Fire Incidents Project Report

INT353 – CA3

Name: Mansi Singh
Registration Number: 12018064
Section: K20CH
Instructor: Mr. Abhijeet Dutta
Dataset Link:
https://www.kaggle.com/datasets/reihanenamdari/fire-incidents

GitHub Link:
https://github.com/singhmansi25/Fire-Incidents-Analysis-Project

Project Introduction & About the dataset

The dataset includes Fire incidents as defined by the Ontario Fire Marshal related to Toronto, having information starting from 2011 to 2018.

Toronto Fire Services (TFS) provides fire protection, technical rescue services, hazardous materials response, and first responder emergency medical assistance in Toronto, Ontario, Canada. Toronto Fire Services is currently the largest municipal fire department in Canada.

This dataset has some interesting columns such as Estimated Dollar Loss, Latitude and Longitude, Exact time, and date etc.

I will use this dataset for Exploratory Data Analysis (EDA) and visualization along with preparing a Tableau Dashboard.

- About Dataset:

Number of columns: 27

Number of rows: 11214

Project Domain

The Fire Incidents dataset belongs to the Fire Safety & Emergency Management domain.

Reason for selecting this dataset

Fire departments play an essential role in ensuring community safety. Beyond their obvious duties of responding to fires, these departments are often the first responders at any kind of emergency. They also play a critical role in fire and accident prevention and safety. Developing the right performance measures, and then tracking those measures consistently over time, allows fire departments to evaluate and improve the quality and effectiveness of their services in all these areas.

Key Performance Indicators

S.No.	Objective	Performance Measure
1.	Place of Incident	1. Area of Origin 2. Latitude 3. Longitude
2.	Loss of Life & Injuries	1. Civilian Casualties 2. Count of Persons Rescued 3. Estimated persons displaced
3.	Business Impacted/ Economy Loss	1. Estimated Loss in Dollars 2. Business Impact 3. Property Use
4.	Possible Cause of Fire	1. Possible Cause 2. Ignition Source 3. Material First Ignited 4. Extent of Fire
5.	Fire Alarm System & Smoke Alarm System	1. Fire Alarm system impact on evacuation 2. Fire Alarm system operation 3. Fire Alarm system presence 4. Smoke alarm at fire origin alarm failure 5. Smoke alarm at fire origin alarm type
6.	Status of Fire Fighters arrival	1. Incident station area 2. Incident ward 3. TFS alarm time 4. TFS arrival time 5. Extent of fire 6. Status of fire on arrival 7. Method of Fire control 8. Last clear time

Approach of Analysis

- Data Collection
- Data Cleaning
- Feature Engineering
- Analysis:
    - Univariate Analysis
    - Bivariate Analysis
    - Multi-variate Analysis
- Data Visualizations

Tools & Technology Used

- Python
    - NumPy
    - Pandas
    - Statistics
    - Date-Time
    - Matplotlib
    - Seaborn
- Google Colab

Goals / Plans

1. Which area is more fire-incident prone?
2. What is the casualty rate?
3. How fire alarm and smoke alarm respond at the time of incident?
4. What business are impacted due to fire?
5. What is the loss per year (in dollars) due to fire?
6. What is the source of ignition in most of the cases?
7. How fire department responds to any incident based on notification time and arrival time?

Data Collection

Dataset is imported from Kaggle. The link is provided in the starting of the report.

Data Cleaning

The Dataset has:

Number of Rows: 11214
Number of Columns: 27

Number of columns with Float type variable: 3
Number of columns with Int type variable: 5
Number of columns with Object type variable: 19

But we see that Columns with Date & Time values are in object format.

We change it into date-time format.

1. Converting columns to right format. (Type Conversion)

Converting Timestamp to Date Time format

The timestamps are string values and need to be converted to datetime objects to do feature engineering.

```
[7] # Converting Columns in object type to Date-time format.
    fire_df['Ext_agent_app_or_defer_time'] = pd.to_datetime(fire_df['Ext_agent_app_or_defer_time'])
    fire_df['Fire_Under_Control_Time'] = pd.to_datetime(fire_df['Fire_Under_Control_Time'])
    fire_df['Last_TFS_Unit_Clear_Time'] = pd.to_datetime(fire_df['Last_TFS_Unit_Clear_Time'])
    fire_df['TFS_Alarm_Time'] = pd.to_datetime(fire_df['TFS_Alarm_Time'])
    fire_df['TFS_Arrival_Time'] = pd.to_datetime(fire_df['TFS_Arrival_Time'])
```

The Dataset has:

Number of columns with Float type variable: 3
Number of columns with Int type variable: 5
Number of columns with Datetime type variable: 5
Number of columns with Object type variable: 14

2. Handling Null values

Handling Null Values

As only Column 'Incident Ward' has null values, let's inspect this column.

```
[10] # analysing unique values in the column.
     fire_df['Incident_Ward'].unique()
```

```
array([18.,  7., 27., 20.,  5., 15., 14., 17., 29.,  9.,  8., 42., 22.,
       36., 43.,  1., 28., 33., 24., 35., 37., 30., 34., 11., 10.,  6.,
       21.,  4., 19., 26., 44., 32., 23., 25., 38., 31.,  2., 13., 12.,
       16.,  3., 40., 41., nan, 39.])
```

As Incident Ward column signifies Ward code of the area where the incident occurred, we can replace null values with 0.

```
[11] # replacing null values 0
     fire_df['Incident_Ward'].fillna(0, inplace=True)
```

3. Outliers Handling

```
[15] # creating a dataframe from dictionary
     explore_df = pd.DataFrame(explore_dict, index = ['Civilian Casualties', 'Count of persons rescued',
                                                      'Estimated Dollar Loss', 'Estimated Number of Persons Displaced',
                                                      'Incident Station Area', 'Incident Ward'])

     explore_df.head()
```

	Mean	Median	Min Value	Max Value	Most Common	Least Common	Variance	St. Dev	Skew	Kurtosis	Mean Absolute Deviation
Civilian Casualties	0.112538	0.0	0.0	15.0	0.0	15.0	2.083276e-01	0.456429	8.455166	149.308247	0.205988
Count of persons rescued	0.062154	0.0	0.0	86.0	0.0	86.0	9.265744e-01	0.962587	67.355363	5747.892245	0.120684
Estimated Dollar Loss	42943.693419	2500.0	0.0	50000000.0	0.0	15600.0	2.850884e+11	533936.667426	76.107737	6892.877165	63425.666540
Estimated Number of Persons Displaced	17.274835	0.0	0.0	999.0	0.0	230.0	1.417807e+04	119.071696	7.929017	61.853362	31.293370
Incident Station Area	288.281880	314.0	111.0	445.0	426.0	346.0	1.155969e+04	107.515995	-0.154630	-1.215131	94.124758

The above statistics show that most values in civilian casualties, persons rescued, dollar loss and persons displaced fall close to 0.

The maximum values are much higher than the mean and median indicating a lot of outliers.

The data is already relatively clean however some cleaning is required:

- The categorical columns have far too many categories and many of these categories have too few observations.
- Outliers need to be removed

4. Analyzing Categorical columns:

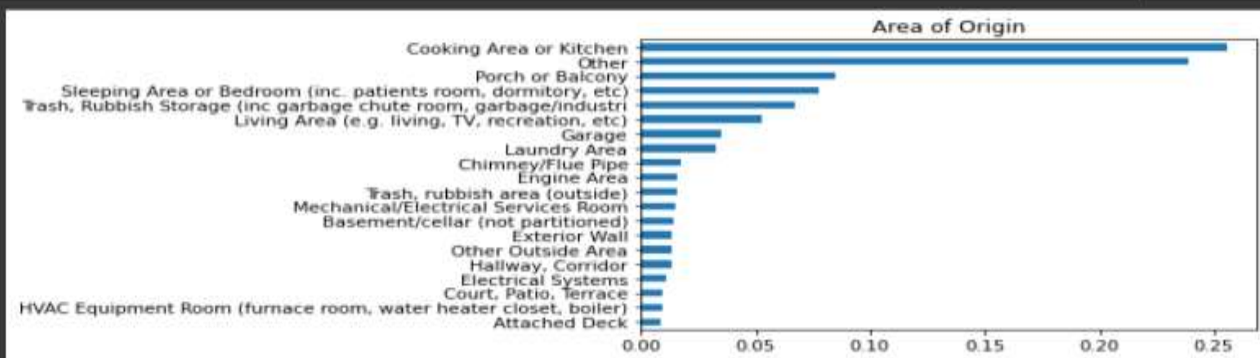
1. Let's analyze Place of origin where fire incident took place and which property got affected.

As 'Area of Origin' has too many categories. Let's find those which are important for analysis.

There is a small drop off in values after 'Attached Deck' where the remaining categories have at least close to 100 observations, this cutoff is arbitrary, and a better cutoff might be possible with further analysis.

I will group all categories with less observations than 'Attached Deck' as well as the 'Other - unclassified' category in a new 'Other' category.

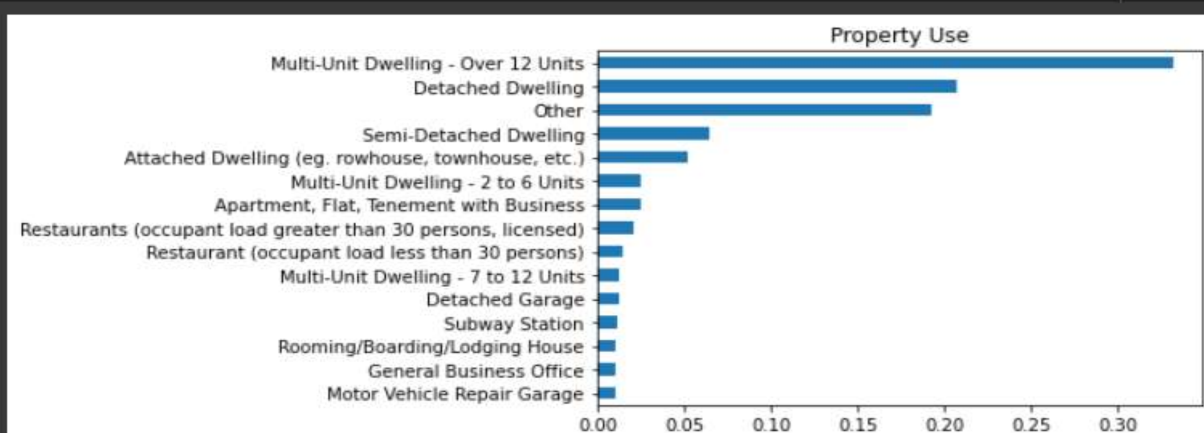
```
# Let's plot a graph for it.
plt.title('Area of Origin')
fire_df['Area_of_Origin'].value_counts( normalize=True, ascending=True).plot(kind='barh')
plt.show()
```



The column has unique values with the most common value being '**Cooking area or Kitchen**' representing 25.6% of entries.

2. The column 'Property Use' has too many categories, let's analyze in detail.

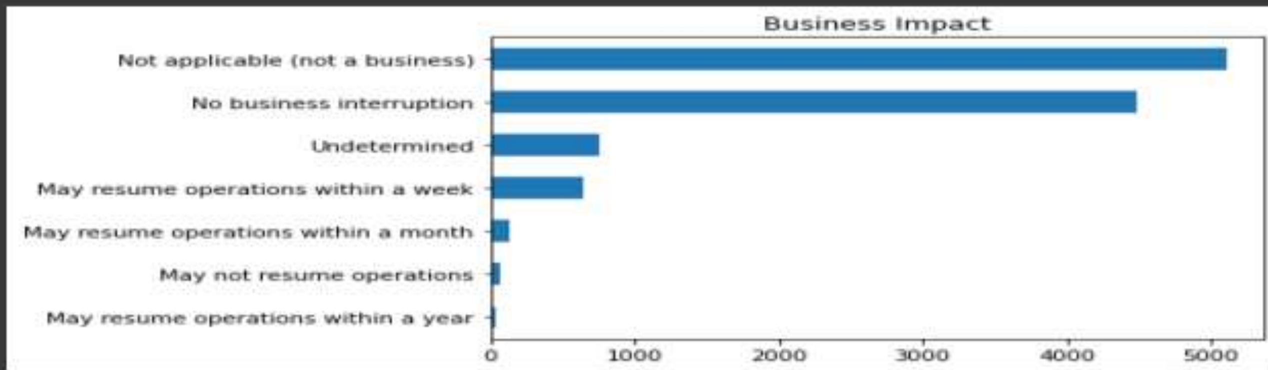
```
# Let's plot a graph for it.
plt.title('Property Use')
fire_df['Property_Use'].value_counts(normalize=True, ascending=True).plot(kind='barh')
plt.show()
```



This describes the use of property that had an incident with most common being '**Multi-unit dwelling**' with 33% while '**Detached dwelling**' consist of 20%.

3. Business Affected due to Fire Incident

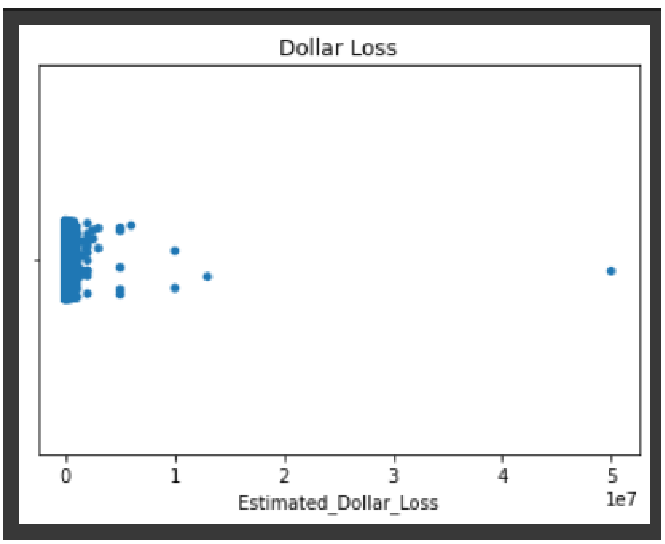
```
# Now we will plot a graph for top values for it.
plt.title('Business Impact')
fire_df['Business_Impact'].value_counts(ascending=True).plot(kind='barh')
plt.show()
```

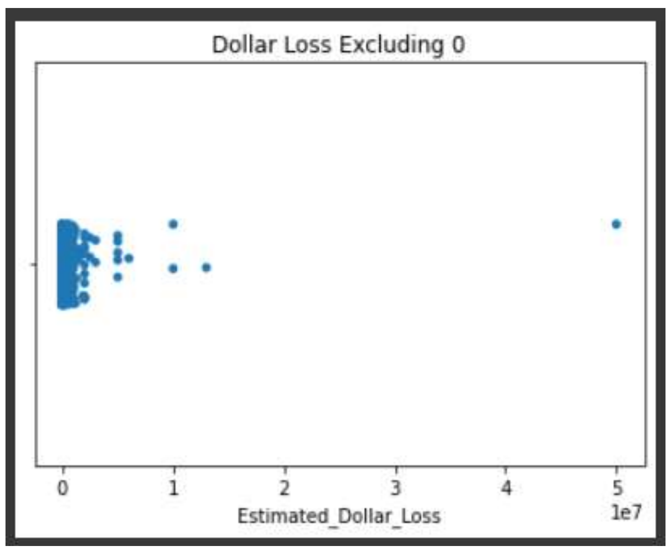


'**Not applicable (no business)**' and '**No business interruption**' represents 86% of entries.

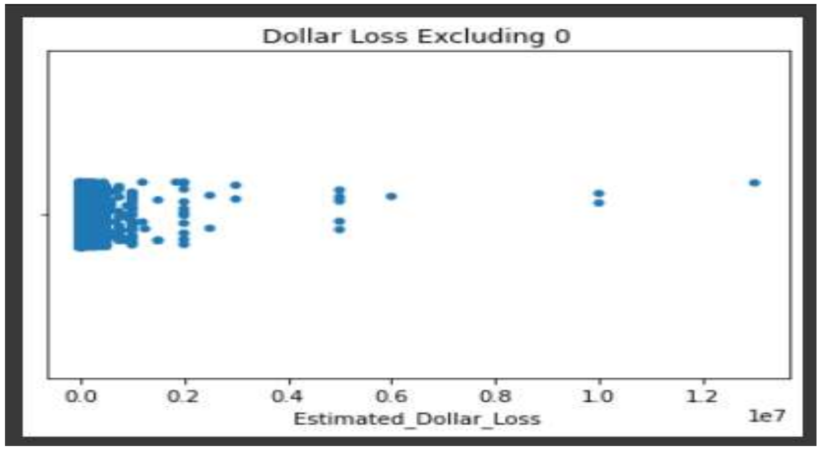
That is business was not impacted in 86% of the fire cases.

4. Dollar Lost due to fire incidents



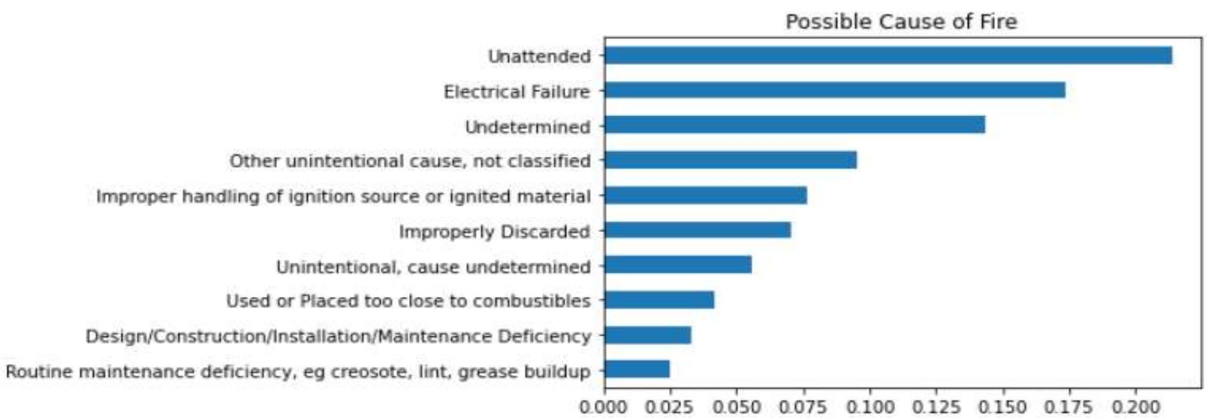


Since most data has a value of 0 the plots don't show the values very well. To show all the values the following plots exclude all 0 values. The dataset has one massive outlier with 50 Million Dollar.

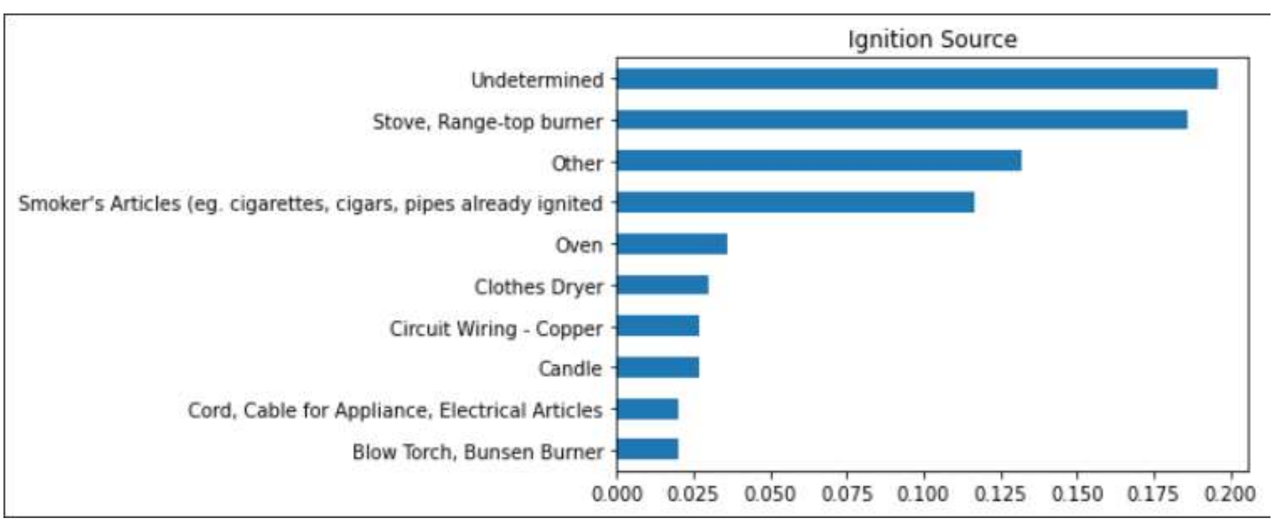


These strange values are removed as well as one massive outlier in the dollar loss column. Now, the range of Estimated dollar loss is greater than 0 and less than 5 Million.

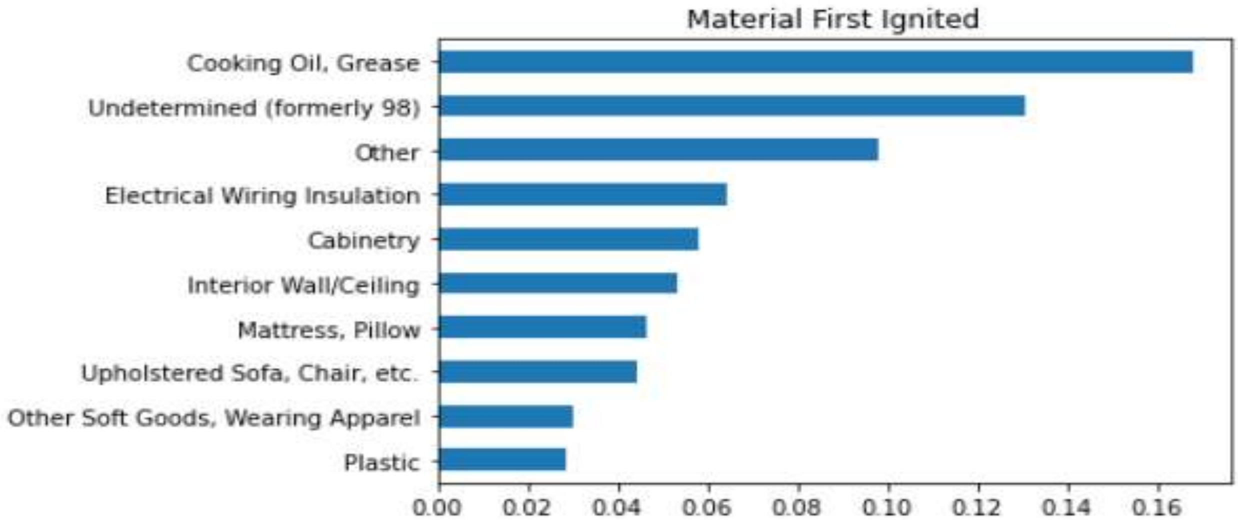
5. Let's analyze **Possible causes of Fire, ignition source, material that caused ignition, and extent of fire**.



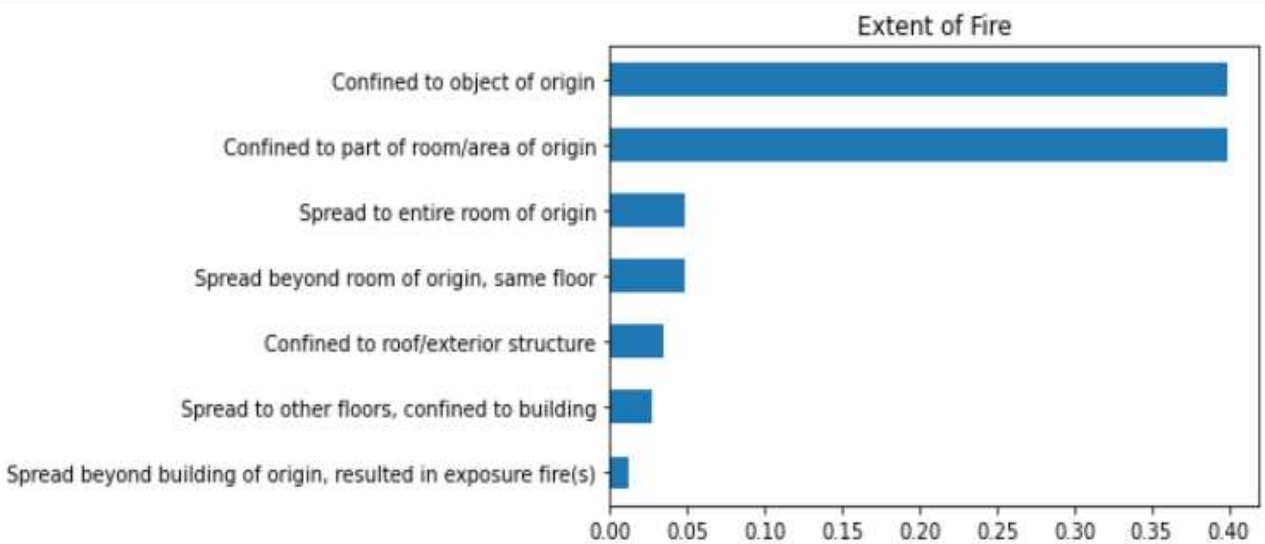
This shows in 21.4% of the cases, the cause of fire is undetermined, while Electrical failure is next highest cause for fire with 17%.



This shows Ignition source is Undetermined in 22% cases, while its Stove/ Burner in 17% cases. We also see that materials that can cause fire are Smoking things, oven, dryer, electrical appliances.

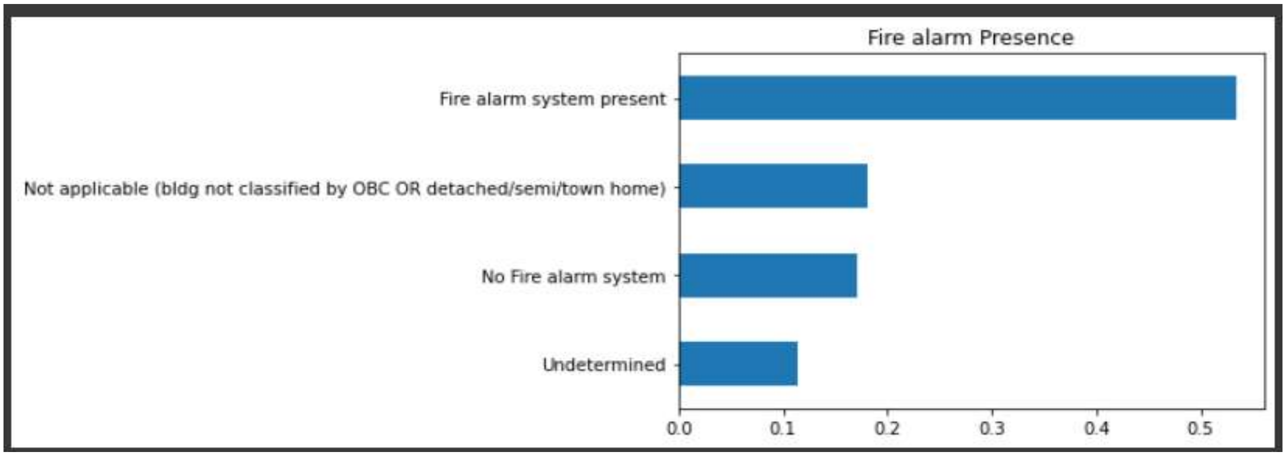


We get to know that Cooking oil, wastes, electrical wirings are some materials that get ignited first.

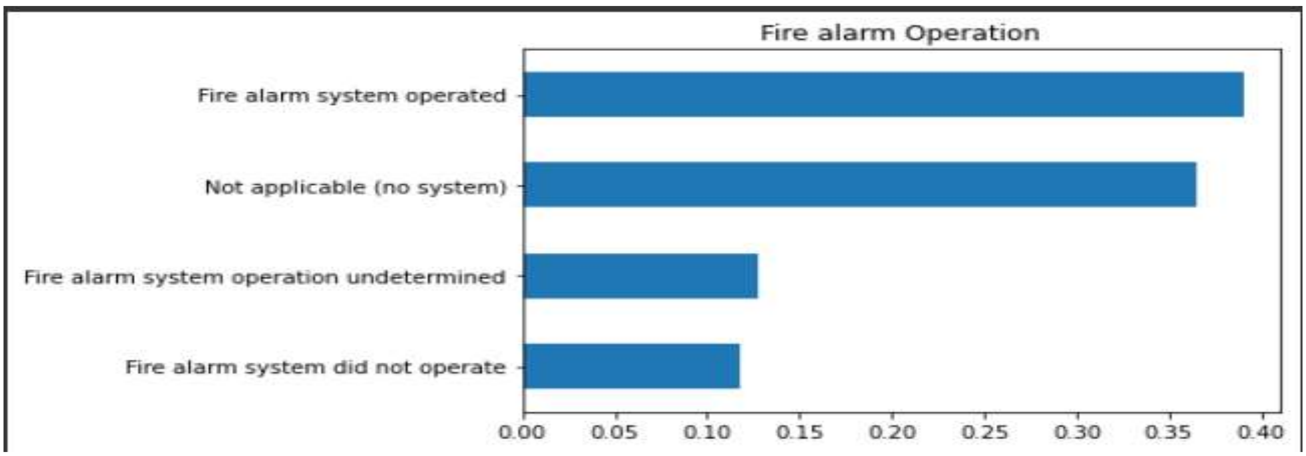


We see that '**confined to object/room of origin**' make up 82% of values. This means the extent of fire is confined to the area of origin most of the times.

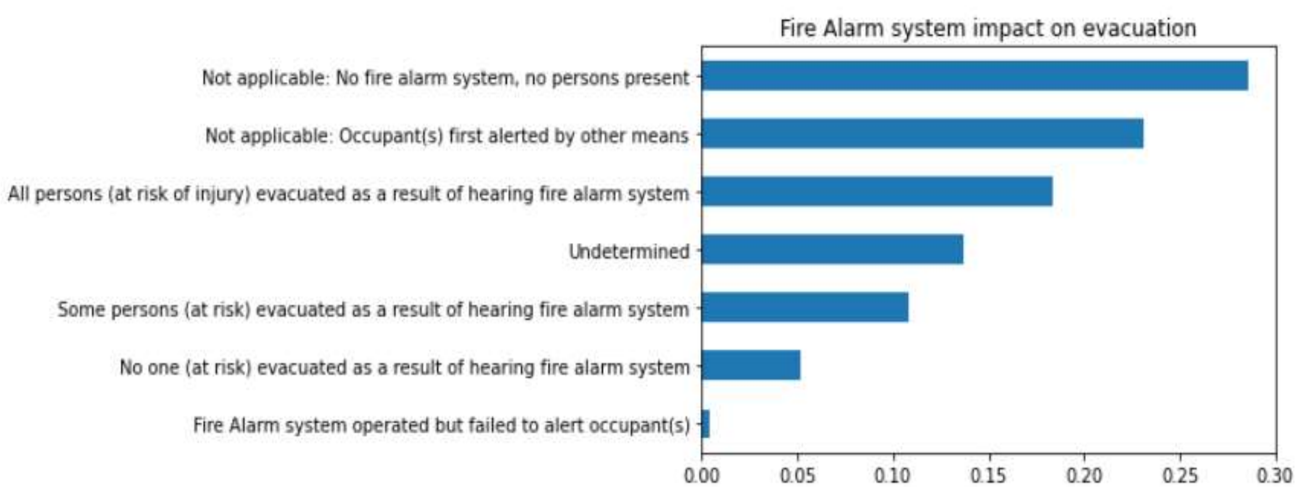
6. Let's Analyses the Fire Alarm System Condition before incident



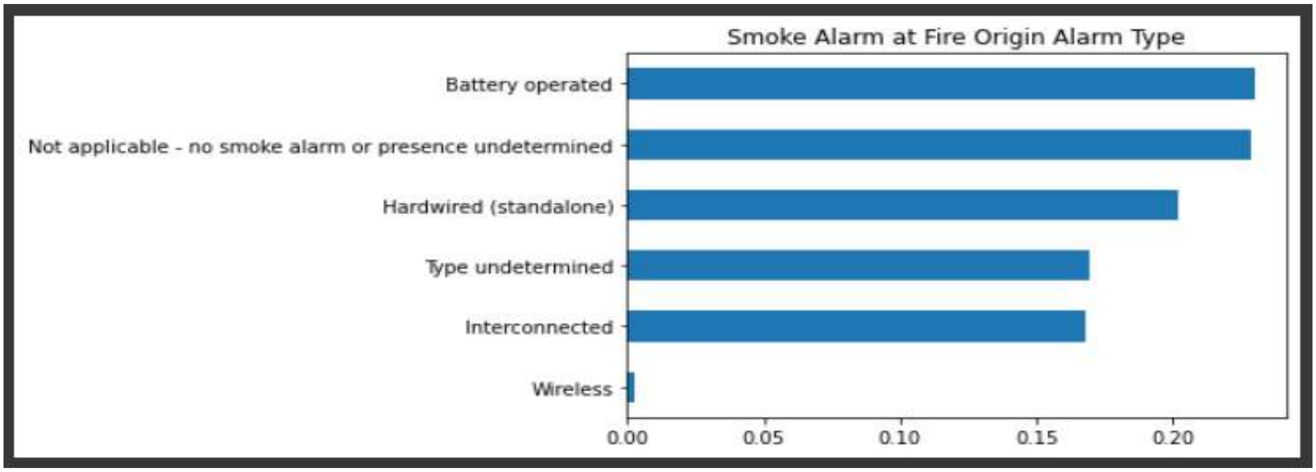
This shows 55% of the Fire alarm was present, however 43% of the times, it wasn't present.



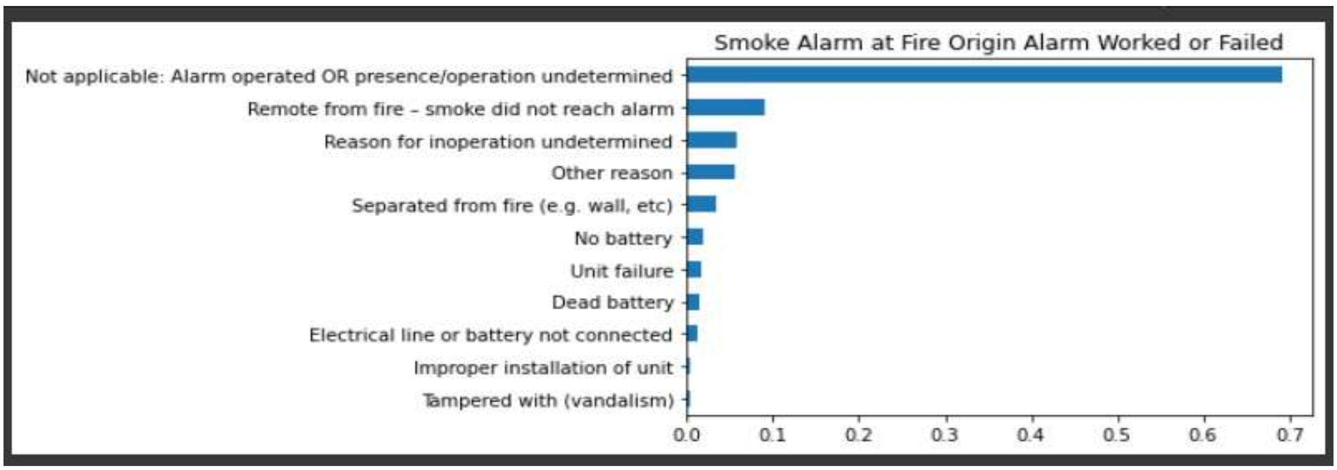
This shows Fire alarm system was present & operated for 40% of the cases. While it didn't operate for 12% of the time.
Fire alarm wasn't present in 35% of the cases.



This shows in 50% of the fire incidents, fire alarm wasn't present.

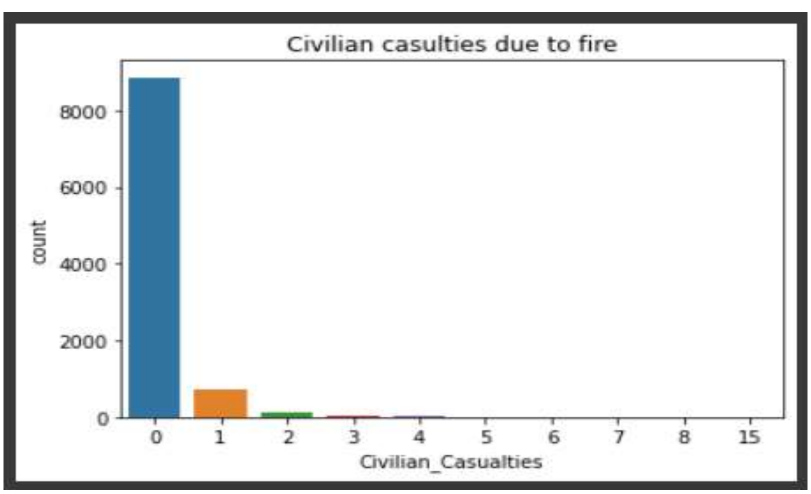


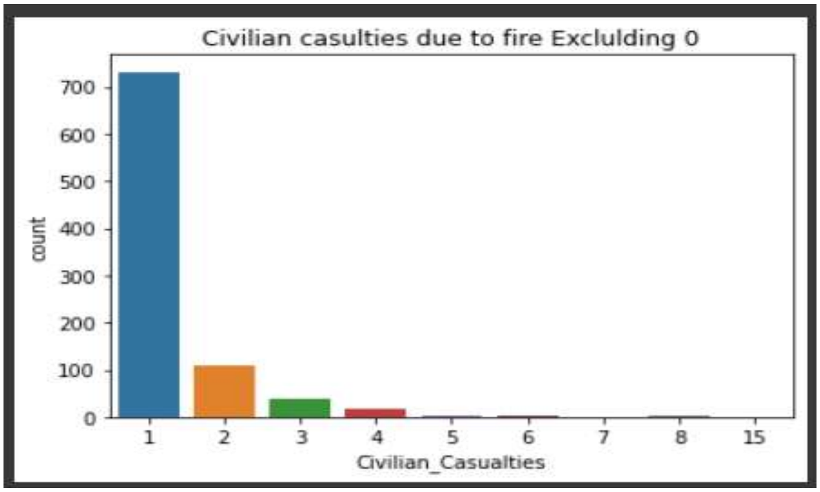
By this we get to know, that 23% of times the Smoke Alarm was Battery operated while 22.85% times it wasn't not present.



This shows 69% of the time alarm operation was undetermined.

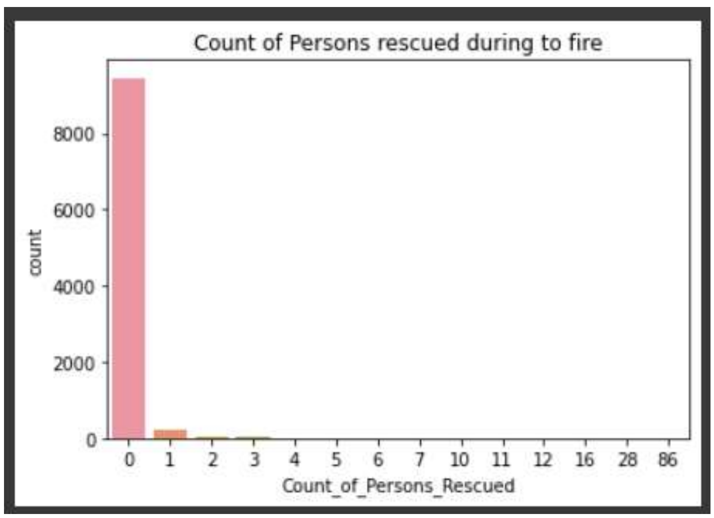
7. Let's analyze Loss of Life or Injuries caused due to Fire incidents.

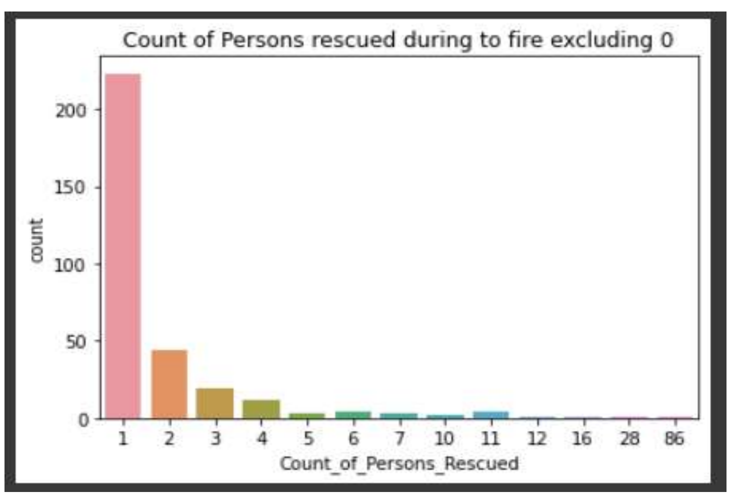




Since most data has a value of 0 the plots don't show the values very well.

To show all the values the following plots exclude all 0 values.

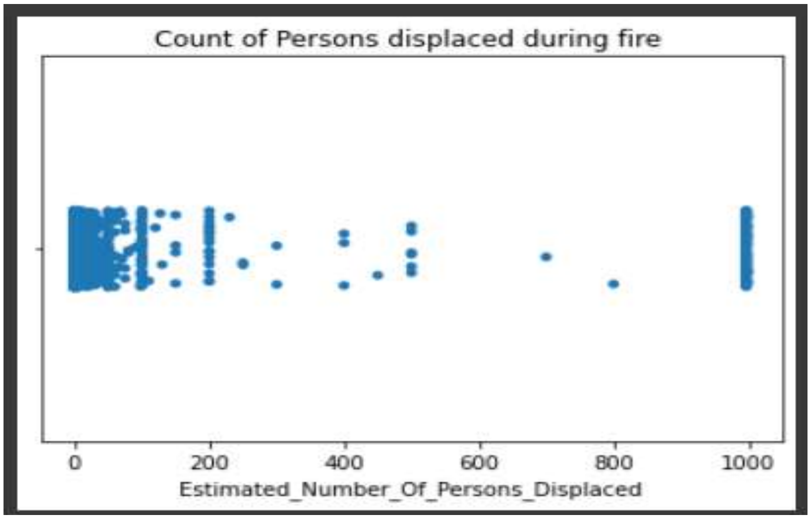




Since most data has a value of 0 the plots don't show the values very well.

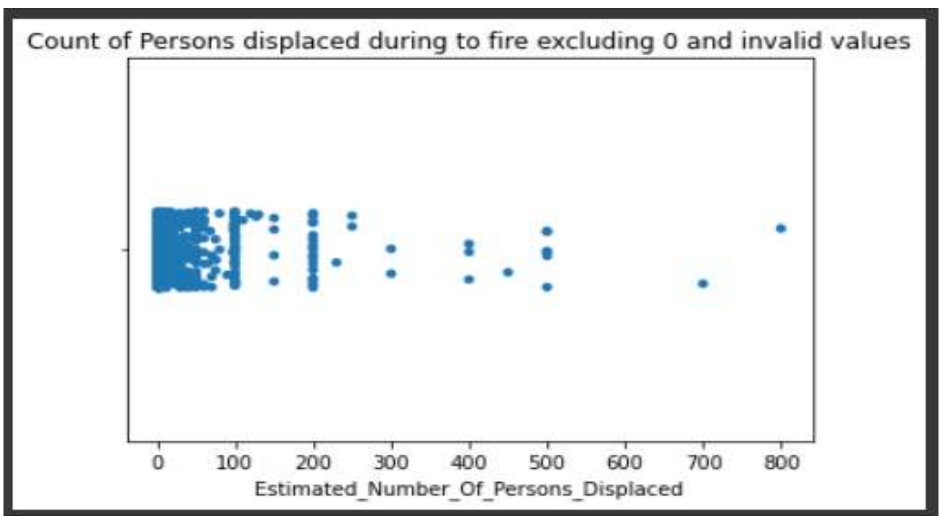
To show all the values the following plots exclude all 0 values.

8. Let's see the persons displaced during fire.



In the persons displaced plot there seems to be a lot of values around 1000 while most other values fall under 200, this is strange.

All these values are exactly 997. there seems to be no relation between civilian casualties, persons rescued, dollar loss and persons displaced when persons displaced is 997, this value appears to be a placeholder or unknown category.

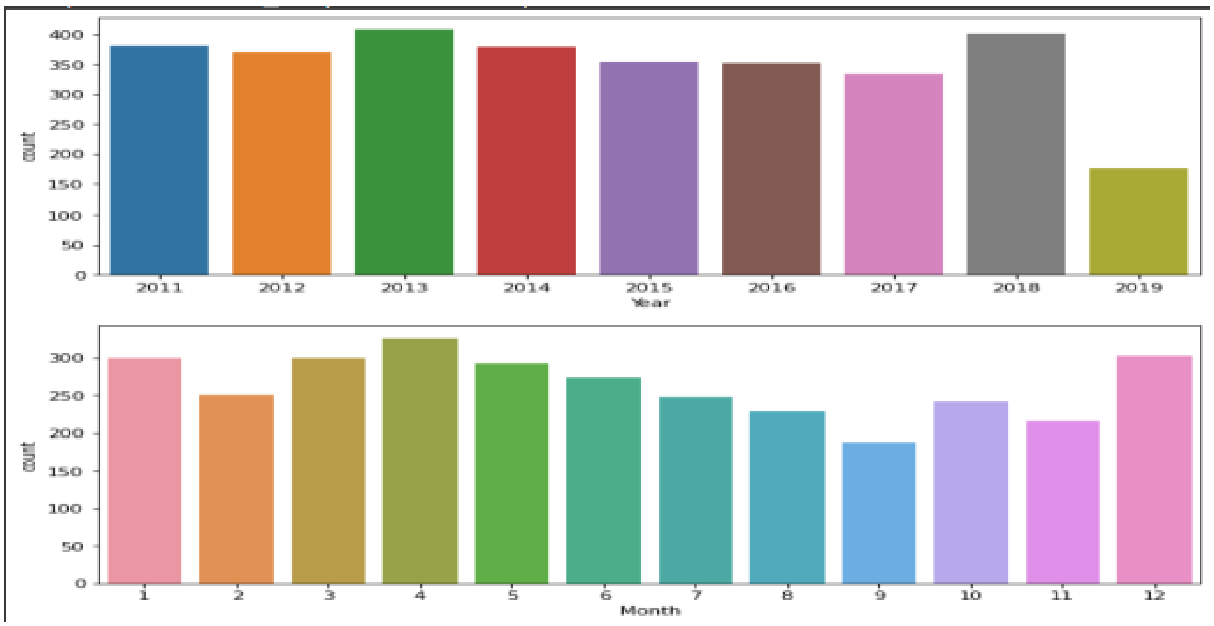


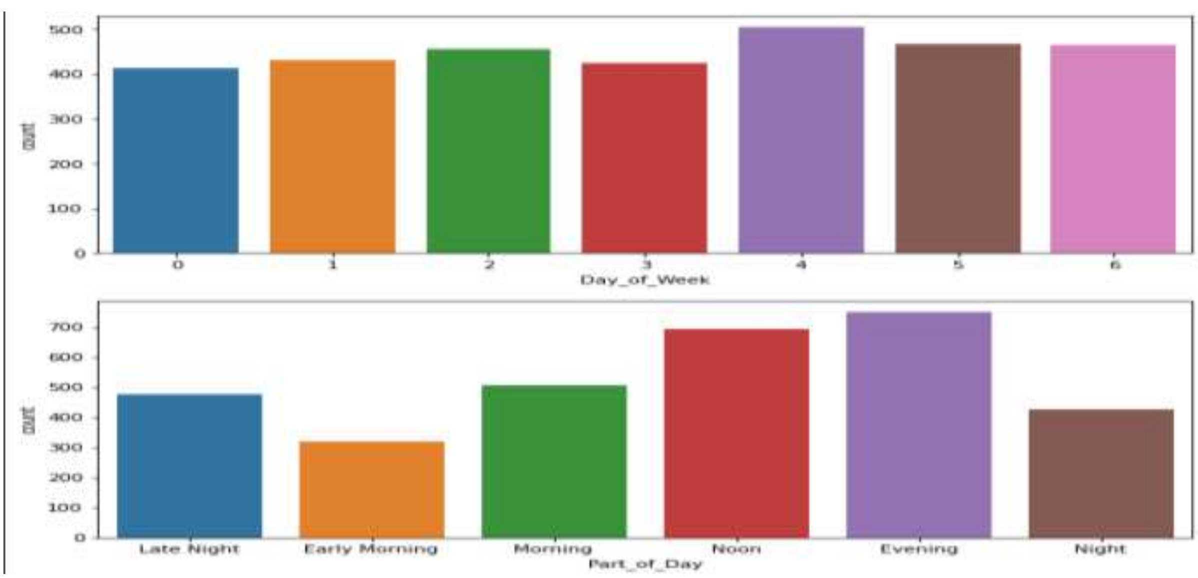
This shows the spread of number of persons displaced during fire incident.

Feature Engineering

Let's use the timestamp columns to engineer some additional features.

	Ext_agent_app_or_defer_time	Year	Month	Day_of_Week	Part_of_Day	Reaction_time_minutes	Extinguish_time_minutes	Time_spent_minutes
1	2018-11-24 07:19:00	2018	11	5	Early Morning	5.18	4.62	330.97
2	2017-02-09 18:02:13	2017	2	3	Evening	3.70	47.53	347.73
3	2012-10-30 00:52:04	2012	10	1	Late Night	2.95	43.83	3825.55
4	2018-07-08 04:35:00	2018	7	6	Late Night	5.07	24.57	750.22
6	2015-03-07 04:49:35	2015	3	5	Late Night	6.25	14.53	158.77

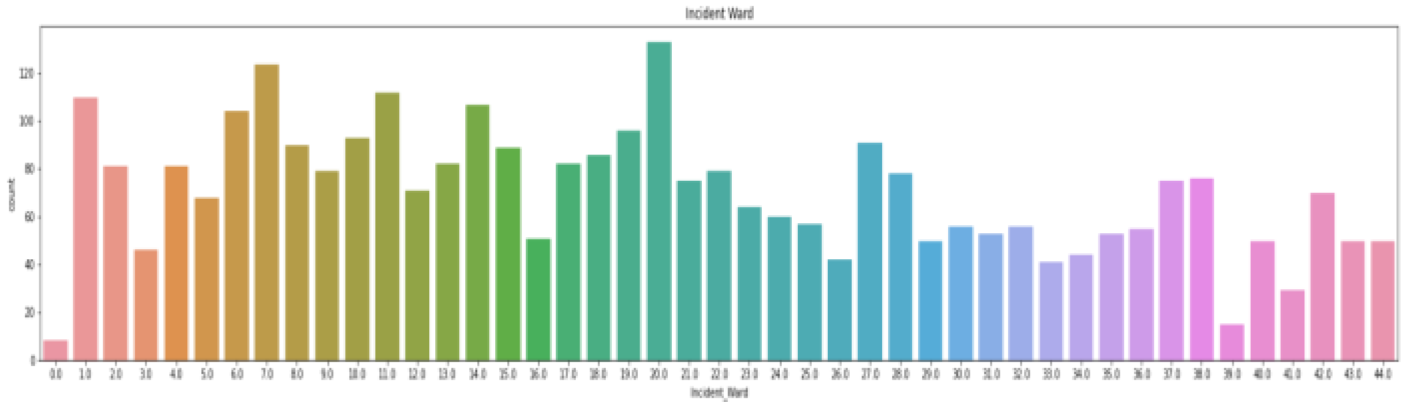




This shows that most incidents occur in May, April, and on Friday in the Noon or Evening.

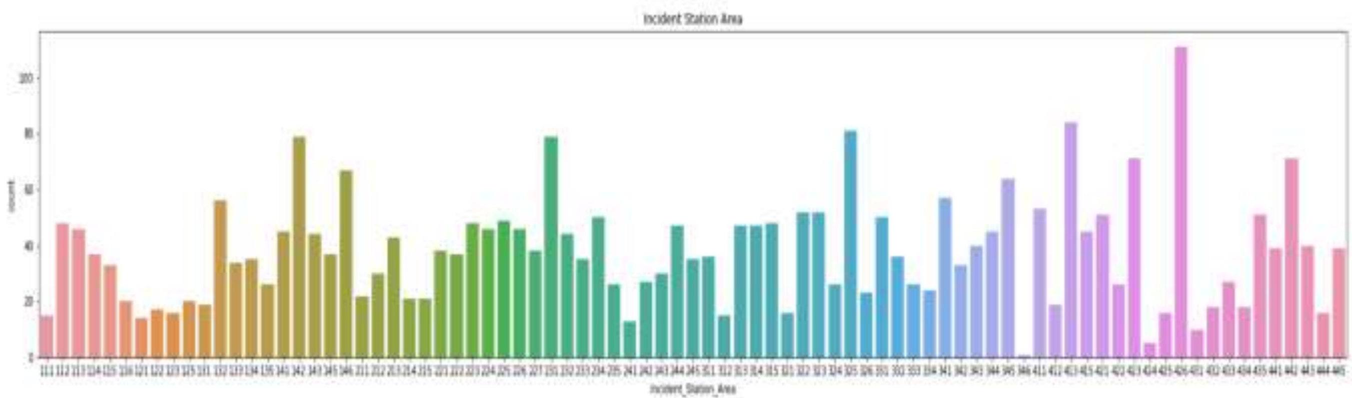
Univariate Analysis

1. Incident Ward

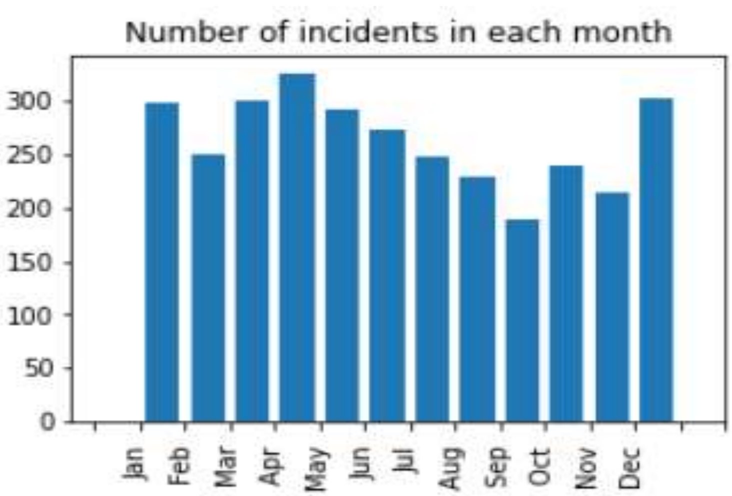


This shows where the incident occurred and its Incident ward. Range is 0 to 44 (0 = Null)

2. Incident Station Area

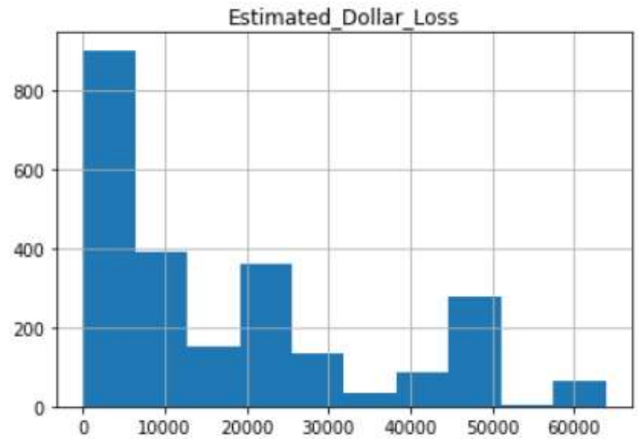


3. Number of incidents month wise



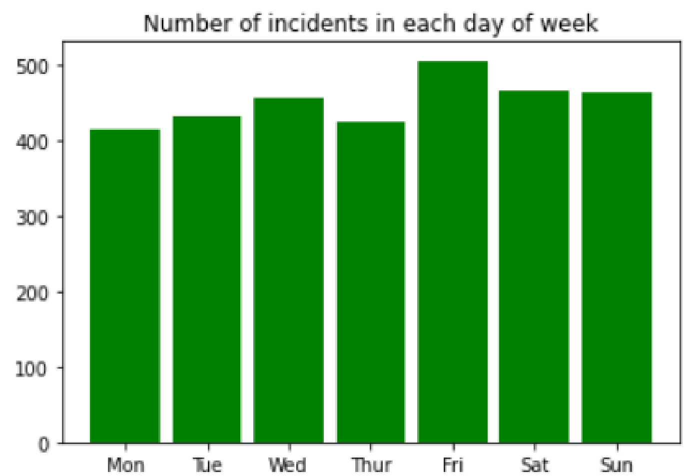
Most of the incidents occurred in month of JAN, MAR, APR, DEC.

4. Estimated Dollar Loss



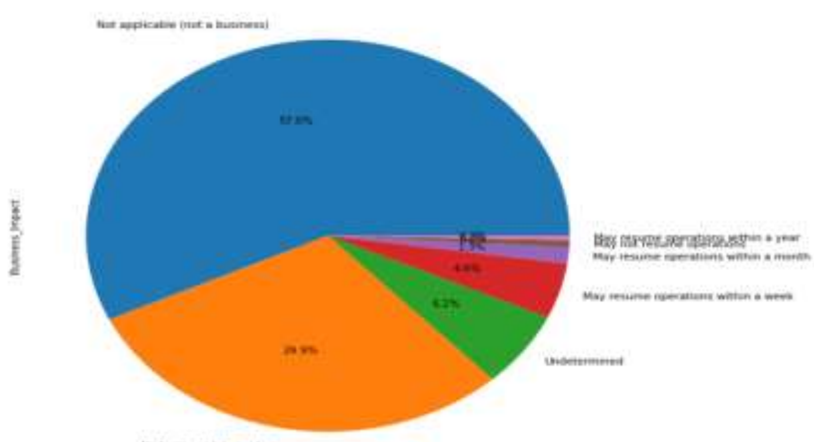
The histogram shows that in most cases the Dollar loss was lesser than 1 Million.

5. Number of incidents in each day of week



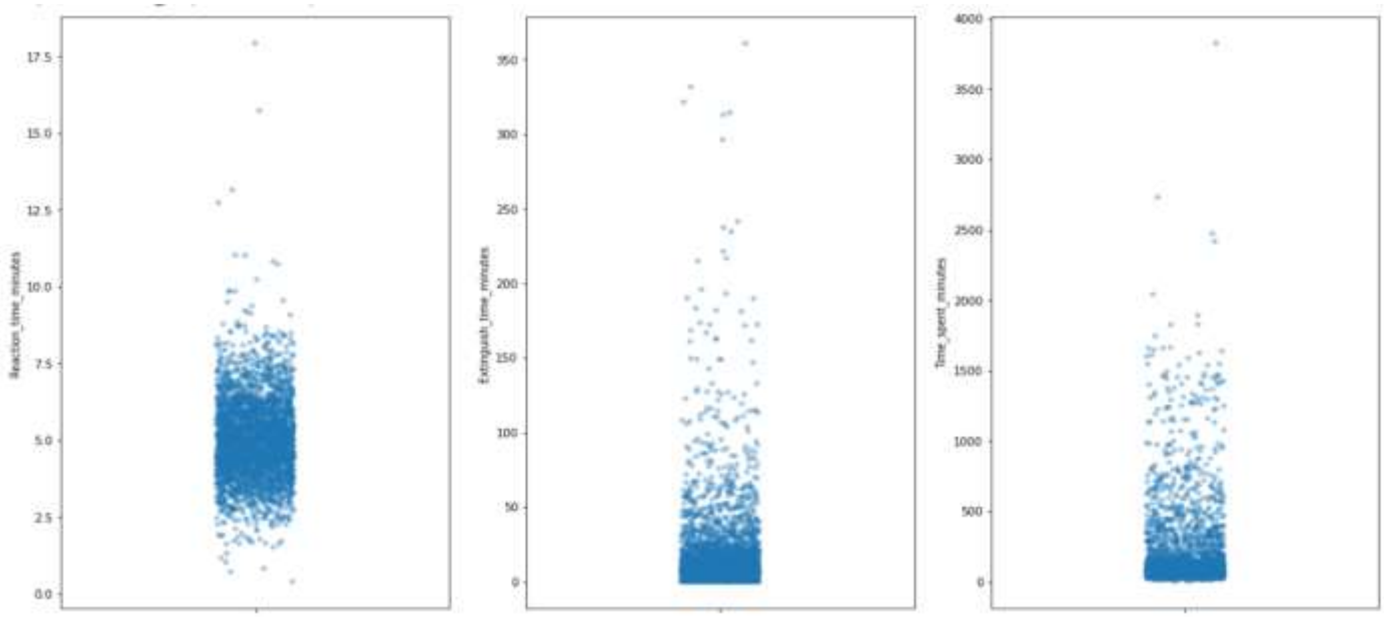
This shows that most incidents occurred on Fridays, followed by Wednesday, Saturday, and Sunday.

6. Business Impact

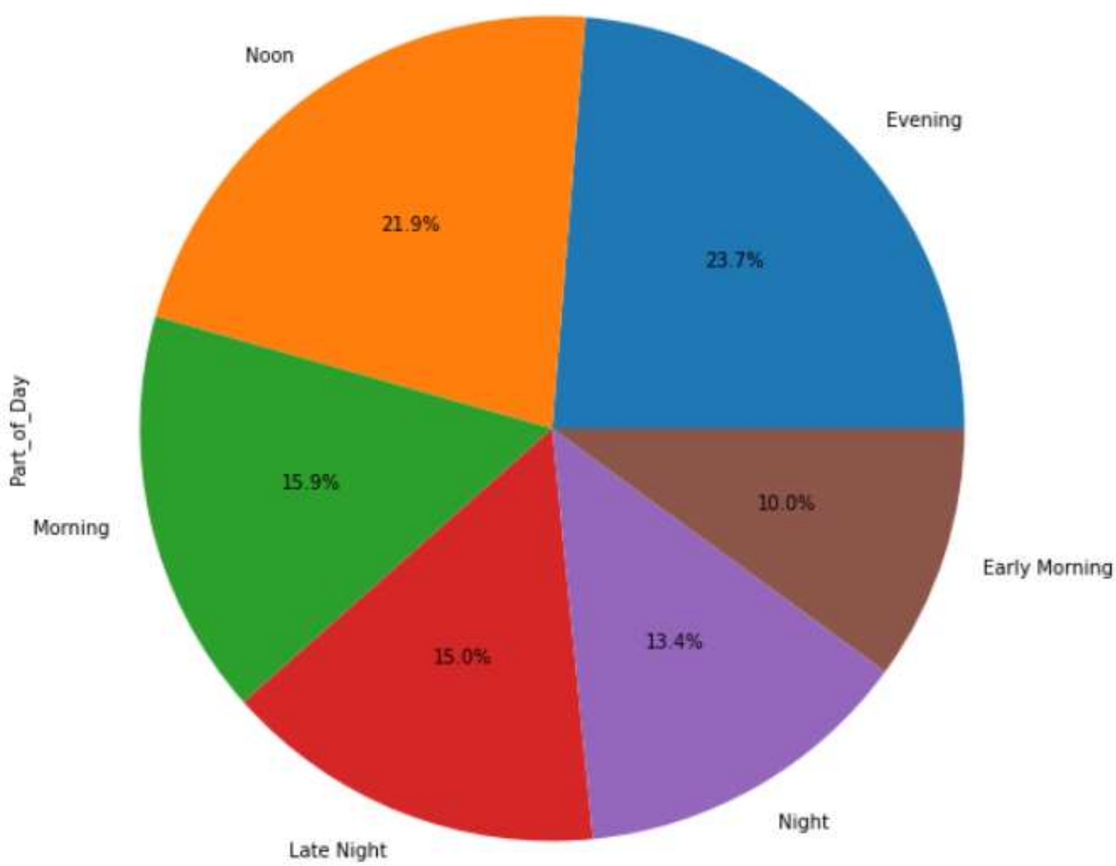


It shows most of fire incidents were not related with any Business.

7. Reaction Time, Extinguish time, Time spent

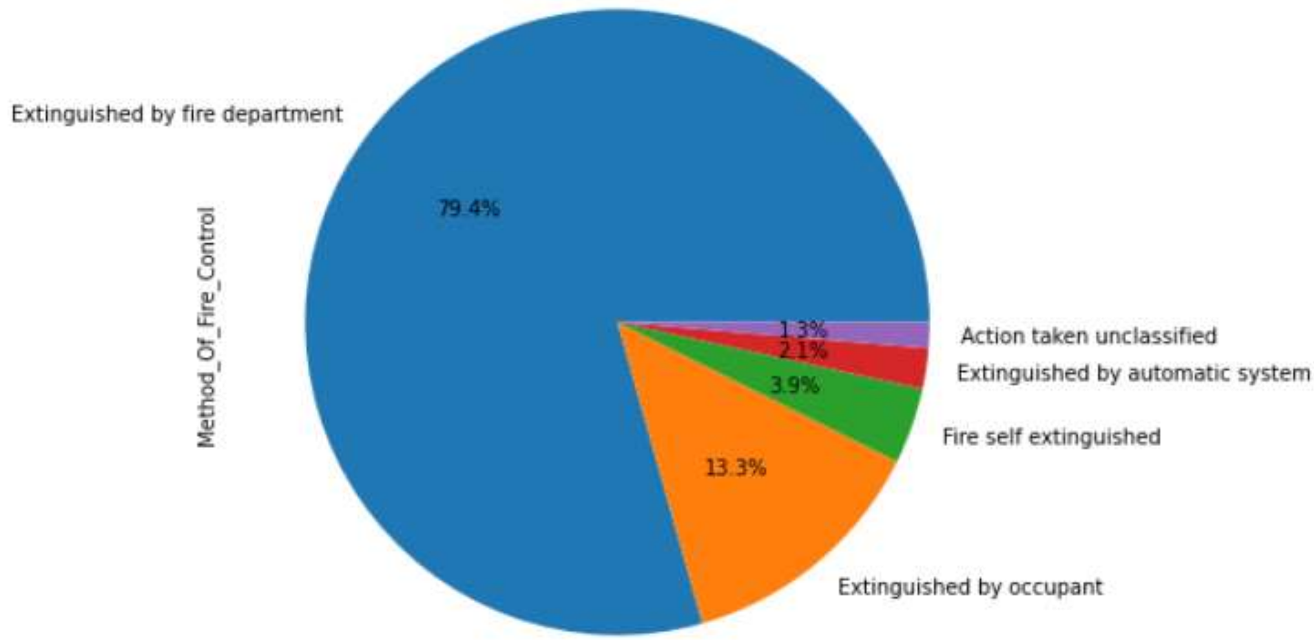


8. Part of Day in pie chart



The pie chart shows that most of the incidents occurred at Evening and Noon.

9. Method of Fire Control

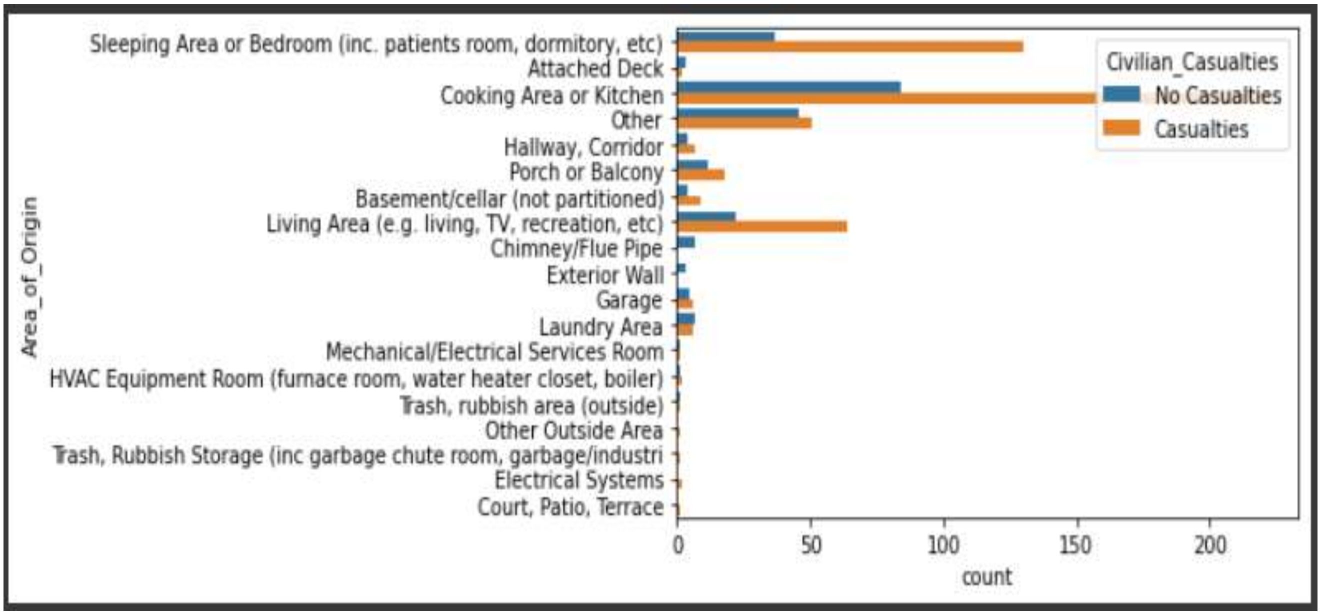


This pie chart shows in 79.4% cases the fire was extinguished by fire department, while in other cases, it was either extinguished by occupant or others.

Bivariate Analysis

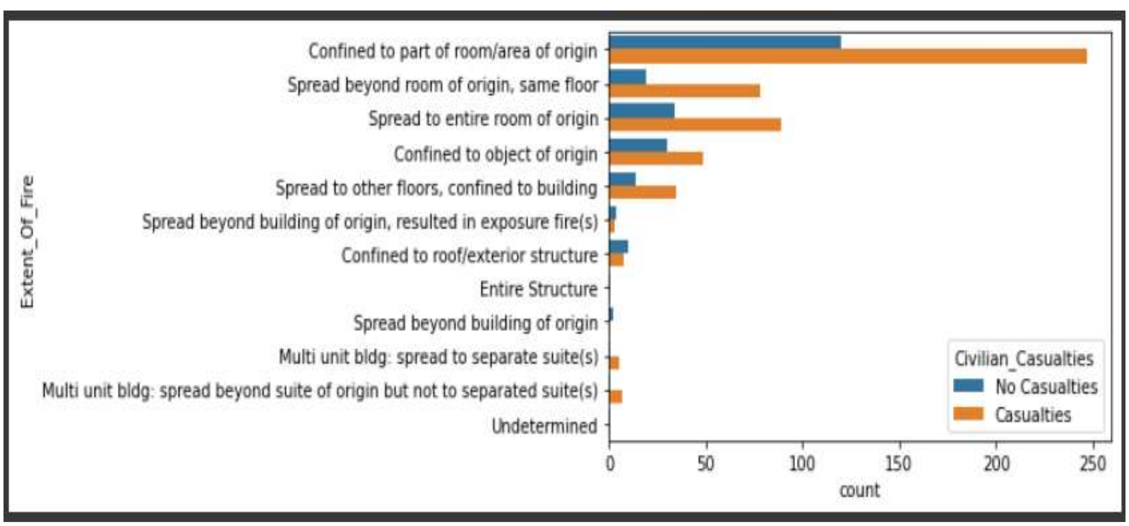
For the data analysis, change the civilian casualties feature to a target column consisting of Casualties or No Casualties, the data analysis will focus on what factors determine whether there will be casualties to help inform firefighters and the public.

1. Area of fire origin vs Casualties



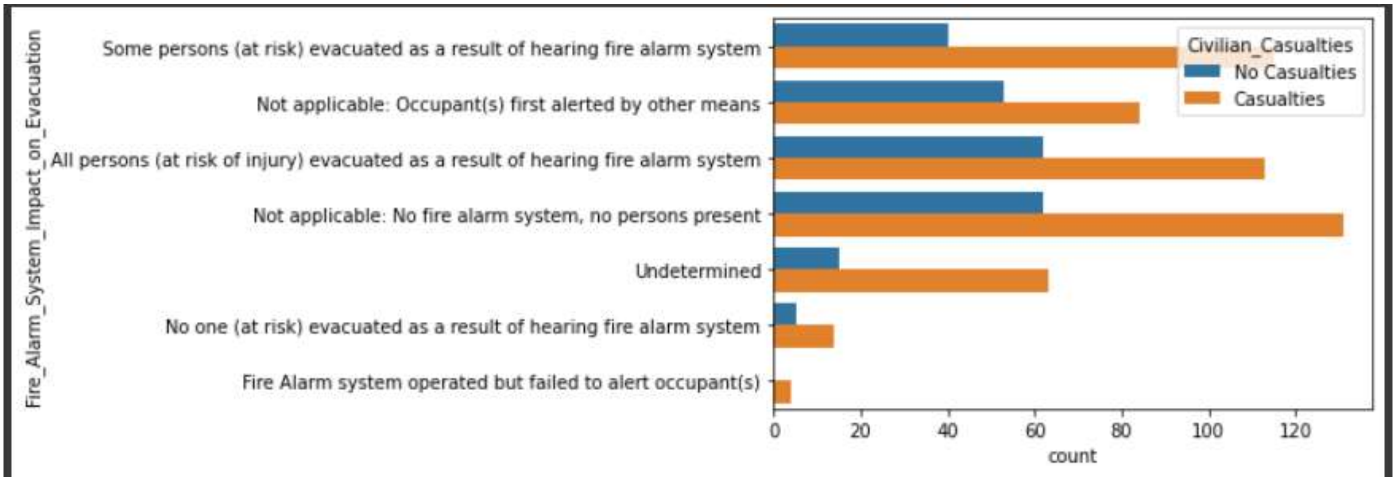
The most dangerous areas are the cooking area and sleeping areas. Obviously, as most of cases take place at evening and noon, and chances o

2. Extent of fire vs Casualties



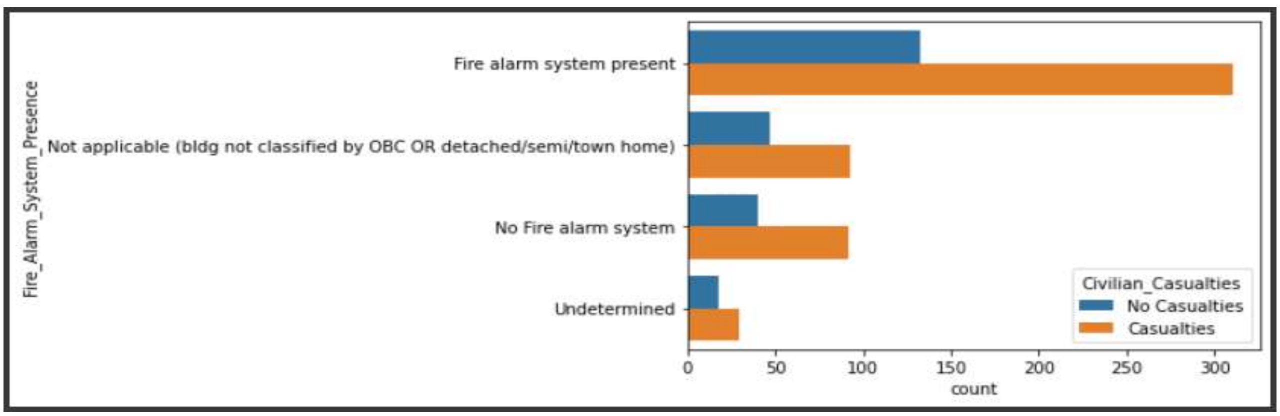
Unsurprisingly incidents that are confined to a smaller area are not as lethal as fires that spread.

3. Fire alarm system impact on evacuation vs Casualties



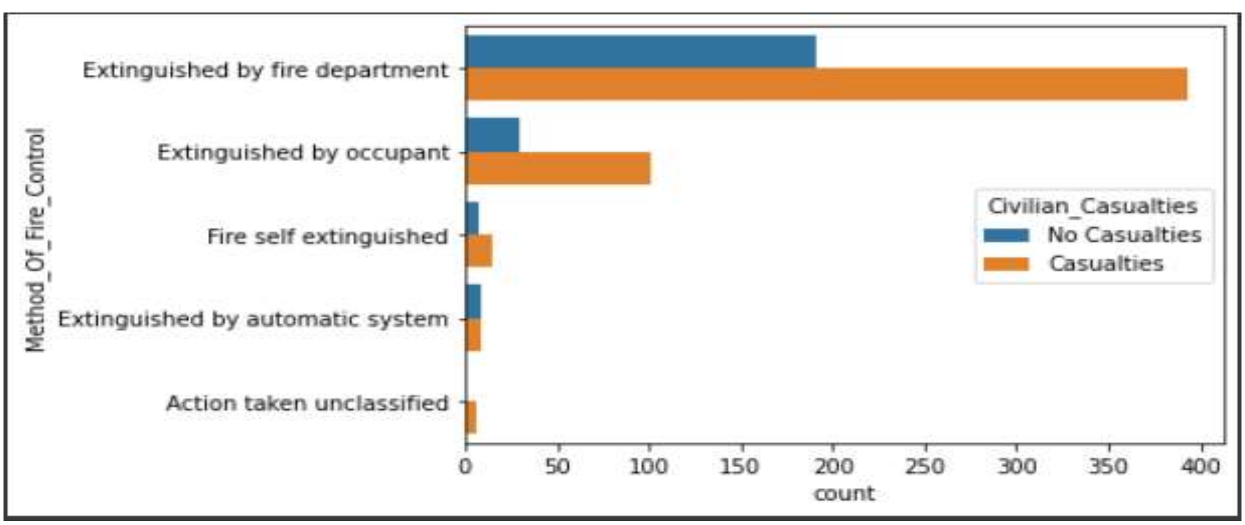
Counterintuitively this plot shows that evacuating people causes more casualties.

4. Fire alarm system presence vs Casualties



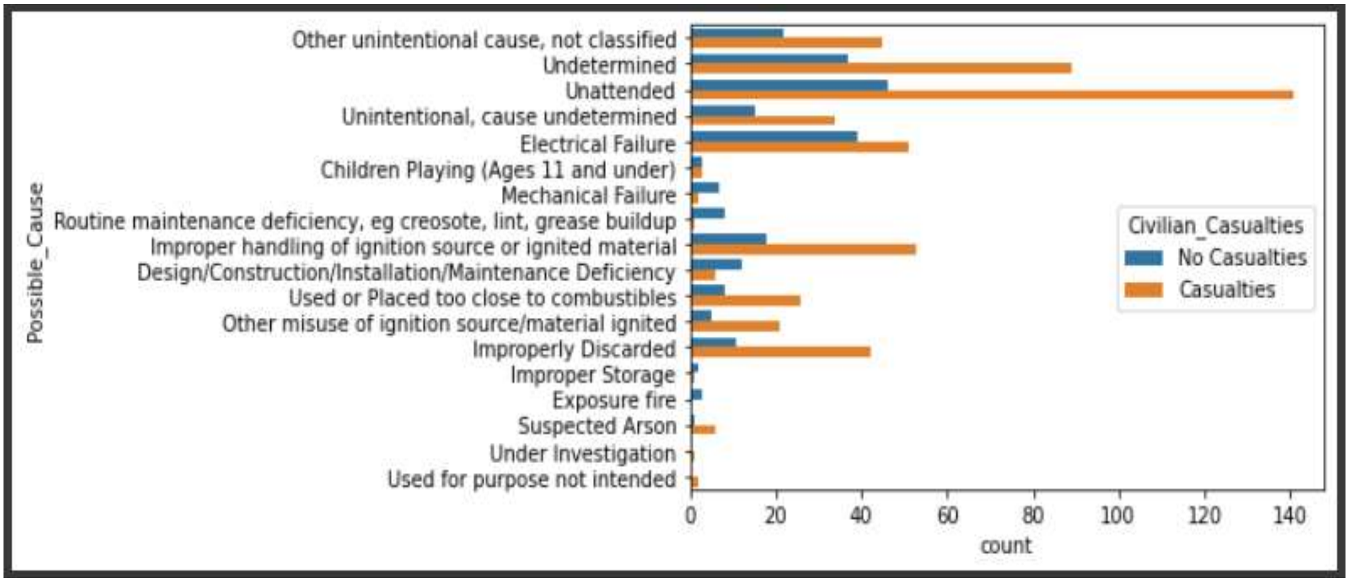
The presence of the fire alarm does not seem to hold statistical significance.

5. Method of fire control vs Casualties



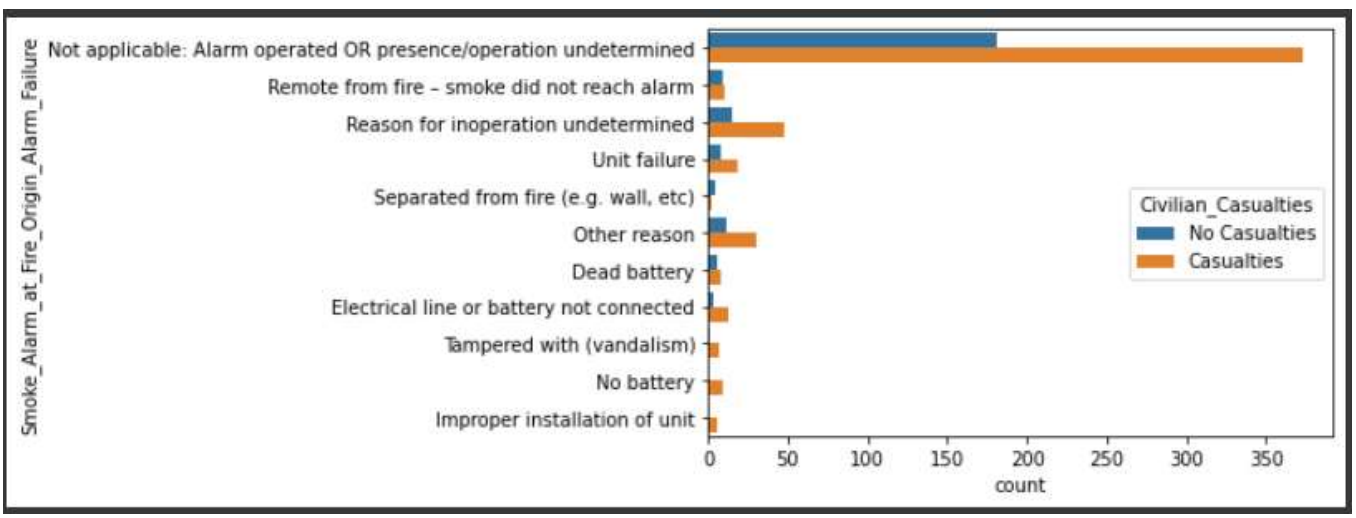
There are more casualties when occupants try to put out the fire themselves. Also, automatic systems reduce casualties a lot.

6. Possible cause vs Casualties



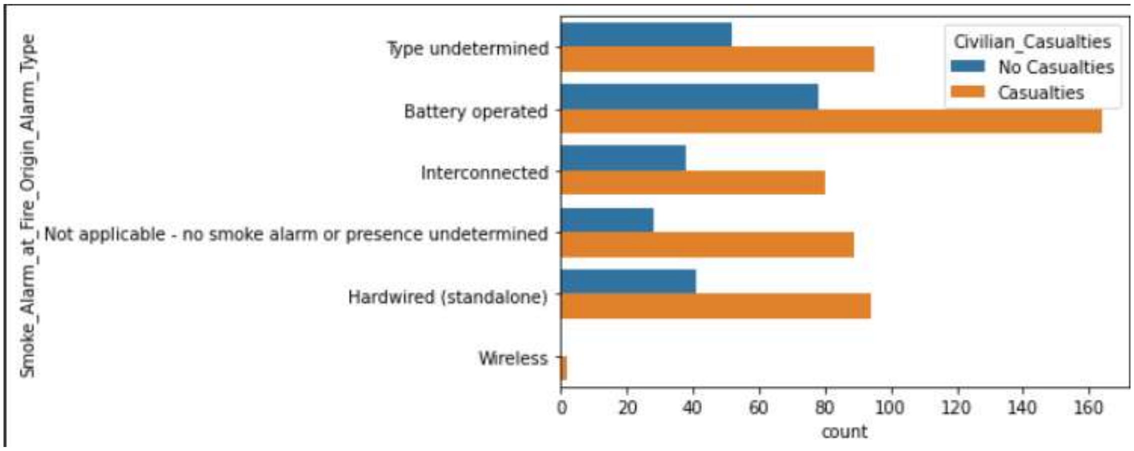
The most common probable causes that result in casualties are not paying attention(unattended) and improper handling of the ignition source.

7. Smoke alarm at Fire origin alarm failure vs Casualties



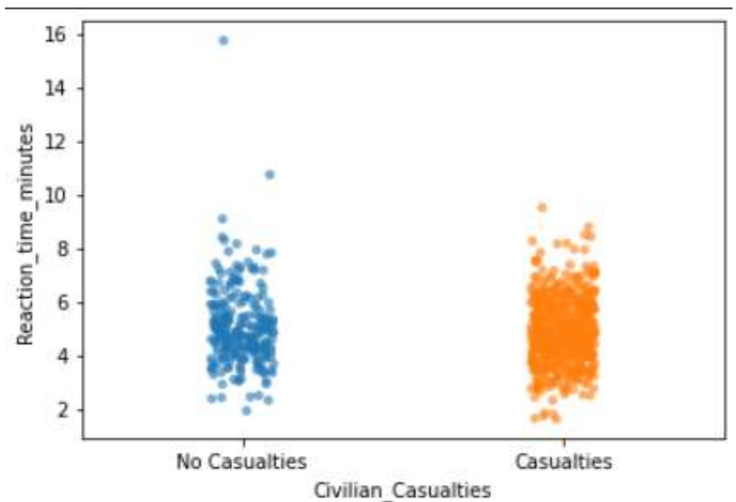
Improper use of a fire alarm is dangerous.

8. Smoke alarm at Fire origin alarm type vs Casualties



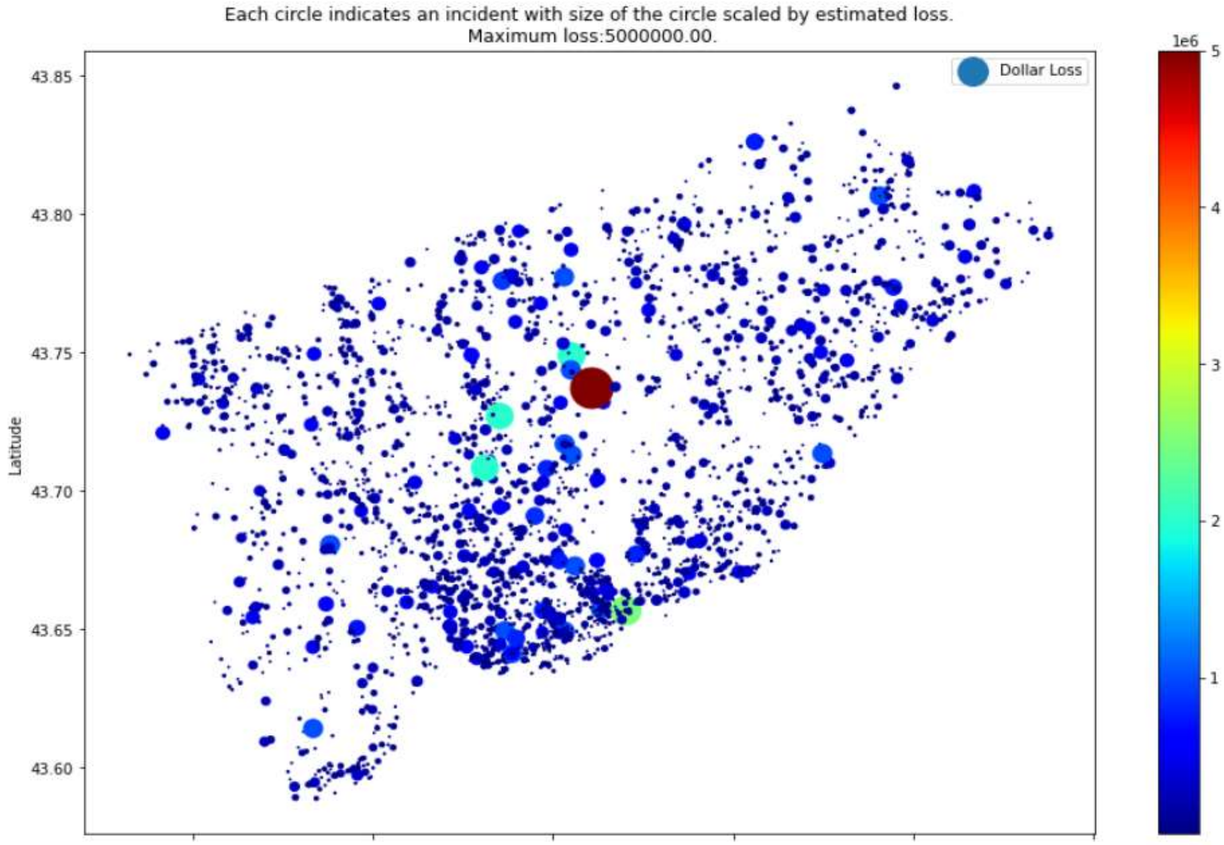
Battery operated fire alarms are more common when there are casualties.

9. Reaction time vs Casualties



Reaction time does not determine casualties.

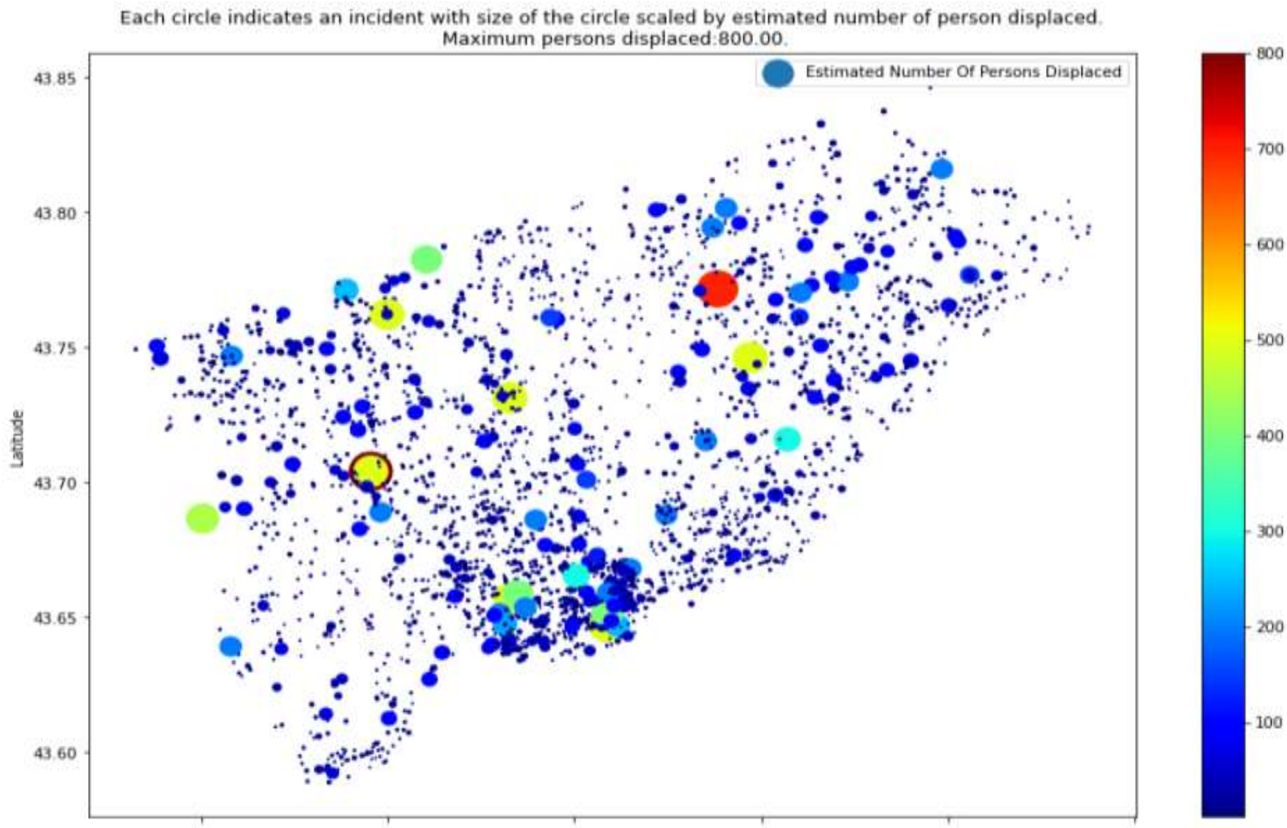
10. Location (Latitude, Longitude) and Dollar loss is available for most of the data. We are going to plot all fire locations and scale the size of each Incident based on Dollar loss.



The graph shows Toronto region with Estimated Dollar Loss. The size and shape of circle shows the amount of Loss.

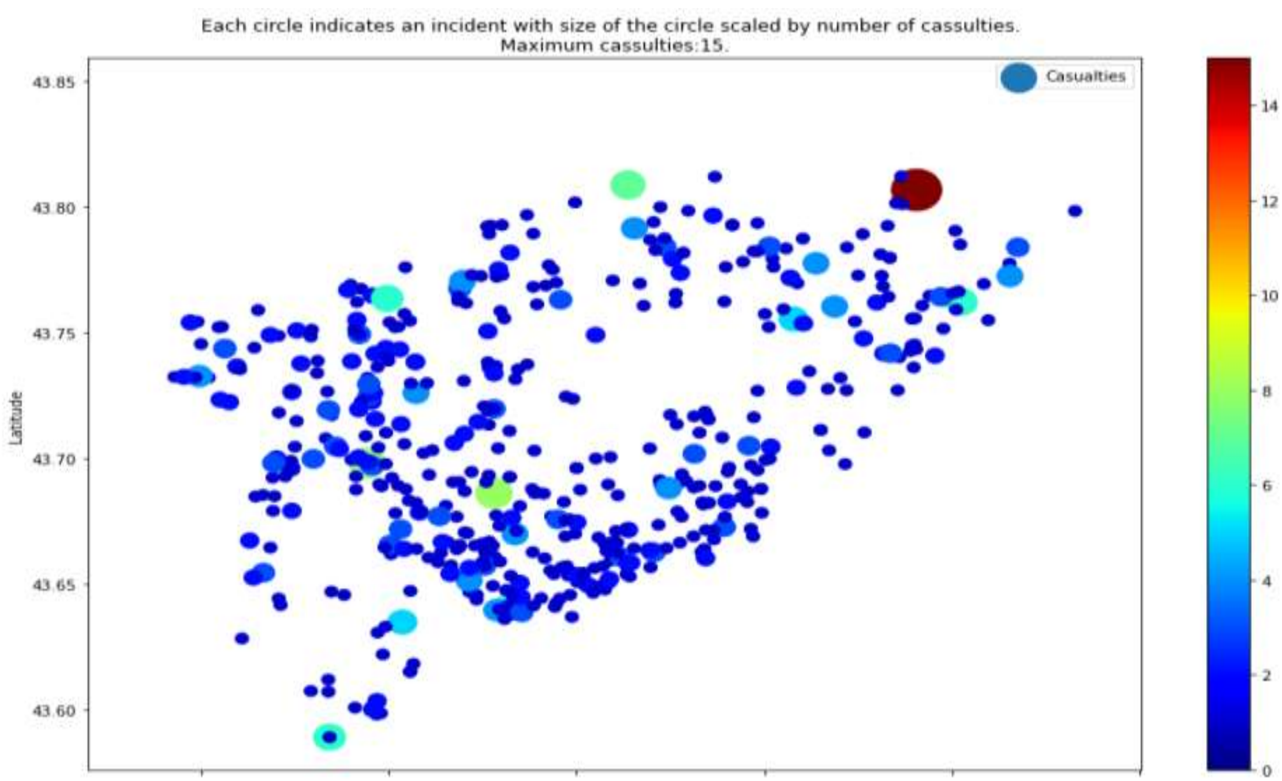
Here we can see a big red dot denoting Maximum Dollar Loss of 5 Million.

11. Estimated number of persons displaced with Location.



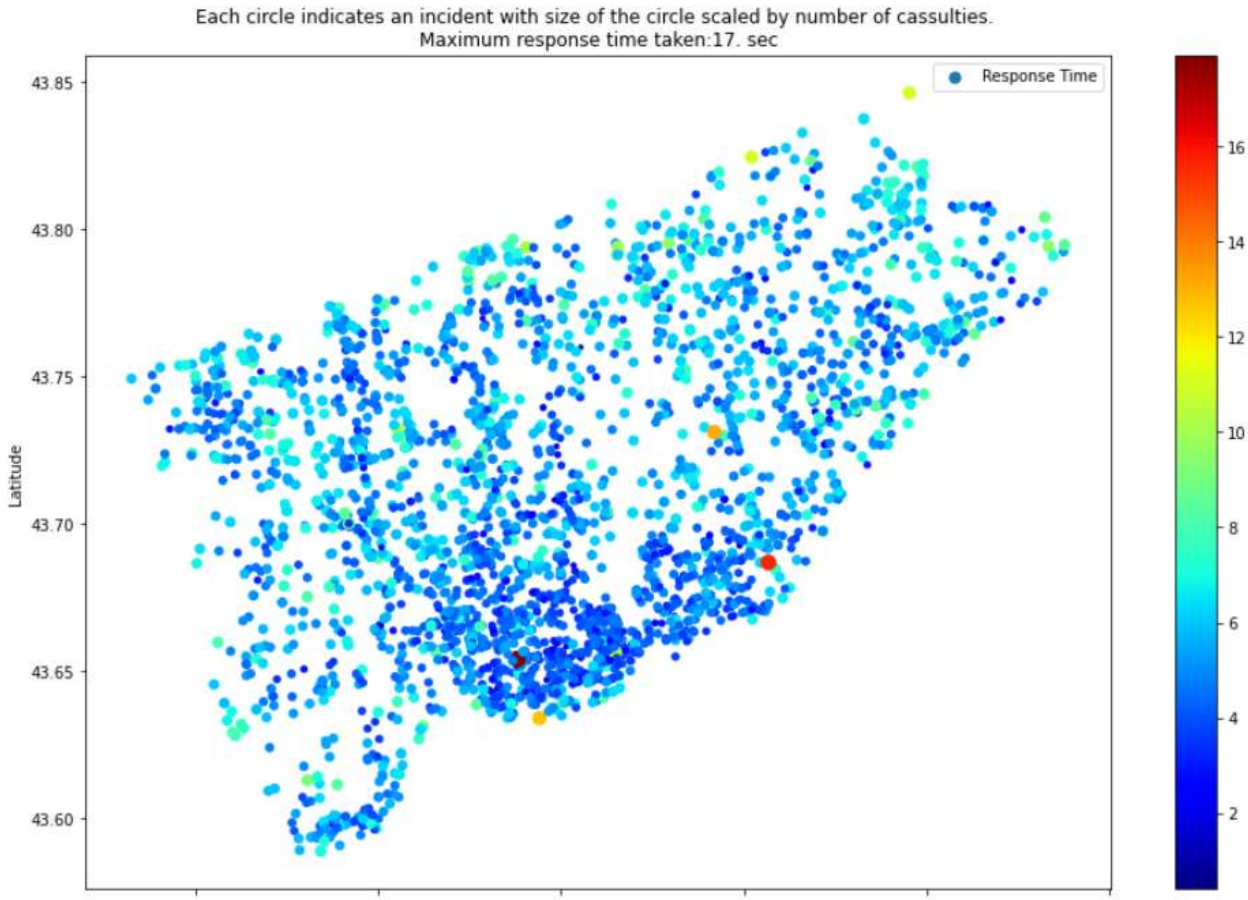
According to size and colour we can determine number of persons displaced.

12. Casualties with Location



We can see that most areas had casualties less than 6 and only 1 region had 15 casualties.

13. Location vs Response time taken in minutes

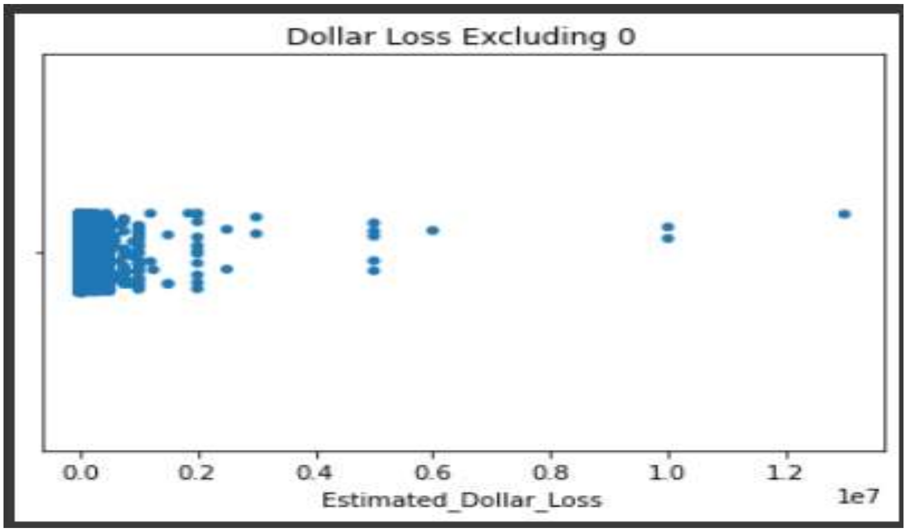


We can see most of the areas of got response within 10 minutes.

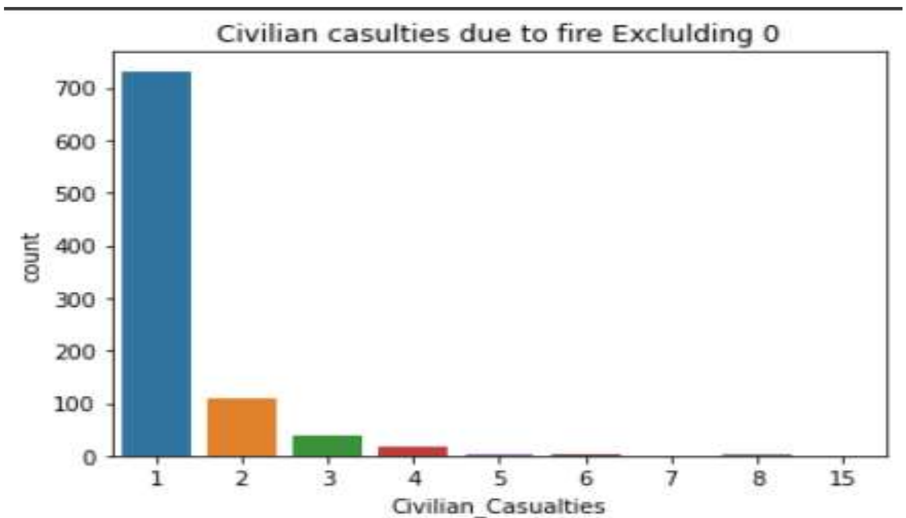
Statistical Inference of Numerical Columns

	Mean	Median	Min Value	Max Value	Most Common	Least Common	Variance	St. Dev	Skew	Kurtosis	Mean Absolute Deviation
Count of persons rescued	0.112538	0.0	0.0	15.0	0.0	15.0	2.083276e-01	0.456429	8.455166	149.308247	0.205988
Estimated Dollar Loss	0.062154	0.0	0.0	86.0	0.0	86.0	9.265744e-01	0.962587	67.355363	5747.892245	0.120684
Estimated Number of Persons Displaced	42943.693419	2500.0	0.0	50000000.0	0.0	15600.0	2.850884e+11	533936.667426	76.107737	6892.877165	63425.666540
Reaction_time_minutes	17.274835	0.0	0.0	999.0	0.0	230.0	1.417807e+04	119.071696	7.929017	61.853362	31.293370
Extinguish_time_minutes	288.281880	314.0	111.0	445.0	426.0	346.0	1.155969e+04	107.515995	-0.154630	-1.215131	94.124758
Time_spent_minutes	19.525058	19.0	0.0	44.0	20.0	0.0	1.390811e+02	11.793265	0.295487	-0.930139	9.953877

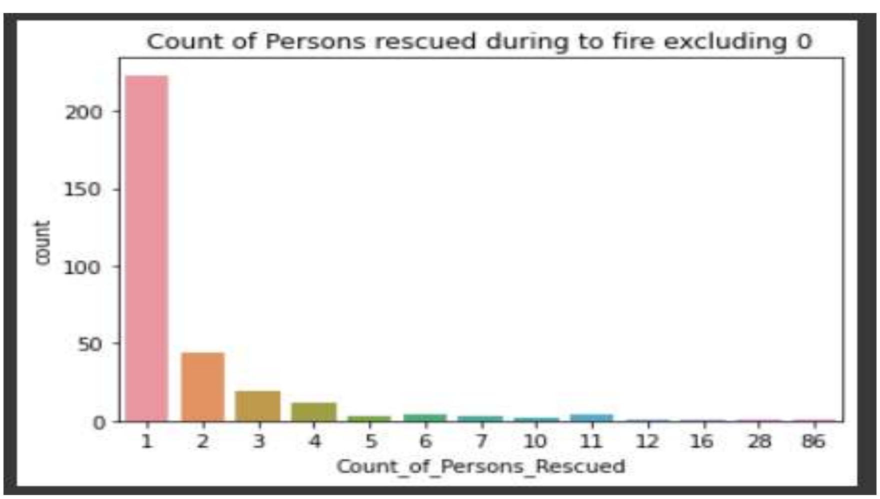
1. Estimated Dollar Loss (After data cleaning)



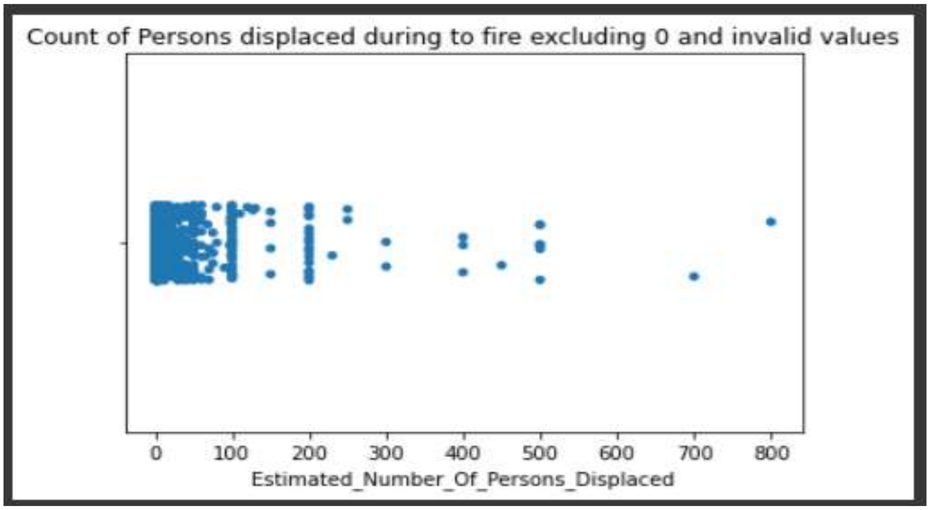
2. Civilian Casualties caused due to fire



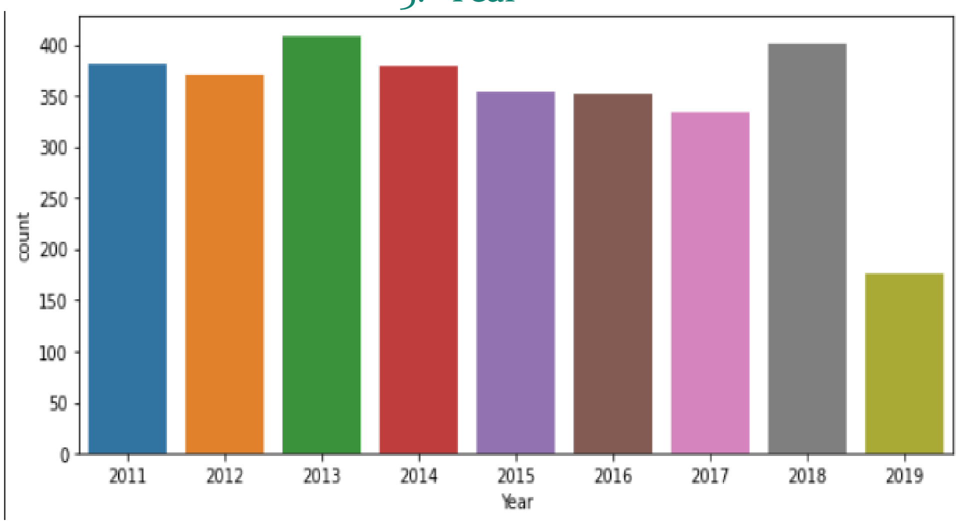
3. Persons rescued during fire



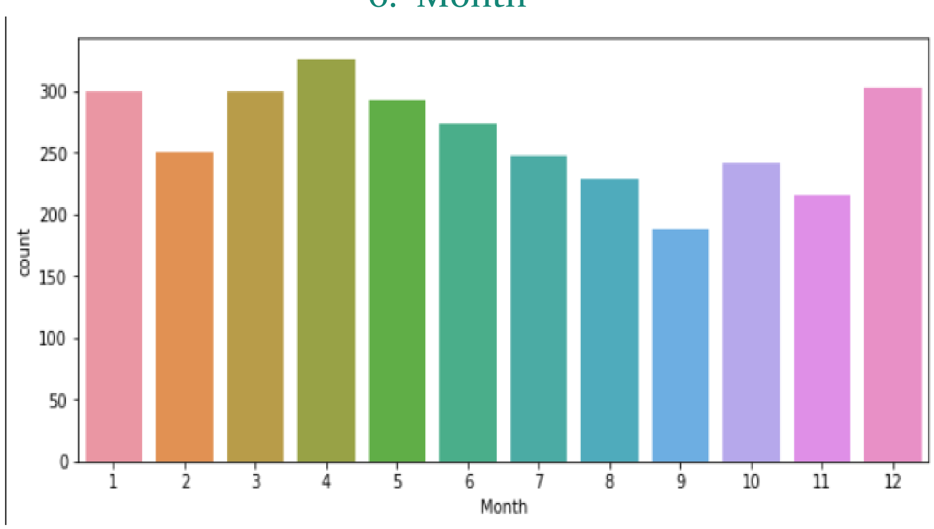
4. Persons displaced during fire



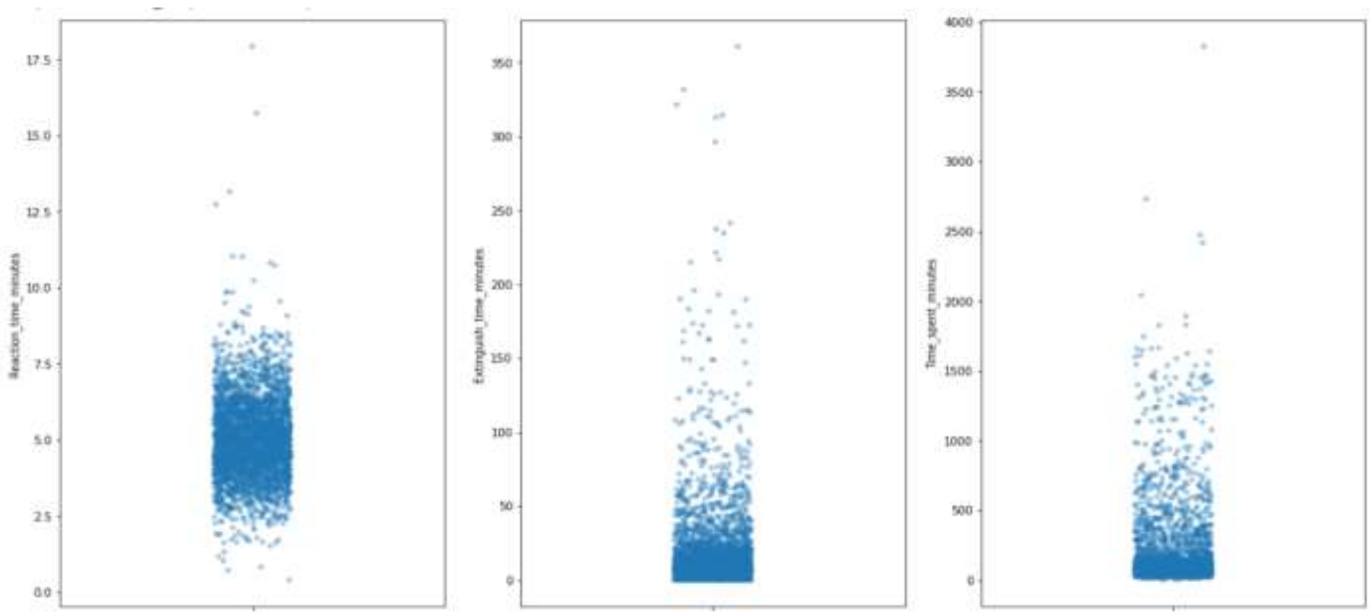
5. Year



6. Month

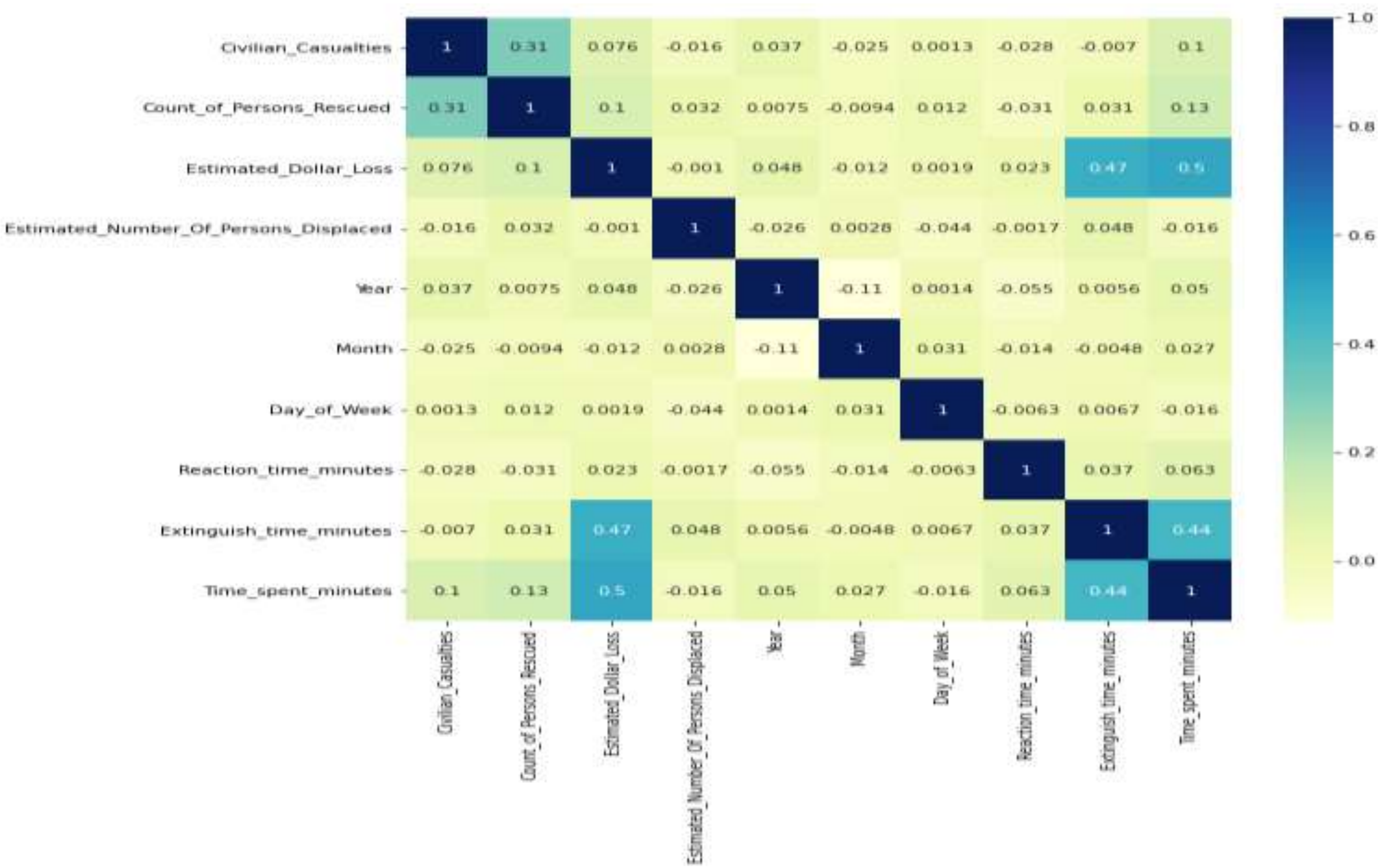


7. Reaction time in Minutes, Extinguish time in minutes, Time taken in minutes

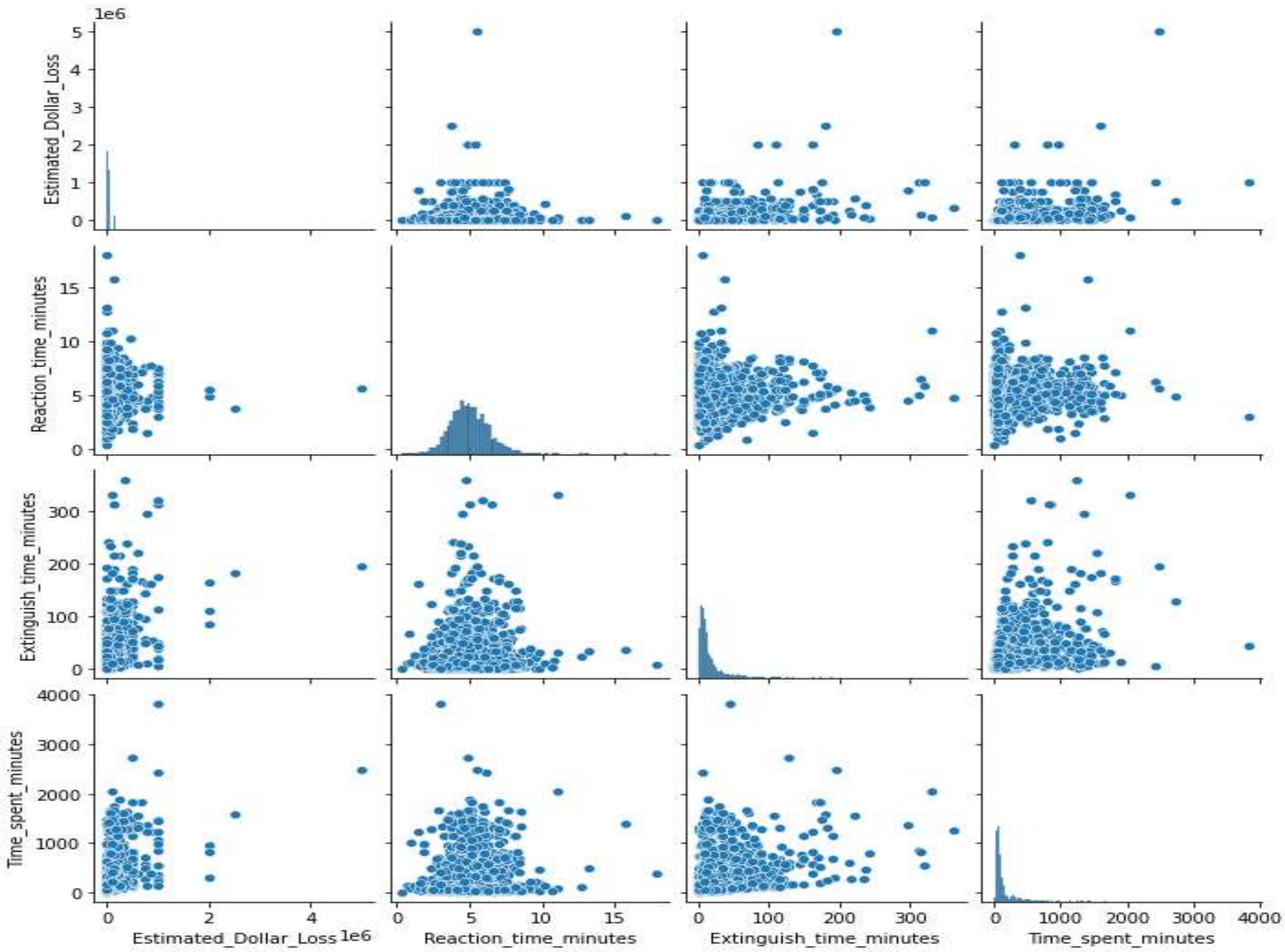


Multivariate Analysis

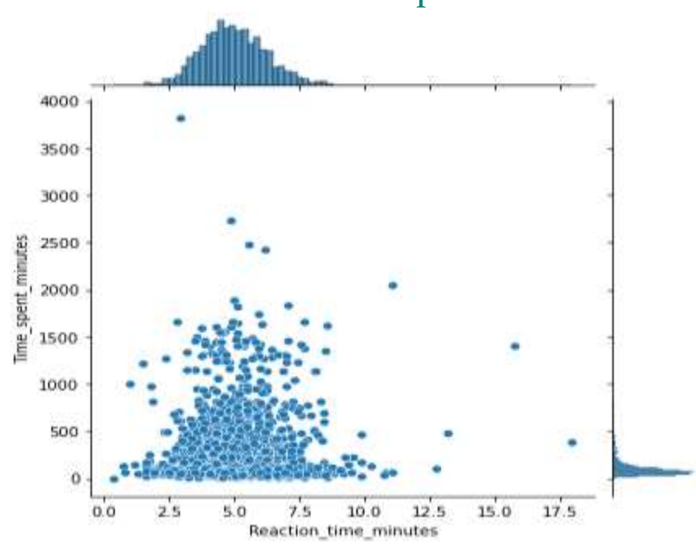
1. Relation between the numerical features – Heatmap



2. Relation between the Estimated Dollar Loss vs Reaction time, Extinguish time, Time spent minutes

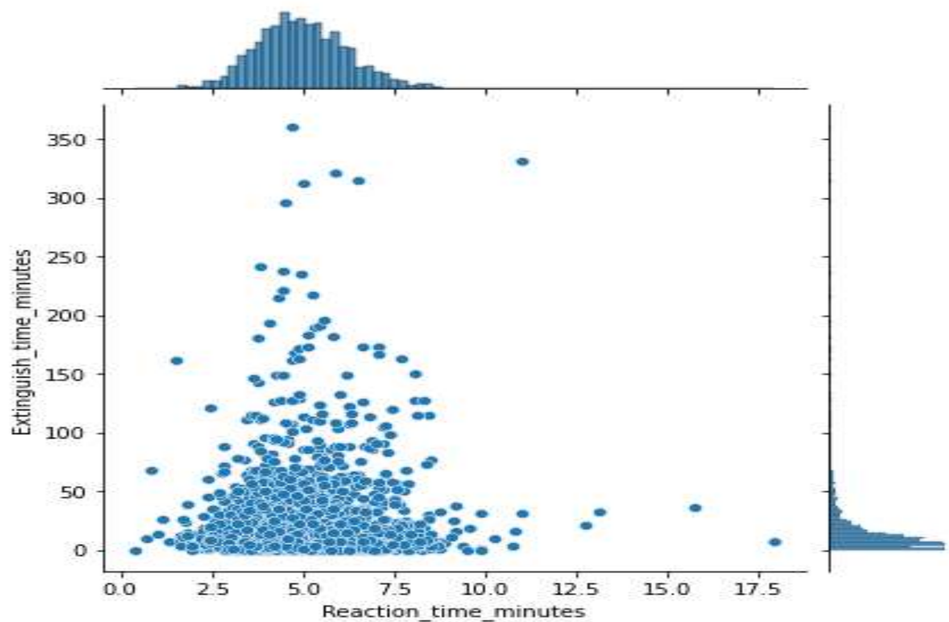


3. Reaction time vs Time spent in minutes



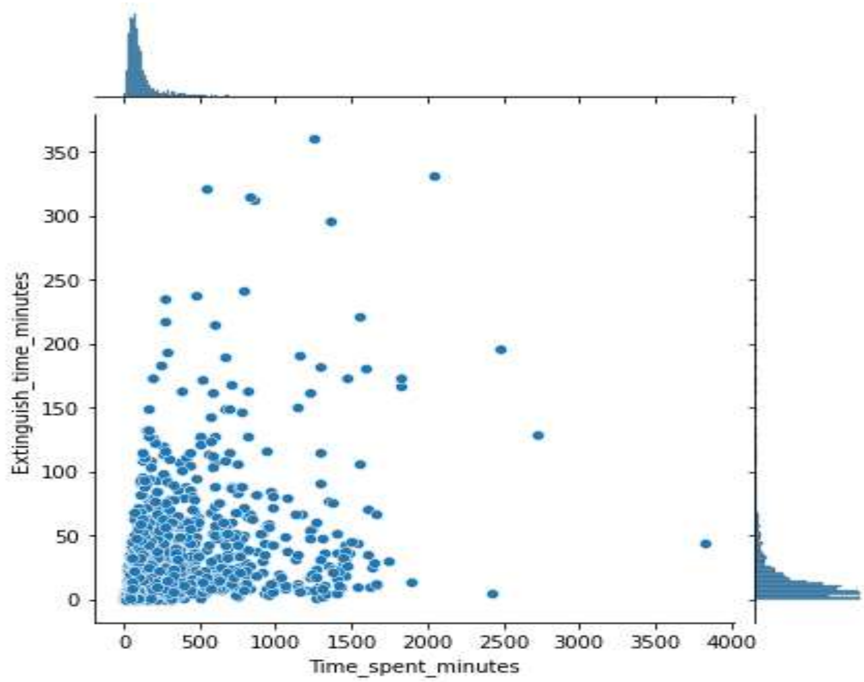
This shows the scattering of time required to reaction time taken by fire fighters and total time spent.

4. Reaction time vs Extinguish time in minutes



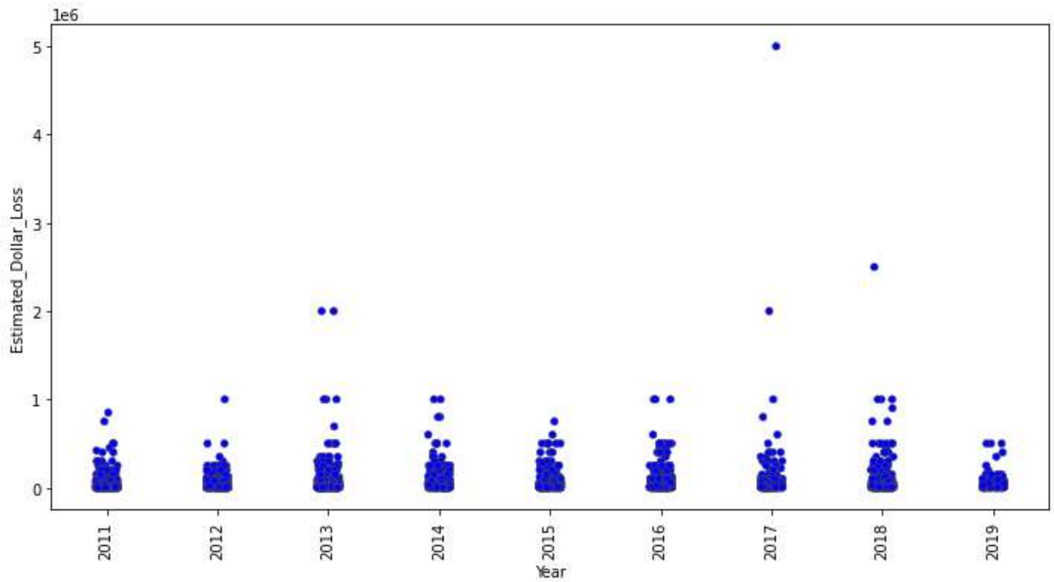
This shows the scattering of time required to extinguish the fire and reaction time required for fire fighters to reach the incident area.

5. Extinguish time vs Time spent in minutes



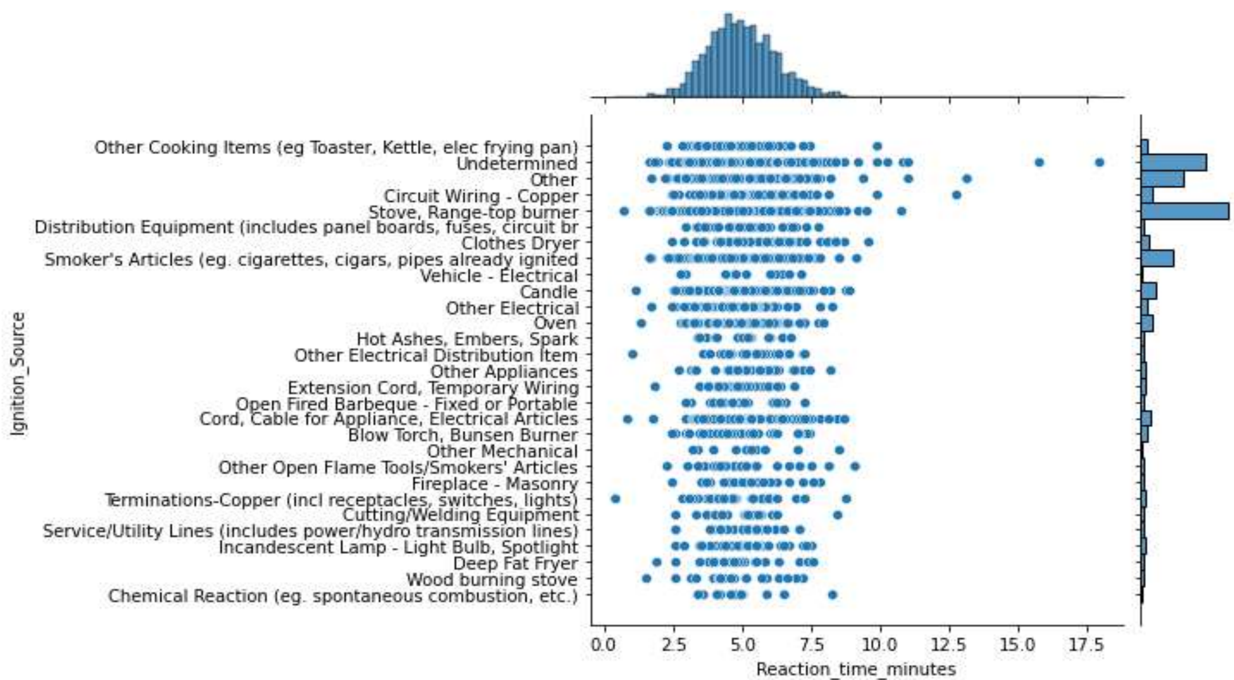
This shows the scattering of time required to extinguish the fire and total time spent.

6. Estimated Dollar Loss vs Year



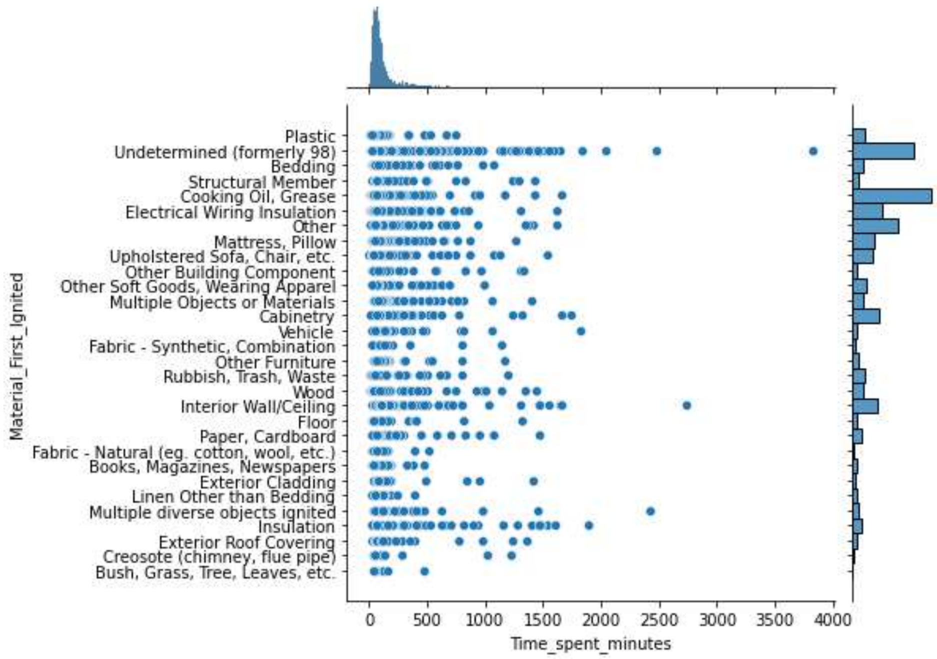
This shows that in 2017, the Estimated Dollar Loss was highest i.e., 5 million. Whereas mostly the dollar loss is below 1 million.

7. Ignition Source vs Reaction Time



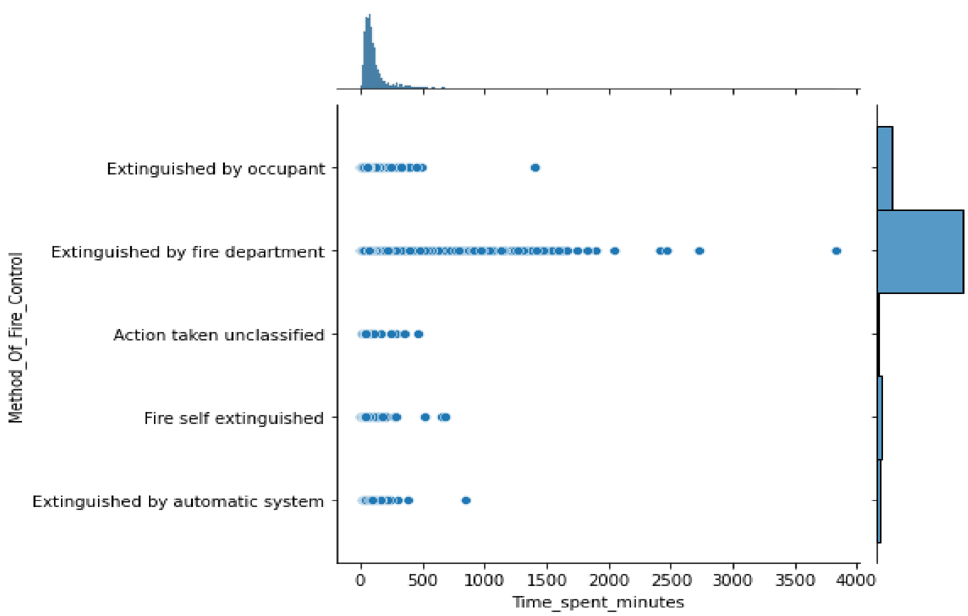
This shows the reaction time in most cases is 1 to 8 minutes. Whereas Stove and burner are Ignition source causing most fires.

8. Material first ignited vs Total time spent



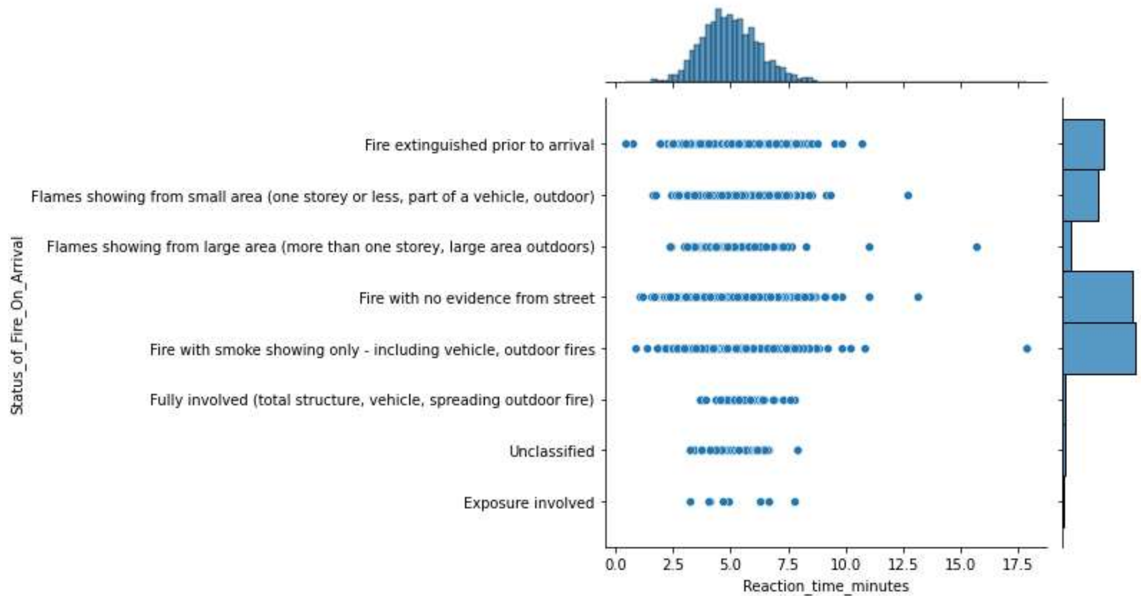
This shows the scattering of material that got ignited first and the total time required to extinguish the fire and clear the area.

9. Method of fire control vs Total time spent



This shows the scattering of method of fire control and the total time required to extinguish the fire and clear the area.

10. Fire status on arrival vs Reaction time in minutes



This shows the scattering of status of fire on arrival of fire team and the reaction time of fire team.

Conclusions

- Cooking area or Kitchen is most the fire-incident prone as most casualties occurred there, followed by Sleeping Area or Bedroom.
- Maximum Civilian Casualty caused during any incident was 15. While casualty was 0 in more than 8000 fire incidents and only 1 casualty was there in more than 700 cases.
- Fires that spread more often result in casualties.
- Evacuating people can result in more casualties.
- Maximum Civilian Casualty caused during any incident was 15. While casualty was 0 in more than 8000 fire incidents and only 1 casualty was there in more than 700 cases.
- Casualties most often occur when improperly handling cooking oil, cabinetry, bedding, and mattresses/pillows.
- Casualties occur more often when people try to put out the fire themselves, automatic systems are safest.
- In 55% of the cases Fire Alarm was present, while in 77% of the cases the Smoke Alarm was present.
- For Fire alarm: In 40% of the cases, it worked but only in 30% of the cases it was able to alert the people.
- For Smoke alarm: In 70% of the cases the operation of alarm was undetermined and in other cases it failed due to reasons such as: Smoke didn't reach the smoke alarm, battery dead or unit failure.
- Risk increases when fire alarms are installed but do not work due to batteries, improper installation etc.
- Casualties occur more often in multi-unit properties.
- In 86% of cases of fire incidents, No Business was affected.
- The maximum Dollar Loss that occurred in one of the incidents was 5 million, with average loss of 60K Dollar.
- The Ignition source in most cases is undetermined whereas its Stove or Burner in 17% cases. The cause of fire is also unknown in 30% cases while it was Electrical failure in 17% cases.
- The mean Reaction time of fire department is 5 minutes, the mean time for Extinguishing the fire is 18 minutes and mean total Time spent is 3 hrs.

THANK YOU