Neighbourhoods in Paris

Capstone Project

Caterina Dolce - 06 November 2020

Moving to a new multicultural city is a complex process

- Moving to a new multicultural city is a complex process that involves many stressful tasks.

- One of them is looking for a new place where to live.

- On top of the challenge of finding a new place to call home, it is very important to find the neighbourhood that matches its own life-style and that has the type facilities that matter the most.

- Close neighbourhood could often be very different

Problem

- When moving to a new city it is not obvious to have a clear picture of each neighbourhood and to have the time to explore them before renting a new place.

- Finding the best-fit neighbourhood in a new city became sometimes more challenging than finds the perfect house

- This task is also challenging for the rental agency, which lacks the information that matters to suggest to a new tenant a neighbourhood based on his life-style and preferences.

This Study

- Focus on the city of Paris.
- To better understand and categorise the 20 arrondissements (neighbourhoods) of Paris based on the venues present in each of them.

Goal

- Help people moving to Paris to find the best neighbourhood to live based on their lifestyle.
- Help the rental agencies to understand better the main characterises of each neighbourhood and how it could suits each different life-style.

How

- With an analysis of each arrondissement
- Clustering arrondissements that are similar by venues type
- Associate each area/cluster with a certain lifestyle in order to help new tenants to find the perfect neighbourhood in Paris.

Data source

- Public data set with Arrondissement data and coordinate: https://www.data.gouv.fr/en/datasets/arrondissements-1/
- Foursquare API, to get the most common venues for each arrondissement of Paris.

Methodology

- The analysis has been performed in the IBM Skills Network Labs and the final notebook has been published in Github.

Methodology

Data cleaning

- Dropping the columns that will not be used
- Renaming the column Index
- Setting the correct Postal Code
- Assign a number to the Arrondissement column
- Separate Latitude and Longitude in 2 new columns
- Convert Coordinate in float

Methodology

Data cleaning

Before

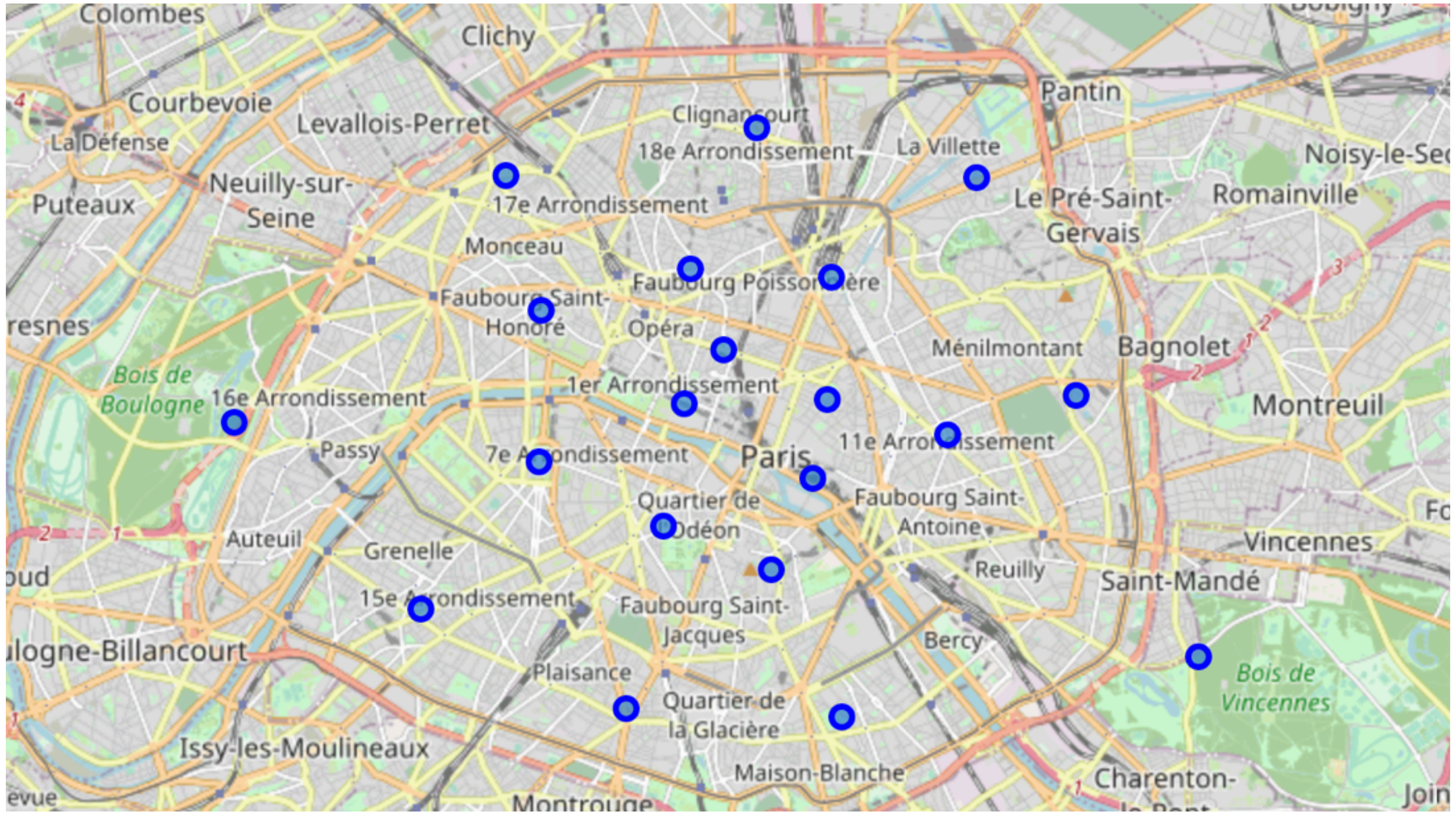
	n_sq_ar	c_ar	c_arinsee	l_ar	l_aroff	n_sq_co	surface	perimetre	geom_x_y	geom
0	750000011	11	75111	11ème Ardt	Popincourt	750001537	3.665442e+06	8282.011886	48.8590592213,2.3800583082	{"type": "Polygon", "coordinates": [[[2.396236...
1	750000003	3	75103	3ème Ardt	Temple	750001537	1.170883e+06	4519.263648	48.86287238,2.3600009859	{"type": "Polygon", "coordinates": [[[2.363828...
2	750000007	7	75107	7ème Ardt	Palais-Bourbon	750001537	4.090057e+06	8099.4		
3	750000005	5	75105	5ème Ardt	Panthéon	750001537	2.539375e+06	6239.1		
4	750000008	8	75108	8ème Ardt	Élysée	750001537	3.880036e+06	7880.5		

After

	Postal Code	Arr	Name	Latitude	Longitude
0	75001	1	Louvre	48.862562701 8	2.33644336205
1	75002	2	Bourse	48.8682792225	2.34280254689
2	75003	3	Temple	48.86287238	2.3600009859
3	75004	4	Hôtel-de-Ville	48.8543414263	2.35762962032
4	75005	5	Panthéon	48.8444431505	2.35071460958

Methodology

Geographical visualisation



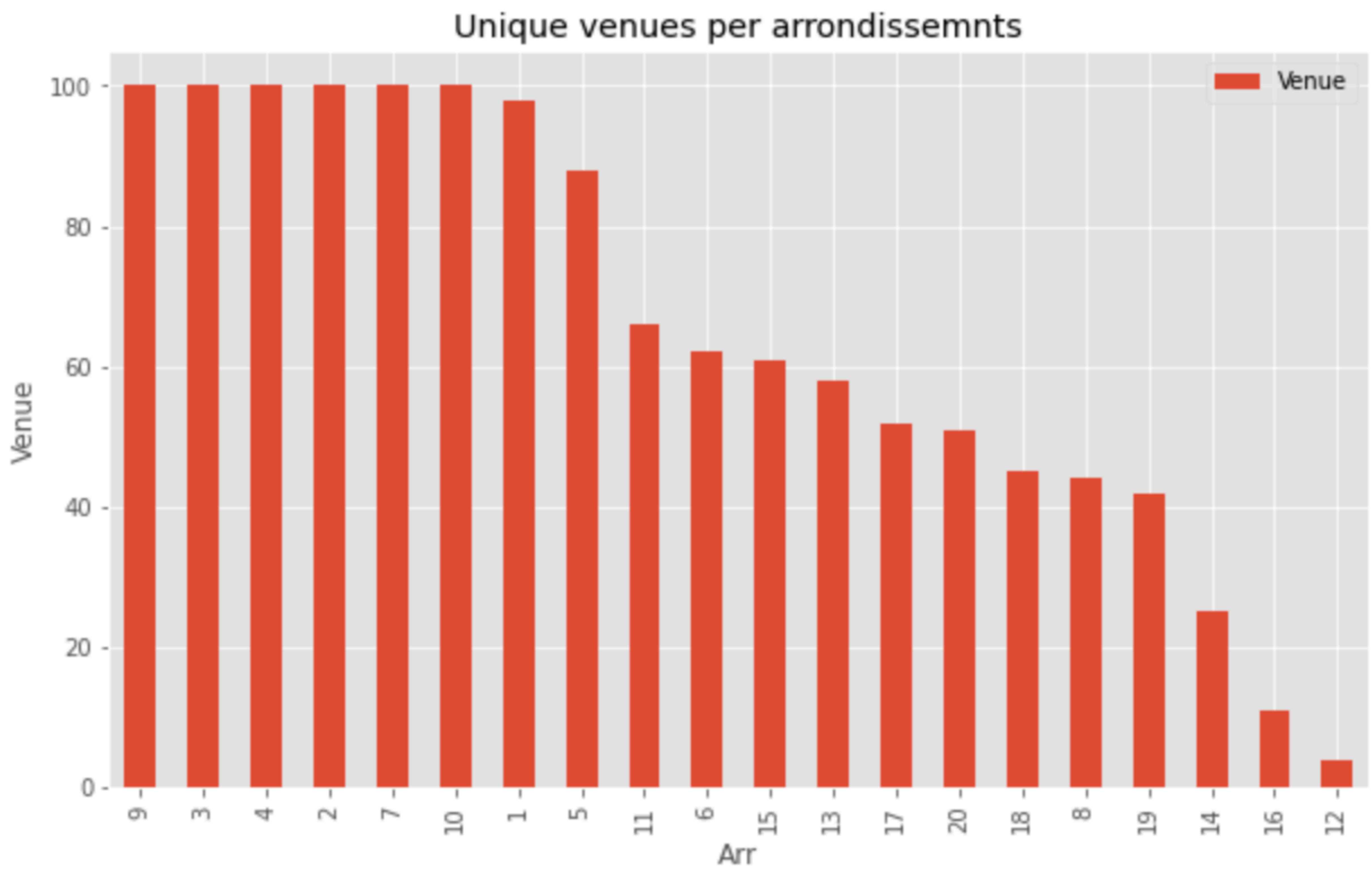
Python folium library to visualise geographic details of Paris and its arrondissement and I created a map of Paris with boroughs superimposed on top

Methodology

Analysis of all arrondissements

6 arrondissements reach the limit of 100 venues

while 7 arrondissement have less than 50 venues

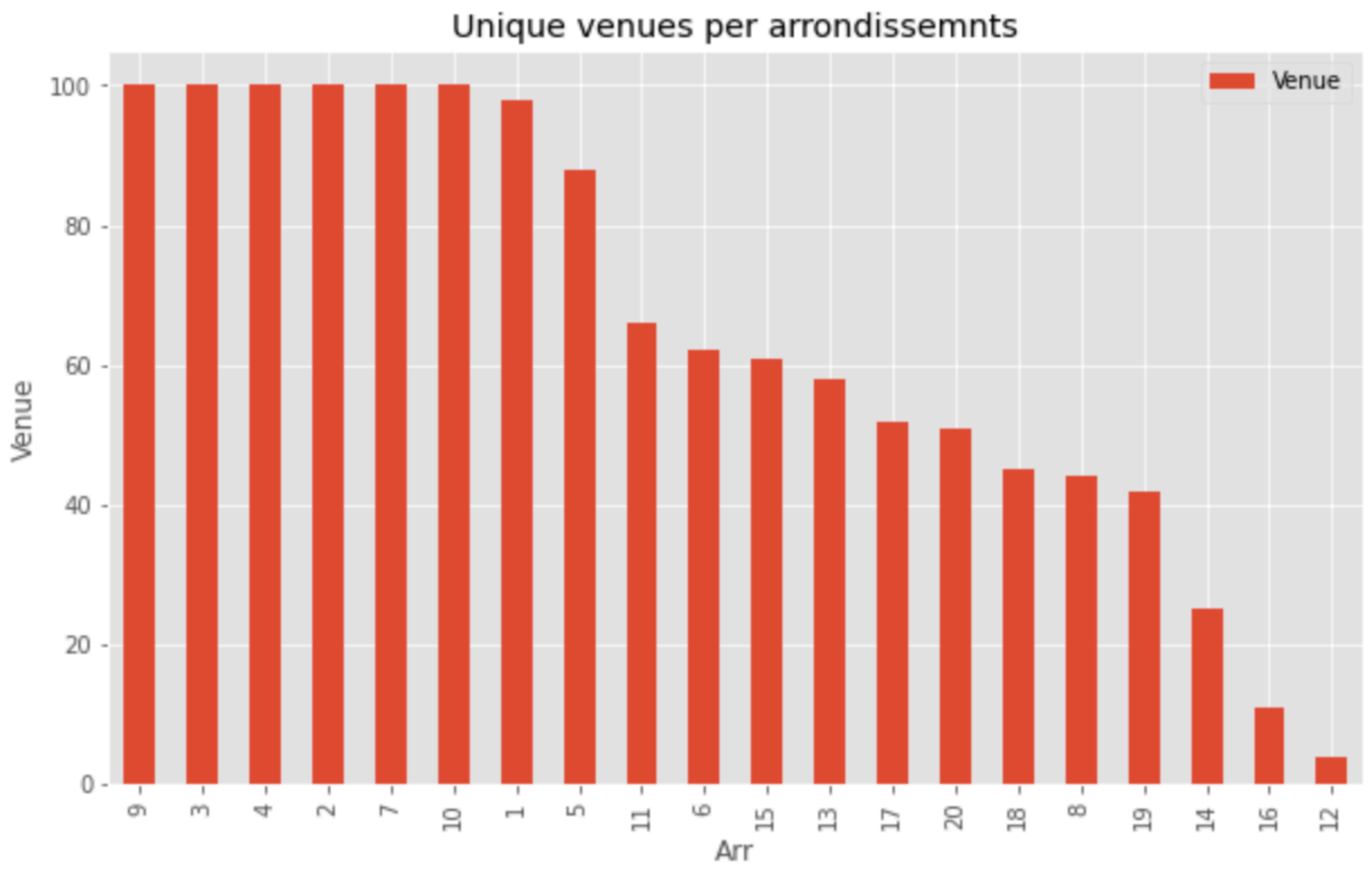


In our analysis we found in total **200 uniques venues** in all Paris.

Methodology

Clustering using k-means

We have some common venue categories in the arrondissement. This reason I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised algorithm. This method doesn't need previous recommendations to build a model. **K-mean method** is good for segmentation. It divides the data into clusters without any cluster-internal structures or label.



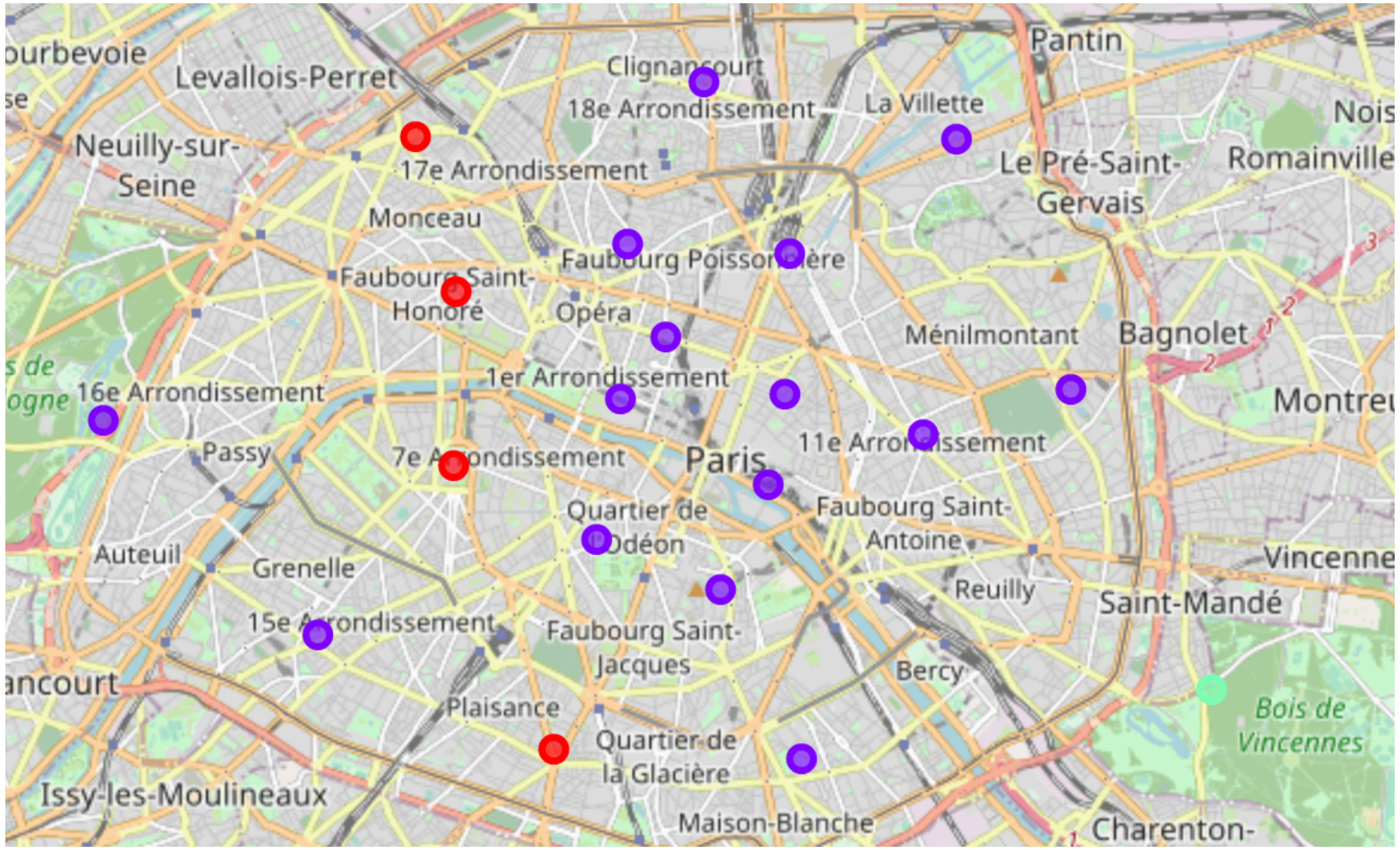
Thanks to the Elbow method 3 was found to be the optimal value for K

Clustering

- Cluster 0 (red dots) = The most popular venues are French restaurants and Hotels. There is not a huge variety of places to hand out. I will describe this cluster as: “Accommodation and French Restaurants”

- Cluster 1 (purple dots) = The second cluster has a bigger variety of international restaurants and wine bars. This second cluster seems to be a more place to socialise and it offer more venues to hand out. I will define “Multiple social venues”

- Cluster 3 (green dot) = we only have one arrondissement that is part of this cluster, that is dominated by Zoos and Supermarket. I will define this area as “Quite residential area"



Conclusion

- Thanks to this analysis, a first clustering of the arrondissement of Paris was done.
- The analysis has been performed used public available data and using Foursquare to get information about the venues present in each location.
- This analysis identified 3 main type of neighbourhoods that can be find in Paris. These 3 clusters are not simply related to their geographical position but they distributed across the city.

Improvement

- For refining the model, more type of data could have be added to have a better idea about the places and the kind of life style and services provided from each arrondissement.

- Other methods of unsupervised learning can be used to solve the problem as agglomerative algorithm or density-based clustering.

Perspective

- This kind of analysis could be used from people that are moving to a new city, to have a better understand of different neighbour of the new place and to help them find easily the perfect location where to live.
- Also, this kind of analysis could be integrated in rental platforms to improve house search engine or used from rental agency to improve the house rent/buy recommendation to new customers.