Group Assignment: Dataset Analysis and Interpretation – Summary Paper

Group 10

August 10, 2020

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 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 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").

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")

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.

Analysis

Since three data sets were being merged (as listed above), downloading and dealing with .csv and .shp files proved to be a challenge. Data across multiple .csv files were concatenated and transformed into one .csv file for a given year range. For the .shp file (geospatial files), Anaconda virtual environment had to be setup to install the geopandas python library due to issues with the Anaconda base environment. The data was then transformed from a .shp to a .csv file for compatibility with pandas. Additional steps were done to clean up the data, as these would otherwise result in incorrect analysis

1. Any fire incident data with latitude and longitude values of 0 were changed to null values.
2. There were duplicate TFS incident numbers (<0.0005% of the dataset) in the data set so these had to be deduped to avoid inflating number of incidents.
3. There were 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.
4. Toronto Weather Climate data was aggregated across 3 weather stations. Several columns were removed due to an overabundance of nulls

Other issues Group 10 encountered were with 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. To further conserve space, the data is present as 3 different files (<30MB) that need to be merged to form the final Dataframe (~200MB).
All components of the analysis can be found in a GitHub repository.

- *Data Science Group 10 GitHub repository - link*

Conclusion

Overall, the results of this analysis were interesting and unexpected. The initial assumption regarding one of the largest organizations providing fire services in Canada, was that majority of calls must be for fires. Surprisingly, the majority of calls to Toronto Fire Services are not for fires. Group 10's analysis has disproved this initial assumption and revealed other meaningful insights concerning the Toronto Fire Services. 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 primarily made by ambulances.

The number of total calls expected in the current month can be predicted using the number of total calls in previous months, allowing the organization to prepare and make better more effective decisions. The analysis concluded that attributes such as time and weather did not impact the calls as believed. 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.