The battle of the neighborhoods

Best place to start a new Italian restaurant in Paris

Yosi Levy

May 05, 2020

Table of contents

1. Introduction

1.1 Background

Paris is the capital and most populous city of France, one of the greatest metropolises over the world. Paris Region has about 18 percent of the population of France. It was ranked as the second most visited travel destination in the world in 2018 and provide many business opportunities. Its economy prospers today thanks to a well-educated workforce, modern infrastructure, and global niches in creative industries, business services, and tourism. Paris is very attractive city and therefore has the most competitive market in France.

In order to find a new location to open a business there, it is important to provide business analysis. This analysis can provide a better understanding of the business environment, reduce risk and narrow down the search for a suitable location.

1.2 Problem description

Paris is famous for its restaurants and haute cuisine. Every French regional cuisine and almost every national cuisine in the world can be found there. Paris has in total 20 boroughs (called arrondissements in French) and 80 neighborhoods (4 neighborhoods for each borough). In order to find an optimal location for a restaurant, various factors need to be studied:

- Italian restaurants in the vicinity
- Total restaurants in the vicinity
- Paris population
- Median household income for the neighborhoods
- Average rent in the area
- Area popularity
- Distance from Paris center

The analysis and recommendations for a location for a new Italian restaurant will focus on general neighborhoods with their establishments, not on specific addresses. Narrowing down the best district options derived from the analysis will allow for either further research to be conducted, advising agents of the chosen district, or on the ground searching for specific sites by the stakeholders.

2. Data

This project focuses on segmenting and clustering the neighborhoods in Paris. Based on the definition of our problem, the factors that will influence our decision and the sources from which the relevant data is are as follows:

- Number of existing restaurants in the neighborhood, type and location was obtained using **Foursquare API**
- Distance of the neighborhood from the city center (Paris center was obtained with GeoPy library)
- Data set of Paris quartier was obtained from **[1]**
- Data set of Paris arrondissements was obtained from **[2]**
- The relevant data of Paris population and median household income for each arrondissements was obtained from **[3]**
- Data set of Average rent in each arrondissements was obtained from **[4]**
- The rank that was given for each arrondissements based on the data was obtained from **[5]**

3. Methodology

In the project presented here we try to determine the suitable areas in which to open an Italian restaurant in Paris, France. Paris is divided to 20 administrative districts, more simply referred to as arrondissements. Every arrondissement is divided to 4 quarters. Each arrondissement has some distinctive and unique features. Therefore, in order to decide upon the best area in which to open an Italian restaurant, we will need to perform a segmentation of the data available on each arrondissement, while taking into account all the specific criteria that are of interest to the stakeholders.

In order to achieve this goal, we first will collect the data of Paris neighborhoods locations, and perform data wrangling for further analysis. Then we will use the Foursquare server in order to find all the restaurants and their type within a 750 meters radius in each neighborhood. After we will identify all the Italian restaurants, we will explore the density of the restaurants in general and of the Italian restaurants in particular in the different arrondissements, by creating Heat maps.

The second step in our analysis will be to gather more information regarding the different arrondissements in Paris, in order to choose the most promising location. We will collect data about population density, average annual household income, average rent prices, and general score of the various arrondissements. The data will be used at a later stage in the segmentation process.

The third and final step will focus on segmentation of the quarters in Paris. This will be achieved by clustering all the data we obtained, using the K-means algorithm, and creating clusters that are comprised of the quarters. K-Means algorithm is one of the most common clustering method of unsupervised machine learning algorithms. We will also present the map of the clustered neighborhoods, using GeoPy and Folium libraries.

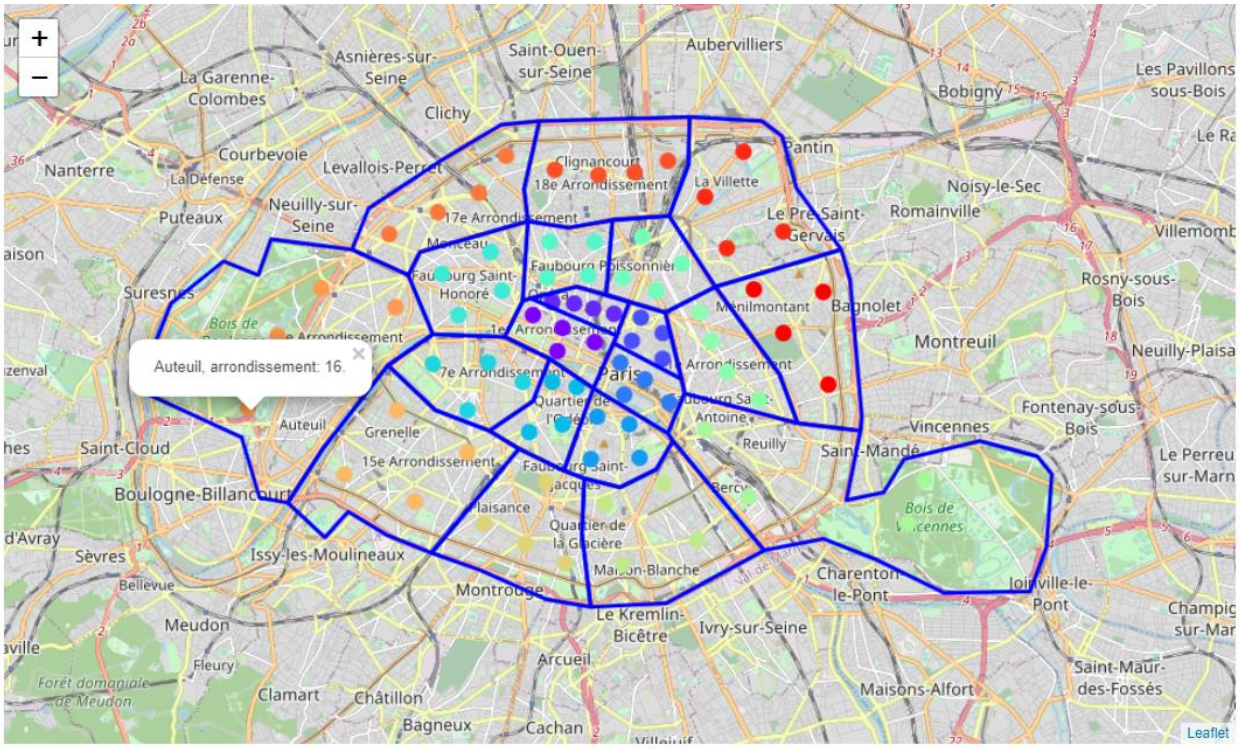
3.1 Exploratory Data Analysis

3.1.1 Boroughs and neighbourhood in Paris

Here is the pandas data frame containing the neighborhoods that exist in each borough of Paris, as well as the latitude and longitude coordinates of each neighborhood.

	quartier_number	quartier_name	arrondissement_number	Latitude	Longitude
0	27	Ecole-Militaire	7	48.850359	2.311031
1	59	Grenelle	15	48.850172	2.291853
2	61	Auteuil	16	48.850622	2.252277
3	12	Sainte-Avoie	3	48.862557	2.354852
4	40	Hôpital-Saint-Louis	10	48.876008	2.368123

In order to visualize geographic details one can use python folium library. Below is the map of Paris with marks for each neighborhood where the name of the neighborhood and the number of borough where the neighborhood is located are superimposed on top.



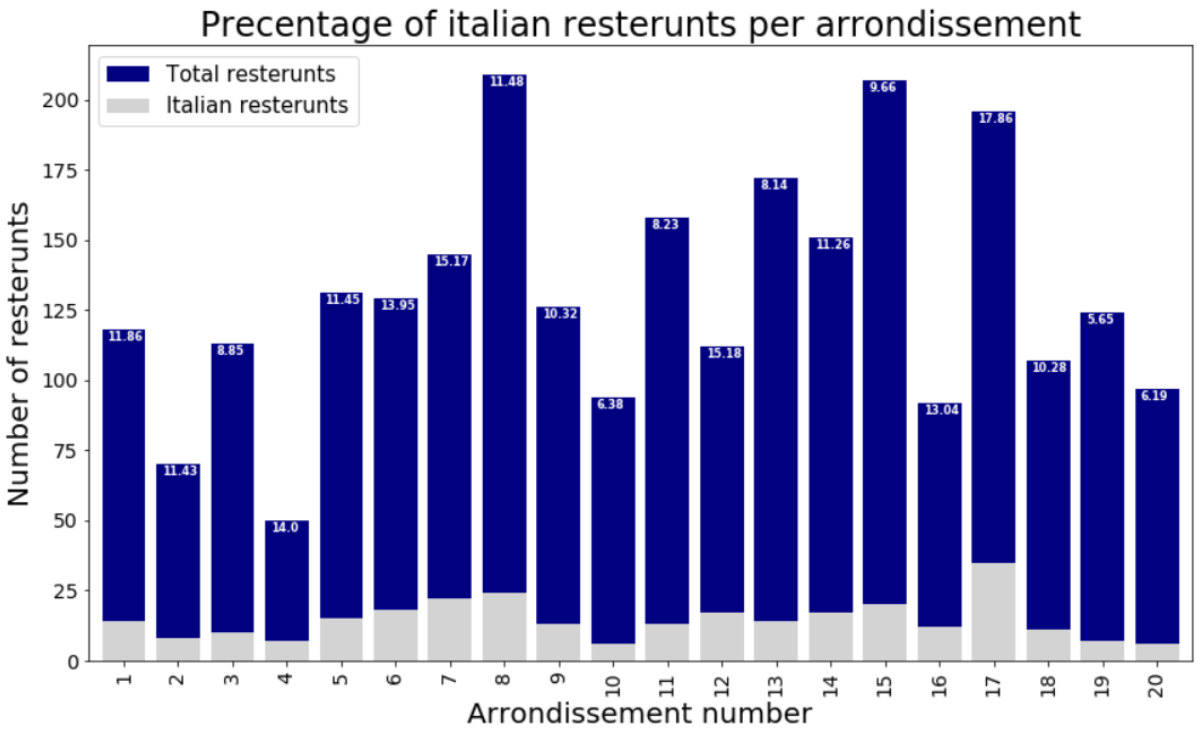
3.1.2 Restaurants in Paris

The dataset, that will be used to get info on the restaurants in each neighborhood of Paris, will be provided from Foursquare API. Foursquare API provides location based experiences with

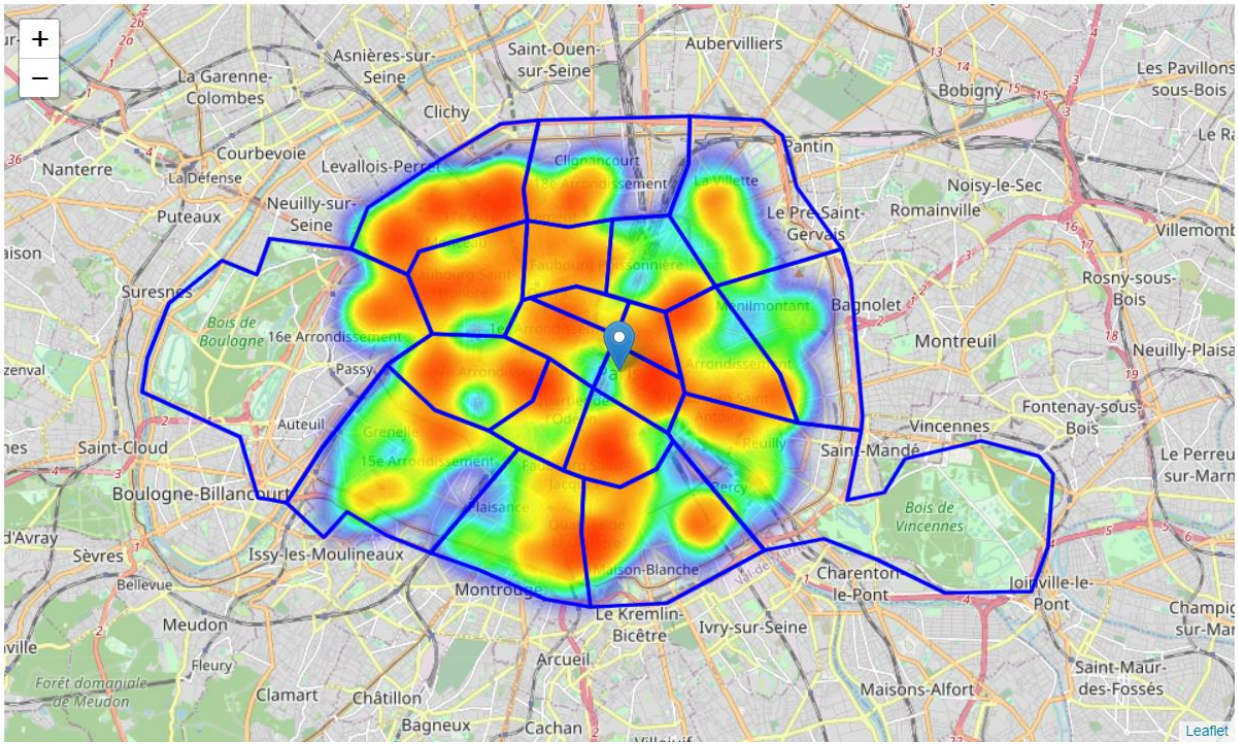
diverse information about venues, users, photos, and check-ins. Here is the head of the list obtained by from Foursquare API, which includes the information of the venues names and category, latitude and longitude, etc.

	quartier_name	arrondissement_number	venue_id	venue_name	venue_lat	venue_lng	venue_categories	is_italian
0	Necker	15	4bb62124f562ef3ba7a92f97	L'Antre Amis	48.8467	2.31022	[(French Restaurant, 4bf58dd8d48988d10c941735)]	False
1	Ecole-Militaire	7	4d714543a8df6dcb224030c1	Lily Wang	48.85	2.31242	[(Asian Restaurant, 4bf58dd8d48988d142941735)]	False
2	Ecole-Militaire	7	4b7261eaf964a5205a792de3	D'Chez Eux	48.8535	2.30951	[(French Restaurant, 4bf58dd8d48988d10c941735)]	False
3	Ecole-Militaire	7	5612e7b7498e3e3b22d91e5e	Escudella	48.8507	2.30897	[(French Restaurant, 4bf58dd8d48988d10c941735)]	False
4	Necker	15	55e09d6d498ebe455fee5d7d	Le Radis Beurre	48.8463	2.30771	[(French Restaurant, 4bf58dd8d48988d10c941735)]	False

Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's see the ratio between the number of Italian restaurants and the total number of restaurants in each arrondissement. We can visualize the percentage (%) of Italian restaurants in each arrondissement using a bar plot graph.



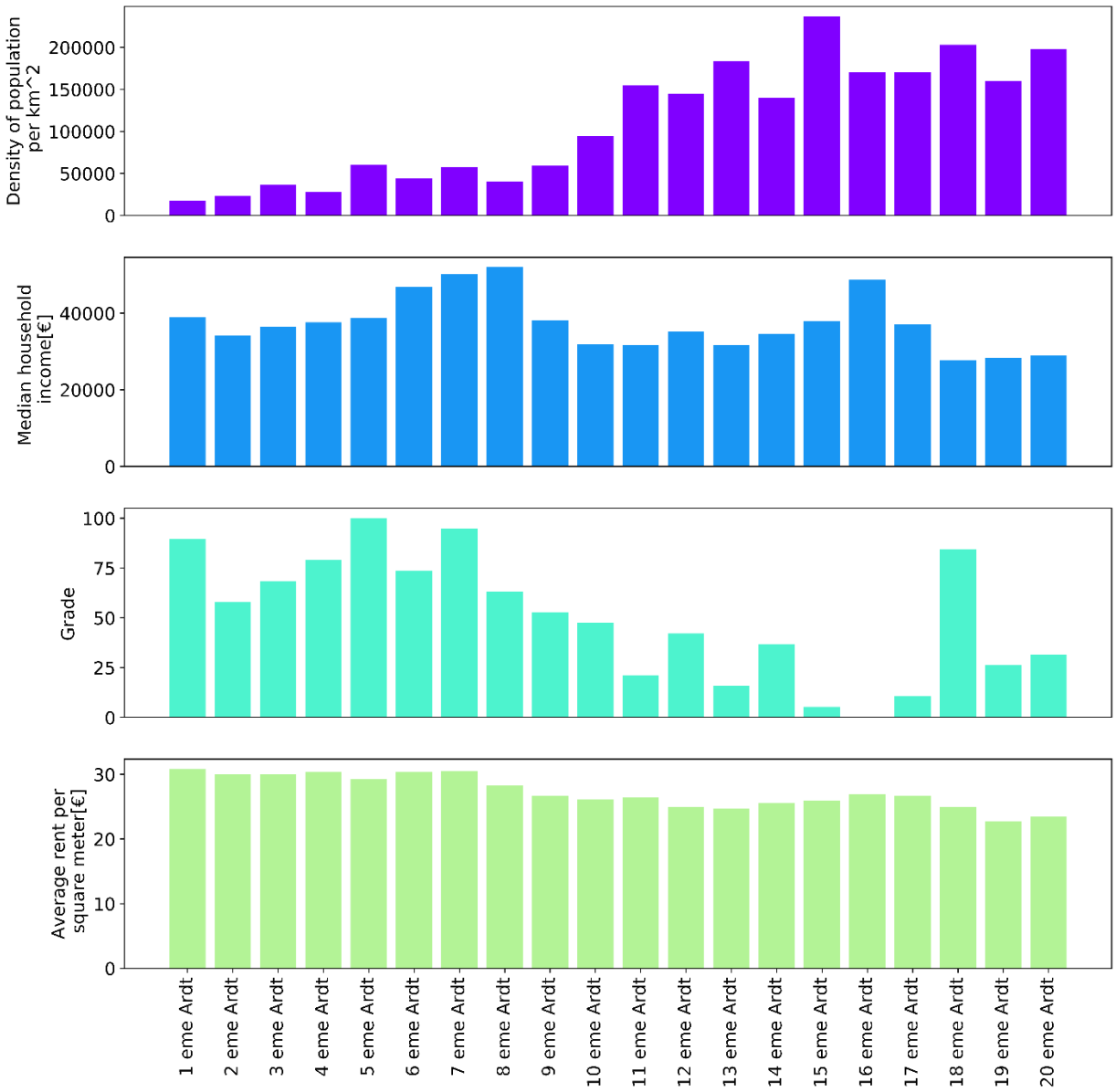
The density of Italian restaurants can be visualized by showing a heat map/density. This map can give some meaningful information about the general distribution of Italian restaurants on Paris, and detect locations that are not already crowded with restaurants.



This map is not so 'hot' (Italian restaurants represent a subset of ~11% of all restaurants in Paris) but it also indicates higher density of existing Italian restaurants directly north and west from Paris center, as well as east-west and west-south. These heat maps can help us choose the best quarters for opening the restaurant after we apply the segmentation of all the quarters.

3.1.3 Socio economic criteria

As mentioned above, information regarding the different arrondissements in Paris should be gathered in order to segment the different neighborhoods of Paris and to find the most promising location. The additional data includes the population density, average annual household income, average rent prices, and general score of the various arrondissements. Let's summarize the data in bar plots.

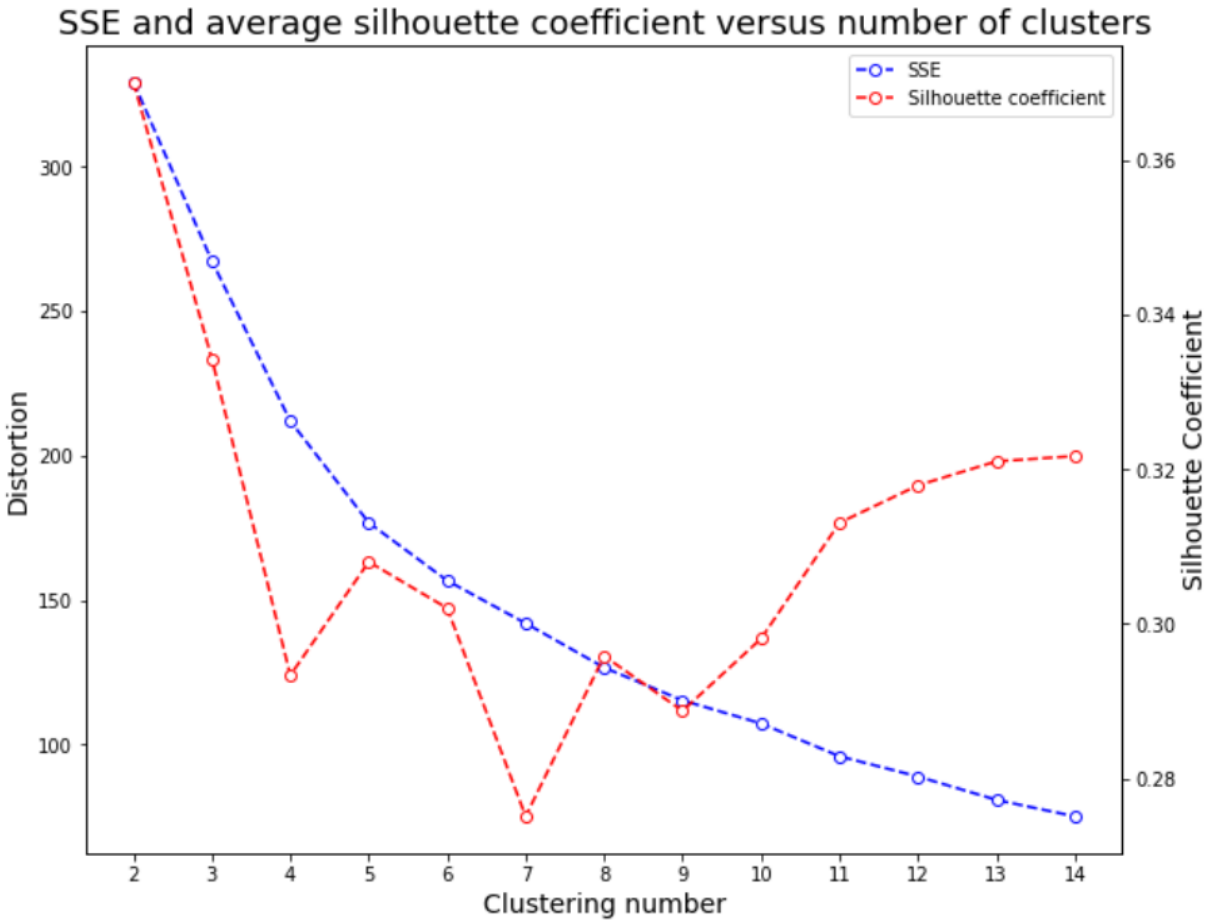


We can observe from these graphs that there is a big variance in the first three parameters, namely, Density of population, Median household income and Grade, between the different arrandosments. In contrast, the Average rent per square meter does not vary significantly between the different arrondissement. We can support this claim by calculating the standard deviation (std) for each parameter.

```
Density of population (mean,std): (111075.00,71572.46)
Median household income (mean,std): (37285.50,7130.57)
Grade (mean,std): (50.00,31.14)
Average rent per square meter (mean,std): (27.23,2.55)
```

3.2 Machine learning

The data mentioned above was collected and then analyzed in order to perform a segmentation of all the 80 neighborhoods of Paris, using the K-means clustering algorithm. By using the elbow and average silhouette methods we have found that the optimal value of K (number of clusters) is 5.



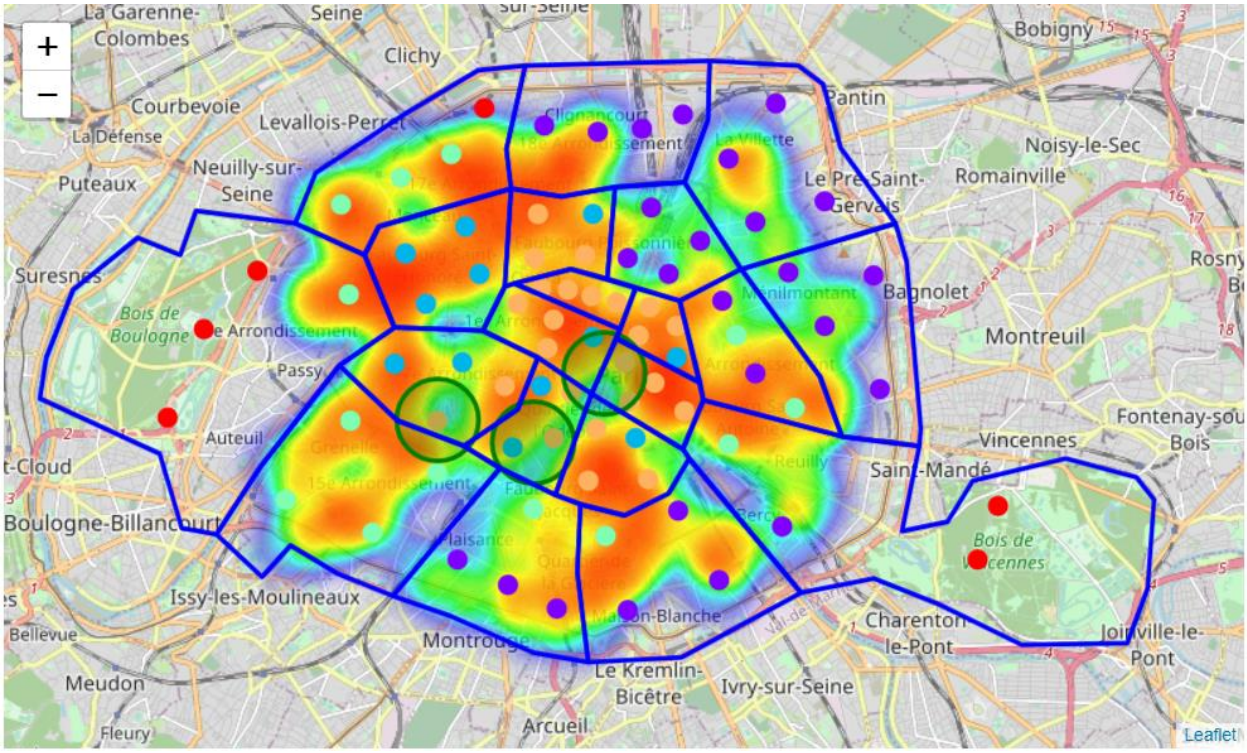
Following the distribution of the neighborhoods to the different clusters, we created a profile for each cluster and chose the most suitable one.

Label	Population density	Income	Rent	Grade	Italian resteruns	Total resteruns	dis
0	161820.000000	30552.400000	24.696000	40.000800	2.320000	28.040000	3403.638000
1	44891.666667	46130.000000	29.316667	75.000000	7.166667	60.000000	2382.900000
2	185114.615385	36583.076923	26.046154	14.574615	7.538462	59.076923	3781.183846
3	38570.833333	39016.666667	29.741667	75.656667	2.045455	16.863636	1462.975833
4	161601.666667	42248.333333	26.200000	15.791667	0.400000	8.200000	6276.915000

It can be observed from the data that clusters 1 and 3 meets most of the criteria. We can label these clusters as follows:

- Cluster 1: The neighborhoods in this cluster are characterized by a **close proximity to the city center**, high **grade**, and high **average household income**. However, these neighborhoods also have a low **population density** and relatively high **number of Italian restaurants**.
- Cluster 3: The neighborhoods in this cluster are characterized by a very **close proximity to the city center**, high **grade**, a relatively low **number of Italian restaurants**, and high **average household income**. However, these neighborhoods also have a very low **population density**.

Finally, let's visualize the resulting clusters on the heat map/density of Italian restaurants. We also created 3 representative zones containing neighborhoods with low number of Italian restaurants, which are fairly close to the city center and belong to clusters 1 and 3. These zones encompass an area included by a radius of ~650 meters (green circles) and their centers should be considered only as a starting point to the stakeholders for exploring potential neighborhoods.



4. Results and Discussion

Our analysis was comprised of a segmentation of the 80 neighborhoods in Paris. The segmentation was performed based on different parameters which are of importance when deciding upon the best location to open an Italian restaurant. The criteria that are of interest to the stakeholders, and thus are important when perfuming the segmentation, are as follows:

- An area with low density of restaurants in general, and of Italian restaurants in particular, in order to reduce the competition with active restaurants.
- An area with high population density, so the restaurant will be widely accessible.
- An area with higher-income neighborhoods, where income level might allow for a greater amount of spending than in lower-income neighborhoods.
- An area with Low rental prices on average, in order to reduce the routine monthly expanses.
- A popular tourist area, in order to increase sales.

All the data was collected and then analyzed in order to perform a segmentation of all the 80 neighborhoods of Paris, using the K-means clustering algorithm. By using the elbow and average silhouette methods we have found that the optimal value of K (number of clusters) is 5. Following the distribution of the neighborhoods to the different clusters, we created a profile for each cluster and chose the most suitable one. The promising clusters, which met most of the criteria, were clusters 1 and 3. These clusters were labeled as follows:

- Cluster 1: The neighborhoods in this cluster are characterized by a close proximity to the city center, high grade, and high average household income. However, these neighborhoods also have a low population density and relatively high number of Italian restaurants.
- Cluster 3: The neighborhoods in this cluster are characterized by a very close proximity to the city center, high grade, a relatively low number of Italian restaurants, and high average household income. However, these neighborhoods also have a very low population density.

In both clusters the average rent is similar, and thus it had no effect the on the decision about the best cluster.

Using the Clustered and Heat maps, we concluded that the following neighborhoods in the city of Paris represent a good starting point for opening an Italian restaurant, and can be of interest to the stakeholders: **Notre-Dame-des-Champs, Halles** (part of cluster 1) and **Ecole-Militaire, Monnaie, Odéon, Saint-Merri, Notre-Dame** (part of cluster 3). These neighborhoods are located

in areas with the lowest density of Italian restaurants and are fairly close to the city center. This of course does not imply that these neighborhoods are actually the optimal locations to open a new restaurant. The purpose of the analysis performed here was to narrow the search down to just a few possible locations that are best suit to match the criteria that was decided upon by the stakeholders. Recommended zones should therefore be considered only as a starting point for more detailed analysis, which should take into consideration more relevant factors.

5. Conclusions

The purpose of this project was to identify the most suitable areas in Paris to open an Italian restaurant, based on pre-selected criteria, in order to aid stakeholders in narrowing down the search for optimal location for a new restaurant. Clustering of the different neighborhoods was performed by calculating the Italian restaurant density, and by collecting relevant data such as the average income, population density and average rent rates. Analysis of the results showed two promising clusters which met most of the criteria.

Final decision on an optimal restaurant location will be made by stakeholders based on the weight they chose to give to each specific criterion of the neighborhoods, while taking into consideration additional factors like attractiveness of each location (proximity to popular landmarks), levels of noise/proximity to major roads, public transportation availability etc.

6. Sources

[1] [Online]. Available: https://opendata.paris.fr/explore/dataset/quartier_paris/information/.

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

[3] [Online]. Available: https://frenchmoments.eu/arrondissements-of-paris/.

[4] [Online]. Available: https://www.statista.com/statistics/769062/rent-the-metre-square-apartments-by-districts-paris-la-france/.

[5] [Online]. Available: https://theearfultower.com/2018/06/06/the-top-paris-arrondissements-for-tourists-ranked-from-worst-to-best/.