



# Predicting California Wildfires

Author: Malcolm Katzenbach



# Outline



- Introduction
- Business Problem
- Method
- Data Understanding
- Exploratory Data Analysis
- Modeling
- Evaluation
- Conclusion / Future Steps

# Introduction

- 2020: Largest Wildfire Season in modern history
- 9,639 incidents of wildfires
- 4,117,855 Acres Burned
- 33 Fatalities
- 10,488 Structures

# Business Problem

- CalFire Red Flag Warnings and Fire Weather Watches
- Based on weather forecast 24-72 hours

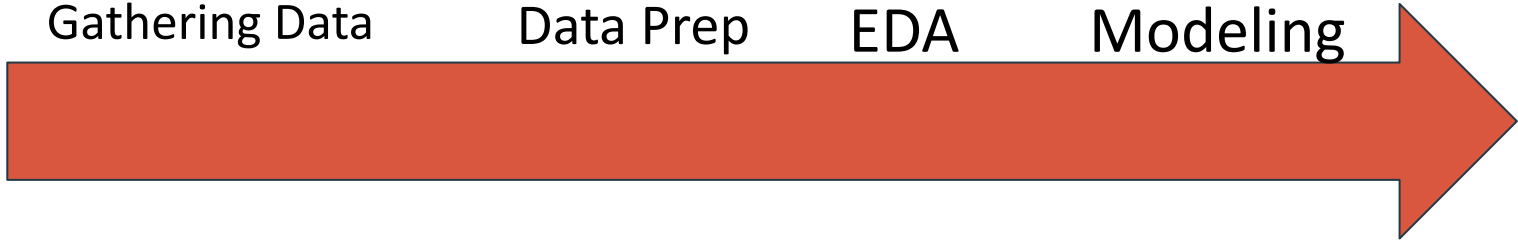
# Method

Gathering Data

Data Prep

EDA

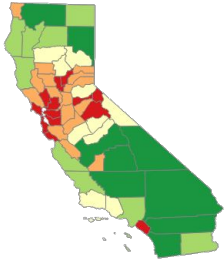
Modeling



# Data Understanding



CalFire Wildfire  
Incidents



Callands



**Anyplace**  
America.com

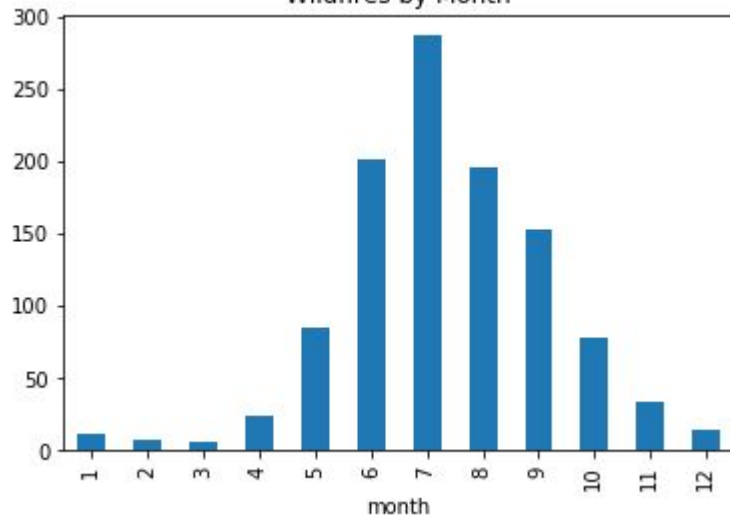
Anyplace  
America



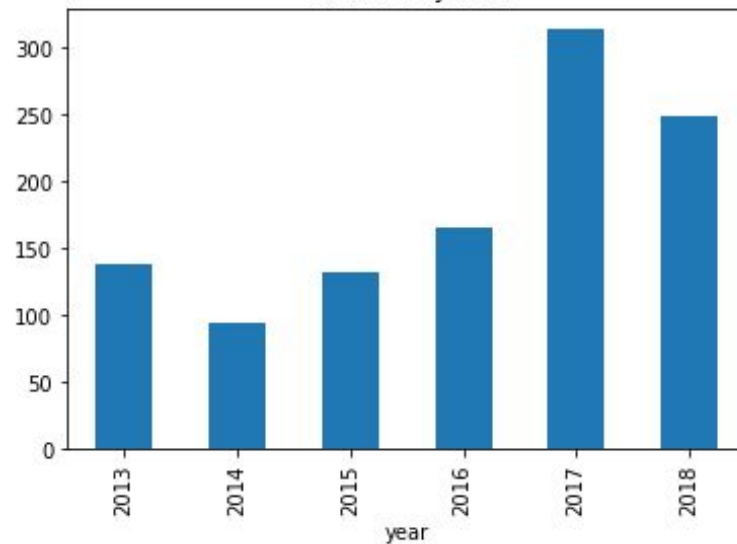
CIMIS  
Weather Underground

# EDA

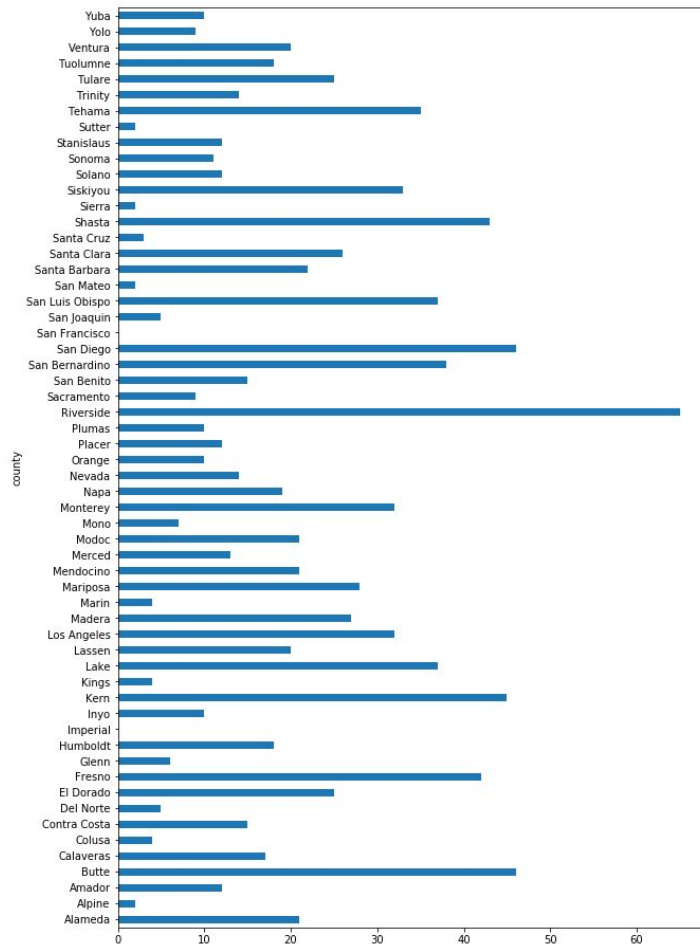
Wildfires by Month



Wildfires by Year



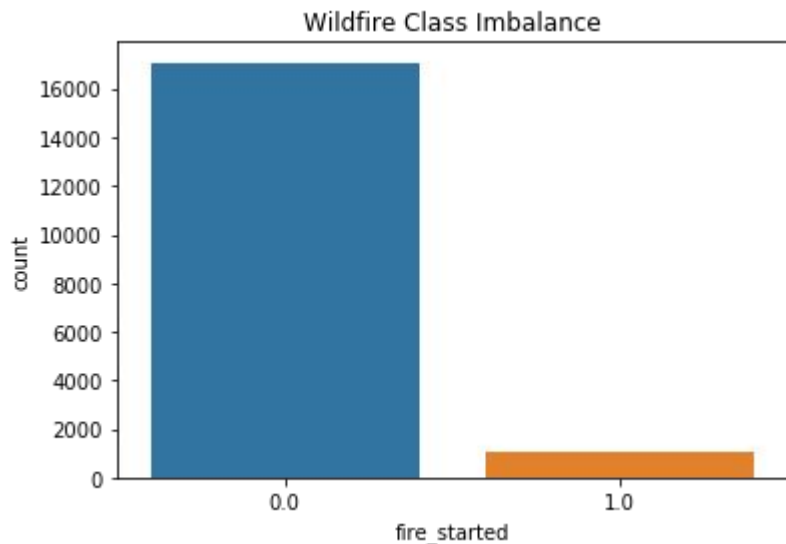
# EDA



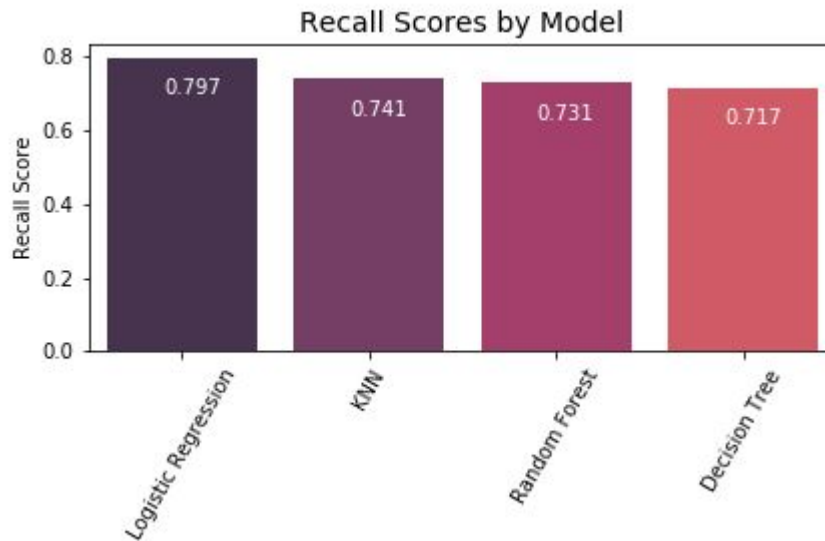
Top Counties:  
Riverside, San Diego, Butte,  
Shasta, San Luis Obispo



# Modeling



UpSampling



Recall - Minimizing False Negatives

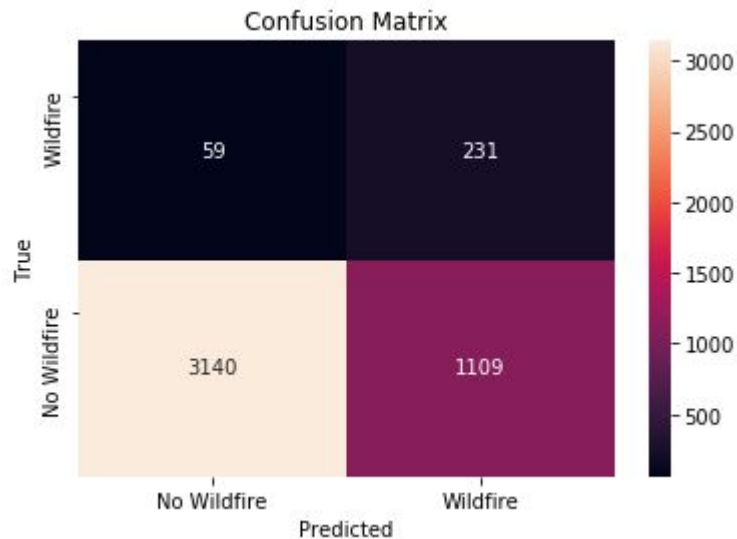
# Evaluation

Top Model: Logistic Regression

Recall: 79.7%

Accuracy: 74.2%

Error: False Positives



# Conclusion and Next Steps

- Recall is low to stand alone
  - Use with CalFire Fire Warning System
- Future Steps
  - Feature Engineering
  - New Models

# Thank You

Author: Malcolm Katzenbach

GitHub: [Malcolm206](#)

Email: [mkatzenbach8@gmail.com](mailto:mkatzenbach8@gmail.com)