New Business Supply Analysis in Toronto

Introduction

Big cities, like Toronto, have a higher chance to possess a saturated market in many neighborhoods, making it more difficult for new businesses to be thriving. It is key for new businesses' success to have a good balance in between demand and supply within the location they start. High demand guarantees that there is a need for the product/service, and low supply guarantees that there's not enough business in the area to satisfy customers' needs. The goal was to find geographical locations of low supply for a specific product/service, therefore offering the least amount of competition to new businesses.

For example, the idea of a new restaurant/coffee shop opening up in the city of Toronto was used. Then, the top 10 neighborhoods with the highest and lowest supply of restaurants/coffee shops were calculated; therefore, areas to avoid and to consider were found. While this approach doesn't guarantee new businesses' success, it offers a range of possibilities for areas of little competition. It is recommended to complement this service with additional market analysis to identify regions of high demand for the product/service offered.

Data

Data Sources

Raw data used to solve this problem was extracted from a Wikipedia page that possessed all the postal codes for the city of Toronto. Each postal code had its respective borough and neighborhoods; therefore a list of all the neighborhoods that belong to the city of Toronto was obtained.

Also, Foursquare's API was used to retrieve information about a specific latitude and longitude. This information contained the different categories of businesses that exist within a neighborhood. Consequently, exploring each neighborhood of Toronto it was found where the highest and lowest numbers of specific venues were located.

Data Cleaning

The original data extracted from Wikipedia had a size of (217, 2) = (rows, columns) (Fig. 1). Then a function was defined that took every entry in the 'Neighborhood' column and found its coordinates ('Latitude' and 'Longitude'). The *geopy.geocoder* library was used to perform the operation, where the only input needed was an address.

Some neighborhood entries had to be deleted, as their name did not produce a reasonable output for latitude and longitude. A final *dataframe* of size (205, 4) was obtained with four columns: 'Borough', 'Neighborhood', 'Latitude', and 'Longitude' for every neighborhood in Toronto (Fig. 1).

	Borough	Neighborhood
0	North York	Parkwoods
1	North York	Victoria Village
2	Downtown Toronto	Regent Park
3	Downtown Toronto	Harbourfront
4	North York	Lawrence Manor

	Borough	Neighborhood	Latitude	Longitude
0	North York	Parkwoods	43.7588	-79.3202
1	North York	Victoria Village	43.7327	-79.3112
2	Downtown Toronto	Regent Park	43.6607	-79.3605
3	Downtown Toronto	Harbourfront	43.6401	-79.3801
4	North York	Lawrence Manor	43.7221	-79.4375

Fig. 1. Raw data extracted from Wikipedia and modified dataframe (only first 5 rows of each dataframe showed)

Methodology

The goal was to find the neighborhoods with the highest and lowest supply of restaurants/coffee shops. Foursquare's API was used to obtain a vast range of venues surrounding a specific location (longitude and latitude). From previous data cleaning, the location coordinates for Toronto's 205 neighborhoods were found. Therefore, a function was defined that took every pair of coordinates and obtained information about the surrounding venues.

Additional parameters needed to make an API called the amount of venues and radius surrounding the specific point. A maximum amount of 100 venues were set and a radial distance of 500 meters. Information extracted about each venue was its name, latitude, longitude, and its category. Finally, the results were formatted into a single *dataframe* of size (5975, 7) (Fig. 2).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.7588	-79.320197	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
1	Parkwoods	43.7588	-79.320197	LCBO	43.757774	-79.314257	Liquor Store
2	Parkwoods	43.7588	-79.320197	Petro-Canada	43.757950	-79.315187	Gas Station
3	Parkwoods	43.7588	-79.320197	Shoppers Drug Mart	43.760857	-79.324961	Pharmacy
4	Parkwoods	43.7588	-79.320197	Pizza Pizza	43.760231	-79.325666	Pizza Place

Fig. 2. Total amount of venues found for every neighborhood (only first 5 rows of dataframe showed)

Determining the amount of restaurants/coffee shops for every neighborhood was done by analyzing the 'Venue Category' entry of every venue. Another function was defined that compared every 'Venue Category' to a group of five to ten *key words* set by the user. The *key words* used to find the amount of restaurants/coffee shops were: Restaurant, Café, Coffee, Place, Food, and Deli (because of their familiarity with the specific venue category). If any of those *key words* were found within the 'Venue Category' column then a restaurant/coffee shop was assumed to be in the neighborhood. To conclude, every value was added giving a total count of venues per neighborhood.

Results

The top 10 neighborhoods with the highest supply of restaurants/coffee shops were:

	Neighborhood	Restaurant	Café	Coffee	Place	Food	Deli	Total	Latitude	Longitude
90	Willowdale	45	0	24	12	6	0	87	43.7615	-79.4109
81	Commerce Court	34	6	11	3	2	3	59	43.6481	-79.379
6	Queen's Park	23	8	19	7	0	0	57	43.6597	-79.3903
70	Design Exchange	27	7	13	3	3	3	56	43.6477	-79.3801
175	First Canadian Place	28	7	9	5	1	3	53	43.6488	-79.3817
69	Toronto Dominion Centre	27	6	10	6	2	2	53	43.6474	-79.3814
11	Don Mills	26	0	10	6	10	0	52	43.7753	-79.3459
46	King	32	5	7	2	3	2	51	43.6489	-79.3778
45	Adelaide	28	6	6	5	3	2	50	43.6505	-79.3795
141	Grange Park	29	5	8	5	0	0	47	43.6522	-79.3923

Fig. 3. Top 10 neighborhoods with the highest presence of restaurants/coffee shops

The top 10 neighborhoods with the lowest supply of restaurants/coffee shops were:

	Neighborhood	Restaurant	Café	Coffee	Place	Food	Deli	Total	Latitude	Longitude
133	Swansea	0	0	0	0	0	0	0	43.6449	-79.4783
21	Port Union	0	0	0	0	0	0	0	43.7755	-79.135
22	Highland Creek	0	0	0	0	0	0	0	43.7901	-79.1733
23	Woodbine Heights	0	0	0	0	0	0	0	43.6999	-79.3191
25	Eringate	0	0	0	0	0	0	0	43.6623	-79.5765
27	Old Burnhamthorpe	0	0	0	0	0	0	0	43.6394	-79.5844
28	Markland Wood	0	0	0	0	0	0	0	43.6312	-79.5854
160	South Steeles	0	0	0	0	0	0	0	43.8162	-79.3145
29	Guildwood	0	0	0	0	0	0	0	43.7552	-79.1982
167	Steeles West	0	0	0	0	0	0	0	43.8162	-79.3145

Fig. 4. Top 10 neighborhoods with the lowest presence of restaurants/coffee shops

Also, a map of the city of Toronto was created to display the results previously obtained (Fig. 3, 4). The green dots represent the lowest supply, therefore guaranteeing little competition for a starting restaurant/coffee shop (Fig. 5). On the other hand, the red dots represent the highest supply which corresponds to a saturated market of restaurants/coffee shops.

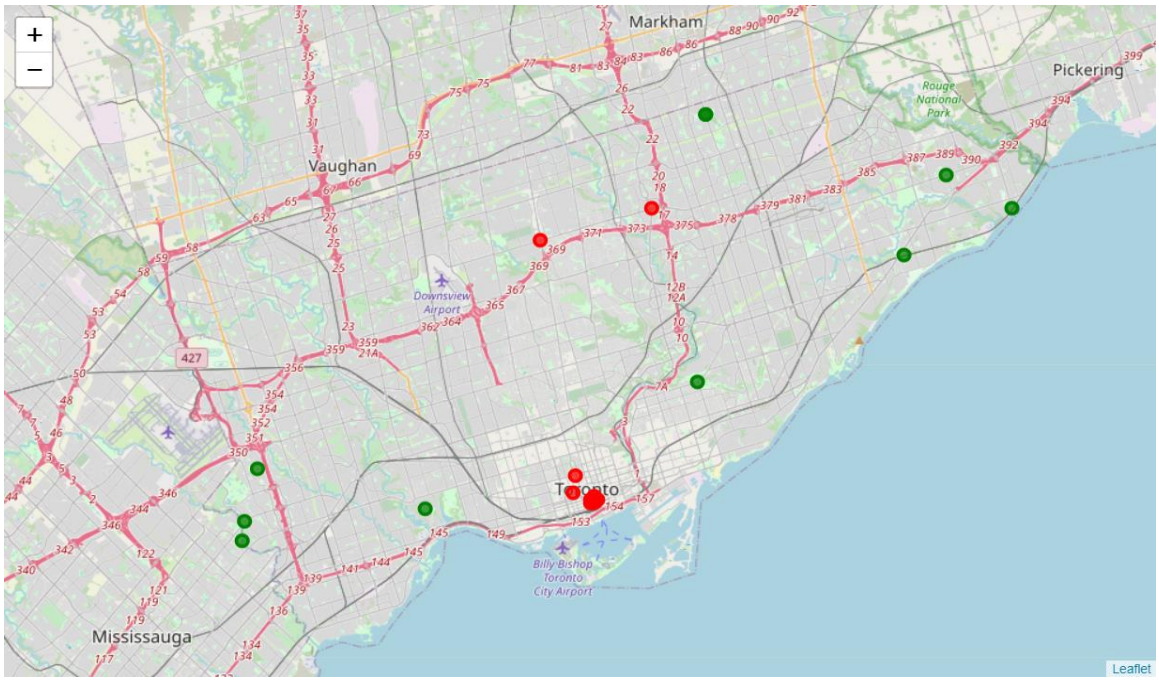

Fig. 5. Map of Toronto showing neighborhoods with the highest and lowest presence of restaurants/coffee shops

Discussion

The methodology procedure could have some improvements to guarantee better accuracy. For example, the maximum amount of venues and radial distance set for the API could have influenced the retrieved data depending on the size and population of each neighborhood. Also, the use of *key words* to calculate the amount of venues did not included every possible 'Venue Category' description, only the most common ones.

The results match what was expected from a big city like Toronto. The neighborhoods closer to Downtown should have a higher supply of venues while those in the surroundings do not (figure). However, there were some outliers to this trend, which may be explained by high population density within surrounding neighborhoods. In addition, locations with low supply of restaurants/coffee shops do not guarantee the success of starting business. This approach should be combined with a market study to confirm high demand for the business within these neighborhoods.

Moreover, additional regression and clustering algorithms could be implemented to improve the results obtained. In neighborhoods with high supply of restaurants/coffee shops, a correlation could be found in between the amount of venues with the presence of other venues like schools, parks, banks, etc. Therefore, the need for more restaurants/coffee shops could be predicted for other neighborhoods with similar characteristics. Also, clustering of neighborhoods could be done on the basis of amount of specific venues.

Conclusion

In conclusion, this approach could benefit new businesses when evaluating ideal location. Depending on the type of business, an area with high competition might be favorable but usually low competition is preferred. Therefore, a supply analysis can be done to determine the neighborhoods to have the lowest supply amount of the specific business in question.