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

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

By:

Hanmo

July 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

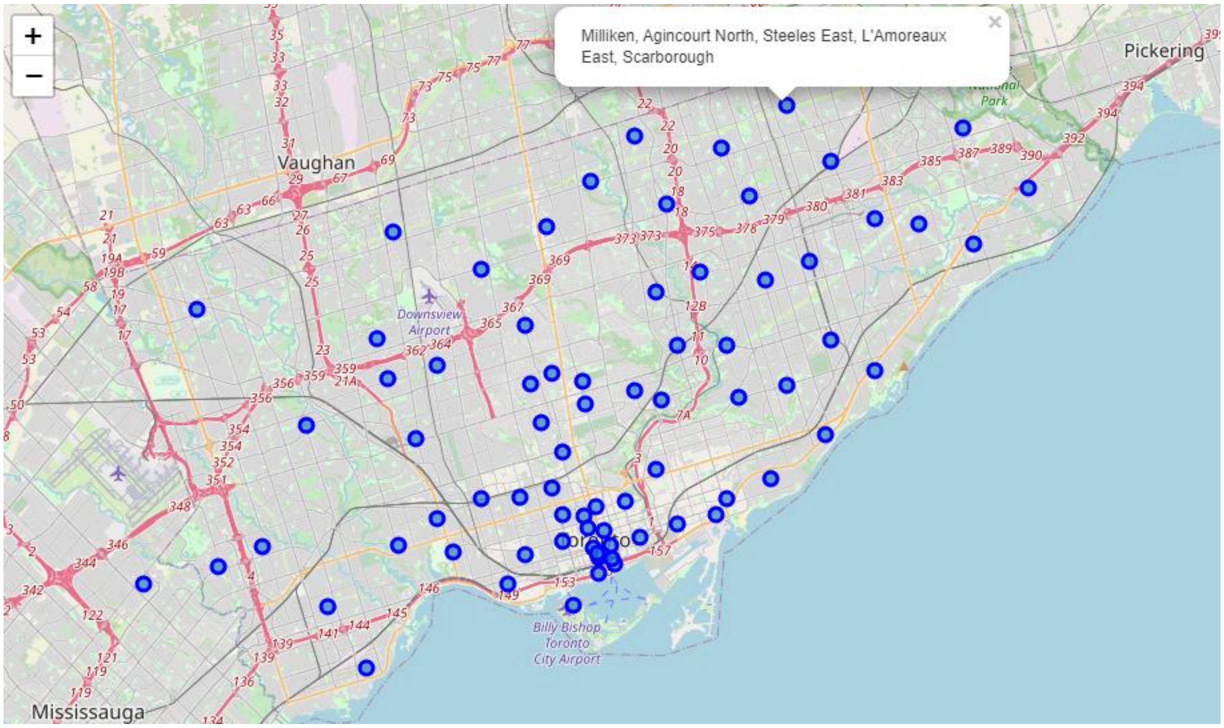

Figure 1: *A map of Toronto with neighbourhoods identified with blue circles.*

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

3. Methodology

Standard data manipulation methods, regression analysis, and correlation analysis was used to:

1. Identify neighbourhoods with maximum population.
2. Create a 3D map of the neighbourhoods' population.
3. Identify neighbourhoods with the maximum number of venues (of any kind).
4. Identify neighbourhoods with the maximum number of food-related venues.
5. Create a 3D map of the neighbourhoods' population and the number of venues.
6. Create a 3D map of the neighbourhoods' population and the number of food-related venues.
7. Create the regression plot for the total population vs the number of venues.
8. Calculate the correlation between the total population and the number of venues in each neighbourhood.
9. Create the regression plot for the total population vs the number of food-related venues.
10. Calculate the correlation between the total population and the number of food-related venues in each neighbourhood.
11. Determine the best neighbourhood for opening a new restaurant.

4. Results

First, we looked for five neighbourhoods in Toronto with the highest population. These neighbourhoods are (see Table 3):

- Waterfront Communities-The Island (population = 65913)
- Woburn (population = 53485)
- Willowdale East (population = 50434)
- Rouge (population = 46496)
- L'Amoreaux (population = 43993)

Table 3: *A data frame showing the neighbourhoods in Toronto with the highest total population.*

	Neighbourhood	Total Population	Pop - Males	Pop - Females	Child 0-14	Youth 15-24	Pop 15 - 64 years
76	Waterfront Communities-The Island	65913.0	33295.0	32635.0	3650.0	7840.0	57635.0
136	Woburn	53485.0	25955.0	27520.0	9625.0	7660.0	35835.0
50	Willowdale East	50434.0	23680.0	26740.0	5920.0	6940.0	38235.0
130	Rouge	46496.0	22475.0	24020.0	7960.0	6700.0	31915.0
116	L'Amoreaux	43993.0	20600.0	23375.0	6120.0	5730.0	28875.0

Then, we better visualize the population distribution in Toronto, we create a 3D map of the neighbourhoods' population as shown in Figure 2, where the size of circles is proportional to the neighbourhoods' populations.

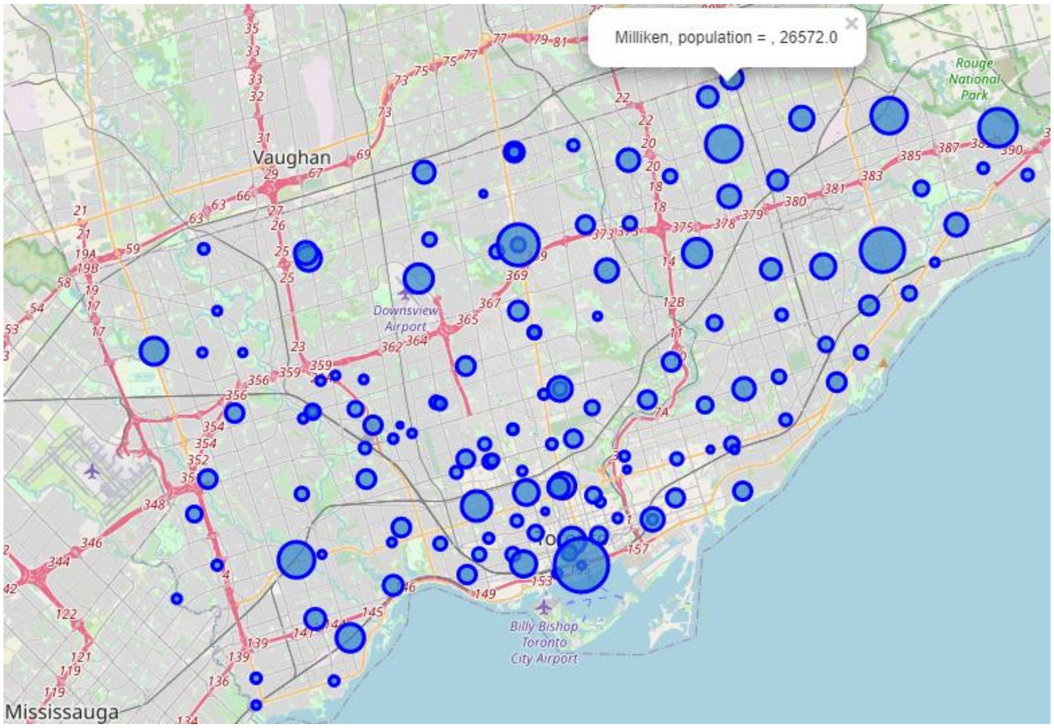

Figure 2: *A 3D map of Toronto where the blue circles show the neighbourhoods. The size of the blue circles (3rd dimension) is proportional to the neighbourhoods' populations.*

Then, we retrieved the venues in each neighbourhood using Foursquare API, as shown in Table 4. The retrieved data showed that there were 268 unique categories of venues in the neighbourhoods.

Table 4: *A data frame showing the retrieved venues in Toronto "West Humber-Clairville" neighbourhood.*

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	West Humber-Clairville	43.722563	-79.597039	Fortinos	43.721438	-79.596291	Grocery Store
1	West Humber-Clairville	43.722563	-79.597039	Mandarin Buffet	43.720360	-79.594387	Chinese Restaurant
2	West Humber-Clairville	43.722563	-79.597039	New York Fries - Woodbine Centre	43.720568	-79.599609	Restaurant
3	West Humber-Clairville	43.722563	-79.597039	TD Canada Trust	43.719630	-79.599896	Bank
4	West Humber-Clairville	43.722563	-79.597039	Subway	43.720448	-79.600085	Sandwich Place

Next, we found neighbourhoods with the highest number of venues (Table 5) and the highest number of food-related venues (Table 6). To find the food-related venues, we search our dataset for venues whose category included at least one of the following keywords: Food, Restaurant, Bakery, Bar, Café, Deli.

Table 5: *A data frame showing the Toronto neighbourhoods with the highest number of venues of any kind.*

Neighborhood	Total Num. Venues
Bay Street Corridor	100
St.Andrew-Windfields	100
Mount Olive-Silverstone-Jamestown	94
Kensington-Chinatown	80
Old East York	79

Table 6: *A data frame showing the Toronto neighbourhoods with the highest number of food-related venues.*

Neighborhood	Num. Food Related
St.Andrew-Windfields	44
Bay Street Corridor	41
Kensington-Chinatown	36
Mount Olive-Silverstone-Jamestown	34
Church-Yonge Corridor	32

In the next step, we created a 3D map of the (1) neighbourhoods' population and number of venues (Figure 3) and (2) neighbourhoods' population and number of food-related venues (Figure 4). In both Figures 3 and 4, the blue and red circles show the population and venues of neighbourhoods, respectively, while their sizes are proportional to the highest population/number of venues (smaller circles show lower population/number of venues). Black and green arrows in Figure 4 show desired neighbourhoods for opening a new food-related business, considering the distribution of population and food-related venues in these neighbourhoods.

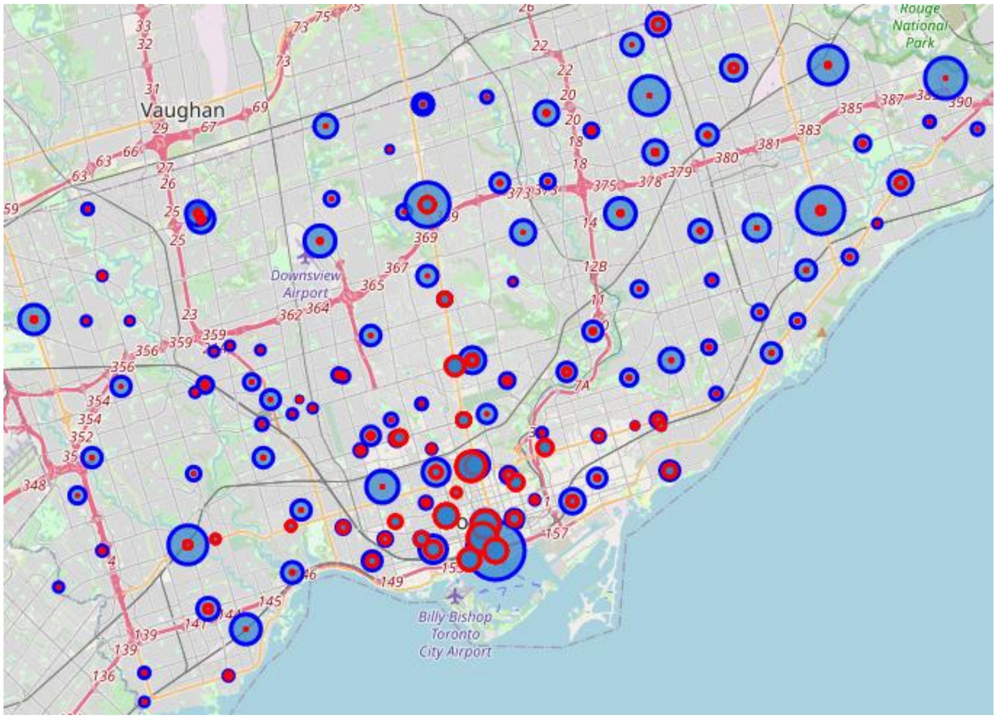

Figure 3: *A 3D map of Toronto where the blue/red circles show the neighbourhoods. The size of the blue/red circles (3rd dimension) is proportional to the neighbourhoods' populations/number of venues.*

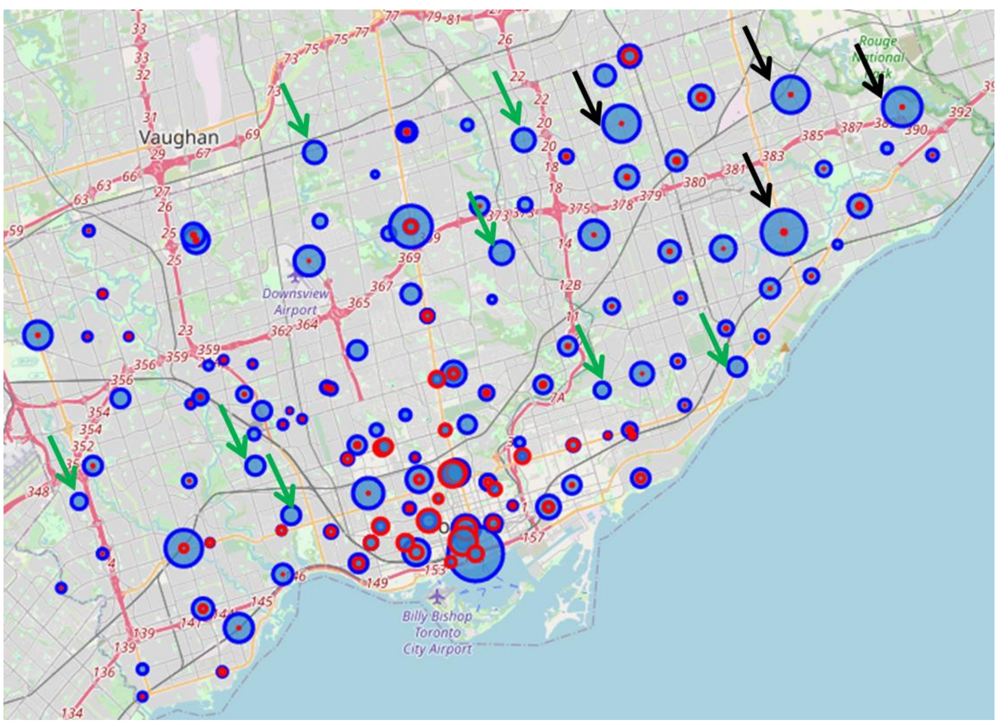

Figure 4: *A 3D map of Toronto where the blue/red circles show the neighbourhoods. The size of the blue/red circles (3rd dimension) is proportional to the neighbourhoods' populations/number of food-related venues.*

Next, we created the regression plot for total population vs number of venues (Figure 5) as well as the regression plot for total population vs number of food-related venues (Figure 6). Both plots show that the population and number of venues are not linearly correlated, meaning that there are neighbourhoods with high population and a few venues, while some neighbourhoods with low population have a high number of venues. We can confirm this by observing the 3D maps in Figures 3 and 4. Another way to confirm this observation is to calculate the correlation between population and the number of venues. Notably, the obtained correlation coefficients were 0.15 and 0.06 for population vs the number of venues and population vs the number of food-related venues, respectively.

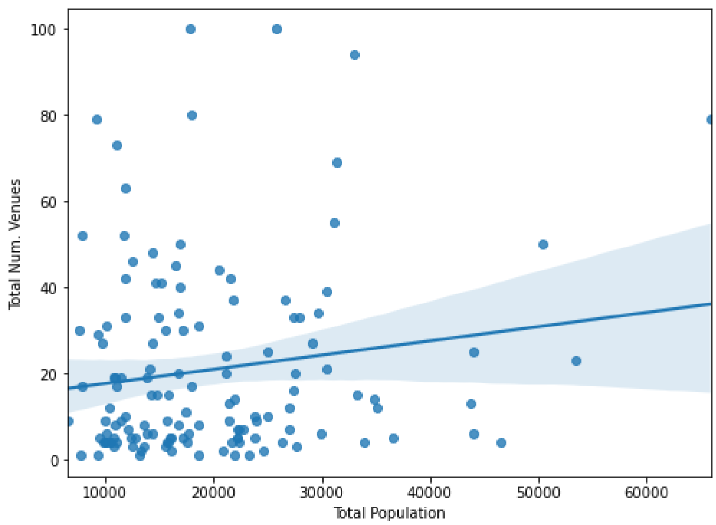

Figure 5: *Scatter plot and regression line for total population vs the total number of venues.*

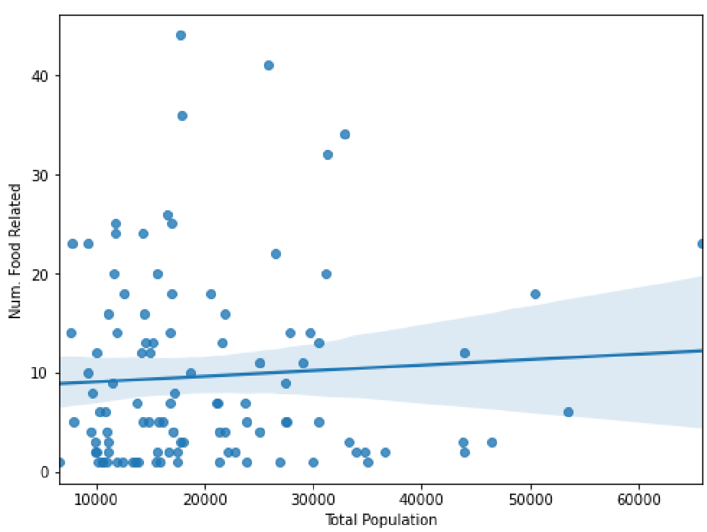

Figure 7: *Scatter plot and regression line for total population vs the total number of venues.*

Finally, to determine the best neighbourhoods for opening a new food-related venue, we calculated the ratio of population over the number of food-related venues for neighbourhoods with at least one food-related venue. Table 7 provides a list of suggested neighbourhoods. According to Table 7, Downsview-Roding-CFB followed by Bendale, Clairlea-Birchmount, High Park-Swansea, L'Amoreaux are the best neighbourhoods with the highest population to the number of food-related venues ratio.

Table 7: *A data frame showing the best Toronto neighbourhoods for opening a new food-related neighbourhood.*

	Neighbourhood	Latitude	Longitude	Total Population	Num. Food Related	Venue Ratio	Pop/Venue Ratio
24	Downsview-Roding-CFB	43.7493	-79.4622	35052.0	1.0	0.227273	35052.0
120	Bendale	43.7535	-79.2553	29960.0	1.0	0.227273	29960.0
113	Clairlea-Birchmount	43.7088	-79.296	26984.0	1.0	0.227273	26984.0
83	High Park-Swansea	43.6363	-79.4754	23925.0	1.0	0.227273	23925.0
111	L'Amoreaux	43.799	-79.306	43993.0	2.0	0.454545	21996.5

Discussions

This project showed the use of data science to create a clear procedure for finding the best location for a new business in the city of Toronto. Here, we analyzed a food-related business in Toronto, but the same idea can be applied to other businesses and cities.

Our analysis showed that the best neighbourhoods for opening a new venue are Downsview-Roding-CFB, Bendale, Clairlea-Birchmount, High Park-Swansea, and L'Amoreaux (Table 7). Also, neighbourhoods identified with green arrows in Figure 4 are good candidates as they have a high population and no food-related venue. To find the best location among these neighbourhoods, one can use secondary measures such as commute time, rental costs, public transport accessibility, and public parking space availability.

Future Directions

As mentioned earlier, the same analysis can be performed to find the best location for other venues. For example, one can determine the best location for a Chinese restaurant using the population of people with Chinese background (as provided in Table 2). Similarly, to find the best location for a long-care facility, information such as the number of seniors living alone (Table 2) can be used. Finally, this project indicates that the venues/facilities are not distributed uniformly (considering the population distribution) among neighbourhoods. Thus, it can help decision-making bodies such as the government of Ontario in their city planning, where all people have access to their desired venues/facilities and do not need to travel along the city to access a specific venue.

Conclusions

This project showed how data science could help find the best location for a business in a city. The proper selection of business location can significantly affect the success or failure of a new business.