

# Coursera Capstone Project

IBM Data Science Professional Certificate

---

## Finding a place to move

by Alina Kleimenova

# Business problem

Moving **IS** stressful, but what if we could take not just our belongings but also **a nice neighbourhood** with us?

In this project we will find and **select the best location** to move in a new city **based on user preferences**



# Who can be our client?

- ✓ Professionals who move because of the new job
- ✓ People who just want to move to a particular city
- ✓ Students who move for studying

# Data

In this study we will help our imaginary user to move from Brussels to Geneva.

Our **input** will be **the current address** (*Chaussee de Wavre 442, Brussels, Belgium*) and **city where the user plans to move** - Geneva.

**Geolocation** of the current address will be extracted with the help of **geopy** package.

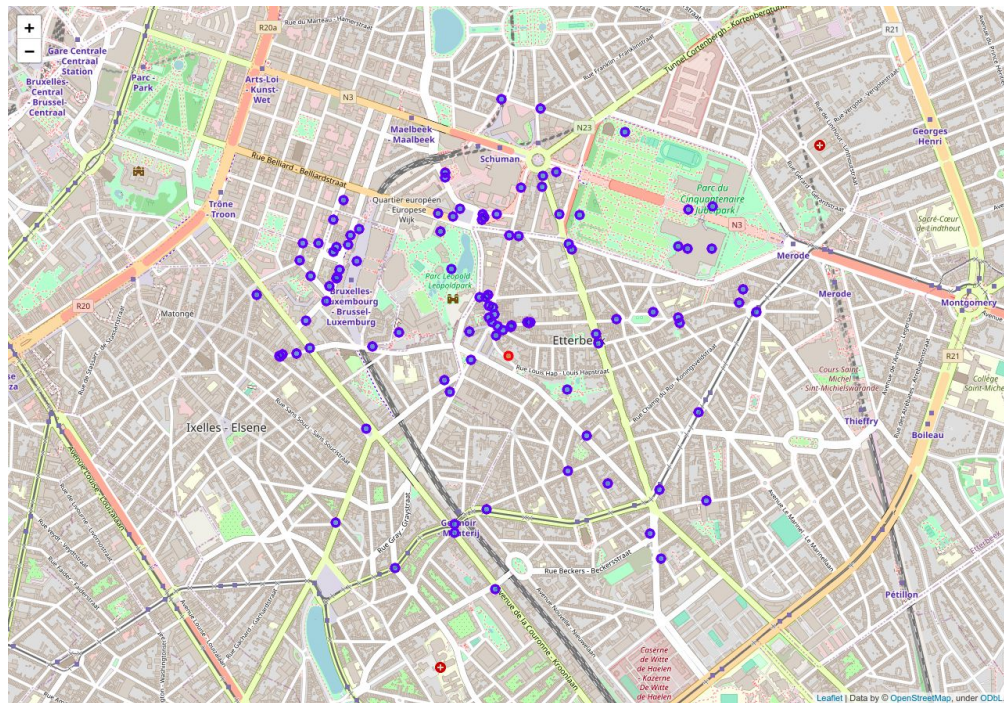
**Data** about boroughs of Geneva (communes) we will take from the [OpenData Swiss](#) website.

**Venues** in the all locations will be extracted using **Foursquare API**.

# Methodology

- Get all necessary **geo data** for both locations
- Extract all **venues** in radius of 1km around the input address and neighbourhood centers using **Foursquare API**
- **Visualise** both locations and their venues using **Folium**
- Find all overlapping venues in the both locations
- Cluster boroughs of Geneva using **k-means clustering** to find similar places
  - Find **best k** using **elbow method**
- Compute **Euclidean distance** between cluster centers and the address in Brussel
- Plot **final map** of Geneva **rated by similarity** to the reference address

# Brussels and venues around



Red dot in the middle is the input address. Blue dots are venues

Top 10 venues around  
(by frequency)

	Venue Category	Category count
30	Italian Restaurant	8
3	Bakery	4
9	Coffee Shop	4
26	Hotel	3
44	Sandwich Place	3
8	Cocktail Bar	3
17	French Restaurant	3
40	Plaza	3
25	History Museum	3
23	Gym / Fitness Center	2

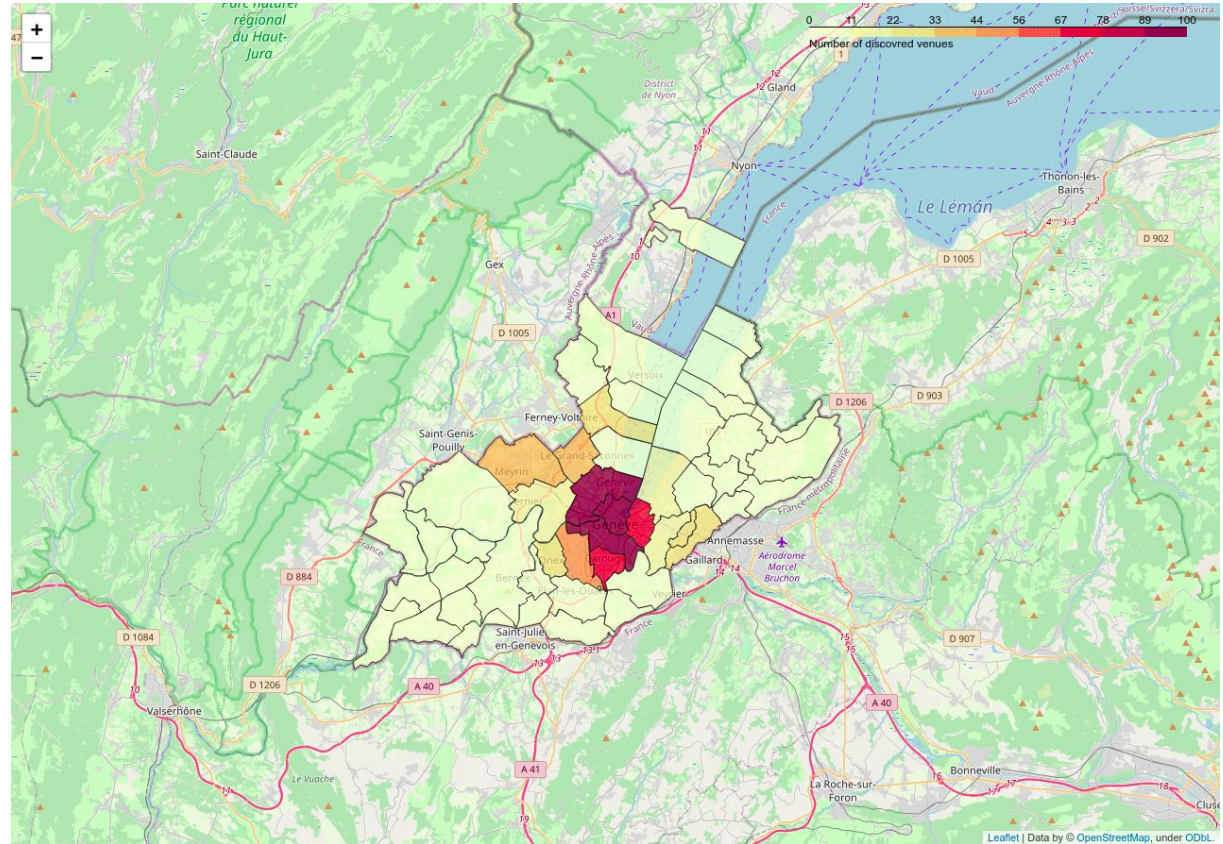
Nice area with a lot of places to visit  
and to go out!



# Geneva and venues around

**Colour code** on the map on the right represents **a total number of venues** retrieved from **Foursquare** in 1 km radius around the borough center. Borough borders are shown in lines.

As we can see, the largest number of venues we can find is near the center of Geneva. It is worth mentioning that the number of venues in the larger boroughs could be improved by segmenting them.



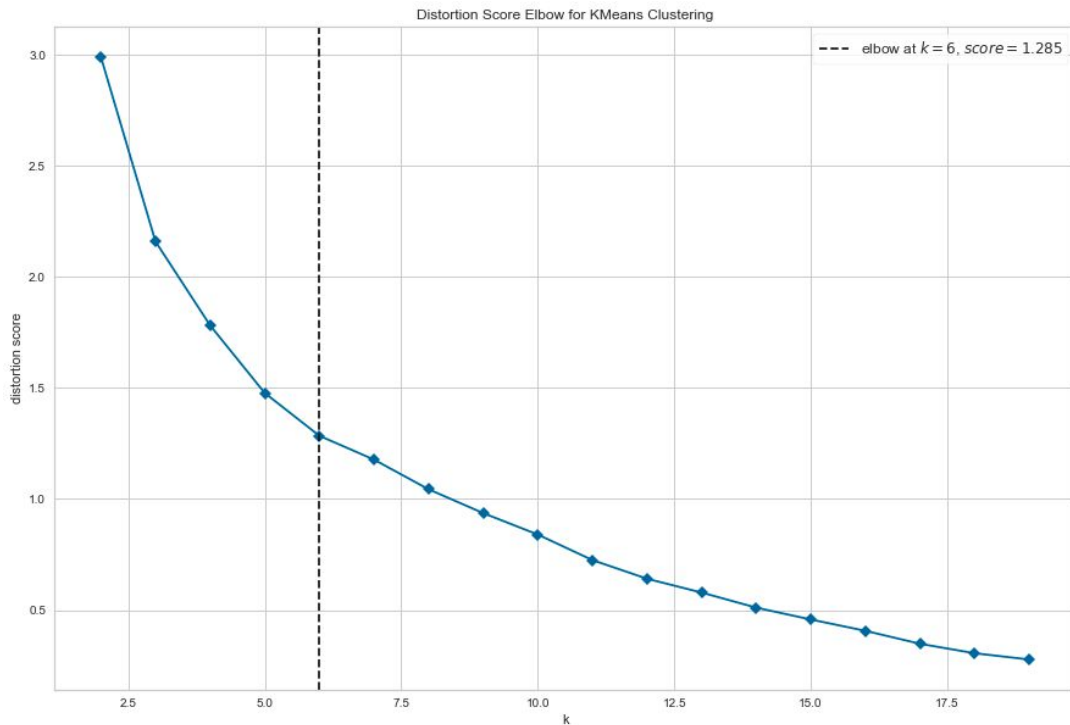
# k-Means clustering

After selecting common features between Brussels and Geneva we are left with **42 venues** out of 177 in Geneva and 62 in Brussels.

Now we can proceed with clustering Geneva boroughs.

To select optimal value for **k** in k-Means we will first run **elbow method**. In this method we will evaluate the distortion of clusters for a range on k from 2 to 20. The scan was done with the help of **yellowbrick** package.

The optimal value of **k** was found to be **6**

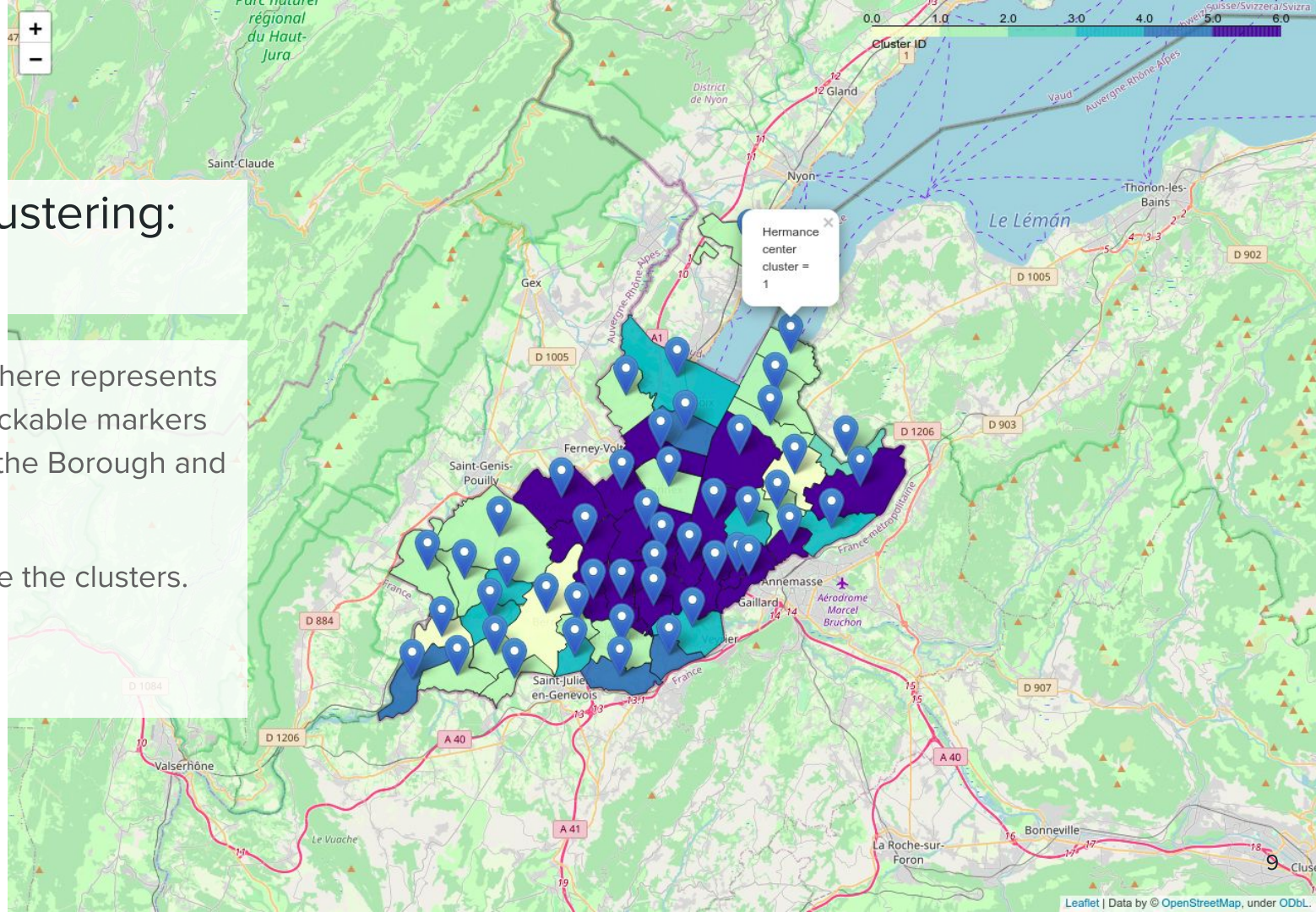




## k-Means clustering: result

The colour code here represents cluster id. The clickable markers contain name of the Borough and the cluster id.

Let's see what are the clusters.



# k-Means clustering: result

**Cluster 0:** Calm boroughs with lots of diner places

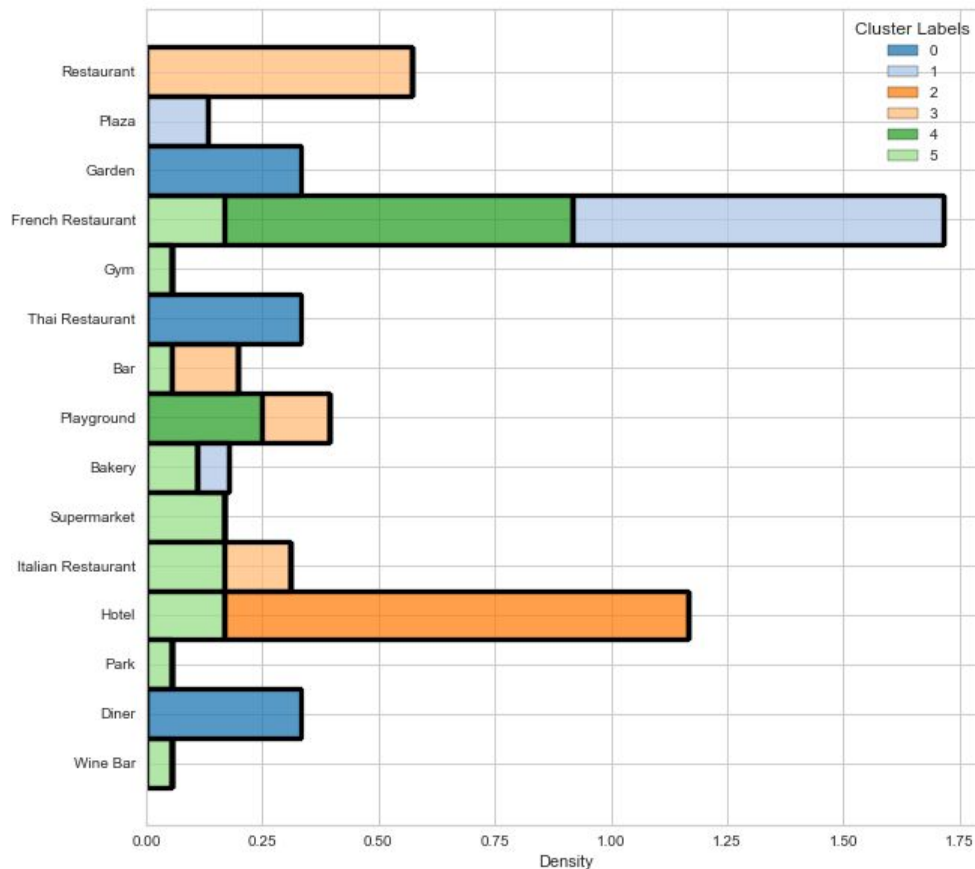
**Cluster 1:** Calm boroughs with lots of french restaurants

**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)

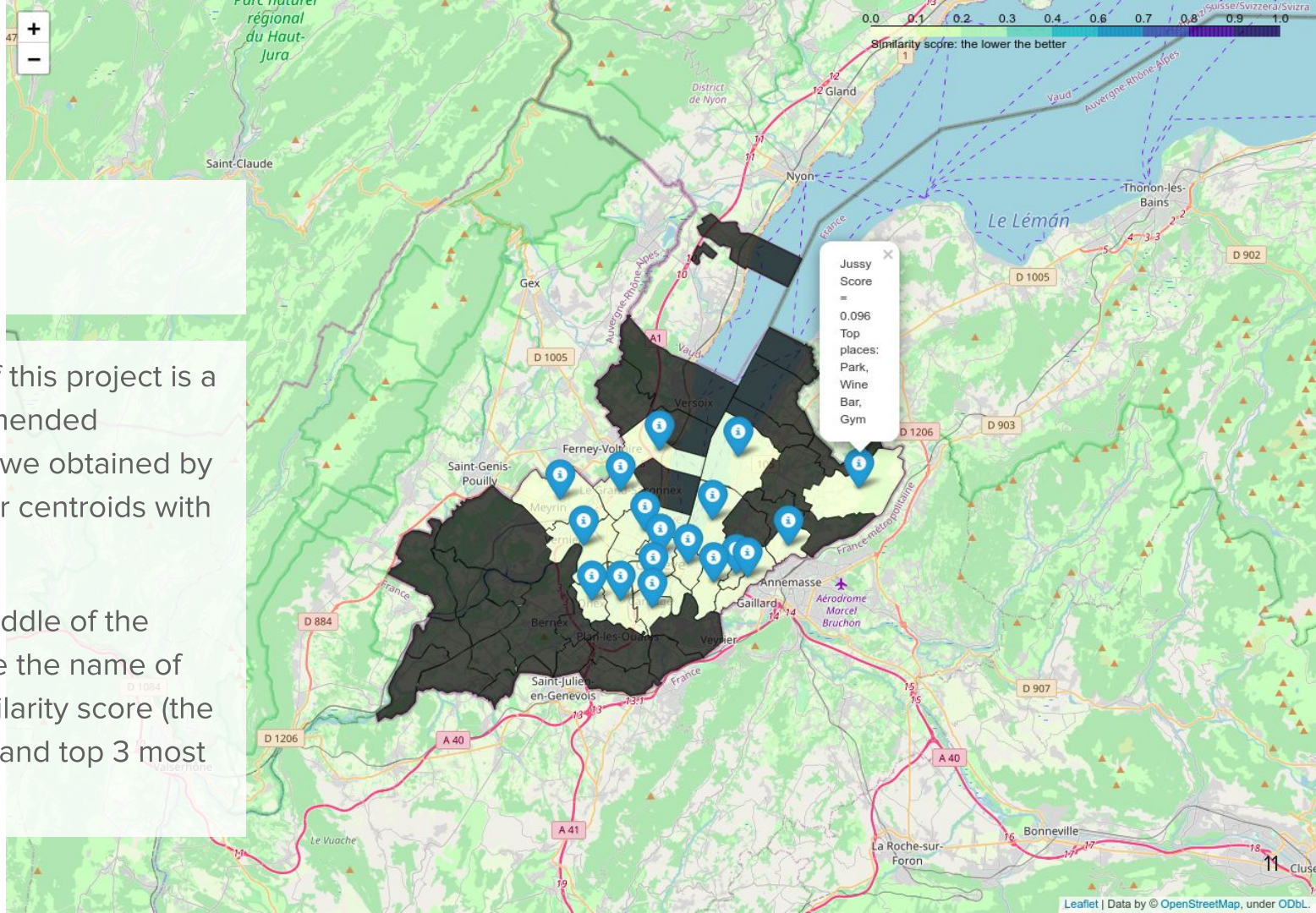




# Result

The final result of this project is a map with recommended boroughs, which we obtained by comparing cluster centroids with initial address.

Markers in the middle of the boroughs provide the name of the borough, similarity score (the lower the better) and top 3 most frequent venues.



# 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 rate** and **crime rate** in the neighbourhoods of interest. Moreover, would be interesting to consider **maximum distance** from a possible workplace/study place etc.

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.

# 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. As an output of the research we provide a map with recommended boroughs containing information about top 3 most frequent venues.

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.