Capstone Project
The Battle of Neighbourhoods

Introduction

The **business problem** that is addressed in this project is the understanding of the city of Toronto and in particular the relationship between the type of venues and the housing prices in Toronto.

Is there a relationship between the price paid for a house in a neighbourhood and the existing restaurants in that same neighbourhood?

Data section

We will use the website https://fr.foursquare.com/ and the API in order to extract the venues in Toronto as previously done for the clustering of Toronto's neighbourhoods.

We will use the website https://www.kaggle.com/ for the dataset of _'House Sales in Ontario'_ that one can find at the below URL.

https://www.kaggle.com/mnabaee/ontarioproperties?select=properties.csv

All the latitude and longitude of Toronto's neighbourhood are available in the file shared by IBM lab at the following address:

Recap of data used:
- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- https://cocl.us/Geospatial_data
- https://fr.foursquare.com/
- https://www.kaggle.com/mnabaee/ontarioproperties?select=properties.csv

Methodology

- **Collecting** the data (Cf. Data Section)
- **Exploring** the data to understand what their content is
- **Clean** the data
    - e.g. drop empty rows and/or columns
    - filter the list of addresses so that our analysis focuses only to the city of Toronto and not the full state of Ontario

- group the data by neighbourhood
- **Process** the data and in particular
  - Extract the list of most visited venues by neighbourhood
  - Get the average price paid by neighbourhood
  - Propose a clustering of venues and housing prices
- **Analyse** the data:
  - Compare the map of venues clustered with the housing prices clustered

Results

Following the week 3 and 4 of the Capstone project we studied the city of Toronto and segmented and clustered the neighbourhoods.

The second part of the notebook tried to understand what can possibly be a segmentation and clustering of the housing prices in the city of Toronto.

We identified 3 clusters of which one represents over 4,704 houses out of a total sample of more than 5,000 houses.

Discussion

Probably the reason why this study is not of particular interest is that we should rank the type of venues and try to distinguish if the quality of the venues may be correlated in some way with the price of the houses in the neighbourhood.

Also, we should probably determine another factor that is the proximity to the venues. This may reveal that more expansive houses are independent homes while less expensive houses are in buildings and that buildings are in the heart of Toronto where the density of venue is.

Conclusion

From these data I don't think there is any conclusion to draw in particular. One can't say that one venue in particular characterise a neighbourhood or that the price of a house might be related in some way with the type of venues close by.