**Creating A Fire Alarm Using Household IOT Sensors**

Gabi Rivera, Sahil Wadhwa, and Liam Richardson

Shiley-Marcos School of Engineering, University of San Diego

**Table of Contents**

**Abstract (Liam)**

Start here….

*Keywords:*

**Creating A Fire Alarm Using Household IOT Sensors (Liam)**

Start here…

Purpose of the Study (Liam) make sure to “clearly define the problem statement” - professor.  
Fire alarms are known to safe lives [insert reference] however there are several known issues which interfere with their efficacy. The first problem being false alarms which can make the fire department less likely to respond quickly just based off of a fire sensor fire alarm, and secondly very often people allow them to run out battery and/or intentionally remove the battery due to the high false alarm rate. Hardware engineers working on IOT sensors have noted that if fused together, the ensemble of many household sensors now present in homes which are placed in thermostats, refrigerators etc could be used as a makeshift fire alarm. This can be useful because having a second “fire alarm” can address both the problems of intermittency, as well as increasing the confidence of a reported fire being legitimate, provided that the artificial fire alarm is tuned properly to reject false positives. Using IOT data which was collected in time synchronous fashion with a known calibrated fire alarm as a fire was lit in a room, we were able to train several models to fuse data from IOT sensors to make an ensemble fire alarm.

**Method**

**General Information of Dataset**

The sensor-fusion smoke detection classification was retrieved from Kaggle under CCO Public Domain. There were two datasets associated in the package which are labeled *train\_dataset.csv* and *test\_dataset.csv*. The train dataset was decidedly used over the test dataset due to the presence of outcome variables. Having an original outcome data to compare with the binary classification train and test results is necessary to create a confusion matrix for performance evaluation. There were 5,000 observations, 14 numerical predictors, and a binary outcome that comprised the train dataset. The outcome variable is labeled *Fire.Alarm* with “Yes” and “No” classes pertaining to fire alarm triggered due to the presence of fire and the absence of fire alarm, consecutively. The other features are measurements of air temperature, air humidity, air pressure, total volatile organic compounds, carbon dioxide (CO2) equivalent concentration, raw hydrogen (H2) gas, raw ethanol, particulate matter < 1.0 µm (PM1.0), particulate matter 1.0 µm < 2.5 µm (PM2.5), sample counter, timestamp UTC seconds, concentration of particulate matter < 0.5 µm (NC0.5), concentration of particulate matter 0.5 µm < 1.0 µm (NC1.0), and concentration of particulate matter 1.0 µm < 2.5 µm (NC2.5).

**Exploratory Data Analysis**

Preliminary data exploratory analysis was used to reduce the predictors down to 8 variables. The first approach was to observe the predictors distribution through a histogram plot followed by a box plot to visualize the outliers as shown in Figures 1-2. A box plot of the fire alarm class distribution was also plotted for each predictor as shown in Figure 3. Skewness score was then calculated to quantify non-normality in the distribution of each predictor. Box cox statistical transformation was then performed to address issues in the predictor’s skewness followed by a <0.5 skewness score improvement selection. The sample counter was the only predictor that benefited from the Box-Cox transformation. Figure 4-5 shows the transformation of sample counter’s scattered distribution into a normal shaped distribution.

Most of the predictors removed were assessed through the correlation matrix analysis. The cutoff was set at 0.75 Pearson correlation coefficient score. This left the smoke alarm dataset with the hour of the day, day of the week, air temperature, total volatile organic compounds, CO2 equivalent concentration, raw hydrogen, concentration of particulate matter 0.5 µm < 1.0 µm (NC1.0), and sample counter predictors. The complete list of predictors had a maximum of 1.0 correlation score while the reduced list of smoke alarm predictors has a maximum of 0.80 correlation score. Figure 6 shows the original dataset correlation plot against the reduced dataset correlation plot.

The second exploratory data analysis step was to extract valuable information about key relationship patterns observed from the entire dataset. A pairwise relationship plot was created to show the general overview of the variable’s correlation and distribution defined by the outcome classes. As shown in Figure 7, absence of alarm (No = 0) proportionally has higher frequency count compared to the frequency of triggered alarms (Yes = 1). This means that the outcome variable is disproportionate and it will need to be rebalanced during data splitting to better represent the minority class of interest. There is also a clear separation between the distribution area of the two fire alarm classes under sample counter area plot. This indicates that this feature can be highly relevant in distinguishing between the two outcome classes and thus justifies retaining the sample counter as one of the predictors.

The proportions of the outcome classes were also explored for each of the remaining predictors as shown in Figure 8. Temperature, concentration of particulate matter 0.5 µm < 1.0 µm (NC1.0), sample counter, and day of the week are the features that show a higher proportion of triggered alarms (Yes = 1) over the absence of alarm (No = 0). The other 4 predictors are showing the opposite trend. In this case, it makes sense that triggered alarm class is highly associated with temperature and particulate matter as those are natural indicators of fire or smoke. The same with CO2’s association with the absence of alarm (No =1) as the compound is a fire suppressant. Raw hydrogen and total volatile organic compounds are both highly flammable and so their high association to absence of alarm (No = 0) is unexpected.

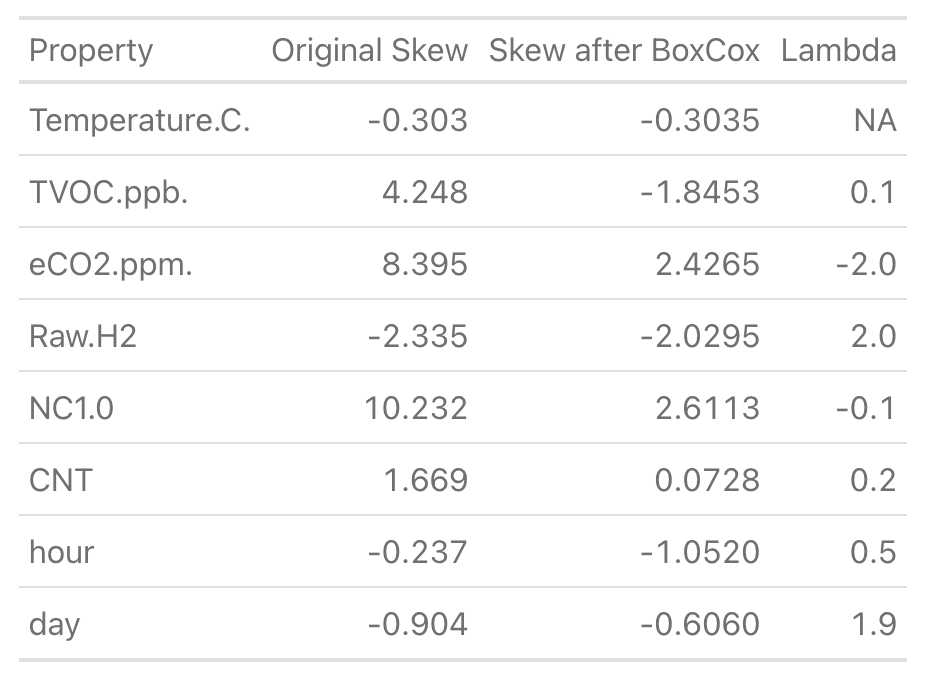
Lastly, day of the week and hour of the day are showing opposite patterns in the fire alarm class proportion. It is possible that certain days of the week are more strongly associated with triggered alarms (Yes = 1) as opposed to hours of the day being highly associated with absence of alarm (No = 0). This suggests that both features capture different aspects of the time-based pattern in the dataset and are worth keeping as predictors. Figure 9 shows a follow up plot to visualize the frequency of fire alarm classes per hour of the day. The pattern is showing a bimodal distribution for triggered alarms (Yes = 1) starting early morning and midday ranges. For the day of the week, the frequency of fire alarm classes per day is shown in Figure 10. The result indicates that alarms were triggered most frequently on Thursdays of the week in the month of June 2022.

**Data Wrangling and Pre-Processing**

During preprocessing, no missing values were identified across the dataset. The next data wrangling step performed was extracting time components from the UTC variable. Month and year were determined to have the same value throughout the UTC variable which was June 2022. These components were omitted while hour of the day and day of the week were kept as predictors in exchange to the UTC variable. As detailed in the EDA section, predictors with high Pearson correlation at >0.75 score were removed from the data frame to avoid features weighing similar information (Figure 6). A skewness assessment was then performed to measure the asymmetry level of the feature’s probability distribution. This is followed by the Box-Cox transformation to reshape non-normal predictors to satisfy normality assumption of statistical models and stabilize variance. As shown in Table 1, the sample counter feature benefited the most at having <0.5 skewness score with λ = 0.2 transformation which is between a log transformation (λ = 0) and a square root transformation (λ = 0.5). The Box-Cox transformed sample counter was used to replace the original distribution and the rest of the predictors still having >0.5 skewness score were kept at their original instance distribution. The last transformation that was applied is Center & Scale to standardize the features for increased comparability and accommodate for algorithms that are especially sensitive to the scale of the data. The train and test set predictors are then transformed to have a mean of zero and standard deviation of 1.

**Table 1**

*Skewness Before and After Box-Cox Transformation (x<0.5)*

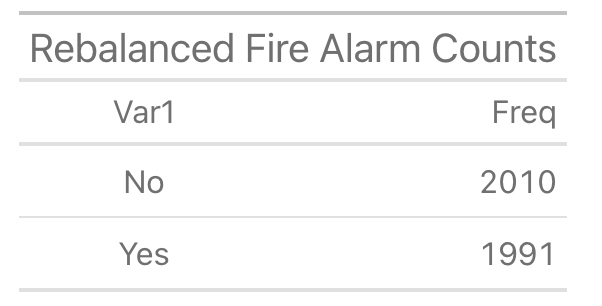


**Data Splitting**

Data splitting was performed by partitioning the preprocessed data set into 80% training set and 20% test set. Then the fire alarm outcome of the training set was evaluated for class imbalance. There was 2,900 absence of alarm (No = 0) count against 1,101 triggered alarms (Yes = 1) count. The major class has around 2.6 times more instances than the minor class. To correct for this class imbalance, random over-sampling technique was employed to generate a synthetic balanced train set by over-sampling the minority class (Yes = 1) and under-sampling the majority class (No = 0). This technique gives equal importance to both classes during modeling and gives the minority class of interest the opportunity for better representation. Table 2 shows the fire alarm class distribution result of the train set after the application of random over-sampling. Reduced, transformed, and rebalanced train set was then fed into selected models for training and tuning. Reduced and transformed test set was later used to assess the performance of each model created.

**Table 2**

*Rebalanced Fire Alarm Outcome Variable*



**Model Strategies (Liam)**

Start here…

**Validation and Testing (Sahil)**

Start here…

**Results and Final Model Selection (Sahil)**

Start here…

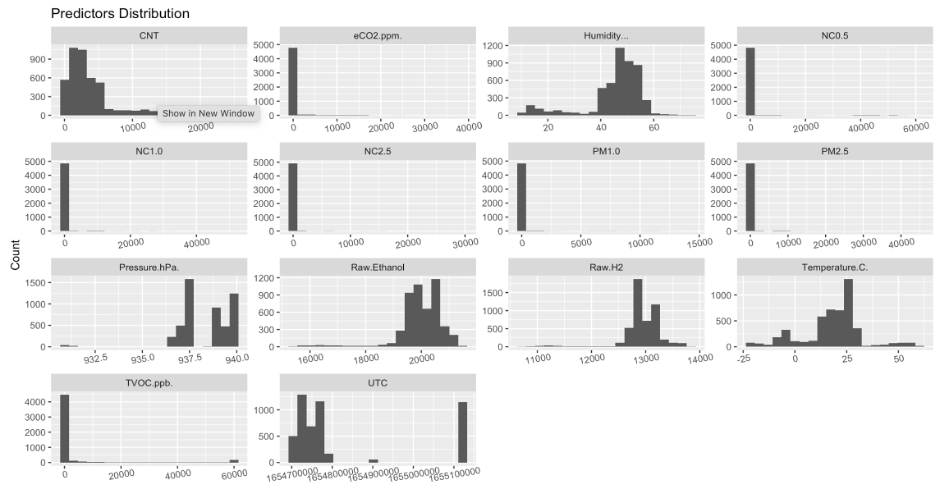
**Discussion and Conclusion (Sahil)**

Start here…

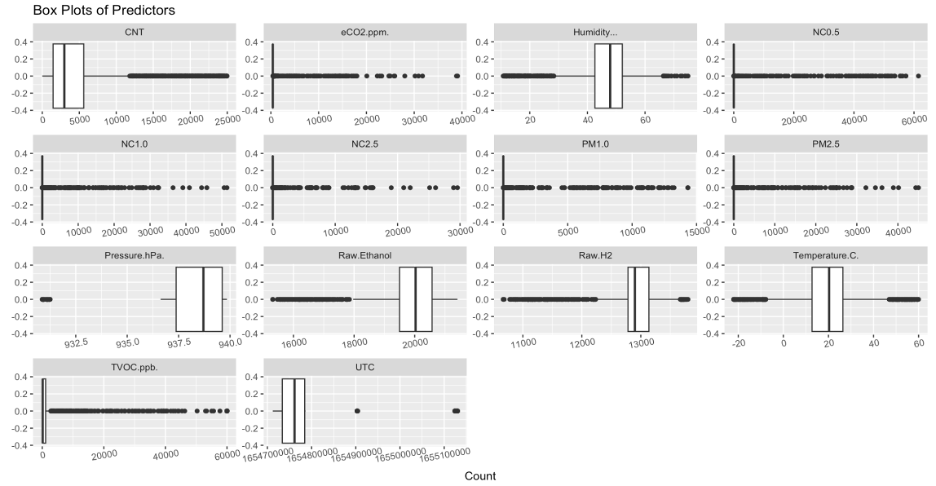
**Figures**

**Figure 1**

*Smoke Alarm Predictors: Data Distribution*

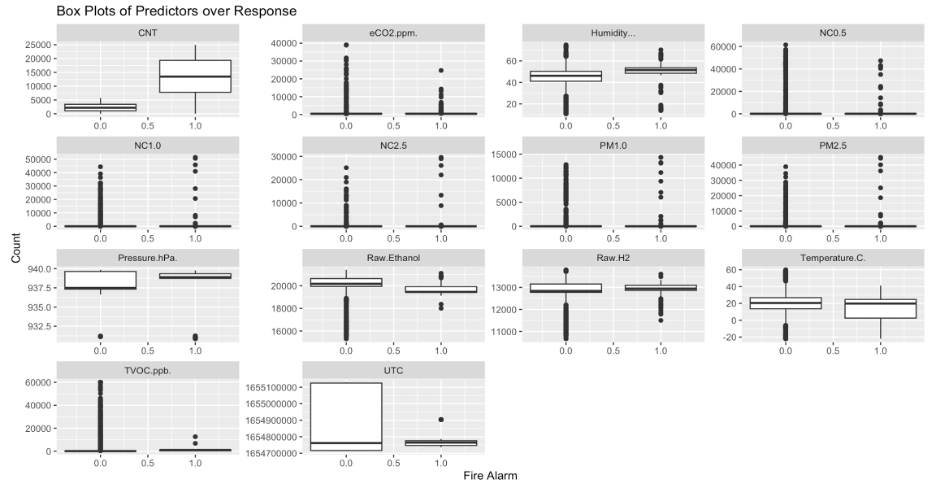


**Figure 2**

*Smoke Alarm Predictors: Box Plots Showing outliers*

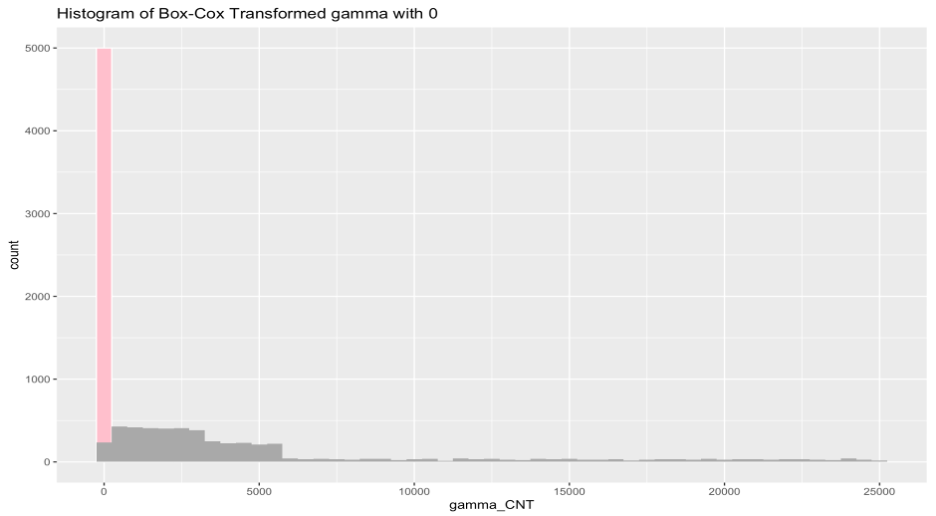
**Figure 3**

*Fire Alarm Outcome Box Plots for each Predictors*



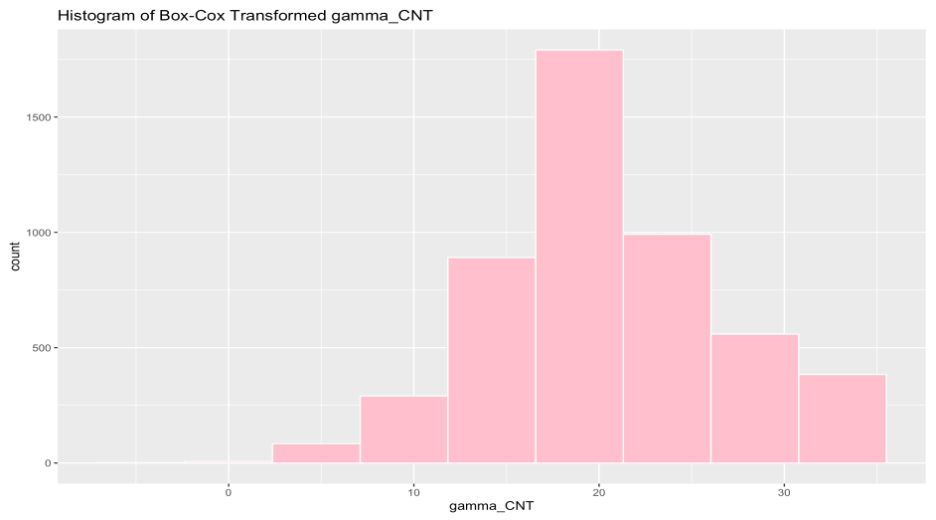
**Figure 4**

*Sample Counter Original Distribution (Dark Gray) vs Box-Cox Transformed Distribution (Pink)*



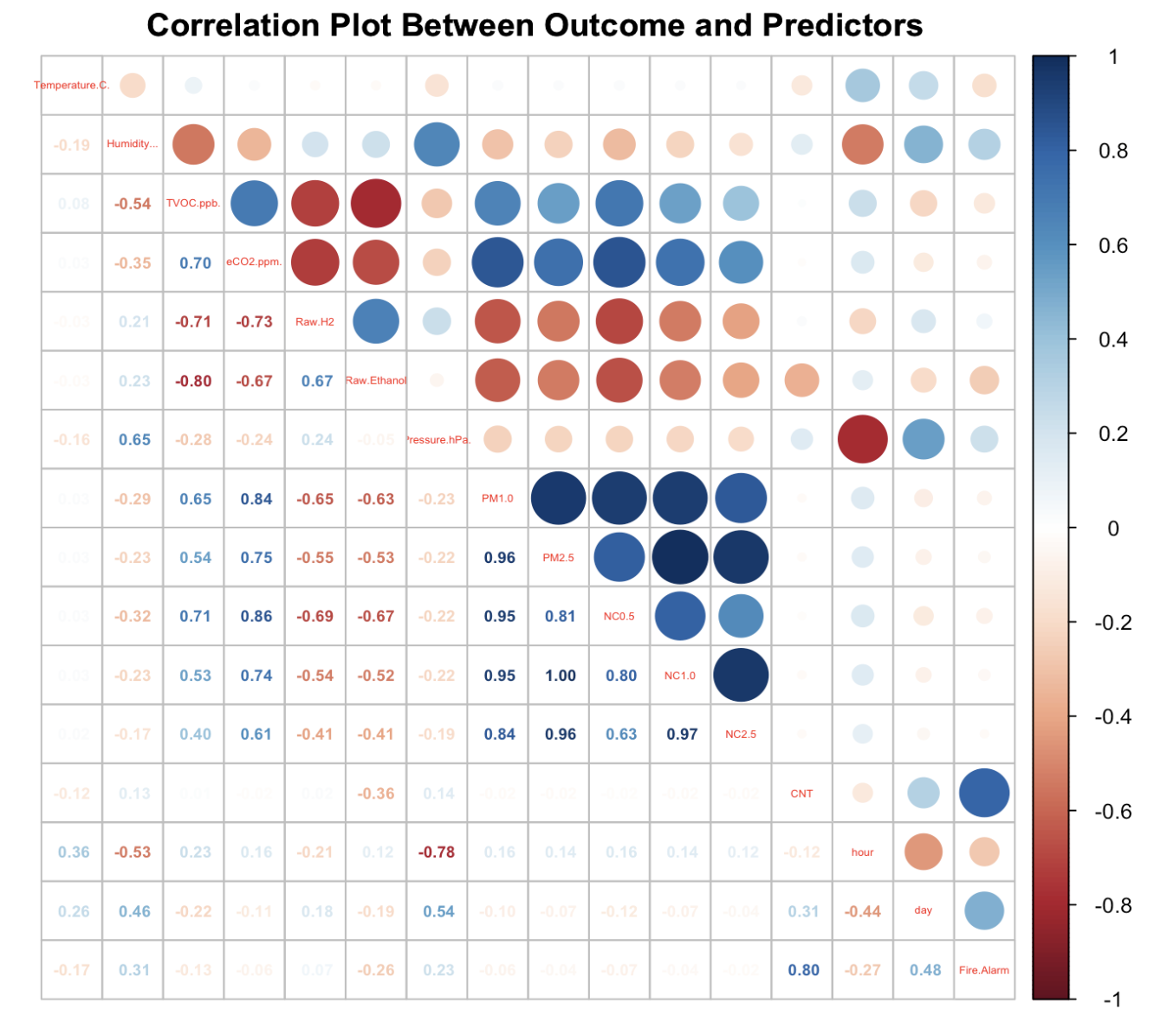
**Figure 5**

*Sample Counter Box-Cox Transformed Distribution*



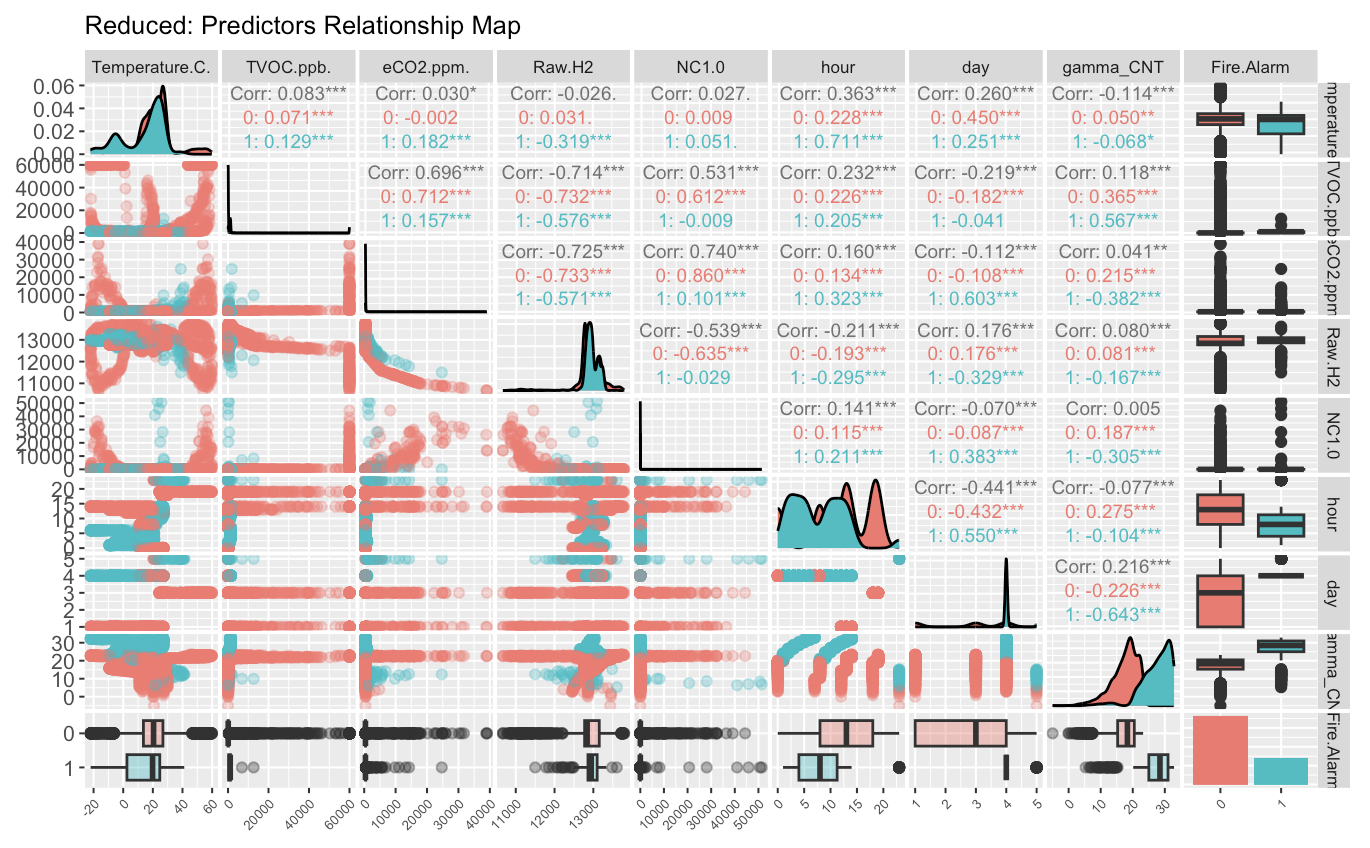
**Figure 6**

*Correlation Plot Comparison: Original vs Reduced Dataset*

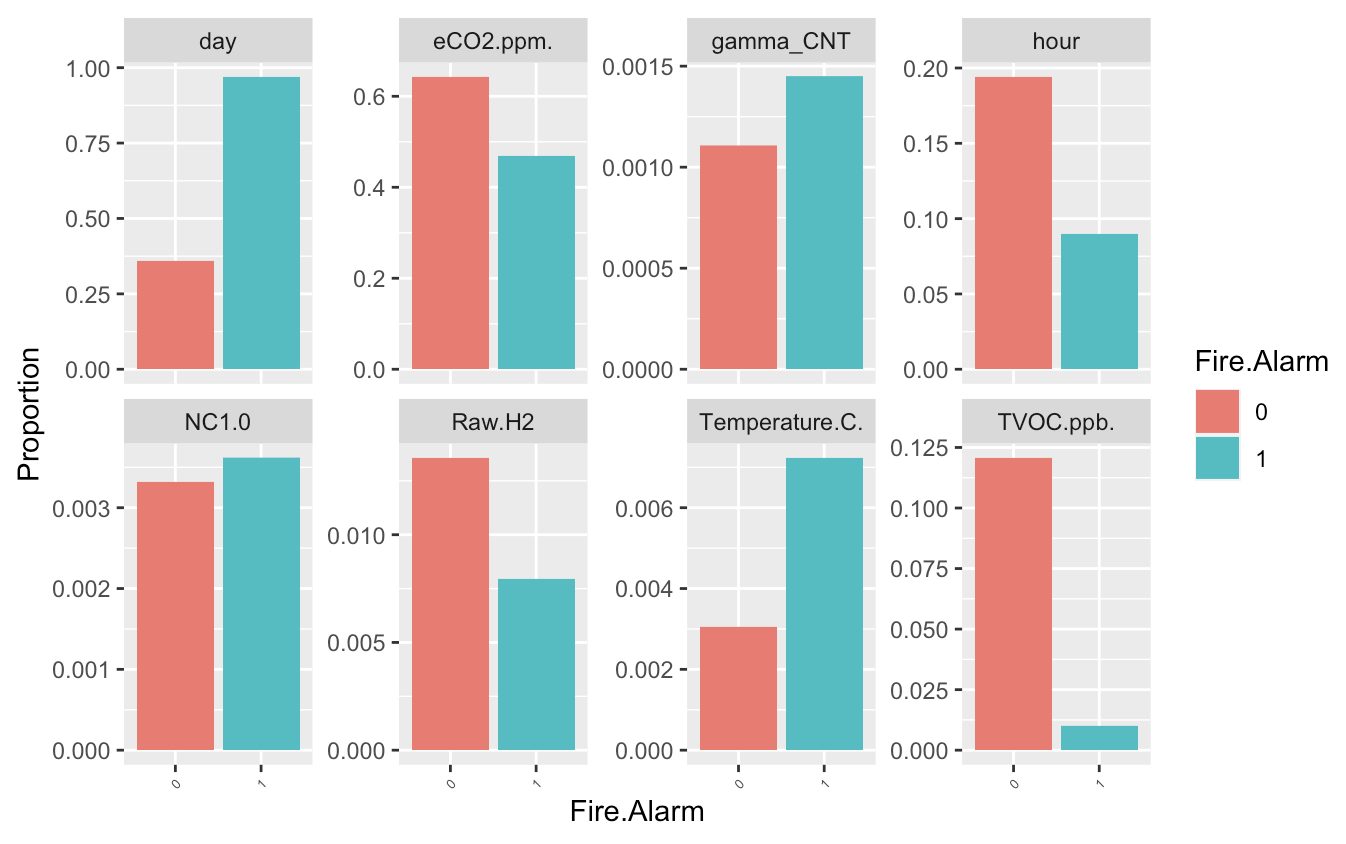


**Figure 7**

*Reduced Smoke Alarm Dataset: Pairwise Relationship Plot*

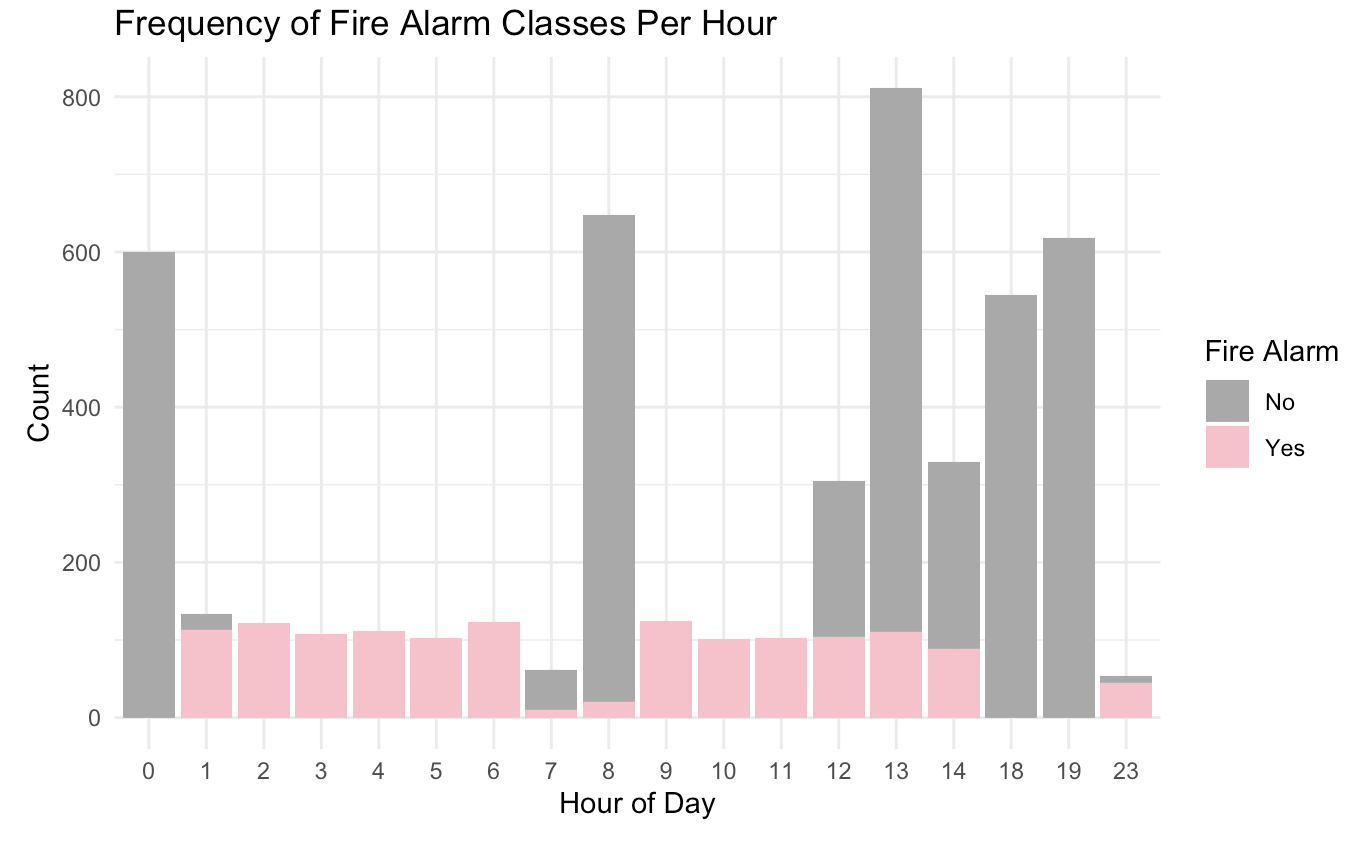
****Figure 8**

*Reduced Smoke Alarm Dataset: Fire Alarm Class Proportion*



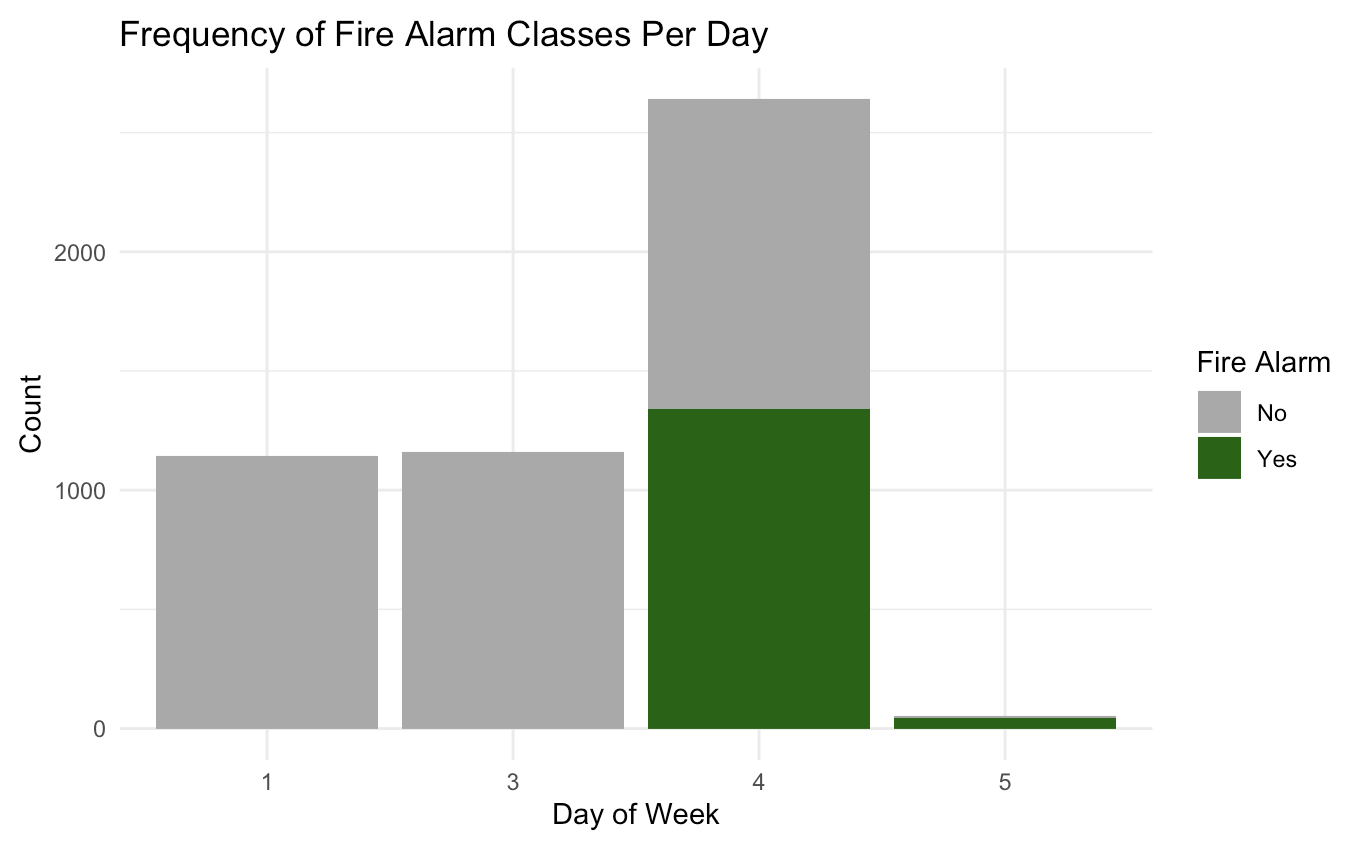
**Figure 9**

*Frequency of Fire Alarm Classes Per Hour*



**Figure 10**

*Frequency of Fire Alarm Classes Per Day*

**

**References**

Start here…

**Appendix**