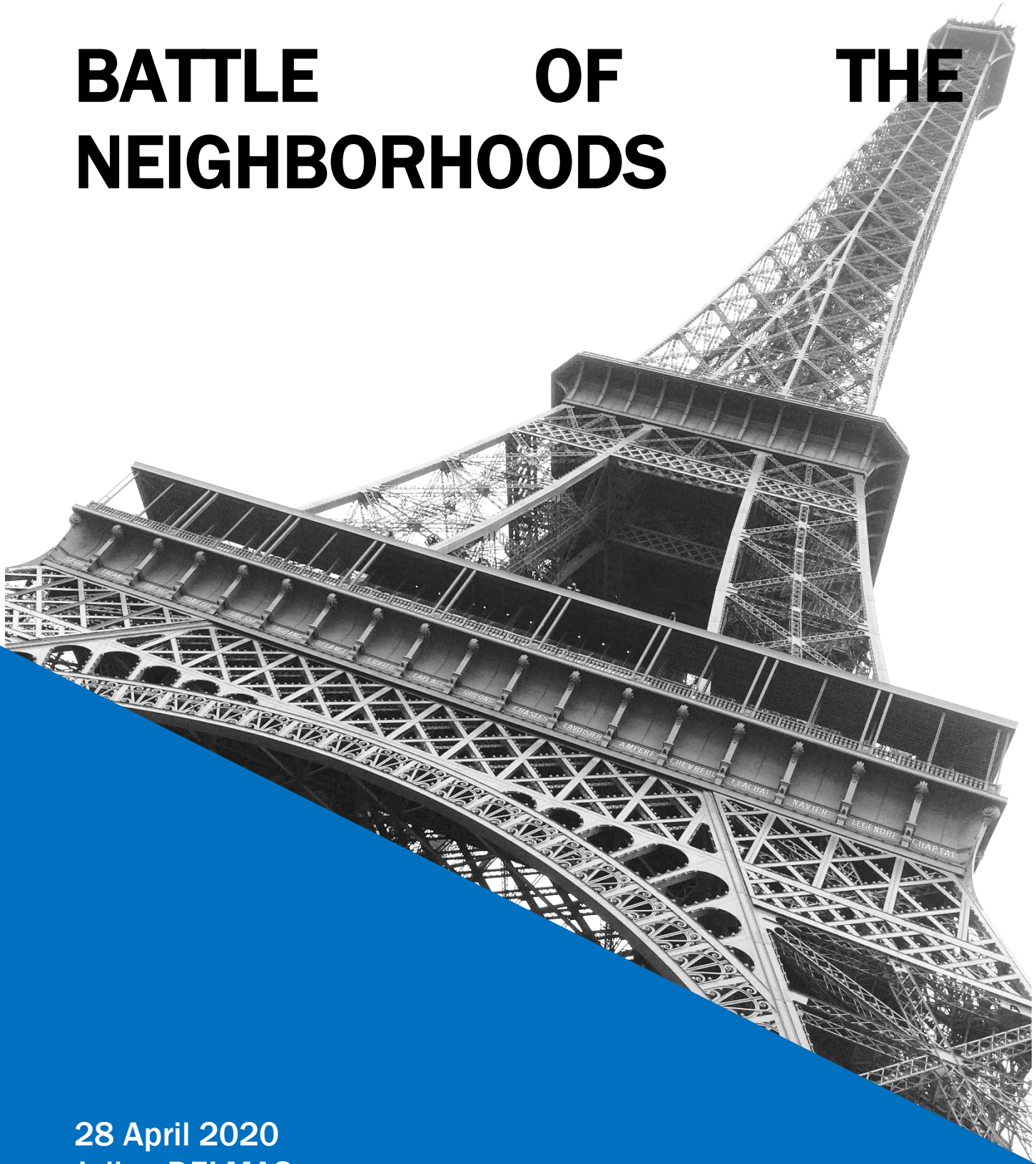
BATTLE OF THE NEIGHBORHOODS



28 April 2020

Julien DELMAS

Best location for a seafood restaurant in Paris, France

Paris is the administrative, economic and demographic capital of France. With over 2.2 million inhabitants, it is also the first touristic destination in the country and it is reknown for its cuisine.

There are hundreds of restaurants in Paris, most of them being French restaurant (mainly "brasserie") or Italian restaurants.

For this capstone project, I decided to suppose that I want to open a seafood restaurant in Paris and I want to find the best place to locate such a restaurant.

This project will be developped in a way that it would be very easy to change the restaurant type (for example change to salad bar, brasserie or fast-food restaurant) by only changing 1 parameter of the code.

Table of Contents

Problem Statement

As stated int the introduction we will try to find the best place to open a new seafood restaurant in Paris.
Paris has a high density of restaurants and finding the best place to open a restaurant is not easy. Consequently, data analysis is probably one of the best ways to approach this problem.

Paris is divided into 20 districts (called "arrondissement" in French). Each of these districts is divided in 4 neighborhoods (called "quartier" in French). Consequently, Paris is divided in 80 neighborhoods.
By analyzing the number of seafood restaurants and their typology, we will try to determine the best neighborhood to open a new seafood restaurant.

The "best place" to open a seafood restaurant can be defined as follows:

- At least one of the following criteria is met regarding competition:
  - ✓ Where there is no or few competition in the seafood restaurant category
  - ✓ Where the competition is on another range of price (cheaper or more premium)
  - ✓ Where the competition is poorly rated
- The population density is high enough to have enough clients (we ignore the impact of tourism for the sake of simplicity)

We will try to answer the following questions:

1. How many seafood restaurants are there in Paris?
2. Where are the best seafood restaurants of Paris located?
3. Where are they located? In which neighborhood / district?
4. In which neighborhood and/or borough should I open a seafood restaurant?

It is important to note that we will develop the project in a way that it would be very easy to change the restaurant type (for example change to salad bar, brasserie or fast-food restaurant) by only changing 1 parameter of the code.

Data Collection Plan

Data about the City of Paris

We will need to obtain the list of neighborhood with as much data as possible.

Use of Paris Opendata

The city of Paris issued numerous data in Opendata format on the website https://opendata.paris.fr. In particular, the page https://opendata.paris.fr/explore/dataset/quartier_paris/information gives data regarding the neighborhoods of Paris.
The last update of this data is from March 2013 which is sufficient due to the absence of administrive reorganization since this time.

The datafile, which is available in CSV, JSON, Excel, GeoJSON, Shapefile and KML, contains the following columns:

- N_SQ_QU: sequencial id of the neighborhood,
- C_QU: Number of the neighborhood,
- C_QUINSEE: Official number of the neighborhood according to a national format (provided by the Institut national de la statistique et des études économiques, INSEE, i.e. thr National Institute of Statistics and Economic Studies),
- L_QU: Name of the neighborhood,
- C_AR: Number of the district,
- N_SQ_AR: Sequencial link between district and neighborhood,
- PERIMETRE: Perimeter of the neighborhood,
- SURFACE: Area of the neighborhood,
- Geometry X Y: Coordinates of the center of the neighborhood,
- Geometry: Polygon of the boundaries of the neighborhood.

We will use the CSV format because it is easily usable with the *pandas* library and the following fields will be used:

- C_QU: Number of the neighborhood,
- C_QUINSEE: Official number of the neighborhood according to a national format,
- L_QU: Name of the neighborhood,
- C_AR: Number of the district,
- SURFACE: Area of the neighborhood.

In addition, the information regarding the geometry of each neighborhood will be used in GeoJSON format to create maps with Folium.

Use of Wikipedia

The Paris Opendata is missing the population by neighborhood. Consequently, I searched for another source of population per neighborhood and the only I found is Wikipedia (https://en.wikipedia.org/wiki/Quarters_of_Paris).
It is fair mentioning that the population data dates from 1999, which is more than 20 years old. Nevertheless, the population of Paris has not changed much within this period (see

https://en.wikipedia.org/wiki/Demographics_of_Paris#/media/File:Paris_Historical_Population.png) and we can assume that it is also the case with neighborhoods.

The fields available in the Wikipedia table are:
- District (district official name and "also called"),
- Neighborhood (number and name),
- Population in 1999,
- Area,
- Map.

We will not use the area as it is already available in the Paris Opendata dataset and we will not use the map as it is given in an image format.

Consequence

Due to the use of 2 sources, we will need to merge the information for both sources. The neihborhood number is the common key to use for the merger.

Thanks to this merger, we will have one database with the following data:
- Neighborhood code (according to INSEE format),
- Neighborhood number (from 1 to 80 according to Paris format),
- Neighborhood name,
- District number of the neighborhood (from 1 to 20),
- District name of the neighborhood,
- District "also called" name of the neighborhood,
- Population (in 1999) of the neighborhood,
- Area of the neighborhood,
- Perimeter of the neighborhood,
- Latitude and longitude of the neighborhood.

The we will be able to add:
- Postal Code which is formed of 75 (numer of the department in France) + 0 + number of the district (from 01 to 20),
- Density of each neighborhood by dividing the population by the area.

In addition, we will have a GeoJSON with the coordinates of the polygon for eahc neighborhood.

Data about the restaurants

We will use Foursquare to obtain the data about the seafood restaurants in Paris.

We will use the "search" method to obtain:
- the list of restaurants for each neighborhood (based on the coordinates of the center of the neighborhood and a radius of 1500 meters which should cover all the neighborhoods (except part of the 2 forests that are at the West and East extremities of the City and contain few restaurants),
- for each restaurant, its id, its name and its location.

Because all the neighborhoods have different sizes and we use only 1 radius, we will have to remove duplicates and check in which neighborhood belong each restaurant. We can perform this last step by using the GeoJSON obtained previously.

Afterwards, we will obtain more details about each seafood restaurant with the "venue" method of Foursquare:

- Price category,
- Rating of the restaurant,
- Number of likes,
- Number of tips,
- Whether the seafood category is the primary category of the restaurant.

Planned methodology

Data visualization and analysis regarding neighborhoods of Paris

In order to better understand the neighborhood of Paris, we will:

- Draw a choropleth map of Paris with the population of each neighborhood to understand where the population is mainly situated,
- Draw a choropleth map of Paris with the density of each neighborhood to take into account the size of each neighborhood,
- Create a scatter plot of the population vs the area per neighborhood,
- Create a bar chart of the density of population per neighborhood,
- Create a bar chart of the density of population per district.

Data visualization and analysis regarding restaurants

In order to better understand the neighborhood of Paris, we will:

- Draw a map of the seafood restaurants to see their location,
- Create a bar chart of the number of seafood restaurants per neighborhood,
- Create a bar chart of the number of seafood restaurants per neighborhood and per million inhabitant,
- Create a bar chart of the number of seafood restaurants per district,
- Create a bar chart of the number of seafood restaurants per district and per million inhabitant,
- Map representing the rating, the price and whether seafood restaurant is their primary category for each seafood restaurant,
- Draw a scatter plot of price vs. rating of seafood restaurants,
- Draw a bar chart of the average rating of seafood restaurants per neighborhood,
- Draw a bar chart of the average price of seafood restaurants per neighborhood,
- Draw a bar chart of the average rating of seafood restaurants per district,
- Draw a bar chart of the average price of seafood restaurants per district.

Clustering of the restaurants

We will cluster the restaurants based on the Density-based spatial clustering of applications with noise (DBSCAN) algorithm.

The main reasons why I chose this clustering algorithm are:

- It can find clusters of arbitrary shapes,
- It is robust to outliers and we may have many outliers in our data,
- Compared to k-Means, we do not need to specify the number of clusters.

We will simulate the DBSCAN algorithm with different sets of input parameters and compare the resuls:

- Coordinates,
- Price,
- Rating,

- Number of tips,
- Number of likes,
- Whether seafood restaurant is the primary category.

And now, let's write Python code and analyze data!

Acknowledgements

Photo of the Eiffel Tower is from Pixabay: https://pixabay.com/fr/photos/paris-tour-effel-otel-resita-676931
It is under Pixabay license allowing its use in this report:

- All content on Pixabay can be used for free for commercial and noncommercial use across print and digital, except in the cases mentioned in "What is not allowed".
- Attribution is not required. Giving credit to the contributor or Pixabay is not necessary but i appreciated by our community.
- You can make modifications to content from Pixabay.

Photo of the seafood platter is from pxhere: https://pxhere.com/fr/photo/820162.
It is under Creative Commons CC0 license.