Assessing the optimal district for a new (japanese) restaurant within the 20 arrondissements of Paris

Koji Rudolf Jaeger

1. Introduction

1.1 Business Case

The recent COVID-19 pandemic has a major impact on the food service industry, leading to many restaurants going out of business.
But eventually the economy will recover, and the freshly vacated locations in major cities, give opportunities for new businesses.

Japanese cuisine has been rapidly gaining popularity in the last decades and usually generates high revenue in the premium segment.
The city of Paris is famous for its fancy restaurants and indicates that the general population or the visiting customers, have a
appreciation for premium food and also the willingness to pay premium.

Given the prequisites, that the distance to the sea and a subsequent quality fish-market is not a deciding factor, but solely the georgraphic location,
a recommendation shall be derived for a japanese restaurant within the 20 districts of Paris.

2.2 Interest

This analysis will be mostly interested for skilled people, in the food-service industry, with the intention of opening their own food-venue in Paris, or major food-service companies aiming to expand their business.

2. Data acquisition and cleaning

2.1 Data Sources

We will be requiring data about the food-service venues, to evaluate possible competition.
For this we will be using the Foursquares data, which we can access with a free account on
https://foursquare.com/.

Additionally, we will be needing demographic data from the 20 districts of Paris. This information is available on public resources such as the official french government website (https://www.data.gouv.fr/en/datasets/arrondissements-1/) or Wikipedia (https://en.wikipedia.org/wiki/Arrondissements_of_Paris)

We will be using the Wikipedia source with Pandas Webscraping, as the .csv file from the government website is causing issues while importing.

2.2 Data Cleaning

The location data from foursquare was imported into the Jupyter Notebook via the URI.
The Data was selected to only include food-venues with a price rating of 3 or higher, which means an average food price >20USD.
Furthermore, the food venues were assigned to each district of Paris for further evaluation.

The Demographic data regarding the arrondissements from Wikipedia had some unwanted data as well as wrong datatypes for the columns.

The following steps were performed to clean this data:

- Remove Informations regarding the majors
- Rename the columns for better addressing
- Convert Area data into floats
- Set datatypes for population and density to int (if not already)
- Split the first row into its individual arrondissements with equal size

We receive the following ("cleaned") table as a result:

|	District	Name	Area	Population	Density
1	5th (Ve) L	Panthéon	2.5410	59631.0	23467.0
2	6th (VIe) L	Luxembourg	2.1540	41976.0	19487.0
3	7th (VIIe) L	Palais-Bourbon	4.0880	52193.0	12767.0
4	8th (VIIIe) R	Élysée	3.8810	37368.0	9628.0
5	9th (IXe) R	Opéra	2.1790	60071.0	27568.0
6	10th (Xe) R	Entrepôt	2.8920	90836.0	31409.0
7	11th (XIe) R	Popincourt	3.6660	147470.0	40226.0
8	12th (XIIe) R	Reuilly	16.3200	141287.0	8657.0
9	13th (XIIIe) L	Gobelins	7.1460	183399.0	25664.0
10	14th (XIVe) L	Observatoire	5.6210	136941.0	24362.0
11	15th (XVe) L	Vaugirard	8.5020	235178.0	27661.0
12	16th (XVIe) R	Passy	16.3000	149500.0	9171.0
13	17th (XVIIe) R	Batignolles-Monceau	5.6690	168737.0	29764.0
14	18th (XVIIIe) R	Butte-Montmartre	6.0050	196131.0	32661.0
15	19th (XIXe) R	Buttes-Chaumont	6.7860	188066.0	27713.0
16	20th (XXe) R	Ménilmontant	5.9840	191800.0	32052.0
17	Paris Centre 1st (Ier)	Louvre	1.3975	25049.0	4481.0
18	2nd (IIe)	Bourse	1.3975	25049.0	4481.0
19	3rd (IIIe)	Temple	1.3975	25049.0	4481.0
20	4th (IVe) R	Hôtel-de-Ville	1.3975	25049.0	4481.0

3. Exploratory Data Analysis

3.1 Data Visualization

In order to get a better understanding for the data, we visualized the location of other competing japanese food venues (marked in dark red), asian food venues (marked in red), seafood venues and other restaurants (grey).
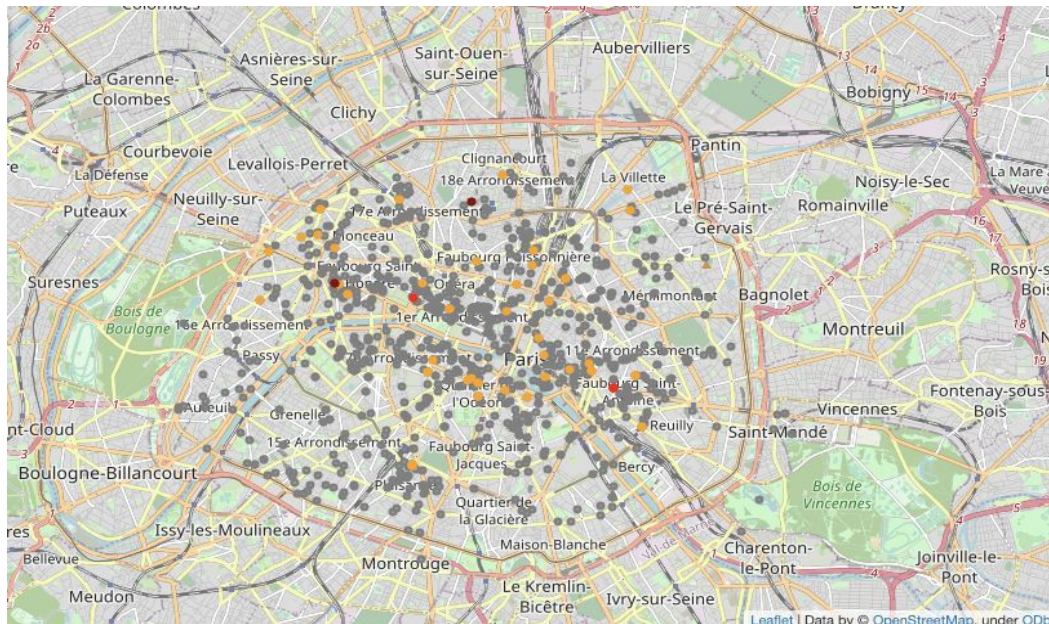
This gives a rough overview of the locations of the food venues in Paris. It can be observed, that there are only very few Japanese restaurants in this map, and even including the similar food venues such as seafood and asian food, the majority of the food venues is marked grey. Let us convert this into a bar-chart to get a better insight:
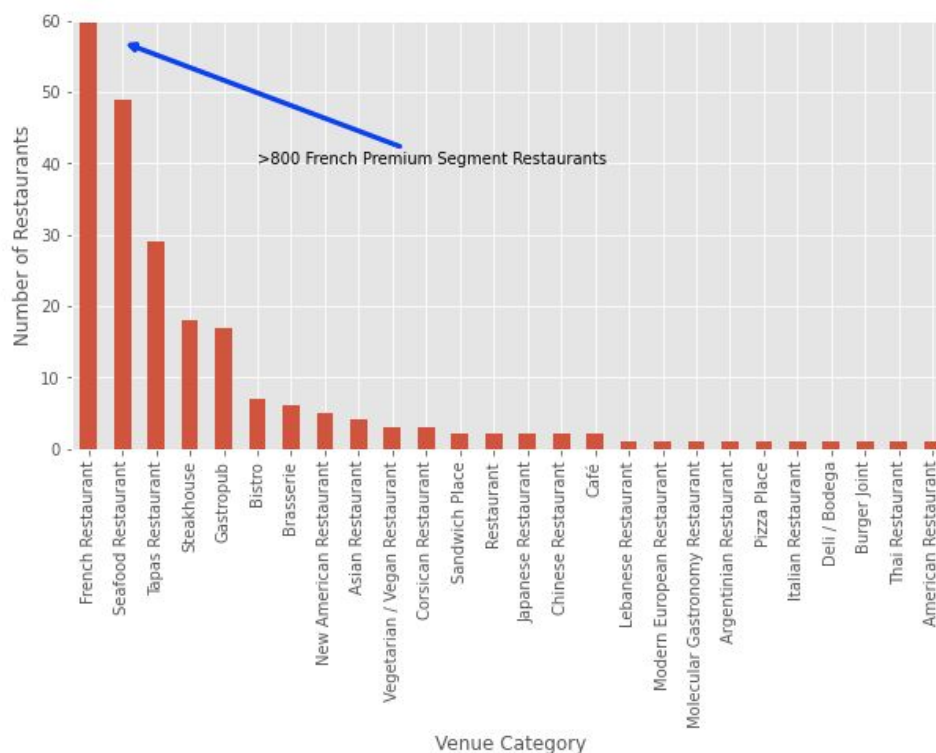
This bar chart gives us a much more detailed overview of the many food-venues in Paris. The majority of the "other" food-venues, seem to be french, as these food venues seem to outnumber any other kind.

At this point, it is indicating that the competition a japanese restaurant is facing, are not the similar restaurants such as asian- or seafood, but the french cuisine.

3.2 Methodology for deriving the optimal location

As seen above, the highest competition for a japanese restaurant seems to be the domestic french cuisine. For easier analysis, we calculated restaurant density for the related food services (asian, seafood and japanese food) as well as the total restaurant count.

	District	Name	Area	Population	Density	Restaurant_density	Rel_Restaurant_density
1	5th (Ve) L	Panthéon	2.5410	59631.0	23467.0	15.741834	1.574183
2	6th (VIe) L	Luxembourg	2.1540	41976.0	19487.0	20.427112	2.785515
3	7th (VIIe) L	Palais-Bourbon	4.0880	52193.0	12767.0	24.461840	0.978474
4	8th (VIIIe) R	Élysée	3.8810	37368.0	9628.0	15.975264	1.288328
5	9th (IXe) R	Opéra	2.1790	60071.0	27568.0	32.583754	0.917852
6	10th (Xe) R	Entrepôt	2.8920	90836.0	31409.0	15.560166	1.037344
7	11th (XIe) R	Popincourt	3.6660	147470.0	40226.0	15.275505	0.818331
8	12th (XIIe) R	Reuilly	16.3200	141287.0	8657.0	2.205882	0.183824
9	13th (XIIIe) L	Gobelins	7.1460	183399.0	25664.0	5.877414	0.000000
10	14th (XIVe) L	Observatoire	5.6210	136941.0	24362.0	6.760363	0.355809
11	15th (XVe) L	Vaugirard	8.5020	235178.0	27661.0	7.174782	0.235239
12	16th (XVIe) R	Passy	16.3000	149500.0	9171.0	5.766871	0.245399
13	17th (XVIIe) R	Batignolles-Monceau	5.6690	168737.0	29764.0	17.110602	1.411184
14	18th (XVIIIe) R	Buttes-Montmartre	6.0050	196131.0	32661.0	5.828476	0.333056
15	19th (XIXe) R	Buttes-Chaumont	6.7860	188066.0	27713.0	6.189213	0.294724
16	20th (XXe) R	Ménilmontant	5.9840	191800.0	32052.0	4.846257	0.000000
17	Paris Centre 1st (Ier)	Louvre	1.3975	25049.0	4481.0	25.044723	0.000000
18	2nd (IIe)	Bourse	1.3975	25049.0	4481.0	30.769231	1.431127
19	3rd (IIIe)	Temple	1.3975	25049.0	4481.0	19.320215	1.431127
20	4th (IVe) R	Hôtel-de-Ville	1.3975	25049.0	4481.0	29.338104	2.146691

In order to now be able to derive a recommendation, it would be helpful to visualize this in a heat map. We will be using the geojson file from the french government website together with a folium choropleth map for this:
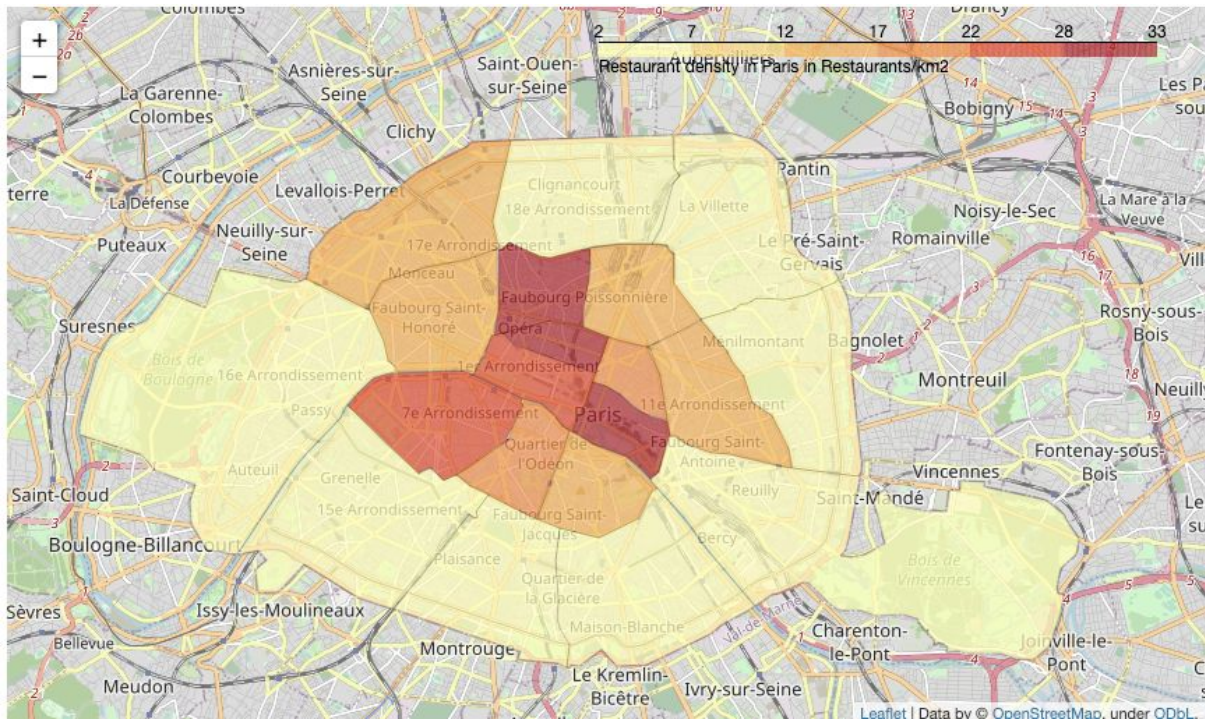
The map indicates a highly competitive district with a dark-red colour, while the less-competitive districts are marked with a light-yellow.

The heatmap shows that the major restaurant density is close to the city centers, especially in the 2 central districts : Hotel-de-ville, Bourse and the center-north district of Opera.

3.3 Side Note: Population Density vs. Restaurant Density

During the analysis, there was an opportunity to investigate the correlation between population density and (premium-) restaurant density, since the data was readily available.
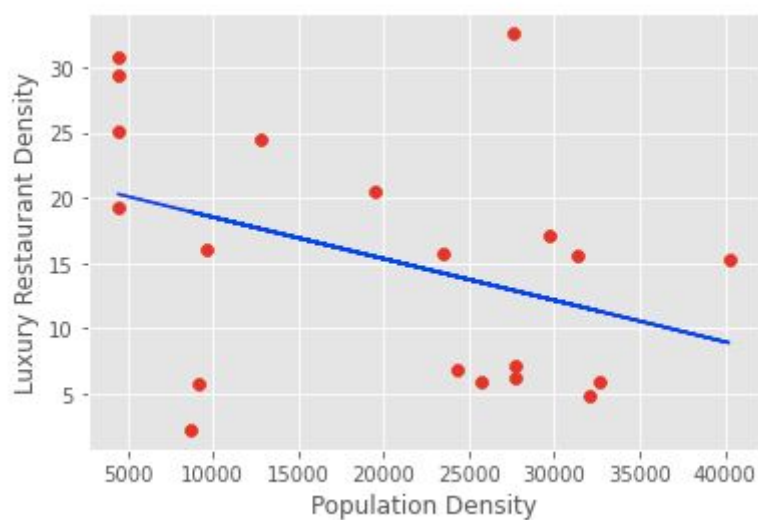
The data is only including the premium segment restaurants, as only they have been extracted from the foursquare database. This data seems to indicate that there is no correlation between population and (premium) restaurant density. Which makes sense, as

there is no necessity for the restaurants to be expensive, and also it is not necessary that the restaurant customers are domestic people.

4. Discussion

The heatmap would indicate that the restaurant density is less, in the rural areas of paris.
This is not a surprising conclusion but pretty much in-line with the common sense.
Opening a restaurant in the yellow-marked areas would indicate good business in terms of not having any competition, but would be a false conclusion, as there will also be not many customers.
At the current state of the work, we can only conclude, that the result is a step in the correct direction, but needs more refinement with more data.

5. Outlook

Data which could further improve this heat-map would be:
- Ratings of the restaurants (so we can calculate a "weighted density")
- Data on Tourists in the area, to distinguish locals from customers
- Data on classification in residential- and commercial districts

Meanwhile, a capable recommendation system will also require a better specification of the restaurant to be evaluated. If the menu of the restaurant is known, the competition can be evaluated much more detailed. Also information regarding the target demography would be valuable for a more detailed assessment. This is a similar issue as the cold-start issue for recommender algorithms.