Business Location Selection Based on Population Distribution: Case Study in City of Toronto

(Capstone Project Final Report for IBM Data Science Professional Certificate)

By:

M. Nazarahari

June 2021

1. Introduction

This project aims to help new entrepreneurs open a new business (e.g., a restaurant) in the best possible location in the city of Toronto by analyzing the population distribution in the city. This is an important decision for any new business, with a significant role in the success or failure of the business. For example, if one opens a restaurant in a neighbourhood where many food services are available, it increases the chance of failure, while an intelligent decision for the location based on population and/or the number of existing food services could drastically improve the chance of success. Thus, in this project, we analyze the relationship between venues and population for the city of Toronto as a direct measure to determine a new food-related business location.

2. Data

The information related to neighbourhoods in Toronto, including postal code, borough, and neighbourhood name, were obtained from [here](#) and stored in a "Python" data frame, see Table 1. This information was used to extract the geographical location, i.e., latitude and longitude, of neighbourhoods using from "Nominatim" geolocation service through the "GeoPy" library in "Python" and then plot the map of the city with neighbourhoods identified, as shown in Figure 1.

Table 1: *A data frame showing the postal code and borough and their associated neighbourhoods in Toronto.*

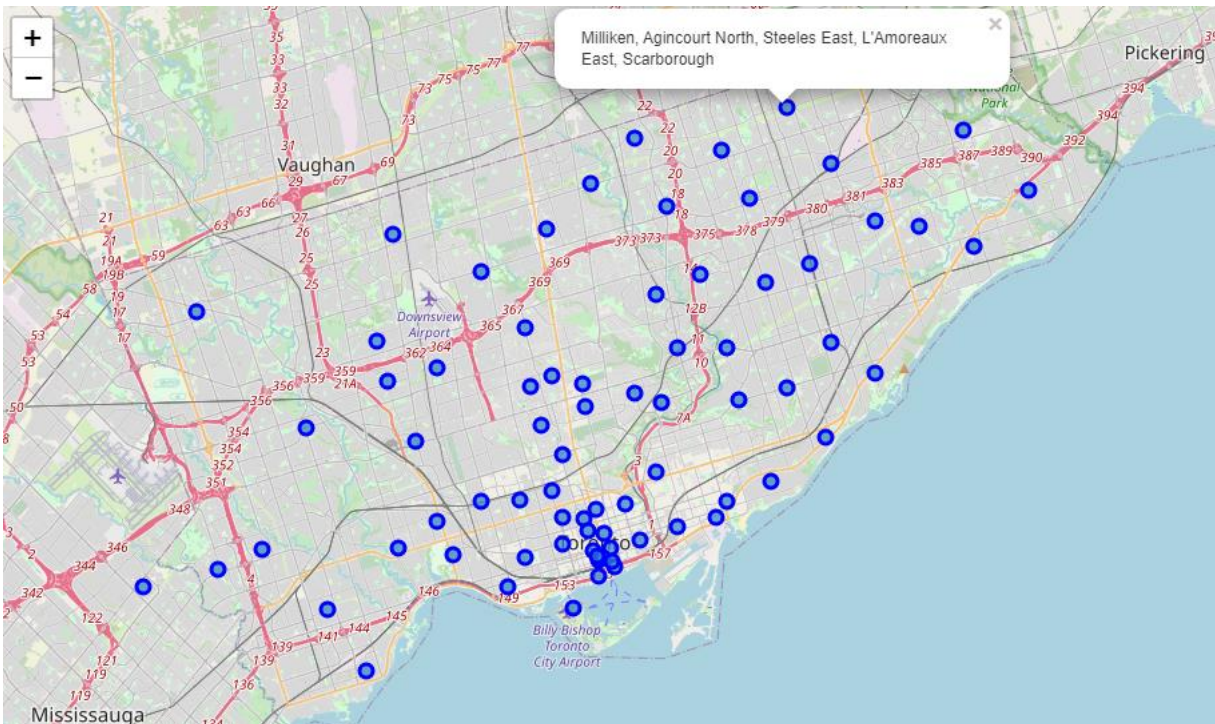|	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494

Figure 1: A map of Toronto with neighbourhoods identified with blue circles.

Also, we obtained the population in each neighbourhood using the "Wellbeing Toronto" provided by the city government (see here). We extracted neighbourhood name and total population (see Table 2) as well as other information such as the population of males/females, which were not used in this study. Then, all venues in each neighbourhood were extracted using Foursquare API. This information was used to analyze the relation between the (1) number of venues in each neighbourhood with its population and (2) number of food-related venues in each neighbourhood with its population.

Table 2: A data frame showing the neighbourhoods in Toronto and their population.

	Neighbourhood	Total Population	Pop - Males	Pop - Females	Child 0-14	Youth 15-24	Pop 15 - 64 years	Chinese	Language - Chinese	Language - Italian	Healthy Food Index	Heritage Sites	Hospital Readmissions	Seniors Living Alone	911 Calls Made by Seniors
0	West Humber-Clairville	33312.0	16625.0	16690.0	5060.0	5445.0	23285.0	470.0	370.0	320.0	23.82	3.0	0.99	160.0	994.0
1	Mount Olive-Silverstone-Jamestown	32954.0	16070.0	16890.0	7090.0	5240.0	22300.0	285.0	170.0	350.0	37.57	1.0	0.81	195.0	740.0
2	Thistletown-Beaumond Heights	10360.0	5055.0	5300.0	1730.0	1410.0	6760.0	110.0	75.0	275.0	42.26	6.0	0.97	105.0	368.0
3	Rexdale-Kipling	10529.0	5130.0	5395.0	1640.0	1355.0	7165.0	165.0	70.0	145.0	23.31	1.0	0.95	115.0	304.0
4	Elms-Old Rexdale	9456.0	4520.0	4935.0	1805.0	1440.0	6370.0	105.0	70.0	190.0	24.71	0.0	0.95	60.0	385.0