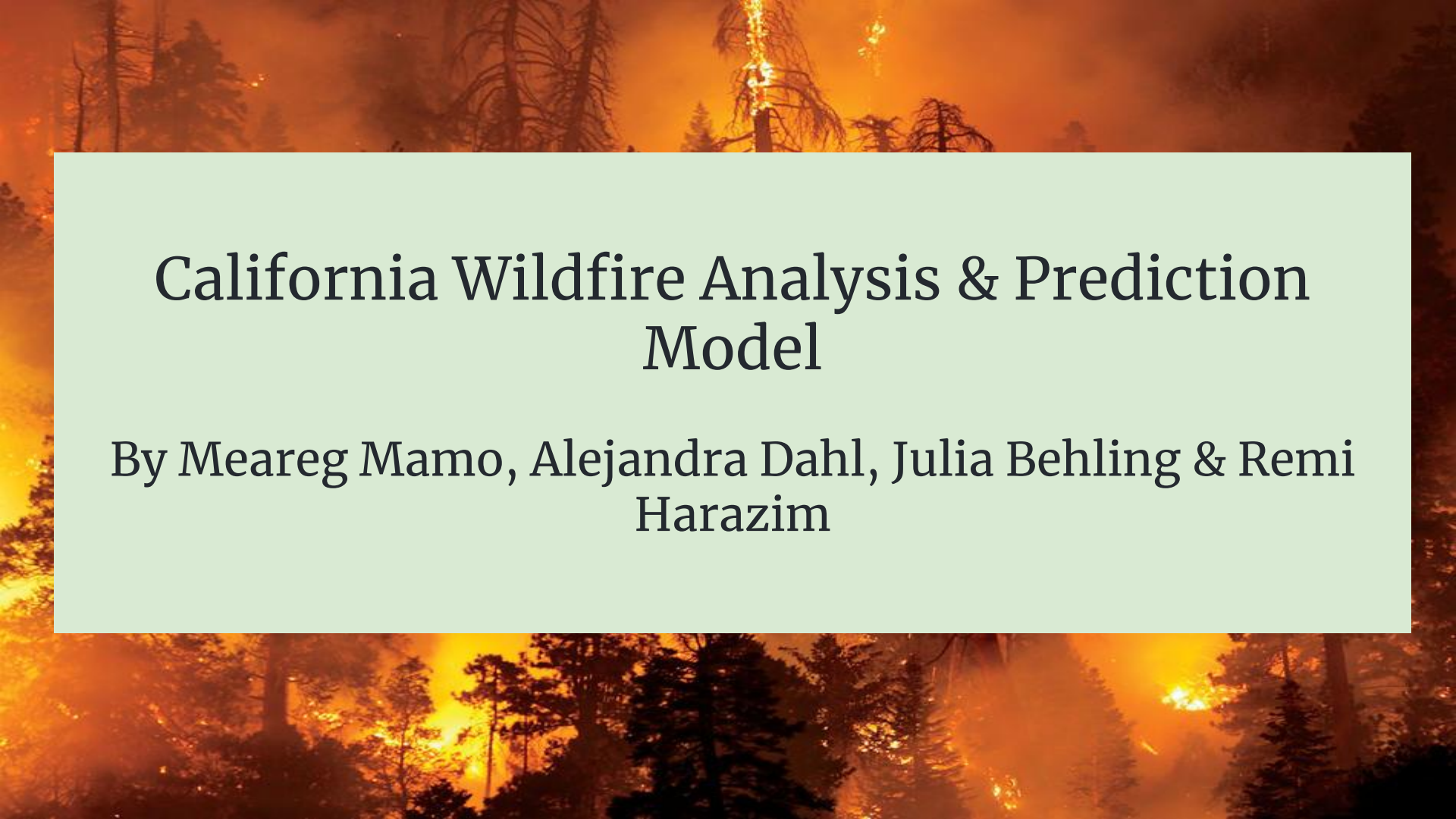
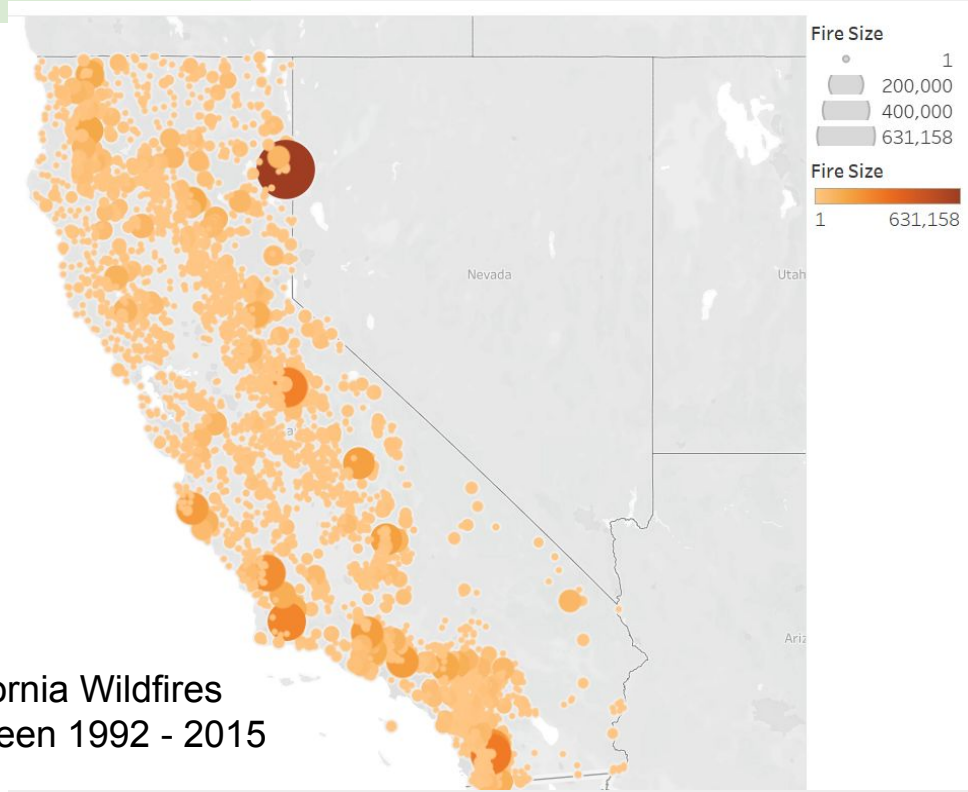


California Wildfire Analysis & Prediction Model

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/capcloudcoder/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

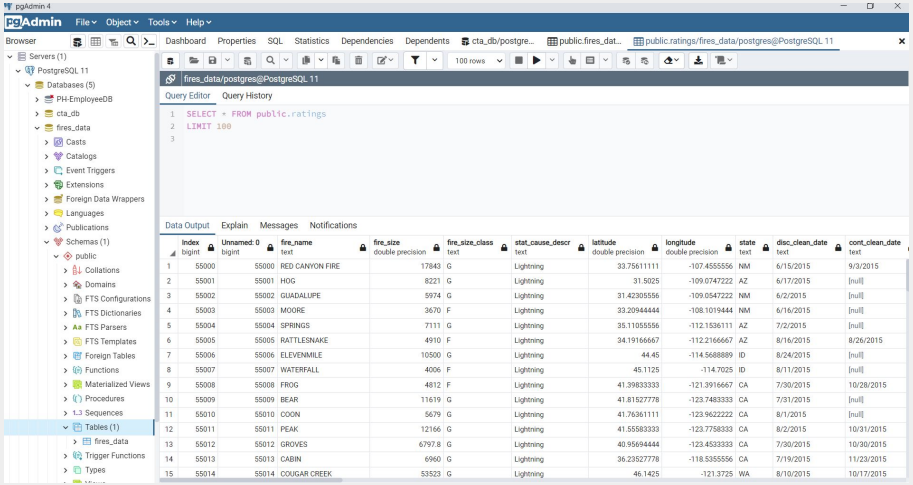


Tableau Dashboard:

- Dashboard Analysis:

https://public.tableau.com/app/profile/julia.behling/viz/California_Wildfires_16457586485670/FinalStory?publish=yes

The Analysis Phase:

Wildfire Size Class Breakdown:

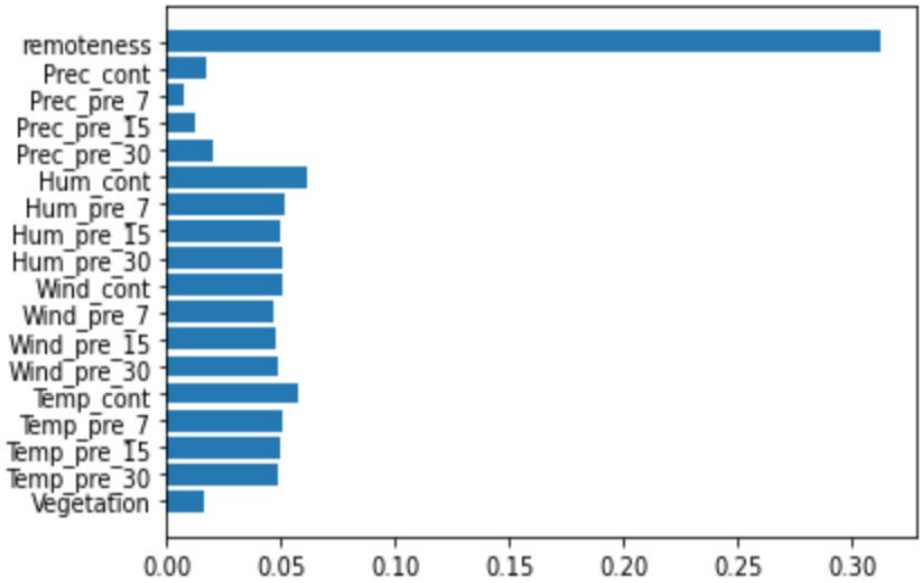
A: < 0.25 ac
B: 0.25 ac - 10.0 ac
C: 10 ac - 100 ac
D: 100 ac - 300 ac
E: 300 ac - 1,000 ac
F: 1,000 ac - 5,000 ac
G: > 5,000 ac

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



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!