Applied Data Science Capstone

5th week (and last) assignment.

Introduction/description

This project aims to help while opening a new restaurant in Terrassa, my city. It will do so by two means:

- Firstly, showing the distribution of restaurants in Terrassa
- Also, showing the common restaurant categories by neighborhood

This two informations can be useful to minimize concurrence of other restaurants geographically and and to choose the theme of the restaurant; adding a new one to an upcoming trend or opening a new kind of restaurant in a neighborhood.

Data required

To achieve the goals described in the introduction, some datasets are required, mainly:

- Geographic data of Terrasa i.e its location, its neighborhoods and boroughs
- Latitude and longitude of the restaurants of Terrassa
- The category of the restaurants, i.e italian food, sushi place, etc.

Folium will be used to generate maps. Nominatim will be used to fetch the location of Terrassa in the map and Foursquare will yield the venues in the desired neighborhoods.

The venues have to be filtered to the ones containing "Restaurant" in their names. Kmeans will cluster the neighborhoods following their most common restaurants.

Methodology

Once all the required data is avaiable in a dataframe format the analysis can be done.
The methodology is focused on clustering the different neighborhoods of Terrassa. The first step has been to place every neighborhood on the map.
After that, I looked for all the venues available in the city and I represented them in another map, coding their kind of cuisine by color.

Also, I wanted to cross information with the wealth of each neighborhood based on the average value/square meter (sqm) of the flats to sell/rent but I could not get access to the required API (I should have tried the analysis straighforward with NY city but I run out of Watson Studio hours by then).

The next step is the classification of neighborhoods with Kmeans using their “cuisines” and their number of restaurants. It can offer different views to the interested person and give advice of where and which kind of restaurant could be profitable in each case.

This project is rather visual, relying almost completely on maps. Although it is simple, is very appealing and comprehensible in my opinion.

Results

The analysis show that there are 93 restaurants available in Terrassa. From those, there are 15 unique categories. The restaurants are displayed on a map, with defined colored labels showing the restaurant category (i.e. mediterranean food in green, italian food in blue etc).

Further analysis showed which ones are the 3 most common venues in each neighborhgood. With this dataframe I used Kmeans to cluster the different neighborghoods. These results help the decision maker to choose where to open a new restaurant and which kind of cuisine should it be. Basic questions that this model could answer are:

1. Is the neighborhood overcrowded with restaurants? Is the area missing restaurants?
2. Which are the most common restaurants in the neighborhood? Do I want to bring a new trend, or sum up to the existing one?
3. Is a specific cuisine missing in a neighborhood from a specific cluster?

Discussion

The first two questions are quite obvious. In my opinion a good insight can be taken from the 3rd question. In this case, in specific clusters containing many neighborhoods, it is normal that the neighborhoods share the most famous restaurants among them in the cluster. However, there is a high probability that some neighborhoods share only 1 or 2 of the most common restaurants. Therefore, a good oportunity can be to open a new restaurant with another cuisine common in the same cluster but not popular in an specific neighborhood.

Focusing the results, It will not surprise anybody that one of the most common kinds of restaurants are the mediterranean and Spanish cuisine. However, I was suprised that many neighborhoods got between the 3 first options vegetarian or vegan options which are certainly a new trend in the city.

I also would like to mention some flaws I realized while coding this notebook.
Firstly, as Terrassa is a smaller city compared to NY or Toronto, the size of data is quite smaller and the analysis is accordingly less robust.
In addition, the “venue category” field in Foursquare might be tricky since I cannot know how are all the restaurants classified in the data base. In the notebook I filter all the venues using the “restaurant” string. However some entries are named “pizza place” which could be also included in the analysis but in this case are ignored.
Last but not least, the “luxury/quality” of the restaurant is also ignored in this analysis. Comparing the price of the sqm between neighborhoods can give insight about how expensive the area is and therefore how expensive the resturants can be. The quality, which is

independent of how expensive is the restaurant, is in my opinion the harder parameter to quantify.

Conclusion

The notebook can help decision makers visualizing the different restaurant options of Terrassa checking their place and style. It offers insight to decide to jump into a positive trendline of new restaurants in the area, or if preferred, opening a traditional restaurant in a place without concurrence.