New York City vs. Toronto City

Restaurant Site Analysis

Mike C

March 2020

Introduction

This is a hypothetical project conducted for a hypothetical client. The goal of this project is to help the client to decide on a restaurant site. Client has a background in pastry and would like to open a dessert shop in either New York City, NY in United States, or Toronto in Canada. We will study the current restaurant scenes in both cities, analyze the commonality and dissimilarity, and see which one would have a better chance to accommodate a new dessert shop.

New York City (NYC) is one of the most popular city in the United States, if not the world. According to Wikipedia, it currently has over 8.4 million people reside in the city, it is the most densely populated major city in the United States. NYC has an estimated gross metropolitan product of $2.0 trillion. Lastly, NYC is divided into five administrative boroughs, they are Brooklyn, Queens, Manhattan, The Bronx and Staten Island.

Toronto, on the other hand, is the capital of Ontario of Canada. It also happens to be the most populous city in Canada according to Wikipedia. The city has a population of approximately 3.0 million. Generally, it is also considered to be the financial capital of Canada. In terms of boroughs, Toronto has 11 boroughs. Details will be provided in later sections.

Data Acquisition and Preparation

This part of the report will go over our process of acquiring, processing, and preparing datasets for further analysis. The overall process of acquiring appropriate data would be that we start with clear list for all boroughs, or municipalities for both cities, and once we have that, we can acquire detailed restaurant venues data through API. Our sources is from the IBM lab, along with Wikipedia for web scraping, and Foursquare for venues data. Therefore, our data can be segmented into two broad categories,

Location Data

This is the first step of data preparation, we started with sets of data that contain the names of neighborhoods, boroughs, and in the case of Toronto, we also have postal codes. Lists like this can be easily find from government website, or wikipedia. After that, we can get the coordinates for different neighborhood, because having latitude and longitudes is much more accurate than having just a name of the neighborhood when it comes to exploring venues.

Venues Data

We used Foursquare for our venues data source, this is a local search-and-discovery platform, which allows us access to venues that have real user-generated feedbacks. We also used the API to acquire only restaurant related information by singling out the categoryID for food category.

Methodology

New York City

The data for NYC was acquired using the json function in python. The original json file can be found alongside the report in the GitHub repository. Once the data is loaded, we proceeded by put into a dataframe, and merged with the provided coordinates from lab. Once this process is done, we are ready to further analysis with the following showing the dataframe.

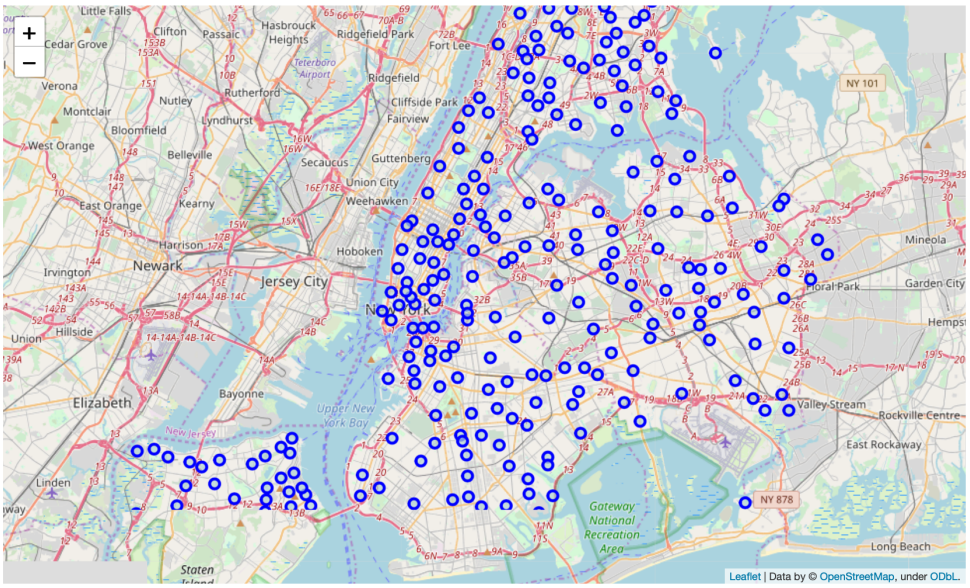
```
In [3]: # load NYC data

        with open('nyc_data.json') as json_data:
            nyc_data = json.load(json_data)

Out[3]: {'type': 'FeatureCollection',
         'totalFeatures': 306,
         'features': [{'type': 'Feature',
           'id': 'nyu_2451_34572.1',
```

	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

Since we have the coordinates needed for different neighborhood, we can easily plot out the neighborhood map as shown below, each dot on the map represents a neighborhood. In total, NYC has 5 boroughs, and 306 neighborhood.



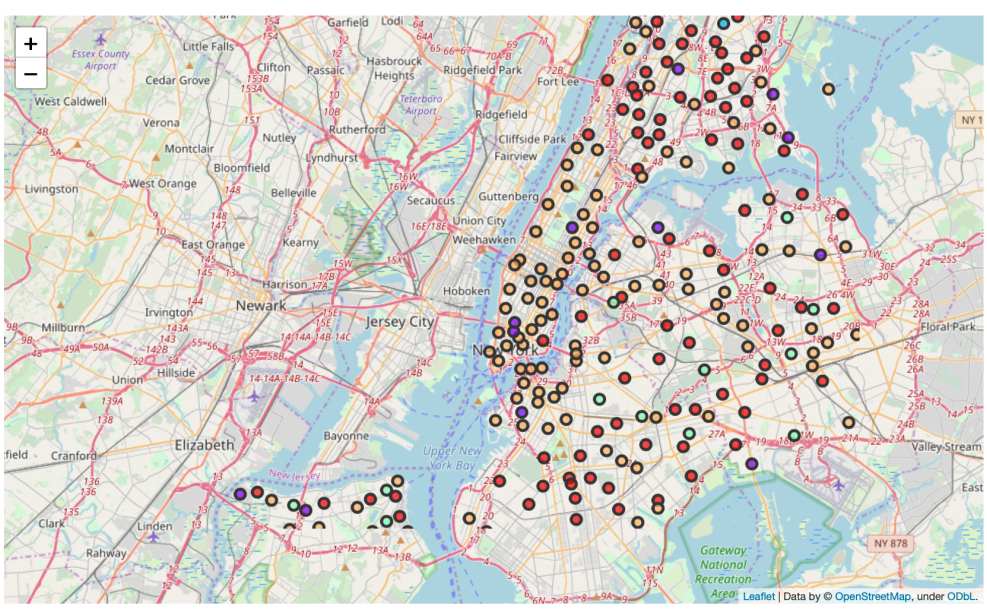
After we have successfully produced a map, we proceeded to connect with Foursquare using API. As mentioned in the introduction, we are only interested in restaurants or food-related venues. After a series of research, we were able to locate the unique identifier for venues that we are mostly interested, by doing this, it saves us a tremendous amount of manual work, having to parse out venues or establishments such as parks, residential buildings, schools, etc that are deemed irrelevant to our analysis, and may or may not interfere with our result. Below shows an image of our returned dataframe, we have the neighborhood name and coordinates, along with all the venues names, location and their categories. Boroughs are not shown here but will be included further on in the process, and Venue Category is what we are mostly interested in. Because individual shop name can be too much information for us to get a good idea of the competition, and sometime, the name is not a good predictor of its product.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
1	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898083	-73.850259	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	SUBWAY	40.890468	-73.849152	Sandwich Place
3	Wakefield	40.894705	-73.847201	Pitman Deli	40.894149	-73.845748	Food
4	Wakefield	40.894705	-73.847201	Central Deli	40.896728	-73.844387	Deli / Bodega

We used one-hot encoding to change the above dataframe, by turning the Venue Category into individual variables, this will help us with the clustering that we are going to perform later on. We then normalize the data to get a new dataframe. This process also allowed us to easily generate the Top 10 Venues by neighborhood as shown below by using a custom sort function. The result table is shown as below,

	Neighborhood	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category	8th Most Common Category	9th Most Common Category	10th Most Common Category
0	Allerton	Pizza Place	Deli / Bodega	Bakery	Chinese Restaurant	Food	Fried Chicken Joint	Mexican Restaurant	Breakfast Spot	Fast Food Restaurant	Donut Shop
1	Annadale	Pizza Place	American Restaurant	Bakery	Restaurant	Sushi Restaurant	Food	Deli / Bodega	Diner	Dumpling Restaurant	Eastern European Restaurant
2	Arden Heights	Pizza Place	Deli / Bodega	English Restaurant	Food	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant	Wings Joint
3	Arlington	Deli / Bodega	American Restaurant	Food	Fast Food Restaurant	Wings Joint	Ethiopian Restaurant	Food Court	Fish & Chips Shop	Filipino Restaurant	Falafel Restaurant
4	Arrochar	Italian Restaurant	Bagel Shop	Pizza Place	Deli / Bodega	Middle Eastern Restaurant	Taco Place	Sandwich Place	Food Truck	Mediterranean Restaurant	Polish Restaurant

In terms of clustering, we set the cluster to 5, and proceed on with the kMeans-cluster, the result automatically separated all venues into 5 clusters, and we produced a map to show different ones. Notice, the different color dots on the map, each color represents a cluster. We can easily see that red and yellow dots are the majority here in NYC. We will go over the results in a later chapter.

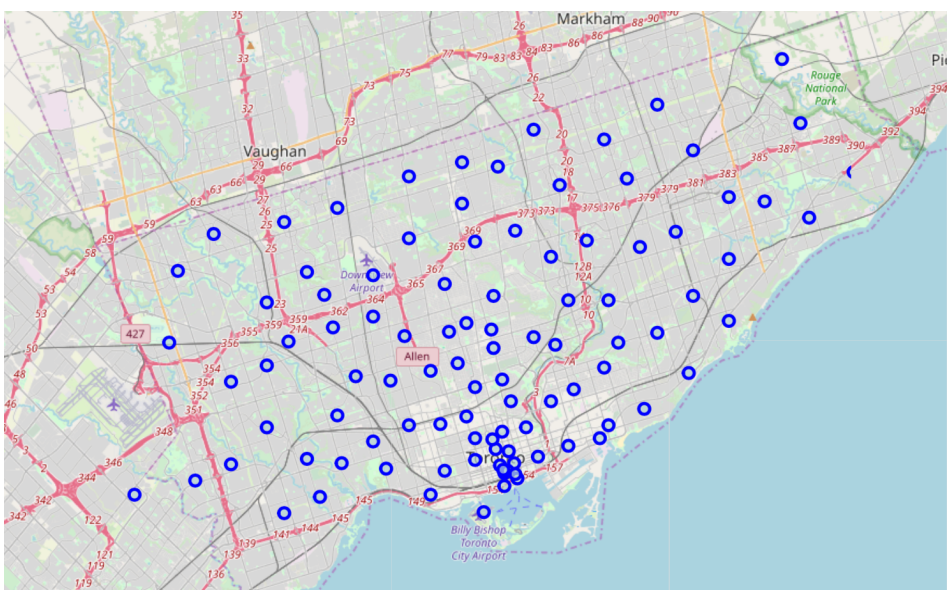


Toronto City

The Toronto City analysis was conducted following a similar process as the NYC. We started with the set of data that we scraped from Wikipedia using the BeautifulSoup package. This is the link that we used to acquired data: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. It is a list consist of all the postal codes and neighborhood names for Toronto. There were some parts of Toronto that has a postal code, but do not have an associated borough or neighborhood. We cleaned our data, and merged with another provided dataset that contains all postal code and their coordinates. We simply join the two dataset and below shows a preview of our processed dataframe. In total, Toronto City has 10 boroughs and 210 neighborhoods.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Port Union, Rouge Hill, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, West Hill, Morningside	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

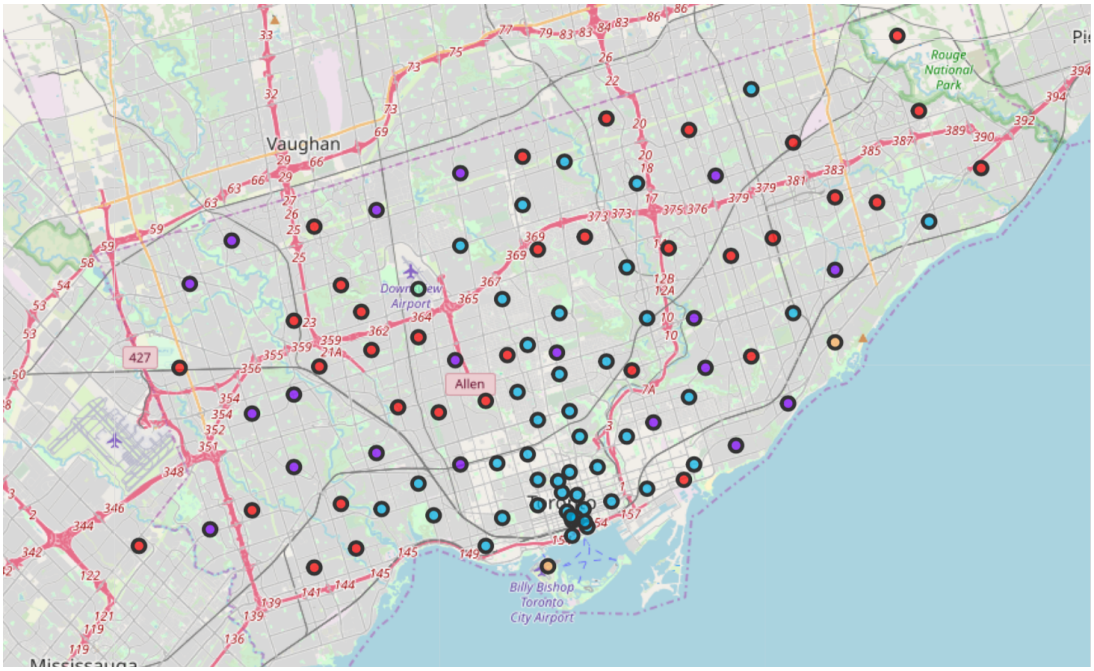
The process for Toronto City analysis follows the same pattern as the NYC, therefore we will not be elaborating on the details once again. First, we have here a map showing all the neighborhoods in Toronto, each dot represents a neighborhood.



We then connected to Foursquare and acquired all the food venues data that we need for analysis. The list contains all food venues names, location and their respected category.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge, Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge, Malvern	43.806686	-79.194353	Meena's Fine Foods	43.804476	-79.199753	Indian Restaurant
2	Guildwood, West Hill, Morningside	43.763573	-79.188711	Doug's Bakery	43.762400	-79.192285	Bakery
3	Guildwood, West Hill, Morningside	43.763573	-79.188711	Sail Sushi	43.765951	-79.191275	Restaurant
4	Guildwood, West Hill, Morningside	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant

One-hot-encoding was once again utilized to help with employing the clustering techniques. We chose to use 5 clusters, and the below map shows the result that we achieved from the kMeans cluster, we can see that red and blue dots are majority here. Once this step is done. We can proceed with our comparative analysis and results.

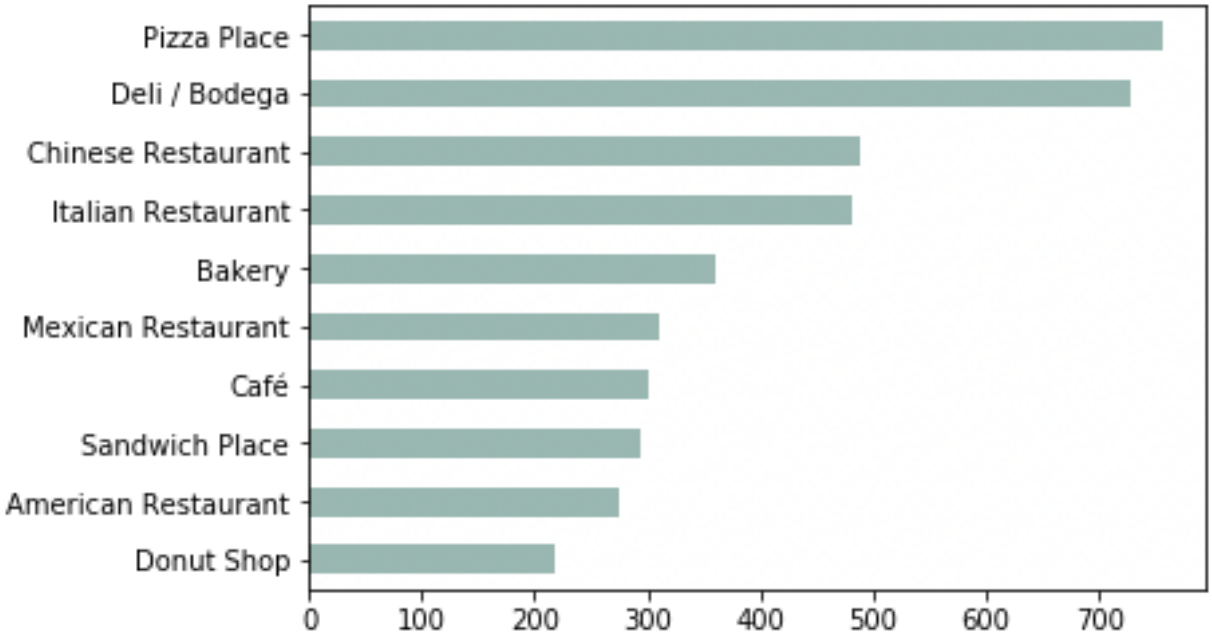


Results

Since our goal is to pick a location to open the dessert shop, we want to be able to understand the general consumer preference by analyzing current restaurant and food venues distribution. We want to know what are the top venues for both cities. We should also find out the difference amongst all of our clusters, and why the algorithm decided to group them together,

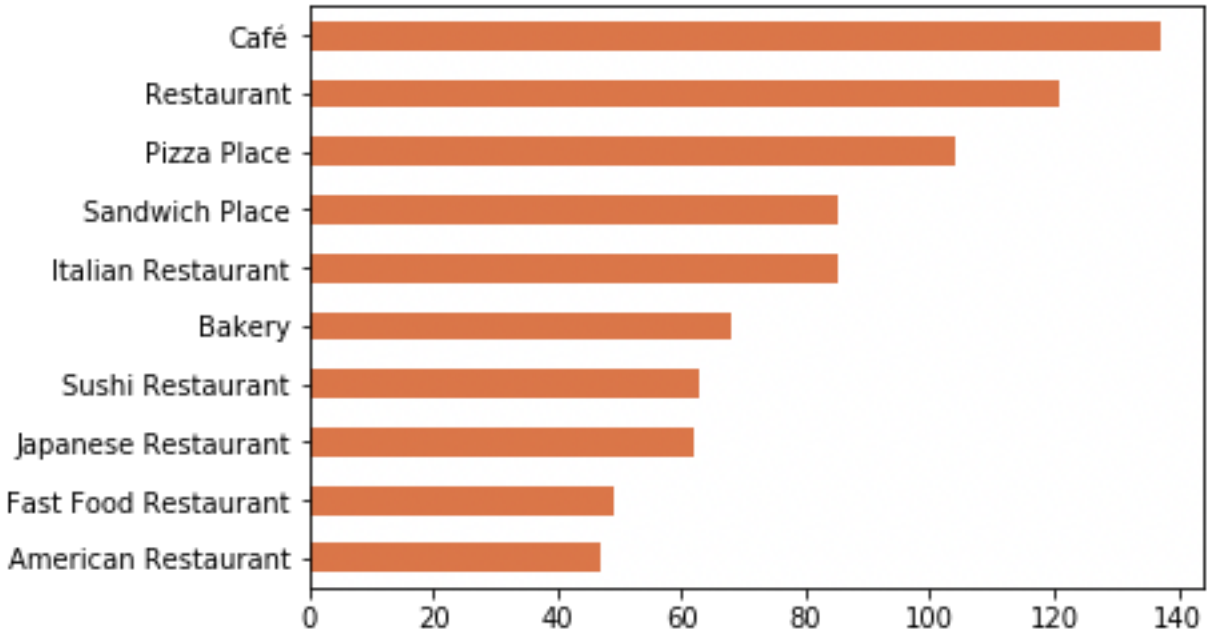
and study those big clusters since they represent significant similarity. Dessert shop is considered a niche product, it is trendy by nature, and should survive in a city that has a greater variety of cuisine since that would translate to better audience acceptance.

Here we have the top 10 venues for NYC, we can see clearly that Pizza place is the top venues here. Indeed, for anyone that has traveled to NYC before, there is practically pizza at every block. As we look through the list, trends staring to emerge. pizza, deli/bodega, bakery, cafe, sandwiches and donut shop are all fast food, this is 60% of just the top ten venues. Amongst these options, over 1400 venues are Pizza and Deli places. The other four options: Chinese, Italian, Mexican and American are not fast food, but they are not consider to be ethnic or specialty food either. Interestingly, American Restaurants are ranked 9th place of top venues in NYC. It is safe for us to concluded that although NYC is known for its fantastic selections of cuisines, ethinic foods, and general overwhelming food scene, the data simply shows that fast food and common cuisines are more popular here.



Let's look at Toronto for a moment, at first glance, we can see some difference here, Cafe is in fact the top venues. Pizza, Sandwich, Bakery, and fast food, these had also show up on the list, but represents a lesser portion of the total venues. We also noticed that Toronto appreciates Italian and American Restaurant, but not so much the Chinese restaurant despite the fact that Toronto has a large presence of Chinese immigrants. Meanwhile, an interesting

observation is the clear preference of Japanese cuisine in Toronto. Both Sushi restaurants and Japanese restaurants are Japanese cuisine. Based on this list, we can have a preliminary assumption that Toronto has a better acceptance for food variety.



Cluster Analysis

Here we show a series of comparison of clusters between New York City and Toronto. We will be showing a few examples of noted cluster for comparison purpose. Three images are shown here, and representing the top most common food venues, 2nd most popular venues, and lastly, the boroughs that are included in the cluster.

NYC Cluster 0

This cluster is mainly consist of Deli and Pizza places, and Chinese food. Most occurrences are from Queens, Bronx and Brooklyn

NYC Cluster 1

This cluster is mainly consist of Italian restaurants, and staten island has the most occurrences here.

NYC Cluster 2

This cluster is a Caribbean restaurant cluster. And Queens and Bronx are the included boroughs here.

NYC Cluster 3

Once again, Deli/Bodega takes the first place in this cluster, however, for the second most popular venue, we have some Ethiopian restaurants mixed in, but not significant enough to dominate the cluster. Queens, Staten Island, Brooklyn and Bronx are included here.

NYC Cluster 4

The last cluster once again consist mainly of Deli/Bodega, Italian Restaurant, Pizza Place. This time, Manhattan appears to be the mainly influenced boroughs.

Toronto Cluster 0

First cluster appears to be a Fast food restaurant cluster, there are some ethnic cuisine mixed in the cluster. Noticeably, we have Indian, Middle Eastern, Filipino, Hakka, and Vietnamese cuisine. Borough affected here is mostly North York.

Toronto Cluster 1

This is the Pizza cluster, clearly winner here dominating at the first most common venue category. And Etobicoke, and North York appear to be mostly included in this cluster.

Toronto Cluster 2

This is a category for Cafe and Asian Restaurants cluster. Mostly Downtown Toronto is shown in this cluster.

Toronto Cluster 3

Interestingly, this is a cluster consist of venue called Snack Place. Interesting but doesn't really tell us many stories here.

Toronto Cluster 0

This is American Restaurant cluster, and Downtown Toronto is the borough influenced here.

Discussion and Conclusion

So after conducting a series of investigation and data analysis, we noticed that New York City and Toronto has a great difference in terms of palette preference. New York City has a great affinity for pizzas, fast foods and street foods. While Toronto has a great preference for cafes, and variety of cuisines. If deeper analysis were to conduct, we would like to bring in demographic informations, and do a further clustering. As of now, despite our initial belief of New York being the top choice and claiming to be a food capital; we believe that the client's dessert shop should open in Toronto which show a better acceptance of niche cuisines.