

Challenge data analysis

Axelle Paquet, Abdellah El Ghilbzouri, Melvin Leroy.

BeCode Liège.

Introduction

- Data cleaning/managing;
- Data analysis;
- Data interpretation.

Data cleaning/managing

Data cleaning/managing

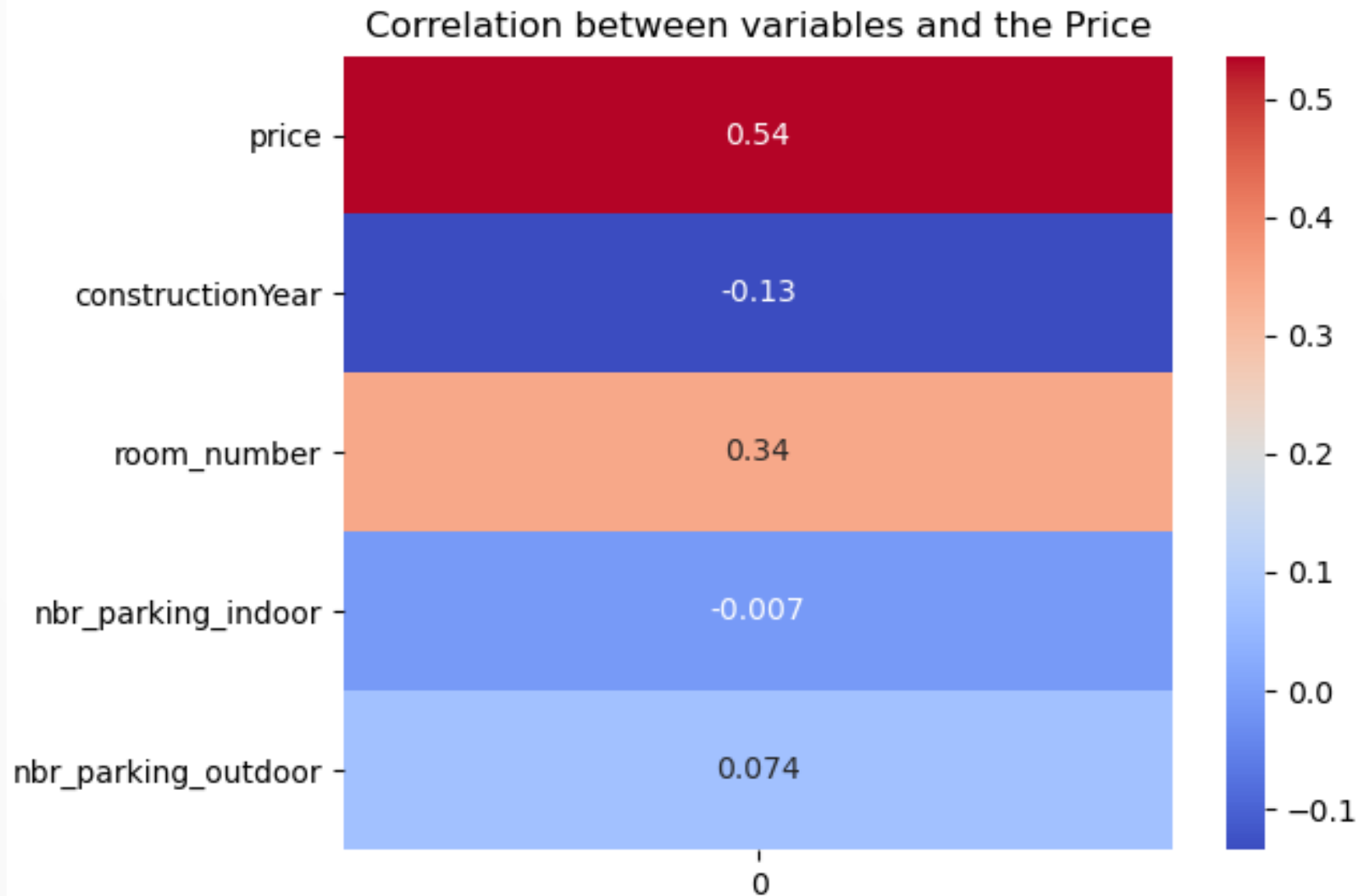
- Rename and drop some columns;
- Delete rows with NAN in prices and duplicated rows;
- Change type some data and create columns from others;
 - Price into floats;
 - Region classification;
 - Province classification;
 - Get all the cities' names the same.
- Create one hot encode columns.

Data analysis

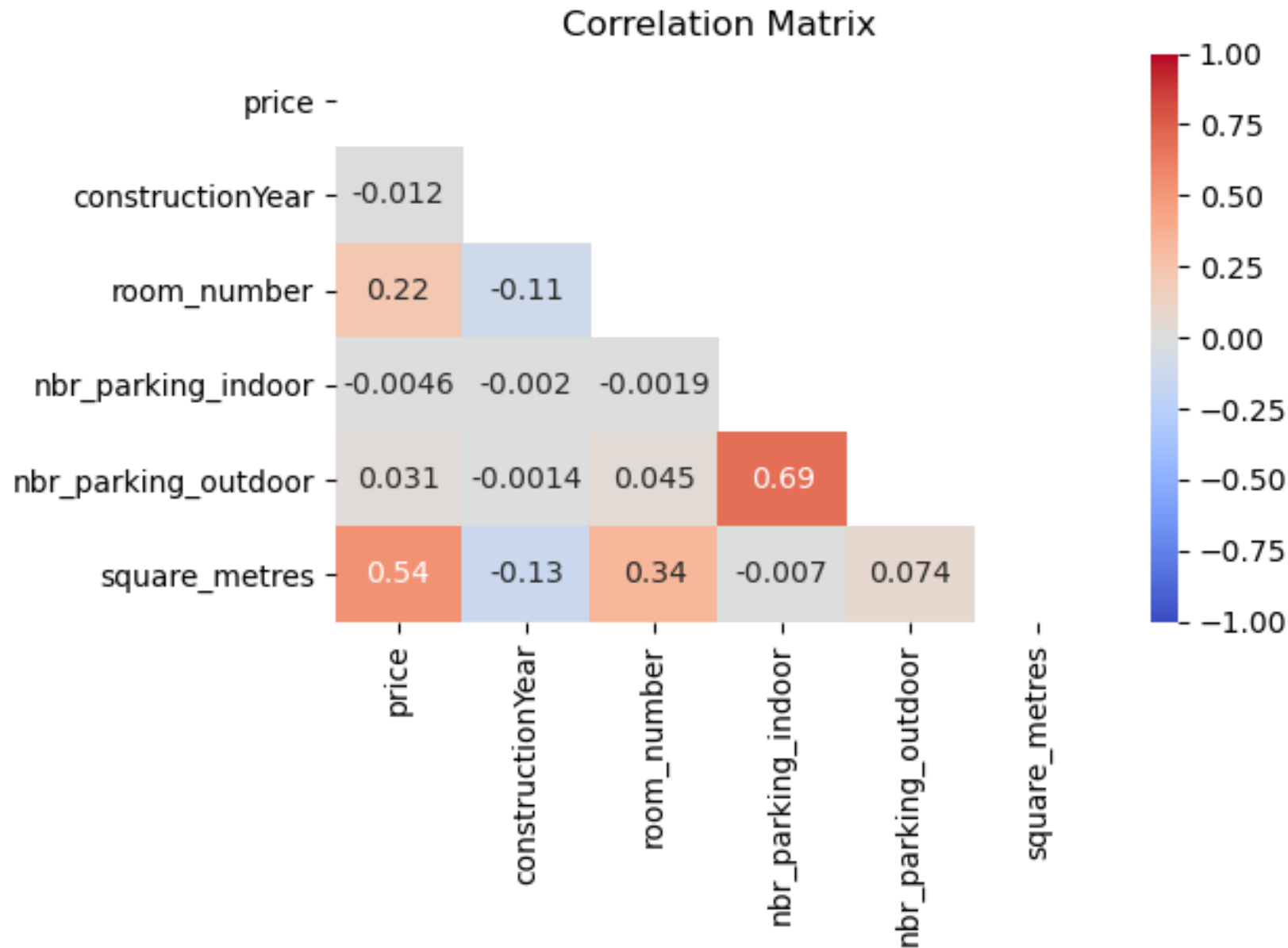
Data analysis

- The target variable:
Price per square meter
- Number of data:
shape = (19855, 73)
=> 1 449 415 values
(/!\ one hot encoding)

Correlation between the variables and the target.



Correlation matrix



Variables with the greatest influence on the target.

the variables with the greatest influence on the target variable:

0

price	0.536121
room_number	0.342929
nbr_parking_outdoor	0.074067

Variables with the least influence on the target

the variables with the least influence on the target variable :

0

constructionYear -0.134425

nbr_parking_indoor -0.007005

Qualitative variables

- Number of quantitative variables:

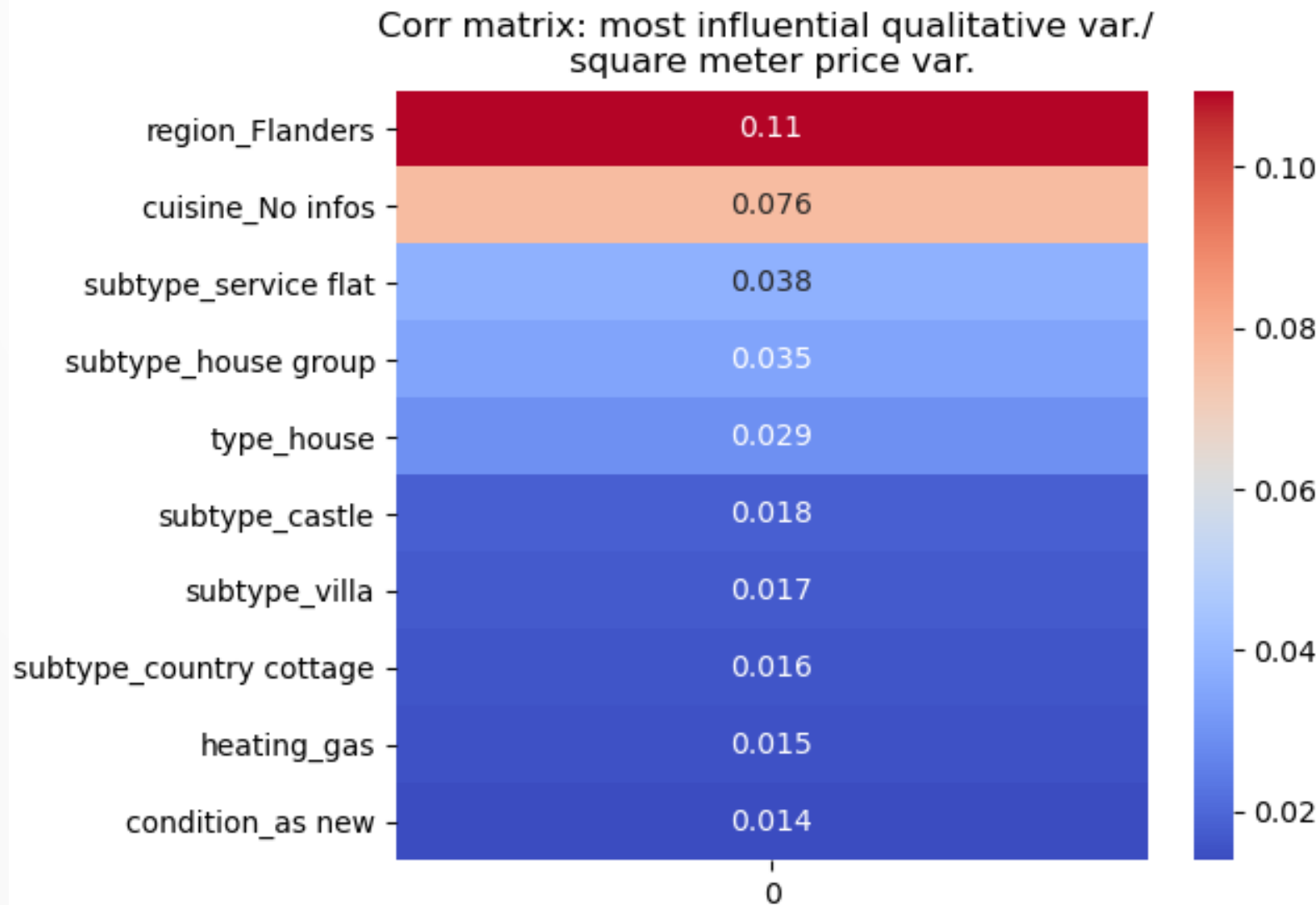
At the start : (19855, 73)

```
number of rows and columns in the quantitative data matrix  
(19855, 6)
```

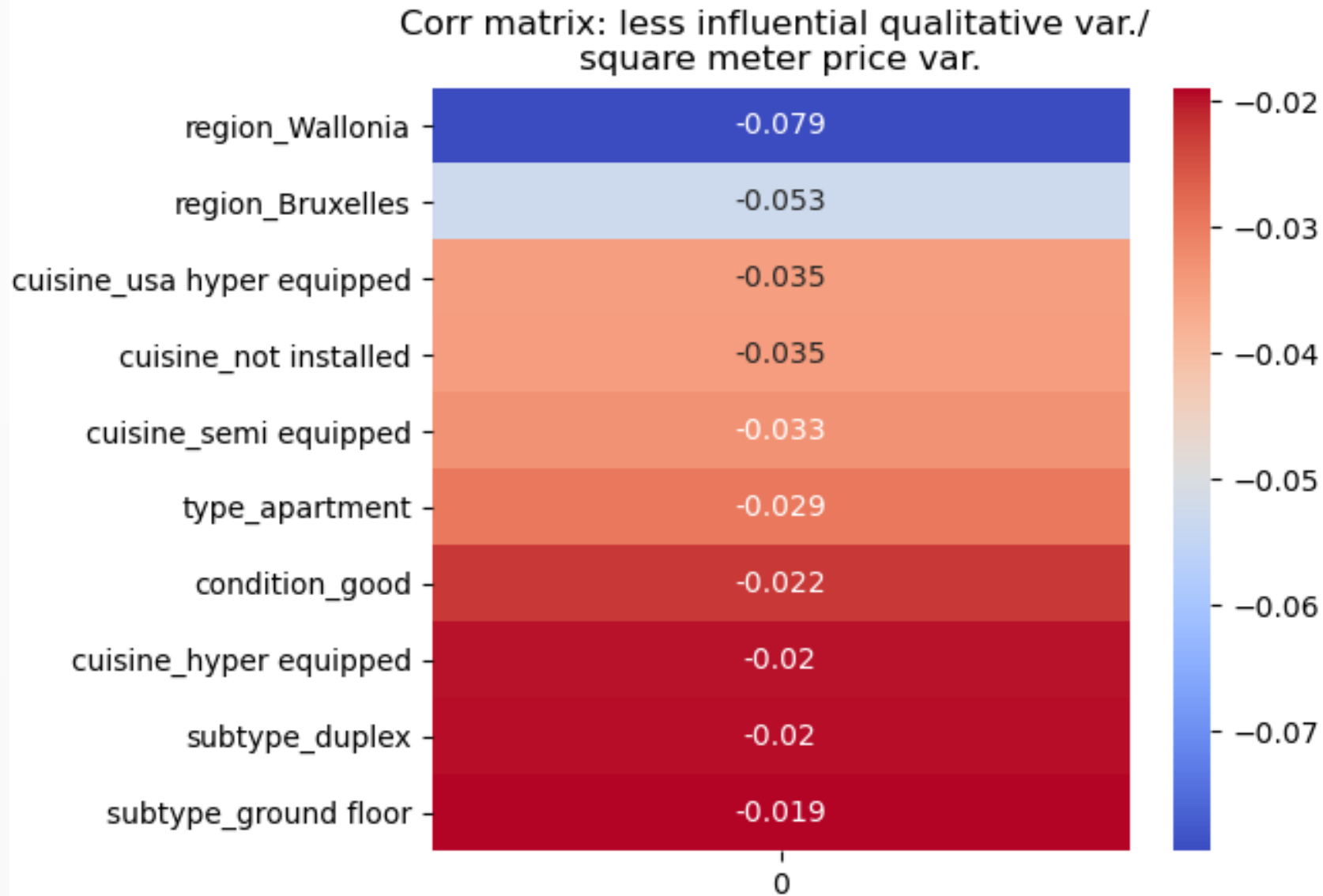
- Qualitative variables transformation:

```
number of rows and columns in the qualitative data matrix  
(19855, 53)
```

Most influential qualitative variables



Least influential qualitative variables



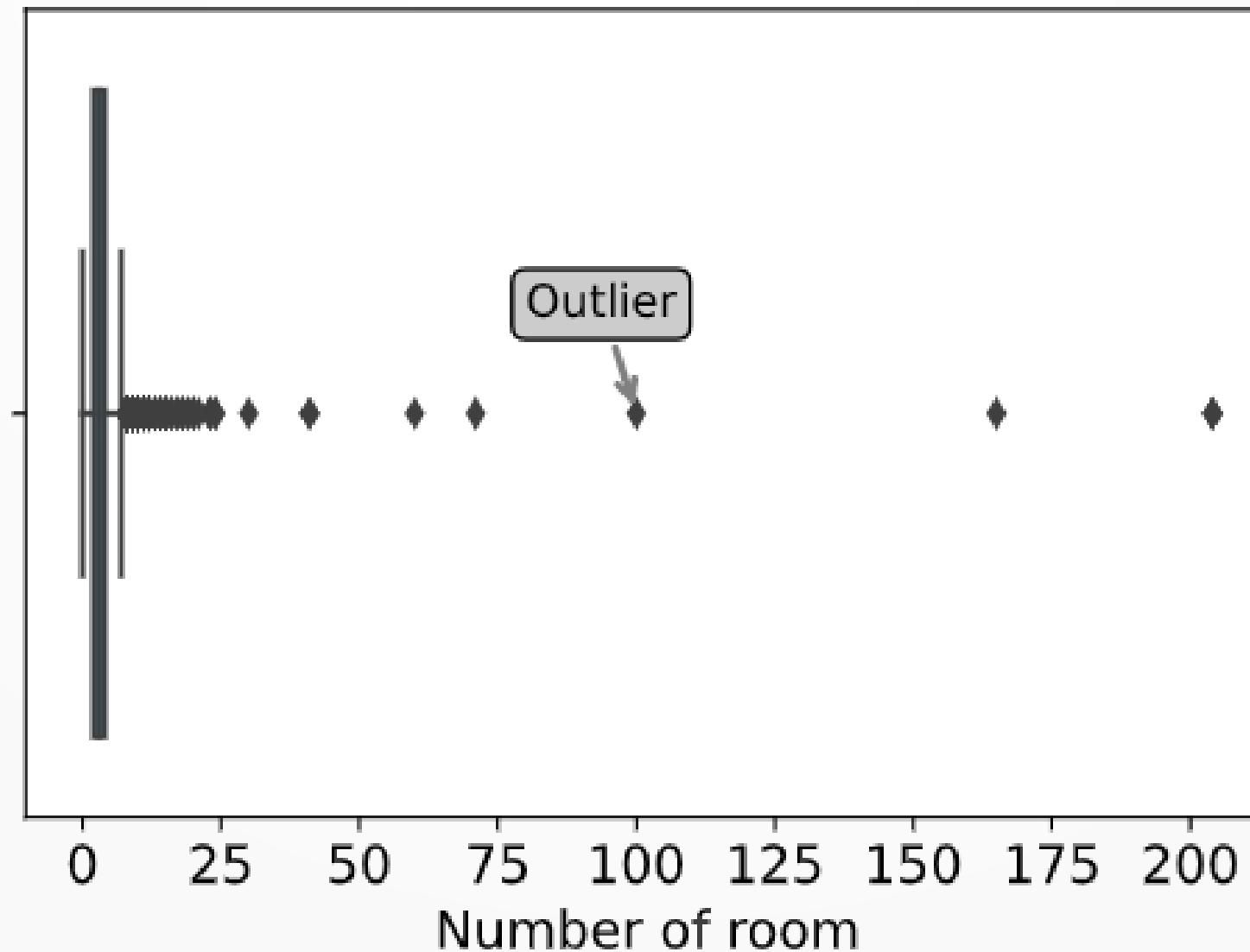
Missing values

```
the Percentage of missing values per column in the quantative variables :  
price                0.000353  
constructionYear     0.000201  
room_number          0.000101  
nbr_parking_indoor    0.645379  
nbr_parking_outdoor   0.722085  
square_metres         0.000000  
dtype: float64
```

Data interpretation

Data interpretation

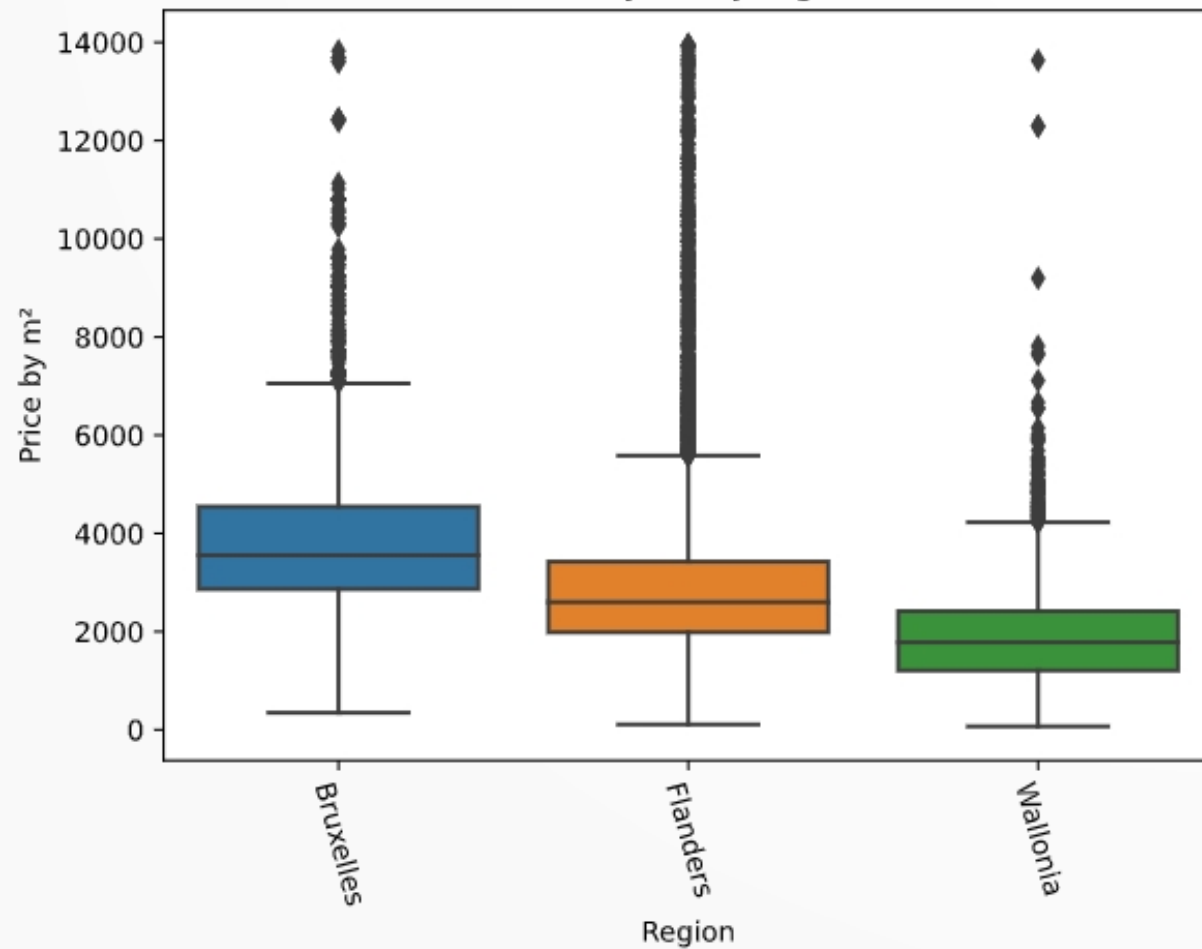
Distribution of room number



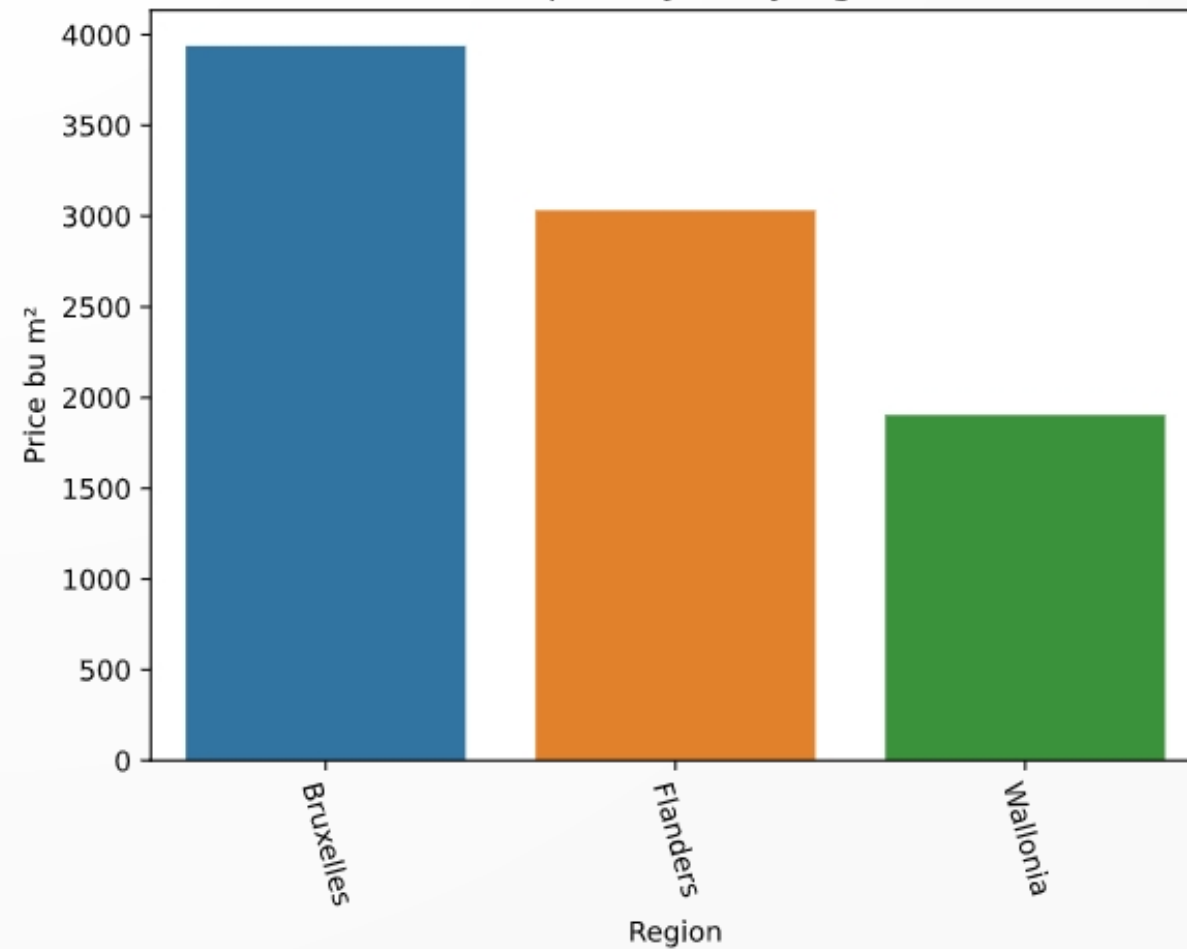
Data interpretation

What are the most and less expensive regions in Belgium?

Price by m² by region



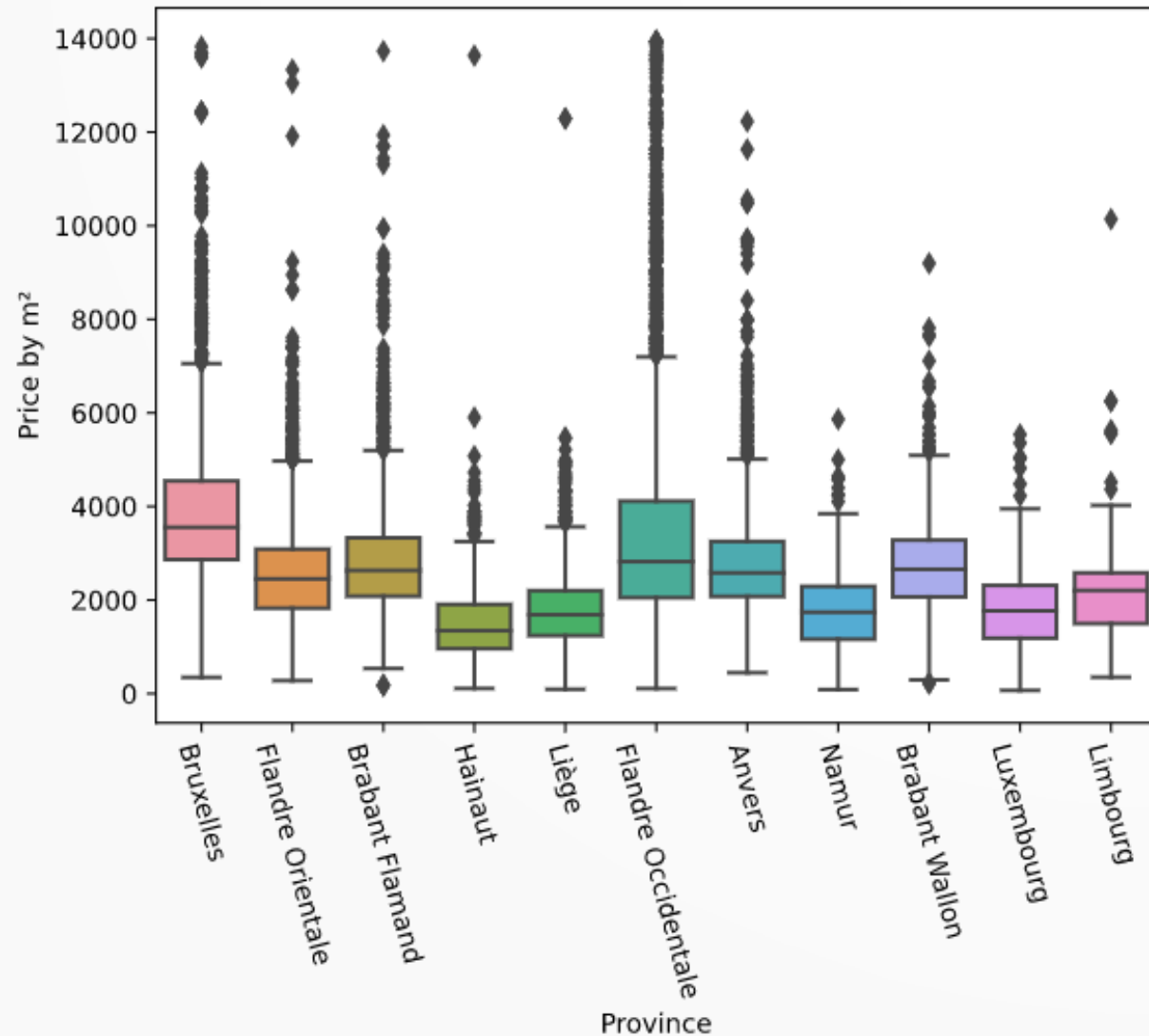
Mean price by m² by region



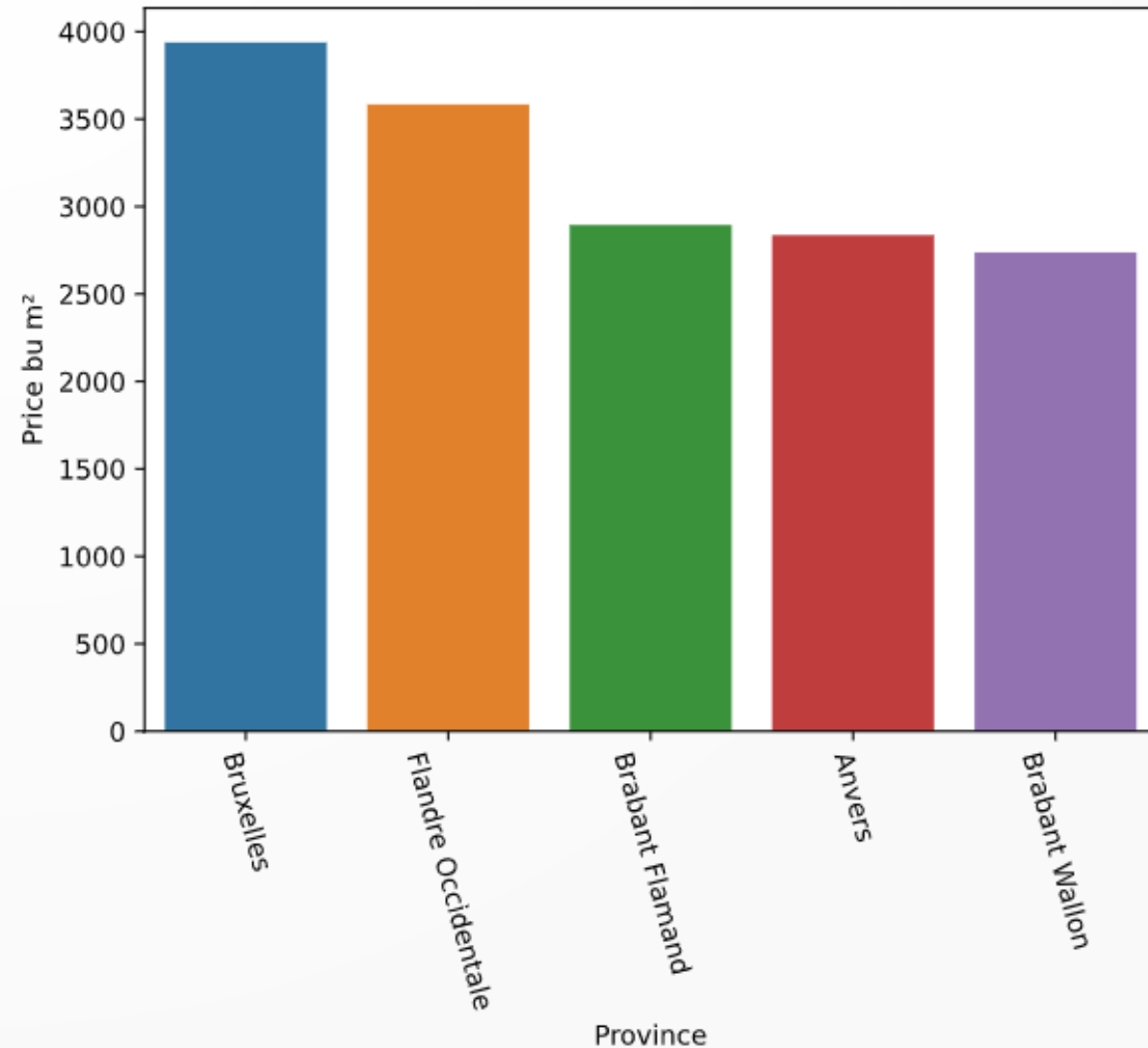
Data interpretation

What are the most and less expensive provinces in Belgium?

Price by m² by province

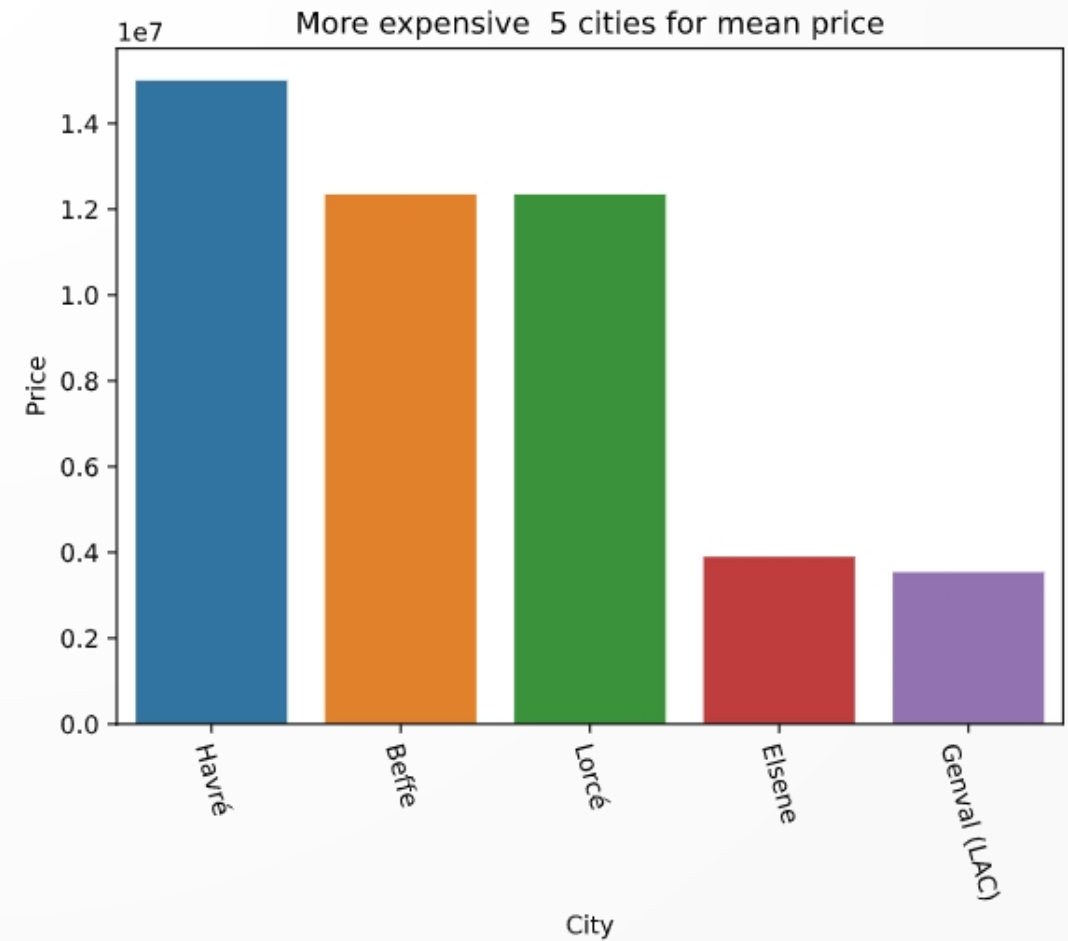
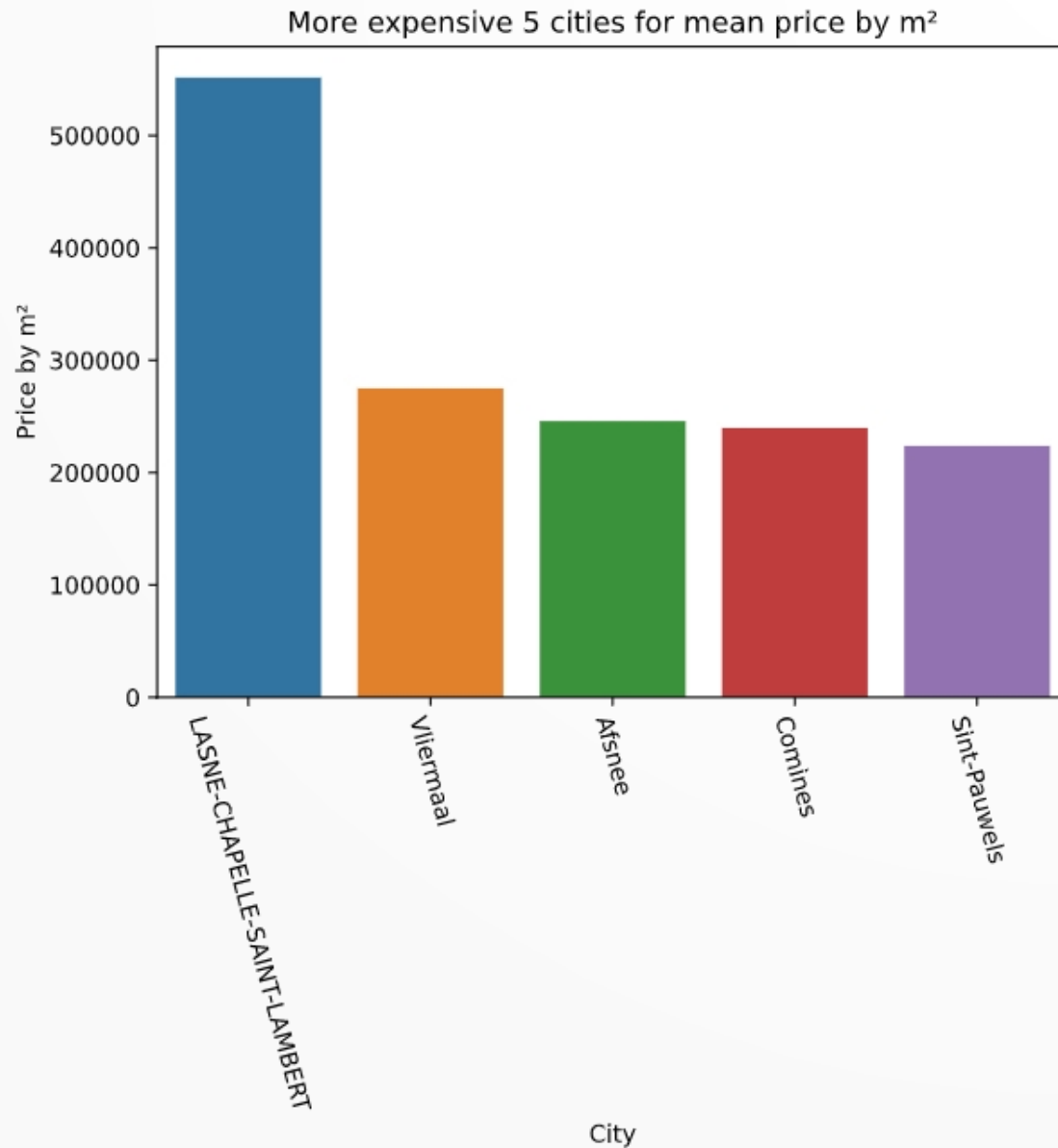


Mean price by m² by province



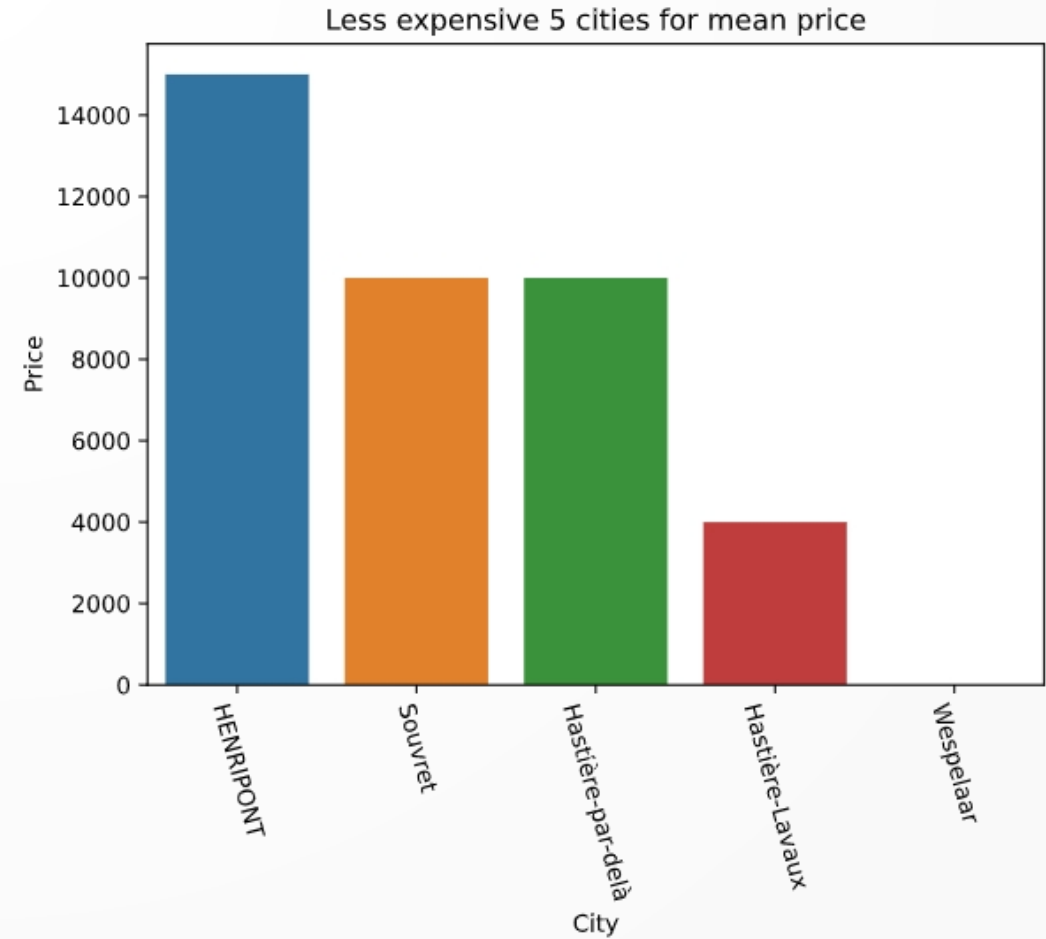
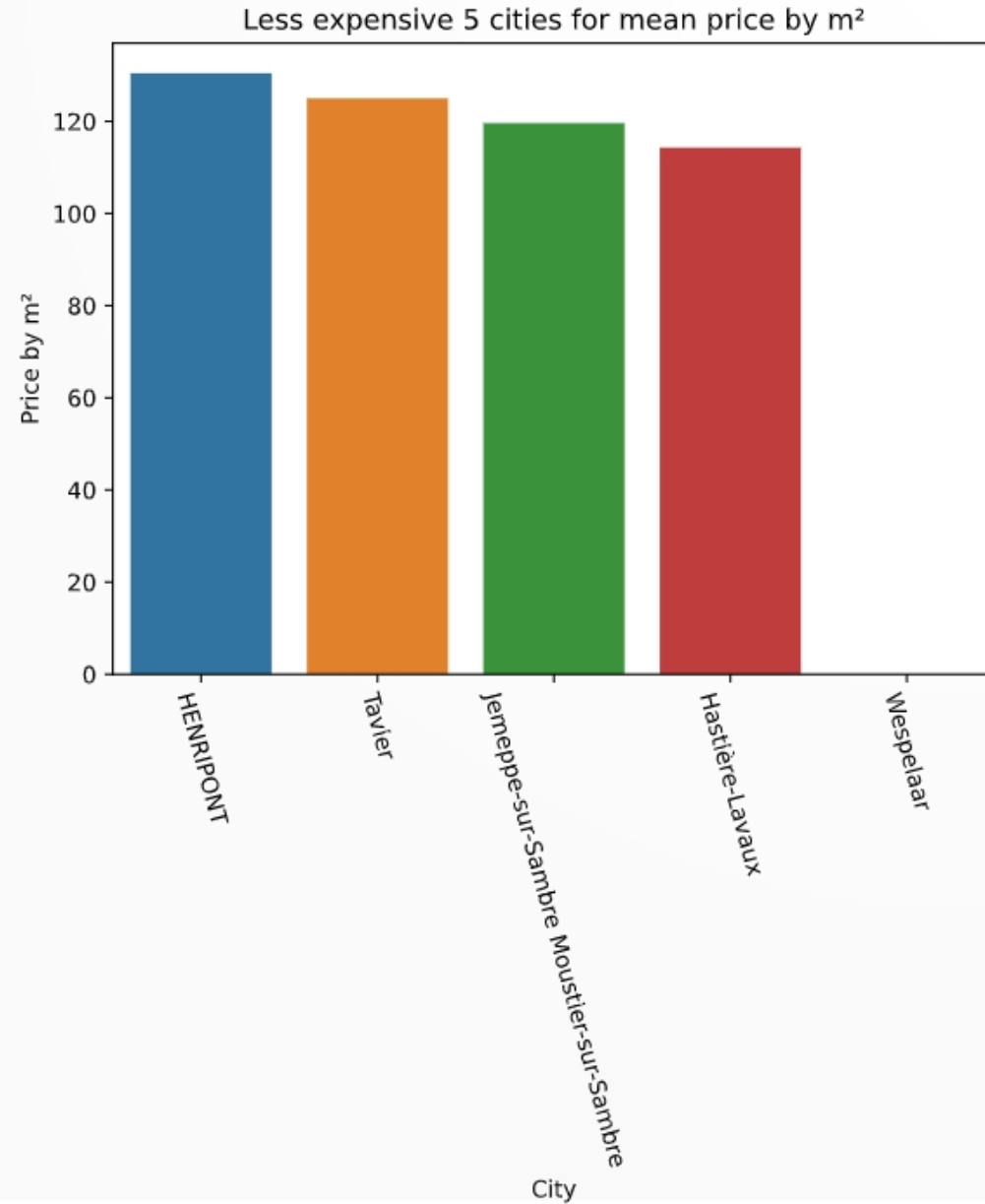
Data interpretation

What are the most expensive municipalities in Belgium?



Data interpretation

What are the less expensive municipalities in Belgium?



Thanks for your attention !

Any questions ?