

Sherlock Homes

London Real Estate Development



London Real Estate Development

Company & Project Overview

Company
London
Project

Data

- Dataset
- Data Cleaning
- EDA - Exploratory Data Analysis

App

- ML models
- Final model: Watson v1.0
- App - Demo

Conclusions

- Conclusions & Learnings



Company & Project Overview

The Company

Focused on B2B & B2C
Construction management & Planning
Real Estate valuation & portfolio management

Sherlock Homes is a Real Estate Developer that combines Digitalization, Construction and Real Estate data-driven focus.

Worth £2.64 trillion, London real estate has shaken off the 2023 correction to stabilize in 2025. Average prices now sit at £652,000, proving the capital's enduring value despite slower growth.

We are in the heart of London and its market, but you won't find us at 221b-Baker Street, though.

The Project

Predictions & Recommendations

We sneake through London to find the best insights on Real Estate. Our model Watson V1.0 predicts dwellings' prices and finds out improvements based on the actual market, ensuring that our clients will get the best opportunities out of their properties.



Data - Datasets & Data Cleaning

Datasets

London House Price Data
from Kaggle
418,201 rows × 30 columns

Features nominal

- postcode ←
- outcode
- property type
- Tenure

Features ordinal

- Energy rates

Features numerical

- | | |
|------------------------|----------------|
| • latitude & longitude | • Living rooms |
| • bathrooms | • Area sqm |
| • Bedrooms | |

Target numerical

- Sold price

London Building Stock Model
from London Datastore
3,887,534 rows × 63 columns

Features nominal

- postcode →
- Conservation Area
- Building range years

Data Cleaning

Major data cleaning
50,000 rows after cleaning

Data > 2023

80,000 rows after cleaning

Null values management:

- Missing rooms replaced by median of similar houses
- Tenure: Unknown
- Energy rates: Not rated

Dropped duplicated by address

New aggregation columns

- Neighbourhood median £/m²
- Size buckets, m²



Data - Features Correlations & ML scope

Correlations

VIF insights + heatmap

✗ Exclude Leasehold

- Tenure Freehold and Leasehold > 18

Pearson r & R² tests

↗ Increasing price impact:

- Floor Area
- Bathrooms quantity
- Bedrooms quantity
- Living rooms quantity
- Neighbourhood value
- Conservated area

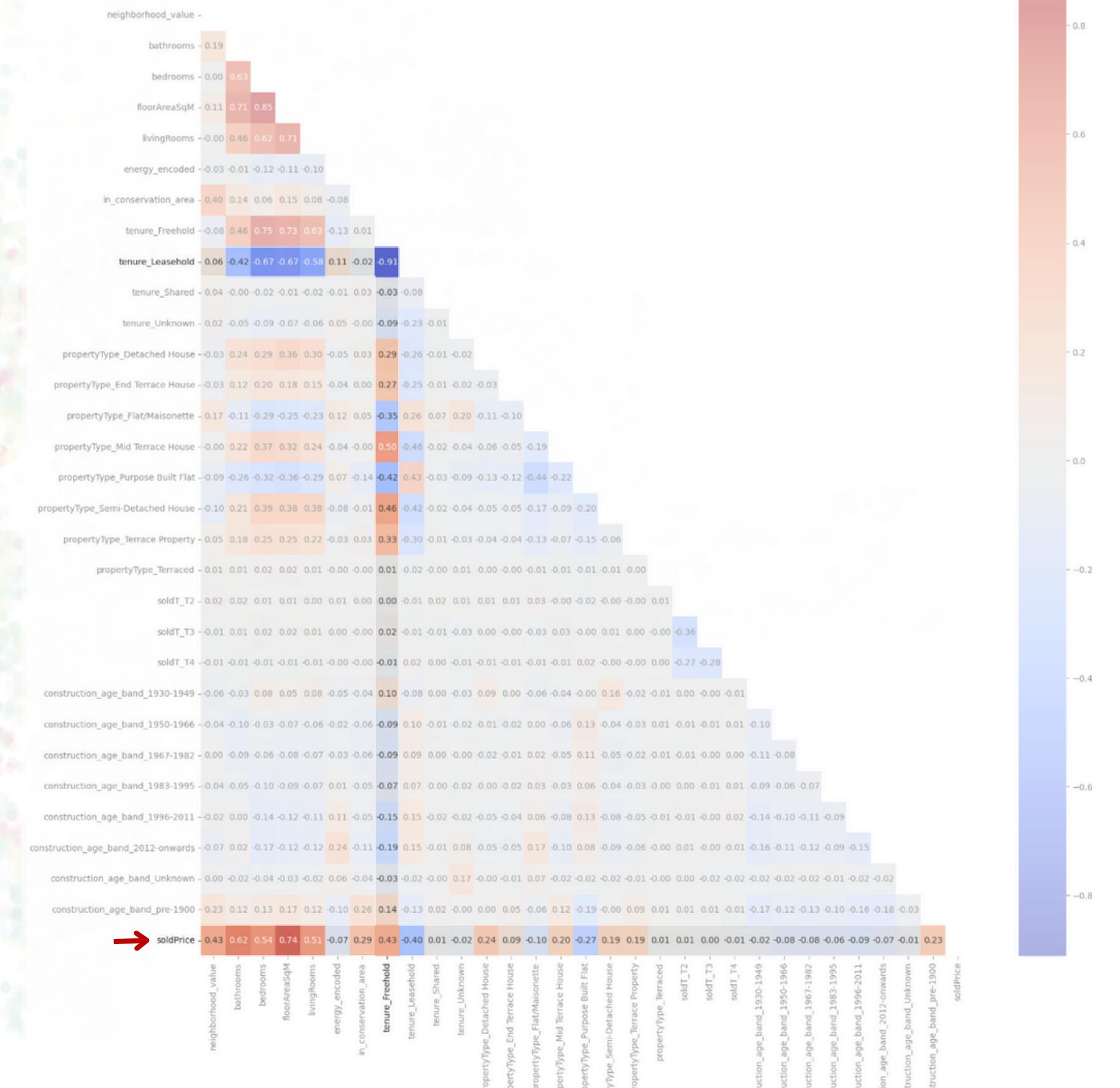
↘ Decreasing price impact:

- Purpose Built Flat
- Energy rate

✗ Noise reduction on data

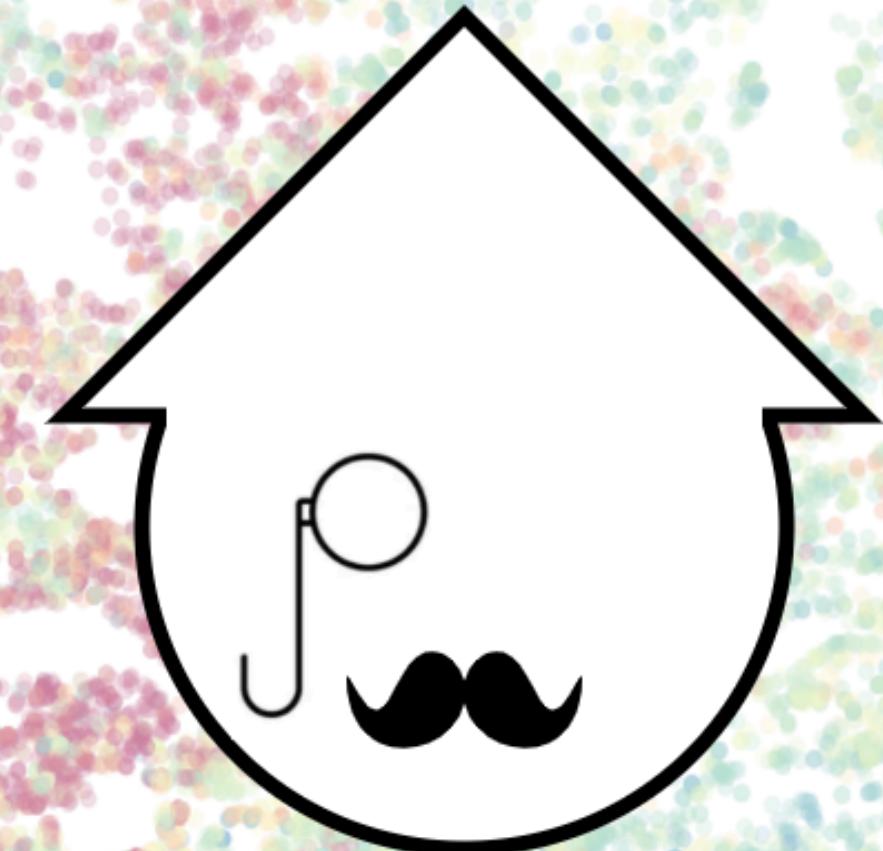
- Selling trimestres & years. Small period of time

Correlation Heatmap (Encoded Features)





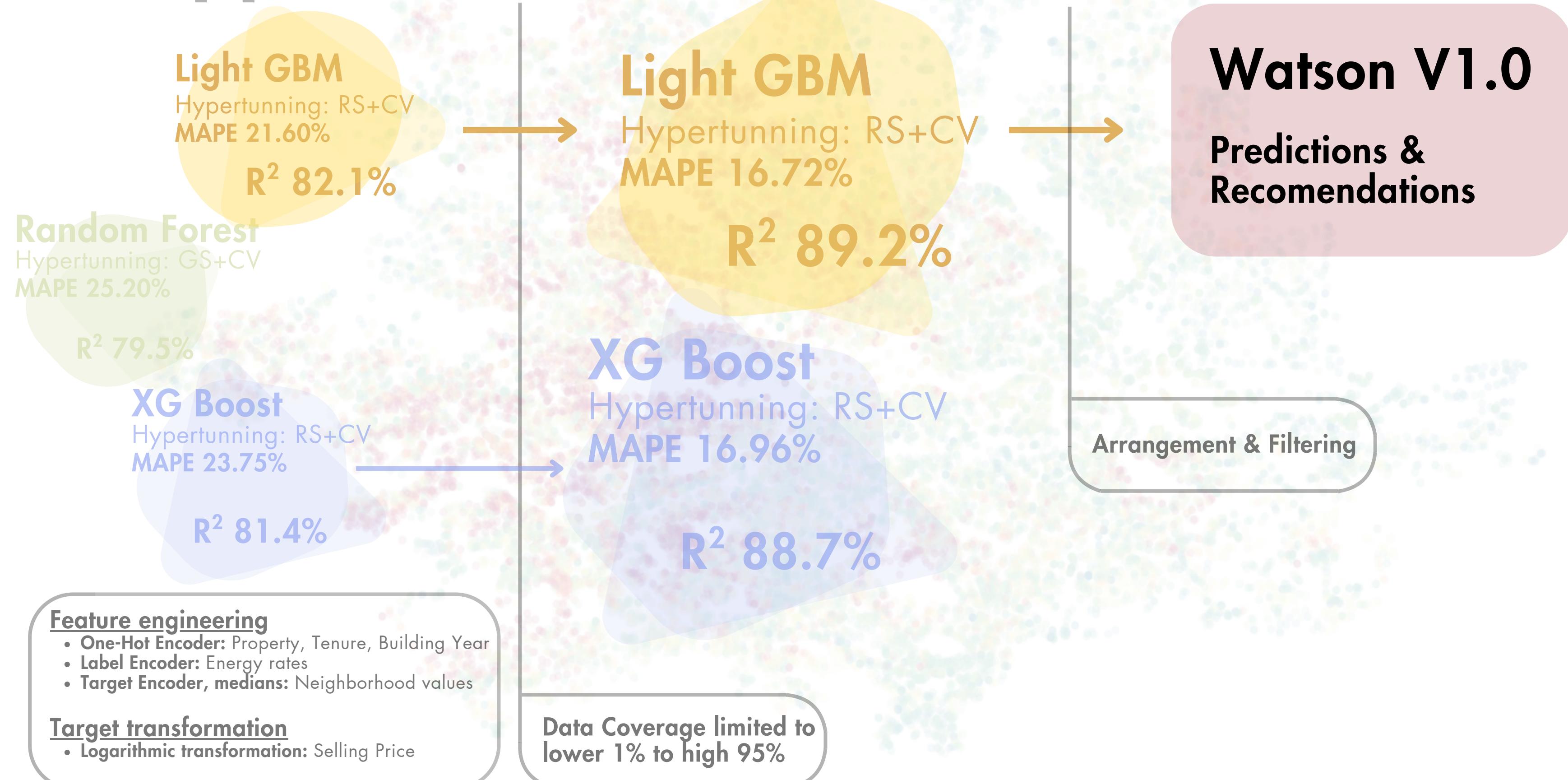
Data - Data Description



**Sherlock Homes
in Tableau**



App - Watson v0.0 to v1.0





App - Watson 1.0

User input → Prediction → Recomendations

Property Details

- postcode
- Bedrooms
- bathrooms
- Living Rooms
- Property Type
- Construction Age
- Tenure
- In Conservation Area
- Energy rate
- Desired rate: customazion of enery rate calculation

Calculates the dwellings' price

Confidence Interval user

- 25% - 75%

Confidence Interval Company

- 10%-90%

Calculates 2 hypothesis

Alternative recommendation

1. checks for all feasible layouts
2. checks for the highest return

Sherlock's recommendation

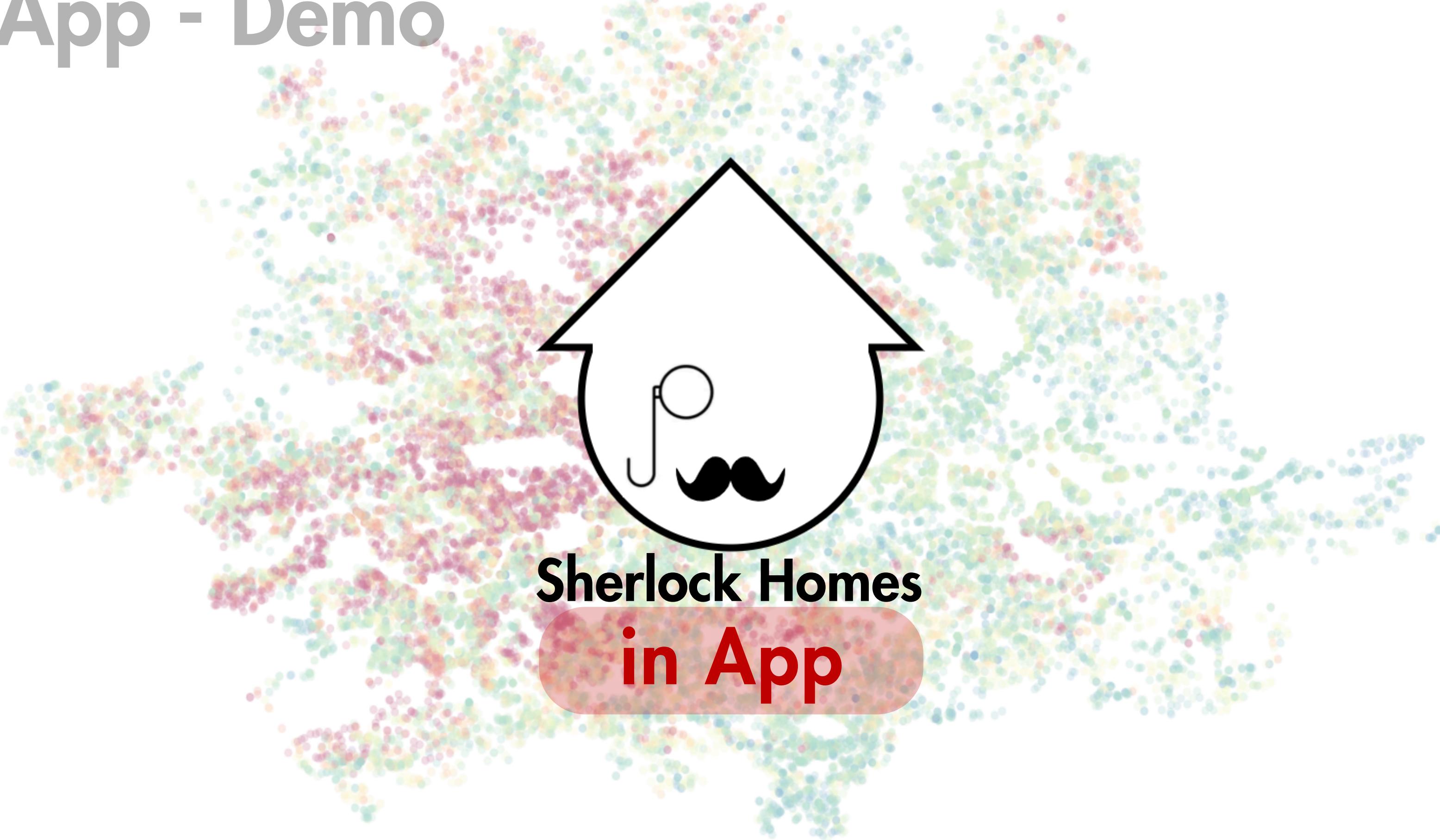
1. checks for all feasible layouts
2. checks for improved and common layouts and EPC within the 50 nearest peers.
3. checks for the highest return layout. If EPC is worse than C, uses C. If it is C, recommends improvement.

Recomendation system

1. Only shows recomendations which prediction is 10% higher than actual price
2. If Sherlock's > Alternative, shows Sherlock's only
3. If Sherlock's < Alternative, shows both



App - Demo



**Sherlock Homes
in App**



Conclusions & Learnings

Real Estate is a complex industry to predict, even for Sherlock Homes

Several factors impacting prices, which makes **similar assets with very different prices**.

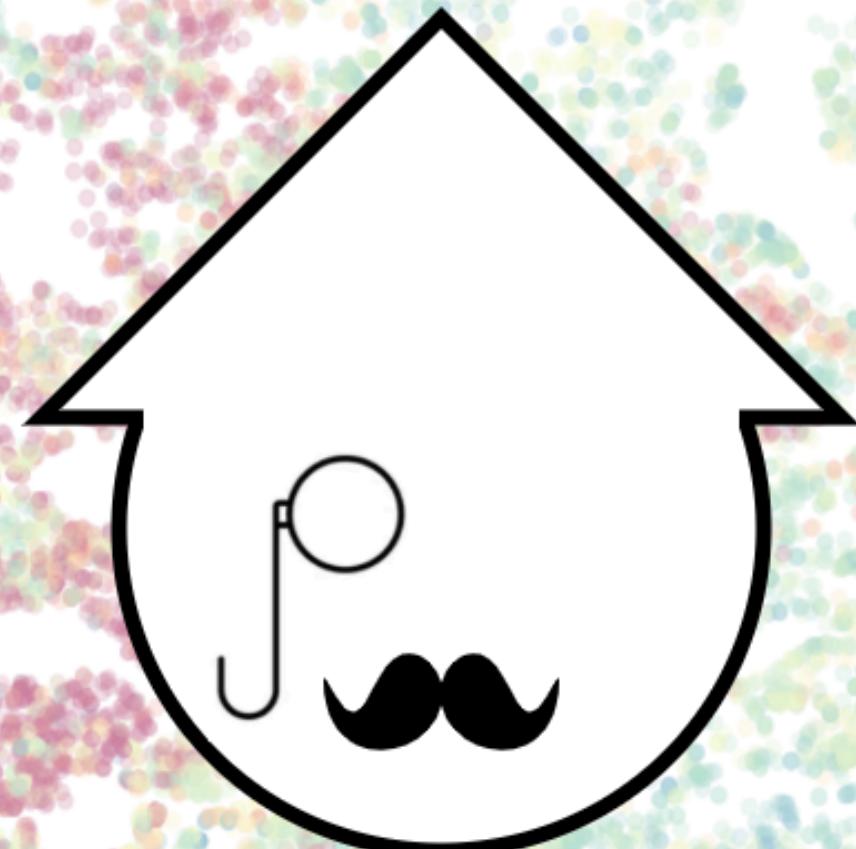
A little bit more of Data is never too much Data

A lot of data is private, specially the one more focused on details. That data would have been important to **decrease the range of the confidence interval**

Outliers: Adress the Real Estate subindustries

To better perform the analysis and predictions we should address the subindustries, **market distribution highly skewed**.

Although in this project we focused only in housing, a **more specific approach** could have been made for the **very low-priced market and the luxurious segmented market**. For example, with 2 ML models specialized on it.



Sherlock Homes
Thank you!