Segmentation tool for neighbourhoods

By: Belal Ibrahim

The problem

The aim of this project is to cluster similar neighbourhoods in the city of toronto together so that people can easily choose where to live based on their interests, lifestyle, and the data of each neighbourhood.
This tool can also be used by real estate agents to find the suitable location for the customer based on their interests and lifestyle.
This project clusters neighbourhoods according to the number venues, like: parks, museums, restaurants, and cafes, each one of them has.

Data

Data required:

- List of neighbourhoods in Toronto
- Locations of neighbourhoods
- Venue data of each neighbourhood

Sources of data:

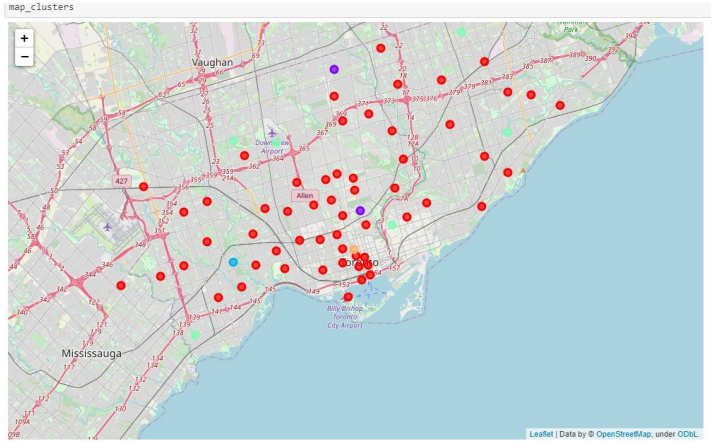
- Foursquare API
- Wikipedia
- This online data set "https://cocl.us/Geospatial_data"

Mythodology

- Web scraping using the pandas module
- Cleaning the data using the pandas module
- Getting the venues data using the Foursquare API
- Clustering the neighbourhoods together using the Kmeans algorithm
- Representing the data on the map using the Folium module

Results

- A dataframe with each neighbourhood's data like cluster number and top venues
- A map with neighbourhoods superimposed on it



Discussion

- Most neighbourhoods are in cluster number 0
- Most neighbourhoods are very similar in the number and type of venues they have

Conclusion:

In conclusion, this tool can be very useful to real estate agents in recommending neighbourhoods to customers or even individuals who want to find a suitable neighbourhood to live in.