

Capstone Project

The Battle of Neighbourhoods

**Studies on the best location for a French
gastronomy restaurant in Paris**

By

Cyrille Limousin

19/11/2019

1 Table des matières

2	Introductory section	3
2.1	Context	3
2.2	Business Cases	4
3	Methodology section	5
3.1	Data	5
3.2	Approach	5
4	Results section	7
4.1	Maps of Paris	7
4.2	Treatment of Kaggle dataset	8
4.3	Bar plot visualization	9
4.4	Boxplot visualization.....	11
4.5	WordCloud.....	13
5	Discussion section	14
6	Conclusion section	15
7	References	16
8	Acknowledgment	16
9	Appendix	16

2 Introductory section

2.1 Context

Catering is a major component of tourism in France and particularly in Paris. The purely gastronomic offer, recently listed as a World Heritage Site, arouses a legitimate expectation on the part of tourists, although it is the showcase for a much more varied catering offer, whether through fast food, sandwich shops, popular and international cuisine, etc. This diversity is an opportunity that should be valued and even encouraged. But it raises difficulties, in particular that of considering the Parisian restaurant market as a whole and the fragmentation of the data available on this subject.

Chinese and American visitors are the biggest spenders on shopping in Paris and Ile-de-France. International customers alone generated around €2 billion for shopping in Paris and Ile-de-France: Chinese and American tourists total more than €500 million, with respectively €265 million and €246 million spent, followed by the Spanish (€85 million) and Japanese (€78 million). Gastronomic catering represents the 6th largest expenditure item for tourists in France.

On the other side Parisian cultural sites welcomed 70.2 million visitors in 2017. This represents an increase of 5.9% over 2016. The cultural attendance survey, conducted each year by the Paris Convention and Visitors Bureau (OTCP), lists the attendance data for Paris museums and monuments (permanent collections and exhibitions) over the past year. The 2017 survey covers 62 sites and 68 temporary exhibitions. Notre-Dame de Paris Cathedral and the Basilica of the Sacré-Cœur de Montmartre are the two most visited monuments in the capital with 12 million and 10 million visitors respectively in 2017. Third in the ranking, the Louvre Museum is growing again with more than 8 million visitors and an increase of +14.8% compared to 2016. It remains the most visited museum in the world. The Eiffel Tower, with 6.2 million admissions in 2017 (+4.6%), has crossed the symbolic barrier of... 300 million visitors since the 1889 Universal Exhibition. The cumulative attendance of the sites considered in this survey reached 70.2 million visitors, an increase of +5.9% compared to 2016.

The Top 5 of the number of visits to Parisian cultural sites in 2017:

- Notre-Dame de Paris Cathedral: 12,000,000
- Basilica of the Sacred Heart of Montmartre: 10,000,000
- Louvre Museum: 8,022,300
- Eiffel Tower: 6,207,303
- Centre Pompidou: 3,337,000

2.2 Business Cases

Our client, a gastronomic chef, would like to open a new gastronomic restaurant. We therefore propose to help him to choose the best site in Paris while taking advantage of the tourist opportunities that the largest cultural sites in Paris can offer and we will study in this analysis the 5 most visited sites in Paris and their neighborhoods.

3 Methodology section

3.1 Data

For this project we will use the Foursquare API to identify the main venue in the neighborhood of the 5th more popular tourism site in Paris, around 500 m to identify the different activity and identify classify them to help our client to better understand the local context.

We also use a data set available on Kaggle based on data from TripAdvisor and allow us to have several complementary data like:

- Cuisine Style: cuisine style(s) of the restaurant, in a Python list object (94 046 non-null)
- Ranking: rank of the restaurant among the total number of restaurants in the city as a float object (115 645 non-null)
- Rating: rate of the restaurant on a scale from 1 to 5, as a float object (115 658 non-null)
- Price Range: price range of the restaurant among 3 categories, as a categorical type (77 555 non-null)

The dataset address is proposed in annex part.

3.2 Approach

First, we will propose a visualization of the selected site on the map of Paris, therefore we will create a first data set containing the localization information. The latitude and longitude of the top 5 tourism site are:

- Notre Dame de Paris : 48.853, 2.349
- Basilica of the Sacred
- Heart of Montmartre : 48.885, 2.337
- Louvre Museum : 48.860, 2.338
- Eiffel Tower : 48.858, 2.294
- Centre Pompidou : 2.352, 48.86

Using foursquare API we will identify the different venues in the 500m around these five locations and focus on the catering environment more precisely on categories like Hotel and restaurant which have the more impact on the activity of our client.

In a third time we will add the data from Kaggle to measure competition in terms of quality and cost and help to choose the best emplacement.

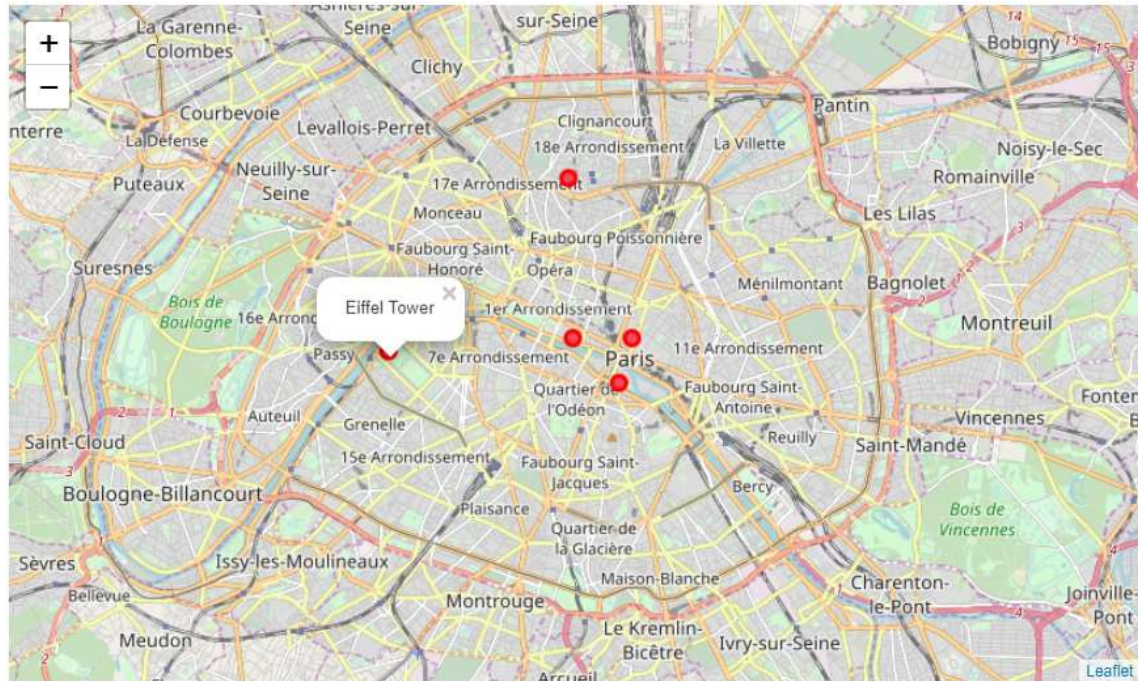
Indeed, one wonders if it is interesting to settle in a neighborhood where the quality of the restaurant is average but at the risk of having a clientele less conducive to a gourmet restaurant.

On the other hand, our client may be preferred to settle in a neighborhood where gastronomy is already "queen" and where he will be sure to find the right clientele.

4 Results section

4.1 Maps of Paris

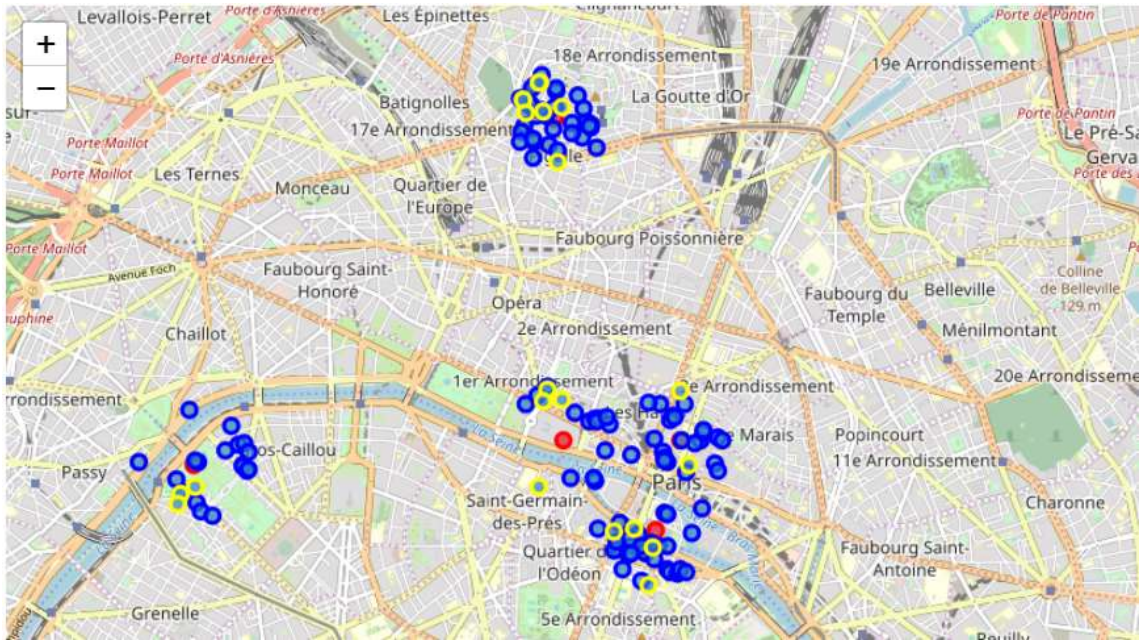
The first step for this studies is to localize the tourism site identified before on the map of Paris :



On this picture we have identified the Eiffel Tower, on the north of the Eiffel Tower you have the Sacred Heart, and on the east, you will find the three others site which are more closer.

In the second step we have used the foursquare API in order to identify all the venue around the 5 tourisms sites.

Based on this selection we have added marks in yellow to identify the "hotel" categories and with blue mark we have place the restaurants and in blue mark for restaurants.



We find that overall the density of hotel (in yellow on the map) is relatively close between the 3 areas with perhaps an advantage in the area of the Sacred Heart. The density of restaurants (in blue on the map) is higher in the area of the Louvre museum, the Centre Pompidou and Notre Dame but this area includes as we have seen 3 tourists' sites so also covers a wider area.

4.2 Treatment of Kaggle dataset

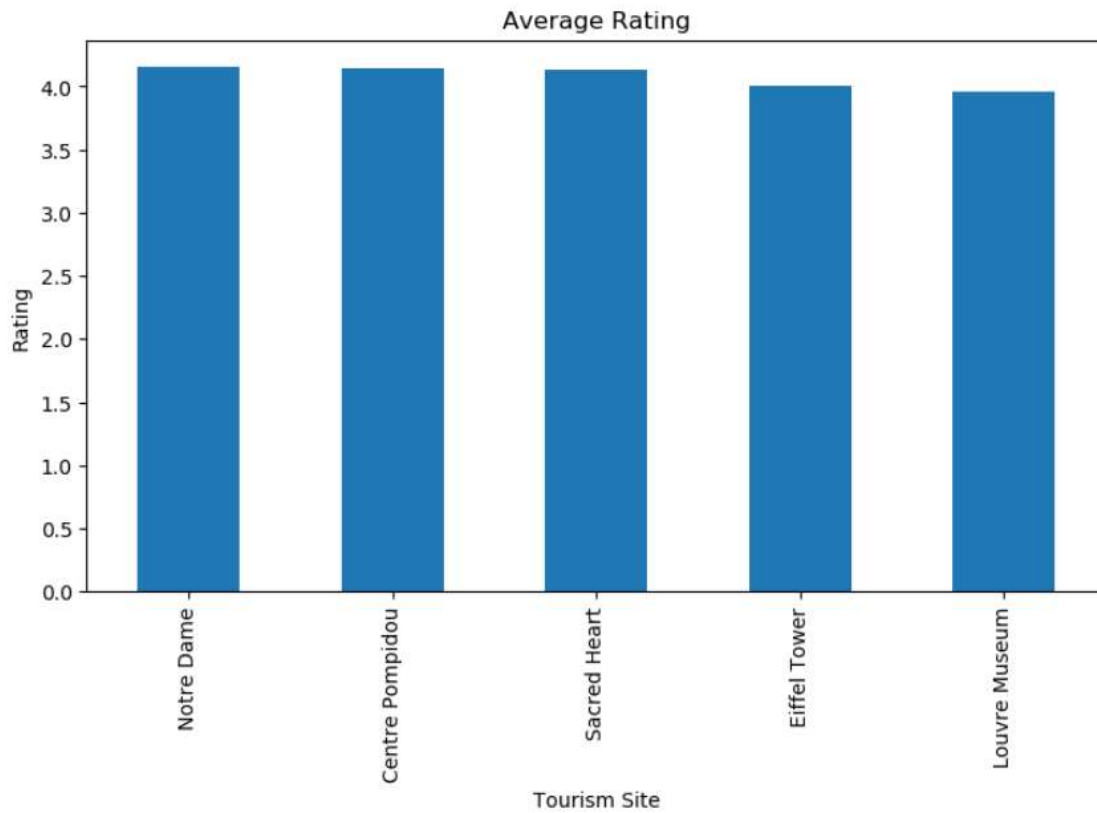
In this next step we will treat the dataset from Kaggle and merge it with the venue identifying by foursquare. We have imported the data added in watson platform, retreated the name of the restaurant by suppressing accent and other specials characters etc... to improve the matching between the data set. We obtain the following dataset:

	site	Venue	Category	City	Cuisine	Ranking	Rating	Price	Reviews	Comments
0	Notre Dame	SOLA	Japanese Restaurant	Paris	['Fusion', 'French', 'Japanese', 'Vegetarian F...]	189.0	4.5	70.0	664.0	Birthday trip', 'A 1 presentation',
2	Notre Dame	LE PETIT CHATELET	French Restaurant	Paris	['French', 'European', 'Vegetarian Friendly']	247.0	4.5	45.0	1454.0	Don't Miss It, 'Good experience',
3	Notre Dame	LA NOUVELLE SEINE	French Restaurant	Paris	['French', 'European']	10231.0	3.5	45.0	208.0	Average food, very average', 'Friendly athmosp...
4	Notre Dame	SOUIRE TAPAS FRANCAISES	Tapas Restaurant	Paris	['French', 'European', 'Wine Bar', 'Vegetarian...]	166.0	4.5	45.0	705.0	Loved the flavours!!', 'Inventive French Small...

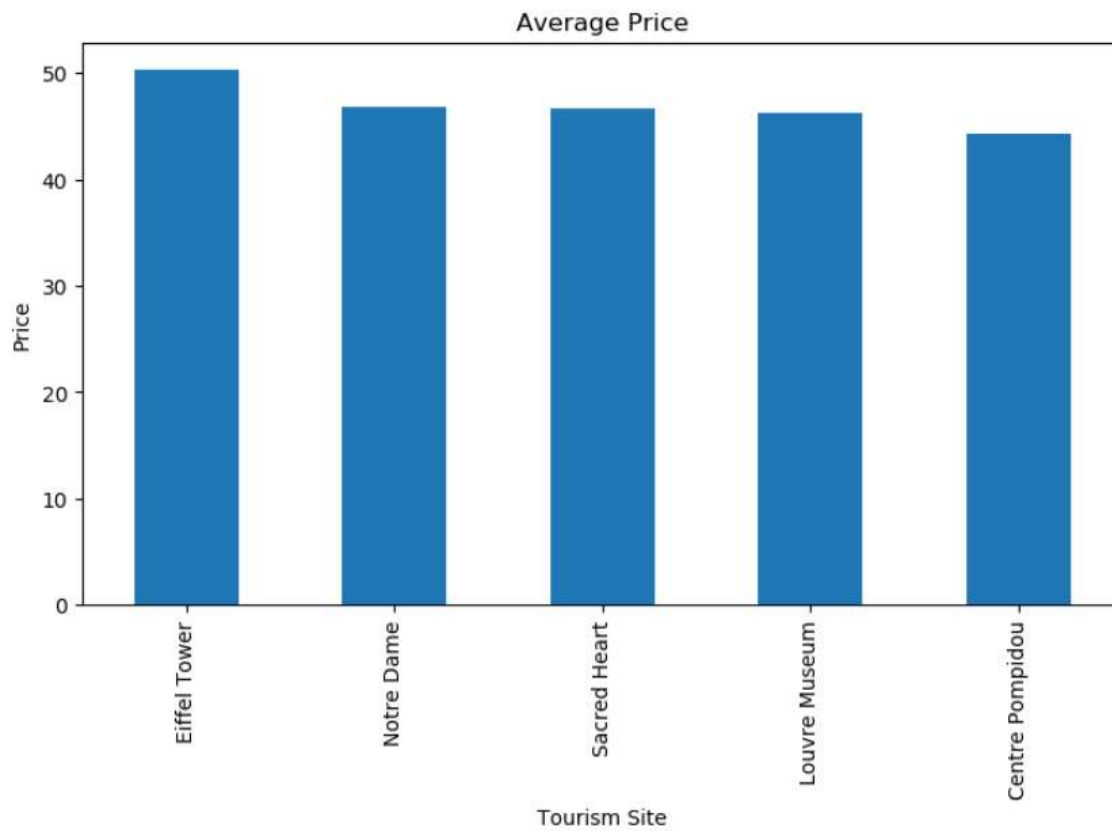
4.3 Bar plot visualization

After retreatment of this new dataset we present the following plot bar :

- Average Rating
- Average Price
- Average number of review

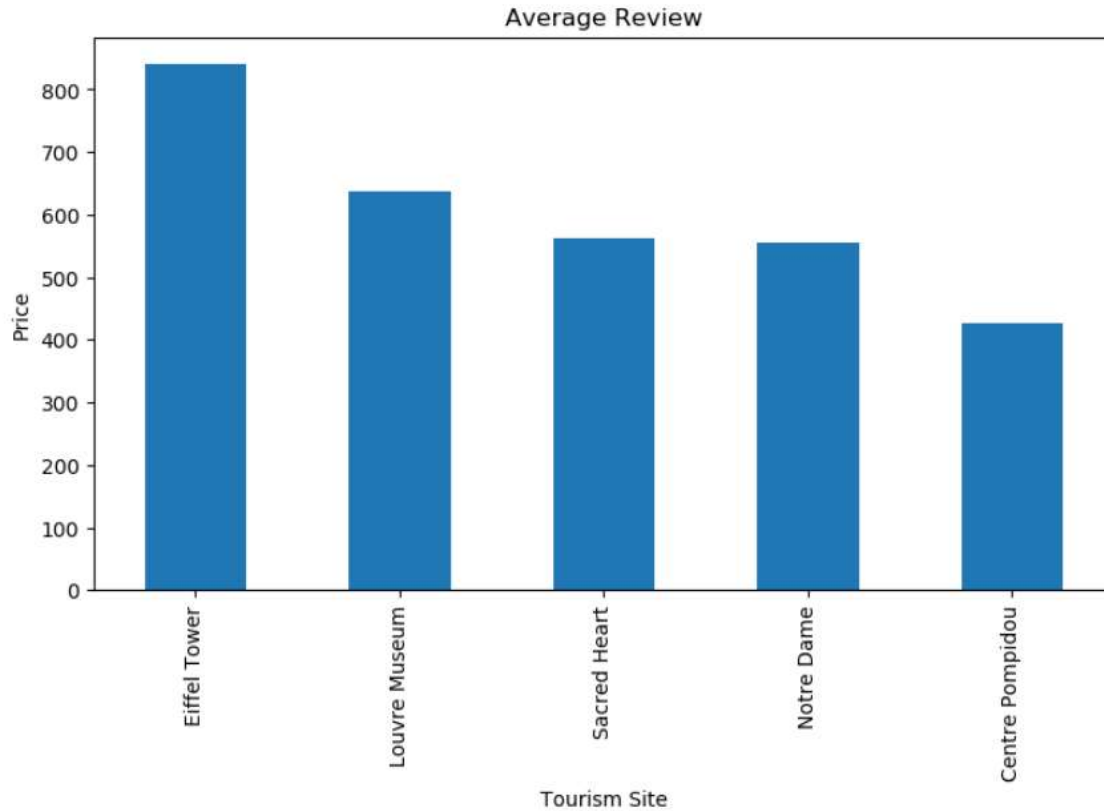


Overall the average ratings are comparable although we find that the first 3 sites Notre Dame de Paris, Centre Pompidou, Sacred Heart have a higher average rating overall.



Our approach to the average price is debatable because we do not have precise figures (table number, average basket, ...).

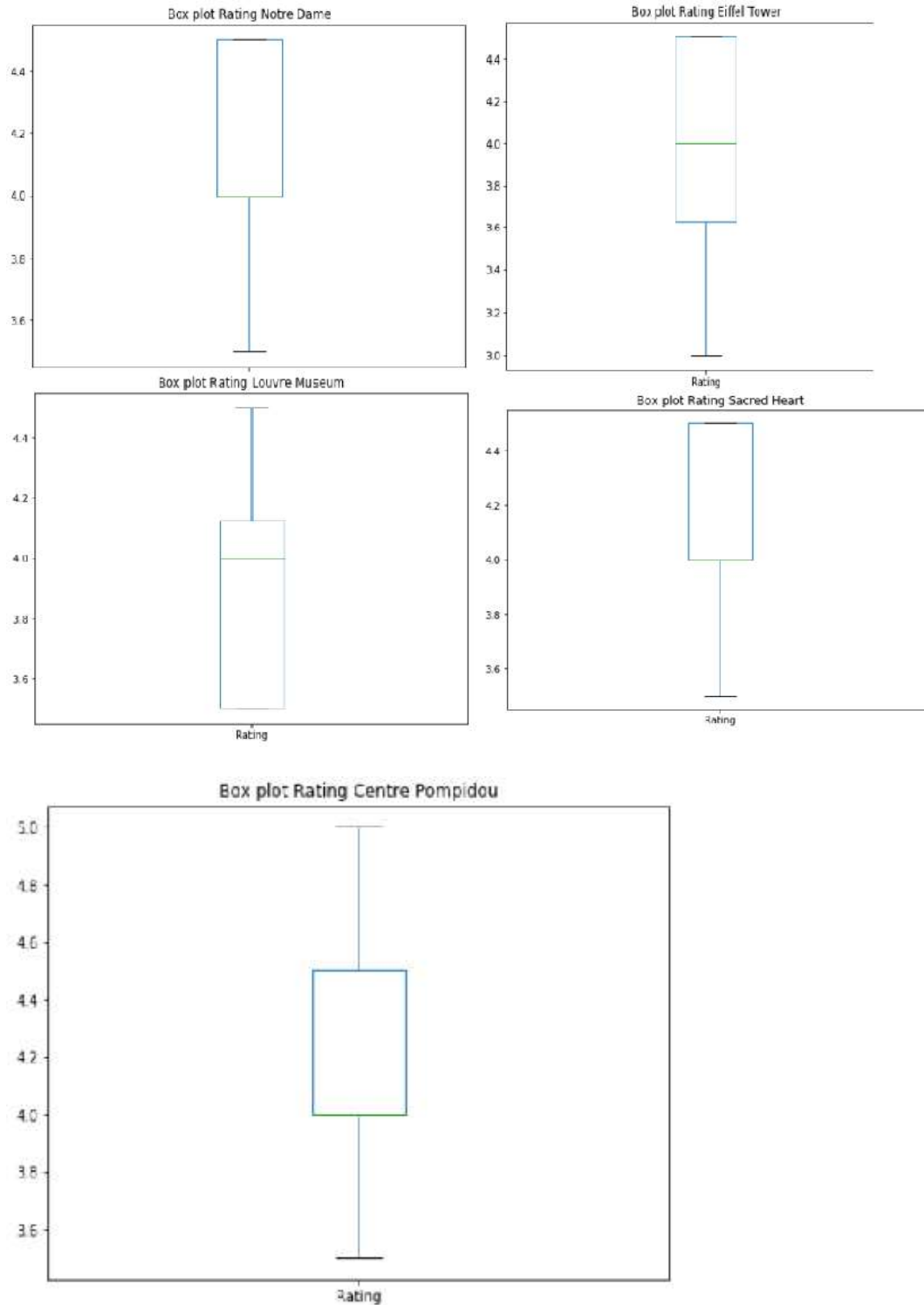
We based on an average price appreciation. However, if overall the results remain close we note however that the site of the Eiffel Tower stands out with a higher average price of restoration.



If we take the average number of recommendations, we recall that this data comes from the trip advisor site it seems quite clear this time that the restaurants of the Eiffel Tower site get scans much more than on the other sites. One can therefore think that the restaurants there are more visited /popular than those of other sites.

We then offer boxplots on the different ratings in order to get a better idea of their distribution.

4.4 Boxplot visualization



The lessons that can be learned is that customer satisfaction is relatively homogeneous with median scores almost always at 4.

On the other hand, there are higher high scores in the area of the Centre Pompidou and the opposite of low notes around 3 in the vicinity of the Eiffel Tower.

4.5 WordCloud

Finally, we propose, as we have the information a wordcloud representation of the comments collected via TripAdvisor:



5 Discussion section

Our chef's goal is to open a gourmet restaurant.

As we saw at the introduction there is a real demand in France for this type of positioning since gastronomy represents the 6th item of expenditure of foreign tourists.

It is obviously not to have enough regular clientele for this kind of establishment, so it seems appropriate for our client to supplement this clientele with the benefits of tourism.

From the beginning of the analysis we can see that the betting center with Notre Dame, the Pompidou Centre and the Louvre Museum is a strong magnet and therefore seems at first the ideal site.

However, this tourist attraction could also have a negative effect as tourists in this area might not take the time to settle down and enjoy a gourmet restaurant.

Added to this is the fact that despite the presence of these three important tourist sites the hotel offer does not seem so much wider than in the other two areas studied, from the moment our consumer chooses to take the metro or VTC transport, access other sites is not much more problematic.

Regarding the Sacred Heart, the overall situation makes that site is even off-center, so on the other hand we notice the presence of a little more Hotel but this time we also notice a very high density of restaurant.

From our analysis by Bar plot or Boxplot the neighborhood of the Sacred Heart never emerges.

Finally, the Eiffel Tower district does not really stand out because of the presence of Hotel. There are 3 hotels within 500 meters of the site, which is on average the same as for the site Notre Dame & Co only the Sacred Heart offers a better offer.

However, what seems to us to be two very significant advantages to our client's positioning is that we have that the average cost of catering around the Eiffel Tower is significantly higher, which can be a definite advantage for our positioning.

The other information that we think is strong for this site is the number of reviews we were able to get via the trip advisor site which there is far superior to other site even though the number of restaurants in this area is clearly not the highest.

6 Conclusion section

Our advice to our client would be to move to the Eiffel Tower site, the two major assets seeming to us to be the average cost of the restoration quite high which will not be too much of a problem. In addition, the density of restaurants in this area seems to us overall lower but above all the number of comments is still much higher, reflecting for us qd even a sharper peak of attendance.

This can also be explained by the fact that the Eiffel Tower is located in a rather upscale area of the capital thus potentially a potentially wider home.

Our second choice will be Châtelet which, as can be seen on the map, is a district in the centre of the 3 sites Centre Pompidou, Musée du Louvre and Notre Dame.

Finally our last choice will be on the sity of the Sacred Heart because of its location on the edge of Paris, the high density of restaurant at the first of the site and therefore the lack of visibility typically on the internet.

7 References

Kaggle dataset used : https://www.kaggle.com/damienbeneschi/krakow-ta-restaurans-data-raw/download/Z7HYhsoNPob8GBTNIhdy%2Fversions%2F27ds2CzTZzEGIkPGCDa9%2Ffiles%2FTA_restaurants_curated.csv?datasetVersionNumber=1

8 Acknowledgment

None

9 Appendix

Notebook address : <https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/7ba0627c-acf5-49b0-bd1a-834a6333bdae/view?projectid=68794967-b21d-42a3-9ac2-2a565c75d9ff&context=analytics>