I am moving out, what's next?

Introduction

The aim of this project is to get a tool to find neighbourhoods with specific characteristics. When people move on to a different city there usually are some essencial venues and services they want on the new neighbourhood. For example, some people may want to have a train station, or scholls near their new home or, maybe, the only want to find a neighbourhood with similar characteristics than they actual neighbourhood. In this project we will try the Foursqare Api to explore the city of Toronto in order to find the more situable neigbourhood for someone moving from a specific place.

Data

The data used to investigate Toronto neighbourhood is extracted from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. This data was extracted from the webpage by using BeautifulSoup library in python.

The data extracted from the webpage consist on the the postal code, the borough and the neighbourhood. The 5 first values are presented below:

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

First of all we delete the not assigned neighbourhood in order to clean the data.

Once the data is clean. We need to extract the longitude and latitude of each neighbourhood in order to be able to extract the data form the Foursquare API. This information is obtained by using the Geocoder library in python and the final data obtained is presented below:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.75188	-79.33036
1	M4A	North York	Victoria Village	43.73042	-79.31282
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.65514	-79.36265
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.72321	-79.45141
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.66449	-79.39302

Using longitude and latitude data in the Foursquare API, the information of the venues in each neighbourhood in Toronto is extracted:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Parkwoods	43.75188	-79.33036	Brookbanks Park	43.751976	-79.332140	Park
Parkwoods	43.75188	-79.33036	PetSmart	43.748639	-79.333488	Pet Store
Parkwoods	43.75188	-79.33036	Brookbanks Pool	43.751389	-79.332184	Pool
Parkwoods	43.75188	-79.33036	Variety Store	43.751974	-79.333114	Food & Drink Shop
Parkwoods	43.75188	-79.33036	The Bing Suites	43.747816	-79.332190	Bed & Breakfast

As we are only interested in the venue category per neighbourhood we select these categories from the total data.

Venue Category	Accessories Store	Airport	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	...
Neighborhood											
Agincourt	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Alderwood, Long Branch	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	...
Bathurst Manor, Wilson Heights, Downsview North	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Bayview Village	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Bedford Park, Lawrence Manor East	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...

Methodology

To find similar neighbourhoods to an original one we first need to stablish the initial conditions. In this case, we assume that the user wants to move from the Riverdale neighbourhood in New York to another with similar characteristics in Toronto.

Using the FourSquare API we find the venues in the Riverdale neigborhood in New York:

1]:

categories	name
Bank	1
Baseball Field	1
Bus Station	2
Food Truck	1
Gym	1
Park	2
Plaza	1

We select the same venues in the Toronto neighbourhoods list:

Venue Category	Bank	Baseball Field	Bus Station	Food Truck	Gym	Park	Plaza
Neighborhood							
Agincourt	NaN	NaN	NaN	NaN	NaN	1.0	NaN
Alderwood, Long Branch	NaN	NaN	NaN	NaN	1.0	NaN	NaN
Bathurst Manor, Wilson Heights, Downsview North	NaN	NaN	NaN	NaN	NaN	1.0	NaN
Bayview Village	NaN	NaN	NaN	NaN	NaN	1.0	NaN
Bedford Park, Lawrence Manor East	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...
Willowdale, Willowdale West	1.0	NaN	NaN	NaN	NaN	1.0	NaN
Woburn	NaN	NaN	NaN	NaN	NaN	1.0	NaN
Woodbine Heights	NaN	NaN	NaN	NaN	NaN	NaN	NaN
York Mills West	NaN	NaN	NaN	NaN	NaN	1.0	NaN
York Mills, Silver Hills	NaN	NaN	NaN	NaN	NaN	NaN	NaN

We get a venue array for every of the neighbourhood in the list we apply the Euclidian distance computation with the original venues array from Riverdale. The final distance value is added to the data table:

categories	Bank	Baseball Field	Bus Station	Food Truck	Gym	Park	Plaza	Distance
Neighborhood								
Agincourt	0.0	0.0	0.0	0.0	0.0	1.0	0.0	3.162278
Alderwood, Long Branch	0.0	0.0	0.0	0.0	1.0	0.0	0.0	3.464102
Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.0	0.0	0.0	1.0	0.0	3.162278
Bayview Village	0.0	0.0	0.0	0.0	0.0	1.0	0.0	3.162278
Bedford Park, Lawrence Manor East	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.605551
...	...	...	...	...	...	...	...	...
Willowdale, Willowdale West	1.0	0.0	0.0	0.0	0.0	1.0	0.0	3.000000
Woburn	0.0	0.0	0.0	0.0	0.0	1.0	0.0	3.162278
Woodbine Heights	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.605551
York Mills West	0.0	0.0	0.0	0.0	0.0	1.0	0.0	3.162278
York Mills, Silver Hills	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.605551

Results

Once the distance between the original neighbourhood and the Toronto neighbourhoods is computed we can obtain the neighbourhood with more similar venues than the original one. These will be those neighbourhood showing lower distance. This neighbourhoods are presented below:

categories	Bank	Baseball Field	Bus Station	Food Truck	Gym	Park	Plaza
Neighborhood							
Harbourfront West, Bathurst Quay, South Niagara, Island airport	1.0	0.0	0.0	0.0	2.0	2.0	0.0
Central Bay Street	1.0	0.0	0.0	0.0	1.0	1.0	2.0
Dufferin, Dovercourt Village	1.0	0.0	0.0	0.0	0.0	2.0	0.0
India Bazaar, The Beaches West	0.0	0.0	0.0	0.0	1.0	2.0	0.0
St. James Town	0.0	0.0	0.0	1.0	2.0	2.0	0.0
St. James Town, Cabbagetown	1.0	0.0	0.0	0.0	0.0	2.0	0.0

From the list the user can select the more suitable neighbourhood to live. he user can select the neighbourhood from the list above taking into account their personal

preferences. For example, analysing the venues in each neighbourhood it seems that the more suitable neighbourhood should be Central Bay Street as it is the more complete neighbourhood from the list.

The user can refine the search by selecting a specific characteristic in the list. For example, none of the above neighbourhood have a bus station near. If the user considers this venue, essential can repeat the search by forcing to have some value in the Bus station column.

Conclusions

In this project a tool for search similar neighbourhoods in different cities is presented. The project uses the Foursquare API and the Euclidian distance to select the neighbourhoods with more similar venues. This allows to the user to explore the new city and to search for different options taking into account his personal preferences.