Hochschule
für nachhaltige Entwicklung
Eberswalde

Final Project Report of Applied Programming in Forestry (WiSe 2024/25)

Impact of Temperature on Wildfire Spread: A Case Study of California (2016–2024)

Submitted by
AHMED MANZIM RIDWAN
Matrikelnummer 23215259

Date: 21-02-2025

Table of Contents

1. Introduction

1.1 Background of the Problem

Wildfires are one of the burning problems of the world right now. Every year wildfires destroy a lot of forest areas. Wildfires are unplanned fires caused by either manmade or natural reasons in wild areas like forests, rangeland, or grasslands. These extreme events are prevalent in the western part of the U.S.A. Although wildfire has no time of its preferred, over the years, previous histories suggest it occurs in summer and fall more ((U.S. Department of Agriculture, n.d.). However, very recently, in January 2025, we saw a massive wildfire in California that couldn't be controlled for several days amidst a massive joint operation of firefighters.

This project report works on two different datasets of California fire events (not only related to wildfires but also includes other types of fire events).

2. Dataset Description

2.1 Overview of Datasets Used

This project utilizes two datasets:

- **CAL FIRE Damage Inspection (DINS) Data:** A dataset containing 130,723 rows and 46 columns detailing fire incidents, damage levels, and structure characteristics.
- **NOAA Weather Data:** A dataset with temperature records across multiple years, including average monthly temperatures for different counties.

2.2 Download Dataset

These are my data source downloading links. However, I've provided the dataset with the codes in the dataset folder.

1. The California fire dataset is downloaded from https://data.ca.gov/dataset/cal-fire-damage-inspection-dins-data.
2. The NOAA dataset is downloaded from https://www.ncei.noaa.gov/cdo-web/datasets/GHCND/locations/FIPS:06/detail. This dataset was obtained with some specific date ranges, and California was set as the Area of Interest location. It took a while to download, but it didn't take too long.

Both of these two are well-known sources.

2.2 Key Features and Preprocessing

- **California Fire dataset**: I filtered to include essential attributes such as fire incident date, location (county, city, coordinates), and fire damage levels.
- **California Weather Summary dataset**: Processed to obtain monthly average temperatures per county, ensuring alignment with fire incident data for correlation analysis. Like the Fire dataset, the dates and missing values have also been handled here.

There were many columns in both datasets, and I've worked only with a few columns that I felt were relevant to my analysis.

3. Methodology

3.1 Data Cleaning and Preprocessing

Here is the summary of the dataset cleaning and preprocessing.

- Converted date fields to datetime format to maintain consistency throughout the analysis and ensure all date formats are recognized. For example, in the American system, often the month comes first, then the date, like 09/25, refers to 25th September, while another zonal tradition might be reading the date first, then the month.
- Aggregated fire incident counts per month, and the county was also done because of checking the monthly counts across various counties.
- Merged datasets on county, year, and month for comparative analysis.
- The dataset contains NaN or blank in several fields. To make it easier to calculate, I Removed inconsistencies such as missing or erroneous values, which were handled in several fields.
- Finally, once the initial cleaning had been done, it was saved as 'dataset_cleaned.csv' for future usage. .to_csv command was used in this case.

3.2 Analytical Approach and Visualizations

- **Fire Trend Analysis:** Fire count trends across months and years. This gave an overview of the trends. To find the pattern of the high volume, which type of property is damaged most, and which state/cities or counties are affected the most or the least. This type of trend analysis was my goal, and to do so, I sometimes calculated based on two or more parameters.
- **Temperature Correlation:** Comparison of fire occurrences with average temperatures. While doing the fire trend Analysis, I observed a pattern that in the first 6

months of the year, there is merely any fire. The first thing that popped into my mind was, "Definitely, there is a summer-winter" impact here, indicating the temperature.

Hence, I got a temperature dataset. It was a huge dataset, starting with data from the 19th century. However, my California fire data was from the 20th century. Hence, I also ensured that the final dataset I'm working on also depicts the effect of temperature.

This NOAA dataset had no GPS cell. Hence, I had no option to compare the two datasets. Then, after further digging, I found that through inner join, something can be obtained from there. Both datasets had county, year, and month columns. Hence, I joined them (inner join) and did my further analysis.

```python
# fix date format type
df_selected['incident_start_date'] = pd.to_datetime(df_selected['incident_start_date'], errors='coerce')

# extract year
df_selected['year'] = df_selected['incident_start_date'].dt.year


# merge fire data with environmental data on 'year'
fire_weather = df_selected.groupby('year')['incident_number'].count().reset_index()
fire_weather.rename(columns={'incident_number': 'Fire Count'}, inplace=True)
fire_weather = fire_weather.merge(df_selected, on='year', how='left')

print(fire_weather.head(2))
```

```
[14]  ✓ 0.1s

   year  Fire Count incident_number incident_name fire_start_location  \
0  2013         280     CASHU 008265        Clover                 NaN
1  2013         280     CARRU 079781        Silver                 NaN
```

Figure 1: Merging two datasets based on yearly column (Author's creation, 2025)

- **County-Level Insights:** For identifying high-risk areas based on fire frequency and temperature variations, a county-based insight was also analyzed based on the county-relevant columns.
- **Visualization Tools:** Seaborn, Matplotlib (static plots), and Plotly, Folium (interactive maps).

4. Breakdown of Coding & Results of Different Files

4.1 dependencies.txt

In case the system doesn't have the following libraries installed initially, it's compiled all together inside this text file.

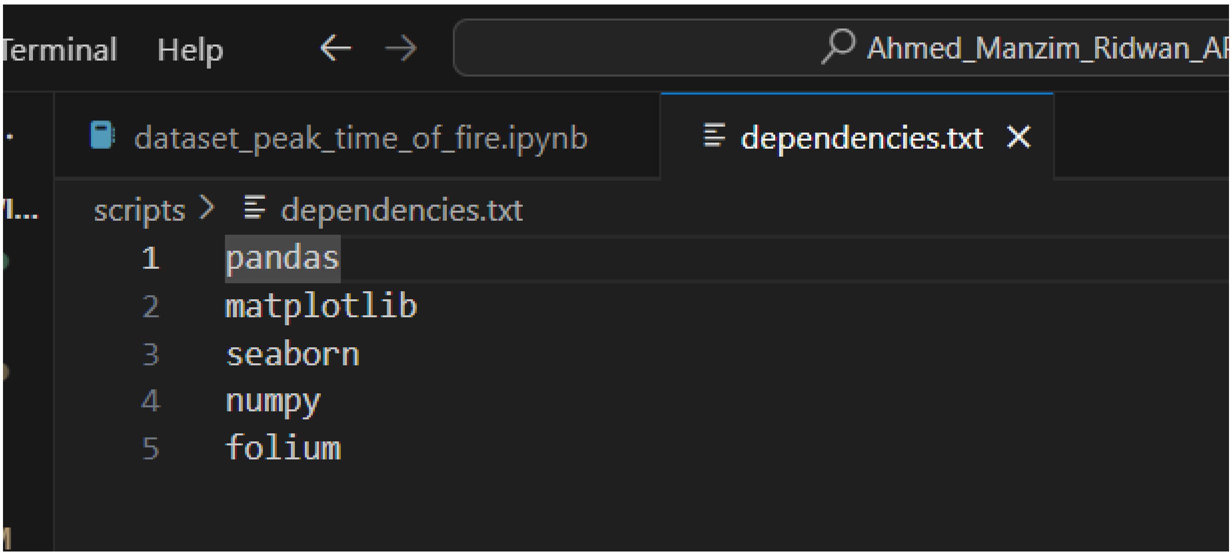


Fig 2: libraries used for this project (Author's creation, 2025)

4.2 Install Requirements

The command !pip install -r dependencies.txt will ensure all the requirements are installed on the device.

4.3 initialCommit.ipynb

Since I'm submitting this project on GitHub, I wanted to test if everything is okay before I start doing the coding. I printed on "*hello world test*" comment and found everything was fine. Finally, I prepared my GitHub setup from bash (terminal). After GitHub was set, I checked my files (git status), and then added my files with git add . followed by adding one commit message and finally pushing this file to GitHub. From next onwards, I put my codes to GitHub every time I think I'm done.

4.5 dataset_cleaning.ipynb

Since the dataset I chose had a lot of cells with missing values, inconsistent date type format, and a few columns had longer names (that could've been difficult to use while coding). Hence, I felt the necessity to cleanse the dataframe. In this file, I did my necessary cleansing.
After the cleaning, I saved this csv file for future analysis.

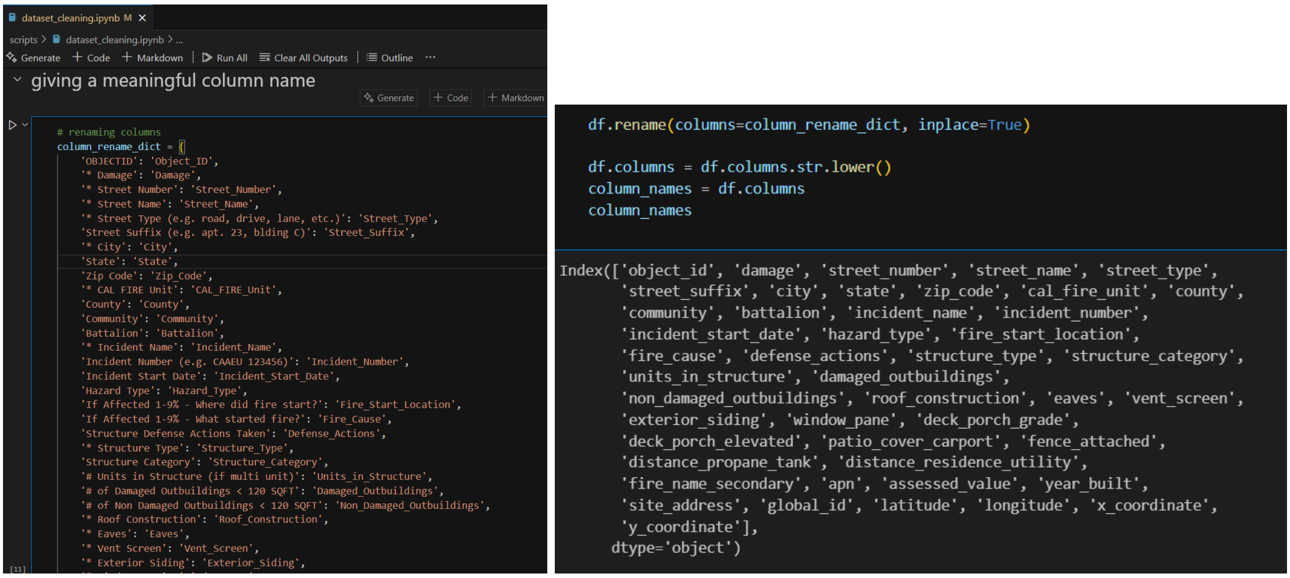


Fig 3: Dataset cleaning part - renaming the column for better readability (Author's creation, 2025)

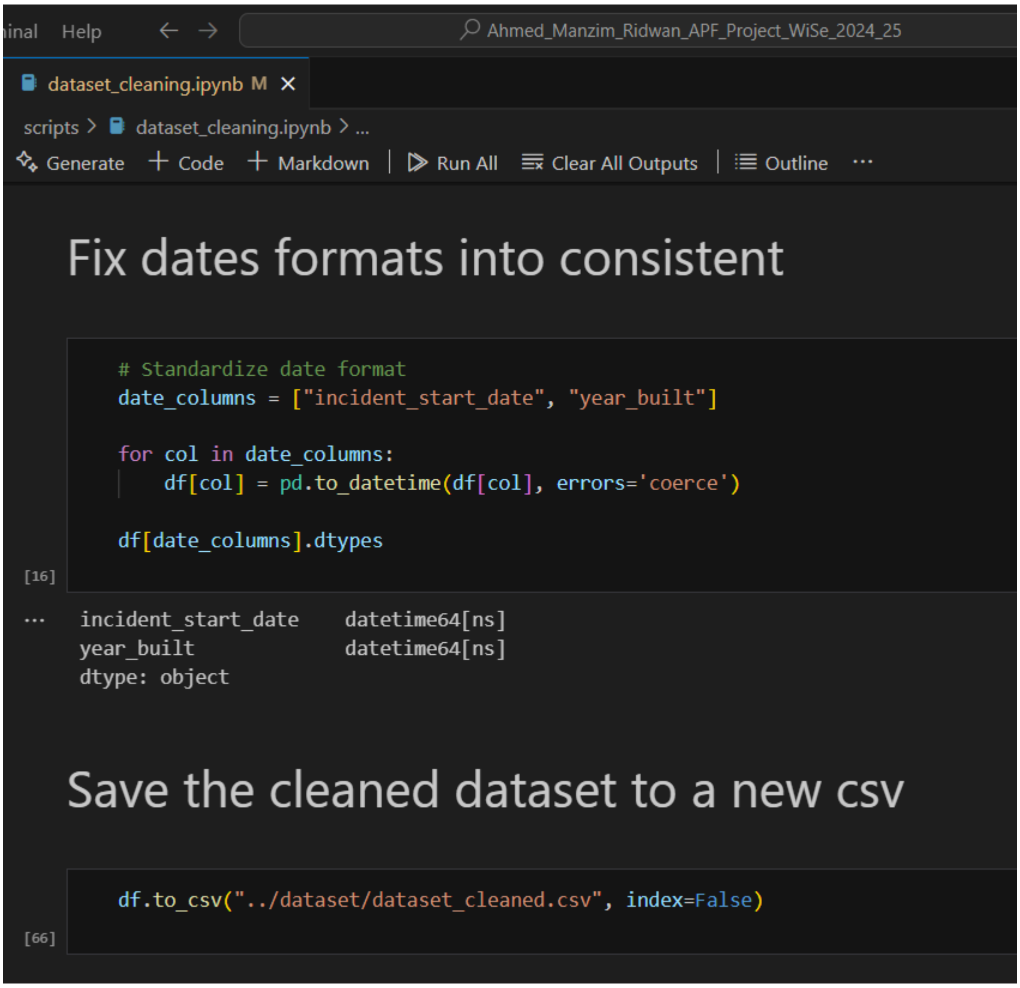


Fig 4: Fixing date type and saving the cleansed dataframe (Author's creation, 2025)

4.6 dataset_peak_time_of_fire.ipynb

Since 2013, I wanted to check the fire occurrence for each month of every year. In a single column and 12 rows, all the incident records have been placed. On the x-axis, there are months, and on the y-axis, there's the fire incident count.
The output provided interesting findings: in the 2nd half of the year, there's more fire occurrence!

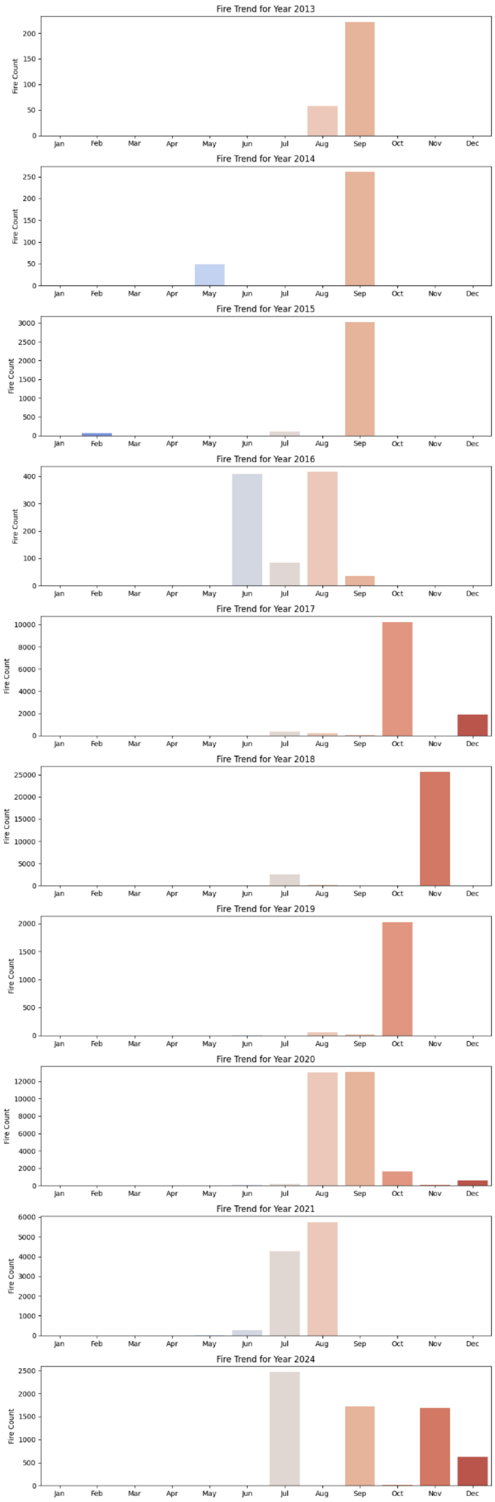


Fig 5: Monthly fire occurrence over the years (from 2013 to 2024) (Author's creation, 2025)

4.7 dataset_noaa_weather.ipynb

To check if there's any relation between the temperature and the fire incidents, I downloaded another dataset from NOAA. Like the preliminary dataset, this dataframe was also cleansed. Later, months and years were extracted since they were required. Then, the temperature trend was observed from 2016.

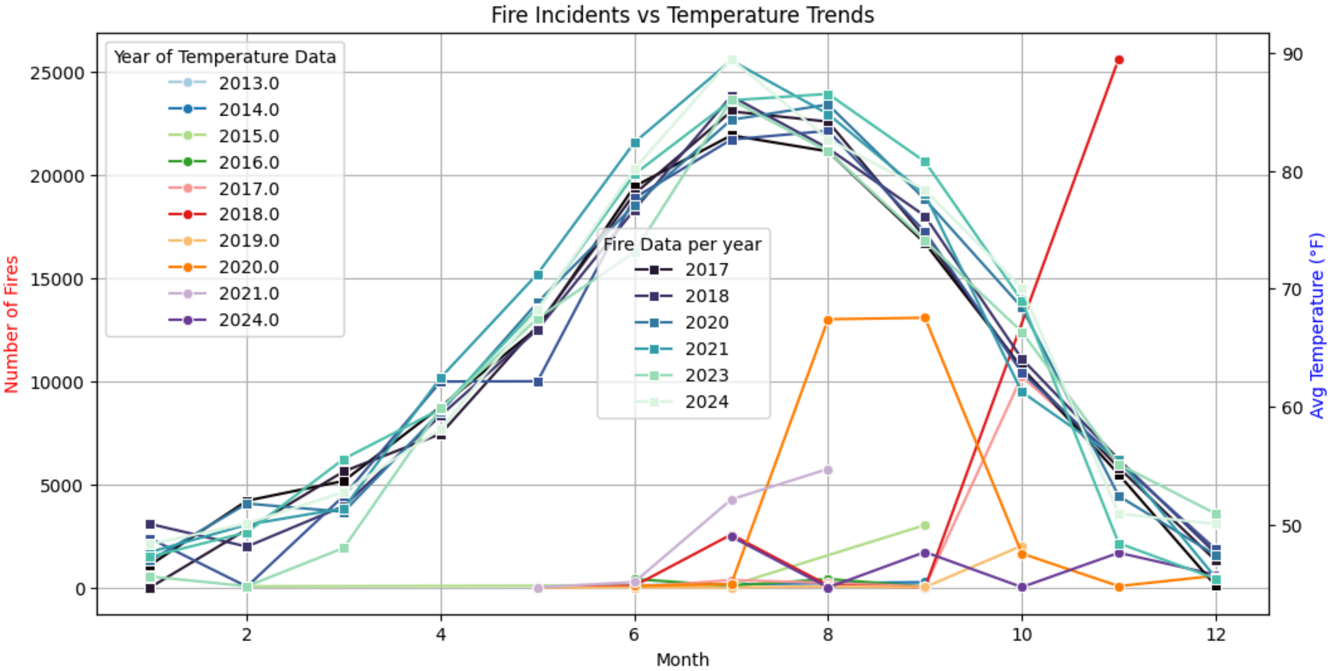


Fig 6: No of fire incidents and avg temperature of each month since 2016 (Author's creation, 2025)

4.8 dataset_loader_function.ipynb

I used this function to load the dataset easily in each of the files. Loaded here and called the function in each file according to their names.

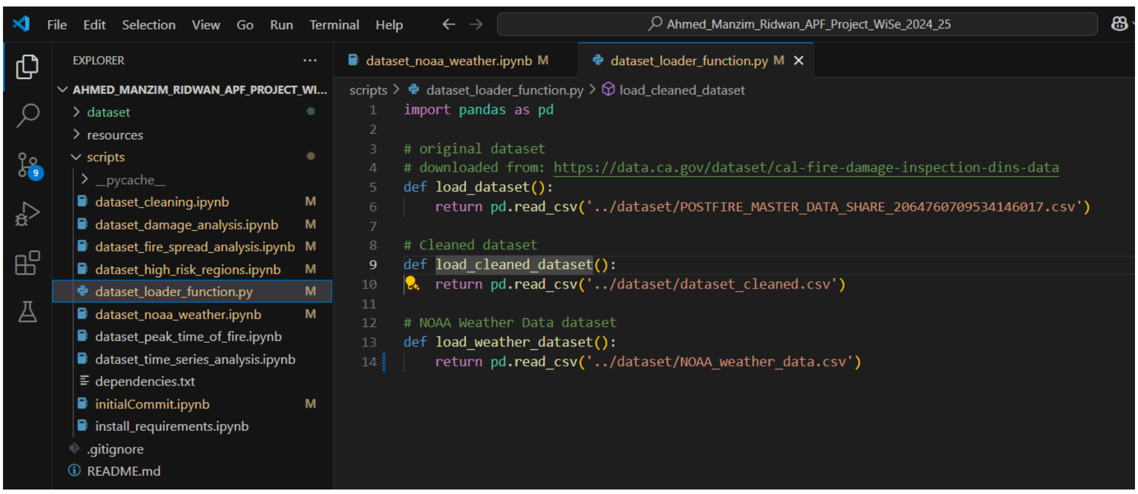


Fig 7.1: Dataset loading function

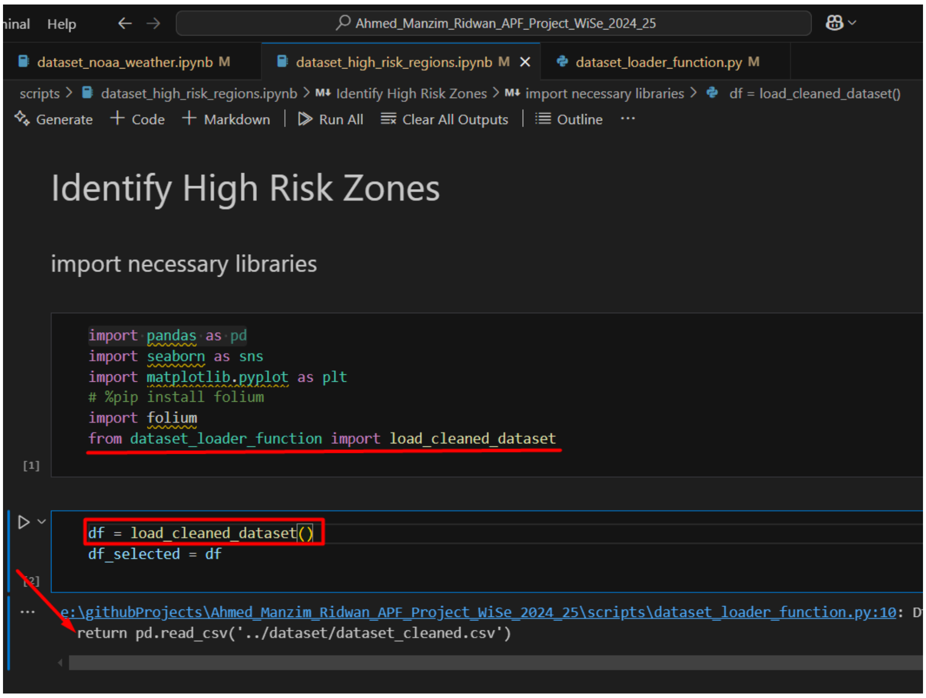


Fig 7.2: Calling the function in another file

4.9 dataset_high_risk_regions.ipynb

In this file, the top 5 states, counties, and cities with the most fire have been analyzed. For California state, top10 riskiest zones have also been shown. Finally, a map (using the folium) has been displayed that shows the fire incident spots.

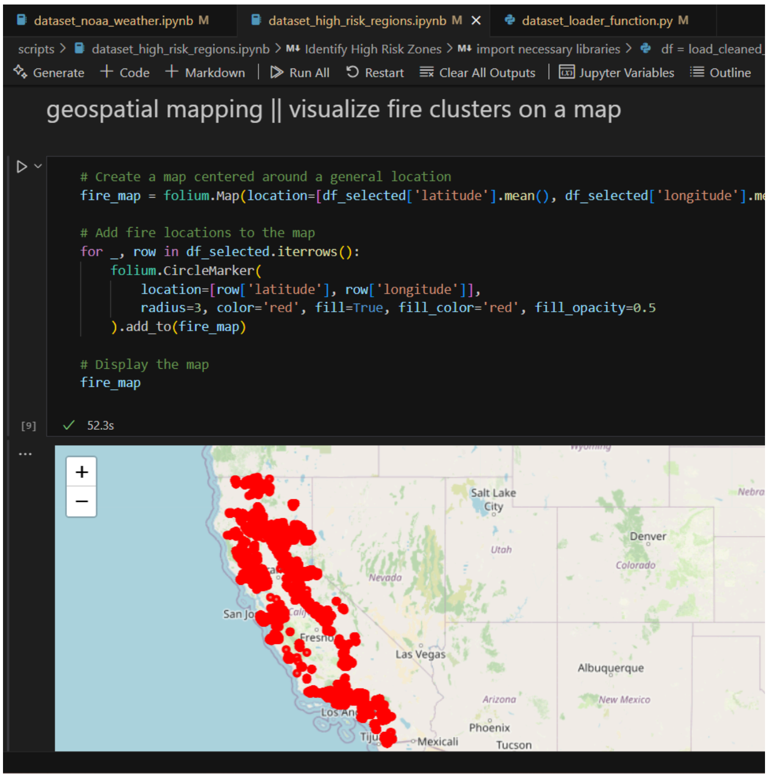


Fig 8: Fire incident spots (Author's creation, 2025)

4.10 dataset_fire_spread_analysis.ipynb

Simple barchart has been portrayed in this file depicting the peak months & year for fire incidents. Also, there's an interactive plot here.

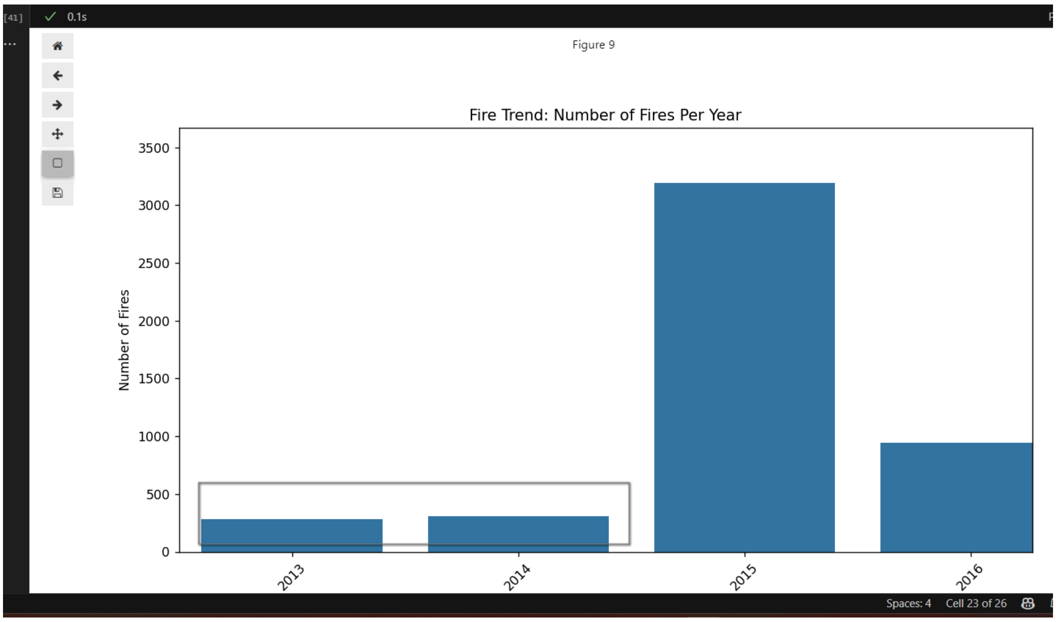
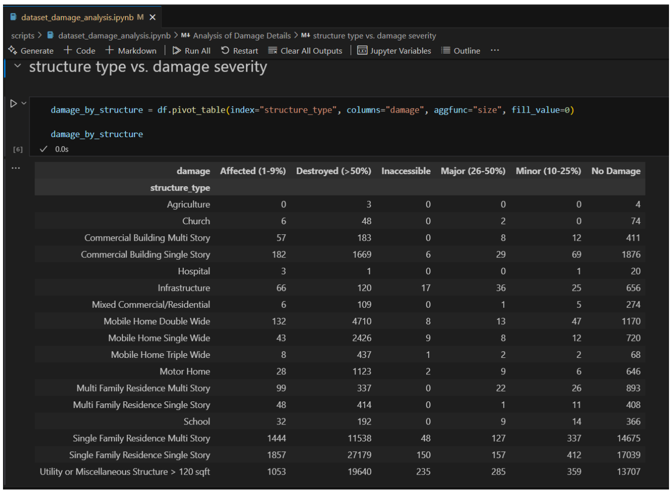


Fig 9: interactive plot (Author's creation, 2025)

4.11 dataset_damage_analysis.ipynb

Finding and visualizing major damage, structure type vs damage severity, financial loss amounts were main goals of this file. For zoom-in-out purposes, multiple interactive plots have also been added here.



dataset_damage_analysis.ipynb M

scripts > dataset_damage_analysis.ipynb > Analysis of Damage Details > structure type vs. damage severity

structure type vs. damage severity

```
damage_by_structure = df.pivot_table(index="structure_type", columns="damage", aggfunc="size", fill_value=0)

damage_by_structure
```

[6] 0.0s

damage / structure_type	Affected (1-9%)	Destroyed (>50%)	Inaccessible	Major (26-50%)	Minor (10-25%)	No Damage
Agriculture	0	3	0	0	0	4
Church	6	48	0	2	0	74
Commercial Building Multi Story	57	183	0	8	12	411
Commercial Building Single Story	182	1669	6	29	69	1876
Hospital	3	1	0	0	1	20
Infrastructure	66	120	17	36	25	656
Mixed Commercial/Residential	6	109	0	1	5	274
Mobile Home Double Wide	132	4710	8	13	47	1170
Mobile Home Single Wide	43	2426	9	8	12	720
Mobile Home Triple Wide	8	437	1	2	2	68
Motor Home	28	1123	2	9	6	646
Multi Family Residence Multi Story	99	337	0	22	26	893
Multi Family Residence Single Story	48	414	0	1	11	408
School	32	192	0	9	14	366
Single Family Residence Multi Story	1444	11538	48	127	337	14675
Single Family Residence Single Story	1857	27179	150	157	412	17039
Utility or Miscellaneous Structure > 120 sqft	1053	19640	235	285	359	13707

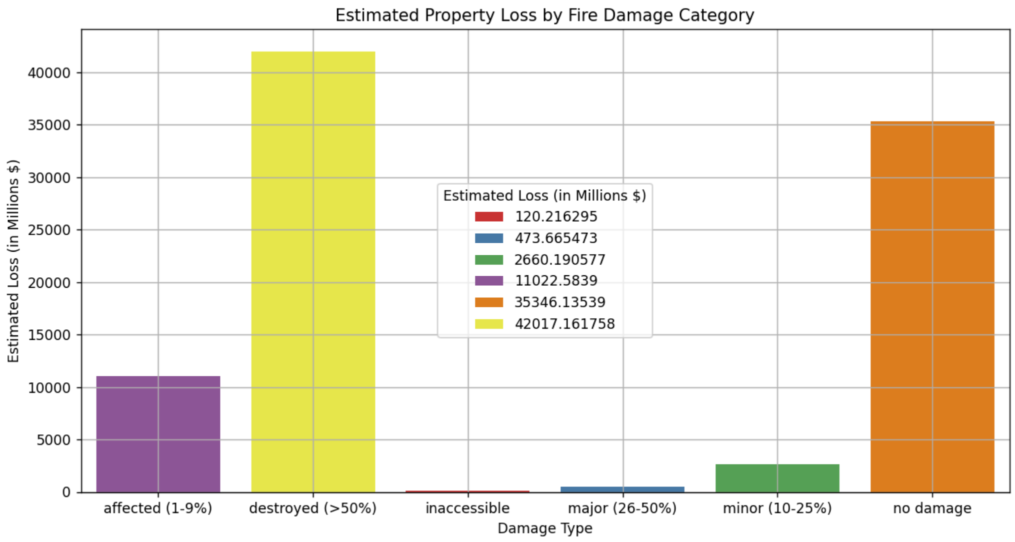


Fig 10: Various findings from dataset_damage_analysis.ipynb file (Author's creation, 2025)

5. Code Execution and Dataset Instructions

Here's the summary of the code execution. More details are on the readme page on the GitHub repository.

5.1 Prerequisite

It's suggested to have iPython, Git and VS Code installed in the system where this repository will be run. Since this project was done in VS Code, in another system, a few lines might not run. For example, `%matplotlib widget` was required for the interactive plots, then again, there is a chance that it won't run on jupyter servers. Hence, additionally, `pip install ipympl` might be required.

5.2 How to Access the Dataset

- The dataset can be downloaded from NOAA and California Open Portal Data.
- Ensure the datasets are placed in the `dataset` folder before running the scripts.

5.2 Running the Scripts

- Open the provided Jupyter Notebook Python scripts in VS Code.
- Execute step-by-step to preprocess, analyze, and visualize data or run all cells.

5.3 Dependencies

Install required libraries using `pip install -r dependencies.txt`

6. Conclusion

This study highlights the strong correlation between temperature variations and wildfire incidents in California. The findings emphasize the importance of proactive measures to mitigate fire risks in high-temperature regions and risky or frequent fire incident spots. Future work could involve integrating additional environmental variables, such as humidity and wind speed, to enhance predictive models.

7. References

- U.S. Department of Agriculture. (n.d.). *Wildfire*. USDA Climate Hubs. Retrieved from https://www.climatehubs.usda.gov/taxonomy/term/398

8. Figures

1. Figure 1: Ahmed Manzim Ridwan (Author, 2025)
2. Figure 2: Ahmed Manzim Ridwan (Author, 2025)
3. Figure 3: Ahmed Manzim Ridwan (Author, 2025)
4. Figure 4: Ahmed Manzim Ridwan (Author, 2025)
5. Figure 5: Ahmed Manzim Ridwan (Author, 2025)
6. Figure 6: Ahmed Manzim Ridwan (Author, 2025)
7. Figure 7.1: Ahmed Manzim Ridwan (Author, 2025)
8. Figure 7.2: Ahmed Manzim Ridwan (Author, 2025)
9. Figure 8: Ahmed Manzim Ridwan (Author, 2025)
10. Figure 9: Ahmed Manzim Ridwan (Author, 2025)
11. Figure 10: Ahmed Manzim Ridwan (Author, 2025)