Project Proposal
Real estate value prediction

Author: Arthur Brouwers

Date: 30/10/2023

Version: 1.1

Table of Contents

Introduction

For the personal challenge in semester 4 of AI, I had to choose someone who is an expert in his or her work. The personal challenge is to write a predictive AI that is useful to your expert. I chose Peter Brouwers as my expert. Peter is an established real estate agent in the area who is enthusiastic about the future of AI. Peter came up with the idea to predict the value of houses and how long they will be for sale with the help of predictive AI. Since Peter is an active real estate agent in and around Den Bosch, I will focus the model on around this place. Predicting the value of the house would benefit Peter with saving time, since right now he does this by hand. Predicting how long the property will be for sale will help both Peter and his customers. It would help Peter in busier times in the real estate market to consider whether putting the house up for sale is even worth his time. On the other hand it would help Peter's customers because Peter can give them an accurate estimate on how long their property will take to be sold.

Domain Understanding

Research

I did my own research about requirements to accurately determine the house price.

First, I asked ChatGPT what the needed requirements are and this was the answer:
To make an accurate prediction on home appraisals and listing duration, I need the following data:
- Features of the house
    - Amount of (bed)rooms
    - Living area and plot area
    - Year of construction of the house
    - Presence of amenities such as a swimming pool, garage, etc.

- Characteristics of the neighborhood
    - Location (e.g. city, neighborhood)
    - Crime rates in the area
    - Proximity to schools, shops, hospitals and public transport
    - Average income and employment rates in the neighborhood

- Historical data
    - Historical sales prices of comparable homes in the area

- Economic factors
    - Interest rates
    - General economic conditions, such as unemployment rates and inflation

- Demographic data
    - Age distribution of the population in the area
    - Family size and household composition

- Time-related data
    - Seasonal trends and market fluctuations
    - Date of listing (when the house was put up for sale)

After asking ChatGPT, I did my own research on the topic and this is what I found:

Determining factors and actors

The price of the houses are determined in an interplay of three markets:

- The housing market.
- The financing and investment market.
- The construction and land market.

The price of houses is strongly determined by the financing options of residential consumers. Not just the income, but most importantly the interest rate plays an important role in this. In addition, the rules for lending by financiers and the mortgage interest deduction are important.

You can view the housing market as a stock market. The additions to the stock in the form of new construction generally mean only a marginal addition to the existing stock. The prices in the existing stock are therefore leading for price development in new construction, especially in regions where there is tension housing market. The difference between house price and construction & land costs are reflected in the residual land price.

The housing market is not a national, but regional market. The weight of the factors may differ between different regions and sometimes even locally. [i]

What determines the value of a house?

Location: This is one of the most important points. The neighborhood is not taken into account necessarily, but at access to services and networks. For example, it matters whether the house is located near a grocery store or the view.

Maintenance condition: The condition of the house also has impact on the value of the house. For example, a house in a bad state and in need of renovation will have a lower value than a house that is ready to move in.

Reference properties: When determining a price, you should also take prices of houses in the neighborhood into account. The sales value of these so-called reference properties affects the value of your home.

WOZ value: This is included because it gives an indication of the value of a house. This value is determined by municipalities based on location and surface area data. The WOZ (Waardering Onroerende Zaken) value is not always 100% reliable as the reference date is always one year behind.

Energy label: A study by Tilburg University has shown that a home with energy label A or B yields an average of € 6,000 more and is for sale for more than a month less than homes with a lower energy label (F or G). This shows that higher energy labels are more in demand. [ii]

The house type: The type of the house also matters. For example, the price of an apartment may differ from a detached house because of multiple different factors.

The living area: The living area of the house (in m2) will also impact the price; the bigger the area the higher the price.

Construction year: The construction year affects the price because of multiple reasons[iii]:
- Age and condition: Older houses tend to require more maintenance and renovations than newer houses, so this ties in with previous mentioned maintenance condition.
- Aesthetic trends: Architectural design trends change over time. Buyers often have preferences for particular styles and houses built in that particular style may command a higher price.
- Energy efficiency: Newer houses are often more energy efficient. This could be because more advances in construction materials and techniques.
- Building Codes and Safety Standards: Building codes and safety standards evolve over time. Newer houses are typically built to more current safety and code standards.
- Infrastructure: The year a house was built is often linked to the infrastructure. Newer developments may have better transportation links, schools, parks and shopping options.
- Historical and cultural significance: In some cases, older houses may be considered historically or culturally significant, which can increase their value. However, this is often the exception rather than the rule.
- Depreciation: Houses depreciate over time. This is a factor in real estate appraisal. The age of a house can affect its depreciation rate and, consequently, its value.
- Market Perception: Buyer and seller perceptions play a role in the impact of the construction year on the house price. Older houses may be seen as having more character and charm, while newer houses are often seen as more convenient and low-maintenance.

Interview

After doing my own research, I interviewed Peter to get his view on how to predict house prices and how he feels about my research.

After the interview, I will combine the gained knowledge together with my research to come up with data requirements.

This is what I got out of the interview:

The important features from most to less important:

1. Location, even neighborhoods
2. Type of house, such as apartment, terraced house or semidetached house
3. Living space area and lot area combined
4. Energy label

There are also other types of area's:

- other indoor space
- Building-related outdoor space
- External storage space
- finishing level of a house

What was a surprising thing to me is that the amount of rooms is less decisive, however it does have impact on the type of buyer, but this is hard to take into account when determining a price. Also, the year of construction should not necessarily be taken into account if the energy label is present. Since energy labels are mandatory, I should not really focus on the year of construction.

External factors to consider are macro-economic circumstances:

- Mortgage interest
- Economic downturn (bear market)
- Inflation
- Employment and Income Levels
- Demographic Trends

Peter also told me it would be a nice challenge to see If I could create a model that predicts what impact sustainable on the price. It is not the priority, but it would be a nice challenge.

I also showed Peter the dataset I found and he said he is satisfied with the data given. I should be able to come up with a model based on the data I have found.

Data Sourcing

I found this dataset for house features: https://www.kaggle.com/datasets/bryan2k19/dutch-house-prices-dataset

The dataset definitely needs some cleaning, but the data is more than useful for my project.
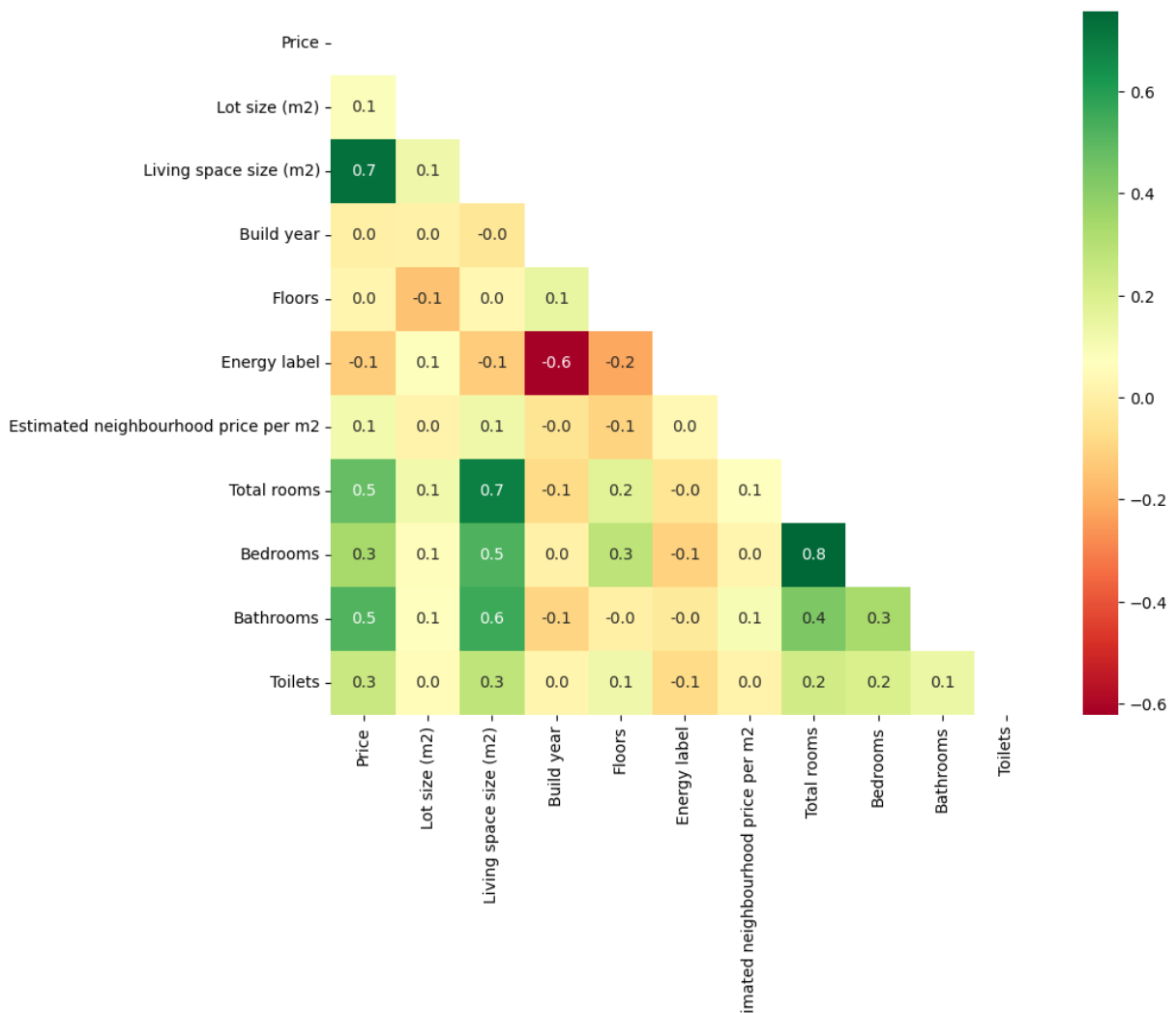
Sources

[i] https://www.tweedekamer.nl/sites/default/files/field_uploads/120912%20PBL%20overzichtspaper_tcm181-232761.pdf

[ii] https://walterliving.com/blog/wat-bepaalt-de-waarde-van-een-huis/#waardehuis

[iii] https://www.obvion.nl/blog/Eigen-huis/taxatie-huis#:~:text=Dat%20doet%20hij%20aan%20de,zijn%20bepalend%20voor%20de%20waarde