

# Capstone project - The Battle of Neighborhoods

Report

By Danijela Zivic

# 1. Introduction/Business Problem

In this project we will analyze existing venues in neighborhoods between Toronto and New York City. As a European tourist visiting Toronto in Canada or New York City in USA which are both megacities for tourist who lives in a city of 10.000 inhabitants. It would be helpful to decide which of those two big cities is the best city in North America to visit first.

With this project the traveler will explore the similarities and dissimilarities between neighborhoods in those two cities, and determine which neighborhoods are best fit for traveler who likes good food for reasonable price. With implementing recommender system to suggest and locating the best places to visit by finding and filtering different venues, especially restaurants based on the number of likes, prices and rating.

## 2. Data

To execute traveler idea we will be retrieving data from:

- Toronto data from Wikipedia page - [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Toronto Latitude and Longitude - [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data) - a csv file that has the geographical coordinates of each postal code
- New York City data, Latitude and Longitude – the link [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572), but better is from file placed on the server [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

First we will load, clean and process the Toronto data from Wikipedia page into a pandas dataframe and with using the Geocoder package and csv file get latitude and longitude of Toronto. Then New York dataset file, already placed on the server, we will simply run a wget command and access the data, then load and transformed into pandas dataframe.

With utilizing the API Foursquare data get the latitude and the longitude coordinates and details about venues, especially restaurants one of each neighborhood in Toronto and New York City separately, based on the largest number of restaurants per neighborhood. With that we will know which two neighborhood we need to focus on.

Then with Foursquare and various python packages we will explore, cluster, create maps and use machine learning models to get a list of recommended restaurants that have the most highest numbers of likes, ratings and for reasonable prices. The bonus that will be gained through the survey is that we can also provide to the traveler a list of nearby attractions.

Finally, the ideal city to visit first can be inferred based on comparison between one of each neighborhood in Toronto and New York City which restaurant have the most likes.

### 3. Methodology

First Toronto data from Wikipedia page was extracted using the BeautifulSoup library to build a pandas dataframe listing the buroughs and neighborhoods, then was cleaned and processed appropriately. The geographical coordinates of each neighborhoods was obtained from the csv file with using the Geocoder package to get latitude and longitude of Toronto. The postal codes for each of the neighborhoods were matched to the information from a geographic csv file and dataframe was updated to the corresponding geographical coordinate for each neighborhood. We also created a Folium map of Toronto with neighborhoods superimposed on top.

Then New York City dataset file, already placed on the server, we had simply run a wget command and with that we had access to the data, then load with JSON library to handle JSON files and transformed into pandas dataframe corresponding geographical coordinate for each neighborhood. We also created a Folium map of New York City with neighborhoods superimposed on top.

We used explore calls from Foursquare Places API to get the latitude and longitude coordinates and details about all of top 100 venues within a radius of 1000 meters in neighborhoods of Toronto and New York City, finishing with printing the size of resulting dataframe for each neighborhoods in Toronto and New York City separately.

Because the resulting dataframe is too big, we focused just on one neighborhood of each cities based on the largest number of restaurants per neighborhood. The Foursquare API was used to query each of the neighborhood for the keyword {Restaurants}. Those are Davisville, Toronto and Elmhurst, New York City. We visualised using Folium a maps of these two neighborhoods. We also explored more about, used machine learning models, Foursquare API calls, various python packages and within various functions to get:

- top 100 venues within a radius of 1000 meters for Davisville and Elmhurst;
- analyze neighborhood and grouping by venues category to get top 10 most common venues;
- for more information about restaurants of this two neighborhoods, we searched, printed and saved separately all of this restaurants in new dataframe;
- pulled some more informations about likes and saved data as top 10 likes;
- based on top 10 restaurants by likes we investigated the rating and prices by those restaurants.

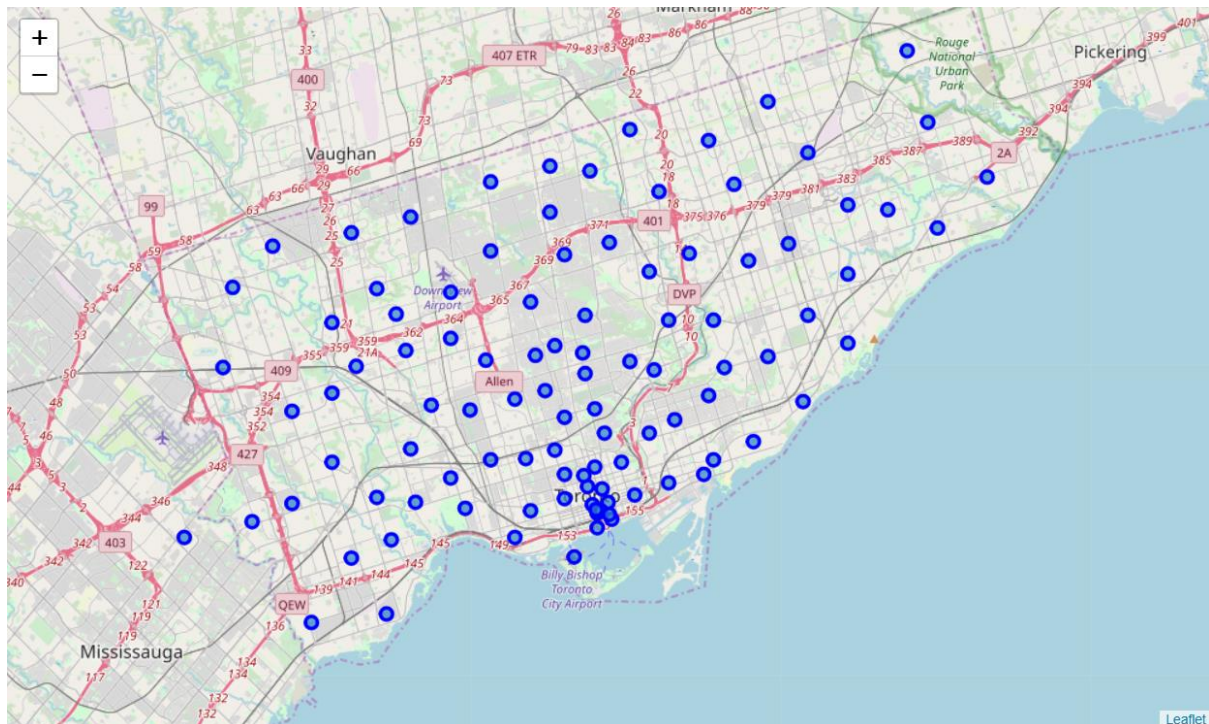
The results of all this, was to get a list of recommended restaurants that have the most highest numbers of likes and create a Folium maps that shows where are those top 10 liked restaurants for Toronto and New York City. The bonus that we have gained through the survey is that we can also provided to the traveler a list of nearby attractions and a map of attractions combined with 10 popular restaurants.

At the end we compared between Davisville, Toronto and Elmhurst, New York City based on top 10 liked restaurants by plotting in bar charts.

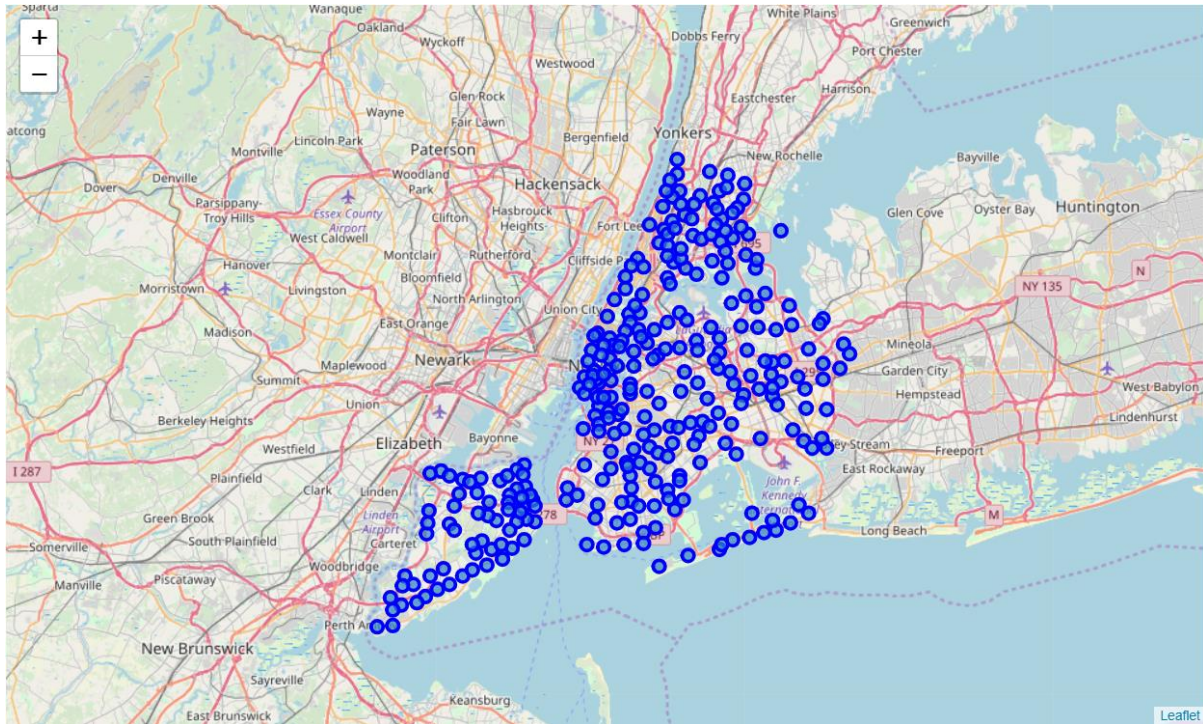
## 4. Results

When we arranged the data, we plotted on a Folium map the geographical coordinates of each neighborhood in Toronto and New York City separately.

*Toronto map of neighborhoods superimposed with the circle marker on top*



*New York City map of neighborhoods superimposed with the circle marker on top*



By exploring top 100 venues in Toronto and New York City neighborhoods with Foursquare API, we have found:

4861 venues and 329 unique categories in 103 neighborhoods of Toronto

20487 venues and 464 unique categories in 306 neighborhoods of New York City

Because we found so many venues we focused just on one neighborhood of each city based on the largest number of restaurants per neighborhood. The Foursquare API was used to query each of the neighborhood for the keyword {Restaurants}. We found that total number of all restaurants in Toronto is 1213 and 4925 in New York City.

Sorting it by decreasing resulted that with 40 restaurants is Davisville neighborhood in Toronto have the most largest number of restaurants per neighborhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Number of restaurants
47	Davisville	43.704324	-79.388790	40
22	Willowdale South	43.770120	-79.408493	39
75	Christie	43.669542	-79.422564	34
41	The Danforth West, Riverdale	43.679557	-79.352188	33
52	Church and Wellesley	43.665860	-79.383160	32

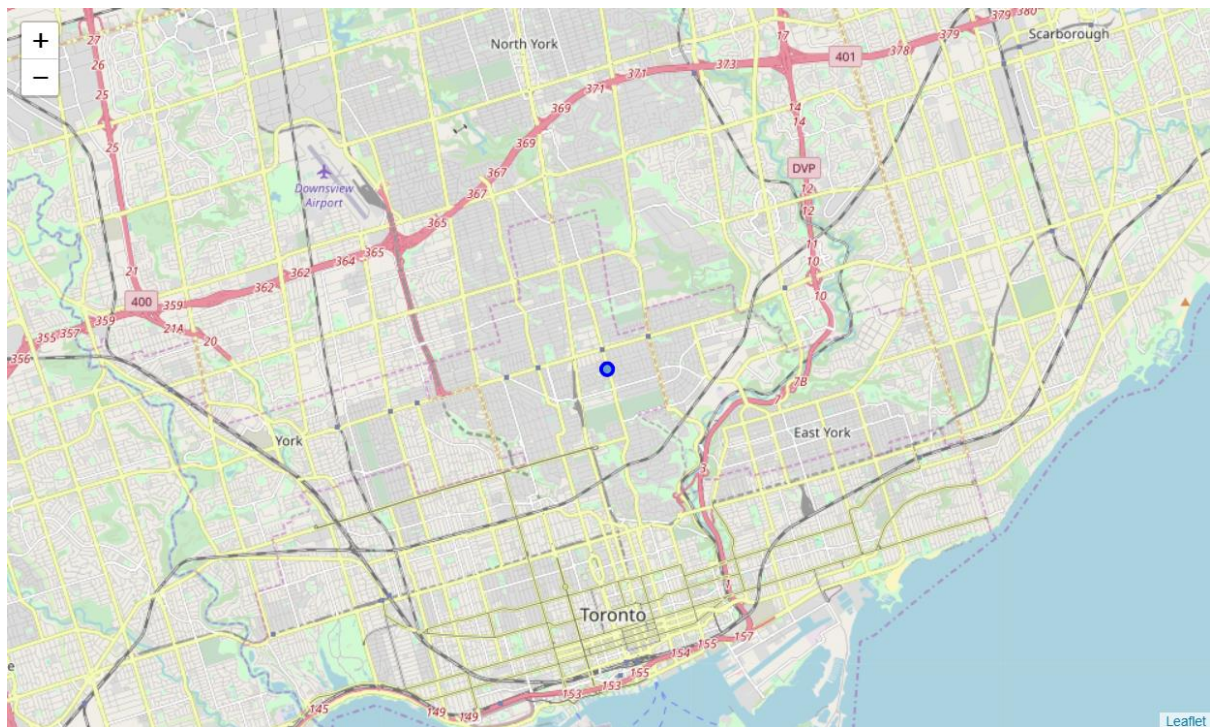
In New York City with 60 restaurants have Elmhurst the most largest number of restaurants per neighborhood.



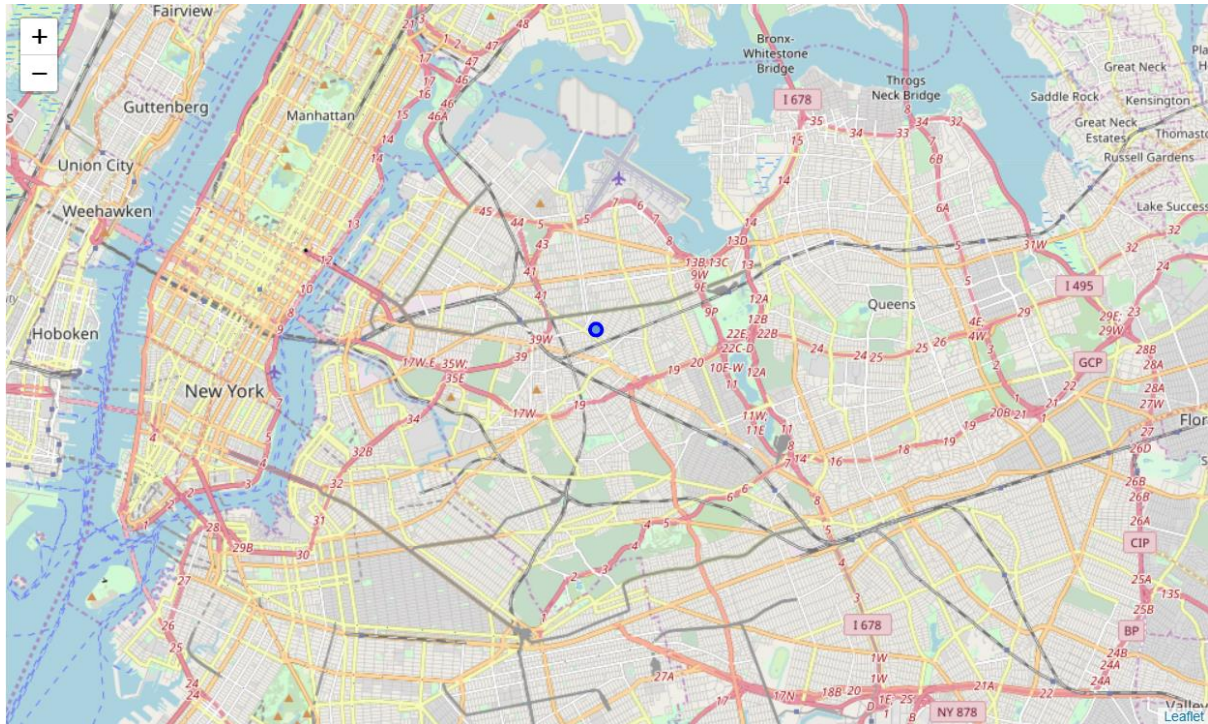
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Number of restaurants
132	Elmhurst	40.744049	-73.881656	60
131	Jackson Heights	40.751981	-73.882821	56
138	Flushing	40.764454	-73.831773	47
130	Woodside	40.746349	-73.901842	43
126	Gramercy	40.737210	-73.981376	42

So we have selected Davisville, Toronto and Elmhurst, New York City for further processing.

Visualising using Folium map of Davisville neighborhood, which is in Central Toronto borough...



... and of Elmhurst neighborhood, which is in Queens borough.



By exploring top 100 venues in Davisville and Elmhurst neighborhoods with Foursquare API calls, analyze neighborhood and grouping by venues category we printed top 10 most common venues.

In Davisville we had also found 57 unique categories and are mostly associated with a restaurant or food - fast food/shops/places (like bakery, dessert, sandwich, taco, salad, ice cream, smoothie), there are also a lot of ordinary shops, stores, pubs, bars, gym and tennis court.

In Elmhurst we found 42 unique categories and are like in Davisville mostly associated with an Asian restaurant or other food places (like snack, dessert, bakery, food truck, ice cream), there are also shops, stores, bars, gym, Nightclub, tea shops, grocery store and Barbershop.

Neighborhood	Davisville	Elmhurst
1st Most Common Venue	Italian Restaurant	Thai Restaurant
2nd Most Common Venue	Coffee Shop	Mexican Restaurant
3rd Most Common Venue	Sushi Restaurant	Bakery
4th Most Common Venue	Indian Restaurant	Chinese Restaurant
5th Most Common Venue	Café	South American Restaurant
6th Most Common Venue	Pizza Place	Latin American Restaurant
7th Most Common Venue	Dessert Shop	Asian Restaurant
8th Most Common Venue	Restaurant	Food Truck
9th Most Common Venue	Mexican Restaurant	Bubble Tea Shop

<b>10th Most Common Venue</b>	<b>Fast Food Restaurant</b>	<b>Vietnamese Restaurant</b>
-------------------------------	-----------------------------	------------------------------

We can see that Italian Restaurant is first most common venue in Davisville, and in Elmhurst is Thai Restaurant.

Then the Foursquare API was used to query of this two neighborhoods for the keyword {Restaurants} and extracted from top 100 venues. With help of that, we have explored for more information about likes, rating and prices of 10 most liked restaurants.

We saved in new dataframe separately for Davisville and Elmhurst.

We can see that in the table are different kinds of restaurants. The Grazie Ristorante – Italian Restaurant is most liked restaurant in the neighborhood and have 124 likes. Rating and Price columns were not perfectly extracted, because restaurants has not yet been rated. Only tree restaurants have more information than the others. We can see by rating that is Tabulè - Middle Eastern Restaurant most promising restaurant with 8.3 rate. According to the 'Moderate' price we assume that is fortible.

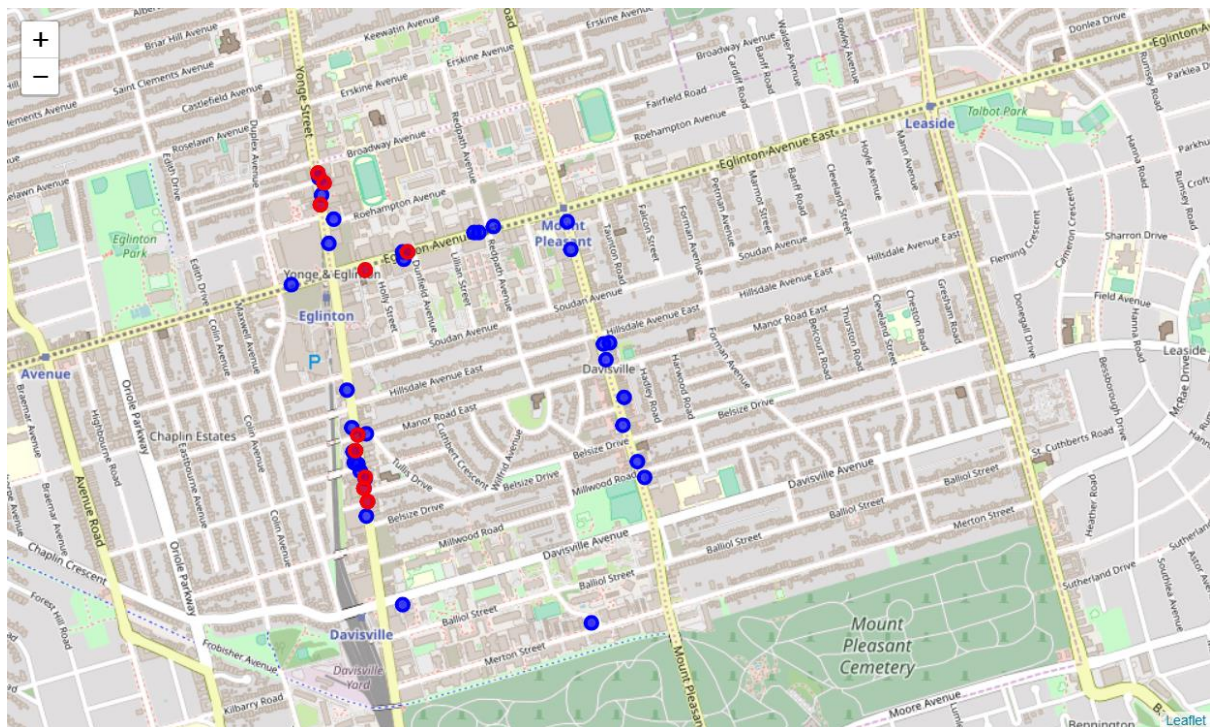
	Venue ID	Venue	Venue Category	Venue Latitude	Venue Longitude	Likes	Rating	Price
12	4ad4c061f964a520a2f720e3	Grazie Ristorante	Italian Restaurant	43.709329	-79.398823	124.0	NaN	NaN
4	521e0c6c04939a8ad55d93d3	Fresh	Vegetarian / Vegan Restaurant	43.707324	-79.395649	98.0	7.5 {'tier': 2, 'message': 'Moderate', 'currency':...}	
9	4aee242ef964a52077d221e3	Tabulè	Middle Eastern Restaurant	43.700932	-79.397078	89.0	8.3 {'tier': 2, 'message': 'Moderate', 'currency':...}	
6	53aca702498e3bbe34efc1bb	Little Sister	Indonesian Restaurant	43.701552	-79.397163	51.0	7.5 {'tier': 2, 'message': 'Moderate', 'currency':...}	
20	4ae24e96f964a5200e8d21e3	Coquine Restaurant	French Restaurant	43.702623	-79.397429	45.0	NaN	NaN
35	50353048e4b0c42300d9c847	A&W Canada	Fast Food Restaurant	43.706837	-79.397167	37.0	NaN	NaN
22	4ad4c060f964a5207ef720e3	Echo Sushi	Sushi Restaurant	43.701276	-79.397192	36.0	NaN	NaN
15	4b6b758f964a520270a2ce3	Five Doors North	Italian Restaurant	43.702236	-79.397526	30.0	NaN	NaN
31	4b01ef6ef964a520ad4622e3	Sushi Rock Café	Sushi Restaurant	43.709089	-79.398641	29.0	NaN	NaN
14	53865482498edb61054992bd	Star King	Vietnamese Restaurant	43.708519	-79.398764	25.0	NaN	NaN

And for Elmhurst. We can see that in the table are most asian restaurants. The JoJu - Vietnamese Restaurant is most liked restaurant in the neighborhood and have 277 likes. For Elmhurst neighborhood Rating and Price columns were also not perfectly extracted, because restaurants has not yet been rated. Only tree restaurants have more information than the others. We can see by rating that is Taste Good Malaysian Cuisine - Malay Restaurant most promising restaurant with 9.3 rate. According to the 'Moderate' price we assume that is fortible.

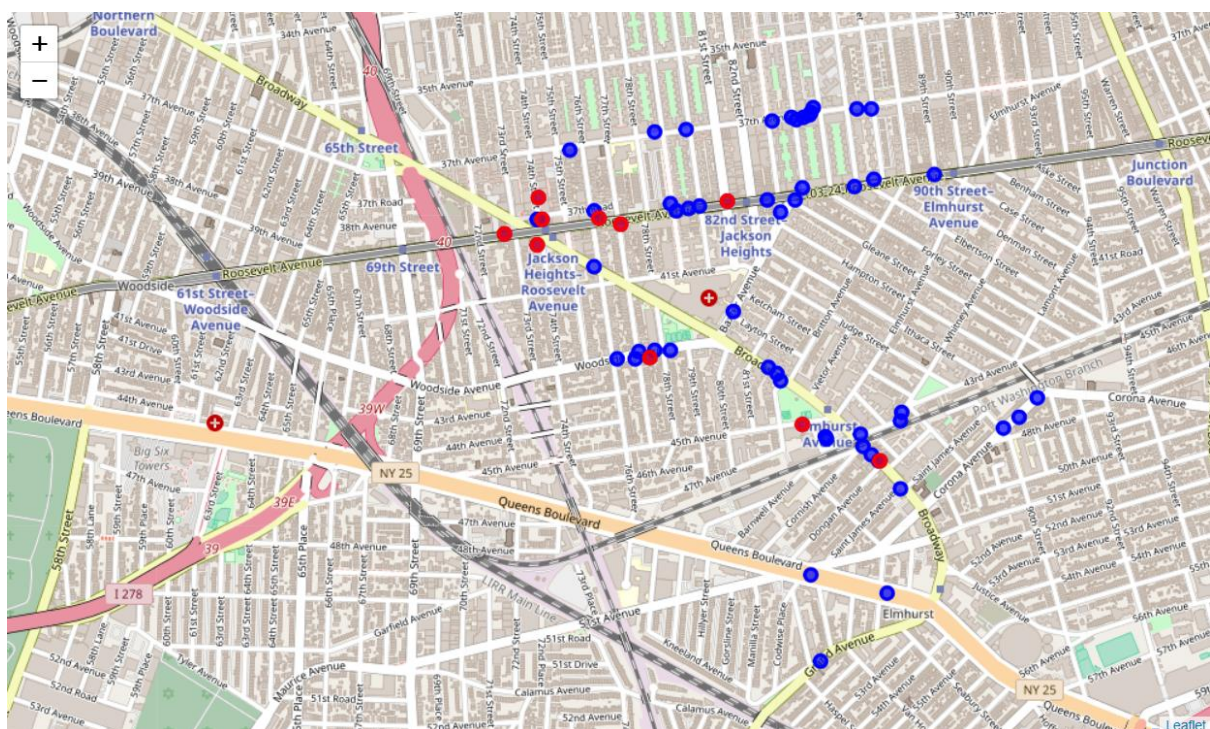
	Venue ID	Venue	Venue Category	Venue Latitude	Venue Longitude	Likes	Rating	Price
3	4e0f605462e1d5a30d0ab7ff	JoJu	Vietnamese Restaurant	40.740689	-73.879389	277.0	8.8 {'tier': 2, 'message': 'Moderate', 'currency':...}	
8	4b26f895f964a5209d8324e3	Ayada Thai	Thai Restaurant	40.743469	-73.887552	272.0	8.7 {'tier': 1, 'message': 'Cheap', 'currency': '\$'}	
37	4db3747a1e7248d1358dc5c8	Phayul	Asian Restaurant	40.747165	-73.891373	151.0	NaN	NaN
10	4bd2b6ef9854d13abe3afc4d	Pollos a la Brasa Mario	South American Restaurant	40.747659	-73.884806	145.0	NaN	NaN
56	3fd66200f964a52009f11ee3	Jackson Diner	Indian Restaurant	40.747753	-73.891517	137.0	NaN	NaN
54	4ab56c26f964a520867420e3	Himalayan Yak	Tibetan Restaurant	40.746767	-73.892697	115.0	NaN	NaN
46	4b4d2b63f964a52032cd26e3	Thai Son	Vietnamese Restaurant	40.746463	-73.891562	112.0	NaN	NaN
1	4aa58ef5f964a520cc4820e3	Taste Good Malaysian Cuisine 好味	Malay Restaurant	40.741663	-73.882177	110.0	9.3 {'tier': 2, 'message': 'Moderate', 'currency':...}	
33	466a4c63f964a52099471fe3	Taqueria Coatzingo	Mexican Restaurant	40.747188	-73.889374	105.0	NaN	NaN
26	53af195d498e0a4a49420b68	Areperia Arepa Lady	Arepa Restaurant	40.747016	-73.888591	98.0	NaN	NaN



Let's visualize this restaurants, top 10 restaurants by likes (red marks) and other Davisville restaurants (blue marks). We can see that most restaurants are on Yonge Street.



And top 10 restaurants by likes (red marks) and other Elmhurst restaurants (blue marks). We can see that most restaurant are nearby of Broadway street.

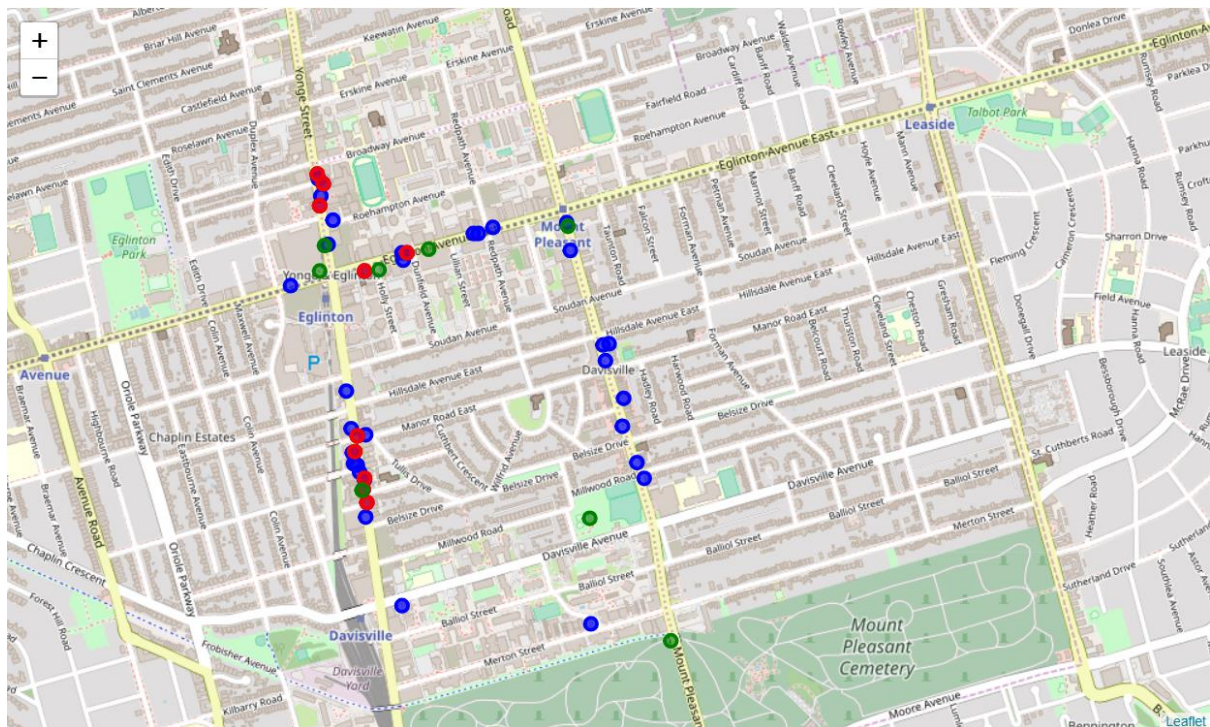


We have also explored about attractions in Davisville neighborhood...



	Venue	Venue Latitude	Venue Longitude	Venue Category
6	June Rowlands Park	43.700517	-79.389189	Park
8	Loblaws	43.707412	-79.394909	Supermarket
13	Granite Brewery	43.707991	-79.389943	Brewery
33	Yonge Eglinton Square	43.706864	-79.398751	Plaza
38	Mysteriously Yours, Mystery Dinner Theatre	43.701222	-79.397251	General Entertainment
59	Mount Pleasant Cemetery	43.697384	-79.386286	Cemetery
65	Cineplex Cinemas Yonge-Eglinton	43.707500	-79.398601	Movie Theater
74	Snakes & Lattes	43.706879	-79.396646	Gaming Cafe

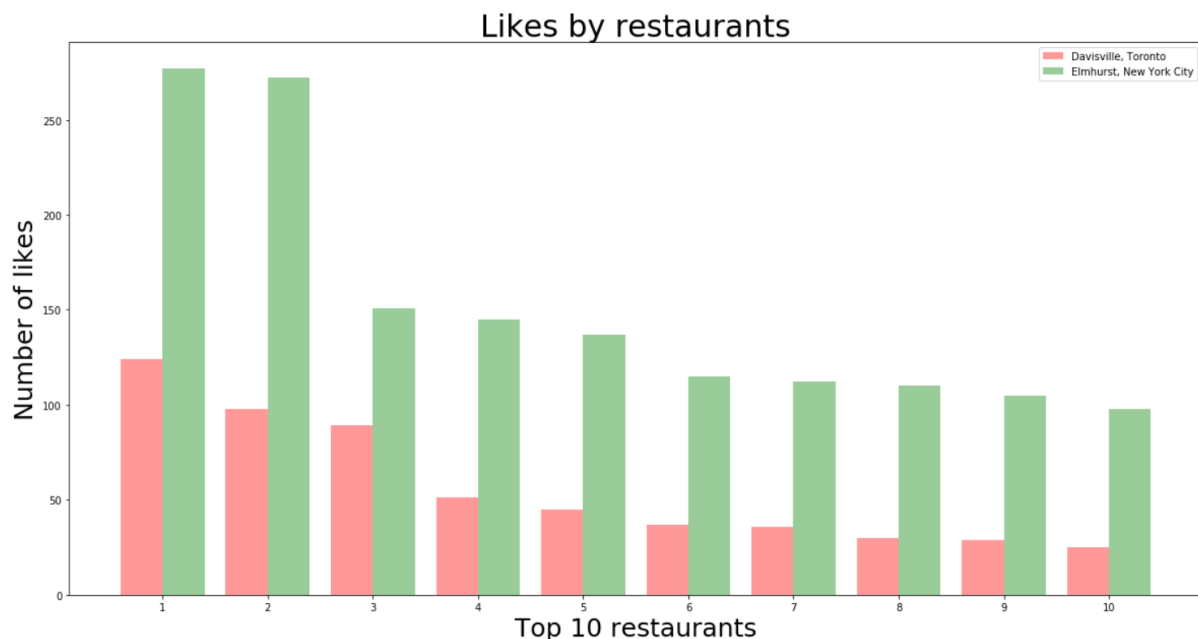
...and let's visualize attractions (green marks) with restaurants (red and blue marks) for Davisville.



The June Rowlands Park at the first look is very big and offers many walking trails to walk and visit cemetery before the traveler decides to eat again.

Elmhurst neighborhood had no nearby attractions. Just a lot of tea rooms, coffee shops and bars, that can the traveler spend some time for relaxing and resting. In the nearby neighborhood is a big park and inside a few Cemeterys, a few minutes walk away. Otherwise traveler need to go to Manhattan to see the biggest and most popular attractions.

At the end we compared between Davisville, Toronto and Elmhurst, New York City based on top 10 liked restaurants by plotting in bar charts.



## 5. Discussion

Toronto has 4861 venues, 329 unique categories and 103 neighborhoods, which has 1213 restaurants. With 40 restaurants is Davisville neighborhood the most largest numbered restaurants per neighborhood. Italian Restaurant is first most common restaurant and Grazie Ristorante is most liked restaurant in the neighborhood. Nearby attractions are June Rowlands Park, Plaza, Cemetery and other entertainments. There are also other food associated places and shops, for drinking pubs and bars, so there's plenty to choose to visit.

New York City has 20487 venues, 464 unique categories and 306 neighborhoods, which has 4925 restaurants. With 60 restaurants is Elmhurst neighborhood the most largest numbered restaurants per neighborhood. Thai Restaurant is first most common restaurant and JoJu is most liked restaurant in the neighborhood. Elmhurst neighborhood had no nearby attractions. Just a lot of tea rooms, coffee shops and bars, that can the traveler spend some time for relaxing and resting. In the nearby neighborhood is a big park and inside few Cemeterys, a few minutes walk away. Otherwise traveler need to go in Manhattan to see the biggest and most popular attractions. Most are asian restaurants and other food venues, ideal for asina food lovers.

Based on 124 likes that have Grazie Ristorante – Toronto Italian Restaurant and 277 likes that have JoJu – New York City Vietnamese Restaurant is JoJu most popular restaurant. All New York City restaurants have more likes than Toronto restaurants.

## 6. Conclusion

For this project, to make best possible recommendation based on popularity, we would like to recommend the traveler from Europe to visit New York City first for the best experience and enjoyment of visits to restaurants and nearby attractions, because Elmhurst offer way more restaurants then Davisville.