

# Battle of Neighborhoods

Pablo Fernandez Cid

01/05/2020

# 1. Introduction

- Currently, for this study case, a person is living in Etobicoke borough, more specifically in the neighborhood of Thistletown, with 10,360 residents and multiple services
- The majority of the residents live in houses, are married and have at least one children, according to the data from Neighborhood Profiles. The person is not married and doesn't have any child, and currently is living alone. He was born in Europe and he enjoys the venues inside his neighborhood.
- Due to a job opportunity in Toronto East, our subject is thinking in moving to a new neighborhood, closer to this new possible job, which is located in Guildwood, East Toronto.
- This analysis could be interesting also for people in the same situation as our subject or people interested in the city of Toronto.

# Problem



## 2. Data Preparation

- Foursquare API: Extraction of information regarding the venues per each Neighborhood
- Neighborhood Profiles: information in order to understand the population, their income and their origin for each neighborhood. Data is extracted from multiple sources and merged by Open Data Toronto. See [link here](#) for more information
- Police Report information: Information regarding police report from 2014 to 2019 ([link here](#)). This data will be used to understand the security of each neighborhood, having in mind the number per each type of crime (Assault, Auto Theft, Break and Enter, Robbery and Theft Over) per year.

## 2.1. Neighborhood Profile

- Extract only the necessary rows:
- Format the numbers and convert them to integer in order to operate with them

Topic	Characteristic	City of Toronto	Thistletown-Beaumont Heights
Immigrants by selected place of birth	Asia	674490	2655
	Americas	212010	1205
	Europe	298270	1140
	Africa	77445	460
	Oceania and other places of birth	3780	10
	Total	1265995	5470

## 2.2. Police Report information

- Some of the Neighborhoods names are different from Neighborhood profile, so these names have been corrected. Also, it is necessary to remove a space at the end of all the Neighborhoods names
- Group by Neighborhood
- Calculate per each Neighborhood the quantity of crimes for a specific year and per 1000 habitants in that Neighborhood.

	Population	Assault	Auto Theft	Break and Enter	Robbery	Theft Over
Neighbourhood						
Agincourt North	29113	2.713564	1.442654	1.923539	1.202212	0.068698
Agincourt South-Malvern West	23757	5.177421	2.609757	4.209286	1.220693	0.631393
Alderwood	12054	2.986560	1.161440	2.156960	0.414800	0.580720
Annex	30526	9.434580	0.982769	6.944899	1.015528	1.375876
Banbury-Don Mills	27695	2.671962	1.516519	2.924716	0.361076	0.505506

## 2.3 Foursquare

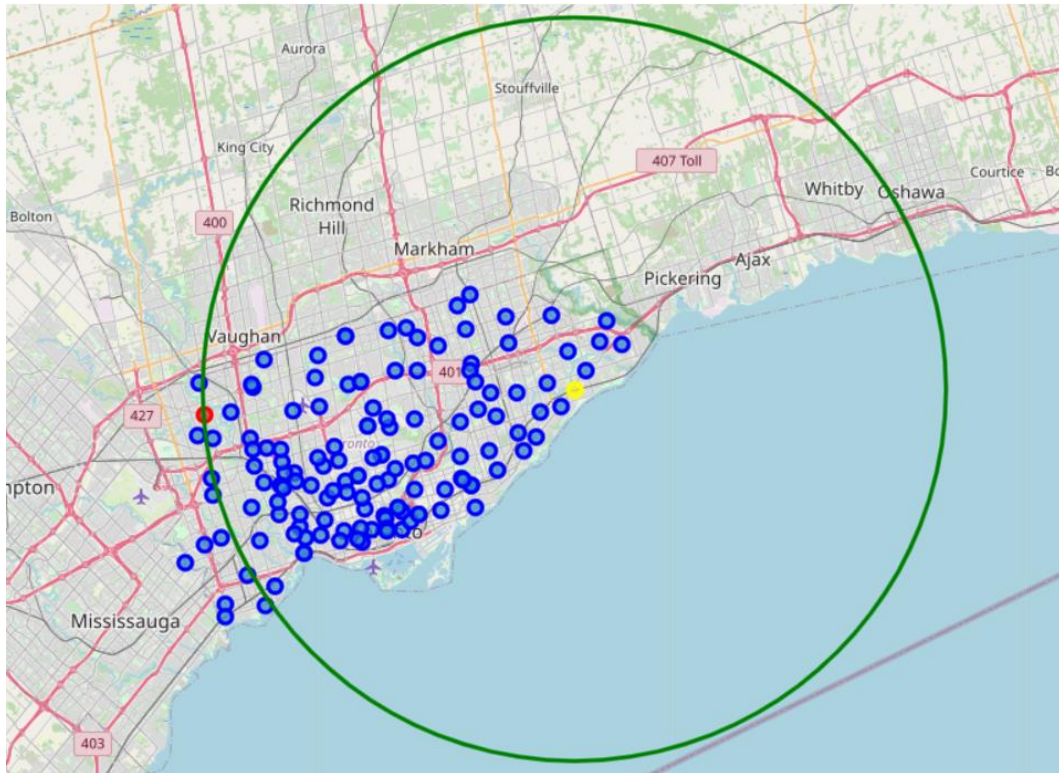
- Get data regarding the location per each Neighborhood using Geolocator.
- From the list of venues, remove the venues not interesting for our subject due to their interest (maintain restaurants and fitness venues), situation(single) and gender (male).
- Calculate the frequency per each kind of venue per each neighborhood

	Latitude	Longitude	Cluster Labels	ATM	Afghan Restaurant	American Restaurant	Amphitheater	Antique Shop	Arcade	Argentinian Restaurant	...	Transportation Service
Neighborhood												
Agincourt North	43.808	-79.2664	5.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0
Agincourt South-Malvern West	43.7892	-79.263	5.0	0.0	0.0	0.066667	0.0	0.0	0.0	0.0	...	0.0
Alderwood	43.6017	-79.5452	5.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0
Annex	43.6703	-79.4071	5.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0
Banbury-Don Mills	43.7348	-79.3572	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...	0.0



# 3. Methodology

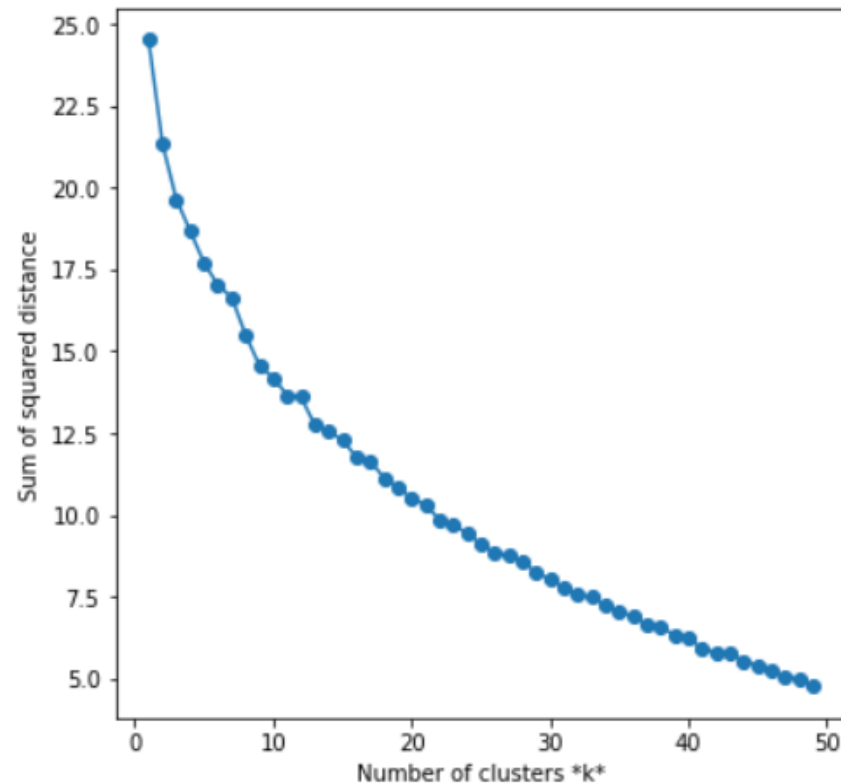
For the election of the best neighborhood, three criteria will be used based on category venues, criminality per neighborhood and profile of the neighborhood, and also having in mind the distance of the new neighborhood to the new job (yellow in map below) compared with the current one (red mark)



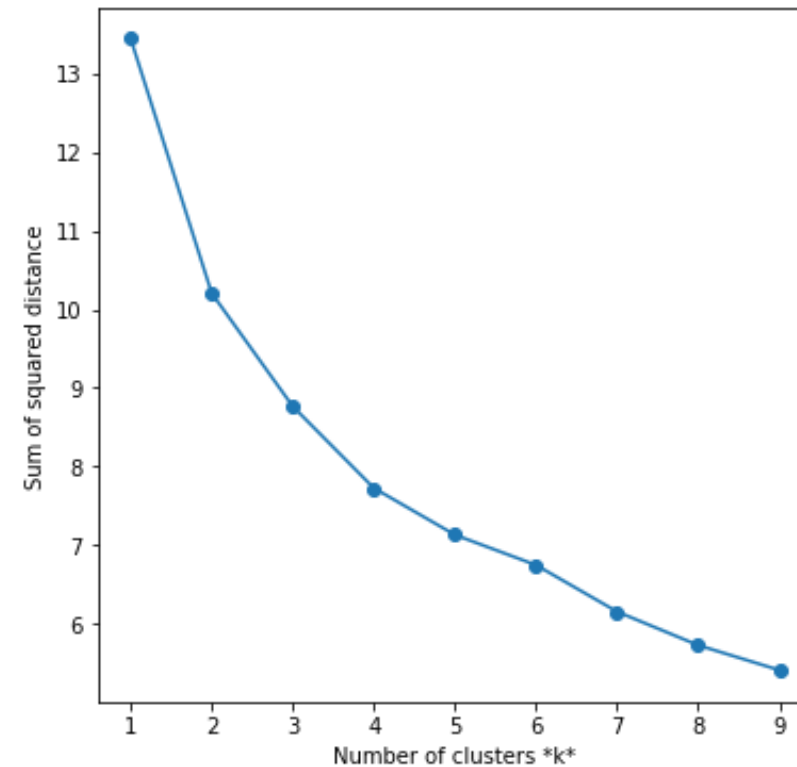


# 3.1. Clustering based on venues

Simplification of the clustering in order to avoid high inertia values (below values without simplification)



Before simplification

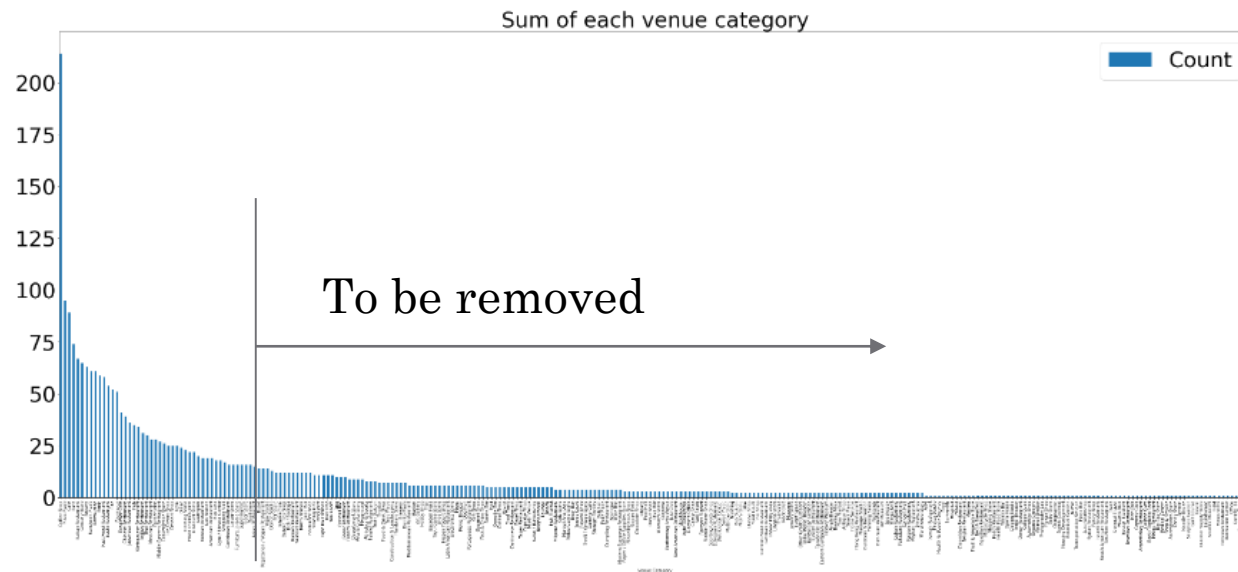


After simplification

# 3.1. Clustering based on venues

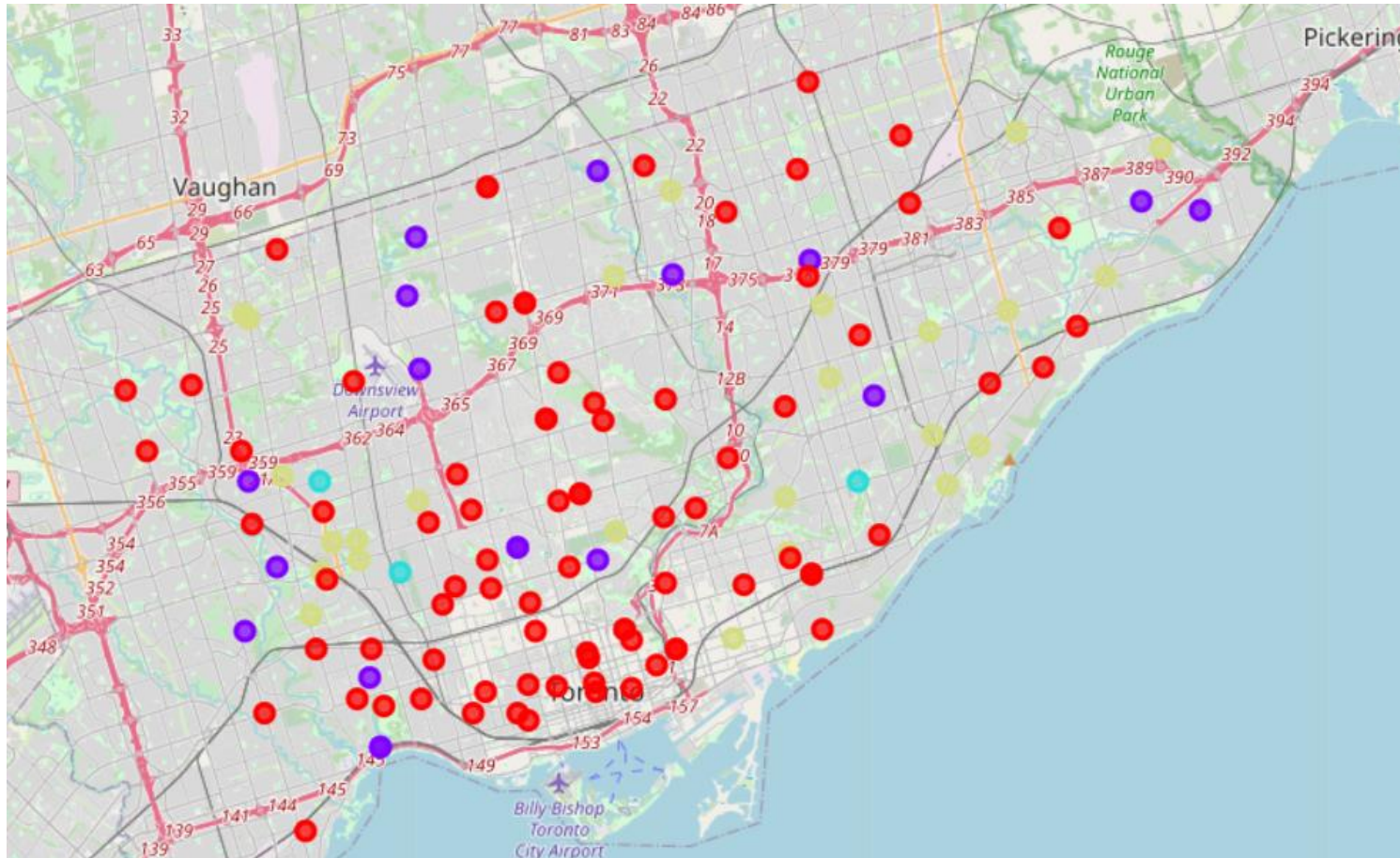
Simplification:

- Remove all the neighborhoods which are farther to the new job neighborhood than the origin neighborhood
- Remove all the venues with a very small frequency.



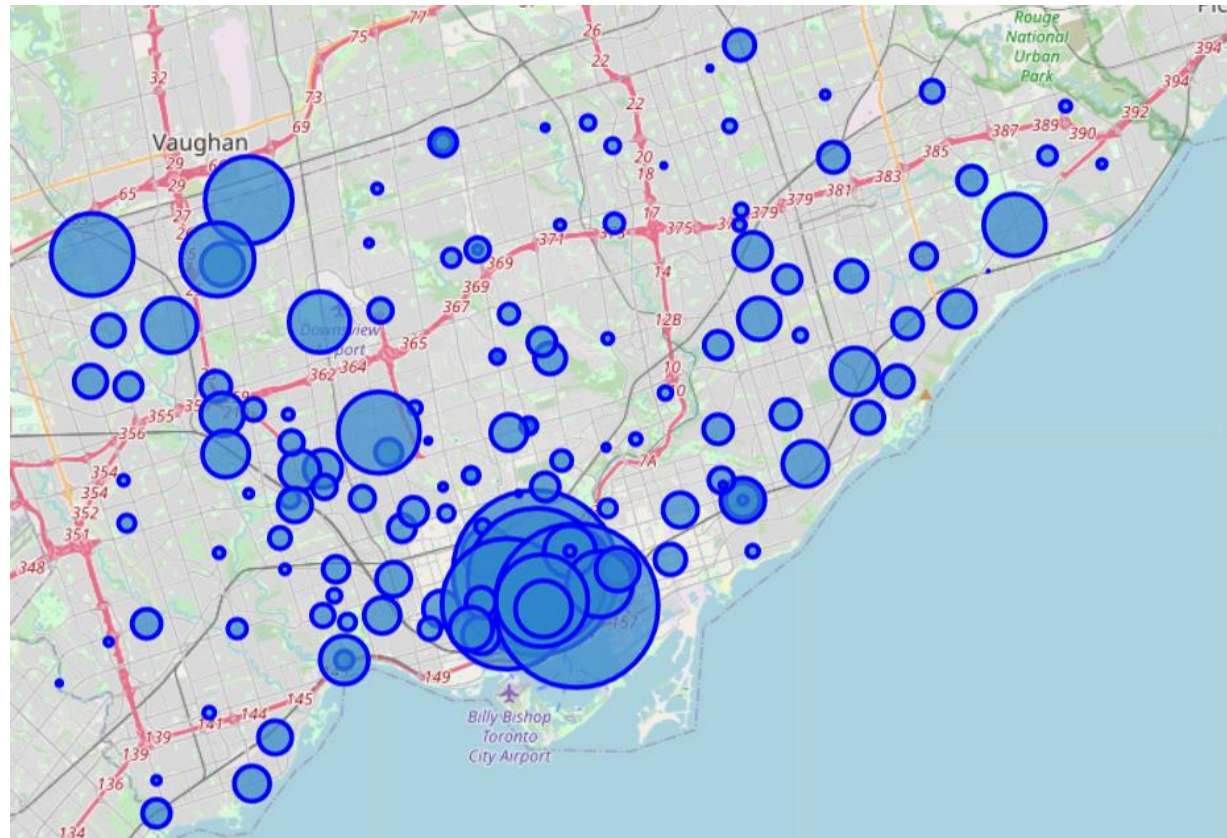
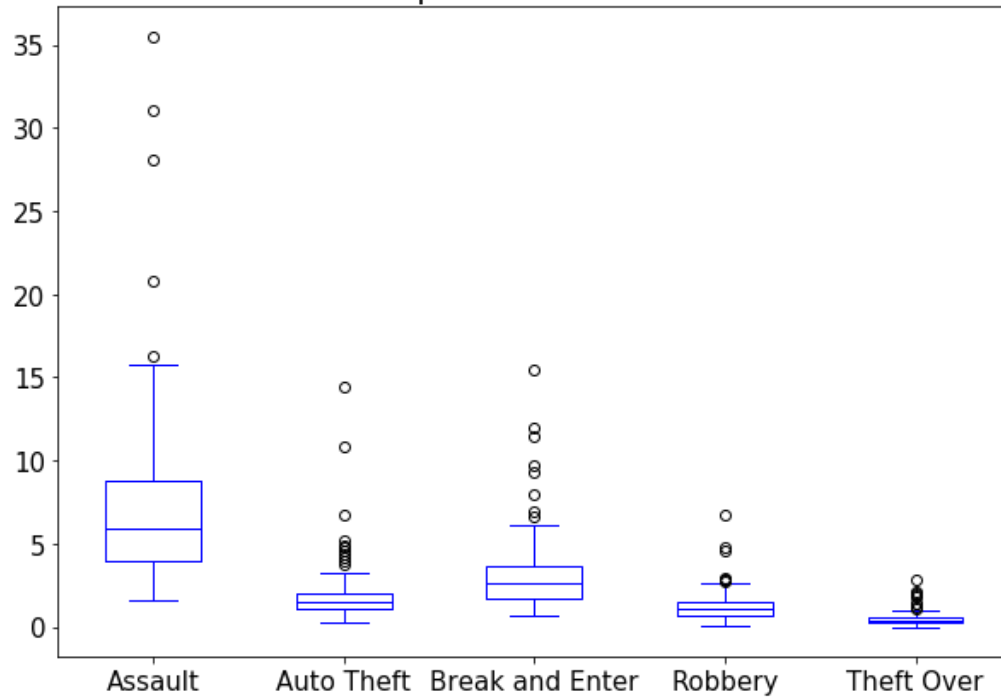
- Group similar categories to reduce to 15 categories

### 3.1. Clustering based on venues



## 3.2. Criminality per Neighborhood

Box Plot per each kind of crime



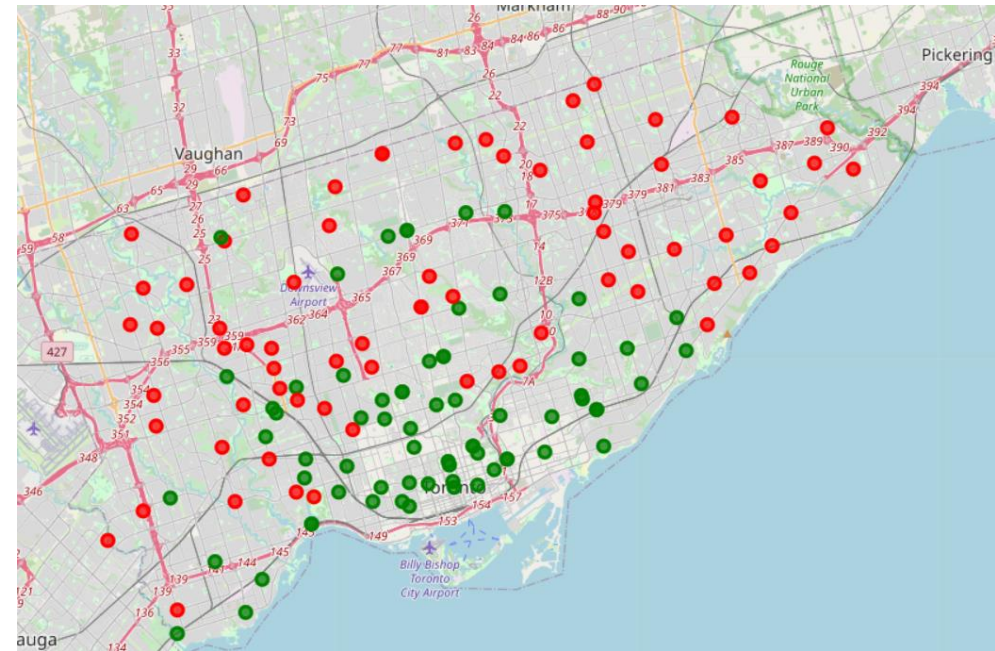


## 3.3. Profiles Neighborhood

- Percentage of people living alone

Neighbourhood	Liv. alone age less than 65 (percentage)
Church-Yonge Corridor	31.716656
Waterfront Communities-The Island	30.714730
Niagara	30.676716
Moss Park	29.308495
Bay Street Corridor	25.448696
Mount Pleasant West	25.355722
South Parkdale	22.037622
Cabbagetown-South St. James Town	20.353072
Annex	20.032104
North St. James Town	19.983884

Top 10

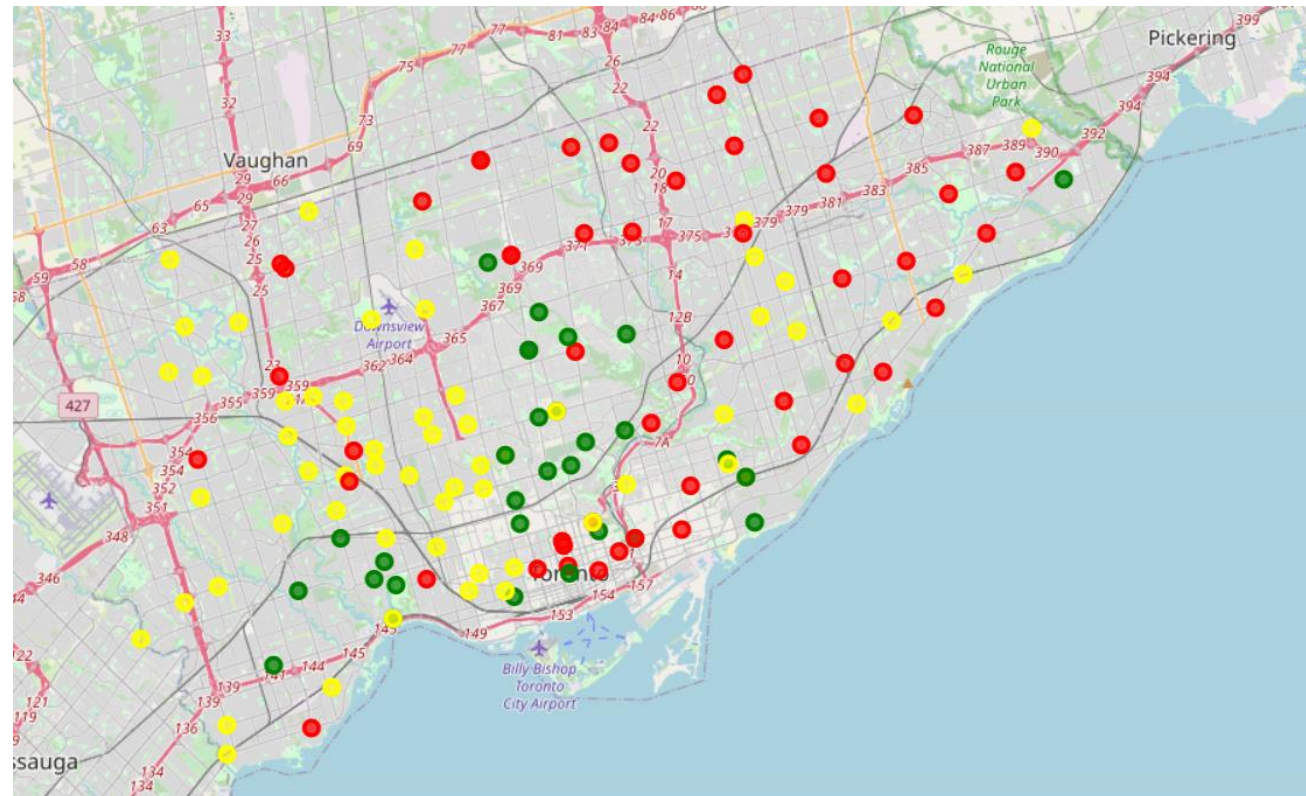


● Neighborhoods above the percentile 50%

## 3.3. Profiles Neighborhood

- Income of the population

- Above 50,000
- Between 20,000 and 49,999
- Below 19,999

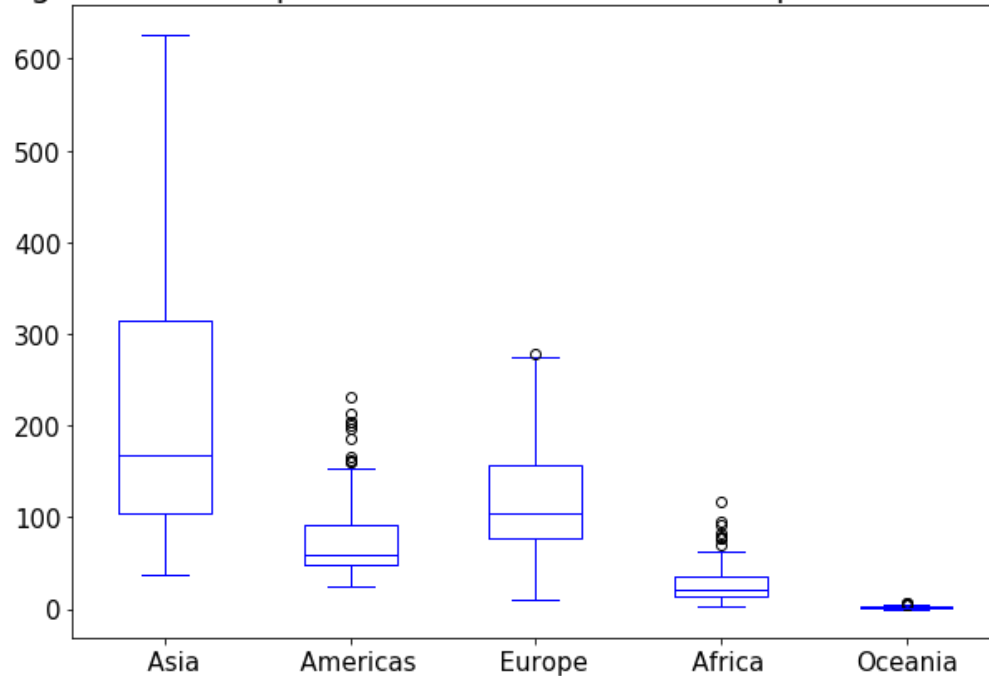




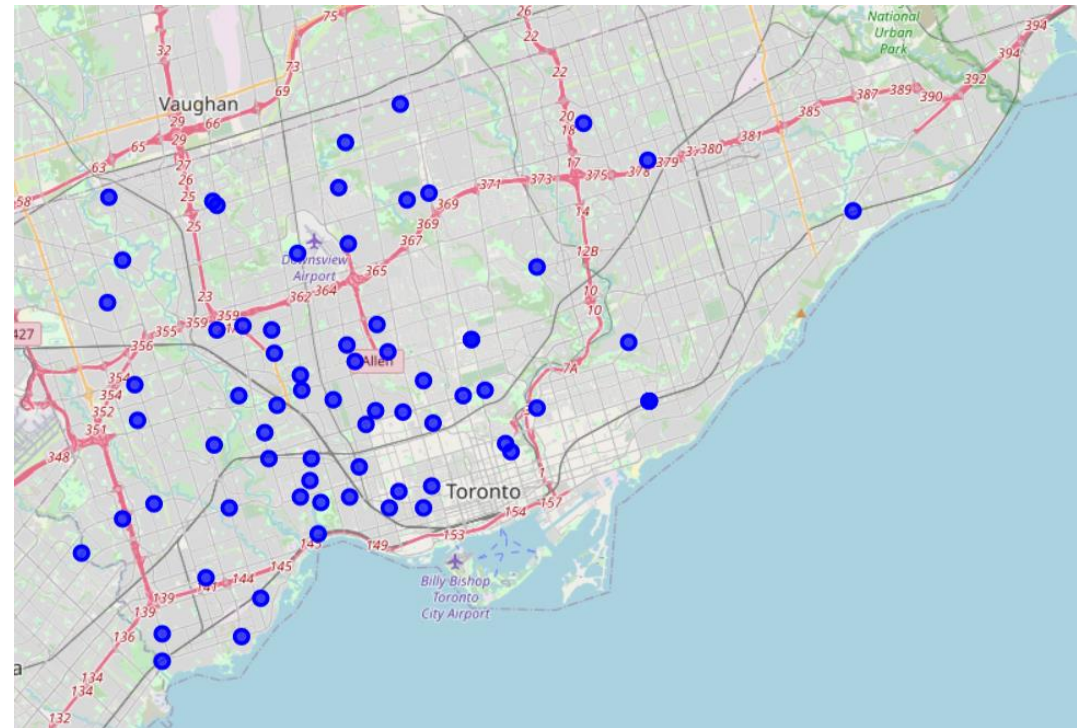
## 3.3. Profiles Neighborhood

- Immigration

Neighborhood Population from each Continent per 1000 habitants



Neighborhoods with European immigration above the percentile 50%



# 4. Results

Neighborhoods with best fit (accomplish all the criteria):

- Briar Hill-Belgravia
- Cabbagetown-South St. James Town
- Dovercourt-Wallace Emerson-Junction
- Dufferin Grove, Junction Area
- Mimico (includes Humber Bay Shores)
- Oakwood Village, Palmerston-Little Italy
- Playter Estates-Danforth
- Rosedale-Moore Park
- Trinity-Bellwoods



# 4. Results

- Category Venues

	Bakery	Bank	Gym	Ice Cream Shop	Italian Restaurant	Mexican Restaurant	Park	Pharmacy	Restaurant	Coffee	Food Store	Fast Food	Asiatic Restaurant
Neighborhood													
Briar Hill-Belgravia	0.142857	0.142857	0.000000	0.000000	0.000000	0.000000	0.142857	0.000000	0.000000	0.285714	0.000000	0.142857	0.000000
Cabbagetown-South St. James Town	0.061224	0.020408	0.000000	0.000000	0.040816	0.000000	0.020408	0.020408	0.061224	0.163265	0.040816	0.081633	0.081633
Dovercourt-Wallace Emerson-Junction	0.062500	0.000000	0.000000	0.000000	0.000000	0.000000	0.062500	0.062500	0.000000	0.375000	0.062500	0.000000	0.000000
Dufferin Grove	0.039216	0.000000	0.000000	0.000000	0.098039	0.039216	0.000000	0.000000	0.078431	0.196078	0.000000	0.019608	0.019608
Junction Area	0.037736	0.000000	0.018868	0.018868	0.075472	0.056604	0.000000	0.000000	0.000000	0.150943	0.056604	0.018868	0.094340
Mimico (includes Humber Bay Shores)	0.000000	0.142857	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.142857	0.428571	0.000000	0.000000
Oakwood Village	0.043478	0.000000	0.000000	0.000000	0.043478	0.086957	0.000000	0.000000	0.000000	0.086957	0.043478	0.130435	0.043478
Palmerston-Little Italy	0.014925	0.000000	0.000000	0.000000	0.059701	0.000000	0.000000	0.000000	0.014925	0.089552	0.000000	0.059701	0.044776
Playter Estates-Danforth	0.038462	0.038462	0.000000	0.000000	0.000000	0.038462	0.000000	0.076923	0.000000	0.153846	0.076923	0.038462	0.038462
Rosedale-Moore Park	0.000000	0.000000	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.500000	0.000000	0.000000
Trinity-Bellwoods	0.047619	0.000000	0.000000	0.023810	0.023810	0.000000	0.023810	0.000000	0.023810	0.071429	0.023810	0.023810	0.047619

# 4. Results

- Crimes and population

	Population	Crimes per 1000	Persons living alone (total)	Liv. alone age less than 65 (percentage)
Neighborhood				
Briar Hill-Belgravia	14257	12.765659	1725	8.066213
Cabbagetown-South St. James Town	11669	23.738110	3315	20.353072
Dovercourt-Wallace Emerson-Junction	36625	14.962457	4655	10.279863
Dufferin Grove	11785	15.528214	2020	13.831141
Junction Area	14366	12.529584	1985	11.067799
Mimico (includes Humber Bay Shores)	33964	14.662584	8205	18.578495
Oakwood Village	21210	13.342763	2650	7.190005
Palmerston-Little Italy	13826	13.742225	2230	12.765804
Playter Estates-Danforth	7804	17.811379	1455	13.582778
Rosedale-Moore Park	20923	13.191225	4170	11.614013
Trinity-Bellwoods	16556	17.516308	2160	10.660788

# 4. Results

- Income per ranges

	Below 19,999	Between 20,000 and 49,999	Over 50,000
Neighborhood			
Briar Hill-Belgravia	4495	5285	1990
Cabbagetown-South St. James Town	3175	3335	3470
Dovercourt-Wallace Emerson-Junction	11555	12305	6920
Dufferin Grove	3645	4155	2325
Junction Area	3760	4300	3505
Mimico (includes Humber Bay Shores)	8770	10445	9735
Oakwood Village	6545	6985	3560
Palmerston-Little Italy	4025	4750	3450
Playter Estates-Danforth	1910	2015	2370
Rosedale-Moore Park	4020	4425	8875
Trinity-Bellwoods	5185	5230	3740

# 4. Results

- Immigrants per each 1000 habitants in the Neighborhood

	Asia	Americas	Europe	Africa	Oceania
Neighborhood					
Briar Hill-Belgravia	234.972294	113.628393	233.218770	17.535246	0.000000
Cabbagetown-South St. James Town	126.831777	68.129231	110.549319	15.853972	1.285457
Dovercourt-Wallace Emerson-Junction	125.733788	69.624573	177.201365	17.883959	1.774744
Dufferin Grove	109.036911	73.398388	176.495545	13.576580	0.848536
Junction Area	114.158430	59.515523	129.124321	11.485452	2.088264
Mimico (includes Humber Bay Shores)	132.934872	63.743964	176.363208	20.610058	1.619362
Oakwood Village	139.321075	130.834512	172.324375	26.638378	0.471476
Palmerston-Little Italy	83.176624	48.821062	152.972660	10.125850	1.446550
Playter Estates-Danforth	63.429011	43.567401	107.637109	11.532547	3.844182
Rosedale-Moore Park	81.728242	54.963437	105.386417	17.922860	1.911772
Trinity-Bellwoods	145.566562	28.388500	167.310945	6.644117	0.906016