, impactful decisions for the benefit of communities.

In this study, limitations impacted the interpretation and depth of the analysis. The economic factors, although examined, didn't exhibit strong correlations with wildfire-related variables. However, these findings should be considered with caution due to the absence of industry-specific or tourism data that might have provided more nuanced insights into the economic impacts of wildfires on specific sectors or local businesses in Williston. This absence limited the comprehensive understanding of the direct economic influence. Therefore, while the study provides valuable insights, its implications should be interpreted considering the scope of available data and the potential influence of unaccounted variables, urging a nuanced and cautious approach to its conclusions.

REFERENCES

1. https://methods.sagepub.com/book/the-data-revolution
2. https://fireadaptednetwork.org/wp-content/uploads/2014/03/economic_costs_of_wildfires.pdf
3. https://www.researchgate.net/publication/268523478_Analysing_the_effects_of_the_2002_McNally_fire_on_air_quality_in_the_San_Joaquin_Valley_and_southern_Sierra_Nevada_California
4. https://www.researchgate.net/publication/23511355_An_Economic_Analysis_of_Bushfires_Management_Programs

DATA

1. https://datacenter.aecf.org/data/tables/10756-annual-unemployment-rate-all-ages?loc=36&loct=2#detailed/2/any/false/1095,2048,574,1729,37,871,870,573,869,36/any/20598
2. https://datacenter.aecf.org/data/tables/10778-median-household-income?loc=36&loct=2#detailed/2/any/false/2048,574,1729,37,871,870,573,869,36,868/any/20623