Fire Incidents Report

Congyu Hang

2023/3/30

Main Goal

The main goal of this report is to investigate the factors that affect the occurrence of fire and the severity of the fire incident from 2017 to 2021. We will analyze the frequency of fire incidents regarding time, space and relative factors that appear in the incidents.

Background

This dataset includes only fire incidents as defined by the Ontario Fire Marshal (OFM) up to December 31, 2021.

This dataset provides more detail than the basic incidents dataset provides for only fire Incidents to which Toronto Fire Service(TFS) responds to. The format is similar to the reporting data sent by TFS to the OFM.

For privacy purposes personal information is not provided and exact address have been aggregated to the nearest major or minor intersection. Some incidents have been excluded pursuant to exemptions under Section 8 of Municipal Freedom of Information and Protection of Privacy Act (MFIPPA).

Incidents with incomplete data may be under investigation or is classified as a no loss outdoor fire.

Variables of Data

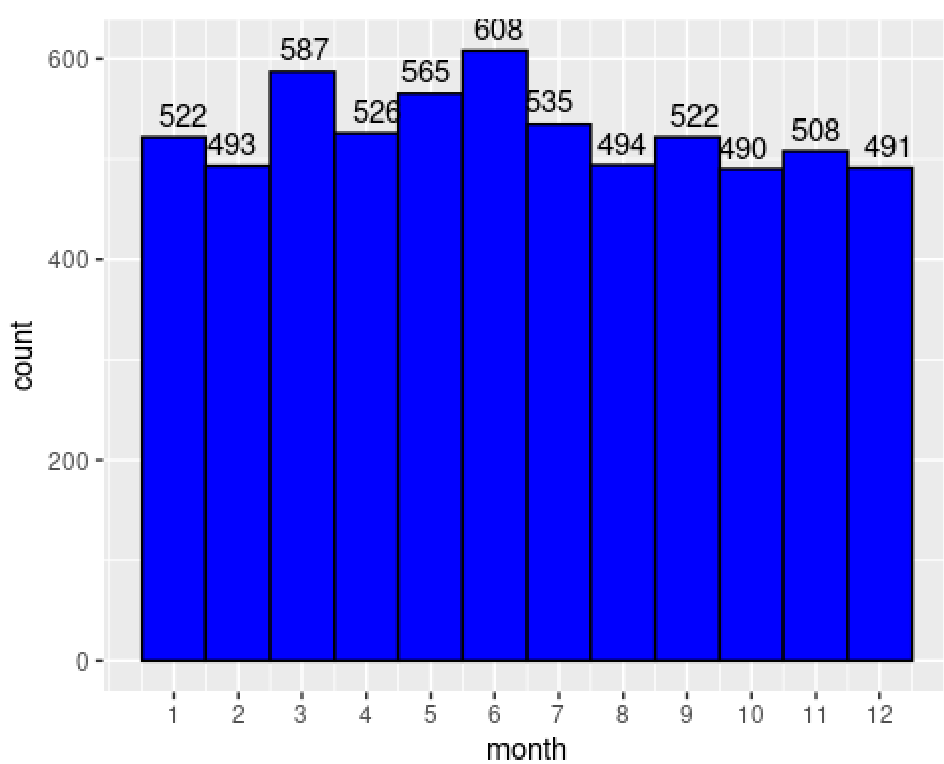
```
Rows: 6,341
Columns: 48
$ X_id                                                            <int> 8…
$ Area_of_Origin                                                  <chr> "…
$ Building_Status                                                 <chr> "…
$ Business_Impact                                                 <chr> "…
$ Civilian_Casualties                                             <int> 0…
$ Count_of_Persons_Rescued                                        <int> 0…
$ Estimated_Dollar_Loss                                           <int> 0…
$ Estimated_Number_Of_Persons_Displaced                           <int> 0…
$ Exposures                                                       <int> N…
$ Ext_agent_app_or_defer_time                                     <chr> "…
$ Extent_Of_Fire                                                  <chr> "…
$ Final_Incident_Type                                             <chr> "…
$ Fire_Alarm_System_Impact_on_Evacuation                          <chr> "…
$ Fire_Alarm_System_Operation                                     <chr> "…
$ Fire_Alarm_System_Presence                                      <chr> "…
$ Fire_Under_Control_Time                                         <chr> "…
$ Ignition_Source                                                 <chr> "…
$ Incident_Number                                                 <chr> "…
$ Incident_Station_Area                                           <int> 1…
$ Incident_Ward                                                   <int> 5…
$ Initial_CAD_Event_Type                                          <chr> "…
$ Intersection                                                    <chr> "…
$ Last_TFS_Unit_Clear_Time                                        <chr> "…
$ Latitude                                                        <dbl> 4…
$ Level_Of_Origin                                                 <chr> "…
$ Longitude                                                       <dbl> -…
```

```
$ Material_First_Ignited                                       <chr> "…
$ Method_Of_Fire_Control                                       <chr> "…
$ Number_of_responding_apparatus                               <int> 6…
$ Number_of_responding_personnel                               <int> 2…
$ Possible_Cause                                               <chr> "…
$ Property_Use                                                 <chr> "…
$ Smoke_Alarm_at_Fire_Origin                                   <chr> "…
$ Smoke_Alarm_at_Fire_Origin_Alarm_Failure                     <chr> "…
$ Smoke_Alarm_at_Fire_Origin_Alarm_Type                        <chr> "…
$ Smoke_Alarm_Impact_on_Persons_Evacuating_Impact_on_Evacuation <chr> "…
$ Smoke_Spread                                                 <chr> "…
$ Sprinkler_System_Operation                                   <chr> "…
$ Sprinkler_System_Presence                                    <chr> "…
$ Status_of_Fire_On_Arrival                                    <chr> "…
$ TFS_Alarm_Time                                               <chr> "…
$ TFS_Arrival_Time                                             <chr> "…
$ TFS_Firefighter_Casualties                                   <int> 0…
$ month_ind                                                    <dbl> 2…
$ year_ind                                                     <dbl> 2…
$ hour_ind                                                     <dbl> 1…
$ extent_severity                                              <dbl> 1…
$ severity_level                                               <dbl> 1…
```

Analysis Regarding Time

The occurrence of fire might be affected by time. The plot indicates the distribution of number of fire incidents with respect to month from year 2017 to 2021, in essence, we would also be able to check the distribution of number of fire incidents with respect to months, seasons and hours of the day.

Month Frequncy Histogram

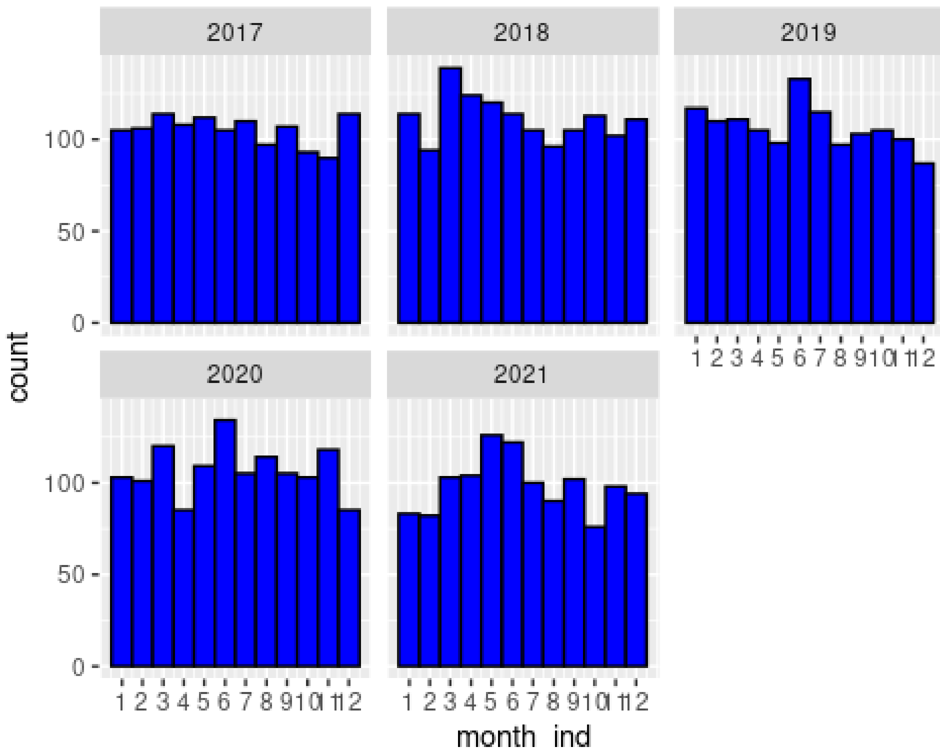


- Spring - March 20 to June 20
- Summer - June 21 to September 21
- Fall (or Autumn) - September 22 to December 20

- Winter - December 21 to March 19

The 3 months that have fire incidents most frequently is June, March and May with values 608, 587, 565 and this also shows that overall, fire incidents happen most frequently in spring from 2017 to 2021.

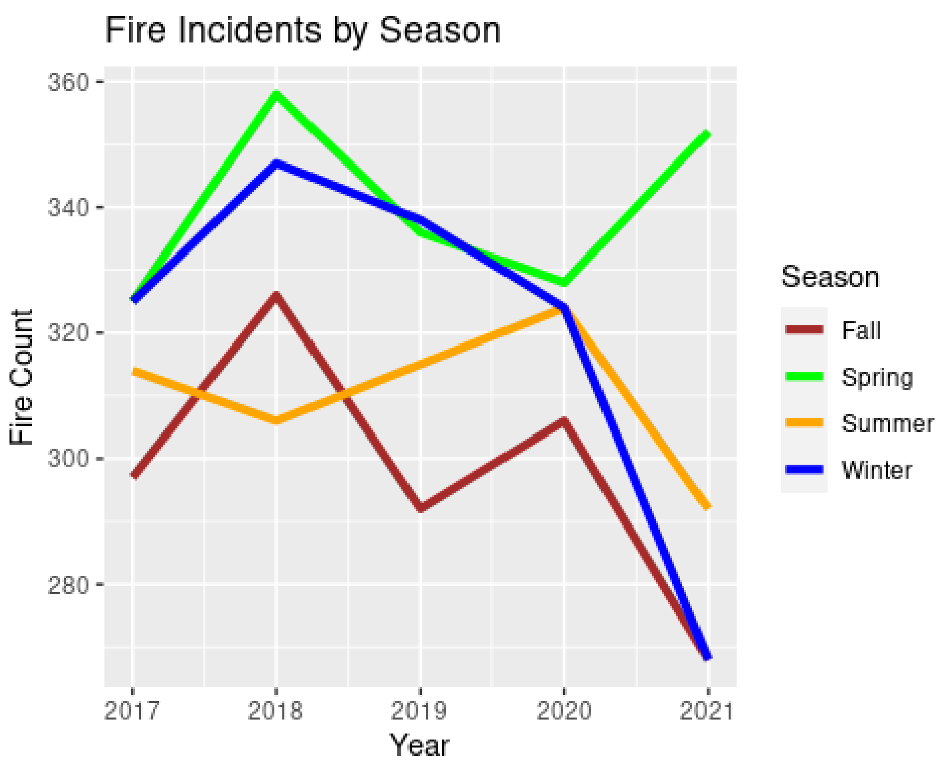
Month Frequency Histogram



The separate histograms with respect to year also indicate that spring has the highest frequency of having fire incidents.

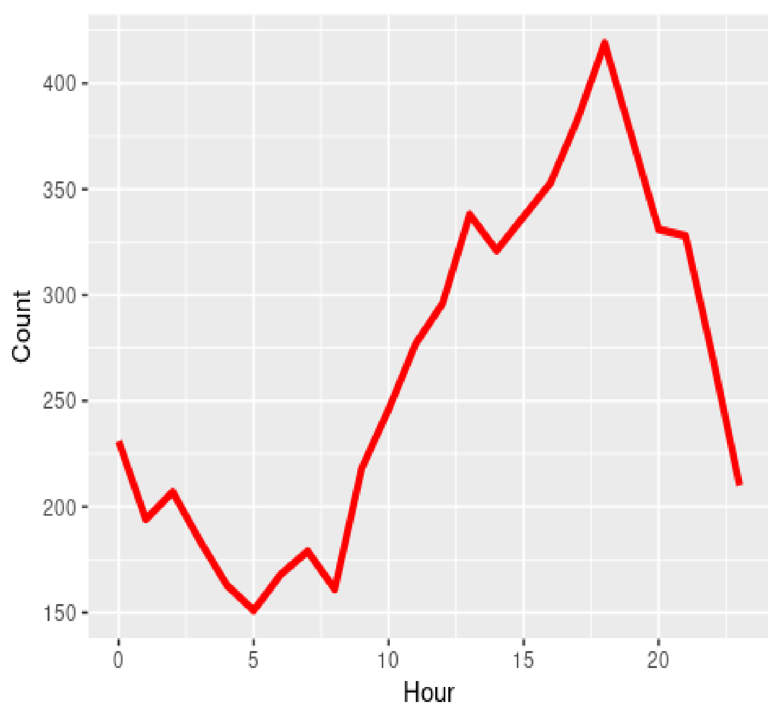
By calculating the total number of fire incidents occurring in each quarter of each year, it indicates that spring has the highest frequency in almost every year (except for 2019 when it was slightly lower than winter). Also, we are able to conclude that the season with the lowest frequency of fire incidents overall is fall. We can show this in the line plot with respect to season as well.

Line Plot of Each Season Regarding Year



The line plot of season's frequency with respect to year conforms to the previous conclusion since it is obvious that the line of spring is approximately above all of the lines of other seasons and the line of fall is almost below all of the other lines.

Frequency of Fire Incidents with respect to Hours



	hour_ind	N
1	18	419
2	17	383
3	19	375
4	16	353
5	13	338
6	15	337
7	20	331
8	21	328
9	14	321
10	12	296

According to the line plot and the table, fire incidents occur most frequently from the afternoon to the evening (15:00 to 20:00).

Conclusion of Fire Incidents w.r.t Time

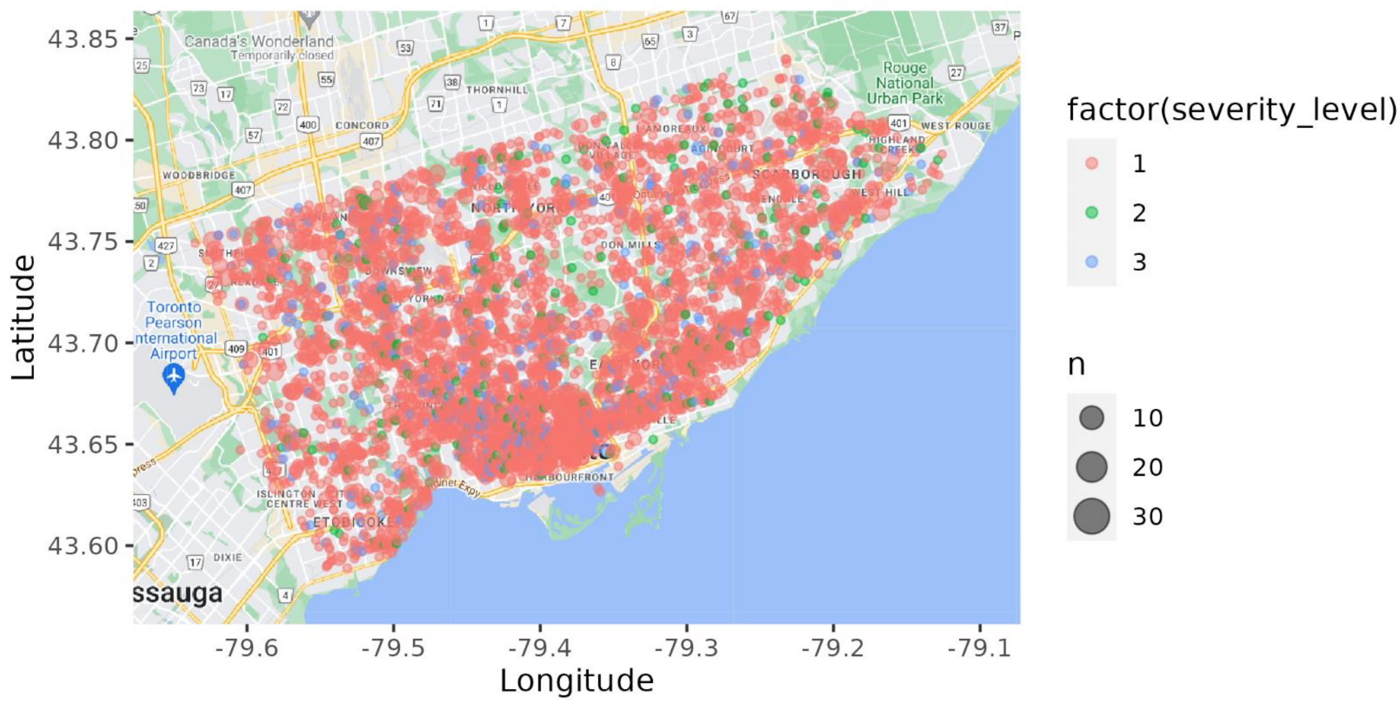
The graphs and table above show that the months that have the highest frequency of fire incidents are June,March and May, which shows that Spring have the highest frequency of fire incidents. The period in the day that have highest frequency of fire incidents is from afternoon to evening (15:00 to 20:00).

Analysis Regarding Space

In this section, the plots are generated based on locations and severity of fire incidents. The severity of fire is allocated to 3 ordinal levels:

level 1 means the extent of fire is within the room or the floor
level 2 means the extent of fire is out of the room and have spread across the room but within the structure
level 3 means the extent of the fire is spread out of the structure

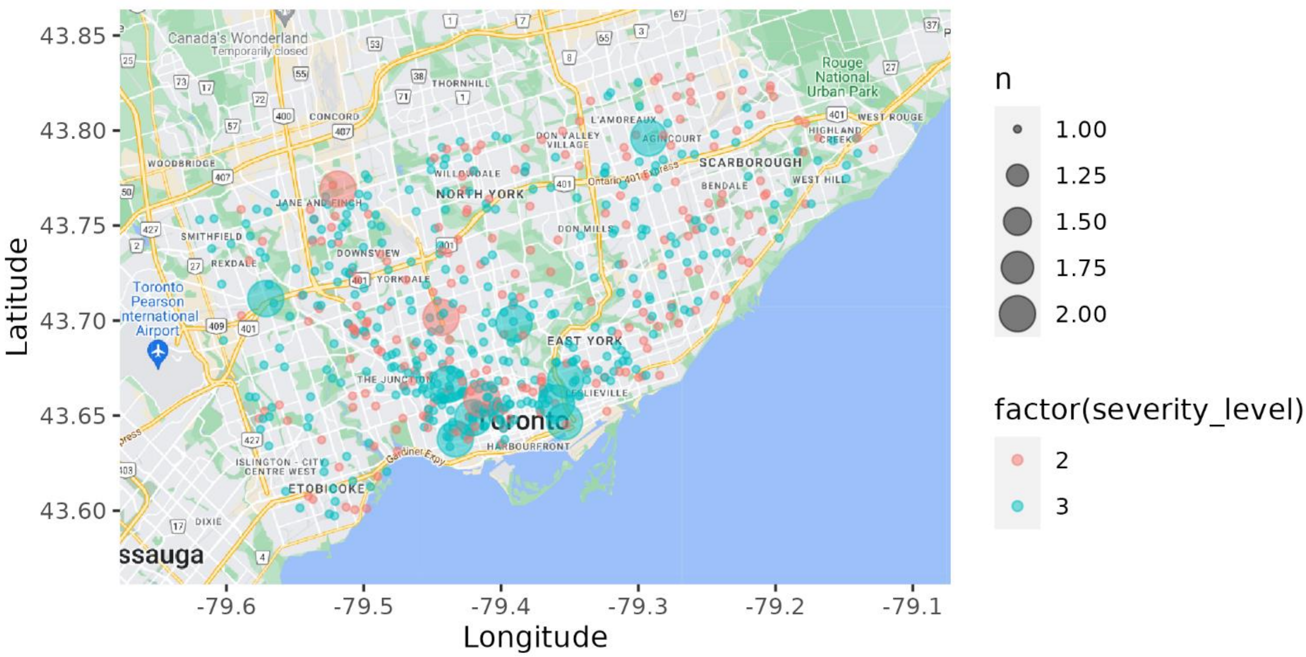
Spatial distribution of fire incidents (all):



Based on the above graph we can conclude that from year 2017 to 2021, fire incidents occur most frequently in downtown area. Secondly, the area close to Lake Ontario (southern region) is more densely populated with points than the north in general. In addition, forest areas (uninhabited) usually have fewer fire incidents. These findings imply that more densely populated areas are more prone to fires.

Spatial distribution of fire incidents (severity_level > 1):

severity_level > 1 represent all fire incidents that spread out of the floor of the ignition area.



In terms of severity, more severe (severity_level greater than 1) fires occur more frequently in southwest Toronto than in other areas.

Conclusion of Fire Incidents w.r.t Space

The graphs above illustrates that southwest area have highest frequency and severity of fire incidents and we can imply that more densely populated areas are more prone to both high frequency and high severity of fires.

Apply Table and Regression To Investigate Factors That Affect Severity of Fire

In this section, take the following variables into consideration:

Sprinkler_System_Operation / Ignition Source /Method_Of_Fire_Control /Status_of_Fire_On_Arrival /Smoke_Alarm_at_Fire_Origin /Month / Hour /Possible_Cause

Ignition Source

- The Ignition Source variable has more than 100 categories, therefore, use functions to classify them to 8 categories based on standards published by Statistics Canada.

https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=794268&CVD=794268&CLV=0&MLV=1&D=1&adm=0&dis=0

- [1] others
- [2] other electrical equipment
- [3] appliances and household equipment
- [4] heat equipment
- [5] cooking equipment
- [6] electrical distribution equipment
- [7] smoker's material and open fire
- [8] exposure

Ignition Source Severity Proportion Table

	ign_classification	sev1_prop	sev2_prop	sev3_prop
1	smoker's material and open fire	0.887	0.0309	0.0825
2	others	0.864	0.0580	0.0780
3	exposure	0.9	0.025	0.075
4	heat equipment	0.917	0.0394	0.0433
5	other electrical equipment	0.922	0.0388	0.0388
6	electrical distribution equipment	0.920	0.0435	0.0367
7	cooking equipment	0.980	0.0131	0.00726
8	appliances and household equipment	0.943	0.0528	0.00407

The table shows that fire incidents with ignition source of smoker's material and open fire, others and exposure have high proportion of causing severest level of fire incidents.

Logistic Regression with Cross Validation

In the logistic regression, consider all cases that have fire not spread across the space not severe, and fire incidents that spread across the room with ignition source inside or wider spread cases severe. Apply cross validation with proportion 0.4,0.6 and use stepAIC to find the model that fits the data best and in this model selection, consider single variables only.

- stepAIC helps to get the model with smallest AIC based on train data set and we test the model's prediction power using test data set.
- The model should include the following terms:

```
logit.mod1 = glm(binomial_severity_level ~ Sprinkler_System_Operation
                 + ign_classification
                 + Status_of_Fire_On_Arrival
                 + Smoke_Alarm_at_Fire_Origin
                 +Possible_Cause, family = binomial, data = fd3_train)
```

Interpretation of Logistic Regression

- "1 - Sprinkler system activated" is ref of Sprinkler_System_Operation
- "appliances and household equipment" is ref of ign_classification
- "1 - Fire extinguished prior to arrival" is ref of Status_of_Fire_On_Arrival
- "1 - Floor/suite of fire origin: No smoke alarm" is ref of Smoke_Alarm_at_Fire_Origin
- "01 - Suspected Arson" is ref of Possible_Cause

The intercept is -4.87303

interpretation: log odds of having severe fire incident is -4.87303 when Sprinkler_System_Operation is "1 - Sprinkler system activated", ign_classification is "appliances and household equipment", Status_of_Fire_On_Arrival is "1 - Fire extinguished prior to arrival", Smoke_Alarm_at_Fire_Origin is "1 - Floor/suite of fire origin: No smoke alarm", Possible_Cause is "01 - Suspected Arson", i.e. log odds of having severe fire incident with all variables of their reference group.

Slope 0.62144 (beta1)

interpretation: log odds ratio(log OR) of having severe fire incident is 0.62144 between Sprinkler_System_Operation is "2 - Did not activate: remote from fire" vs. Sprinkler_System_Operation is "1 - Sprinkler system activated" holding all other variables not changed.

Slope 1.18418 (beta5)

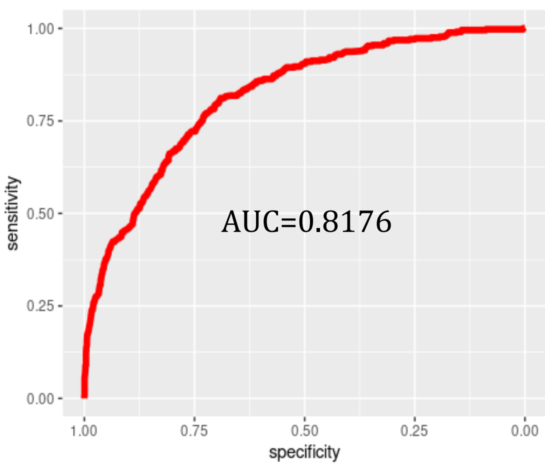
interpretation: log odds ratio(log OR) of having severe fire incident is 1.18418 between Sprinkler_System_Operation is "8 - Not applicable - no sprinkler system present" vs. Sprinkler_System_Operation is "1 - Sprinkler system activated" holding all other variables not changed.

Test of Hypothesis for beta5:

- H0: beta5=0
- H1: beta5!=0
- p-value = 1.79e-06 < 0.05
- At 0.05 level of significance, since p-value = 1.79e-06 < 0.05, we reject the null hypothesis H0: beta5=0.

ROC and AUC:

The curve is plotted based on the model that was constructed by training data set and fitted values of test ing data set since cross validation is applied.



- AUC of this model is 0.8176, which shows that the model distinguishes between severe and not severe fire incidents well.

Bootstrap of beta5 and Test of Hypothesis Using CI

```
boot_func = function(){
  boot.d = fd3_train %>% sample_n(nrow(fd3_train),replace = T)
  m2=glm(binomial_severity_level ~ Sprinkler_System_Operation
         + ign_classification
         + Status_of_Fire_On_Arrival
         + Smoke_Alarm_at_Fire_Origin +Possible_Cause,
         family = binomial, data = boot.d)
  s=coef(m2)[6]
  return(s)
}

output=replicate(1000,boot_func())
quantile(output,c(0.025,0.975))
```

```
##      2.5%     97.5% 
## 0.7606394 1.7258521
```

By sampling with replacement, we constructed the Bootstrap CI with 1000 replication of Bootstrap sampling. The Bootstrap 95%-CI of beta5 is between 0.76 and 1.73, and the estimated beta5 is 1.18418, which is inside the CI, therefore, reject the null hypothesis of beta5=0 since 0 is not in the CI. Moreover, the result conforms to the p-value displayed in the summary of the logistic model.

Conclusion

In conclusion, the analysis of fire incidents in Toronto from 2017 to 2021 reveals several key trends. First, fire incidents occurred most frequently in the spring months, particularly in June, March, and May. Afternoon to evening hours (15:00 to 20:00) registered the highest frequency of fire incidents. Second, the data demonstrates a spatial correlation between fire incidents and population density, with downtown areas and regions near Lake Ontario (southern region) experiencing a higher frequency of fires. In contrast, uninhabited forest areas registered fewer fire incidents.

Additionally, the study found that more severe fires (severity_level greater than 1) tended to occur more frequently in southwest Toronto compared to other areas. This finding supports the inference that more densely populated areas are not only prone to a higher frequency of fires but also to fires of greater severity. Lastly, the ignition sources most commonly associated with the severest level of fire incidents were smoker's material and open fire, followed by other sources and exposure.

According to the Logistic Regression with AUC 0.8176, factors other than space and time that affect severity of fire are Sprinkler System Operation, Ignition Source, Status of Fire on Arrival, Smoke Alarm at Fire Origin and Possible Cause of fire incident. These factors can help to predict whether the fire would be widely spread (in area). The Logistic Regression also illustrate how to protect fire safety, for example, we should install equipment that prevent severe fire incidents such as sprinkler system and smoke alarm inside the buildings and ensure their capabilities of being activated in time.

Understanding these trends can help inform targeted fire prevention and mitigation strategies, ultimately contributing to the enhancement of public safety and the reduction of fire-related damages in Toronto.