Capstone Project - Airbnb - Paris, France

April 29th 2020

Florence D'AMORE

Problem

Airbnb has seen an enormous growth, with the number of rentals listed on its website.

Making it difficult to choose a location.

This project concerns a family who wishes to visit Paris, in France, at walk and by subway.

This family need an apartment, near market, bakery and subway.

The data science can help to prepare their visit.

Data acquisition and cleaning

Data source :

- Inside Airbnb website : listings.csv.gz , Paris, on March 15th 2020. Give the updated listings of rentals by neighborhood; price; type of accommodation; minimum of nights; number of bedrooms…

- Foursquare location data : give the location of market, bakery and subway.

Data cleaning :

- Price are in dollars.

- Keep only the columns interested for this project.

- There are no duplicate values, and values have been made uniform.

- As this project is at destination of a family, in the *room_type* column, only the type 'Entire home/apt' has been selected.

- After cleaning dataset, there are 58177 rows and 16 columns.

Mean of rental price

The mean of rental price is $123.

Figure 1 : Box plot of rental price
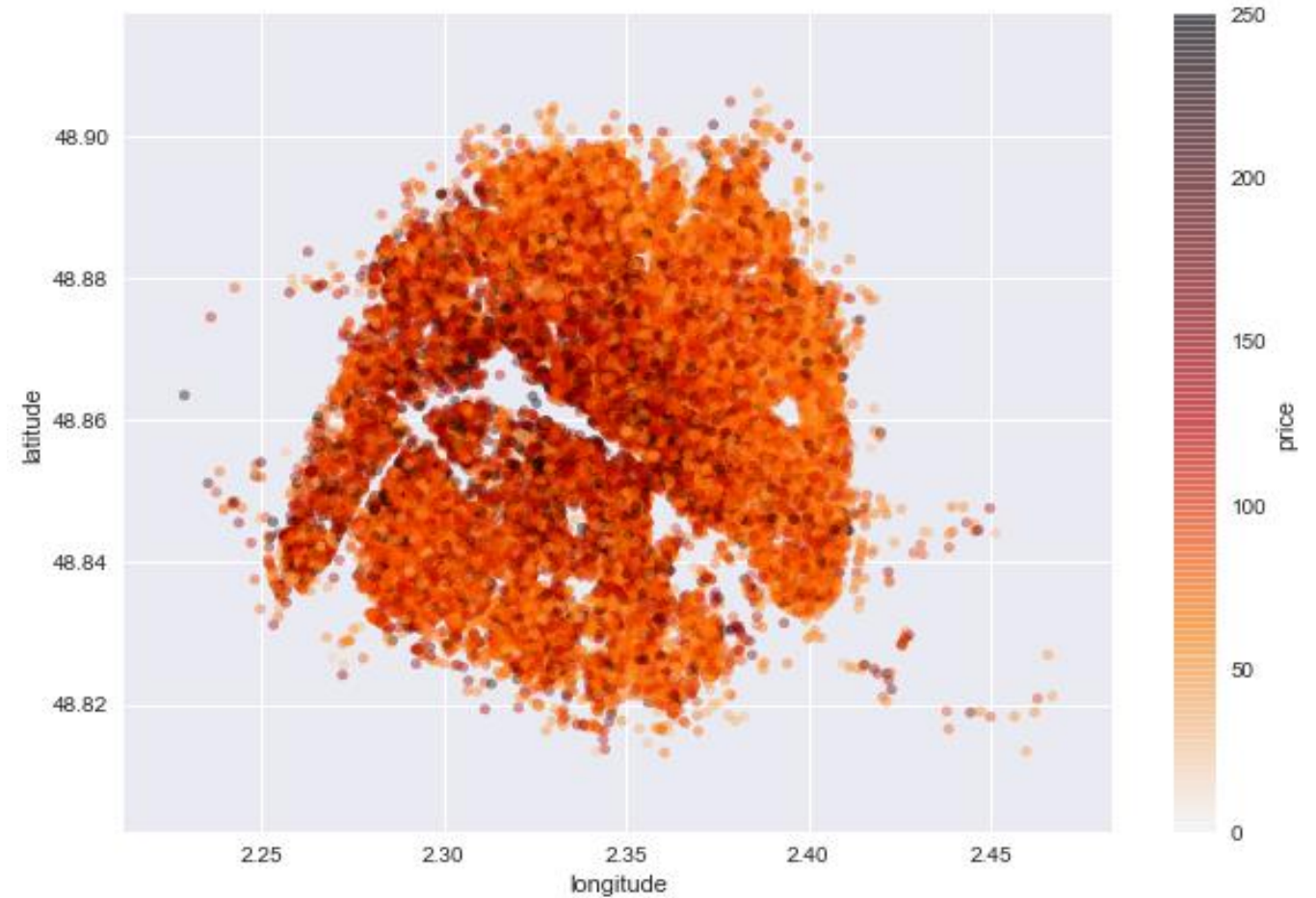
Price by coordinates

Figure 2 : Distribution of rental price according to the coordinates

Rental price by neighborhood

The most central neighborhood of Paris, are the most expensive.
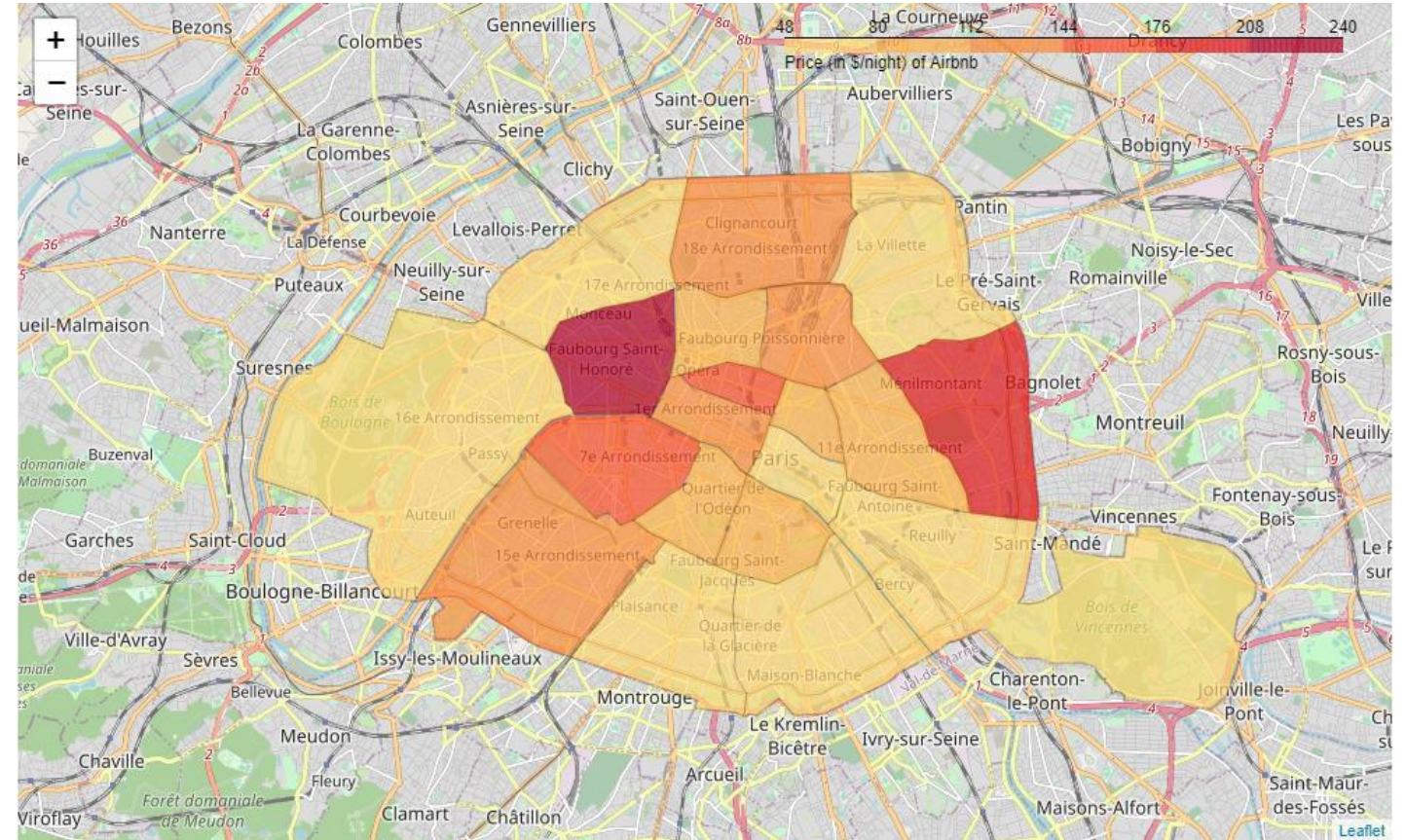

Figure 3 : Choropleth by neighborhood

Rental price by number of bedrooms

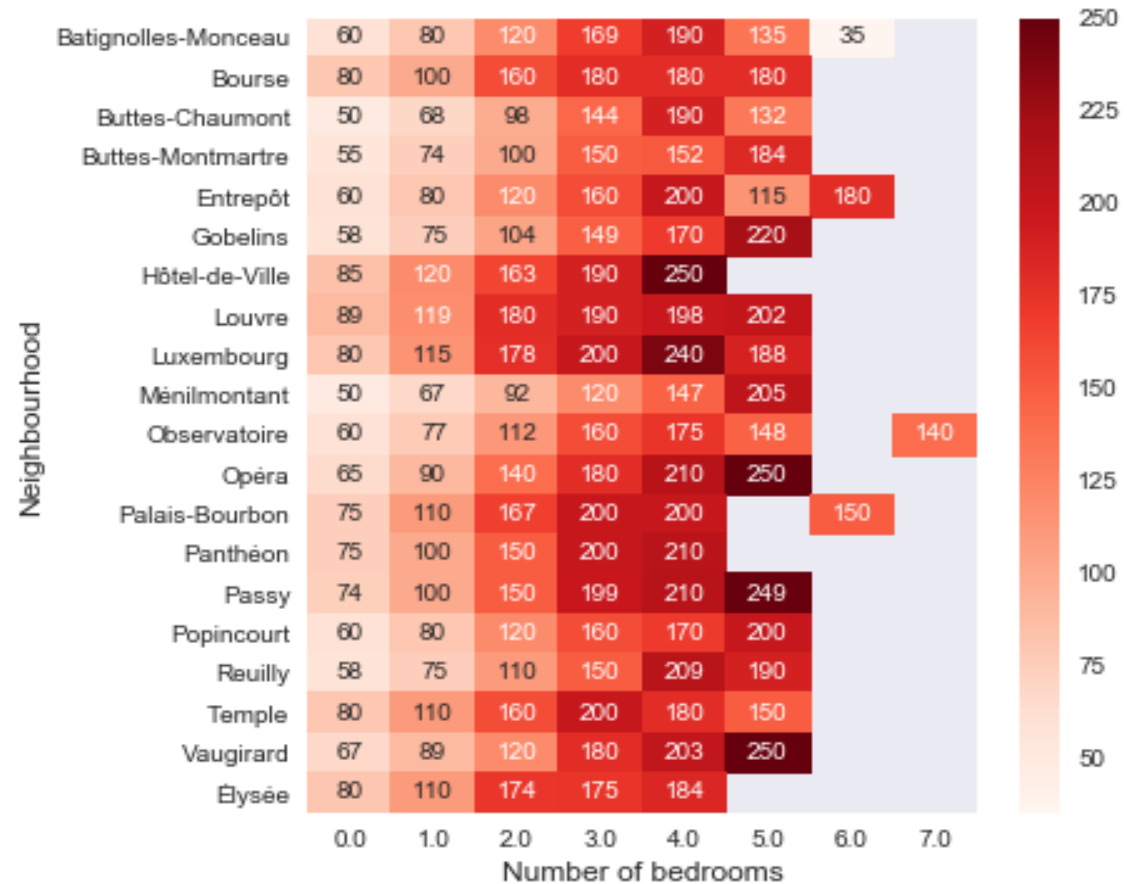The fewer rooms there are in rental; the lower the price is.

Figure 4 : Relationship between price, number of bedrooms and neighborhood

Amenities

The majority of rental has got a kitchen, wifi, washer and TV.

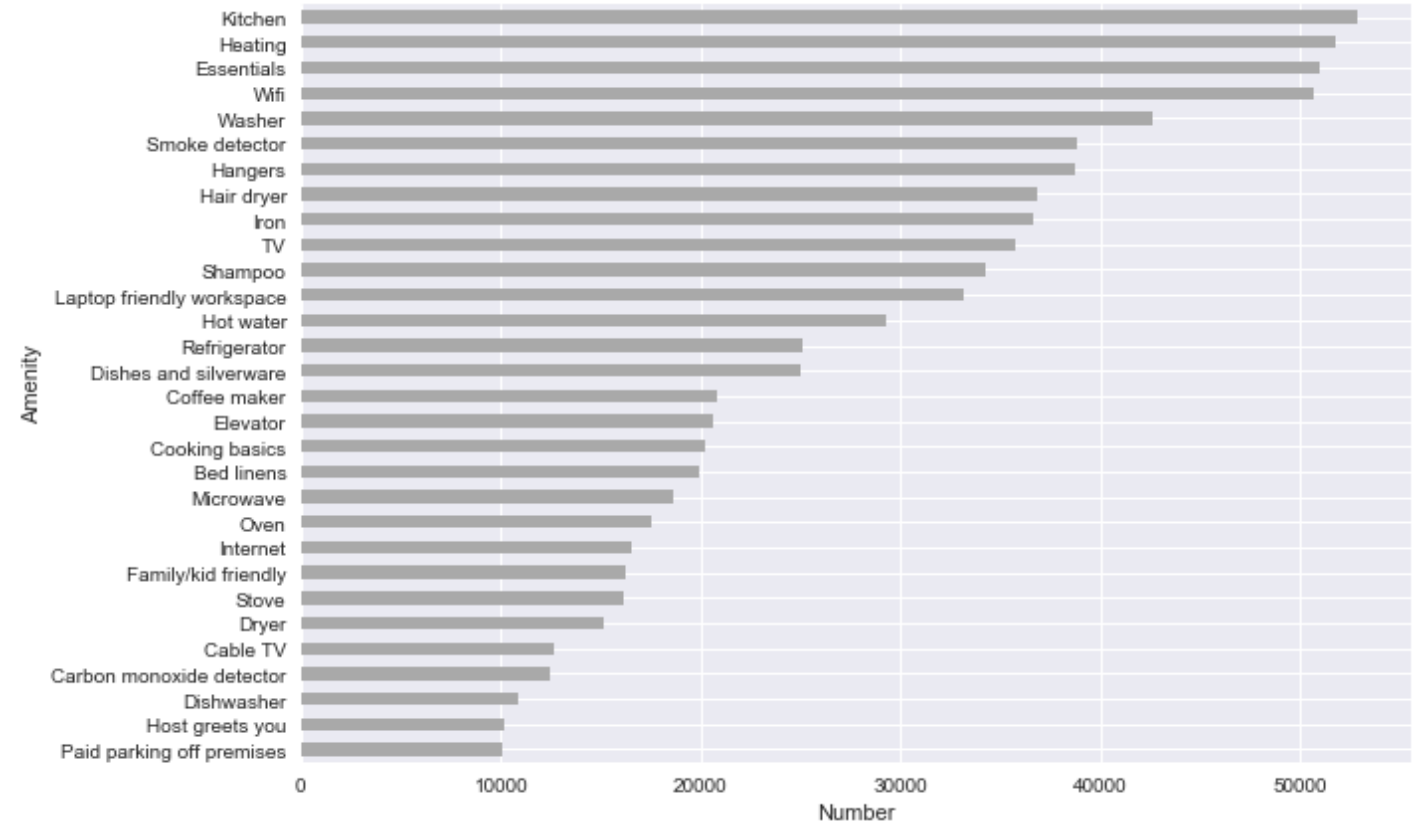Some hosts mentioned that they have carbon monoxide detector.

Figure 5 : Number of amenity

Rental price by property type

The highest rental price is for the houseboat.

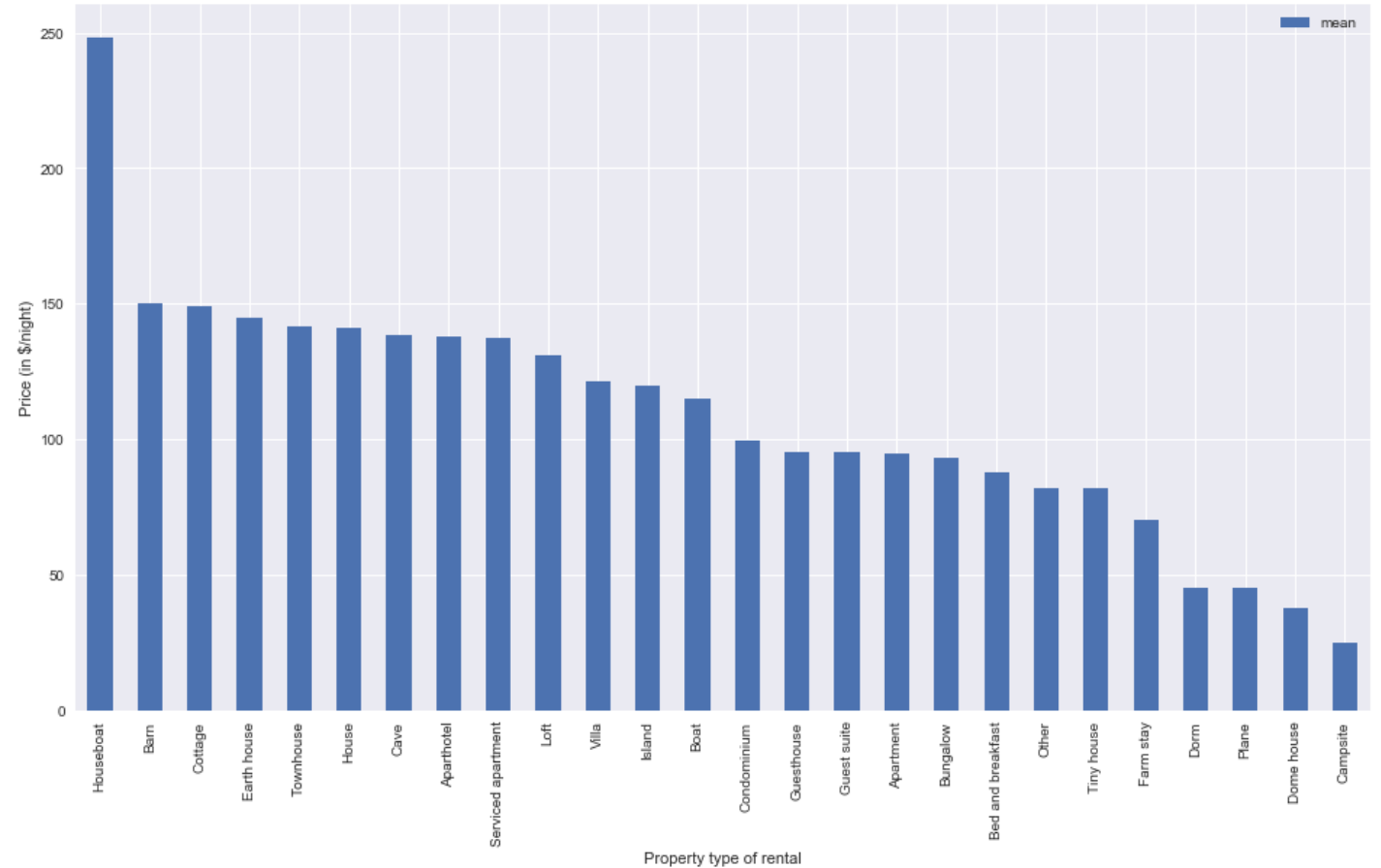And the lower price is for campsite.

Figure 6 : Rental price by property type

Rental price by security deposit and cleaning fee

It seems that the lower rental price, the lower security deposit and cleaning fee are.
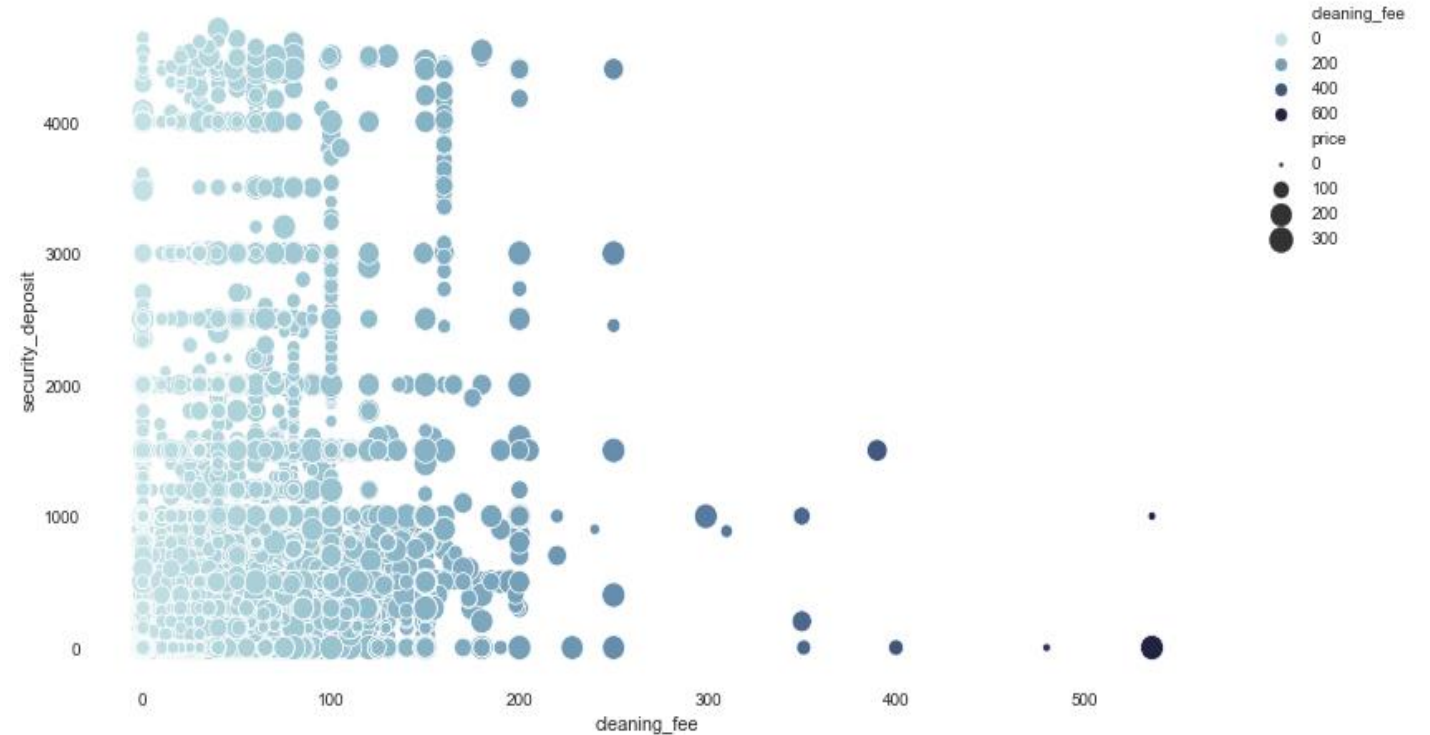
And vice versa.

Figure 7 : Relationship between the rental price by security deposit and cleaning fee

Linear regression : to predict rental area value (Size)

- Splitting data

- Make linear regression and prediction.

The mean of size is 56,94m².

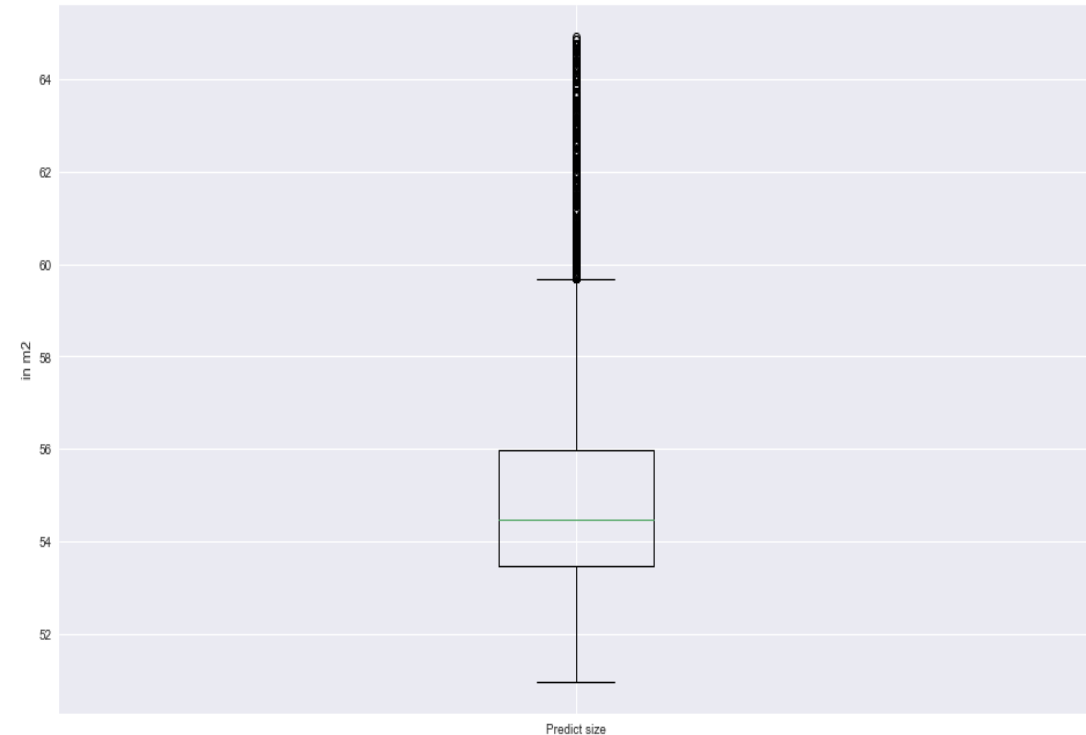|	Predict size (in m²)
Count	25 794.00
Mean	56.94
Std	13.70
Min	50.97
25%	53.47
50%	54.68
75%	56.38
Max	477.95

Figure 8 : Mean of the predict size

Foursquare location data

Localization of bakery, subway and market, near the accommodation.
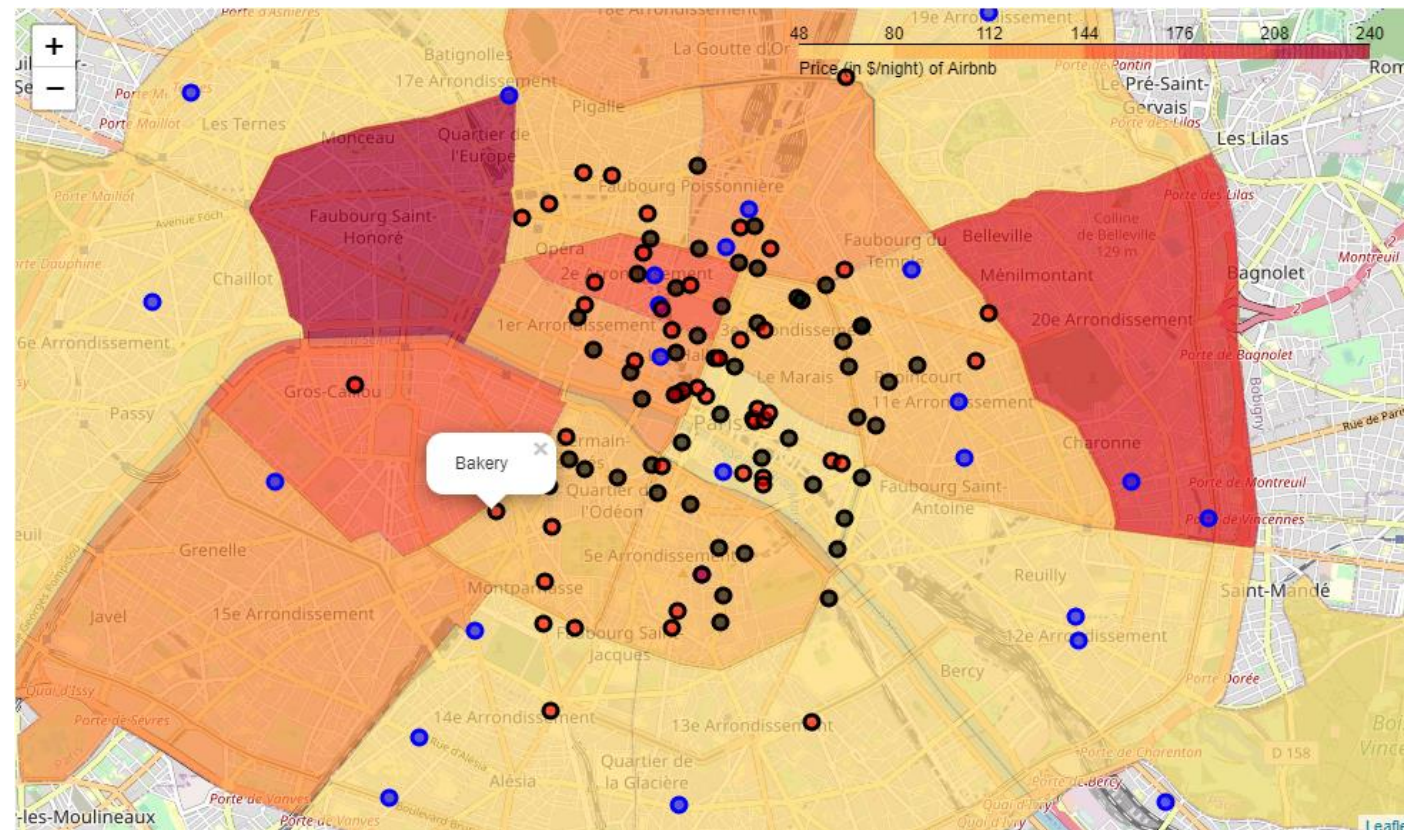

Figure 9 : Choropleth by neighborhood, with bakery, market and subway, in the center of Paris

Conclusion

The rental price depends of :

- the geographic position (In the center of Paris, is the most expensive),

- number of bedrooms,

- the size of accommodation.

The location in the center of Paris have the most amenities (Market, Subway…), near the location.

There are some values outliers, in the dataset. You must take all the information before booking, to avoid unpleasant surprises.

Customer satisfaction values should be added to this project.

Sources

☻ ["Predicting Prices: XGBoost & Feature Engineering"](#) , by Britta Bettendorf

☻ "Predicting the Improvement of NBA players", by Zhenfeng Liu : [report](#) and [presentation](#)

☻ ["Housing Sales Prices & Venues Data Analysis of Istanbul"](#) , by Sercan Yıldız

☻ GeoJSON data in Paris : [arrondissements.geojson](#)