

# Capstone Project - The Battle of the Neighborhoods

## Toronto and York

Ulises Ortega

November 2020

### 1. Introduction: Business Problem

#### Toronto vs York. Where to eat better?

Eventually, at some point in life, we must leave behind the place we have lived in for years, for a job change, looking for better opportunities, or simply to gain new experiences.

There are many elements that we must consider in order to decide about the place we will choose to live the next stage of our life. Since, in general, people change their home to live there for a long time, the resources available in the new home must be analyzed in depth, since the quality of life, that they have, will largely depend on it.

One of the points that must be considered is food, so in this analysis we will focus on distinguishing which city has the best places to eat. The cities considered are those that contain the words Toronto (Downtown Toronto, East Toronto, West Toronto, Central Toronto) and York (North York, East York, York), we will treat each group as a different city.



## 2. Data

We will use the data provided by Wikipedia to obtain the postal code of each city mentioned, we will unify the cities to have only two at the end (Toronto and York).

We will use the latitude and longitude data to print each location on a map.

Then we will use the Foursquare API to obtain the ratings of the restaurant category to finally obtain an average rating for each city and thus be able to select the best one.

With the data in the soup object, we will scrape the HTML document that contains until to get the table element. Then we will iterate in it to get each row and append each in a pandas DataFrame.

|   | Postal Code | Borough          | Neighbourhood                               |
|---|-------------|------------------|---|
| 0 | M1A         | Not assigned     | Not assigned                                |
| 1 | M2A         | Not assigned     | Not assigned                                |
| 2 | M3A         | North York       | Parkwoods                                   |
| 3 | M4A         | North York       | Victoria Village                            |
| 4 | M5A         | Downtown Toronto | Regent Park, Harbourfront                   |
| 5 | M6A         | North York       | Lawrence Manor, Lawrence Heights            |
| 6 | M7A         | Downtown Toronto | Queen's Park, Ontario Provincial Government |
| 7 | M8A         | Not assigned     | Not assigned                                |
| 8 | M9A         | Etobicoke        | Islington Avenue, Humber Valley Village     |
| 9 | M1B         | Scarborough      | Malvern, Rouge                              |

Clean data frame.

```
▶ MI
# Clean dataframe

# Ignore cells with a borough that is Not assigned.
df = df[(df['Borough'] != 'Not assigned')]
print('Not assigned borough =', df[(df['Borough'] == 'Not assigned')]['Borough'].count())

# More than one neighborhood can exist in one postal code area.
print('Repeated postal code areas =', df['Postal Code'].value_counts()[df['Postal Code'].value_counts() > 1].count() )

# Not assigned neighborhood count.
print('Not assigned neighborhood =', df[(df['Neighbourhood'] == 'Not assigned')]['Neighbourhood'].count())

df.reset_index(drop = True, inplace = True)

df.head()

Not assigned borough = 0
Repeated postal code areas = 0
Not assigned neighborhood = 0
```

Get the latitude and the longitude coordinates of each neighborhood.

|   | 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 |

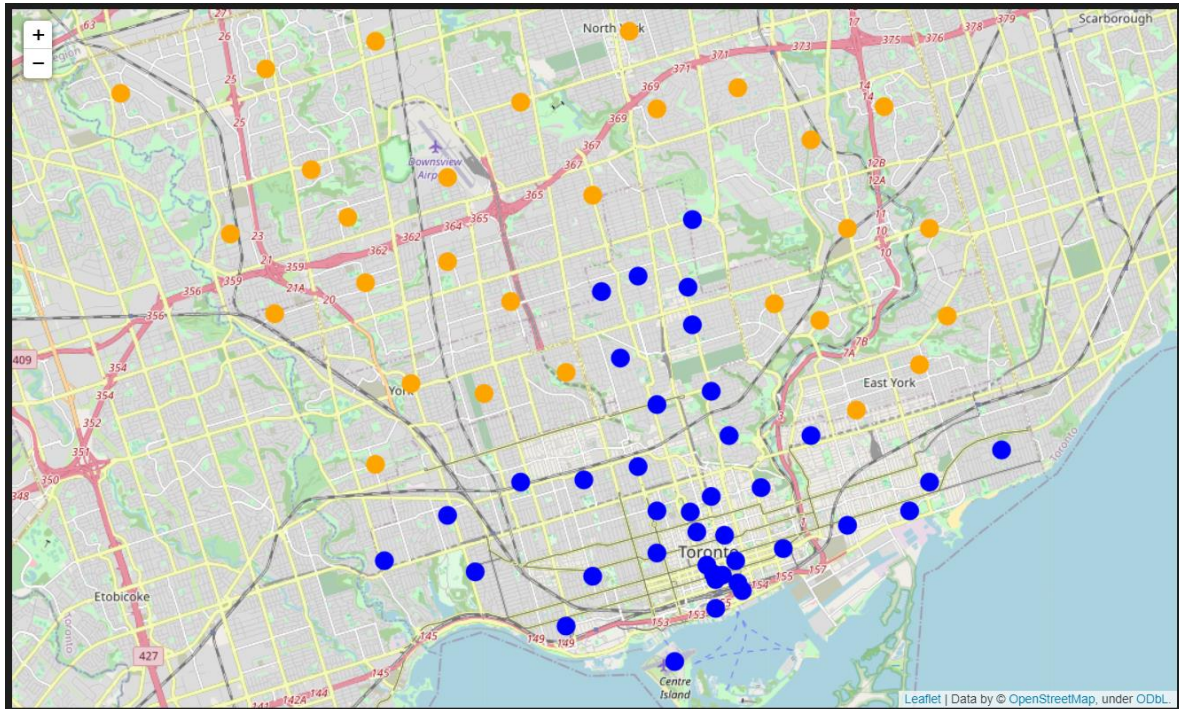
Then, we will obtain the complete dataset with the Latitude and Longitude values for each Neighborhood.

|     | PostalCode | Borough | Neighborhood                                | Latitude  | Longitude  |
|-----|------------|---------|---|-----------|------------|
| 0   | M5A        | Toronto | Regent Park, Harbourfront                   | 43.654260 | -79.360636 |
| 1   | M7A        | Toronto | Queen's Park, Ontario Provincial Government | 43.662301 | -79.389494 |
| 2   | M5B        | Toronto | Garden District, Ryerson                    | 43.657162 | -79.378937 |
| 3   | M5C        | Toronto | St. James Town                              | 43.651494 | -79.375418 |
| 4   | M4E        | Toronto | The Beaches                                 | 43.676357 | -79.293031 |
| ... | ...        | ...     | ...   | ...       | ...        |
| 68  | M3N        | York    | Downsview                                   | 43.761631 | -79.520999 |
| 69  | M6N        | York    | Runnymede, The Junction North               | 43.673185 | -79.487262 |
| 70  | M9N        | York    | Weston                                      | 43.706876 | -79.518188 |
| 71  | M2P        | York    | York Mills West                             | 43.752758 | -79.400049 |
| 72  | M2R        | York    | Willowdale, Willowdale West                 | 43.782736 | -79.442259 |

### 3. Methodology and Analysis

The methodology to follow consist in get restaurants data of Toronto and York from Foursquare API. We will use standard and premium calls to get that information for each restaurant. We will clean and analyze those data to get insights and then put into a Kmeans algorithm to generate clusters and get conclusion. We will support od Folium library too to create visual material and make the analysis more interactive.

Get coordinates of Toronto and York, then create a map with Folium and create a map.



We can see than the map is more populated with Toronto (blue) venues than York (orange).

We will use a function to get the venues of Toronto and York through Foursquare API.

|                      | Borough | Neighborhood                                | Neighborhood Latitude | Neighborhood Longitude | Venue ID                  | Venue Name                 | Venue Latitude | Venue Longitude | Venue Category            |
|----------------------|---------|---|-----------------------|------------------------|---------------------------|----------------------------|----------------|-----------------|---------------------------|
| 0                    | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | 5612b1cc498e3dd742af0dc8  | Impact Kitchen             | 43.656369      | -79.356980      | Restaurant                |
| 1                    | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | 53a22c92498ec91fda7ce133  | Cluny Bistro & Boulangerie | 43.650565      | -79.357843      | French Restaurant         |
| 2                    | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | 51ddecee498e1ffd34185d2f  | El Catrin                  | 43.650601      | -79.358920      | Mexican Restaurant        |
| 3                    | Toronto | Queen's Park, Ontario Provincial Government | 43.662301             | -79.389494             | 4a8355bfff964a520d3falfe3 | Mercatto                   | 43.660391      | -79.387664      | Italian Restaurant        |
| 4                    | Toronto | Queen's Park, Ontario Provincial Government | 43.662301             | -79.389494             | 52d884c5498ecf5c7cafe5ab  | Nando's                    | 43.661728      | -79.386391      | Portuguese Restaurant     |
| ...                  | ...     | ...   | ...                   | ...                    | ...                       | ...                        | ...            | ...             | ...                       |
| 443                  | York    | Willowdale, Willowdale East                 | 43.770120             | -79.408493             | 4bb628bfef159c74e18075f7  | Shawarma Max               | 43.766012      | -79.410844      | Middle Eastern Restaurant |
| 444                  | York    | Willowdale, Willowdale East                 | 43.770120             | -79.408493             | 544ad69a498eeb86b7a4a8d0  | Kinton Ramen               | 43.769684      | -79.413049      | Ramen Restaurant          |
| 445                  | York    | Willowdale, Willowdale East                 | 43.770120             | -79.408493             | 4b4779f0f964a520543326e3  | Pho 88 Vietnamese Cuisine  | 43.770456      | -79.413064      | Vietnamese Restaurant     |
| 446                  | York    | Willowdale, Willowdale East                 | 43.770120             | -79.408493             | 4b8bffc6f964a52076b732e3  | Ichiban Fish House         | 43.765995      | -79.410813      | Japanese Restaurant       |
| 447                  | York    | Willowdale, Willowdale East                 | 43.770120             | -79.408493             | 5bb92d3ec53093002c1d6a1a  | PROJECT:FISH               | 43.769100      | -79.414305      | Sushi Restaurant          |
| 448 rows x 9 columns |         |   |                       |                        |                           |                            |                |                 |                           |

448 rows x 9 columns

Then, only get the Restaurant category from the data.

|   | Borough | Neighborhood                                | Neighborhood Latitude | Neighborhood Longitude | Venue ID                 | Venue Name                 | Venue Latitude | Venue Longitude | Venue Category        |
|---|---------|---|-----------------------|------------------------|--------------------------|----------------------------|----------------|-----------------|-----------------------|
| 0 | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | 5612b1cc498e3dd742af0dc8 | Impact Kitchen             | 43.656369      | -79.356980      | Restaurant            |
| 1 | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | 53a22c92498ec91fda7ce133 | Cluny Bistro & Boulangerie | 43.650565      | -79.357843      | French Restaurant     |
| 2 | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | 51ddecee498e1ffd34185d2f | El Catrin                  | 43.650601      | -79.358920      | Mexican Restaurant    |
| 3 | Toronto | Queen's Park, Ontario Provincial Government | 43.662301             | -79.389494             | 4a8355bff964a520d3fa1fe3 | Mercatto                   | 43.660391      | -79.387664      | Italian Restaurant    |
| 4 | Toronto | Queen's Park, Ontario Provincial Government | 43.662301             | -79.389494             | 52d884c5498ecf5c7cafe5ab | Nando's                    | 43.661728      | -79.386391      | Portuguese Restaurant |

Use a function to get the rating for each restaurant in the dataset. This function uses a Premium call of Foursquare.

|   | Venue ID                 | Venue Name                 | Venue Rating |
|---|--------------------------|----------------------------|--------------|
| 0 | 5612b1cc498e3dd742af0dc8 | Impact Kitchen             | 8.8          |
| 1 | 53a22c92498ec91fda7ce133 | Cluny Bistro & Boulangerie | 8.3          |
| 2 | 51ddecee498e1ffd34185d2f | El Catrin                  | 7.9          |
| 3 | 4a8355bff964a520d3fa1fe3 | Mercatto                   | 8.3          |
| 4 | 52d884c5498ecf5c7cafe5ab | Nando's                    | 8.1          |

Finally, merge Toronto and York venues data with its ratings and drop instances with null values.

|     | Borough | Neighborhood                                | Neighborhood Latitude | Neighborhood Longitude | Venue ID                 | Venue Name                 | Venue Latitude | Venue Longitude | Venue Category            | Venue Rating |
|-----|---------|---|-----------------------|------------------------|--------------------------|----------------------------|----------------|-----------------|---------------------------|--------------|
| 0   | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | 5612b1cc498e3dd742af0dc8 | Impact Kitchen             | 43.656369      | -79.356980      | Restaurant                | 8.8          |
| 1   | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | 53a22c92498ec91fda7ce133 | Cluny Bistro & Boulangerie | 43.650565      | -79.357843      | French Restaurant         | 8.3          |
| 2   | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | 51ddecee498e1ffd34185d2f | El Catrin                  | 43.650601      | -79.358920      | Mexican Restaurant        | 7.9          |
| 3   | Toronto | Queen's Park, Ontario Provincial Government | 43.662301             | -79.389494             | 4a8355bff964a520d3fa1fe3 | Mercatto                   | 43.660391      | -79.387664      | Italian Restaurant        | 8.3          |
| 4   | Toronto | Central Bay Street                          | 43.657952             | -79.387383             | 4a8355bff964a520d3fa1fe3 | Mercatto                   | 43.660391      | -79.387664      | Italian Restaurant        | 8.3          |
| ... | ...     | ...   | ...                   | ...                    | ...                      | ...                        | ...            | ...             | ...                       | ...          |
| 425 | York    | Willowdale, Willowdale East                 | 43.770120             | -79.408493             | 4b832e14f964a520d0fc30e3 | Wendy's                    | 43.768461      | -79.412377      | Fast Food Restaurant      | 6.2          |
| 426 | York    | Willowdale, Willowdale East                 | 43.770120             | -79.408493             | 4bb628fbef159c74e18075f7 | Shawarma Max               | 43.766012      | -79.410844      | Middle Eastern Restaurant | 6.4          |
| 427 | York    | Willowdale, Willowdale East                 | 43.770120             | -79.408493             | 544ad69a498eeb86b7a4a8d0 | Kinton Ramen               | 43.769684      | -79.413049      | Ramen Restaurant          | 6.1          |
| 428 | York    | Willowdale, Willowdale East                 | 43.770120             | -79.408493             | 4b4779f0f964a520543326e3 | Pho 88 Vietnamese Cuisine  | 43.770456      | -79.413064      | Vietnamese Restaurant     | 6.1          |
| 429 | York    | Willowdale, Willowdale East                 | 43.770120             | -79.408493             | 4b8bffc6f964a52076b732e3 | Ichiban Fish House         | 43.765995      | -79.410813      | Japanese Restaurant       | 5.7          |

430 rows x 10 columns

In this point, we will select 'Venue Category' and 'Venue Rating' as features to put in the Kmeans algorithm to create two clusters, which will indicate if each restaurant is in the most rated group or not.

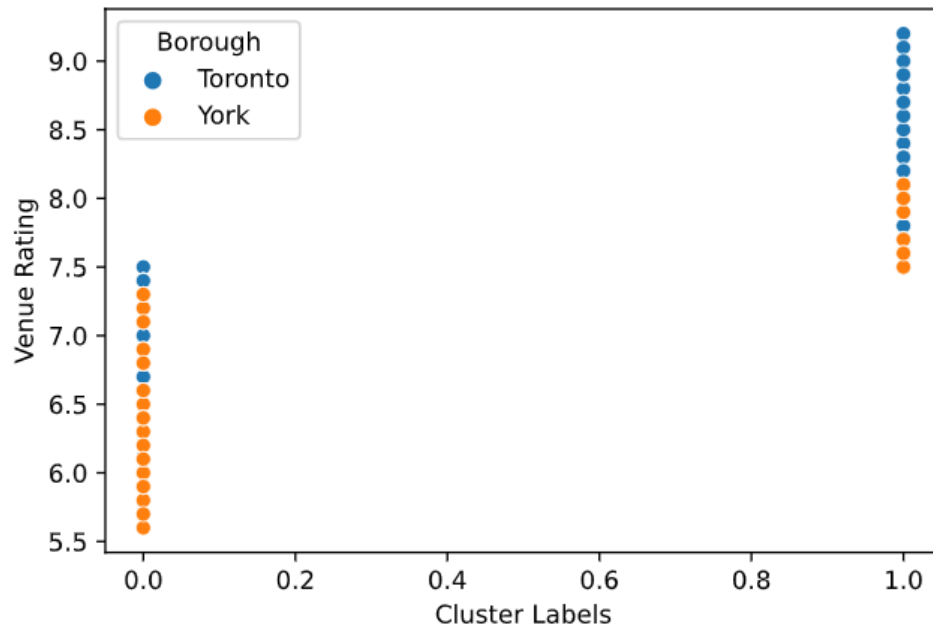
Run a k-means clustering and check cluster labels generated for each row in the data frame.

Create a new data frame with the principal features and its labels.

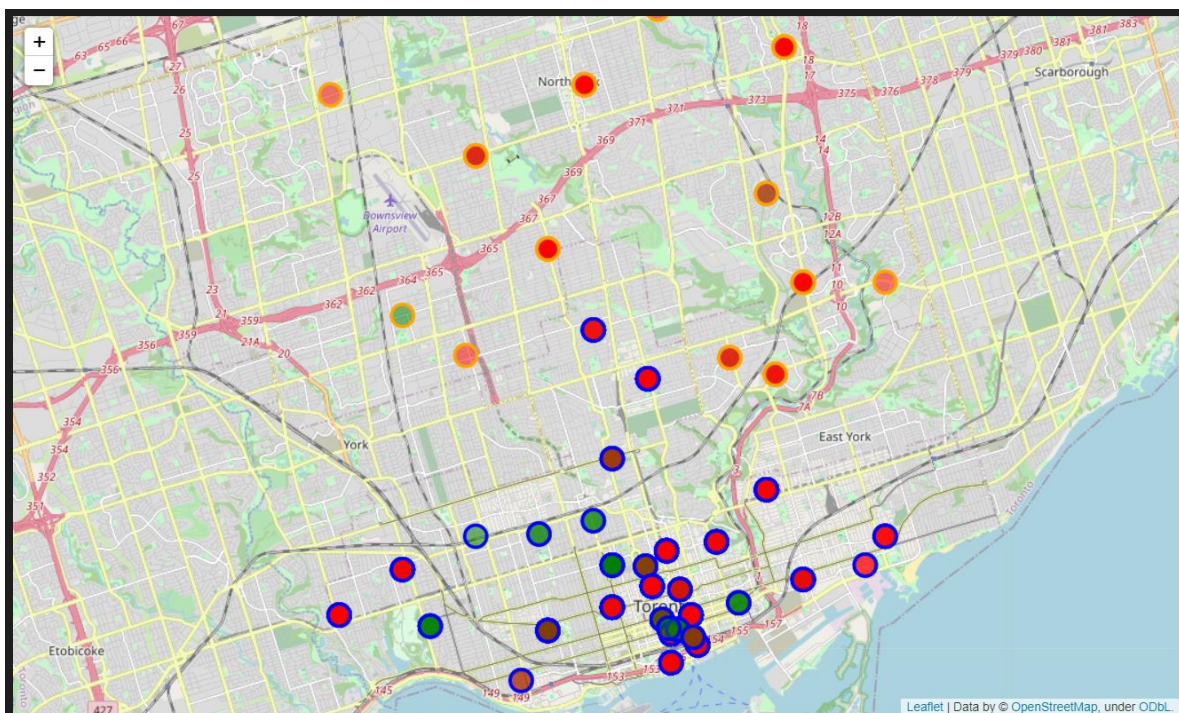
|   | Borough | Neighborhood                                | Neighborhood Latitude | Neighborhood Longitude | Venue Name                 | Venue Rating | Cluster Labels |
|---|---------|---|-----------------------|------------------------|----------------------------|--------------|----------------|
| 0 | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | Impact Kitchen             | 8.8          | 1              |
| 1 | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | Cluny Bistro & Boulangerie | 8.3          | 1              |
| 2 | Toronto | Regent Park, Harbourfront                   | 43.654260             | -79.360636             | El Catrin                  | 7.9          | 1              |
| 3 | Toronto | Queen's Park, Ontario Provincial Government | 43.662301             | -79.389494             | Mercatto                   | 8.3          | 1              |
| 4 | Toronto | Central Bay Street                          | 43.657952             | -79.387383             | Mercatto                   | 8.3          | 1              |



We can see in the scatter plot that the clusters were divided about the 7.5 rating value. Besides, plot indicates that York venues are most rated with low score, while Toronto restaurants are the best ones.



In next map, restaurant of Toronto (blue perimeter) and York (perimeter) are marked with green and red fills. Green corresponds to group with rating greater than 7.5 and red fill to restaurants with rating less than 7.5. We can see that most approved restaurants are in Toronto.



## 4. Results and Discussion

Results obtained in the analysis, following the methodology, demonstrate that most venues are in Toronto region, which give us an idea of than we will hope about the rating of restaurants between York and Toronto. In this case, we can think that the rating average could to be a little more high than the York restaurants, because will have more people in Toronto who can put a rating, and at same time, the rating will be more trustworthy.

In below cells, we can see clearly that Toronto has 266 approved restaurants, that is the venues count in that region. And, in the York side, we have only 10 restaurants to review. Toronto has the advantage. Also, we have the restaurants mean rating of each region, where Toronto has 8.09 and York 7.74 points. Difference is not significantly, but Toronto's mean rating is more trustworthy because it has the support of more consumers.

| Cluster Labels |     |
|----------------|-----|
| Borough        |     |
| Toronto        | 266 |
| York           | 10  |

| Venue Rating |         |
|--------------|---------|
| Borough      |         |
| Toronto      | 8.08985 |
| York         | 7.74000 |

## 5. Conclusion

If you need to choose a Borough to live, Toronto o York, maybe Toronto is the best option for you because you can find the best and most varied restaurants in the region. People say through their ratings that in Toronto are the best restaurants that you can find in that region.