

Toronto neighbourhoods

Analysis of the venue types vs crime data in Toronto

Alexander Oltu - 3 May 2020 - Version 1.0

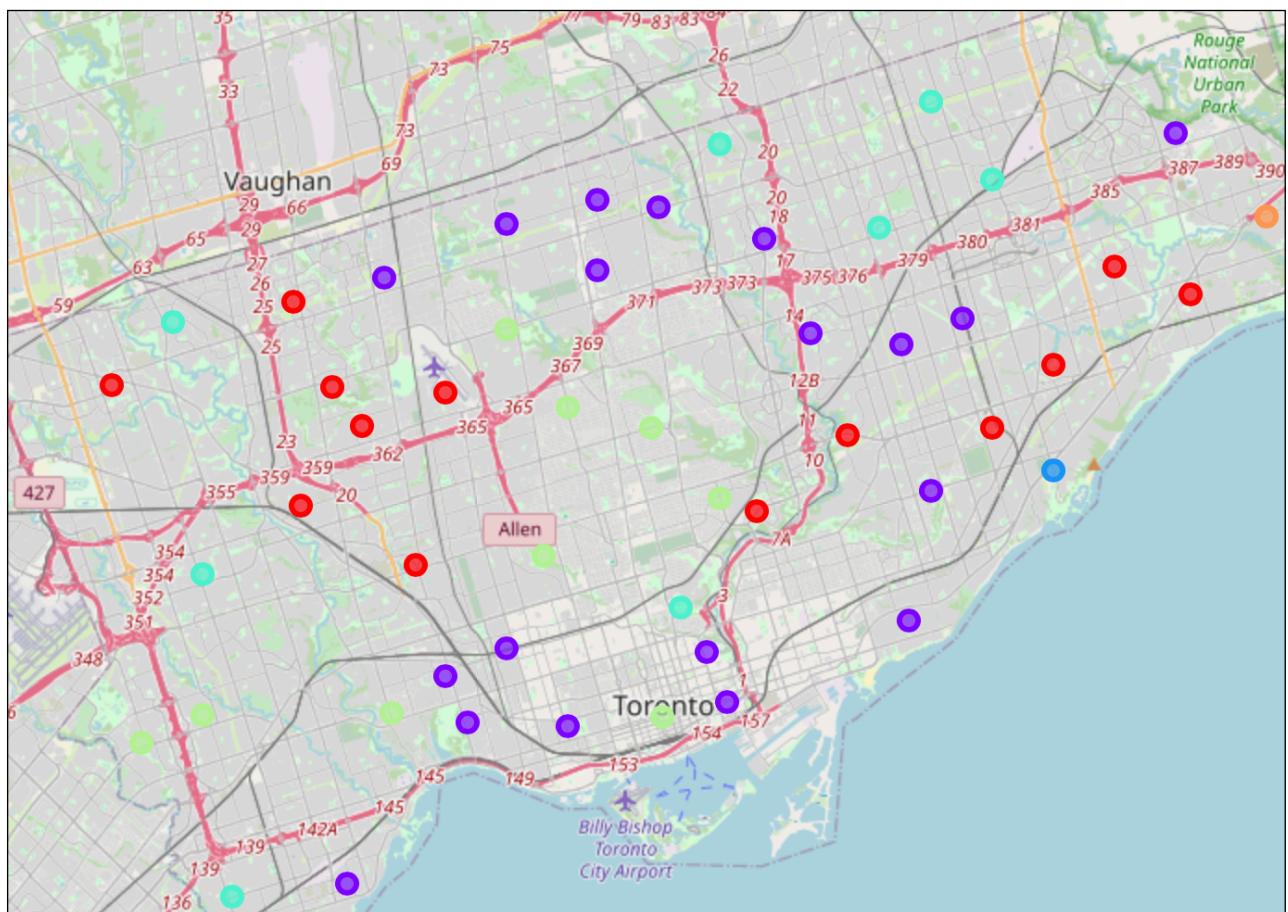


Table of Contents

Table of Contents	2
Introduction	3
Data	3
Methodology	4
Results	6
Cluster 0.....	6
Cluster 1.....	8
Cluster 2.....	9
Cluster 3.....	10
Cluster 4.....	10
Cluster 5.....	11
Discussion.....	12
Conclusion	12

Introduction

A new physical security firm would like to start their service in Toronto. In order to understand risks and offer adequate pricing based on the risks, they would like to have an overview over type of services per neighbourhood and their associate crime rates. The prices will depend based on the risks associated and venue type. They also would like to group those to keep price types to minimum.

The target audience for this report is this security firm which is going to start its business in Toronto.

Data

The following data sources will be used in this study:

1. The neighbourhoods information will be used from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. In order to find information about venues we will use Foursquare API.
3. Crime data will be used from Open Data of Toronto Police Public Service <http://data.torontopolice.on.ca/datasets/>, MCI 2014 to 2019, which states for Major Crime Indicators and includes 5 main categories of crime.

As client would like to minimise type of pricing, we will cluster neighbourhoods and crime data by using k-means.

Methodology

To define neighbourhoods we've used Wikipedia Postal Codes data with the corresponding neighbourhoods. This data was used as a master neighbourhood data with the GPS coordinates of its centroids. Resulted data frame looked like this, with the 103 postal codes:

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

The crime data with the major crime indicators has 5 main crime categories:

- * Assault
- * Break and Enter
- * Robbery
- * Theft Over
- * Auto Theft

In this data Assault prevails with over 50% of all the crime incidents, Break and Enter is 2nd with about 20% of the data. All crime data had over 206 000 records in total. We mapped crime data neighbourhood fields with the postal codes data and resulted data frame was reduced to about 88 000 records. We clearly were not able to map some of the neighbourhoods, quick checks haven't revealed something obvious and improving this mapping was left to the future improvements. The resulted data frame looked like this:

	MCI	Crime Neighbourhood	Crime Lon	Crime Lat	Neighbourhood
1	Assault	Tam O'Shanter-Sullivan (118)	-79.307907	43.778732	Clarks Corners, Tam O'Shanter, Sullivan
2	Break and Enter	Woburn (137)	-79.225029	43.765942	Woburn
12	Assault	Downsview-Roding-CFB (26)	-79.508636	43.720917	Downsview
13	Break and Enter	Bedford Park-Nortown (39)	-79.416420	43.735794	Bedford Park, Lawrence Manor East
14	Assault	Malvern (132)	-79.209801	43.814213	Malvern, Rouge

Next step was to get venue data via Foursquare API. For each of the neighbourhoods centroids we've downloaded all* the venues in 1km radius. The resulted

data frame had 325 venue categories, a bit over 4 800 venues and looked like this:

Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0 Parkwoods	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
1 Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
2 Parkwoods	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	Café
3 Parkwoods	43.753259	-79.329656	A&W	43.760643	-79.326865	Fast Food Restaurant
4 Parkwoods	43.753259	-79.329656	Bruno's valu-mart	43.746143	-79.324630	Grocery Store

Next was to hot encode venue and crime data in each neighbourhood based on the frequency of occurrence of the venue category or crime category and merge these together.

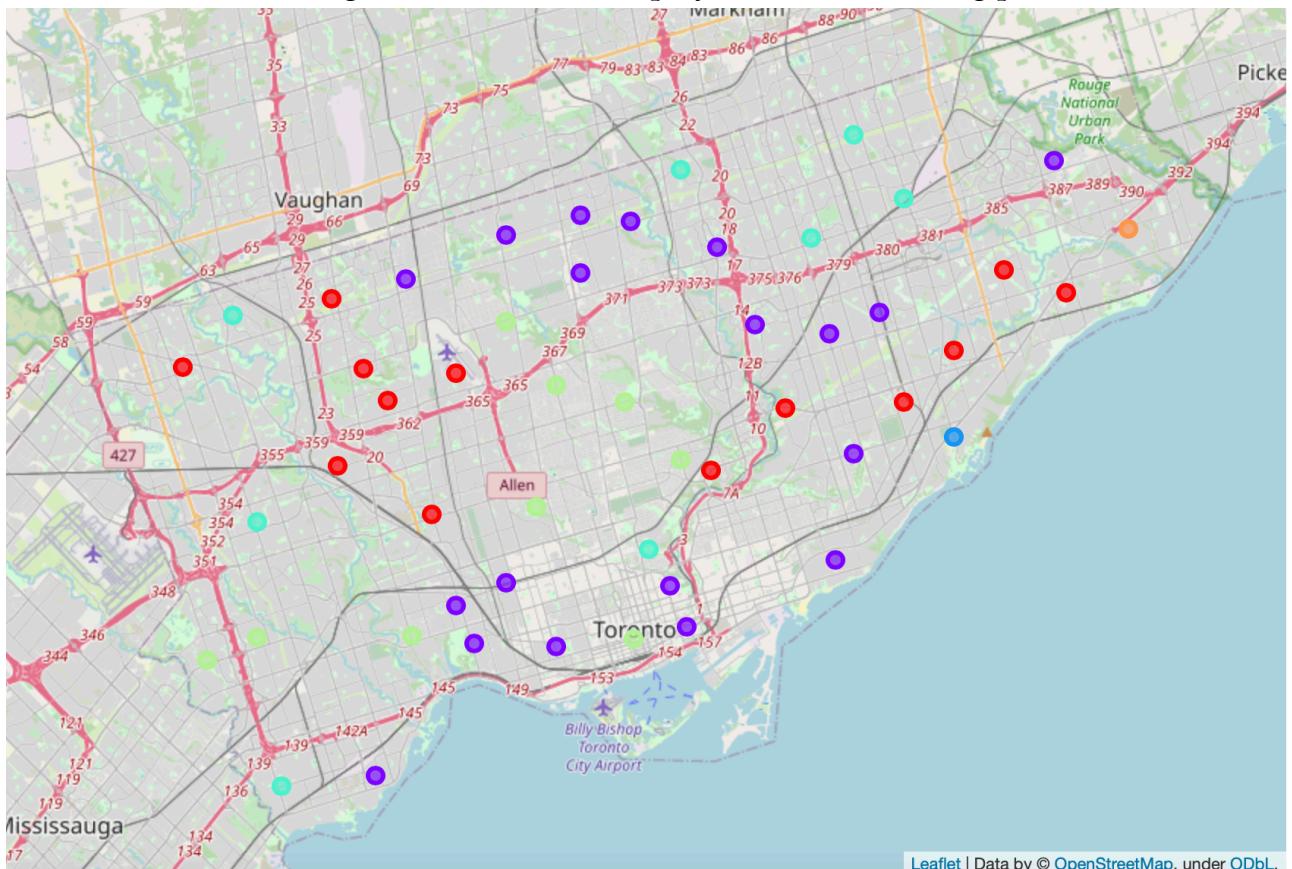
Before grouping we had 97 neighbourhoods with the venue data and 47 neighbourhoods with the crime data. We added venue data only to the neighbourhoods with the crime data as it was important to do clustering including both crime and venue data. The resulted dataset had 47 neighbourhoods and looked like this:

Neighbourhood	Assault	Auto Theft	and Enter	Robbery	Theft Over	ATM	Accessories Store	Afghan Restaurant	Airport	...	Video Store	Vietnamese Restaurant	Warehouse Store	Whisky Bar	Wine Bar
0 Agincourt	0.428485	0.133333	0.290303	0.099394	0.048485	0.000000	0.0	0.000000	0.000000	...	0.000000	0.021739	0.000000	0.00	0.00 0
1 Alderwood, Long Branch	0.462992	0.134646	0.258268	0.100787	0.043307	0.000000	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.00	0.00 0
2 Bathurst Manor, Wilson Heights, Downsview North	0.416781	0.220083	0.258597	0.077029	0.027510	0.000000	0.0	0.000000	0.000000	...	0.037037	0.000000	0.000000	0.00	0.00 0
3 Bayview Village	0.496224	0.132686	0.259978	0.057174	0.053937	0.000000	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.00	0.00 0
4 Bedford Park, Lawrence Manor East	0.212903	0.221774	0.447581	0.062903	0.054839	0.000000	0.0	0.000000	0.000000	...	0.023810	0.000000	0.000000	0.00	0.00 0
5 Clarks Corners, Tam O'Shanter, Sullivan	0.479942	0.100656	0.274252	0.121080	0.024070	0.000000	0.0	0.000000	0.000000	...	0.000000	0.027027	0.000000	0.00	0.00 0

The last step was to apply clustering to this data and analyse results. We used K-means and grouped data to 6 clusters.

Results

Result of clustering with K-Means is displayed on the following picture:



Red - cluster 0

Violet - cluster 1

Blue - cluster 2

Cyan - cluster 3

Green - cluster 4

Orange - cluster 5

Cluster 0

The following neighbourhoods are in cluster 0:

Postal Code	Neighbourhood
M4A	Victoria Village
M1E	Guildwood, Morningside, West Hill
M1G	Woburn
M4H	Thorncliffe Park
M1J	Scarborough Village

M1K	Kennedy Park, Ionview, East Birchmount Park									
M3K	Downsvie									
M3L	Downsvie									
M3M	Downsvie									
M6M	Del Ray, Mount Dennis, Keelsdale and Silverthorn									
M3N	Downsvie									
M9N	Weston									
M9V	South Steeles, Silverstone, Humbergate, Jamest...									

This cluster has the following distribution of category types according to their popularity.

Neighbourhood	1st Most Common	2nd Most Common	3rd Most Common	4th Most Common	5th Most Common	6th Most Common	7th Most Common	8th Most Common	9th Most Common	10th Most Common
Victoria Village	Assault	Break and Enter	Coffee Shop	Auto Theft	Sporting Goods Shop	Gym / Fitness Center	Grocery Store	Intersection	Golf Course	Lounge
Guildwood, Morningside, West Hill	Assault	Pizza Place	Break and Enter	Robbery	Coffee Shop	Fast Food Restaurant	Bank	Auto Theft	Juice Bar	Pharmacy
Woburn	Assault	Park	Coffee Shop	Break and Enter	Robbery	Fast Food Restaurant	Chinese Restaurant	Mobile Phone Shop	Indian Restaurant	Auto Theft
Thornclefe Park	Assault	Break and Enter	Coffee Shop	Robbery	Indian Restaurant	Grocery Store	Auto Theft	Theft Over	Gym	Pizza Place
Scarborough Village	Assault	Ice Cream Shop	Break and Enter	Robbery	Pizza Place	Coffee Shop	Restaurant	Sandwich Place	Fast Food Restaurant	Bowling Alley
Kennedy Park, Ionview, East Birchmount Park	Assault	Discount Store	Break and Enter	Robbery	Chinese Restaurant	Coffee Shop	Grocery Store	Fast Food Restaurant	Auto Theft	Light Rail Station
Downsvie	Assault	Auto Theft	Break and Enter	Robbery	Vietnamese Restaurant	Coffee Shop	Hotel	Grocery Store	Park	Pizza Place
Downsvie	Assault	Auto Theft	Break and Enter	Robbery	Vietnamese Restaurant	Coffee Shop	Hotel	Grocery Store	Park	Pizza Place
Downsvie	Assault	Auto Theft	Break and Enter	Robbery	Vietnamese Restaurant	Coffee Shop	Hotel	Grocery Store	Park	Pizza Place
Del Ray, Mount Dennis, Keelsdale and Silverthorn	Assault	Furniture / Home Store	Robbery	Grocery Store	Break and Enter	Auto Theft	Coffee Shop	Fast Food Restaurant	Gas Station	Sandwich Place
Downsvie	Assault	Auto Theft	Break and Enter	Robbery	Vietnamese Restaurant	Coffee Shop	Hotel	Grocery Store	Park	Pizza Place
Weston	Assault	Break and Enter	Train Station	Coffee Shop	Auto Theft	Robbery	Pharmacy	Diner	Discount Store	Sandwich Place
South Steeles, Silverstone, Humbergate, Jamest...	Assault	Pizza Place	Grocery Store	Robbery	Auto Theft	Break and Enter	Pharmacy	Caribbean Restaurant	Beer Store	Fried Chicken Joint

Cluster 1

The following neighbourhoods are in cluster 1:

Postal Code	Neighbourhood
M3A	Parkwoods
M5A	Regent Park, Harbourfront
M1B	Malvern, Rouge
M4E	The Beaches
M6H	Dufferin, Dovercourt Village
M2J	Fairview, Henry Farm, Oriole
M3J	Northwood Park, York University
M6J	Little Portugal, Trinity
M2K	Bayview Village
M1L	Golden Mile, Clairlea, Oakridge
M2M	Willowdale, Newtonbrook
M2N	Willowdale
M1P	Dorset Park, Wexford Heights, Scarborough Town...
M6P	High Park, The Junction South
M1R	Wexford, Maryvale
M2R	Willowdale
M6R	Parkdale, Roncesvalles
M8V	New Toronto, Mimico South, Humber Bay Shores
M4X	St. James Town, Cabbagetown

With the following distribution:

Neighbourhood	1st Most Common	2nd Most Common	3rd Most Common	4th Most Common	5th Most Common	6th Most Common	7th Most Common	8th Most Common	9th Most Common	10th Most Common
Parkwoods	Assault	Break and Enter	Auto Theft	Park	Robbery	Bus Stop	Pharmacy	Shopping Mall	Convenience Store	Fish & Chips Shop
Regent Park, Harbourfront	Assault	Break and Enter	Coffee Shop	Robbery	Auto Theft	Diner	Café	Pub	Theater	Park
Malvern, Rouge	Assault	Break and Enter	Auto Theft	Fast Food Restaurant	Trail	Robbery	Gym	Bank	Bakery	Chinese Restaurant
The Beaches	Assault	Break and Enter	Robbery	Auto Theft	Pub	Coffee Shop	Pizza Place	Japanese Restaurant	Breakfast Spot	Beach
Dufferin, Dovercourt Village	Assault	Break and Enter	Robbery	Coffee Shop	Café	Park	Auto Theft	Convenience Store	Bakery	Bar

Fairview, Henry Farm, Oriole	Assault	Break and Enter	Clothing Store	Auto Theft	Coffee Shop	Robbery	Theft Over	Sandwich Place	Restaurant	Japanese Restaurant
Northwood Park, York University	Assault	Break and Enter	Auto Theft	Coffee Shop	Robbery	Furniture / Home Store	Pizza Place	Theft Over	Bar	Middle Eastern Restaurant
Little Portugal, Trinity	Assault	Break and Enter	Café	Auto Theft	Robbery	Bar	Restaurant	Vegetarian / Vegan Restaurant	Bakery	Italian Restaurant
Bayview Village	Assault	Break and Enter	Bank	Gas Station	Japanese Restaurant	Auto Theft	Chinese Restaurant	Park	Grocery Store	Restaurant
Golden Mile, Clairlea, Oakridge	Assault	Break and Enter	Robbery	Intersection	Auto Theft	Coffee Shop	Diner	Bus Line	Bakery	Convenience Store
Willowdale, Newtonbrook	Assault	Break and Enter	Auto Theft	Korean Restaurant	Café	Robbery	Pizza Place	Middle Eastern Restaurant	Coffee Shop	Diner
Willowdale	Assault	Break and Enter	Robbery	Auto Theft	Coffee Shop	Korean Restaurant	Pizza Place	Ramen Restaurant	Bubble Tea Shop	Japanese Restaurant
Dorset Park, Wexford Heights, Scarborough Town...	Assault	Break and Enter	Auto Theft	Robbery	Electronics Store	Coffee Shop	Restaurant	Furniture / Home Store	Bakery	Asian Restaurant
High Park, The Junction South	Assault	Break and Enter	Robbery	Auto Theft	Café	Coffee Shop	Bar	Convenience Store	Park	Italian Restaurant
Wexford, Maryvale	Assault	Break and Enter	Auto Theft	Grocery Store	Middle Eastern Restaurant	Robbery	Pizza Place	Burger Joint	Bakery	Intersection
Willowdale	Assault	Break and Enter	Robbery	Auto Theft	Coffee Shop	Korean Restaurant	Pizza Place	Ramen Restaurant	Bubble Tea Shop	Japanese Restaurant
Parkdale, Roncesvalles	Assault	Break and Enter	Robbery	Auto Theft	Café	Pizza Place	Coffee Shop	Sushi Restaurant	Bar	Bakery
New Toronto, Mimico South, Humber Bay Shores	Assault	Break and Enter	Robbery	Park	Café	Auto Theft	Fast Food Restaurant	Bakery	Fried Chicken Joint	Liquor Store
St. James Town, Cabbagetown	Assault	Break and Enter	Robbery	Park	Café	Diner	Restaurant	Gastropub	Japanese Restaurant	Auto Theft

Cluster 2

The following neighbourhoods are in cluster 2:

Postal Code	Neighbourhood
M1M	Cliffside, Cliffcrest, Scarborough Village West

With the following distribution:

Neighbourhood	1st Most Common	2nd Most Common	3rd Most Common	4th Most Common	5th Most Common	6th Most Common	7th Most Common	8th Most Common	9th Most Common	10th Most Common
Cliffside, Cliffcrest, Scarborough Village West	Assault	Pizza Place	Break and Enter	Ice Cream Shop	Beach	Robbery	Park	Cajun / Creole Restaurant	Burger Joint	Hardware Store

Cluster 3

The following neighbourhoods are in cluster 3:

Postal Code	Neighbourhood
M2H	Hillcrest Village
M9L	Humber Summit
M9R	Kingsview Village, St. Phillips, Martin Grove ...
M1S	Agincourt
M1T	Clarks Corners, Tam O'Shanter, Sullivan
M1V	Milliken, Agincourt North, Steeles East, L'Amo...
M4W	Rosedale
M8W	Alderwood, Long Branch

With the following distribution:

Neighbourhood	1st Most Common	2nd Most Common	3rd Most Common	4th Most Common	5th Most Common	6th Most Common	7th Most Common	8th Most Common	9th Most Common	10th Most Common
Hillcrest Village	Assault	Break and Enter	Auto Theft	Robbery	Pharmacy	Park	Coffee Shop	Convenience Store	Bakery	Recreation Center
Humber Summit	Assault	Auto Theft	Electronics Store	Break and Enter	Robbery	Pharmacy	Arts & Crafts Store	Shopping Mall	Italian Restaurant	Park
Kingsview Village, St. Phillips, Martin Grove ...	Assault	Auto Theft	Break and Enter	Pharmacy	Robbery	Mobile Phone Shop	Gas Station	Supplement Shop	Supermarket	Beer Store
Agincourt	Assault	Break and Enter	Auto Theft	Chinese Restaurant	Robbery	Shopping Mall	Theft Over	Bakery	Caribbean Restaurant	Pizza Place
Clarks Corners, Tam O'Shanter, Sullivan	Assault	Break and Enter	Robbery	Coffee Shop	Auto Theft	Sandwich Place	Intersection	Park	Bank	Fast Food Restaurant
Milliken, Agincourt North, Steeles East, L'Amo...	Break and Enter	Assault	Auto Theft	Chinese Restaurant	Robbery	Bakery	Pharmacy	Park	Noodle House	Fast Food Restaurant
Rosedale	Assault	Break and Enter	Coffee Shop	Grocery Store	Park	Robbery	Auto Theft	Metro Station	BBQ Joint	Bank
Alderwood, Long Branch	Assault	Break and Enter	Auto Theft	Discount Store	Pharmacy	Robbery	Convenience Store	Pizza Place	Park	Theft Over

Cluster 4

The following neighbourhoods are in cluster 4:

Postal Code	Neighbourhood
M9B	West Deane Park, Princess Gardens, Martin Grov...
M6C	Humewood-Cedarvale

M9C	Eringate, Bloordale Gardens, Old Burnhamthorpe...									
M4G	Leaside									
M3H	Bathurst Manor, Wilson Heights, Downsview North									
M5H	Richmond, Adelaide, King									
M5M	Bedford Park, Lawrence Manor East									
M4N	Lawrence Park									
M6S	Runnymede, Swansea									

With the following distribution:

Neighbourhood	1st Most Common	2nd Most Common	3rd Most Common	4th Most Common	5th Most Common	6th Most Common	7th Most Common	8th Most Common	9th Most Common	10th Most Common
West Deane Park, Princess Gardens, Martin Grove...	Assault	Auto Theft	Park	Break and Enter	Pizza Place	Robbery	Theft Over	Clothing Store	Mexican Restaurant	Hotel
Humewood-Cedarvale	Assault	Break and Enter	Auto Theft	Pizza Place	Coffee Shop	Robbery	Theft Over	Middle Eastern Restaurant	Soccer Stadium	Café
Eringate, Bloordale Gardens, Old Burnhamthorpe...	Assault	Break and Enter	Auto Theft	Robbery	Coffee Shop	Convenience Store	Cosmetics Shop	Pet Store	College Rec Center	Gas Station
Leaside	Break and Enter	Assault	Auto Theft	Robbery	Sporting Goods Shop	Coffee Shop	Theft Over	Grocery Store	Furniture / Home Store	Sports Bar
Bathurst Manor, Wilson Heights, Downsview North	Assault	Break and Enter	Auto Theft	Robbery	Coffee Shop	Bank	Pizza Place	Middle Eastern Restaurant	Sushi Restaurant	Fried Chicken Joint
Richmond, Adelaide, King	Assault	Break and Enter	Auto Theft	Robbery	Coffee Shop	Café	Hotel	Theft Over	Theater	Restaurant
Bedford Park, Lawrence Manor East	Break and Enter	Auto Theft	Assault	Coffee Shop	Italian Restaurant	Robbery	Theft Over	Bank	Restaurant	Sandwich Place
Lawrence Park	Break and Enter	Assault	Auto Theft	Café	Coffee Shop	Trail	Gym / Fitness Center	College Gym	College Quad	Park
Runnymede, Swansea	Assault	Break and Enter	Auto Theft	Robbery	Coffee Shop	Café	Pizza Place	Bakery	Pub	Italian Restaurant

Cluster 5

The following neighbourhoods are in cluster 5:

Postal Code	Neighbourhood
M1C	Rouge Hill, Port Union, Highland Creek

With the following distribution:

Neighbourhood	1st Most Common	2nd Most Common	3rd Most Common	4th Most Common	5th Most Common	6th Most Common	7th Most Common	8th Most Common	9th Most Common	10th Most Common
Rouge Hill, Port Union, Highland Creek	Assault	Break and Enter	Breakfast Spot	Playground	Burger Joint	Park	Italian Restaurant	Auto Theft	Robbery	Theft Over

Discussion

Based on the distribution and clustering, we suggest the following clusters:

- * Cluster 0 - is dominated by Assault with some Break & Enter, it is Parks, Discount Stores and Pizza Places
- * Cluster 1 - is dominated by Assault, Break & Enter and Robbery. Those are Coffee shops, Banks, Restaurants
- * Cluster 2 - is dominated by Assaults in Pizza Places.
- * Cluster 3 - is dominated by Assault, Auto Thief and Break & Enter in Electronics shops, Coffee shops and Asian restaurants
- * Cluster 4 - is primarily Break & Enter, Assault in Sporting shops, Coffee shops and restaurants
- * Cluster 5 - is Assault and sometimes Break & Enter in Breakfast spots, playgrounds and Burger places

We suggest 6 price categories to begin with. The pricing can be based on the description of the cluster above.

Conclusion

Current study is covering only 47 neighbourhoods and is not taking into account higher or lower criminal rates per neighbourhood into account. Including rest of the neighbourhoods and adjusting weight towards higher crime rate neighbourhoods can improve pricing categorisation. Due to the time constraints of this project it was not possible to perform such studies.

A following followup to improve categorisation is suggested for the project:

1. There are 140 neighbourhoods in total in Crime data and 97 neighbourhoods with the venue data. Due to naming differences in Crime data we were unable to map all the neighbourhoods in the first phase. The next phase would be improve this and map up to 97 neighbourhoods.
2. Another observation is that we have much more crime data over less categories than the venue data, 88 0000 with 5 categories vs 4 800 with 325 categories. This can make our clustering shifted towards crime data. The next follow up could be to apply different weighting or reduce crime data.

-
3. These analysis didn't investigate if there are some neighbourhoods with higher or lower crime rate. For a proper pricing policy it will be important to take this into account with the next research phase.
 4. The crime data is primarily dominated by Assault with over 50% of the crime data and Break and Enter on the 2nd place with about 20%. The next follow up would be to look if that skews our clustering and first investigate if that is dependent on neighbourhood and 2nd see if weights should be adjusted
 5. Try different number of clusters and find optimal