Opening a Japanese restaurant in Paris

Jean-Christophe Loeb

02/12/2020

Contents

1.Introduction

1.1 Background:

Nowadays more and more restaurants open every day in a big city like Paris. The competition is hard and many factors have to be taken in consideration before to start a new business. The location and the type of restaurant is particularly important. Indeed, in a rich and dense area we might have more chance to open a fancy restaurant. Immigration and tourism might also play an important role. The public transportations around the area can also bring more people in.

Analyzing the best suitable location is crucial to success.

1.2 Problem

For this research I have been hired by a Japanese chef that would like to open a Japanese restaurant in Paris.

As people in Paris mostly use public transportation, the restaurant should be not too far from a metro station (less than 200m). I decided then to only investigate the center of Paris (zone 1). My goal will be to find the metro station that don't have currently Japanese restaurants but have a high potential

To find the most suitable location I have chosen to focus on those points:

- Find out the current Japanese restaurants around the different metro station.
- For those Japanese restaurants, get the most common venues (other kind of restaurants, hotels, park….) around and trying to cluster the Japanese restaurants in order to find a pattern for a Japanese restaurant to success.
- Get the average salary for each district of Paris.
- Get the population for each district.

1.3 Interest:

This analysis will interest some Japanese chefs willing to open a business in Paris. They might want to know where are the current famous Japanese restaurants and if there are some zone in Paris which are suitable but currently does not have any famous Japanese restaurants. It will avoid them some harsh competition and give them the opportunity to grow. This analyze could also be easily adapted for other kind of restaurant and city.

2.Data acquisition and cleaning

2.1 Data Source:

- Get the different metro station of Paris and their geolocalisation. We can use https://en.wikipedia.org/wiki/List_of_Paris_M%C3%A9tro_stations and geopy library.
- Get the Japanese places already in places around those metro stations. Using foursquare API explore.
- Get the most common places around those Japanese restaurants also using the foursquare API explore

- Get the population for each district of Paris. We can use: https://en.wikipedia.org/wiki/Demographics_of_Paris
- Get the median salary for each district https://www.apur.org/observatoires_apur/familles/obs/3_3revenus/revenus_tab.htm

2.2 Data cleaning:

We retrieved first the data from 3 different web sources:

- For the population of each district in Paris, I just selected the right chart in the Wikipedia page. Then using slicing I selected only the rows we were interested in, and convert the population number in integer instead of string.

- For the mean salary for each district I used the apur webpage.

    - I selected the right chart and the right rows and columns I was interested in (only the district column and the mean salary).
    - Then I had to remove from the district columns the different letters to fit a number between 1 to 20 (the different district of Paris) and convert the column to int64. Ex: ‘1er’becoming 1.
    - From the salary column remove the character space and convert it in int64. Ex: ‘34 000’ becomes 34000.

- For the list of metro of Paris: I selected only the columns metro station, zone and district.
    - I selected only the rows where column zone=1.
    - I deleted the metro station rows that did not have for district a number. Ex if district column = ‘Levallois’.
    - If a metro station is associated with several districts we decided to only take the first one in the list. Ex: if column district is: 1,18 we will change it for 1.

3.Exploratory data analysis

3.1 Plan

The plan is the following:

- Take the list of metro and look for the GPS coordinates thanks to geopy.
- Explore common venues thanks to foursquare around those metro stations.
- Extract from this venue list the famous Japanese places.
- Explore common venues thanks to foursquare around those Japanese restaurants.
- Group each Japanese restaurants with the 5 most common venues around them.
- Run the clustering algorithm Kmean to find 3 groups.
- Analyzing those groups and try to see a pattern around famous Japanese places.

3.2 Detailed execution

Here are the Data I got after scraping and cleaning it from the different websites.

Demographics Data:

Arrondissement	Area (km2)	Population	Population per km2	Median salary	Salary normalisation	Population normalisation
1	1.826	17268	9457	47561	0.484570	0.000000
2	0.992	22558	22740	31413	0.208951	0.023674
3	1.171	36727	31364	38404	0.328275	0.087082
4	1.601	28068	17532	41225	0.376425	0.048332
5	2.541	61080	24038	52651	0.571448	0.196066
6	2.154	44154	20499	70965	0.884038	0.120320
7	4.088	58166	14228	77759	1.000000	0.183026
8	3.881	39409	10154	73493	0.927186	0.099085
9	2.179	60293	27670	44895	0.439066	0.192544
10	2.892	95436	33000	24950	0.098638	0.349815
11	3.666	156831	42780	26535	0.125691	0.624569
12	6.377	146527	22977	34289	0.258039	0.578457
13	7.146	184235	25782	27183	0.136752	0.747206
14	5.621	142535	25358	34966	0.269594	0.560592
15	8.502	240723	28314	44692	0.435601	1.000000
16	7.846	170239	21698	70532	0.876647	0.684572
17	5.669	171945	30331	39302	0.343603	0.692206
18	6.005	202780	33769	19171	0.000000	0.830198
19	6.786	187799	27674	19753	0.009934	0.763156
20	5.984	199113	33274	20861	0.028845	0.813788

Metro Data:

	Station	Zone	District
0	Abbesses	1	18
1	Alésia	1	14
2	Alexandre Dumas	1	11
3	Alma – Marceau	1	16
5	Anvers	1	9
...	...	...	...
296	Victor Hugo	1	16
300	Villiers	1	8
301	Volontaires	1	15
302	Voltaire	1	11
303	Wagram	1	17

242 rows × 3 columns

I used then geopy to add the latitude and longitude to the metro station data.

I used foursquare for the list of metro to explore venues around:

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}

Station	Latitude	Longitude	Name	Place latitude	Place longitude	Categories
Abbesses	48.884568	2.337929	Al Caratello	48.885248	2.336002	Italian Restaurant
Abbesses	48.884568	2.337929	Place des Abbesses	48.884406	2.338538	Plaza
Abbesses	48.884568	2.337929	Amorino	48.885056	2.336719	Ice Cream Shop
Abbesses	48.884568	2.337929	Chez Toinette	48.884224	2.336904	French Restaurant
Abbesses	48.884568	2.337929	Guilo Guilo	48.885942	2.337048	Japanese Restaurant

Then I took the Japanese restaurants from the previous list and use again the foursquare Api to explore venues around those restaurants.

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}

	Restaurant name	Latitude	Longitude	Name	Place latitude	Place longitude	Categories
0	Guilo Guilo	48.885942	2.337048	Boulangerie Alexine	48.886141	2.334477	Bakery
1	Guilo Guilo	48.885942	2.337048	La Boîte aux Lettres	48.886841	2.338186	Bistro
2	Guilo Guilo	48.885942	2.337048	Al Caratello	48.885248	2.336002	Italian Restaurant
3	Guilo Guilo	48.885942	2.337048	Amorino	48.885056	2.336719	Ice Cream Shop
4	Guilo Guilo	48.885942	2.337048	Chez Toinette	48.884224	2.336904	French Restaurant
...	...	...	...	...	...	...	...
552	Ebis	48.865375	2.332310	Nolinski	48.865367	2.334584	Hotel
553	Ebis	48.865375	2.332310	Brasserie Réjane	48.865486	2.334824	Restaurant
554	Ebis	48.865375	2.332310	Le Roch Hotel & Spa Paris	48.866200	2.332995	Hotel
555	Ebis	48.865375	2.332310	Hôtel Le Pradey	48.864459	2.331654	Hotel
556	Ebis	48.865375	2.332310	Cafés Verlet	48.864088	2.333878	Café

557 rows × 7 columns

I grouped each restaurant name and got the most 5 common venues around those restaurants.

Restaurant name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Ayama	Hotel	Japanese Restaurant	Bakery	Farmers Market	Fruit & Vegetable Store
Blueberry	French Restaurant	Clothing Store	Pastry Shop	Japanese Restaurant	Café
Bon Kushikatsu	Speakeasy	Coffee Shop	Japanese Restaurant	Italian Restaurant	Bar
:ôté Sushi Vaugirard	French Restaurant	Hotel	Gym	Japanese Restaurant	Plaza
Ebis	Hotel	Women's Store	Restaurant	Israeli Restaurant	Plaza
Eizosushi	French Restaurant	Bistro	Hotel	Japanese Restaurant	Supermarket
Fukuyama	Bistro	Tapas Restaurant	Organic Grocery	Vietnamese Restaurant	Sandwich Place
Garden Sushi	Plaza	Bookstore	Japanese Restaurant	Gourmet Shop	Korean Restaurant
Guilo Guilo	French Restaurant	Bakery	Italian Restaurant	Japanese Restaurant	Bistro
Himeji-Jō	Korean Restaurant	Italian Restaurant	Bakery	Hotel	French Restaurant
Hoki Sushi	French Restaurant	Plaza	Convenience Store	Bar	Japanese Restaurant
Jipangue	French Restaurant	Hotel	Brasserie	Brewery	Salad Place
Karaage-Ya Bourse	French Restaurant	Italian Restaurant	New American Restaurant	Pedestrian Plaza	Nightclub

We can already see that many Japanese restaurants are around French restaurants, hotels, coffee places and bakeries.

I decided then to use Kmean algorithm to get a better insight and cluster those Japanese restaurants into 3 groups

3.2 Kmean algorithm for clustering

I decided to cluster and find for the Japanese restaurants around what kind of other venues they usually are. When that will be done we could visualize them using folium map to display the famous Japanese restaurants, the common venues that are usually around Japanese restaurants and the metro station. Maybe we will find that around some metro station there is a potential for Japanese restaurant but currently does not have any. Then we could decide the best choice by looking at the average salary and population for the different metro stations.

First I created a one hot vector of the venue categories and added back the restaurant names.

Then I grouped all the places by restaurant name using the mean for the different categories

Restaurant name	African Restaurant	American Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Auto Dealership	Auvergne Restaurant	BBQ Joint	...	Tech Startup	Thai Restaurant	Theater	St
Ayama	0.0	0.0	0.0	0.000000	0.000000	0.00	0.000000	0.0	0.0	...	0.0	0.000000	0.00	0.0
Blueberry	0.0	0.1	0.0	0.000000	0.000000	0.00	0.000000	0.0	0.0	...	0.0	0.000000	0.00	0.0
Bon Kushikatsu	0.0	0.0	0.0	0.000000	0.000000	0.00	0.000000	0.0	0.1	...	0.0	0.000000	0.00	0.0
Côté Sushi Vaugirard	0.0	0.0	0.0	0.000000	0.000000	0.00	0.000000	0.0	0.0	...	0.0	0.000000	0.00	0.0
Ebis	0.0	0.0	0.0	0.000000	0.000000	0.00	0.000000	0.0	0.0	...	0.0	0.000000	0.00	0.0
Eizosushi	0.0	0.0	0.0	0.000000	0.000000	0.00	0.000000	0.0	0.0	...	0.0	0.000000	0.00	0.0
Fukuyama	0.0	0.0	0.0	0.000000	0.000000	0.00	0.000000	0.0	0.0	...	0.0	0.000000	0.00	0.0
Garden Sushi	0.0	0.0	0.0	0.000000	0.000000	0.00	0.000000	0.0	0.0	...	0.0	0.000000	0.10	0.0

Then I ran on this the Kmean algorithm using 3 different clusters.

4. Results

4.1 The clusters and their meaning.

After running Kmeans we can see three distinct clusters for our Japanese restaurants.

- Around French restaurants
- Around hotels
- Around other restaurants and bar (particularly around other Asian places)

Cluster 1:

Restaurant name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Label
Blueberry	French Restaurant	Clothing Store	Pastry Shop	Japanese Restaurant	Café	0
Guilo Guilo	French Restaurant	Bakery	Italian Restaurant	Japanese Restaurant	Bistro	0
Koko Bistro	French Restaurant	Multiplex	Bar	Historic Site	Canal Lock	0
Sola	French Restaurant	Bookstore	Church	Scenic Lookout	Bakery	0
Wanobi	French Restaurant	Bar	Tech Startup	Bagel Shop	Bookstore	0
Yoshi	French Restaurant	Health Food Store	Thai Restaurant	Bookstore	Chocolate Shop	0
Yoshida	French Restaurant	Italian Restaurant	Bakery	Sandwich Place	Café	0
Yuzu	French Restaurant	Café	Bakery	Historic Site	Salad Place	0

Cluster 2:

Restaurant name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Label
Ayama	Hotel	Japanese Restaurant	Bakery	Farmers Market	Fruit & Vegetable Store	1
Ebis	Hotel	Women's Store	Restaurant	Israeli Restaurant	Plaza	1
Jipangue	French Restaurant	Hotel	Brasserie	Brewery	Salad Place	1
Kiku	Hotel	Candy Store	Karaoke Bar	Restaurant	Corsican Restaurant	1
Kinugawa Vendôme	Hotel	French Restaurant	Women's Store	Israeli Restaurant	Dessert Shop	1
Miss Kō	Hotel	French Restaurant	Japanese Restaurant	Boutique	Pastry Shop	1
Otaku	Hotel	Japanese Restaurant	Gym / Fitness Center	Pizza Place	Historic Site	1
Sushi Star	Hotel	Japanese Restaurant	Sandwich Place	Art Gallery	Auto Dealership	1

Cluster 3:

Restaurant name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Label
Bon Kushikatsu	Speakeasy	Coffee Shop	Japanese Restaurant	Italian Restaurant	Bar	2
Garden Sushi	Plaza	Bookstore	Japanese Restaurant	Gourmet Shop	Korean Restaurant	2
Himeji-Jō	Korean Restaurant	Italian Restaurant	Bakery	Hotel	French Restaurant	2
Kintaro	Japanese Restaurant	Bookstore	Pastry Shop	Supermarket	French Restaurant	2
La Maison du Saké	Chinese Restaurant	French Restaurant	Spa	Japanese Restaurant	Sushi Restaurant	2
Mushimushi	African Restaurant	Asian Restaurant	Spanish Restaurant	Steakhouse	Restaurant	2
Nakagawa	Restaurant	Bakery	Chinese Restaurant	Indian Restaurant	Bar	2
Nana-Ya	Cheese Shop	Japanese Restaurant	Burger Joint	Bakery	French Restaurant	2
Osaka Sushi	Bakery	Chinese Restaurant	Café	Bistro	Rock Club	2
Otakuni	Indian Restaurant	Asian Restaurant	Cosmetics Shop	Café	Japanese Restaurant	2

I will focus more on clusters 1 and 2 as they seem to be very clear, Japanese restaurants are often around French restaurants and hotels.

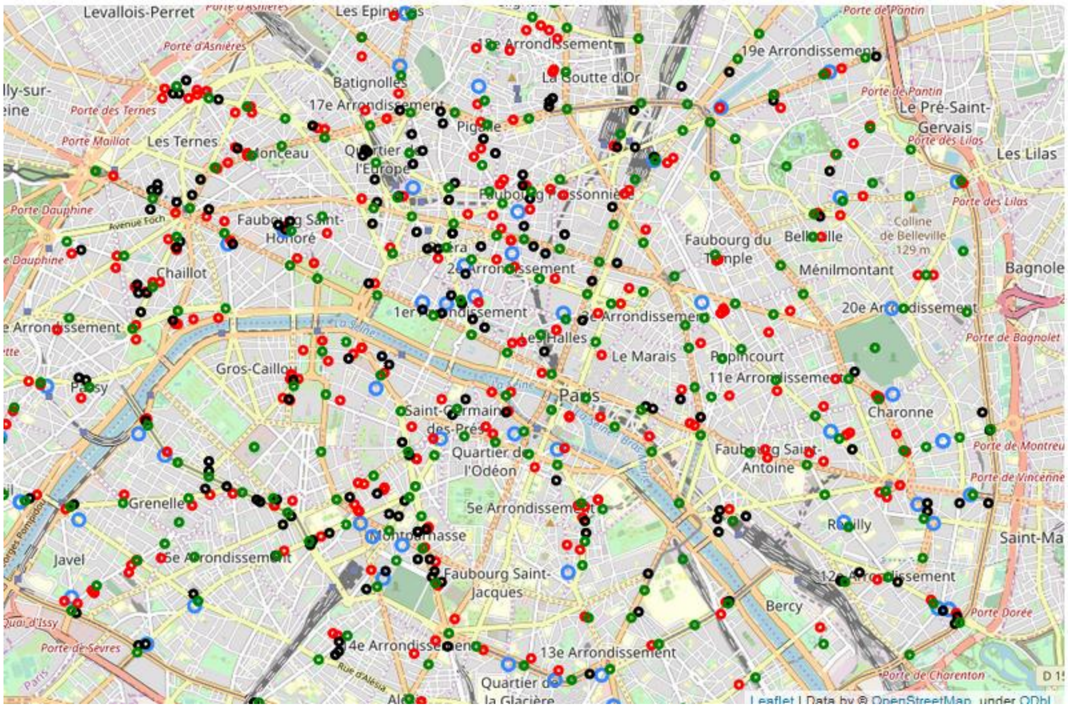
4.2 Map visualization

I decided them to check on a map to see the Japanese restaurants with the French restaurants and hotels alongside with metro stations.

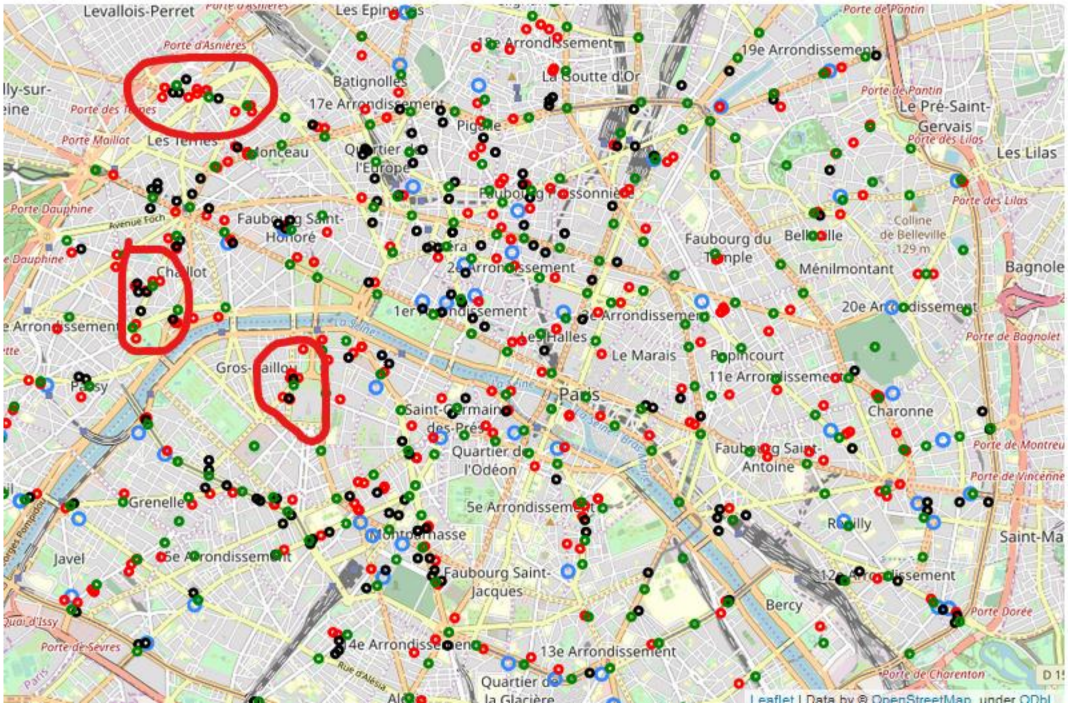
We visualize on a folium map:

- The famous Japanese restaurant in blue.

- The famous French restaurant in black.
- The famous hotel in red.
- The metro station in green.



We can see few spots that could be interesting for us (meaning French restaurant, hotel and metro around but no Japanese restaurant yet):



Those areas are around the metro station: **boissiere**, **porte de Champerret**, **La Tour-Maubourg.**

4.3 Checking the demographics of those districts to refine

Checking the population and the median salary for those metro station could help us refine where we would like to open the restaurant.

Arrondissement	Area (km2)	Population	Population per km2	Median salary	Salary normalisation	Population normalisation	Station
7	4.088	58166	14228	77759	1.000000	0.183026	La Tour-Maubourg
16	7.846	170239	21698	70532	0.876647	0.684572	Boissière
17	5.669	171945	30331	39302	0.343603	0.692206	Porte de Champerret

It seems that **Boissiere** metro station might be a good choice as the salary is pretty high and the population too.

5.Conclusion

Thanks to the data collected and the Kmeans clustering algorithm we found out that most of the famous Japanese restaurants can be clustered into three groups: Around French restaurants, around hotels and around other restaurants and bars.

I would propose to the Japanese chef if he wants to avoid hard competition to try to settle around **Boissiere** metro station as there seem to be many French restaurant and hotel around but no famous Japanese yet and in this district people seems to have a pretty good salary in comparison. Of course there are a lot of factors I did not take in consideration in this project. Here are some factors we could have checked:

- We could have seen the immigrations by nationalities and by district.
- The price for renting a place in each district.
- The different age categories for each districts

Thanks for your time reading my project.