



IBM Data Science Professional Certificate

# Capstone Project **The Battle of the Neighborhoods**

Linbo Huang

# 1. Introduction: Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in **opening an seafood restaurant in Toronto, ON, Canada.**

- ◆ locations that are not already crowded with much more seafood restaurants.
- ◆ locations close to rich and safe neighborhood as possible
- ◆ locations close to areas with more population in the neighborhood

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

## 2.Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of existing seafood restaurants in the neighborhood
- number of criminal case in the neighborhood.
- average income of the neighborhood.
- population of the neighborhood.

Following data sources will be needed to extract/generate the required information:

- *To provide the stakeholders the necessary information we will combine Toronto's newest Census that contains Population, Average income per Neighborhood with Toronot's Neighborhoods shapefile and Foursquare API to collect competitors on the same neighborhoods. Toronto's Census data is publicly available at this website: <https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#8c732154-5012-9afe-d0cd-ba3ffc813d5a>*
- *Toronto Neighborhoods' shapefile is publicly available at this website: <https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#a45bd45a-ede8-730e-1abc-93105b2c439f>*
- *Also, we will get the Toronto Crime data from kaggle.com. we can download the data from this website: <https://www.kaggle.com/alincijov/toronto-crime-rate-per-neighbourhood>*

# Explore Dataset

140 Neighborhoods

Population, Income

	Neighborhood	Population	Income
0	Agincourt North	29113.0	30414.0
1	Agincourt South-Malvern West	23757.0	31825.0
2	Alderwood	12054.0	47709.0
3	Annex	30526.0	112766.0
4	Banbury-Don Mills	27695.0	67757.0



Crime rate

	Neighborhood	Hood_ID	Crime_total
0	Yonge-St.Clair	97	81
1	York University Heights	27	729
2	Lansing-Westgate	38	165
3	Yorkdale-Glen Park	31	426
4	Stonegate-Queensway	16	206



Population, Income, Crime

	Neighborhood	Population	Income	Hood_ID	Crime_total
0	Agincourt North	29113.0	30414.0	129.0	214.0
1	Agincourt South-Malvern West	23757.0	31825.0	128.0	329.0
2	Alderwood	12054.0	47709.0	20.0	88.0
3	Annex	30526.0	112766.0	95.0	604.0
4	Banbury-Don Mills	27695.0	67757.0	42.0	221.0

# Explore Dataset

140 Neighborhoods

Population, Income, Crime

	Neighborhood	Population	Income	Hood_ID	Crime_total
0	Agincourt North	29113.0	30414.0	129.0	214.0
1	Agincourt South-Malvern West	23757.0	31825.0	128.0	329.0
2	Alderwood	12054.0	47709.0	20.0	88.0
3	Annex	30526.0	112766.0	95.0	604.0
4	Banbury-Don Mills	27695.0	67757.0	42.0	221.0

Get location data

dropna()



Population, Income, Crime, Location[]

	Neighborhood	Population	Income	Hood_ID	Crime_total	Address	Coordinates	Latitude	Longitude
0	Agincourt North	29113.0	30414.0	129.0	214.0	Agincourt North, Toronto	(Agincourt North, Scarborough North, Scarborough)	43.808038	-79.266439
1	Agincourt South-Malvern West	23757.0	31825.0	128.0	329.0	Agincourt South-Malvern West, Toronto	(Toronto Fire Station 243, Sheppard Avenue East)	43.789232	-79.263037
2	Alderwood	12054.0	47709.0	20.0	88.0	Alderwood, Toronto	(Alderwood, Etobicoke — Lakeshore, Etobicoke, Toronto)	43.601717	-79.545232
3	Annex	30526.0	112766.0	95.0	604.0	Annex, Toronto	(The Annex, University — Rosedale, Toronto, Goldring)	43.670338	-79.407117
4	Banbury-Don Mills	27695.0	67757.0	42.0	221.0	Banbury-Don Mills, Toronto	(Banbury Road, Don Valley West, North York, Toronto)	43.734804	-79.357243

110 Neighborhoods

# Explore Dataset

## Foursquare Venues Data

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Agincourt North	43.808038	-79.266439	Menchie's	43.808338	-79.268288	Frozen Yogurt Shop
1	Agincourt North	43.808038	-79.266439	Fahmee Bakery & Jamaican Foods	43.810170	-79.280113	Caribbean Restaurant
2	Agincourt North	43.808038	-79.266439	Saravanaa Bhavan South Indian Restaurant	43.810117	-79.269275	Indian Restaurant
3	Agincourt North	43.808038	-79.266439	Shoppers Drug Mart	43.808894	-79.269854	Pharmacy
4	Agincourt North	43.808038	-79.266439	Samosa King - Embassy Restaurant	43.810152	-79.257316	Indian Restaurant

Restaurant Venues

Total Venues  
(8899, 7)

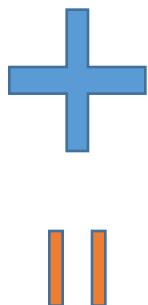


Restaurant Venues  
(2312, 7)

# Explore Dataset

## Foursquare Venues Data

	Seafood Restaurant	Restaurants_sum
Neighborhood		
Agincourt North	0	22
Agincourt South-Malvern West	1	37
Alderwood	3	23
Annex	0	30
Banbury-Don Mills	0	12



## Population, Income, Crime, Location[]

	Neighborhood	Population	Income	Hood_ID	Crime_total	Address	Coordinates	Latitude	Longitude
0	Agincourt North	29113.0	30414.0	129.0	214.0	Agincourt North, Toronto	(Agincourt North, Scarborough North, Scarborough...)	43.808038	-79.266439
1	Agincourt South-Malvern West	23757.0	31825.0	128.0	329.0	Agincourt South-Malvern West, Toronto	(Toronto Fire Station 243, Sheppard Avenue East...)	43.789232	-79.263037
2	Alderwood	12054.0	47709.0	20.0	88.0	Alderwood, Toronto	(Alderwood, Etobicoke —Lakeshore, Etobicoke, To...)	43.601717	-79.545232
3	Annex	30526.0	112766.0	95.0	604.0	Annex, Toronto	(The Annex, University —Rosedale, Toronto, Gold...)	43.670338	-79.407117
4	Banbury-Don Mills	27695.0	67757.0	42.0	221.0	Banbury-Don Mills, Toronto	(Banbury Road, Don Valley West, North York, To...)	43.734804	-79.357243

## Population, Income, Crime, Seafood Restaurant, Restaurant, Location[]

	Neighborhood	Population	Income	Hood_ID	Crime_total	Address	Coordinates	Latitude	Longitude	Seafood Restaurant	Restaurants_sum
0	Agincourt North	29113.0	30414.0	129.0	214.0	Agincourt North, Toronto	(Agincourt North, Scarborough North, Scarborough...)	43.808038	-79.266439	0	22
1	Agincourt South-Malvern West	23757.0	31825.0	128.0	329.0	Agincourt South-Malvern West, Toronto	(Toronto Fire Station 243, Sheppard Avenue East...)	43.789232	-79.263037	1	37
2	Alderwood	12054.0	47709.0	20.0	88.0	Alderwood, Toronto	(Alderwood, Etobicoke —Lakeshore, Etobicoke, To...)	43.601717	-79.545232	3	23
3	Annex	30526.0	112766.0	95.0	604.0	Annex, Toronto	(The Annex, University —Rosedale, Toronto, Gold...)	43.670338	-79.407117	0	30
4	Banbury-Don Mills	27695.0	67757.0	42.0	221.0	Banbury-Don Mills, Toronto	(Banbury Road, Don Valley West, North York, To...)	43.734804	-79.357243	0	12

# Explore Dataset

Visualize the data

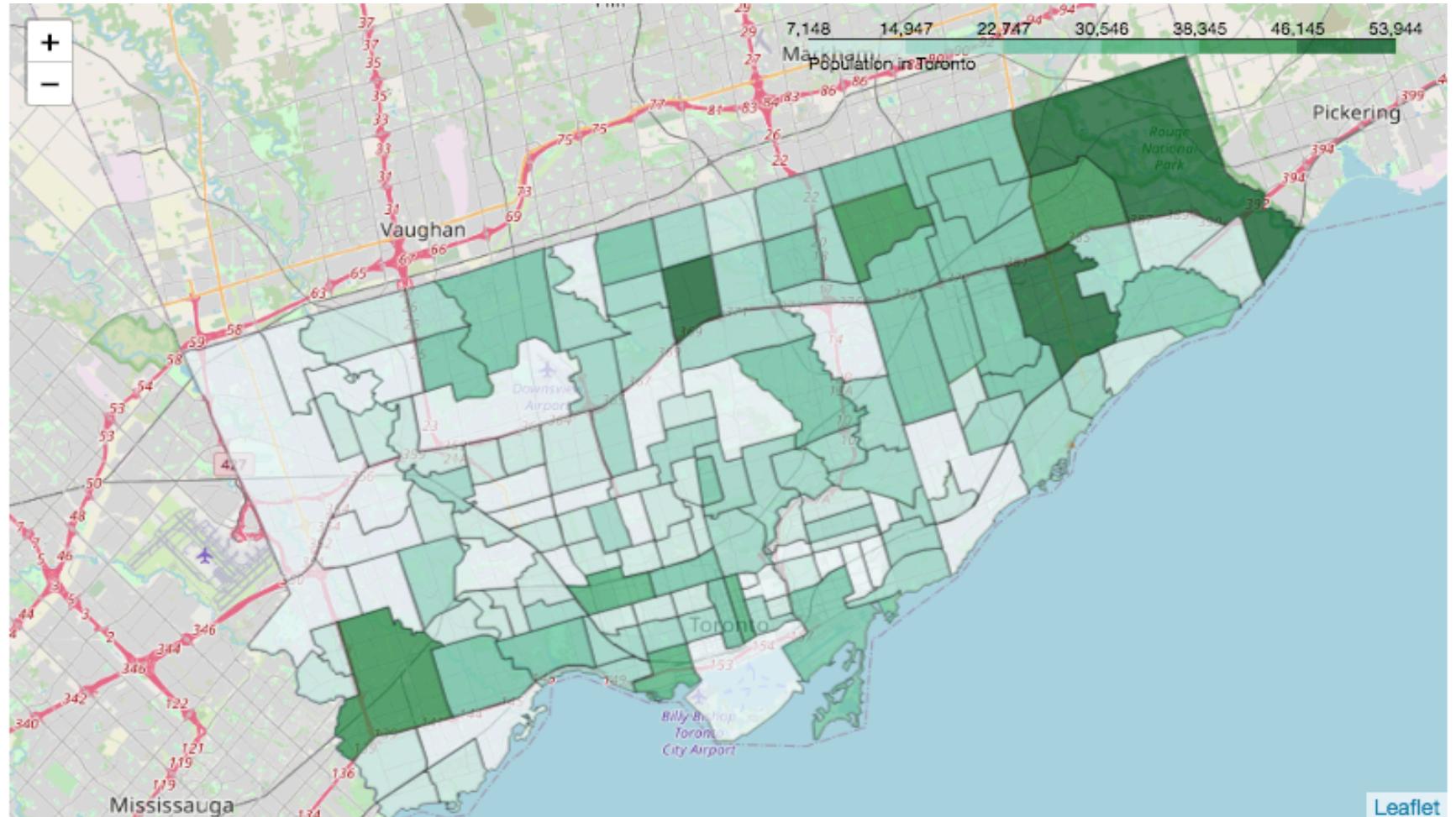


Figure - Population Map in Toronto

# Explore Dataset

Visualize the data

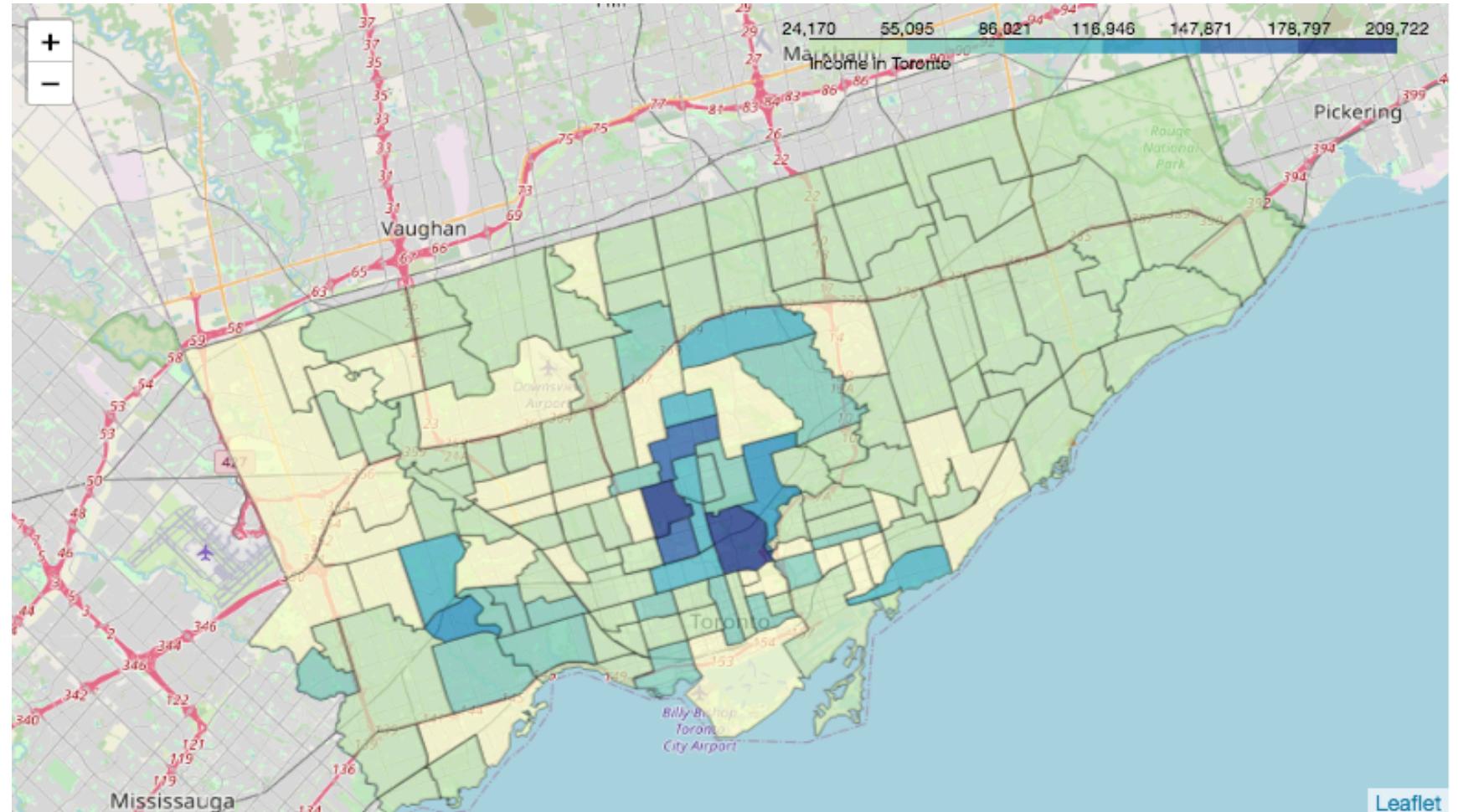


Figure - Income Map in Toronto

# Explore Dataset

Visualize the data

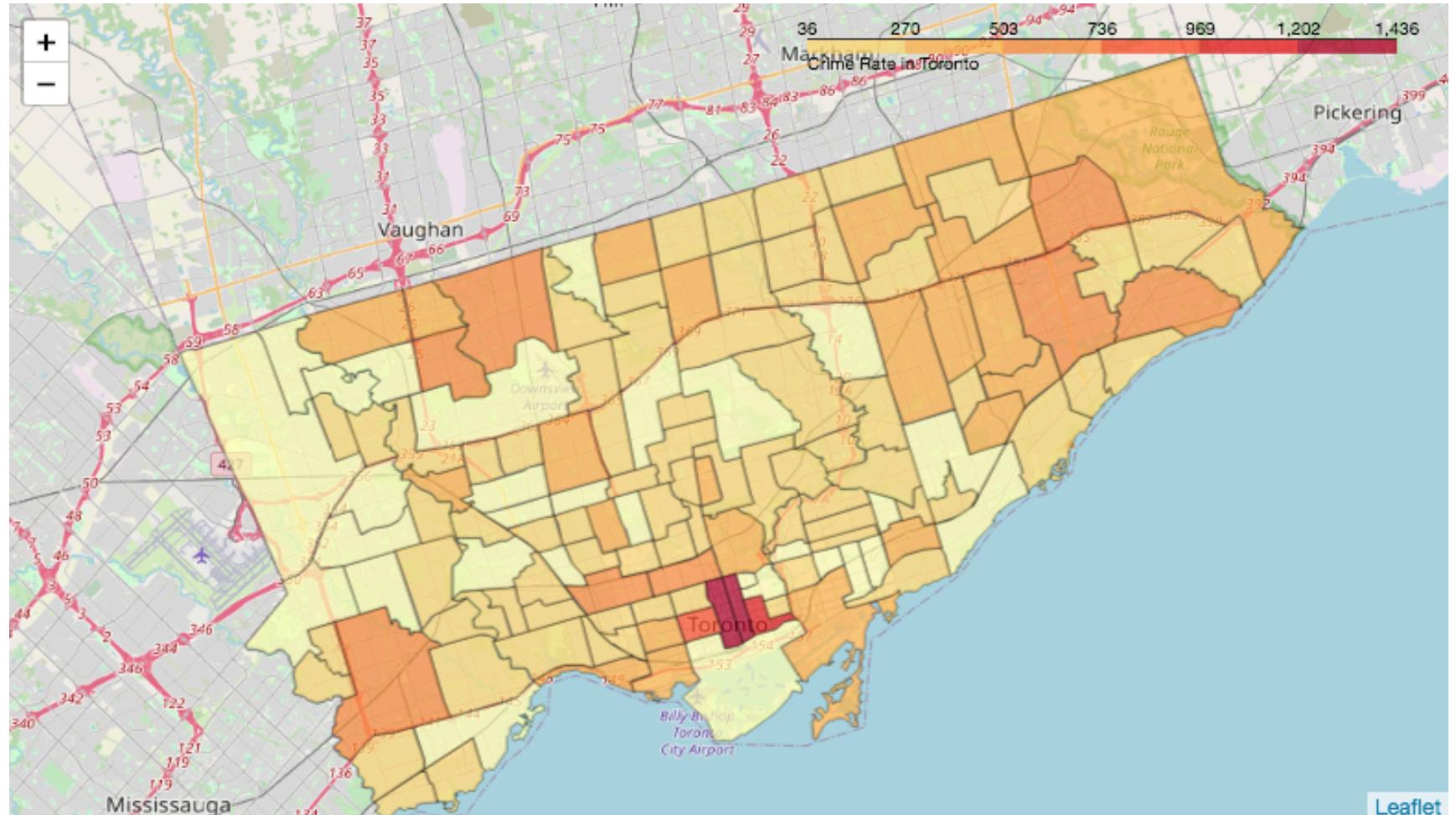


Figure - Crime Map in Toronto

# Explore Dataset

Visualize the data

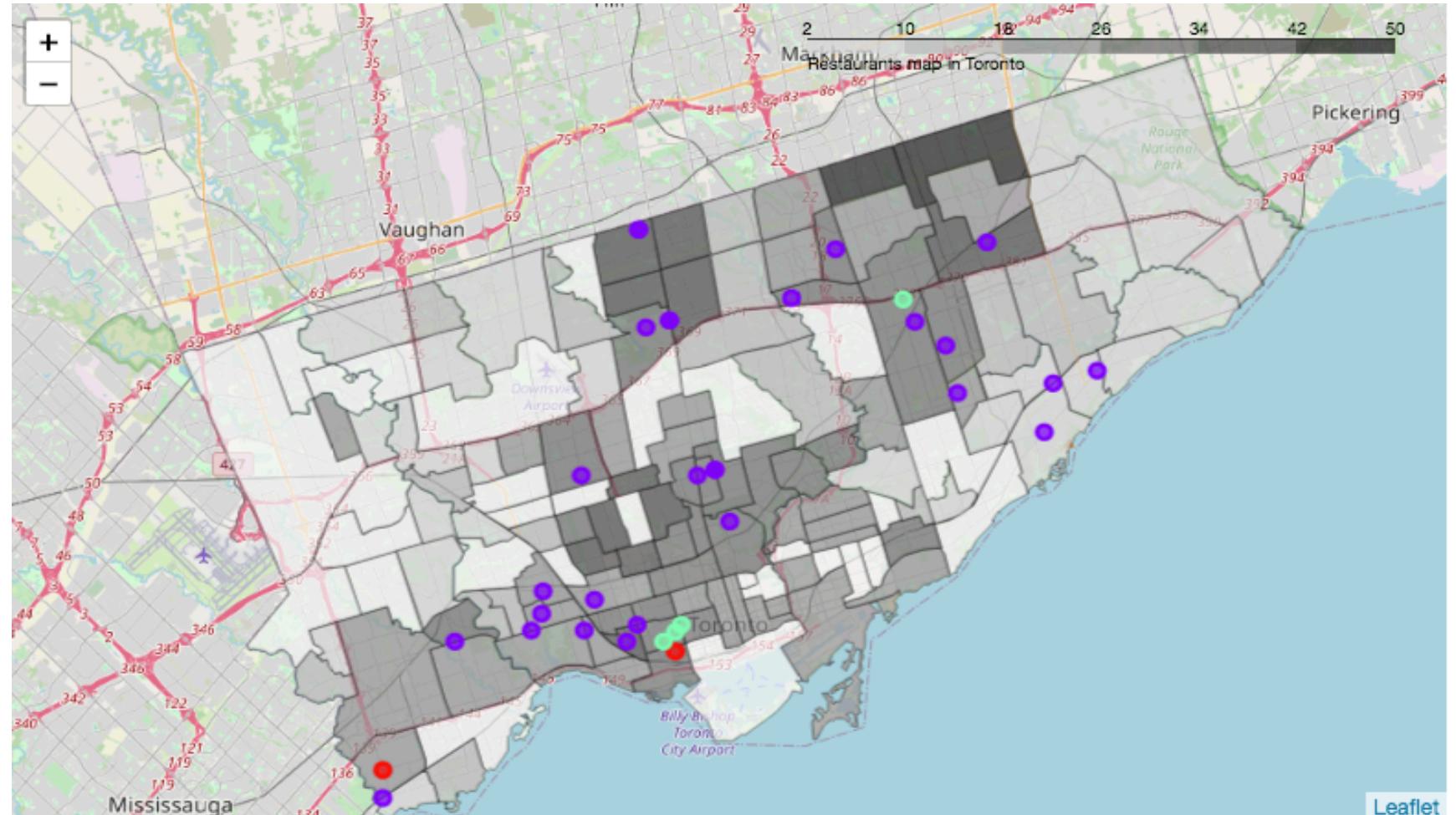


Figure – Restaurants and Seafood Restaurants Map in Toronto

## 3.Methodology

In this project we will direct our efforts on detecting areas of Toronto that have low restaurant density, particularly those with low number of Seafood restaurants.

**In first step** we have collected the required data: location and type (category) of every restaurant within 500m from each neighborhood center. We have also identified seafood restaurants (according to Foursquare categorization).

**Second step** in our analysis will be calculation and exploration of 'restaurant density' across different areas of Toronto - we will use choropleth maps to identify a few promising areas with low number of restaurants in general (and no seafood restaurants in vicinity) .

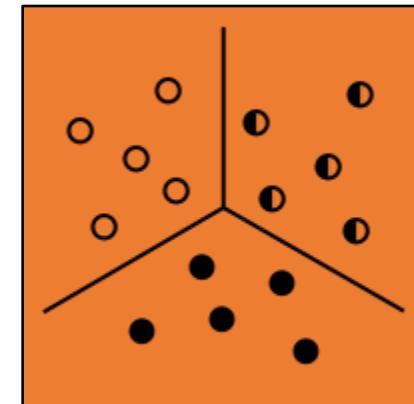
**In third and final step** we will use the data to create clusters of locations , and try to find out a cluster that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with low number of restaurants in the neighborhood, and we want locations without seafood restaurant in radius of 500 meters, we also want locations at lower crime rate and high income neighborhood. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

# Analysis Data

Population, Income, Crime, Restaurant, Seafood Restaurant,

	Population	Income	Crime_total	Restaurants_sum	Seafood Restaurant
0	29113.0	30414.0	214.0	22	0
1	23757.0	31825.0	329.0	37	1
2	12054.0	47709.0	88.0	23	3
3	30526.0	112766.0	604.0	30	0
4	27695.0	67757.0	221.0	12	0

K-Means Clustering



Cluster Labels, Population, Income, Crime, Restaurant, Seafood Restaurant,

Cluster Labels	Neighborhood	Population	Income	Crime_total	Restaurants_sum	Seafood Restaurant	Hood_ID	Latitude	Longitude
0	Agincourt North	29113.0	30414.0	214.0	22	0	129.0	43.808038	-79.266439
1	Agincourt South-Malvern West	23757.0	31825.0	329.0	37	1	128.0	43.789232	-79.263037
2	Alderwood	12054.0	47709.0	88.0	23	3	20.0	43.601717	-79.545232
3	Annex	30526.0	112766.0	604.0	30	0	95.0	43.670338	-79.407117
4	Banbury-Don Mills	27695.0	67757.0	221.0	12	0	42.0	43.734804	-79.357243

# 4. Analysis

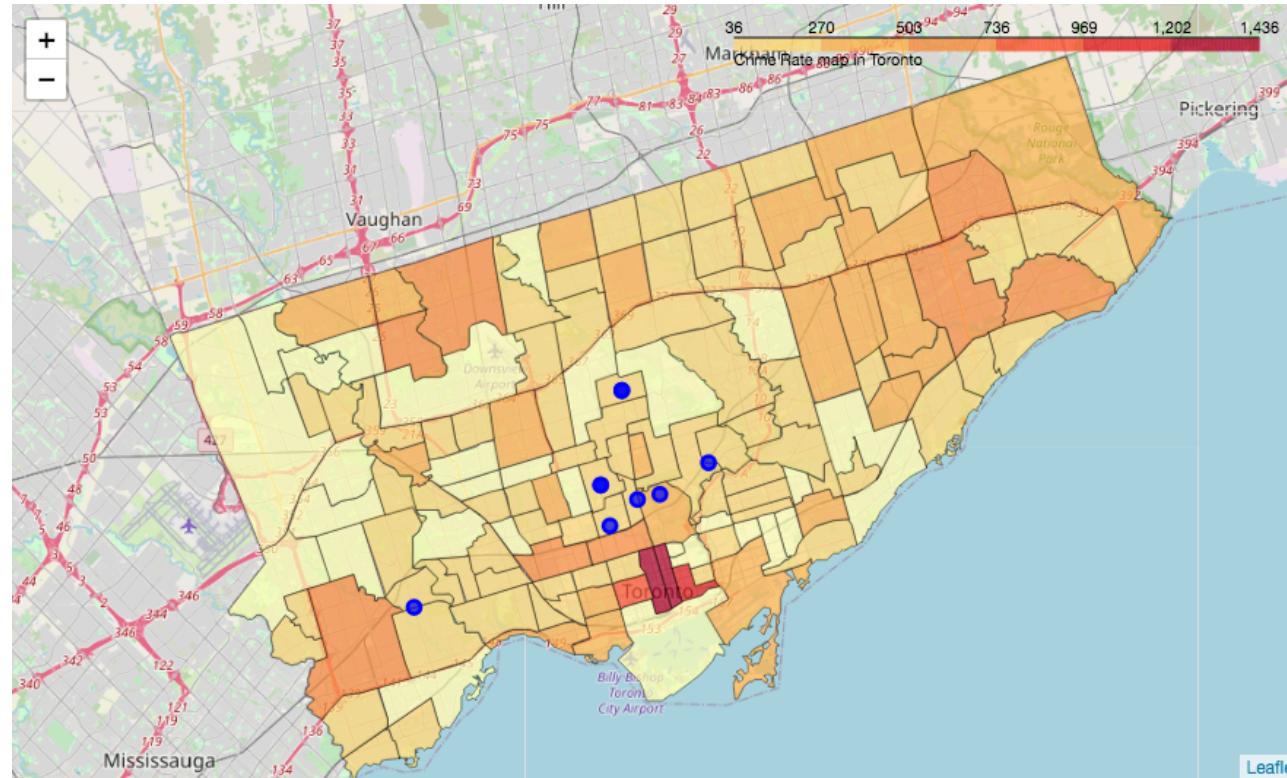
Cluster Labels	Population	Income	Crime_total	Restaurants_sum	Seafood Restaurant	Hood_ID	Latitude	Longitude
mean	0.0	18889.366667	46953.333333	251.066667	32.300000	1.100000	84.033333	43.715292 -79.392572
mean	1.0	15637.927273	46668.236364	185.490909	14.345455	0.090909	65.181818	43.238726 -79.348268
mean	2.0	35642.062500	45612.500000	670.187500	17.625000	0.125000	86.187500	43.725117 -79.356746
mean	3.0	13760.222222	147542.555556	124.333333	30.222222	0.222222	85.888889	43.694625 -79.410104

The clustering results show that the dataset is divided into four categories:

- ① **The first group:** large population, low income, high crime rate and the largest number of restaurants (including seafood restaurants).
- ② **The second group:** small population, low income, a moderate crime rate, and the fewest restaurants (including seafood restaurants).
- ③ **The third group:** the largest population, the lowest income, the highest crime rate, and a small number of restaurants (including seafood restaurants).
- ④ **The fourth group:** a small population, the highest income, the lowest crime rate, and a large number of restaurants (including seafood restaurants).

# 4. Analysis

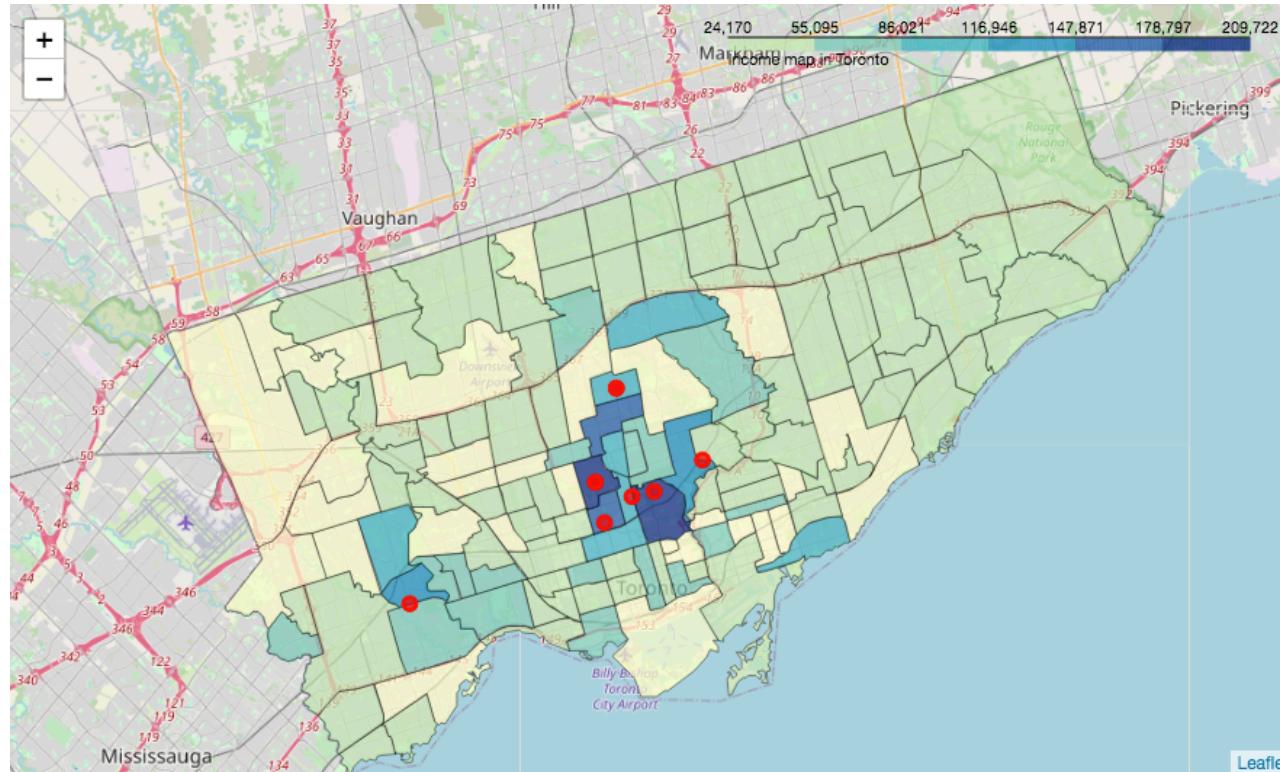
**choose the fourth group which is a group of the highest income and the safest neighborhoods**  
look at these neighborhood on the different choropleth map



(1) Selected Neighborhoods on the Crime Rate Map

# 4. Analysis

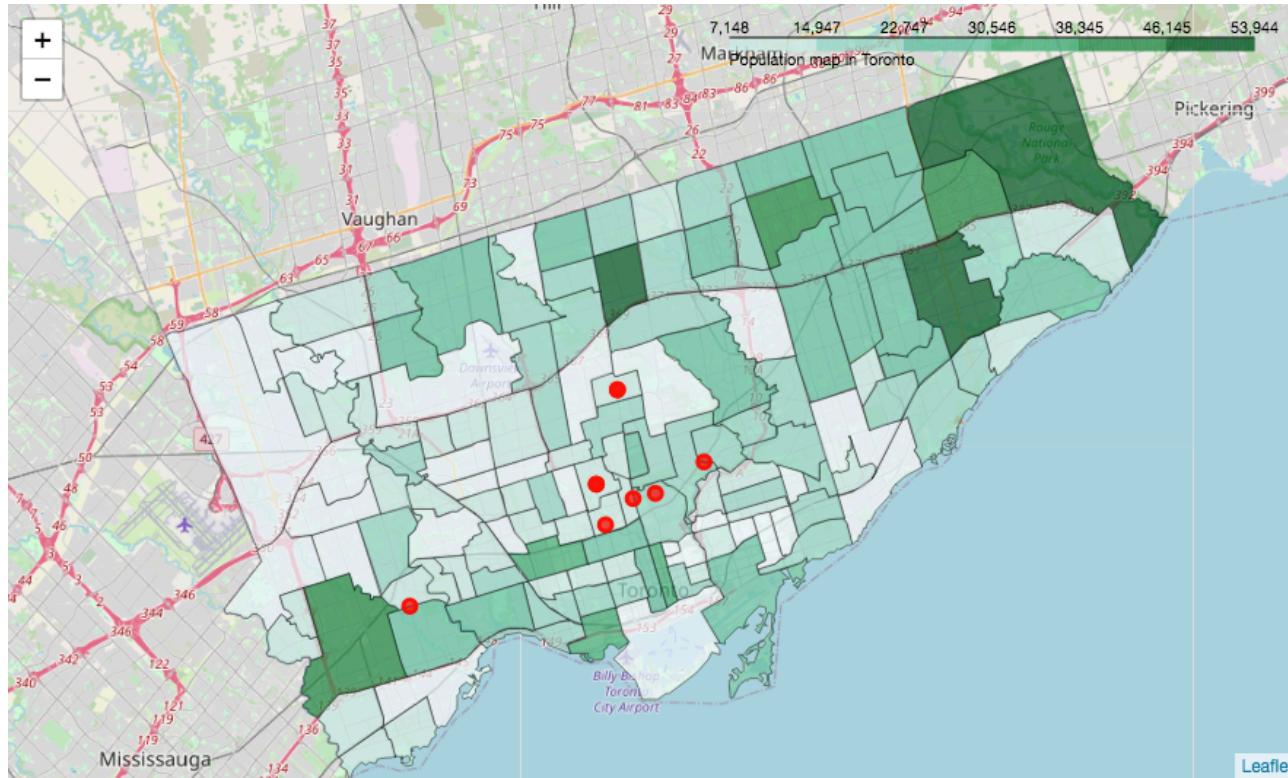
**choose the fourth group which is a group of the highest income and the safest neighborhoods**  
look at these neighborhood on the different choropleth map



**(2) Selected Neighborhoods on the Income Map**

# 4. Analysis

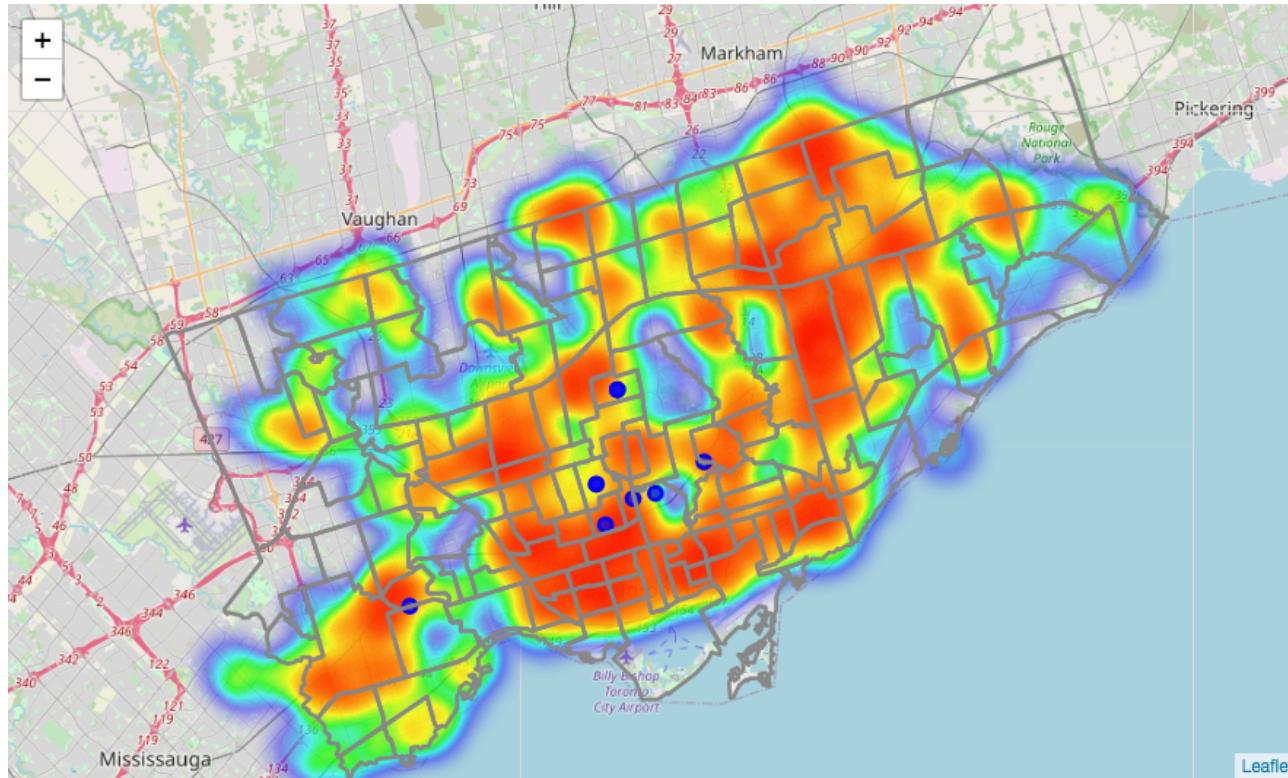
**choose the fourth group which is a group of the highest income and the safest neighborhoods**  
look at these neighborhood on the different choropleth map



**(3) Selected Neighborhoods on the Population Map**

# 4. Analysis

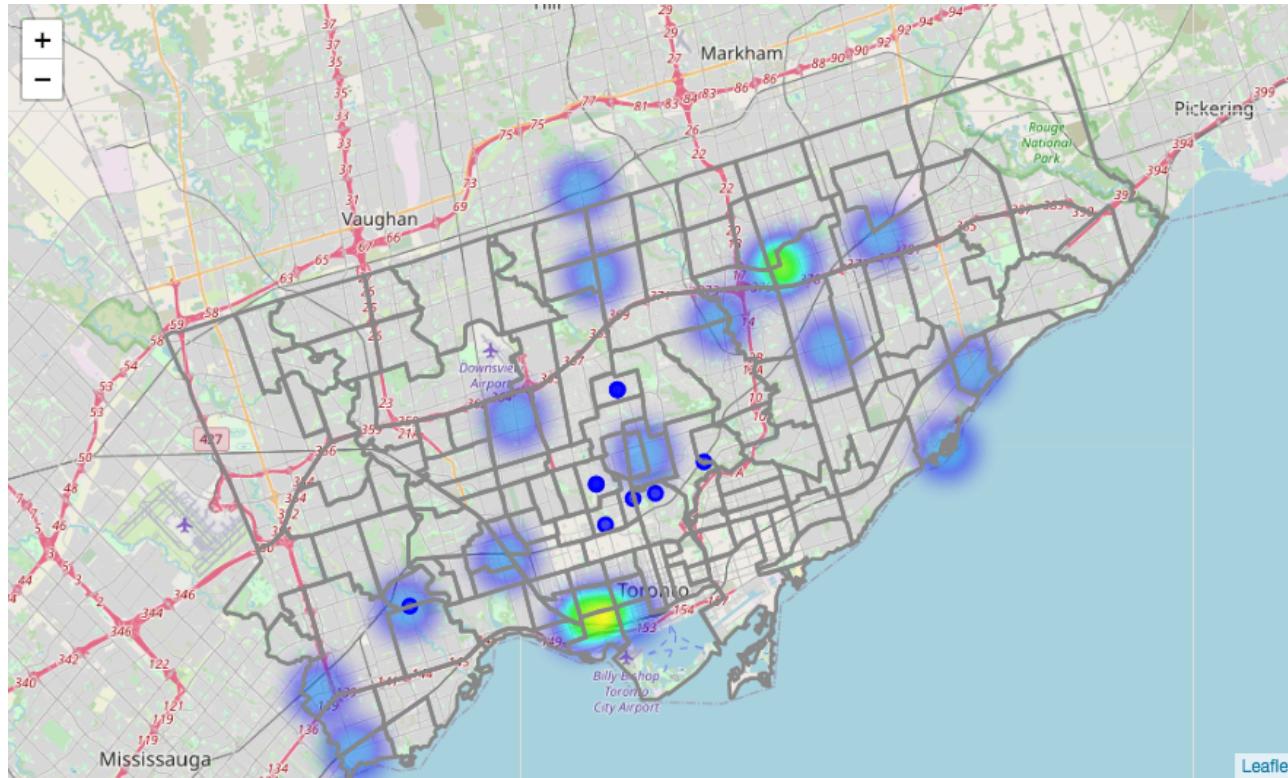
**choose the fourth group which is a group of the highest income and the safest neighborhoods**  
look at these neighborhood on the different choropleth map



(4) Selected Neighborhoods on the Restaurants Map (By Heatmap)

# 4. Analysis

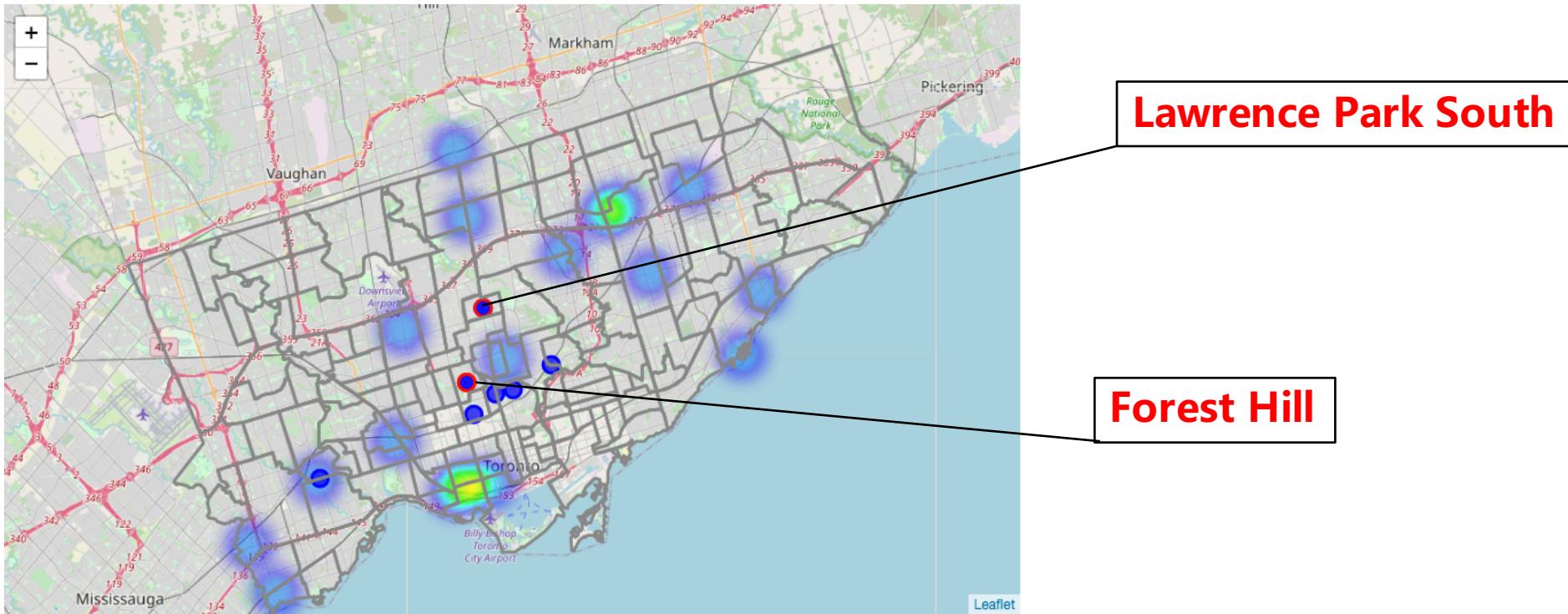
**choose the fourth group which is a group of the highest income and the safest neighborhoods**  
look at these neighborhood on the different choropleth map



(5) Selected Neighborhoods on the Sedfood Restaurants Map (By Heatmap)

# 5. Results

According to several choropleth maps, we will eventually choose a location in **Lawrence Park South** or **Forest Hill** to open a seafood restaurant. There are currently no seafood restaurant in these two neighborhoods, also there are just relatively few restaurants and low crime rate, higher income residents and a moderate population.



# 6. Conclusion

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.

