APPLIED DATA SCIENCE CAPSTONE PROJECT REPORT

INTRODUCTION

Fredericton is the capital city 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
The city offers a wide spectrum of venues and is a government, university, and cultural hub.
We use open data about the city for entrepreneurial uses and will help the citizens of the city to get the most out of the existing resources.
Target Audience: Policymakers, City planners and Developers and the government workers
QUESTIONS:
1) Are the population in an area and the crime level in an area have a correlation or not?
2) What type of locations (coffee shops, police stations) are present in all the different locations of the city?

3) The neighborhood with what location have the highest crime level?

DATA

Foursquare Developers Access to use the data needed data: https://foursquare.com/
Fredericton Census Tract Demographics:
http://data-fredericton.opendata.arcgis.com/datasets/census-tract-demographics--donn%C3%A9es-d%C3%A9mographiques-du-secteur-de-recensement
Fredericton Neighbourhoods:
http://data-fredericton.opendata.arcgis.com/datasets/neighbourhoods--quartiers
Open Data Site: http://data-fredericton.opendata.arcgis.com/

Using the data from these websites we will try to discuss and find the answers to the questions listed in the introduction section. Different data from different sources are used for different works. For eg: We can group the crimes committed in an area using neighborhood data. We can use clusters to find the locations which are present in all the areas of interest.

METHODOLOGY

1) Loading of all the libraries required for the project
2) Downloading all the required datasets from the different mentioned websites.
3) Loading all the datasets into the jupyter notebook to explore them and make the required conclusions.
4) Examing the crime rate in each frequency.
5) Examine the type of crimes in each neighborhood.
6) STudy of the crime rate and the crime type frequency based on the neighbourhood.

7) Understanding the correlation between the crime rate and the population at the particular area

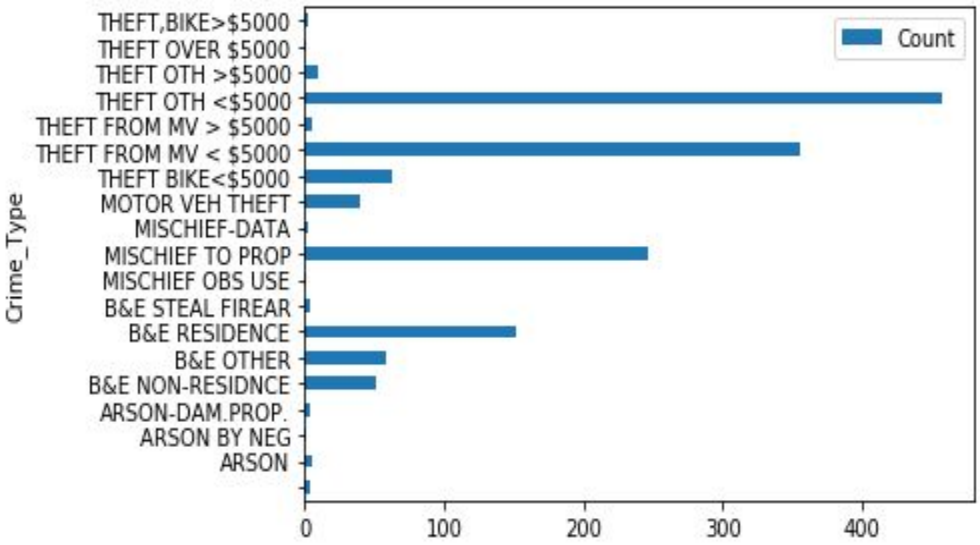
	Neighbourhood	Count
0	Barkers Point	47
1	Brookside	54
2	Brookside Estates	9
3	Brookside Mini Home Park	5
4	College Hill	41

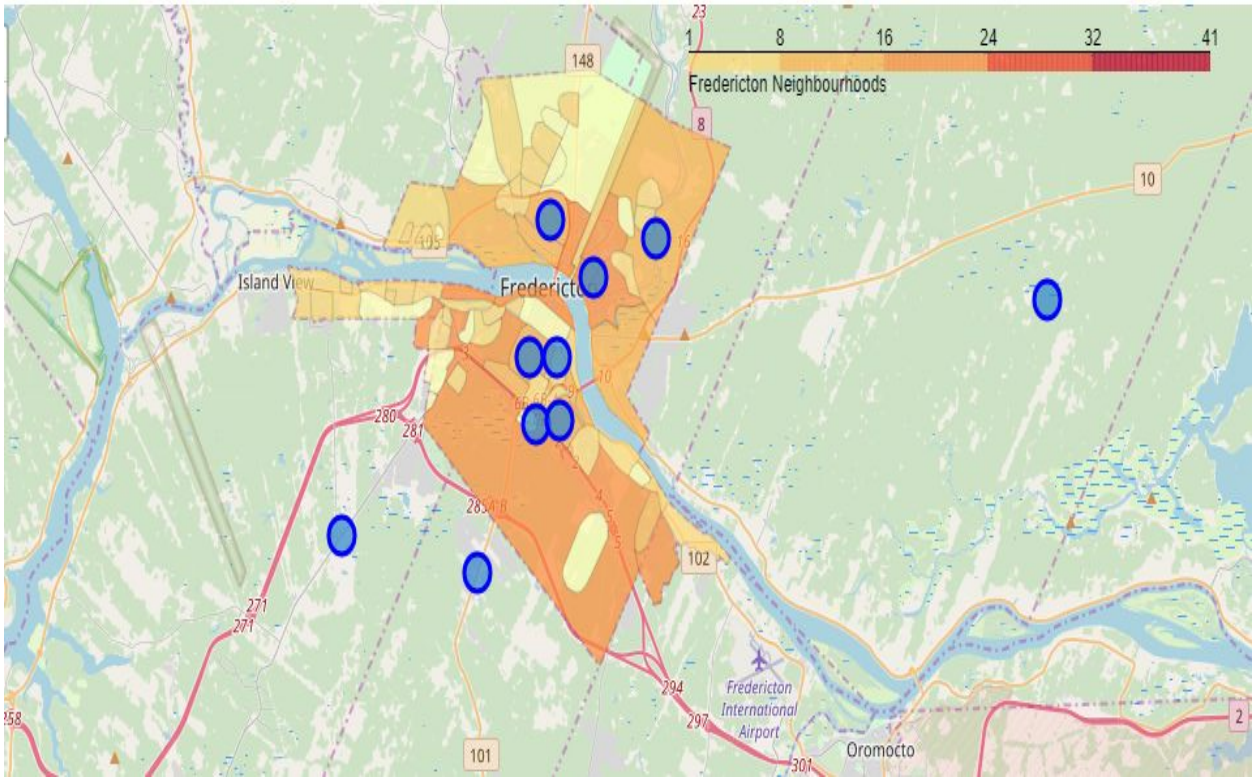
RESULTS

1) Theft from motor vehicles is most prevalent in the same area as the most frequent crimes
2) This area is mostly residential and most do not have garages.
3) From the examined plots, we found out that the most common crime is another theft less than 5k dollars followed by motor vehicle theft less than 5k dollars.
4) We can also see that the average crime rate is around 7 in the neighborhoods by the plotted graphs.
5) As we have examined we can also see that there is a confirmed the correlation between the population in a given area and the crime rate in that area. So our hypothesis comes out to be false.
6) The most common venues in the highest crime neighborhood are coffee shops followed by Pubs and Bars.
7) We were able to determine the top 10 most common venues by location of interest.

8) Statistically, we determined there are no coffee shops within the Knowledge Park clusters.

These are the conclusions and results that we are able to make using the data.


Count
Crime_Type
THEFT,BIKE>$5000
THEFT OVER $5000
THEFT OTH >$5000
THEFT OTH <$5000
THEFT FROM MV > $5000
THEFT FROM MV < $5000
THEFT BIKE<$5000
MOTOR VEH THEFT
MISCHIEF-DATA
MISCHIEF TO PROP
MISCHIEF OBS USE
B&E STEAL FIREAR
B&E RESIDENCE
B&E OTHER
B&E NON-RESIDNCE
ARSON-DAM.PROP.
ARSON BY NEG
ARSON
0
100
200
300
400


1
8
16
24
32
41
Fredericton Neighbourhoods
Island View
Fredericton
Fredericton International Airport
Oromocto
148
8
10
102
101
2

DISCUSSIONS

We are not able to find answers for most of the hypotheses that came to our mind using the existing data. We need more data for creating a more general overview of that area or make any conclusions about that area.

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 behavior is complex requiring thick profile data by individuals and the conditions surrounding the event(s). To be sufficient for reliable future prediction it would need to demonstrate validity, currency, reliability, and sufficiency.

Theft from motor vehicles is most prevalent in the same area as the most frequent crimes. It is 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 neighborhood.

CONCLUSIONS

Using a combination of datasets from the City of Fredericton Open Data project and Foursquare venue data we were able to analyze, discover and describe neighborhoods, 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.