Capstone Project - The Battle of Neighborhoods

1. **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:

1. 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?
2. 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.
3. 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.
4. **Data**
5. Most of the data are available in Foursquare, but need to write new codes to pull them (i.e., ratings, photos, comments).
6. Geographic data is available from previous projects.
7. Need to find the race/immigrants (original) nationality composition for these 2 cities.
8. **Methodology**

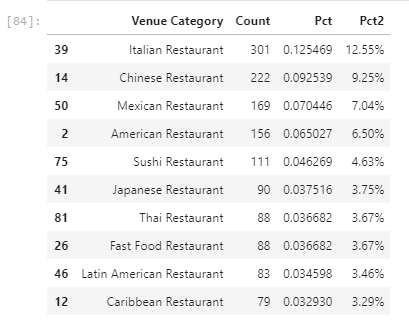
In this report, we gathered restaurants information of each style from Foursquare and investigated the distribution of these different style restaurants. We are wondering whether the distribution of the restaurants is consistent with the distribution of common impression of food preference or ethnic/race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

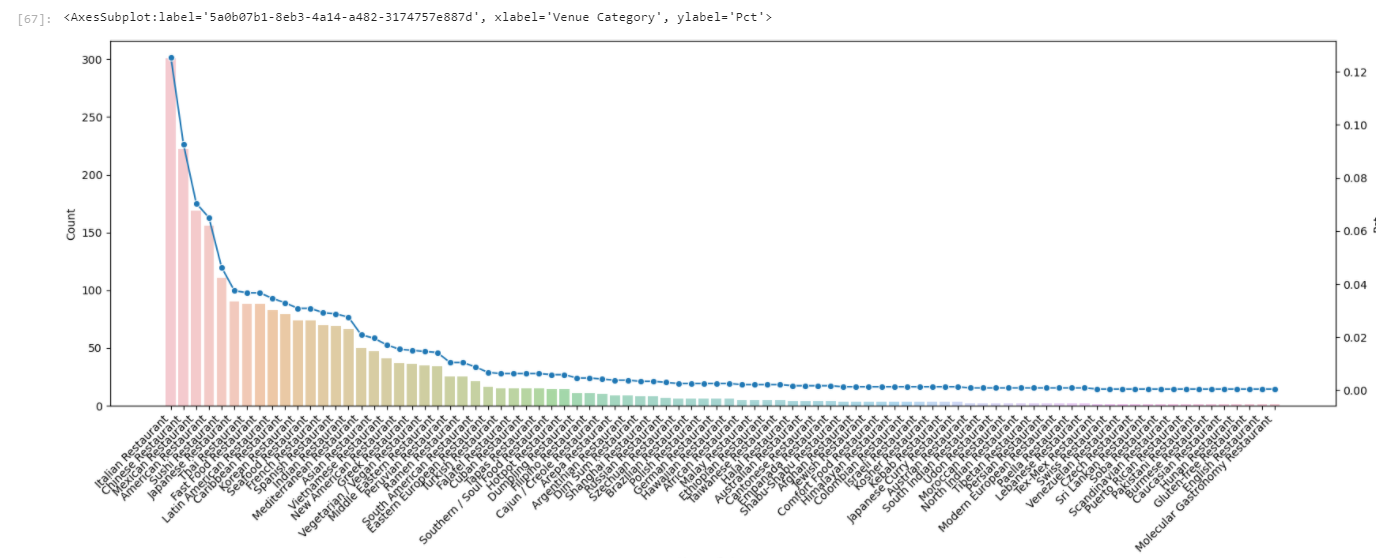
We also tried to use the restaurant style information combined with geographic information to classify neighborhoods in New York City.

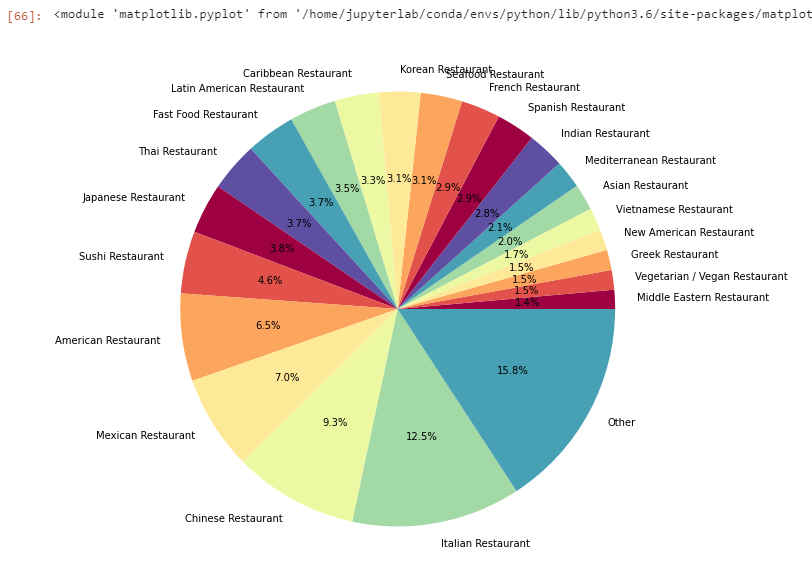
1. **Results and Discussion**

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.

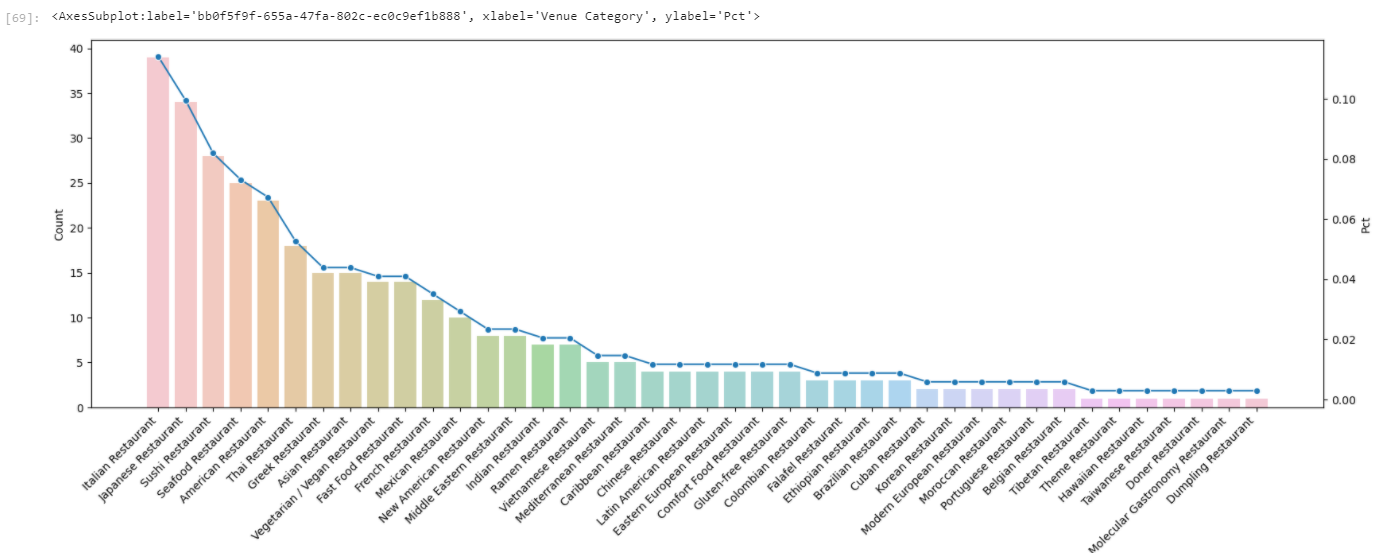


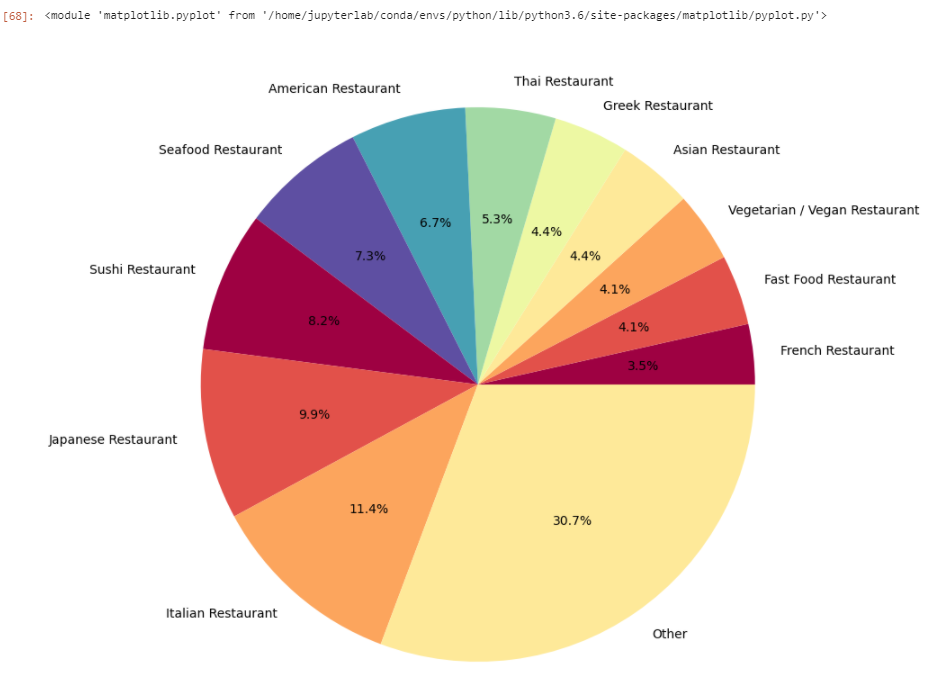




In Toronto, the most numbered restaurant is Italian as well, But the 2nd and 3rd are Japanese and Sushi restaurants, which should be both are Japanese restaurants. The 4th is Seafood, which might include Japanese restaurants. The 5th is American restaurant.

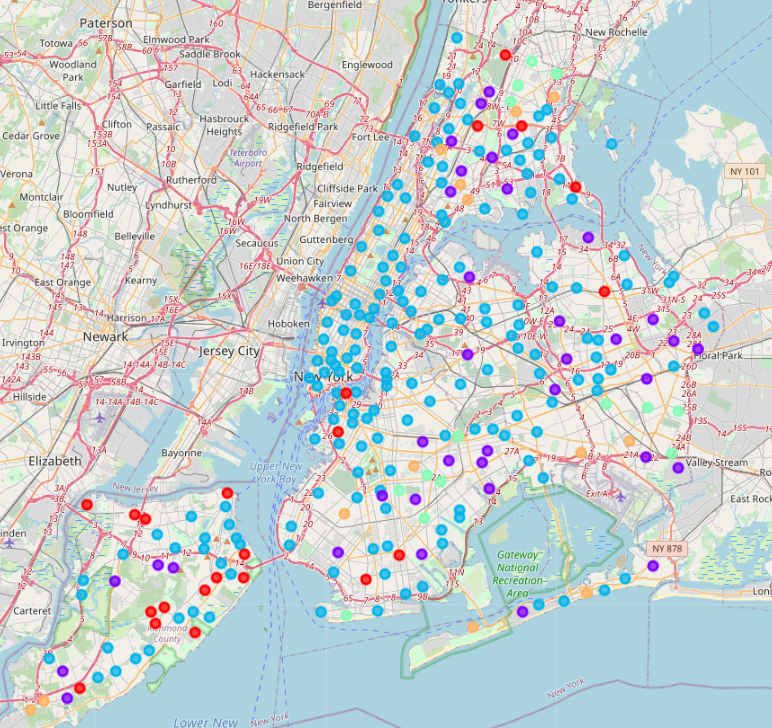






We used the number of each kind restaurants and geographic information to group the neighborhoods by clustering and pinned in a map.





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.



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.

1. **Conclusion**

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