DATA ANALYTICS ON US WILDFIRES

An Analytics Project Report
Presented to
The Academic Faculty

by

Raghuraj Adhiyarath
Shunmuga Sundari Raamakrishnan
Anushri Pendiala Balasubramaniam

In Partial Fulfillment
of the Requirements for the course
MGT6203 in the
Online Master of Science in Analytics

Georgia Institute of Technology
November 2022

TABLE OF CONTENTS

OVERVIEW

Project Context and Research Questions

The number and impact of wildfires are increasing on a yearly basis. These have large impacts on the people, economy, business, and public policy. Being from California, our team had a common interest in analyzing the California wildfire data and its impact on environment data such as Air Quality Index. This led to our shortlisting of wildfire data and analyzing the impact primarily for California.

We identified data from different public agencies and used linear regression models to find correlations of wildfire and other factors like Air Quality Index, seasons, and environmental factors.

The project plans to find correlations between wildfires and either of air quality, environmental factors and seasons using the research questions:

1. RQ1 – Correlation of wildfire, traffic volume and Air Quality Index at county level
2. RQ2 – Correlation of Seasons and wildfires
3. RQ3 – Correlation of environmental factors and wildfires

DATA OVERVIEW

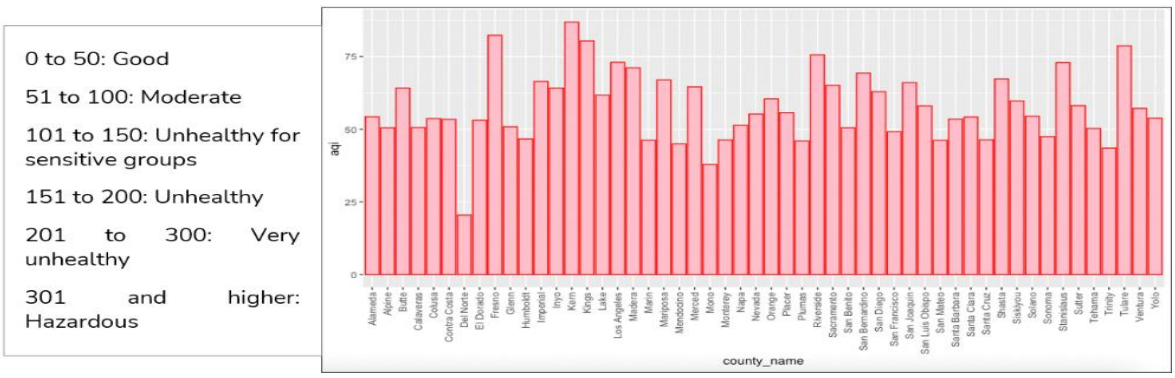
Datasets Used

Dataset details	Used in	Source	Format	Link
Wildfire dataset	All	Forest Service U.S. Department of Agriculture	SQLITE	[1]
Air Quality Data - Daily AQI data by county	RQ1	United States Environmental Protection Agency	CSV	[2]
County level Traffic volume information	RQ1	California Department of Transportation	CSV	[3]
Calendar with Seasons	RQ2	Kaggle	JSON	[4]
Environmental Data	RQ3	California Irrigation Management Information System	CSV	[5]
Downloaded datasets are at MGT6203_Team017	All	One Drive	All	[6]

Key variables and definitions

RQ1 - FIRE_DATE, COUNTY_NAME and PM2.5

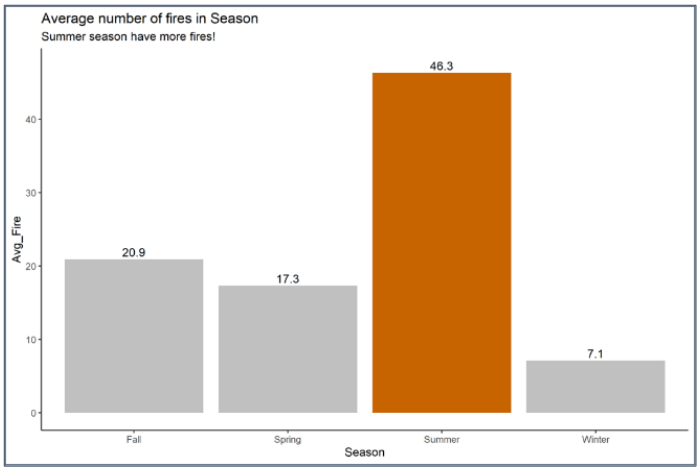
The wildfire dataset had 70,000 rows ranging from 2010 to 2018 with around 200 observations. The environment dataset had 57,718 rows with around 10 observations. As we had to merge datasets from 2 different sources, we merged rows with overlapping years in the 2010-2018 range using FIRE_DATE and COUNTY_NAME columns.

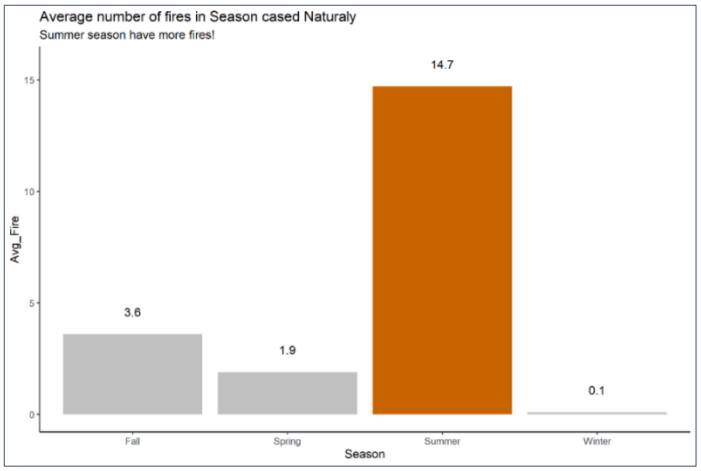


The dataset was filtered to only include parameter “PM2.5” which indicates particulate matter which is 2.5 microns in size that is in the air. Scientific studies have linked increases in daily PM2.5 exposure with increased respiratory and cardiovascular hospital admissions, emergency department visits and deaths. The visualization provide insight into the dataset – level of AQI by county.

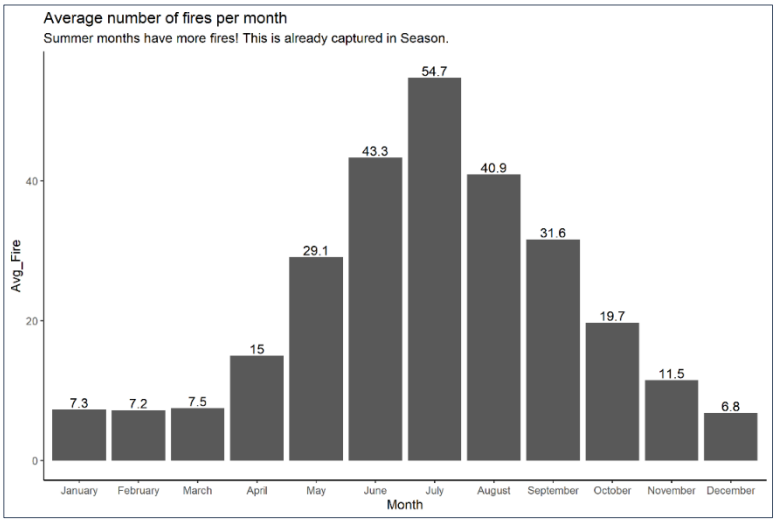
RQ2 – Date, FIRE_COUNT, Summer, and Month

Summer seasons seem to have more fires than other seasons. This trend holds true for naturally caused fires as well.





Moreover, even the adjacent months of Summer (May and September) have more fires than other months.



RQ3 - FIRE_SIZE, Evapotranspiration (ETo) and Precipitation

This model uses FIRE_SIZE with key environmental factors like Evapotranspiration (ETo)[7], Precipitation[7] and many more.

Data cleansing

RQ1 - The datasets were validated for null values and missing values for key variables such as fire size, AQI, county population and traffic volume. Missing county population data for about 11 counties were treated using the mean of the population data for that particular year.

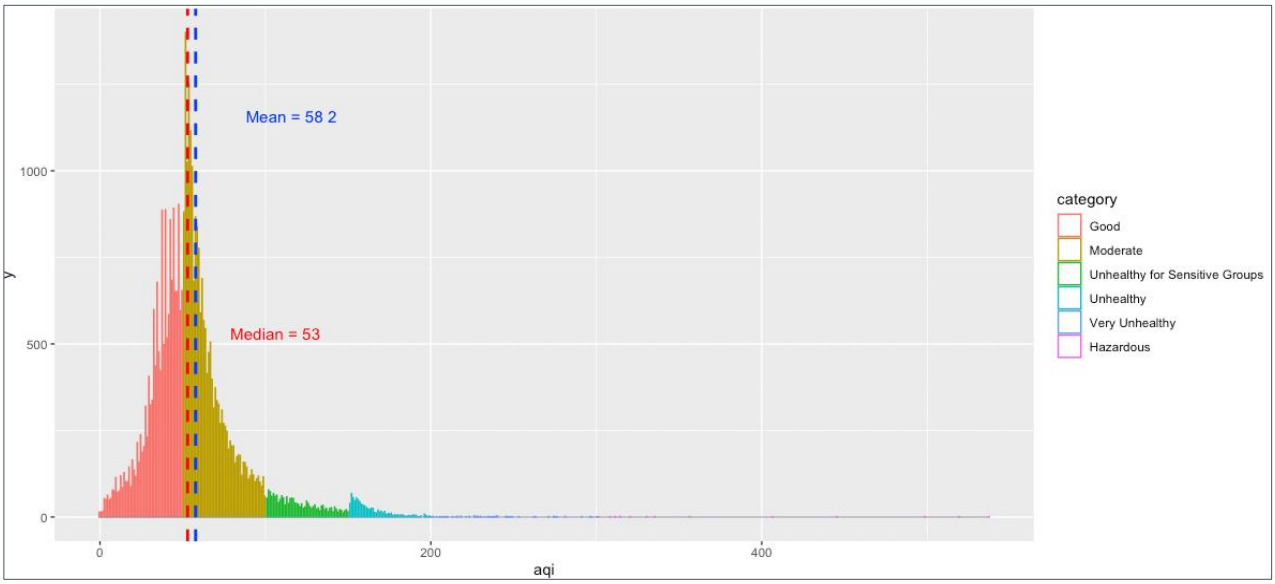
RQ3 – We used knn classification models for county data imputation. This was possible since we already have longitude and latitude for all the records.

Feature Engineering

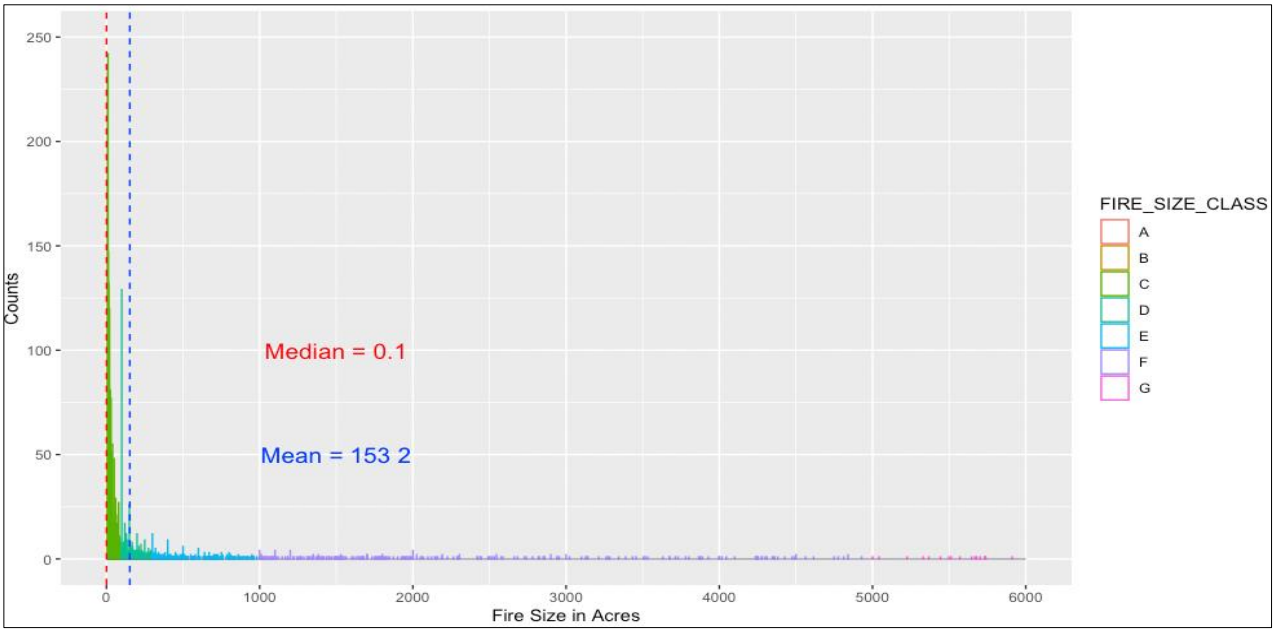
A new variable was introduced 'IsSummer' to denote if the fire was during summer. New flags were also introduced to indicate if the fire was during one of the summer months. We also added weekdays and weekend information before merging with Wildfire data.

Skewness of AQI

The skewness of the AQI data was measured. The histogram plot shows that the AQI data was slightly right skewed as the mean (58.2) is > median (53).



Skewness of Fire Size – Mean (153.2) > Median (0.1) is heavily skewed to the right. Hence, it's best to do a log or sqrt transformation of this variable.



#A=greater than 0 but less than or equal to 0.25 acres, #B=0.26-9.9 acres, #C=10.0-99.9 acres, #D=100-299 acres, #E=300 to 999 acres, #F=1000 to 4999 acres, #G=5000+ acres

MODELING

Model 1 – Wildfire, population, traffic, and Air Quality Index

Our approach is to use a linear regression model to regress variable AQI on Fire Size input variable.

- Null hypothesis – There is no association between fire size (in acres), population count, traffic count and Air Quality Index
- Alternate hypothesis – There is an association between fire size (in acres), population count, traffic count and Air Quality Index

Linear Regression model, $\mathbf{y = b_0 + b_1 \log(x) + e}$

y	=	AQI value
b0	=	intercept for the base value of the AQI with no wildfire
b1	=	co-efficient of the fire-size in acres
x	=	Fire-size in acres burnt
e	=	error term

Model Interpretation

Every 1% increase in X increases the AQI value by 0.01 units. If b1 is not equal to 0, then null hypothesis can be rejected. Alternatively, if b1 = 0, then null hypothesis is true which implies there's no relationship between FIRE SIZE and AQI value.

The coefficient of the model R2 is important to measure the quality of the model. R2 is the proportion of the variance in the dependent variable (AQI) that can be determined from the independent variable (FIRE SIZE). R2 can range from 0 to 1, where 0 indicates that the dependent variable cannot be predicted from the independent variable while 1 indicates that the dependent variable can be predicted without any error.

The residual error term e is measured as a difference between the actual value and predicted value. Sum and mean of the residual error in a regression model = 0. The initial single variable model has the below output for AQI value based on the data with brushfire.

```
> summary(lm_model_fire)

Call:
lm(formula = aqi ~ log(FIRE_SIZE), data = fire_counties_aqi_fire)

Residuals:
   Min     1Q Median     3Q    Max
-90.94 -34.37 -11.40  25.34 533.47

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    71.9062     1.9454  36.962  < 2e-16 ***
log(FIRE_SIZE)  2.2785     0.4115   5.537 3.31e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 45.19 on 3503 degrees of freedom
Multiple R-squared:  0.008675,  Adjusted R-squared:  0.008392
F-statistic: 30.65 on 1 and 3503 DF,  p-value: 3.31e-08
```

As you can see the R2 of the model is very low (.008675) which reveals the model can be improved with respect to independent variables used in the model. The intercept is 71 which indicates the base value of the AQI with Fire Size as 0. <u>The coefficient of FIRE SIZE indicates for every 1% increase in the Fire size acres, the AQI increases by 0.0227 unit.</u>

Addition of traffic volume as an input variable to the model

This model is a multi-variate model that includes both Fire Size and Traffic volume to regress the AQI. Multivariate regression is a method to measure the correlation of more than 1 independent variable on 1 or more dependent variable. The traffic volume is daily volume per county level average data. The data was cleansed, treated for missing values with mean value for the county or the year depending on the missing values of the data. A second regression model was generated to regress aqi based on fire size and traffic volume as the input variables. As shown below, the response of the model has improved with adjusted R2 as 0.1428. The model is still not a great model overall since the value is pretty low.

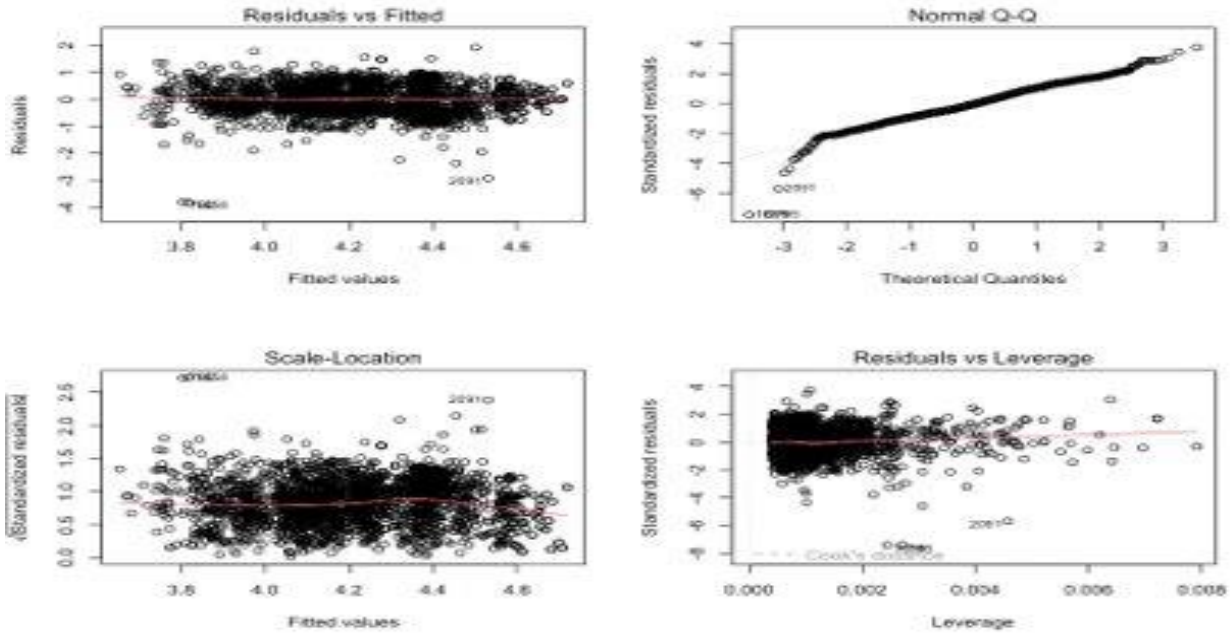
```
> summary(new_model)

Call:
lm(formula = log(aqi) ~ log(FIRE_SIZE) + log(tot_back_aadt),
    data = fire_aqi_traffic)

Residuals:
    Min      1Q  Median      3Q     Max
-3.8171 -0.3328 -0.0277  0.3676  1.9245

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       2.302487   0.095168   24.19  < 2e-16 ***
log(FIRE_SIZE)    0.019281   0.005239    3.68 0.000238 ***
log(tot_back_aadt) 0.119205   0.005936   20.08  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5155 on 2488 degrees of freedom
Multiple R-squared:  0.1435,    Adjusted R-squared:  0.1428
F-statistic: 208.4 on 2 and 2488 DF,  p-value: < 2.2e-16
```

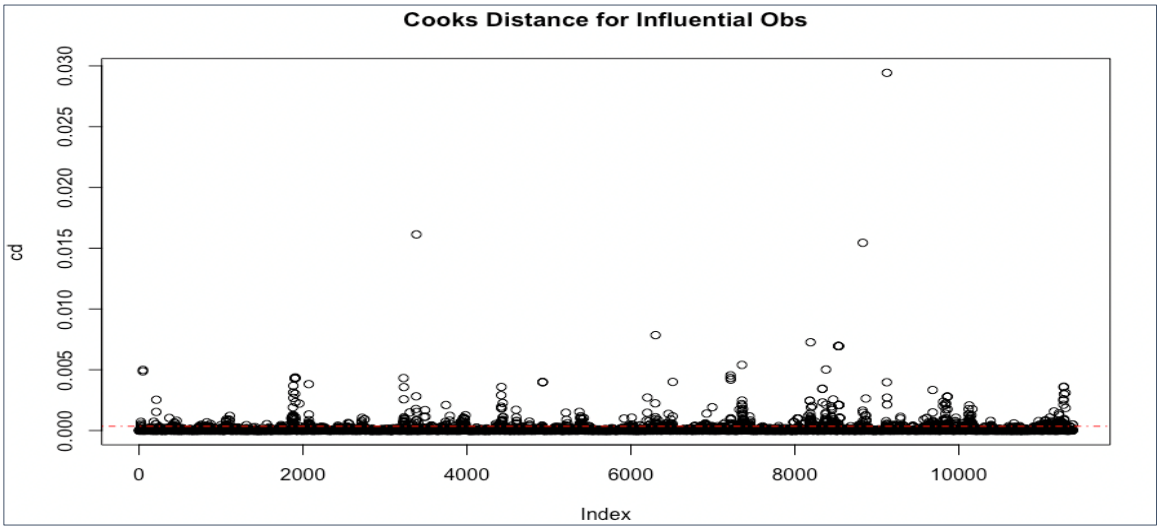


Analysis of the model plots

The 4 plots above show how the variances (residuals) are distributed to confirm the assumptions of a linear model. The Residual Vs Fitted plot shows how the variances are uniformly distributed and are closer to the middle line. The plot indicates the input and output variables show a linear pattern. The Residuals Vs Leverage plot shows any highly influential point which we do not see in the graph. There are no points closer to the cook's distance line and so there are no highly influential points. The Scale-Location plot is used to check the homoscedasticity of the variances (equal or constant) as one of the assumptions of the linear regression. If the red line is roughly horizontal across the plot then the assumption is met. The normal QQ plot shows the variances exhibit a normal distribution. If the points of the plot roughly follow along the line of the plot then this assumption is met as well. Although there are a few points towards the end beginning that deviate from the line, it's not enough points to invalidate the assumption.

Outlier treatment

There are several techniques that are used to identify the highly influential data points. Cook's Distance is primarily used to identify influential data points. It's an estimate of the influence of a data point in the model. It takes into account both residuals and leverage of a point. The general rule is to remove any points that has a cooks distance > 4/n where n is the number of data points or remove points that are 3 times more than the mean of all distances. For our model, we have used the former method as you can see in the below graph the number of data points that lies above the 4/n with a red dashed line.



Once the outliers are removed, we rerun the regression to produce the following output. The Adjusted R2 has increased to 0.1212. The Residual Vs Leverage plot shows that there are no outliers closer to cook's distance.

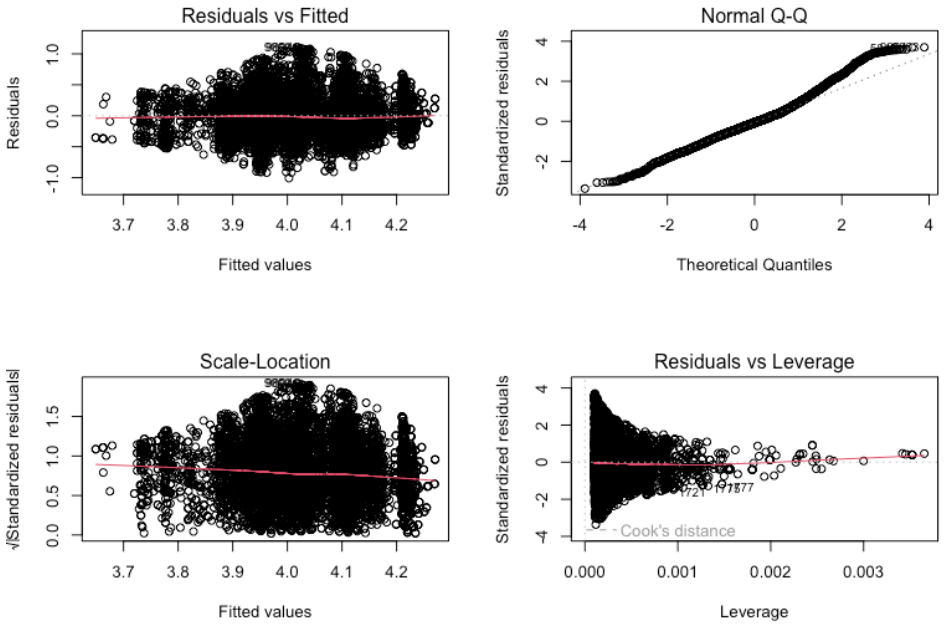
```
> summary(new_model_c)

Call:
lm(formula = log(aqi) ~ log(FIRE_SIZE) + log(tot_ahead_aadt),
    data = new_no_outlier)

Residuals:
     Min       1Q   Median       3Q      Max
-1.00711 -0.18775 -0.02046  0.15692  1.10735

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.052996   0.027158 112.416  < 2e-16 ***
log(FIRE_SIZE)      0.006026   0.001437   4.195 2.76e-05 ***
log(tot_ahead_aadt) 0.063283   0.001709  37.038  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2996 on 9965 degrees of freedom
Multiple R-squared:  0.1214,    Adjusted R-squared:  0.1212
F-statistic: 688.5 on 2 and 9965 DF,  p-value: < 2.2e-16
```



Discarded models

We developed models regressing AQI with fire size, traffic and population as well. But since the model did not show any improvement, the model was discarded from the results.

Insights for business

1. Due to the correlation of Fire, Traffic volume to AQI, we would recommend the state and county to take additional steps towards controlling the air pollution during wildfires and at other times. The county can be proactive in controlling the air pollution by adding additives to vehicle fuels.
2. The hospitals in certain counties where the AQI is high can take additional measures to provide masks. They can educate high-risk patients with respiratory illness to wear masks, use of air purifiers inside the houses/buildings.
3. State, County can also encourage use of electric vehicles to prevent pollution. They can provide better tax break rates for counties where the AQI is high. The traffic can be controlled closer to areas where the hospitals are built.

Model 2 - Seasons and wildfire

Data Preparation

Seasonal Data

The Kaggle source file in JSON with season information was converted to flat format. During this step, included month, weekdays, and weekend flag to the calendar file for 2010 to 2018.

Fire Data

Extracted the number of fires per day in California along with cause of fire. This was then merged with calendar file with seasons for 2010 to 2018.

Linear Regression model

Model using an indicator variable for Summer (isSummer)

Our approach is to use a linear regression model to regress an indicator variable (IsSummer) as input variable and total fire count as dependent variable. Summer season is selected based on data exploration plot.

- Null hypothesis – There is no association between the summer season and number of fires.
- Alternate hypothesis – There is an association between the summer season and total number of fires.

Linear Regression model, $\mathbf{y = b_0 + b_1 * x + e}$

y	=	Average number of fires
b0	=	intercept for the base value of fires in other seasons
b1	=	co-efficient of the summer indicator variable
x	=	indicator variable for summer
e	=	error term

```
Call:
lm(formula = Total_Fires ~ isSummer, data = Calendar_with_Fire)

Residuals:
    Min      1Q  Median      3Q     Max
-36.338 -10.304  -3.338   6.696 138.662

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.3041     0.3023   50.62   <2e-16 ***
isSummer     31.0340     0.5967   52.01   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.6 on 3582 degrees of freedom
Multiple R-squared:  0.4303,    Adjusted R-squared:  0.4301
F-statistic:  2705 on 1 and 3582 DF,  p-value: < 2.2e-16
```

Model using months to validate finding on Summer

For validating the finding of summer indicator variable, we used a linear regression model to regress Month as input variable and total fire count as dependent variable. This model is created for supporting RQ2 on impact of Summer.

- Null hypothesis – There is no association between the summer months and number of fires.
- Alternate hypothesis – There is an association between the summer months and total number of fires.

Linear Regression model, $\mathbf{y = b_0 + b_1 * x + e}$

y	=	Average number of fires
b0	=	intercept for the base value of fires in January
b1	=	co-efficient of the Month variable
x	=	Month variable
e	=	error term

By looking at individual months, we can see that the model holds true for May, June, July, August, September, and October.

```
Call:
lm(formula = Total_Fires ~ Month, data = Calendar_with_Fire)

Residuals:
    Min      1Q  Median      3Q     Max
-34.694  -6.222  -2.222   4.710 144.068

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     7.28374    0.77111   9.446  < 2e-16 ***
MonthFebruary  -0.06151    1.10954  -0.055  0.95579
MonthMarch      0.20160    1.07441   0.188  0.85117
MonthApril      7.75653    1.08225   7.167 9.27e-13 ***
MonthMay       21.82594    1.07189  20.362  < 2e-16 ***
MonthJune      36.00626    1.08048  33.324  < 2e-16 ***
MonthJuly      47.40981    1.07189  44.230  < 2e-16 ***
MonthAugust    33.64852    1.07189  31.392  < 2e-16 ***
MonthSeptember 24.33626    1.08048  22.524  < 2e-16 ***
MonthOctober   12.40013    1.07189  11.568  < 2e-16 ***
MonthNovember   4.17081    1.08315   3.851  0.00012 ***
MonthDecember  -0.49575    1.09628  -0.452  0.65114
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.11 on 3572 degrees of freedom
Multiple R-squared:  0.599,     Adjusted R-squared:  0.5978
F-statistic: 485.1 on 11 and 3572 DF,  p-value: < 2.2e-16
```

Discarded models

For seasonal, we created linear regression models using the impact of seasons on fires caused by humans and naturally occurring fires. These models were discarded as Total Fires was a better model over 10 years of data.

We also used linear regression models for the count of fires against weekday and weekends. But these models were discarded due to poor adjusted R-squared value.

Model interpretation

During summer (IsSummer=1), the total number of fires increases by ~33.5. Since b1 is not equal to 0, then null hypothesis can be rejected. This implies there an association between summer season and the number of fires.

During summer months and summer adjacent months, the total number of fires increases by double digits with a high p-value. The adjusted R-squared value is good at 0.5978. Since the coefficients are not equal to 0, then null hypothesis can be rejected. This implies there an association between summer months and summer adjacent months and the number of fires.

Insights for business

1. Due to the correlation of Season with number of fires, we would recommend an increased investment in the fire departments in the form of seasonal workforce for Summer. Based on budget availability, the adjacent months of Summer (May and September) also should be considered for seasonal workers.

2. Increased sign boards in Summer to improve human behavior so that human caused fires can be reduced.
3. Order medical supplies (or provide alerts to healthcare providers to order supplies) to treat patients with respiratory illnesses during the summer season.

Model 3 – Environmental factors and wildfire

Data Preparation

Environmental Data (2010 to 2018)

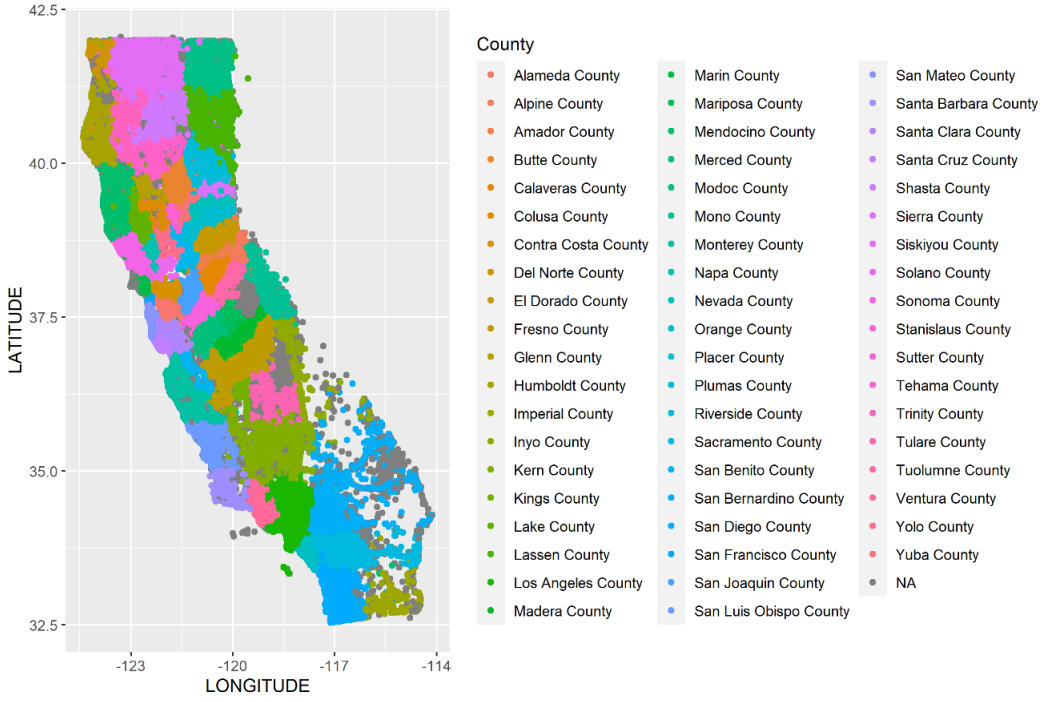
We downloaded 9 years data over 40 different files and merged it together. We noted that some counties had multiple weather stations. So, we used only one station per county to remove bias in the model.

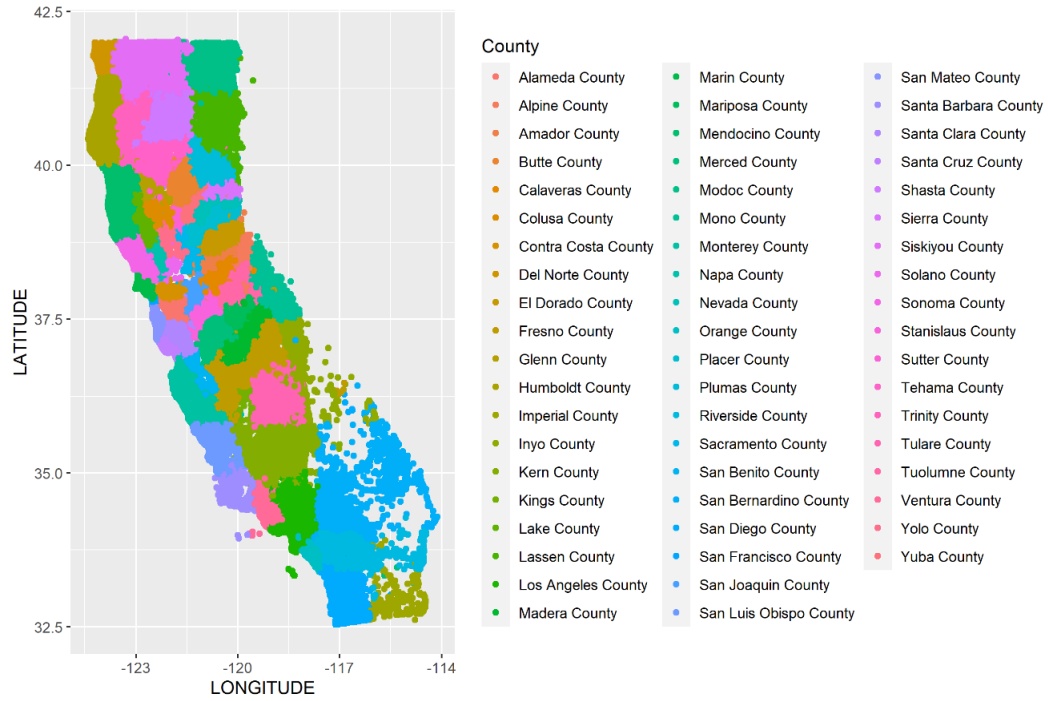
Key Factor	Definition	Link
Evapotranspiration (ET)	A term used to describe the water consumed by plants over a period. ETo is evapotranspiration from standardized grass (ETo) surfaces.	[7][8][9]
Precipitation	Rainfall is measured using tipping bucket rain gauges	[7]
Solar Radiation	The total incoming solar radiation is measured using pyranometers at a height of 2.0 meters above the ground	[7]
Vapor Pressure	The vapor pressure of the atmosphere is the partial pressure exerted by atmospheric water vapor	[7]
Relative Humidity	Relative humidity is the ratio of the actual amount of water vapor in the atmosphere to the amount the atmosphere can potentially hold at a given air temperature	[7]
Dew Point Temperature	Dew point temperature is the temperature to which the atmosphere must be cooled, at constant pressure and water vapor content, in order to reach saturation	[7]
Soil Temperature	Soil temperature is measured at 15 centimeters (6 inches) below the soil surface	[7]

Fire Data (2010 to 2018)

This model requires county information to be accurate so that environmental conditions can be mapped correctly with the fire dataset. However, 39.39% of data had county name as NA. But the datafile had the latitude and longitude data for all records. We used knn classification model (k=3) to map NA values to county names. We used the available county name to latitude and longitude mapping for the training dataset.

The below scatter plots show the dataset before and after county names corrections.





For the environmental data, some counties had multiple weather stations. So, we used only one station per county to remove bias in the model.

Extracted the number of fires per day in California along with cause of fire. This was then merged with calendar file with seasons for 2010 to 2018.

Linear Regression model

Our approach is to use a linear regression model to regress key environmental factors (Evapotranspiration (ETO), Precipitation, Solar Radiation, Avg Vapor Pressure, Avg. Relative Humidity, Dew Point Temperature, Avg. Soil Temperature) as input variable and size of fire as dependent variable.

- Null hypothesis – There is no association between the environmental factors and fire size
- Alternate hypothesis – There is an association between the environmental factors and fire size

Linear Regression model, $\mathbf{y = b_0 + b_i * x_i + e}$

y	=	Size of fires
b_0	=	intercept for the base value of fire size
b_i	=	co-efficient of each of environmental variable
x_i	=	environmental variables

- Evapotranspiration (ET)
- Precipitation
- Solar Radiation
- Vapor Pressure
- Relative Humidity
- Dew Point Temperature
- Soil Temperature

e	=	error term

```
Call:
lm(formula = Fire_Size ~ ETO + Precipitation + SolarRadiation +
    Avg_VaporPressure + Avg_RelativeHumidity + DewPointTemperature +
    Avg_SoilTemperature, data = CA_Fire_Env)

Residuals:
   Min     1Q Median     3Q    Max
  -673   -109    -71    -33 281608

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          191.7384   116.4994   1.646 0.099804 .
ETO                   57.1887    17.0046   3.363 0.000771 ***
Precipitation         -3.0333     5.1655  -0.587 0.557058
SolarRadiation        -1.2333     0.3037  -4.061 4.89e-05 ***
Avg_VaporPressure     50.4975    73.7391   0.685 0.493465
Avg_RelativeHumidity  -0.8892     1.3055  -0.681 0.495816
DewPointTemperature   -3.7727     6.5870  -0.573 0.566811
Avg_SoilTemperature   -3.2099     3.9948  -0.804 0.421683
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2393 on 53992 degrees of freedom
Multiple R-squared:  0.0006396, Adjusted R-squared:  0.00051
F-statistic: 4.936 on 7 and 53992 DF,  p-value: 1.362e-05
```

Discarded models

We also used linear regression models for all the available variables. But this was discarded as most of significant factors were name of the counties. Since this does not have a direct relevance on this research question, these models were discarded

Model interpretation

For each unit increase in ETO, the fire size increases by 57.2 units and one unit increase in Solar Radiation results in a reduction of 1.2 units in fire size. However, the R-squared and adjusted R-squared values are too small to consider the impact of these variables in fire size.

We cannot reject null hypothesis and there may not be association between environmental factors and number of fires.

Insights for business

1. The analytical model using the publicly available data failed to find a correlation between the environmental factors and size of fires. Due to this, we would not recommend any changes to standard operating procedures considering the environmental factors.
2. It was observed that some counties have higher propensity towards large fires while others have a high frequency of fires. We would recommend doing an analytical study with focus on optimized operations using chaining principles. This will be helpful to manage uncertainty through scenario planning and will help to improve the resiliency of state’s preparedness to manage high-consequence events with low probability.

CONCLUSION

Overall, this project intended to find correlations between wildfires and air quality, environmental factors, and seasons.

We concluded that there is a slight positive correlation between the fire size and air quality index, however this model was not improved by the addition of traffic volume and therefore, we could not validate the correlation between traffic volume and air quality. As a result of this correlation, we recommend the state and county to take additional steps towards curbing air pollution on roads, taking proactive measures to provide healthcare education and masks as well as motivate the use of electric vehicles to minimize the occurrence and effects of wildfires. This can be carried out by district and county issued guidelines such as in the Ventura County Air Quality Assessment Guidelines where air quality mitigation measures and incentive programs are actively analyzed and encouraged.[12]

In addition to air quality, we discovered an association between the summer seasons and a high number of wildfire occurrence using historical fire data. Due to this known pattern, we would recommend an increased seasonal investment in the fire department workforce, increase in sign boards and education to the public as well as increased inventory of medical supplies directed to treat respiratory illnesses during the summer season and the months adjacent to summer.

Finally, we also investigated the role of other environmental factors such as ETO, precipitation, solar radiation, vapor pressure and humidity to the prevalence of wildfires. The extremely low R-squared values suggested that there may not be a close association between these environmental factors and wildfires, hence we were unable to recommend any changes to the standard operating procedures of the state and county.

Future Work

Throughout our research, it was observed that some counties display a higher propensity towards large fires while others have a high frequency of fires. Further steps to include moving forward would be the identification of priority counties for immediate action and the use of resilient planning or changing principles for an optimized approach to managing the uncertainty of wildfire occurrences. This approach would aid proactive tackling of larger fires and natural disasters that may occur.

REFERENCES

[1] Short, Karen C. 2021. Spatial wildfire occurrence data for the United States, 1992-2018 [FPA_FOD_20210617]. 5th Edition. Fort Collins, CO: Forest Service Research Data Archive. https://doi.org/10.2737/RDS-2013-0009.5

[2] Air Quality Index (AQI) Datasets - Daily AQI data by county (2010-2022) was downloaded individually from https://aqs.epa.gov/

[3] County level Traffic volume information was downloaded from CALTRANS site - https://dot.ca.gov/programs/traffic-operations/census

[4] Season data from year 1500 to 2700 - https://www.kaggle.com/datasets/donnetew/calendar-dates-and-seasons-15002700

[5] Environmental Data was downloaded in multiple files from https://cimis.water.ca.gov/Default.aspx

[6] The datasets are available in OneDrive at MGT6203_Team017

[7] Refer to Data Types section in CIMIS site - CIMIS (ca.gov)

- Evapotranspiration (ET) - A term used to describe the water consumed by plants over a period. ETo is evapotranspiration from standardized grass (ETo) surfaces.
- Precipitation - Rainfall is measured using tipping bucket rain gauges
- Solar Radiation - The total incoming solar radiation is measured using pyranometers at a height of 2.0 meters above the ground
- Vapor Pressure - The vapor pressure of the atmosphere is the partial pressure exerted by atmospheric water vapor
- Relative Humidity - Relative humidity is the ratio of the actual amount of water vapor in the atmosphere to the amount the atmosphere can potentially hold at a given air temperature
- Dew Point Temperature - Dew point temperature is the temperature to which the atmosphere must be cooled, at constant pressure and water vapor content, in order to reach saturation
- Soil Temperature - Soil temperature is measured at 15 centimeters (6 inches) below the soil surface

[8] Understanding Plant Water Use (Evapotranspiration) - https://coagmet.colostate.edu/extended_etr_about.php

[9] Evapotranspiration article on Wikipedia. - https://en.wikipedia.org/wiki/Evapotranspiration

[10] Inspiration of the project from https://www.kaggle.com/datasets/rtatman/188-million-us-wildfires

[11] Other Kaggle references

- https://www.kaggle.com/datasets/chelseazaloumis/cimis-dataset-with-fire-target
- https://www.kaggle.com/code/skpatel12/california-wildfire-analysis
- https://www.kaggle.com/code/balavashan/forestfir-visualisation

[12] Ventura Country Air Quality Assessment Guidelines. http://www.vcapcd.org/pubs/Planning/VCAQGuidelines.pdf

[13] California Fires Gaining in Intensity Since 2010 https://knoema.com/brpdrvg/california-fires-gaining-in-intensity-since-2010

[14] Aerosol characterization in the Southeastern U. S. using satellite data for applications to air quality and climate http://hdl.handle.net/1853/43589

[15] Interactions between climate variability and air pollution—A study of severe haze and large wildfires http://hdl.handle.net/1853/60692

[16] A comparative analysis of state emergency plans: improving response to vulnerable populations http://hdl.handle.net/1853/29774

[17] Measurements of emissions from agricultural fires and wildfires in the U.S. http://hdl.handle.net/1853/59147

[18] Evaluation of emission uncertainties and their impacts on air quality modeling: applications to biomass burning http://hdl.handle.net/1853/26689

[19] Effects of prescribed burning on air quality in the southeastern U.S. and implications for public health studies http://hdl.handle.net/1853/62661

[20] Predicting behaviors and effects of biomass burning http://hdl.handle.net/1853/54843

[21] Integration of air quality data for improved estimates of PM2.5 source impacts http://hdl.handle.net/1853/55643