Introduction

Background

Canada is known as the most immigrant welcoming and multicultural western country. In 2019, Canada admitted roughly 340,000 new permanent residents, of which 45% chose to settle around Toronto which is the biggest metropolis of Canada and continuously appraised as one of the topmost livable cities in the world.

Problem

In 2020 November, Canada announced a 3-year open work permit and immigration pathway for Hong Kong residents. The policy has drawn strong attention of Hong Kong youths who are desperate for an immediate move. This report aims to study the living environment of Toronto neighbourhoods in 4 areas – house price, community safety, ethnic diversity and neighbourhood venues. Target audience of this report is Hong Kong residents who decided to make a hasty trip to Toronto but have not determined a suitable place to settle down. Particularly I will focus the study on the most populous urban centre - Toronto City. Toronto City is 630 square kilometres in size and of estimated 2.9M population. I will dissect the 140 neighbourhoods in Toronto City.

Data Source

Below are the data sources for this study, with a fraction of the datasets being shown.

1. Wikipedia provides a table of postal codes and neighbourhoods.

Postal Code	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue, Humber Valley Village
M1B	Scarborough	Malvern, Rouge

2. Geocode Python provides latitude and longitude coordinates of the postal codes.

	Postal Code	Latitude	Longitude
1	M1B	43.8066863	-79.1943534
2	M1C	43.7845351	-79.1604971
3	M1E	43.7635726	-79.1887115
4	M1G	43.7709921	-79.2169174
5	M1H	43.773136	-79.2394761
6	M1J	43.7447342	-79.2394761
7	M1K	43.7279292	-79.2620294
8	M1L	43.7111117	-79.2845772
9	M1M	43.716316	-79.2394761
10	M1N	43.692657	-79.2648481

3. Toronto Open Data Portal provides Toronto City's Neighbourhood geojson file for creating neighbourhoods map.

```
{
"type": "FeatureCollection",
"crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:OGC:1.3:CRS84" } },
"features": [
{ "type": "Feature", "properties": { "AREA_SHORT_CODE": 96 }, "geometry": { "type": "Polygon", "coordinates": [ [ [ -79.414693178177814,
43.673910416425898 ], [ -79.414849301228315, 43.674338824792699 ], [ -79.415527912609406, 43.67606998537741 ], [ -79.415786758113697, 43.676730252138597
], [ -79.416438564554596, 43.678332902051103 ], [ -79.416578254386195, 43.678778551234402 ], [ -79.416758394468204, 43.6792484522149 ], [
-79.416982035372911, 43.679824744004698 ], [ -79.417090247125103, 43.680104700401401 ], [ -79.4174797570886, 43.681051830378998 ], [ -79.417666629515793,
43.681569909621203 ], [ -79.417697602085696, 43.68163169892501 ], [ -79.417811656038793, 43.681925277740298 ], [ -79.418121371159799, 43.682710522285703
], [ -79.4182734938897, 43.683096166233497 ], [ -79.418301440907001, 43.683167503583 ], [ -79.418431341169196, 43.683496684199604 ], [
-79.418487075737104, 43.683629484254801 ], [ -79.418541079590497, 43.683758205453799 ], [ -79.418409292256101, 43.683786393574003 ], [
-79.418295528334397, 43.683493715887 ], [ -79.415426533315099, 43.684065660546999 ], [ -79.413520043283199, 43.684445691506198 ], [ -79.413408786159593,
43.684469595764298 ], [ -79.413468695917601, 43.684624553621397 ], [ -79.4124643275333, 43.684828847674503 ], [ -79.411896000969406, 43.684941854512502 ],
[ -79.412602065633905, 43.686685769244299 ], [ -79.412821822381204, 43.687229039671003 ], [ -79.412273186089493, 43.687340632142998 ], [
-79.412309585178903, 43.687430664093597 ], [ -79.412444931077502, 43.6877690817767 ], [ -79.412574976461102, 43.688102670528302 ], [ -79.412602244203299,
43.688203082891008 ], [ -79.412626235544394, 43.688338876854601 ], [ -79.412631283351004, 43.6884334123732 ], [ -79.412643206881896, 43.688516872543701 ],
[ -79.412669085002705, 43.688669014294199 ], [ -79.412179802344795, 43.688772568451498 ], [ -79.412131745035893, 43.688782747996001 ], [
-79.409491162023201, 43.6893416576956 ], [ -79.409515914193193, 43.689437316094597 ], [ -79.4079543147094, 43.689784797617897 ], [ -79.406402601984695,
43.690080679442403 ], [ -79.4048987011668, 43.690395860753497 ], [ -79.403017636954601, 43.690782128177098 ], [ -79.403062238680505, 43.690739751898903 ],
[ -79.403084589916006, 43.690687266286403 ], [ -79.403084702545399, 43.690632727523202 ], [ -79.402338539631316, 43.688721197786599 ], [
-79.401495489974394, 43.686640696640801 ], [ -79.401126389151997, 43.685716070065702 ], [ -79.400841057476001, 43.685027250051903 ], [
-79.400530364766198, 43.684309801023602 ], [ -79.400209294998206, 43.683526704463098 ], [ -79.400168894772705, 43.683415980098403 ], [
-79.399803530204395, 43.682543487581498 ], [ -79.399215422659694, 43.681127631446898 ], [ -79.399065444652393, 43.680766546492201 ], [
-79.398939474684695, 43.680460426561197 ], [ -79.398733162036507, 43.679962228305001 ], [ -79.398562965777202, 43.679554272862703 ], [
-79.398351516059805, 43.679039164810199 ], [ -79.3981409020124, 43.678530736041502 ], [ -79.398012161217096, 43.6782038734666 ], [ -79.3979440512351,
43.678026297710304 ], [ -79.397931367252596, 43.677994961802398 ], [ -79.397997906572996, 43.677964884787102 ], [ -79.398107663056621, 43.677915484430301
], [ -79.3982178162304, 43.677866084320804 ], [ -79.398328366093196, 43.677816684457504 ], [ -79.398439659889505, 43.677767294149199 ], [
-79.3985509707706, 43.677722377401999 ], [ -79.398663057978112, 43.677672969576797 ], [ -79.398775149838102, 43.677628062304201 ], [ -79.398887998060914,
43.677583155590305 ], [ -79.399001242983402, 43.677538249115202 ], [ -79.399114879728003, 43.677488851215699 ], [ -79.3992025285702, 43.677457423924402 ],
[ -79.399290179973605, 43.677416995250098 ], [ -79.399378605070098, 43.677381067848302 ], [ -79.399467426912693, 43.677345149728104 ], [
-79.399556613219403, 43.677313723516598 ], [ -79.399646203799904, 43.6772777969312 ], [ -79.399735781856506, 43.677241879265701 ], [ -79.399826133678403,
43.677210444864301 ], [ -79.399916894643695, 43.677179028752199 ], [ -79.400007618338194, 43.677147594532599 ], [ -79.400099135653207, 43.677116169935701
], [ -79.400190660317904, 43.677080253613802 ], [ -79.400282561897896, 43.677048829205198 ], [ -79.4003752372778, 43.6770218970549 ], [
-79.400467912560003, 43.676994973830503 ], [ -79.400560979653704, 43.676963550212399 ], [ -79.400654014406399, 43.676936636150799 ], [
-79.400747862653418, 43.676905204057803 ], [ -79.400841695965198, 43.676882782194703 ], [ -79.400935916075099, 43.676851359270799 ], [
-79.401030518155693, 43.676828939719187 ], [ -79.401139491412707, 43.676793026939499 ], [ -79.401248444757115, 43.676761616498403 ], [ -79.401357794850597,
43.676730206295701 ], [ -79.401467516888601, 43.676698796308798 ], [ -79.401577635644998, 43.676667404560597 ], [ -79.401687746997595, 43.676640477354098
], [ -79.401798242667795, 43.676613577375498 ], [ -79.401909142446996, 43.676582167975802 ], [ -79.402020022385798, 43.676555259110899 ], [
-79.4021313064015, 43.676523849824697 ], [ -79.402242972549601, 43.676501442076102 ], [ -79.402354633510896, 43.676474533554199 ], [ -79.402466678831402,
43.6764476252511 ], [ -79.402578724051907, 43.676420716837903 ], [ -79.402691555634902, 43.676398300642603 ], [ -79.402803972719994, 43.67637139232491 ],
[ -79.402883149386, 43.676354186087302 ], [ -79.402916791725104, 43.676348975896403 ], [ -79.403029982688196, 43.676326568673197 ], [ -79.403143577680495,
43.676299652019097 ], [ -79.404967016838796, 43.67592080470620 2 ], [ -79.407333814941794, 43.675429007406997 ], [ -79.407847991861999, 43.675322181326607
], [ -79.411360894925096, 43.674595915212201 ], [ -79.412186097033299, 43.674431512860401 ], [ -79.413716181248901, 43.674111501823198 ], [
-79.414674868207698, 43.673914309334698 ], [ -79.414693178177814, 43.673910416425898 ] ] ] } },
```


It also provides Neighbourhood Profiles basing on tabulations of 2016 Census of Population data, which portraits demographic and social characteristics of people in each neighbourhood. Please note that Census of Population is held every 5 years in Canada, I look to a refreshed dataset be coming soon.

	Characteristic	City of Toronto	Agincourt North	icourt South-Malvern West	Alderwood	Annex	Banbury-Don
1	Latin American	77165	170	230	180	405	
2	South Asian	338965	5315	4190	540	1125	
3	Chinese	299465	16460	11110	275	2145	
4	Black	239850	1340	1545	300	935	
5	Filipino	152715	1420	1280	615	350	
6	Arab	36030	275	300	90	300	
7	Southeast Asian	41650	260	155	55	190	
8	West Asian	60320	125	375	20	340	
9	Korean	41640	70	140	65	605	
10	Japanese	13415	60	40	35	255	
11	Visible minority n.i.e.	36975	245	275	140	125	
12	Multiple visible minorities	47670	630	490	180	480	

4. Kaggle provides the latest datasets of Toronto House Price information originated from the Toronto Regional Real Estate Board (TRREB). Data is broken down into 4 house types – Single Family Detached, Single Family Attached, Townhouse, Apartment, of individual neighbourhood in the months of 2015-2019. Apart from house price benchmark of each neighbourhood, TRREB also uses various housing attributes to derive House Price Index HPI which is a measure of weighted house price change.

	Location	CompIndex	CompBenchmark	ompYoYChange	DetachIndex	iBenchmark	YoYChange	AttachIndex	hBenchmark	hYoYChange	HouseIndex	eBenchmark	YoYChange	ApartIndex	Benchmark	oYChange	Date
62	Adjala-Tosorontio	262.0	788400.0	17.28	262.3	791200.0	17.41										2020-08-01
63	Adjala-Tosorontio	264.3	795400.0	19.16	264.7	798400.0	19.34										2020-09-01
64	Adjala-Tosorontio	273.2	822100.0	21.75	273.6	825300.0	21.93										2020-10-01
65	Adjala-Tosorontio	279.9	842300.0	19.46	280.3	845500.0	19.63										2020-11-01
66	Adjala-Tosorontio	292.1	879000.0	27.06	292.6	882600.0	27.22										2020-12-01
67	Ajax	179.8	444300.0	16.0	178.8	478900.0	17.86	187.1	402400.0	15.64	170.5	311600.0	18.16	148.7	252500.0	5.61	2015-07-01
68	Ajax	180.4	445800.0	16.31	179.2	480000.0	15.76	187.6	403500.0	16.81	170.4	311400.0	15.76	151.5	257200.0	8.99	2015-08-01
69	Ajax	182.8	451700.0	15.55	181.4	485800.0	15.1	190.1	408800.0	15.99	172.3	314900.0	14.33	154.3	262000.0	10.53	2015-09-01
70	Ajax	183.8	454200.0	15.74	182.3	488300.0	14.51	191.2	411200.0	15.95	177.1	323700.0	17.13	152.4	258800.0	8.7	2015-10-01
71	Ajax	182.1	450000.0	12.69	180.3	482900.0	11.99	190.3	409300.0	13.14	173.1	316300.0	12.99	151.2	256700.0	5.0	2015-11-01
72	Ajax	182.4	450700.0	13.22	180.8	484200.0	12.86	190.2	409100.0	13.42	173.7	317400.0	12.35	152.9	259600.0	5.74	2015-12-01
73	Ajax	187.3	462800.0	16.05	187.2	501300.0	16.56	196.2	422000.0	16.44	176.9	323400.0	17.15	144.0	244500.0	-0.69	2016-01-01
74	Ajax	189.5	468300.0	14.23	188.1	503700.0	13.04	198.6	427200.0	14.53	177.1	323700.0	22.73	149.6	254000.0	2.05	2016-02-01
75	Ajax	196.1	484600.0	15.76	194.7	521400.0	15.07	205.7	442500.0	16.02	182.2	333100.0	21.71	148.7	252500.0	1.09	2016-03-01
76	Ajax	202.8	501100.0	18.18	201.2	538800.0	17.59	213.0	458200.0	18.99	189.7	346800.0	21.6	151.2	256700.0	1.0	2016-04-01
77	Ajax	210.4	519900.0	20.44	207.2	554900.0	19.63	218.8	470600.0	19.56	195.8	357900.0	18.88	155.6	264200.0	4.99	2016-05-01
78	Ajax	212.3	524600.0	19.81	210.6	564000.0	19.59	222.2	478000.0	20.56	199.8	365200.0	16.84	158.0	268300.0	8.22	2016-06-01
79	Ajax	214.0	528800.0	19.02	213.3	571300.0	19.3	226.3	486700.0	20.95	202.8	370700.0	18.94	161.3	273900.0	8.47	2016-07-01
80	Ajax	216.8	535700.0	20.18	215.0	575800.0	20.04	229.0	492600.0	22.07	202.4	370000.0	18.78	168.0	285300.0	10.96	2016-08-01
81	Ajax	220.8	545600.0	20.79	218.8	585900.0	20.62	232.6	500300.0	22.36	208.1	380400.0	20.78	173.0	293800.0	12.12	2016-09-01
82	Ajax	224.9	555700.0	22.36	222.6	596100.0	22.11	235.8	507200.0	23.33	215.4	393800.0	21.69	177.0	300500.0	16.22	2016-10-01
83	Ajax	226.6	559900.0	24.44	225.0	602600.0	24.79	238.6	513200.0	25.32	217.9	398300.0	25.88	182.9	310600.0	20.97	2016-11-01
84	Ajax	228.0	563400.0	25.0	225.3	603400.0	24.61	238.9	513900.0	25.54	223.4	408400.0	28.61	182.6	310100.0	19.42	2016-12-01

Kaggle also provides 2014 -2019 Crime Rate, data is broken down into individual neighbourhood's figures.

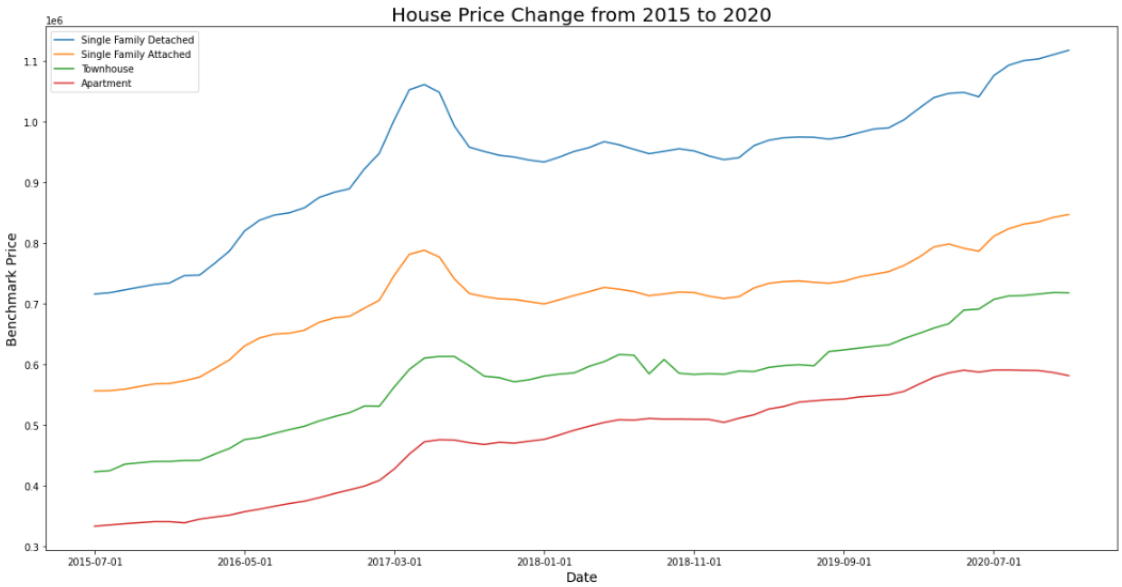
	OBJECTID	Neighbourhood	Hood_ID	Population	Assault_2014	sault_2015	sault_2016	ult_2017	lt_2018	ult_2019	sault_AVG	ult_CHG	Assault_Rate_2019	eft_2014	eft_2015	eft_2016	eft_2017	ft_2018	heft_2019	heft_AVG	ft_CHG	AutoTheft_Rate_2019
1	1	Yonge-St.Clair	097	12528	20	29	39	27	34	37	31	0.09	295.3	2	3	7	2	6	6	4.3	0	47.9
2	2	York University Heights	027	27593	271	296	361	344	357	370	333.2	0.04	1340.9	105	100	105	92	92	144	106.3	0.57	521.9
3	3	Lansing-Westgate	038	16164	44	80	68	85	75	72	70.7	-0.04	445.4	19	22	27	26	16	32	23.7	1	198
4	4	Yorkdale-Glen Park	031	14804	106	136	174	161	175	209	160.2	0.19	1411.8	63	53	41	52	63	61	55.5	-0.03	412.1
5	5	Stonegate-Queensway	016	25051	88	71	76	95	87	82	83.2	-0.06	327.3	34	29	12	32	31	34	28.7	0.1	135.7
6	6	Tam O'Shanter-Sullivan	118	27446	106	90	98	124	133	107	109.7	-0.2	389.9	22	19	22	18	32	25	23	-0.22	91.1
7	7	The Beaches	063	21567	83	108	86	111	94	81	93.8	-0.14	375.6	8	23	7	10	30	20	16.3	-0.33	92.7
8	8	letown-Beaumond Heights	003	10360	56	52	39	51	50	74	53.7	0.48	714.3	10	14	13	23	42	50	25.3	0.19	482.6
9	9	Thorncliffe Park	055	21108	86	105	87	99	93	115	97.5	0.24	544.8	12	11	4	7	9	13	9.3	0.44	61.6
10	10	Danforth East York	059	17180	64	50	44	74	80	83	65.8	0.04	483.1	6	7	12	10	12	9	9.3	-0.25	52.4

5. Foursquare API for exploring common venues in the 140 Toronto City neighbourhoods.

Methodology

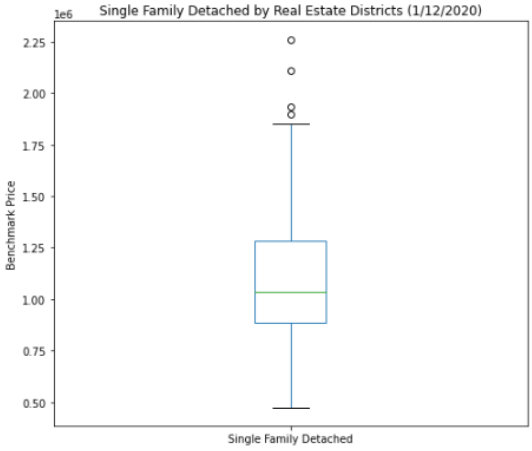
I. Plot Price Trend by Line Chart

I plot the composite benchmark price of the 4 house types in 2015 to 2019 to compare house types and visualize their price trend. This assists house buyers on making buy or not buy decision.

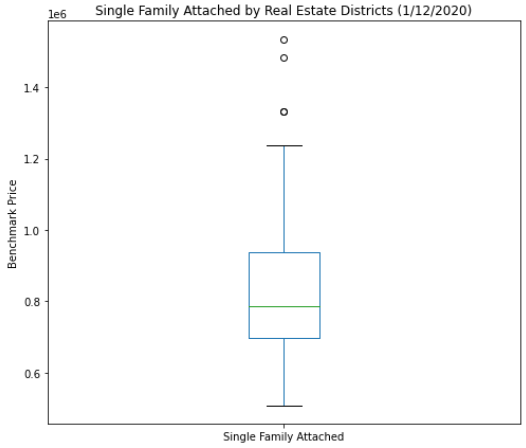


II. Box Plot the Price Range

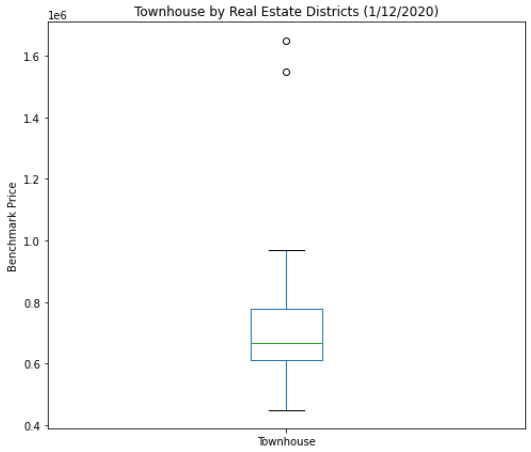
I use box plot to depict the price range of individual house type and see the price distribution. This helps house buyers to understand the affordable house types.



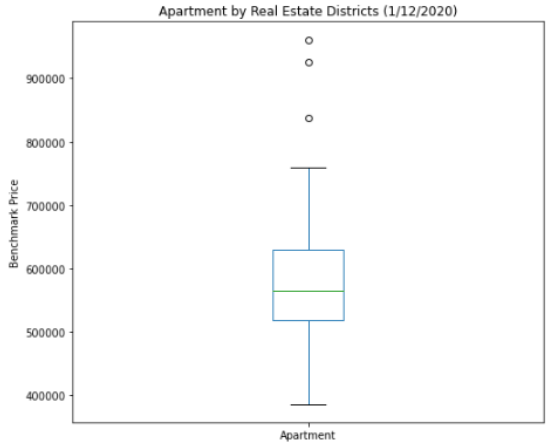
	Single Family Detached
count	7.200000e+01
mean	1.117861e+06
std	3.623379e+05
min	4.747000e+05
25%	8.833500e+05
50%	1.035650e+06
75%	1.284275e+06
max	2.262500e+06



	Single Family Attached
count	7.100000e+01
mean	8.472028e+05
std	2.310988e+05
min	5.082000e+05
25%	6.956500e+05
50%	7.848000e+05
75%	9.360000e+05
max	1.535600e+06



	Townhouse
count	5.400000e+01
mean	7.182222e+05
std	2.168184e+05
min	4.504000e+05
25%	6.121250e+05
50%	6.656500e+05
75%	7.784000e+05
max	1.650100e+06



	Apartment
count	60.000000
mean	581381.666667
std	117569.829941
min	386600.000000
25%	518250.000000
50%	565600.000000
75%	630375.000000
max	961300.000000

III. Rank and Compare Price of Neighbourhoods

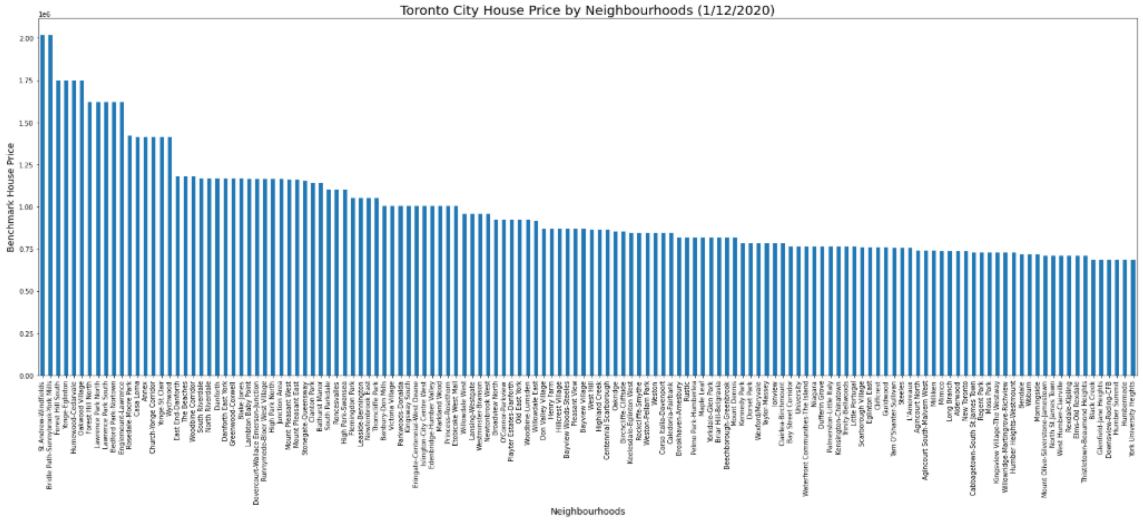
Toronto City is divided into 140 neighbourhoods. TRREB however divides house price data into 35 real estate districts of Toronto City. Although the borders of neighbourhoods and real estate districts may not perfectly align, I manually map the 140 neighbourhoods approximately into 35 real estate districts. Then by merging with house price TRREB house price data, I obtain a table of neighbourhood house price index (HPI) and house price benchmark. Among all neighbourhoods, 2 of them do not have house price data thus they are dropped from the table.

	Neighbourhood	Hood_ID	MLS_District
1	Yonge-St.Clair	97	Toronto C02
2	York University Heights	27	Toronto W05
3	Lansing-Westgate	38	Toronto C07
4	Yorkdale-Glen Park	31	Toronto W04
5	Stonegate-Queensway	16	Toronto W07
6	Tam O'Shanter-Sullivan	118	Toronto E05
7	The Beaches	63	Toronto E02
8	letown-Beaumond Heights	3	Toronto W10
9	Thorncliffe Park	55	Toronto C11
10	Danforth East York	59	Toronto E01
11	Humewood-Cedarvale	106	Toronto C03
12	Islington-City Centre West	14	Toronto W08
13	Danforth	66	Toronto E01
14	Rustic	28	Toronto W04
15	Scarborough Village	139	Toronto E08

Merging Neighbourhoods with TRREB dataset's composite house benchmark price and HPI I obtain 138 neighbourhoods' house price and index. The below bar chart shows the neighbourhoods' benchmark price, from high to low order, enabling house buyers to position their potential destinations among 140 neighbourhoods given a limited budget.

TOP 10 NEIGHBOURHOODS BENCHMARK PRICE IN CAD

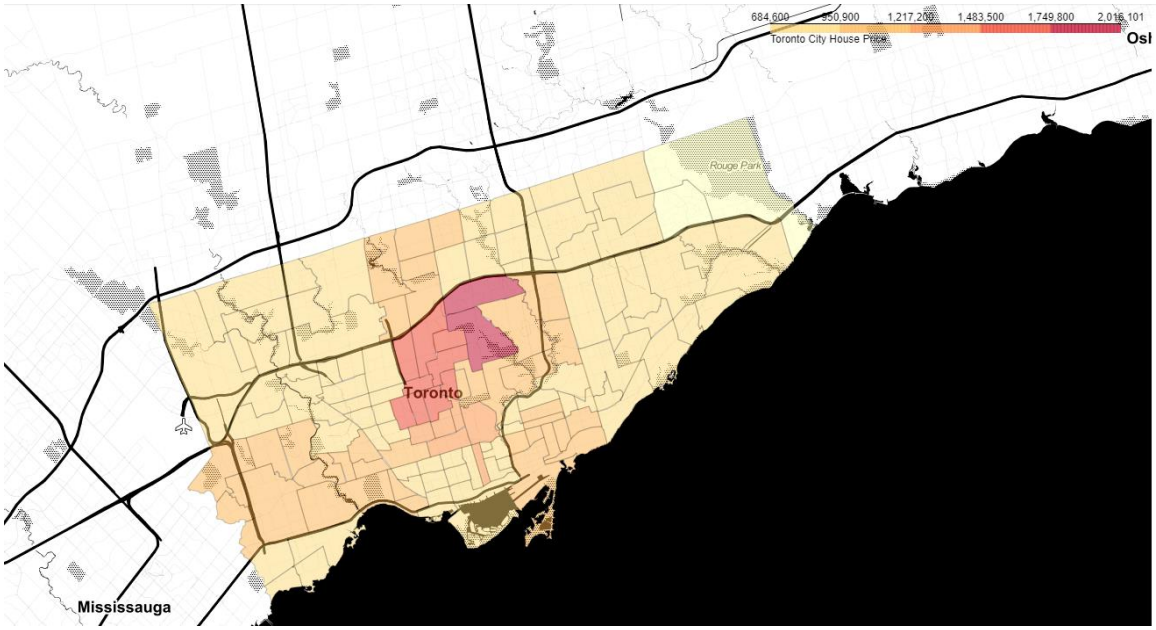
Neighbourhood	Hood_ID	MLS_District	CompBenchmark	CompIndex
St.Andrew-Windfields	40	Toronto C12	2016100.0	236.3
Bridle Path-Sunnybrook-York Mills	41	Toronto C12	2016100.0	236.3
Forest Hill South	101	Toronto C03	1748600.0	334.8
Yonge-Eglinton	100	Toronto C03	1748600.0	334.8
Humewood-Cedarvale	106	Toronto C03	1748600.0	334.8
Oakwood Village	107	Toronto C03	1748600.0	334.8
Forest Hill North	102	Toronto C04	1617800.0	259.9
Lawrence Park North	105	Toronto C04	1617800.0	259.9
Lawrence Park South	103	Toronto C04	1617800.0	259.9
Bedford Park-Nortown	39	Toronto C04	1617800.0	259.9



IV. Visualize House Price Level on Choropleth Map

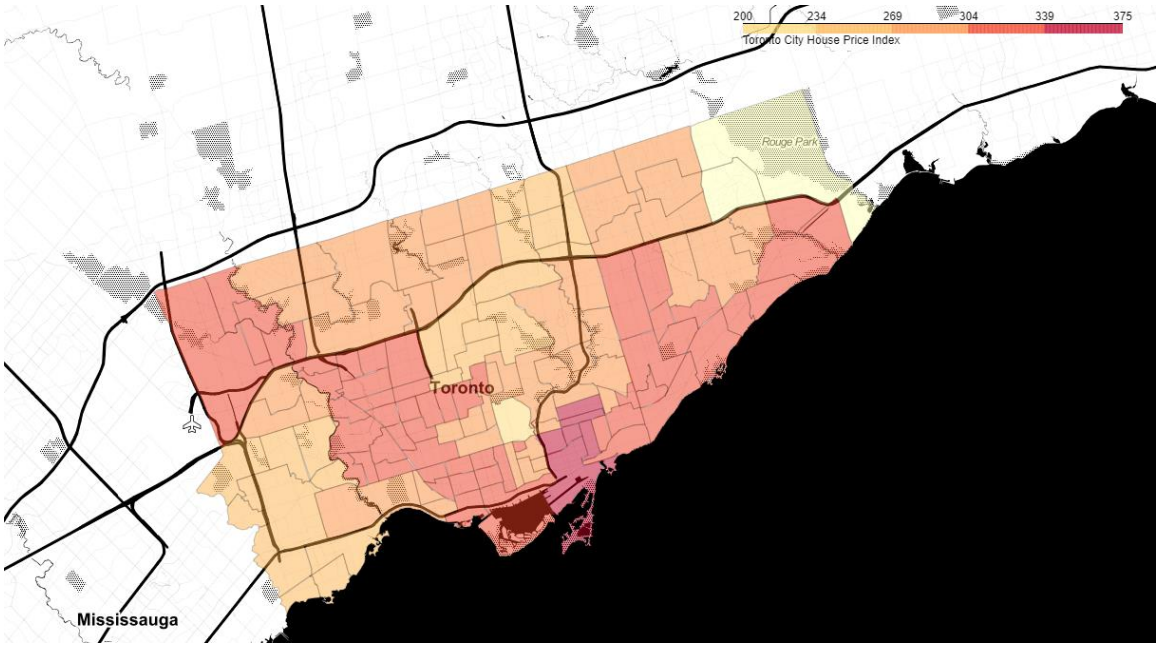
I overlay the 1/12/2020 house benchmark price on Choropleth Map. Benchmark price of Toronto City ranges from CAD 684,600 to 2,016,100. Higher-priced houses are located at the central of Toronto City.

Please note although there are 138 neighbourhoods' price data, Toronto Open Data Portal's neighbourhood geojson file only contains coordinates of 100 neighbourhoods. Neighbourhoods on the outskirts of Toronto City are thus not depicted on the Choropleth Map.



V. Visualize House Price Index (HPI) on Choropleth Map

I also create a Choropleth Map to display the neighbourhoods' HPI as of 1/12/2020. Settlers who concern about price rate change can reference this map for information.



VI. Illustrate Population Distribution of Visible Minority by Pie Chart

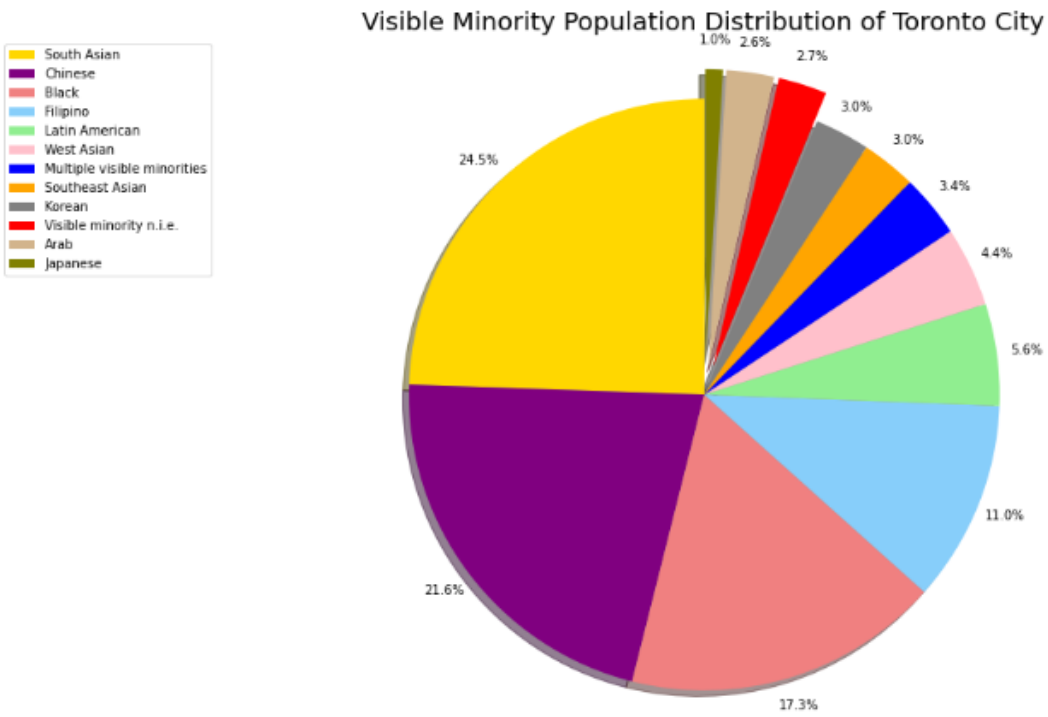
According to 2016 census profile, half of Toronto City's population are identified as visible minorities. A visible minority is defined by the Government of Canada as "persons, other than aboriginal peoples, who are non-Caucasian in race or no-white in color.". Under this definition,

Toronto City population is observed to be contributed by 13 ethnics – Latin American, South Asian, Chinese, Black, Filipino, Arab, Southeast Asian, West Asian, Korean, Japanese, Visible minority n.i.e. (not in elsewhere), Multiple visible minorities, Not a visible minority.

```
Toronto City population: 2691675
Percentage of aboriginals in Toronto City: 48.51%
Percentage of other ethnics in Toronto City: 51.49%
```

Characteristic	City of Toronto
Not a visible minority	1305815
South Asian	338965
Chinese	299465
Black	239850
Filipino	152715
Latin American	77165
West Asian	60320
Multiple visible minorities	47670
Southeast Asian	41650
Korean	41640
Visible minority n.i.e.	36975
Arab	36030
Japanese	13415

The bar chart below illustrates the proportions of each minority group that totally contributes to 51.49% of Toronto City total residences. South Asian, Chinese and Black are identified as the top 3 minority groups.



VII. Highlight Other Ethnics-Dominated Neighbourhoods on Choropleth Map

From the above analysis I learnt that the average visible minority population in Toronto City is 51.49%. Next I will check which neighbourhoods specifically have more condensed minority groups than the overage average, and identify the individual group that dominates the residency.

Toronto Open Data Portal provides breakdown of 140 neighbourhoods' population count of 13 ethnic groups. I transformed the figures into "each ethnic's proportion in the 140 neighbourhoods".

SNAPSHOT OF ETHNICS PROPORTION IN 140 NEIGHBOURHOODS

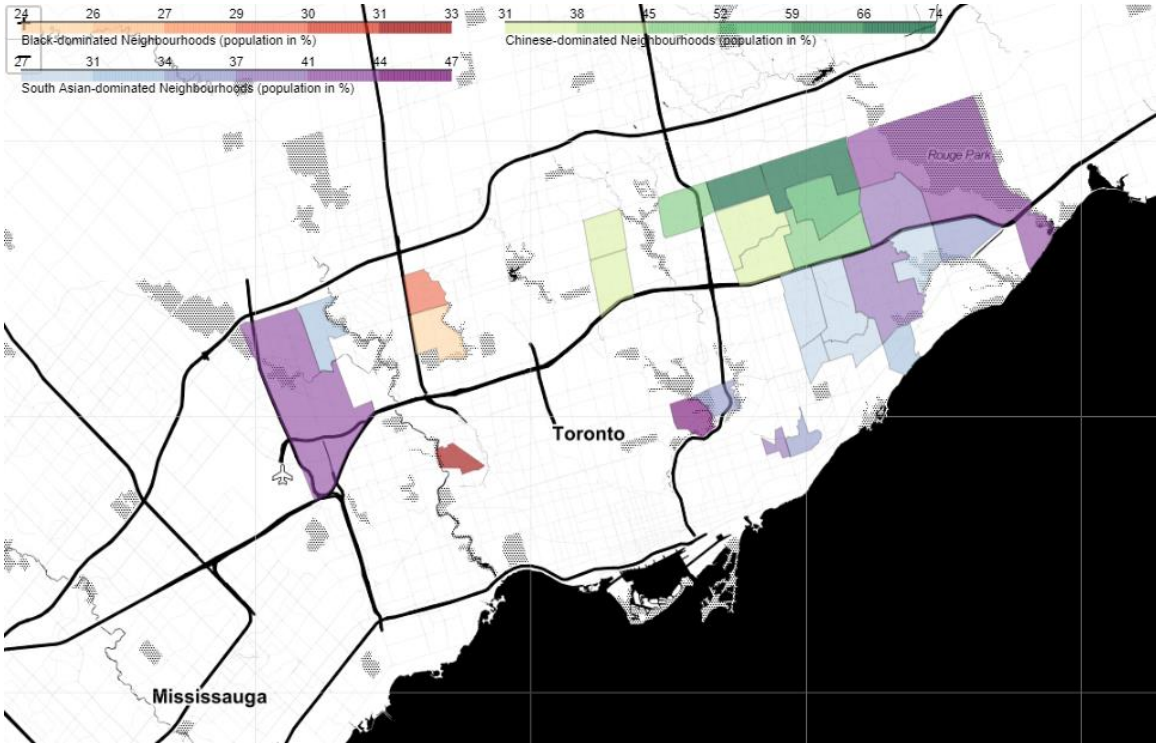
Characteristic	Latin American	South Asian	Chinese	Black	Filipino	Arab	Southeast Asian	West Asian	Korean	Japanese	Visible minority n.i.e.	Multiple visible minorities	Not a visible minority
Neighbourhood													
Agincourt North	0.59	18.43	57.08	4.65	4.92	0.95	0.90	0.43	0.24	0.21	0.85	2.18	8.55
Agincourt South-Malvern West	0.98	17.87	47.38	6.59	5.46	1.28	0.66	1.60	0.60	0.17	1.17	2.09	14.16
Bay Street Corridor	1.80	9.45	29.03	3.03	1.48	5.12	1.33	2.40	5.23	1.31	0.61	1.44	37.78
Bayview Village	1.94	6.43	27.71	3.57	3.62	1.68	0.95	11.54	7.83	0.85	0.50	2.01	31.38
Bayview Woods-Steeles	1.19	5.29	34.45	4.19	1.38	1.03	1.07	8.02	5.61	0.55	0.36	2.05	34.81

Below are the findings, telling the identified neighbourhoods, the biggest minority group in these neighbourhoods and this group's population percentage in the neighbourhoods. Settlers who have preference on either living in close proximity to their own ethnics or bending into the major Canadian aboriginals' community can reference this information.

```
57 neighbourhoods has higher minority percentage than the Toronto City average 51.49%
In 28 out of the 57 neighboorhoods, the biggest ethnic group is not the aboriginals.
Black dominates residency in 3 neighbourhoods
Chinese dominates residency in 9 neighbourhoods
South Asian dominates residency in 16 neighbourhoods
```

	Neighbourhood	Top minority	Percent	Hood_ID	MLS_District
7	Black Creek	Black	29.26	24	Toronto W05
30	Mount Dennis	Black	32.86	115	Toronto W04
16	Glenfield-Jane Heights	Black	24.34	25	Toronto W05
0	Agincourt North	Chinese	57.08	129	Toronto E07
44	Tam O'Shanter-Sullivan	Chinese	32.06	118	Toronto E05
43	Steeles	Chinese	73.14	116	Toronto E05
32	Newtonbrook East	Chinese	31.10	50	Toronto C11
53	Willowdale East	Chinese	33.78	51	Toronto C14
26	L'Amoreaux	Chinese	37.15	117	Toronto E05
28	Milliken	Chinese	71.77	130	Toronto E07
1	Agincourt South-Malvern West	Chinese	47.38	128	Toronto E07
19	Hillcrest Village	Chinese	48.61	48	Toronto C15
22	Ionview	South Asian	27.36	125	Toronto E04
50	West Humber-Clairville	South Asian	42.34	1	Toronto W10
47	Thorncliffe Park	South Asian	47.23	55	Toronto C11
45	Taylor-Massey	South Asian	38.59	61	Toronto E04
6	Bendale	South Asian	28.57	127	Toronto E09
42	Scarborough Village	South Asian	33.57	139	Toronto E08
40	Rouge	South Asian	43.80	131	Toronto E11
35	Oakridge	South Asian	35.26	121	Toronto E06
31	Mount Olive-Silverstone-Jamestown	South Asian	33.79	2	Toronto W10
13	Eglinton East	South Asian	29.15	138	Toronto E08
29	Morningside	South Asian	30.08	135	Toronto E09
15	Flemingdon Park	South Asian	34.90	44	Toronto C11
27	Malvern	South Asian	39.91	132	Toronto E11
18	Highland Creek	South Asian	36.25	134	Toronto E10
11	Dorset Park	South Asian	29.71	126	Toronto E04
55	Woburn	South Asian	40.65	137	Toronto E09

Accordingly, I created 3 individual geojson files to represent the Black, Chinese & South Asian-dominated neighbourhoods. The geojson files are then overlaid on folium as a Choropleth Map to provide a one-view picture of these neighbourhoods' locations and the ethnics' population proportions.



VIII. Explore Neighbourhoods' Nearby Venues by Foursquare API

Toronto City is divided into 140 neighbourhoods which are assigned to 103 postal codes. Below shows a fraction of the merged data of Wikipedia's postal codes to neighbourhoods mapping and Geocode Python's postal codes coordinates.

POSTAL CODES AND COORDINATES

	Postal Code	Neighbourhood	Latitude	Longitude
0	M3A	Parkwoods	43.753259	-79.329656
1	M4A	Victoria Village	43.725882	-79.315572
2	M5A	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
5	M9A	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Malvern, Rouge	43.806686	-79.194353
7	M3B	Don Mills	43.745906	-79.352188
8	M4B	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Garden District, Ryerson	43.657162	-79.378937

Note that among all neighbourhoods, only Downsview has multiple postal codes.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
40	M3K	North York	Downsview	43.737473	-79.464763
46	M3L	North York	Downsview	43.739015	-79.506944
53	M3M	North York	Downsview	43.728496	-79.495697
60	M3N	North York	Downsview	43.761631	-79.520999

I query each postal code's popular venues within 1km proximity. Below is a fraction of the result sets.

	Postal Code	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M3A	Parkwoods	43.753259	-79.329656	4b8991cbf964a520814232e3	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
1	M3A	Parkwoods	43.753259	-79.329656	4e8d9dcdd5fbbbb6b3003c7b	Brookbanks Park	43.751976	-79.332140	Park
2	M3A	Parkwoods	43.753259	-79.329656	57e286f2498e43d84d92d34a	Tim Hortons	43.760668	-79.326368	Café
3	M3A	Parkwoods	43.753259	-79.329656	58a8dcaa6119f47b9a94dc05	A&W	43.760643	-79.326865	Fast Food Restaurant
4	M3A	Parkwoods	43.753259	-79.329656	4bafa285f964a5203a123ce3	Bruno's valu-mart	43.746143	-79.324630	Grocery Store

Describe the result set to learn that:

- 4890 venues are discovered
- There are 3625 unique venues and 339 unique categories which means 1265 venues are overlapped in multiple neighbourhoods' circles.
- Venues are found in 102 postal codes and 98 neighbourhoods. 1 out of total 103 postal codes has no nearby venue discovered.

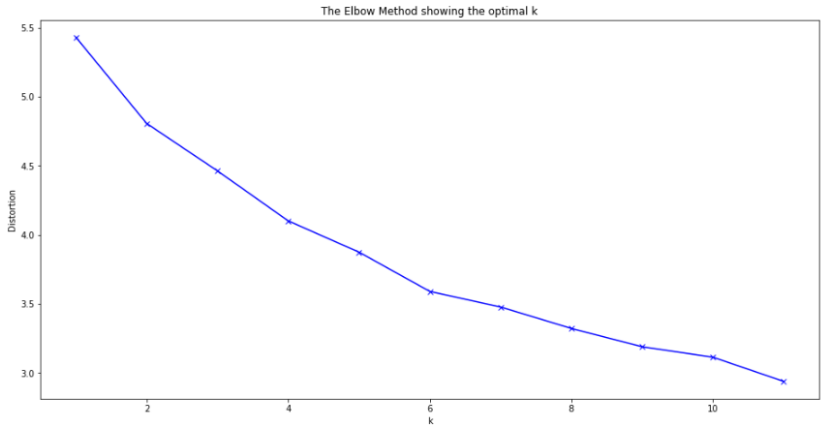
	Postal Code	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
count	4890	4890	4890.000000	4890.000000	4890	4890	4890.000000	4890.000000	4890
unique	102	98	NaN	NaN	3625	2810	NaN	NaN	329
top	M5C	The Annex, North Midtown, Yorkville	NaN	NaN	59cd51c71b0ea516e9e7b3aa	Tim Hortons	NaN	NaN	Coffee Shop
freq	100	100	NaN	NaN	9	98	NaN	NaN	371
mean	NaN	NaN	43.684694	-79.392318	NaN	NaN	43.684377	-79.392624	NaN
std	NaN	NaN	0.044955	0.068780	NaN	NaN	0.044776	0.068715	NaN
min	NaN	NaN	43.602414	-79.615819	NaN	NaN	43.593866	-79.626960	NaN
25%	NaN	NaN	43.651494	-79.419750	NaN	NaN	43.650653	-79.419247	NaN
50%	NaN	NaN	43.668999	-79.384568	NaN	NaN	43.666755	-79.386748	NaN
75%	NaN	NaN	43.709060	-79.360636	NaN	NaN	43.707303	-79.360775	NaN
max	NaN	NaN	43.815252	-79.160497	NaN	NaN	43.821167	-79.148776	NaN

I group the result set by postal code and venue category then transform the data set into a list of unique postal codes and % of each venue category being discovered in the postal code area.

```
Postal code group shape: (102, 330)
```

	Postal Code	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	American Restaurant	Amphitheater	Animal Shelter	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Dealership	Auto Garage	A Worksh
0	M1B	0.0	0.0	0.0	0.0	0.043478	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.043478	0.0	0.000000	0.0	0.0	0.043
1	M1C	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000
2	M1E	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000
3	M1G	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000
4	M1H	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.034483	0.0	0.0	0.000

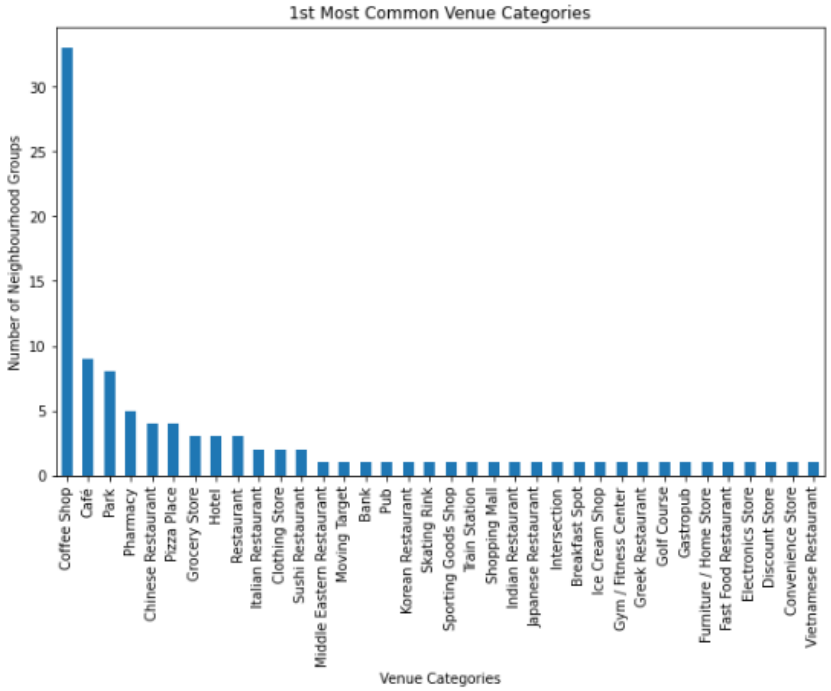
I use K-Means methodology to cluster the neighbourhoods by commonality of their nearby venue categories. I try to find the optimal k of K-Means by elbow method but the elbow is not acute. I decide to use k=6 for clustering.



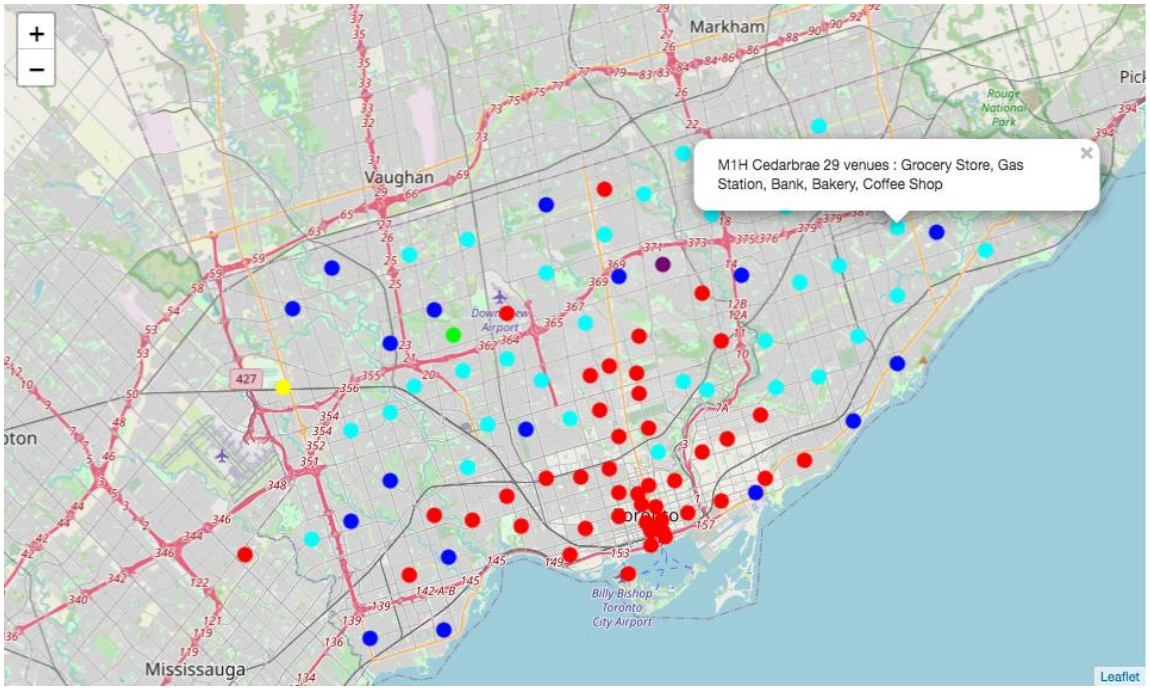
I sort out the 5 most common venue categories and venues count of each neighbourhood. Joining back postal code coordinates and the corresponding neighbourhoods.

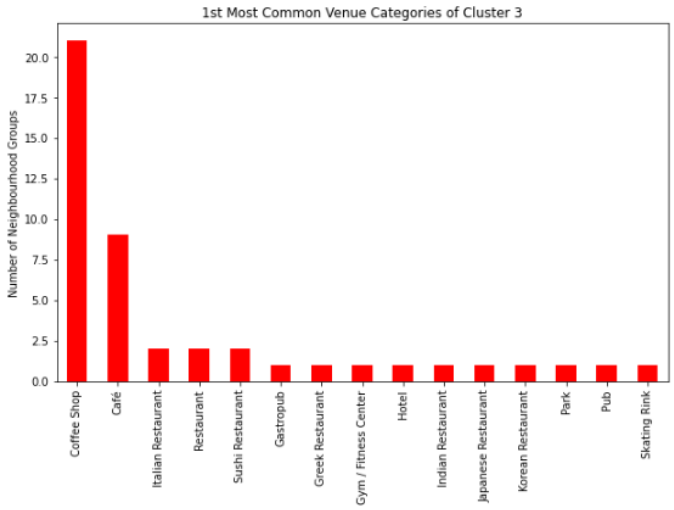
	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Count
0	M3A	North York	Parkwoods	43.753259	-79.329656	4.0	Park	Shopping Mall	Pharmacy	Bus Stop	ATM	28.0
1	M4A	North York	Victoria Village	43.725882	-79.315572	0.0	Coffee Shop	Hockey Arena	Sporting Goods Shop	Portuguese Restaurant	Pizza Place	12.0
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	3.0	Coffee Shop	Restaurant	Café	Park	Theater	100.0
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	0.0	Clothing Store	Coffee Shop	Fast Food Restaurant	Restaurant	Vietnamese Restaurant	46.0
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	3.0	Coffee Shop	Park	Sushi Restaurant	Italian Restaurant	Ramen Restaurant	100.0

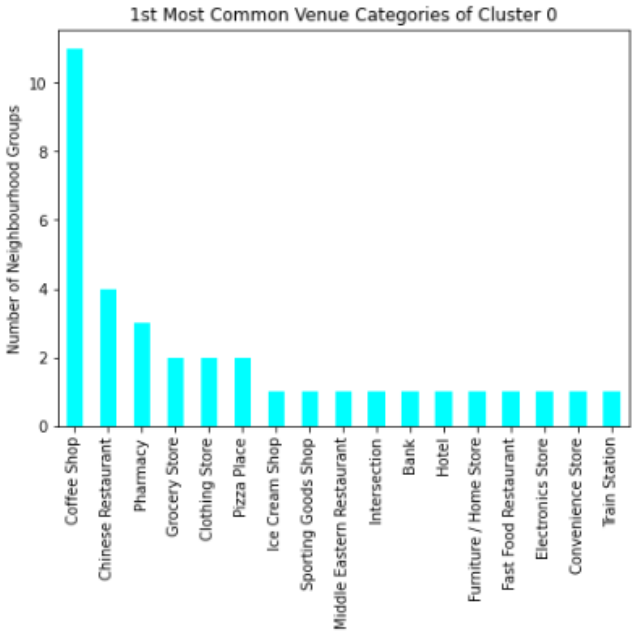
Below bar chart shows the 1st most common venues in all neighbourhoods.

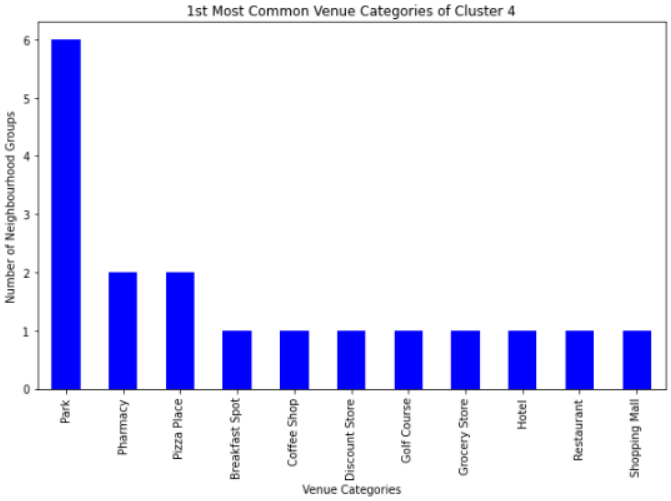


I visualize the neighbourhoods on folium map using different colors to represent the cluster group that they belong to. Markers are labelled to describe neighbourhood name, number of venues and the 1st-5th most common venue categories.









Foursquare API only discovers 3 to 4 venues in LIME, PURPLE, YELLOW clusters. They can be regarded as outliers. K-means uses % of each venue category in a neighbourhood to calculate the clustering. With only a few venues dominating the % of venue categories in one neighbourhood will tend to form a highly distinctive cluster by the neighbourhood itself. This may explain why the elbow chart is flattened and unable to form an acute elbow as k increases.

LIME Cluster

Postal Code	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
M3M	Downsview	43.728496	-79.495697	4d39fb9434ee37047185729b	Hung Long	43.726524	-79.483659	Vietnamese Restaurant
M3M	Downsview	43.728496	-79.495697	4c59fce9d3aee21e1aae6955	Roding Park	43.728655	-79.492918	Baseball Field
M3M	Downsview	43.728496	-79.495697	4c06fb96a0129c74b66dd2c9	Pho Mi Nha Trang	43.722709	-79.498058	Vietnamese Restaurant
M3M	Downsview	43.728496	-79.495697	5b7bc7ca123a19002c11786b	Thai Express	43.724104	-79.488435	Restaurant

PURPLE Cluster

Postal Code	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
M2L	York Mills, Silver Hills	43.75749	-79.374714	54c93d4c498e80ef9b327c09	Swimming Pool	43.750994	-79.374365	Pool
M2L	York Mills, Silver Hills	43.75749	-79.374714	4dd5826ec65bee535a834718	Talara Park	43.765457	-79.377301	Park
M2L	York Mills, Silver Hills	43.75749	-79.374714	4c2e147311cb9c74ef0673b3	St. Andrews Park	43.757309	-79.386616	Park
M2L	York Mills, Silver Hills	43.75749	-79.374714	4e88bdd0d5fb8ad9ceefeb58	Ames Park	43.751868	-79.365473	Park

YELLOW Cluster

Postal Code	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
M9W	Northwest, West Humber - Clairville	43.706748	-79.594054	4be452122457a593c40eaa15	Tim Hortons	43.714657	-79.593716	Coffee Shop
M9W	Northwest, West Humber - Clairville	43.706748	-79.594054	4fa474cce4b02175375ac90b	227 Lounge	43.703184	-79.588193	Lounge
M9W	Northwest, West Humber - Clairville	43.706748	-79.594054	55dc6273498eea1c592572ef	Up Express On Board	43.705154	-79.604830	Moving Target

IX. Analyze Crimes Rates by Crime Type

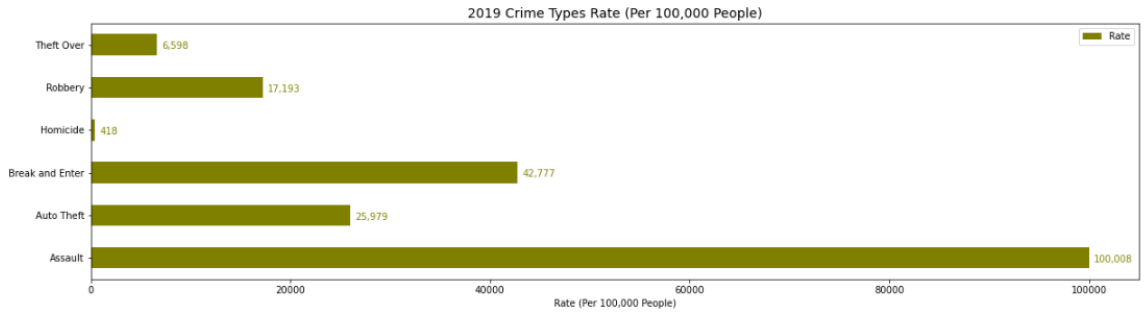
Kaggle provides 2014-2019 crime rates, i.e. number of crimes per 100,000 people, of each individual neighbourhood. I will study the 2019 data as it is the latest. Counts are available for Assault, Auto Theft, Break and Enter, Homicide, Robbery and Theft Over.

FRACTION OF 140 NEIGHBOURHOODS' 2019 CRIME RATES

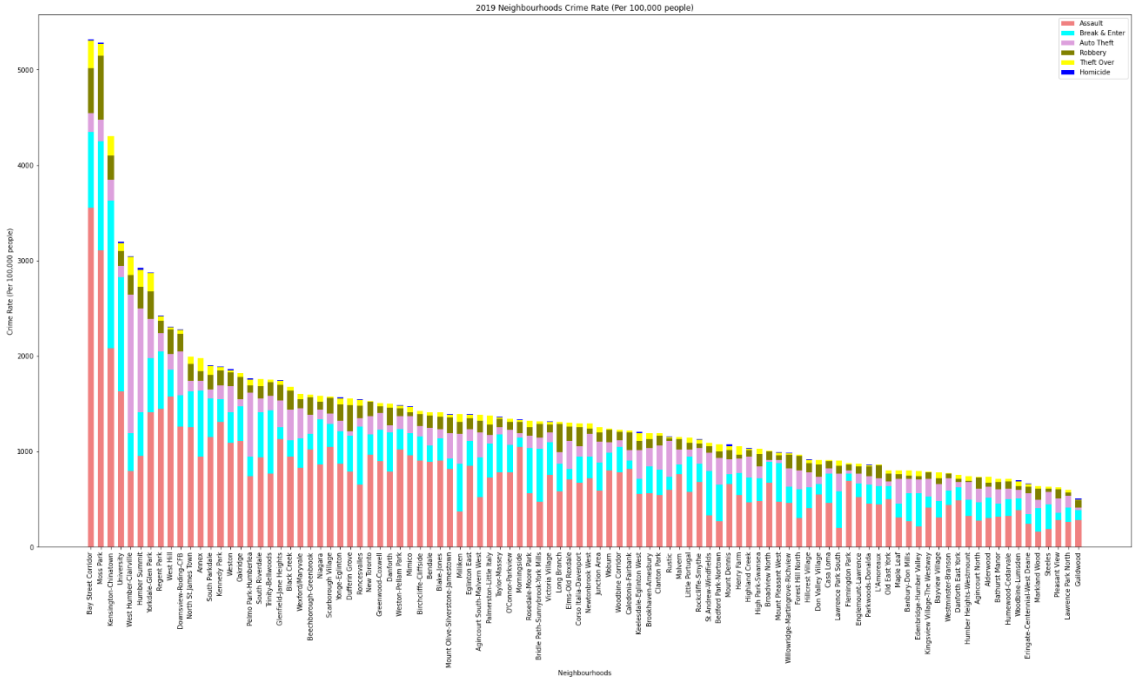
	Neighbourhood	Hood_ID	Population	Assault_Rate_2019	AutoTheft_Rate_2019	BreakandEnter_Rate_2019	Homicide_Rate_2019	Robbery_Rate_2019	TheftOver_Rate_2019	Total_Crime_Rate_2019
0	Yonge-St.Clair	97	12528	295.3	47.9	223.5	0.0	31.9	47.9	646.5
1	York University Heights	27	27593	1340.9	521.9	391.4	0.0	286.3	101.5	2642.0
2	Lansing-Westgate	38	16164	445.4	198.0	241.3	0.0	68.1	68.1	1020.9
3	Yorkdale-Glen Park	31	14804	1411.8	412.1	567.4	6.8	283.7	195.9	2877.7
4	Stonegate-Queensway	16	25051	327.3	135.7	255.5	0.0	87.8	16.0	822.3

Describe the dataset to find that Assault has the highest overall crime rate among the 6 crime types in Toronto City. A comparison of crime types can be clearly presented with the below horizontal bar chart.

	Neighbourhood	Hood_ID	Population	Assault_Rate_2019	AutoTheft_Rate_2019	BreakandEnter_Rate_2019	Homicide_Rate_2019	Robbery_Rate_2019	TheftOver_Rate_2019	Total_Crime_Rate_2019
count	140	140.0000	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000
unique	140	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
top	Mount Dennis	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
freq	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	70.5000	19511.221429	714.346429	185.565000	305.550714	2.987143	122.813571	47.132143	1378.395000
std	NaN	40.5586	10033.589222	498.301136	169.631765	216.778588	4.812197	86.312788	41.478434	797.675784
min	NaN	1.0000	6577.000000	161.100000	26.100000	67.500000	0.000000	7.500000	0.000000	504.200000
25%	NaN	35.7500	12019.500000	392.300000	102.800000	173.575000	0.000000	70.350000	22.750000	869.200000
50%	NaN	70.5000	16749.500000	592.600000	147.200000	258.600000	0.000000	104.200000	37.400000	1232.900000
75%	NaN	105.2500	23854.500000	875.675000	199.250000	361.225000	5.725000	148.575000	54.650000	1532.300000
max	NaN	140.0000	65913.000000	3550.800000	1446.900000	1549.200000	24.200000	668.100000	283.000000	5314.500000



I rank the neighbourhoods by their total crime rates and plot them all in stacked bar charts, in high to low order. People who are sensitive to community safety level can easily spot the neighbourhoods which they would like to avoid. By inspecting the stacked bars, the graph can tell 3 neighbourhoods have exceptionally high Auto Theft rate than other neighbourhoods, they are West Humber-Clairville, Humber Summit and Pelmo Park-Humberlea.

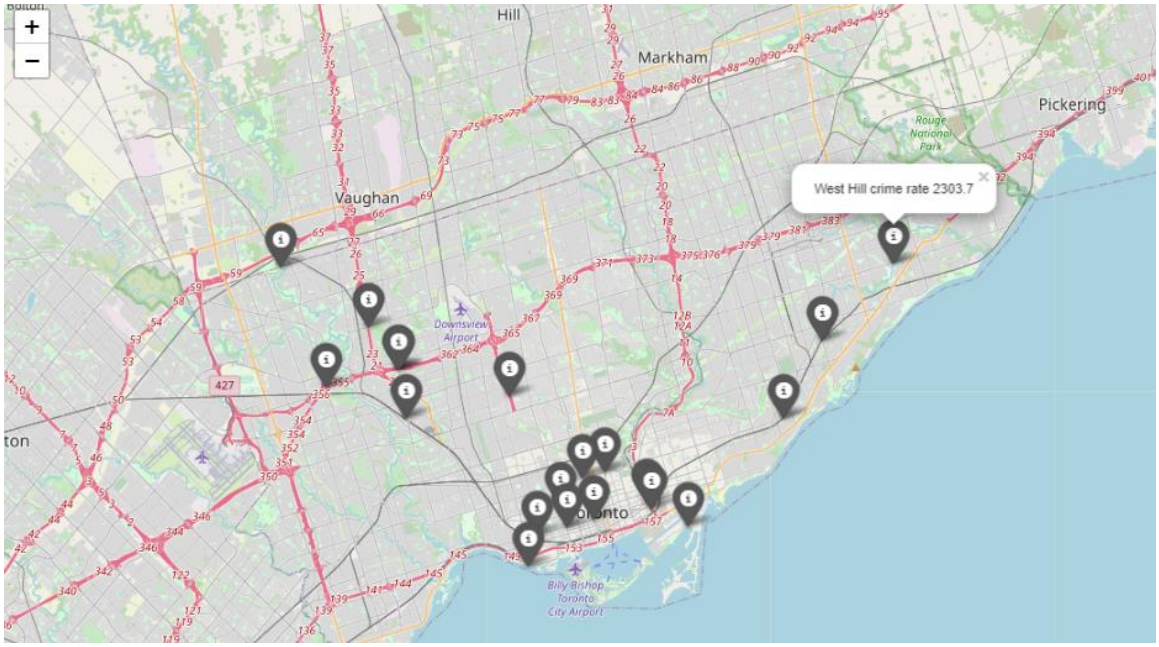


X. Pin the Least Safe Neighbourhoods on Folium Map

Considering that settlers will be more attentive to of high-crime-rate neighbourhoods, I further pin 20 highest-crime-rate neighbourhoods on folium map for anyone to overview the locations of these neighbourhoods on map. The neighbourhoods are pinned on an interactive folium map by markers as well as tips labelling neighbourhood names and crime rate.

20 HIGHEST CRIMES RATE NEIGHBOURHOODS

Bay Street Corridor, Moss Park, Kensington-Chinatown, University, West Humber-Clairville, Humber Summit, Yorkdale-Glen Park, Regent Park, West Hill, Downsview-Roding-CFB, North St.James Town, Annex, South Parkdale, Kennedy Park, Weston, Oakridge, Pelmo Park-Humberlea, South Riverdale, Trinity-Bellwoods, Glenfield-Jane Heights



Results

House Price

Toronto house price shows an upwards trend in past 5 years. Special attention should be paid to apartment which begins to go downwards since 2020. Townhouse's rise is flattening as well.

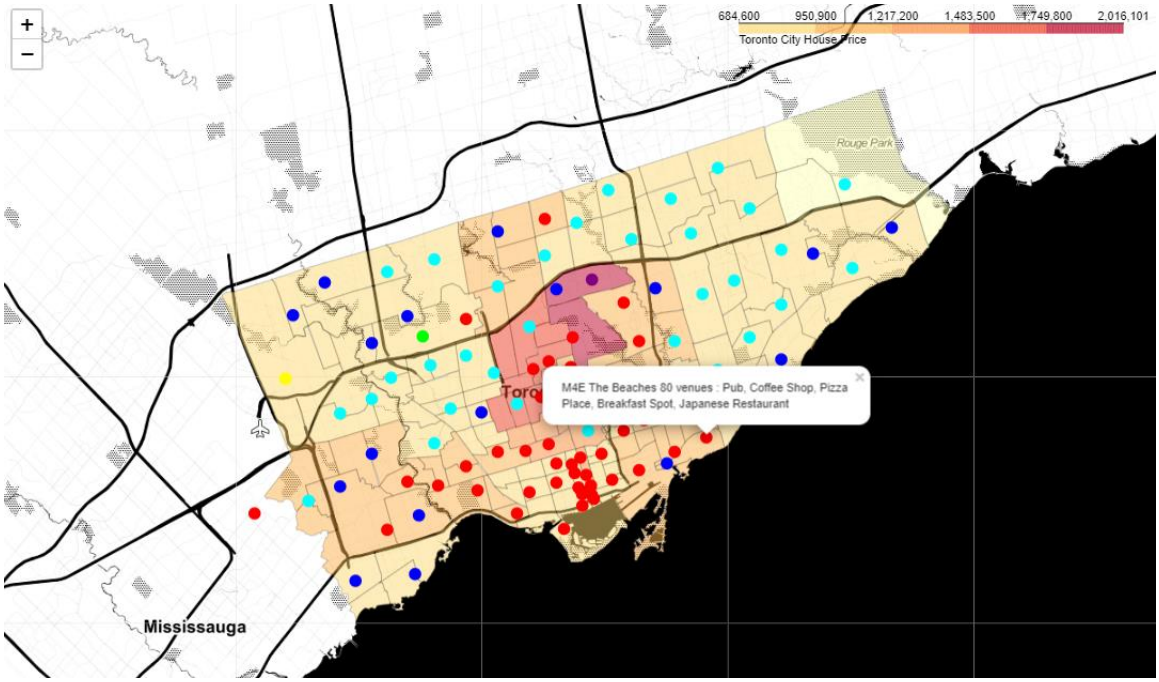
Benchmark price by different house types are summarized in below table. Potential house buyers can reference this to determine their affordable house types.

House Type (as of 1-Dec-2020)	**Mean Price (CAD)**	**Standard Deviation from Mean (CAD)**
Single Family Detached	1,117,861	362,338
Single Family Attached	847,203	231,099
Townhouse	718,222	216,818
Apartment	581,382	117,570

When measuring by the composite house price, the highest-priced houses are located in 17 neighbourhoods at central of Toronto City. Potential house buyers can reference below table and map to determine their affordable destinations. There color dots will be explained in later section. Please note Benchmark Price is used in this study. Researches on HPI or house rate by size will be conducted for comparison in future study.

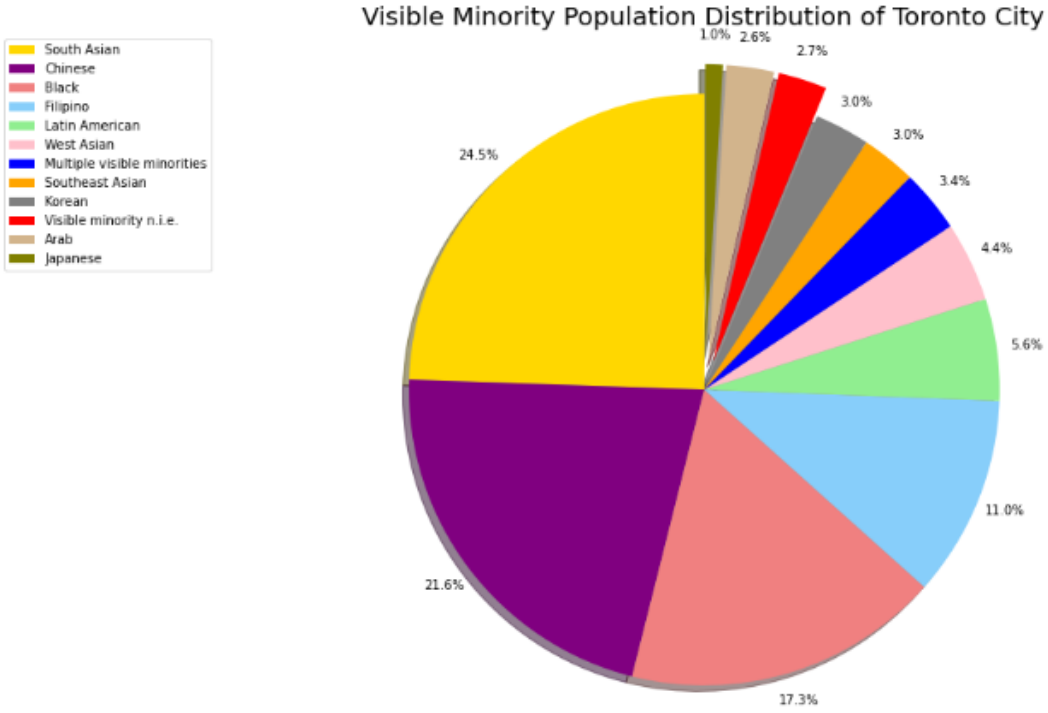
Neighbourhood	**Composite Benchmark Price (CAD)**	**Price Bin**
St.Andrew-Windfields	$2,016,100	
Bridle Path-Sunnybrook-York Mills	$2,016,100	
Forest Hill South	$1,748,600	
Yonge-Eglinton	$1,748,600	
Humewood-Cedarvale	$1,748,600	
Oakwood Village	$1,748,600	
Forest Hill North	$1,617,800	
Lawrence Park North	$1,617,800	
Lawrence Park South	$1,617,800	
Bedford Park-Nortown	$1,617,800	
Englemount-Lawrence	$1,617,800	
Rosedale-Moore Park	$1,423,000	
Casa Loma	$1,413,700	
Annex	$1,413,700	
Church-Yonge Corridor	$1,413,700	
Yonge-St.Clair	$1,413,700	
Wychwood	$1,413,700	

PRICE LEVEL MAP



Ethnics Population & Distribution

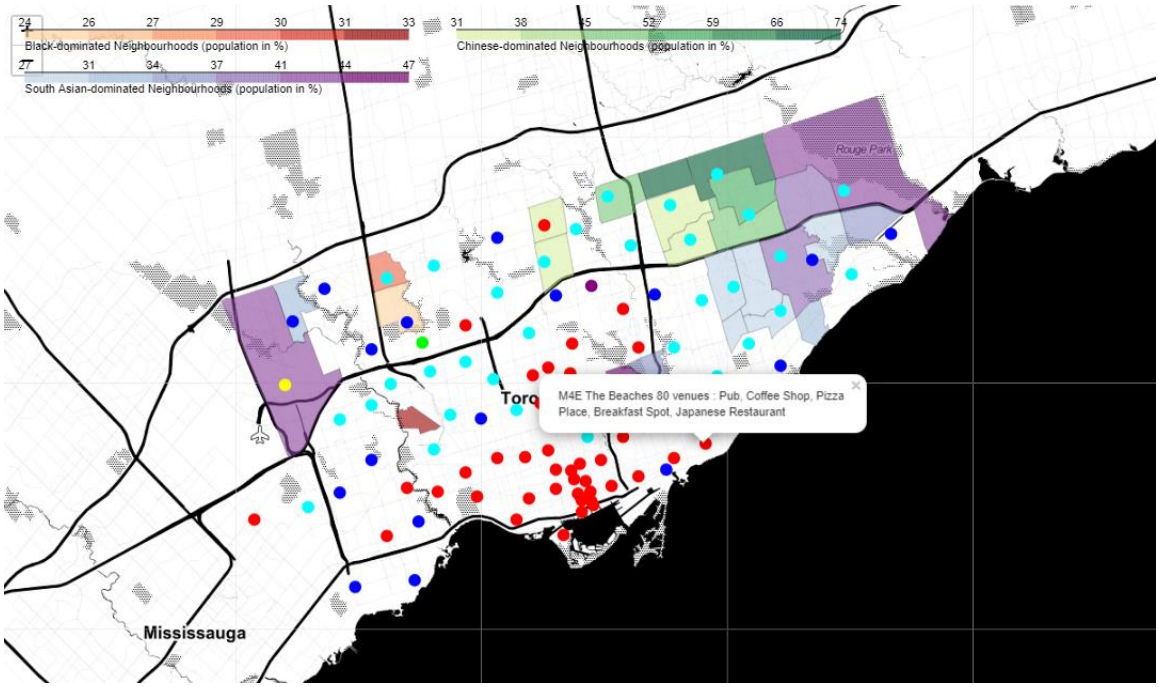
According to 2016 census profile, 51.49% of Toronto City's population are identified as visible minorities. A breakdown of the minority by 12 ethnic groups is shown in below pie chart. **South Asian, Chinese and Black** are the 3 biggest ethnic groups which largely contribute to the 51.49% minority residences, that is equivalent to about **33% of 2.9M Toronto City Total Population**.



In 27 out of the total 140 neighbourhoods, residency % from a single ethnic outreaches that of the aboriginals. Potential settlers who have preference on either living in close proximity to their own ethnics or bending into the major Canadian aboriginals' community can reference below table and map (the colored dots will be explained in next section).

Neighbourhood	Top minority	Percent
Black Creek	Black	29.26
Mount Dennis	Black	32.86
Glenfield-Jane Heights	Black	24.34
Agincourt North	Chinese	57.08
Tam O'Shanter-Sullivan	Chinese	32.06
Steeles	Chinese	73.14
Newtonbrook East	Chinese	31.10
Willowdale East	Chinese	33.78
L'Amoreaux	Chinese	37.15
Milliken	Chinese	71.77
Agincourt South-Malvern West	Chinese	47.38
Hillcrest Village	Chinese	48.61
Ionview	South Asian	27.36
West Humber-Clairville	South Asian	42.34
Thorncliffe Park	South Asian	47.23
Taylor-Massey	South Asian	38.59
Bendale	South Asian	28.57
Scarborough Village	South Asian	33.57
Rouge	South Asian	43.80
Oakridge	South Asian	35.26
Mount Olive-Silverstone-Jamestown	South Asian	33.79
Eglinton East	South Asian	29.15
Morningside	South Asian	30.08
Flemingdon Park	South Asian	34.90
Malvern	South Asian	39.91
Highland Creek	South Asian	36.25
Dorset Park	South Asian	29.71
Woburn	South Asian	40.65

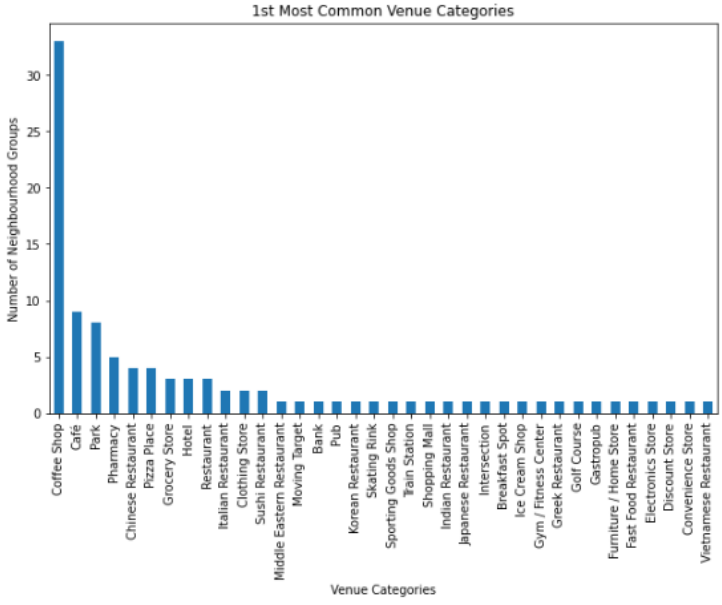
MINORITY DOMINATED NEIGHBOURHOODS MAP



Neighbourhood Features

K-means methodology is applied to cluster neighbourhoods by the venue categories explored by Foursquare API. 3625 unique venues of 339 unique categories are found in Toronto City neighbourhoods. 6 clusters are overlaid on the above maps, representing by dots in different colors. The pins are labeled to describe the neighbourhood name, number of venues found in 1 km proximity and the 1st-5th most common venue categories.

Toronto City neighbourhoods' 1st most common venues are summarized in below bar charts. Coffee Shop completely stands out, following it are Café, Park, & Pharmacy.



RED CLUSTER

It can be seen that coffee shop and cafe are highly centralized in the RED neighbourhoods which are largely located along Yonge Street. RED cluster also major in wide variety of international cuisines.

CYAN CLUSTER

In CYAN neighbourhoods, coffee shop is still the 1st most common venue although the number is just half of that of RED cluster. Dining places appear to have less variety in CYAN cluster, chinese restaurant stands out among all dining options. CYAN cluster is well-occupied by variety of merchandizes including pharmacy, grocery store, clothing store, furniture store, electronics store, convenience store. It also has park, hotel, transportations and intersection. CYAN cluster is more multi-functional.

BLUE CLUSTER

BLUE cluster is specially featured with an abundance of parks. BLUE neighbourhoods also serve larger leisure venues such as discount store, shopping mall and golf course

The LIME PURPLE YELLOW clusters comprised of only 1 neighbourhood each, and are regarded as outliers since only 3 to 4 venues are discovered by Foursquare API.

LIME CLUSTER 2 Vietnamese restaurant, 1 Baseball Field, 1 Restaurant

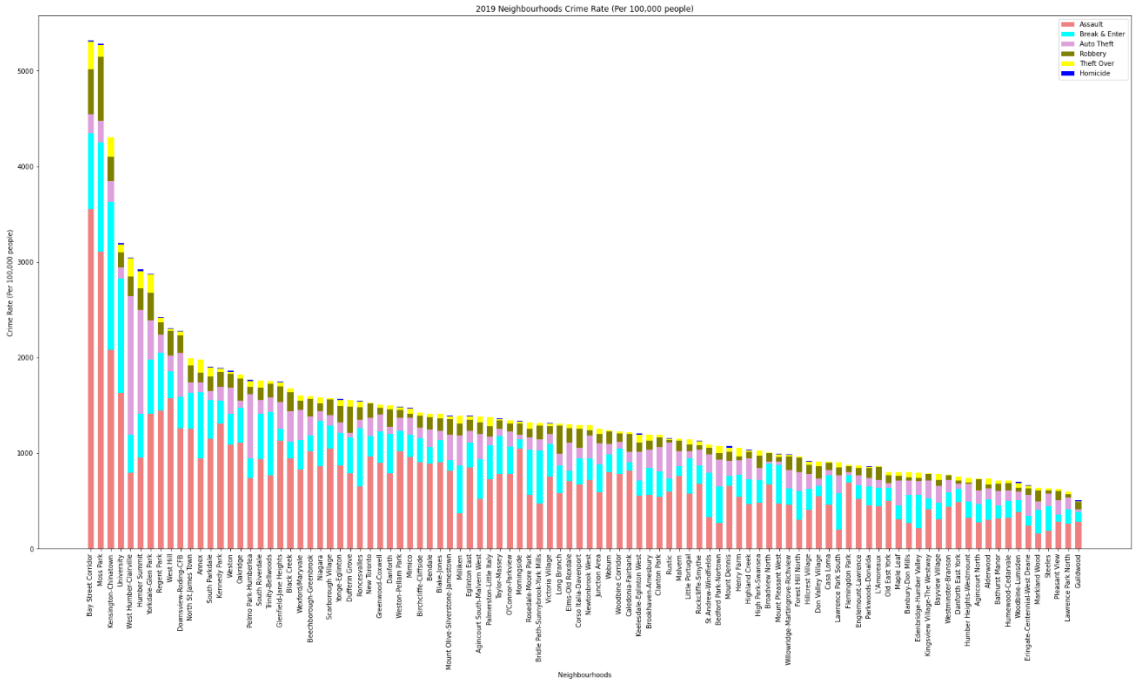
PURPLE CLUSTER 1 Pool, 3 Parks

YELLOW CLUSTER 1 Coffee Shop, 1 Lounge, 1 Moving Target

Community Safety

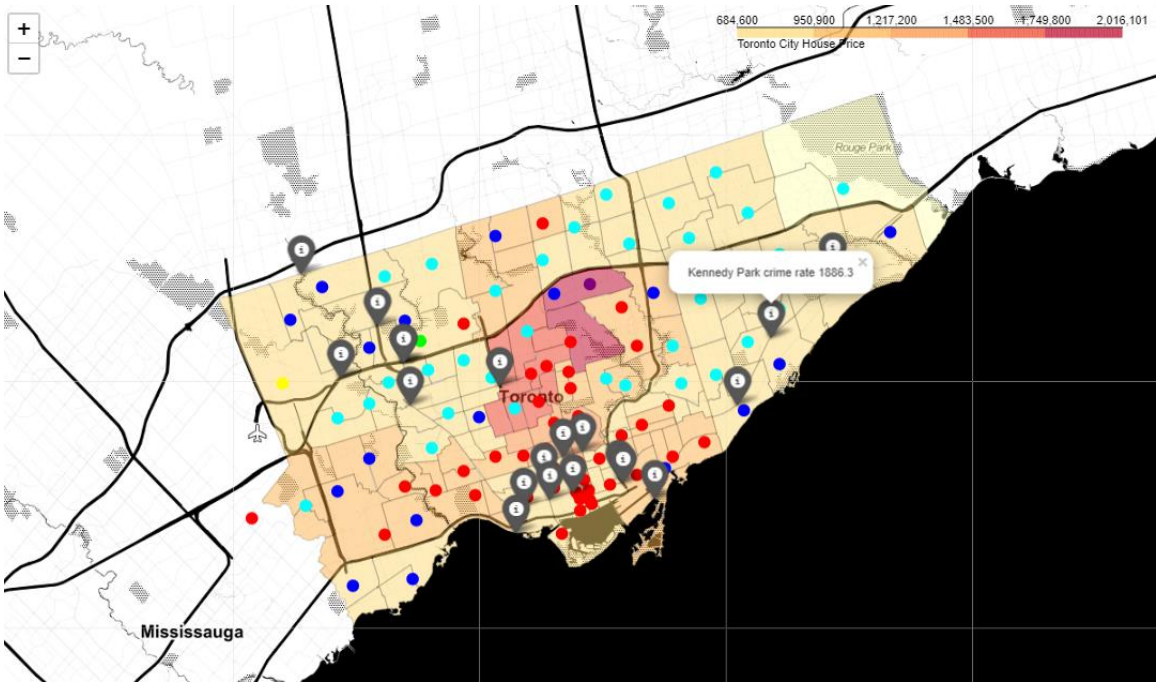
This section is for people who are sensitive to community safety to shortlist the least safe neighbourhoods which they may wan to avoid. 2019 Crime Rate (per 100,000 people) per neighbourhood is laid out in below stacked bar chart which provides 3 insights:

- Assault has the highest overall crime rate among the 6 crime types. Following after are Break and Enter, Auto Theft, Robbery. Crime rates of Theft Over and Homicide are comparatively low.
- 3 neighbourhoods have exceptionally high Auto Theft rate than other neighbourhoods, they are West Humber-Clairville, Humber Summit and Pelmo Park-Humberlea.
- The 20th highest-crime-rate neighbourhoods are: Bay Street Corridor, Moss Park, Kensington-Chinatown, University, West Humber-Clairville, Humber Summit, Yorkdale-Glen Park, Regent Park, West Hill, Downsview-Roding-CFB, North St.James Town, Annex, South Parkdale, Kennedy Park, Weston, Oakridge, Pelmo Park-Humberlea, South Riverdale, Trinity-Bellwoods, Glenfield-Jane Heights.

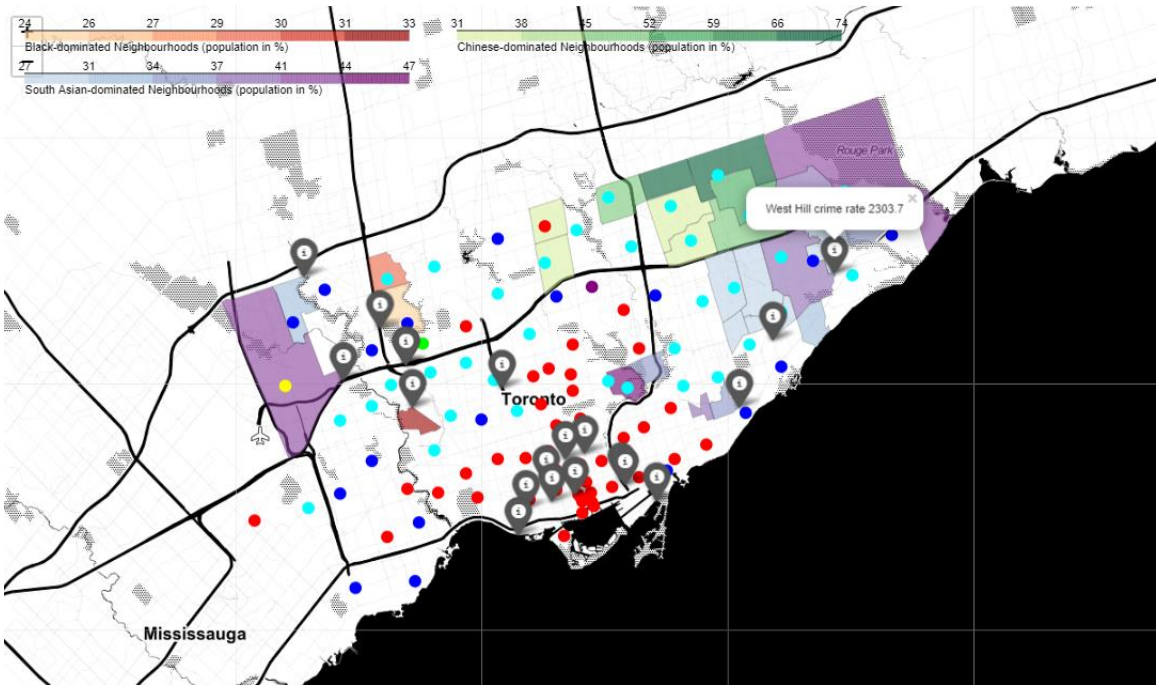


Maps are further enhanced to add markers for the 20 highest-crime-rate neighbourhoods. Markers are labeled to describe the neighbourhood name and the respective 2019 total crime rate. Readers can overview neighbourhood clusters and least safe neighbourhood on either the Price Level Map or the Minority-dominated Map.

PRICE LEVEL MAP



MINORITY DOMINATED NEIGHBOURHOODS MAP



Discussion

Looking at the enhanced Price Level Map above, I would suggest the CYAN cluster neighbourhoods are more favorable for living. The neighbourhoods are found to be featured with multi-functional venues around, fair house price and less crimes. While whether ones prefer to stay close to specific ethnics' zone is very personal consideration.

However, there are a few missing pieces in this study due to the limitation of data sources.

As mentioned earlier, TRREB divides Toronto City's houses into 35 real estate districts. The house price statistics is detailed down to 35 districts but the entire area actually covers 140 neighbourhoods. The research can be more precise if each neighbourhood's house price is available as input for my study.

Besides, Toronto Open Data Portal's neighbourhood geojson file only contains 100 neighbourhoods's coordinates. 40 neighbourhoods on the outskirts of Toronto City are missing from the Choropleth Map sketch.

The statistics of visible minorities is originated from 2016 Census of Population. The recent actual figures could have changed. Canada's Census of Population is usually held every 5 year, I look to a refreshed dataset be coming soon.

In the neighbourhood clustering study, totally 4890 venues are discovered within 1km proximity and among these venues, only 3625 venues are unique. That means 1265 venues are overlapped in more than 1 neighbourhoods' circles. This problem is unavoidable due to difference size and irregular boundary shape of the neighbourhoods. The trial of using 500m

proximity proves that it does not significantly reduce the overlaps, rather it reduces the number of venues by half which does not provide sufficient venue category features and accurate averaged category counts for the clustering algorithm. After all, I determine 1km as an appropriate walkable distance. Although the overlap is not perfect condition, it indeed does not violate the principle in this neighbourhood clustering model.

I try to find the optimal k for K-Means clustering by elbow method but the elbow is not acute. I ended up take k=6. Among the 6 clusters, Foursquare API only discovers 3 to 4 venues in LIME, PURPLE, YELLOW clusters. K-means uses % of each venue category in a neighbourhood to calculate the clustering. With only a few venues dominating the % of venue categories in one neighbourhood will tend to form a highly distinctive cluster by the neighbourhood itself. This may explain why the elbow chart is flattened and unable to form an acute elbow as k increases.

I am wondering if there is another clustering algorithm or a bigger venue dataset that yields better results.

Conclusion

In spite of some limitations, this study achieves overview of the latest house price by type and the price trend of Toronto City; exploration of livable neighbourhood zones in terms of community environment and safety; and distribution of major ethnics. As discussed above, it is proposed to further deepen research on 1) house price by HPI or factoring in house size; 2) and the neighbourhood venues.

When ones ponder immigration and the destination to settle in, the decision is influenced by a basket of factors and concerns. This study only focus on 4 aspects. Statistical researches can possibly be extended to include e.g. general living costs, job opportunities, educations etc.