

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 neighbourhood 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.

The target audience for this report are, city business developers, local authorities, local people (to understand the facilities provided by their neighbourhood), people intending to re-locate to these cities, people that look for new business opportunities.

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 starting 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.

Here are samples of the data of New York that will be used for this analysis. See Figure 1, and 2.

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Figure 1. Sample Data of New York Neighbours with their longitude and Latitude

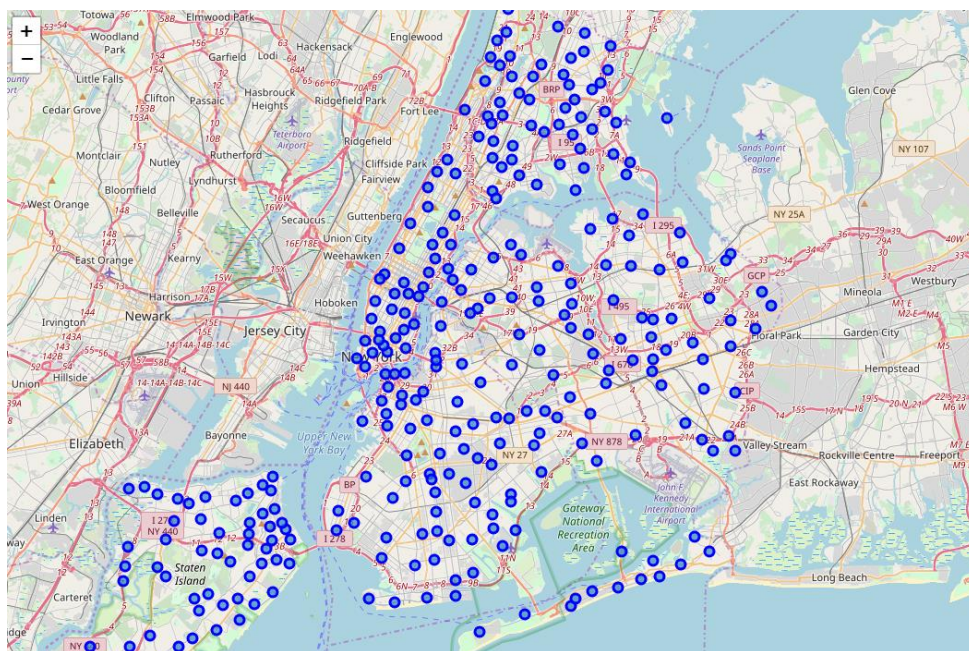


Figure 2. Map showing the New York's neighbourhoods

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.

Here are samples of the data of Toronto that will be used for this analysis. See Figure 3, 4, and 5.

	PostalCode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M5A	Downtown Toronto	Regent Park
4	M6A	North York	Lawrence Heights
5	M6A	North York	Lawrence Manor
6	M7A	Queen's Park	Queen's Park
7	M9A	Etobicoke	Islington Avenue
8	M1B	Scarborough	Rouge

Figure 3. Sample Data of Toronto Neighbours with their postcode

	PostalCode	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

Figure 4. Sample Data of Toronto postcodes with their longitude and Latitude

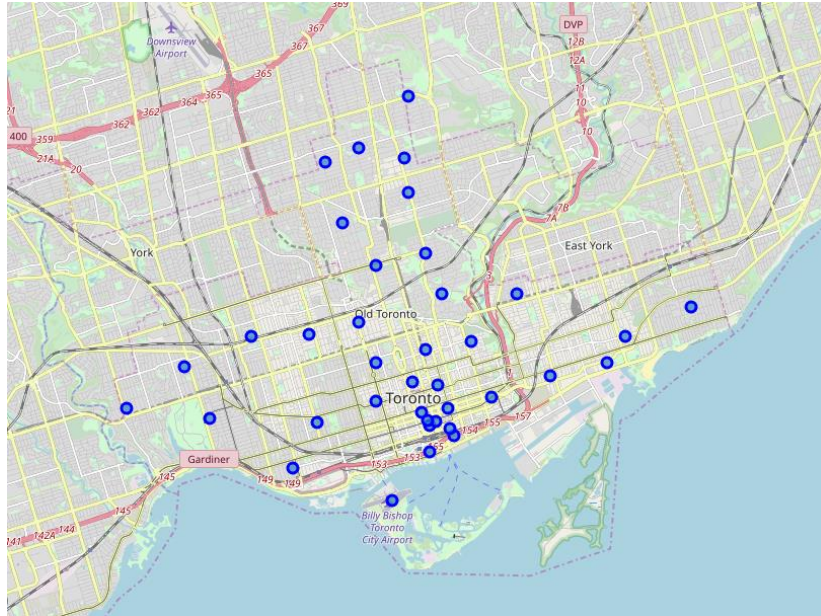


Figure 5. Map showing the Toronto's neighbourhoods

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)