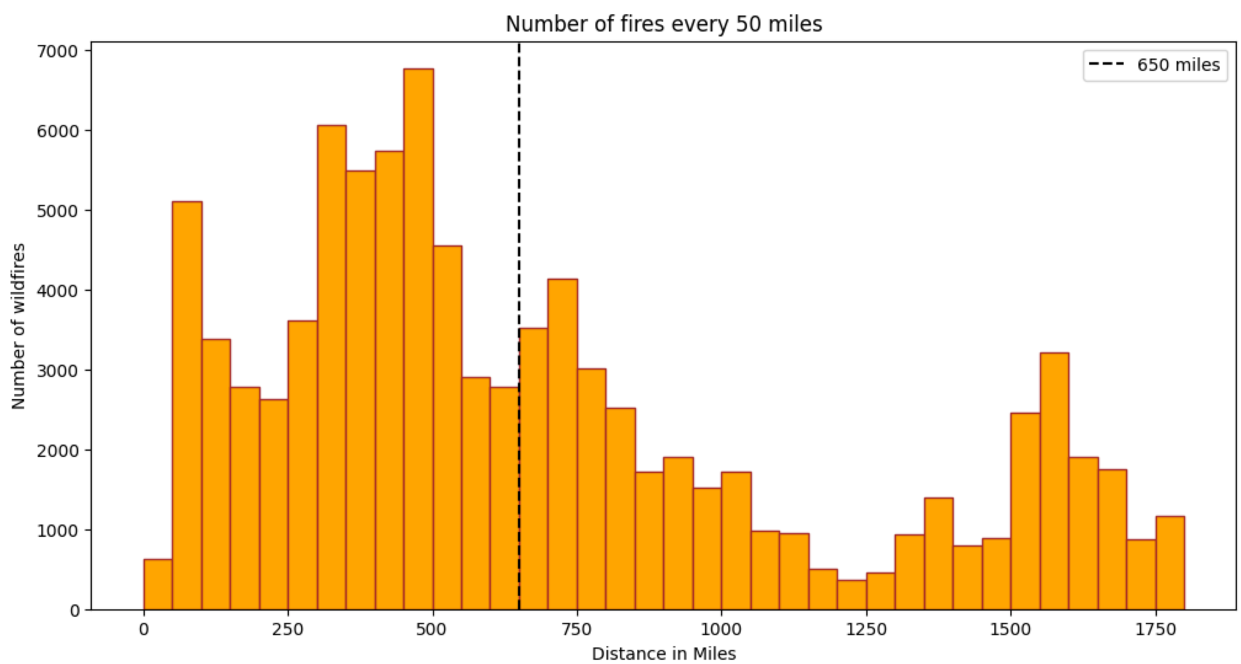
Reflections

Visualizations

Plot 1

A histogram showing the number of fires occurring every 50 mile distance from Stockton, CA for all fires ranging up to 1800 miles away. The histogram indicates the distance cut-off used for my modeling work, i.e, 650 miles.



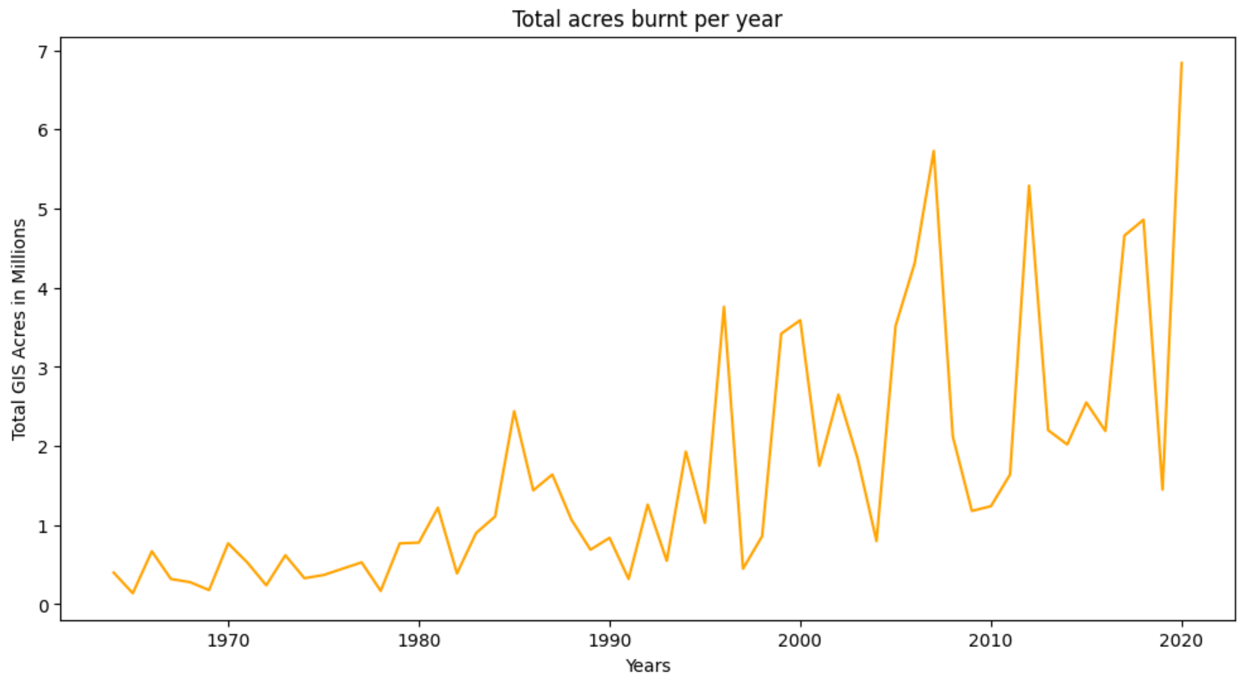
In the above plot, the x-axis represents the distance from Stockton, CA in miles, while the y-axis represents the count of wildfires in that corresponding distance range. The counts are calculated in buckets of 50, i.e., wildfires within 0-50 miles are counted together, 51-100 are counted together, and so on. The dotted line represents the 650-mile cutoff that is used while filtering the data to use for training the model. This dotted line gives us an idea of the scope of the project.

The data for this graph was fetched from the [Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons)](), and the shortest distance from each wildfire to the city was calculated. This distance in miles was used plot the above graph.

From the above histogram, we see that a big proportion of the wildfires within 1800 miles of Stockton, CA are actually below 500 miles. This suggests the proximity of Stockton, CA to wildfire zones, and hints towards the fact that wildfires might have a considerable impact on the air quality in this city.

Plot 2:

A time series graph of total acres burned per year for the fires occurring in 650 miles from Stockton, CA.



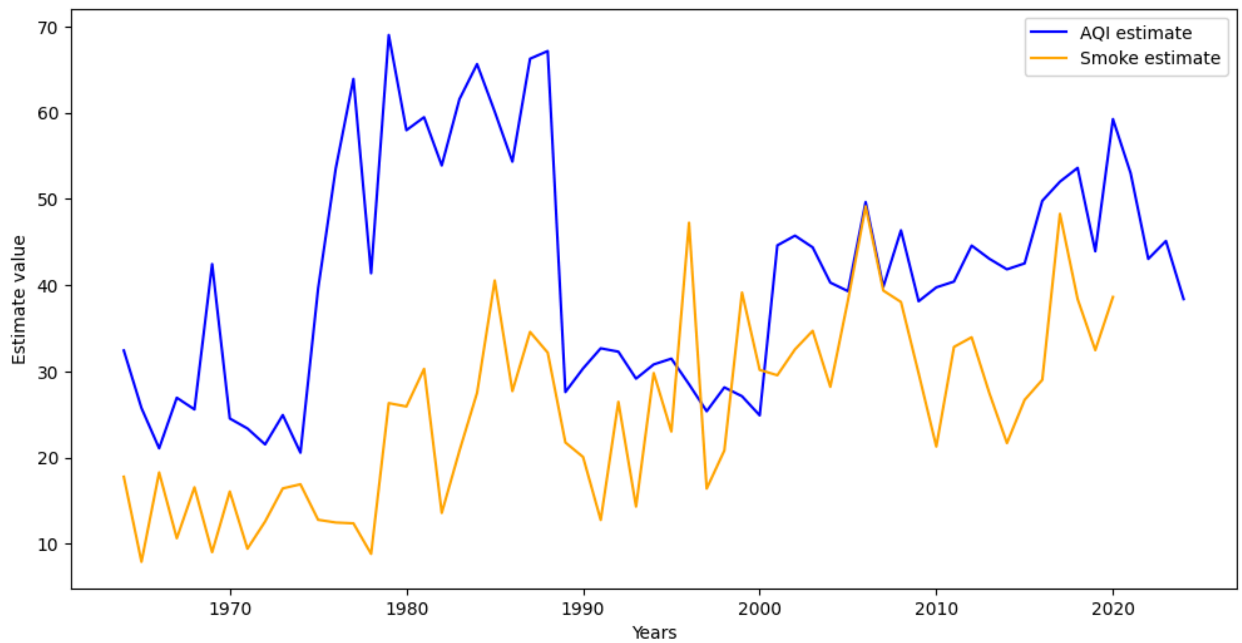
In the above plot, the x-axis represents the years from 1964 to 2020, while the y-axis represents the total area burnt(GIS Acres in Millions) due to wildland fires. The total area burnt is represented in millions to improve the readability of the graph.

The data for this graph was fetched from the [Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons)]. Each wildland fire in this dataset has information on the year in which the fires occurred, and the total area of land burnt (GIS_Acres). Both these fields were used, and the total GIS_Acres burnt in a year is calculated by summing up the acres_burnt for all wildfires in that year.

From the above line graph, we see the recorded area burnt due to wildland fire was fairly smaller before 1990. Towards the mid-1990s, we started to see a big spike in the graph, indicating that there have been large/severe wildfires during those years. Looking at the Wikipedia page of California wildfires, these are the years in which California experienced the deadliest wildfires, which contributed to the spike in the total area burnt. We see that year 2020 has had the highest area burnt

Plot 3:

A time series graph containing the fire smoke estimates and AQI estimates for Stockton, CA.



The time series graph illustrates smoke estimates for Stockton, CA, along with the yearly air quality index (AQI) values over the years. The x-axis spans the years (1964–2024), while the y-axis shows estimated values for AQI and smoke estimates. The yellow line represents smoke estimates, and the blue line represents AQI values.

The data for the above graph comes from 2 sources-

1. Wildland fire data: The smoke estimates at Stockton,CA were calculated using information from wildfires like total area burnt, distance from the city to the wildfire, type of the fire, etc. The original source of the data is [Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons)]
2. AQI data: The annual AQI estimates are calculated by aggregating the daily summaries from different air quality monitoring stations that are near the Stockton,CA city. AQIs of different pollutants are aggregated to find the annual estimates. The original AQI data is fetched using the APIS mentioned at https://aqs.epa.gov/aqsweb/documents/data_api.html

We notice that both the estimates increased with years, but we notice an unusual spike between mid 1970s to late 1980s. Interestingly, though there is a shift in the AQI values, the trend seems to be close to that of the smoke estimates, with both the curves having some peaks and dips in a few years. This suggests that during those years, there might have been another contributing factor to the high value of the air quality index.

Collaboration

Reflecting on part of this project, I've realized just how valuable collaboration is when dealing with complex data challenges. One of our main tasks was to figure out the best way to aggregate AQI data, a process that wasn't straightforward due to many edge cases and unique data sources. Brainstorming with others helped me get the clarity I needed and also provided new insights because everyone was researching different resources. Working with Navya Eedula was especially helpful here. She's detail-oriented, and her structured approach made it easier to understand these complexities and find a clear path forward.

Our work together didn't just involve crunching data; it also helped me better understand what the numbers mean. Navya had done extensive reading about AQI, and both of us were able to have a lot of back-and-forth discussions to ensure we handled cases with missing AQI values properly. This experience taught me a lot about how AQI is derived from pollutant levels, and working alongside someone so methodical really boosted my confidence in reading and interpreting complex documents. This wasn't just about solving a data issue—it was about building knowledge, and each conversation helped us gain a clearer understanding.

Discussions about estimating smoke dispersion were also crucial. Smoke estimates depend a lot on details like the shape of a fire and its distance from a location, so I worked closely with Manasa Shivappa Ronur to dig deeper into these factors. Together, we considered how the fire's shape could impact smoke dispersion—whether a circular shape holds smoke closer or an irregular one lets it spread. Manasa's analytical approach helped us think through how these spatial details would influence our calculations. Her insight into the shape's effect on smoke concentration helped me refine my final smoke estimation model.

Through this research, I learned how wildfires are not the only contributing factor for AQI. The shift in the AQI index from the years 1975 - 1990 shows an interesting factor trend. Also, learning about how AQI indexes are calculated was super interesting. Additionally, while I knew that wildland fires are prominent in california, I was never quite aware of the scare of destruction it causes. Looking at the total acres burnt per year was baffling for me.

Beyond the technical takeaways, this project showed me how important open communication and flexibility are when working together. By openly sharing our methods and adjusting to each other's ideas, we turned individual insights into stronger strategies. Working with Navya and Manasa showed me that collaboration isn't just about dividing up tasks but about combining our different perspectives for a more accurate analysis. All of us supported each other, as we constantly validated and refined our approach.

Overall, this project gave me new technical skills and a deeper appreciation for collaborative work. I learned how valuable it is to check ideas with others, learn from them, and solve tough problems together. Moving forward, I'm even more convinced that the best results come from combining diverse strengths. This project has shown me that collaboration can enhance any research or analysis, and it's a lesson I'll carry with me into future projects.