

Applied Data Science Capstone Week 5

Fredericton Crime Study:

Introduction:

Fredericton is the capital of the Canadian province of New Brunswick. The city is situated in the west-central portion of the province along the Saint John River, which flows west to east as it bisects the city. The river is the dominant natural feature of the area. One of the main urban centres in New Brunswick, the city had a population of 56,224 in the 2011 census. It is the third-largest city in the province after Moncton and Saint John.

Researches and developers are usually interested to decreasing the crime in this city, so there is usually a big question about what neighbourhoods have the highest crime, and Is population density correlated to crime level. So, they usually study use the available data on the open sources sites with focus on this issue. So in this report we will study the open data about this city and its neighbours, and analysis data to be able to give a conclusion that may help the decision maker in this city.

Data:

Using this data will allow exploration and examination to answer the questions. The neighbourhood data will enable us to properly group crime by neighbourhood. The Census data will enable us to then compare the population density to examine if areas of highest crime are also most densely populated. Fredericton locations of interest will then allow us to cluster and quantitatively understand the venues most common to that location.

There are many different open source websites we will use in our study like:

<http://data-fredericton.opendata.arcgis.com/datasets/crime-by-neighbourhood-2017--crime-par-quartier-2017> where Crime types are listed by dates, neighbourhoods and wards. Crimes listed are against property (and not person) and match what is posted on the crime maps put up on the City's Internet site. The tabular data can be related to the Neighbourhood and Ward map datasets also found on this portal. We can download data from this website as csv files, or download all the dataset, it offers data with different attributes, like neighbourhood, crime type, city,...

we will also these different sites for more data:

<https://foursquare.com/>

<https://www.numbeo.com/crime/in/Fredericton>

<http://www.fredericton.ca/en/fredericton-police/crime-map>

Methodology

The methodology of work will start in loading data set, then we will study different descriptive statistics for different parameters like crime type and neighbourhood ,after that we will study the crime types and then pivot analysis of crime type frequency by neighbourhood, then we will study the correlation between population density and crimes. In orders to build a model, we will use k-means statistical algorithm on venues by locations of interest based on findings from crimes and neighbourhood.

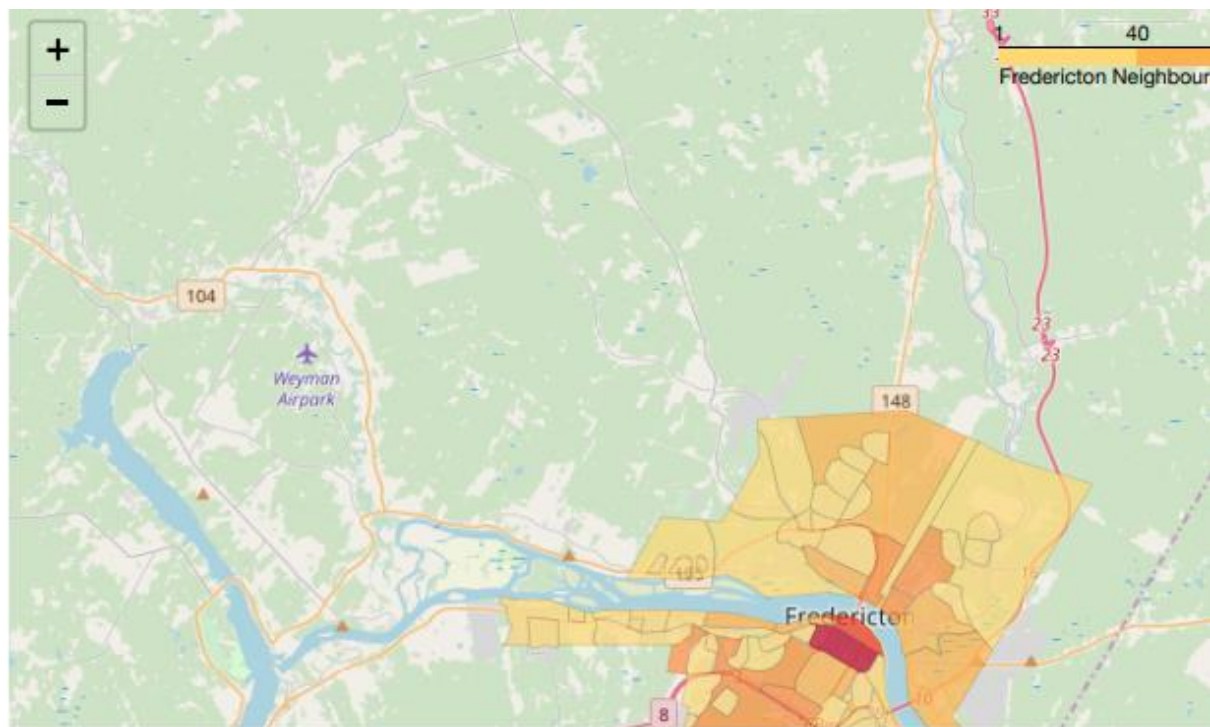
Load Data:

We first download an excel file, "Crime by Neighbourhood 2017" from the City of Fredericton Open Data site is found under the Public Safety domain. This dataset was then uploaded for the analysis. It's

interesting to note the details of this dataset are aggregated by neighbourhood. It is not an exhaustive set by not including all crimes (violent offenses) nor specific location data of the crime but is referenced by neighbourhood. This means we can gain an understanding of the crime volume by type by area but not specific enough to understand the distribution properties. Valuable questions such as, "are these crimes occurring more often in a specific area and at a certain time by a specific demographic of people?" cannot be answered nor explored due to what is reasonably assumed to be personal and private information with associated legal risks. There is value to the city to explore the detailed crime data using data science to predict frequency, location, timing and conditions to best allocated resources for the benefit of its citizens and its police force.

Visualization:

Once the data was prepared, a choropleth map was created to view the crime count by neighbourhood. As expected the region of greatest crime count was found in the downtown and Platt neighbourhoods. Examining the crime types enables us to learn the most frequent occurring crimes which we then plot as a bar chart to see most frequently type. Theft from motor vehicles is most prevalent in the same area as the most frequent crimes. It's interesting to note this area is mostly residential and most do not have garages. It would be interesting to further examine if surveillance is a deterrent for motor vehicle crimes in the downtown core compared to low surveillance in the Platt neighbourhood.



We also study the neighbourhood crime count and have the result in the following table:

| | Neighbourh | Crime_Count |
|----|--------------------------|-------------|
| 0 | Barkers Point | 47 |
| 1 | Brookside | 54 |
| 2 | Brookside Estates | 9 |
| 3 | Brookside Mini Home Park | 5 |
| 4 | College Hill | 41 |
| 5 | Colonial heights | 9 |
| 6 | Cotton Mill Creek | 4 |
| 7 | Diamond Street | 1 |
| 8 | Doak Road | 1 |
| 9 | Douglas | 3 |
| 10 | Downtown | 127 |
| 11 | Dun's Crossing | 18 |
| 12 | Forest Hill | 12 |
| 13 | Fredericton South | 85 |
| 14 | Fulton Heights | 36 |
| 15 | Garden Creek | 13 |

connection to venues using Foursquare data

By Loading the "Fredericton Locations" data enables us to perform a statistical analysis on the most common venues by location. We might wonder if the prevalence of bars and clubs in the downtown region has something to do with the higher crime rate in the near Platt region. Plotting the latitude and longitude coordinates of the locations of interest onto the crime choropleth map enables us to now study the most common venues by using the Foursquare data. Analysing each Location Grouping rows by location and the mean of the frequency of occurrence of each category we venue categories we study the top five most common venues. Putting this data into a pandas data frame we can then determine the most common venues by location and plot onto a map.

Results

The analysis enabled us to discover and describe visually and quantitatively: neighbourhoods in Fredericton, crime frequency by neighbourhood, Crime type frequency and statistics, the mean crime count in the City of Fredericton is 22. Crime type count by neighbourhood. Theft from motor vehicles is most prevalent in the same area as the most frequent crimes. It's interesting to note this area is mostly residential and most do not have garages. It would be interesting to further examine if surveillance is a deterrent for motor vehicle crimes in the downtown core compared to low surveillance in the Platt neighbourhood. 1. Motor Vehicle crimes less than \$5000 analysis by neighbourhood and resulting statistics. The most common crime is Other Theft less than 5k followed by Motor Vehicle Theft less than 5k. There is a mean of 6 motor vehicle thefts less than 5k by neighbourhood in the City. 2. That population density and resulting visual correlation is not strongly correlated to crime frequency.

Causation for crime is not able to be determined given lack of open data specificity by individual and environment. 3. Using k-means, we were able to determine the top 10 most common venues within a 1 km radius of the centroid of the highest crime neighbourhood. The most common venues in the highest crime neighbourhood are coffee shops followed by Pubs and Bars. While, it is not valid, consistent, reliable or sufficient to assume a higher concentration of the combination of coffee shops, bars and clubs predicts the amount of crime occurrence in the City of Fredericton, this may be a part of the model needed to be able to in the future. 1. We were able to determine the top 10 most common venues by location of interest. 2. Statistically, we determined there are no coffee shops within the Knowledge Park clusters. Discussion and Recommendations the City of Fredericton Open Data enables us to gain an understanding of the crime volume by type by area but not specific enough to understand the distribution properties. Valuable questions such as, "are these crimes occurring more often in a specific area and at a certain time by a specific demographic of people?" cannot be answered nor explored due to what is reasonably assumed to be personal and private information with associated legal risks. There is value to the city to explore the detailed crime data using data science to predict frequency, location, timing and conditions to best allocated resources for the benefit of its citizens and its police force. However, human behaviour is complex requiring thick profile data by individual and the conditions surrounding the event(s). To be sufficient for reliable future prediction it would need to demonstrate validity, currency, reliability and sufficiency.

Recommendations

The City of Fredericton Open Data enables us to gain an understanding of the crime volume by type by area but not specific enough to understand the distribution properties. Valuable questions such as, "are these crimes occurring more often in a specific area and

at a certain time by a specific demographic of people?" cannot be answered nor explored due to what is reasonably assumed to be personal and private information with associated legal risks. There is value to the city to explore the detailed crime data using data science to predict frequency, location, timing and conditions to best allocated resources for the benefit of its citizens and it's police force. However, human behaviour is complex requiring thick profile data by individual and the conditions surrounding the event(s). To be sufficient for reliable future prediction it would need to demonstrate validity, currency, reliability and sufficiency. A note of caution is the possibility neighbourhoods names could change. The crime dataset did not mention which specific neighbourhood naming dataset it was using but we assumed the neighbourhood data provided aligned with the neighbourhoods used in the crime data. It may be beneficial for the City to note and timestamp neighbourhood naming in the future or simply reference with neighbourhood naming file it used for the crime dataset.

Conclusion:

Using a combination of datasets from the City of Fredericton Open Data project and Foursquare venue data we were able to analyse, discover and describe neighbourhoods, crime, population density and statistically describe quantitatively venues by locations of interest. While overall, the City of Fredericton Open Data is interesting, it misses the details required for true valued quantitative analysis and predictive analytics which would be most valued by investors and developers to make appropriate investments and to minimize risk.