New Business Supply Analysis in Toronto

Diego Gordon

Capstone Project – Coursera

June 3, 2020

Problem Statement

- 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.
- 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.

Example: new restaurant/coffee shop

- Calculate the top 10 neighborhoods with the highest and lowest supply of restaurants/coffee shops. Therefore, have the knowledge of areas to avoid and to consider for a new business

Data Acquisition and Cleaning

- Data Sources:
  - Raw data extracted from Wikipedia page
  - Foursquare's API to retrieve location information
- Data Cleaning:
  - Wikipedia data had every postal code, borough, and neighborhood for Toronto
  - Used *geopy.geocoder* to obtain the latitude and longitude for every neighborhood and combine it in a single *dataframe*

	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

Determining Number of Venues

- Used Foursquare's API to explore venues for each of the 205 neighborhoods, by inputting the latitude and longitude
- Limit of venues = 100, Radius = 500 meters
- For each of the 5975 venues retrieved, obtain their latitude, longitude, and venue category

	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

- Compare each entry from the 'Venue Category' with a group of five to ten *key words* set by the user.
- *Key words* = ['Restaurant', 'Café', 'Coffee', 'Place', 'Food', 'Deli']

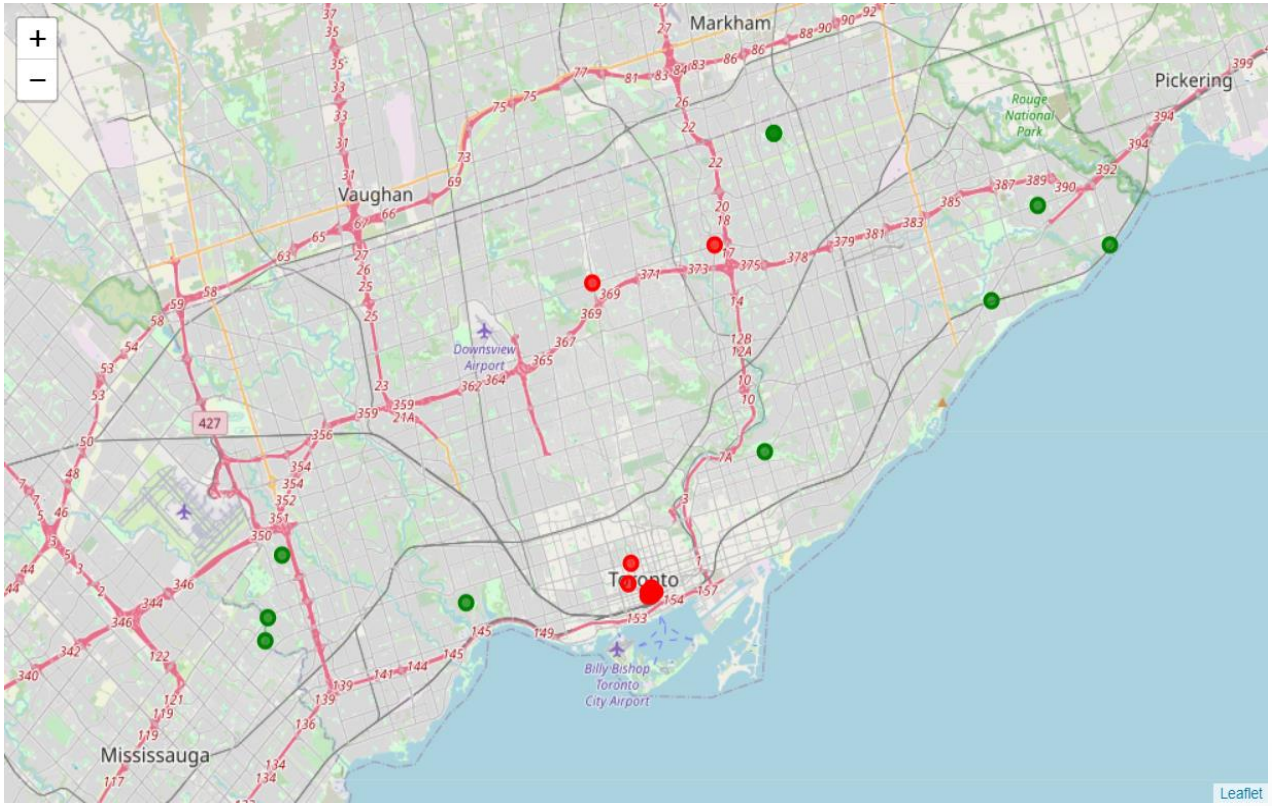
Results

- Top 10 neighborhoods with the highest and lowest supply of restaurants/coffee shops:

	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

	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

- Map of city of Toronto where the green dots represent the lowest supply, therefore guaranteeing little competition for a starting restaurant/coffee shop
- The red dots represent the highest supply which corresponds to a saturated market of restaurants/coffee shops.



Conclusion and Improvements

- This approach could benefit new businesses when evaluating ideal location
- To guarantee new businesses' success, an additional market study should be performed to confirm high demand within neighborhoods with low supply.
- Improvements:
  - Find correlation in between the amount of specific venues with the presence of other venues like schools, parks, banks, etc. Perform regression analysis to estimate the need for more specific venues in neighborhoods with similar characteristics
  - Clustering of neighborhoods on the basis of amount of specific venues and their location