

COMPARING MCDONALD'S BRANCHES IN MANHATTAN AND TORONTO BY CONSUMER FEEDBACK ON FOURSQUARE API

Capstone Project: Battle of the Neighborhoods

IBM DATA SCIENCE SPECIALIZATION – WWW.COURSERA.ORG



Prepared by Mohamad Shammaa

August 2019

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.

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.

BUSINESS PROBLEM

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 ?

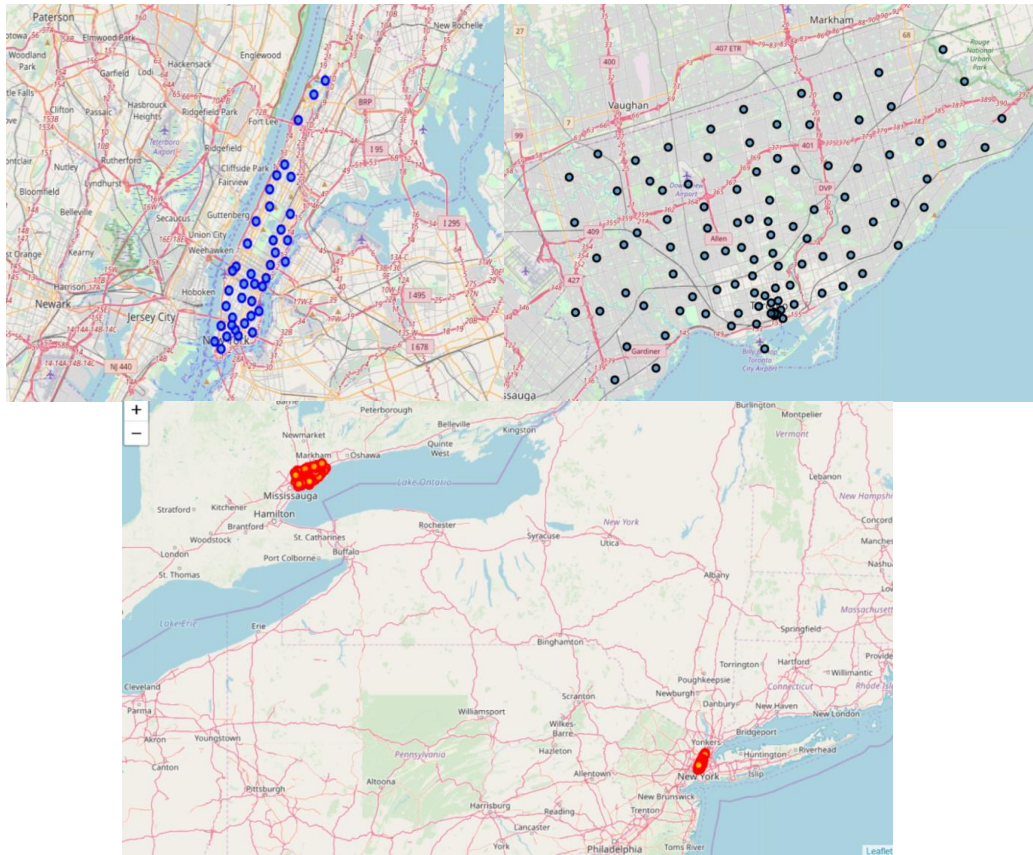
II. DATA

Toronto data from:

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

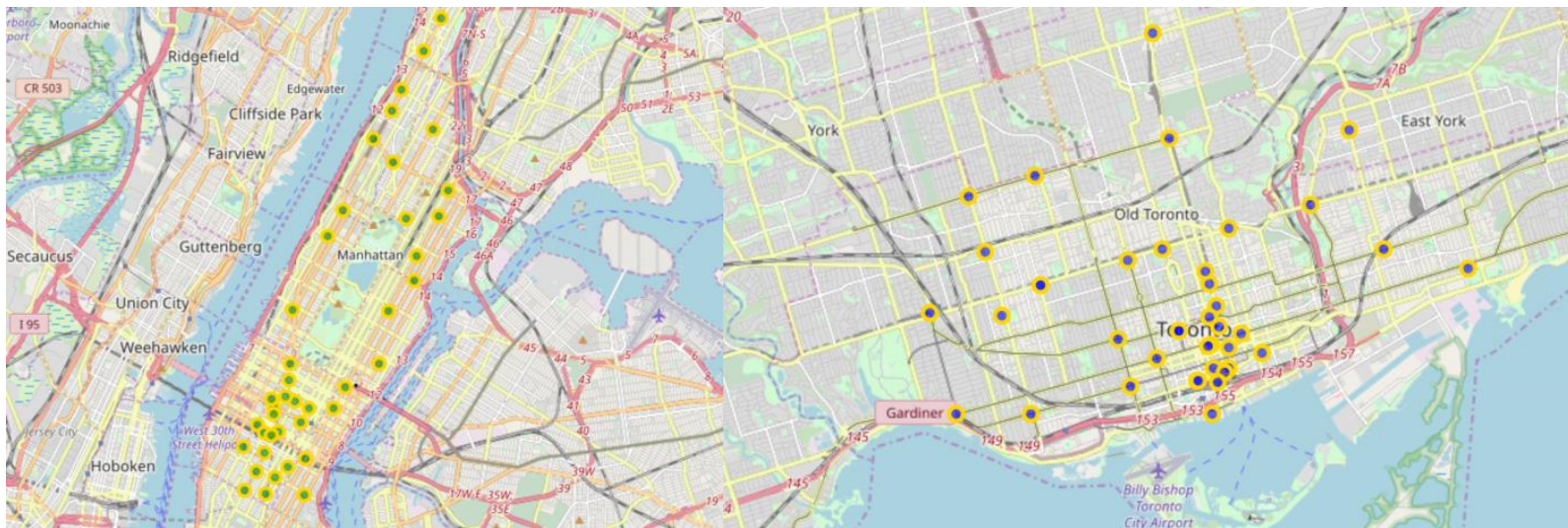
New York data from:

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



I used the FourSquare API to get the coordinates of all McDonalds branches in Manhattan and Toronto within a 5 km radius.

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.

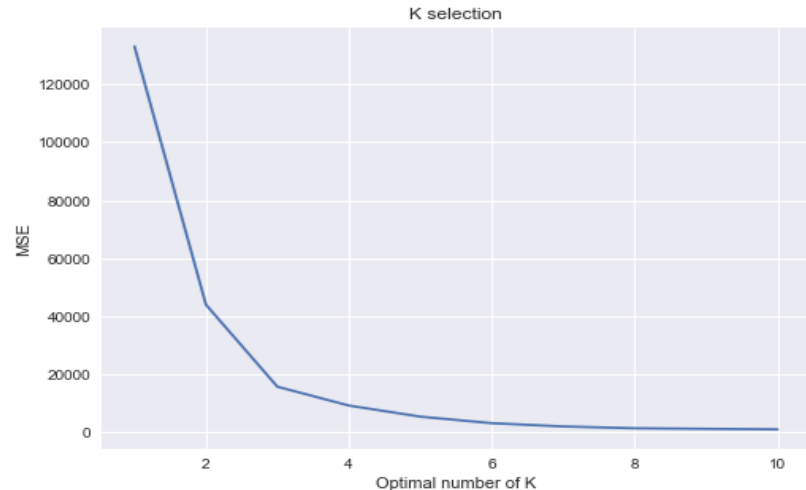
III. METHODOLOGY

The Geocoders library was used to retrieve the geographic coordinates of all districts and neighborhoods in Toronto.

The Foursquare API was used to explore the chain in all neighborhoods of Toronto and Manhattan, and to get the total likes and the average ratings for each branch of the chain.

To compare the branches, I started clustering using the K means with "*k-means++*" as the centroid initialization method and applied clustering based on both the total number of likes and the average rating for each branch.

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.



To evaluate the performance of each branch of the chain, I have created a categorical variable that depends on the quantile distribution of the total likes:

Add a column describes the level of performance depending on the quantile values of likes

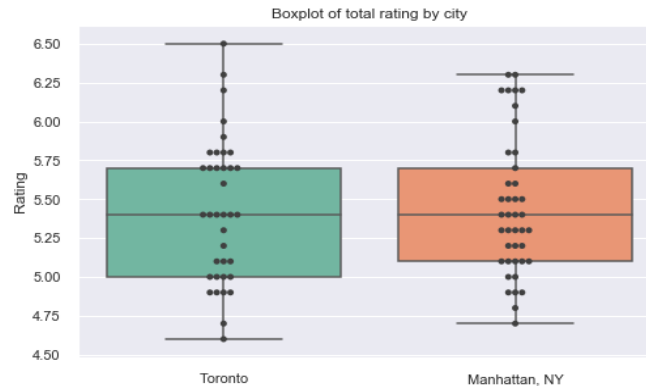
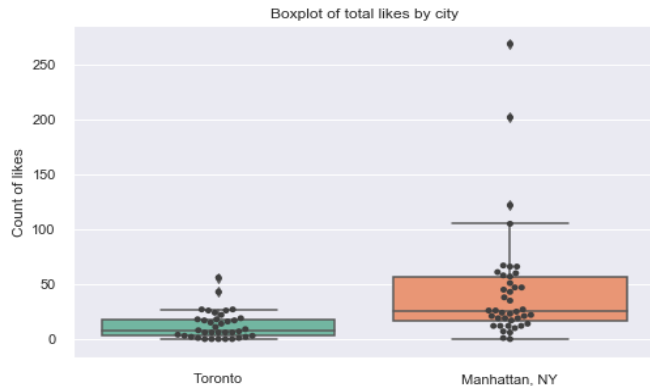
```
df_final['likes'].describe()
```

```
count      78.000000
mean       29.230769
std        41.449366
min         0.000000
25%         6.250000
50%        18.500000
75%        33.000000
max        269.000000
Name: likes, dtype: float64
```

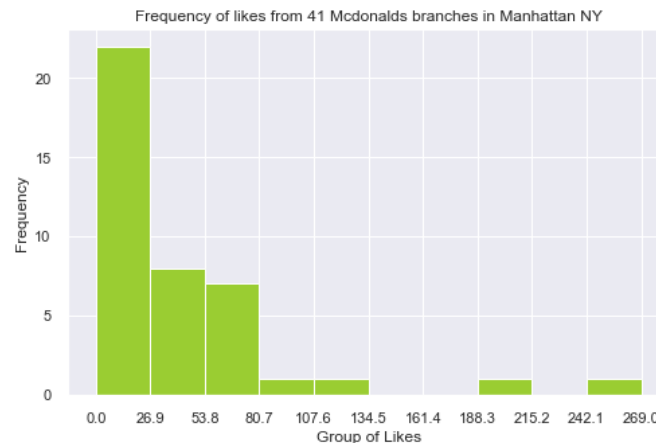
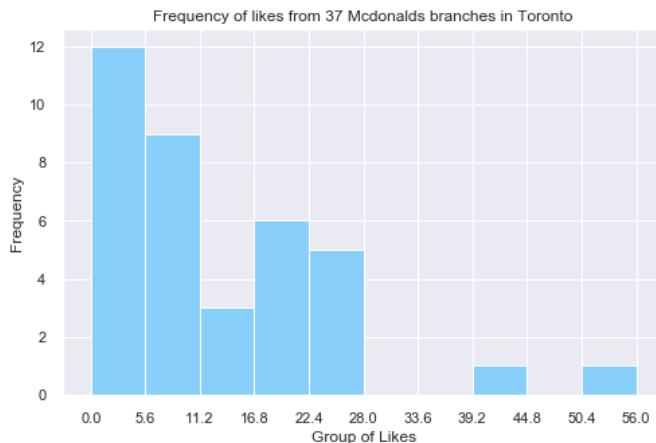
```
poor = df_final['likes']<=7
below_avg = df_final[(df_final['likes']>7) & (df_final['likes']<=19)]
abv_avg = df_final[(df_final['likes']>19) & (df_final['likes']<=34)]
good = df_final[(df_final['likes']>34) & (df_final['likes']<=100)]
very_good = df_final[(df_final['likes']>100) & (df_final['likes']<=150)]
great = df_final['likes']>150
def conditions(n):
    if n['likes']<=7:
        return 'poor'
    if n['likes']<=19:
        return 'below avg'
    if n['likes']<=34:
        return 'abv avg'
    if n['likes']<=100:
        return 'good'
    if n['likes']<=150:
        return 'very good'
    if n['likes']>150:
        return 'great'
df_final['level of performance']=df_final.apply(conditions, axis=1)
```

IV. RESULT

A huge difference between the two cities in terms of likes, where the ratings are very similar for both cities:



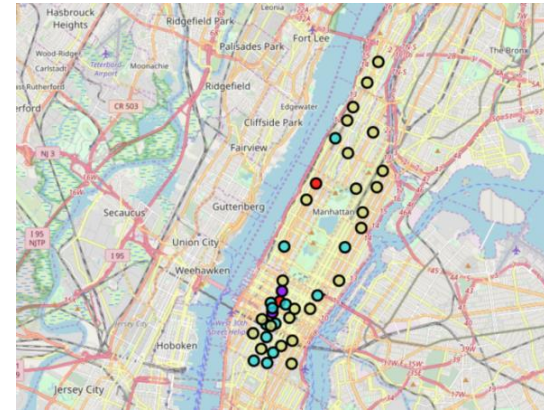
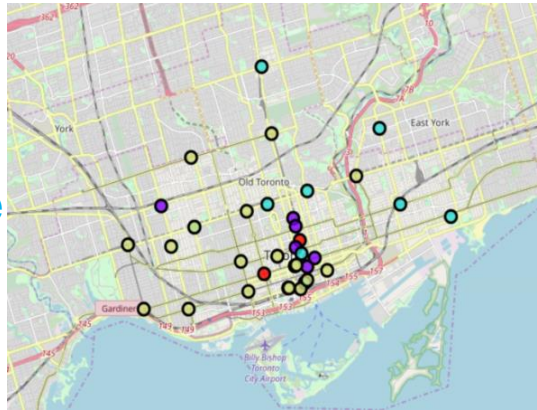
The distribution of likes in Toronto shows that the most branches fall within a major between 0 and 28 likes where the majority of branches in Manhattan got total number of likes between 0 and 81:



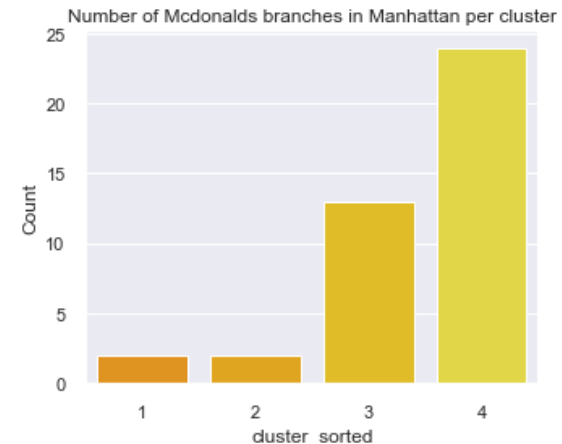
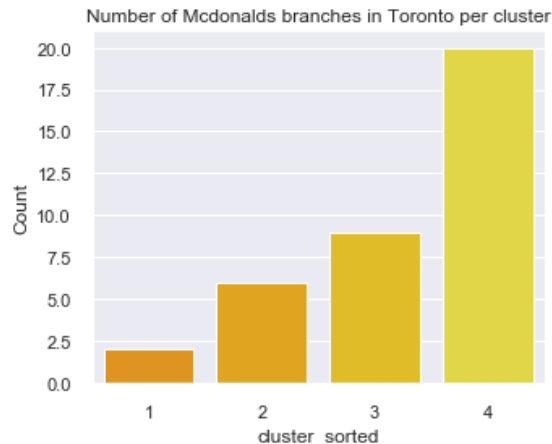
The K-means clustering categorized McDonalds branches in each city in 4 clusters based on likes and ratings:

- 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.

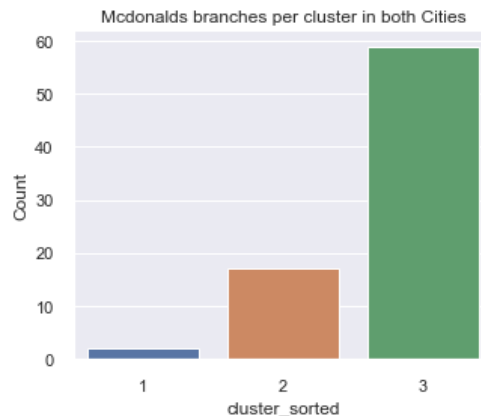
cluster 1 Red
cluster 2 Purple
cluster 3 Light blue
cluster 4 Yellow



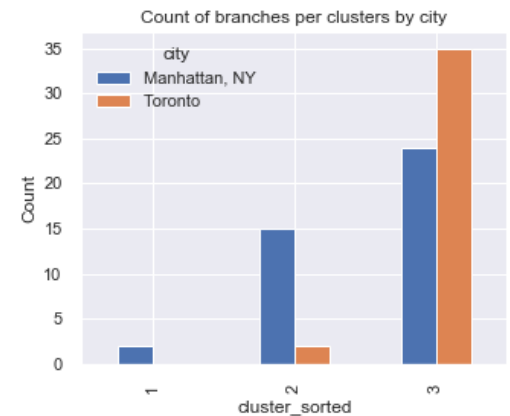
More than half of all branches in each city fall into the category of worst branches:



By 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, belong to the middle group and all other branches belong to the worst group:



cluster_sorted	Count
0	3
1	2
2	1

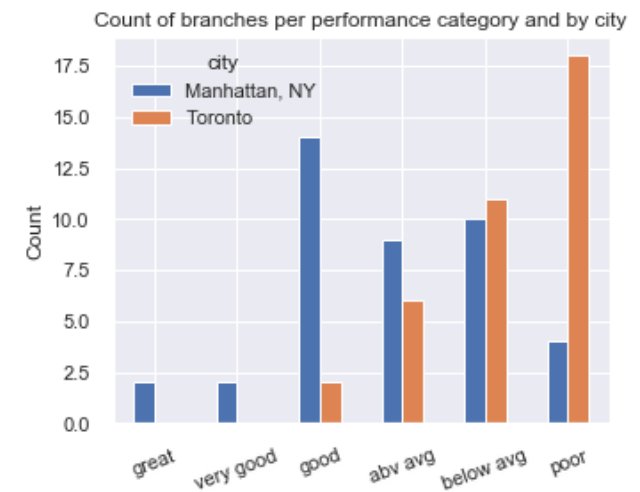


cluster_sorted	city	count
1	Manhattan, NY	2
2	Manhattan, NY	15
	Toronto	2
3	Manhattan, NY	24
	Toronto	35

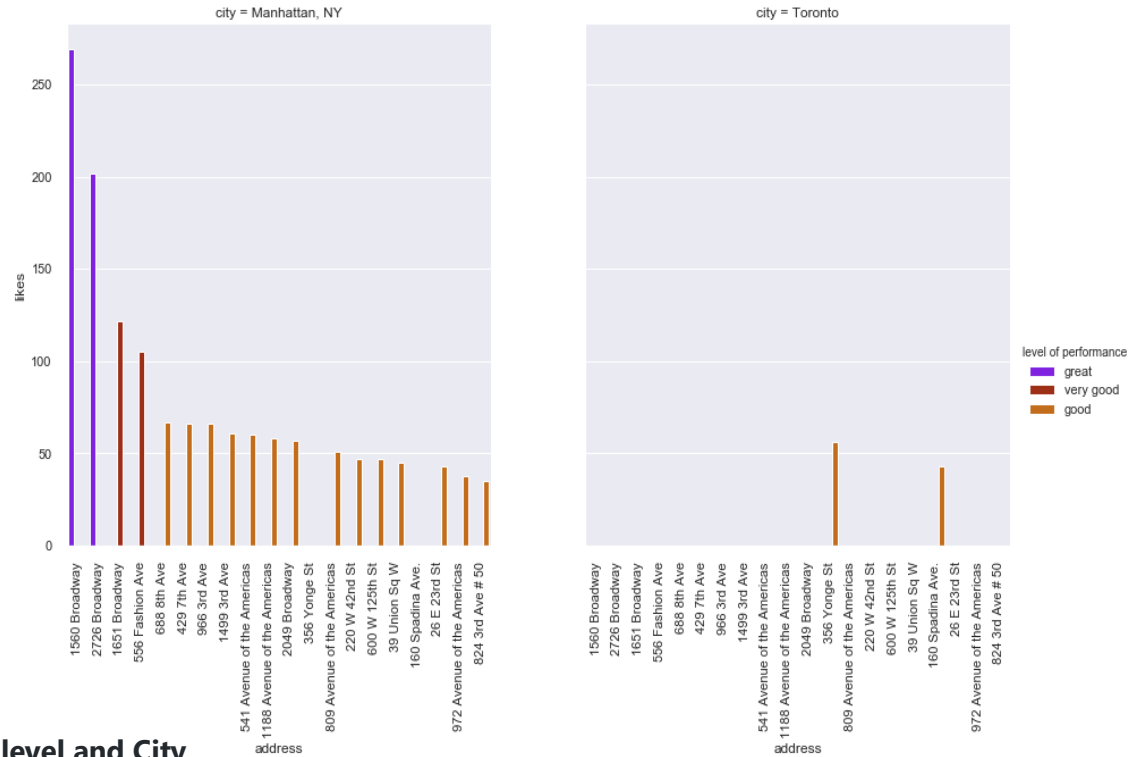
In terms of overall performance, I visualized the performance levels in six groups and found that the majority of branches fall within the level of performance from poor to good:



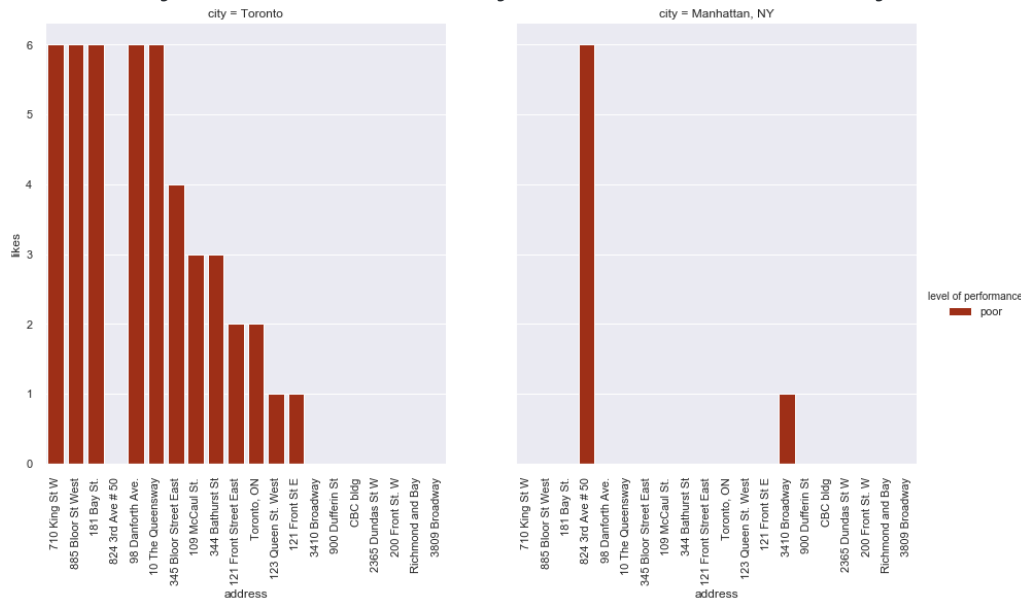
level of performance	city	count
great	Manhattan, NY	2
very good	Manhattan, NY	2
good	Manhattan, NY	14
	Toronto	2
abv avg	Manhattan, NY	9
	Toronto	6
below avg	Manhattan, NY	10
	Toronto	11
poor	Manhattan, NY	4
	Toronto	18



Top twenty McDonalds branches by Performance level and City



Worst 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:

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; where 18 branches 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 the K-Means Clustering, to the data available on Foursquare API regarding the consumer evaluation of McDonald's chain.

I concluded that there is a huge disparity within each city in terms of the consumer feedback.

When comparing 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

It was also 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 are 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 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 could be due 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.

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

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