Sanjay Pramanik

October 22, 2018

Opening an italian restaurant in toronto

IBM Applied Data Science Capstone from Coursera

# Background

A friend of mine will be moving to Toronto next year and is thinking about opening an Italian restaurant. The Italian cuisine is one of the best in the world, and opening a restaurant has always been a big dream of my friend. However, I have never been to Toronto, nor has my friend, so a location has to be chosen from the desk. My friend would like to find a location without any Italian restaurants at this point, in order to avoid competition.

Data from various sources will be used for this project. Wikipedia will be used to extract geological details of the Toronto neighbourhoods. Foursquare location data will be used in order to come up with the best possible location for this restaurant.

# Problem

Opening a restaurant in a new place is challenging, especially if you have never been in such a place. This project will look into the best neighbourhood in Toronto to open an Italian restaurant in. The client wants to find a neighbourhood without any other Italian restaurants.

# Data Collection

The following sources will be used to collect data from:

**Wikipedia**

Wikipedia will be used to extract the neighbourhood data. The website <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> will be the main source for neighbourhoods in Toronto.

**Geospatial data**

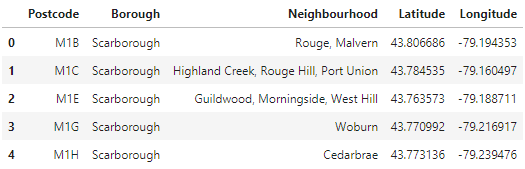
The geospatial data from <https://cocl.us/Geospatial_data/Geospatial_Coordinates.csv> will be used to collect the geospatial coordinates of the different neighbourhoods.

**Foursquare**

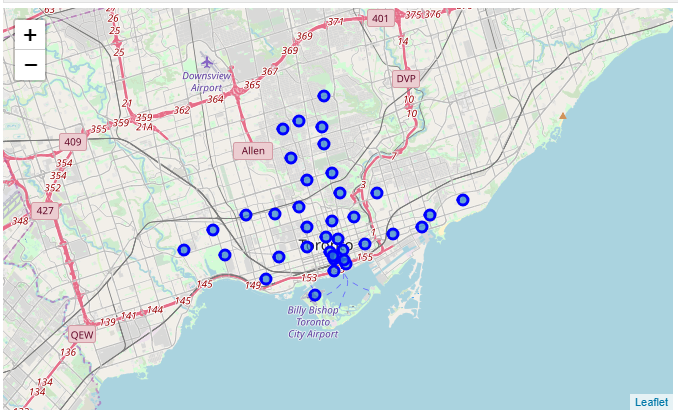
The Foursquare API will be used to find venues on different locations and find top venues per neighbourhood.

# Methodology

After loading all necessary Python packages, a list of Canadian neighbourhoods were extracted from <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>. The data was filtered, an all neighbourhoods of Toronto were listed. The geospatial data was extracted from <https://cocl.us/Geospatial_data/Geospatial_Coordinates.csv> and the postcode, latitude and longitude were added to the neighbourhood data.



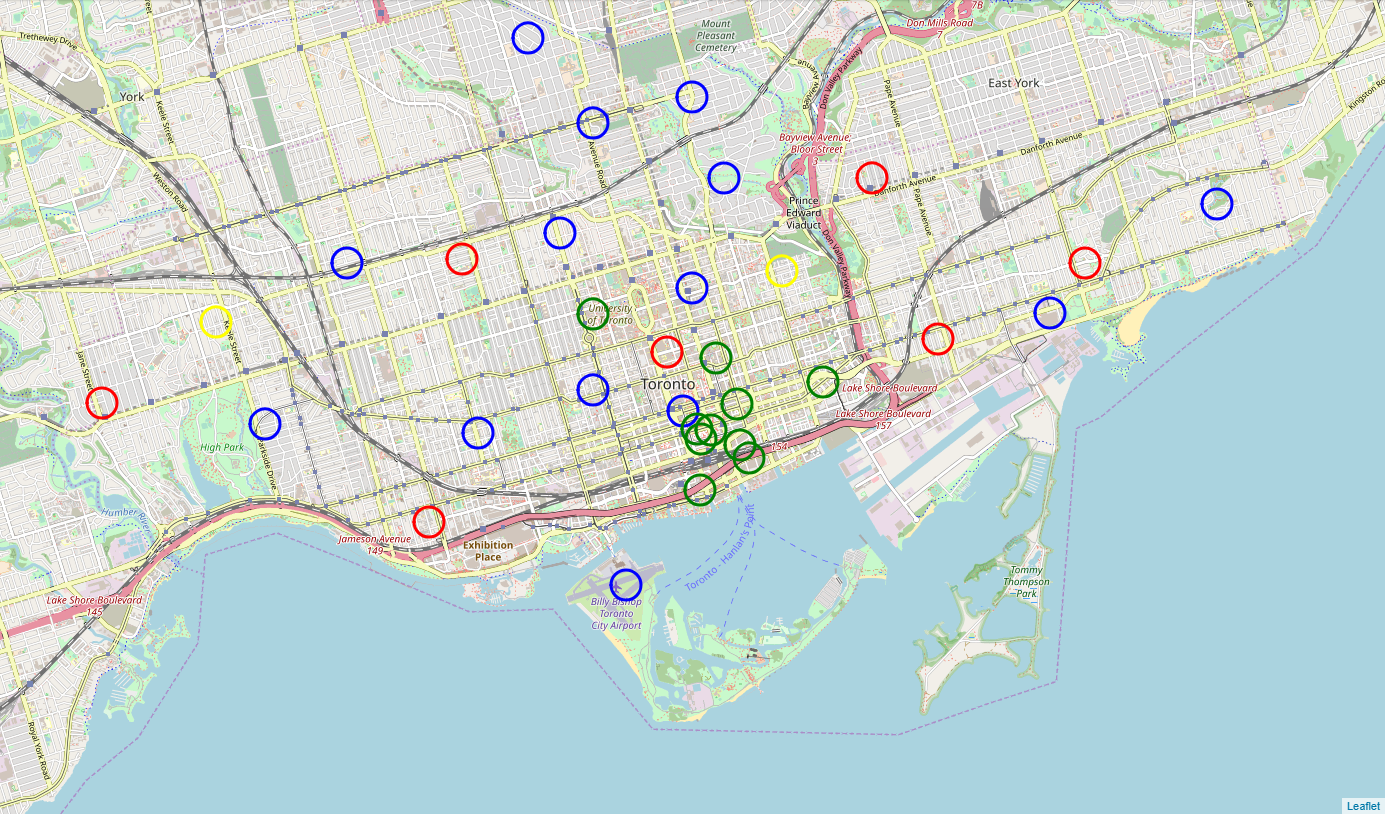
A map was now created of the different neighbourhoods in Toronto.



Now the Foursquare API was used to extract all the Italian Restaurants in Toronto grouped by neighbourhood and sorted by frequency. From this data, it follows that most Italian restaurants are currently located in Christie, Runnymed, Swansea and Davisville, as shown in the following table.



This data was merged with the latitude and longitude data and a heat map was made. Red and yellow indicate that there are already several restaurants in the area. Blue and green indicate that there is no or very little competition in the area.



We were able to extract all green and blue locations, the green locations indicate little competition by other Italian restaurants, the blue ones have almost none to no Italian restaurants present. The recommendation is to look at these neighbourhoods for future investments.

# Results

Surprisingly, there are 28 neighbourhoods in Toronto with little or no Italian restaurants. There is therefore a big market that could be reached by moving into one of those areas.

# Discussion

We have now looked at suburbs that have little or no Italian restaurants. To make a final decision demographics should be taken account as well as suburbs with Italian restaurants that are not highly recommended. It could be that having competition is not the worst thing as long as you have a good concept for a restaurant.

# Conclusion

Data science can be used to make a decision based on only data. We have now made a selection of neighbourhoods without any knowledge of the city. Data science is therefore very powerful.