DATA MANAGEMENT - ECON2306-1

INSIDE AIRBNB

PARIS

Rachelle DAOUDI
Larissa DUTRA
Edom SILESHI

Why this topic ?

- For **ERASMUS Students** who would like to visit Paris while they are still in Europe/close to France

- **Save money** on accommodation to do other activities in the city

The question we tried to answer :

Where to find the cheapest accommodation in Paris?

Data cleaning [Standard]

Removal of specific columns: irrelevant or not necessary for the analysis (host, URL,...)

Deleting rows with missing values: from 'review_scores_rating' and 'review_scores_accuracy'

Price transformation: from text to floating point number

Creation of new variables: 'price_per_person'

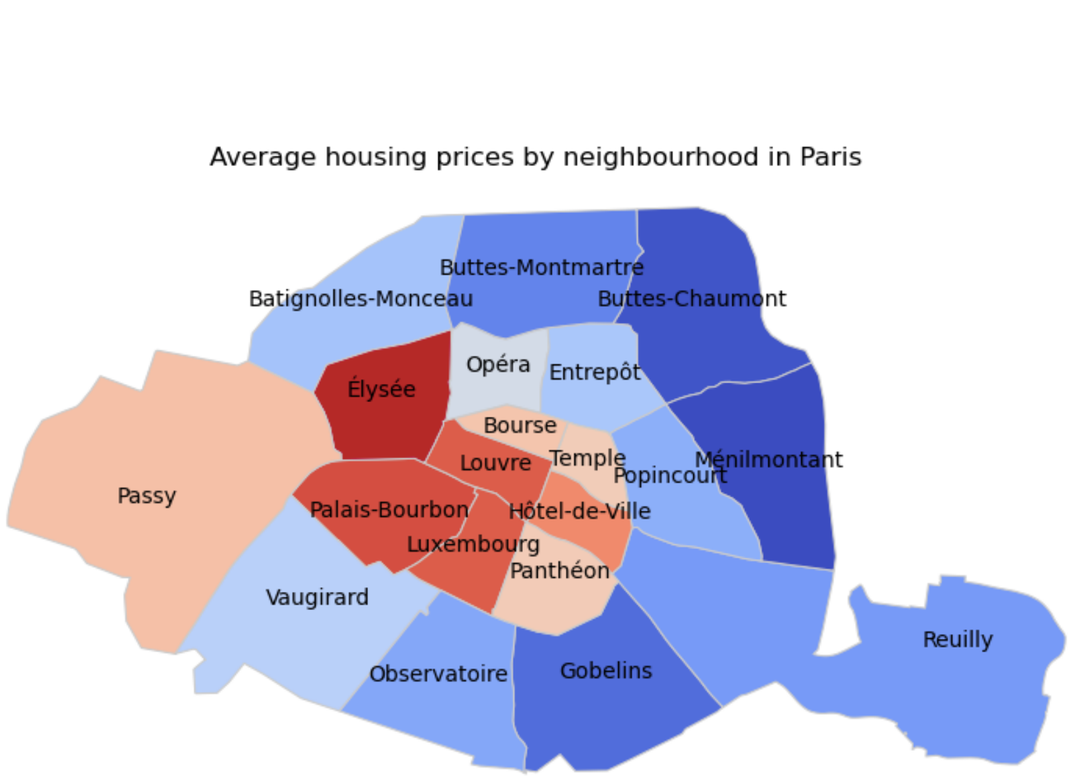
Adjustment of latitude and longitude values: rounded to 7 decimal places for greater accuracy.

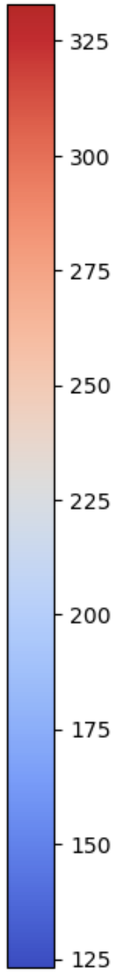
Removal of outliers: price greater than 7000 are considered outliers and are removed.

Filtering based on the "minimum_nights" and "availability_365" conditions:
Lines where "minimum_nights" is greater than 15 are removed and lines where "availability_365" is less than 10.

Changes to the precision of 'review_scores_rating':
The code then rounds 'review_scores_rating' to two decimal places.

Data Visualisation [Advanced]





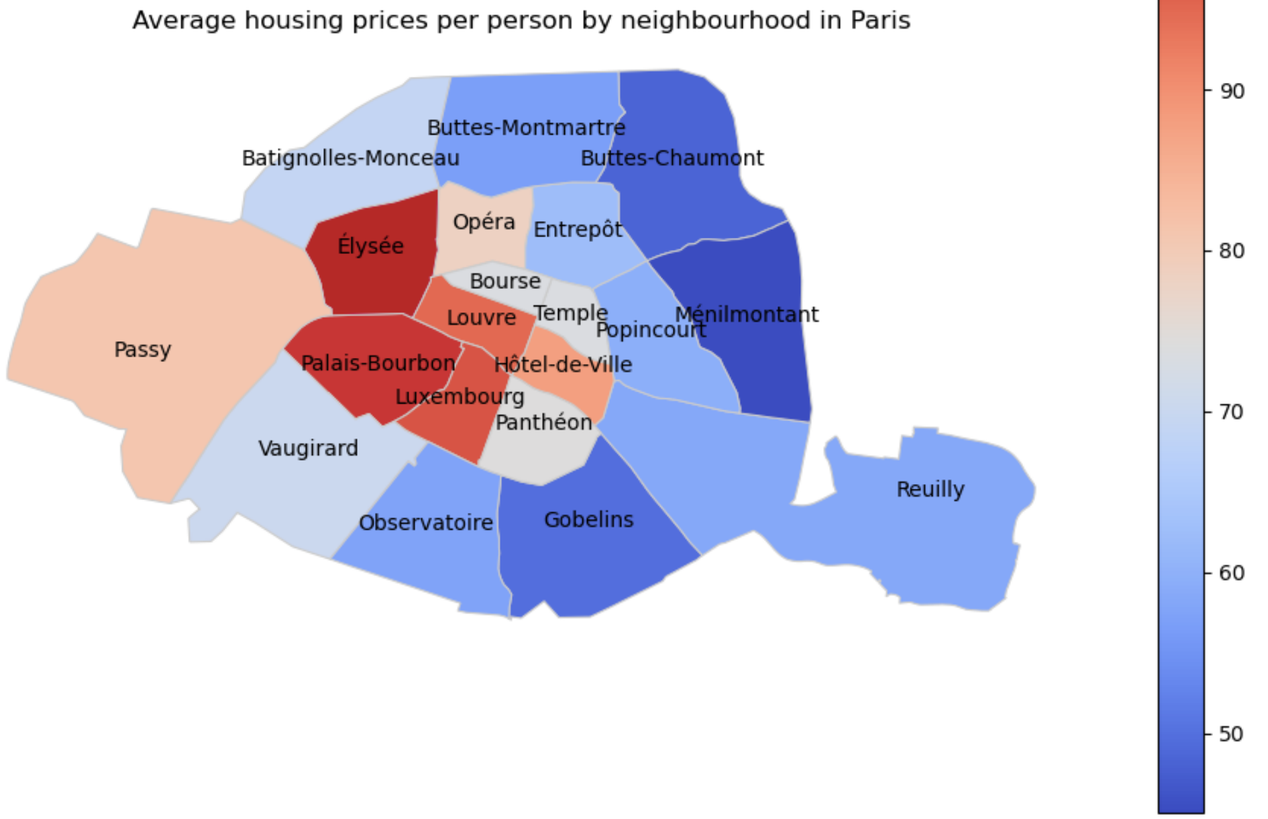
N°1 : Heatmap
Average housing prices by neighbourhood in Paris

- Most expensive : center of Paris (especially Elysée)
- Lowest average price : north-est of Paris

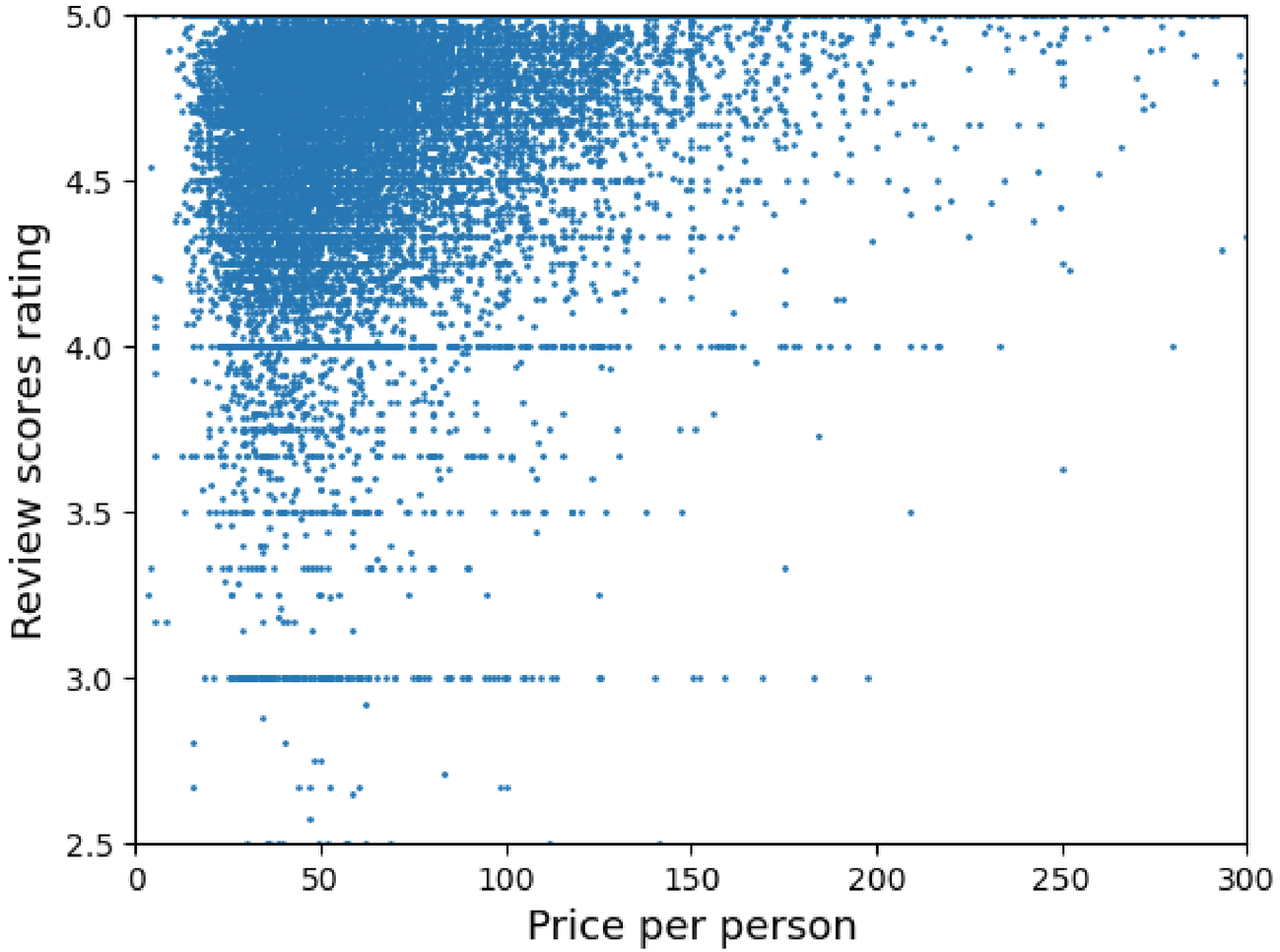
Data Visualisation [Advanced]

N°2 : Heatmap
Average housing prices per person by neighbourhood

- No major difference between prices per person and price by neighbourhood



Data Visualisation [Advanced]



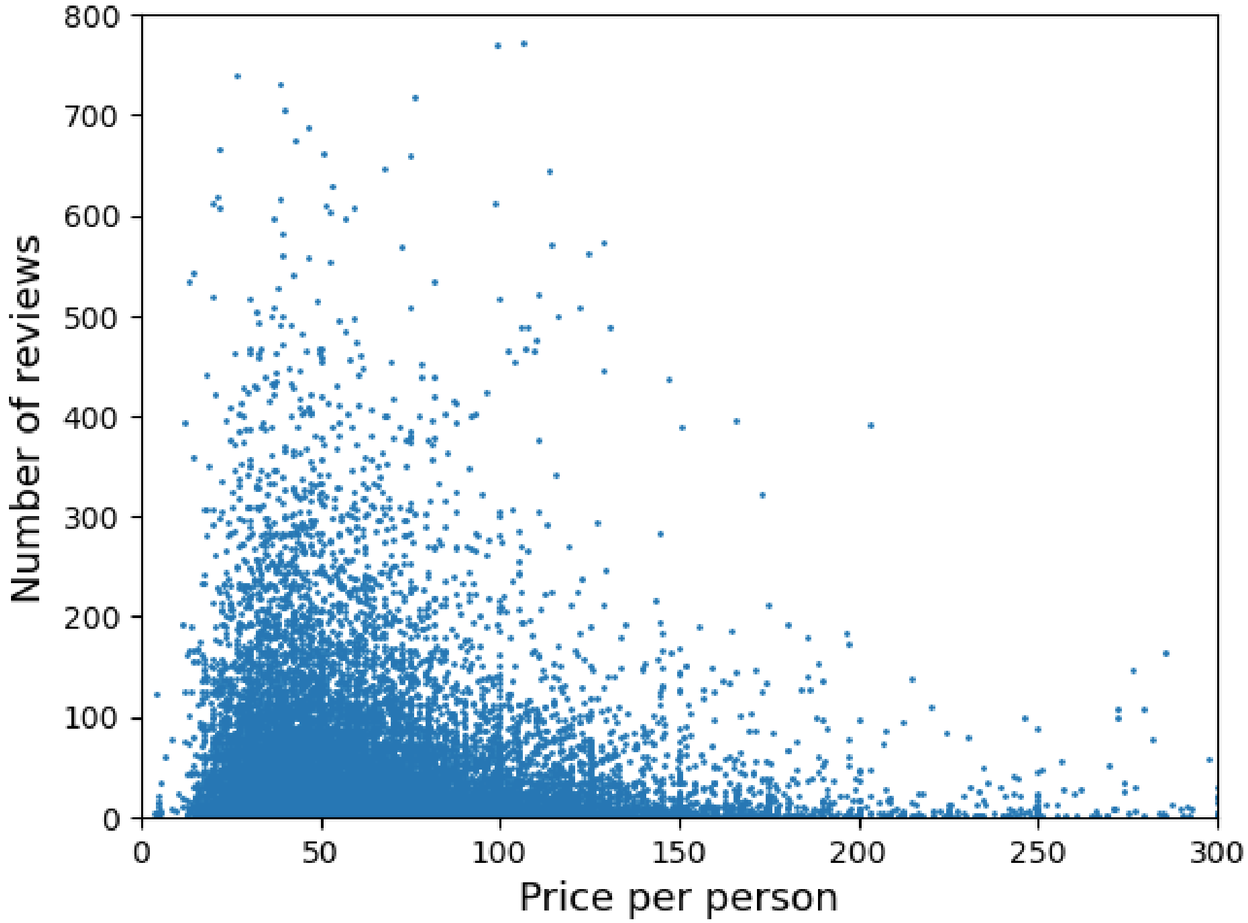
N°3 : Graph
Correlation between "price_per_person" and "review_scores_rating"

- Positive logarithmic correlation between price_per_person and review_scores_rating

Data Visualisation [Advanced]

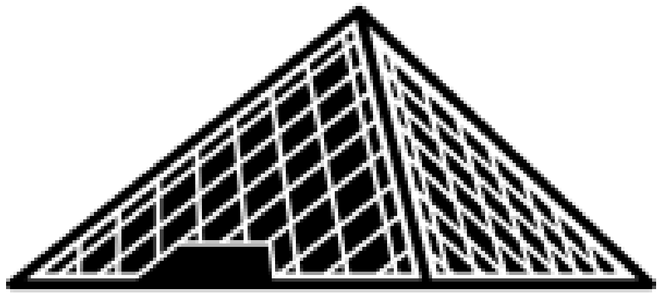
N°3 : Graph
Correlation between "price_per_person" and "number_of_reviews"

- Negative exponential correlation between price_per_person and number_of_reviews



Data Visualisation [Advanced]

Where to rent the cheapest AirBnB accommodation closest to the Louvre ?



- As close as possible to the museum and the least expensive : **Bourse**.
- A bit further and a bit more expensive : **Panthéon** OR **Temple.**

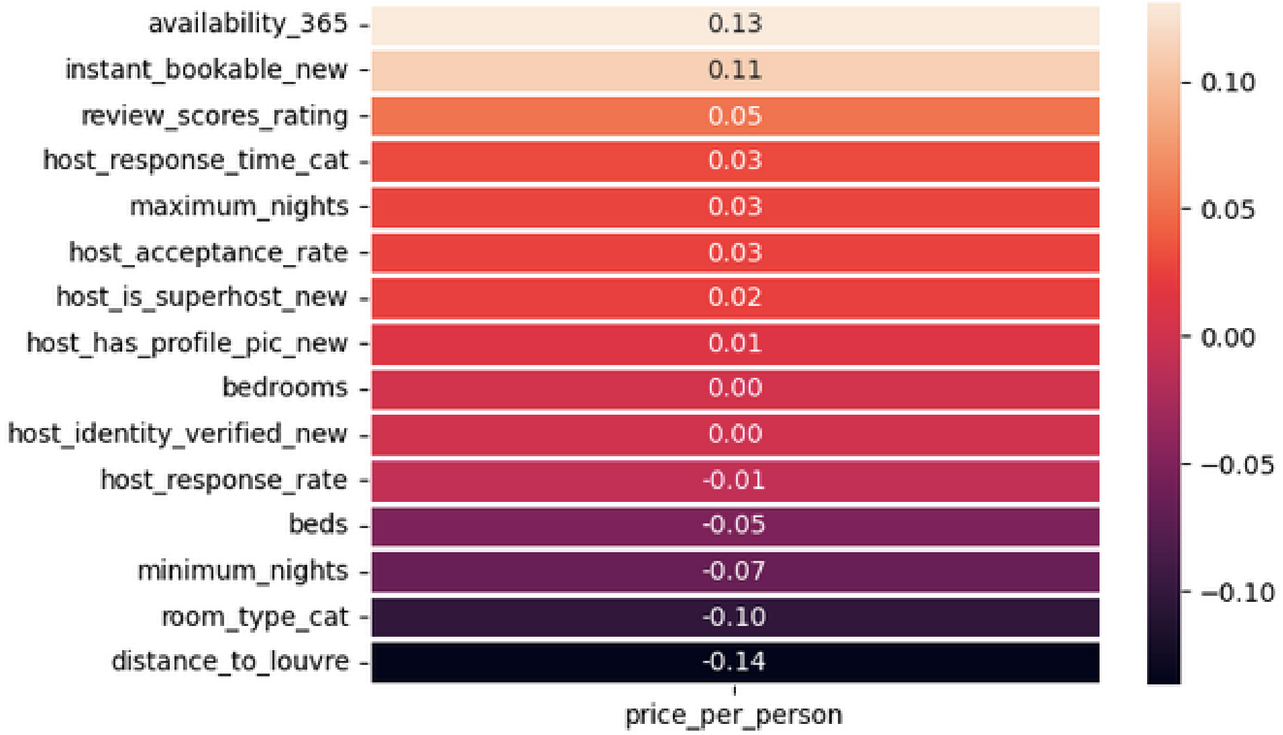
Data Modelling [Standard]

Model to predict price in relation to the distance to the Louvre

RMSE:0.017824216738514 28

Data Modelling [Standard]

The matrix to choose the best variables to be used for the model



- The higher the price, the closer the accomodation is to the Louvre.
- Accomodations that are available for more days in a year might be more expensive.
- The other variables have a weaker correlation with price (they do not significantly influence the price of AirBnB accommodation).

Data Modelling [Standard]

3 models to predict price in relation to other variables

MODEL 1: all main variables considered and "price_per_person"

MSE: 0.0002944855102515150 7
RMSE: 0.0171605801257275 4
R2: 0.1057484383266413 8

MODEL 2: 4 main correlations to "price_per_person": 'instant_bookable_new', 'distance_to_louvre', 'room_type_cat', 'availability_365'

MSE: 0.0003049670806069847 5
RMSE: 0.0174633066916602 24
R2: 0.0739195016459792 7

MODEL 3: only the 2 strongest correlations to 'price_per_person': distance_to_louvre', 'availability_365'

MSE: 0.0003101977772071217 2
RMSE: 0.0176124323155893 82
R2: 0.0580356844540015 35

Model 1 is the most accurate model.

Conclusion

& thank you !