

Price suggestion for Airbnb properties

Final Presentation



Ordinary Linear Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

- y is the response variable = Airbnb price
- x_{1-k} are the predictor variables = features that predict the price
- β_{1-k} are the slope coefficients = change in the mean of y for a one-unit change in x

Spatial Regression

- Accounts for geographical context by incorporating spatial aspects into the regression
 - **through exogenous variables (x)**
 - Spatial Heterogeneity
 - **Spatial Dependence**

Airbnbs in Paris

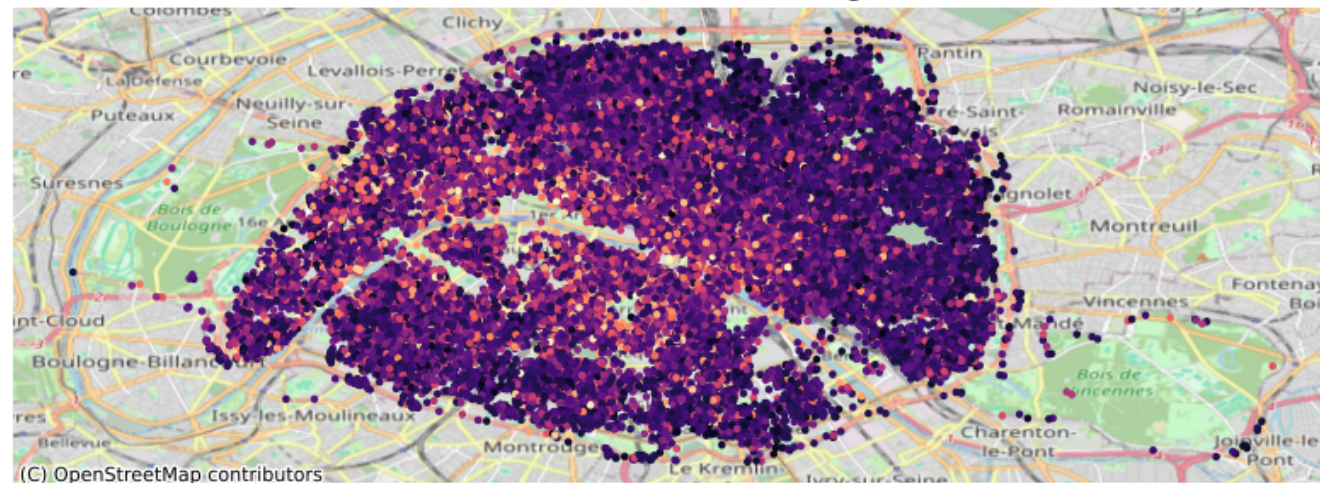
path to higher distribution:
red to yellow to blue to purple

Distribution of Airbnb Listings across Paris



path to higher distribution:
dark purple to light colours

Distribution of the Price of Airbnb Listings across Paris



Baseline Regression Model



Baseline Regression Model is a function that can be invoked with different input data



Helps to continuously introduce new spatial variables into the regression framework



K-fold Cross Validation (k=5) splits data into Training and Validation Sets



Regression is assessed by different Evaluation Metrics on the Validation Set

1. **RMSE**
2. **MAE**
3. **Adjusted R-Squared**

Performing Ordinary Linear Regression

Non-Spatial (without Neighbourhoods)

- accomodates
- bedrooms
- beds
- minimum nights
- number of reviews
- review scores rating
- review scores location
- calculated host listing count
- property type
- room type

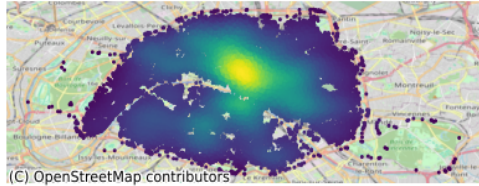
	Non Spatial
RMSE	73.327495
MAE	53.764690
Adjusted R2	0.361417

With Neighbourhoods

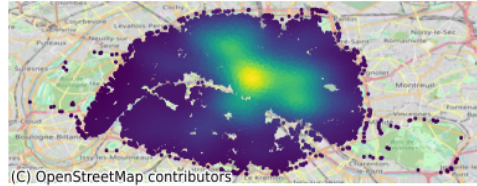
- neighbourhood cleansed
- accomodates
- bedrooms
- beds
- minimum nights
- number of reviews
- review scores rating
- review scores location
- calculated host listing count
- property type
- room type

	Non Spatial	With Neighbourhoods
RMSE	73.327495	69.359654
MAE	53.764690	50.056972
R2 Adjusted	0.361417	0.427211

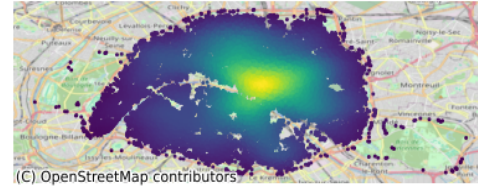
Density of kde_restaurants [km]



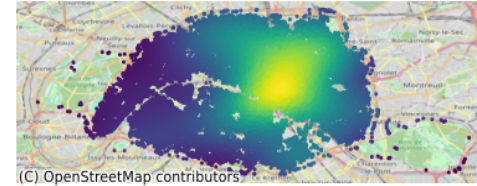
Density of kde_bars [km]



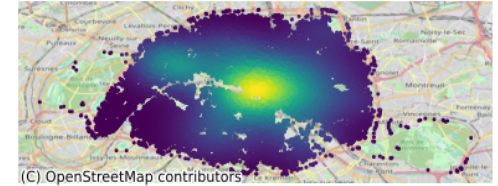
Density of kde_cafe [km]



Density of kde_art [km]



Density of kde_nightclub [km]



Feature 1 : Distance to Amenties

- Restaurants
- Cafés
- Bars
- Art Centers
- Nightclubs

With Distances to Amenities	
RMSE	69.010156
MAE	49.722377
R2 Adjusted	0.432604

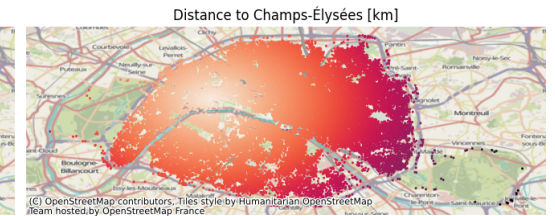
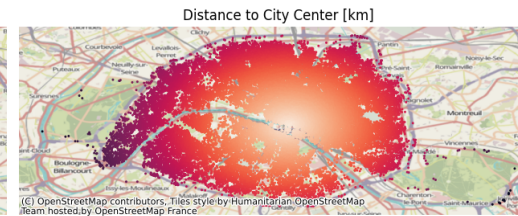
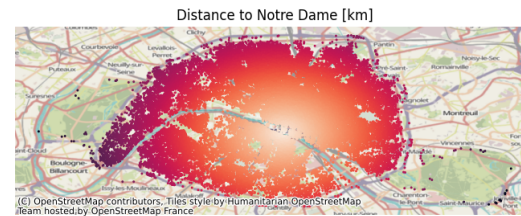
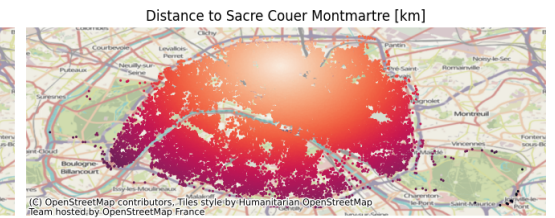
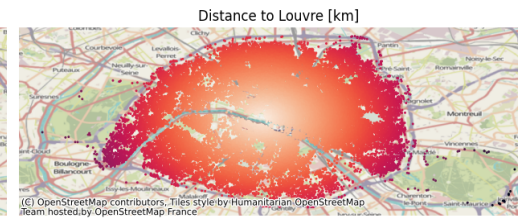
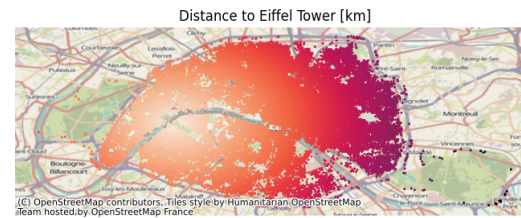
Feature 2 : Kernel Density Estimation of Amenties

- Restaurants
- Cafés
- Bars
- Art Centers
- Nightclubs

With KDE of Amenities	
RMSE	68.643718
MAE	49.310271
R2 Adjusted	0.438250

Feature 3 : Distance to POIs

- Eiffeltower
- Sacre-Cœur (Montmartre)
- Louvre Museum
- Champs-Élysées
- City Center



With Distances to POIs	
RMSE	68.255393
MAE	49.001096
R2 Adjusted	0.444168

SLX Model (Spatial Dependence)

- Spatial Dependence = the value of an observation is influenced by the values of nearby observations due to their spatial arrangement
- incorporates the Spatial Lag of x-variables
- the Spatial Lag captures how the values of a variable in one area are influenced by the values of the same variable in neighbouring areas

Spatial Lag Features

- review_scores_rating
- review_scores_location
- number_of_reviews
- calculated_host_listings_count
- property_type variables
- room_type variables

With SLX for some x-variables

RMSE	68.165666
MAE	48.874893
R2 Adjusted	0.444775

Main Drivers for the price of a property

Price-increasing Property Features

- Accomodates (35.50)
- Bedrooms (21.56)
- Review scores rating (11.79)
- Room type *Hotel Room* (6.97) and *Entire Home/Apt* (0.35)

Price-increasing Neighbourhoods

- Popincourt (8.95)
- Temple (8.47)
- Entrepôt (8.32)

Price-increasing Amenties

- Density of Nightclubs (18.91) and Cafés (12.61)

Price-increasing POIs

- Notre Dame (-56.82)
- Eiffeltower (-18.14)
- Sacré-Cœur (-10.60)

Price-decreasing Property Features

- Minimum number of nights (-7.05)
- Property Type *Private Room in Rental Unit* (-6.94)
- Room type *Shared Rooms* (-4.70)
- Number of Reviews (-4.36)
- Beds (-3.01)

Price-decreasing Neighbourhoods

- Batignolles-Monceau (-4.54)
- Observatoire (-2.71)
- Vaugirard (-2.50)

Price-decreasing Amenties

- Density of Art Centers (-11.65) and Bars (-7.81)

Price-decreasing POIs

- City Center (49.75)
- Louvre (11.11)
- Champs-Élysées (6.31)

Conclusion and Main Findings

Ordinary Linear Regression vs. Spatial Regression

- Spatial Regression performs better than Ordinary Linear Regression across all metrics

Best Performing Model

- Regression Model with all spatial features together performs the best

Continuous Model Improvement

- every additional spatial feature provided a continuous improvement in the model performance

Influence of POIs

- Distances to POIs improved the RMSE and the Adjusted R2 the most

Main Drivers of the price

- **Distances to POIs** (Notre Dame, Eiffeltower, Sacré-Cœur), **Capacity/Size** of the properties and **Density of Nightclubs and Cafés**

Thank you

Final Presentation

