Montana's Economic Resilience: A Deep Dive into Wildfire Effects and Forecasts

Evan Yip

1. Introduction

On August 21, 2023, the U.S. Department of the Interior (DOI), which is responsible for the management and conservation of federal lands and natural resources, released a report showing that visitor spending near communities near national parks contributed $50.3 billion nationwide and supported nearly 380,000 jobs. The state of Montana is home to some of the most widely revered national parks such as Yellowstone National Park, Glacier National Park and Bighorn Canyon. In 2022 alone, nearly 5 million park visitors contributed an astonishing $621 million dollars to the Montana economy. According to the Montana Department of Labor and Industry, Montana is “2nd in the nation for outdoor recreation production as percent of total GDP”, with outdoor recreation generating $2.5 billion of annual gross domestic product (GDP) accounting for 4.4% of total GDP in 2021. Missoula, as the second most populous city in Montana, is strategically positioned to capitalize on these trends. Not only does it offer convenient access to Glacier and Yellowstone National Parks, but it also serves as a regional hub for shopping, dining, and arts and culture. As a result, Missoula is poised to mirror the economic dynamics seen in other communities near national parks, with the potential for increased visitor spending and job support in the foreseeable future.

While it is evident that these economies heavily rely on visitors attracted to outdoor recreation at national parks, it's essential to highlight their susceptibility. The dependence on outdoor recreation and tourism makes Missoula particularly vulnerable to factors that can impact these sectors. Any sort of factor that alters the landscape, wildlife, or experience of visitors at these national parks has a high potential of altering the economy. Wildfires serve as a quintessential example of an external force that can profoundly impact the state's economic stability. The destructive nature of wildfires not only jeopardizes the natural beauty of the parks but also disrupts tourism flow by significantly diminishing air quality through wildfire smoke. According to a study conducted by Utah State University, authored by Man-Keun Kim and Paul M. Jakus, the findings reveal that wildfires in Utah exert detrimental and significant impacts on visitation in four out of the five national parks. These consequences are reflected in economic losses ranging between $2.7 and $4.5 million dollars. This deterioration in air quality resulting in decreased visitation rates to the national parks presents a substantial threat to Montana's economic well-being.

Given these circumstances, it is clear that city council members and residents of Missoula, Montana need to be equipped with the knowledge of the relationships between the economy and changes due to wildfires so that they assess to what degree they need to implement strategic measures for economic resilience. To meet this need, the aim of this study is to perform a retrospective statistical analysis to assess the degree at which Montana’s economy is affected by factors such as air quality, smoke, and national park visitation. The developed statistical models and analysis will not only provide insights into the past, but also serve as invaluable tools for predicting potential future scenarios.

2. Background and Related Work

2.1 Literature Review

Numerous studies have explored the intricate relationships among wildfires, national park visitation, and regional economies. Notably, Michael Kernan's extensive exploration, specifically his study titled "Modeling and Forecasting Glacier National Park Visitation" at the University of Montana, stands out as a comprehensive endeavor in this field. Kernan employs a combination of regression and multivariate time series models to probe into the impacts of wildfires and precipitation on Glacier National Park visitation. Various aspects of Kernan's methodology served as inspiration for designing the system of our study. Regarding datasets, Kernan curated a diverse collection, encompassing national park visitation, wildfire, precipitation, air quality, macroeconomic, and gasoline price data. The national park visitation data, in particular, was sourced from the Integrated Resource Management (IRMA) portal of the National Park Service. Kernan's choice and results inspired me to also leverage this dataset for this study. An inherent challenge in crafting statistical models involves feature engineering to effectively capture the variations and trends in the data. Addressing this challenge, we adopted a design choice inspired by Kernan—specifically, the development of AQI features. Kernan transformed daily AQI measurements into two monthly aggregate values, by counting the number of days in that month where the AQI exceeded 75 and 100. Though the study lacks thorough explanation of why this value is chosen, upon inspection of the U.S. Environmental Protection Agency (EPA) we see that the AQI value of 75 is situated right in the center of the 'Moderate health concern' category and the AQI value of 100 is the lower bound of the 'Unhealthy for Sensitive Groups' range. After feature engineering, Kernan next evaluated correlations between these variables and to identify and take actions to address multicollinearity in the predictor variables. Kernan then developed a Ordinary Least Squares (OLS) model based on AQI, precipitation counts, unemployment rate, gasoline prices lagged one month, and other features. The results of this model highlighted that AQI, unemployment, gasoline prices and precipitation were all statistically significant predictors of national park visitation. The final section of Kernans methods involved using exponential smoothing (ETS) and autoregressive-integrated-moving average models (ARIMA) to develop predictions for Glacier National Park visitations. While these particular models were not incorporated into our final design, Kernan's work paves the way for future exploration of more complex statistical models.

Kernan's work laid a solid foundation for understanding the dynamics between wildfires, park visitation, and economic indicators of Glacier National Park. From his work, we gained valuable abstractions of AQI measurements, correlation analysis frameworks, and clear examples of time-series regression analysis. This study seeks to build and expand upon his work by widening the scope of the analysis to incorporate statewide impacts of four separate national parks, wildfires within a 1250 mile radius of Missoula, and multiple AQI measurement sources to predict the economic impacts at a statewide scale.

Another study that was pivotal for the design of this analysis was Man-Keun Kim et. al's manuscript on the effects of wildfires, gas prices, economic recessions, and national park visitation on changes in Utah's economy. Much like Kernan's approach, the team from Utah State University integrated both park visitation data from Utah's five national parks and information on wildfire activity. In a manner akin to my decision to incorporate AQI abstractions from Kernan's work, I also opted to include Kim et al.'s wildfire feature in estimating wildfire smoke effects. The researchers aggregated daily wildfire data by transforming it into a single variable representing the total burned area per month, providing a comprehensive measure of wildfire activity.

Though this study also employed OLS models to predict park visitation, it distinguished itself from Kernan's studies by directly quantifying the effects wildfires and other features had on southern Utah's economy. Implementing a multi-step bootstrapping process to generate an empirical distribution of wildfire-induced visitations, Kim et al. combined this information with expenditure profiles from the US NPS in 2016 and IMPLAN input-output models to measure the impact of changes in visitations on local economies. Using the IMPLAN models they defined the total economic effect of national park visitation in terms of lodging, camping, groceries, transportation, retail, goods and services as well as income changes. These services were incorporated into my GDP feature where I attempted to select the private industries that best captured these effects. I felt confident in adhering to their definition since they were able to show that the "regional economic impacts of seasonal wildfires at all national parks in Utah are estimated to be between $2.7 and $4.5 million.

2.2 System Design and Data sources

The methodology and outcomes of both referenced studies have been instrumental in refining the system design for this investigation. They have not only offered valuable insights into data sources, feature engineering, correlation analysis, and regression modeling but have also showcased effective methods for predicting economic impacts using wildfire data, AQI data, and national park visitation data. These studies, thus, serve as an indispensable foundational template, providing a structured roadmap for executing a thorough analysis of the statewide economic implications of wildfires and their interconnected factors in Montana.

The final system design incorporates four distinct data sources: AQI data from the EPA, wildfire data from USGS, national park visitation data from IRMA, and GDP data from the Bureau of Economic Analysis (BEA). AQI measurements, encompassing multi-particle data (SO2, NO2, CO, O3, PM10, and PM2.5), were collected from monitoring sites in Missoula, Montana. Wildfire data, sourced from USGS, was meticulously filtered to include only wildfires within a 1250-mile radius of Missoula, Montana. For the study's national park visitation data, records were extracted from 1904 to 2022, covering all national parks and monuments in Montana, including Big Hole National Battlefield, Bighorn Canyon National Recreation Area, Fort Union Trading Post National Historic Site, Glacier National Park, Grant-Kohrs Ranch National Historic Site, Nez Perce National Historic Park, and Yellowstone National Park.

Economic data, vital for assessing the regional impact of national park visitation, will be derived from BEA's Regional Economic Accounts, specifically the March 31, 2023 (Gross Domestic Product by State, 4th Quarter and Annual 2022) files. Following the methodology introduced by Kim, four private industry subgroups were selected: 1. Agriculture, forestry, fishing and hunting, 2. Transportation and warehousing, 3. Arts, entertainment, recreation, accommodation, and food services, and 4. Retail trade. This selection was developed with the goal of capturing an exhaustive measure of the economic landscape influenced by national park activities.

These datasets collectively encompass Gross Domestic Product (GDP) estimates for all states from 1977 to 2022. The integration of these four distinct data sources ensures a comprehensive dataset spanning the years from 1983 to 2020, facilitating a thorough examination of economic trends over this period.

2.3 Research Questions

The primary research question guiding this study is: How do wildfires impact the economic stability of Montana? In pursuit of this overarching query, several sub-questions will be explored. Firstly, is there a robust linear correlation between factors influenced by wildfires, such as smoke, air quality, and national park visitation, and the GDP of private industries affected by the national parks? Secondly, by examining the historical trends of these variables, what insights can be gleaned regarding the future trajectory of Montana's GDP?

3. Methodology

3.1 Feature Engineering

In the initial phases of this project, the focus was on creating effective feature abstractions to encompass three key measurements: Air quality, national park visitation, and GDP. While park visitation and GDP data were obtained from singular sources, the approach to capturing air quality involved gathering AQI measurements from the EPA and generating a smoke estimate using wildfire data from USGS. The subsequent section outlines the design and rationale behind the feature engineering process.

3.1.1 AQI

First was a collection of AQI measurements from monitoring stations in Missoula county. The AQI is actually the maximum value of six major pollutants (SO_2, NO_2, CO, O_3, PM10, and PM2.5). Thus, the data comprises six separate daily measurements across multiple monitoring stations. To consolidate this multivariate data into a singular air quality measurement, I first computed the max pollutant concentration at each station for each day. This gave me an AQI for each station per day. I then computed the max AQI across all monitoring stations for each day, removing the monitoring station dimension. At this point, I have daily AQI measurements from 1977 to 2023. However, given that all of the other data sources only have the time fidelity of years, I need to aggregate these values one additional time. There are a seemingly infinite number of ways that one can aggregate these measures: maximums, minimums, averages etc. To guide this process, I considered what sort of measurements a potential visitor would be most concerned by. Given that potential visitors would likely be deterred from traveling to Montana if they see abnormally high AQI levels, I chose to include max and mean AQI for the year. Additionally, drawing inspiration from Kernan's work I chose to also include two additional measures of AQI. These were AQI75 and AQI100, which consist of the number of days that year where AQI exceeded 100 (Unhealthy for Sensitive Groups) and 75 (Moderate health concern). These AQI features are shown in **Figure 1** below.

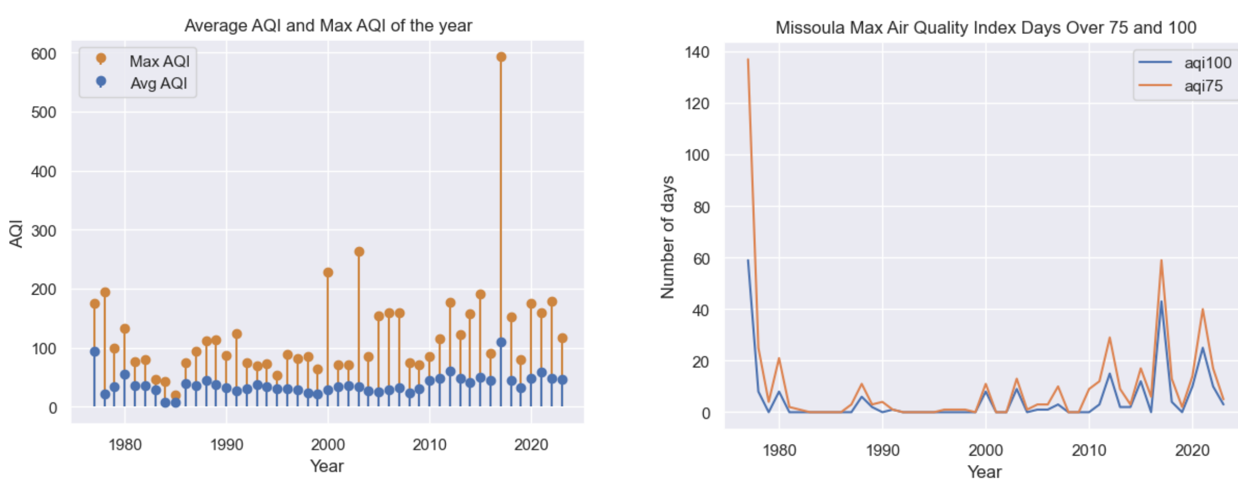

Figure 1: AQI aggregate features plotted against time. AQI measurements were collected from monitoring stations in Missoula county.

3.1.2 Smoke Estimate

The second method employed to gauge air quality involved utilizing wildfire data from USGS. To estimate the smoke emanating from wildfires, four key features were considered: the year, acres burned, distance from Missoula, and the fire type (wild or prescribed). Building upon Kim's approach of summing acres burned, this model expanded to incorporate both distance and wildfire type elements. Not only does incorporating distance into the estimate provide a more physically accurate model of smoke, but it can capture additional human centric behaviors. If potential visitors are residents of towns relatively close to Missoula and the national parks, they may also be experiencing poor air quality. Given that health authorities would likely advise against outdoor activities if air quality is poor, these residents will be further deterred from leaving their homes to venture towards Montana. The resulting formula was devised based on four main assumptions:

1. Fires with a greater acreage produce more smoke.
2. Fires in closer proximity to Missoula transport more smoke to the area.
3. The cumulative effect of multiple fires results in increased smoke generation.
4. Wildfires generate more smoke than prescribed fires.

The formula was as follows:

$$\text{Smoke Estimate(year)} = \sum_{j}^{n} \left(\frac{\text{Scaled Acres Burned}_j\text{(year)}}{\text{Scaled Distance}_j\text{(year)}} \times \text{Wildfire Type Constant} \right)$$

In the formula above, n is the total number of wildfires in that year and j represents a unique identifier for each wildfire for that year. Additionally, the acres and distance features for each wildfire were scaled using standard scaler and subjected to a sigmoid function to normalize their values between 0 and 1. The wildfire type constant takes on the value of 1 if the fire is a wildfire and 0.25 if the fire is a prescribed fire. The value of the constant reflects the assumption that prescribed fires are likely to produce less smoke due to their planned and controlled nature, minimizing their impact on surrounding cities. Simply put, for each year, the calculation involves iterating through all wildfires, dividing the scaled acres burned by the scaled distance from Missoula, and multiplying by the wildfire type constant. The resulting components are then summed to obtain the smoke estimate for that year. This equation was designed to capture the inverse relationship between distance and smoke while also considering the multiplicative impact of fire size and proximity to the city. The derived smoke estimate is shown in **Figure 2** below in the rightmost plot.

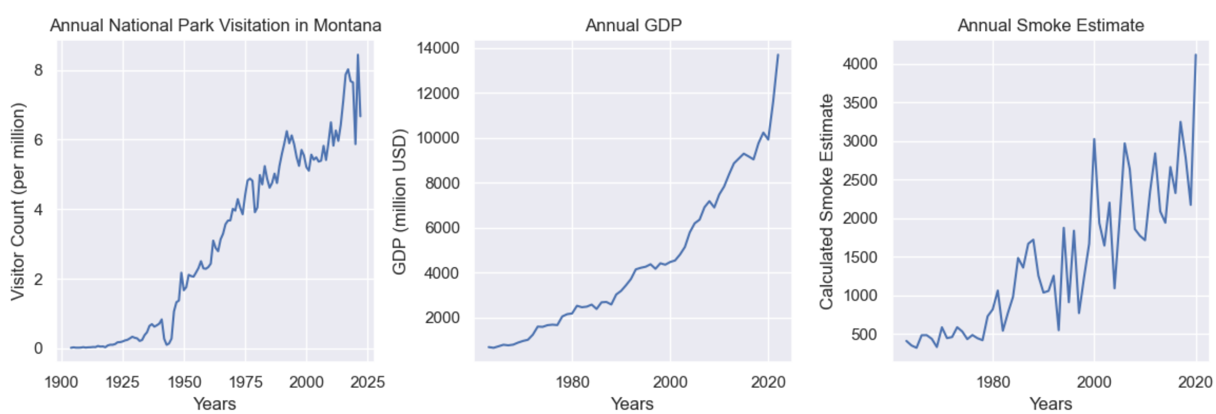

Figure 2: The time series plots of national park visitation, GDP, and smoke estimate features.

3.1.3 National Park visitation

Cleaning and transforming the National Park visitation data which was sourced from the National Park Service was relatively trivial. All that was required was aggregating the total visitation for each year from all of the national parks and monuments in Montana. The aggregated visitation data is displayed in the leftmost plot in **Figure 2**.

3.1.4 Gross Domestic Product

The final data source considered was the Bureau of Economic Analysis, an agency of the US Department of Commerce. This source provides 93 distinct categories of GDP measurements spanning from 1963 to 2022. Given my specific interest in understanding the economic impact of changes in air quality and park visitation due to wildfires on Montana's economy, I opted to focus on specific sectors rather than examining the overall GDP of Montana over time. I chose to narrow down the categories to the following: 1. Agriculture, forestry, fishing and hunting, 2. Transportation and warehousing, 3. Arts, entertainment, recreation, accommodation, and food services, and 4. Retail trade. Shown in **Figure 3** we can see that these industries make up 25% of Montana's GDP in 2022. This selection was informed by the prior work conducted by Kim et al., as mentioned earlier in the literature review. After reducing the number of categories, I aggregated the values across these industries to derive a final GDP-affected feature which is displayed in the center plot of **Figure 2**.

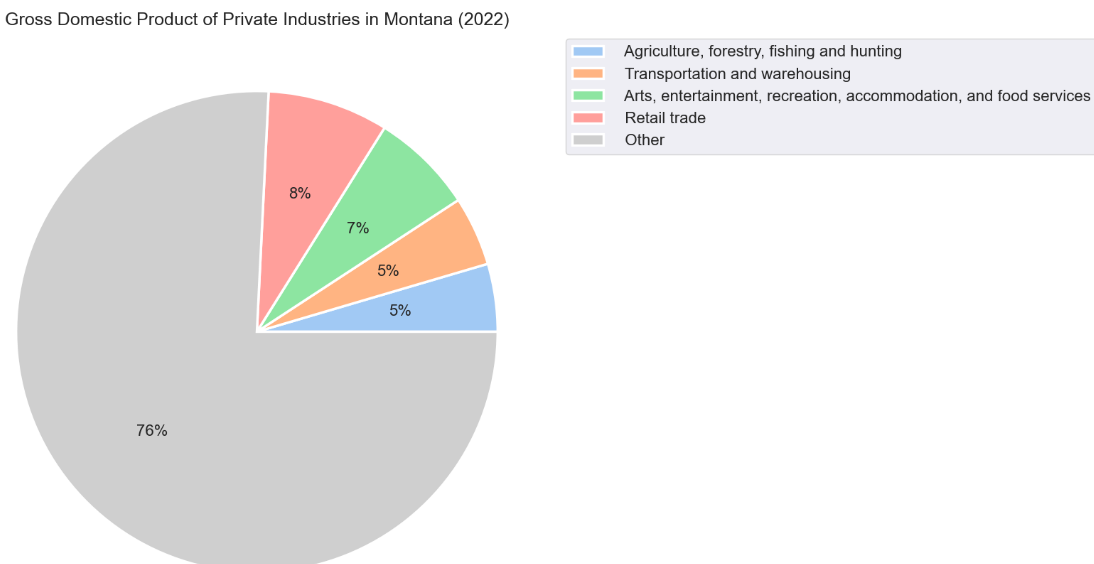

Figure 3: The private industries that were selected for the GDP estimate based on Kim et. al's work.

3.2 Correlation Analysis

In the pursuit of the primary research question—How do wildfires impact the economic stability of Montana?—this correlation analysis is pivotal. It seeks to reveal robust linear correlations between wildfire-influenced factors (smoke, air quality, and national park visitation) and the GDP of private industries linked to national parks. Synchrony is the degree of coordination or simultaneous variation across multiple time series variables. The objective is to quantify the synchrony between these variables, providing insights into the extent to which changes in smoke, air quality, and park visitation contribute to variations in Montana's economic landscape. To comprehensively address these questions, four distinct measurements of synchrony will be employed: Pearson correlation, rolling-Pearson cross correlation, time-lagged cross-correlation (TLCC), and dynamic time warping (DTW). By utilizing these correlation measures, this study aims not only to quantify relationships but also to illuminate the temporal dynamics and nuanced interplay among the variables.

Pearson correlation is one of the most standard measures for quantifying the strength and direction of a linear correlation between two variables. Utilizing this measure, I can determine the degree that Montanas GDP is linearly correlated with the wildfire-influenced factors. Using this insight, Montana's government officials and city council members of Missoula will be able to determine how much they should invest in countermeasures to safeguard their economic well-being. Although this correlation can provide easily interpretable results, there are some limitations. The three main limitations of Pearson correlation is that it cannot measure local synchrony, temporal dynamics and non-linear relationships. The first limitation can be addressed using rolling-window Pearson correlation by taking a specified window of the data, computing the Pearson correlation, and sliding it across all of the data. The benefit of quantifying local synchrony across the years is that we potentially detect yearly seasonal trends or periods of time where two variables are more strongly linearly correlated. If for example, the past smoke was not very predictive of GDP in early years, but became more strongly correlated in later years as the frequency of wildfires increased, rolling Pearson correlation could detect this behavior. The next method, TLCC addresses the temporal dynamics limitation by quantifying measurements of time lagged global synchrony. In the context of understanding wildfire impacts on Montana's GDP, TLCC can reveal if changes in smoke, air quality, or park visitation precede or follow economic shifts. For example, a positive correlation with a time lag might suggest that declines in air quality influence GDP years later, offering a nuanced understanding of the economic repercussions. Finally, DTW, which allows for stretching and compressing the time axis of one time series, can address nonlinear behaviors and patterns in changes in variables that do not align precisely. Similar to how rolling Pearson correlation allows for locality measures of linear correlation, DTW can identify periods of time where one variable is varying faster than the other. For instance, it might capture instances where the economic impact lags behind or accelerates in response to changes in air quality. This flexibility is crucial for understanding the nuanced and dynamic nature of these relationships.

In summary, while the Pearson correlation coefficient is effective for capturing linear relationships at a specific point in time, rolling-Pearson Correlation, TLCC and DTW provide a more comprehensive view of how these relationships evolve over time and accommodate potential delays. This enhanced temporal understanding is critical for crafting policies that proactively respond to the economic effects of wildfires on Montana's GDP.

3.3 Time Series Forecasting

This final phase of the analysis focuses on predicting the future trajectory of Montana's GDP using wildfire-influenced factors. Various statistical models are available for forecasting multivariate time-series data, including OLS, ARIMA, and Exponential Smoothing. I opted for OLS due to its superior transparency and interpretability, essential for communicating insights to a non-technical audience, including city officials and the general public. OLS provides clear coefficients for each wildfire-influenced variable, offering straightforward insights into the strength and direction of relationships. Additionally, OLS can be directly applied to non-stationary data, a key advantage over more complex models like ARIMA, which assume stationarity.

The initial step in developing the OLS model involved utilizing correlation analysis results to select features that optimize R2, adjusted R2, and minimize MSE while minimizing the number of features. Once the optimal feature subset was identified, the subsequent step was to forecast future GDP values. While OLS has many benefits, a notable drawback arises when utilizing multiple features—forecasting the response variable requires predicting the values of the predictor variables. To overcome this challenge, individual OLS models were developed for each wildfire-influenced factor, forecasting values from 2020 to 2030. These forecasted values were then employed to estimate the future trajectory of Montana's GDP.

4. Findings
4.1 Correlation Analysis

The first stage of this analysis was to look into the relationships between predictors and the response variable via correlation analysis. The relationships and plots that follow are based on the training set, the first 90% of the data. From the results of the Pearson correlation shown in **Figure 4** we can see that GDP is most strongly correlated with the year with a Pearson correlation coefficient of 0.96. This implies that 96% of the variation in GDP can be explained by variation in the time. In addition to time, park visitation as well as my smoke estimate also have strong correlations with GDP with coefficients of 0.75 and 0.71 respectively. Interestingly, none of the AQI abstractions had significant coefficients with GDP. As expected, all of the AQI features are strongly correlated with each other, a likely result of them all being derived from the same data. Based on this rough analysis, it seems that developing a linear model based on smoke, park visits and year would yield a highly predictive model.

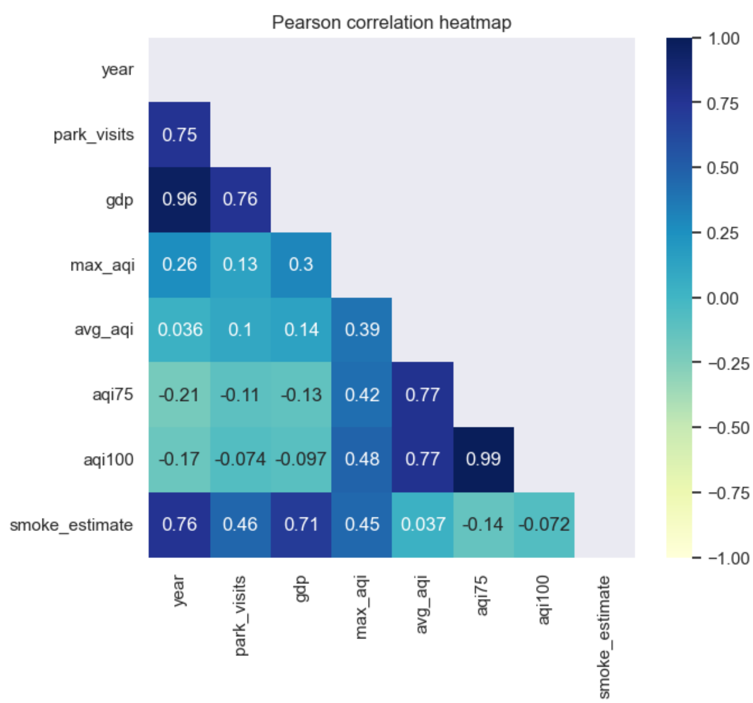

Figure 4: The Pearson correlation coefficients across all features and response variables.

As mentioned in the methodology section, Pearson correlation is a measure of global synchrony. To measure local synchrony, I computed a rolling Pearson correlation against GDP for max AQI, smoke estimate, and park visitation data with a 12 year window. Max AQI was selected over the other AQI features given that it had the highest global Pearson correlation of 0.3. Looking at the results in **Figure 5**, we can see that park visitation exhibited strong linear correlations with GDP during the periods of 1990 to 2000 as well as 2010 onward.

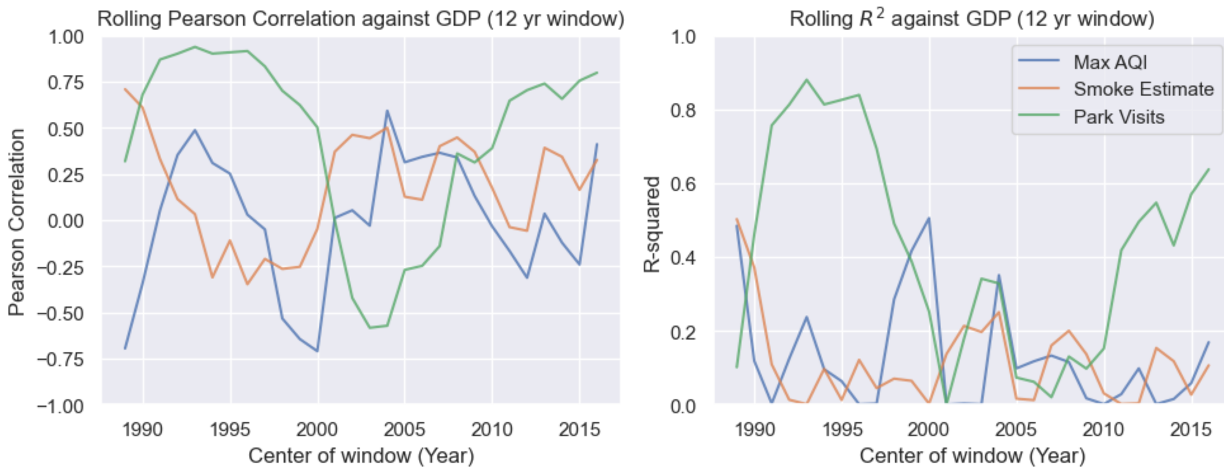

Figure 5: Rolling Pearson Correlation and Rolling R^2 across between Max AQI, Smoke Estimate, and Montana national park visits with a 12 year window.

These plots also highlight that there appears to be a bit of seasonality in the AQI and smoke estimate trends with GDP. Looking at the Rolling R^2 plot to the right of **Figure 5** we can see that both Max AQI and Smoke Estimate exhibit seasonal increases in R^2 over what appears to be 3-4 year intervals. It is difficult to completely interpret what this means since GDP appears to just increase over time, and thus any window of time where either of these variables begins to increase will likely increase the R^2 for that window. It is interesting however, that within these plots it seems that Smoke estimate is not significantly more correlated with GDP than Max AQI. In fact, with a window centered at 2000, Max AQI has a rolling R^2 of nearly 0.5, which is much higher than any of the other windows for Smoke Estimate. The results of these rolling correlations ultimately raise concerns for the assumptions taken to employ Ordinary Least Squares regressions later in this study.

After conducting rolling window correlations, I employed Time-lagged Cross-Correlation (TLCC) to investigate potential temporal dependencies between predictors and the response variable. **Figure 6** displays TLCC plots that reveal a robust, simultaneous correlation between park visitation and GDP, indicating concurrent fluctuations without noticeable lag. In contrast, TLCC analyses for AQI and the smoke estimate produce distinct patterns. Notably, for all AQI plots, the most substantial correlation occurs when GDP lags behind AQI by 7 years, suggesting that GDP increases are associated with subsequent AQI rises. Intriguingly, the TLCC plots for the smoke estimate demonstrate an inverse relationship, with GDP tending to increase 8 years after an elevation in smoke levels. These conflicting behaviors defy expectations, as one might anticipate similar trends between AQI, smoke, and GDP. Interpreting this substantial offset proves challenging. One plausible explanation posits that GDP growth fosters improved facilities, resources, and infrastructure, attracting more visitors to Missoula county. However, the impact on AQI becomes apparent only after a 7-year delay. Regarding the smoke estimate, the delayed effect could signify that heightened smoke levels adversely affect wildlife and Montana's national park landscapes, diminishing their appeal over time. Consequently, fewer visitors arrive, exerting a negative influence on the economy 8 years later.

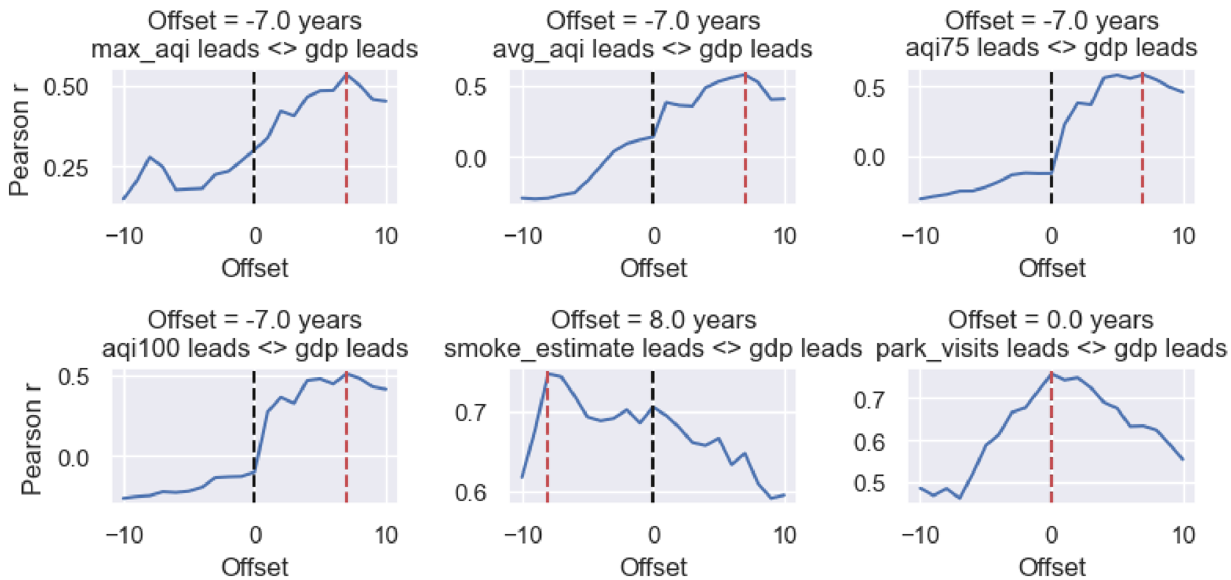

Figure 6: Time-lagged Cross Correlation plots for wildfire-influenced variables against GDP. The black dotted line indicates the center, when the offset is set to 0. The dotted red line highlights the offset at which the peak synchrony occurs.

The final measure of temporal synchronization explored in this study was DTW. Unlike TLCC, DTW excels at capturing both leading and lagging behavior as time varies. The plot depicted in **Figure 7** illustrates the accumulated cost matrix generated by the accelerated DTW algorithm implemented using the dtw Python package. The prominent white line traversing from the bottom-left to the upper-right corner represents the warping path. As explained by Franses and Wiemann (2020), "A non-shifted temporal alignment between the two series would be indicated by a fully diagonal warping path, whereas deviations from the diagonal indicate leading and lagging relationships of the series." According to this definition, Max AQI, Avg AQI, Smoke Estimate, and Park visits exhibit a closer alignment with GDP, as they closely follow the diagonal path. In contrast, paths such as AQI75 and AQI100 reveal that only the final values of these AQI metrics align with GDP. It is noteworthy that, in the Smoke Estimate and Park visits plots, the warping path oscillates back and forth between the bottom-right and uppermost-left corners across data points. This suggests that over time, these variables oscillate between leader and lagger behavior concerning GDP. This underscores the complexity of the relationship between these variables, again, emphasizing the necessity for more sophisticated statistical models to capture these intricate behaviors.

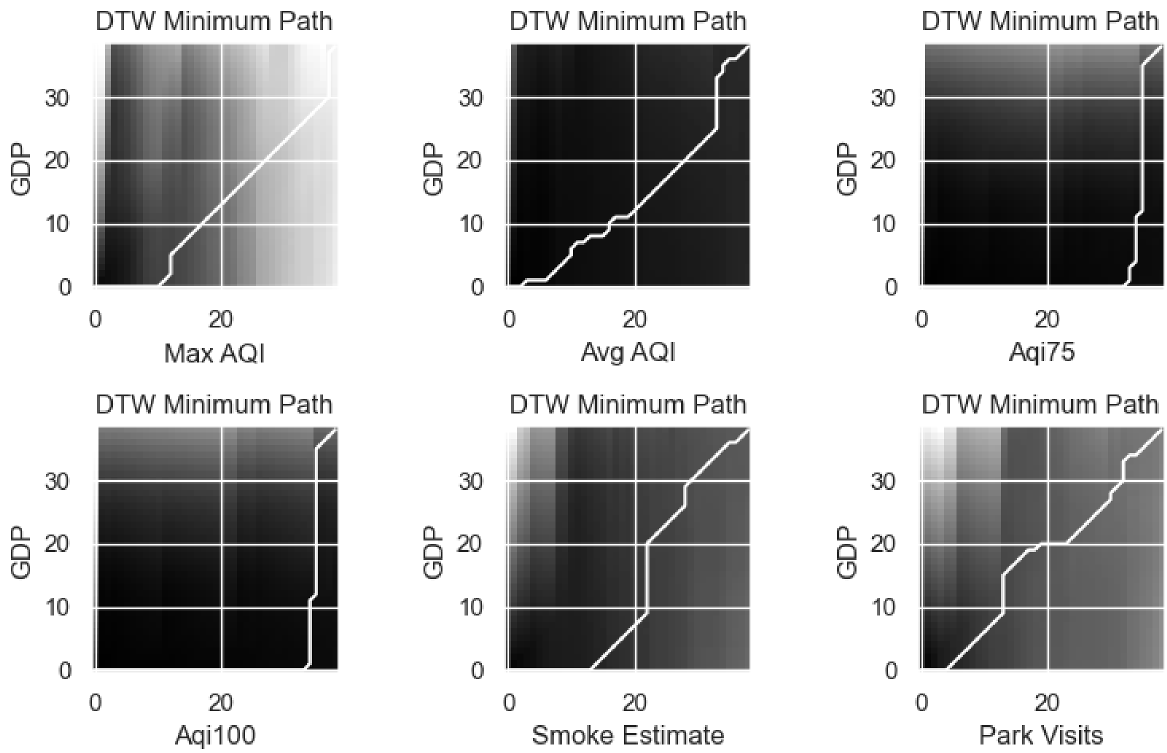

Figure 7: Dynamic Time Warping plots of all wildfire-influenced factors against GDP in Montana across 1977-2015.

4.2 Time-series Forecasting.

In the methodology section, the regression model chosen for GDP prediction based on wildfire-influenced factors is Ordinary Least Squares (OLS). Initial steps involved determining the subset of variables and types of terms (e.g., polynomial, interaction) to include in the OLS model. Recognizing the temporal nature of these relationships, interaction terms were incorporated between each wildfire-influenced factor and the year variable. To assess model performance and complexity, the Akaike Information Criterion (AIC) was compared across four OLS models of varying intricacy. The model with the lowest AIC was selected. **Table 1** presents the outcomes of these models.

Name	**Features**	**Additional Terms**	**AIC**	**Adj. R^2**
Full Model	Year, max_aqi, avg_aqi, aqi75, aqi100, smoke estimate, park_visits	Temporal interaction terms	-123.5	0.978
Subset1	Year, max_aqi, smoke estimate, park_visits	Temporal interaction terms	-125.8	0.977
Subset2	Year, smoke, park_visits	Temporal interaction terms	-127.0	0.977

No interaction	Year, smoke, park_visits	N/A	-84.72	0.928

Table 1: Displays the three OLS models that were evaluated and their corresponding AIC values.

During the process of addressing collinearity concerns highlighted in the correlation analysis, I observed marginal enhancements in AIC, through removing various AQI terms, while preserving a high adjusted R^2 value. Among the models tested, Subset2 exhibited the lowest AIC, and its formula is detailed below.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3$$

This final OLS model incorporates the year, smoke estimate, and park visits as predictor variables. Additionally, there are two interaction terms between the smoke estimate and year, as well as park visits and year, represented by the final two terms in the formula. Examination of the prediction line plot in **Figure 8** reveals that the OLS model reasonably tracks the training data but tends to overpredict on the test set, specifically during the years 2015 to 2020.

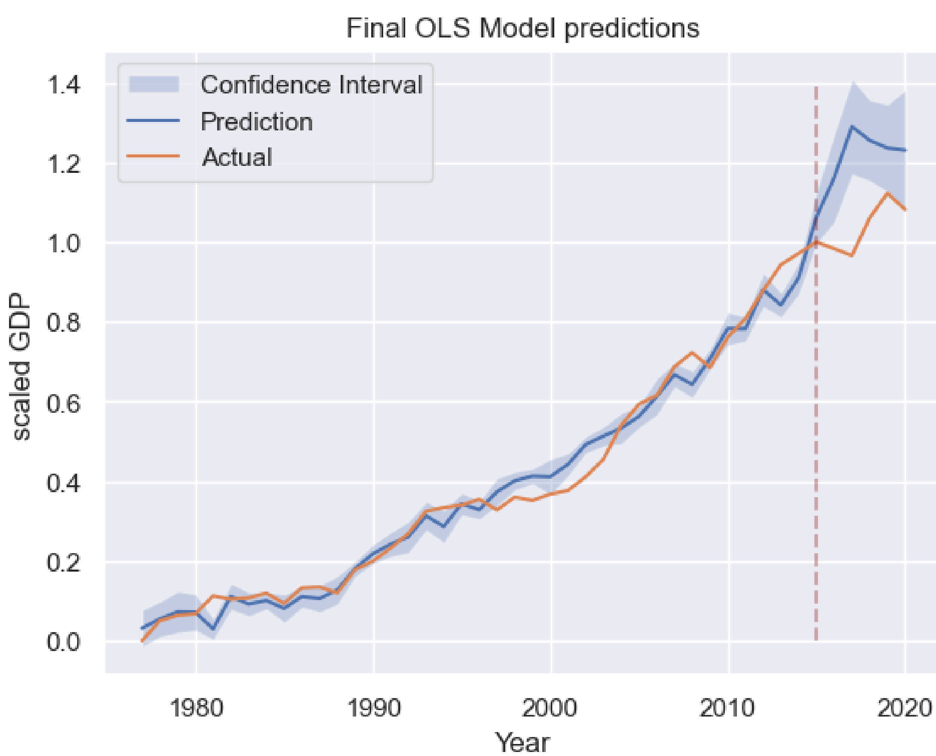

Figure 8: The Final OLS models training and test predictions with confidence intervals. The dashed red line indicates the separation between training and test data.

To forecast Montana's GDP beyond 2020, I initiated the process by predicting data for the predictor variables in the final model. This involved fitting regressions of the predictor variables against the time variable using Ridge Regression. A grid search with polynomial degrees ranging from 1 to 9 was employed. Subsequently, using the best ridge estimators—degrees 1 and 3 for smoke_estimate and park_visits, respectively—I projected their future values for the next decade. The results of the final GDP forecast suggest an increase to approximately $15,500 million USD by 2030, seemingly unaffected by the rising incidence of wildfires. With an AIC of -127.4 and an adjusted R2 value of 0.976, this final model demonstrates high performance. A closer examination of the coefficients in **Table 2** reveals that each term is statistically significant, evident from the p-values below the 0.01 significance level.

Name	**Coefficient**	**p-value**
Intercept	-52.83	1.63e-19
year	0.03	1.28e-19
smoke_estimate	-5.39	4.44e-4
smoke_estimate.year	0.003	4.55e-4
park_visits	-3.12	5.13e-3
park_visits.year	0.002	4.95e-3

Table 2: Shows the features, coefficients and p-values of the Final OLS model fitted on all of the data from 1977 to 2020. Note that all features have statistically significant p-values below the 0.001 alpha level. Values of the coefficients are scaled, and thus do not represent specific physical values.

Interpreting the coefficients for each variable requires accounting for the interaction terms. The coefficient of the Smoke variable (−5.39), represents the effect of Smoke on GDP when the variable Year is zero (at the reference point). This effect is the baseline impact of Smoke on GDP, without considering any interaction with time. For the coefficient of the interaction term Smoke×Year (0.003), it represents the additional effect of Smoke on GDP for each unit increase in Year. In other words, it captures how the impact of Smoke on GDP changes over time. In simpler terms, the negative coefficients for Smoke suggest that, at the reference point, an increase in Smoke is associated with a decrease in GDP. Unexpectedly, the same inverse relationship is observed between Park Visits and GDP. This contradicts the anticipated positive correlation, where an increase in park visits would lead to an increase in GDP—a trend evident in the correlation analysis. While the interpretations of the coefficients may not provide extensive insights, the forecasted plots in **Figure 9** reveal a positive trajectory for all three variables—Smoke, Park Visits, and GDP. These projections indicate an upward trend for these factors over the coming decade.

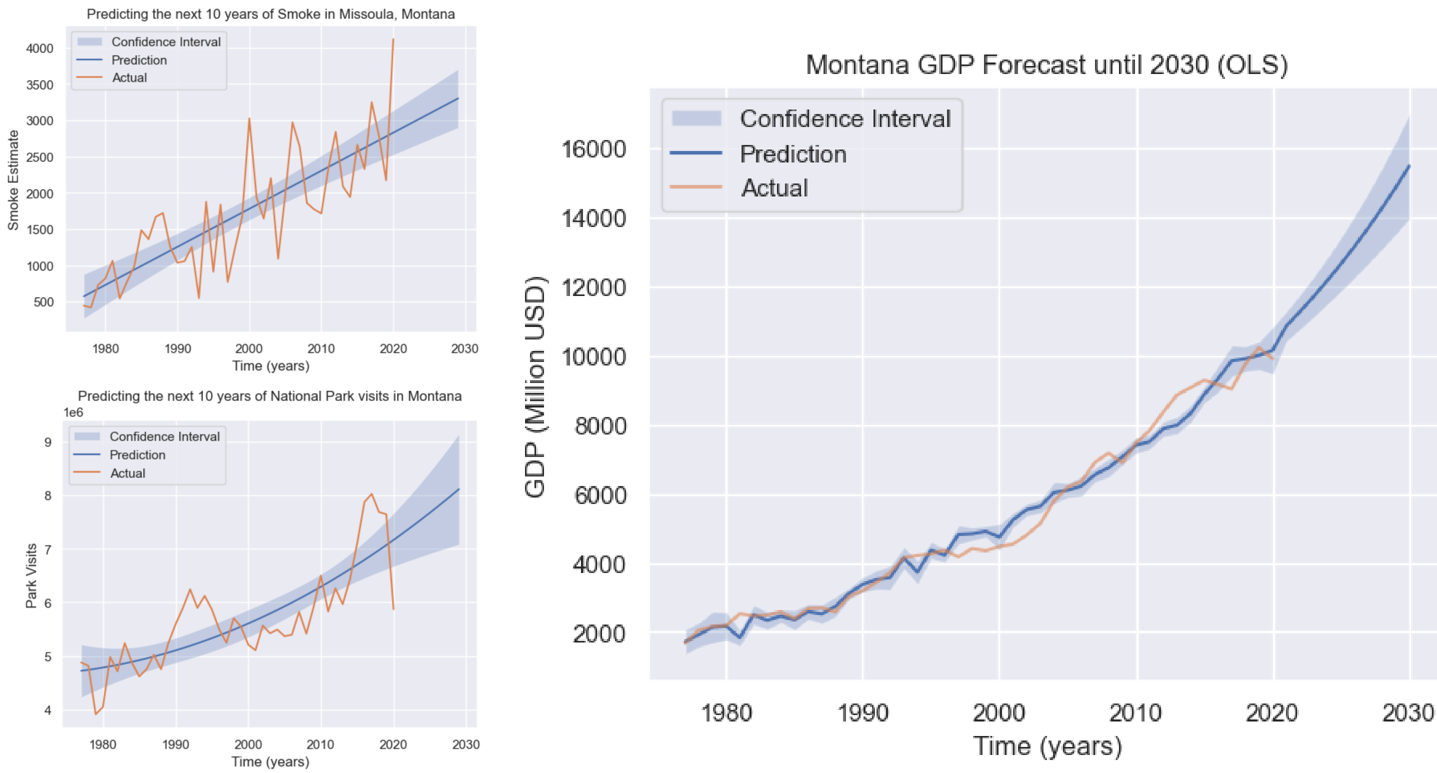

Figure 9: Shown on left are the projected values of Smoke and National Park visitation that were used as inputs into the final OLS model to forecast Montana's GDP to 2030 shown in the plot on right.

5. Implications

The significance of these findings lies in the critical role smoke and park visitation play in influencing Montana's GDP. Despite historical increases in wildfires, national park visitation continues to increase over time. With the development of a forecasting model based on these key predictors, the projection indicates that Montana's GDP is expected to rise to $15,481 million USD by the year 2030. In response to these findings, the Missoula city council, city manager/mayor, and residents can rest assured knowing that the projected upward trajectory of Montana's GDP, highlights the resilience of the local economy.

However, in light of the projected increase in wildfires and smoke, it is advisable for the council to develop and execute a strategic plan within the next year. This plan should include the establishment of additional air quality monitoring stations and the enhancement of park visitation tracking with a focus on visitor expenditures. While existing evidence suggests that wildfires are not projected to significantly affect the economy, it is crucial to continue monitoring for changes in this dynamic, since based on the DTW results, relationships can change over time.

In addition, it is crucial to acknowledge, assess, and address other ways in which residents can be affected by the escalating wildfire activity. These include but are not limited to health concerns, infrastructure damage, water quality issues, and even psychological impacts. It is critical that the council begins allocating additional funds for further research to assess the effects of rising smoke levels on these factors so that they can ensure all aspects of the well-being of Missoula residents.

6. Limitations

In the course of this study, it is crucial to recognize and address several limitations that may impact the robustness and general applicability of the findings. Firstly, while the chosen data sources offered valuable insights, there is a need for increased fidelity, particularly in acquiring more detailed and comprehensive data concerning park visitation and economic activities. Utilizing fine-grained data, on a monthly or even daily scale, could unveil nuanced relationships and enhance the precision of the analysis. Moreover, the reliance on aggregated values for AQI, smoke and even GDP might obscure underlying variations, prompting consideration for a more granular approach at smaller temporal scales.

Secondly, there were specific assumptions that were made to capture key variables that have their limitations. The smoke estimate, for example, is contingent on various assumptions about the factors influencing smoke. One key assumption was that the accumulation of smoke due to multiple wildfires is a term that can be described by the sum of the smoke generated by other fires. A more theoretical and exhaustive approach, informed by additional research on the intricacies of smoke generation, would strengthen the accuracy and reliability of the smoke estimate. Moreover, the use of Montana Statewide GDP projections as a proxy measure for the city of Missoula introduces assumptions regarding the uniformity of economic trends across the state. Future investigations could benefit from incorporating more localized GDP measures specific to Missoula.

Turning to the forecasting measures, it is imperative to acknowledge the limitations associated with dependency on forecasted results. The forecasted GDP relies on additional forecasted outcomes, introducing an element of uncertainty and potential errors. Future research should explore methodologies to minimize cascading errors in forecasting by adopting more integrated modeling approaches. This is further supported by warnings about potential multicollinearity among variables in the fitted OLS models, particularly due to the inclusion of multiple time-dependent predictor variables. Exploring the application of alternative forecasting models, such as ARIMA or Exponential Smoothing, may mitigate issues associated with multicollinearity and enhance the accuracy of GDP forecasts.

In summation, recognizing and addressing these limitations will contribute to the refinement and advancement of future research endeavors in this domain, promoting a more comprehensive understanding of the economic implications of wildfires in Montana.

7. Conclusion

In conclusion, this study aimed to address the central research question: How do wildfires impact the economic stability of Montana? To understand this overarching inquiry, specific sub-questions were examined. Initially, the study investigated the existence of a robust linear correlation between factors influenced by wildfires—namely, smoke, air quality, and national park visitation—and the GDP of private industries affected by national parks. This analysis uncovered significant insights into the interconnected dynamics of these variables. Finally, the study sought to provide valuable insights into the future trajectory of Montana's GDP, shedding light on potential trends and considerations for Missoula's economic stability.

Through the analysis, this study has yielded crucial findings that significantly contribute to understanding the impact of wildfires on the economic stability of Montana. The correlation analysis underscored smoke estimate and park visitation as key variables, revealing their pivotal roles in influencing the state's GDP. The temporal complexity of these relationships was further elucidated through TLCC and DTW, showcasing both lagging and leading behaviors at different timepoints. Finally, through the development of OLS models, I developed predictions for future values of Montana's GDP. Intriguingly, despite historical increases in wildfires, national park visitation demonstrated a consistent upward trend, reflecting the resilience of local tourism. Projections suggest that Montana's GDP is anticipated to reach $15,481 million USD by 2030. This work exemplifies a specific application of data science, showcasing its capacity to uncover the potential impacts of real-world events on human needs. Beginning with a measurable change in wildfires, I hypothesized potential impacts on smoke and national park visitation to develop a model that ultimately predicts how Montana resident's livelihoods would be affected. Extending beyond the statistical conclusions of this study, I provided actionable recommendations to Missoula city council members and residents to further monitor economic challenges posed by wildfires and to fund further research into other sectors that may be affected by the growing number of wildfires in Montana.

8. References

1. *About Missoula*. (n.d.). Destination Missoula. Retrieved December 8, 2023, from https://destinationmissoula.org/about-missoula
2. *AQI Basics | AirNow.gov*. (n.d.). AirNow.gov, U.S. EPA. Retrieved December 8, 2023, from https://www.airnow.gov/aqi/aqi-basics
3. *Combined wildland fire datasets for the United States and certain territories, 1800s-Present—ScienceBase-Catalog*. (n.d.). Retrieved December 11, 2023, from https://www.sciencebase.gov/catalog/item/61707c2ad34ea36449a6b066
4. Franses, P. H., & Wiemann, T. (2020). Intertemporal Similarity of Economic Time Series: An Application of Dynamic Time Warping. *Computational Economics*, *56*(1), 59–75. https://doi.org/10.1007/s10614-020-09986-0
5. Hendrix, L. (2022). *MONTANA'S OUTDOOR ECONOMY*.
6. Keller, M. C. (n.d.). *Post-fire recovery within ravine forest communities of Pattee Canyon, Missoula, Montana*.
7. Kernan, M. J. (n.d.). *MODELING AND FORECASTING GLACIER NATIONAL PARK VISITATION*.

8. Kim, M.-K., & Jakus, P. M. (2019). Wildfire, national park visitation, and changes in regional economic activity. *Journal of Outdoor Recreation and Tourism*, *26*, 34–42. https://doi.org/10.1016/j.jort.2019.03.007
9. *National Park Visitation Sets New Record as Economic Engines | U.S. Department of the Interior*. (2023, August 21). [Press Release]. https://www.doi.gov/pressreleases/national-park-visitation-sets-new-record-economic-engines
10. *National Parks In Montana: Explore The 8 Montana National Parks (2022 Update) - National Park Obsessed*. (2021, October 3). https://nationalparkobsessed.com/national-parks-in-montana-national-parks/
11. *National Parks Near Missoula*. (n.d.). Wren Missoula. Retrieved December 8, 2023, from https://wrenmissoula.com/explore-missoula-hotel-downtown/national-parks-near-missoula-mt/
12. PhD, J. C. (2021, December 9). *Four ways to quantify synchrony between time series data*. Medium. https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9
13. *The National Park Service*. (n.d.). Retrieved December 11, 2023, from https://www.nps.gov/nature/customcf/NPS_Data_Visualization/state.html?state=Montana&code=MT
14. U.S. Bureau of Economic Analysis. (2001, January 1). *Total Gross Domestic Product for Missoula, MT (MSA)*. FRED, Federal Reserve Bank of St. Louis; FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/NGMP33540
15. US EPA, O. (2016, August 30). *Air Data Basic Information* [Data and Tools]. https://www.epa.gov/outdoor-air-quality-data/air-data-basic-information

9. Data Sources:

For accessing the data sources, the methods, links and licenses are all shown below.

1. Combined wildland fire datasets for the United States and certain territories, 1800s-Present
    a. Data Source: USGS ScienceBase-Catalog, FRESC Public Data
    b. License: Public Domain
    c. Link: https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81
    d. Description: Collection of US wildfire data records and their features which include but are not limited to USGS ID, the fire type, the year the fire occurred, number of acres burned, and an object representing the location and size of the wildfire.
2. AQI Dataset
    a. Data Source: US EPA Air Quality System API
    b. License: Public Domain
    c. Link: https://aqs.epa/data/api
    d. Description: The EPA provides an API where users can post requests from specific monitoring stations to get daily measurements of pollutants such as (O_2, SO_2, NO_2 etc).
3. Annual Visitation and Record Year by Park (1904 - Last Calendar Year)
    a. Data Source: IRMA National Park Service Visitor Use Statistics
    b. License: Public Domain
    c. Link: STATS - National Reports (nps.gov)

d. Description: The NPS provides an Integrated Resource Management portal for the public to download datasets of temporal visitation estimates from any national park in the US. Data columns consist of region name, park name, park type, year, and total recorded visitation.
e. Method:
   i. Select Annual Visitation and Record by Park (1904 - Last Calendar Year)
   ii. Generate report using all regions and park types while limiting the Parks to the following subset of national parks that are in Montana (Big Hole National Battlefield, Bighorn Canyon National Recreation Area, Fort Union Trading Post National Historic Site, Glacier National Park, Grant-Kohrs Ranch National Historic Site, Nez Perce National Historic Park, Yellowstone National Park)
   iii. Download report to a csv file.

4. GDP Data
   a. Data Source: Bureau of Economic Analysis - US Department of Commerce
   b. License: Public Domain
   c. Link: [BEA : Regional Economic Accounts - Previously Published Estimates]
   d. Description: The BEA provides GDP estimates for all US states across both private and government sectors. The downloaded tables contain columns such as FIPS code, state name, region, tablename, linecode, industry classification, description, unit and a column for each year of recorded GDP.
   e. Method:
      i. Navigate to Gross Domestic Product by State
      ii. Download the following files
         1. March 31, 2023 (Gross Domestic Product by State, 4th Quarter and Annual 2022)
            a. Download the Annual GDP by State and Industry