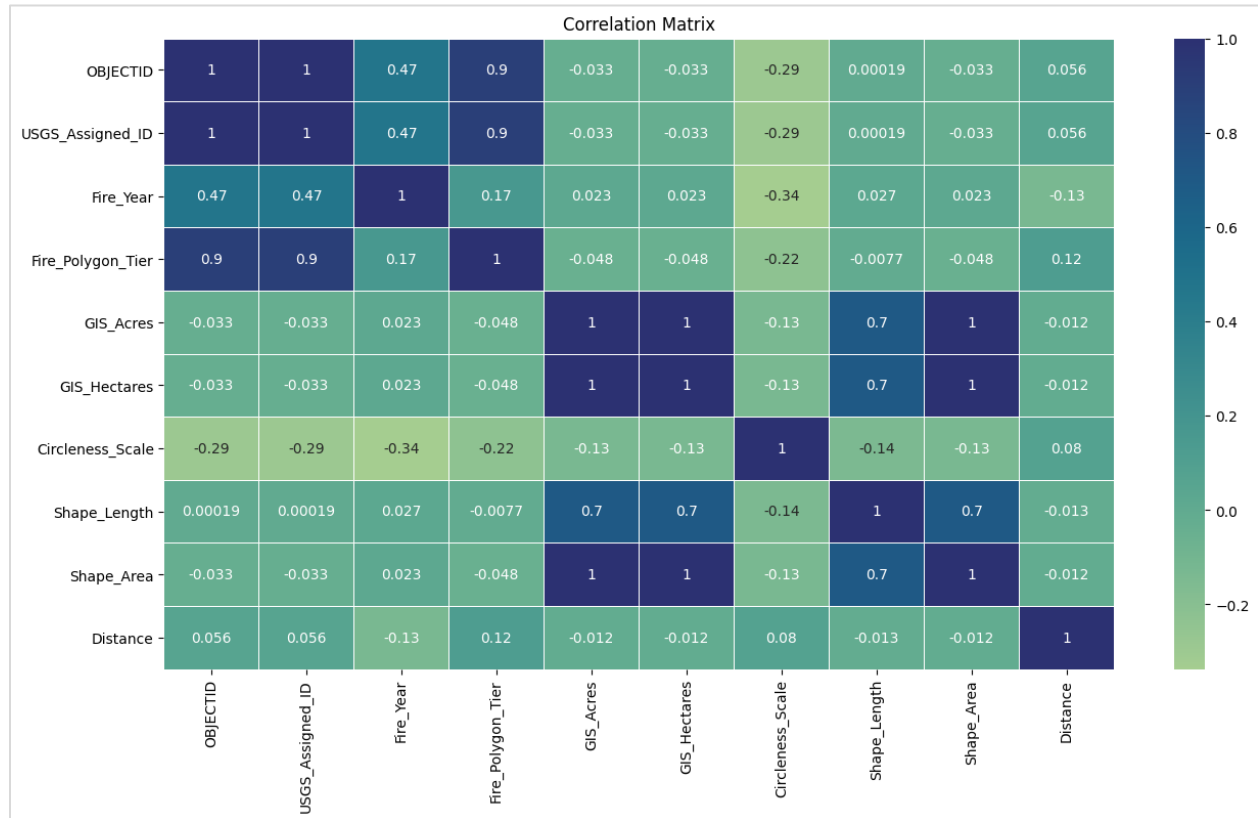


Visualization Interpretations/Explanations:

Visualization 1



The heatmap visualization illustrates the correlations between the different factors with values relevant to the designated city, Minot, covering a span of up to 1250 miles.

The correlation matrix provided furnishes a comprehensive overview of the interrelationships among the variables under consideration, forming a critical foundation for the deliberative process in formulating a novel smoke estimate variable. A salient observation resides in the existence of a perfect positive correlation between OBJECTID and USGS_Assigned_ID, indicative of a direct but redundant association because of the context of our case-study.

Temporal dynamics, as encapsulated by Fire_Year, display a moderate positive correlation with OBJECTID, USGS_Assigned_ID, and Fire_Polygon_Tier. This signifies a discernible temporal influence on these variables, accentuating the imperative to account for the temporal aspect in the envisaged smoke estimate. Contrastingly, the spatial variable Distance manifests minimal correlations with other variables, suggesting a limited impact on the measured parameters. In light of these spatial dynamics, on performing additional scrutiny, we did not consider them in the context of constructing the smoke estimate.

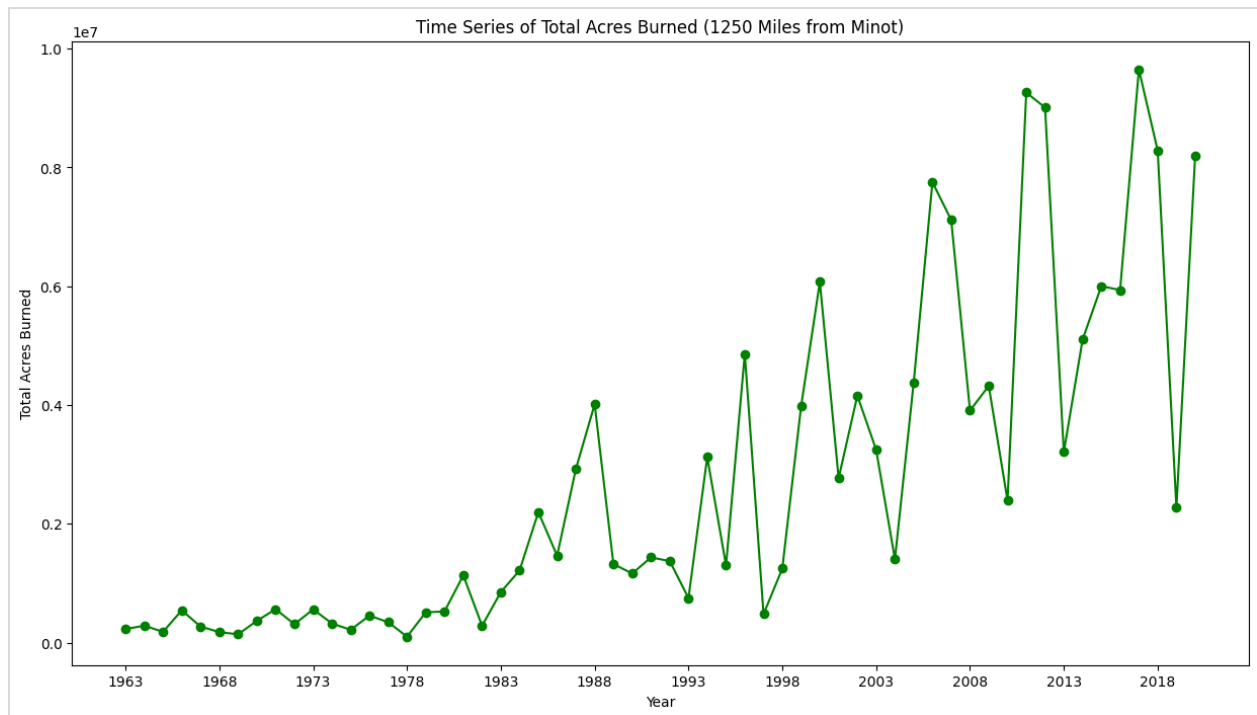
The geospatial metrics, GIS_Acres and Shape_Area, reveal a perfect positive correlation, indicating functional redundancy. Hence, we did not select these variables to avert redundancy in the envisaged smoke estimate. Noteworthy variables, such as Fire_Polygon_Tier, exhibit distinct characteristics and

robust associations, thereby positing them as prospective cornerstones in the construction of the new smoke estimate.

In interpreting the graph, the color map (cmap) serves as a visual conduit delineating correlation strength and direction. The gradation of warmer hues signifies positive correlations, while cooler tones connote negative correlations. Optimization of the color map configuration enhances visual discernibility, and a judicious comprehension of the scale, ranging from -1 to 1, facilitates a nuanced understanding of the magnitude of correlations.

In summation, the presented analysis not only identifies pivotal associations and redundancies but also lays the groundwork for an informed decision-making process in the development of a bespoke smoke estimate variable. The insights gleaned from the correlation matrix, when amalgamated with domain knowledge, constitute a substantive asset as a process to learn more about the data and research.

Visualization 2



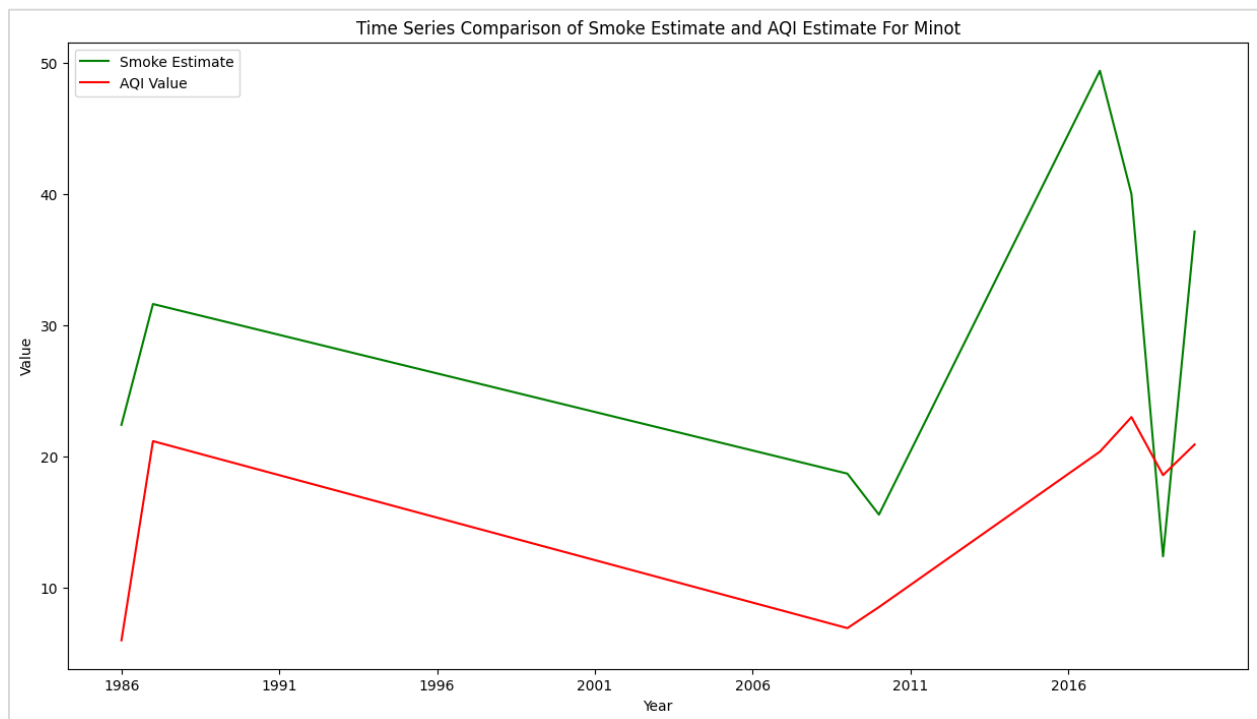
The time series graph presented above illustrates the annual progression of total acres burned by wildfires within a 1250-mile radius from Minot. Serving as a valuable analytical tool, this graph provides a clear depiction of the dynamics of wildfires in the region over a specific timeframe. The x-axis represents the years, while the y-axis indicates the total acres burned for each respective year, facilitating a straightforward comprehension of the relationship between time and the extent of wildfires.

The data underpinning this time series plot is sourced from the original fire dataset, meticulously filtered to include only those fires falling within a distance of up to 1250 miles from Minot and occurring between 1963 and 2020. Noteworthy is the absence of data post-2020 meeting both criteria. The filtered data was subsequently aggregated by year, with the total acres burned calculated for each year, culminating in the

construction of the time series plot. To enhance readability, the x-axis intervals were configured to display years at 5-year increments.

A discerning analysis of the graph reveals a discernible trend—a marked increase in the total acres burned as the years progress. In fact, the total acres burned in 2020 register a magnitude nearly 10^3 times greater than those in 1963. While occasional fluctuations and temporary declines punctuate the trajectory, the overarching pattern remains conspicuous. Various factors contribute to this concerning trend, including rising temperatures, prolonged droughts fostering favorable conditions for wildfires, and the introduction of non-native invasive plant species altering ecosystems, rendering them more susceptible to fires. Moreover, population growth and urbanization amplify the risk of accidental ignitions and fires in the wildland-urban interface. Notably, a nuanced observation from this analysis reveals a cyclical year-on-year trend amidst an overarching upward trajectory when viewed over a more extensive timeframe.

Visualization 3



The line graph depicts a temporal interplay between the smoke estimate, derived from an ARIMA model within a 1250-mile radius of Minot, North Dakota, and the government-provided Air Quality Index (AQI). The correlation coefficient of 0.6555 suggests a moderately positive relationship between the two variables, indicating a discernible alignment in their trends. The X-axis spans the years from 1986 to 2020, offering a chronological context, while the Y-axis quantifies the respective values of the smoke estimate and AQI.

Examining the green line representing the smoke estimate, distinct temporal patterns emerge. A notable peak occurs in 2017, indicating a substantial spike in estimated smoke levels, followed by a decline in 2018 and another increase in 2020. The red line representing AQI values responds to these fluctuations, affirming the correlation. In 2017, the AQI peaks correspondingly to the elevated smoke estimate, signifying a direct impact on air quality.

In alignment with AQI standards, the majority of data points fall within the "good" air quality range (0-50). However, spikes in both the smoke estimate and AQI in 2017 and 2020 breach this threshold, indicating periods of compromised air quality. These peaks prompt a closer investigation into the specific conditions contributing to heightened smoke levels and their subsequent impact on local air quality. Understanding the causative factors behind these peaks can inform targeted interventions for effective air quality management, ensuring a comprehensive and proactive approach to environmental monitoring.

Reflection Report:

An essential takeaway from this assignment underscores the critical role of meticulous data cleaning and exploration. Beyond obtaining results and statistical values, a conscientious validation process becomes pivotal. This involves scrutinizing not just what is present but delving into the nuances of what is absent. The observation regarding the final smoke estimate and AQI graphs, revealing a seeming trend due to missing data from 1987 to 2009, exemplifies the necessity of thorough examination. This gap influences the flat line in the graphs, a crucial insight that surpasses mere statistical values and requires acknowledgment.

Beyond data validation, collaboration brought attention to potential pitfalls such as data gaps and outliers, compelling effective addressing of these issues. An instance involves the identification of a small number of curveRings instances in the fire data, comprising just 35 out of 135,000 data points. Their limited impact led to their exclusion, emphasizing the nuanced handling of data intricacies. The process of datafication, as discussed in class, sometimes overlooks subtleties due to scope and data collection considerations. Recognizing this, we incorporated data smoothing techniques, utilizing moving averages to effectively address outliers within our time series data.

In this intricate task marked by limited subject knowledge, collaboration played a pivotal role. Within the collective cohort setting, a rich exchange of ideas, designs, approaches, and algorithmic considerations contributed to a more robust and insightful solution. The iterative process of experimenting with various algorithms and metrics facilitated a nuanced analysis, culminating in the development of the final visualization.

Addressing the research question unearthed several insights regarding wildfires and their impact on Minot. The upward trend in total acres burned within a 1250-mile radius over six decades underscores the influence of climate change, prolonged droughts, invasive species, and urbanization, collectively intensifying wildfires. A compelling observation from the analysis of fire smoke estimates and the Air Quality Index (AQI) for Minot reveals an absence of a clear trend in the former, while the latter displays a gradual upward trajectory. This transition from "good" to "moderate" air quality underscores the long-term impact of wildfires, necessitating urgent measures for public health and environmental mitigation.

The exploration of the spatial distribution of wildfires through the histogram visualization yielded significant insights. A lower frequency of fires within the initial 500 miles from Minot, likely due to enhanced fire prevention measures in urban wildland interface (UWI) zones, contrasts with a notable spike between 750 and 1250 miles from Minot. This regional disparity emphasizes the need for tailored fire management strategies accommodating varying fire frequencies across distances from the city.

A departure from traditional approaches to understanding wildfire patterns, collaboration introduced the concept of geospatial data analysis, expanding the research dimension. Collaborating with Anish broadened my understanding in this regard, while the collaboration with Harshit Rai proved instrumental in refining our comprehension of air quality and its intricate relationship with wildfires. This holistic approach, combining methodological innovation and collaborative insights, enriched the depth and breadth of our research endeavors.

What I could improve if I had more time: If I could improve one aspect where I struggled and would love to spend more time on: it's the scalability: the data acquisition part took 8 hours to run which is too long and not ideal – there has to be a better way to get the data without such a long processing time.