Project Part 1: Common Analysis

Figure 1: Number of fires vs every 50-mile-distance in Gillette, Wyoming

Number of fires vs every 50-mile-distance in Gillette, Wyoming
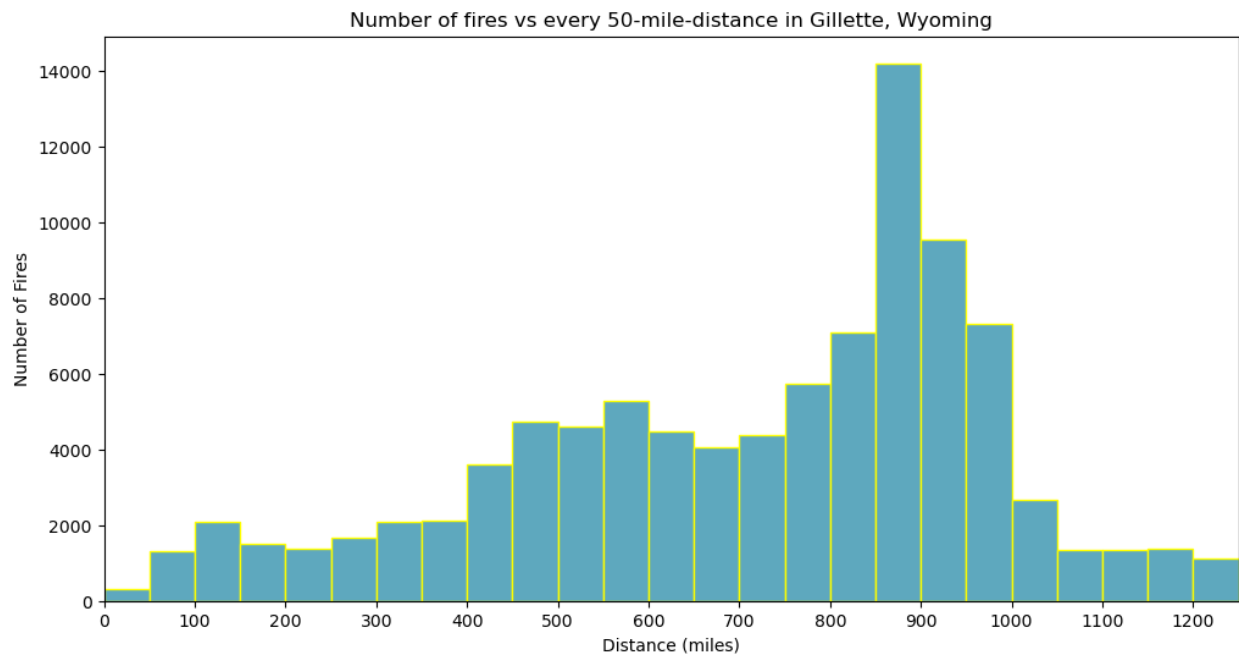
The figure above represents a histogram that displays the number of fires occurring at every 50-mile-distance from Gillette, Wyoming up to a maximum specified distance of 1250 miles. The x-axis represents the distance in miles from Gillette, Wyoming. Each bin on the x-axis represents a range of distances, with labels at the bin edges (e.g., 0-50, 50-100, 100-150 miles, and so on). The y-axis represents the number of fires that occurred within each distance range. It shows the count of fires falling within each bin.

The underlying data used for this visualization is a dataset of fire occurrences. The data is filtered to include only fires that fall within a specified maximum distance from the city (in this case, 1250 miles). The data is then grouped into 50-mile range bins based on the distance from Gillette.

The figure above can be 'read' from left-to-right or vice versa. From the histogram, it is evident that the most significant concentration of fires, numbering approximately 14,000 incidents, occurred in regions situated around 850 to 900 miles away from Gillette, Wyoming. In contrast, areas within a 100-mile radius of the city have experienced fewer than 2,000 fires in the last 60 years. After 1,000 miles, the number of fires declines again. This indicates a possibility that certain fire-prone regions of the country are located about 800-1000 miles from Gillette which contributes to the spike we see on the histogram.

Figure 2: Total Acres Burned per Year within 1250 miles of Gillette, Wyoming
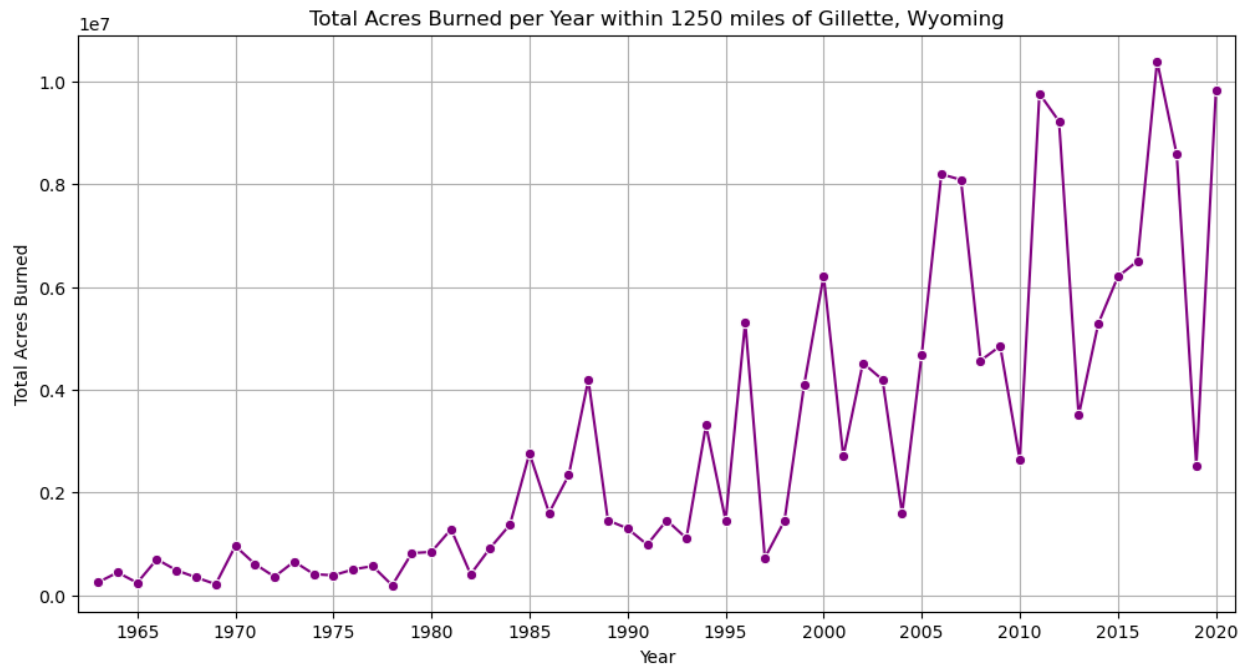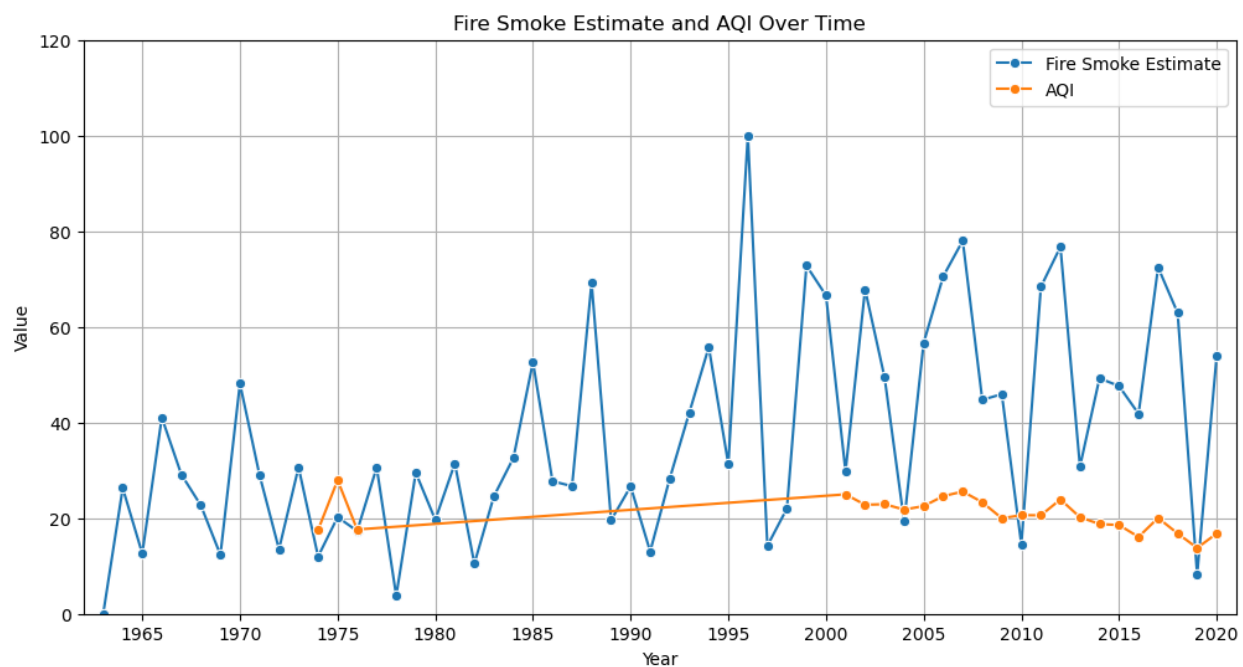

The figure above presents a time series plot that visualizes the total acres burned each year within 1250 miles from Gillette, Wyoming. The x-axis represents the years from 1963 to 2020, displaying each year at five-year intervals, which provides a clear view of the long-term trends. The y-axis shows the total acres burned, illustrating the scale of wildfires' impact on the area.

The underlying data has been processed by filtering the dataset to include only fire incidents within the specified distance from the city. Then, the data is grouped by year and the total acres burned in each year are calculated. This processed data is used to create the time series plot.

The plot is best read left-to-right. It reveals a pattern of wildfire activity that is not continuous but rather characterized by fluctuations from year to year. While the total acres burned display a general upward trend over the analyzed years, the data exhibits distinct periods of increase and decrease in wildfire impact. During some years, such as 2011 and 2017, there are noticeable spikes in the total acres burned, signifying periods of elevated wildfire activity. Conversely, there are years, like 2010 and 2019, when the total acres burned decrease, indicating a reduction in wildfire impact.

This variability strongly suggests that Gillette, Wyoming, is not consistently prone to being near large fires. Instead, the region experiences intermittent periods of heightened wildfire risk, with some years witnessing more significant wildfire events while others enjoy relative respite.

Figure 3: Fire Smoke Estimate and AQI Over Time

The figure above displays a time series graph representing the Fire Smoke Estimate and Air Quality Index (AQI) for Gillette from 1963 to 2023. The x-axis represents the years from 1963 to 2020, displaying each year at five-year intervals, which provides a clear view of the long-term trends. The y-axis represents the value of the two variables being measured. The scale of the y-axis corresponds to the values of the Fire Smoke Estimate which is from 0-100 and AQI which is from 0 – 500.

The underlying data comprises records from 1963-2023, with associated values for Fire Smoke Estimate and AQI Estimate. Fire Smoke Estimate is calculated based on distance and the area burned. AQI Estimate represents air quality data, and its values reflect air pollution levels, which are usually measured using a standardized index. The AQI data for Gillette was quite sparse despite the presence of monitoring stations. Consistent data is only available after 2000.

The plot is best read left-to-right. Rather than a continuous trend, the graph displays fluctuations in Fire Smoke Estimate from year to year. In contrast, the Air Quality Index (AQI) for the city appears relatively stable over the years, with a slight but noticeable decline beginning around 2012. A correlation if 0.29 is present between both variables which seems reasonable considering many factors that affect smoke impact such as wind direction over a course of several days, the intensity of the fire, and its duration are not available to us.

<u>Reflection Statement</u>

Through this assignment, I developed a deeper understanding of air quality monitoring and factors affecting air quality. The collaborative aspect of this assignment played a pivotal role in expanding my knowledge. I collaborated with Mohammad Danish Nadeem on conceptual discussions related to the research question and code related to the final ARIMA model I used to predict the smoke estimate.

A discussion that I thought was very useful when answering the research question was related to how to go about calculating the AQI for the chosen city. We discussed multiple options such as should be consider data from multiple sensors or a few, should gaseous or particle pollutants be prioritized for one reason or another or ways to calculate the AQI for a particular year.

Also useful was discussing the methodologies for calculating the smoke estimates for each year. Through this process, I learnt that I was initially calculating smoke estimates incorrectly as I was trying to come up with an estimate for a year directly. Instead, a better approach was calculating smoke estimates for each fire event and then devising an appropriate way to average them over the course of a year. This collaborative process also raised pertinent questions about how to aggregate these estimates. The debate revolved around whether the yearly smoke estimate should be cumulative, a simple average, or a weighted average. The final choice was a simple average as it seemed to be the most fitting approach for capturing the city's average smoke impact throughout the year.

Discussing the prediction task was also useful. The absence of predictor data for the forecast years of 2021-2049 posed a significant challenge, leading to extensive discussions about potential modeling approaches. Options discussed included regression, Autoregressive Integrated Moving Average (ARIMA), and Seasonal ARIMA (SARIMA). The discussion was useful as we exchanged ideas regarding the advantages and disadvantages of each modeling approach.

Overall, through this assignment, I realized that the utility of collaboration is inherently tied to the nature of the problem at hand. In Homework 1 and Homework 2, we were given straightforward tasks where the path to a solution was relatively clear-cut and so I did not feel a need to collaborate with others. However, for complex problems, such as the one in this assignment, collaboration proves to be a very useful tool. The nature of such problems calls for multifaceted thinking, as there are various routes to tackle the issue, and each may lead to different results. I think this is also something to keep in mind when I join the industry and work on projects in the real world.