


Applied Data Science

at Kodit.io 

Dec 2 2019

www.kodit.io

Nico Rotstein, CTO
nico@kodit.io

Nico Rotstein, CTO at Kodit.io

- PhD in AI
- 7 years of academic research
- 7 years (and counting) in startups
- 2.4 years at Kodit.io as CTO and co-founder

Kodit.io

- Provides convenience to home sellers and buyers
 - Sellers: speed, lease-back
 - Buyers: move-in-ready, self-tours
 - Both: trade-in, flexible moving dates
- Tech-powered lead spotting and qualification
- Under the hood: scalable buy-renovate-sell cycle
- Currently operating in Finland, Spain, Poland

Kodit.io Org Chart

- Management
 - operations
 - finance
- Real Estate
 - acquisition
 - renovation
 - sale
- Tech
 - product (data ⇔ product: requirements)
 - data (data ⇔ software: execution)
 - software

Kodit.io Tech Squads

- Product
 - head of product
 - data analyst
 - 33.33% tools expert
- Data
 - data plumber
 - machine learning engineer
 - two data scientists
- Software
 - architect
 - full-stack
 - back-end
 - front-end

How it works

Purchase Process

- Lead dashboard
 - Inbound
 - Outbound
- Lead qualification
- On-site inspection
- Qualification + inspection = Offer

Tech (Data) Value-add

- Purchase
 - Pricing
 - Renovation needs
 - Sale potential after renovation
- Renovation
 - Initially from inspection
 - Asset management tool
- Sale
 - Pricing
 - Marketing

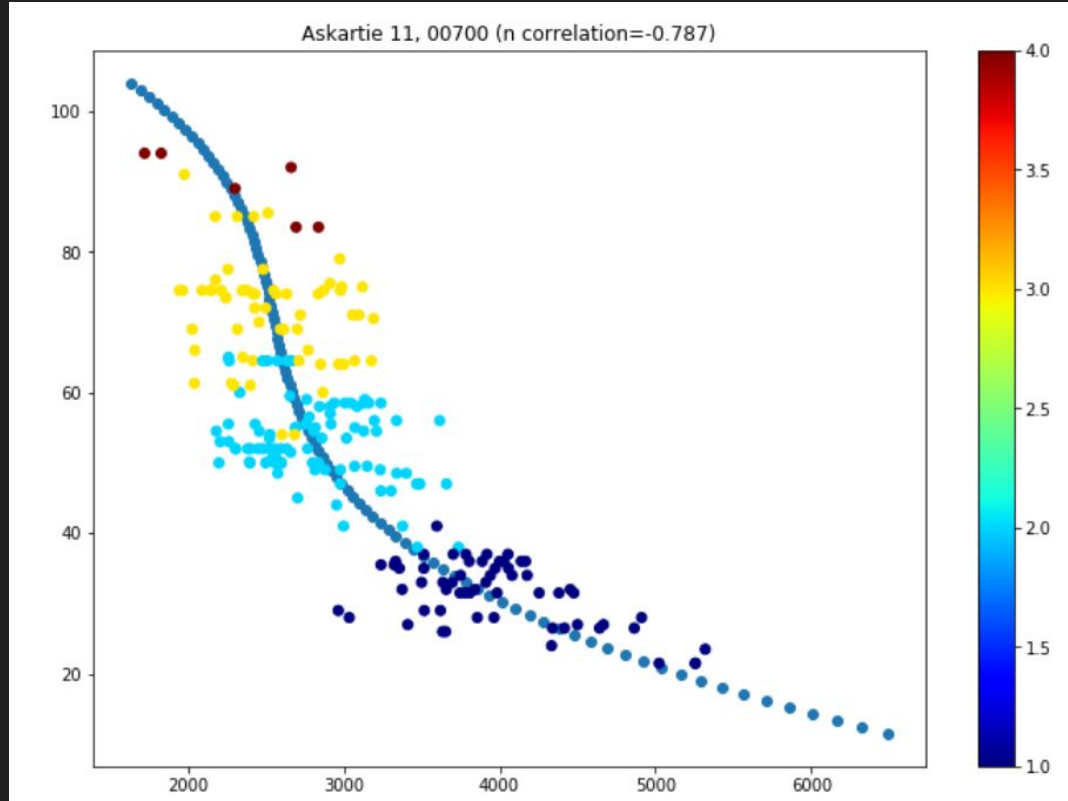
Big data vs Small data

- Big-data issues:
 - Store, clean and handle (if possible!)
 - Model training
 - Model inference
- Small-data issues:
 - Statistical significance
 - Similar-case coverage
 - Amount of non-predictions
- Kodit operates on small data

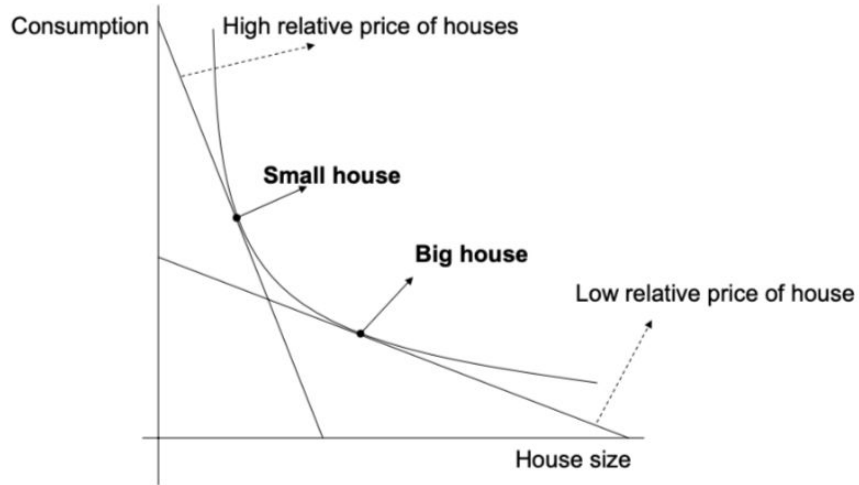
Neighbourhoods

- Helps in dealing with small data
- Similarity notion borrowed from real estate experts
 - Size +/- 10%
 - Within a radius
 - Building from the same period
 - Same number of rooms
- Result: neighbourhood size vs price/sqm curve

Neighbourhood Pricing Curve (failed attempt)

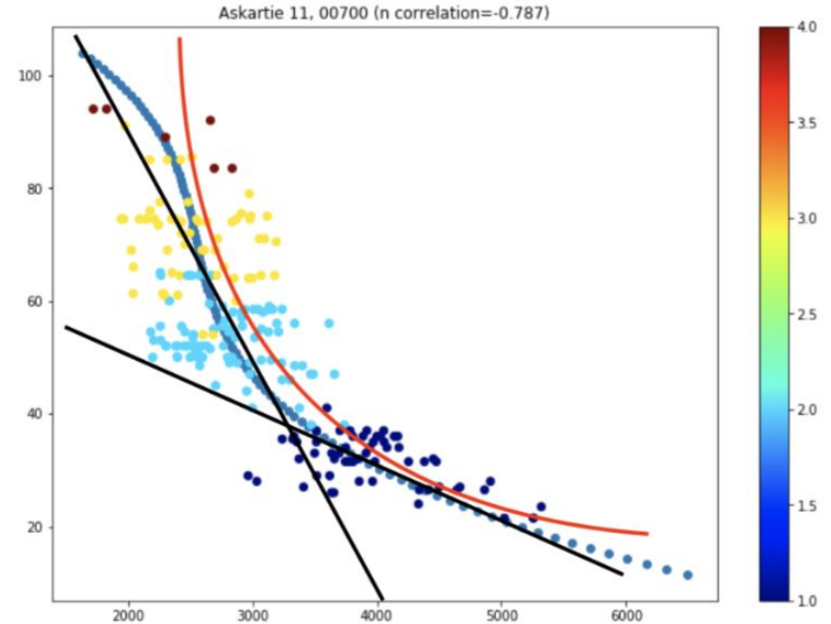


Housing Pricing (with substitution)

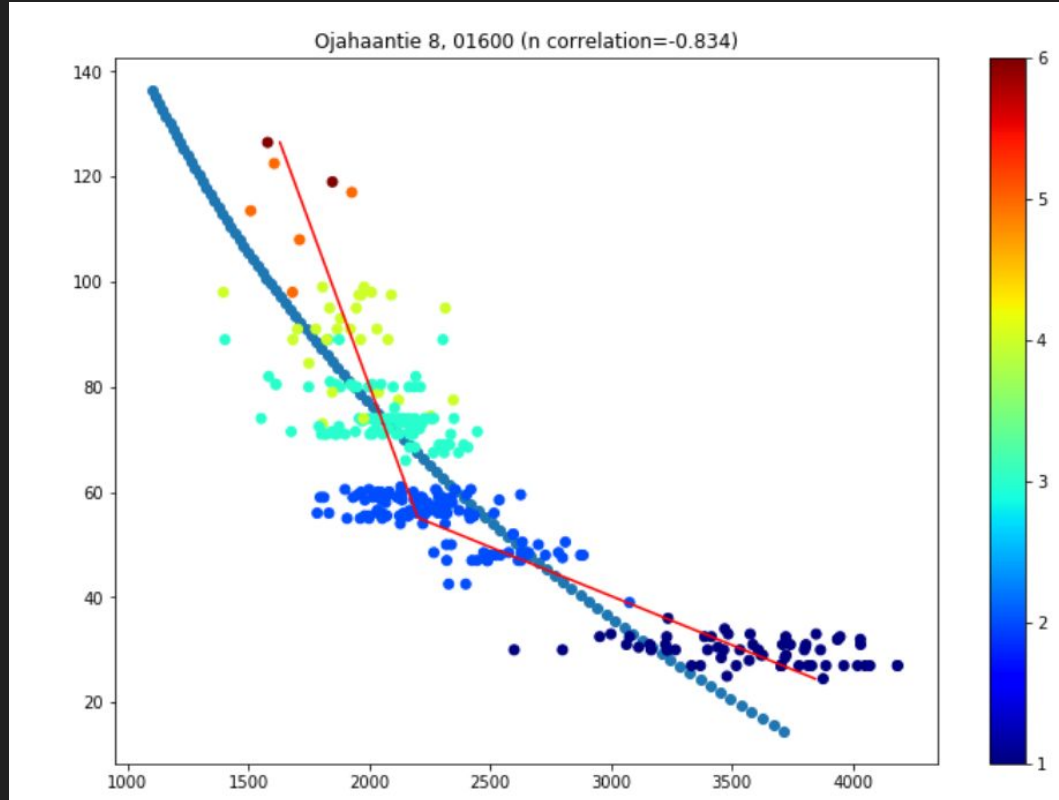


Econ 137 - Summer 2007

10



Neighbourhood Pricing Curve (involving expert)



Business Value From Data

- Intelligence augmentation
 - No one's losing their job
 - Everyone's super-efficient
- Prototype till you drop
 - Startup mentality
 - Improve or ditch
- Continuous biz↔tech conversation
 - Biz says what; tech says how; decisions are joint
 - Incremental improvements

Data-insured Business

- Pricing a home is highly subjective
- Data-based tools set processes and limits
- Predictive models and indicators set expectations
- Tools and indicators make for safer deals

Lead generation

Lead Qualification (1/2)

██████████ katu 30
33900 Tampere
██████████ · 2019-11-21T19:52:38

✓ Pipes
✗ Maint. fee (5.5)
✓ Condition

Features
55 m²
decent · 1960

Ad price
109 000 €
1 982 €/m²

Predicted price
110 000 €
1 993 €/m²

Ratio
0.99

OMG
6.6

Room count	2
Floor	3
Balcony	No
Lot ownership	No
Maintenance fee	303 €/mo (5,5 €/m ²)

Past renovations
2001-2001: radiator
2001-2001: heating
2007-2007:
2009-2009: staircase
2009-2009: paint

Coming renovations
2019-2019: heating

[More](#)

Address Matches

Notes

Contacted

[Discard](#) [Assign](#) [Handle](#) [Snooze](#)

UNASSIGNED
[Create new inspection form](#)

Lead Qualification (2/2)

Opportunity analysis

Pre-renovation price

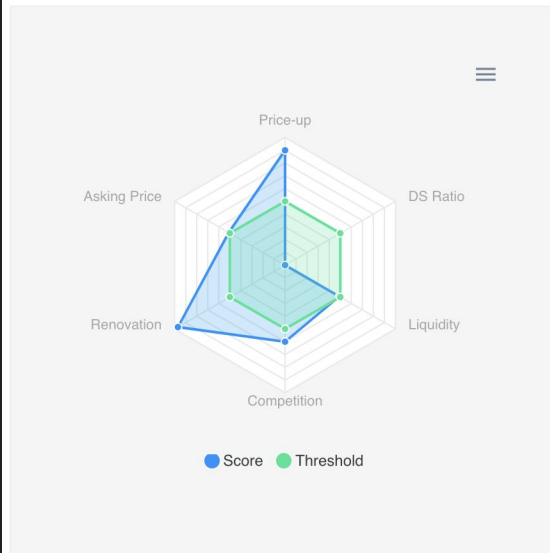
110000 €

1000 € (0.9%) price-up from listing price

Post-renovation price

138000 €

28000 € (25.5%) price-up from pre-renovation price



Price-up

9

DS Ratio

0

Liquidity

4.9

Asking Price

5.1

Renovation

9.7

Competition

6

Total

6.6

Lead Qualification Example 2

Opportunity analysis

Pre-renovation price

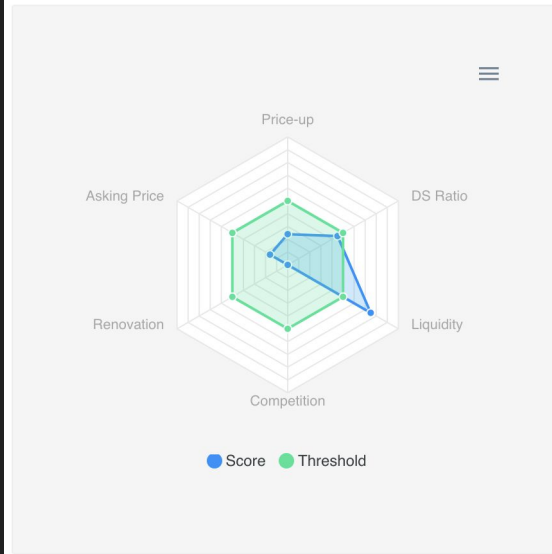
157000 €

22000 € (12.3%) under listing price

Post-renovation price

157000 €

-



Price-up

2.4

DS Ratio

4.5

Liquidity

7.5

Asking Price

1.6

Renovation

0

Competition

0

Total

3.2

Lead Qualification Example 3

Opportunity analysis

Pre-renovation price

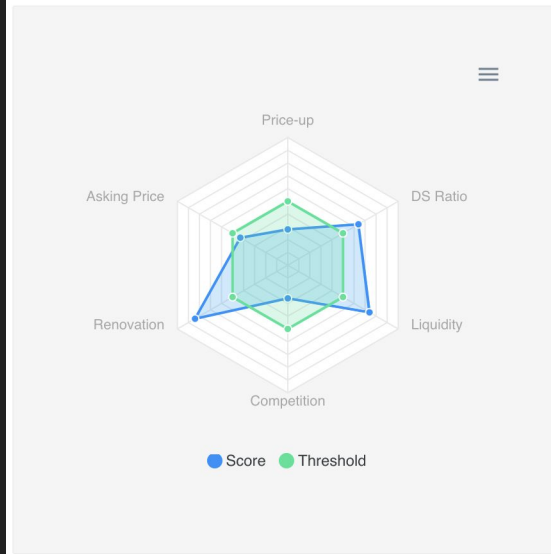
165000 €

5000 € (2.9%) under listing price

Post-renovation price

168000 €

3000 € (1.8%) price-up from pre-renovation price



Price-up

2.8

DS Ratio

6.4

Liquidity

7.4

Asking Price

4.3

Renovation

8.4

Competition

2.6

Total

5.5

Models vs Indicators

- Not everything can be predicted
 - Lack of (local) data
 - Noise
- Enter indicators
 - Plain number crunching
 - Somewhat advanced statistics
 - Biased-to-safety predictive models
- Point predictions could mislead where a range would help

Predictive Model Sample

Predictive Models

- Each market has a number of data sources
- Most relevant: real estate transactions and listings
- Each data source means a new price predictive model

Predictive Model Sample (1/3)

Features

- latitude-longitude
- size
- bedrooms
- bathrooms
- floor
- neighbourhood
- condition
- rating_leads (contacts to ad)
- rating_visits (visits to ad)
- garage
- garden
- sports
- storage
- terrace
- swimming_pool
- air_conditioning

Engineered features

- distance: to the city centre
- distance_metro: minimum
- KNN: latitude, longitude and size

Predictive Model Sample (2/3)

Model creation

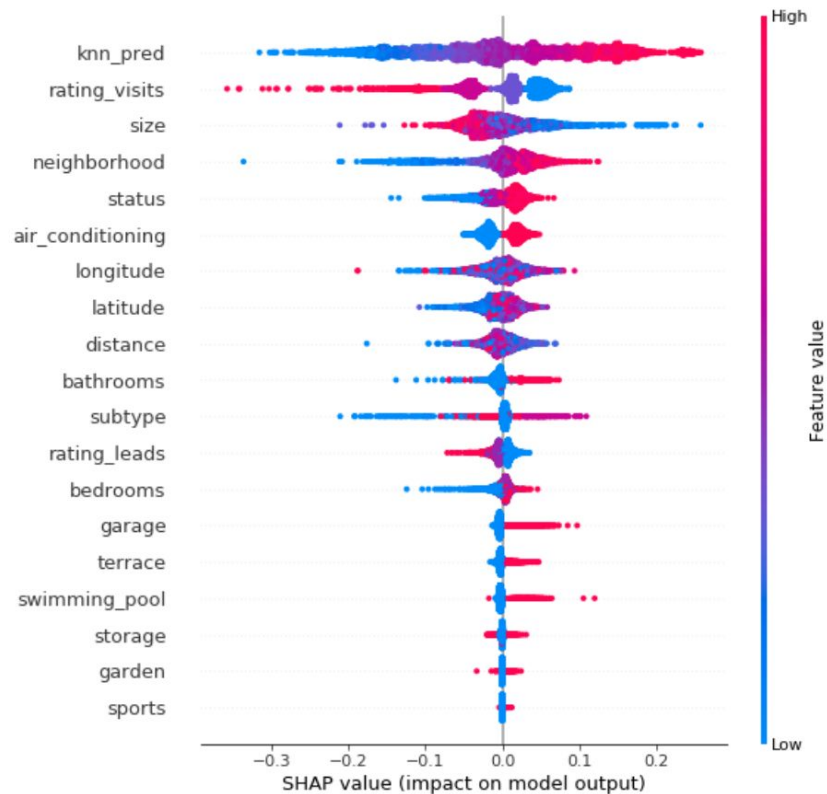
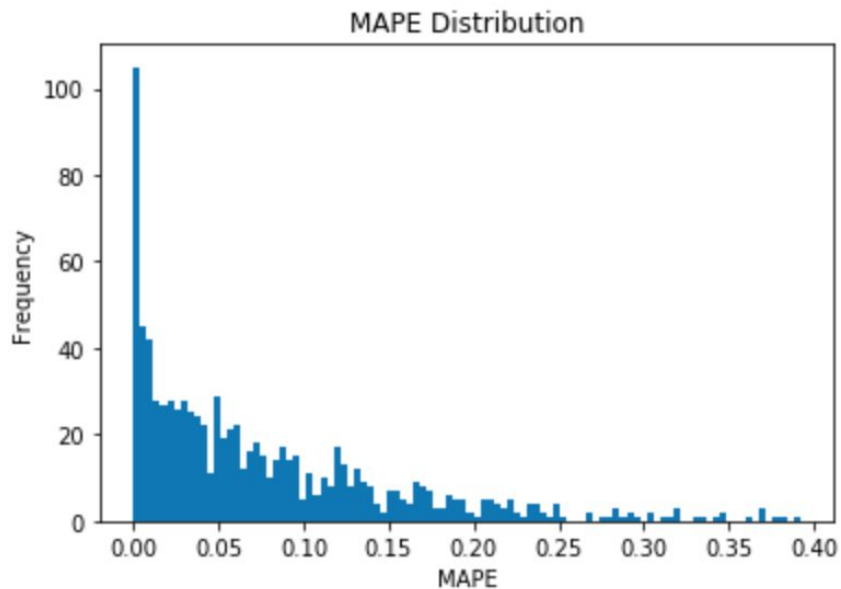
1. Input data
2. Filter and clean
3. Feature Engineering
4. Preprocessing
5. Train
6. Evaluate
7. Interpret

Preprocessing

- MinMax normalisation for numeric variables
- CatBoostEncoder for categorical variables
- Log transformation for the target

Predictive Model Sample (3/3)

Median Absolute Percentual Error: 7%



What's next?

The Data-based Future at Kodit.io

- Instant offers
 - Self-aware predictive models
 - Self-service home-buying and -selling
 - Additional Kodit.io services
- Marketplace
 - Matchmaking buyers and sellers
 - Kodit services within the platform
 - Data-powered everything: Kodit.io as facilitator

Real Estate and Technology

- Transparency
 - More open data
 - Free home valuations
- Convenience
 - Electronic everything
 - Self-service
- New models of ownership

Thanks!