Wildfire Exploratory Analysis

Wildfires have always presented a threat to living beings, property, and the natural ecosystem due to their frequency and intensity. As this continues to be an issue, effective management and mitigation of wildfire risk will require a more strategic allocation of resources. By leveraging historical wildfire data, this project seeks to identify patterns or trends that can help identify hotspots that may require more focus. Understanding the quarterly and yearly trends can also inform the timing of preventive measures like posting public awareness campaigns.

We are looking into multiple states like the Pacific Northwest, as well as Colorado and Texas. Colorado has 15 of the top 20 largest fires that occurred in the last 9 years since 2012. Between 2011 and 2020, the state had an average of 5,618 wildfires per year where 237,500 acres burned, and this only has been increasing (WFCA). According to Western Fire Chiefs Association, climate change has continuously been changing with the increase in temperatures and decrease in humidity, so wildfires have not only increased but also have been growing in intensity.

NOAA Dataset

The dataset we pulled originated from the National Oceanic and Atmospheric Administration (NOAA) Storm Events database (NOAA Storm Events database). The dataset contains years from 2014 to 2024, each row being an occurrence of a wildfire. Some of the columns include city, state, date, start and end time, duration, injuries and deaths, and finally property damage. The dataset also contains a zone ID (column “cz_fips”), which refers to the Public Forecast Zone a wildfire occurs in. This Public Forecast Zone either corresponds to a county, or a subdivided part of one. With these CSV files (11 in total), we create our initial "wildfires" dataset.

We also source information from NOAA Zone-County Correlation Files (NOAA Zone-County). We were motivated to do this in order to obtain location data - through joining via state and zone ID, we can retrieve the latitude and longitude of the centroid of the Public Forecast Zone, giving us a rough measure of where the fire was located. A limitation of this is that the centroid does not correspond to the actual location of the wildfire, but nonetheless it serves as an adequate approximation.

A notable roadblock was the lack of past versions of the Zone-County Correlation File- the current version of the website only holds this file for March and September 2024. We needed this information in order to obtain locations for past fires: some fires in our wildfire table are listed under Public Forecast Zones which have since been redefined. Attempting to join these rows to an up-to-date version of the Zone-County Correlation File will fail, as the zone no longer exists within that file. In order to remedy this, we made extensive use of the Wayback Machine (Wayback Machine). This allowed us to access past snapshots of the Zone-County Correlation File website. Using this method, we obtain Zone-County Correlation Files from 7 timeframes in total: March 2020, November 2020, March 2021, March 2023, September 2024, March 2024, September 2024 (up-to-date version, includes upcoming zone redefinitions). In total, we used 18 CSV and DBX files to create our final dataset.

Data Cleaning

In addition to work done on sourcing, we also conduct extensive data cleaning and preprocessing in order to compile a final version of the dataset for use in visualization and analysis. In particular, we conduct the SQL preprocessing in three stages:

1. **Joining** wildfire information to location information
2. **Remedying** granularity issues
3. **Storing** duration information without using the TIME data type in MySQL

Joining wildfire information to location information

After deduplicating each location table to ensure uniqueness in state, zone ID, latitude and longitude using GROUP BY, we obtain seven location tables, one for each timeframe. Next, we LEFT JOIN each of those seven location tables to the “wildfires” dataset, and use COALESCE, which takes in any number of inputs, returning the first non-null input. We list values for latitude, longitude and zone name from each location table as parameters into the COALESCE function. By listing values in reverse chronological order, if a value does not exist in the most recent location table, we use the next most recent location table as a fallback. Through joining in this manner, out of 3403 rows in the “wildfires” table, we obtain location information for all but 159 of them. With this join, we create an updated “wildfire_360” table.

Remedying granularity issues

Upon conducting initial duration analysis, we found that the maximum duration of one wildfire was 31 days. This struck out as a data issue to us, and upon further inspection we discovered a mistake in our granularity assumptions. While we assumed one row would correspond to one wildfire, we found that wildfires that took place over several months were split into several rows. Using GROUP BY over state, zone ID, and fields “episode_narrative” and “event_narrative” (news reports used to correspond rows that are reported as the same event), we were able to extract earliest start and latest end date, creating a consolidated “wildfire_360_v2” table with 3251 rows.

Storing duration information

Finally, we found the native MySQL TIME data type only stores a maximum of 839 hours, implying we would be forced to find another method to store the duration of a wildfire, which could (and indeed does) surpass this threshold. To remedy this we used the TO_SECONDS function, converting start and end dates to seconds since Jan 1, 1970, taking the difference to retrieve duration in seconds. Next, we divide this value by the number of seconds in a day (86400) and use FLOOR to get the number of complete days within the duration of a wildfire. We also take the duration modulo 86400 to get the remainder of the division, which returns the remainder, stored as a time interval using SEC_TO_TIME. We use this to create a "wildfire_360_v3" table.

Who can benefit from the data?

There are many parties who can benefit from this data. Emergency management departments could use this information to identify hotspots or trends to reallocate resources more effectively and improve evacuation planning. Politicians constantly update or create new policies- this data would give them reason to rethink the current wildfire prevention methods. Researchers and scientists can utilize this data to create their own predictive models to explore trends that go beyond looking at historical data.

How would the data help them to make better decisions?

Analyzing historical wildfire data will provide insights that can enhance the decision-making process for managing and mitigating wildfire risks. For states like Colorado, the analysis can identify high risk areas by seeing where to prioritize. Next steps would be strategically placing resources where needed such as equipment, funding, and firefighter services. This ensures that high risk areas have a quick response time during wildfire incidents.

Data visualizations

United States of America

Figure 1: The maps below, created using the Pyplot and Cartopy libraries, depict the number of wildfires across the United States, using longitude and latitude data to represent state zone centroids. The color of the dots indicates the duration of fires, ranging from 0 to 40 days, with the darkest colors representing the longest durations. The size of the dots reflects the number of fires in each zone, providing an accurate representation of wildfire intensity. Figure 1 illustrates the number of wildfires in the United States of America, particularly in the central to west coast regions, with states like Washington, Oregon, California, Texas, and Colorado experiencing the longest durations.

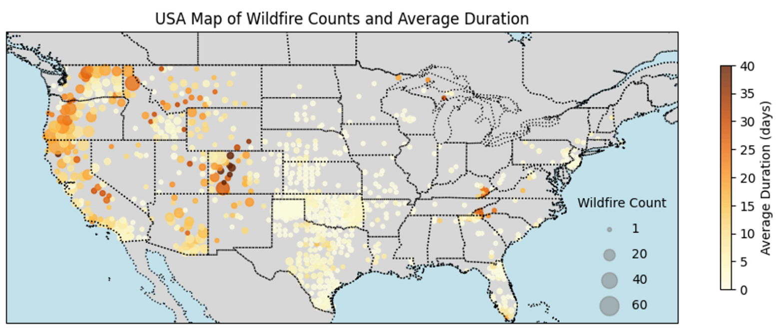


Figure 2: The map below analyzes wildfires based on seasonal variations, highlighting trends in their occurrence. California, Washington, and Oregon see frequent wildfires, likely due to consistent vegetation along the West Coast, while Texas and neighboring states have most fires in winter or spring. In contrast, the West Coast experiences more wildfires during summer or fall. The seasonal classification is based on the highest wildfire counts in each zone.

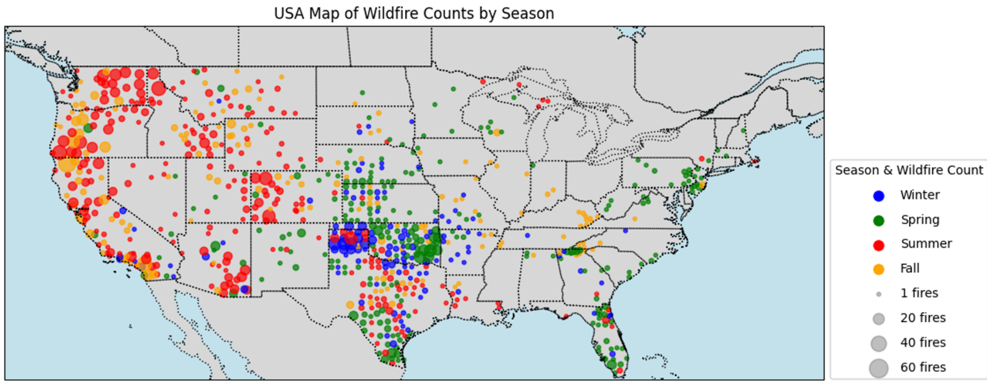


Figure 3: The graph illustrates the number of wildfires over the last ten years, excluding 2024. There was a steady increase in fires from 2014 to 2018, followed by a significant drop in 2019. This decline was primarily due to increased precipitation, cooler temperatures, and a higher snowpack, particularly in California, which often experiences severe wildfire seasons. These factors, along with effective fire management strategies, helped reduce the number and severity of fires that year. However, the number of wildfires surged again in 2020 and remained consistently high, with a notable decrease observed in 2023.

Figure 4: Analysis of seasonal wildfire trends reveals a consistent pattern across the years. Wildfire activity peaks in Q3 (summer) due to high temperatures and low humidity, which dry out vegetation and facilitate rapid fire spread. Q2 (spring) also shows significant wildfire activity due to warming temperatures and winds drying out vegetation, as well as potential agricultural burns. Wildfires are less common in Q1 (winter) and Q4 (fall) due to cooler temperatures, higher moisture levels, and snow in some regions, which act as natural barriers to fire spread.

Figure 4

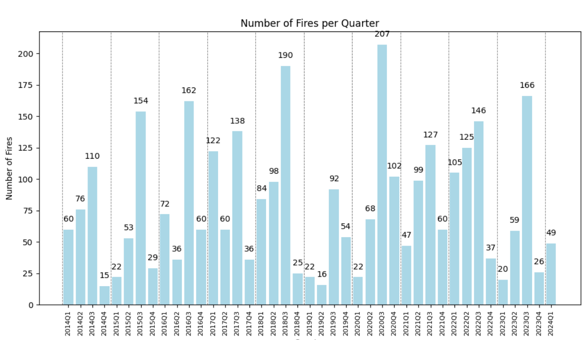


Figure 3

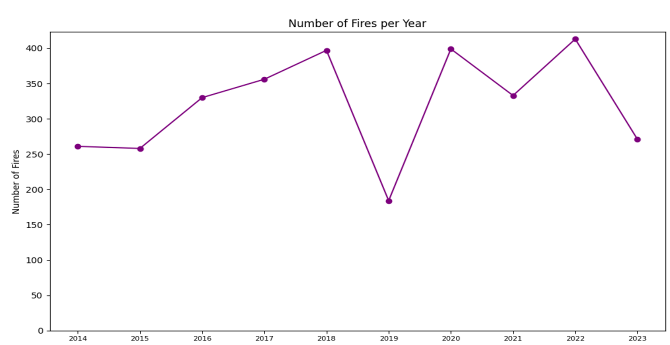
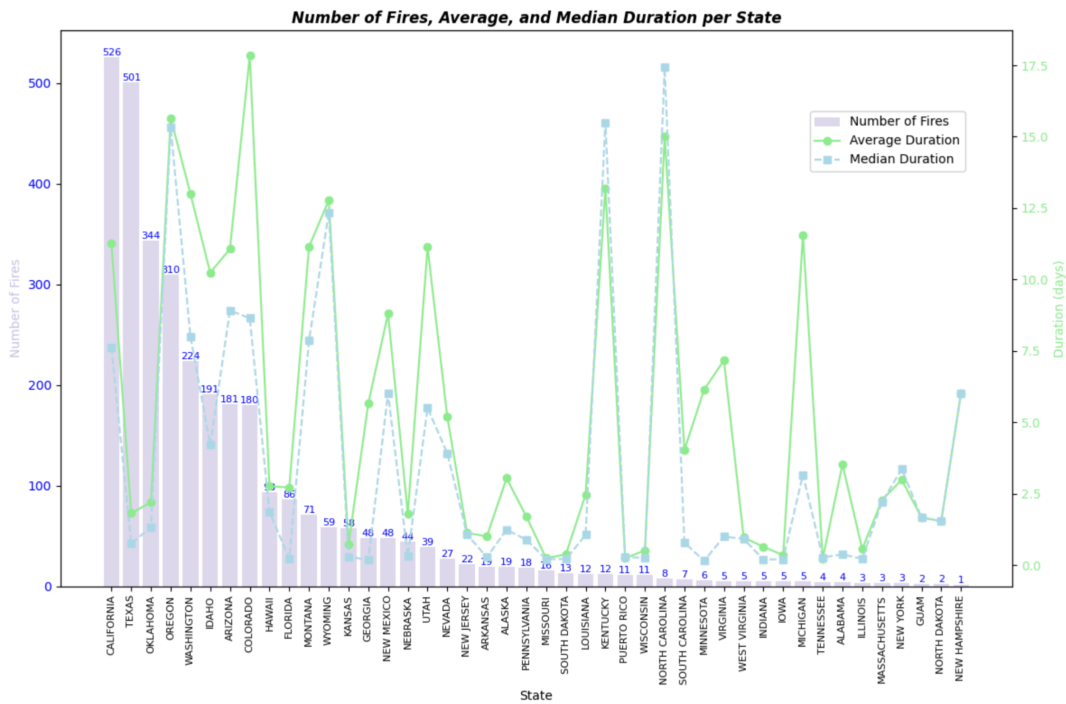


Figure 5: California experienced the highest number of wildfires, followed by Texas. This graph provides a state-by-state analysis of the number of wildfires, as well as the average and median durations for each state. We use this information to select four states for a more in-depth analysis.



CALIFORNIA

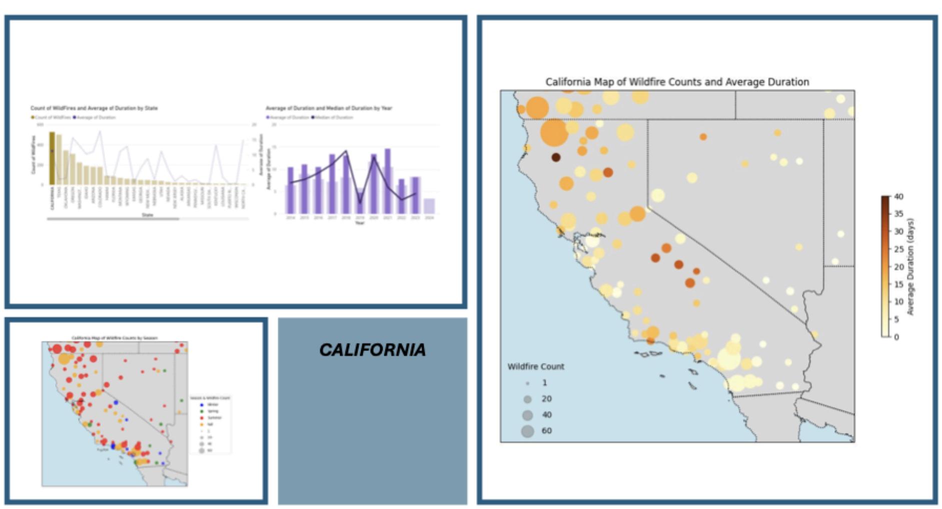


Figure 6: From 2014 to 2024, California experienced a significant surge in wildfire activity, driven by prolonged droughts, rising temperatures from climate change, and abundant vegetation that fueled fires. Key events include the 2017 Northern California fires, the 2018 Camp Fire, and the 2020 wildfires.

In 2019, wildfire activity notably decreased due to an unusually wet winter in 2018-2019, leading to increased precipitation and temporary relief from drought conditions. This surge in vegetation initially increased fire risk but ultimately reduced activity as the moist environment hindered fire ignition and spread. However, this drop in 2019 was short-lived. The overall trend of rising wildfire intensity resumed in subsequent years, influenced by cyclical droughts, changing climate conditions, and forest management practices, underscoring the persistent challenge of managing wildfire risks in California.

TEXAS

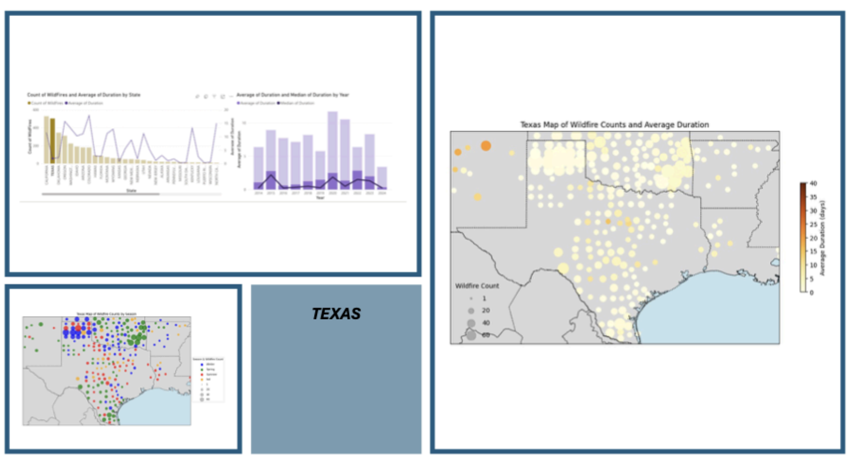


Figure 7: Texas, the state with the second-highest number of wildfires at 501, experiences notably shorter wildfire durations compared to California. The average and median durations for Texas wildfires are as brief as 3 days. This is in part due to the state's diverse vegetation and typically less dense fuel load compared to the heavily vegetated regions of the West Coast.

Wildfires in Texas predominantly occur in winter or spring, which is a period of increased lightning activity, a common ignition source for these fires. Unlike California's long and intense fire seasons, Texas wildfires are often short-lived due to quicker changes in weather conditions and the relatively lower accumulation of vegetation. Additionally, while Texas does experience significant fire events, the shorter duration of wildfires can be attributed to its more variable climate and effective fire management practices that help to control the spread.

The Pacific Northwest (Washington and Oregon)

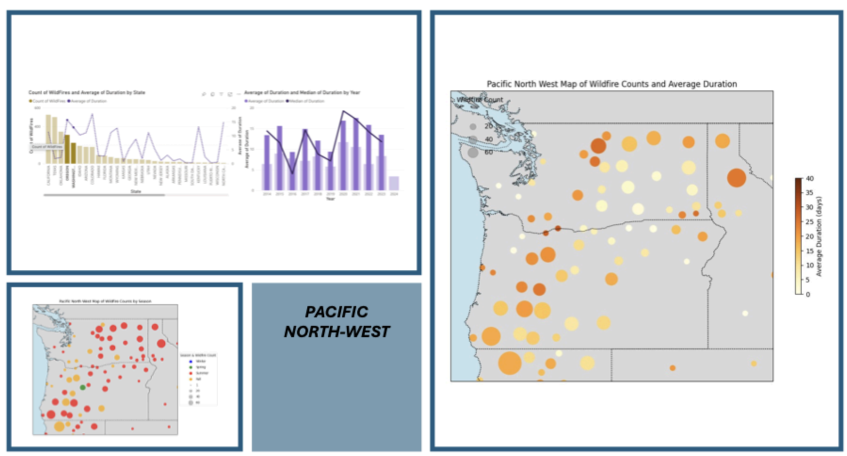


Figure 8: In the Pacific Northwest, particularly Washington and Oregon, wildfire activity is driven by both climate and land management. Private lands, managed by warehouses that farm trees, often have newly planted, single-species trees without natural sap, making them less fire-resistant. These closely planted trees create dense, flammable environments, and reduced vegetation diversity further weakens the ecosystem's resilience to fire. These conditions contribute to the increased frequency and intensity of wildfires in the region.

COLORADO

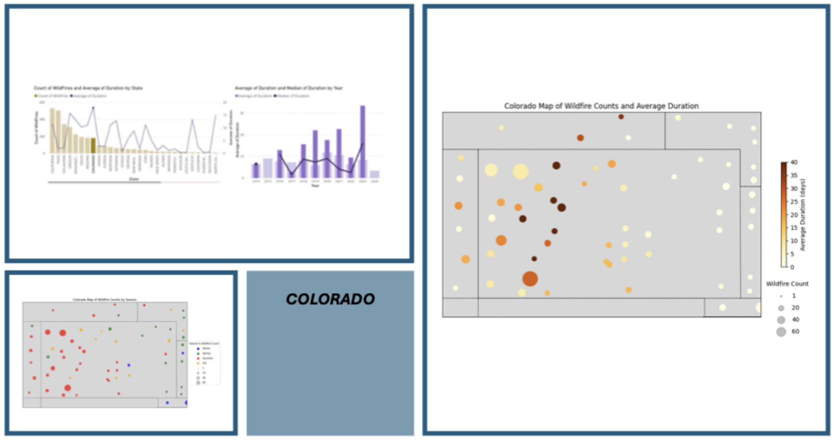
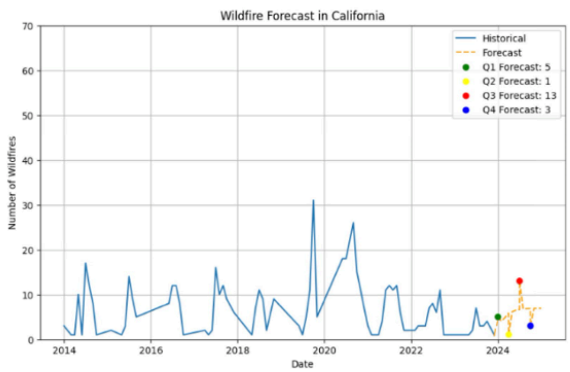
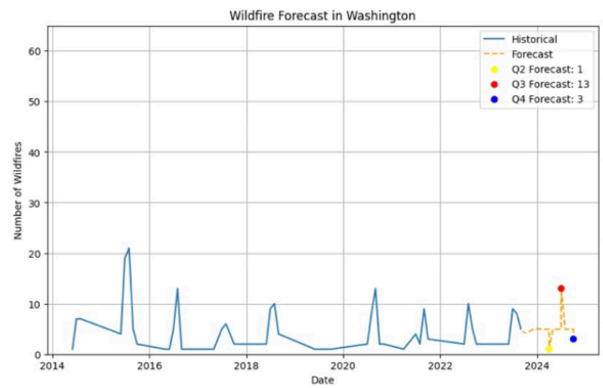


Figure 9: Colorado has faced increasing wildfire activity over the past decade, driven by rising temperatures, prolonged drought conditions, and an accumulation of dry vegetation. The state's fire season typically peaks in late summer and fall, aligning with its semi-arid climate. The Cameron Peak Fire, igniting in August 2020, emerged as one of the most significant wildfire events in Colorado's history. It became the largest wildfire on record, burning over 208,000 acres and persisting for several months until it was contained in December 2020. The fire's extensive duration highlighted the growing intensity and severity of wildfires in the state. Factors such as high temperatures, dry conditions, and strong winds contributed to the fire's intensity, which resulted in widespread destruction of both natural and developed areas. This event underscores Colorado's challenges in managing wildfire risks amidst evolving climate conditions and environmental pressures.

Forecast Analysis



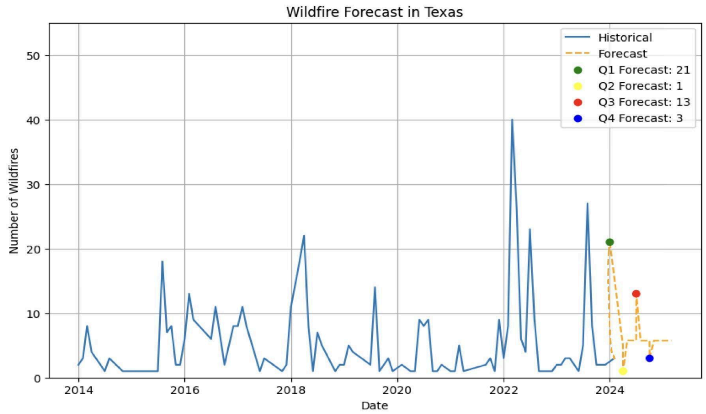
- Forecast Analysis of California Wildfires in 2024



- Forecast Analysis of Washington Wildfires in 2024

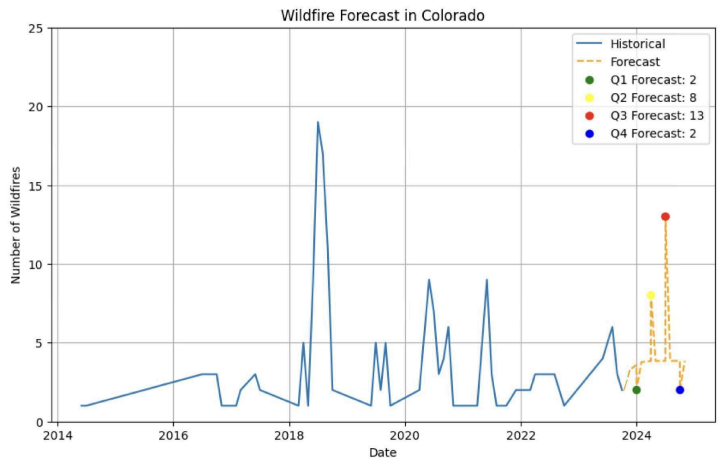
Forecast Analysis of California and Washington Wildfires in 2024

Figure 10 & Figure 11: In 2020, wildfires peaked due to extreme weather conditions. Compared to 2020, wildfire activity in 2021 decreased significantly due to increased precipitation and milder weather. Forecasting using an autoregressive integrated moving average (ARIMA) model, which assumes a linear relationship in the data, suggests varied wildfire activity for 2024. For California, Q3 is expected to be the most active period, aligning with its historical trend of high wildfire occurrences during the hot, dry summer months. Conversely, Q2 is forecasted to have only one wildfire, presenting an opportunity for preventive measures. In Washington, the forecast also predicts a significant increase in wildfires, particularly in Q3, with 13 wildfires anticipated. This aligns with the historical pattern of higher wildfire activity during the summer, driven by hotter and drier conditions.



Forecast Analysis of Texas Wildfires in 2024

Figure 12: The peaks in wildfire occurrences have often been followed by periods of lower activity, demonstrating seasonality and variability. In recent years (2021-2023), there have been significant spikes in wildfire numbers, particularly in 2022, indicating periods of extreme conditions conducive to wildfires. Forecasts for Q1 suggest a substantial increase in wildfires, implying that the early part of the year will be particularly active for wildfires in Texas. This forecasting is based on historical patterns of dry vegetation, favorable weather conditions, and human activities contributing to wildfire risk. As such, Texas is more likely to have wildfires in Q1 than in Q3 of other states. In Q1, dry vegetation and fuels, climate and weather conditions, and human activity led to more wildfires.



Forecast Analysis of Colorado Wildfires in 2024

Figure 13: The period from 2014 to 2017 saw relatively lower wildfire activity in Colorado due to favorable weather conditions, healthy vegetation, and fewer ignition sources. Around 2018, there was a notable peak in wildfire occurrences driven by a combination of environmental and climatic factors. In Colorado, high temperatures, drought conditions, dry lightning, wind events, and human activities increase the likelihood of wildfires in Q3. This period consistently shows higher wildfire risk due to these contributing factors.

What other data would be useful to have?

Geographical Data : Temperature, Rain, Wind, Humidity; parameters of the area where the fire occurred. The higher the temp, the stronger the fire can burn. The higher the humidity, the fewer chances of wood catching fire. Wind can push fires further and affect more areas. And rain can help extinguish the fire or decrease how strong the fire is.

Human Factors: Proximity to cities will help with severity analysis. And human traffic would show that the higher the number of people hiking the risk of starting a wildfire due to human intervention also increases.

What are your key insights and recommendations?

- ➔ Resource Allocation and Preparedness - Discuss opportunities to implement preventive measures during lower activity periods to mitigate the impact of the peak season.
- ➔ Emphasize the need for heightened preparedness and resource allocation; this includes staffing, equipment readiness, and community awareness programs
- ➔ Risk Management and Insurance - Suggest that insurance companies may need to adjust their risk assessments and pricing models based on the forecasted data
- ➔ Government and Policy Implications - Suggest the implementation or reinforcement of regulations aimed at reducing wildfire risks
- ➔ Highlight the importance of securing funding and support for firefighting agencies and emergency response teams
- ➔ Community Impact and Safety - Stress the need for continuous public awareness campaigns to educate residents about wildfire risks and safety measures
- ➔ Encourage communities to update and rehearse evacuation plans to ensure readiness in the event of a wildfire

Any extraordinary obstacles that you overcame

There were limited NOAA wildfire forecast zones so the amount of fires are not truly representative. Any wildfire going beyond 30 days was recorded as a new row. The MySQL TIME data type was inadequate for storing the duration of a wildfire (As discussed in the data cleaning process).

Reflections on how AI tools assisted you in the group project.

AI has played a significant role in wildfire analysis by enabling advanced data visualization techniques. For instance, projecting Seaborn scatter plots on maps allows for a clearer spatial representation of wildfire incidents. It helped understand basemap and cartopy, for this project we picked cartopy. Additionally, normalizing values using sklearn for hue on the map and coordinating these with a color bar or size legend has enhanced the understanding of data distribution and intensity.

Citations

Colorado Department of Public Safety. "Historical Wildfire Information." *Colorado Division of Fire Prevention and Control*, https://dfpc.colorado.gov/sections/wildfire-information-center/historical-wildfire-information. Accessed 29 July 2024.

Internet Archive. "Wayback Machine." *Internet Archive*, https://web.archive.org/. Accessed 29 July 2024.

National Centers for Environmental Information. "Storm Events Database." *National Centers for Environmental Information*, National Oceanic and Atmospheric Administration, https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/. Accessed 29 July 2024.

National Weather Service. "Zone-County Correlation File." *National Weather Service*, National Oceanic and Atmospheric Administration, https://www.weather.gov/gis/ZoneCounty. Accessed 29 July 2024.

Western Fire Chiefs Association. "Colorado Fire Season: In-Depth Guide." *Western Fire Chiefs Association*, 6 June 2023, edited 24 Apr. 2024, https://wfca.com/wildfire-articles/colorado-fire-season/. Accessed 29 July 2024.