Assignment: The Battle of neighborhoods

Finding the right neighborhood in a new city

Introduction

When moving to a new city, for professional or personal reason, it can be quite hard to find a new neighborhood to settle in. In certain circumstances, it might be that you enjoy the neighborhood you are currently living in, but you have to move for professional reasons. Alternatively, you already found a nice neighborhood through recommendation but unfortunately, no housing options are available in that neighborhood at that moment. In these circumstances, it might be beneficial to find similar neighborhoods similar to the one you currently living in or you are interested in.

Business Problem

As indicated in the introduction, several scenarios can be thought of where it can be interesting to find neighbourhoods, in a certain city/location to a neighbourhood/location of choice. This problem can be tackled by analysing the distinct features attributed to a certain region, and compare these features with features from neighbourhoods in a target region of choice. A Content based recommendation algorithm would then be able to recommend similar neighbourhood. Such a setup could prove to be useful for real estate agents or big cities investing in attracting new citizens to help choosing a suitable neighbourhood. For this specific case we will compare my neighbourhood 'Lange Munte, 8500 Kortrijk, Belgium' with neighbourhoods in New York to recommend me similar neighbourhoods. The top 5 recommended neighbourhoods will be show in a map using Folium.

Data

Data from Foursquare will be used to solve the business problem. Venues relevant for the input and target neighbourhoods will be retrieved in order to build the recommender system. Coordinates for the 'Lange Munte' neighbourhood where I live are retrieved from Wikipedia (
50°48'40"NB, 3°17'56"OL, https://nl.wikipedia.org/wiki/Lange_Munte_(Kortrijk)).
New York has 306 neighbourhoods, latitude and longitude coordinate of each neighbourhood will be retrieved from https://geo.nyu.edu/catalog/nyu_2451_34572 and used to query Foursquare. As explained, venues from neighbourhoods will be retrieved from Foursquare and used to build a content based recommendation system as an example (for example, the proximity of schools, restaurants, parks can be an important feature). This application here could be extended to work with any region of choice.

Data Processing

Data containing neighborhood information for New York was downloaded in JSON format from https://cocl.us/new_york_dataset. Next, for each neighborhood 3 the neighborhood name and coordinates were extracted. Subsequently, for each neighbourhood, Foursquare was queried to obtain all venues (with a maximum of 250 neighborhoods) in a radius of 2000 meters from the neighborhoods center. For this project, the coordinates for the neighborhood "Lange Munte" in Kortrijk, Belgium was manually extracted from Wikipedia and all venues were extracted for this neighborhood in a similar fashion (as shown in Figure 1).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Lange Munte	50.811111	3.298889	Fitness Lange Munte	50.806963	3.301564	Gym / Fitness Center
1	Lange Munte	50.811111	3.298889	Sport-K-Fee	50.806926	3.301754	Sports Bar
2	Lange Munte	50.811111	3.298889	Sportcampus Lange Munte	50.806627	3.302055	Athletics & Sports
3	Lange Munte	50.811111	3.298889	Cantor	50.803238	3.293851	Bar
4	Lange Munte	50.811111	3.298889	De Vlasblomme	50.805269	3.288848	Bistro
...	...	...	...	...	...	...	...
90	Lange Munte	50.811111	3.298889	Halte Beneluxpark	50.799539	3.279427	Bus Stop
91	Lange Munte	50.811111	3.298889	Garage Vanderheren	50.795271	3.288437	Auto Garage
92	Lange Munte	50.811111	3.298889	Notaris Declercq	50.813584	3.325812	Notary
93	Lange Munte	50.811111	3.298889	DLPA	Advocaten	50.828716	3.302520	Lawyer
94	Lange Munte	50.811111	3.298889	stortbeton bosschaert	50.795125	3.286513	Construction & Landscaping

Figure 1: venues retrieved from Foursquare for neighborhood 'Kortrijk'

Next, all venues were encoded in a binairy fashion (though onehot encoding) for compatibility with the prediction algorithm. Also, the dataset was transformed to a binairy data set, as it was reasoned that the proximity of one or 3 parks was irrelevant and could induce bias in the dataset. The resulting dataset is represented in the Figure 2.

	Neighborhood	Zoo Exhibit	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Weight Loss Center	Whisky Bar	Windmill	Wine Bar	Wine Shop
0	Allerton	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	Annadale	0	0	0	0	0	0	0	0	1	...	0	0	0	0	0
2	Arden Heights	0	0	0	0	0	0	0	0	1	...	0	0	0	0	1
3	Arlington	0	0	0	0	0	0	0	0	1	...	0	0	0	0	0
4	Arrochar	0	0	0	0	0	0	0	0	1	...	0	0	0	0	0

Figure 2. Dataframe after onehot encoding

Results

A simple content based recommendation system was build using dot multiplication of the venue matrix for all neighborhoods in New York and the "Lange Munte" venue feature set. The top 5 neighborhoods similar to "Lange Munte" are: Little Neck, Fresh Meadows, East Village, Sunnyside and Chinatown (Figure 3)

149	Neighborhood
107	24	Fresh Meadows
157	24	Little Neck
81	23	East Village
265	23	Sunnyside
50	23	Chinatown

Figure 3: Top 5 neighborhoods resembling "Lange Munte"

When looking at a detailed description of the venues in the top5 neighborhoods there is indeed a high resemblance with the venues in "Lange Munte". In addition, the top 5 neighborhoods are visualized in a geographical map in Figure 4.
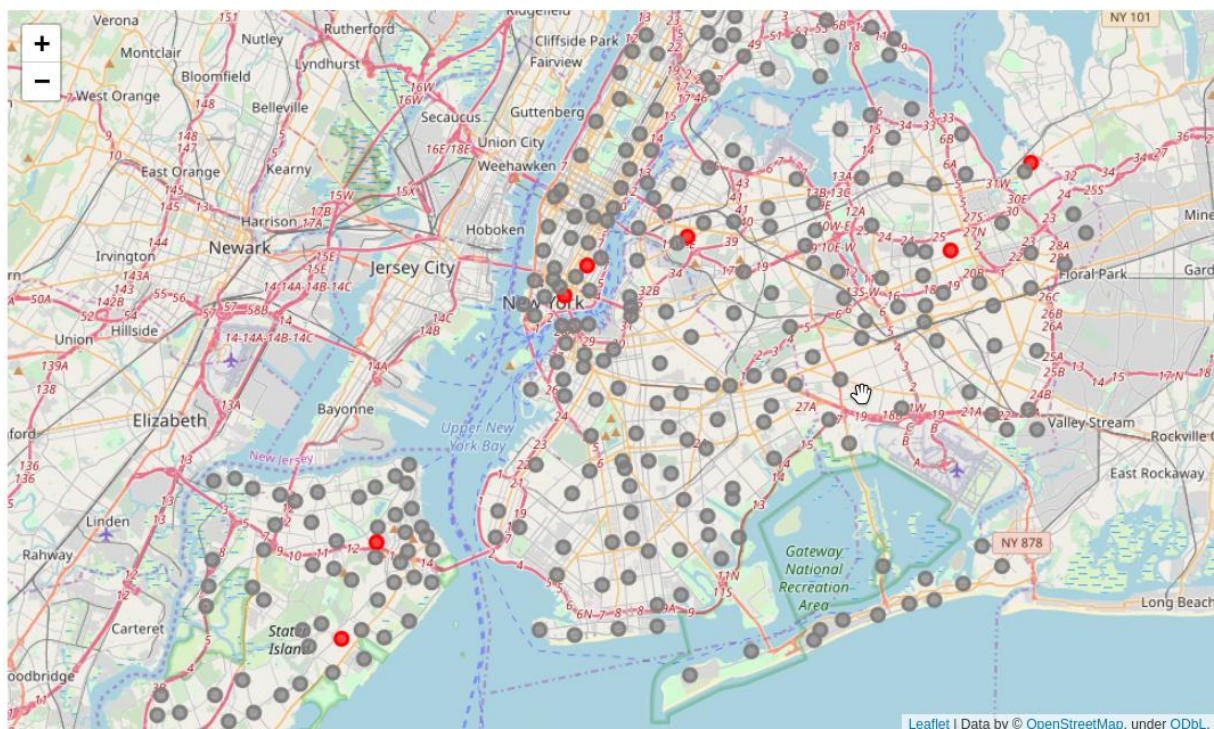

Figure 4: Geographical representation of New York, all neighborhoods centers are highlighted on the map with dots, the top 5 neighborhoods resembling the "Lange Munte" neighborhood are highlighted with red dots.

Conclusion and future persepectives

In this project I've successfully managed to extract neighborhood data for New York. Next, this neighborhood data was used to extract up to 250 most popular venues for each neighborhood in a radius of 2000m, and this data was transformed to feed a recommendation system. As an example, the neighborhood "Lange Munte", from Kortrijk in Belgium was used to find similar neighborhoods in New York. This application can be used by real estate agents, governments or as an application to identify similar neighborhoods for users/costumers. For example, real estate website can use this tool to pre-select neighborhoods for their costumers to look for new housing opportunities.

However, many improvements are possible to improve performance of the made recommendation. Users could input their own preferences (for example additional importance on parks in near proximity) or input some requirements or restrains. Furthermore, such an application could benefit from learning from other users, by storing preferences of other users and use their data to improve recommendation for users in a similar situation. Moreover, this project was limited to the input of one neighborhood, but could be extended to accept multiple neighborhoods as input (for example in Kortrijk, I enjoy the "Lange Munte" and "Morinnehoed" neighborhood.