The truth about data science in real estate

From a classical use case to the everyday challenges

bulwiengesa

Dr. Federica Fusco

Women in Data Science
Regensburg, 14.05.2024

Agenda

Introduction

Use Case: RIWIS Prospect

Challenges

Final remarks

Agenda

Introduction

Use Case: RIWIS Prospect

Challenges

Final remarks

bulwiengesa inside.

In-depth data combined with experience, expertise and competence.
Independent. Objective. Excellent.

40 years of experience covering 4 market cycles for all sectors

More than 1000 variables from **over 30** sources

Over 10 million facts on more than **130,000** objects

60 professional researchers plus a large network of market experts

RIWIS perfectly informed through all investment phases.

Classical data science use cases

Price – purchase/rent – prediction

Property valuation

Forecasting (demographical and economical variables)

Location analysis with GIS

Agenda

Introduction

Use Case: RIWIS Prospect

1. Data
2. Model
3. Accuracy and limitations

Challenges

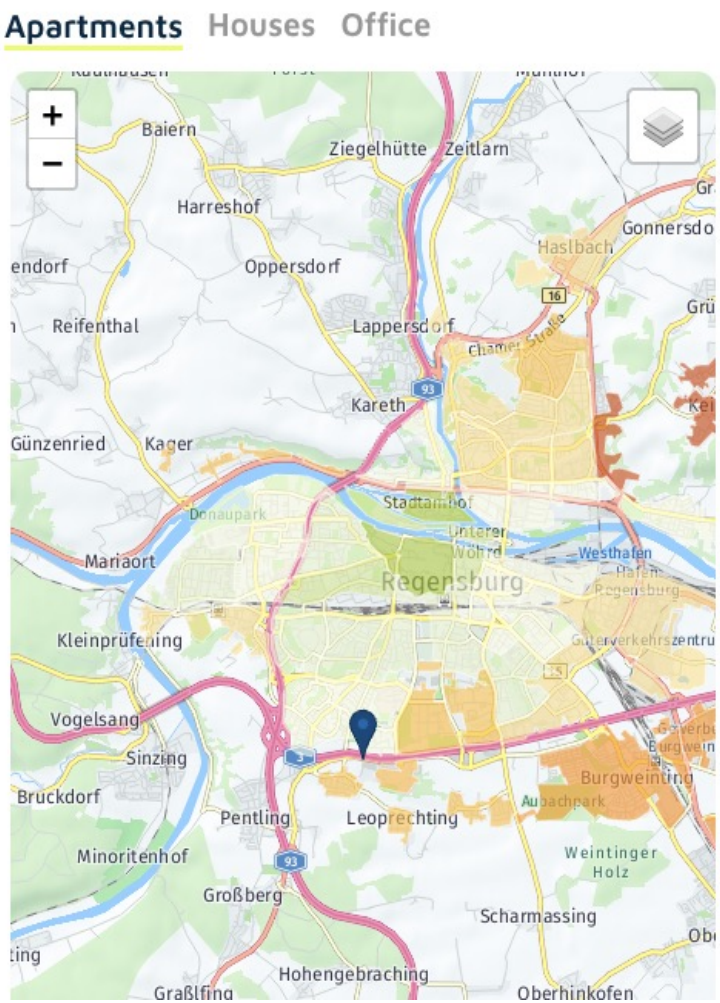
Final remarks

RIWIS Prospect

RIWIS Prospect determines *rental and purchase prices* of flats and houses in just a few seconds.

With this tool we can also estimate very large portfolios.

Prospect is currently available for *residential and office segments*. Logistic is work in progress.

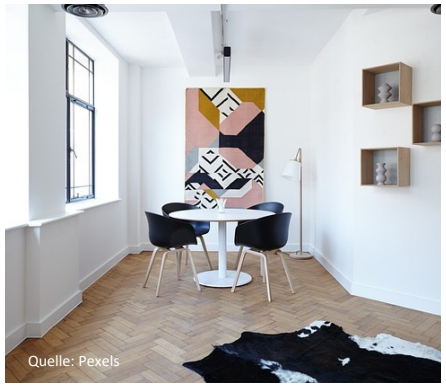


RIWIS Prospect – Data and features

Riwis Prospect uses ca. 10^6 property listings/year.

- Address

- Year of construtions
- ….



- Living space
- Floor
- …

Model

Prospect is a hedonic pricing GAM model

$$\log(\boldsymbol{Price}) = \boldsymbol{f}(\boldsymbol{W}) * \boldsymbol{X} + \boldsymbol{Q} + \varepsilon$$

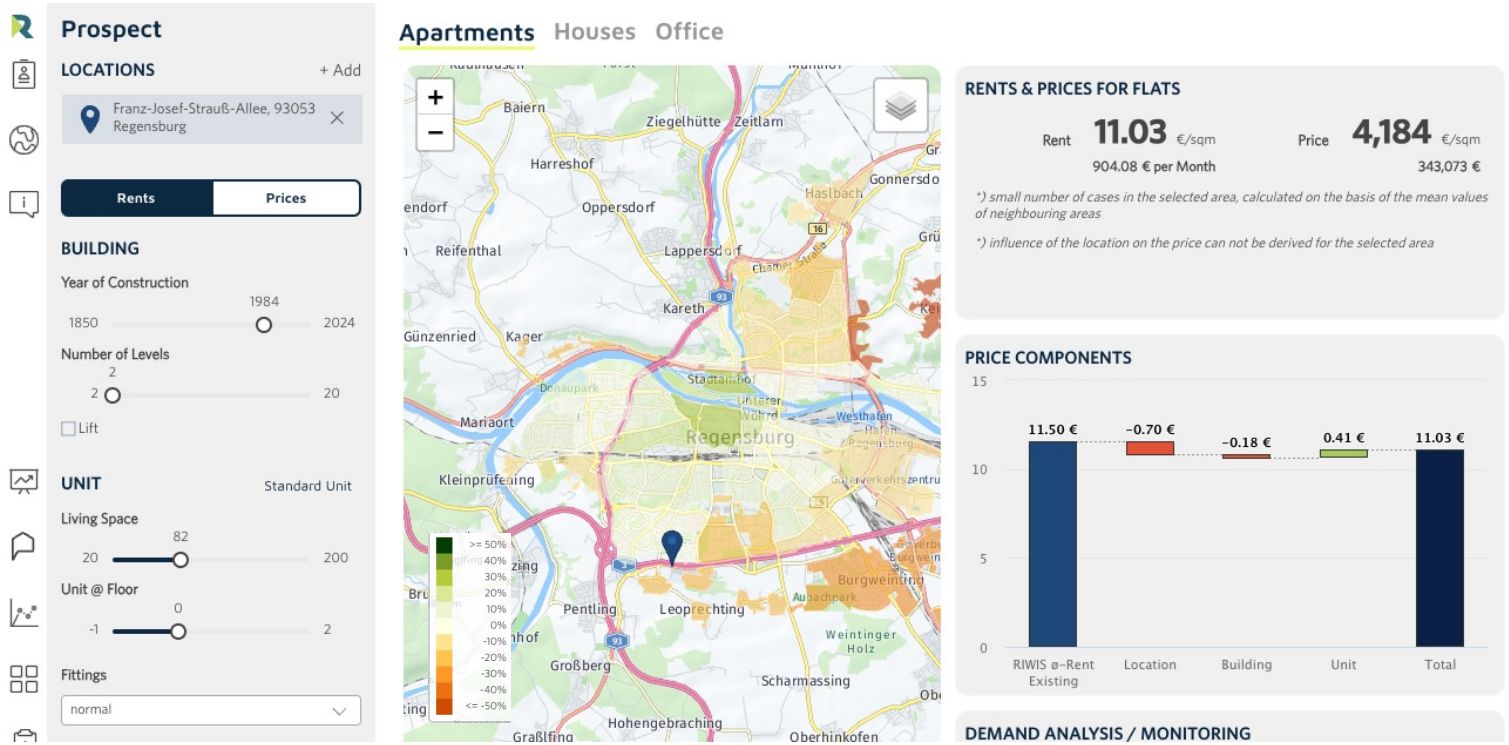
With:

- $\boldsymbol{X}$: features
- $\boldsymbol{W}$: weights
- $\boldsymbol{Q}$: intercept

A separate model is calculated for each district (*Landkreis*) on a quarterly basis.

- *Hedonic model:*
  - used for heterogenous classes of goods, such as buildings
  - Assumption: the composite good can be reduced to its constituent parts
- *log model* to get normal (or at least symmetric) distributed price data
- *GeoGAM*:
  - allows for smooth results within a polygon
  - quick to train, easy to understand and explain.

RIWIS Prospect



Geo polygons are *KGS12 entities* – within each of these the price cannot change.

Limitations and improvements

Limits of the model

- In areas with few listings, only aggregated statements are possible
- The current model has several *intrinsic* limitations, and it doesn't account for e.g., external factors, such as taxes and interest rates.

Improvements

- Feature engineering, again and again…
- Expand features with other data we have available
- Introduce new models (*xgboost*), which might be harder to interpret and handle

Agenda

Introduction

Use Case: RIWIS Prospect

Challenges

Final remarks

Challenges

Stakeholders are **often skeptical** about data science and ML methods – make them easier to access through e.g., ExplainableAI – and are worried about **legal issues.**

Purchase value of the properties is *not accurate*. In addition, some input variables cannot be foreseen (*causal inference*).

Many data sources are still **not digitalized** (e.g., *Bodenrichtwerte*), hosted in outdated systems or no API is available.

Some variables are available at **different geographical** levels, e.g., the unemployment rate.

Data science in real estate

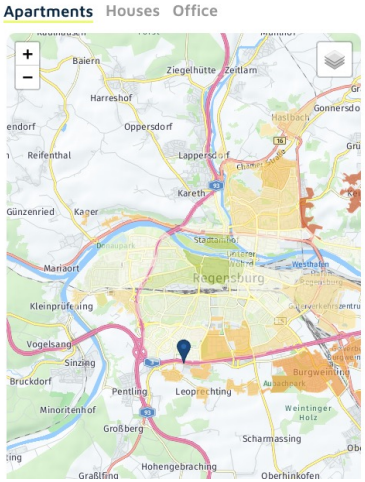
The Role of Data Science in Deriving Greater Insights into Property Market Dynamics

- 2022, CoreLogic Survey
- Only 42% of PropTech companies actively use data science to generate valuable insights to improve their business
- Main issues:
  - Fragmented data sources
  - Integration challenges
  - Lack of internal maturity in DS/ML/AI
  - Data scientist lack domain knowledge
- But the trend is clearly going in the *right* direction

Quelle: gif im Fokus, 1/2024, KI als Game-Changer in der Immobilienwirtschaft, ISSN Print 2198-6894, Online 2198-8013

CoreLogic Survey: The Role of Data Science in Deriving Greater Insights into Property Market Dynamics

Final remarks



Riwis Prospect

The geoGAM in background was presented, with its advantages and disadvantages.

Challenges

Data availability, data integration, acceptance of ML-methods

Key takeaway

Explain, teach, communicate, mentor about data science!

Quelle: gif im Fokus, 1/2024, KI als Game-Changer in der Immobilienwirtschaft, ISSN Print 2198-6894, Online 2198-8013

bulwiengesa

bulwiengesa AG
Nymphenburger Straße 5
80335 München
Telefon +49 89 232376-0
Fax +49 89 232376-76
info@bulwiengesa.de
www.bulwiengesa.de