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”.

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.