Capstone Project
Neighbourhoods in Paris

1. Introduction

1.1 Background

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. People that have lived in multicultural cities know how much each neighbourhood different from another.

1.2 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.

For this study, I will focus on the city of Paris. I would like to better understand and categorise the 20 arrondissements (neighbourhoods) of Paris based on the venues present in each of them.

1.3 Interest

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

The rental agency would find beneficial to have an accurate analysis of each area as this will increase to chance to find the perfect neighbourhood for a new tenant as this will increase the lasting time of a tenant in a property and customer satisfaction. The new mover will find beneficial to have a quick analysis of each arrondissement, in order to find quicker the areas that most suit them.

The goal of this project is to analysis and describe each arrondissement, cluster arrondissements that are similar by venues type, and associate each area/cluster with a certain lifestyle (eg. Family, People that love going out, quiet area, residential area) in order to help new tenants to find the perfect neighbourhood in Paris.

2 Data Description

I will use the neighbourhood's data in the city of Paris taken from a public dataset from the city of Paris [1]. I have dowloaded the csv file and imported in my notebook.

I will explore and cluster them based on the type of venues present in each area. I will find the most popular venues from Foursquare API.

- 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.

3 Methodology

In this section the element of data analysis performed will be described. This could include exploratory data analysis, any inferential statistical testing, and , if any, the machine learnings used.

The analysis has been performed in the IBM Skills Network Labs and the final notebook has been published in Github. The IBM Skills Network Labs is part of the Cognitive Class project at IBM.

3.1 Data Cleaning

The main steps of the data cleaning include are the following:

- 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

With the previous steps we go from the initial Dataframe:

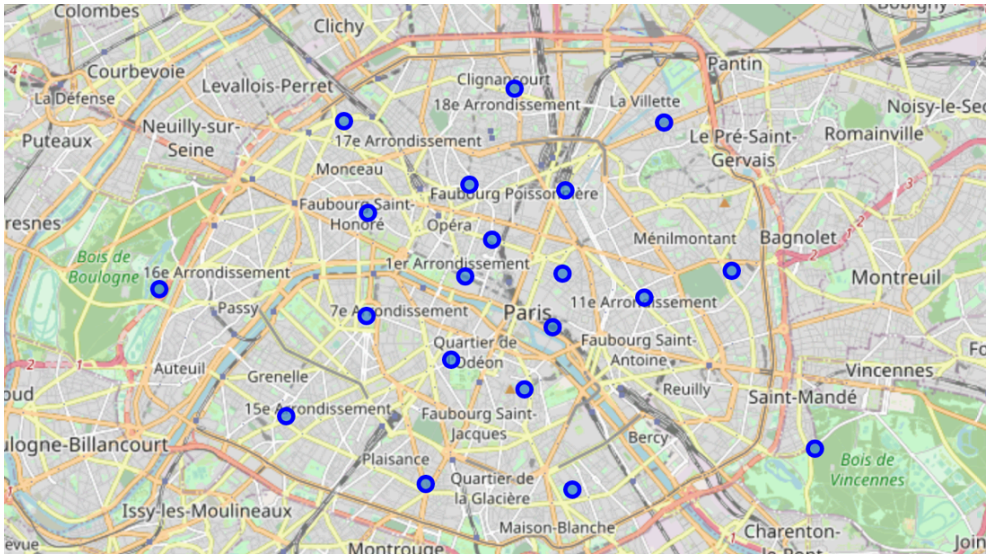
	n_sq_ar	c_ar	c_arinsee	l_ar	l_aroff	n_sq_co	surface	perimetre	geom_x_y	geom
0	750000011	11	75111	11ème Ardt	Popincourt	750001537	3.665442e+06	8282.011886	48.8590592213,2.3800583082	{"type": "Polygon", "coordinates": [[[2.396236...
1	750000003	3	75103	3ème Ardt	Temple	750001537	1.170883e+06	4519.263648	48.86287238,2.3600009859	{"type": "Polygon", "coordinates": [[[2.363828...
2	750000007	7	75107	7ème Ardt	Palais-Bourbon	750001537	4.090057e+06	8099.424883	48.8561744288,2.31218769148	{"type": "Polygon", "coordinates": [[[2.320902...
3	750000005	5	75105	5ème Ardt	Panthéon	750001537	2.539375e+06	6239.195396	48.8444431505,2.35071460958	{"type": "Polygon", "coordinates": [[[2.364433...
4	750000008	8	75108	8ème Ardt	Élysée	750001537	3.880036e+06	7880.533268	48.8727208374,2.3125540224	{"type": "Polygon", "coordinates": [[[2.325836...

To this final cleaned Dataframe:

	Postal Code	Arr	Name	Latitude	Longitude
0	75001	1	Louvre	48.8625627018	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

3.2 Geographic visualisation

I used python **folium** library to visualise geographic details of Paris and its arrondissement and I created a map of Paris with boroughs superimposed on top. I used latitude and longitude values to get the visual as below:



I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as **100 venue** and the radius **500 meter** for each arrondissement from the given latitude and longitude informations. I have used the method **GET**, the **Endpoint Group venus**, *and the* ***Endpoint explore***.

3.3 Arrondissement venues analysis

3.3.1 Analysis of a first arrondissement

Here is a head of the list Venues name, category, latitude and longitude informations from Forsquare API for the **First Arrondissement**. There are in total 98 venues and 45 unique categories.

	name	categories	lat	lng
0	Musée du Louvre	Art Museum	48.860847	2.336440
1	Palais Royal	Historic Site	48.863236	2.337127
2	Comédie-Française	Theater	48.863088	2.336612
3	La Clef Louvre Paris	Hotel	48.863977	2.336140
4	Place du Palais Royal	Plaza	48.862523	2.336688

Also, it was interesting to see what were the most popular categories in this arrondissement. By doing as sense check the venues below make so much sense, as the 1st arrondissement is well known to be a touristic area (this explain the many hotels), but also it is the area with the highest concentration of Japanese restaurant in all Paris. Young folk go there to eat the best Japanese in the city.

	categories	name
0	French Restaurant	9
1	Japanese Restaurant	7
2	Plaza	6
3	Hotel	6
4	Coffee Shop	5
5	Ramen Restaurant	4
6	Café	4
7	Italian Restaurant	4
8	Art Museum	3
9	Korean Restaurant	3

From this simple analysis I could already provide a summary of this arrondissement. I better picture will come from comparing this data with the other arrondissement. I permed a similar analysed for all the other arrondissements.

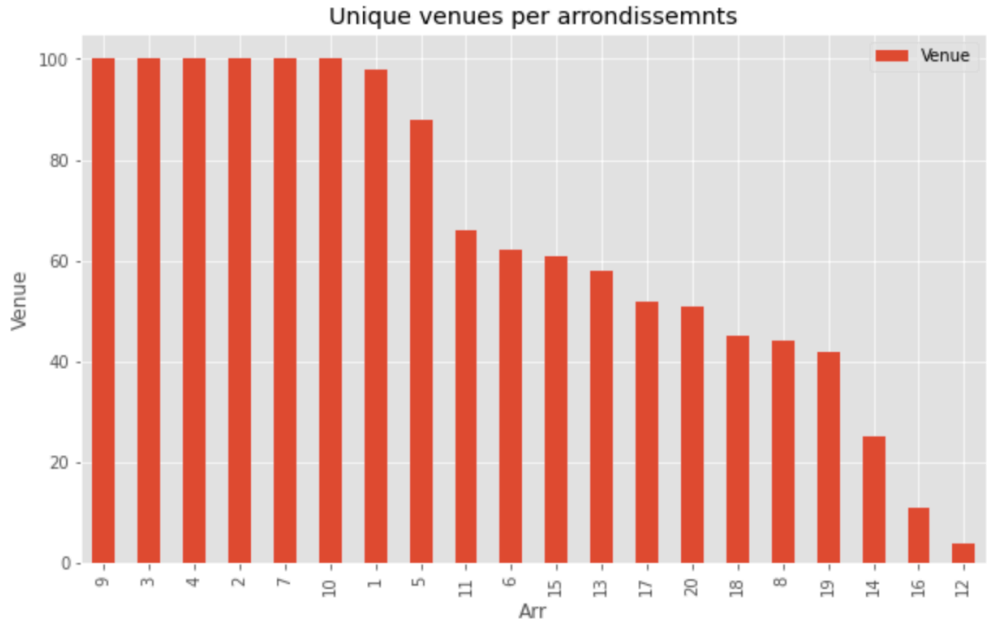
3.3.2 Analysis of all the arrondissements

From the analysis of all the arrondissements (performed applying the analysis done to the first arrondissement to all the other) I noticed that:

- 6 arrondissements reach the limit of 100 venues (2,3,4,7,9,10 arrondissement), while 7 arrondissement have less than 50 venues (12,16,14,19,8,18).

- There are in total **200 uniques venues** in all Paris.

In the Bar Chart below, I report the number of unique venues by arrondissement.



In summary of this graph **200** unique categories were returned by Foursquare, then I created a table which shows list of top **10** venue category for each arrondissement reported in the table below.

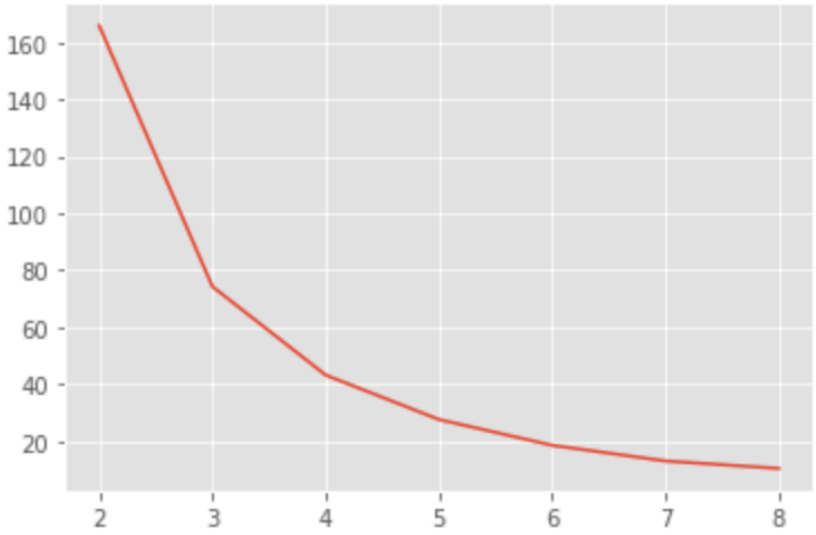
We observe that there are a few venue that are recurrently popular in some arrondissement. In order to help a new Tenant to pick the best arrondissement for his flat search, it would be helpful to cluster the arrondissement by the venues that they have in common. In this way would be much easier to pick the arrondissement where to focus for flat search.

	Arr	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	1	French Restaurant	Japanese Restaurant	Hotel	Plaza	Coffee Shop	Italian Restaurant	Ramen Restaurant	Café	Art Museum	Korean Restaurant
1	2	French Restaurant	Wine Bar	Cocktail Bar	Italian Restaurant	Bistro	Hotel	Salad Place	Bakery	Pizza Place	Furniture / Home Store
2	3	Bakery	French Restaurant	Coffee Shop	Burger Joint	Bistro	Cocktail Bar	Japanese Restaurant	Art Gallery	Café	Clothing Store
3	4	French Restaurant	Ice Cream Shop	Hotel	Clothing Store	Pedestrian Plaza	Italian Restaurant	Cocktail Bar	Pastry Shop	Bistro	Bookstore
4	5	French Restaurant	Science Museum	Hotel	Italian Restaurant	Café	Plaza	Bar	Pub	Coffee Shop	Bakery
5	6	French Restaurant	Bistro	Italian Restaurant	Bakery	Pastry Shop	Wine Bar	Pub	Chocolate Shop	Fountain	Hotel
6	7	Hotel	French Restaurant	Café	Italian Restaurant	Plaza	Cocktail Bar	History Museum	Coffee Shop	Art Museum	Bistro
7	8	French Restaurant	Hotel	Japanese Restaurant	Spa	Cocktail Bar	Art Gallery	Theater	Fast Food Restaurant	Mediterranean Restaurant	Sporting Goods Shop
8	9	French Restaurant	Hotel	Bakery	Bistro	Cocktail Bar	Wine Bar	Lounge	Japanese Restaurant	Pizza Place	Tea Room
9	10	French Restaurant	Hotel	Coffee Shop	Indian Restaurant	Bistro	Café	Italian Restaurant	Pizza Place	Japanese Restaurant	Korean Restaurant

3.4 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.

First, I will run K-Means to cluster the boroughs into 3 clusters because when I analyze the K-Means with elbow method it ensured me the 3 degree for optimum k of the K-Means.



4 Results

The Paris 20 arrondissement have been clustered in 3 groups based on their most common venues.

Here is my partial merged table with cluster labels for each borough.

Postal Code	Arr	Name	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
75001	1	Louvre	48.862563	2.336443	1	French Restaurant	Japanese Restaurant	Hotel	Plaza	Coffee Shop	Italian Restaurant
75002	2	Bourse	48.868279	2.342803	1	French Restaurant	Wine Bar	Cocktail Bar	Italian Restaurant	Bistro	Hotel
75003	3	Temple	48.862872	2.360001	1	Bakery	French Restaurant	Coffee Shop	Burger Joint	Bistro	Cocktail Bar
75004	4	Hôtel-de-Ville	48.854341	2.357630	1	French Restaurant	Ice Cream Shop	Hotel	Clothing Store	Pedestrian Plaza	Italian Restaurant
75005	5	Panthéon	48.844443	2.350715	1	French Restaurant	Science Museum	Hotel	Italian Restaurant	Café	Plaza
75006	6	Luxembourg	48.849130	2.332898	1	French Restaurant	Bistro	Italian Restaurant	Bakery	Pastry Shop	Wine Bar
75007	7	Palais-Bourbon	48.856174	2.312188	0	Hotel	French Restaurant	Café	Italian Restaurant	Plaza	Cocktail Bar
75008	8	Élysée	48.872721	2.312554	0	French Restaurant	Hotel	Japanese Restaurant	Spa	Cocktail Bar	Art Gallery

Below the map that help visualise the cluster of the arrondissement. The red, purple, and green markers represents that 3 group of clusters.

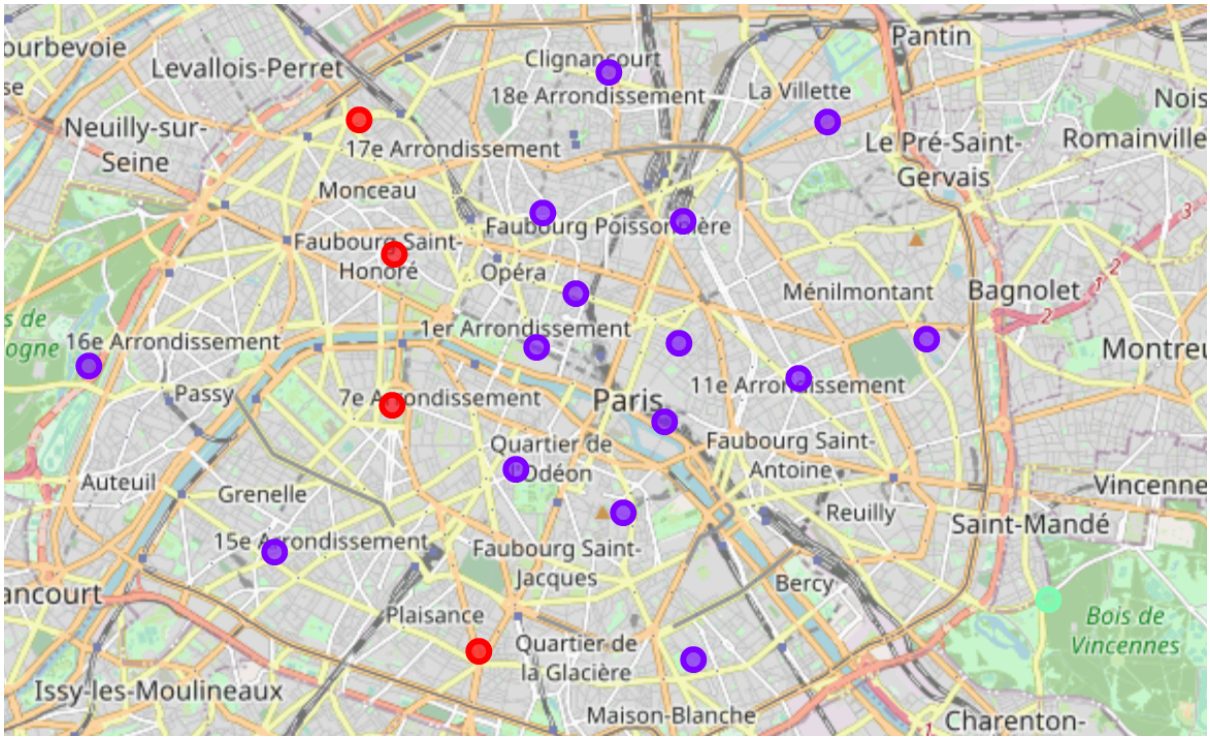
At this point it is interesting to observe the difference in the main clusters.

By analysing the 2 most popular venues in all three clusters we observe than:

- 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"



5 Discussion and Recommendations

Paris is a multicultural city with a huge variety of venues and cultural complexity.

As there is such a complexity, use only the venues to clustering the arrondissement it is only the first step for an accurate classification.

I used a Kmeans algorithm as part of this clustering study I have found an optimum k value of only 3. However for a more accurate description, the data set could be expanded with demographic data of the people already living in that area.

The 3 clusters were reported in the map of Paris, where each of the arrondissement had a cluster related color. This help visualise to identify the similarity among clusters.

I then concluded the analyses by looking at the specification of each cluster, by analysing the most frequent venues by cluster. This last analysis helped to find a label for each of the cluster that could match with a life style searched from new people moving in Paris.

6 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.

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. 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.

In addition of working with a more gradual database, other methods of unsupervised learning can be used to solve the problem as agglomerative algorithm or density-based clustering.

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.

7 References

[1] https://www.data.gouv.fr/en/datasets/arrondissements-1/