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

Hanmo

July 2021

Capstone Project Final Report
IBM Data Science Professional Certificate

Introduction

- Finding best location for a new business have a significant role in the success or failure. For example:
  - Opening a restaurant where population is low and there are some well-established restaurants present.

- **Business question:**
  - Can we find the best location by analyzing the population of neighborhoods along with the number of venues in each neighborhood?

Data

- We performed a case study for the city of Toronto.
  - Wellbeing Toronto: neighborhood names and populations.
  - Geolocation service through the “GeoPy”: latitude and longitude of each neighborhood.
  - Foursquare API: venues in each neighborhood.

	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

Methodology

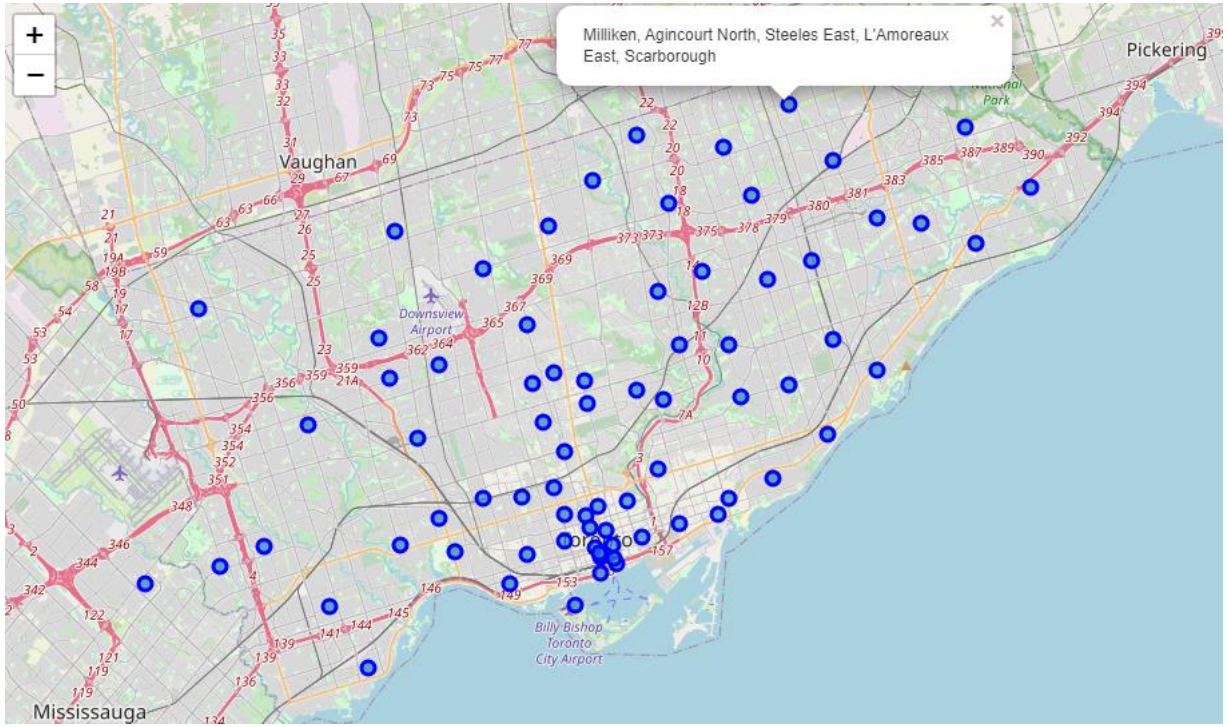
- Standard data manipulation methods, regression analysis, and correlation analysis was used to
  - Identify neighborhoods with maximum population.
  - Create a 3D map of the neighborhoods' population.
  - Identify neighborhoods with the maximum number of venues.
  - Create a 3D map of the neighborhoods' population and the number of venues.
  - Create the regression plot for the total population vs the number of venues.
  - Calculate the correlation between the total population and the number of venues in each neighborhood.
  - Determine the best neighborhood for opening a new restaurant.

Results: Most Populated Neighborhoods

List of five neighborhoods in Toronto with highest population.

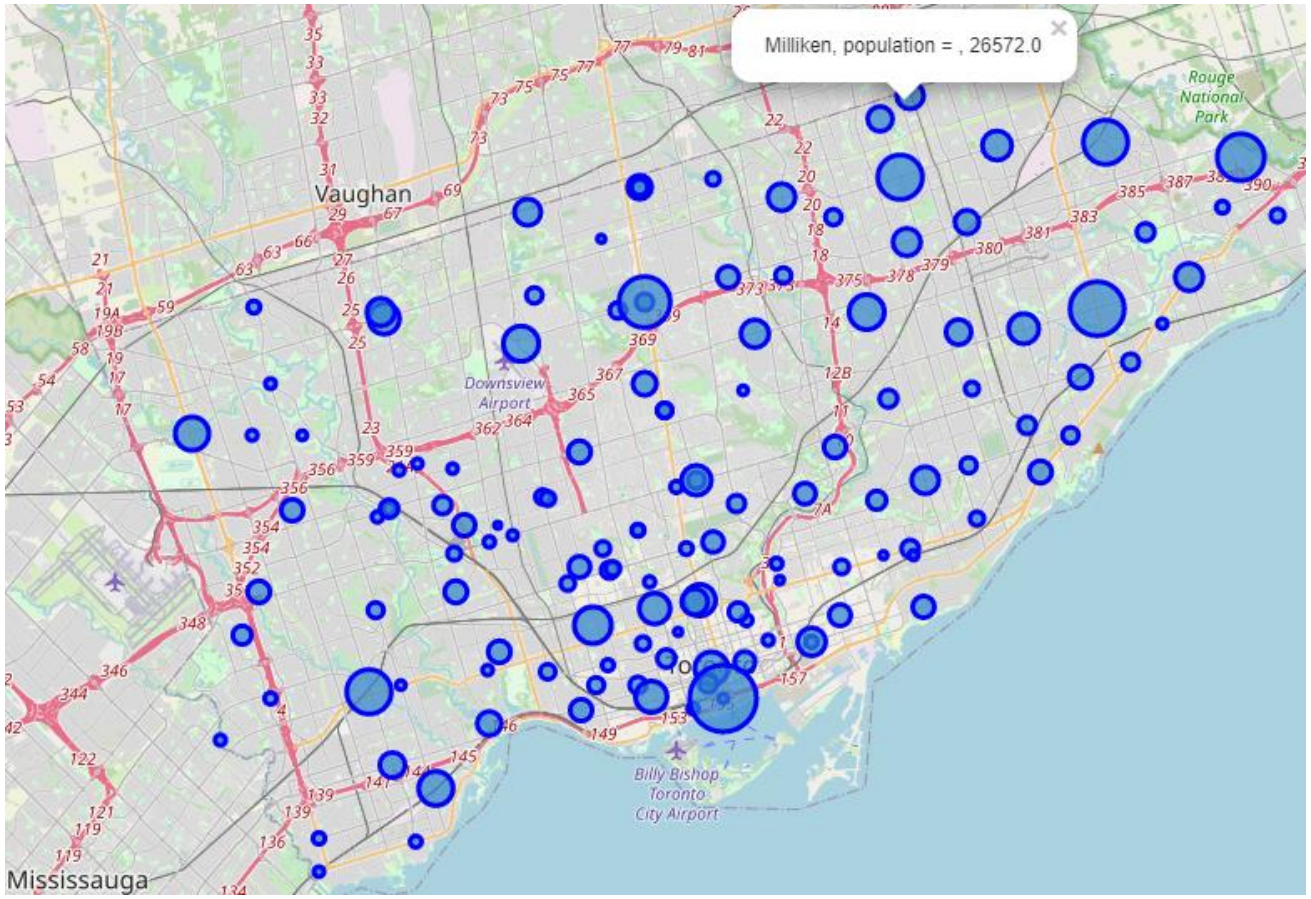
	Neighbourhood	Total Population	Pop - Males	Pop - Females	Child 0-14	Youth 15-24	Pop 15 - 64 years
76	Waterfront Communities-The Island	65913.0	33295.0	32635.0	3650.0	7840.0	57635.0
136	Woburn	53485.0	25955.0	27520.0	9625.0	7660.0	35835.0
50	Willowdale East	50434.0	23680.0	26740.0	5920.0	6940.0	38235.0
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

Map of Toronto neighborhoods.



Results: 3D map of Toronto

Neighborhoods with higher population have a bigger circle.



Results: Neighborhoods with most venues

List of five neighborhoods in Toronto with 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

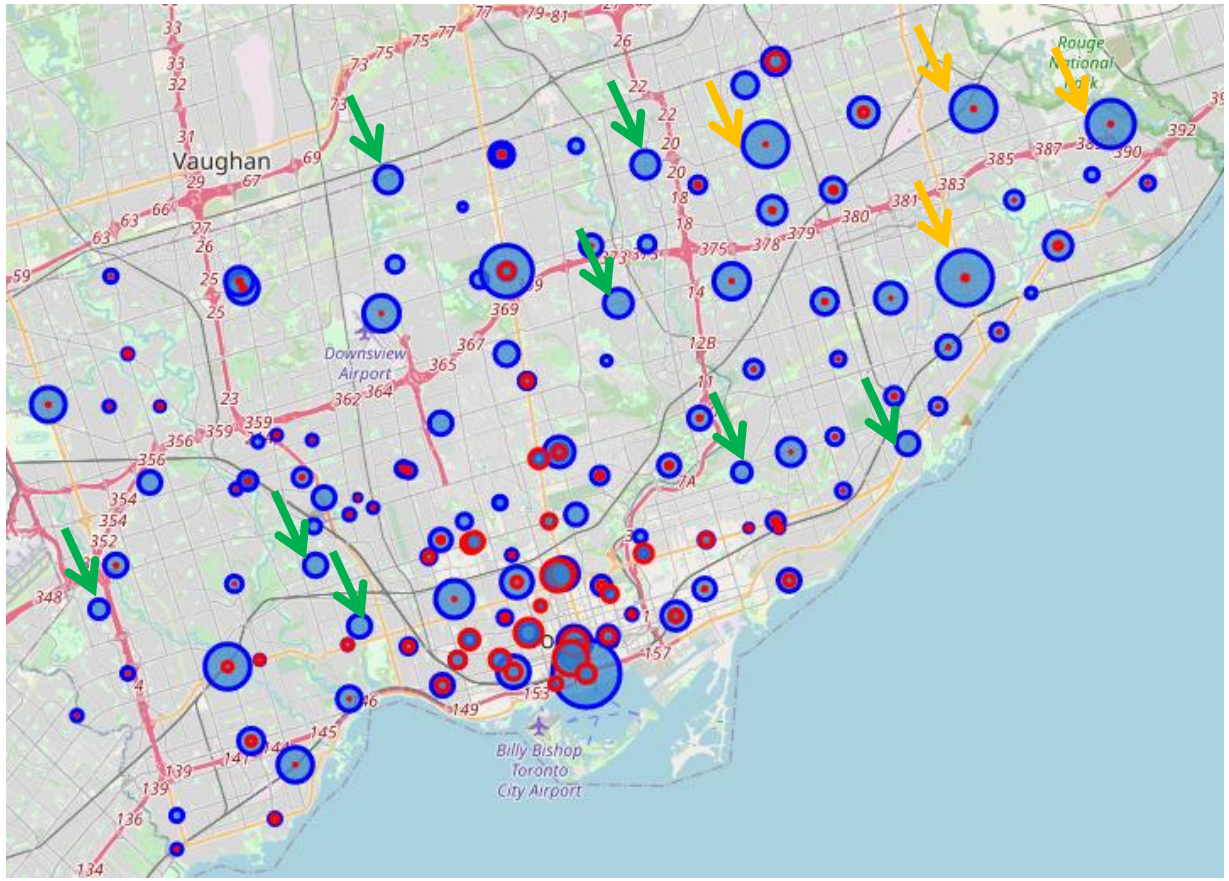
List of five neighborhoods in Toronto with highest number of food-related venues.

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

Results: Population vs Number of Venues

The size of the blue/red circles (3rd dimension) is proportional to the neighborhoods' populations (blue) and number of food-related venues (red).

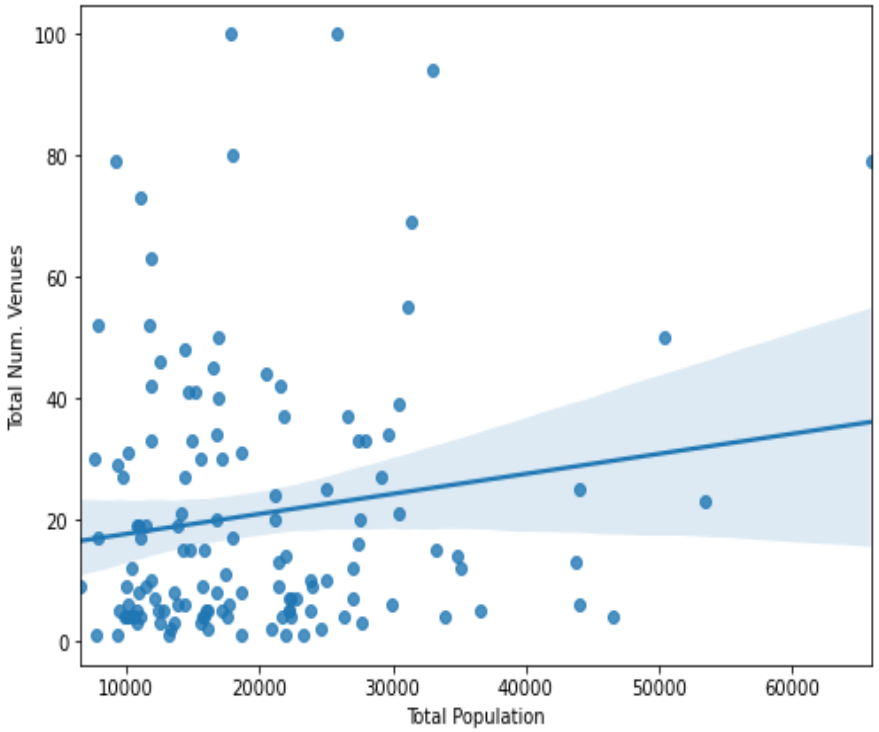
- Orange arrows: neighborhoods with high population and low number of food-related venues.
- Green arrows: neighborhoods with high population and no food-related venue.



Results: Population vs Number of Venues

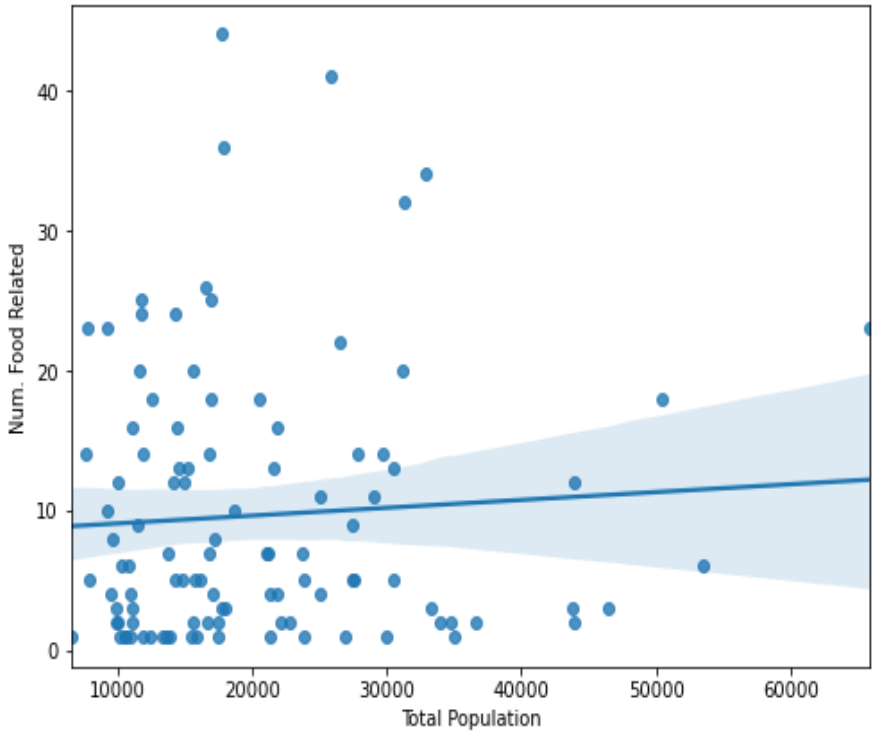
Population of neighborhoods vs number of venues of any kind.

Correlation Coefficient = 0.15



Population of neighborhoods vs number of food-related venues.

Correlation Coefficient = 0.06



Results: Best Neighborhood

Five neighborhoods with highest population to number of venue ratio.

	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

Discussion

- Data science proved to be useful in finding the best location for a new business.

- We performed a case study for a food-related business in Toronto
  - The same idea can be applied to other businesses and cities.

- Best neighborhoods for opening a new venue
  - Downsview-Roding-CFB.
  - Bendale.
  - Clairlea-Birchmount.
  - High Park-Swansea.
  - L'Amoreaux.