Segmenting and Clustering Neighborhoods in Fredricton, Canada

APPLIED DATA SCIENCE CAPSTONE WEEK 5

ABDALLAH MOUBAYED

Opportunity Introduction

- Fredericton is the Capital City of the only Canadian fully-bilingual Province of New Brunswick and is beautifully located on the banks of the Saint John River.
- While one of the least populated provincial capital cities with a population base of less than 60 thousand residents, it offers a wide spectrum of venues and is a government, university and cultural hub.
- As the city grows and develops, it becomes increasingly important to examine and understand it quantitatively. The City of Fredericton provides open data for everyone and encourages entrepreneurial use to develop services for the benefit of its citizens.

Opportunity Introduction

- Developers, investors, policy makers and/or city planners have an interest in answering the following questions as the need for additional services and citizen protection:
  1) What neighborhoods have the highest crime?
  2) Is population density correlated to crime level?
  3) Using Foursquare data, what venues are most common in different locations within the city?
  4) Does the Knowledge Park really need a coffee shop?
- Does the Open Data project have specific enough or thick enough data to empower decisions to be made or is it too aggregate to provide value in its current detail? Let's find out.

In [73]:
```
from IPython.display import Image
from IPython.core.display import HTML
Image(url= "http://www.tourismfredericton.ca/sites/default/files/field/image/frederictc
n.jpg")
```

Out[73]:

Data Description

- **To understand and explore we will need the following City of Fredericton Open Data:**
  1) Open Data Site: http://data-fredericton.opendata.arcgis.com/
  2) Fredericton Neighbourhoods: http://data-fredericton.opendata.arcgis.com/datasets/neighbourhoods—quartiers
  3) Fredericton Crime by Neighbourhood: http://data-fredericton.opendata.arcgis.com/datasets/crime-by-neighbourhood-2017--crime-par-quartier-2017
  4) Fredericton Census Tract Demographics: http://data-fredericton.opendata.arcgis.com/datasets/census-tract-demographics--donn%C3%A9es-d%C3%A9mographiques-du-secteur-de-recensement
  5) Fredericton locations of interest: https://github.com/JasonLUrquhart/Applied-Data-Science-Capstone/blob/master/Fredericton%20Locations.xlsx
  6) Foursquare Developers Access to venue data: https://foursquare.com/

Data Description

- Using this data will allow exploration and examination to answer the questions.
- The neighborhood data will enable us to properly group crime by neighborhood.
- 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.

Methodology

- All steps are referenced below in the Appendix: Analysis section.
- The methodology will include:
  1) Loading each data set
  2) Examine the crime frequency by neighborhood
  3) Study the crime types and then pivot analysis of crime type frequency by neighborhood
  4) Understand correlation between crimes and population density
  5) Perform k-means statistical analysis on venues by locations of interest based on findings from crimes and neighborhood
  6) Determine which venues are most common statistically in the region of greatest crime count then in all other locations of interest.
  7) Determine if an area, such as the Knowledge Park needs a coffee shop.

Data Loading

- After loading the applicable libraries, the referenced geo-json neighborhood data was loaded from the City of Fredericton Open Data site.
- This dataset uses block polygon shape coordinates which are better for visualization and comparison.
- The City also uses Ward data but the Neighborhood location data is more accurate and includes more details. The same type of dataset was then loaded for the population density from the Stats Canada Census tracts.

Data Loading

- The third dataset, an excel file, "Crime by Neighborhood 2017" downloaded from the City of Fredericton Open Data site is found under the Public Safety domain.
- This dataset was then uploaded for the analysis.
- 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.

Data Exploration

- Exploring the count of crimes by neighbourhood gives us the first glimpse into the distribution.
- One note is the possibility neighbourhoods names could change at different times. 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.
- An example of data errors: There was an error found in the naming of the neighbourhood "Platt". The neighbourhood data stated "Plat" while the crime data stated "Platt". Given the crime dataset was most simple to manipulate it was modified to "Plat". The true name of the neighbourhood is "Platt".

Data Visualization

- Once the data was prepared, a choropleth map was created to view the crime count by neighborhood. As expected the region of greatest crime count was found in the downtown and Platt neighborhoods.
- Examining the crime types enables us to learn the most frequent occurring crimes which we then plot as a bar chart to see most frequent 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 neighborhood.

Data Visualization

- After exploring the pivot table showing Crime_Type by Neighborhood, we drill into a specific type of crime, theft from vehicles and plot the choropleth map to see which area has the greatest frequency.
- Again, the Platt neighborhood appears as the most frequent.
- Is this due to population density?
- Visualizing the population density enables us to determine that the Platt neighborhood has lower correlation to crime frequency than I would have expected.
- It would be interesting to further study the Census data and if this captures the population that is renting or more temporary/transient population, given the City is a University hub.

Data Visualization

- 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.
- 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:
  1) Neighborhoods in Fredericton
  2) Crime frequency by neighborhood
  3) Crime type frequency and statistics. The mean crime count in the City of Fredericton is 22.
  4) Crime type count by neighborhood:
     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 neighborhood.

Results

- Motor Vehicle crimes less than $5000 analysis by neighborhood 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 neighborhood in the City.
- The 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.
- 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 neighborhood. **The most common venues in the highest crime neighborhood are coffee shops followed by Pubs and Bars**.

Results

- 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

- 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 behavior is complex requiring thick profile data by individual and the conditions surrounding the event(s).

Discussion

- To be sufficient for reliable future prediction it would need to demonstrate validity, currency, reliability and sufficiency.
- A note of caution is the possibility neighborhoods names could change.
- The crime dataset did not mention which specific neighborhood naming dataset it was using but we assumed the neighborhood data provided aligned with the neighborhoods used in the crime data.
- It may be beneficial for the City to note and timestamp neighborhood naming in the future or simply reference with neighborhood naming file it used for the crime dataset.

Discussion

- Errors exist in the current open data. An error was found in the naming of the neighborhood "Platt". The neighborhood data stated "Plat" while the crime data stated "Platt".
- Given the crime dataset was most simple to manipulate it was modified to "Plat". The true name of the neighborhood is "Platt".
- 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.

Discussion

- 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.
- It would be interesting to further study the Census data and if this captures the population that is renting or more temporary/transient pollution, given the City is a University hub.
- Given the findings of the top 10 most frequent venues by locations of interest, the Knowledge Park does not have Coffee Shops in the top 10 most common venues as determined from the Foursquare dataset.
- Given this area has the greatest concentration of stores and shops as venues, it would be safe to assume a coffee shop would be beneficial to the business community and the citizens of Fredericton.

Conclusion

- 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.
- The Open Data project is a great start and empowers the need for a "Citizens Like Me" model to be developed where citizens of digital Fredericton are able to share their data as they wish for detailed analysis that enables the creation of valued services.
- More details are available in the report.