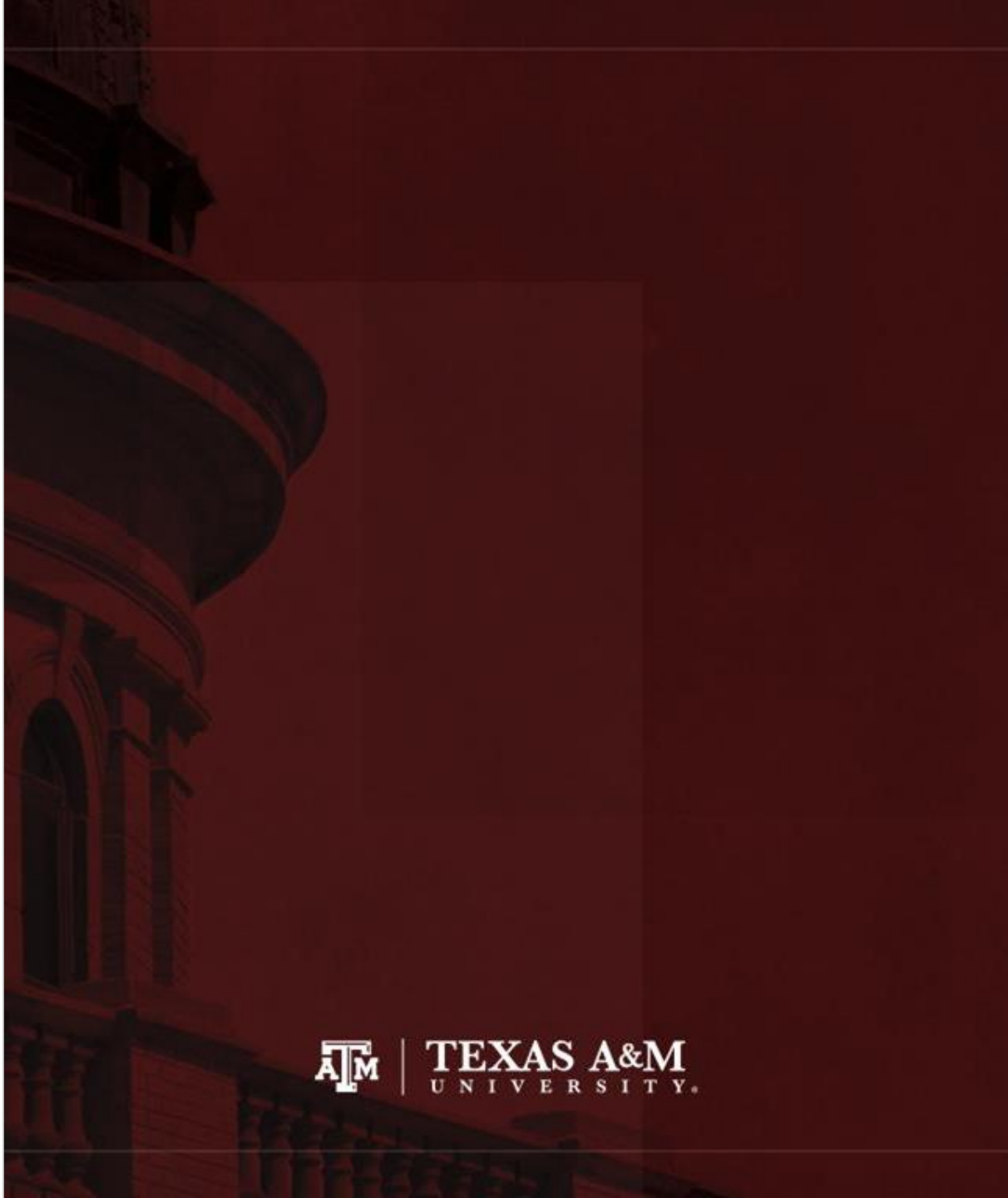
Datanators

2023 Data Science Competition

Rohit Dube, Tushar Pandey, Arpan Pal, Mansi Bezbaruah, Benjamin Warren

Github:
https://github.com/ManaswineeB/datanators



Problem Statement

- Being better prepared for wildfires
- Minimizing the losses and damages

Proposed Solution

1. Allocation of stations for wildfire control preparedness among states and resource assignment among these stations
2. Checking the effectiveness of existing resources and provide recommendations for effective resource allocation.

Data Collection and Preprocessing

Data Sources

- Weather: National Center for Environment Information
  https://www.ncei.noaa.gov/
- Wildfire: National Interagency Fire Center https://www.nifc.gov/
  Texas A&M Forest Service https://tfsweb.tamu.edu/
- Location: State Maps https://simplemaps.com/
  County Maps https://www.opendatasoft.com/
  Location and distance https://project-osrm.org/

Preprocessing and Feature Engineering

- Wildfire location, area burnt and time variables are consolidated with cause of fire, weather and county information

Initial Findings of Analysis

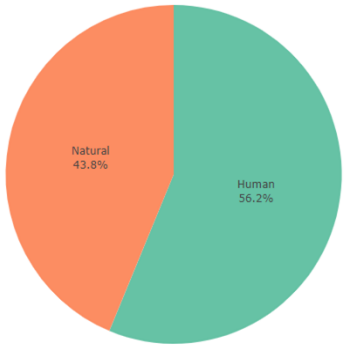


Fig 1: Causes of wildfire

- More than half of the wildfire cause are attributed to humans
- About 90% of wildfires in Texas are caused by humans*

- Naturally occurring wildfires are increasing
- The prediction model would give better results due to weather patterns

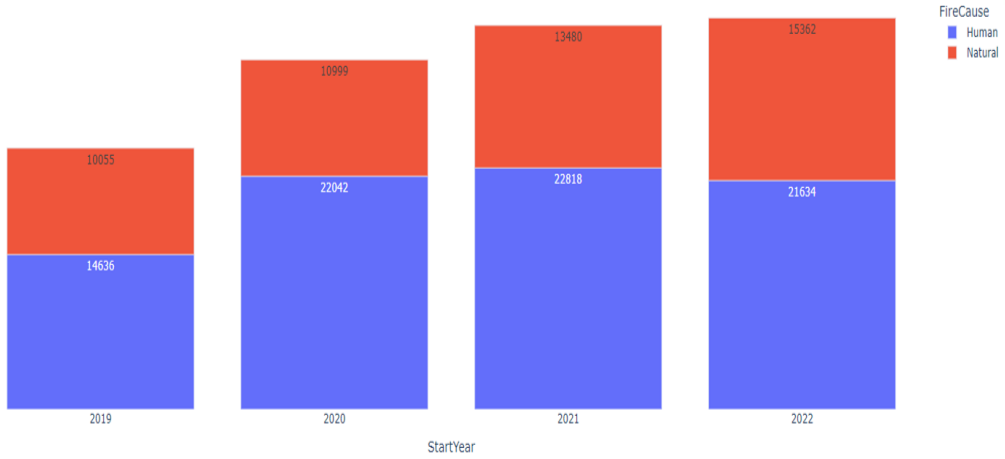


Fig 2: Distribution of Natural and Human wildfires

*refer:https://tfsweb.tamu.edu/HuntingFireSafety/

Effect of Months on Areas burnt

- The intensity of Wildfires reduce during Jan and December for most states
- The higher intensity varies between april to august

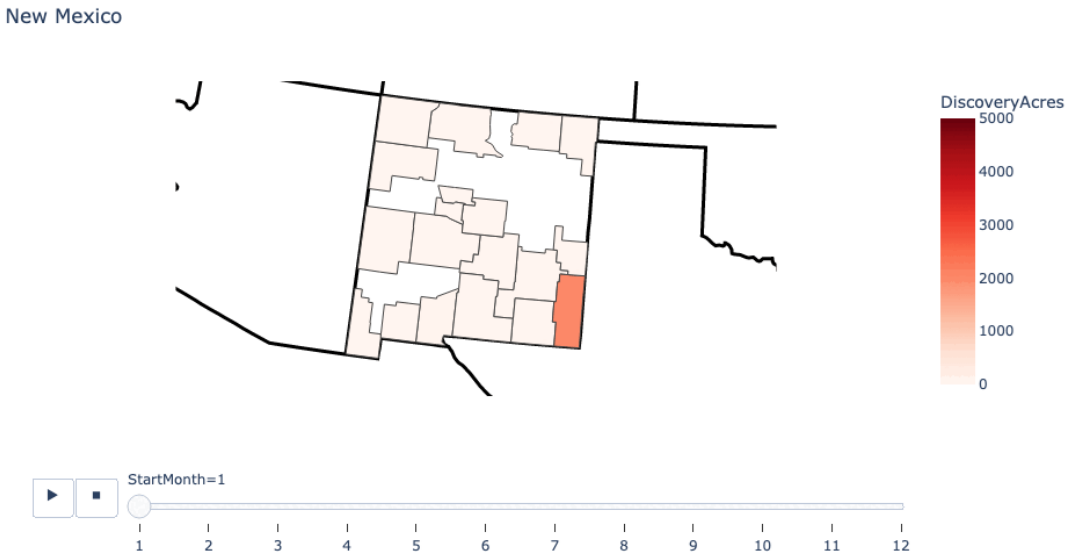


Fig 3: Fires in New Mexico over different months. This shows trend in the area burnt over different months. For the months of May, June and July, the area burnt is more, as well as more densely burnt patches inside the state.

Wildland Fire Likelihood

- Assigning a score describing the likelihood of repetition of wildfire around an incident location.
- Calculated using factors like weather, time, county demographics, location and the intensity of the fire.

$$\mathcal{WFL} = \frac{10 - \hat{A} - \hat{R} + \hat{T} - 0.01(M-12)(4+M) + 0.1(PD)^{-1} + 0.01(\mathcal{M}ax(L) - L)}{100}$$

$\hat{A}$: Normalized area burnt for that incident

$\hat{R}$: Normalized rainfall for the month in that county

$\hat{T}$: Normalized max temp for the month in that county

M : Month of the incident

PD : Population density

L : Latitude of the location

Wildland Fire Likelihood

- The normalized scores are aggregated for different counties, Los Angeles County has the highest WFL score (>12000 incidents)
- The state of California records the highest number of wildfires, most of the top WFL counties are from California.

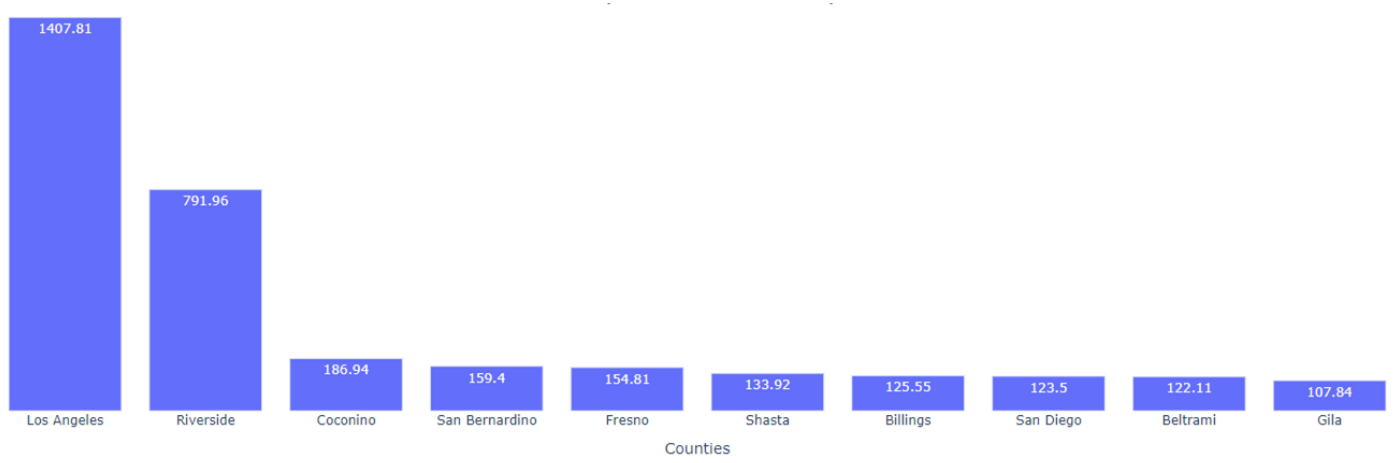


Fig 4: Wildland Fire Likelihood

Modelling as an Integer program

- Historical wildfire location data is used to find appropriate locations for wildfire control stations
- An integer programming model is proposed to find locations for the control stations along with a number for a given county

Minimize $\sum_{i,j} d_{ij}x_{ij}$

Subject to: $\sum_j x_{ij}=1 \ \forall i \ ; \ \sum_i x_{ij} < k y_j \ \forall j \ ; \ \sum_j y_j \leq m$

x_{ij} equals 1 if location i looks over station j,
y_j takes the value 1 when station is built at location j,
d_{ij} is the distance between location i and station j,
m is the maximum stations that can be built

County-level Stations plots

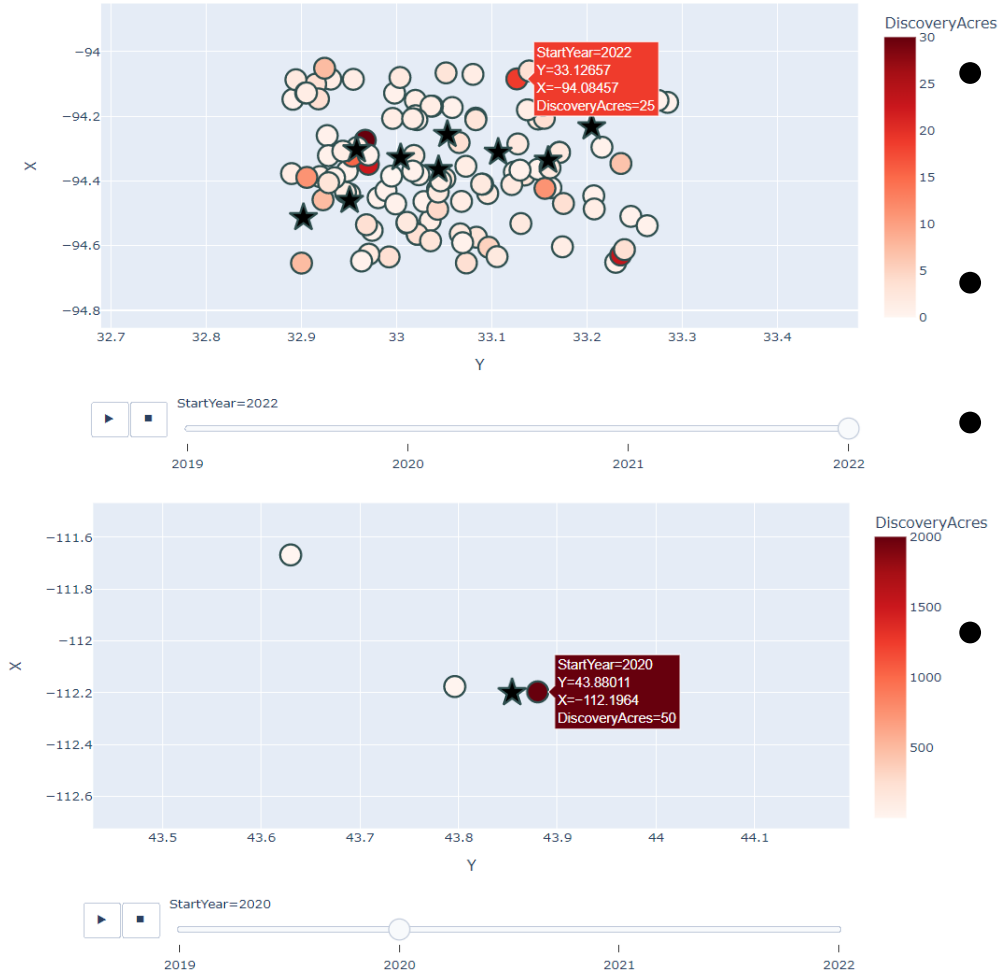


Fig 5: County-wise Station assignment by IP model

- Wildfire locations are plotted with the stations using integer programming
- The dynamic plots help us match the intuition with results
- Cass County(top) has a high volume of wildfires, therefore 9 stations
- Butte county, (bottom) does not have many wildfires, therefore, we get one station in the IP model.

Resource Allocation within State

- Regional division used to reduce the computation
- Requirement of resources proportional to number and intensity of wildfire in the region
- Wildfire stations with high priority should be classified
- Creation of a metric based on the duration of travel from the 19 wildfire stations to the wildfires

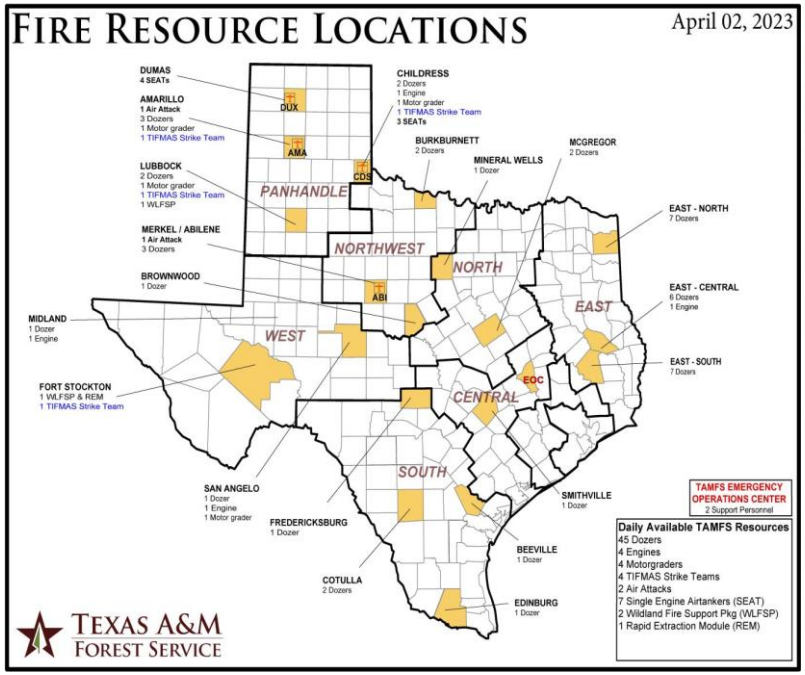


Fig 6: Resource Allocation in Texas

County Wildfire Score

- CWS helps us to identify the counties which were highly affected by wildfires
- The score takes into account wildfire occurrences and areas burnt and helps in identifying repeated occurrences of wildfires

$$CWS = \frac{3 \times \hat{I} + \hat{A}}{4}$$

Î : Normalized number of wildfire occurrences in a county

Â: Normalized Area (acres) burnt in a county

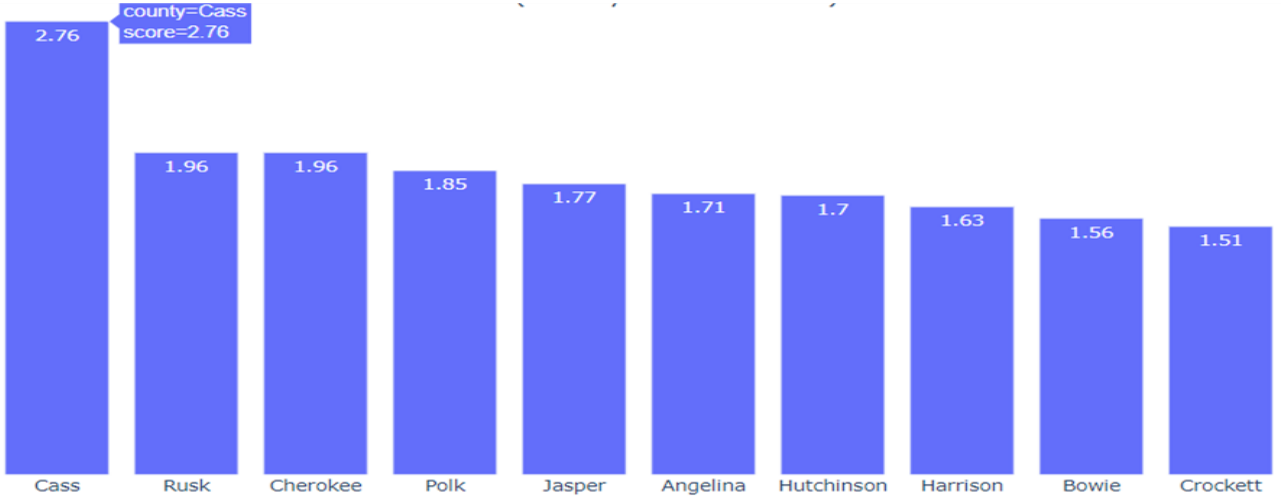


Fig 7: CWS for Counties in Texas

Weighted Wildfire Frequency

The Weighted Wildfire Frequency Score for each station is calculated using this duration of travel weighted by the CWS of the county in which the wildfire was located.

$$\mathcal{WWF}(stn) = \sum_{L_i} d(L_{stn}, L_i) \times \mathcal{CWS}_i$$

WWF(stn): The Weighted Wildfire Frequency Score for the wildfire station i

L_{stn} : The location of wildfire station

L_i : The location of wildfire in the region of wildfire station

$d(L_{stn}, L_n)$: Driving duration between two locations in hours

CWS_i : County Wildfire score of the county of the wildfire location

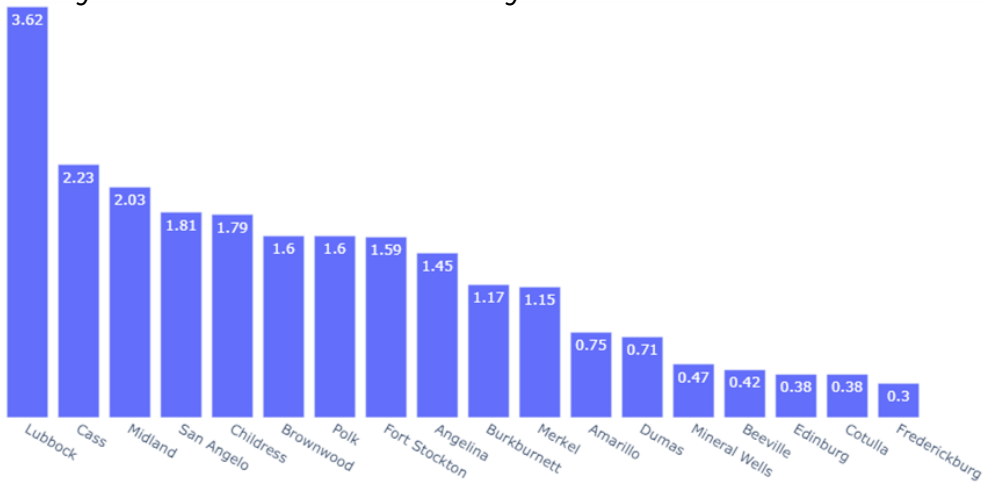


Fig 8: WWF for Stations in Texas

Conclusion and Remarks

- There is a clear seasonal trend in the wildfires across the entire country.
- The counties can set up first response teams based on the likelihood of wildfire in the area.
- The resources can be allocated in a way that minimizes the cost of operation and reduces the “Time-to-action”.
- With external ML models and an accurate cost/budget, the model can be extended to a full scale, self-sufficient model to cut down loss + damages caused due to wildfire across the USA.

Future Work

- Include the cost associated with extinguishing the wildfire based on time.
- Adding a budget constraint to find the suitable number of local first response stations.
- Acquiring the cost sheet for different resources, making a MIP model that allocates the resources into different stations.
- Adding the cost of transfer of resources.

THANK YOU

Special thanks to our sponsors