Coursera Capstone project report

Alina Kleimenova

January 2021

1 Introduction

Moving to another city is very exciting but also very stressful. Searching for a new place to stay for at least a year can be a nightmare, especially in a completely new city. Sometimes so many things need to be taken into account: rental prices, bus/train/metro stops, shops, schools or kindergartens or both nearby and so on. Thus, it would be very helpful to reduce the area of search from the whole city to a group of neighbourhoods matching user's preferences. And to do so we could use our current or past experience.

The aim of this capstone project is to find a neighbourhood (or group of neighbourhoods) in another city based on the location given by the user. Therefore, the question we are going to answer in the end of this project is: in which neighbourhood I should search for an apartment to rent/to buy based on my current preferred area of staying?

This project can be of particular interest for people who change their workplace frequently, such as, for example, people working in academia.

2 Data

For this project we will need the location of the current user address and location of the neighbourhoods in the new city.

Geneva boroughs in GeoJson format are taken from `Système d'information du territoire à Genève (SITG)` [1] database with the help of `Esri ArcGIS Rest API`. Center of each borough is computed using `Shapely Python package` [2] for computational geometry.

Current user address chosen arbitrary and set to *Chaussee de Wavre 442, Brussels, Belgium*, its latitude and longitude data is extracted using `geopy` client.

To compare user's current address with the neighbourhoods in the new city, we used `Foursquare API` [3]. From this API we requested following information for each neighbourhood using its geo-position:

1. Venue
2. Name of the venue

3. Venue Latitude
4. Venue Longitude
5. Venue Category

We took into consideration all types of venues: shops, restaurants, cafe, schools, parks and so on.

3 Methodology

As a first step of this project, we created two tables: one table with home address, its latitude and longitude, and another table with the list of boroughs of Geneva along with their geo-locations (latitude and longitude).

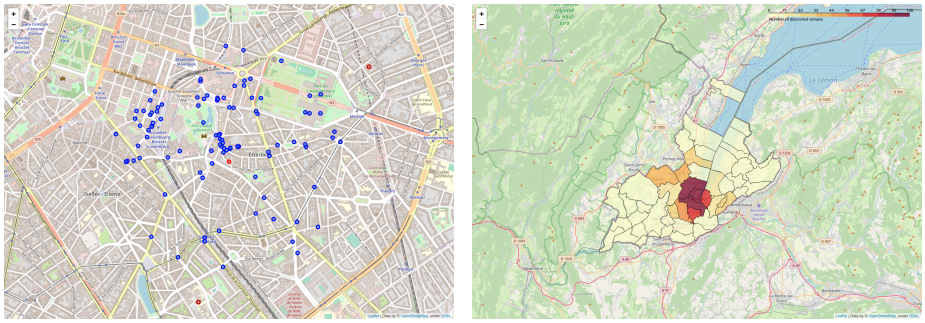

Figure 1: Left: Venues around address in Brussels (red dot); Right: Boroughs on Geneva, colour scale indicates the number of discovered venues.

We then use `Foursquare API` and for each table each location we request all venues in a radius of 1 km from neighbourhood center. We put a limit on number of returned venues to 100. The found venues for both locations are displayed on the map with help of the `Folium` library (Fig. 1).

To proceed with the analysis, we segment home neighbourhood and new city neighbourhoods via categories of venues around their location. We make sure that venue categories are matching between home address data and neighbourhoods. After doing so we have 42 features out of 177 total features in Geneva and 62 total venues in Brussels.

In order to identify similar neighbourhoods in Geneva we used `k-means` clustering algorithm. This approach will allow us demonstrate the user all possible similar areas to the reference address. To define k we used *elbow method* [4]. The best k was found to be equal to 6 (Fig. 2 left). The result of clustering is displayed on the choropleth map (Fig. 2 right). The obtained clusters can be described in the following way:

- **Cluster 0** – Calm boroughs with lots of diner places

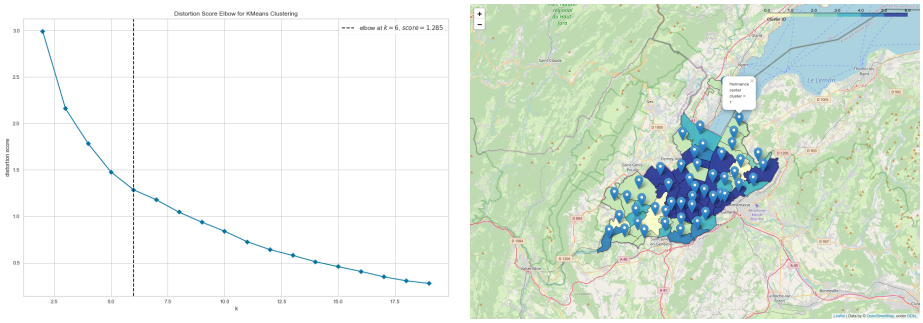

Figure 2: Left: Output of the elbow method performed with help of `yellowbrick` [5]. Right: Output map of Geneva boroughs, colour code here indicates cluster id.Each marker on the map is placed in the center of corresponding borough and contains the name of the borough and its cluster id.

- **Cluster 1** – Calm boroughs with lots of french restaurants
- **Cluster 2** – Calm touristic borough
- **Cluster 3** – Borough with lots of places to go for dinner
- **Cluster 4** – Calm borough for families with kids (lots of playgrounds, few venues)
- **Cluster 5** – 'Central' touristic boroughs (includes areas with the city center and near lake areas)

Finally, we compared centroids of obtained clusters with the feature vector of the home city using Euclidean distance. The cluster with the minimal distance is shown on the output map. For each borough we show top 3 most frequent places.

4 Results

The output map is shown on Fig. 3. As we can see the best matching cluster was found to be **Cluster 5** – central 'touristic'-like areas with a lot restaurants and places to visit. The tables for both, Brussels and selected Geneva cluster, are given in the end of the report.

5 Discussion

In this simple study we took into consideration only the surrounding venues, thus, to be more complete we would suggest to add rental prices, income and crime rates in the neighbourhoods of the interest. Moreover, would be beneficial

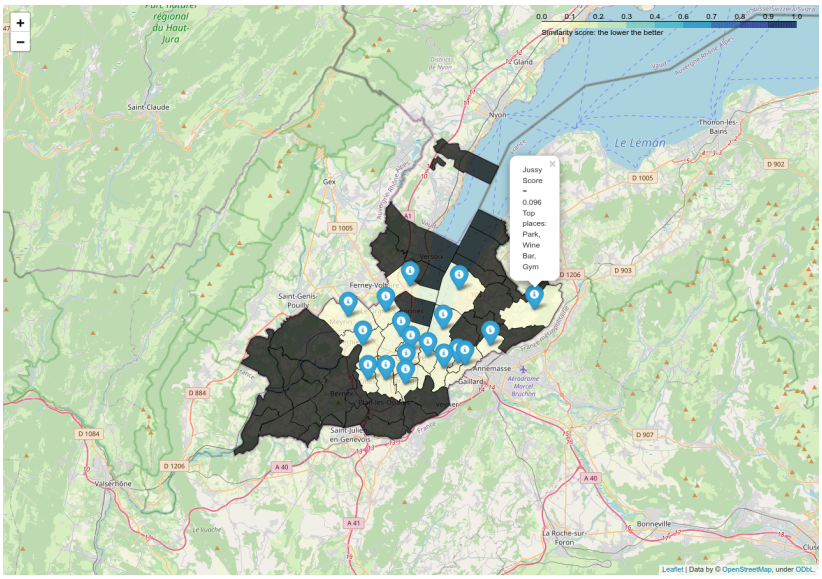

Figure 3: Output map with highlighted suggested cluster. Markers contain name of the borough, similarity score and top 3 most frequent venues.

to consider maximum distance from a possible place of work or study. Additionally, a feature filtering can be used to enhance the most preferred venues around.

Studying the boroughs of Geneva, we saw that some borough centers were very close to each other in comparison with the search radius. This caused overlapping venues in some neighbourhoods. The problem could be solved by segmenting the boroughs and reducing the search radius. The latter could add more precision to the final result, since we can improve number of discovered venues in larger boroughs.

6 Conclusion

The aim of the project was in finding best location to move based on user inputs. Given the home address, Chaussee de Wavre 442, Brussels, Belgium, and the city of moving, Geneva, we found a group of boroughs, where the user can move based on venues around. For this we used `k-means` clustering and Euclidean distance measure. As an result of the research we provide a map with recommended boroughs and an information about top 3 most frequent venues in each borough.

The final choice, however, must include many other factors, such as rental prices, proximity to work and so on, thus the study can be considered as a first step towards a final decision.

This project can be improved in many possible ways, for instance, more data about boroughs can be considered. Potentially this project could be turned into an interactive application which can be helpful for people finding a new home in a new city.

References

[1] Système d'information du territoire à Genève (SITG) au 07/02/2021 URL: ge.ch/sitg/sitg_catalog/geodataid/5661

[2] Shapely package URL: shapely.readthedocs.io

[3] Foursquare Developers URL: developer.foursquare.com

[4] Elbow method on Wikipedia URL: en.wikipedia.org/wiki/Elbow_method_(clustering)

[5] Yellowbrick: Machine Learning Visualization URL: scikit-yb.org

Borough	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Belgium, Brussels, chaussee de wavre 442	50.835607	4.382307	Italian Restaurant	Bakery	Coffee Shop

Table 1: Top 3 features assigned to the address in Brussels

Borough	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Bellevue	46.254665	6.145282	Gym	Restaurant	Italian Restaurant
Carouge	46.181673	6.140352	French Restaurant	Italian Restaurant	Bar
Chêne-Bougeries	46.193563	6.181689	Bakery	Restaurant	Plaza
Chêne-Bourg	46.197378	6.197007	Supermarket	Pizza Place	Restaurant
Collonge-Bellerive	46.252159	6.198609	Italian Restaurant	Wine Bar	Lounge
Cologny	46.222820	6.180841	Bakery	Italian Restaurant	Diner
Genève-Cité	46.206755	6.146280	Hotel	French Restaurant	Plaza
Genève-Eaux-Vives	46.202380	6.164940	French Restaurant	Italian Restaurant	Steakhouse
Genève-Petit-Saconnex	46.217463	6.135547	Hotel	Park	Italian Restaurant
Genève-Plainpalais	46.193499	6.141230	Bar	French Restaurant	Italian Restaurant
Grand-Saconnex	46.235769	6.119188	Hotel	Bar	Supermarket
Jussy	46.237120	6.279120	Park	Wine Bar	Gym
Lancy	46.184896	6.119608	Supermarket	Restaurant	French Restaurant
Meyrin	46.232229	6.079096	Italian Restaurant	Pizza Place	Supermarket
Onex	46.184855	6.100500	Italian Restaurant	Supermarket	Park
Puplinge	46.210838	6.232174	Wine Bar	Lounge	Italian Restaurant
Thônex	46.196086	6.204595	Supermarket	Pizza Place	Restaurant
Vernier	46.210735	6.094633	French Restaurant	Pizza Place	Supermarket

Table 2: Final cluster of Geneva boroughs found to be compatible with the address in Brussels