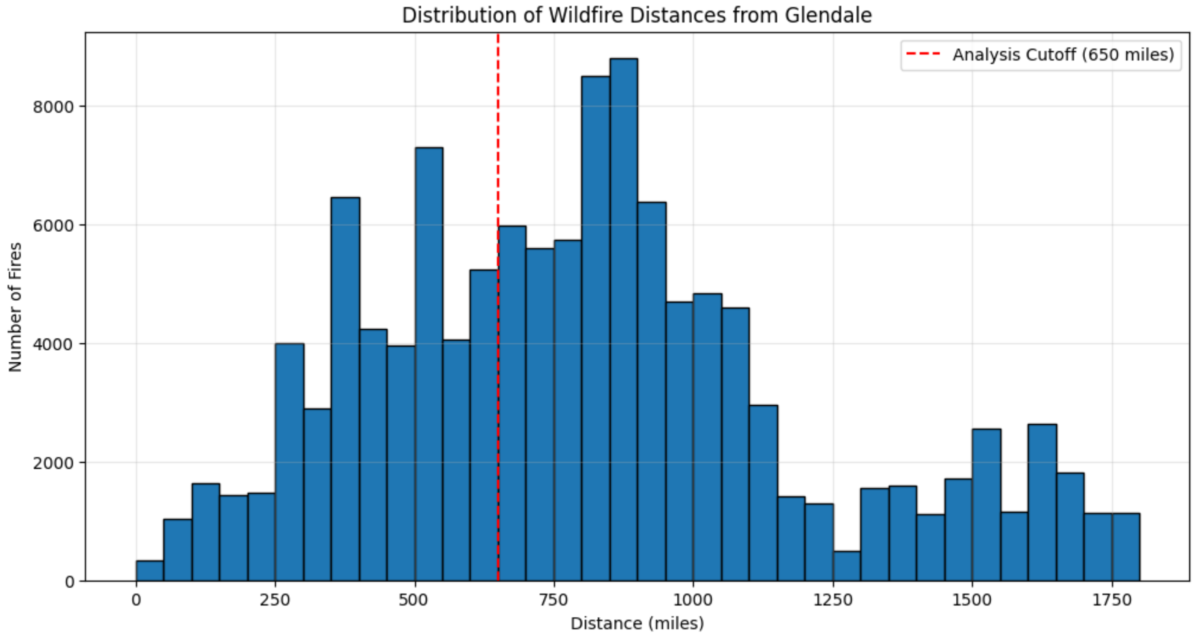
Visualization 1



This histogram represents the distribution of wildfires with respect to their distance from the city of Glendale, Arizona. The x-axis shows the range of distance in miles from Glendale, ranging from 0 to about 1800 miles. The y-axis gives the number of fires observed at each distance interval. Each bar represents a 50-mile range in distance, and the height of each bar shows the number of fires that occurred within that exact distance from Glendale.

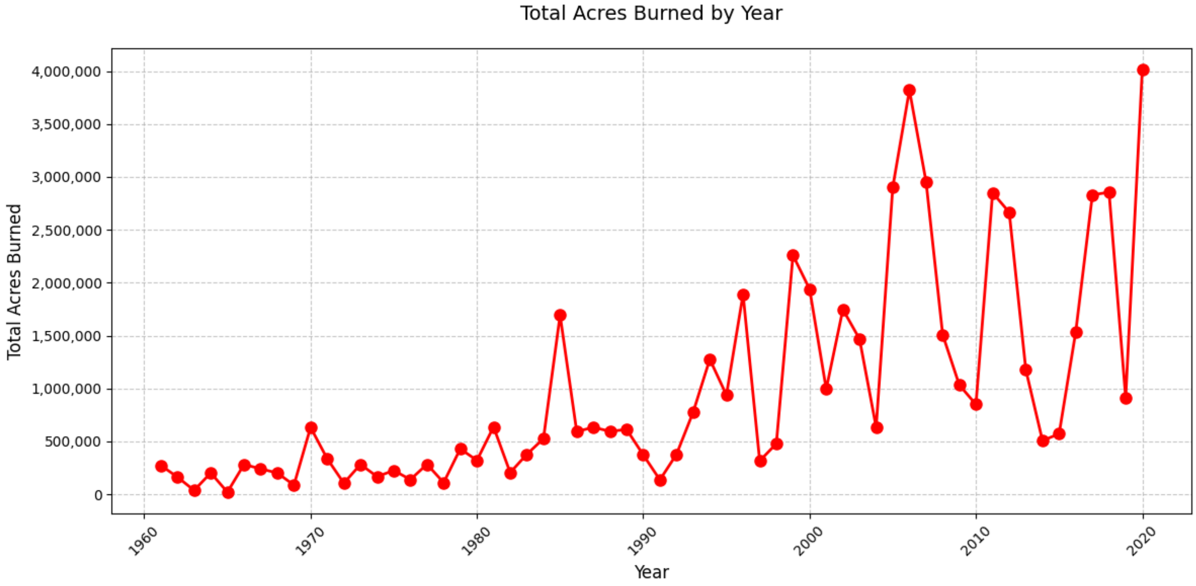
There is also a red dashed vertical line in the histogram at 650 miles labeled as "Analysis Cutoff." This cut-off point gives the threshold distance that may be used to give emphasis on fairly close fires to Glendale. The emphasis in the said cutoff assists the audience in easily distinguishing between the fires that are within the immediate vicinity, hence 0-650 miles, and those that are quite a distance away from it.

The data comes from a JSON dataset that includes geographic coordinates of wildfire perimeters across the United States. These coordinates are initially in the ESRI:102008 coordinate system, so each point is transformed into the EPSG:4326 (latitude and longitude) format for accurate distance calculations. Using Glendale's coordinates (latitude: 33.53, longitude: -112.19) as a reference point, distances to each wildfire perimeter are calculated using the geodesic distance formula, which measures the shortest path between two points on the Earth's surface. Only the shortest distance from Glendale to any point on each wildfire's perimeter is considered. This calculation is done in meters and then converted to miles for readability.

The figure highlights that most wildfires occur within roughly 250 to 1000 miles from Glendale, with a peak around 900 miles. After 1000 miles, the frequency of fires decreases, with a smaller set of fires observed beyond this range.

This visualization is designed to provide insights into the spatial distribution of wildfires concerning Glendale. It can help analysts or decision-makers understand how wildfire activity varies with distance from a specific urban area, potentially guiding resource allocation or risk assessment efforts based on wildfire proximity.

Visualization 2



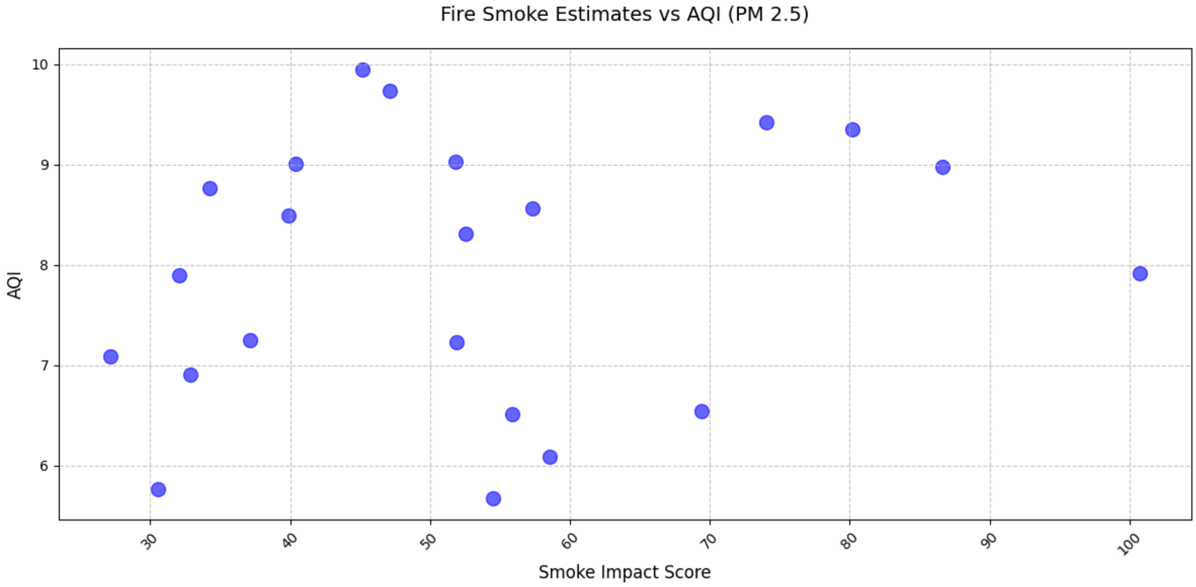
This line chart compares the trend for the total acres burned in wildfires across the United States, starting roughly from the year 1960 to the end of 2020. The year is represented on the x-axis, with labels showing at about one-decade intervals; total acres burned are represented on the y-axis, with commas showing for thousands to improve readability. The wide variation in the impact of wildfires over time is captured by a y-axis scale running from 0 to over 4,000,000 acres of land burned.

Each point represents the total area burned by wildfire in a particular year, identified with a red circle. To give an easy visual trend, the line that connects these has also been colored red for clear contrast to the background. A dashed grid is also provided to help the viewer more easily match each data point to its corresponding value on the y-axis.

The data very clearly shows an upward trend across the decades, with a relative low and stable area burned each year from 1960 through the 1980s. Starting in the 1990s, an increase is seen both in the frequency and the scale of large wildfire events. This trend continued into the 21st century, where peaks in burned acreage occurred around the mid-2000s, the 2010s, and again in 2020. The height has reached just above 4,000,000 acres during just the last few years of this chart, showing indeed that the activity of wildfire has intensified.

This visualization suggests that wildfires have become more frequent and severe over the years, likely due to factors such as climate change, land management practices, and increased temperatures. It provides a straightforward look at how wildfire impact has evolved, enabling researchers, policymakers, and the general public to recognize the growing scale of wildfire threats and the need for proactive measures in fire management and climate adaptation.

Visualization 3



The following scatter plot shows the Smoke Impact Score with the Air Quality Index AQI for particulate matter-PM 2.5, which is a common measure of air pollution especially due to wildfires. The x-axis is the Smoke Impact Score, ranging from 30 to 100, which is a quantified measure of the potential impact that smoke from a wildfire will have on air quality. The AQI values represent the y-axis, ranging between 6 and 10 for this plot, with higher values representing poorer conditions.

Each point is a Smoke Impact Score for location and time paired with the AQI value, but the marker size and the blue color helps them be more visible. The points are a bit transparent so when the points do overlap, it is easier to see that more than one data point exists in that location.

Calculation of the Smoke Impact Score

The Smoke Impact Score in this visualization is derived from a custom function that estimates the smoke impact based on fire size (in acres) and distance to a city (in miles). This function, `estimate_smoke_impact`, was specifically calibrated to correlate with observed PM2.5 levels, making it a useful indicator for air quality conditions due to wildfire smoke.

Key components of the Smoke Impact Score calculation are as follows:

1. **Distance Factor:** The score penalizes fires that are farther away, but instead of using an inverse-square relationship (which would greatly diminish distant fires' impact), it uses an inverse-linear relationship. For distances within 250 miles, the score is maximized, while for distances up to 650 miles, it decreases linearly to zero.

2. **Fire Size Scaling:** The fire size is scaled and normalized using a cube root, which helps moderate the effect of extreme fire sizes. A scaling factor (1e5) adjusts fire size to maintain consistency across a range of values.

3. **Baseline Impact for Large Fires**: For fires over 100,000 acres, a minimum baseline impact is added, ensuring that exceptionally large fires contribute to smoke impact even at greater distances.

This estimation approach allows the Smoke Impact Score to reflect both the intensity (size) and proximity of wildfires, aligning better with observed PM2.5 levels than a simple distance-based model would.

Insights from the Visualization

From the scatter plot, it is observed that Smoke Impact Score versus AQI is somewhat dispersed, with no clear linear relationship. This may be because the variation in AQI depends on other variables besides those of smoke impact emanating from fires, such as weather conditions, topography, and regulatory laws concerning air quality. It will, therefore, imply that while a higher Smoke Impact Score normally equates to higher AQI values, this may not be proportional in all cases.

Reflection Statement

One of the most important learning experiences in this project was coming up with a sound calculation method for Smoke Impact Score so as to effectively relate the impact of wildfires to values of Air Quality Index. At the beginning, I had considered it a relationship that would be straightforward regarding the size of fire versus distance from the fire to AQI. However, as I did start to research the topic and got to play with the model a bit more, the contributing factors of AQI range from extremely involved-not just from fire characteristics alone but also the set of external elements such as wind, topography, and regional air quality policies. This intuition alone completely framed my approach toward the problem at hand and really underlined the need to develop a model that would truly depict the subtlety.

Research and Development of the Smoke Impact Score

A major part of my research focused on identifying factors that could improve the accuracy of the Smoke Impact Score. While distance from the fire to the affected city and fire size were my primary considerations, I found that other variables, such as wind direction and speed, play a critical role in smoke dispersion and could significantly influence AQI readings. Although I couldn't incorporate wind data directly into this project due to data limitations, this finding helped me appreciate the complexity of smoke impact assessment and the limitations of a purely distance-based approach.

I also reused a geodesic distance calculation method to measure distances between wildfires and target locations. This technique, which calculates the shortest path on the Earth's surface, provided a more accurate distance metric than simple Euclidean calculations, particularly over larger distances. This method was instrumental in the model, as accurate distance measurements are essential to calculating a meaningful Smoke Impact Score.

Insights from Collaboration and Data Attribution

This project involved collaboration that enabled me to assess other people's perspectives and share my own ideas that might improve the model. Various insights by fellow students were helpful in the discussion and suggestion over threshold values for fire size and distance penalties that influenced my final model architecture. Application of cube-root normalization to the fire size was suggested by a peer, which is quite useful in such scenarios, as it reduces the effect of extreme values towards a more uniform system of scoring. Indeed, this proved to be a golden idea when the Smoke Impact Score was estimated from its relation and matched the observed AQI pattern much better.

Speaking of which, the data sources I used included publicly available AQI data; then, from this data, I actually discovered what neighboring counties in my city reported to corroborate whether or not AQI readings were correct. In one word, this was localized data that was reliable to use as the basis of validation; it helped me realize that Smoke Impact Score, though generally related to AQI, was not strongly related as expected. This actually made me investigate how other environmental factors influence AQI and further broadened my problem understanding.

Reflections on Findings and Next Steps

One unexpected outcome was realizing the lack of a significant correlation between the Smoke Impact Score and AQI. This prompted me to think critically about other influential factors that were not captured in this model, such as wind conditions and atmospheric pressure. This insight was valuable because it demonstrated the limitations of simplified models and highlighted the need for multi-variable analyses in real-world environmental studies. As a next step, I would like to explore integrating wind data, possibly through available meteorological APIs, to enhance the accuracy of the Smoke Impact Score.

Overall, this collaborative project taught me the value of incorporating multiple perspectives and reusing established techniques, like geodesic calculations, to tackle complex environmental data problems. I gained a deeper understanding of the limitations of single-variable models and developed a strong foundation for future studies on wildfire smoke impact and air quality.