Understanding the Wildfires' Impact on the Regional Socio-Economy for the City of Twin Falls, Idaho

DATA512 - Introduction to Human Centered Data Science

Name: Tanushree Yandra

Date: 12/11/2023

Table of Contents

1. Introduction

More and more frequently, summers in the western US have been characterized by wildfires with smoke billowing across multiple western states. There are many proposed causes for this: climate change, US Forestry policy, growing awareness, just to name a few. Regardless of the cause, the impact of wildland fires is widespread. There is a growing body of work pointing to the negative impacts of smoke on health, tourism, property, and other aspects of society.

Due to its relatively dry conditions especially during summer months, and the presence of forests, grasslands, and vegetation that can act as a fuel for wildfires, Twin Falls, Idaho faces a high susceptibility to wildfires. This makes it crucial to understand the various repercussions on the region's prosperity and well-being. Moreover, the total socio-economic impacts of wildfires go well beyond the cost of damages, as they lead to industry disruption, business closures, economic losses, and affect the job opportunities of people living in the affected regions. This in turn exacerbates economic disparities. All these reasons are the key motivation to pursue this study which is an attempt to solve the real problem of the socio-economic ramifications caused by wildfires.

This analysis thus studies the wildfires in detail and is divided into two parts. The first part focuses on studying the wildland fires within 1250 miles of the city of Twin Falls, Idaho for the last 60 years (1963-2020). A smoke estimate is then created to estimate the wildfire smoke impact which is later modeled to make predictions for the next 30 years (until 2049). The second part of the project further extends this analysis to a specific impact focus - regional socio-economy and assesses how this sector is impacted by wildfires.

From a scientific and practical perspective, studying the impact of wildfire smoke is vital for protecting public health, enabling effective emergency response, guiding policy decisions, fostering community preparedness, and mitigating the various detrimental effects on both health and the environment. This will thus equip policy makers, city managers, city councils, or other civic institutions with data-driven information allowing them to make informed decisions related to public safety measures and resource allocation during wildfire events.

2. Background/Related Work

Extensive research has already been done in the domain of understanding the socio-economic impact of wildfires. However, no concrete research exists specifically for Twin Falls county or the state of Idaho. Some of the works that resonated the most with this study have been discussed here. The paper "Economic footprint of California wildfires in 2018" (*Meier et al., 2023*) found that the economic impacts of wildfires in California in 2018 were substantial, totaling $148.5 billion, roughly 1.5% of California's annual gross domestic product. These economic impacts included the value of destroyed and damaged capital, health costs related to air pollution exposure, and indirect losses due to broader economic disruption cascading along regional and national supply chains. Additionally, the study revealed that the United States beyond California also suffered considerable economic damages ($45.9 billion) related to California's wildfires, indicating that large wildfires in California or other western states have far-reaching economic impacts beyond the immediate areas affected by the fires.

Another paper "The regional economic impact of wildfires: Evidence from Southern Europe" (*Wang et al., 2020*) estimated the impact of wildfires on the growth rate of gross domestic product (GDP) and employment of regional economies in Southern Europe from 2011 to 2018. The results suggest an average contemporary decrease in a region's annual GDP growth rate of 0.11–0.18% conditional on having experienced at least one wildfire. A decrease in the average annual employment growth rate for activities related to retail and tourism (e.g., transport, accommodation, food service activities) of 0.09-0.15% was also found.

The insights from these previous studies guided the formulation of the current study's hypotheses regarding the socio-economic impact of wildfires for Twin Falls, Idaho. Specifically, the aim of this study is to build upon the established findings from prior research in similar contexts, and extend it to four indicators. The first indicator - GDP, a widely used indicator referring to the total gross value added by all industries in the economy (*WDI - Economy, n.d.*), was chosen to understand the overall impact of wildfires on the region's economy. Next, in order to choose socio-economic indicators, the article from Centers for Disease Control and Prevention (*Centers for Disease Control and Prevention, 2023*) was referred to. The top three socio-economic indicators were defined as education, employment, and income. However, since there was no education level data available for the city of Twin Falls, Idaho, only employment and income data were considered. Finally, to understand if job losses and decrease in income in turn aggravate economic disparity, income inequality was chosen as the fourth indicator.

Thus, these four indicators - GDP, unemployment rate, income, and income inequality formed the basis for this study. Specifically, the analysis wanted to focus on four research questions,

1. Do the years affected by heavy smoke experience dips in economic productivity? The aim is to assess the impact on GDP growth during periods of intense wildfires.
2. Do unemployment rates increase during and immediately after wildfire events due to disruptions in economic activities? The goal is to analyze how wildfires affect unemployment rates in periods experiencing severe smoke exposure.
3. Did years with higher smoke exposure experience declines in personal income? The goal is to investigate how wildfires influence personal income per capita.
4. Finally, do low-income groups suffer disproportionately due to wildfires' economic impacts? The aim is to explore if income inequality exacerbates the economic vulnerability of low-income groups during and after wildfires.

To answer the above research questions, the four indicators were sourced from four different datasets from the Federal Reserve Economic Data (FRED), maintained by the Research Department at the Federal Reserve Bank of St. Louis. It is an online database consisting of hundreds of thousands of economic data time series from multiple national, international, public, and private sources. All the four datasets are under the Fred® Services General License. They are under copyright; but, provided that one has not engaged in any prohibited uses, one may use these data series with proper attribution of the source. For more information, please refer to the Terms of Use. The four data sources and their uses have been mentioned below:

1. Decrease in Economic Output: For this, the dataset of "Real Gross Domestic Product: All Industries in Twin Falls County" was used. This is an annual dataset from 2001 to 2021 representing the GDP - a measure of the market value of final goods and services produced within a county in a particular period. The dataset contains only two columns - date in the YYYY-MM-DD format, and the GDP.
2. Unemployment Fluctuations: For this, the dataset of "Unemployment Rate in Twin Falls County" was used. This is a monthly dataset from 1990 to 2023 representing the percentage of unemployed people in the total civilian labor force. The dataset contains only two columns - date in the YYYY-MM-DD format, and the unemployment rate as a percentage.

3. Effect on Personal Income: For this, the dataset of "Per Capita Personal Income in Twin Falls County" was used. This is an annual dataset from 1969 to 2022 representing the personal income of the residents in the area divided by the resident population. The dataset contains only two columns - date in the YYYY-MM-DD format, and the per capita personal income.

4. Income Inequality during Crisis: For this, the dataset of "Income Inequality in Twin Falls County" was used. This is an annual dataset from 2010 to 2021 representing the ratio of the mean income for the highest quintile (top 20 percent) of earners divided by the mean income of the lowest quintile (bottom 20 percent) of earners in a particular county. The dataset contains only two columns - date in the YYYY-MM-DD format, and the income inequality ratio.

To model these datasets and make predictions to understand future implications, no existing model was adapted or adopted. Instead, the model trained on smoke estimates was used to serve this purpose. This model was the Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX) model. It is a part of the statsmodels library, used for time series analysis and forecasting. The SARIMAX model takes several parameters:

1. endog: This represents the endogenous variable, the time series data you want to model.

2. order (p, d, q): These three parameters correspond to the autoregressive (p), differencing (d), and moving average (q) components of the model, similar to the ARIMA model.

3. seasonal_order (P, D, Q, s): The seasonal components of the model, denoted by uppercase P, D, and Q. The 's' parameter defines the seasonal length.

4. exog: This represents exogenous variables, additional independent variables that can impact the endogenous variable. This is an optional parameter.

5. trend: Specifies the trend component in the model. Options include 'n' for no trend, 'c' for a constant term, 't' for a linear trend, and 'ct' for both.

The SARIMAX model can handle seasonal patterns in data, making it suitable for time series with recurring patterns at fixed intervals. It combines autoregressive (AR) and moving average (MA) components to capture dependencies between observations and the impact of previous error terms, which can be vital in time series analysis. It includes differencing to make a time series stationary, which can help stabilize variance and make data more amenable to modeling. Thus, this model is effective for making predictions based on historical trends and other factors.

3. Methodology

3.1 Generating and Modeling the Smoke Estimate

As mentioned earlier, the first part of the study focuses on analyzing the wildland fires within 1250 miles of Twin Falls, Idaho for the last 60 years (1963-2020). A smoke estimate was then created to estimate the wildfire smoke impact which was later modeled to make predictions for the next 30 years (until 2049). As a human-centered research, to foster trust and accountability, the methodology and the decision-making process is transparently communicated in this report, clearly stating all the assumptions that were made during the analysis.

3.1.1 Data Retrieval

This section of the study discusses the data retrieval process. Two datasets were used for the first part of the analysis - Wildland Fires Dataset to get access to the historical data of wildfires, and the Air Quality Index Dataset to validate the smoke estimate.

Wildland Fires Dataset

This dataset was collected and aggregated by the US Geological Survey, and is relatively well documented. Fire polygons in this dataset are available in ArcGIS and GeoJSON formats. The following steps were taken to read and filter the data:

1. Data Reading: The data files were downloaded in the GeoJSON format and a GeoJSON reader was used to read the data. The GeoJSON reader was sampled from the 'wildfire' module (*wildfire.zip, n.d.*) which is a user module including one object, a Reader, that can be used to read the GeoJSON files associated with the wildefire dataset. This module was provided under copyright by the author and cannot be redistributed. For any permissions, please reach out to the author - Dr. David W. McDonald.

2. Coordinates Conversion: The wildfire data was expressed in the ESRI:102008 coordinate system. The most commonly used geographic coordinate system however is 'WGS84', also called 'decimal degrees' (DD). Thus, in order to compute the distance of the wildfire from Twin Falls, the coordinates had to be converted to the DD system. This was done by taking the geometry of a fire feature, extracting the largest ring (i.e., the largest boundary of the fire) and converting all of the points in that ring from the ESRI:102008 coordinate system to DD coordinates.

3. Computing the Geodetic Distance: Since fires are irregularly shaped, the next step was how the notion of 'distance' be defined. For example, should the distance be calculated from the closest point of the fire perimeter or the centroid of the region? For this study, the average distance of all perimeter points to Twin Falls, Idaho was considered because for some very large fires, including just the shortest distance seemed to be biased.

4. Generating the final dataset: The individual wildfire incidents were looped through to get the 'ring' boundary of that specific fire - which is a list of geodetic coordinates. These coordinates were then used to compute the distance between the wildfire and Twin Falls, Idaho. While reading the data, geodetic distance computations were simultaneously being made to filter those wildfires that were within 1250 miles of Twin Falls, Idaho. The code for the same was sampled from a notebook (*wildfire_geo_proximity_example.ipynb, n.d.*) licensed under the Creative Commons CC-BY license.

Thus, the final wildfires dataset had a total of 84319 wildfire incidents in the period 1963-2020 where all the wildfires were within 1250 miles of Twin Falls, Idaho.

Air Quality Index Dataset

US Environmental Protection Agency (EPA) Air Quality Service (AQS) API is a historical API which was used to obtain the Air Quality Index (AQI) data. The AQI gives information on how healthy or clean the air is on any day. Lower AQI values indicate cleaner air. The API does not provide real-time air quality data. The documentation for the API provides definitions of the different call parameters and examples of the various calls that can be made to the API. The following steps were taken to retrieve the yearly AQI data:

1. Retrieving the list of Monitoring Stations: Air quality monitoring stations are located all over the US at different locations. The FIPS code of Twin Falls, Idaho was used to get a list of nearby stations during the API call.

2. Get the Yearly AQI data: The AQI data in the API was available on a daily basis which had to be converted to a yearly value. This could be done by taking an average of all the AQI data for that year or picking the maximum value. Moreover, instead of collecting the entire year's data to estimate the annual AQI, the data was pulled only for the period of fire season that runs from May 1st through October 31st. Since the aim is to understand the smoke estimate's relation with AQI levels, picking the maximum AQI level from the fire season period every year seemed appropriate.

3. Handling Missing Values: The AQI Data was found to be missing for the years 1992, 1993, 1994 and 2014. These values were first replaced with an empty value instead of zero, and filled with the rolling average value of the five previous AQI values. While these values may not be accurate, considering five previous values' rolling average was a reasonable estimate for the missing AQI data.

The code for performing the above tasks was sampled from a sample notebook (*epa_air_quality_history_example.ipynb, n.d.*) provided under the Creative Commons CC-BY license. In the final dataset, it was found that the AQI data for monitoring stations in Twin Falls county was available from 1986 onwards. Thus, even though the wildfires dataset ranges from the year 1963 to 2020, the AQI data is available from 1986 only.

3.1.2 Data Preprocessing

This section discusses the various columns present in the Wildland Fires dataset and relevant assumptions that were made to remove any unnecessary columns.

Dropping Columns not Useful for the Analysis

Several columns not useful for the analysis were identified and dropped from the dataset. The explanation for dropping each of these columns is provided below,

1. OBJECTID: It is a unique identification for the fire polygon and its attributes. The dataset had another column 'USGS_Assigned_ID' which was also a unique identification that provides further consistency. Thus, the OBJECTID column was dropped.

2. Fire_Polygon_Tier: This refers to the tier from which the fire polygon is generated. This feature, although numerical, did not feel like it would add any value to the creation of the smoke estimate and its modeling. Thus, it was dropped.

3. Fire_Attribute_Tiers: This feature has a list of Polygon Tiers consolidated from all the data sources for each fire. This was irrelevant to the analysis, and was hence dropped.

4. GIS_Hectares: This encapsulates the hectares of the fire polygon calculated by using the Calculate Geometry tool in ArcGIS Pro. Since there was another column representing the same value in the units of acres, this column was dropped.

5. Source_Datasets: This column contains all the original source datasets that contributed to either the polygon or the attributes. This was irrelevant for our analysis.

6. Listed_Fire_Types: This includes each fire type listed in the fires from the merged dataset. The 'Fire_Type' column represents the same value and thus, it was not needed.

7. Listed_Fire_Codes: This includes each fire code listed in the fires from the merged dataset. Any feature that had a 'list' of values from the merged dataset were ignored.

8. Listed_Fire_IDs: This includes each fire ID listed in the fires from the merged dataset. Since it was a 'list', it was dropped.

9. Listed_Fire_IRWIN_IDs: This includes each fire IRWIN ID listed in the fires from the merged dataset. This was dropped since it was a 'list'.

10. Listed_Fire_Dates: This includes each fire date listed in the fires from the merged dataset. Since wildfires were being considered on a yearly basis, this was not needed.

11. Listed_Fire_Causes: This includes each fire cause listed in the fires from the merged dataset. It was a 'list' and was hence dropped.

12. Listed_Fire_Cause_Class: This includes each fire cause class listed from the merged dataset. While this seemed important, it could not be quantified and was thus dropped.

13. Listed_Rx_Reported_Acres: This contains each prescribed fire's reported acres listed in the fires from the merged dataset. For the area of the fire, the column 'GIS_Acres' was being used and hence, this column was dropped.

14. Listed_Map_Digitize_Methods: This includes each fire digitization method from the merged dataset. This did not add any value to the analysis and was thus dropped.

15. Listed_Notes: This contains additional notes associated with each fire from the merged dataset. Notes were irrelevant to this study.

16. Processing_Notes: This provides some rationale when the attribute data was altered during the processing and a new attribute was added. These notes were not relevant.

17. Wildfire_Notice and Prescribed_Burn_Notice: These are a notice present in every field indicating the quality of the wildfire/prescribed burn data. These were not needed.

'Assigned_Fire_Type' and 'Wildfire_and_Rx_Flag' columns

The 'Assigned_Fire_Type' column was one of the five types - Wildfire, Likely Wildfire, Unknown - Likely Wildfire, Prescribed Fire, Unknown - Likely Prescribed Fire. The key

difference between Wildfires and Prescribed fires is the intent. A prescribed fire is a planned fire intentionally ignited by park managers to meet management objectives. A wildfire on the other hand, is an unplanned fire caused by lightning or other natural causes, by accidental (or arson-caused) human ignitions, or by an escaped prescribed fire. While prescribed fires are intentional and usually in control, they still do contribute to air pollution. For this analysis, it was assumed that prescribed fires and wildfires contribute to the same amount of pollution for a given land within the same area.

The 'Wildfire_and_Rx_Flag' column is a text flag field indicating that the attributes from the various data sources flagged a fire as both a wildfire and a prescribed fire. This could indicate an error in assigning the fire type, a misassignment of the fire type, or that there were actually two fires that occurred in this area in the same year, one a wildfire and one a prescribed burn. Since both wildfires and prescribed fires were treated in the same manner, this field was ignored.

'Overlap_Within_1_or_2_Flag' column

In the wildfires dataset, fire polygons with near 100% overlap in consecutive years could be the same fire in different datasets with a year value that is correct in one and incorrect in another. This can occur particularly with older fires. There is no way to identify the actual year or which one is correct, if one is in fact incorrect. Therefore, 'Overlap_Within_1_or_2_Flag' column is present to flag areas that burned with >10% overlap of the current fire within 1 or 2 years of the current burn. Each fire that meets this criteria is included in this attribute including the percentage and acres of overlap, the year the overlapping fire occurred, and the overlapping fire's Assigned_USGS_ID. While the overlap flag may or may not be correct, it is assumed that another row pertaining to the same fire exists for those fires that are flagged. Overlapping fires were thus removed since there is another fire already existing in the database with more than 10% overlap.

Near Perfect Circle Wildfires with High Acreage

Some of the fires in the wildfires dataset appear as near perfect circles. This could be from lightning strikes being counted as small fires or other small fires having a point buffered to the area of the fire size because no true polygon was created. A circle-ness index is thus calculated by using the following equation in Field Calculator,

$$\frac{4\pi \; x \; Shape\ Area}{Shape\ Length^2}$$

As values of the circle-ness index approach 1, the shape becomes more circular. A 'Circle_Flag' column is thus present to flag any shapes with a value greater than or equal to 0.98.

Circular fire polygons are highly unlikely to represent the actual area burned. When fire size is less than 1 acre, the risk of mis-assigning the burned area is minimal given the fire size. For any circular polygon greater than 1 acre, the risk of mis-assigning a burned area is too high and hence these are not included in the analysis. The column 'Exclude_From_Summary_Rasters' has a flag - 'Yes' for fires that are circular and greater than 1 acre, and 'No' for non-circular fires and circular fires less than 1 acre. Thus, the fires that had the flag as 'Yes', were removed.

Thus, the final processed wildfires data contained 72608 wildfire instances with 9 features.

3.1.3 Generating and Evaluating the Smoke Estimate

Creating the Smoke Estimate Formula

The next step was to create an annual estimate for wildfire smoke in Twin Falls, Idaho. This estimate was just a number that would eventually be used to build a predictive model. It seemed reasonable that a large fire burning a large number of acres, close to the city would put more smoke into a city than a small fire that is much further away. Thus, the variables 'GIS_Acres' and 'Distance' were used to define the smoke estimate which would then be applied to every fire within 1250 miles of Twin Falls between 1963 and 2020. It is important to note that the column 'GIS_Acres' was converted to square miles since the 'Distance' values were stored in miles.Thus, the two variables had to be plotted to understand their inherent relationship.

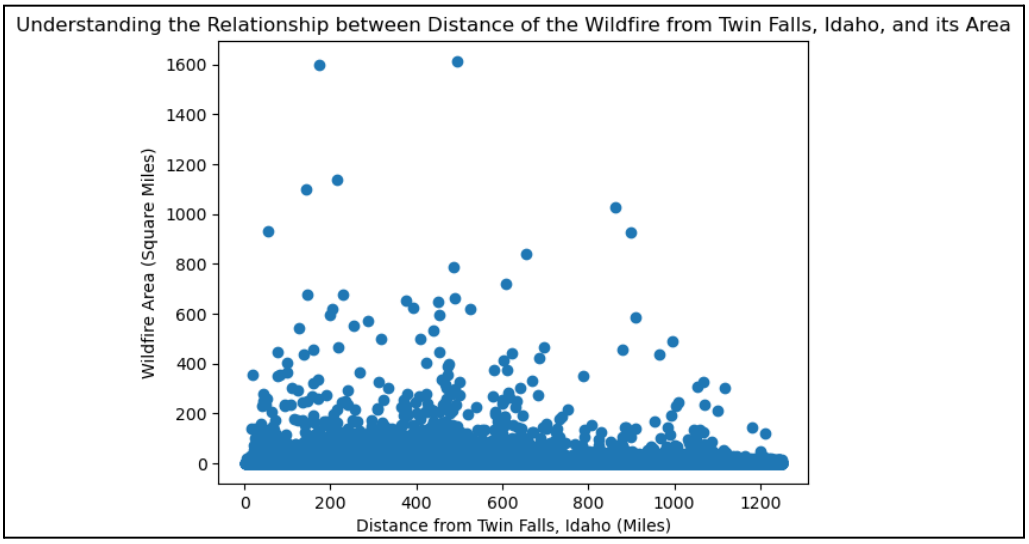

Fig. 1. Analyzing the Relationship between Distance of the Wildfire from Twin Falls and its Area

The relationship in the above figure (Fig. 1) looked pretty random where most of the wildfires were concentrated in smaller areas with no apparent relationship visible. The square of the 'Distance' was plotted instead with the 'GIS_Square_Miles' variable.

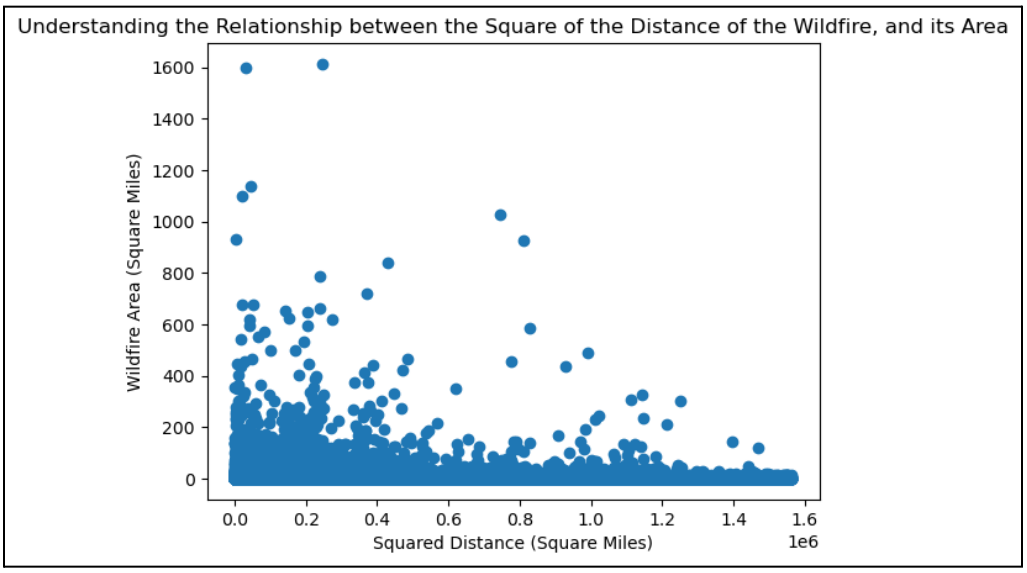

Fig. 2. Analyzing the Relationship between the Square of the Distance of the Wildfire and its Area

From Fig. 2, there seemed to be some improvement in the manner in which the points were scattered. It could be observed that the data points whose square of distance is smaller, tend to have a higher area. The relationship could be approximated to a decaying exponential function. Thus, the smoke estimate that best captured the overall intuition was created using the following formula,

$$Smoke\ Estimate = w_1 \times Area\ of\ Fire \times \exp(-w_2 \times Distance^2)$$

The aim was to not have a very high smoke estimate value and keep it under 50. Based on the existing data points, w_1 and w_2 were chosen in such a way that the overall smoke estimate value stayed under 50. Thus w_1 and w_2 for this analysis were chosen to keep the number of digits of 'Distance_Square' and 'GIS_Square_Miles' under control. Since 'GIS_Square_Miles' is usually in hundreds, w_1 was chosen as 1/100 to reduce the overall value of smoke estimate. Similarly, since 'Distance_Square' was a very high value usually in the hundred thousands, w_2 was chosen as 1/100000 to curb the digits. Thus, the final smoke estimate looked as follows,

$$Smoke\ Estimate = 0.01 \times Area\ of\ Fire \times \exp(-10^{-5} \times Distance^2)$$

Using the above formula, the smoke estimate was calculated for every fire in the dataset.

Analyzing the Smoke Estimate

The next step was to understand how the smoke estimate for each fire would be used to find one smoke estimate for every year. This could have been done by taking the cumulative smoke estimate during each year or by amortizing over the fire season (May 1st to October 31st). To make this decision, the total number of wildfires were plotted for every year.

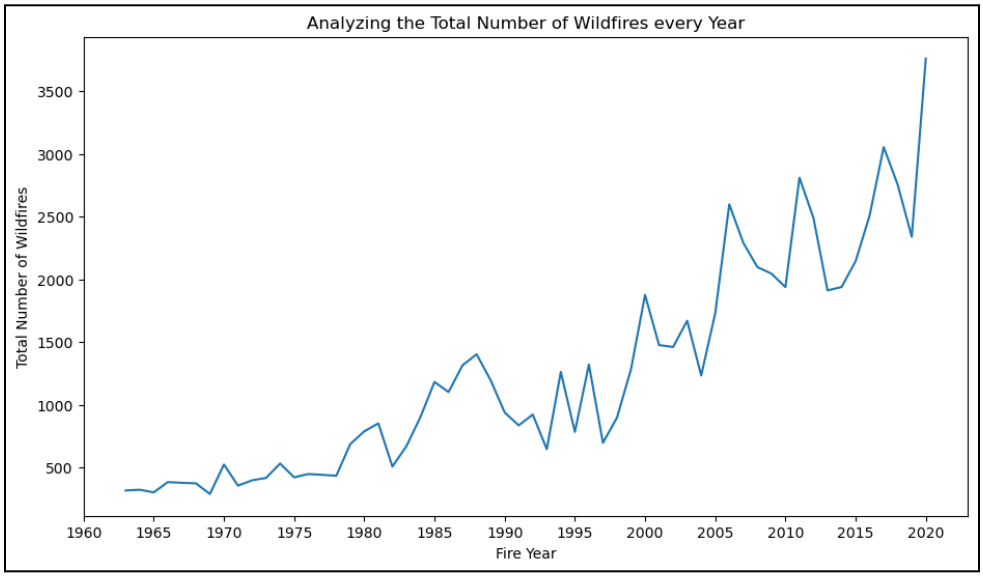

Fig. 3. Analyzing the Total Number of Wildfires every Year

From the above figure (Fig. 3), it was found that the total number of wildfires had increased over time. This change was quite significant from 1963 to 2020. This could also be attributed to the lack of proper reporting measures earlier, but either way, taking an average of the smoke estimate would not account for the total number of wildfires. Since the "smoke emitted" was being analyzed, it made more sense if a cumulative of the smoke estimate was taken. This would take the large number of wildfires in recent times into account and not provide any biased results.

Evaluating the Smoke Estimate's Performance Using US EPA Air Quality Data

The next step was to validate the smoke estimate by comparing it with the AQI data.

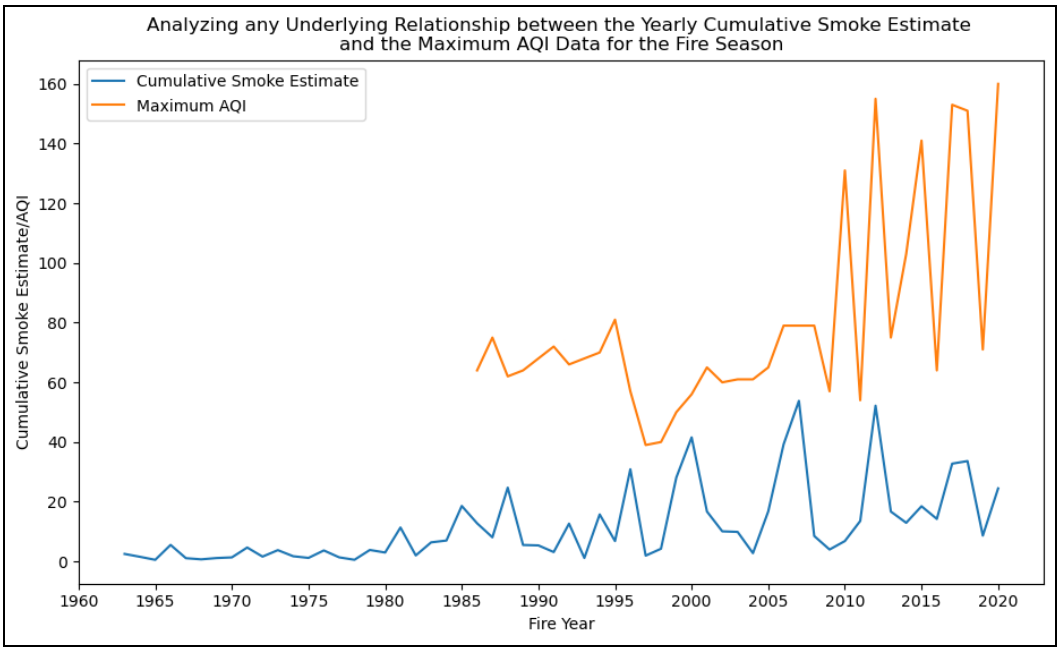

Fig. 4. Validating the Yearly Cumulative Smoke Estimate and Maximum AQI Data for the Fire Season

As evident from the above plot (Fig. 4), the Maximum AQI data for the fire season of every year is missing for the years 1963-1985. However, for the years where the data is available, there seemed to be a somewhat dependent relationship between the two variables - Cumulative Smoke Estimate and the Maximum AQI. The peaks and dips were mostly in sync for both the curves. However, there were a few instances where these didn't match.

The AQI measures air pollution levels, which can arise from various sources, not solely wildfires. While wildfires are a significant contributor to poor air quality and can substantially impact the AQI, other factors and sources also influence air pollution levels. Some of these other sources are industrial emissions, vehicle emissions, construction activities, etc. Thus, while there might be a relationship between both the variables we cannot assume that wildfires are the sole reason for poor AQI. Overall, the curves were quite similar in their trends indicating that our smoke estimate was capturing the data well.

3.1.4 Modeling and Predicting the Smoke Estimate

The plot of the Cumulative Smoke Estimate showed continuous peaks and falls which were difficult to capture using a simple linear regression model. Several models were considered initially to model the smoke estimate including polynomial regression, modeling using a sine function, etc. However, the polynomial regression model was unable to grasp all the ups and

downs of the graph. The model accuracy was found to be very low. Similarly, the sine function which did capture the data better than the regression model, still could not capture the increasing trend of the overall curve. The ups and downs generated from this curve were constant and not increasing with time.

Finally, the moving average model was chosen since it smoothes out short-term fluctuations, making underlying trends more apparent. It captured the ups and downs of the time-series data really well. This model was also easier to implement and interpret, making it feasible for this analysis. The AQI data could also have been added as a feature in this model. However, since the AQI data was missing for more than 20 years, only the smoke estimate was used to compute the rolling average. Thus, the final model that was used for this analysis was the Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX) model.

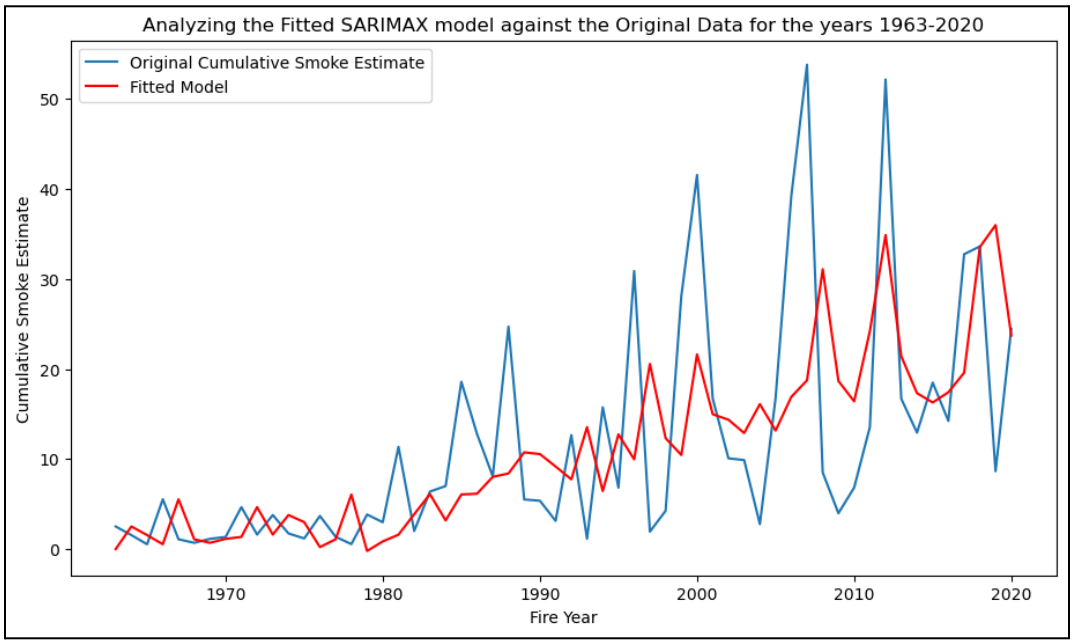

Fig. 5. Comparing the Fitted SARIMAX model with the Original Data

From Fig. 5, it can be seen that the model does not fully capture the data for the current period, and there is quite some scope for improvement. However, overfitting of the data was something that had to be avoided. Thus, this model was chosen and predictions were generated for the smoke estimate for every year for the next 29 years (i.e., 2021-2049). While the study took place in the year 2023, predictions were generated from 2021 onwards since the Smoke Estimate data was available until 2020 only.

3.2 Analyzing Wildfires and Socio-Economic Indicators

The goal of this part of the analysis was to plot the four socio-economic indicators with the smoke estimate to see if any correlation exists. For those indicators where correlation exists, the next step would be to model them and generate predictions until 2049 to see the impact of the smoke from the future wildfires on the regional economy.

All the four datasets - GDP, unemployment rate, personal income per capita and income inequality were mostly clean. The only preprocessing task that had to be performed in these datasets was converting the date to a datetime object and extracting the year, after which plots were plotted to check for correlation with the smoke estimate.

For those indicators where correlation did exist, predictions were generated using the moving average model, specifically the SARIMAX model. This model was chosen because it can handle seasonal patterns in data, making it suitable for time series with recurring patterns at fixed intervals. SARIMAX combines autoregressive (AR) and moving average (MA) components to capture dependencies between observations and the impact of previous error terms, which can be vital in time series analysis. It includes differencing to make a time series stationary, which can help stabilize variance and make data more amenable to modeling. Thus, this model is effective for making future predictions based on historical patterns and other factors.

4. Findings

This section discusses each of the plots, research questions, and predictions obtained in detail. The aim here is to analyze each of the results in detail and draw possible explanations and conclusions from the observations of the plots.

4.1 Initial Exploratory Analysis

Distribution of Wildfires by their Distance from Twin Falls, Idaho

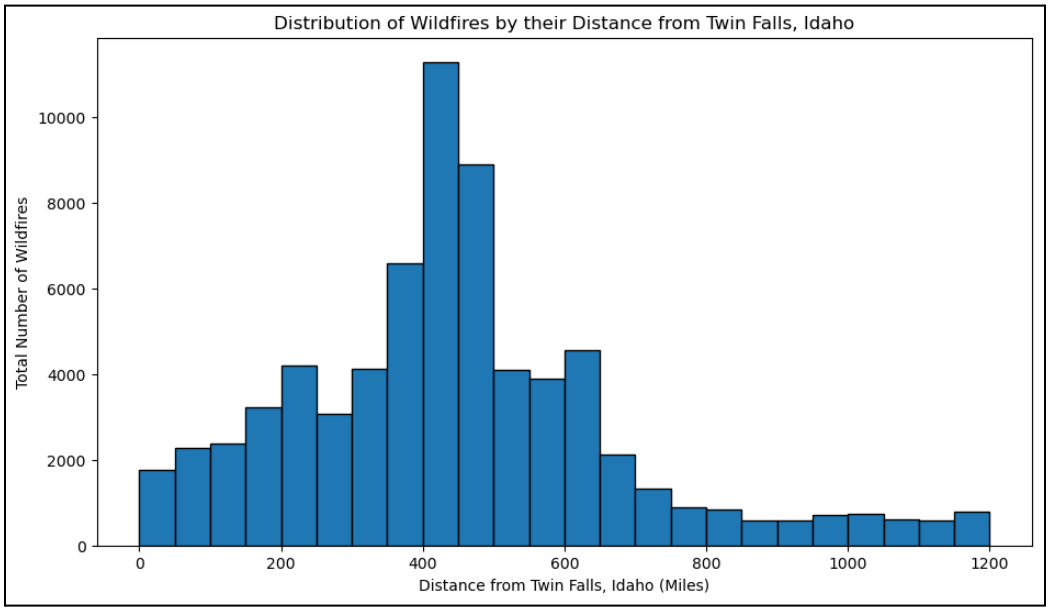

Fig. 6. Histogram of Distribution of Wildfires by their Distance from Twin Falls, Idaho

The above plot (Fig. 6) shows the distribution of the wildfires as a histogram for the total wildfires occurring every 50 miles distance. In essence, each bar covers the total wildfires within 50 miles buckets. To help read the figure, consider the tallest bar in the histogram which corresponds to the bucket of 400-450 miles. It was found that most wildfires are present in the 400-450 miles radius of Twin Falls, Idaho. The x-axis represents the Distance from Twin Falls, Idaho. Note that for this analysis, only the wildfires within the 1250 miles radius of Twin Falls, Idaho were considered. The y-axis represents the Total Number of Wildfires for a particular bucket of distance.

Total Acres Burned by Year

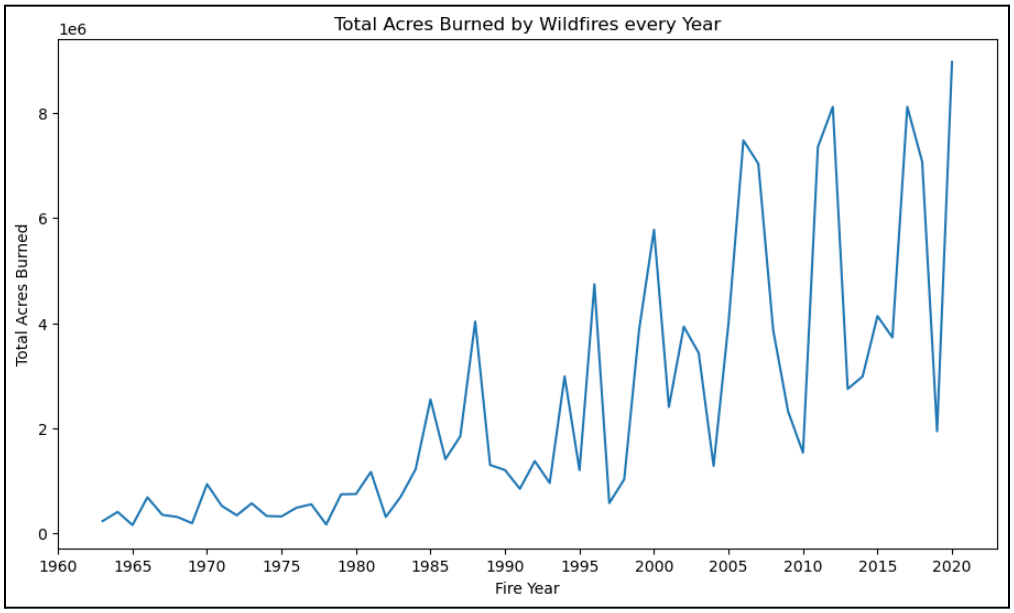

Fig. 7. Total Acres Burned by Wildfires every Year

The above figure (Fig. 7) shows the time series graph of the total acres burned by wildfires per year for the fires occurring within 1250 miles of Twin Falls, Idaho from 1963 to 2020. To read the figure, one can pick any point on the line chart. The corresponding value on the x-axis represents the year and the corresponding value on the y-axis represents the total acres that were burned. Thus, the x-axis represents the year in which the wildfires occurred, and the y-axis represents the total acres that were burned by those wildfires. It was found that the total acres burned go up and down periodically. However, the peaks and the dips steadily increase every time indicating that although the total burned acres do go down at times, the average value of the total burned acres over time has been increasing.

4.2 Smoke Estimate and its Predictions

The smoke estimate data extracted for the wildfires within 1250 miles of Twin Falls, Idaho for the period 1963-2020 were used to generate predictions from 2021 through 2049. The predictions were generated with confidence intervals to indicate the uncertainty.

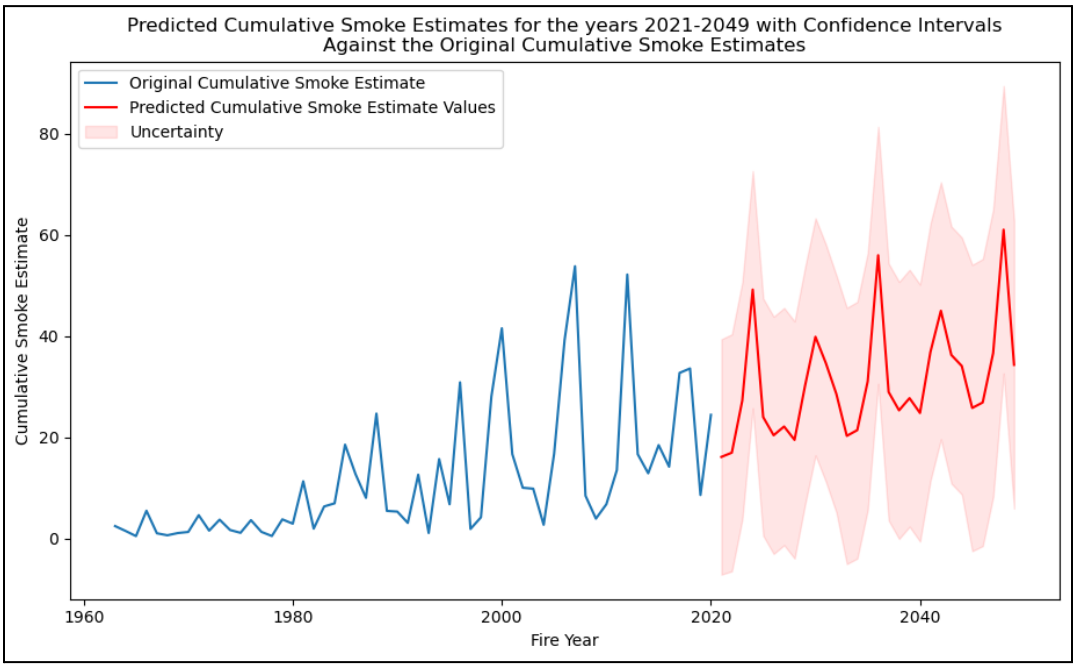

Fig. 8. Yearly Cumulative Smoke Estimate and its Predictions with Confidence Intervals

The above figure (Fig. 8) shows the time series graph for the smoke estimate as well as its predictions. To read the figure, one can pick any point on the line chart. The corresponding value on the x-axis represents the year and the corresponding value on the y-axis represents the smoke estimate. Thus, the x-axis represents the year in which the wildfires occurred, and the y-axis represents the cumulative smoke estimate. From the plot above, there seems to be some sort of periodicity existing. The smoke estimate goes up at a peak every 3-5 years. This periodicity is found to have been maintained in the predictions as well. The height of the peaks alternatively goes up and down with an overall increasing trend. This indicates that the smoke estimate is bound to increase for the next thirty years. This is concerning because some of the peak values are an all-time high indicating that the situation is just getting worse in the future.

4.3 Correlation between the Socio-Economic Indicators and the Smoke Estimate

4.3.1 GDP

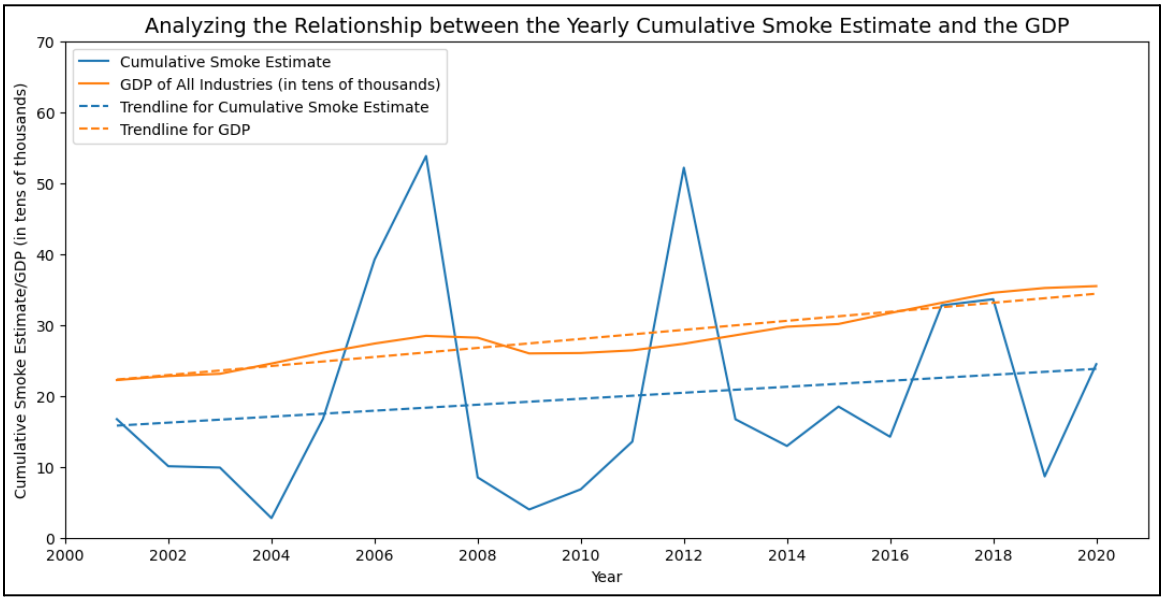

Fig. 9. Analyzing the Relationship between the Yearly Cumulative Smoke Estimate and the GDP

The plot above (Fig. 9) shows the time series graph for the smoke estimate and the GDP along with their trendlines. The x-axis represents the year in which the wildfires occurred, and the y-axis represents the cumulative smoke estimate and the GDP (in tens of thousands). While the wildfires data was available for the period 1963-2020, the GDP data was available only from 2001 to 2021. Thus, the intersection of the two time periods was considered while plotting the above graph, i.e. 2001 through 2020. It can be observed from the plot that the overall trend of the smoke estimate and the GDP is increasing. However, the correlation coefficient of these two variables was found to be very low (0.26) indicating that no strong correlation exists between the yearly smoke estimate and the GDP. Thus, there is no evidence that economic productivity experiences a dip for the years having heavy smoke.

4.3.2 Unemployment Rate

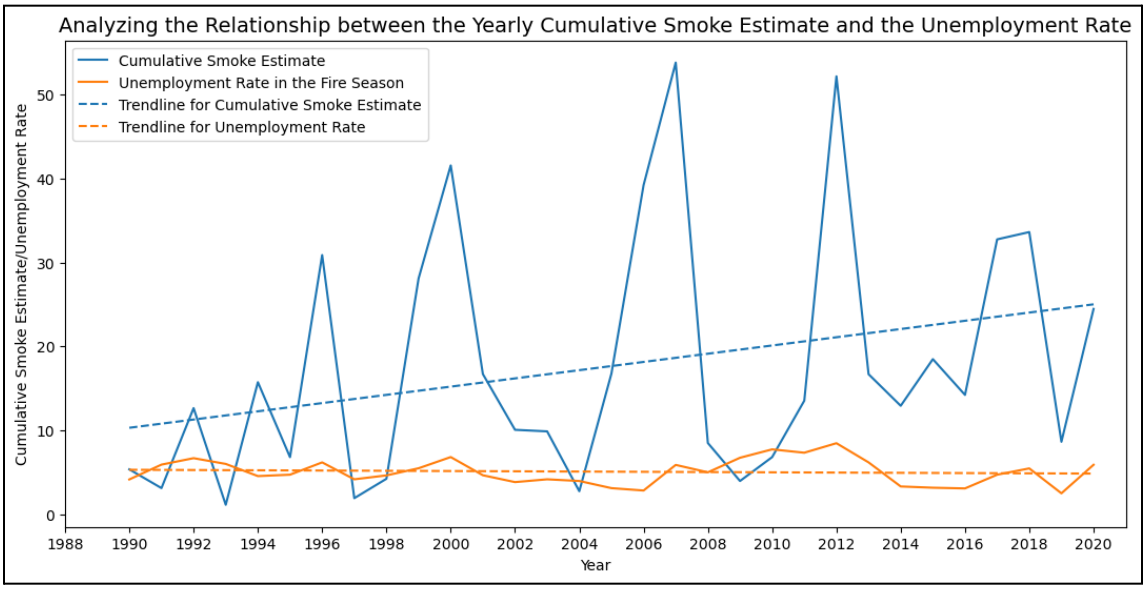

Fig. 10. Analyzing the Relationship between the Smoke Estimate and the Unemployment Rate

In the plot above (Fig. 10), the time series graph for the smoke estimate and the unemployment rate is shown along with their trendlines. The x-axis represents the year in which the wildfires occurred, and the y-axis represents the cumulative smoke estimate and the unemployment rate. While the wildfires data was available for the period 1963-2020, the unemployment data was available only from 1990 to 2023. Thus, the intersection of the two time periods was considered while plotting the above graph, i.e. 1990 through 2020. It can be observed from the plot that the overall trend of the smoke estimate is increasing while that of the unemployment rate is decreasing. However, the correlation coefficient of these two variables was found to be reasonably high (0.46) indicating that a strong positive correlation exists between the yearly smoke estimate and the unemployment rate. In other words, the unemployment rate increases when smoke estimate increases, and vice versa.

Some possible explanations could be that wildfires and heavy smoke might force businesses to temporarily or permanently shut down leading to layoffs and higher unemployment rates. Additionally, areas affected by wildfires often experience reduced tourism due to safety concerns, impacting jobs in the hospitality and service industries. While the correlation between wildfire smoke estimate and the unemployment rate might not imply causation, it does highlight how these two factors could be interconnected through various economic disruptions and challenges faced during wildfires. The relationship could be influenced by the severity, duration, geographical reach of the wildfires, as well as the recovery measures implemented.

4.3.3 Personal Income per Capita

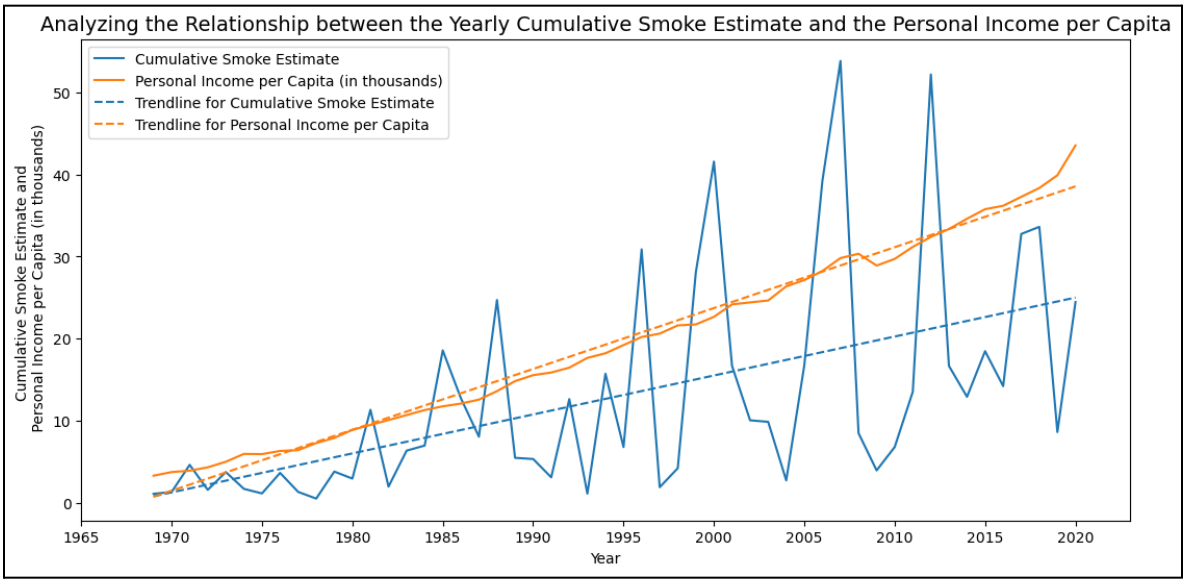

Fig. 11. Analyzing the Relationship between the Smoke Estimate and the Personal Income per Capita

The plot above (Fig. 11) shows the time series graph for the smoke estimate and the personal income per capita (in thousands) along with their trendlines. The x-axis represents the year in which the wildfires occurred, and the y-axis represents the cumulative smoke estimate and the personal income per capita (in thousands). While the wildfires data was available for the period 1963-2020, the personal income per capita data was available from 1969 to 2022. Thus, the intersection of the two time periods was considered while plotting the above graph, i.e. 1969 through 2020. It can be observed from the plot that the overall trend of the smoke estimate and the personal income per capita is increasing. However, the correlation coefficient of these two variables was found to be very low (0.23) indicating that no strong correlation exists between the yearly smoke estimate and the personal income per capita. Thus, there is no evidence that years with higher smoke exposure experience declines in personal income.

4.3.4 Income Inequality

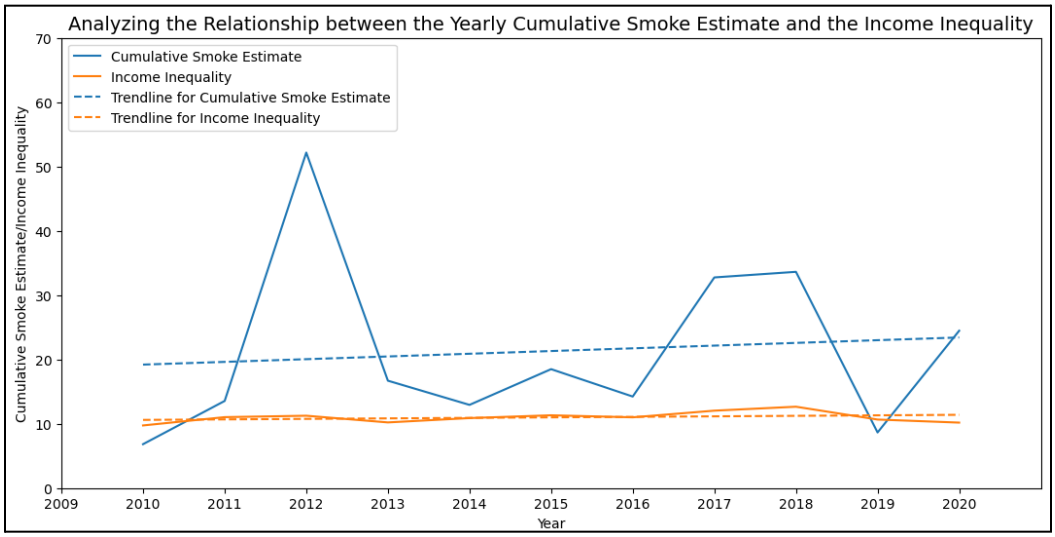

Fig. 12. Analyzing the Relationship between the Smoke Estimate and the Income Inequality

The figure above (Fig. 12) shows the time series graph for the smoke estimate and the income inequality along with their trendlines. The x-axis represents the year in which the wildfires occurred, and the y-axis represents the cumulative smoke estimate and the income inequality. While the wildfires data was available for the period 1963-2020, the income inequality data was available from 2010 to 2021. Thus, the intersection of the two time periods was considered while plotting the above graph, i.e. 2010 through 2020. It is difficult to observe the trend of the income inequality in the above plot since the changes are small and gradual. The plot was thus rebuilt using the square of the income inequality instead, to understand the underlying trend.

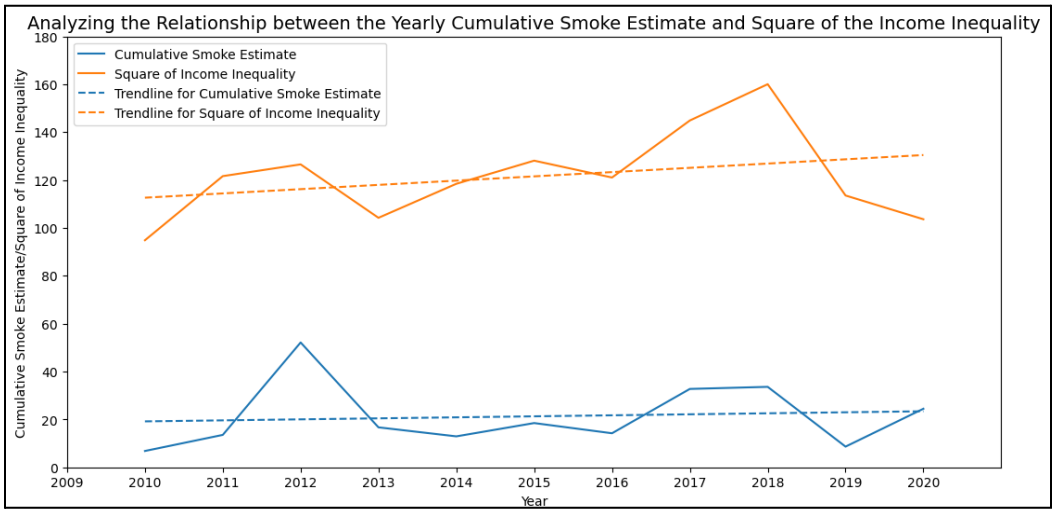

Fig. 13. Analyzing the Relationship between the Smoke Estimate and the Square of the Income Inequality

The trends in income inequality are more evident in the plot above (Fig. 13). It can be observed from the above plot that the overall trend of the smoke estimate and income inequality is increasing. The correlation coefficient of these two variables was also found to be high (0.56) indicating that a strong positive correlation exists between the yearly smoke estimate and the income inequality. In other words, the income inequality increases when the smoke estimate increases, and vice versa.

Some possible explanations could be that individuals in lower-income brackets are more likely to work in sectors or jobs susceptible to wildfire-related disruptions, such as agriculture, outdoor labor, or service industries, leading to job instability or job losses. They also might be disproportionately affected by wildfire smoke due to factors like limited access to healthcare, inability to relocate during wildfires, or living in areas more exposed to smoke. While the correlation between wildfire smoke estimate and the income inequality might not imply causation, it does highlight the complex socioeconomic dynamics following natural disasters. It emphasizes how vulnerable populations or lower-income groups often bear a burden of the impact, facing challenges in income stability, recovery, and resilience. Understanding these dynamics is crucial in designing targeted policies and interventions to address income disparities and support equitable recovery for all affected communities.

4.4 Predictions for Socio-economic Indicators with the Smoke Estimate

Since a reasonably high to strong correlation was observed for the indicators - unemployment rate and income inequality, predictions were generated for these variables for the years 2021 through 2049. These predictions were plotted against the predicted smoke estimate to understand the future impact of wildfires on Twin Falls' socio-economy.

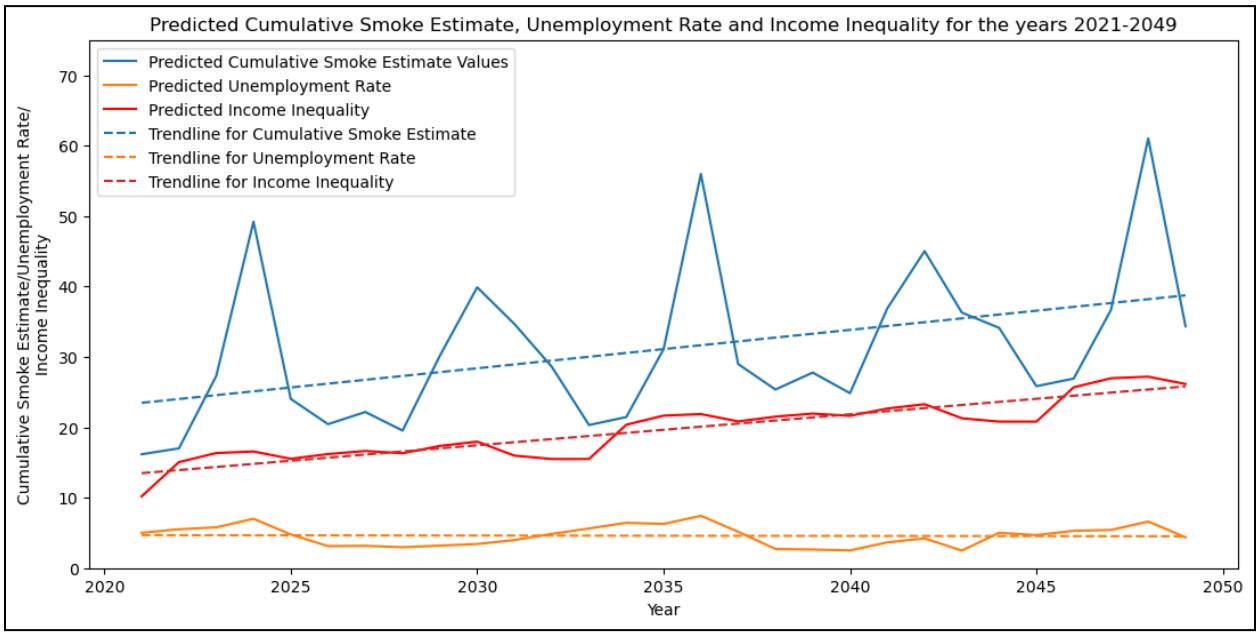

Fig. 14. Generating Predictions for Smoke Estimate, Unemployment rate and Income Inequality from 2021 to 2049

The figure above (Fig. 14) shows the time series graph for the predicted values of smoke estimate, unemployment rate, and the income inequality along with their trendlines for the years 2021 through 2049. The x-axis represents the year in which the wildfires occurred, and the y-axis represents the cumulative smoke estimate, unemployment rate, and income inequality. It can be observed that the income inequality and the smoke estimate have an increasing trend indicating that the situation in the future is going to worsen. The unemployment rate however, seems to be stable with a few ups and downs.

5. Discussion/Implications

From the findings in the previous section, the smoke estimate is predicted to overall increase with the income inequality. Moreover, a strong correlation was found to exist between the smoke estimate and income inequality. Identifying a projected increase in both wildfire smoke and income inequality signals potential challenges for already vulnerable communities. The wildfire smoke is impacting the livelihoods of the non-affluent section of the society. With escalating wildfire smoke, the vulnerable populations with limited resources are being affected due to increased healthcare expenses or job loss. Thus, the rich are becoming richer and the poor are becoming poorer. This increase in income inequality might exacerbate the economic situation of certain demographic groups, potentially amplifying the disparities in the aftermath of wildfires.

The implications of these findings extend beyond mere statistical insights; they underscore pressing concerns that demand immediate attention from local governance, civic leaders, and residents alike. Thus, the goal of this study is not just to uncover correlations or statistical relationships but to translate these findings into human-centric actionable solutions that improve the lives of community members.

The decision-making process in this project was deeply rooted in human-centered data science principles. By centering on community impact, equity, and inclusivity, the analysis aimed to address real-world challenges faced by individuals and communities. The findings guided by these principles seek not only to elucidate statistical relationships but also to empower meaningful action that prioritizes the well-being of all residents. The analysis ensures inclusivity by considering diverse perspectives, ensuring that findings are representative and considerate of all segments of the community, especially vulnerable or marginalized groups. Fairness in decision-making was ensured by focusing on equitable outcomes, aiming to reduce disparities and prioritize interventions that benefit the entire community, irrespective of socioeconomic status or background.

Action Plan

To address the issues discussed earlier, the city council must develop and implement policies focusing on mitigating the impact of wildfire smoke, prioritizing the health of affected communities, and ensuring equitable distribution of resources and assistance. This can be accomplished by fostering collaborations with relevant agencies, community organizations, and experts to strategize disaster preparedness, response, and recovery plans that prioritize inclusivity

and equity. Thus, the mayor and civic leaders must advocate for equitable resource allocation, raise awareness about health risks associated with wildfire smoke exposure, and champion initiatives to address income disparities within the community. This can be done by engaging with residents, encouraging community involvement in resilience-building efforts, and offering support mechanisms for affected individuals and families.

The city residents can also do their part in case of a wildfire. They must stay informed and engaged in local initiatives, participate in disaster preparedness programs, and support neighbors, especially vulnerable groups, during wildfire events. Adaptive measures need to be adopted at the individual level, such as improving home air quality, creating emergency plans, and accessing available support services.

Timeline

The urgency to act upon these findings cannot be overstated. Based on the predictions in Fig. 14, the next big jump in wildfires smoke estimate is predicted to be in the year 2030. Thus, immediate steps should commence within the upcoming fiscal year to develop comprehensive strategies and policies. To prioritize human-centered data science principles, the approach must be iterative, allowing for continuous learning, adaptation, and improvement of strategies based on feedback from the community and the evolving nature of challenges. This will allow for continual review and adaptation of these plans over the next seven years to ensure their relevance and effectiveness.

6. Limitations

While the study was conducted with appropriate assumptions and keeping human-centered data science principles in mind, it does come with its own set of limitations. Acknowledging these limitations is crucial in interpreting the study's findings appropriately, highlighting areas for further research and informing the development of more robust methodologies to capture the true impact of wildfires on regional economies.

Limitations and Assumptions in Data

1. It is important to note that the wildfires dataset may not be the most accurate especially for wildfires in the earlier years due to lack of proper reporting measures. Wildfires might have been underreported or not systematically documented in earlier periods, especially in remote or less populated areas, leading to incomplete or inaccurate datasets.

2. Since fires are irregularly shaped, defining the notion of 'distance' of the fire from Twin Falls, Idaho was challenging. For example, should the distance be calculated from the closest point of the fire perimeter or the centroid of the region? For this study, the average distance of all perimeter points was taken because for some very large fires, including just the shortest distance seemed biased. However, this is an assumption that can affect the ultimate results since the 'Distance' variable factors into the final smoke estimate.

3. The wildfires dataset had the data for both prescribed fires and wildfires. A prescribed fire is a planned fire intentionally ignited by park managers to meet management objectives whereas a wildfire is unplanned caused by natural causes, by accidental (or arson-caused) human ignitions, or by an escaped prescribed fire. While prescribed fires are intentional and usually in control, they still do contribute to air pollution. For this analysis, it was assumed that prescribed fires and wildfires contribute to the same amount of pollution for a given land within the same area. This assumption of equal impact could have led to inaccurate estimations of pollution levels and can affect the overall results.

4. Overlapping fires were removed from the dataset since there is another fire already existing in the database with more than 10% overlap. While the overlap flag may or may not be correct, it is assumed that another row pertaining to the same fire exists for those fires that are flagged. This assumption however holds to be valid, only if the dataset was accurate in terms of documenting the overlapping fires. There is a possibility of overlapping fires still existing in the dataset or "non-overlapping" fires being removed.

5. The AQI data obtained from the US EPA had a few missing values. While these values were dealt by taking the rolling average of the previous five years which makes for a reasonable assumption, it still does create a caveat.

6. The data available for the four socio-economic indicators was widely inconsistent in terms of the timelines for which it was available. While some indicators had a longer period of data, the income inequality dataset had data only for 10 years. Thus, generalizing the relation between such indicators with the smoke estimate was difficult. However, for the scope of this assignment, this factor was neglected. Any correlation existing, even for a 10 year period, was taken into account.

Modeling Assumptions

1. Since the smoke estimate was modeled using the data in-hand which relied on subjective observations or proxy indicators rather than direct measurements of particulate matter or specific pollutants, this could potentially introduce inaccuracies. While the smoke estimate's correlation with AQI data is a positive indicator, these limitations underscore the need for cautious interpretation and further validation.

2. The SARIMAX model was used to model the smoke estimate, unemployment rate and income inequality. It was also used to generate predictions until 2049. However, this model has several assumptions and limitations:
    - It assumes linear relationships between variables, which might not fully capture the complex, nonlinear dynamics of smoke estimate data.
    - The model might be sensitive to outliers or unusual events not accounted for, impacting its predictive ability. This in turn leads to problems when forecasting long-term trends, especially if the data exhibits nonlinear or irregular patterns.

3. Finally, establishing a direct cause-and-effect relationship between wildfires and economic factors is complex. Broader economic trends, market fluctuations, or global events can influence regional economies independently of wildfire impacts, making it challenging to isolate the exact impact of wildfires. Changes in governmental policies, emergency response strategies, or economic interventions during or after wildfires also might influence the economic outcomes observed. Thus, for this analysis, it was assumed that if strong correlation between an indicator and smoke estimate exists for two or more indicators, then the wildfires do have an impact on the socio-economy of Twin Falls, Idaho.

7. Conclusion

In conclusion, the first part of the study focused on studying the wildland fires within 1250 miles of the city of Twin Falls, Idaho for the last 60 years (1963-2020). A smoke estimate was then created to estimate the wildfire smoke impact which was later modeled to make predictions for the next 30 years (until 2049). The second part of the project further extended this analysis to the regional socio-economy, specifically to four indicators - GDP, unemployment rate, personal income per capita, and income inequality. The analysis focused on four research questions,

1. Do the years affected by heavy smoke experience dips in economic productivity? The aim was to assess the impact on GDP growth during periods of intense wildfires.
2. Do unemployment rates increase during and immediately after wildfire events due to disruptions in economic activities? The goal was to analyze how wildfires affect unemployment rates in periods experiencing severe smoke exposure.
3. Did years with higher smoke exposure experience declines in personal income? The goal was to investigate how wildfires influence personal income per capita.
4. Finally, do low-income groups suffer disproportionately due to wildfires' economic impacts? The aim was to explore if income inequality exacerbates the economic vulnerability of low-income groups during and after wildfires.

From the findings, it was found that the majority of the wildfires are within the 400-500 miles radius of Twin Falls, Idaho where the total acres burnt for the last 60 years has overall been increasing. The smoke estimate was found to have a periodicity of 3-5 years where the size of the peaks were also gradually increasing.

When the smoke estimate was plotted with the four chosen indicators, two indicators - unemployment rate and income inequality, were found to have reasonably high to strong correlation. Thus based on the research questions, it was concluded that:

1. There is no evidence that economic productivity experiences a dip for the years having heavy smoke.
2. Unemployment rates were found to increase during and immediately after wildfire events and decrease when the smoke estimate was low.

3. There is no evidence that years with higher smoke exposure experience declines in personal income.

4. Income inequality was found to increase during and immediately after wildfire events and decrease when the smoke estimate was low. Low-income groups thus suffer disproportionately due to wildfires' economic impacts.

When the predictions were generated for the two indicators - unemployment rate and income inequality, with the smoke estimate, it was found that the income inequality and smoke estimate had an increasing trend until 2049. The unemployment rate, on the other hand, was found to be somewhat stable.

The implications of these findings extend beyond mere statistical insights; they underscore pressing concerns that demand immediate attention from local governance, civic leaders, and residents alike. The wildfire smoke is impacting the livelihoods of the non-affluent section of the society. With escalating wildfire smoke, the vulnerable populations with limited resources are being affected due to increased healthcare expenses or job loss. This increase in income inequality might exacerbate the economic situation of certain demographic groups, potentially amplifying the disparities in the aftermath of wildfires.

The council thus needs to take immediate action by implementing policies that address climate change and help mitigate the frequency and severity of wildfires. Targeted support and employment opportunities to low-income areas and marginalized groups can also be provided. The urgency to act upon these findings cannot be overstated. Based on the predictions, the next big jump in wildfires smoke estimate is predicted to be in the year 2030. Thus, immediate steps should commence within the upcoming fiscal year to develop comprehensive strategies and policies.

This study thus demonstrates the application of human-centered data science principles in addressing real-world challenges. By analyzing the impact of wildfires on socio-economic factors, the study prioritized the following principles:

- The study focused on community impact and how data insights affect individuals, emphasizing the practical implications of wildfires on regional economies.

- Keeping equity considerations in mind, the analysis highlighted how vulnerable populations are disproportionately affected by wildfires, thereby informing policies aimed at reducing disparities.

- Findings and methodologies are transparently communicated to stakeholders, fostering trust and accountability in the decision-making process.
- The approach is iterative, allowing for continuous learning, adaptation, and improvement of strategies based on feedback from the community and the evolving nature of challenges.

By employing these principles, the study aims to empower communities, inform policy decisions, and contribute to the creation of targeted interventions. It emphasizes the need for data-driven, community-centered approaches to address the complex challenges posed by natural disasters like wildfires, ultimately striving for equitable and resilient communities.

8. References

[1] Wang, D., Guan, D., Zhu, S., Kinnon, M. M., Geng, G., Zhang, Q., Zheng, H., Lei, T., Shao, S., Gong, P., & Davis, S. J. (2020). *Economic footprint of California wildfires in 2018*. Nature Sustainability, 4, 1–9. https://doi.org/10.1038/s41893-020-00646-7

[2] Meier, S., Elliott, R. J. R., & Strobl, E. (2023). *The regional economic impact of wildfires: Evidence from Southern Europe*. Journal of Environmental Economics and Management, 118, 102787. https://doi.org/10.1016/j.jeem.2023.102787

[3] *WDI - Economy*. (n.d.). Datatopics.worldbank.org. https://datatopics.worldbank.org/world-development-indicators/themes/economy

[4] *Centers for Disease Control and Prevention*. (2023, September 1). Socioeconomic factors | CDC. Centers for Disease Control and Prevention. https://www.cdc.gov/dhdsp/health_equity/socioeconomic

[5] McDonald D., *wildfire.zip*. (n.d.). Google Drive. Retrieved November 1, 2023, from https://drive.google.com/file

[6] McDonald D., *wildfire_geo_proximity_example.ipynb*. (n.d.). Google Drive. Retrieved November 1, 2023, from https://drive.google.com/file

[7] McDonald D., *epa_air_quality_history_example.ipynb*. (n.d.). Google Drive. Retrieved November 1, 2023, from https://drive.google.com/file

9. Data Sources

---

[1] *Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons) - ScienceBase-Catalog*. (n.d.). www.sciencebase.gov. https://www.sciencebase.gov/catalog/item

[2] U.S. Bureau of Economic Analysis. (2001, January 1). *Real Gross Domestic Product: All Industries in Twin Falls County, ID*. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/REALGDPALL

[3] U.S. Bureau of Labor Statistics. (1990, January 1). *Unemployment Rate in Twin Falls County, ID*. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/IDTWIN

[4] U.S. Bureau of Economic Analysis. (1969, January 1). *Per Capita Personal Income in Twin Falls County, ID*. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/PCPI

[5] U.S. Census Bureau. (2010, January 1). *Income Inequality in Twin Falls County, ID*. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/2020RATIO