ImmoWeb Data Exploration
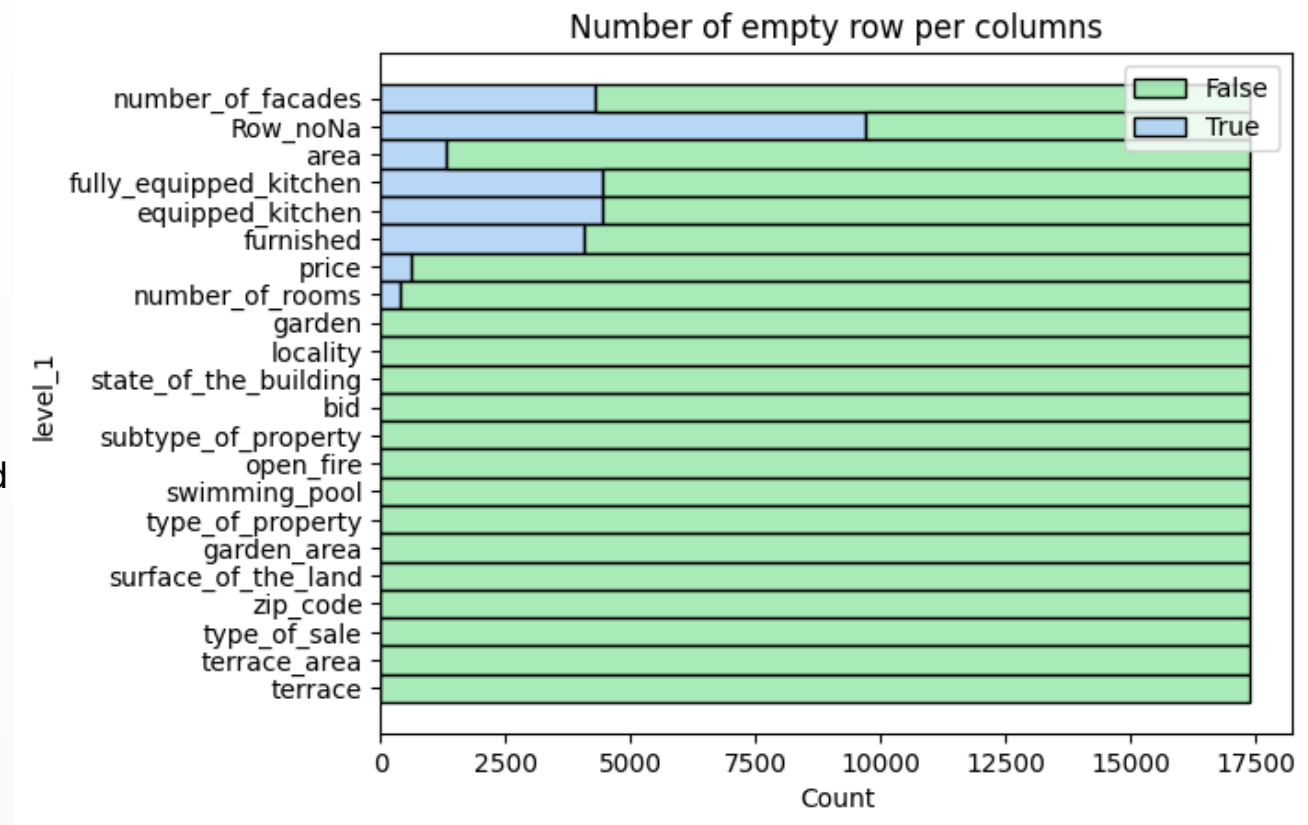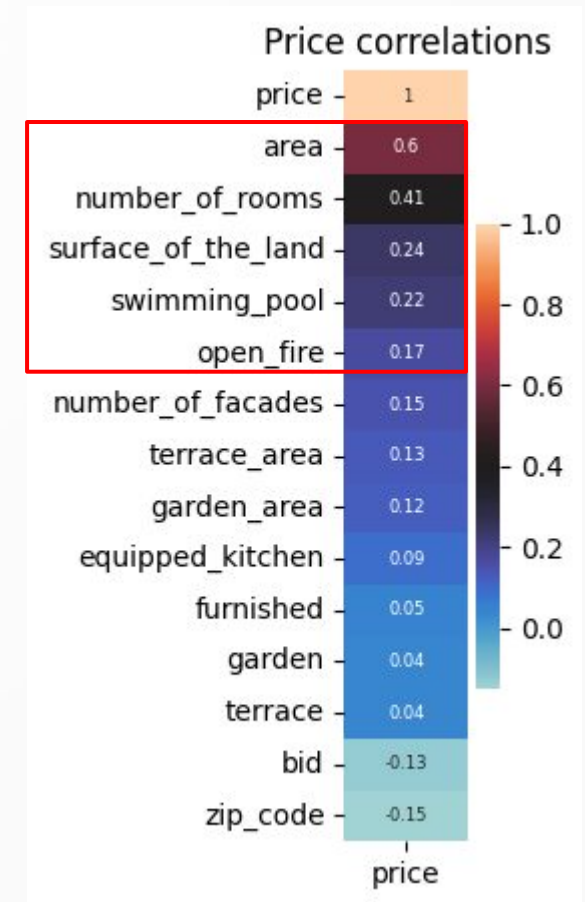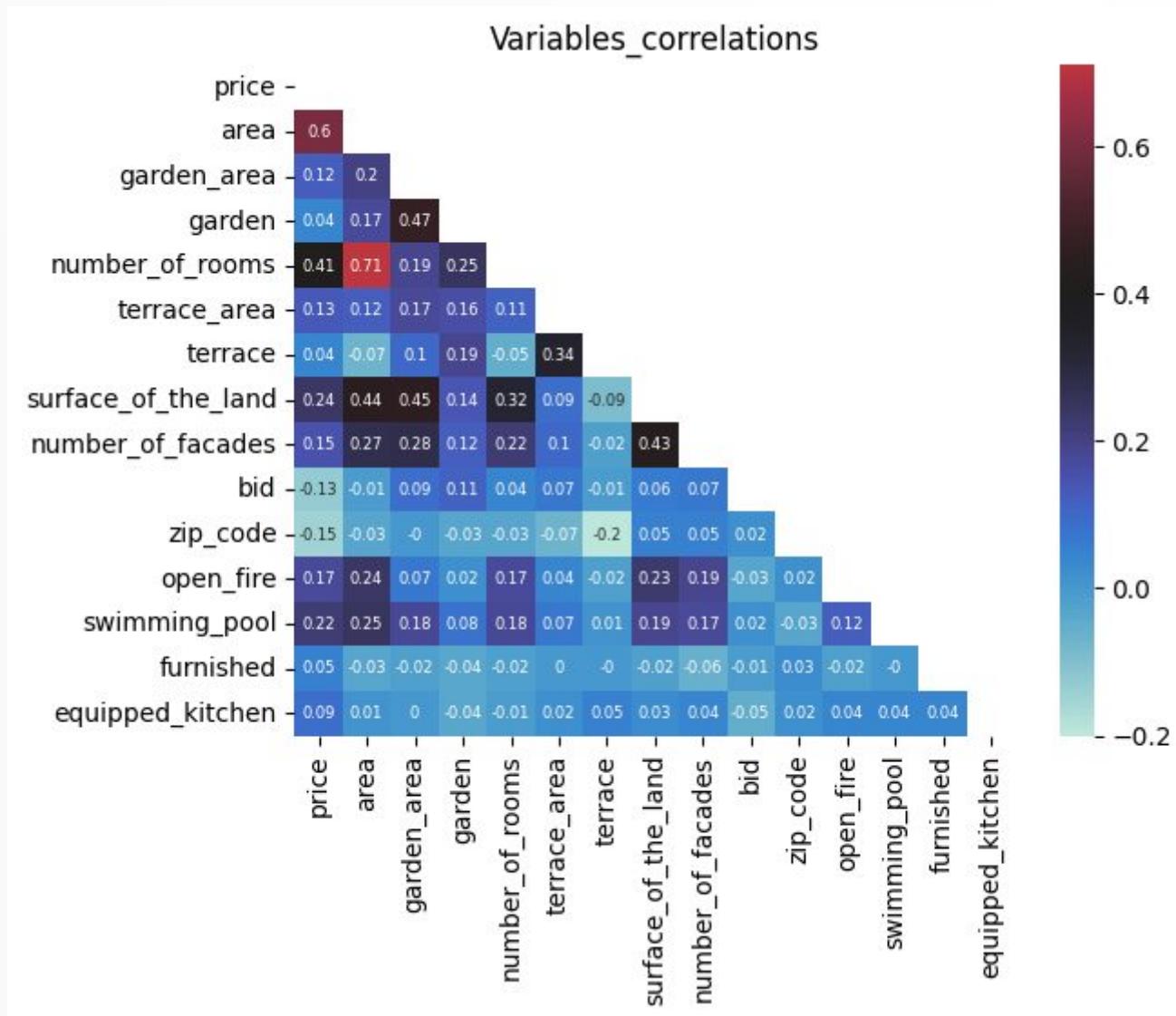
Tableau Team:
Isidora, Louis and Julien
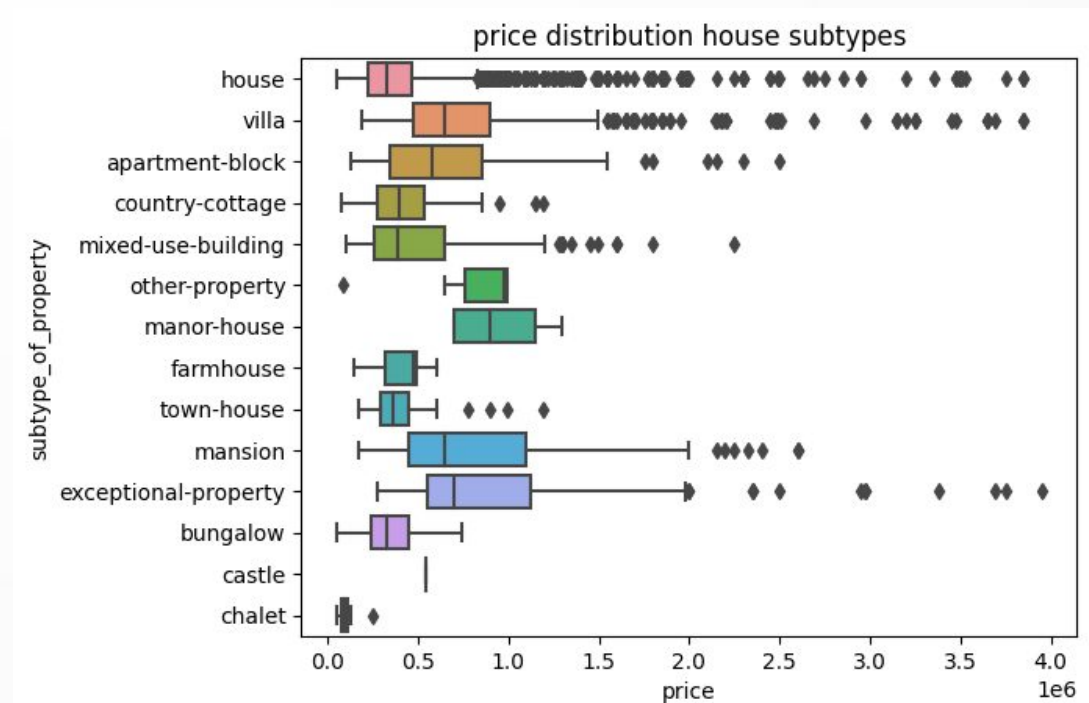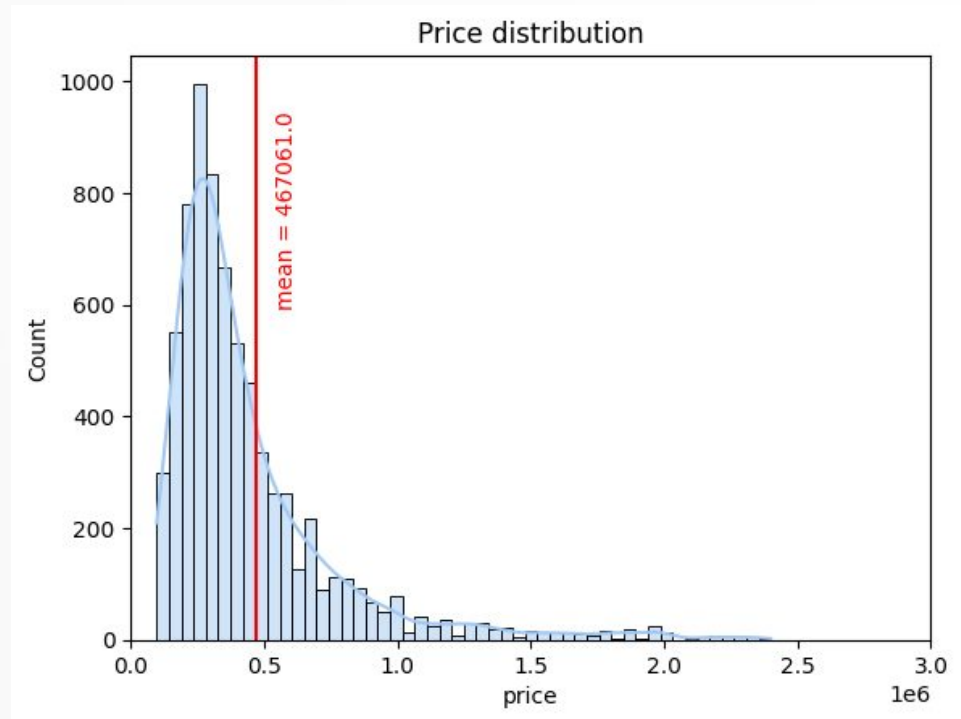
Data cleaning

- Data frame with NaN (17373, 21)

  – Assumption that for some variables if NaN = none

    • garden_area

    • terrace_area

    • swimming_pool

    • surface_of_the_land

    • open_fire

- Data frame without NaN (7688, 21)

- Removing of outliers (tails) (7493,21)

  – The percentile 0.5% en 99.5% have been removed

    • Price

    • Area

    • garden_area

    • number_of_rooms

    • terrace_area

    • surface_of_the_land

Number of empty row per columns

Variables correlation

Variables_correlations

Price correlations

Price distribution

- Most expensive municipalities in BE per meter square.
    1. Knokke   10469 €
    2. Belsele   10359 €
    3. Heist   6172 €
    4. Sterrebeek   5097 €
    5. Leuven   4876 €
    6. Ixelles   4792 €
- Cheapest municipalities in BE per meter square.
    1. Doische   466 €
    2. Vance   643 €
    3. Paliseul   650 €
    4. Erquelinnes   704 €
    5. Stoumont   750 €
    6. Lierde   756 €
- Price ranking among Belgium regions per meter square.
    - Brussels   3743 €
    - Flanders   3142 €
    - Wallonia   1965 €

- Next step to reduce the prices variables.
  - Remove :
    - Furnished, equiped_kitchen.
  - Reduce the classes for some variables:
    - Regroup Subtypes
  - Improve some variables
    - Number of facade (to see with subtype en m2)

nbr of facades VS subtypes