Linkedin : https://fr.linkedin.com/in/david-hinh-75577b137

Email : david_hinh@hotmail.fr

Github : https://github.com/github-dhh/Applied-Capstone-Project-IBM

COURSERA IBM DATA SCIENCE

APPLIED CAPSTONE PROJECT: FINDING THE BEST LOCATION TO OPEN AN IZAKAYA IN PARIS

Photo from Japan-guide.com

By David HINH
18/07/2020

SUMMARY

I) INTRODUCTION

In this study, we will determine the ideal location to open an izakaya in Paris. Izakaya is concept from Japan, which is close to the concept of tapas bar well known in the West. It is a place where you can order foods and drink. People usually meet there in the evening with friends or colleagues in a relaxed atmosphere.

We will make some clusters of neighborhoods in Paris with the K-means methods to determine the best location to open an izakaya in Paris.

Opening a restaurant is a project that requires a lot of work and investment. It requires a preliminary study to maximize the chances of success of the structure. The study of the location and the customers are necessary.

a) Location survey

→ Choose a place where the competition is well "balanced": The area should not have a huge number of Japanese restaurants to avoid competition, but it should also not be empty of restaurants. A group of restaurants makes the area more attractive for customers.

→ Evaluate customer flow in the area.

→ Learn about the competition of the area: Get to know more about the successful restaurants in the area and propose a different offer to avoid direct competition.

b) Customers survey

→ Check the profile of the possible customers in the area to determine if the location is good for the izakaya concept.

c) Hypothesis

For this study, some hypothesis will be made:

→ The main target group for this type of restaurant is mainly the young working population between 18 and 35 years of age. the izakaya should be open to all financially, have an affordable price range.

→ Tripadvisor data will be used for our study. Let's suppose that the data on this ranking website is trustworthy.

→ Based on the Tripadvisor data, a famous restaurant will be defined by its number of reviews and its rating on Tripadvisor. A restaurant is famous when its rating is higher than 4, and its number of reviews higher than 30.

II) NEEDED DATA

→ A dataset of the Japanese restaurants in Paris with the following information:

- Name of the restaurants
- Address
- Price Range
- Number of reviews
- Reviews
- Rating
- Latitude
- Longitude.

This dataset will help us to get to know better the competition.

→ Information about the population of every district in Paris and their profiles:

- Average wage for households
- Age
- Number of inhabitants per district / Density of population

These data will help us to identify the possible customers.

→ Returned location data by Foursquare API or others to get information regarding the Japanese restaurants in the neighborhoods

III) GETTING & CLEANING DATA

a) GETTING DATA

On Kaggle, some datasets are shared by users for data analysis. I found one dataset published, where 20 000+ restaurants were extracted from Tripadvisors in 2018. The dataset is not directly usable and need some cleaning. Also, there is no restaurants' addresses. We will have to complete the data by web scrapping.

URL of the shared dataset:

https://www.kaggle.com/damienbeneschi/krakow-ta-restaurans-data-raw

i) Address extraction

The first step will be to scrap all the restaurants' addresses in Paris and match the data with the existing database shared on Kaggle.

The first difficulty of restaurants' addresses scraping is to get all the URL of all restaurants in Paris. One way to do that is to visit the link for robots http://www.tripadvisor.com/robots.txt and find the general sitemap. then download all the sub website related to french restaurants "fr-restaurant_review". Download all the gz files and keep only the links related to restaurants in Ile de France "Paris_Ile_de_France.html".

ii) Latitude and longitude coordinates

To convert addresses into latitude and longitude coordinates, I used this following french website: https://adresse.data.gouv.fr/csv

It converts CSV files with addresses data into latitude and longitude coordinates.

iii) Information related to the customers profile

All the information related to the customers profile such as average wage for households for each district in Paris, number of population are found on the french website: https://insee.fr/

(The National Institute of Statistics and Economic Studies collects, produces, analyses and disseminates information on the French economy and society).

	Postal Code	District Average Wage Per Household €	District population 2018
0	75001	2862	16545
1	75002	2724	20796
2	75003	2803	35049
3	75004	2733	27146
4	75005	2972	59333
5	75006	3515	42428
6	75007	3896	54133
7	75008	3739	36694
8	75009	2974	59408
9	75010	2195	91770
10	75011	2312	149834
11	75012	2343	142340
12	75013	2023	183216
13	75014	2394	139992
14	75015	2695	234994
15	75016	3509	165487
16	75017	2670	168533
17	75018	1766	197580
18	75019	1552	185654
19	75020	1721	195556

Fig 1 - Average Wage per households & number of inhabitants per district

Age	75001	75002	75003	75004	75005	75006	75007	75008	75009	75010	75011	75012	75013	75014	75015	75016	75017	75018	75019	75020
0-14	11.8	12	11.4	11.4	11.4	10.2	12.1	15.1	14	14.9	12	13.4	13.7	12.6	13.6	14.6	14.3	14	17	15.1
15-29	22.1	26.2	25.9	24.5	29.2	27.6	23.6	23.7	25	23.5	24.7	22.7	23.7	26.3	24.2	20.8	23.9	23.5	21.9	20.9
30-44	22.7	27.3	25.2	21.6	17.3	16.4	17.3	20.2	25.4	26.7	25.4	22.4	20.3	19.9	21.4	17.1	23.1	26.5	21.7	23.1
45-49	20.3	19.8	19.2	19.3	16.9	16.9	18.2	18.7	17.8	18.5	17.8	18.5	19	17.2	17.6	19.7	17.9	18	19.2	20
60-74	14.9	10.6	12.5	15.3	15.5	17.3	17.3	14.8	11.8	11.4	13.6	14.3	15.2	15.1	14.3	16.3	13.5	12.2	13.7	14.4
75+	8.1	4.1	5.8	7.8	9.5	11.5	11.5	7.6	6	4.9	6.5	8.7	8.1	8.9	8.9	11.5	7.3	5.8	6.5	6.5

Fig 2 - Age of the population

b) Cleaning Data

The shared database on Kaggle have to be cleaned to be usable.

	Unnamed: 0	Name	City	Cuisine Style	Rating	Price Range	Number of Reviews	Reviews
0	0	La Meduse	Paris	French - Seafood - European - Vegetarian Frien...	5.0	— $	178.0	You cant figure out from the outside the
1	1	Le Cappiello	Paris	French - Mediterranean - European - Contempora...	5.0	— $	208.0	Incredible! - Amazing food - wonderful atmosph...
2	2	ASPIC	Paris	French - European - Contemporary	5.0		427.0	Second time and just as good - Best dinner in ...
3	3	Les Apotres de Pigalle	Paris	South American - Brew Pub - European - Vegetar...	5.0	— $	1152.0	Wonderful culinary experience - Must go restau...
4	4	Epicure	Paris	French - European - Vegetarian Friendly - Vega...	5.0		2305.0	Very nice place - Speechless

Fig 3 – First cleaning database from Kaggle

c) Merging data

Using the name of the restaurants to merge both existing datasets and selecting only Japanese restaurants in Paris.

	Name	Borough	Neighborhood	Address	Insee Code	Latitude	Longitude
0	La Cucina	Paris 15e	Boulevard Garibaldi	36 Boulevard Garibaldi 75015 Paris	75115.0	48.846662	2.305116
1	Le Reciproque	Paris 18e	Rue Ferdinand Flocon	14 Rue Ferdinand Flocon 75018 Paris	75118.0	48.891625	2.346073
3	Hideout Montparnasse FERMÉ	Paris 14e	Rue de la Gaîté	33 Rue de la Gaîté 75014 Paris	75114.0	48.839248	2.323327
4	Umami Matcha Café	Paris 3e	Rue Béranger	22 Rue Béranger 75003 Paris	75103.0	48.866528	2.363535
5	Le Pré Salé	Paris 13e	Rue du Moulin des Prés	22 Rue du Moulin des Prés 75013 Paris	75113.0	48.828559	2.352699

Fig 4 –Database with latitude & longitude coordinates

	Name	City	Cuisine Style	Rating	Price Range	Number of Reviews	Reviews	Borough	Neighborhood	Address	Insee Code	Latitude	Longitude
941	Tonkatsu Tombo	Paris	Japanese	4.5	– $	117.0	Katsu don is a must! - Great value for money	Paris 15e	Rue de l'Arrivée	14 Rue de l'Arrivée 75015 Paris	75115.0	48.843298	2.321696
2695	Okomusu Paris	Paris	Japanese	4.0	– $	92.0	Nice discovery - Delicious meal!	Paris 3e	Rue Charlot	11 Rue Charlot 75003 Paris	75103.0	48.861460	2.360837
2806	Kazoku	Paris	Japanese	4.0	$	90.0	Great price - good quantity but adequate qu	Paris 14e	Rue de la Tombe Issoire	105 Rue de la Tombe Issoire 75014 Paris	75114.0	48.826053	2.331825
3138	Le Feu de Mars	Paris	Japanese	4.0	– $	82.0	Just great food! - great teppanyaki - cheap price	Paris 13e	Rue Vandrezanne	41 Rue Vandrezanne 75013 Paris	75113.0	48.827681	2.353186
4672	Izakaya Taisho Ken 3	Paris	Japanese	4.0	$	73.0	Simply Stunning - Disappointing	Paris 1er	Rue Sainte-anne	11 Rue Sainte-anne 75001 Paris	75101.0	48.866043	2.335302

Fig 5 – Final database after merging and selecting only Japanese restaurants data

IV) LOCATION OF JAPANESE RESTAURANTS IN PARIS

First, we use the folium library to plot all the Japanese restaurants from the final database on a map. Japanese restaurants are grouped together and are spread all over Paris. some places may still seem empty of Japanese restaurants.

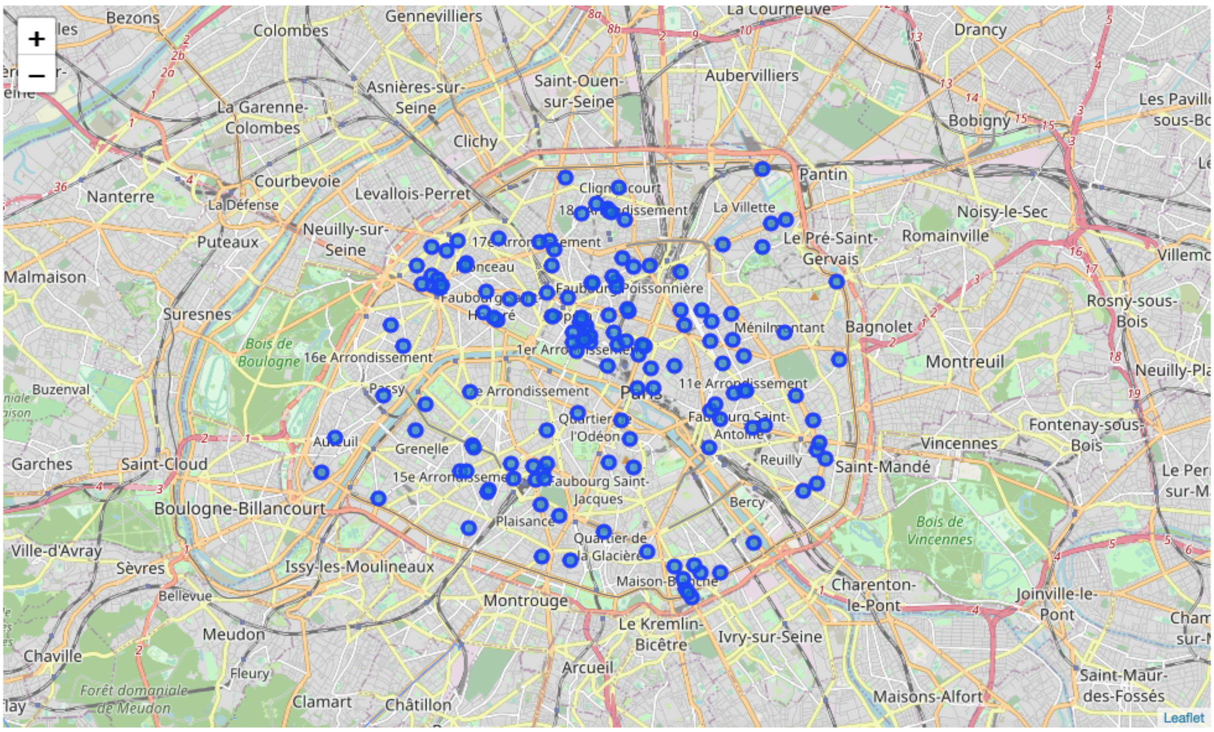

Fig 6 – Location of the Japanese restaurants in Paris

To refine the analysis, we are going to use the Foursquare API to explore the neighborhoods and segment them. First, we will get the mean frequency of Japanese restaurants in different neighborhoods of Paris.

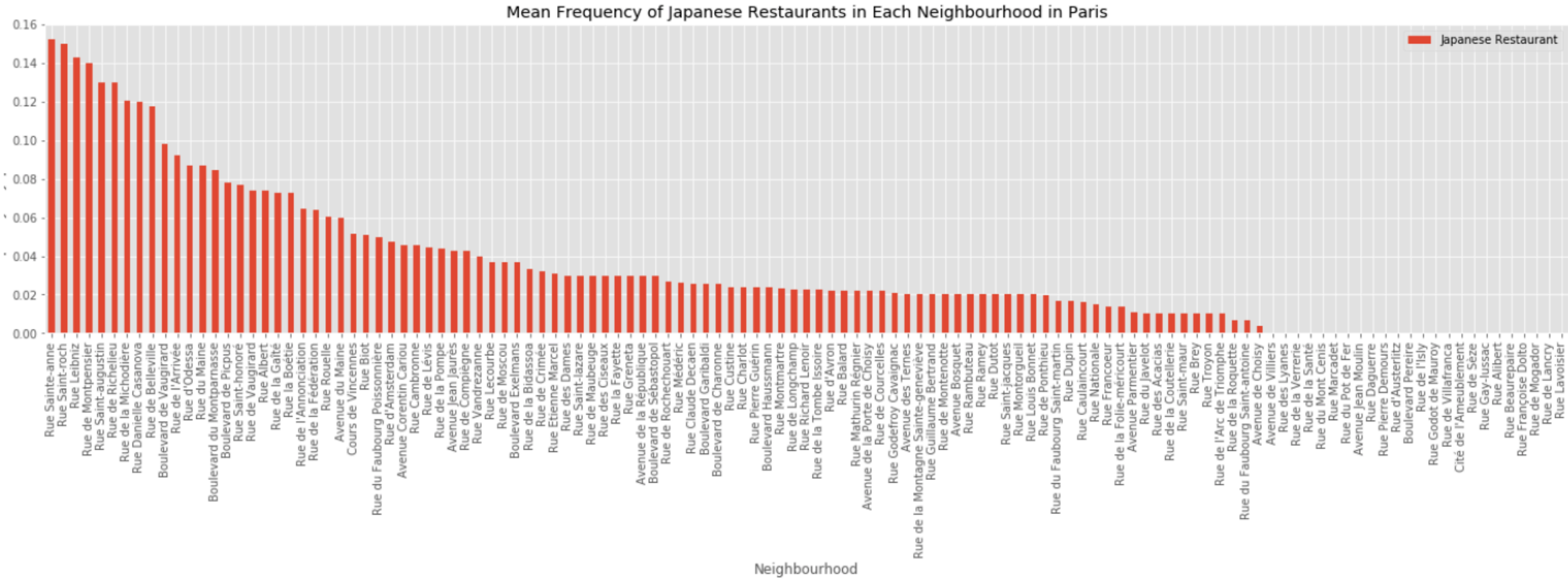

Fig 7 – Mean frequency of Japanese restaurants in each neighborhood in Paris

The distribution of Japanese restaurants is not the same in all Paris districts. Rue Saint Anne has the most Japanese restaurants with 0.15 mean frequency of Japanese restaurants. in our interest, the izakaya opening will take place in a neighborhood where the competition is average.

V) CLUSTERING WITH K-MEANS METHOD

3 criteria will be used to do the clusters for finding potential customers:

- The spending power of the residents: We are looking for people who can afford leisure activities such as going to restaurants. Izakaya must be a place where everyone can go even on a tight budget. Their spending power should ideally be mid/high.

- The number of inhabitants in the district: We are looking for a place where the number of residents is high. Because of the lack of available data on the internet, the population number in the district will be used to assess the customers flow instead of the density of population.

- The competitiveness of the location. The competition must be balanced. The area should not have a huge number of Japanese restaurants to avoid competition, but it should also not be empty of restaurants. A group of restaurants makes the area more attractive for customers.

We have to create a database with the following criteria:

	Neighborhood	latitude	longitude	Postal Code	District Average Wage Per Household €	District population 2018	Japanese Restaurant
0	Rue de l'Arrivée	48.843690	2.322940	75015	2695	234994	0.092105
1	Rue de Vaugirard	48.844827	2.319498	75015	2695	234994	0.074074
2	Avenue du Maine	48.844752	2.320214	75015	2695	234994	0.060000
3	Rue Mathurin Régnier	48.840948	2.306902	75015	2695	234994	0.022222
4	Rue de Villafranca	48.831845	2.303805	75015	2695	234994	0.000000

Fig 8 – Dataset for clustering

Normalizing data makes it possible to compare several data with different orders of magnitude.

	Cluster Label	Neighborhood	latitude	longitude	Postal Code	District Average Wage Per Household €	District population 2018	Japanese Restaurant	Normalized Household Income	Normalized Number of Residents	Normalized Number of Japanese Restaurants
0	5	Rue de l'Arrivée	48.843690	2.322940	75015	2695	234994	0.092105	0.217326	1.455922	1.621794
1	5	Rue de Vaugirard	48.844827	2.319498	75015	2695	234994	0.074074	0.217326	1.455922	1.122245
2	5	Avenue du Maine	48.844752	2.320214	75015	2695	234994	0.060000	0.217326	1.455922	0.732327
3	4	Rue Mathurin Régnier	48.840948	2.306902	75015	2695	234994	0.022222	0.217326	1.455922	-0.314295
4	4	Rue de Villafranca	48.831845	2.303805	75015	2695	234994	0.000000	0.217326	1.455922	-0.929956

Fig 9 – Dataset with normalized data

Let’s select the best K coefficient to do the K-Means clustering. To do that, we will plot the K value against the square error cost.

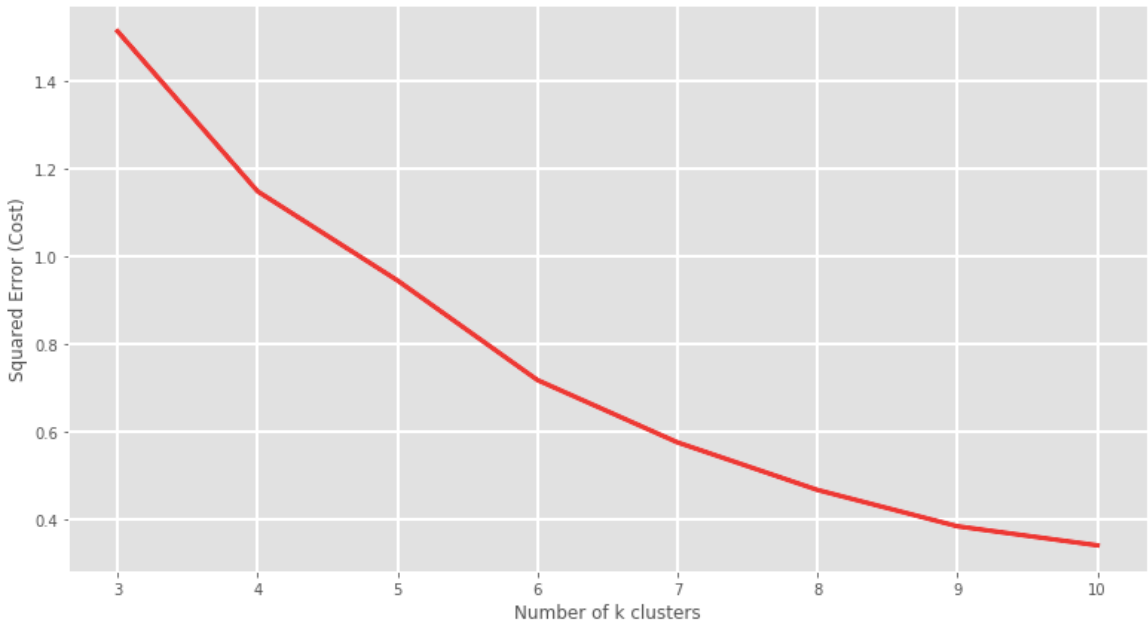

Fig 10 – K value against the square error cost

Taking a very small K value will not minimize the cost function. K = 6 seems to be a good value.

The results of the clustering are plotted on a map:

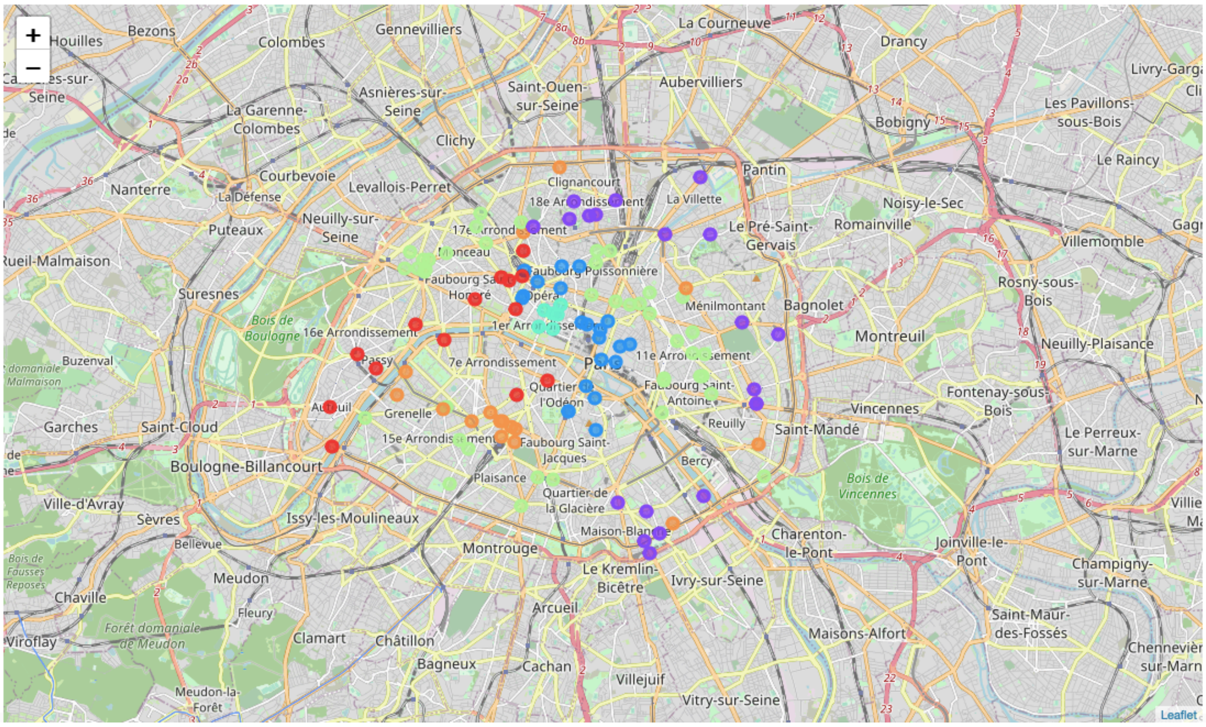

Fig 11 – Japanese restaurants clustering

VI) EXAMINING THE CLUSTERS

Cluster 0:

	Cluster Label	Neighborhood	latitude	longitude	Postal Code	District Average Wage Per Household €	District population 2018	Japanese Restaurant	Normalized Household Income	Normalized Number of Residents	Normalized Number of Japanese Restaurants
82	0	Avenue Bosquet	48.861594	2.301895	75007	3896	54133	0.020000	2.413963	-1.118156	-0.375861
98	0	Rue de Ponthieu	48.870005	2.311514	75008	3739	36694	0.019841	2.126809	-1.366354	-0.380259
99	0	Rue Saint-honoré	48.867928	2.324380	75008	3739	36694	0.076923	2.126809	-1.366354	1.201176
100	0	Rue de Moscou	48.879746	2.326561	75008	3739	36694	0.037037	2.126809	-1.366354	0.096145
101	0	Rue la Boétie	48.874466	2.319677	75008	3739	36694	0.072464	2.126809	-1.366354	1.077632
102	0	Rue Lavoisier	48.873690	2.322177	75008	3739	36694	0.000000	2.126809	-1.366354	-0.929956
103	0	Rue de l'Isly	48.874762	2.326420	75008	3739	36694	0.000000	2.126809	-1.366354	-0.929956
106	0	Rue de l'Annonciation	48.855642	2.280771	75016	3509	165487	0.064516	1.706138	0.466674	0.857445
107	0	Rue de Longchamp	48.864798	2.293033	75016	3509	165487	0.022727	1.706138	0.466674	-0.300303
108	0	Rue de la Pompe	48.858518	2.274850	75016	3509	165487	0.044118	1.706138	0.466674	0.292311
109	0	Boulevard Exelmans	48.839650	2.266909	75016	3509	165487	0.037037	1.706138	0.466674	0.096145
110	0	Rue Pierre Guérin	48.847761	2.266302	75016	3509	165487	0.023810	1.706138	0.466674	-0.270320
114	0	Rue des Ciseaux	48.853243	2.334129	75006	3515	42428	0.030000	1.717112	-1.284746	-0.098814
115	0	Rue Dupin	48.850360	2.324801	75006	3515	42428	0.016667	1.717112	-1.284746	-0.468211

Cluster 1:

	Cluster Label	Neighborhood	latitude	longitude	Postal Code	District Average Wage Per Household €	District population 2018	Japanese Restaurant	Normalized Household Income	Normalized Number of Residents	Normalized Number of Japanese Restaurants
34	1	Avenue de Choisy	48.820245	2.364328	75013	2023	183216	0.003546	-1.011766	0.718999	-0.831712
35	1	Rue Vandrezanne	48.828192	2.356282	75013	2023	183216	0.040000	-1.011766	0.718999	0.178233
36	1	Rue Françoise Dolto	48.829385	2.382952	75013	2023	183216	0.000000	-1.011766	0.718999	-0.929956
37	1	Rue Nationale	48.821872	2.369142	75013	2023	183216	0.014706	-1.011766	0.718999	-0.522533
38	1	Avenue de la Porte de Choisy	48.817737	2.366293	75013	2023	183216	0.022222	-1.011766	0.718999	-0.314295
39	1	Rue du Javelot	48.826193	2.365315	75013	2023	183216	0.010417	-1.011766	0.718999	-0.641365
61	1	Boulevard de Charonne	48.848506	2.399505	75020	1721	195556	0.025641	-1.564126	0.894627	-0.219578
62	1	Rue des Lyanes	48.862755	2.406363	75020	1721	195556	0.000000	-1.564126	0.894627	-0.929956
63	1	Cours de Vincennes	48.848452	2.399542	75020	1721	195556	0.051724	-1.564126	0.894627	0.503047
65	1	Rue d'Avron	48.851465	2.398697	75020	1721	195556	0.022222	-1.564126	0.894627	-0.314295
66	1	Rue de la Bidassoa	48.865210	2.395121	75020	1721	195556	0.033333	-1.564126	0.894627	-0.006465
75	1	Rue Caulaincourt	48.884905	2.329747	75018	1766	197580	0.016393	-1.481821	0.923433	-0.475780
77	1	Rue Francoeur	48.889956	2.342502	75018	1766	197580	0.013889	-1.481821	0.923433	-0.545168
78	1	Rue Custine	48.887311	2.349208	75018	1766	197580	0.024096	-1.481821	0.923433	-0.262372
79	1	Rue du Mont Cenis	48.886439	2.341282	75018	1766	197580	0.000000	-1.481821	0.923433	-0.929956
80	1	Rue Marcadet	48.890309	2.355449	75018	1766	197580	0.000000	-1.481821	0.923433	-0.929956
81	1	Rue Ramey	48.887015	2.347462	75018	1766	197580	0.020000	-1.481821	0.923433	-0.375861

Cluster 2:

	Cluster Label	Neighborhood	latitude	longitude	Postal Code	District Average Wage Per Household €	District population 2018	Japanese Restaurant	Normalized Household Income	Normalized Number of Residents	Normalized Number of Japanese Restaurants
25	2	Rue Charlot	48.860666	2.359843	75003	2803	35049	0.023810	0.414859	-1.389766	-0.270320
26	2	Rue Greneta	48.865408	2.353128	75003	2803	35049	0.030000	0.414859	-1.389766	-0.098814
27	2	Boulevard de Sébastopol	48.862088	2.350445	75003	2803	35049	0.030000	0.414859	-1.389766	-0.098814
28	2	Rue Rambuteau	48.860188	2.356768	75003	2803	35049	0.020000	0.414859	-1.389766	-0.375861
43	2	Rue Etienne Marcel	48.863569	2.350696	75002	2724	20796	0.030612	0.270367	-1.592620	-0.081852
45	2	Rue Montmartre	48.865060	2.344866	75002	2724	20796	0.023333	0.270367	-1.592620	-0.283512
47	2	Rue Montorgueil	48.864615	2.346638	75002	2724	20796	0.020000	0.270367	-1.592620	-0.375861
86	2	Rue de Mogador	48.873431	2.331065	75009	2974	59408	0.000000	0.727619	-1.043080	-0.929956
87	2	Boulevard Haussmann	48.872087	2.338386	75009	2974	59408	0.023810	0.727619	-1.043080	-0.270320
88	2	Rue de Sèze	48.870482	2.326544	75009	2974	59408	0.000000	0.727619	-1.043080	-0.929956
89	2	Rue Godot de Mauroy	48.870080	2.326521	75009	2974	59408	0.000000	0.727619	-1.043080	-0.929956
90	2	Rue de Rochechouart	48.876613	2.344288	75009	2974	59408	0.026667	0.727619	-1.043080	-0.191163
91	2	Rue d'Amsterdam	48.875798	2.326955	75009	2974	59408	0.047619	0.727619	-1.043080	0.389316
92	2	Rue Saint-lazare	48.876767	2.338547	75009	2974	59408	0.030000	0.727619	-1.043080	-0.098814
93	2	Rue du Pot de Fer	48.843066	2.349383	75005	2972	59333	0.000000	0.723961	-1.044148	-0.929956
94	2	Rue Gay-lussac	48.846887	2.340909	75005	2972	59333	0.000000	0.723961	-1.044148	-0.929956
95	2	Rue Gay-lussac	48.846887	2.340909	75005	2972	59333	0.000000	0.723961	-1.044148	-0.929956
96	2	Rue Saint-jacques	48.852066	2.346398	75005	2972	59333	0.020000	0.723961	-1.044148	-0.375861
97	2	Rue de la Montagne Sainte-geneviève	48.849544	2.349012	75005	2972	59333	0.020000	0.723961	-1.044148	-0.375861

Cluster 3 :

	Cluster Label	Neighborhood	latitude	longitude	Postal Code	District Average Wage Per Household €	District population 2018	Japanese Restaurant	Normalized Household Income	Normalized Number of Residents	Normalized Number of Japanese Restaurants
41	3	Rue Sainte-anne	48.867067	2.335991	75002	2724	20796	0.1525	0.270367	-1.592620	3.295014
42	3	Rue Saint-augustin	48.868824	2.338254	75002	2724	20796	0.1300	0.270367	-1.592620	2.671657
44	3	Rue Danielle Casanova	48.867657	2.333318	75002	2724	20796	0.1200	0.270367	-1.592620	2.394610
46	3	Rue de Richelieu	48.866739	2.337334	75002	2724	20796	0.1300	0.270367	-1.592620	2.671657
116	3	Rue de Montpensier	48.863883	2.336132	75001	2862	16545	0.1400	0.522770	-1.653122	2.948705
117	3	Rue Saint-roch	48.864220	2.331394	75001	2862	16545	0.1500	0.522770	-1.653122	3.225752

Cluster 4:

	Cluster Label	Neighborhood	latitude	longitude	Postal Code	District Average Wage Per Household €	District population 2018	Japanese Restaurant	Normalized Household Income	Normalized Number of Residents	Normalized Number of Japanese Restaurants
3	4	Rue Mathurin Régnier	48.840948	2.306902	75015	2695	234994	0.022222	0.217326	1.455922	-0.314295
4	4	Rue de Villafranca	48.831845	2.303805	75015	2695	234994	0.000000	0.217326	1.455922	-0.929956
8	4	Boulevard Garibaldi	48.847795	2.302753	75015	2695	234994	0.025641	0.217326	1.455922	-0.219578
9	4	Rue Balard	48.845455	2.277199	75015	2695	234994	0.022222	0.217326	1.455922	-0.314295
10	4	Rue Dutot	48.839262	2.309423	75015	2695	234994	0.020000	0.217326	1.455922	-0.375861
14	4	Cité de l'Ameublement	48.850575	2.385576	75011	2312	149834	0.000000	-0.483183	0.243895	-0.929956
15	4	Rue de la Roquette	48.853718	2.370599	75011	2312	149834	0.006667	-0.483183	0.243895	-0.745258
16	4	Rue Saint-maur	48.858686	2.383386	75011	2312	149834	0.010000	-0.483183	0.243895	-0.652909
17	4	Rue Guillaume Bertrand	48.863183	2.379417	75011	2312	149834	0.020000	-0.483183	0.243895	-0.375861
18	4	Rue Louis Bonnet	48.870375	2.376418	75011	2312	149834	0.020000	-0.483183	0.243895	-0.375861
19	4	Avenue Parmentier	48.858810	2.379499	75011	2312	149834	0.010989	-0.483183	0.243895	-0.625508
20	4	Rue Godefroy Cavaignac	48.853994	2.381870	75011	2312	149834	0.021053	-0.483183	0.243895	-0.346699
21	4	Rue du Faubourg Saint-antoine	48.853277	2.370530	75011	2312	149834	0.006536	-0.483183	0.243895	-0.748879
22	4	Rue Richard Lenoir	48.854193	2.382507	75011	2312	149834	0.022727	-0.483183	0.243895	-0.300303
23	4	Avenue de la République	48.867013	2.366115	75011	2312	149834	0.030000	-0.483183	0.243895	-0.098814

Cluster 5:

	Cluster Label	Neighborhood	latitude	longitude	Postal Code	District Average Wage Per Household €	District population 2018	Japanese Restaurant	Normalized Household Income	Normalized Number of Residents	Normalized Number of Japanese Restaurants
0	5	Rue de l'Arrivée	48.843690	2.322940	75015	2695	234994	0.092105	0.217326	1.455922	1.621794
1	5	Rue de Vaugirard	48.844827	2.319498	75015	2695	234994	0.074074	0.217326	1.455922	1.122245
2	5	Avenue du Maine	48.844752	2.320214	75015	2695	234994	0.060000	0.217326	1.455922	0.732327
5	5	Boulevard de Vaugirard	48.841637	2.319769	75015	2695	234994	0.098039	0.217326	1.455922	1.786193
6	5	Rue Rouelle	48.850308	2.287211	75015	2695	234994	0.060241	0.217326	1.455922	0.739003
7	5	Rue Lecourbe	48.844861	2.310567	75015	2695	234994	0.037037	0.217326	1.455922	0.096145
11	5	Boulevard du Montparnasse	48.846604	2.316556	75015	2695	234994	0.084337	0.217326	1.455922	1.406587
12	5	Rue de la Fédération	48.855026	2.289944	75015	2695	234994	0.063830	0.217326	1.455922	0.838430
13	5	Rue Cambronne	48.847396	2.301759	75015	2695	234994	0.045455	0.217326	1.455922	0.329350
30	5	Rue d'Odessa	48.843182	2.324402	75014	2394	139992	0.086957	-0.333205	0.103820	1.479150
31	5	Rue du Maine	48.840502	2.323970	75014	2394	139992	0.086957	-0.333205	0.103820	1.479150
40	5	Rue Albert	48.823910	2.373544	75013	2023	183216	0.074074	-1.011766	0.718999	1.122245
50	5	Rue Biot	48.883728	2.326604	75017	2670	168533	0.050633	0.171601	0.510026	0.472815
64	5	Rue de Belleville	48.872110	2.377414	75020	1721	195556	0.117647	-1.564126	0.894627	2.329423
76	5	Rue Leibniz	48.896966	2.337975	75018	1766	197580	0.142857	-1.481821	0.923433	3.027861
84	5	Boulevard de Picpus	48.840156	2.400156	75012	2343	142340	0.078261	-0.426484	0.137238	1.238239

Analysis of the clusters:

	Spending power	Population of the district(s)	Competition
Cluster 0	High	Low & High	Average
Cluster 1	Low	High	Low
Cluster 2	High	Low	Low
Cluster 3	High	Low	High
Cluster 4	Mid	High	Average
Cluster 5	Mid	High	High

For the concept of Izakaya, we are looking for a populated district, where residents have mid/high spending power and there is average competition. The cluster number 4 seems to be the most appropriate.

VII) MORE ANALYSIS ON CLUSTER 4

Postal Code	Cluster Label	Neighborhood	latitude	longitude	District Average Wage Per Household €	District population 2018	Japanese Restaurant	Normalized Household Income	Normalized Number of Residents	Normalized Number of Japanese Restaurants
75010	8	8	8	8	8	8	8	8	8	8
75011	11	11	11	11	11	11	11	11	11	11
75012	2	2	2	2	2	2	2	2	2	2
75014	3	3	3	3	3	3	3	3	3	3
75015	5	5	5	5	5	5	5	5	5	5
75017	12	12	12	12	12	12	12	12	12	12

Fig 12 – Number of neighborhoods in the cluster 4 in each district of Paris

Let's use scrapped data from Tripadvisor to see what offer the existing successful Japanese restaurants near the same location are offering. By hypothesis, a successful restaurant has a rating higher than 4, and a number of reviews higher than 30.

	Name	City	Cuisine Style	Rating	Price Range	Number of Reviews	Reviews	Borough	Neighborhood	Address	Insee Code	Latitude	Longitude
2806	Kazoku	Paris	Japanese	4.0	$	90.0	Great price - good quantity but adequate qu	Paris 14e	Rue de la Tombe Issoire	105 Rue de la Tombe Issoire 75014 Paris	75114.0	48.826053	2.331825
3737	Shook	Paris	Japanese	4.5	– $	71.0	A good plan - Always a pleasure coming here	Paris 11e	Rue de la Roquette	8 Rue de la Roquette 75011 Paris	75111.0	48.853485	2.370625
3848	Zenzan	Paris	Japanese	4.0	– $	70.0	Average - but high price - Very good	Paris 17e	Rue Brey	4 Rue Brey 75017 Paris	75117.0	48.876165	2.296262
4575	Kyobashi	Paris	Japanese	4.5	– $	65.0	Super friendly staff - and delicious food - Mo...	Paris 11e	Rue Saint-maur	117 Rue Saint-maur 75011 Paris	75111.0	48.866291	2.376661
4001	Toyama	Paris	Japanese	4.0	– $	63.0	Jap style restaurand - Very nice sushi restaurant	Paris 14e	Rue d'Odessa	14 Rue d'Odessa 75014 Paris	75114.0	48.842297	2.324454
4003	Toyama	Paris	Japanese	4.0	– $	63.0	Jap style restaurand - Very nice sushi restaurant	Paris 11e	Boulevard de Charonne	109 Boulevard de Charonne 75011 Paris	75111.0	48.856084	2.394406
2687	Yamada	Paris	Japanese	4.5	– $	60.0	Very good food - Fantastic - good -value sushi	Paris 10e	Rue du Faubourg Saint-martin	188 Rue du Faubourg Saint-martin 75010 Paris	75110.0	48.878671	2.362421
4033	Teppanyaki Sushi	Paris	Japanese	4.0	– $	60.0	Thought of best Japanese food	Paris 17e	Rue de Lévis	88 Rue de Lévis 75017 Paris	75117.0	48.884722	2.311877
5435	Naoki	Paris	Japanese	4.5	– $	59.0	Authentic - Excellent Japanese	Paris 11e	Rue Guillaume Bertrand	5 Rue Guillaume Bertrand 75011 Paris	75111.0	48.863356	2.379793
4533	Robata	Paris	Japanese	4.0	– $	57.0	Very nice food stop - Great service and good food	Paris 12e	Rue Claude Decaen	96 Rue Claude Decaen 75012 Paris	75112.0	48.838740	2.396308
2456	Yoki	Paris	Japanese	4.5	– $	47.0	Excellent - reasonably - priced sushi - Small r...	Paris 14e	Rue du Maine	18 Rue du Maine 75014 Paris	75114.0	48.840675	2.322375

Fig 13 – Successful restaurants in the same district than neighborhoods in the cluster 4

What emerges is that all restaurants offer a rather low-price range. The next step of analysis would be to check the offer of each successful restaurant.

	Age	75001	75002	75003	75004	75005	75006	75007	75008	75009	75010	75011	75012	75013	75014	75015	75016	75017	75018	75019	75020
0	0-14	11.8	12	11.4	11.4	11.4	10.2	12.1	15.1	14	14.9	12	13.4	13.7	12.6	13.6	14.6	14.3	14	17	15.1
1	15-29	22.1	26.2	25.9	24.5	29.2	27.6	23.6	23.7	25	23.5	24.7	22.7	23.7	26.3	24.2	20.8	23.9	23.5	21.9	20.9
2	30-44	22.7	27.3	25.2	21.6	17.3	16.4	17.3	20.2	25.4	26.7	25.4	22.4	20.3	19.9	21.4	17.1	23.1	26.5	21.7	23.1
3	45-49	20.3	19.8	19.2	19.3	16.9	16.9	18.2	18.7	17.8	18.5	17.8	18.5	19	17.2	17.6	19.7	17.9	18	19.2	20
4	60-74	14.9	10.6	12.5	15.3	15.5	17.3	17.3	14.8	11.8	11.4	13.6	14.3	15.2	15.1	14.3	16.3	13.5	12.2	13.7	14.4
5	75+	8.1	4.1	5.8	7.8	9.5	11.5	11.5	7.6	6	4.9	6.5	8.7	8.1	8.9	8.9	11.5	7.3	5.8	6.5	6.5

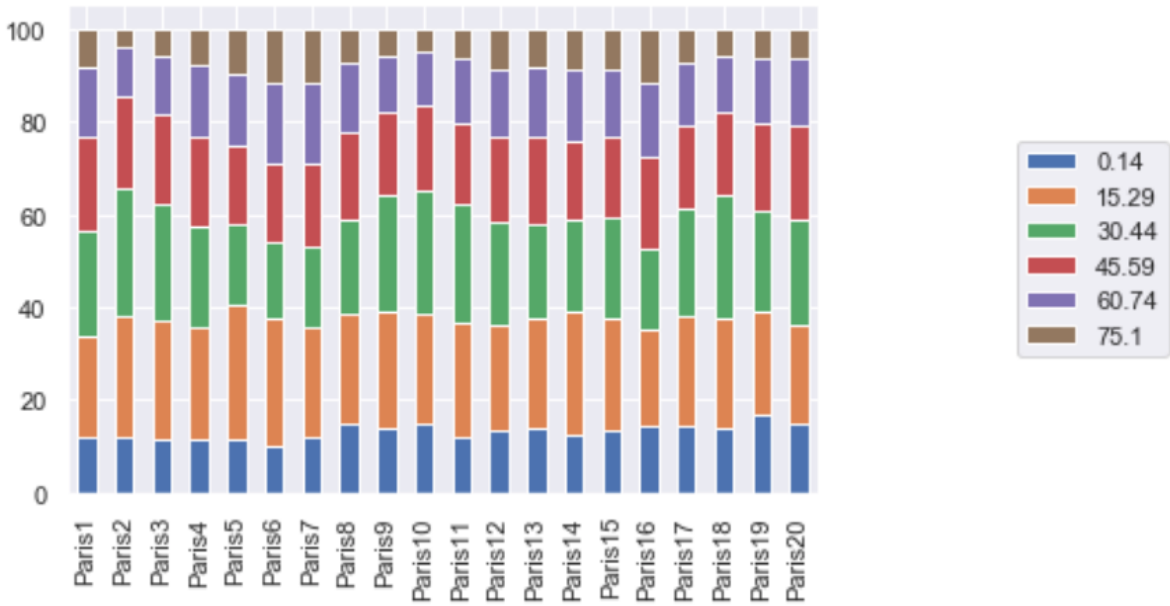

Fig 14 – Age distribution of the population in Paris

After analyzing the age of the population of the district, it seems that the age of the population is quite homogeneous in every district. most of the population in Paris belongs to the young working class, which is good for Izakaya concept. the age criterion therefore has little impact on the choice of location.

VIII) CONCLUSION

In this study, I tried to find the optimal location to open a restaurant in Paris based on 3 criteria: The purchasing power, the population of the district, and the plausible competition in the neighborhood in case of opening. An izakaya should be a place you can frequent for any budget. I particularly focused on a population with medium/high purchasing power who will be the target customers

In the absence of data, I used the number of residents in the borough as a criterion to study the possible flow of customers. Favoring a place with a large number of residents is therefore a necessity. The age of the population is quite homogeneous in every district, so it has no impact on the choice of the location.

Finally, an average competition was favored to take full advantage of the attractiveness of the Japanese restaurant group in the area to attract a maximum of customers.

Cluster 4 seems to correspond perfectly to our expectations.

IX) TO GO FURTHER

With more data, for example the population density per m2 for each neighborhood, we could be more accurate about the customer flow. The number of residents in a neighborhood is not relevant, but it can give us a potential number of clients.

The location of parking lots or public transport close to the chosen location, the more accessible the location is, the more customers there will be. If there are a lot of offices around the location that have their own self to assess the number of potential customers at lunchtime.

X) DATA SOURCES

Restaurants dataset from Kaggle : https://www.kaggle.com/damienbeneschi/krakow-ta-restaurans-data-raw

All demographic data are from Insee: https://insee.fr/

Codes and data are available on github: https://github.com/github-dhh/Applied-Capstone-Project-IBM