Segmenting and Clustering Neighborhoods in Frederiction, NB

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

Out[73]:

Data

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/

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.

Methodology

All steps are referenced beleow in the Appendix: Analysis section.

The methodology will include:

1. Loading each data set
2. Examine the crime frequency by neighbourhood
3. Study the crime types and then pivot analysis of crime type frequency by neighbourhood
4. Understand correlation between crimes and population density
5. Perform k-means statisical analysis on venues by locations of interest based on findings from crimes and neighbourhood
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.

Results

The analysis enabled us to discover and describe visually and quantitatively:

1. Neighbourhoods in Fredericton

2. Crime freqency by neighbourhood

3. Crime type frequency and statistics. The mean crime count in the City of Fredericton is 22.

4. 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 deterant for motor vehicle crimes in the downtown core compared to low surveillance in the Platt neighbourhood.

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 occuring 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 neighbhourhoods, 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 quantitiatve 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.

What is the crime count by neighbourhood?

```
In [128]:  crime_data = crime_df.groupby(['Neighbourhood']).size().to_frame(name='Count').rese
           t_index()
           crime_data
```

Out[128]:

	Neighbourhood	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
16	Garden Place	4
17	Gilridge Estates	3
18	Golf Club	7
19	Grasse Circle	1
20	Greenwood Minihome Park	2
21	Hanwell North	8
22	Heron Springs	3
23	Highpoint Ridge	5

```
In [89]:  world_geo = r'world_countries.json' # geojson file

          fredericton_1_map = folium.Map(location=[45.97, -66.65], width=1000, height=750,zoo
          m_start=12)

          fredericton_1_map
```

Out[89]:

```
In [90]:    fredericton_geo = r.json()

            threshold_scale = np.linspace(crime_data['Crime_Count'].min(),crime_data['Crime_Cou
            nt'].max(), 6,dtype=int)
            threshold_scale = threshold_scale.tolist()
            threshold_scale[-1] = threshold_scale[-1]+1

            fredericton_1_map.choropleth(geo_data=fredericton_geo, data=crime_data,columns=['Ne
            ighbourh', 'Crime_Count'],
                key_on='feature.properties.Neighbourh', threshold_scale=threshold_scale,fill_co
            lor='YlOrRd', fill_opacity=0.7,
                line_opacity=0.1, legend_name='Fredericton Neighbourhoods')

            fredericton_1_map
```

Out[90]:
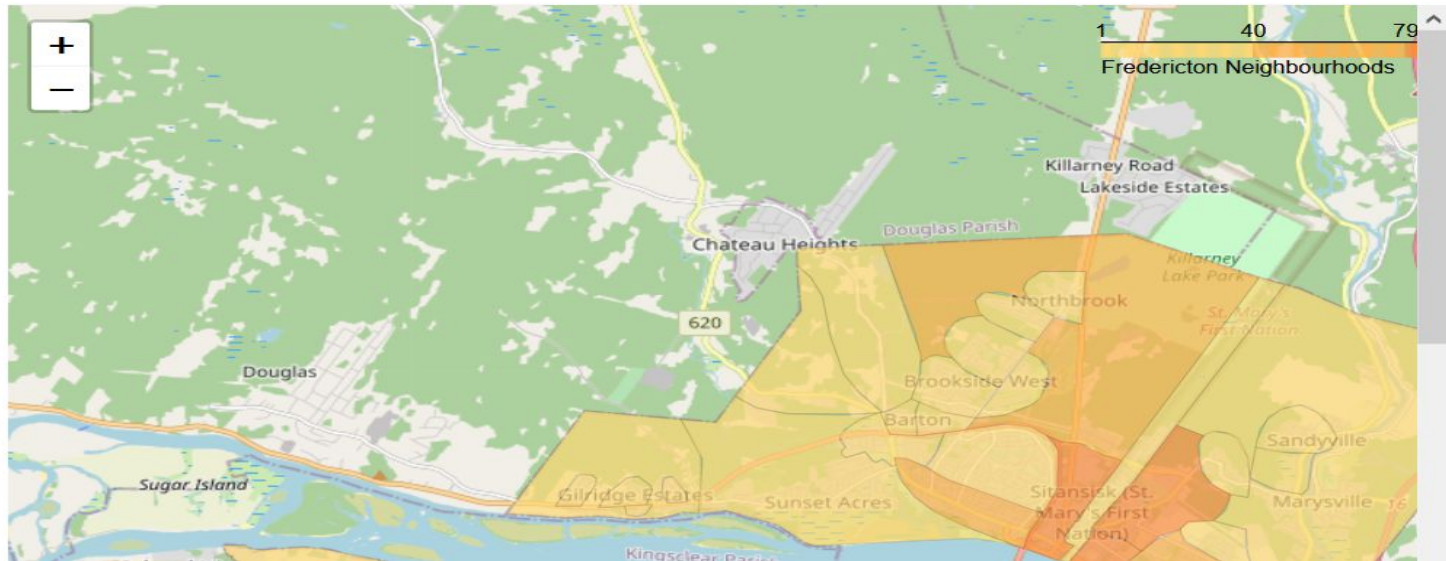
Examine Crime Types

```
In [131]: crimetype_data = crime_df.groupby(['Crime_Type']).size().to_frame(name='Count').res
          et_index()
          crimetype_data
```

Out[131]:

|	Crime_Type	Count
0		4
1	ARSON	5
2	ARSON BY NEG	1
3	ARSON-DAM.PROP.	4
4	B&E NON-RESIDNCE	51
5	B&E OTHER	58
6	B&E RESIDENCE	151
7	B&E STEAL FIREAR	3
8	MISCHIEF OBS USE	1
9	MISCHIEF TO PROP	246
10	MISCHIEF-DATA	2
11	MOTOR VEH THEFT	40
12	THEFT BIKE<$5000	63
13	THEFT FROM MV < $5000	356
14	THEFT FROM MV > $5000	5
15	THEFT OTH <$5000	458
16	THEFT OTH >$5000	9
17	THEFT OVER $5000	1
18	THEFT,BIKE>$5000	2

```
In [97]: ## Motor Vehicle Crime <$5000 Count
         fredericton_geo = r.json()
         threshold_scale = np.linspace(mvcrime_data['MVCrime_Count'].min(), mvcrime_data['MV
         Crime_Count'].max(),6,dtype=int)
         threshold_scale = threshold_scale.tolist()
         threshold_scale[-1] = threshold_scale[-1]+1

         fredericton_c_map.choropleth(geo_data=fredericton_geo,data=mvcrime_data,columns=['N
         eighbourh', 'MVCrime_Count'],key_on='feature.properties.Neighbourh',
             threshold_scale=threshold_scale, fill_color='YlOrRd',fill_opacity=0.7,line_opac
         ity=0.1,legend_name='Fredericton Neighbourhoods')
         fredericton_c_map
```

Out[97]:
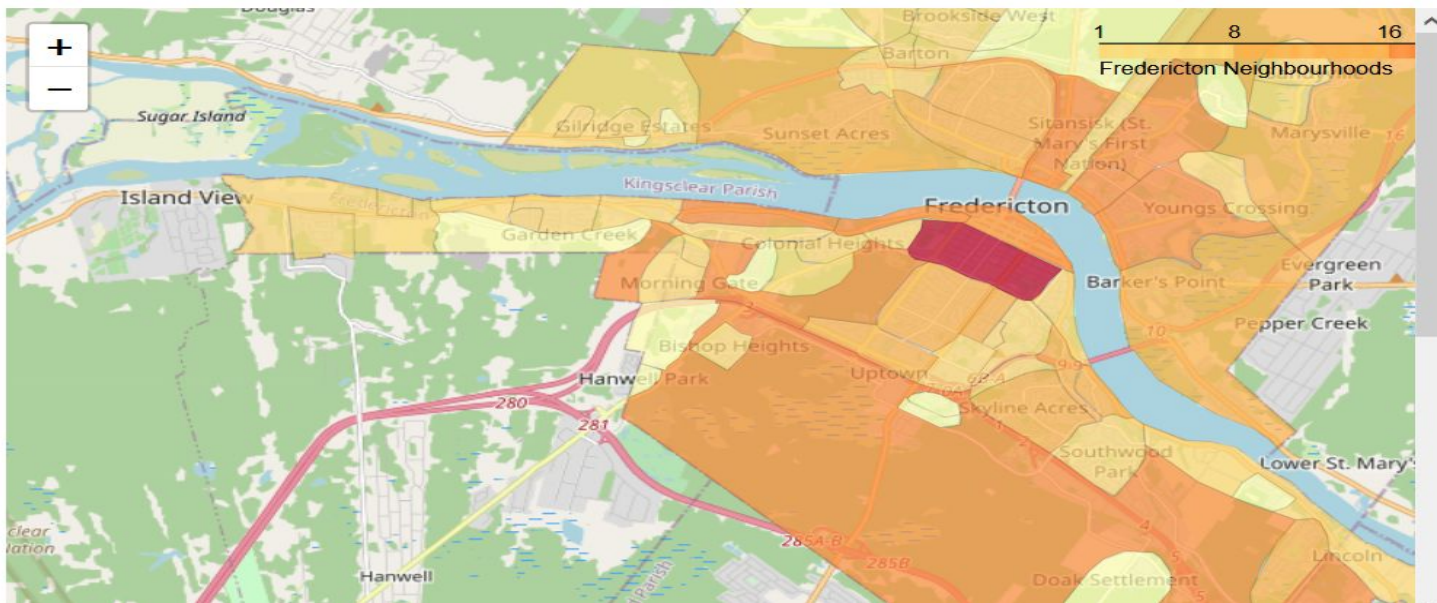
```
In [100]: # Population Density
          world_geo = r'world_countries.json' # geojson file
          fredericton_d_map = folium.Map(location=[45.94, -66.63], width=1200, height=750,zoo
          m_start=12)
          fredericton_d_map

          threshold_scale = np.linspace(demog_df['DBpop2011'].min(),demog_df['DBpop2011'].max
          (),6,dtype=int)
          threshold_scale = threshold_scale.tolist()
          threshold_scale[-1] = threshold_scale[-1]+1

          fredericton_d_map.choropleth(geo_data=demog_geo,data=demog_df,columns=['OBJECTID','
          DBpop2011'],key_on='feature.properties.OBJECTID',
              threshold_scale=threshold_scale,fill_color='PuBuGn',fill_opacity=0.7, line_opac
          ity=0.1,legend_name='Fredericton Population Density')
          fredericton_d_map
```

Out[100]: