Capstone Project
The Battle of Neighborhoods in Paris

Method for determining the best location for a new restaurant

Christophe Coudé
December 14, 2020

Background and Business problem

About 12,000 restaurants in Paris, few place to open new ones

Two strategies (among others):

- finding a place where there are very few restaurants of his specialty in order to attract customers of the local neighborhood who want to change cuisine
- setting up his restaurant where there are already many restaurants of his own specialty because even if there is competition, this means that there is an already existing clientele.

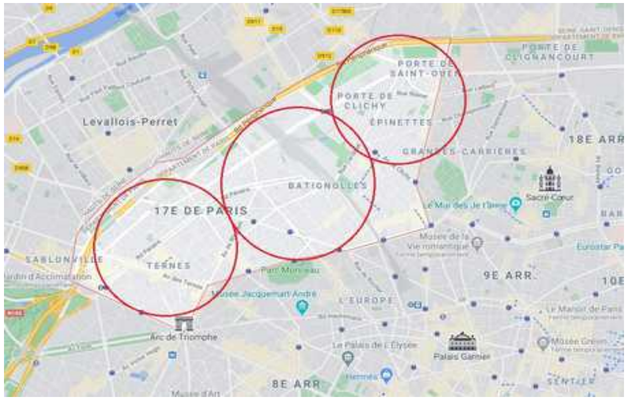
Aim of the project:

offering to new restaurant owners a global vision of the distribution of restaurants in Paris to help him/her find the best location, by

- finding where the restaurants are located and to know their specialty with the use of *Foursquare API*
- using machine learning to bring out the general culinary trends and tastes of each neighborhood, for example the unsupervised learning method *Clustering*
- visualizing the position of the various restaurants in the choosen district, for example with *Folium*

Data acquisition and cleaning

- Paris venues descriptions requested with *Foursquare API*
- List of Foursquare venues categories collected at *Foursquare*
- Paris GPS coordonates scrapped from *latitude.to*
- Paris districts GPS coordonates picked up on *Google Map*
- Paris maps drawn with *Folium*



	Arrd	Arrd_lat	Arrd_lng	Venue	Venue lat	Venue lng	Category region	Cuisine
0	1	48.865789	2.327739	Charly Bun's	48.867215	2.332305	America	American
1	1	48.865789	2.327739	Razowski	48.867273	2.332588	America	American
2	1	48.865789	2.327739	Will'n Joy	48.871317	2.325540	America	American
3	1	48.865789	2.327739	Ferona	48.868458	2.324971	America	Argentinian
...	...	...	...	...	...	...	...	...
12268	20	48.855189	2.406624	Gook	48.847705	2.407768	Miscellaneous	FastFood
12269	20	48.855189	2.406624	Vitamin	48.851584	2.399250	Miscellaneous	FastFood
12270	20	48.855189	2.406624	Le Délice D'Avron	48.851358	2.398775	Miscellaneous	FastFood
12271	20	48.855189	2.406624	le marechal	48.857639	2.410185	Miscellaneous	FastFood
12272	20	48.855189	2.406624	Le Pasha	48.847516	2.410533	Miscellaneous	FastFood

12273 rows × 8 columns

12 273 venues, 69 categories, 8 regions

Arrd	Arrd_lat	Arrd_lng	Width
1	48.862994	2.335592	2.12
2	48.868651	2.342876	1.93
3	48.863132	2.359781	1.82
4	48.854801	2.358448	2.05
5	48.844402	2.350722	2.16

	Arrd	Arrd_lat	Arrd_lng	rad
0	1	48.865789	2.327739	500
1	1	48.863079	2.336408	500
2	1	48.861244	2.345377	500
3	2	48.869364	2.333671	150
4	2	48.868489	2.341267	300

46 circles

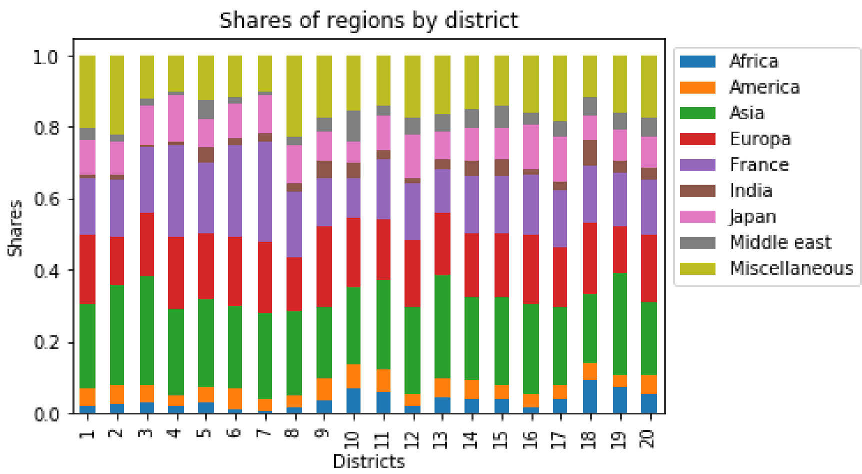
Methodology

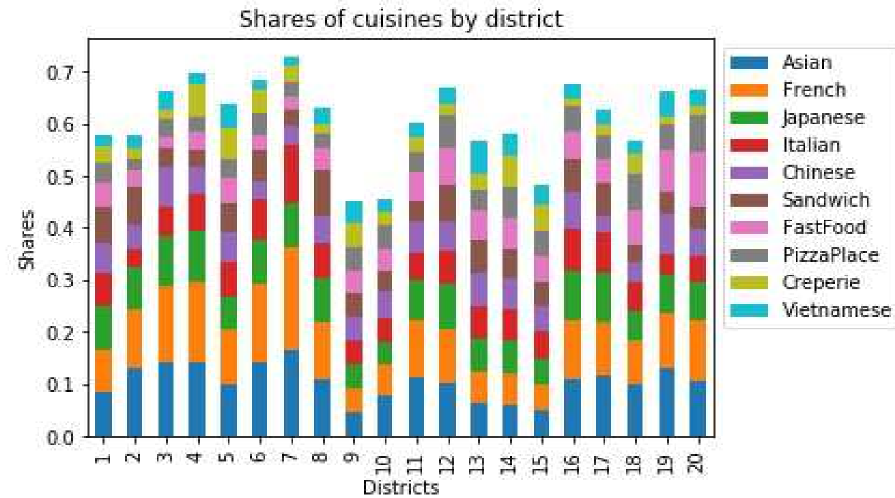
- First step : retrieve the data, GPS coordinates of the districts, list of the different restaurants with their category and their GPS coordinates.

- Second step : analyze the distribution of restaurant categories in the different districts in a vertical way, district by district, then transversally across the districts thanks to a clustering in order to highlight the major trends in distribution by grouping them together.

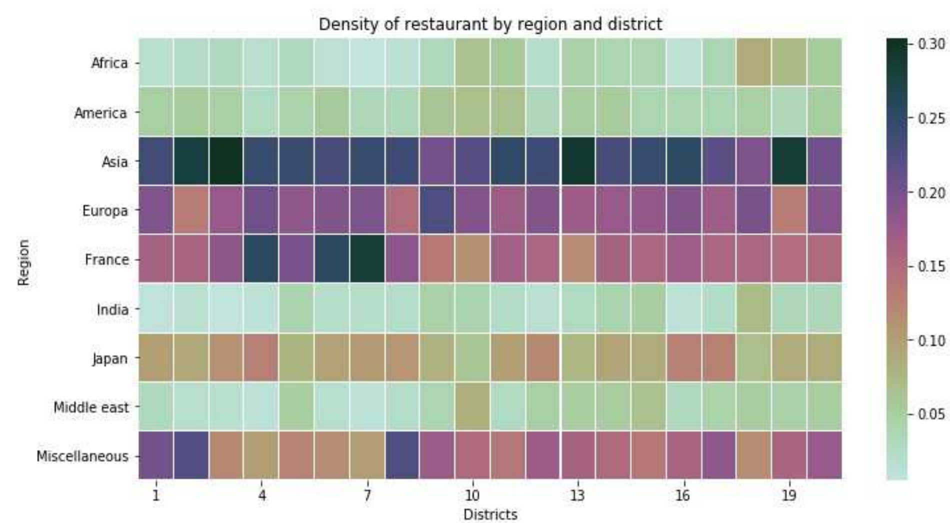
- Third step : study the distribution of restaurants in choosen district by visualizing where the groups of restaurants are located, the areas where there are few or no restaurants, and where the already present Indian restaurants are. The restaurateur will thus have a precise idea of the location where he could set up his restaurant.

Exploratory Data Analysis





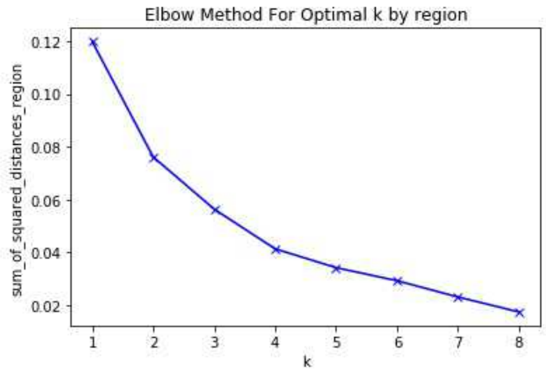
Arrd	Average_distance	Area	Density	Arrd	Average_distance	Area	Density
1	0.687064	183	3.754449	11	0.840872	367	2.291205
2	0.413809	99	4.179888	12	0.985709	637	1.547424
3	0.392223	117	3.352337	13	0.851786	715	1.191309
4	0.517666	160	3.235415	14	0.953627	564	1.690829
5	0.635183	254	2.500719	15	1.050954	848	1.239332
6	0.616876	215	2.869189	16	1.390533	791	1.757943
7	0.732535	409	1.791038	17	1.044189	567	1.841603
8	0.767314	388	1.977612	18	0.769694	601	1.280689
9	0.515735	218	2.365756	19	0.994141	679	1.464125
10	0.702536	289	2.430919	20	1.142961	598	1.911305



Clustering, data

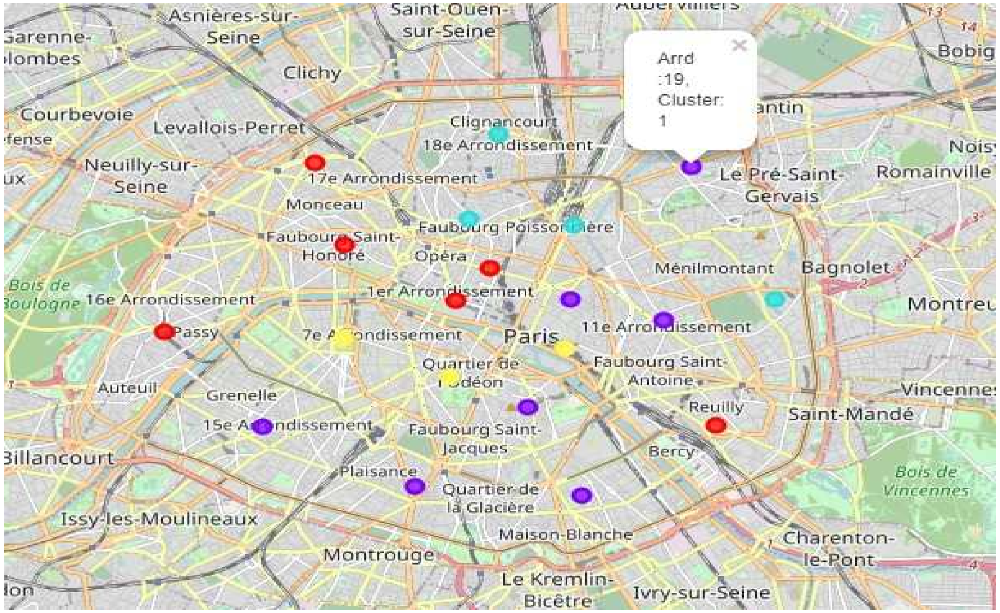
Arrd	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
1	French	Asian	Japanese	Sandwich	Italian	SaladPlace	Chinese	FastFood
2	Asian	French	SaladPlace	Japanese	Sandwich	Thai	Chinese	Italian
3	French	Asian	Japanese	Chinese	Italian	VegetarianVegan	Vietnamese	Sandwich
4	French	Asian	Japanese	Italian	Creperie	Chinese	Spanish	Tapas
...	...	...	...	...	...	...	...	...
8	French	Asian	Sandwich	Japanese	SaladPlace	Italian	Chinese	FastFood
...	...	...	...	...	...	...	...	...
12	Asian	French	Japanese	Sandwich	FastFood	PizzaPlace	Italian	Chinese
13	Vietnamese	Sandwich	Asian	Chinese	French	Italian	Japanese	Thai
14	Chinese	PizzaPlace	Japanese	Asian	Italian	French	FastFood	Creperie
15	FastFood	French	Japanese	Thai	Italian	Asian	PizzaPlace	India
...	...	...	...	...	...	...	...	...
18	Asian	French	India	African	FastFood	PizzaPlace	Italian	Japanese
19	Asian	French	FastFood	Chinese	Japanese	African	PizzaPlace	Vietnamese
20	French	Asian	FastFood	Japanese	PizzaPlace	Chinese	Italian	Sandwich

Clustering, region



We can see the following clusters :

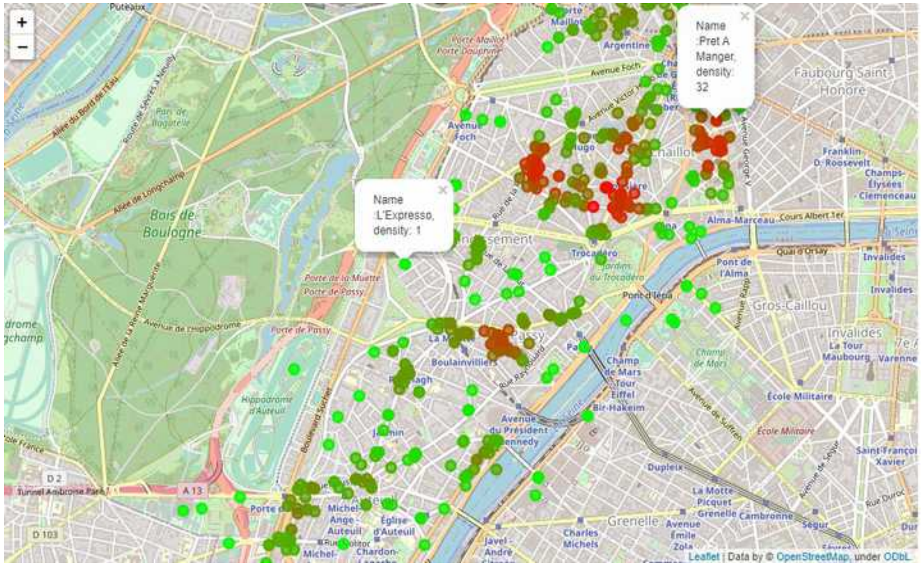
- cluster 0 : districts 1, 2, 8, 12, 16, 17 (red)
- cluster 1 : districts 3, 5, 11, 13, 14, 15, 19 (violet)
- cluster 2 : districts 9, 10, 18, 20 (turquoise)
- cluster 3 : districts 4, 6, 7 (yellow)

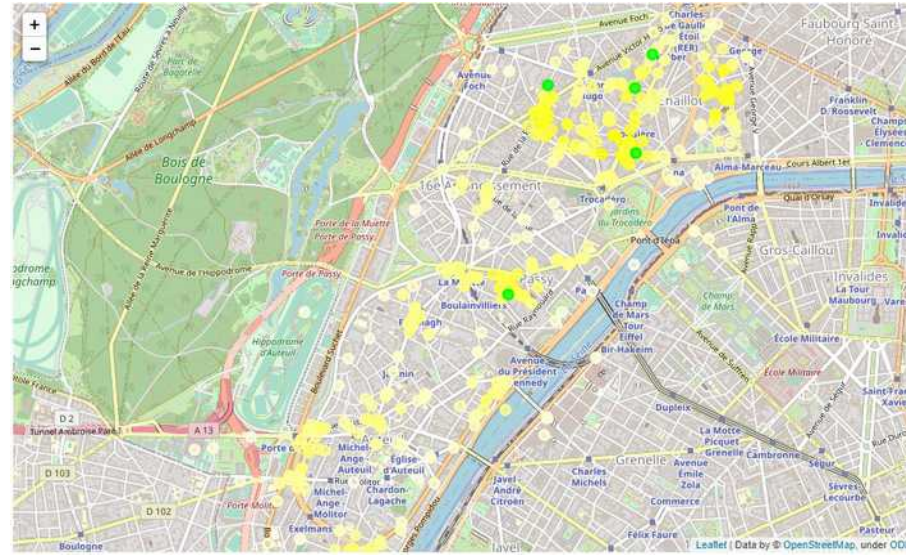


Arrd	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	3	France	Asia	Europa	Japan	Miscellaneous
6	3	France	Asia	Europa	Miscellaneous	Japan
7	3	France	Asia	Europa	Japan	Miscellaneous

Arrd	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	2	Europa	Asia	Miscellaneous	France	Japan
10	2	Asia	Europa	Miscellaneous	France	Middle east
18	2	Europa	Asia	France	Miscellaneous	Africa
20	2	Asia	Europa	Miscellaneous	France	Japan

Looking for the best location





Green points : low restaurant density

to

Red points : high restaurant density

Pale yellow points : low general restaurant density

to

Bright yellow points : high general restaurant density

Green points : low Indian restaurant density

to

Red points : high Indian restaurant density

Results, difficulties, discussion and possible improvements

- Paris is a big city with a lot of restaurants. The method made it possible to get an idea of the distribution and density of restaurants of all nationalities across all districts.
- Due to the large number of requests, I have limited the number of categories to 69. Normally, all categories should be used. Likewise, one should check what « general » categories like Asia or Japan are accounting for because some Japanese restaurants were in both national category as well as in Asia but others were not in Asia.
- Foursquare using circles to bring back the venues, it is difficult to be exhaustive in the network of districts. This can be awkward when the density of restaurants is high at the crossroads of districts. It would necessary to increase the number of requests.
- Despite this, the method made it possible to correctly define the distribution of restaurants and their density. The clustering highlighted a number of constants in the distribution of cuisines and regions across the districts.
- Thus the « fast cuisine » cluster (number 0) brings together neighborhoods where people do not have much time to eat lunch because they have a train to catch or because they work in places where there are a lot of offices and a lot of queues.
- The cluster 3 brings together neighborhoods that have a fairly old historical and gastronomic tradition (marais, Saint-Germain-des-Près, Invalides, Tour Eiffel) where many French restaurants are located.
- Asia trusts most of the first place ; It would be even more the case in adding inside the categories Japan and India.
- The owner therefore chose the 16th district and we saw that locating the restaurants as well as their density, and separating the India category from the others, allowed him to choose the right location corresponding to the strategy he had chosen. But the method would also allow him to respond to any other strategy.

Conclusion

- This method allows any restaurateur to analyze the distribution of restaurants and culinary trends by district (density) but also across districts (clustering) in Paris. And it is also valid in any city for an owner coming from another continent for the choice of the city. It is also valid for any other category of venues.
- For example, if a person wanted to live in an area where there is a lot of nightlife, or where there are a lot of services, or even on the contrary a peaceful area with few shops. Everything is possible.
- Well, of course, this is only the data part of location, density, distribution of the different categories. The future restaurateur must also do a market study depending on the location chosen to see if his business will be viable (local taxes, popularity of restaurants in the neighborhood, level of tourism and so on).