Presentation Notes:

California Wildfire Analysis & Prediction Model

Meareg:

Intro Slide–

Hello and welcome to our presentation on California Wildfire Analysis. (Introduce ourselves)

meareg

Now we are going to talk about why we selected this topic.

Tools and Technology Slide– we use for this project

Here are the different tools and technology that we used for our project. We used Microsoft Excel for the initial data exploration. We used Python and Jupyter Notebook to write our code.

For our machine learning model and data cleaning ,database load procedure

We loaded the data into Pandas dataframes for analysis, and we used Scikit-Learn for our machine learning models. We chose to host our data in Postres SQL and we imported it using SQL Alchemy. All of our work was pushed to github so we could all access and collaborate with each other. And lastly, we created our dashboard using the public version of Tableau.

Remi:

Data Analysis Phase Slide–

The data analysis phase of our project consisted of three machine learning models and the assembly of graphs and visualizations in tableau. Our first machine learning model was a logistic regression using an sklearn classifier. We trained the model using all of the weather readings, ecological factors and locational data for about 1800 California wildfires and had it predict the eventual fire size class. Then it would compare the prediction to its actual size for

accuracy. With this model we reached a 75% accuracy score in predicting all six fire size classes, but when we narrowed it down to just G-class fires, we reached 90% accuracy. The hope with this model is that in the future it could be adapted to take in current data and weather readings and predict the risk and severity of a potential wildfire. That way, adequate resources could be deployed in time. This model confirmed that, to some extent, fire size can be predicted using only these factors. So, next we wanted to know which factor was the best predictor. So we fed the same data into a Random Forest Feature Selection model and had it output each factors' weight on determining fire size. This is represented on the graph to the right. As it turns out, remoteness was the biggest factor in predicting wildfire size at 30% weight. Humidity was the next best predictor, and surprisingly precipitation was the least useful. So next we wanted to know if there was a strong, linear correlation between remoteness and fire size, so we put those two factors into a linear regression model. This ended up not being the case. There were too many outliers in the data like large fires that were close to cities and smaller fires that were farther away

Alejandra:

Julia:

Why We picked this topic:

- **Effect on Human Health and welfare:**
    - Our data specifically worked with fire information from 1992 – 2015.
        - Total quantity of 3,844 documented fires in that timespan. Since then there has been an exponential increase in documented wildfires.
        - In 2021, there were 8,619 documented fires with a total burned acreage of over 2,500,000.
    - This is incredibly dangerous for the millions of people who call California home
        - In recent years: property loss, personal injury, and loss of life have become a concerning risk for individuals all across the western united states.
    - Being able to predict fire size and severity would be incredibly beneficial to those living with such uncertain conditions. Prediction models such as this can be utilized to determine risk levels and create preventative action plans, evacuation methods, etc before the disaster can strike.

Dashboard:

Total Californian Wildfires.

- This map is based on the location and size of each documented wildfire between 1992 - 2015.
    - Size is in hectares which for reference is 1 hectare = 2.5 acres.
- You can see where the sierra nevada range cuts across and the biome changes

Wildfire relief stations per county

- Southern california has much more than northern
    - Traditionally, this area is more arid and has a higher population density. AKA more frequency in fires.
    - However in recent years the north has had higher-severity fires in both size and intensity. Being a mixed conifer forest, there is way more combustible material compared to southern counterparts.

Wildfires by year:

- This data is not showing significant increase.
    - Want to point out that fire is a healthy and natural factor in an environment. It acts as a way to refresh a landscape, re-add nutrients, and encourage new growth.

Wildfire by size class

- Size class B is most common which is between 0.25 - 10 acres of affected area. G is however closely behind.
- Biomaterial build-up, overstocking, and environmental factors, and fire suppression techniques cause such a build up of combustible material that when it does occur, it is significantly harder to put out.

Conclusions and takeaways:

3 primary takeaways from the dashboard:

Size Class Frequency:

· The data indicates a high rate of Frequency in class B size fires with a runner up being Class G fires.

· Most likely due to climatic factors and customary fire suppression techniques.

Total Wildfires:

· Our data is somewhat outdated and does not predict the current quantity wildfires total average every year.

Wildfires per county:

· More wildfire relief stations clustered throughout the southern region – a more traditionally arid and dense region.

· However there are more large-scale fires occurring throughout the northern area due to heat-stress, overstocking, excess woody debri, and ecological factors, current resources cannot keep up.