|  |  |
| --- | --- |
| **Exploring Toronto Neighbourhoods for Best Place to Open Ramen Place**  **Capstone Project - The Battle of the Neighborhoods (Week 2)** | Dan Liu  Applied Data Science Capstone |

Table of Contents

[1.0 Introduction 1](#_Toc17550010)

[1.1 Business Problem 1](#_Toc17550011)

[1.2 Target Audience 2](#_Toc17550012)

[2.0 Data 3](#_Toc17550013)

[2.1 Neighbourhood Profiles 3](#_Toc17550014)

[2.2 Neighbourhoods 4](#_Toc17550015)

[2.3 Foursquare 4](#_Toc17550016)

[2.4 2018 Toronto Detached Prices by Neighbourhood 6](#_Toc17550017)

[2.5 Neighbourhood Crime Rates 8](#_Toc17550018)

[3.0 Methodology 9](#_Toc17550019)

[4.0 Analysis 10](#_Toc17550020)

[5.0 Results 11](#_Toc17550021)

[6. 0 Discussion & Recommendations 12](#_Toc17550022)

[7.0 Conclusion 13](#_Toc17550023)

# 1.0 Introduction

## 1.1 Business Problem

This purpose of this project is explore the City of Toronto neighbourhoods for the optimal location to open a Ramen Restaurant.

The following criteria were selected by the stakeholders to determine initial potential locations

1. High population
2. High population density area to allow for highest potential number of initial customers when opening
3. High percentage of population with income greater than $30,000
4. High percentage of population who are living alone
5. High percentage of population who are aged between 15 - 54 years old
6. High population percentage of immigrants originating from East Asian countries
7. Neighbourhood with high residential real estate price
8. Neighbourhood with lower number of existing vendors that are similar to Ramen
9. High Average Total income
10. Low crime rate for patrons
11. Low crime rate for owners and investors

The following weighting will be used to calculate weighted score for each neighbourhood for each criteria.

|  |  |  |
| --- | --- | --- |
| Criteria | Qualitative Weight | Quantitative Weight |
| 1 | Low | 0.025 |
| 2 | High | 0.15 |
| 3 | Medium | 0.1 |
| 4 | High | 0.15 |
| 5 | Medium | 0.1 |
| 6 | Low | 0.025 |
| 7 | Low | 0.025 |
| 8 | Medium | 0.1 |
| 9 | Medium | 0.1 |
| 10 | Medium-High | 0.125 |
| 11 | Medium | 0.1 |

After initial exploration of the available data, it also determined to limit the geographical area to be close to where the owner of the ramen place resides to reduce traveling time:

* South of St. Clair Street
* East of Roncesvalles Avenue
* West of Coxwell Avenue
* Excluding the following areas to avoid existing competitions: Bay Corridor Yonge-Church Corridor.

This reduces the total number of neighbourhoods being examined from 140 to 32. By concentrating on the neighbourhoods close to the downtown core we are also targeting those who may work in downtown are looking to find dinner spots on their commute home to the suburbs.

## 1.2 Target Audience

The content of this project is intended following group of stakeholders

* Person who is interested in opening their own ramen restaurant in Toronto
* Investors who are interested in partnering to open a ramen restaurant or other similar type of restaurant

# 2.0 Data

Most of the data to be used to determine the best neighbourhood to open a Ramen Restaurant will be obtained from Toronto Open Data project:

The following particular datasets will be utilized in the project.

## 2.1 Neighbourhood Profiles

<https://open.toronto.ca/dataset/neighbourhood-profiles/>

This data set was last refreshed in July 2019 and contains the data for each of City of Toronto's 140 social planning neighbourhoods of Toronto collected during the latest Census of Population in 2016. Data collected contains: age and sex, families and households, language, immigration and internal migration, ethnocultural diversity, Aboriginal peoples, housing, education, income, and labour.

These social planning neighbourhoods were developed by the City of Toronto to help government and community organizations with local planning by providing socio-economic data at a meaningful geographic area. The boundaries of these social planning neighbourhoods are consistent over time, allowing for comparison between Census years

The Census data to provide a portrait of the demographic, social and economic characteristics of the people and households in each City of Toronto neighbourhood, where in this case we are particularly interested in characteristics of people in favour over households.

The data was sourced from a number of Census tables released by Statistics Canada.

Particularly, we are interested in the following data points for each neighbourhoods

* Population and dwellings
  + Population, 2016
  + Population density per square kilometre
* Age characteristics
  + Youth (15-24 years)
  + Working Age (25-54 years)
* Family characteristics of adults
  + Persons living alone (per cent)
* Income of individuals in 2015
  + Total - Employment income groups in 2015 for the population aged 15 years and over in private households - 100% data
    - Under 5,000(𝑖𝑛𝑐𝑙𝑢𝑑𝑖𝑛𝑔𝑙𝑜𝑠𝑠),5,000(includingloss),5,000 to 9,999,9,999,10,000 to 19,999,19,999,20,000 to 29,999,29,999,30,000 to 39,999,39,999,40,000 to 49,999,49,999,50,000 to 59,999,59,999,60,000 to 69,999,69,999,70,000 to 79,999,79,999,80,000 and over
* Total income: Average amount ($)
* Immigrants by selected place of birth
  + Total - Selected places of birth for the immigrant population in private households - 25% sample data
    - Asia
      * China, Hong Kong, Japan, Korea; South, Philippines, Taiwan, Viet Nam

The percentage is calculated by tabulating the total of each subcategory and dividing by the total respondents.

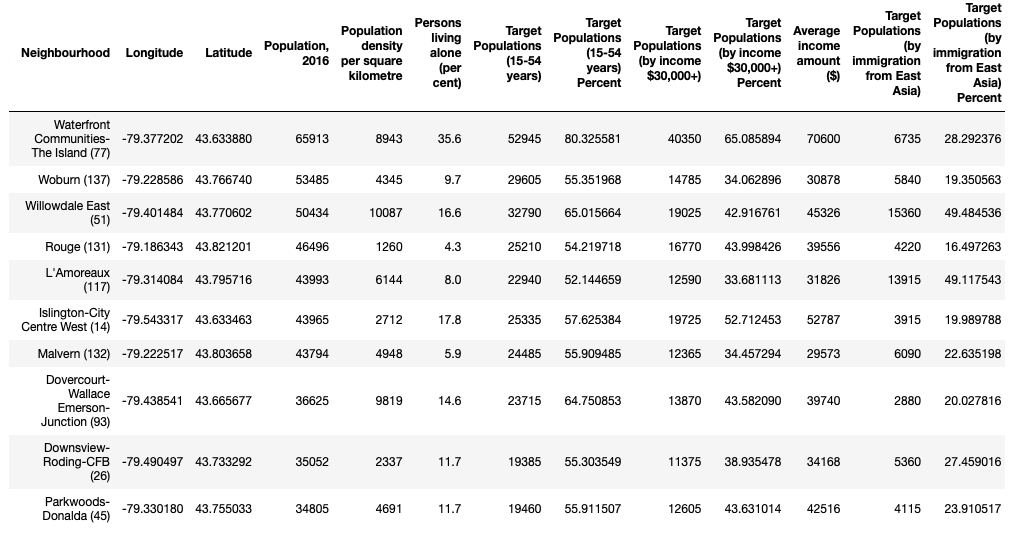


Figure 1 Data from Neighbourhood Profile and Neighbourhoods Boundary

## 2.2 Neighbourhoods

<https://open.toronto.ca/dataset/neighbourhoods/>

This dataset in .geojson format contains boundaries as well as the geographic coordinates (latitude and longitude) of the 140 City of Toronto neighbourhoods for social planning purposes. The geographic coordinates will be used to call the Foursquare API to determine the upto 10 food vendors within the center of each neighbourhood. The boundaries information will be used to generate choropleth map with Folium on a Leaflet map. The file was converted to .json file for the purpose of this project and easy reading into



Figure 2 Data from Neighbourhood Boundaries

## 2.3 Foursquare

<https://developer.foursquare.com/>

Foursquare API is used to collect data on close by venues within a specific radius of a given geographic coordinate. For the purpose of this project we are only looking at food related venues, by pass ‘food’ to the parameter ‘Section’ in the Get Venue Recommendations API call. (<https://developer.foursquare.com/docs/api/venues/explore>) The data returned will be used to explore existing make up for the food venue landscape of each neighbourhood.

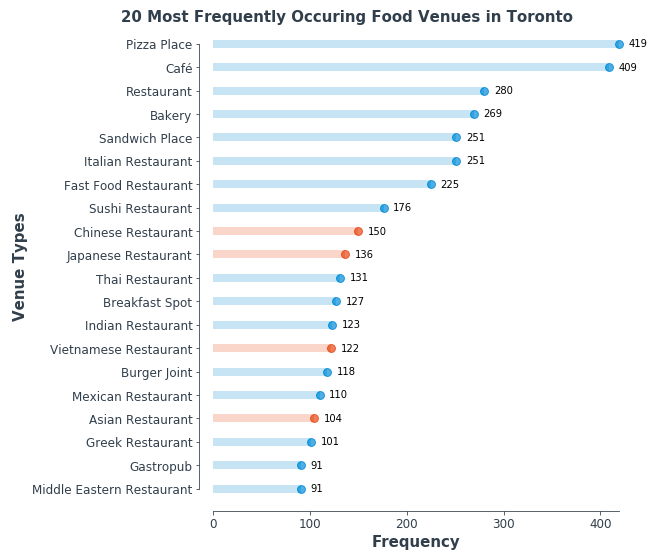
****

Figure 3 20 Most Frequently Occurring Food Venues in Toronto

****

Figure 4 Top 10 Food Venue from Foursquare (sample)

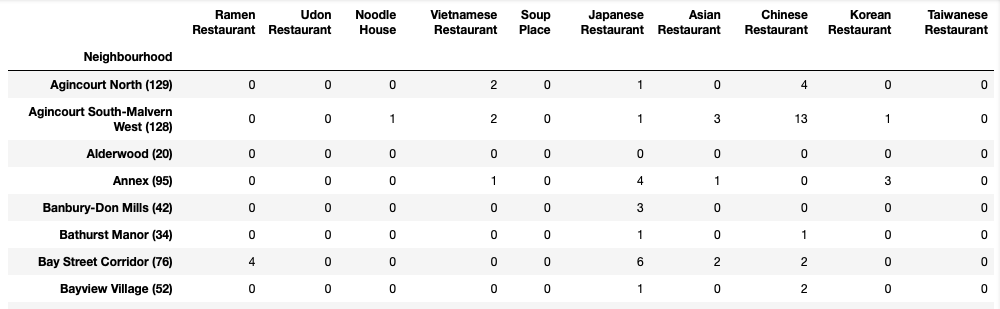
****

Figure 5 Foursquare Data - Venues Similar to Ramen

## 2.4 2018 Toronto Detached Prices by Neighbourhood

<https://docs.google.com/spreadsheets/d/1GizBkDvGaYm5AAPJ7O0y9eq-n7i81to6WpjejvYFzw0/edit#gid=754169703>

This data set contain average detached house prices by 144 TREB (Toronto Real Estate Board) neighbourhoods aggregated by Scott Ingram. As Scott mentioned, detached house prices were used to keep things simple and more apples-to-apples between neighbourhood. (<http://www.century21.ca/scott.ingram/blog/Toronto_s_most_and_least_expensive_neighbourhoods_in_2018>)

In order for the data to be used, the TREB neighbourhood must be converted to the 140 City of Toronto social planning neighbourhoods. For neighbourhood that does not have any sales, the overall average price for Toronto will be used. This data can be used to determine the state of the real estate price.

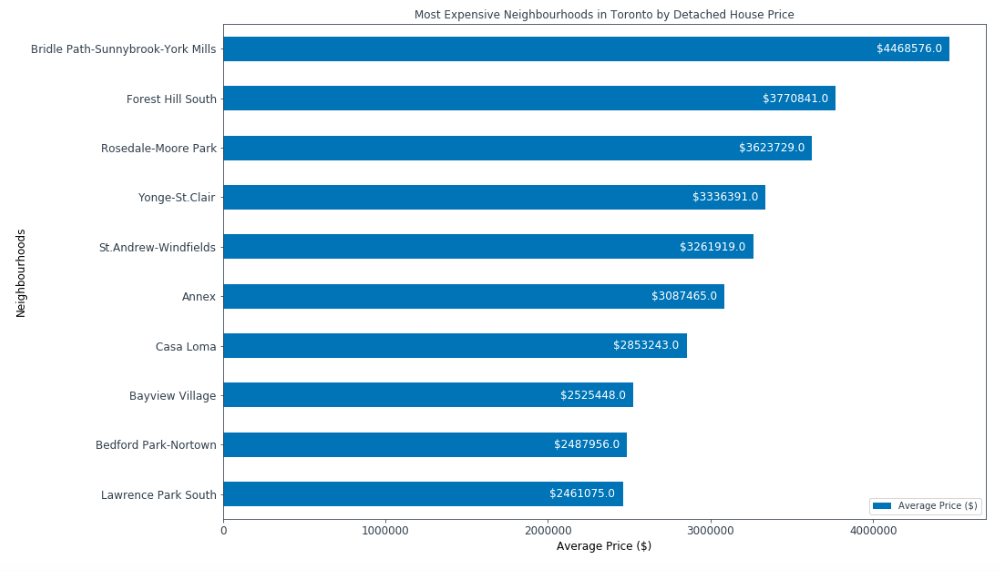


Figure 6 Most Expensive Real Estate Neighbourhoods in Toronto

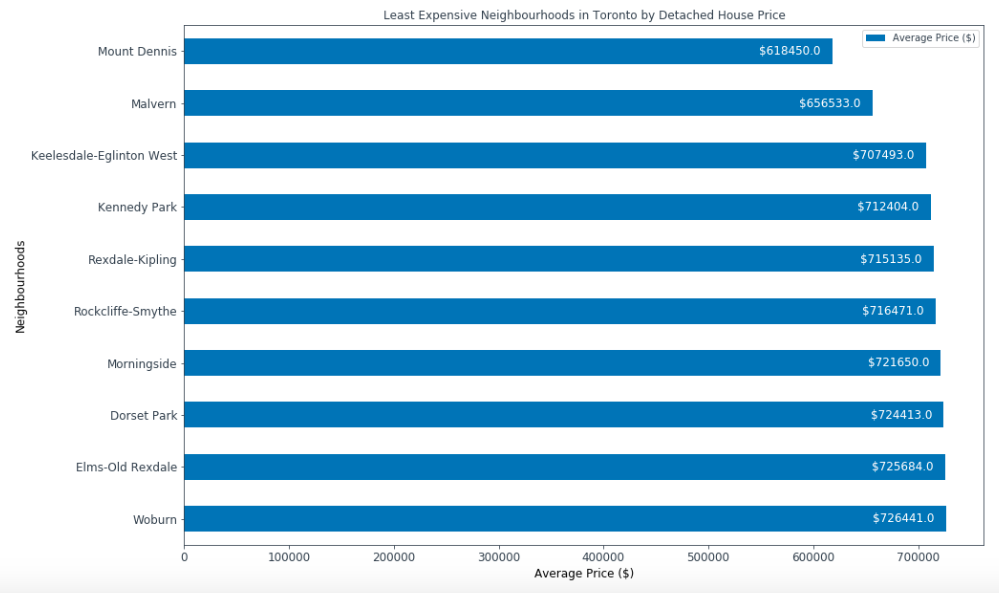


Figure 7 Least Expensive Real Estate Neighbourhoods in Toronto

## 2.5 Neighbourhood Crime Rates

<http://data.torontopolice.on.ca/datasets/neighbourhood-crime-rates-boundary-file->

This dataset contains Crime Data by Neighbourhood. the 2014-2018 period. Counts are available for Assault, Auto Theft, Break and Enter, Robbery, Theft Over and Homicide. Data also includes four year averages and crime rates per 100,000 people by neighbourhood based on 2016 Census Population.

The crime rates will be used to calculate the safety rate for the potential patrons and the owners and investors. For patrons we will consider Auto Theft, Robbery, and Homocide. For owners and investors we will consider Theft Over.

*Definition of Crimes*

Theft Over: The act of stealing property in excess of $5,000 (excluding auto theft).

Robbery: The act of taking property from another person or business by the use of force or intimidation in the presence of the victim.

Auto Theft: The act of taking or another person's vehicle (not including attempts). Auto Theft figures represent the number of vehicles stolen.

Break and Enter: The act of entering a place with the intent to commit an indictable offence therein.

Homicide/Murder: The homicide category includes the offences of First Degree Murder, Second Degree Murder, and Manslaughter. A homicide occurs when a person directly or indirectly, by any means, causes the death of another human being. Deaths caused by criminal negligence, suicide, or accidental or justifiable homicide (i.e self-defence) are not included. Homicide data is compiled based on the Homicide Squad Case List Log. Count is based on offence (i.e each deceased victim)

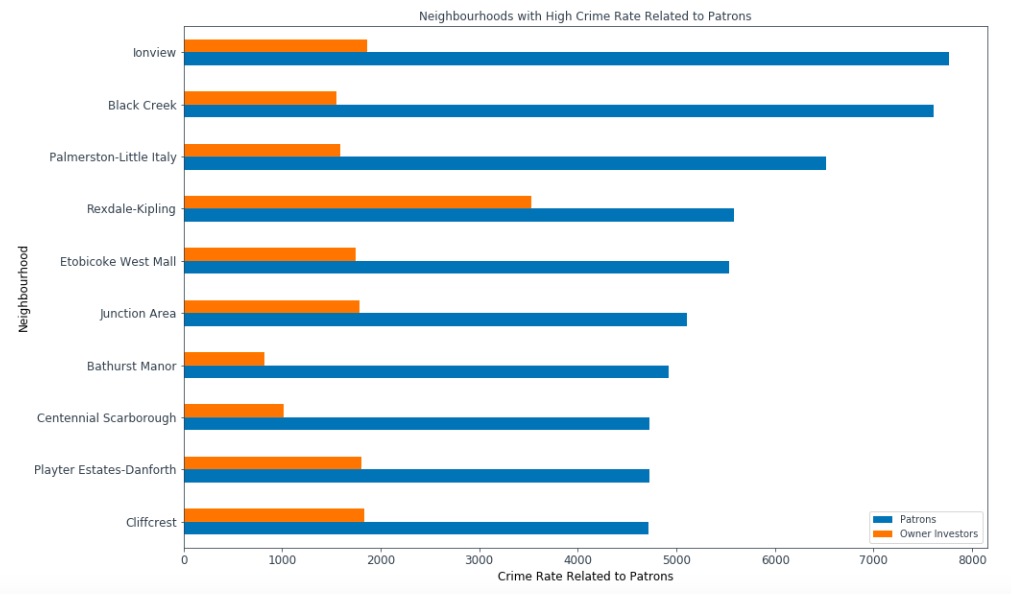


Figure 8 High Crime Rate Neighbourhood in Toronto

# 3.0 Methodology

To Narrow down the scope of this project, we will limit the neighbourhoods examined to the following geographical areas. This reduces the total number of neighbourhoods being examined from 140 to 32.

* South of St. Clair Street
* East of Roncesvalles Avenue
* West of Coxwell Avenue
* Excluding the following areas to avoid existing competitions: Bay Corridor Yonge-Church Corridor.

By concentrating on the neighbourhoods close to the downtown core we are also targeting those who may work in downtown are looking to find dinner spots on their commute home to the suburbs.

A decision matrix will be used to determine potential locations with the following criteria:

1. High population
2. High population density area to allow for highest potential number of initial customers when opening
3. High percentage of population with income greater than $30,000
4. High percentage of population who are living alone
5. High percentage of population who are aged between 15 - 54 years old
6. High population percentage of immigrants originating from East Asian countries
7. Neighbourhood with high residential real estate price
8. Neighbourhood with lower number of existing vendors that are similar to Ramen
9. High Average Total income
10. Low crime rate for patrons
11. Low crime rate for owners and investors

A standardized score for each criteria will be calculated for each neighbourhood by dividing the value from the neighbourhood profiles by the average of the neighbourhoods examined.

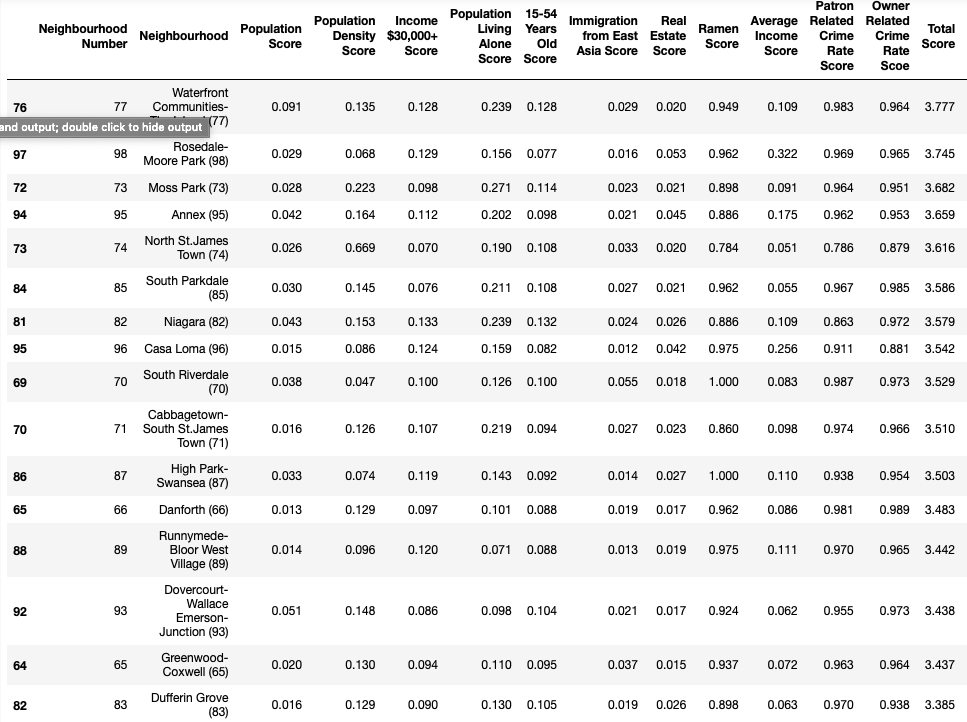
The following weighting will be applied to calculate weighted score for each neighbourhood for each criteria

|  |  |  |
| --- | --- | --- |
| Criteria | Qualitative Weight | Quantitative Weight |
| 1 | Low | 0.025 |
| 2 | High | 0.15 |
| 3 | Medium | 0.1 |
| 4 | High | 0.15 |
| 5 | Medium | 0.1 |
| 6 | Low | 0.025 |
| 7 | Low | 0.025 |
| 8 | Medium | 0.1 |
| 9 | Medium | 0.1 |
| 10 | Medium-High | 0.125 |
| 11 | Medium | 0.1 |

The neighbourhoods with top combined weighted score will be further examined and recommended.

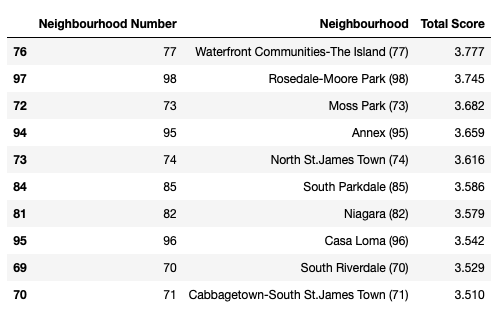
# 4.0 Analysis

A decision-making matrix was calculated using the criteria and dividing the value from the neighbourhood profiles by the average of the neighbourhoods examined.

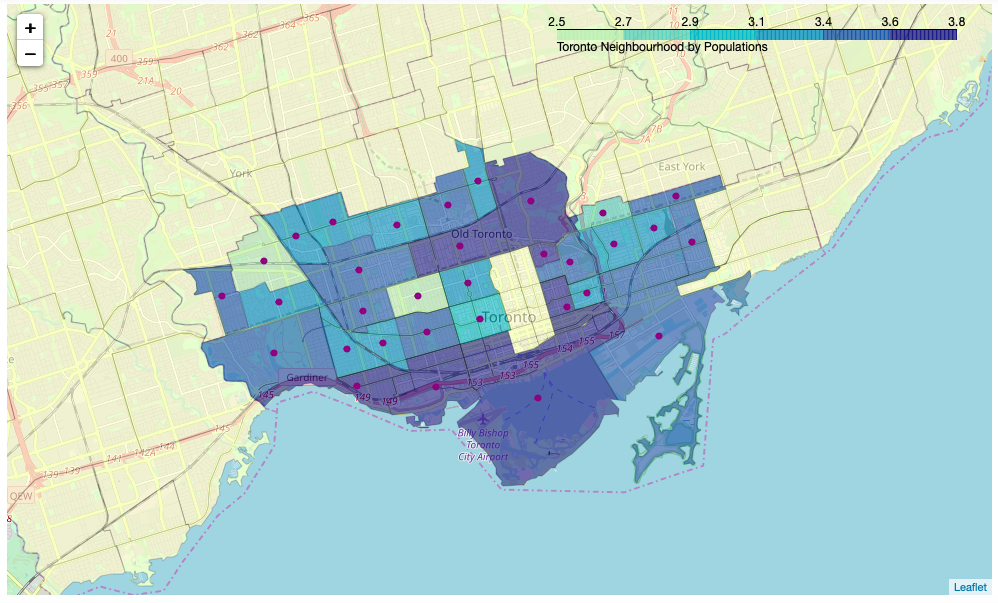


# 5.0 Results

Top 10 Neighbourhood to open Ramen Restaurant in Toronto



Choropleth Map of the Total Scores of the selected neighbourhoods



# 6. 0 Discussion & Recommendations

Based on the results calculated we can make the following observations and recommendations:

* Based on the weighted criteria selected the most optimal neighborhood to open a ramen restaurant would be Waterfront Communities — The Island. With a total score of 3.777. This area has low number of existing food establishments that are similar to ramen, low crime rates, high population, and individuals with income more than $30,000.
* Close in second place for recommendation is Rosedale-Moore Park. This area has high average income, low crime rate, and low number of existing food establishments that are similar to ramen.
* In a surprise outcome, Moss Park came in third as recommended place to open a ramen restaurant, even though the neighbourhood itself has a bad reputation within the city of Toronto. The neighbourhood has relative low crime rate, high percentage of people living alone as well as high percentage of 15-54 years old populations.
* Annex is the fourth recommended neighbourhood
* The following neighbourhoods should be avoided based on the calculated scores:
  + Weston-Pellam Park
  + North Riverdale
  + Kensington-Chinatown
  + Playter Estates-Danforth
  + Junction Area
  + Palmerston-Little Italy

This is due to higher crime rate and higher number of existing similar to ramen food establishments.

# 7.0 Conclusion

Purpose of this project was to help stakeholders identify optimal neighbourhoods in Toronto to open a ramen restaurant. Four main data sources were used: Foursquare, Toronto Neighbourhood Profiles, Toronto Average Detached House Sale Price 2018, and Toronto Crime Rate by Neighbourhood.

A total of 11 criteria were selected: population, population density, population living alone, population aged between 15-54, population with greater than $30,000 income, average income, existing food establishment similar to ramen, detached house prices and crime rate affecting patrons and owner.

Each selection criteria were assigned a weight. The neighbourhood data were standardized by dividing it by the mean of each category. The weights were then multiplied to calculated the weighted score. The neighbourhood with highest total weighted score is recommended for the best location to open the ramen restaurant.

The final decision on restaurant location will be made by stakeholders based on other factors that could not be considered for this specific project such as availability of commercial real estate space and price, parking availability, traffic, and etc..

Due to the limited scope of this of the analysis, only a small section of the neighbourhoods were selected and with a small set of features. With further features selected the 140 neighbourhoods could be clustered into groups using machine learning algorithm to determine groups of neighbourhood with similar characteristics. However, with the current the model, it is easily replicated for future census data as well as updated Foursquare data.