

The Coursera logo, featuring the word "coursera" in a blue, lowercase, sans-serif font.

Applied Data Science Capstone Project: Indonesian Flavours in the West

Abraham Suriadi Halim

December 2020

Executive Summary

The Indonesian Minister of Tourism and Creativity Economy, Wishnutama Kusubandio, believed that one way to increase tourism to Indonesia is by cultivating the Indonesian culture worldwide. One way to do this is by opening Indonesian restaurants. While tourism in Indonesia is packed with most foreigners from Asia, one can take interest in opening such restaurant in the West.

Two cities this paper looks deeper into are Toronto and New York City. They are both diverse cities. Both has its similarities and differences – which leads to the question – which of these two is a better city for one to open an Indonesian restaurant in?

This paper works with the Toronto and New York City data provided by IBM's data science course 'Applied Data Science Capstone'. Using these data and the data science techniques from 'IBM Data Science Professional Certificate', one can find that Toronto is a better city for an Indonesian restaurant to open.

Of course, there are other factors that need to be taken into consideration. However, this paper only focuses on the geographical location of both cities. This is done by utilising Foursquare API to obtain information about the neighbourhoods and its venues for both cities.

Table of Contents

Executive Summary	2
1. Introduction	4
1.1. Background	4
1.2. Problem	4
2. Data	4
3. Methodology	5
4. Results	10
5. Discussion	10
6. Conclusion	10

1. Introduction

1.1 Background

In a recent interview prepared by a well-known Indonesian Doctor of Psychology (Deddy Corbuzier) Wishnutama Kusubandio, the country's current Minister of Tourism and Creative Economy, stated that one way to aid Indonesia's economy is through the propagation of its culture worldwide. One way of doing this is through promoting Indonesian cuisine. He believed that by doing so, this will eventually increase tourism and in effect, aid the nation's economy.

In 2019, research statistics show that the top five sources of visitors to Indonesia are Singapore, Malaysia, China, Australia and Japan – countries in the Australasian region. Meanwhile, travels to Indonesia from North America has been relatively few. New York City and Toronto are two financial capitals in North America and because of this, both locals and immigrants from all over the continent generally aim to work there. This makes both cities diverse. With the geographical interest and diversity that both cities possess, opening an Indonesian restaurant in such location may attract a significant number of customers.

1.2 Problem

The problem that will pose as the main question of this paper is “which city – Toronto or New York – would be a better financial capital for Indonesian restaurants to open?” Knowing both cities' similarities and dissimilarities on a larger scale, this paper seeks to delve deeper into the cities' venues data using Foursquare API to find an answer to this problem.

2. Data

The two main datasets that will be processed are the New York City and Toronto data provided in ‘Applied Data Science Capstone’ course available on Coursera. These contain neighbourhood datasets of both cities. Data analysis and visualisation techniques will be restricted to the materials provided to students throughout the courses listed on ‘IBM Data Science Professional Certificate’. Data such as venues and latitude & longitude values are obtained are from utilising Foursquare API and the GeoPy library on Python. Additional information and/or processed data relating to this paper will be referenced and they can be found at the end of this document.

3. Methodology

After retrieving both New York and Toronto data available from the course, both datasets are initially cleaned up and organised. By doing this, data on both cities' borough, neighbourhood, latitude and longitude are obtained. Firstly, the number of neighbourhoods per borough are visualised. They are shown in Figure 3.1 and 3.2 below.

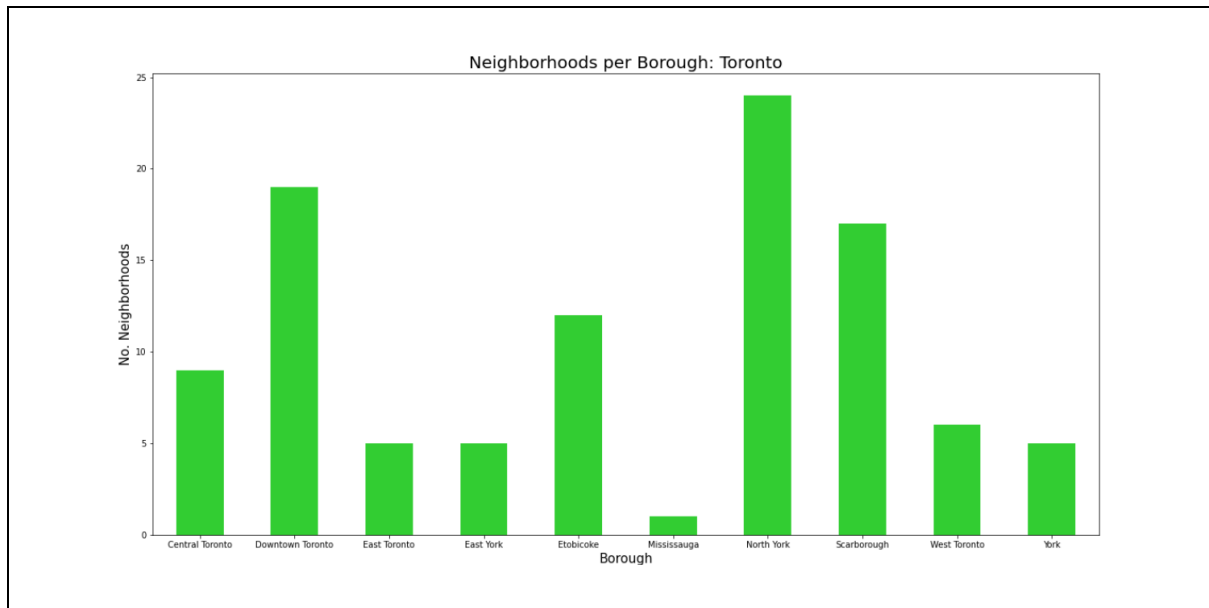


Figure 3.1 – Number of neighbourhoods per borough in Toronto

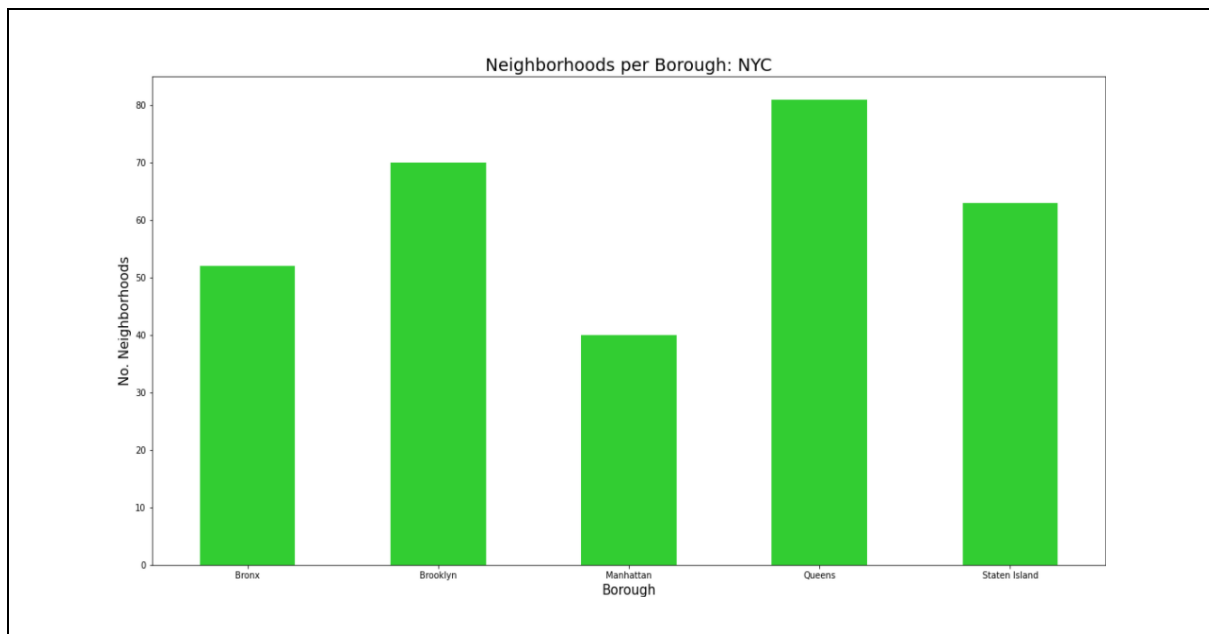


Figure 3.2 – Number of neighbourhoods per borough in New York City

Proceeding, data is further analysed to find the number of Indonesian restaurants and whereabouts in each respective city they are located at. Each neighbourhood's nearby venues are evaluated.

Unnamed: 0	Borough	Neighborhood	ID	Name
------------	---------	--------------	----	------

Figure 3.3 – Venues listed as an ‘Indonesian Restaurant’ in Toronto

Unnamed: 0	Borough	Neighborhood	ID	Name
0	0	Queens	Elmhurst	53c99461498e31473d141ea2 Sky Cafe

Figure 3.4 – Venues listed as an ‘Indonesian Restaurant’ in New York City

Due to the lack of data of Indonesian restaurants in both cities, more exploring had to be done. For now, data of all restaurants in both cities will be analysed. This can be obtained by defining a function that; creates the API request URL, makes the GET request and returns necessary information.

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):
    venues_list=[]

    for name, lat, lng in zip(names, latitudes, longitudes):
        # print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]["groups"][0]["items"]

        # return only relevant information for each nearby venue
        venues_list.append([
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name'] for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                            'Neighborhood Latitude',
                            'Neighborhood Longitude',
                            'Venue',
                            'Venue Latitude',
                            'Venue Longitude',
                            'Venue Category']

    return(nearby_venues)
```

Figure 3.5 – Function to obtain data on nearby venues

After requesting for the data from Foursquare, both datasets are narrowed down in a way that their Venue Categories containing the word ‘restaurant’.

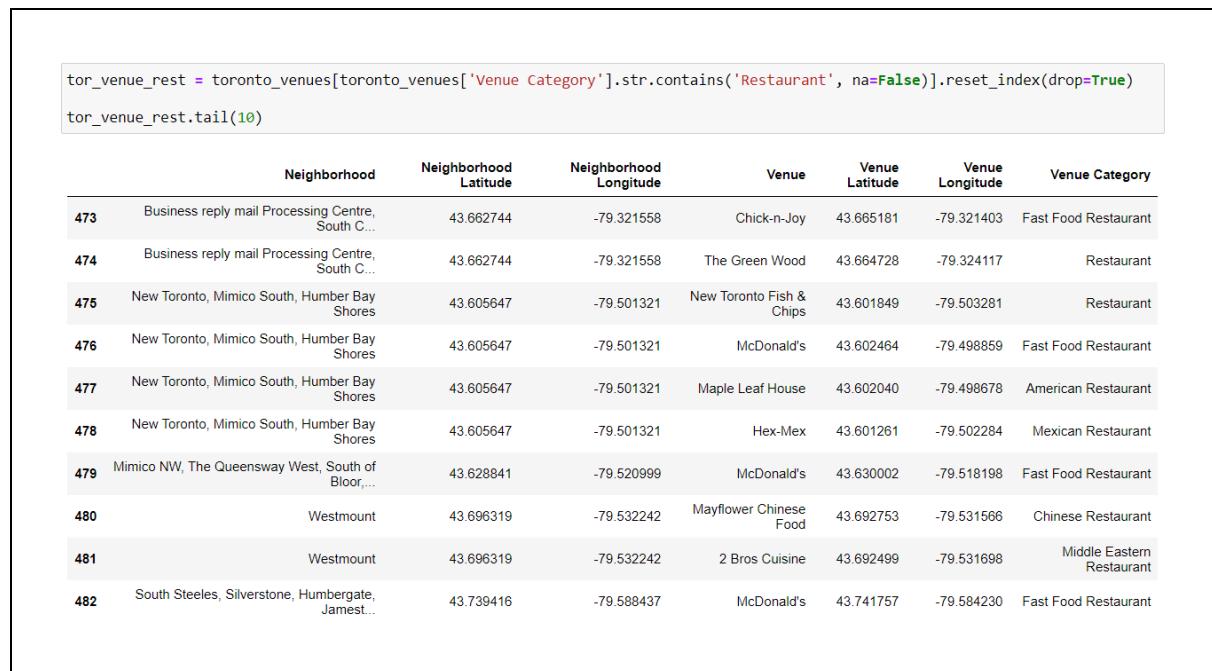


Figure 3.6 – Venues in Toronto containing the word ‘restaurant’

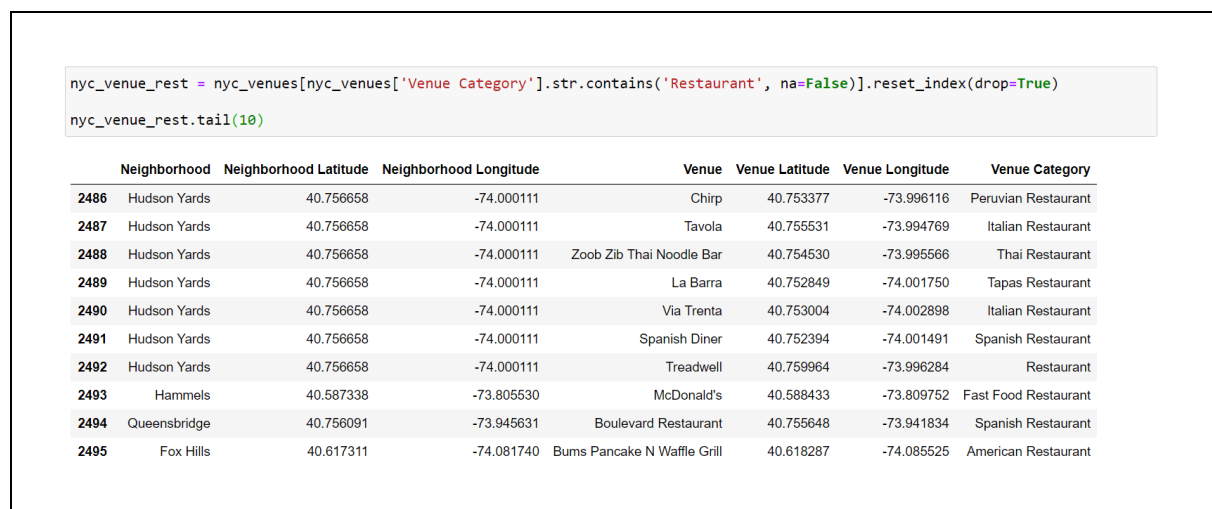


Figure 3.7 – Venues in New York City containing the word ‘restaurant’

With this data, one can carry out one-hot encoding and determine the mean frequency of visit of each restaurant in every neighbourhood for both cities. The figures below show code and the mean frequency of visit to some of each city's neighbourhood.

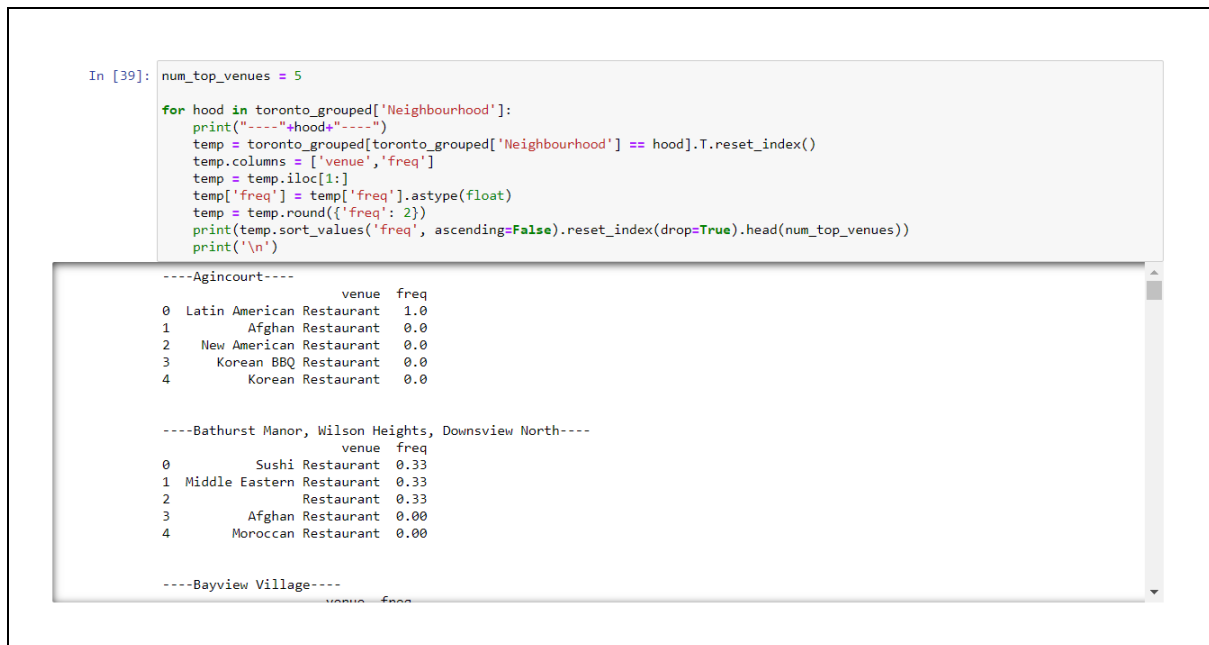


Figure 3.8 – Mean frequency of Visit, Toronto



Figure 3.8 – Mean frequency of Visit, New York City

Continuing, the dataset of each city will now be clustered into 5 clusters. This is done firstly by merging the location data (which includes names of the neighbourhoods, latitude and longitude) with the ‘top venues’ dataset. The ‘top venues dataset’ is created by defining a function that returns the most common venues in each neighbourhood. Folium is then utilised to visualise the clusters in each city.

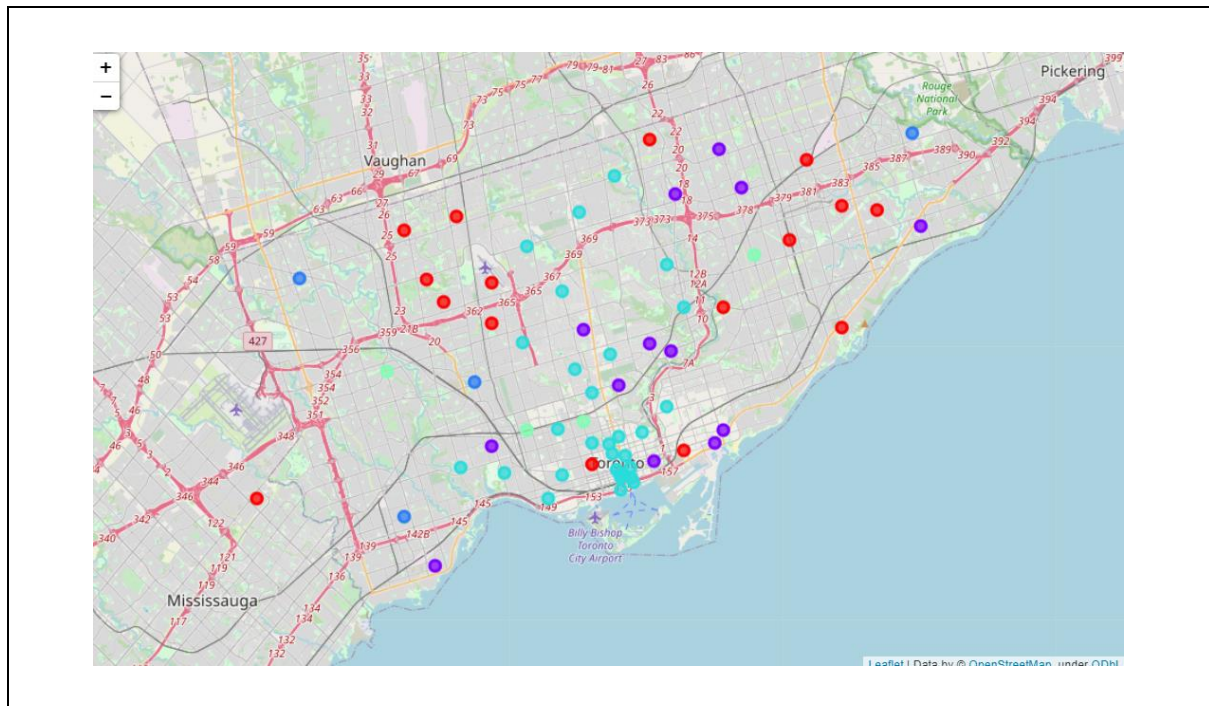


Figure 3.9 – Toronto Folium Map

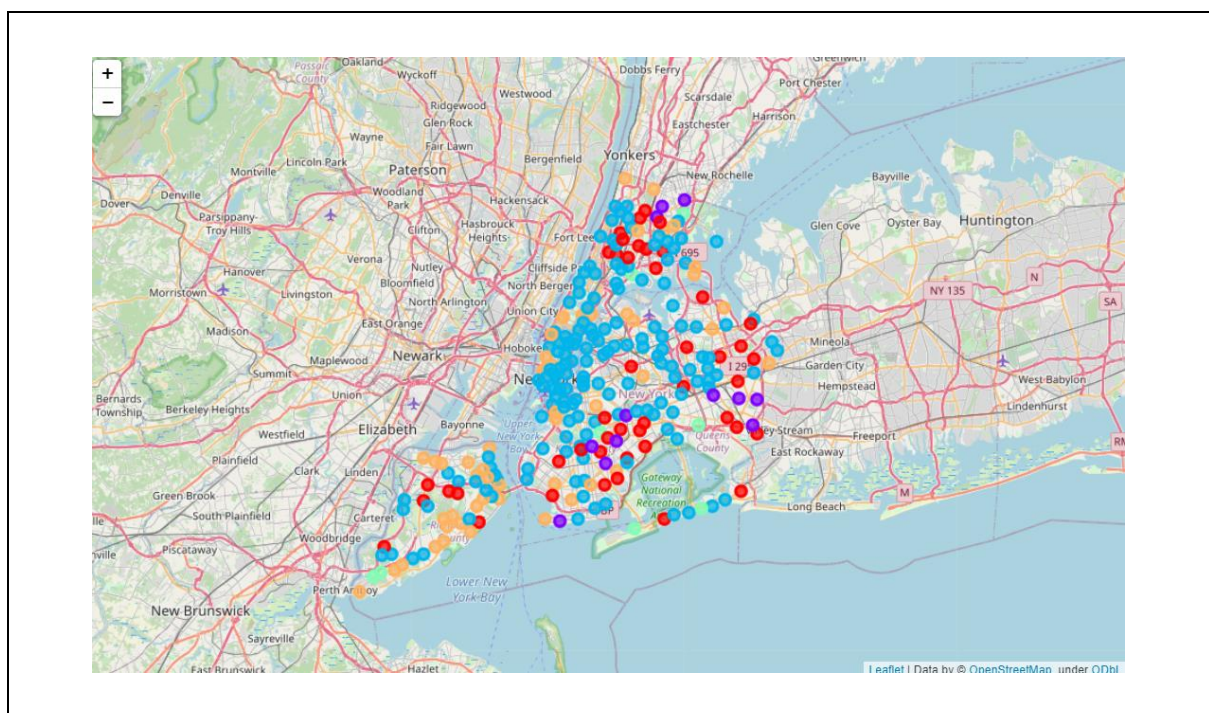


Figure 3.10 – New York City Folium Map

4. Results

From the utilised dataset and the methodology carried out as described above, the following results can be drawn (please note that ‘Asian restaurants’ below refer to the variety of venues labelled as restaurants under Asian influence. For example, Chinese restaurant, Malay restaurant, Filipino restaurant, et cetera):

- There is only one Indonesian restaurant in New York City and none in Toronto
- By inspection, cluster 1 (Red) in Toronto has more Asian restaurants as top venues as compared to other clusters in the city
- By inspection, cluster 1 (Red) in New York City also has more Asian restaurants as top venues compared to other clusters in the city
- In Toronto, Asian restaurants can easily be spotted as common venues all over its different neighbourhoods
- In New York City, however, customers in general prefer restaurants under the Western influence such as American restaurants, Fast food restaurants and Spanish restaurants

5. Discussion

In choosing a location to which city could be a better to opening an Indonesia restaurant, the dataset above is not sufficient. There are more factors to consider. For example, in both cities; the population of Indonesian residents or immigrants, the attractiveness of neighbourhoods to both locals & tourists, the price of rent/land and more.

6. Conclusion

Based merely on the geographical datasets of Toronto & New York City and the analysis done, Toronto is a better city to open an Indonesian restaurant. The city does not yet have such restaurant and its citizens have stronger tendencies to visit restaurants with an Asian influence.