DATA 512

Wildfire Smoke Impact Analysis Report - Lubbock, Texas

Chakita Muttaraju

11/29/2024

1. Introduction

This document outlines the analysis performed to estimate the impact of smoke from wildfires on the city of Lubbock, Texas.

Lubbock, Texas, faces significant wildfire risks, with about 59% of its buildings considered at high risk for wildfires. This risk is exacerbated by climate change, which is expected to increase the frequency of dangerous fire weather days in the area [1]. In recent years, the Texas Panhandle, including areas around Lubbock, has experienced severe wildfires. For instance, in early 2024, a series of wildfires burned over one million acres across the region, marking one of the largest wildfire events in Texas history[2][3] The impact of these wildfires has been devastating, leading to the destruction of homes and significant loss of livestock. For example, during the 2024 wildfires, more than 15,000 livestock were lost, and numerous homes were destroyed[3][4]. The fires also caused extensive damage to infrastructure, including water wells, further complicating recovery efforts[4]. Investigation of the recent fires attributed the scale of the fire to inadequate response coordination among local, state and federal entities and also to underfunded volunteer fire departments.

The aim of this analysis is to draw attention to the long-term and far reaching negative consequences of poorly handled wildfires. We first estimate the impact of just the smoke from nearby wildfires by devising a heuristic formula. We then attempt to evaluate the effectiveness of our smoke estimate by comparing it to AQI values. Finally we use our smoke estimate to find the impact of these wildfires on employment in the leisure and hospitality sectors of Lubbock. This is an important question to ask due to the revenue that the hospitality industry rakes in and contributes towards the economy. In 2023, visitors to and within Texas spent over $90 billion, creating an economic impact that supported 1.3 million Texas jobs and bolstered the state's economy. [5]

Background

Related work

A closely related research was conducted by Resources for the Future scholars Margaret Walls and Matthew Wibbenmeyer [6] where they analyzed the impacts of wildfires on California local employment growth using quarterly data on employment at the county level from the US Bureau of Labor Statistics and data on individual business establishments from the National Establishment Time Series Database. The analysis was limited to large and damaging fires, using the size of the fire in acres and a measure of the population that the fire impacted. They found that in smaller geographic areas that are closer to where the fires occur (within 3.7 miles of the fire perimeter), the employment growth rate falls by 1.3 percentage points in the year of the fire.

Research Questions

The main research question for my analysis is : "Do wildfires have a negative impact on the employment rates in leisure and service sectors?"

- The null hypothesis H_0 is "Wildfires have no impact on the employment rates in leisure and service sectors in Lubbock"
- The alternative hypothesis H_a is "Wildfires have a negative impact on the employment rates in the leisure and service sectors in Lubbock"

I framed my hypothesis that wildfires affect employment in leisure and hospitality sectors negatively based on the results of the research by Margaret Walls and Matthew Wibbenmeyer.

Data

The analysis draws from three comprehensive data sources to examine the relationship between wildfires and employment in Lubbock, Texas.

The primary dataset comes from the **USGS Combined Wildland Fire Dataset [7]**, which provides detailed geographic information for wildfires across the United States from the 1800s to present day. The data is structured in GeoJSON format, containing polygon geometries expressed in the ESRI:102008 coordinate system, and includes crucial information such as fire type (categorized as Wildfire, Likely Wildfire, Unknown-Likely Wildfire, Unknown-Likely Prescribed Fire, or Prescribed Fire), burn area measurements in both acres and hectares, fire names, and precise geometric boundaries. For analytical precision, the data was filtered to focus on fires occurring between 1961 and 2021, within 1800 miles of Lubbock (though smoke estimates specifically considered only fires within 650 miles), and during the fire season (May to October).

The second dataset comes from the **US EPA Air Quality System (AQS) API [8]**, providing historical Air Quality Index measurements for Lubbock, which serves as a validation metric for our smoke impact calculations. To quantify smoke impact, we developed a formula that accounts for fire size, distance from Lubbock, and fire type severity, expressed as: Smoke estimate = α * (1/distance) * β * size + γ, where α and β are both weighted at 0.5 to give equal importance to distance and fire size, while γ represents a severity modifier based on fire type (ranging from 0 for prescribed fires to 4 for confirmed wildfires). Distance calculations were performed using the PyProj module to account for Earth's curvature, with measurements taken from the average of all perimeter points to provide a more accurate representation of the fire's impact.

The third dataset comprises employment statistics from the **U.S. Bureau of Labor Statistics** [9], specifically focusing on two sectors in Lubbock: Service-Providing and Leisure/Hospitality employment. These datasets cover the same period (1961-2021) and include comprehensive employment figures for civilian workers aged 16 and above across all demographics, providing a granular view of employment trends at the city level. This employment data was chosen for its direct relevance to the research questions, its city-specific focus, and its accessibility as public data, requiring only proper citation of the Bureau of Labor Statistics as the source.

Model selection

I chose to use simple, explainable and open source models such as ARIMA for time series forecasting and linear regression for modeling the relationship between smoke impact and number of employees in the leisure and service industries.

Methodology

This research utilizes a transparent, human-centered approach to investigate the associations between wildfire smoke exposure, air quality, and economic impacts in Lubbock. Our approach emphasizes reproducibility and interpretability through the use of openly available datasets and explainable statistical methods, supplemented with rigorous effect size calculations and confidence intervals to provide a complete understanding of the relationships and their practical significance.

The basis of our analysis is a carefully crafted smoke estimate model that is based on three key variables: the size of the wildfire in GIS acres burned, the distance from the wildfire to Lubbock, and the classification type of the wildfire. These are combined into the smoke estimate using a weighted formula of the form: $\text{Smoke\_estimate} = \alpha * (1/\text{distance}) * \beta * \text{size} + \gamma$, where the constants α and β are set to 0.5 to give equal weighting to distance and size considerations. This equal weighting was chosen based on the principle that both factors significantly influence smoke dispersion, and in the absence of domain-specific evidence suggesting otherwise, an equal weighting provides a balanced initial approach that can be refined as more data becomes available. The γ parameter is a type-based severity modifier, which uses an ordinal scale from 0 for prescribed fires to 4 for confirmed wildfires. This reflects the typical higher intensity and emissions from uncontrolled wildfires compared with managed burns.

To validate our smoke estimates and understand their relationship with air quality, we conducted a comprehensive correlation analysis using three complementary statistical methods. We used Pearson's correlation coefficients to detect linear relationships and Spearman and Kendall's correlations to detect any nonlinear relationships. This multi-faceted approach was selected particularly because each has different strengths: Pearson's correlation is best under linear relationships but assumes normality, whereas Spearman and Kendall's correlations are more robust against outliers and thus can detect monotonic relationships. We computed effect sizes for each correlation coefficient and 95% confidence intervals to show both the magnitude and uncertainty of such relationships. This comprehensive approach helps us to capture the full range of possible relationships and also gives stakeholders clear measures of statistical certainty.
In forecasting, we applied two established time series models: exponential smoothing and ARIMA. These models were chosen after considering alternatives carefully, as they do exceptionally well for annual data without seasonality and offer interpretable parameters. We have chosen the ARIMA models because of their capability to capture both autoregressive and moving average components of time series data, making them well-suited for complex environmental data. We supplement these forecasts with prediction intervals to provide stakeholders with a clear understanding of forecast uncertainty.

The economic impact analysis is focusing on employment in leisure and service sectors, using both monthly and annual correlation analyses. Cross-correlation functions were implemented to identify potential lag effects of the economic impacts, which many times occur after the environmental event. Data standardization was done using scikit-learn to ensure scaling consistency across measures for valid comparisons. Linear regression models were favored for the predictions of employment because of their interpretability and well-understood statistical properties. Each regression model is accompanied by measures of effect size, R^2 and standardized coefficients, with 95% confidence intervals, which give a complete picture of both the magnitude and reliability of the predicted relationships.

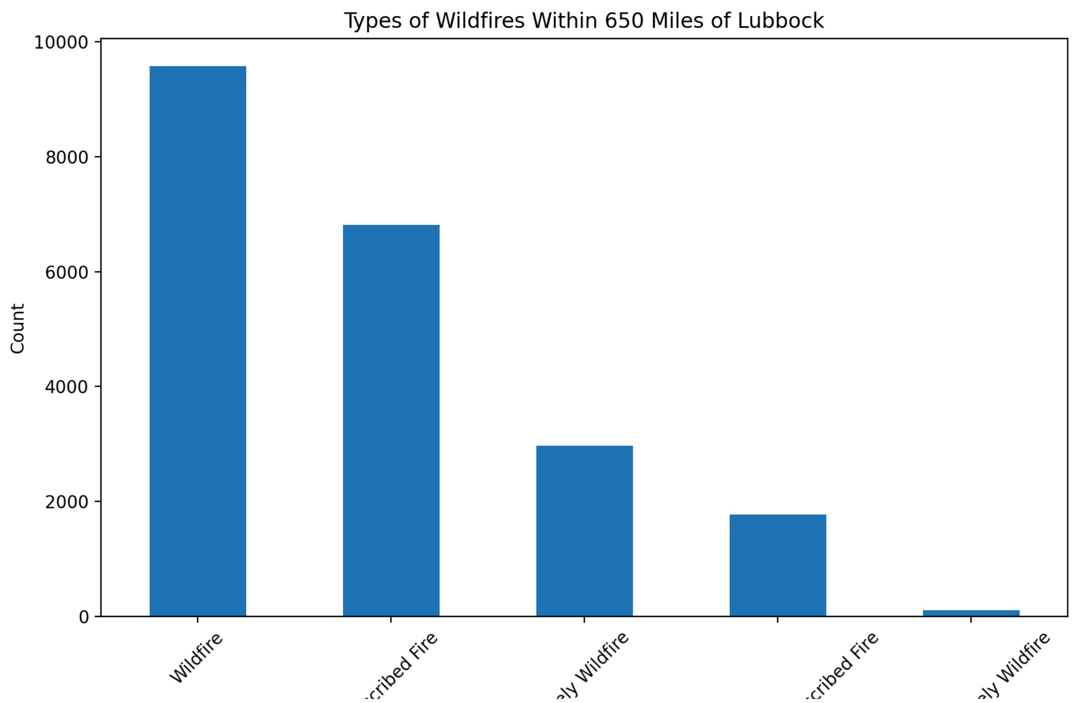
The strength of our methodology in particular lies in its commitment to reproducibility and transparency. All analyses are done on open-source datasets that have licenses allowing their use, implemented in Python code using standard libraries. This entire workflow is documented and reproducible with the code and parameters provided. We deliberately chose models and methods that are interpretable while retaining statistical rigor. For example, while more sophisticated machine learning models may capture more patterns, we decided to use traditional statistical techniques that allow for transparency of relationships and effects for easier actionability by policy makers.

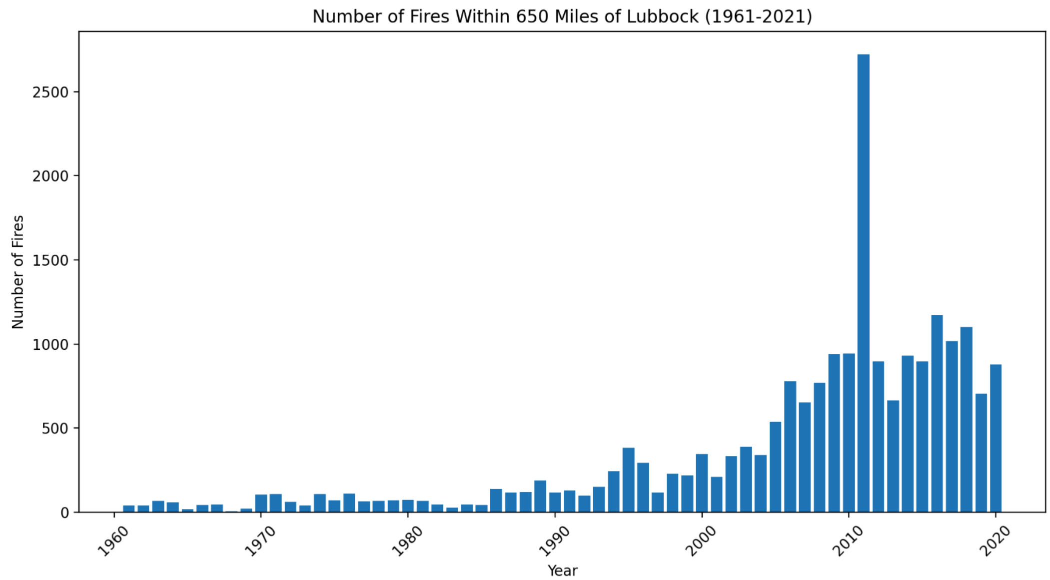
This methodology balances scientific rigor with practical utility. Our approach emphasizes human understanding through explainable models while maintaining statistical validity through established methods. By incorporating effect sizes and confidence intervals throughout the analysis, stakeholders have clear measures of both the magnitude and uncertainty of our findings. The use of openly available data and tools means our research can be verified and built upon by other researchers, while the focus on interpretable results places the findings within reach of the policymakers and other stakeholders.

Findings

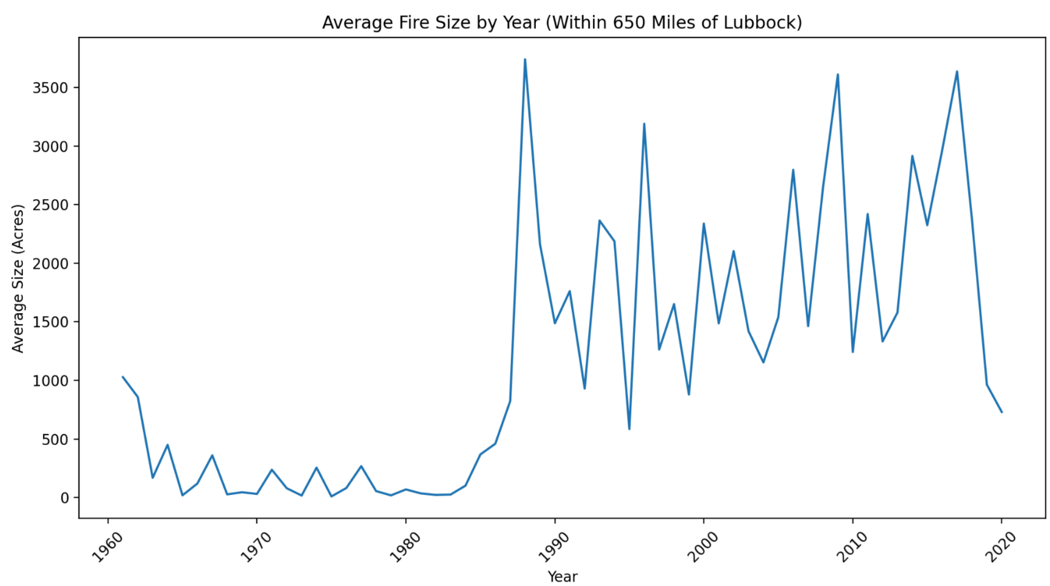
Exploratory Data Analysis

From the exploratory data analysis, we see that the most frequent types of wildfires within 650 miles of Lubbock are Wildfires. This contributes to a high smoke estimate since we weight the “wildfire”class of fires highest.





We notice that the number of fires within 650 miles of Lubbock sharply increases around 2011. This can be attributed to the fact that in 2011, a series of destructive wildfires in Texas occurred in the fire season. During 2011 in Texas, around 31,453 fires had burned 4,000,000 acres or 16,190 square kilometres(about double the previous record), 2,947 homes (1,939 of which were destroyed over the Labor Day weekend), and over 2,700 other structures. 47.3% of all acreage burned in the United States in 2011 was burned in Texas [10]



The average fire size has seen a sharp increase since the 1980s and has shown some periodicity ever since. The low values prior to 1980 could be attributed to the lack of precise records of wildfire occurrences. We notice peaks for years when Texas experienced significantly widespread fires (eg: 1998, 2011, 2017)

Air Quality Analysis

A test of the relationship between yearly smoke estimates and Air Quality Index (AQI) found no significant association. Pearson's correlation coefficient ($r = -0.010$) indicated a slight negative relationship while the non-parametric measures (Spearman's $\rho = 0.058$, Kendall's $\tau = 0.048$) both provided slight positive relationships. All these findings provide evidence that our smoke estimates contain information on dimensions of wildfire impact not captured well in the standard AQI analyses.

Employment Impact Analysis

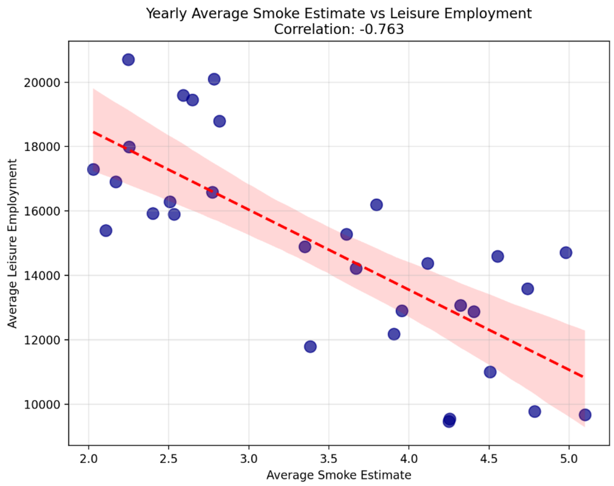
Monthly Trends

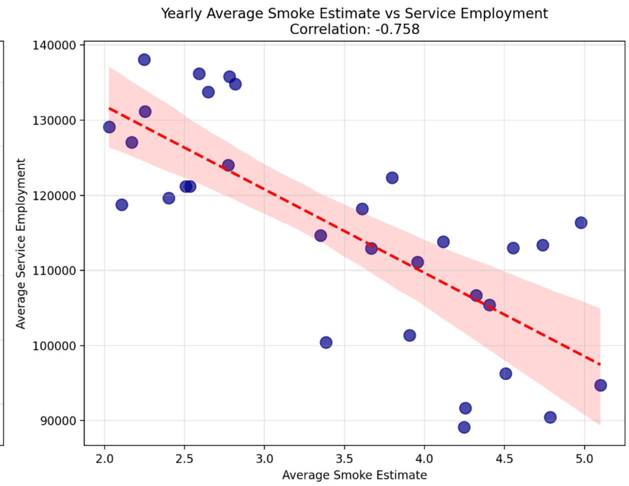
Analysis of monthly data from 1990-2021 showed consistent, statistically significant negative correlations between smoke estimates and employment in both leisure and service sectors. Pearson correlations for leisure ($r = -0.133$, $p < 0.001$) and service employment ($r = -0.134$, $p < 0.001$) showed immediate impacts with strongest correlations at zero lag. Spearman correlations suggested more robust relationships ($\rho = -0.232$ and $\rho = -0.230$ for the leisure and service sectors, respectively, $p < 0.001$), suggesting non-linear associations between exposure to smoke and levels of employment.

Annual Cycles

The aggregation of data to annual levels showed much stronger associations. The smoke estimates correlated strongly negatively with leisure employment ($r = -0.763$, $p < 0.001$) and with service employment ($r = -0.758$, $p < 0.001$). The R^2 value indicates that the smoke estimates explain almost 58% of the variance in leisure employment and 57.5% in service employment. The strong correlation found between leisure and service employment ($r = 0.990$)

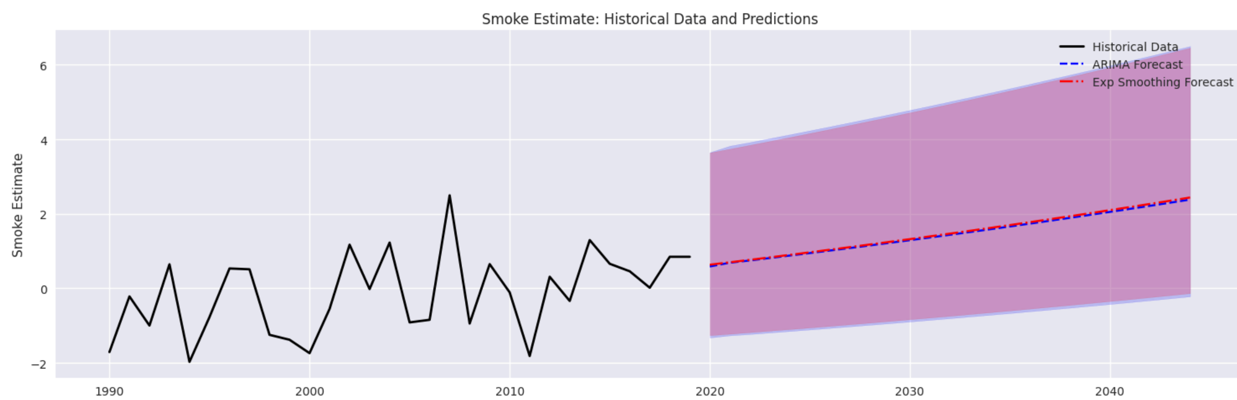
shows that both sectors have similar responses to environmental stressors.





Forecasting Evaluation

Time series modeling with ARIMA predicts the smoke estimates to increase through 2045, along with projected decreases in both employment sectors. However, the wide confidence intervals in these forecasts suggest a great deal of uncertainty in long-term predictions, reflecting the complex interplay of environmental and economic factors.



Importance

These results suggest that exposure to wildfire smoke may have significant economic impacts on the leisure and service sectors in Lubbock, at least at an annual level of analysis. The stronger correlations in annual data compared to monthly data suggest that the employment impact of smoke exposure may be cumulative rather than immediate. However, the lack of correlation with AQI measurements underscores the complexity of quantifying air quality impacts from wildfire events.

The discrepancy between monthly and yearly correlations is worthy of further investigation and might reflect longer-term adaptation strategies in these economic sectors. These results should be interpreted within the broader context of economic and environmental factors affecting employment patterns in the region.

Limitations

The current review, though extensive, operates within certain critical parameters and limitations that need to be recognized in the interpretation of the results:

Data Distributions

We make assumptions about the properties of the data distribution such as normality assumptions where appropriate and also that the data is independent to the best of our knowledge. However, these assumptions make it so that our analysis is only a rough guideline rather than a precise prediction/forecast. Extensive survey and statistical analysis is required to get a more precise idea.

Economic Data Complexity

However, the employment statistics in both the leisure and service industries are influenced by a much more complex set of elements than just the environmental factors. More major economic events, especially recessions, have shown massive impacts on the course of employment trends. For example, the financial crisis of 2008 and the COVID-19 pandemic of 2020 caused unprecedented disruptions in these sectors. Finally, seasonal changes, regional economic regulatory shifts, and broader market dynamics all impact employment disparities, rendering it very difficult to isolate the exact impact of wildfire smoke.

Environmental Confounding Variables

The diversified climate of Texas harbors different environmental challenges that may impact economic activities. In fact, from 1980 to 2024, the state has faced 19 major drought and heat wave events, accounting for more than 50% of weather-related deaths in Texas. Such extreme weather patterns can significantly alter outdoor activities and attractions and may pose confounding variables within our analysis of the impacts of wildfire smoke. Other meteorological events, such as thunderstorms and tornadoes, also have similar effects on the economic activities of leisure and services sectors, further complicating the effort to attribute variations solely to exposure to wildfire smoke.

Smoke Estimate Model Limitations

Our smoke estimate model, while providing a structured approach to quantifying wildfire impact, has several limitations.
Wind Dynamics: The current model does not consider the speed and direction of winds in the dynamical system governing smoke dispersion. Stronger winds will therefore move the smoke to farther distances or in completely erratic directions, which can lead to under- and overestimates of smoke exposure in defined regions.
Atmospheric Conditions: The model does not account for variables such as humidity, temperature inversions, and atmospheric stability—all of which can have a strong influence on smoke behavior and concentration.

These caveats mean that the smoke estimates presented should be taken as approximate indicators rather than precise quantifications of smoke exposure. Additionally, the model provides a basic framework that can be refined in future work by incorporating additional meteorological and atmospheric variables. Grasping these limitations is important for the proper interpretation of our findings and points to possible ways future research might be conducted.

Conclusion

References

[1] https://climatecheck.com/texas/lubbock

[2] https://www.texastribune.org/2024/02/27/texas-panhandle-wildfires-evacuations/

[3]https://www.texaspolicyresearch.com/investigative-committee-unveils-findings-on-2024-panhandle-wildfires

[4]https://www.texastribune.org/2024/05/01/texas-panhandle-wildfires-report/

[5] https://www.travelstats.com/dashboard/texas

[6]https://www.resources.org/common-resources/how-do-wildfires-affect-local-economies/

[7]

[8]

[9]

[10]https://en.wikipedia.org/wiki/2011_Texas_wildfires

Data Sources

The data used for this analysis was the USGS data for Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons) , specifically the GeoJSON_Files.zip file

We use the Air Quality index as a metric to compare and sanity test our smoke estimate. We get the AQI data from the US EPA Air Quality System (AQS) API.

The following data is sourced from the U.S Bureau of Labor Statistics [2] website.

1. **Employed and Office of Employment and Unemployment Statistics : Service-Providing : https://data.bls.gov/dataViewer/view/timeseries/SMU48311800700000001**

2. **Employed and Office of Employment and Unemployment Statistics : Leisure and Hospitality : https://data.bls.gov/dataViewer/view/timeseries/SMU48311807000000001**