

IMMOELIZA DATA ANALYSIS

Nicole Pretorius
Vera Ilias
Hussain

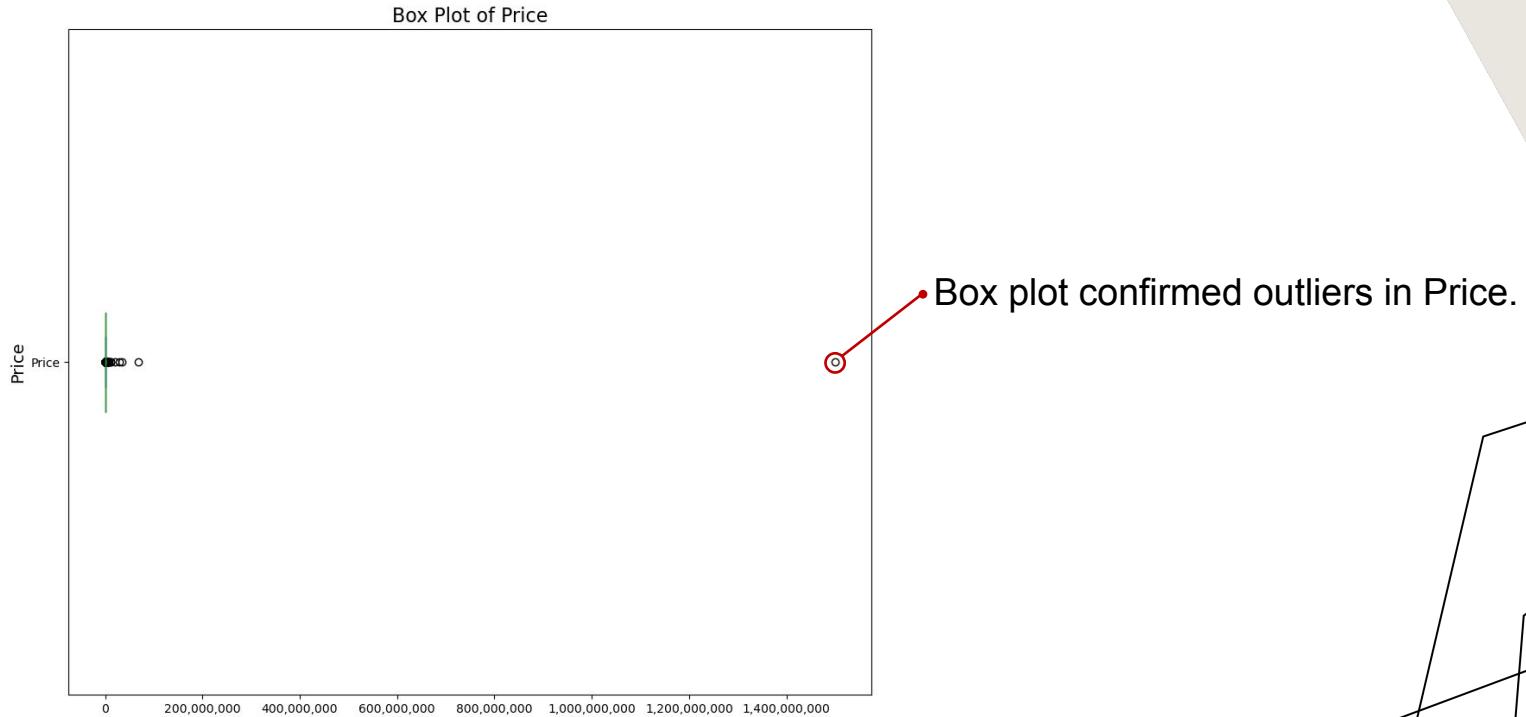
ABOUT SELECTED DATA

- 13 707 rows
- 20 columns
- Type: Apartments , Houses and other

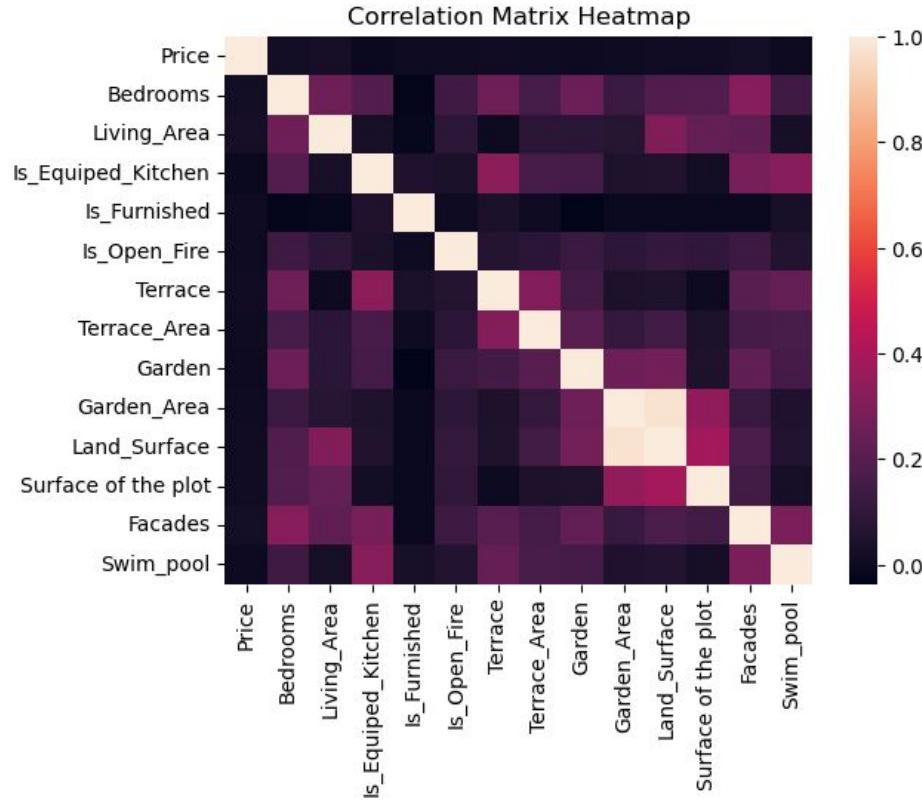
But it has *problems!*

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13707 entries, 0 to 13706
Data columns (total 20 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Id               13707 non-null   int64  
 1   Locality         13707 non-null   object  
 2   Type             13707 non-null   object  
 3   SubType          13707 non-null   object  
 4   Sale_Type        13707 non-null   object  
 5   Price            13707 non-null   int64  
 6   Bedrooms         13707 non-null   int64  
 7   Living_Area     13707 non-null   int64  
 8   Is_Equiped_Kitchen 13707 non-null   bool   
 9   Is_Furnished    13707 non-null   bool   
 10  Is_Open_Fire    13707 non-null   bool   
 11  Terrace          13707 non-null   bool   
 12  Terrace_Area    13707 non-null   int64  
 13  Garden           13707 non-null   bool   
 14  Garden_Area     13707 non-null   int64  
 15  Land_Surface    13707 non-null   int64  
 16  Surface of the plot 13707 non-null   int64  
 17  State            13707 non-null   object  
 18  Facades          13707 non-null   int64  
 19  Swim_pool        13707 non-null   bool  
dtypes: bool(6), int64(9), object(5)
memory usage: 1.5+ MB
```

OUTLIERS



OUTLIERS



Initial correlation matrix suggested presence of outliers.

WHICH VARIABLES WOULD YOU DELETE AND WHY ?

Sale_Type(c): Only value is “Private”.

Surface of the plot(c): Over 50% empty values. Some values are smaller than Land_Surface which isn’t logical. Overall unreliable.

“Other” category in Type (r): Only interested in houses and apartments - not garages, covered parking, offices, etc.

STEP 1 : DATA CLEANING

- ✓ Fixed data types
- ✓ Dropped duplicates (92 rows)
- ✓ Dropped:
 - Column “Sales_Type”
 - Category “Other” in “Type”
 - Column “Surface of the plot”
- ✓ Imputed 0 values:
 - Living_Area
 - Land_Surface
- ✓ Changed Booleans to binary values

CLEANED DATASET

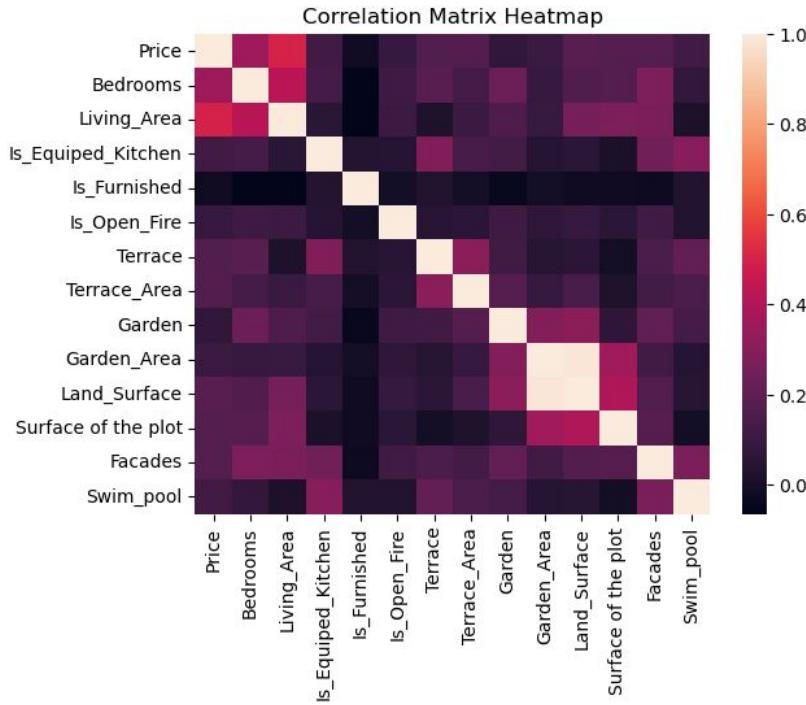
Int	Cat	Cat	Cat	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Cat	Int	Int
Id	Locality	Type	SubType	Price	Bed rooms	Living _Area	Is_Equippe d_Kitchen	Is_Furnished	Is_Open_Fire	Terrace	Terrace _Area	Garde n	Garden _Area	Land Surface	State	Facades	Swimming _pool
2031 6087	Ertvelde	House	house	329000	2	127	1	0	0	1	0	0	0	146	Good	3	0
2031 6069	Hornu	House	house	180000	3	138	0	0	0	1	1	1	1	140	As new	2	1
2031 6048	Beersel	House	house	369000	6	230	0	0	0	1	0	0	0	230	To be done up	3	1

- Columns: 18
- Rows: 10,684
- Number of Qualitative Variables: 4
- Number of Quantitative Variables: 14
- Missing Values: 0



STEP 2: ANALYSIS

WHAT IS THE CORRELATION BETWEEN THE VARIABLES AND THE PRICE?

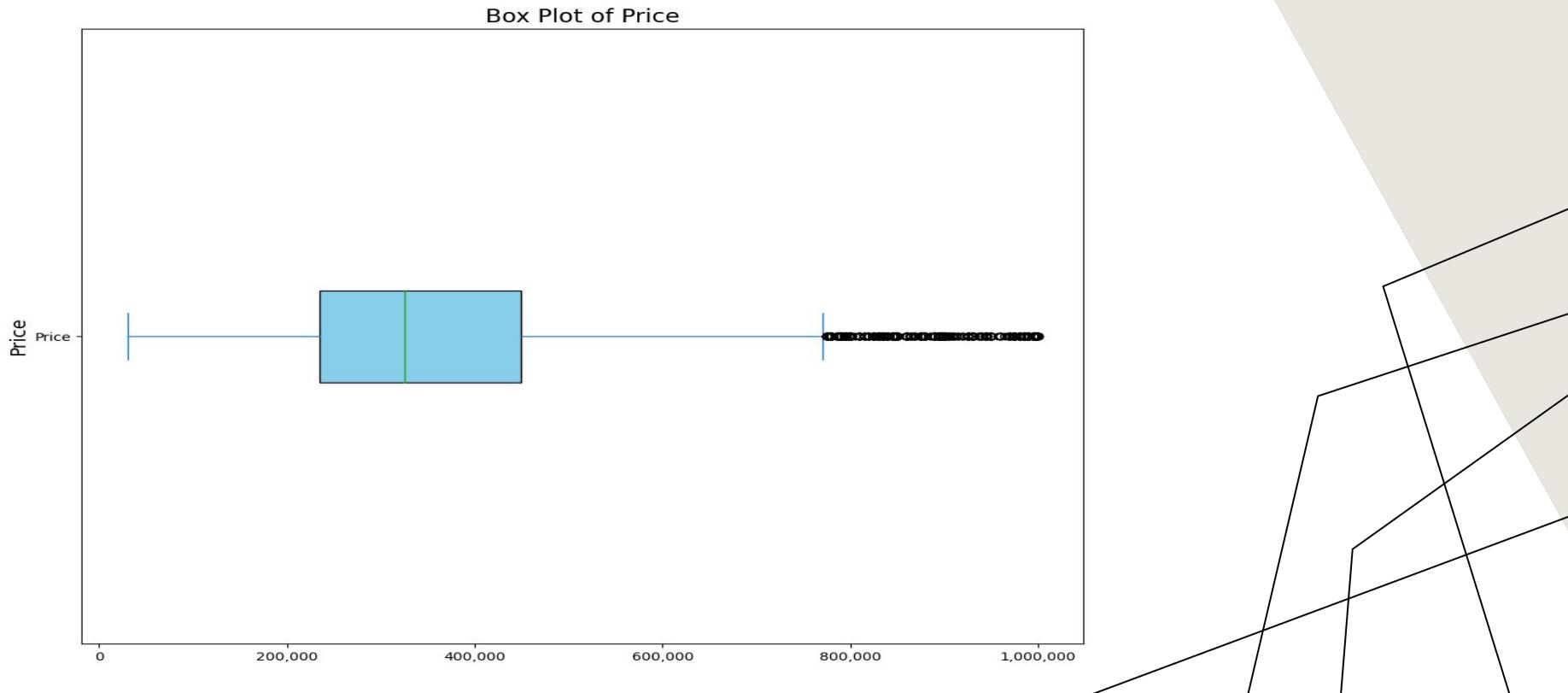


Which variables have the greatest influence on the price & how are they correlated?

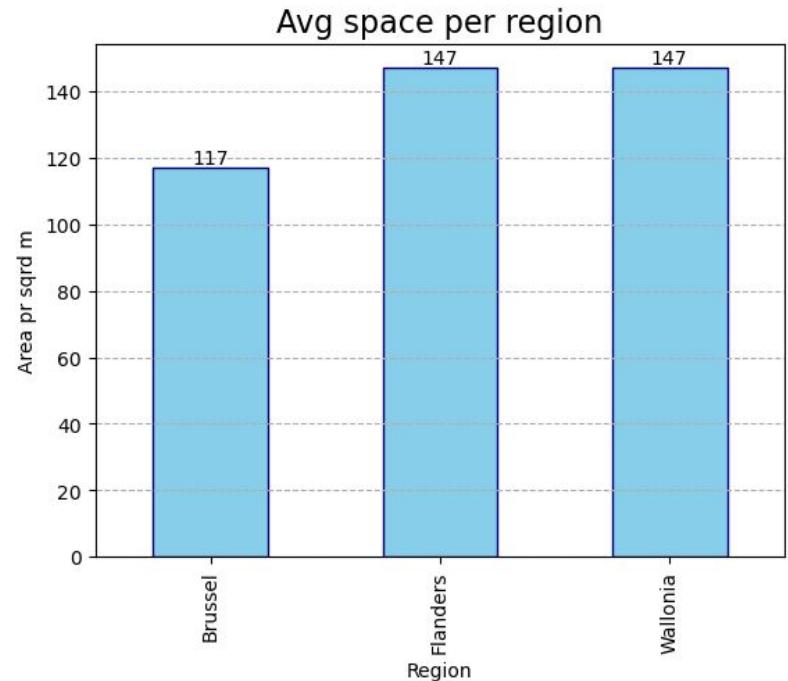
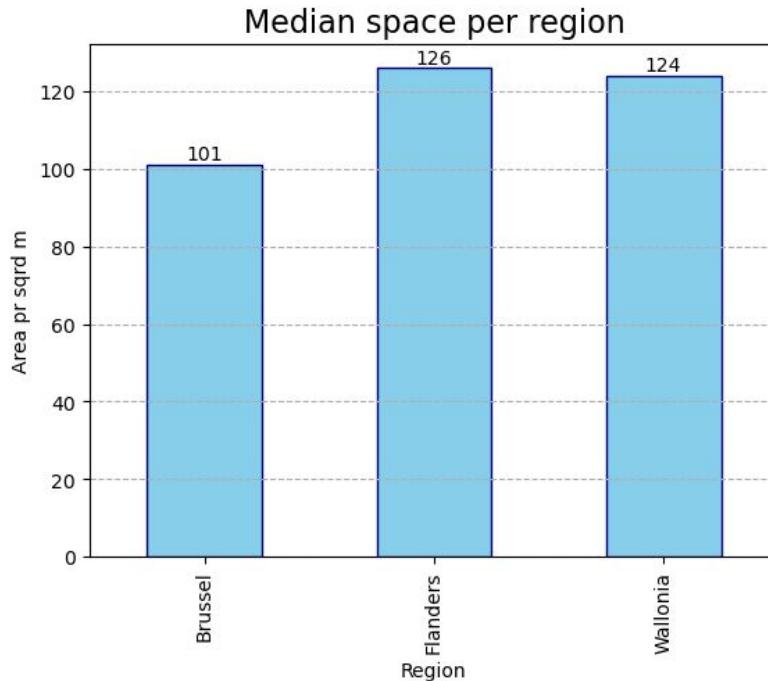
The matrix shows a positive correlation between Price and both Living Area and Bedrooms.

This suggests that properties with more rooms and greater living space tend to be more expensive.

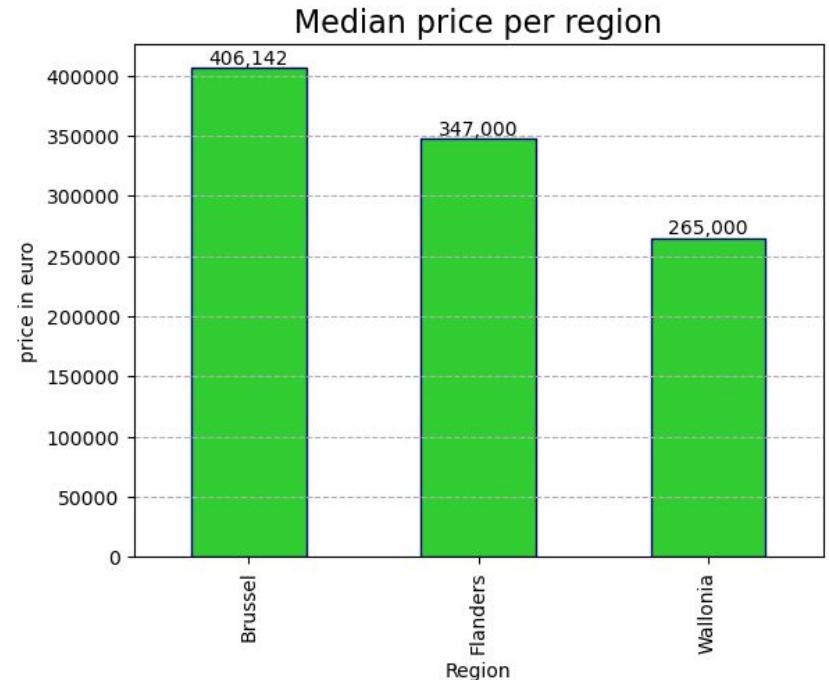
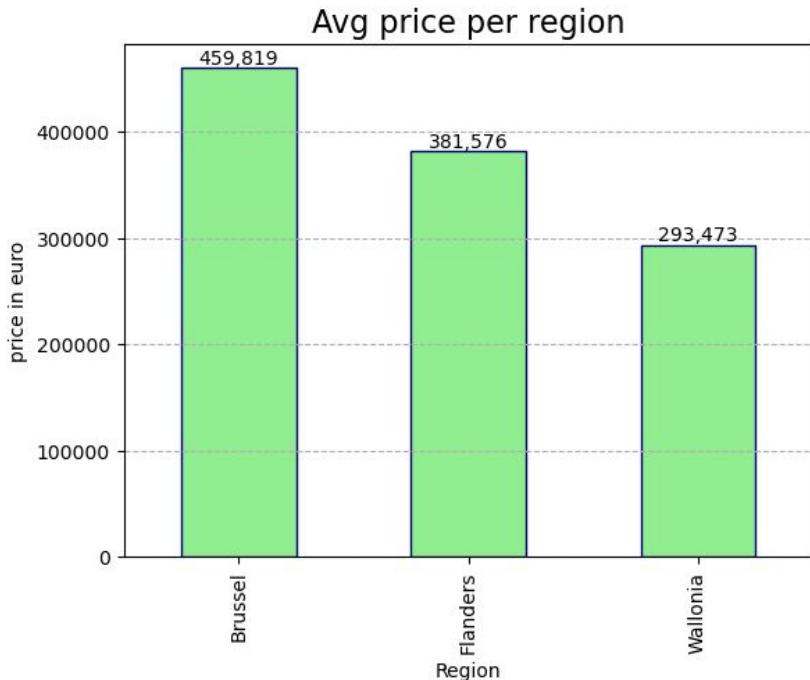
ABOUT OUR CLEANED DATA



ABOUT OUR CLEANED DATA



ABOUT OUR CLEANED DATA

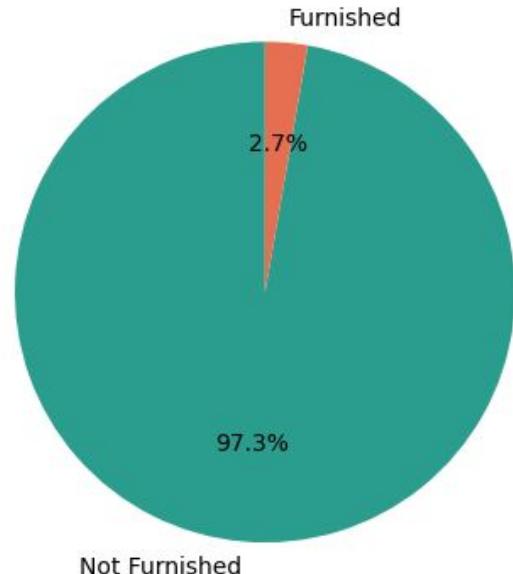


WHAT IS THE CORRELATION BETWEEN THE VARIABLES AND THE PRICE?

**Which variables have the least influence
on the price & how are they correlated?**

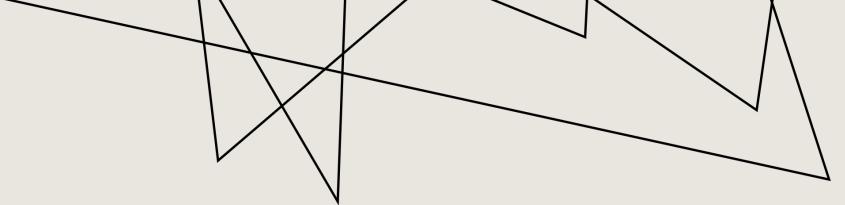
Whether the property is furnished or not has the strongest negative correlation due to the fact that the majority of properties are unfurnished as a standard.

Percentage of Furnished vs. Unfurnished Properties

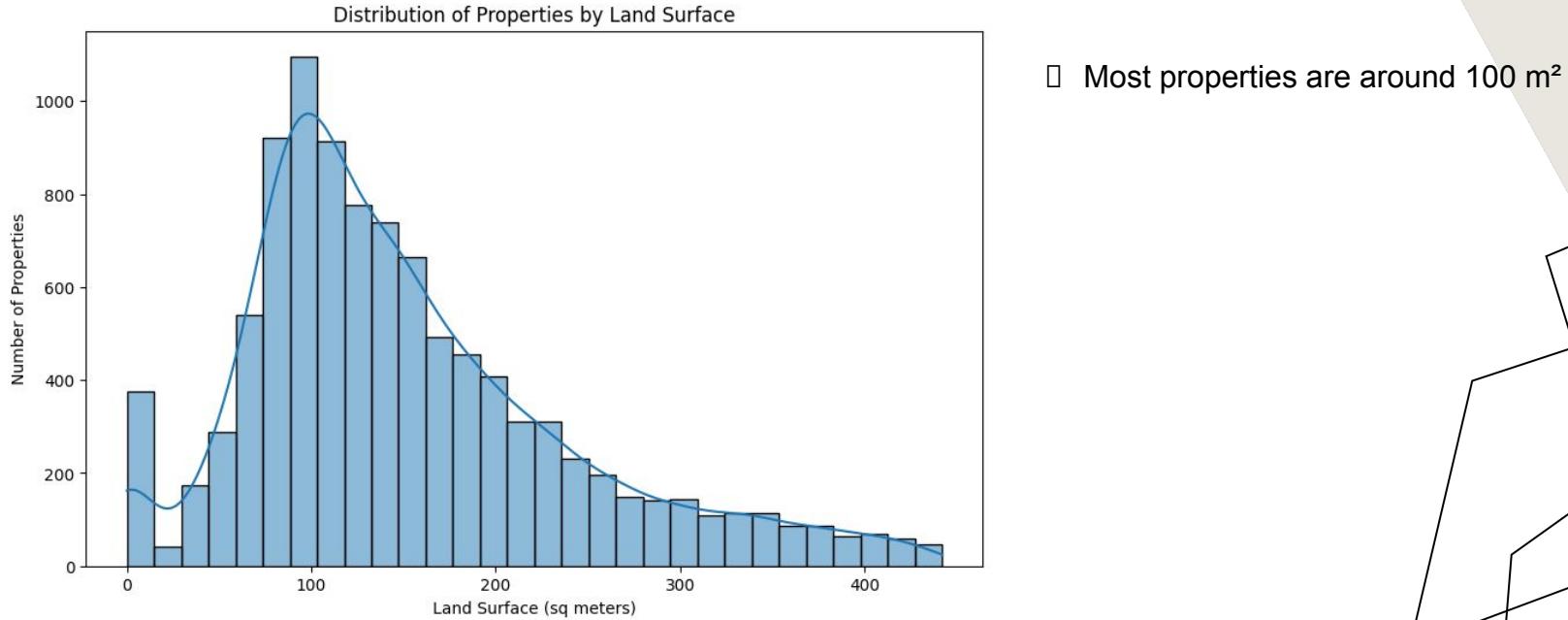




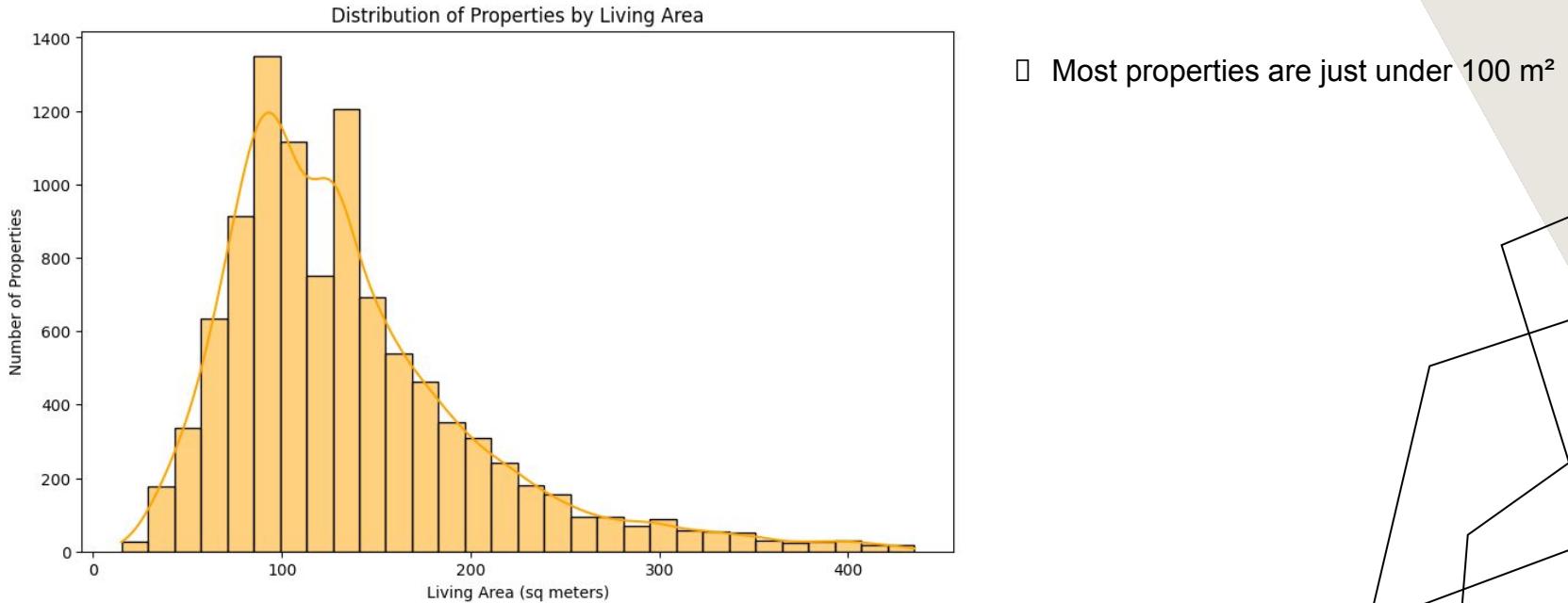
STEP 3: INTERPRETATION



REPRESENT THE NUMBER OF PROPERTIES ACCORDING TO THEIR SURFACE USING A HISTOGRAM.



REPRESENT THE NUMBER OF PROPERTIES ACCORDING TO THEIR SURFACE USING A HISTOGRAM.



WHICH 5 VARIABLES ARE THE MOST IMPORTANT AND WHY?

Price: Central point of analysis.

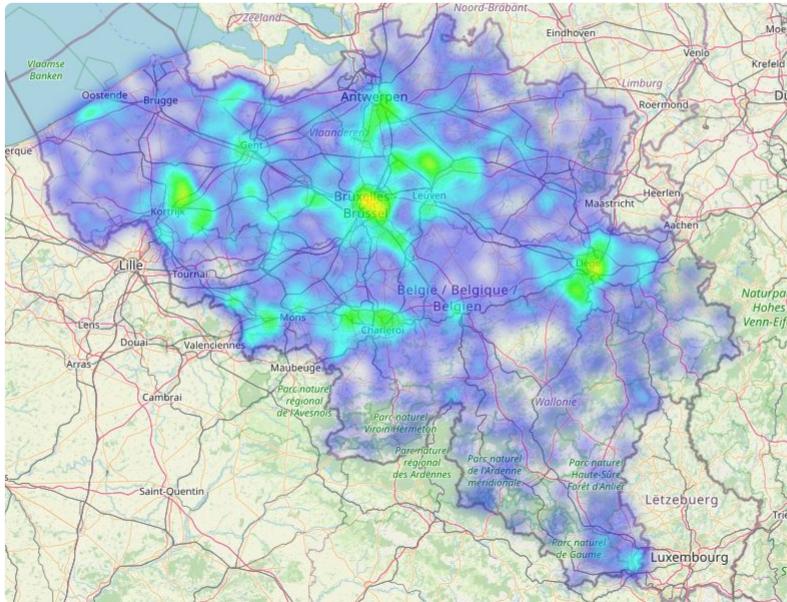
Locality: Key decision-making factor + can significantly impact price.

Living_Area and/or Land_Surface: Size of building/plot can have significant impact on price.

Type: Houses tend to be more expensive than apartments.

State: Can signal whether the buyer might face higher long-term costs than the initial price implies.

HEATMAP OF PROPERTY PRICE DISTRIBUTION IN BELGIUM

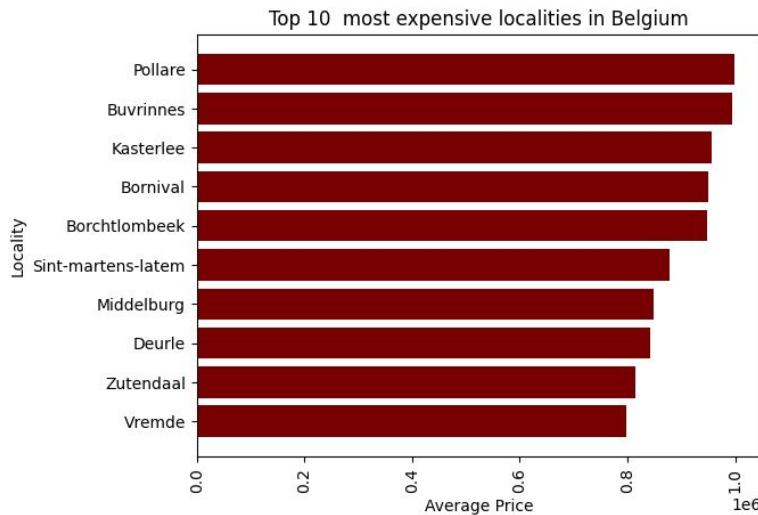


- Highest property prices in Brussels.
- Expensive properties cluster in cities.
- Cities outprice rural areas.
- More data available for Flanders than Wallonia.

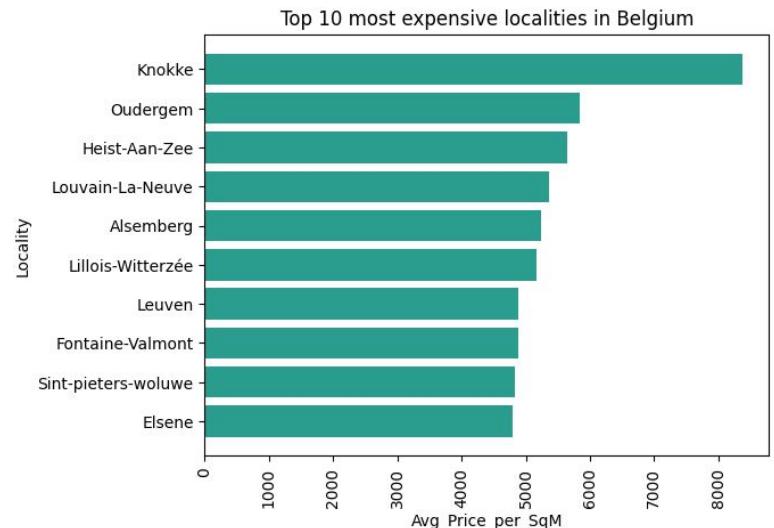
WHAT ARE THE **MOST** EXPENSIVE MUNICIPALITIES IN BELGIUM?

Living Area

Sorted by Avg Price:



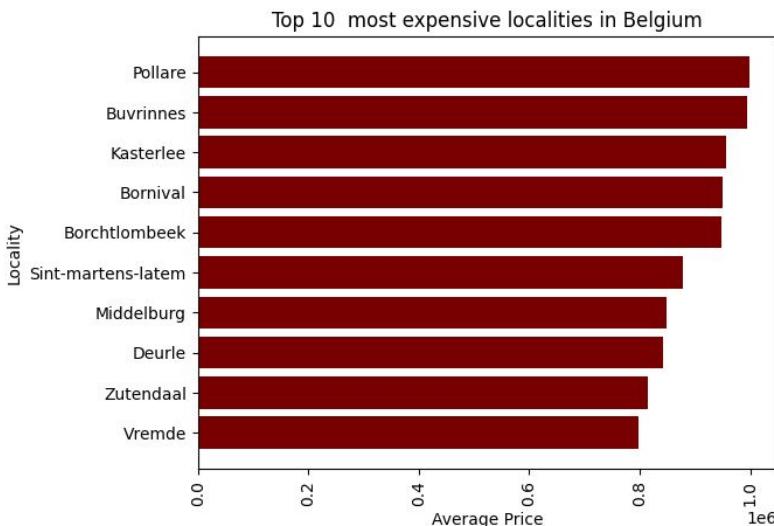
Sorted by Avg Price per m²:



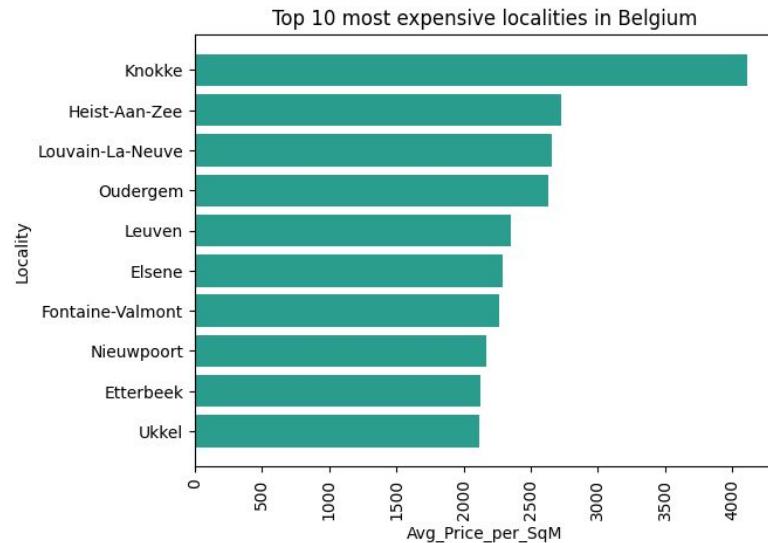
WHAT ARE THE **MOST** EXPENSIVE MUNICIPALITIES IN BELGIUM?

Land Surface +
Living Area

Sorted by Avg Price:



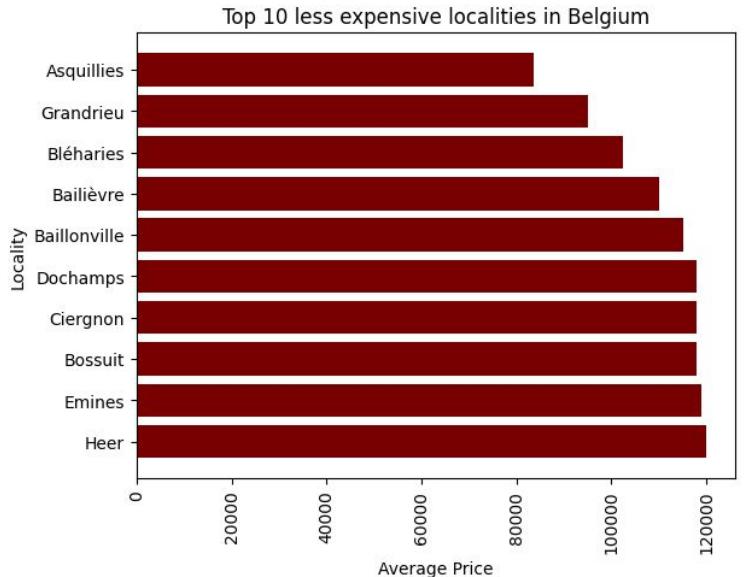
Sorted by Avg Price per m²:



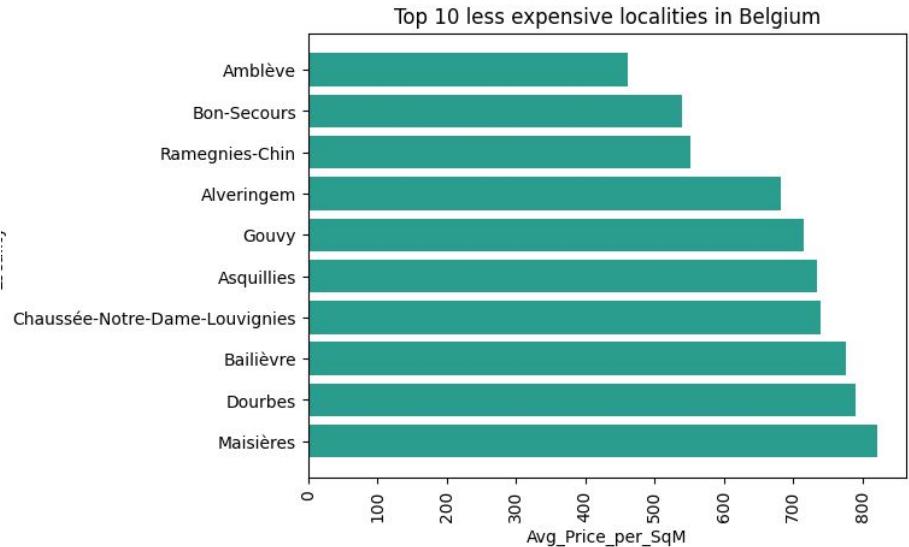
WHAT ARE THE **LESS** EXPENSIVE MUNICIPALITIES IN BELGIUM?

Living Area

Sorted by Avg Price:



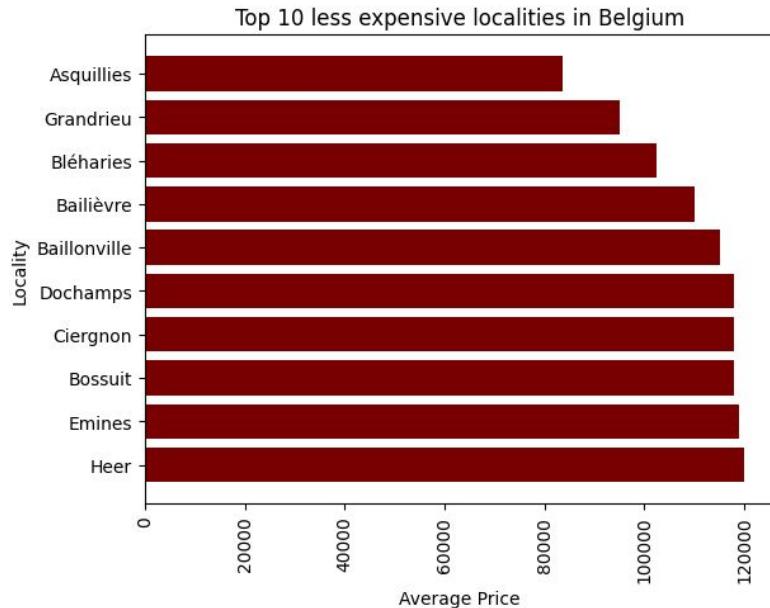
Sorted by Avg Price per m²:



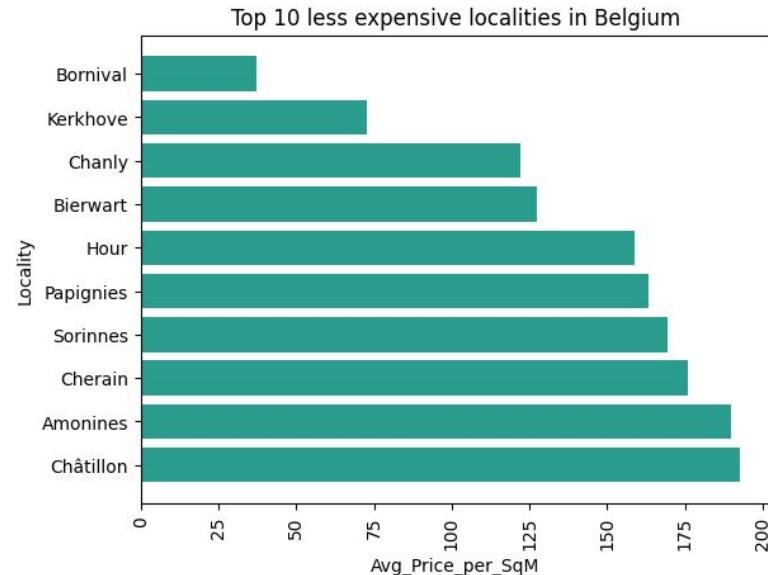
WHAT ARE THE **LESS** EXPENSIVE MUNICIPALITIES IN BELGIUM?

Land Surface + Living Area

Sorted by Avg Price:



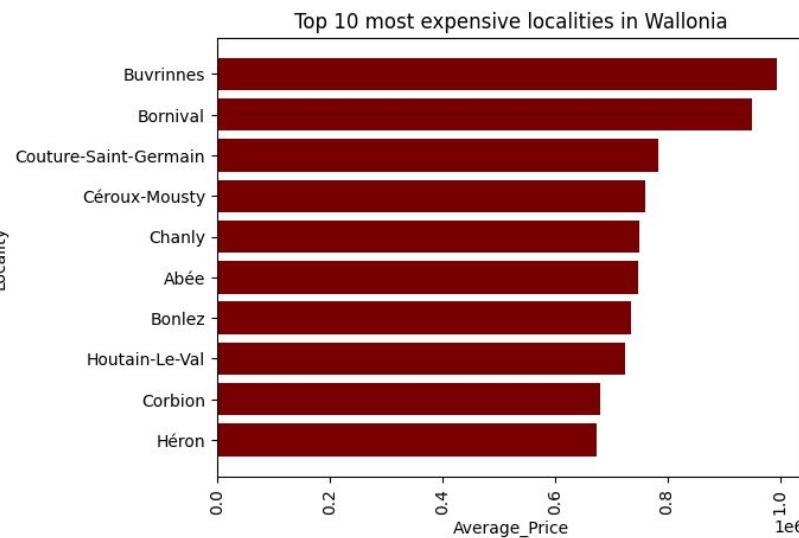
Sorted by Avg Price per m²:



WHAT ARE THE **MOST** EXPENSIVE MUNICIPALITIES IN WALLONIA?

Living Area

Sorted by Avg Price:



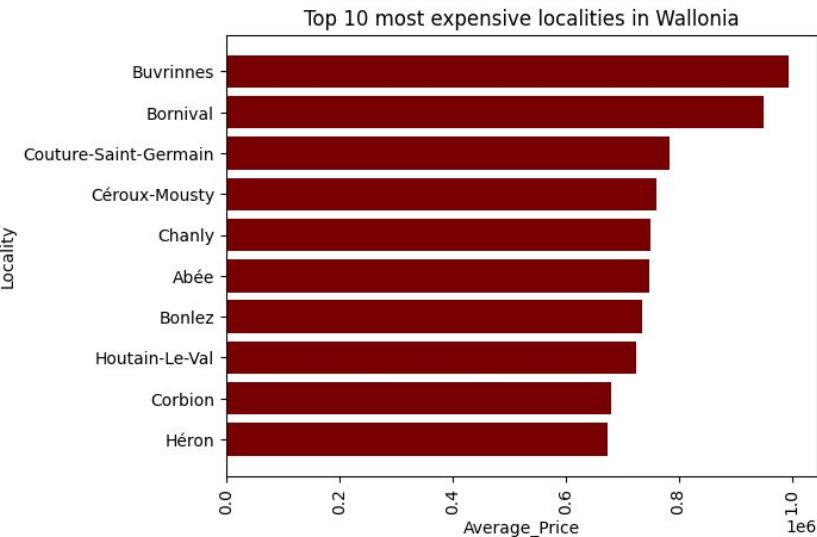
Sorted by Avg Price per m²:



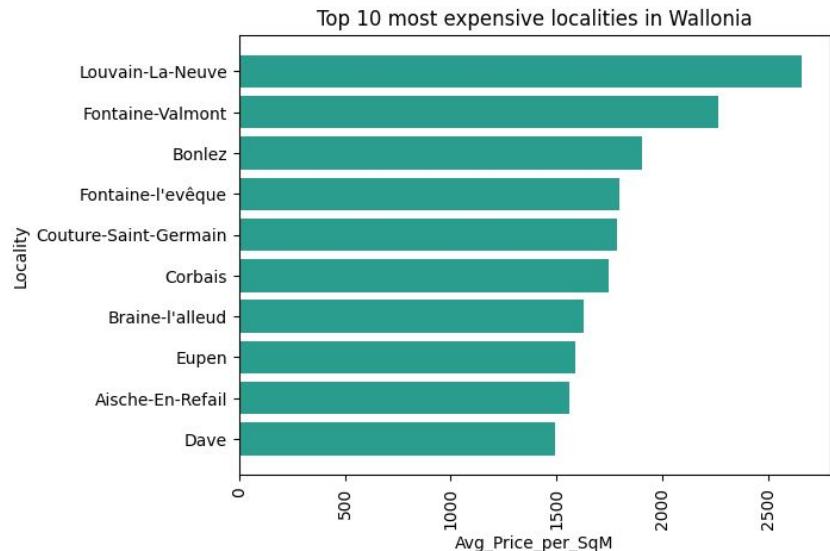
WHAT ARE THE **MOST** EXPENSIVE MUNICIPALITIES IN WALLONIA?

Land Surface + Living Area

Sorted by Avg Price:



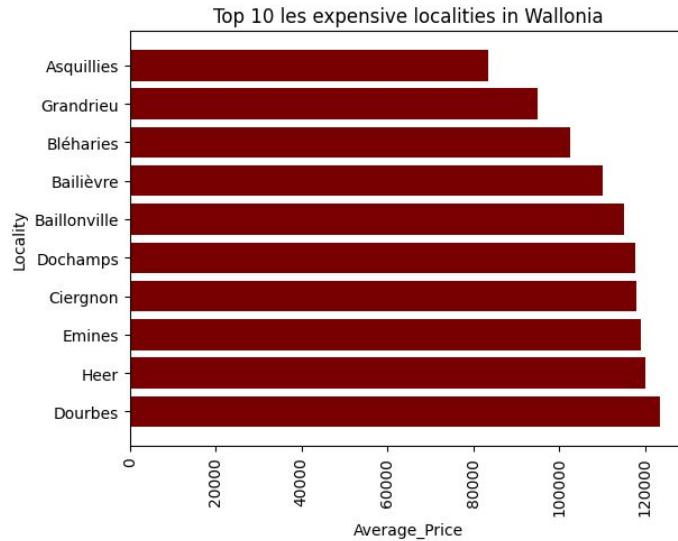
Sorted by Avg Price per m²:



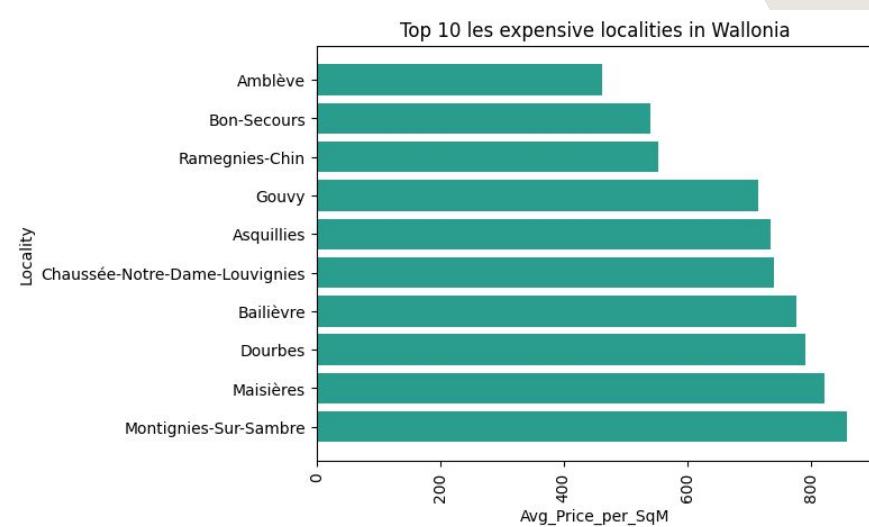
WHAT ARE THE **LESS** EXPENSIVE MUNICIPALITIES IN WALLONIA?

Living Area

Sorted by Avg Price:



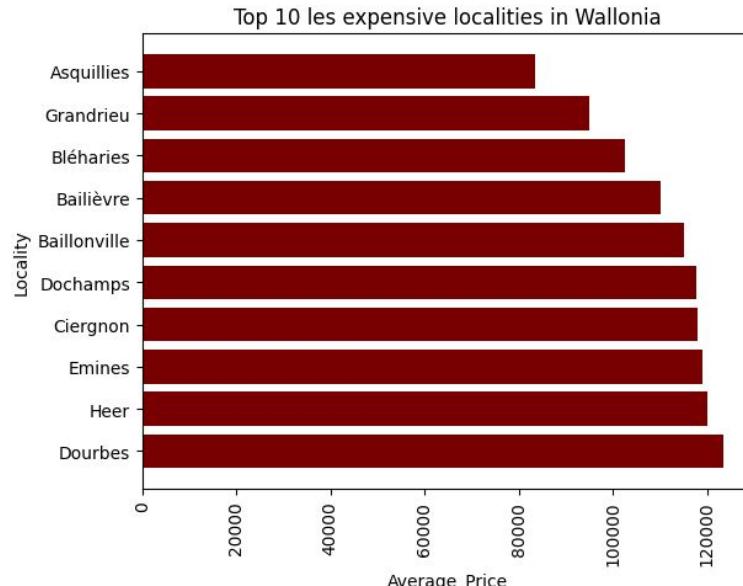
Sorted by Avg Price per m²:



WHAT ARE THE **LESS** EXPENSIVE MUNICIPALITIES IN WALLONIA?

Land Surface + Living Area

Sorted by Avg Price:



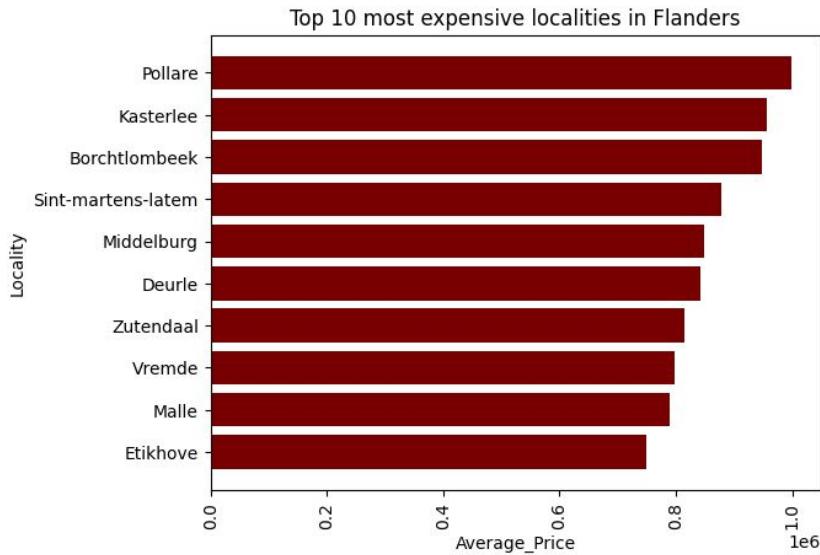
Sorted by Avg Price per m²:



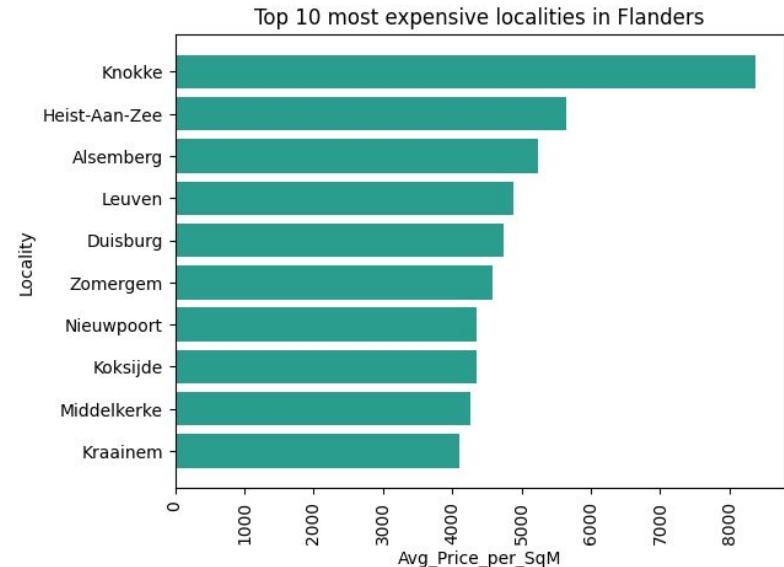
WHAT ARE THE **MOST** EXPENSIVE MUNICIPALITIES IN FLANDERS?

Living Area

Sorted by Avg Price:



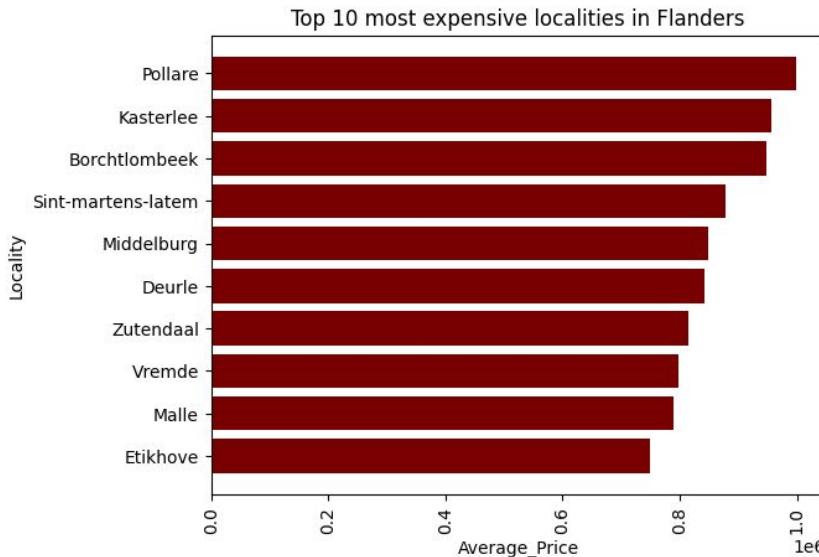
Sorted by Avg Price per m²:



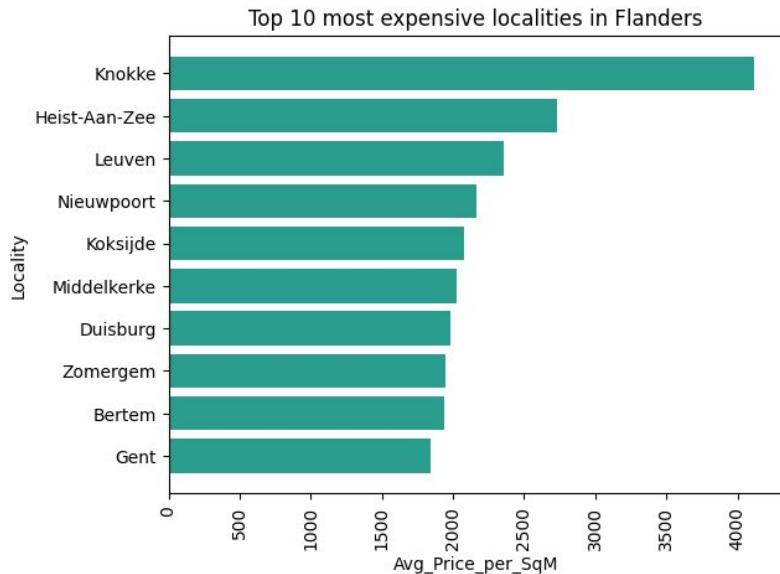
WHAT ARE THE **MOST** EXPENSIVE MUNICIPALITIES IN FLANDERS?

Land Surface + Living Area

Sorted by Avg Price:



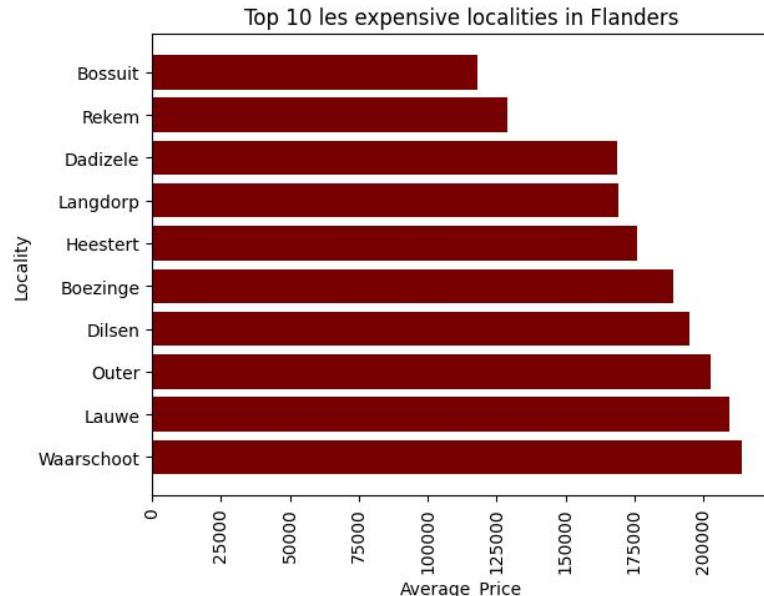
Sorted by Avg Price per m²:



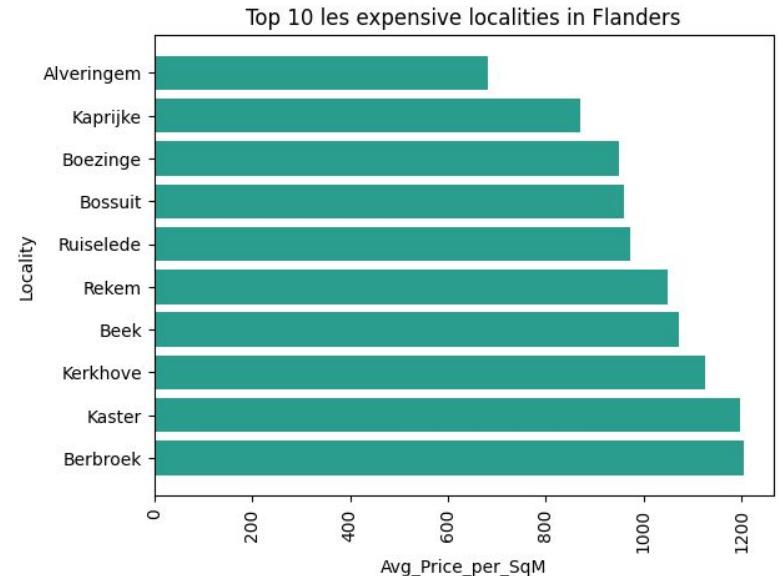
WHAT ARE THE **LESS** EXPENSIVE MUNICIPALITIES IN FLANDERS?

Living Area

Sorted by Avg Price:



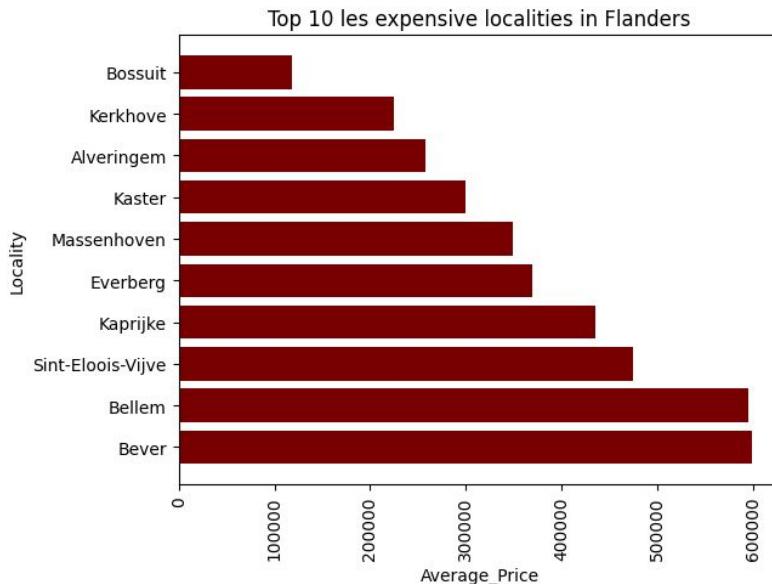
Sorted by Avg Price per m²:



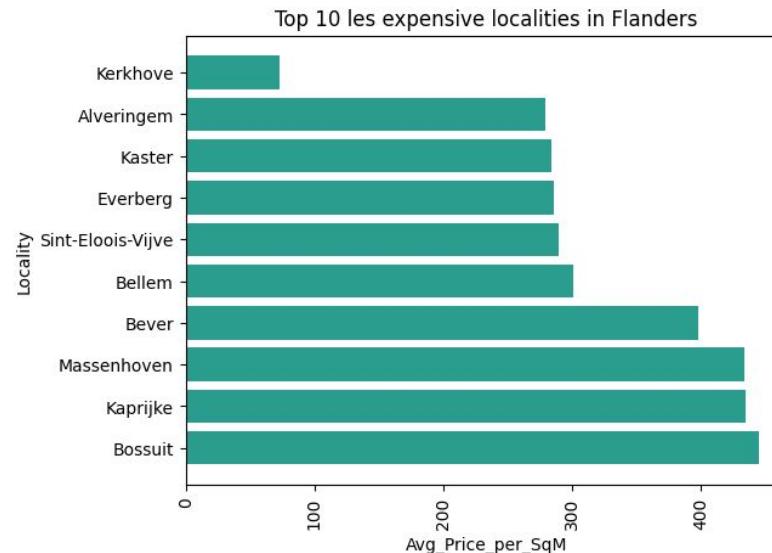
WHAT ARE THE **LESS** EXPENSIVE MUNICIPALITIES IN FLANDERS?

Land Surface + Living Area

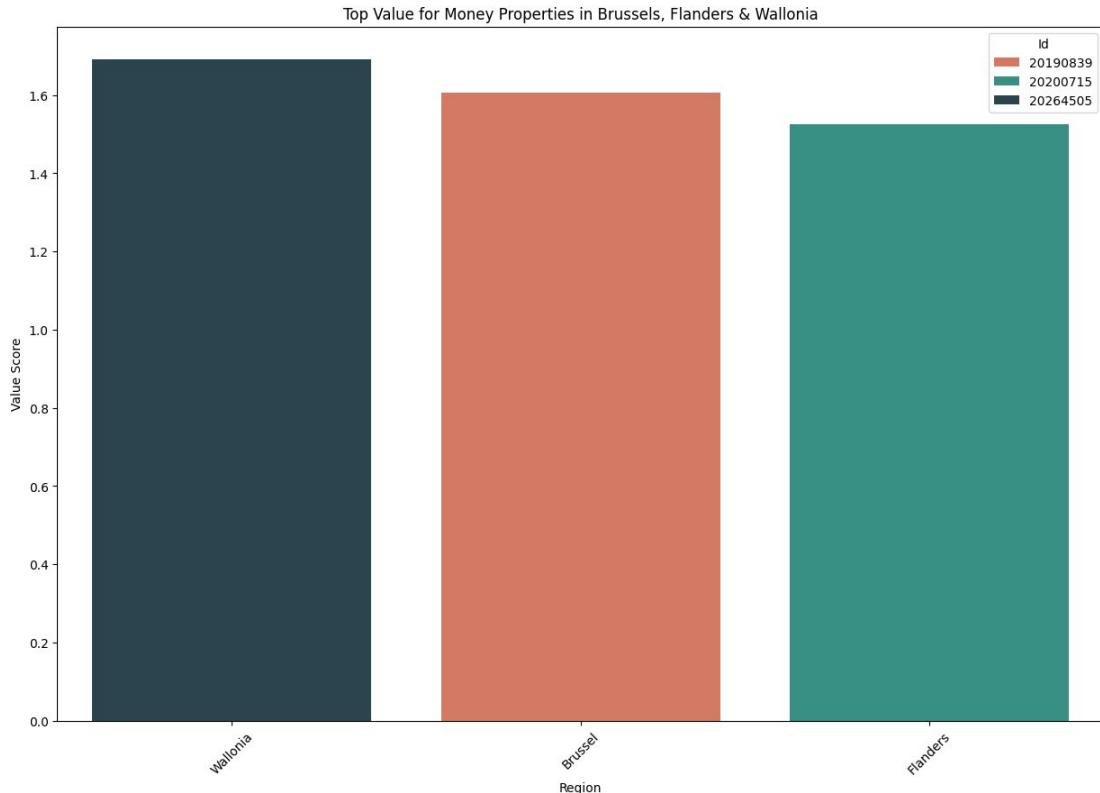
Sorted by Avg Price:



Sorted by Avg Price per m²:



WHICH PROPERTY IS THE BEST VALUE FOR MONEY?



WHICH PROPERTY IS THE BEST VALUE FOR MONEY?

Wallonia:

<https://www.immoweb.be/en/classified/apartment/for-sale/saint-gilles/1060/20264505>

€269,000 | 785 m²

Commercial and residential building.

Two-storey house and attic as well as several warehouses on the floors.

In need of renovation.

Large garden and swimming pool.



WHICH PROPERTY IS THE BEST VALUE FOR MONEY?

Brussels:

<https://www.immoweb.be/en/classified/apartment/for-sale/saint-gilles/1060/20190839>

€219,000 | 149 m² | 2 bedrooms

Spacious unfurnished apartment with an equipped kitchen, front and back terrace.



WHICH PROPERTY IS THE BEST VALUE FOR MONEY?

Flanders:

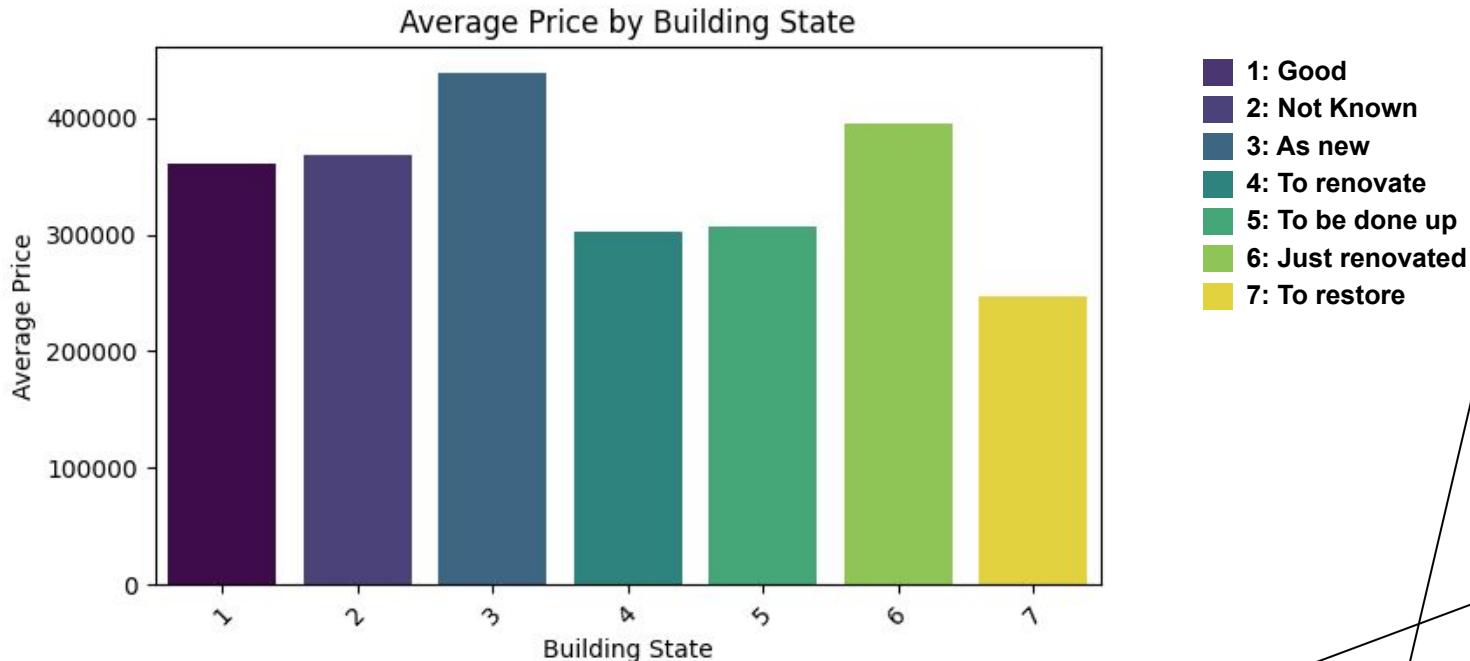
<https://www.immoweb.be/en/classified/apartment/for-sale/saint-gilles/1060/20200715>

€345,000 | 119 m² | 3 Bedrooms

Newly renovated townhouse with equipped kitchen, terrace and garden.



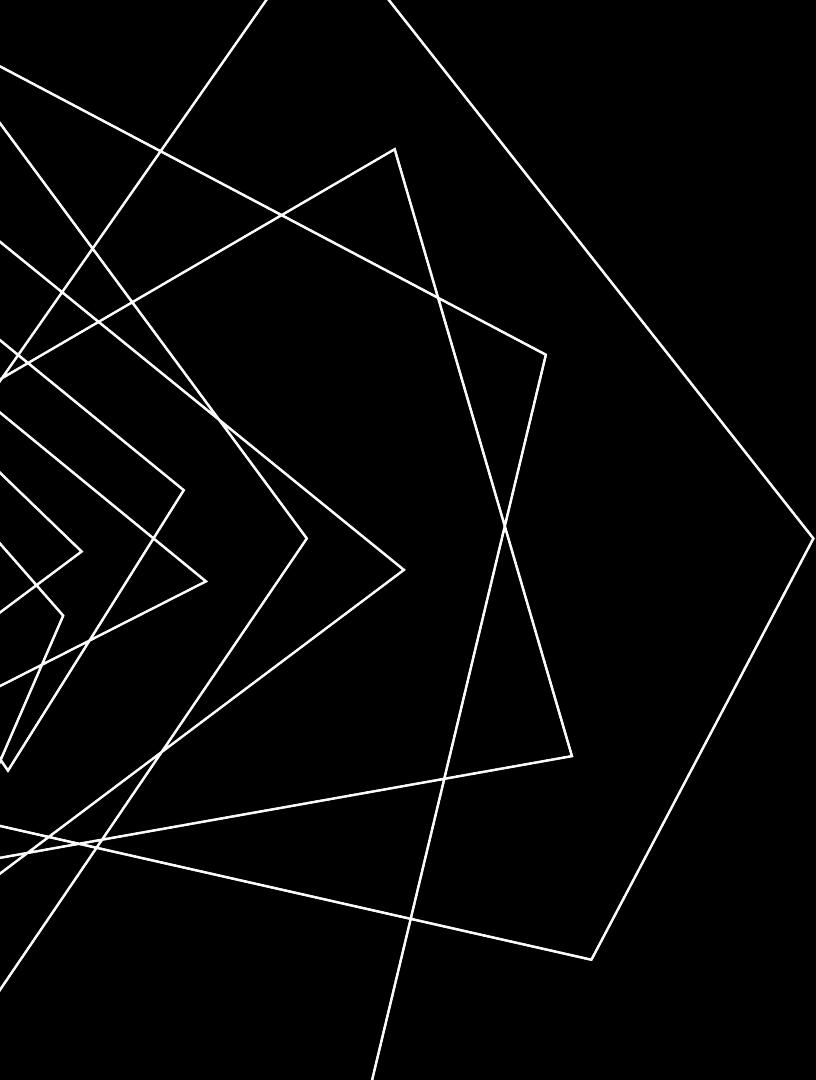
BUILDING STATE VS. PRICE





CONCLUSION

This analysis assists decision-makers in prioritizing locations based on investment potential and affordability.

An abstract graphic on the left side of the slide consists of several thin, white, intersecting lines forming various geometric shapes like triangles and parallelograms. The lines are mostly straight, with some slight curves, and they overlap each other to create a sense of depth and complexity.

THANK YOU!

Questions?