

A black and white photograph of a wildfire scene. In the foreground, a firefighter is visible, holding a hose and battling the flames. The ground is covered in fire and smoke. In the background, several large trees stand against a dark, smoky sky. The overall atmosphere is one of intense danger and emergency response.

LINEAR REGRESSION

CALIFORNIA WILDFIRES

BACK STORY

California Department of Forestry and Fire Protection (CalFire)

Acres burned Google Maps area tool

<https://www.daftlogic.com/projects-google-maps-area-calculator-tool.htm#>

BACK STORY



8k

wildfires

BACK STORY



8k
wildfires



5.8k
structures

BACK STORY



8k
wildfires

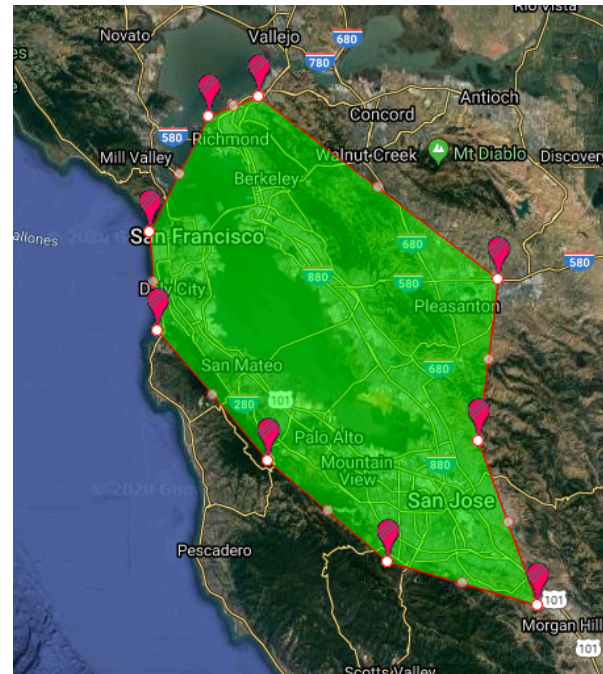


5.8k
structures



900k
acres burned

BACK STORY



<https://www.daftlogic.com/>



900k
acres burned



\$400B

AccuWeather estimated the total damage and economic loss caused by wildfires cost California \$400 billion in 2018 and \$85 billion in 2017.

<https://www.accuweather.com/en/weather-news/california-wildfires-will-cost-tens-of-billions-accuweather-estimates/612548>

OBJECTIVE

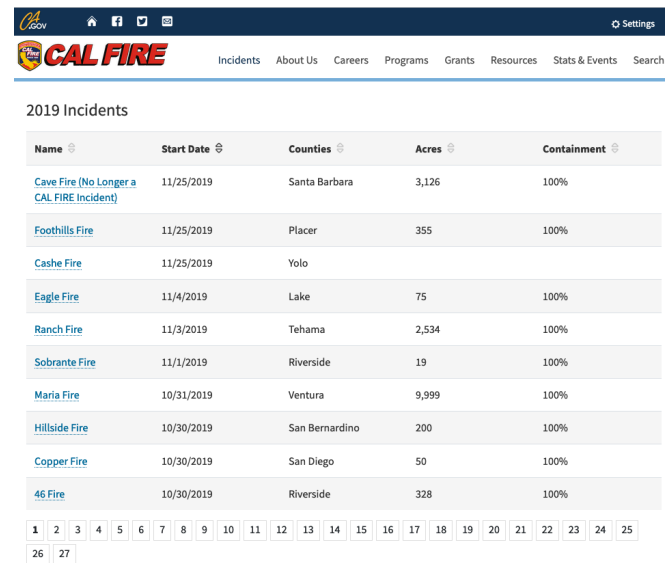
Predict acres burned using
linear correlation

Predict acres burned given a California wildfire by using linear regression to identify relationships between acres burned and **weather & time**

By accomplishing this, we can then look into how many resources might be needed to combat wildfires

DATA SOURCES

METHODOLOGY



The screenshot shows the CAL FIRE website's 'Incidents' page. It features a table titled '2019 Incidents' with columns for Name, Start Date, Counties, Acres, and Containment. The table lists 12 incidents, including Cave Fire, Foothills Fire, Cashe Fire, Eagle Fire, Ranch Fire, Sobrante Fire, Maria Fire, Hillside Fire, Copper Fire, and 46 Fire. A pagination bar at the bottom shows page 1 of 27.

Name	Start Date	Counties	Acres	Containment
Cave Fire (No Longer a CAL FIRE Incident)	11/25/2019	Santa Barbara	3,126	100%
Foothills Fire	11/25/2019	Placer	355	100%
Cashe Fire	11/25/2019	Yolo		
Eagle Fire	11/4/2019	Lake	75	100%
Ranch Fire	11/3/2019	Tehama	2,534	100%
Sobrante Fire	11/1/2019	Riverside	19	100%
Maria Fire	10/31/2019	Ventura	9,999	100%
Hillside Fire	10/30/2019	San Bernardino	200	100%
Copper Fire	10/30/2019	San Diego	50	100%
46 Fire	10/30/2019	Riverside	328	100%

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
26 27



Using Python – my approach was to scrape all wildfire data available from Cal Fire (**last 7 years**): acres burned, started, location by latitude/longitude
<https://www.fire.ca.gov/>

Using the Dark Sky API, I made requests get weather data: temperature, humidity, windspeed, etc.

My final dataset included about 1300 rows where each row represented a wildfire along with weather attributes

METHODOLOGY



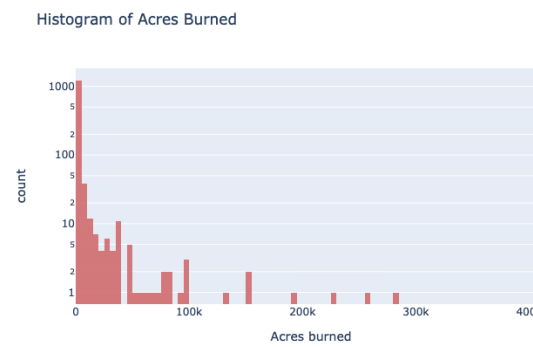
METHODOLOGY



TARGET VARIABLE

METHODOLOGY

Transform target variable → acres burned

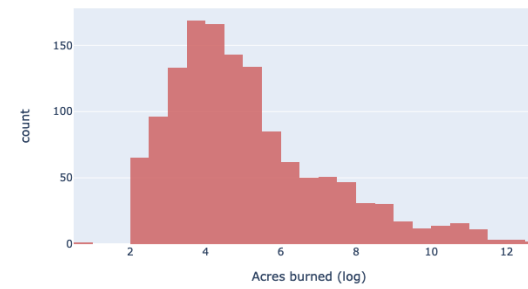


Acres burned is highly skewed right

METHODOLOGY

Transform target variable → acres burned

Histogram of Acres Burned (log)



METHODOLOGY

1. Split my data into 80/20 train/test
2. Standardized the data so that all features are on the same scale
3. OLS Model – mainly to see summarized results (p-values)
 1. p-values helped identify features to drop
 2. R^2 after dropping features
4. Lasso Model
 1. Helped zero out coefficients
5. Prediction

METHODOLOGY

TRAIN/TEST

METHODOLOGY

TRAIN/TEST

STANDARDIZE

METHODOLOGY

TRAIN/TEST

STANDARDIZE

OLS

METHODOLOGY

TRAIN/TEST

STANDARDIZE

OLS

LASSO

METHODOLOGY

TRAIN/TEST

STANDARDIZE

OLS

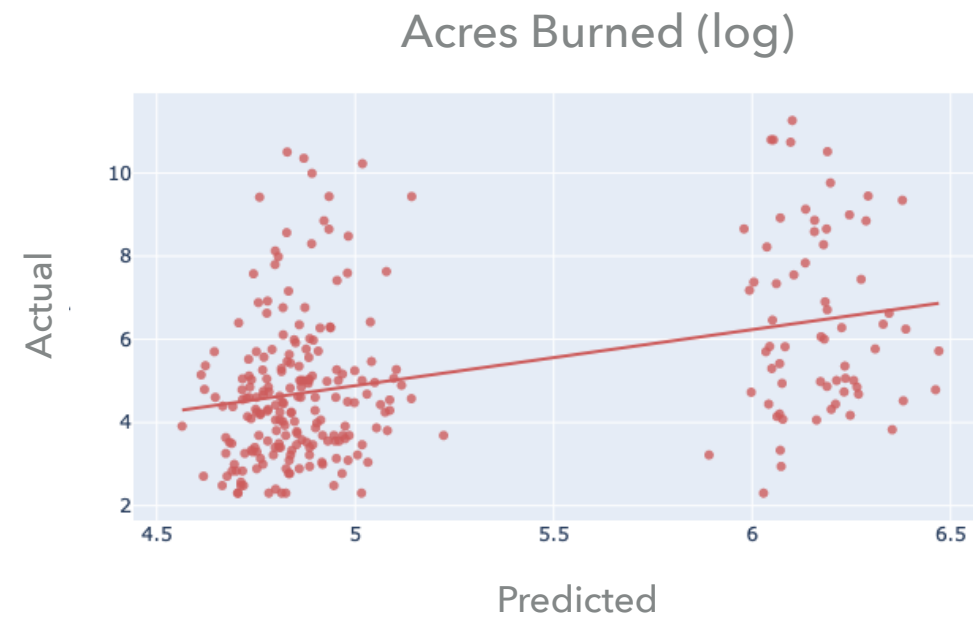
LASSO

PREDICTION

MODEL

RESULTS

R^2 score = 0.14



Ineffective model

Two clusters (bi-modal distribution?)

```
x = y_predict (lasso_model.predict(X_te)
```

```
y = y_test
```

FUTURE WORK

- ▶ Larger weather date range
- ▶ Categorical features
- ▶ Cross validation
- ▶ Model comparison

Need more data, particularly weather data to include a larger date range

Categorical features such as **season, location**

“We don’t really have a fire season anymore; it’s really a fire year.”

Cal Fire Deputy Director

Quote

<https://www.accuweather.com/en/weather-news/california-wildfires-will-cost-tens-of-billions-accuweather-estimates/612548>



Questions:

Why did I only choose Lasso?

- mainly concerned with trimming features
- helped zero out coefficients

Number of features

- started with 13 and ended with 5