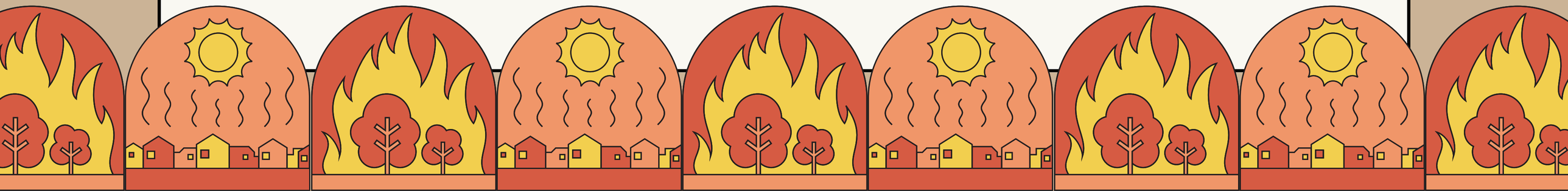
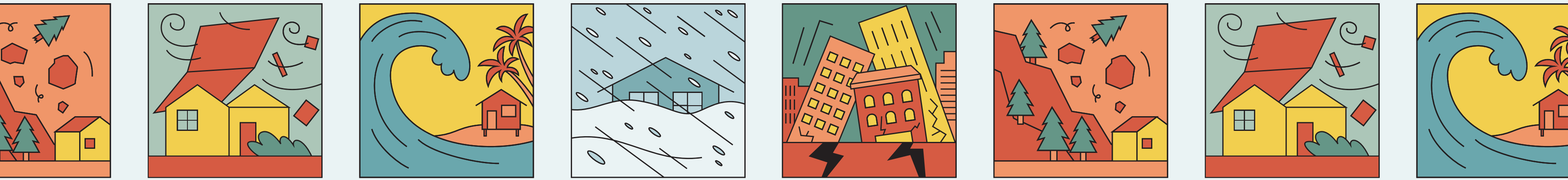
Predicting California Wildfires

Zoë Langhoff Weinstein
Environmental Data Science



Inspiration

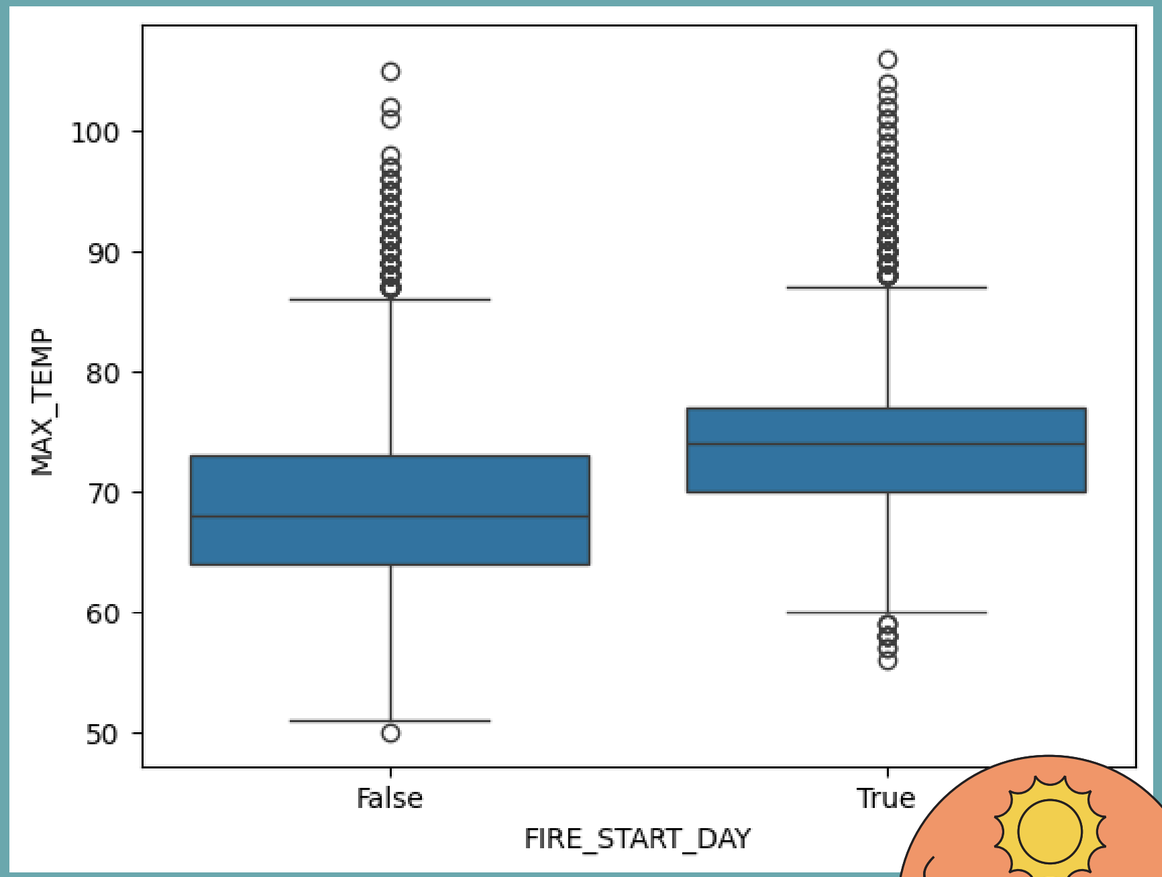
- Natural disasters can be extremely devastating especially because they often come unannounced with little time to prepare
- The goal of this research is to attempt to predict wildfires in California in light of the recent fire in the Palisades

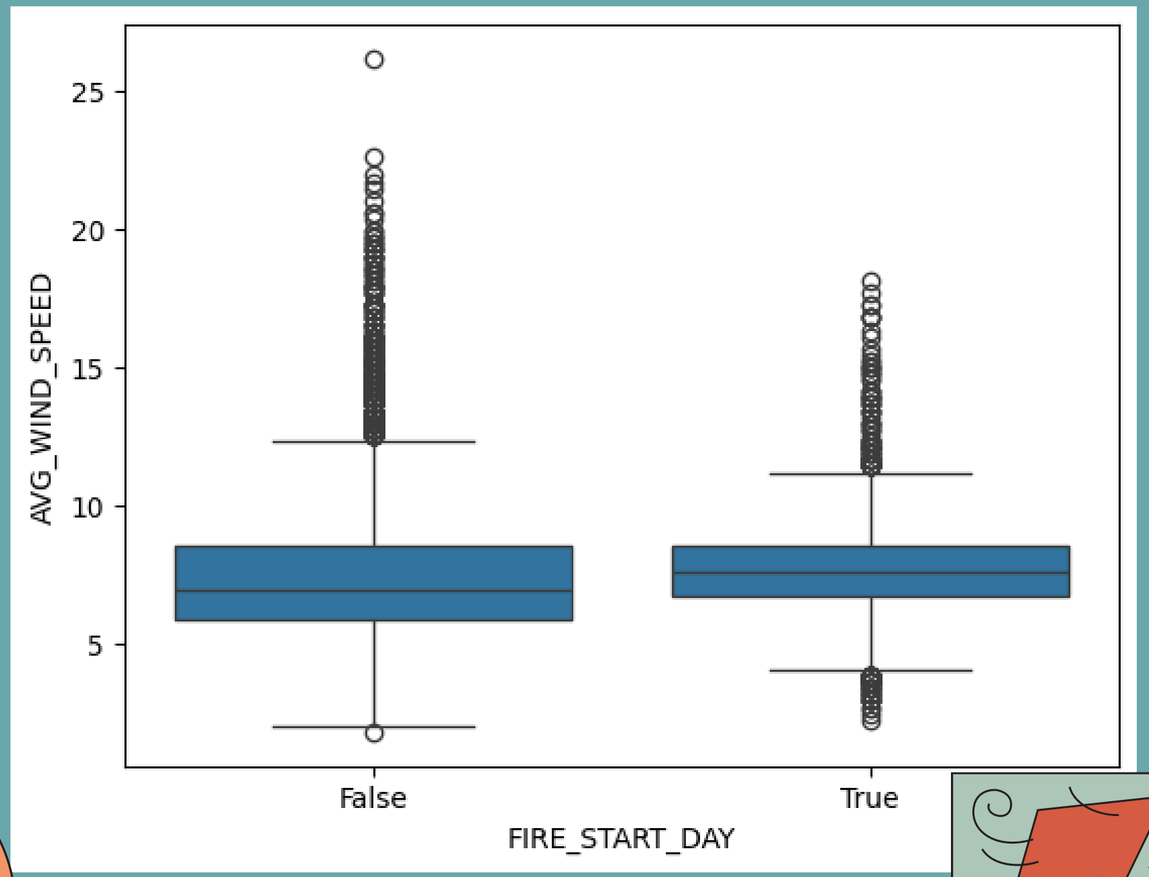


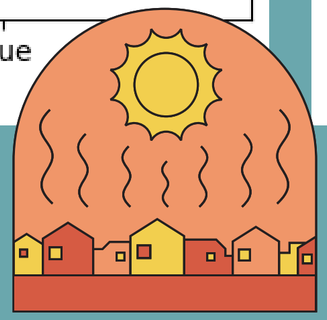
Data

- CAL Fires and NOAA Climate Data Online, from 1984-2024, just under 15,000 rows of data
- Key variables of interest:
  - Dependent variable: Binary variable for fire started
  - Independent variables: Precipitation, maximum temperature, average wind speed, temperature range, wind temperature ratio, month, season, precipitation in the last seven day, average wind speed in last seven days, day of the year (from 1-365)

Temperature and Wind







Logistic Regression

Logistic Regression Classification Report

	precision	recall	f1-score	support
False	0.72	0.87	0.78	1938
True	0.72	0.51	0.60	1344
accuracy			0.72	3282
macro avg	0.72	0.69	0.69	3282
weighted avg	0.72	0.72	0.71	3282

- Logistic regression has an easier time predicting when a fire will not happen
- Struggles with predicting correctly when a fire will happen (51% correct)

Random Forest adn XGBoost Classification

Random Forest Classification Report

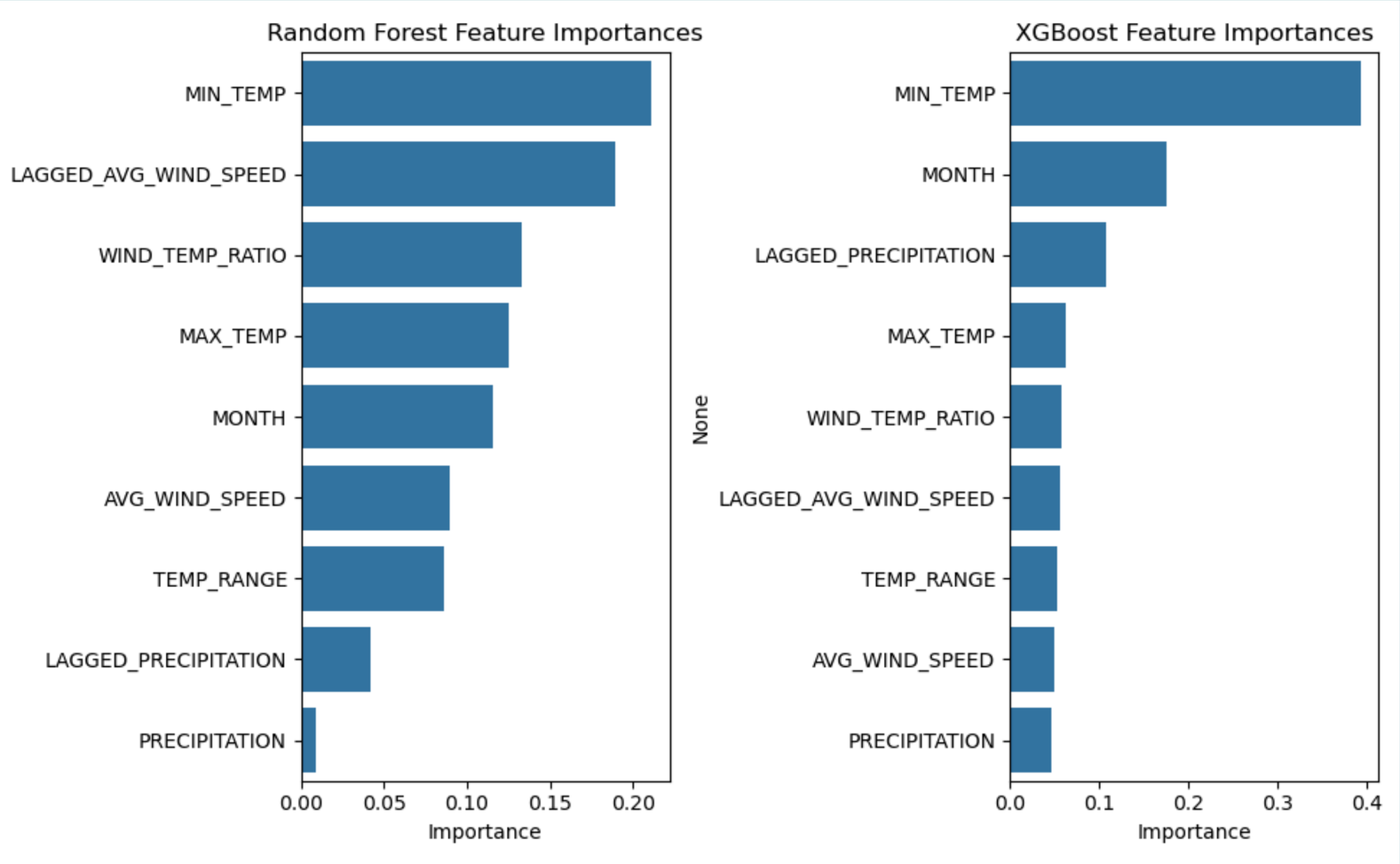
	precision	recall	f1-score	support
False	0.72	0.86	0.78	1938
True	0.72	0.51	0.60	1344
accuracy			0.72	3282
macro avg	0.72	0.69	0.69	3282
weighted avg	0.72	0.72	0.71	3282

XGBoost Classification Report

	precision	recall	f1-score	support
False	0.73	0.86	0.79	1938
True	0.72	0.53	0.61	1344
accuracy			0.72	3282
macro avg	0.72	0.70	0.70	3282
weighted avg	0.72	0.72	0.72	3282

- Random forest classification performs similarly to logistic regression
- XGBoost Model performs similarly as well, but slightly better in the True recall, meaning it correctly predicted a fire will happen more than the RF and Logit Model

Feature Importance



- Minimum temperature is most important in both models
- Precipitation is least important
- In XGBoost, lagged precipitation is third most important

Conclusion

- Predicting wildfires is challenging for numerous reasons, fires differ across space and time
- This research split data up by year and did not factor in location, further research could include longitude and latitude and compare three biomes of CA
- XGBoost Model performs best overall, but still only have a 53% chance of correctly predicting a fire will start on a specific day