Datascience Capstone Project

The battle of neighbourhoods: Toronto vs New York

# Rohan Tiwari

# April 2020

Datascience Capstone Project

The battle of neighbourhoods: Toronto vs New York

# **Problem Statement**

# This work explains how the neighbourhood of Toronto is different, or similar to that of New York City. We will get to see the most common venues for both.

**Project Background:**

Toronto and New York being the financial capital of Canada and the US respectively are one of the densely populated cities in the world. Both these cities see a lot of diversity resulting from the movement of a lot of immigrants from several parts of the world for work and settlement. These are one of the most immigrant-friendly cities, still different in so many aspects, which we are going to observe in this work.

The purpose of this whole exercise is for submission of the final capstone project for the "IBM Data Science" course on Coursera as well as to showcase my data science skills in the real-world application.

**Data and Resources to be used**

1. List of postal codes of Canada Wiki: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> for access to neighbourhood data of Toronto region.

2. List of postal codes of New York Wiki <https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm> for access to neighbourhood data of New York region.

3. Geographical coordinates of the neighbourhoods: <http://cocl.us/Geospatial_data> for getting the longitude and latitude data for the neighbourhoods.

4. Foursquare database: [https://Foursquare.com](https://foursquare.com/) to be used in order to explore the desired neighbourhood data for various restaurant details and access the JSON files. This data shall be utilised to map the Indian restaurants in various locations.

**Problem solving strategy:**

The idea is to compare Toronto and the New York region for their similarities difference in terms of the most common venues they house. I will specifically compare the number of Indian, Chinese, and American Restaurants in both cities as well as list down the 10 most common venues in both cities Neighbourhood wise. The outcome of this study will help tourists and new immigrants have an overview of the common venues in both cities and chalk out the differences between both, which might further help them in their decision of travel or immigration choice.

Step-wise approach of problem-solving:

**Step-1:**Web scraping of the neighbourhood data from postal codes of Canada Wiki-link. Clean the data by removing the missing values and store the data in a python Dataframe consisting of three columns namely: PostalCode, Borough, and Neighbourhood

**Step-2:**Web scraping of the neighbourhood data from postal codes of New York Wiki-link. Clean the data by removing the missing values and store the data in a python Dataframe consisting of three columns namely: PostalCode, Borough, and Neighbourhood

**Step-3:** Take the help of long-lat data from the geospatial data wikilink and append the geographical coordinates in the above dataframes to get new respective dataframes for further analysis.

**Step-4:** Getting location data using the Foursquare API. It will be used to retrieve information of the common venues in Toronto and New York neighbourhoods. The API will return a JSON file which will be further converted into a Python Dataframe.

**Step-5:** Exploratory Data Analysis

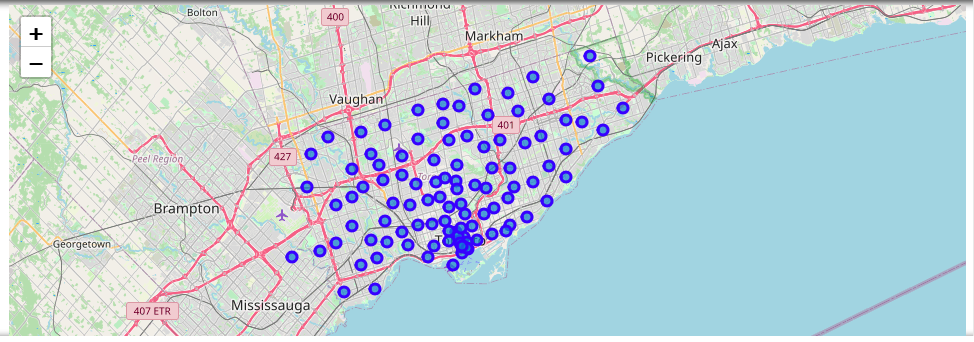
• Plot the prepared location data on the map for visualisation purposes.

• Plot and find the relationship between Neighbourhoods and Various Restaurants for both cities.

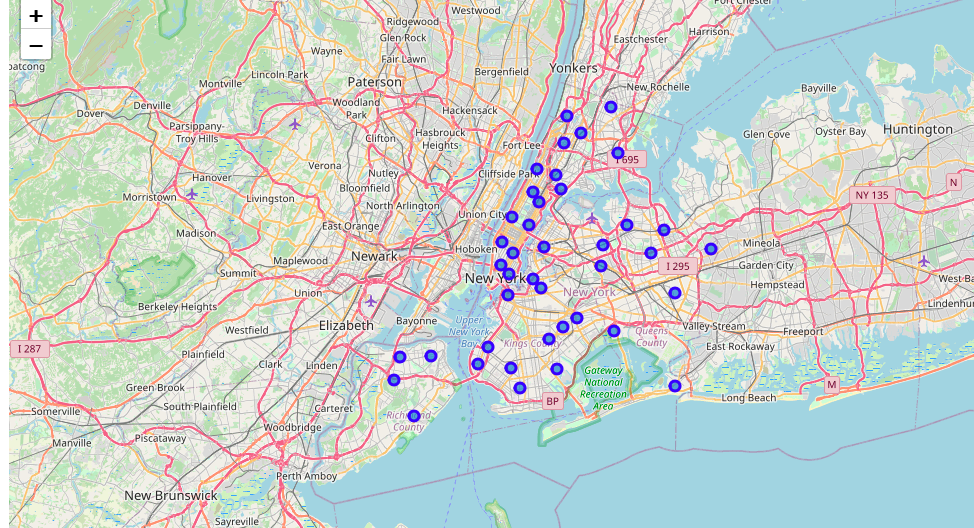
• List down the 10 most common venues for both cities.

**Step-6:**Lastly we will discuss the results based on the above findings and provide a snapshot of both cities which will help travellers and immigrants in making their choice.

A visual glimpse of the neighbourhood of Toronto and New York under the analysis, created using Folium Maps:

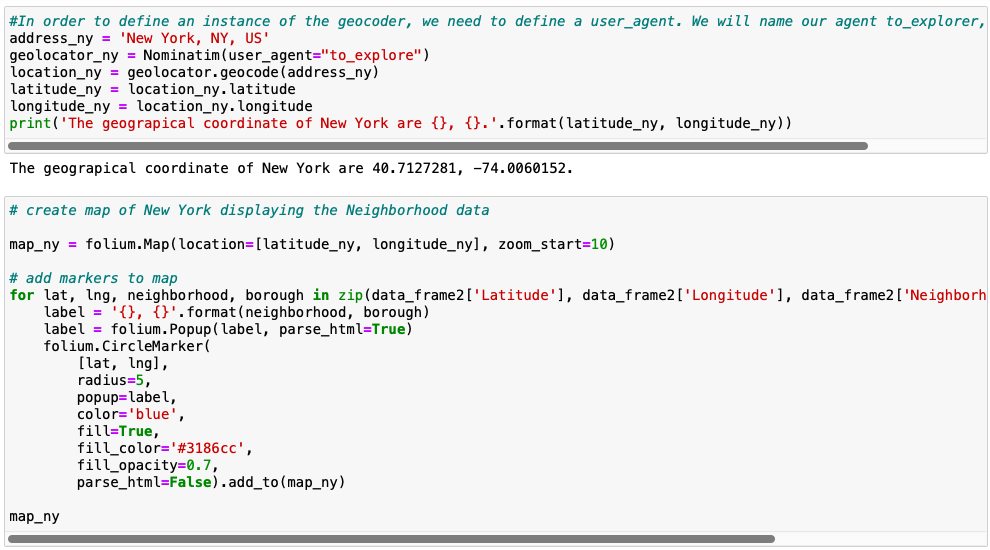


Toronto Neighborhood

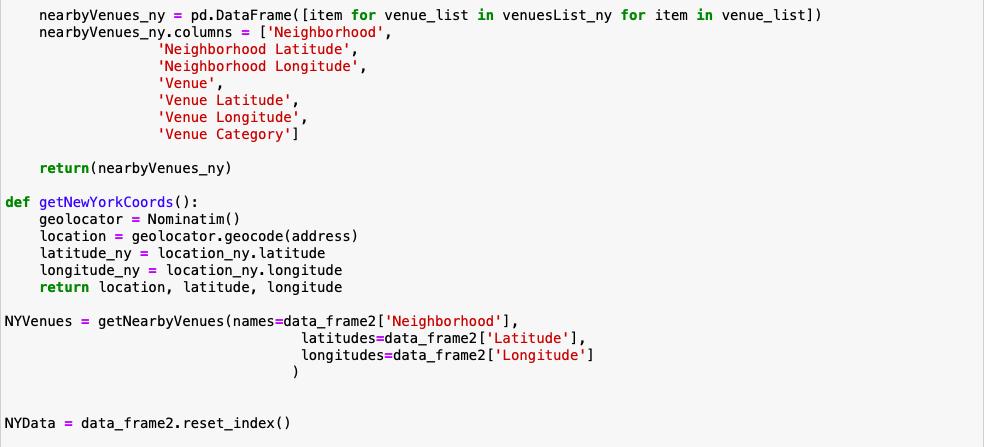
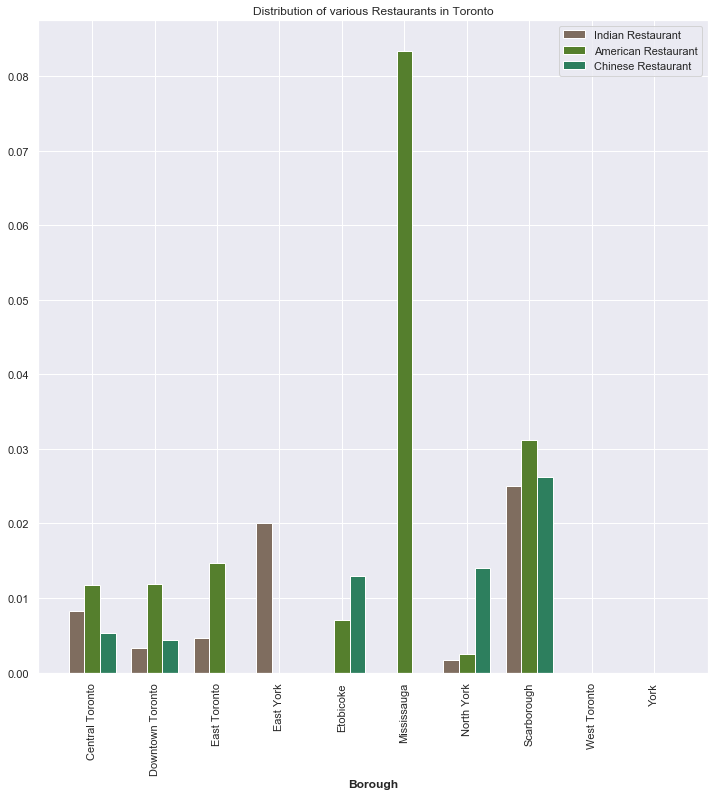


New York Neighbourhood

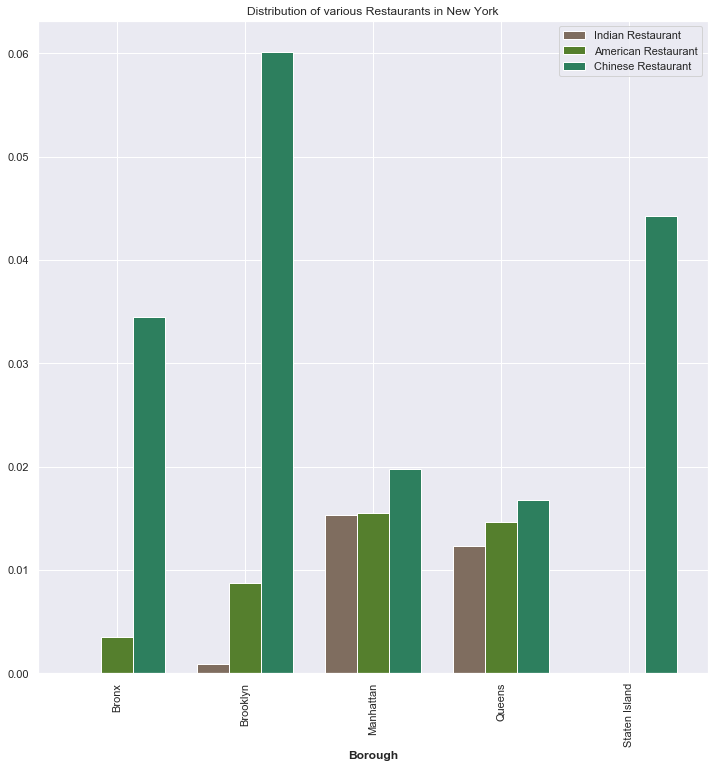
These interactive maps were created using Geolocator and Folium Maps. Sharing a snapshot of the code below:



Webscraping was done using the Foursquare API with the following code:

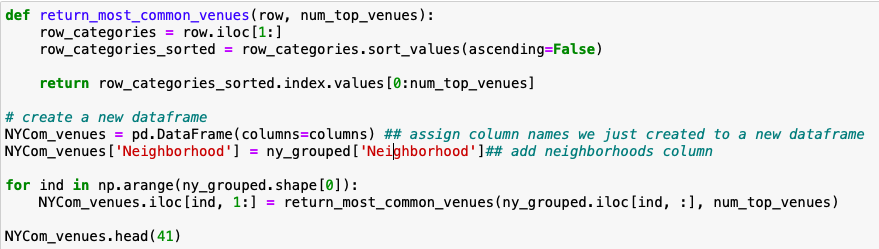
Distribution of various restaurants in Toronto Region (plotted with help of matplotlib library)

In Toronto City most Indian Restaurants could be found in Central Toronto, Scarborough, or East York, while Mississauga has a bunch of American cuisine restaurant options, and for Chinese cuisines, North York and Scarborough are the best regions to look for.

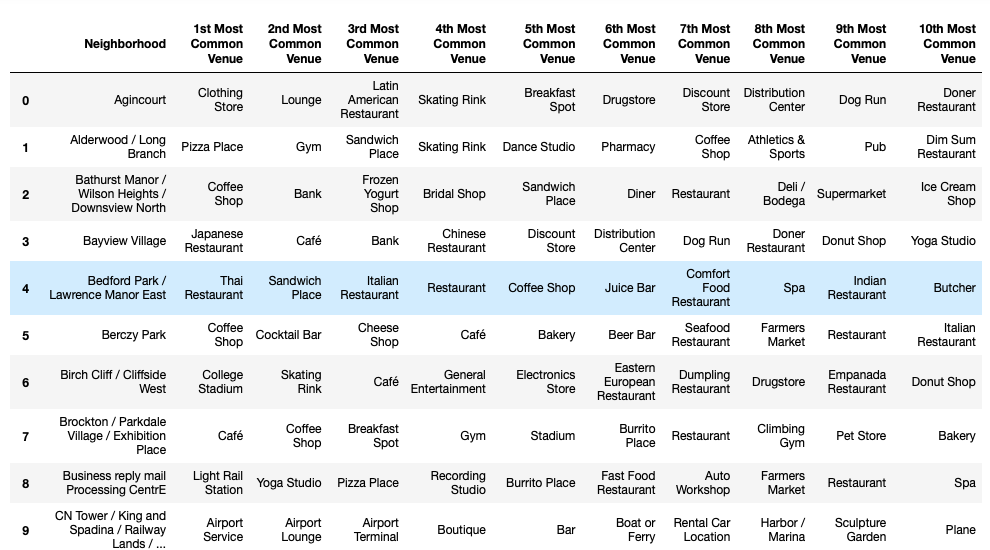


In New York City both Indian and American Restaurants are available in Manhattan and Queens area, however most Chinese cuisine restaurants are available in Brooklyn and Staten Island area.

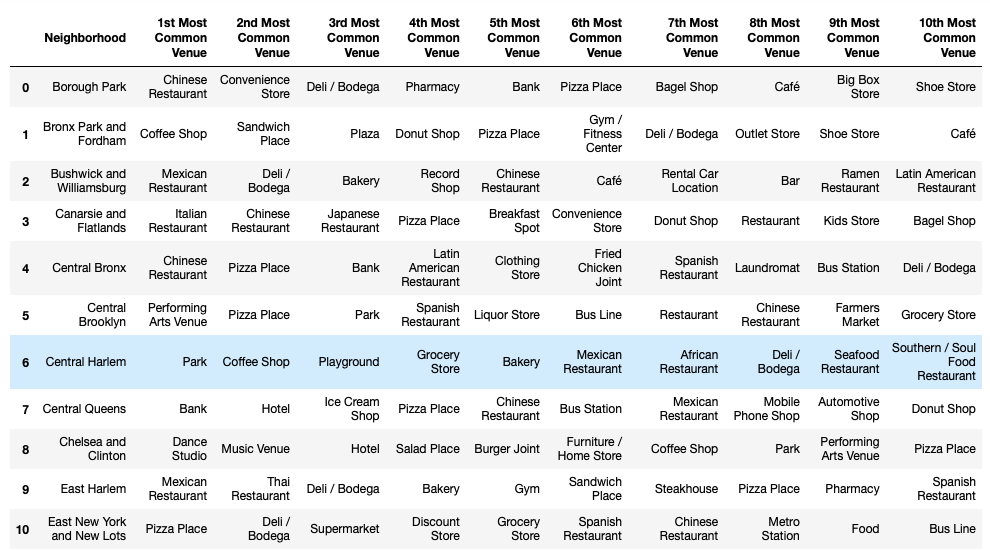
**I have also deduced a list of 10 most common venues for both cities:**

**This was done with the following code:**

# Most common venues for Toronto City



# Most common venues for New York City



# **Final Conclusion**

After carefully going through the most common venues data above one can find out the most common venues in the desired neighborhood area of the respective city. This data might help travelers and new immigrants to find a place to stay or hang-out.