August 10, 2020

Group Assignment: Dataset Analysis and Interpretation – Part 1

Group 10

Group Members

1. Nidhi, Punja, npunja - (npunja@uwaterloo.ca)
2. Judith, Roth, j5roth - (j5roth@uwaterloo.ca)
3. Iman, Dordizadeh Basirabad, idordiza - (idordiza@uwaterloo.ca)
4. Daniel Adam, Cebula, dacebula - (dacebula@uwaterloo.ca)
5. Cynthia, Fung, c27fung - (c27fung@uwaterloo.ca)
6. Ben, Klassen, b6klasse - (b6klasse@uwaterloo.ca)

Objective

The Toronto Fire Services is an all hazards emergency response organization providing its city protection from a wide range of dangers and threats. This includes protection against loss of life, property, and the environment from the effects of fire, illness, accidents, and all other hazards ("Fire Services"). Through preparedness, prevention, public education and emergency responses, the Toronto Fire Services delivers its service with an emphasis on quality services, efficiency, effectiveness, and safety ("Fire Services").

The Toronto Fire Services is the largest fire service in Canada and the fifth largest in North America ("Fire Services"). Maintaining such a large organization draws interest on how the organization operates. By using data science, providing insights on organizational operations can help make appropriate business decisions to achieve the type of service it strives for. Thus, Group 10 has decided to analyze the types of calls received by the Toronto Fire Department from 2011 to 2018. Being such a large organization and responding to a variety of emergency calls, Group 10 will investigate what attributes can influence the likelihood of the top 10 types of emergency calls. Some questions that will guide the analysis in this investigation are as follows: "What are the top 10 types of emergency calls and how were the calls made?", "Are fires the main reason for calls received by the Toronto Fire Department?", "Does geography affect the number of calls received?", "Was the building being heated or cooled?", and "Does the weather or time of year effect calls?".

Data Preparation

Data Sources

A. Toronto Fire Services Basic Incident Details

- *Open Data Toronto - link - webpage - download link*

- **Description**

  Dataset provides information similar to what is sent to the Ontario Fire Marshall relating to incidents to which Toronto Fire Services respond to. The amount of information is trimmed and includes only Fire incidents as defined by Ontario Fire Marshall and covers the year range between 2011 – 2018 ("Fire Services Basic Incident Details"). For privacy purposes, personal information is not provided, and exact addresses have been aggregated to the nearest major / minor intersection. Incident exclusion has been made pursuant under Section 8 of Municipal Freedom of Information of Privacy Act ("Fire Services Basic Incident Details").

- **Features**

  Data contains the reasons as to the existence of Toronto Fire Services (TFS). Fire, medical and emergency related calls are all indicated here; as well as the source of the call and other like metrics. One can determine the main purpose of TFS calls and appropriate resource properly.

B. Toronto Fire Services Station Locations

- *Open Data Toronto - link - webpage - download link*

- **Description**

  The .shp file contains the location of all fire stations and station numbers within the City of Toronto by latitude and longitude (Decimal Degrees) ("Fire Station Locations").

- **Features**

  The location of the TFS Fire Stations is important to determine how many calls are taken by each station and if they should be proportioned fewer or more resources as a result.

C. Government of Canada Toronto Historical Climate Weather

- *Government of Canada - link - documentation*

- **Description**

  Historical Weather, climate, data, and related information for numerous locations across Canada. 3 weather stations were found for Toronto city that have climate weather observations from 2010 - 2020. The first is located near University of Toronto St. George Campus, the second is located on Toronto Centre Island in Billy Bishop Toronto City Airport and the third is in North York near York University. The weather and climate readings have been aggregated from the 3 locations for the 10-year range.

- **Features**

The Toronto Weather and Climate are important to determine how the temperature, precipitation, and heating / cooling energy use impact TFS incidents and if resources should be proportioned in climate conditions.

Preparation Process

All elements pertaining to the data preparation and analysis can be found in a GitHub repository.

- *Data Science Group 10 GitHub repository - link*

A. Data Retrieval

The data is retrieved in the 01-DATA_RETRIEVAL.ipynb Jupyter notebook. Data is retrieved and stored in folders as multiple data types dynamically when the Jupyter notebook cells are executed.

Downloading and dealing with .csv and .shp files proved to be a challenge. Data across multiple .csv files are concatenated and transformed into 1 .csv file for a given year range. More specifically, with the .shp file (geospatial files) some additional steps had to be undertaken. For instance, Anaconda virtual environment had to be setup to install the geopandas python library due to issues with the Anaconda base environment. Details on this step can be found in 02-Geopandas_and_Merging.ipynb. The data was transformed from a .shp file to a .csv file for compatibility with pandas.

B. Data Munging / Cleaning

For Toronto Fire Incident Data, cleaning was done in Jupyter notebook 02-GEOPANDAS_AND_MANIPULATIONS.ipynb. Several steps were done to clean up the data, as follows:

- First, any fire incident data with latitude and longitude values of 0 were changed to be null values.
- Second, the TFS incident number is set as the index and any duplicated TFS incident numbers were removed (<0.0005% of the dataset). This means each row will have a unique TFS Incident Number.
- Third, there are missing Incident Station Areas (TFS Fire Stations) that were imputed by their distance to the Latitude and Longitude of the call. The Haversine formula was instrumental in calculating the distance to each TFS Fire Station and returning the smallest distance as the TFS Fire Station Number.
- Fourth, Toronto Weather Climate data was aggregated across 3 weather stations. Several columns were removed due to an overabundance of nulls, as seen in 01-DATA_RETRIEVAL.ipynb Jupyter notebook.
- Fifth, TFS Fire Locations had the number of the fire station extracted from the name and set as the index.

C. Data Storage

GitHub was the main form of data storage. Due to GitHub file storage concerns, all .csv files were saved as .csv.bz2 files using the bz2 compression algorithm available on pandas. This compresses the size of the data enough to fit in a GitHub repository.

D. Data Merging

To further conserve space, the data is present as 3 different files (<30MB) that need to be merged to form the final Dataframe (~200MB).

Analysis

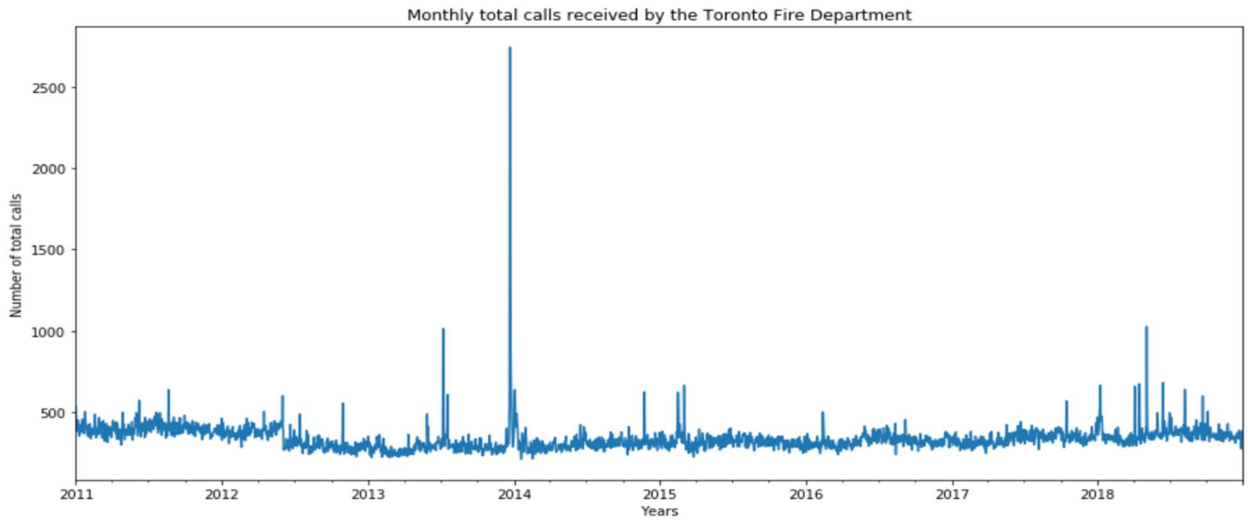
Overview

Between 2011 and 2018, The Toronto Fire Department (TFD) received 975,113 calls. The top 10 types of calls, representing approximately 86%, are as follows:

Code	Description	Count	% of total calls received	% of top 10
89	Other Medical	538,503	55%	65%
32	Alarm System Equipment - Accidental activation (exc. code 35)	56,956	6%	7%
31	Alarm Equipment - Malfunction	51,192	5%	6%
62	Vehicle Collision	49,299	5%	6%
34	Human - Perceived Emergency	33,319	3%	4%
35	Human - Accidental (alarm accidentally activated by person)	25,236	3%	3%
66	Persons Trapped in Elevator	21,346	2%	3%
33	Human - Malicious intent, prank	20,718	2%	2%
22	Pot on Stove (no fire)	19,903	2%	2%
24	Other Cooking/toasting/smoke/steam (No Fire)	17,488	2%	2%
	Total:	**833,960**		

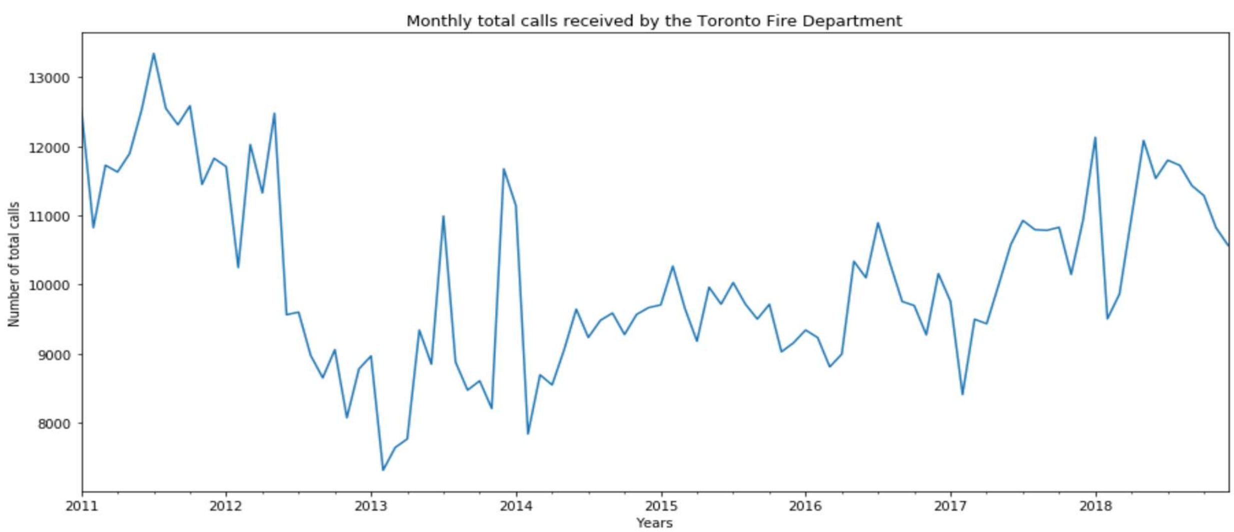
It is important to highlight that the top 10 reasons, 86% of calls received by the TFD, are not related to fire. The majority of calls received by the TFD, between 2011 and 2018, are medical (55% of the total calls received). Unexpectedly, fires are not the main reasons for calls to the TFD.

The graph below shows the number of totals calls received by the TFD during 2011 and 2018. At the end of 2013, we see a spike that indicates an outlier[1] in the graph below. After further



investigation, this outlier relates to an increase in calls received on December 22, 2013 due to a severe ice storm. The ice storm caused power outages from falling trees, as well as other emergencies in which the TFD needed to be called.

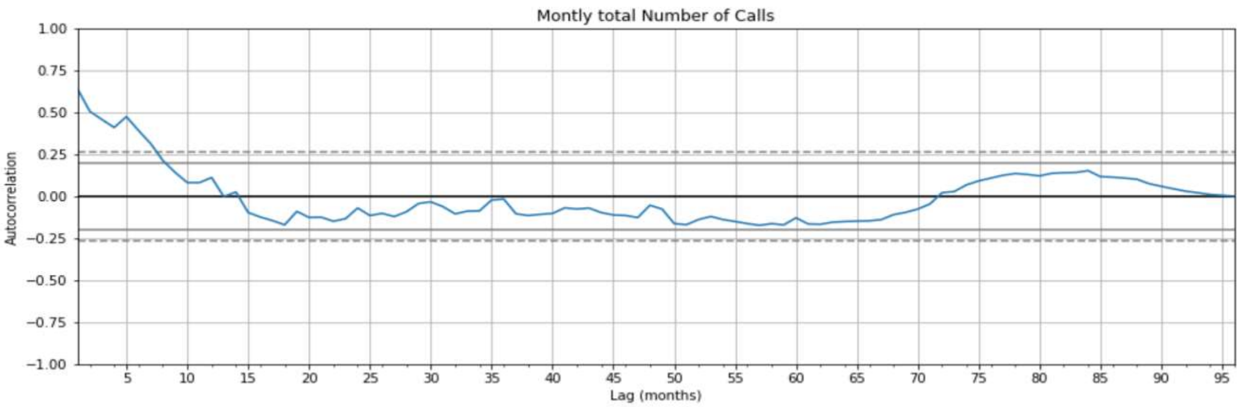
After removing the calls from December 22, 2013, the graph changes as shown below.



From the graph above, it can be observed that there is a downtrend of the total number of monthly calls received during the beginning of 2011 to the end of 2012. It can also be observed that there is a slight upward trend of the total number of monthly calls from the beginning of 2013 to the end of 2018.

[1] An outlier is an observation that lies an abnormal distance from other values.

The autocorrelation[2] of the total number of monthly calls was plotted. The graphic below reveals that there is a statistically significant positive autocorrelation for a lag of about 7 months or less. The autocorrelation is high for a lag of about 2 months. After 7 months lag, there is no statistically significant autocorrelation. This means that we can



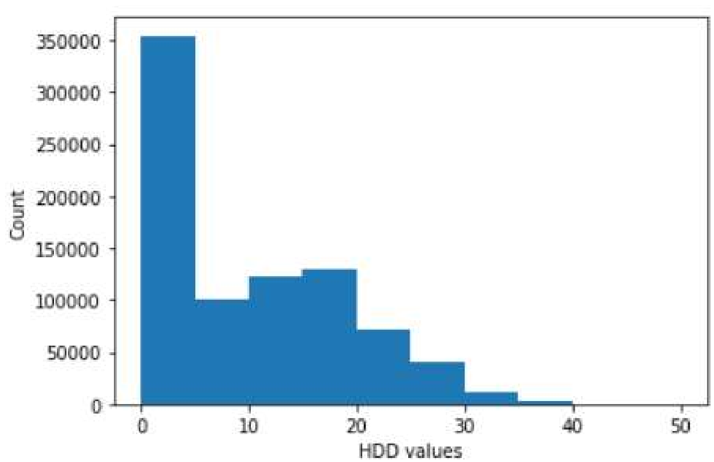
predict the number of total calls to be received based on the number of the total calls received in the previous months.

Weather

Weather and weather-related metrics did not have an impact on the number of emergency calls the fire department received. The main reason could be that most of the top 10 reasons for emergency calls are medical or minor issues/accidents unrelated to weather (except for “62 - Vehicle Collision”)

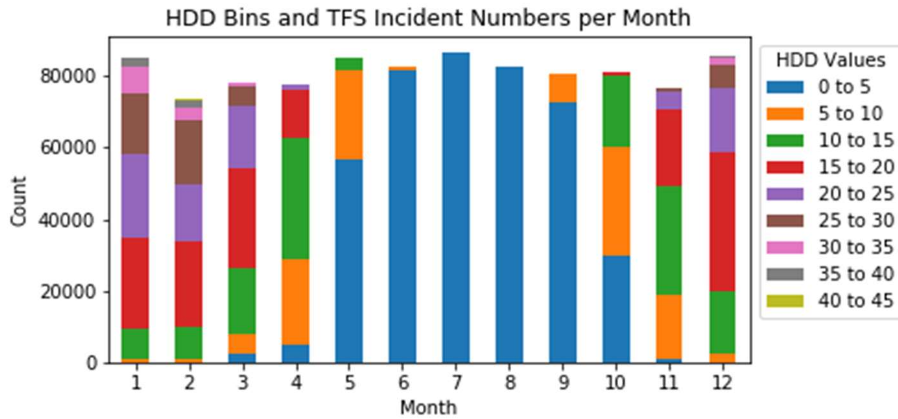
Heating Degree Day (HDD):

HDD values and the number of emergency calls were analyzed, and as shown in the chart below, about 42% of the emergency calls occur on days with minimal to no energy requirements (0-5 HDD).



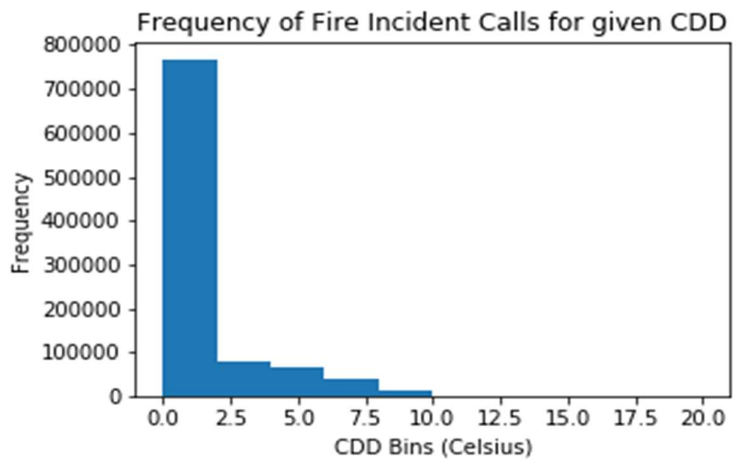
[2] It indicates how well a linear relation represents the relation between a value of a series and the same value separated by an interval of time. If the autocorrelation is significant (close to 1 or -1) indicates that the previous values of a series will be helpful in predicting the current value.

Looking at the data by month, there is no discerning trend to show a relationship between HDD and the number of emergency calls.

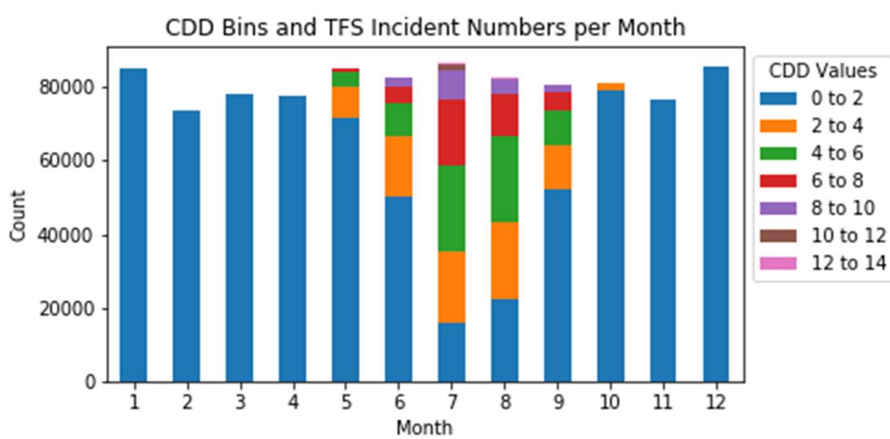


Cooling Degree Day (CDD):

As observed below, close to 80% of the emergency calls occur on days where there were no requirements for cooling.



One thing to note, as air conditioning/cooling is mostly required during the summer months, the chart still shows there is no relationship between emergency calls and CDD values.



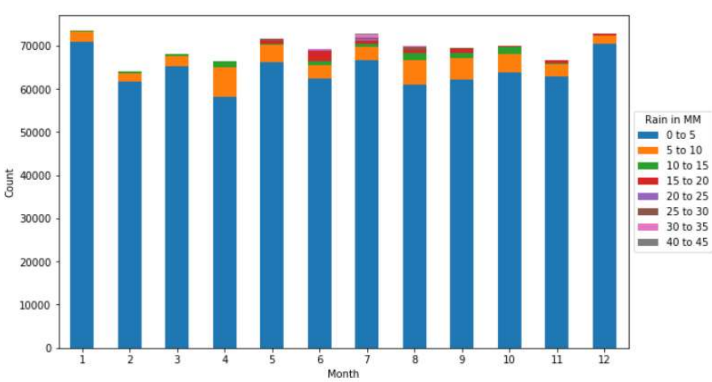
Rain and Snow

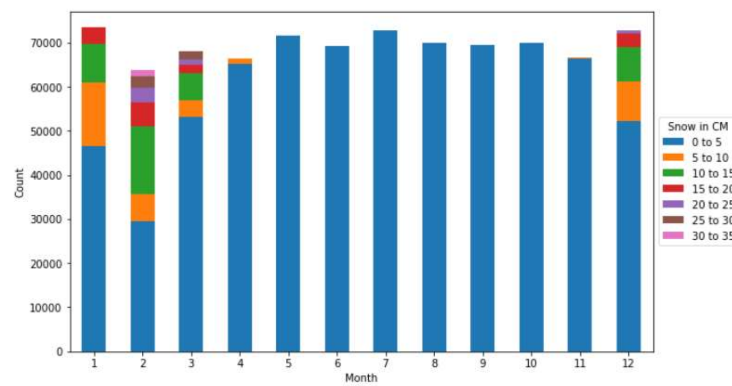
Rain and Snow were also investigated. Based on the results, rain and snow do not show any influence on the number of emergency calls. The data was grouped into bins of 5mm, for both rain and snow. As we can see from the table below, over 90% of calls occurred during 0mm-5mm of rain, and 85% of calls occurred when there is 0mm-5mm of snow.

RAIN_BINS	COUNT	PERCENT
0 to 5	770881	92.433659
5 to 10	43383	5.201905
10 to 15	9177	1.100382
15 to 20	6610	0.792582
20 to 25	766	0.091848
25 to 30	1677	0.201083
30 to 35	1234	0.147965
35 to 40	0	0.000000
40 to 45	255	0.030576
45 to 50	0	0.000000

SNOW_BINS	COUNT	PERCENT
0 to 5	735934	88.268581
5 to 10	34505	4.138561
10 to 15	38252	4.587979
15 to 20	14032	1.683011
20 to 25	5250	0.629690
25 to 30	4370	0.524142
30 to 35	1401	0.168037
35 to 40	0	0.000000

As visualized below, it is easy to see that majority of calls occurred when there was little rain and snow.

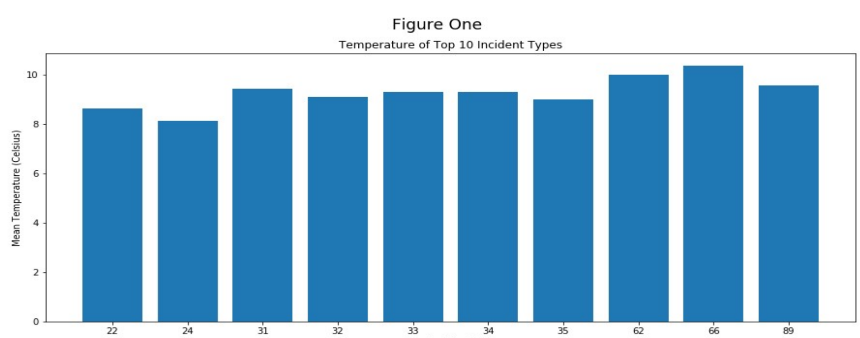




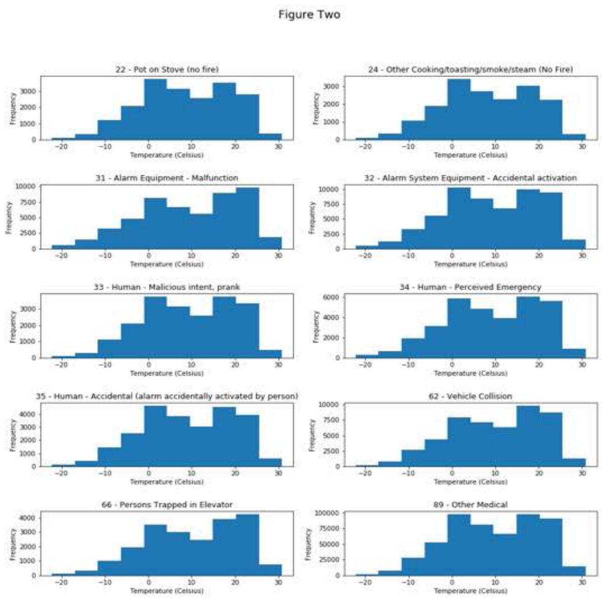
Temperature

Incident Type vs. Mean Temperature

The mean temperature varies modestly among the top 10 incident types, ranging from 8.13 to 10.34 degrees Celsius, as depicted below (Figure 1).



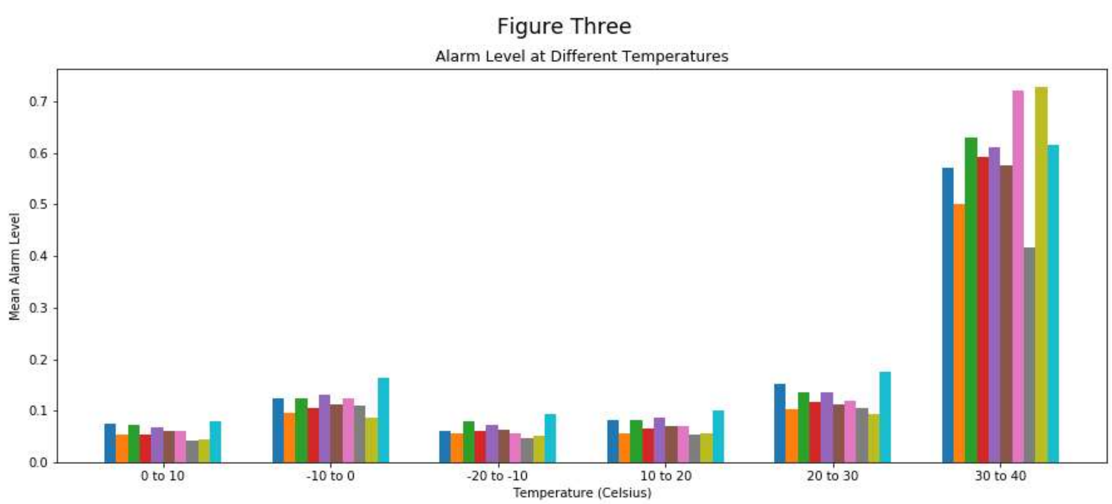
Additionally, as shown below (Figure 2), the frequency distributions for each of these types of incidents are remarkably similar. They depict peaks at around 0 and 20 degrees Celsius. However, given that the temperature in Toronto is usually between 0 and 20 degrees Celsius (evidenced by the table shown below), this is not surprising.



	Number of Incidents
Mean Temperature (Celsius)	
0 to 20	571114
20 to 40	203561
-20 to 0	199344

Alarm Level vs. Mean Temperature

As the chart reveals below, there is a noticeable spike in the severity of the emergency when the temperature is between 30 and 40 degrees Celsius. Perhaps this indicates that hotter temperatures exacerbate fires and other emergencies.



Time and Date

Various measurements of time and date were investigated to see if there was any impact on the volume of emergency calls the TFD received.

Months of the year

The chart below shows the probability of each call type per month. As it can be observed, there are no obvious outlier(s). This means that regardless of which month it is, the probability of receiving each call type is relatively the same.

Final Call Type	22	24	31	32	33	34	35	62	66	89
Month										
1	9.0	9.0	10.0	9.0	8.0	9.0	9.0	7.0	7.0	9.0
2	8.0	8.0	8.0	8.0	8.0	7.0	8.0	7.0	7.0	8.0
3	9.0	8.0	7.0	8.0	8.0	8.0	9.0	7.0	8.0	8.0
4	8.0	8.0	7.0	8.0	8.0	8.0	8.0	7.0	7.0	8.0
5	8.0	9.0	8.0	8.0	9.0	8.0	8.0	8.0	8.0	9.0
6	8.0	7.0	8.0	8.0	9.0	8.0	8.0	9.0	9.0	8.0
7	7.0	6.0	10.0	9.0	8.0	9.0	8.0	9.0	11.0	9.0
8	7.0	7.0	9.0	8.0	8.0	8.0	8.0	9.0	9.0	8.0
9	8.0	8.0	9.0	8.0	8.0	8.0	8.0	9.0	8.0	8.0
10	9.0	9.0	8.0	9.0	9.0	8.0	9.0	9.0	9.0	8.0
11	9.0	9.0	7.0	8.0	8.0	8.0	9.0	9.0	8.0	8.0
12	9.0	10.0	8.0	8.0	9.0	9.0	9.0	9.0	8.0	9.0

Day of the Week

In general, the probability seems to be evenly distributed for all call types throughout the different days of the weeks. This means that the probability of receiving each call type is about the same for all days of the week.

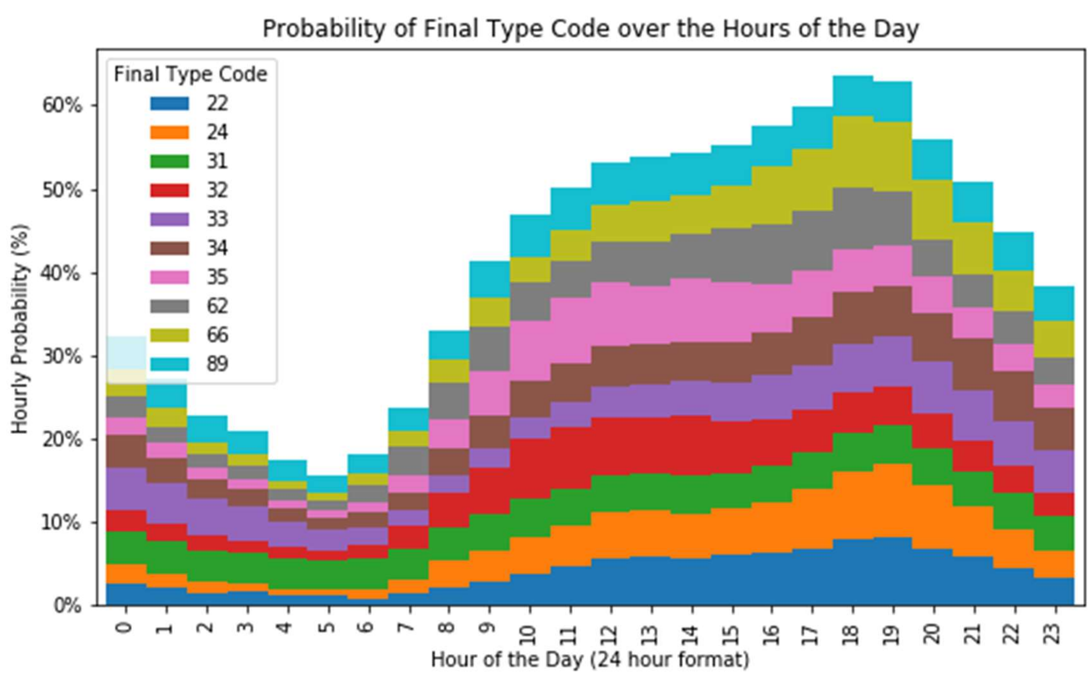
Final Call Type	22	24	31	32	33	34	35	62	66	89
Days of the weeks										
Friday	14.0	14.0	14.0	15.0	15.0	15.0	16.0	16.0	14.0	15.0
Monday	13.0	14.0	15.0	15.0	13.0	15.0	15.0	13.0	13.0	14.0
Saturday	16.0	16.0	14.0	12.0	17.0	15.0	12.0	14.0	18.0	14.0
Sunday	17.0	16.0	14.0	10.0	16.0	14.0	10.0	13.0	16.0	14.0
Thursday	13.0	14.0	14.0	16.0	13.0	14.0	16.0	15.0	12.0	14.0
Tuesday	13.0	13.0	15.0	16.0	13.0	14.0	16.0	15.0	13.0	14.0
Wednesday	14.0	13.0	14.0	16.0	13.0	14.0	16.0	14.0	13.0	14.0

Furthermore, the following points were discovered when analyzing the data by days of the week:

- 48% of calls code 32 (Alarm System Equipment - Accidental activation (exc. code 35)) comes during the middle of the workweek (Tuesday through Thursday). The day that received the least of this type of calls is Sunday, with only 10%.
- 30% of calls code 31 (Alarm Equipment – Malfunction) comes on Mondays and Tuesdays. This means that there is a slightly higher probability of receiving a code 31% call, during two consecutive days, for the first two workdays (Mondays and Tuesdays).
- 79% of calls code 35 (Human Accidental (alarm accidentally activated by person)) comes during the workweek (Monday to Friday) distributed evenly over each day.
- 34% of calls code 66 (Persons Trapped in Elevator) comes during weekends. This is the highest probability in two consecutive days.
- 48% of calls code 33 (Human - Malicious intent, prank) comes during Fridays through Sundays. This is the highest probability on three consecutive days.
- 33% of calls code 22 (Pot on Stove (no fire)) comes during the weekends. This is the highest probability in two consecutive days.
- 32% of calls code 24 (Other Cooking/toasting/smoke/steam (No Fire)) comes during the weekends. This is the highest probability in two consecutive days.

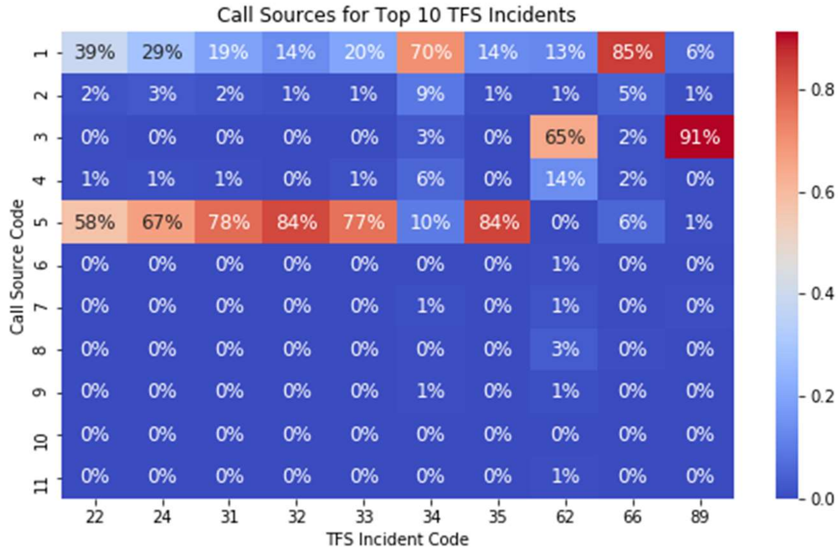
Hours of the day

Most of the top 10 call types increase in frequency after about 8 AM until about 10 PM. The exemptions to this are call type codes 31 and 33. Calls with type code 31 (Alarm Equipment – Malfunction) are consistent and frequent throughout the day. Calls of type code 33 (Human - Malicious intent, prank) are least frequently occurring from 5 AM to 9 AM.



Source of Calls

Additionally, analysis on where the emergency call came from was done as call source was available in the data. By looking into where the call came from, it provided rather interesting findings.



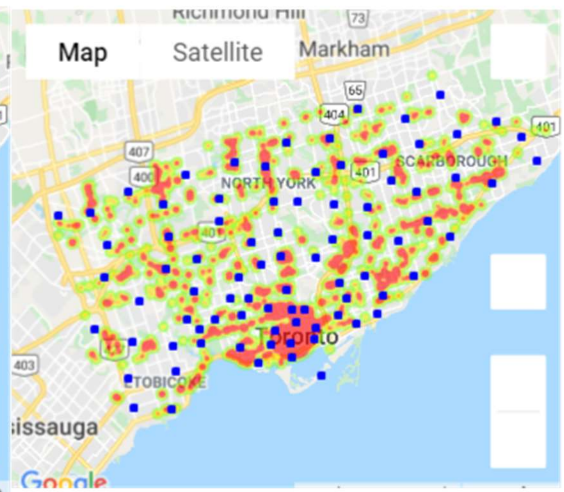
To better interpret the data, a heat map was used. The following observations were made pertaining to the top 10 types of calls:

- For the final call types listed below, at least 60% of the calls were from a telephone from a monitoring agency, and the remaining 40% of the calls were made through 911.
  - 22 - Pot on Stove (no fire)
  - 24 - Other Cooking/toasting/smoke/steam (No Fire)
  - 31 - Alarm Equipment – Malfunction
  - 32 - Alarm System Equipment - Accidental activation (exc. code 35)
  - 33 - Human - Malicious intent, prank
  - 35 - Human - Accidental (alarm accidentally activated by person)
- The high frequency of calls from a monitoring agency would indicate that the nature of these types of calls could be from detectors being set off that are wired to systems to alert the monitoring agency in which a service is provided (i.e. smoke from burning a pot setting off a fire alarm detector automatically contacting the monitoring agency)
- For 33 - Human – Perceived Emergency, there is a bit more variety from where the calls come. 70% of calls are from 911, and the remaining 30% are from a telephone from civilian (other than 911), from police services, and Telephone from a Monitoring Agency, with each representing about 10%.
- In the case of the calls code 62 (Vehicle Collisions), the distribution is different. 60% of calls are from an ambulance, while the remaining calls are mainly from 911 or police services.
- Calls of code 66 (Persons Trapped in Elevator) are mainly from 911 calls at 90%, with the remaining coming mostly from a Telephone from a Monitoring Agency.

- Interestingly, people trapped in an elevator made the top 10, and most people when trapped in an elevator call 911. The remaining is from a telephone from a monitoring agency, as perhaps they have cameras in the elevator if the trapped individual cannot call out themselves.
- Finally, calls of code 89 (Other Medical), are 90% from an ambulance, and remaining from 911.

Geospatial Data

The geospatial data (latitude and longitude) provided by the TFS Incidents were aggregated and plotted as a heat map. The top 1000 geospatial locations were observed due to Google Maps API cost concerns (out of >80,000). The heat map was then compared to the TFS Fire Station Locations to compare if the hot spots correspond to TFS Fire Station Locations. This is a qualitative analysis.



As you can see the heatmap (red, left and right image) correspond very well to the TFS Fire Station locations (blue, right image). The jupyter notebook (06-GEOSPATIAL.ipynb) provides the means to further explore the dataset by different incident type and even by the call event type; all that is required is a Google Maps API key to do so. As we can see from the heatmap, stations are appropriately located throughout the city.

Conclusion

The initial assumption, when thinking of one of the largest organizations providing fire services in Canada (“Fire Services”), is the number of fire calls they respond to. Group 10 analyzed Toronto Fire Department incident data from 2011 to 2018 that revealed many unexpected discoveries. Unlike the initial assumption, most calls received by the TFD do not actually pertain to fire incidents. More specifically, 86% the top 10 calls received by the TFD are related to the following:

- medical, (55%)
- alarm malfunction, (6%)
- alarm accidental activation, (5%)
- pranks, (5%)
- perceived emergency,(3%)
- alarm accidentally activated, (3%)
- vehicle collision, (2%)
- persons trapped in an elevator, (2%)
- pot on the stove (no fire) (2%)
- and other cooking (no fire) (2%).

This could be explained by medical/paramedic and fire services being house in the same stations across Toronto, giving the expectation of having a positive correlation.

Also discovered, the number of total calls expected in the current month can be predicted using the number of total calls in previous months. This insight becomes a great tool for the organization, as they can try and forecast and/or make prediction models for better business decisions.

It was also concluded that the month and the weather do not impact the number of calls received by the TFD. This was an unexpected finding, as one would expect to have significantly higher call volumes during December due to weather and holidays.

It was established that it was reasonable to expect calls related to accidental alarm activations by a person to be higher during the workweek than on weekends. This was confirmed through the analysis as the average probability of this type of call during work weekdays was higher. In addition, 48% of calls that were with “Human - Malicious intent” or a prank and came more often during Fridays through Sundays. This insight was also reasonable as people are more available not during work or school hours and could get together with friends to formulate and act out pranks and mischievous behaviour. Most of these calls were received between 5 PM and 2 AM, and it was exceptionally low between 5 AM and 9 AM.

The analysis also revealed that numerous calls are made by monitoring agencies. This could be explained by calls being initiated from home security systems contacting a monitoring agency automatically when alarms are set off. (i.e. smoke from a burning pot setting off fire alarm detector automatically contacting the monitoring agency). Six of the top ten final call types revealed this finding. They were as follows:

- 22 - Pot on Stove (no fire)
- 24 - Other Cooking/toasting/smoke/steam (No Fire)
- 31 - Alarm Equipment – Malfunction
- 32 - Alarm System Equipment - Accidental activation (exc. code 35)
- 33 - Human - Malicious intent, prank
- 35 - Human - Accidental (alarm accidentally activated by person)

Moreover, when a person gets into a vehicle collision, the individual is inclined to call 911 or other police services. Based on the seriousness of the accident, an ambulance and the fire department are called. The fire department is called to make sure the crash is safe to approach or to assist in the extraction of the victim(s) from being trapped in the vehicle. This is supported by the finding that more than 80% of the calls for this type of incident is made from an ambulance, 911 or police services.

Overall, the results of this analysis were interesting and unexpected. The majority of calls to Toronto Fire Services are surprisingly not for fires. As a well-rounded organization providing services for all hazardous emergencies, the majority of their calls are for other medical reasons in which these calls are made by ambulances. Regardless of what time of the year, day of the week, rain, or shine, hot or cold, the Toronto Fire Services get the same amount of calls 365 days a year.

References

“Fire Services.” *City of Toronto*, www.toronto.ca/city-government/accountability-operations-customer-service/city-administration/staff-directory-divisions-and-customer-service/fire-services. Accessed Jul. 2020.

“Fire Services Basic Incident Details.” *City of Toronto Open Data*, Fire Services, 2020, open.toronto.ca/dataset/fire-services-basic-incident-details. Accessed Jul. 2020.

“Fire Station Locations.” *City of Toronto Open Data*, Fire Services, 2017, open.toronto.ca/dataset/fire-station-locations. Accessed Jul. 2020.

“Historical Climate Data.” *Government of Canada*, Government of Canada, 4 Dec. 2019, climate.weather.gc.ca.