

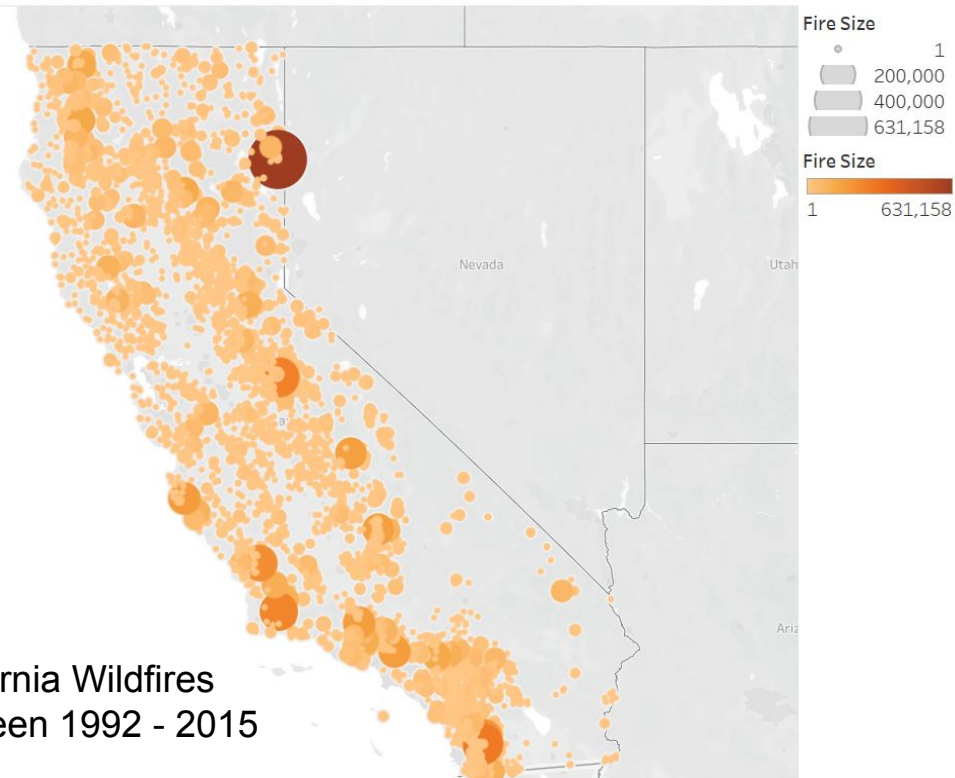
# California Wildfire Analysis & Prediction Model

By Meareg Mamo, Alejandra Dahl, Julia Behling & Remi  
Harazim

# Why We Picked This Topic:

- Effect on Human Health and Welfare
- Impact on Air Quality and Ecological/Landscape Dynamics
- Insurance claim and probability models
- Prediction models for public safety
- Personal Experiences

California Wildfires  
Between 1992 - 2015



# Dataset Sources:

Two Data Sources were utilized to increase the number of comparable variables:

Variables:

- Ecological Factors (Vegetation type)
- Weather measurements (Temp, Precipitation, Wind speed, etc.)
- “Remoteness” or proximity to fire relief resources
- Funding based on County

Kaggle

U.S. Wildfire Data (and other attributes):

- <https://www.kaggle.com/cloudcoder/us-wildfire-data-plus-other-attributes>

Data.gov

CAL Fire Facilities for Wildland Fire Protection:

- <https://catalog.data.gov/dataset/cal-fire-facilities-for-wildland-fire-protection>

# Tools and Technology Utilized for The Project:

## And How Each Were Applied

- Microsoft Excel
  - Initial Data Exploration
- Python
  - Jupyter Notebook
  - Pandas
  - Sklearn
- Postgres SQL
  - Hosting Database
- Github
  - Organization & Collaboration
- Tableau
  - Visualization & Analysis



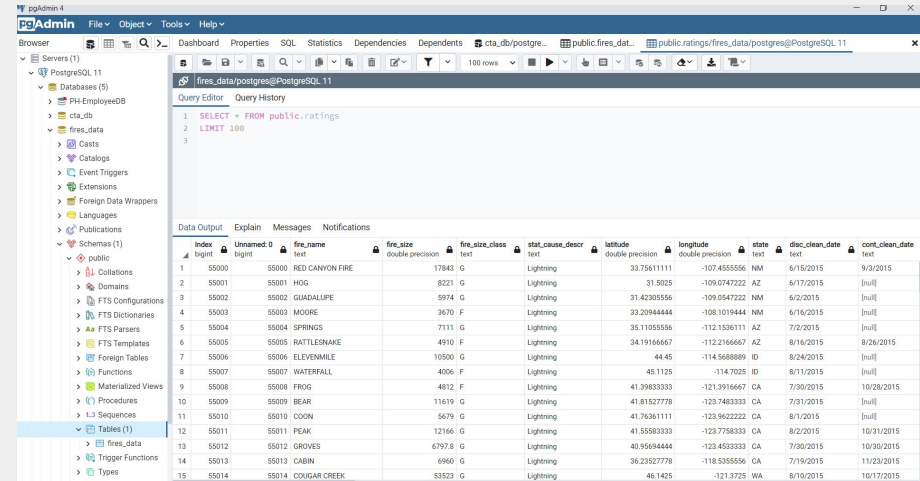
# Questions We Hope to Answer:

- Can past wildfires be used to predict future ones?
- Which Ecological factors are the best predictors?
- Does funding/availability of resources have an effect on fire fighting ability?

# The Data Exploration Phase:

## ETL Process

- Original CSV file from Kaggle.com
- Data cleaning process using Pandas dataframes
- Export to Postgres SQL Database



The screenshot shows the pgAdmin 4 interface. On the left, the 'Servers' tree is expanded to show a PostgreSQL 11 connection. The 'Databases' tree is also expanded, showing a database named 'public'. The 'Tables' tree is expanded, showing a table named 'fires\_data'. The 'Query Editor' is open, showing a SQL query: `SELECT * FROM public.ratings`. The 'Query Output' tab is active, displaying a table with 15 rows and 11 columns. The table contains data about fires, including fire\_id, fire\_name, fire\_size, fire\_class, start\_cause\_desc, latitude, longitude, state, disc\_clean\_date, and cont\_clean\_date.

Index	fire_id	fire_name	fire_size	fire_class	start_cause_desc	latitude	longitude	state	disc_clean_date	cont_clean_date
1	55000	RED CANYON FIRE	17843	G	Lightning	33.75611111	-107.4555556	NM	6/15/2015	9/3/2015
2	55001	HOG	8221	G	Lightning	31.5025	-109.0747222	AZ	6/17/2015	[null]
3	55002	GUADALUPE	5974	G	Lightning	31.42305556	-109.0547222	NM	6/2/2015	[null]
4	55003	MOORE	3670	F	Lightning	33.20944444	-108.1019444	NM	6/16/2015	[null]
5	55004	SPRINGS	7111	G	Lightning	35.11055556	-112.1836111	AZ	7/2/2015	[null]
6	55005	RATTLESNAKE	4910	F	Lightning	34.19166667	-112.2166667	AZ	8/16/2015	8/26/2015
7	55006	ELVEENMILE	10500	G	Lightning	44.45	-114.5688889	ID	8/24/2015	[null]
8	55007	WATERFALL	4006	F	Lightning	45.1125	-114.7025	ID	8/11/2015	[null]
9	55008	FROG	4812	F	Lightning	41.39833333	-121.3916667	CA	7/30/2015	10/28/2015
10	55009	BEAR	11619	G	Lightning	41.81927778	-123.7483333	CA	7/31/2015	[null]
11	55010	COON	5679	G	Lightning	41.76361111	-123.9622222	CA	8/1/2015	[null]
12	55011	PEAK	12166	G	Lightning	41.55833333	-123.7758333	CA	8/2/2015	10/31/2015
13	55012	GROVES	6797.8	G	Lightning	40.95694444	-123.4533333	CA	7/30/2015	10/30/2015
14	55013	CADEN	6960	G	Lightning	36.23527778	-118.5555556	CA	7/19/2015	11/20/2015
15	55014	COUGAR CREEK	83523	G	Lightning	46.1425	-121.3725	WA	8/10/2015	10/17/2015

# Tableau Dashboard:

- Dashboard Analysis:

[https://public.tableau.com/app/profile/julia.behling/viz/California\\_Wildfires\\_16457586485670/FinalStory?publish=yes](https://public.tableau.com/app/profile/julia.behling/viz/California_Wildfires_16457586485670/FinalStory?publish=yes)



# The Analysis Phase:

## Machine Learning Models

- Logistic Regression Model
  - Sklearn classifier
  - Random Forest Feature Selection
- Linear Regression Model
  - Fire size vs remoteness

## Data Visualization

- Tableau
  - Wildfire Size & Comparative Factors

Wildfire Size Class  
Breakdown:

A: < 0.25 ac

B: 0.25 ac - 10.0 ac

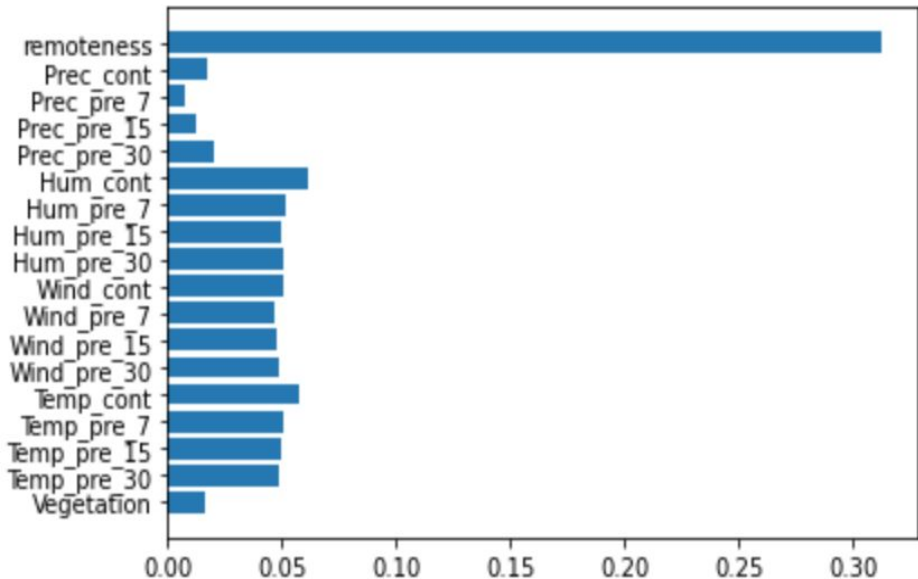
C: 10 ac - 100 ac

D: 100 ac - 300 ac

E: 300 ac - 1,000 ac

F: 1,000 ac - 5,000 ac

G: > 5,000 ac





# Conclusion & Takeaways:

- Dashboard Analysis:
  - Size Class Frequency
  - Wildfires By Year
  - Wildfire Relief Disparity
- Consistent wildfire size prediction
- Distance is the best predictor

# What We Would Have Done Differently

- More recent data (2016 - 2021)
- Analytical comparison of vegetation types
- Statistical analysis on G-class fires



# Are there any Questions?

Thank you for listening!