

Understanding the Socioenvironmental Impact of LA City Wildfires

Vivian Sultan

College of Business and Economics
California State University
Los Angeles, USA
email: vsultan3@calstatela.edu

Xuehan (Harvey) Jing

School of Civil and Environmental Engineering and Earth
Sciences
Clemson University
Clemson, SC, USA
email: xuehanj@clemson.edu

Sirisha Mahesh

College of Business and Economics
California State University
Los Angeles, USA
email: smahesh@calstatela.edu

Christopher Lee

College of Business and Economics
California State University
Los Angeles, USA
email: clee@calstatela.edu

Abstract—This paper outlines a research plan to investigate how wildfires in Los Angeles affect different communities and the environment. We aim to use spatial data to develop equitable prevention and response strategies. The research will combine geographic information system datasets with demographic data, land-use records, and historical fire information to identify patterns that can support better preparedness and urban-resilience planning. The proposal emphasizes that wildfires are not just natural disasters; they are social events that worsen existing inequalities, particularly impacting marginalized communities through home destruction, poor air quality, evacuation challenges, and insurance burdens. It also highlights significant such environmental damage as deforestation and biodiversity loss.

Keywords—Wildfire; ArcGIS; California.

I. INTRODUCTION

This project focuses on exploring the socioenvironmental impacts of wildfires in Los Angeles, specifically how wildfire frequency, intensity, and spatial distribution intersect with population vulnerability, infrastructure exposure, and climate patterns. The key research question is, “How do wildfires in Los Angeles disproportionately impact certain communities, and how can spatial data be used to inform equitable prevention and response strategies?”

Los Angeles faces increasing wildfire risks due to climate change, prolonged droughts, and the expansion of the wildland–urban interface (WUI). These wildfires are critical because they reveal and exacerbate existing inequalities, disproportionately affecting marginalized communities through property destruction, hazardous air quality, and financial burdens. Wildfires also cause significant environmental damage, including deforestation, biodiversity loss, and soil degradation. We must understand these long-term consequences to guide urban planning, inform emergency management, and protect critical infrastructure. This issue also serves as a key case study for how large urban areas can plan for and adapt to new environmental realities as climate-related disasters become more frequent and severe, necessitating proactive, data-driven preventive strategies rather than reactive responses.

The research is intended for multiple key stakeholders who can act on the findings. The primary audience, urban

planners and policymakers, can help develop land-use regulations and emergency preparedness policies. We hope that such emergency-management agencies as the Los Angeles Fire Department, the California Department of Forestry and Fire Protection (Cal Fire), and the Los Angeles Office of Emergency Management will use such data as this to enhance readiness and create equitable evacuation, shelter, and recovery plans. Public health officials could also use these data to anticipate and manage the health implications of wildfires, including air quality and mental health effects on vulnerable populations. Perhaps academic researchers and students could use the model and methodology for interdisciplinary analysis. Nonprofits and community organizations could use also this analysis, especially those serving low-income and immigrant neighborhoods, to support advocacy for better access to resources and protection. Finally, this research could increase public awareness and individual preparedness through accessible maps and data storytelling.

Section 2 reviews the available literature. Section 3 outlines the data used in this research, and Section 4 details the applications used. Section 5 discusses the code used. Section 6 presents the analysis. Section 7 concludes.

II. LITERATURE REVIEW

Davies et al. [1] highlighted that environmental disasters often disproportionately affect minority and impoverished communities globally, and wildfires in the United States show similar patterns of outsized impacts on vulnerable populations. The research develops a social-ecological approach to characterize fire vulnerability, considering both wildfire potential and communities’ socioeconomic attributes. Davies et al. found that census tracts with majority Black, Hispanic, or Native American populations experienced approximately 50% greater vulnerability to wildfire compared to other census tracts. This underscores the need for management strategies that extend beyond technical solutions to incorporate socioeconomic and political considerations, especially tailoring emergency planning and mitigation strategies to address diverse populations and overcome such barriers as language differences.

The LA City Wildfires GeoHub [2] serves as a critical application for community awareness regarding wildfire

incidents and their potential impacts in Los Angeles. Designed to provide real-time status and information, this hub displays a list of fires with detailed information about their impacts on human and ecosystem populations. The data integrated into this application is sourced from authoritative entities like the Environmental Protection Agency and the U.S. Forest Service, ensuring reliability. For urban planners and emergency-management agencies, this GeoHub is an invaluable tool for understanding current wildfire situations, enabling more informed decision making and resource allocation.

Reid et al. [3] assessed the various health effects of wildfire smoke exposure, noting that consistent evidence links wildfire smoke to general respiratory issues, asthma exacerbation, and chronic obstructive pulmonary disease. Growing evidence also suggests increased respiratory-infection risk and overall mortality. For public health officials and the general public, this highlights the profound health implications of wildfires and the need for protective measures, though further research is still needed to clarify specific cardiovascular outcomes and identify particularly susceptible populations.

The Governor's Office of Emergency Services [4] outlined California's comprehensive strategy for reducing disaster losses and building resilience against a wide array of natural and human-caused hazards, including wildfires. Recognizing that disasters are increasing in frequency and intensity due to climate change and urban expansion, the plan emphasizes a proactive, integrated, and equitable approach to risk reduction. It detailed hazard assessments, identifies vulnerabilities across state assets, critical facilities, community lifelines, and disproportionately impacted equity-priority communities, and set forth goals, objectives, and mitigation actions to guide statewide efforts. The document highlighted existing state capabilities, such as strong building codes and various funding programs, and outlined processes for local government planning coordination to ensure unified and effective mitigation programs aimed at protecting life, property, and the environment.

This article uses satellite data and land-use analysis [5] to show that the fastest-growing U.S. regions are those bordering wildlands, leading to increased wildfire ignition and damage potential. For Los Angeles, with its flammable ecosystems, this study highlights the risks of expanding residential zones into fire-prone areas and demonstrates how spatial data can predict high-concern areas.

Shi et al. [6] argued for embedding social equity into all aspects of climate-resilience planning and identified how adaptation measures can unintentionally worsen inequality if social vulnerability is not integrated into spatial planning. This frames the project's intent to use wildfire risk data in conjunction with sociodemographic analysis to ensure proposed strategies are equitable and just.

The U.S. Forest Service [7] uses fire modeling, census data, and remote sensing to create national-scale wildfire risk maps. Its interactive layers, which include wildfire likelihood, flame length potential, and housing exposure, offer a technical framework adaptable to Los Angeles's neighborhood-level data using such geographic information system (GIS) software as ArcGIS or QGIS.

Cutter et al. [8] provided a valuable model for combining social data with environmental risk to develop composite indices. Their approach, dataset normalization, weighted scoring, and index mapping, offers methods that can be incorporated to visualize vulnerability alongside fire intensity or frequency for Los Angeles.

III. DATA REVIEW

The City of Los Angeles [2] is a primary source for wildfire data. It provides current status and potential impact of wildfire incidents and displays a list of fires with details on their effects on human and ecosystem populations. The data in this hub is sourced from authoritative entities and includes layers from such agencies as the Environmental Protection Agency, the National Weather Service, and the U.S. Forest Service. Its interactive layers, wildfire likelihood, flame length potential, and housing exposure, offer a technical framework for analysis.

We also use the Social Vulnerability Index (SVI) developed by the Centers for Disease Control and Prevention based on data from the U.S. Census Bureau. The SVI incorporates 15 census variables, including poverty, lack of vehicle access, and overcrowding, to rank vulnerability by census tract. This methodology is ideal for overlaying with fire hazard zones to identify communities most at risk in Los Angeles. Existing studies have shown how SVI can be used in conjunction with fire risk data to understand disproportionate impacts.

Land-use records from LA City's planning department provide information on how land is used across Los Angeles, including urban development, WUI zones, and other land classifications. Understanding land use is crucial as rapid growth in WUI zones increases wildfire risk and damage potential.

Historical fire occurrence layers (from Cal Fire or similar authoritative sources) provide records of past wildfire ignitions, perimeters, and behavior. This historical data is essential for identifying patterns, understanding fire frequency, and conducting spatial clustering analyses to predict wildfire ignition hotspots [9].

Climate data from the National Oceanic and Atmospheric Administration and state-level climate assessments, including temperature, precipitation, and drought, is vital as these factors influence wildfire frequency, intensity, and spatial distribution. Climate change is recognized as increasing wildfire risks.

Infrastructure exposure from LA City's Public Works and utility providers (e.g., transportation networks, utilities, emergency-services facilities) are used to assess exposure to wildfires. This helps understand how wildfires interact with the city's social geography and infrastructure.

California fire perimeters from the California Natural Resources Agency (CNRA) Geportal provide the geographical boundaries (perimeters) of historical wildfires across California. This foundational layer offers precise spatial and temporal information about past fire events. Each perimeter typically includes such attributes as fire name, year, acres burned, and sometimes ignition cause. This dataset is directly relevant to understanding wildfire frequency, intensity, and spatial distribution in Los Angeles.

For emergency management agencies, this data is indispensable for identifying areas with a history of recurrent fires, informing fuel management and response planning. Urban planners and policymakers can use it to assess WUI expansion risks and guide zoning decisions in fire-prone areas. Academic researchers and students will find it crucial for developing predictive models and understanding historical fire ecology. For the general public and local residents, visualizing past fire perimeters can significantly increase local-risk awareness.

IV. APPLICATION REVIEW

The analysis of California wildfire data, as outlined across `Capstone3.ipynb`, `Capstone4.ipynb`, and `Capstone5.ipynb`, leverages a powerful suite of Python libraries and functions to comprehensively understand wildfire patterns and their socioenvironmental impacts, and to inform equitable prevention and response strategies. This multifaceted approach is tailored to provide actionable insights for urban planners, emergency management agencies, public health officials, academic researchers, nonprofits, and the general public, all of whom are critical stakeholders in building resilient communities.

At the foundational level, data management and preparation are primarily handled by the `pandas` library. Across all notebooks, `pd.read_csv()` and `pd.read_excel()` are instrumental in loading the raw wildfire perimeter data from sources like the CNRA Geoportal. Once loaded, `pandas` DataFrames become the central structure for organizing this complex tabular data. Crucial initial steps involve data cleaning and transformation: `df.dropna()` handles missing values in such critical columns as latitude, longitude, and acres burned, ensuring data quality for subsequent analyses.

`pd.to_datetime()` converts date columns, such as `Started`, into `datetime` objects, which is essential for any temporal analysis, such as identifying monthly or yearly trends.

Furthermore, `pandas` facilitates feature engineering by allowing the creation of new columns—for example, extracting month names from start dates or generating binary flags like `FatalitiesOccurred`, `CausedDamage`, and `FullyContained`. This data-preparation phase ensures that the raw information is robust, complete, and in the correct format for deeper statistical and geospatial exploration.

Following data preparation, exploratory data analysis and visualization play a vital role in revealing key patterns and insights. The `matplotlib.pyplot` and `seaborn` libraries are central to this phase. `matplotlib` provides granular control over plot customization, allowing for clear titles, axis labels, and overall layout adjustments using functions like `plt.figure()`, `plt.title()`, and `plt.tight_layout()`. Building on this, `seaborn` offers high-level interfaces for creating attractive and informative statistical graphics. For instance, `sns.countplot()` is employed to visualize the number of wildfires by month or year, effectively highlighting seasonal peaks and long-term trends. This information is indispensable for emergency management agencies to allocate seasonal resources and for the general public to understand peak risk periods. `sns.heatmap()` is used to generate correlation matrices between numerical variables

like `AcresBurned` and response resources (e.g., `PersonnelInvolved`, `CrewsInvolved`, `Engines`), revealing how fire size relates to resource demand. Furthermore, `scipy.stats` contributes to a deeper statistical understanding of fire-size distribution, with `skew()` and `kurtosis()` functions quantifying the asymmetry and *tailedness* of the `GIS_ACRES` data. This is crucial for emergency management agencies and policymakers to understand the probability of outlier megafires, while `pandas.DataFrame.describe()` and sorting functions identify top-10 largest fires (Figure 1), providing concrete examples of extreme events for all stakeholders.

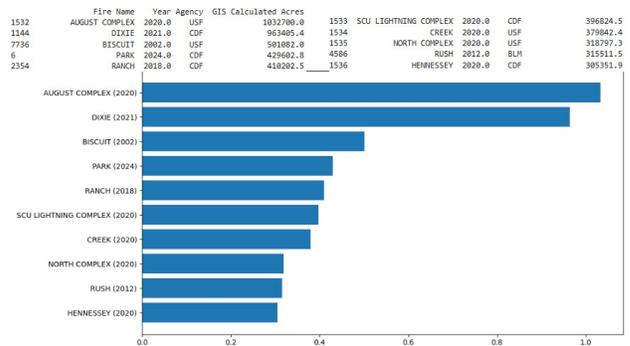


Figure 1. Top-10 largest wildfires in California by GIS-calculated millions of acres burned

For advanced spatial analysis, the `geopandas` library is a critical extension to `pandas`, enabling direct manipulation and analysis of geospatial data. `geopandas` treats the wildfire perimeters (from the CNRA dataset) as geographical objects (polygons) rather than just tabular data, fundamental for creating `GeoDataFrames` and performing essential spatial operations. For example, `geopandas` facilitates spatial joins and overlay analyses, allowing the project to combine fire perimeter data with such spatially explicit datasets as the SVI from the U.S. Census Bureau or critical infrastructure layers. This enables the identification of specific geographic areas where wildfire risk intersects with vulnerable populations or critical infrastructure, directly addressing the core research question of disproportionate impacts on communities.

Finally, for predictive modeling and map-based decision support, the `scikit-learn` library, combined with `folium`, forms a powerful predictive analytics pipeline. `sklearn.model_selection.train_test_split()` is used to correctly partition data into training and testing sets, ensuring the model's performance is evaluated on unseen data. Classification models, `sklearn.linear_model.LogisticRegression` and potentially `sklearn.ensemble.RandomForestClassifier` for comparison, are trained to predict such critical wildfire outcomes as whether a fire is likely to result in fatalities or structural damage. The `class_weight = balanced` parameter in `LogisticRegression` is particularly important for handling imbalanced datasets, common when predicting rare but critical events like fatalities.

Model performance is rigorously evaluated using `sklearn.metrics.classification_report()` with key metrics like precision, recall, and F1 score to assess the model's reliability in predicting these adverse outcomes. The `folium` library then plays a crucial role in visualizing these

predictions on interactive maps. By plotting predicted outcomes (e.g., red markers for predicted fatalities, green for no fatalities) on a base map of California, folium creates intuitive map-based decision-support tools, from which emergency management agencies and urban planners can identify predicted hotspots for future destructive or fatal incidents, allowing for proactive resource allocation, targeted education, and infrastructure protection. This comprehensive analytical framework, integrating data management, exploratory statistics, advanced geospatial analysis, and predictive modeling, is designed to generate actionable intelligence, fostering data-driven and equitable wildfire prevention and response strategies in Los Angeles for all target audiences.

V. CODE REVIEW

The comprehensive analysis of California wildfire data, as meticulously orchestrated across `Capstone3.ipynb`, `Capstone4.ipynb`, and `Capstone5.ipynb`, represents a multifaceted approach to understanding the socioenvironmental impact of wildfires and informing equitable strategies for the Los Angeles region. This integrated analytical pipeline, leveraging a powerful suite of Python libraries, is designed to provide explicit, actionable insights for a diverse target audience, including urban planners, emergency management agencies, public health officials, academic researchers, nonprofits, and the general public.

The initial phase of this robust analysis focuses on data acquisition, cleaning, and preparation, primarily powered by the pandas library, with crucial support from geopandas for spatial data. The process begins with the `pd.read_csv()` and `pd.read_excel()` functions used to load raw wildfire data, including the California Fire Perimeters dataset, into DataFrame structures. This foundational step ensures that all subsequent operations are performed on a well-structured tabular dataset. Data integrity is paramount, hence the rigorous cleaning. `df.dropna(thresh = threshold, axis = 1)` meticulously prunes columns containing an excessive percentage of missing values (e.g., more than 90%), preventing the inclusion of sparse or irrelevant data that could skew results. Furthermore, `df.dropna(subset = [Latitude, Longitude, AcresBurned])` explicitly removes rows with missing geographical coordinates or fire size, as these are nonnegotiable for spatial and quantitative analyses. `pd.to_datetime()` transforms raw date strings into proper date-time objects, a critical step for enabling any form of time-series analysis, such as identifying seasonal trends or multiyear patterns. For geospatial analysis, geopandas then converts standard DataFrames with latitude and longitude into GeoDataFrames. This transformation assigns a geometric property (points for fire ignitions, polygons for perimeters) and a coordinate reference system (CRS) to each record, making the data spatially usable. This is fundamental for accurately mapping wildfire incidents and for subsequent overlay analyses with other spatial datasets like social-vulnerability indices or infrastructure layers, allowing urban planners to assess WUI expansion risks and emergency management agencies to understand geographical spread.

Following preparation, exploratory data analysis and feature engineering delve deeper into the dataset's characteristics. pandas facilitates powerful feature engineering, such as extracting the month name from the started date-time column to categorize fires seasonally, enabling a clear visualization of peak fire activity. Crucially, such binary flags as `FatalitiesOccurred` (1 if any fatalities, 0 otherwise), `CausedDamage` (1 if structures destroyed, 0 otherwise), and `FullyContained` (1 if containment reached 100%, 0 otherwise) are engineered from raw numerical counts. These transformations are vital as they convert continuous or count data into clear, actionable target variables suitable for classification models, directly addressing specific research questions about fire impacts. `df.describe()` provides comprehensive descriptive statistics for numerical columns like 'GIS_ACRES' (acres burned), including count, mean, standard deviation, and quartiles, giving academic researchers and emergency management agencies a quick overview of fire scales and variability. To further characterize fire-size distribution, `scipy.stats.skew()` and `scipy.stats.kurtosis()` functions calculate skewness and kurtosis of GIS_ACRES. A high positive skewness, for instance, indicates that while most fires are small with a long tail of a few extremely large (mega) fires. High kurtosis suggests more extreme outlier fire sizes. Understanding these distribution shapes is critical for emergency management agencies to plan for outlier events, not just the average. Additionally, `pandas.DataFrame.sort_values()` and `head(10)` are used to explicitly identify and display the 10 largest fires by acres burned, highlighting extreme historical events for policymakers and the general public to grasp the immense scale of potential damage.

Visualizations are paramount in translating complex data into understandable insights for the diverse target audience. `matplotlib.pyplot` provides the base for plotting controls (`plt.figure()`, `plt.title()`, `plt.xlabel()`, `plt.ylabel()`, `plt.xticks()`, `plt.tight_layout()`) to ensure clarity and professional presentation. Building upon this, seaborn creates high-level statistical graphics. `sns.countplot()` generates bar charts showing the number of wildfires by month or fire frequency by year. These visualizations unequivocally confirm seasonal peaks (e.g., July–October in California), allowing emergency management agencies to precisely calibrate seasonal resource allocation and public health officials to anticipate periods of increased smoke exposure and related health risks. `sns.heatmap()` coupled with `df.corr()` helps visualize correlation matrices between AcresBurned and operational resources (PersonnelInvolved, CrewsInvolved, Engines). A strong positive correlation, explicitly displayed with numerical annotations and a color gradient, shows that larger fires are directly associated with greater resource deployment, offering insights into resource demand dynamics for emergency management and policymakers. `sns.scatterplot()` also helps visualize such relationships as fire size vs collection method, helping academic researchers understand potential biases or characteristics associated with different data-collection approaches for fires of varying scales.

Finally, the most impactful phase involves predictive modeling and geospatial decision support. The scikit-learn library provides the machinery for building predictive models. `sklearn.model_selection.train_test_split()` divides the features and target variables into training and testing sets, ensuring the model's performance is rigorously evaluated on unseen data for robust generalization.

`sklearn.linear_model.LogisticRegression` (and potentially `sklearn.ensemble.RandomForestClassifier` for comparative analysis as hinted in the proposal) is used to build classification models that predict binary outcomes, such as whether a wildfire is likely to result in fatalities. A critical detail is the `class_weight='balanced'` parameter in `LogisticRegression`, which explicitly addresses the inherent data imbalance (fatalities are typically rarer than no-fatality incidents). This weighting ensures the model gives sufficient attention to the minority class, preventing bias and improving its ability to correctly identify these critical, albeit infrequent, events. Model performance is then thoroughly assessed using `sklearn.metrics.classification_report()`, which explicitly provides precision, recall, and F1-score for each class (fatalities vs. no fatalities). High recall for the fatalities class is particularly crucial for emergency management agencies, as it indicates the model's effectiveness in minimizing false negatives (missing actual fatality events), which can have severe consequences. The crowning visualization for decision support is achieved using the `folium` library.

`folium.Map()` creates interactive base maps of California, onto which `folium.CircleMarker` objects are added for each predicted wildfire incident. These markers are dynamically colored (e.g., red for predicted fatalities, green for no fatalities) and include interactive popups displaying the prediction. This geospatial visualization directly translates complex model outputs into an intuitive and actionable format, allowing urban planners, emergency management agencies, and community organizations to visualize predicted hotspots for future destructive or fatal incidents. The use of `tempfile` and `webbrowser` further streamlines this by allowing the interactive maps to be saved as hypertext markup language files and opened in a browser for broader dissemination and collaborative decision making.

In essence, the synergistic application of these functions and packages across the notebooks creates a powerful, data-driven framework. From meticulous data preparation and detailed statistical profiling to advanced spatial visualizations and predictive modeling, the analysis aims to uncover hidden patterns, forecast critical outcomes, and provide tangible, geographically explicit insights. This comprehensive approach helps develop equitable, proactive, and resilient wildfire-prevention and response strategies in Los Angeles, directly addressing the multifaceted socioenvironmental challenges posed by wildfires.

VI. ANALYSIS

Figure 2 depicts wildfire occurrences across different months of the year, clearly indicating a pronounced seasonality in wildfire activity, unequivocally marking June through September as the peak wildfire season. Conversely,

the winter and early spring show a less favorable environment for large-scale fire ignition and spread.

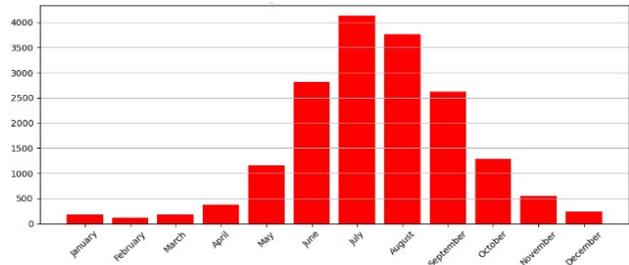


Figure 2. Number of Wildfires by Month All Years Combined (1878–2025)

A. Implications of Seasonality

This trend carries significant implications. Emergency management agencies can use this information for strategic resource allocation, ensuring that personnel, equipment, and aerial assets are optimally deployed and on heightened alert during the peak fire months. This data supports proactive operational planning and prepositioning of resources in high-risk areas to improve response and initial-attack effectiveness.

Urban planners and policymakers can leverage these insights to inform the timing of vegetation management, defensible-space initiatives, and public outreach, targeting efforts before the peak season. It also reinforces the need for stricter building codes and land-use regulations in the WUI.

Public health officials can anticipate periods of increased hazardous air quality, allowing them to proactively issue health advisories, activate clean-air shelters, and distribute N95 masks, especially in vulnerable communities. This visualization directly supports their efforts to manage wildfires' public health implications, both respiratory and mental health effects. Nonprofits and community organizations can use this clear seasonal pattern to time their preparedness workshops, evacuation drills, and resource distribution efforts, ensuring communities, particularly marginalized ones, are well informed and equipped before the most dangerous months arrive.

For academic researchers and students, this plot serves as a foundational piece of exploratory data analysis, confirming known seasonal patterns and potentially prompting further investigation into the specific climatic or anthropogenic drivers that make these months so prone to wildfires.

Finally, for the general public and local residents, this visualization provides an intuitive understanding of when their wildfire risk is highest, a powerful reminder for individual preparedness—maintaining defensible space, having an evacuation plan, and staying informed about air quality—reinforcing the idea that wildfires are social events that reveal and exacerbate existing inequalities.

B. Detailed Visualizations

The following visualizations offer critical insights into wildfires' geographical distribution and temporal trends, with particular relevance to understanding the context for Los Angeles. Each map serves a distinct purpose.

Figure 3 showcases the sheer extent and historical spread of wildfires across the entire state. This map conveys the widespread nature of fire activity. This broad view is essential for academic researchers and students to understand the macro-level distribution of fire events and for emergency management agencies to grasp the extensive historical fire activity that shapes landscape-level risk. For the general public and local residents, it serves as a powerful reminder of California’s inherent fire risk.

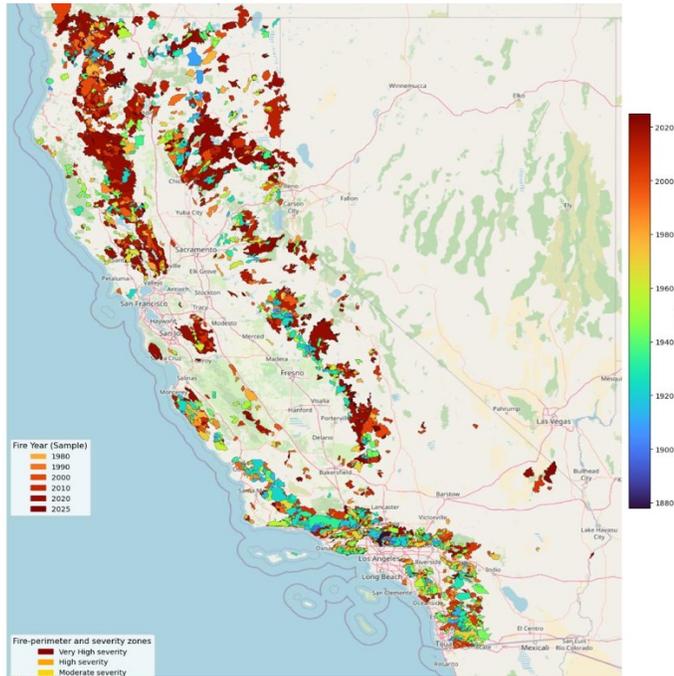


Figure 3. Largest wildfires in California by year (1878–2025)

To generate the visualization in Figure 3, we merged a wildfire perimeter map with a California wildfires basemap. The basemap provides the essential geographic context (state outline, major cities, roads, etc.) upon which these selected large-fire perimeters are overlaid. This combination offers a clear and impactful visual representation of where California’s most significant historical wildfires occurred with their size and year explicitly symbolized to convey scale and temporal context.

Figure 3 is particularly impactful as it combines temporal and spatial dimensions to highlight extreme events. This map explicitly reveals that exceptionally large fires are not just a recent phenomenon; they have occurred throughout history. For emergency management agencies, this visualization helps to identify regions consistently prone to mega-fires, informing strategies for prepositioning resources and developing large-scale evacuation plans. Urban planners and policymakers can use it to understand the historical scale of destruction and integrate long-term resilience measures into land-use planning, particularly in areas historically impacted by large fires. Public health officials can assess the long-term exposure risks in regions frequently experiencing large fires. Nonprofits and community organizations can leverage this figure to advocate for increased resources and awareness in

communities that have repeatedly borne the brunt of major wildfires. Academic researchers and students gain insight into the spatial distribution of historical fire extremes, which can inform studies on climate change impacts and fire ecology. For the general public and local residents, it underscores the immense scale and destructive potential of wildfires that recur over generations.

Figures 4 and 5 offer critical insights into the temporal trends of wildfire occurrences across the state. While both depict wildfire frequency over time, their distinct methodologies—explicit annual counting versus 10-year binning—serve different analytical purposes and convey unique messages to our diverse target audience.

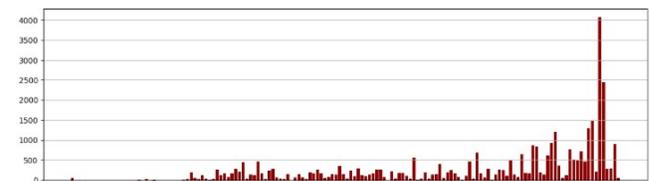


Figure 4. Total acres burned in California by year (1880–2020)

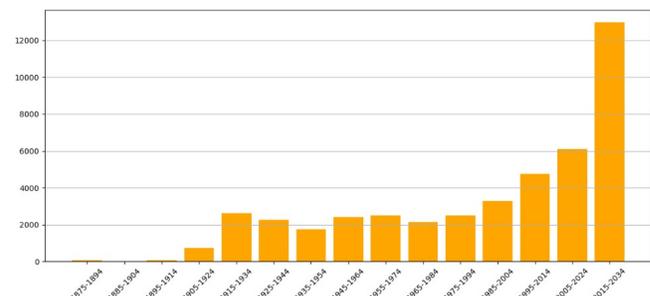


Figure 5. Total acres burned by wildfires in California (grouped by 10-year periods)

Figure 4 presents an explicit, year-by-year count of wildfires, providing a granular view of fire frequency and severity, allowing for the identification of specific years with unusually high or low fire activity and serving as a direct historical record. For emergency management agencies, this explicit annual count is invaluable for dissecting past wildfire seasons in detail, correlating fire numbers with specific climatic events (e.g., drought years), and evaluating the effectiveness of annual mitigation strategies or policy changes. Academic researchers and students can use this chart to identify precise anomalies for deeper investigation into their causes. However, for urban planners or the general public, the year-to-year variability can sometimes obscure broader, long-term patterns.

In contrast, Figure 5 aggregates wildfire counts into 10-year intervals, offering a smoothed visualization of trends over a longer period. This binning strategy effectively filters out short-term fluctuations, allowing a more macroscopic trend to emerge. This visualization is particularly powerful for observing long-term increases or decreases in wildfire frequency, indicating shifts in climate patterns or the cumulative impact of such factors as urban expansion into the WUI. This aggregated view informs long-term planning,

infrastructure investments, and broad policy adjustments aimed at addressing decades-long shifts in fire risk. Public health officials can observe decadal changes in fire frequency, which can correlate with long-term air-quality degradation or public health burdens. Nonprofits and community organizations can underscore the escalating nature of the wildfire problem in their advocacy efforts, highlighting the need for sustained and growing investment in resilience. While it sacrifices annual precision, this binned visualization excels at revealing the persistent, underlying patterns of wildfire activity, providing compelling evidence of a changing wildfire landscape that demands strategic, long-term responses from all stakeholders.

Figure 6 spatially and temporally visualizes significant wildfire events across California. Each colored polygon on the map represents the geographic perimeter of an individual wildfire. This map allows users to discern clear seasonal patterns and regional concentrations of major wildfires over time. The visualization directly aligns with the peak wildfire season identified in other temporal analyses, reinforcing the profound influence of hot, dry conditions during these periods. The spatial distribution of these fires, particularly their concentration in specific areas like Northern California, the Sierra Nevada, and Southern California mountain ranges, is clearly visible against the shaded topographic basemap.

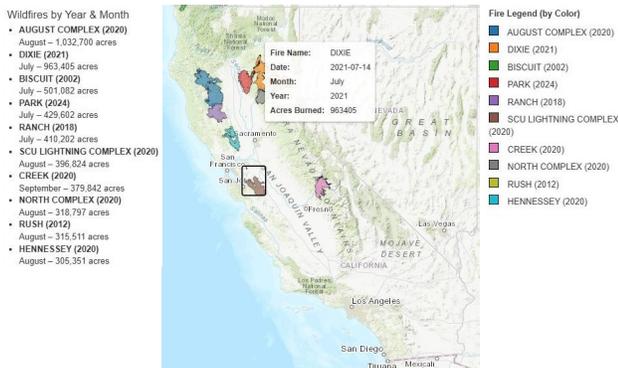


Figure 6. Wildfires by year and month (1878–2025)

This map carries substantial implications. Emergency management agencies can understand where and when historical mega-fires have occurred, informing proactive resource prepositioning and the development of large-scale, seasonally tailored evacuation plans for high-risk areas. Urban planners and policymakers can discern long-term fire patterns and their spatial footprints to guide resilient land-use planning, particularly in WUI zones, and to prioritize infrastructure hardening in historically impacted regions. Public health officials can assess long-term exposure risks in regions frequently experiencing large fires and plan for recurring public health advisories related to air quality. Nonprofits and community organizations can advocate for increased resources and sustained awareness campaigns in communities that have repeatedly borne the brunt of major wildfires. Academic researchers and students gain invaluable insight into the spatial distribution and temporal clustering of historical fire extremes, which can inform studies on climate

change impacts, fire ecology, and evolving fire behavior. Ultimately, the general public and local residents gain an understanding of wildfires’ historical scale and seasonality, the destructive potential of these events, and the imperative for ongoing preparedness.

Figure 7 provides a comprehensive breakdown of how different ignition sources have collectively contributed to wildfires in California over a century and a half, vividly underscoring the overwhelming role that human activity plays in wildfire ignition and subsequent acreage burned. The most dominant contributor, “Other Human Cause,” accounts for the highest cumulative burned area. This broad category encompasses a wide range of unspecified or miscellaneous human activities; a diverse array of human-related actions have been a major driver of extensive wildfire damage. Lightning ranks as the second most impactful cause, responsible for more than 11 million acres burned, making it the most significant natural-ignition source. While often smaller in scale, such intentional causes as arson, campfires, smoking, and illegal fireworks contribute significantly to fire ignition and spread. The limited acreage attributed to the indeterminate categories demonstrates fire investigators’ skills at categorizing fires.

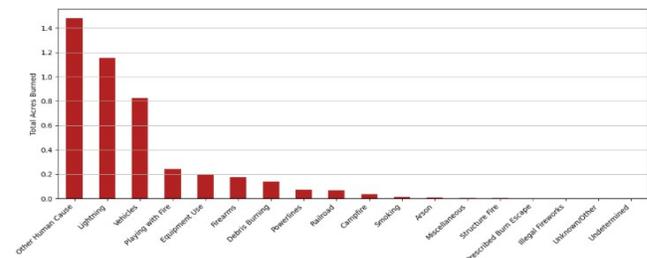


Figure 7. Total acres (10⁷) burned by wildfire cause (1878–2025)

This analysis of wildfire causes carries profound implications. Urban planners and policymakers receive clear evidence that human behavior is the predominant factor in large-scale wildfire destruction. This understanding calls for the development of targeted prevention strategies, stricter regulations, and community-outreach programs focused on reducing human-caused ignitions, particularly those falling under the other-human-cause umbrella. Emergency management agencies can use this breakdown to allocate resources more effectively for prevention campaigns, public-education initiatives, and specific enforcement actions aimed at high-risk human activities.

Public health officials can use this information to understand the human origins of smoke events, which can guide public health messaging and interventions aimed at behavioral change to reduce ignitions. Nonprofits and community organizations gain powerful evidence to support advocacy for comprehensive prevention programs. This empowers them to push for greater investment in community education and mitigation. Academic researchers and students find this chart a foundational piece for further study into the specific contexts and behaviors contributing to other-human-cause fires, or to model the impact of different prevention strategies. Finally, for the general public and local residents,

this visualization serves as a compelling reminder of their individual and collective responsibility in the effort to prevent wildfires, emphasizing that the majority of burned acres originate from human actions.

Figure 8 presents major wildfire events across California categorized by their underlying cause. Each colored polygon on the map represents the geographic perimeter of an individual wildfire with the cause indicated through color. The panel to the left lists the top five wildfire causes by total acreage burned with significant fires under each category.

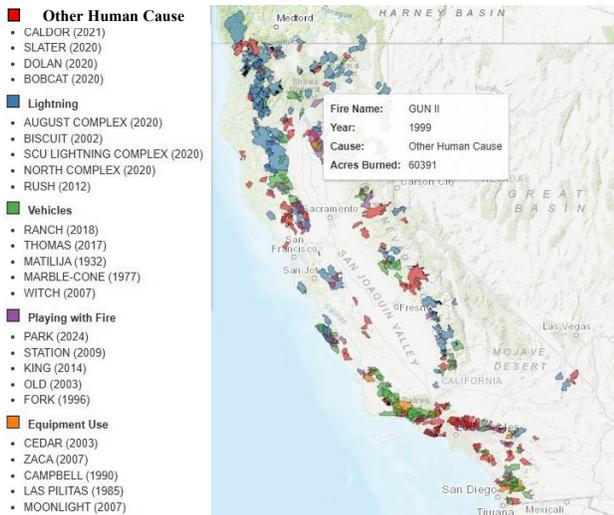


Figure 8. Wildfires by cause and year (1878–2025)

This map serves as a powerful analytical tool, revealing spatial patterns in wildfire causation and linking specific regions to dominant ignition sources. The visualization prominently displays a concentration of human-caused wildfires (red, green, purple, and orange) across the Central Valley, Southern California, and along major transportation corridors, directly reflecting the profound role of population density and infrastructure in wildfire ignitions. Conversely, natural causes (in blue), primarily lightning, dominate in the northern and central mountainous regions, such as the Sierra Nevada and Klamath ranges, aligning with areas prone to natural electrical activity. The background is lightly shaded to provide geographical context, aiding users in visualizing regional fire clusters relative to California’s diverse terrain.

This map carries substantial and actionable implications. Urban planners and policymakers gain invaluable insights into the specific human-related activities that trigger devastating fires in and around urban development, reinforcing the need for targeted education campaigns, stricter enforcement of fire-safe practices, and specialized land-use planning in areas prone to human-caused ignitions. Emergency management agencies can strategically allocate resources for prevention and response by understanding the prevailing causes in different regions. Public health officials can use this causal data to identify specific human behaviors that lead to widespread smoke events, informing public health messaging. Nonprofits and community organizations can leverage this compelling visual evidence to advocate for

specific policy changes or community programs tailored to address dominant ignition sources.

Academic researchers and students can use this map as a foundation for further investigation into the environmental, social, and policy factors contributing to specific fire causes and their spatial distribution. Finally, this visualization provides an intuitive and explicit understanding of the diverse origins of wildfires, encouraging the public to adopt fire-safe behaviors and recognize the specific risks prevalent in their local environments.

Figure 9 presents a statistical overview of the relationships between key wildfire variables, acres burned (fire size), cause code (ignition source), collection method (how data was gathered), and management objective (the fire’s management goal, e.g., suppression or resource benefit). Each cell displays the Pearson correlation coefficient.

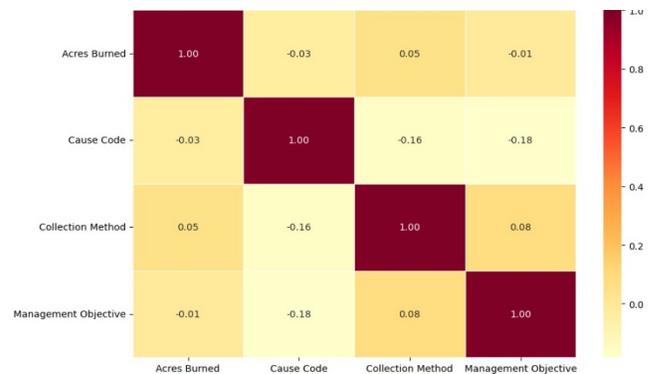


Figure 9. Correlation matrix: Fire size vs cause and objectives

The most critical insight derived from this matrix pertains to the relationships with acres burned. The correlation between acres burned and cause code is -0.03 , indicating a very weak and almost negligible negative linear relationship: The specific coded cause of a wildfire (e.g., lightning, equipment use, arson) has little direct linear influence on how large the fire ultimately becomes. Similarly for collection method and management objective.

These findings carry significant implications. For emergency management agencies, the most striking takeaway is that the ultimate cause, collection method, and initial management objective of a wildfire do not strongly predict how large the fire will grow. This strongly suggests that other dynamic factors not present in this matrix likely exert a far more substantial influence on fire growth and ultimate size. This insight directs resources and research toward these unexamined variables for more effective fire prediction and containment. Regulators should recognize that, while prevention efforts targeting specific causes remain vital (as human causes dominate total acres burned), policy changes based solely on cause or initial management strategy may not directly influence acres burned.

This matrix challenges academic researchers’ and students’ assumptions and explicitly points to areas where more complex modeling is needed. It suggests that linear relationships between these specific variables and fire size

are weak, prompting the exploration of nonlinear relationships, interaction effects, or the inclusion of more nuanced environmental and operational data. This analysis underscores the complexity of wildfire dynamics, informing nonprofits' and community organizations' advocacy efforts that move beyond simple cause-and-effect narratives to promote comprehensive, multifactor risk-mitigation strategies. Finally, for the public, the matrix reveals that identifying a fire's cause might be important for prevention, but it provides little insight into how large a fire will become, reinforcing the message that preparedness must consider a wide array of unpredictable factors.

Figure 10 presents a spatial visualization of wildfire events across California, explicitly grouping them into three broad cause categories—natural causes, human causes, and unknown causes—based on predictions from a random forest classification model. Each colored region on the map represents a predicted wildfire event, with its shade indicating the most likely cause as determined by the model.

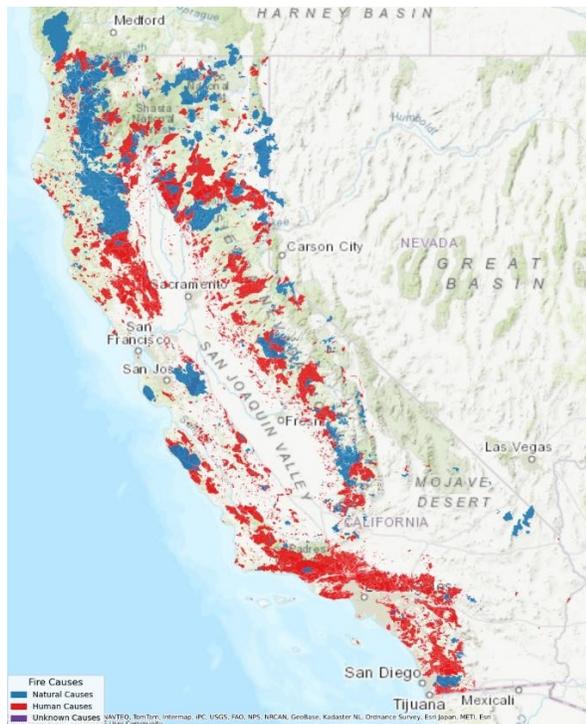


Figure 10. Predicted wildfire causes by category (1878–2025)

This map helps explicitly identify spatial patterns in wildfire causation, informing fire-prevention strategies, resource allocation, and public policy. The visualization reveals a prominent concentration of human-caused wildfires across the Central Valley, Southern California, and around major transportation corridors. This pattern reflects the significant role of population density, infrastructure, and human activity in triggering ignitions in these regions. Conversely, natural causes are predicted to dominate in the northern and central mountainous regions, such as the Sierra Nevada and Klamath ranges. Fires categorized under

unknown causes (purple areas) appear sporadically across the state, pointing to data gaps or ambiguous reporting.

This map carries actionable implications. Urban planners and policymakers gain invaluable insights into the specific types of ignitions most prevalent in different geographic areas, especially near urban and WUI zones, informing such highly targeted prevention policies as stricter regulations on equipment use in high-risk human-caused areas or specialized public awareness campaigns addressing common human behaviors that lead to ignitions. Emergency management agencies can strategically allocate resources for prevention and response by understanding the prevailing causes in different regions.

Public health officials can use this cause data to identify specific behaviors that lead to widespread smoke events, guiding public health messaging and interventions aimed at behavioral change. Nonprofits and community organizations can leverage this compelling visual evidence to advocate for specific policy changes or community programs tailored to address the dominant ignition sources in their respective regions.

Academic researchers and students can use this map as a foundation for further investigation into the environmental, social, and policy factors contributing to specific fire causes and their spatial distribution. Ultimately, this visualization provides an intuitive and explicit understanding of the diverse origins of wildfires in the state and local areas, encouraging the public to adopt fire-safe behaviors and recognize the specific risks in their local environments.

Figure 11. provides a crucial spatial and seasonal forecast of wildfire events across the state, explicitly categorizing them by their most likely season of occurrence based on a machine-learning model. Each colored polygon on the map represents a predicted wildfire event, with its shade indicating season.

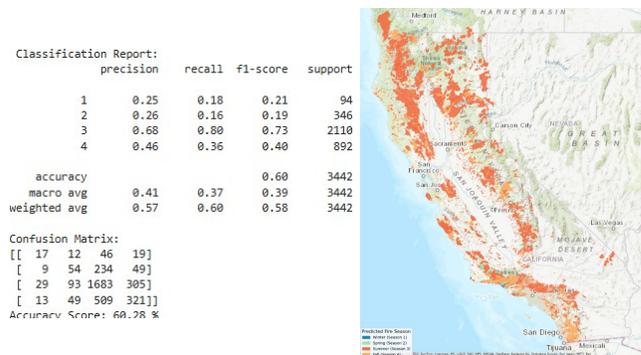


Figure 11. Predictive analysis: Random forest model

This visualization vividly reveals clear and actionable seasonal patterns. It explicitly shows that summer and fall fires dominate large regions of Central and Southern California, particularly within the San Joaquin Valley, Sierra Nevada foothills, and coastal ranges. This aligns perfectly with California's historical peak fire seasons, driven by prolonged dry conditions, high temperatures, and seasonal winds. Conversely, spring and winter fires appear less frequent and are more localized. The map serves as a powerful analytical tool for

understanding wildfire seasonality and regional concentration, derived from the model's learning of such historical fire attributes as cause, management objective, agency, and geographic unit.

For our diverse target audience, this map, coupled with the underlying model's performance, carries significant prescriptive implications. Emergency management agencies gain invaluable foresight: firefighting personnel, equipment, and air support can be strategically prepositioned in summer- and fall-prone regions before peak fire months. This proactive approach optimizes resource allocation and enhances early-response capabilities. Urban planners and policymakers can leverage these explicit spatial-seasonal risk patterns to inform the timing of prescribed burns and vegetation-management efforts, prioritizing these activities during lower-risk winter and spring months to reduce fuel loads. They can also enhance infrastructure hardening, such as grid resilience, in areas predicted to experience seasonal fire surges.

Public health officials can anticipate periods of heightened smoke exposure based on these predicted seasonal shifts, allowing for targeted public health advisories and the activation of clean air shelters during predicted high-risk seasons in specific regions. Nonprofits and community organizations receive a clear visual tool to intensify community preparedness campaigns and evacuation planning efforts, especially in summer and fall zones, well in advance of the predicted peak seasons. This explicit mapping helps them effectively advocate for resources in communities with high seasonal wildfire risk.

Academic researchers and students can use this predictive map as a foundation for further model refinement, exploring how incorporating additional climate data or socioeconomic factors might improve seasonal predictions. Ultimately, for the general public and local residents, this visualization provides an intuitive and explicit understanding of when and where the greatest wildfire threats are predicted seasonally, empowering them to take timely individual-preparedness actions and reinforcing the concept of aligning response strategies with spatially seasonal risk patterns to proactively reduce wildfire impacts.

C. Prescriptive Analysis Summary

The comprehensive analysis of California wildfire data throughout this project provides a robust, data-driven foundation for prescriptive actions aimed at mitigating the socioenvironmental impact of wildfires and fostering equitable resilience in Los Angeles and across the state. Recognizing that wildfires are complex phenomena exacerbated by climate change and urban expansion, the insights derived from meticulously cleaned data, temporal and spatial visualizations, and predictive modeling offer clear opportunities for proactive fire management and policy planning for all stakeholders.

Strategic resource allocation and operational timing stand as paramount prescriptive opportunities. The consistent evidence of a pronounced wildfire season, explicitly peaking in summer and fall (July, August, September, October), as demonstrated in Figure 2, calls for a strategic reorientation of firefighting personnel, equipment, and air support. These

critical resources should be prepositioned in high-risk, summer- and fall-prone regions, such as the Sierra Nevada, San Joaquin Valley, Southern Coast Ranges, and Inland Empire, as highlighted in Figure 14. Furthermore, the correlation analysis indicating that larger fires demand more resources underscores the importance of a swift and decisive initial attack. By understanding seasonal and spatial risk patterns, emergency management agencies can optimize readiness, ensuring that resources are deployed proactively, rather than reactively, to improve response times and initial attack effectiveness during peak fire periods.

Proactive policy timing and targeted community preparedness campaigns are equally vital. Given that other human cause and lightning account for the vast majority of total acres burned, with human causes overwhelmingly dominating, policy interventions must be acutely focused. Prescribed burns and vegetation-management efforts, crucial for reducing fuel loads, should be strategically prioritized during the lower-risk winter and spring months when environmental conditions are more favorable. This aligns with the seasonal predictions, allowing for controlled fire management activities without exacerbating risk. Simultaneously, community preparedness campaigns, including public outreach and evacuation planning, should intensify significantly before the onset of summer and fall, particularly in specific regions with high seasonal risk.

These campaigns must be specifically tailored to address the prevalent human ignition causes identified (e.g., safe equipment use, proper debris burning, preventing recreational fires) and to reach vulnerable communities—those with high SVI scores—to ensure equitable access to preparedness resources and clear evacuation plans.

Enhancing infrastructure resilience and fostering interagency coordination complete this prescriptive framework. Utility companies and local governments must prioritize infrastructure hardening, focusing on grid resilience and establishing robust defensible spaces around critical assets, particularly in areas predicted to experience seasonal fire surges. Figures 3 and 8 explicitly show the locations and causes of devastating fires, providing empirical justification for these investments. Furthermore, the complex and widespread nature of California's wildfire challenge necessitates unprecedented interagency coordination. Agencies overseeing jurisdictions with overlapping seasonal or causal threats should actively collaborate on joint-response protocols, mutual-aid agreements, and resource-sharing initiatives. This collective effort, informed by the detailed spatial and temporal analyses, ensures a unified and effective approach to wildfire mitigation across administrative boundaries.

This overarching analysis of California wildfire data, encompassing historical trends, spatial patterns, causes, and predictive modeling, provides a compelling mandate for a paradigm shift from reactive disaster management to proactive, data-driven resilience. By explicitly aligning response strategies with spatial-seasonal risk patterns and understanding the drivers of wildfire activity, urban planners, emergency management agencies, public health officials, nonprofits, and the general public can collectively optimize when, where, and how to deploy prevention and response strategies throughout the year. This prescriptive framework is designed not only to reduce the devastating impact of

wildfires on property and the environment but, more importantly, to protect human lives and foster truly equitable and resilient communities in the face of California's evolving and increasingly challenging wildfire reality.

VII. CONCLUSION

The comprehensive analysis of California wildfire data detailed throughout this project unequivocally reinforces that wildfires are profound socioenvironmental events extending far beyond mere natural occurrences. Our research, grounded in the understanding that Los Angeles faces escalating wildfire risks due to climate change, prolonged droughts, and urban expansion into the WUI, has systematically explored how fire frequency, intensity, and spatial distribution intersect with population vulnerability, infrastructure exposure, and climate patterns. The literature review explicitly highlighted that communities of color and other marginalized groups experience disproportionately greater vulnerability to wildfires, a finding central to our core research question: "How do wildfires in Los Angeles disproportionately impact certain communities, and how can spatial data be used to inform equitable prevention and response strategies?" This foundational understanding has driven every stage of our data-driven approach to support better preparedness, targeted risk mitigation, and more inclusive urban resilience planning for all stakeholders.

The exploratory data analysis has yielded critical insights into the temporal and spatial dynamics of wildfires, painting a clear picture of California's evolving fire landscape. Such visualizations as Figure 2 unequivocally demonstrate a pronounced seasonality, with a significant surge in incidents during late summer and autumn (July, August, September, October). This explicit monthly pattern is crucial for emergency management agencies to strategically allocate resources and for public health officials to anticipate periods of hazardous air quality. Furthermore, comparing Figures 4 and 5 reveals that while year-to-year counts can fluctuate, the binned view distinctly highlights a macroscopic trend of increasing wildfire frequency over longer periods, indicating the long-term impact of climate change and WUI expansion. This long-term perspective is vital for urban planners and policymakers to inform sustainable land-use regulations and infrastructure investments. Additionally, Figure 7 shows that other human cause and lightning are the leading contributors to burned acreage, with human causes significantly surpassing natural ones, underscoring the critical need for targeted prevention strategies that address human behavior.

Our advanced geospatial analysis and predictive modeling provide actionable intelligence crucial for proactive decision making. Figure 3 visually confirms the occurrence of exceptionally large fires throughout history, emphasizing their cyclical and increasingly frequent nature. This spatial-temporal understanding is invaluable for emergency management agencies in identifying regions consistently prone to megafires and for urban planners in integrating long-term resilience measures. The correlation matrix exploring fire size versus resource deployment offers insights into the reactive nature of current responses, indicating that larger fires generally require more personnel and equipment. More

critically, the predictive analysis, including Figures 13 and 14, use machine learning to forecast the most likely cause or season of wildfires based on historical data. Figure 13 explicitly highlights concentrations of human-caused fires near population centers and natural causes in mountainous regions. These predictive visualizations, derived from rigorous models, enable emergency management agencies to align resource allocation, policymakers to time prevention campaigns (e.g., prescribed burns in winter/spring), and community organizations to intensify public outreach before peak risk periods.

In conclusion, this comprehensive analytical framework, encompassing meticulous data acquisition and cleaning, robust exploratory data analysis, advanced spatial visualization, and machine-learning-driven predictive modeling, provides unparalleled insights into the socioenvironmental impact of wildfires. By explicitly identifying historical trends, current risk patterns, and predicting future wildfire characteristics, this analysis offers tangible, geographically explicit intelligence. For our diverse audience of urban planners, emergency management agencies, public health officials, academic researchers, nonprofits, and the general public, these findings are not merely academic; they are a critical toolkit for developing truly equitable, proactive, and resilient wildfire prevention and response strategies in Los Angeles. This work underscores the imperative for continued data-driven efforts to mitigate wildfire risks, protect vulnerable communities, and foster enduring urban resilience in the face of California's evolving environmental realities.

REFERENCES

- [1] I. P. Davies, R. D. Haugo, J. C. Robertson, and P. S. Levin, "The Unequal Vulnerability of Communities of Color to Wildfire," *PLOS ONE*, vol. 13, no. 11, art. e0205825, 2018, doi:10.1371/journal.pone.0205825.
- [2] City of Los Angeles, *LA City wildfires—GeoHub*, 2025, [Online]. Available from: <https://lacitywildfires-lahub.hub.arcgis.com> 2025.06.16 2026-01-24
- [3] C. E. Reid et al., "Critical Review of Health Impacts of Wildfire Smoke Exposure," *Enviro. Health Persp.*, vol. 124, no. 9, pp. 1334–1343, 2016, doi:10.1289/ehp.1409277.
- [4] Governor's Office of Emergency Services, *2023 California State Hazard Mitigation Plan*, 2023. [Online]. Available from: https://www.caloes.ca.gov/wp-content/uploads/Hazard-Mitigation/Documents/2023-California-SHMP_Volume-1_11.10.2023.pdf 2026-01-24
- [5] V. C. Radeloff, D. P. Helmers, H. A. Kramer, and S. I. Stewart, "Rapid Growth of the US Wildland-Urban Interface Raises Wildfire Risk," *Proc. Nat. Acad. Sci.*, vol. 115, no. 13, pp. 3314–3319, 2018. doi:10.1073/pnas.1718850115.
- [6] L. Shi et al., "Equity in Climate Change Adaptation Planning," *Nat. Clim. Chang*, vol. 6, no. 2, pp. 131–137, 2016, doi:10.1038/nclimate2841.
- [7] U.S. Forest Service, *Wildfire Risk to Communities*, 2020. [Online]. Available from: <https://wildfirerisk.org> 2026-01-24
- [8] S. L. Cutter, C. G. Burton, and C. T. Emrich, "Disaster Resilience Indicators for Benchmarking Baseline Conditions," *J. Homel. Secur. Emerg. Manag.*, vol. 7, no. 1, art. 51, 2010, doi:10.2202/1547-7355.1732.
- [9] A. D. Syphard et al., "Predicting Spatial Patterns of Fire on a Southern California Landscape," *Int. J. Wildland Fire*, vol. 21, no. 7, pp. 914–925, 2012, doi:10.1071/WF07087.