

Choose the best place for Japanese restaurant

Ilyes Touzene

I) Introduction

a. Background

Paris is the most visited capital in the world. About 50 million tourists visited Paris in 2018. The city is known for its monuments, as well as for its gastronomy.

A huge number of visitors and residents of Paris take advantage of the multitude of restaurants, shops and activities that the city offers. The city is also known for its ethnic, cultural diversity which allows a cultural richness.

Thanks to this wealth, Paris has seen the emergence of many different restaurants such as Asian restaurants, Italian restaurants, etc...

b. Problem

Paris offers many restaurants. Suppose I want to open a Japanese restaurant. I need to know where is the best place to open my restaurant. That is to say analyze the restaurant in each borough and look at the proportion of Japanese restaurant, also take into account the data on the population (age, young people, median income), the price for the rental of the room

c. Interest

The interest of this project is that I will be able to make the best choice to open my restaurant while minimizing competition, by providing a targeted service in relation to the population of the neighborhood, so I will be able to adapt my offers and be competitive.

II) Data

Based on definition of our problem, factors that will influence our decision are:

- i. Number of Japanese restaurants in Borough
- ii. Characteristic of people in Borough
- iii. Price for the rental

So the following will be needed to solve the probleme :

1. https://fr.wikipedia.org/wiki/Liste_des_quartiers_administratifs_de_Paris contains the informations of the name and the different neighborhood in each borough

Quartiers administratifs depuis 1860 [modifier | modifier le code]

Arrondissement ^{1, n° 1}	Quartiers	Population en 1999 (hab.) ²	Superficie (ha) ²	Densité hab/km ²	Plan
1 ^{er} arrondissement dit « du Louvre »	1 ^{er} Saint-Germain-l'Auxerrois	1 672	86,9	1 924	
	2 ^{er} Halles	8 984	41,2	21 806	
	3 ^{er} Palais-Royal	3 195	27,4	11 661	
	4 ^{er} Place-Vendôme	3 044	26,9	11 316	
2 ^{er} arrondissement dit « de la Bourse »	5 ^{er} Galign	1 345	18,8	7 154	
	6 ^{er} Vivienne	2 917	24,4	11 955	
	7 ^{er} Mail	5 783	27,8	20 802	
	8 ^{er} Bonne-Nouvelle	9 595	28,2	34 514	
3 ^{er} arrondissement dit « du Temple »	9 ^{er} Arts-et-Métiers	9 560	31,8	30 063	
	10 ^{er} Enfants-Rouges	8 562	27,2	31 478	
	11 ^{er} Archives	8 609	36,8	23 394	
	12 ^{er} Sainte-Avoye	7 501	21,3	35 216	
4 ^{er} arrondissement dit « de l'Hôtel-de-Ville »	13 ^{er} Saint-Merri	6 523	31,3	20 840	
	14 ^{er} Saint-Gervais	10 587	42,2	25 088	
	15 ^{er} Arsenal	9 474	48,7	19 454	
	16 ^{er} Notre-Dame	4 087	37,9	10 784	

2. <https://www.apur.org/dataviz/portraits-metropole-grand-paris-donnees/?fbclid=IwAR13J2vJTG6ZDpsGJgicSaOkJN1EILnH-GGtyiEDD6yUtypq9cqIC50l7k>

Characteristic of people in each borough. We choose only certain informations and put in a csv

Postal Code	Municipal population	Young people	Old people	Student	Median income	Employment density	Sale price
75001	16545,00	15,30	18,70	8,50	31843,00	328,00	11290,00
75002	20796,00	15,70	10,10	8,90	30025,00	603,00	11270,00
75003	35049,00	15,60	13,20	9,60	30988,00	267,00	11240,00
75004	27146,00	15,70	17,10	10,10	30515,00	260,00	12300,00
75005	59333,00	17,80	18,80	16,50	32950,00	211,00	11330,00
75006	42428,00	19,00	22,70	15,50	38448,00	206,00	12530,00
75007	54133,00	18,60	20,90	10,80	41949,00	174,00	12400,00
75008	36694,00	20,70	16,50	9,90	39774,00	440,00	9890,00
75009	59408,00	18,80	13,00	8,20	32771,00	530,00	9910,00
75010	91770,00	19,70	11,10	7,40	25154,00	291,00	8850,00
75011	149834,00	16,10	14,20	8,50	26253,00	223,00	9190,00
75012	142340,00	18,90	11,20	8,00	26729,00	74,00	8620,00
75013	183216,00	19,60	17,40	9,90	23538,00	177,00	8380,00
75014	139992,00	17,70	18,20	12,10	27233,00	147,00	9430,00
75015	234994,00	18,60	17,50	9,10	30227,00	201,00	9140,00
75016	165487,00	20,60	22,60	9,40	38299,00	70,00	9820,00
75017	168533,00	19,70	15,30	8,10	29872,00	188,00	9400,00
75018	197580,00	18,70	12,90	7,10	20942,00	147,00	8560,00
75019	185654,00	23,50	13,80	6,60	19137,00	132,00	7630,00
75020	195556,00	20,90	14,50	6,40	20632,00	128,00	7890,00

I choose to keep : Number of people, the proportion of young people, old people and student. The median income (euros per year) to analyze if the borough is rich. The employment density (employment per hectare) to see if there are lot of worker in the borough and finaly, the sale price for the rental (Euros/m^2).

3. Foursquare API to find the restaurant in each Borough
4. Geopy to find the location of each borough.

III) Methodology:

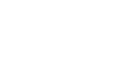
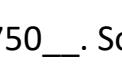
The goal of this part is to prepare the data and the cluster to the Analysis part. We have to find the best data, and the best transformation to clustering our data.

So :

- a) We will first extract the data from wikipedia to recover the Boroughs and their neighborhoods.
- b) We will add the postal code to later search for locations with Geopy.
- c) Next, we will extract the population data and add it to the dataframe to identify the characteristics of our Boroughs.
- d) We use Foursquare to find the most Common Venues and We filter by restaurant to keep only the most common restaurant in the Borough to compare it.
- e) Finally with all these data, we will perform a K-Means in order to cluster our boroughs and be able to analyze them according to their characteristics in the Analysis part

a) Data from Wikipedia :

Quartiers administratifs depuis 1860 [modifier | modifier le code]

Arrondissement ^{1,n 1}	Quartiers	Population en 1999 (hab.) ²	Superficie (ha) ²	Densité hab/km ²	Plan
1 ^e arrondissement dit « du Louvre »	1 ^e Saint-Germain-l'Auxerrois	1 672	86,9	1 924	
	2 ^e Halles	8 984	41,2	21 806	
	3 ^e Palais-Royal	3 195	27,4	11 661	
	4 ^e Place-Vendôme	3 044	26,9	11 316	
2 ^e arrondissement dit « de la Bourse »	5 ^e Gaillon	1 345	18,8	7 154	
	6 ^e Vivienne	2 917	24,4	11 955	
	7 ^e Mail	5 783	27,8	20 802	
	8 ^e Bonne-Nouvelle	9 595	28,2	34 514	
3 ^e arrondissement dit « du Temple »	9 ^e Arts-et-Métiers	9 560	31,8	30 063	
	10 ^e Enfants-Rouges	8 562	27,2	31 478	
	11 ^e Archives	8 609	36,8	23 394	
	12 ^e Sainte-Avoye	7 501	21,3	35 216	
4 ^e arrondissement dit « de l'Hôtel-de-Ville »	13 ^e Saint-Merri	6 523	31,3	20 840	
	14 ^e Saint-Gervais	10 587	42,2	25 088	
	15 ^e Arsenal	9 474	48,7	19 454	
	16 ^e Notre-Dame	4 087	37,9	10 784	

We extract from Wikipedia with the package Beautiful Soup to obtain this data. This data is composed of Borough with their Neighborhood

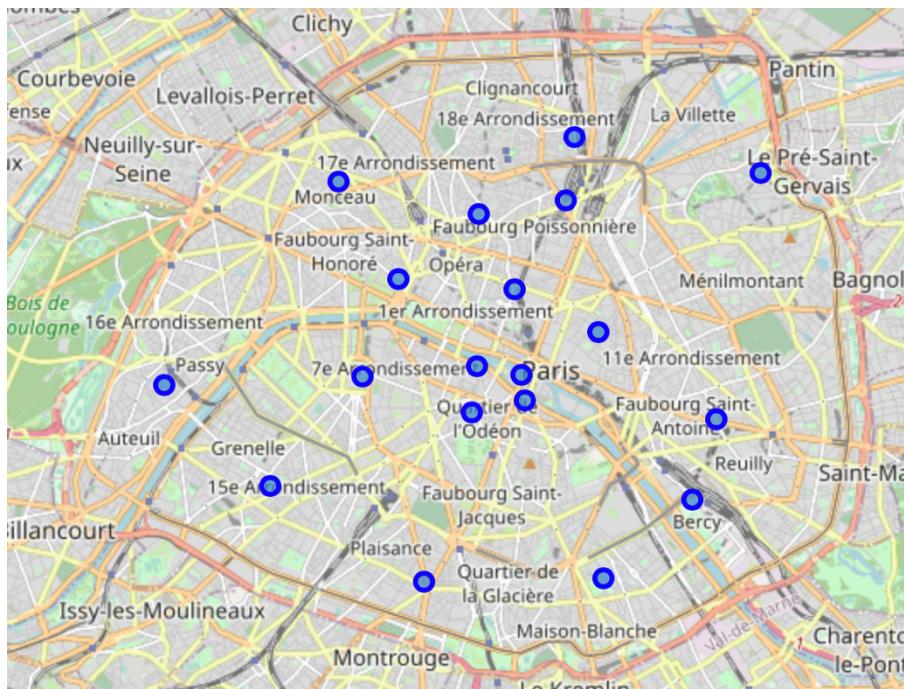
Borough	Neighborhood_1	Neighborhood_2	Neighborhood_3	Neighborhood_4
0 1er arrondissement de Paris	Quartier Saint-Germain-l'Auxerrois	Quartier des Halles	Quartier du Palais-Royal	Quartier de la Place-Vendôme
1 2e arrondissement de Paris	Quartier Gaillon	Quartier Vivienne	Quartier du Mail	Quartier de Bonne-Nouvelle
2 3e arrondissement de Paris	Quartier des Arts-et-Métiers	Quartier des Enfants-Rouges	Quartier des Archives	Quartier Sainte-Avoye
3 4e arrondissement de Paris	Quartier Saint-Merri	Quartier Saint-Gervais	Quartier de l'Arsenal	Quartier Notre-Dame (Paris)
4 5e arrondissement de Paris	Quartier Saint-Victor	Quartier du Jardin-des-Plantes	Quartier du Val-de-Grâce	Quartier de la Sorbonne

b) Postal Code :

We show that in Paris, the postal code for each borough begin by 750__. So for the first Borough, the postal code is 75001, and so on.

Borough	Neighborhood	Postal Codes
0 1er arrondissement de Paris	Quartier Saint-Germain-l'Auxerrois, Quartier d...	75001
1 2e arrondissement de Paris	Quartier Gaillon, Quartier Vivienne, Quartier ...	75002
2 3e arrondissement de Paris	Quartier des Arts-et-Métiers, Quartier des Enf...	75003
3 4e arrondissement de Paris	Quartier Saint-Merri, Quartier Saint-Gervais, ...	75004
4 5e arrondissement de Paris	Quartier Saint-Victor, Quartier du Jardin-des-...	75005

With Geopy, we find the longitude and the latitude for each Borough. The Postal Code is usefull to find this location.



Good, we have a map with Folium with the different Borough and their locations.

c) Extract the characteristic of Borough's people

Chiffres clés
T1-Paris

Jai

2e arrondissement

	Année	Indicateur	Note de lecture
DEMOGRAPHIE / SOCIETE			
Population municipale	2015	20 796	En 2015, l'arrondissement compte 20 796 habitants
Densité de population	2015	210	En 2015, l'arrondissement compte 210 habitants à l'hectare (bois compris)
Naissances	2017	276	En 2017, il y a eu 276 naissances domiciliées dans l'arrondissement
Décès	2017	88	En 2017, il y a eu 88 décès domiciliés dans l'arrondissement
Jeunes de moins de 20 ans	2015	15,7	En 2015, 15,7% des habitants ont moins de 20 ans
Personnes âgées de 65 ans ou plus	2015	10,1	En 2015, 10,1% des habitants ont 65 ans ou plus
Population immigrée	2015	21,2	En 2015, 21,2% des habitants sont immigrés
Population étrangère	2015	16,9	En 2015, 16,9% des habitants sont de nationalité étrangère
Etudiants résidents	2015	8,9	En 2015, 8,9% des habitants sont étudiants
Employés et ouvriers	2015	21,2	En 2015, les employés et ouvriers représentent 21,2% de la population active
Cadres et professions intellectuelles supérieures	2015	51,4	En 2015, les cadres et professions intellectuelles supérieures représentent 51,4% de la population active
Population non diplômée	2015	15,0	En 2015, 15% des personnes de 15 ans ou plus non scolarisées n'ont pas de diplôme
Population diplômée du supérieur	2015	67,3	En 2015, 67,3% des personnes de 15 ans ou plus non scolarisées sont diplômées du supérieur
Salaire net horaire moyen des femmes	2015	20,4	En 2015, le salaire net horaire moyen des femmes s'élève à 20,4€
Salaire net horaire moyen des hommes	2015	27,2	En 2015, le salaire net horaire moyen des hommes s'élève à 27,2€
Revenu médian	2015	30 025	En 2015, le revenu disponible médian par unité de consommation est de 30 025€ par an
Bénéficiaires de l'Allocation Adulte Handicapé (AAH)	2017	1,3	En 2017, les bénéficiaires de l'Allocation Adulte Handicapé représentent 1,3% des 20-64 ans
Bénéficiaires de l'Allocation d'Education de l'Enfant Handicapé (AEEH)	2017	2,0	En 2017, les bénéficiaires de l'Allocation d'Education de l'Enfant Handicapé représentent 2% des moins de 20 ans
MENAGE / FAMILLE			
Nombre de ménages	2015	12 010	En 2015, l'arrondissement compte 12 010 ménages
Taille moyenne des ménages (nb de personnes/ménage)	2015	1,72	En 2015, chaque ménage compte en moyenne 1,72 personne
Ménages ayant une voiture ou plus	2015	20,3	En 2015, 20,3% des ménages ont une voiture ou plus
Ménages d'une personne	2015	56,4	En 2015, 56,4% des ménages sont composés d'une seule personne
Familles avec enfant(s)	2015	17,0	En 2015, les familles avec enfant(s) de moins de 25 ans représentent 17% des ménages
Familles nombreuses (3 enfants ou plus)	2015	13,1	En 2015, 13,1% des familles avec enfant(s) de moins de 25 ans ont 3 enfants ou plus
Familles monoparentales	2015	24,0	En 2015, 24% des familles avec enfant(s) de moins de 25 ans sont des familles monoparentales
LOGEMENT / HABITAT			
Nombre de logements	2015	17 200	En 2015, l'arrondissement compte 17 200 logements

There is a lot of data in this database, so we have to choose the relevant data to better characterize our borough. For that we ask the question: what would be useful to know for a person who wants to open a restaurant. We would all like to know the proportion of different people in our Borough. There are elderly, young, students, It is important to know this to target our customers. Is the Borough populated? That is to say there are a lot of potential customers? We must also take into account the economic context of. Borough: people's income can influence our pricing strategy. Finally, the price of the rent is also a criterion for all businesses.

With all these ideas, here is the dataset that we extract from the database

	Borough	Neighborhood	Postal Codes	Latitude	Longitude	Postal Code	Municipal population	Young people	Old people	Student	Median income	Employment density	Sale price
0	1er arrondissement de Paris	Quartier Saint-Germain-l'Auxerrois, Quartier d...	75001	48.857366	2.336815	75001	16545.0	15.3	18.7	8.5	31843.0	328.0	11290.0
1	2e arrondissement de Paris	Quartier Gaillon, Quartier Vivienne, Quartier ...	75002	48.867460	2.344156	75002	20796.0	15.7	10.1	8.9	30025.0	603.0	11270.0
2	3e arrondissement de Paris	Quartier des Arts-et-Métiers, Quartier des Enf...	75003	48.861751	2.361134	75003	35049.0	15.6	13.2	9.6	30988.0	267.0	11240.0
3	4e arrondissement de Paris	Quartier Saint-Merri, Quartier Saint-Gervais, ...	75004	48.856079	2.345537	75004	27146.0	15.7	17.1	10.1	30515.0	260.0	12300.0
4	5e arrondissement de Paris	Quartier Saint-Victor, Quartier du Jardin-des-...	75005	48.852752	2.346343	75005	59333.0	17.8	18.8	16.5	32950.0	211.0	11330.0

d) Foursquare : Find the restaurant in each Borough

With Foursquare, we find the most common Venues in each Borough. We Filter by restaurant to keep only the restaurant and we take only the 10 most common restaurant in each Borough

42] :

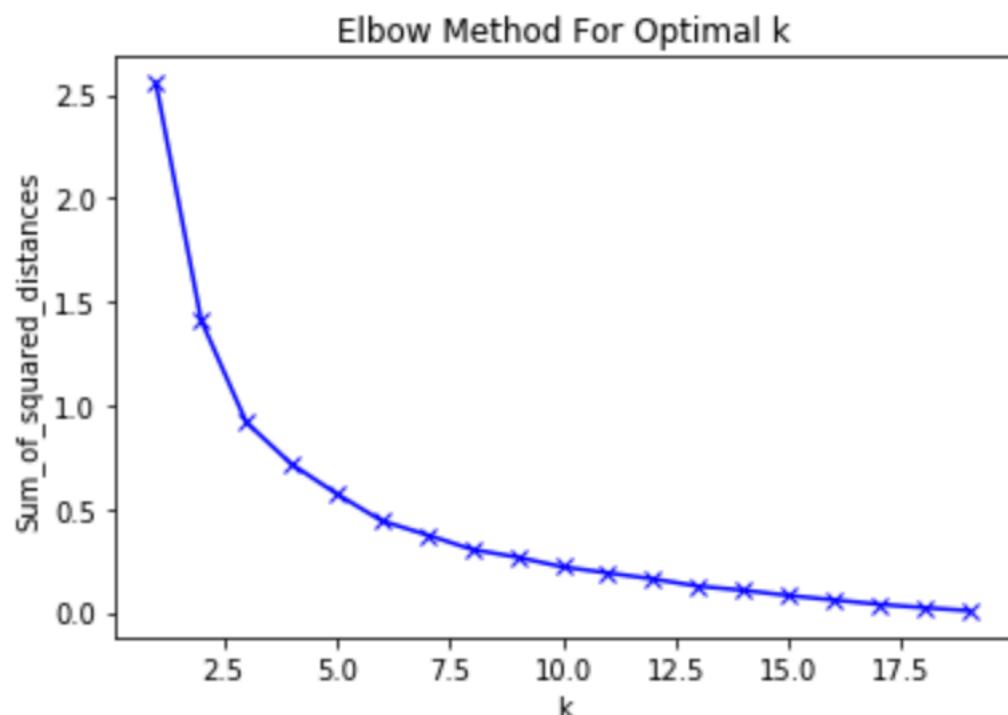
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Quartier Gaillon, Quartier Vivienne, Quartier ...	French Restaurant	Italian Restaurant	Thai Restaurant	Japanese Restaurant	Restaurant	Indonesian Restaurant	Israeli Restaurant	Lebanese Restaurant	Mexican Restaurant	Ch'ti Restaurant
1	Quartier Saint-Georges (Paris), Quartier de la...	French Restaurant	Japanese Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Modern European Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Falafel Restaurant	Venezuelan Restaurant
2	Quartier Saint-Germain- l'Auxerrois, Quartier d...	French Restaurant	Japanese Restaurant	Restaurant	Seafood Restaurant	Ramen Restaurant	Breton Restaurant	Modern European Restaurant	Asian Restaurant	Tapas Restaurant	Argentinian Restaurant
3	Quartier Saint-Lambert, Quartier Necker, Quart...	French Restaurant	Italian Restaurant	Lebanese Restaurant	Indian Restaurant	Restaurant	Scandinavian Restaurant	Moroccan Restaurant	Persian Restaurant	Russian Restaurant	Japanese Restaurant
4	Quartier Saint-Merri, Quartier Saint-Gervais, ...	French Restaurant	Japanese Restaurant	Falafel Restaurant	Restaurant	Auvergne Restaurant	Lebanese Restaurant	Moroccan Restaurant	Portuguese Restaurant	Tapas Restaurant	Alsation Restaurant
5	Quartier Saint-Thomas- d'Aquin, Quartier des In...	French Restaurant	Italian Restaurant	Basque Restaurant	Korean Restaurant	Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Auvergne Restaurant	Indonesian Restaurant
6	Quartier Saint-Victor, Quartier du Jardin-des-...	French Restaurant	Japanese Restaurant	Lebanese Restaurant	Restaurant	Tapas Restaurant	Mexican Restaurant	Falafel Restaurant	Chinese Restaurant	Indian Restaurant	Hawaiian Restaurant

e) Clustering :

With K-means, we clustering the Borough to find similarity between Borough.

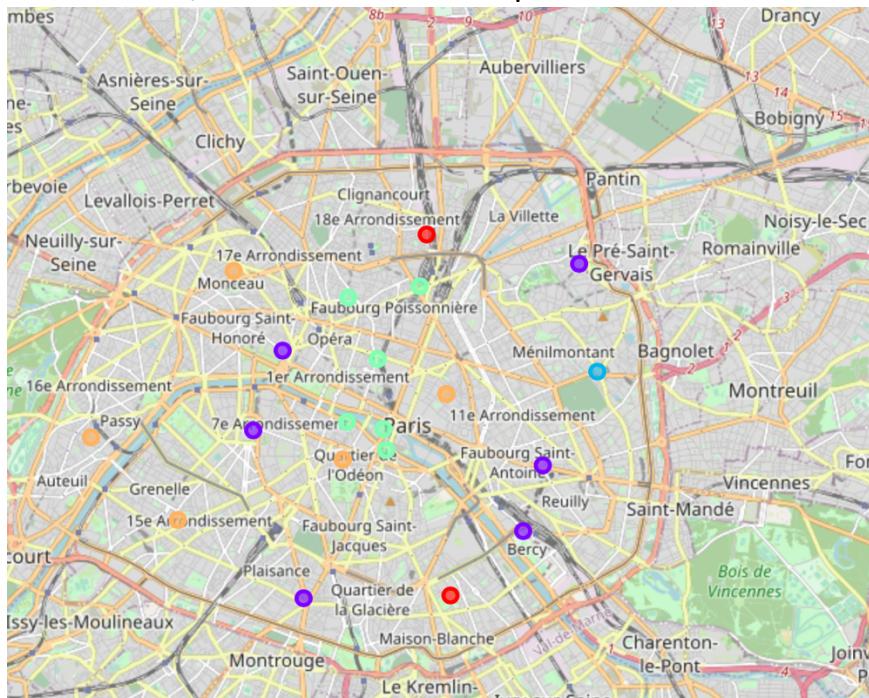
This will allow us to understand what type of different cluster exists in Paris and guide our choice of location for the restaurant.

In first, we try to find the best K-cluster for our dataset :



The elbow is for K = 5

With Folium, we actualize our map with the new cluster



We find this cluster with different color.

For example, the cluster 1 is like this :

Neighborhood	Municipal population	Young people	Old people	Student	Median income	Employment density	Sale price	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Co...
12 Quartier de la Salpêtrière, Quartier de la Gar...	183216.0	19.6	17.4	9.9	23538.0	177.0	8380.0	0	Vietnamese Restaurant	Asian Restaurant	Thai Restaurant	Chinese Restaurant	French Restaurant	Rest
17 Quartier des Grandes-Carrières, Quartier de Cl...	197580.0	18.7	12.9	7.1	20942.0	147.0	8560.0	0	Indian Restaurant	French Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Rest

Now, we have our cluster (1 to 5). We can analyze them in the next part

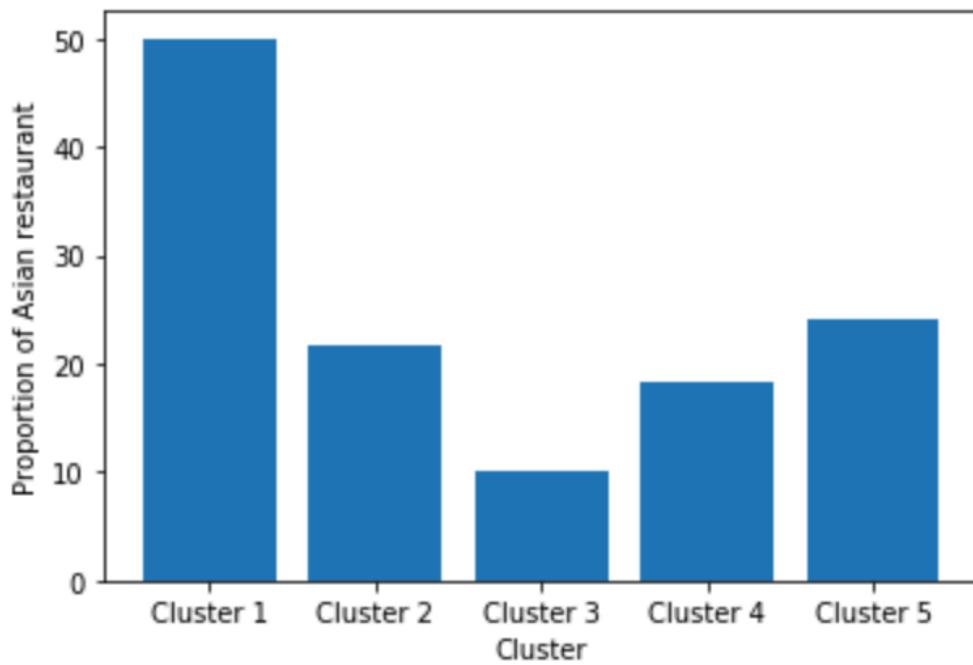
IV) Analysis

We can check the mean of each Cluster

	Municipal population	Young people	Old people	Student	Median income	Employment density	Sale price	Cluster Labels	Cluster name
0	190398.000000	19.150000	15.15	8.500000	22240.000000	162.000000	8470.000000	0.0	Cluster 1
1	118107.833333	19.250000	15.80	9.316667	30179.166667	198.333333	9526.666667	1.0	Cluster 2
2	195556.000000	20.900000	14.50	6.400000	20632.000000	128.000000	7890.000000	2.0	Cluster 3
3	45833.000000	17.166667	14.80	9.933333	30543.000000	370.500000	10825.000000	3.0	Cluster 4
4	129298.200000	18.700000	18.26	10.340000	33566.800000	186.400000	10426.000000	4.0	Cluster 5

We saw dispersion on each Cluster.

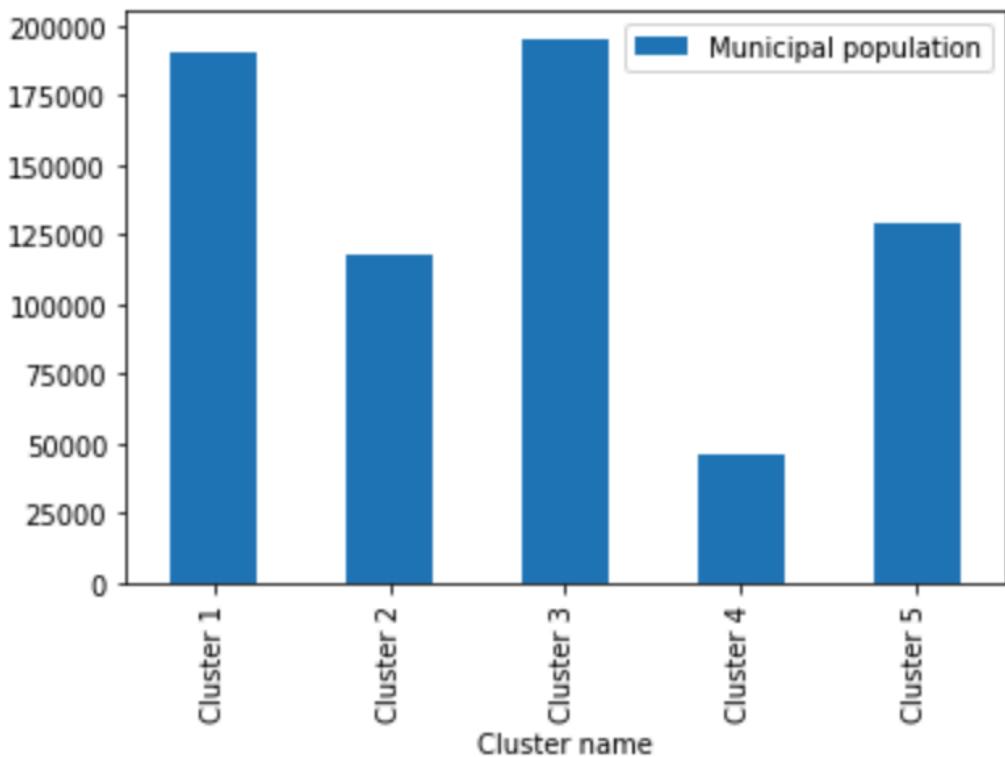
Asian restaurants are competitors for our Japanese restaurant because in the mind of Parisians, the food is similar. So we will check the proportion of Asian restaurants in each Cluster to determine the one in which there is the least amount of Asian restaurants



In the cluster 1, we have a large proportion of Asian restaurants. The Cluster 3 and 4 have a small proportion of Asian restaurant. So it's important for us to keep in mind because the competition in Paris is very hard.

Observation of features :

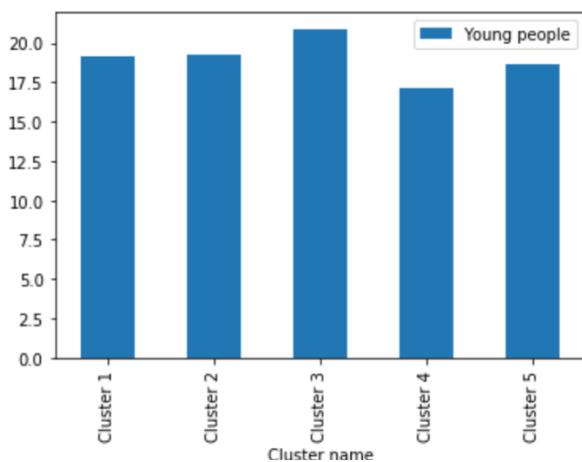
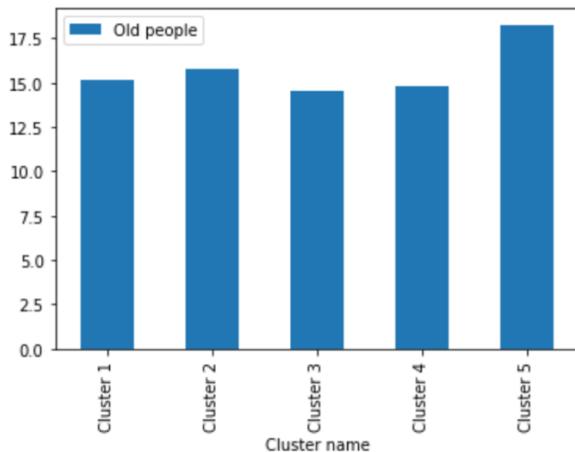
- Municipal Population :



The population in cluster 1 and 3 are very important, but in the cluster 4, the population is very low.

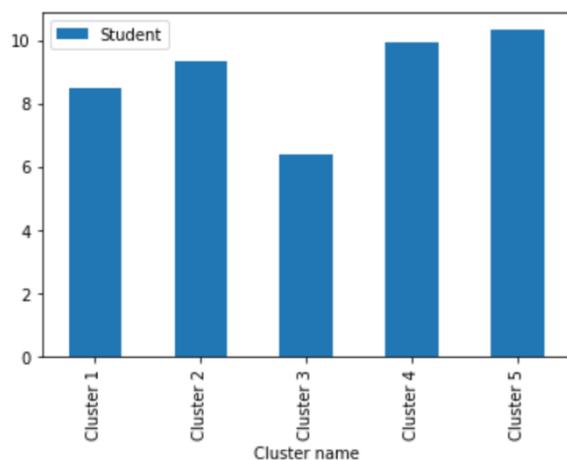
It's important for a restaurant to have a large of people around his place.

- Young and Old people



The proportion of old and young people is very similar. So it's not important for our study

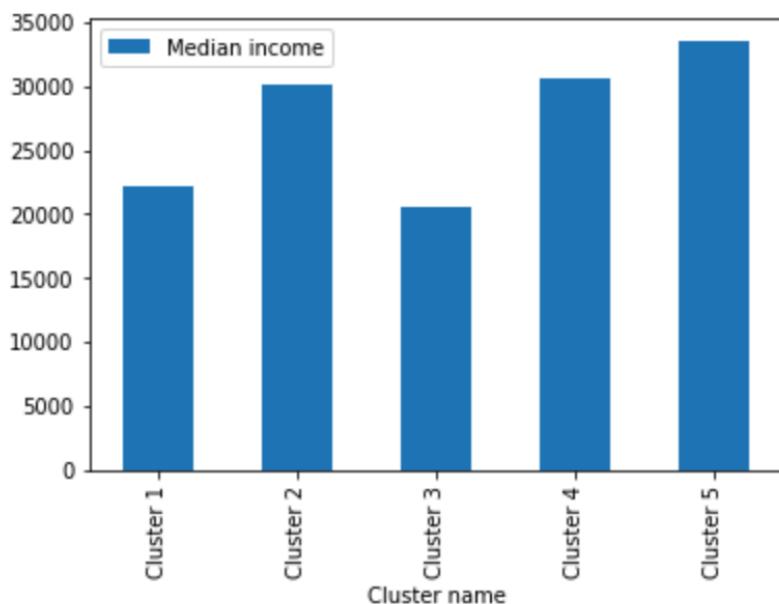
- Student



The proportion of student is low in the cluster 3 and it's constant in the other (e.g 10 % of the population).

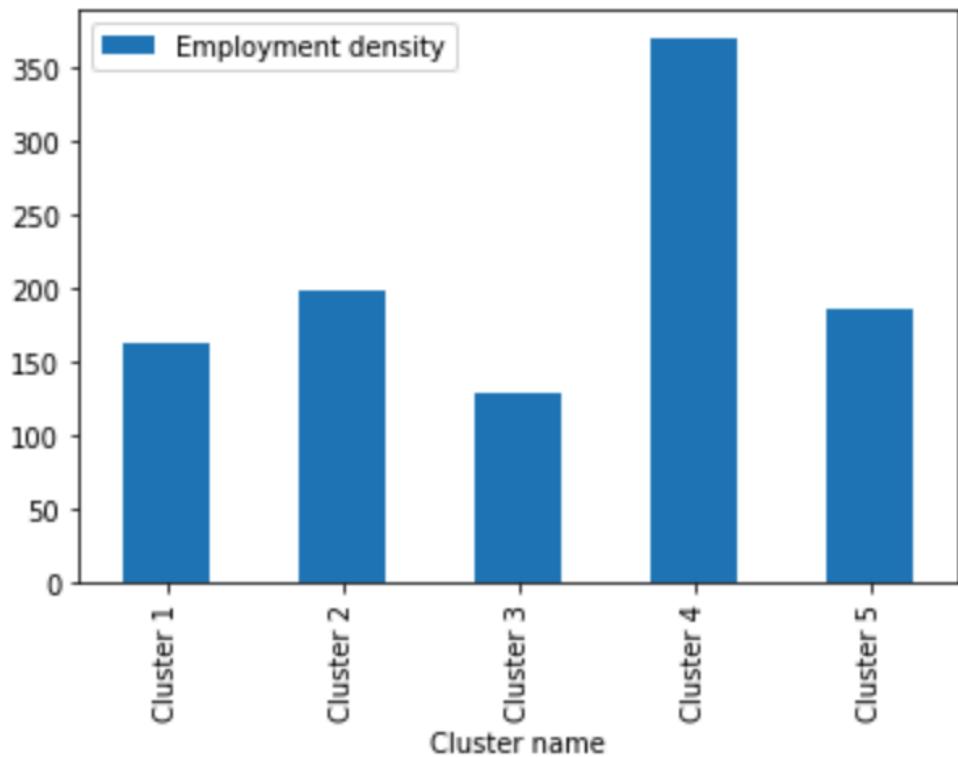
It's important to keep in mind because we can offer student discount to attract this population

- Median income

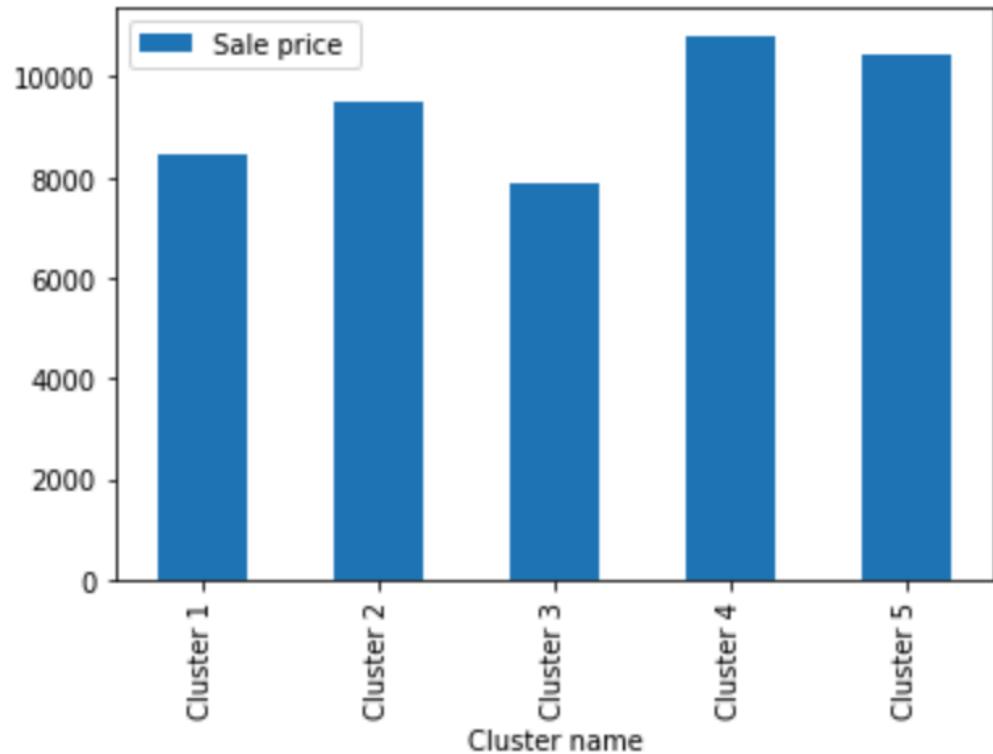


The median income is low for cluster 1 and 3 and very high for the other cluster. We keep in mind because the price of food is directly correlated with the median income of the population.

- Employment density



The cluster 4 has a high employment density. Is due to a lot of Firm in the cluster. But in cluster 3, the density is very low. It's important for our study because a lot of workers lunch around their office.



The Sale price of rental is very high in cluster 4 and 5. Is due to the attractivity of the different neighborhood, because they have a lot of monuments. In cluster 3 and 1, the rental is low.

V) Result and Discussion

By analyzing the different clusters more precisely, we realize that there are disparities. First of all, some districts by their history, have monuments, restaurants, activities, which attract different populations and therefore create a disparity in Paris. We also observed a higher proportion of Asian restaurants in certain cluster (eg cluster 1). Cluster 3 has the lowest proportion of Asian restaurants.

Some clusters are more populated than others, due to the fact that some districts are more residential, and others are business districts. The

proportion of old people and young people is equivalent in all the clusters so this should not guide our decision.

The number of students remains fairly high in each cluster, which can push us to offer special offers for students and thus make our business more attractive. The median income in the cluster is quite high because we are in Paris, but we can observe that in cluster 3 and 1, the income is lower. This can influence the price of food, knowing that the higher the income, the more we will be able to offer better quality products and therefore more expensive.

The density of employment can also play a role in our restaurant positioning. People working around the restaurant will tend to eat in the same neighborhood. This can be an important economic argument. We observe that cluster 4 has a high density of employment. Conversely, cluster 3 has a lower density.

Finally, the rental price influences our decision. A high price will generate high costs. However, we observe that cluster 3 has the cheapest rent compared to the others. It's important to take it into the equation

Concerning the clustering model, we used a simple k-means model, we used the elbow method to minimize inertia. This method is simple to implement but it is not effective in all cases. We could start over with another clustering algorithm and see if we get the same results.

VI) Conclusion

Our goal was to choose the best location for our Japanese restaurant. We recovered the paris data with the different districts, then we added data on the characteristics of the populations in these Boroughs in order to be able to cluster them more easily.

Finally we created these clusters and we analyzed them separately. What emerges from this analysis is that we must position ourselves on the type of restaurants that we want to open.

For example, if we want to reach people with low budgets, families with their children, and with low competition and cheap rent, we will choose the neighborhoods of cluster 3.

On the other hand, if we want to create a gourmet Japanese restaurant , and touch the executives who work in the offices and who earn a very good living, we will rather choose the neighborhoods of cluster 4. Finally, several choices are possible depending on the type of restaurant we want to open