The Battle of the Neighborhoods

Jiajie Li

July 17, 2019

1.   Introduction

1.1 Background

Running a restaurant is not an easy job. There are many factors, such as service, price, and taste, that might affect the restaurant. Among all the factors, location is to be considered as one of the main factors. Having an optimal location could help raise customers' awareness and hence attract customers; therefore, it is vital to the success of restaurant at the first stage. Many managers spend thousands of money hiring consultants to find a location for their restaurants. Therefore, it is advantageous for managers or restaurants owners to accurately predict whether a location is closed to their targeted customers and has fewer competitors. This information can be used to choose a location for their restaurants.

1.2 Business Problem

In this project we will try to find an optimal location for an Italian restaurant in Toronto. Specifically, this report will be targeted to stakeholders interested in opening an Italian restaurant in Toronto, Canada.

Since there are lots of restaurants in Toronto, we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Italian restaurants in vicinity. we would also prefer locations as close to Downtown Toronto as possible, assuming that first two conditions are met. Furthermore, we would like to have the restaurant as close as possible to the neighborhoods whose residents live Italian food.

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

2.   Data acquisition and cleaning

2.1 Data Sources

Based on definition of our problem, factors that will influence our decission are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Italian restaurants in the neighborhood, if any
- distance of neighborhood from city center

We decided to scrape the Wikipedia page that lists all the neighborhoods in Toronto and transform the data into a pandas dataframe.

Following data sources will be needed to extract/generate the required information:

- geographical coordinates of the neighborhoods using the **CSV file provided by IBM**
- coordinate of Toronto will be obtained using **Foursquare API**
- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**

2.2 Data cleaning

Data scraped from Wikipedia were transformed into one table. There were several problems with the datasets. Firstly, there were a lot of "not assigned" values, I decided to drop rows with a borough that is "not assigned". Secondly, there was a cell that has a borough but a "not assigned" neighborhoRood, I decided to make the cell the same as the borough. Thirdly, some rows had the same postal code and I combined these rows into one row. Fourthly, the dataset lacked the latitude and the longitude coordinates of each neighborhood. In order to use the Foursquare API we need to have the information, so we used IBM's csv file that has the geographical coordinates of each postal code to get the information.

3.   Exploratory Data Analysis

3.1 Density of restaurants

According to our analysis, on average Italian restaurant can be found within 5000m from every Neighborhood, that's very far. Especially in areas other than the Downtown Toronto, the density of restaurants is even lower.
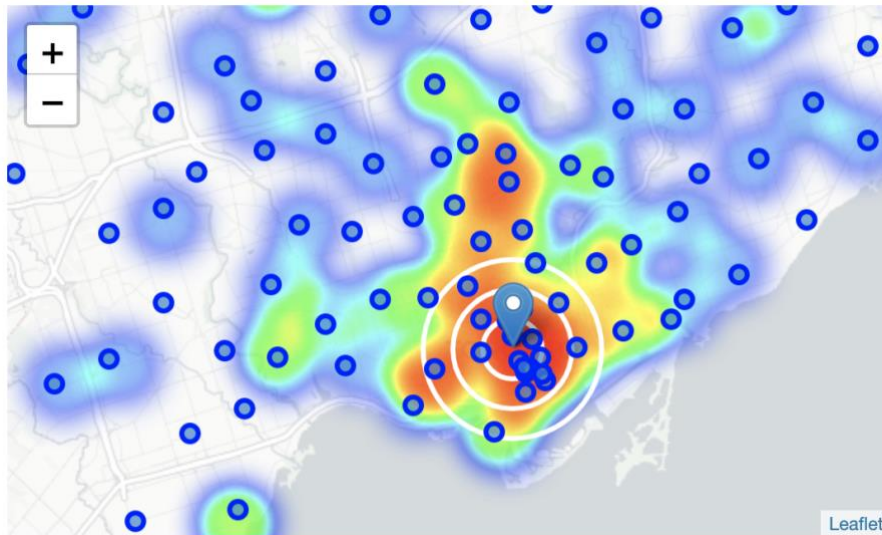

Figure 1. heatmap of restaurants in Toronto

Regarding the Italian restaurants, Italian restaurants represent a subset of ~9.1% of all restaurants in Toronto. Therefore, it is not surprised to see the map below, which is 'cooler' than the first map. It also indicates higher density of existing Italian restaurants directly in the south of Toronto, with closest pockets of low Italian restaurant density positioned east, west and North from city center.
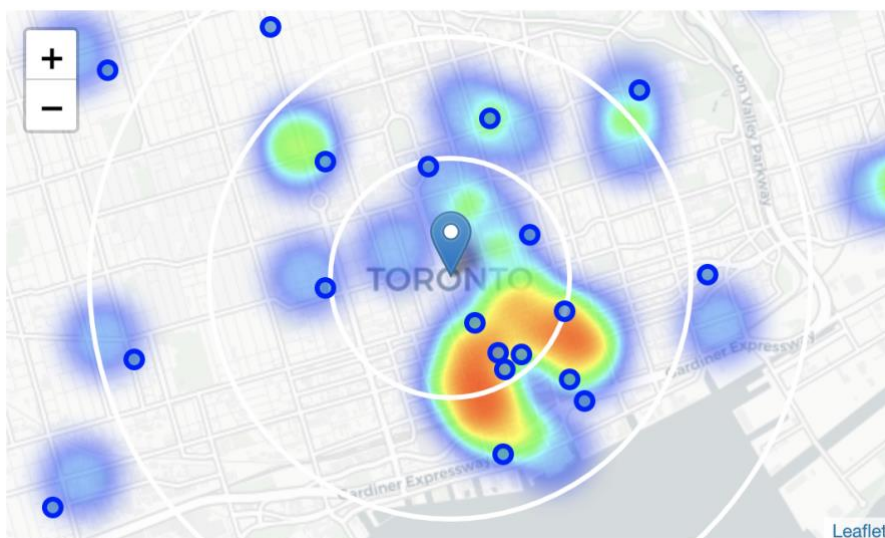

Figure 2. heatmap of Italian restaurants in Toronto

3.2 Neighborhoods that prefer Italian Restaurants

With the help of Foursquare, we can find the top 10 venues for each neighborhood. Based on the heatmap and the table above, we will now focus our analysis on Neighborhoods Deer Park, Forest Hill SE, Rathnelly and South Hill. There is no restaurant in these areas, and the cloest Italian restaurant is 3084 meters away. However, residents in these neighborhoods like Italian Restaurant, so we think these areas might be ideal for a new Italian restaurant.

Table 1 List of Neighborhoods that have 0 Restaurants within 400m of the center and the distance to the closest Italian restaurant is further than 3km

	Postcode	Borough	Neighborhood	Restaurants in area	Distance to Italian restaurant	Result
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	0	17128.57181	TRUE
26	M3B	North York	Don Mills North	0	3308.440237	TRUE
49	M4V	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hi...	0	3084.504403	TRUE
73	M6C	York	Humewood-Cedarvale	0	3703.697054	TRUE
74	M6E	York	Caledonia-Fairbanks	0	4047.503205	TRUE
97	M9M	North York	Emery, Humberlea	0	7061.250453	TRUE

4.   Analyze the targeted Neighborhoods

Deer Park, Forest Hill SE, Rathnelly and South Hill are in the heart of Toronto. Analysis of popular travel guides and web sites often mention Deer Park as beautiful, interesting and Commercial.

"*sandwiched between Forest Hill on its western flank and Moore Park to the east, Deer Park is utterly unlike either of them—it's more commercial, a fast-changing community dominated by apartment dwellers.*" (Robert Fulford)

We defined new, narrower region of interest, which included low-restaurant-count parts of Deer Park, Forest Hill SE and Rathnelly and found locations with no more

than two restaurants in radius of 250 meters, and no Italian restaurants in radius of 400 meters.
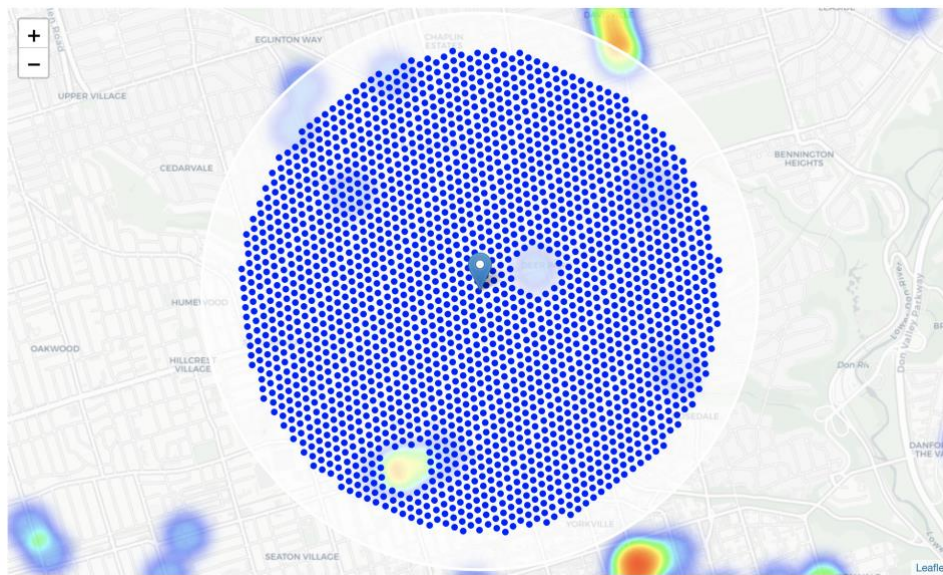

Figure 3. locations with low number of restaurants and no Italian restaurants nearby

The above map clearly indicates zones with low number of restaurants in vicinity, and no Italian restaurants at all nearby. We then used k-means to cluster these locations and create 15 centers of zones containing good locations.
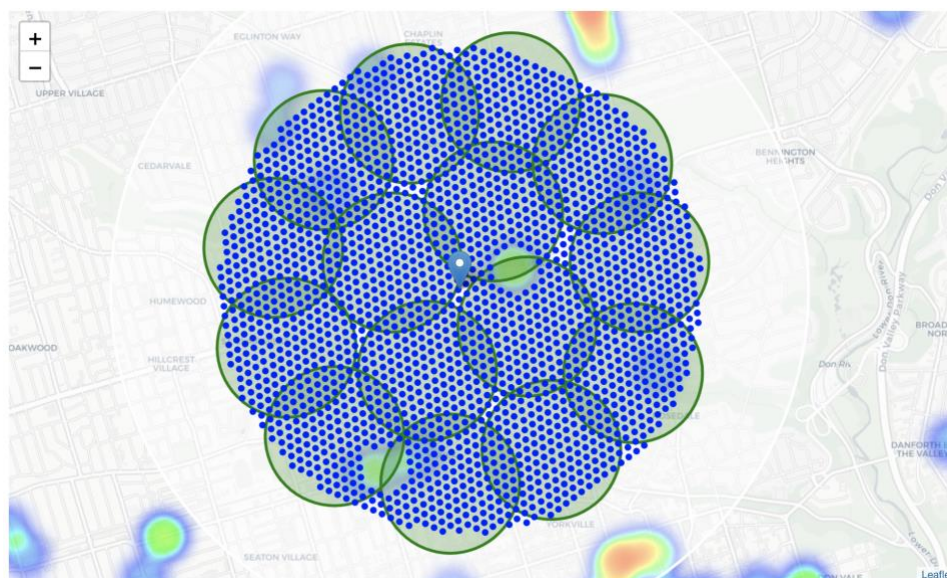

Figure 4. Clusters with low number of restaurants and no Italian restaurants nearby

We now have a bunch of locations fairly close to Deer Park, Forest Hill SE, Rathnelly and South Hill, and we know that each of those locations has no more than two restaurants in radius of 250m, and no Italian restaurant closer than 400m. Any of

those locations is a potential candidate for a new Italian restaurant, at least based on nearby competition. Below are the list and the map of these 15 addresses.

Table 2 List of 15 candidate addresses

| Address	Distance from the center
30, Hillsboro Avenue, Yorkville, University—Rosedale, M5R 1J3	1.9km
438, Russell Hill Road, Forest Hill, Toronto—St. Paul's, M5P 2T6	1.9km
134, Inglewood Drive, Moore Park, University—Rosedale, M4T 2B4	1.9km
97, Wells Hill Avenue, St. Clair West, Toronto—St. Paul's, M6H 3L9	1.9km
16, Oriole Gardens, Deer Park, Toronto—St. Paul's, M4V 2K5	0.9km
30, Woodlawn Avenue West, University—Rosedale, M4T 1W5	0.8km
584, Davenport Road, Tarragon Village, Toronto—St. Paul's, M5R 1K9	2.0km
7, Thornwood Road, Rosedale, University—Rosedale, M4W 1R1	2.0km
oliver mowat, 1, Moore Avenue, Moore Park, University—Rosedale, M4T 2B4	2.0km
188, Heath Street West, Forest Hill, Toronto—St. Paul's, M4V 1R8	0.8km
71, Davisville Avenue, Chaplin Estates, Toronto—St. Paul's, M4S 2Y9	2.0km
33, Russell Hill Road, South Hill, Toronto—St. Paul's, M4V 1R8	0.9km
29, Millbank Avenue, Forest Hill, Toronto—St. Paul's, M5P 2W3	2.0km
99, Admiral Road, The Annex, University—Rosedale, M2R 1K5	2.0km
121, Chaplin Crescent, Chaplin Estates, Toronto—St. Paul's, M5P 2K4	1.9km
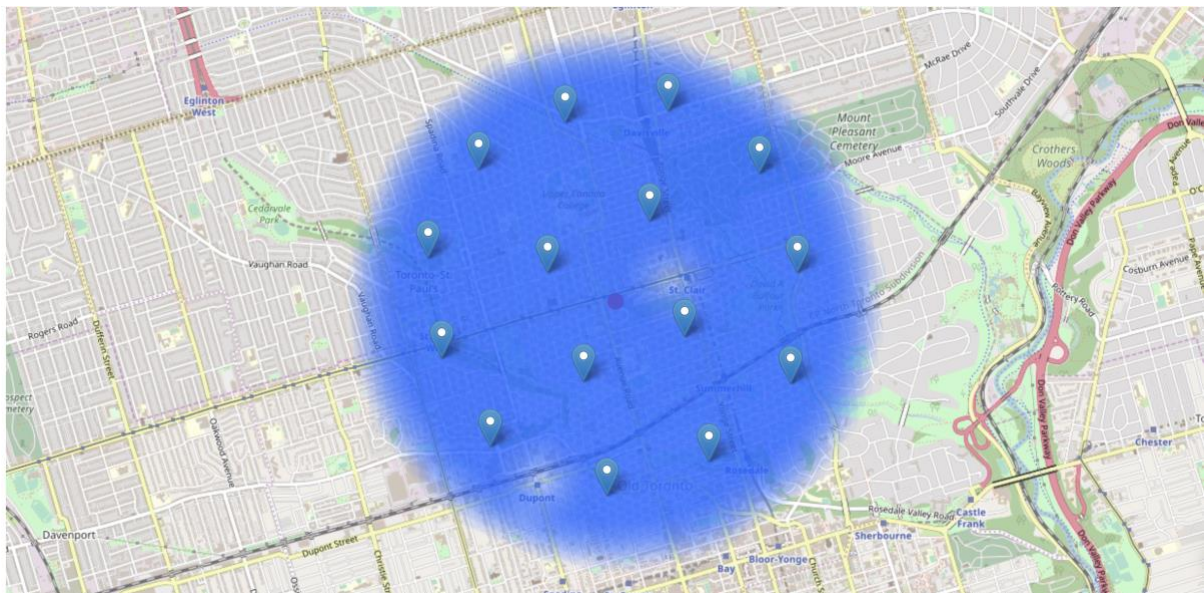
Figure 5. Map of the 15 clusters

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Italian restaurants nearby, all zones being fairly close to city center (all less than 2km from Deer Park,

Forest Hill SE, Rathnelly and South Hill). However, these centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations. All of the zones are located in Deer Park, Forest Hill SE, Rathnelly and South Hill, which we have identified as interesting due to being popular with residents, fairly close to Downtown Toronto and well connected by public transport.

5.   Results and Discussion

Our analysis shows that although there are 651 restaurants in Toronto, there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected south from Toronto, so we focused our attention to areas north, east and west, corresponding to neighborhoods Deer Park, Forest Hill SE, Rathnelly and South Hill, which offer a combination of preference for Italian restaurant among residents, closeness to city center, strong socio-economic dynamics and a number of pockets of low restaurant density.

After directing our attention to this narrower area of interest, we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Italian restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Italian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Downtown Toronto but not crowded with existing restaurants (particularly Italian) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which

would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition, but also other factors taken into account and all other relevant conditions met.

6. Conclusion

Purpose of this project was to identify areas close to Downtown Toronto with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. By calculating restaurant density distribution and filtering neighborhoods based on venues preferences from Foursquare data we have first identified several neighborhoods that justify further analysis (Deer Park, Forest Hill SE, Rathnelly and South Hill), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

7. Future directions

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.