Capstone Project : Where open a new japanese shop in Paris

1. Introduction

We are a consulting society specialized in data science applied to business. Our client ask us where he should install his society. He chose Paris but didn't decide now the neighborhood. We need to make a study to find the neighborhood that suits the best. This project is very important for my client because he would have more possibilities of succeeding if the neighborhood has been chosen wisely. In the other hand, a bad choice would make a bad start and would be detrimental for the company. We must compare different neighborhood and find the ones that fits the best with the client's expectation. The client's business will be a shop specialized in Japanese food.

2. Data

2.1 Data from open data

We got all Paris' neighborhood from a csv file. This is open data that we can get with this link : https://opendata.paris.fr/explore/dataset/quartier_paris/export/?dataChart=eyJxdWVyaWVzIjpbeyJjb25maWciOnsiZGF0YXNldCI6InF1YXJ0aWVyX3BhcmlzIiwib3B0aW9ucyI6e319LCJjaGFydHMiOlt7ImFsaWduTW9udGgiOnRydWUsInR5cGUiOiJjb2x1bW4iLCJmdW5jIjoiQVZHIiwieUF4aXMiOiJuX3NxX3F1IiwiY2NpZW50aWZpY0Rpc3BsYXkiOnRydWUsImNvbG9yIjoiIzAwMzM2NiJ9XSwieEF4aXMiOiJsX3F1IiwibWF4cG9pbnRzIjo1MCwic29ydCI6IiJ9XSwidGltZXNjYWxlIjoiIiwiZGlzcGxheUxlZ2VuZCI6dHJ1ZSwiYWxpZ25Nb250aCI6dHJ1ZX0%3D. We also got another dataset about Paris' borough.

2.2 Data from Foursquare

We will massively use Foursquare in this analysis. We assume that people interested in Japanese food in Paris are also interessed in other Asian food. We will then focus on data relative to Asian food in Paris. It will be shop, restaurant... We will also cluster the different neighborhood in Paris and validate our analysis based on the clustering with data visualization. That means that we need neighborhood venues. Then we will choose the neighborhood that fits the best with all information that we will have gathered. We get also Asian restaurant and bubble tea shop that will help us for a further visualization.

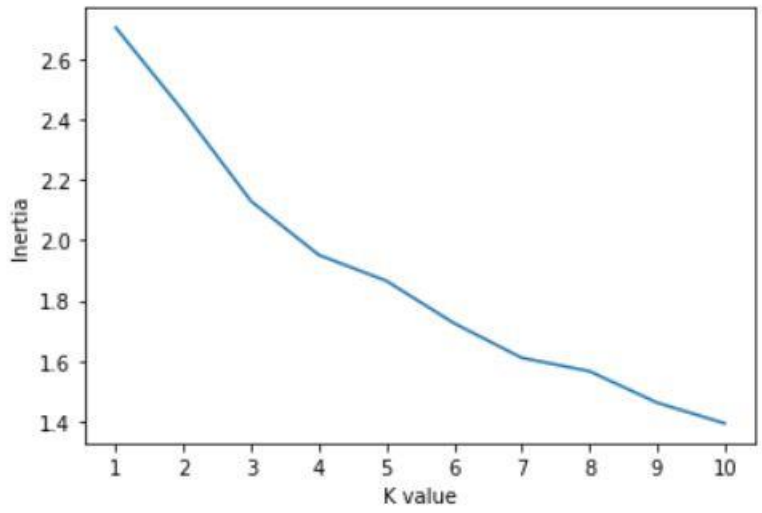
3. Methodology

3.1 Introduction

In a first time we got all the info we need thanks to open data and foursquare. Then we make a clustering by merging them with foursquare's venue data. We identify the cluster that contain the most the Asian restaurant, then we confront this with a data visualization of Asian venue. We will have good insight about the borough where we should implement the Japanese shop.

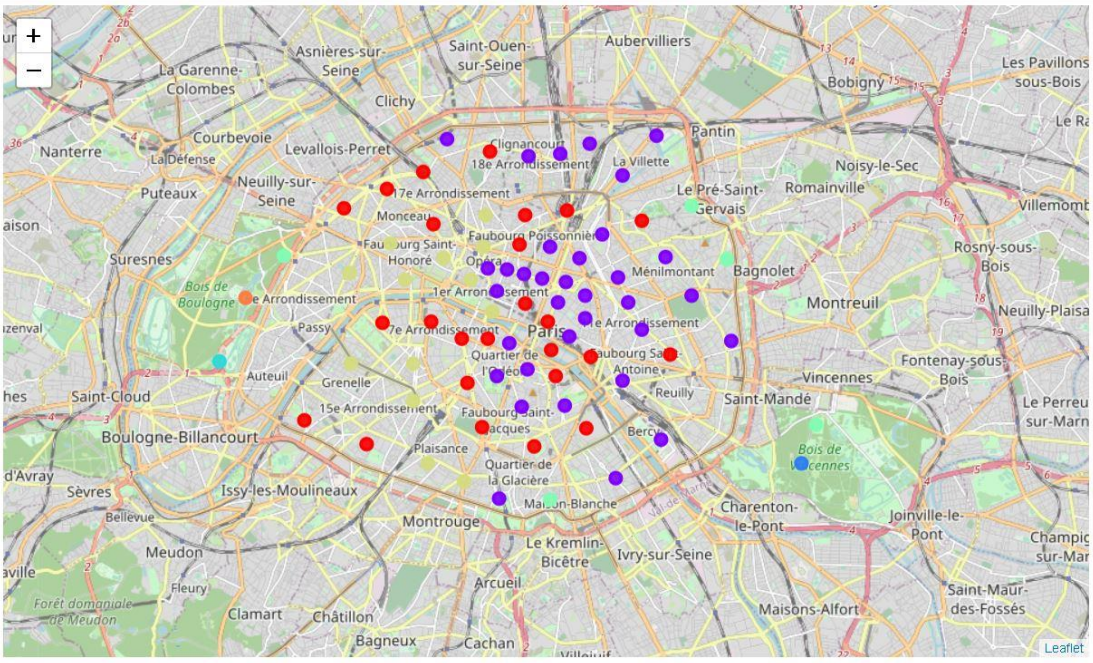
3.2 Clustering Analysis

As we must make clusters, we use the KMeans algorithms. We need first to get the best K then we create a plot of K value VS inertia. We get the following plot :



We could have chosen 4, but we chose 7 as the slope is higher between 5 and 6 then between 4 and 5. Then we initialize with random parameter the algorithm and get the minimal inertia for K=7. Then we plot the map with the clusters:

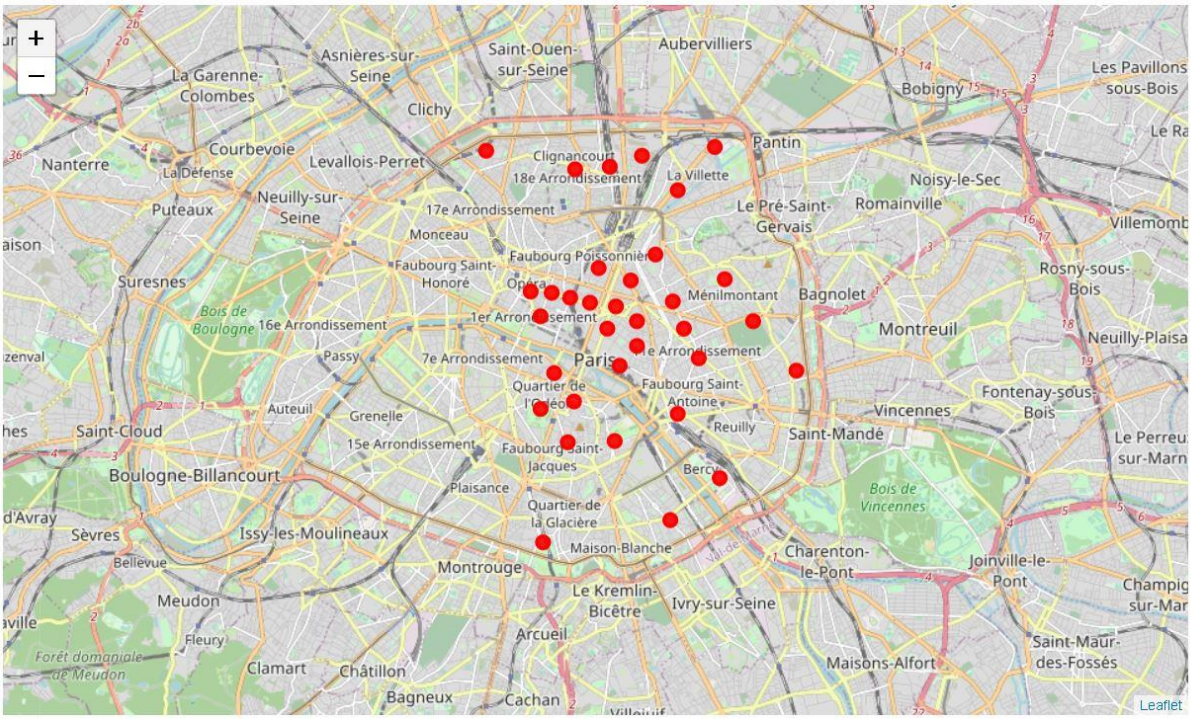
Out[211]:



We plot the value of each cluster, and he seems that the cluster that fit the most with our business problem is the second:

	Borough	Latitude	Longitude	Cluster Labels	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
1	12	48.835209	2.386210	1	Hotel	Bakery	Italian Restaurant	Supermarket	Pastry Shop	Museum	Japanese Restaurant	French Restaurant	Breakfast Spot
5	2	48.869307	2.333432	1	Japanese Restaurant	Hotel	French Restaurant	Chocolate Shop	Coffee Shop	Wine Bar	Pastry Shop	Korean Restaurant	Asian Restaurant
7	4	48.855719	2.358162	1	French Restaurant	Clothing Store	Hotel	Pastry Shop	Gourmet Shop	Cultural Center	Pedestrian Plaza	Furniture / Home Store	Ice Cream Shop
9	10	48.873618	2.352283	1	Hotel	French Restaurant	Bistro	Vegetarian / Vegan Restaurant	Bakery	Japanese Restaurant	Bar	Restaurant	Fast Food Restaurant
11	19	48.887661	2.374468	1	Hotel	Bar	Café	French Restaurant	Asian Restaurant	Italian Restaurant	Multiplex	Food Truck	Bistro
13	5	48.841940	2.356894	1	French Restaurant	Science Museum	Hotel	Garden	Bakery	Italian Restaurant	Greek Restaurant	Moroccan Restaurant	Museum
14	10	48.871245	2.361504	1	Coffee Shop	French Restaurant	Asian Restaurant	Theater	Cambodian Restaurant	Breakfast Spot	Cocktail Bar	Hotel	Bakery
15	11	48.857064	2.380364	1	French Restaurant	Bar	Bistro	Cocktail Bar	Café	Italian Restaurant	Pastry Shop	Japanese Restaurant	Supermarket
21	5	48.841684	2.343861	1	French Restaurant	Bar	Hotel	Creperie	Asian Restaurant	Chinese Restaurant	Café	Bistro	Bakery
23	20	48.863719	2.395273	1	Bakery	Bistro	Bar	Japanese Restaurant	Plaza	French Restaurant	Wine Bar	Italian Restaurant	Park
24	18	48.894012	2.364387	1	Chinese Restaurant	Supermarket	French Restaurant	Soccer Field	Gym Pool	Farmers Market	Diner	Asian Restaurant	Mexican Restaurant
25	3	48.862557	2.354852	1	French Restaurant	Art Gallery	Café	Hotel	Chinese Restaurant	Restaurant	Bakery	Garden	Gastropub
26	10	48.876008	2.368123	1	French Restaurant	Bar	Coffee Shop	Wine Bar	Café	Asian Restaurant	Pizza Place	Restaurant	Bistro
27	20	48.871531	2.387549	1	Bar	French Restaurant	Burger Joint	Café	Italian Restaurant	Japanese Restaurant	Chinese Restaurant	Thrift / Vintage Store	Bakery
31	19	48.895556	2.384777	1	French Restaurant	Asian Restaurant	Hotel	Movie Theater	Tram Station	Rock Club	Café	Shopping Mall	Science Museum

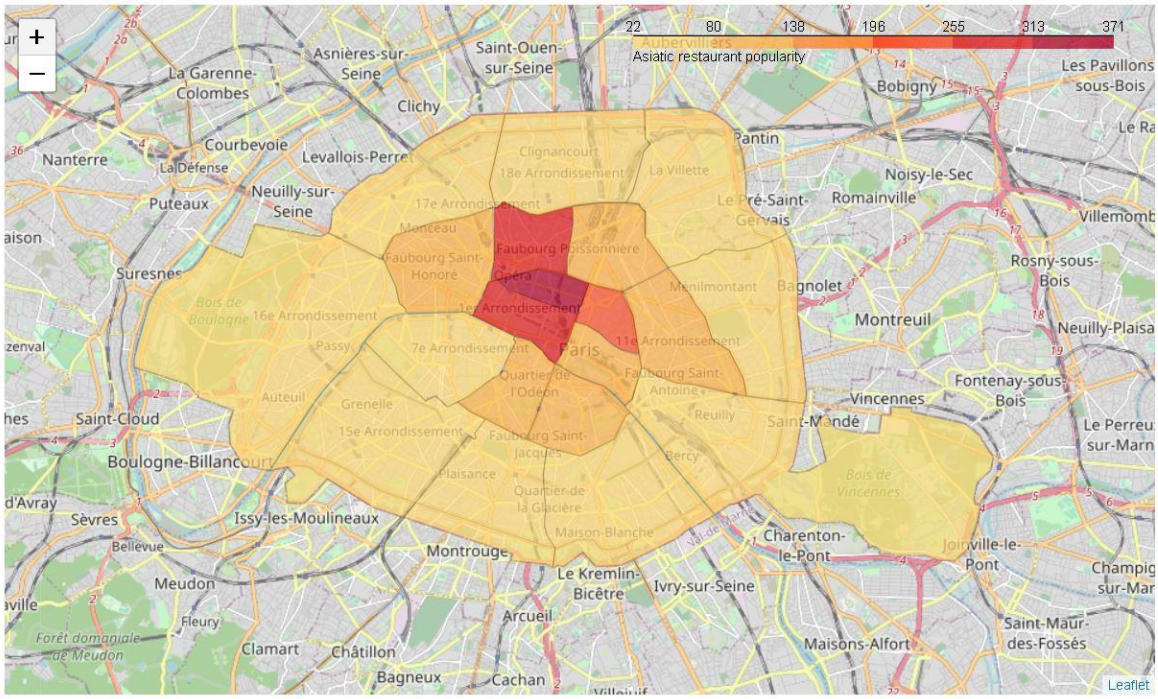
We can see plenty restaurants relative with Asian restaurant in this cluster. We can plot a map containing only that cluster:



There is a majority of clusters with Asian restaurants' venue in first, second, third, seventh and eleventh borough.

Let's verify this with a venue by borough analysis.

3.3 Borough Analysis



We can see that our assumption is verified thanks to that map visualization.

4. Result and discussion

First at all, we can see that the results are conclusive. The clusters that we made were relevant for our business problem, and there is a real contribution of the First, second and third borough, that indicates that we should make a shop in one of them. So, the results are mostly positive, but it hides a big problem with data. The data that foursquare gather are data from tourist. Then, we will have only the touristic venue, and nothing about people living in Paris. For example, the thirteenth borough has a big asian venue contribution, but this contribution can't be seen with this data. Then, we lose a big part of the information, and that could lead to a failed business, because there could be have more interesting borough or neighborhoods for the shop. We need then to make further exploration and get a dataset that would allow us to have a view on local behavior about restauration. With this exploration, we could conclude that we must invest on first borough (the most interesting), but we also saw that this exploration is insufficient because of the lack of local behavior information. The second problem is that the first, second and third borough have especially really expensive rent, that could be problematic if we pick one of them.

5. Conclusion

The purpose of this analysis was to find the best neighborhood where we should implement the Japanese shop. This analysis is efficient if we only consider tourist information, but local behavior could be interesting too. Then we can't really conclude properly, all we need is further investigation to gather more data source and analyse them.