# Capstone Project - The Battle of Neighborhoods

## Introduction/Business Problem

With the huge amount of data and feedbacks in Foursquare, as well as the geographic information of NYC and Toronto, I am curious about the following questions:

- We all know both NYC and Toronto are big multicultural cities, but is there any difference of taste in these 2 cities? What are the most favored and least favored styles of food in NYC and Toronto? Are the preferences of food consistent with their race/immigrants (original) nationality composition?
- As social network apps are more and more popular nowadays, the feedbacks on apps like Foursquare, Yelp, etc. are more and more important to restaurants. Higher rating, more comments and more fancy photos are great help to promote a restaurant and attract more customers. I would like find out what kind of restaurants have the highest ratings, the most comments and the most photos. In addition, I want to find out do more feedbacks (comments and photos) imply extremer (high or low) ratings.
- Cluster the neighborhoods by the number of feedbacks. Currently, I assume restaurants in business areas will have less feedbacks. Because people go there either just grab something to eat during breaks or have business dinners, they do not have time/mood to take pictures and leave comments or it's inappropriate to do so in a business dinner. On the other hand, restaurants in leisure/fun/living areas may have more feedbacks, because people usually go there with family and friends and they like taking pictures and commenting on the foods/atmospheres under such scenario. I do not live either city and I do not know too much about them, yet I am relatively more familiar with NYC, so I will only study NYC for this item.

## Data

- Most of the data are available in Foursquare, but need to write new codes to pull them (i.e., ratings, photos, comments).
- Geographic data is available from previous projects.
- Need to find the race/immigrants (original) nationality composition for these 2 cities.

## Methodology

- In this report, we gathered restaurants information of each style from Foursquare and investigated the distribution of these different style restaurants. We are wondering whether the distribution of the restaurants is consistent with the distribution of common impression of food preference or ethnic/race in Toronto and New York city.

- We also tried to use the restaurant style information combined with geographic information to classify neighborhoods in New York City.

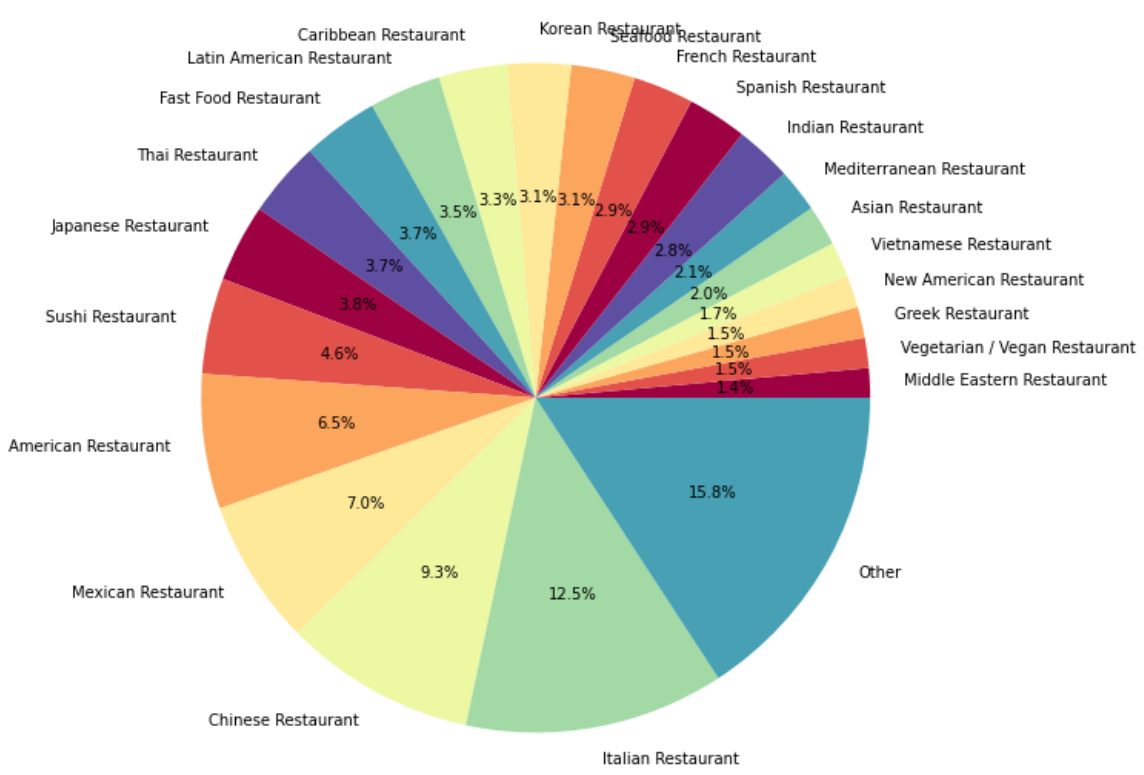
## Results and Discussion

- There are totally 2399 restaurants in New York city and 342 restaurants in Toronto after data cleaning.
- In New York city, Italian restaurants out numbers other flavored restaurants. Chinese and Mexican restaurants are the 2nd and 3rd.

[84]:

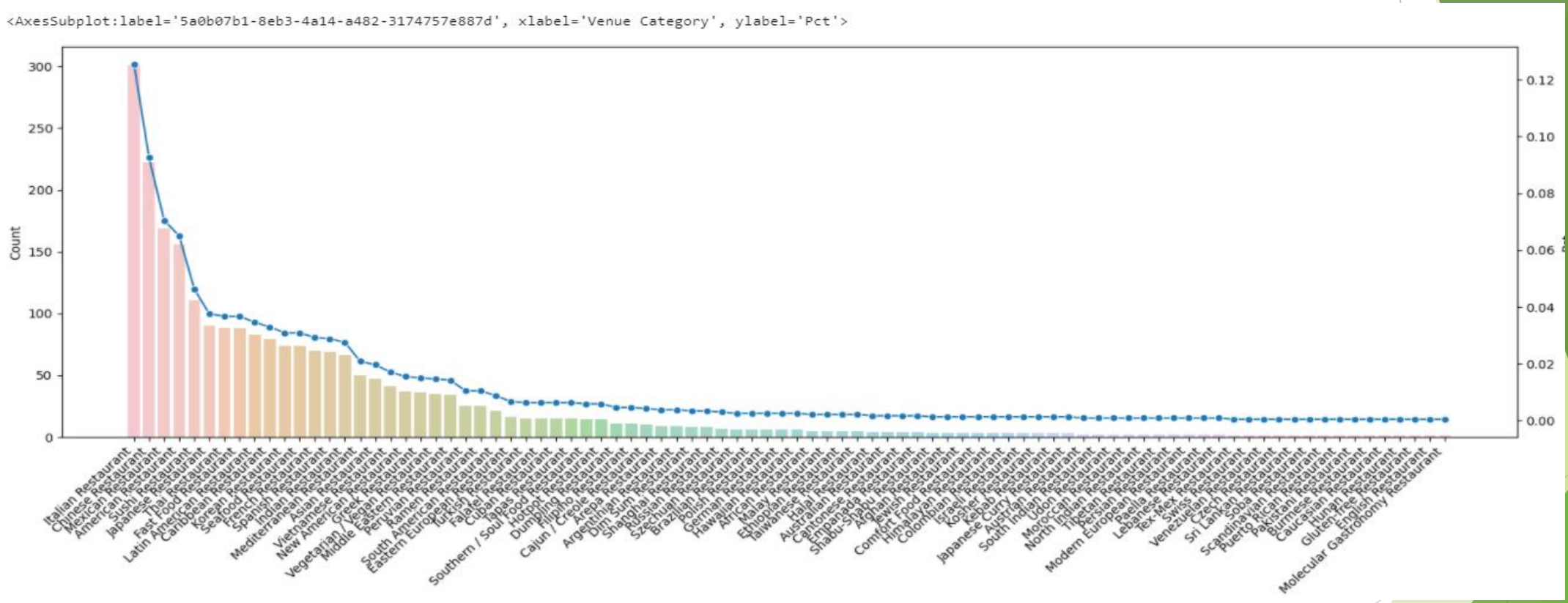
| | Venue Category | Count | Pct | Pct2 |
|---|---|---|---|---|
| 39 | Italian Restaurant | 301 | 0.125469 | 12.55% |
| 14 | Chinese Restaurant | 222 | 0.092539 | 9.25% |
| 50 | Mexican Restaurant | 169 | 0.070446 | 7.04% |
| 2 | American Restaurant | 156 | 0.065027 | 6.50% |
| 75 | Sushi Restaurant | 111 | 0.046269 | 4.63% |
| 41 | Japanese Restaurant | 90 | 0.037516 | 3.75% |
| 81 | Thai Restaurant | 88 | 0.036682 | 3.67% |
| 26 | Fast Food Restaurant | 88 | 0.036682 | 3.67% |
| 46 | Latin American Restaurant | 83 | 0.034598 | 3.46% |
| 12 | Caribbean Restaurant | 79 | 0.032930 | 3.29% |

[66]: <module 'matplotlib.pyplot' from '/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/matplot



# Results and Discussion

[67]: <AxesSubplot:label='5a0b07b1-8eb3-4a14-a482-3174757e887d', xlabel='Venue Category', ylabel='Pct'>



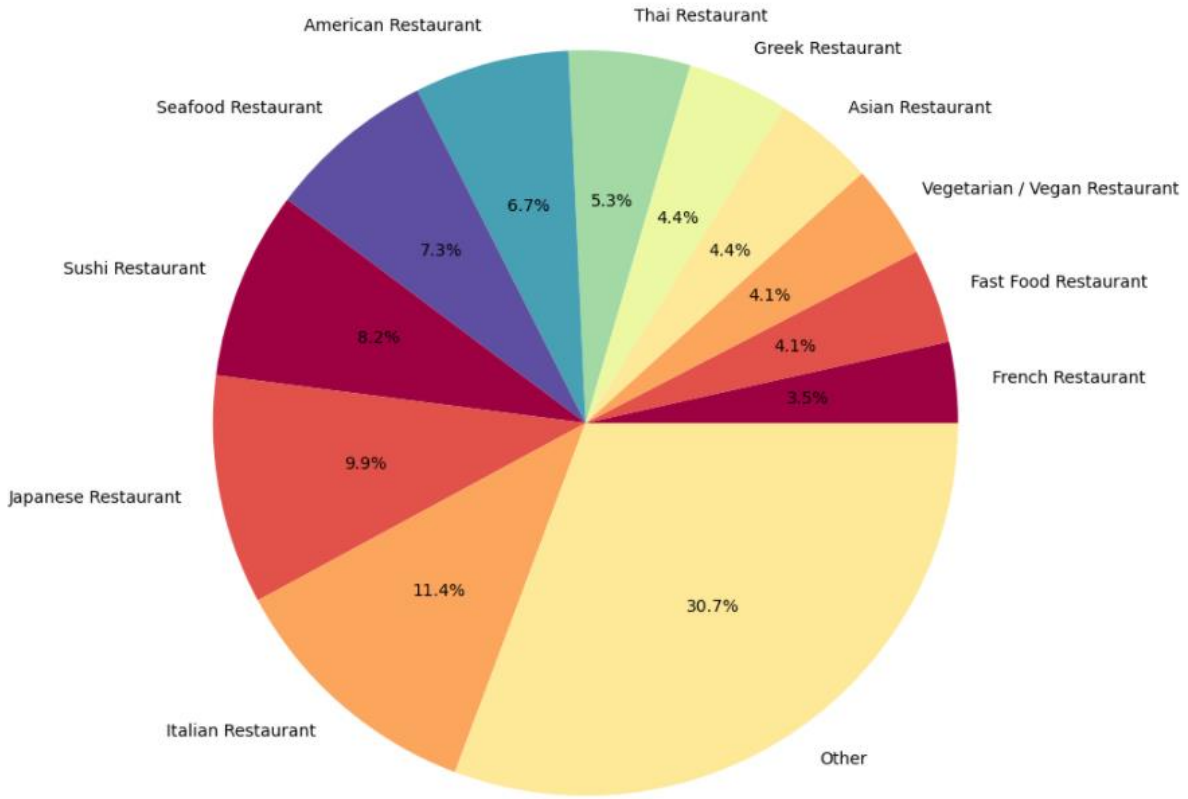
## Results and Discussion

- In Toronto, the most numbered restaurant is Italian as well, But the 2nd and 3rd are Japanese and Sushi restaurants, which should be both are Japanese restaurants. The 4th is Seafood, which might include Japanese restaurants. The 5th is American restaurant.

[13]:

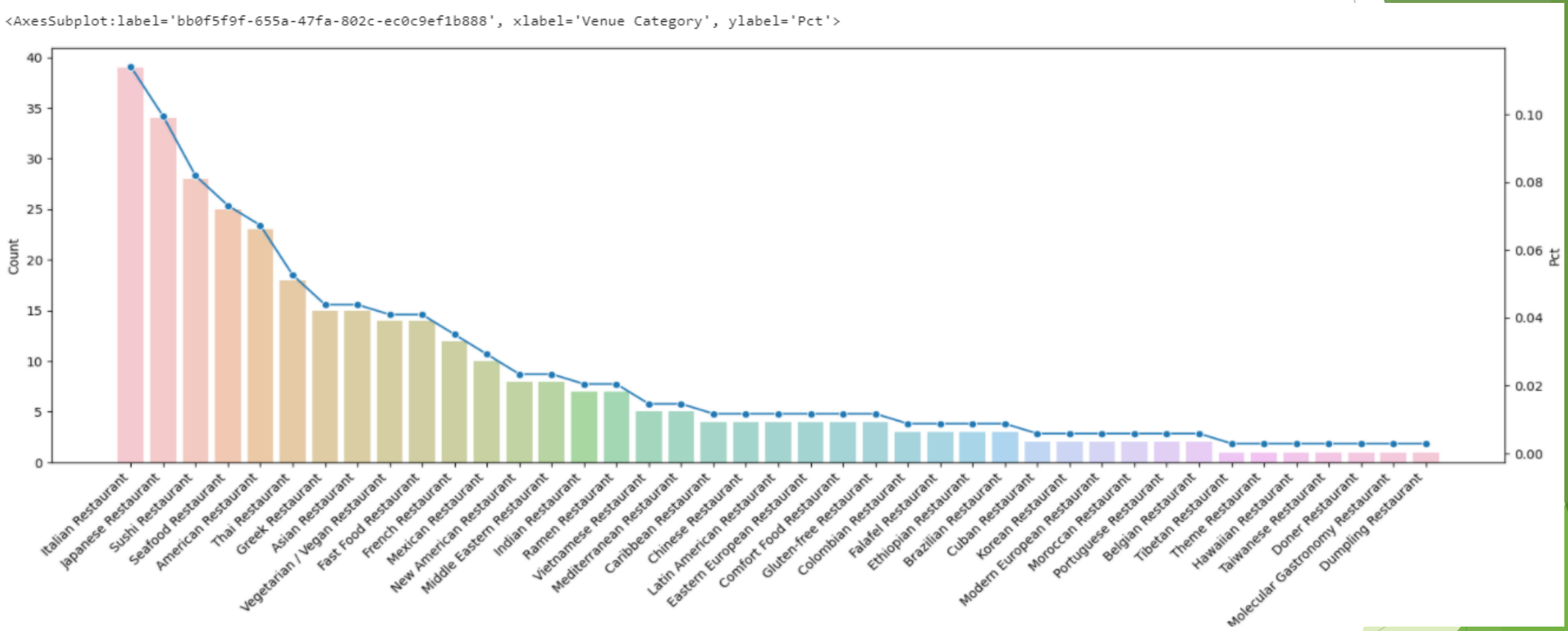
| | Venue Category | Count | Pct | Pct2 |
|---|---|---|---|---|
| 20 | Italian Restaurant | 39 | 0.114035 | 11.40% |
| 21 | Japanese Restaurant | 34 | 0.099415 | 9.94% |
| 34 | Sushi Restaurant | 28 | 0.081871 | 8.19% |
| 33 | Seafood Restaurant | 25 | 0.073099 | 7.31% |
| 0 | American Restaurant | 23 | 0.067251 | 6.73% |
| 36 | Thai Restaurant | 18 | 0.052632 | 5.26% |
| 17 | Greek Restaurant | 15 | 0.043860 | 4.39% |
| 1 | Asian Restaurant | 15 | 0.043860 | 4.39% |
| 39 | Vegetarian / Vegan Restaurant | 14 | 0.040936 | 4.09% |
| 14 | Fast Food Restaurant | 14 | 0.040936 | 4.09% |

[68]: <module 'matplotlib.pyplot' from '/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/matplotlib/pyplot.py'>



# Results and Discussion

[69]: <AxesSubplot:label='bb0f5f9f-655a-47fa-802c-ec0c9ef1b888', xlabel='Venue Category', ylabel='Pct'>



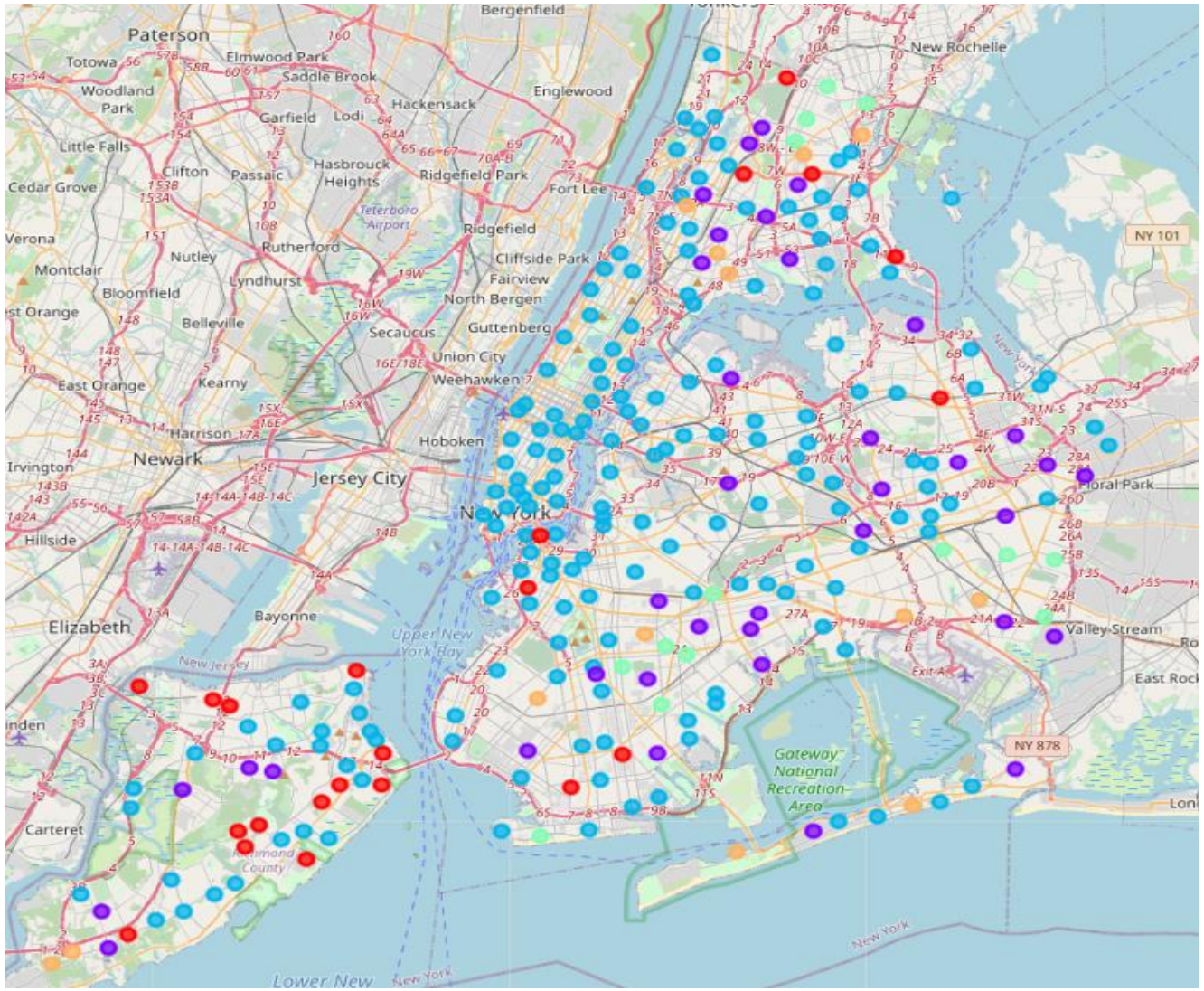
## Results and Discussion

- We used the number of each kind restaurants and geographic information to group the neighborhoods by clustering and pinned in a map.

[93]:

| | Borough | Neighborhood | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Bronx | Wakefield | 40.894705 | -73.847201 | 3.0 | Caribbean Restaurant | Vietnamese Restaurant | Empanada Restaurant | Ethiopian Restaurant | Falafel Restaurant | Fast Food Restaurant | Filipino Restaurant | French Restaurant | German Restaurant | Gluten-free Restaurant |
| 1 | Bronx | Co-op City | 40.874294 | -73.829939 | 4.0 | Fast Food Restaurant | Vietnamese Restaurant | Venezuelan Restaurant | English Restaurant | Ethiopian Restaurant | Falafel Restaurant | Filipino Restaurant | French Restaurant | German Restaurant | Gluten-free Restaurant |
| 2 | Bronx | Eastchester | 40.887556 | -73.827806 | 3.0 | Caribbean Restaurant | Seafood Restaurant | Fast Food Restaurant | Chinese Restaurant | Vietnamese Restaurant | Gluten-free Restaurant | Ethiopian Restaurant | Falafel Restaurant | Filipino Restaurant | French Restaurant |
| 5 | Bronx | Kingsbridge | 40.881687 | -73.902818 | 2.0 | Latin American Restaurant | Mexican Restaurant | Spanish Restaurant | Fast Food Restaurant | Seafood Restaurant | Chinese Restaurant | Caribbean Restaurant | Vietnamese Restaurant | Ethiopian Restaurant | Falafel Restaurant |
| 6 | Manhattan | Marble Hill | 40.876551 | -73.910660 | 2.0 | Seafood Restaurant | American Restaurant | Vietnamese Restaurant | Greek Restaurant | English Restaurant | Ethiopian Restaurant | Falafel Restaurant | Fast Food Restaurant | Filipino Restaurant | French Restaurant |

## Results and Discussion



## Results and Discussion

- Italian restaurants ranked number 1 in both New York city and Toronto in terms of the number of restaurants.  Chinese and Mexican restaurants are also popular in New York city. Japanese restaurants are popular in Toronto. I think the distribution makes sense and is consistent with the common impression.
- In the map of New York city, we can find although Italian restaurants out numbered Chinese restaurants, the Chinese restaurants neighborhoods (blue dot) seems covering more places than Italian restaurants neighborhoods (purple dot).
- An interesting finding is in Little Italy, the 1st common venue is Chinese restaurant, the 2nd is Mediterranean, the 3rd is Italian.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 121 | Little Italy | Chinese Restaurant | Mediterranean Restaurant | Italian Restaurant | Vietnamese Restaurant | Hotpot Restaurant | French Restaurant | Seafood Restaurant | Japanese Restaurant | Thai Restaurant | Asian Restaurant |

- One thing we can improve in the future research is group the venue categories more accurate. For example, in Toronto, the 2nd numbered restaurant is Japanese restaurant and the 3rd is Sushi. They both should be classified as Japanese restaurant.

## Conclusion

- we have some really interesting finds by description statistics and clustering. The result is consistent with common impression.