# The Battle of Neighbourhoods (New York vs Toronto) Data Analysis Report

### Author S Cavin

## Introduction

The objective of this report is to describe the findings from the data analysis of two important cities in the world, New York, and, Toronto, and their neighbourhood facilities/amenities provided by their local businesses.

Both cities share similar characteristics, such as:

* Large cities in two major countries in North America, Canada and United States of America;
* Toronto is the capital of Ontario, and the most populous city in Canada, with a population of 2,731,571 in 2016;
* Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world;
* New York is the most populous city in the United States of America, with an estimated 2018 population of 8,398,748;
* New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports;

The findings will enable to understand, their neighbourhoods similarities and differences, the top businesses that are flourishing (in numbers) in those cities, an understanding of the population and their behaviour pattern, along with type of businesses that are popular (and the least ones) in their neighbourhoods.

## Data

The data, used for this analysis, is described for each city, as below.

### For New York (United States of America)

The dataset is provided by coursera through a web link for download (<https://cocl.us/new_york_dataset>); the origin of this dataset comes from a free web link available in the URL location <https://geo.nyu.edu/catalog/nyu_2451_34572>

This dataset stores information of the New York neighbourhoods, which are 5 boroughs in total and 306 neighbourhoods. The data is required to be segmented and processed before staring the intended analysis. The dataset also contains the latitude and longitude coordinates of each neighbourhood.

The location (latitude, longitude), is important as it helps in the analysis, due to its use by the Foursquare site, who is the provider of insight information of each neighbourhood regarding top businesses, amenities provided in the neighbourhood, used by the local population.

For the Foursquare site, a subscription is used to obtain relevant data for the neighbourhood analysis.

### For Toronto (Canada)

The dataset is provided by Wikipedia in the following web link

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

This webpage contains information of the postcodes and neighbourhood of Toronto. Although this data is openly available, it requires to be extracted from the Wikipedia webpage in a suitable data frames format, this to be used for the city’s data analysis.

For the extraction of the relevant information from the Wikipedia web page, a library is used called BeautifulSoup, which facilitates the extraction of data.

Once the dataset is extracted with information of neighbourhood of Toronto, an important piece of information for each neighbourhood is still required to accomplish the data analysis task for the city, which is the longitude, and latitude. This information is provided by coursera through an open URL access, which is located in <http://cocl.us/Geospatial_data>

A merge of data should occur between the information provided by the web page in Wikipedia, and the geospatial data provided by course.

After the merge, the data will be suitable to be used for the data analysis tasks, and enable the comparison with the findings of the New York City.

Similar to the New York description, the location (latitude, longitude), is important as it will help in the analysis, due to its use by the Foursquare site, that provide insight information of each neighbourhood regarding top businesses, amenities provided in the neighbourhood, used by the local population.

For the Foursquare site, a subscription use to obtain relevant data for the neighbourhood analysis.

## Methodology section

*(which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why)*

## Results section

*(where you discuss the results)*

## Discussion section

*(where you discuss any observations you noted and any recommendations you can make based on the results)*

## Conclusion section

*(where you conclude the report)*