

Choosing the best spot to open a Pizzeria in Paris.

Mickael Tessier

28th of January 2021

1. Introduction

1.1. Background

The client wants to open a new pizzeria franchise (Italian restaurant) in Paris. The client has numerous franchises in different cities in Europe and north America and is recognized to serve high quality food. The business department evaluate the French market as very promising and are very excited to open a new restaurant in Paris. From the experience they acquired by opening restaurants in Europe, they know what the best conditions for the location are. The project is to use the marketing department inputs for evaluating where is the best location for opening the first company's restaurant in Paris using machine learning and data scouting methods.

1.2. Problem

Discussion with the company marketing department led to several inputs that can be expressed as:

- **Condition 1:** It has been noticed that many clients go to location where several options are available in a relatively small area. Therefore, the restaurant should open in an area where other pizzerias are around. It has been evaluated that at least 3 pizzerias should be found in a range of 1 km around the location.
- **Conditions 2:** It has been noticed that many clients look at the rate that other customers gave to the place and that the company's restaurant is performing more when the average rate of the neighboring restaurants is lower.

These two conditions are used to find the best location to open the restaurant in Paris.

2. Data Section

2.1. Data sources

I used FourSquare to obtain the data that I need for the project. Basically, two API calls were needed. The first one was a search of all the pizzeria located in a radius of 5 km around the center of Paris. The second call was to get the client's note in the identified restaurants.

2.2. Data cleaning

Many variables were not needed (*e.g.* address, restaurant type, country, city, *etc*), those have been remove from the dataset. Two datasets were combined into one dataframe, the first contained the restaurant name, ID, latitude, longitude and the second one the client evaluation.

2.3. Obtained data

Data are displayed in 1 table in the next page of the report. It consists of 5 columns, the name of the restaurant, the latitude (lat) and longitude (lng), the ID and the reviewing note. It should be noted that some restaurants did not have a note. The code can be found on the following link:

https://github.com/mtessier44/Coursera_Capstone/blob/main/Capstone-W4.ipynb

name	lat	lng	id	Note
O'scià Pizzeria Napoletana	48.864821	2.347524	4c9894df07916dcb8d01ed1b	7.6
Pizzeria Italiana Da Michele	48.867739	2.356261	51430960e4b0455b2d070f45	None
Pizzeria d'Istanbul	48.837677	2.359962	4dc445371f6ef43b89f7e46e	None
Pizzeria Mazarine	48.853989	2.337987	4c6fb714d274b60c3d57d70d	None
Pizzeria Pepone	48.853511	2.337967	5131ded7e4b04a7c69396261	7.3
Pizzeria Maccaron	48.859454	2.344401	5371073c498e8b0d316a4db8	None
Pizzeria Luka	48.862087	2.351836	4e59fe822271886714fb58d5	None
Pizzeria Luciana	48.845297	2.342616	573eecb1498ebf4b4efd1b8c	None
Pizzeria Nino	48.839535	2.299931	523741de11d2a8888a604602	None
Pizzeria La Gaine	48.859215	2.354996	4e59e059483b865169e59dba	None
Pizzeria Popolare	48.868074	2.343379	58bc834d7d0f6d34315f5cfe	8.3
Pizzeria Ghost	48.861029	2.358219	4e58fd782271886714ec7f9d	None
Pizzeria d'Auteuil	48.848773	2.265839	4b8d6621f964a5206ff932e3	7.9
Pizzeria Sette	48.871270	2.353464	53f7b51b498e5ac6875be50a	8.2
Mama Pizzeria	48.859778	2.402545	513130f5e4b0ed2721b6e1fa	8.0
Pizzeria Renato	48.866348	2.364821	4b8bffe9bf964a52082b732e3	7.5
Pizzeria Les Comédiens	48.844731	2.320639	4ba66d6cf964a520635139e3	6.7
Pizzeria Saint-Lazare	48.876788	2.337441	4b41e2a0f964a520c7c925e3	None
pizzeria	48.884677	2.333735	58ce831d18fd2b1c6a40c5ec	None
Pizzeria Romanella	48.838223	2.356450	51043554e4b08881ebbec640	None
Pizzeria Le Coq	48.848458	2.374179	4b60951cf964a520c3ee29e3	7.4
Pizzeria Don Giacomo	48.853623	2.370256	53f79936498ef939417afa78	6.5
Pizzeria Sette	48.856550	2.373932	5b5e12e160255e002c0d5d51	None
Pizzeria Napoletana	48.853396	2.378799	4e0328be7d8bf11e9f3c1383	None
Pizzeria Miramar	48.870489	2.320640	543eb095498e12bc703d80a5	None
Pizzeria Via Maria	48.836997	2.318552	4d946494ca3f6ea8f502c834	6.4
Pizzeria Sandro	48.847132	2.304382	4c2b9a5b8abca59301720020	6.4
La Pizzeria di Rebellato	48.868593	2.281747	4ed289aae5faa5ec0559eb2e	7.2
Pizzeria Saint Georges	48.877220	2.354398	4e972515b634527c67a6595a	None
Pizzeria Chez Hino	48.840488	2.312244	4eef943377c82b92fa7121d1	None

Table 1: Data used for the project

3. Methodology

3.1. Data visualization

The first step consists of visualizing the data on a map. For that folium was used to generate a map and the different restaurants were displayed according to their positions (cf. Figure 1). We can already observe that some neighborhoods are isolated with only one restaurant while in others districts several pizzerias are close to each other.

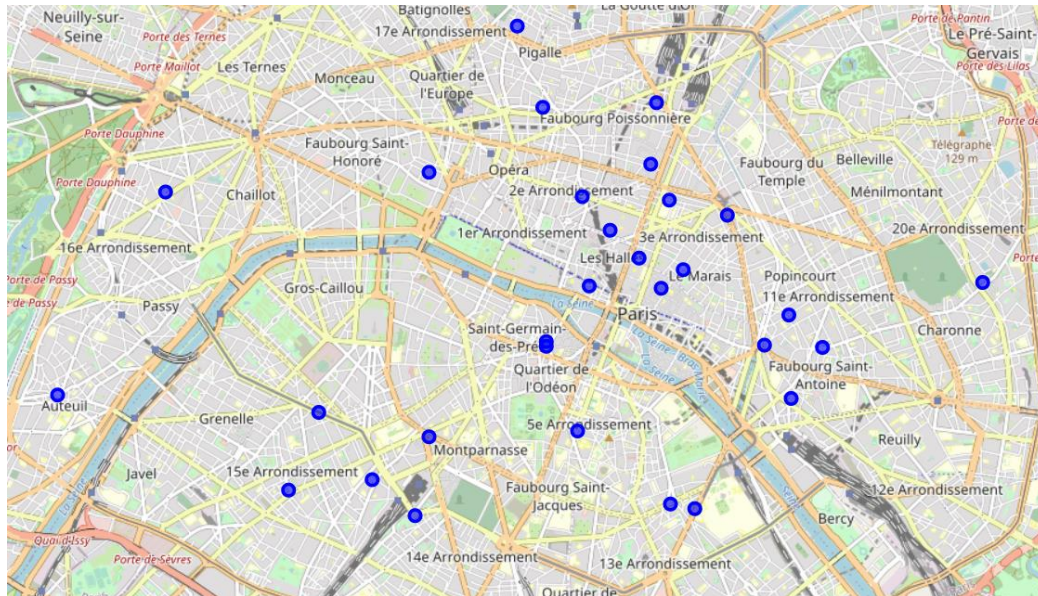


Figure 1: Positions of pizzerias in Paris (blue dots)

3.2. Machine learning: Clustering

To identify districts, we used the machine learning tool sk-learn. We clustered restaurants in five cluster according to their latitude and longitude (Table 2). These clusters therefore correspond to five different districts that we analyze in the next section. The clusters are displayed in the next map (Figure 2).

District	Latitude	Longitude
District 1	48.848025	2.346564
District 2	48.848773	2.265839
District 3	48.853905	2.382424
District 4	48.840254	2.311418
District 5	48.871665	2.329171
District 6	48.866741	2.354234

Table 2: District's coordinates

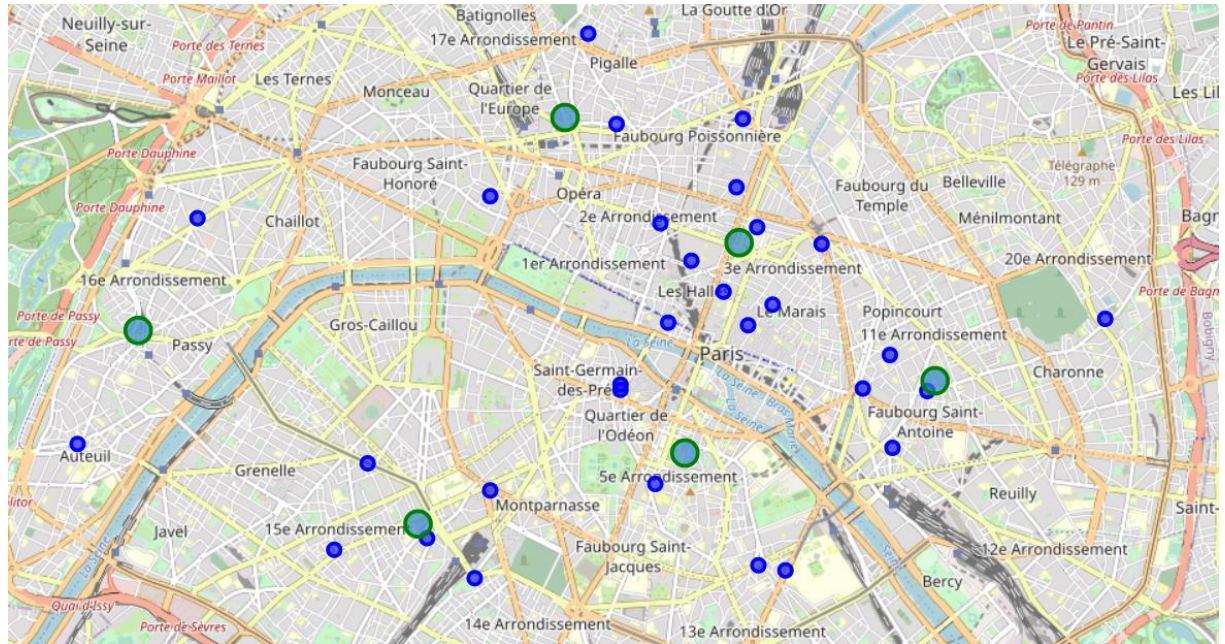


Figure 2: Positions of selected districts in Paris (green dots)

3.3. Neighboring search

The first condition is that the chosen spot must have at least 3 other pizzerias in less than 1 km. For that we used Haversine formula (https://en.wikipedia.org/wiki/Haversine_formula) to measure distances between the districts' centers and the restaurants position in each district. This method allows to obtain the following table:

District	Latitude district	Longitude district	Number of neighbours
District 1	48.866423	2.353877	8
District 2	48.858683	2.273793	0
District 3	48.848025	2.346564	3
District 4	48.841777	2.311150	5
District 5	48.854361	2.379942	4
District 6	48.877318	2.330605	2

Table 3: District's position and number of restaurants

We can see that district 2 and district 6 have less than 3 restaurants. They do not respect condition 1 and therefore these spots cannot be selected to open the new restaurant.

3.4. Average Note per district

To implement the second condition, we must make the average of the note of each restaurant in each district (except district 2 and district 5 that do not respect condition 1). We therefore obtain the following table:

District	Average note
District 1	7.900000
District 3	7.300000
District 4	6.500000
District 5	7.366667

Table 4: average note per district

4. Results

From table 4, we can conclude that the best spot to open a new pizzeria in Paris is district 4. The spot is plotted in the map of Paris in Figure 3.

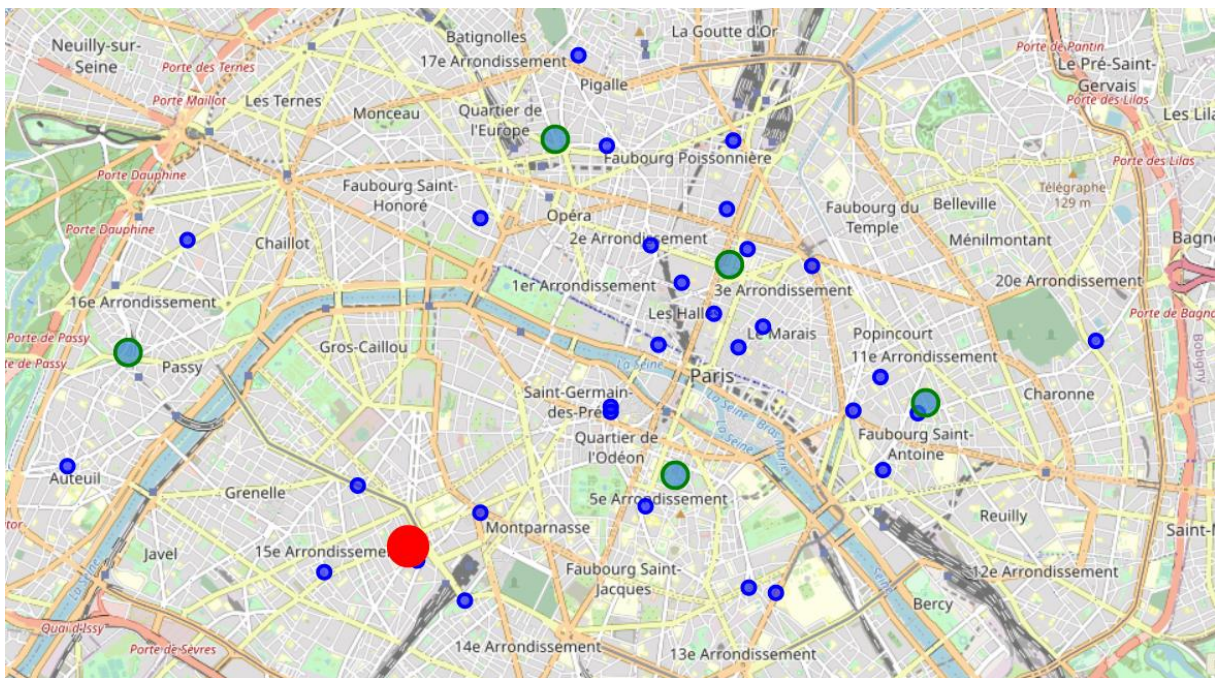


Figure 3: Best spot to open the restaurant (red dot)

5. Discussion

We deduce from condition 1 and condition 2 given by the marketing department is district 3. We can also see from table 4 that district 4 is a good spot as the note is very close to the one of district 4. This spot is in the 14th arrondissement of Paris.



Figure 4: Picture of the 14th arrondissement close to the identified spot

6. Conclusion

We have founded the best spot to open a pizzeria in Paris respecting the conditions given by the marketing department (i.e. more than 3 restaurants in less than 1 km distance and lowest reviews in the neighborhood). The spot is in the 14th arrondissement of Paris.