How can analytics be used to manage the risks associated with Wildfires in California?

Environmental Analytics Final Project

BEMM465

Table of Contents

Introduction	3
Overview of California Wildfire Incident	3
Data Exploration	4
Economic Loss Occurred	9
Analytics to Manage Risks	11
Use Case Scenario	13
Ethics	14
Recommendations	15
Limitations	17
Conclusion	17
References	18

Introduction

Wildfires are uncontrollable infernos that ravage through the natural landscapes of forests, grasslands, and brush, as well as human-made structures like homes and businesses. Typically instigated by natural causes such as lightning strikes, these conflagrations can also arise out of human negligence through activities such as campfires, cigarette butts, and arson. The environmental and economic toll of these disasters can be immense, with the destruction of habitats and damage to ecosystems often leading to devastating impacts on industries such as agriculture, forestry, and tourism. In addition, wildfires can have serious human repercussions, causing loss of life, injury, and the displacement of entire communities.

In response to these risks, a range of measures have been developed to control and mitigate their impact, including fuel reduction, fire suppression, and prescribed burning techniques. The latter involves the careful use of small, controlled fires to remove combustible material from vulnerable areas, ultimately limiting the likelihood of larger, more destructive blazes. Though natural wildfires play an integral role in many ecosystems, the recent increase in frequency and severity of these events has been widely attributed to a range of human factors, including climate change, population growth, and changes in land use. Given their significant impact, ongoing study of the causes and potential solutions for wildfires is imperative in developing effective strategies to manage the risks associated with these dangerous emergencies.

Overview of California Wildfire Incident

California has been plagued by numerous catastrophic wildfires throughout its history, with its fire season persisting longer and intensifying due to factors such as climate change and prior fire management techniques. The season of wildfires in the state usually takes place during the dry and gusty months of summer and autumn, with fires sparked by lightning strikes, electrical lines, or human activity. The impact of these wildfires on California's communities and ecosystems is enormous, with homes and businesses decimated, wildlife habitats obliterated, and air quality drastically affected. In recent times, some of the state's most significant and catastrophic wildfires have gained national and international publicity, including the 2018 Camp Fire, which emerged as the deadliest and most destructive blaze in California's history.

Various strategies, including prescribed burns, fuel reduction projects, and early warning systems, are employed in California to manage the risks associated with wildfires. The state agency responsible for fire prevention, suppression, and response, the California Department of Forestry and Fire Protection (CAL FIRE), directs these efforts. The California Wildfire Incidents dataset offers important insights into the reach and effects of wildfires in the state, as well as informing management strategies for these catastrophic incidents.

The questions that can be configured are:

- What is the average trend of a wildfire over the years in California?
- What are the causes of a wildfire in California?
- How has the number of structures destroyed or damaged due to wildfires changed over time?

These questions can be answered by performing various data analytics techniques such as data visualization, regression analysis, time series analysis, and descriptive statistics and high profile policy forming organization such as yours can do the needful according to the result of the different analysis.

Data Exploration

We are conducting an analysis based on data that we have found regarding the California Wildfires. The data used for this study is taken from a dataset found in Kaggle.com.

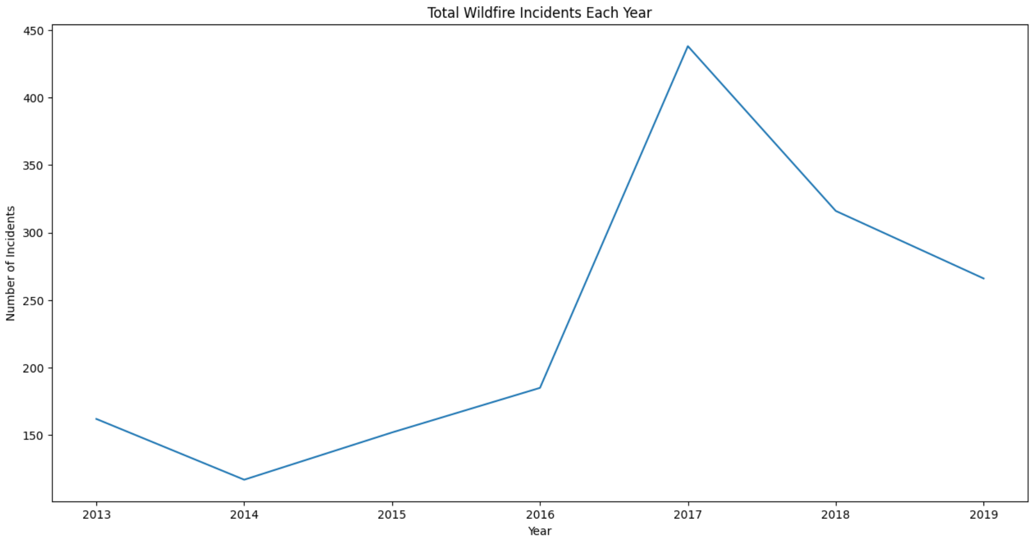
This dataset comprises a comprehensive list of Wildfires that have plagued California from 2013 to 2020. Not only does the dataset provide the location of where the wildfires occurred - including county name, latitude, and longitude values - but it also contains valuable information regarding the start of these wildfires. The data enables experts to generate insights into which locations within California are most at risk of fire outbreaks, the time when wildfires are most likely to occur, as well as allowing them to analyse the frequency and severity of these devastating events.

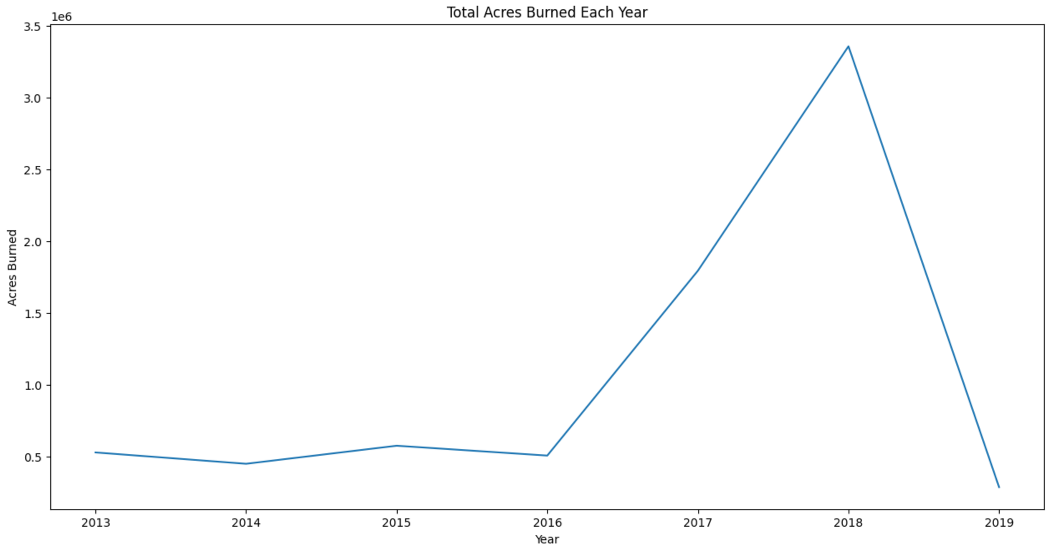
Here is a small snippet of the dataset taken:

AcresBurned	Active	AdminUnit	ArchiveYear	CalFireIncident	Counties	CountyIds	Extinguished	Featured	Final
257314	FALSE	Stanislaus National Forest/Yosemite National Park	2013	TRUE	Tuolumne	55	2013-09-06T18:30:00Z	FALSE	TRUE
30274	FALSE	USFS Angeles National Forest/Los Angeles County/CAL FIRE	2013	TRUE	Los Angeles	19	2013-06-08T18:30:00Z	FALSE	TRUE
27531	FALSE	CAL FIRE Riverside Unit / San Bernardino National Forest	2013	TRUE	Riverside	33	2013-07-30T18:00:00Z	FALSE	TRUE
27440	FALSE	Tahoe National Forest	2013	FALSE	Placer	31	2013-08-30T08:00:00Z	FALSE	TRUE
24251	FALSE	Ventura County Fire/CAL FIRE	2013	TRUE	Ventura	56	2013-05-11T06:30:00Z	FALSE	TRUE

We have performed exploratory data analysis (EDA) to understand the distribution of various features in the dataset and identified trends and patterns in the data that may help in managing risks associated with wildfires that will help in risk management.

1. Trends in the number of wildfires: We can examine the number of wildfires in California over time to identify any long-term trends in wildfire frequency. This information can be used to develop wildfire risk management plans that consider the likelihood of wildfire events occurring in the future.



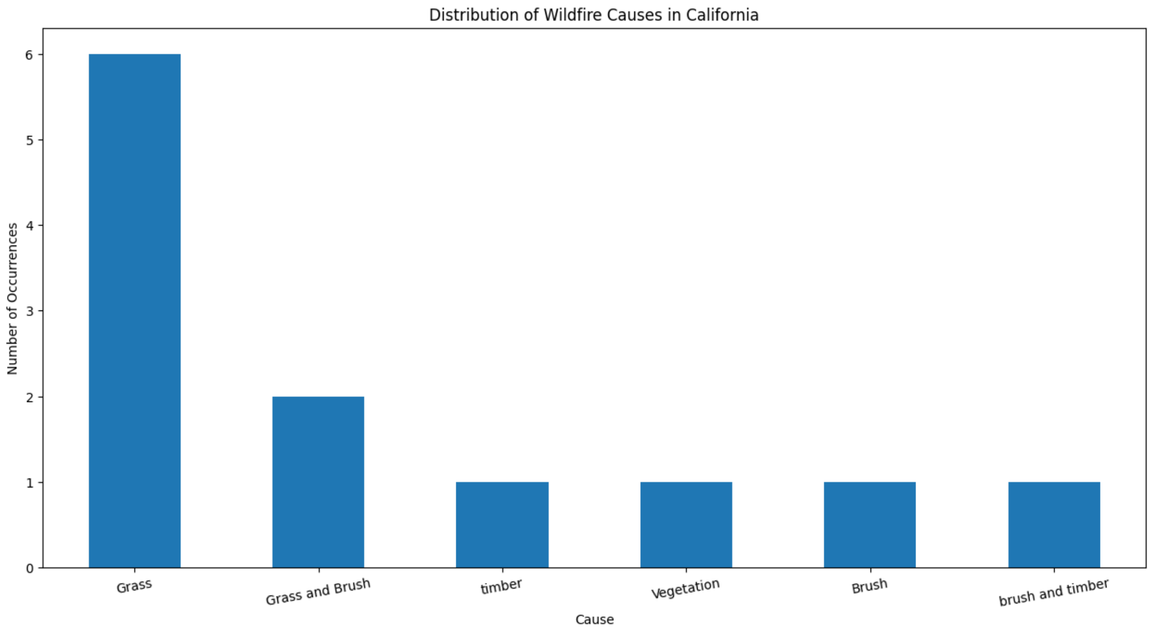


Insights:

The line plot's highest peak, which represents the total number of wildfire incidents that happened in 2017, indicates that 2017 had the most incidents ever recorded in the dataset. This might be because of various things, such the climate that year, how many people were present in places that were prone to wildfires, or modifications to fire management procedures. In a similar vein, the fact that 2018 has the highest peak in the line plot for the total number of acres burned indicates that this year saw the most wildfire damage in the dataset. Again, a variety of reasons, such as modifications in land use, climatic circumstances, or the efficiency of firefighting efforts, may have contributed to this. Overall, our observations emphasise how crucial it is to monitor and control wildfire outbreaks because they can have a large negative impact on both the ecosystem and human societies.

We can better understand the elements that contribute to these disasters and create measures to limit their consequences by analysing trends and patterns in wildfire data.

2. Causes of wildfires: We can identify the most common causes of wildfires in California and analyze any trends in the causes of wildfires over time. This information can be used to develop strategies to reduce the incidence of wildfires caused by human activities or natural factors. To ascertain the leading trigger behind the recent surge of wildfires in California, we can leverage the 'FuelType' column in the dataset and conduct an analysis of the frequency of occurrence for each plausible factor. We can then construct a highly informative graphical presentation through deploying a bar chart to convey a topographical depiction of the distribution of the contributing factors.

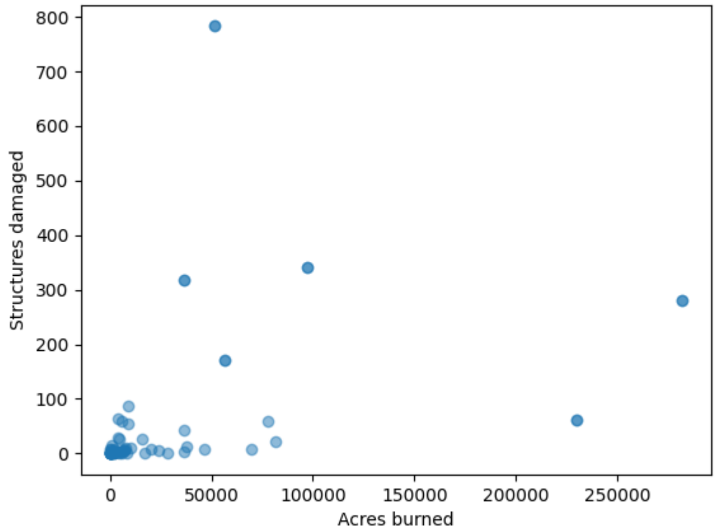


Insights:

In the bar graph produced by this code, "grass" has the highest bar as the output reason, indicating, according to the statistics, that grass fires are the most common cause of wildfires in California.

This knowledge can be useful for figuring out what causes wildfires to break out in California and for formulating plans to lessen or minimise the harm they cause. Researchers and decision-makers who are researching the effects of climate change, land management techniques, or other factors on wildfire frequency and severity in California may find it useful as well.

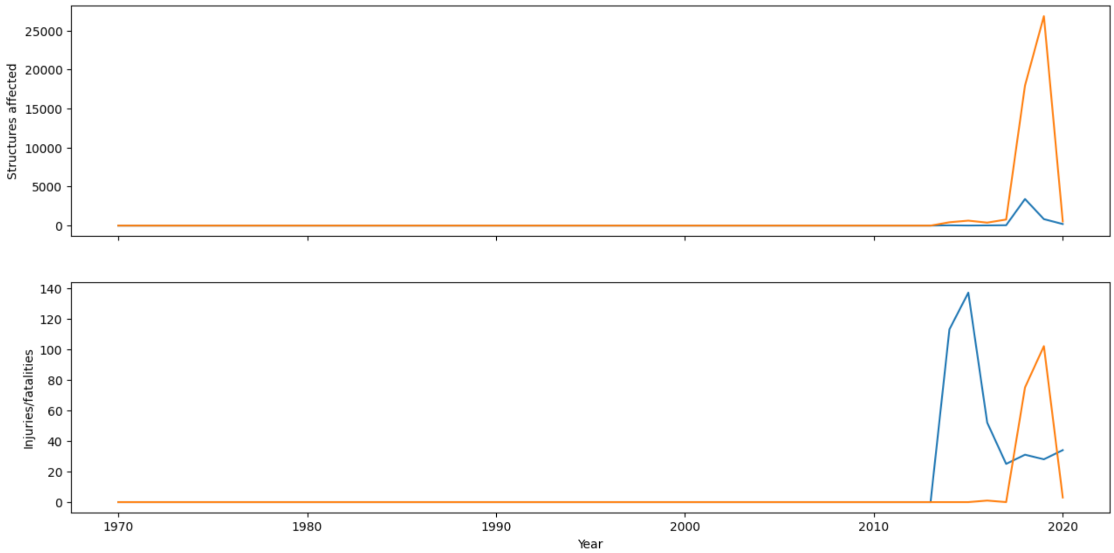
3. Patterns in wildfire size and severity: We can analyze the size and severity of wildfires and the structured damaged accordingly to identify any patterns or trends in the spread of wildfires. This information can be used to develop effective strategies for managing the spread of wildfires, including resource allocation and evacuation planning.



Insights:

The scatter plot reveals an apparent concentration of data towards the bottom-left corner while exhibiting a sporadic spread throughout. It indicates a weak positive correlation between Acres Burned and Structures Damaged, where the likelihood of an increase in the number of structures being damaged is higher when the quantity of acres burned rises. However, there is a great degree of variation in the relationship, suggesting that there might be additional influential factors that need to be taken into account. The presence of a cluster in the bottom left corner of the scatter plot implies the existence of a significant number of data points characterized by low values of both Acres Burned and Structures Damaged. This observation presents a possibility of several factors contributing to this pattern, such as small fires that can be easily contained or incidences that occur in areas with fewer structures. The seemingly arbitrary dispersion of residual data points indicates a considerable degree of variability within the association between Acres Burned and Structures Damaged. For instance, there might exist wildfires wherein numerous structures incur damage, albeit a comparatively minimal expanse is engulfed in flames. Conversely, there could be scenarios where vast regions are consumed by fire, yet only a handful of structures suffer damage. Employing a scatter plot to visually portray these interconnections aids in discerning any latent patterns or aberrations embedded within the dataset.

4. Impacts of wildfires on people and infrastructure: We can analyse the impacts of wildfires on people and infrastructure to identify areas that are at higher risk of damage or destruction in the event of a wildfire. This information can be used to develop evacuation plans and allocate resources to minimize the impact of wildfires on people and infrastructure.



Insights:

The red line in the top plot indicating structures affected shows a peak at 25000, while the blue line shows a peak below 5000. This suggests that some years witness significantly greater damage to buildings due to wildfires in California. The impact of various factors such as weather patterns, fuel supply, land management techniques, and other elements that affect wildfire behaviour may have contributed to this phenomenon.

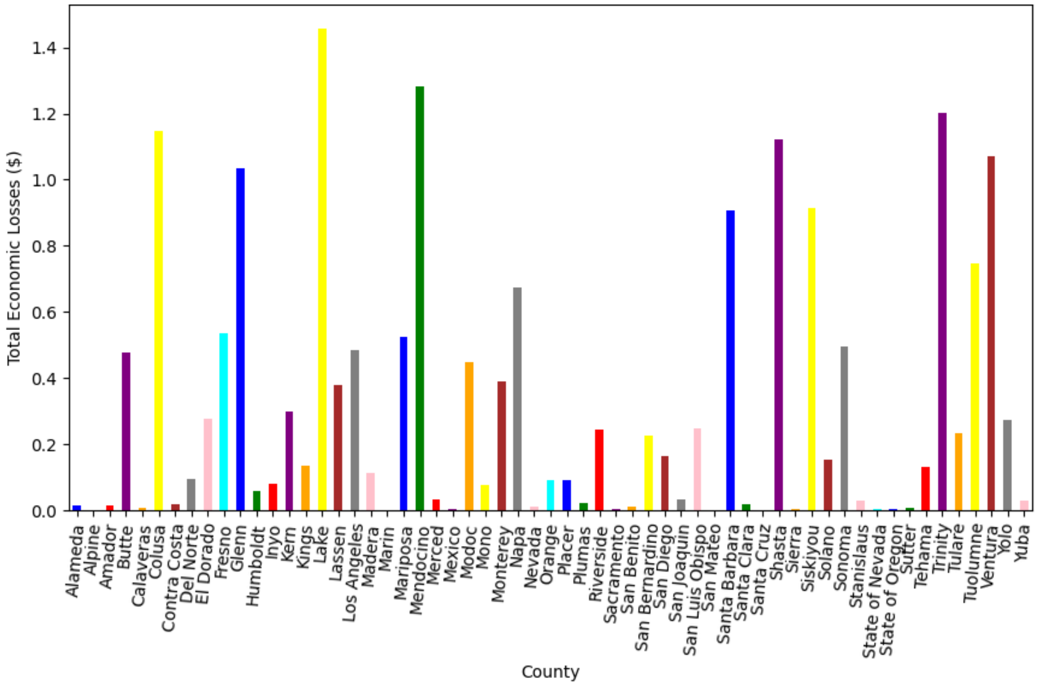
Concurrently, the azure trajectory displayed in the subordinate chart, pertaining to injuries and fatalities, reaches an apogee at 140, while the crimson curve ascends to a zenith of 100. This insinuates that particular year's witness California's wildfires inflicting a substantially heightened number of casualties and mortalities in comparison to dissimilar time periods. A plausible explanation for this phenomenon may be attributed to analogous factors that govern wildfire tendencies and the peril they pose to human communities.

The above-mentioned graphical representation can facilitate the identification of recurrent trends and patterns in the aftermath of wildfires, thereby aiding in making informed decisions pertaining to formulation of policies and strategies for managing wildfire hazards. For instance, it may become necessary to intensify resource allocation and mitigation measures in regions that are particularly susceptible to wildfire damage, or to prioritize emergency responses in areas that have witnessed high rates of fatalities and injuries in the past.

Through the analysis of trends and patterns present within the California wildfire dataset, we can develop sophisticated risk management strategies that not only account for the likelihood of wildfire outbreaks on people, infrastructure, and the environment but also consider their potential impact.

Economic Loss Occurred

Examining the aggregate economic impact of the wildfires across California by analyzing the total economic loss within the dataset can yield invaluable insights into the financial ramifications for the state as a result of these incidents. A comprehensive assessment of the total economic loss can provide significant implications for understanding the full scope and consequences of these devastating fires.

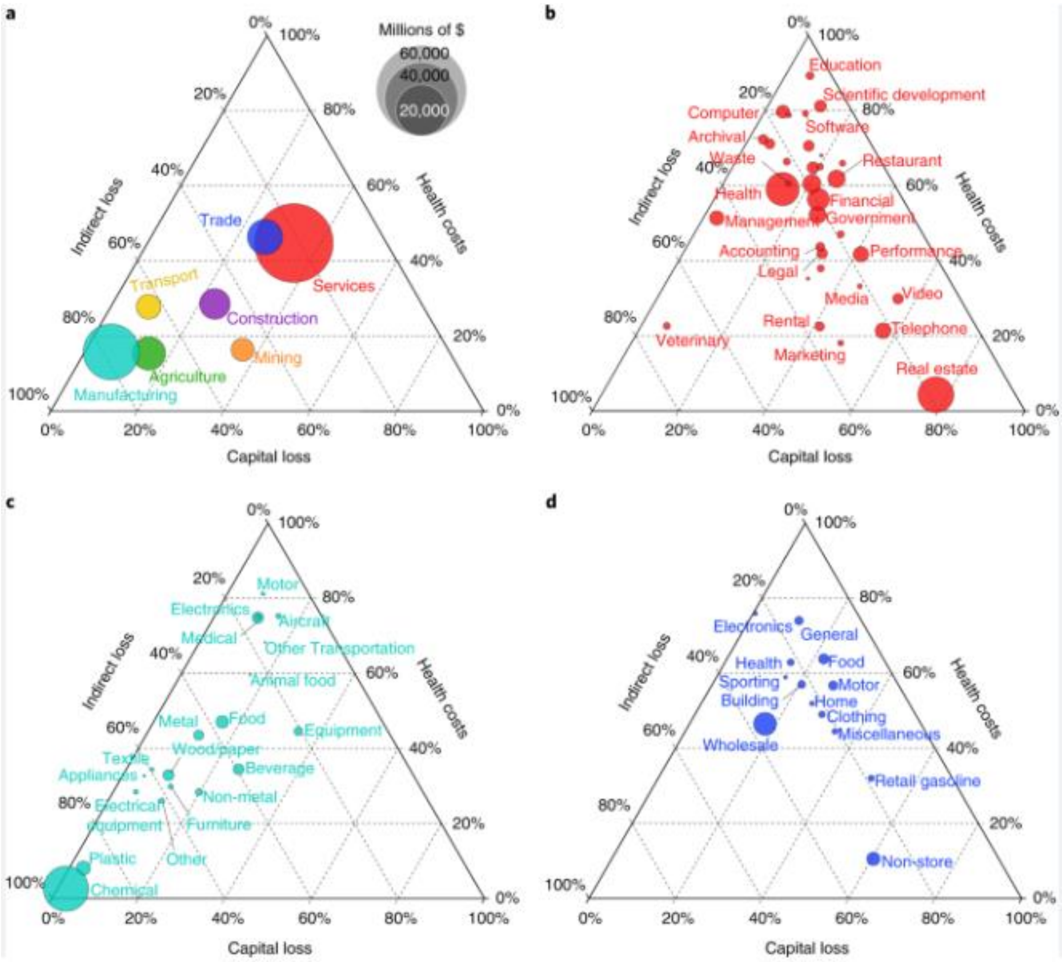


The extent of economic losses varies significantly across different counties. An analysis of the bar chart that illustrates total economic losses by county reveals that some counties have suffered considerably more severe economic losses than others, implying that these specific counties are more susceptible to the economic impact of wildfires and hence require more substantial resources for effective mitigation of these impacts.

Elevated financial detriments transpire in relation to more extensive conflagrations: A comprehensive examination of the dataset, accounting for each wildfire occurrence, facilitates the discernment of underlying correlations or progressions. It appears plausible that wildfires of increased magnitude engender a more substantial economic influence, due to the potential amplification of destruction inflicted upon properties and infrastructure, in addition to impeding economic operations within the impacted vicinity.

Property damage is not the only kind of economic losses: The dataset's overall economic losses can cover a wide range of expenses, including the price of fighting the fire, the loss of economic activity brought on by it, and the damage the fire does to the surrounding environment and ecosystems. Understanding the many costs that go into the total economic losses might help to comprehend the financial effects of wildfires more fully in California.

Economic footprint of California Wildfires in 2018:

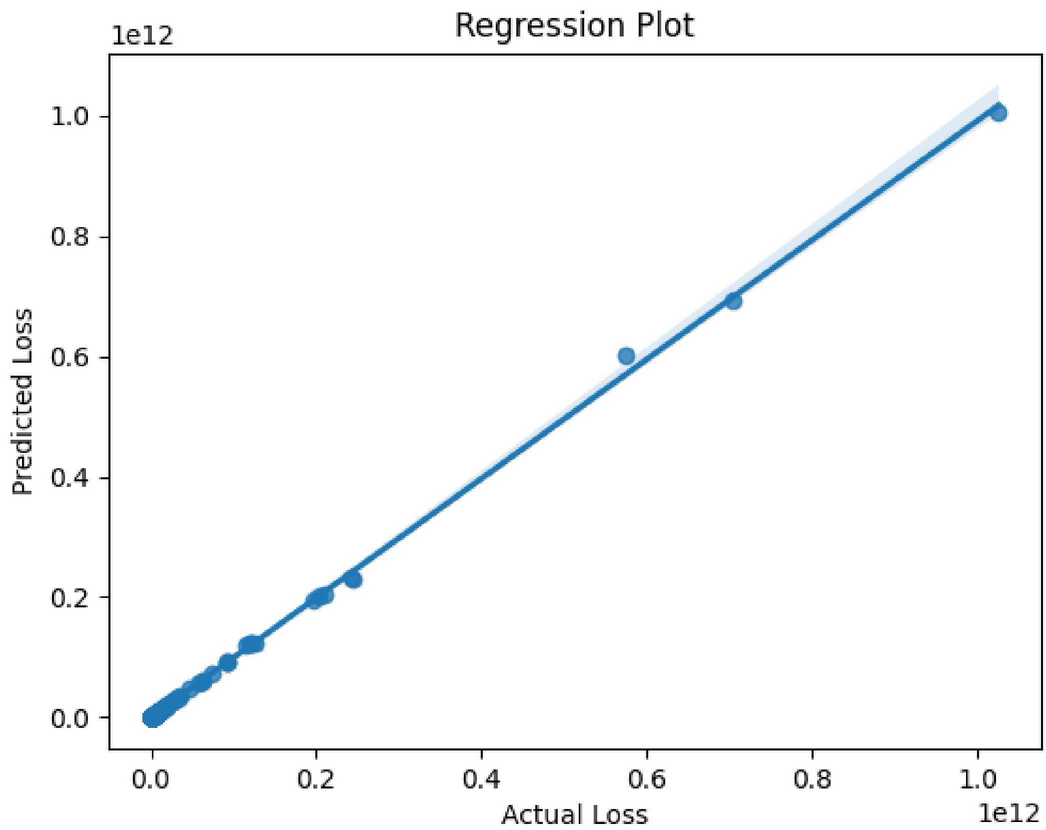


Using an amalgamation of physical, epidemiological, and economic models, we have endeavoured to estimate the extensive economic ramifications of California wildfires in 2018. These ramifications, including the devaluation of destroyed and damaged capital, the health tolls linked with air pollution exposure, and indirect losses due to a general economic disruption that propagates throughout regional as well as national supply chains, have resulted in staggering economic damages of $148.5 (126.1–192.9) billion. This outcome amounts to approximately 1.5% of California's gross domestic product, with $27.7 billion (19%) in capital losses, $32.2 billion (22%) in health costs, and $88.6 billion (59%) of indirect losses (all denominated in US$). The outcomes we offer also expose the fact that most economic impacts related to California wildfires tend to be indirect and commonly affect industry sectors as well as locations that are distant from the fires themselves. Incidentally, 52% of the indirect losses—31% of total losses—in 2018 were discovered to be outside of

California. Therefore, our new findings and exceptional methodology present a novel threshold for decision makers who are assigned with the task of safeguarding lives, ensuring that key production sectors are not decimated, and subsequently preventing future wildfires from ravaging California.

Analytics to Manage Risks

- Using Random Forest Regression for predictions:



Using a Random Forest regression model, we can predict the total loss from wildfires using the features of the Acres Burned, number of Injuries, the Structures Damaged and the Structures Destroyed.

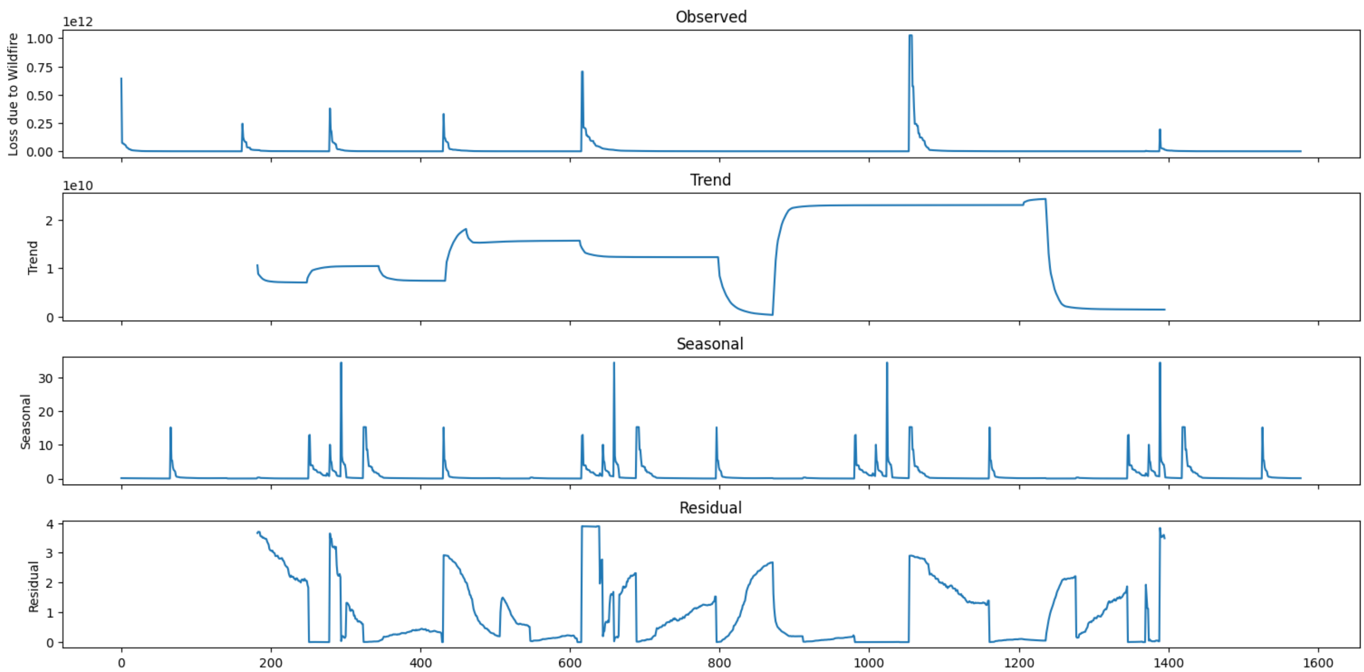
Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared are the metrics for assessing the model's performance. The Total Loss values' 99.9% variability is explained by the model, according to the R-squared value of 0.999. The Random Forest regression model has achieved a high R-squared value of 0.999, indicating that it is an excellent fit for the data and can explain almost all of the variability in the Total Loss values. However, the MSE and RMSE values are also quite high, with the MSE being in the order of 3.425x10^18 and the RMSE being around 1.85 billion. This indicates that there may be some outliers or extreme values in the dataset that are driving up the error metrics.

As can be seen from the regression figure, there is a linear relationship between the predicted and real Total Loss values, demonstrating that the model is successfully identifying the fundamental patterns in the data. The results show that the Random Forest model can accurately forecast the Total Loss values for wildfires and is a strong fit for this dataset, according to the results.

Risk assessors and emergency management teams in your organization can benefit from these analyses by having a better understanding of the trends and patterns of tornadoes as well as their possible effects on communities. They may make better use of this knowledge to create readiness and response strategies for disasters, including early warning systems, evacuation protocols, and emergency supplies. By being aware of the dangers posed by tornadoes, communities can take proactive measures to lessen their effects and lower the likelihood of fatalities and property destruction.

- Using time series forecasting to predict future values based on the data:

By using time series forecasting techniques, we could make predictions about the likelihood and severity of future wildfire incidents in California, based on historical data and any patterns or trends identified in the dataset.



Based on the California wildfire incident dataset on Kaggle, we can use time series analysis and forecasting techniques to gain insights and make predictions about future wildfire incidents in the state. Some potential insights and predictions that could be gained from this dataset include:

Trend: The dataset includes information on wildfire incidents from 2013 to 2020, which could reveal any long-term trends or patterns in wildfire occurrence and severity over this time period. If the trend is increasing over time, it indicates that losses due to wildfires are generally increasing over the years, while a decreasing trend indicates that losses are decreasing. Here, we cannot see any significant change in trend pattern.

Seasonality: The dataset includes information on the date and time of each wildfire incident, which could reveal any seasonal patterns in wildfire occurrence and severity, such as increased incidents during certain months or seasons. The seasonal component shows the repetitive pattern of the time series over a fixed period. In this case, a period of 365 days is used, which could reveal seasonal patterns in wildfire losses, such as increased losses during certain months of the year.

Residuals: The residual component displays the irrational variations in the time series which cannot be accounted for by the trend or seasonal components. The model might not be capturing all the patterns in the time series if the residuals are big.

Use case scenario

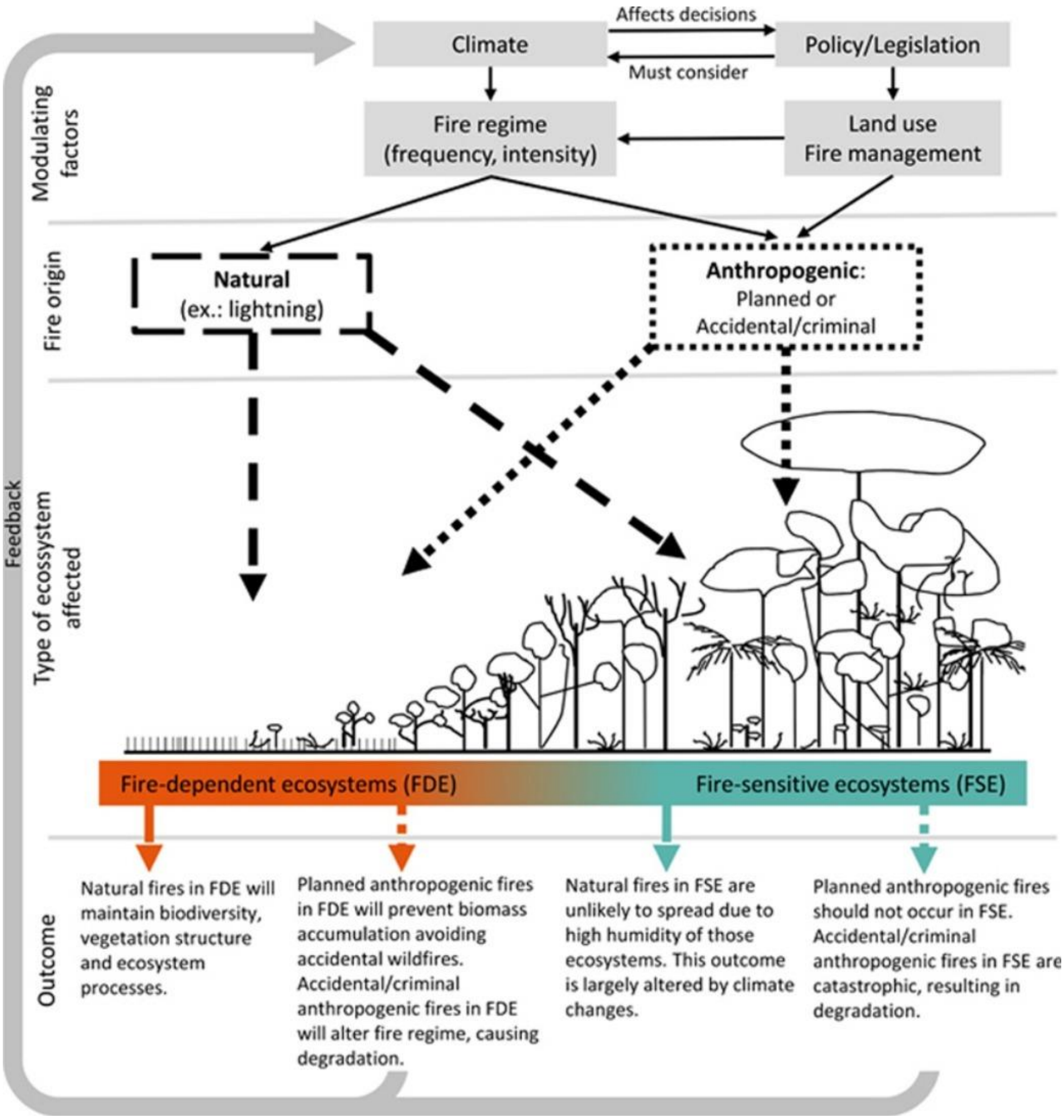
Let us consider a use case scenario that takes into account the analysis done so far and the recommendations for better risk management for California wildfires, where the state government of California decides to implement a new risk management strategy for wildfires based on the insights gained from the analysis of historical wildfire data. The strategy involves the following steps:

1. To evaluate the risk of wildfires in various parts of the state, the state government will create a framework that considers a variety of risk factors, including weather, vegetation, human activity, and infrastructure.

2. To detect the beginning of wildfires more effectively, the state government will incorporate cutting-edge technologies like satellite photography, artificial intelligence, and machine learning algorithms into its early warning systems. This will facilitate the rapid mobilisation of firefighting assets and the containment of wildfires.

3. The state government will increase its capacity for preparedness and response by making investments in cutting-edge firefighting gear, training programmes for firefighters, and a centralised command system for overseeing firefighting operations.

The state government can considerably lessen the effects of wildfires and safeguard the lives and property of its residents by putting this risk management approach into place.



Ethics

The use of data was properly managed throughout this investigation. The study did not collect or use any personal information. As part of data cleaning, Nan values were removed from the dataset. There are no specific individuals who were the focus of the study or the data, and no societal, political, or legal issues are involved.

The sources of information utilised in the paper are openly disclosed, as are any potential conflicts of interest that might have affected the research.

The social responsibility of providing the information has been taken into account in the report. This entails addressing how the flames have affected the ecosystem and the larger community, as well as highlighting methods to aid in relief efforts and stop wildfires in the future.

The report and analysis was written with consideration for the affected individuals. To do this, I have refrained from using words that could be considered insulting or inconsiderate towards people who have lost their homes or loved ones.

The recommendations that is being suggested is viewed as educational material and a notion that could be used in the future. The planned system's implementation and assessment are subject to public debate.

Recommendations

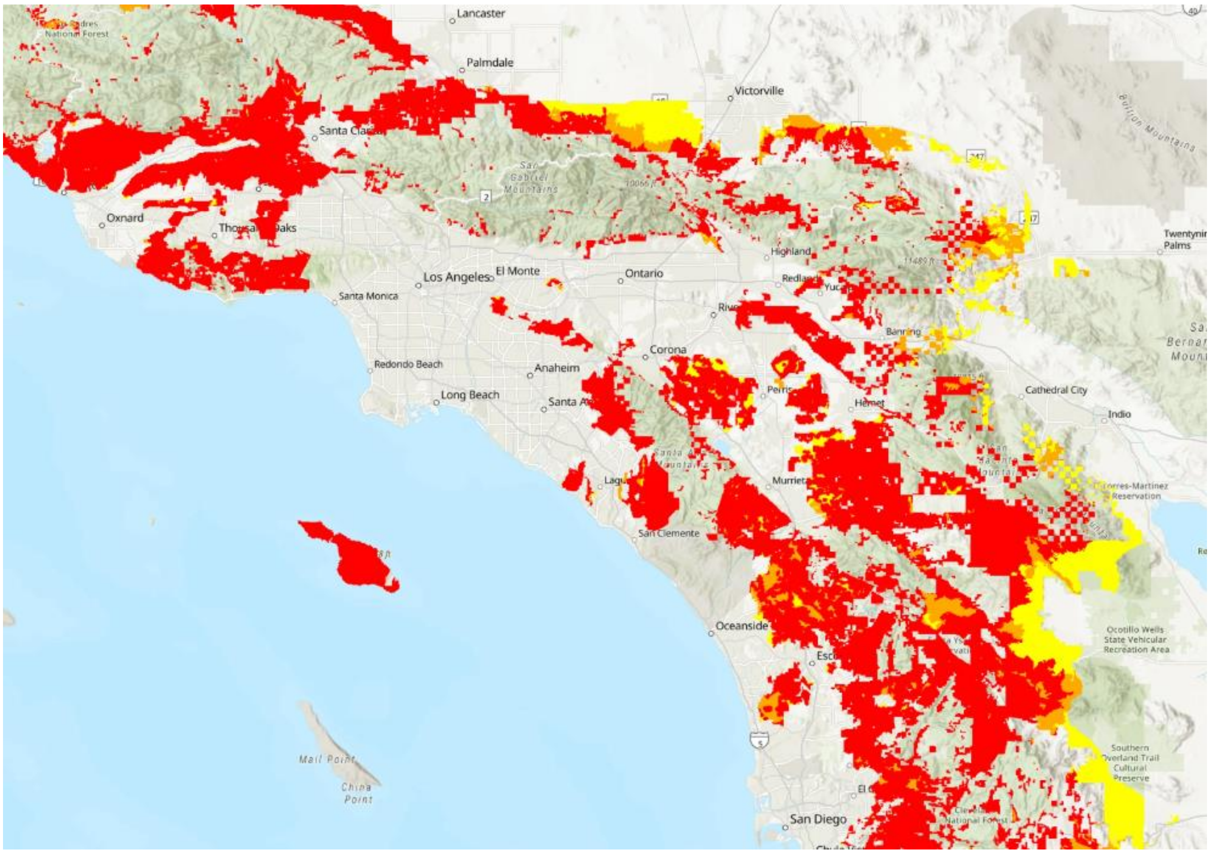
Based on the analysis done, we can make the following recommendations:

- To enhance fire prevention efforts, it's crucial to educate the public on how to prevent wildfires caused by human activities. This can be achieved by implementing campfire restrictions and increasing fines for violators. Additionally, it can be beneficial to promote the use of equipment that reduces the risk of wildfires such as flame-retardant clothing and spark arresters. Providing educational materials and outreach programs through different media channels can also be effective in increasing awareness and preventing wildfires.
- Elevate the allocation of resources towards combating wildfires: The alarming escalation in both intensity and occurrences of wildfires in California necessitates the urgent augmentation of funds designated for firefighting endeavours. This encompasses the acquisition of additional personnel and the procurement of advanced firefighting apparatus.
- A substantial proportion of the wildfires were caused by natural phenomena, such as lightning strikes. To mitigate the impact of such disasters, it is imperative to formulate and execute superior land management policies, including controlled burns, aimed at reducing combustible materials.
- The analysis has demonstrated a continued escalation in the number of individuals impacted by wildfires throughout successive years, thus highlighting the imperative nature of enhancing evacuation procedures to ensure utmost safety for the affected populace. Such a feat can be achieved through the establishment of improved emergency protocols, enhanced communication channels, and the delivery of timely information to the public.
- Changes in temperature, humidity, and other factors that can be signs that a wildfire is brewing can be found using these systems' sensors, cameras, and other gadgets. The Alert Wildfire Network is one instance of an early warning system. For the purpose of spotting wildfires, this network is made up of cameras that are carefully positioned in risky locations. The cameras have software that can instantly warn firefighters and other emergency personnel when a fire starts and can detect the beginning of a fire. Firefighters can make better decisions on how to combat the blaze with the help of the system's live feeds of the fire.
- To pinpoint regions that are at a high danger of wildfires, this system can gather and analyse data on vegetation, terrain, weather, and other aspects. Using sensors on satellites, aircraft, drones, or other objects, remote sensing technology collects data

from a distance. GIS, on the other hand, describes the procedure of examining and interpreting such data in order to produce maps and other visualisations that aid in the discovery of patterns and trends. Maps that depict areas at high risk of wildfires can be made using the data gathered from remote sensing technology and GIS.

It is feasible to produce risk maps that display the regions most likely to be harmed by wildfires using remote sensing technology and GIS. Land managers and emergency responders can use these maps to plan for prevention and mitigation measures. Land managers, for instance, can use the data to prioritise vegetation management initiatives or perform prescribed burning to lower the risk of wildfire initiation.

Cal Fire’s Fire Hazard Severity Zone map shows that most of Southern California is classified as a very high fire hazard. The red colour shows the areas that the state is responsible for:



Data on vegetation cover, slope, and other topographic characteristics, for instance, can be used to pinpoint regions that are vulnerable to the start or spread of wildfires. Furthermore, information on weather parameters like wind, humidity, and temperature can be used to determine when conditions are favourable for wildfires.

Additionally, real-time monitoring of the spread of wildfires can be done using GIS and remote sensing. This can be done by tracking the development of a wildfire using satellite data or aerial imagery. By using this information, evacuation orders can be made more accurate, and firefighters' efforts can be more efficiently directed.

Limitations

While there are several options for managing the risk of wildfire incidents in California, there are also some restrictions that need to be taken into account. The following restrictions should be remembered:

- GIS and remote sensing technologies can only map areas that are visible to the sensors or satellites. Like early warning systems, they only work in the areas where they are deployed. This implies that some places that are vulnerable to wildfires may not be covered by this technology.
- Upfront expenditures for hardware and software for remote sensing and GIS systems installation might be high. In a similar vein, constant maintenance and monitoring are necessary to maintain the effectiveness of early warning systems. For certain communities or organisations, the price of these systems may be unaffordable.
- The accuracy of remote sensing and GIS systems can be impacted by weather, technological issues, and other reasons. Similarly, communication or power disruptions may have an impact on early warning systems. This implies that these systems could not be always trustworthy.

  Overall, while technical risk management strategies for California wildfire incidents can be successful, it is important to consider their limitations. It is feasible to create more thorough and efficient risk management techniques by taking into consideration these constraints.

Conclusion

The wildfire problem in California is a multifaceted issue that necessitates a comprehensive risk management approach. Employing technical solutions like geographic information system (GIS) and remote sensing as well as early warning systems prove effective in pinpointing high-risk areas, monitoring the progression of wildfires, and alerting emergency responders promptly. However, these technical solutions come with limitations such as limited coverage, high costs, dependence on technology, and the propensity for human error.

In order to adequately address the risk posed by wildfires in California, a multifaceted strategy that encompasses both technical solutions and community education and preparedness, as well as vegetation management and firefighting efforts, is imperative. By adopting such an approach, the detrimental consequences of wildfires to both the environment and surrounding communities can be substantially mitigated. Consequently, coordinated efforts between government agencies, community organizations, and individuals are necessary for effective implementation of these strategies to safeguard California against the catastrophic effects of wildfires.

References

- Ananthu, R. (2021). California Wildfire Incidents (2013-2020) [Dataset]. Kaggle. https://www.kaggle.com/datasets/ananthu017/california-wildfire-incidents-20132020
- Datacamp.(2023). Random Forest Classification with Scikit-Learn. https://www.datacamp.com/tutorial/random-forests-classifier-python
- Machinelearningmastery.(2020). How to Decompose Time Series Data into Trend and Seasonality. https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/
- Brinklow, A. (2021, August 3). California's New Fire Hazard Map Is Out. LAist. https://laist.com/news/climate-environment/californias-new-fire-hazard-map-is-out
- Fulton, E. A., Boschetti, F., Lenton, A., Little, L. R., Perry, R. I., & Dambacher, J. M. (2021). Five priorities for using ecosystem services effectively in policy and decision-making. Nature Sustainability, 4(1), 53-62. https://www.nature.com/articles/s41893-020-00646-7