# Capstone Project - The Battle of the Neighborhoods (Week 2)

## Comparing cities Project: New York Vs Toronto

Marc Arold ROSESMOND

July 2020

## Abstract

The globalization of the world had had a great effect on our standard of living. We can now travel all over the world for diverse reasons. Consequently, the tourism industry has grown tremendously. Every cities, every country makes many efforts for having the best place to visit. Meanwhile all of those efforts, there some countries witch are above in terms of beauty, organizations, hotels, and restaurant. The differences, however are difficult to see and to analyze. This work try to analyze those difference for two cities that people think are best in the world, New York and Toronto. We use the Forthsqare API for the venues that we will use and Wikipedia for the location data.

## Background

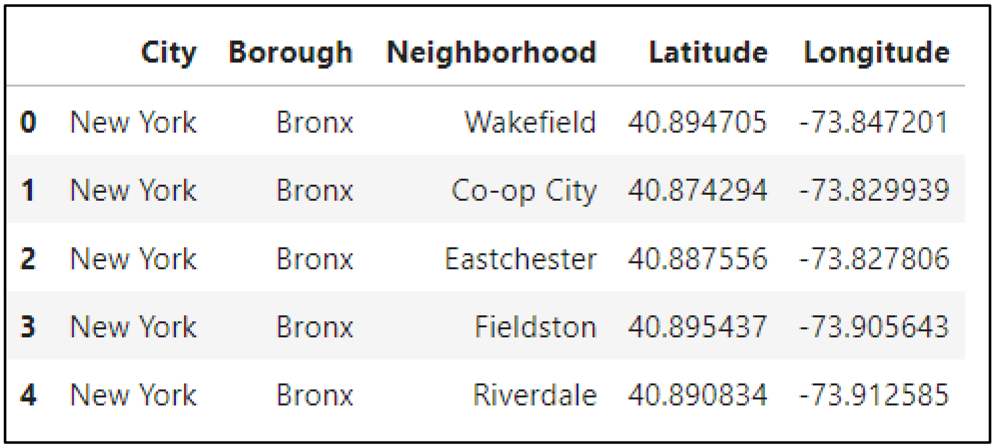
The comparison of cities has become crucial in the management of the tourism industry because globalization has put enormous pressure on them. In fact, the free movement of people has led to the emergence of a whole travel culture and thus to form a spectacular competition. Cities in a country compete with each other, each leveraging their strength to attract the most tourism. Thus, they built a whole social and cultural infrastructure in order to win this increasingly fierce competition.

The notion of competition therefore calls for knowing your competitor and assessing his potential strengths and weaknesses. It is therefore necessary for a given city, determined those which bring it closer to another city which makes it compete and also the factors of dissimilarity in order to establish a rational tourism policy to increase profits in this sector.

## Data Preparation

To arrive at our analysis, we will use relocated data. In fact, we are going for the five cities to look for the districts in each of these cities, as well as the districts, the districts and the GPS coordinates. For each district, we will use the Foursquare API to access 100 Venues within a radius of 500 meters. Once we have the first base which contains as variable, city, district, neighborhoods, GPS data of the neighborhoods. We will have to find for each of these districts a set of coming that characterizes them as well as the nature of it. Next, we will do a classification analysis from this data to classify the cities.

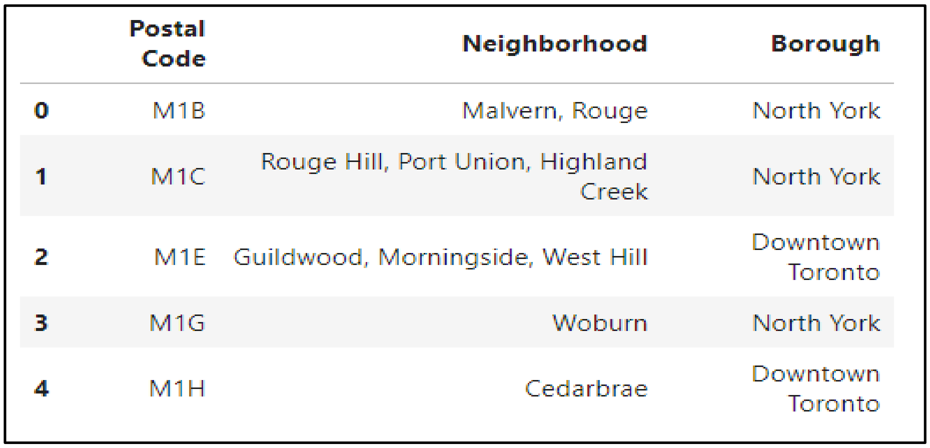
The following table presents a brief shot of the data for New York City.



| | City | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| **0** | New York | Bronx | Wakefield | 40.894705 | -73.847201 |
| **1** | New York | Bronx | Co-op City | 40.874294 | -73.829939 |
| **2** | New York | Bronx | Eastchester | 40.887556 | -73.827806 |
| **3** | New York | Bronx | Fieldston | 40.895437 | -73.905643 |
| **4** | New York | Bronx | Riverdale | 40.890834 | -73.912585 |

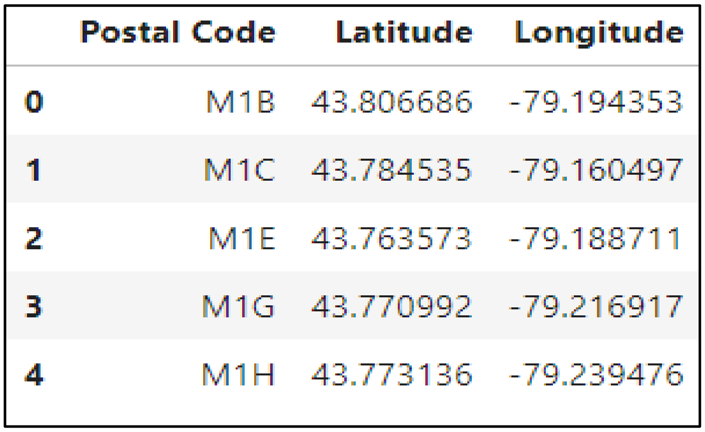
The process underlying these data is quite simple. First, we download the file *newyork_data* witch allows us to put data in a data frame we've called neighborhood_newyork. After that, we populate this blank data frame with their corresponding data with a *for* loop. And finally, we add a Colum call city with all the lines populate with "New York". This last step is really important because it will allows us to identify the cities, whether New York or Toronto.

The process for acquiring the data of Toronto is more difficult. In fact, we have a Wikipedia page with the postal code, the borough and their correspondent neighborhoods.



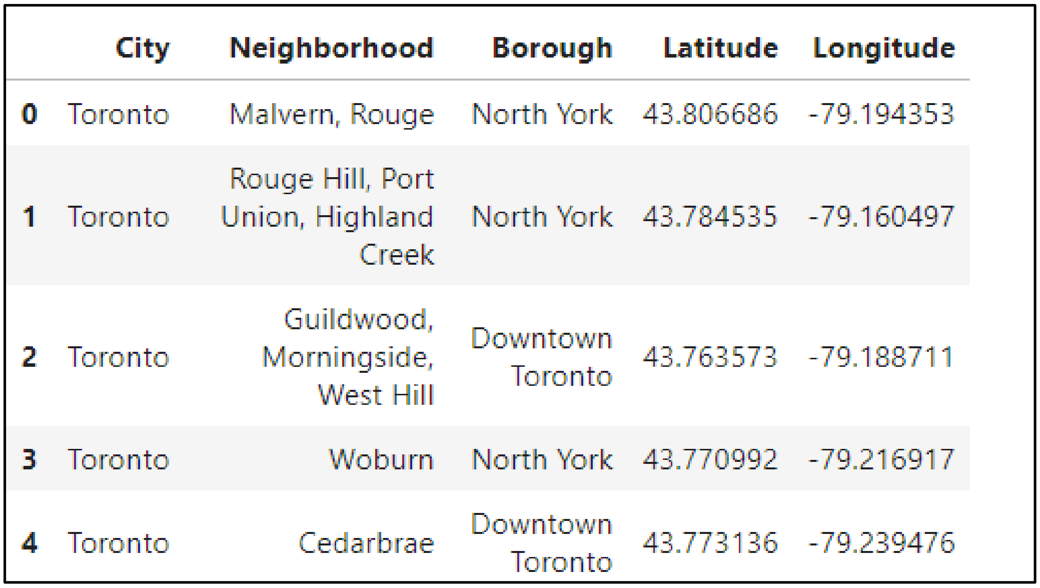
| | Postal Code | Neighborhood | Borough |
|---|---|---|---|
| **0** | M1B | Malvern, Rouge | North York |
| **1** | M1C | Rouge Hill, Port Union, Highland Creek | North York |
| **2** | M1E | Guildwood, Morningside, West Hill | Downtown Toronto |
| **3** | M1G | Woburn | North York |
| **4** | M1H | Cedarbrae | Downtown Toronto |

But it's not enough, we also need their location coordinates. For that, we have in our disposal another database with postal code, and their coordinates.



| | Postal Code | Latitude | Longitude |
|---|---|---|---|
| **0** | M1B | 43.806686 | -79.194353 |
| **1** | M1C | 43.784535 | -79.160497 |
| **2** | M1E | 43.763573 | -79.188711 |
| **3** | M1G | 43.770992 | -79.216917 |
| **4** | M1H | 43.773136 | -79.239476 |

After that, we merge the two databases on their primary keys "Postal code". Take a look on the final database for the Toronto city after we've added the Column City.

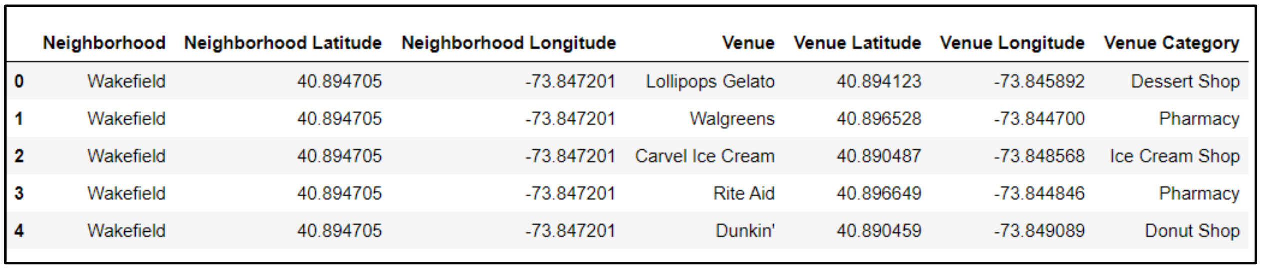


| | City | Neighborhood | Borough | Latitude | Longitude |
|---|---|---|---|---|---|
| **0** | Toronto | Malvern, Rouge | North York | 43.806686 | -79.194353 |
| **1** | Toronto | Rouge Hill, Port Union, Highland Creek | North York | 43.784535 | -79.160497 |
| **2** | Toronto | Guildwood, Morningside, West Hill | Downtown Toronto | 43.763573 | -79.188711 |
| **3** | Toronto | Woburn | North York | 43.770992 | -79.216917 |
| **4** | Toronto | Cedarbrae | Downtown Toronto | 43.773136 | -79.239476 |

The last step is to combine the two data sets into a one a final data set consisting of the following variables: City, Neighborhood, Borough, Latitude, and Longitude.

## Analysis of the cities based on their venues

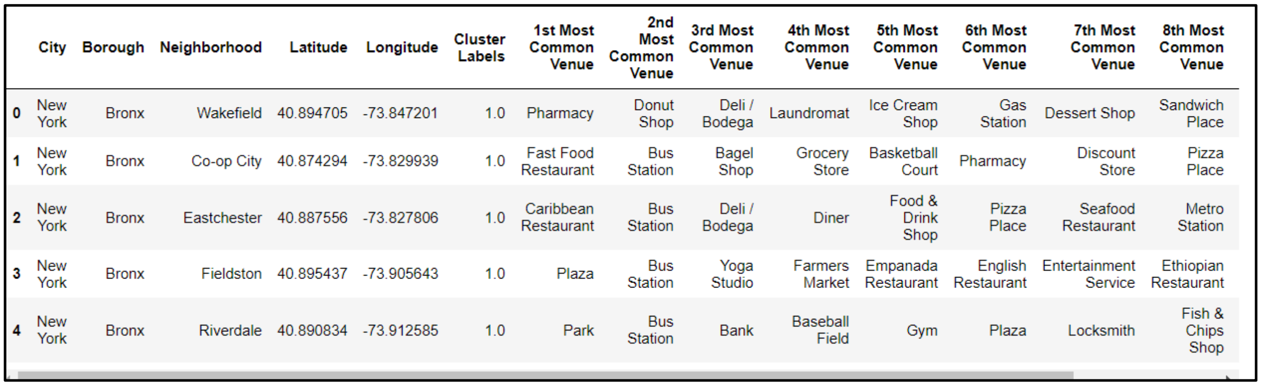
The fundamental aspect of our analysis is to find the points of similarity between the two cities according to their characteristics. To find the characteristics of these, we go by a request to the API of Fourthspare to find the arrivals. Near each arrival there are points of interest that we have grouped together with their contact details and the category in which they are located. The table below summarizes the characteristics:



| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Wakefield | 40.894705 | -73.847201 | Lollipops Gelato | 40.894123 | -73.845892 | Dessert Shop |
| 1 | Wakefield | 40.894705 | -73.847201 | Walgreens | 40.896528 | -73.844700 | Pharmacy |
| 2 | Wakefield | 40.894705 | -73.847201 | Carvel Ice Cream | 40.890487 | -73.848568 | Ice Cream Shop |
| 3 | Wakefield | 40.894705 | -73.847201 | Rite Aid | 40.896649 | -73.844846 | Pharmacy |
| 4 | Wakefield | 40.894705 | -73.847201 | Dunkin' | 40.890459 | -73.849089 | Donut Shop |

For example, the Wakefield district of respective coordinates, 40.894705 of latitude and -73.847201 of longitude contains the lollipops Gelato of respective coordinates 40.894123 of latitude and -73.845892. We can also place this venue in the Dessert Shop category.

To analyze the similarities of each city we perform a cluster analysis which will place each district in a category. No two districts are alike, even if they are not in the same city, if they are in the same cluster. Conversely, even if they are in the same city, if they are in different clusters, they are considered to be different. The table below shows the final result of the cluster analysis with the addition of a cluster column to the initial base to classify the districts of two cities.



| | City | Borough | Neighborhood | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | New York | Bronx | Wakefield | 40.894705 | -73.847201 | 1.0 | Pharmacy | Donut Shop | Deli / Bodega | Laundromat | Ice Cream Shop | Gas Station | Dessert Shop | Sandwich Place |
| 1 | New York | Bronx | Co-op City | 40.874294 | -73.829939 | 1.0 | Fast Food Restaurant | Bus Station | Bagel Shop | Grocery Store | Basketball Court | Pharmacy | Discount Store | Pizza Place |
| 2 | New York | Bronx | Eastchester | 40.887556 | -73.827806 | 1.0 | Caribbean Restaurant | Bus Station | Deli / Bodega | Diner | Food & Drink Shop | Pizza Place | Seafood Restaurant | Metro Station |
| 3 | New York | Bronx | Fieldston | 40.895437 | -73.905643 | 1.0 | Plaza | Bus Station | Yoga Studio | Farmers Market | Empanada Restaurant | English Restaurant | Entertainment Service | Ethiopian Restaurant |
| 4 | New York | Bronx | Riverdale | 40.890834 | -73.912585 | 1.0 | Park | Bus Station | Bank | Baseball Field | Gym | Plaza | Locksmith | Fish & Chips Shop |

The first results in this table show both the membership of each neighborhood in a cluster but also the 10 most frequented arrivals in each neighborhood. It becomes clear that in the Wakefield district, the most frequented place is the pharmacy and this district is in the first cluster.

## Conclusion

Our study, although simple, is very revealing, in fact, it makes it possible to make a rigorous classification of two big cities, New York and Toronto in order to elucidate their potential resemblance and differences according to the categories of places of frequentation of the two cities. We can conclude that our study is of great importance as it will not only allow public decision makers in both cities to take the comparative advantages of these cities and make more rational policy recommendations in terms of tourism policies.