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.

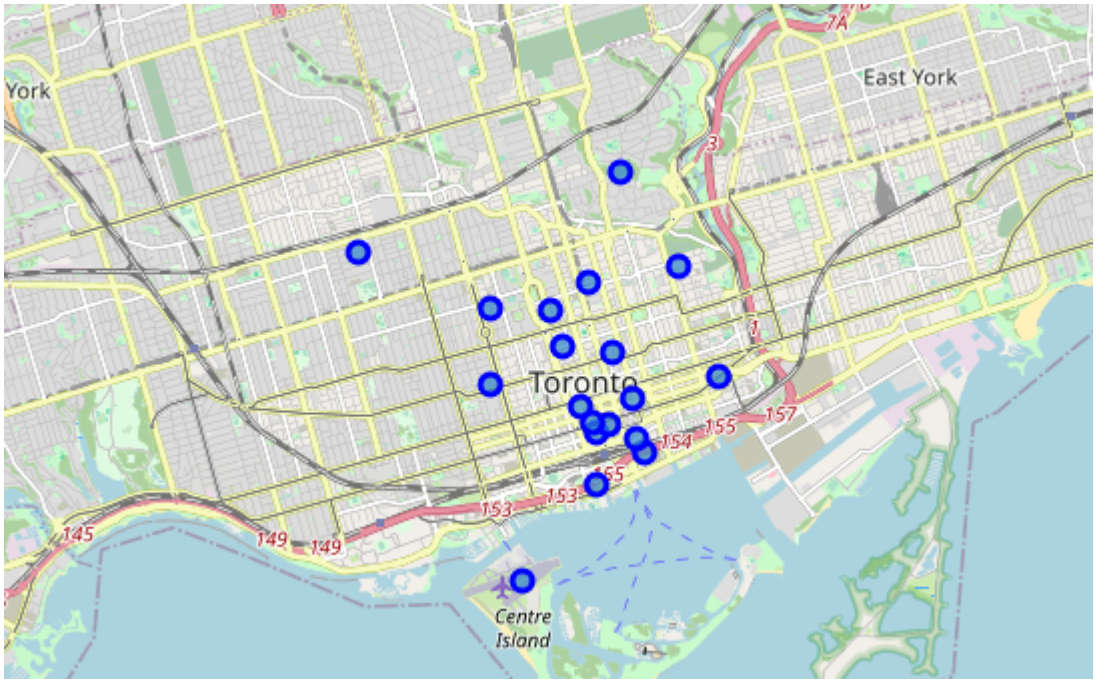
Procedures

- *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*

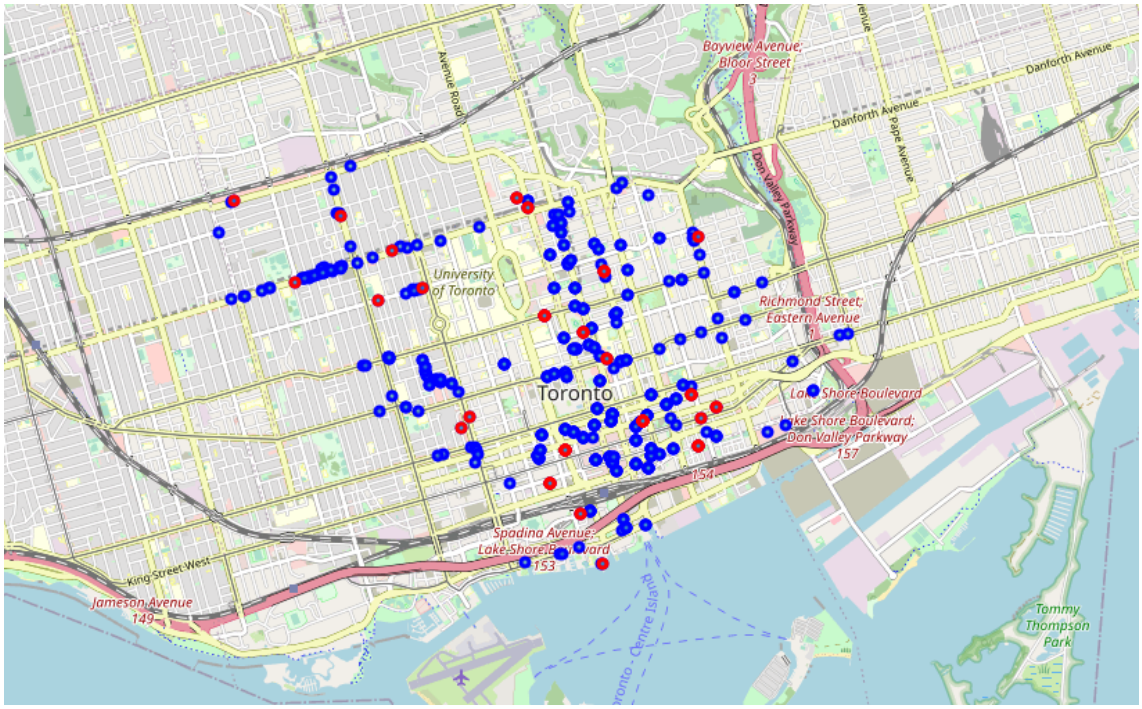
- *Restricting Neighborhoods that are part of Downtown Toronto*

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

The first part of our study was aimed at collecting and preparing the data so that we could analyze Our area of interest was defined as Downtown Toronto and within this area we were able to collect all the restaurants, including Italians, within a radius of 1 km from neighborhoods centers.

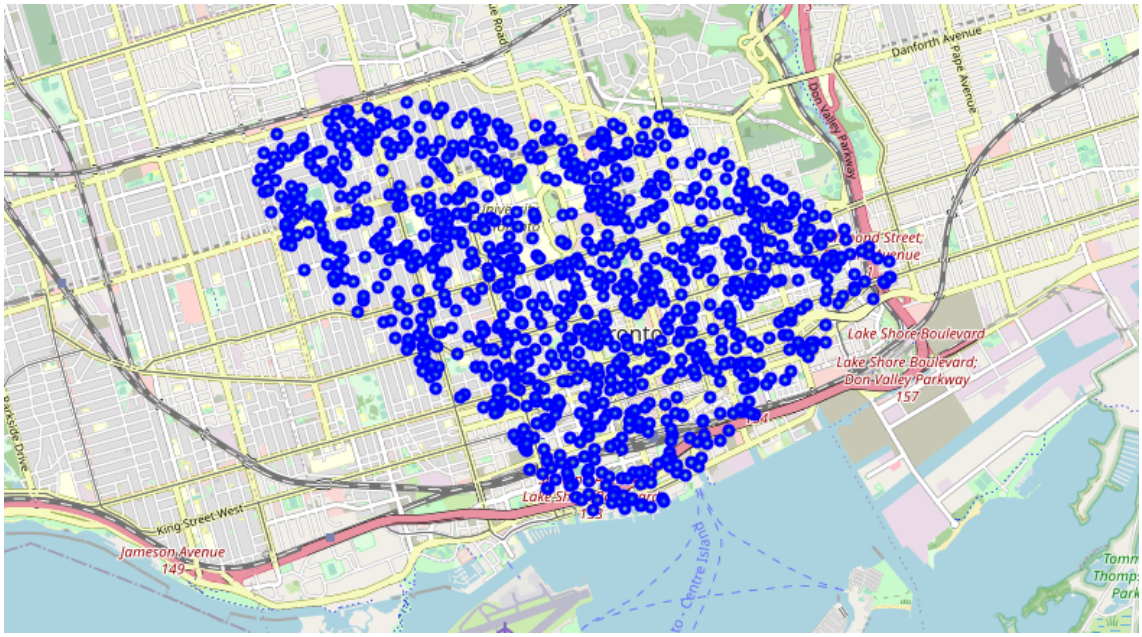
Then we will go into the analysis section, which will be divided into two parts. The first will be responsible for randomly generating 1000 locations within our area of interest and filtering them according to the following parameters: having less than four and at least two restaurants (possible gastronomic centers) within 250 meters and no Italian restaurant in 400 meters.

In the second part we will use the machine learning model called k-means clustering, which will help us to identify promising zones, by grouping the ideal locations and generating centers for each of the 18 clusters. Finally, we will convert the clusters centers to addresses using reverse geocoding.

Analysis

- *Creating random locations inside our interest area*

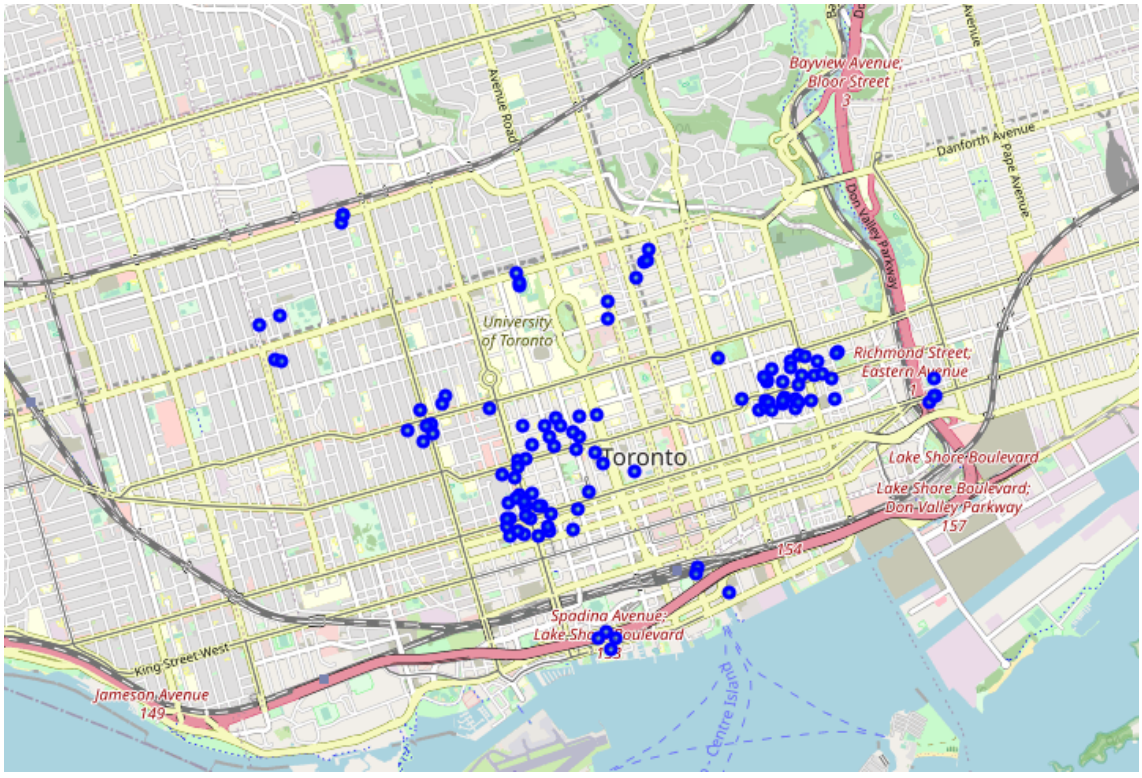
- *Figure 4. Map of random generated locations*



- *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*

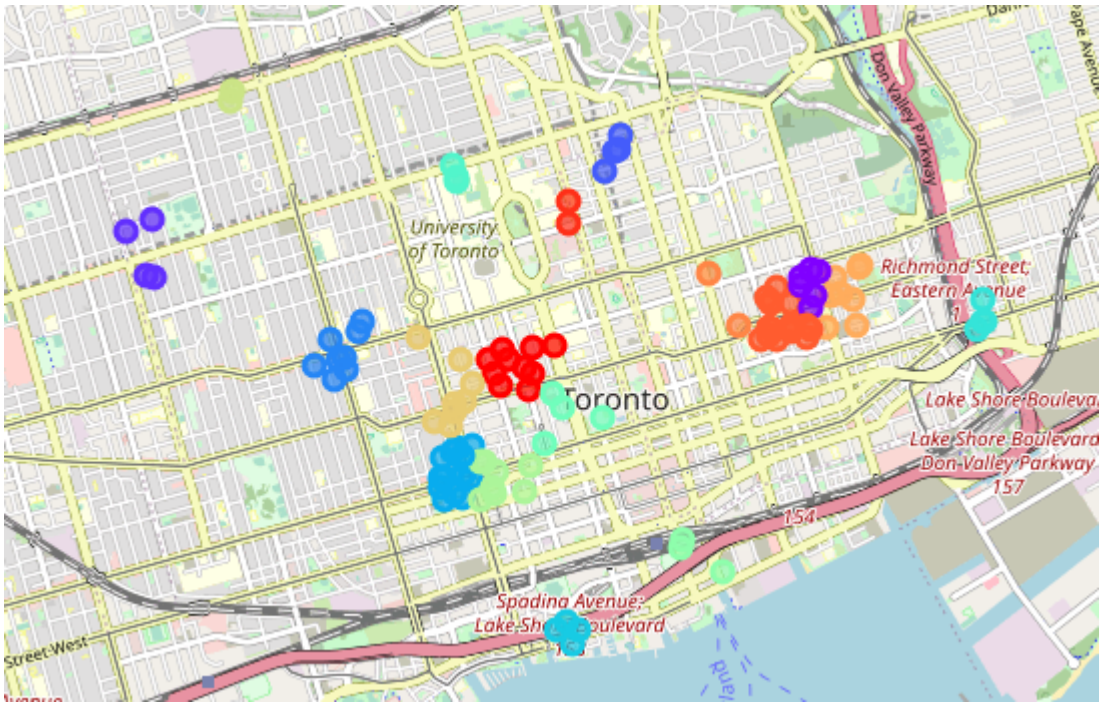
- *Figure 5. Map of good locations*



- *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
```

Results and Discussion

We can observe that although there are a relatively large number of venues in the Downtown Toronto area, there are still possible zones to be explored with low concentration of restaurants.

Our result was to generate 18 addresses with potential for stakeholders to invest in an Italian restaurant (based on proximity to other restaurants). We have acquired information about those locations, however it does not indicate that these addresses are really ideal for opening an italian restaurant.

Future studies are necessary, because even with low competitiveness there may be other factors that make unsustainable for new restaurants to be setted in those locations.

Conclusion

The aim of this project was to identify possible promising areas for opening an Italian restaurant in Toronto. By collecting information about the city we could see that the Downtown area was the most interesting for the study, because it has the biggest touristic center, busy streets etc. With the downtown Toronto region defined as our zone of interest, we generated 1000 random locations within that zone and filtered them to meet some parameters about nearby restaurants. So, we were able to group these locations into zones or clusters, which addresses are the final product of the work, to be used in the future as a starting point for stakeholder evaluation.

It is worth mentioning that although it is a starting point, the addresses obtained should not be readily considered ideal areas for opening the restaurant, because there are many other factors to be considered such as: proximity to attractions, commerce, price, availability of location, and the socio-economic dynamics of the neighborhood.