Business Case/ Introduction

The aim of the chapter is to define an objective for the analysis.
The kind of analysis that is performed is going to address a specific problem related to restaurant businesses.

The business case takes into consideration an audience formed by a group of investor that is keen to start a new restaurant business in Toronto.
The gathering of entrepreneurs is aiming to find the right spot in the city to open a new restaurant specialized in Italian cuisine.

I/our company has the role of advising the client regarding which spot would be more favorable for opening kind of activity of this type.

The analysis will have to take into consideration the current presence of similar businesses, in particular those under the classification of Italian Restaurant and Pizza places..
Although a higher number of restaurants is a good sign of the possible economic good response, if the number of Italian restaurants is too high that might be counter productive as the market may be tougher to penetrate.

This kind of analysis would be definitely helpful for our client to have a clear vision of which Neighborhoods share the same characteristics and could be possibly host their investment.

It is important to underline that the actual decision of which Neighborhood is going to be be the absolute best does not concern this investigation and it is strongly based on the clients criteria for the choice which could be viable and due to many circumstances.

Data

To complete the task, it is important to have enough information about the amount of restaurants of a certain area, their distribution and geographical persistency.

The Foursquare API outputs are fit for this purpouse, as these can be extracted based on Latitude and Longitude (Easting and Northing) of the various Neighborhoods composing Toronto’s urban texture.

The coordinates of the Boroughs have been extracted based on a Wikipedia page showing the Lat and Long for the various postal codes of the city. These data have been scrapped and thoroughly processed for our use.

The outputs from the Foursquare API have been extracted based on searches using the “explore” function of the framework, with the coordinates as an Input, allowing to gather data of the places relevant to our analysis.

The reasonable amount of data have been sorted and tabulated using various Py libraries and frameworks (Pandas, Numpy, Matlotlib, Folium).

The tool used to produce the final clustering is called Kmeans and it is part of the SciPy Library.

Methodology

The methodology used for the analysis included many different sorts of data analysis tecniques and tools.

Jupyter Notebooks have been used to prepare the code given their great flexibility.
Python has been choosen as a main language for analysis process.
The libraries used for the purpose included Pandas, Numpy, Matplotlib, JSON, SkLearn, Folium.

Foursquare API (FAPI from now on) whas been used as a support for the analysis

The complete code used for the purpouse is visible inside the Github repository attached with this file.

The first step has been to scrap the info out of the Wikipedia page related to Toronto's postal codes, reporting the coordinates for each on of these.
These coordinates have been later used to perform a series of queries through FAPI.

The main purpose of each query was to explore the area for each neighborhood to get an idea of the main local activities of the area.
Because of the scope of the analysis, we ere particularly keen to search for restaurant.

Once we had the data for each Neighborhood, these have been sorted and the table has been re-arranged to only show binary values for each activity associated to neighborhood.
To make it clearer, if a certain neighborhood had a cafè but did not have a pizza place, you would have a row for that hood containing a column for bth activities and a 1 as a value for cafè, a 0 as a value for pizza place.
This is done by simply applying the *Onehot* function to the original table containing the extracted values.

Once this was done, a second table has been created which only contained the columns for all the hoods in Toronto, "Italian Restaurant" and "Pizza Place".
This has been used to fit the Kmeans model, so that it would take into consideration those values to create the appropriate clusters.

The final clusters have been created and showed on the map generated via Folium.

Results

The results indicated a clear segmentation of the various clusters as indicated in the image below, representing the various neighborhoods which have been color coded according to the cluster these belong to.

Figura 1: Cluster Color Coded Neighborhoods

It is clearly noticeable a difference between the city centre neighborhoods which do have similar characteristics and the areas a bit more further away from the centre which seem to be more diverse.

Conclusion

The results seem to be reliable and can be used as an additional guide for a potential owner (or a group of them) to investigate which area of the city best suites their needs.
To make the analysis a bit wider it might be a good move to integrate a study to check which areas have an higher cost per square meter of land, so that by crossing the two results it might be possible to find a good balance between the reactivity of the area and its actual financial value.