Location & Time. Primary factors in Predicting Fire:

By Syeda Aziz, Sara Opris, Mohammed Siddiqui

Project Overview

Emergency services like fire departments exist to provide support and minimize negative 8 outcomes on lives and structures whenever possible. Although according to data, fire dispatches only make up about 4% of calls to fire departments nationally *(U.S. Fire Administration, Aug. 2018)*, our project aims to provide a predictive model that will enable fire departments and emergency resources to further minimize this number by being able to predict locations of incidence within the city of Seattle. With this information, we can assist stakeholders such as city planning officials, first responders, and medical staff in allocating resources according to risk.

Currently, the city of Seattle operates 33 fire departments that are equally split across 6 battalions within the city to provide coverage based upon the following geographic sections: the downtown/Central Area, north and northeast Seattle, northwest Seattle, south and southeast Seattle and West Seattle (*Scoggins, Harold D., 2020)*. When an emergency is initially called into 911, the 911 dispatcher will re-route the call to the Fire Alarm Center (FAC) who collects address information from the caller, and then calls in a local fire department based upon the closest corresponding battalion. (*Fire Alarm Center - Fire*, City of Seattle)
The goal of our project is to analyze the 911 dispatch data post-pandemic time from Seattle to see if we can ascertain any areas of increased incidence, or other factors that may help predict where fires are more likely to occur. Utilizing this data, we can provide supporting information for the city planners and the fire department to make determinations on how to better apply resources around areas or times of focus.

Previous related studies have been done, one of focus being the Firebird project that is implemented by the Atlanta Fire Department. While Firebird uses SVM and Random Forest models to predict buildings that are likely to catch fire, this project differs from our own in a few key regards. Firebird focuses mainly on the commercial buildings and generates a risk score for buildings for the fire department to then utilize as a method for prioritizing fire inspections. Our project will focus on both commercial and residential areas within Seattle and will utilize the 911 dispatch data from the years 2020-2022 to identify areas that are at higher risk for fire.

Description of Work
While our project initially looked to analyze all the data available from the 911 dispatch data, we soon found out that we had some technological limitations to running through data that large. Due to these issues, we had to re-evaluate our goal and find a way to minimize our data in order to be able to run the statistical analyses required for our research. Our first method of minimizing our dataset was to limit our research to the pandemic years of 2020-2022 instead of using the full data set of 2004 – present. This would allow us to uncover any trends in economic recovery and protect us from data drift if we used data over too long of a time frame.

After limiting our research time frame, we proceeded to clean our data by dropping all rows with missing or NA values. Additionally, we separated our original data file into a secondary smaller data frame that we will be using to perform logistic regression, which includes the following variables: Date time, Report location, and incident type. We split out the datetime columns into year, month, day and hour so that we can leverage more granular analysis. We also used Longitude and Latitude at the 2^{nd} degree decimal point due to computational limitations. Lastly, we created a new response variable from incident type which was 1 if the incident type was fire or 0 if otherwise. After treating all our predictors and response variables as Factors, we were able to run our model.

Next, we went on to perform the following analyses on our dataset: Logistic Regression, Decision Tree, and Random Forest Model.
For the logistic regression, we had initially planned on doing a linear regression but decided to instead change the incident type in our dataset into a categorical variable of 1 for type equaling fire, and 0 for all other incidence types. After doing so, we ran our glm function with Type (fire) as our response and our other variables including Longitude, Latitude, Month, Date, and Time as our predictors. As we were limited with our feature-set, we tried to determine via classification how critical the above factors were in predicting fires and then to evaluate our results against the context of the Tableau visualization we created and a Google maps search of neighborhoods. In doing so, we can paint a better picture and provide analyses to support the Seattle Fire Department. We used multiple factor weights in testing our model as initially, our model was amply predicting 0 for every instance and we wanted to punish false negatives.
For the decision tree model, we applied a cost matrix to penalize false negatives by a factor of 10. Our tree was simple, and performance was similar to our first glm in which we simply predicted 0 for every instance. To improve the performance of our decision tree, we decided to build a random forest model.

Finally for our random forest model, we decided to train a model across 300 iterations (limited by computational ability) and leverage all 4 variables per iteration via the mtry parameter. Our random forest model did take some time to train and gave us insight into the importance of each feature in predicting the result.

Results and Interpretation

Logistic Regression Results

As our first logistic regression model simply guessed 0 for each instance, we built out a separate glm with iterated through different weights for the response variable factors. We received the AUC for each of them. Unfortunately, the value of the AUC was roughly 0.545 for all of them and increased only by the hundredths as we optimized the weights which indicated to us that the model was barely better than guessing. This led us to try working with other supervised classification models like decision trees and random forests.

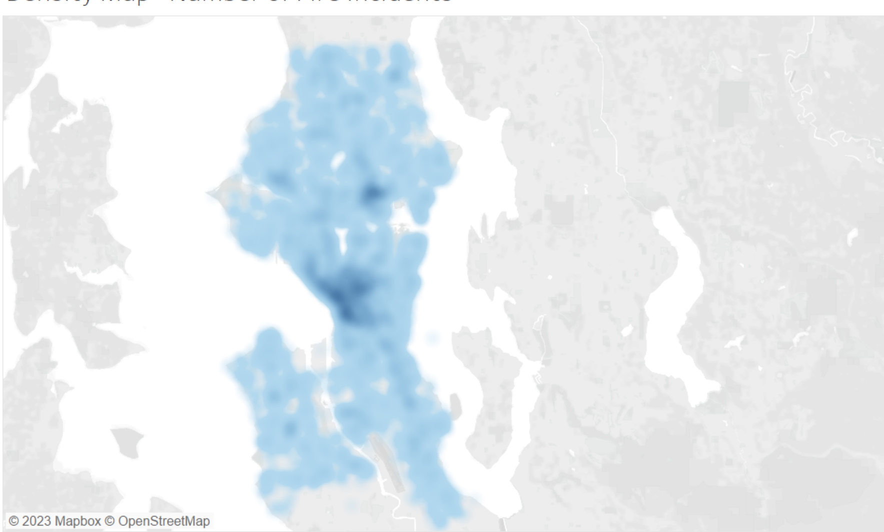
Decision Tree and Random Forest Results

Interestingly, our random forest model concluded that Latitude followed by Longitude (or Location) were the most important features in its predictions. The margin of difference between the location features and the time features was stark. Following this result, we checked the density of fire specific events across a map in Tableau and saw clusters of fire activity in certain neighborhoods like the port area.

We had previously used Tableau to perform EDA. After running our statistical models above, we revisited our visualizations and created a few different graphical representations. The first type of visualization that we ran depicts density maps of Fire based 911 calls spread across the city of Seattle which we have broken up against the 3 years within our study, 2020-2022.

Figure 2A – Density Map of Fire Incidents 2020

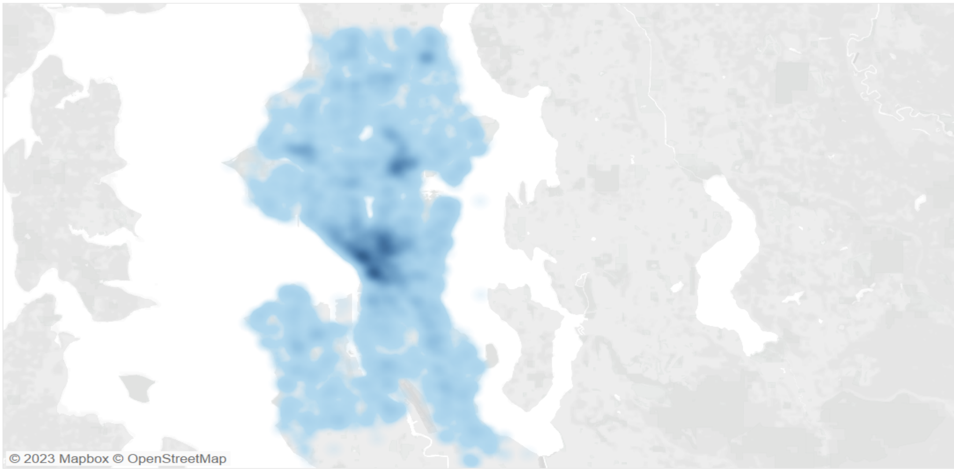
Density Map - Number of Fire Incidents



Map based on Longitude and Latitude. The data is filtered on Type, which keeps 53 of 250 members. The view is filtered on Datetime Year, which excludes 2003 and 2023.

Figure 2B – Density Map of Fire Incidents 2021

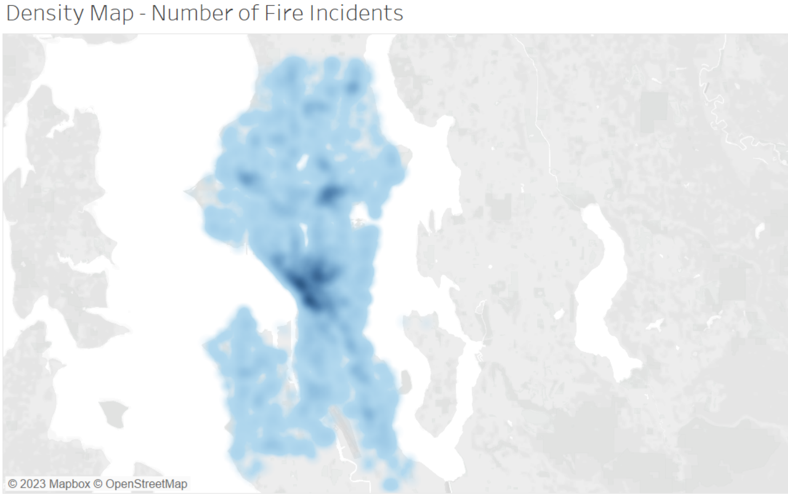
Density Map - Number of Fire Incidents



Map based on Longitude and Latitude. The data is filtered on Type, which keeps 53 of 250 members. The view is filtered on Datetime Year, which excludes 2003 and 2023.

Figure 2C- Density Map of Fire Incidents 2022

Density Map - Number of Fire Incidents



Map based on Longitude and Latitude. The data is filtered on Type, which keeps 53 of 250 members. The view is filtered on Datetime Year, which excludes 2003 and 2023.

From the above density maps, it would appear that density has slightly increased over the span of our timeframe, but the amount is arguably negligible from the map alone, and therefore not a clear indicator.

The second visualization we ran was frequency of fire incidences based upon location, which we organized as a graph. From this we can see that a couple of location points drive up the majority of 911 calls for fire, which suggests there are specific points within the city that are much more likely to have a fire than other random areas in the city.

Figure 3A – Locations and Number of Incidents

Locations and Number of Incidents

Report Location	
POINT (-122.330809 47.602114)	239
POINT (-122.376149 47.638695)	166
POINT (-122.333001 47.623222)	164
POINT (-122.299024 47.584234)	154
POINT (-122.3134 47.598322)	132
POINT (-122.345399 47.612716)	117
POINT (-122.322726 47.679478)	116
POINT (-122.325207 47.604377)	114
POINT (-122.330457 47.609088)	105
POINT (-122.327854 47.618497)	104
POINT (-122.331449 47.602813)	92
POINT (-122.328965 47.599968)	91
POINT (-122.293193 47.571853)	90
POINT (-122.319413 47.605346)	87
POINT (-122.362969 47.553321)	82
POINT (-122.34681 47.698721)	81
POINT (-122.32821 47.608054)	80
POINT (-122.304237 47.586234)	79
POINT (-122.296531 47.732881)	77
POINT (-122.387209 47.545436)	75
POINT (-122.345064 47.731129)	69

Count of Repor..
0 239

From the above figure, we tried to showcase the number of incidents based on their exact location or coordinates. So, from Google Maps, the point (-122.330809 47.602114) Pioneer Square has the highest number of incidents reported over the years with 239 reports filed. Then the second highest number of incidents took place in 15th Ave W in Downtown Seattle with 166 reports. In this way, we tried to depict the number of incident cases based on their exact coordinates using Tableau visualization.

Conclusion & Business Implications

In our initial report, we noted that there seemed to be an increase in incidences in the months of July, August, and December. In conducting our research, we have found that our logistic regression model does appear to consider those months as statistically significant towards a higher rate of fire calls. When looking at the p-values provided from our logistic regression, we can observe that there is a higher significance not only for those months, but for all the months from July – December. This conclusion is also supported by secondary research that found similar results, with July, August, and October having an increase in emergency fire response (*U.S. Fire Administration, Aug. 2018*). Based on that information, our recommendations would be to focus emergency fire preparedness on those months. This can be done by implementing things such as increased staffing or providing educational fire prevention seminars during that period.

Additionally, we did find that certain hours and days of the month were more prone to fire calls than others as shown in the output.txt file within Github. We also saw, as evidenced below, that certain locations, especially in key business and social districts, are more prone to fires. Based on our findings, we would recommend planners to ensure they can readily access the densest locations within a brief time frame as well as remain staffed between the late am to late pm hours where incidents seem to be most common.

We would recommend further study to compare the incident data for other types of calls to the ones we identified for fires. Additionally, we'd investigate other predictors for fire incidents to evaluate the relationship. For example, type of building, location, wealth, temperature, cause of fire, the impact of the fire on life and property damages, and other features might provide data which can help both for prediction and visual analysis. The long-term goal would be to derive insights which can assist the Seattle Fire Department and policy makers more effectively staff resources and prevent future fires as well as respond to new incidents more efficiently.

While our analysis was limited to few predictors and data from the years 2020 – 2020 which coincides with the Covid-19 pandemic, we will have trouble generalizing to 'normal' or post-pandemic years. We relied heavily on support from EDA as well as external context to interpret our models and understand what was going on. With more granular data and computational resources, we may be able to provide concrete recommendations for city planners and the fire departments with a higher level of confidence.

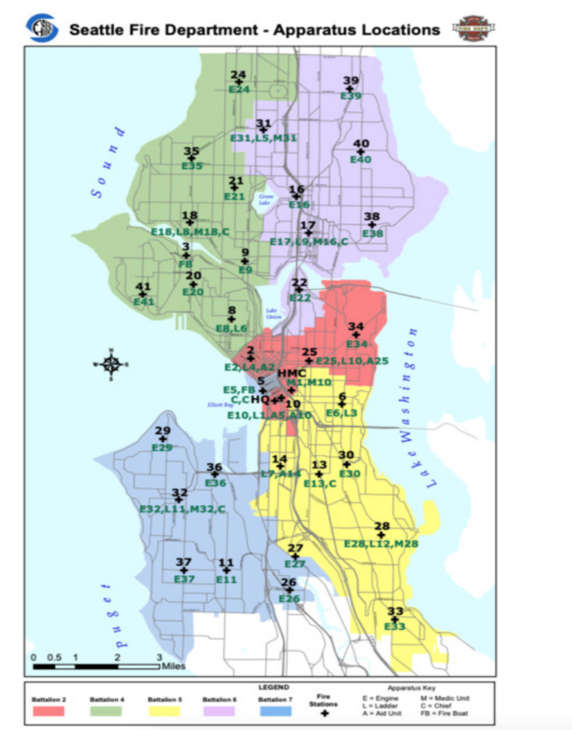
Figure 4A – Logistic Regression Coefficients

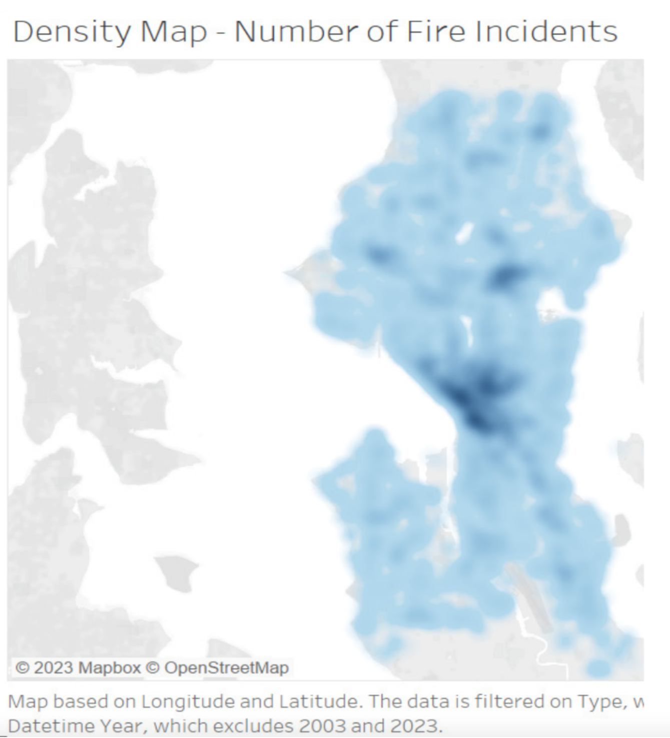
```
Call:
glm(formula = Type_Fire ~ Latitude + Longitude + month + day +
    time, family = binomial, data = model_df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8302  -0.4877  -0.4598  -0.4266   2.3783

Coefficients: (1 not defined because of singularities)
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.423e+00  1.230e-01 -19.704  < 2e-16 ***
Latitude47.2    -8.958e+00  6.241e+01  -0.144  0.88587
Latitude47.3     1.158e+00  9.235e-01   1.253  0.21003
Latitude47.4    -3.719e-01  2.272e-01  -1.637  0.10167
Latitude47.5    -2.283e-02  1.247e-01  -0.183  0.85477
Latitude47.6     1.785e-01  1.246e-01   1.433  0.15174
Latitude47.7    -1.738e-01  1.261e-01  -1.378  0.16812
Longitude-121.7 -8.937e+00  1.135e+02  -0.079  0.93722
Longitude-121.8 -9.112e+00  1.977e+02  -0.046  0.96323
Longitude-121.9 -1.997e-01  2.063e+02  -0.001  0.99923
Longitude-122.0 -4.783e-04  1.098e+00   0.000  0.99965
Longitude-122.1 -9.289e+00  5.404e+01  -0.172  0.86351
Longitude-122.2  1.337e-01  5.514e-02   2.425  0.01532 *
Longitude-122.3  8.154e-02  5.331e-02   1.529  0.12617
Longitude-122.4         NA         NA      NA       NA
month2           1.812e-02  3.075e-02   0.589  0.55572
month3           5.967e-02  2.999e-02   1.989  0.04665 *
month4           6.009e-02  3.024e-02   1.987  0.04694 *
month5           5.498e-02  2.967e-02   1.853  0.06385 .
month6           2.653e-02  2.953e-02   0.898  0.36895
month7           1.632e-01  2.847e-02   5.733 9.90e-09 ***
month8           1.691e-01  2.850e-02   5.932 2.98e-09 ***
month9           2.091e-01  2.866e-02   7.293 3.02e-13 ***
month10          2.350e-01  2.843e-02   8.266  < 2e-16 ***
month11          1.370e-01  2.925e-02   4.684 2.82e-06 ***
month12          1.315e-01  2.865e-02   4.592 4.40e-06 ***
```

Additionally, our models lead us to conclude that there are areas within Seattle that have a higher prevalence of emergency fire calls. As shown below, we can compare our density map against fire stations within Seattle, to identify which areas should be prioritized (*"Seattle Fire Department - Apparatus Locations."*). From this we can note that fire stations 2,5,10 and 25 appear to have the most calls. Based on this information, our recommendations would be to increase the focus of funding and staffing at these locations, and secondary focus as well on the fire stations 17 & 18 that also have a higher density when compared to the rest of the city.

Figure 5A – Fire Density compared to Fire Stations – Seattle

From the density map below, we tried to show the side-by-side comparison of our data with the Seattle fire station data. It seems that both the maps have similarities with each other, which mainly focused on the incidents in Downtown Seattle. We did provide a supporting visualization report to support our cases regarding this data.





Citations

- "Fire Alarm Center." *Fire Alarm Center - Fire*, City of Seattle, https://www.seattle.gov/fire/about-us/about-the-department/resource-management/fire-alarm-center.
- Scoggins, Harold D. "Seattle Fire Department 2020 Adopted Budget." City of Seattle, 2020.https://www.seattle.gov/documents/Departments/FinanceDepartment/20adoptedbudget/SFD.pdf
- "Seattle Fire Department - Apparatus Locations." City of Seattle. https://www.seattle.gov/documents/Departments/Fire/About/SeattleFireMap.pdf
- https://www.seattle.gov/fire/about-us/about-the-department/resource-management/fire-alarm-center
- Challands, Neil. "The Relationships between Fire Service Response Time and Fire Outcomes - Fire Technology." *SpringerLink*, Springer US, 10 Oct. 2009, https://link.springer.com/article/10.1007/s10694-009-0111-y.
- "Fire Department Overall Run Profile as Reported to the National Fire Incident Reporting System." U.S. Fire Administration, Aug. 2018, https://nfa.usfa.fema.gov/downloads/pdf/statistics/v19i5.pdf
- Perry, Jonathan. "Trust in Public Institutions: Trends and Implications for Economic Security | DISD." *United Nations*, United Nations, July 2021, https://www.un.org/development/desa/dspd/2021/07/trust-public-institutions/.
- Yamashita, Kiyoshi. "Understanding urban fire: Modeling fire incidence using classical and geographically weighted regression." Published-2008. https://www.proquest.com/openview/de9ef0134be570da0896c3cb0568cc51/1?pq-origsite=gscholar&cbl=18750