

# Finding best location to open a Food Court in Paris, France

M.G April 2020

## 1. Introduction / Business understanding

My client is a 35 years old food court owner in Paris. He opened his first food court last year in Paris 14eme. His idea for this food court was to provide different kinds of food based on what's most popular in his neighborhood. This probably was a good idea as his business is very successful now. So he decided to open another food court in another place in Paris. He asked for my counseling services to be helped finding where he should open his new business.

His strategy is the following: He wants for his new food court to be the same as the one he already has in Paris 14eme. That means he wants to provide the same food offer exactly.

He provided me with a list of 19 different neighborhoods in Paris where he has the possibility to buy a property to install his new food court.

The question he wants me to answer for him is the following:

**From those 19 places provided by my client, what are those that look the most like his actual business, based on the restaurant offer in the neighborhood?**

## 2. Data and Analytics approach

### Data requirement

The necessary data for this project, based on the above stated requirements, are:

- 1) The name of the place my client has as actual Food Court.
- 2) The list of the names of the places my client prospects to open his new business.
- 3) GPS coordinates for each place mentioned above.
- 4) The list of all restaurant categories near each of the places mentioned above

### Data collection

The list of the names of the places my client prospects to open his new business is provided (by my client) below:

Place	Corresponding to
Paris 1er	Prospection for new business
Paris 2eme	Prospection for new business
Paris 3eme	Prospection for new business
Paris 4eme	Prospection for new business
Paris 5eme	Prospection for new business
Paris 6eme	Prospection for new business
Paris 7eme	Prospection for new business
Paris 8eme	Prospection for new business
Paris 9eme	Prospection for new business
Paris 10eme	Prospection for new business
Paris 11eme	Prospection for new business
Paris 13eme	Prospection for new business
<b>Paris 14eme</b>	<b>Actual business</b>

Paris 15eme	Prospection for new business
Paris 16eme	Prospection for new business
Paris 17eme	Prospection for new business
Paris 18eme	Prospection for new business
Paris 19eme	Prospection for new business
Paris 20eme	Prospection for new business

For each of this place, GPS localization will be retrieve using [api-adresse.data.gouv.fr](https://api-adresse.data.gouv.fr/), the French government geoloc API:

place_name	latitude	longitude
Paris 1er	48.8621	2.33621
Paris 2eme	48.8677	2.34116
Paris 3eme	48.8625	2.35932
Paris 4eme	48.8544	2.35692
Paris 5eme	48.8453	2.35058
Paris 6eme	48.8491	2.33049
Paris 7eme	48.8547	2.31208
Paris 8eme	48.8736	2.3116
Paris 9eme	48.877	2.33789
Paris 10eme	48.8761	2.36227
Paris 11eme	48.8602	2.38184
Paris 13eme	48.8302	2.365
Paris 14eme	48.8301	2.32303
Paris 15eme	48.842	2.29477
Paris 16eme	48.8564	2.26482
Paris 17eme	48.8878	2.30484
Paris 18eme	48.892	2.34868
Paris 19eme	48.8873	2.38771
Paris 20eme	48.8626	2.39675

For each localization, the list of the most popular restaurant in a 500 meters radius will be retrieve through Foursquare API.

	0	1	2	3	4	...
<b>Paris 1er</b>	Italian Restaurant	Café	Café	Italian Restaurant	Bakery	...
<b>Paris 2eme</b>	French Restaurant	Bistro	Italian Restaurant	French Restaurant	Pizza Place	...
<b>Paris 3eme</b>	Sandwich Place	Sandwich Place	Okonomiyaki Restaurant	Sandwich Place	Burger Joint	...
...	...	...	...	...	...	...

### Analytics approach

Using the collected data, I will calculate the number of restaurant by category for each localization. After data preparation I will use the resulting dataframe as input to K-Means clustering algorithm in order to obtain the clusters of the neighborhood including the one of my client actual business.

## 3. Methodology

### Data preparation

It has consisted in transforming the features columns into dummies variables and applying a function to each row to get the probability for each features (so that the sum of each row is equal to 1).

	<b>Afghan Restaurant</b>	<b>African Restaurant</b>	<b>American Restaurant</b>	<b>Argentinian Restaurant</b>	<b>Asian Restaurant</b>	<b>Auvergne Restaurant</b>	...
<b>Paris 1er</b>	0	0	0	0	0	0	...
<b>Paris 2eme</b>	0	0	0	0	0	0	...
<b>Paris 3eme</b>	0	0	0	0	0.014706	0.014706	...
<b>Paris 4eme</b>	0	0	0	0	0.014706	0	...
<b>Paris 5eme</b>	0	0	0	0.014706	0	0.014706	...
<b>Paris 6eme</b>	0	0	0	0	0	0	...
<b>Paris 7eme</b>	0	0	0	0	0.039216	0	...
<b>Paris 8eme</b>	0	0	0	0	0.021277	0	...
...	...	...	...	...	...	...	...

## Data exploratory

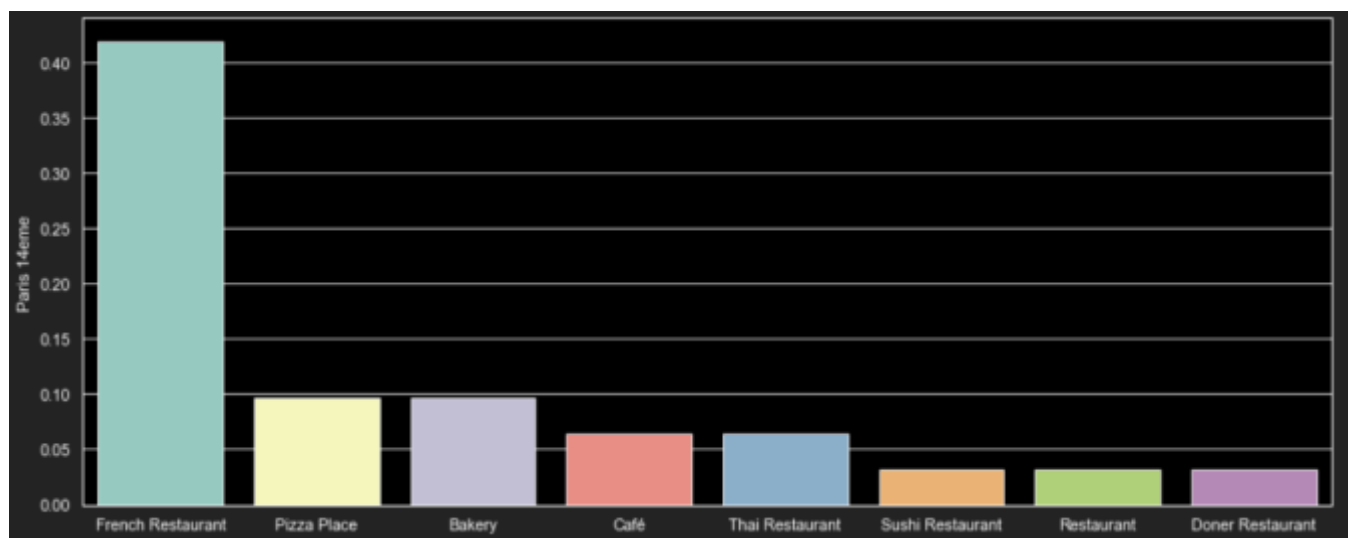
Let's first visualize the different places on a map using the folium Python Library:



The food court currently owned by my client in Paris 14eme is indicated as the actual food court. All other positions are prospected place for my client to open a new business.

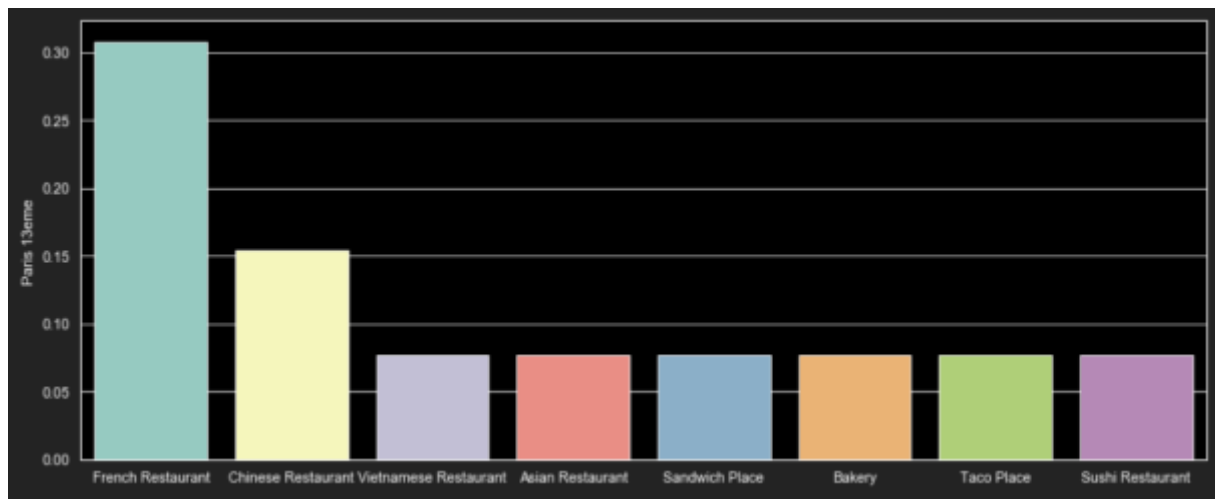
The blue circles represent the position of a place and the red circles represent the area in which we collected the restaurant categories present in the area.

Let's explore my client business neighborhood in Paris 14eme, and keep the 8 first categories of restaurant:



We can see it is mostly even almost half French Restaurant, and then Pizza place, Bakery, café and Thai restaurants.

Let's explore another place randomly in the list, still keeping only the 8 first categories:



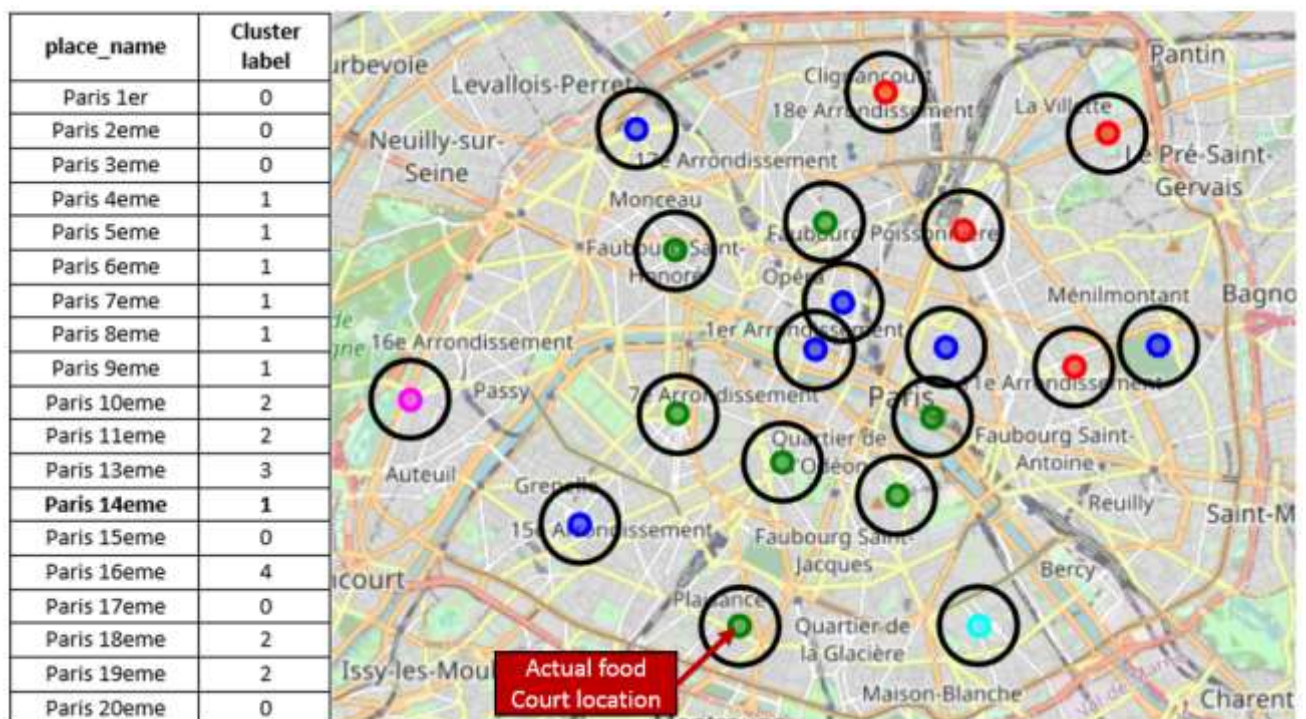
We can see that French restaurant stay first but after the other categories have changed compare to Paris 14eme.

There is not much data exploratory required more because the data are quite simple.

We will now proceed the K-Means clustering algorithm with a k-value arbitrary chosen at 5, in order to cluster the different places. Here we expect the actual business place to be cluster with other places (and not alone).

## 4. Results

Hereafter the resulting clusters:



We can see that the current business is in the green cluster which is composed of 7 places including the current business place.

## 5. Discussion

Based on the results of this study, foursquare might not be the best venue API to use in Paris as far it seems that not all venues are known by Foursquare. This study would have been more accurate using google. Map API.

## 6. Conclusion

In this study the aim was to find similar neighborhood as the one that contain the current food court of my client, based on its restaurant offer. The clustering method used here, permitted to cluster my client current business neighborhood with 6 other neighborhoods.

We can then conclude that among the 19 places prospected by my client, 6 of them would be a good place to open a new restaurant based on my client strategy explained in the introduction.