Midnight in Paris

A Data Science Project

Juan Manuel Gelos

August 5, 2020

Table of Contents

1. Introduction

1.1. Background

Midnight Company is a (Fantastic) tematic restaurant looking to open a branch in Paris. They are interested in finding the neighborhood that best suits their charisma: love for antiques.

The restaurant prepares cuisine from all countries, offering a thematic setting and dishes based on different golden periods of each country

Its main concern is to find a place where people and tourists usually enjoy existing restaurants, in areas surrounded with antiques shop

1.2 Problem

So, the big question to solve is: *What is the best neighborhood in Paris fitting:*

- *In first place, has enough antiques shops that assure attraction on target customers*
- *In second, has an important presence of restaurants that assure the flow of customers on streets*
- *In third place, the selected Neighborhood is better than others with similar cuisines*

This project aims to classificate Paris neighborhoods in order to make the best recommendation

1.3 Interest

The principal interested is, of course, the Midnight Company. Others who could be interested are antiques shops in Paris and existing restaurants

2. Data acquisition and cleaning

2.1. Data Sources

For this project, we need to analyze neighborhoods in Paris, and use geolocalization to associate neighborhoods with restaurants and antiques shops

We will use Foursquare (www.fourtsquare.com) geolocated data API to collect data about restaurants and antiques shops in Paris to analize and segment neighborhoods to find the best

Paris Neiborhoods Names (called Arrondissements) will be scrapped from WikiPedia (https://en.wikipedia.org/wiki/Arrondissements_of_Paris)

Also, each Neighborhood geolocalization data will be obtained using Nominatim GEOPY geocoder

2.2. Data Cleaning

In Paris, Neighborhood are called Arrondissement. Data scrapped from Wikipedia looks like this:

	Arrondissement (R for Right Bank, L for Left Bank)	Name	Area (km2)	Population(2017 estimate)	Density (2017) (inhabitants per km2)	Peak of population	Mayor	2020-2026
0	Paris Centre 1st (Ier) / 2nd (IIe) / 3rd (IIIe...	Louvre, Bourse, Temple, Hôtel-de-Ville	5.59 km2 (2.16 sq mi)	100196	17924	before 1861	Ariel Weil (PS)	NaN
1	5th (Ve) L	Panthéon	2.541 km2 (0.981 sq mi)	59631	23477	1911	Florence Berthout (DVD)	NaN
2	6th (VIe) L	Luxembourg	2.154 km2 (0.832 sq mi)	41976	19524	1911	Jean-Pierre Lecoq (LR)	NaN
3	7th (VIIe) L	Palais-Bourbon	4.088 km2 (1.578 sq mi)	52193	12761	1926	Rachida Dati (LR)	NaN
4	8th (VIIIe) R	Élysée	3.881 km2 (1.498 sq mi)	37368	9631	1891	Jeanne d'Hauteserre (LR)	NaN

As you can see, Paris Centre are divided in 4 Neighborhood, so I must made some data transformation to split their names.

In order to obtain correct geolocalization it was tested all neighborhood names with a search string in the format: “Neighbordhood, Paris, France” and display all the results in a map

In order to improve cuisines classification, we will combine “French restaurants” with “Restaurants” because it reflects better the cuisines segmentation.

We also made some cleaning, filtering and grouping to allow count on different restaurants and antiques shops.

3. Methodology

3.1. Data exploratory analysis

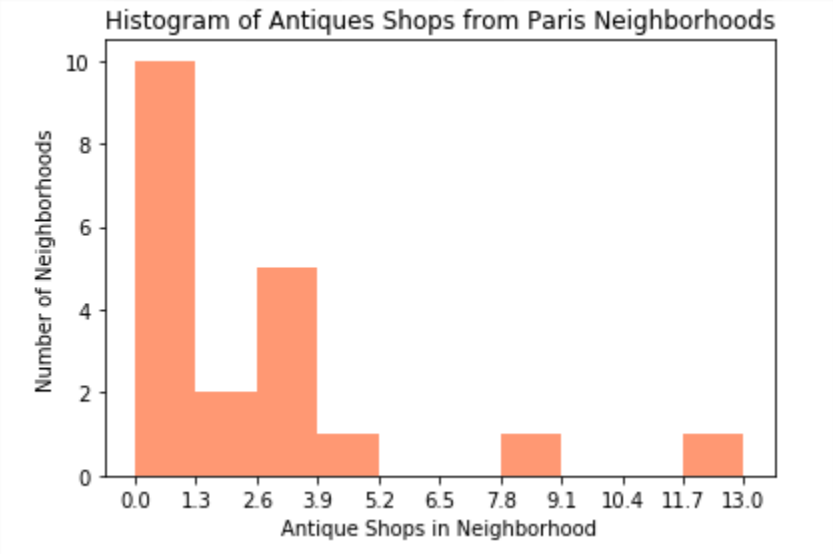
With Data Collected and prepared we principally focused on two mainly features:

- Total Antique Shops in the Neighborhood
- Total Restaurants in the Neighborhood

Neighborhood	Latitude	Longitude	AntiqueShops	Restaurants
Louvre	48.861147	2.338028	3	28
Bourse	48.868630	2.341474	3	68
Temple	48.866500	2.360708	2	66
Hôtel-de-Ville	48.856426	2.352528	8	68
Panthéon	48.846191	2.346079	1	63
Luxembourg	48.850433	2.332951	13	42
Palais-Bourbon	48.861596	2.317909	3	18
Élysée	48.846644	2.369830	0	15
Opéra	48.870645	2.332330	3	65
Entrepôt	48.876106	2.359910	0	63
Popincourt	48.858416	2.379703	5	42
Reuilly	48.839615	2.395752	2	23
Gobelins	48.832397	2.355583	1	31
Observatoire	48.829567	2.323962	3	20
Vaugirard	48.841370	2.300383	1	41
Passy	48.857505	2.280983	1	39
Batignolles-Monceau	48.881452	2.316667	1	35
Butte-Montmartre	48.890012	2.346467	1	34
Buttes-Chaumont	48.878396	2.381201	0	17
Ménilmontant	48.866708	2.383374	1	28

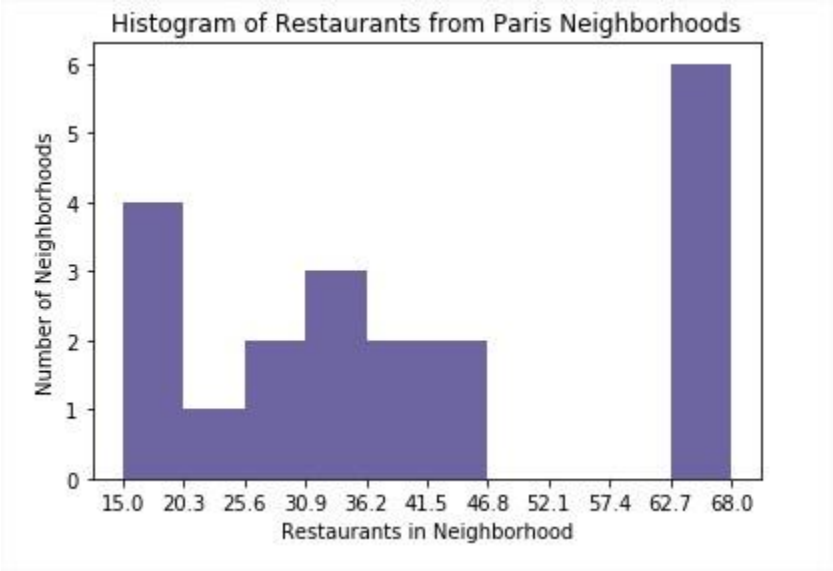
3.2. Analyzing Antique Shops Density by Neighborhood

In the Histogram, we clearly see that most Neighborhoods has fewer Antiques Shops, then you have two Neighborhoods with a good density of antiques shops (8 and 13 shops respectively)



3.3. Analyzing Antique Shops Density by Neighborhood

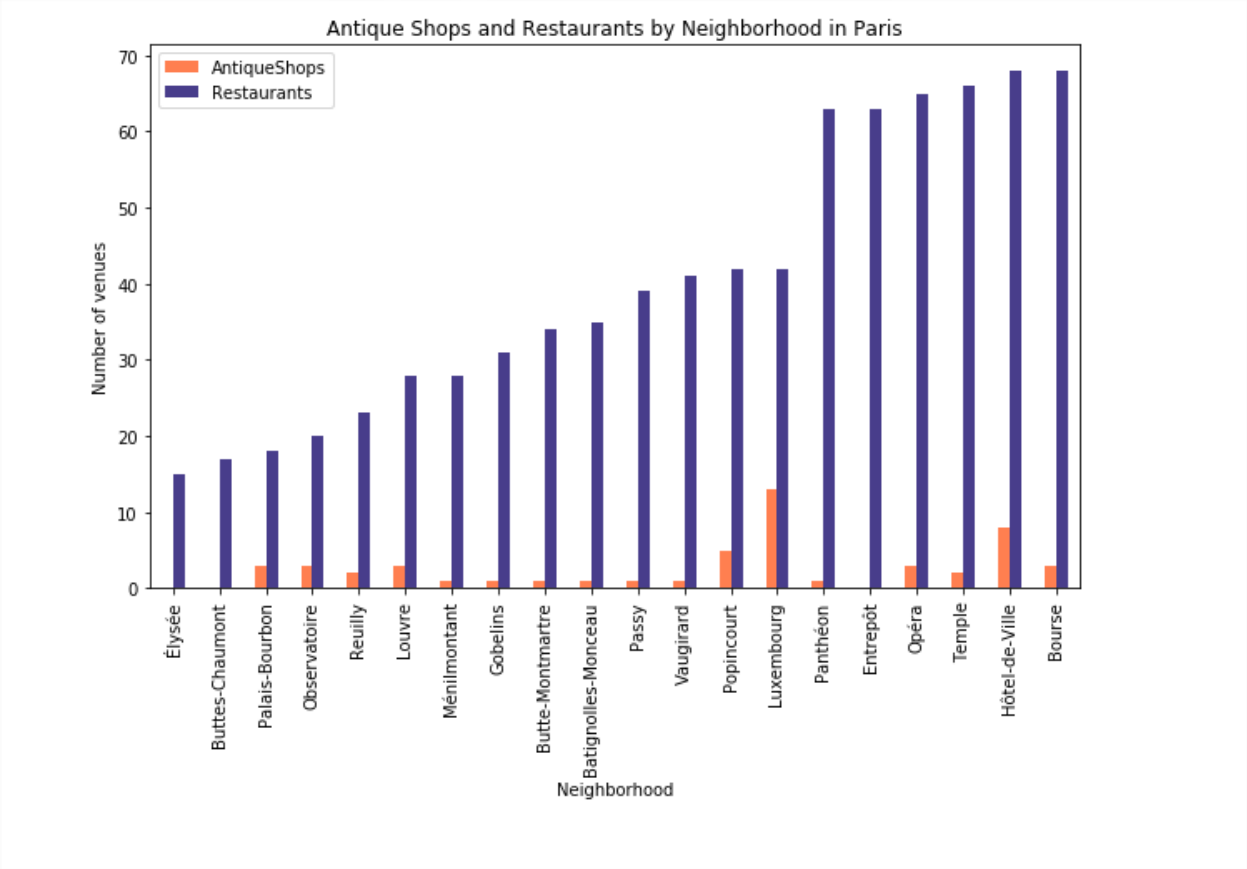
In the Histogram, we clearly see that most Neighborhoods has fewer Antiques Shops, then you have two Neighborhoods with a good density of antiques shops (8 and 13 shops respectively)



3.4. Comparing Antiques Shops & Restaurants

At first sight it seems that this features are no correlated, but in order to see it clearly were going to compare them. If no correlation exists, it will useful to see diferences between the two main features that the company needs to balance in order to achieve his objective

Lets seem them together in a Bar Chart

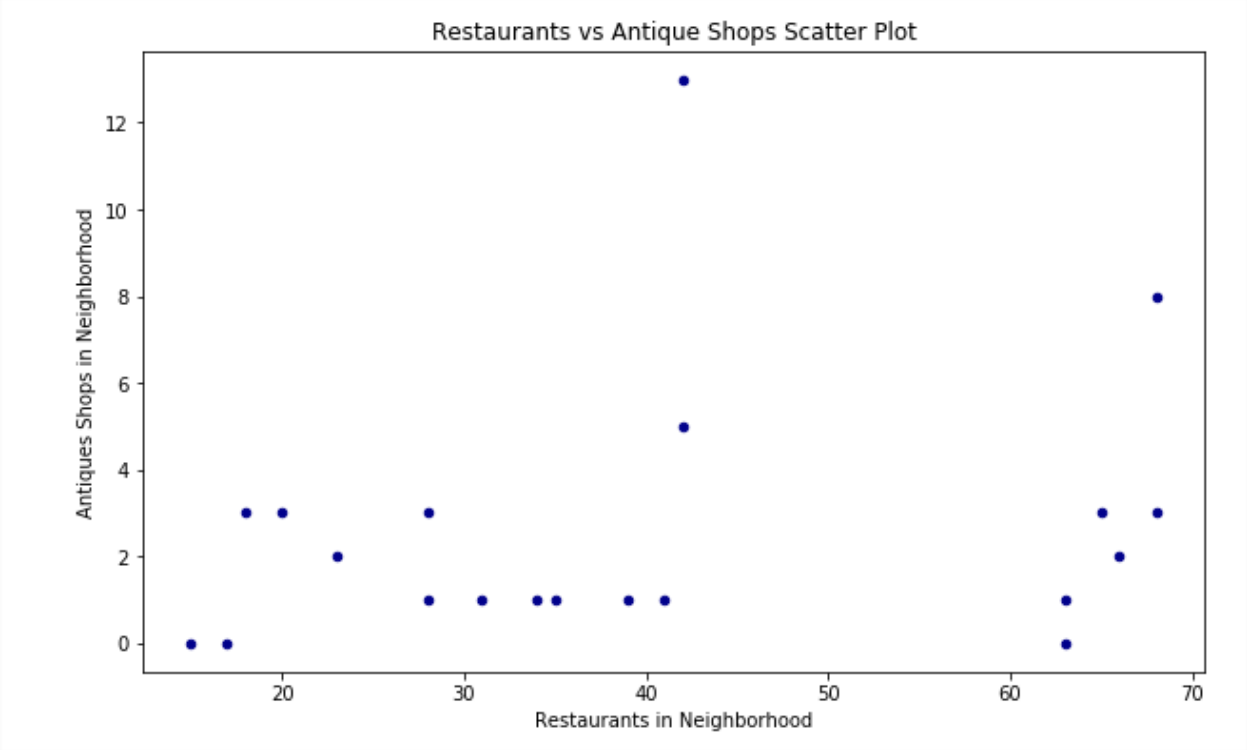


We can observe two facts:

* Luxembourg has most Antique Shops in Paris, with a high count of Restaurants

* Hotel-de-ville has a second position in Antique Shops count, bus is in the top off rest aurant counts

So far, we can say there is no correlation between this features. Let’s demonstrate this with a scatter plot:



This Scatter plot show there is NO CORRELATION between Restaurants and Antique Shops by Neiborhood in Paris

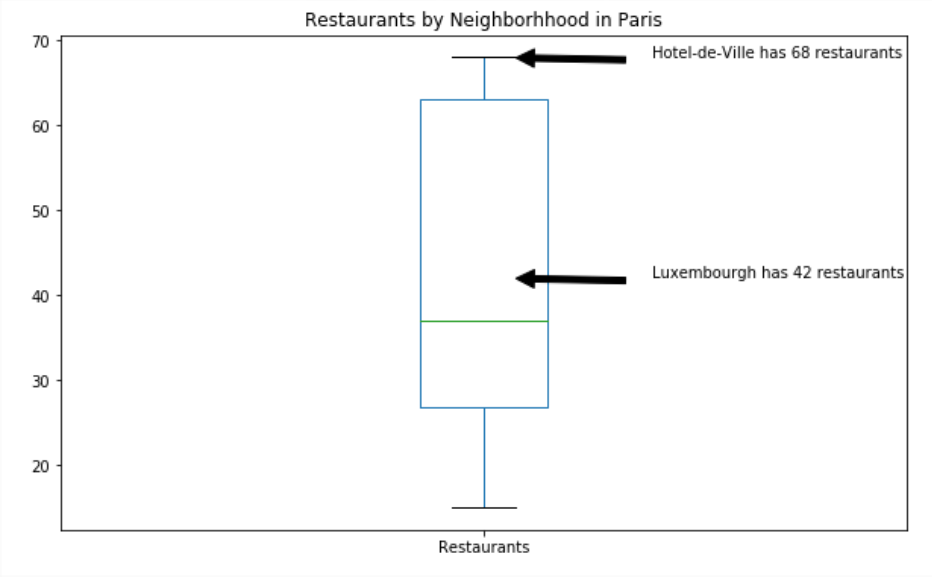
3.5. Comparing Candidate Neiborhoods: Luxembourg vs Hotel-de-ville

In previous section we observe the main characteristics that put this Neiborhood in our sight. Lets compare them in contrast with all the Neighborhood

In this box plot, we can observe thath Luxemburg, with 13 Antique Shops is an outliner, is really in the top of the hill on Antique Shops count, well above average on whole Paris

Hotel-de-Ville is also an outliner, bun in lesser way.

So, what about Restaurants count?

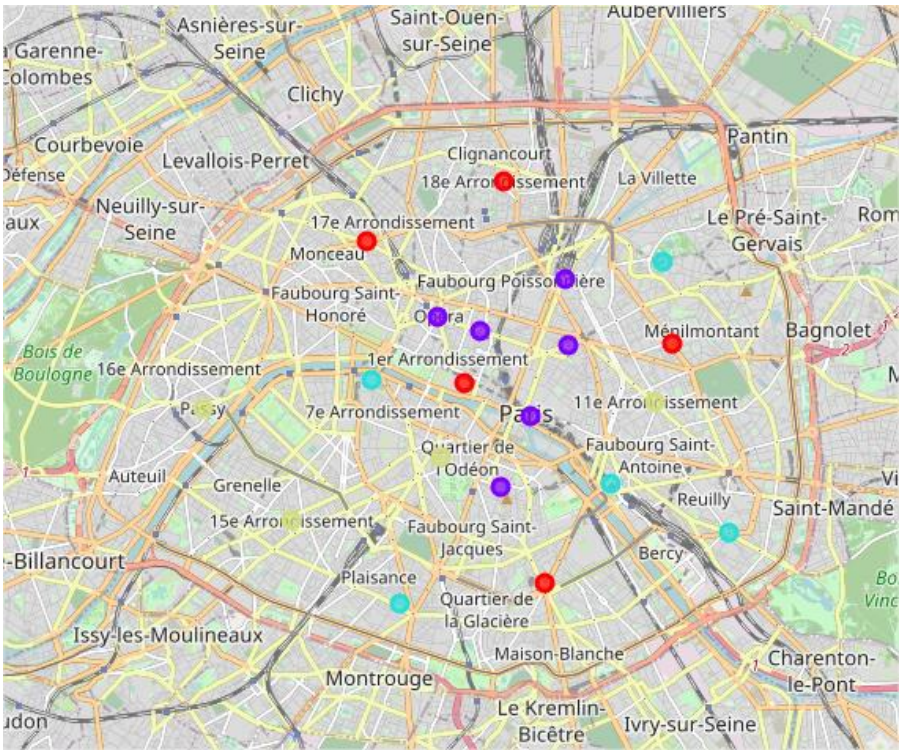


In this box graph we can easily see that Luxembourg is much lower than Hotel-de-Ville (in number of restaurants), but still above average, within the upper quartile

3.6. Food Clusters

Food Cluster is important for the company because it defines the overall kind of food his competitors offer in each Neighborhood. Clustering Neighborhoods by this data bring us the capacity to compare the cuisine offer in wich the company prefer to set his branch

For this purpouse, where going to apply Machine Learning to calculate Clustering in four Clusters of similar restaurant categories offer as seen in the next image:



Were interested in the clusters in wich our candidate belongs, as we can see in the next section

3.7. Analizing and Comparing Luxembourg's and Hôtel-de-Ville's Clusters

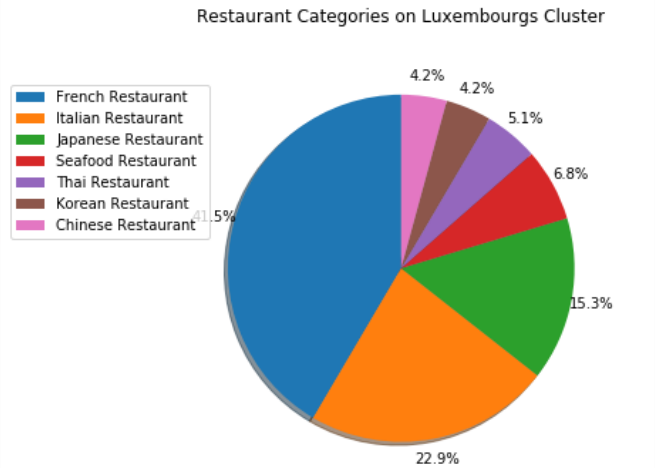
The Luxembourg cluster is made up as follows:

Neighborhood	Latitude	Longitude	AntiqueShops	Restaurants	Cluster
Luxembourg	48.850433	2.332951	13	42	3
Popincourt	48.858416	2.379703	5	42	3
Vaugirard	48.841370	2.300383	1	41	3
Passy	48.857505	2.280983	1	39	3

As we can see, Luxembourg is the best choice between his cluster

This top 7 restaurant categories in the cluster are:

French Restaurant	49
Italian Restaurant	27
Japanese Restaurant	18
Seafood Restaurant	8
Thai Restaurant	6
Korean Restaurant	5
Chinese Restaurant	5



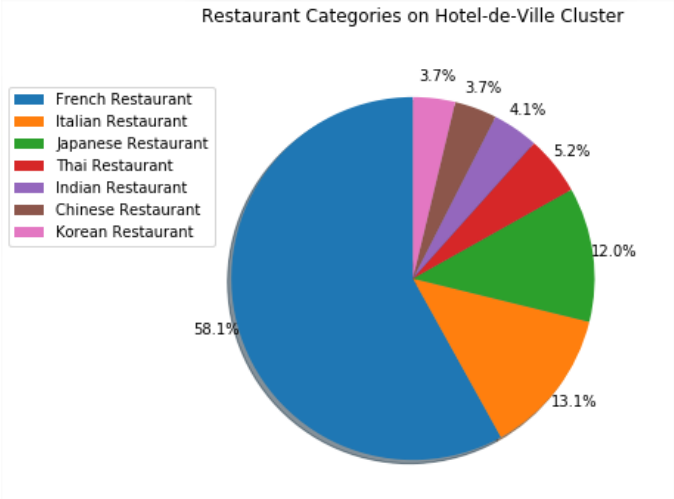
In the other hand, The Hôtel-de-Ville's Cluster is made up as follows:

Neighborhood	Latitude	Longitude	AntiqueShops	Restaurants	Cluster
Bourse	48.868630	2.341474	3	68	1
Temple	48.866500	2.360708	2	66	1
Hôtel-de-Ville	48.856426	2.352528	8	68	1
Panthéon	48.846191	2.346079	1	63	1
Opéra	48.870645	2.332330	3	65	1
Entrepôt	48.876106	2.359910	0	63	1

Note that, again, Hotel de Ville is the best choice between his own cluster

This top 7 restaurant categories in the cluster are:

French Restaurant	155
Italian Restaurant	35
Japanese Restaurant	32
Thai Restaurant	14
Indian Restaurant	11
Chinese Restaurant	10
Korean Restaurant	10



As we can see, Both Clusters offers similar types bus different proportions, in this case HotelDeVille's Cluster has more French and Regular Restaurants, and Luxembourg will be a more influenced by Italian Restaurants

7. Results

As a result we observe that although there is no relationship between the factors of Antique Shops and Restaurants, one can observe neighborhoods in which the concentration of antique shops is much greater than it could be, configuring an outliner that can be very advantageous for the company.

The grouping of neighbourhoods into clusters indicates that the Luxembourg neighbourhood is the one that best meets the above-mentioned factors, and has a cuisine offer similar to other similar neighbourhoods

8. Discussion

In this project we discovered that the distribution of antique shops in Paris is not uniform,.

This allows us to observe that two districts of Paris can be highlighted as high concentration of antique shops.

In Luxembourg, we have a high presence of restaurants, always above the average of Paris.

We finally discarded hotel-villa because in the comparison, it is always a priority for the company to associate the brand "Midnight" with sectors of antique shops, although in hotel-villa it is of greater flow of customers

Another important observation obtained is the fact of classifying the type of restaurants that will make up the competition, distinguishing mainly standard restaurants (and French food), secondly Italian food and thirdly Oriental food.

We trust that this observation will be useful for the company to establish priorities when configuring the menu and the thematic environment in order to attract customers.

9. Conclusion

For all the above reasons, we recommend Luxembourg neighborhood as the best option for the new branch of "Midnight in Paris"