



SEPTEMBER 2, 2018

THE BATTLE OF NEIGHBORHOODS REPORT

COURSERA – CAPSTONE PROJECT

TIMBAKE DANG



TABLE OF CONTENTS

BUSINESS PROBLEM	3
Executive Summary	3
DATA - STATISTICS & REPORTS	4
Top 20 Venues Recommended By Category in State/city, neighborhoods	4
Top 10 venues have the most interesting (likes) by category in STATE/ citiES, neighborhoods	7
Top 5 venues have the highest rating by citiES, neighborhoods.....	10
CONCLUSION	14
Data Sets Was Gathered & Backup	15
SOURCE CODE – CAPSTONE PROJECT.....	16

BUSINESS PROBLEM

EXECUTIVE SUMMARY

There are many existing venues in neighborhoods in New York state as well as Toronto city. How can we know which venue by category and city have the most interesting from people as well as have the most number of users checked-in? Then how can we know the competition of the venues of the same category in a city or between neighborhoods in a city bases on the statistic number of users, visits, checked-ins, likes, rating?

Through by answering this questions, we can have insight about the competition between venues of the same category in a state, a city or in a neighborhood, then we can find out the reasons could make these venues to be attractive to the people.

Finally, after we know reasons, and have the accurate statistic data about the existing competitors, the behavior of users, the user culture, the demands are not served, or are not served well, the needs in future, we will develop strategies for changes in the market, sustainable growth as well as expansion business to new market.

DATA - STATISTICS & REPORTS

The data used in these statistics and reports was resources retrieved from:

- Toronto data: Wikipedia (Toronto postal code) combined with neighborhoods, venues data by using Foursquare APIs.
- New York data: included neighborhoods, venues data gathered from <https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json>

Before using Foursquare APIs to get neighborhoods, venues data, I used Geocoder library to get coordinates (latitude, longitude) for each neighborhood, instead of Google Geoencoding APIs.

Then I combined 2 data frames Toronto and New York together in order to perform comparison, statistic and create reports.

TOP 20 VENUES RECOMMENDED BY CATEGORY IN STATE/CITY, NEIGHBORHOODS

TABLE 1 TOP 20 CATEGORIES HAS THE MOST VENUES IN NEW YORK (STATE), TORONTO (CITY)

	New York	Toronto	Total
Pizza Place	434.0	49.0	483.0
Coffee Shop	264.0	198.0	462.0
Italian Restaurant	322.0	48.0	370.0
Deli / Bodega	274.0	18.0	292.0
Bakery	231.0	48.0	279.0
Café	164.0	103.0	267.0
Bar	223.0	41.0	264.0
Chinese Restaurant	204.0	25.0	229.0
Sandwich Place	186.0	41.0	227.0
Park	171.0	45.0	216.0
American Restaurant	176.0	31.0	207.0
Grocery Store	179.0	24.0	203.0
Mexican Restaurant	170.0	19.0	189.0
Pharmacy	169.0	18.0	187.0
Donut Shop	168.0	2.0	170.0
Gym / Fitness Center	144.0	21.0	165.0
Ice Cream Shop	141.0	20.0	161.0
Restaurant	87.0	61.0	148.0
Sushi Restaurant	113.0	30.0	143.0
Hotel	96.0	45.0	141.0

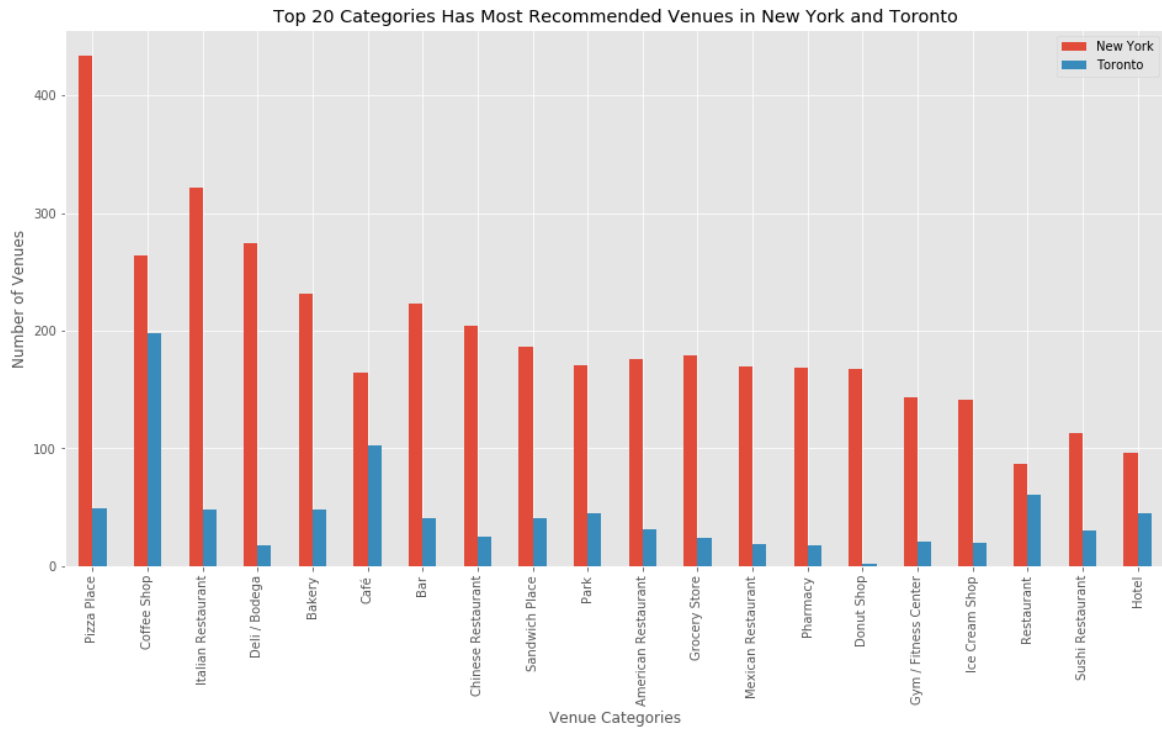
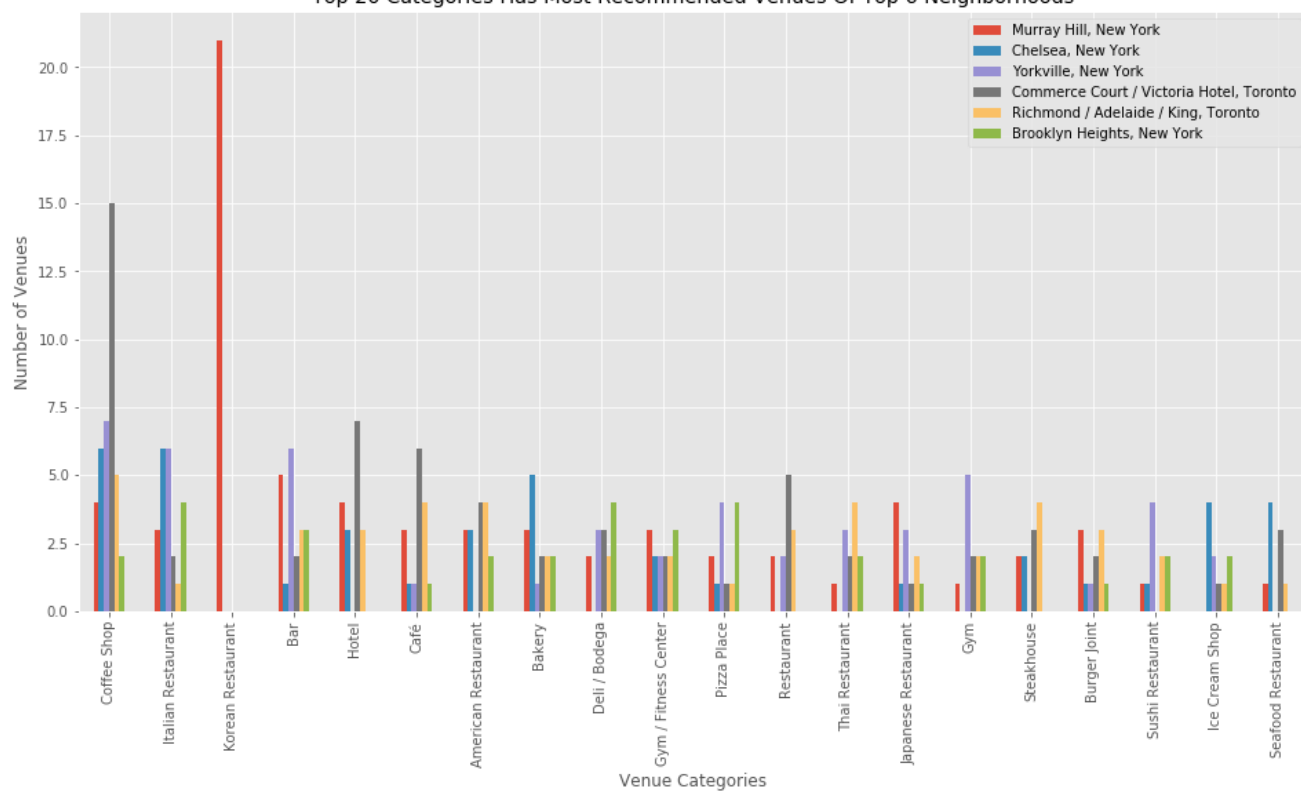


TABLE 2 TOP 20 CATEGORIES HAVE THE MOST VENUES IN 6 NEIGHBORHOODS

	Murray Hill, New York	Chelsea, New York	Yorkville, New York	Commerce Court / Victoria Hotel, Toronto	Richmond / Adelaide / King, Toronto	Brooklyn Heights, New York	Total
Coffee Shop	4	6	7	15	5	2	39
Italian Restaurant	3	6	6	2	1	4	22
Korean Restaurant	21	0	0	0	0	0	21
Bar	5	1	6	2	3	3	20
Hotel	4	3	0	7	3	0	17
Café	3	1	1	6	4	1	16
American Restaurant	3	3	0	4	4	2	16
Bakery	3	5	1	2	2	2	15
Deli / Bodega	2	0	3	3	2	4	14
Gym / Fitness Center	3	2	2	2	2	3	14
Pizza Place	2	1	4	1	1	4	13
Restaurant	2	0	2	5	3	0	12
Thai Restaurant	1	0	3	2	4	2	12
Japanese Restaurant	4	1	3	1	2	1	12
Gym	1	0	5	2	2	2	12
Steakhouse	2	2	0	3	4	0	11
Burger Joint	3	1	1	2	3	1	11
Sushi Restaurant	1	1	4	0	2	2	10
Ice Cream Shop	0	4	2	1	1	2	10
Seafood Restaurant	1	4	0	3	1	0	9

Top 20 Categories Has Most Recommended Venues Of Top 6 Neighborhoods



TOP 10 VENUES HAVE THE MOST INTERESTING (LIKES) BY CATEGORY IN STATE/ CITIES, NEIGHBORHOODS

From above data, I'm going to use FOURSQUARE API get detail information of venues such as likes, rating, the number of users, etc... However, because API endpoint Details of FOURSQUARE is PREMIUM, this endpoint limits only 50 calls per day. It certainly doesn't adapt for my above data set with 12,374 different venues in New York and Toronto. Thus, I decided to use API endpoint Likes as an alternative method to get the number of likes.

Besides, because the limitation of Foursquare Regular Accounts as well as my targeting market is New York state first, I decided to get data frame of venues by top 10 categories and in New York only. This filtering helped me to decrease the size of data set to 3916 records (venues).

In this project, I only retrieve the number of User Likes from Foursquare for first 500 venues in my data set.

TABLE 3 TOP 10 CATEGORIES HAVE THE MOST TOTAL USER LIKES IN NEW YORK STATE

	Category Name	Likes
0	Italian Restaurant	1134
1	Pizza Place	1040
2	Bar	816
3	Donut Shop	551
4	Mexican Restaurant	449
5	Deli / Bodega	376
6	Pharmacy	368
7	American Restaurant	346
8	Bakery	297
9	Park	293

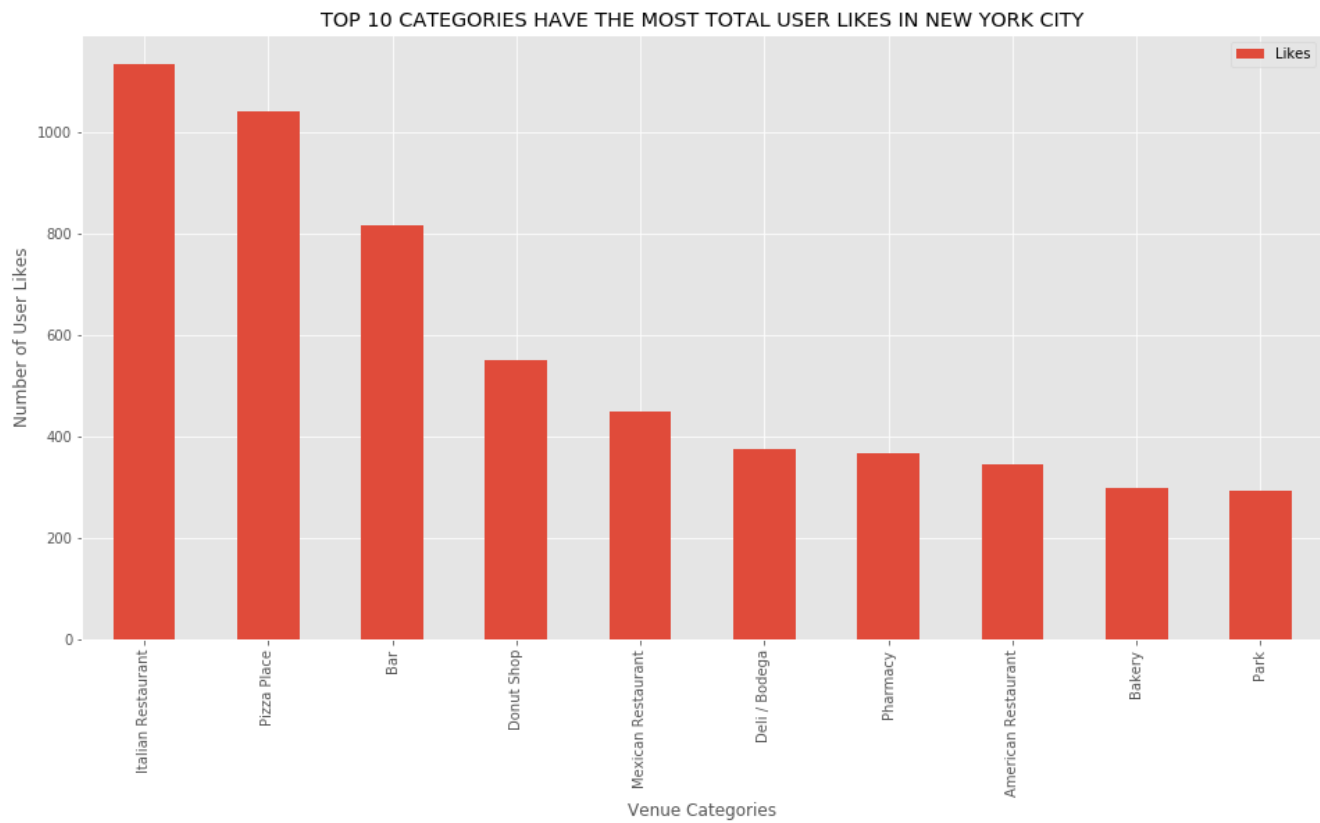
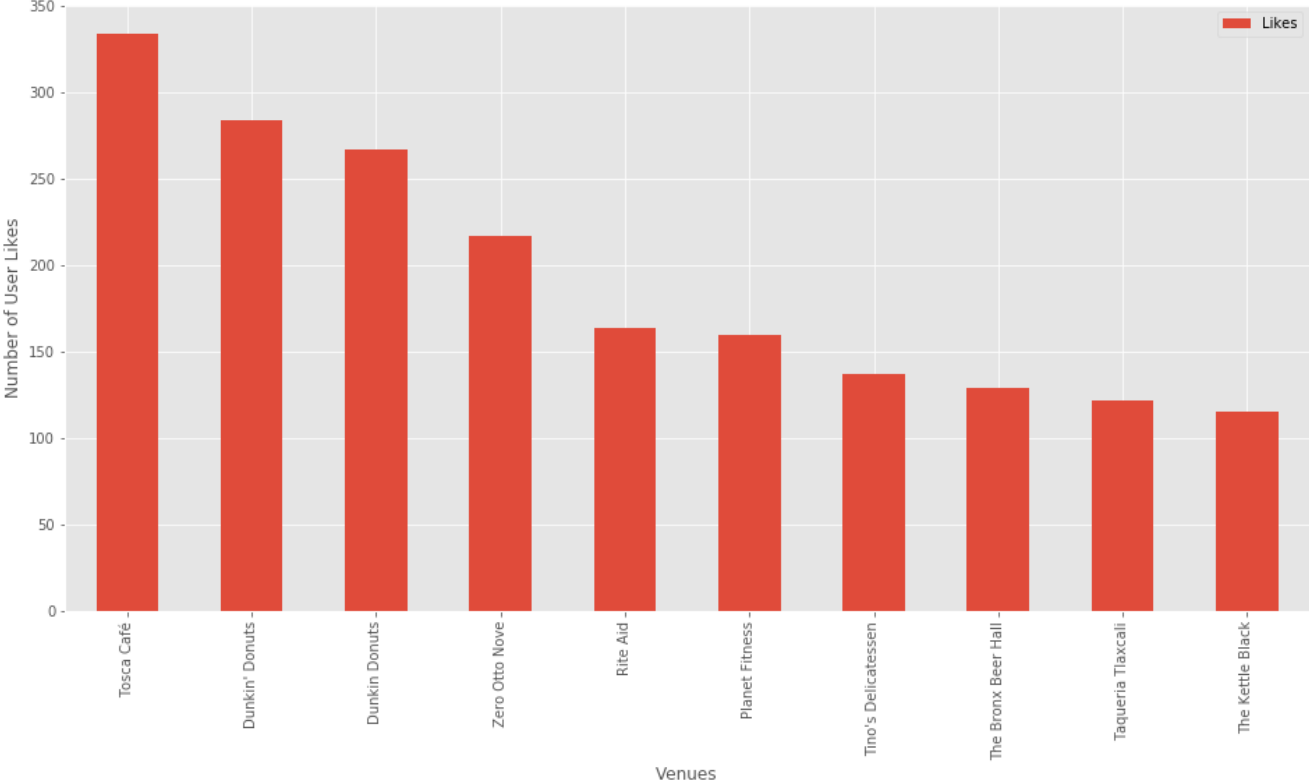


TABLE 4 TOP 10 VENUES HAVE THE MOST TOTAL USER LIKES IN NEW YORK CITY

	Venue Name	Likes
0	Tosca Café	334
1	Dunkin' Donuts	284
2	Dunkin Donuts	267
3	Zero Otto Nove	217
4	Rite Aid	164
5	Planet Fitness	160
6	Tino's Delicatessen	137
7	The Bronx Beer Hall	129
8	Taqueria Tlaxcali	122
9	The Kettle Black	115

TOP 10 VENUES HAVE THE MOST TOTAL USER LIKES IN NEW YORK CITY



TOP 5 VENUES HAVE THE HIGHEST RATING BY CITIES, NEIGHBORHOODS

TABLE 5 ORDER LIST OF VENUES HAVE HIGHEST RATING BY CITY

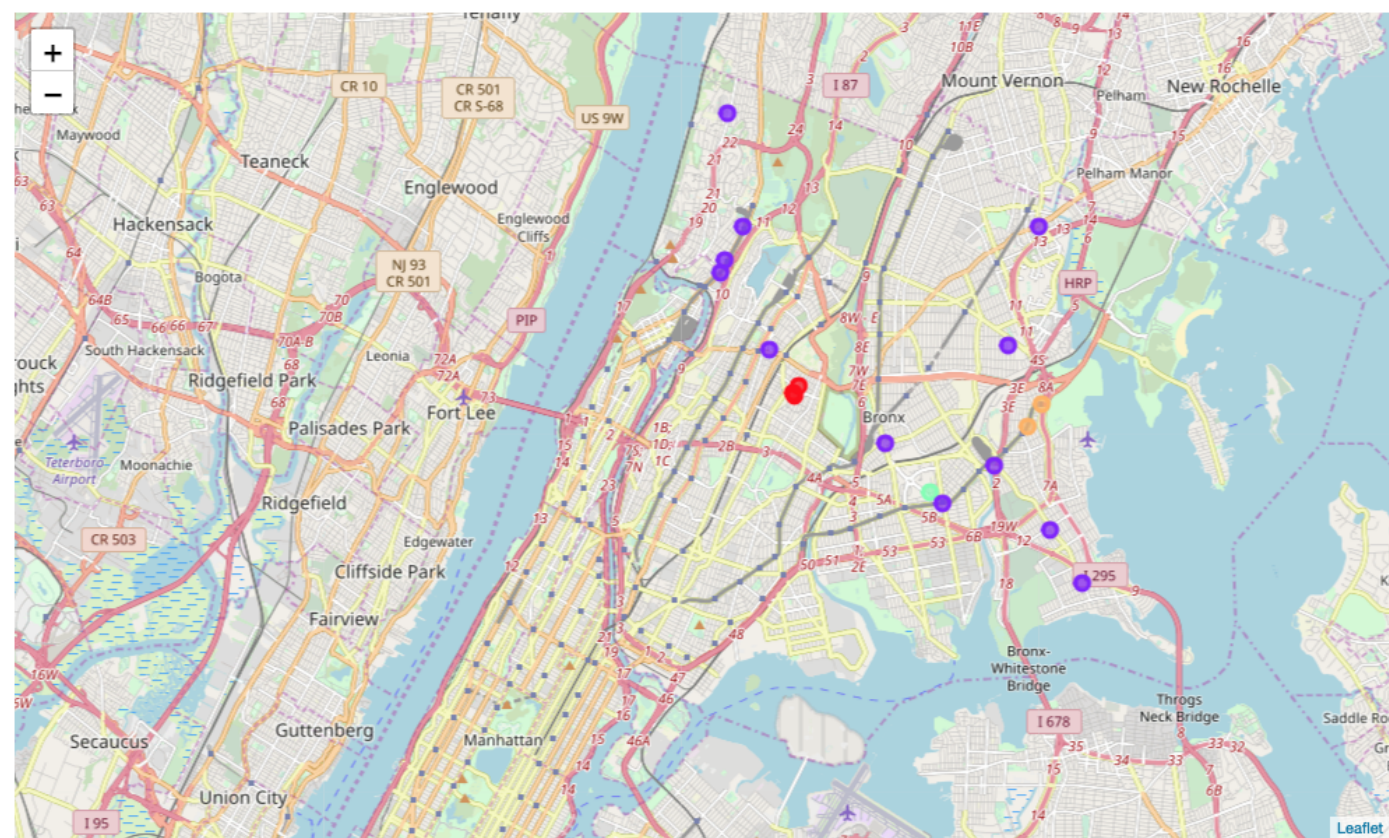
		Venue ID	Venue Latitude	Venue Longitude	Rating
City	Venue Name				
Bronx	Zero Otto Nove	4aabd3e6f964a5204a5a20e3	40.854714	-73.888388	9.4
	Tino's Delicatessen	4acf80aef964a52025d420e3	40.855882	-73.887166	9.4
	Taqueria Tlaxcali	4b4929d2f964a520656826e3	40.836098	-73.854948	9.3
	The Bronx Beer Hall	50e600a4526287052c53bb99	40.854198	-73.888233	8.6
	Tosca Café	4b4b9f1df964a5201fa226e3	40.819204	-73.817467	8.4
	Planet Fitness	4f5ce789e4b0a4baa3d5dbb6	40.863298	-73.835568	8.0
	Dunkin' Donuts	4b5357adf964a520319827e3	40.876993	-73.906507	7.5
	Dunkin' Donuts	4ba92d21f964a5209e113ae3	40.852560	-73.827359	7.5
	Rite Aid	4bafb541f964a520a4193ce3	40.829232	-73.825341	7.4
	Dunkin Donuts	4bfb0096bbb7c9284e8f0743	40.845346	-73.865942	7.3
	Planet Fitness	462d15a3f964a520e7451fe3	40.848290	-73.830875	7.3
	Dunkin Donuts	4e38a2cfbd415d9f63d71183	40.879146	-73.905260	7.3
	Dunkin' Donuts	4b3a2cecf964a520e56125e3	40.906489	-73.904592	7.1
	Dunkin Donuts	4ba3782cf964a520b43f38e3	40.885481	-73.828015	7.0
	Dunkin Donuts	4c0bc649bbc676b096fb4bd5	40.834105	-73.851683	6.9
	Dunkin' Donuts	4b647ca3f964a52030b72ae3	40.841109	-73.839253	6.6
	Rite Aid	4b82201df964a52022cb30e3	40.885481	-73.900814	6.4
	Planet Fitness	4b9c08adf964a5201c4336e3	40.862619	-73.894390	6.3
Brooklyn	The Kettle Black	4a9880f5f964a520692c20e3	40.622839	-74.031411	8.1
	Dunkin' Donuts	4b991f43f964a520d06235e3	40.621868	-74.028789	7.4

TABLE 6 ORDER LIST OF VENUES HAVE HIGHEST RATING BY NEIGHBORHOOD

		Venue ID	Venue Latitude	Venue Longitude	Rating
Neighborhood	Venue Name				
Bay Ridge - Brooklyn	The Kettle Black	4a9880f5f964a520692c20e3	40.622839	-74.031411	8.1
	Dunkin' Donuts	4b991f43f964a520d06235e3	40.621868	-74.028789	7.4
Baychester - Bronx	Planet Fitness	4f5ce789e4b0a4baa3d5dbb6	40.863298	-73.835568	8.0
Belmont - Bronx	Zero Otto Nove	4aab3e6f964a5204a5a20e3	40.854714	-73.888388	9.4
	Tino's Delicatessen	4acf80aef964a52025d420e3	40.855882	-73.887166	9.4
	The Bronx Beer Hall	50e600a4526287052c53bb99	40.854198	-73.888233	8.6
Eastchester - Bronx	Dunkin Donuts	4ba3782cf964a520b43f38e3	40.885481	-73.828015	7.0
Fordham - Bronx	Planet Fitness	4b9c08adf964a5201c4336e3	40.862619	-73.894390	6.3
Kingsbridge - Bronx	Dunkin Donuts	4e38a2cfbd415d9f63d71183	40.879146	-73.905260	7.3
	Rite Aid	4b82201df964a52022cb30e3	40.885481	-73.900814	6.4
Marble Hill - Bronx	Dunkin' Donuts	4b5357adf964a520319827e3	40.876993	-73.906507	7.5
North Riverdale - Bronx	Dunkin' Donuts	4b3a2cecf964a520e56125e3	40.906489	-73.904592	7.1
Parkchester - Bronx	Taqueria Tlaxcali	4b4929d2f964a520656826e3	40.836098	-73.854948	9.3
Pelham Bay - Bronx	Dunkin' Donuts	4ba92d21f964a5209e113ae3	40.852560	-73.827359	7.5
	Planet Fitness	462d15a3f964a520e7451fe3	40.848290	-73.830875	7.3
Schuylerville - Bronx	Rite Aid	4bafb541f964a520a4193ce3	40.829232	-73.825341	7.4
Throgs Neck - Bronx	Tosca Café	4b4b9f1df964a5201fa226e3	40.819204	-73.817467	8.4
Unionport - Bronx	Dunkin Donuts	4c0bc649bbc676b096fb4bd5	40.834105	-73.851683	6.9
Van Nest - Bronx	Dunkin Donuts	4bfb0096bbb7c9284e8f0743	40.845346	-73.865942	7.3
Westchester Square - Bronx	Dunkin' Donuts	4b647ca3f964a52030b72ae3	40.841109	-73.839253	6.6

Now, let's analyze each venue to cluster them by rating!

TABLE 7 - 5 VENUE CLUSTERS BY RATING



Belmont Cluster

	City	Neighborhood	Venue Name	Venue Address	Category Name	Rating	Likes	Users	Visits
0	Bronx	Belmont	Zero Otto Nove	2357 Arthur Ave (at E 186th St)	Pizza Place	9.4	217	0	0
2	Bronx	Belmont	Tino's Delicatessen	2410 Arthur Ave (E. 187 St.)	Italian Restaurant	9.4	137	0	0
3	Bronx	Belmont	The Bronx Beer Hall	2344 Arthur Ave (btwn E186th St & E 184th St)	Bar	8.6	129	0	0

Venues in separate neighborhoods in Bronx

	City	Neighborhood	Venue Name	Venue Address	Category Name	Rating	Likes	Users	Visits
1	Bronx	Throgs Neck	Tosca Café	4034 E Tremont Ave (at Miles Ave)	Bar	8.4	167	0	0
6	Bronx	Baychester	Planet Fitness	1780 E Gun Hill Rd	Gym / Fitness Center	8.0	76	0	0
7	Bronx	Marble Hill	Dunkin' Donuts	5501 Broadway (W 230th St)	Donut Shop	7.5	65	0	0
10	Bronx	Fordham	Planet Fitness	309 E Fordham Rd	Gym / Fitness Center	6.3	35	0	0
11	Bronx	Westchester Square	Dunkin' Donuts	2973 E Tremont Ave	Donut Shop	6.6	34	0	0
12	Bronx	Eastchester	Dunkin Donuts	4090 Boston Rd (Conner Street)	Donut Shop	7.0	32	0	0
13	Bronx	Van Nest	Dunkin Donuts	699 Morris Park Ave (White Plains Road)	Donut Shop	7.3	30	0	0
14	Bronx	Kingsbridge	Dunkin Donuts	209 W 231st St	Donut Shop	7.3	29	0	0
15	Bronx	Unionport	Dunkin Donuts	2172 Westchester Ave	Donut Shop	6.9	26	0	0
17	Bronx	Kingsbridge	Rite Aid	5825-35 Broadway (at 238th St)	Pharmacy	6.4	24	0	0
18	Bronx	Schuylerville	Rite Aid	3590 E Tremont Ave	Pharmacy	7.4	22	0	0
19	Bronx	North Riverdale	Dunkin' Donuts	508 259th St (Riverdale Ave and 259th St) (Riv...	Donut Shop	7.1	20	0	0

Although Dunkin's Donuts is dominating market in group Donut Shop, it only has rating at good level with 7.5 maximum, and have 65 likes.

In group Gym / Fitness Center, there is only Planet Fitness in 2 neighborhoods. But these is a huge gap about rating and user likes between Baychester and Fordham.

Bay Ridge Cluster

We still see only Dunkin' Donuts in group Donut Shop, but it's rating is at good level, and the number of likes is low.

	City	Neighborhood	Venue Name	Venue Address	Category Name	Rating	Likes	Users	Visits
5	Brooklyn	Bay Ridge	The Kettle Black	8622 3rd Ave (at 87th St.)	Bar	8.1	115	0	0
16	Brooklyn	Bay Ridge	Dunkin' Donuts	8704 4th Ave	Donut Shop	7.4	24	0	0

Pelham Bay Cluster

	City	Neighborhood	Venue Name	Venue Address	Category Name	Rating	Likes	Users	Visits
8	Bronx	Pelham Bay	Dunkin' Donuts	3276 Westchester Ave	Donut Shop	7.5	51	0	0
9	Bronx	Pelham Bay	Planet Fitness	3060 Westchester Ave	Gym / Fitness Center	7.3	49	0	0

CONCLUSION

At the beginning of study, I'm going to perform market researching by gathering data, analyzing data, statistic, clustering, and visualizing data so that I can make decision to choose the specific business what I'm interested in.

However, because Foursquare regular account has limit while there is no any clear, specified business, the information that I got is not enough to have the good overall picture about the market.

Thus, I need to find out and focus on a specific business that I'm really love such as organic foods imported from ASEAN, ASEAN restaurant. Then, I'll perform researching again using this methodology.

Base on the information that I have from researching, I will produce the strategies to develop New York market.

<https://github.com/timbake/TheBattleOfNeighborhoods>

- NewYorkRVenues.csv: was from IBM
<https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json> combined with venues info retrieved from Foursquare using API endpoint Explore.
- TorontoRVenues.csv: was scrapped coordinates form Wikipedia
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M , then also combined with venues info retrieved Foursquare.
- RecommendedVenues.csv: was merged from 2 data set dfNewYorkRVenues (NewYorkRVenues.csv) and dfTorontoRVenues (TorontoRVenues.csv).
- 500NewYorkFilteredRVenues.csv
- 20NewYorkRVenuesDetail.csv

SOURCE CODE – CAPSTONE PROJECT

https://github.com/timbake/data-science-capstone-coursera/blob/master/The_Battle_Of_Neighborhoods.ipynb