Capstone Project - The Battle of Neighborhoods

Predicting Optimal location in Paris for a new Japanese restaurant

September 20, 2019

CAMARA Idrissa

1. Introduction

1.1 Background

My brother is a restorer and has a Japanese restaurant in the north of Paris (France). He wants to expand his business and create a chain of Japanese restaurants in France at first and then in Europe. For his second restaurant, he stays in Paris and he would like to detect the optimal location for this second restaurant. He would like to be in a neighborhood similar to the first restaurant area with some constraints that are:

- low Japanese restaurants in vicinity (< 5)
- neighborhood with two middle school at least
- neighborhood with 3 bike station at least
- neighborhood close to city center as possible

1.2 Problem and Interest

The challenge is to find the optimal location for a japanese restaurant in a city like Paris or New York. I believe this is a relevant challenge with valid questions for anyone want to have a chain of restaurants in to other large city in US, EU or Asia. The same methodology can be applied with other type of restaurant.

2. Data acquisition and cleaning

2.1 Data acquisition
The following data is required to answer the issues of the problem:

- List of Boroughs and neighborhoods of Paris with their geodata (latitud and longitud)
- List of middle schools Paris with their geodata (latitud and longitud)
- List of bike station in Paris with their address location
- List of existing japanese restaurants in the neighborhood
- distance of neighborhood from city center

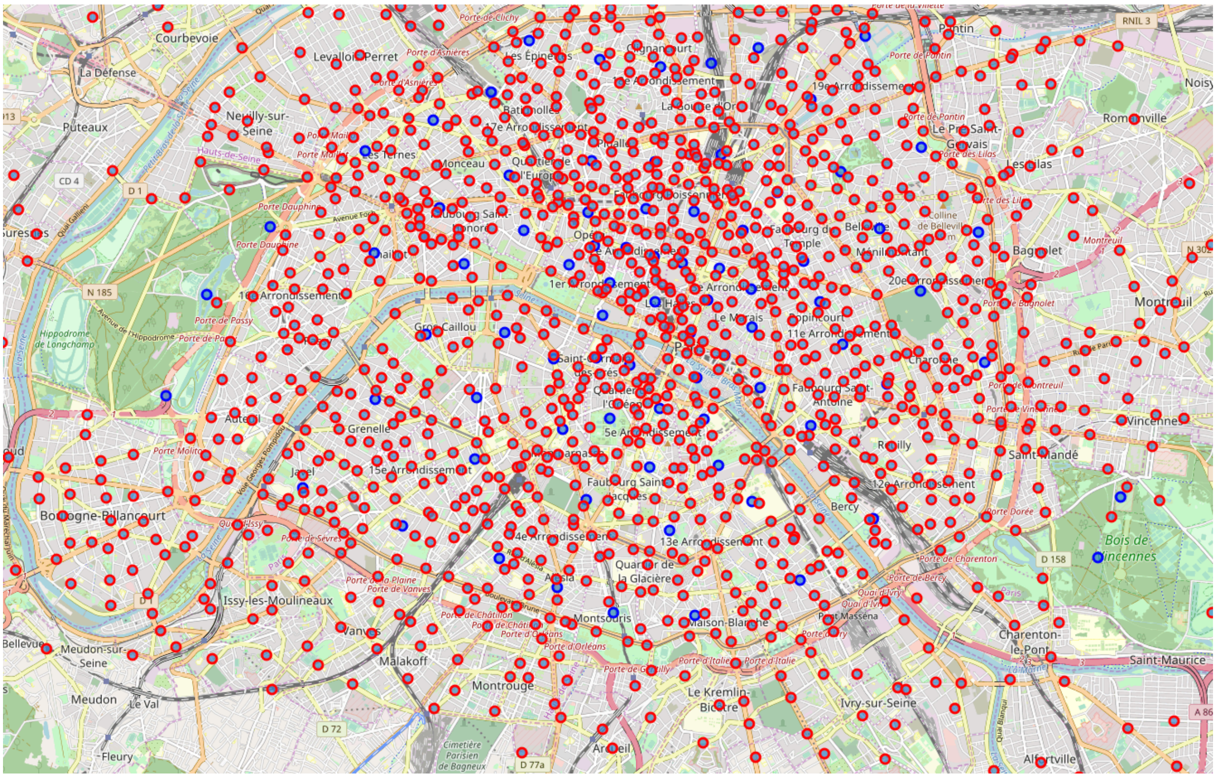
Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Foursquare List of japanese restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- coordinate of Paris center will be obtained using Foursquare
- List of Boroughs and neighborhoods of Paris with their geodata (latitud and longitud) are available for free at the Paris City Hall website (https://opendata.paris.fr/explore/dataset/quartier_paris/table/)
- List of middle schools Paris with their geodata (latitud and longitud) are available for free at the Paris City Hall website (https://opendata.paris.fr/explore/dataset/secteurs-scolaires/table/?disjunctive.id_projet&disjunctive.zone_commune&disjunctive.annee_scol)
- List of bike station in Paris with their address location are available for free at the Paris City Hall website (https://opendata.paris.fr/explore/dataset/velib-emplacement-des-stations/table/)

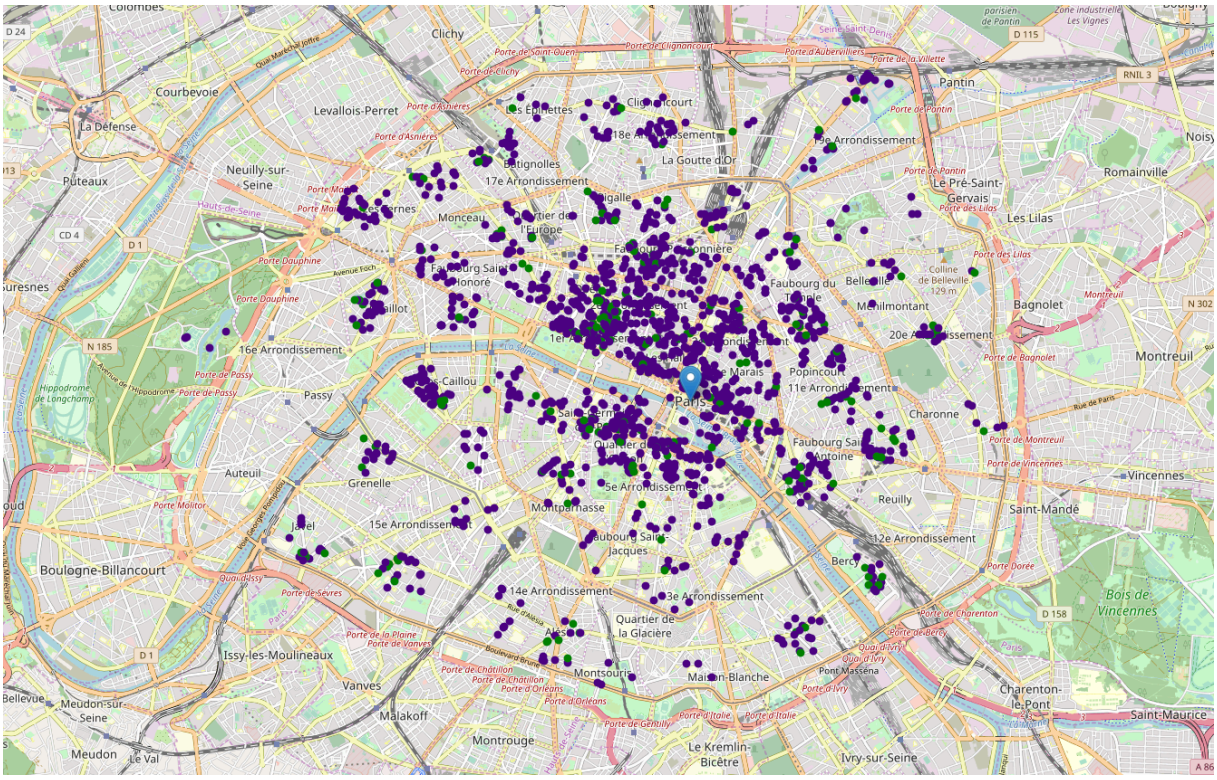
2.2 Data cleaning
Data downloaded from Paris City Hall website sources into CSV format.
There were a lot of missing values but these dont affect our analysis. Fortunately, I have latitude and longitude in these data but these are in string format so, I decided convert into float before any jobs.

3. Exploratory Data Analysis

Paris neighborhood (blue) and bike station (red)



Paris Restaurants (Indigo) and Paris Japanese restaurant (green)



Calculation of nomber of japanese restaurant in neighborhood

Out[34]:

	L_QU	Latitude	Longitude	Japanese Restaurants in area
0	Saint-Gervais	48.855719	2.358162	4
1	Chaussée-d'Antin	48.873547	2.332269	0
2	Saint-Ambroise	48.862345	2.376118	2
3	Necker	48.842711	2.310777	0
4	Sainte-Avoie	48.862557	2.354852	4
5	Rochechouart	48.879812	2.344861	2
6	Folie-Méricourt	48.867403	2.372965	6
7	Saint-Fargeau	48.871035	2.406172	1
8	Batignolles	48.888482	2.313856	4
9	Val-de-Grâce	48.841684	2.343861	2

Calculation of nomber of Bike station and Middle schools in neighborhood

Out[50]:

	L_QU	Latitude	Longitude	Japanese Restaurants in area	Number Byke Station	Number Middle Schools
1	Chaussée-d'Antin	48.873547	2.332269	0	9	4
2	Saint-Ambroise	48.862345	2.376118	2	6	8
3	Necker	48.842711	2.310777	0	4	5
5	Rochechouart	48.879812	2.344861	2	8	15
7	Saint-Fargeau	48.871035	2.406172	1	5	5
11	Plaisance	48.830317	2.315305	0	5	8
13	Saint-Victor	48.847664	2.354093	2	6	6
18	Mail	48.868008	2.344699	2	8	9
22	Grenelle	48.850172	2.291853	3	4	8
23	Saint-Germain-l'Auxerrois	48.860650	2.334910	1	4	5
28	Gare	48.827527	2.372398	3	3	13
29	Pont-de-Flandre	48.895556	2.384777	2	5	5
30	Plaine de Monceaux	48.885044	2.302910	2	3	5
31	Archives	48.859192	2.363205	3	3	9
32	Invalides	48.858515	2.316445	0	4	4
34	Combat	48.878639	2.380127	0	4	2
35	Montparnasse	48.837623	2.331784	0	3	4
36	Saint-Merri	48.858521	2.351667	0	10	7
37	Notre-Dame	48.852896	2.352775	1	6	5
39	Champs-Elysées	48.867074	2.308652	2	5	4
40	Saint-Vincent-de-Paul	48.880735	2.357471	1	6	12
41	Hôpital-Saint-Louis	48.876008	2.368123	2	8	11
45	Porte-Saint-Denis	48.873618	2.352283	1	5	11
50	Ternes	48.881178	2.289964	2	5	11
52	Amérique	48.881638	2.395440	0	3	11
53	Belleville	48.871531	2.387549	1	6	14
55	Porte-Saint-Martin	48.871245	2.361504	0	8	9
56	Maison-Blanche	48.823128	2.352433	0	4	7
59	Arsenal	48.851585	2.364768	2	3	9
62	Charonne	48.854760	2.407430	1	5	17

4. Methodologiy

In this project we will direct our efforts on detecting areas of Paris that have low japaneses restaurant density (< 5), close to two middle school at least and close to one bike station at least. Given a neighborhood, we will limit our analysis to area ~1km around neighborhood center.

In first step we have collected the required data: location and type (category) of every restaurant within Paris, middle school in Paris and bike station in Paris. We have also identified Japanese restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration,

- number of Japanese restaurants in 1km arround the center each neighborhood.
- number of middle school in 1km arround the center each neighborhood.
- number of bike station in 1km arround the center each neighborhood

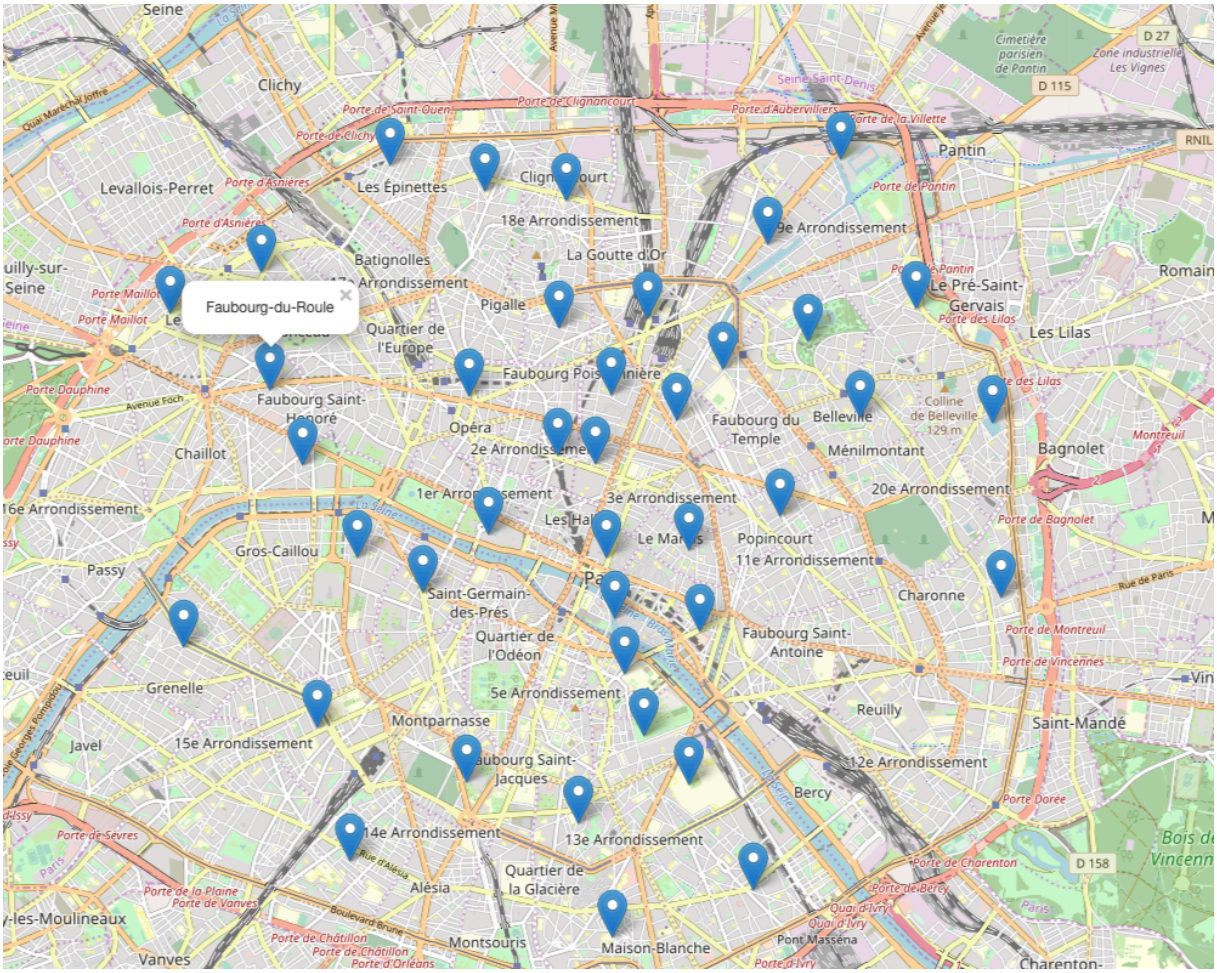
In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without Japanese restaurants in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

5. Results and Discussion

Our analysis shows us a very high number of restaurants in Paris and a very high concentration of Japanese restaurant in the center of Paris. We then focused on the outermost neighborhoods while staying in Paris. We held the neighborhoods with a low number of Japanese restaurants (<= 3), a large number of bike stations (> = 3) and with a number of colleges also raised to give more chance to our future Japanese restaurant.

First of all, we retrieved all the data on a site set up by the city of Paris (data are available for free). We then proceeded to a data processing before retrieving the list of restaurants (especially Japanese restaurants) with foursquare. The data collected on bike stations and schools, also allowed us to make an aggregation to determine the optimal location for our next restaurant.

Result of all this is 40 zones containing largest number of potential new restaurant locations based on number of and distance to existing japanese restaurant - both restaurants in general and Japanese restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on neighborhoods in Paris it is entirely possible that there is a very good reason for small number of restaurants in any of those neighborhoods, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.


Faubourg-du-Roule
Seine
Clichy
Levallois-Perret
Les Épinettes
18e Arrondissement
9e Arrondissement
Batignolles
La Goutte d'Or
Pigalle
Quartier de l'Europe
Faubourg Saint-Honoré
Opéra
Faubourg du Temple
Belleville
Colline de Belleville 129 m
Ménilmontant
Bagnolet
Chaillot
20e Arrondissement
3e Arrondissement
Popincourt
11e Arrondissement
Gros-Caillou
Passy
Saint-Germain-des-Prés
Quartier de l'Odéon
Faubourg Saint-Antoine
Charonne
Grenelle
5e Arrondissement
Montparnasse
Reuilly
Saint-Mandé
Javel
15e Arrondissement
12e Arrondissement
Bercy
14e Arrondissement
13e Arrondissement
Quartier de la Glacière
Alésia
Montsouris
Maison-Blanche
Vanves
Pantin
Le Pré-Saint-Gervais
Les Lilas
Romainville
Porte de Clignancourt
Porte de Saint-Ouen
Porte de Clichy
Porte d'Asnières
Porte Maillot
Porte Dauphine
Avenue Foch
Porte d'Aubervilliers
Porte de la Villette
Porte de Pantin
Porte des Lilas
Porte de Bagnolet
Porte de Montreuil
Porte de Vincennes
Porte Dorée
Porte de Charenton
Porte de Bercy
Quai d'Ivry
Pont Masséna
Porte d'Ivry
Porte de Sèvres
Porte de la Plaine
Porte de Vanves
Porte de Châtillon
Porte d'Orléans
Boulevard Brune
Rue d'Alésia
Rue de Paris
Cimetière parisien de Pantin
Zone industrielle Les Vignes
D 27
D 115
D 158
RNIL
Bois de Vincennes

6. Conclusion

Purpose of this project was to identify Paris neighborhood close to 2 middle school at least, close to 3 bike station at least and with low number of japanese restaurants in order to aid stakeholders in narrowing down the search for optimal location for a new Japanese restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general neighborhood that justify further analysis (external of Paris), and then generated extensive collection of locations which satisfy some basic requirements. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.