

Toronto VS New York, who is the more multicultural city in Catering

1. Introduction/Business Understanding

1.1 Description of the problem

New York City, the Empire State, is well-known as the BEST city on this planet. However, as a foodie who living around Greater Toronto Area, I confidently propose Toronto City as a decent rivalry to challenge New York on who is the more multicultural city in catering aspect.

1.2 Discussion of the background

According to the report, Toronto's 2020 population is now estimated at 6,196,731, it ranked as the fourth biggest city in North America. The racial composition is also diverse: White: 50.2% ,East Asian:12.7% (10.8% Chinese, 1.4% Korean, 0.5% Japanese) South Asian: 12.3% Black: 8.5% Southeast Asian: 7.0% (5.1% Filipino) Latin American: 2.8% and others.

The most common ancestry groups were: English (12.9%), Chinese (12.0%), Canadian (11.3%), Irish (9.7%), Scottish (9.5%), East Indian (7.6%), Italian (6.9%), Filipino (5.5%), German (4.6%), French (4.5%), and Polish (3.8%). Other common groups include Portuguese, Jamaican, Jewish, Ukrainian and Russian.

Toronto is known for being the most multicultural in the world, and the city promotion motto just like what it represents, "Diversity our strength", people living at Toronto are proud of having ethnic and authentic foods from all over the world. Toronto features ethnic supermarkets, farmers markets, cultural specific or diet specific restaurants and festivals that celebrate cultural diversity. Whether you want to try something new or eat something that reminds you of home. Different ethnic neighborhoods throughout the cit include Chinatown, K-town, Greektown, Little India and Little Italy.

On the other hand, New York is also famous for diversity and the melting pot of different races and cultures. More than 8 million population in New York composited by 65.7% white, 15.9% black, 7.3% Asian, and 7.4% of other races.

Both two cities have the incredible cultural diversity and has become the classic rivalry on who is the better city when it comes to food and catering, I will utilize

the data by scraping data from Internet and request venues by using location based system's API to generate the distribution of the restaurants and food courts, and also to see its diversity and scale, the customer ranking and viewing is also an aspect to evaluate its impact.

In general, the criteria will be:

- 1 The total number of different cuisines, and the per capita.
- 2 The total number of each diverse cuisines in each neighborhood
- 3 The different types of food in each neighborhood

2. Data Requirements

The demographic data of New York and Toronto

Datasource: <https://www.macrotrends.net>

Description: The demographic data of two cities is the foundation of this project and give us a general idea of the assumption that New York and Toronto is comparable on diversity in catering.

The GeoJson Data of all neighborhoods and communities at Toronto and New York

Datasource: <https://catalog.data.gov/dataset/city-boundary-1e013/resource/a8d680fd-ba89-4c0c-bdbd-61841bcd4568>

Description: The GeoJson files cover most of cities' neighborhoods in North America by their postal code or names.

The geographic information of each neighborhoods in New York and Toronto

Datasource: Google Maps Geocoding API

Description: To obtain the latitude and longitude coordinates of each neighborhood at two cities.

Toronto and New York data that contains list districts (Wards) along with their latitude and longitude:

Datasource : [https://en.wikipedia.org/wiki/](https://en.wikipedia.org/wiki/Special:wards_of_Toronto#List_of_special_wards)

[Special_wards_of_Toronto#List_of_special_wards](https://en.wikipedia.org/wiki/Special:wards_of_Toronto#List_of_special_wards)

Description: We will Scrap New York and Toronto districts (Wards) Table from Wikipedia and get the coordinates of these 23 major districts using geocoder class of Geopy client.

Restaurants and food courts in each neighborhood of New York and Toronto:

Data source: Foursquare APIs

Description : By using this API we will get all the venues in each neighborhood. We can filter these venues to get only restaurants and food courts.

The detail and types of each restaurant and food court in each neighborhood of New York and Toronto:

Data source: Foursquare APIs

Description : By using this API we will get all the detail, type and menu of each venue.

3. Methodology

Three stages focused exploring the geographic data from two cities base on all neighborhoods, and requesting all the venues related catering and food by using Foursquare API. The returned venues will also be cleaned and divided into 7 cuisine cultures including Asian food, Indian food, European food, Latin American, Middle Eastern, African and Mediterranean. The diversity score is assigned for representing the diversity of each community, and the criteria is the number and diversity of the restaurant. Finally the choropleth folium map would generate diversity level of each neighborhood of two cities, it would be easy to see the comparison between two cities even among the communities directly.

Stage A

(1) Scrape and save the data from Wikipedia page to request all neighborhoods and boroughs of two cites. The dataset of geographic information from Google Maps Geocoding API append two columns which are latitude and longitude coordinates for each neighborhoods. Then using Pandas library to transform the two DataFrames and they are showing like:

New York City

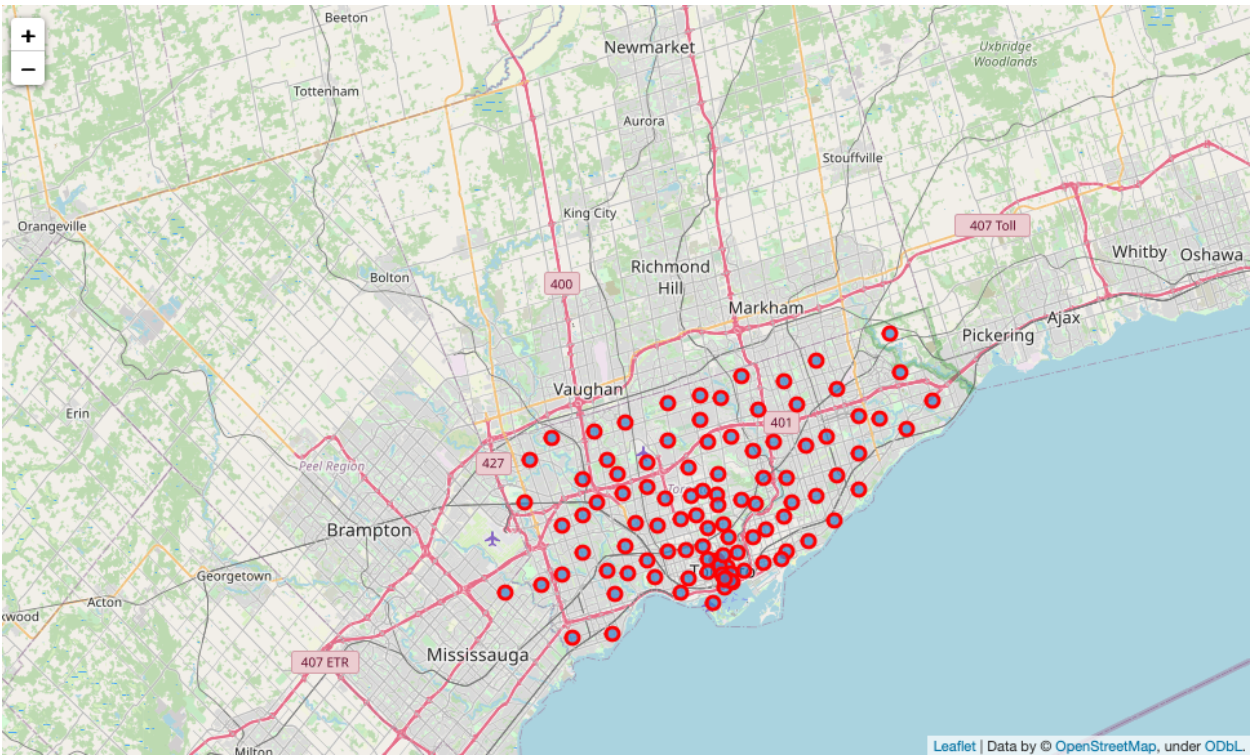
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446
10	Bronx	Baychester	40.866858	-73.835798

The City of Toronto:

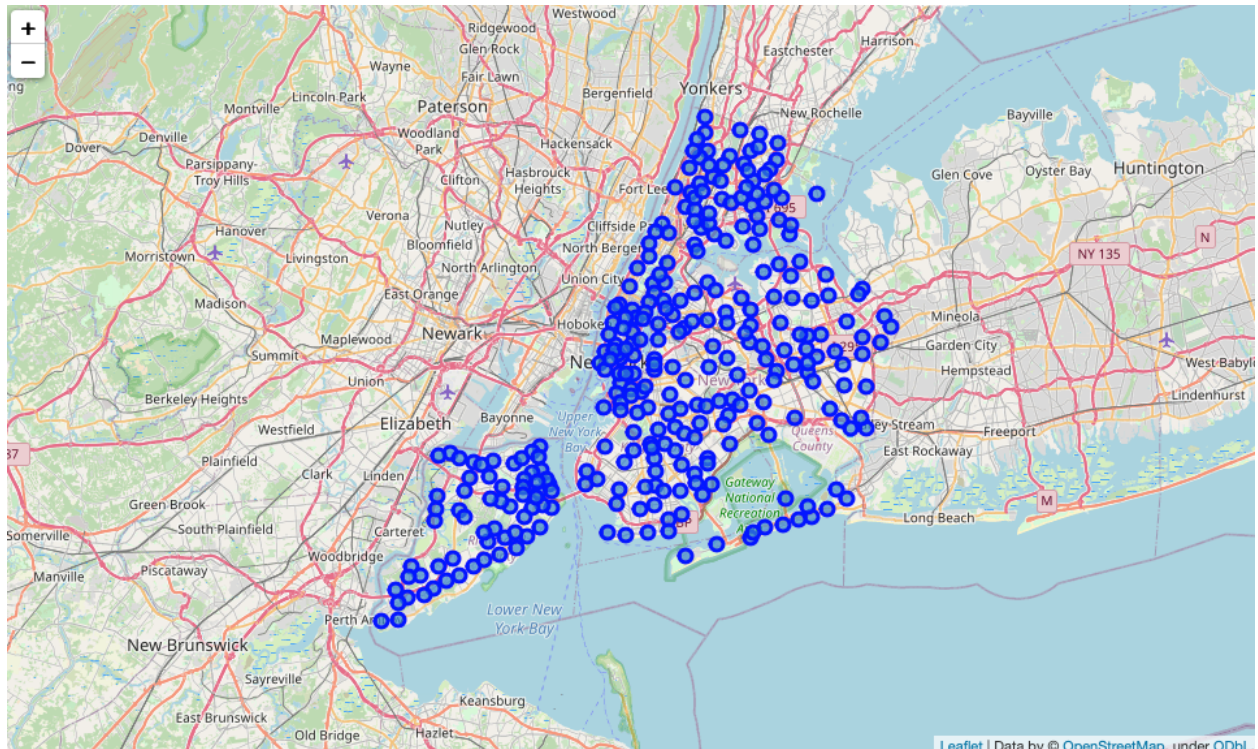
	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073

(2) Generate the clear views of each neighborhood for two cites by showing it on choropleth folium map, we also notice that New York has 306 neighborhoods comparing the number of 102 at Toronto from our dataset, it reminds us to use mean value and per capita since New York is objectively a larger city and has bigger population and communities than Toronto.

New York City



The City of Toronto:



Stage B

- (1) Create a function to return all venues for each neighborhood in New York and Toronto by utilizing the Foursquare API:

```
toronto_venues = getNearbyVenues(names=df_to['Neighborhood'],
                                  latitudes=df_to['Latitude'],
                                  longitudes=df_to['Longitude']
                                )
```

```
newyork_venues = getNearbyVenues(names=df_ny['Neighborhood'],
                                   latitudes=df_ny['Latitude'],
                                   longitudes=df_ny['Longitude']
                                 )
```

- (2) By using Groupby function and Count attribute from Panda library return the dataset with each neighborhood and its number of venues nearby:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Allerton	34	34	34	34	34	34
Annadale	13	13	13	13	13	13
Arden Heights	4	4	4	4	4	4
Arlington	8	8	8	8	8	8
Arrochar	21	21	21	21	21	21
Arverne	20	20	20	20	20	20
Astoria	100	100	100	100	100	100
Astoria Heights	12	12	12	12	12	12
Auburndale	19	19	19	19	19	19
Bath Beach	50	50	50	50	50	50

(3) Hot encoding to transform our list of established all categories of venues, return the mean value of number of venues of each neighborhood

```
ny_ground = newyork_onehot.groupby('Neighborhood').mean().reset_index()
ny_ground.head(5)
```

	Neighborhood	Yoga Studio	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Terminal	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum
0	Allerton	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0
1	Annadale	0.0	0.0	0.0	0.0	0.0	0.0	0.076923	0.0	0.0	0.0	0.0	0.0	0.0
2	Arden Heights	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0
3	Arlington	0.0	0.0	0.0	0.0	0.0	0.0	0.125000	0.0	0.0	0.0	0.0	0.0	0.0
4	Arrochar	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0

(4) New York has 426 unique categories and Toronto has 264, the categories related to restaurant and catering are the targets we need. We filtered the dataset by setting the key strings are “Restaurant” and “Food”, and we assigned the returned dataset as ny_restaurant and to_restaurant, let’s take New York as an example:

```
filt = [x for x in ny_ground.columns[ny_ground.columns.str.contains('Restaurant')]]
```

```
ny_restaurant.head(5)
```

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cambodian Restaurant	Cant Rest
0	Allerton	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	Annadale	0.0	0.0	0.076923	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Arden Heights	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Arlington	0.0	0.0	0.125000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Arrochar	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(5) The return dataset represent the categories of the restaurants and we need to divide them into different types of food, we set up 7 groups which are Asian Food, Indian Food, European Food, Latin American Food, Middle Eastern Food, African Food, Mediterranean Food, and the crucial point for this stage is that we know there is no ultimate solution to separate different types food into exact group, and the division does not represent all type of food in all cultures, and I will propose several solutions and discuss them at discussion part in report. Finally we have the all restaurants with 7 groups for each neighborhood.

to_restaurant

	Neighborhood	Asian Food	Indian Food	European Food	Latin American Food	Middle Eastern Food	African Food	Mediterranean Food
0	Agincourt	0.000000	0.000000	0.000000	0.250000	0.000000	0.000000	0.000000
1	Alderwood, Long Branch	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.095238	0.000000	0.000000	0.000000	0.047619	0.000000	0.000000
3	Bayview Village	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Bedford Park, Lawrence Manor East	0.120000	0.040000	0.120000	0.000000	0.000000	0.000000	0.000000
5	Berczy Park	0.035088	0.017544	0.035088	0.000000	0.000000	0.000000	0.000000
6	Birch Cliff, Cliffside West	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	Brockton, Parkdale Village, Exhibition Place	0.000000	0.000000	0.041667	0.000000	0.000000	0.000000	0.000000
8	Business reply mail Processing Centre, South C...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Stage C

(1) The diversity score is the evaluation for measuring the catering diversity of a neighborhood base on the mean value of number of restaurant from different culture.

```
ny_restaurant_score = ny_restaurant[['Neighborhood', 'diversity_score']]
ny_restaurant_score.head(5)
```

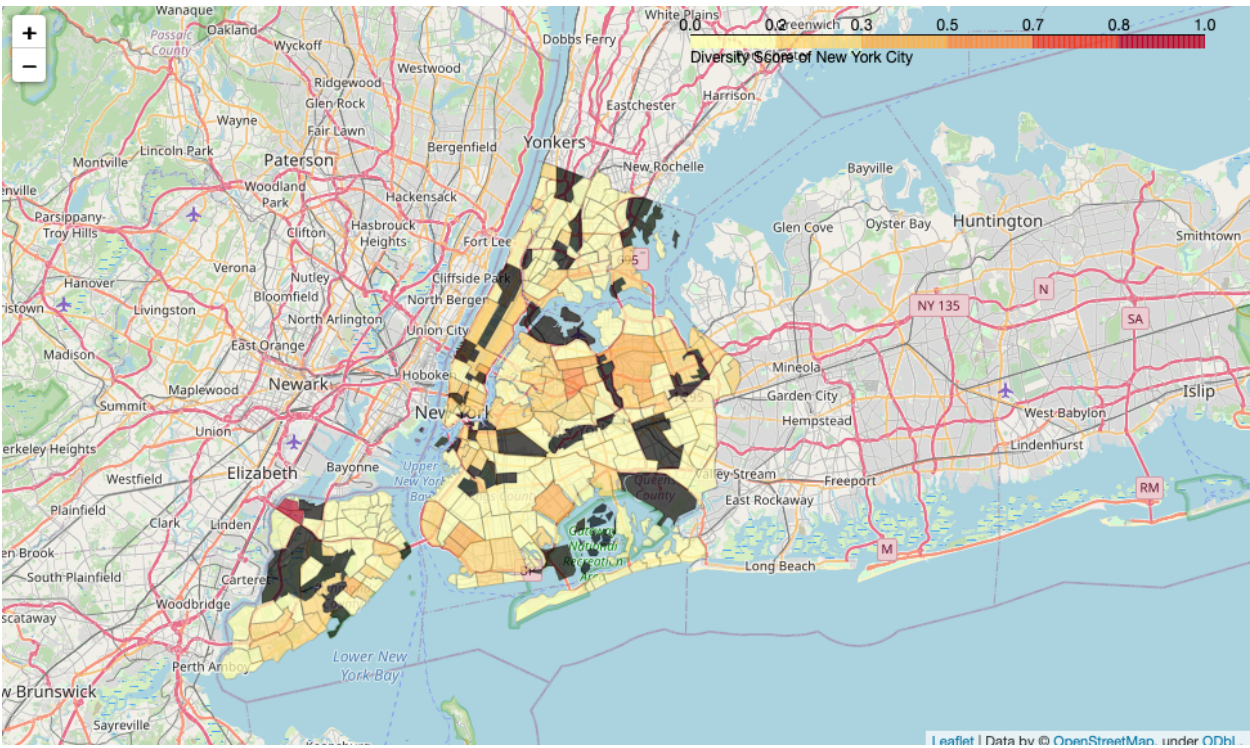
	Neighborhood	diversity_score
0	Allerton	0.058824
1	Annadale	0.153846
2	Arden Heights	0.000000
3	Arlington	0.000000
4	Arrochar	0.190476

(2) At this stage, we collected the Geo Json files from two cites that would clearly locate each neighborhood based on their names or postal codes.

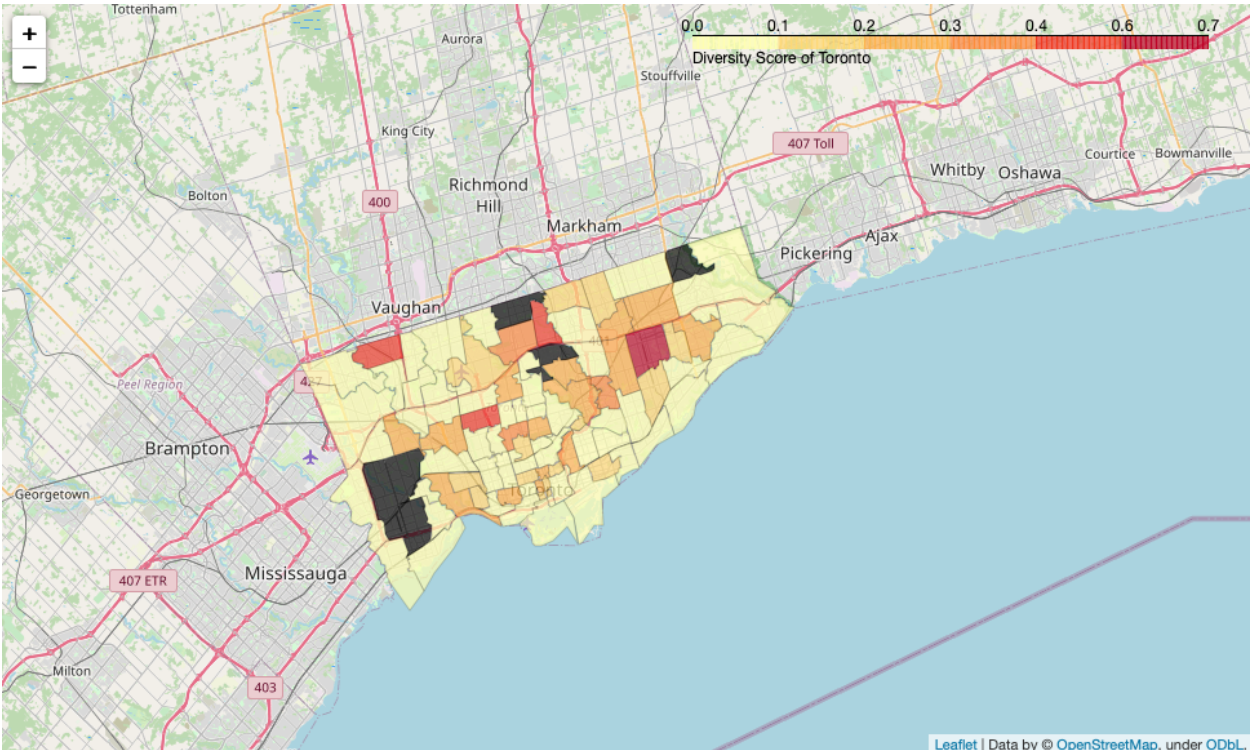
4. Results

(1) The Chopleth map shows the diversity level of two cities among all neighborhoods, the legend of the map is the diversity score of the city and the dark areas are due to NaN value which means we currently have no information in those neighborhoods.

New York City:



City of Toronto:



5. Discussion

From the results part that presented above, there are several observations and recommendations need to be discussed, mainly from the stage B about data cleaning and the standard of setting diversity score.

- (1) We notice the limitation and narrow perspective to divide all kinds of food into 7 groups, the classification could be more scientific and reasonable. The division we used in this report is most likely based on geographic reason, but we do realized that Chinese food, Japanese food, Korean food and Taiwanese food are so different even they are assigned into the Asian Food group, and the same situation also happened in other groups.
- (2) Most of the catering venues returned from using Foursquare API are restaurants, and we differentiate them by their names. For example, we assumed Sushi restaurant is Japanese food and belong to Asian Food group, we did not eliminate the possibility that the name of the restaurant do not represent its type and mistakenly used by us.
- (3) The diversity score is not objective enough, the different parameters should be considered when it comes to the size and impact of different restaurant. We use the same parameters in this report and set it as 1, but it should be mentioned that diversity score is not only about quantity, the quality and reviews of the restaurant is also the key factors.
- (4) Missing data is due to two reasons: 1) The venues data is missing from Foursquare and the get none information when using request function. 2) The neighborhood geo json from Google API is not exactly same with Wikipedia, since the neighborhood's name and postal code are from Wiki and geographic information is from Google API, so they do not match 100%.

6. Conclusion

In conclusion, we assume that New York has more restaurants from more different types food, it covers most neighborhoods among Bronx, Manhattan, queen, Brooklyn. Toronto is also a diverse city in catering and multicultural in Downtown, Scarborough and North York areas. Beside the missing value of the data and set diversity score as the evaluation of our assumption, we think New York is the more diverse city on catering aspect comparing Toronto.