Battle Of Neighborhoods- Windsor, Ontario

BY SHREYA

Windsor, South-Western Ontario

Introduction: Business Problems

Windsor is a city in southwestern Ontario, Canada, on the south bank of the Detroit River directly across from Detroit, Michigan. Located in Essex County, it is the southernmost city in Canada and marks the southwestern end of the Quebec City–Windsor Corridor. The city's population was 217,188 at the 2016 census, making it the third-most populated city in Southwestern Ontario, after London and Kitchener. The Detroit–Windsor urban area is North America's most populous trans-border conurbation, and the Ambassador Bridge border crossing is the busiest commercial crossing on the Canada–United States border.

Windsor is a major contributor to Canada's automotive industry and is culturally diverse. Known as the "Automotive Capital of Canada", Windsor's industrial and manufacturing heritage is responsible for how the city has developed through the years. Owing to this virtues it offers a wide spectrum of venues and is a governement, university and cultural hub.

As the city grows and develops, it becomes increasingly important to examine and understand it quantitiatively. The City of Windsor provides open data for everyone and encourages entrepreneurial use to develop services for the benefit of its ciitzens.

Developers, policy makers and/or city planners have an interest in answering the following questions:

1) What neighbourhoods 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?

5) What other landmarks and facilities can be useful?

Data

To understand and explore we will need the following City of Windsor Open Data:

Open Data Site: http://data-windsor.opendata.arcgis.com/

Windsor Neighbourhoods: http://data-windsor.opendata.arcgis.com/datasets/neighbourhoods--quartiers

Windsor Crime by Neighbourhood: http://data-windsor.opendata.arcgis.com/datasets/crime-by-neighbourhood-2017--crime-par-quartier-2017

Fredericton Census Tract Demographics: http://data-windsor.opendata.arcgis.com/datasets/census-tract-demographics--donn%C3%A9es-d%C3%A9mographiques-du-secteur-de-recensement

Windsor locations of interest: https://github.com/old-school-kid/Coursera_Capstone/blob/master/Windsor%20Locations.xlsx

Foursquare Developers Access to venue data: https://foursquare.com/

Using this data will allow exploration and examination to answer the questions asked above. 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. Windsor 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.

Loading the data

After loading the applicable libraries, the referenced geojson neighbourhood data was loaded from the City of Windsor 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 Neighbourhood 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.

The third dataset, an excel file, "Crime by Neighbourhood 2017" downloaded from the City of Windsor 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 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

Exploring the data

Exploring the count of crimes by neighborhood gives us the first glimpse into the distribution.

One note is the possibility neighborhoods names could change at different times. 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.

An example of data errors: There was an error 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".
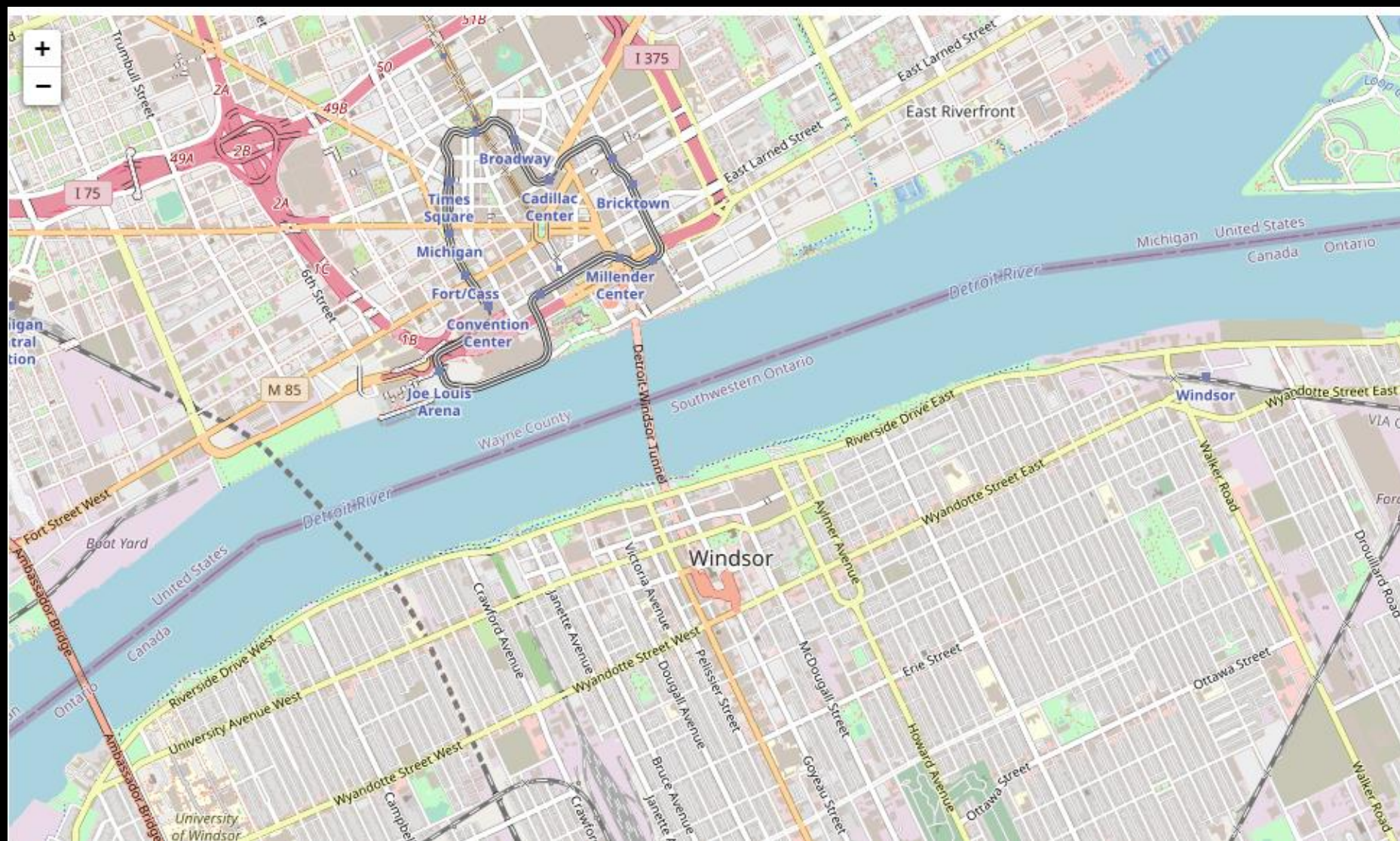
The map of Winsdor, Ontario

First Visualization Of Crime

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 occuring crimes which we then plot as a bar chart to see most frequenty 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 deterant for motor vehicle crimes in the downtown core compared to low surveillance in the Platt neighbourhood.

Crime per Neighbourhood
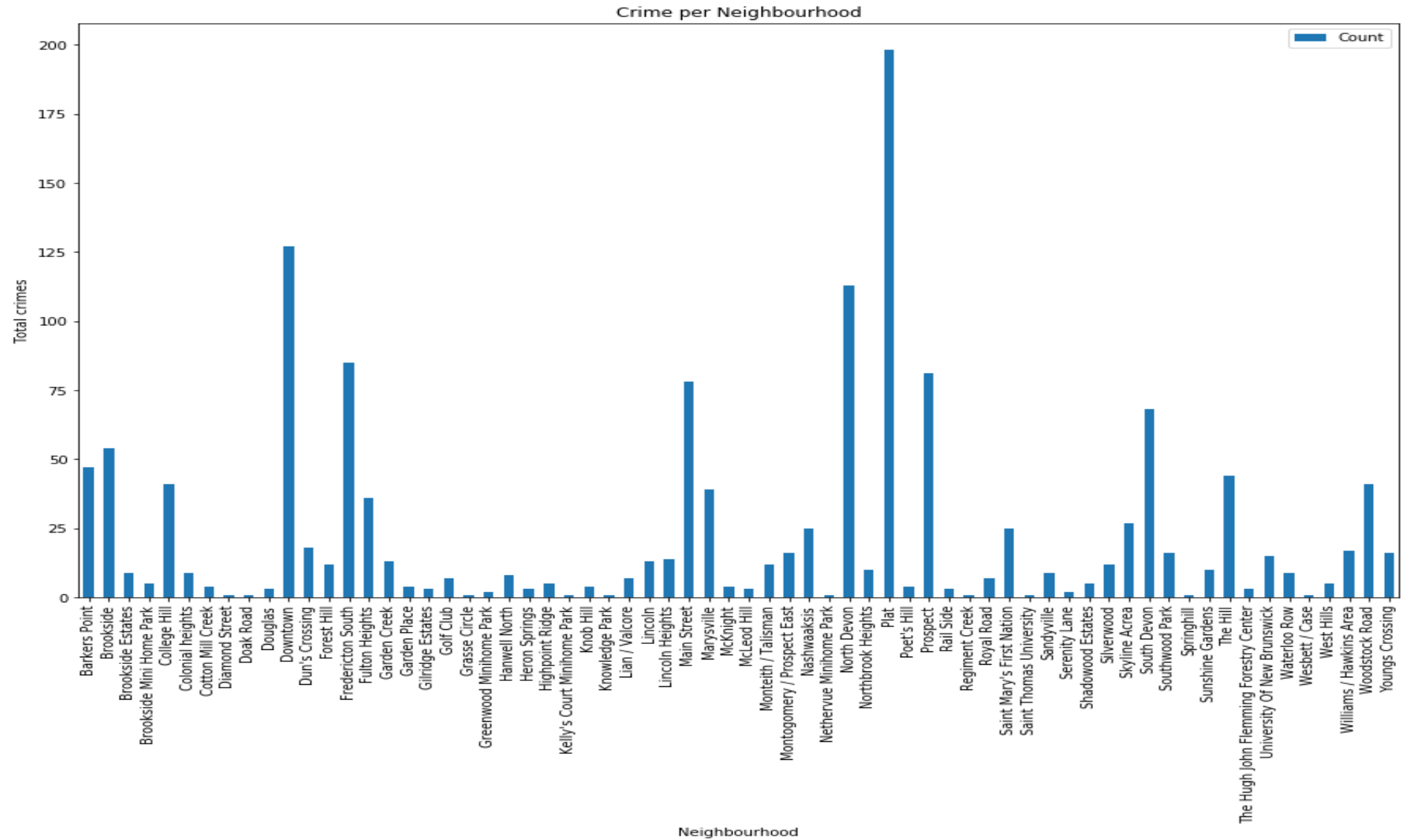
Choropleth map showing Crime Count of the neighborhoods
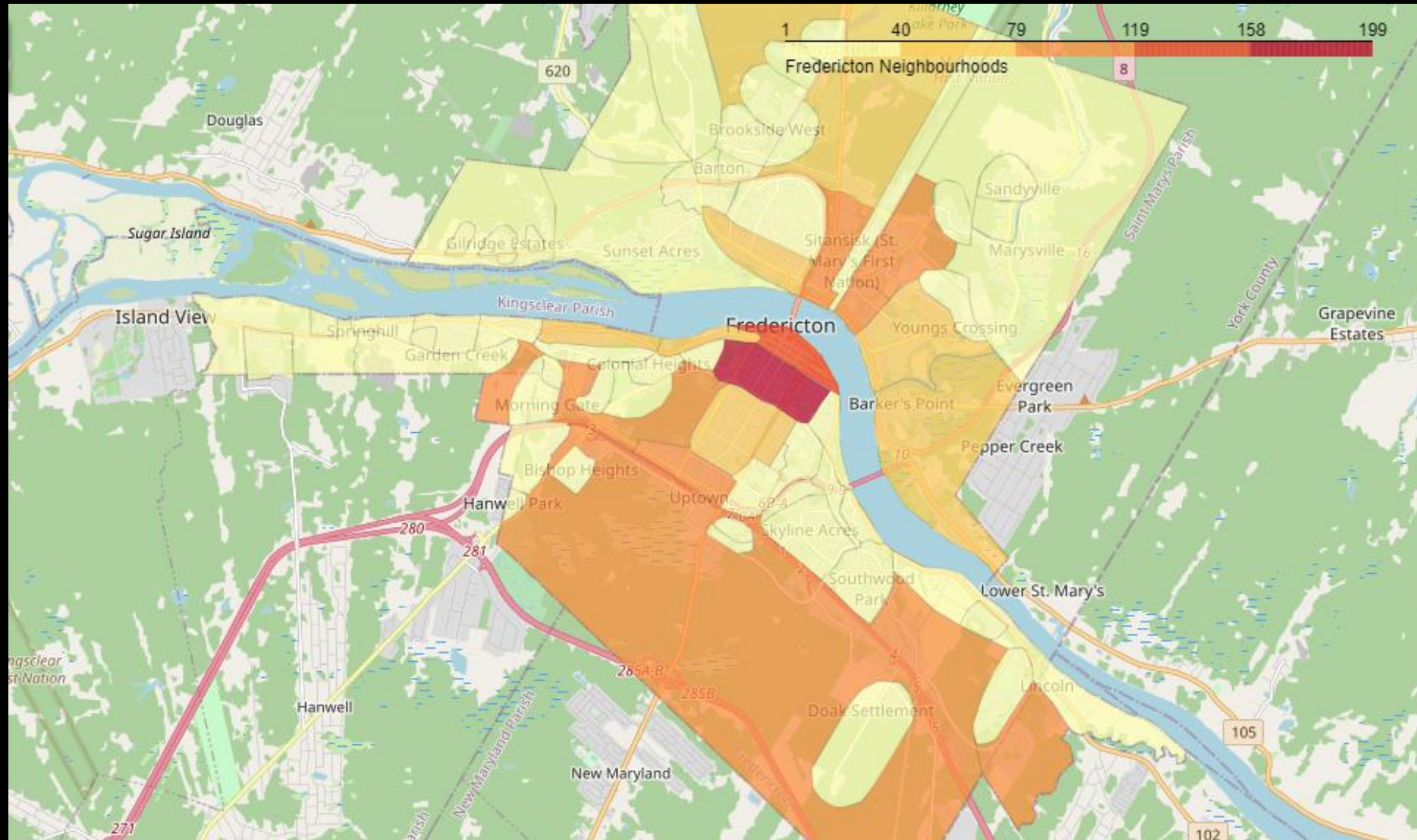
Importing data from Foursquare API for specific locations

Loading the "Windsor 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.

The Specific Locations

	Location	Latitude	Longitude
0	Ambassador Bridge	42.3072	-83.0711
1	Windsor Riverfront	42.3202	-83.0387
2	Peche Island	42.3455	-82.9278
3	Coventry Islands	42.3294	-82.9791
4	Sand Point Beach	42.3379	-82.9183
5	University of Windsor	42.3043	-83.0660
6	Saint Claire College	42.2485	-83.0203
7	Devenwood Conservation Area	42.2644	-82.9803
8	Jackson Park	42.2951	-83.0227
9	Little River Corridor	42.3243	-82.9246
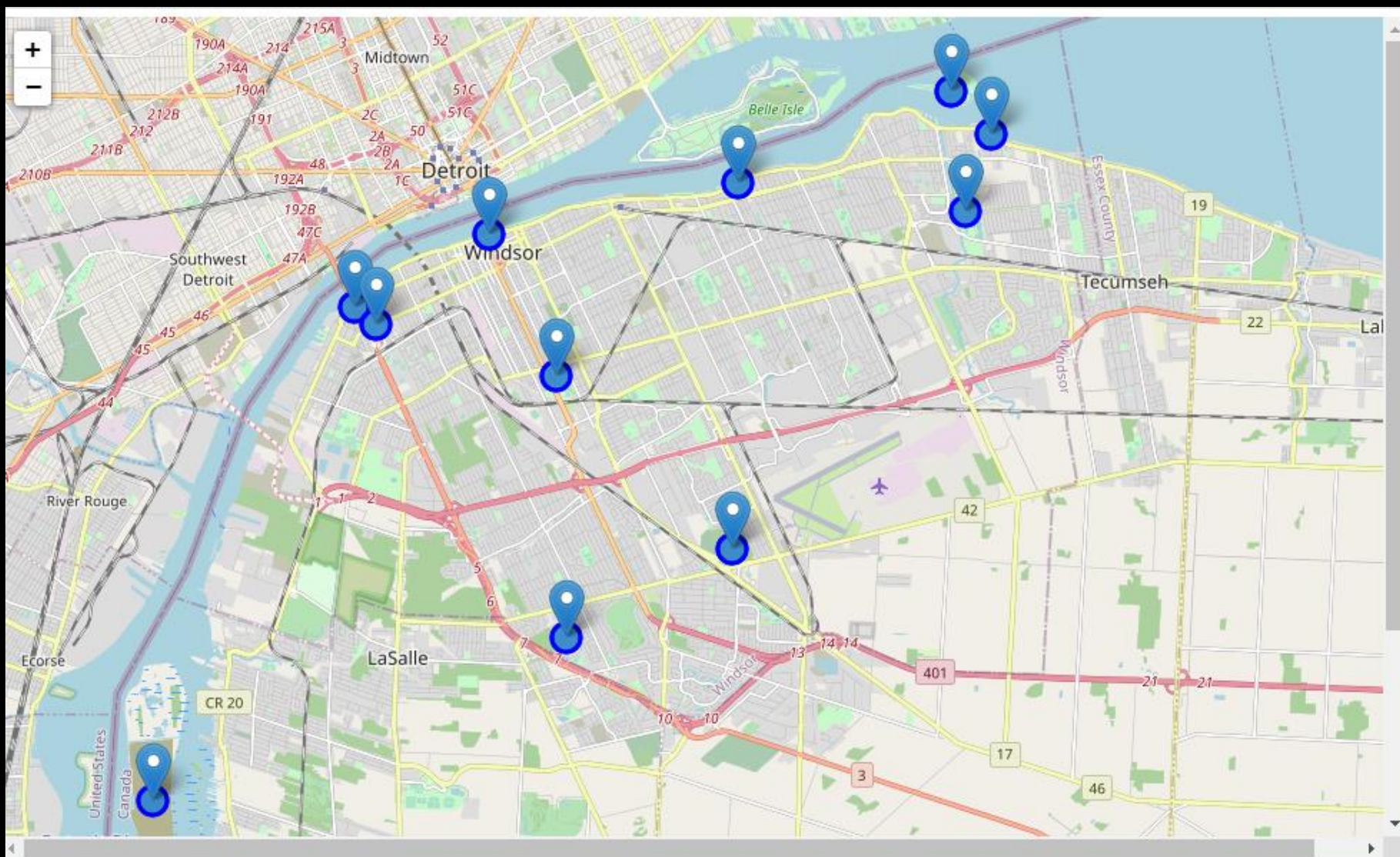
Analyzing the data from Foursquare

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

```
----Peche Island----
                          venue  freq
0               Harbor / Marina  0.50
1                         Trail  0.25
2             Seafood Restaurant  0.25
3                   Men's Store  0.00
4                          Park  0.00
5   Paper / Office Supplies Store  0.00


----Saint Claire College----
                          venue  freq
0   Construction & Landscaping   0.29
1                  Home Service  0.29
2                   Pizza Place  0.14
3                   Coffee Shop  0.14
4                           Pub  0.14
5                      Pharmacy  0.00


----Sand Point Beach----
                   venue  freq
0                  Trail  0.33
1         Ice Cream Shop  0.17
2      Athletics & Sports  0.17
3                  Beach  0.17
4     Seafood Restaurant  0.17
5      African Restaurant  0.00


----University of Windsor----
                 venue  freq
0     Convenience Store  0.10
1           Pizza Place  0.10
2           Coffee Shop  0.08
3            Restaurant  0.05
4      Asian Restaurant  0.05
5                   Pub  0.05


----Windsor Riverfront----
                venue  freq
0                  Bar  0.07
1                Hotel  0.06
2          Coffee Shop  0.06
3                 Café  0.04
4               Lounge  0.04
5    Italian Restaurant  0.03
```

```
----Ambassador Bridge----
                 venue  freq
0           Pizza Place  0.09
1     Convenience Store  0.09
2           Coffee Shop  0.09
3      Asian Restaurant  0.06
4                   Pub  0.06
5            Restaurant  0.06


----Coventry Islands ----
                  venue  freq
0                 Garden  0.09
1                 Museum  0.09
2    Sporting Goods Shop  0.04
3          Bowling Alley  0.04
4                  Plaza  0.04
5    Fast Food Restaurant  0.04


----Devenwood Conservation Area----
                          venue  freq
0                          Hotel  0.13
1          Furniture / Home Store  0.13
2                 Ice Cream Shop  0.07
3                    Gas Station  0.07
4    Paper / Office Supplies Store  0.07
5                    Coffee Shop  0.07


----Jackson Park----
                  venue  freq
0               Pharmacy  0.07
1    Fast Food Restaurant  0.07
2      Gym / Fitness Center  0.07
3               Pet Store  0.04
4                    Gym  0.04
5             Coffee Shop  0.04


----Little River Corridor----
                    venue  freq
0             Coffee Shop  0.17
1               Gym Pool  0.17
2                   Trail  0.17
3                    Park  0.17
4           Knitting Store  0.17
5     General Entertainment  0.17
```

Mean of the Frequency of Occurrence of each Category for each Location

	Location	African Restaurant	American Restaurant	Aquarium	Art Gallery	Asian Restaurant	Athletics & Sports	Bakery	Bank	Bar	Beach	Beer Garden	Beer Store	Boat or Ferry	Boc
0	Ambassador Bridge	0.000	0.029412	0.000000	0.000000	0.058824	0.000000	0.000000	0.029412	0.029412	0.000000	0.029412	0.000000	0.000000	0.
1	Coventry Islands	0.000	0.000000	0.043478	0.000000	0.000000	0.000000	0.000000	0.043478	0.000000	0.000000	0.000000	0.043478	0.000000	0.
2	Devenwood Conservation Area	0.000	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
3	Jackson Park	0.000	0.037037	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
4	Little River Corridor	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
5	Peche Island	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
6	Saint Claire College	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
7	Sand Point Beach	0.000	0.000000	0.000000	0.000000	0.000000	0.166667	0.000000	0.000000	0.000000	0.166667	0.000000	0.000000	0.000000	0.
8	University of Windsor	0.025	0.025000	0.000000	0.000000	0.050000	0.000000	0.000000	0.025000	0.025000	0.000000	0.025000	0.025000	0.000000	0.
9	Windsor Riverfront	0.000	0.000000	0.000000	0.011236	0.000000	0.011236	0.011236	0.011236	0.067416	0.000000	0.000000	0.011236	0.011236	0.

|	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ambassador Bridge	Coffee Shop	Pizza Place	Convenience Store	Restaurant	Asian Restaurant	Park	Pub	Sandwich Place	Middle Eastern Restaurant	Duty-free Shop
1	Coventry Islands	Garden	Museum	Harbor / Marina	Sandwich Place	Food Truck	Fast Food Restaurant	English Restaurant	Mexican Restaurant	Coffee Shop	Grocery Store
2	Devenwood Conservation Area	Furniture / Home Store	Hotel	Paper / Office Supplies Store	Ice Cream Shop	Rental Car Location	Discount Store	Department Store	Miscellaneous Shop	Sandwich Place	Home Service
3	Jackson Park	Pharmacy	Gym / Fitness Center	Fast Food Restaurant	Business Service	Convenience Store	Pizza Place	Pet Store	Park	Coffee Shop	Middle Eastern Restaurant
4	Little River Corridor	Coffee Shop	Park	General Entertainment	Knitting Store	Gym Pool	Trail	Art Gallery	Food	Deli / Bodega	Department Store
5	Peche Island	Harbor / Marina	Trail	Seafood Restaurant	Farmers Market	Construction & Landscaping	Convenience Store	Deli / Bodega	Department Store	Discount Store	Duty-free Shop
6	Saint Claire College	Home Service	Construction & Landscaping	Pizza Place	Pub	Coffee Shop	Whisky Bar	Ethiopian Restaurant	Convenience Store	Deli / Bodega	Department Store
7	Sand Point Beach	Trail	Ice Cream Shop	Athletics & Sports	Beach	Seafood Restaurant	Whisky Bar	Fast Food Restaurant	Convenience Store	Deli / Bodega	Department Store
8	University of Windsor	Pizza Place	Convenience Store	Coffee Shop	Asian Restaurant	Tea Room	Middle Eastern Restaurant	Restaurant	Pub	Fast Food Restaurant	Japanese Restaurant
9	Windsor Riverfront	Bar	Hotel	Coffee Shop	Café	Lounge	Italian Restaurant	Burger Joint	Restaurant	Middle Eastern Restaurant	Outdoor Sculpture

Results

The analysis helped us discover visually and quantitatively Neighborhoods in Fredericton

1) Crime frequency by neighborhood

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

3) 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 neighborhood.

4) **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.

5) 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.

6) 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.

7) 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.

8) We were able to determine the top 10 most common venues by location of interest.

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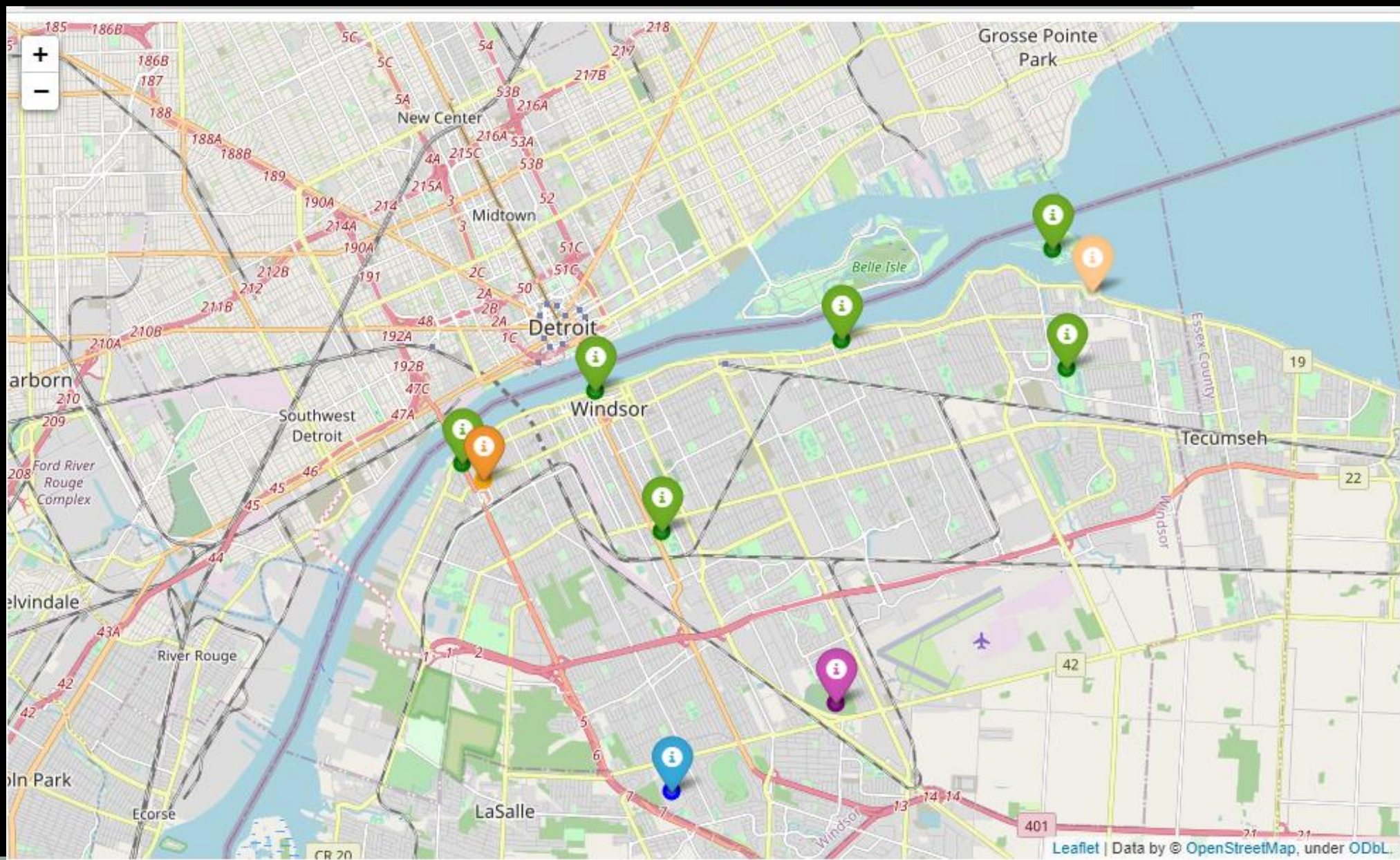
Conclusion

Using a combination of datasets from the City of Windsor 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 Windsor 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 Windsor are able to share their data as they wish for detailed analysis that enables the creation of valued services.

Reference

Google Map

Foursquare API

Wikipedia for Windsor

Github Repository