Explore the Housing Rental and Sale Prices & Venues Data Analysis of Paris

Letian DAI

March 25, 2020

1. Introduction

Figure 1 : Paris at night [1].

Paris, capital of France, is one of the largest metropolises in the world with a population of **2,148,271 habitants** (official estimate, 1 January 2020) in an area of 105 square kilometers (**41 square miles**) [2]. Paris ranks as **4th most expensive** real estate city in the world from the data of UK branch of the American commercial real estate company CBRE [3]. Because I have been in Paris for more than 5 years and I like to live in this beautiful city with a lower cost, it is very useful for me and the people like me to know the local prices of renting or buying a house in Paris. This information is also useful for potential investors of property in Paris. That is my motivation to create a map and information chart to present the real estate index in Paris and the density of venue in each district.

2. Data Acquisition and Cleaning

2.1. Data sources

To consider the problem, I need to acquire the geographic and location data of each borough in Paris from the data base of French official site, the venues of given borough in Paris from **Foursquare API** , latest Average Housing Rental Price (Avg_HRP) and Average Housing Sale Price (Avg_HSP) in each borough of Paris is sourced from the site **Seloger**.

2.2. Data cleaning

The official geographic and location data of each borough in Paris is written in French with accent character (For example, the name of borough in French 'Opéra'). I need to replace them with English without accent 'Opera'. Otherwise it might cause some problems in Jupyter notebook. There are also some empty cells that I dropped them directly. I only selected the useful columns which contain 'Borough' (Name of borough), 'Postcode' (Postal code), 'Latitude' and 'Longitude'. The geojson of Paris will be used to fill in the color which is related to the price information. In order to make the link between different data frame file, I used the 'Postcode' as their common column. The featureidkey 'id' in geojson is the last two numbers inside the 'Postcode'.

The data of price was scraped from the site **Seloger**. The price is written in string not integer. So I need to grab the part of number and change its format from string to integer.

I used the database in the repository in my study. The cleaned data has the main components *Postcode*, *Borough*, *Latitude, Longitude, Average House Sale Price (Avg_HSP)* and *Average Housing Rental Price (Avg_HRP)* information in each district in Paris.

	Postcode	Borough	Latitude	Longitude	Avg_HSP	Avg_HRP
0	75001	LOUVRE	48.862630	2.336293	14 869 €/m²	39 €/m²
1	75002	BOURSE	48.867903	2.344107	13 501 €/m²	39 €/m²
2	75003	TEMPLE	48.863054	2.359361	14 198 €/m²	39 €/m²
3	75004	HOTEL-DE-VILLE	48.854228	2.357362	14 858 €/m²	41 €/m²
4	75005	PANTHEON	48.844509	2.349859	14 533 €/m²	39 €/m²
5	75006	LUXEMBOURG	48.848968	2.332671	16 862 €/m²	42 €/m²
6	75007	PALAIS-BOURBON	48.856083	2.312439	15 892 €/m²	40 €/m²
7	75008	ELYSEE	48.872527	2.312583	14 366 €/m²	39 €/m²
8	75009	OPERA	48.876896	2.337460	12 696 €/m²	37 €/m²
9	75010	ENTREPOT	48.876029	2.361113	11 442 €/m²	36 €/m²
10	75011	POPINCOURT	48.859415	2.378741	11 804 €/m²	35 €/m²
11	75012	REUILLY	48.835156	2.419807	10 940 €/m²	33 €/m²
12	75013	GOBELINS	48.828718	2.362468	10 918 €/m²	32 €/m²
13	75014	OBSERVATOIRE	48.828993	2.327101	11 644 €/m²	33 €/m²
14	75015	VAUGIRARD	48.840155	2.293559	11 707 €/m²	33 €/m²
15	75016	PASSY	48.860399	2.262100	12 981 €/m²	35 €/m²
16	75017	BATIGNOLLES-MONCEAU	48.887337	2.307486	12 253 €/m²	35 €/m²
17	75018	BUTTE-MONTMARTRE	48.892735	2.348712	10 895 €/m²	33 €/m²
18	75019	BUTTES-CHAUMONT	48.886869	2.384694	9 643 €/m²	30 €/m²
19	75020	MENILMONTANT	48.863187	2.400820	9 995 €/m²	30 €/m²

Table 1 : Data frame of borough in Paris with the average housing rental and sale price.

3. Exploratory Data Analysis

3.1. Visualize the geographic details of each borough

I visualize geographic details of each borough inside Paris using python folium library with the each latitude and longitude data. I created a map of Paris with boroughs with blue marker superimposed on the top as shown below:

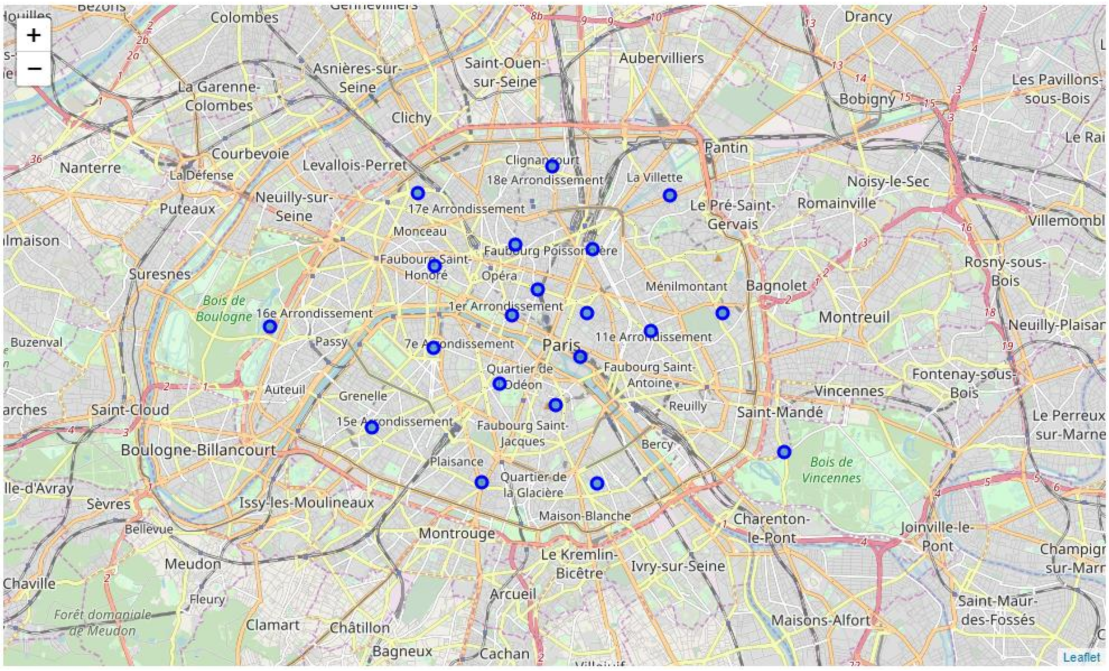

Figure 2 : Map of Paris with blue marker at each borough center.

3.2. Explore the boroughs with Foursquare API

I used the Foursquare API to explore the boroughs and segment them. I set the limit as **100 venues** and the radius **500 meter** for each borough from their corresponding latitude and longitude data. Here is the head of list name (Venues' names), catagories, latitude (lat) and longitude (lng) information from Foursquare API in Paris.

	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	Cour Napoléon	Plaza	48.861172	2.335088
4	Place du Palais Royal	Plaza	48.862523	2.336688

Table 2 : The head of list of venue returned by Foursquare API.

In summary of this data **1342** venues were returned by Foursquare API (**version 20200325**). Here is the head of merged table of boroughs and venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	LOUVRE	48.86263	2.336293	Musée du Louvre	48.860847	2.336440	Art Museum
1	LOUVRE	48.86263	2.336293	Palais Royal	48.863236	2.337127	Historic Site
2	LOUVRE	48.86263	2.336293	Comédie-Française	48.863088	2.336612	Theater
3	LOUVRE	48.86263	2.336293	Cour Napoléon	48.861172	2.335088	Plaza
4	LOUVRE	48.86263	2.336293	Place du Palais Royal	48.862523	2.336688	Plaza

Table 3 : Merged table of boroughs and venues.

I counted the venue number in each borough using **groupby method** and plotted the dataframe of borough and venue count number in the graph below. On one hand, we can see that Temple, Bourse, Entrepôt, Palais-Bourbon, Opéra, Hôtel-de-Ville and Louvre have reached the **100** limit of venues. On the other hand, Passy and Reuilly have less than 20 venues. The number of venue depends on the latitude and longitude information of borough and the radius of area we set.

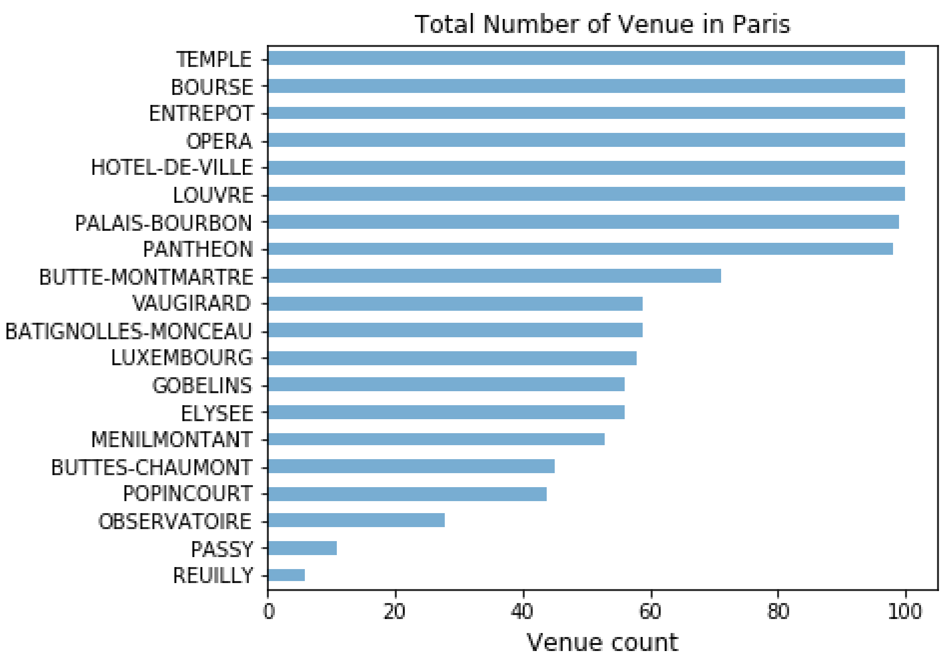

Figure 3: Total number of venue in each borough in Paris.

In summary of my venue results, **204** unique categories were returned by Foursquare. I listed the head of top **10** venue category in each borough in the following table:

	Borough	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	BATIGNOLLES-MONCEAU	Hotel	French Restaurant	Italian Restaurant	Café	Japanese Restaurant	Bakery	Bistro	Plaza	Diner	Portuguese Restaurant
1	BOURSE	French Restaurant	Cocktail Bar	Italian Restaurant	Wine Bar	Coffee Shop	Hotel	Bakery	Thai Restaurant	Bistro	Pastry Shop
2	BUTTE-MONTMARTRE	French Restaurant	Bar	Pizza Place	Hotel	Restaurant	Italian Restaurant	Plaza	Bistro	Café	Vietnamese Restaurant
3	BUTTES-CHAUMONT	French Restaurant	Bar	Supermarket	Bistro	Seafood Restaurant	Beer Bar	Hotel	Café	Canal	Burger Joint
4	ELYSEE	French Restaurant	Hotel	Bakery	Spa	Corsican Restaurant	Cocktail Bar	Plaza	Mediterranean Restaurant	Brasserie	Cycle Studio

Table 4 : Head of top 10 venue category in each borough in Paris.

3.3. Class the boroughs into 7 clusters

We have some common venue categories in boroughs. In this reason, I used unsupervised leaning **K-means algorithm** to cluster the boroughs (K-means algorithm is one of the most common cluster method of unsupervised leaning algorithms).

First, I run the K-means algorithm to cluster the borough to obtain the optimal K value for K-means algorithm. Because it is not obvious to pick the optimal K value, I decide to choose K = 7. It means I will cluster the boroughs into 7 clusters.

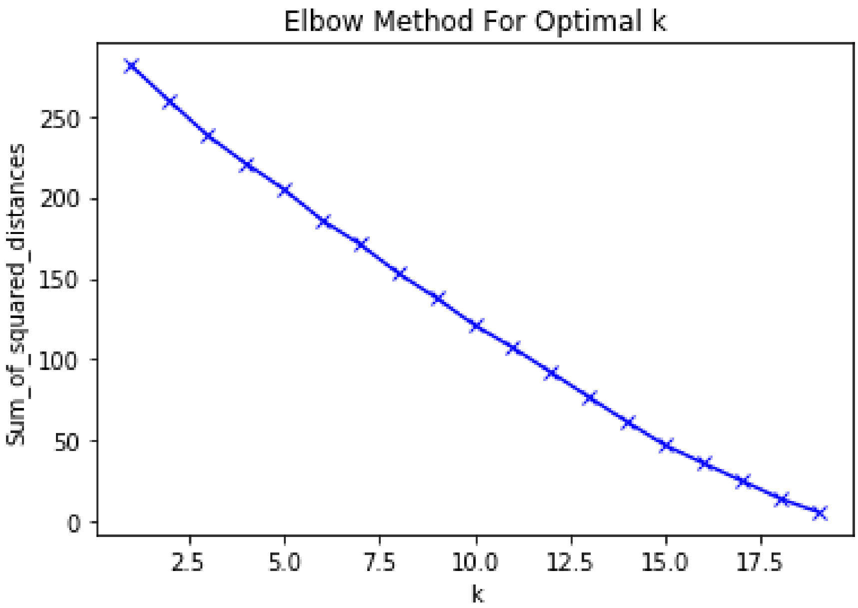

Figure 4 : Sum of square distances for each K value in the K-means algorithm clustering.

Here is the head of my merged table with cluster labels for each borough in Paris.

Borough	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
LOUVRE	48.862630	2.336293	1	French Restaurant	Hotel	Japanese Restaurant	Plaza	Café	Coffee Shop	Bakery	Art Museum	Historic Site	Udon Restaurant
PANTHEON	48.844509	2.349859	1	French Restaurant	Hotel	Italian Restaurant	Bar	Bakery	Café	Plaza	Coffee Shop	Wine Bar	Pub
OPERA	48.876896	2.337460	0	French Restaurant	Hotel	Bistro	Bar	Wine Bar	Bakery	Lounge	Cocktail Bar	Pizza Place	Theater
ENTREPOT	48.876029	2.361113	1	French Restaurant	Bistro	Hotel	Coffee Shop	Café	Pizza Place	Bar	Japanese Restaurant	Asian Restaurant	Indian Restaurant
PASSY	48.860399	2.262100	2	Lake	Pool	French Restaurant	Plaza	Bus Station	Bus Stop	Art Museum	Bike Rental / Bike Share	Park	Boat or Ferry

Table 5 : Head of merged table with cluster labels and each borough in Paris.

4. Results

4.1. Visualization of each cluster on the map

Using the merged table with cluster labels for each borough in Paris, we can visualize the resulting clusters in the map as following:

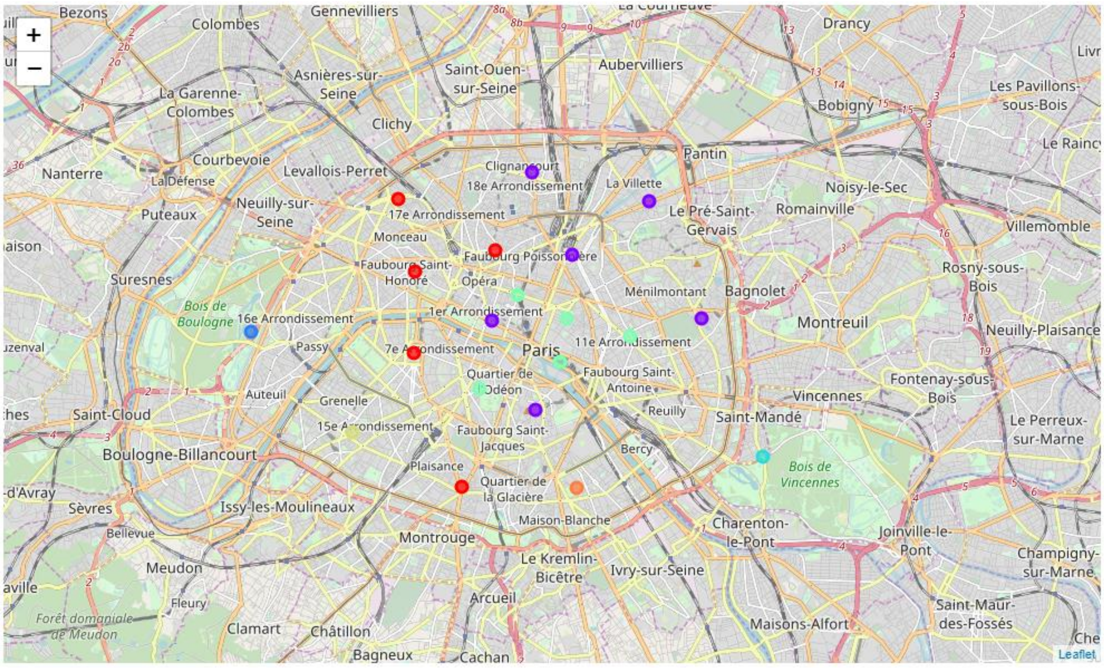

Figure 5 : Boroughs of 7 clusters in different colors.

4.2. Analysis of each cluster

I used the **groupby** method by '*Cluster Labels*' to obtain the most frequent in **1st Most Common Venue** in each cluster.

Cluster Labels	count	unique	top	freq
0	5	2	French Restaurant	3
1	6	2	French Restaurant	5
2	1	1	Lake	1
3	1	1	Zoo Exhibit	1
4	5	4	French Restaurant	2
5	1	1	Italian Restaurant	1
6	1	1	Vietnamese Restaurant	1

Table 6 : The describe information of merged table groupby 'Cluster Labels'.

When we examine above table we can get the most frequent in **1st Most Common Venue** of each cluster as follows:

Cluster Labels	top
0	French Restaurant
1	French Restaurant
2	Lake
3	Zoo Exhibit
4	French Restaurant
5	Italian Restaurant
6	Vietnamese Restaurant

Table 7 : The most frequent in 1st Most Common Venue of each cluster.

4.3. Visualization of each borough's area on the map

I can use the **Geojson** and **Plotly** to fill the color of each borough in Paris with meaningful color. The geojson open data file of Paris can be downloaded from [French government site](). I first showed the map of Paris with each *Postcode* as below:

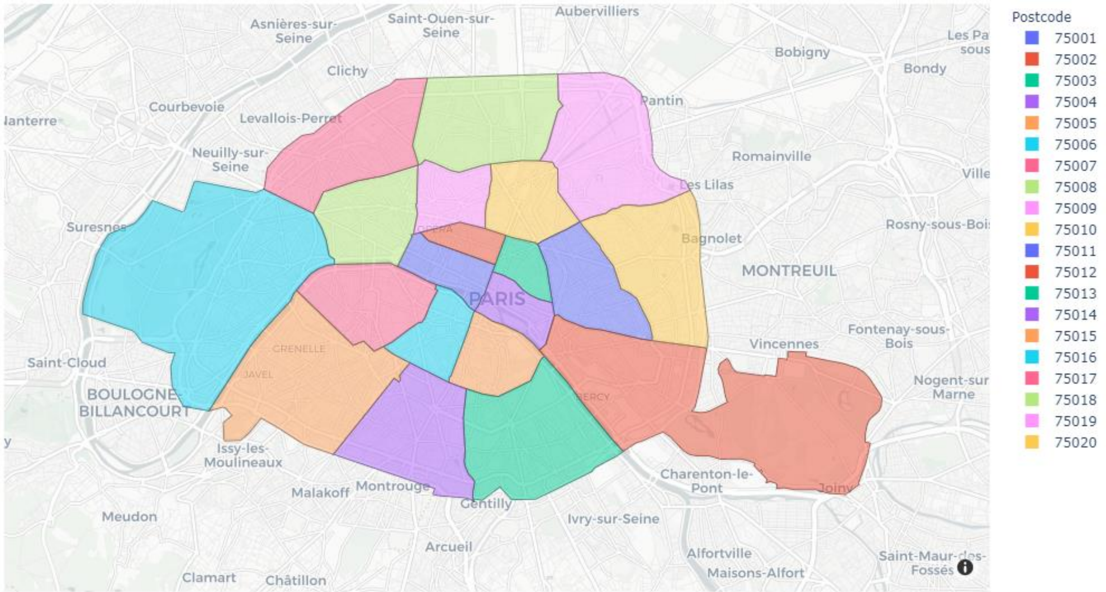

Figure 6 : Map of Paris with painted borough.

Because we have clustered the boroughs into 5 clusters, I plotted the map of each borough in Paris with *Cluster Labels* as below:

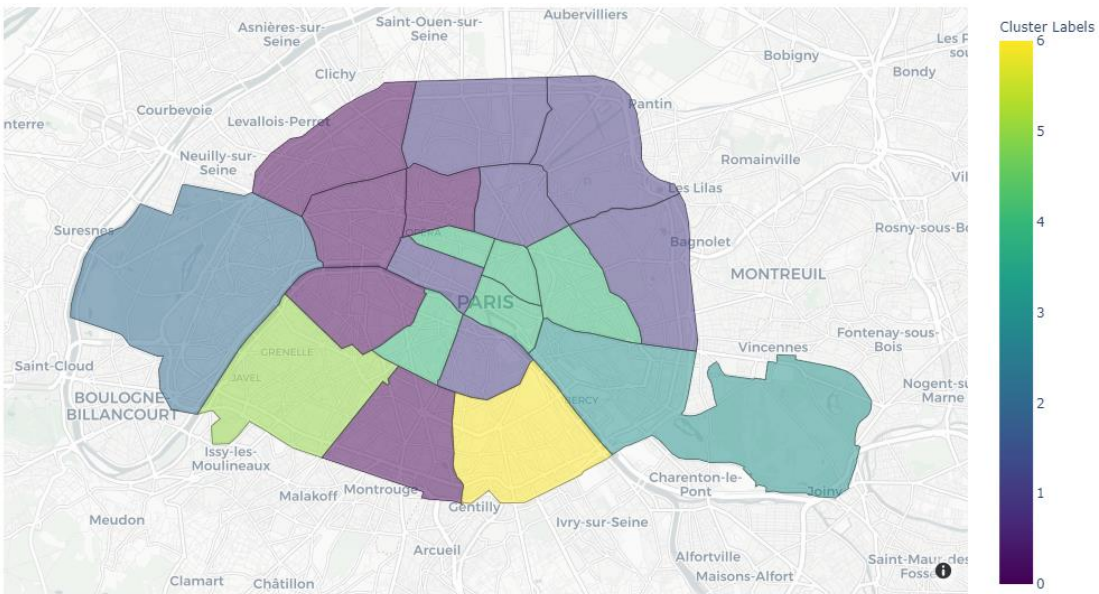

Figure 7 : Map of Paris with painted borough in each cluster.

According to the data of *Average House Sale Price (Avg_HSP)* and *Average Housing Rental Price (Avg_HRP),* I plotted the map of Paris with each *Avg_HSP* and *Avg_HRP* as below:

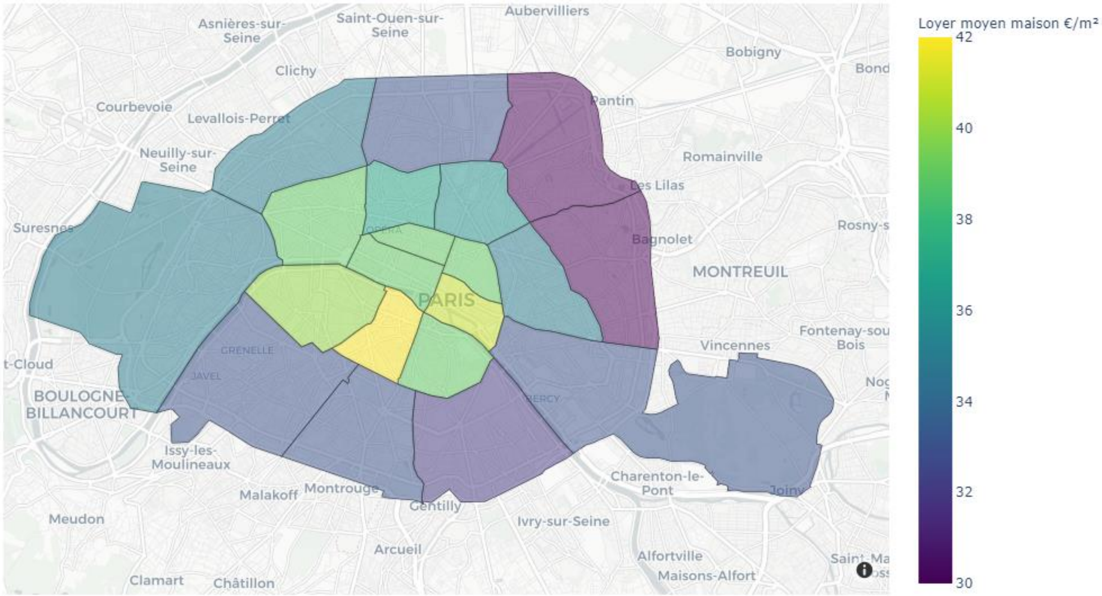

Figure 8 : Average Housing Rental Price (Avg_HRP) €/m² of Paris in 2019.

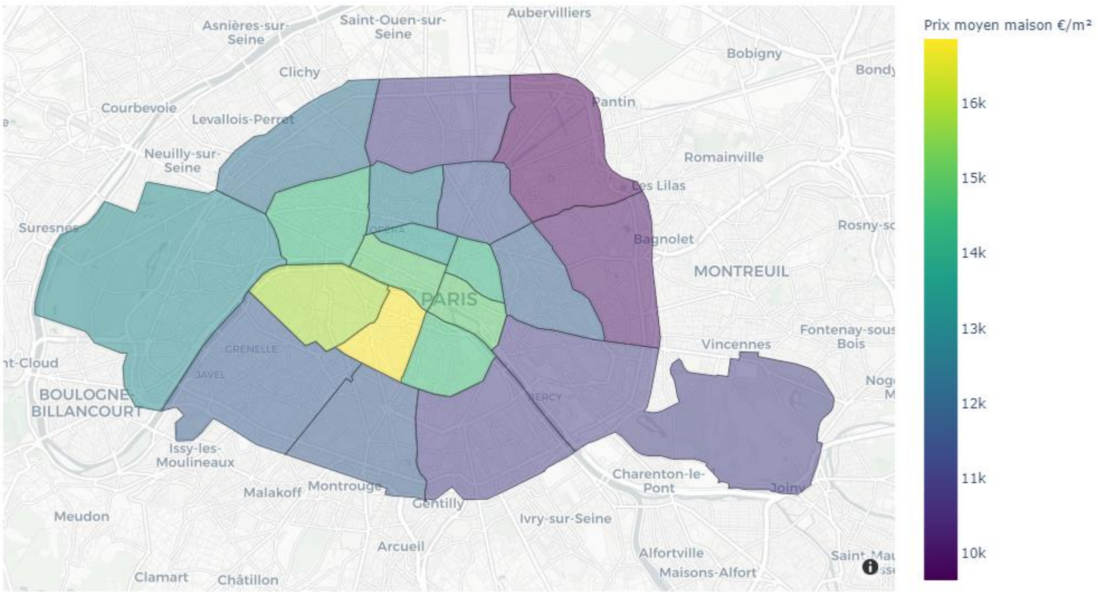

Figure 9 : Average Housing Sale Price (Avg_HSP) €/m² of Paris in 2019

5. Discussion

Using the **Foursquare API**, a very useful API, we obtained the information of the **Most Common Venue** in each borough. We observed that the **1st Most Common Venue** is the category of **French Restaurant** for **16** boroughs out of 20.

I visualized the data using the **Plotly** and **Geojson**. It is obviously to find the correlation between the Average Housing Rental Price and Average Housing Sale Price thorough the visualization. I used the linear regression to fit the scatter points as below:

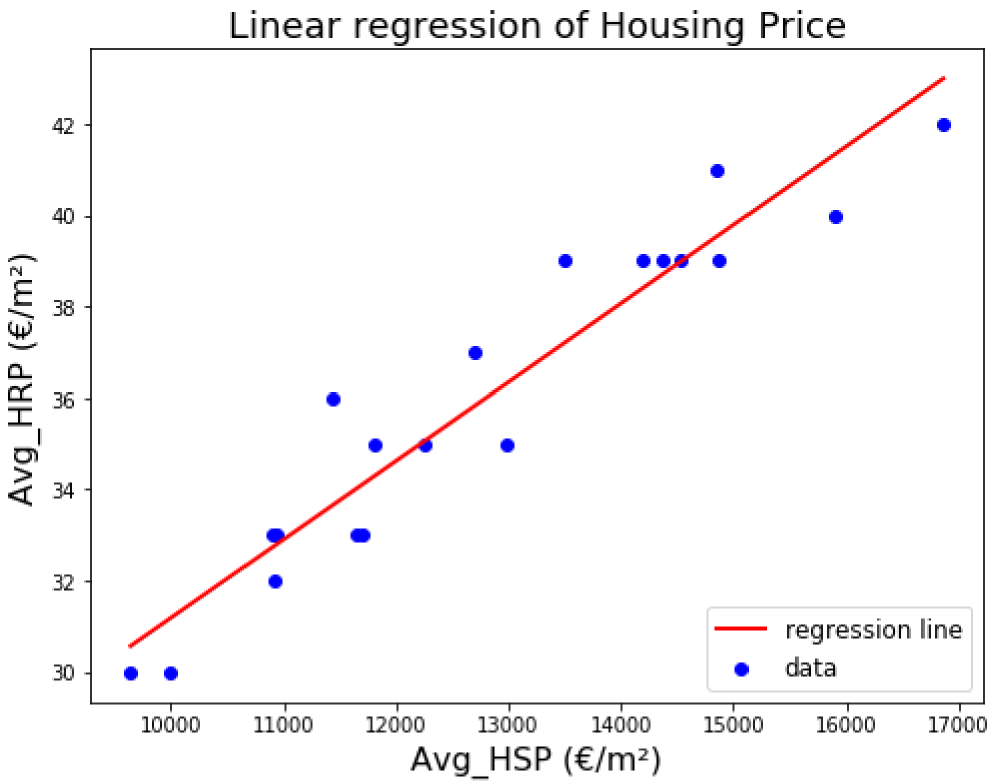

Figure 10 : Linear regression of average Housing Sale Price to average Housing Rental Price in 2019 in Paris.

The borough in the central part of Paris has relatively higher Housing Rental and Sale Price. Not only for investors and habitants, but also for city managers can manage the city more regularly by using similar data analysis types or platforms.

Finally, I obtained the visualizing data about the clustering information and the Housing Rental and Sale Price on the Paris map. In the future studies, web or telephone applications can be carried out to direct investors or users. The update of data can also be accessed dynamically from specific platforms or packages.

6. Conclusion

In this study, I analyzed the relationship between the Housing Sale Price and Housing Rental Price in Paris and the cluster of boroughs using the venues returned by **Foursquare API**. I have found almost linear relation between the Housing Sale Price and Housing Rental Price in Paris and the relation between the price and geographic location in Paris. This information can be used for the habitants and investors to think about the choosing the best place to live in Paris with reasonable cost.

7. Reference

[1]http://eskipaper.com/images/paris-evening-1.jpg

[2]https://en.wikipedia.org/wiki/Paris

[3]https://parispropertygroup.com/blog/2015/paris-ranks-4th-expensive-city-buy-real-estate/