August 2021

# Analysis of the location of eating places in Paris

A Capstone project

Ian SOBRAL

# Contents

# 1 Introduction

Paris has about 2.2 million inhabitants for a surface area of 105 km2, reaching a density of 20,754 inhabitants per km2. Which is a lot, comparing to other cities like New York (10,716.36/km2) or Tokyo (6,373 inhabitants/km2). Not for nothing is it one of the densest cities in Europe.

With such a high population density, one can imagine that there is also a very large number of restaurants in a small area. For those who are interested, it may therefore be difficult to decide on a location to open a new eating place in Paris.

With this in mind, the objective of this project is to find an optimal location for a stakeholder interested in opening a eating place (restaurant, café, fast food, etc.) in Paris and also to find out which categories of eating places are the most common in Paris and where they are located.

In order to do this analysis, we will use data science and machine learning techniques to decide the best spots and the restaurants to avoid if the investor wants to open a new business.

## 2 Data description

The data used in the project were taken from the following sources:

- The coordinates of the borders of Paris were obtained from the "Paris Data" website [1]. A geojson file containing the coordinates of the edges of Paris has been downloaded from this website to avoid taking information from outside Paris.
- The informations about the coordinates systems used in the coordinate conversion functions were obtained from the open-source epsg.io web service [2].
- The data on eating places (addresses, coordinates, name of the establishment, etc.) was obtained from the Foursquare API [3].
- The latitude and longitude of Paris center were retrieved using Nomitatim API [4].

# 3 Methodology

The Python programming language with its libraries and the APIs presented on the previous section were used for the data extraction, generation, cleaning and analysis.

We will end this section by applying the "K-means" method (a machine learning unsupervised learning clustering algorithm) to find the best places to open a restaurant at Paris.

## 3.1 Python libraries and packages used for the data extraction and preparation

The Python libraries and packages that were used on this project for the data extraction and preparation were :

- NumPy : Library intended to manipulate matrices or multidimensional arrays and also mathematical functions applied on them.
- Json : Package used to manipulate (open, read and write) json files.
- Pandas : Library for data manipulation and analysis.
- Matplotlib : Library for plotting and data visualisation.
- Requests : Library used for making HTTP requests.
- Folium : This package is an interface between python and the leaflet library and it is used to data visualisation on the world map.
- Math : Standard module that provides access to mathematical functions.
- Sklearn : Library featuring various machine learning algorithms.
- KMeans from sklearn.cluster : A type of clustering algorithm.
- Geopy : Python client used for geocoding.
- Nominatim from geopy.geocoders : A tool used to retrieve coordinates from an addres and vice versa .
- Pyproj : Library for coordinate transformations library and cartographic projections.
- CRS from pyproj: A coordinate reference system manager.
- Transformer from pyproj: Tool used for coordinate transformation.
- Shapely : Package for computational geometry.
- Point from shapely.geometry : Creates a shapely point object.
- Polygon from shapely.geometry : Creates a shapely polygon object.
- Cascaded Union from shapely.ops : Combines shapely geometry objects

## 3.2 Setting the boundaries of Paris

First, we downloaded the data of the limits of Paris as a geojson file from the "Paris Data" website [1].

The city of Paris is divided in twenty districts called "arrondissements". From the downloaded geojson file, we have the coordinates and information for those 20 arrondissements.

To manipulate the data inside of this file, 2 dictionaries were created :

- paris_ar_geom : Coordinates of the arrondissements' borders.
- paris_ar_prop : "Properties" of the arrondissements. For example, the arrondissements' surface, coordinate center, name and others.

For example, we can display the properties of the 13th arrondissement of Paris. By doing : paris_ar_prop[13], we obtain :


```
{'n_sq_co': 750001537,
 'perimetre': 11546.5465264,
 'l_ar': '13ème Ardt',
 'surface': 7149311.09107136,
 'geom_x_y': [48.8283880317, 2.36227244042],
 'n_sq_ar': 750000013,
 'l_aroff': 'Gobelins',
 'c_arinsee': 75113,
 'c_ar': 13}
```


These properties have been taken directly from the geojson file. Not all of them were used in the project.

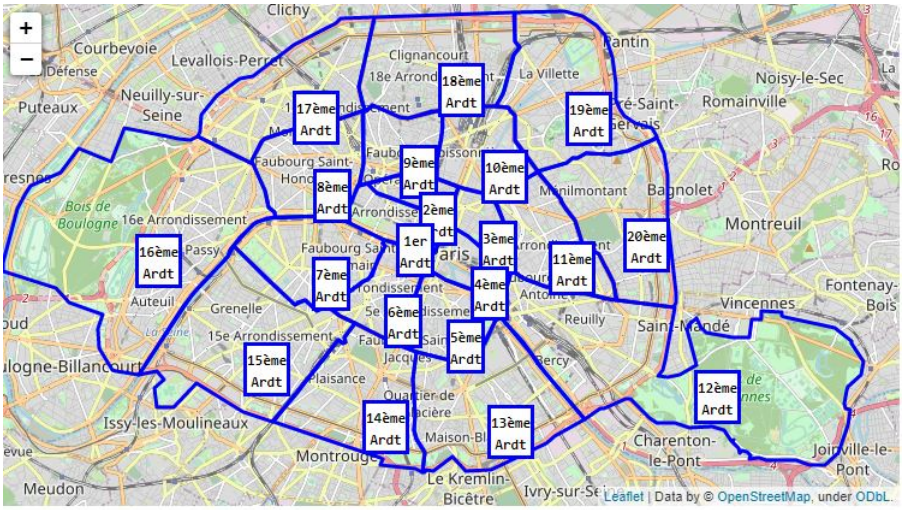
Using the nominatim geocoder to gather the latitude and longitude of Paris city center, we have :

Coordinates of Paris, France :

- Latitude : 48.8566969
- Longitude : 2.3514616

These coordinates will be used with the folium package to center the map visualization in the middle of Paris.

Using the dictionaries paris_ar_geom and paris_ar_prop with folium, we can visualise the arrondissements placement on the map :




Using those same dictionaries, we can also display only one arrondissement at a time (the 13th arrondissement for example):


Now that we have extracted the coordinates of the borders of Paris from the geojson file, we can use the shapely package to create Polygon objects for each arrondissemnt. The objective of creating those Polygon objects is to merge all of them with the "cascaded_union()" method in order to create a unique Polygon for the entire Paris area and to finally extract only the coordinates of the borders of Paris. We did this because the geojson file did not provide the coordinates of the Paris borders only.

In this way, we can display the outer limits of Paris and not only the limits of each arrondissement :


## 3.3 How to know if a point is inside or outside the boundaries of Paris

The geojson files were used to the visualization of the borders of Paris and its arrondissemts. The Polygon objects were used to merge all the arrondissements of Paris in order to extract only the coordinates of the borders of Paris.

Another utility of the Polygon objects is to know if a coordinate point is inside or outside Paris or one of its arrondissements.

To exemplify this, we have created two distinct Point objects :

- point1 : is the point of the center of the 13th arrondissement
- point2 : is the point of the center of Paris

Using the "contains" method, we can check if those points are inside the 13th arrondissement. For example, doing "paris_ar_polygon[13].contains(point1)" we obtain "True" and doing "paris_ar_polygon[13].contains(point2)" we obtain "False". This means that the point 1 is inside the 13th arrondissement and the point 2 is not. With these objects, we will be able to know if, for example, from which arrondissement a restaurant belongs to.

## 3.4 Coordinate systems conversion functions

To do some mathematical operations on a map, we have to convert the coordinates from WSG84 ( World Geodetic System 1984) : the coordinate system used in GPS (Global Positioning System) to UTM (zone 31) (Universal Transverse Mercator) : a 2D cartesian, coordinate system.

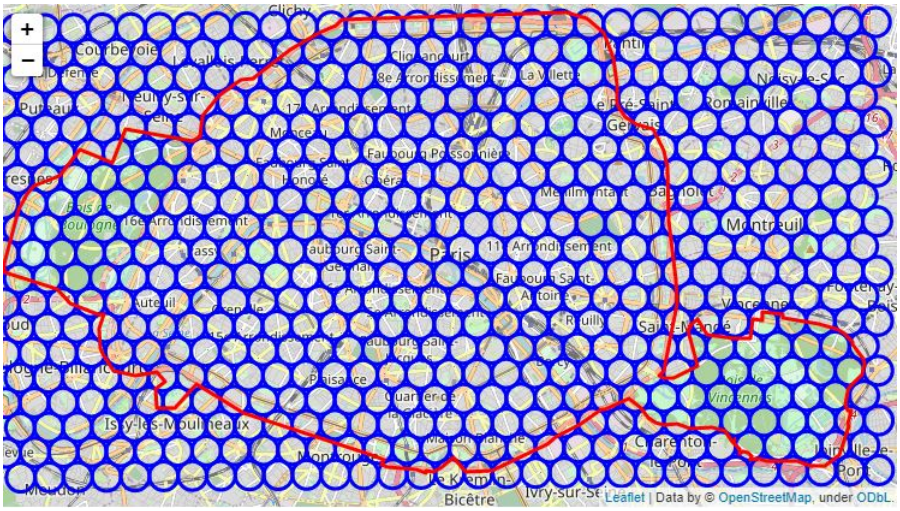
Two function were created to deal with these conversions with the Pyproj library :

- gps_to_utm(lat,lon) : Converts a point on a WSG84 coordinate system to a UTM coordinate system.
- utm_to_gps(x, y) : Converts a point on a UTM coordinate system to a WSG84 coordinate system.

## 3.5 Foursquare API

The Foursquare API works on a specific coordinate point and gathers locations within a specific radial distance from that point.

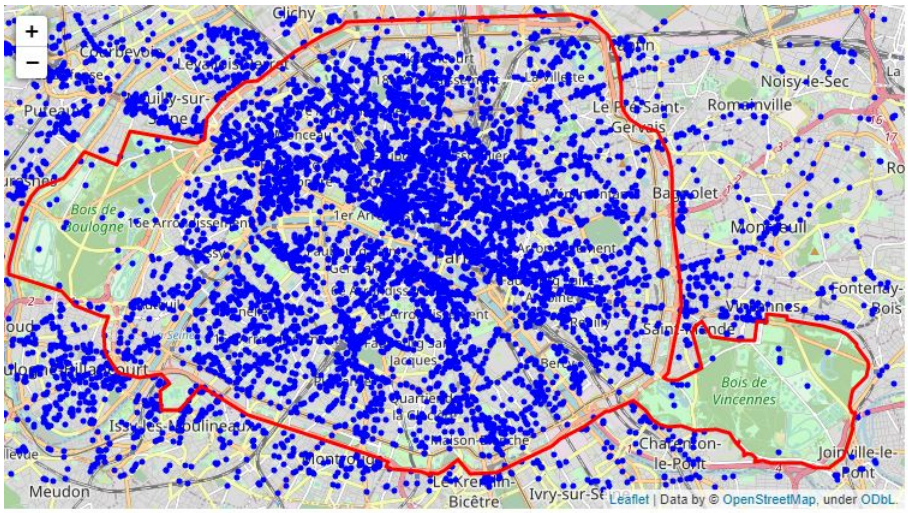
In this way, we will need several points to take the maximum number of venues at Paris. To do so, a hexagonal grid has been placed on the map of Paris :




Each circle has a diameter of 600 meters.

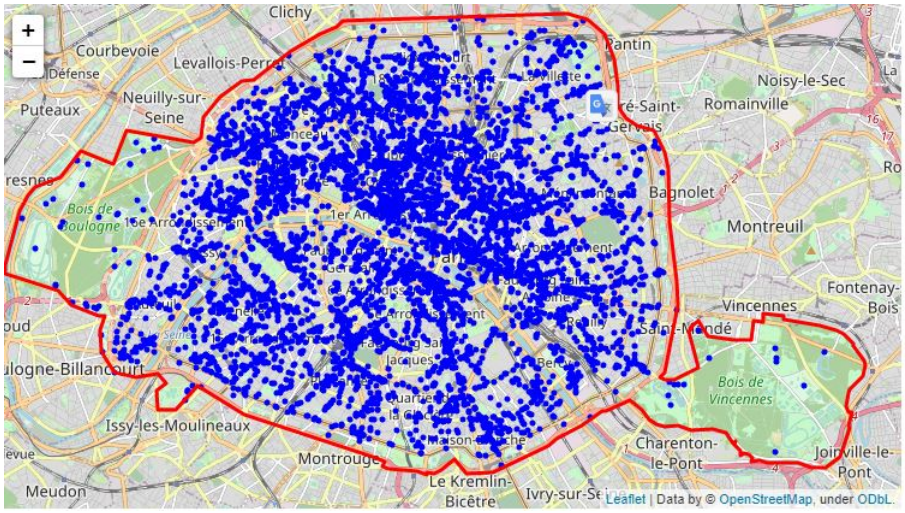
We can now use the Foursquare API at the center of each of these points to retrieve the desired eating places locations.

Plotting these locations on the map, we have :




From the Foursquare API, we retrieved more than 9000 venues. However, some of these venues are out of Paris. This is normal, since the hexagonal grid was extended beyond the borders of Paris. We collected the locations like this so as not to lose any information around the borders of Paris.

As we are only interested on the eating places that are in inside Paris, we can remove the points that we are not interested using the methods presented on the section 3.3. Filtering only the results that are inside Paris and displaying them on the map, we have :




There are 6215 restaurants detected by Foursquare inside Paris.

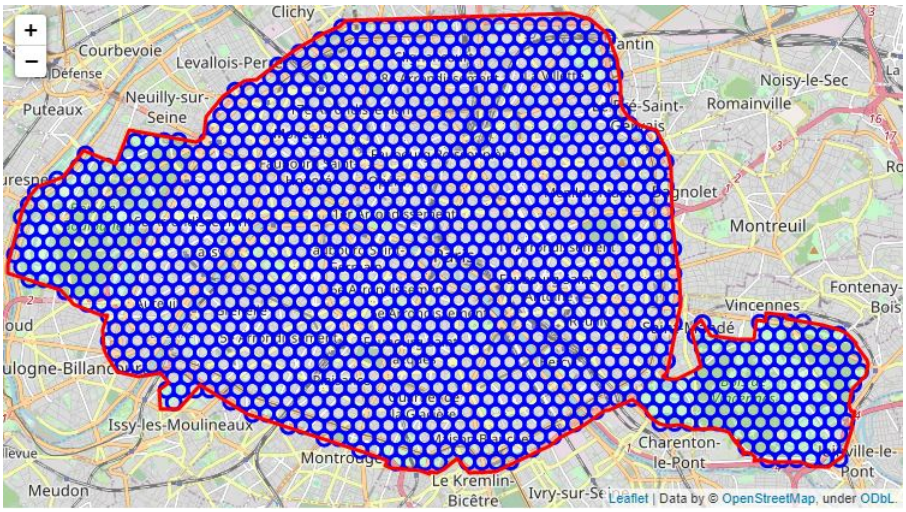
We can now create a data frame with only the restaurants inside Paris. Displaying the firt 5 rows of this data frame :

| | venue_id | venue_name | categorie_id | categorie_name | address | lat | lon | distance | ardt |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 5b8a69172f97ec002ca160d6 | Mali Buffet Thai | 4bf58dd8d48988d149941735 | Thai Restaurant | 75013 Paris, France | 48.817560 | 2.365998 | 241 | 13 |
| 1 | 4b102aeef964a5206e6a23e3 | La Rotonde | 4bf58dd8d48988d16d941735 | Café | 7 place du 25 aout 1944, 75014 Paris, France | 48.822026 | 2.325494 | 295 | 14 |
| 2 | 4cb34a961463a143da7eb3a9 | Bistrot 32 | 4bf58dd8d48988d110941735 | Italian Restaurant | 32 boulevard Jourdan, 75014 Paris, France | 48.821533 | 2.333947 | 272 | 14 |
| 3 | 4d0f45807177b1f714704a22 | Le Chalet du Parc | 4bf58dd8d48988d147941735 | Diner | 26-28 boulevards des Maréchaux, 75014 Paris, F... | 48.821360 | 2.334616 | 305 | 14 |
| 4 | 4ce7ecc4678aa0935951e5ea | L'Epi d'Orléans | 4bf58dd8d48988d16a941735 | Bakery | 107 boulevard Jourdan, 75014 Paris, France | 48.822406 | 2.328029 | 323 | 14 |

## 3.6 K-means algorithm

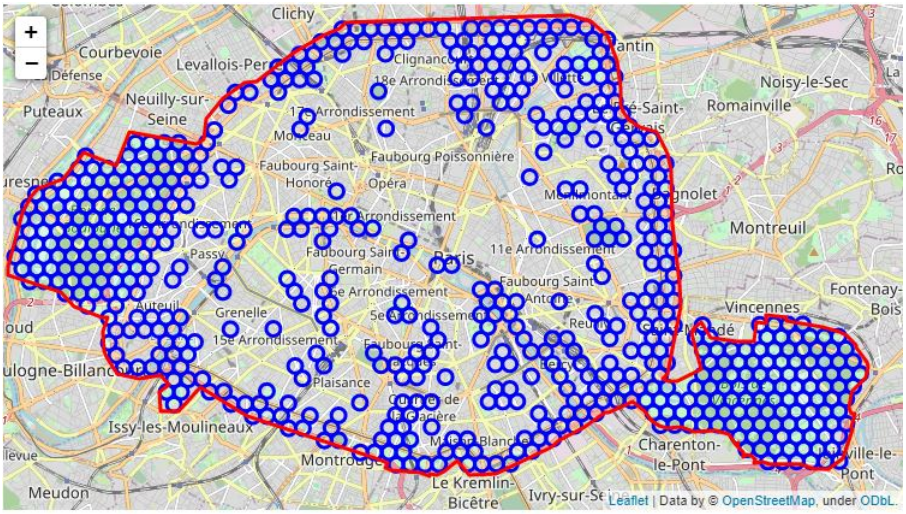
In order to find the less dense places regarding the quantity of eating places, the K-means algorithm was used to cluster these areas.

To do so, a new hexagonal grid was created as we already did before, but this time with a smaller distance between points (this time the diameter of each circle is 300 meters):



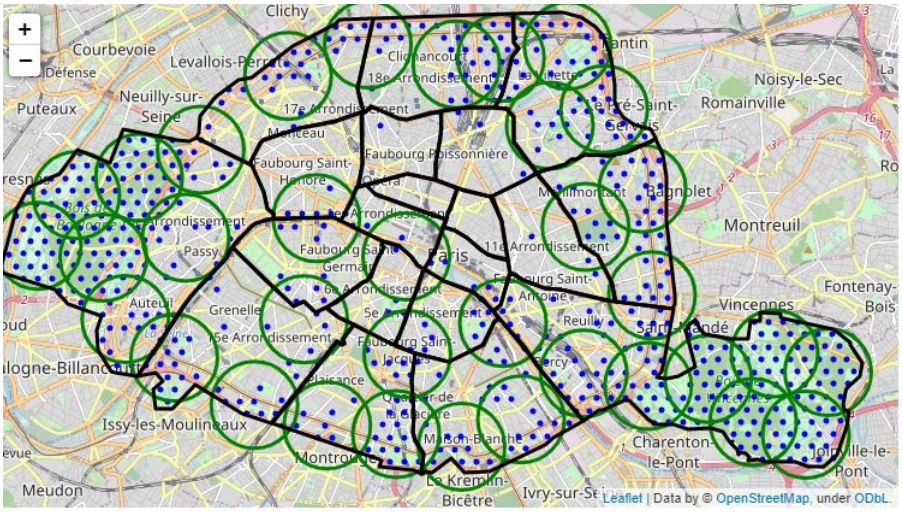

With this new grid in place, we are interested on removing the circles from it which there are more than 2 restaurants inside. We can do it by calculating the distance from the center of each circle of the grid to all the eating places.

After the filtering procedures, we can display the remaining circles of the grid :




Each of these circles have less than 3 eating places on a radius of 150 meters.

Now we can use the Kmeans method to cluster the less dense areas and find coordinates of the center of the clusters :



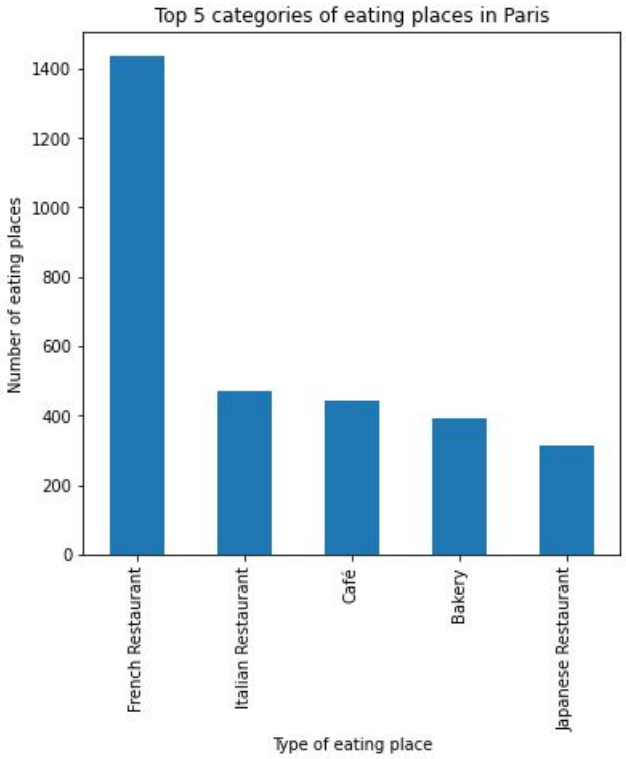

We used a k = 35 on this method. Each green circle represents a cluster.

# 4 Results and discussion

Now that we did all the data extraction and cleaning, we can start the data analysis.

## 4.1 Top categories of eating places

1. Top 5 categories of eating places in Paris :




As we can see on the bar graph above, the French restaurant is the most common type of restaurant in Paris. This is normal, because Paris is known for having a large amount of typical French restaurants and also because Paris is a very touristic city and tourists frequently go to the typical restaurants of the country they are visiting.

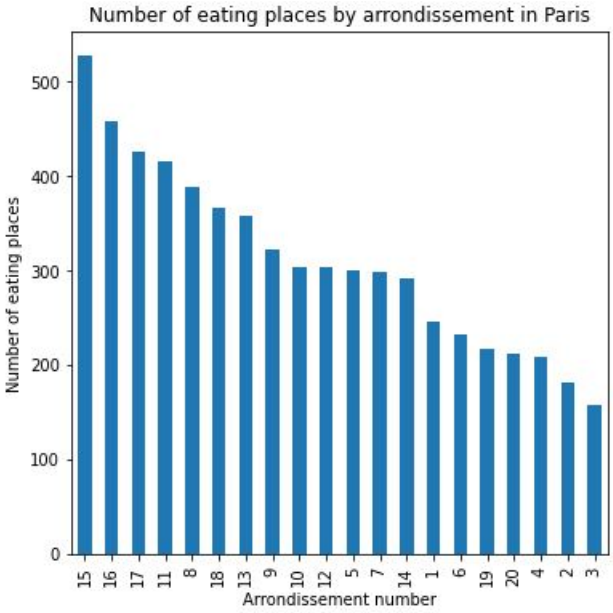
In second place we have Italian restaurants. This type of restaurant is very popular in the world and in France, especially because of the proximity of France and Italy.

In third and fourth position, we have the cafes and bakeries, which are present throughout the country, in all cities of France.

In fifth position, we have Japanese restaurants. Even though this is a very popular type of restaurant in the world, it is interesting to know that they are so common in Paris.

So, if a stakeholder wants to open a new restaurant in Paris and is looking for less competition, these types of restaurants should be avoided.

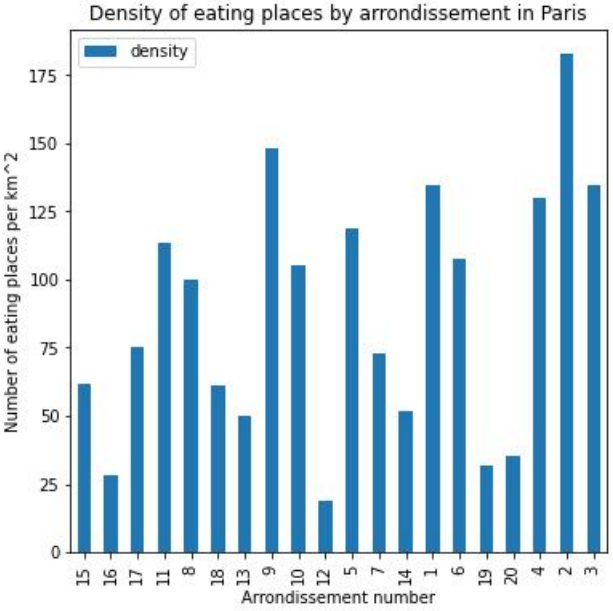
2. Number of eating places by arrondissement :




This time, if a stakeholder is looking to open a new restaurant on a specif arrondissement in Paris, the number of restaurants by region could be an interesting information.

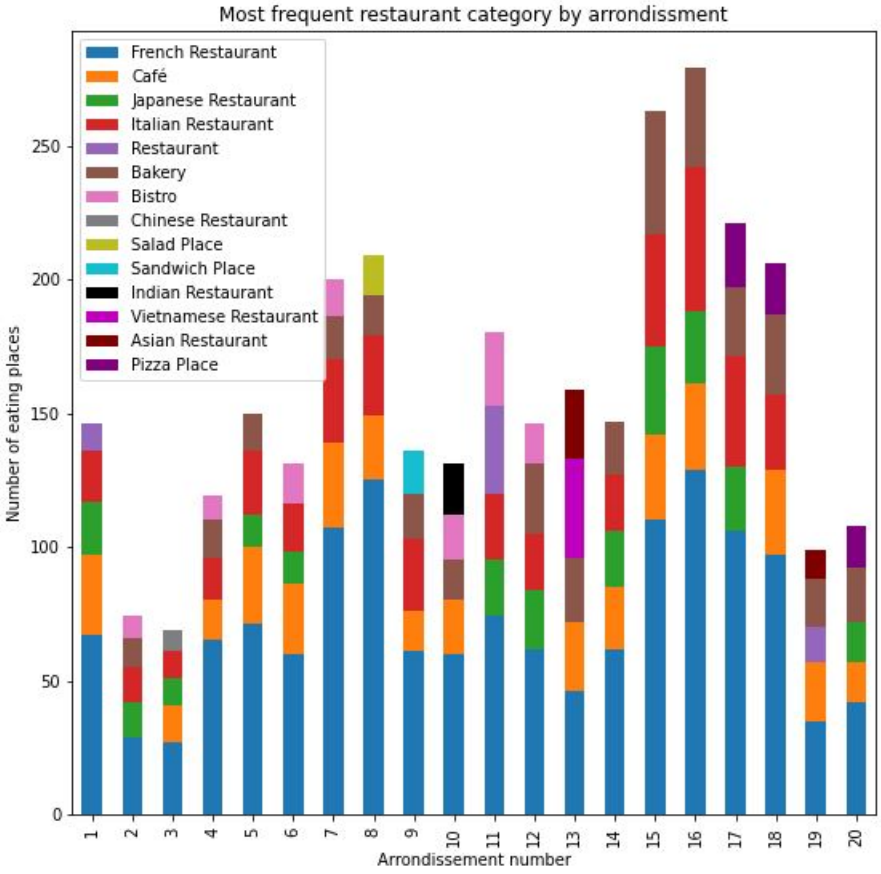
However, since there are some arrondissements that have a greater area than others (for example, the 15th arrondissement is has a larger surface than the 3rd arrondissement), a better way to visualize which area has more arrondissements is by dividing the number of venues by the surface area (check the next analysis for the results).

3. Density of eating places by arrondissement :



Comparing the results from this section with the last section, we can see some remarkable differences. For example, the second arrondissement is one of the areas with the fewest restaurants, but it has the highest amount of eating places per $km^2$. So even though the 15th arrondissement has the most restaurants, it is potentially the best area to open a restaurant since it has the lowest density of eating places.

4. Most frequent restaurant category by arrondissment :




As in the last two previous analysis, if we are interested in opening a restaurant in a specific arrondissement, it may be interesting to know which restaurants are the most frequent by arrondissement.

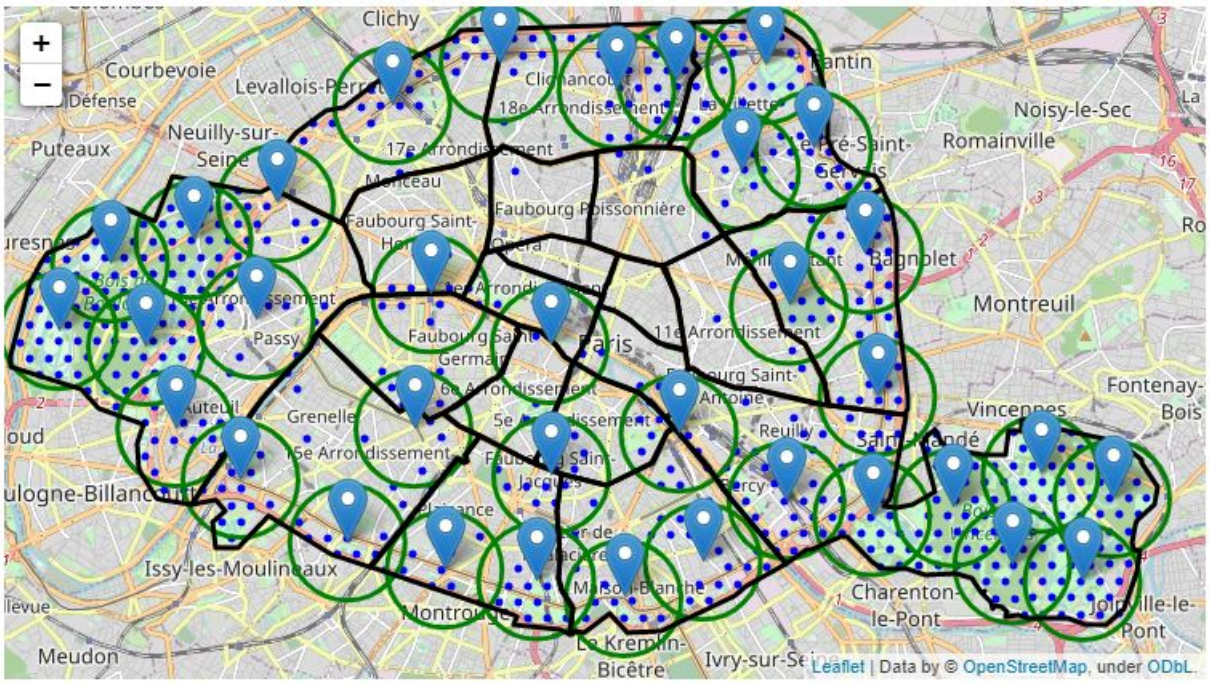
For example, the 10th arrondissement has an important number of Indian restaurants and the 13th has an important number of Vietnamese restaurants. So, in those districts, it would be interesting to avoid opening these kind of eating places.

## 4.2 Best places in Paris to open a new eating place based on the less dense areas

We can take directly the addresses of the center each cluster from the Kmeans method. The addresses are :

| Addresses |
|---|
| Monument aux mères françaises, Boulevard Kellermann, Quartier de la Maison-Blanche, 13e Arrondissement, 75013 |
| 27, Rue Marceline Desbordes-Valmore, Quartier de la Muette, 16e Arrondissement, 75116 |
| Wonderland, Cours de Vincennes, Quartier de Charonne, 20e Arrondissement, 75020 |
| Allée d'Andrézieux, Quartier de la Goutte-d'Or, 18e Arrondissement, 75018 |
| Oiseaux du Lac des Minimes, Route de la Cascade, Quartier de Picpus, 12e Arrondissement, 75012 |
| Jardin méditerranéen, Parterre Francais, Quartier de la Muette, 16e Arrondissement, 75016 |
| 13, Passage Boudin, Campagne à, Quartier Saint-Fargeau, 20e Arrondissement, 75020 |
| Résidence Jacques Balmat, Quartier des Batignolles, 17e Arrondissement, 75017 |
| École maternelle d'application Erlanger, Boulevard Exelmans, Hameau Boileau, Quartier d'Auteuil, 16e Arrondissement, 75016 |
| Les Magnolias, Route de la Pyramide, Quartier de Picpus, 12e Arrondissement, 75012 |
| Hôpital Cochin, 27, Rue du Faubourg Saint-Jacques, Quartier du Montparnasse, 14e Arrondissement, 75014 |
| 2, Square Pasteur, Quartier Necker, 15e Arrondissement, 75015 |
| Halle Gabriel Lamé, Rue Baron Le Roy, Quartier de Bercy, 12e Arrondissement, Reuilly, 75012 |
| Port d'Austerlitz, Rue Berenice Abbott, Quartier de la Salpêtrière, 13e Arrondissement, 75013 |
| Cimetière du Père-Lachaise, 8, Boulevard de Ménilmontant, Quartier du Père-Lachaise, 20e Arrondissement, 75020 |
| Quartier du Pont-de-Flandre, 19e Arrondissement, 75019 |
| Pelouse de Saint-Cloud, Avenue de Saint-Cloud, Quartier d'Auteuil, 16e Arrondissement, 75016 |
| Pont Alexandre III, Quartier des Invalides, 7e Arrondissement, 75007 |
| Parc Montsouris, Boulevard Jourdan, Quartier du Parc-de-Montsouris, 14e Arrondissement, 75014 |
| Avenue des Tribunes, Quartier de Picpus, 12e Arrondissement, Ferme de, 75012 |
| 134 bis, Avenue de Saint-Ouen, Quartier des Épinettes, 17e Arrondissement, 75017 |
| Boulevard du Général Martial Valin, Quartier de Javel, 15e Arrondissement, 75015 |
| Rue Trolley de Prévaux, Quartier de la Gare, 13e Arrondissement, 75013 |
| Allée de Longchamp, Quartier de la Muette, 16e Arrondissement, 75016 |
| Route de la Ferme, Quartier de Picpus, 12e Arrondissement, 75012 |
| Église luthérienne Saint-Pierre, Rue Manin, Quartier du Combat, 19e Arrondissement, 75019 |
| Tristan Tzara, Rue de l'Évangile, Quartier de la Chapelle, 18e Arrondissement, 75018 |
| Monument à Émile Levassor, Porte Maillot, Quartier de la Porte-Dauphine, 16e Arrondissement, 75116 |
| 14, Rue de Périgueux, Quartier d'Amérique, 19e Arrondissement, 75019 |
| Promenade Maurice-Boitel, Quartier du Bel-Air, 12e Arrondissement, Reuilly, 75012 |
| Hippodrome de Longchamp, Avenue de l'Hippodrome, Quartier d'Auteuil, 16e Arrondissement, 75016 |
| G20, Avenue Albert Bartholomé, Expo Porte de Versailles, Quartier Saint-Lambert, 15e Arrondissement, 75015 |
| Boulevard Brune, Alésia, Quartier du Petit-Montrouge, 14e Arrondissement, 75014 |
| Palais Conti, Quai de Conti, Quartier de la Monnaie, 6e Arrondissement, 75006 |
| Allée Cavalière, Quartier du Bel-Air, 12e Arrondissement, 75012 |

Displaying the map with the clusters and the markers that indicate the center of each cluster (where the addresses were taken from) :




These addresses are potentially good places to open a new restaurant. However, we must pay attention to this analysis because it indicates the "Bois de Vincenne" and the "Bois de Boulogne" as potential locations for the opening of a new restaurant. As these places are parks, it is not too feasible to open a new business over there.

# 5 Conclusions

In this study, I analyzed how eating places are distributed in Paris and in its "arrondissements". I used several Python libraries to perform data cleaning and extraction and the Foursquare API to retrieve the location, categories and names of venues. The "Paris Data" website provided a geojson file with the boundaries of Paris, which was useful to delineate the study area.

Furthermore, I used the Kmeans machine learning method from the Sklearn library to create clusters on the areas with fewer restaurants and gathered the addresses for each cluster center.

Also, with this analysis, we can tell the most frequent types of restaurants in Paris by arrondissment and on the entire city.

Finally, as it was already mentioned, this study can be useful for anyone who is interested in opening a new eating place at Paris and wants to know which kind of restaurant to choose and also the best locations to do so.

# References

[1] "Paris Data" webstie - Arrondissements : https://opendata.paris.fr/explore/dataset/arrondissements/

[2] Klokan Technologies' open-source web service : https://epsg.io/

[3] Foursquare API : https://developer.foursquare.com/

[4] Nominatim API : https://nominatim.org/