IMMO ELIZA:
DATA ANALYSIS

INTRODUCTION:
TEAM & PROJECT

The Team:
Kristin Nuyens
Nancy Van den Steen
Bryan Maina
Anna Lalova

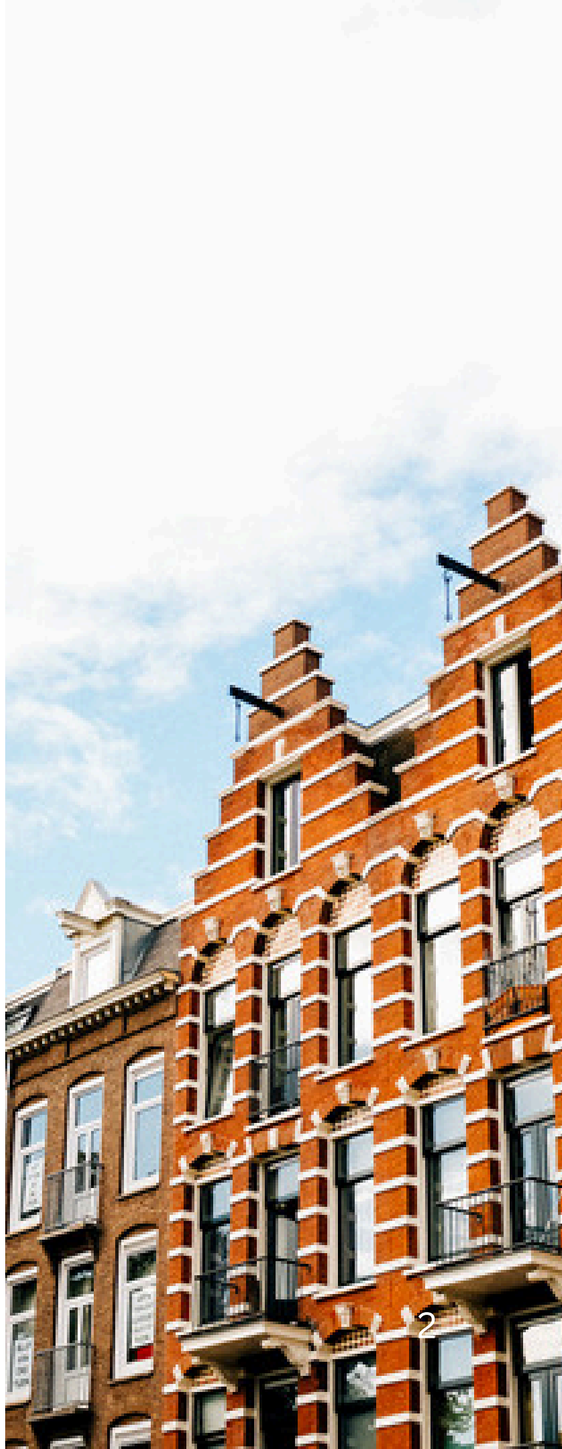
The Project:
ImmoEliza wants to become market leader in Belgian real estate
Strategic focus: improving accuracy & speed of property valuation

Our task:

- Build predictive model for property prices
- Data Analysis (N2) deliverable: understanding dataset before modeling

Guiding questions:

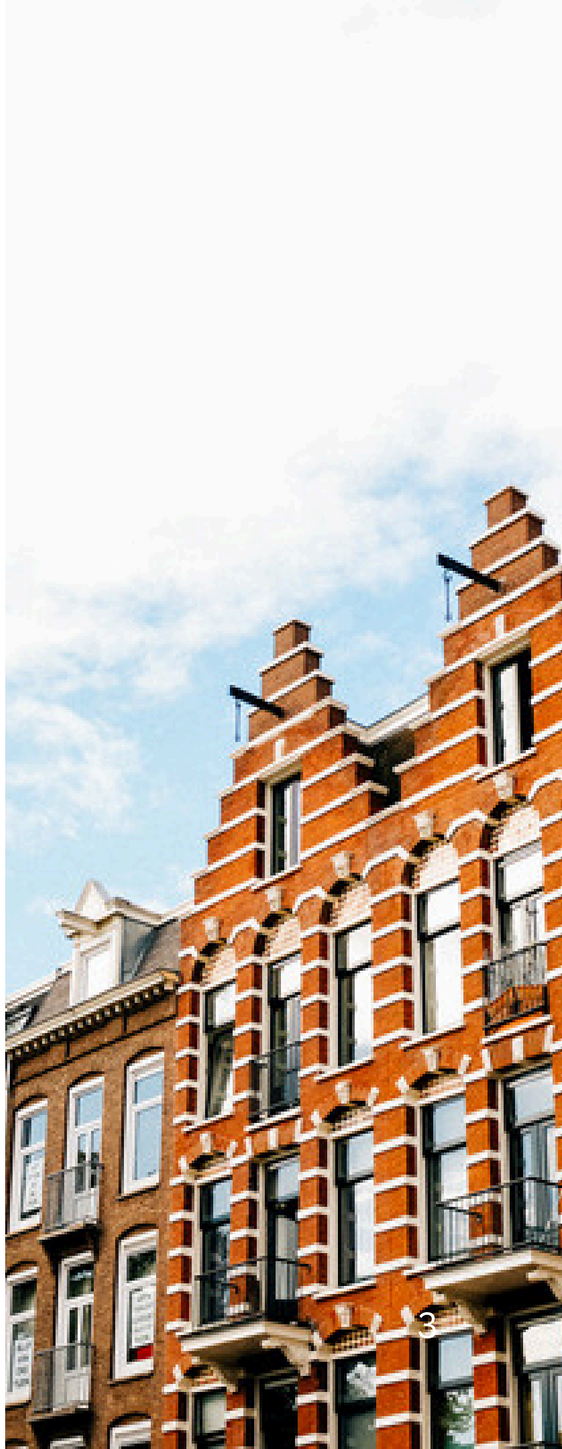
1. What notable insights emerge from Belgian real estate data?
2. Which variables have strongest influence on property prices?



IMMO DATA:
SCOPE & STRUCTURE

Dataset’s basic characteristics

- total number of observations: 75508
- total variables: id + 29
- data types:
  - quantitative continuous 10
  - quantitative discrete 3 (zip code, nmbrs)
  - nominal/categorical 9 (e.g. subproperty type)
  - binary 7 (flags)
- target variable: PRICE



DATA QUALITY ASSESSMENT & PROCESSING

Key Insight

– Missingness –

- Not random but clustered by property type, subproperty type, region, & EPC
- Some missing values carry strong predictive information (e.g., land surface)
- Preserving or flagging missing values helps models capture price signals

→ Missing values themselves can help predict price

- The dataset is ready for analysis while preserving predictive signals
- Any subsequent modelling can account for meaningful missingness without losing important information

[UNDER QUESTION] DISTRIBUTION DIAGNOSTICS:

UNUSUAL DATA POINTS (OUTLIERS)

IF WE KEEP THIS SLIDE, TO BE MADE LESS TECHNICAL:

Introduce how you detected outliers in key numerical variables (skewness coefficients and checked also with IQR). Show which variables exhibited strong skewness (price, surface_land_sqm, total_area_sqm, garden_sqm, terrace_sqm, terrace_sqm, nbr_frontages, nbr_bedrooms) and why that matters for modeling (alters the correlation and regressions results). Explain your chosen remedies: top-capping extreme values, log transformations. Clarify how these adjustments affect model stability and interpretability.

XXXXXXXXXXXXX

XXXXXXXXXXXXXX

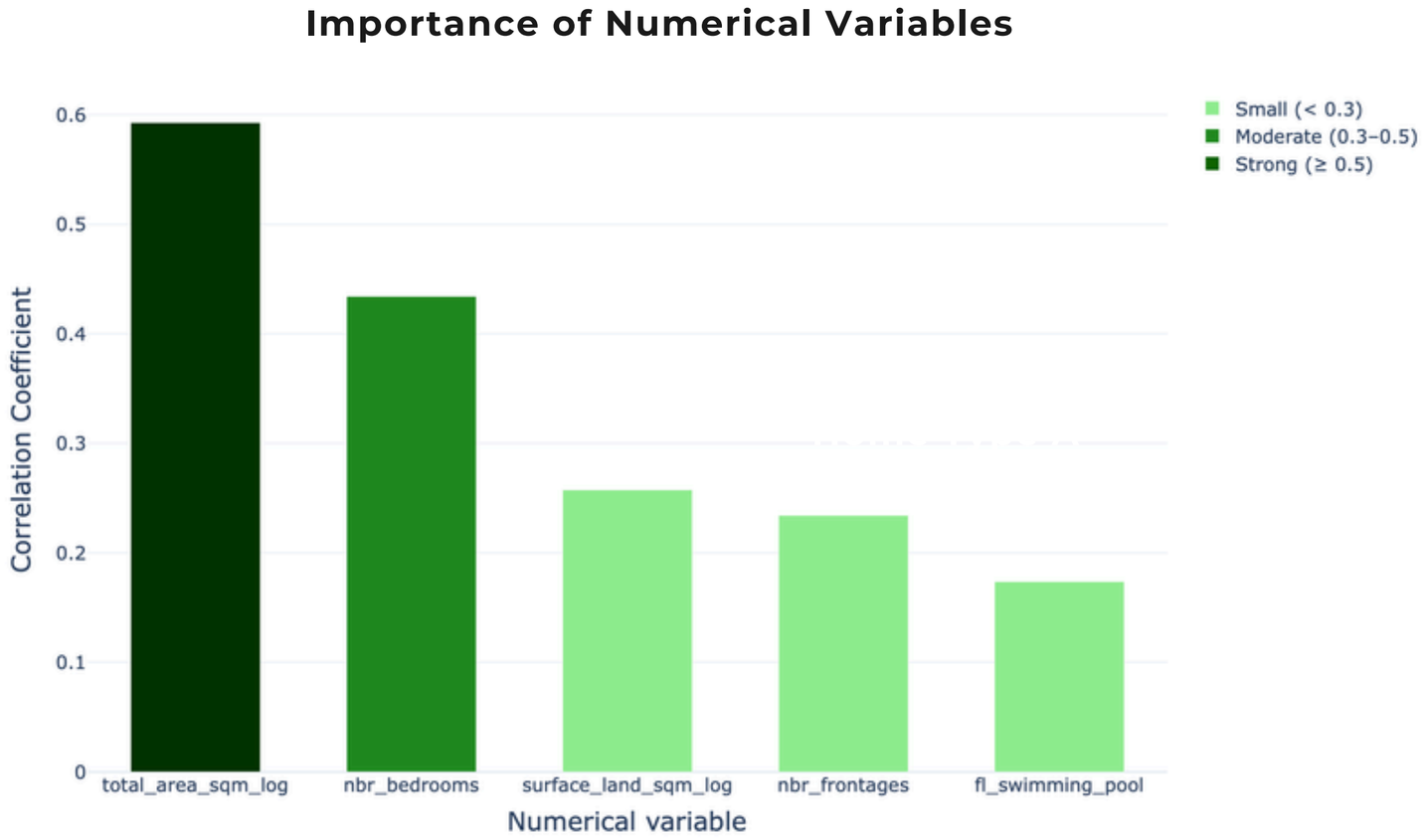
PRICE LANDSCAPE:
SUMMARY & VISUAL INSIGHTS

xxxxxxxxxxxxx

[Provide descriptive statistics for the price variable (mean, standard deviation,). Add visual comparisons across categorical groups such as regions or subtypes (e.g boxplots, histograms, or density plots). The goal is to reveal structural differences in price distributions and identify meaningful segmentation patterns.]

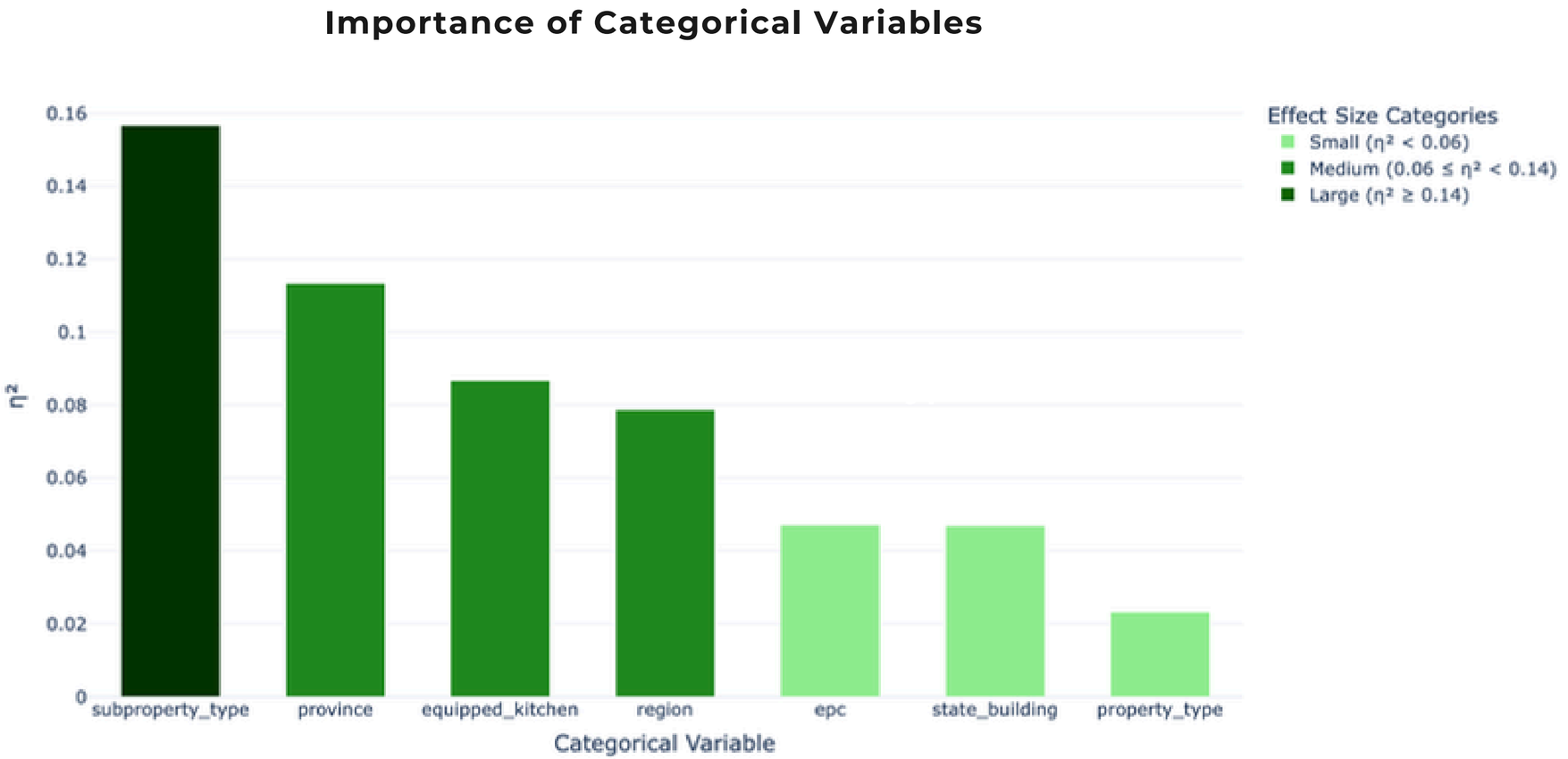
UNDERSTANDING DRIVERS OF PRICE:

NUMERICAL PREDICTORS



- Living area – highly valued as it provides core usable space.
- Bedrooms – matter as they determine practical capacity for families.
- Land size & façades – moderately important, adding value in suburban/detached markets.
- Interior space – prioritized over exterior/structural features, reflecting buyer preference for functional living areas.

UNDERSTANDING DRIVERS OF PRICE:
CATEGORICAL PREDICTORS



- Sub-property type: strongly impacts price, capturing high market segmentation;
- Province vs Region - there are stronger differences in the provinces (e.g. specific local conditions);
- Equipped kitchen - possible proxy for renovation level and quality;
- Energy performance certificate (epc)
- State of the building - possible self-reported bias;
- Property type - the broad distinction explains little variance;

UNDERSTANDING DRIVERS OF PRICE:
WITHIN PROVINCES AND SUB-PROPERTY TYPES

Top three drivers at sub-property type level

- Apartments & Studios

Total area (very strong), **Bedrooms** (strong), Locality (large to very large), Terrace (moderate)

- Houses, Duplexes, Villas, Townhouses, Mansions

Total area (very strong), Locality (very large), **Bedrooms** (strong), Province/Region (very large for luxury types)

- Special types (Farmhouse, Penthouse, Apartment Block, Country Cottage, Castle, Chalet, Loft, Bungalow)

Total area (very strong), Locality (very large), Cadastral income (strong to very strong), Outdoor/amenity features (terrace, garden, land)

Top three drivers at province level

- Antwerp, Brussels, Limburg, Flemish Brabant

Total area (very strong), **Bedrooms** (strong to very strong), Sub-property type (very large), **Cadastral income** (strong)

- East Flanders, Liège, Hainaut, Namur, Walloon Brabant

Total area (very strong), **Bedrooms** (strong), Cadastral income (moderate to strong), Frontages/land/terrace (moderate)

- West Flanders

Total area (very strong), **Bedrooms** (strong), **Cadastral income** (very strong), Surface land (strong)

KEY INSIGHTS & RECOMMENDATIONS

Conclusions:

- Property prices in Belgium (globally and across provinces and sub-property types) are primarily driven by **living area.**
- **Sub-property type** matters mostly at global level.
- **Interior space** prioritised over exterior.
- Localisation matters.

Recommended actions:

- Tailor pricing and marketing by sub-property type.
- Highlight key interior features and area.
- Highlight province-specific amenities (e.g swimming pool in Hainaut, terraces in Brussels).
- Use province-level targeting.
- Validate self-reported features or consider objective inspections.

THANK YOU
FOR YOUR
ATTENTION