REAL ESTATE VALUATION USING LINEAR REGRESSION

Sravanthi Tarani
Christophe Schellinck
Philippe Fimmers
Mikael Dominguez

Datacleaning:

- Webscraping Immoweb

Score of our model/prosperity index

Dataset quality

- Improve consistency of data
- Data on Immoweb

Data on Immoweb (e.g.)

- Completeness of data
- Total m^2 ↔ habitable m^2
- 4 rooms ↔ 3 rooms
- Price all units ↔ 1 unit

Linear Model implementation

Model training

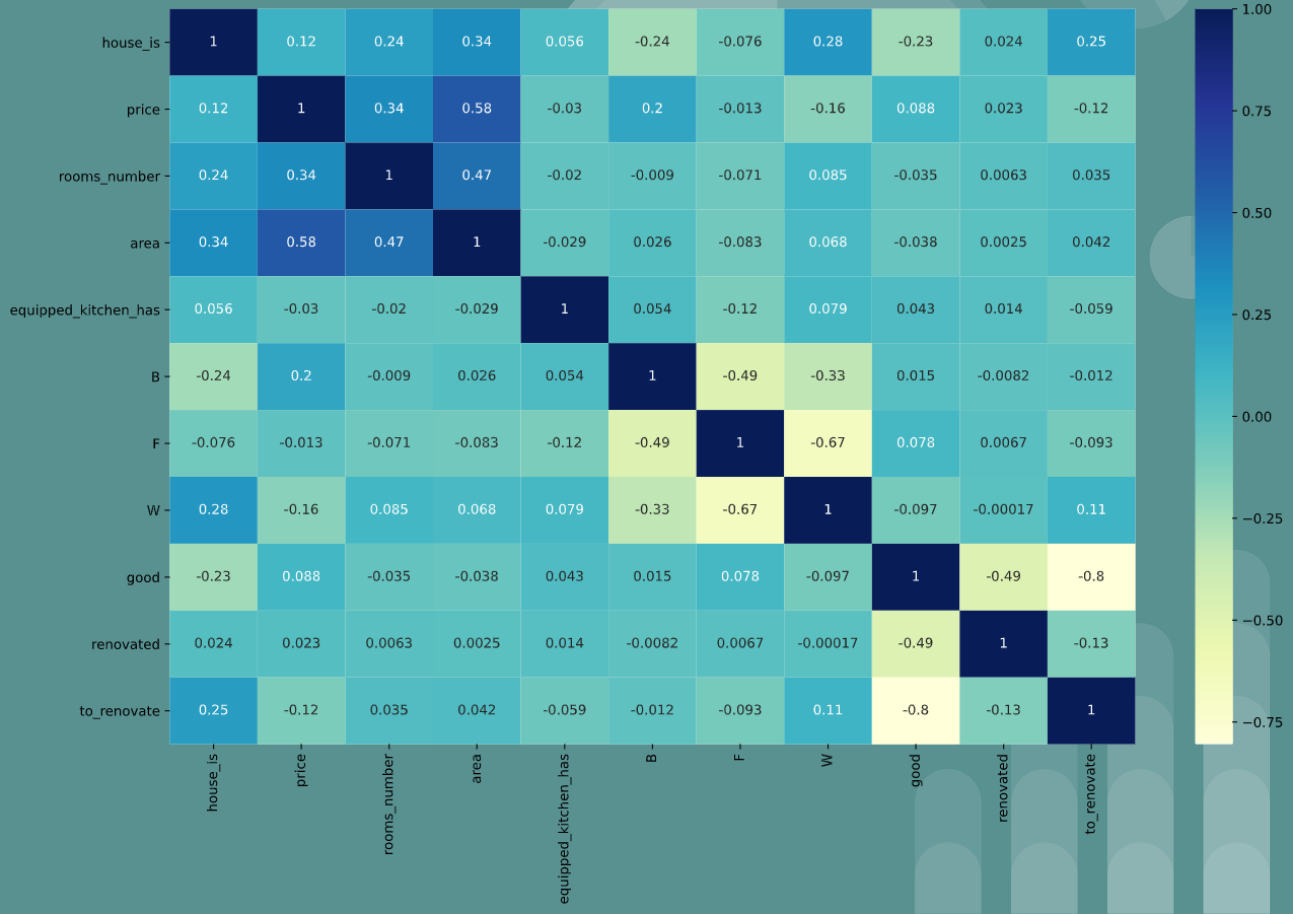
- Volume of data
- Clean data
- Numerical data
- Nomalized data

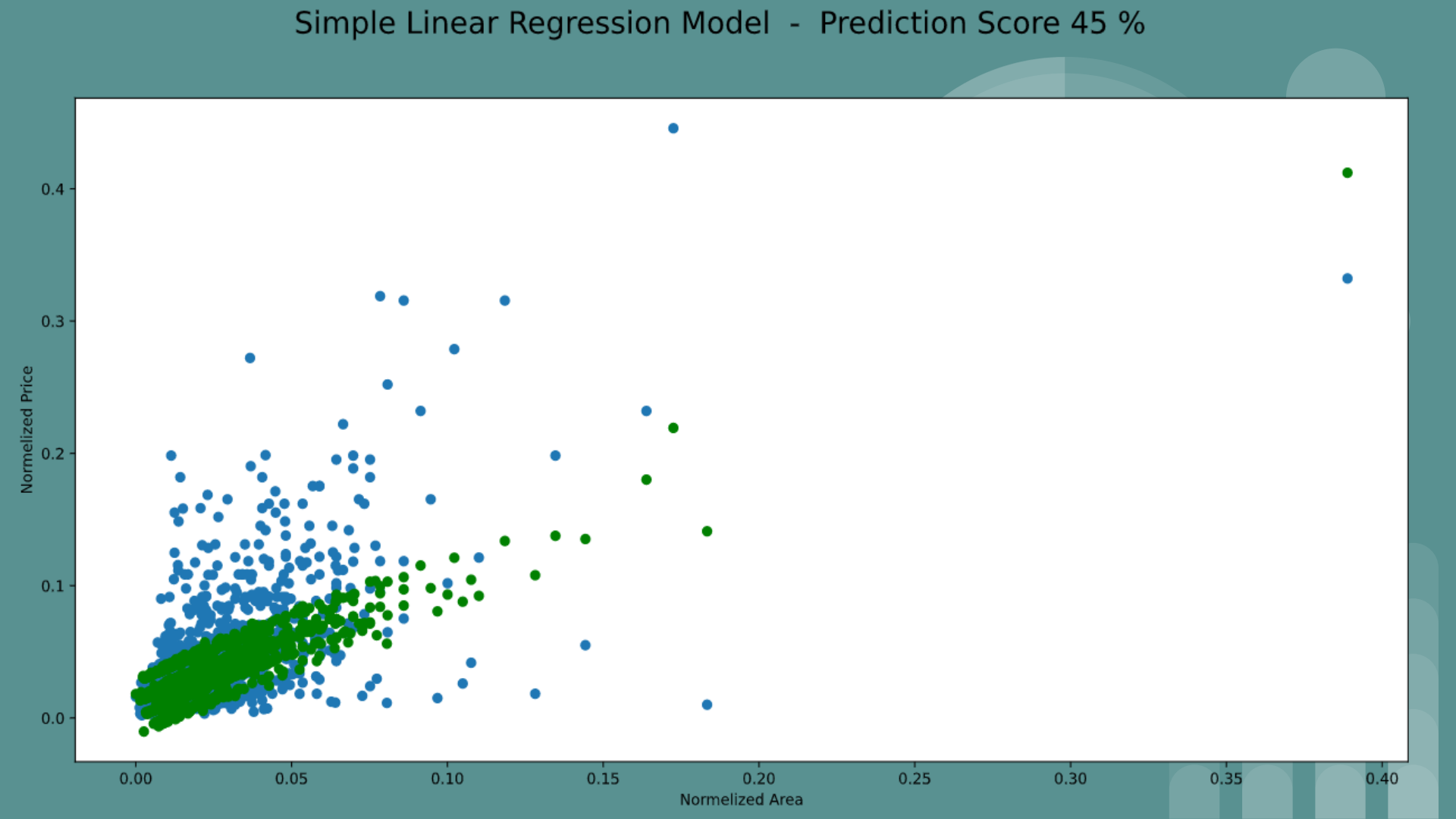
Data selected in the model

Data Set shape (11288, 10)

Target : Price

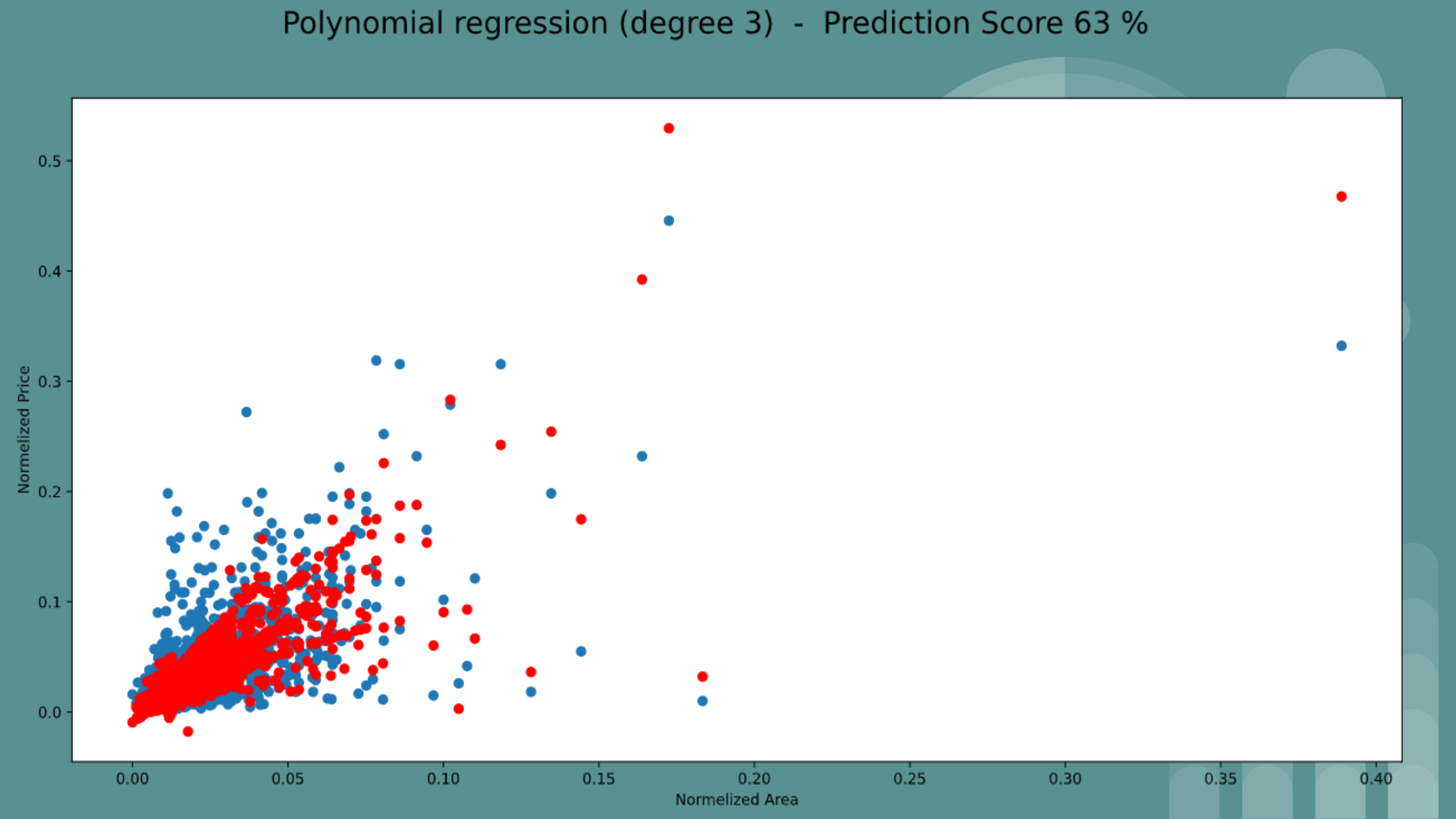
Features : m = 10
- House is
- Rooms numbes
- Area
- Equiped kitchen has
- B (Brussels)
- F (Flanders)
- W (Wallonia)
- Good
- Renovated
- To_renovated




Simple Linear Regression Model - Prediction Score 45 %
Normelized Price
Normelized Area
0.0
0.1
0.2
0.3
0.4
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40

Improve the model

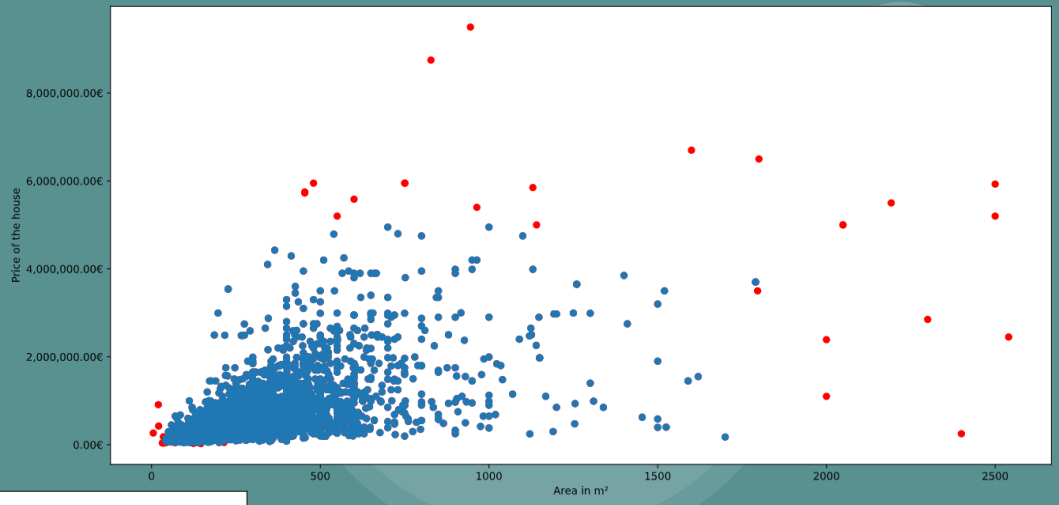
- Making hypotheses
- Analyse again the features
- Change model


Polynomial regression (degree 3) - Prediction Score 63 %
Normelized Price
Normelized Area
0.0
0.1
0.2
0.3
0.4
0.5
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40

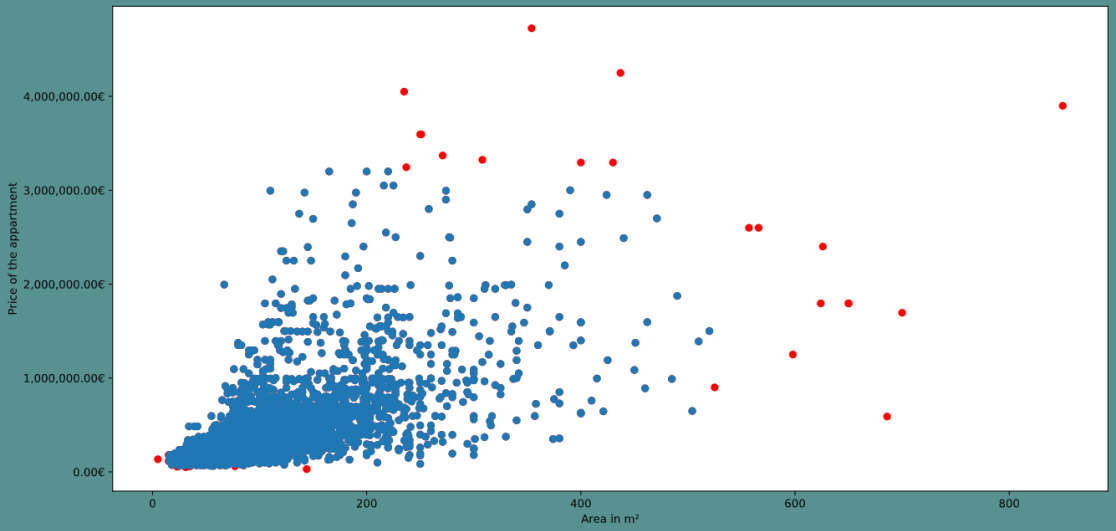
Hypothese:

- Remove outlier

Outlier house



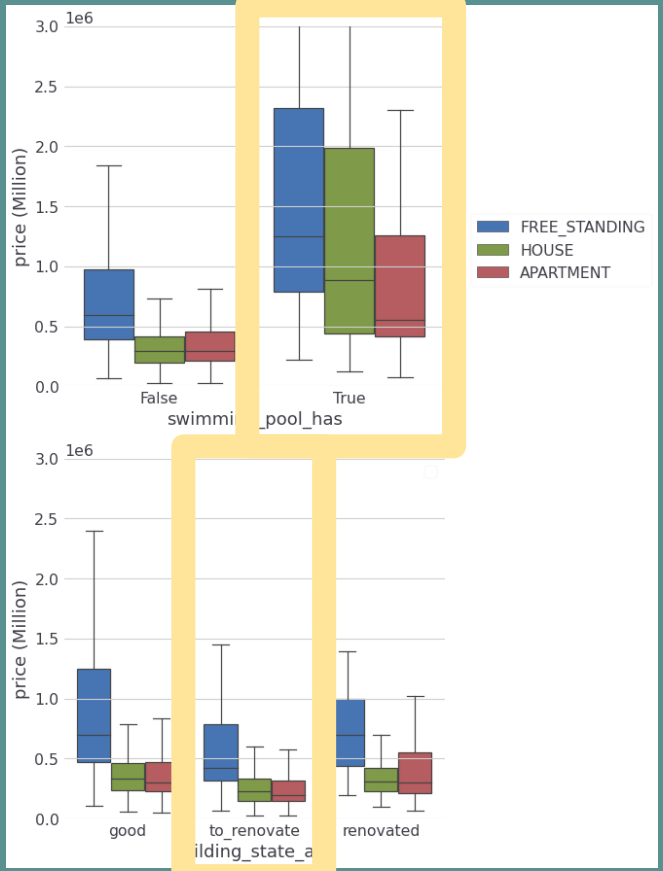
Outlier appartment



- The score loose 5 %

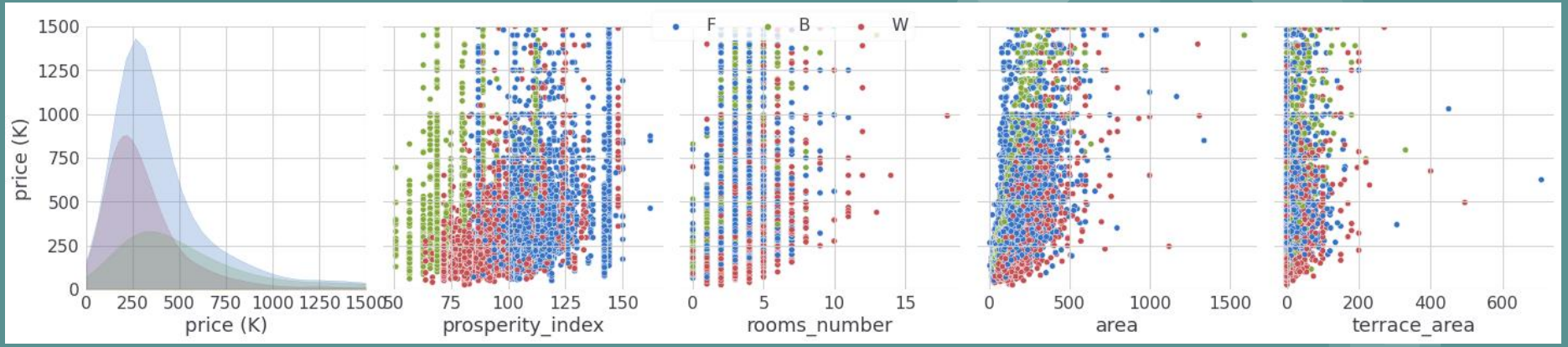
How to reduce model complexity?

Price sensitivity to attributes



- Sensitivity to 3 logical property categories
- Swimming pool and need for renovation influence the price
- No effect found for:
  - Terrace
  - Garden
  - Equipped kitchen
  - Open fire
- A constant for Swimming pool and to_renovate

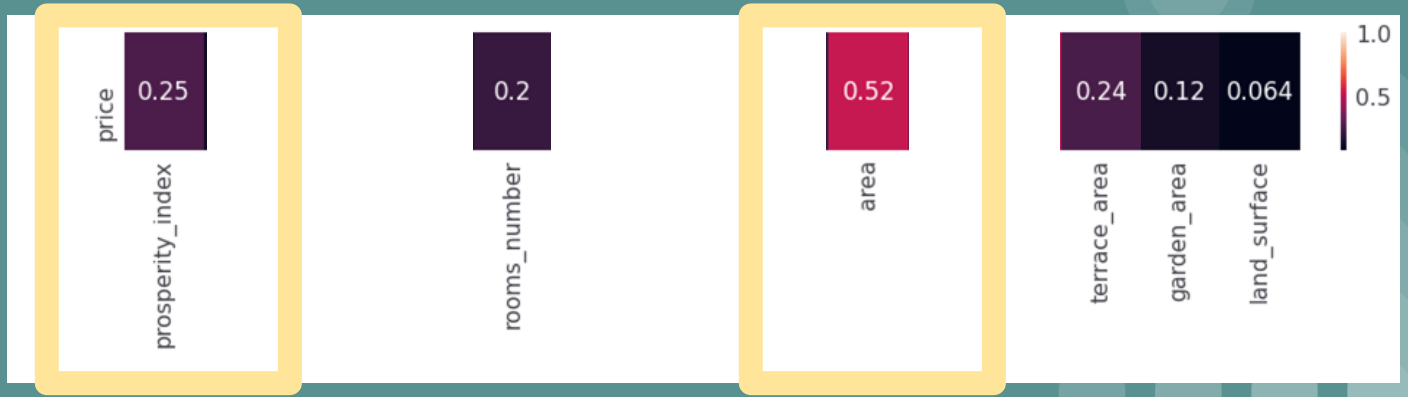
Correlations with price



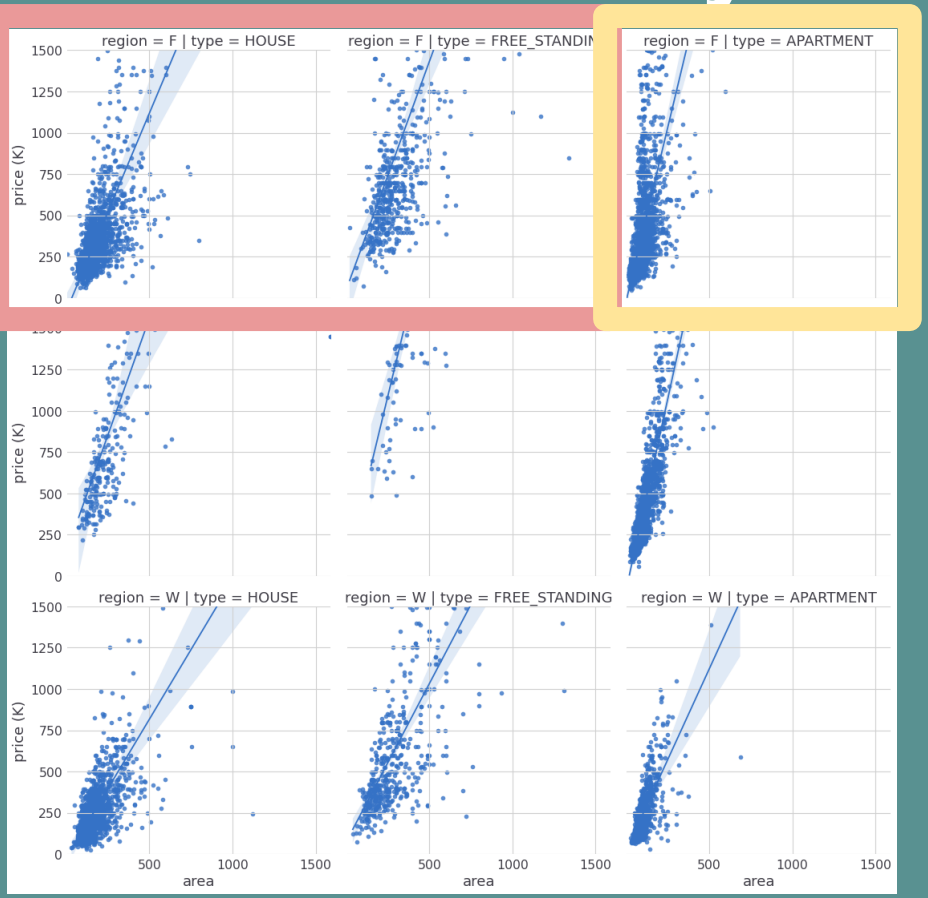
Wallonia

Flanders

Brussels

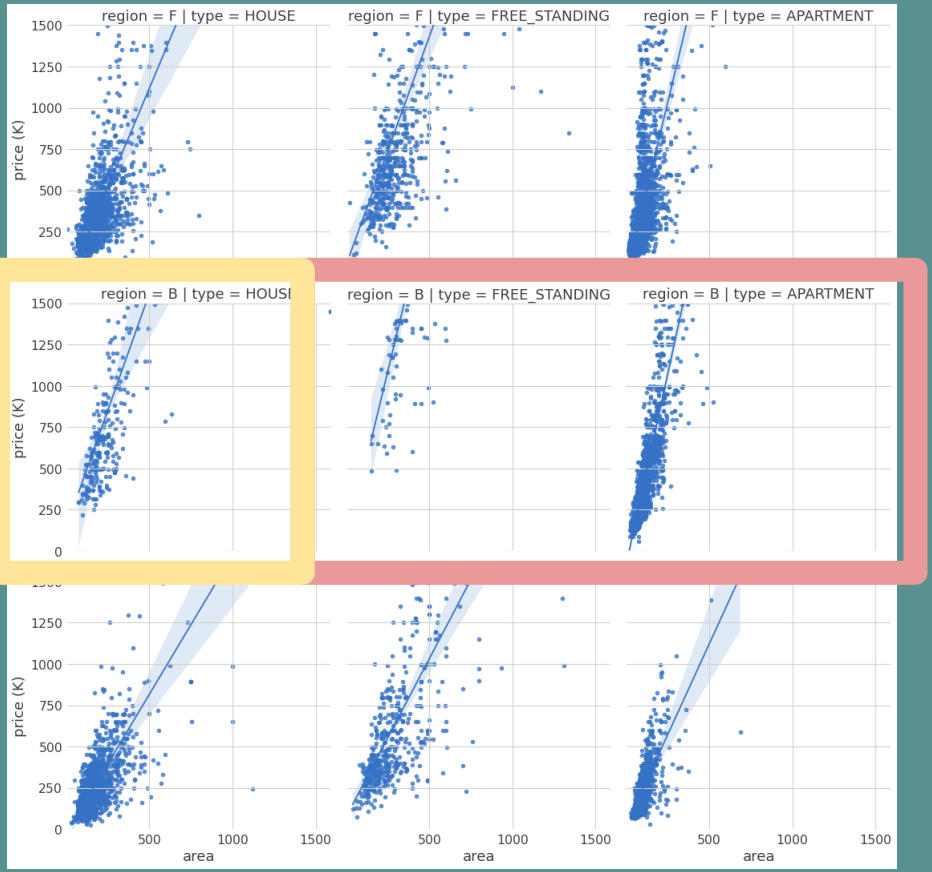


Price sensitivity to area (m^2)



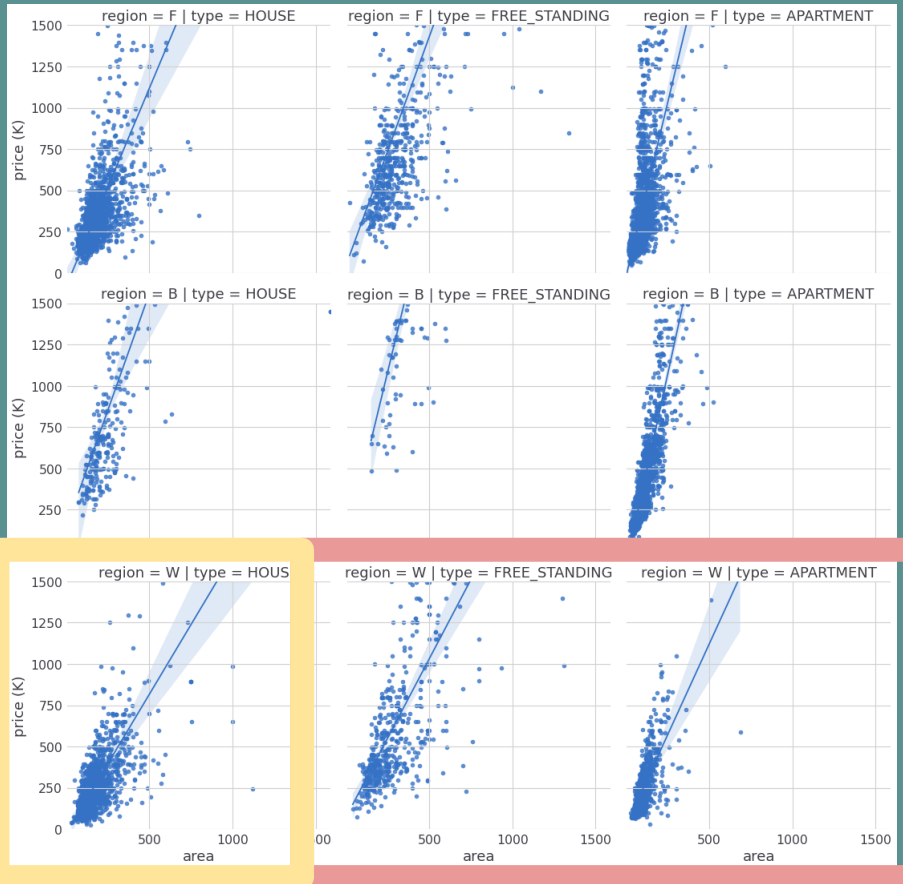
- Lowest in Wallonia
- Highest in Brussels
- Depens on property type
  - 2 factors for Flanders

Price sensitivity to area (m^2)



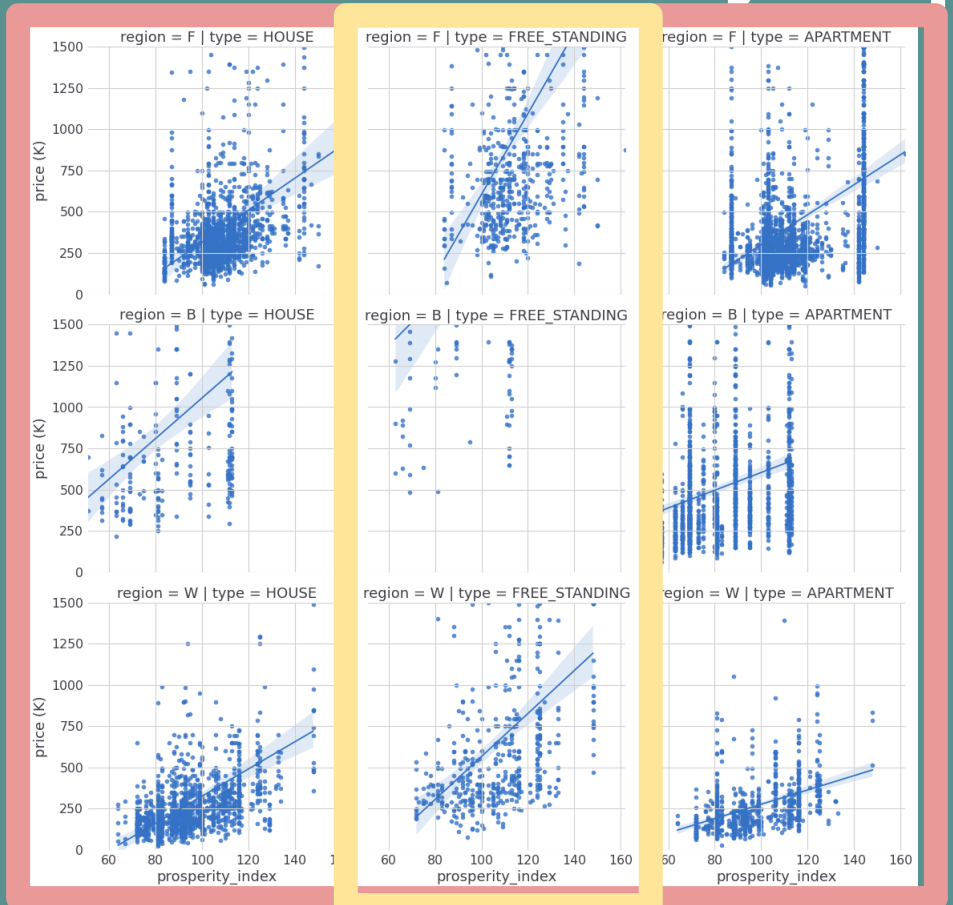
- Lowest in Wallonia
- Highest in Brussels
- Depens on property type
  - 2 factors for Flanders
  - 2 factors for Brussels

Price sensitivity to area (m^2)



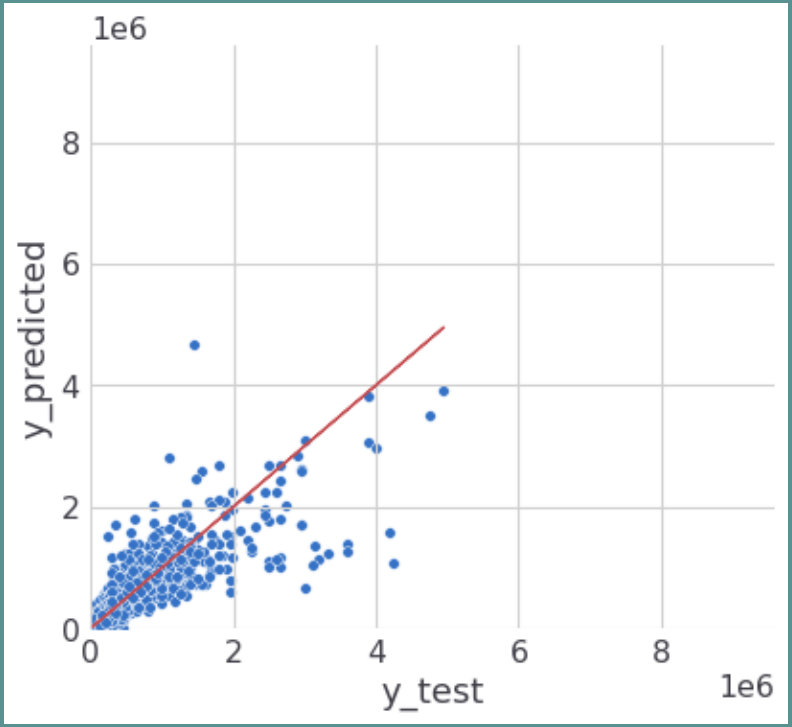
- Lowest in Wallonia
- Highest in Brussels
- Depens on property type
  - 2 factors for Flanders
  - 2 factors for Brussels
  - 2 factors for Wallonia

Price sensitivity to prosperity



- Highest for free standing houses
  - 2 factors
- Highest intercept in Brussels
  - 1 constant per region

Simplified 3rd degree Regression model



- 13 factors + INTERCEPT
  - Found coefficients are very large
- R^2 (test & train): 62%