**The Battle of Neighbourhoods**

***Introduction/Business Problem***

**Problem Statement-**

**Whenever someone looks for opening up a restaurant or any other business venture the main problem faced by them is choosing a location for the set up.**

**But in this era of high knowledge availability we have come up with a solution where using the neighbourhood, location of different points we can optimize the best location problem.**

**For this purpose, I am using Toronto data set to come up with best possible solution for the problem as stated above.**

**Solution for the problem-**

In this project we will try to find an optimal location for an Italian restaurant. 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 on the basis of

1. not already crowded with restaurants.
2. areas with no restaurants in vicinity.
3. close to city centre (downtown Toronto) as possible, assuming that first two conditions are met.

**About Data-**

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

For this problem we'll be required to explore, segment, and cluster the neighbourhoods in the city of Toronto. However, the neighbourhood data is not readily available on the internet. For the Toronto neighbourhood data, a Wikipedia page exists [[https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M]](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M%5D) that had all the information I need to explore and cluster the neighbourhoods in Toronto. I scraped the Wikipedia page and wrangled the data, cleaned it, and then read it into a pandas data frame so that it is in a structured format to proceed with my analysis.  
Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario’s north-western shore. It's a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower. Toronto also has many green spaces, from the orderly oval of Queen’s Park to 400-acre High Park and its trails, sports facilities and zoo.

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

* number of existing restaurants in the neighbourhood (any type of restaurant)
* number of and distance to Italian restaurants in the neighbourhood, if any
* distance of neighbourhood from city centre

We decided to use regularly spaced grid of locations, centred around city centre, to define our neighbourhoods.

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

* For coordinated we'll be using a csv file available with postal code and latitude and longitude information.
* number of restaurants and their type and location in every neighbourhood will be obtained using **Foursquare API**
* coordinate of centre will be obtained using **Google Maps API geocoding**.



Fig: Dataset finally being used for EDA and modelling

***Methodology***

* In this project we will direct our efforts on detecting areas of canada that have low restaurant density, particularly those with low number of Italian restaurants. We will limit our analysis to area ~6km around city centre.
* In first step we have collected the required **data: location and type (category) of every restaurant close to Downtown Toronto**. We have also **identified Italian restaurants** (according to Foursquare categorization).
* Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas near downtown - we will use **heatmaps** to identify a few promising areas close to downtown with low number of restaurants in general (*and* no Italian restaurants in vicinity) and focus our attention on those areas.
* In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no more than two restaurants in radius of 250 meters**, and we want locations **without Italian restaurants in radius of 400 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighbourhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

After out analysis:

* Total number of restaurants: 483
* Total number of Italian restaurants: 40
* Percentage of Italian restaurants: 8.28%

***This is good number to start as no is very less.***

* Looking good. So now we have all the restaurants in area within few kilometres from Downtown Toronto, and we know which ones are Italian restaurants! We also know which restaurants exactly are in vicinity of every neighbourhood candidate centre.
* This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Italian restaurant!

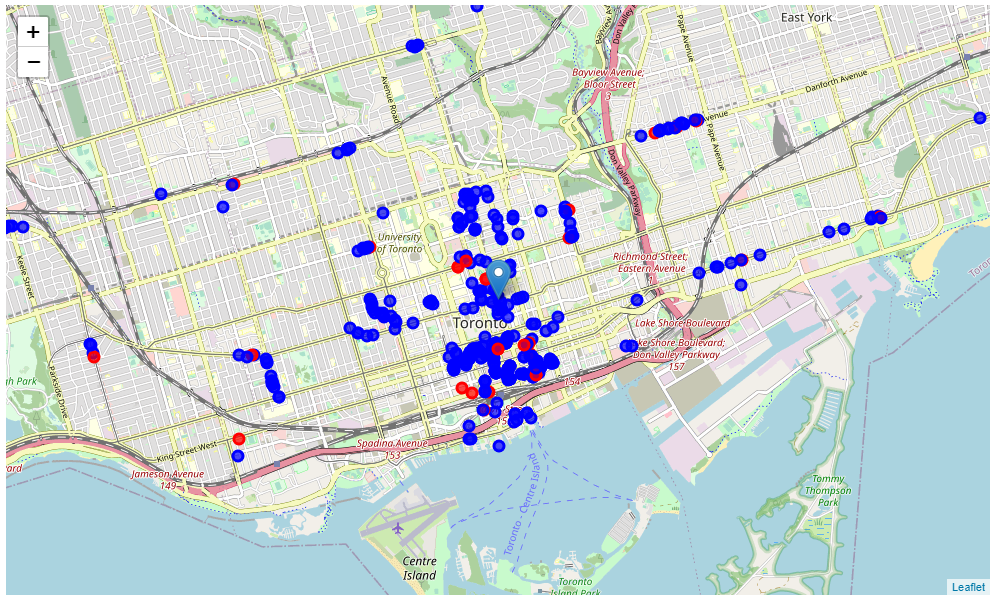
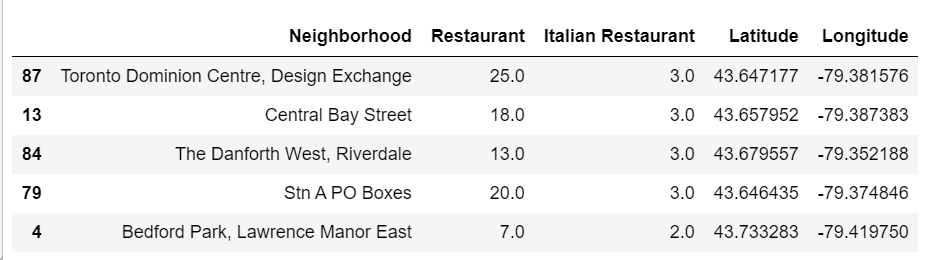


Fig: Red shows italian restaurants, blue shows restaurants

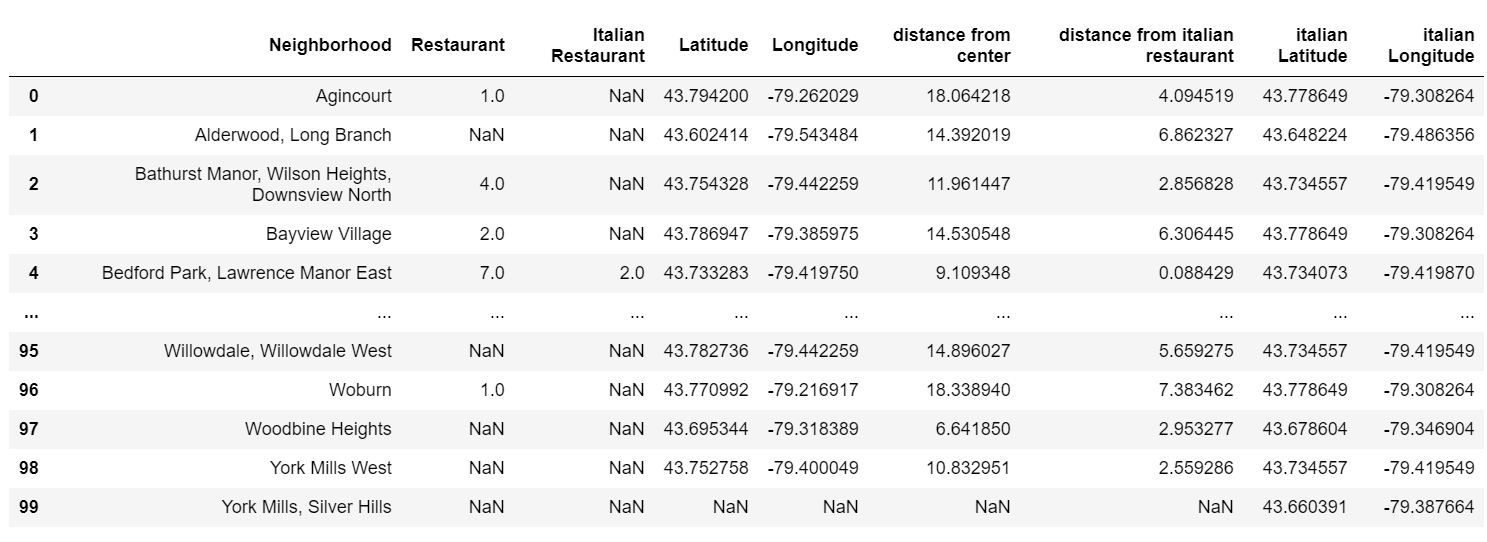
The data is now for location with no. of Italian restaurant and restaurants. Along with coordinates.



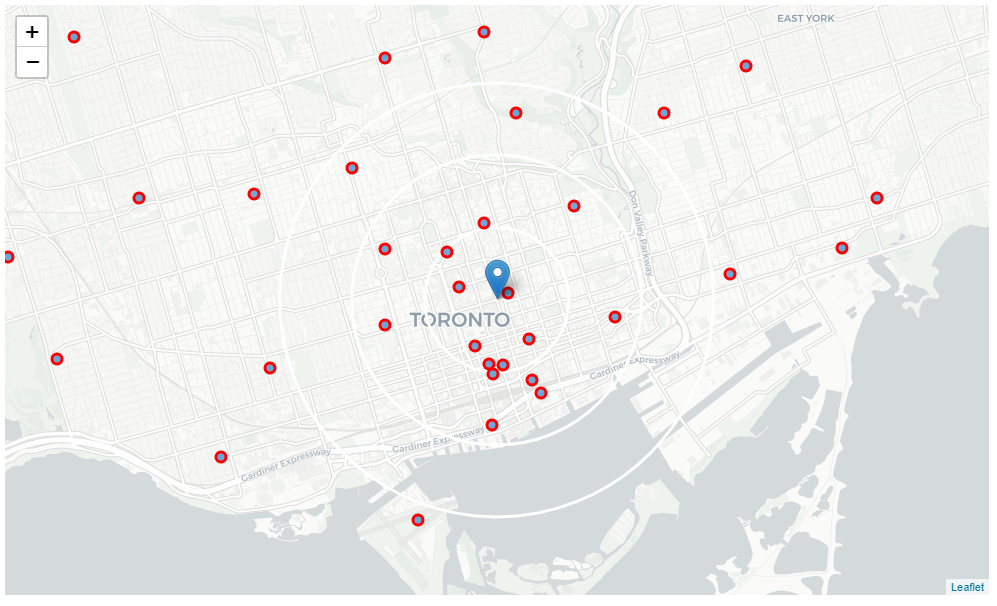
Number of localities with no italian restaurant is 74

Number of localities with no restaurant is 36

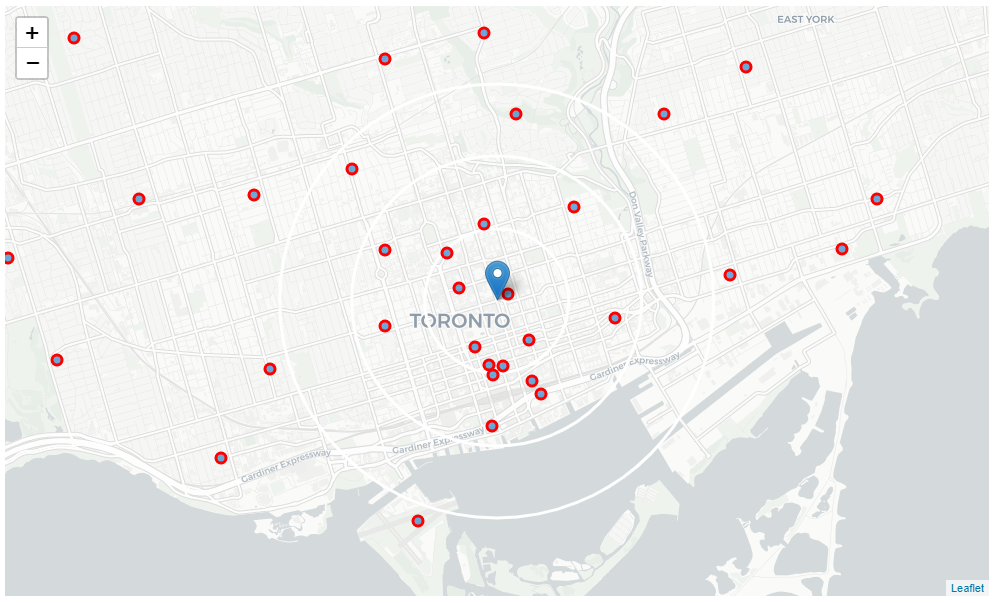
After this I added latitude and longitudes of nearest Italian restaurants as well as distance from centre and Italian restaurants.



Restaurants in radius of 2000,3000 and 4000 around downtown toronto.



Italian restaurants in radius of 2000,3000 and 4000 around downtown Toronto.



Locations with no more than two restaurants nearby: 22

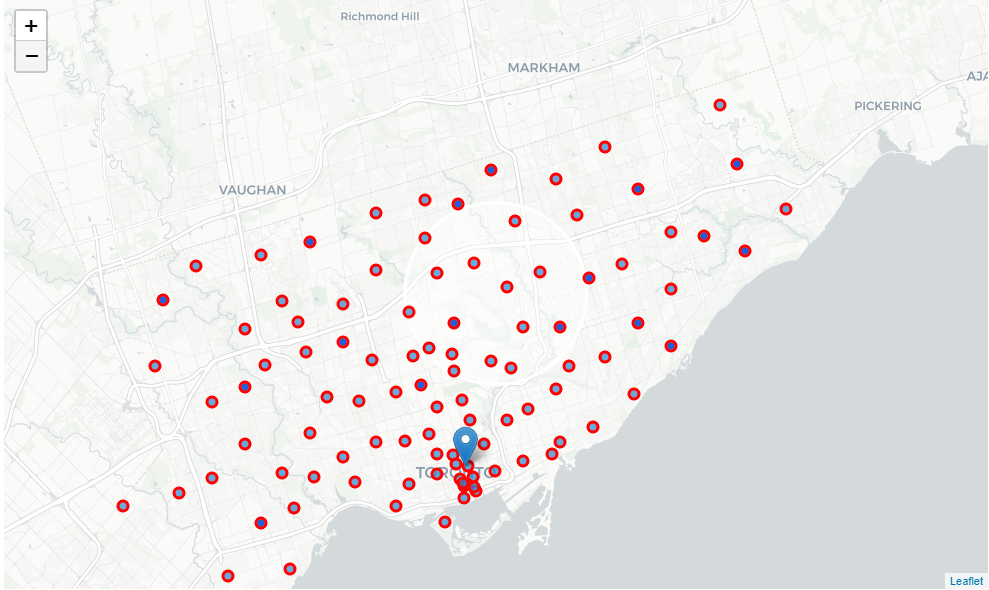
Locations with no Italian restaurants within 200m: 58

Locations with both conditions met: 17

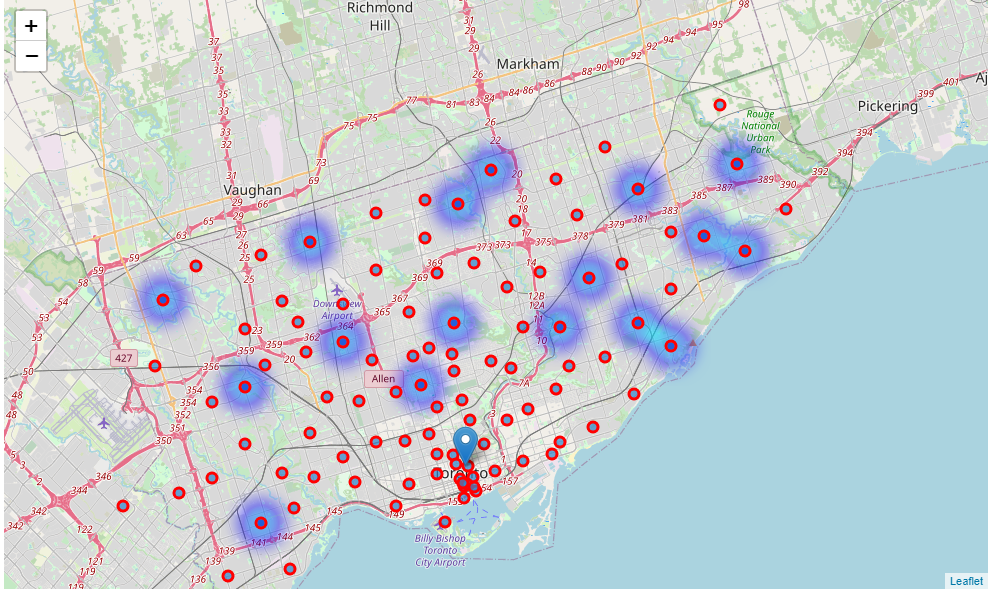
According to above conditions good locations are:



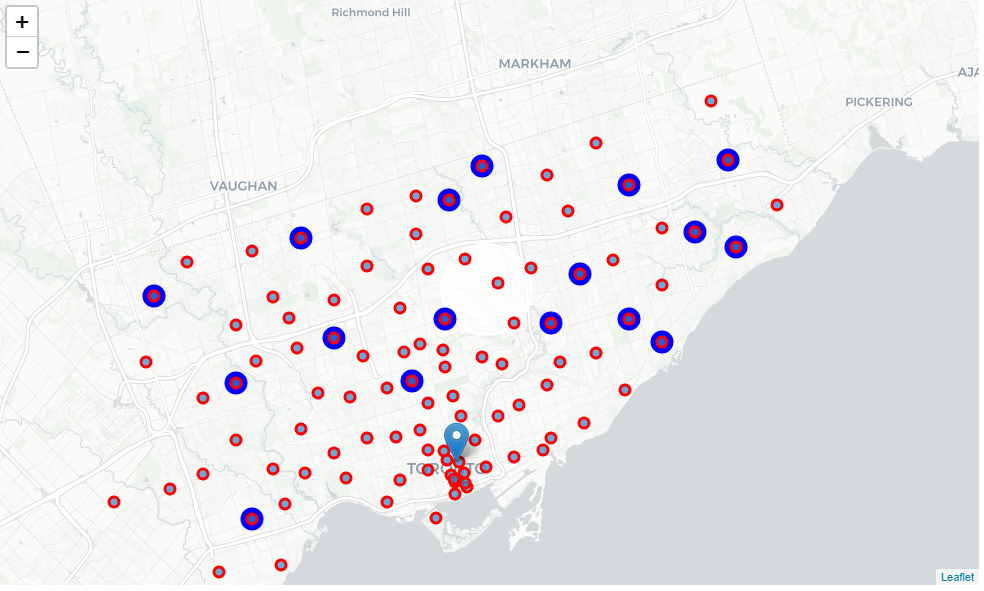
All the restaurants and Italian restaurant.



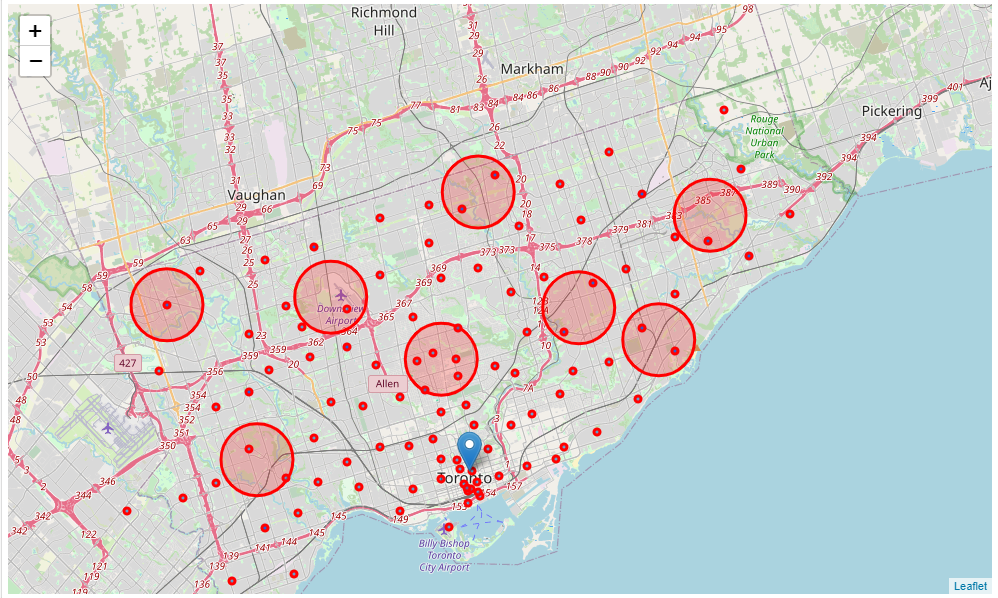
All the good locations centres:



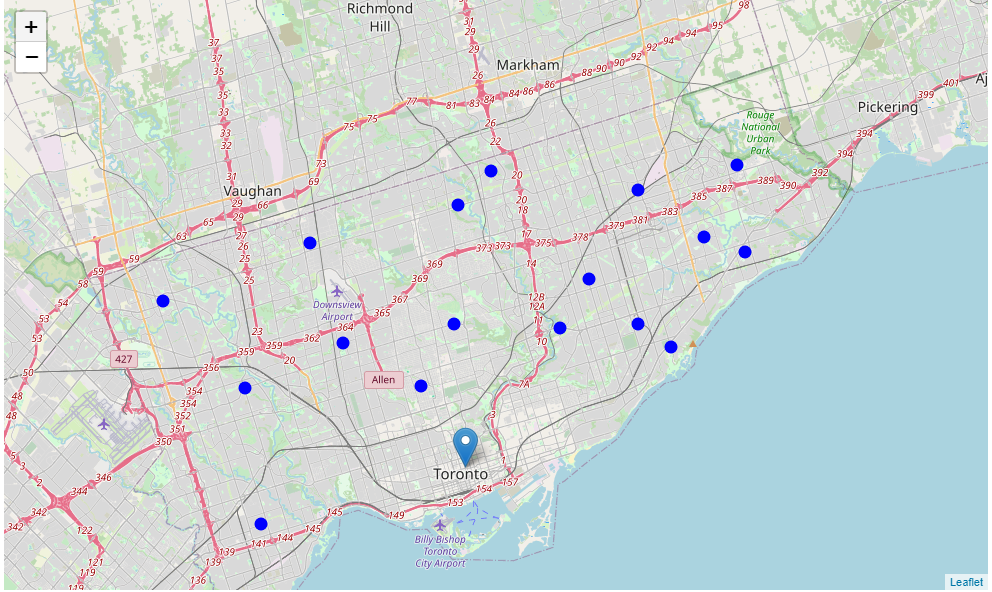
All the good locations:



Best possible clusters:



Best locations from these clusters:



## *Results and Discussion*

Our analysis shows that although there is a great number of restaurants in Toronto (~800 in our initial area of interest which was 12x12km around downtown toronto), there are pockets of low restaurant density fairly close to city center. We have high availablity near airport where there is less no of restaurants.

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 8 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.

## *Conclusion*

Purpose of this project was to identify Toronto area close to downtown 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 from Foursquare data we have first identified general boroughs that justify further analysis , 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.

Final decission 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.