

BY QBICLE.COM



**IMMO ELIZA**

**Project : BREPP\_AI**

Belgium Real Estate Price Predictor - Powered by AI

**SPRINT 2 Report: Data Visualisation**

**16/02/2024**

# AGENDA

- Context
- The dataset and its first output
- The approach
- Features/Price charting
- Global correlation matrix
- Principal component analysis
- Conclusion & Next Steps

# CONTEXT

OBJECTIVE : Design a feature predictor for the Belgian real estate market based on a machine learning model

4 SPRINTS :

- DATA COLLECTION AND PREPROCESSING
- DATA VISUALIZATION
- DEVELOP THE MACHINE LEARNING MODEL
- DEPLOYMENT THROUGH API

# A dataset reworked to be quantitative

- 25K unique properties
  - 44% houses
  - 56% apartments
- 28 actionable parameters
  - of which 24 are now quantitative (cluster to boolean)
- Rather high NaN values (25% average)

# 6 categories for 28 parameters

# Quantitative

# Qualitative

```
graph TD; Root[Geoloc, General, Energy, Access, Interior, Exterior] --> Price[Price]; Root --> Loc[locality, province, region]; Root --> Gen[type, subtype, year, new, condition, frontage, tenement, floodSafety]; Root --> Eng[energyClass, heating, doubleGlazing]; Root --> Acc[elevator, disabled]; Root --> Int[livingSurface, furnished, bedrooms, bathrooms, kitchen, openFire]; Root --> Ext[plotSurface, garden, terrace, parking, pool]
```

The diagram illustrates the classification of real estate features. It starts with six main categories: Geoloc, General, Energy, Access, Interior, and Exterior. Each category is represented by an orange rounded rectangle. Below each category, there is a list of blue rounded rectangles representing sub-features or specific characteristics.

- Geoloc:** Price, locality, province, region
- General:** type, subtype, year, new, condition, frontage, tenement, flood safety
- Energy:** energy class, heating, double glazing
- Access:** elevator, disabled
- Interior:** living surface, furnished, bedrooms, bathrooms, kitchen, open fire
- Exterior:** plot surface, garden, terrace, parking, pool

# First outputs : real estate market facts

Property  
Features →

- Price Mean/Median: 475/340 KEUR
- Age Mean/Median : 1976 /1978
- 19% considered as new (2021+)
- 23% of the goods to be renovated
- 33% in good energy class (B or better)
  - 95% have double glazing
- 5% in a flooding zone
- 28% accessible (disabled people)
- 6% tenement property

# First outputs : real estate market facts

- 50% parking
- 40% terrace
- 20% garden
- 2.5% swimming pool

← Exteriors  
Features

Interiors  
Features →

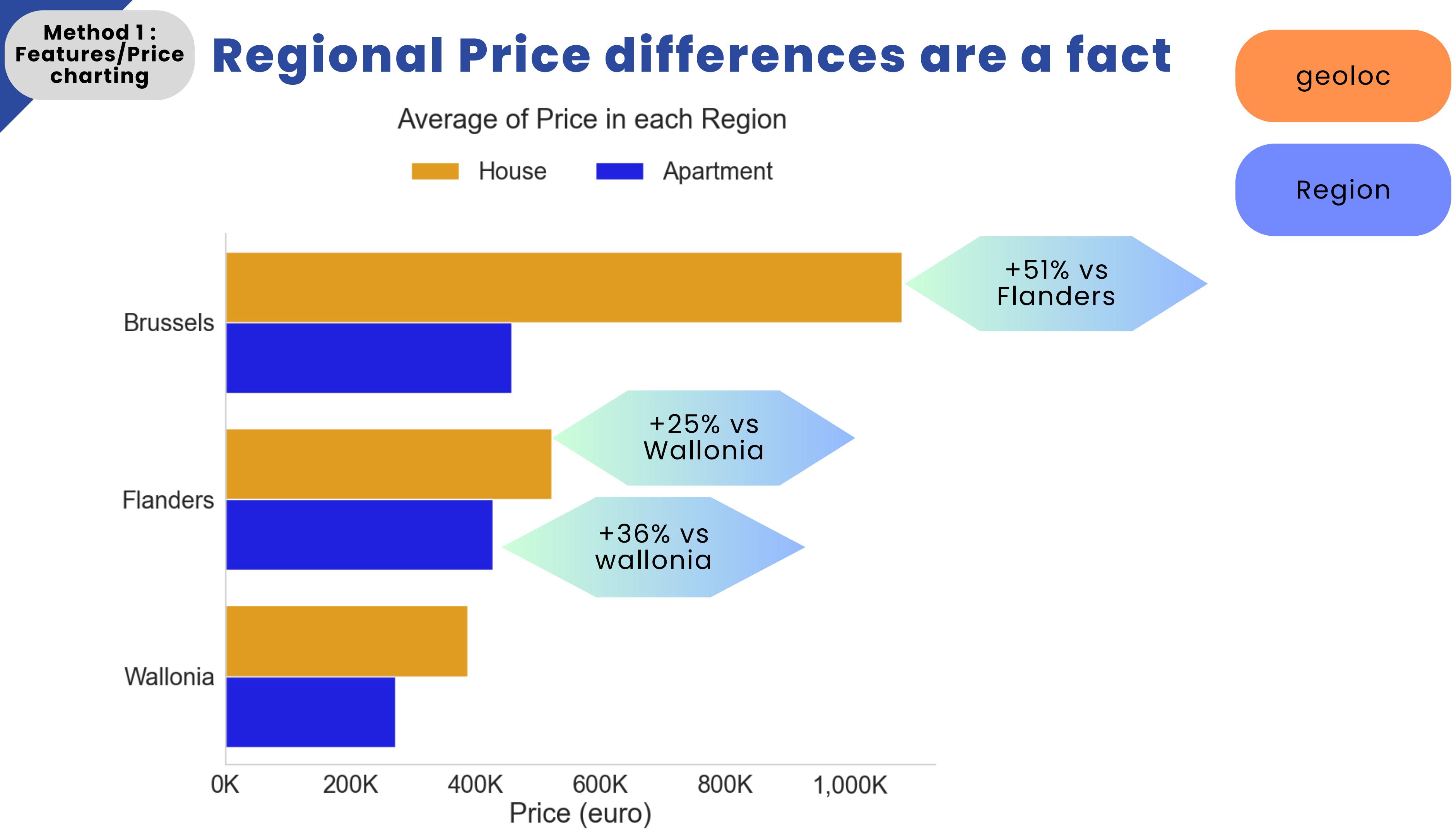
- Average living surface 170 m<sup>2</sup>
- Avg bedroom/bathroom: 3.0/1.4
- 95% of equipped kitchen
- 3% of open fire
- 2.5% furnished

# The triple down approach

**28 Features**

**3 correlation methods used in synergy**





Method 1:  
Features/Price  
charting

# Driven by provincial disparities...

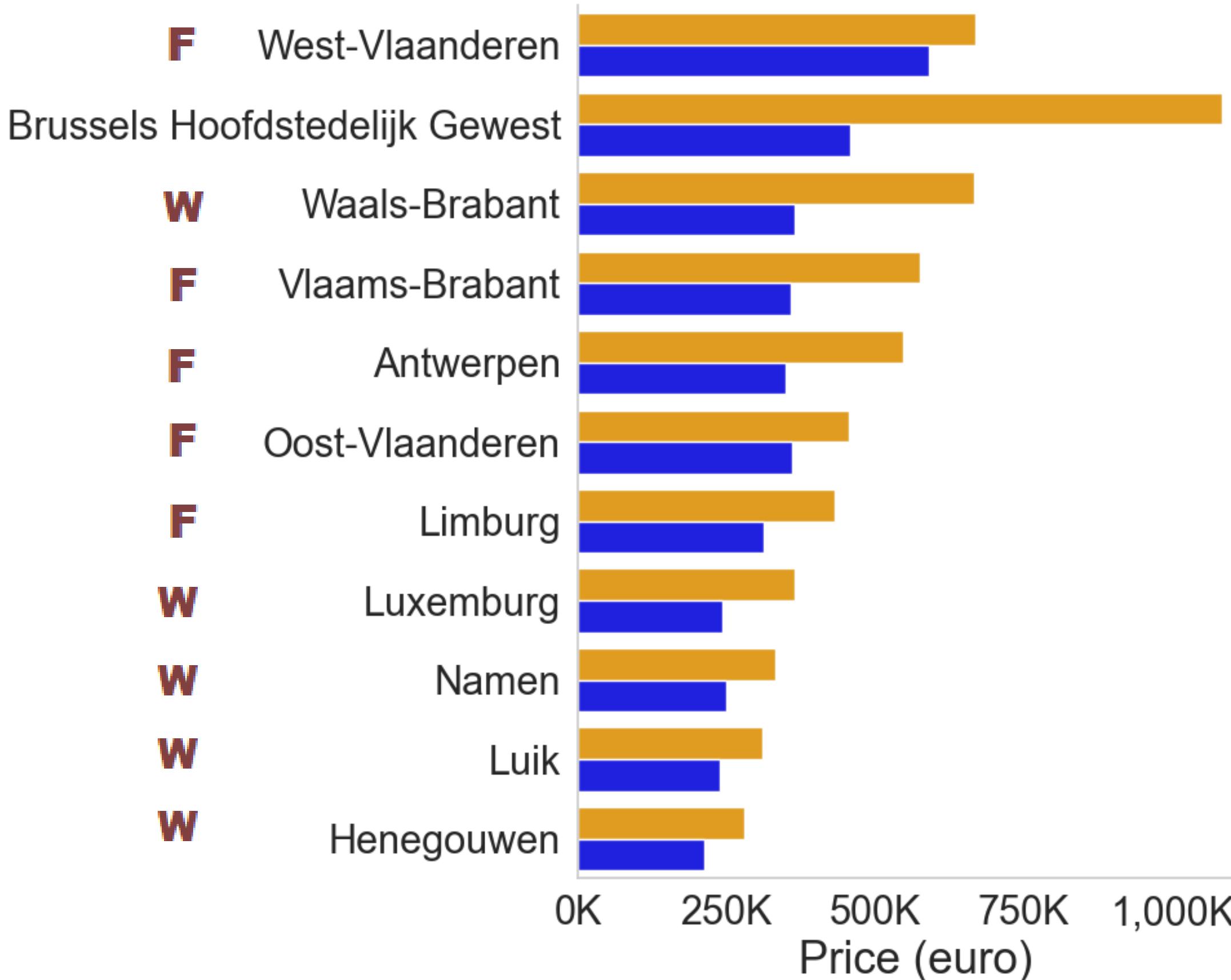
Average of Price in each Provinces

House Apartment

geoloc

Province

**W** for Wallonia  
**F** for Flanders



Method 1:  
Features/Price  
charting

# ... and confirmed by local disparities as well : the big 10X stretch

geoloc

Locality

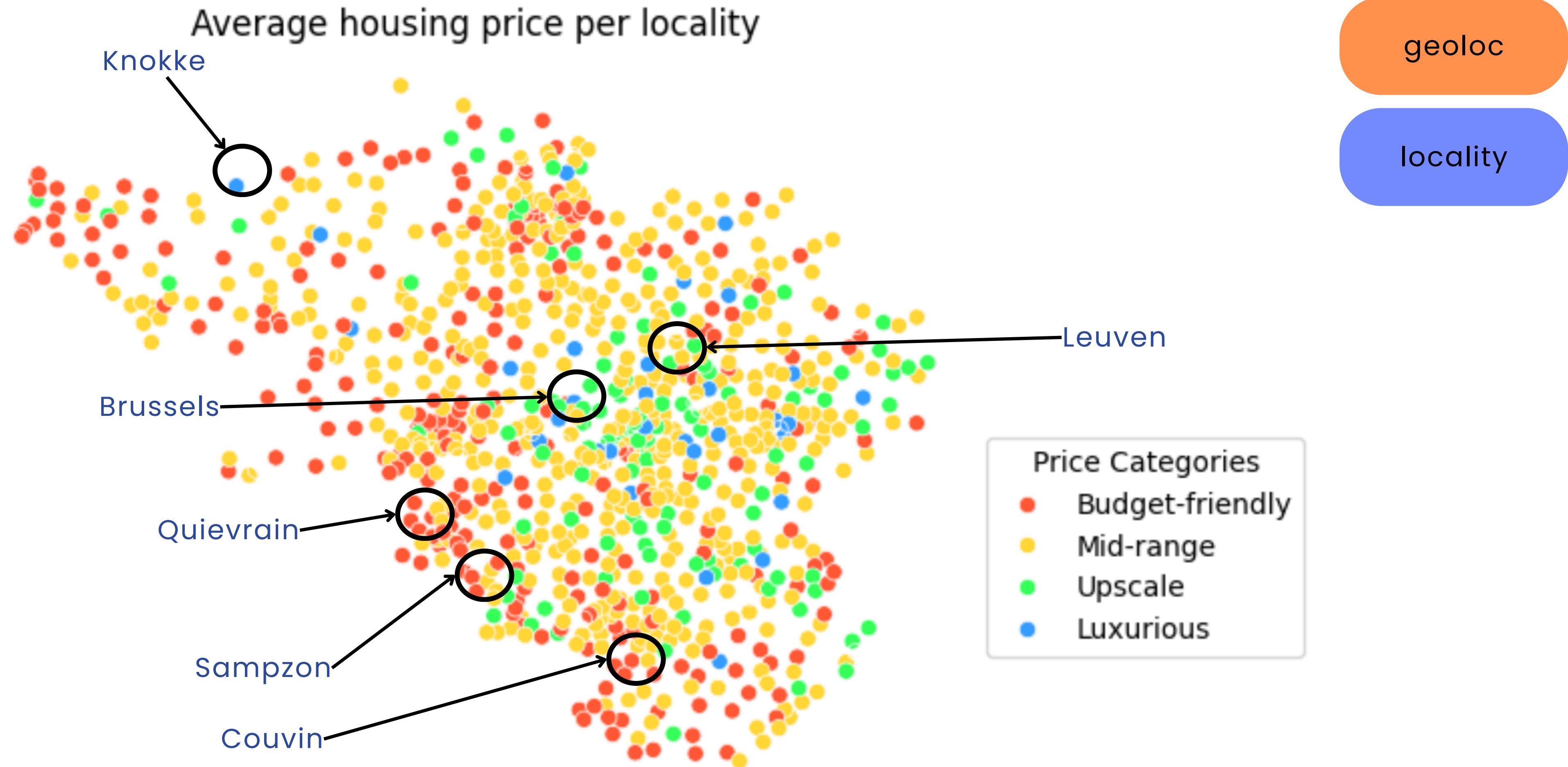
Top 3 Cheapest localities			
Postal code	Locality	province	Price (per m2)
5660	Couvin	Namen	1,008
7120	Sampzon	Henegouwen	1,018
7380	Quievrain	Henegouwen	1,046

Top 3 Most Expensive localities			
Postal code	Locality	province	Price (per m2)
8300	Knocke-Heist	West-Vlaanderen	11,393
8301	Knocke-Heyst	West-Vlaanderen	7,713
3000	Leuven	Vlaams-Brabant	6,718

\* Only Localities with 10 or more listings were selected for this analysis

Method 1:  
Features/Price  
charting

# All this reflects in locality heatmap BUT...



Method 1:  
Features/Price  
charting

# ... but localities impact price even on narrower geography as well

geoloc

Locality

Top 3 Most Cheapest localities in Brussels		
Postal code	Locality	Price (per m2)
1081	Koekelberg	2,628
1080	Molenbeek-Saint-Jean	2,618
1130	Haren	2,520

Top 3 Most Expensive localities in Brussels		
Postal code	Locality	Price (per m2)
1050	Ixelles	4,748
1180	Uccle	4,531
1000	Bruxelles	4,406

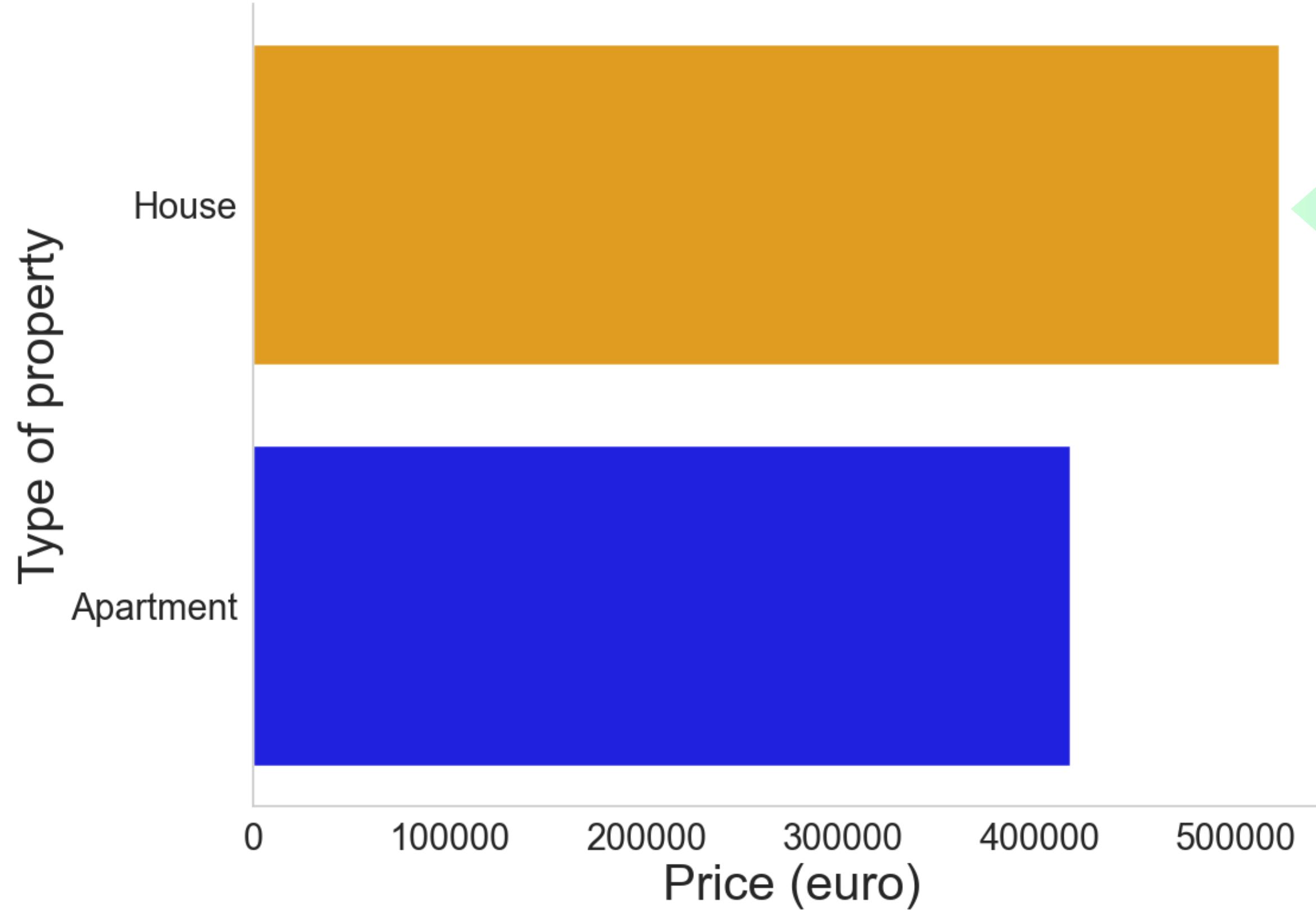
Method 1:  
Features/Price  
charting

# Property Type: A serious impact

general

Type

Property Type Vs Price



Method 1:  
Features/Price  
charting

# Features : The 9 silver bullets ?

general

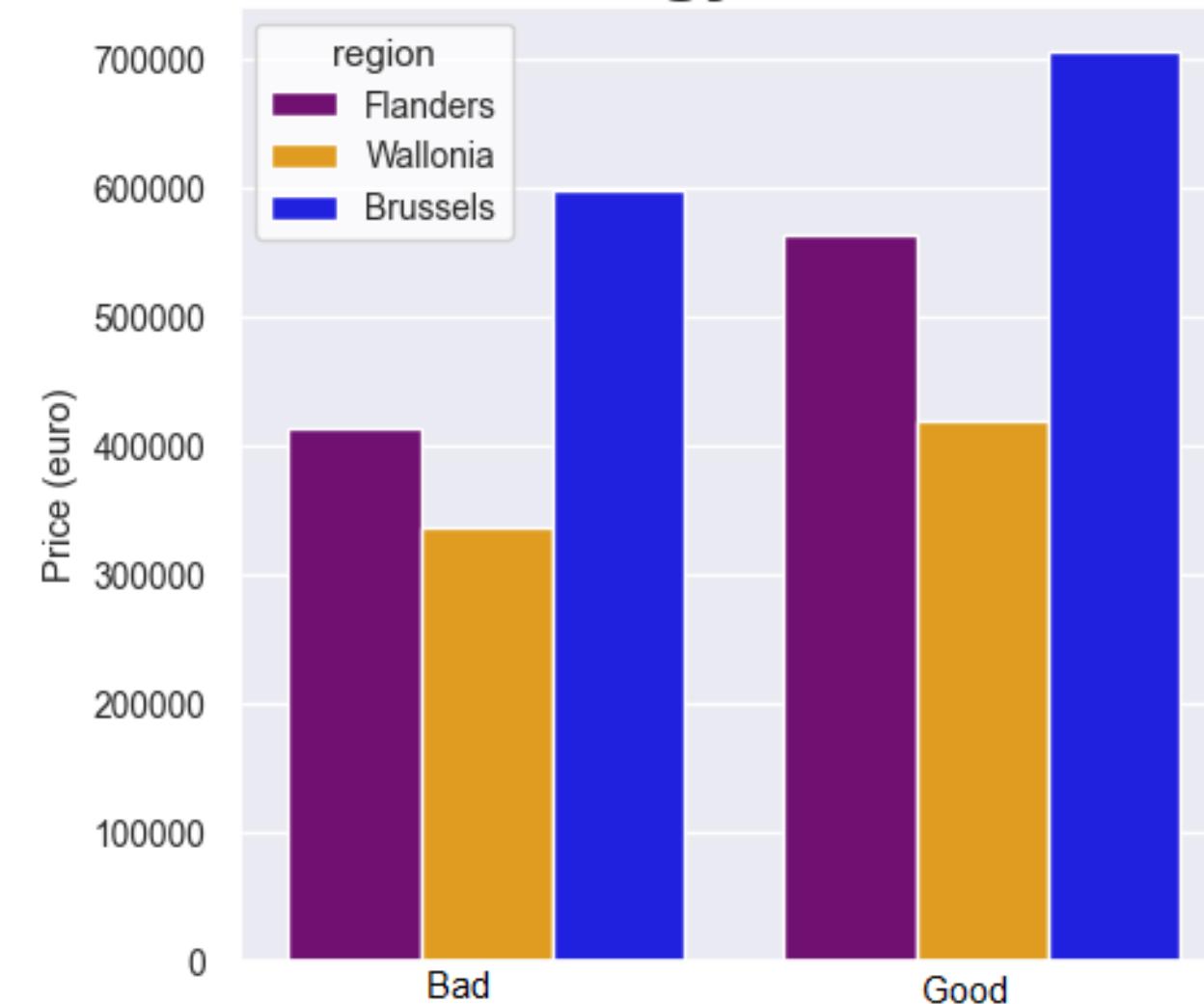
energy class

Condition

Tenement

## Features impact on Price

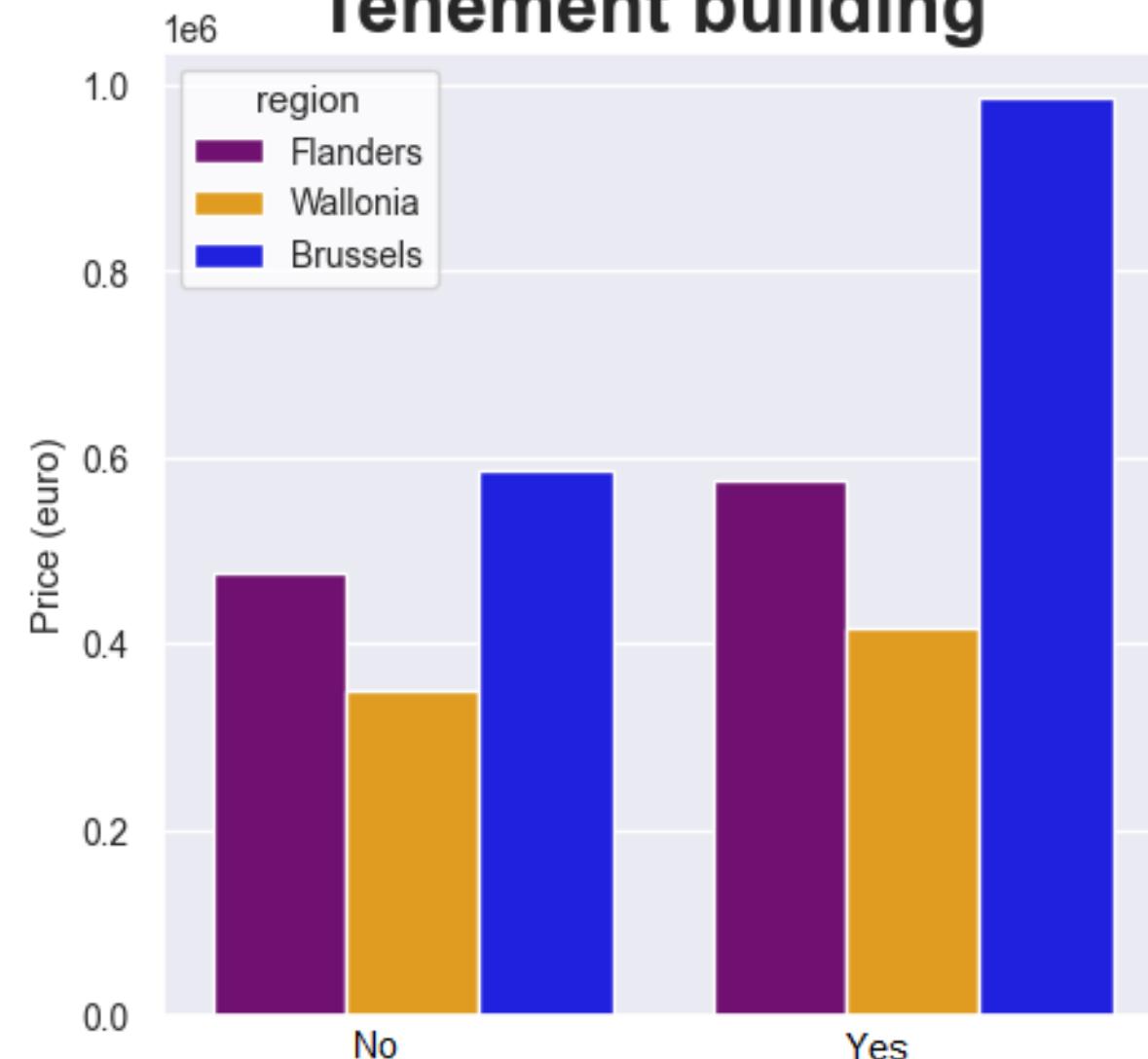
Energy class



Building condition



Tenement building



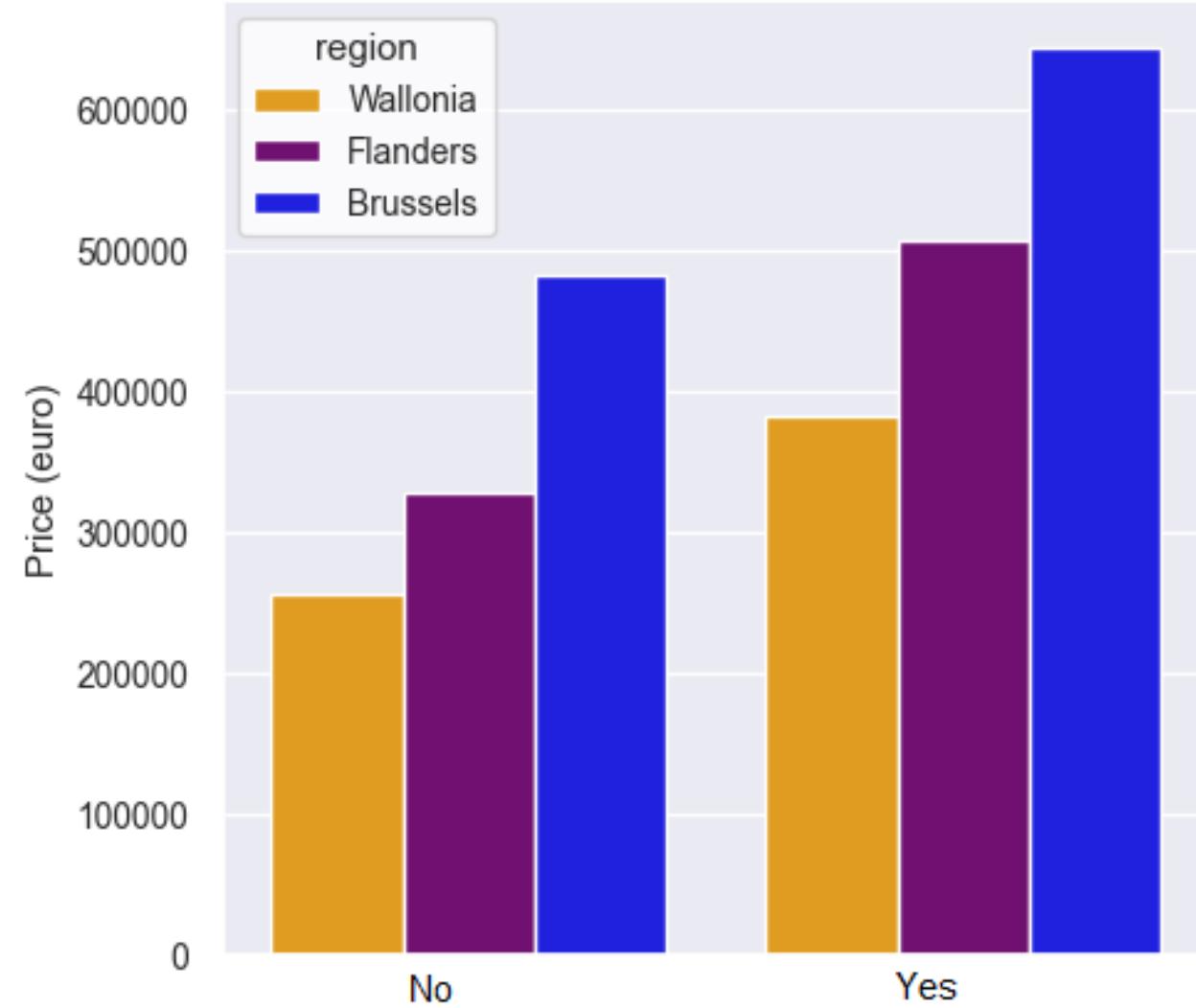
**Method 1:  
Features/Price  
charting**

# Features : The 9 silver bullets ?

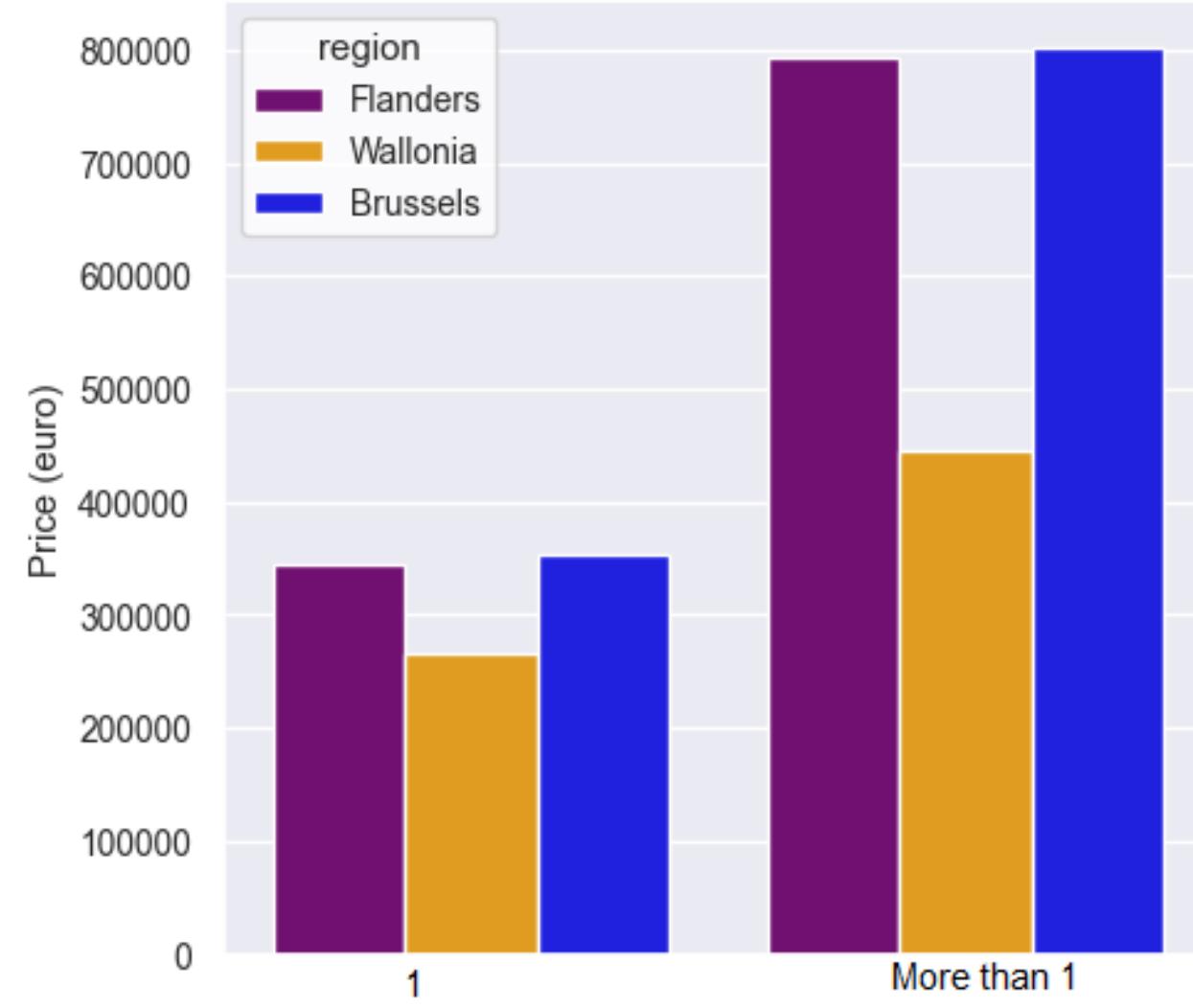
- furnished      Bedrooms
- Interior      Bathrooms
- Surface
- Open fire
- kitchen

## Features impact on Price

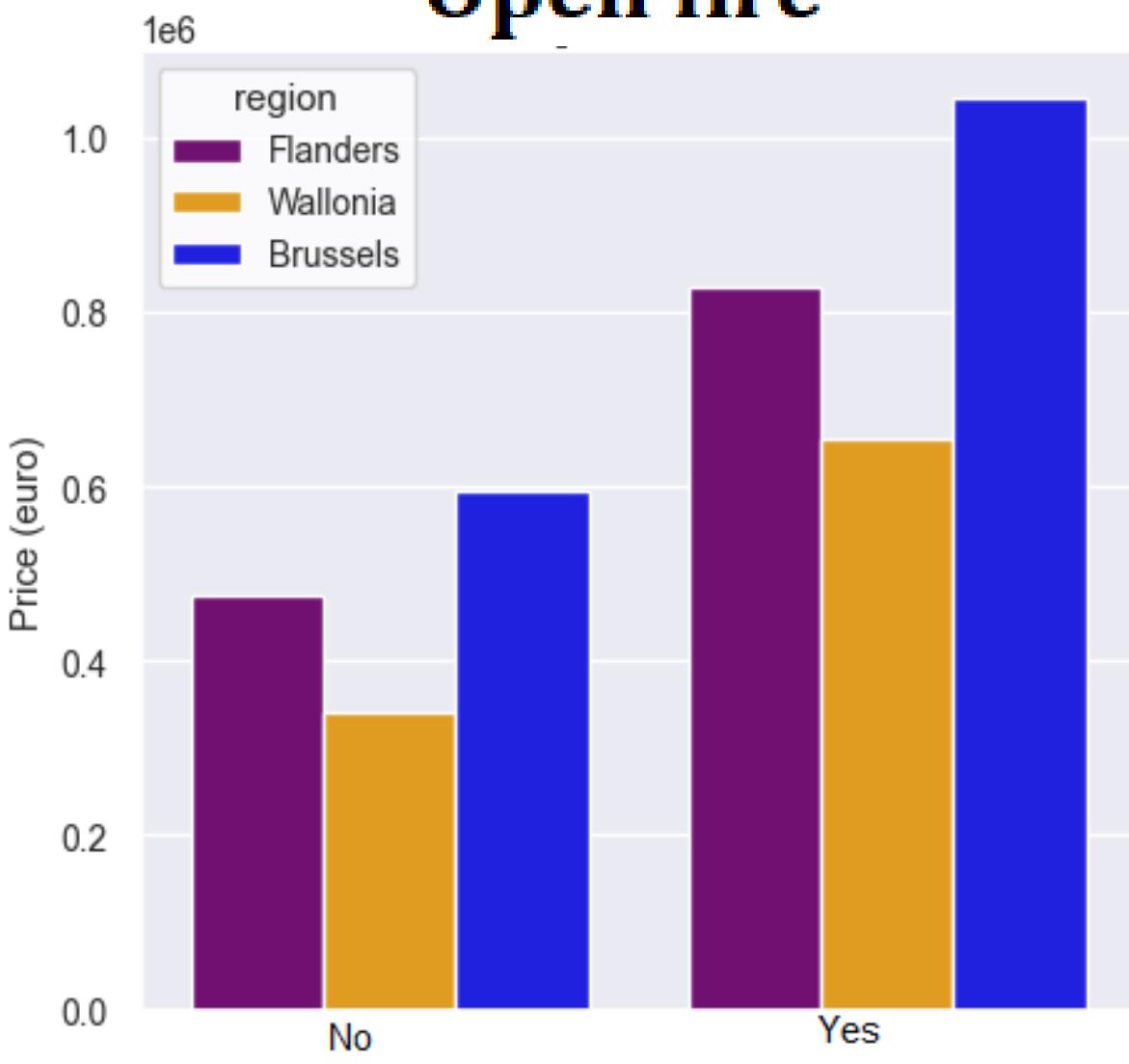
**Kitchen equipped**



**Bathrooms total nb**



**Open fire**

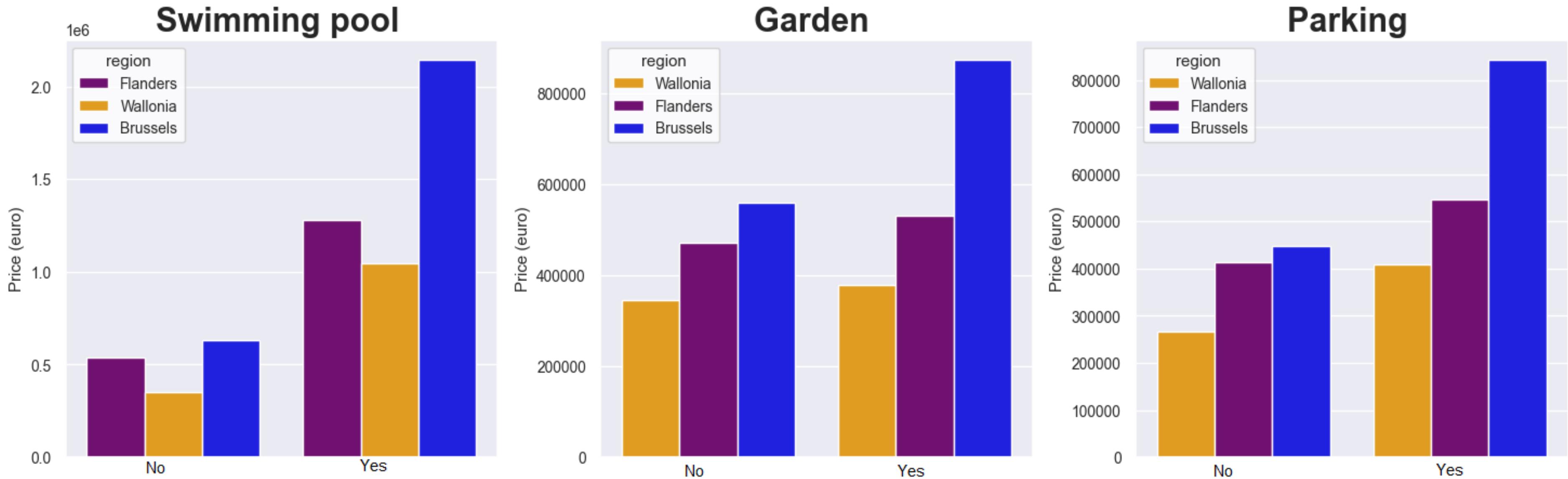


Method 1:  
Features/Price  
charting

# Features : The 9 silver bullets ?

- Exterior
- Plot surf.
- parking
- Surf. tot
- garden
- pool

## e Features impact on Price



## **Method 2 : Features correlation matrix**

# 8 bullets confirmed by correlation matrix

- Living surface, bedrooms & bathroom nb
  - Swimming pool
  - Terrace and garden surface
  - Num of frontages > house vs appment
  - Parking
  - Energy class
  - Building conditions
  - Open fire

## Non conclusive :

- Kitchen equipped,
  - Tenement building

# Correlating variables

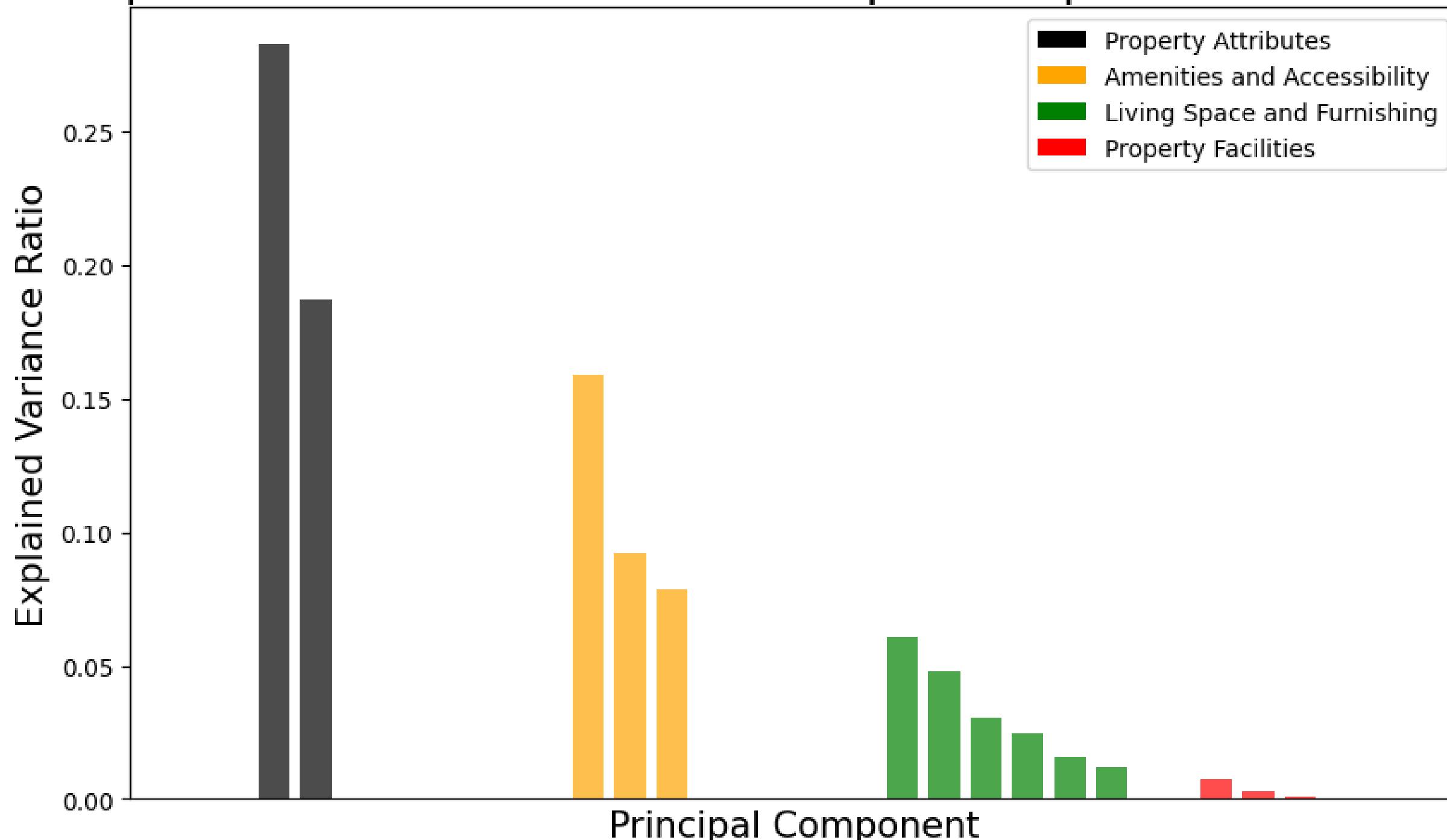
# Highest correlation to Price

	Price (euros)	Construction year	New Construction boolean	Building condition boolean	Swimming pool boolean	Parking boolean	Cooker boolean	Price data boolean	Gated boolean	Terrace surface (m²)	Homes total id boolean	Plot surface (m²)	Number of terraces	Open fire	Kitchen equipped boolean	Living surface (sqm)	Draught glazing boolean	Accessible for disabled people boolean	Furnished boolean	No. of bedrooms	Price (euros)	Construction year	New Construction boolean	Building condition boolean	Swimming pool boolean	Parking boolean	Cooker boolean	Price data boolean	Gated boolean	Terrace surface (m²)	Homes total id boolean	Plot surface (m²)	Number of terraces	Open fire	Kitchen equipped boolean	Living surface (sqm)	Draught glazing boolean	Accessible for disabled people boolean	Furnished boolean	No. of bedrooms
Construction year	0.011	1	0.014	0.033	0.012	0.032	0.014	-0.016	0.0068	0.22	0.0068	0.28	0.0015	0.33	0.14	0.17	0.12	0.074	0.49	0.0564	0.058	0.037	0.38																	
New Construction boolean	-0.013	2.54	1	0.27	-0.087	-0.077	0.23	0.0068	-0.053	-0.063	-0.04	0.21	-0.02	0.05	-0.053	-0.057	4.7e-05	-0.13	0.005	0.21	-0.075	0.13																		
Building condition boolean	0.12	0.30	0.22	1	0.007	0.021	0.16	0.016	0.024	0.052	0.025	0.13	0.097	0.02	0.005	-0.052	0.29	0.043	0.07	0.021	0.11	0.09																		
Swimming pool boolean	0.32	0.012	0.087	0.067	1	0.18	0.052	0.0068	0.1	0.062	0.25	0.058	0.059	0.22	0.26	0.13	0.039	0.21	0.031	0.0673	0.0546	0.2																		
Parking boolean	0.14	0.11	-0.027	0.021	0.18	1	-0.13	-0.025	0.088	0.16	0.13	0.079	0.12	0.033	0.33	0.089	0.025	0.18	0.06	0.1	-0.095	0.17																		
Gated boolean	-0.016	0.43	0.23	0.16	-0.052	-0.13	1	-0.018	0.013	0.48	0.17	0.12	0.19	0.066	-0.18	-0.09	0.073	-0.32	0.025	0.39	0.16	-0.38																		
Plot safe boolean	0.0086	0.003	0.0688	0.014	0.0061	0.0095	0.016	1	0.034	0.0016	0.0015	0.0018	0.012	0.04	0.0063	0.023	0.011	-0.011	0.058	0.0084	0.011	0.021																		
Garden surface (sqm)	0.22	-0.071	-0.053	-0.064	0.1	0.088	0.013	-0.034	1	-0.19	-0.023	0.079	0.34	0.14	0.038	-0.051	0.25	0.022	0.059	0.543	0.17																			
Garden boolean	0.046	-0.13	0.083	0.083	0.083	0.016	0.46	-0.016	1	0.12	0.21	0.16	0.092	0.08	0.029	0.13	0.025	0.13	-0.077	0.2																				
Roof surface (sqm)	0.28	0.033	-0.04	0.029	0.26	0.13	-0.17	0.00015	0.19	0.12	1	0.15	0.11	0.15	0.053	0.029	0.28	0.004	-0.011	-0.001	0.22																			
Terrace boolean	0.0015	0.18	0.21	0.13	-0.056	0.079	0.12	0.019	-0.022	0.21	1	0.12	-0.012	-0.065	0.024	0.051	0.074	0.064	0.15	-0.076	-0.076																			
Bathrooms total int boolean	0.23	0.069	-0.02	0.097	0.055	0.12	-0.19	-0.012	0.075	0.16	0.15	0.12	1	0.039	0.049	0.097	0.032	0.25	0.011	0.043	-0.032	0.33																		
Plot surface (sqm)	0.14	-0.026	-0.05	-0.02	0.22	0.033	0.055	-0.016	0.84	-0.211	0.11	-0.012	0.036	1	0.087	0.015	0.14	-0.005	0.034	0.0087	0.083																			
Number of fridges	0.17	0.087	-0.0033	-0.006	0.36	0.33	-0.18	-0.0002	0.14	0.062	0.18	-0.065	0.019	0.097	1	0.11	0.019	0.21	0.0045	-0.11	-0.049	0.19																		
Open fire	0.12	-0.068	-0.087	-0.072	0.13	0.089	-0.09	0.073	0.038	0.08	0.063	0.034	0.087	0.016	0.11	1	4.5e-05	0.11	0.0074	-0.011	-0.037	0.1																		
Kitchen equipped boolean	0.074	0.11	4.7e-05	0.29	0.039	0.025	0.073	0.011	-0.001	-0.025	0.026	0.051	0.032	0.015	0.019	4.5e-05	1	0.0054	0.038	0.01	0.048	-0.011																		
Living surface (sqm)	0.49	-0.2	-0.13	-0.040	0.31	0.18	0.32	0.011	0.29	0.12	0.28	0.014	0.25	0.14	0.21	0.11	0.0014	1	-0.011	-0.009	0.068	0.62																		
Draught glazing boolean	0.0084	0.049	0.029	0.07	0.031	0.016	0.025	-0.0008	0.022	0.026	0.024	0.054	0.061	0.009	0.0049	0.0014	0.028	0.011	1	0.028	0.019	0.0034																		
Accessible for disabled people boolean	0.058	0.26	0.21	0.071	0.0073	0.1	0.29	-0.0564	0.059	0.13	-0.031	0.15	0.043	0.004	0.011	-0.021	0.01	-0.066	0.078	1	0.031	-0.031																		
Furnished boolean	0.037	0.057	-0.075	0.11	0.0048	-0.095	0.16	0.011	0.043	-0.377	-0.001	-0.078	-0.002	0.007	0.049	-0.032	0.046	-0.066	0.019	0.031	1	-0.091																		
No. of bedrooms	0.38	-0.23	-0.13	-0.08	0.2	0.17	0.38	-0.0021	0.17	0.2	0.22	-0.078	0.33	0.083	0.19	0.1	-0.011	0.012	-0.034	0.11	-0.031	1																		

- Construction year /Building condition / Access / Elevator / Terrace / Energy Class / Living surface/ Number of Bedroom(<0),
  - Number of Bathroom / Number of Bedroom / Living surface
  - Terrace surface / Living surface, Swimming pool, Number of bedroom, Garden
  - Parking to number of frontage

# Quantifying features' impact on pricing for houses

Explained Variance Ratio of Principal Components for Houses



**Construction year:** 22.78%  
**Building condition:** 11.84%

**Energy class :** 9.11%  
**Double glazing:** 7.08%  
**Accessible if disabled:** 6.42%

**Living surface (sqm):** 6.04%  
**Furnished:** 5.30%  
**Nb of Bedrooms:** 4.96%  
**Nb of Bathrooms:** 4.31%  
**Kitchen equipped:** 4.00%  
**Open fire:** 3.52%



# Conclusions : 5 crucial parameters

- **Geographical location** (locality/province/region):
  - Cultural and economic factors are key influencers.
- **Type of property:** Houses/Apartments
  - Necessity for separate modeling due to disparities?
- **Energy efficiency rating:**
  - Increasing global and political significance.
- **Year of construction/building condition:**
  - Influence on housing quality and renovation costs.
- **Living area/number of bedrooms/number of bathrooms:**
  - Property dimensions and layout crucial for market value.

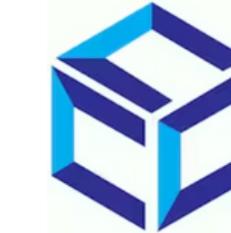
# Conclusions : 4 desirable parameters



- **Parking space:**
  - Still a practical necessity for buyers
- **Swimming pool:**
  - Swimming pool indicate a premium for luxury amenities
- **Equipped kitchen:**
  - Equipped kitchens are the standard
- **Open fire:**
  - Synonym of comfort and ambiance

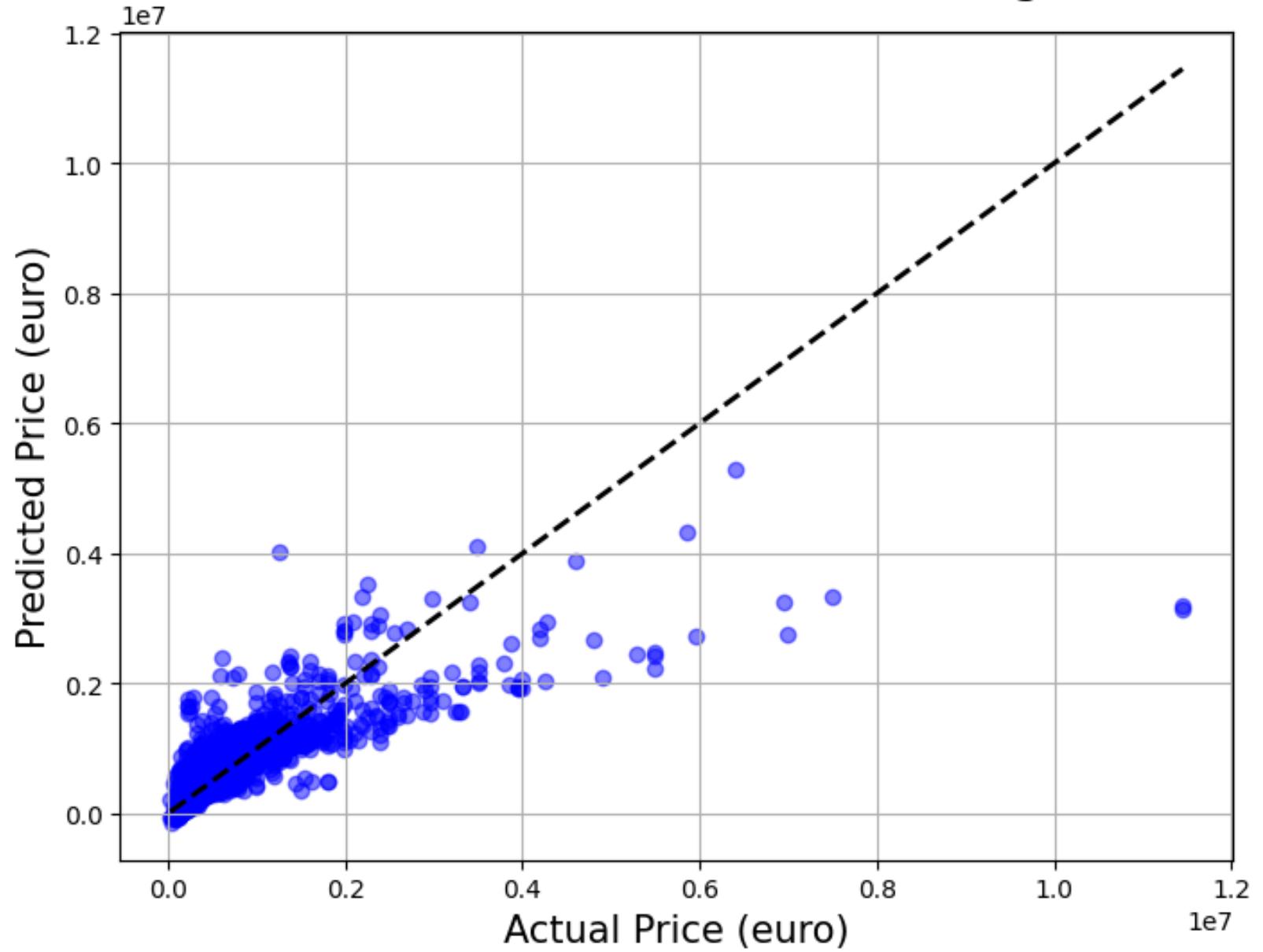
# Next Step : Price Prediction

BY QBICLE.COM

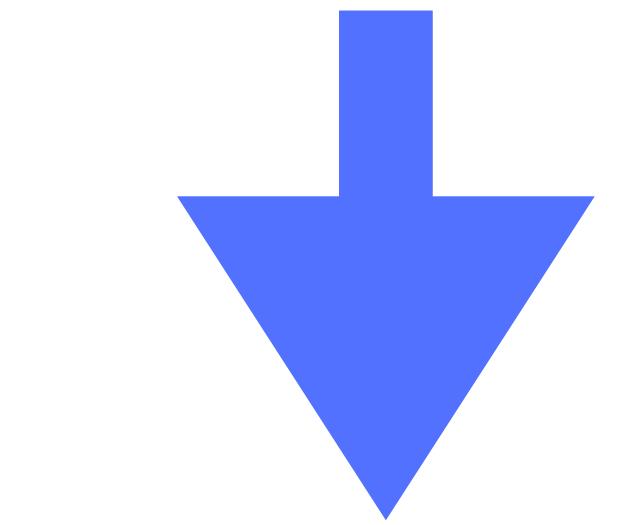


Ready for sprint 3: Create the machine learning model

Actual vs Predicted Prices (Linear regression)



Regression Model Performance:  
R-squared: 0.664  
Mean Absolute Error: 133,798.43



Room for improvement

**IMMO ELIZA  
PROJECT: BREPP\_AI**

BY QBICLE.COM



THANK YOU !

&

SEE YOU NEXT WEEK !

Mahak, Maarten, Sylvain

**IMMO ELIZA  
PROJECT: BREPP\_AI**

BY QBICLE.COM

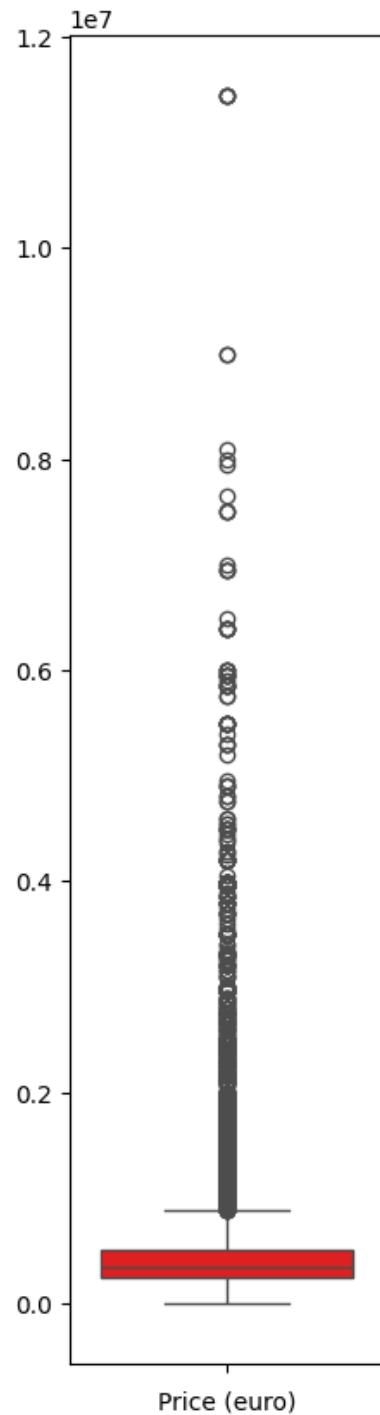
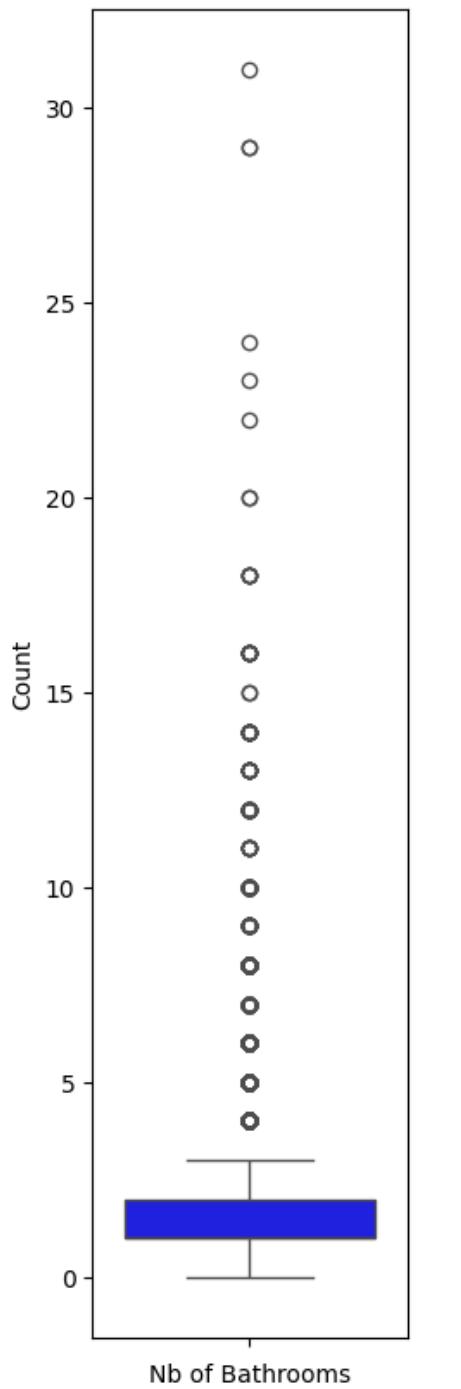
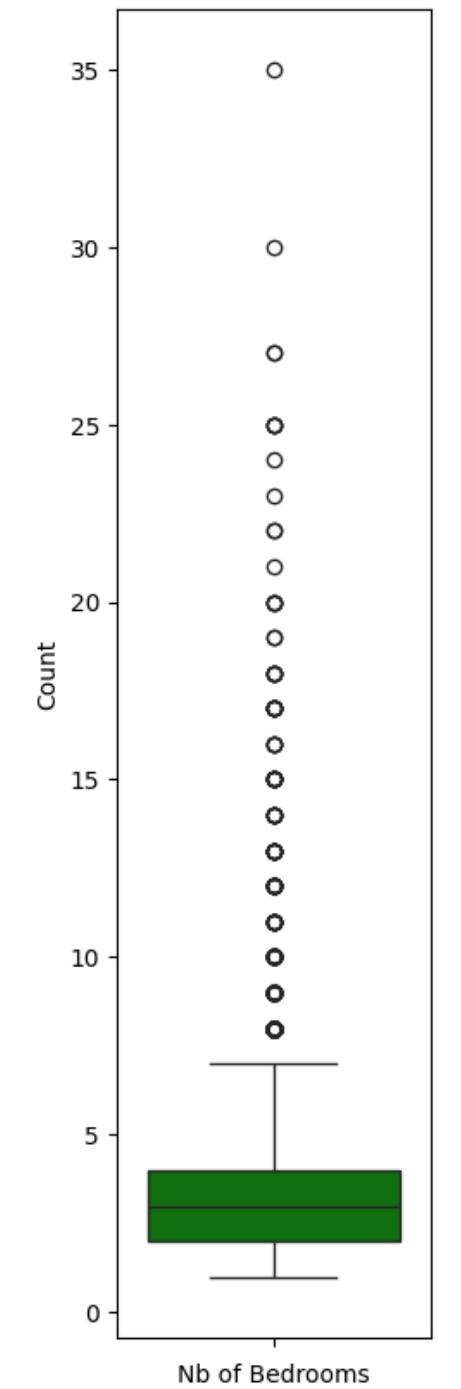
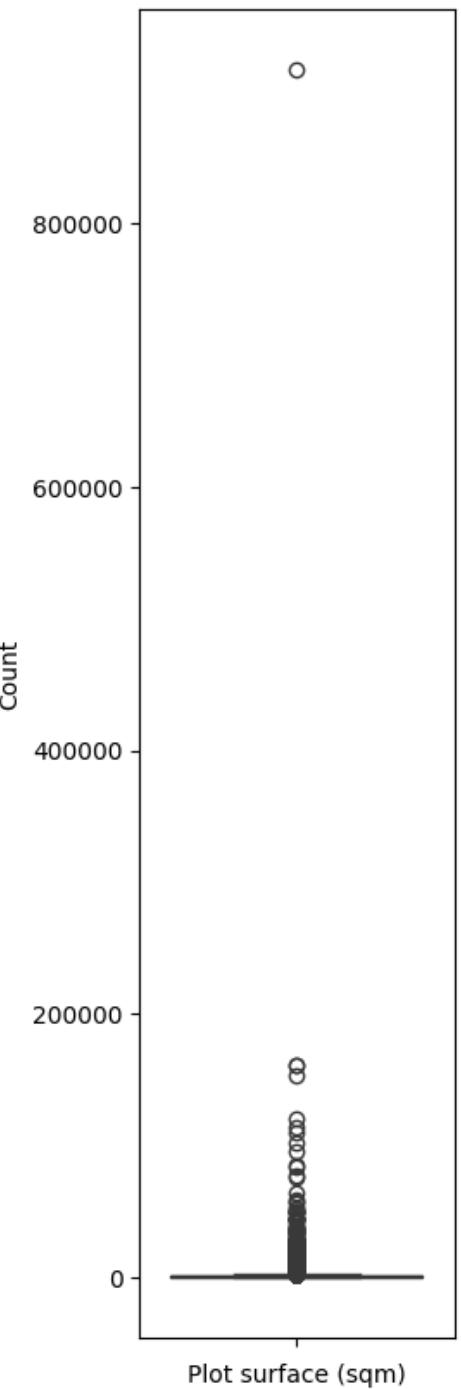
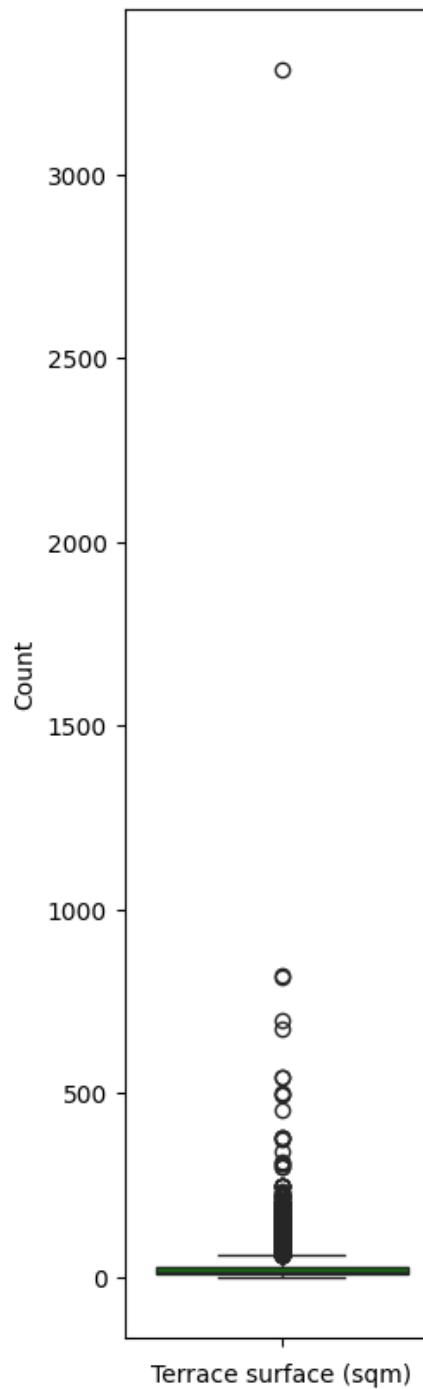
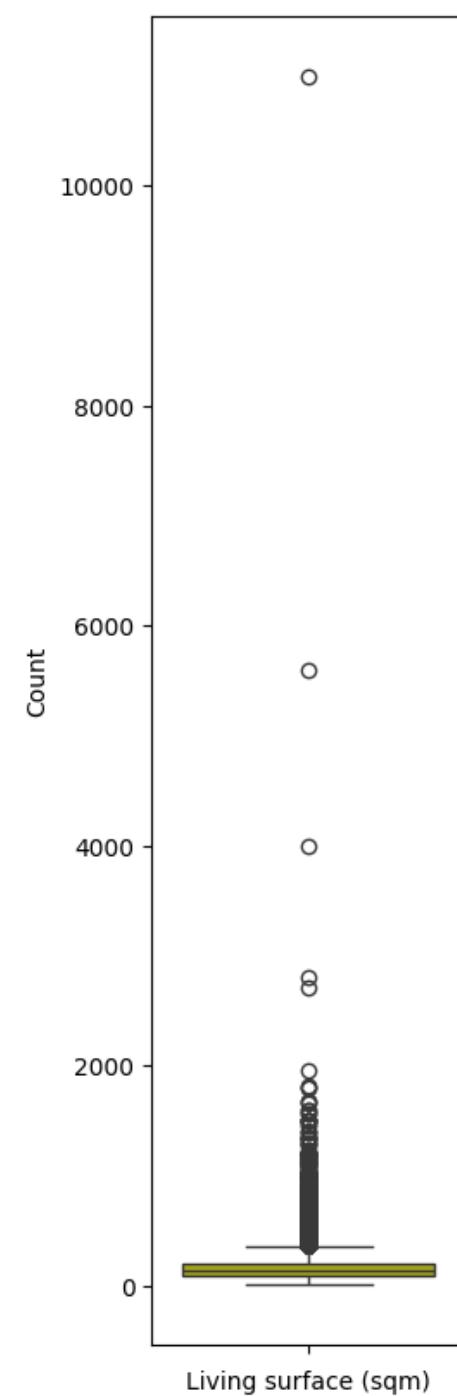


back up

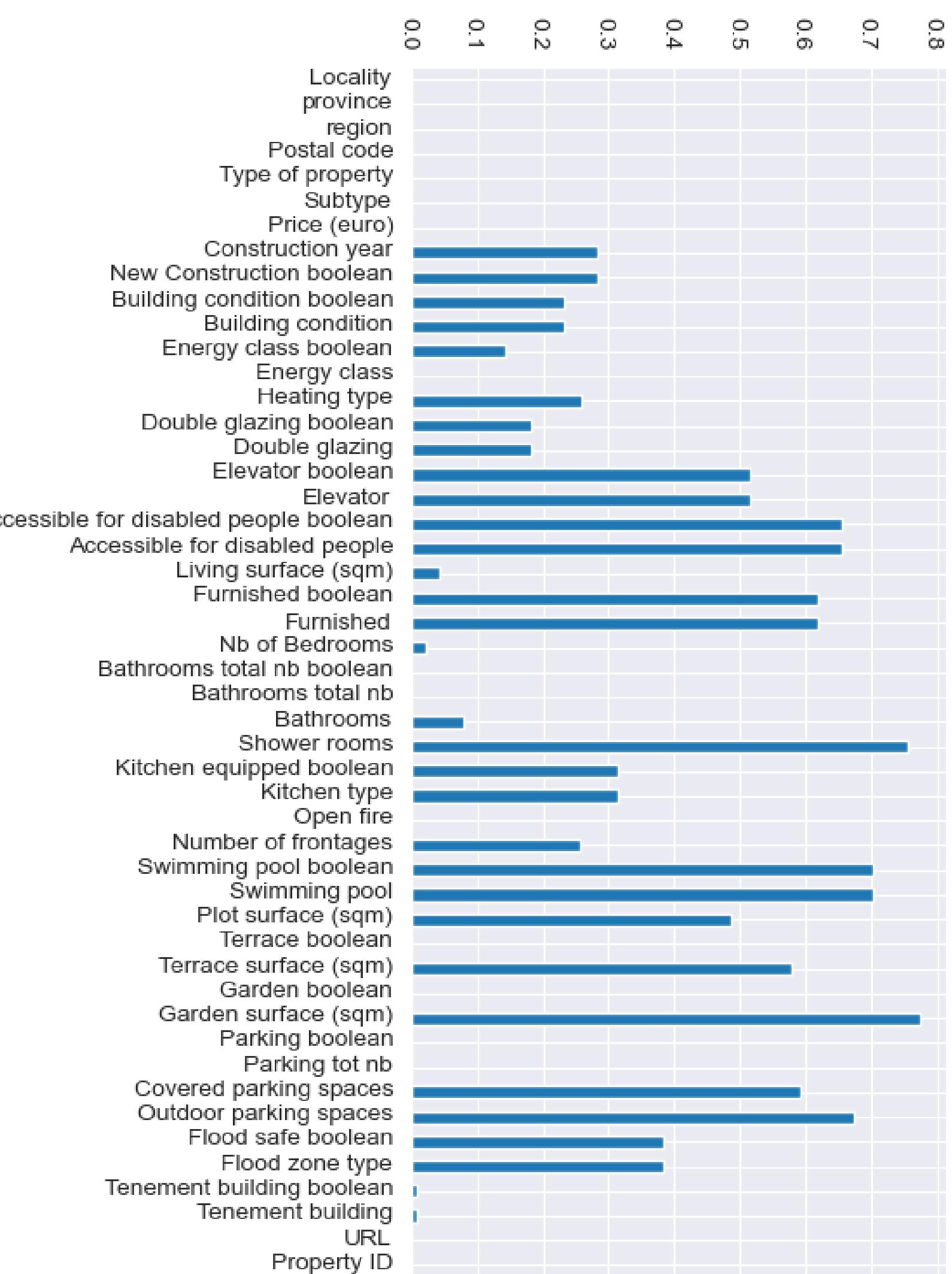
Mahak, Maarten, Sylvain

**IMMO ELIZA  
PROJECT: BREFP\_AI**

# Parameters with Outliers



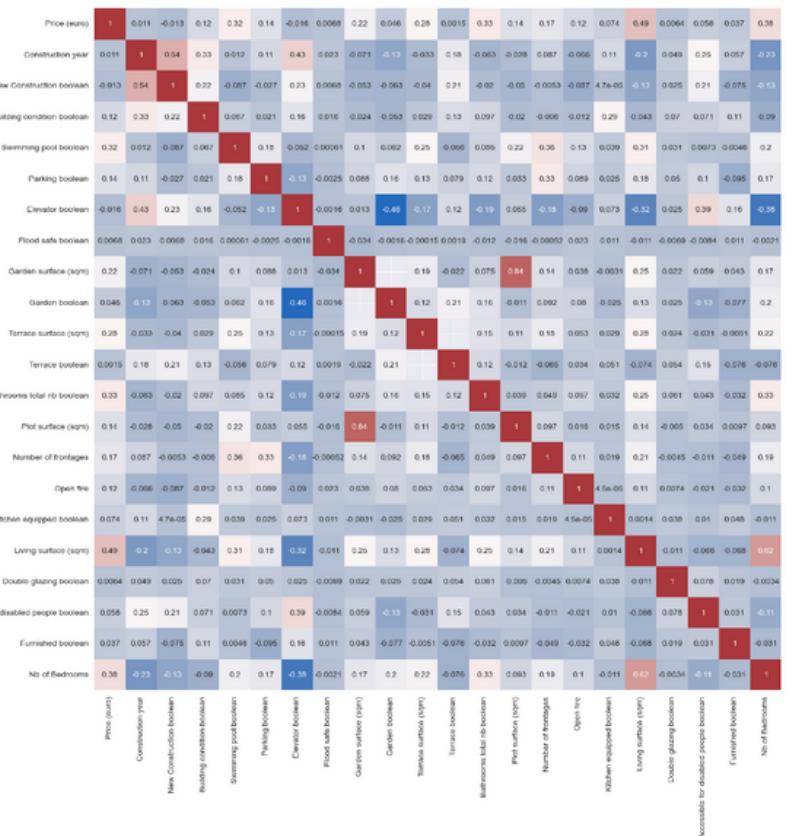
# Nan



# Intervariable correlation

## Examples of intervariable correlation

- construction year /building condition / access / elevator / terrace / nrj / living surface/ bedroom nb(<0),
- bathroom nb / bedroom nb / living surface
- terrace surface / living surface, swimming pool, nb of bedroom, garden
- parking to nb of frontage



# Conclusion: Impactful Parameters

BY QBICLE.COM



## MAIN PARAMETER

- Type of Building : Houses or Appartement >> 2 arms
- Geoloc (locality/province/region)
- Energy class
- Construction year /building condition
- Living surface /nb of bedrooms / nb bathrooms

## SIMPLE PRICE CORREL

- Price Analysis :
- Energy Class
- Good Building condition
- Tenement Buildings
- Equipped kitchen
- No. of Bathroom /No. of bedrooms/Living Surface
- Open Fire
- Swimming Pool
- Garden
- Parking

## PRINCIPAL COMPONENT ANALYSIS

- Construction year: 22.78%
- Good Building condition : 11.84%
- Energy class (A+,A & B) : 9.11%
- double glazing and
- Accessible for disabled people: 6.42

## More crispy recommendation

- cocktail in the pool ?
- kitchen equiped chicken happy !
- Belgium Hygge (said "hooga" is of Danish) > light a fire
- Belgium car lover > parking

## • correlation MAP

**0.10 - 0.15**

- Good building condition
- Parking present
- Plot surface
- Open fire

**0.16 - 0.30**

- Garden surface
- Number of frontage
- Terrace surface

## Above 0.30

- Number of bathroom
- Number of bedroom
- Living surface
- Swimming pool

**IMMO ELIZA  
PROJECT: BREFP\_AI**