Exploring Toronto

1. Introduction

1.1 Problem

A pizza restaurant owner has two restaurants located in Manhattan, in the Carnegie Hill and Hamilton Heights Neighborhoods, and he wants to open a new restaurant in Toronto. He is investigating more about the neighborhoods were his restaurants are located, in order to have a profile of the neighborhood he is looking for in Toronto. What the owner needs is to find similar neighborhoods in Toronto where he can open his new restaurant.

2. Data Acquisition

2.1 Data description

Neighborhood profile will be created from places of interest in each neighborhood, this will help us to explain to the owner why people is going to these neighborhoods. To perform this search, we will use the Foursquare API, but first is necessary to obtain the Neighborhood coordinates.

2.2 Data sources

New York neighborhood information can be found on the following link on the internet, in JSON format: https://geo.nyu.edu/catalog/nyu_2451_34572

Toronto Neighborhoods can be found on Wikipedia:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The places we are looking for Neighborhood profile will be obtained using Foursquare API, and to perform this we will need the coordinates of each Neighborhood.

2.3 Data cleaning

Data cleaning is necessary to format all the information in Pandas data frame structure. All the description of how this was achieved will be described in this section.

2.3.1 New York Data

New York data is obtained in JSON Format. So, the first step is to transform from JSON format to a Pandas Data frame structure.

The neighborhoods where the pizza restaurants are located are marked in red in the following map of neighborhoods in Manhattan.

```
{'type': 'Feature',
 'id': 'nyu_2451_34572.1',
 'geometry': {'type': 'Point',
  'coordinates': [-73.84720052054902, 40.89470517661]},
 'geometry_name': 'geom',
 'properties': {'name': 'Wakefield',
  'stacked': 1,
  'annoline1': 'Wakefield',
  'annoline2': None,
  'annoline3': None,
  'annoangle': 0.0,
  'borough': 'Bronx',
  'bbox': [-73.84720052054902,
   40.89470517661,
   -73.84720052054902,
   40.89470517661]}}
```


Figure 1. JSON Format of NY Neighborhoods information

To perform this structure transformation, we will use libraries in Python that will help us to extract the information needed:

1. Borough
2. Neighborhood
3. Latitude coordinate
4. Longitude coordinate

And after all the processing we end up with the following data structure:

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

Table 1. Data frame of NY Neighborhoods.

After keeping only, the neighborhoods of Manhattan, these are printed in a map of Manhattan, giving a different color to the two neighborhoods that are in our interest.

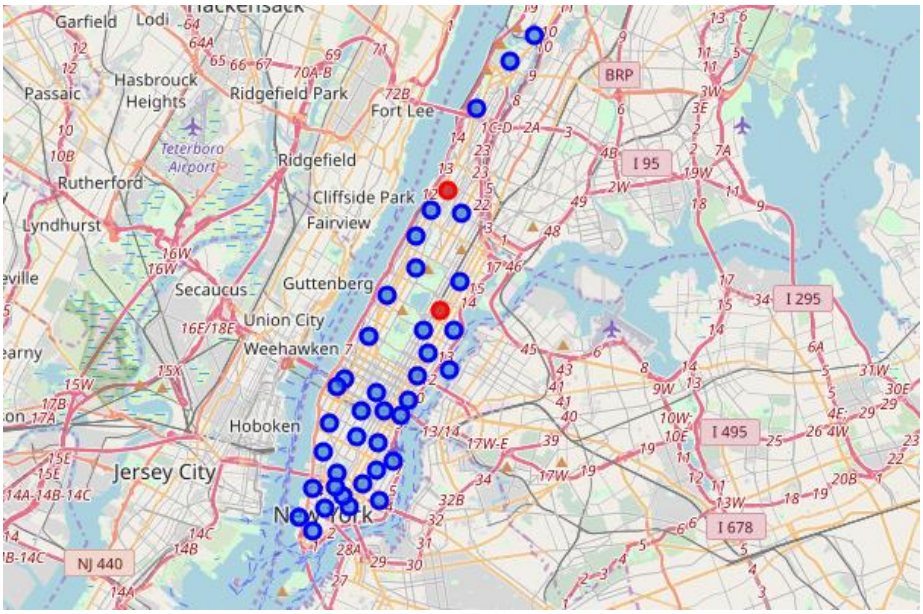

Figure 3. Map of Manhattan Neighborhoods

After this step is necessary to keep the necessary information for the purpose of this research, and this will be keeping the information of the neighborhoods Carnegie Hill and Hamilton Heights. After this we will use Foursquare API to obtain the top 100 venues per neighborhood and we will look at the data in order to determine how the places will help us to describe each neighborhood.

The parameters for getting the list of the places in each neighborhood will be the top 100 venues in a radius of 500 meters. After request and format the data provided by Foursquare (JSON format).


```
{'meta': {'code': 200, 'requestId': '5cdc7c4d4c1f6753b174bff2'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Marble Hill',
  'headerFullLocation': 'Marble Hill, New York',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 25,
  'suggestedBounds': {'ne': {'lat': 40.88105078329964,
    'lng': -73.90471933917806},
   'sw': {'lat': 40.87205077429964, 'lng': -73.91659997808156}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4b4429abf964a52037f225e3',
       'name': "Arturo's",
       'location': {'address': '5198 Broadway',
```


Figure 4. Information obtained from Foursquare API

We will use Python libraries to obtain the information needed from each venue, such as:

- Venue
- Venue Latitude
- Venue Longitude
- Venue Category

The quantity of venues for each neighborhood is showed in the following image:

Neighborhood	Venue
Carnegie Hill	100
Hamilton Heights	60

Table 2. Quantity of venues per neighborhood

2.3.2 Toronto

Toronto information is obtained in HTML format, so after processing the information, the list of Neighborhoods per Postal Code is obtained.

	Borough	Neighborhood	Latitude	Longitude
97	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280
99	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558

Table 3. Toronto neighborhoods

3. Methodology

As explained before, the venues in the neighborhood will help us to describe the neighborhood. So, first we will explore what type of venues are the most common in each neighborhood. To perform this, we will use the data frame with all the venues per neighborhood that was obtained in previous steps.

```
----Carnegie Hill----
                 venue  freq
0          Pizza Place   6.0
1          Coffee Shop   5.0
2                 Café   4.0
3       Cosmetics Shop   4.0
4          Yoga Studio   3.0
5        Grocery Store   3.0
6                  Gym   3.0
7  Japanese Restaurant   3.0
8    French Restaurant   3.0
9            Bookstore   3.0
```

Table 4. Carnegie Hill top 10 venues

We observe that pizza place occupies the first place in the top10 venues list in Carnegie Hill. And we also observe that the 50% of the top10 list are restaurants.

We also plot the quantity of venues in the neighborhood to see the distribution:

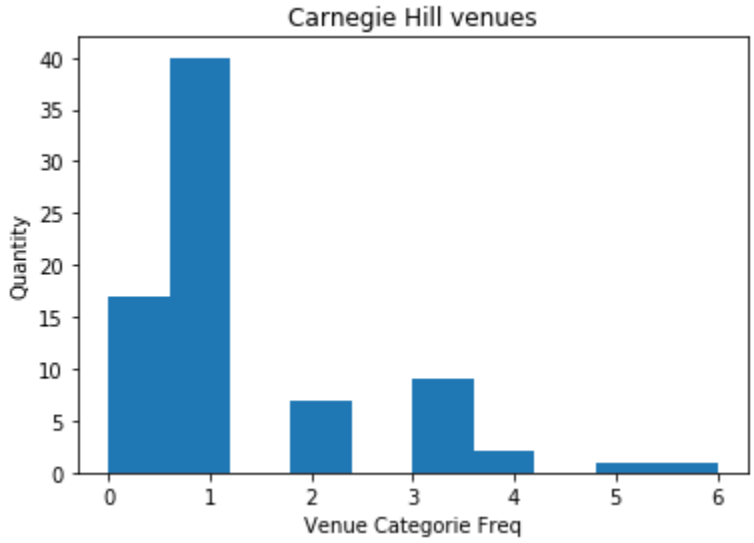

Figure 4. Carnegie Hill Venue Categories Frequency distribution

We observe that 15 venue categories are not present in this neighborhood, 40 categories appear only one time, 6 venue categories appear twice, almost 10 categories appear 3 times, and less than 5 categories appears 4, 5 and 6 times, each.

```
----Hamilton Heights----
                venue  freq
0  Mexican Restaurant   5.0
1         Pizza Place   4.0
2         Coffee Shop   4.0
3               Café   4.0
4   Indian Restaurant   2.0
5  Chinese Restaurant   2.0
6    Sushi Restaurant   2.0
7              School   2.0
8      Sandwich Place   2.0
9        Liquor Store   2.0
```

Table 5. Hamilton Heights top 10 venues

And we plot the histogram to see the distribution in Hamilton Heights:

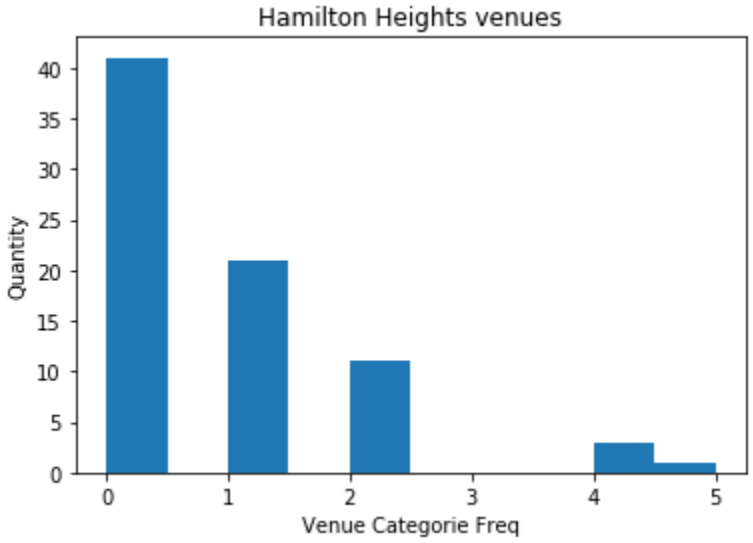

Figure 5. Hamilton Heights Venue Categories Frequency distribution

We observe that 40 venue categories are not present in this neighborhood, 20 categories appear only one time, 10 venue categories appear twice, and less than 5 categories appears 4, and 5 times, each.

In Hamilton Heights we have pizza places in second place, and for this neighborhood, 80% of the top 10 venues are restaurants, so practically what we are looking for is neighborhoods with these similarities.

The next step is to add these two neighborhoods to the list of neighborhoods in Toronto, and we will use unsupervised learning algorithm k-means clustering, in order to cluster all the neighborhoods and we will observe which neighborhoods get clustered with these two neighborhoods.

	Borough	Neighborhood	Latitude	Longitude
97	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280
99	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558
0	Manhattan	Hamilton Heights	40.823604	-73.949688
1	Manhattan	Carnegie Hill	40.782683	-73.953256

Table 6. Complete neighborhoods information

We will perform the same processing we made on the first two neighborhoods, first we will get the top 100 venues in a radius of 500 meters using Foursquare API and then we will count how many venues categories we encounter in each neighborhood.

	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Steakhouse	American Restaurant	Thai Restaurant	Hotel	Burger Joint	Bar	Gym	Bakery
1	Berczy Park	Coffee Shop	Cocktail Bar	Steakhouse	Café	Cheese Shop	Restaurant	Bakery	Farmers Market	Seafood Restaurant	Beer Bar
2	Brockton, Exhibition Place, Parkdale Village	Coffee Shop	Café	Breakfast Spot	Burrito Place	Restaurant	Bar	Stadium	Furniture / Home Store	Italian Restaurant	Climbing Gym
3	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Yoga Studio	Auto Workshop	Garden Center	Garden	Fast Food Restaurant	Farmers Market	Comic Shop	Park	Recording Studio
4	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Service	Airport Lounge	Airport Terminal	Plane	Boat or Ferry	Sculpture Garden	Boutique	Airport Gate	Airport	Harbor / Marina
5	Cabbagetown, St. James Town	Coffee Shop	Bakery	Italian Restaurant	Café	Restaurant	Market	Pub	Pizza Place	Pet Store	Pharmacy
6	Carnegie Hill	Pizza Place	Coffee Shop	Cosmetics Shop	Café	Yoga Studio	Bookstore	Japanese Restaurant	Spa	Gym	Grocery Store
7	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Burger Joint	Japanese Restaurant	Ice Cream Shop	Middle Eastern Restaurant	Spa	Bar	Sandwich Place
8	Chinatown, Grange Park, Kensington Market	Café	Vegetarian / Vegan Restaurant	Bar	Coffee Shop	Bakery	Mexican Restaurant	Dumpling Restaurant	Chinese Restaurant	Vietnamese Restaurant	Burger Joint

Table 7. Most common venues in Toronto Neighborhoods, including currently pizza neighborhoods

4. Results

After applying the unsupervised learning algorithm (k=4), we observe the distribution:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels
2	Downtown Toronto	Harbourfront, Regent Park	43.654260	-79.360636	2
9	Downtown Toronto	Ryerson, Garden District	43.657162	-79.378937	3
15	Downtown Toronto	St. James Town	43.651494	-79.375418	1
19	East Toronto	The Beaches	43.676357	-79.293031	0
20	Downtown Toronto	Berczy Park	43.644771	-79.373306	2
24	Downtown Toronto	Central Bay Street	43.657952	-79.387383	1

Table 8. Toronto neighborhoods clustering

The distribution of the clustering the neighborhoods is as follows (This includes the two neighborhoods of Manhattan):

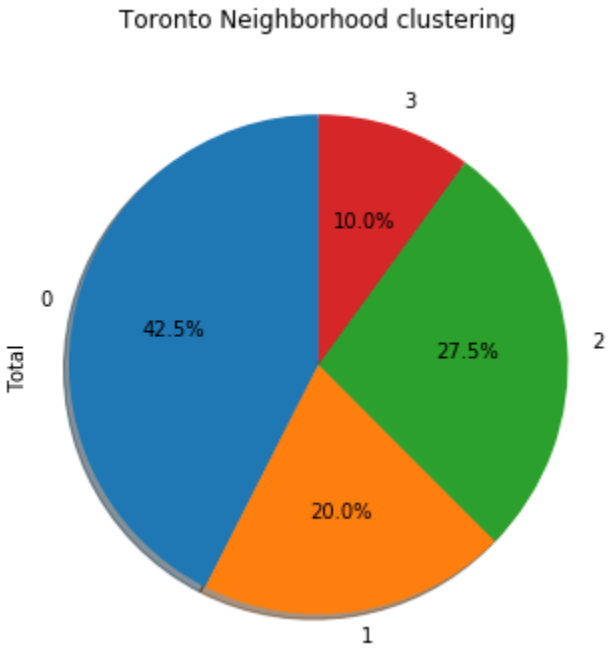

Figure 6. Toronto Neighborhood clustering

The current neighborhoods where the pizza restaurants are located were clustered together with other two neighborhoods of Toronto.

	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
9	Ryerson, Garden District	Coffee Shop	Clothing Store	Café	Cosmetics Shop	Middle Eastern Restaurant	Japanese Restaurant	Tea Room	Sporting Goods Shop	Fast Food Restaurant	Restaurant
84	Chinatown, Grange Park, Kensington Market	Café	Vegetarian / Vegan Restaurant	Bar	Coffee Shop	Bakery	Mexican Restaurant	Dumpling Restaurant	Chinese Restaurant	Vietnamese Restaurant	Burger Joint
0	Hamilton Heights	Mexican Restaurant	Coffee Shop	Pizza Place	Café	Cocktail Bar	Indian Restaurant	Liquor Store	Deli / Bodega	Yoga Studio	Chinese Restaurant
1	Carnegie Hill	Pizza Place	Coffee Shop	Cosmetics Shop	Café	Yoga Studio	Bookstore	Japanese Restaurant	Spa	Gym	Grocery Store

Table 9. Neighborhoods of interest

5. Discussion

After the clustering we can confirm that the algorithm recommends 2 neighborhoods in Toronto for opening the new pizza restaurant. These two neighborhoods have similar venues in them.

Let's observe the top10 venues in each recommended neighborhood:

```
----Chinatown, Grange Park, Kensington Market----
                         venue  freq
0                          Café   8.0
1  Vegetarian / Vegan Restaurant   6.0
2                           Bar   5.0
3                        Bakery   4.0
4            Mexican Restaurant   4.0
5          Dumpling Restaurant   4.0
6                   Coffee Shop   4.0
7         Vietnamese Restaurant   3.0
8            Chinese Restaurant   3.0
9                  Burger Joint   2.0
```

Table 10. Top 10 venue categories in Chinatown, Grange Park and Kensington Market

In the first neighborhood we observe that has 80% venue restaurants in the top 10 venue, which is similar to Hamilton Heights that has 80% restaurants venues in the top 10.

```
----Ryerson, Garden District----
                         venue  freq
0                  Coffee Shop   8.0
1              Clothing Store   7.0
2              Cosmetics Shop   4.0
3                        Café   4.0
4  Middle Eastern Restaurant   3.0
5       Fast Food Restaurant   2.0
6                     Theater   2.0
7            Bubble Tea Shop   2.0
8         Sporting Goods Shop   2.0
9         Japanese Restaurant   2.0
```

Table 11. Top 10 categories in Ryerson, Garden District

For the second neighborhood we observe that 50% of the top10 venue categories has to be with restaurants, similar to Carnegie Hill neighborhood.

Based on this information we can conclude that the clustering was successful and the recommendations for these two neighborhoods can be made to the owner.

6. Conclusion

What we were trying on this problem was to understand the data of each neighborhood based on venues in neighborhood. This could be one valuable resource for making recommendations for the owner, but is necessary to remember that not only the places around matter since the culture of each country can impact in the development of new business. But for the purpose of this problem I am happy with the results.