Introduction

This analysis aims to uncover the complex ties between smoke occurrences and economic aspects such as employment, income fluctuations, and industry resilience. It's crucial to grasp how air quality intersects with employment trends and socioeconomic markers within Salina. By examining smoke estimates and AQI, it seeks to reveal their potential impact on employment rates, industries, and income-related measures, holding significant implications for public health, policymaking, and community welfare.

The motivation behind this exploration arises from the necessity to understand the diverse effects of air quality on a community. Recognizing the connections between air quality and employment dynamics, industry patterns, and income variations is pivotal for making well-informed decisions. This analysis aims to unveil vulnerabilities, strengths, and the potential impact on policies or support strategies.

This investigation tackles real-world challenges by venturing into a relatively unexplored area: the relationship between air quality and employment dynamics, industry performance, and income markers. By scrutinizing these correlations, it aims to shed light on previously unexplored links and their potential effects on community welfare. This inquiry fills a gap in comprehending the nuanced impacts of air quality on different aspects of a community's socioeconomic landscape.

Understanding the intricate ties between air quality and socioeconomic indicators holds immense importance. It can assist policymakers in crafting targeted measures to shield vulnerable groups from adverse air quality effects. Furthermore, it can inform resource allocation, public health tactics, and policy frameworks to ensure community resilience and well-being under varying air quality circumstances. Ultimately, this analysis is crucial as it provides actionable insights that can shape policies and support systems, safeguarding public health and socioeconomic stability in Salina.

Background

My research delves into the intricate interplay among wildfires, air quality measures, and diverse economic facets such as employment, income variations, and industry resilience within Salina. Expanding on prior work like Meier et al.'s "The regional economic impact of wildfires: Evidence from Southern Europe"[1], my study aims to unveil the nuanced effects of wildfires on local economies. It endeavors to reveal correlations between shifts in air quality due to smoke incidents and employment fluctuations across different industries.

Inspired by Diaz's "Economic Impacts of Wildfire" (2012)[2], my investigation goes beyond conventional indicators, seeking a holistic comprehension of wildfire effects. This study prompted me to move beyond industry behavior, focusing on how ordinary individuals are affected through socioeconomic metrics.

Moreover, Link-Herrera's "Assessing Local Government Actions in Response to Wildfire" (2019)[3] informs my research by spotlighting post-wildfire local government strategies. This study motivates an exploration of the socioeconomic implications of air quality changes and effective responses from local governments. It steers the creation of a predictive framework estimating funds needed by individuals affected by Salina wildfires for the next 25 years.

During the development of Course Project - Part 2, I explored an established model titled "The Effects of Large Wildfires on Employment and Wage Growth and Volatility in the Western United States" authored by Max Nielsen-Pincus, Cassandra Moseley, and Krista Gebert[4]. This model utilized sophisticated statistical methods, specifically GARCH models and regression analysis, to predict shifts in employment and wages across 413 counties in the western US from 2004 to 2008. Their analysis relied on Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW) data, alongside county attributes, examining significant wildfire suppression events reported by the USDA Forest Service, and socioeconomic characteristics at the county level.

In my project, I aimed to extend this investigation to Salina by leveraging datasets like the US Census Bureau's Income and Industry Data spanning from 2010 to 2021. Like the referenced model, I utilized BLS data to enrich the analysis. The Census Bureau's dataset offered crucial insights into Salina's employment landscape, industries, occupations, household income, and structures. Moreover, I integrated BLS Data Finder Historical Employment and Unemployment data, spanning from 1990 to 2023, with a focus on Salina's Monthly Employment and Unemployment records. Additionally, I incorporated the Employment by Industries dataset obtained from Data USA, leveraging specific columns such as Industry and Workforce by Industry and Gender tailored to Salina. These datasets were instrumental in establishing links between economic markers and occurrences of smoke incidents in Salina.

The decision to incorporate these datasets was guided by their relevance in understanding the impact of wildfires on employment, industry dynamics, income trends, and household structures within Salina. By building upon prior research and utilizing these datasets, I expanded and refined the analysis within Salina's socioeconomic landscape. Furthermore, drawing inspiration from the regression modeling technique employed in the referenced study, I adopted a similar methodology to scrutinize and interpret the complex interplay between wildfire occurrences and various socioeconomic factors in Salina.

My research inquiries aim to unravel the multifaceted impacts of wildfire-induced air quality changes on employment rates, industry-specific trends, and income dynamics). These Research Questions encompass:

1. Investigating the correlation between air quality indicators (Smoke Estimate and AQI) and Salina's employment or unemployment rates.
2. Analyzing the connection between smoke estimates and employment counts across diverse Salina industries.

3. Examining how smoke estimate levels relate to various income indicators across income brackets, including public assistance income and households benefiting from food stamps or SNAP programs.

Through these inquiries, my goal is to contribute to a comprehensive understanding of how wildfires reshape Salina's socioeconomic landscape, drawing insights from various economic sectors impacted by these environmental occurrences.

Methodology

The approach I used was diverse, blending analytical tools with ethical and practical factors guiding the study's setup.

Analytical Tools
Firstly, talking about the analytical side of things. I chose tools like Pandas, NumPy, and Scikit-learn to handle data and perform regression, correlation analysis, and create clear visualizations using Matplotlib, Seaborn, and Plotly Express. I picked these methods because they fit well with the study's goals. Linear regression, correlation analysis, and heatmaps were vital to exploring connections between air quality, employment trends, and income indicators in Salina.

Constructing the Smoke Estimate
I'd like to talk a little about how I came up with my Smoke Estimate. We have two primary pieces of information from the GIS data: the number of acres burned by the wildfire GISAcres and the distance between the wildfire and the city shortest_dist.

The first step is to acknowledge that several elements, including as plant type, fire intensity, wind direction, weather, and terrain, affect how much smoke is created and spread. This intricate problem is made simpler by introducing a proportional connection. It is expected that the amount of smoke that reaches the city is inversely related to the distance from the fire to the city and directly proportional to the size of the fire (measured in acres).

<u>Smoke Factor</u>
To account for the variation in smoke production per acre due to factors like vegetation type and fire intensity, a constant factor called Smoke_Factor is introduced. This factor simplifies various elements that affect the amount of smoke produced per acre. This factor takes into account Acres Burned and Distance from Salina. More about this:

- <u>Acre Classification</u>
  We categorize wildfires into three size categories: "small fire," "medium fire," and "large fire." These categories are determined based on the number of acres burned by the wildfire. For example, a "small fire" could represent fires with fewer acres burned, while a "large fire" could represent massive fires.

- Distance Classification
  Wildfires are also categorized by their proximity to the city, resulting in classifications like "close fire," "intermediate fire," and "far fire."

Based on these classification values, we then define a mapping of Smoke_Factor based on combinations of acre classification and dist classification. For instance:

- A "small fire" that is "close to the city" may have a Smoke Factor of 0.1.
- A "medium fire" that is "far from the city" may have a Smoke Factor of 0.3.
- A "large fire" that is "close to the city" may have a Smoke Factor of 0.9.

Overlap Factor

While thinking about the formula for Smoke Estimate, I consulted with my grandfather, previously the director of the Geological Survey of India. He surmised that the extracted data had notoriously little to help with the building of an accurate smoke estimate, but that aside from coming up with a Smoke Factor, I needed to take advantage of the Overlap Flags column. It was particularly valuable, as it contained the Overlap Percentage of the fires as well as the years they occurred. With this in mind, I also ended up consulting a study called "Spatial interactions among short-interval fires reshape forest landscapes" by Brian J. Harvey, Michele S. Buonanduci, Monica G. Turner[5] which goes into detail about the ecological change these reburns bring, apart from amplifying the effect of the actual fire.

Thus, I decided to introduce the Overlap_Factor, which considers the cumulative effects of previous fires and the extent to which they overlap with the current fire's location and time. This factor helps account for how multiple fires, over time and space, may contribute to the overall smoke production and dispersion. This factor takes into account both the time since the previous burn Years_Since_Previous_Burn and the extent of the overlap between the wildfires Overlap_Percentage which are extracted from the OverlapFlags column.

Finally, we construct the formula:

Smoke_Estimate = (Burned_Acres/Distance_from_City) * Smoke_Factor * (1 + Overlap_Factor)

The Overlap_Factor is computed using information from the 'OverlapFlags' column. If no overlap information is available, it is set to 0. If overlap information is present, the code extracts the years since the previous burn and the overlap percentage. The Overlap_Factor is then computed as

(Years_Since_Previous_Burn + 1) * (1 + (Overlap_Percentage / 100))

Some additional readings that contributed to this estimation are as follows:

1. Wildland fire growth prediction method based on Multiple Overlapping Solution[6]
2. Vegetation response to a short interval between high-severity wildfires in a mixed-evergreen forest[7]
3. Integrating multiple factors to optimize watchtower deployment for wildfire detection[8]

4. [Fire behavior and smoke modeling: Model improvement and measurement needs for next-generation smoke research and forecasting systems][9]

Method of each research question

<u>Research Question 1: Relationship between Air Quality Indicators and Employment/Unemployment Rates</u>

- Calculated correlations between Smoke Estimate, AQI, and Salina's Employment/Unemployment rates.
- Created and visualized linear regression models to represent relationships between Smoke Estimate, AQI, and Employment/Unemployment rates.

<u>Research Question 2: Relationship between Smoke Estimates and Employment across Industries</u>

- Analyzed Smoke Estimate correlations across different industries to assess relationships.
- Constructed and visualized linear regression models for selected industries to showcase the relationship between Smoke Estimates and Employment Counts.

<u>Research Question 3: Impact of Smoke Estimate on Income Indicators</u>

- Explored correlations between Smoke Estimate and different income brackets, visualized via a heatmap.
- Examined correlations between Smoke Estimate and income-related metrics like retirement income, median household income, cash public assistance income, and households benefiting from food stamps or SNAP programs.

<u>Forecasted Data Creation:</u>

- Imported and processed forecasted Smoke Estimate values for the future (up to 2045).
- We used historical Smoke Estimate data and linear regression to predict Mean Public Assistance Income for the upcoming 25 years.
- Generated a data frame with predicted Mean Public Assistance Income for the next 25 years and stored it as a CSV file. The forecasted_df contains the forecasted Mean Public Assistance Income for the subsequent 25 years based on predicted Smoke Estimates.

Talking about Human Centered Ethics, I'd first like to go over the assumptions of a model I've used extensively in my analysis - a linear regression model. This assumes:

- Linearity: It presumes a straight-line relationship between predictors and outcomes.
- Independence: Each observation is independent, with residuals not linked.
- Homoscedasticity: Consistent variance among residuals along the prediction range.

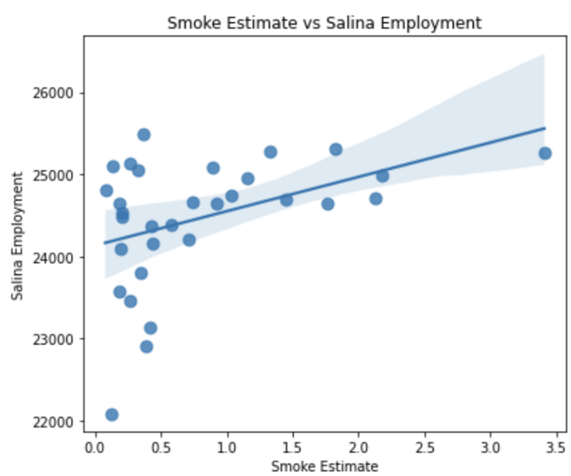
- Normal Residuals: Residuals ideally follow a normal distribution.
- No Multicollinearity: Predictors should be independent.
- No Autocorrelation: Sequential data should have uncorrelated residuals.

If any one of these assumptions were violated in my data analysis, the results would not be accurate at all. Further, my analysis stops at correlation and linear regression. It would not be fair to solely rely on these conclusions to arrive at my conclusions - as causation is very different from correlation.

It would be more ethical to reach these conclusions after making more extensive visualizations, assessing the model's residuals via Residual Analysis, Model Evaluation using R-squared or RMSE, and thorough assumption checking.

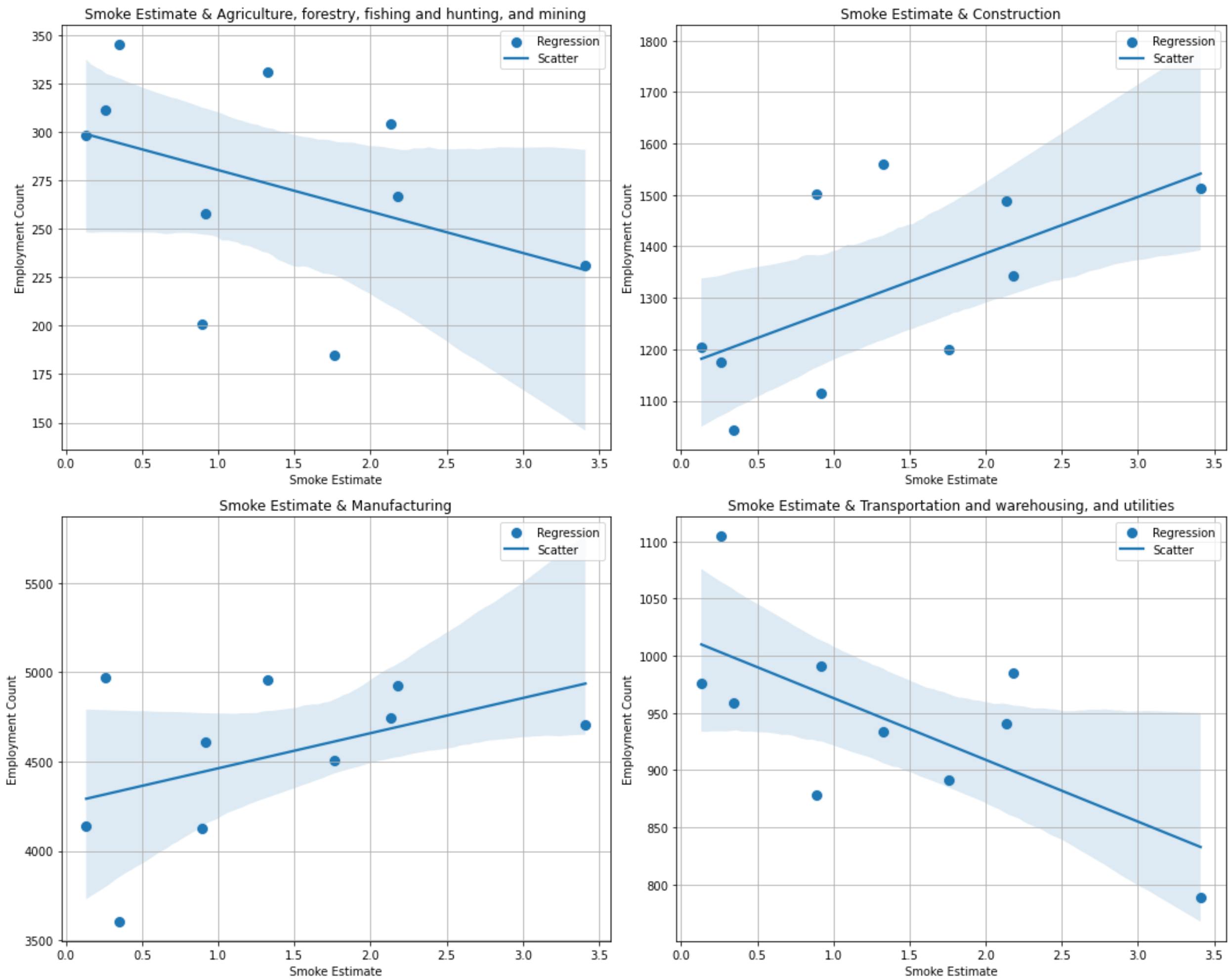
Findings

We see a positive link between Smoke Estimate and Salina's jobs.



Air quality indicators (Smoke Estimate and AQI) and employment rates in Salina reveal positive correlations. This implies an association between increased air quality indicators and higher employment rates. The linear regression models support this by highlighting a noticeable positive relationship between both Smoke Estimate and AQI with Salina's employment rates, as shown by the regression lines. Focusing on Smoke Estimate for the sake of this analysis, we justify this boost in employment by thinking along the lines of - Higher smokiness can be correlated with job boosts in Healthcare, Firefighting and Emergency Services and Construction industries.

Construction and Manufacturing are wildfire-resilient industries, while Transportation and Agriculture are not.

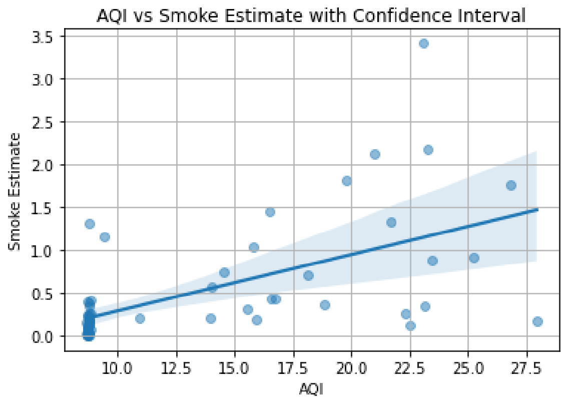


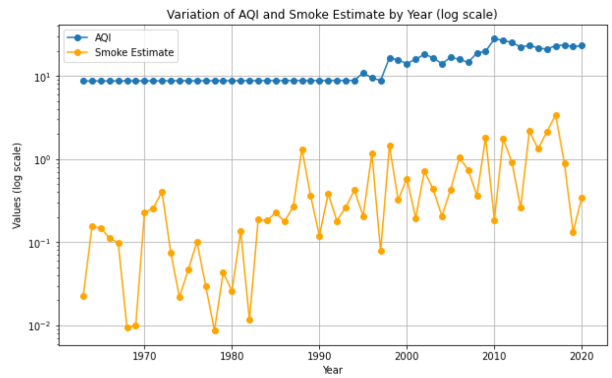
We see that the industries of Construction and Manufacturing exhibit strong positive correlations with smoke estimates. This means that higher smoke estimates align with increased employment count within these sectors. After a wildfire, construction employment rises due to reconstruction needs, like repairing infrastructure and homes. Meanwhile, Manufacturing sees increased demand for building materials, boosting employment in factories producing these supplies for rebuilding efforts.

Transportation and Warehousing, and Utilities displays a notably strong negative correlation with smoke estimates. Higher smoke estimates coincide with decreased employment count within these industries. Wildfires can disrupt transportation routes, damage utility infrastructure, and hinder warehouse operations due to safety concerns, causing a decline in employment in these sectors.

Agriculture, forestry, fishing and hunting, and mining showcase a weak negative correlation with smoke estimates. These industries might experience a minor decrease in employment due to temporary disruptions caused by wildfires, such as interrupted production or work stoppages.

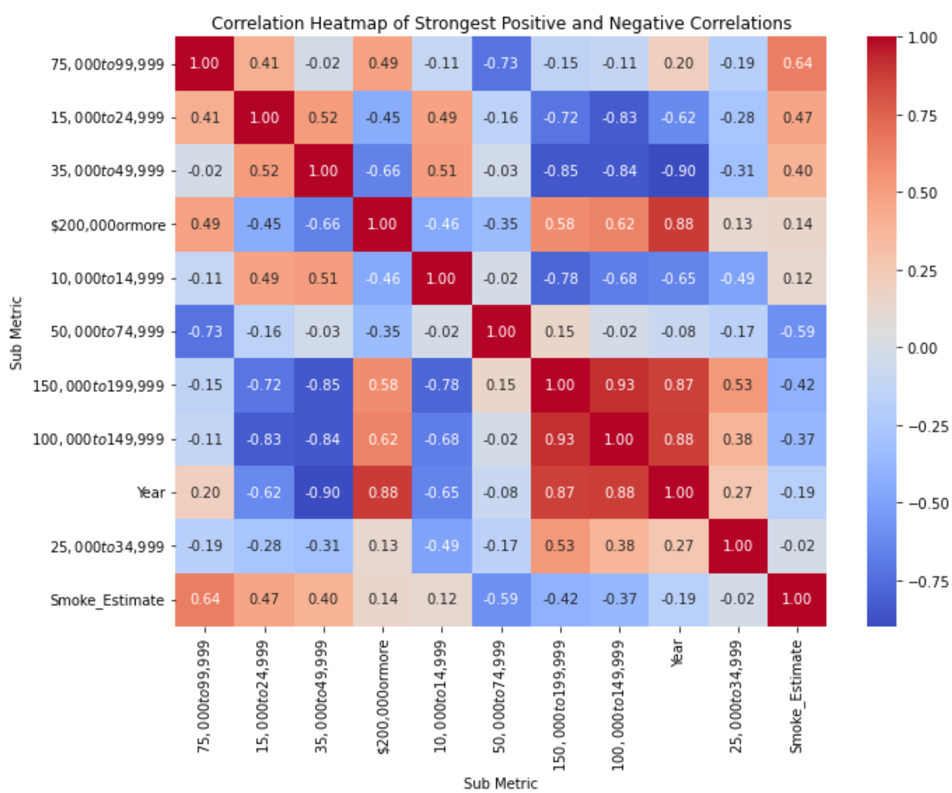
Smoke Estimate and AQI are moderately positively correlated.





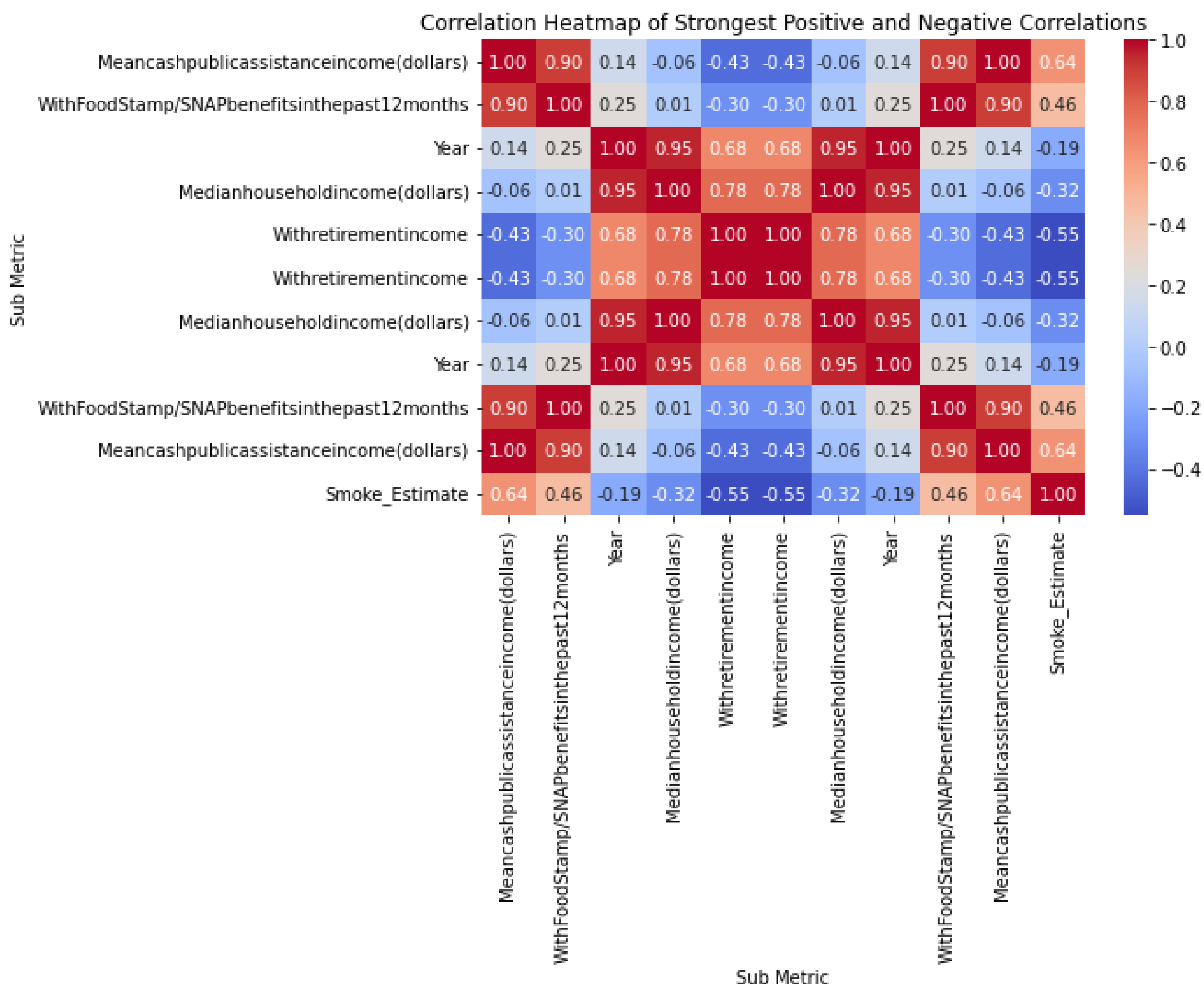
The AQI is constant during the years 1970-1995 due to the imputed values using NNLS we have set up. We see that the AQI and Smoke Estimate are (moderately) positively correlated. This means that as Smokiness increases, so does the AQI - and this is a good sign. The relationship is not as strong as I would have liked it to be though, and we see that there are quite a lot of outliers outside of the confidence interval in the first graph.

People in higher income brackets remain very weakly affected by wildfires, while people in the income brackets between $50,000-74,999 suffer.



Higher-income brackets often possess resources that help them alleviate the immediate consequences of wildfires, such as better insurance coverage, available options for relocation, or properties in safer areas. In contrast, the $50,000-$74,999 bracket might confront difficulties like fewer resources for relocation or reconstruction, resulting in a more significant impact on their financial security due to wildfires. Additionally, apart from financial aspects, socio-economic factors like access to support networks, property location, and job stability could shield higher-income brackets from immediate wildfire effects. Conversely, the $50,000-$74,999 bracket might lack some of these advantages, intensifying challenges during wildfires due to both financial limitations and limited socio-economic resources.

There is increased aid given by the government during wildfires.



During wildfires, governments typically enhance assistance initiatives, particularly targeting affected regions. Services such as food stamps and public aid see a rise in recipients amid these periods. This surge aims to aid individuals and households grappling with urgent issues like displacement, property loss, or financial stress due to the calamity, intending to relieve the socio-economic burden on those impacted by the wildfires.

Discussion/Implications

The findings from this analysis are significant for Salina's community because they uncover how air quality, particularly smoke incidents, intertwines with employment rates, income disparities, and industry resilience. Understanding these connections holds immense importance for public health, policymaking, and the well-being of the community.

In light of these findings, the city council, mayor, and residents should consider several actions.

- Firstly, they should prioritize strategies that support industries like Construction and Manufacturing post-wildfires, recognizing their potential for employment boosts. Similarly, for sectors like Transportation, Warehousing, and Utilities, implementing measures to mitigate employment declines due to wildfire disruptions becomes crucial.

- To address the vulnerabilities observed in the $50,000-$74,999 income bracket, the city council should focus on tailored support programs, considering both financial aid and community resources for affected households. This requires immediate attention to fortify their financial stability during and after wildfire events.

- Regarding a concrete plan, the city council and managers should aim to devise and implement strategies within the next six months to a year. It's crucial to act swiftly, especially considering the potential economic and social impacts of wildfires in Salina.

The findings underscore the necessity for targeted measures to shield vulnerable groups, prioritize resilient industries, and allocate resources effectively post-wildfires. Such actions will not only safeguard the community but also foster socio-economic stability amid environmental challenges.

Limitations

Data Handling Challenges:

- Incomplete Data and 0 Values: Handling data marked as 0 like in the Overlap Factor column introduces potential bias, assuming a specific value for missing data.
- Imputation Technique (NNLS for AQI): While useful, Non-Negative Least Squares (NNLS) imputation might oversimplify complex data relationships.

Modeling Techniques Limitations:

- Linear Regression Analysis: Assumes linear relationships, potentially misrepresenting nonlinear patterns.
- Exclusive Reliance on Correlation Analysis: Overlooks intricate variable interdependencies, necessitating more detailed statistical techniques.

Data Preprocessing Drawbacks:

- Handling Null Values: Dropping null values in critical datasets reduces sample size, impacting representativeness.

- SARIMA Model for Forecasting: May overlook non-stationarity and structural changes in time series data.

Data Source Reliability Concerns:

- United States Census Bureau Data: Risks associated with potential errors, inconsistencies, or data manipulation.

External Economic Factors Complexity:

- Impact on Economic Trends: Disentangling the effects of external factors like inflation or governmental policies within complex economic scenarios poses a significant challenge.

Regional Specificity Challenges:

- Shift in Dynamics: Previous research lacks specificity to Salina, potentially altering the economic narrative.

Scope and Relevance of Smoke Estimate Data:

- Geographic Consideration: Concerns about the direct applicability of smoke estimate data to Salina's economic scenario due to expansive coverage.

Absence of Population Growth Consideration:

- Unaccounted Variable: Failure to incorporate population growth in the model overlooks its potential influence on economic dynamics.

Conclusion

Let's wrap up by discussing how the findings from all three research questions are connected. First, let's start by reiterating the three research questions:

1. Investigating the correlation between air quality indicators (Smoke Estimate and AQI) and Salina's employment or unemployment rates.

2. Analyzing the connection between smoke estimates and employment counts across diverse Salina industries.

3. Examining how smoke estimate levels relate to various income indicators across income brackets, including public assistance income and households benefiting from food stamps or SNAP programs.

These questions create a linked narrative, revealing the intricate impact of wildfires and air quality changes in Salina, Kansas. While an initial strong positive correlation existed between Smoke Estimate and Employment, further investigation unveiled that this association doesn't uniformly apply across all Salina industries. Industries like Construction and Manufacturing thrive due to post-wildfire demands for reconstruction, while Transportation faces setbacks from disrupted routes and safety issues. Agriculture experiences temporary employment declines due to interrupted production. Similarly, varying impacts extend to individuals - the affluent remain largely unaffected by leveraging their resources and property, whereas the lower and upper middle-class face heightened vulnerability due to limited financial means and socioeconomic circumstances. Nevertheless, the analysis highlights a positive aspect:

government efforts to increase aid during heightened smoky periods were identified using this data.

This study's connection between air quality, socioeconomic elements, and government actions provides vital insights for policymaking and community welfare. It highlights how environmental shifts affect a community's economic health, aiding in tailored support for vulnerable groups and strengthening resilient industries after wildfires. Yet, limitations in handling data, modeling techniques, and regional focus emphasize the importance of careful analysis. This study is a foundational step in grasping the complex link between natural calamities, socioeconomic indicators, and a community's ability to bounce back. It tries to use Data Science for good in a city that has been plagued by natural disasters time and time again.

References

1. Sarah Meier, Robert J.R. Elliott, Eric Strobl, The regional economic impact of wildfires: Evidence from Southern Europe, Journal of Environmental Economics and Management, Volume 118, 2023, 102787, ISSN 0095-0696, https://doi.org/10.1016/j.jeem.2023.102787.
2. Diaz, John M. 2012. Economic impacts of wildfire. SFE Fact Sheet 2012-7. Joint Fire Science Program, Southern Fire Exchange. 4 p.
3. Link-Herrera, D. (2019). Assessing Local Government Actions in Response to Wildfire. Retrieved from https://scholarworks.calstate.edu/concern/theses/9593tx096?locale=de
4. Nielsen-Pincus, Max & Moseley, Cassandra & Gebert, Krista. (2013). The Effects of Large Wildfires on Employment and Wage Growth and Volatility in the Western United States. Journal of Forestry. 111. 404-411. 10.5849/jof.13-012.
5. Spatial interactions among short-interval fires reshape forest landscapes | Northern Rockies Fire Science Network. (2023). Retrieved from https://www.nrfirescience.org/resource/25417
6. Bianchini, G., Denham, M. M., Cortés, A., Margalef, T., & Luque, E. (2010, December 1). Wildland fire growth prediction method based on Multiple Overlapping Solution. https://doi.org/10.1016/j.jocs.2010.07.005
7. Donato, D. C., Fontaine, J. B., Robinson, W. D., Kauffman, J. B., & Law, B. E. (2008). Vegetation response to a short interval between high‐severity wildfires in a mixed‐evergreen forest. Journal of Ecology, 97(1), 142-154. https://doi.org/10.1111/j.1365-2745.2008.01456.x
8. Fuquan Zhang, Pengcheng Zhao, Shuwen Xu, Yin Wu, Xubing Yang, Yan Zhang, Integrating multiple factors to optimize watchtower deployment for wildfire detection, Science of The Total Environment, Volume 737, 2020, 139561, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2020.139561.
9. Liu Y, Kochanski A, Baker KR, Mell W, Linn R, Paugam R, Mandel J, Fournier A, Jenkins MA, Goodrick S, Achtemeier G, Zhao F, Ottmar R, French NH, Larkin N, Brown T, Hudak A, Dickinson M, Potter B, Clements C, Urbanski S, Prichard S, Watts A, McNamara D. Fire behavior and smoke modeling: Model improvement and measurement needs for next-generation smoke research and forecasting systems. Int J Wildland Fire.

2019;28(8):570. doi: 10.1071/wf18204. PMID: 32632343; PMCID: PMC7336523.

Data Sources

1. Wildland Fire Polygons Fire Feature Data Open Source GeoJSON Files.
2. US Census Bureau Income and Industry Data.
3. BLS Data Finder Historical Employment and Unemployment data.
4. Data USA.