EDA PROJECT:

How Growing Population Fuels California Wildfire

Andy Sun

MSDS 430: Python for Data Science

Table of Contents

Introduction

Wildfires have become a growing threat to the environment that we live in, and we had one of the worst fire seasons in recent years. One of the biggest wildfires in the past years, the 2019-2020 Australian Bushfires, burned more than 46 million acres (Center for Disaster Philanthropy, 2020). Similarly in the U.S, 253,321 acres were burned in the Californian landscapes in 2019 (Center for Disaster Philanthropy, 2020). Although wildfires can be very catastrophic, it is important to understand that wildfire is a natural process. Wildfires burn down the underbrush and understory of the forest and allows new plants to flourish. As a result, more biodiversity is created in the environment, and it evens out the playing field of the ecosystem. Additionally, many plants have evolved to adapt to wildfires. The Sequoia trees and its reproductive system depend on the heat of the wildfires; the heat of the fire can activate the seed cone to release Sequoia's seed for its reproduction (Save the Redwoods League, 2020). Hence, wildfires might be devastating for people, but it's not necessarily a disaster for forest ecosystems.

It is natural and beneficial to have natural wildfires, however, wildfires caused by humans are more dangerous since it can create an unbalanced ecosystem and essentially slow the recovery of the forest. To protect nature and the people, I believe we should explore other factors that might worsen the wildfires. I hypothesize that there is a positive correlation between the growing population and the increasing wildfires.

Finding reliable data for my analysis is paramount. The wildfire dataset I used for this analysis is from the California Department of Forestry and Fire Protection. This dataset is compiled annually for the public use by The Fire and Resource Assessment Program (FRAP), CAL FIRE, the United States Forest Service Region 5, the Bureau of Land Management, and the National Park Service (CAL FIRE, 2021). In this analysis, I wanted to focus on the wildfires in the western United States, specifically the State of California. To support my hypothesis, the other dataset that I am using is the Population data from the U.S Census Bureau. This dataset collects information from respondents directly by surveys or from other sources such as the local state and government (US Census Bureau, 2019).

Data Preparation

The wildfire dataset includes fires that occurred in California between the years 1950 and 2017. "The fire perimeter database represents the most complete digital record of fire perimeters in California." according to the Cal fire website and it consists of 14,847 rows. To prepare the data for this study, I removed the irrelevant fields and analyzed the key attributes (Table 1). Additionally, I incorporated descriptions of the cause of the fire for better interpretation of the code (Table 2). Lastly, I removed all the records that had "null" listed within the cause. Eventually, only 14,805 records were left in my dataset and outliers were also kept in the dataset since all wildfires are relevant in my analysis.

On the other hand, the population dataset consists of 121 rows, and it includes the population of California from 1900 to 2020. No changes were required for this dataset since the data is clean and intuitive.

Analysis

I started the analysis by comparing the growth of the California population and the number of wildfires in California over time. The data confirms that both population and wildfire in California have grown steadily over the years (See Figure 1 and 2). This indicates a positive correlation between the number of wildfires and the population of California. Furthermore, I categorized the cause into "cause type" to distinguish fire caused by human, natural fire, or unknown. I discovered that 62.5% of the wildfires are caused by human, 26.1% are the unknown cause, and 11.4% are from natural fires (See Figure 3). This also tells us that people have an immense impact on the number of wildfires.

I also created heatmaps to show how wildfires are distributed each year. Based on the total wildfire heatmaps (Figure 4), the annual wildfire season is between May and October. Most natural fires occur consistently within the fire season with some outliners (Figure 5). However, the fire caused by humans appear to happen more often and expand beyond the fire season each year (Figure 6). Moreover, I analyzed the top 100 fires from the dataset and observed that these large fires are becoming more common in recent years (Figure 7).

Conclusion

While the population in California has grown significantly in the past years, we also see the number of wildfires rising each year. Although natural fires are becoming more frequent, there is an exponential increase in the number of wildfires caused by human that were reported. The data demonstrates a high correlation between the population and the increasing wildfires reported in California over the years.

Although this analysis supports the hypothesis, I believe it would be interesting to investigate more on population in California. For example, the house construction data will show if fireproofing a house would make high risk locations less vulnerable. The income level of the population would also help me to understand if there is a correlation between resource distribution and these high-risk areas. With the additional data, my analysis will have a holistic view of the wildfire problem.

Appendix

Table 1: Wildfire Attribute Definitions

Column Name	Data Type	Description
YEAR_	Integer	Fire Year
FIRE_NAME	Object	Name of the fire
ALARM_DATE	Date	Alarm date for fire
CAUSE	Integer	Reason fire ignited
GIS_ACRES	Float	GIS calculated area in acres
SHAPE_AREA	Float	Area in square meters

Table 2: Cause Definition

Cause Code	Description
1	Lightning
2	Equipment Use
3	Smoking
4	Campfire
5	Debris
6	Railroad
7	Arson
8	Playing with Fire
9	Miscellaneous
10	Vehicle
11	Power Line
12	Firefighter Training
13	Non-Firefighter Training
14	Unknown/Unidentified
15	Structure
16	Aircraft
17	Volcanic
18	Escaped Prescribed Burn
19	Illegal Alien Campfire

Figure 1

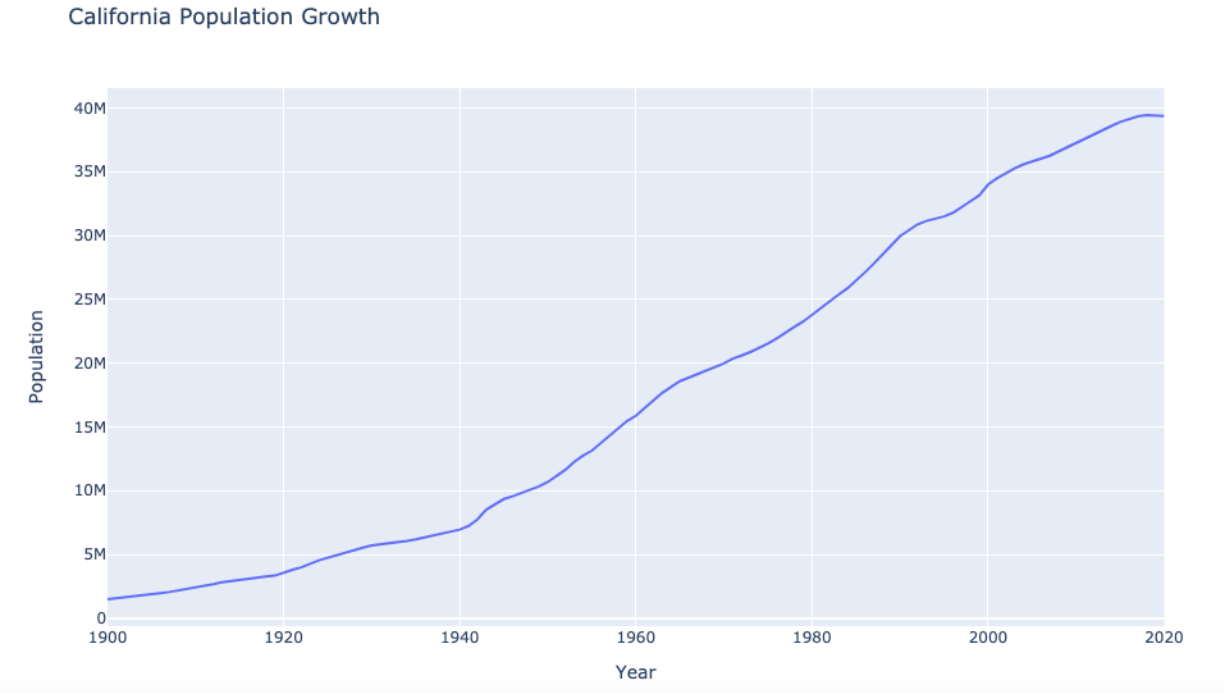


Figure 2

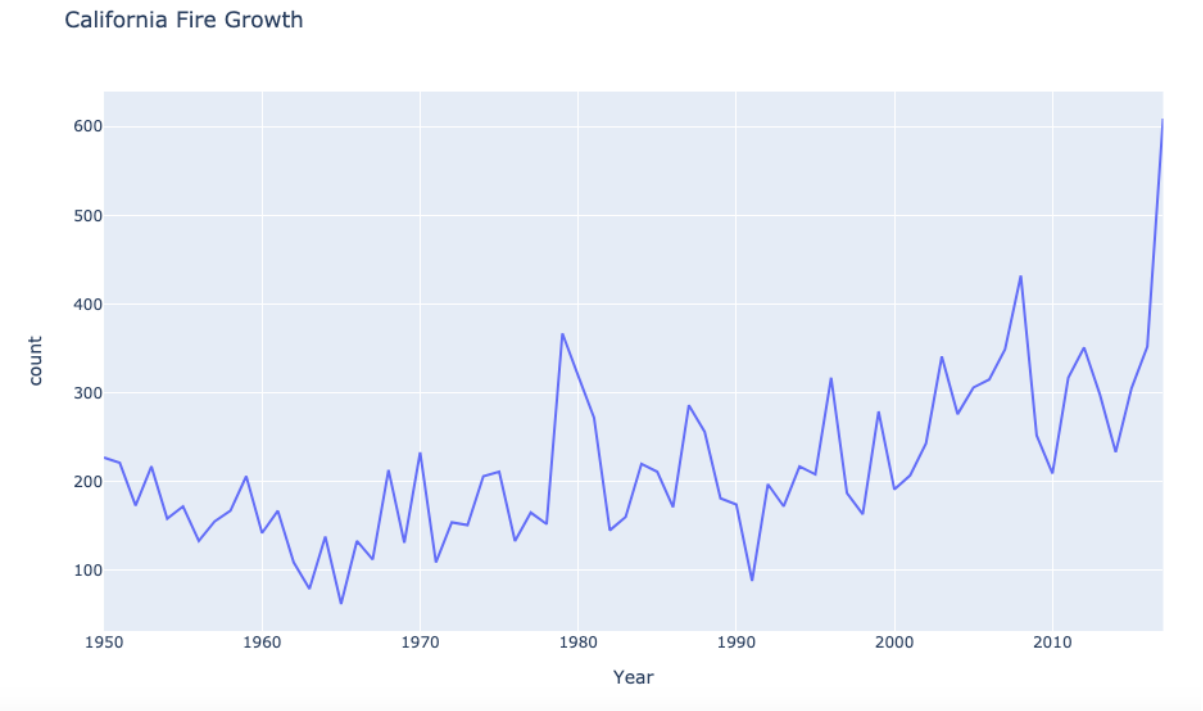


Figure 3

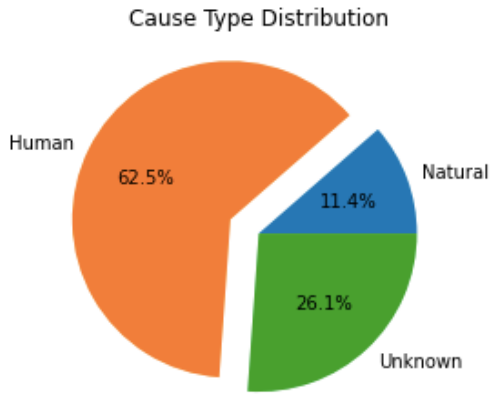


Figure 4

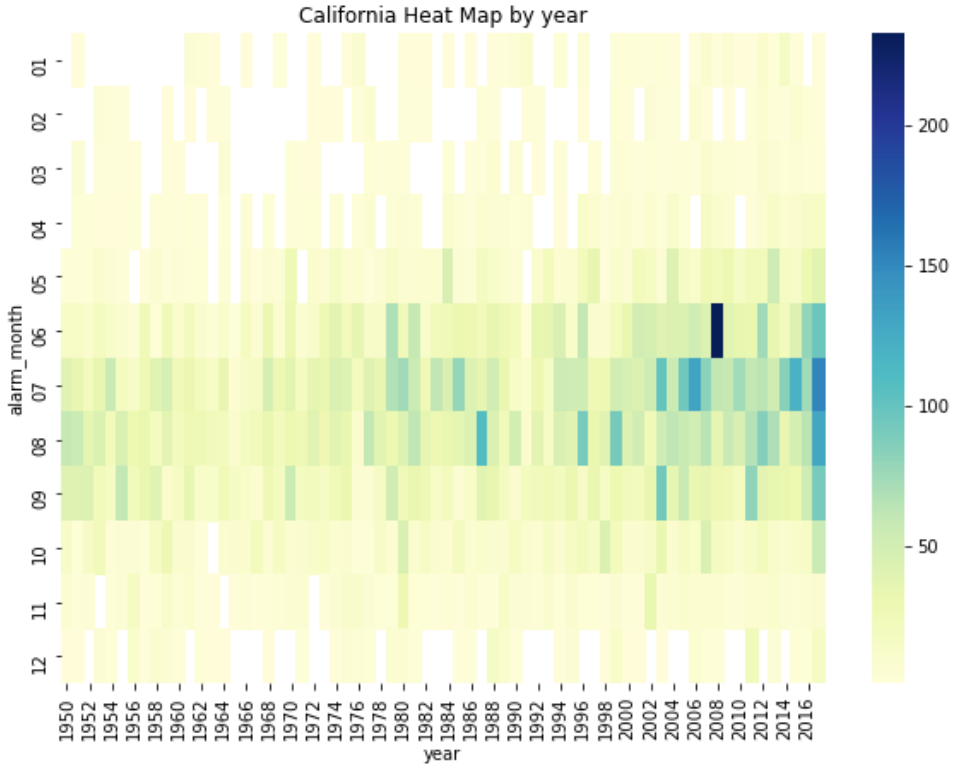


Figure 5

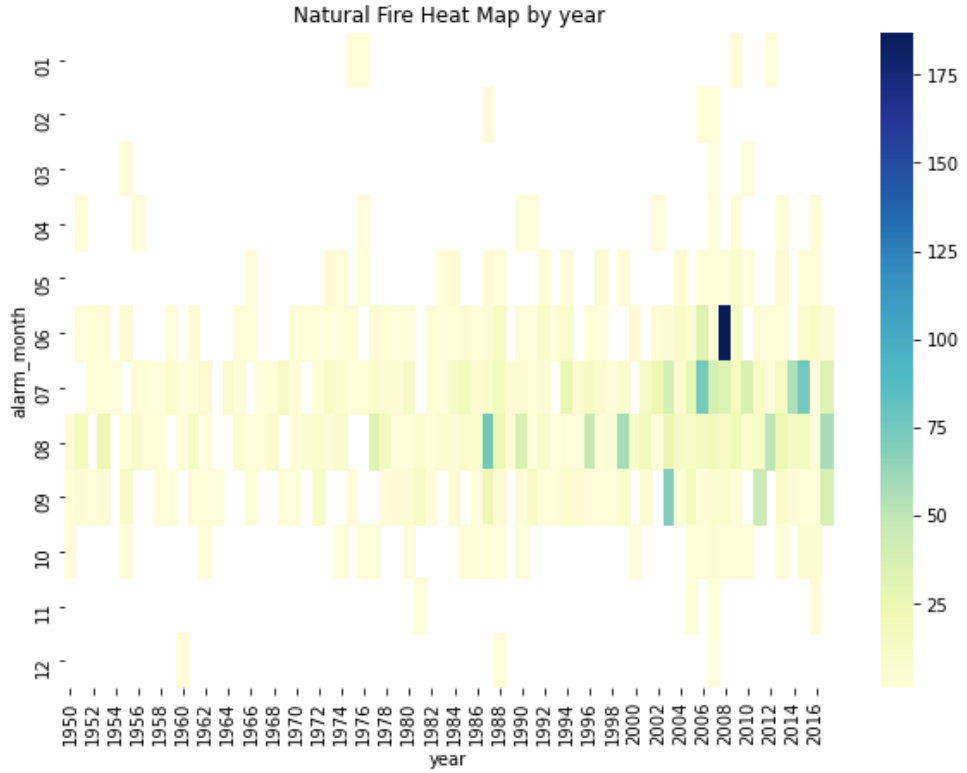


Figure 6

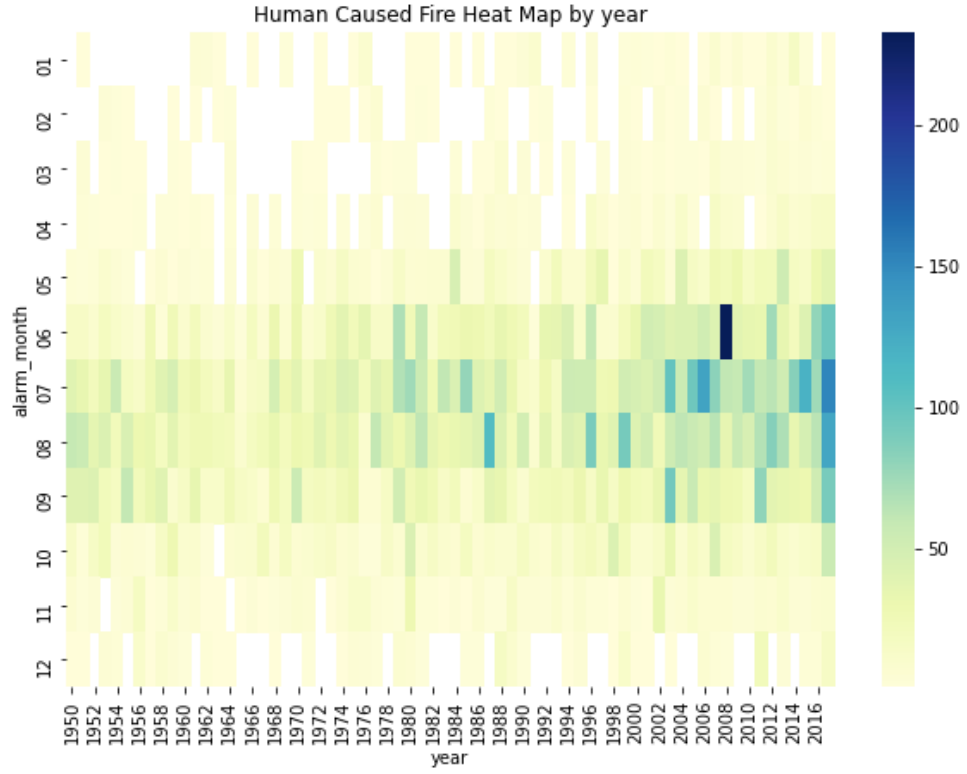
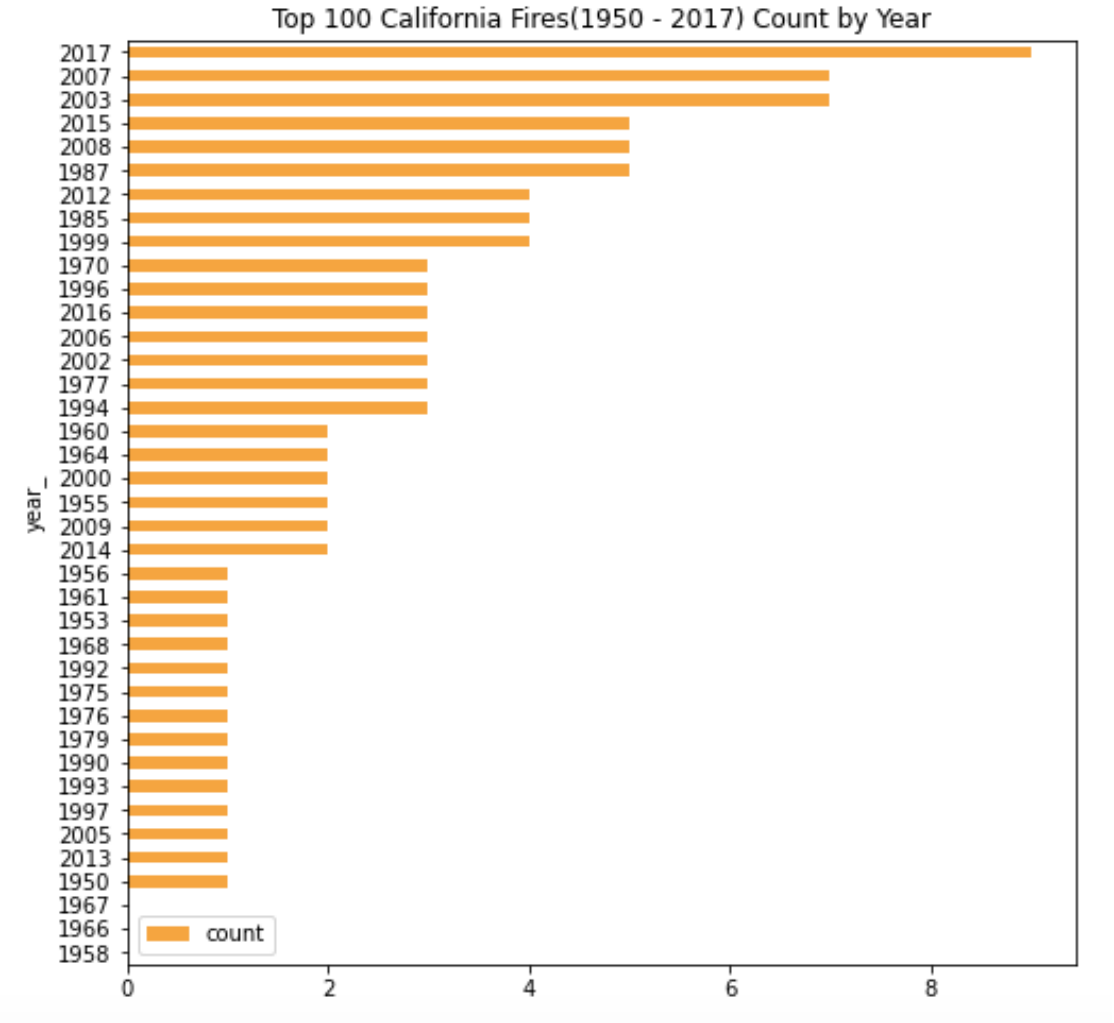


Figure 7


Top 100 California Fires(1950 - 2017) Count by Year
year_
2017
2007
2003
2015
2008
1987
2012
1985
1999
1970
1996
2016
2006
2002
1977
1994
1960
1964
2000
1955
2009
2014
1956
1961
1953
1968
1992
1975
1976
1979
1990
1993
1997
2005
2013
1950
1967
1966
1958
count
0
2
4
6
8

Reference

US Census Bureau. (2019, May 02). Combining Data – A General Overview. Retrieved June 05, 2021, from https://www.census.gov/about/what/admin-data.html#:~:text=The%20Census%20Bureau%20uses%20data%20from%20a%20variety%20of%20sources.&text=Some%20data%20are%20collected%20from,censuses%20and%20surveys%20we%20conduct.&text=Primary%20sources%20for%20additional%20data,well%20as%20some%20commercial%20entities.

Center for Disaster Philanthropy. (2020, Oct). 2019-2020 Australian Bushfires. (n.d.). Retrieved June 05, 2021, from https://disasterphilanthropy.org/disaster/2019-australian-wildfires/

Center for Disaster Philanthropy. (2020, July). 2019 California Wildfires. Retrieved June 05, 2021, from https://disasterphilanthropy.org/disaster/2019-california-wildfires/

California Department of Forestry and Fire Protection (CAL FIRE). (n.d.). Fire Perimeters. Retrieved June 05, 2021, from https://frap.fire.ca.gov/frap-projects/fire-perimeters/

Save the Redwoods League. (2020, December 3). *Giant Sequoia*. Save the Redwoods League. https://www.savetheredwoods.org/redwoods/giant-sequoias/?gclid=Cj0KCQjw5PGFBhC2ARIsAIFIMNdqjm9ymN8pkX9F8-tDXXhTsiBHOShBAxdtp1dlPfBiWaG2c-fOiPgaAjf8EALw_wcB.

Dataset:
https://github.com/BuzzFeedNews/2018-07-wildfire-trends/tree/master/data Californina fire reports 1950 - 2017
https://www.macrotrends.net/states/california/population California Historical Population 1950 - 2020