Real Estate Price Modeling with GAMs

Fabian Hollinetz

2026-01-08

Overview

This report presents a Generalized Additive Model (GAM) for real estate price prediction. The model is built on a clean, reproducible pipeline combining dbt-based feature engineering and a Python GAM implementation.

Experiment Configuration

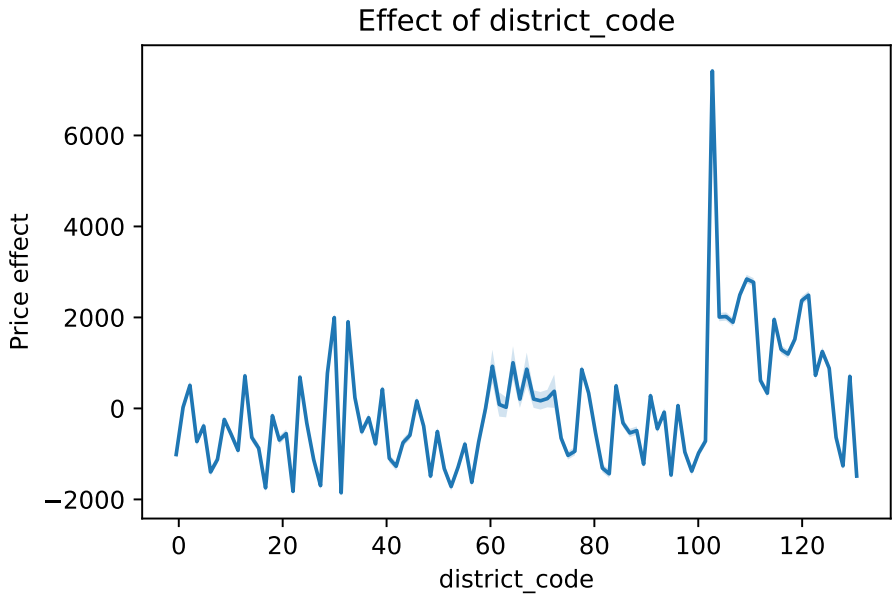
This report corresponds to the baseline experiment:

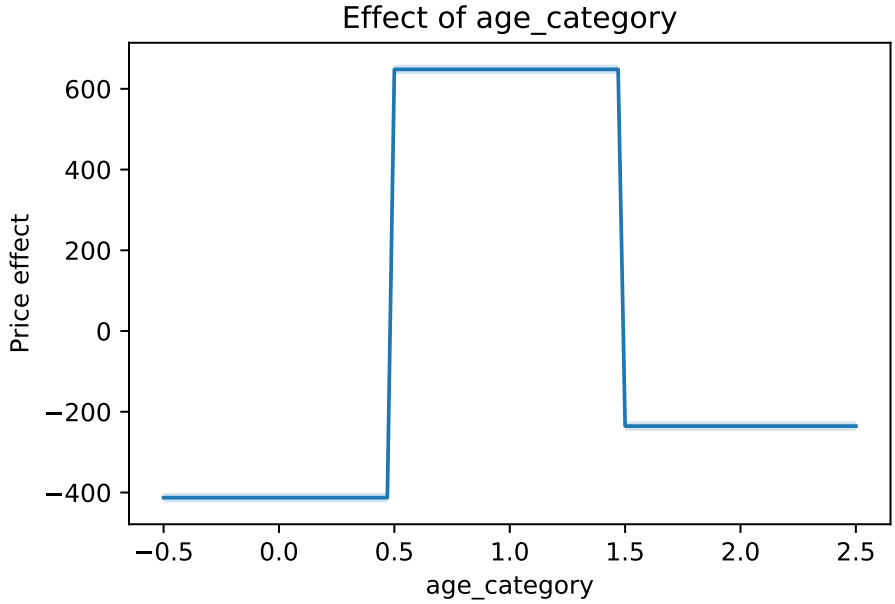
- Experiment: `baseline_gam.yaml`
- Model: LinearGAM
- Target: Price per m²

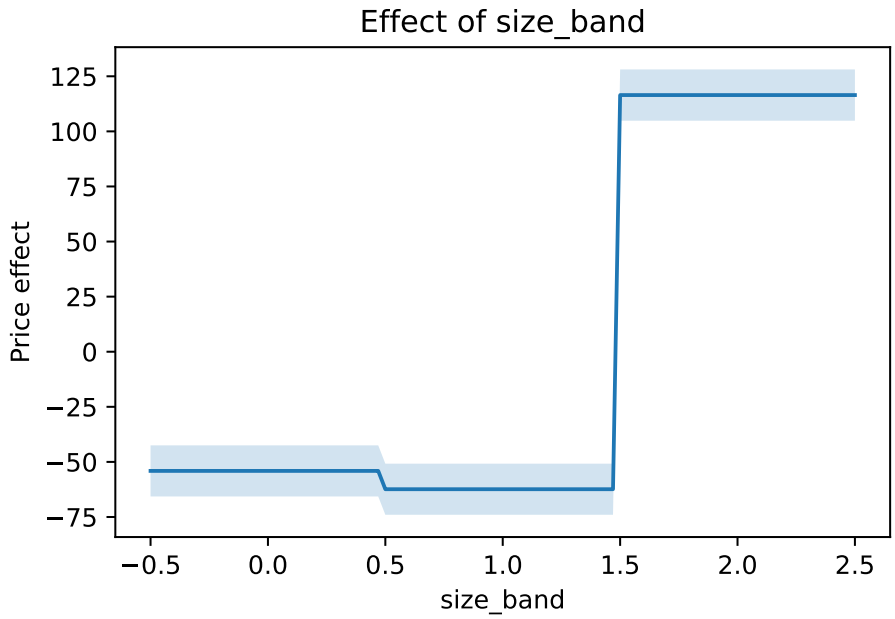
Data Loading

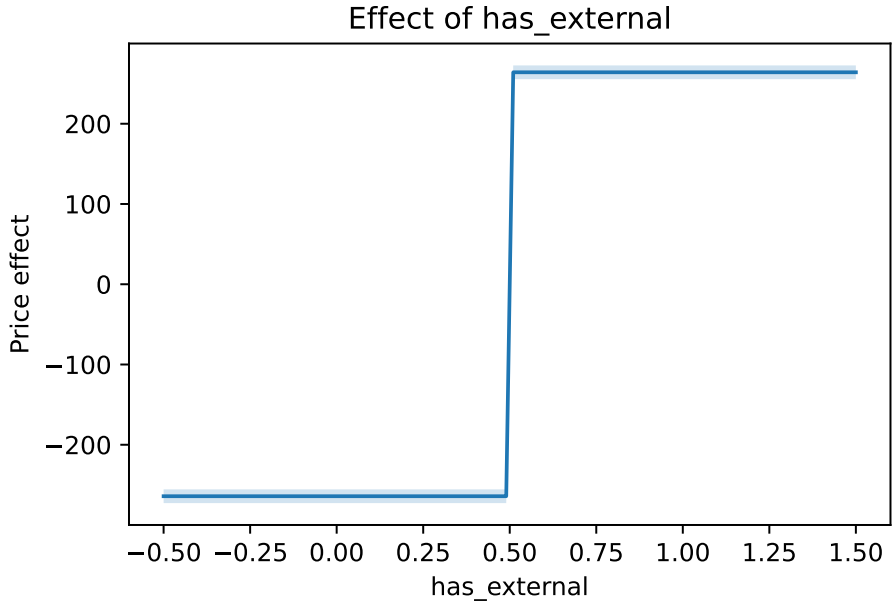
```
LinearGAM
============================================= ==========================================
Distribution:                        NormalDist Effective DoF:
Link Function:                     IdentityLink Log Likelihood:
Number of Samples:                        20502 AIC:
                                                AICc:
                                                GCV:
                                                Scale:
                                                Pseudo R-Squared:
=============================================================================================
Feature Function                  Lambda               Rank         EDoF         P > x
```

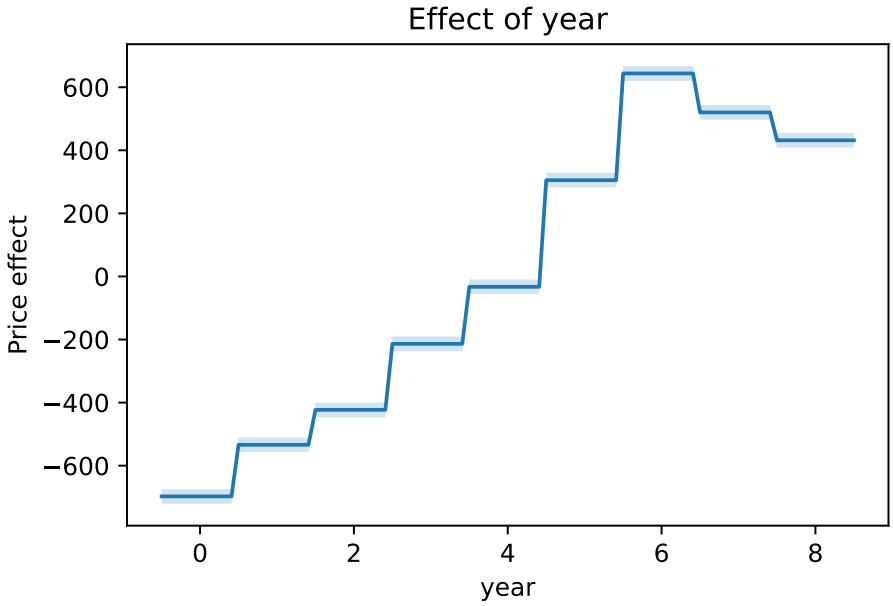
```
================================= ================== ============ ============ ==========
f(0)                              [0.6]              131          130.1        1.11e-16
f(1)                              [0.6]              3            2.0          1.11e-16
f(2)                              [0.6]              3            2.0          1.11e-16
f(3)                              [0.6]              2            1.0          1.11e-16
f(4)                              [0.6]              9            8.0          1.11e-16
intercept                                            1            0.0          1.11e-16
=============================================================================================
Significance codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

WARNING: Fitting splines and a linear function to a feature introduces a model identifiabili
         which can cause p-values to appear significant when they are not.

WARNING: p-values calculated in this manner behave correctly for un-penalized models or mode
         known smoothing parameters, but when smoothing parameters have been estimated, the
         are typically lower than they should be, meaning that the tests reject the null too
None
{'r2': OrderedDict([('explained_deviance', np.float64(0.8677604585353504)), ('McFadden', np.

/home/nuntiusten/dev/real-estate-gam-dbt/real_estate_gam/run_gam.py:53: UserWarning: KNOWN B

Please do not make inferences based on these values!

Collaborate on a solution, and stay up to date at:
github.com/dswah/pyGAM/issues/163

  print(gam.summary())
```

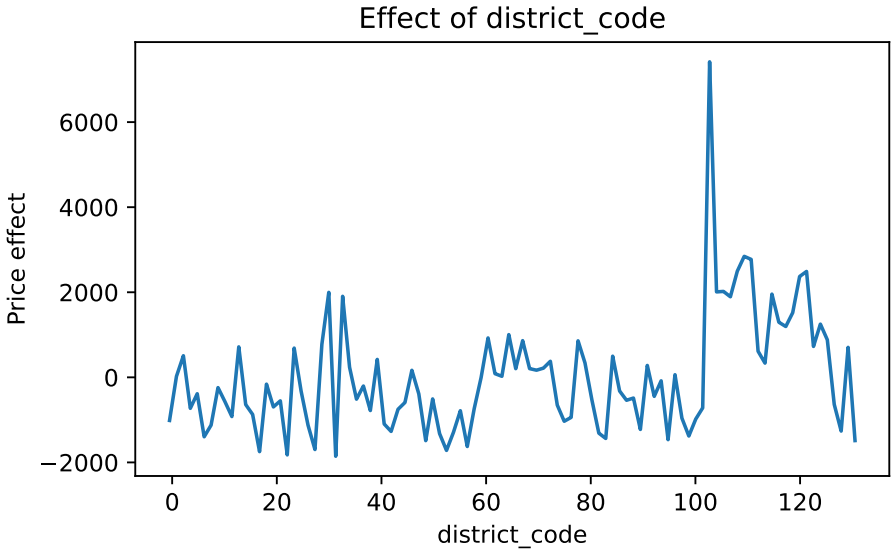

Effect of district_code
Price effect
6000
4000
2000
0
−2000
0
20
40
60
80
100
120
district_code

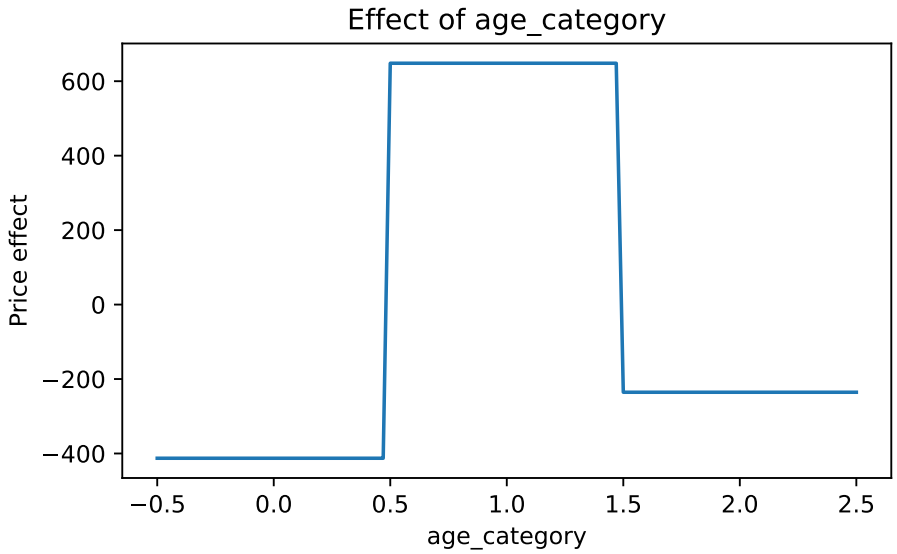

Effect of age_category
Price effect
600
400
200
0
−200
−400
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
age_category

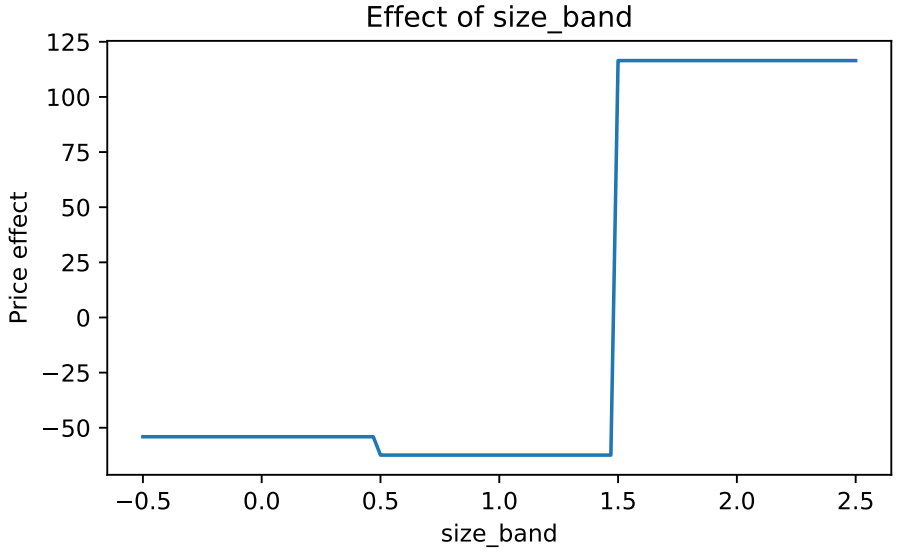

Effect of size_band
Price effect
125
100
75
50
25
0
−25
−50
−75
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
size_band

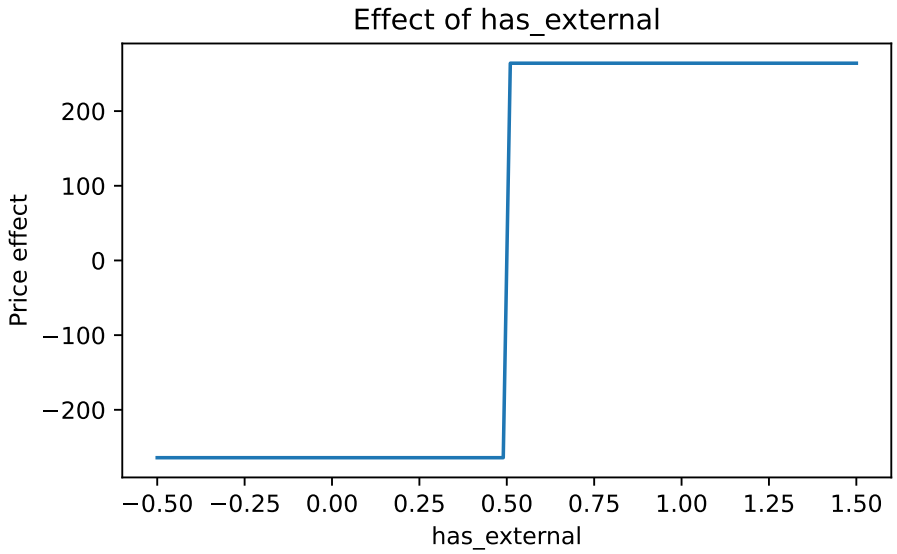

Effect of has_external
Price effect
200
100
0
−100
−200
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
1.25
1.50
has_external

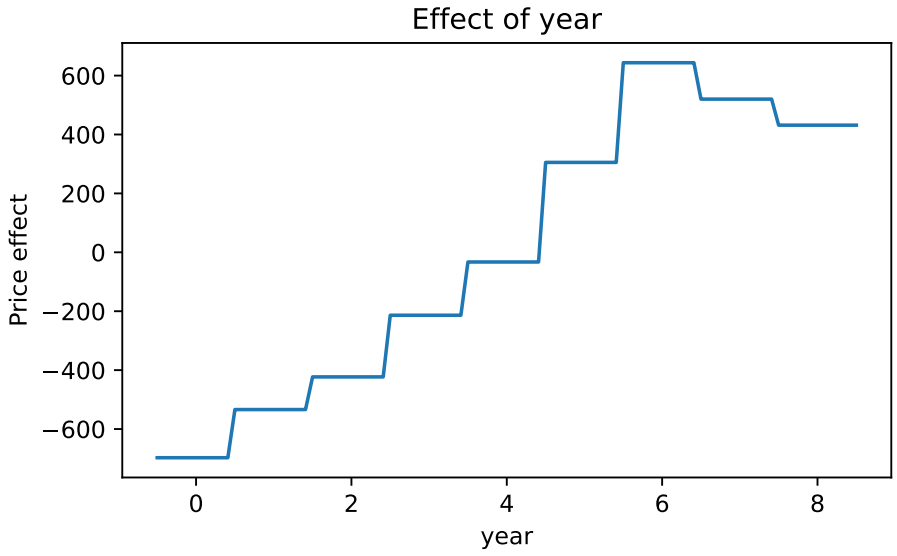

Effect of year
Price effect
600
400
200
0
−200
−400
−600
0
2
4
6
8
year


Effect of district_code
Price effect
6000
4000
2000
0
−2000
0
20
40
60
80
100
120
district_code


Effect of age_category
Price effect
600
400
200
0
−200
−400
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
age_category


Effect of size_band
Price effect
125
100
75
50
25
0
−25
−50
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
size_band





Key Takeaways

- The GAM captures strong non-linear effects.
- Categorical variables show clear price differentials.
- The model achieves solid explanatory power while remaining interpretable.

Reproducibility

This report was generated from a fully reproducible pipeline. To reproduce: python3 -m src.run_gam quarto render reports/gam_report.qmd

Source code and configuration are available in the public repository.