Shake Shack Location Prediction

1. Introduction

Shake Shack is a burger brand established in New York in 2004 is now very popular. Not only famous in its domestic market, but It also has overseas branches in Asia, Middle East, and Britain. However, there is no branches in their nearest neighbor, Canada. Obviously, determine a suitable location is the crucial problem for fast food operations. This project aimed at comparing New York and Toronto and look for a suitable area for Shake Shack to open a new branch.

Geolocation data and venue surrounding data is the most valuable data source to predict which location in Toronto is the most similar to already existed Shake Shack locations. Obviously, Toronto is the equivalent of New York City. Stakeholders would be very interested in the best location to open new branches if they wish to go to Canada.

2. Data acquisition and cleaning

2.1 Data sources

Geolocation data and venue data are used in this project. To get 'Shake Shack in New York' from Foursquare API, New York City coordinate data are obtained first. And Toronto's borough data are acquired from Wikipedia, and their nearby venue are obtained from Foursquare API.

2.2 Data cleaning

New York coordinate data are first obtained through geopy module. Then 'Shake Shack' Data downloaded from Foursquare API are cleaned and filtered by its city name, 'New York City'. Toronto borough data are acquired from Wikipedia and cleaned by beautiful soup module. Nearby venue data are obtained by Foursquare API and grouped by the city name.

Table 1. Shake Shack location filtered by city name

	name	categories	address	cc	city	country	cross Street	distance	formattedAddress	labeledLatLngs	lat	lng	neighborhood	post
0	Shake Shack	Burger Joint	200 Broadway	US	New York	United States	at Fulton St	339	[200 Broadway (at Fulton St), New York, NY 10038]	[{'label': 'display', 'lat': 40.71070286896417...	40.710703	-74.009024	NaN	
1	Shake Shack	Burger Joint	215 Murray St	US	New York	United States	btwn West St & North End Ave	791	[215 Murray St (btwn West St & North End Ave),...	[{'label': 'display', 'lat': 40.71532572032048...	40.715326	-74.014750	NaN	
3	Shake Shack	Burger Joint	E 23rd St and Madison Avenue, Madison Square Park	US	New York	United States	Madison Ave & E 23rd St	3535	[E 23rd St and Madison Avenue, Madison Square ...	[{'label': 'display', 'lat': 40.74148371088094...	40.741484	-73.988218	Flatiron	
6	Shake Shack	Burger Joint	691 8th Ave	US	New York	United States	at W 44th St	5276	[691 8th Ave (at W 44th St), New York, NY 10036]	[{'label': 'display', 'lat': 40.75833417228198...	40.758334	-73.988988	Theater District	
7	Shake Shack	Burger Joint	1700 Broadway	US	New York	United States	West 53rd Street	5991	[1700 Broadway (West 53rd Street), New York, N...	[{'label': 'display', 'lat': 40.763495, 'lng':...	40.763495	-73.982419	NaN	
8	Shake Shack	Burger Joint	152 E 86th St	US	New York	United States	btwn Lexington & 3rd Ave	8563	[152 E 86th St (btwn Lexington & 3rd Ave), New...	[{'label': 'display', 'lat': 40.77909852079875...	40.779099	-73.954686	NaN	
9	Shake Shack	Burger Joint	366 Columbus Ave	US	New York	United States	at W 77th St	7964	[366 Columbus Ave (at W 77th St), New York, NY...	[{'label': 'display', 'lat': 40.78067635406833...	40.780676	-73.976443	NaN	

Table 2. Shake Shack in New York viewed on map
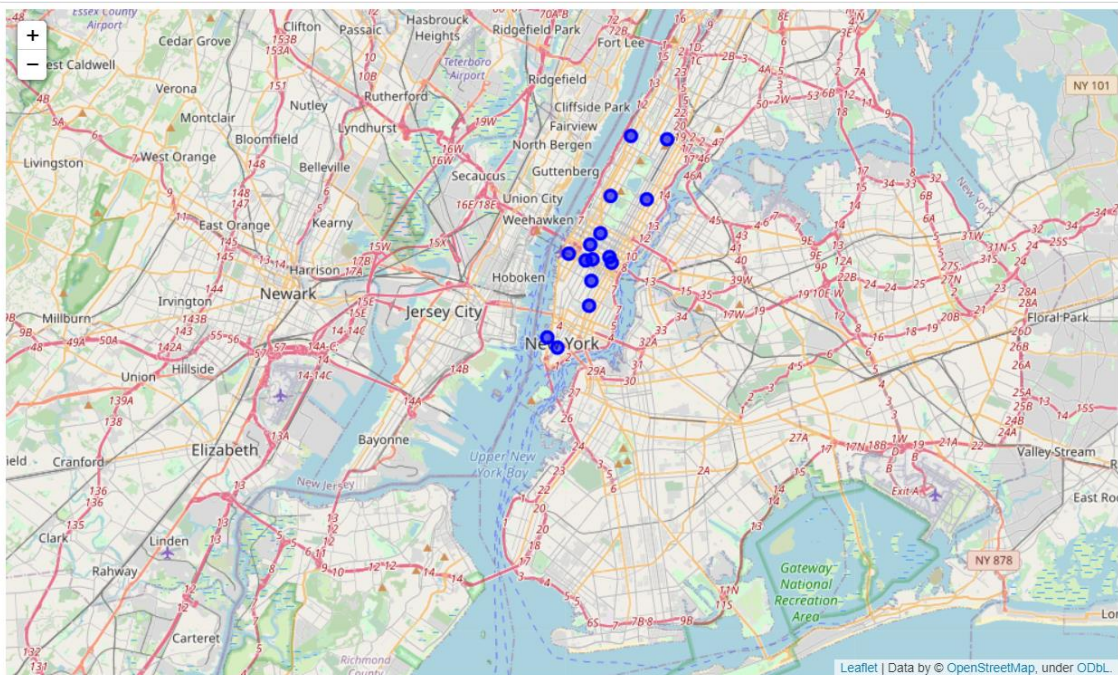
3. Venue data collecting, grouping, and encoding

3.1 Venue data collecting and grouping

By using Foursquare API, Shake Shack in New York are grouped together and renamed to Shake Shack. And their nearby venue's categories are obtained. Similarly, venue category data in Toronto are obtained and grouped by the borough name.

Table 3. Shake Shack nearby venues

	Name	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Shake Shack	40.710703	-74.009024	The Bar Room at Temple Court	40.711448	-74.006802	Hotel Bar
1	Shake Shack	40.710703	-74.009024	Pisillo Italian Panini	40.710530	-74.007526	Sandwich Place
2	Shake Shack	40.710703	-74.009024	The Beekman, A Thompson Hotel	40.711173	-74.006702	Hotel
3	Shake Shack	40.710703	-74.009024	For Five Coffee Roasters	40.709554	-74.010576	Coffee Shop
4	Shake Shack	40.710703	-74.009024	Joe's Pizza	40.710178	-74.007769	Pizza Place

3.2 one-hot encoding

One-hot encoding method is useful to covert categorical types into numerical data, which can be further processed by machine learning algorithms. In this project, venue category data are processed through one-hot encoding and grouped by the location names. There are 15 groups in the dataset, 14 of them are Toronto boroughs and 1 is all Shake Shack in New York grouped together. Also, the mean of one-hot encoded venue category data can be used to calculate distance between Shake Shack and boroughs in Toronto. Which can represent the similarity to Shake Shack.

Table 4. One-hot encoded venues in each group

	Name	Accessories Store	Adult Boutique	African Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Vegetarian / Vegan Restaurant	Video Game Store	Video Store	Vietnamese Restaurant
0	Central Toronto	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.009009	...	0.000000	0.000000	0.000000	0.009009
1	Downtown Toronto	0.000000	0.00092	0.00000	0.000920	0.00092	0.00184	0.00276	0.00092	0.011960	...	0.011040	0.001840	0.000000	0.003680
2	Downtown Toronto Stn A	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.000000	...	0.010309	0.000000	0.000000	0.000000
3	East Toronto	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.029126	...	0.000000	0.000000	0.000000	0.000000
4	East Toronto Business	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.000000	...	0.000000	0.000000	0.000000	0.000000
5	East York	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.000000	...	0.000000	0.000000	0.014085	0.000000
6	East York/East Toronto	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.000000	...	0.000000	0.000000	0.000000	0.000000
7	Etobicoke	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.015625	...	0.000000	0.000000	0.000000	0.000000
8	Etobicoke Northwest	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.000000	...	0.000000	0.000000	0.000000	0.000000
9	Mississauga	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.071429	...	0.000000	0.000000	0.000000	0.000000
10	North York	0.004184	0.00000	0.00000	0.004184	0.00000	0.00000	0.00000	0.00000	0.008368	...	0.000000	0.004184	0.000000	0.008368
11	Queen's Park	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.000000	...	0.000000	0.000000	0.000000	0.000000
12	Scarborough	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.010309	...	0.000000	0.000000	0.000000	0.010309
13	Shake Shack	0.000740	0.00000	0.00148	0.000000	0.00000	0.00000	0.00000	0.00000	0.026647	...	0.005922	0.000740	0.000000	0.005181
14	West Toronto	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.000000	...	0.018987	0.000000	0.000000	0.012658
15	York	0.000000	0.00000	0.00000	0.000000	0.00000	0.00000	0.00000	0.00000	0.000000	...	0.000000	0.000000	0.000000	0.000000

4. KMeans Modeling and Visualization

Kmeans algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping clusters where each data point belongs to only one group. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid is at the

minimum. In this sense, Toronto borough and Shake Shack grouped together into a pool and clustered by KMeans. Similar locations will assign into the same group and thus we know which location is the most suitable for a new branch.

Table 5. Cluster labels for each group

Out[15]:

	Name	Cluster Labels	Latitude	Longitude
0	Central Toronto	13	43.698631	-79.397703
1	Downtown Toronto	6	43.652234	-79.381877
2	Downtown Toronto Stn A	6	43.646435	-79.374846
3	East Toronto	9	43.670279	-79.338488
4	East Toronto Business	5	43.662744	-79.321558
5	East York	10	43.706382	-79.347744
6	East York/East Toronto	0	43.685347	-79.338106
7	Etobicoke	8	43.652527	-79.539176
8	Etobicoke Northwest	2	43.706748	-79.594054
9	Mississauga	4	43.636966	-79.615819
10	North York	1	43.755758	-79.401845
11	Queen's Park	7	43.662302	-79.389494
12	Scarborough	11	43.762253	-79.267948
13	Shake Shack	1	40.753427	-73.984470
14	West Toronto	12	43.651391	-79.449396
15	York	3	43.691853	-79.470383

Shake Shack in New York are purple labeled, so that if we zoom into Toronto, we notice that the North York is also purple labeled. Thus, the North York is the most similar to Shake Shack branch locations.

Table 6. Cluster visualized on map.
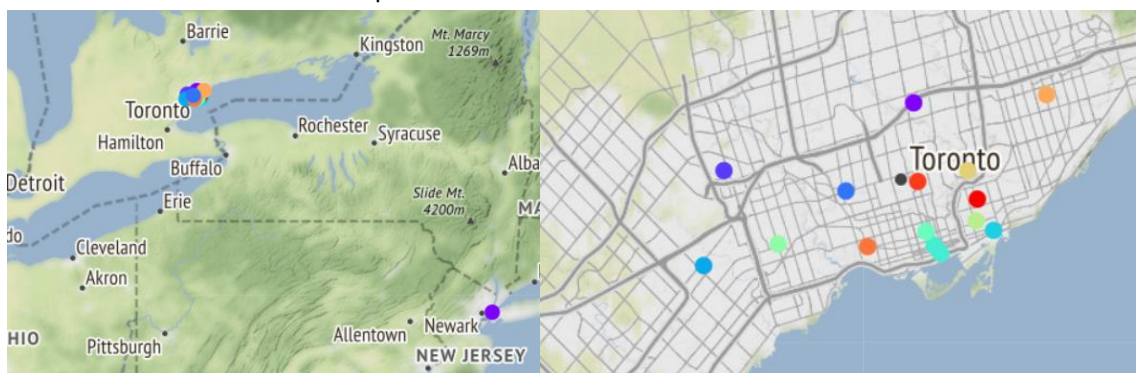
5. Future directions

5.1 Sales weighted Shake Shack group

Shake Shack in New York is arbitrarily grouped together in this project. However, not all branches are performed well. Thus, sales weighted is more reasonable by give more weight to outperformed branch locations.

5.2 drill into detailed locations

Since we know that North York is the most similar to Shake Shack locations in New York, we can further drill into the communities and blocks in North York by using the same method. By iterations, we can finally find the most similar location of desired levels.