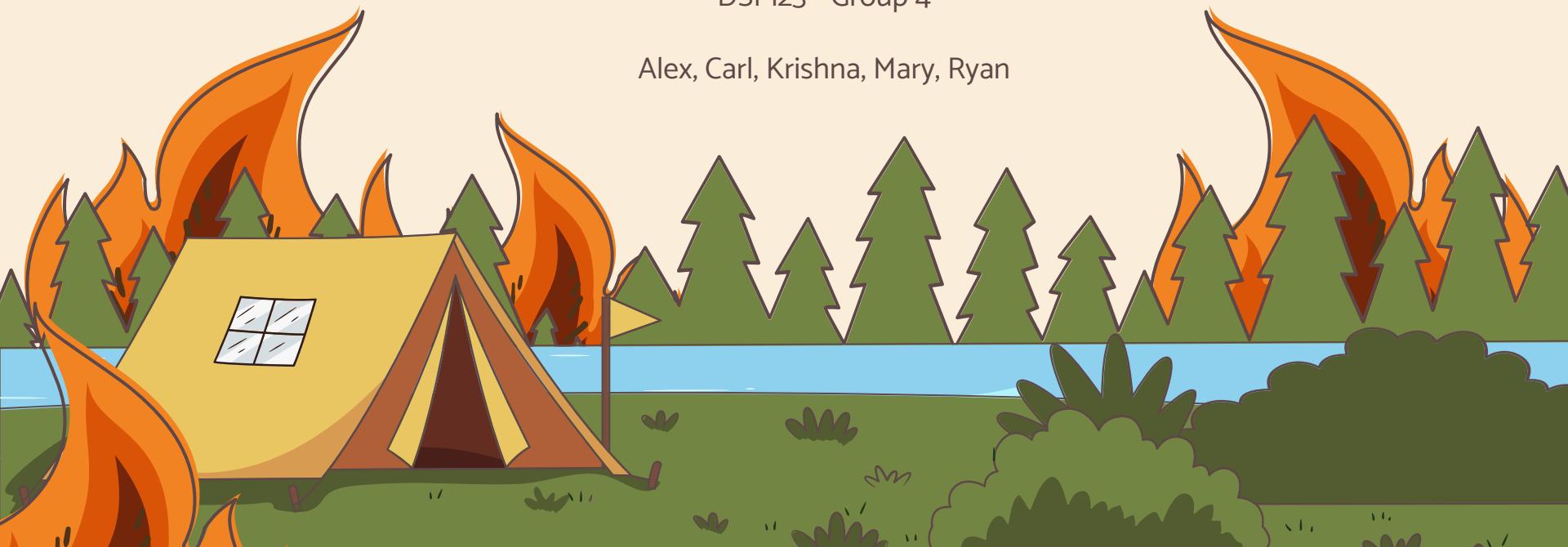


Wildfires in California

DSI 123 - Group 4

Alex, Carl, Krishna, Mary, Ryan



Our Group



Alex



Carl



Krishna



Mary



Ryan

Table of contents

01

Problem

What problem are we trying to solve

02

Preprocessing

What did we do to prepare the data for modeling?

03

Modeling

What models did we create?

04

Data Visuals

What does our data look like?

05

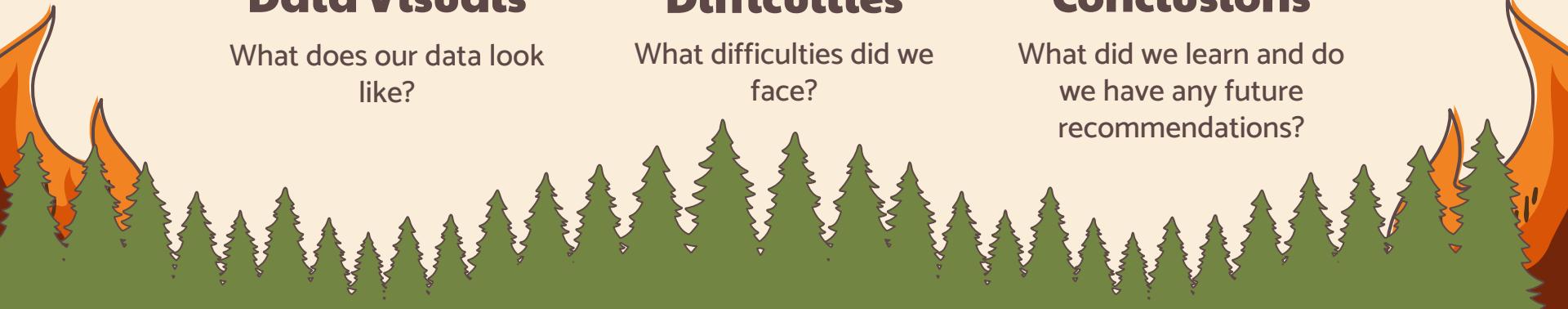
Difficulties

What difficulties did we face?

06

Conclusions

What did we learn and do we have any future recommendations?



01

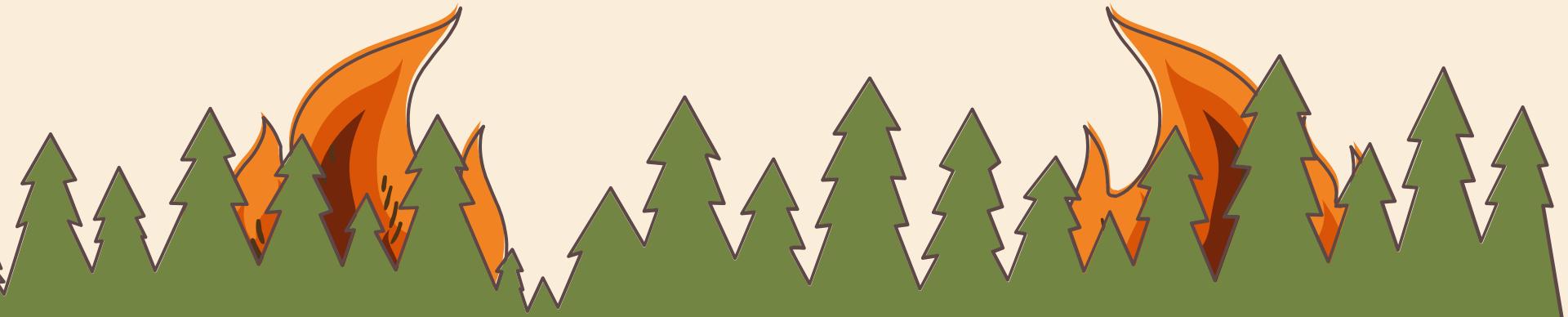
Problem



Problem Statement

The project aims to create a model to accurately identify the cause of wildfires using various features, with the dataset containing information on the statistical cause of fires. The importance of this lies in the potential catastrophic consequences of wildfires and the need to prevent them.

Accurate prediction can help identify prevention strategies, allocate resources, and improve response time. Identifying the cause can also lead to a better understanding of contributing factors and steps to mitigate their impact.



U.S. Wildfires!

We started with a spatial database of
wildfires that occurred in the United
States from 1992-2013.

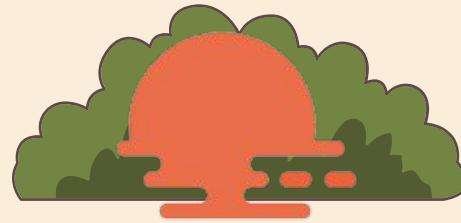


API's



SoilGrids API

API that uses soil profile and covariate data to model the spatial distribution of soil properties.



OpenMeteo API

API that accepts geographical coordinates (longitude and latitude) and returns the terrain elevation for those points

02

Preprocessing



EDA & Feature Engineering

1. Explored trends in the data
2. Removed nulls
3. Removed duplicates
4. Created new bins

Goal: redefine our target variable, which are the causes of wildfires.

Target Variables

Lightning
Miscellaneous
Equipment Use
Arson
Campfire
Missing/Undefined
Debris Burning
Smoking
Children
Powerline
Railroad
Fireworks
Structure

**Lightning
Negligence
Infrastructure
Arson**

03

Models



Models

Baseline



**Logarithmic
Regression**



KNN



XGBoost



Neural Net

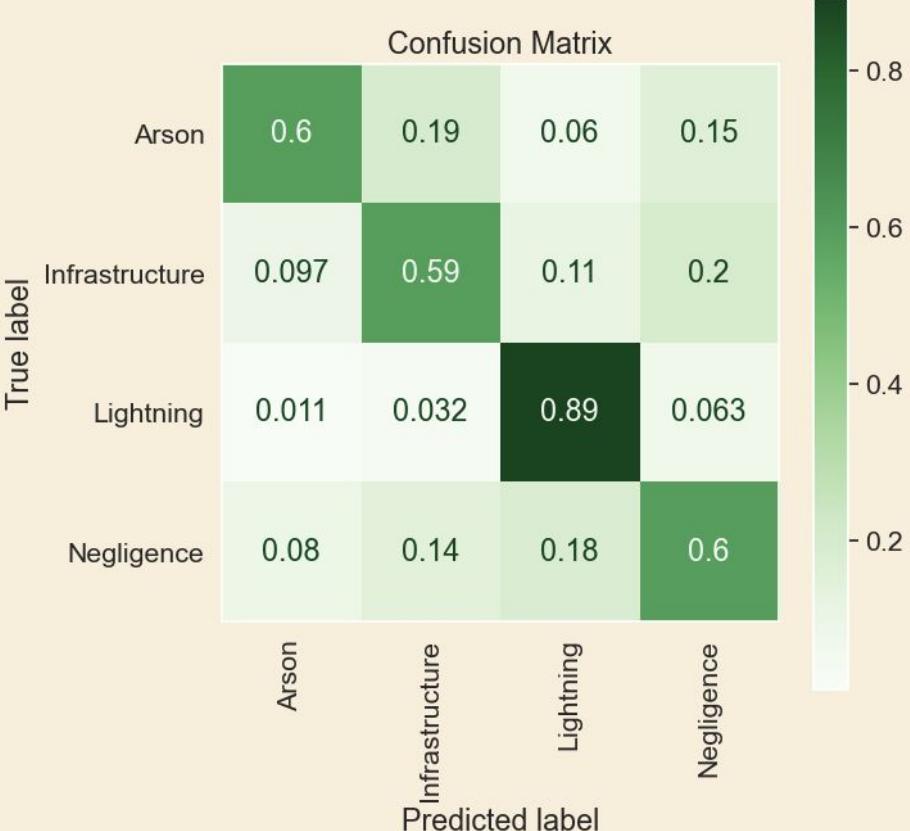


**Random
Forest**



Confusion Matrix

RandomForestClassifier
n_estimators=100,
max_depth=None,
min_samples_split=2,

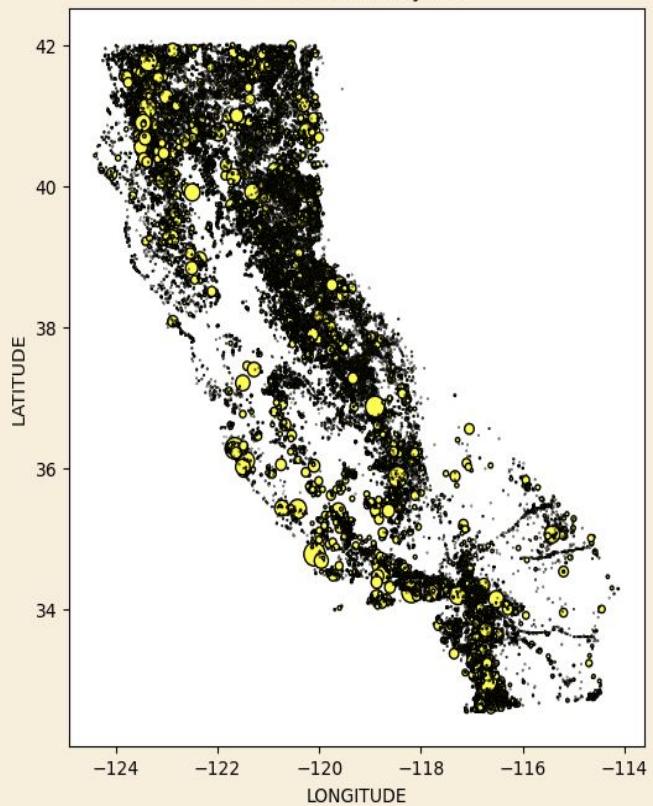


04

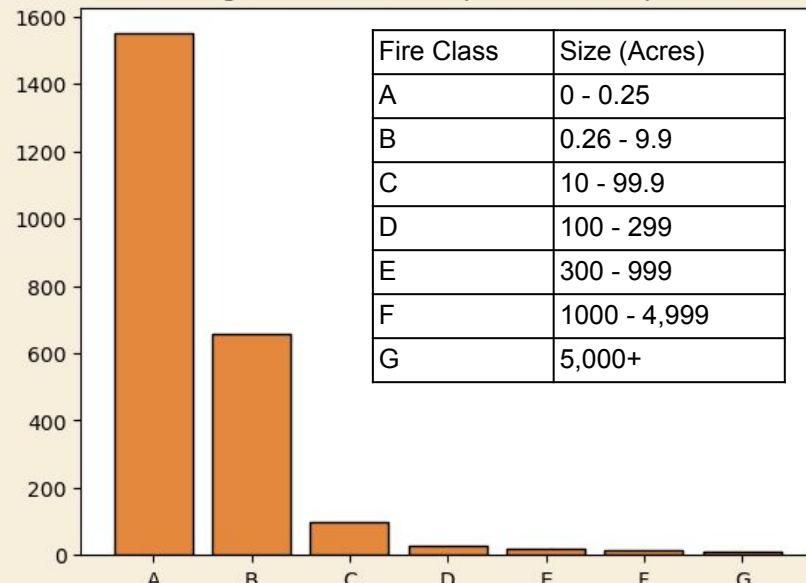
Data Visuals



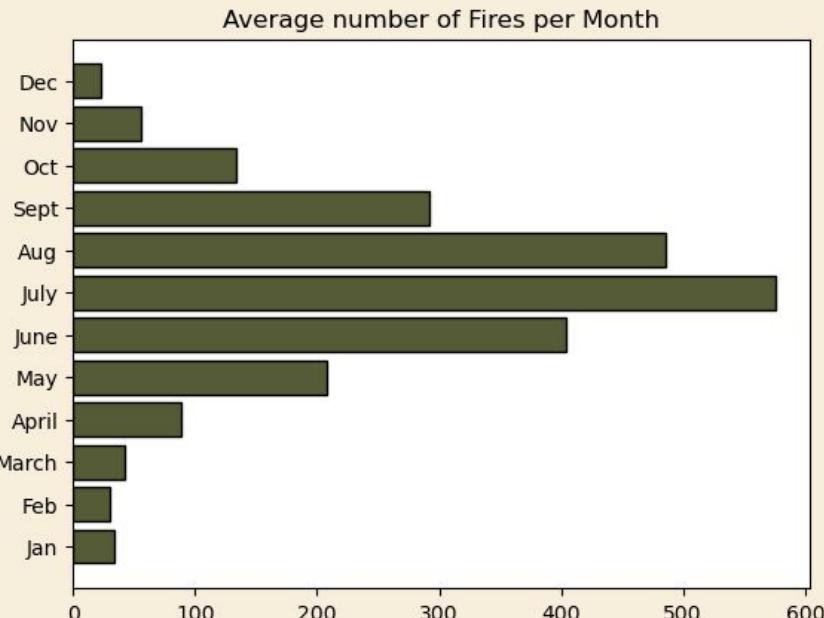
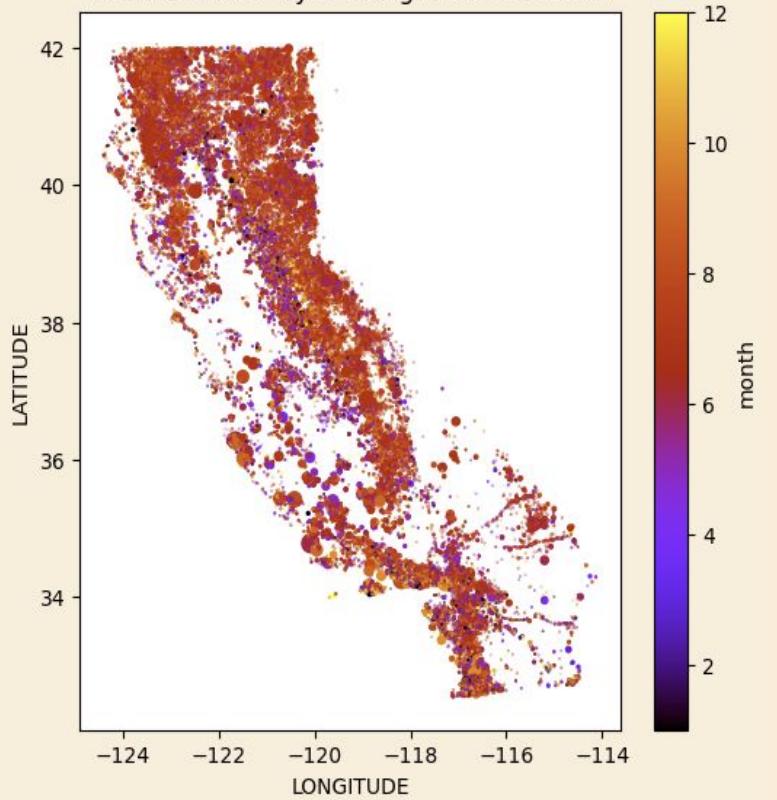
California Fires by Size



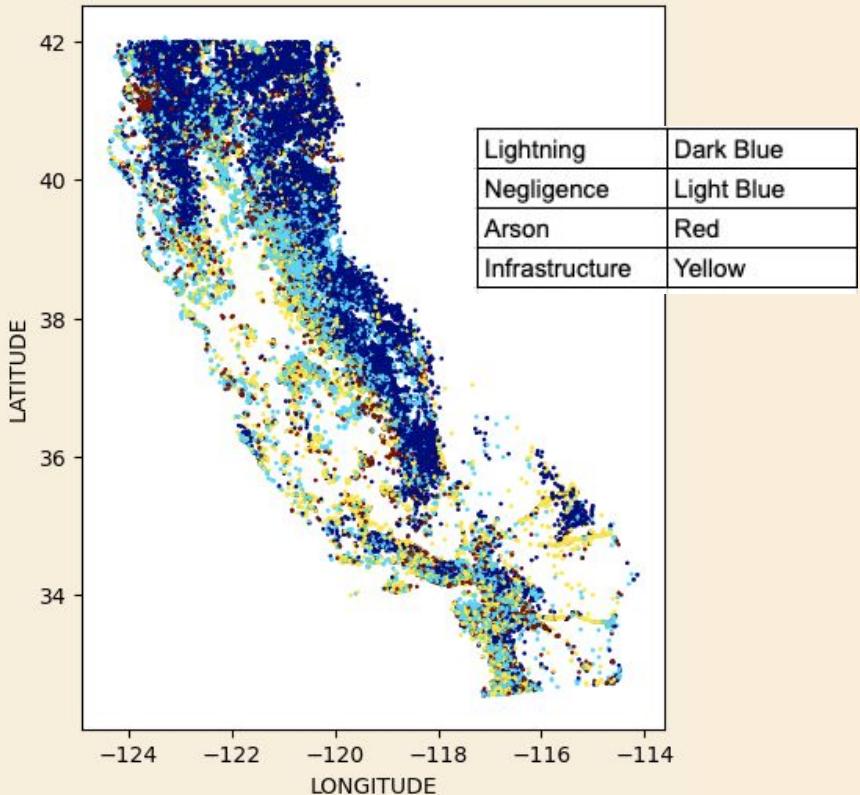
Average Number of Fires per Class Size per Year



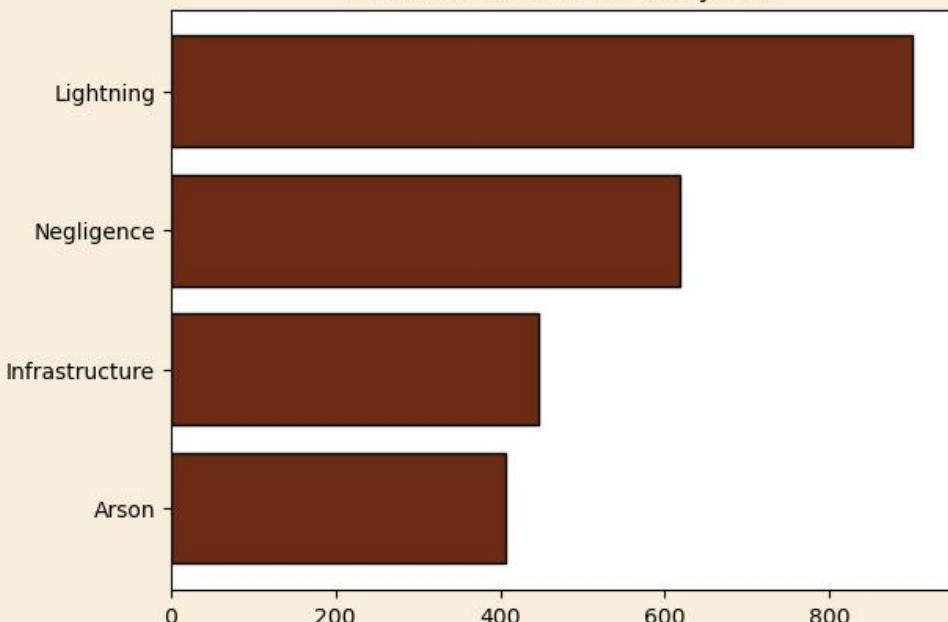
California Fires by Starting Month and Size



California Fires by Cause



Causes of California Fires by Year



05

Difficulties



Difficulties

- Navigating a file too large for GitHub.
- Reducing data from 1.88 million data points to ~55,000.
- Calling API's and formatting final data frame for modeling.
- Identifying target variable.
- Handling uncertain classes.



06

Conclusions and Recommendations



Conclusions

- Wildfires are a global threat, and this project developed a predictive model that accurately identifies the cause of wildfires in California with a 71% accuracy rate, using various features such as soil composition, elevation data, and geo coordinates.
- The model can be further improved by incorporating additional data resources and refining the "Miscellaneous" cause description in the dataset.
- A more accurate model can help public officials and first responders develop more effective strategies to manage and prevent wildfires, ultimately reducing their impact.



Recommendations

- Educate the community about the most common causes of wildfires in the area, such as lightning strikes, human activity, and equipment failure.
- Encourage citizens to report any suspicious activity or behavior that may lead to a fire.
- Train local firefighters and first responders to identify the cause of a fire and collect evidence for investigation, which will further improve this dataset
- Use technology such as drones and cameras to observe high-risk areas and gather data on environmental factors that could contribute to a fire's occurrence.
- Gather additional data from the provided geo-coordinates that may be more pertinent to identifying the cause of a wildfire



References

- 1. SlidesGo (Presentation Template)
- 2. Medium - How Do Neural Nets Really Work? By Chris Landschoot (<https://tinyurl.com/2p8rst4z>)
- 3. Kaggle (1.88 million wildfires dataset)
(<https://www.kaggle.com/rtatman/188-million-us-wildfires>)
- 4. Open-Meteo API (<https://open-meteo.com/>)
- 5. SoilGrids API (<https://soilgrids.org/>)





Thanks!
Do you have any
questions?

