Capstone Project
The Battle of Neighborhoods

Introduction

Paris, city and capital of France, is by far the country's most important center of commerce and culture. The 'City of Light' (la Ville Lumière), as commonly known, is one of the world's most important and attractive cities. It is particularly appreciated for its gastronomy which enjoys an enviable reputation.

The choice of the location for a restaurant is the starting point, but also one of the most crucial decision. The success of the business can depend on this unique parameter.

The objective of this analysis is to assist an entrepreneur wiling to open a new restaurant in Paris while being not familiar with it, in finding a good location, with regards to the type of venues nearby, the customer base, and the property prices. This data-driven recommender can be seen as a starting point for anyone who is struggling to find the perfect place for their business.

I. Analytic Approach

In order to determine the best location for opening a restaurant, we will first get the list of the neighborhoods and their geolocation, so that we can explore them using the Foursquare API and identify the different types of venues that are located nearby. We will then cluster the neighborhoods using k-means algorithm to better understand their particularities and identify which neighborhoods fit the most with a restaurant business.

We will then compare the neighborhoods using population data, so that we can shortlist the neighborhoods in which it would be relevant for opening a restaurant. For that we will determine a "target population" based on the age of our potential customer base. The assumptions on the "target population" are formulated later.

Our analysis will consider the median rent of the neighborhoods as it can be one of the main costs for a restaurant. We will make the hypothesis of a lease for this analysis.

In summary, we will use for our analysis the following data:

DATA	SOURCE
List of Paris neighborhoods	Data.gouv.fr, french platform of public data
List of venues in each neighborhood	Foursquare API
Population size and characteristics for each neighborhood	Data.gouv.fr, french platform of public data
Rent data of each neighborhood	Data.gouv.fr, french platform of public data

II. Methodology

The city of Paris is composed of 20 borough and 80 neighborhoods.
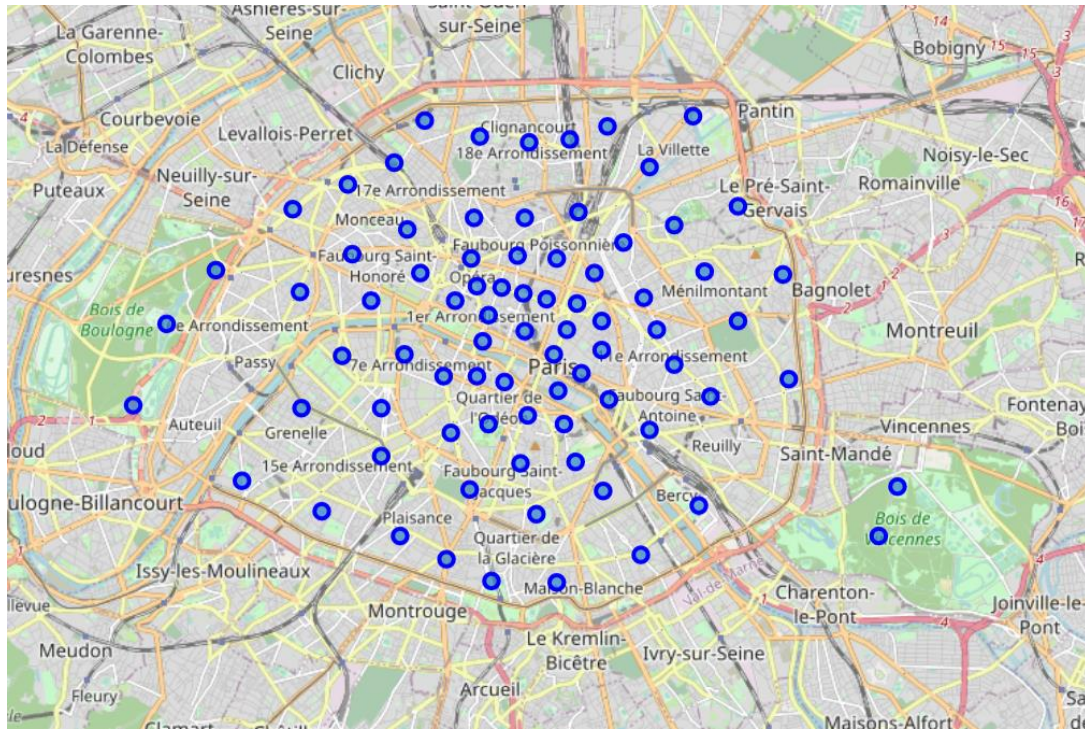
The city's population is quite young as we can see in the following figure, with the 25-39 years representing more than a quarter of the total population.


Population age structure of Paris
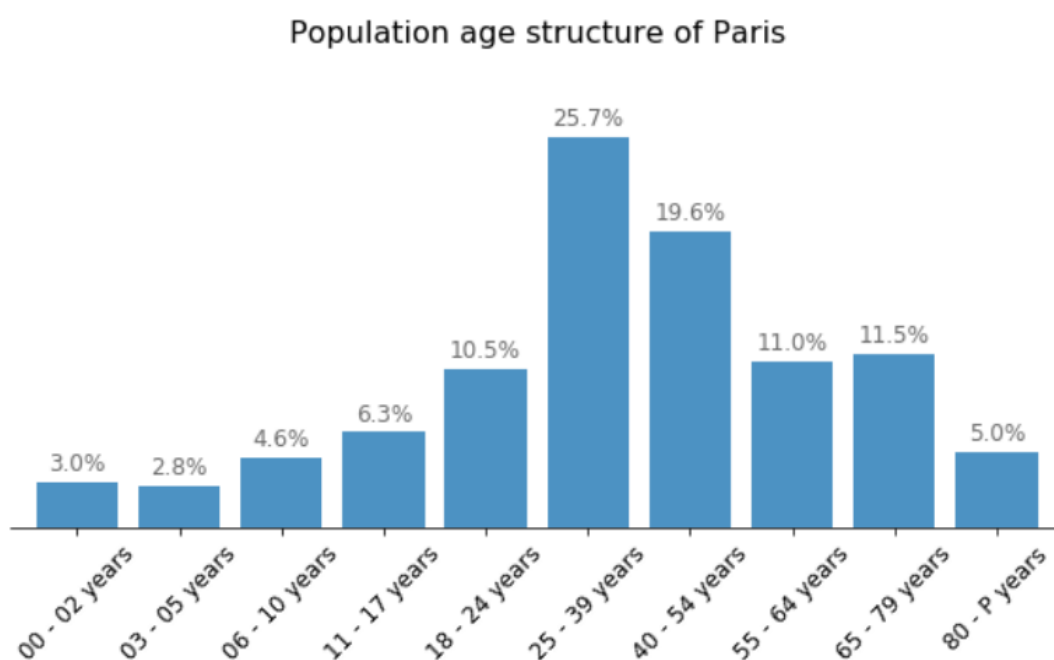
We calculated a target population based on the population ages, to reflect what customer base we are targeting for our business. This target population will be used later to compare the number of restaurants available per population in each neighborhood.

Below are the assumptions we made, which show that we are considering a restaurant concept that should attract a rather young population.

POPULATION SEGMENT	TARGET POPULATION PERCENTAGE
11-17 years	5 %
18-24 years	20 %
25-39 years	50 %
40-54 years	20 %

Using the Foursquare API, we explored each one of the neighborhoods to identify the different types of venues nearby. The objective of this approach is to be able to cluster the neighborhoods with regards to the type of venues that we find the most in it. in order to identify if there is a cluster that would be more appropriate to our business.
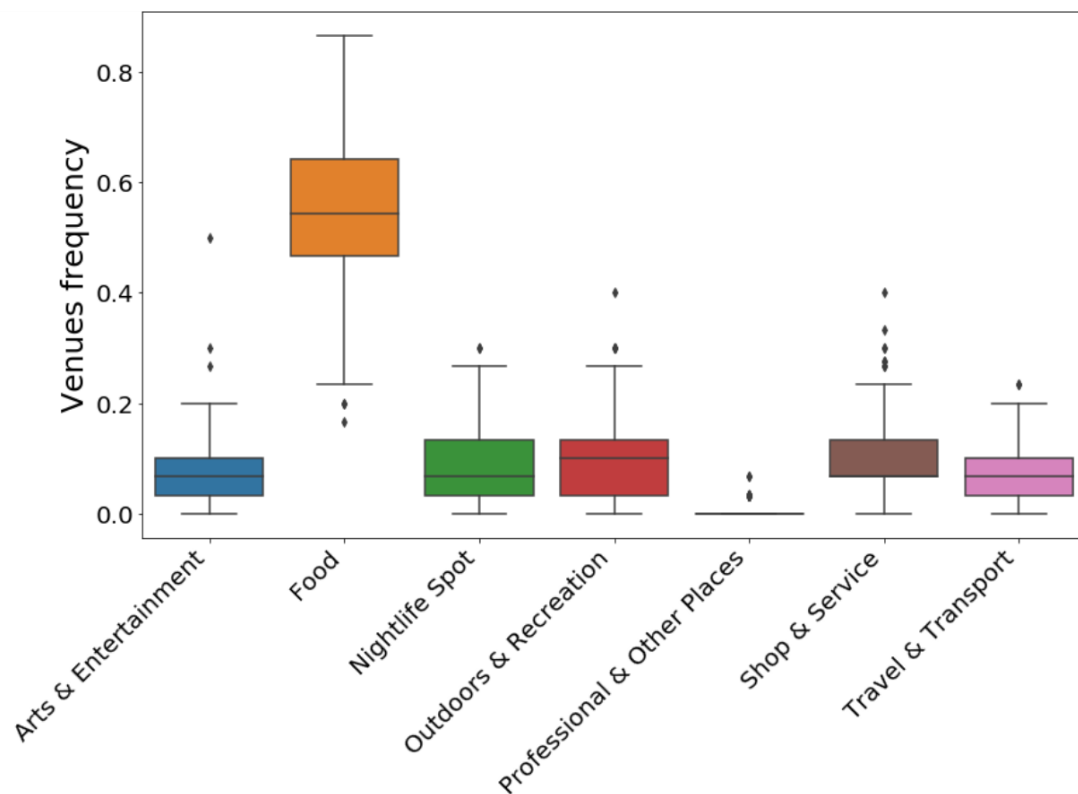
We can see that "Food" venues has a large place in Paris venues, as we could predict.

Before going through k-means clustering, we must choose the number of clusters that we want.

The elbow method consists on quantifying the sum of squared distances between the neighborhoods data, for different values of clusters k. As a result of the method, by increasing the number of clusters, the error will be mechanically reduced. But at some point, the gain from the additional cluster becomes insignificant, and adds more complexity than anything else.

Elbow Method to determine the optimal number of clusters k
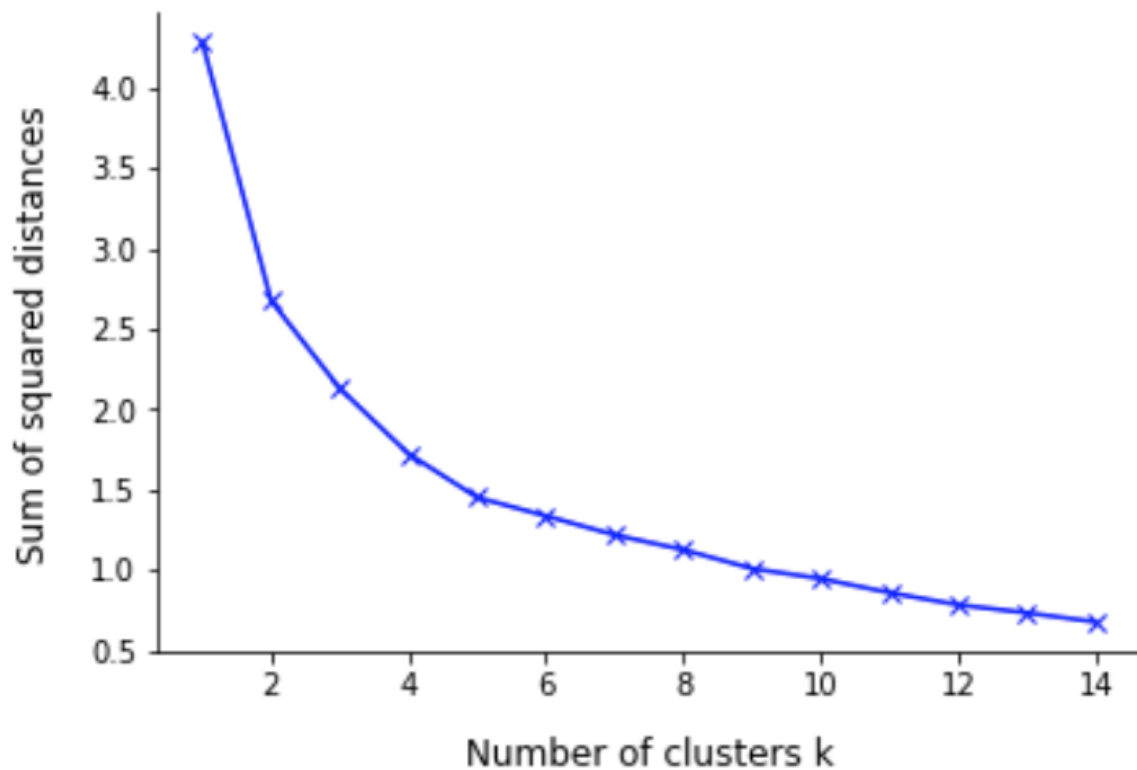
We can see that a suitable number of clusters would be between 3 and 5. After this value, the information gained by adding clusters is too marginal and will make our analysis too complex.
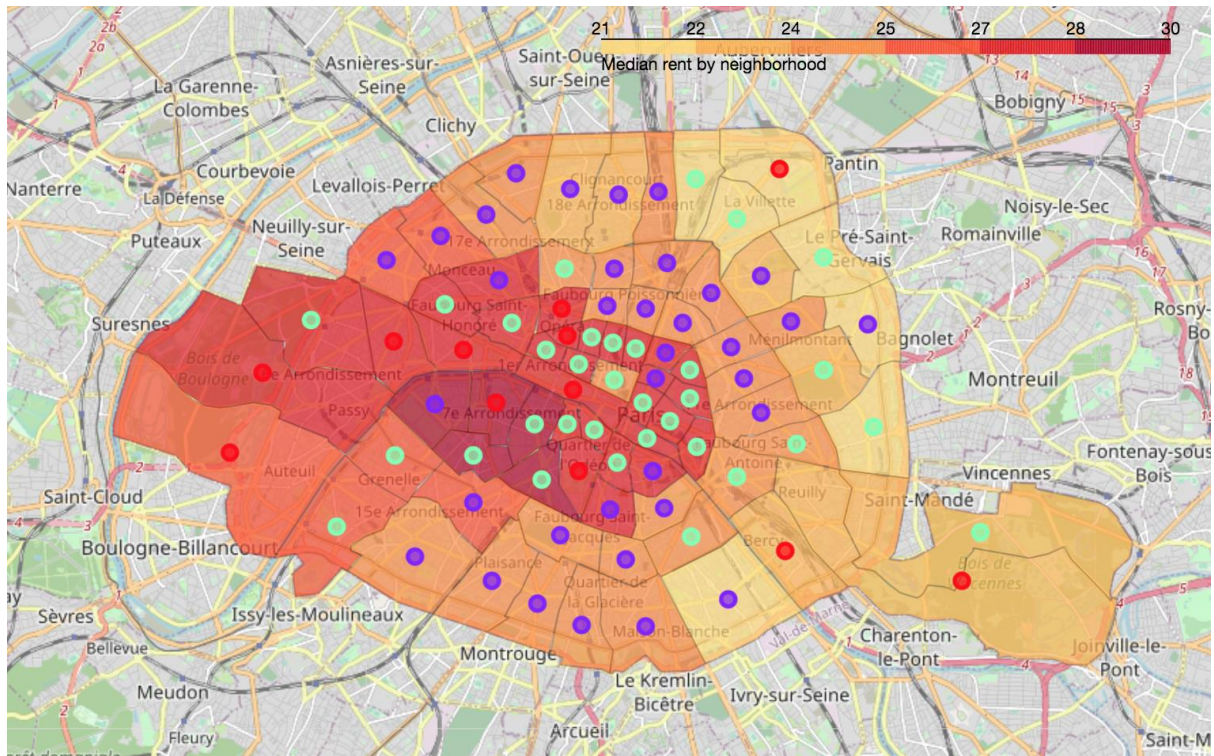
For the sake of our analysis, we will consider 3 clusters as we can understand and interpret them more easily.

In the next part, we will go through the results obtained by the clustering.

III. Results

The map below represents the different clusters, as well as the rent price by neighborhood.

- Cluster 0
- Cluster 1
- Cluster 2

We can see that most of the neighborhoods of cluster 1 are located whether in high rent price area, whether in the extreme sides of the city.

Neighborhoods of clusters 0 and 2 are more spread around the city.

Below is the result of the k-means clustering with 3 clusters.
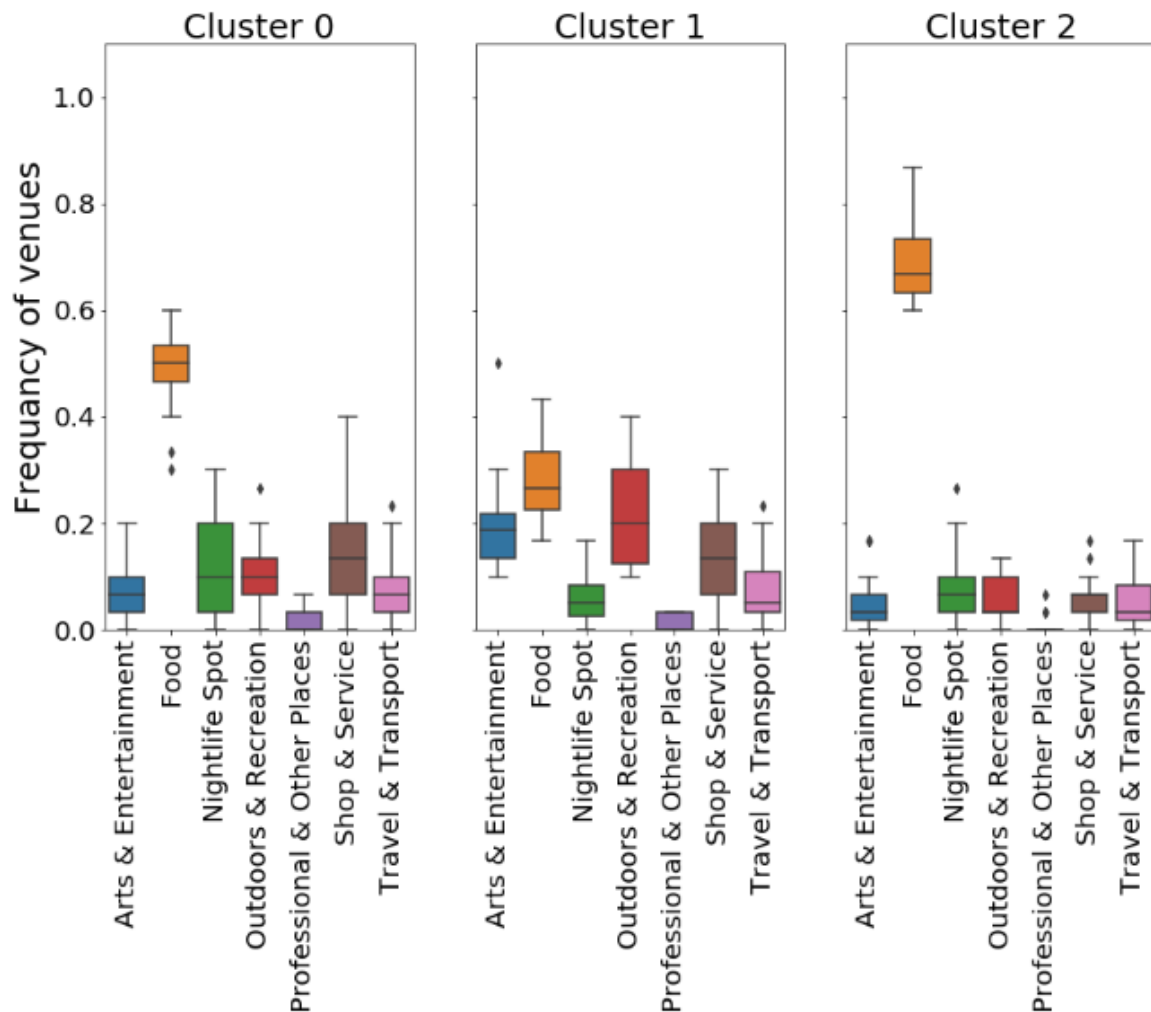
We can see that cluster 0 and 2 are dominated by "Food" venues, whereas the cluster 1 is more balanced. We can imagine that the neighborhoods of cluster 1 would be more like residential or cultural districts, with some restaurants in which the customer base would be for the vast majority the neighborhood community.

For this reason, we will discard cluster 1 from our analysis as it does not seem to be suited to our business.

In order to complete our analysis which is based only on absolute numbers of venues, hence not considering the size of each neighborhood, we compared their number of restaurants per population.

Number of restaurants per 1000 population
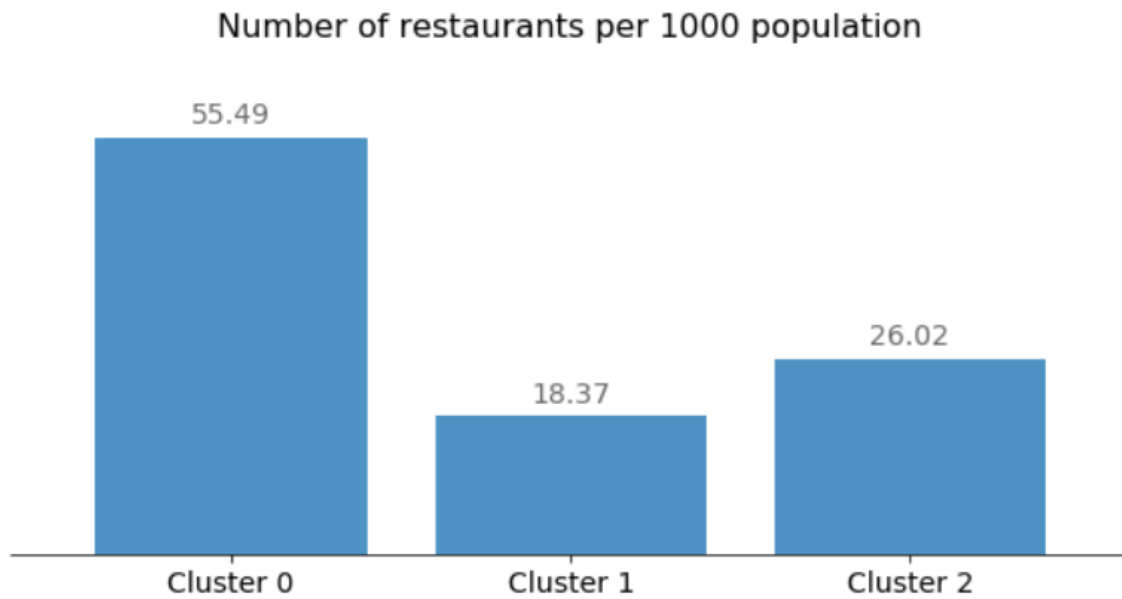
The results show that cluster 0 contains the largest number of restaurants per population. It would then be challenging to succeed in those neighborhoods considering the fierce competition we would face, especially if we want to open a new restaurant, and not just buy an existing one (in which case, this cluster would be a good option to consider).

The most secured option would then be neighborhoods in the cluster 2.

By selecting neighborhoods of cluster 2 that are located in areas where the median rent price is below 25€ per m^2, we could refine our list of potential neighborhoods.

Our last analysis was to compare them with regards to the number of restaurants per population, in order to exclude outliners.

Number of restaurants per 1000 population
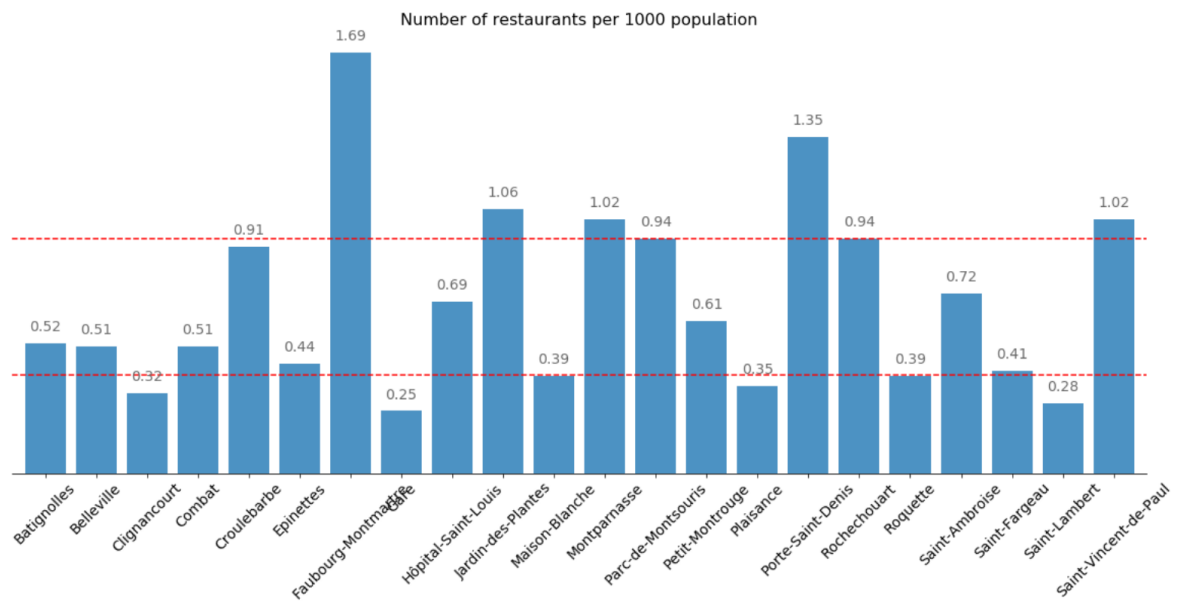
By keeping neighborhoods in which the number of restaurants lies between the 1st and 3rd quartile, we get our final list of potential neighborhoods for opening a restaurant in Paris.

Median rent by neighborhood
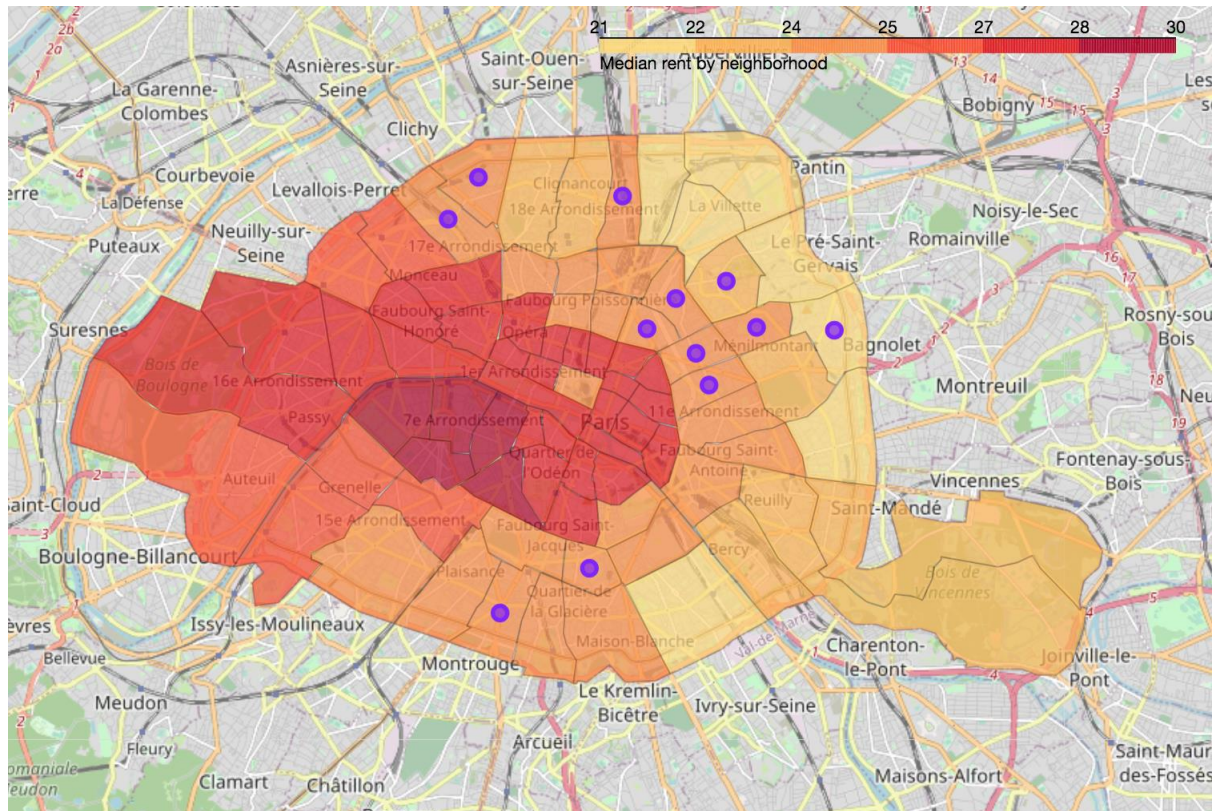
	Neighborhood
0	Batignolles
1	Belleville
2	Combat
3	Croulebarbe
4	Epinettes
5	Hôpital-Saint-Louis
6	Parc-de-Montsouris
7	Petit-Montrouge
8	Rochechouart
9	Saint-Ambroise
10	Saint-Fargeau

IV. Discussion

Our analysis recommends opening a restaurant in neighborhoods included in cluster 2 and we have been able to make a shortlist of 11 neighborhoods considering rent prices and population size.

But as we stated earlier, cluster 0 could have been a good opportunity for buying an existing restaurant. Indeed, the large number of restaurants per population shows that these neighborhoods are able to attract people from other neighborhoods.

One additional analysis that we could have done if we had the data would be to compare the pedestrian flow between neighborhoods. This

information would have been very important as it is a direct measure of the attractivity and the visibility of a neighborhood.

An observation that we can make is about the number of venues categorized "Professional & other places", which seems to be abnormally low.

Indeed, there is an important business district located just outside of the city, which explains that we don't have that many professional places inside the city.

But there should be nonetheless a significant number of them that the API should have returned. These missing data can be explained by the fact that people do not usually "check in" in these kind of places when using the Foursquare application, which results in small number of venues of this category returned by the application.

That being said, this missing information is not crucial to our project, as there is no real business district inside the city, but a rather scattered professional places throughout it.

Conclusion

Our analysis enabled us to select 11 neighborhoods (on a total of 80) that we consider to be the most suitable for opening a new restaurant in the city of Paris. This analysis has been conducted using different characteristics, such as the types of venues nearby, the population size and age, and the rent prices.

This analysis is not perfect as it neglects considerations of pedestrian flow which are key for a restaurant success. Nonetheless it is a good starting point for an entrepreneur who is not familiar of the city or who want to have a more data driven recommendation.