Capstone Project

The Battle of the Neighborhoods

Applied Data Science Capstone by ***IBM/Coursera***

AUTHOR: FAVIO PEÑALOZA

Table of contents

Introduction: Business Problem

“RETAIL Inc.” a department store (B2C Store) from the city of Toronto, which seeks to create a new Physical Store. But the question for the directive board is: Where it would be the best place to place the new Physical Store?

And there are several critical points to analyses and take the decision:

- Is essential to be in a neighbourhood with several points of attractions nearby.
- Another critical point is that they should be near to the customers that already purchase on the online store, to offer a Fulfilment Centre near to the customers.
- And the last point of this evaluation is the crime rates over the zone, is important to be in a relatively safe place.

These points are what makes “RETAIL Inc.” ask for help to the data science department.

To solve this puzzle…

Introduction: Business Problem

RETAIL Inc. first needs to answer the following questions

1. What are the top 5 of neighbourhoods with the most points of attractions nearby?
2. What are my most common destination neighbourhoods for my delivery to home orders?
3. What are the most dangerous neighbourhoods in the city?

Once that we have an answer to those questions, we can answer the **main question:**

Where it would be the best place to place our new Physical Store?

Data Section

RETAIL Inc. has the questions next questions:

1. What are the top 5 neighbourhoods with the most points of attractions nearby?
2. What are my most common destination neighbourhoods for my delivery to home orders?
3. What are the most dangerous neighbourhoods in the city?

For each question, it will be necessary the followings data repositories.

Data Section

Question 1:

We will be using the foursquare's API for the creation of a data set for each neighbourhood in the city of Toronto with their corresponding POI (Points Of Interest).

Question 2:

Using a data extraction tool, a CSV file was created from previous orders. These orders will be a data set that will allow mapping a heat map for the city of Toronto in which shows the most common destinations for delivery home orders.

Question 3 :

Using data from "Toronto Police Service Public Safety Data Portal" we can create an analyse a data set from Crime records.

After the completion of the previous questions, we can the previous data set and create a merged data set for answering the **main question**.

Where it would be the best place to place our new Physical Store?

Methodology section

For each question, it would be necessary to perform: Preprocessing stage (Understanding and Preparation) Processing stage (Modeling and Evaluation)

And then for the final question, the following steps will be executed.

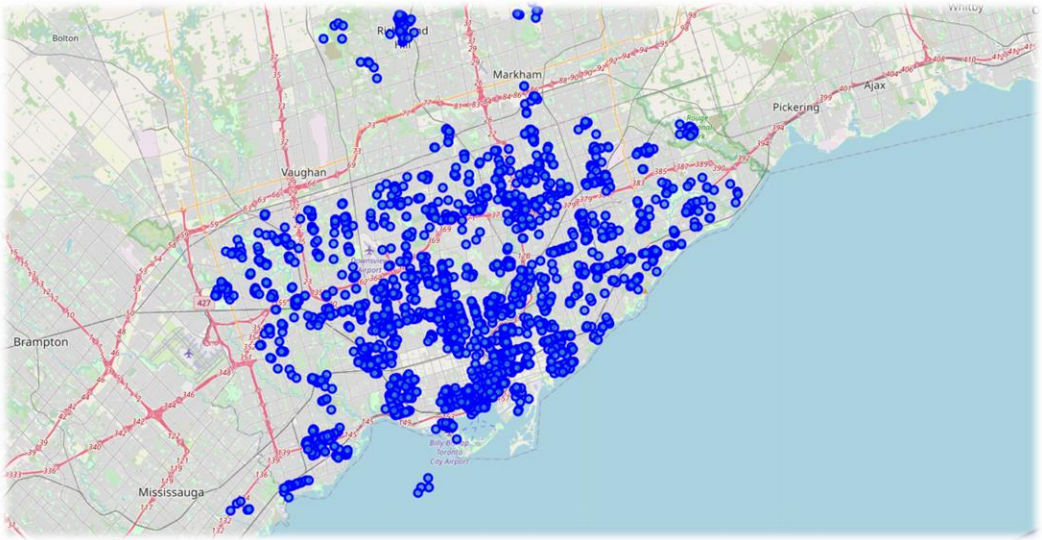
1. **Merge and Discard the unnecessary data:** because of the previous questions had different scopes and objectives, some data may not be needed, and some data needs to be merged.
2. **Data normalization:** Each question before it will have different scales and values, so is necessary first to normalize the data to proceed.
3. **Processing and feeding of the Model:** after all data is normalized, we need to train, feed, and execute our model to evaluate the accuracy of it.

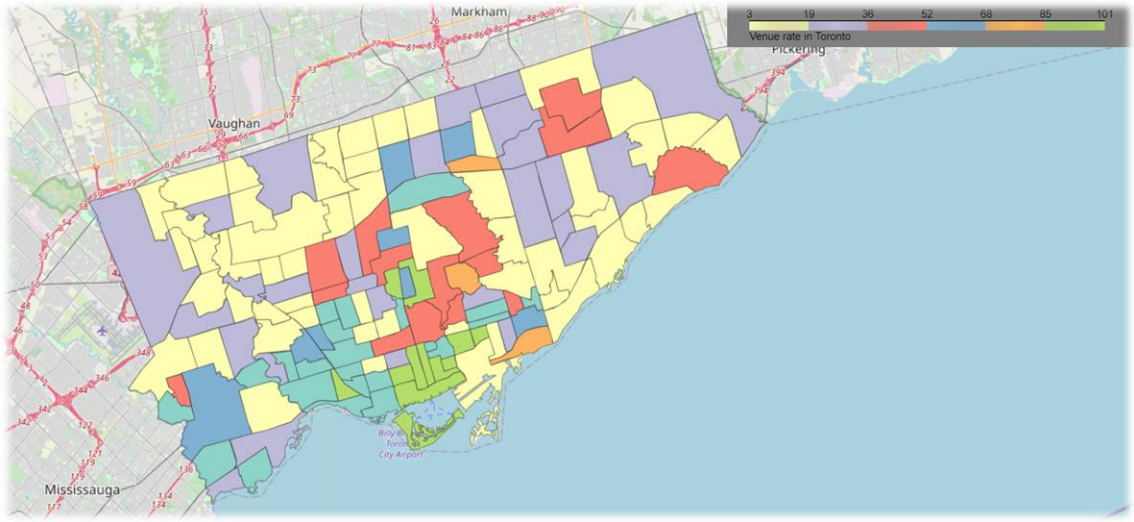
DIVIDE AND RULE (Latin: divide et impera), or **DIVIDE AND CONQUER**, in politics and sociology is gaining and maintaining power by breaking up larger concentrations of power into pieces that individually have less power than the one implementing the strategy. But in what we are trying to do, it means that we can decompose a big problem in small problems, that we can solve one by one.

So let's Start...

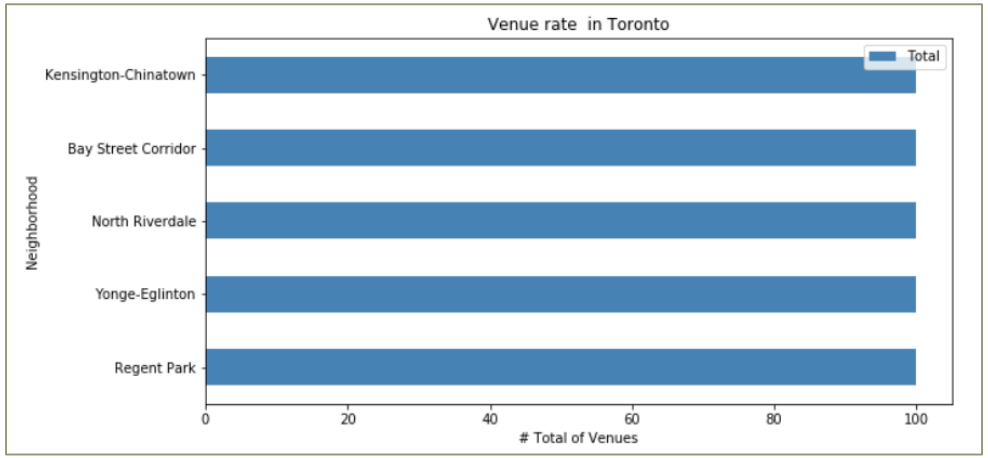
Question 1: What are the top 5 neighborhood's with the most points of attractions nearby?

We will be using the foursquare's API for the creation of a data set for each neighborhood in the city of Toronto with their corresponding POI (Points Of Interest).





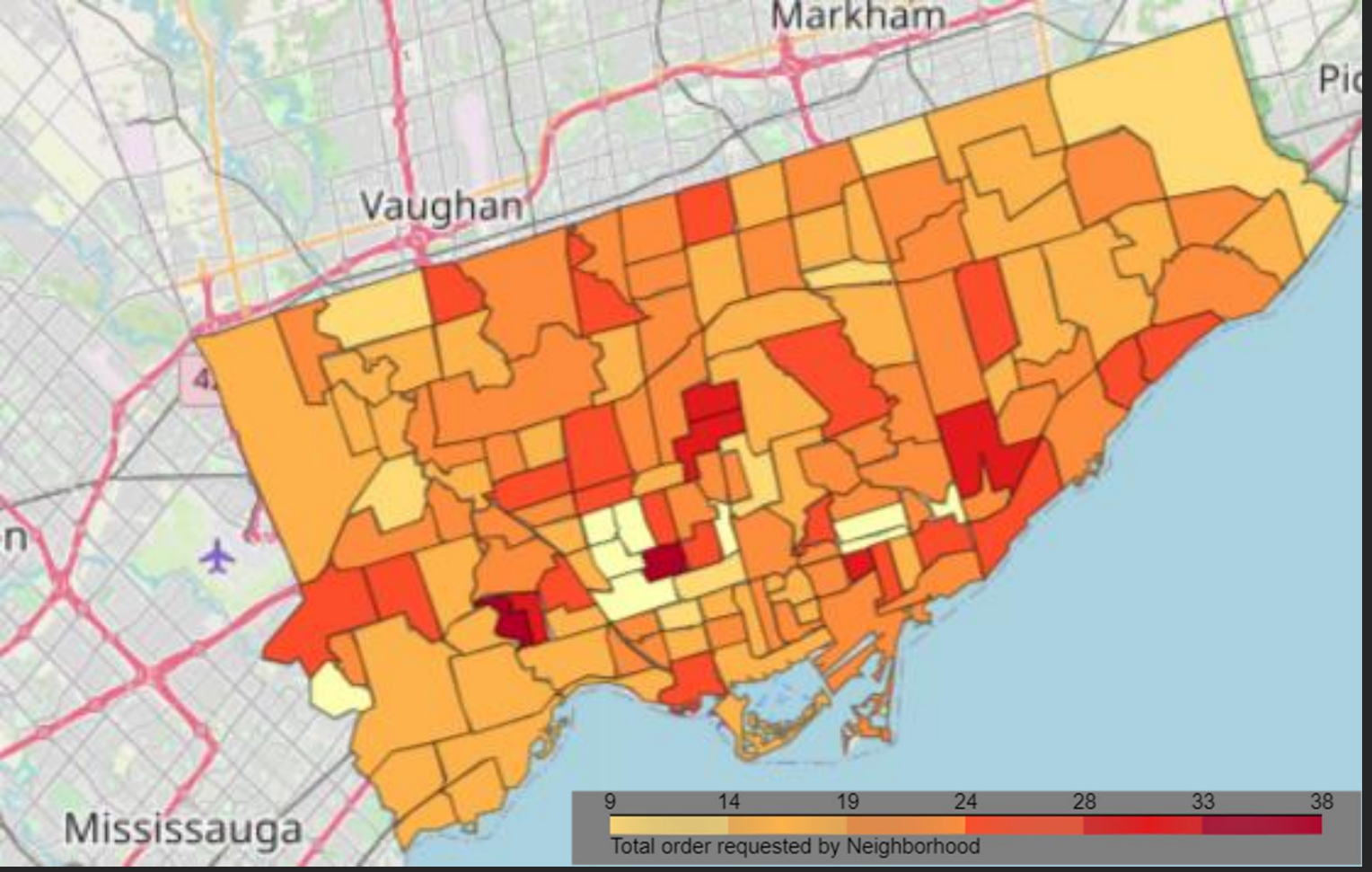
Question 1: What are the top 5 neighborhood's with the most points of attractions nearby?



Using this information we are going to separate the neighborhoods (Top 5) that have more attraction points (Venue)

Moreover, we can get a better picture of each element of the top 5

Neighborhood	Total	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
Regent Park	100	Coffee Shop	Restaurant	Café	Park	Bakery	Thai Restaurant	Gastropub	Pub	Breakfast Spot	Bar
Yonge-Eglinton	100	Coffee Shop	Italian Restaurant	Sushi Restaurant	Café	Bakery	Japanese Restaurant	Restaurant	Pizza Place	Gastropub	Fast Food Restaurant
North Riverdale	100	Greek Restaurant	Coffee Shop	Café	Vietnamese Restaurant	Pub	Ice Cream Shop	Italian Restaurant	Park	Bakery	Fast Food Restaurant
Bay Street Corridor	100	Coffee Shop	Café	Seafood Restaurant	Restaurant	Japanese Restaurant	Hotel	Gastropub	Cosmetics Shop	Tea Room	Plaza
Kensington-Chinatown	100	Coffee Shop	Gym	Café	Park	Mexican Restaurant	French Restaurant	Yoga Studio	Beer Bar	Bar	Italian Restaurant

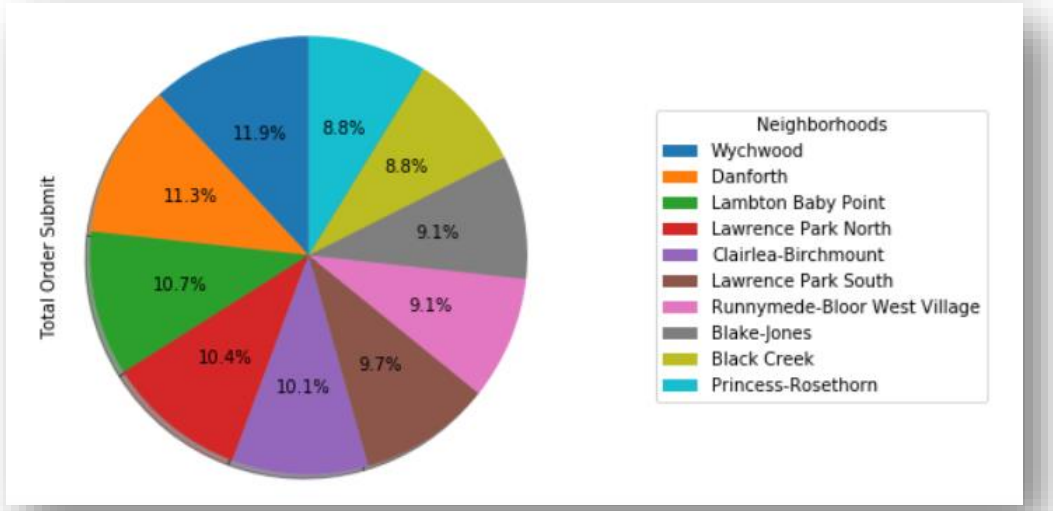


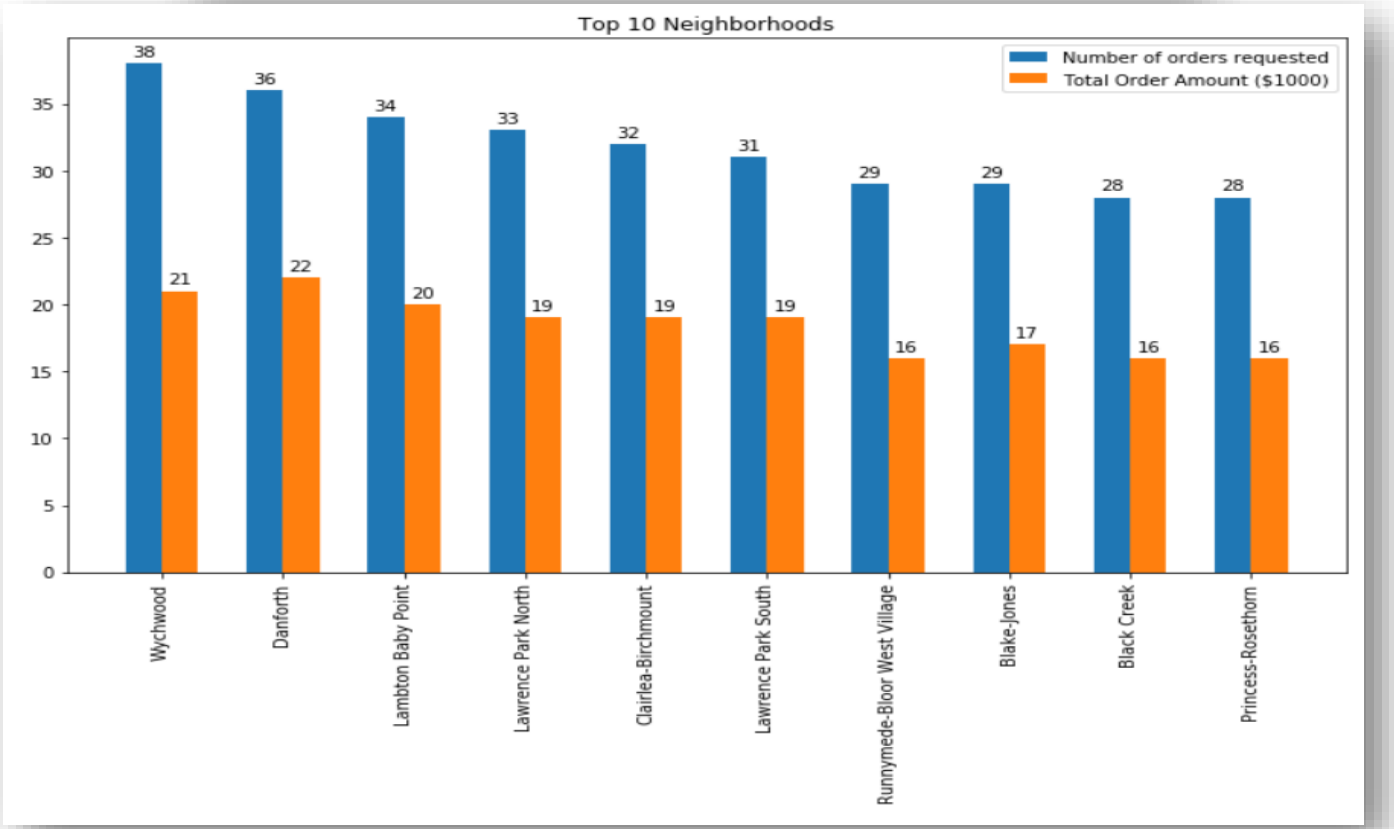
Neighborhood	Total Order Submit	Total Order Amount
Wychwood	38	20146.0
Danforth	36	21438.0
Lambton Baby Point	34	19903.0
Lawrence Park North	33	18888.0
Clairlea-Birchmount	32	18408.0
Lawrence Park South	31	18992.0
Runnymede-Bloor West Village	29	15829.0
Blake-Jones	29	16392.0
Black Creek	28	15827.0
Princess-Rosethorn	28	15833.0

Using a data extraction tool, a CSV file was created from previous orders. These orders will be a data set that will allow mapping a heat map for the city of Toronto in which shows the most common destinations for delivery home orders.

Question 2: What are my most common destination neighborhood's for my delivery to home orders?

Question 2: What are my most common destination neighborhood's for my delivery to home orders?





Question 3 : What are the most dangerous neighborhood's in the city?

Using data from "Toronto Police Service Public Safety Data Portal" we can create and analyze a dataset from Crime records DB.

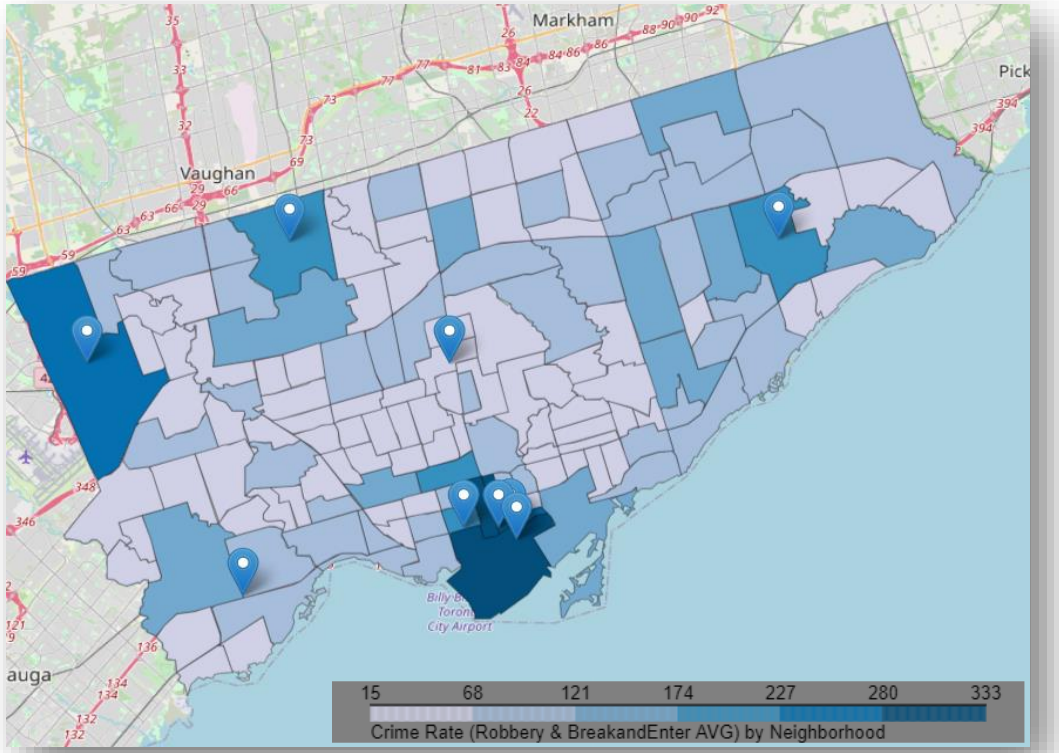
Neighbourhood	Population	Crime Rate	LONGITUDE	LATITUDE
Waterfront Communities-The Island	65913	329.5	-79.368286	43.643870
Church-Yonge Corridor	31340	324.2	-79.371410	43.649425
Bay Street Corridor	25797	280.0	-79.377839	43.649201
Moss Park	20506	274.0	-79.371410	43.649425
West Humber-Clairville	33312	229.6	-79.596356	43.716180
Kensington-Chinatown	17945	214.8	-79.396364	43.648785
York University Heights	27593	189.0	-79.488883	43.765736
Woburn	53485	188.7	-79.228586	43.766740
Annex	30526	188.3	-79.404001	43.715854
Islington-City Centre West	43965	158.1	-79.513671	43.621299

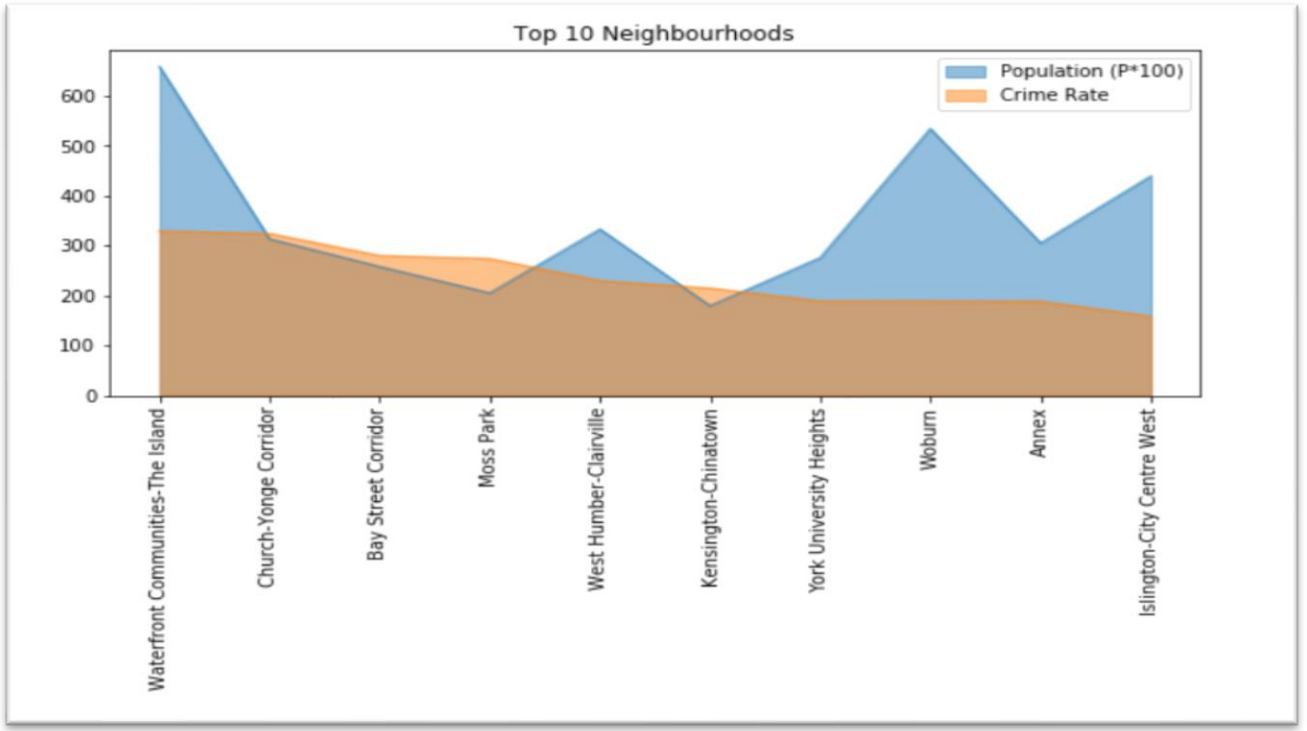
Because there is a lot of crime types and is out of this analysis some of those.

We sum only average values for **robbery** and **break & enter**

This is in order to satisfy our stakeholders definitions

Question 3 : What are the most dangerous neighborhood's in the city?





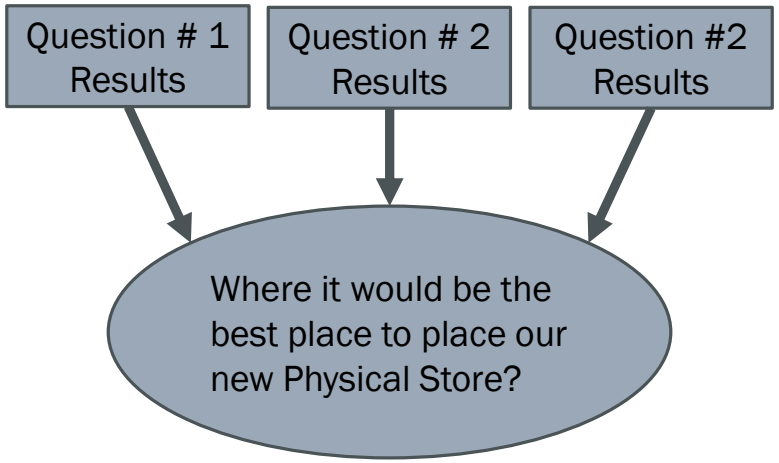
Neighbourhood	Annex	Bay Street Corridor	Church-Yonge Corridor	Islington-City Centre West	Kensington-Chinatown	Moss Park	Waterfront Communities-The Island	West Humber-Clairville	Woburn	York University Heights
Counter	188.3	280.0	324.2	158.1	214.8	274.0	329.5	229.6	188.7	189.0
Population	30526.0	25797.0	31340.0	43965.0	17945.0	20506.0	65913.0	33312.0	53485.0	27593.0

In order to show a clean comparison between Population and Crime Rate, numbers of the population are in relation to 100.

Result Section

The One Million Question:

Where it would be the best place to place our new Physical Store?



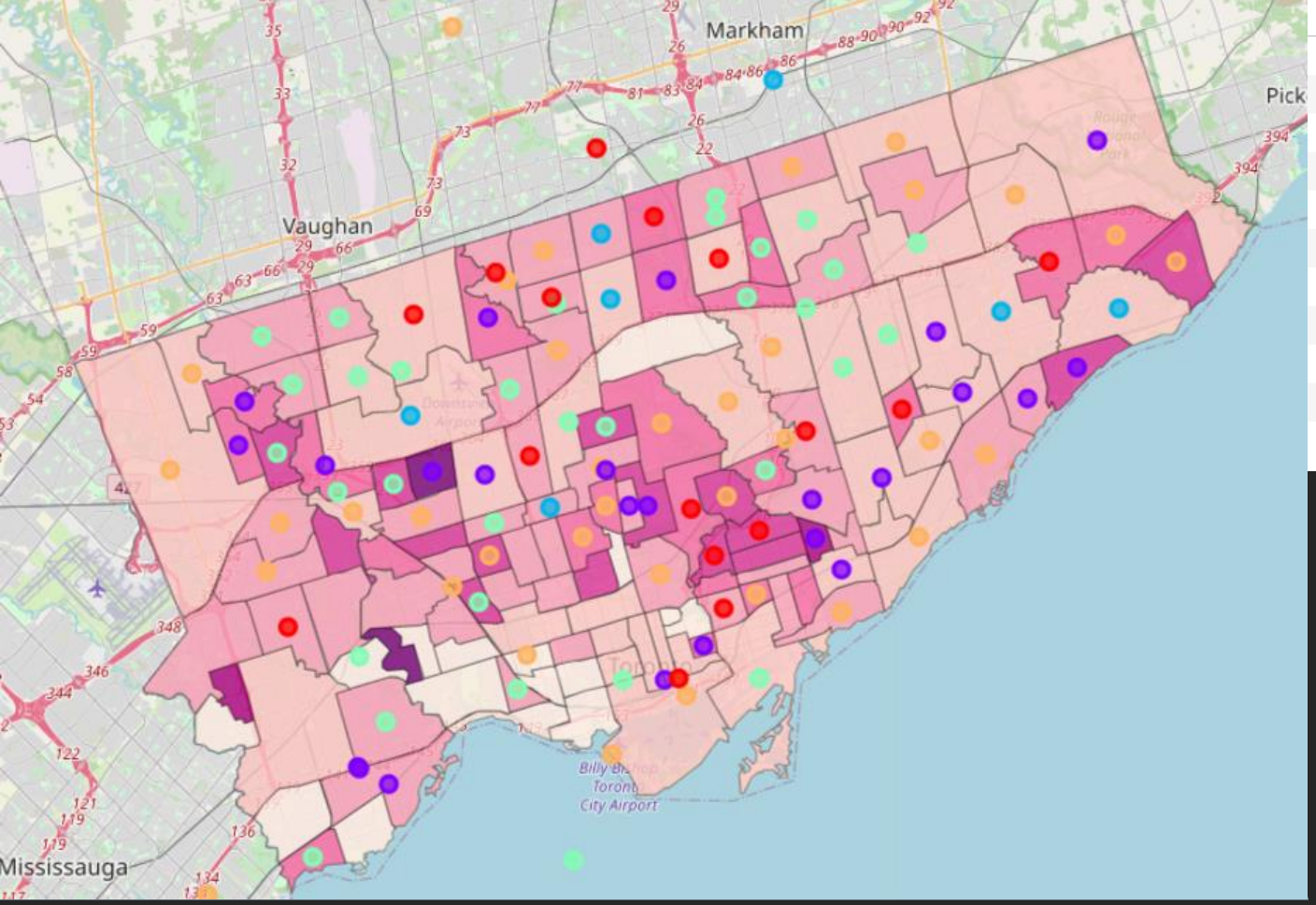
	Neighborhood	Venue	Population	Crime Rate	Longitude	Latitude	Total Order Amount	Total Order Submit	Scoring
0	Maple Leaf	15	10111	17.9	-79.480758	43.715574	11328.0	18	37
1	Lambton Baby Point	29	7985	17.8	-79.512561	43.656624	19903.0	34	35
2	Woodbine-Lumsden	42	7865	21.2	-79.311164	43.694107	10490.0	19	28
3	Etobicoke West Mall	50	11848	22.7	-79.513199	43.620495	12814.0	23	27
4	Broadview North	36	11499	27.0	-79.355630	43.688825	16100.0	26	25
...	...	...	...	...	...	...	...	...	...
119	Moss Park	100	20506	274.0	-79.371410	43.649425	13997.0	22	3
120	York University Heights	23	27593	189.0	-79.488883	43.765736	11681.0	20	3
121	Waterfront Communities-The Island	100	65913	329.5	-79.368286	43.643870	7753.0	17	2
122	Bay Street Corridor	100	25797	280.0	-79.377839	43.649201	9153.0	16	2
123	Church-Yonge Corridor	100	31340	324.2	-79.371410	43.649425	13603.0	21	2

So now after processing the data for each question, we can merge in a dataset with the specific characteristics that we need for our analysis.

This 124 records will become the fuel for your engine...

Note the Scoring is calculated with the next formula :

(Venue+(Total Order Amount/Total Order Submit)+(Population/1000))/Crime Rate



Neighborhood	Venue	Population	Crime Rate	Longitude	Latitude	Total Order Amount	Total Order Submit	Scoring
Maple Leaf	15	10111	17.9	-79.480758	43.715574	11328.0	18	37
Lambton Baby Point	29	7985	17.8	-79.512561	43.656624	19903.0	34	35
Woodbine-Lumsden	42	7865	21.2	-79.311164	43.694107	10490.0	19	28
Etobicoke West Mall	50	11848	22.7	-79.513199	43.620495	12814.0	23	27
Broadview North	36	11499	27.0	-79.355630	43.688825	16100.0	26	25
...	...	...	...	...	...	...	...	...
Moss Park	100	20506	274.0	-79.371410	43.649425	13997.0	22	3
York University Heights	23	27593	189.0	-79.488883	43.765736	11681.0	20	3
Waterfront Communities-The Island	100	65913	329.5	-79.368286	43.643870	7753.0	17	2
Bay Street Corridor	100	25797	280.0	-79.377839	43.649201	9153.0	16	2
Church-Yonge Corridor	100	31340	324.2	-79.371410	43.649425	13603.0	21	2

So, after applying a "Classification" algorithm we were able to have 5 different Clusters.
Each clusters represents neighbours similarly according to our parameters:

- Venue
- Population
- Crime Rate
- Longitude
- Latitude
- Total Order Amount
- Total Order Submit
- Scoring

Where it would be the best place to place our new Physical Store?

Result Section

Classification

Cluster number: 0

	Neighborhood	Venue	Population	Crime Rate	Longitude	Latitude	Total Order Amount	Total Order Submit	Scoring
16	Broadview North	36	11499	27.0	-79.355630	43.688825	16100.0	26	25
82	Old East York	25	9233	27.0	-79.335488	43.696781	10231.0	19	21
19	Casa Loma	16	10968	31.6	-79.408007	43.818521	16617.0	28	20
7	Bayview Woods-Steeles	15	13154	32.9	-79.382118	43.796802	10118.0	17	19
70	Morningside	16	17455	38.4	-79.207041	43.782399	11201.0	17	18
64	Leaside-Bennington	41	16828	40.8	-79.366072	43.703797	9876.0	15	18
52	Ionview	25	13641	40.1	-79.272470	43.735364	9640.0	15	17
88	Princess-Rosethorn	4	11051	43.7	-79.544559	43.666051	15833.0	28	13
115	Willowdale West	5	16936	48.8	-79.427558	43.771210	11159.0	19	12
106	Victoria Village	14	17510	52.2	-79.314874	43.728489	10799.0	20	11
78	North Riverdale	100	11916	66.7	-79.351260	43.671995	12515.0	20	11
36	Englemount-Lawrence	32	22372	69.2	-79.437409	43.720345	12621.0	19	10
110	Westminster-Branson	16	26274	59.7	-79.452418	43.778813	12691.0	23	10
28	Don Valley Village	64	27051	82.2	-79.353644	43.783294	7016.0	14	7
80	Oakridge	4	13845	97.8	-79.279708	43.974081	12532.0	21	6
122	York University Heights	23	27593	189.0	-79.488883	43.765736	11681.0	20	3
21	Church-Yonge Corridor	100	31340	324.2	-79.371410	43.649425	13603.0	21	2

Result Section

Classification

Cluster number: 1

	Neighborhood	Venue	Population	Crime Rate	Longitude	Latitude	Total Order Amount	Total Order Submit	Scoring
67	Maple Leaf	15	10111	17.9	-79.480758	43.715574	11328.0	18	37
119	Woodbine-Lumsden	42	7865	21.2	-79.311164	43.694107	10490.0	19	28
38	Etobicoke West Mall	50	11848	22.7	-79.513199	43.620495	12814.0	23	27
44	Guildwood	13	9917	22.7	-79.195055	43.748829	14384.0	26	25
74	Mount Pleasant East	100	16775	36.9	-79.384924	43.704852	8256.0	12	22
103	Thistletown-Beaumond Heights	19	10360	34.0	-79.563491	43.737988	10398.0	18	18
90	Rexdale-Kipling	14	10529	36.0	-79.566228	43.723725	9226.0	15	18
85	Pelmo Park-Humberlea	16	10722	41.1	-79.528282	43.717515	11412.0	19	15
89	Regent Park	100	10803	50.0	-79.360509	43.659992	9826.0	15	15
4	Bathurst Manor	11	15873	40.6	-79.456055	43.764813	15966.0	27	15
6	Bayview Village	27	21396	49.0	-79.377117	43.776361	13663.0	20	15
51	Humewood-Cedarvale	7	14365	40.3	-79.427683	43.913698	12827.0	25	13
101	Taylor-Massey	4	15683	67.0	-79.295901	43.949982	11052.0	18	9
96	Scarborough Village	19	16724	66.3	-79.216813	43.738652	16012.0	27	9
75	Mount Pleasant West	58	29658	86.2	-79.393360	43.704435	14946.0	22	9
79	O'Connor-Parkview	18	18675	61.0	-79.312228	43.706800	10657.0	22	9
69	Mimico	29	33964	77.1	-79.500137	43.615924	10216.0	18	8
32	East End-Danforth	61	21381	108.7	-79.299359	43.684174	13101.0	25	6
34	Eglinton East	16	22776	97.3	-79.245598	43.740922	11052.0	20	6
123	Yorkdale-Glen Park	46	14804	94.8	-79.457108	43.714672	15025.0	28	6
94	Rouge	24	46496	104.6	-79.186343	43.821201	7888.0	13	6
10	Bendale	16	29960	123.4	-79.257400	43.760366	8747.0	16	5
22	Clairlea-Birchmount	12	26984	135.2	-79.281382	43.713592	18408.0	32	5
53	Islington-City Centre West	53	43965	158.1	-79.513671	43.621299	12279.0	18	5
2	Annex	47	30526	188.3	-79.404001	43.715854	5988.0	9	4
5	Bay Street Corridor	100	25797	280.0	-79.377839	43.649201	9153.0	16	2

Result Section

Classification

Cluster number: 2

	Neighborhood	Venue	Population	Crime Rate	Longitude	Latitude	Total Order Amount	Total Order Submit	Scoring
26	Danforth	10	9666	58.0	-79.329819	43.840255	21438.0	36	11
40	Forest Hill North	28	12806	46.0	-79.428143	43.704218	7442.0	16	11
76	Newtonbrook East	19	16097	58.8	-79.405937	43.791536	12438.0	24	9
114	Willowdale East	57	50434	120.9	-79.401484	43.770602	9031.0	15	6
108	West Hill	38	27392	147.7	-79.176676	43.767490	14293.0	23	5
117	Woburn	28	53485	188.7	-79.228586	43.766740	9320.0	14	4
31	Downsview-Roding-CFB	8	35052	143.5	-79.490497	43.733292	11223.0	19	4

Result Section

Classification

Cluster number: 3

	Neighborhood	Venue	Population	Crime Rate	Longitude	Latitude	Total Order Amount	Total Order Submit	Scoring
60	Lambton Baby Point	29	7985	17.8	-79.512561	43.656624	19903.0	34	35
35	Elms-Old Rexdale	11	9456	24.0	-79.548983	43.721519	10297.0	18	25
95	Rustic	18	9941	23.9	-79.498091	43.711609	10587.0	20	23
48	Humber Heights-Westmount	11	10948	29.2	-79.522416	43.922328	13963.0	23	22
62	Lawrence Park North	56	14607	30.3	-79.403978	43.730060	18888.0	33	21
112	Weston-Pellam Park	53	11098	33.7	-79.460244	43.673962	8741.0	14	20
45	Henry Farm	83	15723	36.0	-79.341241	43.771144	7366.0	13	18
39	Flemingdon Park	36	21933	41.0	-79.332646	43.715930	15016.0	23	17
87	Pleasant View	19	15818	31.6	-79.334948	43.786982	6419.0	15	15
86	Playter Estates-Danforth	17	7804	39.9	-79.354887	43.797005	10589.0	19	15
118	Woodbine Corridor	30	12541	47.0	-79.315407	43.767729	10026.0	16	14
65	Long Branch	30	10084	44.5	-79.533345	43.592362	9003.0	16	14
72	Mount Dennis	5	13593	41.3	-79.499989	43.881441	12269.0	22	14
120	Wychwood	17	14349	47.8	-79.425515	43.769193	20146.0	38	12
33	Edenbridge-Humber Valley	22	15535	50.8	-79.522458	43.708859	9811.0	18	11
47	Hillcrest Village	22	16934	53.3	-79.354804	43.802988	13083.0	23	11
83	Palmerston-Little Italy	4	13826	67.5	-79.418409	43.591571	6556.0	9	11
23	Clanton Park	14	16472	60.6	-79.446303	43.741978	11262.0	17	11
12	Black Creek	16	21737	68.0	-79.521979	43.764890	15827.0	28	9
91	Rockcliffe-Smythe	12	22246	72.7	-79.494420	43.747905	12862.0	21	9
99	Stonegate-Queensway	7	25051	73.5	-79.501128	43.635518	10684.0	17	9
92	Roncesvalles	100	14974	74.5	-79.442992	43.646123	11681.0	22	9
50	Humbermede	21	15545	64.9	-79.542367	43.743430	7722.0	14	9
49	Humber Summit	8	12416	74.4	-79.556175	43.758920	7955.0	12	9
14	Briar Hill-Belgravia	38	14257	58.2	-79.452851	43.699024	11760.0	24	9
100	Tam O'Shanter-Sullivan	30	27446	90.4	-79.302919	43.780130	14113.0	20	8
42	Glenfield-Jane Heights	9	30491	89.9	-79.513465	43.745636	7911.0	16	6
8	Bedford Park-Nortown	49	23236	105.5	-79.420227	43.731486	13714.0	23	6
1	Agincourt South-Malvern West	40	23757	107.1	-79.265612	43.788658	8637.0	14	6
29	Dorset Park	33	25003	113.6	-79.278908	43.759274	12517.0	25	5
59	L'Amoreaux	29	43993	116.5	-79.314084	43.795716	10843.0	22	5
113	Wexford/Maryvale	32	27917	127.5	-79.298637	43.748572	13173.0	22	5
97	South Riverdale	14	27876	157.8	-79.335651	43.649292	13809.0	23	4
57	Kensington-Chinatown	100	17945	214.8	-79.396364	43.648785	14268.0	21	4

Result Section

Classification

Cluster number: 4

	Neighborhood	Venue	Population	Crime Rate	Longitude	Latitude	Total Order Amount	Total Order Submit	Scoring
20	Centennial Scarborough	5	13362	26.7	-79.150843	43.782376	9167.0	15	24
104	Thorncliffe Park	72	21108	36.5	-79.349984	43.707749	16288.0	22	23
9	Beechborough-Greenbrook	4	6577	28.6	-79.479473	43.932165	12268.0	21	21
27	Danforth East York	8	17180	33.0	-79.331403	43.894681	10702.0	17	20
18	Caledonia-Fairbank	27	9955	29.3	-79.455212	43.688569	12877.0	24	20
13	Blake-Jones	100	7727	35.5	-79.337394	43.676173	16392.0	29	19
46	Highland Creek	6	12494	45.1	-79.177472	43.790775	12295.0	19	15
121	Yonge-Eglinton	100	11817	47.5	-79.403590	43.704689	11541.0	20	15
15	Bridle Path-Sunnybrook-York Mills	10	9266	52.1	-79.378904	43.731013	12282.0	17	14
41	Forest Hill South	35	10732	45.9	-79.414318	43.694526	12552.0	21	14
81	Oakwood Village	54	21210	57.3	-79.439785	43.885656	13198.0	20	13
58	Kingsview Village-The Westway	17	22000	54.5	-79.547863	43.698993	7779.0	12	13
55	Keelesdale-Eglinton West	8	11058	42.7	-79.471437	43.857275	9368.0	17	13
17	Brookhaven-Amesbury	20	17757	55.6	-79.485589	43.701326	15929.0	25	12
63	Lawrence Park South	41	15179	55.9	-79.406039	43.717212	18992.0	31	12
54	Junction Area	58	14366	56.3	-79.471440	43.678898	15026.0	25	12
25	Corso Italia-Davenport	17	14133	52.5	-79.447469	43.776609	15140.0	26	12
37	Eringate-Centennial-West Deane	7	18588	43.3	-79.580445	43.580168	12903.0	28	11
102	The Beaches	77	21567	69.6	-79.299601	43.671050	14676.0	23	11
61	Lansing-Westgate	7	16164	53.5	-79.424748	43.754271	10596.0	22	9
98	Steeles	22	24623	66.4	-79.321207	43.812959	6776.0	12	9
77	Newtonbrook West	4	23831	79.7	-79.431422	43.785830	14520.0	20	9
24	Cliffcrest	12	15935	72.0	-79.235530	43.721121	14301.0	22	9
116	Willowridge-Martingrove-Richview	25	22156	70.3	-79.554221	43.683645	11149.0	21	8
0	Agincourt North	46	29113	83.7	-79.266712	43.805441	11029.0	18	8
43	Greenwood-Coxwell	17	14417	59.0	-79.324318	43.726123	12684.0	27	8
93	Rosedale-Moore Park	44	20923	87.2	-79.379669	43.682820	13071.0	22	8
84	Parkwoods-Donalda	17	34805	94.7	-79.330180	43.755033	9878.0	17	7
11	Birchcliffe-Cliffside	15	22291	93.8	-79.265093	43.694682	15548.0	27	7
105	University	20	7607	92.8	-79.401180	43.625062	11136.0	18	7
56	Kennedy Park	22	17123	91.8	-79.260382	43.725556	13057.0	21	7
73	Mount Olive-Silverstone-Jamestown	18	32954	109.2	-79.587259	43.746868	15096.0	23	6
66	Malvern	18	43794	103.0	-79.222517	43.803658	10623.0	20	6
3	Banbury-Don Mills	45	27695	88.2	-79.349718	43.737657	11840.0	25	6
111	Weston	22	17992	91.7	-79.515723	43.702716	8489.0	18	6
30	Dovercourt-Wallace Emerson-Junction	100	36625	157.5	-79.438541	43.656769	17233.0	28	5
68	Milliken	14	26572	141.0	-79.275009	43.820691	9144.0	16	4
109	West Humber-Clairville	27	33312	229.6	-79.596356	43.716180	9823.0	17	3
71	Moss Park	100	20506	274.0	-79.371410	43.649425	13997.0	22	3
107	Waterfront Communities-The Island	100	65913	329.5	-79.368286	43.643870	7753.0	17	2

Conclusion Section

After analysing the cluster information we can take as good prospect clusters 3 and 1.
From those clusters, the following neighbourhoods:

Neighborhood	Venue	Population	Crime Rate	Longitude	Latitude	Total Order Amount	Total Order Submit	Scoring
Maple Leaf	15	10111	17.9	-79.480758	43.715574	11328.0	18	37

Neighborhood	Venue	Population	Crime Rate	Longitude	Latitude	Total Order Amount	Total Order Submit	Scoring
Lambton Baby Point	29	7985	17.8	-79.512561	43.656624	19903.0	34	35

These two neighborhoods are the recommended options for the new store location.

“It’s one small step for man, one giant leap for mankind.”

- NEIL ARMSTRONG