# Capstone Project - The Battle of the Neighborhoods (Week 2)

### Applied Data Science Capstone by IBM/Coursera

## Introduction

Through this study we will analyze the neighborhoods of Toronto and identify the best area, theoretically, to open an Italian restaurant. The purpose of this analysis is to serve as a basis to introduce and attract potential investors from the food industry and / or local families looking for a new investment, thus stimulating the city's economy. The location in which a business is established is fundamental to its success, especially when it comes to restaurants, so it is important to study the various possible locations in order to reduce the risks of the investment failing.

## Data

To carry out this study we will mainly use FourSquare, a technology company that built a massive dataset of location data. We will also use a table containing the postal codes of Canada, imported from Wikipgia, as well as a csv file with its coordinates. These will be the necessary tools to explore and analyze the different neighborhoods of Toronto to identify different parameters, such as: number of restaurants, proximity to Italian restaurants, among others.

### Procedures

#### Importing a Dataframe from a Wikipegia page containing postal codes of Canada’s boroughs

#### Preparing and cleaning the data (removing NaN values etc)

#### Reading a csv file with coordinates data and merging it with our dataframe

#### Restricting 'Borough' for places that contain 'Toronto'

#### Using geopy library to get the latitude and longitude values of Toronto (for map purposes)

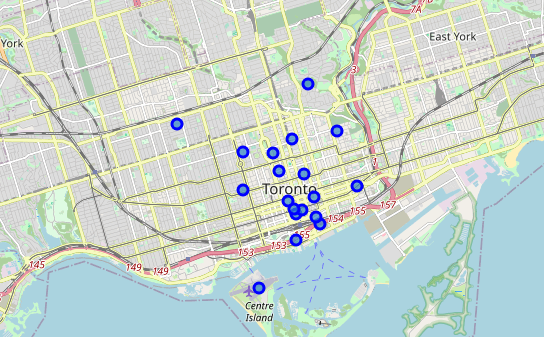
#### Figure 1. Map of Toronto neighborhoods

#### C:\Users\Silvio\Desktop\1.png

#### Restricting Neighborhoods that are part of Downtown Toronto

Downtown Toronto is a buzzing area filled with skyscrapers, restaurants, nightlife, and an eclectic mix of neighborhoods. It’s also home to iconic attractions like the CN Tower, St. Lawrence Market, and the Royal Ontario Museum, with exhibits on natural history. Bloor Street is an upscale shopping area, and the Eaton Centre is a huge, multistory mall. On the lake, the Harbourfront area has parks and cultural venues. Because of the characteristics listed above we chose this area as our interest zone which we will analyze in the next section.

#### Figure 2. Map of Downtown's neighborhoods



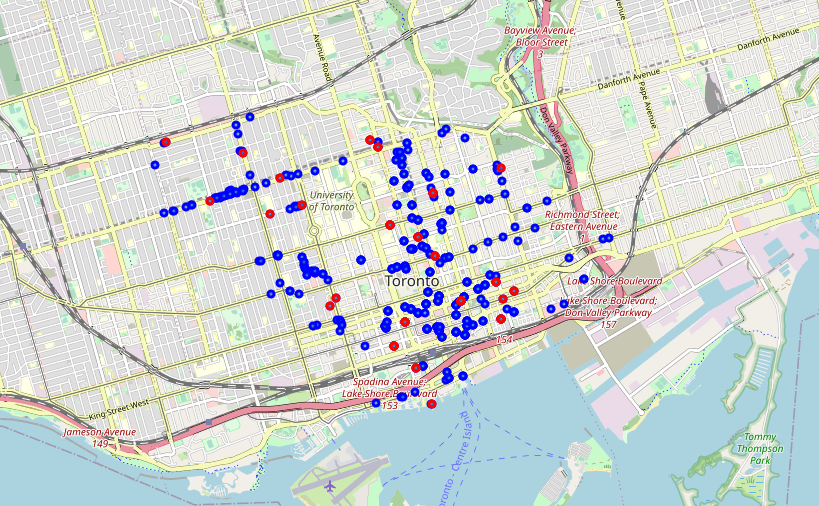
#### Exploring Neighborhoods with FourSquare

#### Firstly we got all the venues in a radius of 1000 meters from the center of each neighborhood

#### Then the venues were filtered for restaurants

#### Same step as above but now for Italian restaurants

#### Figure 3. Map of Toronto with blue circles representing all restaurants and red circles representing Italian restaurants



## Methodology

In the first step we collected the required data: location and type (category) of every restaurant within 1km from each neighborhood center in the Downtown of Toronto. We have also identified Italian restaurants (according to Foursquare).

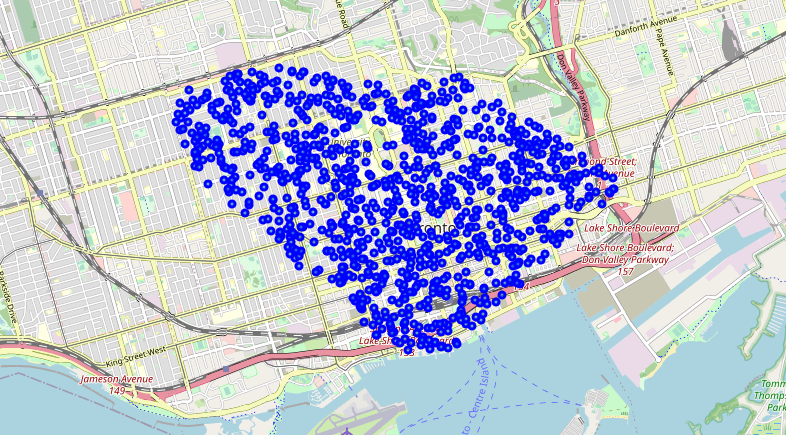
Second step in our analysis will be calculation and exploration of numbers of restaurants across 1000 random generated locations inside our areas of interest - we will filter those locations to identify promising areas with low number of restaurants in general (and no Italian restaurants in vicinity) and focus our attention on those areas.

In the 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 four and at least two restaurants in radius of 250 meters (looking for possible gastronomic centers), 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 / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

## Analysis

#### Creating random locations inside our interest area

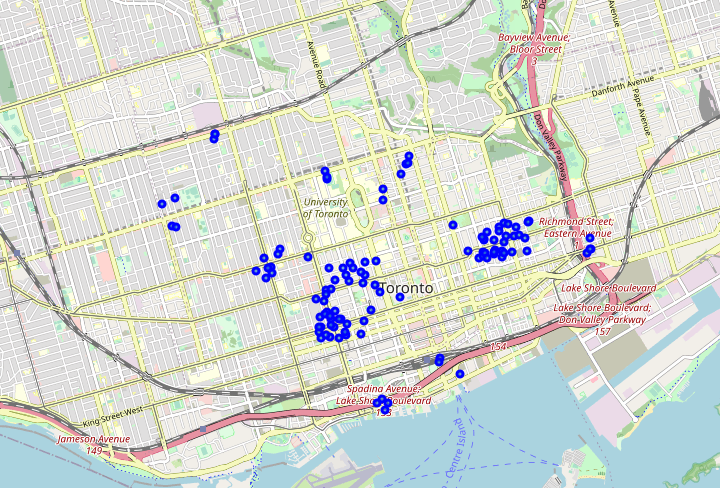
#### Figure 4. Map of random generated locations



#### Preparing the data acquired before to create coordinates lists and defining a function to measure distance between two locations

#### Filtering the locations that fit the established parameters

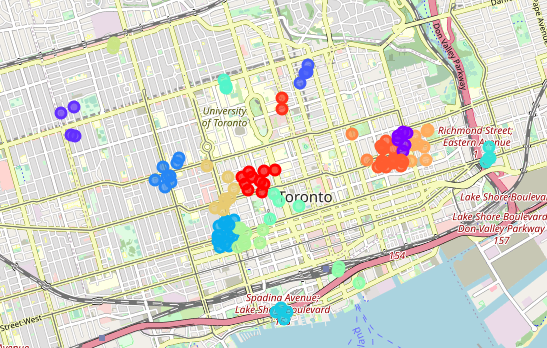
#### Figure 5. Map of good locations



#### Clustering good locations to find it centers

Were defined 18 clusters, and numeric labels for each of them.

#### Figure 6. Map of Toronto with the good locations clustered in different colors



#### Using reverse geocoding to identify the address of each cluster center

==============================================================

Addresses of centers of areas recommended for further analysis

==============================================================

162 McCaul St, Toronto, ON M5T 1W4, Canada

350 Parliament St, Toronto, ON M5A 2Z7, Canada

850 A Bloor St W, Toronto, ON M6G 1M2, Canada

45 Charles St E, Toronto, ON M4Y 1S2, Canada

431 College St, Toronto, ON M5T 1T1, Canada

3 Cameron St, Toronto, ON M5V 2A9, Canada

318 Queens Quay W, Toronto, ON M5V 3A7, Canada

630 Queen St E, Toronto, ON M4M 1G3, Canada

110 Devonshire Pl, Toronto, ON M5S 2C9, Canada

225 Simcoe St, Toronto, ON M5G 1S4, Canada

18 Lake Shore Blvd W, Toronto, ON M5E 1Z8, Canada

134 Peter St, Toronto, ON M5V 2H2, Canada

888 Palmerston Ave, Toronto, ON M6G 2S2, Canada

307 Spadina Ave, Toronto, ON M5T 2E6, Canada

500 Dundas St E, Toronto, ON M5A 3V3, Canada

368 George St, Toronto, ON M5A 2N3, Canada

142 Seaton St, Toronto, ON M5A 2T3, Canada

1000 Bay St, Toronto, ON M5S 3A6, Canada

This concludes our analysis. We have created 18 addresses representing centers of zones containing locations with low number of restaurants and no Italian restaurants nearby. Although it's important to emphasize that the 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 Downtown Toronto boroughs, which we have identified as interesting due to being popular with tourists, close to city center and well connected by public transport.

## Results and Discussion

Our analysis shows that although there is a great number of venues in Toronto, there are pockets of low restaurant density in Downtown Toronto. After directing our attention to this more narrow area of interest we first created a dense list of location candidates; those locations were then filtered so that those with more than four and less than two (to identify possible gastronomic centers) 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 18 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. The purpose of this analysis was to only provide info on areas in 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 Downtown Toronto areas 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. We 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 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, price, social and economic dynamics of every neighborhood