

# The Battle of Neighborhoods: New York City and Toronto

Soumya Ranjan Behera

1st February, 2021

## Part I: Introduction

In this project, we will be studying about the neighborhoods of two of most multicultural and cosmopolitan cities in the world: New York City (NYC), United States of America and Toronto, Canada. We will be investigating on what kinds of businesses are common in both cities, what kinds of businesses are more common in one of the two cities than the other city, and what kinds of businesses are not common in both cities.

Doing this project will enable us to get a better understanding of similarities and differences between the two cities which will make it known to business people what types of businesses are more likely to thrive in both cities, what are the neighborhoods that are suitable for each type of business, and what types of businesses are not very desirable in each city. This allows business people to take better and more effective decisions regarding where to open their businesses.

## Part II: Data for the Project

In this section, we will be discussing about the dataset used in this project. To be able to do this project, two types of data are needed:

- **Neighborhood Data:** Lists the names of the neighborhoods of NYC and Toronto and their latitude and longitude coordinates. We have this data provided through the "IBM Data Science Professional Certificate" course and also, we need to scrape some data from the internet.
- **Venues Data:** Lists the top 100 venues (restaurants, cafes, parks, museums, etc.) in each neighborhood of the two cities. The data should list the venues of each neighborhood with their categories. This data will be retrieved from Foursquare which is one of the world largest sources of location and venue data. Foursquare API will be utilized to get and download the data.

### **1. Neighborhood Data**

For each city, data that describes the names of its neighborhoods and their coordinates is needed.

#### **1.1 New York City**

A dataset that specifies the neighborhood data for New York City was provided through the “Applied Data Science Capstone” course which is provided by IBM. The dataset is originally a JSON file that specifies the name of each neighborhood, its coordinates—latitude and longitude, its borough, and other data too. The below snippet shows a part of this JSON file.

```

1  {"type": "FeatureCollection",
2   "totalFeatures": 306,
3   "features": [{"type": "Feature",
4    "id": "nyu_2451_34572.1",
5    "geometry": {"type": "Point",
6     "coordinates": [-73.84720052054902, 40.89470517661]},
7    "geometry_name": "geom",
8    "properties": {"name": "Wakefield",
9     "stacked": 1,
10    "annoline1": "Wakefield",
11    "annoline2": null,
12    "annoline3": null,
13    "annoangle": "OE-11",
14    "borough": "Bronx",
15    "bbox": [-73.84720052054902, 40.89470517661, -73.84720052054902, 40.89470517661]}},
16   {"type": "Feature",
17    "id": "nyu_2451_34572.2",
18    "geometry": {"type": "Point",
19     "coordinates": [-73.82993910812398, 40.87429419303012]},
20    "geometry_name": "geom",
21    "properties": {"name": "Co-op City",
22     "stacked": 2,
23     "annoline1": "Co-op",
24     "annoline2": "City",
25     "annoline3": null,
26     "annoangle": "OE-11",
27     "borough": "Bronx",
28     "bbox": [-73.82993910812398, 40.87429419303012, -73.82993910812398, 40.87429419303012]}},

```

Part of the JSON file containing NYC Neighborhood data

To be able to use the data of this JSON file in the later parts of this project, it would be processed and stored in a Pandas dataframe. In total, the JSON file contains data on 306 neighborhoods.

## 1.2 Toronto

For Toronto, there is no dataset that contains all the needed neighborhood data as was the case for NYC. In the “Applied Data Science Capstone” course, a dataset was provided mapping the Toronto postal codes to their respective latitude and longitude coordinates. Below is a snippet showcasing the same.

Postal Code	Latitude	Longitude
M1B	43.8066863	-79.1943534
M1C	43.7845351	-79.1604971
M1E	43.7635726	-79.1887115
M1G	43.7709921	-79.2169174
M1H	43.773136	-79.2394761
M1J	43.7447342	-79.2394761
M1K	43.7279292	-79.2620294
M1L	43.7111117	-79.2845772
M1M	43.716316	-79.2394761

**Toronto's Postal Codes with their coordinates**

We will be needing another dataset enlisting the names of the neighborhoods and their postal codes which can be used in combination with the above dataset to produce the desired results. There is a Wikipedia page titled “List of postal codes of Canada: M” containing the information about the postal codes with neighborhood and borough name associated with it. The postal codes in Canada that starts with the letter M are the postal codes of Toronto city. We will be extracting the relevant data from the webpage with the Pandas `read_html()` function. The URL for the Wikipedia page is ‘[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)’.

Postal Code ↕	Borough ↕	Neighbourhood ↕
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights

**Toronto's Postal Codes with Neighborhood and Borough names from Wikipedia**

In the above list on the Wikipedia page, there are 77 records out of 180 where the “Borough” variable has the value “Not assigned”; for these 77 records, the “Neighborhood” variable also has the value “Not assigned”; In the above figure, the first two rows shows some examples

of these records. Thus, these records will be deleted because they don't carry meaningful information regarding Toronto neighborhoods.

## **2. Venues Data**

For both of the cities, we need data that provides information about the different venues in a particular neighborhood and the categories of these venues. Venues data will be retrieved from Foursquare which is a popular source of location data. Foursquare API service will be utilized to access and download venues data.

To retrieve data from Foursquare using their API, a URL should be prepared and used to request data related a specific location. An example URL is the following:

```
https://api.foursquare.com/v2/venues/search?&client_id=1234&client_secret=1234&v=20180605&ll=40.89470517661,-73.84720052054902&radius=500&limit=100
```

where search indicates the API endpoint used, client\_id and client\_secret are credentials used to access the API service and are obtained when registering for Foursquare developer account, v indicates the API version to use (i.e, it should be the current date), ll indicates the latitude and longitude of the desired location, radius is the maximum distance in meters between the specified location and the retrieved venues, and limit is used to limit the number of returned results if necessary.

A function is created that takes as input the names, latitudes, and longitudes of the neighborhoods, and returns a dataframe with information about each neighborhood and its venues. It creates an API URL for each neighborhood and retrieves data about the venues of that neighborhoods from Foursquare. Below is a screenshot of the data after retrieving venue data for each neighborhood from Foursquare.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

**Processed sample venue data for a neighborhood retrieved from Foursquare**

## 2.1 New York City

From Foursquare, the data retrieved for the NYC contained more than 23,000 venues in it. For each venue, venue name, category, latitude, and longitude were retrieved. Different numbers of venues were found in different neighborhoods: For example, data about 81 venues were returned for Williamsburg, 86 venues for Yorkville neighborhood and 57 venues for Midtown. Each venue belongs to one of 578 unique categories.

## 2.2 Toronto

Similar to what has been done for NYC, data was retrieved from Foursquare describing the venues present in the different Toronto neighborhoods. We got information about 7,700 venues in Toronto. Different numbers of venues were found in different neighborhoods: for example, data about 84 venues were returned for Weston neighborhood and 70 venues in Lawrence Park neighborhood. Each venue belongs to one of 504 unique categories.