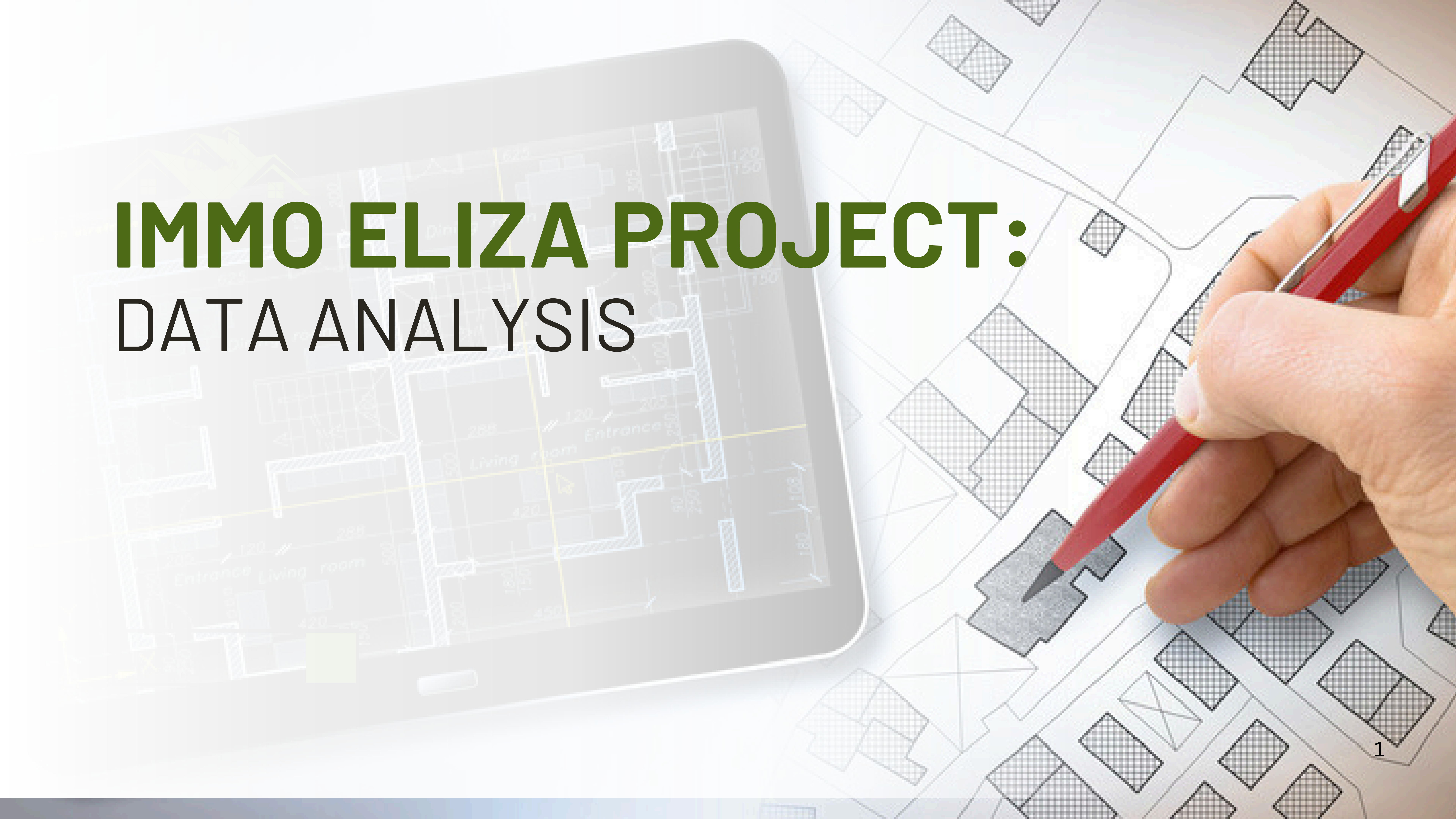


IMMO ELIZA PROJECT:

DATA ANALYSIS

INTRODUCTION:
TEAM AND PROJECT

Our Team:

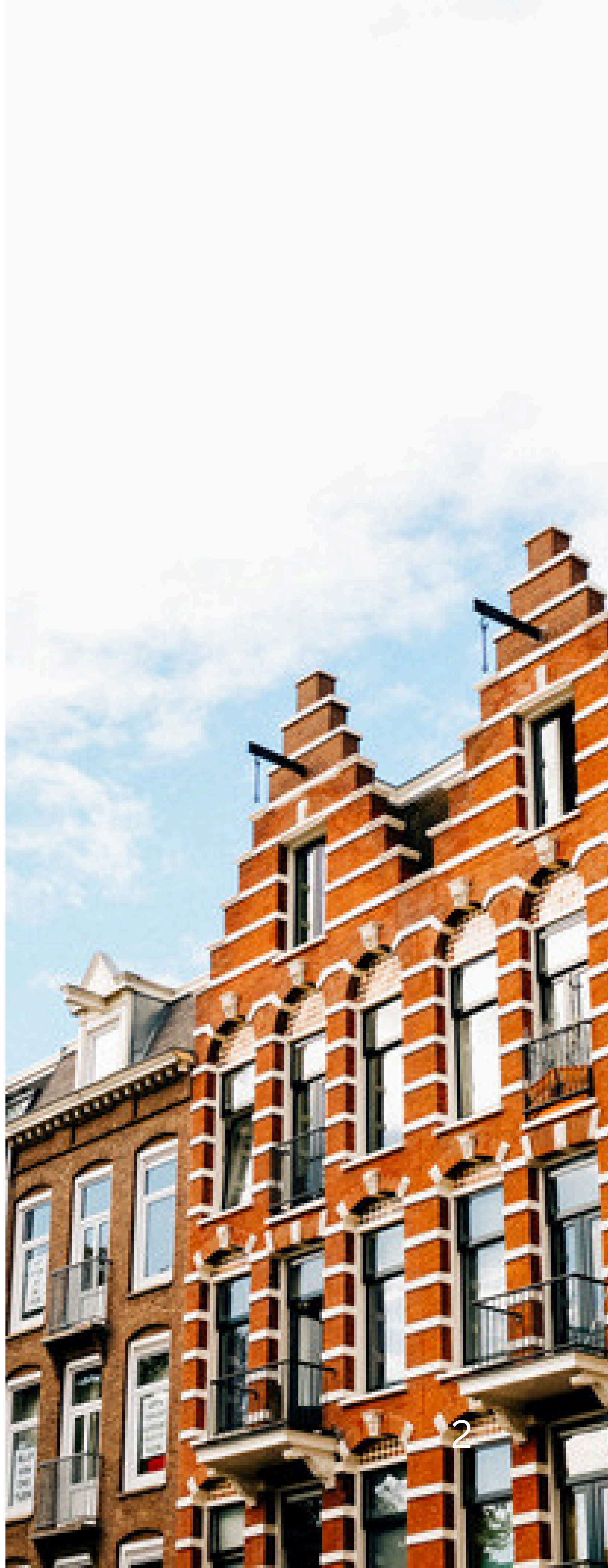
Kristin Nuyens
Nancy Van den Steen
Bryan Maina
Anna Lalova

The Project:

- ImmoEliza is working toward becoming a leader at the market
- Strengthening market position depends on more accurate and faster estimation of property value
- Our role - build a predictive model
- Deliverable N2 - Data Analysis

- Guiding questions:

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

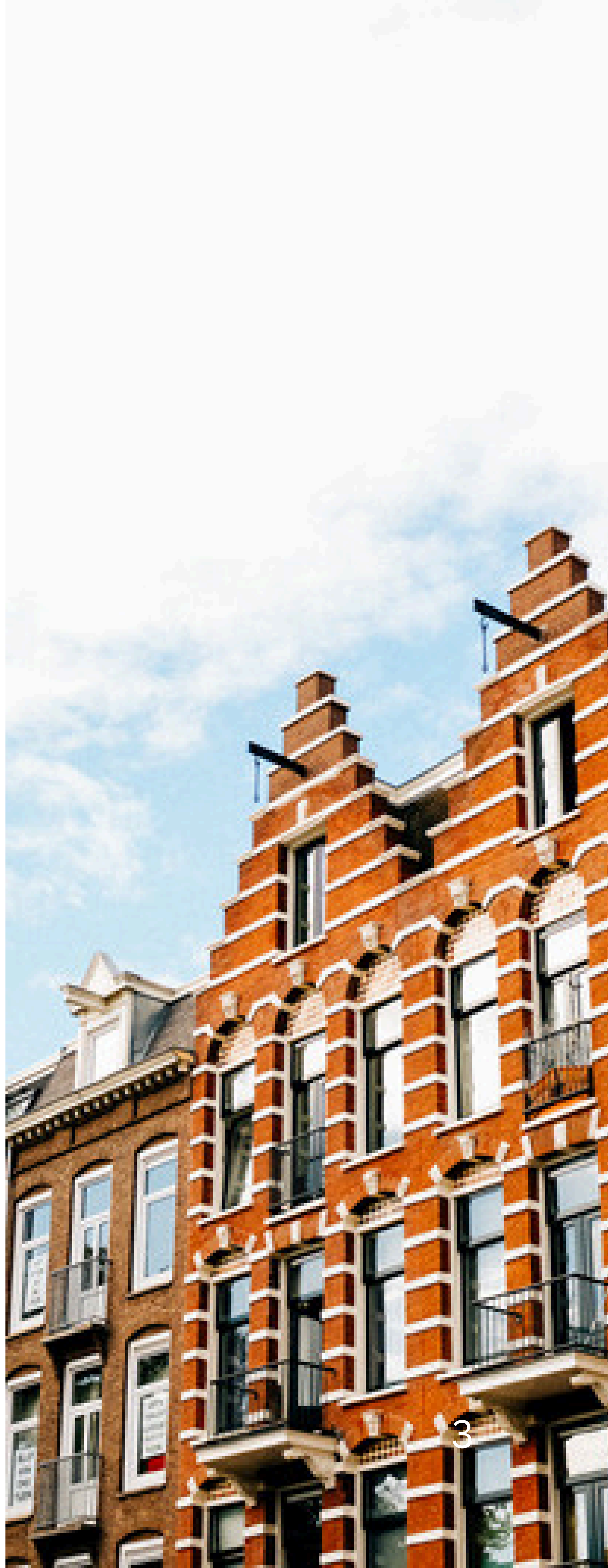


IMMO DATA:

SCOPE AND STRUCTURE

xxxxxxxxxxxxxx

[Explain the dataset’s basic characteristics: total number of observations (.), and data types (qualitative: continuous (8), discrete(3), qualitative: nominal(9), ordinal(1), binary(7)). Clarify the target variable (price). The goal is to give the audience a solid orientation before diving into analysis.]

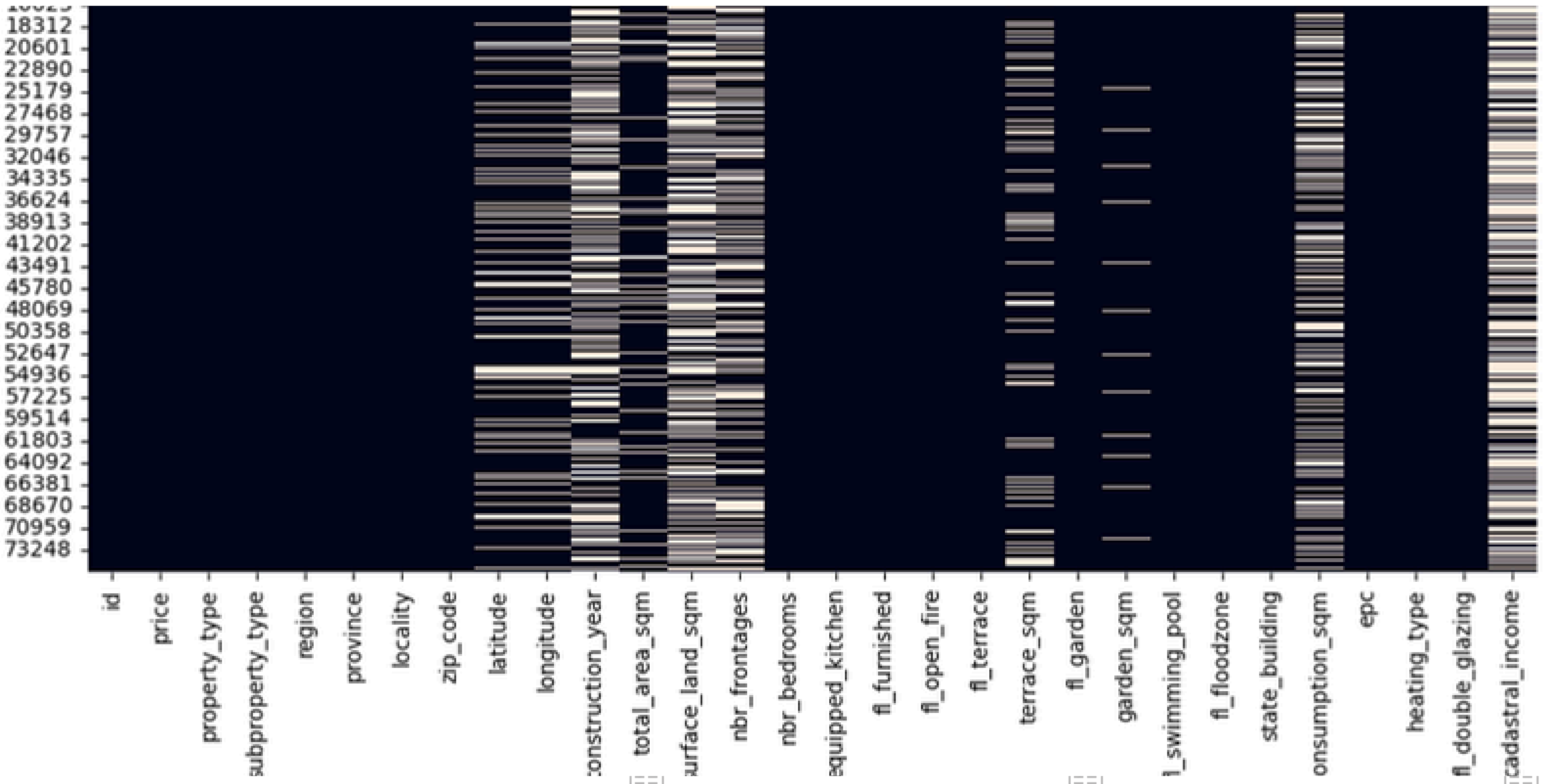


DATA QUALITY ASSESSMENT & PROCESSING

Step	Action	Why
Initial Inspection	Check column types (dtypes) and previewed rows	Correctly identifies numeric vs categorical fields
Remove Empty Rows	Drop rows entirely empty in critical columns or with missing id	Guarantees remaining rows have meaningful data
Remove Duplicates	Drop duplicate rows (excluding id)	Prevents double-counting and bias
Handle Missing Values	Convert "MISSING" strings and empty text to NaN	Ensures consistent representation of missing data
Strip Text Columns	Remove leading/trailing whitespace from all string columns	Prevents categorical inconsistencies (e.g., " A " vs "A")
EPC Mapping	Standardize energy performance certificate codes regionally; unmapped/missing labeled "MISSING"	Maintains categorical consistency without losing rows

DATA QUALITY ASSESSMENT & PROCESSING

Number of observations (rows): 75511
Number of features (columns): 30
Proportion of missing values per column (%):
cadastral_income 59.55
surface_land_sqm 48.01
construction_year 44.22
primary_energy_consumption_sqm 35.18
nbr_frontages 34.89
Variables deleted? None:
Missingness... not random (e.g. surface_land_sqm/apartment)



DISTRIBUTION DIAGNOSTICS: OUTLIERS [& SKEW]

XXXXXXXXXXX

xxxxxxxxxxxxxx

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.

XXXXXXXXXXXXXX

XXXXXXXXXXXXXXX

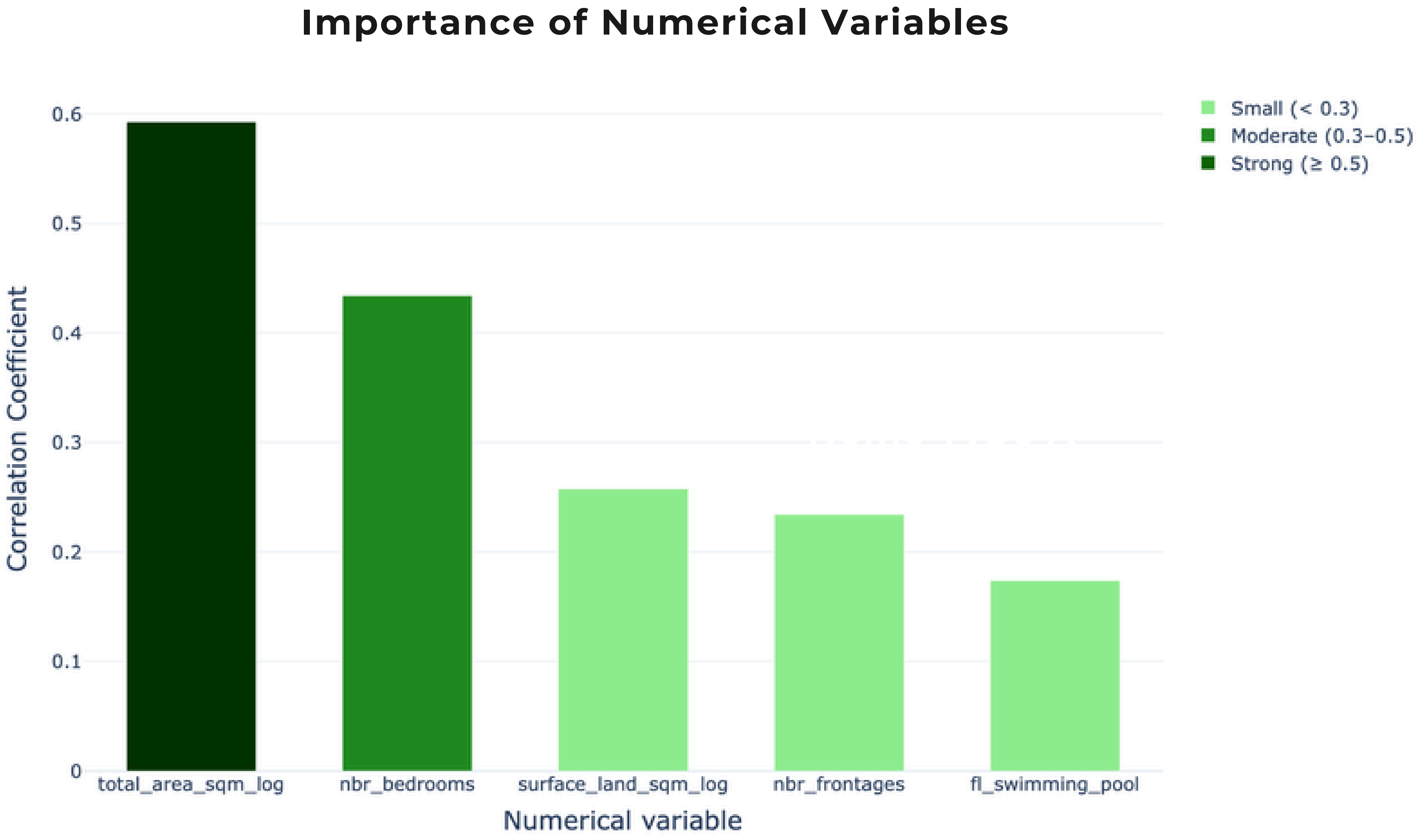
PRICE LANDSCAPE:
SUMMARY & VISUAL INSIGHTS

XXXXXXXXXXXXXX

[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

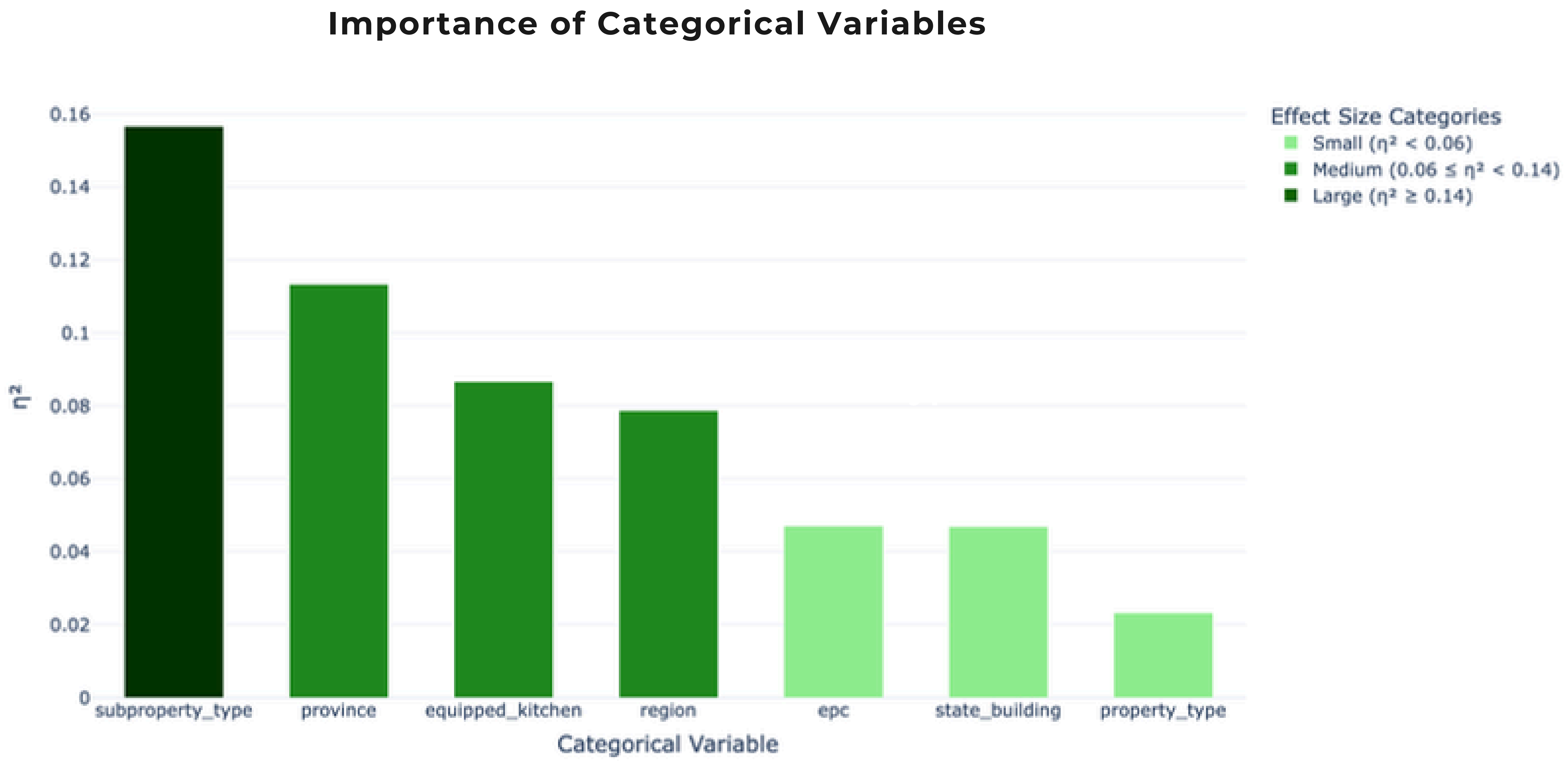


Importance of Numerical Variables

- 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 PRECITORS



- 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;
- 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 property sub-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-type, and location**.
- Sub-property type matters most globally.
- **Interior space** prioritised over exterior.
- Location (province level) matters.

Recommended actions:

- Tailor pricing and marketing by sub-property type.
- Highlight key interior features.
- Add region-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