

**Coursera IBM Data Science Capstone Project:**  
**Opening a New Italian Restaurant in Toronto**



**Report prepared by Tommaso Colli**

**June 2020**

### Introduction:

For this Capstone Project I am evaluating the opening of a new Italian restaurant in Toronto, and the area where it would be best to open it.

The basis of this valuation is the following: there is a high concentration of Italians in Toronto; namely, over 180,000 Italians, accounting for 7% of the overall population.

Furthermore, Italian cuisine is considered the most appreciated cuisine in the world and, as an Italian living abroad, I know that many times the authenticity of the food provided by Italian restaurants abroad is debatable. So, a new Italian restaurant opened under the supervision of an Italian might be a great recipe for success.

Through a quick search, I have found out that most of the Italian population of Toronto lives in Etobicoke Centre.

This Project will have to prove whether that location might be ideal, or if another location would be more suitable.

### Goal:

As a further clarification, the goal of this Project is to find the most suitable location for a new Italian restaurant in Toronto, Canada.

### Target Audience:

The entrepreneur willing to proceed in the opening of a new Italian restaurant in Toronto.

### Data:

The data necessary for this Project is:

- The list of neighborhoods of Toronto
- The latitude and longitude of each neighborhood
- The list and location of existing Italian restaurants in Toronto

## Methodology:

The first step would be to obtain the list of neighborhoods of Toronto, Canada. This information can be extracted from the Wikipedia page:

“[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)”.

The web scraping will be done utilizing pandas html table scraping method, in order to obtain tabular data from the web page and create a dataframe.

After that, it will be necessary to add the coordinates of each neighborhood to the list of names and postal codes obtained from Wikipedia. This step will be done with the use of Foursquare.

The subsequent step is to visualize on a map each neighborhood, correctly positioned with the related coordinates, using Folium package.

Afterwards, Foursquare API will aid in obtaining the list of top 100 venues within a 500 meters radius, in order to focus the attention on Italian restaurants. With Foursquare I can obtain name, category, latitude and longitude of every venue.

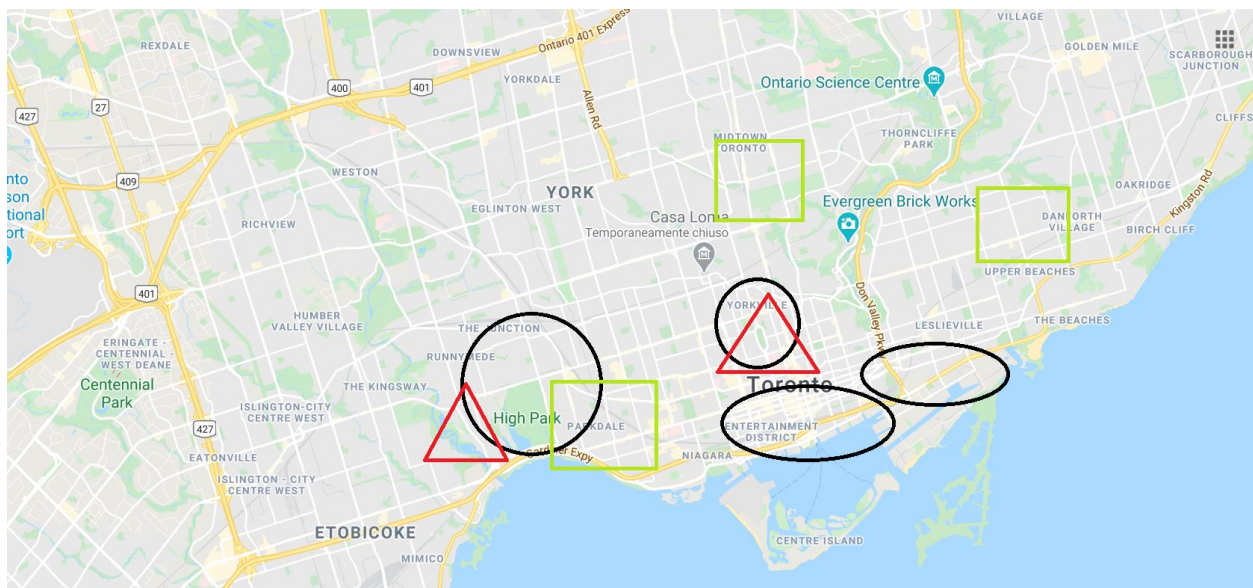
The analysis can then proceed by grouping by neighborhood, in order to do clustering later on.

Finally, the clustering is done using k-means, a method that identifies k number of centroids, and then allocated every data point to the nearest cluster, while keeping the centroids as small as possible.

Based on all the aforementioned steps, I will be able to suggest the most suitable location for a new Italian restaurant in Toronto, Canada.

## Results:

Here below a picture that shows the map of Toronto and the main areas where other Italian restaurants are located.



Through k-means clustering we divided Toronto into 5 clusters.

However, for practical reasons, in the map above, I only highlighted cluster 1,2 and 4, being the more relevant ones. In particular, the ovals in black mark cluster 1. The rectangles in green mark cluster 2. The triangles in red mark cluster 4. So, in order of relevance we have cluster 2, then cluster 4, and finally cluster 1. Cluster 0 had no Italian restaurants, and cluster 3 was the 4<sup>th</sup> more relevant.

### Recommendations:

As anticipated in the results, most of the Italian restaurants are located in cluster 2, then 4, 1, and finally cluster 3. The areas to focus on, in order of priority for the opening of a new Italian restaurant would then be around High Park, or the area of Yorkville / Central Bay Street.

### Limitations and Suggestions for Future Research:

It has to be considered that the main assumption of this project is that opening a new Italian restaurant in an area that has already some other Italian restaurants would be a good idea.

The second valuation has been to identify the area that has the highest concentration of Italian population: Etobicoke.

The final valuation has been to consider the average prices of real estate in different areas of Toronto.

For future research, it could be useful to consider areas that have a lower presence of Italian restaurants but might be still promising for the purpose of opening a new Italian restaurant.

### Conclusion:

I would suggest the entrepreneur to consider opening the new Italian restaurant either in the area of High Park, or in the area of Yorkville / Central Bay Street. In addition, by checking the real estate prices in Toronto, I can add that the closer we go to Etobicoke, the lower the prices of store would be. In that area we could therefore take advantage of both a high Italian population, as well as more affordable real estate. The decision to locate in this area could work for a less posh, "home-cooked" style of restaurant.

On the other hand, if the idea is to open a very chic, high-end restaurant, it could be worth it to explore a more central area, which could benefit from a population of higher-spenders, who probably work and/or live in that area.