

Capstone Project: Battle of the Neighborhoods
IBM Data Science Specialization – www.coursera.org

Comparing McDonald's Branches in Manhattan and Toronto by Consumer Feedback on FourSquare API

Prepared by Mohamad Shammaa

August 2019

Table of Contents

	Page
I. INTRODUCTION	3
II. DATA	4
III. METHODOLOGY	5
IV. RESULT	6
V. Discussion	9
VI. CONCLUSION	10
VII. References	11
VIII. Appendix	11

I. INTRODUCTION

McDonald's Corporation is an American fast food company, founded in 1940. It is the world's largest restaurant chain by revenue, serving over 69 million customers daily in over 100 countries across 37,855 outlets as of 2018.¹



Different factors, as cultural background of costumers, location and way of dealing with customers as well as flavors and diversity of meals between branches might contribute positively or negatively to the formation of the final consumer opinions about a particular branch of the chain.

In my study I will focus on two large cities with different cultural background namely Toronto and Manhattan in New York.

Toronto (Population: 2.93 million 2017²), is a business-minded, conscientious, socially progressive, and pluralistic city. It is a city of many museums, theatres, festival events and sports activities. Toronto's cosmopolitan atmosphere is reflected in its cuisine, with many world cooking styles represented.³

Manhattan (Population: 1.629 million 2017⁴), often referred to locally as the City, is the most densely populated of the five boroughs of New York City and its economic and administrative center, cultural identifier, and historical birthplace. Manhattan has been described as the cultural, financial, media, and entertainment capital of the world.⁵

Business Problem

Suppose we were commissioned by McDonald's management to evaluate the performance of all McDonald's branches in both Manhattan and Toronto, and we were also asked to compare the two cities in terms of performance of all branches and then to classify them according to public satisfaction degree.

One way to do that would be to evaluate each branch based on the overall consumer satisfaction. To fulfill our mission, we either have to ask a group of people for their own opinion after having eaten in a particular branch of the chain, or simply we could compare the total number of likes and the average ratings each branch has received from its customers.

Based on the above, this research will attempt to answer the following question:

Are all McDonald's branches equal in terms of final consumer opinion and which branches have been performing better than the others in Toronto and Manhattan ?

¹ <https://en.wikipedia.org/wiki/McDonald%27s>

² United Nations

³ <https://en.wikipedia.org/wiki/Toronto>

⁴ United Nations

⁵ <https://en.wikipedia.org/wiki/Manhattan>

II. DATA

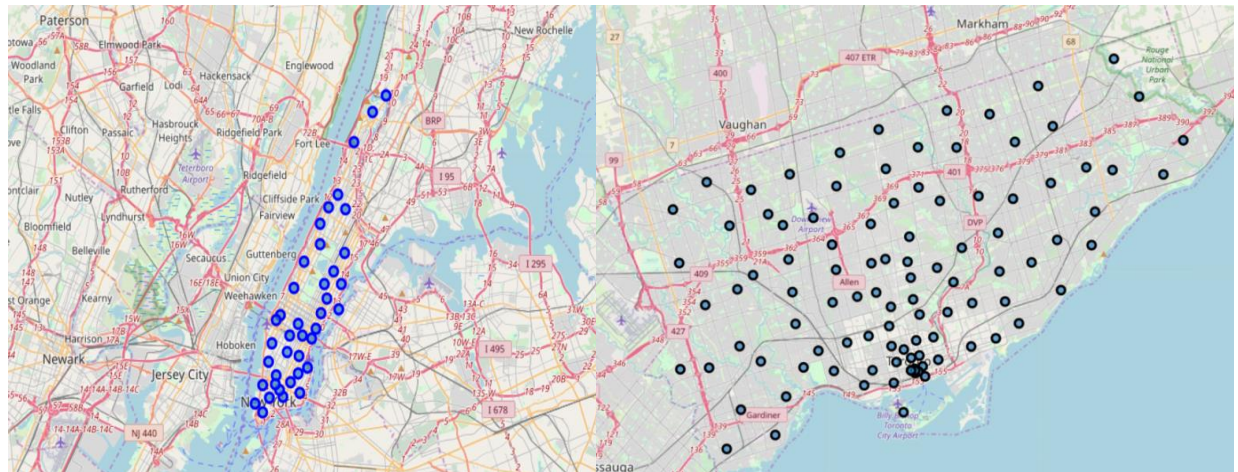
I used in this research Toronto regional data published on Wikipedia and New York's official data from NYU to review all areas and neighborhoods in both cities.

Toronto data:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

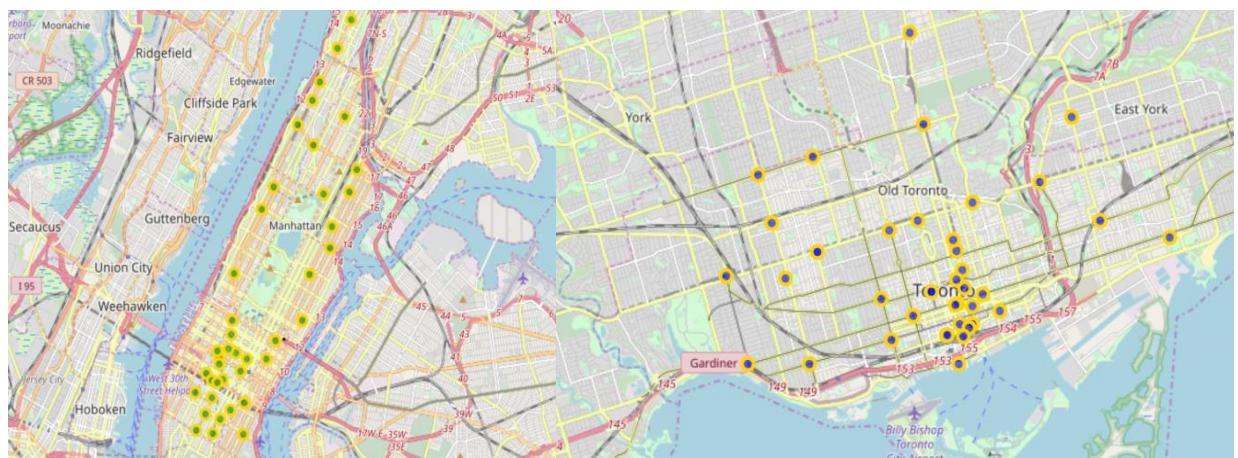
New York data:

<https://geo.nyu.edu/download/file/nyu-2451-34572-geojson.json>



Regarding the essence of this research and in order to achieve my task, I utilized the FourSquare API to get the coordinates for all McDonald's branches once in Manhattan and once in Toronto. Initially, a search query with McDonald's name within a 5 km radius was sent.

The following map shows all McDonalds branches to be studied; 41 in Manhattan and 40 in Toronto:



In addition, FourSquare API was used again to get both the total number of likes and the average ratings for each of the previous McDonalds branches.

III. METHODOLOGY

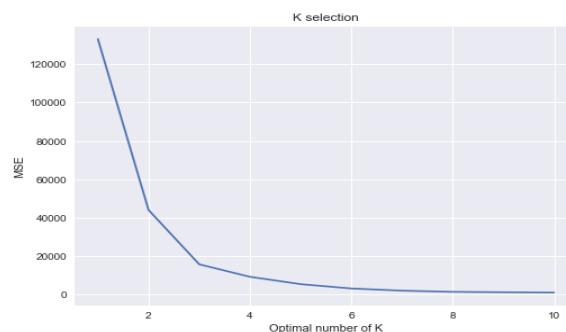
Firstly, the Geocoders library was used to retrieve the geographic coordinates of all districts and neighborhoods in Toronto and Manhattan and to determine their position on the world map for later visualization.

Secondly, the Foursquare API was used to explore the chain in all neighborhoods of Toronto and Manhattan. Once the raw data was obtained through the search query with McDonald's as input and within a 5km radius, the data were wrangled and filtered by determining, for example, 'city'=='Toronto' and 'categories'=='Fast Food Restaurant', in order to exclude any irrelevant data. Subsequently, all branches that consist of 'nan' values have been also removed from the data. After doing the same for Manhattan, I visualized the remaining 81 McDonalds branches in both cities.

The next step was getting the number of total likes and the average ratings for each branch of the chain in these two cities by passing the Id numbers of our branches in the Foursquare request. I then dropped three branches from Toronto data that had no rating values.

To compare the branches, I started clustering using the K means with "k-means++" as the centroid initialization method⁶ and applied clustering based on both the total number of likes and the average rating for each branch. K-Means Clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

To find the similarities between the two cities, I clustered them again, but this time together after determining the optimal number of k that was 3.

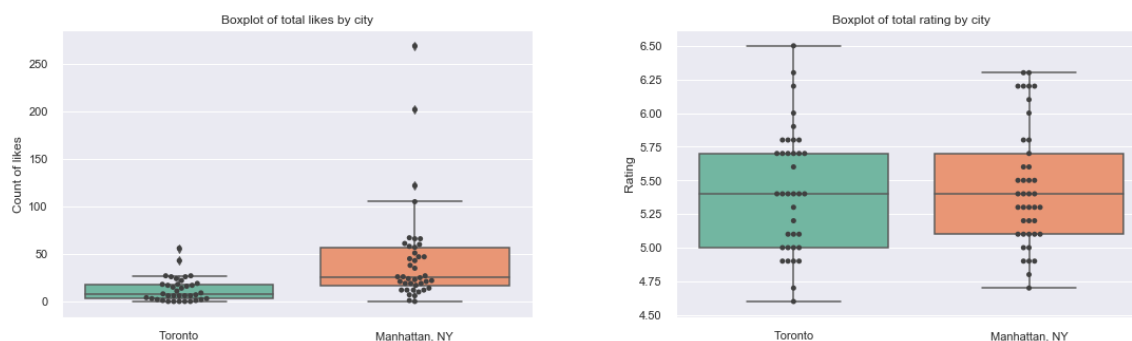


Finally, and to evaluate the performance of each branch of the chain, I created a categorical variable depends on the quantile distribution of the total number of likes.

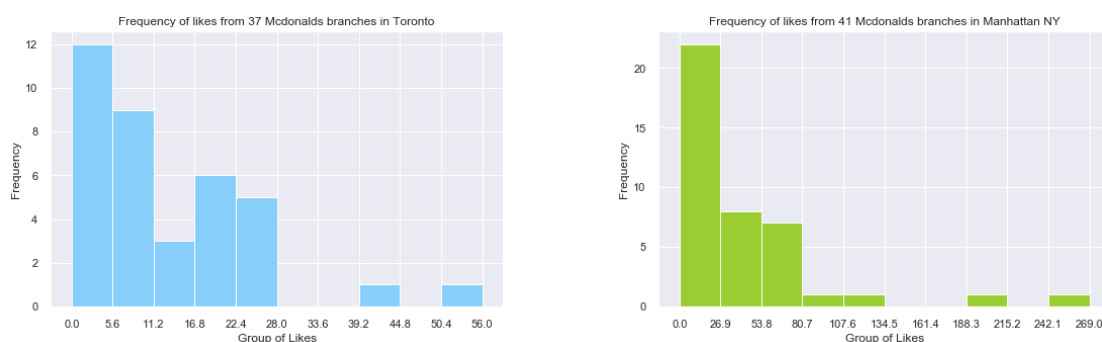
⁶ **k-means++** selects initial cluster centers for k-means clustering in a smart way to speed up convergence.

IV. RESULT

The descriptive statistics shows a huge difference between the two cities in terms of likes, where the descriptive statistics for the ratings are very similar for both cities as we can see through the following Boxplots:

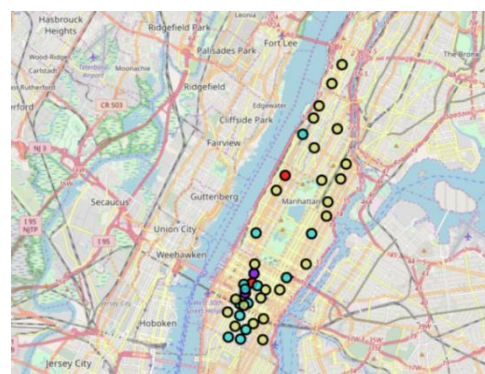


In addition, the distribution of likes in Toronto shows that the most branches fall within a major between 0 and 28 likes where the majority of branches in Manhattan got total number of likes between 0 and 81, as follows:



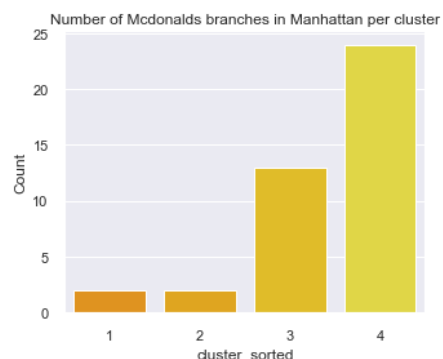
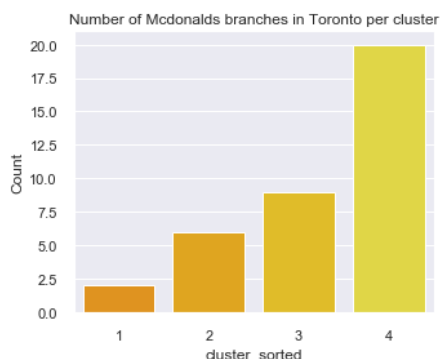
The K-means clustering categorized McDonalds branches in each city in 4 clusters based on the total number of likes and the average ratings:

- Cluster 1 (the best group): McDonalds branches with the most number of likes and the highest rating values.
- Cluster 2: McDonalds branches with the average number of likes and rating values.
- Cluster 3: McDonalds branches with modest number of likes rating values.
- Cluster 4 (the worst group): McDonalds branches with the least number of likes and the lowest rating values.

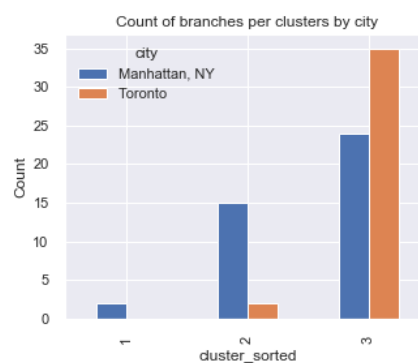
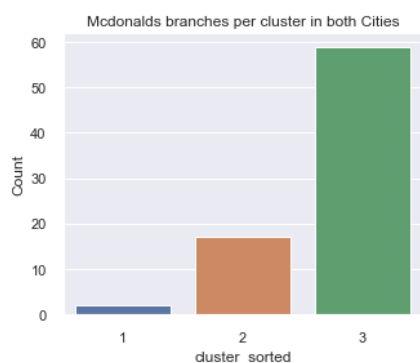


The results of the clustering are visualized in the previous map with cluster 1 in red color, cluster 2 in purple, cluster 3 in light blue and cluster 4 in yellow color.

It was also found by clustering each city individually that more than half of all branches in each city fall into the category of worst branches:



After clustering the two cities at the same time, It turned out that only two branches which are in Manhattan fall into the best group and 17 branches, of which only two are in Toronto, belong to the middle group and all other branches belong to the worst group, as follows:



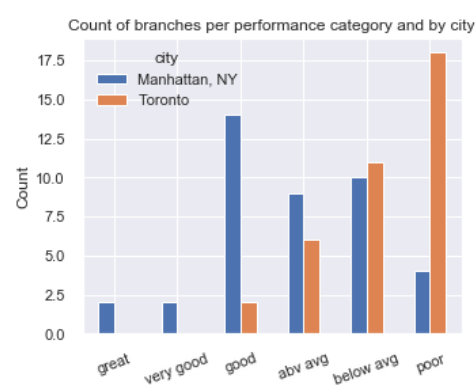
cluster_sorted	Count
0	3
1	2
2	1

cluster_sorted	city	count
1	Manhattan, NY	2
2	Manhattan, NY	15
	Toronto	2
3	Manhattan, NY	24
	Toronto	35

In terms of performance, I visualized the performance levels in six groups and found that the majority of branches fall within the level of performance from poor to good:

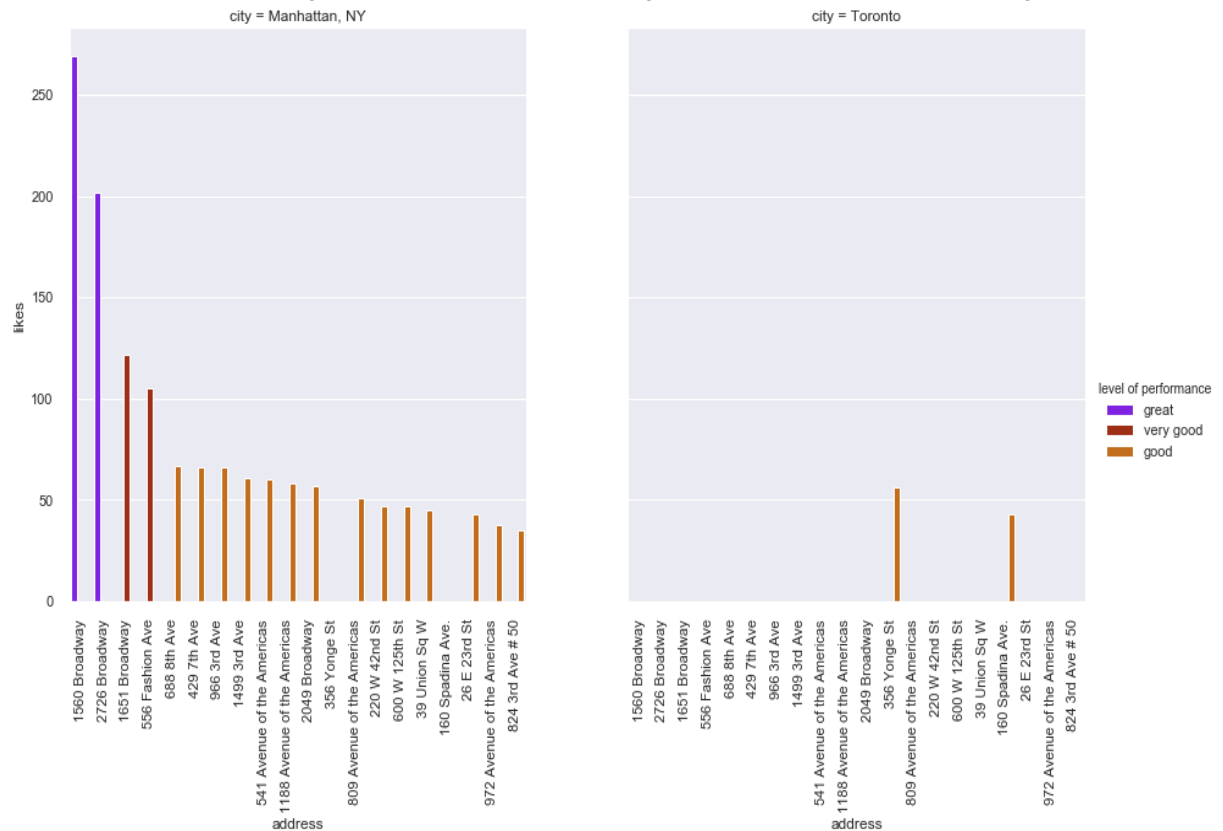


level of performance	city	count
great	Manhattan, NY	2
very good	Manhattan, NY	2
good	Manhattan, NY	14
	Toronto	2
abv avg	Manhattan, NY	9
	Toronto	6
below avg	Manhattan, NY	10
	Toronto	11
poor	Manhattan, NY	4
	Toronto	18



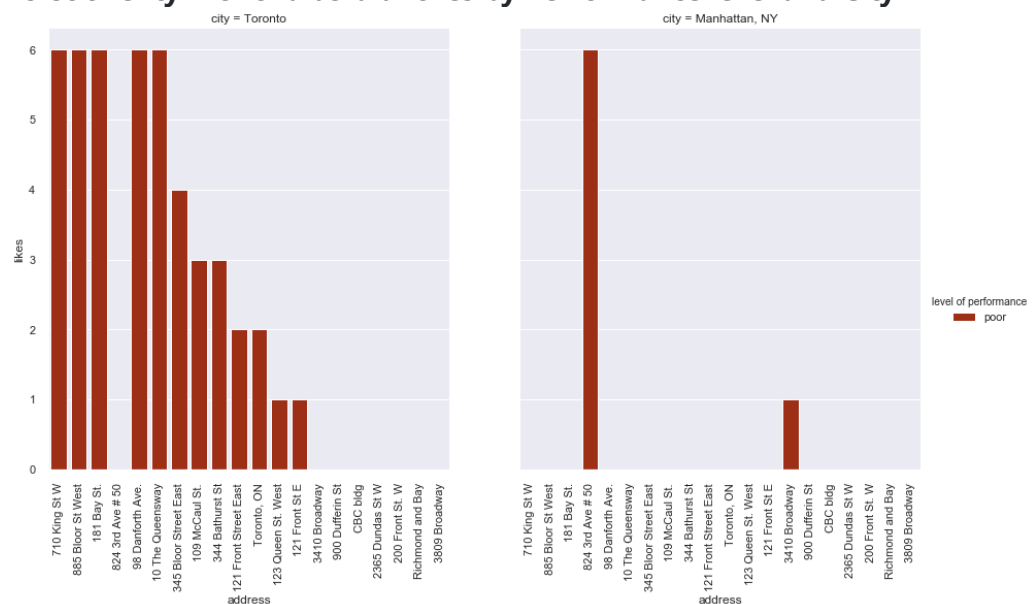
By sorting all branches according to their level of performance and selecting the best twenty (and then the worst twenty) of all studied branches, I could visualize the results through bar plots as follows:

Top twenty McDonalds branches by Performance level and City



It also turned out that 18 branches out of the worst twenty branches are located in Toronto and that the best eleven branches are in Manhattan:

Worst twenty McDonalds branches by Performance level and City



V. Discussion

I found that the average number of likes in Manhattan equals about four times the average number of likes in Toronto with 3 very high outliers in Manhattan.

I discovered that more than 30 branches in Toronto got less than 28 likes and the best branches in Toronto got less than 60 likes. On the other hand about a third of the branches in Manhattan got between 26 and 80 likes and 4 branches got more than 100 likes.

By clustering each city individually with $K = 4$ and sorting the branches from best to worst, it was found that more than half of all branches in each city fall into the worst category. It should be noted, however, that the branches of the best group in Toronto equals the third cluster in Manhattan in terms of total likes.

By clustering both cities simultaneously, It was found that only two branches in Manhattan fall into the first cluster (the best group) and 17 branches, of which only two are located in Toronto, belong to the middle group and all other branches belong to the third group (the worst group), 24 of which are in Manhattan and 35 are in Toronto.

It turned out that the majority of branches fall within the level of performance from poor to good and only four branches which are in Manhattan fall within the level of very good to great. Furthermore, the majority of poorly performing branches are in Toronto.

It turned out that 18 out of the worst twenty branches are in Toronto and that the best eleven branches are in Manhattan.

VI. CONCLUSION

Through this study, I have sought to answer the following question "Are all McDonald's branches equal in terms of final consumer opinion and which branches have been performing better than the others in Toronto and Manhattan?" by applying one of the Machine Learning techniques, known as K-Means Clustering, to the data available on Foursquare API regarding the consumer evaluation of McDonald's chain in Manhattan and Toronto.

With regard to the first part of the previous question, I concluded that there is a huge disparity within each city, where McDonalds branches fall into four groups in both Manhattan and Toronto in terms of the consumer feedback.

When I compared the two cities simultaneously, I found out that the number of likes and therefore the average level of performance in Toronto was significantly lower than the average level of performance in Manhattan, despite the similarity between Toronto and Manhattan in terms of the average rating values.

After grouping all branches in both cities in three clusters, it was found that although the number of branches of the third group is almost equal in both cities, the first group, which is considered the best group, includes only two branches located in Manhattan as well as 15 branches of the 17 branches in the second mid-performance group are located in Manhattan.

This means either McDonald's in Toronto is not very popular although the chain has more than 40 branches there or there are other reasons why the majority of consumers do not share their opinion about their own experience.

This low evaluation may be attributed, in my opinion, either to the decrease in the quality of meals and services or simply to the decrease in the demand for this type of food due to the wide variation in the cultural background of the population of Canada in general. This could also be attributed to the limited spread of Foursquare API in Canada compared to the United States.

Therefore, the results are not final and should not be generalized prior to further relevant research.

VII. References

<https://www.coursera.org/learn/applied-data-science-capstone>

<https://www.foursquare.com/>

<https://en.wikipedia.org/wiki/Toronto>

<https://en.wikipedia.org/wiki/Manhattan>

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

<https://www.geo.nyu.edu/>

VIII. Appendix

The following table shows the overall results ranked from best to worst and by city:

city	address	likes	ratings	level of performance	cluster_sorted
Manhattan, NY	1560 Broadway	269	6.3	great	1
	220 W 42nd St	202	6.2	great	1
	1651 Broadway	122	6.2	very good	2
	1188 Avenue of the Americas	105	5.4	very good	2
	972 Avenue of the Americas	67	5.3	good	2
	966 3rd Ave	66	6.2	good	2
	429 7th Ave	66	6.3	good	2
	490 8th Ave	61	5.5	good	2
	809 Avenue of the Americas	60	5.8	good	2
	600 W 125th St	58	5.6	good	2
Toronto	2049 Broadway	57	5.5	good	2
	356 Yonge St	56	6.5	good	2
	541 Avenue of the Americas	51	5.5	good	2
	556 Fashion Ave	47	5.5	good	2
	946 8th Ave	47	5.4	good	2
	39 Union Sq W	45	5.3	good	2
	West 103	43	5.6	good	2
	160 Spadina Ave.	43	5.4	good	2
	688 8th Ave	38	5.2	good	2
	824 3rd Ave # 50	35	5.7	good	3
Toronto	1185 Dupont St	27	5.6	abv avg	3
	123 Yonge St	27	5.1	abv avg	3
Manhattan, NY	2726 Broadway	27	5.4	abv avg	3
Toronto	595 Bay St	26	5.4	abv avg	3

Manhattan, NY	2549 Broadway	26	5.3	abv avg	3
Toronto	552 Yonge St	26	5.7	abv avg	3
Manhattan, NY	145 E 125th St	26	5.2	abv avg	3
	480 3rd Ave	25	5.4	abv avg	3
Toronto	470 Yonge St.	24	5.7	abv avg	3
Manhattan, NY	336 E 23rd St	24	5.1	abv avg	3
	14 E 47th St	23	5.2	abv avg	3
Toronto	220 Yonge St	22	6.3	abv avg	3
Manhattan, NY	4040 Broadway	22	4.8	abv avg	3
	1286 1st Ave	21	5.0	abv avg	3
	1499 3rd Ave	21	5.1	abv avg	3
Toronto	345 Bloor Street East	19	5.0	below avg	3
Manhattan, NY	3543-49 Broadway	19	6.0	below avg	3
	809-811 8th Ave	19	5.3	below avg	3
	2142 3rd Ave	19	4.9	below avg	3
Toronto	29-31 Woodward Ave.	18	5.4	below avg	3
	833 Bloor Street West	18	5.4	below avg	3
	127 Church St.	17	5.7	below avg	3
Manhattan, NY	1872 3rd Ave	17	6.1	below avg	3
Toronto	28 Eglinton Ave. E.	17	5.4	below avg	3
	1045 Pape Ave	16	6.2	below avg	3
	1221 King St. W	15	4.9	below avg	3
Manhattan, NY	151 W 34th St	14	5.1	below avg	3
Toronto	1000 Gerrard St E	14	4.9	below avg	3
Manhattan, NY	26 E 23rd St	12	5.3	below avg	3
	2379 Adam Clayton Powell Jr Blvd	12	5.1	below avg	3
	18 E 42nd St	12	4.7	below avg	3
	686 Avenue of the Americas	12	5.0	below avg	3
Toronto	192A Bloor St W	11	4.9	below avg	3
Manhattan, NY	354 W 125th St	10	4.9	below avg	3
Toronto	710 St Clair Ave W	9	5.7	below avg	3
	121 Front St E	8	5.0	below avg	3
	11 St. Clair Ave. E.	7	5.2	poor	3
Manhattan, NY	401 Park Ave S	7	6.2	poor	3
Toronto	2365 Dundas St W	6	4.9	poor	3
	10 The Queensway	6	5.1	poor	3
	200 Front St. W	6	5.3	poor	3
Manhattan, NY	1871 2nd Ave	6	4.9	poor	3
Toronto	1168 St. Clair Ave. W.	6	5.0	poor	3
	181 Bay St.	6	5.7	poor	3
	61 Front St. W.	4	4.6	poor	3
	344 Bathurst St	3	5.1	poor	3
	109 McCaul St.	3	5.0	poor	3
	98 Danforth Ave.	2	4.7	poor	3
	Toronto, ON	2	6.0	poor	3
Manhattan, NY	3410 Broadway	1	5.8	poor	3
Toronto	78 St. Patrick Street	1	5.7	poor	3
	Richmond and Bay	1	5.9	poor	3
Manhattan, NY	3809 Broadway	0	5.1	poor	3
Toronto	121 Front Street East	0	5.8	poor	3
	900 Dufferin St	0	5.4	poor	3
	123 Queen St. West	0	5.8	poor	3
	CBC bldg	0	5.8	poor	3
	710 King St W	0	5.8	poor	3